

A case for a negative-strand coding sequence in a group of positive-sense RNA viruses

Journal:	<i>Virus Evolution</i>
Manuscript ID	VEVOLI-2019-067.R1
Manuscript Type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Dinan, Adam; University of Cambridge, Department of Pathology Lukhovitskaya, Nina; University of Cambridge, Department of Pathology Olendraite, Ingrida; University of Cambridge, Department of Pathology Firth, Andrew; University of Cambridge, Department of Pathology;
Keywords:	RNA virus, overlapping genes

SCHOLARONE™
Manuscripts

We thank the reviewers and editor for their careful assessment of our manuscript and for their kind and helpful comments and suggestions. Please find below our detailed responses.

Associate Editor: Darren Obbard

I agree with the reviewers, and I enjoyed reading the manuscript very much. However, in addition to those points raised by the reviewer, I would also like to ask for a couple of clarifications or small additions.

(1) Narnaviruses classically lack a virion. However, is there experimental evidence for this in any narna-like virus infecting non-fungi, or is this just a generalization from the *saccharomyces* viruses?

To the best of our knowledge, this has not been experimentally investigated outside of the two *Saccharomyces cerevisiae* narnaviruses. The lack of an obvious capsid coding sequence (except in ourmia-like viruses, as discussed in the ms and in Dolja & Koonin, 2018, PMID 29103997) suggests that they do lack a virion, but one could envisage other scenarios, such as a helper virus providing capsid proteins, or unidentified capsid-encoding segments. We have added the text "Unlike classical narnaviruses and mitoviruses, ourmiaviruses encode a capsid protein on one of their genome segments, and some of the "ourmia-like" viruses encode a capsid protein in a separate ORF on the RdRp-encoding RNA.", and "In none of the sequences did we observe an additional ORF that might encode a capsid protein. Nonetheless, it is possible that some of these viruses might encode a capsid protein on an undetected genome segment (cf. Lye et al. 2016) or in an extra ORF 5' or 3' of the RdRp ORF that has been missed as a result of an incomplete assembly, or might exploit a helper virus for the provision of a capsid protein."

(2) I was confused by the note that there was 'Scant conservation' in the protein encoded by the rORF, as in the results this was described as "Overall, the amino acid sequences predicted to be encoded by rORFs were rather more divergent than the corresponding set of RdRps (mean pairwise identities 20.1% and 26.6%, respectively)." To me this implies that it is almost as highly conserved (20%) as the RdRp (27%). Based on closer pairs only (say RdRp of 50% protein identity), is there good evidence for some conservation in the rORF protein (i.e. more than expected based on the AA identity seen in the RdRp, the codon usage, and conditioning on the absence of rORF stops?)

Regarding the first point: These levels of global identity (20.1%, 26.6%) are very much in the "twilight zone", i.e. the region where protein homology inference based on sequence alignment identity becomes uncertain. For example Rost (1999, PMID 10195279) analyzed more than a million sequence alignments between protein pairs of known structures and found that, "above a cut-off roughly corresponding to 30% sequence identity, 90% of the pairs were homologous; below 25% less than 10% were". The point we wished to make at the referenced part of the ms was not so much a statement on global conservation (which is minimal) but rather whether or not more local regions of conservation could be detected. From Figure 7 we can see that the RdRp has many more, stronger regions of amino acid conservation than the rORF. Whereas the 20.1% and 26.6% global values don't distinguish whether or not conservation is localized to the same regions in each sequence pair, Figure 7 is based on multiple sequence alignments and thus highlights localized regions that are more conserved than average across many sequences.

To clarify this, we have now reworded "We found scant evidence for amino acid conservation in the rORF." to "In contrast to the RdRp, we found little evidence for conserved amino acid motifs in the rORF protein (**Figure 7**)."

The second point is non-trivial. One way would be to apply synplot2 to the RdRp ORF of more closely related sequences than currently shown in Figure 7. However, this does not directly address the question and does not distinguish between rORF amino acid conservation and, e.g., embedded functional RNA structures. Instead we performed some sequence randomization

experiments to attempt to address this point with, as suggested, pairs of more closely related sequences. "Conditioning on the absence of rORF stops" was achieved by restricting each analysis to the part of the RdRp ORFs that are overlapped by the rORFs. A simple way to control for the amino acid identity and codon usage in the RdRp ORF might be to shuffle all the RdRp codons used for each amino acid X between all sites in the RdRp ORF where amino acid X is present – i.e. shuffle the RdRp nucleotide sequence while preserving its amino acid sequence order and codon usage statistics. Then the % amino acid conservation between the two RdRps will be preserved and one can assess how the shuffled RdRp codons have affected the rORF % amino acid conservation. This approach is however flawed because the two sequences are not phylogenetically independent: e.g. if sequence 1 has a CUG codon at position i , then sequence 2 is also more likely than random to have a CUG codon at position i (e.g. if the ancestral sequence had a CUG at that site). Thus it is essential to shuffle alignment columns not separate sequences. Of course, if alignment RdRp codon columns are shuffled, then the % amino acid identities for both the RdRp ORF and the rORF will be unchanged from the original sequence pair since the rORF is in phase with the RdRp ORF so has the same codon columns. To move forwards, we decided to look at conservation of amino acid k -mers, figuring that functionally important amino acid motifs often occur in clumps. We count the % of k -mers that are conserved between sequence 1 and 2, for $k = 1, 2, 3, \dots$, for each of the RdRp and rORF, and for the WT pairwise alignment and shuffled alignments. We shuffle RdRp codon columns between sites that encode the same amino acid pair (i.e. if sequence 1 has amino acid X and sequence 2 has amino acid Y at positions i and j , then, in our shufflings, it is permissible to switch the sequence 1 and 2 codons at position i with the sequence 1 and 2 codons at position j). This will have no effect on the RdRp amino acid conserved k -mer counts nor the rORF amino acid conserved 1-mer counts, but it may change the rORF amino acid conserved k -mer counts for $k > 1$.

When applied to, for example, the sequence pair KP642119 and KP642120 (61% amino acid identity in the RdRp), 100 randomizations, we get the following results:

	k :	RdRp					rORF				
		1	2	3	4	5	1	2	3	4	5
WT		482	331	233	172	130	261	105	41	21	12
randomized (means)		482	331	233	172	130	261	90.6	32.7	12.1	4.7

As noted above, only the last 4 of the 10 columns can have different values between WT and the randomizations. The result is that, randomizing the sequence decreases the number of conserved amino acid 2-, 3-, 4- and 5-mers between the two sequences in the rORF polypeptide. This might be taken as evidence that the rORF polypeptide is subject to some purifying selection at the amino acid level. Nonetheless, we can think of various caveats with this approach that may render this observation an artifact.

Thus we have not included any change in the ms for this point (see also our response to the related point 3 below).

(3) related to 2 above, is there any detectable homology at all between different 'clusters' of rORFs (i.e. clades of the the RdRp tree that encode the rORF)? If so, wouldn't this imply multiple losses rather than gains? What about similar properties for the rORF protein (predicted secondary structure? conserved amino-acid properties or composition (after having accounted for any that is enforced by encoding the RdRp on the other strand)?

Regarding the first part:

We split the rORF sequences into five main clades (based on Figure 2), aligned each clade with MUSCLE, and then used HHpred to compare each clade with each other clade. The 10 clade-versus-clade comparisons had e-values ranging from 0.019 to $8.9e-16$. Thus, at first glance, this analysis indicates that there is homology between different clusters of rORFs. However, to what extent this reflects constraints imposed by the RdRp and/or common ancestry as opposed to true purifying selection on the rORF amino acid sequence is difficult to ascertain. As a kind of control, we did the same analysis for seven non-rORF containing clades (again based on Figure 2), where

we generate “fake” rORF amino acid sequences by translating the RdRp ORF region in the reverse reading direction, and change any stop codons to the ambiguous amino acid code “X”. Many of the non-rORF versus non-rORF interclade HHpred comparisons have comparable e-values to the rORF versus rORF interclade comparisons. Thus the apparent homology between different clusters of rORFs may indeed be due to constraints on the RdRp and/or common ancestry. Actually, in these analyses, it is difficult to control for differences in the number and divergence of sequences in the within-clade alignments, and differences in the divergences between the different clades of the different comparisons. Thus, we do not think this analysis is conclusive one way or the other, and so we have not added these analyses to the manuscript.

Regarding the second part:

We searched for evidence of known motifs (ELM - Eukaryotic Linear Motif server) and transmembrane regions (TMHMM, TMPred) in individual sequences, and between related sequences at various levels of identity, but did not find anything of particular note.

We have now also looked at various sequence properties: pI, predicted helix/sheet/turn/coil composition, and hydrophobicity plots, using the 26 representative RdRp and rORF sequences.

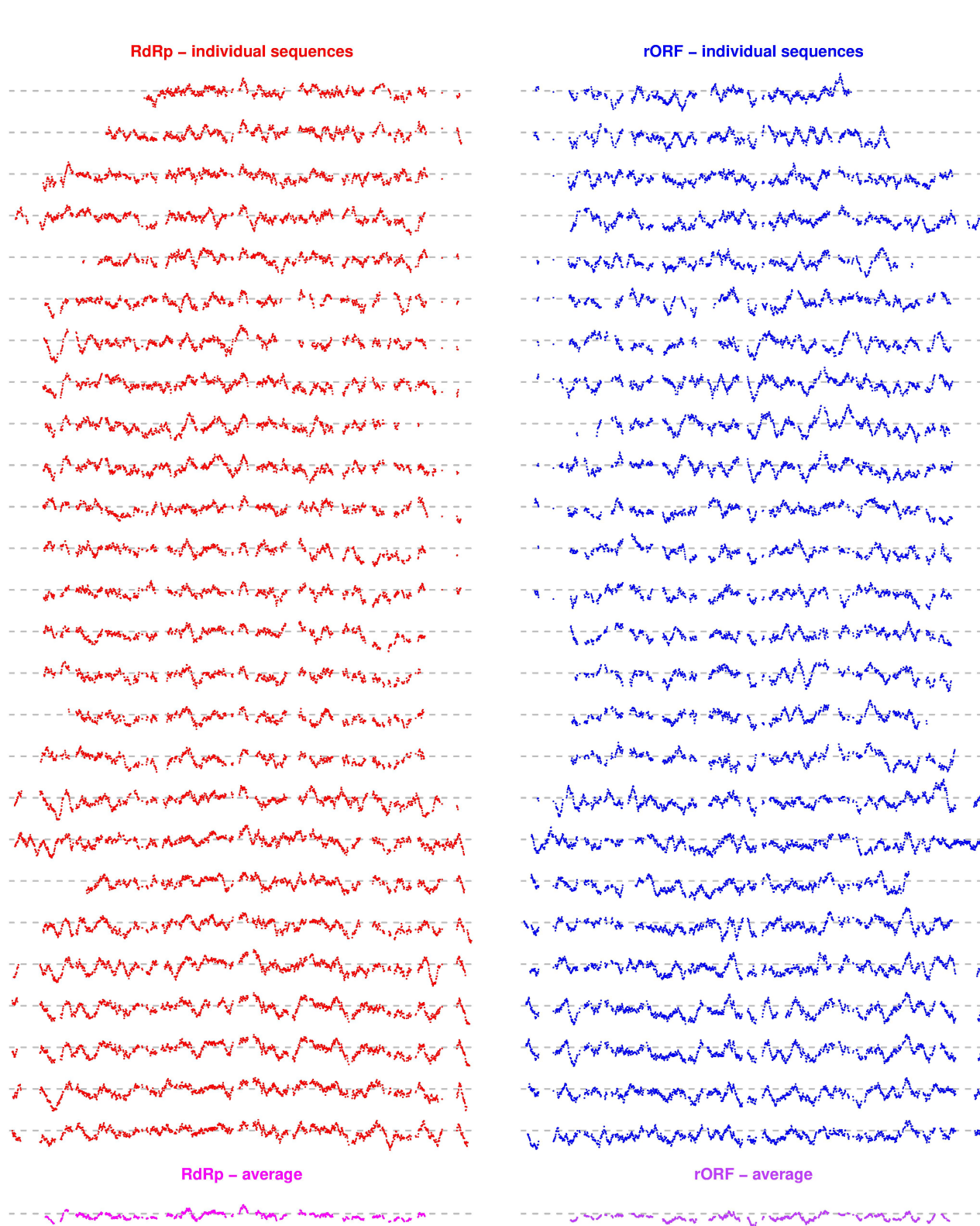
Isoelectric point (pI) is potentially interesting. Both the rORF polypeptide and the RdRp are quite basic (rORFs 9.82 ± 0.98 , RdRps 10.06 ± 0.59 ; cytoplasmic proteins have an average pI value of ~5–6). This is consistent with a hypothesis that the rORF protein might, like the RdRp, bind the viral RNA to protect it and maintain it in a single-stranded form in the cytoplasm. We have now added to the ms the text "Both the RdRp and rORF amino acid sequences are highly basic (isoelectric points: rORFs – 9.82 ± 0.98 , RdRps – 10.06 ± 0.59 ; means $\pm$ standard deviations). The high isoelectric point of the RdRp is consistent with it binding viral nucleic acid to form ribonucleoprotein complexes. It is possible that the rORF protein might also bind viral nucleic acid, perhaps (given its expected low expression level) with a specificity for the negative-strand."

We predicted helix/sheet/turn/coil composition with the garnier program in the EMBOSS suite. This has a fairly low prediction accuracy (~65%), but we preferred to use a single-sequence method, and then simply look at overall proportions of positions called as helix, sheet, turn or coil. The results are (means $\pm$ standard deviations over the 26 representative rORFs and RdRps):

	rORFs	RdRps
helix	0.191 ± 0.048	0.314 ± 0.051
sheet	0.259 ± 0.038	0.242 ± 0.034
turns	0.310 ± 0.033	0.229 ± 0.033
coil	0.240 ± 0.037	0.214 ± 0.027

Thus the RdRps have more predicted helix than the rORFs and the rORFs have more predicted turns. Even so, we decided not to add this to the ms as it does not obviously lead anywhere.

We calculated hydrophobicity with the pepwindow program (19-aa window) in the EMBOSS suite, which produces Kyte-Doolittle plots. We then transformed the plot for each sequence to alignment coordinates, so that profiles could be compared across the 26 different rORF and RdRp sequences. We also plotted the mean (for positions in the alignment with <20% gaps), with the idea being, perhaps, that if the rORF has a less-conserved hydrophobicity profile, then the mean should be flatter. These data are shown in the plot on the following page (red = RdRp, blue = rORF, pink = RdRp mean, purple = rORF mean). However, once again, there does not seem to be anything to be gained by adding this analysis to the ms.



1
2 **Finally, I note that a substantial amount of the material in lines 340-370 describe results,**
3 **and I think it would be better if all results appeared in the results section (with**
4 **accompanying methods in the methods section).**

5
6 The parts describing the SRA taxonomy analyses are simply reporting data shown in the NCBI
7 databases, so these are not really something that can be described in our Results/Methods
8 sections. Similarly much of the other material here is a discussion of previously published results.
9 We have however rearranged the EVE/WGS material as suggested, adding a new (small) Results
10 section ("EVEs related to rORF-containing narnaviruses are found in insect WGS datasets"),
11 besides corresponding details in Methods.

12
13 We also added another sentence to the Discussion in light of a very recent paper: "On the other
14 hand, it has also been suggested that arthropod-associated narnaviruses may instead be infecting
15 trypanosomatids – common intracellular parasites of arthropods (Harvey et al., 2019)
16 (trypanosomatid DNA – including EVEs – could potentially also contaminate WGS datasets of their
17 hosts)."

18
19 -----
20
21 **Reviewer: 1**

22
23 **It's a pleasure to read this paper. The authors presented strong bioinformatics evidence**
24 **that in some alphanavirus, the negative strand encode functional ORFs. The performed**
25 **phylogenetic analysis which separate nanaviruses into two proposed genera. The showed**
26 **that in about half of the alphanaviruses, the negative strand contained an ORF that**
27 **spanned most of the sequences. The length of the ORFs, their R0 reading frame in relation**
28 **to the reading frame of RdRp, and the codon avoidance of UUA, CUA, and UCA, provide**
29 **strong evidence that these ORFs are functional.**

30
31 **It will be exciting if the authors or someone else can provide experimental evidence that the**
32 **rORFs are indeed expressed at the protein level. Since their sequences are divergent, and**
33 **they only exist in some of the alphaviruses, it's difficult to imagine that they have essential**
34 **function. However, if they lack of essential function, how do the viruses maintain such**
35 **rORFs? I look forward to follow-up studies.**

36
37 **A few minor comments:**

38
39 **1. Do any of the rORFs-encoded proteins have any similarity to known proteins?**

40
41 To check for significant homology to known proteins, we performed blastp searches against the
42 NCBI non-redundant (nr) protein database, using each of the protein sequences predicted to be
43 encoded by an rORF separately as a query. Apart from other rORF-containing narnaviruses, no
44 significant homology was detected. We also applied HHpred, using rORF amino acid alignments
45 as queries (five alignments, one for each of the main rORF clades in Figure 2, with – for
46 definiteness – IABX01132835, KX883605, KX883500, excluding KX883602 being one of these
47 clades), searching against PDB_mmCIF70_11_Oct and Pfam-A_v32.0, but again found no
48 potential homologues (best e-value = 56). This is not surprising, since overlapping genes normally
49 have to evolve via overprinting of an ancestral gene (in this case, clearly the RdRp), and not via
50 gene duplication and divergence, and thus they nearly always exhibit novel protein folds (Keese &
51 Gibbs, 1992, PMID 1329098; Rancurel et al, 2009, PMID 19640978). We have added details of
52 these analyses to the ms Results (top of "Analysis of amino acid and nucleotide conservation in
53 the rORF" section) and Methods.

54
55
56 **2. Page 9 lines 194-196: The authors restricted the rORFs finding to within 200 nt of 5' end.**
57 **Though the provided justification, it would be beneficial if they repeat this part of analysis**
58 **without this restriction and present it as a supplemental figure.**
59
60

We have included the additional figure in the Supplementary Material (new Supplementary Figure 5), showing the longest stop codon-free regions (i.e. ORFs), as a percentage of sequence length, regardless of the 5' start position of the ORF. Similar to the original version (Figure 4), there is a clear segregation of rORF-containing and rORF-lacking taxa, based on R0-frame ORF lengths. As would be expected, when the 5' start position of the ORF is not restricted, the maximal relative ORF lengths increase for many sequences. In particular, two additional sequences have an R0-frame ORF occupying ~36% of the respective sequences. However, this value (36%) is still considerably closer to the range for non-rORF-containing sequences (previously 5–25%) than it is to the range for rORF-containing sequences (94–100%). Hence, our argument – that a clear distinction can be drawn between rORF-containing and rORF-lacking taxa, based on R0-frame ORF length – is unchanged.

3. Page 10, lines 221-222, double check the codon locations. Since they are in the same reading frame, the differences should be multiples of 3. The numbers presented are not.

We apologise for the confusion here. Actually these numbers are in units of codons rather than nucleotides (i.e. a difference of 1 = 1 codon). Therefore, the differences in positions need not be a multiple of three. To help clarify this, we have amended the text as follows:

- "at positions 105 and 373" → "at **codons** 105 and 373"
- "at positions 81, 134 and 157" → "at **codons** 81, 134 and 157"
- "at position 72" → "at **codon** 72"

4. Page 12, lines 262-265, the authors used the virus GC contents to justify their conclusion. The GC contents of their hosts (if known) should be used instead since the viruses use their host machinery for translation.

This is not possible, since the hosts are only really known for a very small subset of narnaviruses (specifically, those for which direct infection has been shown in a laboratory setting). Others might come from fungi or protozoans that infect the main transcriptomic target such as insects or plants. However, the point of this section in the ms is that the presence or absence of the rORF cannot be explained by variation in overall GC3 content of the virus genome. This conclusion is based on a direct comparison of rORF-containing and non-rORF-containing alphanarnaviruses, and it stands regardless of the host taxa.

Reviewer: 2

This paper is close to perfection, it is brilliantly written and its technical level is without question. The authors cannot do miracles, though: the message of the paper boils down to finding long rORF, which arise due to the avoidance of stop codons. There is no information whatsoever as to what the function of the encoded proteins could be. However, I believe that these thoroughly documented findings as such are of great interest and "Virus Evolution" is a perfect journal for presenting them.

We greatly appreciate the reviewer's kind words.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 **ARTICLE**

2

3

4

5

6 **A case for a negative-strand coding sequence in a group of positive-sense RNA**

7

8

9 **viruses**

10

11

12

13 Adam M. Dinan¹, Nina I. Lukhovitskaya¹, Ingrida Olendraite¹, and Andrew E. Firth¹

14

15

16 1. Division of Virology, Department of Pathology, University of Cambridge, Tennis Court Road,

17

18 Cambridge, CB2 1QP, UK

19

20

21

22 7

23

24

25 **Correspondence to:** Andrew E. Firth, aef24@cam.ac.uk

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

9 ABSTRACT

10 Positive-sense single-stranded RNA viruses form the largest and most diverse group of eukaryote-
11 infecting viruses. Their genomes comprise one or more segments of coding-sense RNA that
12 function directly as messenger RNAs upon release into the cytoplasm of infected cells. Positive-
13 sense RNA viruses are generally accepted to encode proteins solely on the positive strand.
14 However, we previously identified a surprisingly long (~1000-codon) open reading frame (ORF)
15 on the negative strand of some members of the family *Narnaviridae* which, together with RNA
16 bacteriophages of the family *Leviviridae*, form a sister group to all other positive-sense RNA
17 viruses. Here, we completed the genomes of three mosquito-associated narnaviruses, all of which
18 have the long reverse-frame ORF. We systematically identified narnaviral sequences in public data
19 sets from a wide range of sources, including arthropod, fungal and plant transcriptomic datasets.
20 Long reverse-frame ORFs are widespread in one clade of narnaviruses, where they frequently
21 occupy >95% of the genome. The reverse-frame ORFs correspond to a specific avoidance of CUA,
22 UUA and UCA codons (i.e. stop codon reverse complements) in the forward-frame RNA-
23 dependent RNA polymerase ORF. However, absence of these codons cannot be explained by other
24 factors such as inability to decode these codons or GC3 bias. Together with other analyses, we
25 provide the strongest evidence yet of coding capacity on the negative strand of a positive-sense
26 RNA virus. As these ORFs comprise some of the longest known overlapping genes, their study
27 may be of broad relevance to understanding overlapping gene evolution and *de novo* origin of
28 genes.

29 INTRODUCTION

30 Traditionally, viruses have been divided between seven Baltimore classes based on the nature of
31 the nucleic acid of their genomes and their replicative intermediates. The seven classes are
32 positive-sense, negative-sense and double-stranded RNA viruses, single-stranded and double-
33 stranded DNA viruses, retroviruses and pararetroviruses (Baltimore 1971). Of these, the single-
34 stranded positive-sense or (+)ssRNA viruses comprise the largest and most diverse group of
35 eukaryote-infecting viruses (Dolja and Koonin 2011). The group includes many important human
36 and animal pathogens (such as dengue, Zika, yellow fever, hepatitis C, foot-and-mouth disease,
37 polio, chikungunya, SARS and MERS viruses) besides the majority of plant viruses.

38 To fully understand the molecular biology of viruses, it is crucially important to know their coding
39 capacity. In recent years, a number of “hidden” protein-coding open-reading frames (ORFs) have
40 been discovered in the genomes of various (+)ssRNA viruses (Chung et al. 2008; Loughran et al.
41 2011; Fang et al. 2012; Firth 2014; Smirnova et al. 2015; Napthine et al. 2017; Lulla et al. 2019).
42 Such genes tend to be very short and/or to overlap previously known coding ORFs, explaining why
43 they have escaped detection prior to the application of sensitive comparative genomic methods.
44 However, all of these experimentally verified novel coding ORFs are in the positive sense. To our
45 knowledge, no negative-sense coding ORF has *ever* been demonstrated in *any* (+)ssRNA virus.
46 Thus, an unanswered question in virology is whether the negative strand of (+)ssRNA viruses can
47 encode proteins.

48 In 2013, by acquiring and analyzing RNA transcriptomic datasets for several mosquito and other
49 dipteran species, we identified two “mosquito-associated” narna-like viruses (Cook et al. 2013).
50 Similar to other narnaviruses, one strand contains a single long ORF that encodes a protein inferred
51 by homology to be the RNA-dependent RNA polymerase (RdRp). However, we were surprised to

notice that both sequences also contain a reverse-frame ORF (rORF) covering nearly the entire sequence of ~3000 nt. We also identified related rORF-containing sequences in public *Puccinia striiformis* and *Uromyces appendiculatus* transcriptome shotgun assembly (TSA) data sets. The extreme divergences between these sequences and the mosquito-associated sequences (~22% amino acid identity in the RdRp sequence) effectively rule out a region of such length being preserved free of stop codons by chance, and thus we hypothesized that the rORF represented a *bona fide* protein-coding sequence.

As currently defined by the International Committee on Taxonomy of Viruses (ICTV), the family *Narnaviridae* contains two genera: *Mitovirus* and *Narnavirus*. Both genera contain single-stranded positive-sense RNA viruses which are non-encapsidated and hence are expected to be transmitted either vertically through cell division or horizontally during host mating. Mitoviruses replicate in the mitochondria of host cells whereas narnaviruses replicate in the cytoplasm. These viruses were originally described as infecting fungi, but related viruses have since been observed in transcriptomic data sets derived from diverse organisms. The narnaviral positive strand normally contains a single long ORF that covers most of the genome and encodes an RdRp which catalyses viral replication. The narnaviral RdRp is highly divergent from those of other eukaryotic RNA viruses, and shows closer homology to the RdRps of RNA bacteriophages in the family *Leviviridae* (Rodriguez-Cousiño et al. 1991; Esteban et al. 1992; Wolf et al. 2018). Detailed comparative genomic and phylogenetic analyses suggest that the *Narnaviridae* are descended from a levivirus-like bacteriophage which may have been carried within the bacterial progenitor of mitochondria at the point of eukaryogenesis, followed by loss of the capsid protein giving rise to capsidless (“naked”) RNA elements (Koonin et al. 2015; Wolf et al. 2018). The escape of a group of these viruses into the cytosol may have given rise to the *Narnavirus* genus (Koonin and Dolja 2014; Wolf et al. 2018).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

76 The prototypical and best studied narnaviruses are the *Saccharomyces cerevisiae* 20S and 23S
77 RNA viruses (ScNV-20S and ScNV-23S, respectively) (reviewed in Wickner et al. 2013). ScNV-
78 20S persistently infects most laboratory strains of the yeast *Saccharomyces cerevisiae*, whereas
79 fewer strains carry ScNV-23S. Their genomes have no 3' poly(A) tail and it is not known if they
80 have a 5' cap structure (Rodríguez-Cousiño et al. 1998). The genomes do not encode capsids, but
81 form ribonucleoprotein complexes with the RdRp in a 1:1 stoichiometry in the host cell cytoplasm;
82 the RdRp interacts with both the 5' and 3' ends which may help protect the viral RNA from
83 degradation by host exonucleases (Solórzano et al. 2000; Fujimura and Esteban 2004; Fujimura
84 and Esteban 2007). Under suitably inducing conditions such as heat shock and nitrogen starvation,
85 copy numbers of either virus can reach 100,000 copies per cell (Kadowaki and Halvorson 1971;
86 Wejksnora and Haber 1978; Esteban et al. 1992). Approximately 98–99% of viral RNA in the cell
87 is in a single-stranded positive-sense form, whereas the remainder exists as single-stranded
88 negative-sense replication intermediates (Rodríguez-Cousiño et al. 1991; Fujimura et al. 2005).
89 When cells are grown at high temperature, double-stranded forms (known as W for ScNV-20S and
90 T for ScNV-23S) accumulate, but these appear to represent by-products and not replication
91 intermediates (Wesolowski and Wickner 1984; Rodríguez-Cousiño et al. 1998; Fujimura et al.
92 2005).

93 To further investigate the presence of the rORF, and now that many more sequences are available
94 in public sequence databases, in this work we present a comprehensive comparative analysis of
95 narnaviral genomes, in which we assess the prevalence, distribution and sequence features of
96 rORFs. Two major clades of narnavirus are identified, for which we propose the establishment of
97 the genera *Alphanarnavirus* and *Betanarnavirus*, with the former clade containing all sequences
98 with long rORFs. Overall codon usage is similar in alphanarnaviruses with and without long
99 rORFs, but the former display a highly specific avoidance of CUA, UUA and UCA codons over

large regions of the RdRp ORF, corresponding to an absence of stop codons in the rORF. We explore possible reasons for the avoidance of CUA, UUA and UCA and conclude that the most plausible explanation is selection to maintain an rORF, indicating that the rORF is functional.

RESULTS

Completion of narnaviral genomic sequences

We completed three genomic sequences for previously described rORF-containing narnaviruses (Cook et al. 2013), and used these genomes as a reference set for comparative analyses. The sequences with GenBank accession numbers KF298275.1, KF298276.1 and KF298284.1 were extended at their 5' termini by 479 nt, 10 nt and 364 nt, respectively, and at their 3' termini by 43 nt, 9 nt and 9 nt, respectively (**Figure 1A**). The latter two sequences are the same length (3093 nt) and encode RdRps which are 97.95% identical at the amino acid level; hence, they were considered to represent a single viral species. Given that the samples from which these sequences were originally obtained were isolated from mosquitoes in the *Ochlerotatus* genus, we name the viruses *Ochlerotatus*-associated narna-like virus 1 (ONLV1; sequence KF298275.1) and ONLV2 (sequences KF298276.1 and KF298284.1) (**Figure 1A**). The AUG codons of the three RdRp ORFs each start at the 7th nucleotide of the respective sequences (contexts: GUCAUGA, GUUAUGG and GUUAUGG), the same position as that of the RdRp AUG in ScNV-23S (GenBank: NC_004050.1). The RdRp-encoding ORFs are 3045, 3075 and 3075 nucleotides in length, respectively. The RdRp stop codons (UGA, UAA and UAA) end at positions 14 nt, 13 nt and 13 nt,

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

respectively, from the 3' termini of the genomes. Hence, the length of the 3' UTR mirrors closely that of ScNV-20S (12 nt; GenBank: NC_004051.1) but is shorter than that of ScNV-23S (59 nt).

An expanded narnaviral phylogeny

We employed tblastn (Camacho et al. 2009) and hmmsearch (Finn et al. 2011), using hidden Markov model based profiles (pHMM) to identify sequences encoding proteins closely related to narnaviral RdRps in the NCBI non-redundant nucleotide (nr/nt) and TSA databases. In total, 124 unique such sequences were identified – 46 in the nr/nt database and 78 in the TSA database. The holobiont sources of these TSA sequences were phylogenetically diverse and included 21 arthropod, 8 fungal and 10 plant species. To place these sequences in context, we also selected a number of mitovirus and ourmiavirus sequences as outgroups, besides additional narna-like virus sequences from Shi et al. (2016), giving 141 sequences in total (**Supplementary Table S1**). We oriented all sequences (by reverse complementing if necessary) so that the RdRp ORF was on the forward strand.

To assess the relationships among sequences, an alignment of the predicted RdRp proteins was generated using MUSCLE (Edgar 2004) (**Supplementary File S1**), and a Bayesian Markov chain Monte Carlo (MCMC)-based tree was constructed from this alignment using MrBayes (Huelsenbeck and Ronquist 2001; **Figure 1B**). As expected, the free-floating genus *Ourmiavirus*, which includes plant-associated viruses with tripartite genomes, grouped unambiguously with the narnaviruses (Turina et al. 2017; Rastgou et al. 2009; Wolf et al. 2018), whereas mitoviruses formed a sister clade (**Figure 1B**). A relatively small, but highly diverse set of fungal- and arthropod-derived sequences clustered with the ourmiaviruses (**Figure 1B**). Recent evidence shows that these “ourmia-like” viruses display a range of genomic architectures, which can be segmented

(as in the case of the *bona fide* ourmiaviruses) or non-segmented, occurring in mono- and dicistronic forms (reviewed in Dolja and Koonin 2018). For example, botrytis ourmia-like virus, which appears to have a non-segmented genome, clusters unambiguously with ourmiaviruses rather than with the narnaviruses (Donaire et al. 2016; **Figure 1B**). Unlike classical narnaviruses and mitoviruses, ourmiaviruses encode a capsid protein on one of their genome segments, and some of the “ourmia-like” viruses encode a capsid protein in a separate ORF on the RdRp-encoding RNA (Dolja and Koonin 2018).

Two major clades of narnaviruses

The narnaviral sequences predominantly fell into one of two major clades, both of which had Bayesian posterior probabilities of 0.95 (**Figure 1B**). Based on these data, we propose that the genus *Narnavirus* be subdivided to form two new genera, which we name *Alphanarnavirus* and *Betanarnavirus*, besides additional unclassified sequences.

The alphanarnaviral clade contains the prototypical narnaviruses, ScNV-20S and ScNV-23S, as well as sequences associated with a range of other fungal taxa, including members of the divisions Ascomycota, Basidiomycota and Entomophthoromycota, and the divergent subphylum Mucoromycotina (**Figure 2**). The alphanarnaviral sequences ranged in length from 1805 nt to 3874 nt (**Supplementary Figure S1**), and the pairwise amino acid identities of the corresponding set of RdRps ranged from 16.8% to 99.8% (**Supplementary File S2**). Putative viral sequences containing an rORF all clustered within the alphanarnaviral clade, and none in the betanarnaviral clade. However, the rORF-containing sequences appear not to form a monophyletic clade, but instead cluster in several regions of the phylogeny, and are found in sequences derived from fungi, arthropods and plants (**Figure 2: red bars**). The core RdRp catalytic regions – motifs A to E in the

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

167 palm domain and motifs F and G in the fingers (Wu et al. 2015; te Velhuis 2014) – are well-
168 conserved despite the overall high degree of sequence divergence (**Supplementary Figure S2**).

169 The betanarnaviral clade includes several viruses of unicellular eukaryotes, including the
170 oomycete-infecting *Phytophthora infestans* RNA virus 4 (PiRV4; Cai et al. 2012) and the
171 protozoan-associated *Leptomonas seymouri* narna-like virus 1 (Lye et al. 2016), as well as
172 sequences from red algae, brown algae, myxozoa and arthropods (**Figure 3**). The betanarnaviral
173 sequences ranged in length from 2215 nt to 3610 nt (**Supplementary Figure S1**), and the pairwise
174 RdRp amino acid sequence identities ranged from 15.2% to 99.9% (**Supplementary File S3**). In
175 the case of *Leptomonas seymouri* narna-like virus 1, the genome may be bipartite, with the RdRp
176 being encoded on the longer (L) segment, although the functional association of the putative
177 segments has not been shown experimentally (Lye et al. 2016).

178 Notably, in both clades, a number of TSA sequences are highly divergent from other sequences in
179 the phylogenies (**Figure 2** and **Figure 3**), suggesting that further sampling will continue to reveal
180 new clades. Conversely, other groups were disproportionately well sampled (e.g. nine Zhejiang
181 mosquito virus 3 sequences with >95% pairwise amino acid identity were found; Shi et al. 2017).
182 In none of the sequences did we observe an additional ORF that might encode a capsid protein.
183 Nonetheless, it is possible that some of these viruses might encode a capsid protein on an
184 undetected genome segment (cf. Lye et al. 2016) or in an extra ORF 5' or 3' of the RdRp ORF that
185 has been missed as a result of an incomplete assembly, or might exploit a helper virus for the
186 provision of a capsid protein.

187
188 **Genomic architecture and terminal regions**

1
2
3
4 189 The terminal sequences of alphanarnaviral and betanarnaviral genomes were found to be dissimilar
5
6 190 – the former group having short runs of G and C residues at the 5' and 3' termini, respectively, and
7
8
9 191 the latter having A/U-rich termini with considerably longer 3' untranslated regions (UTRs)
10
11 192 (**Supplementary Figure S3**). Local RNA secondary structures are predicted to occur at the 5' ends
12
13
14 193 of the ONLV1 and ONLV2 genomes, coincident with a reduction in synonymous site variation in
15
16 194 the RdRp ORF (see below). The putative structures are large (92 nt and 103 nt), and the G-rich 5'
17
18 195 terminus forms an integral component of the stem duplex (**Supplementary Figure S4A**), as it does
19
20 196 in the *Saccharomyces* narnaviruses (Rodríguez-Cousiño et al. 1998; Fujimura and Esteban 2007).
21
22
23 197 Meanwhile, the stop codons of the RdRp ORFs are predicted to be situated within shorter RNA
24
25 198 stem-loop structures at the genomic 3' termini (**Supplementary Figure S4, B and C**).
26
27
28 199

200 **Identification of ORFs in narnaviral genomes**

201 We next determined for each sequence the longest stop codon-free region (ORF) in each of the
202 three possible reading frames on each of the positive and negative strands, restricting to ORFs that
203 begin within 200 nt of the 5' end of the positive or negative strand, as appropriate (**Figure 4**; see
204 **Supplementary Figure S5 for a similar analysis without the 5'-proximity restriction**). We used this
205 approach because, under normal circumstances, non-5'-proximal ORFs are not expected to be
206 translated, and the inclusion of spurious long non-5'-proximal ORFs might dilute the signal from
207 translatable 5'-proximal ORFs. We designated the reading frame of the RdRp ORF as frame “F0”
208 (for forward orientation, 0 frame), followed by frames “F+1” and “F+2” in the same orientation;
209 whereas the reverse complement of the set of codons in frames F0, F+1 and F+2 were designated
210 “R0” (reverse orientation, 0 frame), followed by “R+1” and “R+2” respectively.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

In line with the relatively short lengths of narnaviral untranslated regions (UTRs), the RdRp-encoding (i.e. frame F0) stop codon-free regions occupied 93.8–100.0% (median 99.5%) of alphanarnaviral sequences and 70.5–100.0% (median 97.0%) of betanarnaviral sequences (**Figure 4**). Several sequences in the data set are likely to be incomplete and, as a result, to have misannotated start codons; for example, the genome of Hubei narna-like virus 18 (KX883517.1; Shi et al. 2016) is annotated as containing a 155-nt 5' UTR, but this UTR contains no stop codons in-frame with the RdRp, and a blastx search of the sequence shows that it encodes amino acid sequence that shows close homology to the RdRp of the related Hubei narna-like virus 19 (e-value = 4×10^{-12}).

Strikingly, other than the RdRp ORF, 5'-proximal stop-codon free regions occupying >25% of the genome were present only in the R0 frames and only in a subset of alphanarnaviral sequences, whereas stop (UAG, UAA, UGA) codons in all other frames were relatively common (**Figure 4**). The absence of stop codons in frame R0 in this subset directly mirrors an absence of the reverse-complementary codons (CUA, UUA, UCA) in the RdRp-encoding (F0) frame.

In a single case – Wenling narna-like virus 7 (GenBank: KX883602.1; Shi et al. 2016) – an rORF of intermediate length (531 codons; 58.5% of the sequence) was found (**Figure 4**), compared with an 890-codon stop-free RdRp-encoding region on the forward strand. The 5' half of this RdRp ORF includes several codons corresponding to R0-frame stops: including two CUA (reverse-strand = UAG) codons, at **codons** 105 and 373; three UCA (reverse-strand = UGA) codons at **codons** 81, 134 and 157; and one UUA (reverse-strand = UAA), at **codon** 72. However, the closely related sequences of Wenling narna-like virus 8 (KX883605.1), Beihai narna-like virus 24 (KX883500.1), and the TSA sequence from *Caridina multidentata* (IABX01132835.1) all contain rORFs occupying >90% of the respective sequences (**Figure 2**).

234

235 Avoidance of CUA, UUA and UCA codons in RdRp-encoding ORFs

236 To enable a representative assessment of codon usage in narnaviral ORFs, sequences predicted to
237 encode RdRp proteins with >90% amino acid identity were clustered using CD-HIT (Fu et al.
238 2012) and a single sequence from each cluster was retained for further analysis. After this step, 53
239 alphanarnavirus and 29 betanarnavirus representative sequences were left. An rORF occupying at
240 least 90% of the sequence was present in 26 of the 53 alphanarnavirus sequences.

241 Among alphanarnaviruses, codon usage (as a proportion of total codons) in the RdRp ORF was
242 broadly similar in sequences with and without long rORFs, with the exception of the three forward-
243 orientation codons that introduce reverse-orientation stops (i.e. CUA, UUA and UCA) (**Figure 5**).
244 Effectively by definition, each of these three codons is excluded from large portions of sequences
245 with long rORFs (mean usage per associated amino acid in RdRp ORF = 0.0037, 0.0054 and
246 0.0056, respectively), but are 17- to 26-fold more common in those alphanarnaviruses without long
247 rORFs (mean usage per associated amino acid = 0.0973, 0.0896, and 0.1277, respectively) (**Figure**
248 **6A**). These three codons were also relatively common in the RdRp ORFs of the betanarnaviruses
249 (mean usage per associated amino acid = 0.1323, 0.1452 and 0.1784, respectively). CUA and UUA
250 both encode leucine (Leu) whereas UCA encodes serine (Ser). Examination of codon usage bias
251 for these two amino acids shows that sequences with long rORFs specifically avoid these codons
252 and use proportionately more of each alternative codon to encode these two amino acids (**Figure**
253 **6A**).

255 Comparison of codon usage across species

1
2
3
4 256 Like all (+)ssRNA viruses, narnaviruses are dependent upon host tRNA pools and translation
5
6 257 machinery for their gene expression. This raises the possibility that the paucity of CUA, UUA and
7
8 258 UCA codons in the rORF-containing sequences could at least partially reflect an adaptation to
9
10 259 particular host codon usage patterns.
11
12
13
14 260 Using the RefSeq database from NCBI, as tabulated in the latest release of the codon usage table
15
16 261 database (CUTD) (Athey et al. 2017), we assessed the relative proportions of the different leucine
17
18 262 and serine codons across species. We found that global usage of CUA, UUA and UCA (per
19
20 263 associated amino acid) scales directly and inversely with GC content at the third position of codons
21
22 264 (GC3) and only species with extremely high GC3 had bias against CUA, UUA and UCA as
23
24 265 extreme as that in the rORF-containing narnaviruses (**Figure 6B**). Most taxa with mean overall
25
26 266 CUA, UUA and UCA usage as low as that of rORF-containing narnaviruses are very high-GC
27
28 267 bacteria and archaea (mean GC3 content = 95.2%). The eukaryotic taxon with lowest mean overall
29
30 268 usage of these three codons (0.0107, cf. mean for rORF-containing narnaviral sequences = 0.0049)
31
32 269 was a fungus, *Rhodotorula graminis* (phylum Basidiomycota), which has a GC3 content of 90.4%.
33
34 270 In contrast, the mean GC3 contents of rORF-containing and non-rORF-containing alphanarnaviral
35
36 271 sequences were considerably lower (61.1% and 59.9%, respectively; **Figure 6B**). Thus, selection
37
38 272 against CUA, UUA and UCA codons in rORF-containing alphanarnaviruses cannot readily be
39
40 273 explained by GC bias or host codon usage bias.
41
42
43
44
45
46
47 274 To investigate whether codon usage could be used as an indicator of the likely host taxonomic
48
49 275 group or groups for narnaviruses, and motivated by previous work (Kapoor et al. 2010), we
50
51 276 performed a principal component analysis (PCA) of codon usage statistics for host groups that
52
53 277 frequently co-occur with narnaviruses in transcriptomic datasets. The PCA allowed clear
54
55 278 segregation of arthropods from ascomycetes and basidiomycetes (two major fungal phyla) and
56
57 279 streptophytes (land plants and some green algae), but the latter three groups did not clearly
58
59
60

segregate from each other (**Supplementary Figure S6**). Narnaviral codon usage statistics were projected onto the resulting principal component space, and did not clearly segregate with a single host group, regardless of whether or not leucine and serine codons were included in the analysis (**Supplementary Figures S6 and S7**). Chordates were excluded from this analysis, as their codon bias is heavily influenced by the avoidance of CpG- and UpA-ending codons, and they group distinctly from other phyla and from the narnaviruses.

Analysis of amino acid and nucleotide conservation in the rORF

To check for homology to known proteins, we queried rORF amino acid sequences and alignments with blastp and HHpred respectively but, apart from other rORF-containing narnaviruses, no significant homologies were detected. This is not surprising, since overlapping genes normally have to evolve via overprinting of an ancestral gene (in this case, the RdRp), and not via gene duplication and divergence, and thus they nearly always exhibit novel protein folds (Keese and Gibbs 1992; Rancurel et al. 2009).

Both the RdRp and rORF amino acid sequences are highly basic (isoelectric points: rORFs – 9.82 ± 0.98 , RdRps – 10.06 ± 0.59 ; means $\pm$ standard deviations). The high isoelectric point of the RdRp is consistent with it binding viral nucleic acid to form ribonucleoprotein complexes. It is possible that the rORF protein might also bind viral nucleic acid, perhaps (given its expected low expression level) with a specificity for the negative-strand.

To assess conservation within the rORF of the representative narnavirus sequences, the predicted RdRp amino acid sequences were aligned and then back-translated to RNA. Overall, the amino acid sequences predicted to be encoded by rORFs were rather more divergent than the

1
2
3
4 302 corresponding set of RdRps (mean pairwise identities 20.1% and 26.6%, respectively). Codon-
5
6 303 based alignments of RdRp ORFs and rORFs were generated, based in both cases on the RdRp
7
8 304 amino acid alignments, and amino acid conservation and synonymous site variation of codons
9
10
11 305 (Firth 2014) were assessed (**Figure 7**).

12
13
14 306 The core RdRp functional motifs (A to G; see above) localise to a region of fewer than 400 amino
15
16 307 acids; with additional short conserved motifs (R&UP and Pxx[L/V]GGx[G/N]xP; U = I, L, V or
17
18 308 M, & = I, L, V, M, A, P, G, F, W or Y) being found further upstream and downstream, respectively
19
20 309 (**Figure 7A**). Increases in synonymous site conservation were observed in the rORF directly
21
22 310 opposite RdRp motifs A and E, and – to a lesser extent – motifs B, C, D and F, and at the
23
24 311 Pxx[L/V]GGx[G/N]xP motif (**Figure 7B**) indicating that these conserved motifs in the RdRp
25
26 312 constrain synonymous site variation in the rORF.

27
28
29
30
31 313 The largest increase in synonymous site conservation in the RdRp ORF lies between codons 23 and
32
33 314 37 (**Figure 7A**), coinciding with the location of a known RdRp interaction site and a *cis*-acting
34
35 315 replication signal in the ScNV-20S genome (Fujimura and Esteban 2007). This region also contains
36
37 316 the predicted 5' structures shown in **Supplementary Figure S4A**. Increased synonymous site
38
39 317 conservation was independently observed in this region for betanarnaviruses (not shown). The
40
41 318 most highly conserved region in the putative rORF-encoded amino acid sequence was also directly
42
43 319 opposite the RdRp motif B, in a 15-amino acid window of the alignment centred on position 510
44
45 320 (**Figure 7B**), where the rORF-encoded consensus sequence is QvxExExxPREREAH.

46
47
48
49
50
51 321
52
53
54 322 **EVEs related to rORF-containing narnaviruses are found in insect WGS datasets**

Endogenized virus elements (EVEs) represent virus fragments which have been spuriously reverse-transcribed and integrated into host genomes; EVEs can provide evidence as to the true host(s) of virus groups (Katzourakis and Gifford 2010). To identify endogenized narnavirus-derived sequences, we queried the ONLV1 and ONLV2 RdRp sequences against NCBI arthropod whole genome shotgun (WGS) datasets. Nine matches were found, all with e-values $\leq 5 \times 10^{-15}$: BHEC01027060.1 (*Coccinella septempunctata*), JJNS02035057.1 (*Homalodisca vitripennis*), SJPC01005419.1 (*Cataglyphis niger*), NJRP01000653.1 and NJRP01005509.1 (*Aphaenogaster floridana*), LBMM01010399.1 (*Lasius niger*), ADOQ01003831.1 and ADOQ01001783.1 (*Linepithema humile*) and NJRK01001360.1 (*Aphaenogaster rudis*), all of which derive from insect datasets. To confirm that these EVEs clustered most closely with rORF-containing narnaviruses, we performed reciprocal tblastn analysis against the NCBI virus nr/nt database. In all nine cases, the top virus match was one of KX883548, MH213236, KX883539 or KP642119, all of which are indeed insect-associated rORF-containing narnaviruses.

DISCUSSION

In this study, we have shown that narnaviral sequences form two major clades, which we suggest to be formally recognized as a pair of genera. We show that sequences with long rORFs are restricted to the proposed genus *Alphanarnavirus*, which also contains the prototypical *Saccharomyces cerevisiae* 20S and 23S narnaviruses. The proposed genus *Betanarnavirus* contains viruses associated with diverse unicellular eukaryotes, including the oomycete *Phytophthora infestans* and the trypanosomatid *Leptomonas seymouri*. The two proposed genera differ in terms of their relative codon usage preferences and in the composition of their genomic termini, with alphanarnaviruses having GC-rich termini and a higher overall GC3 content in codons, and the betanarnaviruses

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

346 having AU-rich termini. These variations might, to some extent, reflect divergence in host taxa.

347 Although some of the available sequences are likely incomplete, there is evidence that members of

348 both proposed genera exhibit complementarity between the genomic termini. In ScNV-20S and

349 ScNV-23S, the 5'-GGGG and CCCC-3' genomic termini, besides 5' and 3' RNA structures, are

350 essential for efficient virus replication (Esteban and Fujimura 2003; Fujimura and Esteban 2004;

351 Esteban et al. 2005). Since narnaviral RNAs are non-polyadenylated and probably non-capped,

352 they are vulnerable to host mRNA degradation pathways. However, the viral RdRp binds to the 3'-

353 terminal CCCC and an adjacent RNA stem-loop structure in the positive strand and this is thought

354 to stabilize the genome and protect the 3' end from host 3' exonuclease degradation (Fujimura and

355 Esteban 2004; Fujimura and Esteban 2007). Similar elements also exist at the 3' end of the negative

356 strand (Esteban et al. 2005; Fujimura and Esteban 2007). Meanwhile, the positive-strand 5'-

357 proximal RNA structure protects the genome from the host SKI1/XRN1 5' exonuclease (Esteban et

358 al. 2008). These stability elements are thought to be particularly important for narnaviruses due to

359 the lack of capsids or membrane-associated replication that other (+)ssRNA viruses use to protect

360 their genomes (Fujimura and Esteban 2007).

361 The diversity of sources of co-clustering narnaviral genomes suggests either that horizontal transfer

362 of narnaviruses between divergent hosts has occurred, or that at least some of these sources are not

363 the *bona fide* host species, potentially reflecting contamination from parasitic or commensal

364 organisms, gut contents, or external debris. For example, in **Figure 2**: two TSA sequences from the

365 fly-infecting fungus *Entomophthora muscae* (GENC01006608.1 and GEND01011317.1) cluster

366 within a clade of arthropod-derived sequences; a TSA sequence (GFKT011160020.1) from the

367 spider *Nephila clavipes* has 95% nucleotide identity to ScNV-20S (AF039063.1); and two TSA

368 sequences (GGCO01105932.1 and GGCO01034162.1) from the barley plant *Hordeum vulgare*

369 cluster with sequences derived from the Basidiomycota obligate plant pathogenic fungi *Uromyces*

1
2
3
4 370 *appendiculatus* and *Puccinia striiformis*. Since ScNV-20S is a well-characterized virus of
5
6 371 *Saccharomyces cerevisiae*, it is likely that the *Nephila clavipes* sequence derives from fungal
7
8
9 372 contamination. The *Entomophthora muscae* sequences were obtained from fungus-infected *Delia*
10
11 373 *radicum* cabbage flies, and indeed the NCBI SRA taxonomy analysis webpages (alpha version, 30
12
13 374 May 2019) for the corresponding RNA-Seq libraries show 4–16 times as much fly RNA as fungal
14
15 375 RNA (within the only 8–9% of reads that were taxonomically identified); thus these TSAs may
16
17
18 376 derive from infected insect cells. Similarly, the *Hordeum vulgare* TSAs may derive from infected
19
20 377 fungal cells since the NCBI SRA taxonomy analysis webpages show contamination with
21
22 378 Opisthokonta (i.e. animal/fungi) RNA at ~4% the level of plant RNA. Thus, it is important to apply
23
24 379 caution when attempting to infer the hosts of TSA sequences.

25
26
27
28 380 Given the co-occurrence of fungal sequences in arthropod transcriptomic data sets, it was originally
29
30 381 suggested that the identified narnavirus sequences might derive from fungal contaminants
31
32 382 (Chandler et al. 2015; Cook et al. 2013). More recently, evidence was put forward that at least
33
34 383 some narnaviruses **may be** true arthropod viruses: for example, the group of alphanarnaviruses
35
36 384 comprising KF298275.1, KF298276.1, KF298284.1, Zhejiang mosquito virus 3, KP642119.1 and
37
38 385 KP642120.1 (**Figure 2**) come from four different studies and five different Culicinae mosquito
39
40 386 species (Chandler et al. 2015; Cook et al. 2013; Shi et al. 2016; Shi et al. 2017); moreover, in
41
42 387 several samples, viral RNA accounts for >0.1% (in one case, >2%) of total non-ribosomal RNA
43
44 388 reads, which **may be** unlikely if the virus is infecting a contaminant (Shi et al. 2016; Shi et al.
45
46 389 2017). Very recently, another narnavirus in this clade (MK628543.1; Culex narnavirus 1; 97% nt
47
48 390 identity to KP642120.1) was found to persistently infect a *Culex tarsalis* cell culture and also give
49
50 391 rise to typical 21-nt viral siRNAs with equal coverage of both strands, indicative of active infection
51
52 392 (Göertz et al. 2019). **An ONLV2-related EVE has been previously reported for arthropods (Shi et**
53
54 393 **al. 2016), and we also identified sequences that clustered with rORF-containing narnaviruses in**
55
56
57
58
59
60

WGS assembly data sets from several species of ant, the seven spot ladybird and the glassy-winged sharpshooter. On the other hand, it has also been suggested that arthropod-associated narnaviruses may instead be infecting trypanosomatids – common intracellular parasites of arthropods (Harvey et al., 2019) (trypanosomatid DNA – including EVEs – could potentially also contaminate WGS datasets of their hosts). Other rORF-containing alphanarnaviruses come from fungal (or plant) samples that have no obvious association with arthropods (< 0.8% metazoan-mapping reads in total per sample), such as the *Uromyces appendiculatus* and *Puccinia striiformis* samples, indicating that presence of the rORF is likely not a unique adaptation to arthropod-association.

Since our original identification of long rORFs in the ONLV1 and ONLV2 genomes, besides sequences from *Uromyces appendiculatus* and *Puccinia striiformis* transcriptomes (Cook et al. 2013), a number of other studies have discovered narnaviral genomes containing similar rORFs (Chandler et al. 2015; Shi et al. 2016; Shi et al. 2017; Viljakainen et al. 2018; Göertz et al. 2019). These rORF-containing sequences appear not to form a monophyletic group, although they do show considerable phylogenetic clustering. The rORF occurs exclusively in the R0 frame relative to the RdRp ORF, and large sections of the latter therefore exhibit a specific exclusion of the three codons CUA, UUA and UCA, which are however widely used in the RdRp ORF of non-rORF-containing narnaviruses. Nonetheless, these codons are still used occasionally in rORF-containing narnaviruses, exclusively within the 5'-most extremity (3%) of the RdRp coding region (at or upstream of the rORF stop codon). This indicates that selection against these codons is not due to an inability for these codons to be decoded by the host translational machinery (e.g. due to the lack of cognate tRNAs). This is also supported by our analysis of cellular organism codon usage which showed that the vast majority of organisms with GC3 in the same range as rORF-containing narnaviruses have CUA, UUA and UCA mean codon usage values well above those of the rORF-

containing narnaviruses (**Figure 6B**). Thus, the presence of the long rORF in highly divergent alphanarnaviruses cannot be explained as an artefact of RdRp ORF codon usage or GC bias.

To our knowledge there is no experimental evidence for protein-coding ORFs in the negative strand of any (+)ssRNA virus. A recent bioinformatic study used single-sequence randomization procedures to identify ORFs overlapping previously annotated ORFs in RNA virus sequences that are statistically significantly longer than expected by chance, with the assumption being that such ORFs are likely to be functional (Schlub et al. 2018). The authors identified statistically significantly long negative-strand stop-codon-free regions overlapping positive-strand ORFs in sweet potato virus 2 (NC_017970.1; family *Potyviridae*), Macrophomina phaseolina tobamo-like virus (NC_025674.1; family *Virgaviridae*), Nhumirim virus (NC_024017.1; family *Flaviviridae*), hibiscus chlorotic ringspot virus (NC_003608.1; family *Tombusviridae*), hydrangea ringspot virus (NC_006943.1; family *Alphaflexiviridae*), and Scrophularia mottle virus (NC_011537.1; family *Tymoviridae*). However it remains unclear how these ORFs would be translated. In eukaryotes, translation normally relies on recruitment of pre-initiation ribosomes to the 5' end of mRNAs, followed by scanning and initiation at the first AUG in a good initiation context. Thus, the majority of *de novo* protein-coding ORFs tend to evolve towards the 5' ends of transcripts, frequently overlapping the 5' end of an ancestral protein-coding ORF where they can be translated via a process known as leaky scanning (Firth and Brierley 2012). If protein-coding ORFs exist in the negative strand of (+)ssRNA viruses, one might expect the majority to have initiation sites close to the 5' end of the negative strand. However for the six ORFs above there are, respectively, 171, 101, 115, 8, 3 and 100 intervening AUG codons between the 5' end of the negative strand and the first in-frame AUG codon in the ORF, which would appear to rule out 5'-end-dependent scanning as a translation mechanism. 5'-distal ORFs on positive-sense transcripts are often translated via ribosomal frameshifting or stop codon readthrough, but that still requires 5'-proximal initiation in

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

the pre-frameshift/pre-readthrough ORF (Firth and Brierley 2012). 5'-distal ORFs are occasionally translated via an internal ribosome entry site (IRES), but IRESes are normally complex elements and it would be difficult for them to evolve within protein-coding sequences. Simple inefficient IRESes are a possibility and, compounded with the low availability of negative strand, would lead to extremely low expression levels of any resulting proteins. Splicing is unknown in (+)ssRNA viruses as they all replicate cytoplasmically. Subgenome-sized negative-sense transcripts, where they are produced, are normally 5' co-terminal with the full-length negative strand so would not normally provide access to internal ORFs, although the production of other classes of negative-sense transcripts is plausible (Sztuba-Solińska et al. 2011). Given the pronounced evolutionary adaptations that might be required to express such negative-strand ORFs, in the absence of experimental data they would be more plausible if they were conserved in related species. However this has not been demonstrated. For example, the Nhumirim virus rORF is not conserved in the next most closely related flavivirus genome sequence, that of Barkedji virus (MG214906.1; 71% aa identity in the forward ORF) where the rORF region is disrupted by eight stop codons. Similarly, the sweet potato virus 2 rORF is not conserved in sweet potato virus G (NC_018093.1; 74% aa identity in the forward ORF) where the rORF region is disrupted by nine stop codons. Indeed it is not even conserved in other isolates of sweet potato virus 2 – e.g. KP729268.1 and KP115618.1 each have a stop codon. Thus we contend that the alphanarnavirus rORF is currently the only plausible candidate for a negative-strand coding ORF in (+)ssRNA viruses.

The replication of most (+)ssRNA viruses is membrane-associated (den Boon and Ahlquist 2010; Romero-Brey and Bartenschlager 2014; Shulla and Randall 2016; Ertel et al. 2017). Infection typically results in extensive rearrangement of host membranes to produce membrane-bound mini-organelles within which genome replication occurs. These structures co-localize viral RNA and proteins and may also shield viral positive:negative strand duplexes from recognition by host

dsRNA-recognizing antiviral factors such as RIG-I, MDA-5, PKR and the RNAi machinery. Sequestering replication within membranous compartments may also play a role in separation of translation from replication: positive-sense mRNAs are extruded into the cytoplasm for translation, whereas the negative strand remains protected within the membranous compartment for replication. This prevents ribosomes translating 5' to 3' from colliding with RdRps tracking 3' to 5' on the same template. Thus, at least at later time points, the negative strand is expected to be generally sequestered away from the translational machinery. Further, the negative strand of (+)ssRNA viruses often lacks the 5' structures (such as a 5' cap, 5' covalently linked viral protein of the genome, or IRES structure) required to efficiently recruit ribosomes. Narnavirus replication is, however, quite atypical among eukaryote-infecting (+)ssRNA viruses, and this may explain why they – possibly uniquely – appear to have evolved negative-strand coding capacity. Narnavirus replication is thought to occur entirely within the cytoplasm, not associated with cellular membranes (Solórzano et al. 2000; Fujimura et al. 2005). As mentioned above, the narnavirus RdRp is more closely related to the RdRp of bacteriophages in the family *Leviviridae* than it is to the RdRp of other (+)ssRNA viruses. Similar to leviviruses, the RNA in narnavirus replication intermediates is essentially single-stranded (Blumenthal and Carmichael 1979; Dobkin et al. 1979; Zinder 1980; Takeshita and Tomita 2012; Fujimura et al. 2005; Wickner et al. 2013). Each newly synthesized RNA is released before a new round of replication commences and resting complexes comprising a negative-sense or positive-sense single-stranded RNA bound to a single copy of the viral RdRp are present within the cytoplasm (Solórzano et al. 2000). Thus, for narnaviruses, the positive and negative strands likely have similar accessibility to the host cell translational machinery.

The rORF is present in many but not all alphanarnaviruses. Surprisingly, the distribution of the rORF does not appear to be monophyletic. Thus it may have evolved multiple times, or it may have

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

been present ancestrally and lost from some lineages. The observation that the rORF is always in the R0 frame relative to the RdRp ORF may be evidence against independent evolution in multiple lineages, although it could also be a consequence of codon bias, e.g. if long ORFs are by chance more likely to occur in the R0 frame as a result of RdRp ORF codon usage. Such random ORFs could provide “seeds” for the evolution of longer overlapping genes (Belshaw et al. 2007). Another surprising feature is that the rORF, where present, is nearly always full-length. If the sole function of the rORF were to encode an additional protein, one might expect to see a variety of rORF lengths in different lineages whereas we only see a single example of an rORF beginning within 200 nt of the 5' end of the negative strand with length in the range 24–94% of the full sequence length (**Figure 4**). **In contrast to the RdRp, we found little evidence for conserved amino acid motifs in the rORF protein (Figure 7).** It is possible that the very high divergence between most of our sequences might obscure a weak conservation signature (in contrast to the RdRp which is the most highly conserved RNA virus protein known). However it is also possible that the protein product of the rORF is not itself functionally important. A possible alternative explanation is that rORF translation in itself might facilitate replication e.g. by disassociating dsRNA, or by increasing negative strand RNA stability (Bicknell and Ricci 2017). Alternatively, if the negative strand is unavoidably accessible for translation, exon-junction-complex-independent nonsense mediated decay (Kurosaki et al. 2019) may provide strong selection pressure against non-full-length rORFs.

Although, as discussed above, the negative strands of most (+)ssRNA viruses are normally expected to be occluded in membranous compartments, it is currently unclear whether at early timepoints – before extensive membranous compartments have formed – negative strands produced in a first round of replication might be exposed to the cytoplasm and potentially available for translation (Shulla and Randall 2015). It is also not inconceivable that some (+)ssRNA viruses might have evolved mechanisms to allow some negative strands access to the translational

machinery even at late timepoints. If so, it is possible that some (+)ssRNA viruses might have evolved ribosome recruitment elements within their negative strands. Thus it is perhaps too early to dismiss the possibility of negative-strand ORFs in other (+)ssRNA virus lineages. In most cases, however, any proteins encoded on the negative strand would be expected to be expressed at much lower levels than positive-strand encoded proteins simply because of the huge disparity in positive:negative RNA abundance during virus infection (typically of order 100:1; Novak and Kirkegaard 1991; Kopek et al. 2007; Irigoyen et al. 2016). Even for narnaviruses, less than 1–2% of ScNV-20S viral RNA in infected cells is negative-sense (Rodriguez-Cousiño et al. 1991; Fujimura et al. 2005), indicating that the rORF, where present, is likely to be expressed at a much lower level than the RdRp.

In conclusion, we have provided the first strong evolutionary evidence for reverse-strand coding capacity in a group of positive-sense RNA viruses. The alphanarnavirus rORFs are exceedingly long compared to most known overlapping genes (Pavesi et al. 2018; Brandes and Linial 2016). Thus their study is not only of interest to the evolution and molecular biology of viruses, but also of broad relevance to understanding the evolution of overlapping genes and the *de novo* origin of genes. During revision of this manuscript, a related study was also published (DeRisi et al. 2019).

METHODS

The NCBI non-redundant nucleotide (nr/nt) and transcriptome shotgun assembly (TSA) databases were downloaded on August 31, 2018. Narnaviral sequences were identified in these databases using, tblastn (version 2.8.0) (Camacho et al. 2009), with the RdRp protein sequence of ScNV-20S and the updated RdRp sequence of ONLV1 (completed in this study) as queries. In total, 81 sequences were found by combined searches using these queries (46 in the nr database and 35 in

the TSA database). Hidden Markov model (HMM)-based searches were carried out using hhsearch (Finn et al. 2011), based on an alignment of RdRps from rORF-containing narnaviruses, leading to the identification of a further 43 TSA database sequences.

To identify narnavirus-derived EVEs, we queried ONLV RdRp sequences AGW51766.2 and AGW51768.2 against the NCBI WGS database on 18 December 2019, using the NCBI tblastn interface, with default parameters except the taxonomy ID was set to “Arthropoda” and the expect threshold and word size were set to 0.1 and 6 respectively. Match sequences were downloaded, the match region extracted, translated, and reciprocally queried against the NCBI nr/nt database with tblastn, with the taxonomy ID set to “viruses” and other parameters as above.

Phylogenetic trees were generated with MrBayes (version 3.2.7) (Huelsenbeck and Ronquist 2001), using a mixed substitution model with sampling across fixed amino acid rate matrices (aamodelpr = mixed) and 5,000,000 generations. All other parameters were set as defaults.

To check for homology to known proteins, we performed blastp searches against the NCBI non-redundant (nr) protein database, using each of the protein sequences predicted to be encoded by an rORF separately as a query. We also applied HHpred (Zimmermann et al. 2018), using rORF amino acid alignments as queries (five alignments, one for each of the main rORF clades in **Figure 2**, with – for definiteness – IABX01132835, KX883605, KX883500, excluding KX883602 being one of these clades), searching against PDB_mmCIF70_11_Oct and Pfam-A_v32.0.

To check for potential capsid-encoding ORFs, all positive-strand ORFs ≥ 150 codons (between flanking stop codons and/or sequence ends), excluding the RdRp ORF itself, were identified and used as queries in HHpred. Among the 29 ORFs found in 124 narnaviral sequences, none were found to have matches to viral capsid proteins; indeed none had significant e-value matches to any known proteins (the minimum e-value was 10).

RNA structures in the genomic terminal regions were predicted by scanning full-length narnaviral genomes with RNALfold (version 2.4.9) (Lorenz et al. 2011), allowing a maximum base-pair span of 150 nt.

For all comparisons of alphanarnaviruses “with” and “without” rORFs, sequences in which the longest stop codon-free region (beginning within the 5′-most 200 nt) occupied at least 90% of the genomic sequence were assigned to the former group, whereas other sequences were assigned to the latter group.

A codon-based multiple sequence alignment of RdRp ORFs was produced by aligning the RdRp amino acid sequences with MUSCLE (version 3.8.31) (Edgar 2004) and then backtranslating to a nucleotide sequence alignment using the tranalign program in the EMBOSS suite (version 6.6.0.0) (Rice et al. 2000). The alignment was then mapped to ONLV2 sequence coordinates by removing alignment positions that contained a gap character in the ONLV2 sequence. Synonymous site variation was assessed using synplot2 (Firth 2014), with a window size of 15 codons. Comparison of amino acid sequences was performed using plotcon from the EMBOSS suite (version 6.6.0.0) (Rice et al. 2000) with a BLOSUM62 substitution matrix and a window size of 15 amino acids. When analysing synonymous site **conservation** in the rORF, the reverse complement of the RdRp alignment was used, and aligned nucleotides on the reverse strand were converted to the corresponding protein sequences using AMAS (Borowiec 2016).

Principal component analysis (PCA) of codon usage across species was performed using relative codon abundances (i.e. normalized by the abundances of the associated amino acids). The principal components were calculated using the RefSeq sequences of cellular organisms, and narnaviral codon abundances were projected onto the resulting principal component space.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

582 **ACKNOWLEDGEMENTS**

583 We thank Chris McCormick, Katherine Brown, Valeria Lulla and Hazel Stewart for useful
584 discussions. Work was supported by Wellcome Trust grant [106207] and European Research
585 Council grant [646891] to A.E.F.

586

587 **DATA AVAILABILITY**

588 Data available in supplementary material and in GenBank accession numbers KF298275.2,
589 KF298276.2 and KF298284.2.

For Review Only

FIGURES

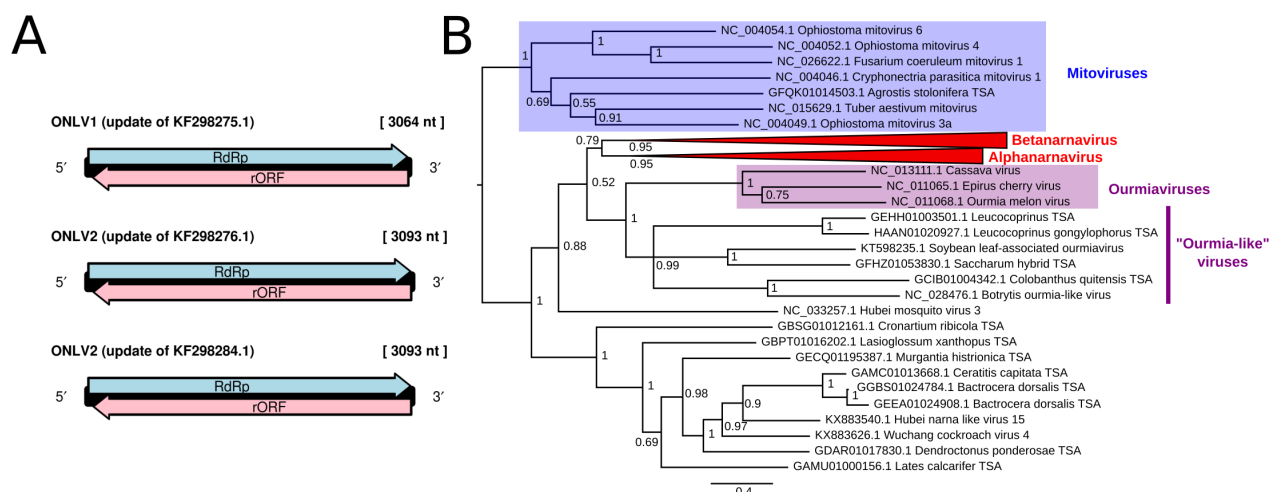


Figure 1. Narnavirus genome structure and taxonomy. **(A)** Updated narnaviral genome sequences: genome structures of ONLV1 and ONLV2. **(B)** Bayesian phylogeny of narnaviruses and selected outgroup sequences. Alphanarnaviruses and betanarnaviruses form clades with posterior probabilities of 0.95. The tree is rooted with the mitoviral clade as an outgroup. Accession numbers are for the nucleotide sequences from which the corresponding protein sequences were derived.

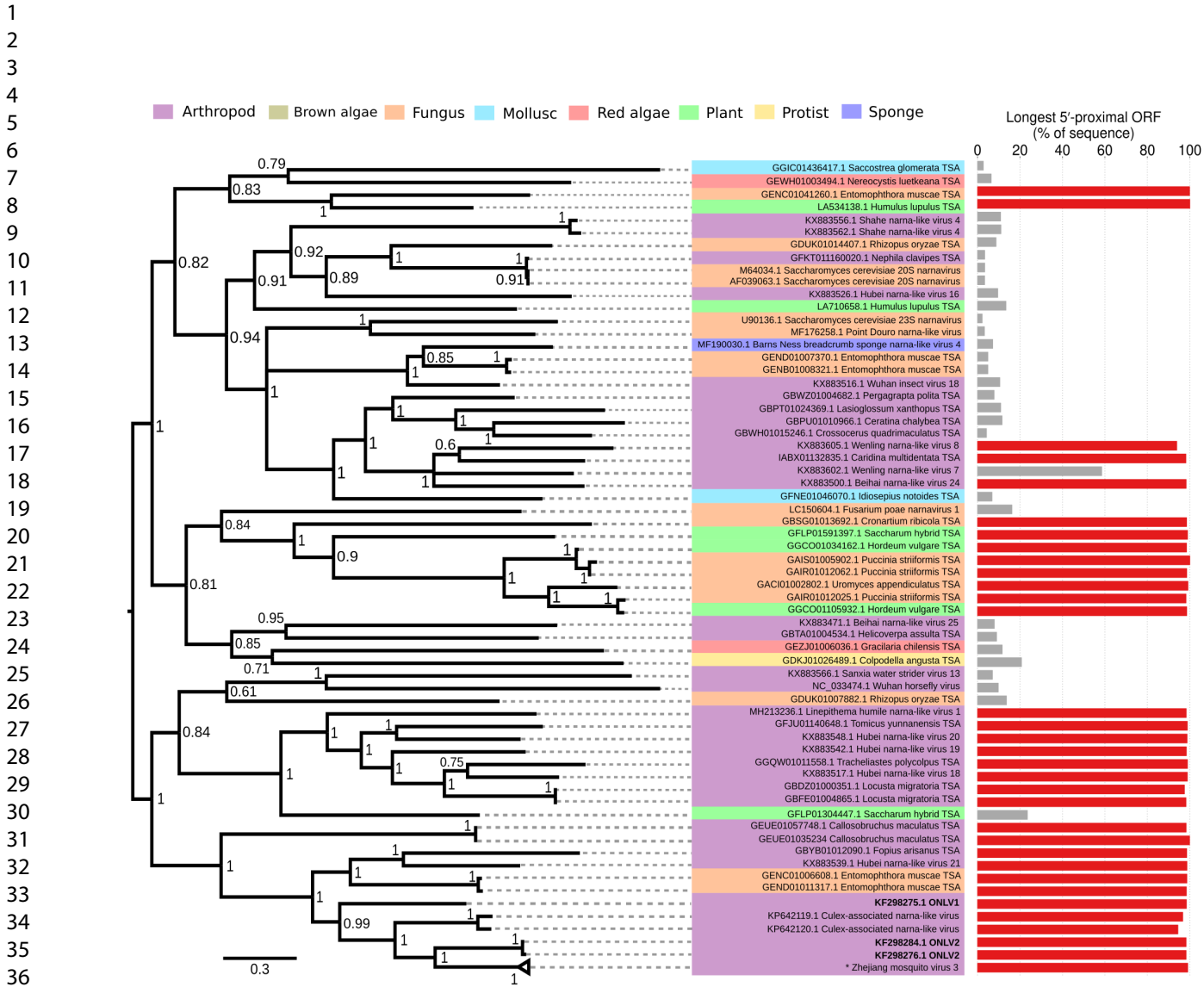


Figure 2. Midpoint-rooted Bayesian phylogenetic tree of alphanarnaviruses. The longest 5'-proximal ORF in the negative-strand R0 frame is shown in the bars to the right. Sequences with R0-frame ORFs occupying over 90% of the sequence are indicated with red bars. Nine highly similar sequences for Zhejiang mosquito virus 3 (indicated with an asterisk) are collapsed to a single taxon. Accession numbers are for the nucleotide sequences from which the corresponding protein sequences were derived.

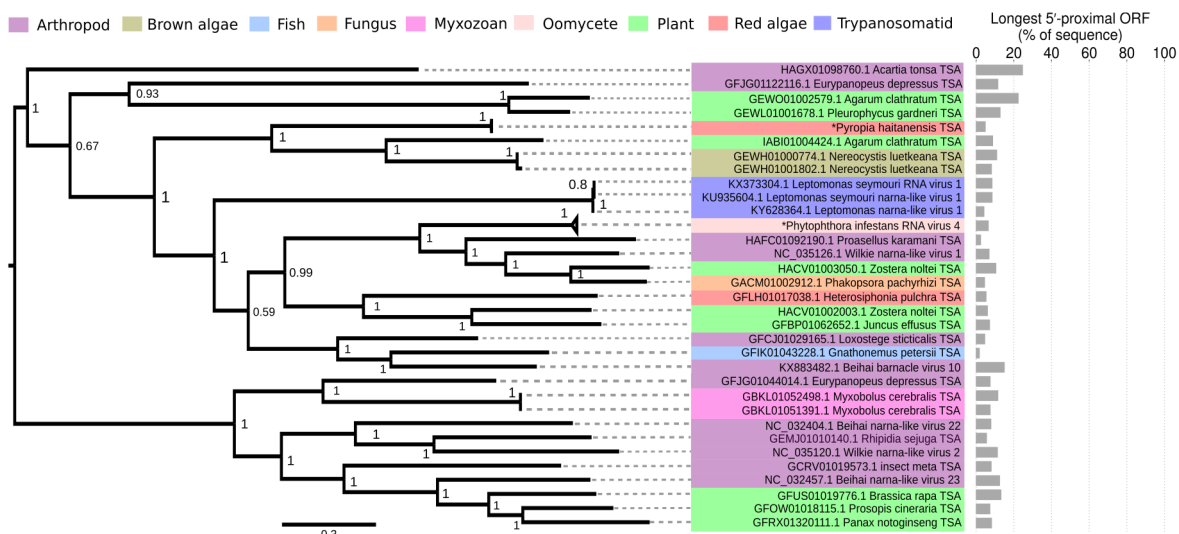


Figure 3. Midpoint-rooted Bayesian phylogenetic tree of betanarnaviruses. As in Figure 2, the longest 5'-proximal ORF in the negative-strand R0 frame is shown in the bars to the right. Five highly similar TSA sequences for *Pyropia haitanensis* and six highly similar sequences for *Phytophthora infestans* RNA virus 4 (both indicated with asterisks) are collapsed to single nodes. Accession numbers are for the nucleotide sequences from which the corresponding protein sequences were derived.

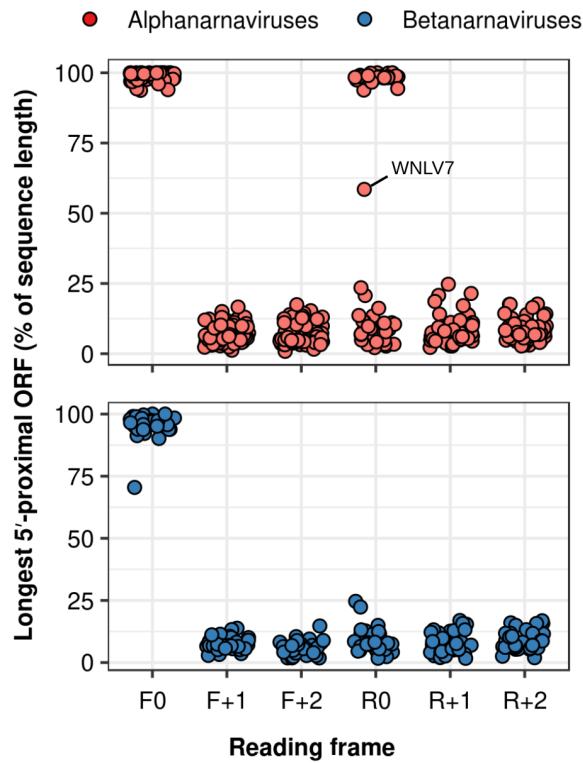


Figure 4. The longest 5'-proximal stop codon-free regions in each of the three possible positive-strand and negative-strand reading frames, for alphanarnaviruses (red) and betanarnaviruses (blue), as a percentage of the sequence length. Wenling narna-like virus 7 (WNLV7) has an intermediate-length R0-frame ORF, as indicated. Mean values are plotted for nodes with high levels of representation in the underlying data set (i.e. Zhejiang mosquito virus 3, *Phytophthora infestans* RNA virus 4, and TSA sequences from *Pyropia haitanensis*).

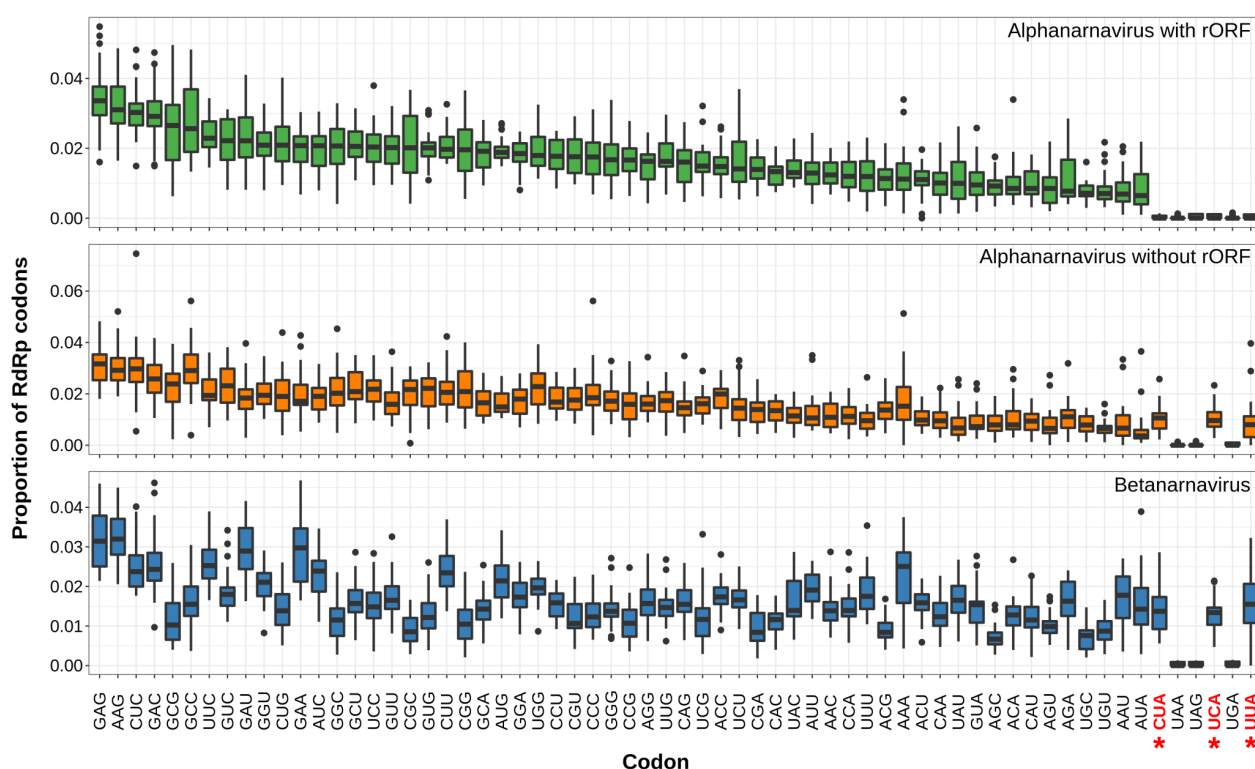


Figure 5. Box plots of codon usage (proportion of total codons) in RdRp ORFs. The upper and lower hinges are located at the first and third quartiles, respectively; the median values are indicated as horizontal lines; and the whiskers extend from the hinges to the furthest data points within 1.5 times the interquartile range (IQR) of the hinges. Data more than 1.5 times the IQR from the hinges are drawn as individual points (outliers). Codons are ordered according to their median frequency in alphanarnaviruses that contain the rORF. Alphanarnaviruses that contain the rORF show a specific avoidance of the three codons marked with red asterisks (CUA, UUA and UCA), which correspond to the reverse complements of stop codons.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

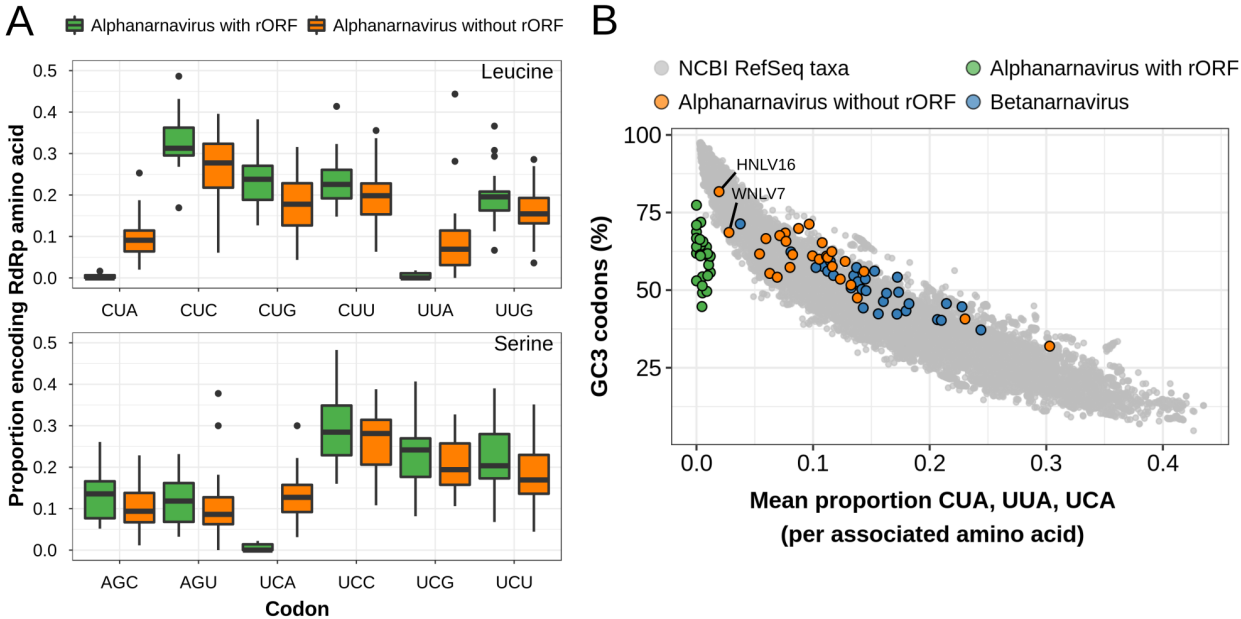


Figure 6. Specific selection against CUA, UUA and UCA codons in rORF-containing alphanarnaviruses. **(A)** Box plots of codon usage bias (proportion of codons encoding the amino acid) for leucine and serine. Box plots are drawn as in Figure 5. **(B)** Comparison of mean usage of CUA, UUA and UCA (per associated amino acid) with the third-position GC (GC3) content of codons. Prokaryotic and eukaryotic taxa in NCBI RefSeq are shown as grey points. The RdRp ORFs of non-rORF alphanarnaviruses, rORF alphanarnaviruses and betanarnaviruses are indicated with orange, green and blue points, respectively. The Wenling narna-like virus 7 (WNLV7) and Hubei narna-like virus 16 (HNLV16) sequences are indicated.

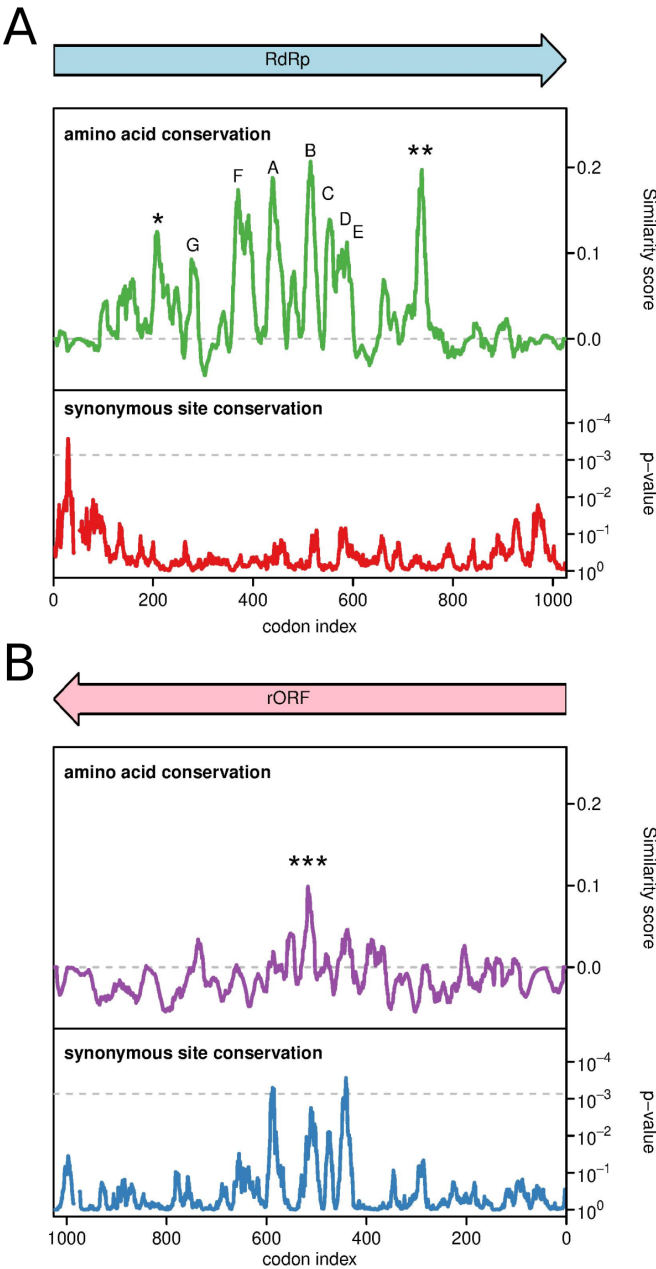


Figure 7. Amino acid conservation and synonymous site conservation in rORF-containing alphanarnaviruses. **(A)** RdRp amino acid conservation, and the corresponding synonymous site conservation analysis. In the latter, the grey dashed line indicates an approximate 5% false positive threshold after correcting for multiple tests (i.e. ~69 non-overlapping/independent 15-codon windows in the 1027-codon ORF). **(B)** The equivalent plots for the rORF. In each case, conservation was assessed in a 15 codon/amino acid window. Motifs A–G in the RdRp amino acid sequence are indicated with letters. Additional consensus motifs indicated with asterisks are as

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

645 follows: * = R&UP, ** = Pxx[L/V]GGx[G/N]xP and *** = QvxExExxPREREAH; U = I, L, V or
646 M, & = I, L, V, M, A, P, G, F, W or Y.

For Review Only

647 **REFERENCES**

- Athey J, Alexaki A, Osipova E, Rostovtsev A, Santana-Quintero LV, Katneni U, Simonyan V, Kimchi-Sarfaty C. 2017. A new and updated resource for codon usage tables. *BMC Bioinformatics* 18:391–391.
- Baltimore D. 1971. Expression of animal virus genomes. *Bacteriol Rev* 35:235.
- Belshaw R, Pybus OG, Rambaut A. 2007. The evolution of genome compression and genomic novelty in RNA viruses. *Genome Res* 17:1496–1504.
- Bicknell AA, Ricci EP. 2017. When mRNA translation meets decay. *Biochem Soc Trans* 45:339–351.
- Blumenthal T, Carmichael GG. 1979. RNA replication: function and structure of Qbeta-replicase. *Annu Rev Biochem* 48:525–548.
- den Boon JA, Ahlquist P. 2010. Organelle-like membrane compartmentalization of positive-strand RNA virus replication factories. *Annu Rev Microbiol* 64:241–256.
- Borowiec ML. 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4:e1660–e1660.
- Brandes N, Linial M. 2016. Gene overlapping and size constraints in the viral world. *Biol Direct* 11:26–26.
- Cai G, Myers K, Fry WE, Hillman BI. 2012. A member of the virus family *Narnaviridae* from the plant pathogenic oomycete *Phytophthora infestans*. *Arch Virol* 157:165–169.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

Chandler JA, Liu RM, Bennett SN. 2015. RNA shotgun metagenomic sequencing of northern California (USA) mosquitoes uncovers viruses, bacteria, and fungi. Front Microbiol 6:185–185.

Chung BY-W, Miller WA, Atkins JF, Firth AE. 2008. An overlapping essential gene in the *Potyviridae*. Proc Natl Acad Sci USA 105:5897.

Cook S, Chung BY-W, Bass D, Moureau G, Tang S, McAlister E, Culverwell CL, Glücksman E, Wang H, Brown TDK, et al. 2013. Novel virus discovery and genome reconstruction from field RNA samples reveals highly divergent viruses in dipteran hosts. PLoS One 8:e80720–e80720.

DeRisi JL, Huber G, Kistler A, Retallack H, Wilkinson M, Yllanes D. 2019. An exploration of ambigrammatic sequences in narnaviruses. Sci Rep 9:17982.

Dobkin C, Mills DR, Kramer FR, Spiegelman S. 1979. RNA replication: required intermediates and the dissociation of template, product, and Q beta replicase. Biochemistry 18:2038–2044.

Dolja VV, Koonin EV. 2011. Common origins and host-dependent diversity of plant and animal viromes. Curr Opin Virol 1:322–331.

Dolja VV, Koonin EV. 2018. Metagenomics reshapes the concepts of RNA virus evolution by revealing extensive horizontal virus transfer. Virus Research 244:36–52.

- 1
2
3
4 Donaire L, Rozas J, Ayllón MA. 2016. Molecular characterization of Botrytis ourmia-like virus, a
5
6 mycovirus close to the plant pathogenic genus *Ourmiavirus*. *Virology* 489:158–164.
7
8
9
10 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput.
11
12 *Nucleic Acids Res* 32:1792–1797.
13
14
15 Ertel KJ, Benefield D, Castaño-Diez D, Pennington JG, Horswill M, den Boon JA, Otegui MS,
16
17 Ahlquist P. 2017. Cryo-electron tomography reveals novel features of a viral RNA
18
19 replication compartment. *Elife* 6:e25940.
20
21
22
23 Esteban LM, Rodriguez-Cousiño N, Esteban R. 1992. T double-stranded RNA (dsRNA) sequence
24
25 reveals that T and W dsRNAs form a new RNA family in *Saccharomyces cerevisiae*.
26
27 Identification of 23 S RNA as the single-stranded form of T dsRNA. *J Biol Chem*
28
29 267:10874–10881.
30
31
32
33 Esteban R, Fujimura T. 2003. Launching the yeast 23S RNA narnavirus shows 5' and 3' cis-acting
34
35 signals for replication. *Proc Natl Acad Sci U S A* 100:2568–2573.
36
37
38
39 Esteban R, Vega L, Fujimura T. 2005. Launching of the yeast 20 S RNA narnavirus by expressing
40
41 the genomic or antigenomic viral RNA in vivo. *J Biol Chem* 280:33725–33734.
42
43
44
45 Esteban R, Vega L, Fujimura T. 2008. 20S RNA narnavirus defies the antiviral activity of
46
47 SKI1/XRN1 in *Saccharomyces cerevisiae*. *J Biol Chem* 283:25812–25820.
48
49
50 Fang Y, Treffers EE, Li Y, Tas A, Sun Z, van der Meer Y, de Ru AH, van Veelen PA, Atkins JF,
51
52 Snijder EJ, et al. 2012. Efficient –2 frameshifting by mammalian ribosomes to synthesize
53
54 an additional arterivirus protein. *Proc Natl Acad Sci USA* 109:E2920.
55
56
57
58
59
60

- Finn RD, Clements J, Eddy SR. 2011. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 39:W29–W37.
- Firth AE. 2014. Mapping overlapping functional elements embedded within the protein-coding regions of RNA viruses. *Nucleic Acids Res* 42:12425–12439.
- Firth AE, Brierley I. 2012. Non-canonical translation in RNA viruses. *J Gen Virol* 93:1385–1409.
- Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28:3150–3152.
- Fujimura T, Esteban R. 2004. The bipartite 3'-cis-acting signal for replication is required for formation of a ribonucleoprotein complex in vivo between the viral genome and its RNA polymerase in yeast 23 S RNA Virus. *J Biol Chem* 279:44219–44228.
- Fujimura T, Esteban R. 2007. Interactions of the RNA polymerase with the viral genome at the 5'- and 3'-ends contribute to 20S RNA narnavirus persistence in yeast. *J Biol Chem* 282:19011–19019.
- Fujimura T, Solórzano A, Esteban R. 2005. Native replication intermediates of the yeast 20 S RNA virus have a single-stranded RNA backbone. *J Biol Chem* 280:7398–7406.
- Göertz GP, Miesen P, Overheul GJ, van Rij RP, van Oers MM, Pijlman GP. 2019. Mosquito small RNA responses to West Nile and insect-specific virus infections in *Aedes* and *Culex* mosquito cells. *Viruses* 11:271.
- Harvey E, Rose K, Eden JS, Lawrence A, Doggett SL, Holmes EC. 2019. Identification of diverse arthropod associated viruses in native Australian fleas. *Virology* 535:189–199.

- Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.
- Irigoyen N, Firth AE, Jones JD, Chung BY-W, Siddell SG, Brierley I. 2016. High-resolution analysis of coronavirus gene expression by RNA sequencing and ribosome profiling. *PLoS Pathog* 12:e1005473–e1005473.
- Kadowaki K, Halvorson HO. 1971. Appearance of a new species of ribonucleic acid during sporulation in *Saccharomyces cerevisiae*. *J Bacteriol* 105:826.
- Kapoor A, Simmonds P, Lipkin WI, Zaidi S, Delwart E. 2010. Use of nucleotide composition analysis to infer hosts for three novel picorna-like viruses. *J Virol* 84:10322–10328.
- Katzourakis A, Gifford RJ. 2010. Endogenous viral elements in animal genomes. *PLoS Genet* 6:e1001191–e1001191.
- Keese PK, Gibbs A. 1992. Origins of genes: "big bang" or continuous creation? *Proc Natl Acad Sci U S A* 89:9489–9493.
- Koonin EV, Dolja VV. 2014. Virus world as an evolutionary network of viruses and capsidless selfish elements. *Microbiol Mol Biol Rev* 78:278.
- Koonin EV, Dolja VV, Krupovic M. 2015. Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology* 479–480:2–25.
- Kopeck BG, Perkins G, Miller DJ, Ellisman MH, Ahlquist P. 2007. Three-dimensional analysis of a viral RNA replication complex reveals a virus-induced mini-organelle. *PLoS Biol* 5:e220–e220.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Kurosaki T, Popp MW, Maquat LE. 2019. Quality and quantity control of gene expression by nonsense-mediated mRNA decay. *Nat Rev Mol Cell Biol* 20:406–420.

Lorenz R, Bernhart SH, Höner Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA Package 2.0. *Algorithms Mol Biol* 6:26.

Loughran G, Firth AE, Atkins JF. 2011. Ribosomal frameshifting into an overlapping gene in the 2B-encoding region of the cardiovirus genome. *Proc Natl Acad Sci USA* 108:E1111.

Lulla V, Dinan AM, Hosmillo M, Chaudhry Y, Sherry L, Irigoyen N, Nayak KM, Stonehouse NJ, Zilbauer M, Goodfellow I, et al. 2019. An upstream protein-coding region in enteroviruses modulates virus infection in gut epithelial cells. *Nat Microbiol* 4:280–292.

Lye L-F, Akopyants NS, Dobson DE, Beverley SM. 2016. A narnavirus-like element from the trypanosomatid protozoan parasite *Leptomonas seymouri*. *Genome Announc* 4:e00713–16.

Naphtine S, Ling R, Finch LK, Jones JD, Bell S, Brierley I, Firth AE. 2017. Protein-directed ribosomal frameshifting temporally regulates gene expression. *Nat Commun* 8:15582.

Novak JE, Kirkegaard K. 1991. Improved method for detecting poliovirus negative strands used to demonstrate specificity of positive-strand encapsidation and the ratio of positive to negative strands in infected cells. *J Virol* 65:3384–3387.

Pavesi A, Vianelli A, Chirico N, Bao Y, Blinkova O, Belshaw R, Firth A, Karlin D. 2018. Overlapping genes and the proteins they encode differ significantly in their sequence composition from non-overlapping genes. *PLoS One* 13:e0202513–e0202513.

Rancurel C, Khosravi M, Dunker AK, Romero PR, Karlin D. 2009. Overlapping genes produce proteins with unusual sequence properties and offer insight into de novo protein creation. *J Virol* 83:10719–10736.

Rastgou M, Habibi MK, Izadpanah K, Masenga V, Milne RG, Wolf YI, Koonin EV, Turina M. 2009. Molecular characterization of the plant virus genus *Ourmiavirus* and evidence of inter-kingdom reassortment of viral genome segments as its possible route of origin. *J Gen Virol* 90:2525–2535.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* 16:276–277.

Rodríguez-Cousiño N, Esteban LM, Esteban R. 1991. Molecular cloning and characterization of W double-stranded RNA, a linear molecule present in *Saccharomyces cerevisiae*. Identification of its single-stranded RNA form as 20 S RNA. *J Biol Chem* 266:12772–12778.

Rodríguez-Cousiño N, Solórzano A, Fujimura T, Esteban R. 1998. Yeast positive-stranded virus-like RNA replicons: 20 S and 23 S RNA terminal nucleotide sequences and 3' end secondary structures resemble those of RNA coliphages. *J Biol Chem* 273:20363–20371.

Romero-Brey I, Bartenschlager R. 2014. Membranous replication factories induced by plus-strand RNA viruses. *Viruses* 6:2826–2857.

Schlub TE, Buchmann JP, Holmes EC. 2018. A simple method to detect candidate overlapping genes in viruses using single genome sequences. *Mol Biol Evol* 35:2572–2581.

- Shi M, Lin X-D, Tian J-H, Chen L-J, Chen X, Li C-X, Qin X-C, Li J, Cao J-P, Eden J-S, et al. 2016. Redefining the invertebrate RNA virosphere. *Nature* 540:539.
- Shi M, Neville P, Nicholson J, Eden J-S, Imrie A, Holmes EC. 2017. High-resolution metatranscriptomics reveals the ecological dynamics of mosquito-associated RNA viruses in Western Australia. *J Virol* 91:e00680–17.
- Shulla A, Randall G. 2015. Spatiotemporal analysis of hepatitis C virus infection. *PLoS Pathog* 11:e1004758–e1004758.
- Shulla A, Randall G. 2016. (+) RNA virus replication compartments: a safe home for (most) viral replication. *Curr Opin Microbiol* 32:82–88.
- Smirnova E, Firth AE, Miller WA, Scheidecker D, Brault V, Reinbold C, Rakotondrafara AM, Chung BY-W, Ziegler-Graff V. 2015. Discovery of a small non-AUG-initiated ORF in polioviruses and luteoviruses that is required for long-distance movement. *PLoS Pathog* 11:e1004868–e1004868.
- Solórzano A, Rodríguez-Cousiño N, Esteban R, Fujimura T. 2000. Persistent yeast single-stranded RNA viruses exist in vivo as genomic RNA·RNA polymerase complexes in 1:1 stoichiometry. *J Biol Chem* 275:26428–26435.
- Sztuba-Solińska J, Stollar V, Bujarski JJ. 2011. Subgenomic messenger RNAs: Mastering regulation of (+)-strand RNA virus life cycle. *Virology* 412:245–255.
- Takeshita D, Tomita K. 2012. Molecular basis for RNA polymerization by Q β replicase. *Nat Struct Mol Biol* 19:229–237.

- Turina M, Hillman BI, Izadpanah K, Rastgou M, Rosa C, ICTV Report Consortium. 2017. ICTV virus taxonomy profile: *Ourmiavirus*. J Gen Virol 98:129–130.
- te Velhuis AJW. 2014. Common and unique features of viral RNA-dependent polymerases. Cell Mol Life Sci 71:4403–4420.
- Viljakainen L, Holmberg I, Abril S, Jurvansuu J. 2018. Viruses of invasive Argentine ants from the European Main supercolony: characterization, interactions and evolution. J Gen Virol 99:1129–1140.
- Wejksnora PJ, Haber JE. 1978. Ribonucleoprotein particle appearing during sporulation in yeast. J Bacteriol 134:246.
- Wesolowski M, Wickner RB. 1984. Two new double-stranded RNA molecules showing non-mendelian inheritance and heat inducibility in *Saccharomyces cerevisiae*. Mol Cell Biol 4:181.
- Wickner RB, Fujimura T, Esteban R. 2013. Viruses and prions of *Saccharomyces cerevisiae*. Adv Virus Res 86:1–36.
- Wolf YI, Kazlauskas D, Iranzo J, Lucía-Sanz A, Kuhn JH, Krupovic M, Dolja VV, Koonin EV. 2018. Origins and evolution of the global RNA virome. MBio 9:e02329–18.
- Wu J, Liu W, Gong P. 2015. A structural overview of RNA-dependent RNA polymerases from the *Flaviviridae* family. Int J Mol Sci 16:12943–12957.
- Zimmermann L, Stephens A, Nam SZ, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. 2018. A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. J Mol Biol 430:2237–2243.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Zinder ND. 1980. Portraits of viruses: RNA phage. Intervirology 13:257–270.

For Review Only

SUPPLEMENTARY MATERIAL

A case for a negative-strand coding sequence in a group of positive-sense RNA viruses

Adam M. Dinan, Nina I. Lukhovitskaya, Ingrida Olendraite and Andrew E. Firth

Division of Virology, Department of Pathology, University of Cambridge, Tennis Court Road,
Cambridge, CB2 1QP, UK

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Supplementary Table S1. Full list of sequences included in this study. Taxonomic information is taken from NCBI GenBank, where applicable.

Organism	Database	Accession	Taxonomy of source
Barns Ness breadcrumb sponge narna-like virus 4	GenBank	MF190030.1	Viruses, ssRNA viruses, ssRNA positive-strand viruses, no DNA stage, Narnaviridae
Beihai barnacle virus 10	GenBank	KX883482.1	Viruses, unclassified RNA viruses
Beihai narna-like virus 22	GenBank	NC_032404.1	Viruses, unclassified RNA viruses
Beihai narna-like virus 23	GenBank	NC_032457.1	Viruses, unclassified RNA viruses
Beihai narna-like virus 24	GenBank	KX883500.1	Viruses, unclassified RNA viruses
Beihai narna-like virus 25	GenBank	KX883471.1	Viruses, unclassified RNA viruses
Botrytis ourmia-like virus	GenBank	NC_028476.1	Viruses, ssRNA viruses, ssRNA positive-strand viruses, no DNA stage, Ourmiavirus
Cassava virus C	GenBank	NC_013111.1	Viruses, ssRNA viruses, ssRNA positive-strand viruses, no DNA stage, Ourmiavirus
Cryphonectria parasitica mitovirus 1-NB631	GenBank	NC_004046.1	Viruses, ssRNA viruses, ssRNA positive-strand viruses, no DNA stage, Narnaviridae
Epirus cherry virus	GenBank	NC_011065.1	Viruses, ssRNA viruses, ssRNA positive-strand viruses, no DNA stage, Ourmiavirus
Fusarium coeruleum mitovirus 1	GenBank	NC_026622.1	Viruses, ssRNA viruses, ssRNA positive-strand viruses, no DNA stage, Narnaviridae
Fusarium poae narnavirus 1	GenBank	LC150604.1	Viruses, ssRNA viruses, ssRNA positive-strand viruses, no DNA stage, Narnaviridae
Hubei mosquito virus 3	GenBank	NC_033257.1	Viruses, unclassified RNA viruses
Hubei narna-like virus 15	GenBank	KX883540.1	Viruses, unclassified RNA viruses
Hubei narna-like virus 16	GenBank	KX883526.1	Viruses, unclassified RNA viruses
Hubei narna-like virus 18	GenBank	KX883517.1	Viruses, unclassified RNA viruses
Hubei narna-like virus 19	GenBank	KX883542.1	Viruses, unclassified RNA viruses
Hubei narna-like virus 20	GenBank	KX883548.1	Viruses, unclassified RNA viruses
Hubei narna-like virus 21	GenBank	KX883539.1	Viruses, unclassified RNA viruses
Leptomonas Narna-like virus 1	GenBank	KY628364.1	Viruses, ssRNA viruses, ssRNA positive-strand viruses, no DNA stage, Narnaviridae
Leptomonas seymouri Narna-like virus 1	GenBank	KU935604.1	Viruses, ssRNA viruses, ssRNA positive-strand viruses, no DNA stage, Narnaviridae
Leptomonas seymouri RNA virus	GenBank	KX373304.1	Viruses, unclassified RNA viruses
Linepithema humile narna-like virus 1	GenBank	MH213236.1	Viruses, ssRNA viruses, ssRNA positive-strand viruses, no DNA stage, Narnaviridae
Narnaviridae environmental sample	GenBank	KP642119.1	Viruses, ssRNA viruses, ssRNA positive-strand viruses, no DNA stage, Narnaviridae
Narnaviridae environmental sample	GenBank	KP642120.1	Viruses, ssRNA viruses, ssRNA positive-strand viruses, no DNA stage, Narnaviridae
Ochlerotatus-associated narna-like virus 1	GenBank	KF298275.2	Viruses, environmental samples
Ochlerotatus-associated narna-like virus 2	GenBank	KF298276.2	Viruses, environmental samples
Ochlerotatus-associated narna-like virus 2	GenBank	KF298284.2	Viruses, environmental samples
Ophiostoma mitovirus 3a	GenBank	NC_004049.1	Viruses, ssRNA viruses, ssRNA positive-strand viruses, no DNA stage, Narnaviridae
Ophiostoma mitovirus 4	GenBank	NC_004052.1	Viruses, ssRNA viruses, ssRNA positive-strand

			viruses, no DNA stage, Narnaviridae
Ophiostoma mitovirus 6	GenBank	NC_004054.1	Viruses, ssRNA viruses, ssRNA positive-strand viruses, no DNA stage, Narnaviridae
Ourmia melon virus	GenBank	NC_011068.1	Viruses, ssRNA viruses, ssRNA positive-strand viruses, no DNA stage, Ourmiavirus
Phytophthora infestans RNA virus 4	GenBank	JN400241.1	Viruses, ssRNA viruses, ssRNA positive-strand viruses, no DNA stage, Narnaviridae
Phytophthora infestans RNA virus 4	GenBank	JN400242.1	Viruses, ssRNA viruses, ssRNA positive-strand viruses, no DNA stage, Narnaviridae
Phytophthora infestans RNA virus 4	GenBank	JN400243.1	Viruses, ssRNA viruses, ssRNA positive-strand viruses, no DNA stage, Narnaviridae
Phytophthora infestans RNA virus 4	GenBank	KU295722.1	Viruses, ssRNA viruses, ssRNA positive-strand viruses, no DNA stage, Narnaviridae
Phytophthora infestans RNA virus 4	GenBank	KU295726.1	Viruses, ssRNA viruses, ssRNA positive-strand viruses, no DNA stage, Narnaviridae
Phytophthora infestans RNA virus 4	GenBank	KU295727.1	Viruses, ssRNA viruses, ssRNA positive-strand viruses, no DNA stage, Narnaviridae
Point-Douro narna-like virus	GenBank	MF176258.1	Viruses, unclassified viruses
Saccharomyces 20S RNA narnavirus	GenBank	AF039063.1	Viruses, ssRNA viruses, ssRNA positive-strand viruses, no DNA stage, Narnaviridae
Saccharomyces 23S RNA narnavirus	GenBank	U90136.1	Viruses, ssRNA viruses, ssRNA positive-strand viruses, no DNA stage, Narnaviridae
Saccharomyces cerevisiae	GenBank	M64034.1	Eukaryota, Fungi, Dikarya, Ascomycota
Sanxia water strider virus 13	GenBank	KX883566.1	Viruses, unclassified RNA viruses
Shahe narna-like virus 4	GenBank	KX883556.1	Viruses, unclassified RNA viruses
Shahe narna-like virus 4	GenBank	KX883562.1	Viruses, unclassified RNA viruses
Soybean leaf-associated ourmiavirus 1	GenBank	KT598235.1	Viruses, ssRNA viruses, ssRNA positive-strand viruses, no DNA stage, Ourmiavirus
Tuber aestivum mitovirus	GenBank	NC_015629.1	Viruses, ssRNA viruses, ssRNA positive-strand viruses, no DNA stage, Narnaviridae
Wenling narna-like virus 7	GenBank	KX883602.1	Viruses, unclassified RNA viruses
Wenling narna-like virus 8	GenBank	KX883605.1	Viruses, unclassified RNA viruses
Wilkie narna-like virus 1	GenBank	NC_035126.1	Viruses, unclassified viruses
Wilkie narna-like virus 2	GenBank	NC_035120.1	Viruses, unclassified viruses
Wuchang cockroach Virus 4	GenBank	KX883626.1	Viruses, unclassified RNA viruses
Wuhan horsefly Virus 3	GenBank	NC_033474.1	Viruses, unclassified RNA viruses
Wuhan insect virus 18	GenBank	KX883516.1	Viruses, unclassified RNA viruses
Zhejiang mosquito virus 3	GenBank	KX883461.1	Viruses, unclassified RNA viruses
Zhejiang mosquito virus 3	GenBank	KX883537.1	Viruses, unclassified RNA viruses
Zhejiang mosquito virus 3	GenBank	KX883538.1	Viruses, unclassified RNA viruses
Zhejiang mosquito virus 3	GenBank	MF176257.1	Viruses, unclassified RNA viruses
Zhejiang mosquito virus 3	GenBank	MF176278.1	Viruses, unclassified RNA viruses
Zhejiang mosquito virus 3	GenBank	MF176306.1	Viruses, unclassified RNA viruses
Zhejiang mosquito virus 3	GenBank	MF176344.1	Viruses, unclassified RNA viruses
Zhejiang mosquito virus 3	GenBank	MF176365.1	Viruses, unclassified RNA viruses
Zhejiang mosquito virus 3	GenBank	MF176385.1	Viruses, unclassified RNA viruses
Acartia tonsa	TSA	HAGX01098760.1	Eukaryota, Metazoa, Ecdysozoa, Arthropoda

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Agarophyton chilense	TSA	GEZJ01006036.1	Eukaryota, Rhodophyta, Florideophyceae, Rhodymeniophycidae
Agarum clathratum	TSA	GEWO01002579.1	Eukaryota, Stramenopiles, PX clade, Phaeophyceae
Agarum clathratum	TSA	IABI01004424.1	Eukaryota, Stramenopiles, PX clade, Phaeophyceae
Agrostis stolonifera	TSA	GFQK01014503.1	Eukaryota, Viridiplantae, Streptophyta, Embryophyta
Bactrocera dorsalis	TSA	GEEA01024908.1	Eukaryota, Metazoa, Ecdysozoa, Arthropoda
Bactrocera dorsalis	TSA	GGBS01024784.1	Eukaryota, Metazoa, Ecdysozoa, Arthropoda
Brassica rapa subsp. chinensis	TSA	GFUS01019776.1	Eukaryota, Viridiplantae, Streptophyta, Embryophyta
Callosobruchus maculatus	TSA	GEUE01035234.1	Eukaryota, Metazoa, Ecdysozoa, Arthropoda
Callosobruchus maculatus	TSA	GEUE01057748.1	Eukaryota, Metazoa, Ecdysozoa, Arthropoda
Caridina multidentata	TSA	IABX01132835.1	Eukaryota, Metazoa, Ecdysozoa, Arthropoda
Ceratina chalybea	TSA	GBPU01010966.1	Eukaryota, Metazoa, Ecdysozoa, Arthropoda
Ceratitis capitata	TSA	GAMC01013668.1	Eukaryota, Metazoa, Ecdysozoa, Arthropoda
Colobanthus quitensis	TSA	GCIB01004342.1	Eukaryota, Viridiplantae, Streptophyta, Embryophyta
Colpodella angusta	TSA	GDKJ01026489.1	Eukaryota, Alveolata, Colpodellidae, Colpodella
Cronartium ribicola	TSA	GBSG01012161.1	Eukaryota, Fungi, Dikarya, Basidiomycota
Cronartium ribicola	TSA	GBSG01013692.1	Eukaryota, Fungi, Dikarya, Basidiomycota
Crossocerus quadrimaculatus	TSA	GBWH01015246.1	Eukaryota, Metazoa, Ecdysozoa, Arthropoda
Dendroctonus ponderosae	TSA	GDAR01017830.1	Eukaryota, Metazoa, Ecdysozoa, Arthropoda
Entomophthora muscae	TSA	GENB01008321.1	Eukaryota, Fungi, Zoopagomycota, Entomophthoromycotina
Entomophthora muscae	TSA	GENC01006608.1	Eukaryota, Fungi, Zoopagomycota, Entomophthoromycotina
Entomophthora muscae	TSA	GENC01041260.1	Eukaryota, Fungi, Zoopagomycota, Entomophthoromycotina
Entomophthora muscae	TSA	GEND01007370.1	Eukaryota, Fungi, Zoopagomycota, Entomophthoromycotina
Entomophthora muscae	TSA	GEND01011317.1	Eukaryota, Fungi, Zoopagomycota, Entomophthoromycotina
Eurypanopeus depressus	TSA	GFJG01044014.1	Eukaryota, Metazoa, Ecdysozoa, Arthropoda
Eurypanopeus depressus	TSA	GFJG01122116.1	Eukaryota, Metazoa, Ecdysozoa, Arthropoda
Fopius arisanus	TSA	GBYB01012090.1	Eukaryota, Metazoa, Ecdysozoa, Arthropoda
Gnathonemus petersii	TSA	GFIK01043228.1	Eukaryota, Metazoa, Chordata, Craniata
Helicoverpa assulta	TSA	GBTA01004534.1	Eukaryota, Metazoa, Ecdysozoa, Arthropoda
Heterosiphonia pulchra	TSA	GFLH01017038.1	Eukaryota, Rhodophyta, Florideophyceae, Rhodymeniophycidae
Hordeum vulgare	TSA	GGCO01034162.1	Eukaryota, Viridiplantae, Streptophyta, Embryophyta
Hordeum vulgare	TSA	GGCO01105932.1	Eukaryota, Viridiplantae, Streptophyta, Embryophyta
Humulus lupulus var. lupulus	TSA	LA534138.1	Eukaryota, Viridiplantae, Streptophyta, Embryophyta
Humulus lupulus var. lupulus	TSA	LA710658.1	Eukaryota, Viridiplantae, Streptophyta, Embryophyta
Idiosepius notoides	TSA	GFNE01046070.1	Eukaryota, Metazoa, Lophotrochozoa, Mollusca
insect metagenome	TSA	GCRV01019573.1	unclassified sequences, metagenomes, organismal metagenomes
Juncus effusus	TSA	GFBP01062652.1	Eukaryota, Viridiplantae, Streptophyta, Embryophyta

Lasioglossum xanthopus	TSA	GBPT01016202.1	Eukaryota, Metazoa, Ecdysozoa, Arthropoda
Lasioglossum xanthopus	TSA	GBPT01024369.1	Eukaryota, Metazoa, Ecdysozoa, Arthropoda
Lates calcarifer	TSA	GAMU01000156.1	Eukaryota, Metazoa, Chordata, Craniata
Leucoagaricus gongylophorus	TSA	HAAN01020927.1	Eukaryota, Fungi, Dikarya, Basidiomycota
Leucocoprinus sp. HH-2015a	TSA	GEHH01003501.1	Eukaryota, Fungi, Dikarya, Basidiomycota
Locusta migratoria	TSA	GBFE01004865.1	Eukaryota, Metazoa, Ecdysozoa, Arthropoda
Locusta migratoria manilensis	TSA	GBDZ01000351.1	Eukaryota, Metazoa, Ecdysozoa, Arthropoda
Loxostege sticticalis	TSA	GFCJ01029165.1	Eukaryota, Metazoa, Ecdysozoa, Arthropoda
Murgantia histrionica	TSA	GECQ01195387.1	Eukaryota, Metazoa, Ecdysozoa, Arthropoda
Myxobolus cerebraalis	TSA	GBKL01051391.1	Eukaryota, Metazoa, Cnidaria, Myxozoa
Myxobolus cerebraalis	TSA	GBKL01052498.1	Eukaryota, Metazoa, Cnidaria, Myxozoa
Nephila clavipes	TSA	GFKT011160020.1	Eukaryota, Metazoa, Ecdysozoa, Arthropoda
Nereocystis luetkeana	TSA	GEWH01000774.1	Eukaryota, Stramenopiles, PX clade, Phaeophyceae
Nereocystis luetkeana	TSA	GEWH01001802.1	Eukaryota, Stramenopiles, PX clade, Phaeophyceae
Nereocystis luetkeana	TSA	GEWH01003494.1	Eukaryota, Stramenopiles, PX clade, Phaeophyceae
Panax notoginseng	TSA	GFRX01320111.1	Eukaryota, Viridiplantae, Streptophyta, Embryophyta
Pergagraptia polita	TSA	GBWZ01004682.1	Eukaryota, Metazoa, Ecdysozoa, Arthropoda
Phakopsora pachyrhizi Thai1	TSA	GACM01002912.1	Eukaryota, Fungi, Dikarya, Basidiomycota
Pleurophycus gardneri	TSA	GEWL01001678.1	Eukaryota, Stramenopiles, PX clade, Phaeophyceae
Proasellus karamani	TSA	HAFC01092190.1	Eukaryota, Metazoa, Ecdysozoa, Arthropoda
Prosopis cineraria	TSA	GFOW01018115.1	Eukaryota, Viridiplantae, Streptophyta, Embryophyta
Puccinia striiformis f. sp. tritici	TSA	GAIR01012025.1	Eukaryota, Fungi, Dikarya, Basidiomycota
Puccinia striiformis f. sp. tritici	TSA	GAIR01012062.1	Eukaryota, Fungi, Dikarya, Basidiomycota
Puccinia striiformis f. sp. tritici	TSA	GAIS01005902.1	Eukaryota, Fungi, Dikarya, Basidiomycota
Pyropia haitanensis	TSA	GADD01004427.1	Eukaryota, Rhodophyta, Bangiophyceae, Bangiales
Pyropia haitanensis	TSA	GFOL01000028.1	Eukaryota, Rhodophyta, Bangiophyceae, Bangiales
Pyropia haitanensis	TSA	GFOL01000185.1	Eukaryota, Rhodophyta, Bangiophyceae, Bangiales
Pyropia haitanensis	TSA	GFOL01002916.1	Eukaryota, Rhodophyta, Bangiophyceae, Bangiales
Pyropia haitanensis	TSA	GFOL01002977.1	Eukaryota, Rhodophyta, Bangiophyceae, Bangiales
Rhipidia sejuga	TSA	GEMJ01010140.1	Eukaryota, Metazoa, Ecdysozoa, Arthropoda
Rhizopus oryzae	TSA	GDUK01007882.1	Eukaryota, Fungi, Fungi incertae sedis, Mucoromycotina
Rhizopus oryzae	TSA	GDUK01014407.1	Eukaryota, Fungi, Fungi incertae sedis, Mucoromycotina
Saccharum hybrid cultivar	TSA	GFHZ01053830.1	Eukaryota, Viridiplantae, Streptophyta, Embryophyta
Saccharum hybrid cultivar	TSA	GFLP01304447.1	Eukaryota, Viridiplantae, Streptophyta, Embryophyta
Saccharum hybrid cultivar	TSA	GFLP01591397.1	Eukaryota, Viridiplantae, Streptophyta, Embryophyta
Saccostrea glomerata	TSA	GGIC01436417.1	Eukaryota, Metazoa, Lophotrochozoa, Mollusca
Tomicus yunnanensis	TSA	GFJU01140648.1	Eukaryota, Metazoa, Ecdysozoa, Arthropoda

Tracheliastes polycolpus	TSA	GGQW01011558.1	Eukaryota, Metazoa, Ecdysozoa, Arthropoda
Uromyces appendiculatus	TSA	GACI01002802.1	Eukaryota, Fungi, Dikarya, Basidiomycota
Zostera noltei	TSA	HACV01002003.1	Eukaryota, Viridiplantae, Streptophyta, Embryophyta
Zostera noltei	TSA	HACV01003050.1	Eukaryota, Viridiplantae, Streptophyta, Embryophyta

For Review Only

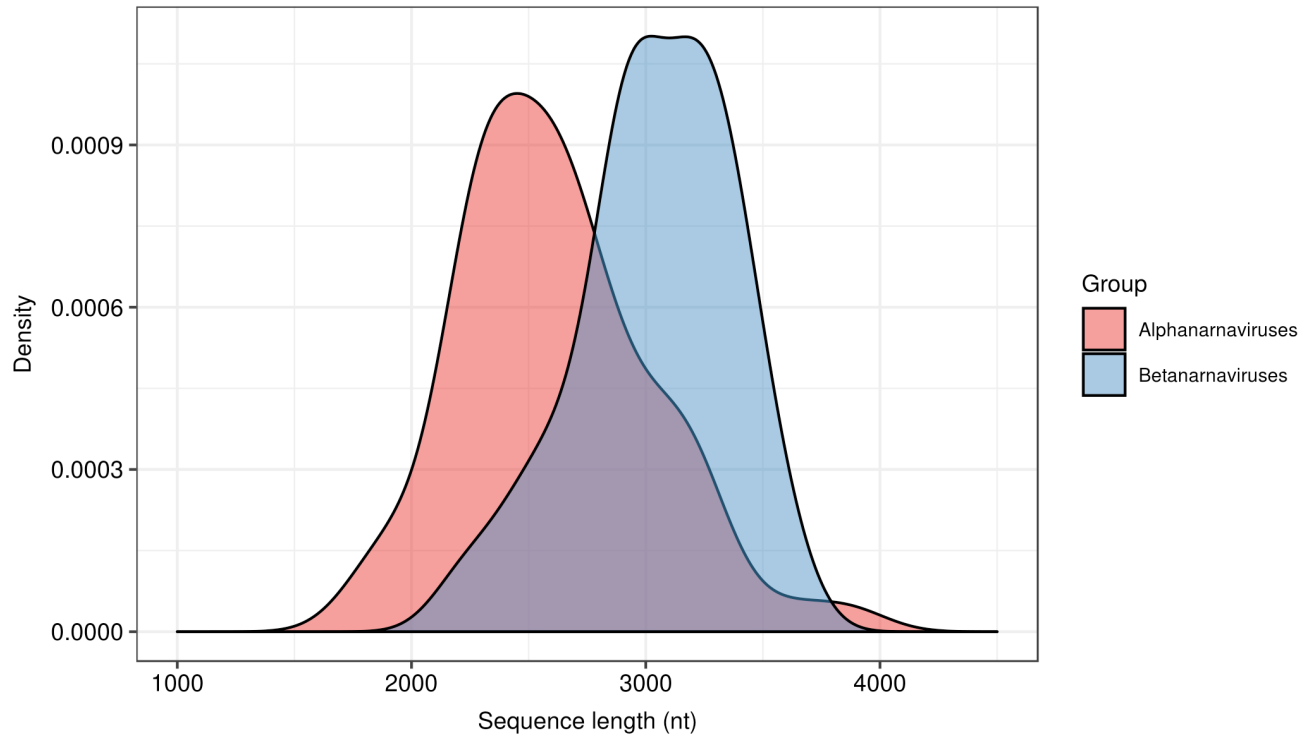


Figure S1. Gaussian kernel density estimates of the sequence length distributions for alphanaviruses ($n = 64$) and betanaviruses ($n = 33$) included in the analysis.

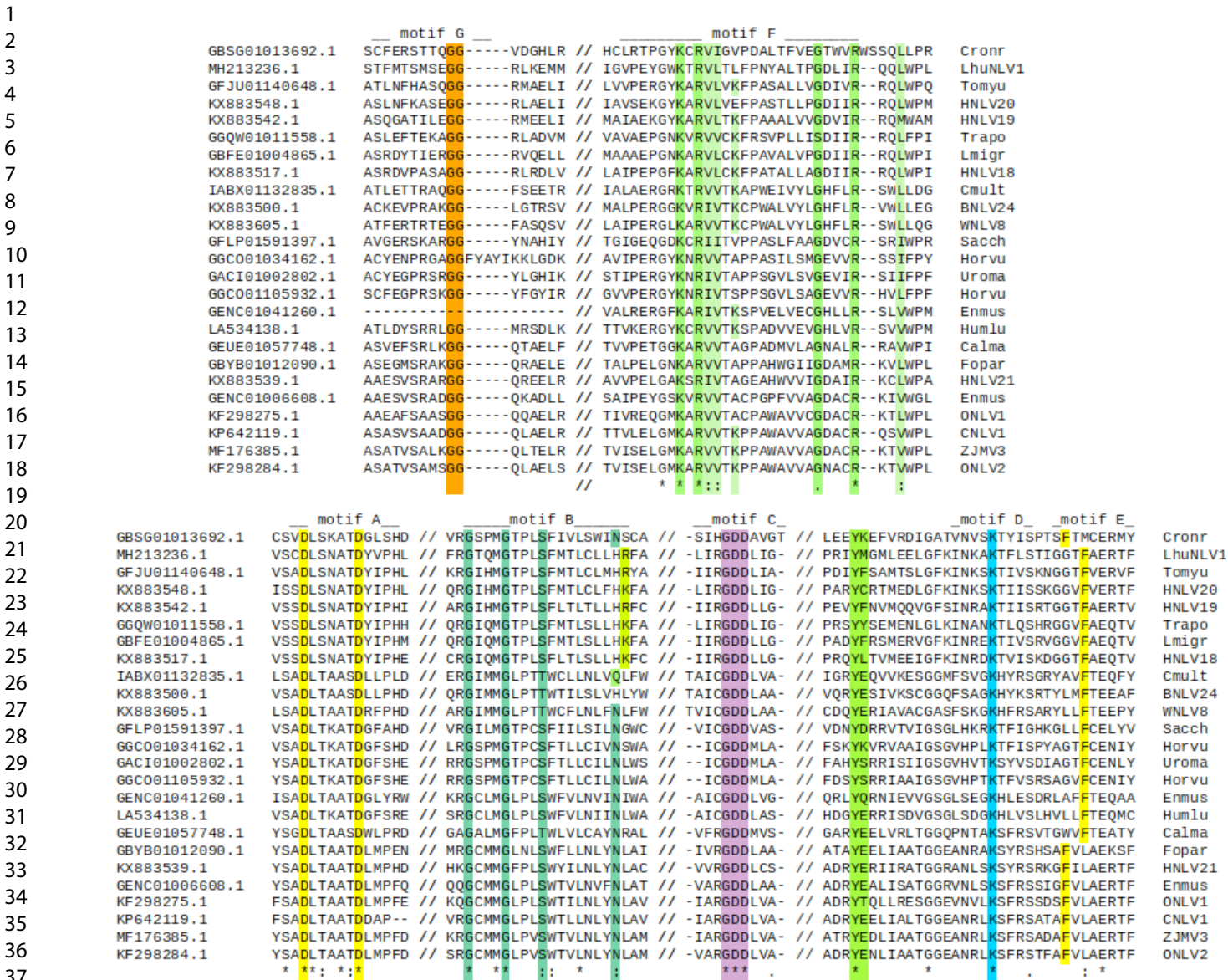


Figure S2. Alignment of RdRp protein-coding sequences of rORF-containing alphanarnaviruses. Known RdRp functional motifs are indicated and key functional residues previously described are highlighted. Key: Cronr = *Cronartium ribicola* TSA; LhuNLV1 = *Linepithema humile* narna-like virus 1; Tomyu = *Tomicus yunnanensis* TSA; HNLV = Hubei narna-like virus; Trapo = *Tracheliaestes polycolpus* TSA; Lmigr = *Locusta migratoria* TSA; BNLV = Beihai narna-like virus; Cmult = *Caridina multidentata* TSA; WNLV = Wenling narna-like virus; Sacch = *Saccharum* hybrid TSA; Horvu = *Hordeum vulgare* TSA; Uroma = *Uromyces appendiculatus* TSA; Enmus = *Entomophthora muscae* TSA; Humlu = *Humulus lupulus* TSA; Calma = *Callosobruchus maculatus* TSA; Fopar = *Fopius arisanus* TSA; ONLV = Ochlerotatus-associated narna-like virus; CNLV = Culex-associated narna-like virus; ZJMV3 = Zhejiang mosquito virus 3.

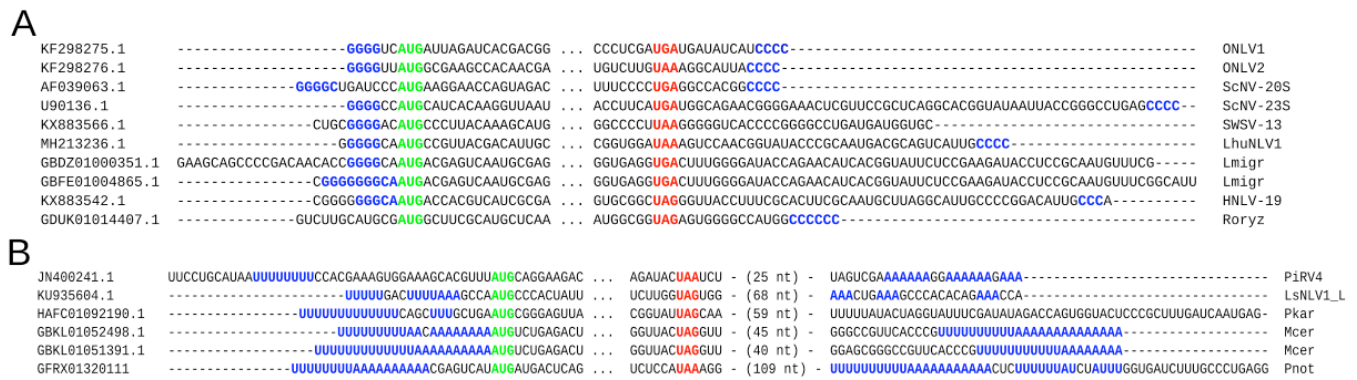


Figure S3. Alignments of (A) alphanarnaviral and (B) betanarnaviral genomic termini. The start and stop codons of the RdRp ORFs are shown in green and red, respectively. Complementary G/C- and A/U-rich regions are indicated in blue. Key: ONLV = Ochlerotatus-associated narna-like virus; ScNV-20S = *Saccharomyces cerevisiae* 20S narnavirus; ScNV-23S = *Saccharomyces cerevisiae* 23S narnavirus; SWSV-13 = *Sanxia water strider virus strain 13*; LhuNLV1 = *Linepithema humile* narna-like virus 1; Lmigr = *Locusta migratoria* TSA; HNLV-19 = Hubei narna-like virus 19; Roryz = *Rhizopus oryzae* TSA; PiRV4 = *Phytophthora infestans* RNA virus 4; LsNLV1_L = *Leptomonas seymouri* narna-like virus 1 segment L; Pkar = *Proasellus karamani* TSA; Mcer = *Myxobolus cerebralis* TSA; Pnot = *Panax notoginseng* TSA.

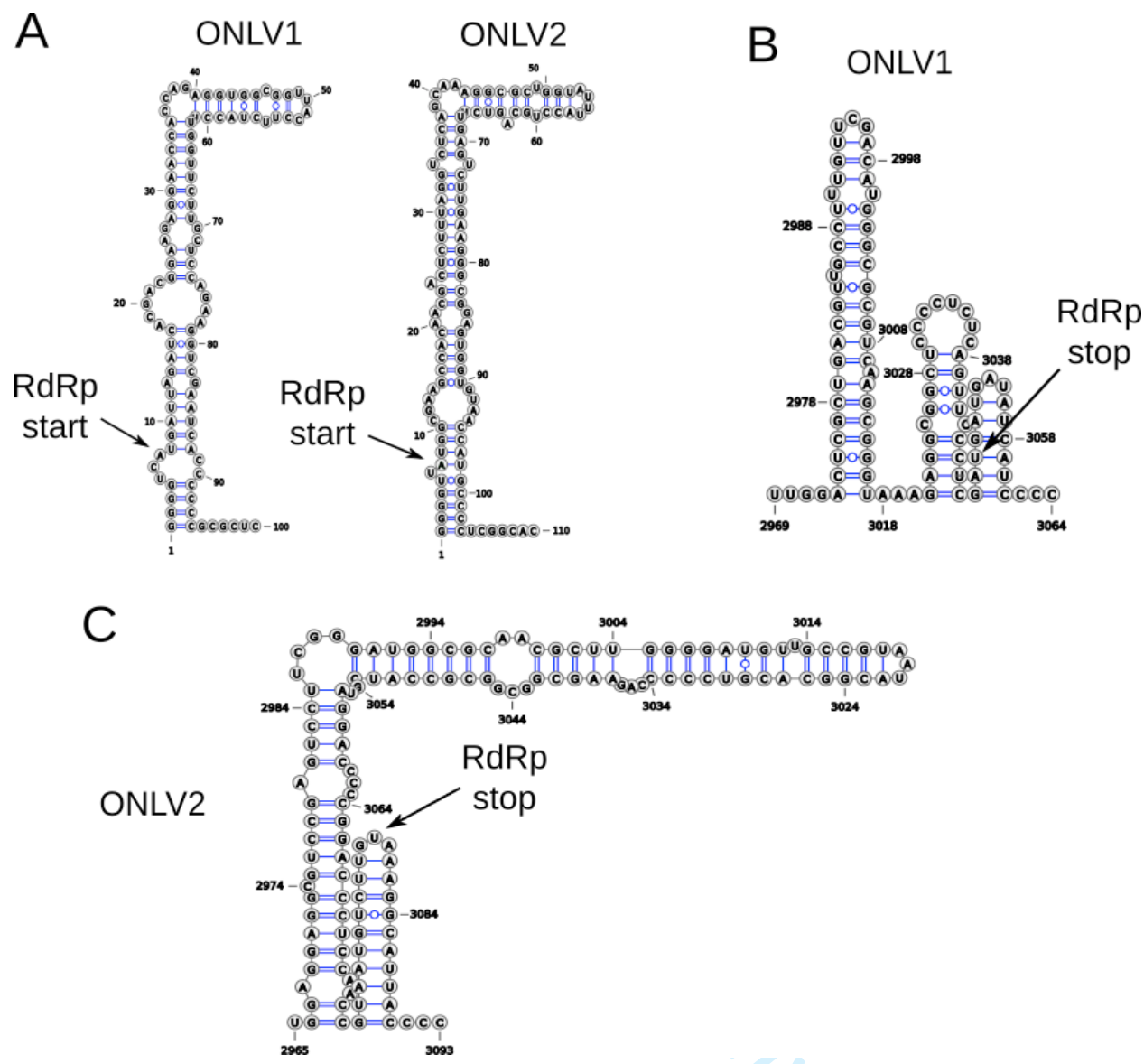


Figure S4. Predicted RNA secondary structures at the terminal regions of alphanarnaviral genomes. (A) 5' termini of *Ochlerotatus*-associated narna-like virus (ONLV) 1 and ONLV2; (B) 3' terminus of ONLV1; and (C) 3' terminus of ONLV2.

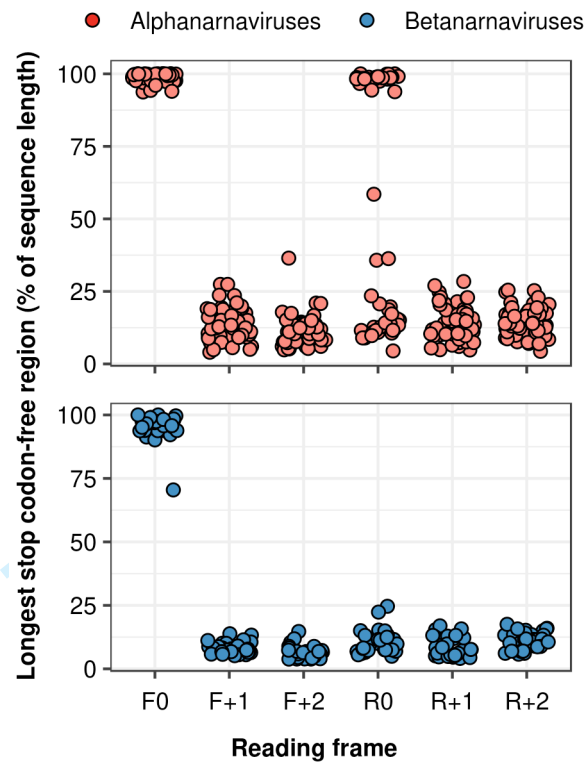


Figure S5. The longest stop codon-free regions in each of the three possible positive-strand and negative-strand reading frames (without the 5'-proximity restriction), for alphanarnaviruses (red) and betanarnaviruses (blue), as a percentage of the sequence length.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

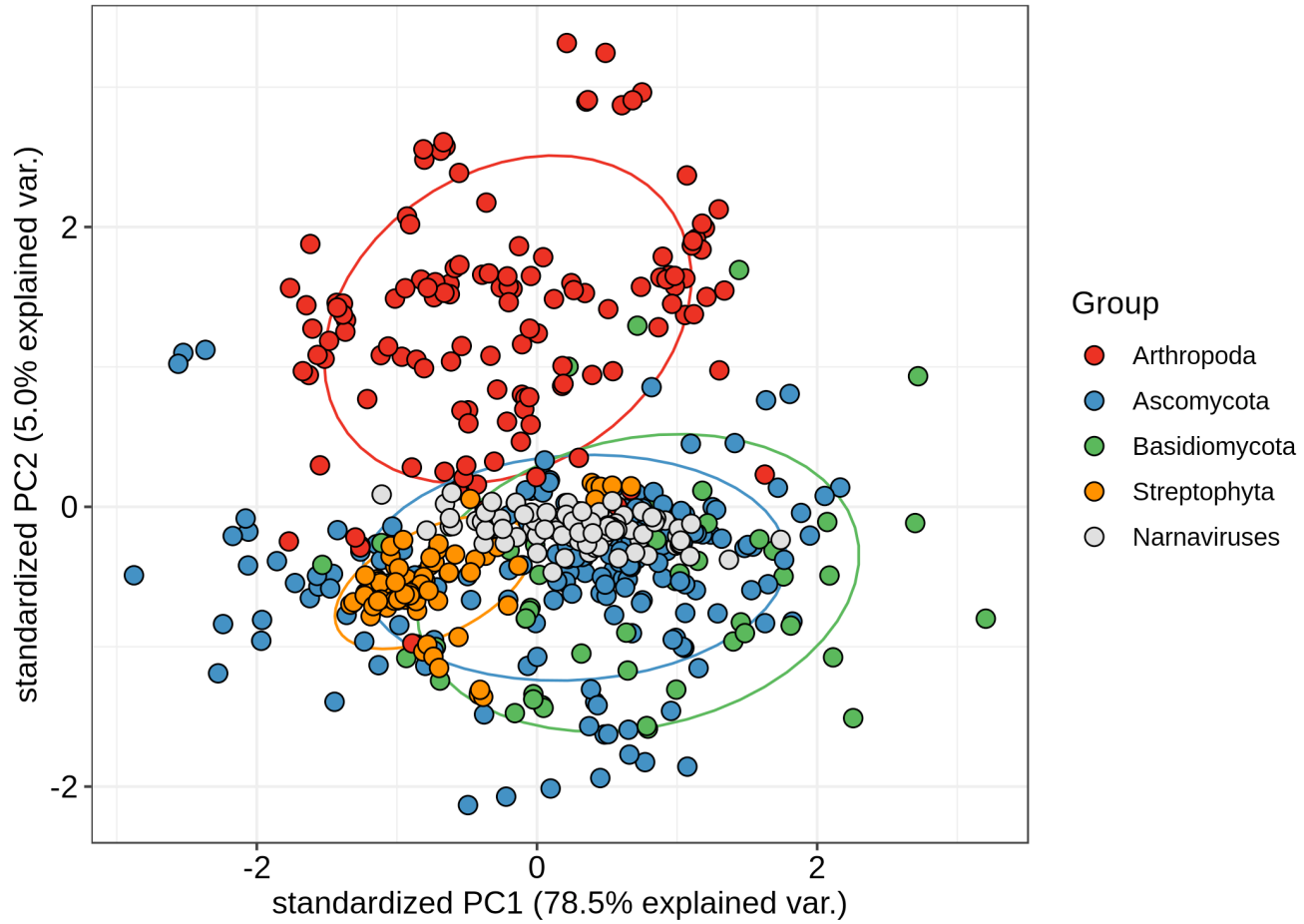


Figure S6. Principal component analysis of codon usage (per associated amino acid) in phyla that frequently co-occur with narnaviruses. Codon usage across NCBI RefSeq genomes was extracted from the latest release of the codon usage table database (CUTD). Ellipse contours are drawn for each group at a normal probability of 0.68.

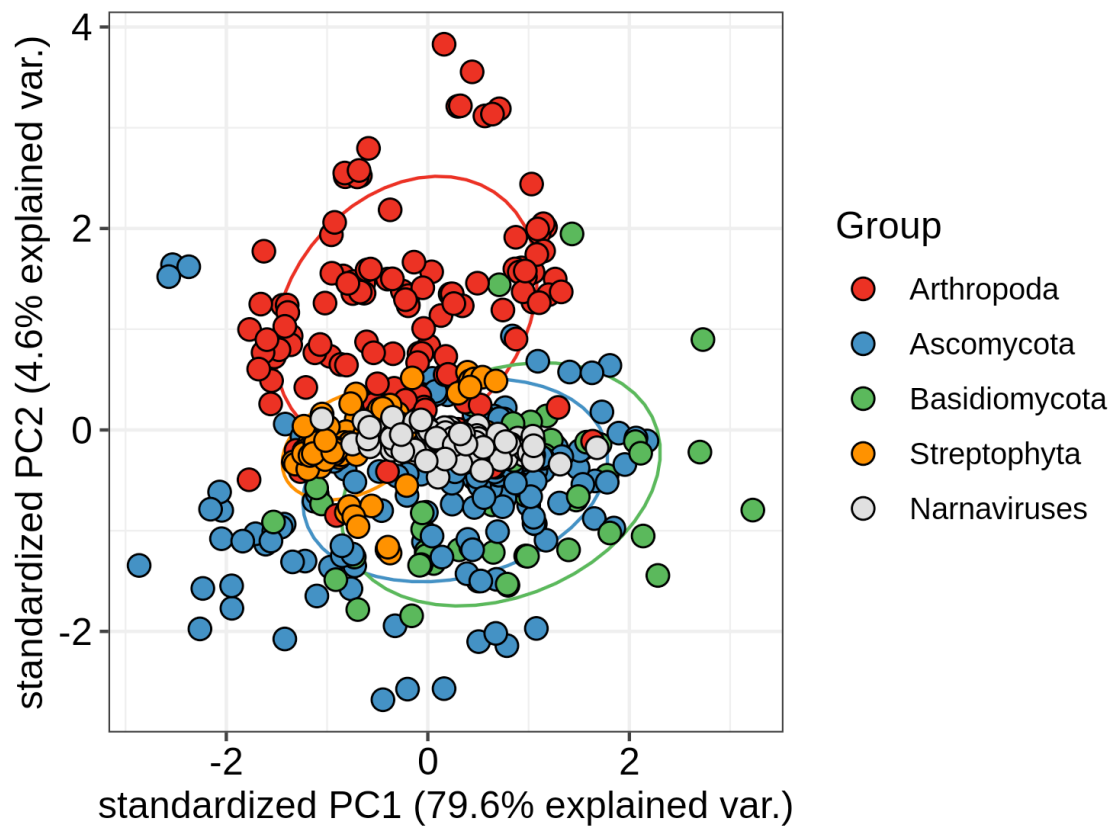


Figure S7. Principal component analysis of codon usage (per associated amino acid) in phyla that frequently co-occur with narnaviruses, excluding leucine and serine codons. Codon usage across NCBI RefSeq genomes was extracted from the latest release of the codon usage table database (CUTD). Ellipse contours are drawn for each group at a normal probability of 0.68.

>NC_004046.1

-----MAMIIHDPVYKLWYA-----XRAKSNLPG
LAPH-----KTESITTVLSLTNVKNKQAT
-----NSILLNKITDLVLGLKAFSIKS
KG----VRTT-----LRVNXHWITPKEFPRFVKLVVWC-----TR
TQEHEDSFMKIIG-KCDHIWQT-----AGPNFLFKYLK-----EVMRLSVRRIA---
-----NIELEPSKKIFVKLNK-----FRFPNIPLPICDQIIRDQ
---NDQVLWASKR-----LIICLLTILSVHRV-LPTKVVPDYSTIVDP-----
----FTGVSKTIDQK--LLRKAHLLNIKRVKQL-----KLKIT--GSMKAG
PN--GKIS--LLT-----SSVDALSFITQPTKIIFTYLD FSVRVYKFRGLLLWMWM--
-----M
CILL-----ITLPYAI--VSFML
GALI-----
----PIMGKLSV--VYDQA-GKARIVAITNSWIQTA-FYSLH-----LHVFK-LL
KNIDQ-----DGTFDQERPFKLLI-----KWLNEPTQ-----
----KF-YGF-----DLTAATDRLPI--DLQVDILNIIFKN-----
-SPGS-SWRSLLRIKYKS---PQGF--LTY-----
-----AVGQPMGAYSSFAMLALTHHVI----V---
QVA-----ALNSGFTTRF-----
-----TDY--CILGDD--IVIA--HDTVA--SEYLKLMET-LGLSISSGKSV
IS--SEF-TEFAKKL-----KGRNNF--DIFYR
SWF-----
SIIHFEKQ-----
-----ILHLCTV-----FELLRRGVCELY-----DLY
PQYINKL-----PKIYLR-----YNLLID-----W
VVVAFTNQILIGDRPRA--DGIR--L-----FDYFV-----
----GLEVIP-PL-----LRIMLHTIKKDWNGLWNSIKYTLNKGFFVVSQVRVGLPD-
----WTELFL-----LPILPSTYIIIRDY-----
-----CRSFNDLTKLFGWWL
LRF-----EESYQVSILD-----VIDRLAH
TSIPNLDIHDKKKVKLTLD--NLYKLSLIVNIPS-----GG
ARRY-----IEFLRFNGLKSPLIVERYIKDGI-----
RIEKPLTL-----
-----QGLHRSGDIQLGFKIS-----

>NC_004054.1

-----MKLKQLKLMKN-----

-----KTYQ-----IIRILLIVFFPSI-----KR
QTVILNIFMSKIN-KMIKN-----NGTLFTVKYLK-----ELRLHITKYISGEP
YRN-----SLNRVSVDKDGF-----TLCKELKVLVNGTY--LEK
---RFVLTITLSKLLIP--QKSESIPFSTKSITDHWSG-IDNISNEE----LD-----
----KSCSELNISTREVQWDVK-----NF-----KLLTKA

GPHG-----PQSLTWYHTIKLYDFNQWLGII-----
-----G
ILPK-----SVLDLFT--ETLSY
ASKL-----VLPEIKS
NIKSTKLIRRLSI--VHDPE-CKERVIAIFDYGSQMV-LKPIA-----DVLFD-LL
RNIPS-----DRTFTQSP-FFTHT-----DLDNKS-----
----KF-WSI-----DLSSATDRFPI--VFQKRVLQKILGK-----
----QMTDSWERIMIG---SKFL--A-----
-PDGD-----TVSYNCGQPMGAQSSWPMFTLAHHVI----V---
RVA-----ANRCGLSNF-----
-----DKY--IILGDD--IVIN---NDNVA--LKYMEIMND-FKVEISRNKTH
VS--NDT-YEFAKRW----I-----K--NKMEFF--PLPIR

GIVDNINN-----
-----KYIIFNILYSFFVEKGNTFLNKDTLL-----VCV
SKFIQLH--SLTLKKPIG-----LNKVKG-----ILYPFNFMLR
Y-RQNLCCTNEE--IRIFL--GSST---C-----KRDDYML-----
----PISAKDVSLELT---RVISAALVG-----MAYNAEKSLKNIYFDLKDLS--P-
----WIGDGF-----KTGKHPK-----
-----VM-IQSIYNSVKSLSDFGL
KMA-----QNKLTLAAMD-----
-----SLLL--DLDSISSSERIKY-----
-----IQMKQNICLSQKVRKELR-----FD
PLQMEQKA-----RA-----
-----MMLVK--HMKDLEG-----

>NC_004052.1

-----MKRNNLQIIIKR-----

-----LILH-----IFKINLSVEIDKF-----
----LGFINHLRKS-----NGLLYTIKYM-----AAKLHITRYMCGKP
LYS-----NNENVALDKTGFP-LRF-----WYLKRLVNDNPRAL-----
---LTLLTYTRRI---VP-----NKSESKARIVKLST-ITDPYKGK-----
----VYTIPKWFILD---FISK--YNLSSTKPI-----YTDND--HYLSIK
GSPN-GKASMSSLYS-----IISFNSSNIRYLFNIVGDYQLVLNKFYQ-----

-----DLSQFYT--KYINR
DKLG-----
-----LGKLSI--VHDPE-LKERVIAIMVDYTTQFA-LRPIH-----NILLN-NL
SKLPCD-----RTFTQDPF-HKWND-----DHKERY-----
-----HSL-----DLAATDRFPI--FLQQKLISLIFND-YE-----
-----FGKNWRNL-LV---DRNY--D-Y-----
-QGIS-----YRYSVGQPMGAYTSWAAFTLTHHLV----V---
HWA-----AELAGLKNF-----

1
2
3 -----KDY--IILGDD--IVIK---NNKVA--QIYINLMTK-WGVDISLSKTH
4 VS--YDT-YEFAKRW-----IKNGKE--ISGIS
5 LK-----
6 GILTNI RH-----
7
8 -----IHVVYM-----NIFTYLQRIPSLNVDILTC--VGKLYGY-----LLI
9 RNRIKSP---NTIKRSLY-----DFHHSI-----RYSFGL-----LNYEEIRNY
10 L-HNKFPFDNYYAWPERLVHSKLN--EI-----FKLEMVE-----
11 ----SAKSFSKDFMN----QSTMLINTVTDNEIMVQWPLYK---GFMNHIEK LK--D-
12 ----YIKSKQ-----NQHDIDL LDLMQNL-----RFQNLDSIVKK-----
13 -----LRNSYTNLIMLDKFWK
14 SA-----FNREYRDLERESILTIEKQESNMMSRIWDMALSYR
15 TSPMSYSTLT FETDSEFYM---MPSIWDMASSSTTGPKP-----FT
16 TATF-----KTDFVGLSTFDDKLLKDLENLKID-----IT
17 LRTGKYTN-----KTQPLE
18 S-----KIHQH--PIECIEDNKVPNFNNIK-----
19 -----
20
21 >NC_026622.1
22 -----
23 -----MP-----LRDSHWSFN
24 SISN-----YSVRLPPDKWMRSLIKNYNS
25 -----
26 -----IMKN-----IKLHLLLVKRLITCIFI-----DH
27 TIITLFIKEIYRLWDK-----NGFTFMIKYMK-----SVRLHITRYICKKP
28 IHS-----NSSNVSVDKSGFP-SKF----LYLKPLIKEVKDIKI----
29 ---LLTLLSFTRG-----LKP--SKKEDSNIKYDLSS-ITSP-----
30 ----FKGTSGKSV PQ--WFIK--SFVETNKLYKKIP-----EYSVKDH-----YL
31 STKG-GPCG-----KSTWSSTKSHLFYKQDLILNIQNIFREGFKELFFT----
32 -----P
33 FLKN-----MHLSY
34 GVTR-----
35 -----WPNGKLSI--VKDPE-GKRRVIAMVDYHSQLA-LKSIH-----NDLLD-LL
36 SKFKCD-----RTFTQDPL-HNWSN-----NKE-----
37 ----SY-FSL-----DLSSATDRFPV--ELQKRLLAEIYQD-DK-----
38 -FATN--WMELL---LN---RDYI--G-----
39 -PGGE-----ICRYSVGQPMGAYSSWAAFTLTHHLV----V---
40 SWS-----AYKALKTKNF-----
41 -----DQY--IILGDD--IVIK---NNEIA--NIYRGQMMR-MGVDISMPKTH
42 IS--KDT-YEFAKRW-----IKDGKE--ISGIP
43 LK-----
44 GLINNIKH-----
45 -----LKIVYTIINDYLIKVP-----SN-----VSL
46 SSWLIFE--K-----IFVGFP-----IYSKNT-----SVKRYISRKY
47 I-RSLKDFALS--VRFSM-----GLTT-----
48 -----PYELRQ----YLSQM WIDESK----TEFKAIPNENIIRQYFEGILVNG-
49 ----LADQAK-----KTILQINDQLYSF-----EKLEQKERKSL-----
50 -----VY-SGVIFGLMNRL-----
51
52
53
54
55
56
57
58
59
60

-----TKLEEQCTKWK-----ED
ESTIVDILNFFSTPSVDSL---SRKDRDCMIRLDF-----L
DSLW-----KKSFKKHFSEQRFPDSYYRDLE-----TK
TLYGSLDS-----LDITSSGIEVR-
-----PIWNT--ELKRFNSRSIGALEMFISDNE-----

>GFQK01014503.1

-----PLDVISLLRSK-----YVVIDKMVS
LSDK-----IRTTLHVPFAINNMLAHGRV
-----GRLARGIRRTNDF
VS---YVLK-----LWKAHGTTFTIKWLKASHVAIAK-----CL
GQNKLLSLRTLEQDLNLPRLV-----NGLPKIIPFEDRRMRQGHVPTIRFWLGLFN
LY-----RILKAPAVPKLAT-ITS-----PWTGQFKP-----
---FLELCRIARR---P-----DFNFFCFLKGYRHIDLQHVQPTE-----
----LVSSSSASPSNKISWQGFLTDLHLLREHN-----PVVWS--AIKRYI
HLTG-GHTF-LSYLEHALEIDDIFSKYDFYRDSSSSEGHLSKTGKKIL-----

-----VADHSMT--KAAIA
QHGL-----
----PTGDGLSQFAIKEEAAGKVRVFALVDLITQST-LKPLH-----DQLFS-LL
RKIPC-----DGTFNQEA-ALDRA-----MFKATKA-----
----NCAFSF-----DLSAATDRIPV--LLTQNILANILNS-SE-----
----AARLWKII-MT---HREF--G-F-----L
QPTAE-----KLGIDHNVGFRYAVGQPMGALSSWAGLAITHHWI----V---
QVA-----AREAYYDLEGWY-----
-----ENY--EVLGDD--IVIF---DSLVA--DKYLGIMSS-IGTDINITKSI
VSRNRPV-FEFAKRI----C-----
-----WGS-----
GIVSGISL-----
-----NQVRAS-----WN--VGSRLAS-----V
LQFARLG-----LLETGS-----SLMVAL-----LSRDAFS-K
G-KAIPNVKTE--SPRSQKARALG---L-----LALLG-----
-----ERYQKGILSL--KEVTRILINP--KDYLDEGNAIA-IP-
----LQFAAQ-----LAYRAL-----VDPQKPHPV-----
-----PSS-VEERDRIWGFR-EPFL
-----KMELMEAITSE-----
-----KQLLISEMV-----SL
SFNW-----TDKL-----FYPLCYDSSKE-----RVP-----ND
DLPVDYLH-----
-----AYWEL--SE-----

>NC_015629.1

-----MMNPFIRLGSLLMGKTTK--

-----ELIIGLYSCVQTLHKEYL-----SQ
GSRGLAIRLKASSTYLMKSW-----GDPLLDCKAFG-----
-----PTMSLTPRGIPIW-IPS-----AWRCQISSRNLAIF-----
-RVTMTLCNLYRV---LP--YQG---KLKLSTITDDRKG-STPDLMFS-----
----FIKDKFAPNITNRVPYPG--IFS WDPKPLL-----TRGPG--AQKEKN
TMSGFGMAL-----RAIHETGMLDHLKAYALYTGLNP-----

-----QLESLKVIWEGTRK
HCSI-----
----GPIGKLAF--KQEPG--KVRVFAMVDCITQWF-LHPLH-----KYLFS-VL
RTTKE-----DATFDQEKGINLVR-----LALSKKLDKSV-----
-----FSF-----DLAATDRLPM--DIQMVLNSLTPFWLAEKAVRH
GVKGG--LGDAWADL-LV---DRDY---YLPR-----
-WSGYARDS-----KVRVAVGQPMGALSSWAMLALTHMI----V---
QFA-----AASVGVGTGWF-----
-----KEY--MVLGDD--IVIY---NSEVA--KAYSTLMGT-LGVGISDTKSL
TS--KIGVFEFAKRL----M-----D---LEGPCQ--GLPLA
EFAAARF-----NLS-----
ILFQSFRS-----
--RTLYPKISTFMRFLGFGYKVLGSLGMRLGDMRKRSYFETVAYSPLVTD-----KSV
NKWSEFF--KRFSTYDLY-----PLVEAC-----VYKAYH-----LIPALTK-F
D-RWN-----LWRTL-----F-----PHFDDKK-----
----ILLIP-EVAG----RHLESLFTSMLDNKVAKYKADHSVYVNRLNRAEVPTND-
----WEAIDT--LLSIMETSPPTSDLLQDP-----DMNELGFGLSV-----
-----SN-YDEDNKS KTC SLSPY
TVFTKI-----PQLLDSIDDWV-----EPIHPDKH
PEVSADVNRPLIHKPGKVF--VKGIATGKNENNK-----
-----VEITKMLQDKKFLAYIIESGKITI-----YS
PAFQAMIR-----
-----ASWAK--AQSGSK-----

>NC_004049.1

-----MKRLTLSQNKSNQLTNNCLS

-----NVGYITKQLFPHWIRLLV-----WS
LQLSPAPYKKFGS-RIAILWKA-----NGVSFTVQYLK-----ECTRIVQHFVSGHP
VFV-----TDVMPIGLAGG---LPT-----IIPGTLRTLLRSKD--SST
IRGVLSTLAVYRIMK-MP-----CVLKLESITDPFKG-ISDTLPKS-----E-----
----IINGLASLGFEIPKGRSK--HLLTSLNPII-----YLLSA--GPNHSI
SMMG-----IW-----KDIYAWYVSPLFPTLLSFIGRMNRGNVL-----
-----I
DLLR-----AEVSYWE--ATGVK
PSVS-----

1
2
3 ----PLDLKLGKLAIKEEAAGKARVFAMADSITQSV-MAPLN-----SWVFS-KL
4 KDLPM-----GTFNQQAPLNRLVQ-----LYQDGLLHDV-----
5 ----EF-YSY-----DLSSATDRLPM--AFQKQIISVLFSGS-----
6 ----KFAKDWATLLVG----RDW---Y-----
7 -LKDI-----PYRYSVGQPMGALSSWAMLALSHHVI----V---
8 QIA-----AMRVGKLPFTNY-----
9 -----ALLGDD--IVIA--DKAVA--TSYHMIMTQILGVEINLSKSL
10 VS--NNS-FEFAKRL----V-----T---MDGEVS--AVGAK
11 -----
12 -----
13 NLLVALKS-----
14 -----RWGIS-----SVI
15 LDLYNKG--LALS-----EQDLRQ----RFSSIP-----T
16 V-SKQFGVDKL--LWLVL--GPFG---F-----IPSKDGLS-----
17 ----AFMKLNRSLSL----VDMHILLSCVDE---AKFDLDK-----KT-
18 ----WEANIQ-----ETVHTLLRFG-----MLSEPAGFEVF-----
19 -----SDFTSSP-LYSFIRGQFGN-KLSA
20 LV-----QDKPVRRLIFD-----GPLLHFNFY
21 TEGWCDGLMEHLTKKIQSD--SQETVSPSNPFKD-----
22 -----DKVILPLRGNIGIFFKHVLALMAE-----RD
23 PATVMRWM-----
24 -----
25 -----
26 -----
27 -----
28 >HAGX01098760.1
29 -----
30 -----
31 -----
32 -----APPARLPLE-----
33 -----
34 -----
35 -----
36 -----FLEGVRLAQPKSDG----
37 ---HLSLAASATV--EYPRSKGG-----
38 ----RAAAYFTWAKA-----RMKELGLKAP-----
39 ----ETVA-----ETATHWLYSQEAQQEYYSECLGQ-----
40 -----
41 -----
42 -----AEAWL
43 ESGE-----
44 ----PVPHRLAV--VGELG--KARVITVAPTWFVQV-LITMQ-----KILMH-SL
45 RFIPEV-----SAGVFSRD-DAYKA-----YAEIPDVST-----
46 --GDYHF-FMS-----DLKDATN-GPPKSGIVDTVLPWLFDQ-VGA-ESP-
47 -IPQR--DLEIALSLLMS---ERKI--T-V-KN-----
48 -D-YG--IP-----REFITRRGILMGDPMTKVILTIAYYLI----V---
49 KAV--C-----ADVEREHPGCHTWF-----
50 -----RIKGDDVVILVVGVSPEASSTEKHFLRRISD-AQFRVSELDTF
51 VS--HA-AFYCEGL----F-----LAPRGL--EIEKI
52 SRARRFGH-----
53 AYADPVKG-----
54 -----
55 -----
56 -----
57 -----
58 -----
59 -----
60 -----

-----RYMVPF-----WKVDPGGKDPLVLP--LSKLESI-----VKR
LSWFVPD-----SVQWRR-----TYSCAM-----TATKS
F-RLERDPSLF--LPPKQ--GGIG--LL-----PDAGIRE-----
----LSTADWNQFIWSWAVDLRATVLRDKIPEFSDPGVPRFVARSSNHAFSHHLRVDP
QVIPAEWEETFL-----HLGDVDERLSTR-----LAEYLGRSPNYSSDISV
ASYAESVDRFSTV-----HL-GRPLIKRLKTQGSYL
SLEKVGSIPTDLIEAVCQYRSNEARHRRIDLRFMRSDLIGPRLVVFKEPEDPEVIDLEA
IKQASSPGEARTLVQAGDP---FMDPDAQILRALYLADFKSVYPIVTNDRRLCFEARRTL
SSRW-----PGRDFVVARSRTPLLDSRSGLEML---SRHLRTLVEGSDSHLWWD
PLRTVELF-----PSLSVTDGERFLRTSDSDWSESLMDITNVQY
ELVKSQ LAPLIQRGDQHWEE--ILAQYETVGVVFPESVQDLELTPGPFQLSWRPRLREP
RRLKVNHLQKIWRLRP-----
>GFJG01044014.1

-----IRWKP-----PIPSIGRDG
PNVLLSGDETPTQKLEIWSLTYGDVCRSMFEKR-----IPRTALRRAGIPWTNKRPN
-----RRLANLTADQRSHFILRQ
LYPRMSGAFKA-----SLKKQSLGQIKSWFHTLDGLILPLL-----LE
GAEDLRLIDRVTK-HVMENCA-----QNYRAFLDRLK-----SLRKRFRKWVLSHV
NTP-----YKGFPGECHDLAP-ILD-----LWWSILQPHQGDRR--KFI
Q--VTCMCTQTRATGLVP-----ARMMQESVKKFEAT-VSKEGPTI-----
----QLDGEVLMETV-----AACRLVNPETAKL-----SGGPA--ACFEKP
RSKG-GQTQ---AIL-----EICKQTINYHYEPEGSTLRRVETPTHE-----
-----V
NLPD-----DILDYCI-EQLRM
SPES-----
----MKTRVSV--VNEPS--KARVITAASVSQVY-MSMVA-----HMLQP-VI
RSPRII-----SGLRSSR-HLWNF-----LRKTMHPQNPEWDRLAGQ-----
---GAIFALST-----DLEEATDFGNI--SVADQIWKAIVET-----VPE
SVKPY--FILAHEAY-TR---ERDL--L-M-----
-PDGR--WV-----KKKRGWLMGDPCTKAILSLAQEYIFRVATK---
RWG--K-----
-----VFMG--SLVGDDCIMLSR---YKEAL--EDYLQEYTR-CGFKISEKDTF
IT--TNA-MYYCEEA---S-----L---VPSDSM-VDNLTV
QIKR-----GMK-----
DAILYIDV-----
-----PRLRLM-----VATRPDQPSRHSYSdT---GRTELL-----GRE
SNWAYRW--SEKLYL-----PFR LGT-----LLQRAL-----V
P-RSVDTISPF--LPISM--GGDG--AFYADPQ-----LTRDLIE-----
----RFAWSI-GEVR----YRIRRVLSGQRNPYYLRTSNPIGGTDKFHIIMDMASIE-
-----ELVPRE-----CRIPVDDMEHCSML-----MSFKGTFIMEPIRTAYRA
LQSRYWESILIEKRIPSD-----ISIPHRVTKT-GHTEGAVPLSAQEIAD
FLKLI-----EDRPKMVQEWPFVDMRKAYIMDYMHLGWKRERDP
TPQEEWDLYLSGQIDLMGA---SRALYNFITRGCTDSGEPIEIPKVIRDRLNLFFSTDPV
VKLE-----VRRMELKPESHLILVSRDYRLAFE-----IQ
KMKPETYI-----TLVDPIIHL LGFN PW

1
2
3 TDDGDKTRVIEDSGAILWAD--
4 LNLFKDGYPVNRRVELVYSACMNCEPTRYGYSRTADWR
5 RLHLNSRIGTYFAYRHGGWILQKVPRERRPSHEAE-----
6 >GBKL01052498.1
7
8 -----
9 -----QKKMSETIRWKP-----PLPSIGRNG
10 PTVVVSKASTLTEKLVYTLETGICVSFFEKR-----IPRSQVLRLNISWRETEPES
11 -----RRLSSLSPDQKTAVILSNIWPKM
12 SP-----HFIT-----LVRAAPVTAIRNWFQTLTGILLIPAIL-----FG
13 LDNEYGLIDRITK-SCLERMA-----QNYSLFLREIK-----EVRKRFRLLFFESD
14 QRF-----GTLRYPGPSYGLTSILE-----FWRRILTPLIGNQREF---
15 ---LRAVVTIGQT--RATGLVD--GATLQKSLDKFVRT-VTTPPEET-----PC-----
16 -----EVDEVAVLRLG-----PLFARADPSRAKL-----SGGPA--SCFESV
17 RSQG-GQTG--AIV-----EICKRSINFSYGIDLVRTDIPEHRICTSA-----
18 -----
19 -----DLFSWAV--EELKI
20
21 NPET-----
22 -----SFAVRAVV--VNEPG--KGRVVTAASIAVLLV-MGVMG-----HIFQP-TI
23 WTERT-----SSGLKKTR-HLWTF-----LKEGLHTEGLAWKLGTP-----
24 ---VLALST-----DLEECTDFGHP--RLSTAVMQVMAKF-----
25 -CPEV--LRPLVEFA-TTVYHSEPI--H-L-----
26 -PDGS-----IRTKRRGWLMDPFTKFMLSAMQEYVYQRMLI---
27 RWG--P-Q-----
28 -----FVT--TVVGDD--LVVL--SSNKPLL--EDYLSQYKV-AGFKVSILDTM
29 IS--DQI-MYYCEEA----S-----Q--IPQKAS-KAL-TV
30 SLK-----RGE-----
31 GEFCYLDY-----
32 -----PRFRLL-----IPTHPDVDRHSYTD--TGRVELL-----NRE
33 FSWVVKW-----KTGGEE-----AFERAV-----LLQKIIV--
34 P-RSSETSHPF--LPVWL--GGDG--AY---H-----PDPGYVF-----
35 ---EYVRSH-SWNVGEYAWRLQRYLEDTEVAYTRTSRSGPSVHRYHNIMALAEIDG-
36 ---LVPEEL-----RFPVVDEEREMW-----LAFKGTHILTPPEAAFKF
37 LKARYWLTVLEGEIPEET-----KISKVIIP-LGNVAPELTRK-SISE
38 LLDRM-----TERKPRLEEWK-----WLVDMNSNAICADYMTLSW
39 ERLKPNPGEWKLFTTEGLL---DMHSAGMSEIYHAITHGDIDIPQGILSRMNVFFSSDPL
40 IYR--LRKDVPDEENIALISRDFKLACKIQQMFPKRG-----VAL-----VD
41 PLYVLGM-----IP-----EEIDSRFKIIEDPGA
42 IYWTDINCFQDGAPREQWSD--
43 TVEDVYSGRLAQRWHPKPGHRILNRRIGVYFAWTSKRR
44 QGGGY-----
45 >GBKL01051391.1
46 -----
47 -----FFFLKKKMSETIRWKP-----PLPSIGRNG
48 PTVVVSKASTLTEKLVYTLETGICVSFFEKR-----IPRSQVLRLNISWRETEPES
49 -----RRLSSLSPDQKTAVILSNIWPKM
50 SP-----HFIT-----LVRAAPVTAIRNWFQTLTGILLIPAIL-----FG
51
52
53
54
55
56
57
58
59
60

LDNEYGLIDRITK-SCLERMA-----QNYSLFLREIK-----EVRKRFRLFFESD
QRF-----GTLRYPGPSYGLTSILE-----FWRRILTPLIGNQR-----
-EFLRAVVTIGQT--RA--TGLVDGATLQKSLDKFVRT-VTTPPEET-----PC-----
-----EVDEVAVLRLG-----PLFARADPSRAKL-----SGGPA--SCFESV
RSQG-GQTG--AIV-----EICKRSINFSYGIDLVRTDIPEHRICTSA-----

-----DLFSWAV--EELKI
NPET-----
-----SFAVRVAVV--VNEPG--KGRVVTAASIA YLLV-MGVMG-----HIFQP-TI
WTERT-----SSGLKKTR-HLWTF-----LKEGLHTEGLAWEKLGTP-----
-----VLALST-----DLEECTDFGHP--RLSTAVMQVMAKF-----
-CPEV--LRPLVEFA-TTVYHSERPI--H-L-----
-PDGS-----IRTKRRGWLMGDPFTKFMLSAMQEYVYQRMLI---
RWG-----
-----PQFVTTVVGGDD--LVVL-SSNKPLL--EDYLSQYKV-AGFKVSILDTM
IS--DQI-MYYCEEA----S-----Q---IPQKASKALTVSL
K-----RGE-----
GEFCYLDY-----
-----PRFRLL-----IPTHPDVDRHSYTD--TGRVELL-----NRE
FSWVVKW-----KTGGEE----AFERAV-----LLQKIIV--
P-RSSETSHPF--LPVWL--GGDG-----AYH-----PDPGYVF-----
-----EYVRSH-SWNVGEYAWRLQRYLEDTEVAYTRTSRSGPSVHRYHNIMALAEIDG-
-----LVPEEL-----RFPVVDEEREMWLAF-----KGTHILTPPEAAFKFL
KARYWLT VLEGEIPE-----ETKISKVIIP-LGNVAPELTRK-SISE
LLDRM-----TERKPRLEEWK-----WLVDMSNAICADYMTLSW
ERLKP NPGEWKLFTEGLL---DMHSAGMSEIYHAITHGDIDIPQGILSRMNVFFSSDPL
IIYR-----LRKDVPDEENIALISRDFKLACK-----IQ
QMFPRKGV-----ALVDPLLYVLGMIPEEIDSRFKIIEDPGA
IYWTDINCFQDGAPREQWSD--
TVEDVYSGRLAQRWHPKPGHRILNRRIGVYFAWTSKRR
QGGGY-----
>NC_032404.1

-----MTNSCNNIPINMLLKEYNLQVGPCVSIFTK-----RVPIWWLKR
SNIR-----FYHDKVPFIKNLQPDDRVRV
-----LLSSIWTPRDRP
LSFAFRAKLKE-----LSTKGSINTIESWWNTVQGCCLPIMVG TN----LW
TDQIPNWCNRITK-WAFESCS-----RNYSLFQSHIK-----AERKLMRQYFMETG
IST-----LEGYKSSRREVVT-YVS-----VSQEVLRNLSTVRE--REV
---MIGLLTQTRA--TG--LAD--QRMINASLEKFLSV-VSH--PCA-----
-----HITLNQKVMDSMVPRISG--YIMGYTKL-----SSGPS--AVFSHP
RTRG-GQTA--AIA-----ELSQR YITYSYDLHTLEKTTVERTR-----
-----I
DNAR-----KVLDYCI--HRAIH
RRTE-----
-----VLKVRVST--VVEPG--KARVITVGTLEYVQI-LSVIA-----HIFKE-SA

1
2
3 NSQRT-----RSGMKASR-HLWNF-----LWKDLDPNRNSIWETISGE-----
4 ---GTIYGLST-----DLETATDYGNM--MSGQIWDLLIQA-ATRVPG--
5 -VPIK--LLHLGKELYQS---DRPL---F-----
6 -SAGK-----LIATKKRGWLMGDPMTKILFTLIQEVISRSVGR---
7 TYR-----
8 -----IAC--SIVGDD--LVVL-SRERSGL--KEYLRLLET-LDLKISWKDTC
9 IS--DQY-IYYCEEL-----ARIPQS-AGVSLR
10 SLAR-----KNK-----
11 DYIGYIDV-----
12 -----PKLRTL-----IPTRGESDAFSNTN--LGRFSLQ-----GKS
13 SEYTSRV-NTGYYP-----VNERAL-----LLQCMF-----I
14 A-EDTSCIHPh--IPIEM--GGGG--SLVNDPF-----YVQRSIS-----
15 -----LCKYQ-REIT-----YRIRSILSH-----NGWGFRFVHGDRPDRVTHKY-HR-
16 -----FGKIIQDPRVFPEGCIHKPRNKAQSEL-----LASVRSSKIIPPVDAFMR
17 LQRMYYFRGLFSGTG-----EVPDPDE-YLKSLEIPDGITKPSI
18 GYTKLNPLTLDEVSTFL-----SLWVKRGLAFRDDIPFWINREKLDINHPLVIELFK
19 KPDMPTWTTEDLIEESEKLSRLAENPWEIPDTLTPALSLVLDQDAILINILERKIKKITN
20 IGRLVVVTLDVRLLRMNRILRKRKSLDRIILIDPF-----LNLL
21 GMAPDFNP-----DIYLEDQGGITYVE
22 D-----MYFQDGMPPVEVEEHMFSDPVYFKLCKPDLPIYIAASVIHTKRF
23 SPKYRACGANLLMS-----
24 >GEMJ01010140.1
25 -----
26 -----ENILTDHWRP-----PLPLTGRFG
27 TRKIIENLNDRSNSYLLKQYFLSS--TSCTN-----IFQKRIPVFWLKWNGIPFKH
28 -----TYVPYLR-----DLPKYYQSKFLIESLYPRMSPRNF
29 CN-----RLQK-----LCYEGKLHIVEQWFNTAEGCVLPLFLLPD-----LW
30 KDVNLKLIDRVTK-WSLEWCA-----CNYAHFISILK-----SFKKAIRKYFALHG
31 NL-----DGFTHPEDRTMVP-YLD-----LVRSLKDDTLERK--LQ-
32 ---VILTCTQTRA--TG--LSN--HQMMMNSLKKFKAT-VTT--PPV-----
33 ---HIEMDQLLLSE--ILNK--RVLRFSEPLSRI-----SAGPS--AVQGYS
34 RKSG-GQTK--ALA-----ELSTKYIDYDYDLRTLTIKEERVQRR-----
35 -----I
36 SSAA-----DVLNYSI--HQALT
37 ERKR-----
38 -----VSTVTVTL--VNEPS--KARTITVATLESVTL-LNVAM-----HIFAP-TC
39 NSQPV-----RSGMRANR-HLWQL-----SWKDLHPSNTLWDKLDHK--HNDT
40 GYNPLWA-LSS-----DLESATDYGNP--SVAKQVWRKLIDI-LSVFPD--
41 -APIG--LLELCCEIHTG---PRDV--L-Y-----
42 --EGE-----P-FCVTQRGWLMGDPMTKVILTIVQEYVL-KQFQ--
43 QLI--P-A-----
44 -----VG--SVVGDD--LIIL-SYSRAHL--EMYPMSLEA-LDFKVSMDTF
45 VS--KRL-MFFCEEg-----S-----I---VPQTVR-DTIVCR
46 VKRPTWKKG-----YKP-----
47 ISCGYIDY-----
48 -----PRIRLL-----IPVRPDQDRLSYTT--LGKFSLL-----GRE
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 VIWAHSI-----NEPLFP-----VMERAV-----VLQNIIL--
4 P-QSDTLSPF--LPIEI--GGNG-S---YCPD-----PDFLK-----
5 -SAVEKRSRSA-SQCR----YRMCQLLNHE----YGFRYLR--SERRNKVLHKN-VV-
6 -----VARLVE-----DESVFPDSVIRAE-----SDEHKTILS-----
7 -----SLKAYISPEEAFMKLEYQKYFTQVLSGTGDIPSNE-YQRDHS�PTNLGGVDV
8 EPNYSKFLKIWH-----KPGFTFKNSWKFFVNPKKLGHNDPLSVDLKFKPDL
9 QVKPDPGDLLGRTLEDFLN---RVQAWD--MYFPD-----HLL-ERLEL----VV
10 ESDS----LLVALLKARLQRGETAQTLCVLS-NDKK-----LLT--RLVTIWD
11 QLRPGECF-----LEVVLVHPDVYLTGMYSWQEYTM
12 PPIMVLKDPSG---MFWES--NNFNACQYWEVCWEVDKYQPLPGVATFTTLFYLRPGR
13 RHPQDDPVQEITPGRVFPHPMGIQDVETSGADTPQEHIEPYVPPHLRK
14 >NC_035120.1
15
16 -----
17 -----MPPLPLIGRLGPRMAVSSDLLT-----TLGDEKISS
18 SVPL-----EVLLERYFTSF--GGCAN-----LFYKKVPISFLRKNSVRFH
19 -----EEVPYLSNLS-----RAMQFRYLRKSVWPQMSPVF
20 YQ----RLYR-----AVSAGRLTTIESWFNTAEAMCLPFFLLG-----VD
21 SHTEYSCVDHITR-WSLANCA-----NGYASFIKTIK-----TFKKCLRKHLSGE
22 PI-----DTRFHGSSRTCSP-YLK-----WAKVVLGSIPEQQT--ALR
23 AQ-AVALLTQTRA--SG--LAD--HRMMMQSITKFKSV-VTT--PYEHIQMDEP-----
24 ----VLSRMIHPRMRRCVLNNA--KLGTGPS-----AIRGFP
25 RSSG-GQNL---ALA-----KTCTMTIDYTYDLRSLSVTSTENRR-----
26 -----I
27 SNSR-----DVLDYCI--SQALN
28 NRQK-----
29 ----VTEVTLVS--VNEPS--KARTITVGDYALVQL-LNVAA-----HIFKD-VC
30 CTQPV-----RSGMRADR-HLFNF-----VWKDLHPQNTLWDDMGWSYETKGM
31 P---IHALSS-----DLETATDYANP--SVGRQIWDCLISG-LEIQYP--
32 -ESSPRALLELCRDLHVG---PRTV---Y-----
33 -YQKI-----FFCTKLRGWLMGDPMTKVILT LAQEYV----LFRS
34 NAG-----RGPTGRLVG-----
35 -----SIVGDD--LVIL-SRLRHHL--GWYLDDLRS-LDFRVSDDDTF
36 IS--SDF-MFYCEET----S-----R---VPQGPQ-QSVVAR
37 TKY-----SHG-----
38 TSCGYIDT-----
39 -----PRIRLL-----IPTRPDED RFSNTN--LGRFSLL-----GKE
40 YQWCLGN-----NSDLAP-----LFRAI-----GYQNCLV--
41 P-QDADTQCPF--MPVEA--GGNG--SYTTDTA-----FWSKLVL-----
42 ----RRSKDP-RTTH----FRVNQLLSNE--YAYRWIRSDRPLRGLAHRYSVVR-QV-
43 ----FEDETR----FPEGSVVKAHEHEERM-----ILRSFKQFLSPETA FMKL
44 EQARWFNGLFDGRGVIPPNMQYDREILLPSVGQSGPETLDPPV-IRRFLERWRNP-GFT-
45 -----YRDDYRFLVRR-----DFLPNEDPLFIEVDPNEP
46 DKLSASVVPDDLIGKTFEE---AVNKILTADLYLDDNLWSRLPLLIESDSMVMARWRELL
47 RRDTYVETLVLISTDRKLLLRMIDLFTVMRPDGYIVEAILMHPALFLTGTWGTGYMDYDQY
48 N
49 GIPGRVVI-----QDPGSIFHQDLYLPQRRS W WEG---

-----TVMLE--PFRTQEREGVTTIYHGYMEVRPELDRKPLPRSNASRIW
RHRE-----
>GCRV01019573.1

-----RHVGCAIMQSADRSWAP-----PLPSLGGKV
GPRI-----MVNPNLSLSDKLKA-----YYLELSSSPSMWIRRVPYAT
-----LRAVGYSKGAKHLVDLPWLYQMKV
LL----KTLYFKASRHFVAKLESACRAGRINHIKQWWITADAVVLPYLLEL-----KE
NDADYQVVDNLTR-FALESCA-----NNYSGFIGRLK-----KVKKLIRKSLA---
-----LDQEIPFMRDMYP-----YIKQYKASSPSEGG--IHA
MAVHCNLWCQTRG--AG--LAD--SKMVNKAETKFLDC-IQRPKVSI-----
----KLDRQILKDDT--LCCR--AANGVHAHL-----SAGPK--ACFEST
REEG-GQTA-----FLSKLARTRVVTHSYNYETLERVPLRRR-----
-----V
GSPK-----DLLDYCI--DWSLN
EPVL-----
----RKVVRLHT--VAEPA--KARVITVAPFAYNRL-MGVCA-----HMIAP-TL
RSTPL-----RSGMTASR-HLWGF-----THDTLHPGDKLWEQGIGC-----
----EA-LST-----DWEVATDNGNV--SVARQVWSEIIRH-LSKIDG--
-APLG--FLNLCKTLYCE---DRYV---L-----
-HKGK-----IHRKTRGWFMGDMMTKVIITLVHDYC----A---
RVA---L-L-----
-----KCY--SLVGDD--FVAM--GYNENL--QAYLQEVAR-LGIVISDDDTY
IS--PRF-MYYCEEM----C-----L---VPHTTS-QLP-VV
QIA-----MGK-----
SKISYIDT-----
-----ARLRLM-----LPCKQENQIFSATQ--VGRYSLL-----GKE
TRWALST-----HREVAP-----IFERGV-----VLQHLLL--
P-RDQDTQCPF--TPIEL--GGDG--SYTSDSN-----FFKNVIE-----
----KKSKNP-AETH----YRIASLMRGA----FGYRYIC--TERVDQVSHKY-RM-
-----YVPLFQELVKYLPPEEAVIPIHDGNRTL-----TTLRARLLETPTORTILRI
VKEAYYRELFGKVPVAKL-----PTLRLEAPPMPSTGG-ATPFYSASRFL-EHWR
NPGFK-----FRDQDPYLVRT-----DIALTLDHMSLGWDFSIR
GPPKRALEDWLSQNLSLFD--GNANEVLAKLLSNEDLP-----EGAKSRLNLFEE
SDGW----VLNDLRSRAEIPPGPIYLISNDIRLGAD-----VSRVMEGKQDPR
EVYCVRPILSIIGRMDEVEDLVPAP-----ALSLEDPGSVSFETL
VTLGEDF-----PEWAF--NPLLLVKTGFSGVFRIVPEIYAD-----

>NC_032457.1

-----MSGHFQRALSKLRSYDYRT

IQ----RWWY-----TVDGLVLPLLLTH-----PD
PDSQMETIDRLIK-FGFENCA-----NNYSHFISRMK-----RLKKLLRKEFAEG-
-----TDLTSGQLFVDLSTYRK-----VFYRYLDLTLERSATE--

---CVMAWTQARA---AG--LAD--RRMIEKSLEKAEKT-LSTPRRVYLRGRWRE-----
----IHAYQGFIPV---LEPEVPRGPLLDVVRVANKVRGETAHISTGPH--ACLEIK
QAFG-GQAAGIAYLC-----KRDVLTVSYDPVTLERTFHPARPCASAQ-----

-----DLLSWAI--SRALD
APVH-----
----TRMVRIVA--VSEPS--KARTVTVGALAYQII-LGVVS-----KIFQPALA
SGISRTGLEGTRNLYEALNNDP--DPSNG-----LWGPLRDMMDRTGKFP-----
----VFALTS-----DLEEATDYGDG--AVARQILQALLMR-CRDIPG--
-FPLG--LAVLAKSLFLS---SRIIRPGY-----
-GHGL-----VHWFRKRNGWLMGDRMTKVVLTLAHEIGILSAGI---
QFA-----
-----RICGDD--VFAL-SKAPAQL--QRYHKVMTD-LGFKISEDHDF
IS--RRI-LFYCEEV---S-----L---VPQDAK-DLPSVC
N-----RRS-----
EPSCYVDY-----
-----PRIRLL-----LPIKVETNALSYPD--TGRFHLL-----GKE
MRWVFQN--SSQQAD-----PFIRAS-----LLQHIA-----I
A-MPRDILSPF--LPQEL--AGDG--AF--PH-----SAEFLLS-----
---VIERKSINYDECL---YRIHSLTHGK---WGFYLR--ADNINEVVHKY-HQ-
----MVPKLG-----VLESVLPPDAVVRAS-----EVLISLKVKG-----
-----LETPEKT-----
-----FFRLYRSYYWY-----
-----KVLHGMKAP--VLNFDedrTRVHG-----
-----GREFKLDIPPDVLVERFY-----TT
WRDSGFTF-----

>GFUS01019776.1

-----NHVRWSPTLPSYGGRLCPRV-----LIPANLPDF
ASKLEAYYLHREGSSSIFVRRVDV--GWLRK-----HGERLIAQGPFLADLPGTV
-----LYKYLIASLGSLTMSAHF
QR-----SLSQ-----LCLAGKMGAIKRWFTTDGVVPLLLSV-----ED
PDQHYADIDRLTK-FSLENCA-----NNYSHFIGRLK-----RVRRLLRKDYA---
-----EMGTSAKSLTD-MTT-----YRGMYLKHAPKNRD--QYR
DVQYLMCWSQTRA---TG--LAD--ERMIRKSLDKAEIT-LSTPRRPRRSIPRED-----
----WVQQSTKLPLGEPVIDRA--LLDSVCERIYRRC--DGRTGHLSAGPN--ACIESP
RNKG-GQTG-----AISRLTRMKVLHSTYDFKTLEVTRHPDRR-----
-----V
GSVQ-----DLLSWCI--QESLE
HSTY-----
----VKCVKLHA--VAEPS--KARTITVCTLPYLLI-VGSLA-----KLLQP-SI
ASDIT-----RGGLESSR-NFWNF-----LYQDLDPADVWHRLRDE---KAE
DGDGKTYALSS-----DLEEATDYGDI--GVARQILDGILRS-ASRIPG--
-FPVG--LALLAKTLFLG---KRFC---F-----R

SENGK-----WRYFVKRSGWLMGDRLTKVILTLAHELAIASGL---
HIA-----
-----RICGDD--VTAF-SKRSRL--ARYLVVIRE-YGFKVSEDDSY
IS--SRL-AFYCEEA-----AIPPVRFDLPSV
CT-----RRK-----
TVPCYVDY-----
-----PRIRLL-----IPTSSEVKAYTYTD--VGRFAQL-----GKE
MRWILQG-----EIHPARD----VYVRAS-----MLQHLLV--
P-QPADTLCPF--TPVEV--GGDG--AF---PH-----SENFFLE-----
--VLAKKARNP-TEIR----FRLHQLHRGV----WGYRLVR--HQYLDNVIHKH-HM-
-----LVPKIE-----ELRSQIRECVPDALIEGDEMLNSISVSGLETPERT
FFRMWRAWYWSEIFR--GHDPPETILEISRECNLPRSEVTLTLD-ALGFYRVWRGKGFTFQ
NRVDYLVTIKMVQPLDYMNLGWNFGVRDLPTENPDA---DQGRFLLNEENSSSFIEYIRT
GRPMDQVVRDSLHRMVESD--SFLKWEFSRRCHA-----DIPQLILLVSTDVNL
GIDF----QLAATRRPDYEDAPLVFLIKPTDYLYG-----FM
SHYESWAA-----DLTQHWDVIVD
FGCVVYNQFTLDSPEFEWFEDPDVVHNPVRRDASLRVYYLSLKEGEGGSERCSLTTQPA
S
QWRKGFFPSQYVREGLLGN-----
>GOW01018115.1

-----SIMSQKRWSPTLPSFGGRLCPRV-----YLPEGYPSD
YCSR-----LEQFYLERSSSSSIFVKKVDVTWLKAKGHQDQLFLDGRFRAPSLGDL
-----EWTVLYKYLMMSGISQTMSLNF
QR----KLSD-----LCRYGKIGMIKRWYQTVDGLVLPLLSV-----ED
PDLHYAEIDKLIK-FGFENCA-----NNYSHFISRMK-----SLKRLLRKDYA---
-----LDTESSKVLSDVST-----YRKIYLRFFPAGRGVIPAD
AIRYVLIWSQTRA---SG--LAD--DRMCHQSLVKCEKT-LSTPRRARRIAPPEE-----
-----RRYSYRGLPLGEPVVPKI--HLSRICERVARRLRGETAHL--SAGPN--ACLEST
RQEG-GQTG-----YISRVCRSKSLHAEYDFRTLECKQCTPRPVRSQD-----

-----DLLSWAI--QEVLE
HPTY-----
----VRCVRLHS--VAEPS--KARTITVCTLPYLLI-VGVLS-----KLLQP-AI
ASDIT-----RAGLHASR-NLWKF-----LYEDLDPGADLWFTLRDQFKVRGQ
ISNLPILALSS-----DLSEATDYGDL--GVARQILAGLLHS-CSEVPG--
-FPVG--LGQLAKTLFIG---KRFC--LRC-----
-ENGR-----YRFFVKRNGWLMGDRLTKVILTLSHEIA----L---
RES--G-L-----
-----YIA--RICGDD--VIAF-SRSRDDL--ERYHQVVTD-LGFKVSEEDHY
IS--SKL-AFYCEEA----C-----IPPTY-RELPQV
AI-----RRK-----
TGSYIDY-----
-----PRIRLL-----IPTCVETHAHSYTD--IGRFSLL-----GRE
TRWVYQN-GAPHVAA-----LFKRAS-----LLQHIH-----V
P-RTADTMCPF--TPLQI--GGDD--SF-----PFSGRFL-----

-----EAVVSKKARNESETR-FRMSSLMTGK-----WGYRLVR--NANLDNVVHKY-HT-
-----LVPKLK-----ELRNLLPEHAILEGN-----QFLLSIKIDGVESPERT
FFRIWRAWYWHQILKGSNPPE---LVLGVRKEANLPQTKVNID-FEYFFHIWKDLGFTFE
NVPDYLWSTRDVHVLDYMNMGWWF-
GTPPVGKTTDPIDFIDIGGRFLLNEENLESFINHVKD
NREL PQVVRSTIYKYVESD---SYIKWEFSKRDDL-DSLTRIIFLVSSDLKLGAELARMA
RSRY--SSNLADAPVVVVVPCWVYSYGMM-DIPIDWS--TNSMSCPDPIVIVDQGSITFD
QMVNSPEF-----SFFDKPRVT
-----TRWYRSNPWLGTQCLDIVEDGTPSEVWSLTTQH-----
PSQWRPGHFPYVPLRT-----
>GFRX01320111.1

-----FFKKKTSHMMTQNRWAPTLPSTFGGRWAPRV-----AIEAEYSAS
YLSR-----LERYFLRRQSCLSMFSHRVEAHALRECFGKDILTGVMVTLKGVTPL
-----EQLLPDELRYRLVKSFGKISFSKNF
LR----KLKS-----LCYSKDLATIKRWWQTVDGLVYPLLLSV-----ED
PDLYYAEIDHLIK-FGLENCA-----NNYAHFISRMK-----SLKRLLRKDYAE--
-----MKDSAKALTDVST-YRK-----IMMKFFPKDEANPH----L
TVQYLLCWTQTRA---SG--LAD--RHMIRKSLDKCEAT-LSEVRRIARISTPET-----
----REYSYRGLPEGEPELPQD--ILNDIAWRVSDKVK--GETAHL SAGPN--ACIEST
RQTG-GQTG---YIK-----HLARMKDSLHTRYNFCTLEPEPITPSPVRSQN-----

-----DLLSWAI--QEALE
HPTY-----
----VRCVRLHA--VPEPS--KARTITVCSLPYLLI-VGVMA-----KLLQP-AI
ASDIS-----RGGLHASR-NLWNF-----LFNDLDPSLDLWGYLRDDGDPKVP
----IMGLSS-----DLSEATDYGDI--GVARQIFSRVLEK-CSSVPC--
-FPTA--LGNLAKTLFIG---RRYC--LRY-----
-HDGK-----GYRFFVKRNGWLMGDRITKVILT LAHEIV----V---
RSS-----
-----GLFCARICGDD--VIAF-SRRREDL--ERYLSEVTR-LGFKVSWEDSY
IS--SRL-AFYCEEA-----CIPLQRFREIPSV
CT-----RRG-----
DQSMYVDY-----
-----PRIRLL-----IPTRVETQAYSFTD--TGRFSLL-----GKE
SRWVHQNKLSQTK-----LFHRAS-----LLQHLH-----V
S-QSPDTLCPY--LPIEL--GGDG--AF--P-----PSHTFLM-----
----EVVRAKARDLRETH--FRMLQLLSK-----WGFRFVR--SMNLNDVVHKY-HL-
-----LVPKIK-----ELRELIPPESIIEGN-----ELLGSIKIAGVESPEKT
FFRIWRAWYWHQILQGTGREPPKLELGVS RHFPREETEVEEMTEEHFHQFFHHWKDSGFTF
R
DTPDYLVDTRYIHVKDYMN MNWGF GIRPAWVS YRPTMEQVGGRMLLNETNMESFIDH
IKD
GRPLNPIVGRTLHKYVEAD---SYILSECARNWSR-LNPPDMCVLVSGDKRLACELARIQ
RAYY-----VPERDIYDKVPWMMVDPNH YLFG-----MM
DPYEDYAE-----AYSSRFEVVV

DFGAVAYSAYLTGGVDMFYENPRIDTKYLWNNPQLGVFEASLDEKSATSKELVSLTTQ
PR
DQVLKRRIPWNFGSGLRGATMSP-----
>GFJG01122116.1
-----DRPNNKLG
EAPGPN DHTGKRPRSDTQEMSDAPPVKRPKLAVRRLLYNVVESV-----IKTAKRTKI
SPPEPQLPDVAQHTEESQYKCDIS---VVENN-----VLREICTRLAPYEYKGIKVR
-----EFRNHPAHRRWYF---MIGRRLTDRTNQFVQSLTFRDYST
LE----QIWR-----AVMHTVLLNPFDPSRF-----DP
DL--IESVRSFKL-WCITYASSHK----NGIVKLGKALK-----TLASWMQWFMLDNQ
ANTCPNRVEIVPGWNRSTKRPTFPWFSGHLS-HCR-----DFNRVKSDRDLH-----
--LLAQMRTFGRA---LP--CPP--KILVDEGLKATLDV-LATE-RKT-----PD
HVLECISSAVSYLNTS---WKIS--NMVRES-----HLSISVS--ASYDHT
QEK G-GSAA---TQR-----EILSPFLDHKVDQLLDGNVPIKGNRLFIDCFGNIMLHGF
SMGRSIRTCRLVDGLYATHATVYGREEISKNLGKPMFGRRTGNVLLLASIMKVQNYGH
FE
PSPK-----GFLRTDD--NQINI
PYWD-----ESIIDEGLPS
FIEEKPFPCRFIA--LAEPG-FKVRPLTIGYTPMMNV-LKNMR-----FMMQS-VL
LHDPRL-----RVGLRSTN-KLWTV-----LKLLKQKKLE-----YG-----
--EGLVL-QSS-----DYKSATDYIPL--DVIRTIWQTICDG-LTK-AHP-
-FRVF--RSLIWH-----PRQI---W-ADIPD-----
-DQ GK--YT-----RQEQQHRCGSFMGEPMSYMTLCILNLAV----E---
LLS--Q-Y-----RLRTLSTLKPIEYDGLNVLHGRGI-----
-----NLDVIIVVGDD--VLAF-RRDMGII--HIFHEVATQ-CGFIFSEGKDG
YS--SRI-GIICEDH-----LLVDRR-----
-----ENP-----
VTIRFLDI-----
-----VKSRL-----TNVARTHGDNRASVFGKGRML-----NMV
LDYLPNTMFKEHVWLAYR-----SVLERT-----IRLKLS-----D
L----DLPWF--LPPSL--GGLG-----LQLSG-----
----DIPDSG-----WKFICFL EARMSSA-ASWQDQLALVWSFRHLSNRQPHG-
-----WEPEQG--TVEMAVECLKELKHAPKEQ-----SLDPNLEKVIW-----
-----STDH-LRRFLRISGAAAPFVG
SAVQLMH-----DHEYRTEAEK-----LGLVPLADVID
QFERLTTFQSALEGTA EHR-FKTIQQWIRRSKKYW-----PKNMMNVR
KPLW-----FTNLSDFERRMIRRFDFGINIDLC-----SR
SVNYGPNL-----RIV
Y-----DVWSQRRRGMPRIGLGGNPLSLKT-----

>NC_013111.1

-----MSKRPLAACLSVEETSFYAKTEVSK-----LCGLVDRMG
KLPR-----TYAGDDL PIDDNLKPVARSV
-----LAPILGTRLTRERTEYICSLSSQ
LVVGISDVTYI-----LLGSALSVKEIRWQSQLFAIL-----AI

1
2
3 AGKLEAAFKSHTG-FLLATCLG-----DRYDSPDPSSS-----NYGGIAPGRIF---
4 -----PPRFQRKIW-MRC-----VLRRRPQDVTLASS-----
5 ---LVNIKRVAPP---VP-----DSFVQESLNKNRER-LSKP-QSD---CPDP-----
6 ----ILDMLERQVVRTVDEIVYNAKKEGWNRKIARSA-----FPSQS--ASFARS
7 IASG-GQLG---EIV-----PENWSLRFPTLLAMVERCQEVVPI-----
8 -----Y
9
10 GWHE-----SPETVLT--ENLRG
11 HIDA-----
12 ----CLEVRRSP--VLEP--FKVRTITMGPAEPYFK-ARRIQ-----GVLWD-LL
13 KHTRCTHLP----NRPIHESD-ISFFC-----SRRGDAVFPDEE-----
14 ----TFFVSG-----DYSAAATDCLSP--VLSTVAVDRLCDHLLSPENQV-
15 -LDPVHPWRALFHRVLVG---HKIM--E-G-----
16 -KRGE-----EVEIAAQTWGQLMGSPLSFPILCIVNLAI----T---
17 RAS-----IEHTAKRQLYL-----
18 -----EECGILVNGDD--ILFK--LPLRGL--ARWDFLVTR-GGLHPSPGKNF
19 VS--KDY-AVLNSTI----Y-----RVGRTA-----
20 -----
21 ELIPTIKG-----
22 -----NLINGT-----ACRGIERPRDGSLYFSD-----ASK
23 YSWGTTIG--DRARELIK-----GFPNHI----ADQLLS--RFLISMKPHLDLFP
24 R-ISWWAHPQY-----GGLG---L-----PVTRPGD-----
25 -----FLPVH-----RRVGAFLAC-----GGRRSQE-----FRMNMQ----
26 -----WLSNPVKQFNAFTSQYLMDIARSLSVP-----VKQIYNDETEP-----
27 -----E-FFLEREIMYSALRFGV
28 EFPTNA-----RGNRELLAAWR-----HFYRQLE
29 RRALRTRITDGDLPERKAD---RKGLFLLSPEKLI-----AGPRFKYVYDWACKNQG
30 GNIWSSSSRGYVLFDKHRVLESSALPFGKDEVKSSVS-----THFT
31 PVNEYRQR-----
32 -----IFFWT-----
33 -----
34 >NC_011065.1
35 -----
36 -----MSNLADHEQVRTYSAYCSRRLCGLTD-----RIGKLPECN
37 LAPA-----FLKEL-----KLPGLPVPQTLYLKATRWT
38 -----LSDRDEEREY
39 VP----VLIS-----HLITAVYDVVKFLLGRQIT-----RK
40 DIHYMVG VYTLLA-----MEGRLEKSFK-----EHTAWLVARAMGDV
41 APEVPPTDGISA-----GKIFSPRLNRIIW-MRC-----VLRRRPQDVSL-----
42 ---AHSLYQVKRV---AP--PLV--TKLVNKASEEMLRR-VTKEEDGEHHILRR-----
43 ----LEGEIERTVDEL VKNARA--ERWTGNIRGQP-----FPAQS--ASYEHG
44 VREG-GQLH---SLL-----GTGHGLNPHLIAMIQRGQYVYPVYGWD-----
45 -----E
46 SVGP-----NLKNVLE----MV
47 DCPE-----
48 ----FLPTRRSP--VLEP--FKVRTITMGPARHYFR-ARQIQ-----GVLWD-LL
49 KGSRST----HLPNRIQESD-ISFFV-----QRAGDRIFPGE-----EA-----
50
51
52
53
54
55
56
57
58
59
60

1
2
3 ----LY-VSG-----DYSAATDNLRP--SLSMTCVNRLCDH-LSQSGSAL
4 DAAHP--LRMLFQRV-LV---GHAI---M-----
5 -DGKRGEES-----EAGRQSWGQLMGSPVSFPVLCIINLAA----T---
6 RWV-----LEQVSGRIVSL-----
7 -----NQSGILVNGDD--ILFR--MPYRGY--ALWNRVVTA-AGLTPSPGKNF
8 VT--SRY-AVLNSEV----Y-----D---TSSRTA-----
9 -----
10 -----
11 DPVPFLKI-----
12 -----NLMYGT-----LARGCERRTSAELLYGDDLEAGGTL-----GHR
13 CRALIKG-FKPEQQ-----DYLITR-----FIRACQ-----PF
14 L-SQAPGVSWF--LHPAY--GGIG--L-----P-----
15 ----SVRRPRIEHH----LRVAAYLSC-----GGKDQAEAQVFMHWLQPS-KE-
16 ----FNRATL-----LRAMEVARELHVPT-----VQVSPEELKEAPE-----
17 -----RFEIP-VKQELLSALTGTEF
18 EVTD-----QGQSARLREWN-----RYYRLLV
19 RRSCLKTYVRGSKDTTGLGL--HAMGWEKAMRGPK-----FRT
20 VFDW-----SARSGGYSMLRKTTSYSRC-----PFS
21 SSESILP-----
22 -----SGGVR--VCYSFLNTQ-----
23 -----
24 >NC_011068.1
25 -----
26 -----MEPRQ-----ELLDPERVR
27 EEAK-----RIVR--WLCGL-----VDRGTGKLPAGDYQGKILNTV
28 -----SEICKRSKLPCGELAEALTKGRL
29 TR----SLVD-----YSELLISNLVVGDFVLEIYL-----GK
30 QSVRLSDIREMACKYTFYAINR-----RLEDYIK-----FQTAWLQARAMRDV
31 TP-----EPEAPSWLNEGFG-RCF--SALLNRKVHLRCVLRAR--PND
32 VSLAASLYQVKRV--AP--PLP--DDQIEKNLEKSLDR-LTKD-EEP---AGVD---EP
33 ----FLEDLKRECKRTVDELVQ--NARREGWNRKISR-----DCFPSQS--AAFENP
34 ISKG-GQLG-----QLVKENNTPRLPVLLGMFEYKGRVTPV-----
35 -----Y
36 GWAD-----DGDITLS--DEELG
37 REVP-----
38 ----AALKCRRSP--VLEP--FKVRVITMGPAVQYYR-ARRVQ-----GCLWD-LL
39 KHTRCT----HLPNRPVEESD-IGFYV-----RRRGADLFRGE-----EV-----
40 ----PY-VSG-----DYSAATDNLHP--DLSLSVVDRCVCDHLLSDDNRPL
41 DPVSP--WRVLFHRV-LV---GHRI---Y-----
42 -DGNSSRNT-----EVAAQSWGQLMGSPVSFPVLCIVNLAV----T---
43 RYV-----LEKACGRIVTL-----
44 -----EESGILVNGDD--ILFR--CPERTI--PFWTRMVTI-AGLSPSPGKNF
45 VS-----YRYCQ-----LNSELY--DMSG-
46 -----SRA-----
47 EYLPFIKA-----
48 -----NLIYGT-----LARGCERKRAADLC--YGDITTE-----GGT
49 FGHRARA--LIK-----GFGPDM-----QDRLMS-----RFLHSIKGF
50
51
52
53
54
55
56
57
58
59
60

1
2
3 L-EKIPVSWF--IHPRY--GGLG-----LPLTR-----
4 ----PVTHNP-----YHLRIAAYLSCGGEQSQEARCM-----MQ-
5 ----WLSAPT-----KSFNAATLLRI-----LEVARNCKVPF-----
6 -----RKVP-FALLHRAEAAGVDLEA
7 L-----FRKALLRSAPR-----LGVEY
8 PSNESGDMQRLGDWRRFFR---DVGRKAARTRCRS-----GTADER
9 KGLF-----LMSPDNAVKGPQYEIFDWASHNMGGN-----I-----WD
10 PSYKFRAD-----
11 -----PFSSS--ESDEPRAKEIGPNRGPE-----
12 -----
13
14 >GADD01004427.1
15 -----
16 -----SPKGPRVYPGRLGRAQM-----GAPAASAPG
17 GSPE-----ALIDFDLGPISWDLKRW-----LPDNPVVALWSVRGKANADD
18 -----QRRSFNLMTNMAIRRSSLEKFRK
19 LRHAERKTLSD-----LTSGLLDNLIASNPER-----IR
20 TARTDPDVRKILT-WAFSLGL-----YNVQHAISEWK-----KFCTLLKHRAEQSE
21 TPWPEPLSDFPGFGRDACPDCPALWARLCPW-LKI-----IWSQGASSKSDMT-----
22 ---RICHLATSRN---LP--TAS--AAEREKTLSSHASV-LTSTPIPD-----IE-----
23 ----RSSMLRMLSRRIGEMVKG--KRKGRSSGHC-----SLAQS--ACIEQP
24 AGKG-GRAF---YIQ-----SEFQTWASRKQTETISETTWFGQPFRLKEGLPMWVTMCR
25 TED-----A
26 SNPT-----MAIGESRQDFNMNFDDFRHEDFLAGLDKATG--YQLLQ
27 WSLE-----EGIRRGCLLGDPRSSSGLA
28 ISKEKLPEIKADI--VGEPG-GKSRVMTIGEGWLTIF-LQPAS-----HHLAG-LI
29 EHHEAA-----WAGFKRGW-QASEW-----AKRYRRKQHWT-----GE-----
30 --D--YV-LSS-----DLKTATDYCRH--DFSLSLLLGFLGEG-SG--QG--
31 -DNPY--LRAAAELL-CS---PRYY--T-G-PV-----
32 -EGSI--GP-----TSRGILMGDPGSKIVLTLFNLAAEEEEAFV---
33 RWA-----TPRMIPTISVGMVTIADSATSW-----
34 -----KIRNLAVSGDD--HTAR--GPKSYL--KGITQAHIA-NGMEVSESSNF
35 LS--PIG-GFYCEEA---L-----L---FRGLDL--QKLRA
36 YSVPFWDI-----
37 PYLEHMHV-----
38 -----DSLKLRLCL-----SPVMKGDATRDEKNPA--IGKASTF-----TGM
39 LRWFAGG-----WACTR-----PIFSAR-----FEQRFK-----T
40 F-LPE-GILRY--LREL--GGVAAPCF-----HLTAGMLE-----
41 ---SSLTKLD-PSVR---SAINLVLAGLAGE----RGELAKL-LAKIPANTSMRG-IP-
42 -----ENLARD-----QVTQILETF-----PGVQGRIDE-----
43 -----SQ-LKAMADDQMGMG-QWWG
44 DAT-----ARDRARFARRN-----NLISVVDALN
45 DIERPYLFREFLAPGQHAV---HEGPSQFKTRIYD-----
46 TTPW-----PVRVNRFRSLALSAPKPDLLDHHHS-----AE
47 GVSDGGLV-----GMIQKGKGLPP
48 P-----PRMYI--PERVIFGDDLVTLQTPL-----
49 -----
50
51
52
53
54
55
56
57
58
59
60

>GFOL01002916.1

-----AISEWK-----KFCTLLKHRALQSE
TPWPEPLSDFPGFGRDACPDCPALWARLCPW-LKI-----IWSQGASSKSDMT-----
---RICHLATSRN---LP--TAS--AAEREKTLLSHASV-LTSTPIPD-----IE-----
----RSSMLRMLSRRIGEMVKG--KRKGRSSGHC-----SLAQS--ACIEQP
AGKG--GRAF---YIQ-----SEFQTWASRKQTETISETTWFGQPFRLEGLPMWVTMCR
TED-----A
SNPT-----MAIGESRQDFNMNFDDFRHEDFLAGLDKATG--YQLLQ
WSLE-----EGIRRGCLLGDPRSSSGLA
ISKEKLPEIKADI--VGEPG-GKSRVMTIGEGWLTIF-LQPAS-----HHLAG-LI
EHHEAA-----WAGFKRGW-QASEW-----AKRYRRKQHWT-----GE-----
--D--YV-LSS-----DLKTATDYCRH--DFSLSLLLGFLGEG-SG--QG--
-DNPY--LRAAAELL-CS---PRYY--T-G-PV-----
-EGSI--GP-----TSRGILMGDPGSKIVLTLFNLAEEEEAFV---
RWA-----TPRMIPTISVGMVTIADPATSW-----
-----KIRNLAVSGDD--HTAR--GPKSYL--KGITQAHIA-NGMEVSESSNF
LS--PIG-GFYCEEA----L-----L--FRGLDL--QKLRA
YSVPFWDI-----
PYLEHMHV-----
-----DSLKLR-----CLSPV-MKGDATRDEKNPA--IGKASTF-----TGM
LRWFAGG-----WACTR-----PIFSAR-----FEQRFK-----T
F-LPE-GILRY--LPREL--GGVAAPCF---H-----LTAGMLE-----
---SSLTKLD-PSVR---SAINLVLAGE----RGELAKL-LAKIPANTSMRG-IP-
-----ENLARD-----QVTQILETF-----PGVQGRIDE-----
-----SQ-LKAMADDQMGD-QWWG
DAT-----ARDRARFARN-----NLISVVDALN
DIERPYLFREFLAPGQHAV---HEGPSQFKTRIYD-----
TTPW-----PVRVNRFRSLALSAPDLLDHHHS-----AE
GVSDGKLV-----GMIQKGKGLPP
P-----PRMYI--PERVIFGDDLVTLQTPL-----

>GFOL01000185.1

-----SFNLMTNMA
-----IRRSSLEKFRKLRHAERKTL

-----SDLTS-----GLLDNLIASNPER-----IR
TARTDPDVRKILT-WAFSLGL-----YNVQHAISEWK-----KFCTLLKHRALQSE
TPWPEPLSDFPGFGRDACPDCPALWARLCPW-LKI-----IWSQGASSKSDMT-----
---RICHLATSRN---LP--TAS--AAEREKTLLSHASV-LTSTPIPD-----IE-----
----RSSMLRMLSRRIGEMVKG--KRKGRSSGHC-----SLAQS--ACIEQP

AGKG-GRAF---YIQ-----SEFQTWASRKQTETISETTWFGQPFRLKEGLPMWVTMCR
TED-----A
SNPT-----MAIGESRQDFNMNFDDFRHEDFLAGLDKATG--YQLLQ
WSLE-----EGIRRGCLLGDPFRSSSGLA
ISKEKLPEIKADI--VGEPG-GKSRVMTIGEGWLTIF-LQPAS-----HHLAG-LI
EHHEAA-----WAGFKRGW-QASEW-----AKRYRRKQHWT----GE-----
--D--YV-LSS-----DLKTATDYCRH--DFSLSLLLGFLLEG-SG--QG--
-DNPY--LRAAAELL-CS---PRYY--T-G-PV-----
-EGSI--GP-----TSRGILMGDPGSKIIVLTLFNLAEEEEAFV---
RWA-----TPRMIPTISVGMVTIADPATSW-----
-----KIRNLAVSGDD--HTAR--GPKSYL--KGITQAHIA-NGMEVSESSNF
LS--PIG-GFYCEEA---L-----L---FRGLDL--QKLRA
YSVPFWDI-----
PYLEHMHV-----
-----DSLKLK-----CLSPVMKGDATRDEKN-PA--IGKASTF-----TGM
LRWFAGG-----WACTR-----PIFSAR-----FEQRFK-----T
F-LPE-GILRY--LREL--GGVA---A---PC-----FHLTAGMLE-----
---SSLTKLD-PSVR---SAINLVLAGE----RGELAKL-LAKIPANTSMRG-IP-
-----ENLARD-----QVTQILETF-----PGVQGRIDE-----
-----SQ-LKAMADDQMGD-QWWG
DAT-----ARDRARFARN-----NLISVVDALN
DIERPYLFREFLAPGQHAV---HEGPSQFKTRIYD-----
TTPW-----PVRVNRFRSLALSAPDLDHHS-----AE
GVSDGKLV-----GMIQKGKGLPP
P-----PRMYI--PERVIFGDDLVTLQTPL-----

>GFOL01000028.1

-----SPKGPRVYPGRLGRAQM-----GAPAASAPG
GSPE-----ALIDFDLGPISWDLKRW-----LPDNPVVALWSVRGKANADD
-----QRRSFNLMTNMAIRRSSLEKFRK
LRHAERKTLSD-----LTSGLLDNLIASNPER-----IR
TARTDPDVRKILT-WAFSLGL-----YNVQHAISEWK-----KFCTLLKHRALQSE
TPWPEPLSDFPGFGRDACPDCPALWARLCPW-LKI-----IWSQGASSKSDMT-----
---RICHLATSRN--LP--TAS--AAEREKTLSSHASV-LTSTPIPD-----IE-----
----RSSMLRMLSRRIGEMVKG--KRKGRSSGHC-----SLAQS--ACIEQP
AGKG-GRAF---YIQ-----SEFQTWASRKQTETISETTWFGQPFRLKEGLPMWVTMCR
TED-----A
SNPT-----MAIGESRQDFNMNFDDFRHEDFLAGLDKATG--YQLLQ
WSLE-----EGIRRGCLLGDPFRSSSGLA
ISKEKLPEIKADI--VGEPG-GKSRVMTIGEGWLTIF-LQPAS-----HHLAG-LI
EHHEAA-----WAGFKRGW-QASEW-----AKRYRRKQHWT----GE-----
--D--YV-LSS-----DLKTATDYCRH--DFSLSLLLGFLLEG-SG--QG--
-DNPY--LRAAAELL-CS---PRYY--T-G-PV-----
-EGSI--GP-----TSRGILMGDPGSKIIVLTLFNLAEEEEAFV---
RWA-----TPRMIPTISVGMVTIADPATSW-----

-----KIRNLAVSGDD--HTAR--GPKSYL--KGITQAHIA-NGMEVSESSNF
LS--PIG-GFYCEEA----L-----L---FRGLDL--QKLRA
YSVPFWDI-----
PYLEHMHV-----
-----DSLKLRL-----CLSPVMKGDATRDEKN-PA--IGKASTF-----TGM
LRWFAGG-----WACTR-----PIFSAR-----FEQRFK-----T
F-LPE-GILRY--LPREL--GGVA---A---PC-----FHLTAGMLE-----
---SSLTKLD-PSVR---SAINLVLAGE----RGELAKL-LAKIPANTSMRG-IP-
-----ENLARD-----QVTQILETF-----PGVQGRIDE-----
-----SQ-LKAMADDQMGD-QWWG
DAT-----ARDRARFARN-----NLISVVDALN
DIERPYLFREFLAPGQHAV---HEGPSQFKTRIYD-----
TTPW-----PVRVNRFRSLALSAPDLLDHHHS-----AE
GVSDGKLV-----GMIQKGKGLPP
P-----PRMYI--PERVIFGDDL VTLQTPL-----

>GFOL01002977.1

-----PAASAPG
GSPE-----ALIDFDLGPISWDLKRW-----LPDNPVVALWSVRGKANADD
-----QRRSFNLMTNMAIRRSSLEKFRK
LRHAERKTLSD-----LTSGLLDNLIASNPER-----IR
TARTDPDVRKILT-WAFSLGL-----YNVQHAISEWK-----KFCTLKHRALQSE
TPWPEPLSDFPGFGRDACPD PALWARLCPW-LKI-----IWSQGASSKSDMT-----
---RICHLATSRN--LP--TAS--AAEREKTLSSHASV-LTSTPIPD-----IE-----
-----RSSMLRMLSRRIGEMVKG--KRKGRSSGHC-----SLAQS--ACIEQP
AGKG-GRAF---YIQ-----SEFQTWASRKQTETISETTWFGQPFR LKEGLPMWVTMCR
TED-----A
SNPT-----MAIGESRQDFNMNFDDFRHEDFLAGLDKATG--YQLLQ
WSLE-----EGIRRGCLLGDPFRSSSGLA
ISKEKLPEIKADI--VGEPG-GKSRVMTIGEGWLTIF-LQPAS-----HHLAG-LI
EHHEAA-----WAGFKRGW-QASEW-----AKRYRRKQHWT-----GE-----
--D--YV-LSS-----DLKTATDYCRH--DFSLSLLLGFLEG-SG--QG--
-DNPY--LRAAAELL-CS---PRYY--T-G-PV-----
-EGSI--GP-----TSRGILMGDPGSKI VLTFLNLAAEEEEAFV---
RWA-----TPRMIPTISVGMVTIADPATSW-----
-----KIRNLAVSGDD--HTAR--GPKSYL--KGITQAHIA-NGMEVSESSNF
LS--PIG-GFYCEEA----L-----L---FRGLDL--QKLRA
YSVPFWDI-----
PYLEHMHV-----
-----DSLKLRLCL-----SPVMKGDATRDEKNPA--IGKASTF-----TGM
LRWFAGG-----WACTR-----PIFSAR-----FEQRFK-----T
F-LPE-GILRY--LPREL--GGVAAPCF-----HLTAGMLE-----
---SSLTKLD-PSVR---SAINLVLAGE----RGELAKL-LAKIPANTSMRG-IP-
-----ENLARD-----QVTQILETF-----PGVQGRIDE-----
-----SQ-LKAMADDQMGD-QWWG

DAT-----ARDRARFARRN-----NLISVVDALN
DIERPYLFREFLAPGQHAV---HEGPSQFKTRIYD-----
TTPW-----PVRVNRFRSLALSAKPDLLDHHHS-----AE
GVSDGKLV-----GMIQKGKGLPP
P-----PRMYI--PERVIFGDDLVTLQTPL-----

>IABI01004424.1

-----LDPLPPCPPVA AVISTEGNGFVSRQQSE-----PLARSPKDV
GPPKGDLSFVPTYGGFNAGIVDRPGGLATHL-----AWRNPLSEKILRRSRNGKPY
-----GVVAYSNGVK-----PSLREAILRRHPEICTYPDVVIRR
LSEDG--HRFN-----RIRGAYLSIEDAIVCA-----VP
EILSDDPVFPLLRRWVLTNGAVH-----EGVDQVIKMWK-----SMIVSWRWSSLRSR
TEQPLAVPGSIFFN---EKGHFVGFPEESI-LSS-----VMEDIFITGITSKH--DGT
---RVCHLCSARG---MP--PAG--QKKSEDSLREHHET-LTGQ-RDD---ISQE----
----RLDLVKKLARRIGRKVAT--VLSKEKGDFRSDG-----HVSATTS--ASVEDT
VKSG-GRAN---TLA-----TQFIPWAKKVATEPLEGNDILGNAVRERPGHARWRTTCR
EVPVPFERDCDCAGEDGVRPNSSNGLPAVHSLEANNREECPNLGLFSSIWRMPE----F
DMPT-----NLDNPLTGVDDESIG--KQMLQ
WAHH-----QGCSAGYIEGPLVGSREEFNTSKVVK-
----PIPVVRSV--VREPG-GKARVVSKNEAWFTTF-LQPFG-----HHSVG-LL
SRLPST-----RAGLSAAA-QGYEW-----AKSASKSTGIDGASG-----
----MLFLSS-----DLKTATDFCDH--RVSKSLLEGFYEG-VGTNP---
-DNPY--IRLCIEIL-CS---PRIL-----
-EDGS-----ITSRGIMMGEPGCKAVLTLFNLVA----E---
EEA---F-----TAHCQNYPEILDLAVVPLRKSW-----
-----RVP--ACAGDD--HIAY--GPETYL--RRITRNHHS-NGMQVSWPSNF
VS--PIA-AIYCEEM----I-----LKLPHLRFEVGYL
PKADYRIT-----
VHIDAMKV-----
-----RLLSPY-----RKETSGGPNDSDKTNPAVGKSSSF-----TKK
LDWQPKG-WELARK-----PFLERF-----RSRMMQ-----Y
L-PHIEAGYQY--LPRYL--GGLG-----VPSVIPS-----
----SLRTL SADLSP----RHVAAIEAIMDGRGTPAIRRCL--SRFASNDEARGIDL-
----SNTAAE-----DENIALMKRAIQGF-----DTPEEGVLTFE-----
-----A-VREKLGLSFDQAENMR
-----RRDFNRKAGRE-----LQLISITDALR
EIERPYITVELLK GKASEP-----KSRFDV-----
-RPW-----KVR YAKLQDDLETCGV PKNPNNVLV-----AS
TLDEVFNA-----VITGNDEYI-
-----PKERISRVISAFGVEVVR-----

>GEWH01000774.1

-----KFNL DSTPCARVAAALDDARYGEA-----QATRY SIPD
PDATGVELVTRIPWRNPLTEKSEVSFGGKR-----RQRRKPLKRPSYHSMIKPSQ

-----RFRILSVTPIIKELSRRLNLWLSE
DG----DRFN-----RVRGCFLGIEDAIYSSC-----PE
LRDDDPIWSKLLK-WTFATGSSRH----SGVDTVIKSWK-----SFATWVKWAALRSE
TEEPAV-----IRGFPPFDENGFTGFPADSQFSELIEPILRTGITSKGE--
---RICHLITSRG--LP--PAG--QSKADASLDEHFKV-LTSDSGIS----DE----
-----RVEEVRVIAKRIGKLVRGCIVKEGEDRFRSDGHV-----SATSS--ACVEST
VKQG-GRAA---SMS-----ANLLTFLNEIPKTETKGVDILNNPFIETGDLPRWRTTCR
DEV-----N
ELSRDCDCFETCLNKWECKNNTWSSSYWSIPDFESDLENMYDNPLMGVDHAIG--
PQMVQ
WAYL-----EGRYRGVFTGPRVGDQKQDFT
CTVSRPPQIIRSV--VAEPG-GKARVVTKGEDWLTVF-LQPFS-----HHTIG-LL
RRLPSA-----RAGLSAAA-MGYEW-----TKALSRKEGLD----KE-----
--DNLYF-LSS-----DLKTATDYCDH--RISKALLEGFYSG-IGW-YK--
-DHSY--IKLAIDLL-CS---SREC-----
-EGEF-----TKRGILMGDPGTKAVLTLFNLVAEEESFQ--
DWK-----RKNPEREVKW-----
-----RCP--AFAGDD--HVAL--GPKEYV--SQITVNHAR-NGMKVGVESNF
VS--KIG-AFYCEEL-----LAHVNG--TKFEV
GYLPKAEY-----EST-----
IHVDAMKI-----
-----RLMSPY-----RKETSAAPNDSDKANPAIGKASSF-----YKK
LSWFQDG---WKSACA-----PLLQRF-----RLRMMD-----Y
L-PAENQVYQY--LPYYL--GGLDMPLLEGNPE-----SVSDWTE-----
----LIPQLD-LAHQ----SAIQMVIDGKADSWLLRCLSRFASNDYARGIDLSN----
-----RYAEDS-----IRGLLQIKKISTQE-----TESYTYEQVVE-----
-----ELV-KTGISKTSIEDMR--
-----KRDFGRKAQDQ-----LKLISMSSALN
LIEKPYIVTNLLRGEVDSQKQKYDLVPWKTRYNR-----FIKDLESVGVSAE
GYEY-----KTGMTKDDIFNRFVSGKDIYPER-----RIT
SILTAYKV-----
-----GVWSTPSLLTETKLS-----

>GEWH01001802.1

-----PILRTGITS
KGEGR-----ICHLITSRGLPPAGQSKADA

-----SLDEHFKVLTSDS-----GI
SDERVEEVRVIAK-RIGKLVRG-----CIVKEGEDRFRSDGHVSATSSACVESTV---
-----KQGGRAASMSAN-LLT-----FLNEIPKTETKGVD-----
---ILNNPFIETG--DLPRWRTT--CRDEVNELSRDCDC-----
----FETCLNKWECKNNTWSSSYWSIPDFESDLE-----NMYDN--PLMGVD
HAIG-PQMV-----QWAYLEGRYRGVFTGPRVGDQ-----

-----QQDFL

TCTV-----
---SRPPQIIRSV--VAEPG-GKARVVTKGEDWLTVF-LQPFS-----HHTIG-LL
RRLPSA-----RAGLSAAA-MGYEW-----TKALSRKEGLDKEDN-----
----LYFLSS-----DLKTATDYCDH--RISKALLEGFYSG-IGW----
-YKDHSYIKLAIDLL-CS---SREC-----
--EGE-----FTKRGILMGDPGTKAVLTLFNLVAEEESFQ---
DWK-----RKNPEREVKW-----
-----RCP--AFAGDD--HVAL--GPKEYV--SQITVNHAR-NGMKVGWESNF
VS--KIG-AFYCEEL-----LAHVNG-TKFEVG
YL-----PKA-----
EYESTIHV-----
-----DAMKIRLM-----SPYRKETSAAPNDSKANPAIGKASSF-----YKK
LSWFQDG---WKSACA-----PLLQRF-----RLRMMD-----Y
L-PAENQVYQY--LPYYL--GGLDMPLLEGNPE-----SVSDWTE-----
----LIPQLD-LAHQ----SAIQMVIDGK----ADSWLLRCLSRFASNDYARG-ID-
-----LSNRYA-----EDSIRGLLQIKKIS-----TQETESYTY-----
-----EQV-VEELVKTGISKTSIED
MR-----KRDFGRKAQDQ-----LKLISMSSALNLIE
KPYIVTNLLRGEGVDSQKQ--KYDLVPWKTRYNR-----FIKDLESVGVS
GYEY-----KTGMTKDDIFNRFVSGKDIYIPER-----RIT
SILTAYKV-----
-----GVWSTPSLLTETKLS-----

>KX373304.1

-----MPTIRNELNPRV-----SFFAEKHGK
GTFK-----CPLSLRHYLGDSRHLRKKCL
-----LHPRAGLLKEV
LG----KSWK-----TFKCFKTSSIVRAHELYLCYLDICVCL-----DP
EILSIESNRSIGV-KVFNTAC-----FSTSTATSLYK-----EFVENVLILLK---
-----CKEGEERWTRQNP-F-KRM-----MTWKIFQDAIASVQLSKGK
SLERVLSVLTRN---FP--EPD--GKSIQKKKVDFIGI-ISQQPPKG--WKSEE-----
----YRSHFLESIEEIANECES---TTVSDCHI-----SVTAA--GSLKKT
VREG-GKFA---EMI-----EEVKGFLAETPSDEELYEFANLKWTCNTNEPRWKTFGII
GEINPMSAMT-----L
FTEGVDFLNEVPNQYLGPIDFSGGSLPPGIPEEDLFITDLRPFRMGLGAQFG--NQLLL
YCCV-----F
YKKDELPMIRASP--VLEGG-DKVRWITMASWRDLVI-QQAAA-----TIFRS-LM
ESHKEM-----KPIFSRAN-LAWVY-----LNKAREIGPSD-----
----IC-YVS-----DYSSATDTVDR--EFAEFILTNFIKRFRNRLSDP-
-LLNF--LELGIRNA-VS---PKVV--I-F-----
-PTGE-----RITSSRGVFMGEPMSKVILTLMFTI----G---
KAA-----KSIYKLRFPSMEKLTFW-----
-----APG---DD--LVAT--GPTEYI--DIYSELAKI-LGQILNHSKV
KS--RTV-FKLCEQW---F-----WVPGLKSSVG--TWAIT
TDPGKY-----RES-----

AWVDTVKL-----
-----KLLGAM-----SLSNHSFEEERNEVIGKAKAL-----SKI
LRWLPSD-TY-----PLEYKK-----LLRTWF-----LCRFEPKLPRTD
S-KTF---AFLMLPGHL--GGFD---LLLDDQ-----EIREAYQ-----
----KVSSFSRILFN----NYYDDFIVG-----SYLRNMLRNRSYRGFALSED----
-----YNLTVE-----KLVKYYIDLLEEA-----PYEEYKDSKVE-----
-----GLS-FRETDRALFKAGFLTE
NRIRDIV-----SRPLAFGSIWG-----RTLKHSPYNT
TPISVRLAKLWKNLTGDTE---RWKNISLKGEISE-----
-----DDFVVLCKENKGPVLKVYRRNDSL-----ED
EVLIGLPS-----MKVSLGILPHDG
Y-----ISW-----

>KU935604.1

-----MPTIRNELNPRV-----SFFAEKHGK
GTFK-----CPLSLRHYLGDSRHLRKKCL
-----LHPRAGLLKEV
LG---KSWK-----TFKCFKTSSIVRAHELYLCYLDICVCL-----DP
EILSIESNRSIGV-KVFNTAC-----FSTSTATSLYK-----EFVENVLILLK---
-----CKEGEERWTRQNP-F-KRM-----MTWKIFQDAIASVQLSKGK
SLERVLSVLTRN---FP--EPD--GKSIQKKKVDFIGI-ISQQPPKG--WKSEE-----
----YRSHFLESIEEIANECES---TTVSDCHI-----SVTAA--GSLKKT
VREG-GKFA---EMI-----EEVKGFLAETPSNEELYEFANLKWTCNTNEPRWKTFGII
GEINPMSAMT-----L
FTEGVDFLNEVPNQYLGPIDFSGGSLPPGIPEEDLFTDLRPFMRMGLGAQFG--NQLLL
YCCV-----F
YKKDELP MIRASP--VLEGG-DKVRWITMASWRDLVI-QQAAA-----TIFRS-LM
ESHKEM-----KPIFSRAN-LAWVY-----LNKAREIGPSD-----
----IC-YVS-----DYSSATDTVDR--EFAEFILTNFIKRFRNRLSDP-
-LLNF--LELGIRNA-VS---PKVV--I-F-----
-PTGE-----RITSSRGVFMGEPMSKVILTLMFTI----G---
KAA-----KSIYKLRFPSMEKLTFW-----
-----APG---DD--LVAT--GPTEYI--DIYSELAKI-LGQILNHSKVF
KS--RTV-FKLCEQW---F-----WVPGLKSSVG--TWAIT
TDPGKY-----RES-----
AWVDTVKL-----
-----KLLGAM-----SLSNHSFEEERNEVIGKAKAL-----SKI
LRWLPSD-TY-----PLEYKK-----LLRTWF-----LCRFEPRLPRTD
S-KTF---AFLMLPGHL--GGFD---LLLDDQ-----EIREAYQ-----
----KVSSFSRILFN----NYYDDFIVG-----SYLRNMLRNRSYRGFALSED----
-----YNLTVE-----KLVKYYTDLLEEA-----PYEEYKDSKVE-----
-----GLS-FRETDRALFKAGFLTE
NRIRDIV-----SRPLAFGSIWG-----RTLKHSPYNT
TPISVRLAKLWKNLTGDTE---RWKNISLKGEISE-----
-----DDFVVLCKENKGPVLKVYRRNDSL-----ED

1
2
3 EVLIGLPS-----MKVSLGILPHDG
4 Y-----ISW-----
5 -----
6
7 >KY628364.1
8 -----
9 -----MPTIRNELNPRV-----SFFAEKHGK
10 GTFK-----CPLSLRHYLGDSRHLRKKCL
11 -----LHPRAGLLKEV
12 LG----KSWK-----TFKCFKTSSIVRAHELYLCYLDICVCL-----DP
13 EILSIESNRSIGV-KVFNTAC-----FSTSTATSLYK-----EFVENVLILLK---
14 -----CKEGERWTRQNP-FKR-----MMTWKIFQDAIASVQLSKGK
15 SLERVLSVLTRN---FP--EPD--GKSIQKKKVDFIGI-ISQQPPKG--WKSEE-----
16 ----YRSHFLESIEEIANECES---TTVSDCHI-----SVTAA--GSLKKT
17 VREG-GKFA---EMI-----EEVKGFLAETPSNEELYEFANLK-WTCNTNEPRWKTFGI
18 IGEINPMSAMT-----L
19 FTEGVDFLNEVPNQYLGPIDFSGGSLPPGIPEEDLFITDLRPFRMGLGAQFG--NQLLL
20 YCCV-----F
21 YKKDELPMIRASP--VLEGG-DKVRWITMASWRDLVI-QQAAA-----TIFRS-LM
22 ESHKEM-----KPIFSRAN-LAWVY-----LNKAREIGPS-----
23 ----DICYVS-----DYSSATDTVDR--EFAEFILTNFIKRFRNRLSDP-
24 -LLNF--LELGIRNA-VS---PKVV--I-F-----
25 -PTGE-----RITSSRGVFMGEPMSKVILTLMFTI----G---
26 KAA-----KSIYKLRFPSMEKLTFW-----
27 -----APG---DD--LVAT--GPTEYI--DIYSELAKI-LGQILNHSKVF
28 KS--RTV-FKLCEQW----F-----W---VPGLKS--SVG-T
29 WAITTDPGKY-----RES-----
30 AWDVTVKL-----
31 -----KLLGAM-----SLSNHSHFEEERNEV--IGKAKAL-----SKI
32 LRWLPSD-----TYPEYKK-----LLRTWF-----LCRFEPKLPRTD
33 S-KTF--AFLM--LPGHL--GGFD--LLLDDQ-----EIREAYQ-----
34 ----KVSSFSRILFN----NYYDDFIVG-----SYLRNML--RNRSYRGFALS-ED-
35 ----YNLTVE-----KLVKYY-----TDLLEEAPY-----
36 -----EEYKDSKVEG-----LS-FRETDRALFKA-GFLT
37 ENRIR-----DIV-----SRPLAFGSIWG-----RTLKHSP
38 YNTTPISVRLAKLWKNLTG---DTERWKNISLKGE-----IS
39 EDDF-----VVLCKENKGPVLKVYRRNDSLE-----DE
40 VLIGLPSM-----KVSLGILPHDG
41 Y-----ISW-----
42 -----
43 >GFLH01017038.1
44 -----
45 -----SIGTLMHGSQTAHLGTRVGSIHWTV-----VLPQPHDVG
46 LKLR-----TSSKPPERIS-----FKGEQISSKKVFGEKGRSGL
47 -----RAYLCTTIGIGVKHAEVIVSRKTSEIRR
48 MF----RQWE-----SIIDNF-LLNPKFFELRF-----
49 ----QPAWKVLRAFLGVAP-----FSVQCTTRMYK-----EFVEFLKQEYFEP-
50
51
52
53
54
55
56
57
58
59
60

-----LEGEKISLPKDNL-FKP-----LVEVVRSVLNKDR--KTD
YV-HFAHICTTRN---LA--PGD--AKEQDNSIREFIEL-TTTEFEVS-----KL-----
-----DLSNLSLAMDRIVSICLN--SYESGGYLG AHM-----SISES--AEYNVP
QSKG-GKAA---AIL-----AEAGEILNRIADKDQKVETPFRS-YTHRQGYPLWMCIA
VG VKDP-----V
EWLS-----PKEMTIVDRPPPREGLDKTTG--EQLFV
AAWI-----A
FNESRSIKVRAVS--VPEPG-GKSRVVTVSQWWLPVI-QQPLS-----HLLSG-VL
SKHPSA-----YSVFRRQD-QAWQS-----LKLASRIKHTQTMSGFAW-----
-----LHS-----DFTSATDAVPH--KVAYCIMERLIMR-----
-CFPN--FKPVLRIL--G---EREV--E-----
-VKKD-----TFFTKRGIFMGEPFSKLVLVGSILAI-----E---
ELC-----YSIEKSISLRRTFTKPSNW-----
-----RFM--HVG GDD--HRCY--GPVSYL--KRLTRLCKQ-FGYIVSPTKHI
MT--RRA-GMYTEKC-----IYYHDR--VINMK
VQQINDNV-----EKS-----
VWVDSIKL-----
-----RLLSPF-----TKPLDGRDDRNIAIGKAKTL-----SRS
IQWFPMG--TELE-----TIARLA-----IERYKL-----RFSTMIPSS
N-RKTMTAIR--LPDSV--GGLG--L-----VYQPEVE-----
----DFRNLP-EIFW----WALNVIHAGHP---KNYKVRSVLQNTWQNFSGYEQLK-
----YLSDFK-----SQIEDYPNMVGVT-----FNALKEEFNEH-----
-----FNGD-NTEFVDFLREK-NIIQ
-----IHEFMKLIERP-----HKFY
KLMTRSRIKHFN TL DN RTR---IARTWEALENLKS---EITDPIPLIKEMEVIQGA AKLA
RHDW--FIKLDEPSEAAFI STELSEKQLDDEEYFTFM-----LD
GTISMRTI-----
-----PFSEH--ILYNAPELKLHNMKIQN-----

>HACV01002003.1

-----ILNIETGIPRDPAGVQDCSEGSSTLRLGSP-----SESVRAPLG
PVPSTKSRQWGRGLLVPTLSPDSVLQNPVRPIYEVKWGGVSLSGGITAYGSRKWRQRLR
K
-----YLPTVWGVSVRTAGILSGRSRSYLIQ
VF-----RNLD-----SLIDNLM LFDRTLFD-----QN
NAESRSVITRFVR-KVLTVSA-----FSIGDCIAKWK-----DFVNHFVTVSSDAT
-----VSSTVSLKGNCFK-FLL-----NEGKFKRLVNDKPS--TRE
DYMDIAHVISTRN---LA--NGG--VKAQQA AVSQFRKT-TSTEFKVS-----PE-----
----NLKKLRLVASRLGRIVTS--IRGSNKLRGHC-----SLNGS--GTLDVP
VNQG-GRAT--DALL-----DLKKFLSHIPPSGESRTYPWGEVWF--PSGIPRWCAFGK
LENFTD-----A
EWMS-----LANREI DRGALRNGCDSTTG--RMVMA
VAYE-----CYRPF
EESGTPIPIRQAA--VSEPG-GKVRMVTTGPWWLAVL-QQPLC-----HGLRE-IL
AYHPSA-----HSCLMRCD-QAWQS-----LHVLERLGISKLEDGMAV-----

-----LSS-----DLKEATDAIPH--SVGKVLLSGFLEG-LG-----
-REDH---DWITNF-IG---ERTV---F-----
CEDGE-----IFTLRRGVMMGEPLSKICLVLLGLAM----E---
ELA--F-----SEHEGLSLHRRDTPRCHW-----
-----RAF--HLGGDD--HLAV--GPVTYL--RSITSNHKL-FGSVISPVKHR
MS--SIY-VVYTEKV-----LRIKGAVLNLKAS
QVDGNP-----GNC-----
TFIDSVKV-----
-----RLLSPF-----SKATDSQNDKNIAIGKVRGI-----AAT
LKYFPSE-----SIKRTV-----FDRVAY-----RFNHYIGSE
R-HRTIRAVMS--LPSEL--GGLG--F-----AFDDR-----
----YLGRLP-PIFN----RAVRTVINGGG---VGFRVRRILGQIFANRCPRGLVSN
ALMVDDYVEQMI-----EQFLDYPGSVDMV-----SSGDLWARLDE-----
-----NGLMS-FREFLEKASHE-GWIS
-----LHDIPRIAEP-----FLMRKILE
GSEVPEYFRTASNAQRIAE--CWSNSLEAIDDVLD-----PTGGE
LSPL-----EIREASRRAKHTVFVNLNEMTCAA-----LVD
PTSEAYDP-----
-----GDPWISADFIDLSMRELLTYGEPSTVPL-----

>GFBP01062652.1

-----LTLRRYMPPTQLTAPPGRFEGGGV-----HSPGLLVPN
LTPD-----HVVVN--PVRPI-----WEHRFGGVCFASLGSKQKWR
-----TSFARTLRVELD-----MGRRASETLANRGLSYIKR
VL----KNLD-----SLIDNLILFSPSMFDR-----ST
PGVIYDTLKRILR-VTLKVSTY-----SVEDQISQWK-----DFVNHFFTLVSKSQ
VEKQICL-----RHNCYKWMLEQDP-FKS-----LLSKIDMTTKDWT-----
---SIAHLISTRN---FA--PGG--ESAKLKAIRAFKET-TTS--GFS---VSEE-----
----NLRKLKLATLR--VTGI--SLHLSEEKRLSGG-----HVSLNSS--GTLDVS
VAKG-GRAA---DAL-----LDLKQFLDEEPVQGETRVYPWGE-LYYPPLRRWRAFGP
LQD-----D
DGPF-----MSKSRDYDLIDREAMLNGCGEYTG--VMIFT
VAYE-----CYRVYLNSGE-
----LIPRQAT--VSEPG-GKARIVTTGPWWLAVL-QQPLC-----HGLRE-LI
GYHPSA-----HSCLLRAD-QAWQS-----LHVWAKCSDNCIPEGEAV-----
-----LSS-----DLKEATDAIPH--SVGQVLIRGCLEA-IG-----
-ATKW---LWMADL-IG---PRLV---L-----
-AEDK-----DVFTLRRGVMMGEPLSKICLILLGLTM-----E---
EIA--F-S-----EYSSLKLRKFTPAQGW-----
-----RAF--HLGGDD--HLAK--GPMRYL--KSITSYHRL-FGSIVSPYKHR
IS--VRA-VTYTEKI---L-----V---FENRIL--NMPVE
QVNSNI-----DKS-----
IFIDSIKV-----
-----RLLSPF-----TKATDTMNDKNVAIGKCIGI-----AGT
LQWMHDR-----SLKRTL-----IDRIHY-----RFRDFIAGP

1
2
3 H-HRTLSAIQS--LPTTL--GGLG---L-----SVDTK-----
4 ----YLENLPEPYNRAIREIIKGGVLAYKARMVMNSTFVNKNLRGIKSDDYVTELITQ-
5 ----WVEYPS-----SVNAMNFKEAIAVV-----DPEGNNFRHN-----
6 -----LWLLKKRN-IVSMNDIPQIAESAYV
7 -----MRRLLEQSQSGKYFRTERLRDRVSKTWEGLTKLDL
8 PESGDVLSDELKTGTRLA---KQALFIDLTTMGT-----IAVA
9 DSTN-----PDYLPDDPWLA AEFIDIPFGQSIK-----YG
10 QPNMGIDL-----
11 -----DHPFP--EEKDLAC-----
12 -----
13
14 >JN400242.1
15 -----
16 -----RS-----TLPQSWERG
17 D-----RYHSDLLKRDLMIRLTRDLG
18 -----LSGRHADRLVNRPAIHLKR
19 IE----EFIS-----GLVDSFWLGDPRTFAL-----GS
20 PD--FKTFRTIVR-KIFAVGT-----SNLGMLMANWK-----EWTNWFTHSVCETG
21 LREEAKLS-----NHNMFHRLARLSV-VQK-----ALSGGVPLMELGS-----
22 ---IVAHLTSTRQ---MP--YMG--LPTEVKAKADFIEI-IST--PFH---VEPS-----
23 ----HHRQMVQCAARMGRLCMS--LRNGVPVSDRSSHF-----SATSS--GELDHS
24 LTRG-GQAQ---ALK-----DAIDRWLIPASGSTYMEDTPFGV-AEHREGVPLWKTFLV
25 DQE-----TQLELIFSDFGDS-----L
26 DWIK-----DVPGRVYGLDDYTG--RQILY
27 VAWK-----EMED-
28 ----IPHIRAST--VPELG-NKARIVTLFAFWLNVL-QAPLS-----HIMKE-VL
29 KYHPSC-----FASFTRGE-QAWHA-----ASGLGRLNPRSVAG-----
30 ----YSVLSS-----DLKNATNAQHI--ALTRDMLRAFIYS-SG-----
31 -LVVSKAYVNLVLDLDT-IC---PRLV--E-----
32 -LDGE-----FTLSCRGIMMGEAIAKPSMTLLNLVV-----E---
33 GLA---F-----LKYENKLVLNNTDKAAPSRRW-----
34 -----RCY--HVGGDD--HLAV--GPDQYL--DLITHNHEL-SGSIISPDKHA
35 KS--RKM-VKYCERV---L-----L---VQNLQY--NTP-D
36 RKPF-----EHG-----
37 LIVDSIKV-----
38 -----RLLEKG-----QSTLIAKDNKN-VA--VGKSQQL-----VKS
39 LDWLPKQ---LYSKGFIR-----SIQHFLF-----IKRMGS-----LLPNRNR-D
40 E-EAFHSVC---LPKIL--GGYG--LG-----LAEDLLT-----
41 ----HLSKSK-PEIQ---QTVSAVLYGIETRRARNILSRL--NTTISDRSIQG-TK-
42 ----WYEDEM-----VQQFREY-----PDMVGAIDG-----
43 -----
44 -----
45 -----REMRR-----
46 --RF-----
47 -----
48 -----PAST-----
49 -----
50
51
52
53
54
55
56
57
58
59
60

>JN400243.1

-----RS-----TLPQSWERG
D-----RYHSDLLKRDLMIRLTRDLG
-----LSGRHADRLVNRPAIHLKR
IE----EFIS-----GLVDSFWLGDPRTF-----AL
GSPDFKTFRTIVR-KIFAVGT-----SNLGMLMANWK-----EWTNWFTHSVCETE
LREEAKLS-----NHNMFHRLARLSV-VQK-----ALSGGVPLMELGS-----
---IVAHLTSTRQ---MP--YMG--LPTEVKAKADFIEI-ISTPFHVE-----PS-----
----HHRQMVQCAARMGRLCMS--LRNGVPVSDRSSHF-----SATSS--GELDHS
LTRG-GQAQ--ALK-----DAIDRWLIPASGSTYMEDTPFGV-AEHREGVPLWKTLFV
DQETQLELIFSDFGDS-----L
DWIK-----DVPGRVYGLDDYTG--RQILY
VAWK-----
-EMEDIPHIRAST--VPELG-NKARIVTLSAFWLNVL-QAPLS-----HIMKE-VL
KYHPSC-----FASFTRGD-QAWHA-----ASGLGRLNPRSVAG-----
----YSVLSS-----DLKNATNAQHI--ALTRDMLRAFIYS-SG-----
-LVVSKAYVNLVLDLDT-IC---PRLV--E-----
-LDGE-----FTLSCRGIMMGEAIAKPSMTLLNLVV-----E---
ELA-----FLKYENKLVLNNTDKAAPSRRW-----
-----RCY--HVGGDD--HLAV--GPDQYL--DLITHNHEL-SGSIISPDKHA
KS--RKM-VKYCERV----L-----L---VQNLQY--NTPDR
KPF-----EHG-----
LIVDSIKV-----
-----RLLEKG-----QSTLIAKDNKNVA--VGKSQQL-----VKS
LDWLPKQ--LYSKGFIR-----SIQHLF-----IKRMGS-----LLPNRNR-D
E-EAFHSVC---LPKIL--GGYGLGLA-----EDLLT-----
----HLSKSK-PEIQ----QTVSAVLYGIETRRARNILSRL--NTTISDRSIQG-TK-
----WYEDEM-----VQQFREY-----PDMVG AIDG-----

-----REMRR-----
--RF-----

-----PAST-----

>JN400241.1

-----MQEDLILVQPD TDWVGGVNRSWRS-----LLPVLNGVE
YILP-----RG--SLVGL-----KLHGRSTLPQSWERGDYHS
-----DLLKRDLMIRL-----TRDLGLSGRHADRLVNRPAIHLKR
IE----EFIS-----GLVDSFWLGDPRTFAL-----GS
PD--FKTFRTIVR-KIFAVGT-----SNLGMLMTNWK-----EWTNWFTHSVCETE
LREEAKLS-----NHNMFHRLARLSV-VQK-----ALSGDVPLMELGS-----
---IVAHLTSTRQ---MP--YMG--LPTEVKAKADFIEI-IST--PFH--VEPS-----
----HHRQMVQCAARMGRLCMS--LRNGVPVSDRSSHF-----SATSS--GELDHS

LTRG-GQAQ---ALK-----DAIDKWLPASGSTYMEDTPFGV-AEHREGVPLWKTFLV
DQE-----TQLELIFSDFGDS-----L
DWIK-----DVPGRVYGLDDYTG--RQILY
VAWK-----EMED-
----IPHIRAST--VPELG-NKARIVTLSAFWLNVL-QAPLS-----HIMKE-VL
KYHPSC-----FASFTRGD-QAWHA-----ASGLGRLNPRSVAG-----
----YSVLSS-----DLKNATNAQHI--ALTRDMLRAFIYS-SG-----
-LVVSEAYVNLVLDL-IC---PRLV--E-----
-LDGE-----FTLSCRGIMMGEAIAKPSLTLLNLVV-----E---
ELA---F-L-----KYENKLVLNNTDKAAPSRRW-----
-----RCY--HVGGDD--HLAV--GPDQYL--DLITHNHEL-SGSIISPDKHG
KS--RKM-VKYCERV---L-----L---VQNLQY--NTPDR
KPF-----EHG-----
LIVDSIKV-----
-----RLLEKG-----QSTLIAKDNKNVA--VGKSQQL-----VKS
LDWLPKQ---LYSKGFIR-----SIQHLF-----IKRMGS-----LLPNRNR-D
E-EAFHSVC---LPKIL--GGYGLGLA-----EDLLT-----
----HLSKSK-PEIQ---QTVSAVLYGIETRRARNILSRL--NTTISDRSIQG-TK-
-----WYEDEM-----VQQFREYPDMVGAI-----DGREMRRRFPA-----
-----ST-YQESKAHAASA-GWLP
-----IDQLARRVTRG-----HLFGELMSGGAKRQNVFK
TRNWIQEFTRLRKRLEEEV---GYLPLPEYADLEVKTLRNALGA-----MT
TDLF-----IDSRQVTSFDKGVGDDSFDFDGE-----LL
RTYEAHQF-----NLVVGLQFIG
LDSTC-----NAFTR--RVRRRY-----

>KU295726.1

-----MQEDLILVQPD TDWVGEVNRSWRS-----LLPVLNGVE
YILP-----RG---SLVGL-----KLHGRSTLPQSWERGDRYHS
-----DLLKRDLMIRL-----TRDLGLSGRHADRLVNRPAIHLKR
IE-----EFIS-----GLVDSFWLGDPRTFVL-----GS
PD--FKTFRTIVR-KIFAVGT-----SNLGMLMTNWK-----EWTNWFTHSVCETE
LREEAKLS-----NHNVFHRLARLSV-VQK-----ALSGGVPLMELGS-----
---IVAHLTSTRQ---MP--YMG--LPTEVKAKADFIEI-IST--PFH--VEPS-----
----HHRQMVQCAARMGRLCMS--LRNGVPVSDRSSHF-----SATSS--GELDHS
LTRG-GQAQ---ALK-----DAIDRWLPASGSTYMEDTPFGV-AEHREGVPLWKTFLV
DQE-----TQLELIFSDFGDS-----L
DWIK-----DVPGRVYGLDDYTG--RQILY
VAWK-----EMED-
----IPHIRAST--VPELG-NKARIVTLSAFWLNVL-QAPLS-----HIMKE-VL
KYHPSC-----FASFTRGD-QAWHA-----ASGLGRLNPRSVAG-----
----YSVLSS-----DLKNATNAQHI--ALTRDMLRAFIYS-SG-----
-LVVSEAYVNLVLDL-IC---PRLV--E-----
-LDGE-----FTLSCRGIMMGEAIAKPSLTLLNLVV-----E---
ELA---F-L-----KYENKLVLNNTDKAAPSRRW-----

-----RCY--HVGDD--HLAV--GPDQYL--DLITHNHEL-SGSIISPDKHG
KS--RKM-VKYCERV----L-----L---VQNLQY--NTPDR
KPF-----EHG-----
LIVDSIKV-----
-----RLLEKG-----QSTLIAKDNKNVA--VGKSQQL-----VKS
LDWLPKQ---LYSKGFIR-----SIQHLF-----IKRMGS-----LLPNRNR-D
E-EAFHSVC---LPKIL--GGYGLGLA-----EDLLT-----
----HLSKSK-PEIQ---QTVSAVLYGIETRRARNILSRL--NTTISDRSIQG-TK-
-----WYEDEM-----VQQFREYPDMVGAI-----DGREMRRRFPA-----
-----ST-YQESKAHAASA-GWLP
-----IDQLARRVTRG-----HLFGELMSGGAKRQNVFK
TRNWIQEFTRLRKRLLEEVEV---GYLPLPEYADLEVKTLRNALGA-----MT
TDLF-----IDSRQVTSFDKGVGDDSFDFDGE-----LL
RTYEAHQ-----NLVVGLQFIG
LDSTS-----NAFTR--RVRGRY-----

>KU295722.1

-----MQEDLILVQPD TDWVGGVNRSWRS-----LLPVLNGVE
YILP-----RGS LVGLKLHGRSTLPQSWERGDYHSDLLKRDLMIRL TRDLG
-----LSGRHADRLVNRP AIHLKR
IE----EFIS-----GLVDSFWLGDPRTF-----VL
GSPDFKTFRTIVR-KIFAVGT-----SNLGMLMTNWK-----EWTNWFTHSVCETE
LREEAKLS-----NHNMFHRLARLSV-VQK-----ALSGGVPLMELGS-----
---IVAHLTSTRQ---MP--YMG--LPTEVKAKADFIEI-ISTPFHVE----PS-----
----HHRQMVQCAARMGRLCMS--LRNGVPVSDRSSHF-----SATSS--GELDHS
LTRG-GQAQ---ALK-----DAIDRWLIPASGSTYMEDTPFGV-AEHREGVPLWKTFLV
DQETQLELIFSDFGDS-----L
DWIK-----DVPGRVYGLDDYTG--RQILY
VAWK-----
-EMEDIPHIRAST--VPELG-NKARIVTLSAFWLNVL-QAPLS-----HIMKE-VL
KYHPSC-----FASFTRGD-QAWHA-----ASGLGRLNPRSVAG-----
----YSVLSS-----DLKNATNAQHI--ALTRDMLRAFIYS-SG-----
-LVVSEAYVNLVLDLDT-IC---PRLV--E-----
-LDGE-----FTLSCRGIMMGEAIAKPSLTLLNLVV----E---
ELA---F-L-----KYENKLVLNNTDKAAPSRRW-----
-----RCY--HVGDD--HLAV--GPDQYL--DLITHNHEL-SGSIISPDKHG
KS--RKM-VKYCERV----L-----L---VQNLQY--NTPDR
KPF-----EHG-----
LIVDSIKV-----
-----RLLEKG-----QSTLIAKDNKNVA--VGKSQQL-----VKS
LDWLPKQ---LYSKGFIR-----SIQHLF-----IKRMGS-----LLPNRNR-D
E-EAFHSVC---LPKIL--GGYGLGLA-----EDLLT-----
----HLSKSK-PEIQ---QTVSAVLYGIETRRARNILSRL--NTTISDRSIQG-TK-
-----WYEDEM-----VQQFREYPDMVGAI-----DGREMRRRFPA-----
-----ST-YQESKAHAASA-GWLP

```
-----IDQLARRVTRG-----HLFGELMSGGAKRQNVFK
TRNWIQEFTRLRKRLLEEVEV---GYLPLPEYADLEVKTLRNALGA-----MT
TDLF-----IDSRQVTSFDKGVGDDSFDFDGE-----LL
RTYEAHQ-----NLVVGLQFIG
LDSTC-----NAFTR--RVRGRY-----
-----
>KU295727.1
-----
-----MQEDLILVQPD TDWVGGVNRSWRS-----LLPVLNGVE
YILP-----RG--SLVGL-----KLHGRSTLPQSWERGDRYHS
-----DLLKRDLMIRL----TRDLGLSGRHADRLVNRPAIHLKR
IE----EFIS-----GLVDSFWLGDPRTFVL-----GS
PD--FKTFRTIVR-KIFAVGT-----SNLGMLMANWK-----EWTNWFTHSVCETE
LREEAKLS-----NHNMFHRLARLSV-VQK-----ALSGGVPLMELGS-----
---IVAHLTSTRQ---MP--YMG--LPTEVKAKADFIEI-IST--PFH--VEPS-----
----HHRQMVQCAARMGRLCMS--LRNGVPVSDRSS-----HFSATSS--GELDHS
LTRG-GQAQ---ALK-----DAIDRWLIPASGSTYMEDTPFGV-AEHREGVPLWKTFLV
DQE-----TQLELIFSDFGDS-----L
DWIK-----DVPGRVYGLDDYTG--RQILY
VAWK-----EMED-
----IPHIRAST--VPELG-NKARIVTLSAFWLNVL-QAPLS-----HIMKE-VL
KYHPSC-----FASFTRGD-QAWHA-----ASGLGRLNPRSVAG-----
----YSVLSS-----DLKNATNAQHI--ALTRDMLRAFIYS-SG-----
-LVVSEAYVNLVLDLDT-IC---PRLV--E-----
-LDGE-----FTLSCRGIMMGEAIAKPSLTLLNLVV----E---
ELA---F-L-----KYENKLVLNNTDKAAPSRRW-----
-----RCY--HVGGDD--HLAV--GPDQYL--DLITHNHEL-SGSIISPDKHG
KS--RKM-VKYCERV----L-----L---VQNLQY--NTPDR
KPF-----EHG-----
LIVDSIKV-----
-----RLLEKG-----QSTLIAKDNKNVA--VGKSQQL-----VKS
LDWLPKQ---LYSKGFIR-----SIQHLF-----IKRMGS-----LLPNRNR-D
E-EAFHSVC---LPKIL--GGYGLGLA-----EDLLT-----
----HLSKSK-PEIQ---QTVSAVLYGIETRRARNILSRL--NTTISDRSIQG-TK-
----WYEDEM-----VQQFREYPDMVGAI-----DGREMRRRFPA-----
-----ST-YQESKAHAASA-GWLP
-----IDQLARRVTRG-----HLFGELMSGGAKRQNVFK
TRNWIQEFTRLRKRLLEEVEV---GYLPLPEYADLEV-----KTLRNALGAMT
TDLF-----IDSRQVTSFDKGVGDDSFDFDGE-----LL
RTYEAHQ-----NLVVGLQFIG
LDSTC-----NAFTR--RVRGRY-----
-----
>HAFc01092190.1
-----
-----MRELSYKLIRVTIPRGATGKSVSEAYRA-----ILPCLLPTD
QEDV-----YNYTLPLGIFKL-----KFKGQSPLPQLWEKADSVYL
```

-----SEREKKIFRFHL-----KTGLGLGKRVTEVFLSRPRSHYLR
TF----GFID-----GLVDAFWAADPKFF-----LV
GSKDYKIFKSIVR-KIHKVGC-----TDLGRLVDDWK-----DWGTRIFHQVARTK
TIGSLPEAK-----PENIFRILSELSW-IKS-----ILAKDVKDLDSLGDG--N--
---LLASIRSSRQ---FP--YMG--AKTEKKSLERFKEV-VST--PFI---VDPE-----
-----TSWQMVATARRIGGICRA---IRTRPIPDMAAHF-----SVSSS--GEYTYT
INKG-AFSQ---AVF-----DGIVPYLMRIIDTSSEETPFGLV--HYQSGKEVWRYLFR
DEEISQGEFLDNR-----K
GFPK-----CQLPRLGLDETFG--KQVLW
AAWK-----ASLEFQGK-
----PIPVRAEI--VPEMG-NKARVVTVCWWLQIL-QAPLA-----HACIA-MM
KWHPSV-----YSSFHKMD-QAWCA-----SVGLHKVKDPDLTKY-----
----WL-LSS-----DLKDATNAQSI--DLTKAMLKAFLLG-SNLLSGRE
AYVNY--CLELLGSTRLG---SREV--H-F-----
-KDKT-----VVLSKSGIMMGEPLAKVGLTLLNLSV----E---
ELA---L-----LTHIGRLDLLKPDPTPKLKW-----
-----RFC--HIGGDD--HLAW--GPEPYL--EKITELHYK-VGSHLSPGEHG
HS--RIF-VRYTEKV---L-----NVMNMKGSPMDAQVNKI
DAFKR-----
IMVDGIKI-----
-----RLLESG-----QSTMISKDNKNVA--IGKAGQL-----VRS
IQWLEPK-IGFTESRIDSIR----DLFCLR-----LNRFLP-----SQHRNWRCYQ
A-TK-----LPRVL--GGFGLGIT-----PSEQYQA-----
-----YVDSP-SPIR----WVVS KWKAGEPIVDELQILSFL--NSNVSTRGIASVEA-
-----YRKEIL-----DQLQTYPNLVNAK-----SWHDICELFPV-----
-----EQDAP-SSMIVDKAAEAGYLT-
-----FEKFAEMSTRG-----AIFQ
NLLLATEEVKIFQTRPYIQ--TMRRVTHQLQLRG-----VDQFGDTPIE
QADF-----EEMLKEIGRPIFFNIDQVTTCDIG-----EYD
PLKPEEET-----
-----WNFQDTTFRELYTKGTPSLIVGLSFLGLRY-----

>NC_035126.1

-----MVS-----GLVDALWCADEFVF-----VL
GSIDLKTALVR-KIFCVGT-----FNLGLLVDDYK-----AWANRLFHKAASSTQ
TIGQMEPLR-----CNNIFKMLDSIEY-IKA-----IQDSGSDCSFINLQ-----
---KLSHLVSSRQ---LP--YMG--LKTEIKSLKGFKDV-VTSDYRPT----LE-----
-----TLDLLTKAARRIGGICRS--IRKDRPIPDAAEHI-----SMTCS--GEATTS
IIDG-GQAC---AVR-----EGFERIMGRIAESDLLEDTPFGV--ASHREGYALWRTIFR
EEPVEGGSLYDPLY-----Q
GYPK-----DQPGRYLGLDRVLG--KQLLY
VAWK-----ESILNP-

-----TIVLRASI--VPEMG-NKARVVTMSPYWVQLL-QAPLA-----HLTIA-GM
RLHPSV-----FSSFARQD-QAWEA-----AKGISRLNIDNWEDL-----
----HV-LSS-----DLKDATNAQQQ--LLTRTMLRAFMNG-YG-----
-MIHRNTYSELALST-IV---ERLI--T-F-----
-EDQT-----SVWATTGIMMGEPIAKPSLTLLNLAI----E---
ELA---F-----LTYTNRLDLLSDLPSPYRAW-----
-----RFI--HIGGDD--HLAI--GPKKYL--ELITNYHLL-AGSHIDPGKHG
YS--QVA-VKYTERI-----LSVKNFRFKAPFD
YDNY-----PKS-----
MIVDSVKV-----
-----RLLERG-----QSTMQKKDNKNVA--IGKSRMM-----SGC
LDWLPSD--PAYW-----PSDKKI-----SIRNLF----INRMGPLLPSKS
L-HPKCYHSVQ--LPSIL--GGFN---L-----GLKSEIY-----
----QAYTLAPPIR----WVLNKAAGVDVTAELKILSLL--NRNISDRGTDPLQE-
-----YRASII-----DQLYSYPQLIGAI-----TWKEILSKFPP-----
-----QENNV-RWSIYAAREAH-----
-----YLSIEEFAEWSTRGNTFVQLIMGRGKNNVFNTRPY
VRTFARVWTELEALQTDLYGDSPLGEGEFNKTLRG-----LSRCLYFDTSQSTTID
IGYY-----KPDDPDREEFEFVDATYLEAFDRM-----LP
TLVVGKKF-----
-----IGFRN-----

>HACV01003050.1

-----AECMNFCGKLIQAQHLRDCVKGRRNGSP-----WDKLLPVLG
NPNGQQS--NTLLY---YHMPSG---SLLSQ-----TFKGRSLRKQKFEK-ESKYF
-----SPRTKYGLFVLL-----TRDIGVTKQIAKILVDRPVSHFKR
IQ----EFIN-----GLVDSLWLANEQVFLE-----CS
QE--LLIVKKTIR-KLFCVGT-----FNLKDLVDQWK-----EWGNYLFHTVARTE
TIGPLKAVA-----KNNIFRTLNGIPY-ISR-----MYHTN--EPDMLL-----
--QHLSHLVSSRQ--MP--YMG--FKTELDsrkkfkev-LLS--SYE--PPKD-----
----FLQKISMAARRIGGICRT---LRP-K-LNSGE-----THISVTSS--GTITHS
IEQG-GQAA--AVM-----EAMKRILLVPEESFEEDTPFGT-AVHKKGIPLWKTLYR
SSN----DDMSSYEFLEPYNL-----I
KEMS-----G---RFRGLDRVTG--AQIMY
VAWRELEATPILRAEVVPENMGNKARHVTLsAYWLNVLQSPLCHLLIDSMKYHWRELEA
T-
----PI-LRAEV--VPENMG-NKARHVTLsAYWLNVL-QSPLC-----HLLID-SM
KYHPSV-----FSSFNRQD-QAFEa-----VKGMCLKKELQLLPGHAV-----
-----LSS-----DLKDATNAQQW--SVTKAILKGYISG-ANLS---
-FSDR--YINLVLDL-IG---PRLV--K-F-----
-CDGT-----TVLSQVGIMMGEAIKPSLTLLNLSI----E---
ELA---F-----LEYNNATELLTSNDPAPYKQW-----
-----RFL--HIGGDD--HLAI--GPVCYL--NQITHNHML-AGSHIDPGKHG
FS--QIC-VRYTERL-----INIRNFQYKQPFH
RDDYSKS-----

IIVDSVKV-----
-----RLLERG-----LSTEMKRDKNKA-AIGKSQTL-----GGT
LEWLPID---DRFWT-----ETKKES-----IRALFV---ERMGELLPRKAV-N
P-RAYAAIH---LPQYF--GGYG-----
-----LGFKRE-----
-----WGKYLR-----DSPAPHRGLVM-----

-----LAHLGVDVRHH-----
-----LKVFRKLNSNLS-----

>GACM01002912.1

-----NFTGNLIQAHLSECAGKDRKRSPWDR-----LLPVLGNPN
KVDVNRQFYLMPSGSLSETYKGSSLMRQKFH-----KRERFFNPRQKAGLRILTR
-----DIGLSQRVAVQLVNRPIVHYKR
IE----EFIS-----GFVDSLWLAEEQIF-----LF
DGKWIKLIRKLIR-KIFCVGT-----YNLGSLVDQWK-----EWGNNLFHTLAEST
TIGVLRKPA-----ENNIFRRLDTIDY-ISA-----VYRGDKSMKLMQ-----
---HVSHLISSRQ---MP--YMG--SSTEKAWDKFQSV-LTSDFVPS-----EK-----
----TIVQLQMAARRVGGICRS---IRSRRIPDGV AHI-----SVTSS--GEYSYP
IAKG-AQAA--AVK-----AAMERILTVVPLEDQEEDTPFGL-VRHHKGIPIWKT LFR
TEPLDTE-----L
PFLA-----PYALIKEQEGRFAGLDRVTG--KQIMY
VAWK-----EYRP-
----LPVLRAEV--VPEMG-NKARLVTISDYWLNIL-QSPLS-----HVLID-AM
KFHPSV-----FSSFHRQD-QAFE A-----TKGLCSLKRKSLKGEAV-----
-----LSS-----DLQDATNAQQW--SVTIAMLRGFIQG-FGLS---
-FRPE--YVELVLST-IG---PRLV--C-F-----
-RDEI-----SILSKVGIMMG EAI AKPSLTLLNLSI-----E---
ELA---F-L-----RHCRAEELLYSVEPAPYRDW-----
-----RYI--HIGGDD--HLVK--GPIPYL--NLITQIHLD-AGSHIDPGKHG
FS--RIC-VKYTERL----L-----N--LSNLEH--GCP-F
DPSDY-----SRS-----
TIVDSVKV-----
-----RLLERG-----QSTLLKKDNKNVA--IGKSTQL-----GGC
LEWLPID--DRFYTETKK-----ASIRAL-----FIERMG-----SLLPRKAV-N
P-RAFAAIIH---LPTIV--GGYG-----LGMSS-----
----ELQKFLEASPE----PHKGLLMKAFCGVNVKEDLKIF--RVLNTNTSDRG-VE-
-----NIQQFQ-----QKII---DQLSEY-----PQMVNAMDW-----
-----WELKRKFPDPS-----GN-SKRTIALAADA-GILS
-----FEEFAKRATR G-----NLFQALLMGRKDLKVFN
TSPFVRTYKNIVWDEAESR---GLLTWSEGFNLNNDEIAIAIK-----NI
APQW-----YFDINQETAMDTGHWDPENPETET-----WD

FKDSTYID-----KYTEGFPSFD
V-----GFKVL--GLRH-----

>GFCJ01029165.1

-----SHLISTRHFPTGGKSVE

-----RQ
AVSDFVEVTTTLK-----NVDSDFLKKVR-----SASTIIGARL----
-----KENSRLNLQTHFS-----
---LSASGSLNKT--IS--EGGR-GQEVKEAIEENLKFI-----PIE-----
-----DKEIKILG-----WTVN--DIAGRPRWQT-----WFRDP-----
-----DWEN-----PEPDVPFGELLKDTLLGYPLYRQ-----

-----GFDEVLG--KQIMA
LAAN-----TFLEE
VSAGRRPSVRVIT--VPEPG-GKARIVTTGPWWLYIL-EQPFS-----HAMRS-LL
VHHPSA-----IAGLCRAD-QAWQY-----LRLLSNAKKYLGKSN-----
----YC-LSS-----DLKSATDAIPF--EICEAVFEGFFDG-IGQ----
-SPTM--SPVIAKLLTCG---RECI---L-----
-KSGF-----SYTTVRGIFMGEPMAKPILTLVNLVY-----E---
EIALRDF-----LKVKKQKPISVPW-----
-----RCF--AVAGDD--HIAV--GPLPYL--KMITENHLK-GGSIISPDKHG
LS--RVL-VRYCEKL-----IDLTRLKTDWTME
SINKDTLGY-----RNG-----
PFVDCVKV-----
-----RLISPA-----SKGNESFNEKNTAIGKGLTL-----GAT
LRWLNKD-HFPTKWV-----NMVRDR-----FFNRMG-----TLLPSRES--
-----GVFWHLLLPPAL--GGLG---LWIEDDVPLLNNRLPQPSRALVR-----
-----DIDANH-PDLK----KNLKLFEKELTSNSTFRGFKLGD--EPFEQCIDTAKEIL-
-----QSTKCT-----LRDLETEFPVLKGL-----NPRGQIAFLRT-----
-----KE-YVTEEELTKSIKRPFL
-----FNQILAGTSKR-----KPFN
TEALKHRYAKIWESLYKIEPNFELDEVVVRKALKY-----
-REY-----QDLYSIQERELIPFRNSEVFNNA-----IE
EITLGLPA-----
-----FSVEWSFAGVASDTQDQSTKNQPT-----

>GFIK01043228.1

-----RPFFCFNMSSFIACDLRAPAGRKHRYRL-----YIPRLFENG
DYIL-----PAKYDSVLNGCFL-----WKRKNLTVNVRGLVIQVLCR
-----DFGFSKRFAEVFASRNISEIRR
IE----QTIH-----AVLDCLLLYDTKLF-----I
ENRYYSLLKHIVR-RIMSVGT-----FSIGLVISYWK-----DFTNDIWMTVSEAE

YQGEISID-----KRNFFYEIERVLC-LKS-----LKGEKLDKVQLQ-----
---RFAHLTSTRG---LP--CGD--KQSEKRAILKFKEH-VTTPTQVS-----DY-----
----VVEDYGRAAGLIARKCAR--RFEGFVRAAHI-----SLATA--GSLDTT
VSEG-GRGE--EIR-----RAIFPMLTFIPKESSIVKTIPIFD-LNEVAGVERWRTWCR
PEPY-----M
DFPD-----KAFGTPNGETLGGCEVFIQGFDEAIG--EQILA
CAYF-----AYQQDKGDQV-
----EILTRVLT--IEEPG-GKSRIVTTAPWYVYVL-EQPVA-----HIMRL-VV
SNHPSA-----ESGLTRSD-QAWAY-----LHLLERAKDHLQDDFKG-----
-----LSS-----DLVASTDLIPK--PILEVLYTEFNKG-AH-----
-LVGP--LFEISLDL-LR---RQRF--CLF-----
-RDGS-----NCLTARGIFMGEPLTKALLTLYNLCT-----E---
EIA-----IRKYLSYDFIQPVKVPW-----
-----RCF--AVAGDD--HIAV--GPVEYL--RLITQTHLA-AEAQLSAGKHA
VS--SVA-LRYCEKM---L-----D---LRKVKS--DWKVK
T-----INS-----
SLENYLNS-----
-----PFVDSV-----KVRLSPCSKTNESFNDNRNTAIGKGKSL-----ART
LSWLNST--LFPTKWV-----SMVRDR-----FFSRMG-----PLLPSRSS--
-----GVYWHLLLPETL--GGLG--L--GC-----P-EDIPD-----
----LLGKLP-TLSK----SFLSEIAEGKITDLYRAFRSLLSNSTYRGYEMATTEIE-
-----LAKSLL-----FDSKWLFPCSKMREL-----RESYPEIEE-----
-----KP-VQNQLSYLMKQ-GWMR
-----YDRLEDTFLRG-----FLFKEILTGNVVKAFN
SQKIKKRYSDLWDRFYDGD--CVLTEEQLKKAWK-----FRDFFDFYQVKNE
VKAW-----IRKKLVCTALQELTYGLPDLTIR-----PS
LIAMRYHD-----EELRDEDKVLT
EALKASDI-----TTWED--AITHEVLKDLKAKLDSESQVEPKREALTNIVSESCDKKE
G-----
>KX883482.1

-----MQSTNLLTSKLIQTGCPRAPAGKLGRGYRI-----FIPHLYENG
DYLLPG-----RNHSLKGVVVRNRMVKPNL
-----LISKLTLMGFGKRFATLVSLRPISHIER
AE----QSVL-----GVIDSLLLYDSELFITDTG-----
----YKLLKFIVK-KTFKVAI-----FNCALVTKLWK-----EFSFYVWAHVTTETV
-----TREKPVRKRENFF-FSL-----VSEQDIEVLKLTNLN---KK
TLTRLAHLTSSRQ--FP--PGE--RQQEIKSLKDFESH-VTEPYIGN-----PV-----
-----FLRRIRLASRV---VGRKARRLGMPKPLTDSHL-----SLAAA--GSFYTS
VKDG-GRAE---ELL-----NMLFKYLAYVPKESCEIKTPFYT-LKDVEGVERWRTWCR
PAVY-----D
HFPN-----VSFGNLLPETLMGFETYQGFDEILG--MQILA
CSYL-----EQEEH
LKCQNIPVRVLT--ITEPG-SKARIVTTGPCWLYTL-QQPVS-----HVLRG-FL
GQHPSA-----AAGLTMSD-QAWQF-----LFLLEKAKSHF-----GD-----
---DFSC-LSS-----DLTAATDVIPL--CILKELYEGFLEG-LRI-----

1
2
3 -EGPF--LNIVGKMI-QL---PRLC--S-V-----
4 -EKIN-----SYFLNSRGIFMGEPIAKVLLTLLNLSC-----E---
5 EIA---I-R-----DYLGVDFRASVQVRW-----
6 -----RCF--SVAGDD--HIAV--GPVGYL--DKITETHIK-AGSKISDLKHG
7 HS--RIA--VRYCEKILDVRNF-----K---GSWSKY--TIN-D
8 STEAY-----LAS-----
9 PFVDSIKV-----
10 -----RLLSPA-----AK----NVLSFNEKN-TA--FGKGKSL-----GRT
11 LQWLNKD-----CF-DSKWI-SLVRDW-----FFKRMS---S----LLPDRSS-G
12 V-YWH---LL--LPEHL--GGLG--L--GT-----E-RDFED-----
13 ----LIVRLP-SPSR----TLLKSIEDGN--PN-MGHIRLF--SGFTSNTTYRG-YK-
14 ----IPETEK-----EIAALFFQGILEEE-----KSDKFENLV-K----
15 -----LYRYED-----LI-PNTAIRLLRTQ-GWVR
16 -----KDYLEDKILRP-----FLFKEILTGRAKVKAFNT
17 EHLKSRYARLWDLTYRGEE---TISPQTLQVCFKP-PAYHLYYK--MGEKLEL-----PF
18 RGRV-----LKVN-----LLE-EGLLGMP-----DL-SIP-----WE
19 VIGDISGP-----L-----AP-----KRGNHARSRV-
20 -----IKSPR--RRNS-----
21 -----
22 >GEHH01003501.1
23 -----
24 -----EV-----GEQDRGPVK
25 SPLR-----MERTINKSKDVSLAPAKAVP
26 -----
27 -----MLWA-----KIRSVLEDVGCGKGLP-----
28 -----WRV-----PDKGFLAEKLLK-----RVAEQERKETGSGR
29 RA-----QNRELTGAASGFL-FRK-----VIPGTVGKTEAEVD-----
30 ---YLRSLGSDHR--HK--CGS--NCRWKLDVRRKVEE-LF---PIG-----
31 ----WDSEYKEYVER-----AYVGRS--ACLEVN
32 KKSG-GFAR--FVR----DMEWDD-----
33 -----
34 -----WVEGYWG--EGDIF
35 ESWR-----REVELGKEER-
36 -----KVSV--LLEDG--KTRTVTVSSGKQAWL--QPFH-----QMMYD-RL
37 ARQKWL-----VRGEVTRKQ-LKRMV-----EGSGLM-----
38 -----VSG-----DYEAAATDNFCP--EHSRWLMEVISSR-CR-----
39 -YVPVDLLQLMENRF-DG---GFLV--G-----
40 -KTGR-----VERRKGQMMGDYLSFPFLCLVNYLT-----
41 AWR-----SLGENESL-----
42 -----MINGDD--IAFR--APEHEI--EKWEKGVAE-AGLSLSKGKTL
43 KS--SWM-FSLNSTF----W-----VEGKTGMTKLRAV
44 -----
45 RAKTFLKE-----
46 -----SER
47 GMWARFD-----EVRKGWSG--
48 -----KDRQI---IG-----RSFLE-----
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 -----WRAHQK-----KKWDDVGGDMIWLVRKTKGKGSE-
4 -----WKRTVP-----AKVRSTIIDL-----YERCALEGVRG-----
5 -----GG-KNFCEDDPRSLYDLRP
6 LSYDE-----EGEPGGRFEWP-----GLYKGHRSEN
7 RQKKVIYFKPIEVVVRLKG---MKRSCPIRVRRVK-----NKE
8 RMGW-----VKKENPEHGIPAVISGEES-----N
9 ECSGEQGP-----
10 -----
11 -----
12 -----
13 >HAAN01020927.1
14 -----
15 -----
16 -----EKE-----GEQDQGPKV
17 SPLR-----MERTKNKSEDVSLAPAKAVP
18 -----
19 -----MLWA-----KIRSSLRK-----
20 ----VGCEGSLPS-WL-----PSKGWLAENLK-----KVAEKERESTGSGV
21 RR-----QNRELTGAASGFL-FRK-----VIPGTVTKAQAEDA-----
22 ---YLRLSGEKHR--HK--CGS--DCRWKSDVRRK---VEDLFPIG-----
23 ----WDSRYEYVER-----AYVGKS--ACLEVN
24 AKKG-GIRS-----WLAEQEFEFFWQGAWGEGE-----
25 -----
26 -----
27 -----EFEEYAE--LCRLG
28 GEER-----
29 -----SVKV--LLDDG--KTRTVTVSSGKQLWL--QPLH-----QLLYD-HL
30 AREKWL-----VRGEVTRKQLE-----RMAKGH-----
31 ----GVMVSG-----DYEAAATDNFCP--EHSRWLMELISTR-CRY-----
32 -VPDG--VLKLAESRFVG---GFLI--G-----
33 -KAGS-----AMRCKGQMMGDYLSFPFLCLVNYLT-----
34 AWR-----ALGENRTL-----
35 -----MINGDD--IAFR--APNDRI--RRWEELVEE-GGLVLSKGKTL
36 KS--DWM-FSLNSMF----W-----VQSGSRMKKLRVV
37 -----
38 RAKTFLKE-----
39 -----
40 -----SER
41 GMPARLD-----EVTKGWSG--
42 -----RSKEI-----MCRSFLE-----
43 -----WRAHQK-----KNWTPVEGGIVELVVRTGAKGGV-
44 ----WKRSVP-----AKVRSKIIVDM-----YERCALEGERG-----
45 -----G-GRVFEEGDPRSLEYMR
46 PLEYWE-----EREPGGRFDWP-----GLYHGYQ
47 KGKMKEKKVIYFKPLEIVV---KLKGFKRGRPIRV-----QKVKKRR
48 DMGW-----IKIENPEHGIPAVINGEESNECGD-----RQ
49 GSVHAEGS-----
50 -----SVPPWAAV-----
51 -----
52 -----
53 >KT598235.1
54
55
56
57
58
59
60

-----MGSENVV
GSAS-----RNGRTLPLFASMTDTSYYK

-----NNRYSSVLYLP-----KG
TPRPCRDFMARTT-IVVQRAT-----SSVRRIHGPGVY-----
-----PTFSPSRIHTCAD-LLK-----QLKQFLATELDGASQEA--
---QMAFQSVKKL---LP--ASC--KCLKHGMLGDLRVR-LSRPPPSL---PVG-----
----YLSFARKVSREI--FTKG--WDHRWGSKVS-----TFSPSLGSCIGKS
RKHG-GQLS---ELS-----VAGQQAQWQESLNLPLPG-----

----SLEGELL--VDSSG--KPRPLTRFVSEAATL--RPLH-----GLLYD-TL
SKQPWL-----LRGEITAEKLRGAGF-----DRAREEPL-----
-----TSG-----DYKSATDNLSI--EVAETILDVAWSSAKY-----
-VPAS--VFRYALAA-QR---PSLS---Y-----
-EDDE-----GLISTFVPTRGQMMGSYLCFPLLCLQNYIA----F---
RYA-----EYVSGVEG-----
-----TPV--LINGDD--ILFQ--SELSFS--KAWMGIVGD-LGLEVEPTKTS
VS--TEY-GSLNSTL-----LRWGPQ--GL---

AVVKTIRM-----
-----GMLREV-----SHPSNLGTTALQFARVGPR--NTWLLNF-----EEF
LSWHVST---IVKWRCVASDM---GFTGRL-----ALRAWS-RFRGGRLLRDDVL-H
Q-MKIDRLPSAH-CPHNIVMGSEEFVTV-----PEESLTK-----
----ELKRDAAVWMASRK--WELGREYTAFKQGKVVSERANA-----TRIPNLL-QD-
----WRTSAR-----ELKEASSAVMRR-----KEYYWNHKKVLC-----
-----HGDFASRGLL-----
-----VRDRPSVSTWR-----
-----WV--GLRRTWWRDRVAR-----EGTRVP
KVLW-----EALHPPLSPFSRDDVVGDLLKQKTGGI-----FS
SAYLALKQ-----
-----AGVES--MVGQFFALVGP-----

>GFHZ01053830.1

-----SFSPRRCRAYRARTTELVR

-----ALAD-----WAYIFGYPTPALGA-----LS
DGGSCQLQFSASVK-GLLASCP-----SSVEEEVMAWQ-----SVKKLL-----
-----PDSCPCMNEPLLRG-----VVS GFRRPAVS-----
-----LP--SGY--LSFVRREVRRLF-----PKA-----
----WDSGLYESNVL-----SCSPGLSATTESS
RTEG-GCLS-----GWTGSHAYFLECCLRETA FEV-----

1
2
3
4
5
6
7 -----RDSASLLV--VQSAG--KPRPLTKFSDSTLVL--KPLH-----QTIYD-RL
8 RRSRWL-----SVGDVSSETL-ARAGF-----VKVEGE-----
9 -----VL-TSG-----DYKSATDGLSI-EVAEVILETLLLNAVS-----
10 -VPRS--VCAAALDI-LR---PTLF---H-KGLCP-----
11 -EGLE-----PRVGQMMGSFSLFPLLCLQNRIA----F---
12 LFA---L-----RRCGLTGPQSEK-----
13 -----VPC--IINGDD--ILFR--SAPALS--ERWMALVSE-LGLEVERTKTS
14 VA--AEY-GSLNSTL-----LRWRGT--NL---
15
16
17 RVAPTLRF-----
18 -----GRLRGS-----GYVTSL-----SRE
19 FRMFVEG-----LKGGRR-----FRAAVV-----F
20 F-RRHLGSLRS--VRLSL--LELG---F-----RGRLA-----
21 -----ARMSELFSLVPPGPPSEVVPLP---PVGHNVQLT-AD-
22 -----VATFVD-----ADSLGPELTAANAR-----ETASWKFSVSY-----
23
24 -----SG-DRERAAVRYCM-ALSR
25 VRPPEP-----FVSFRERPLWR-----
26 -----LLT--PESPAASRSRVLS-----
27 ---W-----FLSPLPAEPPKVAL-----MNR-----LL
28 ACQHELFD-----PP-----
29 -----PSYES--SVAF-----
30
31
32 >GEWO01002579.1
33
34
35
36
37
38
39 -----YSRSKG-----GKFSEFRDRVV---
40 -----DEFTFTPINEI-VGH-----
41 -----MP--KID-----LID-----
42 ----VTGNVFLRAED---WVYG-----TTIRDC--AYPDL
43 RDEG-ASADTRIGLF-----GLLWAVLEV LKDPDIETCTND-----
44 -----M
45 AWP-----GLGSGIS--PQIFL
46 DEGT-----
47 -----VRAKVVS--LPEPG-YKARVITIVPLAISII-GSVAR-----HLLDD-MV
48 QSDMT-----KLGLLSKDKL-YSFLR-----KMNGGKISGTADVSDSCK-----
49 ---FAYA-ESV-----DLTTATDTPPL--SSVRSLLQG TIAT-YLY-DKH-
50 -AGAD--FCQFAASL-STL---PRTF--V-L-----
51 -PGGK--VT-----NPNDHRSGIMMGEGLSGTFLNVMSSAVRFLLRP---
52 TID---V-FSDYDGT TAVDADDFIGCNRDRIQEFLDGRERPMFEPD-----
53 -----PTS--TQSGDD--VLVFDNLAPGMA--RRYLVLLYR-IFGLLPSGTTF
54
55
56
57
58
59
60

1
2
3 FS--SEFYGTFCEEV----I-----V---RTHDSR--GW---
4 -----
5 TYVDTTKS-----
6 -----KLFQTT-----SVEGPEKLLGRIGAI-----ASS
7 LKYSGE--ALL-----LRAAER-----VDAMIS-----GHPTLSD-R
8 I-LRY-DLPLG--LPECL--GGI-----PH-----PNGSLSD-----
9 ----FVACLD-----LVSLEILTGA STVPDAVLLEIE-FSRFFDDLPSDQ-IH-
10 ----LVEDTV-----RNVLEVIGGW-----ESPSFVDSTIP-----
11 -----FQV-FKRENFLPQGD LKYE V
12 -----WVEVAEAVSRR-----HGLISIRSAI
13 EKISSMIRASIFLSGGVVP--PVNPLIRRRNRRQ-----KL
14 LTRL-----RQHSQSIGAVDNSTPRTVVKRLRT-----CE
15 NRYFRFQD-----ISDIFELHG
16 Y-----PSFVV--PFSLS-----
17 -----
18 >GEWL01001678.1
19 -----
20 -----LSDEF-----VPLGVWQVR
21 PSRQ-----CGSDLLPSCMAFEGDVL---
22 -----
23 -----QSWR-----AVEAALLLAGASAPHCSEVHELLLR-----TF
24 VRGGIRAFALTLKGWGKAVRGL-----LSLEPLDTHLV-----GYTRLPKNCVCALR
25 REGMGES-----DIRDFLDFNGKLGGIPP-----FLLFSEFNKVLLPGRRLSD
26 IQRIASLTTLGRS--AP--PAT--QEVLRGAVDAFVKN-VTT-----
27 ----HARTFDLFTIG--WEFG--QYAKGKSPVRSRI-----GLNTH--SSYDYS
28 RSLG-GKFA--EFR-----DRVIDDFLFWPIEEVVERKPNHDLIDITGNVFIRREDWS
29 AGTTLRDCA-----Y
30 PLLE-----RDEGSSADSRIGLLGLLWAVLEV LK
31 DPNI-----VAETVGMVQPGLGSRDPPQ
32 IKLGGFVPAKVVS--LPEPG-FKARVITVVPLAVSII-GSAAR-----HLLDN-MV
33 WADDMS-----RIGLLSKN-KLYDFMR-----VLNFGRIEGNATDGVSCN-----
34 --HLPSV-ESV-----DLTTATDTPSL--QAVGSLLSGATNG-YL-----
35 -EDSI--AFEF SQMA-VSLATSSREF--T-L-----
36 -PGGR--KL-----SYLLHESGIMMGEGLSGTFLNMSSLVRFL LPE---
37 AVI--AFWDYKGN SVAEADV FID LNSGWIQLFLDGQGRA-----
38 -----RFEPDPHSTQSGDDVILFND--LRPGEA--RRYLVL LYRIFGLKPSETTFY
39 SS--GTF-GTFCEEV----C-----V---KTKSSN--GW---
40 -----
41 TFVDTVKA-----
42 -----RLFQVT-----ST--EGPEKIL-----SRI
43 SAISDAL--KYNRDDTLI-----LRAAEA-----VDAMIS-----THPTLSD-R
44 I-FRY-GLPLG--LPECL--GGI-----PH-----P-IGLASDFVACLDEIS
45 LEILTGA STVP-DAVL-----LEIEFSQYFD--DLPDAQVELI--RKLIQNLLGTI-DG-
46 ----WERPSG-----LVTAEPFQVYTRER-----
47 -----IL-PQGD LQYSTWV-GIAT
48 -----AISERHGLSSV-----RTVV
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 EKISAQVRASVFLSGGSVP---PVNPLIRRRRRRD-----RL
4 LKKL-----REHTTSIGGVDLSTPKT-----V---VER-----LR
5 SCEDRYIR-----F-----SDVQDVFELYG
6 L-----PSFVI--PFSLS-----
7 -----
8
9 >GEZJ01006036.1
10 -----
11 -----AGGQDNVPQTGGPTRPARNYLLSSVK-----SLASVGRKV
12 LG-----KPRTLNLDDGVLAQMTSGRH
13 -----KKPLSKPQKKHLRE
14 LR---RNWS-----ATRGFLELTFGVKKGK-----IF
15 APNNRKAIREQVS-FLLEVLE-----SGLQVTAK-----RWSSLLRSKSI---
16 -----YQGTSDSPCTGIG-----TRGKTPMA-----
17 ---LFLASTYSRA---CVVDPPT--KDEVQTMVLAAIER-ISTREPLP---EPD-----
18 ----VLRHLKGFIADL--FRSQ--RMKPRRAPVLP-----VPAPT--SCYERS
19 GSEG-GSII---ALT-----RFGRQDLSEAVHEHIARLFAESQRRAPQEERHDYNWTDL
20 DEPLY-----E
21 GIPR-----DPQEMIRCLDRQGAMDLMMC
22 PPPP-----PRLTLPETFKKAV
23 KDSSRPEERRAKALPIVQQD-GKIRVATLHMSGVVWS-ARAMS-----AFLLP-RL
24 RNLTFS-----KDMLRNRT-VQLES-----HRSDDL-----
25 ----LV-YSA-----DLSKSTDPISI--ELSRFVLQEITRH-TG-----
26 -KPTW--WNDALAV-IT---NHQI--E-Y-----
27 -TADS--VK-----KTVMSECGALMGLGPGWTVLSVLNAFA----A---
28 EAA-----GARQGDY-----
29 -----AVCGDD--LIGL--WPTEVC--DRYEGNLAS-LRLKANKEKSF
30 RS--THR-GVFCERL----V-----LRSDNR--ASARS
31 CV-----
32 RLGEASGA-----
33 -----RAIDGG-----KGRVVV-----DAL
34 REVAS-----SSEAGR-----IIRQTA-----YRTIMRMA
35 L-SQT-----AKGSMKEGGGG-----IGQAD-----
36 ----AVTLIS--YLL----HGPTQLYKVETDSRIAAMKTLLRKAPNLSTGVPVE-DV-
37 ----LTAGRA-----QIELNRRLETA-----ERGPRPKYLPS-----
38 -----QS-VRRRIHSRRKDAKKAI
39 ADYGG-----VINAVKGLLAS-----
40 ---GCYVRENNRIRDLTH--AVRGRAYTRALRI-----A
41 RSSW-----RTKVCPhAAKEAFEVSFPEHQPRV-----TS
42 DLEPMP-----
43 -----QAWDS--RLAD-----
44 -----
45 >GDKJ01026489.1
46 -----
47 -----STWGGMVAKTGLGSTSRK-----TAASGSISV
48 CPPR-----STQQKPSGKVAKKVSHNLSQ
49 -----LTTLGSLLLVWQDLVKSNAKAL
50
51
52
53
54
55
56
57
58
59
60

1
2
3 TRVQRALHLYK-----AAADGKGNNLKVFREKKRRSRTMK-----LR
4 FKAietaflpilhawllv-----NGGASCCSRPN-----NIGKWSFTSY---
5 -----MMAMLTDSKYLKSITHQ-----WREESIRHSNTRVA-----
6 -KRLFQASTLGRA---VP-FPHT--MLQEDRAYQGAIDA-WTQGPAFV---SKQ-----
7 -----HLDQIAEFISK---LPIR--RSKEKPGLHDL-----MLQNK--SCIENT
8 QAHG-GIAH---YLYTQEVEENKKRRDARLFKAGDLIQGQATTIRTKILMESFSRGEELD
9 SRL-----E
10
11 GQRA-----TMSQFKRLQQUEE
12 DIR-----QYRYQDLLDKARARFE
13 NTCHATQPMNPVI--VRELK--KLRCVTVHPADEVIT-ARTIT-----TKLLR-VL
14 RGCYST-----RDTLRGQC-PTVRT-----ETGR-----
15 -----PVLISA-----DLSSATDHVDH--RLAQFIGEKLLDK-MG-----
16 -LESS--KPLLKSI-FS---PKEL--R-F-----
17 -TKKE-----YKKHNGTLTSRGVHMGLGPSWVILCLINMYA-----A---
18 HNA-----GYNKDEV-----
19 -----QICGDD--LIAL--GSERAA--QRYQQTME-LGMVVNETKSF
20 VG--KR--GVFCERF-----LEIKVD-----
21 -----RIG-----
22
23 CYASALDV-----
24 -----GHIAEA-----SAVRTTSDPD--KSRLAAV-----ERL
25 YKERHLP-----TLARKV-----ADQLAP-----RGVGK--
26 -----GPIPM--GGNG-----RGAPT-----
27 -----TP-----QLVHFMKCGPIRLVQTQFSPKL--ELNASRVPIKDGIP-
28 -----WSGLRT-----EMMRHDAQMS-----YFKNESKPLTI-----
29 -----IR-AEHFSRLARGY-----
30 -----VRSTATVHLKS-----QSGTYLSLLPRAA
31 RQKVLLLLKRFTDSSNTIN--SNLSLPKATRAAL-----AKALTGPT
32 AERF-----VNLSTASNLLGLHHCPDADIDLA-----LM
33 RSRAKMTN-----
34 -----DGIRL--GPKEAELRRITNLDLL-----
35
36
37
38
39 >KX883566.1
40
41 -----MPLQ-----SMPSQKVNF
42 GSLS-----AVAASLRRLAREGRTVGSSQ
43 -----YATFARLAEKLRGLSPRQR
44 RR-----CLSTASWVAGGLSVSTD-----RS
45 KGSSLGHRLAVVK-FAIKLWNNII----SDVEQFVVDYK-----RVCNSAAYAARGY
46 -----TSIVAAGKCRDYL-HRA-----FGPCGIIPHKGRRS-----
47 ---LFALASTTRC---LP--PASLTPSRVKIELEALKSR-LTSVPPPVS---SGA-----
48 -----LSELVSFSHD-----LLRDAGPPVQVEVTGN-----ALGPPGKTSLDRA
49 QLYG-----SYAFKGVIMSKTLLMMSKSAIPPDVPFKE-----
50
51
52 -----DLEENVQ-----
53
54
55 -----RVSL--LAEAG-DKVRAITVPSSLAVQLAGRSIN-----RALIR-CL
56
57
58
59
60

RR-----SDVFGPGF-GKRPG-----TLMSGANLGDK-----
----HI-VSC-----DLTAASDYLNQ--DVAIAVLADYLRs-HA-----
--SGI--WYPWASAL-LG---PQRI--R-S-----
-RDDT--YT-----TSRGVLMGTPLAWPILSICHAFA----C---
YKA--L-G-----
-----NTYQFLC--RIVGDD--GLLI--CNASQY--RAYVSVMNr-LGFVINEHKTF
QG--KQC-GILAGRL----Y-----V--LKTRRA--KFQ--
-----DLN-----
SPTKTMAW-----
-----RMVDCQ-----APYVPELLAARGAQRFRSDAP-----TAP
IRVPGLS-----QRENKA----ALRAFK-----ILNPRTFTDA
M-LRA-GRP---TPLTT-----V-----QRRGLLV-----
----LKNAAPSSGAR-----SLWKSSL--AKRGHYASLMT-SY-
----LRQLMD-----INLPACSKAPVNGS-----YLARMSGTVSQ-----
-----L-VTARLNLPKDS-RMIL
RPSYV-----LSRLGKRLDKK-----MH
TPKQSSVLRLRLRGLEAEMV---PPSHVSVMEKLLR-----
-----SMVGPDDGARPISGVIGDFSQ LGI-----TS
PTELYPQF-----
-----LHWIE--EGAP-----

>GGIC01436417.1

-----KNQRIS
NKKK-----IKKNNISKSKIYKKIVSPEV
-----KNYIDKLGLKTYKIED
ME----KIKD-----KEVEVLKDYVDEITHQHIELKDER-----KR
SRNECVKDIRLFV-YLVRsILA-----SGPNQLIKGLK-----EKLTsVRRYCMGDI
EEQEIVDKHFSQVLARDQEKRHLSKKERQQR-RLDVRSDRQKILGLYPQWLELPGIRKFT
RHQHIQMTFISRS---LP--EGT--RNVKDEAIEGHKKM-VSSGGKMD----EE-----
----IRRDLVRWVRE--YCRG--KVSDIQLLGLAGKW-----SATPS--SCLEAT
RKEG-GHLMGSYQLYNMIS--PCRHTLFNQSRSEDIERMNWERHGGDRFCsVSVSLVATSL
MIHESEPLKEFVTKYLHEAKQIPSLPNVIQERHNLKIAPGFSFRSKTDCKTFVKQQKRRL
GIKD-----KVVDEKLGRMVDNFD
DRRI-----LSPFLPSGDQSEVHSTLGGILSTEEFSDYAKS
CFDEGGLRSRVSA--IAERG-WKSRVVTNHSELEIQS-GNIIR-----NLLNS-IL
SKDKRF-----LGSRRTHFD-DQEV L-----DAMGRNSVGG-----HP-----
----LV-LSE-----DMTAATDSIYM--DVFATIWN TLCST-IG---L-
-KFSS--IHIAGLIM-CA---PRLY--E-Y-KDL-----
-----TKVRTNRGIPMGTPISWPLLNLANMFC----W---
ERA-----NIDRAFEII-----
-----PPV--TILGDD--LLGV--ADLTYI--NKYSDNLKR-IGGISNPRKRV
LS--VFY-GLIGERI----Y-----Q---FDVEDN--EKSLV
IKHKETQLEAEQTNYQVHAEIAGKSSIKTLLLSEDGKRKDLVELGSSDYRISSSQARYTE
SLKMMLED-----QDCHTGRRSKIVYELEIVKYIESDLIKEDRKDIEEK
LRFVVKKD TEDRQKEIENQIINICKIYKYCYGKGRSFSGII---GPFENY-----KGI

VEWLFGS--IVIMIRGVT-----YSISTR-----RKDCVTSKDGIIRSRRPLWVR-Q
T-NGKLEKKAYEEGPVNLDPGFDG----VELHLNQYQKKFKTLLRQWHS-----
----PRRNRGLVISK----RLKCTSLKGIVSGTNKLGATDINKGIKPRWTCMGSIEH-
----FVLESL----GDSNEVRQILNRISI-----FHPDIIRELRGHRLEPFI
PESMGGGGLVYKNKLDSTRWPCSSVFKKYINKILFGQNVKDYL-
EKINTSLVISEKQDWW
ELASEMGRADFEDVKIVERLLHPSIEGITIGNHQD---LVTENILQYQGDFIRQMGPDP
FCKKVSKLVRGKTKQAMKK---EYESWASMRGFKGSIEEFIRFRDIYLSRFMAITIRQEE
LSTY-----LRRTKGGFPLIVPKEHHLNY-----ITTKTNTSQTQDQ
PRGQAETV-----LNADVTVVITDGLLETIV
PSE-----PQRRQ--LIAAAGRSMHLTHILPGLYFYTNPAVWFKMLESTP---

>GCIB01004342.1

-----THVNFFRVS
HDMV-----IVSKLMRHGAACAAGRSIAK

-----ALNTASKIMT-----RE
FCPSYELFAPLKV-----RTCSDVGDWAG-----EWSKECLRSL----
-----PRRKRDRFRMAVKGTKR-----IFDVPCRRCD EASK-----
---LLAIENWDRI---TS--VKT--SPVDETRLELLRKH-----
----VRELSSGW-----WRKIESCRKDTCNPVLPD-----SYVPDRQGCLETK
SIWG-GTLA-----

-----TRDDECLGDHSLVR
RGT-----
-----VKTKG--KFRVVTMQSARVKRV-LRPLH-----NALYD-SL
SKDAWL-----VRGDVTRDDF-RAVAE-----TMGKNE-----
----YL-ISG-----DYEQATDNIYQ--DAVKAVVDVLSEG-EM-----
-L-EE--ERECLLGS-FH---NLRW--I-S-C-----
-DFGA--KQ-----HEIVRGSMMGNLMSFPILCLLNKAC-----H---
DIA-----SDLVYGEGVR-----
-----RVG--RFNGDD--CLFV--GGDSMY--KKWREVT SW-FGLVVNEKKTG
RS--RKT-CELNSQI-----YRVSTN-----

NLIPKIVL-----
-----NFLRPL-----SNEPGECLSAVL--TGISSCR-----PAV
QQWVVNV-----VMRYEI-----SLKGFA-----LSNLSRYW
I-RVLK KKKWF---RTLVDWGA-----
-----TTHTSIHARADKPLARGHLVLE--RDL PQVVGPPP-HP-
----SRVEEA-----TIACAE-----VSRLWTAKWQG-----
-----VSGLVVKPK-----
-----VDRPTFRARYE-----
-----KKRDIPLPKTKFI-----GV
IVRW-----GWLWPSEILGKFSTDNVD-----YG
LLDTYPSS-----

1
2
3 -----PYLMLISTYR-----
4 -----
5 >NC_028476.1
6 -----
7 -----
8 -----
9 -----MPTSNDMEAGCKATRSIRES
10 -----
11 -----LLRATQLI-----SR
12 EFNLPVPVGPFG-----TSCMELRSQWD-----TYAKWSCLQFV---
13 -----SKKEERRKT-LVQ-----LIKSGCRLFDGDCR-----
14 -----TCD--PKLKELEKRKWVEQ-MSTARPRVAYYKGLTRST--
15 -----WLGRLKDSARE---LITG--WGRNLEGCRMEDY-----IPDQS--GCFEAK
16 SIEG-GTLS---VSQ-----
17 -----
18 -----
19 -----DFASF-----
20 -----
21 -----PNEVRLGT--AKTKG--KIRVVTMQGANTKKV-LRPVH-----SALYD-YL
22 AGFGWL-----VRGDVTAADFEAIE-----DKQGDE-----
23 -----KF-ISG-----DYEQATNHINI--DSVQAIISVIAEE-----
24 -PLLSDEEREVLIRS-FR---DVTV---Y-----
25 KNGS-----FLCKVRNGSMMGNLVSFPLLCILNKCC----Y---
26 DMS-----REIESEENGVPYCP-----
27 -----RVG--RFNGDD--CAFC--GTDRFF--EIWRETTSI-FGLVVQEKKTG
28 IS--SRW-IELNSES-----FDSLKH--RFVQK
29 -----
30 NFLSYLRV-----
31 -----
32 -----TRDTPGDLLSEIV--SGTKGFK-----GST
33 KMWLINH--VLRYEIAIR-----GVCAST-----IPRKLF-----LLLIKRVW
34 F-RKTLNPL---PPFPT--TGVD-----RSIRQ-----
35 -----TVVSPPLPSFI---PLIDDIETLVRKSHVKKWVGVSIDGRRARCSNEM-NP-
36 -----DHDHFR-----PMVHGSLFFPFISSE-----YSPTPLSSSLR-----
37 -----GSLLRQRKVAFDLSS
38 KPSA-----SVRFSRKHSWG-----
39 ---FTMSELCKDTLTTF--GETWLMPPKSLCG-----ST
40 LPYY-----HPLKIKCDFLTTPVTPYPPVPS-----LLTTKTAEHAA
41 TIYNLERT-----SSSILRNSRVRIA
42 KADS-----KLWSF--VPDNYIGSRLVV-----
43 -----
44 >LC150604.1
45 -----
46 -----
47 -----MASQFCDSIAR-----RCGEVIRIT
48 LHR-----NDHTTSPPHGGPRLEGNET
49 -----FPPAKRDKH
50 AK---RVMQ-----AFYAAYIAMG-----YH
51 CHLRDRSLTRFFEKLYIRAWH-----EGLDRVCPDIK-----KQVTLLRSALI---
52 -----AETGVSPDVLGKL-LPR-----SWANLAHPRK-----
53
54
55
56
57
58
59
60

---LFQLTCLGRA---LP--RPT--QKVISSKREALDG-FTSPASVT---TDIN-----
----VLHELVPMALNLRYSADTTIFRGSSSTEQSTLGP-----VFGTS--ATLFNS
RKEG-GRAG---IIH-----DACRSRPRQVGAIVALKRQA AIRKTAAPL-----

-----RFDIRSLKSEPLIG
EDSR-----RVLNDVVRH
RFPNNLGESEIAV--VPGLG-FKARVVTKSHPTVVA-GHQLR-----RRLYP-TL
YQSRVF-----SRALGDKP-KALRF-----ARRSG-----
----AVFFSA-----DLSRATDGLSH-----HTVGVFCRN-AE-----
-IDPD---IIFRNM-----S-----
-VDGH-----ALKRGIFMGLPMSWTILSYIHRAV----C---
DSV-----DPIQNY-----
-----YLKGDD--LVAH--WTARQI--SLYRERVAS-VGMPLNESKTF
VG--PRK-AVFCEGY---Y-----E-----

SSKASFNK-----
-----KRM RTE-----IWLVRQPSISLRRFFPDS-----EAA
ALRLSEG--LDEV-----DYNLRS-----QYNRLV-----YCFYPNYIR--
I-SRKIRLNPF--LPPQL--GGSG--L-----LIPDLNS-----
-----QVHSVFD R---MRLNSIFGGV---KRFNASIPVGGSSSNMKQVN---

-----IL-LSRVRMVPRGS-----
-----GLTCSHFDKWL-----EFA
TSAAAFRDAKIGIFPKDPS---DHTFFAEAKRISR-----DRRR
GVDF-----PIDYRTVRSVCRSLEPLLSTVPLR-----
---HTCE-----

>NC_033257.1

-----M
VR---RAYD-----LKRPLPYSQVEAAKPC-----A
HQ--VKTWKEFTN-FLLDLVA-----HGKSALVRRPG-----STPFTRSFTSF---
-----IVSRCAWSRGRGK-WAS-----
--IAFSLGQLKRC---FP--ALP--DEMRRREALDKHQEL-LGSA-AEP---FVEF---PI
----VDRIIQQYFPV--NWARD--TSVKVYN-----AIPNT--ACIE--
---G-GSCR---DYL-----ASTGPKPA-----
-----RGC-----T
PIPS-----SGQ---ILAKRRY--ESIFD
KLYM-----QRLQP-
----LPTLKAVA--VPDAG--KFRIITIDSVSAKCL--QPVQ-----RALLS-HL
SRFPEFEFC---TSTFQGELPPKWGK-----MAKGEKL-----
-----LSV-----DYSNATDGVGDG--SFMSELMDRIIDRS GS-----
-FELLALREQAVQETTLG----RRI-----

1
2
3 -SNDL-----CSFWQRRGTLMGSLVSFPLLCLWNASI----I---
4 REA---G-M-----
5 -----KKF--IVCGDD--ALAY--ATNTQK--KRWADYSAK-LGLVQSPGKTF
6 HS--RKF-GTFCSKV-----VGTEHL-----
7 -----NAG-----
8 RILAPMSG-----
9 -----ENYNRT-----
10 -----PKQFRH----IYKRYC-----F
11 Y-PHGEFRSLY--GPREI--GGLG-----GEVD-----
12 ----DIDVPS-RRAL----RKAQAHLAR----QGTKYQL-----GMLDMQD-
13 ----ENHAGI-----YTTAWSACDPT-----
14 -----
15 -----PQRLTPRLRKG-----
16 -----GSLDYRTKPIR-----
17 ---W-----FTR-----WD
18 RKSRSRRF-----
19 -----
20 -----
21 -----
22 -----
23 -----
24 >GBSG01013692.1
25 -----
26 -----RMVRHTNKRLT-----SFVKSPPCG
27 AG-----SDSRKNDRTLHRGRRKSVEQ
28 -----
29 -----QVWT-----ACYSSLVHS-----GW
30 DNTNMAWF--LHD-WVSKTLYS-----RGVRFTVGELK-----KLSHGVRFASL---
31 -----RATGTVPDAPRI-PKK-----VYSAFLYLSVREKR--N--
32 ---AFAFSRLSRG--LP--CPP--KEDEIKALRDAKEMSVSYHPTSG-----
33 ----KAIEAIMWH----ILQE--RGGKTRLPD-----RLPNTLSSCFERS
34 TTQG-GVDG---HLR-----HLGFGELINIQCQYMGSCSSADMIRAKYSSFATDTLGQF
35 CLKSI-----C
36 DVAD-----NFDSTKD--DEALR
37 CLGV-----LKLRSDRP-
38 ----DPRCKSHC--LRTPG-YKCRVIGVPDALTFVE-GTWVR-----WS--S
39 QLLPRK-----HFDPSKNT-CPAVK-----GRTASDK-----
40 ----RY-CSV-----DLSKATDGLSH--DAVRAVIHALAVS-GR---IR-
41 -KEDL--SMSLRSLG-VG---DTEM--I-W-SWRIPKE-----
42 -PVEE--LR-----CVRGSPMGTPLSFIVLSWINSCA----T---
43 ESF-----
44 -----THS--SIHGDD--AVGT-MYNTAEL--EEYKEFVRD-IGATVNVSKTY
45 IS--PTS-FTMCERM----Y-----APEKTK-----
46 -----
47 GMFTAFCP-----
48 -----PS--CPIPGGV-----VPV
49 PPQAHIP-----NLYLRR-----AERVQR-----TIHPWVSKLP
50 I-LR-----LPTSV--GGFG---Y-----TGYGL-----
51 ----KVSKA-----TRCRLATACS-----RNIADVT--KEILKQSTFRG-EG-
52
53
54
55
56
57
58
59
60

```
1 -----LFPRSA-----VPTPKNSSKYYKYR-----NLIHKDKRFST-----
2 -----VG-SEEFEDVVLAS-DFVT
3 -----FREKVILQSYL-----AIGGKLGRGSS
4 ERKPERTRARALFKGKAPC---NIAPLTVKRGIGA-----
5 -----LSKLSDRVKNQKVRVRTDVACVIR-----
6 -GRTHSK-----
7 -----
8 -----
9 -----
10 -----
11 -----
12 >GBSG01012161.1
13 -----
14 -----EYFKIPDFNDALSR-----GRATVLQFA
15 HSRT-----YPNSVPIDVPSVKHGSALSI
16 -----WVKPIARKDIHR
17 AI---SVCQ-----AVVSGFVVVYPKFACLHLKR-----QA
18 DVQACVPLVRILR-FILNLYSR-----GAEYKVSEYLYK-----GMCNWLRYMACPSN
19 LRI-----SPPPPRMPAGPFT-VLGKLDTSIFRGQLCSMKISTR--V--
20 ---LLCMSRMGRA--LP--PGG--KRVMAETIQKHKDI-VCDKDIRS--PPIS-----
21 ----LVNAYGRFCKN---LRG--RRSLLTEPRVSFPL-----SSGVDGVLQTHIP
22 RRVG-GAEY-----VVLGALGCSSRGSGFSNAKELERNLRPLTML-----
23 -----
24 -----TLNEFERTLSDIHK
25 GEVI-----RR
26 LDQPVTRIARMVA--IPDRGGFKCRGITVSDSVIQQH-CTPLR-----RALYRRVL
27 SRLEC-----SRVLSGDYSHSTYS-----SRVASRRVGGRRSTP-----
28 ----YVLLSA-----DLTAATDYFPF--EFQEVLSNSNMFYH-----
29 -KEKSDTQVKAWMWL-SR---PHTI--M-Y-----PKV
30 DEDGF-----PVQGVTPLVQTRGTLMGTVPSWFHLNCLNSFMIR---A---
31 SLA---C-LSE---LVKRGRRNILLTESAYDADREMLARELQ--SILPTITY-----
32 -FPRSE-----S--VICGDD--LQAV--LPIGAS--LIYEILIGM-YGGRLSVGKHY
33 V-----F-----R--LPREPA--AG---
34 -----PQV-----
35 PFTVFAET-----
36 -----FRTY-----SSSRAKNVSAP-----PL--RSVTTLM-----S
37 ADWKDIG---DRFRDIEL-S----CSSPAQ-----RRSLLSLAFLYNKKTLTNIRR-T
38 L-PGAPALPLF--VPRTW--GGMG---A---PD-----V-KGDPF-----
39 ----RTGVNL-PQLK----RMIVNL-----RSVINDP--VRFIFEVLKIR-WS-
40 ----WYAGNK-----
41 -----
42 -----SRVGLWV-----
43 -----NRIHD-----
44 -----
45 -----E-----
46 -----
47 -----
48 -----
49 -----
50 -----
51 -----
52 -----
53 -----
54 >GBPT01016202.1
55 -----
56 -----
57 -----
58 -----
59 -----
60 -----
```

-----GVMSKKFLEFLIPYVNSDALSRRGDF-----LCFHWQRPT
APHK-----ARGLLATGASLFVKPERPAE
-----IEEFPFLRCALIPDFPNSVV
RD----RCAK-----TINGLLVTLVALWPTCQFGLPYT-----LS
SAIALGPWIELVA-WTMNAFSQ-----NGEKFVAERLK-----GMAAWARYLASNNQ
LYPPPKP-----LTRFPLAVKGRFS-PSK-----LFKGAIASKATSLR--S--
---LLRLSRVSRA---LP--PAG--TEIQAEALASHLDL-CQDEPHVP----PA-----
----IAGLFTKAAAQ--ICAG--RLEGIEDARI-----ALSNS--ACFENT
RING-GAMS---YLR-----SMLWEYHPHAFYSRSSDYGDVIRQSRPNEGYPDEVVETR
A-----V
FHKT-----CVHLAFKSTYG--QPRVY
QDTV-----PGDLLGFDE
EFPAPNHRVVA--IPDRGGFKARVVTAGPAALQAL-AHNVR-----RVVYAQVL
AKIPTR-----HALREGGV-ELFFA-----DIRSQRFFAKHWEPYFGP-----
----TVLLSS-----DLTAATDRFPH--SLTLATNMGMESC-LSEDQRI-
-SANW---CAWRSL-SG---PQLL--V-Y-----PSY
ETKGENNEI-----VDFSVSXSCGNLMGTAPSWFLLNMFNLTFE---A---
AWS---I-WLN---PTFRSGFTGRQL-FSLTPREELLPIWEELKKDVLAEAAARDSFEFK
RIPREYGGRLRLDSY--ILVGDD--LGAV--CPYAVA--LLYEILIEA-CYGELSQGKHF
VQ-----PYREGC----F-----L---LVAEEF--AT---
-----VRS-----
GVLVHERI-----
-----EHLRGI-----ASTQTSFDPR-----DKK
CLWAQIG--STLASAVAH-----IDDLGR-----ANSLLS-----MASMSLGTWRAT
L-LKK-GLPIY--IPEAL--GGLG---W--PH-----P-KGLEA-----
----GINRTA-IRCI----EHYHLL-----RGFRSDP--VGFSIEIAKTR-AC-
----WHTSIE-----VNHGLPLDLMRGH-----LNGIPVARDRD-----
-----GRL-EEQDEERPIGLLSIDS
PSSMRFSDFLSLLCMEGICFRYL--VREPLDHAVRK-----AGLSMLPIREESALVSGA
PSPLLNMQVDRDARESFSI--PLASSRYRRRLRELSLVNAGYDVQLLSPRTGYVEM-MT
QDMW-----LLEALHILPCAVIADLFPHVCH-----LC
HVTGQFVF-----
-----DAAIG--GRHA-----

>GECQ01195387.1

-----TMTSIK-----ATLSLPIVD
SVGL-----SRGGS---YLGFY-----NQRPATSQPGTSVLVLT PAS
-----FVVTRPMPGYPHLRIPKIGNGCV
RR----RLAT-----ILNGIGAAFVALWPSAQYGLQPN-----WA
NKEAMRAIVSLYK-WTISSFLG-----SGEAYVAVRLK-----GLCAWARYLAAEKQ
LYPPPAPL-----ERFPFADMEGRFQ-PSL-----LFRGSIAAMAESRS--C--
---LLRLSRMSRA---LP--PGD--SDASLAAIADHKAL-SLSQPTVA----ME-----
----LLPSFRHVAAR---VSTG--RMAGIENPMI-----SLSSS--ACYQAS
RKQG-GSAN---YLR-----QRLWNFFPDTWCPMDERRPAVAPSTFGRTVQAITGMPW
E-----V

PLVQ-----ERFHTTCEWDSADG--SDLPL
KRAT-----QPWASGHS
WDNPRHPHTHIVSA--VPDRGGFKTRVVTAGPAGLNTM-AQAVR-----TIVYSRVL
MSTPTK-----WGLINDGV-LKWFK-----QLPDLHYVDPRELGD-----
----WVLLSS-----DLKSATDRFPF--SVVEAINDGFEEESLTGEQRQ--
-CPNW---IAWRSL-SG---PQLL--H-YGPRVSHIHPPFD-----
-DDGV-----RVCGSCITTSCGNLMGTVPSTWHLNIYNLTAIR---L---
AWS---L-WTW---PWFRRRAPYL--ERISDRSDLEPKMQNLRELILEALRRPACHPF
RFPSKW---KMNELT--ALVGDD--LGAL--CPWAVA--IVYECIIHM-TNGRLSVGKHY
VQPYREGSYLLIAEE-----FAILRS-----

GRMTHLHL-----
-----DHLRGI-----AQTVLSLTHD-----ANV
PAWASAG-QLLQNSLSVPT-----PEEPVA-----RYRALV---SFAHVS LASLRQR
L-MRA-GLPVY--LPTYV--GGLG---W---PH-----P-RGVTA-----
----GVERTS-----RRALLHYHCL----RGFRDDP--VRF AVEISKTA-RM-
-----WASVTE-----RNHFAPIVDALQSY-----LRRIPVSRDIH-----
-----GQL-YAATSECPIGLHNPEL
EHKQIR-----LTELLNALAVQ-----GIGVRYLVTEPFE
EWLANNPVTSGELAAMAQQ--SLSELPAIGQVLE-----FDKREPTISSAARA
YHRR-----VDRLLSLRTGGGLQLLPRTSYGDL-----FE
QDREYLD S-----LFVAVTPELEE QYPFLF
K-----PRGGLVL SHNPLALSSAVEMDSPSLPNDE-----

>GAMU01000156.1

-----RLADV GSP-

-----NFSLSFSSCLERSRKD-----
-----GGAGHYLREVL-----
-----WTPMIVG-----
-----TP-----
----LWSSLRMFTDR-----YLPPP--GSLT-P
GMEGIGHDV-----IALRHTMYDHLATEVYYHHW-----

-----EKGVS
QPRR-----
----RLDHVVHC--IPDRGGFKVRVVTAGPAVHQSL-AHAVR-----KVLRYRTVL
PRLPTR-----HALRSGGL-VRWFR-----GLVLESSDITIRLGD-----
----WVALSS-----DLKSATDLFPF--YLVEAINDGLADS-LSTTQRA-
-SPGW---VSWRSL-SG---PQTL---WTSEVRKTSLVKGGVP-----
-PEDN-----SDWGHCCCLVDGTSSGNLMGTAPSWVLLNIFNYSLFR---L---
AWA---V-WST---PLYRRRLWGLLGRAIPREEMKTRGTWDIVKQKALRYIGDPGFYPF
RYDGPQ---RLNRLT--VLIGDD--LGAL--CPFAVA--VLYEIFIEC-AGGRTSPGKHY
V-----MPWEDGS-----Y-----I---LIAEEV--AL---

1
2
3 -----IRD-----
4 GRANFLQV-----
5 -----EALRGL-----AGGAIVA--DTRSRK
6 VPWSQIG--ENISTTLRTAR----PSVSTA-----LCSYAH-----MQLSHWR-K
7 L-LQQNGLPIY--LPTFV--GGLG--W--PH-----P-RGLDY-----
8 ----GIQRTS-VLAL----RAYQVL-----RGYRQKP--LRFLTEIAKIR-ST-
9 -----WLVGQS-----CADFSRLYDLILSK-----LEVLQESRDDH-----
10 -----GMLAP-LDDLHPIGLPSPGIKS
11 IP-----VQKFLDDIALE-----GASCQYLCS
12 PCREAPPPATPRIIPVEPE---AWGRWLPTVRLRPWDLIELPEPPKVIKKTVRS---AS
13 KAYY-----ARARALVSLRAGRIRWLQNRKDIPSF-----LRE-----DE
14 ELLGSYHI-----TVNDLVREIFPHLS
15 EV-----SDWDE--PTPWPY-----
16 -----
17
18
19 >GAMC01013668.1
20 -----
21 -----KVMTTNIRE-----SVISTPVFN
22 SEGL-----RGGG--IVRFF-----IERSTTYPsrcSDKpvcPGS
23 -----YVGpyrgNLpsfKvppHHMRMSE
24 IM----RLAQ-----ILNAQVAafVLVWPACQYGLTAN-----YA
25 NVQAILPLVNLIR-WTCSSYIH-----NGEQFVAPRLK-----LFCAWARYKSGPQL
26 HHPPEGI-----SRFPFCdVNGVPD-FGK-----VFIGSIADKSSRPK--T--
27 ---LFRFSRMGRA--FP--PGD--INVQRQALADHREL-MTSEFSTP-----QV-----
28 ----VLNAFYKVIKT--VTKD--CLEGIENPGI-----SLSNS--SCWERE
29 RRRG-GAAE--EIT-----STLLEYyPLVPYLSRLEGSTFESYKEACQGLGYDETFIG
30 MSTGP-----E
31 GLPH-----EVSDVTGLAFDAIV
32 QKEF-----SGNVsRR-
33 ----LATHKVVP--VPDRGGfKVRVITAGQARIQCL-AQNVR-----KVIYRLVL
34 PKTPSK-----WSLVENGv-LHFLR-----QLRRPAGPQADTLGA-----
35 ----WVAMSC-----DMKSATDRFPH--DVVETINDAIESN-LSPpQAE-
36 -CANW---VSWRML-SG---PQLL--R-Y-----
37 -PGES-----AENQ---LTrTTCGNLMGTAPSWVHLNLFNLTLIR---T---
38 SWS---L-WAS---ERWRRQLMRFFKESSWNTGDiasVDKTGLRDkILiLADKRfNPF
39 VFPKIW--KFNELT--CIVGDD--LGAS--CPFAVA--VIYEiLEI-CNGRTSfGKHY
40 VQ-----LWQDGA-----F-----V---LLAEDW--GI---
41 -----VKE-----
42 QSLEVQPC-----
43 -----ESLRGI-----VAATHQYDSR-----QAM
44 PAWHQIG--TTLSASV-----QRVRPA-----ARRAVC---SLGHTALYDWRS-Y
45 L-LRV-GLPVY--LPTSV--GGLG--W--PH-----P-RGLEy-----
46 ----ALDRLS-----IKTKLAYQAL----RGFRQEP--KNFVFQILKLR-SS-
47 ----WLSSER-----NSAYTRLLSTLKSy-----FNGfSVAKTAE-----
48 -----GQLAPLD-ETNHiIGLfPT-RWAI
49 SSQESR-----CRRLTDVIDKL-----VLN
50 GLPVGYMCENISLTFDKPY---REASLAakRFSRE-----
51
52
53
54
55
56
57
58
59
60

-----RDRLVRLRQGHMKYVKKGSELAQD-----IRS-----DE
ELLDSLVI-----FIGDEFENTFPNLS
S-----VVLDDQ--PTLD-----

>GGBS01024784.1

-----INRE-----QILAVPVFN
SESL-----RGGS--FVRFF-----LERSATYPSRCSDRPVCPEA
-----FLGRYRGSLPHFIVPTHKLSKQK
LA----RLSQ-----ILNAQVAAFVLVWPACQYGLTVN-----YP
NAQAVQPIVELVR-WTCSSYIH-----NGEQYVASRLK-----LLAAWARYKAGPQL
HAPPGKL-----LRFPFCDSGVPD-FRK-----LFVGSISDKATRPM--T--
---LFRFSRMSRA---FP--PGD--KIIQEALADHKAL-MTS--SFT---TQPQ-----
----LLRAFKTVIQT--VSKD--CLSGVSEPGV-----SLSNS--SSWERA
RKRG-GAAE---EIT-----SSLLRHYPLVQYLTKEGTSFEDYEAACRGVGFDESIG
MSTGP-----E
GLPH-----EVSDILAMAHSACI
RDEF-----PDQVERR-
----LAKHNVIP--VPDRGGFKVRVITAGQARVQCL-AQNVR-----KIIYRNIL
PKTPSR-----WSLIEDGV-LQFLR-----QLRKPTGPQADAL--GA-----
----WVVLSC-----DMKSATDRFPH--DVVEAMNDSLESN-LAPGDSR-
-SANW---VCWRML-SG---PQQL--R-Y-PEM-----
-KEDD-----PLLLTTCGNLMGTAPSWVHLNLFNLTLIR---T---
AWS---L-WAS---ERWRRMLLPLVNKG--QHCDITMLDKQALKDKISVILQHDFRNP
IFPKIW---KFNEMT--CIVGDD--LGAV--CPFGVA--CIYEILEM-CNGMTSFGKHY
VQ-----PWQDGA----F-----M---LLAEDW--GI---
-----VKD-----
DYLLVQPC-----
-----ESLRGI-----VAVTHQYDSR-----QAM
PAWHQIG--TTLTASV-----NRVRKS-----ARRAVC--SLGHTALRDWRS-Y
L-LRV-GLPVY--LPSSV--GGLG--W--PH-----P-RGVEY-----
----ALDRLS-----IKTKLAYSAL-----RGFRQEP--KNFVFKVLELR-SS-
----WLSTER-----NSAYTCLLNTLKS-----FFGFSVAKTPE-----
-----GQLAPLD-GDRYIVGLYPT-RWAI
MYQESR-----CRRLIDVIDKL-----VLH
GLPVGYMCENISLTFDKPY---REASLAAKRFRK-----
-----RDELLNLRKGHLLNVKSGSELAQD-----IQE-----DE
LLLSSLYI-----YIGDQFEQTFPHLS
A-----VVLDS--VSDD-----

>GEEA01024908.1

-----RHTNMPTINREQILAVP-----VFNSESLRG
GSFVRFF-----LERSATYPSRCSDRPVCPEA
-----FLGRYRGSLPHFIVPTHKLSRQK
LA----RLSQ-----ILNAQVAAFVLVWPACQYGLTVN-----YP

1
2
3 NAQAVQPIVELVR-WTCSSYIH-----NGEQYVASRLK-----LLAAWARYKAGPQL
4 HAPPGKL-----LRFPFCSDSGVDP-FRK-----LFVGSISDKATRPM-----
5 --TLFRFSRMSRA---FP--PGD--KIIQEALADHKAL-MTSSFTTQ-----PQ-----
6 ----LLRAFKTVIKT--VSKD--CLSGVSEPGV-----SLSNS--SSWERA
7 RKRGR-GAAE---EIT-----SSLLRHYPLVQYLTKLEGTSFEDYEAACRGVGFDESIG
8 MSTGP-----E
9 GLPH-----EVSDILAMAHSACI
10 RDEF-----PD
11 QVERRLAKHNVIP--VPDRGGFKVRVITAGQARVQCL-AQNVR-----KIIYRNIL
12 PKTPSR-----WSLIEDGV-LQFLR-----QLRKPTGSQADALGA-----
13 ----WVVLSC-----DMKSATDRFP--DVVEAMNDSLESN-LAPGDTR-
14 -SANW---VCWRML-SG---PQLL--R-Y-----
15 -PEMK-----EDDP---LLTTCGNLMGTAPSWVHLNLFNLTLIR---T---
16 AWS---L-WAS---ERWRRMLPLVNKG--QHCDITALDKQALKDKISVILQHDFRNP
17 IFPKIW---KFNEMT--CIVGDD--LGAV--CPFGVA--CIYEMILEM-CNGMTSFGKHY
18 VQ-----PWQDGA----F-----M---LLAEDW-----
19
20
21
22
23 GIVKDDHL-----
24 -----LVQPCESLRGIVAVT---HQYDSR-----QAM
25 PAWHQIG-TTLTASV-----NRVRKS-----ARRAVC---SLGHTALHDWRS-Y
26 L-LRV-GLPVY--LPSSV--GGLG---W---PH-----P-RGVEY-----
27 ----ALDRLS-----IKTKLAYSAL-----RGFRQEP--KNFVFKVLELR-SS-
28 -----WLSTER-----NSAYTCLLNTLKS---FFGFSVAKTPE-----
29 -----GQLAP-LDGDKHIVGLYPTRWA
30 IMYQESR-----CRRLIDVIDKL-----VLH
31 GLPVGCMCENISLTFDKPY---REASLAARKFTRK-----RDELLNLRKGHLLN
32 VKSG-----SELAKDIQEDELSSLYIIGDQ-----FE
33 QTFPHLSA-----
34 -----VVLDS--VSDD-----
35
36
37
38 >GDAR01017830.1
39
40 -----CTPVFNGEALRAGNSVTFFQR-----RALNFPAQ
41 SGPQSALDY-----CIRTVEEVPTSHQSGLLPRC
42 -----RIPAAFPFLGNTKKEKAI
43 RN----RVAI-----CCNAAVAAWITVWPKSQWNLALN-----AA
44 NSEAMCGLRDLVA-WIVRTYSG-----AGEKFVAERLK-----AFCSWLRYLAAPNQ
45 LHTPPGPV-----TRFPFSDSEGTIV-PSK-----LFRGSIARCATTSK-----
46 --SLLRFSRMARA---LP--PGG--PWAAEEALSAHRTA-MQLVPEKV-----EP-----
47 ----LALIFTSIVSKVTPSALL--KKADFEKAFA-----TLSSS--SCFEVS
48 RGDG-GAAE---VIR-----TLVRDTPDLDSAVITDDGDVFRPTLFDDSGELRQGGTI
49 FASGGERFDYPH-----S
50 GLPD-----EHDYLNKVVHRLV
51 ESVF-----GAPRASELYSQR
52 EYDRVVPKHRVIC--VPDRGAFKVRVITAAPAAIQSL-AHNVR-----KAIYHHVL
53 PQLPTK-----WALHEGGM-VDFLK-----QLRKPELKEDLGP-----
54
55
56
57
58
59
60

-----WVALSC-----DMKTATDLFPH--YLIEAINDGIEGN-LPHSAKR-
-CANW---QAFRSL-SG---RQKL--A-Y-----
--GGS-----PEAP--EDWVENTCGNLMGTAPSWCLLNLFNFALFRTAWA---
LWA---Q-----KPIRKKFGRDLMDSFAEIPRDSQKWSFLRDLRQVLPEHSRSKR
LFWYPKEVPKFNELV--ALCGDD--LTAC--CPYGVA--CLYELLLEL-ANGRASPGKHY
VQPWVPGSLLLLAE-----F-----AIVQDN-----

GEVHHVRA-----
-----ESVRSF-----AAVTLYSDTS-----DKR
VEWAQIG--SILSTAV-----EKARPS-----VRRGLC---SLAHLVSGWRAV
L-LRS-GLPVY--LPTQL--GGLG--W---PH-----P-RGFEY-----
----ALERTS-----VRSLQAYSVV----RGFKSDP--VQFCFKLTKMH-SL-
-----WCRPGI-----ASAYSELTHTLDFL-----FRDIPVSRDQF-----
-----GAI-GPLSAEFPLGTQ---
-----VAGLRSAPLFD-----CI
DAVIKRCISFQMASTRLSL--VEGEAAEPKRVST-----AAKQ
YRRF-----VSDLLRLRTGDVLPNRR-----

>KX883540.1

-----MALT-----YELCMPVFN
GDAL-----RVGS---IVNFF-----CRRSCTYPARQSDGPKCPKD
-----FISREILKSSLPHFRVPGPACSHKL
RR----RLAQ-----TLNAFAAAAFISVWPSSQFGLCPT-----PP
NLTAMWALCKFLH-WVCKVYVG-----NGEPWLAPVLR-----GIATWARYHASTTH
LHLPPEGI-----FRTLSSADGSID-FGK-----LFTGSISRKATSPS--C--
---LLRFSRFARA---LP--PGD--DWTCAQGLVDHQKA-MRDNPTVP-----EQ-----
----LTKLFKDVATRI--SCQP--LLKDVGTPGF-----TLSMS--SCLERT
RSQG-GAAE--EVK-----EAVWEIFPGLRRIHEGVTEEDTELSRVPCLDNPLYSD
VEYYYSFY-----E
GLPE-----HVPRVER--LTYLA
CCEK-----VFPPC
RLERRVPSHRIIP--VPDRGGFKIRVITAGEALTQSL-AHQVR-----KVYRKVL
PTLPTV-----WGIREGTV-RNFIE-----QLRMPSSDPRLGR-----
----WVALSC-----DMKAATDRFPH--PLIEAINDGLESN-LSESQRR-
-CPAW---LAWRSL-SG---PQRL--L-Y-----
-PGGD-----VVLSSCGNLMGTAPSWALLNMYNYALFR---L---
AWS---I-WST---GWSRRKFPELAAREVLPSRQSLDGSWEPLKSRVLRVLSDSRFHPW
NFPKGL---KFNELA--ALVGDD--LAAA--CPFGVA--MLYELILEL-CNGKPSAGKHY
V-----MPWIDGS---Y-----L---LLAEFF--GF---
-----VKG-----
DRLVHLHA-----
-----EFLRGI-----VTGTNCVDTR-----SRN
PEWAQIG--LSLAASI-----AGCRNE-----TRRALS---SFSHLMLADLRK-T

1
2
3 L-LRF-GLPIY--LPASV--GGLG---W---PH-----P-RGLEY-----
4 ----ALERTS-AKVL----RAYSVV-----RGFRQDP--VQFILIARLR-SS-
5 ----WMRPEL-----RTPFSNLLSLLSGY-----FRSLKVS RDEQ-----
6 -----GRL-APCTPESPIGLPSLDG
7 -----TSKNFFDCIDS-----IGVH
8 GVTVAFVAGEELPLLEQKE--RFRTIGASRRAYG-----QA
9 ISDL-----LRRREGHV PPLPRRSDYAAL-----IAE-----DE
10 KLVRDCHL-----YLGESLEAQL-
11 -----PFLQG--VLGKSPEEALEDIGEAE-----
12 -----
13
14 >KX883626.1
15 -----
16 -----ME-----YILCTPVVN
17 GDGL-----RAGS--FLRFF-----IRRSRAYPARLSDGPVGLSS
18 -----FVSDVSQGQLPQIKIPPMESPAV
19 RR----RVAS-----ICNAFLAAWIEVWPKAQYGLEAN-----LA
20 NILPFRSLAHLIK-WVIRTYLG-----NGETYLRGQLK-----LFAAWLRFLAGDQL
21 HLPPSPL-----RRFPFSEDGRLD-PTL-----LFTGSISAKAGSPA-----
22 --ALLAFSRFARA--LP--AGN--KQAEVEALHEHFKL-MQS--TPE---VVPP-----
23 ----LAKHFREVASK--VGKD--SLSDIEKPGM-----TLSFS--SSLERS
24 RAEG-GAAE--EIR-----EVTWDYFPGLRRMHFDVDMQTDVYVSSFEIDPKTDDL VV
25 DGT-----PYDYGFTFG-----E
26 GLPG-----ECERVVHMAHAACI
27 YKAF-----GTEHLRGKIPLSLPRV-
28 ----IPKHRVIA--VPDRGGLKVRVITAGPACLQSL-AHNVR-----KVIYRSVL
29 SKTPTQ-----WGIIEHGP-QRFLE-----QLKLPADRDP ELGP-----
30 ----WVALSC-----DMRSATDNFPH--YLVEAINDGLEQN-LPEDIRS-
31 -SPNW---VAWRSL-SG---PQTL---FYKDILN-----
32 -EDGE-----SSEITSTCGNLMGTAPSWALLNIFNLSIFR---L---
33 AWS---L-WSSRDFRRRCRRQFGQESDSAVARRIFAHTGQ--LRELVLREISRDKFHPW
34 RYPKGY---SFNELT--CLVGDD--LAAA--CPLMVA--ALYETLLGL-AGGSVSTGKHY
35 VQPWEDGNFLLVAEE-----FGRIEG-----
36 -----
37 QGIRRLHF-----
38 -----EHLRAF-----TQISSSVDTR-----SKK
39 ETWAQLG--SALDSAV-----RSCRST-----HSRALC----SFGHIASSGLRQR
40 M-LKL-GLPVY--LPTSV--GGLG---W---PH-----P-RGLLV-----
41 ----GLSRTS-ALAL----RAYRVL-----RGFRSDP--VKFSCEIARLR-AT-
42 ----WYLSEI-----SHKYLTLLEIVRGY-----CTFPVSRDKDGKVAP-----
43 -----LTEECP-IGFPNEDFPDSVNLF D
44 FIEEVV-----ISGISAGFISGEETFSAGHISQRAHKAFRIVYTVG
45 SGTSQTSTRSRASHTIQAA---ARSFRRACRRLDLRQGNIQGAMSYS DYMNC LAEDEEL
46 LARY-----HVLKNCDRVRSVLPSFATPLQE-----DD
47 QMSPPQEH-----
48 -----LGRSCVLHV-----
49 -----
50
51
52
53
54
55
56
57
58
59
60

>KX883471.1

-----KSAPNGKDTRLKPAKPAVIKIDLSHP-----AGSRVDPKV
ADKNKPRLEFASLDSLQVSLAR-----SSLGVQRISTIPLSPDPSKS
-----AAQLFRRFAARGKQNIK
FQ----RLWS-----TVLNAVSLAFKIDQFDLRT-----PQ
GRIAYTDLQCAFR-HLLARFTT-----VGQDSVRVVA-----FWALQARQMAL---
-----NSAVATSGHRPN-LVK-----FLRGRLDIRRASQR--T--
---KFLFSGVGRA---LP--FAKIDAIRIDAEIKAHRAR-VTVPGTTD-----PL-----
----ILGKYAEWLQKKLNLLDS--RADLPSMKLNIAPL-----TAARS--SAVGAP
SSKG-GRYG--FYK-----SLFEQGPQFQERILEHA-----

-----IKEVE
DCPT-----
----PLPVKLTV--IPDKG-DKARTVSTTLPGIQVL-GEQLN-----QSMNV-LL
RRLAPS-----AIPLSEN-PVSRR-----LNNHG-----
----TM-LST-----DLAASDYIDH--TVLKVTTRLVARK-LG-----
---WQFNSKLVRL-TG---PFMF--TID-----
-QEGV-----KDSYLGLRGAPMGLPLVWPTLSLMNMFC----A---
EYR-----ISPNSK-----
-----RSY--AVLGDD--HVAR--WPKADI--DQYFRNMSD-IGLKINKAKTF
LS--KRD-AVFAEDL----Y-----S---LVNGKV--KR---

TVVPKLSA-----
-----VMSTRQQ-----GTD
PQIFAI--ASLQSALA-----NAKNFK-----ERKQIK--KIFFSQWKKLTR-Q
I-RRR-QVPLQ--LPTQL--GGLGIAPF-----
----KLSKLT-RKQR-----SRLAYLLSL-----PEYASKS--ARSDQRSYFRSLDQ-
-----FKNANR-----FAHAIRNKDKLVKE-----AARQVIEQISD-----
-----EEERT-SRSFNATGCAVEDCAP
QLPE-----GGVPAEAVLKE-----LLALKLTRLTQ
EGKVQATDRSVSSGTVLRR---LTRVWERLPLCLK-----PV
GARG-----YTKLLSLCVQSSQILSTQSVAKLI-----TT
NLPLYAYV-----GVFYKDGGMPIQC
S-----PNSND--AYDSATAG-----

>NC_033474.1

-----MTVTSVPVVAQVRIPTLRLLA-----AVLQRSPVL
RKGS-----LRESLLRLRSLSLKAEERLD
-----PIERKF
RE----ELLR-----WVDAALYMSLPMYNRTISGS-----HQ
VEQYLLALWKLH-YVYK-----NSIPDVVKTLK-----DLFFRYLDAGLQSD
REF-----RRRRFTYLGRHP-LKG-----FFAIAVKGLTRES-----
---ARTIGLAGRC---IPSLPLS--KERRLEEYLNKER-LTTPMPTI----PACRGCG
HLPPEHPDKRLAAGPGCQRLSRDH-IVESVIKFAEVFARPVKHLPPMSYPADNMVYIPGP

LHHG-ETRP---WLM-----GVTKPAVEAISDAFDSFDEWKEAVNSPKLPYQLGIWSVH
PEA-----I
ENPP-----ERDPAVGLTRVISEGESLKT
PKRR-----ANVWYDRKDLRVWYSERHAPSV
LHPKDPEAERRDDHRILPPG-YGLRALKRSPIAPILN-DVVLSGEDQCLVPVHKVLF-IP
EAGAKV-----RVATVPSV-VSLQTAARQINGIVLRALKHCPTFHLGLKNS-----
----KC-FSSKTFAMPDGAKFYSDLSAASDRLSQ-PLSLSVLWPFIREA-EK---AG-
-LE-N--AFEVARLS-CG---PQRI--I-L-----
-EDPE--VMELLADPKYSAHLPPFE---TEAGLLMGTPLAWPALSLLHLFAALEAGI---
DIH-----
-----RVF--IVGDDGGLIAT---APEY--QKYYHIMTNILGLRVNREKTH
ES--PMM-GVVAQKV-----LKIRPA--PCP-H
AGTPRGKLGIGPL-----
TVWDFVRH-----
-ELNLPNESPVYWGK-----VELDQSHCYVPAA---LT--ARSRRRR-----DPH
APWNPPE--WSDVARWK-----RWRER-----LPKALR--L---VYPSLVS--
Q-SRYRGLRME--APAIL--GGCG-----PL-----TEGAVRK-----
----GLTPV-----QRAGAYAIAR--EIMHGTRKDL--SIDVGPDLTRT-IP-
----RAVILA-----ANEVFNELPRSETR-----GTGLPKEAA-----
-----LRSL-TSFFVRVAWEG-GFRD
ANSYL-----GRSLGHRVVRH-----P
GRLLEAKGRIFLRSRGKRA--VPKPLTVVQAFKR-----LPKPVDVSPGVLKG
LNRL-----IHDLDLGGRQKRRLDYVLSAGR-----DDA-----EE
GTSSSSVQ-----VP-----EAMSSLLNRLR
HGWG-----KVWGQ--IQSLGSGGKGNLTSTHHEA-----

>GFLP01304447.1

-----FLALGLSSG
LYLTS-----LDSKHFSTLAFELDNRKRSR

-----TFHDTIGDFRPR-----
----KLFACLIT-YQM-----QNPEILNEE-----KLRAVMPEVSI---
-----STNRSRSSKFPVT-LGS-----
-----LP--HPS--ETVVEETTRKYLEV-VTGPKSSI-DSPPLD-----
----LLTSIEEFGRK---YVVD--RLHRVKDKVS-----WVPNLNTASLHVT
AREG-GRMT-----ELLLGADTVLKLLEYTDPSEPNA-----

-----DLEEAIV--QTSIE
IAKL-----
KALDGNFEAEVMA--VGEAG-YKARILCKYSAHILLV-GDIIR-----RQLWP-LV
SEEPWFDDEKDYSFSEKLERAF-HDLPF-----WTGYHEP-----
----EV-VST-----DLSSATDFIPF--EVARALWSGILSG-LGVG---
-FDSW--IFQHVITW-AG---PTKV-----
-KVGD-----SVHTTNRGIPMGTPLSFMTLCLLHHYA----V---
KST-----EKETGME-----

-----GPY--VIRGDD--FIGV--LPNKEV---YFRKMEE-LGFVVNHSKTF
 IS--KVP-GIPVRPS----L-----R---SDTSNE--DTV--
 -----RRS-----
 SACGVFAE-----
 -----TVVQFNFREHPV----TSGK-----RKP
 FTWQPIR-----LSDVTV-----
 -----IPDV-----PFAGLVR-----
 ---EEKSSI-----KRSQLVRYG-----EWYRQIF---RRPDGTLLDRGR-
 ---WVKKAS-----RVVFTAIRRLSKDL-----IMRAVESGVPI-----

 -----HTPLEIGGAGI-----
 -----PD---ARGRLTLPRDFVT-----RS
 KITY-----LLNHTGNSRVRSARAT-----LR
 SAAAEELGG-----
 -----KIWSS--FDKV-----

 >MH213236.1

 -----MTASLRVYRWTLSTDITAWP-----VTPGVGGVG
 S-----WISVIPGPATPCVEHMTDRR

 -----AIWT-----AVAITLTAVNPKGYRN-----
 ---IIPPKLLAK-WVLR-----APVQTIVKGLK-----SLAHEYRSKAV---
 -----LGGPLHFVPEVPLVFQK-----WLKGHVCHATQ-----
 ---YAQLGRLGRL---MP--KPW--QKITSKAVHEWIRV-NTRRPDGPPIHEPSVDNER--
 ---MSNDIYRFAHE---FATQ--RKDRLTTEVNFT-----PSESA--STFMTS
 MSEG--GRLK---EMMN-----DARDLIGEFADFETIVNLAAREGGYKPPPPRAQ-----
 -----C
 GMQP-----VLDELILENQIQMA
 DSKE-----
 -----FSPRVIG--VPEYG-WKTRVLTLPNYALTP-GDLIR-----QQLWP-LV
 SEEWL-----DADRVPSW-NLFNN-----MLSRSLDRVGTW-----
 -----VSC-----DLSNATDYVPH--LYCKALWAGIYDA-LGVS---
 -EGSW--VRRYTEKM-FS---PMVI-----
 -AGTK-----GRITFRGTQMGTPLSFMTLCLLHRFAVVRSGF---
 EYC-----
 -----PH--LIRGDD--LIGCFHKP-----RIYMGMLEE-LGFKINKAKTF
 LS--TIG-GTFAERT-----FRAERKLIKVRH
 GYFFSH-----
 RKVTAGQW-----
 -----AYTRAT-----LLQDIPLRGLIKVA-TTGPNRYR-----SKL
 LMLSEIV---ENVAMGFR-----PKRRAL-----IRRLAT-----QGRGDILIA
 A-LKS-GIPLS--LPREV--GGAG---L-----
 -----PTVSEILRD-----RIYIDLE-----TRRRLG---
 -----WMISHV-----PKGPQLVAK-----LRRRASGGLIG-----
 -----FT-ERYLEGIKNISLGLRV

P-----VTEPERTAHYL-----A
AKRRTIHLKPWEVVHRPPC---SFREWGRTFRSFG-----KV
GPRW-----ALGGDRLSLVSDRLKVATACIVYVPIY-----QGFR
GPIRRISD-----
-----GIWTEDPGQNEGYRHQPSDCSAMS-----

>GFJU01140648.1

-----LAKPMRWFLP-----KNPLSWPAI
PMVE-----VGTASDVVKALTREDVKTVL

-----SFIE-----PCRVINSIH-----RR
PWLDLRPYRSFVN-WVCCMVHE-----KQEQAIAIAIK-----DLSFAVRMWAIE--
-----GRFPETWKSRSRSH-LPH-----TFVRWVKSFLCHTG--T--
---RVQFARLGRC---LP--VGS--RAVMDTARANFLAV-VHTE-PKD---VSVD-----
-----IADELYEFGRS-----FIVSEGLKPVSWE-----PNEST--ATLNFH
ASQG-GRMA--ELI-----YDARLPLHLLEEFTP-----
-----I
TEPN-----ELLDEVIVETAIEK
SRIE-----
----DWEPELLV--VPERG-YKARVLVKFPASALLV-GDIVR-----RQLWP-QI
EQAEWL-----DLDREVNS-EKLTYS-----FLRRCIDPNTGLG-----
----QC-VSA-----DLSNATDYIPH--LYAQALWAGILEK-VD-----
-APDF--VKGYVEKM-FS---PMKI-----
--EGR-----MTKRGIHMGTPLSFMTLCLMHRYA----L---
-----RKAGLHR-----
-----HPH--IIRGDD--LIAI--IPQP---DIYFSAMTS-LGFKINKSKTI
VS--KNG-GTFVERV----F-----K--FEREDV--IIPH-

PILRLPPQ-----
-----SVITRA-----LEVVDIPLAGMIWPS--ERGSKL-----RTI
GRWYNQC-Q-----SRLTPR-----MKRTVS--MLIYMRHKDILA-K
A-TKL-RLPIS--TPVEL--GGCG---L-----PNRH-----
----GRNDFDAPHWY----RAIVGWAAC-----HESMAKKFSGRVTRLDGKDPNPP-
----YYIQNI-----RENQSRSTYE-----YDPTTTRYRG-----

-----MKRRFQIFRWA-----GAL
DEDRFKARVEREQSRFISL---PLNQWYEPFKNLK-----RV
RPRW-----VPRRRADTDRLIRRIKMFNAE-----YVL
PYTSRVQF-----
-----LGSDSYKTGGDT-----

>KX883548.1

-----MTKSMRRFYLP-----HEPWRWPVT
PTPE-----VGRGSVPTLADRDMR-----

-----DF
AR----RVWH-----SVVTPIRMVTSGRLSR-----
----IAEFRSFAT-WMVRKVAA-----GERSVVCRELK-----NLSFSFRVWAV---
-----EDRWTfHSDQ-LPR-----SFVRALKGLAPCSE-----
--ARAQFARIGRS--LP--TGD--GKVVEAAIRSYQRI-VVGR-PNP---SWD-----
----VADELYAFGKR---WAE--RAHLVRGRVNWAP-----NENT--ASLNFK
ASEG-GRLA--ELI-----ADSAAVIEPLLEFTP-----
-----E
GEGS-----ELLDEVLIETSIEK
ARTE-----
----RPIPKVIA--VSEKG-YKARVLVEFPASTLLP-GDIIR-----RQLWP-ML
EEEEWI-----DMDVVPNE-ERLKA-----FLRHSVDHSA-----
----TC-ISS-----DLSNATDYIPH--LYAQALWRGVLD A-IP-----
-APEW--VG DYTDKM-FG---P WDL--L-F-----
-PDGK-----TVTSQRGIHMGTPLSFMTLCLFHKFA----V---
ERS-----GFS---H-----
-----HPH--LIRGDD--LIGL---FPDP--ARYCRTMED-LGFKINKSKTI
IS--SKG-GVFVERT---F-----R---FTTGRE--PRRF-
-----ERR-----
SLASFLPD-----
-----RMV-----IRETTELKDVPVAGLLNPS--VEGSL-----RSV
GRWYGQL-----PDLSAR-----RRKRLH---TVIWSVFGDTLR-K
A-RCS-GIPIS--TPVQL--GGCG---I-----PNRH-----
----GRLDIDASFKM---RRSMALLNK-----RSVARQF--AYAVSAIDNLPHTD-
----FDRLFL-----KHHEESQRSRSVYT-----YDPQSTGYYQS-----
-----ASRR-FKMYRLLGGGA-ESVK
-----RRGLREGKLHK-----
-----PA---GFRRWV--RLFSN-----LKMS
CPRY-----AVGRRANSSRLETYLKISFGTYIPV-----AS
PGYSIFGP-----GNGEGGVAHQ A
-----PQLVV--ALP-----

>KX883542.1

-----MTTSSRYFDLP-----SNILSWPVT
PGVV-----GAG-----HWHQLPVSALTDRDGVRYSN

-----SVWK-----AVVTNIVLAHNKKFRN-----
----VVP AKIFAG-WLLRCSRNA-----GG RYYVAKQLK-----QLSFEYRQWSVT--
-----GRRPTPKSDVPRT-FQK-----WLAGHVCHAEA-----
---RAQIARIGRC--LP--EGD--TLVQSKAIKAHRSI-VTRP-PCD---VNME-----
----LADDIYYWARK---FSTD--NAHLFKEEVS WA-----PNENS--ASQGAT
ILEG-GRME---ELI-----RDAKAESDFESVKSLAEGDF-----
-----D
PWVS-----KALDDAVYMTACTV
PPKK-----

-----EWIPDVMA--IAEKG-YKARVLTkFPAAALVV-GDVIR-----RQMWA-MV
SDKPWA-----DMDRQPDD-VKFLS-----TIRSALDRHG-----
----IC-VSS-----DLSNATDYIPH--IYAQALWAGLLEP-HH-----
-VMPY--INSYCAIM-FS---SMKL--R-Y-----
-PDGT-----VVTsARGIHMGTPLSFLTLTLLHRFC----V---
-----EKAGYGS-----
-----YPH--IIRGDD--LLGI--FPRPEV---YFNVMQQ-VGFSINRAKTI
IS--RTG-GTFAERT----V-----R--FSHSLA--TQE-L
SNP-----LKR-----
TLGQFIPV-----
-----NISSV-----KVLQDLPVGGVVRAT--PGKGSLV-----KAL
GRWFSQT-----SNIV-----PRQRRK-----AYRAIG-----LNHGDLV--
I-RLSTTVPPH--LPLDL--GGAG--L--PD-----RKGRV-----
----GLNGVP-----FAIRAAIGHA--ASHHETAVKL--TGLIARTDGVS-RG-
----FVDVFS-----KKDF-----REQWSKSIW-----
-----TTE-PQETEFYQRTR-RYWR
-----YFGHRDRPSRP-----
-----I---SFRQWRSGLMTLP-----RV
KARW-----VPRSNSDPSRL-----ANR-----IR
SMTGTYIP-----YTQPTTG
ST-----YQWAA--KCG-----

>GGQW01011558.1

-----AEHYERFHP-----VNPLSWPVI
PGVV-----RAGGWYNLPSEALTGEGLKF

VK----SVWE-----AFVVVfVASHKKTIKN-----
----LVPPKIFLR-WVMSCSKNE-----TSRDFCCEQLK-----KLAFIYRRWAVE--
-----DKYPTDHCSIPSS-FLK-----WLKGNMSSCTQ-----
---RAQIARIARC--LP--KGT--TVRMERAIegHKKI-VSIPARVD----ID----
----LADEIYHFAAK---YTAE--RKEKLIKVSFH-----PRENT--ASLEFT
EKAG-GRLA---DVM-----ANASDWLRRRPIALSEHLYTNKSYLDE-----

-----AIVEAAV--KQDLA
YPYY-----
-----RAEVVA--VAEPG-NKVRVVCKFRSVPLLI-SDIIR-----RQLFP-IF
EEDENM-----DFDQDIRS-EKFGK-----MIRRCLHNEG-----
----VI-VSS-----DLSNATDYIPH--HYAQALWHGILEQ-FD-----
-APEW--VLKHVLRM-LS---PVEF--N-Y-----
-PDGS-----VIMSQRGIQMGTPLSFMTLSLLHKFA----V---
ERS-----RNGE-----
-----SSY--LIRGDD--LIGV---FKYP--RSYYSEMEN-LGLKINANKTL
QS--HRG-GVFAEQT----V-----K--VSWKKS--TRTSY
HR-----
TLYDYFPA-----

-----AETDEV-----SRIWVLDDIPMKGLLHVD--PGVYGSK----LRMI
GRWYSQW--SSYPICSR-----KR-----AYRAIG-----IAHKELIK--
T-ARFLGLPSS--TPLEL--GGCG--L--P-----NRKGALA-----
-----LDASFLHRGQ----IGYAASHSDSRFSHFVGKLDGIV-
-----KDDGLL-----EQSIERSRRSLGSS-----VFETDPF-----

-----NTEFYLKARRA-----
---RAYLSPDYLKERERCT---PLSQWKKSFRKLP-----TV
KARW-----VPKRDKDSFGIDRLIDLIKI-----FN
GRYVPFTP-----

>GBDZ01000351.1

-----SSPDNTGAMTSQCETTYGLP-----ISVLSWPVN
PGVV-----GAASW-----IPNPRTALTGVERGRALAYA

-Q----GVWY-----ALTSNIEAVHPRPFKS-----
---LIPAKILAG-WVVRCSATA-----AGRTQCSGLLK-----RFAFLYRLWAT---
-----TDTKPLVDGIPK-CPQ----LWIRSLCSHASV-----
---RAQIARLGRM---LP--QGD--AVTVRKALKKHSV-VTQSTRAN-----VF-----
---VADEIYAFAKE---YMEN--REEHLREKVSFHP-----TEGT--ASRDYT
IERG-GRVQ---ELL-----DNAWAMMRKPGYIDYQKYPDKS-----

-----RIDEFIV--QAAID
ADKR-----
----LHHYRVSVMAAAEPG-NKARVLCKFPAVALVP-GDIIR-----RQLWP-IV
ASDPDL-----DFDQDPRS-EKFLQ-----MIRRTVDREG-----
----TI-VSS-----DLSNATDYIPH--MYAQALWRGLLES-FD-----
-APSW--VQEHINRM-FS---PMEM--T-Y-----
-PDGK-----KVWSQRGIQMGTPLSFMTLSLLHKFA----V---
QAS-----GHGLSAH-----
-----IIRGDD--LLGV---FTSP--ADYFRSMER-VGFKINREKTI
VS--RVG-GVFAEQT---V-----R---VRYTPR--RQP--
-----KKI-----
TLADFMGP-----
-----PREDVESV-----TVLNDIPLKGAVHVD---TKGSVL-----RGL
GRWYAQW--SEQYP-----AKCGRT-----AHRAVM-----RVHQALIR-D
A-KRW-RVPLH--TPLEL--GGAG--I-----PDRRGRL-----
----GVADLS-FDMR-----RKLGYACSHFDRSYARAVRRLD-----GGST-DP-
----FEEHFE-----QMLKKA-----PLGSCVYEYDP-----
-----YA-TEYYAACRRKY-YLTK
-----VTRPARKDLNP-----
-----P---PLRDWL--KQFDN-----CPAV
QPRW-----AVSKHADVTLLIHLKMFSGA-----YV
PFTTPETS-----

1
2
3 -----ILARR--KVR-----
4 -----
5 >GBFE01004865.1
6 -----
7 -----
8 -----GGAMTSQCETTYGLP-----ISVLSWPVN
9 PGVV-----GAASWIPNPRTALTGVERGR
10 -----ALAY
11 AQ----GVWY-----ALTSNIEAVHPRPKS-----
12 ----LIPAKILAG-WVVRCSATA-----AGRTQCSGLLK-----RFAFLYRLWAT---
13 -----TDTKPGLVDGIPK-CPQ-----LWIRSLCSHASV-----
14 ---RAQIARLGRM--LP--QGD--AVTVRKALKKHRSV-VTQSTRAN----VF-----
15 ----VADEIYAFAKE--YMEN--REEHLREKVSFH-----PTEGT--ASRDYT
16 IERG-GRVQ--ELL-----DNAWAMMRKPGYIDYQKYPDKS-----
17 -----
18 -----
19 -----RIDEFIV--QAAID
20 ADKR-----
21 ----LHHYRVGVMAAAEPG-NKARVLCKFPAVALVP-GDIIR-----RQLWP-IV
22 ASDPDL-----DFDQDPRS-EKFLQ-----MIRRTVDREG-----
23 ----TI-VSS-----DLSNATDYIPH--MYAQALWRGLLES-FD-----
24 -APSW--VQEHIDRM-FS---PMEM--T-Y-----
25 -PDGK-----KVWSQRGIQMGTPLSFMTLSLLHKFA----V---
26 QAS-----GHGLSAH-----
27 -----IIRGDD--LLGV---FTSP--ADYFRSMER-VGFKINREKTI
28 VS--RVG-GVFAEQT----V-----R--VRYTPR--RQP--
29 -----KKI-----
30 TLADFMGP-----
31 -----PREDVESV-----TVLNDIPLKGAVHVD---TKGSVL-----RGL
32 GRWYAQW--SEQYP-----AKCGRT-----AHRAVM---R---VHQALIR-D
33 A-KRW-RVPLH--TPLEL--GGAG---I--PD-----RRGRL-----
34 ----GVADLS-FDMR----RKLGYACSHFDRSYARAVRRLD-----GGST-DP-
35 ----FEEHFE-----QMLKKA-----PLGSCVYEYDP-----
36 -----YA-TEYYAACRRKY-YLTK
37 -----VTRPARKDLNP-----
38 -----P--PLRDWL--KQFDN-----CPAV
39 QPRW-----AVSK---HADVTLLIHRLKMFSGA-----YV
40 PFTTPETS-----
41 -----ILARR--KVR-----
42 -----
43 >KX883517.1
44 -----
45 -----
46 -----
47 -----
48 -----
49 -----
50 -----
51 -----
52 -----MAVVIEAIH-----RR
53 PFTNLVPAKMFIAG-WLIRASRG-----GGQKYVCQQIK-----SLAFDYRVWSI---
54 -----TAKKPNYSGD-IPR-----AFRNWLSSLCCHAE-----
55
56
57
58
59
60

--TRAQFARVARM---LP--VGN--KLVIDQALRKHRSV-VTR--RPK---LDIG-----
----LAHQIYGYARD---YIHE--RKHSLSERVSFHP-----TEGT--ASRDVP
ASAG-GRLR---DLV-----DNAHRVMSQRPIVLSEYLYSHKK-----

-----YFDDAIV--EAAIY
TDEK-----
----VSTYNVDVLA�PEPG-FKARVLCKFPATALLA-GDIIR-----RQLWP-IF
ESDPNM-----DFDQDIRS-EKFQN-----VIRRLNRDG-----
----TI-VSS-----DLSNATDYIPH--EYAKALWAGILDA-FD-----
-FPEW--VENYLERM-FA---PIRM--S-Y-----
-PDGV-----TVTSCRGIQMGTPLSFLTLSLLHKFC----V---
HKS-----GHE--R-----
-----SPY--IIRGDD--LLGV---FSSP--RQYLTVMEE-IGFKINRDKT
IS--KDG-GTFAEQT---V-----K---VTWKAK--ERD-P
L-----ARP-----
TLYDFIVT-----
-----DQKV-----VSSITNLDDIPFKGLIHL--NKGGR-----RQV
GRWYAQW--SPYYP-----PRKGKV-----AYRAIR-----RTIGNVL-R
I-ARSLRIPIT--CPMEL--GGCG--I--PN-----K-RGTMQ-----
----LDANFQ-HRSR----VGYAASHESH--NFQLAVRKLD----IGNASDLL-ED-
----YRQHVA-----DLPKGK-----SVYMFDPYS-----
-----SED-FSGYRRKRMLL-NNTG
NYSAY-----HTKPVPLHRWL-----
-----AVFSRCKE--NRARWVPSRRCG-----SD
LRRL-----IDR-----IK
MLSPIYVP-----YTTP-----
-----NAQIL--SIRI-----

>GFLP01591397.1

-----AYSTTMRELLP-----SMLTYVRVG
ESHP-----SPSERTGPGSRPRDRKSRKA
-----GSRGKR
QRASREDLVWQ-----AVWLGFIAGISYQ-----
----LGCWRTRE--WVKRSIRR-----DGVHNTCSKLL-----DLAVSIRGYVL---
-----SDVQPPMVPGVPN-----RIQRWLRAEVDKDT--LS-
---GLAFTRFARA---LP--KPD--FRETANAMASHLDL-ITDRRVVP-----EG-----
----IQDQIEQHVVVR--MFGG--QFVGILRHPST-----PISNH--AVGERS
KARG-GYNAHIYDCA-----GDPEFLGVDPPLDRIPAYPPTRALWRLADSPPGE-----

-----GIPEYMASASALSA
ARIV-----EQFARDTS-
----RVVHSATG--IGEQQ-DKCRIITVPPASLFAA-GDVCR-----SRIWP-RV
QQSDAR-----IFGNDEDA-KSSL-----QMGLPR-----GS-----
----VY-VSA-----DLTKATDGAFAH--DAVRAVLRGLARC-G-----
-VDNH--TVKLMSET-LGVGDRQHYV--K-Y-KLAELPRRKREWAVQRFETIEE-----

1
2
3 -DGVS--TH-----VLVPLVRGILMGTPCSFIILSILNGWCC----M---
4 PLG-----
5 -----PRT--VICGDD--VASA--CTPEAV--DNYDRRVTV-IGSGLHKRKTF
6 IG--HKG-LLFCELY-----
7 -----
8
9 VLPGYGEH-----
10 -----RCFEPV-----PLKSLAKDG-----DGT
11 LDTSDFD-----TFIWKR-----MDRACR-----VLWRDVRA-K
12 A--RRLNRWPQ--LPVEL--GGLG--H-----PSSG-----
13 ----KMGAVP-GSVR----NRLATLIKVGVPVKKIARWTVAAGPLDWRSFRSDKE-VA-
14 ----WSLFES-----TDAAVSQLEAD-----SRYVDTSFVSYS-----
15 -----HE-ANQYVSVLANQ-----
16 -----LYSAHGGKFVN-----
17 ---EFNRKAMKPGKIQYP--GVGSLQYSAKAPM-----SMV
18 AREY-----LAKLHAGGEYLPTDSVRKIRR-----
19 --RIGKST-----
20 -----
21 -----
22 -----
23 -----
24 >GGCO01034162.1
25 -----
26 -----SYTKNMKVRLSSHLTRLEFKD-----STAPSPRKS
27 GGCK-----PLSKIRTKHWHSDKERTPEQ
28 -----
29 -----IVWE-----ALWAALVVSDSVN-----DL
30 DVSSMRGSFYLRD-WVARSSR-----RTVGYVVSVLK-----EELNRLRAISL---
31 -----GSDPPPIAHG-FPK-----KLDTFLOKQVFTPLG-----
32 ---LFAFSGVSRG--FP-TCVD--KVKVSETLVDHLVN-VANPAPAV---PIN-----
33 ----FLRSIEQFVPR--YIQKSVCYDRKLADKMHDPSPVVC----KSSPS--ACYENP
34 RGAG-GFYAYIKKLG-----DKISEGQNDLLADIRAGVKPRPEDICHPIKGIKASYD
35 PDR-----
36 -----TQDLLSPKEMSDIL
37 DHFY-----FYSGCNEYL
38 DNPDKIEIVHRVAV--IPERG-YKNRVVTAPPASILSM-GEVVR-----SSIFP-YV
39 KAHPSC-----DIVKDGDMPVVC-----PAVDGA-----
40 ----KL-VSA-----DLTKATDGFHSH--DVIRSVGVGLRNA--G-----
41 -FPQH--VVDFFVDT-LGAGERPHYA--E-Y--DVHAILRRHSKNTVKYNSLCSFLQSYG
42 WDGIS--SK-----VKVPMLRGSPMGTPTCSFTLLCIVNSWA----L---
43 DHC---S-----
44 -----NGP--KICGDD--MLAY--ITAKEF--SKYKVRVAA-IGSGVHPLKTF
45 IS--PYA-GTFCENI----Y-----I--KDPCDS--KFP-S
46 VVAFGRAIKCPIKAMLY-----PQK-----
47 GSNGRIDW-----
48 -----PDEIVHQV-----FTGETRMSKEKVCLSSFND--EPEYHVQ-----
49 RRFEAGP--GDFGYNDGPSV----FRLGRK-----VHRVLR-----TLRRDQIRSA
50 L-KK--GRYPW--LPACL--GGLN--Y-----P-RKNPK-----
51 ----GVRGLD-KHCR----ARLFQFTHSD--LSVVELGKFM-----SSLD-NA-
52
53
54
55
56
57
58
59
60

-----KRPLPE-----HRELEFISDMVADA-----EINDAIVEDTL-----
-----ERQGECS-FVDSYSLARHQ-GSVR
--SNF-----FRSMGGKLDLS-----ASSYSIFNAA
RIRWPAVKGELGTYDKRTP--FSLIQDLNLLR-----
-----APRYRVLSSFAEKLPGVEHPKSNR-----WH
HLHGA-----
-----GGWND--GIL-----

>GAIS01005902.1

-----E
WRLQ-----ALSKIRTKHWHSDKERTPEQ

-----IVWE-----ALWAALVVSDSYN-----DL
DVSSMRGSFYLRD-WVARSSR-----RTVGYVVSVLK-----EELNRLRAISL---
-----GSDPPPIAHG-FPK-----KLDTFLLQKVFTPLG-----
---LFAFSGVSRG--FP-TCVD--KVKVSETLVDHVMN-VATPAPAV---PIN-----
----FLRSIEQFVPR--YIQK--SVRYDRRLADKMHDPSPV--VCKSSPS--ACYENP
RGAG-GFYAYIKKLG-----DKISEGQNDLLADIRAGVKPRPENICHPILRGIAKASYD
-----P
DRTQ-----DLLSPKE--MSSIL
DHFY-----FYSGCNEYL
NNPDKDIIHRVAV--IPERG-YKNRVVTAPPASILSM-GEVVR-----SSVFP-YV
KAHPSC-----DIVKDGDGDM-PVVCF-----PAVDGA-----
----KL-VSA-----DLTKATDGFSH--DVIRSVGVGLRNA--G-----
-FPQH--VVDFFVDT-LGAGDKPHYA--E-Y--DVHTILRRHSKNTVKYDRLCGFLQSYG
WDGHS--SK-----VKVPMLRGSPMGTPCSFTLLCIINSWA----L---
DHC-----SNGP-----
-----KICGDD--MLAY--ITAKEF--TKYKVRVAA-IGSGVHPLKTF
IS--PYA-GTFCENI-----YIKDPCDSEFP--

SVVAFGRA-----
-----IKCPIKAMLYPQ-----KGS-----NGR
IDWPDEI-----VHQVFT-----GETRMSKEKVR
L-SSFTDEPEYH-VQRKFEAGPGD-----FGYND-----
---CPSVFRLG-----RKVRRVL-----RTLRRDQ-----IRSALKKGRYP-
-----WLPACL-----
-----GG-LNYPRRNPKGVMG---
-----LSGLCRARLFQ-----FTHSDLS
VVELGKFMSTLDNAKRPLP--EHRELEFISDMVA-----
-----DAEINDAIVEDTLEKQGECSFVDS-----YS
LARHQGSV-----
-----RSNFFRSMGGKLDLSAE-----

>GAIR01012062.1

1
2
3 -----YTKNMKVRLSSHLTRLEFKD-----STAPSPRKS
4 GGCK-----PLSKIRTKHWHSDKERTPEQ
5 -----
6 -----IVWE-----ALWAALVVSDSYN-----DL
7 DVSSMRGSFYLRD-WVARSSR-----RTVGYVVSVLK-----EELNRLRAISL---
8 -----GSDPPPIAHG-FPK-----KLDTFQLQKVFTPLG-----
9 ---LFAFSGVSRG---FP-TCVD--KVKVSETLVDHVMVN-VATPAPAV---PIN-----
10 ----FLRSIEQFVPR---YIQK--SVRYDRRLADKMHDPSPVVC--KSSPS--ACYENP
11 RGAG-GFYAYIKKLG-----DKISEGQNDLLADIRAGVKPRPENICHPILRGIAKASYD
12 PDR-----
13 -----TQDLLSPKEMSSIL
14 DHFY-----FYSGCNEYL
15 NNPDKDIIHRVAV--IPERG-YKNRVVTAPPASILSM-GEVVR-----SSVFP-YV
16 KAHPSC-----DIVKDGDMPVVC-----PAVDGA-----
17 ----KL-VSA-----DLTKATDGFSH--DVIRSVGVGLRNA--G-----
18 -FPQH--VVDFFVDT-LGAGDKPHYA--E-Y--DVHTILRRHSKNTVKYDRLCGFLQSYG
19 WDGHS-----SKVKVPMLRGSPMGTPCSFTLLCIINSWA----L---
20 DHC---S-----
21 -----NGP--KICGDD--MLAY--ITAKEF--TKYKVRVAA-IGSGVHPLKTF
22 IS--PYA-GTFCENI----Y-----IKDPCD-SEFPSV
23 VAFGRA-----IKC-----
24 PIKAMLYP-----
25 -----QKGSNGR-----IDWPDEIVHQVFTGETRMSKEKVRLSSF-----TDE
26 PEYHVQR-KFEAGPGDFGYNDPCSVFRLGRK-----VRRVLR-----TLRRDQIRSA
27 L-KK--GRYPW--LPACL--GGLN-----YPRRNP-----
28 ----GVMGLS-GLCR----ARLFQFT-----HSDLSVVELGKFMSTLDNAK-RP-
29 ----LPEHRE-----LEFISDMVADA-----EINDAIVEDTL-----
30 -----EKQGECS-FVDSYSLARHQGSVRS
31 NF-----FRSMGGKLDLS-----ASSYSIFNAA
32 RIRWPAVKGELGTYVKRTP--FSLIIQDLNLLR-----
33 -----APRYRVLSSFAEKLPGVEHPKSNR-----WH
34 HLHGGWGG-----
35 -----GMMGY-----
36 -----
37 >GACI01002802.1
38 -----
39 -----VRLSSHLTR
40 V-----DFKDLPTPLPVKGGGSKPRS
41 -----TIISRDPRHRRP
42 IQ----IVWE-----AVWAALIISDSYD-----GR
43 DVSSLRGSFIFRR-WVARSSY-----RSVGYVVAILK-----LEMNRLRAISL---
44 -----KSSPVPGVPEGFP---K-----RLDQFLLDNISPG-----
45 ---LFTLSTMTRG---FP-VTTD--KSDKAKALLEHLEN-ISTPCAPV---PTP-----
46 ----YLRSIDFVPRMFNESHFAKNARDFTKNVQFPAPDTI----KLSSS--ACYEGP
47 RSRG-GYLG---HIKKEGDDDAKDRIANGYQEILKDIRDGVRRPPENDTHPLLFNIAKRS
48 YDP-----D
49
50
51
52
53
54
55
56
57
58
59
60

NLIS-----DLSPSRL--AEIMS
YYYY-----WNNCDKYMK
SGHRDEQVHRVST--IPERG-YKNRIVTAPPSGVLSV-GEVIR-----SIIFP-FV
RDHPSC-----TLVRDGD-LRGARF-----YDRTGC-----
----RL-YSA-----DLTKATDGFH--EAIEAVGRGLLS--YG-----
-FPTSVVEHFLGTLG-VG---SKHY--AEY--NVADLLQGHKEGSRKYNQVCSYLMEKG
WDGNS--KV-----MRVLMRRGSPMGTPCSFTLLCILNLWS----L---
DRA--C-----
-----NGP--RICGDD--MLAY--ITPHDF--AHYSRRISI-IGSGVHVTKSY
VS--DIA-GTFCENL----Y-----VRSQTQ--D---

GIVSFERG-----
-----IRCPKSVLFPQ-----KGS-----NGR
IDWPDVPHQVFTGETRM-----EOPYRQ-----VATVVP-----TEQGYVHTR
QRELHQGDFLFEDLPLTV--KIG-----
-----VRVARVL-----RTLRRDT-----IRACKKGRNP-
-----WAPACL-----GGLNYPRKNAKGK--GFPKWFRANLY-----
-----SF-THSDLSVPE-----
-----LSKVLAQIDCE-----
-----KKPDSLSKELE--FVSTMVADTEINE-----SL
VPEI-----HDRVGDCAFVDMYALSRHYGGIKS-----NFFR
SMGGRLSV-----
-----ASSYS-----

>GAIR01012025.1

-----SYMK-----NMKVRLSSH
LTRV-----DFKDLPAVPPVKGGARKPRS
-----TIKSRDPANRRP
IQ----IVWE-----AVWAALVVADPSDG-----IR
DVSSLRGSFHLKD-WVVRSSH-----RSVKYTVETLK-----AEMTRLRAHSL---
-----LSTEIPVPHG-FPR-----KLDRFLRAAFAPMG-----
---LFTLSTMTRG---FP-VSTD--PKQRSALMGHMENVMSPPPSVP-----AH-----
----YLRSEIEFVPR--MFKESSYKRSDDLRIVNPAPVCN---KTSSS--SCFEGP
RSKG-GYFG---YIR-----TKGDKAGEGKDIAKEIREGTTLPPEKTCLPLFRGIAEKS
YT-----M
EFPA-----PLDEF---ADIIS
HYYF-----LNDCELYN
RDPSQQLVHRIGV--VPERG-YKNRIVTSPPSGVLSA-GEVVR-----HVLFP-FV
KSHPSK-----EIVRDGDM-VNASF-----YNRPYC-----
----SL-YSA-----DLTKATDGFH--EAILAVAQGLKT--VG-----
-FPTS--VIDIFLST-LG---VSEI--H-Y-AEYSIKDLMAGLRVGSRKYKNRASELMS
RGWDGFTDK-----LRIAMRRGSPMGTPCSFTLLCILNLWA----L---
DAA--K-N-----
-----GPK--ICGDD--MLAY--ITPSEF--NSYSRRIAA-IGSGVHPTKTF
VS--RSA-GVFCENI----Y-----IRREDS-DGIACF

1
2
3 NRGVRIPIKSILF-----PQK-----
4 GSNGRIDW-----
5 -----PDAMVHQMFTGETRLSESYSVVELVPPEEQGYKR--IRHRELT-----DGD
6 RFYEDLP-HT-----MRIGVR-----IRRVLE-----TLRRDAIRAC
7 L-KRN-RVPW---APACL--GGLN-----YPRKNPK-----
8 ----GVKDFP-AWFR----RDLYNFTHSDLVGELSKVIAAIDCQKKPSKGSEELGM-
9 ----LSTMVA-----ENDINEAIRDDIGE-----DVGECVSVDVY-----
10 -----TL-ARRAYGSARARF-----
11 -----FTSMGGTLDTK-----ASSYSAFC
12 VPKIDWPKGSGGCYDCKTP--MELVIKDLNLLR-----
13 SPNY-----
14 RITRDFS-----
15 -----
16 -----
17 -----
18 -----
19 >GGCO01105932.1
20 -----
21 -----YMK-----NMKVRLSSH
22 LTRV-----DFKDLPAVPPVKGGARKPRS
23 -----TIKSRDPANRRP
24 IQ----IVWE-----AVWAALVVADPSDG-----IR
25 DVSSLRGSFHLKD-WVVRSSH-----RSVKFTVEILK-----AEMTRLRALSL---
26 -----LSTEVPVPHG-FPR-----KLDRLRAAFAPMG-----
27 ---LFTLSTMTRG---FP-VSTD--PKQRSAELMGHMENVMSPPPSVP-----AH-----
28 ----YLRSEEFVPR---MFKESSIYKRSDDPKIVNPAPVCN---KTSSS--SCFEGP
29 RSKG-GYFG---YIR-----TKGDKAGEGKDIAKEIREGTTLPPEKTCLPLFRGIAEKS
30 YT-----M
31 EFPA-----PLDEF---ADIIS
32 HYYF-----LNDCEDYLN
33 RDSSQQLVHRIGV--VPERG-YKNRIVTSPPSGVLSA-GEVVR-----HVLFP-FV
34 KSHPS-----EIVRDGDM-VNASF-----YSRSSC-----
35 ----SL-YSA-----DLTKATDGFSH--EAILAVAQGLKTV-G-----
36 -FPTS--VIELFLST-LG---VSEI--H-Y-AEYSIKDLMAGLRVGSRKYNRASELMS
37 RGWDG--FT-----DKLRIAMRRGSPMGTPCSFTLLCILNLWA----L---
38 DEA---K-N-----
39 -----GP--KICGDD--MLAY--ITPSEF--DSYSRRIA-IGSGVHPTKTF
40 VS--RSA-GVFCENI----Y-----IRREDS-RGIACF
41 DRGVRIPIKSILF-----PQK-----
42 GSNGRIDW-----
43 -----PDTIVHQMFTGESRMSESYSVVELVPPEEQGYKR--IRHRELT-----DGD
44 RFYEDLP-HT-----KKIGVR-----IRRVLR-----TLRRDAIRAC
45 L-KRN-RVPW---APACL--GGLN-----YPRKNPK-----
46 ----GVRDFP-AWFR----RDLYNFTHSDLVGELSKVIAAIDCQKKPSKGSEELDM-
47 ----LSTMVA-----ENDINEAIRDDIGE-----DVGECVSVDIY-----
48 -----TL-ARRAYGSARARF-----
49 -----FTSMGGTLDTK-----SSSYSAFC
50 VPKIDWPKGSGGCYDCKTP--MELVIKDLNLLR-----
51
52
53
54
55
56
57
58
59
60

-----SPKYRITRDFSGKIQGAEHLKSSR-----WH
LLRGA-----
-----GGWND--GDL-----

>GBTA01004534.1

-----LVPEI-----

-----KEFT-----
---LKIVSKMYQDWKSKTEKFNE---QCEETGRKKINLDLSVPEWSPNLHASAC---
-----IENPIKNGGQQAY-MHK-----
---LFQVAHTGVG-----EARNARAFAMRIRN-----
---VLQQEMIWLKSMGDWQVG-----NELPS---TTD
RKEG-GVEE-----IPKDDFRFNTSMPE-----

-----GIPDYVS-----
-GSY-----
----LPVGRASV--VYEKG-MKARVVTVHHA AFVAL-SCALN-----RTLIN-LL
KCYTPS-----KASLLGRD-IELKT-----LKR DGLEG-----
-----FSA-----DLSTATDLIPF--EVARAVVDGICEA-LN-----
-WGPH--IKEIMYYA-TG---PFRI--E-YDRA-----
-PRSK-----DSIVPDYIISRRGILMGLGT SWPILSILNSFC-----A---
EYN--C-----SKEGV-----
-----RNY--RICGDD--LAAL--WDKNDI--GNYIGNLES-VGLKLNDKKT V
TS--DQG-LVFTEEL----Y-----S---IKSRTH--INRKG
ERSPD-----QKD-----
LYLSRFRK-----
-----AIVSVI-----VDAKHNSTTKFTNPE-----ERL
PRYLTIG--DCSHAIDEFCP----LWQKSR-----AQKALL-----LMNKKCIR-E
I-FRA-GLDPF--LPKYL--GGAG-----VK-----
----TVGNEKRHPLM----KHL SALIQLPYEDIQLVCRRLQ-----RP-
----WLSSSV-----NKETSAIFS-----
-----KT-LNLFEKLPRAEHGSAI
PYKEA-----VRKFTGAVLAR-----
-----YSLNFGSCQTN--KVCTLKFIRKQVN-----KA
RKTL-----FSKAPRRYHYMSLLK-----FS
NALEKLEP-----

>GEWH01003494.1

-----TLNS-----SVAEKSETG
KRPR-----SRMEKLLRRCTERDMTTRQL

-----LETL-----GLRLDSLQLGQDLVQAFSTLLDVER-----GN

1
2
3 KYFERRWFIRLAQ-WLLRAIDSFTDKVKAEGISSAMTWIK-----DALTHLRACAL---
4 -----DAGKVTDYPFPSK-APR-----VERRRM-----
5 ---LVQASRCSRS---LR--AAT--GDKLEQAAQSHKEI-LTSPFKTP-----EP-----
6 ----ELLSFREYVRA-----KTRASTVQVG-----EVSAS--ASLDTK
7 GSGG-GSLE---EVRGLVD--SLRSRPFTYPDLLELDRTFPHEMLGGHVLLNRPKVQERF
8 RIGQVGHSSSGTL-----L
9 GVPG-----DQVLFLYDQSCEISLDEWEFLREQLFA
10 IVAC-----VLDVDLDS-
11 ----RPACRQCV--IAERG-WKTRVVTPIAASNLY-LTVAN-----KGLLS-II
12 GEHREL-----DNSQRGRPCDNLDW-----SFGKRY-----
13 ----ALTFS-----DLKSATDYMPH--DLCLVAAEELTRG-----
14 -WPAY--LRRLVLRV--VG---SYDL--R-C-----
15 -LDGG-----TVETRRGILMGVPTSWPILSLYSGWL-----
16 HYR-----SSSDGWF-----
17 -----AVCGDD--YLGC--HTYASY--HKYLRARER-TGAKGSPGKDI
18 LS--TESTGIFAEEL-----VTVGRC-----
19 -----
20 RVLRTVPI-----
21 -----RPLTGI-----PK-----GDT
22 PSWAMGEAISVLLSEA-----GFSEER-----SGRLVS-----RRFGSEMDK
23 L-RRR-GIDPK--GPRWALAGGFP-----GKPG-----
24 -----PRVL---ETARRVLS-----QGEDITL-----GWLTAIG-IA-
25 -----WSDAPS-----RDLIQGM-----
26 -----
27 -----VRDKLGGLKGQ-----
28 -----
29 -----
30 -----
31 -----TWRAE-----
32 -----
33 -----
34 -----
35 -----
36 -----
37 >GDUK01007882.1
38 -----
39 -----KVEKWLP-----PFSVFGGSS
40 MASR-----SRPLHVGRIRDLRVEMQSQH
41 -----FPGASMTADQI
42 AQ----QFWN-----VVRSHLLAAFPLVSGLPVA-----KR
43 SGRNARFGLVLLQYHLLRRLHL-----GGVASLVRWLK-----ALAFSCRDSVV---
44 -----SDKKLPRE---FSR-----YFHGVLNPSALS-----
45 ---LTTLAKLGRALPECDAKMGG--GSVIRKALRQHRAD-MTSARVKQ---FGFN-----
46 ----PMDRLDLAVAK---ETASFLLRGRLGSPRM-----HDLSS--ASADVS
47 SREG-GLAA--GLR-----PSWARMEEIGPVPWALGHALAEWQAALLRGARNTAWAT
48 AEEERG-----I
49 VLSNGSFLPGPLVLLASPDGLWVTRRSEYIRSVLSFHEGLGPVRVTGRFGTTVQVDHLTV
50 QFWH-----VAGEPDLKRFFA
51 SCARDLPKAKVAL--VRERG-WKVRVVTSLTDQVIA-GHALR-----DLIWP-VL
52 QDDPCF-----DLTGDP-----RIEGMPVDQG-----
53
54
55
56
57
58
59
60

<https://mc.manuscriptcentral.com/vevolu>

1
2
3 I-QRH-NLGLF--VPREL--GGFG---L---PH-----P-KGWAR-----
4 ----GVTSA--RWAH----KRLAVFFTAT----RGAAVKI-----GCLGYDP-
5 ----WVVDPN-----TGRLWGEVSKRLASR-----EKSCLVLLPPE----RV
6 AEPTADRTPESEWYQYVPN-----YG-GYYLAVVPGAGRGVAQ
7 PVTEAV-----AQEMALESAR-----RGLGLDPVVDKPQ
8 EKSRSKSGTSDGCKSKAKKP---SRSALTIISKHNR-AMRKTGKSRFPEKFLPKRALRPE
9 KLQM-----ALMRHKMYVQHGSFSGPSPLHIV-----FN
10 SEKGRGPP-----RRPGGWKYYH
11 DPKQGRSKTSSD---PVPRS--PAPPLVGGVPLDPAWENSWPSRKSNTS-----
12
13 -----
14
15 >GEUE01035234.1
16 -----
17 -----
18 -----
19 -----
20 -----
21 -----
22 -----
23 -----
24 -----
25 -----
26 -----
27 -----
28 -----
29 -----
30 -----
31 -----KATV--VPETG-GKARVVTAGPADMVLA-GNALR-----RAVWP-IV
32 MGDDSIDTEEAVAQEEVSAGDK-ELLAR-----LLGGGCS-----
33 ----W--YSG-----DLTAASDWLPR--DVCLTIWDAIAEA-AG-----
34 -FAKGSPEYLLGRKL-LG---NVVI-E-Y-----
35 -ADGT-----EVTGAGALMGFPLTWLVLCAYNRAL----V---
36 SMV---S-K-----
37 -----KAIQRT--VFRGDD--MVSQ--MTPEEG--ARYEELVRL-TGGQPNTAKSF
38 RS--VTG-WVFTEAT----Y-----Q---LTTWHT--TL---
39 -----TRA-----
40
41 DLAERGVT-----
42 --LEHGLGRAGTLRAI-----TVRRVPDCSLRHLLRQR-----VAS
43 QPYLAIA--GPAVSGAVA-----PLLASRLADVRAVGNRLV---RRWMRVNAARMVE-S
44 I-QRH-NLGLF--VPREL--GGFG---L---PH-----P-KGWAR-----
45 ----GVTSA--RWAH----KRLAVFFTAT----RGAAVKI-----GCLGYDP-
46 ----WVVDPN-----TGRLWGEVSKRLASR-----EKSCLVLLPPERVAEPTA
47 DRTPESEWYQY-----VPNYG-GYYLAVVPGAGRGVAQ
48 PVTEAV-----AQEMALESAR-----RGLGLDPVVDKPQ
49 EKSRSKSGTSDGCKSKAKKP---SRSALTIISKHNR-AMRKTGKSRFPEKFLPKRALRPE
50 KLQM-----ALMRHKMYVQHGSFSGPSPLHIV-----FN
51 SEKGRGPP-----
52 -----RRPGGWKYYHDPKQGRSKTSSDPVPRSPAPL-----
53 -----
54
55
56
57
58
59
60

>GBYB01012090.1

-----ASARLGRS
GKWKLPNLPEPVATAVTPPAAAARLLSNLRKDAFPRGVTGNYLS-----RLEAAGIPA
CSVSRHPRGSEPHPLRKYTPGRGDPACSPLTS-----LPHTPLPTSPSRPGRARGGP
-----AGERPRKRGHSRDPPGGCPLGCGAAADCRD
LR----RIWD-----SFRLCILSLIPRLQRSGLDLSN-----RR
DYRMAVSILKVRS-WLSINWRS-----SGRNETIRQIK-----AYAEFCRARALD--
-----AEQSRRGAPRGFP-VRA-----FEHGLGSLLRGASS--GRT
A--LAQLARLGRS---MP--PAD--QAILTAALEQHRRV-LTTADPGV---PLT-----
----LLDGLFWASS--WSGS--RPASLRNSVVA-----SFSSS--ASEGMS
RAKG-GQRA---ELE-----QLARAEIDAIRADIAGHPDTPDGPFFDVDEGDYAGLYFVE
-----W
DLQT-----VTDSAMARVREEAA
DTRY-----
----AIPVRATA--LPELG-NKARVVTAPPAHWGII-GDAMR-----KVLWP-LL
ETDPRIDLSGRRSLDGAASQFH-DQVVK-----SLRGAA-----GQ-----
----WM-YSA-----DLTAATDLMPE--NVILALWHGVLHG-LG---IP-
-EESF--FARAGDKI-LG---CVDV--S-Y-PDLAA-----
--PGE--KP-----IVVRSMRGCMMGLNLSWFLNLNLYNLAI----V---
DIA--C-L-----GGIAALTDPDQDGLDEAEVRRIVGL-----
-----APA--IVRGDD--LAAA--LTERQA--TAYEELIAA-TGGEANRAKSY
RS--HSA-FVLAEKS----F-----L---VDREVR--PLS-G
GQKRGFISFPAGLAVDAPMLA-----RDP-----
RALEALGF-----
---GGPLTGKDLGGATWTDAAVEVVSCLTALQDIPVRHLIP--APSKEGL-----
PVYVSLP---AAAADVLYAA----EEEGSP-----LFPGMC---RAVLSVNSEYVQ-R
Y-RDY-RIPLL--LPREV--GGAG--F--PH-----P-GGFSK-----
----ALASGG-RGHW----LRATLRVTTY--GVEQRARRHL-----DE-DV-
----WRVDGS-----NSDRRAAARLIRAR-----EAAAMNSGQLG-----
-----NMVPVPL-EDEV
-----TREVAFESLWR-----DLFCPADRKERCTR
RDRGVRLATISKRLSKADM---AARLAFNDRFLV--RS-----LV
RAEY-----QADVAMARGAESVYVPLSERDPRR-----LMPLSDDHPSLV
GLHRRRYR-----AGRASSHGRGNGCGGVSQVG
GVP-----DNWES--RLDDDFVYGGLDDNEAGISGHCLPKRSTLERAVRSRGQ
GAVAVRTLRSGLMVLADSIHYTESRSKRRGKPPTS-----
>KF298275.1

-----GVMIRSRRKRNHQRWRLP-----STLVLAPEG
RITP-----RAHSQPNEAGGAQG-----AAQSNVAQVTAERAANKPTQ
-----ANSQRTRHARAPKPDPEQM
GQ-----ATWD-----AFRLAIWALIPSLQKKGLALKG-----PR
NLATVRGIQSVHK-WISVSMAR-----SGPEWTAKQVK-----EYSNYCRARSL---
-----RDARTPCPRGFP-VKA-----FEVGLIKAAISDPN--GGR
PA-LAQLARLGRA---MP--VAT--DRVGRQALEKHRET-LSRATDVP----PA-----
----LIARFGAWAEQ---WARQ--QADKTGLTKEGSRLFQAVET--KFSNS--AAEAFS

AASG-GQQA---ELR-----ENYQRTMVDVQAEYLLATEGFVWDPEVPTLGAEDSDLMN
IT-----L
CARE-----EAEDYYL--RRHRW
RGVD-----
----GMPVRATI--VREQG-MKARVVTACPAWAVVC-GDACR-----KTLWP-LL
AADPRVDLSGARPVAEDLDRFN-DQIAL-----SLVGSi-----TP-----
----EF-FSA-----DLTAATDLMPF--EVSNALWAGLCRG-LD---LL-
-PTSQ--ISRIGHAL-LG---PVEV--S-Y-PDLAA-----
--KGE--PI-----PRILSKQGCMMLPLSWTILNLYNLAV-----A---
DFA-----VVAPSVVRVGV-----
-----APV--IARGDD--LVAA--YLPEEA--DRYTQLLRE-SGGEVNVLKSF
RS--SDS-FVLAERT----F-----KVTLRD--AVPLP
VHRRGFELRARN-----THA-----
PLLTQVPQ-----PSGTSRSDP
NVLGGGPAAESILGKTRETRGKVVSIMFDDCPLRTL VGKG--PGYAAGA-----VI
PAYISVP---SAASASLA-----EFEGKR-----FYPALA---QGLLSVHKSLVG-E
F-RRS-AIPVF--YPREL--GGGG--F--PH-----P-RGFGH-----
----AVCSAG-LLGR----KRAAFAMTC-----FTHKARK-----KHGLGDDP-
----WASRHR-----LAELERARQQLSAE-----EARAQTRAASEVPGAPRV
IPDDWEPKV-----WD-QIKDLEPPRGA-RIPV
PMEDAV-----VQRAAHGETWA-----
----SMILQEGVGQRDRYP---SLGTIA--KRLQA---IRKVTAERKFSNRFLPKRGKLS
RERL-----LERVAMARGGETRYVPYVDFGGGG-----MV
LVCRGFDT-----YAAARYEPES
Q-----PERYP--ERRQKL R PGTEGEEMVEVPVAVRTTRAGLGLADVAFVR
HGRVKRVKRRLLPSQFPR-----
>KP642119.1

-----MVLSGTCAQPNPSGVASA-----VKAVASPSV
GASA-----KLRKGPVVGTDRI-----PKSSLSSPRGHVGRKPRETV
-----KPKPDPLQQQ
AQ----RAWD-----VWRLALFALMPSLQKKGLDVRKAKPAKKPV----TA
QHATQRAIQVLYK-WLASSFVR-----SGPEWTCKQIK-----EYANYCRSRL---
-----GDERTDAPRGFP-VKG-----FEAGFIKACIMDPN--AGG
TA-LAQLARVGRA---MP--MGT--TRVAVAALRKHRDT--LTSTTVVP-----EE-----
----TQVNLRLWAER---WTRD--RLAAGARAREGTAT-----SFSRA--ASASVS
AADG-GQLA---ELR-----QLPRFVDHRQMLRDLLLEE QGFDLDSDD-----
-----I
FLSA-----TCDEGDD--LQLLA
DCAI-----ETARDRCMEGS-
----PIPMVATT--VLELG-MKARVVTKPPAWAVVA-GDACR-----QSVWP-LL
ESDRRIDLSGARPNTESLDRFH-DNLAH-----SLVGAATP-----
----QF-FSA-----DLTAATDDAPL-----
-----RLLGRYL-LG---PVTV--G-Y-PTLAAIEEHDR-----L
YQAGE-----PTEIVSVRGCMMLPLSWTLLNLYNLAV-----A---
DMA---C-----SPVGVVQVGV-----

-----APA--IARGDD--LVAA--IPPGEA--DRYEELIAL-TGGEANRLKSF
RS--ATA-FVLAERT-----FRVETQEIPDRVR
GLPRGWRW-----RRVSRFGDPALP
PLLTREHQ-----
--PGLRGQPSRNLEGA----REVVGVRMSVDLPLRALIAGS---ASFAGG-----AAV
PTYISLP--PAATSCLS-----EFEGTR-----FYPAVA---RGLLSVHRDLVS-E
M-RRS-AVPLF--YPREL--GGGG--F--PH-----P-GGFGA-----
----GLASAG-LLGL----QRAGLALTS-----HGWAARK-----KAALLE-DP-
----WVPRRQ-----NAQLNRAREQLLAS-----ERRAWAQSVLKRGETVAS
LEAAGLAERAKELKL-----ALATP-VPEYERVPRGA-RIAV
PLEDET-----IRQAARAGLWE-----
DAFLSPTVDDGGRSKTKYP--SLGDIARRLKRIR----EVTSKCKFDPRFVPNKDRVQ
RDSF-----FEKLRMLRGSETVLPVDSRPVRV-----TVWD
PARHRADP-----ELDNDDCGY--
-----PGRRR--LLRALDREEGGDPELVPDPVPLLQRGAGATLGDLVVFR
PAKSRHGPKSRAGTGARREKRSR RTPQHP-----
>KP642120.1

-----MLATGVMAKPQRPKVPNRWFLPKPSVGASAKLRKGPVV
GTDR-----IPKSSLSSPRGHVGRKPRKT
-----VKPKPDPLQGQ
AQ----RAWD-----VWRLALFALMPSLQKKGLDARKAKLAKKPV----TA
QRATQRAIQVLYK-WLASSFVR-----SGPEWTCKQIK-----EYANYCRSRL---
-----GDERTDAPRGFP-VKG-----FEAGFIKACIMDPNAGGTA
---LAQLARVGRA--MP--MGT--TRVAVAALRKHRDT-LTSTTVVP----EG----
----TQVNLRLWAER--WTRD--RLAAGARAREGTTT-----SFSRA--ASASVS
AADG-GQLA--ELR-----QLPRFVDHRQMLRDLLLEEQGFDLGDGDD-----
-----I
FLSA-----TCDEGDD--LQLLA
DCAI-----ETARDR
CMEGSPIPMVATT--VLELG-MKARVVTKPPAWAVVA-GDACR-----QSVWP-LL
ESDRRIDLSGARPNTESLDRFH-DNLAH-----SLVGAA-----TP-----
----QF-FSA-----DLTAATDLMPF--SVSWSLWDGLCDG-LG---AA-
-ADAP--LRLGGRYL-LG---PVTV--G-Y-PTLAAMEEHDR-----L
YQAGE--PT-----EIVSVRGCMMLPLSWTLLNLYNLAV----A---
DMA--C-----SPLGVVQVGV-----
-----APA--IARGDD--LVAA--IPPGEA--DRYEELIAL-TGGEANRLKSF
RS--ATA-FVLAERT-----F-----R---VETQEI--PDRVR
GLPRGWRW-----RRVSRFGDPALP
PLLTREHQ-----
--PGLKGQPSRNLEGA----REVVGIRMSVDLPLRALIAGS---ASFAGG-----AAV
PTYISLP--PAATSCLS-----EFEGTR-----FYPAVA---RGLLSVHRNLVS-E
M-RRS-AVPLF--YPREL--GGGG--F--PH-----P-GGFGA-----
----GLASAG-LLGL----QRAGLALTS-----HGWAARK-----KAALLE-DP-
----WMPRRQ-----NAQLNRAREQLLAS-----ERRAWAQSVLKRGETVAS
LEAAGLAERAKELKL-----ALATP-VPEYERVPRGA-RVAV

1
2
3 PLEDET-----IRQAARAGLWE-----
4 DAFLSPAVDDGGRSKTKYP---SLGDIARRLKRIR----EVTSKCKFDPRFVPNKDRAQ
5 RDSF-----FEKLRMLRGSETVLVPVDSRPVRV-----TVWD
6 PARHRADP-----ELDNDDCGY--
7 -----PGRRR--LLRALDREEGGDPELVPDPVPLLQRGAGATLGDLVVFR
8 PAKSRHGPKSRAGTGARREKRSR RTPQHP-----
9 >KF298284.1
10
11 -----
12 -----GVMAKPQRLFRSQRRWYLP-----AVLSLEERS
13 GVTMPLGTFPVKTGKVAKDGPCHP---VPQGS-----AKRANRHPNSPPTRRPKPER
14 -----RARTQRTPTQPGVELQRQ
15 AQ----HIWD-----TFRLTIWVLLPSAQKKGIA-----VS
16 AKATIRSIKGVYD-WIRNSMIR-----SGVEWTSKAIK-----EYANYCRARAL---
17 -----GDDRVRGRPRGFP-VRT-----FERGYIAAAVKCPG--QGR
18 LH-LAQIARIGRA---MP--KGT--RNVGISSLRKHRGV-LSQ--PME---TDKD-----
19 ----LLENLRSWATQ---WAQE--RVSERGTVVNPSV-----SFSRS--ASATVS
20 AMSG-GQLA--ELS-----QLESVKAHHRLIRELLLEEQGFENPAGTSDDS-----
21 -----I
22 GLLT-----VADEVAD--LQILA
23 DCAI-----SDALDAAREQI-
24 ----PIPMTATV--ISELG-MKARVVTKPPAWAVVA-GNACR-----KTVWP-LL
25 EADRRIDLSGLRPTAEVLDEFH-DNLAH-----SLVGAR-----TT-----
26 ----QF-YSA-----DLTAATDLMPF--DVSRALWNGLCDG-LG---AS-
27 -PSAP--LRSLGSYL-LG---PVRV--S-Y-PDLAALKERDR-----L
28 FKAGE--ET-----EILSSRGCMMGLPVSWTVLNLYNLAM----A---
29 DMA--C-----TPAGVPTLLGV-----
30 -----APA--VARGDD--LVAA--IPEFAA--DRYENLIAA-TGGEANRLKSF
31 RS--TFA-FVLAERT----F-----EVG-VKSVKDLTALTRR
32 GFKVRQVSRAP-----RTA-----
33 PLLCRFDE-----
34 ---LGPQQIPAFIQEM--DRPVEVVALRMSCDLPIRSLLGGM---AGFAGG-----EAV
35 PNYVAIP---SAAAACLA-----EFEGTR-----LYGAIC---HGLLSVHRDLVK-E
36 M-RRS-AIPLF--YPRVL--GGGG--F--PH-----P-RGFAA-----
37 ----AVASAG-PLGL----QRAGLYLTA-----SGYKARR-----KVGIGH-DP-
38 ----WVPLAD-----SRRMDVARDRLVAT-----DRRLWAKSVLDRLPDLLD
39 LEDPYLDKRLSERAE-----ALSVP-CVGFETVPRGA-RVAV
40 PLDDAV-----ISEAATIAAWE-----
41 ----DAVLGPCPSAERRYYP---TLGGIAKRLKAIR----QKTVEARFPMKFVPMKRMG
42 REGF-----FNRVKLLLASSETVLPYSNRNTAR-----VTVFD
43 PERHCTVQ-----EPEDEGRVEGDLG
44 L-----PIRRR--ELRERHGGGVRLRDGATLGDAVIRHVPQKRRRHRRT
45 PGPSNRNVL-----
46 >KF298276.1
47 -----
48 -----GVMAKPQRLFRSQRRWYLP-----AVLSLEGRS
49 GVTMPLGTLPAKAGRVAKVGPCHPD PQGS-----AKRANQHPNSPSTRPNPER
50
51
52
53
54
55
56
57
58
59
60

-----RARAQRTPTQPGVELQRQ
AQ----HIWD-----TFRLTIWVLLPSAQKKGIA-----VS
AKATIRSIKGVYE-WIRNSMIR-----SGVEWTSKAIK-----EYANYCRARAL---
-----GDDRVRGRPGFP-TGL-----FERGYIAAALKCPG--QGR
LH-LAQLARVGRA---MP--KGT--RNVGISSLRKHRGV-LSQPMETD-----KD-----
-----LLENLRSWATQ--WARE--RVSESGTVVNPSV-----SFSRS--ASATVS
AMSG-GQLA---ELS-----QLESVKAHHRLIRELLLEEQGFENPAGTSDDS-----
-----I
GLLT-----VADEVAD--LQILA
DCAI-----SDALDA
AREQIIPMTATV--ISELG-MKARVVTKPPAWAVVA-GNACR-----KTVWP-LL
EADRRIDLSGLRPTAEVLDEFH-DNLAH-----SLVGAR-----TT-----
-----QF-YSA-----DLTAATDLMPF--DVSRALWNGLCDG-LG---AS-
-PSAP--LRSLGSYL-LG---PVRV--S-Y-PDLAALKERDR-----L
FKAGE--ET-----EILSSRGCMMGLPVSWTVLNLNLYNLAM----A---
DMA---C-----TPMGVPTLLGV-----
-----APA--VARGDD--LVAA--IPEFAA--DRYENLIAA-TGGEANRLKSF
RS--TFA-FVLAERT----F-----EVGVKSVKDL-TALTRR
GFKVRQVSRAP-----RTA-----
PLLCRFDE-----
LGPQQIPAFIQEMDRP----VEVVALRMSCDLPIRSLLGGM---AGFAGG-----EAV
PNYVAIP--SAAAACLA-----EFEGTR-----LYGAIC--HGLLSVHRDLVK-E
M-RRS-AIPLF--YPRVL--GGGG--F--PH-----P-RGFAA-----
-----AVASAG-PLGL----QRAGLYLTA-----SGYKARR-----KVGIGH-DP-
-----WVPLAD-----SRMDVARDRLVAT-----DRRLWAKSVLDRLPDLLD
LEDPYLDKRLSERAE-----ALSVP-CVGFETVPRGA-RVAV
PLDDAV-----ISEAATIAAWE-----
----DAVLGPCPSAERRYPP--TLGGIAERLKAIR----QKTVEARFPMKFVPMKRMG
REGF-----FNRVKLLLASETVLVPYSNRNTAR-----VTVFD
PERHCTVQ-----EPEDEGRVEGDLG
L-----PIRRR--ELRERHGGGVRLRDGATLGDVAVIRHVPQKRRRHRRT
PGPSNRNVL-----
>MF176344.1

-----MSLPERVGGTIPRTPAGSREP-----ASAQVGRPQ
PSPK-----GLARK---TNHTQ-----IPPTIVDPKSSPRRARPPKK
-----ACAPGKATTKPPASAQQKQ
AQ----HIWD-----AFRLSLWVLLPRLQKHGLS-----TK
AVATIEAIKALYS-WIGNSVAR-----SGVEFTAKQVK-----EYANYCRARAL---
-----KAQQCEPPRGFP-IRA-----FERGYIRAMLACPG--QGR
LH-LAQLARVGRA---MP--IGT--AKVAIASLRKHREV-LSQ--PMD---TEPA-----
-----LVEKLRAWAEI--WTRE--RLGAGGRTLTPNAA-----HFSRS--ASATVS
ALKG-GQLT---ELR-----QLPAVAEHYALIQELLGDQLDPV-TGEPFVWDPYTFDAD
GGA-----I
GLLT-----SASEADD--LQILA
DCAL-----RTATEHAADNT-

-----PIPMTATV--ISELG-MKARVVTKPPAWAVVA-GDACR-----KTVWP-LL
EGDRRIDLSGVRPTAEVLDTFH-DNLAH-----SLVGAR-----TT-----
-----QF-YSA-----DLTAATDLMPF--DVSRAIWDGLCDG-LG---AT-
-ATAP--LRTLGRYL-LG---PVQV--S-Y-PDLRALPASSK-----L
YVAGE--RV-----ECLSKRGCMMLPVSMTVLNLYNLAM----A---
DLA---C-----TPEGSPVLVNV-----
-----APA--IARGDD--LVAA--IPAGEA--TRYEDLIAA-TGGEANRLKSF
RS--ADA-FVLAERT----F-----EVGVLRPNV--ELRQR
GYVVRRTY-----RTA-----
PLLAQFDS-----
AELHGGTGDARRLGGA----PEVVAIMACDVPIRSLGSG--PRTAGA-----NPV
PDYVSIP--PAAAACLA-----EFEGTR-----LYRSVA--EGLMSVHRGLVA-D
L-RRS-AVPLF--YPREL--GGAG--F--PH-----P-KGFAA-----
----AVASAG-ELGL----KRALGLTT----AGHKAQQ-----KVGLLS-CP-
-----WSPKLT-----DRRAKEARRRLIAS-----DQRAWANSVLARHPDLPR
GGSAWTDRAAGELRK-----ALSIP-ARGFETVPRGA-RIAV
ALEDV-----IREVAAAAAWE----DAFLGPGARGWSGVPPPMQ
EPSRKRSARRKEESSGMYP--SLGEVARRLKTIR----SQTQAARFNRTFLPKRAVQE
RARF-----FDRLKALRELETVLVPAQSAGARL-----VVFD
PVTHRVSE-----SAGEIDGGEGRMG
F-----PLWRH--TLQERKKGKRGAGQGEAAVPPVATTRSGLSLGDVMAVR
RLPRRPRGVRRLPNPGGAHSRNV-----
>MF176385.1

-----MSLPERVGGTIPRTPAGSHEP-----ASAQVGRPQ
PSPK-----GLARK--TNHTQ-----IPPTIVDPKSSPRRVHPPKK
-----VCAPGKATTKPPASAPQKQ
AQ----HIWD-----AFRLSLWVLLPRLQKHGLS-----TK
AVTTIEAIKALYS-WIGNSVAR-----SGVEFTAKQVK-----EYANYCRARAL---
-----KAQQCEPPRGFP-IRA-----FERGYIHAMLACPG--QGR
LH-LAQLARVGRA--MP--IGT--TKVAIASLRKHREV-LSQ--PMD--TEPA-----
----LVEKLRAWAEI--WTRE--RLGAGGRTLTPAA-----HFSRS--ASATVS
ALKG-GQLT--ELR-----QLPAVAEHYALIQELLGDQLDPV-TGEPFVWDPYTFDAD
GGA-----I
GLLT-----SASEADD--LQILA
DCAL-----RTATEHAADD-
-----PIPMTATV--ISELG-MKARVVTKPPAWAVVA-GDACR-----KTVWP-LL
EGDRRIDLSGVRPTAEVLDTFH-DNLAH-----SLVGAR-----TT-----
-----QF-YSA-----DLTAATDLMPF--DVSRAIWDGLCDG-LG---AT-
-ATAP--LRTLGRYL-LG---PVQV--S-Y-PDLRALPASSK-----L
YAAGE--RV-----ECLSKRGCMMLPVSMTVLNLYNLAM----A---
DLA---C-----TPEGSPVLVNV-----
-----APA--IARGDD--LVAA--IPAGEA--TRYEDLIAA-TGGEANRLKSF
RS--ADA-FVLAERT----F-----EVGVLRPNV--ELKQR
GYVVRRTY-----RTA-----
PLLAQFDS-----

1
2
3 AELHGGTGDARRLGGA----PEVVAIMACDVPIRSLG---PRTAGA-----NPV
4 PDYVSIP---PAAAACLA-----EFEGTR-----LYRSVA---EGLLSVHRGLVA-D
5 L-RRS-AVPLF--YPREL--GGAG---F---PH-----P-KGFAA-----
6 ----AVASAG-ELGL----KRALGLTT-----AGHKAQQ-----KVGLLS-CP-
7 -----WSPKLT-----DRRAKEARRRLIAS-----DQRAWANSVLARHPDLPR
8 GGSAWTDRAAGELRK-----ALSIP-ARGFETVPRGA-RVAV
9 ALEDAV-----IREVAAAAAWE----DAFLGPGARGWSGVPPPMQ
10 EPSRKRSARRKEESSGMYP---SLGEVARRLKTIR----SQTQAARFNRTFLPKRAVQE
11 RARF-----FDRLKALRELETVLVPAQSAGARL-----VVFD
12 PVTHRVSE-----SAGEIDGGEGRMG
13 F-----PLWRH--TLQERKGKRGAGQGEAAVPPVATTRSGLSLGDVMTVR
14 RLPRRPRGVRRLPNPGGAHSRN-----
15 >MF176278.1
16
17
18

19 -----MSLPERVGRTPPTPAGSHEP-----ASAQVGRPQ
20 PSPK-----GLARK---TNHTQ-----IPPTIVDPKSSPRRARPPKK
21 -----VCAPGKATTKPPASAQQKQ
22 AQ----HIWD-----AFRLSLWVLLPRLQKHGLS-----TK
23 AVATIEAIKALYS-WIGNSVAR-----SGVEFTAKQVK-----EYANYCRARAL---
24 -----KAQQCEPPRGFP-IRA-----FERGYIHAMLACPG--QGR
25 LH-LAQLARVGRA--MP--IGT--TKVAIASLRKHREV-LSQ--PMD---TEPA-----
26 ----LVEKLRAWAEI--WTRE--RLGAGGRTLNPAA-----HFSRS--ASATVS
27 ALKG-GQLT---ELR-----QLPAVAEHYALIQELLGDQLDPV-TGEPFVWDPYTFDAD
28 GGA-----I
29 GLLT-----SASEADD--LQILA
30 DCAL-----RTATEHAADNT-
31 ----PIPMATATV--ISELG-MKARVVTKPPAWAVVA-GDACR-----KTVWP-LL
32 EGDRRIDLSGVRPTAEVLDTFH-DNLAH-----SLVGAR-----TT-----
33 ----QF-YSA-----DLTAATDLMPF--DVS RAMWDGLCDG-LG---AT-
34 -ATAP--LRTLGRYL-LG---PVQV--S-Y-PDLRALRASSK-----L
35 YVAGE--RV-----ECLSKRGCMMLPISWTVLNLYNLAM-----A---
36 DLA--C-----TPEGSPVLVNV-----
37 -----APA--IARGDD--LVAA--IPAGEA--TRYEDLIAA-TGGEANRLKSF
38 RS--ADA-FVLAERT----F-----EVGVLRPNV--ELRQR
39 GYVVRRTY-----RTA-----
40 PLLAQFDS-----
41 AELHGGTGDARRLGGA----PEVVAIMACDVPIRSLG---PRTAGA-----NPV
42 PDYVSIP---PAAAACLA-----EFEGTR-----LYRSVA---EGLMSVHRGLVA-D
43 L-RRS-AVPLF--YPREL--GGAG---F---PH-----P-KGFAA-----
44 ----AVASAG-ELGL----KRALGLTT-----AGHKAQQ-----KVGLLS-CP-
45 -----WSPKLT-----DRRAKEARRRLIAS-----DQRAWANSVLARHPDLPR
46 GGSAWTDRAAGELRK-----ALSIP-ARGFETVPRGA-RVAV
47 ALEDAV-----IREVAAAAAWE----DAFLGPGARGWSGVPPPMQ
48 EPSRKRSARRKEESSGMYP---SLGEVARRLKTIR----SQTQAARFNRTFLPKRAVQE
49 RARF-----FDRLKALRELETVLVPAQSAGARL-----VVFD
50 PVTHRVSE-----SAGEIDGGEGRMG
51
52
53
54
55
56
57
58
59
60

1
2
3 F-----PLWRH--TLQERKGKRGAAQGEAAVPVPVATTRSGLSLGDVMTVR
4 RLPRRPRGVRRLPNPGGAHSRNV-----
5 >MF176306.1
6
7 -----
8 -----MSLPERVGRTIPRTPAGSHEP-----ASAQVGRPQ
9 PSPK-----GLARK---TNHTQ-----IPPTIVDPKSSPRRAHPPKK
10 -----VCAPGKATTKPPASAAQKKQ
11 AQ----HIWD-----AFRLSLWVLLPRLQKHGLS-----TK
12 AVATIEAIKALYS-WIGNSVAR-----SGVEFTAKQVK-----EYANYCRARAL---
13 -----KAQQCEPPRGFP-IRA-----FERGYIHAMLACPG--QGR
14 LH-LAQLARVGRA---MP--IGT--TKVAIASLRKHREV-LSQ--PMD---TEPT-----
15 ----LVEKLRAWAEI--WTRE--RLGAGGRTLNPAA-----HFSRS--ASATVS
16 ALKG-GQLT---ELR-----QLPAVAEHYALIQELLGDQLDPV-TGEPFVWDPYTFDAD
17 GGA-----I
18 GLLT-----SASEADD--LQILA
19 DCAL-----RTATEHAADNT-
20 ----PIPMATV--ISELG-MKARVVTKPPAWAVVA-GDACR-----KTVWP-LL
21 EGDRRIDLSGVRPTAEVLDTFH-DNLAH-----SLVGAR-----TT-----
22 ----QF-YSA-----DLTAATDLMPF--DVS RAMWDGLCDG-LG---AT-
23 -ATAP--LRTLGRYL-LG---PVQV--S-Y-PDLRALPASSK-----L
24 YVAGE--RV-----ECLSKRGCMMLPISWTVLNLYNLAM----A---
25 DLA--C-----TPEGSPVLVNV-----
26 -----APA--IARGDD--LVAA--IPAGEA--TRYEDLIAA-TGGEANRLKSF
27 RS--ADA-FVLAERT----F-----EVGVLRPNV--ELRQR
28 GYVVRRTY-----RTA-----
29 PLLAQFDS-----
30 AELHGGTGDARRLGGA----PEVVAIMACDVPIRSLG--PRTAGA-----NPV
31 PDYVSIP--PAAAACLA-----EFEGTR-----LYRSVA---EGLMSVHRGLVA-D
32 L-RRS-AVPLF--YPREL--GGAG---F--PH-----P-KGFAA-----
33 ----AVASAG-ELGL----KRAGLALT-----AGHKAQQ-----KVGLLS-CP-
34 ----WSPKLT-----DRRAKEARRRLIAS-----DQRAWANSVLARHPDLPR
35 GGS AWTDRAAGELRK-----ALSIP-ARGFETVPRGA-RVAV
36 ALEDAV-----IREVAAAAAWE----DAFLGPGARGWSGVPPPMQ
37 EPSRKRSARRKEESSGMYP--SLGEVARRLKTIR----SQTQAARFNRTFLPKRAVQE
38 RARF-----FDRLKALRELETVLVPAQSAGARL-----VVFD
39 PVTHRVSE-----SAGEIDGGEGRMG
40 F-----PLWRH--TLQERKGKRGAAQGEAAVPVPVATTRSGLSLGDVMTVR
41 RLPRRPRGVRRLPNPSGAHSRNV-----
42 >MF176257.1
43 -----
44 -----MSLPERVGGTIPRTPAGSHEP-----ASAQVGRPQ
45 PSPK-----GLARK---TNHTQ-----IPPTIVDPKSSPRRVHPPKK
46 -----VCAPGKATTKPPASARQKKQ
47 AQ----HIWD-----AFRLSLWVLLPRLQKHGLS-----TK
48 AVATIEAIKALYS-WIGNSVAR-----SGVEFTAKQVK-----EYANYCRARAL---
49 -----KAQQCEPPRGFP-IRA-----FERGYIHAMLACPG--QGR
50
51
52
53
54
55
56
57
58
59
60

LH-LAQLARVGRA---MP--IGT--TKVAIASLRKHREV-LSQ--PMD---TEPA-----
 ----LVEKLRAWAEI---WTRE--RLGAGGRTLTPNAA-----HFSRS--ASATVS
 ALKG-GQLT---ELR-----QLPAVAEHYALIQELLGDQLDPV-TGEPFVWDPYTFDAD
 GGA-----I
 GLLT-----SASEADD--LQILA
 DCAL-----RTATEHAADNT-
 ----PIPMATATV--ISELG-MKARVVTKPPAWAVVA-GDACR-----KTVWP-LL
 EGD RRIDLSGVRPTAEVLDTFH-DNLAH-----SLVGAR-----TT-----
 ----QF-YSA-----DLTAATDLMPF--DVS RAMWDGLCDG-LG---AT-
 -ATAP--LRTLGRYL-LG---PVQV--S-Y-PDLRALPASSK-----L
 YVAGE--RV-----ECLSKRGCMMLPVSWTVLNLNLYLAM----A---
 DLA--C-----TPEGSPVLVNV-----
 -----APA--IARGDD--LVAA--IPAGEA--TRYEDLIAA-TGGEANRLKSF
 RS--ADA-FVLAERT---F-----EVGVLRPNV--ELRQR
 GYVVRRTY-----RTA-----
 PLLAQFDS-----
 AELHGGTGDARRLGGA----PEVV AIRMACDVPIRSLG GGG--PRTAGA-----NPV
 PDYVSIP---PAAACL A-----EFEGTR-----LYRSVA---EGLMSVHRGLVA-D
 L-RRS-AVPLF--YPREL--GGAG--F---PH-----P-KGFAA-----
 ----AVASAG-ELGL----KRA GLALT T-----AGHKAQQ-----KVGLLS-CP-
 ----WSPKLT-----DRRAKEARRRLIAS-----DQRAWANSVLARHPDLPR
 GGS AWTDRAAGELRK-----ALSIP-ARGFETVPRGA-RVAV
 ALEDAV-----IREVAAAAAWE----DAFLGPGARGWSGVPPPMQ
 EPSRKRSARRKEESSGMYP---SLGEVARRLKTIR----SQTQAARFNRTFLPKRAVQE
 RARF-----FDRLKALRELETVLVPAQSAGARL-----VVFD
 PVTHR VSE-----SAGEIDGGEGRMG
 F-----PLWRH--TLQERKGKRG AQGQEA AVPV PVATTRSGLSLGDVMTVR
 RLPRRPRGVRRLPNPGGAHSRNV R-----
 >MF176365.1

 -----MSLPERSVGTIPRTPAGSHEP-----ASAQVGRPQ
 PSPK-----GLARK---TNHTQ-----IPPTIVDPKSSPRRVHPPKK
 -----VCAPGKATTKPPAS AQKQ
 AQ----HIWD-----AFRLSLWVLLPRLQKHGLS-----TK
 AVATIEAIKALYS-WIGNSVAR-----SGVEFTAKQVK-----EYANYCRARAL---
 -----KAQQCEPPRGFP-IRA-----FERGYIHAMLACPG--QGR
 LH-LAQLARVGRA---MP--IGT--TKVAIASLRKHREV-LSQ--PMD---TEPA-----
 ----LVEKLRAWAEI---WTRE--RLGAGGRTLTPNAA-----HFSRS--ASATVS
 ALKG-GQLT---ELR-----QLPAVAEHYALIQELLGDQLDPV-TGEPFVWDPYTFDAD
 GGA-----I
 GLLT-----SASEADD--LQILA
 DCAL-----RTATEHAADNT-
 ----PIPMATATV--ISELG-MKARVVTKPPAWAVVA-GDACR-----KTVWP-LL
 EGD RRIDLSGVRPTAEVLDTFH-DNLAH-----SLVGAR-----TT-----
 ----QF-YSA-----DLTAATDLMPF--DVS RAMWDGLCDG-LG---AT-
 -ATAP--LRTLGRYL-LG---PVQV--S-Y-PDLRALPASSK-----L

YVAGE--RV-----ECLSKRGCMMLPVSWSVTLNLYNLAM----A---
DLA---C-----TPEGSPVLVNV-----
-----APA--IARGDD--LVAA--IPAGEA--TRYEDLIAA-TGGEANRLKSF
RS--ADA-FVLAERT----F-----EVGVLRPNV--ELRQR
GYVVRRTY-----RTA-----
PLLAQFDS-----
AELHGGTGDARRLGGA----PEVVAIRMACDVPIRSLLGGG---PRTAGA-----NPV
PDYVSIP---PAAAACLA-----EFEGTR-----LYRSVA---EGLMSVHRGLVA-D
L-RRS-AVPLF--YPREL--GGAG---F---PH-----P-KGFAA-----
----AVASAG-ELGL----KRAGLALT-----AGHKAQQ-----KVGLLS-CP-
----WSPKLT-----DRRAKEARRRLIAS-----DQRAWANSVLARHPDLPR
GGSAWTDRAAGELRR-----ALSIP-ARGFETVPRGA-RVAV
ALEDV-----IREVAAAAAWE----DAFLGPGARGWSGVPPPMQ
EPSRKRSARRREESSGMYP---SLGEVARRLKTIR----SQTQAARFNRTFLPKRAVQE
RARF-----FDRLKALRELETVLVPAQSAGARL-----VVFD
PVTHRVSE-----SAGEIDGGEGRMG
F-----PLWRH--TLQERKGKRGAGQGEAAVPVPVATTRSGLSLGDVMTVR
RLPRRPRGVRRLPNPSGAHSRNV-----
>KX883461.1

-----MSLPEGVSGAIPRTPAGSHEP-----ASAQVGRPQ
PGPE-----GLARK---TNHTQ-----IPPTIVDPKSSPRRAHPPKK
-----ARAPGKATTKPPASVKQKQ
AQ---HIWD-----AFRLSLWVLLPRLQKHGLS-----TK
AVATIEAIKALYS-WIGNSVAR-----SGVEFTAKQVK-----EYANYCRARAL---
-----RAQQCEPPRGFP-IRA-----FERGYILAMLACPG--QGR
LH-LAQLARVGRA---MP--IGT--TKVAIASLRKHREV-LSQ--PMD---TEPA-----
----LVEKLRAWAEI---WTRE--RLGAGGRTLTPAA-----HFSRS--ASATVS
ALKG-GQLT---ELR-----QLPAVAEHYALIQELLGDQLDPV-TGEPFVWDPYTFDAD
GGA-----I
GLLT-----SASEADD--LQILA
DCAL-----RTATEHAADNT-
----PIPMATATV--ISELG-MKARVVTKPPAWAVVA-GDACR-----KTVWP-LL
EGDRRIDLSGVRPTAEVLDTFH-DNLAH-----SLVGAR-----ST-----
----QF-YSA-----DLTAATDLMPF--DVS RAMWNLCDG-LG---AT-
-ATAP--LRTLGLYL-LG---PVQV--S-Y-PDLRALPASSK-----L
YVAGE--RV-----ECLSKRGCMMLPVSWSVTLNLYNLAM----A---
DLA---C-----TPEGSPVLVNV-----
-----APA--IARGDD--LVAA--IPAGEA--TRYEDLIAA-TGGEANRLKSF
RS--ADA-FVLAERT----F-----EVGVLRNSV--ELKQR
GYVTRRTF-----RTA-----
PLLAQFDS-----
AELHGGTGDARRLGDA----PEVVAIRMACDVPIRSLLGGG---PRTVGA-----NPV
PDYVSIP---PAAAACLA-----EFEGTR-----LYRSVA---EGLMSVHRGLVA-D
L-RRS-AIPLF--YPREL--GGAG---F---PH-----P-KGFAA-----
----AVASAG-ELGL----KRAGLALT-----AGHKAQQ-----KVGLLS-CP-

-----WSPKLT-----DRRAREARRRLIAS-----DQRAWANSVLARHPDLPR
GGSAWTDRAAGELRK-----ALSVP-AREFETVPRGA-RVAV
ALEDV-----IREVAAAAAWE-----DAFLGPGARGWSGVPPPTQ
EPSRKRSARQKEESLGMYP---SLGEVARRLKTIR----SQTQAARFNRTFLPKRAVQE
RARF-----FDRLKALRELETVLVPAQAAGARL-----VVFD
PVTHRVPE-----SAGEIDGGEGRMG
F-----PLWRH--TLQERKGKRGAGQGEAAVPVPVATTRSGLSLGDVMTVR
RLPRRPRGVRRLPNRS GAHSRNVR-----
>KX883537.1

-----MSLPEGVSGAIPRTPAGSHEP-----ASAQVGRPQ
PSPK-----GLARK---TNHTQ-----IPPTIVDPKSSPRRAHPPKK
-----ARAPGKATTKPPASVQQKQ
AQ---HIWD-----AFRLSLWVLLPRLQKRGLS-----TK
AVATIEAIKALYS-WIGNSVAR-----SGVEFTAKQVK-----EYANYCRARAL---
-----RAQQCEPPRGFP-IRA-----FERGYILAMLACPG--QGR
LH-LAQLARVGRA---MP--IGT--TKVAIASLRKHREV-LSQ--PMD---TEPA-----
----LVEKLRAWAEI---WTRE--RLGAGGRTL TNPA-----HFSRS--ASATVS
ALKG-GQLT---ELR-----QLPAVAEHYALIQELLGDQLDPV-TGEPFVWDPYTFDAD
GGA-----I
GLLT-----SASEADD--LQILA
DCAL-----RTATEHAADNT-
----PIPM TATV--ISELG-MKARVVT KPPAWAVVA-GDACR-----KTVWP-LL
EGDRRIDLSGVRPTAEVLDAFH-DNLAH-----SLVGAR-----ST-----
----QF-YSA-----DLTAATDLMPF--DVS RAMWNGLCDG-LG---AT-
-ATAP--LRKLGLYL-LG---PVQV--S-Y-PDLSALPASSK-----L
YVAGE--RV-----ECLSERGCMMGLPISWTVLNLYNLAM----A---
DLA---C-----TPEGSPVLVNV-----
-----APA--IARGDD--LVAA--IPAE EA--TRYEDLIAA-TGGEANRLKSF
RS--ADA-FVLAERT----F-----EVGVLRPNV--ELKQR
GYVTRRTY-----RTA-----
PLLAQFDS-----
AELHGGTGDARRLGDA----PEVV AIRMACDVPIRSLLGGG---PRTVGA-----NPV
PDYVSIP---PAAAACLA-----EFEGTR-----LYRSVA---EGLMSVHRGLVA-D
L-RRS-AIPLF--YPREL--GGAG--F--PH-----P-KGFAA-----
----AVASAG-ELGL----KRAGLALT-----AGHKAQQ-----KVGLLS-CP-
----WSPKLT-----DRRAREARRRLIAS-----DQRAWANSVLARHPDLPR
GGSAWTDRAAGELRK-----ALSVP-ARGFETVPRGA-RVAV
ALEDV-----IREVAAAAAWE-----DAFLGPGARGWSGVPPPMQ
EPSRKRSARQKEESLGMYP---SLGEVARRLKTIR----SQTQAARFNRTFLPKRAVQE
RARF-----FDRLKALRELETVLVPAQAAGARL-----VVFD
PVTHRVSE-----SAGEVDGGEGRMG
F-----PLWRY--TLQERKGKRGAGQGEAAVPVPVATTRSGLSLGDVMTVR
RLPRRPRGVRRLPNRS GAHSRNVR-----
>KX883538.1

1
2
3 -----MSLPEGVSGAIPRTPAGSHEP-----ASAQVGRPQ
4 PSPK-----GLARK---TNHTQ-----IPPTIVDPKSSPRRAHPPKK
5 -----ARAPGKATTKPPASVQQKQ
6 AQ---HIWD-----AFRLSLWVLLPRLQKRGLS-----TK
7 AVATIEAIKALYG-WIGNSVAC-----SGVEFTAKQVK-----EYANYCRARAL---
8 -----RAQQCEPPRGFP-IRA-----FERGYIHAMLACPG--QGR
9 LH-LAQLARVGRA---MP--IGT--TKVAIASLRKHREV-LSQ--PMD---TEPA-----
10 ----LVEKLRAWAEI---WTRE--RLGAGGRTLTPAA-----HFSRS--ASATVS
11 ALKG-GQLT---ELR-----QLPAVAEHYALIQELLGDQLDPV-TGEPFEWDPYTFDAD
12 GGA-----I
13 GLLT-----SASEADD--LQILA
14 DCAL-----RTATEHAADNT-
15 ----PIPMATATV--ISELG-MKARVVTKPPAWAVVA-GDACR-----KTVWP-LL
16 EGDRRIDLSGVRPTAEVLDAFH-DNLAH-----SLVGAR-----ST-----
17 ----QF-YSA-----DLTAATDLMPF--DVS RAMWNGLCDG-LG---AT-
18 -ATAP--LRKLGLYL-LG---PVQV--S-Y-PDLSALPASSK-----L
19 YVAGE--RV-----ECLSERGCMMLPVSWTVLNLYNLAM----A---
20 DLA--C-----TPEGSPVLVNV-----
21 -----APA--IARGDD--LVAA--IPAEAA--TRYEDLIAA-TGGEANRLKSF
22 RS--ADA-FVLAERT----F-----EVGVLRPNV--ELKQR
23 GYVTRRTY-----RTA-----
24 PLLAQFDS-----
25 AELHGGTGDTRRLGDA----PEVVAIMACDVPIRSLLGGG---PRTVGA-----NPV
26 PDYVSIP--PAAAACLA-----EFEGTR-----LYRSVA--EGLMSVHRGLVA-D
27 L-RRS-AIPLF--YPREL--GGAG--F--PH-----P-KGFAA-----
28 ----AVASAG-ELGL----KRAGLALT-----AGHKAQQ-----KVGLLS-CP-
29 ----WSPKLT-----DRRAKEARRRLIAS-----DQRAWANSVLARHPDLPR
30 GGS AWTNRAAGELRK-----ALSVP-SRGFETVPRGA-RVAV
31 ALEDAV-----IREVAAAAAWE-----DAFLGPGARGWSGVPPPMQ
32 EPSRKRSARQKEESLGMYP---SLGEVARRLKTIR----SQTQAARFNRTFLPKRAVQE
33 RARF-----FDRLKALRELETVLVPAQAAGARL-----VVFD
34 PVTHR VSE-----SAGEIDGGEGRMG
35 F-----PLWRH--TLQERKGKRG AQGQEA AVPVVATTRSGLSLGDVMTVR
36 RLPRRPRGVRRLPNRSGAHSRNV-----
37 >KX883539.1
38
39 -----
40 -----
41 -----
42 -----
43 -----
44 -----
45 -----
46 -----
47 -----
48 -----
49 -----MGREETVKRVK-----QYAEYCRARSID--
50 -----ARQAEAGPPRGFP-FHA-----LEGGSFKALLRCPV--QGP
51 AA-LAQVARLGRS---MP--LPT--HKQQGRALIEHCNE-LTAGPHEE---IPVE-----
52 ----LLDAFYWTRT---WGAG--KGAALRDEVTV-----NFSSA--AAESVS
53 RARG-GQRE---ELR-----LLAAPHFEAIREALAI DDIDFRVEEGDYAGI-----
54 -----Y
55
56
57
58
59
60

SEED-----DLEIITQAAIERVR
ADIR-----AGR-
----RPDVRAAV--VPELG-AKSRIVTAGEAHWVVI-GDAIR-----KCLWP-AL
EAEPRIDLSGRRDVAGTAAHLH-DEVAK-----ALTGAE-----AV-----
----EF-YSA-----DLTAATDLMPH--DLIQAIWRGLMEG-LG---VS-
-DDDL--MFTAGLAL-LG---PVNL--H-H-PDLD----D-----
LDRPG--KY-----AYGVTHKGCMMGFPLSWYILNLYNLAC----A---
DLA---L-----SAERGRFLAEVAPRLVGA-----
-----APA--VVRGDD--LCSA--HTPAEA--DRYERIIRA-TGGRANLSKSY
RS--RKG-FILAERT----F-----L---VTTART--KCSAA
RSRLGTVVVRPGVH-----PRA-----
PLLTRAPD-----AFES
GALGYRSIDPRSRRRG---AATVVGLTMLS DVPVRHLMPS-----DS-----EGL
PAYVTLP--PAASAVVD-----EIKDIK-----RREAIC---YGILSVNAPLVR-Q
F-NRF-KIPF--YPREL--GGAG---F---PH-----P-KGFGS-----
----AIRSAG-GMGH----LRATLAMTTFTRRARQAYQLDR-----DP-
----WICDRQ----VEFERSRAQTGLIAE-----ERRAAESGQLG-----
-----LATSVP-LEDAVAARAAEGQLWA
EVMLPQSRE-----LASARRGRRGPRLSSVRKGLSDVFNALAKVHAKFD
PKFLANPRIDKETWMDRVA---
MAKGSSTVYRVVRPSGRRVVMHGIRDTHREIRAEREML
KGGW-----TFGGQRDLRQNTAAGLEGGEIA-----VP
VPTQRAGA-----
-----TLWDV--LSVSCKRARKRKKGNPRL-----

>GENC01006608.1

-----RVSQGS DKA WFLP-----DNVRWTQIV
GNLR-----PPGESAGTAEMVESLPASQK
-----SKVGRKKEKFPGRATKAA
GQ----RIWN-----AFRLAFWAIAPSAQRKGLNPRS-----QR
QARSIQAFRTLEK-WIAISLLK-----CGTEWTCKAVK-----DYSEYCRARAL---
-----RAKQCVRPREFP-VTG-----FEIGLVRGAINDPLAGAEA
---LAQLARLGRA--MP--QGT--DKQVIRALEKHRDT-LLGETRVD----EE-----
----TLDRYHAWAER--WSSD--RRRFAPERVAV-----KFSDS--AAESVS
RADG-GQKA---DLL-----ENARPHFEALRAALAESQGW DAPDDYIGLYSEED-----

-----DVETIAQMALETVR
AAVR-----
--SGEPV PARASA--IPEYG-SKVRVVTACPGPFVVA-GDACR-----KIVWG-LL
EADPRVDLSGRRPLGDSISQFQ-DEVAR-----SLVGAV-----TT-----
----EY-YSA-----DLTAATDLMPF--QLSAAIWTGLCDG-LG---ES-
-QDSD--LRSIGLHL-LG---PVQV--R-Y-PDLK-----EG-----S
VEEGF--RV-----PTIISQQGCMMGLPLSWTVLNVFNLAT----A---
DLA---I-----GAPEDRVTDLIGR-----
-----APA--VARGDD--LAAA--FLPEEA--DRYEALISA-TGGRVNLSKSF

1
2
3 RS--SIG-FVLAERT----F-----MVKRSV--DRYQR
4 RAPRQAVLIQPHRPEGIP-----QNA-----
5 PILAQDPS-----
6
7 --APTMSGVGRDVRGH-----RTESLRMLPDCSLKGLLGGS-----AST-----DGA
8 PRYVTLP--VSAASSLY-----EFEGTR-----FYGALC--RGLLSVHKDLLG-T
9 F-REF-GVPKF--YPREL--GGGG--F--PH-----P-RGYLR-----
10 ----AMESAG-GIGH----LRATLNITRFTESSRRAAALHM-----DP-
11 -----WRTERD-----SRTLERARSQLIAE-----ESRAREAGLLG-----
12 -----SF-VSKFQEEAE-----
13 -----VTRAAQGSWE-----TTLLSQKKRDEK
14 FPRLKDVRHRMLEITEKTKAIKFNRRFLLRRQLHR-----DEYLQIASHARGA
15 PTRY-----VRMERAPGLVALRDIVPGQQA-----PN
16 SFQTGTVP-----QHGNARGNRG
17 ERAEMSLVPTLRSRASLGQVARILPARASRKKRGARRGPNRPSDSSRGE-----
18 -----
19
20 >GEND01011317.1
21 -----
22 -----DLSRQGS DKA WFLP-----DNVRWTQIV
23 GNLR-----PPGESAGTAEMVESLPASQK
24 -----SKVGRKKEKFPGRATKAA
25 GQ----RIWN-----AFRLAFWAIAPSAQRKGLNPRS-----QR
26 QARSIQAFRTLEK-WIAISLLK-----CGTEWTCKAVK-----DYSEYCRARAL---
27 -----RAKQCVRPREFP-VTG-----FEIGLVXGAINDPLAGAEA
28 ---LAQLARLGRA--MP--QGT--DKQVIRALEKHRDT-LLGETRVD----EE----
29 ----TLDRYHAWAER--WSSD--RRRFAPERVAV-----KFSDS--AAESVS
30 RADG-GQKA--DLL-----ENARPHFEALRAALAESQGW DAPDDYIGLYSEED-----
31 -----
32 -----DVETIAQMALETVR
33
34 AAVR-----
35 --SGEPV PARASA--IPEYG-SKVRVVTACPGPFVVA-GDACR-----KV VWG-LL
36 EADPRVDLSGRRPLGDSISQFQ-DEVAR-----SLVGAV-----TT-----
37 ----EY-YSA-----DLTAATDLMPF--QLSAAIWTGLCDG-LG--ES-
38 -QDSD--LRSIGLHL-LG---PVQV--R-Y-PDLK----EG-----S
39 VEEGF--RV-----PTIISQQGCMMLPLSWTVLNVFNLAT----A---
40 DLA---I-G-----APEDRVTDLIGR-----
41 -----APA--VARGDD--LAAA--FLPEEA--DRYEALISA-TGGRVNLSKSF
42 RS--SIG-FVLAERT----F-----MVKRSV--DRYQR
43 RAXRQAVLIQPHRPEGIP-----QNA-----
44 PILAQDPS-----
45 --APTMSGVGRDVRGH-----RTESLRMLPDCSLKGLLGGS-----AST-----DGA
46 PRYVTLP--VSAASSLY-----EFEGTR-----FYGALC--RGLLSVHKDLLG-T
47 F-REF-GVPKF--YPREL--GGGG--F--PH-----P-RGYLR-----
48 ----AMESAG-GIGH----LRATLNITRFTESSRRAAALHM-----DP-
49 -----WRTERD-----SRTLERARSQLIAE-----
50 -----
51 -----
52
53
54
55
56
57
58
59
60

```

-----ESRAREAG-----
-----
-----
-----
-----
-----
-----
-----
-----
>KX883556.1
-----
-----MQEVNIS-----TLTRQGVSN
DRPLSATEKVDADCMKILHESSGTDPTAARG-----APHPSTEVEGEESRQRGRSK
-----AAPAPAREGSGADHRSK
NPAADNVSVWG-----PGGRALLSDSLSSVLSLGPSSSTRITGLRGV-----TL
AIFTAEQDEV LARS LAYRTCRHLARRALHEGVEAVLSVMG-----RWFSSARESFL---
-----NGAKNPSLNPV-MRR-----FSPGTRKDRVTC-----
---VAQFSYGKRA---LP--VAS--DRVAQESLRVHERR-LTTPLSTP-----DP-----
----VLADFRRFCRA---WAE--HLDPNPLPLGVV-----GLSQA--ASYTHS
VRKG-GHMA---RCR-----EILEEMNASNYCEPPTRPP-----
-----G
LLPQ-----EWWDGWVK--ERLLT
HCRE-----RYSTPGRE-
----PPKGTAVI--VYERG-LKARIVTKIETEACIL-GHQAR-----MRLVS-AM
RKMPLE-----SALSGNH-KEFLT-----AFNGQEG-----
----WI-LSS-----DLTAASDLLPL--DLVSAGVEGICDS-GR-----
-LLPD--EELGLRVC-TG---PFDL--E-W-----
-VRGS-----PRRSSTGILMGAPPTWCLLSLIHLFW----L---
EQA-----RRVYGNSGGQ-----
-----LPA--RIFGDD--LVAA--FDKRQK--LAYERSVTV-CHGLLSKKGKHS
FH--DSH-GVFLERL-----FVASRHIVEVDHT
NRGPL-----
QVMGRLTT-----
-----RHNKIKSI-----SLLRTLPLRPLTVLQVSVAAGRRHVSK-----GTL
PTPLAVG--GVSDGLMQH-----GFPEL-----IRRCQI-----ALWPGLPA-F
Y-RKL-GIPAF--LPTLL--GGGG--LI-----LPEGWEA-----
----PITRFS-KLVR-----ASTTALVTG-----NGRLSYVPLARLPPVQLQMA----
----LEEADQ-----LLPAHPHRVSHV-----TPKPWLGHIAW-----
-----HD-MGVWVNDFRITA-----
-----AMVAQEGYLLQ-----
-MPMECLGRPLNITVGRWK---DQVWKTRLALCK-----
-KAW-----WASVTRARYTIARAVRKCTEWPHV-----WL
PGSPDPDH-----PGVYGI--
-----PMYIDSKPVMEAKAAIALRQGRPTS-----
-----
>KX883562.1
-----
-----MKILHESSGTDPTAARGAPH
PSTE-----VEGEESRQRGRSKAAPAPAR
-----EGSGADHRSKNPAADNVSV

```

1
2
3 WPGGGRALLSD-----SLLSSVLSLGPSSSTRLLITGLRGV-----TL
4 AIFTAEQDEV LARSLAYRTCRHLARRALHEGVEAVLSVMG-----RWFSSARESFL---
5 -----NGAKNPSLNPVMRR-----FSPGTRKDRVTC-----
6 ---VAQFSYGKRA--LP--VAS--DRVAQESLRVHERR-LTTPLSTP-----DP-----
7 ----VLADFRRFCRA--WAEG--HLDPNPLPLGVV-----GLSQA--ASYTHS
8 VRKG-GHMA-----RCREILEEMNASNYCEPPTRPPGLLPQE-----
9 -----
10 -----
11 -----WWDGWVK--ERLLT
12 HCRE-----RYSTPGRE-
13 ----PPKGTAVI--VYERG-LKARIVTKIETEACIL-GHQAR-----MRLVS-AM
14 RKMP EL-----SALSGNH-KEFLT-----AFNGQEG-----
15 ----WI-LSS-----DLTAASDLLPL--DLVSAGVEGICDS-GR-----
16 -LLPD--EELGLRVC-TG---PFDL--E-W-----
17 -VRGS-----PRRSSTGILMGAPPTWCLLSLIHLFW----L---
18 EQA-----RRVYGNSGGQ-----
19 -----LPA--RIFGDD--LVAA--FDKRQK--LAYERSVTV-CHGLLSK GKHS
20 FH--DSH-GVFLERL-----FVASRHIVEVDHT
21 NRGPL-----
22 QVMGRLTT-----
23 -----RHNKIKSI-----SLLRTLPLRPLTVLQVSVAAAGRRHVSK-----GTL
24 PTPLAVG--GVSDGLMQH-----GFPEL-----IRRCQI-----ALWPGLPA-F
25 Y-RKL-GIPAF--LPTLL--GGGG--LI-----LPEGWEA-----
26 ----PITRFS-KLVR----ASTTALVTG-----NGRLSYVPLARLPPPV LQMA----
27 ----LEEADQ-----LLPAHPHRVSHV-----TPKPWLGHIAW-----
28 -----HD-MGVWNDFRITA-----
29 -----AMVAQEGYLLQ-----
30 -MPMECLGRPLNITVGRWK---DQVWKTRLALCK-----
31 -KAW-----WASVTRARYTIARAVRKCTEWPHV-----WL
32 PGSPDPDH-----PGVYGI--
33 -----PMYIDSKPVMEAKAAIALRQGR-----
34 -----
35 >GENC01041260.1
36 -----
37 -----
38 -----
39 -----
40 -----
41 -----
42 -----
43 -----
44 -----
45 -----LRSKVDQWFDQPSEFGDTL-----FR
46 DFPSFRDDTRFTN-----QGDENAAQY-----EYA-----
47 -----
48 ---MSLVAADENS--LP---E--AKLPSRGVTEYVVD-VTADVHLE----RER-----
49 ----ISKIVRDISLR--W-----LA
50 KLTG-----
51 -----
52 -----
53 -----
54 -----
55 ----PPRASVVA--LRERG-FKARIVTKSPVELVEC-GHLLR-----SLVWP-ML
56
57
58
59
60

RNDPHV-----HGSLSGARLESFIN-----DLKERPLVNPAWVGRPSI-----
-----ISA-----DLTAATDGLYR--WAISAVWEGVCDG-AG-----
-FGED--VRKMGRVL-LG---PMKM--S-Y-----
-PASA--GV-----DEFTSKRGCLMGLPLSWFVLNVINIWA----A---
KSA--V-E----K-----SCRGRVGKPAD-----
-----FRF--AICGDD--LVGI--MPARAQ--RLYQRNIEV-VGSGLSEGKHL
ES--DRL-AFFTEQA----A-----T---FTRVTR--SA---
-----PRG-----
TLAEKAGL-----
-----REEYLEPV-----EMVGAIPVKSIVHPSHFAGLRQGP-----AAT
PSWAAAG--PALSSLAV-----PWVHDS-----TLKVVS--RLGRYLRPEW---R
L-LASSGIPPE--LPRQL--GGGG--F--PA-----RIPRRGFLA-----
-----APKQWRS----ALTCALLTGD----HKWLSRL-----R-NS-
-----WLTTGV-----YGDVMAEALSMAEE-----EFQDYPLAELE-----
-----IP-LSEVRTTRFTGGGLAI
TEA-----DLRTRVATVWA-----
---PALTIASPFIEGRXY--CERFHSFRKRLSL-----MVK
RKRL-----PGHLQPAQFNDTEAVSRLISLSEG-----RS
FLVPRESL-----PPGLNIDFIG
TRRPCG-----DGEDR--AIEVLT-----

>LA534138.1

-----RGVDREARE-----

---LLQFSSFARA---LP--PAD--DLVCKEALRDHREI-LNT--PVS--TPVA-----
----VVSELSRYATR--WGQR--MARFANTSVAKL-----SASSS--ATLDYS
RRLG-GMRS---DLKGM---VDGWFEEDPASPTEFHPYPTFSDPTRFTATGEMNRAMHP
APT-----R
GSVE-----YVVNVLEDPDRLERSRVARIIR
DVSL-----RRF
RAREGPLPCKATT--VKERG-YKCRVVTKSPADVVEV-GHLVR-----SVVWP-ML
EHDPRV-----QASLDGGR-LEEVE-----KGFSRQIECPASLGS-----
----LLLUSA-----DLTKATDGFSR--ESIYAVWDGVCEG-AD-----
-LPSD--VHALGRRI-LG---PMSV--E-Y-----
-EDDE--LE-----DLESSRGCLMGLPLSWFVLNIINLWA----C---
ESS--I-R-----ECCHRLSLPSEVCKDL-----
-----LRF--AICGDD--LASV--FPAAEH--DGYERRISD-VGSGLSGDKHL
VS--LHV-LLFTEQM----C-----W---FEFEER--PAPPY

TLLAWLKP-----
-----GKAVPEGFRATYTCFTA VRMIDYVPVRSLVHPGHFAIKRVMG---PISFEL

1
2
3 PSWATSG--PAITSAIP-----EWIGVE-----KRRRIH---QVVKVLRPEARA--
4 --LASVGIPPF--VPREI--GGGG---F-----PPARPGR-----
5 ----VLRDCP-RSYR----LFLFSLLVGDQKKAETVARRII--NLWRTCGVAG-DL-
6 ----LSDALA-----EAEFEVE-----QKPLWKSPSDG-----
7 -----SL-AERLSTVIDDHLTMTE
8 D-----DALLRLASVWA-----PALGIAGFQQGRP
9 YCKKFGAFQYRALVGAG--SARITGRGNPLRN-----ED
10 EAAV-----ALRLKEICAGPVVFVPRDRIPPGL-----GVV-----VQ
11 GMCSRKRA-----
12 -----VGLHR--KIRSNRMESA-----
13 -----
14
15
16 >LA710658.1
17 -----
18 -----PPQKKTLDHNP-----RVDVWWNAC
19 EAVL-----GSRGAIRPKGRGATRQVLPR
20 -----STRQRAAFLSVSLPSAAHTKL
21 LS----LVWE-----ALLGVAAAYRPLPVWGS-----RF
22 GNKRCLELQRLAL-WLVRTVDG-----QGIGATIAWLK-----SAASGARNTCV---
23 -----TGAALLPSTRFL-VRK-----ILIGDTRLES-----
24 ---LDQLSFLSRS--LP--EGD--SVVESRTLIAHRDN-MTTSFRTS-----PN-----
25 ----LISNLRGFAMA--FATR--HLKKEDLREEV-----VPTPS--ASLDSS
26 RKEG-GCRE---EVR-----RMFMRWLIECPHVVPHDAWNQAIFS-----
27 -----S
28
29 EWLD-----PSEVYGLRRAQATS
30 HVAR-----AAAYK
31 TAAAELEHRVLT--VPERG-WKRRVVSAPPATFA-GTVLN-----RAMLR-GV
32 RRHGPC-----AMFLRGDR-KGAVE-----QAVSHASVGC-----
33 ----RF-VST-----DLTAATDRLPL--DLVRAVVDGLCDGWSG-----
34 -LPPV--WAEALYAL-TG---TQTL--R-Y-----
35 -PWGQ-----TIQESGVLMLGPSWPIMSIIHAWW----V---
36 ELA-----ASRVGIPRSVAR-----
37 -----RTT--AIGGDD--LLGA--WPPRLE--ESYRRLVLE-TNGKPSKGKDF
38 SS--DTS-GNFTEMT----F-----WVVGEA-----
39 -----
40
41 DGTPQIRW-----
42 -----SAAIPTKGLVGTS----IDEL-----GAA
43 YESLGSE-----PGRCLR-----GRRVLK-----ALRPHAWR--
44 V-CREAGVSVN--APRL--GGAG---L--P-----PLRGLA-----
45 ----RVDFK-KWHA----LALGKFLYG-----SGQDQIP--FSPPSWVEAADPAV-
46 ----WEARQV-----AEQRLRAE-----AEIGIVSF-----
47 -----D-TRPLCGTPGSK-HV--
48 -----VQSIDQMAWW-----AGARVFS
49 DTPFPPVATEMVSLKKYHR--LIKRWCSSRTKEG-----IPSSLAVKSGR
50 NSRF-----ALTAKARRNRDRWVIRNALFDNGI-----PV
51 DPRHRPII-----
52 -----
53
54
55
56
57
58
59
60

>U90136.1

-----MHHKVNVTQREVHFPMD-----LLQACGASA
PRPV-----ARVSRATDLDRRYRCVLS---LPEERARSVGCKWSSTRAAL
-----RRGLEELGSREFRRRLRL
AD----DCWR-----AICAAVCTGRKFPSFSVTD-----RP
ARARLAKVYRMGR-RLLVGTV-----CRGESVVSDLK-----QECADLRRVIF---
-----EGSTRIPSSSLW-----GLVGVLGWTSPER-----
---AMQLTFIGRA---LP--YGS--PDVERRALASHAAT-LSIPAECHE---PN-----
----YLVAEEQFAKS---WADD--NLPRKFRIYPI-----AVQES--SCMEYS
RAQG-GLLQ---SFR-----KGFVGYDPAAPSADPDDL-----

-----ELAKERG--FSRIR
ASWY-----STFRYRGEL
KSTNQSLEARVAV--VPERG-FKARIVTTHSASRVTF-GHQFR-----RYLLQ-GI
RRHPAL-----VDVIGGDH--RRAVE-----TMDGDFGLLRPDG-----
----RL-LSA-----DLTSASDRIPH--DLVKAILRGIFSDPDR-----
-RPPGTSLADVFDLV-LG---PYHL--H-Y-----
-PDGS-----EVTVRQGILMGLPTTWPLLCLHLFW----V---
ELS--D-W-----A-----PARPNHSRGEVLG-----
-----ESF--RICGDD--LIAW--WRPERI--ALYNQIAVD-CGAQFSAGKHL
ES--KTW-GIFTEKV----F-----T---VKPVKM--KVRVR
SEPSLKGYVFSRSSAFSCRMGGKGITGI-----RAARLYTIGAMP
RWSRRIRD-----
--VYPGSLEHRTASQRYGEPVTYRFRGWSSAIPLRWAVRAP---TRTVGN---PVQSL
PDWFTVG--PAASSVAAD-----SNAFGA-----VSRVLR-----RMFPGLPR-K
L-ASA-GIPPY--LPRVF--GGGGLVKS-----TGLTT-----
----KIGAVASRRWM----SRIGHDLYRS-----RERKSTL-----G-RV-
----WTLSTS-----PAYAASLHEVEKFM-----DRPDILTRKC-----
-----RNP-MLKHARELGLFEEVF-
-----ESRVGGGILWA-----SLNGKALVES
HSPSILQVSRNLRRSLACP---SGGFLRPSAPIGK-----LVQRHTLPR
GTVW-----FLESSATDSARQGGMGLPPPPPPPLGG-----GGMAGPPPPPPFM
GLRPESSV-----
-----PTSVP--FTPSMFSERLAALESFGRPPPS-----

>MF176258.1

-----MPLVAR-----PTSMRVATRSEVSRRVNLA

-----ETWK-----AICAAVATRKWPKFNLD-----QT
DRRRFCTFQALAR-HLLLSALA-----RGPVETIKRVK-----SWTAQIREDVF---
-----LHGRCPSA-----FRGLVGALGWTD--NS-
---HLQLSYIGRA---LP--YGD--ATICEHALKSHADV-LGSLGVTP----PS-----

1
2
3 ----LLGAARGWAKE---WAIK--NRLSLDLIREP-----PIQES--SCLEYS
4 RREG-GYAR---AIK-----EHTARCLTLPPSSVRHSMK-----
5 -----E
6
7 GTRA-----SNPGRVL--KDLAF
8 ADLS-----
9 -----NTARVVA--VPERG-WKARIVTAHAAARVAF-LHSLR-----HGLFA-AL
10 RNDRRT-----KLVAEGLH-REAVE-----ALFATGVDPRC-----
11 ----TI-VSA-----DLSAASDTLHG--DLLQAIVEGLGDA-LGC----
12 -TSKF--MDEFTNAA-VG---SYTL--S-Y-----
13 -PDGS-----TVTTQRGALMGLPTTWPLLCLTHLFW----C---
14 DQS---L-----VQLHGRGGCREKETPGSSR-----
15 -----EAE--VICGDD--LAAA--WRPDRV--AAYEEIAVL-CGAVFSPGKHL
16 KS--KVY-GIFTEDI----Y-----S---VQVSCV--PTRVT
17 REVSKTWTGPVDKVPMLVNELGQLLRWRAQRYKVGKTGAFQFP----PRN-----
18 VTLSYVSK-----
19
20 -----ETGLKMNPV-----SVLFREWSTAIPLRWAVRAP--KHAPGMR-----GDL
21 PPWVTVP--LAAHAVASAH-----PGRWSK-----VCRVVK-----LCYPGMAK--
22 F-FASHGIPPY--LPRVL--GGGG--L-----PTPLG-----
23 ----DRVKIG-RVAS-----RRIRKALGG-----ALYRSTDPTTFGSIWTTALS-PA-
24 ----YALALA-----DAQDLDAMHPDA-----FKVRQTRSLRP-----
25 -----P-FKDNGNYADFV----
26 -----EMRAGKGSWA-----LRQGAQLVGRSFTSP
27 ARVSNALRTKVRKALAKRG---YLRSSAPVSKLIE-----RASVKPS
28 DHMW-----IRERPVL SFGRFGGR-----MS
29 GKKEEPVS-----
30 -----FQWKK--RESVKTGS-----
31
32
33
34 >GDUK01014407.1
35
36 -----SCMRWLRMLKTGTSP-----AADAAGSVG
37 HS-----GPTESWTRSGSAVARAPLGS
38 -----DVLASGMSLEKRKALASRIK
39 RANLRLRSLLG-----TWDAGLTAFFGRQIRRG-----EV
40 ADSVVVWYDHVRS-WLVRVAAT-----QGVEVAAAELK-----KFSSDARRAWVEGS
41 -----PPRHFWRSCPEW-LRR-----KQGL-----
42 ---WAQLSYLGRA---LP--LGV--DFHEKKALDQHYAD-LTSSFSCS-----EE-----
43 ----VLASARAFAER---WARS--HLSRAPGPEDWLEF-----PSGSS--ATFGRK
44 RREG-GFAT-----DCQILLESAAEEVDVTPPDHVPV-----
45 -----G
46 PWSG-----DVTAFRLVTAALVE
47 AHGR-----
48 --RDTLPKGRVAV--VPERG-HKVRVVSAMERHALIL-GELAR-----KRLFK-GL
49 RKWGPL-----AAVLRGDS-RSGVL-----RLLGVPG-----
50 ----EF-VSA-----DLRAASDLIPL--DLASAIVDGLEAS-GRL----
51 -MPAE---IWGLRLGTG---PQEL--T-W-----
52 -PSGK-----TAVTSRGILMGLPTTWALLNLHGWA-----W---
53
54
55
56
57
58
59
60

1
2
3 SQA-----AKAVPVQSLPRASAARGGR-----
4 -----PALSCAICGDD--LVGV--APPVVI--DAYEAAMRS-IGAEMSAGKHF
5 RS--TRR-AVFLEEL----W-----E---FRGSAV--SVVDG
6 NPIYRTIR-----EKG-----
7 KRVRQV-----
8 -----NSTRLRAD-----RLVRATAIPLKGMVSAPGVPGR-----IDP
9 PDWWVAG--LAESALADR-----GFSRAK-----ISSVSR-----SLRPDLPV-R
10 F-RAS-GIPPF--LPRAL--GGAG--L-IAPH-----
11 ----TRCEAP-RLHR----KAIATLI-----WGYDPDSTLAFDRVWSDSVP-GD-
12 ----WRELAI-----DMVDAE-----FTPEVTRVLPV-----
13 -----GSRP-TGSWVAIGHPDQ----
14 -----VREKAVADNAS-----ALSSSLGSPDPVVG
15 YPTIPVLARRLRKVREDLA----SRWASAKPTKR-----PVDELLLRQRAAD
16 LRVW-----VPLTDQTSRWNDWLYVSATSPYKKS-----WR
17 RLVISSVS-----REDLPAVT
18 -----EWWLPGGGRIPSSTMDPSPPEWR-----
19 -----
20 >GFKT011160020.1
21 -----
22 -----PMKEPVDCRLS-----TPAGFSGTV
23 PPP-----GRTKAARQGTIPVRRSRGSA
24 -----SALPGKIYGWSRRQRDR
25 FA----MLLS-----SFDAALAAYSGVVVSRT-----RS
26 LPPSLRLFRMTMRKWSVTARG-----NGVEFAIASAK-----EFSAACRAGWISGT
27 VP-----DHFFMKWLPEPVR-----RKSG-----
28 ---WAQLSFIGRS---LP--EGG--DRHEIEALANHKAA-LSSSFEVP----AD----
29 ----VLTSLRNYSED---WARR--HLAADPDPSLLCEP-----CTGNS--ATFERT
30 RREG-GFAQSITDLV-----SSSPTDNLLPLESMFPGPTQGG-----
31 -----
32 -----ALPVHVLEVSLSR
33 RNS-----
34 ----DPKGRVSV--VRERG-HKVRVVSAMETHVL-GHAAR-----RRLFK-GL
35 RRERRL-----KDTLKGDF-EATTK-----AFVGCAG-----
36 ----TV-ISS-----DMKSASDLIPL--SVASAIVDGLEAS-GR-----
37 -LLPV--EVAGLRAC-TG---PQHL--V-Y-----
38 -PDGS-----EITTRRGILMGLPTTWAILNLMHLWC-----W---
39 DSA-----DRQYRLEGHPFRAT-----
40 -----VRSDCRVCGDD--LIGV--GPDSLL--RSYDRNLGL-VGMILSPGKHF
41 RS--NRR-GVFLERL----L-----E---FQTRKT--VYEHA
42 VIYRKVG-----HRR-----
43 VPVDRSHI-----
44 -----PVVTRV-----TVLNTIPLKGLVRAS-----VLGR-----DDP
45 PVWWAAA---VAESSLLS-----DYPRKK-----IFAAAR-----TLRPGLSR-Q
46 F-RRL-GIPPF--LPREL--GGAG--L-----VGSSD-----
47 ----RVDAP-AFHR----KAISLV-----WGSDATAAYSFIRMWQGFEG-HP-
48 ----WKTAAS---QETDTWFADYKVTRPGK-----MYPDRYGFLDG-----
49
50
51
52
53
54
55
56
57
58
59
60

-----ES-LRTKSTMLNSA-----
-----VYETFLGPDPE-----ATHYPSL
RIVASRLAKVRKDLVSRWP--SVRPVG--KDLGT-----ILEAFEESKL
CTLW-----VTPYDASGYFDDSLLLMDESUYQRR-----FR
QLVIAGLM-----REDRMGDLL
F-----PSWLP--PSTVVS-----

>M64034.1

-----MKEPVDCRLS-----TPAGFSGTV
PPP-----GRTKAARPGTIPVRRSRGSA
-----SALPGKIYGWSRRQRDR
FA----MLLS-----SFDAALAAYSGVVVSRT-----RS
LPPSLRLFRAMTRKWLSTARG-----NGVEFAIASAK-----EFSAACRAGWISGT
VP-----DHFFMKWLPEPVR-----RKSGL-----
---WAQLSFIGRS---LP--EGG--DRHEIEALANHKAA-LSSSFEVP----AD-----
----VLTSLRNYSED--WARR--HLAADPDPSLLCEP-----CTGNS--ATFERT
RREG-GFAQSITDLV-----SSSPTDNLPPLESMPFGPTQGG-----

-----ALPVHVLEVSLSRV
HNGS-----
----DPKGRVSV--VRERG-HKVRVVSAMETHELVL-GHAAR-----RRLFK-GL
RRERRL-----RDTLKGDF-EATTK-----AFVGCAG-----
----TV-ISS-----DMKSASDLIPL--SVASAIVDGLEAS-GR-----
-LLPV--EIAGLRAC-TG---PQHL--V-Y-----
-PDGS-----EITTRRGILMGLPTTWAILNLMHLWC-----W---
DSA-----DRQYRLEGHPFRAT-----
-----VRSDCRVCGDD--LIGV--GPDSLL--RSYDRNLGL-VGMILSPGKHF
RS--NRR-GVFLERL----L-----E---FQTRKT--VYEHA
VIYRKVG-----HRR-----
VPVDRSHI-----
-----PVVTRV-----TVLNTIPLKGLVRAS-----VLGR-----DDP
PVWWAAA--VAESSLLS-----DYPRKK-----IFAAAR-----TLRPGLSR-Q
F-RRL-GIPPF--LPREL--GGAG--L-----VGPSD-----
-----RVDAP-AFHR-----KAISLV-----WGSDATAAYSFIRMWQGFEG-HP-
-----WKTAAS---QETDTWFADYKVTRPGK-----MYPDRYGFLLDG-----
-----ES-LRTKSTMLNSA-----
-----VYETFLGPDPE-----ATHYPSL
RIVASRLAKVRKDLVNRWP--SVKPVG--KDLGT-----ILEAFEESKL
CTLW-----VTPYDASGYFDDSLLLMDESUYQRR-----FR
QLVIAGLM-----REGRMGDLL
F-----PNWLP--PSTVVV-----

>AF039063.1

-----MKEPVDCRLS-----TPAGFSGTV

1
2
3 PPP-----GRTKAARPGTIPVRRSRGSA
4 -----SALPGKIYGWSRRQRDR
5 FA----MLLS-----SFDAALAAYSGVVVSRT-----RS
6 LPPSLRLFRAMTRKWSVTARG-----NGVEFAIASAK-----EFSAACRAGWISGT
7 VP-----DHFFMKWLPEPVR-----RKSGL-----
8 ---WAQLSFIGRS---LP--EGG--DRHEIEALANHKAA-LSSSFEVP----AD-----
9 ----VLTSLRNYSED---WARR--HLAADPDPSLLCEP-----CTGNS--ATFERT
10 RREG-GFAQSITDLV-----SSSPTDNLPPLESMPFGPTQGG-----
11 -----
12 -----ALPVHVLEVSLSR
13 HNGS-----
14 ----DPKGRVSV--VRERG-HKVRVVSAMETHVL-GHAAR-----RRLFK-GL
15 RRERRL-----RDTLKGDF-EATTK-----AFVGCAG-----
16 ----TV-ISS-----DMKSASDLPL--SVASAIVDGLAS-GR-----
17 -LLPV--EIAGLRAC-TG---PQHL--V-Y-----
18 -PDGS-----EITTRRGILMGLPTTWAILNLMHLWC----W---
19 DSA-----DRQYRLEGHPFRAT-----
20 -----VRSDCRVCGDD--LIGV--GPDSLL--RSYDRNLGL-VGMILSPGKHF
21 RS--NRR-GVFLERL----L-----E---FQTRKT--VYEHA
22 VIYRKVG-----HRR-----
23 VPVDRSHI-----
24 -----PVVTRV-----TVLNTIPLKGLVRAS----VLGR-----DDP
25 PVWWAAA---VAESSLLS-----DYPRKK-----IFAAAR-----TLRPGLSR-Q
26 F-RRL-GIPPF--LPREL--GGAG--L-----VGPSD-----
27 -----RVDAP-AFHR----KAISLV-----WGSDATAAYSFIRMWQGFEG-HP-
28 -----WKTAAS---QETDTWFADYKVTRPGK-----MYPDRYGFLDG-----
29 -----ES-LRTKSTMLNSA-----
30 -----VYETFLGPPDP-----ATHYPSL
31 RIVASRLAKVRKDLVNRWP--SVKPVGKDLGTL-----EAFEESKL
32 CTLW-----VTPYDASGYFDDSLLLMDESUYQRR-----FR
33 QLVIAGLM-----REGRMGDL
34 F-----PNWLP--PSTVVS GFP-----
35 -----
36 >MF190030.1
37 -----
38 -----MGNKVGRFSSRSCDR-----YLKRAPLPG
39 DRKG-----RVRGRVIDKLEVP-----RTRSRKSKGSRMRGARPPMA
40 -----ESYKIKPVGWRKGFMTPGRE
41 IE----GLWV-----ALIATLTSISPRICRFNMRC-----RA
42 DRDRHGAMVRLAV-WFTRRVD-----RSSYDLLKSAK-----KLTAGFREASI---
43 -----LRRPYNWVDAPPG-WAK-----RFRPILKGIRLSEA-----
44 -V-LDQLSYMGRA---LP--KGG--PRVLQEALKDHKDC-MLKPSPLP-----EE-----
45 ----VFGAARDFACR---WADR--YVASGGRLSHP-----GTTES--SVLGST
46 RRDG-GISS-----SLCGLTYSVKPLPMDVQDEVRLVSSLKCDTPSEC---
47 -----
48 -----DVDAALR--TTSIR
49
50
51
52
53
54
55
56
57
58
59
60

EWAI-----DESSPDV-
----IPENKVLA--LPERG-YKTRVATIPDLPRQIL-GGFVR-----SYLIP-AL
RRTPEC-----ADTLRGDD-ERAAC-----RIGSSQRSFST-----
----DMIVSS-----DLTAASDTLSR--ELLALTAGLVDS-GK-----
-FPSW--FGPALRQL-VG---PMLL--R-Y-----
-PDGS-----SITSQRGTLMLGHTWSMLSLSHLFW----L---
DYS-----NKTVHKTIARSVE-----
-----RTA--AVCGDD--LIVR--APWRWI--NSYHRIAKM-CGAQFSSGKHF
IS--RER-YCFLEQL----W-----T---LKSKEIVPAYTRA
LDK-----RIP-----
KGGDQFPP-----
----LGTDPKSVRRV--MKFHIPSKCVRSTAIPLAGLTSGAGSDPEKPLV-----VGT
PTWAKLG--LAVENTSRT-----PVAARK-----VRAVSR-----ALYPRLPR-W
M-RKH-GIEPY--LPRAY--GGGG--L-----LPPGNVS-----
----HLGALA-PQVR----KLVTVTAYG-----RCPKTAARLSRIYLPNRTA-TN-
-----FTEAQE-----HTDIRFKRVGAI-----LCRRGKVPF-----
-----NGHAVK-LGTLTDAPERA-ALAI
-----ARELSMVLFPF-----PRNRK
KSTISPFQVGRAFHSA MRK---AIKKWPSAHPCSA-----KVTRRVLLQKARAL
EEEF-----VWWSPRRRIPEEGFGGWG-----LA
PAAQRVLK-----
-----DSLWSSDC-----

>KX883516.1

-----MKPPRPTEVHKDFRLP-----VLLRACRVG
GTWS-----RDVSRYLVPALQDGVGV---RPGEGGNPQTRPRKPAKSVD
-----LERTTRYDLTVRWRKEFRSRGRL
VE----GAWA-----AVHAALVAAHPDLGRADMST-----RA
GRDRIWAWRRLAV-WMVRWID-----RDPRDMVRGWK-----QLAATFREASI---
-----MQKPYDWAAAPAG-FPK-----RLRPVLAGVRLTPP-----
-V-LDQLSYIGRA--LP--QGD--AWVQRKALDDHREC-LSSA-PRP---LANH-----
----VLAAARDFARS--WASR--HVRPSGRISHP-----KVLES--ATIQEP
RRSG-GVAA--HLC-----KQGFDLPPFEGLTDEL RDQIRAMGESAPSDCDLEA----
-----I
GRST-----AFRDWAL--GQLCC
PDAS-----
----PPRSRVEA--IPERG-FKARVVTVPDL PESIL-GSALR-----CYLLP-AL
RRTPEC-----ADVLSGKE-ELAAM-----RIASSRRPIE-----GD-----
----WI-VSS-----DLTAASDTLSF--ELLQALVDGLEES-GT-----
-LPPW--AIRALKVL-VG---PQEL--V-Y-----
-PDGS-----VLVTRRGTLMLGHTWSILSLSHLFW----L---
DWA-----HRSVDPRVADVVR-----
-----RTA--AVCGDD--LVVR--APLKWV--QSYHQIAEA-SGARFSTGKHF
VA--KRR-YVFLERI----W-----S---LKSSLQ--RVRLP
RERTR-----RAG-----

RWVREFPP-----
----LGARPVKVRPT-----VVRSPRSVPSTAIPLAGLTAGEVDCHFPGVRNLVL
PNWVRLG--MAVETLLHS-----PEAARK-----VRKVVR-----ALYPGLPR-W
M-RRH-GVQPY--LPRFL--GGGG--LL-----PPGRNVT-----
----RIGRLP-RPVA----KAVTTWVYG---RSPRENLSL--AKIYLPTRRI-TN-
-----YAVGEQ-----HTAVRFQQVSAVLCRA-----GKKPFDGHGRR-----
-----LGTLSEAPEAA-ALAI
-----ARELSMVLLPE-----PSEERAYR
RLLRPSQVGQTFRRLVL-R---ANRRWPGSKPICP-----RAPRRALLARCKAV
EEGF-----VWWANPRLLTQFGFGGWG-----LE
PAAQRVVK-----
-----DALGWSTFRPTPGRGTH-----

>GEND01007370.1

-----LDHEPPTMGSSKKTARETHVSNVSLH-----ALLRNPRVG
GSWS-----KDVSL--YLAPA-----RRSGVQGESGERGLPSRTGL
-----ATHANLEHRGAS-----RRGQGHAKRGPEPRRKAPKGGYTR
QR----VCWS-----REFKSSGRKLEAGWNALTAALAACHPAALRFDLSVRA
DRDRVGALKRLAV-ACVRRLD-----RGVGELMAELK-----SLATTFRTASI---
-----RQVPYSWDTCPAG-LRR-----HLEAALRGFRLTTE-----
--TLDQLSFLGRA--LP--PGD--STVEERALRVHRDV-ITSD-PAP---IPPG-----
----ILKAAEAFACT--WALT--YLHRGGRLAHP-----KTMES--STITSR
RREG-GVAA-----DLCMKSMDDLPLPETVQAEVRRQ---VLELEGPLPSECD
LSA-----A
FRTT-----ALRDWAV--TTLTE
PGYV-----
----PTH-KVLP--IPEAG-WKCRIATVPELPASVA-GGLCR-----SYLLP-AL
RRTPEC-----ADVLRGEE-DKAAR-----RVAASRRPIE-----GD-----
----WI-VSS-----DLTAASDTLAF--GLLQALVRGLALS-GR-----
-LPPW--VVRSLVL-VG---PQRL--T-Y-----
-PDGE-----TVITARGTLMGLGHTWSLLSLSHLFW----L---
EAA-----HASVPRPVSEQVR-----
-----LTA--AVCGDD--LVVR--APLGWI--RRYEEVASL-TGAQFSQTKHF
VA--KRR-FVFLEEL----Y-----T---LRGTQQ-AASTVL
GPTS-----RGG-----
RLVRAWPA-----
PGTRPPKVRVTRLRTY-----ASGTVKSTAIPLAGLVAGDVMTHQVMA-----DSV
PLWSKIG--LAVESLAST-----RERALR-----VRRVTR-----HLYPALPS-W
L-LRH-GIRPT--LPRAL--GGGG--LL-----PNSGNCM-----
----RLGRLP-----RRVAKMVTAVTVGRCPRTAARL--SRIYLPLSRRA-TN-
-----FAEAKV-----HTDIRFKRVGAV-----LSRSGREPW-----
-----NGKGVR-LGTLTEAPEQA-ALAI
-----ARELSMVLLPE-----PKKGGSSH
NAALSPFQVGKQFLALTK---ALRRWPGAKPVSA-----RAPRKSVDTRTRAV
EEAF-----VWWAPSHRIPEEGFGGWG-----LA

1
2
3 PAAQRVLK-----
4 -----DSLGWSA--PSP-----
5 -----
6
7 >GENB01008321.1
8 -----
9 -----LDHEPPTMGSSKKTARETHVSNVSLH-----ALLRNPRVG
10 GSWS-----KDVSL--YLAPA-----RRSGVQGESGERGLPSRTGL
11 -----ATHANLEHRGAS-----RRGQGHAKRGPEPRRKAPKGGYTR
12 QR----VCWS-----REFKSSGRKLEAGWNALTAALAACHPAALRFDLSVRA
13 DRDRVGALKRLAV-ACVRRLD-----RGVGELMAELK-----SLATTFRTASI---
14 -----RQVPYSWDTCPAG-LRR-----HLEAALRGFRLTTE-----
15 --TLDQLSFLGRA--LP--PGD--STVEERALRVHRDV-ITSD-PAP---IPPG-----
16 ----ILKAAEAFACT--WALT--YLHRGGRLAHP-----KTMES--STITSR
17 RREG-GVAA-----DLCMKSMDDLPLPETVQAEVRRQ---VLELEGPLPSECD
18 LSA-----A
19 FRTT-----ALRDWAV--TTLTE
20 PGYV-----
21 -----PTH-KVLP--IPEAG-WKCRIATVPELPASVA-GGLCR-----SYLLP-AL
22 RRTPEC-----ADVLRGEE-DKAAR-----RVAASRRPIEGD-----
23 ----WI-VSS-----DLTAASDTLAF--GLLQALVRGLALS-GR-----
24 -LPPW--VIRSLEVL-VG---PQRL--X-Y-----
25 -PDGE-----TVITARGTLMGLGHTWSLLSLSHLFW----L---
26 EAA-----HASVPRPVSEQVR-----
27 -----LTA--AVCGDD--LVVR--APLGWI--RRYEEVASL-TGAQFSQTKHF
28 VA--KRR-FVFLEEL----Y-----T---LRGTQQ-AASTVL
29 GPTS-----RGG-----
30 RLVRAWPA-----
31 PGTRPPKVRVTRLRTY-----ASGTVKSTAIPLAGLVAGDVMTHQVMA-----DSV
32 PLWSKIG--LAVESLAST-----RERALR-----VRRVTR-----HLYPALPS-W
33 L-LRH-GIRPT--LPRAL--GGGG--LL-----PNSGNCM-----
34 ----RLGRLP-----RRVAKMVTAVTVGRCPRTAARL--SRIYLPLSRRA-TN-
35 ----FAEAKV-----HTDIRFKRVGAV-----LSRSGREPW-----
36 -----NGKGVR-LGTLTEAPEQA-ALAI
37 -----ARELSMVLLPE-----PKKGGSSH
38 NAALSPFQVGKQFLALTK---ALRRWPGAKPVSA-----RAPRKSVS DRTRAV
39 EEAF-----VWWAPSHRIPEEGFGGWG-----LA
40 PAAQRVLK-----
41 -----DSLGWSA--PSP-----
42 -----
43 >GBWZ01004682.1
44 -----
45 -----GS RTP-----WRKSLPQMN
46 RGS R-----
47 -----QK
48 VR----ATWS-----VIRAALAASGFVRCDLST-----LA
49 ARDSANGLCRLAW-WLWRTTAH-----SGIEYTLKLLK-----TWLSETRRWVV---
50
51
52
53
54
55
56
57
58
59
60

-----EEVPKPKRRGG-MAR-----FLHGNFRSMVDSNA-----
-K-ALQLSYAGRA---LP--VGG--PSVILRSLRGHQDA-FASNPVTP-----PE-----
----LLTSARIFARE---WAKR--HLSSSPSFARA-----GLNES--SSLEST
RAKG-GVLA---QTR-----ERIYAIPDEALRERKSSLQSREEFADLSSC-----

-----DLDIVAR--ESLLR
DSAL-----EELNRSLENG-
----NPQAEVRY--CLERG-LKCRVVTKSPWALVYL-AHSVR-----HELLN-AL
RNDKRT-----KHVLKGDH-QEAVE-----DVFSTPRRID-----NR-----
----QI-VSA-----DLTSASDLLPL--DLVGALVDGLRDT-AR---LW-
-WPEW--MLKVLGIV-TG---PMDL-----WYPAE-----
-LGED--EL-----TIVGTRCGILMGVPTTWSILSLVQLFW----A---
EEA----W-----RSRFPNPAWGHPSMT-----
-----PET--AICGDD--LVGW--WPKSVI--KTYEAVAKR-CGAQFSAGKHF
HS--RTH-AVFTEEI---Y-----R--LRSISE--ETPRC
PDWESLPKLQQRWVRSGTKRR-----PRK-----
RLVNFLRT-----
-----ARLGGP-----VGKGVVKYPSIPLAGLARPSVLPGSYQTV-----
PWWIAIG--PTYEAVAAQ-----TGNPSA-----ARRVLR-----TLHPNAWR-W
M-RER-GIFPE--LPREL--GGGG---L---P-----PLRGSAD-----
----VLKLP-PFLR----FGVAKLLYG-----TKFRDLR--SPGALWSTTFA-PE-
----WRVMAA-----DHAEVK-----FKTSAVNCHPG-----
-----RVP-FHGKGFELGGVKALED
VTTRL-----AGELTLMMEHP-----KDGR
QAFSVVPRFIARSVRKFYs---ARGKVARRSKMKGIDPSGVPRT-----AL
LKRW-----ASLGQTRVWFSEPAAGGEGFVLGI-----PG
QVKRRIAP-----A-----
-----LGWAR--P-----

>GBPT01024369.1
-----LRM
AVRKPSTQAARERVTTLSLVEAALIGSRLVPLAGIVRPARARAL-----RSDRGATLN
PKRRGVGQPLEPIWAQGSTAVPIPNSSSPHKACLKPISRQQDGKPPKASAMKKGHKIKD
-----HFRRFRLSPRSSEKVRASANPPKEDHDVDWSSVRRLYPEWGRISKRGHRQ
LK----TVWA-----AFRAALAAVGSTRVSLRT-----NS
GREALRKWCKLAH-WLVRTTGL-----SGITHVLATLK-----AWTATVRRWV---
-----EGVREQDRQGG-LAK-----FLAGSLRFLVTRDD-----
-A-AVQLSYVGRA---LP--QGT--LSVQRKALVSHKKA-LGEDYRTP-----RH-----
----ILREAREFAAG---WGDR--YLRSVPLMNC-----GFSQS--ACLEVP
TSKG-GTLEATLERT-----NALQTDVNWSKAVADWRDTLKSS-----
-----I
PTLP-----DCDREILARGTALR
DRSI-----VELREA
VSQSGHPRAAVAV--LPERG-SKARVVTKSPWALVQL-AHPIR-----GWLLS-GL
RRDPRI-----AKVLEGDH-HGAAT-----GLVDVPRLADGR-----
----EF-LSA-----DLTSASDLIPL--DLAAALIEGLMEG-TGDT----

1
2
3 -LPDW--AREVLRLA-TG---PMLV--D-YSTAWP-----
4 -PSGG-----RRPLPGPCEFISRRGALMGLPTTWSLLCILQLFW-----A---
5 DRA-----W-----RLTFPNIAQRTDPGNPGMT-----
6 -----PAT--IICGDD--LAAW--WPKQVS--DAYEQTUVS-CGGIFSVGKHF
7 RS--RAV-VVFTELL----FHAEKSTVRALAGAGPDKISAPAI---ASWERE--GRPFL
8 MKVRRPHAKSTARRRRSER-----GRR-----
9 PMIPKMTR-----
10 ---GIGPHDLSDHRRV-----VTRLVPFKSIPLKGLLEAEGGGPLADNH-----VRQ
11 PRWVNIG--PVADRISKM-----TGNPYA-----VRRVLR-----TLFPHYWG-W
12 F-QKH-GFVPT--LPRFL--GGPG--L--P-----PRTGDLD-----
13 -----RLALP-RDLL----LKVRTLVHG-----TKWHLR-----SPK-AP-
14 -----WTQMG-----GFYRQMVAGH-----AEARLQTHARC-----
15 -----ARVGRVP-YNGKGVLLGGV-DVYE
16 DL-----ATRFENEATVG-----IFG
17 CRDDSRSAFKIRPGQVSQ---ALRKFWKGVKVPTGEKHRLGGEERRRGRLLNKTSARNL
18 ERRW-----IELTSSRAWHADPMAAQEGFVLGL-----NR
19 RVKLDLAI-----KLGQTGRAGGT
20 ES-----PGPSA--EEKSPAGDAPAASGSPTAS-----
21 -----
22 >GBPU01010966.1
23 -----MPSN
24 KPNTAGVSLEKLTLEAYLLGSALVPLSDSLPSRSGGAVTPRP-----NVGPRNPVG
25 PRRS-----TLHLPLSGTRSSGANPI-----KIPLLPKFPARSSGKMRLI
26 -----DLIKTSSKSVSSKRKT-----PVLRSKKNKGAKRKGRSSAFSHKW
27 VRDNRGRVRWDWTKLRALAPELASLDRKARGLCEQAWTAVRAALAADGRRRATLMT
28 SSGR
29 TA--VRSYCRLAK-WLVHTTAF-----SGITHALKTLK-----DWVSLARQYVV---
30 -----EGVRPKDRKLG-LAR-----FFRGDFRHYVDSK-----
31 -A-LLQLSYIGRA--LP--LGG--PSVVRKALIKHKEV-YSTEHTP-----AH-----
32 ----LLEAARRFASE--FAQD--RLGVVPLLSS-----GFSES--ACLERP
33 AAKG-GSAA--HTH-----AVYNSLVSAPEWEEAKREWLAEIAKEDLNIPDCDRDIA
34 VR-----V
35 AILR-----DFSLRKL--KESLE
36 ETSG-----
37 ----IPKAEVTV--VTERG-LKARVVTKSPWALVQL-GHPVR-----AWLLS-AL
38 RKDPRT-----SGVLEGDH-LNAAK-----GVVDAPRLAD-----TR-----
39 ----RF-LSA-----DLTAASDRLPR--DLVSALVEGLIEG-SG---ET-
40 -LPEW--ARWVMRVS-VG---PMLL--D-Y-SKAWNPCLG-P-----P
41 LPGDD--KI-----I--SRRGILMGLPLTWVLLSLTQMFW-----A---
42 SQG----W-----KAILAHMTRSAGPSAT-----
43 -----PST--VICGDD--LLGW--WPQVVI--DQYQSVALA-CDAQFSEGKHF
44 VS--TRR-AVFTELM----F-----E---ARIRPA--SNVSP
45 RIALWER-----RRF-----
46 PRLTFAPT-----KKPKWRKVL
47 LGPQPQCPGQRDTRRV-----VTGLRRTPLLTLRGMVEPE--AGLDSSY-----VRQ
48 PRWINLG--LVYQEVAQAQ-----SGRPAA-----ARRVLR-----TLYPGYWS-W
49
50
51
52
53
54
55
56
57
58
59
60

Y-RQK-GFQPT--LPRFL--GGPG---L---P-----PLTGDVN-----
-----RLHLP-KPLR-----LGVA AFLYS-----TPWRDIR---CPSSLWSTGTAK-
-----WVREMA-----ADSAEWKFHKFAKS-----SRVGGKVPY-----
-----GGLGVL-LGGPQAVEETTTRLAN
ELSLIL-----DCDPKESSRWK-----KKP
REVAKALNGFYARVLKRKP---GVRNWVRKGVSKGVVPG-----AL
EARW-----RRVQNAYAWYSHPTAAAEFGC-----LAL-----PQ
GVKSNLSV-----ALRQSGVPARV
ERKKS-----KMARE--PSLDGPDDLGLQSLAEFPPLPSAGASRRR-----

>GBWH01015246.1

-----GRA---LP--VGT--QVVIRRSLEAHKVN-LGEEYHVP-----EP-----
-----ILERARAWAKC---FAKD--KLRSVPLITC-----GFTQS--ACLENP
SSKG-GSFAITLSRC-----ATLETETEEQRRGRTQHEWKKSIDKEVLDAVP-----

-----DCDAHIATRТАALR
ELAL-----EEIRKQ
VEQHGYPMAEVAA--LPERG-MKARIVTKSPWSLVTЛ-SHHIR-----VWLLH-GL
RSDVRT-----SGVLDGKE-KEAVE-----GLVDVPRЛAD-----SR-----
-----KF-LSA-----DLTTASDLLPL--DLVAAIVDGLVEG-SE---GT-
-LPАW--AQEVFRLS-TG---PMVL--D-Y-SKAWDGSV-----P
LKGPT--QM-----I---SRRGILMGLPTTWLЛSLVQLFW-----
SDT---A-W-----RNGFPEIMARRGFSAPS-----
-----PAT--VICGDD--LAAW--WPEKVI--EEYEALART-CGALFSAGKHY
RS--PKN-LVFTEVI----FRSVVSTAQDHPPKRSFWERERLPFALIGKTVR--SSNVG
AKAGQVEGRNLTKKE-----RRAKLRQSERTI
TLVRRDGV-----
-----RRVSSLSSKDDRRVVRKLAVMPAVPLRGLVEPE--LGLNESF-----IRQ
PRWVNЛG---PAVDCAIQI-----SGLPHA-----VRRILK-----ILYPGYWA-W
F-RQQ-GFEPT--LPRFL--GGPG---L---P-----PLTGSPA-----
-----RLRLP-RRLR-----VAIGILLYA-----TPWKEIR-----SPAАV-
-----WTCLGA-----GSVRESAREHTEWK-----FSKSAVACRKG-----
-----KVP-FNGNGAVLGGHSAVED
VTTRM-----YNELALILDPE-----DRR
SRFKVKPREVSSAVRKFYD---RLLPRDPKQRVKT-----HPYITRGIAPSVPSAL
RRRW-----EGILGQRTWYSTPGASSEGFCLGL-----AA
GVKHESL-----
-----ALGHSGRRTRRSPS-----

>KX883605.1

-----MVARLVEDRRRDNRSGAAAKQ-----PASRLGGFM
GHRR-----GDRQRVAALFPWRRRLKGSE

-----WT-----RLSEAWRVIQAAIAAGFLTVGLED-----PA
SREFLRGLCRLAW-WMVKTGHY-----TGVERVLSDLK-----GFAGKARLWVM---
-----EETPPTLRRGS--FAR-----WFRGIGRSLVDRDE-----
-R-AMQLSFIGRA---LP--AGG--PRVVAQNLKTFREV-VTA--PYE---TAPH-----
----LLSEMEAFKA--WASR--YLQPAEIPPP-----WMGEG--ATFERT
RTEG-GFASQSVEMA-----GLFDQAYRVDADWDELIRAAEV-----
-----A
GMTD-----CDVDIACR--EAQTR
EVAM-----GTLSRLRR-
----PIAVEALA--IPERG-LKARVVTCKPWALVYL-GHFLR-----SWLLQ-GL
RRDPRI-----SAVLEGEHQRAVLD-----LLKSETAEVTD-----
----SILLSA-----DLTAATDRFPH--DLVQACTRGILEG-AR-----
-CPLNSDWTQVFADL-TG---PLLV-----V
PPRGE-----GSDYLTARGIMMGLPTTWCFLNLFNLFW----S---
ERA----W-----SLCPRVNPVIHRASLFRPGHHART-----
-----PAT--VICGDD--LAAV--WPHDVC--DQYERIAVA-CGASFSGKGHF
RS--ARY-LLFTEEP----Y-----EVLVSETERR--HGRAV
ERPQEW-----
TLADYLFP-----
-----SRTRSALEPSGRGSGWHYTGFRRMKSVPLRGLVRPT---HKPFDH-----HTI
PPWAALP--FAVQASVRA-----SGDGAR-----VRRILK-----VLHPGIWA-W
A-RNN-GFSPT--SPTL--GGYG--L--P-----PVHGRRG-----
----EGYLLP-PWHR----RGISVLLRH-----TKWKDLI--SPSTAWTLIESGSH-
----WRGMGE-----EHTRHKLAVS-----AVACKRGKVPF-----
-----EGKGSE-LGTVDCVEQLTTRMAR
-----ELRMLLGPQR-----PRDEFVAPRQIAVRV
RKMFRGLEANKGVPGGRLP--SVSLVPASGSVQR-----L
LSEY-----ESRLTERCYFSAPGPAREGFVLWL-----PG
SARRDLAV-----A-----
-----LGWTR--SSAHADD-----

>KX883602.1

-----MPRNKARHRWVCLYELTSLVATLLP-----ARPGTGPYL
GGTR-----PNGPDTGRFLPRF-----PRHPNIRCPKERGNAATGKT
-----TNGNVNGILSKVLRRRPL
GN----QRWT-----RLEEAWRTILAAIAACGFVRLDSSR-----RA
HREALRGLSRLST-WIVKTAAL-----SGTEEVRLRELK-----TWSARLRAHV---
-----NSLPPHKRRRG--FAR-----YFQGVLRSLIDKDS-----
-R-ALQMSYIGRA---LP--VGS--FRETGAIRKFREV-TQV--PFE---THPD-----
----LLDSIRDWGAH---WAAR--FLDTPAMPPL-----WRSSS--GCVDFS

```

1  RLQG-GLA-----EAAVRAADFAAAEYESREDWDDLRNTAAE-----
2  -----V
3  PEMT-----DCDVDIAVREAQLR
4  DVAL-----TVA AHLPD-
5  ----LPPVEVCT--VPERG-MKCRIVTKAPWCLIHI-GHFLR-----SWLFG-GL
6  RKDRRV-----QGVLAGDH---AG-----ALRGQTERVYSQKPTEEE-----
7  ----NICLSA-----DLTAATDLFPQ--DLVAALVDGLLSG-AR-----
8  -KPPTAEIQEVFRKL-TG---PLNA--R-L-----
9  -PDGD-----GFAIRRGIMMGFPTTWLLNLVNLFW----S---
10  ETA----W-----SASGSLPFQDVALSAQPAAS-----
11  -----PRT--IVCGDD--LASV--WPPKVA--DRYEEVAEA-CGAKFSVGKHY
12  RS--RDY-ILFTEEI----F-----K---VSWEKQEYSLMDS
13  AKP-----RRT-----
14  TLADFFPV-----
15  -----PRAPRKSTMRKRYRVQAKLAPVRCVPLRGLVRPL---HAPKDR-----RLR
16  PSWAALP---AAVEAAMD-----LGAPVA-----VRRVLR-----VLHPGIWC-W
17  A-RKR-GFSPT--LPLVV--GGYG---L---P-----PLRGSVR-----
18  ----RVSLP-KWLR----YGLAAWLLN-----SKWKDLE--GPSRCWEGVCRPVS-
19  ----WQTMAR-----EHADVRIASS-----TFAVRKGRTPP-----
20  -----PGFHL-LGGLEAIQNLSVKFEG
21  ELCV-----LLGTEEGSLKR-----QFGLAP
22  RRVANSLRKFFARMVSDHP---SRRPFPSSLGRTA-----L
23  VQRW-----RSRAEERALYSTPRSAVEGLVLSL-----PV
24  RAKRLLAE-----
25  -----ALGWV-----
26  -----
27  >IABX01132835.1
28  -----
29  -----PLPKQQHRWVHLS-----RVSLVGAAS
30  VGAR-----RGAPLSIFFEERASTDRQKGRNLPESAADYCLAEIARVSR
31  -----LAWTPDDTGIPKGPLYKVF
32  VSRPMPRALRS-----SLVETWRVVLSSLAAAGFIRLDVSK-----WA
33  GREALHGLCDLAW-WIIRTSHL-----TGIDLMFSQLK-----GYCSALRLAVV---
34  -----SEIPPAERRRG--FAR-----FFRGVLRSCVDRDA----
35  -R-ALQISYIGRA---LP--KAS--PAQCTAALENFRAV-LKRPHTTS----PE----
36  ----ILDRA YQFAVE---WAGI--HLTSPEIPNP-----FLGAG--ATLETT
37  RAQG-GFSE---ETR-----RKAGQFDWQLGAPDQEMWDERMMLLNE-----
38  -----I
39  GSLT-----DCDVDIAVREAMTR
40  DVVL-----DRIHRMGD-
41  ----FPEVSAIA--LAERG-RKTRVVT KAPWEIVYL-GHFLR-----SWLLD-GL
42  RRDSRV-----STVLMGDH-AEAVK-----RLSREALYCR-----GD-----
43  ----EI-LSA-----DLTAASDLLPL--DLVRAITDGLWEG-TRKG----
44  -CRPEERFRLVLDRL-TG---PLRV--E-M-----
45  -PDLY-----DAFISERGIMMGLPTTWCLLNLVQLFW----C---
46  EFA----W-----MSGQQYRHMHDPLRGSPRSS-----

```

1
2
3 -----PNT--AICGDD--LVAY--WPSDII--GRYEQVVKE-SGGMFSVGKHY
4 RS--GRY-AVFTEQF----Y-----D---CKVTRFSAPGPAL
5 RKPGER-----
6 TLADFLFP-----
7
8 -----CMPSKGGEITPWNQVGEVRNVPTLPLRGLVRPSHGPGDRHLM-----
9 PSWAALP-LAVEHTMMYR-----PENVKA-----LRRALR-----ILHPGIWC-W
10 A-RSR-GFAVG--LPTSL--GGYG--L-----PTLSGCP-----
11 ----SRCRLP-KWVQ----VGVAALLFK-----VPWRDII--SPARYWQMVRQNLP-
12 ----HSTMAA-----DHARYKFETS-----AVACRPGKSPF-----
13 -----NGNGHG-LGGLGAPDEVAVRLAN
14 EL-----FMILEDLP--RR-----PELRVRP
15 RTIASGVRKFYQRLTKRYP--VGIRGVKRVGKTWM-----DL
16 QRQW-----DSRQNERLWWTNPRSASEGFLGL-----PR
17 SAKRRLAS-----
18 -----ALGWV-----
19
20 -----
21 >KX883500.1
22
23 -----
24 -----
25 -----
26 -----
27 -----
28 -----
29
30 -----MEAVPVTSRRRCG-FTK-----YFRGSLRSIIDRDC-----
31 --RALQFSYVGRA---LP--EAS--PGQCDAAALTKYREE-LSQPATTP-----DS-----
32 ----ILDSAEWARR---WASR--FLSGPVVPPT-----WLSGG--ACKEVP
33 RAKG-GLGT---RSV-----DVARRATDAICEPSEPEWESLRERLNAVEGLT-----
34 -----
35 -----DCDVDIVCRERLVS
36 NQAL-----VELSELPR-
37
38 ----PIAVDAMA--LPERG-GKVRIVTKCPWALVYL-GHFLR-----VWLLE-GL
39 RRDERT-----KSVLDGDH-AGSVG-----GLVGGAGVRLLE-----
40 ----HLAVSA-----DLTAASDLLPH--DLCQAIVRGVLRG-SR-----
41 -TRPAPHLDEVWADL-IG---PLTI--Q-M-----
42 -PDGS-----EFVNQRGIMMGLPTTWTILSLVHLYW-----A---
43 EWA---W-A-----STVPRHISLSGLPSAA-----
44 -----PRT--AICGDD--LAAV--WPTSVV--QRYESIVKS-CGGQFSAGKHY
45 KS--RTY-LMFTEEA----F-----R---LRVEMR--DYDTR
46 RVDGRRP-----GEA-----
47 TLADFLPS-----
48
49 ----WMVEEPSRDSRW-----CTGVDHLPLVPLRGLVRPL---HLPKDR-----QPL
50 PVWVAVP--HCVEAACTQ-----SGDAGA-----VRRILR-----VLHPGIHR-W
51 A-RSR-GFAVN--APTSL--GGYG--L---P-----PVRGDPK-----
52 ----RARLP-KWFR----WGISSLRA-----TPWKDLI--NPARAWAPVMRDIP-
53 ----HRDMGA-----DHASEKLKVG-----AVAARKGRVPF-----
54 -----M-GRGADLGYYDAEDRLTV
55
56
57
58
59
60

1
2
3 RL-----SNELLFILDRD-----TLPSSFFLQGP
4 RKVSGAVRKMFKQHLRRFP---IPEGSVVASETPM-----ASL
5 RRRW-----EEVRTARSWWSSPRGAAEGFVLGIS-----SS
6 AKRAVASA-----
7 -----LGWAS--MEPAHSGGPVSD-----
8 -----
9
10 >GFNE01046070.1
11 SDLSGVRRPQTRKVLLSRITYVEATRLQAARCRSVLPSVSGRRRTPAVCRSRGGWVTRR
12 R
13 KCRPVAPASNAFLSSRTGGGLSLHVADSTLSSLSQPPSISLLLS-----SKAFSTGLA
14 HSPTLIPLLLMQDPTWPSDRPSPAAPAKAVTTAKHADPRPPGRAPAPQPPNPHPPGPK
15 GRVPARKVGP RPAGAGGGPGAGDAARKHQPPRAPFGTQREHAVVPGLVGTPTRTGQAF
16 RT
17 LC----GAWD-----ALRACIAAAGFVEFDLRE-----RS
18 SRDAVGGLAQLSG-WLTTATRG-----SGLGYTIKTLK-----TWSAQFRSWVF---
19 -----NRTPTHKRRGG---LAR-----FLRGKLRSLVDRDE-----
20 --RALQMSYLGRA---LP--VGS--DGVCEKSLKNHRDA-LTRPHPTD-----PD-----
21 ----LLARAYRFARL---WADR--HLPEVPNPSHV-----RLGHG--ACKEAT
22 RATG-GFAR-----AIVHLAGRDLPENTARQTIMDRLRQEEDLGPE SITEC-
23 -----
24 -----DLDSFVA--ESEL I
25 RAVV-----KPL
26 LDAPRPPRHS AIP--VAERG-FKCRVVTKSPAALTAG-LHHIR-----TWLMA-GL
27 RSDKTS-----STTISGDHRAAVEA-----ILAGDPVPREG-----
28 ----YCFLSA-----DLTAASDLLPR--DLLIALVDGLSDADHS-----
29 -APQW--YFDLLRLA-CG---PQWV--S-Y-----
30 -PWGD-----TILTERGTMGLPHTWCMLSLVHLFW----I---
31 DQA--V-W-----TTPATRIPPALLL-----
32 -----KRS--AICGDD--LAAH--WPDRVI--GRYEDVATR-CGAEF SVGKHY
33 KS--RDY-VCFTESL----AKARCNPVGCNPPTCSGSGGRVPRY--RRTRPR--PAPQT
34 PIKRPYGSASPIPSPLASAVAADPH PALLAKSLGLTPPASAPPSRSIPTPTSKPKGRSR
35 PLLPNLPPKGRRRGGRPV ALRAPDSPPRDPFPYPPTVPGRPDGPRWPRGPKRWPRLG
36 AGSGAKLPRRRRPPRSNRRIMRRVVGVRWSNAVPVRGLIRPNVVPFGREAL--
37 GDTMGPA
38 PAWAVIG---PAFWAIVG-----ENPPPP-----RLNAAR---RALKALHPHLDA-W
39 G-RHNFGCLAT--VPREL--GGLG---L---P-----KSKGDAL-----
40 ----KVGKLP-KWLR----QALAVDLYG-----TRAGESP--PPIGTWSGMKPGHH-
41 ----RSM AEE-----HSRARLQRM-----TIRCRA GKVPF-----
42 -----KGAP-TPATLLGPEVD-----
43 -----VEELSTIAMS R-----
44 ---SLTYILGAELLGDRL---SVNPRRLADRIR-----SMVSR
45 STRW-----AGSRGICQSAPWAKLIARHRLVQNWV-----LW
46 LKPGADEG-----FAGALIDKAAKRS
47 LAGT-----LGWAI--PRASAGGSPARPTSQELSGP-----
48 -----
49 >KX883526.1
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 -----
5 -----MP-----PRKSNGRKG
6 GR-----APRTCVALEAVLSAGAAAGL
7 -----
8 -----TS
9 AEVRTLPIGVLA-----SEDNPDLGALK-----ALSAAYRAHWI---
10 -----ATGDP-LAG-----
11 ----QLSFVGRA---LP--PAP--ASARLAALVQHRED-LTSTFRT-----PS-----
12 ----VLALAREFATR---WARR--HLARAAVPLEEAS-----WPTPS--SCYERS
13 ASKG-GLLR--HLL-----ECTSFEPAPPGFDGLPDGPVTQLWVQSA-----
14 -----
15 -----
16 -----RFREYGLRGLERCR
17 SAAG-----
18 ----PPTHRVAV--LSERG-LKTRVVTVGPAWLQVL-GHCVR-----QRLLR-AL
19 RSNRGT-----YAPLAGARDDEIWA-----SLEGGHSE-----
20 ----AV-VST-----DLTRASDLLPL--DLLSAIIDGLQES-GK-----
21 -MSTL--EVEILRIS-AG---PQRL--E-Y-----
22 -PHGW-----GTYSSSRGALMGLPTTWCLLSLVHLFW----M---
23 DQV--R-----LTSRAGPMRAG-----
24 -----HRF--SVCGDD--ALLA--TTEAGA--QRYSA LVRL-CGGQPSEGKHY
25 VSGGVRCRGRGVFLEKL----L-----E---FDTEGG--RL---
26 -----
27 -----TGM-----
28 RRLAAMPV-----
29 -----KGVTSV-----SLPRDFTGALPIQCRS----KGLRLLV-----TLD
30 SVLEDLG-----PAGLVP-----VRNWVR-----TRASWLGRF
31 A-REQALLTPG--APLRQ--GGYA-----FGDPN-----
32 -----DRTPEEERR----FRAISLLPRPSIGA-AALRSVD-----PL-
33 -----WRLASR-----ETLAGE-----DLPLSSGKVVE-----
34 -----
35 -----IGRHPPLS-----
36 -----GEGPLLPTQWR-----
37 -----QSELDRLAEGYP--AVAIQSQEERHTS-----AC
38 VALF-----AALKSLMGPPLEVYTFR-----VSQ-----LR
39 AAVLKATA-----TG-----
40 -----EALRE--SLGPTA-----
41 -----
42 -----
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

>GEZJ01006036.1

-----AGG
QDNVPQTGGPTRPARNYLLSSVKSLASVGRKV----LGKPRTLNLDDGVLAQMTSGRHK
KPLSKPQKKHL-----

-----RELRRNW-----SATRG
FLELTFGVKLGKIF-----APNNRKAIREQVS----FLLEVLE-----
---SGL---QVTAKRWSSLLRSKSI-----YQGTS
DSPC-TGIGTRGKTPMALF-----LASTYSRACVVDP-----PTKDEV-QTMV
LAAIERISTREPL-----PEPD---VLRHLKGF---IADLFRSQRMKPRRAPVLP---
-----VPAPTSCYERSGSEGGSI-----IALT-RFG
RQDLSEAVHEHIARLFAESQRRAPQEERHDYNWTDLDEPLYEGIPRDPQEMIRCLDRQ-G
A-----MDLMMC-PPPPP-----RLLT
LPETFKKAV---KDSSRPE-----ER-----
-----RA---KALPIVQQDGKIRVATLHMSGVV-WSARAMS--AFLPRLRNLT
FS-----KDMLRNRTVQLESHRRSD-----DLLVYSADLSKSTDPI--SIEL
SRFVLQEITRHTG---KPTWWNDALAVITNHQI-EYTA-----
-----DSVKKTVM---SECGALMGLGPGWTVLSVLNAFAAEAAGA-----
-----R-Q--GDYAVCGDD-LIGLWPTEVCDRYEGNLASLRLKANKEKSFRS-
--THR-GVFCERLVLRSNDRASAR-----
-----SCVRLG
EASGA-----

-----RAIDGGKGRVV
V-----DALREVAS-----SSEAHRIIRQTAYRTIMRMALSQ-
TAKGSMKE-GGGGIGQADAVTLISYLLHGPTQLYKVEDSRIAAMKTLR-----
KAPNLSTGVPVEDVLTAGRAQI-ELNRRLETAERGPRPKYLPSQSVR-----
-----RRIHSRRKDAKKAIDYGGVINAVK-----GLLASGCYVR
ENNRIR-----RDLTHAVRG-----RAYTRALRIARSSWRTKV-----
CPHAAKEAFEVS-----FPEHQPRVTSLEPMP-----
-----QAWDSRLAD-----

>GGIC01436417.1

----KNQRISNKKKIKNNISKSKIYKKIVS----PEVKNYIDKLGLKTYKIEDMEKI
KDKEV-----

-----EVLKDYV-----DEITH
QHIKELKDERKR-----SRNECVKDIRLFV-----YLVRSIL-----
-A-SGPNQLIKGLKEKLTSVRRYCMGDIEEQEIVDKHFSQVLARDQEKRHLSKKERQRR
LDVRSRQKILGLYPQWLELPGIRKFTRH--QHIQMTFISRLP-----EGTRNV-KDEA
IEGHKKMVSSGGK-----MDEE---IRRDLVRW---VREYCRGKVSIDIQLLAG---
-----KWSATPSSCLEATRKEGGHLMGSYQLYNMISPCRHTLF
NQSRSEDIERMNWERHGGDRFCSVSLVATSLMIHESEPLKEFVTKYLHEAKQIPSLP--
-----NVIQERHNLKIAPGFSFRSKTD

CKTFVKQQKRRRLGIKDKVWDEKLGRMVDNFDDRRILSPFLPSGDQSEVHSTLGGILSTE
E
FSDYAKSCFDEGGLRSRVSAIAERGWSRVVTNHSELEI-QSGNIIR--NLLNSILSKDK
RFL-----GSRRTHFDDQEVLDAMGRNSVGG----HPLVLSEDMTAATDSI--YMDV
FATIWN TLCSTIG---LKFSSIHAGLIMCAPRLY-EYKDL-----
-----TKVR---TNRGIPMGTPISWPLLNLANMFCWERANI-----
-----DRAFEI-I--PPVTILGDD-LLGVADLT YINKYSDNLKRIGGISNPRKRVLS-
--VFY-GLIGERIYQFDVEDNEKSLVIKHKETQLEAEQTNYQVHAEIAGKSSIKTLLLSE
DGKRKDLVELGSSDYRISSSQARYTESLKMMLLEDQDCHTGRRSKIVYELEIVKYIESDLI
KEDRKDIEEKLRFVVKKDTEDRQKEIENQIINICKIYKYCYGKGRSFSGIIGPFENYKG
IVEWLFGSIVIMIRGV TYSISTRKDCVTSKDGIISSRRPLWVRQTNGKLEKKAYEEGPV
NLDPGFDGVELHLNQYQKKFKTLLRQWHSPRRNRGLVISKRLKCTSLK-
GIVSGTNKLGA
TDINK---GIKPRWTC-MGSIEHFVLES--LGDSNEVRQILNRIRSIFHPDIIRELRGH-
RLEPFIPE-SMGGGGLVYKNKLDSTRWPCS-SVFKKYINKILFGQNVKDYLE-----
KINTSLV--ISEKQDWWELASE-MG-RADFEDVKIVERLLHPSIEG-----
-----ITIGNHQDLVTENILQYQGDFIRQ--MGDPDFCKKVSKLVRG
KTKQAM-----KKEYESWASMRGF---KGSIEEFIRFRDIYLSRF-----
--MAITIRQEEL-----STYLRRTKGGFPLVP-----KEHTHLNYITTK
TNTSQTDQPRGQAETVLNADVTVVITDGLLETIVPSEPQRRQLIAAA---GRSMHLTHIL
PGLYFYTNPAVWFKMLESTP-----
>GDKJ01026489.1

---STWGGMVAKTGLGSTSRKTAASGSISVC----PPRSTQQKPSGKVAKKVSHNLSQ
LTTLGSLLLVWQDLVKS-----

-----KALTRVQ-----RALHL
YKA-AADGKGNNLKVFREKKRRSRTMKLRFKAIETAFLPILHA----WLLV-----
---NGGASCCSRPNNIGKWMSF-----TSYMM
AMLT-DSKYLKSITHQWREESIRHSNTRVAKRLFQASTLGRAVP----FPHTMLQEDRA
YQGAIDAWTQGPA---F--VSKQ---HLDQIAEF---ISKLPIRRSKEKPG--LHD---
-----L-----MLQNKSCIENTQAHGGIA-----HYLYTQEV
EENKKRRDARLFKAGDLIQGQATTIRTKILMESFSRGEELDSRLLEGQRATMSQFKRL-Q
Q-----EELEDI-RRQYR-----YQDL
L--DKARARFENTCHATQPM-----
-----NPVIVRELG-KLRCVTVHPADEV-ITARTIT--TKLLRVLRGCV
ST-----RDTLRG-QCPTVRTETG-----RPVLYSADLSSATDHV--DHRL
AQFIGEKLLDKMG----LESSKPLLKSIFSPKEL-RFTKKEY-----
-----KKHNGTLTSRGVHMGLGPSWVILCLINMYAAHNAGY-----
-----N-K--DEVQICGDD-LIALGSERAAQRYQQTMKELGMVVNETKSFVG-
--KR--GVFCERFLEIKVDRIGCYASAL-----
-----DVGHIA
EASAV-----

-----RTTSDPKSRLAAVERLYKE---RHLPTLARKVADQLAPRG
VGK-----GPIPMGGNGRGAPTPQLVH-----FMKCGPIRLVQTQ-

1
2
3 ----FSPKLELNASRVPIKDGIPWSGLRTEM MRHDAQMSYFKNESKPLTIIRAE-----
4 HFSRLAR--GYVRSTATVHLKS-QSGTYLSLLPRAARQKVKLLKRF-----
5 -----TSSNTINSNLSLPKATRAALA-----
6 -----KALTGPTAE-----RFVNLSTASNLLGLHH-----
7 -CPDADIDLALM-----RSRAKMTNDGIRLGPK-----
8 -----EAELRRITNLDL---L-----
9 -----

10
11 >KX883566.1
12 -----

13 -----MPLQSMPSQKVNFGSLSAVA----ASLRRRLAREGRTVGSSQYATFAR
14 LAEKLRLGLSPRQRRR-----
15 -----
16 -----CLSTASW-----VAGGL
17 SVSTTDRSKGS-----SLGHRLAVVKFAI----KLWNNII-----
18 ---SDVEQFVVDYKRVCSAAYA AAV-----RG
19 YTSIVAAGKCRDYLHRAFGPCGIIPHKGRR-SLFALASTTRCLP---PASLTPSRVKIE
20 LEALKSRLTSVPP---P--VSGA---LLSELVSF---SHDLLRDAGPPVQVEVTGN---
21 -----ALGPPGKTSLDRAQLYGSYA-----
22 --FKGVIMSKTL-----L
23 M-----MSKSAI-PPDVP-----FKED
24 LEENVQ-----
25 -----RVSLLAEGDKVRAITVPSSLAVQLAGRSIN--RALIRCLRRSD
26 VF-----GPGFGKRPGTLMMSGANLG-----DKHIVSCDLTAASDYL--NQDV
27 AIAVLADYLRSHA----SGIWYPWASALLGPQRI-RS-----
28 -----RD--DTYT--TSRGVLMGTPLAWPILSICHAFACYKALG-----
29 -----NTY-Q--FLCRIVGDD-GLLICNASQYRAYVSVMNRLGFFVINEHKTFQG-
30 --KQC-GILAGRLYVLK-----
31 -----TRRAKF
32 QDLNS-----PT-----
33 -----

34 -----KTMAWRMVDCQA-----PYVP---ELLAARGA---
35 -----QRFRGSDA-PTAPIRVPGLS-----QRENKAALRAFK-ILNPRTFTDAMLRA
36 GRPTPLTT-VQ-----RRGLLVLKNAAP---SSGARSLWKSSLAKRGHYASLMTSYLR
37 QLMDINL--PACSKAPVNGSYL-AR--MSGTVSQLVTARLNLPKDSR-----
38 -----MILRPSYVLSRLGKRLDKK-----
39 -----MHTPKQSSV-----LRLLRGLEAE--MVPPSH-----
40 ----VSVMEKL-----LRSMVGPDDGARPISG-----
41 -----VIGDFSQGLI-----TSPTELYPQFL---HWIEEGAP--
42 -----

43
44 >KX883471.1
45 -----

46 -----KSAPNGK
47 DTRLKPAKPAVIKIDLSHPAGSRVDPKVADKN----KPRLEFASLDSLQVVSLLARSSL
48 GVQRISTIPLSPDPSKSAA-----
49 -----

50 -----QLFRRFAARGKQNIQAFQRLW-----STVLN
51 AV--SLAFKIDQ-FDLRTPQ-----GRIAYTDLQCAFR-----HLLARFT-----
52
53
54
55
56
57
58
59
60

-T-VGYQDSVRVVAFWALQARQMAL-----NSAVA
TSG--HRPNLVKFLRGRDLIRRASQR----TKFLFSGVGRALP---FAKIDAIR-IDAE
IKAHRARVTV-PG---T--TDPL---ILGKYAEWLQKKLNLLDSRADLPSMKLNIA---
-----PL-----TAARSSAVGAPSSKGGRY-----GFYK---
--SLFEQGPQFQ-----ERILEH-A
I-----KEVEDC-PTPLP-----V---

-----KLTVIPDKGDKARTVSTTLPGIQ-VLGEQLN--QSMMVLLRRLA
PS-----AIPLSENPDVSRRLNNG-----HGTMLSTDLSAASDYI--DHTV
LKVTTRLVARKLG----WQFNSKLVRLLTGPFMF-TIDQ-----
-----EGVKDSYL---GLRGAPMGLPLVWPTLSLMNMFCAEYRIS-----
-----PNS-K--RSYAVLGDD-HVARWPKADIDQYFRNMSDIGLKINKAKTFLS-
--KRD-AVFAEDLYSLVNGKVKR-----

-----TVVPKLSAVMSTRQQ---
-----GTDPQIFA-IIASLQSALAN--AKNFKERKQIKKIFF-SQWKKLTRQIRRA-
QVPLQLPT-QLGGLGIAP----FKLSKLT-RKQRSRLAYLLSL---PEYASKSARSDQR
SYFRSLD--QFKNANRFAHAIR-NKDKLVKEAARQVI-EQISDEEERTSR-----
---SFNATGCAVEDCAPQLPEGGVPAEAVLKELLALKLTRL-----TQEGK
VQATDR-----SVSSGTVLR--RL---TRVWERLPLCLKPVGARG---YTKLLS
LCVQSSQILSTQ-----SVAKLITTNLPLYAV-----
---GVFYKDGGMPIQCS-----PNSNDAYDSAT---AG-----

>LC150604.1

-----MASQFCDSIA----RRCGEVIRITLHRSNDHTTSPPH
GGPRLEGNLETFPFAKRD-----

-----KHAKRVM-----QAFYA
AYIAMGYHCHLR-----DRSLTRFFEKLYI----RAWH-----
---EGLDRVCPDIKKQVTLLRS-AL-----IAETG
VSPDVLGKLLPRSWANLAHPR-----KLFQLTCLGRALP----RPTQKV-IESS
KREALDGFTS-PA---SVTTDIN--VLHELVP--ALNLRRYSKADTTIFRGS---
-----SSTEQSTLGPVFGTSATLFNSRKEGGRA-----GIIH---
--DACRSRPRQV-----GAIVALKRQA AIRKTAA-P
LR-----FDIRSL-KSEPL-----IGED
SRRVLNDVRHRFPNNLGES-----
-----EIAVVPGLGFKARVVTKSHPIV-VAGHQLR--RRLYPTLYQSR
VF-----SRALGDKPKALRFARRS-----GAVFFSADLSRATDGL--SHHT
VGVFCRNAEIDPD-----IIFRNMSV-----
-----DG--HALK----RGIFMGLPMSWTILSYIHRAVCDSVDP-----
-----I--QNYYLKGDD-LVAHWTAARQISLYRERVASVGMPLNESKTFVG-
--PRK-AVFCEGYYESKASFNKK-----
-----RMRTEI

1
2
3 WLVQR-----
4 -----
5 -----
6 -----PSISLR-----
7 -----RFFPDSEA-AALRLSEGLDE--VDYNLRSQYNRLVY-CFYPPNYIRISRKI-
8 RLNPFLPP-QLGGSGLLIPDLNSQVHS---VFDRMRLNSIFGGVKRFNASI-----
9 -----PVGGSSSNMKQVN-IL---LSRVRMVP-----
10 -----RGSGLTCSHFDKWLEFATSAAAFR-----
11 -----DAKIGIFPK-----DPSDHTFFAEAKRISRDR-----
12 ---RRGVDFPID-----YRTVRSVCRSLEPLLS-----
13 -----TVPLRHTCE-----
14 -----
15 -----
16 >GBSG01013692.1
17 -----
18 -----RMVRHTNKRLTSFVKSP-----CGAGSDSRKNDRTLHRGRRK---
19 -----
20 -----
21 -----SVEQQVW-----TACYS
22 SLVHS-----GWDNTNMAWFLHD----WVSKTLY----
23 -S-RGVRFTVGELKKLSHGVRFASLRA-----TGTVP
24 DAPRIPKKVYSAFLYLSVREKR-----NAFAFSRLSRGLP----CPPKED-EIKA
25 LRDAKEMSVSYHP-----TSGK---AIE-----AIMWHILQERGGKT--RLP---
26 -----D-----RLPNTLSSCFERSTTQGGVD-----GHLR---
27 --HLGFGEINI-----QCQYMGSCSSADMIRAKYSSFATDTLGQF-C
28 L-----KSICDV-ADNFD-----STKD
29 DEALRCLGVCLKLRSDRPD-----PR-----
30 -----C-----KSHCLRTPGYKCRVIGVPDALTF-VEGTWVRWSSQLLPRKHFD
31 SK-----NTPPAVKGRAS-----DKRYCSVDLSKATDGL--SHDA
32 VRAVIHALAVSGR--IRKEDLSMSLRSLGVGDTEM-IW-----SWRIP-----
33 -----KEPVEELR--CVRGSPMGTPLSFIVLSWINSCTESF-----
34 -----THSSIHGDDAVGTMYNTAELEEEYKEFVRDIGATVNVSKTYIS-
35 --PTS-FTMCERMYAPE-----
36 -----KTKGMF
37 TAFCP-----
38 -----
39 -----PSCPIPGGVVPVPPQ---
40 -----AHIPNLYL-----RRAERVQR-TIHPWVSKLPILR-
41 ----LPT-SVGGFGY--TGYGLKVSKAT-RCRLATACSRNIADVTKEILKQSTFRGEG
42 LFPRSAV--PTPKNSSKYYKYR-NL---IHKDKRFSTVGS-----
43 -----EEFEDVVLASDFVTFREKVILQSY-----
44 -----LAIGGKLGR--GS---SERKPERTRARALFKGKA----PCNI
45 APLTVKRGIGAL-----SKLSDRVKNQKVRVR-----
46 -----TDVACVIRGRT---THSK-----
47 -----
48 >GFLP01591397.1
49 -----
50 -----AYSTTMRELLPSMLTYVRVGESHPSPS-----ERTGPGSRPRDRKSRKAGSRGKR
51
52
53
54
55
56
57
58
59
60

1
2
3 QRA-----
4 -----
5 -----SREDLVW-----QAVWL
6 GFIAF-----GISYQLGCWRTRE----WVKRSIR----
7 -R-DGVHNTCSKLDLAVSIRG-----YVLSD
8 VQPP-MVPGVPNRIQRWLRAEVDKDTL---SGLAFTRFARALP----KPDFRE-TANA
9 MASHLDLITDRRV-----VPEG---IQDQIEQHVV--VRMFGGQFVGILRH--PST---
10 -----PISNHAVGERSKARGGYN-----AHYIY---
11 ---DCAGDPEFL-----GVDPPDRIPAYPPTRALWRLADSP
12 PG-----EGIPEY-MASAS-----ALSA
13 ARIVEQFAR---DTSRVVH-----
14 -----SATGIGEQQDKCRIITVPPASLF-AAGDVCR--SRIWPRVQQSD
15 AR-----IFGNDEDAKSSLSQMGLPR-----GSVYVSADLTATDGF--AHDA
16 VRVLRGLARCG----
17 VDNHTVKLMSETLGVGDRQHYVKYKLAELPRRKREWAVQRFET
18 IEE-----DGVSTHVL-VPLVRGILMGTPCSFIILSILNGWCCMPLG-----
19 -----PRTVICGDD-VASACTPEAVDNYDRRVTVIGSGLHKRKTFIG-
20 --HKG-LLFCELYVLPGYGEHRCF-----
21 -----
22 -----
23 -----
24 -----EPVPLK-SLAKDGDGT--
25 -----LDTSDFDTFIW-----KRMDRACR-VLWRDVRARARL-
26 NRWPQLPV-ELGGLGHP---SSGKMGA VP-GSVNRNLATLIKVGVPVKKIAR-----
27 ----WT--VAAGPLDWRSFRS-DKEVAWSLFESTDAAVSQLEADSR-----
28 -----YVDTSFVSYSHEANQYVSVLANQLY-----
29 -----SAHGGKFVN-----EFNRKAMKPGKIQYPGVG-----S
30 LQYSAKAPMSMV-----AREYLAKLHAGGEYLP-----
31 -----TDSVRKIRRI--GKST-----
32 -----
33 >GGCO01034162.1
34 -----
35 -----SYTKNMKVRLSSHLTRLEFKDSTAPS----PRKSGGCKPLSKIRTKHWHSDKE
36 R-----
37 -----
38 -----TPEQIVW-----EALWA
39 ALVVSDSVNDL-----DVSSMRGSFYLRD----WVARSSR----
40 ---RTVGYVVSVLKEELNRLRAISL-----G
41 SDPPPIAHGFPKKLDFTLQKVFTPL-----GLFAFSGVSRGFP----TCVDKVKVSET
42 LVDHLNVNANPAP---A--VPIN---FLRSIEQF--VPRYIQKSVCYDRKLADKM---
43 -----HDPSPVVCKSSPSACYENPRGAGGFY-----AYIK---
44 --KLGDKISEGQ-----NDLLADIRAGVKPRPEDICHPIKGIKASYDPDRTQDLLSP
45 -----KEMSDI-LDHFY-----FYSG
46 CNEYLDNPDKEIVH-----
47 -----RVAVIPERGYKNRVVTAPPASIL-SMGEVVR--SSIFPYVKAHP
48 SC-----DIVKDGDMPVVCFAVD-----GAKLVSADLTATDGF--SHDV
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 IRSVGVLNRAG----FPQHVVDFFVDTLGAGERPHYAEYDVHAILRRHSKNTVKYNSL
4 CSFLQSYGWDGISSKVK-VPMLRGSPMGTPCSFTLLCIVNSWALDHCS-----
5 -----NGPKICGDD-MLAYITAKEFSKYKVRVAAIGSGVHPLKTFIS-
6 --PYA-GTFCENIYIKDPCDSKFP-----
7 -----SVVAFG
8 RAIKC-----PIKAMLYPQKGSNG
9 RIDWP-----
10 -----DEIVHQVFTGETRMSKEKVCLSSF--NDEP---EYHVQRRFEAG
11 PGDFG-Y-NDCPSVFR-LG-----RKVHRVLR-TLRRDQIRSALKK-
12 GRYPWLPA-CLGGLNY--PRKNPKGVRGLD-KHCRARLFQFTHSDLSVVELGKF---MS
13 SLDNAKR--PLPEHRELEFISD-MV----ADAEIND-AIVEDTLER-----
14 -----QGECSFVDSYSLARHQGSVRSNFF-----
15 -----RSMGGKLDL---S---ASSYSIFNAARIRWPAVK-----GEL
16 GTYDKRTPFSLI-----IQDLNLLRAPRYRVL-----
17 ----SSFAEKLPGVEHPK-----SNRWHHLHGAG---GWNDGIL---
18 -----
19 >GAIS01005902.1
20 -----
21 -----EWRLQALSK----IRTKHWHSDKER-----
22 -----
23 -----TPEQIVW-----EALWA
24 ALVVSDSYNDL-----DVSSMRGSFYLRD----WVARSSR-----
25 ---RTVGYVVSVLKEELNRLRAISL-----G
26 SDPPPIAHGFPPKKLDTFLQKVFTPL-----GLFAFSGVSRGFP----TCVDKVKVSET
27 LVDHVMNVATPAP---A--VPIN---FLRSIEQF---VPRYIQKSVRYDRRLADKM---
28 -----HDPSPVVCKSSPSACYENPRGAGGFY-----AYIK---
29 --KLGDKISEGQ-----NDLLADIRAGVKPRPENICHPILRGIKASYDPDRTQDLLSP
30 -----KEMSSI-LDHFY-----FYSG
31 CNEYLNNPDKDIIH-----
32 -----RVAVIPERGYKNRVVTAPPASIL-SMGEVVR--SSVFPYVKAHP
33 SC-----DIVKDGDMPVVCFAVD-----GAKLVSADLTAKATDGF--SHDV
34 IRSVGVLNRAG----FPQHVVDFFVDTLGAGDKPHYAEYDVHTILRRHSKNTVKYDRL
35 CGFLQSYGWDGHSSKVK-VPMLRGSPMGTPCSFTLLCIINSWALDHCS-----
36 -----NGPKICGDD-MLAYITAKEFTKYKVRVAAIGSGVHPLKTFIS-
37 --PYA-GTFCENIYIKDPCDSEFP-----
38 -----SVVAFG
39 RAIKC-----PIKAMLYPQKGSNG
40 RIDWP-----
41 -----DEIVHQVFTGETRMSKEKVRLSSF--TDEP---EYHVQRKFEAG
42 PGDFG-Y-NDCPSVFR-LG-----RKVRRVLR-TLRRDQIRSALKK-
43 GRYPWLPA-CLGGLNYP-RRNPKGVMG-LS-GLCRARLFQFTHSDLSVVELGKF---MS
44 TLDNAKR--PLPEHRELEFISD-MV----ADAEIND-AIVEDTLEK-----
45 -----QGECSFVDSYSLARHQGSVRSNFF-----
46 -----RSMGGKLDL-----
47 -----SAE-----
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6 >GAIR01012062.1
7
8 -----YTKNMKVRLSSHLTRLEFKDSTAPS-----PRKSGGCKPLSKIRTKHWHSDKE
9 R-----
10
11 -----TPEQIVW-----EALWA
12 ALVVSDSYNDL-----DVSSMRGSFYLRD-----WVARSSR-----
13 ---RTVGYVVSVLKEELNRLRAISL-----G
14 SDPPPIAHGFPPKKLDFTLQKVFTPL-----GLFAFSGVSRGFP-----TCVDKVKVSET
15 LVDHVMNVATPAP---A--VPIN---FLRSIEQF---VPRYIQKSVRYDRRLADKM---
16 -----HDPSPVVCKSSPSACYENPRGAGGFY-----AYIK----
17 --KLGDKISEGQ-----NDLLADIRAGVKPRPENICHPILRGIAKASYDPDRTQDLL-S
18 P-----KEMSSI-LDHFY-----FYSG
19 CNEYLNPNPKDIIH-----
20 -----RVAVIPERGYKNRVVTAPPASIL-SMGEVVR--SSVFPYVKAHP
21 SC-----DIVKDGDMPVVCFAVD-----GAKLVSAADLTAKATDGF--SHDV
22 IRSVGVLRLNAG----FPQHVVDFVDTLGAGDKPHYAEYDVHTILRRHSKNTVKYDRL
23 CGFLQSYGWDGHSSKVK-VPMLRGSPMGTPCSFTLLCIINSWALDHCS-----
24 -----NGPKICGDD-MLAYITAKEFTKYKVRVAAIGSGVHPLKTFIS-
25 --PYA-GTFCENIYIKDPCDSEFP-----
26 -----SVVAFG
27 RAIKC-----PIKAMLYPQKGSNG
28 RIDWP-----
29 -----DEIVHQVFTGETRMSKEKVRLSSF--TDEP---EYHVQRKFEAG
30 PGDFG-Y-NDCPSVFR-LG-----RKVRRVLR-TLRRDQIRSALKK-
31 GRYPWLPA-CLGGLNYP-RRNPKGVMG-LS-GLCRARLFQFTHSDLSVVELGKF----MS
32 TLDNAKR--PLPEHRELEFISD-MV-----ADAEIND-AIVEDTLEK-----
33 -----QGECSFVDSYSLARHQGSVRSNFF-----
34 -----RSMGGKLDL---S---ASSYSIFNAARIRWPAVK-----GEL
35 GTYVKRTPFSLI-----IQDLNLLRAPRYRVL-----
36 ----SSFAEKLPGVEH-----PKSNRWHLHG---GWGGGMMGY-
37
38
39 >GACI01002802.1
40
41 -----
42 -----VRLSSHLTRVDFKDLPTPL-----PVKGGGSKPRSTIISRDPRHR--
43
44
45 -----RPIQIVW-----EAVWA
46 ALIISDSYDGR-----DVSSLRGSFIFRR-----WVARSSY-----
47 ---RSVGYVVAILKLEMNRLRA-----ISLKS
48 SPVPGVPEGFPKRLDQFLDNISPF-----GLFTLSTMTRGFP-----VTTDKSDKAKA
49 LLEHLENISTPCA---P--VPTP---YLRSEDFF---VPRMFNESKFAKNARDFTK---
50 -----NVQFPAPDTIKLSSSACYEGPRSRGGYL-----GHIK----
51 --KEGDDDAKDRIANGYQEILKDIRDGVRPPPEND-THPLLFNIAKRSYDPDNLISDL-S
52
53
54
55
56
57
58
59
60

P-----SRLAEI-MSYYY-----YWNN
CDKYMKSG--HRDEQVH-----
-----RVSTIPERGYKNRIVTAPPSGVL-SVGEVIR--SIIFPFVRDHP
SC-----TLVRDGDRLGARFYDRT-----GCRLYSADLTATDGF--SHEA
IEAVGRGLLSYG----FPTSVVEHFLGTLGVGSK-HYAEYNVADLLQGHKEGSRKYNQV
CSYLMEKGWDGN-SKVMRVLMMRGSPMGTPCSFTLLCILNLWSLDRAC-----
-----NGPRICGDD-MLAYITPHDFAHYSRRISIIGSGVHVTKSYVS-
--DIA-GTFCENLYVRSQTQD-----
-----GIVSFE
RGIRC-----PIKSVLFPQKGSNG
RIDWP-----
-----DPVVHQVFTGETRMEEPYRQVA---TVVPTEQGYVHTRQRELH
QGDFLF--EDLPLTVK-IG-----VRVARVLR-TLRRDTIRACLK-
GRNPWAPA-CLGGLNYP--RKNAGVKGF-KWFRANLYSFTHSDLSVPELSKVL-
QIDCEKK--PDSLSKELEFVST-MV-----ADTEINESLVPEIHDR-----
-----VGDCAFVDMYALSRHYGGIKSNFF-----
-----RSMGGRLSV-----ASSYS-----

>GAIR01012025.1

-----SYMKNMKVRLSSHLTRVDFKDLPAVP-----PVKGGARKPRSTIKSRDPANR--

-----RPIQIVW-----EAVWA
ALVVADPSDGIR-----DVSSLRGSFHLKD----WVVRSSH-----
---RSVKYTVEILKAEMTRLRAHSL-----L
STEIPVPHGFPRKLDRLRAAFAPM-----GLFTLSTMTRGFP-----VSTDPKQRS
LMGHMENVMSPPP---S--VPAH---YLRSEEF---VPRMFKESSYKRSDDLRI---
-----VNPAPVCNKTSSSSCFEGPRSKGGYF-----GYIR---
--TKGDKAGEGK-----DIAKEIREGTTLPPEKTCLPLFRGIAEKSYTMEFPA-P
L-----DEFADI-ISHYY-----FLND
CEDYLNDRDPSQQLVH-----
-----RIGVVPERGYKNRIVTSPPSGVL-SAGEVVR--HVLFPFVKSH
SC-----EIVRDGDMVNASFYNRP-----YCSLYSADLTATDGF--SHEA
ILAVAQGLKTVG----FPTSVIDIFLSTLGVSEI-HYAEYSIKDLMAGLRVGSRKYNR
ASELMSRGWDGFTDKLR-IAMRRGSPMGTPCSFTLLCILNLWALDAK-----
-----NGPKICGDD-MLAYITPSEFNSYSRRIAAIGSGVHPTKTFVS-
--RSA-GVFCENIYIRREDSD-----
-----GIACFN
RGVRI-----PIKSILFPQKGSNG
RIDWP-----
-----DAMVHQMFTGETRLSESYVVE---TLVPEEQGYKRIRHREL
DGDRF-Y-EDLPHTMR-IG-----VRIRRVLE-TLRRDAIRACLK-
NRVPWAPA-CLGGLNY--PRKNPKGVKDFP-AWFRRLDLYNFTHSDLVGVGELSKV---IA

1
2
3 AIDCQKK--PSKGSEELGMLST-MV-----AENDINE-AIRDDIGED-----
4 -----VGECVSDVYTLARAYGSARARFF-----
5 -----TSMGGTLDT--K---ASSYSAFCVPKIDWPKGS-----G
6 GCYDCKTPMELV-----IKDLDNLLRSPNYRIT-----
7 -----RDFS-----
8 -----
9 -----
10 >GGCO01105932.1
11 -----
12 -----YMKNMKVRLSSHLTRVDFKDLPAVP-----PVKGGARKPRSTIKSRDPANR--
13 -----
14 -----
15 -----RPIQIVW-----EAVWA
16 ALVVADPSDGIR-----DVSSLRGSFHLKD----WVVRSSH-----
17 ---RSVKFTVEILKAEMTRLRA-----LSLLS
18 TEVP-VPHGFPRKLDRLRAAFAPM-----GLFTLSTMTRGFP----VSTDPKQRSAE
19 LMGHMENVMSPPP---S--VPAH---YLRSIEEF---VPRMFKESSYKRS--DDP---
20 -----KIVNPAPVCNKTSSSSCFEGPRSKGGYF-----GYIR---
21 --TKGDKAGEGK-----DIAKEIREGTTLPPEKTCLPLFRGIAEKSYSYTFEPA-P
22 L-----DEFADI-ISHYY-----FLND
23 CEDYLNDRDSSQQLVH-----
24 -----RIGVVPERGYKNRIVTSPPSGVL-SAGEVVR--HVLFPFVKSH
25 SC-----EIVRDGDMVNASFYRS-----SCSLYSADLTATDGF--SHEA
26 ILAVAQGLKTVG----FPTSVIELFLSTLGVSEI-HYAEYSIKDLMAGLRVGSRKYNR
27 ASELMSRGWDGFTDKLR-IAMRRGSPMGTPCSFTLLCILNLWALDEAK-----
28 -----NGPKICGDD-MLAYITPSEFDSYSRRIAIGSGVHPTKTFVS-
29 --RSA-GVFCENIYIRREDSR-----
30 -----GIACFD
31 RGVRI-----PIKSILFPQKGSNG
32 RIDWP-----
33 -----DTIVHQMFTGESRMSESYSVVE---TLVPEEQGYKRIRHRELT
34 DGDRF-Y-EDLPHTKK-IG-----VRIRRVLR-TLRRDAIRACLKR-
35 NRVPWAPA-CLGGLNYP-RKNPKGVRD-FP-AWFRRDLYNFTHSDLGVGELSKV----IA
36 AIDCQKK--PSKGSEELDMLST-MV-----AENDINE-AIRDDIGED-----
37 -----VGECVSDIYTLARAYGSARARFF-----
38 -----TSMGGTLDT--K---SSSYSAFCVPKIDWPKGS-----G
39 GCYDCKTPMELV-----IKDLDNLLRSPKYRIT-----
40 ----RDFSGKIQGAHLK-----SSRWHLRLRGAG--GWNDGDL---
41 -----
42 >MH213236.1
43 -----
44 -----MTASLRVYRWTLSTDITAWPVTPGV-----GGVGSWISVIPGPATPCVEHM--
45 -----
46 -----TDRRAIW-----TAVAI
47 TLTAVNPK-----GYRNIIPPKLLAK-----WVLR-----
48 ---APVQTIVKGLKSLAHEYRSKAV-----L
49
50
51
52
53
54
55
56
57
58
59
60

GGPLHFVPEVPLVFQKWLKGHVCHAT----QYAQLGRLGRLMP----KPWQKI-TSKA
VHEWIRVNTRPDGPIHE--PSVDNERMSNDIYRF---AHEFATQRKDRLTTEVNFT---
-----PSESASTFMTSMSEGGRL-----KEMM----
--NDARDLIGEF-----AFDETIVNLAAREGG
Y-----KPPRAQCGMQP-----VLDE
L--ILENQIQMADSKEFSP-----
-----RVIGVPEYGWKTRVLTLPNYAL-TPGDLIR--QQLWPLVSEEE
WLDAD----RVPSWNLFNNMLSRSLDR-----VGTWVSCDLSNATDYV--PHLY
CKALWAGIYDALG--VSEGSWVRRYTEKMFSPMVI-AG-----T-----
-----KG--R--I---TFRGTQMGTPLSFMTLCLLHRFAVVRSGF-----
-----E-Y--CPHLIRGDD-LIGCFHKP--RIYMGMLEELGFKINKAKTFLS-
--TIG-GTFAERTFRAERKLIKVN-----
-----RHGYFF
SHRKV-----

-----TAGQWAYTRATLL---QDIPLR-GLIKVATTGPN
-----RYRSKLLM-LSEIVENVAMG--FRPKRRALIRRLAT-QGRGDILIAALKS-
GIPLSLPR-EVGGAGLPTVSEILRDRIYID-LETRRRLGWMISHVPGKGPQLVAK-----
LRRRASGGLIGFTERYLEGIKN-IS-----
-----LGLRVPVTEPERTAHYLAARRTI-----HL
KPWEVV-----HRPPCSFRE---W---GRFTRSFSGKVGPRWALGG-----
-----DRLSLV-----SDRLKVATACIVYVPI-----
-----YQGFRGPIRRISD-----GIWTEDPGQNE---GYRHQPSDCS
AMS-----
>GFJU01140648.1

-----LAKPMRWFLPKNPLSWPAIPMVE---VGTASDVVKALTR-----

-----EDVKTVL-----SFIE
PCRIVINSIHR-----PWLDRPYRSFVN----WVCCMVH-----
-E-KQEQAIAATAIKDLSFAVRMWAI-----EGRFP
ETWKSRSAPHLPTFVRWVKSFLCHTG----TRVQFARLGRCLP-----VGSRAV-MDTA
RANFLAVVHTEPK---D--VSVD---IADELYEF---GRSFI---VSEGLKPVSWE---
-----PNESTATLNFHASQGGRM-----AELI---
--YDARLPLHLL-----
-----EEFTPI-TEPNE-----LLDE
V--IVETAIEKSRIEDWEP-----
-----ELLVPERGYKARVLVKFPASAL-LVGDIVR--RQLWPQIEQAE
WLDLD----REVNSEKLTFLRRCIDPNTG-----LGQCVSADLSNATDYI--PHLY
AQALWAGILEKVD---APDFVKGYVEKMFSPMKI-----
-----EG--R--M---TKRGIHMGTPLSFMTLCLMHRYALRKAGL-----
-----H-R--HPHIIRGDD-LIAIPQP--DIYFSAMTSLGFKINKSKTIVS-
--KNG-GTFVERVFKFEREDVI-----
-----IPHPIL
RLPPQ-----SV-----

-----ITRALEV---VDIPLA-GMIWPSEGRSK
L-----RTIGRWYN-QCQSRLTPRMK-----RTVSMLIY-MRHKDILAKATKL-
RLPISTPV-ELGGCGLPNRHGRNDFDAP---HWYRAIVGWAACHESMAKKFSG-----
RVTRLDG--KDPNPPYYIQNIR-ENQSRSTYEYDPTTTRYRGMKRR-----
-----FQIFRWAGALDEDRFKARVEREQS-----
-----RFISLPLNQ---W---YEPFKNLKRVRPRWVPRR-----
----RADTDRL-----IRRIKMFNAEYVLPYT-----
-----SRVQFLGSDSY---KTGGDT----

>KX883548.1

-----MTKSMRRFYLPHEPWRWPVTPTPE----VGRGSVPTLADRDMR-----

-----DFARRVW-----HSVVT
PIRMVTSGRLSR-----IAEFRSFAT----WMVRKVA-----
-A-GERSVVCRELKNLSFSFRVWAV-----E
DRWTFHSDQLPRSFVRALKGLAPCSE-----ARAQFARIGRSLP----TGDGKV-VEAA
IRSYQRIVVGRPN-----PSWD---VADELYAF---GKRWAEERAHLVRGRVNW---
-----PNENTASLNFKASEGGRL-----AELI----
--ADSAAVIEPL-----
-----LEFTPE-GEGSE-----LLDE
V--LIETSIEKARTERP-----IP-----
-----KVIAVSEKGYKARVLVEFPASTL-LPGDIIR--RQLWPMLEEEE
WIDMD----VVPNEERLKAFLRHSVDH-----SATCISSDLSNATDYI--PHLY
AQALWRGVLD AIP---APEWVG DYTDKMF GPWDL-LF-----P-----
-----DG--KTVT---SQRG IHMG TPLSFMTLCLFHKFAVERS GF-----
-----S---H--H PHLIRGDD-LIGLFPDP--ARYCRTMEDLGFKINKSKTIIS-
--SKG-GVFVERTFRFTTGREPRF-----
-----ERRSLA
SFLPD-----

-----RMVIRETTEL---KDVPVA-GLLNPSVEGSL
L-----RSVGRWYGQLPDL-----ARRRKRLHTVIW-SVFGDTLRKARCS-
GIPISTPV-QLGGCGIPNRHGRLDIDAS---FKMRRSMALLNKRSVARQFA-----
YAVSAIDNLPHTDFDRLFLKHH-EESQRSRSVYTYDP-----
-----QSTGYYSASRRFKMYRLLGGGAE-----
-----SVKRRGLRE-----GKLHKPAGFRRWVRLFSNLKMSCPR
YAVGRRANSSRL-----ETYLKISFGTYIPVAS-----
PGYSIFGPGNGEGGVA-----HQAPQLVVALP-----

>KX883542.1

-----MTTSSRYFDLPSNILSWPVTPGV-----VGAGHWHQLPVSALTDRDGV---

-----RYSNSVW-----KAVVT
NIVLAHNK-----KFRNVVPAKIFAG----WLLRCSR-----
-NAGGRYYVAKQLKQLSFYRQWSV-----T
GRRPTPKSDVPRTFQKWLGHVCHAE----ARAQIARIGRCLP----EGDTLV-QSKA
IKAHRSIVTRPPC---D--VNME---LADDIYYW---ARKFSTDNAHLFKEEVSWA---
-----PNENSASQGATILEGGRM-----EELI---
--RDAKAESDFE-----S
V-----KSLAEGDFDPWV-----SKAL
DDAVYMTACTVPPKKEWIP-----
-----DVMAIAEKGYKARVLTKFPAAAL-VVGDVIR--RQMWAMVSDKP
WADMD----RQPDDVKFLSTIRSALDR-----HGICVSSDLSNATDYI--PHIY
AQALWAGLLEPHH---VMPYINSYCAIMFSSMKL-RY-----P-----
-----DG--TVVT---SARGIHMGTPLSFLTLLHRFCVEKAGY-----
-----G-S--YPHIIRGDD-LLGIFRPEV--YFNVMMQQVGFSINRAKTIIS-
--RTG-GTFAERTVRFSHSLATQE-----
-----LSNPLK
RTLQG-----

-----FIPVNISSVKVL---QDLPVG-GVVRATPGKGS
LV-----KALGRWFS-QTSNIVP-----RQRRKAYRAIG-LNHGDLVIRLSTT-
-VPPHLPL-DLGGAGLPDRKGRVGLNG-VP-FAIRAAIGHAASHHETAVKLTG-----
LIARTDG--VSRGFVDVFSKKD-FREQWSKSIWTTEP-----
-----QETEFYQRTTRYWRYFGHRDRPSR-----
-----PISFRQWRS-----GLMTLPRVKARWVPRS-----
----NSDPSRL-----ANRIRSMGTGYIPYTQ-----
-----PTTGSTY-----QWAAKCG---

>GGQW01011558.1

-----AEHYE--RFHLPVNPLSWPVIPGV-----VRAGGWYNLPSEALT---GEG

-----KFVKS VW-----EAFVV
---VFVASHKK-----TIKNLVPPKIFLR----WVMSCSK-----
-NETSRDFCCEQLKKLAFIYRRWAV-----EDK--
--YPTDHCSIPSSFLKWLKGNMSSCT----QRAQIARIARCLP----KGTTVR-MERA
IEGHKKIVSI-PA---R--VDID---LADEIYHF---AAKYTAERKEKLIEKVSFH---
-----PRENTASLEFTEKAGGRL-----ADVM---
--ANA-----SDWLRRR-P
I-----ALSEHL-YTNKS-----YLDE
A--IVEAAV---KQDLAYP-----YY-----
-----RA---EVVAVAEPGNKVRVVKFRSVPL-LISDIIR--RQLFPIFEDE
NMDFD----QDIRSEKFGKMIRCLHN-----EGVIVSSDLSNATDYI--PHHY
AQALWHGILEQFD---APEWVLKHVLRMLSPVEF-NY-----P-----
-----DG--SVIM---SQRGIQMGTPLSFMTLSLLHKFAVERSRN-----

-----G-E--SSYLIRGDD-LIGVFKYP--RSYYSEMENLGLKINANKTLQS-
--HRG-GVFAEQTVKVSWKSTRTS-----
-----YHRTLY
DYFPA-----AE-----

-----TDEVSRIWVL---DDIPMK-GLLHVDPGVYG
SKL----RMIGRWYS-QWSSYYPICSR-----KRAYRAIG-IAHKELIKTARFL-
GLPSSTPL-ELGGCGLPNRKALDAS---FLHRGQIGYAASH-SDSRFS-----
HFVGKLD--GIVKDDGLLEQSI-ER-----SRRSLGSSVFETDPF-----
-----NTEFYLKARRARAYLSPDYLKE-----
-----RERCTPLSQ--W--KKSFRKLPTVKARWVPKR-----
--DKDSFGIDRL-----IDLIKIFNGRYVPFTP-----

>GBDZ01000351.1
-----S
SPDNTGAMTSQCETTYGLPISVLSWPVNPGRV-----VGAASWIPNPRALTGVERGRAL

-----AYAQGVW-----YALTS
NIEAVHPR-----PFKSLIPAKILAG----WVVRCSA-----
-TAAGRTQCSGLLKRFAFLYRLWAT-----T
DTKPGLVDGIPKCPQLWIRSLCSHAS----VRAQIARLGRMLP-----QGDAVT-VRKA
LKKHRSVVTQSTR-----ANVF--VADEIYAF--AKEYMENREEHLREKVSFH---
-----PTEGTASRDYTIERGGRV-----QELL----
--DNAWAMMRKP-----
-----GYIDYQKYPDKS-----RIDE
F--IVQAAIDADKRLHHYRV-----
-----SVMAAAEPGNKARVLCKFPAVAL-VPGDIIR--RQLWPIVASDP
DLDFD----QDPRSEKFLQMIRRTVDR-----EGTIVSSDLSNATDYI--PHMY
AQALWRGLLESFD---APSWVQEHINRMFSPMEM-TY-----P-----
-----DG--KKVW--SQRGIQMGTPLSFMTLSLLHKFAVQASGH-----
-----G-L--SAHIIRGDD-LLGVFTSP--ADYFRSMERVGFKINREKTIVS-
--RVG-GVFAEQTVRVRYTPRRQP-----
-----KKITLA
DFMGP-----PR-----

-----EDVESVTVL---NDIPLK-GAVHVDTKGSV
L-----RGLGRWYA-QWSEQYPAKCG-----RTAHRVAVM-RVHQALIRDAKRW-
RVPLHTPL-ELGGAGIPDRRGRLGVAD-LS-FDMRRKLGACSH-FDRSYA-----
RAVRRLD--GGSTDPFEEHFEQ-ML-----KKAP-----
-----LGSCVYEYDPYATEYYAACRRKYY-----LTKV
TRPARK-----DLNPPPLRD---W--LKQFDNCPAVQPRWAVSK-----
----HADVTLL-----IHLKMFSGAYVPFTT-----
-----PETSILARRKV--R-----

-----GGAMTSQCETTYGLPISVLSWPVNPVG-----VGAASWIPNPRALTGVERGRAL

-----AYAQGVW-----YALTS
NIEAVHPR-----PFKSLIPAKILAG----WVVRCSA-----
-TAAGRTQCSGLLKRF AFLYLRLWAT-----T
DTKPGLVDGIPKCPQLWIRSLCSHAS----VRAQIARLGRMLP----QGDAVT-VRKA
LKKHRSVVTQSTR-----ANVF---VADEIYAF--AKEYMENREEHLREKV SFH---
-----PTEGTASRDYTIERGGRV-----QELL----
--DNAWAMMRKP-----
-----GYIDYQKY PDKS-----RIDE
F--IVQAAIDADKRLHHYRV-----
-----GVMAAAEPGNKARVLCKFP AVAL-VPGDIIR--RQLWP IVASDP
DLDFD----QDPRSEKFLQMIRRTVDR-----EGTIVSSDLSNATDYI--PHMY
AQALWRGLLESFD---APSWVQE HIDRMFSPMEM-TY-----P-----
-----DG--KKVW---SQRGIQMGTPLSFMTLSLLHKFAVQASGH-----
-----G-L--SAHIIRGDD-LLGVFTSP--ADYFRSMERVGF KINREKTIVS-
--RVG-GVFAEQTVRVRYTPRRQP-----
-----KKITLA
DFMGP-----PR-----

-----EDVESVTVL----NDIPLK-GAVHVDTKGSV
L-----RGLGRWYA-QWSEQYP AKCG-----RTAHRAVM-RVHQALIRDAKW-
RVPLHTPL-ELGGAGIPDRRGRLGVAD-LS-FDMRRKLGYACSH-FDRSYA-----
RAVRRLD--GGSTDPFEEHF EQ-ML-----KKAP-----
-----LGSCVYEYDPYATEYYAACRRKYY-----LTKV
TRPARK-----DLNPPPLRD---W---LKQFDNCPAVQPRWAVSK-----
----HADVTLL-----IHRLKMFSGAYVPFTT-----
-----PETSILARRKV---R-----

-MAVVIEAIHRR-----PFTNLVPAKMFAG----WLIRASR-----
 -GPGGQKYVCQIKSLAFDYRVWSI-----T
 AKKPNYSGDIPRAFRNWLSSLCCCHAE----TRAQFARVARMMLP-----VGNKLV-IDQA
 LRKHRSVVTRRPK-----LDIG---LAHQIYGY---ARDYIHERKHSLSERVSFH---
 -----PTEGTASRDVPASAGGRL-----RDLV---
 --DNAHRVMSQR-----P
 I-----VLSEYL-YSHKK-----YFDD
 A--IVEAAIYTDEKVVSTYNY-----

-----DVLAIEPGFKARVLCKFPATAL-LAGDIIR--RQLWPIFESDP
NMDFD----QDIRSEKFQNVIRSLNR-----DGTIVSSDLSNATDYI--PHEY
AKALWAGILDAFD---FPEWVENYLERMFAPIRM-SY-----P-----
-----DG--VTVT---SCRGIQMGTPLSFLTLSLLHKFCVHKSGH-----
-----E-R--SPYIIRGDD-LLGVFSSP--RQYLTVMEEIGFKINRDKTVIS-
--KDG-GTFAEQTVKVTWKAKERDPL-----
-----ARPTLY
DFIVT-----

-----DQKVVSITNL---DDIPFK-GLIHLDNKGGR
L-----RQVGRWYA-QWSPYYPKRG-----KVAYRAIR-RTIGNVLRIARSL-
RIPITCPM-ELGGCGIPNKRGTMQLDAN--FQHRSRVGYAASHESHNFQ-----
-----LAVRKLDIGNASD-LLEDYRQHVDLPKGKSVYMFDPY-----
-----SSEDFSGYRRKRMLLNNTGNYSAY-----
-----HTKPVPLHR-----WLAVFSRCKENRARWVPSR-----
--RCSGSDLRRL-----IDRIKMLSPIYVPYTT-----
-----PNAQILSIRI-----

>GFLP01304447.1

-----FLALGLSSGLYLTSLD----SKHFSTLAFELDNRKRS-----

-----RTFHDTI-----GDFRP
RKLFAKLITYQM-----QNPEILNEEKLRA-----
-----VMPEVSISTNRSRS-----
-----SKFPVTLGSLP----HPSETV-VEET
TRKYLEVVTG-PKSSIDS--PPLD---LLTSIEEF--GRKYVVDRLHRVKD--KVS---
-----WVPNLNTASLHVTAREGGRM-----TELL-----
--LGADTVLKLL-----
-----LEYTDP-SEPNA-----DLEE
A--IVQTSIEIAKLKALDG-----NF-----
-----EA---EVMAVGEAGYKARILCKYSAHIL-LVGDIIR--RQLWPLVSEEP
WFDDEKDYSFSEKLERAFLDLPFWTGYH-----EPEVVSTDLSATDFI--PFEV
ARALWSGILSGLG--VGFDWIFQHVITWAGPTKV-KVGD-----
-----SVHT---TNRGIPMGTPLSFMTLCLLHHYAVKSTEK-----
-----ETGM-E--GPYVIRGDD-FIGVLPNKEV--YFRKMEELGFVVNHSKTFIS-
--KVP-GIPVRPSLRSDTSNEDTVRRSSACGVFAETVVQFNF-----
-----REHPVT
SGKRK-----PF-----

-----TWQPIRLSDVTVI---PDVPFA-GLVREEKSSIK
RSQL---VRYGEWYRQIFRRPDGTLLDRGRWVKKASRVVFTAIR-RLSKDLIMRAVES-
GVPIHTPL-EIGGAGIPDARGRLTLPRD---FVTRSKITYLL-----
-----NHTGNSRVRSARA-TL-----

-----RSAAAELGG-----KIWSS-----
-----FDKV-----

>NC_033474.1

----MTVTSVPVVAQVRIPTLRLLAAVLQRSP----VLRKGSRESLLRLRSLSLKAE
RLDPIERKF-----

-----REELLRW-----VDAAL
YMSLPMYNRTIS-----GSHQVEQYLLALW----KLLHYVY----
-K-NSIPDVVKTLKDLFFRYLDAGLQS-----DREFRRRRF
TYLG-KRHPLKGFFAIAVKGLTRE-----SARTIGLAGRCIP---S-LPLSKERRLEE
YLNLERLTT-PM-----PTIP---ACRGCGHL---PPEHPDKRLAAGPGCQRLSRDH
IVESVIKFAEVFARPVKHLPPMSYPADNMVYIPGPLHHGETR-----PWL----
--GVTKPAVEAISDAFDSFDEWKEAVNSPKLPYQLGIWSVHPEAIENPPERDPAVGLTRV
ISEGESLKTPKRRANVWYDRKDLRVWYSERHAPSVLHPKDPEAERRDDHR-----ILPP
GYGLRALKRSPIAPILNDV-----VLSGEDQCLVP-----
-----VH---KVLFIPEAGAKVRVATVPSVVS LQTAARQIN--GIVLRALKHC
TF-----HLGLKNSKCFSSKTFAMPD-----GAKFYSDLSAASDRLSQPLSL
S-VLWPFIREAEK--AGLEN-AFEVARLSCGPQRI-ILEDPEVMELLADPKY-----
-----SAHLPFFE---TEAGLLMGTPAWPALSLHLFAALEAGI-----
-----D-I--HRVFIVGDDGGLIATAPEYQKYHIMTNILGLRVNREKTHES-
--PMM-GVVAQKVLKIRPAPCPH-----
-----AGTPRG
KLGIG-----PL-----
-TVWD-----
-----FVRHELNLNPNESPVYWGKVELDQSHCYVP---AALTARSRRRR
DPHAP---WNPPEWSD-VA-----RWKRWRERLRLPKALR-LVYPSLVQSRYR-
GLRMEAPA-ILGGCGPLTEGAVRKGLTPVQ-RAGAYAIAREIMHGTRKDLSDVGPDLTR
TIPRAVI--LAANEFNELPRS-ET----RGTGLPKAAALRSLTSF-----
-----FVRVAWEGGFRDANSYLGRSLGHR---VVRHPGRLLLEAKGRIF
LRSRGK-----RAVPKPLTV-----VQAFKRLPKPVDVSPGV LKGLNRLIHD
LDLGGRQKRRLDYVLSAGRNDDAEEGTSSSSVQ-----V
PEAMSSLLNRLRHGWG-----KVWGGIQSLGS---GGKGNLTSTH
HEA-----
>GBTA01004534.1

-----LVPEIKEFTL-----

-----KIVSKMY-----QDWKS
KTEKFNEQCEET-----GRKKINLDLSVPE----W-----
-----SPNLHASACIE-----
--NPIKNGGQQAAYMH-----KLFQVAHTG-----VGEARN-ARAF
AMRIRNVLQQEMI-----WLKSMGDW-----

-----QVGNELPSTTDRKEGGVE-----
--EIPKDDFRFN-----
-----TSMPEG-IPD-----YVSG
S--YLPVG-----
-----RASVVYEKGGMKARVVTVHHAAFV-ALSCALN--RTLINLLKCYT
PS-----KASLLGRDIELKTLKRD-----GLEGFSADLSTATDLI--PFEV
ARAVVDGICEALN---WGPHIKEIMYYATGPFRI-EYDRAPRSKDS----I-----
-----VP--DYII---SRRGILMGLGTSPILSILNSFCAEYNCS-----
-----KEG-V--RNYRICGDD-LAALWDKNDIGNYIGNLESVGLKLNDKKTVTS-
--DQG-LVFTEELYSIKSRTHINR-----

-----KGERSPDQKDL YLSRFKRAIVSVIVDAKHNSTTKFT
NPE-----ERLPRYLT-IG-DCSHAIDE--FCPLWQKSRAQKALL-LMNKKCIREIFRA-
GLDPFLPK-YLGGAGVKTGVNEKR-----HPLMKHLSALIQLPYEDIQLVCR-----
RLQRPWL---SSSVNKETSAIFS-KTLNLFELPRAEH-----
-----GSAIPYKEAVRKFTGAVLARY-----
-----SLLNFGSCQ-----TNKVCTLKFIRKQVNKAR-----
KTLFSKAPRRYH-----YMSLLKFSNALEKLEP-----

>GEWH01003494.1

-----TLNSSVAEKSETGK-----RPRSRMEKLLRRCTERDMTTRQL
LETGLRL-----

-----DSLQLG-----QDLVQ
AFSTLLDVERGN-----KYFERRWFIRLAQ-----WLLRAIDSFTDK
VKAEGISSAMTWIKDALTHLRACAL-----
DAGKVTDYPFPSKAPRVERRR-----MLVQASRCSRSLR-----AATGDK-LEQA
AQSHKEILTS-PF---K--TPEP---ELLSFREY---VRAKTRASTVQVGE-----
-----VSASASLDTKGSGGGS-----EEVR-----
--GLVDSLRSRPFTYPDLLELDRTFPHEMLGGHVLLNRPKVQERFRIGQVGHSSSGTLLG
VPGDQVLFLY-----DQSCEISLDEWE-----FLRE
QLFAIVACVLDVDLDSRPAC-----
-----RQC VIAERG WKTRVTPISAASN-YLLTVAN--KGLLSIIGEHR
EL-----DNSQRGRPCDNLDSFGKR-----YALTFSADLKSATDYM--PHDL
CLVAAEELTRG-----WPAYLRRLVLRVSGSYDL-RC-----L-----
-----DG--GTVE---TRRGILMGVPTSPILSLYSGWLHYRSSH-----
-----D--GWFAVCGDD-YLGCHTYASYHKYLRARERTGAKGSPGKDILS-
--TESTGIFAEELVT-----

-----VGRCRVL----RTVPIR-PLTGIPK----

-----GDTPSW-A-MG-EAISVLLS---EAGFSEERSGRLVS-RRFGSEMDKLRRA-
GIDPKGPR-WALAGGFPGKPG-PRVLETAR-RVLSQGEDTIL---GWLT-----
ALGIAWS--DAPSRDLIQGMVR-DK-----
-----LGGL-----

-----KGQTWRAE-----

>GDUK01007882.1

-----KVEKWLPPFSVFGGSSMASRSR---PLHVGRIRDLRVEMQSQHFPGAS
MTAD-----

-----QIAQQFW-----NVVRS
HLLAAFPPLVSGLPVAKRS-----GRNARFGLVLLQY----HLLRRLH----
-L-GGVASLVRWLKALAFSCRDSVV-----
----SDKKLPREFSRYFHGVLNPSAL---SLTTLAKLGRALPECDAKMGGGSV-IRKA
LRQHRADMTSARVKQFGFNPMDDL---DLAVAKET---ASFLLRGRLGSPRM-----
-----HDLSSASADVSSREGGLA-----AGLRPSWA
RMEEIGPVPWALGHALAEWQAALLRGARNTAWATAEEERGIVLSNGSFLPGPLVLLA-
S
P-----DGLWVT-RRSEYIRSVLSFHEG
LGPVRVTGRFGTTVQVDHLTVQFVHVAGEPDLKRFFASCARD----LP-----
-----KA---KVALVRERGWKVRVVTSLTDQV-IAGHALR--DLIWPVLQDDP
CF-----DLTGDPRIEGMPVDQ-----GEVVL SADLSRATDVV--SIGL
ALAVFDQWNRIDP---RFDRQWHDLARATVSGYRI-QY-----P-----
-----DG--RYVR---SSAGWLMGHPSTWALLCLVHKAIAEMAGF-----
-----KKYRIRGDD-LLAVGSRESVEEYFTLISKY-FEVNVKKSFSV-
--PNG-GTFAEDTFVVRNGQLRLIA-----

-----HAAPPR-GWLAGEKASLA
AAYTQ-F-GRLPRKWR-----MAYRHSL-SSGKHQIAEFRKH-
RIPLFLPR-VLGGVGVPHPRGVGGAVA-----GSVLVSKMVTGAG-----
SVSTAWV--NRMAYRQREALYS-EAREVTRMLHQDGSLSSTVLTEE-----
-----VINEITLKSNIAMGCTWP-----EGWGDKT
LLGELR-----DPSYASIGR-----DWALFRAKCAGIRPPPHK-----
--VGTGWSEQRL-----ADAIKSSLGGMYPY-----
-----RPDPPIRLS-----

>GEUE01057748.1

---QRMKQPAKPGKFYLPNLVESEVPKVTLDLILGGKGGAVIRDLPTLSRVRPARPMPS
AKGKSAEDN CNVLGNDNNGSKRAKTRRRDES WERSD-----

1
2
3 -----KDANVFW-----GAWL
4 LLLDISSIYERGLDLKNPS-----ERRSAREALSVRG----WIRKSVH-----
5 -F-NGVDFVQKEFKKLAHYRHRAV-----GD
6 VRDPPAPGLLDRHLSGIPRSVLKGPIEGRLRACAFIGRLGRCGP----PPSKKV-EERA
7 LKEHRRVLCRRFK-----TPKV---QLASIRKS---TKAWCKDLNEVGPSVVAFS---
8 -----SAASVEFSRLKGGQT-----AELF----
9 --DHAKGAIQMY-----QTLLYGEP
10 I-----EGSNSTGLRDED-----IRIT
11 VQVALDSAIEKAKSRSPLRV-----
12 -----KATVVPETGGKARVVTAGPADMV-LAGNALR--RAVWPIVMGDD
13 SIDTEEAVAQEEVSAGDKELLARLLGG-----GCSWYSGDLTAASDWL--PRDV
14 CLTIWDAIAEAAG--FAKGSPEYLLGRKLLGNVVI-EYA-----
15 -----DG--TEVT---TGAGALMGFPLTWLVLCAYNRALVSMVSK-----
16 -----KA-I--QRTVFRGDD-MVSQMTPEEGARYEELVRLTGGQPNTAKSFRS-
17 --VTG-WVFTEATYQLTTWHTTLTRADLAE-----
18 -----RGVTLE
19 HGLGR-----
20 -----
21 -----AGTLRAITVRRV----PDCSLR-HLLRQRVASQP
22 YLA----IAGPAVSGAVAPLLASRLAD---VRAVGNRLVRRWMR--VNARMVESIQRH-
23 NLGLFVPR-ELGGFGLPHPKGWARGVTSA--RWAHKRLAVFFTATRGAAVKIGC-----
24 LGYDPWVVDPNTGRLWGEVSKR-LASREKSCLVLLPPERVAEPTADR-----TPESW
25 EQYVPNYGGYYLAVVPGAGRGVAQPVTEAVAQEMALESARW---
26 RGLGLDPVVDKPKQES
27 RKSGTSDGCKSKAKKPSRSALTIISK-----NHRAMRKT LGKSRFPEKF-----
28 LPKRALRPEKLQ---MALMRHKMYVQHGS LFGG-----
29 ---PSPLHIVFNSEKGRGPPRRP-----GGWKYYHDPKQ--GRSKTSSDPV
30 PRSPAPPLVGGVPLDPAWENSWPSRKSGNTS-----
31 >GEUE01035234.1
32 -----
33 -----
34 -----
35 -----
36 -----
37 -----
38 -----
39 -----
40 -----
41 -----
42 -----
43 -----
44 -----
45 -----
46 -----
47 -----
48 -----
49 -----
50 -----
51 -----KATVVPETGGKARVVTAGPADMV-LAGNALR--RAVWPIVMGDD
52 SIDTEEAVAQEEVSAGDKELLARLLGG-----GCSWYSGDLTAASDWL--PRDV
53 CLTIWDAIAEAAG--FAKGSPEYLLGRKLLGNVVI-EYA-----
54 -----DG--TEVT---TGAGALMGFPLTWLVLCAYNRALVSMVSK-----
55 -----
56 -----
57 -----
58 -----
59 -----
60 -----

-----KA-I--QRTVFRGDD-MVSQMTPEEGARYEELVRLTGGQPNTAKSFRS-
--VTG-WVFTEATYQLTTWHTTLRADLAE-----
-----RGVTLE
HGLGR-----

-----AGTLRAITVRRV---PDCSLR-HLLRQRVASQP
YLA----IAGPAVSGAVAPLLASRLAD---VRAVGNRLVRRWMR--VNARMVESIQRH-
NLGLFVPR-ELGGFGLPHPKGWARGVTSA--RWAHKRLAVFFTATRGA AVKIGC-----
LGYDPWVVDPNTGRLWGEVSKR-LASREKSCLVLLPPERVAEPTADR-----TPESW
EQYVPNYGGYYLAVVPGAGRGVAQPVTEAVAQEMALESARW---
RGLGLDPVVDKPKQEK
RKSGTSDGCKSKAKKPSRSALTIISK-----NHRAMRKT LGKSRFPEKF-----
LPKRALRPEKLQ---MALMRHKMYVQHGS LFGG-----
---PSPLHIVFNSEKGRGPPRRP-----GGWKYYHDPKQ---GRSKTSSDPV
PRSPAPPL-----
>GBYB01012090.1
ASARLGRSGKWKL PNLPEPVATAVTPPAAAARLLSNLRKDAFPRGVTGNYLSRLEAAGI
P
ACSVSRHPRGSEPHPLRKYTPGRGDPAC SPLTSLPHTPLTSPSRPGRARGGPAGERPRK
RGHSRDPPGGCPLGCGAAADC-----

-----RDLRRIW-----DSFRL
CILSLIPRLQRSGLDLSNRR-----DYRMAVSILKVRS----WLSINWR-----
-S-SGRNETIRQIKAYAEFCRARAL-----DAEQS
RRGAPRGFPVRAFEHGLGSLLRGASSGRT--ALAQLARLGRSMP----PADQAI-LTAA
LEQHRRVLTTADP---G--VPLT---LLDGLFW---ASSWSGSRPASLRNSVVAS---
-----FSSASEGMSRAKGGQR-----AELE---
--QLARAEIDAI-----RADIAGHPDTPDGFDFVDEGD
Y-----AGLYFV-EWDLQ-----TVTD
S--AMARVREEAADTRYAIPV-----
-----RATALPELG NKARVVTAPPAHWG-IIGDAMR--KVLWPLLETDP
RIDLSGRSLDGAASQFHDQVVKSLRGAA-----GQWMYSADLT AATDLM--PEN
ILALWHGVLHGLG--IPEESFFARAGDKILGCVDV-SYDPLAA-----
-----PG--EKPIVVRS MRGCMMLNLSWFLNLNLAIVDIACLG GIAALTD PQD
GGLDEAEVRRIVG-L--APAIVRGDD-LAAAL TERQATAYEELIAATGGEANRAKSYRS-
--HSA-FVLA EKSFLVDREVRPLSGGQKRGFISFPAGLA-----
-----VDAPML
ARDPR-----ALEALGFGGPLTG-

-----KDLGGATWTDAAVEVV SCLTAL----QDIPVR-HLIPAPSK---
-----EGLPVYVS-LPAAAADVLYAAEEEGSPLFPGMCRAVL-SVNSEYVQRYRDY-
RIPLLLPR-EVGGAGFPHPGGFSKALASGG-RGHWLRATLRVT TYGVEQRARRH-----
LDEDVWR--VDGSNSDRRAAAR-LIRAREAAAMNSGQ-----
-----LGNMVPVPLEDEV TREVAFESLWR-----DLFCPADRKER
CTRRDR-----GVRLATISK-----RLSKADMAARRLAFNDRF-----
--LVRSLVRAEY-----QADVAMARGAESVYVPLSERDPRRLMPLSDDHPSLVGLHRRR

YRAGRASSHGRGNGCGGVSQVGGVP-----DNWESRLDDDFVYGGLDDNEAGIS
GHCLPKRSTLERAVRSRGQGA VAVRTLRS GMVLADSIHYTESRSKR RGKPPTS-----
>KF298275.1

-
GVMIRSRRKR NHQRWRLPSTL VLAPEGRITPRAHSQPNEAGGAQGAAQSNVAQVTAER
A
ANKPTQANSQRTRHARAPKPDPE-----

-----QMGQATW-----DAFRL
AIWALIPSLQKKGLALKGPR-----NLATVRGIQSVHK-----WISVSMA-----
-R-SGPEWTAKQVKEYSNYCRARSLR-----DARTP
CP---RGFPVKAFEVGLIKAAISDPNGGRP-ALAQLARLGRAMP----VATDRV-GRQA
LEKHRETLSRATD-----VPPA---LIARFGAW---AEQWARQQADKTGL--TKE---
-----GSRLFQAVETKFSNSAAEAFSAASGGQQ-----AELR----
--ENYQRTMVDV-----QAEYLLATEGFVWDPEV--
-----PTLGAE-DSDLN-----NITL
C--AREEAE---DY YLRRH-----RWRGVD-----GM-----
-----PV---RATIVREQGMKARVVTACPAWAV-VCGDACR--KTLWPLLAADP
RVDLSGARPV AEDLDRFNDQIALSLV GSI-----TPEFFSADLTAATDLM--PFEV
SNALWAGLCRGLD--LLPTSQISRIGHALLGPVEV-SYPD LAA-----
-----KG--EPIPRILSKQGCMMGLPLSWTILNLNLAVADFAVV-----
----APSVVRVG-V--APVIARGDD-LVAA YLPEEADRYTQLLRESGGEVNV LKSFRS-
--SDS-FVLAERTFKVTLRDAVPLPVHRRGFELRARN-----
-----THAPLL
TQVPQ-----PSGTSRSDPNVLGG
GPA-----
-----AESILGKTRETRGKVVSIMKF----DDCPLR-TLVGKGPGYAA
G-----AVIPAYIS-VPSAASASLAE--FEGKRFYPALAQGLL-SVHKSLVGEFRRS-
AIPVFYPR-ELGGGGFPHPRGFGHAVCSAG-LLGRKRAAFAMTCFTHKARKKKHG-----
LGDDPWA--SRHRLAELERARQ-QL-SAE EARAQTRAASEVPGAPRV-----IP
DDWEPKVWDQIKDLEPPRGARIPVPMEDAVVQRAAHGETWA-----SMILQEG
VGQRDR-----YPSLGTIK-----RLQAIKVT AERKFSNRF-----
LPKRGKLSRERL-----LERVAMARGGETRYVP-----
---YVDFGGGGMV LVCRGFDTYAAARYEPESQPERYPERRQKLRP----GTEGEEMVEV
PVAVRTTRAGLGLADVA FVRHGRVKRVKRR LPSQFPR-----
>KP642119.1

-MVLSGTCAQPNPSGVASAVKAVASPSVGASA-----KLRKGPVVGTD RIPKSSLSSPRG
HVGRKPRETVKPKPDPLQ-----

-----GQAQRAW-----DVWRL
ALFALMPSLQKKGLDVRKAKPAKKPVTA--QHATQRAIQVLYK-----W LASSFV-----
-R-SGPEWTCKQIKEYANYCRSRSL-----GDE
RTDAPRGFPVKGF EAGFIKACIMDPNAGGT-ALAQLARVGRAMP----MGTTTRV-AVAA
LRKHRDTLTSTTV-----VPEE---TQVNLRLW---AERWTRDRLAAGARAREGT---

-----ATSFRAASASVSAADGGQL-----AELR----
--QLPRFVDHRQ-----MLRDLLLEEQGFDLDSDDIF
L-----SATCDE-GDDLQ-----LLAD
C--AIETARDRCMEGSPIM-----
-----VATTVLELGMKARVVTKPPAWAV-VAGDACR--QSVWPLLESDR
RIDLSGARPNTESLDRFHDNLAHSLVGAA-----TPQFFSADLTAATDDA--PLRL
-----LGRYLLGPVTV-GYPTLAAIEEHDRLYQ-----
-----AGEPTEIV---SVRGCMMGLPLSWTLLNLYNLAVADMACS-----
----PVGVVQVG-V--APAIARGDD-LVAAIPPGEADRYEELIALTGGEANRLKSFRS-
--ATA-FVLAERTFRVETQEIPDRVRGLPRGWRWRRVSRFGDP-----
-----ALPPLL
TREHQ-----PG-----

-----LRGQPSRNLEGAREVVGVRMS---VDLPLR-ALIAGSASFAG
G-----AAVPTYIS-LPPAATSCLE--FEGTRFYPAVARGLL-SVHRDLVSEMRRS-
AVPLFYPR-ELGGGGFPHPGGFGAGLASAG-LLGLQRAGLALTSHGWAARKKAA-----
LLEDPWV--PRRQNAQLNRARE-QLLASERRAWAQSVLKRGETVASL---EAAGLAERAK
ELKLALATPVPEYERVPRGARIAVPLEDETIRQAARAGLWE-----DAFLSPTVD
DGGRSKTK-----YPSLGDIA-----RLKRIREVTSKCKFDPRF-----
VPNKDRVQRDSF-----FEKLRLRGSETVLVP-----VDSRPVRVTVD
PARHRADPELDNDDCGY-----PGRRLRLALD---REEGGDPELV
PDPVPLLQRGAGATLGDLVVFRPAKSRHGPKSRAGTGARREKRSRTPQHP-----
>KP642120.1
-----MLA
TGVMAKPQRPKVPNRWFLPKPSVGASAKLRK-----GPVVGTDRIKSSLSSPRGHVGR
KPRKTVKPKPDPLQ-----

-----GQAQRAW-----DVWRL
ALFALMPSLQKKGLDARKAKLAKKPVTA--QRATQRAIQVLYK-----WCLASSFV-----
-R-SGPEWTCKQIKEYANYCRSRL-----GDE
RTDAPRGFPVKGFEGFIKACIMDPNAGGT-ALAQLARVGRAMP-----MGTTTRV-AVAA
LRKHRDTLTSTTV-----VPEG---TQVNLRLW---AERWTRDRLAAGARAREGT---
-----TTSFRAASASVSAADGGQL-----AELR----
--QLPRFVDHRQ-----MLRDLLLEEQGFDLGDDI-F
L-----SATCDE-GDDLQ-----LLAD
C--AIETARDRCMEGSPIM-----
-----VATTVLELGMKARVVTKPPAWAV-VAGDACR--QSVWPLLESDR
RIDLSGARPNTESLDRFHDNLAHSLVGAA-----TPQFFSADLTAATDLM--PFSV
SWSLWDGLCDGLG--AAADAPLRLGLGRYLLGPVTV-GYPTLAAMEEHDRLYQ-----
-----AGEPTEIV---SVRGCMMGLPLSWTLLNLYNLAVADMACS-----
----PLGVVQVG-V--APAIARGDD-LVAAIPPGEADRYEELIALTGGEANRLKSFRS-
--ATA-FVLAERTFRVETQEIPDRVRGLPRGWRWRRVSRFGDP-----
-----ALPPLL
TREHQ-----PG-----

-----LKGQPSRNLEGAREVVGIRMS---VDLPLR-ALIAGSASFAG

G-----AAVPTYIS-LPPAATSClse--FEGTRFYPAVARGLL-SVHRNLVSEMRRS-
AVPLFYPR-ELGGGGFPHPGGFGAGLASAG-LLGLQRAGLALTSHGWAARKKAA-----
LLEDPWM--PRRQNAQLNRARE-QLLASERRAWAQSVLKRGETVASL---EAAGLAERAK
ELKLALATPVPEYERVPRGARVAVPLEDETIRQAARAGLWE-----DAFLSPAVD
DGGRSKTK-----YPSLGDiar-----RLKRIREVTSKCKFDPRF-----
VPNKDRAQRDSF-----FEKLRMLRGSETVLVP-----VDSRPVRVTVWD
PARHRADPELDNDDCGY-----PGRRLLRALD---REEGGDPELV
PDPVPLLQRGAGATLGDLVVFRPAKSRHGPKSRAGTGARREKRSR RTPQHP-----
>KF298284.1

-----G
VMAKPQRLFRSQQRRWYLPAVLSLEERSGVTM----PLGTFPVKTGKVAKDGPCHPVPQ
GSAKRANRHPNSPPTRRPKPERRARTQRTPTQPGVELQ-----

-----RQAQHIW-----DTFRL
TIWVLLPSAQKKGIAVS-----AKATIRSIKGVYD----WIRNSMI-----
-R-SGVEWTSKAIKEYANYCRARAL-----GDD
RVGRPRGFPVRTFERGYIAAAVKCPGQGRL-HLAQIARIGRAMP----KGTRNV-GISS
LRKHRGVLSQ-PM---E--TDKD---LLENLRSW---ATQWAQERVSErgTVVNPS---
-----VSFSRSASATVSAMSGQL-----AELS-----
--QLESVKAHHR-----LIRELLLEEQGFENPAGTSDDSIG
L-----LTVADE-VADLQ-----ILAD
C--AISDALDAAREQIIPM-----
-----TATVISELGMKARVVTKPPAWAV-VAGNACR--KTVWPLLEADR
RIDLSGLRPTAEVLDEFHDNLAHSLVGAR-----TTQFYSADLTAATDLM--PFDV
SRALWNGLCDGLG--ASPSAPLRSLGSYLLGPVRV-SYPDLAALKERDRLFK-----
-----AGEETEIL---SSRGCMMLPVSWTVLNLNLAMADMACT-----
---PAGVPTLLG-V--APAVARGDD-LVAAIPEFAADRYENLIAATGGEANRLKSFRS-
--TFA-FVLAERTFEVGVKSVKDLTALTRRGFKVRQVSRAP-----
-----RTAPLL
CRFDE-----

-----LGPQQIPAFIQEMDRPVEVVALRMS---CDLPIR-SLLGGMAGFAG
G-----EAVPNYVA-IPSAACLAEL--FEGTRLYGAIHGGL-SVHRDLVKEMRRS-
AIPLFYPR-VLGGGGFPHPRGFAAAVASAG-PLGLQRAGLYLTASGYKARRKVG-----
IGHDPWV--PLADSRMDVARD-RLVATDRRLWAKSVLDRLPDLLDL---EDPYLDKRLS
ERAEALSVPCVGFETVPRGARVAVPLDDAIVISEAATIAAWE-----DAVLGPC
PSAERR-----YPTLGGIAK-----RLKAIRQKTVEARFPMKF-----
VPNMKRMGREGF-----FNRVKLLLAsetVLVP-----YSNRNTARVTVFDPERHC
TVQEPEDegrVEGDLGL-----PIRRRELRErh---GGGVRVLRDG
ATLGDVAVIRHVPQKRRRHRRTPGPSNRNVL-----
>KF298276.1

-----G
VMAKPQRLFRSQQRRWYLPAVLSLEGRSGVTM----PLGTLPAGRVAKVGPCHPD PQ
GSAKRANQHNPSPSTRPNPERRARAQRTPTQPGVELQ-----

-----RQAQHIW-----DTFRL

1
2
3 TIWVLLPSAQKKGIAS-----AKATIRSIKGVYE----WIRNSMI-----
4 -R-SGVEWTSKAIKEYANYCRARAL-----GDD
5 RVGRPRGFPTGLFERGYIAAALKCPGQGRL-HLAQLARVGRAMP----KGTRNV-GISS
6 LRKHRGVLSQ-PM---E--TDKD---LLENLRSW---ATQWARERVSERGTVVNPS---
7 -----VSFSRSASATVSAMSGGQL-----AELS-----
8 --QLESVKAHHR-----LIRELLLEEQGFENPAGTSDDSIG
9 L-----LTVADE-VADLQ-----ILAD
10 C--AISDALDAAREQIIPM-----
11 -----TATVISELGMKARVVTKPPAWAV-VAGNACR--KTVWPLLEADR
12 RIDLSGLRPTAEVLDEFHDNLAHSLVGAR-----TTQFYADLTAAATDLM--PFDV
13 SRALWNGLCDGLG--ASPSAPLRSLGSYLLGPVRV-SYPDLAALKERDRLFK-----
14 -----AGEETEIL---SSRGCMMLPVSWTVLNLNLYLAMADMACT-----
15 ---PMGVPTLLG-V--APAVARGDD-LVAAIPEFAADRYENLIAATGGEANRLKSFRS-
16 --TFA-FVLAERTFEVGVKSVKDLTALTRRGFKVRQVSRAP-----
17 -----RTAPLL
18 CRFDE-----
19 -----
20 -----LGPQQIPAFIQEMDRPVEVVALRMS---CDLPIR-SLLGGMAGFAG
21 G-----EAVPNYVA-IPSAACLAEE--FEGTRLYGAICHGLL-SVHRDLVKEMRRS-
22 AIPLFYPR-VLGGGGFPHPRGFAAAVASAG-PLGLQRAGLYLTASGYKARRKVG-----
23 IGHDPWV--PLADSRMDVARD-RLVATDRRLWAKSVLDRLPDLLDL---EDPYLDKRLS
24 ERAEALSVPCVGFETVPRGARVAVPLDDAIVISEAATIAAWE-----DAVLGPC
25 PSAERR-----YPTLGGIAE-----RLKAIRQKTVEARFPMKF-----
26 VPMKRMGREGF-----FNRVKLLLASETVLVP-----YSNRNTARVTVFDPERHC
27 TVQEPEDRVEGDLGL-----PIRRRELRRH---GGGVRVLRDG
28 ATLGDVAVIRHVPQKRRRHRRTPGPSNRNVL-----
29 >MF176344.1
30 -----MSL
31 PERVGGTIPRTPAGSREPASAQVGRPQSPKGL---ARKTNHTQIPPTIVDPKSSPRRA
32 RPPKKACAPGKATTKPPASAQQ-----
33 -----
34 -----KQAQHIW-----DAFRL
35 SLWVLLPRLQKHGLSTK-----AVATIEAIKALYS----WIGNSVA-----
36 -R-SGVEFTAKQVKEYANYCRARAL-----KAQ
37 QCEPPRGFPPIRAFERGYIRAMLACPGQGRL-HLAQLARVGRAMP----IGTAKV-AIAS
38 LRKHREVLVSQ-PM---D--TEPA---LVEKLRAW---AEIWTRERLGAGGRTLTPN---
39 -----AAHFSRSASATVSALKGGQL-----TELK-----
40 --QLPAVAEHYA-----LIQELLGDQLDPVTGEPFVWDPYTFDADGGAI-G
41 L-----LTSASE-ADDLQ-----ILAD
42 C--ALRTATEHAADNTPIPM-----
43 -----TATVISELGMKARVVTKPPAWAV-VAGDACR--KTVWPLLEGDR
44 RIDLSGVRPTAEVLDTFHDNLAHSLVGAR-----TTQFYADLTAAATDLM--PFDV
45 SRAMWDGLCDGLG--ATATAPLRTLGRYLLGPVQV-SYPDLRALPASSKLYV-----
46 -----AG--ERVE-CLSKRGCMMLPVSWTVLNLNLYLAMADLACT-----
47 ---PEGSPVLVN-V--APAIARGDD-LVAAIPAGEATRYEDLIAATGGEANRLKSFRS-
48 --ADA-FVLAERTFEVGVLRPNV-ELRQRGYVVR---R--TY-----
49
50
51
52
53
54
55
56
57
58
59
60

-----RTAPLL
AQFDS-----AE-----

-----LHGGTGDARRLGGAPEVVAIRMA----CDVPIR-SLLGGGPRTAG
A-----NPVPDYVS-IPPAACLAEE--FEGTRL YRSVAEGLM-SVHRGLVADLRRS-
AVPLFYPR-ELGGAGFPHPKGFAAAVASAG-ELGLKRAGLALT TAGHKAQQKVG-----
LLSCPWS--PKLTDRRAKEARR-RLIASDQRAWANSVLARHPDLPRG---GSAWTDRAAG
ELRKALSIPARGFETVPRGARIAVALEDAVIREVAAAAAWEDAFLGPGARGWSGVPPPM
Q
EPSRKRSARRKEESSGMYP SLGEVAR-----RLKTIRSQTQAARFNRTF-----
LPKRAVQERARF-----FDRLKALRELETVLVP-----AQSAGARLVVFDPVTHR
VSESAGEIDGGEGRMGF-----PLWRHTLQERK---GKRGAQQGQEA
AVPVPVATTRSGLSLGDVMAVRRL-----PRRPRGVRRLPNPGGAHSRNV
>MF176385.1
-----MSL
PERVGGTIPPTPAGSHEPASAQVGRPQSPKGL---ARKTNHTQIPPTIVDPKSSPRRV
HPPKKVCAPGKATTKPPASAPQ-----

-----KQAQHIW-----DAFRL
SLWVLLPRLQKHGLSTK-----AVTTIEAIKALYS----WIGNSVA----
-R-SGVEFTAKQVKEYANYCRARAL-----KAQ
QCEPPRGFPPIRAFERGYIHAMLACPGQGRL-HLAQLARVGRAMP----IGTTKV-AIAS
LRKHREVL SQ-PM---D--TEPA---LVEKLRAW---AEIWTRERLGAGGRTL TNP---
-----AAHFSRSASATVSALKGGQL-----TEL R-----
--QLPAVAEHYA-----LIQELLGDQLDPVTGEPFVWD PYTFDADGGAI-G
L-----LTSASE-ADDLQ-----ILAD
C--ALRTATEHAADDTPIPM-----
-----TATVISELGMKARVVT KPPAWAV-VAGDACR--KTVWPLLEGDR
RIDLSGVRPTAEVLDTFHDNLAHSLVGAR-----TTQFYSADLT AATDLM--PFDV
SRAIWDGLCDGLG--ATATAPLRTLGRYLLGPVQV-SYPDLRALPASSKLYA-----
-----AG--ERVE-CLSKRGCMMLPVSWTVLNLYNLAMADLACT-----
---PEGSPVLVN-V--APAIARGDD-LVAAIPAGEATRYEDLIAATGGEANRLKSFRS-
--ADA-FVLAERTFEVGVLRRPNV-ELKQRGYVVR---R--TY-----
-----RTAPLL
AQFDS-----AE-----

-----LHGGTGDARRLGGAPEVVAIRMA----CDVPIR-SLLGGGPRTAG
A-----NPVPDYVS-IPPAACLAEE--FEGTRL YRSVAEGLL-SVHRGLVADLRRS-
AVPLFYPR-ELGGAGFPHPKGFAAAVASAG-ELGLKRAGLALT TAGHKAQQKVG-----
LLSCPWS--PKLTDRRAKEARR-RLIASDQRAWANSVLARHPDLPRG---GSAWTDRAAG
ELRKALSIPARGFETVPRGARVAVALEDAVIREVAAAAAWEDAFLGPGARGWSGVPPP
MQ
EPSRKRSARRKEESSGMYP SLGEVAR-----RLKTIRSQTQAARFNRTF-----
LPKRAVQERARF-----FDRLKALRELETVLVP-----AQSAGARLVVFDPVTHR
VSESAGEIDGGEGRMGF-----PLWRHTLQERK---GKRGAQQGQEA
AVPVPVATTRSGLSLGDVMTVRRL-----PRRPRGVRRLPNPGGAHSRN--

>MF176278.1

-----MSL
PERVGRTPPRTTPAGSHEPASAQVGRPQPSPKGL---ARKTNHTQIPPTIVDPKSSPRRA
RPPKKVCAPGKATTKPPASAQQ-----

-----KQAQHIW-----DAFRL
SLWVLLPRLQKHGLSTK-----AVATIEAIKALYS----WIGNSVA-----
-R-SGVEFTAKQVKEYANYCRARAL-----KAQ
QCEPPRGFPPIRAFERGYIHAMLACPGQGRL-HLAQLARVGRAMP-----IGTTKV-AIAS
LRKHREVLQS-PM---D--TEPA---LVEKLRAW---AEIWTRERLGAGGRTLTPN---
-----AAHFSRSASATVSALKGGQL-----TELR----
--QLPAVAEHYA-----LIQELLGDQLDPVTGEPFVWDPYTFDADGGAI-G
L-----LTSASE-ADDLQ-----ILAD
C--ALRTATEHAADNTPIPM-----
-----TATVISELGMKARVVTKPPAWAV-VAGDACR--KTVWPLLEGDR
RIDLSGVRPTAEVLDTFHDNLHSLVGAR-----TTQFYSADLTAATDLM--PFDV
SRAMWDGLCDGLG--ATATAPLRTLGRYLLGPVQV-SYPDLRALRASSKLYV-----
-----AG--ERVE-CLSKRGCMMLPISWTVLNLNLYLAMADLACT-----
---PEGSPVLVN-V--APAIARGDD-LVAAIPAGEATRYEDLIAATGGEANRLKSFRS-
--ADA-FVLAERTFEVGLRRPNV-ELRQRGYVVR---R--TY-----
-----RTAPLL
AQFDS-----AE-----

-----LHGGTGDARRLGGAPEVVAIRMA---CDVPIR-SLLGGGPRTAG
A-----NPVPDYVS-IPPAACLAEE--FEGTRLYSVAEGLM-SVHRGLVADLRRS-
AVPLFYPR-ELGGAGFPHPKGFAAAVASAG-ELGLKRAGLALTTAGHKAQQKVG-----
LLSCPWS--PKLTDRAAKEARR-RLIASDQRAWANSVLARHPDLPRG---GSAWTDRAAG
ELRKALSIPARGFETVPRGARVAVALEDAVIREVAAAAAWEDAFLGPGARGWSGVPPP
MQ
EPSRKRSARRKEESSGMYPSSLGEVAR-----RLKTIRSQTQAARFNRTF-----
LPKRAVQERARF-----FDRLKALRELETVLVP-----AQSAGARLVVFDPVTHR
VSESAGEIDGGEGRMGF-----PLWRHTLQERK---GKRGAGGQEA
AVPVPVATTRSGLSLGDVMTVRRL-----PRRPRGVRRLPNPGGAHSRNV
>MF176306.1

-----MSL
PERVGRTPPRTTPAGSHEPASAQVGRPQPSPKGL---ARKTNHTQIPPTIVDPKSSPRRA
HPPKKVCAPGKATTKPPASAQQ-----

-----KQAQHIW-----DAFRL
SLWVLLPRLQKHGLSTK-----AVATIEAIKALYS----WIGNSVA-----
-R-SGVEFTAKQVKEYANYCRARAL-----KAQ
QCEPPRGFPPIRAFERGYIHAMLACPGQGRL-HLAQLARVGRAMP-----IGTTKV-AIAS
LRKHREVLQS-PM---D--TEPT---LVEKLRAW---AEIWTRERLGAGGRTLTPN---
-----AAHFSRSASATVSALKGGQL-----TELR----
--QLPAVAEHYA-----LIQELLGDQLDPVTGEPFVWDPYTFDADGGAI-G
L-----LTSASE-ADDLQ-----ILAD

C--ALRTATEHAADNTPIPM-----
-----TATVISELGMKARVVTKPPAWAV-VAGDACR--KTVWPLLEGDR
RIDLSGVRPTAEVLDTFHDNLAHSLVGAR-----TTQFYSADLTAATDLM--PFDV
SRAMWDGLCDGLG--ATATAPLRTLGRYLLGPVQV-SYPDLRALPASSKLYV-----
-----AG--ERVE-CLSKRGCMMLPISWTVLNLNLYLAMADLACT-----
----PEGSPVLVN-V--APAIARGDD-LVAAIPAGEATRYEDLIAATGGEANRLKSFRS-
--ADA-FVLAERTFEVGVLRRPNV-ELRQRGYVVR---R--TY-----
-----RTAPLL
AQFDS-----AE-----

-----LHGGTGDARRLGGAPEVVAIRMA---CDVPIR-SLLGGGPRTAG
A-----NPVPDYVS-IPPAACLAEE--FEGTRLYSVAEGLM-SVHRGLVADLRRS-
AVPLFYPR-ELGGAGFPHPKGFAAAVASAG-ELGLKRAGLALTTAGHKAQQKVG-----
LLSCPWS--PKLTDRAAKEARR-RLIASDQRAWANSVLARHPDLPRG---GSAWTDRAAG
ELRKALSIPARGFETVPRGARVAVALEDAVIREVAAAAAWEDAFLGPGARGWSGVPPP
MQ
EPSRKRSARRKEESSGMYPSSLGEVAR-----RLKTIRSQTQAARFNRTF-----
LPKRAVQERARF-----FDRLKALRELETVLVP-----AQSAGARLVVFDPVTHR
VSESAGEIDGGEGRMGF-----PLWRHTLQERK---GKRGAAQQGEA
AVPVPVATTRSGLSLGDVMTVRRL-----PRRPRGVRRLPNPSGAHSRNV
>MF176257.1
-----MSL
PERVGGTIPRTPAGSHEPASAQVGRPQSPKGL---ARKTNHTQIPPTIVDPKSSPRRV
HPPKKVCAPGKATTKPPASARQ-----

-----KQAQHIW-----DAFRL
SLWVLLPRLQKHGLSTK-----AVATIEAIKALYS----WIGNSVA-----
-R-SGVEFTAKQVKEYANYCRARAL-----KAQ
QCEPPRGFPPIRAFERGYIHAMLACPGQGRL-HLAQLARVGRAMP-----IGTTKV-AIAS
LRKHREVLQS-PM---D--TEPA--LVEKLRAW--AEIWTRERLGAGGRTLNP---
-----AAHFSRSASATVSALKGGQL-----TELK---
--QLPAVAEHYA-----LIQELLGDQLDPVTGEPFVWDPYTFDADGGAI-G
L-----LTSASE-ADDLQ-----ILAD
C--ALRTATEHAADNTPIPM-----
-----TATVISELGMKARVVTKPPAWAV-VAGDACR--KTVWPLLEGDR
RIDLSGVRPTAEVLDTFHDNLAHSLVGAR-----TTQFYSADLTAATDLM--PFDV
SRAMWDGLCDGLG--ATATAPLRTLGRYLLGPVQV-SYPDLRALPASSKLYV-----
-----AG--ERVE-CLSKRGCMMLPVSWTVLNLNLYLAMADLACT-----
----PEGSPVLVN-V--APAIARGDD-LVAAIPAGEATRYEDLIAATGGEANRLKSFRS-
--ADA-FVLAERTFEVGVLRRPNV-ELRQRGYVVR---R--TY-----
-----RTAPLL
AQFDS-----AE-----

-----LHGGTGDARRLGGAPEVVAIRMA---CDVPIR-SLLGGGPRTAG
A-----NPVPDYVS-IPPAACLAEE--FEGTRLYSVAEGLM-SVHRGLVADLRRS-
AVPLFYPR-ELGGAGFPHPKGFAAAVASAG-ELGLKRAGLALTTAGHKAQQKVG-----

1
2
3 LLSCPWS--PKLTDRRAKEARR-RLIASDQRAWANSVLARHPDLPRG---GSAWTDRAAG
4 ELRKALSIPARGFETVPRGARVAVALEDAVIREVAAAAAWEDAFLGPGARGWSGVPPP
5 MQ
6 EPSRKRSARRKEESSGMYP SLGEVAR-----RLKTIRSQTQAARFNRTF-----
7 LPKRAVQERARF-----FDRLKALRELETVLVP-----AQSAGARLVVFDPVTHR
8 VSESAGEIDGGEGRMGF-----PLWRHTLQERK---GKRGAQGQEA
9 AVPVPVATTRSGLSLGDVMTVRRL-----PRRPRGVRRLPNPGGAHSRNV
10 >MF176365.1

11
12 -----MSL
13 PERVSGTIPRTPAGSHEPASAQVGRPQSPKGL---ARKTNHTQIPPTIVDPKSSPRRV
14 HPPKKVCAPGKATTKPPASAQQ-----
15 -----

16
17 -----KQAQHIW-----DAFRL
18 SLWVLLPRLQKHGLSTK-----AVATIEAIKALYS----WIGNSVA-----
19 -R-SGVEFTAKQVKEYANYCRARAL-----KAQ
20 QCEPPRGFPPIRAFERGYIHAMLACPGQGRL-HLAQLARVGRAMP-----IGTTKV-AIAS
21 LRKHREVL SQ-PM---D--TEPA---LVEKLRAW--AEIWTRERLGAGGRTL TNP---
22 -----AAHFSRSASATVSALKGGQL-----TEL R-----
23 --QLPAVAEHYA-----LIQELLGDQLDPVTGEPFVWDPTFDADGGAI-G
24 L-----LTSASE-ADDLQ-----ILAD
25 C--ALRTATEHAADNTPIPM-----
26 -----TATVISELGMKARVVT KPPAWAV-VAGDACR--KTVWPLLEGDR
27 RIDLSGVRPTAEVLDTFHDNL AHSLVGAR-----TTQFYSADLTAATDLM--PFDV
28 SRAMWDGLCDGLG--ATATAPLRTLGRYLLGPVQV-SYPDLRALPASSKLYV-----
29 -----AG--ERVE-CLSKRGCMMLPVSWTVLNLN LAMADLACT-----
30 ----PEGSPVLVN-V--APAIARGDD-LVAAIPAGEATRYEDLIAATGGEANRLKSFRS-
31 --ADA-FVLAERTFEVGVLRRPNV-ELRQRGYVVR---R--TY-----
32 -----RTAPLL
33 AQFDS-----AE-----
34 -----

35
36 -----LHGGTGDARRLGGAPEVV AIRMA---CDVPIR-SLLGGGPRTAG
37 A-----NPVPDYVS-IPPAACLAEE--FEGTRL YRSVAEGLM-SVHRGLVADLRRS-
38 AVPLFYPR-ELGGAGFPHPKGFAAAVASAG-ELGLKRAGLALT TAGHKAQQKVG-----
39 LLSCPWS--PKLTDRRAKEARR-RLIASDQRAWANSVLARHPDLPRG---GSAWTDRAAG
40 ELRRALSIPARGFETVPRGARVAVALEDAVIREVAAAAAWEDAFLGPGARGWSGVPPP
41 MQ
42 EPSRKRSARRREESSGMYP SLGEVAR-----RLKTIRSQTQAARFNRTF-----
43 LPKRAVQERARF-----FDRLKALRELETVLVP-----AQSAGARLVVFDPVTHR
44 VSESAGEIDGGEGRMGF-----PLWRHTLQERK---GKRGAQGQEA
45 AVPVPVATTRSGLSLGDVMTVRRL-----PRRPRGVRRLPNPSGAHSRNV
46 >KX883461.1

47
48 -----MSL
49 PEGVSGAIPRTPAGSHEPASAQVGRPQPGPEGL---ARKTNHTQIPPTIVDPKSSPRRA
50 HPPKKARAPGKATTKPPASVKQ-----
51 -----

52
53 -----KQAQHIW-----DAFRL
54
55
56
57
58
59
60

SLWVLLPRLQKHGLSTK-----AVATIEAIKALYS-----WIGNSVA-----
-R-SGVEFTAKQVKEYANYCRARAL-----RAQ
QCEPPRGFPPIRAFERGYILAMLACPGQGRL-HLAQLARVGRAMP-----IGTTKV-AIAS
LRKHREVL SQ-PM---D--TEPA---LVEKLRAW---AEIWTRERLGAGGRTL TNP---
-----AAHFSRSASATVSALKGGQL-----TEL R-----
--QLPAVAEHYA-----LIQELLGDQLDPVTGEPFVWDPYTFDADGGAI-G
L-----LTSASE-ADDLQ-----ILAD
C--ALRTATEHAADNTPIPM-----
-----TATVISELGMKARVVT KPPAWAV-VAGDACR--KTVWPLLEGDR
RIDLSGVRPTAEVLDTFHDNLAHSLVGAR-----STQFYSADLTAATDLM--PFDV
SRAMWNGLCDGLG--ATATAPLRTLGLYLLGPVQV-SYPDLRALPASSKLYV-----
-----AG--ERVE-CLSKRGCMMGLPVSWTVLNLN LAMADLACT-----
---PEGSPVLVN-V--APAIARGDD-LVAAIPAGEATRYEDLIAATGGEANRLKSFRS-
--ADA-FVLAERTFEVGVLRRSNVELKQRGYVTRRTF-----
-----RTAPLL
AQFDS-----

-----AELHGGTGDARRLGDAPEVV AIRMA---CDVPIR-SLLGGGPRTVG
A-----NPVPDYVS-IPPAACLAEE--FEGTRL YRSVAEGLM-SVHRGLVADLRRS-
AIPLFYPR-ELGGAGFPHPKGF AA AVASAG-ELGLKRAGLALT TAGHKAQQKVG-----
LLSCPWS--PKLTD RRAREARR-RLIASDQRAWANSVLARHPDLPRG---GSAWTDRAAG
ELRKALSV PAREFETVPRGARVA VALEDAVIREVAAAAAWEDAFLGP GARGWSGVPPP
TQ
EPSRKRSARQKEESLGMYP SLGEVAR-----RLKTIRSQTQAARFNRTF-----
LPKRAVQERARF-----FDRLKALRELETVLVP-----AQAAGARLVVFDPVTHR
VPESAGEIDGGEGRMGF-----PLWRHTLQERK---GKRGAQQQE A
AVPVPVATTRSGLSLGDVMTVRRLPRRPRGVRRLPNRSGAHSRNV R-----
>KX883537.1
-----MSL
PEGVSGAIPRTPAGSHEPASAQVGRPQPSPKGL---ARKTNHTQIPPTIVDPKSSPRRA
HPPKKARAPGKATTKPPASVQQ-----

-----KQAQHIW-----DAFRL
SLWVLLPRLQKRGLSTK-----AVATIEAIKALYS-----WIGNSVA-----
-R-SGVEFTAKQVKEYANYCRARAL-----RAQ
QCEPPRGFPPIRAFERGYILAMLACPGQGRL-HLAQLARVGRAMP-----IGTTKV-AIAS
LRKHREVL SQ-PM---D--TEPA---LVEKLRAW---AEIWTRERLGAGGRTL TNP---
-----AAHFSRSASATVSALKGGQL-----TEL R-----
--QLPAVAEHYA-----LIQELLGDQLDPVTGEPFVWDPYTFDADGGAI-G
L-----LTSASE-ADDLQ-----ILAD
C--ALRTATEHAADNTPIPM-----
-----TATVISELGMKARVVT KPPAWAV-VAGDACR--KTVWPLLEGDR
RIDLSGVRPTAEVLDAFHDNLAHSLVGAR-----STQFYSADLTAATDLM--PFDV
SRAMWNGLCDGLG--ATATAPLRKLGLYLLGPVQV-SYPDLSALPASSKLYV-----
-----AG--ERVE-CLSERGCMMGLPISWTVLNLN LAMADLACT-----
---PEGSPVLVN-V--APAIARGDD-LVAAIPAE EATRYEDLIAATGGEANRLKSFRS-

--ADA-FVLAERTFEVGVLRPNVELKQRGYVTRRTY-----
-----RTAPLL
AQFDS-----

-----AELHGGTGDARRLGDAPEVVAIRMA---CDVPIR-SLLGGGPRTVG
A-----NPVPDYVS-IPPAACLAEE--FEGTRL YRSVAEGLM-SVHRGLVADLRRS-
AIPLFYPR-ELGGAGFPHPKGFAAAVASAG-ELGLKRAGLALTTAGHKAQQKVG-----
LLSCPWS--PKLTDRRAREARR-RLIASDQRAWANSVLARHPDLPRG---GSAWTDRAAG
ELRKALSVPARGFETVPRGARVAVALEDAVIREVAAAAAWEDAFLGPGARGWSGVPPP
MQ
EPSRKRSARQKEESLGMYP SLGEVAR-----RLKTIRSQTQAARFNRTF-----
LPKRAVQERARF-----FDRLKALRELETVLVP-----AQAAGARLVVFDVPVTHR
VSESAGEVDGGEGRMGF-----PLWRYTLQERK---GKRGAQGQEA
AVPVPVATTRSGLSLGDVMTVRRLPRRPRGVRRLPNRSGAHSRNVR-----
>KX883538.1
-----MSL
PEGVSGAIPRTPAGSHEPASAQVGRPQSPKGL---ARKTNHTQIPPTIVDPKSSPRRA
HPPKKARAPGKATTKPPASVQQ-----

-----KQAQHIW-----DAFRL
SLWVLLPRLQKRGLSTK-----AVATIEAIKALYG----WIGNSVA-----
-C-SGVEFTAKQVKEYANYCRARAL-----RAQ
QCEPPRGFPIRAFERGYIHAMLACPGQGRL-HLAQLARVGRAMP-----IGTTKV-AIAS
LRKHREVL SQ-PM---D--TEPA---LVEKLRAW---AEIWTRERLGAGGRTL TNP---
-----AAHFSRSASATVSALKGGQL-----TEL R---
--QLPAVAEHYA-----LIQELLGDQLDPVTGEPFEWDPYTFDADGGAI-G
L-----LTSASE-ADDLQ-----ILAD
C--ALRTATEHAADNTPIPM-----
-----TATVISELGMKARVVT KPPAWAV-VAGDACR--KTVWPLLEGDR
RIDLSGVRPTAEVLDAFHDNLAHSLVGAR-----STQFYSADLTAATDLM--PFDV
SRAMWNGLCDGLG--ATATAPLRKLGLYLLGPVQV-SYPDL SALPASSKLYV-----
-----AG--ERVE-CLSERGCMMGLPVS WTVLNLYNLAMADLACT-----
---PEGSPVLVN-V--APAIARGDD-LVA AIPAE EATRYEDLIAATGGEANRLKSFRS-
--ADA-FVLAERTFEVGVLRPNVELKQRGYVTRRTY-----
-----RTAPLL
AQFDS-----

-----AELHGGTGDTRRLGDAPEVVAIRMA---CDVPIR-SLLGGGPRTVG
A-----NPVPDYVS-IPPAACLAEE--FEGTRL YRSVAEGLM-SVHRGLVADLRRS-
AIPLFYPR-ELGGAGFPHPKGFAAAVASAG-ELGLKRAGLALTTAGHKAQQKVG-----
LLSCPWS--PKLTDRRAKEARR-RLIASDQRAWANSVLARHPDLPRG---GSAWTNRAAG
ELRKALSVPSRGFETVPRGARVAVALEDAVIREVAAAAAWEDAFLGPGARGWSGVPPP
MQ
EPSRKRSARQKEESLGMYP SLGEVAR-----RLKTIRSQTQAARFNRTF-----
LPKRAVQERARF-----FDRLKALRELETVLVP-----AQAAGARLVVFDVPVTHR
VSESAGEIDGGEGRMGF-----PLWRHTLQERK---GKRGAQGQEA

1
2
3 AVPVPVATTRSGLSLGDVMTVRRLPRRPRGVRRLLPNRSGAHSRNVR-----
4 >KX883539.1
5 -----
6 -----
7 -----
8 -----
9 -----
10 -----
11 -----
12 ---MGREETVKRVKQYAEYCRARSI-----DARQA
13 EAGPPRGFPFHAELEGGSFKALLRCPVQGPA-ALAQVARLGRSMP----LPTHKQ-QGRA
14 LIEHCNELTAGPH--EE--IPVE--LLDAFYW---TRTWGAGKGAALRDEVTVN---
15 -----FSSAAESVSRARGGQR-----EELR---
16 --LLAAPHFEBI-----REALAIDIDFRVEEGD
17 Y-----AGIYSE-EDDLE-----IITQ
18 A--AIERVRADIRAGRPPDV-----
19 -----RAAVVPELGAKSRIVTAGEAHWV-VIGDAIR--KCLWPALEAEP
20 RIDLSGRRDVAGTAAHLHDEVAKALTGAE-----AVEFYSADLTAATDLM--PHDL
21 IQAIWRGLMEGLG--VSDDDLMTAGLALLGPVNL-HHPDLDDLDLDR-----
22 -----PGKYAYGV---THKGCMMSGFPLSWYILNLYNLACADLALS-----AER
23 GRFLAEVAPRLVG-A--APAVVRGDD-LCSAHTPAEADRYERIIRATGGRANLSKSYRS-
24 --RKG-FILAERTFLVTTARTKCSAARSRLGTVVVRPGVH-----
25 -----PRAPLL
26 TRAPD-----AFESGALG-----
27 -----
28 -----YRSIDPRSRRRGAATVVGLTML---SDVPVR-HLMPSPTS---
29 -----EGLPAYVT-LPPAASAVVDE--IKDIKRREAICYGIL-SVNAPLVVRQFNRF-
30 KIPFYPR-ELGGAGFPHPKGFGSAIRSAG-GMGHLRATLAMTTFTRRARQAYQ-----
31 LDRDPWI---CDRQVEFERSR-AQTGLIAEERRAAESGQ-----
32 -----LGLATSVPLEDAVAARAAEGQLWA-----EVMLPQSR
33 ELASARRGRR-----GPRLSSVRK----GLSDVFNALAKVHAKFDPKF-----
34 -LANPRIDKETW-----MDRVAMAKGSSTVYRV-----VRPSGRRVVMHGIRD
35 HREIRAEREMKGGWT-----FGGQRDLRQNTAAGLEGGEIAPV
36 PTQRAGATLWDVLSVSCKRARKRKKGNPRL-----
37 >GENC01006608.1
38 -----
39 ----RVSQRGSDKAWFLPDNVRWTQIVGNLR----PPGESAGTAEMVESLPASQKSKV
40 GRKKEKFPGRATK-----
41 -----
42 -----AAGQRIW-----NAFRL
43 AFWAIAPSAQRKGLNPRSQR-----QARSIQAFRTLEK----WIAISLL-----
44 -K-CGTEWTKAVKDYSEYCRARAL-----RAK
45 QCVRPREFPVTGFEIGLVRGAINDPLAGAE-ALAQVARLGRAMP----QGTDKQ-VIRA
46 LEKHRDTLLGETR-----VDEE---TLDRYHAW---AERWSSDRRRFAPERVAVK---
47 -----FSDSAAESVSRADGGQK-----ADLL---
48 --ENARPHFEAL-----RAALAESQGWDA PD-D
49 Y-----IGLYSE-EDDVE-----TIAQ
50
51
52
53
54
55
56
57
58
59
60

1
2
3 M--ALETVRAAVRSGEPVPA-----
4 -----RASAIPEYGSKVRVVTACPGPFV-VAGDACR--KIVWGLLEADP
5 RVDLSGRRPLGDSISQFQDEVARS LVGAV-----TTEYY SADLT AATDLM--PFQL
6 SAAIWTGLCDGLG--ESQSDLSIGLHLLGPVQV-RYPDLK----EGSVE-----
7 -----EG--FRVPTIISQQGCMMLPLSWTVLNVFNLATADLAIG-----
8 --APEDRVTDLIG-R--APAVARGDD-LAAAF LPEEADRYEALISATGGRVNLSKSFRS-
9 --SIG-FVLAERTFMVKRSVDRYQRRAPRQAVLIQPHRPEGIP-----
10 -----QNAPIL
11 AQDPS-----
12 -----
13 -----APTMSGVGRDVRGHRTESLRML----PDCSLK-GLLGGSAST--
14 -----DGAPRYVT-LPVSAASSLYE--FEGTRFYGALCRGLL-SVHKDLLGTFREF-
15 GVPKFYPR-ELGGGGFPHPRGYLRAMESAG-GIGHLRATLNITRFTESSRRAA-----
16 LHMDPWR--TERDSRTLERARS-QLIAEESRAREAGL-----
17 -----LGSFVSKFQEEAEVTRAAQGS LWE-----TTLLSQ
18 KKRDEK-----FPRLKDVRH-----RMLEITEKTKAIKFNR RF-----
19 -LLRRQLHRDEY-----LQIASHARGAPTRYVRMERAPGLVALRDIVPGQQAPNSFQTG
20 TVPQHGNARGNRGERAEMSLVPTLRSRASLGQVARILPARASRKKRG---ARRGPNRPS
21 DSSRGE-----
22 >GEND01011317.1
23 -----
24 ----DLSRQGS DKA WFLPDNVRWTQIVGNLR----PPGESAGTAEMVESLPASQKSKV
25 GRKKEKFPGRATK-----
26 -----
27 -----AAGQRIW-----NAFRL
28 AFWAIAPSAQRKGLNPRSQR-----QARSIQAFRTLEK----WIAISLL-----
29 -K-CGTEWTCKAVKDYSEYCRARALR-----AKQCV
30 --RP-REFPVTGFEIGLVXGAINDPLAGAE-ALAQLARLGRAMP-----QGTDKQ-VIRA
31 LEKHRDTLLG-ET---R--VDEE---TLDRYHAW---AERWSSDRRRFAPE--RVA---
32 -----V-----KFSDSA AESVSRADGGQK-----ADLL---
33 --ENARPHFEAL-----RAALA-ESQGWDAPD-D
34 Y-----IGLYSE-EDDVE-----TIAQ
35 M--ALETVR---AAVRSGE-----PV-----
36 -----PA---RASAIPEYGSKVRVVTACPGPFV-VAGDACR--KVVWGLLEADP
37 RVDLSGRRPLGDSISQFQDEVARS LVGAV-----TTEYY SADLT AATDLM--PFQL
38 SAAIWTGLCDGLG--ESQSDLSIGLHLLGPVQV-RYPDLK----EGSVE-----
39 -----EG--FRVPTIISQQGCMMLPLSWTVLNVFNLATADLAIG-----
40 --APEDRVTDLIG-R--APAVARGDD-LAAAF LPEEADRYEALISATGGRVNLSKSFRS-
41 --SIG-FVLAERTFMVKRSVDRYQRRAXRQAVLIQPHRPEGIP-----
42 -----QNAPIL
43 AQDPS-----
44 -----
45 -----APTMSGVGRDVRGHRTESLRML----PDCSLK-GLLGGSAST--
46 -----DGAPRYVT-LPVSAASSLYE--FEGTRFYGALCRGLL-SVHKDLLGTFREF-
47 GVPKFYPR-ELGGGGFPHPRGYLRAMESAG-GIGHLRATLNITRFTESSRRAA-----
48 LHMDPWR--TERDSRTLERARS-QLIAEESRAREAG-----
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9 >KX883556.1
10 -----MQEVNISTLTRQGVSNDRP
11 LSATEKVDADCMKILHESSGTD PNTAARGAPH----PSTEVEGEESRQRGRSKAAPAPA
12 REGSGADHR SKNPA-----
13
14 -----ADNVSVWGP GGRALLSDSLLS
15 SVLSLGPSSSTR LITGLRGVTLAIFTAEQDEV LARSLAYRTCR----HLARRAL-----
16 -H-EGVEAVLSVMGRWFSSARE-----SFLNG
17 AKNPSLNPVMRRFSPGTRKDRVT-----CVAQFSYGKRALP----VASDRV-AQES
18 LRVHERRLTT-PL---S--TPDP---VLADFRRF---CRAWAEGHLDPNPLPLGVV---
19 -----GLSQAASYTHSVRKGGHM-----ARCR---
20 --EILEEMNASN-----YCEPPTRP
21 P-----GLLPQE-WWDGW----VKER
22 ---LLTHCRERYSTPGRE-----PP-----
23 -----KG---TAVIVYERGLKARIVTKIETEAC-ILGHQAR--MRLVSAMRKMP
24 EL-----SALSGNHKEFLTAFNGQ-----EGWILSSDLTAASDLL--PLDL
25 VSAGVEGICDSGR---LLPDEELGLRVCTGPFDL-EW-----V-----
26 -----RG--SPRR---SSTGILMGAPPTWCLLSLIHLFWLEQARR-----
27 -----VYGNSGG-Q--LPA RIFGDD-LVA AFDKRQKLAYERSVTVCHGLLSK GKHSFH-
28 --DSH-GVFLERL FVASRHIVEVDHT-----
29 -----NRGPLQ
30 VMGRL-----
31
32 -----TTRHNKIKSISLL----RTLPLR-PLTVLQVSVA A
33 GRRHV-SKGTLP TPLA-VGGVSDGLMQH----GFPPELIRRCQI-ALWPGLPAFYRKL-
34 GIPAF LPT-LLGGGG LILPEGWEAPITRFS-KLVRASTTALVTGNGRLSYV-----
35 PLARL---PPPV LQMALEEAD-QLLPAHPHRVSHVTPKPWLGHIAW-----
36 -----HDMGVW NDFRITAAMVAQEGYLLQ-----MPME
37 CLGRPL-----NITVGRWKD-----QVWKTRLALCKKAWWASV-----
38 ----TRARYTI-----ARAVRKCTE WPHVWLP-----
39 -GSPDPDHPGVY GIPMYIDSKPV-----MEAKAAIALRQ---GRPTSR----
40
41 >KX883562.1
42 -----
43 -----MKILHESSGTD PNTAARGAPHPSTE----VEGEESRQRGRSKAAPAPAREGS
44 GADHR SKNPA-----
45
46 -----ADNVSVWGP GGRALLSDSLLS
47 SVLSLGPSSSTR LITGLRGVTLAIFTAEQDEV LARSLAYRTCR----HLARRAL-----
48 -H-EGVEAVLSVMGRWFSSARE SFL-----NG
49 AKNPSLNPVMRRFSPGTRKDRVT-----CVAQFSYGKRALP----VASDRV-AQES
50
51
52
53
54
55
56
57
58
59
60

1
2
3 LRVHERRLTT-PL---S--TPDP---VLADFRRF---CRAWAEGHLDPNPLPLGVV---
4 -----GLSQAASYTHSVRKGGHM-----ARCR---
5 --EILEEMNASN-----YCEPPTRP
6 P-----GLLPQE-WWDGW-----VKER
7 ---LLTHCRERYSTPGREPP-----
8 -----KG---TAVIVYERGLKARIVTKIETEAC-ILGHQAR--MRLVSAMRKMP
9 EL-----SALSGNHKEFLTAFNGQ-----EGWILSSDLTAASDLL--PLDL
10 VSAGVEGICDSGR---LLPDEELGLRVCTGPFDL-EW-----V-----
11 -----RG--SPRR---SSTGILMGAPPTWCLLSLIHLFWLEQARR-----
12 -----VYGNSGGQ--LPARIFGDD-LVAAFDKRQKLAYERSVTVCHGLLSKKGKHSFH-
13 --DSH-GVFLERLRFVASRHIVEVDHT-----
14 -----NRGPLQ
15 VMGRL-----
16 -----
17 -----TTRHNKIKSISLL---RTLPLR-PLTVLQVSVAA
18 GRRHV-SKGTLPPLA-VGGVSDGLMQH----GFPELIRRCQI-ALWPGLPAFYRKL-
19 GIPAFLLPT-LLGGGGGLILPEGWEAPITRFS-KLVRASTTALVTGNGRLSYV-----
20 PLARL---PPPVLMQALEEAD-QLLPAHPHRVSHVTPKPWLGHIAW-----
21 -----HDMGVWVWDFRITAAMVAQEGYLLQ-----MPME
22 CLGRPL-----NITVGRWKD-----QVWKTRLALCKKAWWASV-----
23 ----TRARYTI-----ARAVRKCTEWPHVWLP-----
24 -GSPDPDHPGVYGYIPMYIDSKPV-----MEAKAAIALRQ--GR-----
25 -----
26 >GENC01041260.1
27 -----
28 -----LRSKVDQWFDQPS-----
29 -----
30 -----
31 -----EFGDTLF-----
32 -----R-----DFPSFRDDTRFTN-----
33 ---QGDENAAQYFY-----
34 -----AMSLVAADENSLP-----EAKL-PSRG
35 VTEYVVDVTADVH-----LERER--ISKIVRDI--SLRWLAKLTGPPRA-----
36 -----
37 -----
38 -----
39 -----SVVALRERGFKARIVTKSPVELV-ECGHLLR--SLVWPMLRNDP
40 HV-----HGSLSGARLESFINDLKERPLVNPVWVGRPSIISADLTAAATDGL--YRWA
41 ISAVWEGVCDGAG---FGEDVRKMGRVLLGPMKM-SYPAS-----A-----
42 -----GV--DEFT---SKRGCLMGLPLSWFVLNVINIWAAKSAVE-----
43 --KSCRGRVGKPA-D--FRFAICGDD-LVGIMPARAQRLYQRNIEVVGSGLSEGHLES-
44 --DRL-AFFTEQAATFTRVT-----
45 -----RSAPRG
46 TLAEK-----
47 -----
48 -----
49 -----
50 -----
51 -----
52 -----
53 -----
54 -----
55 -----
56 -----
57 -----
58 -----
59 -----
60 -----

-----AGLREEYLEPVEMV---GAIPVK-SIVHPSHFAGL
RQGP---AATPSWAA-AGPALSSLAVP--WVHDSTLKVVSR LGR-YLRPE-WRLLASS-
GIPPELPR-QLGGGGFPARIPRRGFLAAPK-QWRSALTCALLTGDHKWLS-----
RLRNSWLT TG VYGDVMAEALSM-AEEEFQDYPLAELEIPLSEVRTTR-----
-----FTGGGLAITEADLRTRVATVWAPA-----LTIA
SPFIEG-----RXYCERFHS-----FRKRLSLMVKRKRLPGHL-----
--QPAQFNDTEA-----VSRLISLSEGRSFLVP-----
---RESLPPGLNIDF-----IGTRRPCGDGE---DRAIEVLT--

>LA534138.1

---RGV-----DRAR-
-----ELLQFSSFARALP----PADDLV-CKEA
LRDHREILNT-PV---S--TPVA---VVSELRSY---ATR WGQRMARFANT--SVA---
-----KL-----SASSSATLDYSRRLGGMR-----SDLK---
--GMVDGWFEEP-----ASDVTEFHPYPTFSDPTRFTATGEMN RAMHPAPTRGSV-E
Y-----VVNVLE-DPDLE-----RSRV
ARIIRDVSLRRFRAREG-----PL-----
-----PC---KATTVKERGYKCRVVT KSPADV V-EVGHLVR--SVVWPMLEHDP
RV-----QASLDGGRLEE VFKGFS DRQIECPASLG SLLV SADLT KATDGF--SRES
IYAVWDGVCEGAD---LPSDVHALGRRILGPMSV-EYEDD-----E-----
-----LE--DLES--SSRGCLMGLPLSWFVLNIINLWACESSIR-----EC
CHRLSLPSEVCKD-L--LRFAICGDD-LASVFPAEAHDGYERRISDVGSGLSDGKHLVS-
--LHV-LLFTEQMCWFEEFERPA-----
-----PPYTLL
AWLKP-----

-----GKAVPEGFRATYTCFTA VRMI---DYVPVR-SLVHPGHFAIK
RVMGPIS-FELPSWAT-SGPAITSAIPE--WIGVEKRRRIHQVVK-VLRPE-ARALASV-
GIPPFVPR-EIGGGGFPPAR-PGRVLRDCP-RSYRLFLFSLLVGDQKKAETVAR-----
RIINLWRTCGVAGDLLSDALAEAEFEVEQKPLWKSPSDGSLAERLST-----
-----VIDDHLTMTEDDALLRLASVWAPA-----LGIAGF
QQGRPY-----CKKFGAF AK-----QYRALVGAGSARITGRG-----
-NPLRNEDEAAV-----ALRLKEICAGPVVFP-----
---RDRIPPGLGVVVQGM-----CSRKRAVGLHRKIRSNRMESA---

>LA710658.1

-----PPQKKT LHDNPRVDVWWNACEAV-----LGSRGAIRPKGRGATRQVLP RST
RQRAAFLSVSLPSAAHT-----

-----KLLSLVW-----EALLG
VAAAYRPLPVWGSRF-----GNKRCLELQRLAL----WLVRTVD-----
-G-QGIGATIAWLKSAASGARN-----TCVTG
A---ALLPSTRFLVRKILIGDTRLE----SLDQLSFLSRSLP----EGDSVV-ESRT
LIAHRDNMTT-SF---R--TSPN---LISNLRGF---AMAFATRHLKKEDLREEVV---
-----PTPSASLDSSRKEGGCR-----EEVR----
--RMFMRWLIEC-----PTHVVPDHAUNQAIFSEWLDPS-E
V-----YGLRRAQATSHV-----ARAA
A--YKTA--ETLEH-----
-----RVLTVPERGWKRRVVSAPPATFAT-VAGTVLN--RAMLRGVRRHG
PC-----AMFLRGDRKGAVEQAVSHASV-----GCRFVSTDLTAATDRL--PLDL
VRVVDGLCDGWS---GLPPVWAEALYALTGTQTL-RY-----P-----
-----WG--QTIQ---SESGVLMGLGPSWPIMSIIHAWWVELAAS-----
---RVGIPRSVA-R--RTTAIGGDD-LLGAWPPRLEESYRRLVLETNGKPSKGKDFSS-
--DTS-GNFTMTFWVGEA-----
-----DGTPQI
RWS-----

-----AAIPTK-GLVGTSID---
-----ELGAAYES-LGSEPGRCLRG-----RRVLK-ALRPHAWRVCREA-
GVSVNAPR-LLGGAGLPPLRGSLARVD-FK-KWHALALGKFLYG-SGQDQIP-----
FSPPSWV--EAADPAVWEARQV-AE---QRLRAEAEIGIVSFDTR-----
-----PLCGTPGSKHVVSISDQMAWW-----
AGARVFSD-----TPFPVATE-----MVSLKKYHRLIKRWCSRTKEGIPSSL
AVKSGRNSRFAL-----TAKARRNRDRWVIR-----
---NALFDNGIPVD-----PRHRPII-----

>U90136.1

-MHHKVNVTQREVHFPMDLLQACGASAPRPV----ARVSRATDLDRRYRCVLSLPEER
ARSVGCKWSSTRAALRRGLE-----

-----ELGSREFRRRL-----RLADDCW-----RAICA
AVC-TGR-KFPS-FSVTDRP-----ARARLAKVYRMGR----RLLVGVV-----
-C-RGE-SVVSDLKQECADLRRVIF-----EGSTR
IPSS-SLWGLVGVLGW-----TSPE-----RAMQLTFIGRALP-----YGSPDV-ERRA
LASHAATLSI-PA---E--CHPN---YLVAAEQF---AKSWADDNLPRKFR--IYP---
-----I-----AVQESSCMEYSRAQGGL-----QSFR---
--KGFVGY-----DP-A
A-----PSADPD-DLELA-----KERG
F--SRIRASWYSTFRYRGEL-----KSTNQS-----LE-----
-----A-----RVAVVPERGFKARIVTTHSASRV-TFGHQFR--RYLLQGIRRH
AL-----VDVIGGDHRRRAVETMDGDFGLLRP---DGRLLSADLTSASDRI--PHDL
VKAILRGIFSDPDR-RPPGTSLADVFDLVLGPYHL-HY-----P-----
-----DG--SEVT---VRQGILMGLPTTWPLLCLHLFWVELSDW-----
APARPNHSRGRFVL-G--ESFRICGDD-LIAWWRPRIALYNQIAVDCGAQFSAGKHLES-

1
2
3 --KTW-GIFTEKVFTVKPVKMKVRVRSEPSLKGYVFSRSSAFSCRMGGKGITGI-----
4 -----RAARLY
5 TIGAM-----PRWSRRIRDVY---
6 -----
7 -----PGSLEHRTASQRYGEPVTVYRFGRWSSAIPLR-WAVRAPTRTVG
8 NPV-----QSLPDWFT-VGPAASSVAAD-----SNAFGAVSRVLR-RMFPGLPRKLASA-
9 GIPPYLPR-VFGGGGLVKSTGLTTKIGAVASRRWMSRIGHDLYRSRERKS-----
10 TLGRVWT--LSTSPAYAASLHE-VEKFMDRPDIILTRKCRNPML-----
11 -----KHARELGLFEEVFESRVGGGILWA-----
12 -----SLNGKALVE-SHSPSILQVSRNLRRSLACPSGGF-----
13 --LRPSAPIGKL-----VQRHTLPRGT-VWFLE-----
14 ---SSATDSARQGGMGL-----PPPPPPPLGGG---GMAGPPPPPF
15 MGLRPESSVPTSPFTPSMFSERLAALESFGRPPPS-----
16 >MF176258.1
17 -----
18 -----MPLVARPT----SMRVATRSEVSRRV-----
19 -----
20 -----
21 -----
22 -----
23 -----
24 -----NLAAETW-----KAICA
25 AVATRKWPKFNLDFT-----DRRRFCTFQALAR-----HLLLSAL-----
26 -A-RGPVETIKRVKSWTAQIRE-----DVFL-
27 ----HGRCPPSAFRGLVGALGWTAN----SHLQLSYIGRALP-----YGDATI-CEHA
28 LKSHADVLGSLGV-----TPPS---LLGAARGW---AKEWAIKNRLSLDL--IRE---
29 -----P-----PIQESSCLEYSRREGGYA-----RAIK---
30 --EHTARCLTL-----P
31 S-----SVRHSM-KEGTR-----ASNP
32 GRVLKDLAF---ADLSNTA-----
33 -----RVVAVPERGWKARIVTAHAAARV-AFLHSLR--HGLFAALRNDR
34 RT-----KLVAEGLHREAVEALFATGVDP-----RCTIVSADLSAASDTL--HGDL
35 LQAIVEGLGDALG---CTSKFMDEFTNAAVGSYTL-SY-----P-----
36 -----DG--STVT--TQRGALMGLPTTWPLLCLTHLFWCDQSLV-----QLHG
37 RGGCREKETPGSS-R--EAEVICGDD-LAAAWRPDRVAAYEEIAVLCGAVFSPGKHLKS-
38 --KVY-GIFTEDIYSVQVSCVPTRVTREVSKTWTGPVDKVPM---NLVNELGQ-----
39 -----LLRWRA
40 QRYKV-----GKTGAFQFPPR---
41 -----
42 -----NVTLSYVSKETGLKMNPVSVLFREWSTAIPLR-WAVRAPKHAPG
43 MR-----GDLPPWVT-VPLAAHAVASA---HPGRWSKVCRVVK-LCYPGMAKFFASH-
44 GIPPYLPR-VLGGGGGLPTPLGDRVKIGRVASRRIRKALGGALYRSTDPT-----
45 TFGSIWT--TALSPAYALALAD-AQDL DAMHPDAFKVRQTRSLRPP-----
46 -----FKDNGNYADFVEMRAGKGS LWA-----LRQGAQLVGR
47 SFTP-S-----P---ARVSN-----ALRTKVRKALAKRGYLR-----
48 ---SSAPVSKL-----IERASVKPSDHMWIRE-----
49 ---RPVLSFGRFGGRMSGKKEEPVS-----FQWKKRESVKT---GS-----
50 -----
51 >GDUK01014407.1
52
53
54
55
56
57
58
59
60

-----SCMRWLR
MLKTGTSPAADAAGSVGHSGPTESWTRSGSAV-----ARAPLGSDVLASGMSLEKRLKALA
SRIKRANLRL-----

-----RSLLGTW-----DAGLT
AFFGRQIRRGEV-----ADSVVVWYDHVRS-----WLV RVAA-----
-T-QGVEVAAAELKKFSSDARRAWV-----
-----EGSPPRHHFWRSCPEWLRRKQG---LWAQLSYLGRALP-----LGVDHFH-EKKA
LDQHYADLTS-SF---S--CSEE---VLASARAF---AERWARSHLSRAPGPDWL---
-----EFPSGSSATFGRKRREGGFA-----TDCQ---
--ILLESAAEVD-----
-----VTPPDH-VPVGP-----WSGD
VTAFRLVTAALVEAHGRRDT-----LP-----
-----KG---RVAVVPERGHKVRVVSAMERHAL-ILGELAR--KRLFKGLRKWG
PL-----AAVLRGDSRSGVLRLLGV-----PGEFVSADLRAASDLI--PLDL
ASAIVDGLEASGR---LMPAEIWGLRLGTGPQEL-TW-----P-----
-----SG--KTAV--TSRGILMGLPTTWALLNLHGWAWWSQAAC-----AVPVQ
SLPRASAARGGRP-A--LSCAICGDD-LVGVAPPVIDAYEAAMRSIGAEMSAGKHFRS-
--TRR-AVFLEELWEFRGSAVSVV-----
-----DGNPIY
RTIRE-----KG-----

-----KRVRRQVVNSTRLLRADRLVRA---TAIPLK-GMVSAPGVPGR
-----IDPPDWWV-AGLAESALADR---GFSRAKISSVSR-SLRPDLVRFRRAS-
GIPPFLPR-ALGGAGLIAPHTRCEAP---RLHRKAIATLIWGYDPDSTL-----
AFDRVWS--DSVPGDWRELAID-MV---DAEFTPEVTRVLPVGSR-----
-----PTGSWVAIGHPDQVREKAVADNASA-----LSSSLGSPSD
VVG-----YPTIPVLAR-----RLRKVREDLASRWASAK-----
---PTKRPVDEL-----LLLWRQRAADLRVWVP-----
---LTDQTSRWNDWLYVSATSPYK-----KSWRRLVISSV---SREDLPAVTE
WWLPGGRISSSTMDPSPPEWR-----
>GFKT011160020.1

-----PMKEPVDCRLSTPAGFSGTVPP---PGRTKAARQGTIPVRRSRGSASA
LPGKIYGWSRRQR-----

-----DRFAMLL-----SSFDA
ALAAYSGVVVSRR-----GTRSLPPLRLFRMTMRKWLSTAR-----
-G-NGVEFAIASAKEFSAACRAGWI-----
-----SGTVPDHFHMKWLPEPVRKSG---LWAQLSFIGRSLP-----EGGDRH-EIEA
LANHKAALSS-SF---E--VPAD---VLTSLRNY---SEDWARRHLAADPDPSLLC---
-----EP-----CTGNSATFERTRREGGFA-----QSIT---
--DLVSSSPTDN-----LL-P
L-----ESMPFG-PTQGG-----ALPV
H--VLEVSL---SRYRNGS-----DP-----
-----KG---RVSVVRERGHKVRVVSAMETHEL-VLGHAAR--RRLFKGLRRER

1
2
3 RL-----KDTLKGDFEATTKAFVGC-----AGTVISSDMKSASDLI--PLSV
4 ASAIVDGLEASGR---LLPVEVAGLRCTGPQHL-VY-----P-----
5 -----DG--SEIT---TRRGILMGLPTTWAILNLMHLWCWDSADR-----
6 -QYRLEGHPFRATVR--SDCRVCGDD-LIGVGPDSLRSYDRNLGLVGMILSPGKHFRS-
7 --NRR-GVFLERLLEFQTRKTVY-----
8 -----EHAVIY
9
10 RKVGH-----
11 -----
12 -----RRVPVDRSHIPVVTRVTVL---NTIPLK-GLVRASVLGR-
13 -----DDPPVWWA-AAVAESSLLSD-----YPRKKIFAAAR-TLRPGLSRQFRRL-
14 GIPPFLPR-ELGGAGLVGSSDRVDAP----AFHRKAISSLVWGSDATAAY-----
15 SFIRMWQ--GFEGHPWKTAASQ-ETDTWFADYKVTRPGKMYPDR-----
16 -----YGFLDGESLRTKSTMLNSAVYETF-----LGPDP
17 ATH-----YPSLRIVAS-----RLAKVRKDLVSRWPSVR-----
18 ---PVGKDLGTI-----LEAFEEKLCTLWVTP-----
19 -----YDASGYFDDSLLLMDESVMYQ-----RRFRQLVIAGLMREDRMGDLLFPS
20 WLPSTVVS-----
21 >M64034.1
22 -----
23 -----MKEPVDCRLSTPAGFSGTVPP----PGRTKAARPGTIPVRRSRGSASA
24 LPGKIYGWSRRQR-----
25 -----
26 -----DRFAMLL-----SSFDA
27 ALAAYSGVVVSR-----GTRSLPSSLRLFRAMTRKWLSVTAR-----
28 -G-NGVEFAIASAKEFSAACRAGWI-----
29 -----SGTVPDHFMMKWLPEPVRKSG---LWAQLSEIGRSLP-----EGGDRH-EIEA
30 LANHKAALSS-SF---E--VPAD---VLTSLRNY---SEDWARRHLAADPDPSLLC---
31 -----EP-----CTGNSATFERTRREGGFA-----QSIT---
32 --DLVSSSPTDN-----LP-P
33 L-----ESMPFG-PTQGQ-----ALPV
34 H--VLEVSL---SRYHNGS-----DP-----
35 -----KG---RVSVVRERGHKVRVVSAMETHEL-VLGHAAR--RRLFKGLRRER
36 RL-----RDTLKGDFEATTKAFVGC-----AGTVISSDMKSASDLI--PLSV
37 ASAIVDGLEASGR---LLPVEIAGLRCTGPQHL-VY-----P-----
38 -----DG--SEIT---TRRGILMGLPTTWAILNLMHLWCWDSADR-----
39 -QYRLEGHPFRATVR--SDCRVCGDD-LIGVGPDSLRSYDRNLGLVGMILSPGKHFRS-
40 --NRR-GVFLERLLEFQTRKTVY-----
41 -----EHAVIY
42 RKVGH-----
43 -----
44 -----RRVPVDRSHIPVVTRVTVL---NTIPLK-GLVRASVLGR-
45 -----DDPPVWWA-AAVAESSLLSD-----YPRKKIFAAAR-TLRPGLSRQFRRL-
46 GIPPFLPR-ELGGAGLVGSDRVDAP----AFHRKAISSLVWGSDATAAY-----
47 SFIRMWQ--GFEGHPWKTAASQ-ETDTWFADYKVTRPGKMYPDR-----
48 -----YGFLDGESLRTKSTMLNSAVYETF-----LGPDP
49 ATH-----YPSLRIVAS-----RLAKVRKDLVNRWPSVK-----
50
51
52
53
54
55
56
57
58
59
60

1
2
3 ---PVGKDLGTI-----LEAFEESKLCTLWVTP-----
4 ----YDASGYFDDSLLLMDESUYQ-----RRFRQLVIAGLMREGRMGDLLFPN
5 WLPPSTVVV-----
6 >AF039063.1
7
8 -----
9 -----MKEPVDCRLSTPAGFSGTVP-----PGRTKAARPGTIPVRRSRGSASA
10 LPGKIYGWSRRQR-----
11 -----
12 -----DRFAMLL-----SSFDA
13 ALAAYSGVVVSR-----GTRSLPPSLRLFRAMTRKWLSVTAR-----
14 -G-NGVEFAIASAKEFSAACRAGWI-----
15 ----SGTVPDHFMMKWLPEPVRRKSG---LWAQLSFIGRSLP----EGGDRH-EIEA
16 LANHKAALSS-SF---E--VPAD---VLTSLRNY---SEDWARRHLAADPDPSLLC---
17 -----EP-----CTGNSATFERTRREGGFA-----QSIT---
18 --DLVSSSPTDN-----LP-P
19 L-----ESMPFG-PTQGQ-----ALPV
20 H--VLEVSL---SRYHNGS-----DP-----
21 -----KG---RVSVVRERGHKVRVVSAMETHL-VLGHAAR--RRLFKGLRRER
22 RL-----RDTLKGDFEATTKAFVGC-----AGTVISSDMKSASDLI--PLSV
23 ASAIVDGLEASGR---LLPVEIAGLRCTGPQHL-VY-----P-----
24 -----DG--SEIT---TRRGILMGLPTTWAILNLMHLWCWDSADR-----
25 -QYRLEGHPFRATVR--SDCRVCGDD-LIGVGPDSSLRSYDRNLGLVGMILSPGKHFRS-
26 --NRR-GVFLERLLEFQTRKTVY-----
27 -----EHAVIY
28 RKVGH-----
29 -----
30 -----RRVPVDRSHIPVVTRVTVL---NTIPLK-GLVRASVLGR-
31 -----DDPPVWWA-AAVAESSLLSD-----YPRKKIFAAAR-TLRPGLSRQFRL-
32 GIPPFLPR-ELGGAGLVGPSDRVDAP-----AFHRKAISSLVWGSDATAAY-----
33 SFIRMWQ--GFEGHPWKTAASQ-ETDTWFADYKVTRPGKMYPDR-----
34 -----YGFLDGESLRTKSTMLNSAVYETF-----LGPDPD
35 ATH-----YPSLRIVAS-----RLAKVRKDLVNRWPSVK-----
36 ---PVGKDLGTI-----LEAFEESKLCTLWVTP-----
37 ----YDASGYFDDSLLLMDESUYQ-----RRFRQLVIAGLMREGRMGDLLFPN
38 WLPPSTVVSGBP-----
39 >MF190030.1
40 -----MGNK
41 VGRFSSRSCDRYLKRAPLPGDRKGRVGRVIDKLEVPRTSRKSKGSRMRGARPPMAES
42 Y
43 KIKPVGWRKGFMTGP-----
44 -----
45 -----REIEGLW-----VALIA
46 TLTSISPRICRFNMRCRA-----DRDRHGAMVRLAV----WFTRRVD-----
47 ---RSSYDLLKSAKKLTAGFREASI-----LRRPY
48 NWVD-APPGWAKRFRPILKGIRLSEA----VLDQLSYMGRALP----KGGPRV-LQEA
49 LKDHKDCMLKPSP-----LPEE---VFGAARDF---ACRWADRYVASGGRLSHPG---
50
51
52
53
54
55
56
57
58
59
60

-----TTESSVLGSTRRDGGIS-----SSLC----
--GLTYSVKPLP-----MDVQDEVRALVSSLK
C-----DTPSEC-DVDAA-----LRTT
S--IREWAIDESSPDVIPEN-----
-----KVLALPERGYKTRVATIPDLPRQ-ILGGFVR--SYLIPALRRTP
EC-----ADTLRGDDERA AKRIGSSQRSFS-----TDMIVSSDLTAASDTL--SREL
LLALTAGLVDSGK---FPSWFGPALRQLVGPMLL-RY-----P-----
-----DG--SSIT--SQRG TLMGLGHTWSMLSLSHLFWLDYSNK-----
---TVHKTIARSV-E--RTAAVCGDD-LIVRAPWRWINSYHRIAKMCGAQFSSGKHFIS-
--RER-YCFLEQLWTLKSKEIVPAYTRALDK-----
-----RIPKGG
DQFPP-----LG-----

-----TDPRKSVRRVMKFHIPSKCVRS---TAIPLA-GLTSGAGSDPE
KPLV---VGTPTWAK-LGLAVENTSRT----PVAARKVRAVSR-ALYPRLPRWMRKH-
GIEPYLPR-AYGGGGLLPP-GNVSHLGALA-PQVRKLVTVTAYGRCPKTAA-----
RLSRIYL--PMNRTATNFTEAQ-EH----TDIRFKRVGAILCRRGK-----
-----VPFNHGA VKLGTLT DAPER AALAIARE-----LSMVLFP PPR
NRK-----KSTISPFQV-----GRAFHSAMRKA IKKWPSAH-----
-PCSAKVTRRVL-----LQKARALEEEFVWWSP-----
---RRRIPEEGFGGWGLA-----PAAQRVLKDSL---RWSSDC---

>KX883516.1

---MKPPRPTEVHKDFRLPVLLRACRVGGTWS-----RDVSRYLVPALQDGVGV RPEGEGG
NPQTRPRKPAKSVDLERTR-----

-----YDLTVRWRKEFRSRG-----RLVEGAW-----AAVHA
ALV-AAHPDLGR-ADMSTRA-----GRDRIWAWRRLAV-----WMVRWID-----
---RDPRDMVRGWKQLAATFREASI-----MQKPY
DWAA-APAGFPKRLRPVLAGVRLTPP----VLDQLSYIGRALP----QGDAWV-QRKA
LDDHRECLSSAPR---P--LANH---VLAAARDF---ARSWASRHVRPSG---RIS---
-----HPKVLESATIQEPRRSGGVA-----AHL C---
--KQGFDL-----PPFEGLTDEL RDQIRAM-G
E-----SAPSDC-DLEAI-----GRST
A--FRDWALGQLCCPDAS-----PP-----
-----RS---RVEAIPERGFKARVVTVPDL PES-ILGSALR--CYLLPALRRTP
EC-----ADVLSGKEELAAMRIASSRRPIE-----GDWIVSSDLTAASDTL--SFEL
LQALVDGLEESGT---LPPWAIRALKVLVGPQEL-VY-----P-----
-----DG--SVLV---TRRG TLMGLGHTWSILSLSHLFWLDWAHR-----
---SVDPRVADV V-R--RTAAVCGDD-LVVRAPLKWVQSYHQIAEASGARFSTGKHFVA-
--KRR-YVFLERIWSLKSSLQRVRLPRERTR-----
-----RAGR WV
REFPP-----

-----LGARPVKVRPTVVRSIPSRSVPS---TAIPLA-GLTAGEVDCHF

GPVRN---LVLPNWVR-LGMAVETLLHS-----PEAARKVRKVVR-ALYPGLPRWMRRH-
GVQPYLPR-FLGGGGLLPGRNVTRIGRLP-RPVAKAVTTWVYGRSPRENL-----
SLAKIYL--PTTRRITNYAVGE-QH-----TAVRFQQVSAVLCRAGK-----
-----KPFDPGHGRRLGTLSEAPEAAALAIARE-----LSMVLLPEPS
EERAYRR-----LLRPSQVGQ-----TFRRLVLRANRRWPGSK-----
-PICPRAPRRAL-----LARCKAVEEGFVWWAN-----
---PRLLTQFGFGGWGLE-----PAAQRVVKDAL---GWSTFRPTPG
RGTH-----
>GEND01007370.1
-----LDHEPPT
MGSSKKTARETHVSNVSLHALLRNPRVGGWSW-----KDVSLYLAPARRSGVQGESGERG
LPSRTGLATHANLEHRGASRRGQGHAKRGPEPRRKAPKGGYTRQRCVCSREFKSSG----

-----RKLEAGW-----NALTA
ALAACHPAALRFDLSVRA-----DRDRVGALKRLAV-----ACVRRLD-----
---RGVGELMAELKSLATTFTASI-----RQVP
YSWDTCPAGLRRHLEAALRGFRLTTE-----TLDQLSFLGRALP-----PGDSTV-EERA
LRVHRDVITSDPA---P--IPPG---ILKAAEAF---ACTWALTYLHRGGRLAHPK---
-----TMESSTITSRRREGGVA-----ADLC---
--MKSMDLDPLP-----ETVQAEVRRQVLELEG
-----PLPSEC-DLSAA-----FRTT
A--LRDWAVTTLTEPGYVPTH-----
-----KVLPIPEAGWKCRIATVPELPAS-VAGGLCR--SYLLPALRRTP
EC-----ADVLRGEEDKAARRVAASRRPIE-----GDWIVSSDLTAASDTL--AFGL
LQALVRGLALSGR---LPPWVVRSLVLVGPQRL-TY-----P-----
-----DG--ETVI--TARGTLMGLGHTWSLLSLSHLFWLEAAHA-----
---SVPRPVSEQV-R--LTAAVCGDD-LVVRAPLGWIRRYEEVASLTGAQFSQTKHFVA-
--KRR-FVFLEELYTLRGTQQAASVTLGPTS-----
-----RGGRLV
RAWPA-----PG-----

-----TRPPKVRVTRLRTYASGTVKS---TAIPLA-GLVAGDVMTHQ
VMA-----DSVPLWSK-IGLAVESLAST-----RERALRVRRVTR-HLYPALPSWLLRH-
GIRPTLPR-ALGGGGLLPNSGNCMLRGLP-RRVAKMVTAVTVGRCPRTA-----
RLSRIYL--PLSRRATNFAEAK-VH-----TDIRFKRVGAVLSRSGR-----
-----EPWNGKGVRLGTLTEAPEQAALAIARE-----LSMVLLPEPK
KGGSSH-----NAALSPFQV-----GKQFLLALTKALRRWPGAK-----
-PVSARAPRKSV-----SDRTRAVEEAFVWWAP-----
---SHRIPEEGFGGWGLA-----PAAQRVLKDSL---GWSAPSP---

>GENB01008321.1
-----LDHEPPT
MGSSKKTARETHVSNVSLHALLRNPRVGGWSW-----KDVSLYLAPARRSGVQGESGERG
LPSRTGLATHANLEHRGASR-----
-----RGQGHAKRGPEPRRKAPKGGYTRQR-----
-----VCWSREFKSSG-----RKLEAGW-----NALTA

ALA-ACHPAALR-FDLSVRA-----DRDRVGALKRLAV----ACVRRLD-----
---RGVGELMAELKSLATTFRTASI-----RQVPY
SWDT-CPAGLRRHLEAALRGFRLTTE----TLDQLSFLGRALP----PGDSTV-EERA
LRVHRDVITSDPA---P--IPPG---ILKAAEAF---ACTWALTYLHRGGR--LAH---
-----P-----KTMESSTITSRREGGVA-----ADLC---
--MKSMDL-----DPLPETV-----QAEV-----RRQVLEL-E
G-----PLPSEC-DLSAA-----FRTT
A--LRDWAV---TTLTEP-----GYV-----PT-----
-----H-----KVLPIPEAGWKCRATVPELPAS-VAGGLCR--SYLLPALR RTP
EC-----ADVL RGEEDKAARRVAASRRPIE----GDWIVSSDLTAASDTL--AFGL
LQALVRGLALSGR---LPPWVIRSLEVLVGPQRL-XY-----P-----
-----DG--ETVI--TARGTLMGLGHTWSLLSLSHLFWLEAAHA-----
---SVPRPVSEQV-R--LTAAVCGDD-LVVRAPLGWIRRYEEVASLTGAQFSQTKHFVA-
--KRR-FVFLEELYTLRGTTQQAAS TVLGPTS-----
-----RGGRLV
RAWPA-----PG-----

-----TRPPKVRVTRLR TYASGTVKS---TAIPLA-GLVAGDVMTHQ
VMA----DSVPLWSK-IGLAVESLAST----RERALRVRRVTR-HLYPALPSWLLRH-
GIRPTLPR-ALGGGGLLPNSGNCMRLGRLP-RRVAKMVTAVTVGRCPR TAA-----
RLSRIYL--PLSRRATNF AEAK-VH----TDIRFKRVGAVLSRSGR-----
-----EPWNGKGVRLGTLTEAPEQAALAIARE-----LSMVLLPEPK
KGGSSH-----NAALSPFQV-----GKQFLLALTKALRRWPGAK-----
-PVSARAPRKSV-----SDRTRAVEEAFVWWAP-----
---SHRIPEGFGGWGLA-----PAAQRVLKDSL---GWSAPSP---

>GBWZ01004682.1

-----GSRTPWRKSLPQMNRGSR-----

-----QKVRATW-----SVIRA
ALAASGFVRCDLSTLA-----ARDSANGLCRLAW----WLWRTTA-----
-H-SGIEYTLKLLKTWLSETRRWVV-----EE
VPKPKRRGGMARFLHGNFRSMVDSNA----KALQLSYAGRALP----VGGPSV-ILRS
LRGHQDAFASNPV-----TPPE---LLTSARIF---AREWAKRHLSSSPSFARAG---
-----LNESSLESTRAKGGVL-----AQTR---
--ERIYAIPDEA-----LRERKSSLQSREE
F-----ADLSSC-DLDIV-----ARES
L--LRDSALEELNRSLENGNPQA-----
-----EVRYCLERGLKCRVVTKSPWALV-YLAHSVR--HELLNALRNDK
RT-----KHVLKGDH QEAVEDVFSTPRRID----NRQIVSADLTSASDLL--PLDL
VGALVDGLRDTAR--LWWPEWMLKVLGIVTGPM DL-WYPA-----
-----ELGEDEL TIVGTRCGILMGVPTTWSILSLVQLFWAE EAWR-----
SRFPNPAWGHPSM-T--PETAICGDD-LVGWWPKSVIKTYEAVAKRCGAQFSAGKHFHS-
--RTH-AVFTEEIYRLRSISEETPRCPDWESLPKLQQRWVRSGTKRR-----

-----PRKRLV
NFLRT-----

-----ARLGGPVGKGVVKY----PSIPLA-GLARPSVLPGS
Y-----QTPVWWIA-IGPTYEAVAAQ----TGNPSAARRVLR-TLHPNAWRWMRER-
GIFPELPR-ELGGGGLPPLRGSADV LK-LP-PFLRFGVAKLLYGTKFRDLR-----
SPGALWS--TTFAPewRVMAAD-HAEVKFKTSaVNCHPGRVPFHGKG-----
-----FELGGVKALEDVTTRLAGELTLMMEHP-----K
DGRQAF-----SVVPRFIAR----SVRKfYSARGKVARRSKMKG-----
-IDPSGVPRTAL-----LKRWASLGQTRVWFSE-----
---PAAAGEGFVLGIP-----GQVKRRIAPAL---GWARP-----

>GBPT01024369.1
-----L
RMAVRKPSTQAARERVTTLSLVEAALIGSRL-----VPLAGIVRPARARALRSDRGATL
NPKRRGVGQPLEPIWAQGST-----AVPIPNPSSSPHKACLKPIRQQDGKPPKA
SAMKKGHKIKDHFRFRFLSPRSSEKVRASANPPKEDHDVDWSSVR-----
-----RLYPEWGRISKRGHRQLKTVW-----AAFRA
ALAAVGSTRVSLRTNS-----GREALRKWCKLAH----WLVRTTG----
-L-SGITHVLATLKAWTATVRRWV-----EGVRE
QD---RQGG LAKFLAGSLRFLVTRDD----AAVQLSYVGRALP----QGTLSV-QRKA
LVSHKKALGE-DY---R--TPRH---ILREAREF---AAGWGDRYLRSVPL--MNC---
-----GFSQSACLEVPTSKGGTL-----EATL---
--ERTNALQTDV-----NWSKAVADWRDTLKS-S
I-----PTLPDC-DREIL-----ARGT
A--LRDRSIVELREAVSQSG-----HP-----
-----RA---AVAVLPERGSKARVVTKSPWALV-QLAHPIR--GWLLSGLRRDP
RI-----AKVLEGDDHHGAATGLVDVPR LAD----GREFLSADLTSASDLI--PLDL
AAALIEGLMEGTG--DTLPDWAREVLR L ATGPMLV-DYSTAWPPSGGRRPLP-----
-----GP--CEFI--SRRGALMGLPTTW SLLCILQLFWADRAWR-----L
TFPNIAQRTDPGNPGMTPATIICGDD-LAAWWPKQVSDAYEQTVVSCGGIFSVGKHFRS-
--RAV-VVFTELLFHA EKSTVRALAGAGPDKISAPIASWER-----
-----EGRPFL
MKVRR-----PHAKSTARRRRSER
GRRPMIPKM-----
-----TRGIGPHDLSDHRRVVTRLVPF----KSIPLK-GLLEAEGGGPL
ADNH---VRQPRWVN-IGPVADRISKM----TGNPYAVRRVLR-TLPHYWGW FQKH-
GFVPTLPR-FLGGPGLPPRTGDL DRLA-LP-RDLLLKVRTL VHGTKWHLLR-----
SPKAPWT--QMGAGFYRQMVAG-HAEARLQTHARC ARVGRVPYNGKG-----
-----VLLGGVDVYEDLATRFENEATVGIFGC-----RDD
SRRSAF-----KIRPGQVSQ----ALRKFWKGVKVPTGEKHRLG---GEERRR
GRLLNKTSARNL-----ERRWIELTSSRAWHAD-----
---PMAAQEGFVLGLN-----RRVKLDLAIKLGQTGRAGGTESPG
PSAEKSPAGDAPAASGSPTAS-----
>GBPU01010966.1
-----MPSNK

<https://mc.manuscriptcentral.com/vevolu>

VAAIVDGLVEGSE--GTLPAWAQEVFRLSTGPMVL-DYSKAWDGSVPL-----
-----KGP-TQMI---SRRGILMGLPTTWLTLVQLFWSDTAWR-----N
GFPEIMARRGFSAPS--PATVICGDD-LAAWWPEKVIEEYEALARTCGALFSAGKHYS-
--PKN-LVFTEVIFRSVSTAQDHPPKRSFWERERLPFALIGKTVRSSNVGAKAGQVEGR
NLTKKERRAKLRQSE-----RTITLV
RRDGV-----

-----RRVSSLSSKDDRRVVRKLAVM---PAVPLR-GLVEPELGLNE
SF-----IRQPRWVN-LGPAVDCAIQI----SGLPHAVRRILK-ILYPGYWAWFRQQ-
GFEPTLPR-FLGGPGLPPLTGSPARLR-LP-RRLRVAIGILLYATPWKEIR-----
SPAAVWT--CLGAGSVRESARE-HT-----EWKFSKSAVACRKGKV-----
-----PFNGNGAVLGGHSAVEDVTTRMYNEL-----ALILDPE
DRRSRF-----KVKPREVSS----AVRKFYDRLLPRDPKQRVKT----HPYIT
RGIAPSPVVSAL-----RRRWEGILGQRTWYST-----
---PGASSEGFCLGLA-----AGVKHELSSLALGHSGRRTRRSPS-

>KX883605.1

-----MVARLVEDRRRDNRSQGAAAKQPA----SRLGGFMGHRRGDRQRVAALFPW
RRRLKGSEW-----

-----TRLSEAW-----RVIQA
AIAAAGFLTVGLEDPA-----SREFLRGLCRLAW----WMVKTGH-----
-Y-TGVERVLSDLKGFAGKARLWVM-----EE
TPPTLRRGSFARWFRGIGRSLVDRDE----RAMQLSFIGRALP----AGGPRV-VAQN
LKTREVVT-A-PY---E--TAPH---LLSEMEAF---AKAWASRYLQPAEIPPPWM---
-----GEGATFERTRTEGGFA-----SQSV---
--EMAGLFDQAY-----RVDADWDELIRAAE
V-----AGMTDC-DVDIA-----CREA
Q--TREVAMGTL SRLRRPIAV-----
-----EALAI PERGLKARVVT KCPWALV-YLGHFLR--SWLLQGLRRDP
RI-----SAVLEGEHQRAVL DLLKSETAEVT---DSILLSADLTAATDRF--PHDL
VQACTRGILEGAR--CPLNSDWTQVFADLTGPLLV-----VPP-----
-----RGE GSDYL---TARGIMMGLPTTWCFNLNLFNFWSERAWS--LCPRVNPVI
HRASLFRPGHHAR-T--PATVICGDD-LAAVWPHDVC DQYERIAVACGASFSKGKHFRS-
--ARY-LLFTEEPYEV LVSETERRHG-----
-----RAVERP
QEWTL-----AD-----Y
LFPSR-----
-----TRSALEPSGRGSGWHYTGFRRM---KSVPLR-GLVRPTHK PFD
H-----HTIPPWAA-LPFAVQASVRA----SGDGARVRRILK-VLHPGIWAWARNN-
GFSPTSPT-ELGGYGLPPVHGRRGEGYLLP-PWHRRGISVLLRH-TKWKDLI-----
SPSTAWT--LIESGSHWRGMGE-EH----TRHKLAV-SAVACKRGK-----
-----VPFEGKGSELGT VDCVEQLTTRMARELRMLLGPE-----R
QPRDEF-----RVAPRQIAV--RV---RKMFRGLEANKGVPGGRL-----PSV
SLVPASGSVQRL-----LSEYESRLTERCYFSA-----

---PGPAREGFVLWLP-----GSARRDLAVAL---GWTRSSAHAD
D-----
>KX883602.1
-----MPRNKA
RHRWVCLYELTSLSVATLLPARPGTGPYLGGR----PNGPDTGRFLPRFPRHPNIRCPK
ERGNAATGKTTNGNVNGIL-----

-----SKVLRRRPLGNQRWTRLEEAW-----RTILA
AIAACGFVRLDSSRA-----HREALRGLSRLST----WIVKTAA-----
-L-SGTEEVRLRELKTWSARLRA-----HVVNS
LPPHKRRRGFARYFQGVLRSLIDKDS----RALQMSYIGRALP----VGSFRE-TEGA
IRKFREVTQV-PF---E--THPD---LLDSIRDW---GAHWAARFLDTPAM--PPL---
-----WRSSSGCVDFSRLQGGLA-----
--EAAVRAADFA-----AAEVESREDWDDLRLNTAAE
V-----PEMTDC-DVDIA-----VREA
Q--LRDVAL---TVAAHLPD-----LP-----
-----PV---EVCTVPERGMKCRIVTKAPWCLI-HIGHFLR--SWLFGGLRKDR
RV-----QGVLAGDHAGALRGQTERVYSQKPTEE-ENICLSADLTAATDLF--PQDL
VAALVDGLLSGAR--KPPTAEIQEVFRKLTGPLNA-RLP-----
-----DG--DGFA---IRRGIMMGFPTTWTLNLVNLFWSETAWS-----ASGSL
PFQDVALSAQPAA-S--PRTIVCGDD-LASVWPPKVADRYEEVAEACGAKFSVGKHYRS-
--RDY-ILFTEEIFKVSWEKQEYSLMDSAKP-----
-----RRTTLA
DFFPV-----PR-----

-----APRKSTMRKRYRVQAKLAPV---RCVPLR-GLVRPLHAPKD
R-----RLRPSWAA-LP-AAVEAAMD----LGAPVAVRRVLR-VLHPGIWCWARKR-
GFSPTLPL-VVGGYGLPPLRGSVRRVS-LP-KWLRYGLAAWLLN-SKWKDLE-----
GPSRCWE--GVCRPVSWQTMAR-EH----ADVRIAS-STFAVRKGR-----
-----TPPPGF---HLLGGLEAIQNLSVKFEGELCVLL-----GTEEG
SLKRQF-----GLAPRRVAN--SL--RKFFARMVS--DHPSRR-----
-PFPSSLGRTAL-----VQRWRSRAEERALYST-----
---PRSAVEGLVLSLP-----VRAKRLLAEL---GWV-----

>IABX01132835.1

---PLPKQQHRWVHLSRVSLVGAASVGARRG----APLSIFFEERASTDRQKGRNLNPE
SAADYCLAEIARVSRLAWT-----
-----PDDTGIPKGPLYKVFVSRPMPRALR-----
-----SSLVETW-----RVVLS
SLAAAGFIRLDVSKWA-----GREALHGLCDLAW----WIIRTSH-----
-L-TGIDLMFSQLKGYCSALRL-----AVVSE
IPPAERRRGFARFFRGVLRSCVDRDA----RALQISYIGRALP----KASPAQ-CTAA
LENFRAVLKR-PH---T--TSPE---ILDRA YQF---AVEWAGIHLSPEI--PNP---
-----FLGAGATLETTRAQGGFS-----EETR---
--RKAGQFDWQL-----GAPDQEMWDERRMLLNE

I-----GSLTDC-DVDIA-----VREA
M--TRDVVL---DRIHRMGD-----FP-----
-----EV---SAIALAERGRKTRVVTKAPWEIV-YLGHFLR--SWLLDGLRRDS
RV-----STVLMGDHAEAVKRLSREALYCR----GDEILSADLTAASDLL--PLDL
VRAITDGLWEGTRKGCRPEERFRLVLDRLTGPLRV-EMPDLY-----
-----DAFI--SERGIMMGLPTTWCLLNLVQLFWCEFAWM-----SGQQY
RHMHDPLRGSPRS-S--PNTAICGDD-LVAYWPSDIIGRYEQVVKESGGMFSVGKHYRS-
--GRY-AVFTEQFYDCKVTRFSAPGPALRKP-----
-----GERTLA
DFLFP-----CM-----

-----PSKGGEITPWNQVGEVRNV---PTLPLR-GLVRPSHGPGD
R-----HLMPSWAA-LPLAVEHTMMY---RPENVKALRRALR-ILHPGIWCWARSR-
GFAVGLPT-SLGGYGLPTLSGCPSRCR-LP-KWVQVGVAALLFK-VPWRDII-----
SPARYWQ--MVRQNLPHSTMAA-DH----ARYKFET-SAVACRPGK-----
-----SPFNGNGHGLGGLGAPDEVAVRLANELFMILEDL-----
PPRRPEL-----RVRPRTIAS--GV--RKFYQRLTK--RYPVGI-----R
GVKRVGKTWMDL-----QRQWDSRQNERLWWTN-----
---PRASEGFVLGLP-----RSAKRRLASAL---GWV-----

>KX883500.1

-----MEAVP
VTSR-RRCGFTKYFRGSLRSIIDRDC----RALQFSYVGRALP----EASPGQ-CDAA
LTKYREELSQ-PA---T--TPDS---ILDSAEW---ARRWASRFLSGPV---VPP---
-----T-----WLSGGACKEVPRAKGGGLG-----TRSV---
--DVARRATDAI-----CEPSEPEWE-----SLRE-----RLN-A
V-----EGLTDC-DVDIV-----CRER
L--VSNQAL---VELSEL-----PRP-----IA-----
-----V-----DAMALPERGGKVRIVTKCPWALV-YLGHFLR--VWLLEGLRRDE
RT-----KSVLDGDHAGSVGGLVGGAGVRLL---EHLAVSADLTAASDLL--PHDL
CQAIVRGVLRGSR--TRPAPHLDEVWADLIGPLTI-QM-----P-----
-----DG--SEFV---NQRGIMMGLPTTWILSLVHLYWAEWAWA-----S
TVPRHISLSGLPSAA--PRTAICGDD-LAAVWPTS SVVQRYESIVKSCGGQFSAGKHYKS-
--RTY-LMFTEEAFRLRVEMRDYDTRRVDGRRP-----
-----GEATLA
DFLPS-----

-----WMVEEPSRDSRWCTGVDHL---PLVPLR-GLVRPLHLPKD
R-----QPLPVWVA-VPHCVEAACTQ----SGDAGAVRRILR-VLHPGIHRWARSR-
GFAVNAPT-SLGGYGLPPVRGDPKRAR-LP-KWFRWGISSLLRATPWKDLI-----

1
2
3 NPARAWA--PVMRDIPHRDMGA-DHASEKLKVGAVAARKGRVPMGR-----
4 -----GADLG YDAEDRLTVRLSNELLFI-----LDRD
5 TLPSSF-----FLQGPRKVS-----GAVRKMFKQHLRRFPIPE-----G
6 SVVASETPMASL-----RRRWEDEVRTARSWSS-----
7 ---PRGAAEGFVLGIS-----SSAKRAVASAL---GWASMEPAHS
8 GGPVSD-----
9 >GFNE01046070.1
10
11 -----SDLSGVRRPQTRKVLLS
12 RITYVEATRLQAARCRSVLPSVSGRRRTPAVC-----RSRGGWVTRRRKCRPVAPASNAF
13 LSSRTGGGLSLHVADSTLSSLSQPPSISLLSSKAFSTGLAHSPTLIPLLLMQDPTWPSD
14 RPSPAAPPAKAVTTAKHADPRPPGRAPAPQPPNPHPPGPKGRVPARKVGRPPAGAGG
15 G
16 PGAGDAARKHQPPRAPFGTQREHAVVPGLVGTPTGQAFRTL CGAW-----DALRA
17 CIAAAGFVEFDLRS-----SRDAVGGLAQLSG-----WLT TATR-----
18 -G-SGLGYTIKTLKTWSAQFRS-----WVFNR
19 TPTHKRRGGLARFLRGKLRSLVDRDE-----RALQMSYLGRALP-----VGSDGV-CEKS
20 LKNHRDALTR-PH---P--TDPD---LLARAYRF---ARLWADRHLPEVPNPSHVR---
21 -----LGHGACKEATRATGGFA-----RAIV---
22 --HLAGRDLPEN-----TEARQTIMDRLRQEEDL-G
23 P-----ESITEC-DLDSFVAESELIRAV
24 VKPLLDAPR-----PP-----
25 -----RH---SAIPVAERGFKCRVVTKSPAALT-AGLHHIR--TWLMAGLRSDK
26 TS-----STTISGDHRAAVEAILAGDPVPRE---GYCFLSADLTAASDLL--PRDL
27 LIALVDGLSDADH---SAPQWYFDLLRLACGPQWV-SYP-----
28 -----WG--DTIL---TERGTLMGLPHTWCMLSLVHLFWIDQAVW-----
29 -TTPATRIPPALL-L--KRSAICGDD-LAAHWPD RVIGRYEDVATRCGAEF SVGKHYKS-
30 --RDY-VCFTESLAKARCNVGCNPPTCSGSGGRVPRYRRTRPRPAPQTPIKRPYGSASP
31 IPSPSLASAVAADPH PALLAKSLGLTPPASAPPSRSIPTPTSKPKG-----RSRPLL
32 PNLPPKGRRRGGRPVALRAPDPSPRPDPFPYPPTVPGRPDGP RWPRGPKRW PRLGAGS
33 GAKLP-----
34 -----RRRRPPRSNRRIMRRVVGVRWS---NAV PVR-GLIRPNVVPFG
35 REALGDTMGPAWAV-IGPAFWAIVGE--NPPPPRLNAARRALK-
36 ALHPHLDAWGRHNF
37 GCLATVPR-ELGGLGLPKSKGDALKVGKLP-KWLRQALAVDLYGTRAGESP-----
38 PPIGTWS--GMKPGHHRSM AEE-HSRARLQRMTIRCAGKVPFKGAP-----
39 -----TPATLLGPEVDVEELSTIAMSRS LTYI-----LGAE
40 LLGDRL-----SVNPRRLAD-----RIRSMVSRSTRWAGSR-----
41 -GICQSAPWAKL-----IARHRNLVQNWVLWLK-----
42 ---PGADEGFAGALID-----KAAKRSLAGTL---GWAIPRASAG
43 GSPARPTSQELSGP-----
44 >KX883526.1
45
46 -----
47 -----MP-----PRKSNGRKG-----
48 -----
49 -----
50 -----GRAPRTC-----VALEA
51
52
53
54
55
56
57
58
59
60

1
2
3 VLSAGAAAGLTS-----AEVRTLPIGVLAS-----
4 ---ED-NPDLGALKALSAAYRAHWIA-----
5 ----TGDP-----LAGQLSFVGRALP----PAPASA-RLAA
6 LVQHREDLTS-TF---R--TPPS---VLALAREF---ATRWARHRLARAAV--PLE---
7 -----E-----ASWPTPSSCYERSASKGGLL-----RHLL----
8 --ECTSFEPAPP-----G
9 F-----DGLPDGPVTQLW-----VQSA
10 R--FREYGLRGLERCRSAAG-----PP-----
11 -----TH---RVAVLSERGLKTRVVTVGPAWLQ-VLGHCVR--QRLLRALRSNR
12 GT-----YAPLAGARDDEIWASLEGGH-----SEAVVSTDLTRASDLL--PLDL
13 LSAIIDGLQESGK---MSTLEVEILRISAGPQRL-EYPHW-----
14 -----G---TYS---SSRGALMGLPTTWCLLSLVHLFWMDQVRL-----
15 ---TSRAGPMRA-G--HRFSVCGDD-ALLATTEAGAQRYSALVRLCGGQPSEGKHYVSG
16 GVRRCR-GVFLEKLLEFDT-----
17 -----EGGRLT
18 GMRRL-----
19 -----
20 -----AAMPVK-GVTSVSLPRDF
21 TGALPI--QCRSKGLR-LLVTLDSDVLED---LGPAGLVPVRNWVR-TRASWLGRFAREQA
22 LLTPGAPL-RQGGYAFGDPNDRTPEEE---RRFRAISLLPRPSIGAAALR-----
23 SVDPLWR--LASRETLAGEDLP-----LSSGKVVEIGRHPPLSGE-----
24 -----GPLLPTQWRQSELDRLAEG-----
25 -----YPAVAIQSQ-----EERHTSACVALFAALK-----
26 --SLMGPPLEVY-----TFRVSQLRAAVLKATA-----
27 -----TGEALRESL---GPTA-----
28 -----
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 >HAGX01098760.1
4 -----
5 -----
6 -----
7 -----
8 -----
9 -----
10 -----APPARLPL-----EF
11 LEGVRLAQPK-----SDGHLSLAASATVEYPRSKGGRAA-----
12 AYFTW-----AKARMKELGLK-AP--ETVAET-----ATHWL-----
13 -----
14 -----YSQEA AQEYYSECLGQ--AEAWLE-----
15 -----SGEPVPHRLAVVGELG-KARVITVAPTWFVQVLI
16 TMQKILMHSLRFIPEVSAGVFSRDDAYK-----AYAEIPDVSTG--DY-----
17 HFFMSDLKDATNGPPKSGIVDTVLPWLFQVGA---ESPIQRDLEIALSLL----MSER
18 KI--TVKN-DYGIPREF-ITRRGILMGDPMTKVILTIAIY-----YLIVKAVCAD-----
19 -----VEREHPGCHTWFRIK----G-----DDVVILVVGVSPEA
20 SSTEKHFLRRISDAQFRVSELDTFVSH-AAFYCEGLFL-APRGLE-----IEKISRARR
21 FG---HAYADPVKGRYMPFWK----VDPGGKDPLVLPLSKLESIVKRLSWFVPDS----
22 VQWRRTYSCAMTATKSFRLERDPSL-----FLPPKQGGIGLLPDAGIRELSTADWNQF
23 IWSWAVDLRATVLRDKIPEFSDPGVPRFVARSSNHAFSHHLRVDPRQVIPAEWEET----
24 -FLHLGDVDERLSTRLAEYLGRSPNYSSDISVASYAESVDRFSTVHLGRPL----IKRL
25 KTQGSYLGSLKVGSI---PPPTDLIEAVCQYRSNEARHRRIDLRFMRSDLIG-----
26 -----PRLVVFKPEDPEVIDLEAIKQAS
27 SPGEARTLVQAGDPFMDPDAQILRALYLADFKSVPYPIVTNDRR--LCFEARRTLSSRWPG
28 RDFVVARSRTPLLDSRSGLEMLSRHLRTLRVEGSDSHLWWDPLRTVELFPSLSVTDGER
29 F
30 LRRTSDSDWSESLMDITNVQYELVKSQ LAPLIQRGDQHWEEILAQYETVGVVFPRESSVQ
31 D
32 LELTPG-----PFQLSWRPRLREPRRLKVNHLQKIWRLRP-----
33 >GFJG01044014.1
34 -----IRWKPPIPSIGRDGPNVLLS-GDET
35 PTQKLEI--WSLTYGDVCRSMFEKRIPTALRRAGIPWTNKR-----PNTRRLANLTAD
36 QRSHFILR-----QLYP--RMSGAFKASLKKQSLG-----QIKSWFHTLDGLILP
37 LLLEGAED-----LRLIDRVTKHVMENCAQ--NYRAFLDRLKSLRKRFRKWV
38 LSH-----VNTPYKGFPGECHDLAPIL---DLWWSILQPH--QGDRRKFIQVT
39 CMCTQ-TRATGLVPARMMQESVKKFEATVSKEG---PTI-----QL
40 DGEVLMETVAACRLVNP-----ETAKLSGGPAACFEKPRSKGGQTQAILEICKQ
41 TINYHYEPEGSTLRRVETPTH-E---VNLPDD----ILDYC-----
42 -----I-EQLRMSP-----
43 -----
44 -----ESMKTRVSVVNEPS-KARVITAASVSQVYMS
45 MVAHMLQP VIRS-PRIISGLRSSRHLWNFLRKT MHPQNPEWDRLAGQGAI-----
46 FALSTDLEEATDFGNIS-VADQIWKAIVETVPE-----S-VKPYFILAHEAY----TRER
47 DL--LMPD-G----RWVKKKRGWLMGDPCTKAILSLAQ-----EYIFRVA---
48 -----TKRWGKVFMGSLVG-----DDCIML----SRY
49 KEALEDY LQEYTRCGFKISEKDTFITTAMYYCEEASL-VPSDSMVDNLT-VQIKRGMKD
50
51
52
53
54
55
56
57
58
59
60

1
2
3 ----AILYIDVPRLRLMVATR----PDQP-SRHSYSDTGRTELLGRESNWAYR-W----
4 SEKLYLPFRLGTLQALVPRSVDT---ISPFLPISMGGDGAFYA---DPQLTRDLIER
5 FAWSIGEVRIRI--RRVLSGQRNP-YYLRTSNPIGGTDKFHIIMD-MASIEELVPRECR
6 IPVDDMEHCSMLMSFKGTFFIMEPIRTAYRALQSRYWESILIEKRIPS-----DISI
7 PHRVTKTGHTEGAVPLSA--QEIADFLK-----LIEDRPMVQEWPFVLVDMR
8 KAYIMDYMHLGWKRERDPTPQEEWDLYLSGQIDLMGASRALYNFITRGCTDSGEPIEP
9 K
10 VIRDRNLFFSTDPVVKLEVRRMELK---PESHLILVSRDYR--LAFEIQKMKPETYIT
11 L-----VDPIIHLGFPWTDD-----GDKTRVIEDSGAIL-WAD
12 LNLFKDG--YPVNRRLVELVYSACMNCEPTRYGYSRTADWRRLHLNSRIGTYFAY-----
13 --RHGG-----WILQKVPRERRPSHEAE-----
14 >GBKL01052498.1
15 -----QKKMSETIRWKPLPSIGRNGPTVVVS-KAST
16 LTEKLKV---YTLETGICVSFFEKRIPRSQVLRLNISWRETE-----PESRRLSSLSPD
17 QKTAVILS-----NIWP--KMSPHFITLVRAAPVT-----AIRNWFQTL DGLLIP
18 AIL---FGLDNE-----YGLIDRITKSCLERMAQ---NYSLFLREIKEVRKRFRLLFF
19 FESD-----QRFGTLRYPGPSYGLTSIL---EFWRRILTPL--IGNQREFLRAV
20 VTIGQ-TRATGLVDGATLQKSLDKFVRTVTTPPE---ETP-----CE
21 VDEV--AVLRLGPLFAR-----ADPSRAKLSGGPASCFCFESVRSQGGQTGAIVE----
22 ICKRSINFSYGIDLVRTDIPE-HR--ICTSAD----LFSWA-----
23 -----VEELK-----
24 -----I
25 N-----PETSFAVRVAVVNEPG-KGRVVTAASIAAYLLVMG
26 VMGHIFQPTIWTER-TSSGLKKTRHLWTFLEGLHTEGLAWEKLGTTPV-----
27 LALSTDLEECTDFGHPRLSTAVMQVMAKFCPE----V-LRPLVEFATTYV---HSER
28 PI--HLPD-G----SIRTKRRGWLMGDPFTKFMLSAMQ-----EYVYQRM----
29 -----LIRWGPQFVTTVVG-----DDLVL----SSN
30 KPLLEDYLSQYKVAGFKVSILDTMISDQIMYYCEEASQ-IPQKASKALT--VSLKRGE-
31 ---EFCYLDYPRFLLIP-----THPDVDRHSYTDTRVELLNREFSWVVK-W----
32 KTGGEAFAERAVLLQKIIVPRSET---SHPFLPVWLGGDGAYHP---DPGYVFEYVRS
33 HSWNVGEYAWRL--QRYLEDTVEV-AYTRTSRSGPSVHRYHNIMA-LAEIDGLVP-EEL
34 RFPVVDEEREMWLA FKGTHTLTPPEAAFKFLKARYWLT VLEGEIPE-----ETKI
35 SKVIPLGNVAPELTR---KSISELLD-----RMTERKPRLEEWKWLVDMS
36 NAICADYMTLSWERLKPNGE-----EWKLFTEGLLD MHSAGMSEIYHAITHGDIDIPQ
37 GILSRMNVFFSSDPLIYRLRKDVPD-----EENIALISRDFK--LACKIQQMFPKRGVA
38 LVDPL-----LYVLGMIPEEIDSRFK-----IIEDPGAIIY-WTD
39 INCFQDG-----APREQWSDTVEDVYSGRLAQR
40 WHPKPGHRILNRRIGVYFAWTSKRRQGGGY-----
41 >GBKL01051391.1
42 -----FFFLKKKMSETIRWKPLPSIGRNGPTVVVS-KAST
43 LTEKLKV---YTLETGICVSFFEKRIPRSQVLRLNISWRETE-----PESRRLSSLSPD
44 QKTAVI-----LSNIWPKMSPHFITLVRAAPVT-----AIRNWFQTL DGLLIP
45 AIL---FGLDNE-----YGLIDRITKSCLERMAQ---NYSLFLREIKEVRKRFRLLFF
46 FESD-----QRFGTLRYPGPSYGLTSIL---EFWRRILTPL--IGNQREFLRAV
47 VTIGQ-TRATGLVDGATLQKSLDKFVRTVTTPPE---ETP-----CE
48 VDEV--AVLRLGPLFAR-----ADPSRAKLSGGPASCFCFESVRSQGGQTGAIVE----
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 ICKRSINFSYGIDLVRTDIPE-HR--ICTSAD----LFSWA-----
4 -----VEELK-----
5 -----I
6
7 N-----PETSFAVRVVVNEPG-KGRVVTAASIAAYLLVMG
8 VMGHIFQPTIWTER-TSSGLKKTRHLWTFLKEGLHTEGLAWEKLGTPV-----
9 LALSTDLEECTDFGHPR-LSTAVMQVMAKFCPE-----VLRPLVEFATTVY----HSER
10 PI--HLPD-G----SIRTKRRGWLMGDPFTKFMLSAMQ-----EYVYQRM---
11 -----LIRWGPQFVTTVVG-----DDLVL----SSN
12 KPLLEDYLSQYKVAGFKVSILDTMISDQIMYYCEEASQ-IPQKASKALT--VSLKRGE-
13 ---EFCYLDYPRFLLIP-----THPDVDRHSYTD TGRVELLNREFSWVVK-W----
14 KTGGEAFAERAVLLQKIIVPRSET---SHPFLPVWLGGDGAYHP---DPGYVFEYVRS
15 HSWNVGEYAWRL---QRYLEDTVEV-AYTRTSRSGPSVHRYHNIMA-LAEIDGLVP-EEL
16 RFPVVDEEREMWLA FKGTHILTPPEAAFKFLKARYWLT VLEGEIPE-----ETKI
17 SKVIPLGNVAPELTR---KSISELLD-----RMTERKPRLEEWKWLVDMS
18 NAICADYMTLSWERLKPNGE-----EWKLFTEGLLD MHSAGMSEIYHAITHGDIDIPQ
19 GILSRMNVFFSSDPLIYRLRKDVPD-----EENIALISRDFK--LACKIQQMFPKRGVA
20 LVDPL-----LYVLGMIPEEIDSRFK-----IIEDPGA IY-WTD
21 INCFQDG-----APREQWSDTVEDVYSGRLAQR
22 WHPKPGHRILNRRIGVYFAWTSKRRQGGGY-----
23 >NC_032404.1
24
25 -----MTNSCNNIPINMLLKEYNLQV-----
26 -----GPCVSIFTKRVPIWWLKR SNIRFYH-----DKVPFIKNLQPD
27
28 DRVRYLLSSIWTPRDRPLSFAFRAKL KELSTKGSIN-----TIESWWNTVQGCCLP
29 IMVGTNLWTDQI-----PNWCNRI TKWAFESCSR---NYSLFQSHIKAERKLMRQYF
30 METG-----ISTLEGYKSSRREVVTYVSVSQEVLRLNLSTV-----REREVM
31 IGLLTQTRATGLADQRMINASLEKFLSVVSHPCA---HIT-----LN
32 QKVMDSMVPRISGYIMG-----YTKLSSGPSAVFSHPRTRGGQTAAIAE---
33 LSQRYITYSYDLHTLEKTTVERTR--IDNARK----VLDYC-----
34
35 -----
36 -----IHRAIHRR-----
37
38 -----TEVLKVRVSTVVEPG-KARVITVGTLEYVQILS
39 VIAHIFKESANSQR-TRSGMKASRHLWNFLWKDLDPRNSIWETISGEGTI-----
40 YGLSTDLETATDYG NMS-MSGQIWDLLIQAATR-VP-GV-PIKLLHLGKELY----QSDR
41 PL--FSAG-K----LIATKKRGWLMGDPMTKILFTLIQ-----EVISRSV---
42 -----GRTY-RIAC-SIVG-----DDLVL----SRE
43 RSGLKEYLRLLETLDLKISWKDTCISDQYIYYCEELAR-IPQSAGVSL--RSLARKNKD
44 YI---GYIDVPKLRTLIP-----TRGESDAFSNTNLGRFSLQGKSSEYTSRVN---
45 -TGYYPVNERALLQCMFIAEDTSC----IHPHIPIEMGGGGS LVN----DPFYVQRSIS
46 LCKYQREITYRI---RSILSHNGWGFRFVHGDRPDRVTHKYHRFGKIIQDPRVFPE-GCI
47 IKPRNKAQSELLASVRSSKIIPPVDAFMRLQRMYYFRGLFS--GTGEVPDPDEYLKSLEI
48 PDGITKPSIGYTKLNP---LTLDEVSTFLSL-----WVKRGLAFRDDIPFWINRE
49 KLDINHPLVIELFKKPDMPWT-----TEDLIEESEKLSLRLA----ENPWEIPD
50 TLTPALSLVLDQDAILINILERKIKK-ITNIGR-LVVVTLDVR--LLIRMNRILRKRGKS
51 LDRIILIDPFLNLLGMAPDFNPDIYLEDQGGGITYVEDMYFQDGMVPVEVEEHMFSD-PVY
52 FKLCKPD--LPI-----YIAASVIHTK
53 RFPSPK-----YRACGANLLMS-----
54
55
56
57
58
59
60

>GEMJ01010140.1

-----ENILTDHWRPPLPLTGRFGTRKII--ENLN
DRSNSYLLKQYFLSSTCTNIFQKRIPVFWLKWNGIPFKH-----TYVPYLRDLPKY
YQSKFLIESLY---PRMSPRNFCNRLQKLCYEGKLH-----IVEQWFNTAEGCVLP
LFL---LPDLWKDVN---LKLIDRVTKWSLEWCAC---NYAHFISILKSFKKAIRKYF
ALHG-----NLDGFTHPEDRTMVPYL---DLVRSLLKDD---TLERKLQVI
LTCTQ-TRATGLSNHQMMMNSLKKFKATVTTTPPVH--IEM-----DQ
LLLSEILNKRVLRFSEP-----LSRISAGPSAVQGYSRKSGGQTKALAE----
LSTKYIDYDYDLRTLTIKEER-VQRRISSAAD-----VLNYS-----

-----IHQALTER-----
-----KRVSTVTVTLVNEPS-KARTITVATLESVTLN
VAMHIFAPTCNSQP-VRSGMRANRHLWQLSWKDLHPSNTLWDKLDHKHNDT-GY---
NPL
WALSSDLESATDYGNP-SAKQVWRKLIDILSV-FP-DA-PIGLLELCCEIH----TGPR
DV--LYEG-E-----PFCVTQRGWLMGDPMTKVILTIVQ-----EYVLKQF----
-----QQLIP-----AVG--SVVG-----DDLIL-----SYS
RAHLEMYPMSMLEALDFKVSMEDTFVSKRLMFFCEEESI-
VPQTVRDTIVCRVKRPTWKKG
YKPISCGYIDYPRIRLLIP-----VRPDQDRLSYTTLGKFSLLGREVIWAHSIN----
-EPLFPVMERAVVLQNIILPQSDT---LSPFLPIEIGNGSYCP---DPDFLKSAVEK
RSRSASQCRYRM---CQLLNHE-YGFRYLRSENRNKLHKNVVVARLVEDESVFPP-DSV
IRAESDEHKILSSL--KAYISPEEAFMKLEYQKYFTQVLS--GTGDIPS-----NFEY
QRDHS�PT---NLGGVDVEPNYSKFLKI-----WHKPGFTFKNSWKFFVNP
KLGHNDPLSVDLKFPPDLQVK-----PDPGDL
LGRTLEDFLNRVQAWDMYFPDHLER-LE-----LVVESDSSL--VALLKARLQRGETAQ
T-----LCLVSNDDKLLTRLVTIWDQLRPGECPLVVLVHPDVYLTGMYS-WQE
YTMTPI--
MVLKDPGSMFWESNNFNACQYWEVCWEVDKYQPLPGVATFTTLFYLRPGR
RHPQDD-----PVQEITPGRVFPHMGIQDVETSGADTPQEHIEPYVPPHLRK
>NC_035120.1
-----MPPLPLIGRLGPRMAVSSDLLTTLGDEKISSSVPL-EVLL
ERYFTSF-----GGCANLFYKKVPISFLRKNSVRFH-----EEVPYLSNLSRA
MQFRYLKRSVWP--QMSP--VFYQRLYRAVSAGRLT-----TIESWFNTAEAMCLP
FFL---LGVDSTE-----YSCVDHITRWLANCAN---GYASFIKTIKTFKKCLRKHL
FSGE-----PIDTFRHGSSRTCSPYL---KWAKVVLGSIPEQQTA-LRAQAV
ALLTQ-TRASGLADHRMMMQSITKFKSVTTTPYEH--IQM-----DE
PVLSRMIHPRMRRCVLN-----NAKLGTGPSAIRGFPRSSGGQNLALAK----
TCTMTIDYTYDLRSLSVTSTENRR--ISNSRD-----VLDYC-----

-----ISQALNN-----
-----RQKVTEVTLVSVNEPS-KARTITVGDYALVQLN
VAAHIFKDVCTQP-VRSGMRADRLHFNFWKDLHPQNTLWDDMGWSYETK-GM----PI
HALSSDLETATDYANPS-VGRQIWDCLISGLEIQYPESS-PRALLELCRDLH----VGPR
TV--YYQK-I-----FFCTKLRGWLMGDPMTKVILTIAQ-----EYVLFRRNAGR
-----G-PTGR-LVG--SIVG-----DDLVL-----SRL

RHHLGWYLDDLRSLDFRVSDDDTFISSDFMFYCEETSR-VPQGPQSVVA-RTKYSHGT-
----SCGYIDTPRIRLLIP-TR----PDE--DRFSNTNLGRFSLLGKEYQWCLGNN----
-SDLAPLFRRRAIGYQNCLVPQDADT----QCPFMPVEAGNGSYTT---DTAFWSKLVL
RSKDPRTTHFRV---NQLLSNEYAY-RWIRSDRPLRGLAHRYSVVRQVFEDETRPEGSV
VKAHEEEERMILRSF--KQFLSPETAFAFMKLEQARWFNGLFDGRGVIPP-----NMQY
DREILLPSVGQSGPET----LDPPVIRRFL--ER-----WRNPGFTYRDDYRFLVRRD
FLPNEDPLFIEVDPNEPDKLS-----ASVVPDDL
IGKTFEEAVNKILTADLYLDDNLWSR-LPLLIESDSMVMARWR--ELLRRDTYVETLVLI
STDRKLLLR---MIDLFTVMRPDGYIVE-----AILMHPALFLTGWGTGYM-DYD
QYNGIPG-
RVVIQDPGSIFHQDLYLPQRRSWWEGTVMLEPFRTQEREGVTTIYHGYMEVR
PELDRK-----PLPRSNASRIWRHRE-----
>GCRV01019573.1
-----RHVGCAIMQSADRSWAPPLPSLGGKVGPRI MVNPNLS
LSDKLKA---YYLELSSSPSMWIRRVPYATLRAVGYS-----GKAKHLVDLPWL
YQMKVLLKTLYF--KASR--HFVAKLESACRAGRIN-----HIKQWWITADAVVLP
YLL---ELKENDAD-----YQVVDNLTRFALESCAN---NYSGFIGRLKKVKKLIRKSL
ALDQ-----EIPFMRDMYPYIKQYKASSPSEGGIHAMAVHC
NLWCQ-TRGAGLADSKMVNKAETKFLDCIQRPK----VSI-----KL
DRQILKDTTLCCRAANG-----VHAHLSAGPKACFESTREEGGQTAFLSKL---
ARTRVVTHSYNYETLERVPLR-RR--VGSPKD-----LLDYC-----

-----IDWSL-----N
E-----PVL RKVVRLHTVAEPA-KARVITVAPFAYNRLMG
VCAHMIAPTLRSTP-LRSGMTASRHLWGFTHDTLHPGDKLWEQGI-----GC-----
EALSTDWEVATDNGNVS-VARQVWSEIIRHLSK-ID-GA-PLGFLNLCKTLY---CEDR
YV--LHKG-----KIH RKTRGWFMGDMMTKVIITLVH-----DYCARV
ALL-----KCY--SLVG-----DDFVAM-----GY
NENLQAYLQEVARLGIVISDDDTYISPRFMYYCEEMCL-VPHTT-----SQLPVVQIA
MGKSKISYIDTARLRLMLP-----CKQENQIFSATQVG RYSLLGKETRWALSTH---
-REVAPIFERGVVLQHLLLP RDQDT---QCPFTPIELGGDGSYTS---DSNFFKNVIEK
KSKNPAETHYRI---ASLMRGA-FGYRYICTERVDQVSHKYRMYVPLFQELVKYLP-EEA
VIPIHDGNRTLTLR-ARLLET PQRTILRIVKEAYYRELK--GVPVAKLP-----TLRL
EAPPPMPSTGGATPFY---SASRFLEH-----WRNPGFKFRDQDPYLVRTD
IALTLDHMSLGWDFSI RGP PK--RALEDWLSQNLSLFDGNANEVLAKL----LSNEDLPE
GAKSRLNLFESDGWVLNDR---SR-AEIPPGPIYLISNDR--LGADVSRVMEGKQDP
R-----EVYCVRPILSIIGRMD-----EVEDLVPAPALSL-EDP
GSVSFET--LVT-----LGEDFPEWAFNPLLLV
KTKFSG-----VFRIVPEIYAD-----
>NC_032457.1

-----MSG--HFQRALSKLCRSYDYR-----TIQRWWYTV DGLVLP
LLL---THPDPDSQ-----METIDRLIKFGFENCAN---NYSHFISRMKRLKKLLRKEF
AEGTD-----LTSGQLFVDLSTYRKVFY-----RYLDLTL SER-----SATECV

MAWTQ-
ARAAGLADRRMIEKSLEKAEKTLSTPRRVYLRGRWREIHAHYQGFIPVLEPEVP
RGPLLDVVRVANKVRG-----ETAHISTGPHACLEIKQAFGGQAAGIAY----
LCKRDVLTVSYPVTLERTFH-PARPCASAQD----LLSWA-----
-----ISRALDAP-----

-----VHTRMVRIVAVSEPS-KARTVTVGALAYQIILG
VVSQIFQPALASGI-SRTGLEGTRNLYEALNNDPNSGLWGPLRDMDRD--GK---FPV
FALTSDLEEATDYGDLA-VARQILQALLMRCRD-IP-GF-PLGLAVLAKSLF----LSSR
II--IRPG-YGHGLVHWFRKRNGWLMGDRMTKVVLTLAH-----EIGIL-----
-----SAGI-QFA--RICG-----DDVFAL----SKA
PAQLQRYHKVMTDLGFKISEDHFISSRILFYCEEVSL-VPQDAK-DLPS-VCNRRSE--
----PSCYVDYPRIRLLP-----IKVETNALS YTD TGRFHLLGKEMRWVWFQNS---S
QQADPFIRASLLQHIAIAMPDILS-----PFLPQELAGDGAFPH---SAEFLLSVIER
KSINYDECLYRI--HSLTHGKWGF-RYLRADNINEVVHKYHQMVPKLVLESVLP----
PDAVVRASEVLISLKVKGLETPEKTFFRLYRSYYWYKVLHGMKAPV-----
-----LNFDEDRT-----RVHGGREFKLDIP-----

-----PDVLVERFYT-TWRD-----
-----SGFTF-----

>GFUS01019776.1
-----NHVRWSPTLPSYGGRLCPRVLI-PANL
PDFASKL-EAYYLHREGSSSIFVRRVDVGWLRKHGHERLI-----AQGPFLADLPGT
VLYKYLIASLGSL-TMSA--HFQRSLSQLCLAGKMG-----AIKRWFQTTDGVVLP
LLL-SVEDPDQH-----YADIDRLTKFSLENCAN---NYSHFIGRLKRVRRLLRKDY
AEMG-----TSAKSLTDMTTYRGM-----YLKHAPKNRDQYRDVQYL
MCWSQ-TRATGLADERMIRKSLDKAEITLSTPRRPRRSIPREDWVQQSTK-LPLGEPVID
RALDSVCERIYRRC DG-----RTGHLSAGPNACIESPRNKGQGTGAISRL---
TRMKVLHSTYDFKTLEVTRHPDRR--VGSVQD----LLSWC-----

-----IQESLEHS-----
-----TYVKCVKLHAVAEPs-KARTITVCTLPYLLIVG
SLAKLLQPSIASDI-TRGGLESSRNFWNFLYQDLDPADVWHRLRDEKAED-GD---GKT
YALSSDLEEATDYGDIG-VARQILDGILRSASR-IP-GF-PVGLALLAKTLF----LGKR
FC--FRSE-NGK-W-RYFVKRSGWLMGDRLTKVILTLAH-----ELAIR-----
-----ASGL-HIA--RICG-----DDVTAF----SKS
RSRLARYLVVIREYGFKVSEDDSYISSRLAFYCEEAA--IPPVRFLDLPS-VCTRRT--
----VPCYVDYPRIRLLIP-----TSSEVKAYTYTDVGRFAQLGKEMRWILQGE---
IHPARDVYVRASMLQHLLVPQADT---LCPFTPVEVGGDGAFPH---SENFFLEVLA
KARNPTEIRFRL---HQLHRGVWGY-RLVRHQYLDNVIHKHHMLVPKIEELRSQIR-ECV
PDALIEGDELMLNSISVSGLETPERTFFRMWRAWYWSEIFRGHDPPEI-----TLEI
SRECNLPR---SEVTL---DALGFYRV-----WRGKGFTFQNRVDYLVITK
MVQPLDYMNLGWNFGVRDLPT---ENPDADQGRFLLNEENSSSFIEYI---RTGRPMDQ
VVRDSLHRMVESDSFLKWEFS---RRCHADIPQLILLVSTDVN--LGIDFQRLAATTRPD

-----YEDAPLVFLIKPTDYLYGFMSHYESWAADLTQHWDVIVDFGCVV-YNQ
FTLDSPE-----FEWFEDPDVVHNPVRRDASLRVYYLSLKEGE
GGSERC-----SLTTQPASQWRKGFFPSQYVREGLLGN-----
>GFW01018115.1
-----SIMSQKRWSPTLPSFGGRLCPRVYL-PEGY
PSDYCSRLEQFYLERSSSSSIFVKKVDVTWLKAKGHQDQLFLDGR---FRAPSLGDLEWT
VLYKYLMMSGISQ--TMSL--NFQRKLSDLCRYGKIG-----MIKRWYQTV DGLVLP
LLL-SVEDPDLH-----YAEIDKLIKFGFENCAN---NYSHFISRMKSLKRLLRKDY
ALDT-----ESSKVLSDVSTYR-----KIYLRFFPAGRGVIPA-DAIRYV
LIWSQ-TRASGLADDRMCHQSLVKCEKTLSTPRRARRIAPPEERRY SYRG-LPLGEPVVP
KIHL SRICERVARRLRG-----ETAHLSAGPNACLESTRQEGGQTGYISRV---
CRSKSLHAEYDFRTLECKQCTPRP--VRSQDD----LLSWA-----

-----IQEVLEH-----
-----PTYVRCVRLHSAEPS-KARTITVCTLPYLLIVG
VLSKLLQPAIASDI-TRAGLHASRNLWKFLYEDLDPGADLWFTLRDQFKVR-GQISNLP
LALSSDLSEATDYGD LG- VARQILAGLLHSCSE-VP-GF-PVGLGQLAKTLF----IGKR
FC--LRCE-NGR-Y-RFFVKRNGWLMGDRLTKVILTLSH-----EIAL-----
-----R-ESGL-YIA--RICG-----DDVIAF----SRS
RDDLERYHQVVTDLGFKVSEEDHYISSKLAFYCEEACI-PPTTYR-ELPQ-VAIRRK TG-
----SIYIDYPRIRLLIP-----TCVETHAHSYTDIGRFSLLGRETRWVYQNG---
APHVAALFKRASLLQHIHVPRTADT---MCPFTPLQIGGDDSF PF---SGRFLEAVVSK
KARNESETRFRM---SSLMTGKWGY-RLVRNANLDNVVHKYHTLV PKLKELRNLLP----
EHAILEGNQFLLNSIKIDGVESPERTFFRIWRAWYWHQILKGSNPPEL-----VLGV
RKEANLPQ---TKVN-----IDFEYFFHI-----WKDLGFTFENVDPDYL VSTR
DVHVLDYMNMGWWFGTPPVGK-TTPDIFDIGRFLNEENLESFINHV----KDNREL PQ
VVRSTIYKYVESDSYIKWEFS-KRDD-LDSLTRIIFLVSSDLK--LGAELARMARSRYSS
N-----LADAPVVVVVPCWVYS-----YGMMDIPIDWST
NSMSCPD--PIVIVDQGS LTFDQMVNSPEFSFFDKPRVTTRWYRSNPWLGTQCLDIVEDG
TTPSEV-----WSLTTQHPSQWRPGHFPYVPLRT-----
>GFRX01320111.1
-----FFKKKTSHMMTQNRWAPTLP SFGGRWAPRVAI-EAEY
SASYLSRLERYFLRRQSCLSMFSHRVEAHALRECFGKDILTGYMVT LKGVTPRLEQLLP
D
ELRYRLVKSFG---KISFSKNFLRKLKSLCYSKDLA-----TIK RWWQTV DGLVYP
LLL-SVEDPDLY-----YAEIDHLIKFGLENCAN---NYAHFISRMKSLKRLLRKDY
AEMK-----DSAKALTDVSTYR-----KIMMKFFPKD--EANPHLTVQYL
LCWTQ-TRASGLADRH MIRS LDKCEATLSEVRR IARIST-PETREYSYRGLPEGEPELP
QDILNDIAWRVSDKVKG-----ETAHLSAGPNACIESTRQTGGQTGYIKHLA--
RMKDSLHTRYNFCTLEPEPITPSP--VRSQND----LLSWA-----

-----I-QEALEH-----
-----PTYVRCVRLHAVPEPS-KARTITVCSLPYLLIVG
VMAKLLQPAIASDI-SRGGLHASRNLWNFLFNDLDP SLDLWG YLRDDGDPKVPI-----
MGLSSDLSEATDYGDIG- VARQIFSRVLEKCSS-VP-CF-PTALGNLAKTLF----IGRR
YC--LRYH-D-GKGYRFFVKRNGWLMGDRITKVILTLAH-----EIVVR-----

-----SSGL-FCA--RICG-----DDVIAF-----SRR
REDLERYLSEVTRLGFKVSWEDSYISSRLAFYCEEACI-PLQRFR-EIPS-VCTRGRD--
----QSMYVDYPRIRLLIP-----TRVETQAYSFTDTGRFSLLGKESRWVHQNK----
LNSQTKLFHRASLLQHLHVSQSPDT---LCPYLPIELGGDGAFPP---SHTFLMEVVRA
KARDLRETHFRM---LQLLSKWWGF-RFVRSMLNNDVVHKYHLLVPKIKELRELIP---
PESIIEGNELLLGSIKIAGVESPEKTFRIWRAWYWHSILQTGREPP-----KLEL
GVSRRHFPREETEVE-----MTEEHFHQFFHH-----WKDSGFTFRDTPDYLVDTR
YIHVKDYMNMNWGFGIRPAWVSYPRTMEQVGGRMMLNETNMESFIDHI----
KDGRPLNP
IVGRTLHKYVEADSYILSECARNWSR-
LNPPDMCVLVSGDKRLACELARIQRAYVPERD
I-----YDKVPWVMMVDPNHVLFMMMDPYEDYAEAYSSRFEVVVDFGAVA-YSA
YLTGGVD-----MFYENPRIDTKYLWNNPQLGVFEASLDEKS
ATSKELVSLTTQPRDQVLKRRIWPNFGSGLRGATMSP-----
>GFJG01122116.1
-----DRPNKLGEAPGPNHTGKRPRSDTQEMSDAPPVKRPKLAVRRLLYNVVE
SVIKTAK-RTKISPPPEQLPDVAQHTEESQYKCDISVVENNVLREICTRLAPYEYKGIKV
REFRNHPAHRRWYFMIGR--RLTDRTNQFVQSLTFR-----DYSTLEQIWRAVMHT
VLLNPDFDPSRFDPLD---IESVRSFKLWCITYASSHKNIGIVKLKGKALKTLASWMQWFM
LDNQANTCPNRVEIVPGWNRSTKRPTFPWFSGHLSHCRDFNRVKSD-----RDLHLL
AQMRTFGRALPCPPKILVDEGLKATLDVLATER-----KT
PDHVLECISSAVSYLNT---SWKISNMVRESHLSISVSASYDHTQEKGGSAATQRE----
ILSPFLDHKVDQLLDGNVPIKGNRLFIDCFGNIMLHGFSMGRSIRTCRLVDGLYATHATV
YGREEISKNLGKPMFGRRT-----GNVLLLASI-MKVQNYG-----
-----HFEPSPKGFLRTDDNQINIPYWDE-----SIIDEGLPSF
I-----EEKPFPFCRFIALAEPGFKVRPLTIGYTPMMNVLK
NMRFMMQSVLLHDPRLRVGLRSTNKLWTVLK-----LLKQKKLEYGE--GL-----
VLQSSDYKSATDYIPLD-VIRTIWQTICDGLTK-----AHPF-RVFRSLI----WHPR
QIWADIPDDQGKYTRQEQQHRCGSFMEPMSYMTLCILN-----LAVELLSQYRLRTLS
LTLKPIEYDGLNVLH-GRGI-NLDVIIIVG-----DDVLAF-----RRD
MGIIHIFHEVATQCGFIFSEGKDGYSRIGIICEDHLL-VDRREN----P-VTIR-----
-----FLDIVKSRLLTNVAR---TH---GDNRASVFGKGRMLNMVLDYLPNTM--FK
EHVWLAYRSVLERTIRLKLSDLDL-----PWFLPPSLGGLGLQLS-----GDIPDS
GWKFICFLEARM---SSAASWQDQL-ALVWSFRHL-SNRQPHGWEPEQGTVEHAVE----
-----CLKEL--KHAPKEQSLDPNLEKVIWSTDHLRRFLRISGAAAPFVGSAVQL
MHDHEYRTEAEKLG---VPLADVID-----QFERLTTFQSALE-----

--GTAEHRFKTIQQWIRRSKK-YWPKNMMNVRKPLWFTNLSDF--ERRMIRRFDFGINID
-----LCSRSVNYGPNLRIVY-----DVWSQRRRGM-PRI
GLGGNPL-SLKT-----

>GEWO01002579.1

-----YRSRKGKGFSEFRD----
RVVDEFTFTPINEIVGHMPKI-D---LIDVTGNVFLRAEDWVYGTTIRDCAYPDLLRDEG
ASAD-----TRIGLFGLL-WAVLEVL-----
----KDPDIETCTNDMAWPGLGSGISPQIF-----
L-----DEGTVRAKVVSLEPEPGYKARVITIVPLAISIIGS
VARHLLDDMVQSDMTKLGLLSKDKLYSFLRKMNGGKISGTADVSDSCKFA-----
YAESVDLTTATDTPPLS-SVRSLLQGTIATYLY---DKHAGADFCQFAASLS----TLPR
TF--VLPG-GKVT--NPNDHRSGIMMGEGLSGTFLNMSSAVRFLLRPTIDVFSDYDGT
AVDA----DDFIGC-NRDR-IQE--FLDGRERPMFEPDPTSTQSGDDVLVF---DNLAP
GMARRYLVLLYRIFGLLPSGTTFFSSEFYGTFCEEVIV-RTHDSR-GWT-----
-----YVDTTKSKLF-----QTTS-VEGPEKLLGRIGAIASSLKYSGDEA---
LLLRAAERVDAMISGHPTLSDRILR--YDLPLGLPECLGGIPHPNG-----SLSDFVAC
LDLVSLEILTGA---STVPDAVLLEIEFSRFFDDLPSDQ---IHLVEDTVRNVLE---
-----VIGGW-----ESPSFVDSTIPFQVFKRENFLPQG-----DLKY
EVWVEVAEAVSRRHGL---ISIRSAIE-----KISSMIRASIFLS-----

----GGVPPVNPLIRRRNR--RQK-LL-----TRLR--QHSQSIGAVDNSTPR
T-----VVKRLRTCENRYFR-----FQDIS-DIF
ELHGYPS--FVV-----
--PFSL-----S-----
>GEWL01001678.1
-----LSDEFVPLGVWQVRP-SRQC
GS DL-----LPSCMAFEGDVLQSWRAVEAALL-----AGASAPHCSEVH
ELLR-----TFVR--GGIRAFAL-----TLKGWGKAVRGLLS-
----LEP-----LDTHLVGYTRLPKNCVCALRREGMGESDIRDF-----
-----LDFNGKLG GIPFLF-----SEFNKVLLPG---RRLSDIQRI
ASLTTLGRSAPPATQEVLRGA VDAFVKNVTTH-----AR
TFDLFTIGWEFGQYAKG-----KSPVRSRIGLNTHSSYDYSRSLGGKFAEFRD----
RVIDDFLFWPIEEVVERKPNH-D---LIDITGNVFIRREDWSAGTTLRDCAYP LLERDEG
SSAD-----SRIGLLGLL-WAVLEVL-----
----KDPNIVAETVGMVQPGLGSRDPPQIK-----
-----LGGFVPAKVVSLEPEGFKARVITVVPLAVSIIGS
AARHLLDNMVWADDMSRIGLLSKNKLYDFMR-----VLNFGRIEGNATD--GVSCNHL
SVESVDLTTATDTPSLQ-AVGSLLSGATNGYLE---DSIAFEFSQMAVSLA----TSSR
EF--TLPG-GRKL--SYLLHESGIMMGEGLSGTFLNMSSLRFLRPEAVIAFWDYKGNS
VAEA----DVFIDL-NSGW-IQL--FLDGQGRARFEPDPHSTQSGDDVILF---NDLRP
GEARRYLVLLYRIFGLKPSSETTFYSSGTFGTFC EEVCV-KTKSSN-GWT-----
-----FVDTVKARLF-----QVTS-TEGPEKILSRISAISDALKYNRDDT---
LILRAAEAVDAMISTHPTLSDRIFR--YGLPLGLPECLGGIPHPIG-----LASDFVAC
LDEISLEILTGA---STVPDAVLLEIEFSQYFDDL PDAQ---VELIRKLIQNLLG---
-----TIDGW-----ERPSGLVTAEPFQVYTRERILPQG-----DLQY
STWVGIATAISERHGL---SSVRTVVE-----KISAQVRASVFLS-----

----GGSVPPVNPLIRRRRR--RDR-LL-----KKLR--EHTTSIGGVDLSTPK

1
2
3 T-----VVERLRSCEDRYIR-----FSDVQ-DVF
4 ELYGLPS--FVI-----
5 --PFSL-----S-----
6
7 >GADD01004427.1
8 -----SPKGPRVYPGRLGRAQMGAASAPGGSPEALI-DFDL
9 GPISWDL-----KRWLPDNPVVAL-----WSVRGKANADDQ
10 RRSFNL-----MTNM--AIRRSSLEKFRKLRHA-----ERKTLSDLTSGLLDN
11 LIA---SNPERIRTART---DPDVRKILTWAFSLGLY---NVQHAISEWKKFCTLKHA
12 LQSET---PWPEPLSDFPGFGRDACPDPCPALWARLCPWLKIIWSQG--ASSK-SDMTRI
13 CHLAT-SRNLPTASAAEREKTLSSHASVLTSTPI---PDI-----ER
14 SSMLRMLSRRIGEMVKG-----KRKGRSSGHCSLAQSACIEQPAGKGGRAFYIQS----
15 EFQTWASRKQTETISETTWFG-QP--FRLKEG----LPMWV--TMCRTEDASN-----
16 -----PTMAIGES---RQDFN-----
17 -----MNFDDEFHEDFLAGLDKATGYQLL--QWSLEEGIRRGCLLGDPF--RSSSGLAI
18 S-----KEKLPEIKADIVGEPGGKSRVMTIGEGWLTIFLQ
19 PASHHLAGLIEHHEAAWAGFKRGWQASEWAK-----RYRRKQHWGTGED-----
20 YVLSSDLKTATDYCRHD-FSLSLLLGFLEGGSGQ----G-DNPYLRAAAELL---CSPR
21 YY--TGPVEG-----SIGPTSRGILMGDPGSKIVLTLFN-----LAAEEEEAFVRWATPR
22 MIPTISVGMVTIADS-ATSW-KIRNLAVSG-----DDHTAR-----GP
23 KSYLKGITQAHIANGMVSESSNFLSPIGGFYCEEALL-FRGLDLQKLRA-YSVPFWDIP
24 YL--EHMHVDSLKLRLCLSPVMK---GDAT-RDEKNPAIGKASTFTGMLRWFAGGW---
25 ACTRPIFSARFEQRFKTFLPEGILR-----YLPRELGGVAAPCF---HLTAGMLESS
26 LTKLDPSVRSAL--NLVLAGERG--ELAKLLAKIPANTSMRGIPE--NLARDQVT----
27 -----QILETF---PGVQGRIDESQLKAMADDQMGD-QWWG-----DATA
28 RDRARFAR---RNN-L---ISVVDALN-----DIERPYLFREFLA-----
29 -----PGQHAVHE
30 GPSQFKTRIYDTPWPVRVNR--FRS-LA----LSAKPDLLD--HHHSAEGVSDGKLVG
31 M-----IQKGKGLPPPPRMYP-----ERVIFGDDL
32 VTLQTPL-----
33
34 >GFOL01002916.1
35 -----
36 -----
37 -----
38 -----AISEWKKFCTLKHA
39 LQSET---PWPEPLSDFPGFGRDACPDPCPALWARLCPWLKIIWSQG--ASSK-SDMTRI
40 CHLAT-SRNLPTASAAEREKTLSSHASVLTSTPI---PDI-----ER
41 SSMLRMLSRRIGEMVKG-----KRKGRSSGHCSLAQSACIEQPAGKGGRAFYIQS----
42 EFQTWASRKQTETISETTWFG-QP--FRLKEG----LPMWV--TMCRTEDASN-----
43 -----PTMAIGES---RQDFN-----
44 -----MNFDDEFHEDFLAGLDKATGYQLL--QWSLEEGIRRGCLLGDPF--RSSSGLAI
45 S-----KEKLPEIKADIVGEPGGKSRVMTIGEGWLTIFLQ
46 PASHHLAGLIEHHEAAWAGFKRGWQASEWAK-----RYRRKQHWGTGED-----
47 YVLSSDLKTATDYCRHD-FSLSLLLGFLEGGSGQ----G-DNPYLRAAAELL---CSPR
48 YY--TGPVEG-----SIGPTSRGILMGDPGSKIVLTLFN-----LAAEEEEAFVRWATPR
49 MIPTISVGMVTIADP-ATSW-KIRNLAVSG-----DDHTAR-----GP
50
51
52
53
54
55
56
57
58
59
60

1
2
3 KSYLKGITQAHIANGMVSESSNFLSPIGGFYCEEALL-FRGLDLQKLRA-YSVPFWDIP
4 YL--EHMHVDSLKLRLCLSPVMK---GDAT-RDEKNPAIGKASTFTGMLRWFAGGW----
5 ACTRPIFSARFEQRFKTFLPEGILR-----YLPRELGGVAAPCF---HLTAGMLESS
6 LTKLDPSVRSAL--NLVLAGERG--ELAKLLAKIPANTSMRGIPE--NLARDQVT---
7 -----QILETF---PGVQGRIDESQLKAMADDQMGD-QWWG-----DATA
8 RDRARFAR---RNN-L---ISVVDALN-----DIERPYLFREFLA-----
9 -----PGQHAVHE
10 GPSQFKTRIYDTTPWPVRVNR--FRS-LA----LSAKPDLLD--HHHSAEGVSDGKLVG
11 M-----IQKGKGLPPPPRMYIP-----ERVIFGDDL
12 VTLQTPL-----
13 -----
14
15
16 >GFOL01000185.1
17 -----
18 -----
19 --SFNL-----MTNM--AIRRSSLEKFRKLRHA-----ERKTLSDLTSGLLDN
20 LIA---SNPERIRTART---DPDVRKILTWAFSLGLY---NVQHAISEWKKFCTLKHLRA
21 LQSET---PWPEPLSDFPGFGRDACPDCPALWARLCPWLKIIWSQG--ASSK-SDMTRI
22 CHLAT-SRNLPTASAAEREKTLLSHASVLTSTPI---PDI-----ER
23 SSMLRMLSRRIGEMVKG-----KRKGRSSGHCSLAQSACIEQPAGKGGRAFYIQS----
24 EFQTWASRKQTETISETTWFG-QP--FRLKEG----LPMWV--TMCRTEDASN-----
25 -----PTMAIGES---RQDFN-----
26 -----MNFDDFRHEDFLAGLDKATGYQLL--QWSLEEGIRRGCLLGDPF--RSSSGLAI
27 S-----KEKLPEIKADIVGEPGGKSRVMTIGEGWLTIFLQ
28 PASHHLAGLIEHHEAAWAGFKRGWQASEWAK-----RYRRKQHWGTGED-----
29 YVLSSDLKTATDYCRHD-FSLSLLLGFLGSGQ----G-DNPYLRAAAELL---CSPR
30 YY--TGPVEG-----SIGPTSRGILMGDPGSKIVLTLFN-----LAAEEEEAFVRWATPR
31 MIPTISVGMVTIADP-ATSW-KIRNLAVSG-----DDHTAR-----GP
32 KSYLKGITQAHIANGMVSESSNFLSPIGGFYCEEALL-FRGLDLQKLRA-YSVPFWDIP
33 YL--EHMHVDSLKLRLCLSPVMK---GDAT-RDEKNPAIGKASTFTGMLRWFAGGW----
34 ACTRPIFSARFEQRFKTFLPEGILR-----YLPRELGGVAAPCF---HLTAGMLESS
35 LTKLDPSVRSAL--NLVLAGERG--ELAKLLAKIPANTSMRGIPE--NLARDQVT---
36 -----QILETF---PGVQGRIDESQLKAMADDQMGD-QWWG-----DATA
37 RDRARFAR---RNN-L---ISVVDALN-----DIERPYLFREFLA-----
38 -----PGQHAVHE
39 GPSQFKTRIYDTTPWPVRVNR--FRS-LA----LSAKPDLLD--HHHSAEGVSDGKLVG
40 M-----IQKGKGLPPPPRMYIP-----ERVIFGDDL
41 VTLQTPL-----
42 -----
43
44 >GFOL01000028.1
45 -----SPKGPRVYPGRLGRAQMGAAPAASAPGGSPEALI-DFDL
46 GPISWDL-----KRWLDPNPVVAL-----WSVRGKANADDQ
47 RRSFNL-----MTNM--AIRRSSLEKFRKLRHA-----ERKTLSDLTSGLLDN
48 LIA---SNPERIRTART---DPDVRKILTWAFSLGLY---NVQHAISEWKKFCTLKHLRA
49 LQSET---PWPEPLSDFPGFGRDACPDCPALWARLCPWLKIIWSQG--ASSK-SDMTRI
50 CHLAT-SRNLPTASAAEREKTLLSHASVLTSTPI---PDI-----ER
51 SSMLRMLSRRIGEMVKG-----KRKGRSSGHCSLAQSACIEQPAGKGGRAFYIQS----
52
53
54
55
56
57
58
59
60

```
1 EFQTWASRKQTETISETTWFG-QP--FRLKEG-----LPMWV--TMCRTEDASN-----
2 -----PTMAIGES---RQDFN-----
3 -----MNFDDFRHEDFLAGLDKATGYQLL--QWSLEEGIRRGCLLGDPF--RSSSGLAI
4 S-----KEKLPEIKADIVGEPGGKSRVMTIGEGWLTIFLQ
5 PASHHLAGLIEHHEAAWAGFKRGWQASEWAK-----RYRRKQHWGTGED-----
6 YVLSSDLKTATDYCRHD-FSLSLLLGFLGSGQ-----GDNPYLRAAAELL---CSPR
7 YY--TGPVEG----SIGPTRGILMGDPGSKIVLTLFN-----LAAEEEEAFVRWATPR
8 MIPTISVGMVTIADP-ATSW-KIRNLAVSG-----DDHTAR-----GP
9 KSYLKGITQAHIANGMEVSESSNFLSPIGGFYCEEALL-FRGLDLQKLRA-YSVPFWDIP
10 YL--EHMHVDSLKLRLCLSPVMK---GDAT-RDEKNPAIGKASTFTGMLRWFAGGW----
11 ACTRPIFSARFEQRFKTFLEPILR-----YLPRELGGVAAPCF---HLTAGMLESS
12 LTKLDPSVRSAL--NLVLAGERG--ELAKLLAKIPANTSMRGIPE--NLARDQVT----
13 -----QILETF---PGVQGRIDESQLKAMADDQMGD-QWWG-----DATA
14 RDRARFAR---RNN-L---ISVVDALN-----DIERPYLFREFLA-----
15 -----PGQHAVHE
16 GPSQFKTRIYDTPWPVRVNR--FRS-LA----LSAKPDLLD--HHHSAEGVSDGKLVG
17 M-----IQKGKGLPPPPRMYIP-----ERVIFGDDL
18 VTLQTPL-----
19 -----
20 -----
21 >GFOL01002977.1
22 -----PAASAPGGSPEALI-DFDL
23 GPISWDL-----KRWLPDNPVVAL-----WSVRGKANADDQ
24 RRSFNL-----MTNM--AIRRSSLEKFRKLRHA-----ERKTLSDLTSGLLDN
25 LIA--SNPERIRTART---DPDVRKILTWAFLGLY---NVQHAISEWKKFCTLKXRA
26 LQSET---PWPEPLSDFPGFGRDACPDCLPALWARLCPWLKIIWSQG--ASSK-SDMTRI
27 CHLAT-SRNLPTASAAEREKTLSSHASVLTSTPI--PDI-----ER
28 SSMLRMLSRIGEMVKG-----KRKGRSSGHCSLAQSACIEQPAGKGGRAFYIQS----
29 EFQTWASRKQTETISETTWFG-QP--FRLKEG-----LPMWV--TMCRTEDASN-----
30 -----PTMAIGES---RQDFN-----
31 -----MNFDDFRHEDFLAGLDKATGYQLL--QWSLEEGIRRGCLLGDPF--RSSSGLAI
32 S-----KEKLPEIKADIVGEPGGKSRVMTIGEGWLTIFLQ
33 PASHHLAGLIEHHEAAWAGFKRGWQASEWAK-----RYRRKQHWGTGED-----
34 YVLSSDLKTATDYCRHD-FSLSLLLGFLGSGQ-----GDNPYLRAAAELL---CSPR
35 YY--TGPVEG----SIGPTRGILMGDPGSKIVLTLFN-----LAAEEEEAFVRWATPR
36 MIPTISVGMVTIADP-ATSW-KIRNLAVSG-----DDHTAR-----GP
37 KSYLKGITQAHIANGMEVSESSNFLSPIGGFYCEEALL-FRGLDLQKLRA-YSVPFWDIP
38 YL--EHMHVDSLKLRLCLSPVMK---GDAT-RDEKNPAIGKASTFTGMLRWFAGGW----
39 ACTRPIFSARFEQRFKTFLEPILR-----YLPRELGGVAAPCF---HLTAGMLESS
40 LTKLDPSVRSAL--NLVLAGERG--ELAKLLAKIPANTSMRGIPE--NLARDQVT----
41 -----QILETF---PGVQGRIDESQLKAMADDQMGD-QWWG-----DATA
42 RDRARFAR---RNN-L---ISVVDALN-----DIERPYLFREFLA-----
43 -----PGQHAVHE
44 GPSQFKTRIYDTPWPVRVNR--FRS-LA----LSAKPDLLD--HHHSAEGVSDGKLVG
45 M-----IQKGKGLPPPPRMYIP-----ERVIFGDDL
46 VTLQTPL-----
47 -----
48 -----
49 -----
50 -----
51 -----
52 -----
53 -----
54 -----
55 -----
56 -----
57 -----
58 -----
59 -----
60 -----
```

>IABI01004424.1
-----LDPLPPCPPVAAVISTEGNGFVSRQQSEPLARS-PKDV
GPPKGDL-----SSFVPTYGGFNAGIVDRPGGLATHLAWRNPLSEKILRRSRNGKPY
GVVAYSNGVKP---SLRE--AILRRHPEICTYPDVVIRRLSEDGHRFNIRGAYLSIEDA
IVC---AVPEILSD-----DPVFPLLRWVLTNGAV-HEGVDQVIKMWKS MIVSWRWSS
LRSRT---EQPLAVPGSIFNEKGHFVGFPEESILSSVMEDIFITG--ITSK-HDGTRV
CHLCS-ARGMPPAGQKKSEDSLREHHETLTGQRDD--ISQ-----ER
LDLVKKLARRIGRKVATVL-SKEKGDFRSDGHVSATTSASVEDTVKSGGRANTLAT----
QFIPWAKKVATEPLEGNDILG-NA--VRERPG-----HARWR--TTCREVPVPFERDCDC
AGEDGVRPNSSNGLPAVHSLEANNREECPNLGLFSSI-WRMPEFD-----
-----MPTNLDNPLTGVD SIGKQML--QWAHHQGCSAGYIEGPLVGSREEFNTSK
V-----VKPIPVVR-SVVREPGGKARVVSKNEAWFTTFLQ
PFGHHSVGLLSRLPSTRAGLSAAAQGYEWAK-----SASKSTGIDGAS-GM-----
LFLSSDLKTATDFCDHR-VSKSLLEGFYEGVGT----NPDNPYIRLCIEIL----CSPR
IL-----E-----DGSITSRGIMMGEPGCKAVLTLFN-----LVAEEEAFTAHCQNY
PEIL-----DLAVVPLRKSW-RVP--ACAG-----DDHIAY-----GP
ETYLRRITRNHHSNGMQVSWPSN FVSPIAAIYCEEMILKLPHLRF-----EVGYLPKAD
YR--ITVHIDAMKVRLLSPYRKETSGGPND-SDKTNPAVGKSSSFTKKLDWQPKGW----
ELARKPFLERFRSRMMQYLPHEAG-----YQYLPYRLGGLGVPSV-----IPSSLRTL
SADLSPRHVAAI---EAIMDGRGTP-AIRRCLSRFASNDEARGIDLSNTAAEDENI----
-----ALMKRA-----IQGFDTP EEGVLTFEAVREKLGLSFDQAE-----NMRR
RDFNRKAG---RELQL---ISITDALR-----EIERPYITVELLK-----

GKASEPKSRFDVRPWKVRYAK-LQDD-LETCG--VPKNPNNVL--VASTLDEVFNA----
-----VITGNDEYIPKERISR-----VISA FGVEV-VR-

>GEWH01000774.1
-----KFNL DSTPCARVAAALDDARYGEAQATRY SIPDPDAT
GVELVTR---IPWRNPLTEKSEVSFGGKRRQRRKPL-----KRPSYHSMIKPS
QRFRLS-----VTPIIKELSRRQLNWLSEDGD-----RFNRVRGCFLGIEDA
IYS---SCPELRDD-----DPIWSKLLKWT FATGSSRHSGVDTV IKS WKSFATWVKWAA
LRSET---EEPAVIRGFPFF-DENGFTGFPADSQFSELIEPILRTG--ITSK-GEGTRI
CHLIT-SRGLPPAGQSKADASLDEHFKVLTSDSG---ISD-----ER
VEEVRVIAKRIGKLVRGCI VKEGEDRFRSDGHVSATSSACVESTVKQGGRAASMSA----
NLLTFLNEIPKTETKGVDILN-NP--FIETGD-----LPRWR--TTCRDEVNELSRDCDC
FET-----CLNKWECKNNTWSSSY-WSIPDFE-----
-----SDLENMYDNPLMGVDHAIGPQMV--QWAYLEGRYRGVFTGPRVGD KQQDFLTC
T-----VSRPPQIIRSVVAEPGGKARVVTKGEDWLT VFLQ
PFSHTIGLLRRLPSARAGLSAAAMGYE WTK-----ALSRKEGLDKED-NL-----
YFLSSDLKTATDYCDHR-ISKALLEGFYSGIGW----YKDHSYIKLAIDLL----CSSR
EC--E---G-----EFT--KRGILMGDPGTKAVLTLFN-----LVAEEESFQDWK---
-----RKNPER-EVKW-RCP--AFAG-----DDHVAL-----GP
KEYVSQITVNHARNGMKVGWESNFVSKIGAFYCEELLAHVNGTKF-----EVGYLPKAE
YE--STIHVDAMKIRLMSPYRKETSAAPND-SDKANPAIGKASSFYKKLSWFQDGW----
KSACAPLLQRFRLRMMDYLP AENQV-----YQYLPYYLGGLDMPLL-EGNPESVSDWTEL

1
2
3 IPQLDLAHQSAI---QMVIDGKADS-WLLRCLSRFASNDYARGIDLSNRYAEDSIR----
4 -----GLLQIKKISTQETESYTYEQVVEELVKTGISK-TSIE-----DMRK
5 RDFGRKAQ---DQLKL---ISMSSALN-----LIEKPYIVTNLLR-----
6 -----G
7 EGVDSQKQKYDLVPWKTRYNR-FIKD-LESVG--VSAEGYEYK--TGMTKDDIFNR----
8 -----FVSGKDIYIPERRITS-----ILTAYKVGW-WST
9 PSLLTET--KLS-----
10 -----
11 -----
12 >GEWH01001802.1
13 -----
14 -----
15 -----
16 -----
17 -----
18 -----PILRTG--ITSK-GEGTRI
19 CHLIT-SRGLPPAGQSKADASLDEHFKVLTSDSG---ISD-----ER
20 VEEVRVIAKRIGKLVKGCIVKEGEDRFRSDGHVSATSSACVESTVKQGGRAASMSA----
21 NLLTFLNEIPKTETKGV DILN-NP--FIETGD----LPRWR--TTCRDEVNELSRDCDC
22 FET-----CLNKWECKNNTWSSSY-WSIPDFE-----
23 -----SDLENMYDNPLMGVDHAIGPQMV--QWAYLEGRYRGVFTGPRVGDQKQDFTLC
24 T-----VSRPPQIIRSVVAEPGGKARVVTCKGEDWLT VFLQ
25 PFSHHTIGLLRRLPSARAGLSAAAMGYEWT-----ALSRKEGLDKED-NL-----
26 YFLSSDLKTATDYCDHR-ISKALLEGFYSGIGW----YKDHSYIKLAIDL----CSSR
27 EC-----EGFTKRGILMGDPGT KAVLT LFN-----LVAEEESFQDWK---
28 -----RKNPER-EVKW-RCP--AFAG-----DDHVAL-----GP
29 KEYVSQITVNHARNGMKVWESNFVSKIGAFYCEELLA-HVNGT--KFEV-GYLPKA--E
30 YE--STIHVDAMKIRLMSPYRKETS AAPND-SDKANPAIGKASSFYKKLSWFQDGW----
31 KSACAPLLQRFRLRMMDYLP AENQV----YQYLPYYLGGLDMPLL-EGNPESVSDWTEL
32 IPQLDLAHQSAI---QMVIDGKADS-WLLRCLSRFASNDYARGIDLSNRYAEDSIR----
33 -----GLLQIKKISTQETESYTYEQVVEELVKTGISK-TSIE-----DMRK
34 RDFGRKAQ---DQLKL---ISMSSALN-----LIEKPYIVTNLLR-----
35 -----G
36 EGVDSQKQKYDLVPWKTRYNR-FIKD-LESVG--VSAEGYEYK--TGMTKDDIFNR----
37 -----FVSGKDIYIPERRITS-----ILTAYKVGW-WST
38 PSLLTET--KLS-----
39 -----
40 -----
41 >KX373304.1
42 -----MPTIRNELNPRVSF-----
43 -----FAEKHGKGTFCPLSLR-----HYLGDSRHL
44 RKKCLLHP-----RAGLLKEVLGKSWKTFKCFKTS-----SIVRAHELYLCYLDI
45 CVC---LDPEILS-----IESNRSIGVKVFNTACF---STSTATSLYKEFVENVLILL
46 KCKE-----GEERWTRQNPFRMMTW---KIFQDAIASV--QLSKGKSLERV
47 LSVLT-TRNFPEPDGKSIQKKKVDFIGIISQPPKGWKSE-----EY
48 RSHFLESIEEIANECES-----TTVSDCHISVTAAGSLKKT VREGGKFAEMIE----
49 EVKGFLAETPSDEELYEFANL-K---WTCNTN-----EPRWK--TFGIIGEINPMS---
50 -----AMTLFTEGV-DFLNEVPNQYLGPIDFSGGSLP
51 PGIPEEDL FITDLRPF RMGLGAQFGNQLL--LYCC-----VFY
52
53
54
55
56
57
58
59
60

K-----KDELP MIRASPVLEGGDKVRWITMASWRDLVIQQ
AAATIFRSLMESHKEMKPIFSRANLAWVYLN-----KAREIGPSD-----
ICYVSDYSSATDTVDRE-FAEFILTNFIKRFRN--RLSDPLLNFLELGIRNA---VSPK
VV--IFPT-G----ERITSSRGVFMGEPMSKVILTLM-----FTIGKAAKSIYKLRF
-----PKSMEK-LTFW--AP----G-----DDL VAT-----GP
TEYIDIYSELAKILGQILNHSKVFKSRTVFKLCEQWFW-VPGLKS-SVGT-WAITTDPGK
YR--ESAWVDTVKLKLLGAMSL----SNHSHFEERNEVIGKAKALSKILRWLPST--YP
LEYKKLLRTWFLCRFEPKLPRTDSK--TFAFLMLPGHLGGFDLLD-----DQEIREA
YQKVSSFSRILF---NNYYDDFIVG---SYLRNMLRNRSYRGFAL---SEDYNL----
-----TVEKLV-----KYYIDLLEEAPYEEYKD-SKVE-----GLSF
RETDRALF--KAG-F---LTENRIRD-----IVSRPLAFGSIWG-----

--RTLKHSPYNTTPISVRLAK-LWKN-LT-----GDTERWK--NIS-LKGEISED FV
V-----LCKENKGPVLKVYR-----RND SL-EDE
VLIGLPS--MKV-----SLGIL---
--PHDG-----YISW-----
>KU935604.1
-----MPTIRNELNPRVSF-----
-----FAEKHGKGTFCPLSLR-----HYLGDSRHL
RKKCLLHP-----RAGLLKEVLGKSWKTFKCFKTS-----SIVRAHELYLCYLDI
CVC---LDPEILS-----IESNRSIGVKVFNTACF---STSTATSLYKEFVENVLILL
KCKE-----GEERWTRQNPFKRMMTW---KIFQDAIASV--QLSKGKSLERV
LSVLT-TRNFPEPDGKSIQKKKVDFIGIISQPPKGWKSE-----EY
RSHFLESIEEIANECES-----TTVSDCHISVTAAGSLKKTVREGGKFAEMIE----
EVKGFLAETPSNEELYEFANL-K---WTCNTN-----EPRWK--TFGIIGEINPMS---
-----AMTLFTEGV-DFLNEVPNQYLGPIDFS GGS LP
PGIPEEDLFITDLRPFRMGLGAQFGNQLL--LYCC-----VFY
K-----KDELP MIRASPVLEGGDKVRWITMASWRDLVIQQ
AAATIFRSLMESHKEMKPIFSRANLAWVYLN-----KAREIGPSD-----
ICYVSDYSSATDTVDRE-FAEFILTNFIKRFRN--RLSDPLLNFLELGIRNA---VSPK
VV--IFPT-G----ERITSSRGVFMGEPMSKVILTLM-----FTIGKAAKSIYKLRF
-----PKSMEK-LTFW--AP----G-----DDL VAT-----GP
TEYIDIYSELAKILGQILNHSKVFKSRTVFKLCEQWFW-VPGLKS-SVGT-WAITTDPGK
YR--ESAWVDTVKLKLLGAMSL----SNHSHFEERNEVIGKAKALSKILRWLPST--YP
LEYKKLLRTWFLCRFEPRLPRTDSK--TFAFLMLPGHLGGFDLLD-----DQEIREA
YQKVSSFSRILF---NNYYDDFIVG---SYLRNMLRNRSYRGFAL---SEDYNL----
-----TVEKLV-----KYYTDLLEEAPYEEYKD-SKVE-----GLSF
RETDRALF--KAG-F---LTENRIRD-----IVSRPLAFGSIWG-----

--RTLKHSPYNTTPISVRLAK-LWKN-LT-----GDTERWK--NIS-LKGEISED FV
V-----LCKENKGPVLKVYR-----RND SL-EDE
VLIGLPS--MKV-----SLGIL---
--PHDG-----YISW-----
>KY628364.1
-----MPTIRNELNPRVSF-----
-----FAEKHGKGTFCPLSLR-----HYLGDSRHL

RKKCLLHP-----RAGLLKEVLGKSWKTFKCFKTS-----SIVRAHELYLCYLDI
CVC---LDPEILS-----IESNRSIGVKVFNTACF---STSTATSLYKEFVENVLILL
KCKE-----GEERWTRQNPFRMMTW---KIFQDAIASV--QLSKGKSLERV
LSVLT-TRNFPEPDGKSIQKKKVDFIGIISQQPPKGWKSE-----EY
RSHFLESIEEIANECES-----TTVSDCHISVTAAGSLKKTVREGGKFAEMIE----
EVKGFLAETPSNEELYEFANL-K---WTCNTN-----EPRWK--TFGIIGEINPMS----
-----AMTLFTEGV-DFLNEVPNQYLGPIDFS GGSLP
PGIIP EEDLFITDLRPFRMGLGAQFGNQLL--LYCC-----VFY
K-----KDELPMIRASPVLEGGDKVRWITMASWRDLVIQQ
AAATIFRSLMESHKEMKPIFSRANLAWVYLN-----KAREIGP-SD-----
ICYVSDYSSATDTVDRE-FAEFILTNFIKRFRN--RLSDPLLNFLELGIRNA----VSPK
VV--IFPT-G----ERITSSRGVFMGEPMSKVILTLM-----FTIGKAAKSIYKLRF
-----PKSMEK-LTFW--AP----G-----DDL VAT-----GP
TEYIDIYSELAKILGQILNHSKVFSRTVFKLCEQWFW-VPGLKS-SVGT-WAITTDPGK
YR--ESAWVDTVKLKLLGAMSL---SNHSHFEERNEVIGKAKALSKILRWLPSDT--YP
LEYKKLLRTWFLCRFEPKLPRTDSK--TFAFLMLPGHLGGFDLLD-----DQEIREA
YQKVSSFSRILF---NNYYDDFIVG---SYLRNMLRNRSYRGFAL---SEDYNL---
-----TVEKLV-----KYYTDLLEEAPYEEYKD-SKVE-----GLSF
RETDRALF---KAG-F---LTENRIRD-----IVSRPLAFGSIWG-----

--RTLKHSPYNTTPISVRLAK-LWKN-----LTGDTERWK--NIS-LKGEISED FV
V-----LCKENKGPVLKVYR-----RND SL-EDE
VLIGLPS--MKV-----SLGIL---
--PHDG-----YISW-----
>JN400242.1

-----RSTLPQSWE-----RGDRYHSDLLKR
DLMIRL-----TRDL--GLSGRHADRLVNRPAI-----HLKRIEEFISGLVDS
FWL---GDPRTFALGSPD--FKTFRTIVRKIFAVGTS---NLGMLMANWK EWTNWFTHSV
CETG-----LREEAKLSNHNMFHRLARL---SVVQKALSGG--VPLM-ELGSIV
AHLTS-TRQMPYMGLPTEVKAKADFI EISTPFH---VEP-----SH
HRQMVQCAARMGRLCMS-L-RNGVPVSDRSSHFSATSSGELD HSLTRGGQAQALKD----
AIDRWLIPASGSTYMEDTPFG-V---AEHREG-----VPLWK--TLFVDQETQLEL----
-----IFSDFGDSL-DWIKDVP-----
-----GRVYGLDDYTGRQILYVAWKE-----
-----MEDIPHIRASTVPELGNKARIVTLSAFWLNVLQA
PLSHIMKEVLKYHPSCFASFTRGEQAWHAAS-----GLGRLNPRSA--GY-----
SVLSSDLKNATNAQHIA-LTRDMLRAFIYSSGL---VV-SKAYVNLVLD T-----ICPR
LV--EL-D-G----EFTLSCRGIMMGEAIAKPSMTLLN-----LVVEGLAFLKYENKL
VLLN----TDKAAP-SRRW-RCY--HVGG-----DDHLAV-----GP
DQYLDLITHNHELSGSIISPDKHA KS RKMVKYCERVLL-VQNLQ-----YNTPD RK-P
FE--HGLIVDSIKVRLLEKGQS---TLIA-KDNKNVAVGKSQQLVKSLDWLPKQL--YS
KGFIRSIQH LFIKRMGSLLPNRNRDEEAFHSVCLPKILGGYGLGLA-----EDLLTH
LSKSKPEIQQT V---SAVLYGIETR-RARNILSRLNTTISDRSIQG---TKWYED----
-----EMVQQF-----REYPDMVG AIDGREMRR--RFP-----AST-

1
2
3
4
5
6
7
8
9 >JN400243.1
10
11 -----RSTLPQSWE-----RGDRYHSDLLKR
12 DLMIRL-----TRDL--GLSGRHADRLVNRPAI-----HLKRIEEFISGLVDS
13 FWL---GDPRTFALGSPD--FKTFRTIVRKIFAVGTS---NLGMLMANWKEWTNWFTHSV
14 CETE-----LREEAKLSNHNMFHRLARL---SVVQKALSGG--VPLM-ELGSIV
15 AHLTS-TRQMPYMGLPTEVKAKADFIEIISTPFH---VEP-----SH
16 HRQMVQCAARMGRLCMS--LRNGVPVSDRSSHFSATSSGELDHSLTRGGQAQALKD----
17 AIDRWLIPASGSTYMEDTPFG-V---AEHREG----VPLWK--TLFVDQETQLEL----
18 -----IFSDFGDSL-DWIKDVP-----
19 -----GRVYGLDDYTGRQILYVAWKE-----
20
21 -----MEDIPHIRASTVPELGNKARIVTLSAFWLNVLQA
22 PLSHIMKEVLKYHPSCFASFTRGDQAWHAAS-----GLGRLNPRVA--GY-----
23 SVLSSDLKNATNAQHIA-LTRDMLRAFIYSSGL----VV-SKAYVNLVLDL----ICPR
24 LV--EL-D-G----EFTLSCRGIMMGEAIAKPSMTLLN-----LVVEELAFKLYENKL
25 VLLN----TDKAAP-SRRW-RCY--HVGG-----DDHLAV-----GP
26 DQYLDLITHNELSGSIISPDKHAKSRKMVKYCERVLL-VQNLQ-----YNTPDRLK-P
27 FE--HGLIVDSIKVRLLEKGQS---TLIA-KDNKNVAVGKSQQLVKSLDWLPKQL--YS
28 KGFIRSIQHLFIKRMGSLLPNRNRDEEAFHSVCLPKILGGYGLGLA-----EDLLTH
29 LSKSKPEIQQTV---SAVLYGIETR-RARNILSRLNTTISDRSIQG---TKWYED----
30 -----EMVQQF-----REYPDVMVGAI DGREMRR--RFP-----AST-
31
32
33
34
35
36
37
38
39
40 >JN400241.1
41 -----MQEDLILVQPD TDWVGGVNRSWRSLL-PVLN
42 GVEYI-----LPRGSLVGLKLHGRSTLPQSWE-----RGDRYHSDLLKR
43 DLMIRL-----TRDL--GLSGRHADRLVNRPAI-----HLKRIEEFISGLVDS
44 FWL---GDPRTFALGSPD--FKTFRTIVRKIFAVGTS---NLGMLMTNWKWTNWFTHSV
45 CETE-----LREEAKLSNHNMFHRLARL---SVVQKALSGD--VPLM-ELGSIV
46 AHLTS-TRQMPYMGLPTEVKAKADFIEIISTPFH---VEP-----SH
47 HRQMVQCAARMGRLCMS-L-RNGVPVSDRSSHFSATSSGELDHSLTRGGQAQALKD----
48 AIDKWLIPASGSTYMEDTPFG-V---AEHREG----VPLWK--TLFVDQETQLEL----
49 -----IFSDFGDSL-DWIKDVP-----
50 -----GRVYGLDDYTGRQILYVAWKE-----
51 -----MEDIPHIRASTVPELGNKARIVTLSAFWLNVLQA
52 PLSHIMKEVLKYHPSCFASFTRGDQAWHAAS-----GLGRLNPRVA--GY-----
53 SVLSSDLKNATNAQHIA-LTRDMLRAFIYSSGL----VV-SEAYVNLVLDL----ICPR
54
55
56
57
58
59
60

LV--EL-D-G----EFTLSCRGIMMGEAIAKPSLTLLN-----LVVEELAFCLKYENKL
VLLN-----TDKAAP-SRRW-RCY--HVGG-----DDHLAV-----GP
DQYLDLITHNHELSGSIISPDKHKGSRKMKVYKVCERVLL-VQNLQ-----YNTDPDRK-P
FE--HGLIVDSIKVRLLEKQGS---TLIA-KDNKNVAVGKSQQLVKSLDWLPKQL--YS
KGFIRSIQHLEFIKRMGSLLPNRNRDEEAFHSVCLPKILGGYGLGLA-----EDLLTH
LSKSKPEIQQTV---SAVLYGIETR-RARNILSRLNTTISDRSIQG---TKWYED----
-----EMVQQF-----REYPDMVGAI DGREMRR--RFP-----ASTY
QESKAHAA---SAG-W---LPIDQLAR-----RVTRGHLFGELMS-----

-GGAQRQNVFKTRNWIQEFTR-LRKR-LEEEVGYLPLPEYADL--EVKTLRNLGAMTTD
L-----FIDSRQVTSFDKGVGD-----DSFDFFDGEL-LRT
YEAHQPN--LVV-----GLQFIGLD---
--STCN-----AFTRRVRRRY-----
>KU295726.1
-----MQEDLILVQPD TDWVGEVNRSWRSLL-PVLN
GVEYI-----LPRGSLVGLKLHGRSTLPQSWE-----RGDRYHSDLLKR
DLMIRL-----TRDL--GLSGRHADRLVNRPAI-----HLKRIEFISGLVDS
FWL---GDPRTFVLGSPD--FKTFRTIVRKIFAVGTS---NLGMLMTNWKWETNWFTHSV
CETE-----LREEAKLSNHNVFHRLARL---SVVQKALSGG--VPLM-ELGSIV
AHLTS-TRQMPYMG LPT EVKAKADFIETSTPFH--VEP-----SH
HRQMVQCAARMGRLCMS-L-RNGVPVSDRSSHFSATSSGELDHSLTRGGQAQALKD----
AIDRWLIPASGSTYMEDTPFG-V---AEHREG----VPLWK--TLFVDQETQLEL----
-----IFSDFGDSL-DWIKDVP-----
-----GRVYGLDDYTGRQILYVAWKE-----
-----MEDIPHIRASTVPELG NKARIVTLSA FWLNVLQA
PLSHIMKEVLKYHPSCFASFTRGDQAWHAAS-----GLGRLNPRSVA--GY-----
SVLSSDLKNATNAQHIA-LTRDMLRAFIYSSGL---VV-SEAYVNLVLD T-----ICPR
LV--EL-D-G----EFTLSCRGIMMGEAIAKPSLTLLN-----LVVEELAFCLKYENKL
VLLN-----TDKAAP-SRRW-RCY--HVGG-----DDHLAV-----GP
DQYLDLITHNHELSGSIISPDKHKGSRKMKVYKVCERVLL-VQNLQ-----YNTDPDRK-P
FE--HGLIVDSIKVRLLEKQGS---TLIA-KDNKNVAVGKSQQLVKSLDWLPKQL--YS
KGFIRSIQHLEFIKRMGSLLPNRNRDEEAFHSVCLPKILGGYGLGLA-----EDLLTH
LSKSKPEIQQTV---SAVLYGIETR-RARNILSRLNTTISDRSIQG---TKWYED----
-----EMVQQF-----REYPDMVGAI DGREMRR--RFP-----ASTY
QESKAHAA---SAG-W---LPIDQLAR-----RVTRGHLFGELMS-----

-GGAQRQNVFKTRNWIQEFTR-LRKR-LEEEVGYLPLPEYADL--EVKTLRNLGAMTTD
L-----FIDSRQVTSFDKGVGD-----DSFDFFDGEL-LRT
YEAHQPN--LVV-----GLQFIGLD---
--STSN-----AFTRRVGRY-----
>KU295722.1
-----MQEDLILVQPD TDWVGGVNRSWRSLL-PVLN
GVEYI-----LPRGSLVGLKLHGRSTLPQSWE-----RGDRYHSDLLKR
DLMIRL-----TRDL--GLSGRHADRLVNRPAI-----HLKRIEFISGLVDS
FWL---GDPRTFVLGSPD--FKTFRTIVRKIFAVGTS---NLGMLMTNWKWETNWFTHSV
CETE-----LREEAKLSNHNMFHRLARL---SVVQKALSGG--VPLM-ELGSIV

AHLTS-TRQMPYMGLPTEVKAKADFIIEISTPFH---VEP-----SH
HRQMVQCAARMGRLCMS-L-RNGVPVSDRSSHFSATSSGELDHSLTRGGQAQALKD----
AIDRWLIPASGSTYMEDTPFG-V---AEHREG-----VPLWK--TLFVDQETQLEL----
-----IFSDFGDSL-DWIKDVP-----
-----GRVYGLDDYTGRQILYVAWKE-----
-----MEDIPHIRASTVPELGNKARIVTLSAFWLNVLQA
PLSHIMKEVLKYHPSCFASFTRGDQAWHAAS-----GLGRLNPRSA--GY-----
SVLSSDLKNATNAQHIA-LTRDMLRAFIYSSGL----VV-SEAYVNLVLDT-----ICPR
LV--EL-D-G----EFTLSCRGIMMGEAIAKPSLTLLN-----LVVEELAFKLYENKL
VLLN----TDKAAP-SRRW-RCY--HVGG-----DDHLAV-----GP
DQYLDLITHNHELSGSIISPDKHGKSRRKMVKYCERVLL-VQNLQ-----YNTPDPRK-P
FE--HGLIVDSIKVRLLEKGQS---TLIA-KDNKNVAVGKSQQLVKSLDWLPKQL--YS
KGFIRSIQHLFIKRMGSLLPNRNRDEEAFHSVCLPKILGGYGLGLA-----EDLLTH
LSKSKPEIQQTV---SAVLYGIETR-RARNILSRLNTTISDRSIQG---TKWYED---
-----EMVQQF-----REYPDPMVGAI DGREMRR--RFP-----ASTY
QESKAHAA---SAG-W---LPIDQLAR-----RVTRGHLFGELMS-----

-GGAKRQNVFKTRNWIQEFTR-LRKR-LEEEVGYLPLPEYADL--EVKTLRNLGAMTTD
L-----FIDSRQVTSFDKGVGD-----DSFDFFDGEL-LRT
YEAHQPN--LVV-----GLQFIGLD---
--STCN-----AFTRVRGRY-----
>KU295727.1
-----MQEDLILVQPD TDWVGGVNRSWRSLL-PVLN
GVEYI-----LPRGSLVGLKLHGRSTLPQSWE-----RGDRYHSDLLKR
DLMIRL-----TRDL--GLSGRHADRLVNRPAI-----HLKRIEFISGLVDS
FWL---GDPRTFVLGSPD--FKTFRTIVRKIFAVGTS---NLGMLMANWKEWTNWFTHSV
CETE-----LREEAKLSNHNMFHRLARL---SVVQKALSGG--VPLM-ELGSIV
AHLTS-TRQMPYMGLPTEVKAKADFIIEISTPFH---VEP-----SH
HRQMVQCAARMGRLCMS-L-RNGVPVSDRSSHFSATSSGELDHSLTRGGQAQALKD----
AIDRWLIPASGSTYMEDTPFG-V---AEHREG-----VPLWK--TLFVDQETQLEL----
-----IFSDFGDSL-DWIKDVP-----
-----GRVYGLDDYTGRQILYVAWKE-----
-----MEDIPHIRASTVPELGNKARIVTLSAFWLNVLQA
PLSHIMKEVLKYHPSCFASFTRGDQAWHAAS-----GLGRLNPRSA--GY-----
SVLSSDLKNATNAQHIA-LTRDMLRAFIYSSGL----VV-SEAYVNLVLDT-----ICPR
LV--EL-D-G----EFTLSCRGIMMGEAIAKPSLTLLN-----LVVEELAFKLYENKL
VLLN----TDKAAP-SRRW-RCY--HVGG-----DDHLAV-----GP
DQYLDLITHNHELSGSIISPDKHGKSRRKMVKYCERVLL-VQNLQ-----YNTPDPRK-P
FE--HGLIVDSIKVRLLEKGQS---TLIA-KDNKNVAVGKSQQLVKSLDWLPKQL--YS
KGFIRSIQHLFIKRMGSLLPNRNRDEEAFHSVCLPKILGGYGLGLA-----EDLLTH
LSKSKPEIQQTV---SAVLYGIETR-RARNILSRLNTTISDRSIQG---TKWYED---
-----EMVQQF-----REYPDPMVGAI DGREMRR--RFP-----ASTY
QESKAHAA---SAG-W---LPIDQLAR-----RVTRGHLFGELMS-----

-GGAKRQNVFKTRNWIQEFTR-LRKR-LEEEVGYLPLPEYADL--EVKTLRNLGAMTTD
L-----FIDSRQVTSFDKGVGD-----DSFDFFDGEL-LRT

1
2
3 YEAHQPN--LVV-----GLQFIGLD---
4 --STCN-----AFTRRVGRY-----
5 >H AFC01092190.1
6
7 -----MRELSYKLIRVTIPRGATGKSVSEAYRAIL-PCLL
8 PTDQEDV----YNYTLPLGIFKLKFKGQSPLPQLWEK-----ADSVYLSEREKK
9 IFRFHL-----KTGL--GLGKRVTEVFLSRPRS-----HYLRTFGFIDGLVDA
10 FWA---ADPKFFLVGSKD--YKIFKSIVRKIIKVGCT---DLGRLVDDWKDWGTRIFHQV
11 ARTKT-----IGSLPEAKPENIFRILSEL---SWIKSILAKDVKDLDLDS-LDGNLL
12 ASIRS-SRQFPYMGAKTEKKSLERFKEVVSTPFI---VDP-----ET
13 SWQMVATARRIGGICRA---IRTRPIPDMAAHFSVSSSGEYTYTINKGAFSQAVFD----
14 GIVPYLMRI-IDTSSEETPFG-L---VHYQSG----KEVWR--YLFRDEEISQ-----
15 -----GEFL-DNRKGF-----
16 -----KCQLPRLGLDETFGKQVLWAAWKA-----SLE
17 F-----QGKPIPVRAEIVPEMGNKARVVTVCWWLQILQA
18 PLAHACIAMMKWHPVSVYSSFHKMDQAWCASV-----GLHKVKDPDLT--KY-----
19 WLLSSDLKDATNAQSID-LTKAMLKAFLGSLN----LSGREAYVNYCLELLGSTRLGSR
20 EV--HFKD-K----TVVLSKSGIMMGEPLAKVGLTLLN----LSVEELALLTHIGRL
21 DLLL----KPDPTP-KLKW-RFC--HIGG-----DDHLAW-----GP
22 EPILEKITEHYKVGSHLSPGEHGSRIFFRYTEKVLN-VMNMGKSPMDA-QVNKIDAF-
23 ---KRIMVDGIKIRLLESGQS---TMIS-KDNKNVAIGKAGQLVRSIQWLEPKI-GFT
24 ESRIDSIRDLFCLRLNRFLPSQHRNWRCYQATKLPRVLGGFGLGIT-----PSEQYQA
25 YVDSPSPIRWVV---SKWKAGEPIV-DELQILSFLNSNVSTRGIAS---VEAYRK---
26 -----EILDQL-----QTYPNLVNAKSWHDICE--LFPVEQ-----DAPS
27 SMIVDKAA---EAG-Y---LTFEKFAE-----MSTRGAIFQNLL-----
28
29 --ATEEVKIFQTRPYIQTMR--VTHQ-LQLRG--VDQFGDTPI--EQADFEEMLEIGRP
30 I-----FFNIDQVTTCDIGEYD-----PLKPEEETWNFQDTTF-REL
31 YTKGTPS--LIV-----GLSFLGLRY--
32
33 >NC_035126.1
34
35
36
37 -----MVSGLVDA
38 LWC---ADEFVFLVLSID--LKTLKALVRKIFCVGTF---NLGLLVDDYKAWANRLFHKA
39 ASTQT-----IGQMEPLRCNNIFKMLDSI---EYIKAIQDSG-SDCSF-INLQKL
40 SHLVS-SRQLPYMGLKTEIKSLKGFKDVVTSYR---PTL-----ET
41 LDLLTKAARRIGGICRS-I-RKDRPIPDAAHISMTCSGEATTSIIDGGQACAVRE----
42 GFERIMGRIAESDLEDTPFG-V---ASHREG----YALWR--TIFREEPVE-----
43 -----GGSLYDPLYQGYPKDQP-----
44 -----GRYLGLDRVLGKQLLYVAWKE-----S
45 I-----LNPTIVLRASIVPEMGNKARVVTMSPYWWVQLLQA
46 PLAHLTIAGMRLHPSVFSSFARQDQAWAAK-----GISRLNIDNWE--DL-----
47 HVLSSDLKDATNAQQQL-LTRTMLRAFMNGYGM----IH-RNTYSELALST-----IVER
48 LI--TFED-Q----TSVWATTGIMMGEPIAKPSLTLLN----LAIEELAFITYNRL
49 DLLD----SDLPSP-YRAW-RFI--HIGG-----DDHLAI-----GP
50 KKYLELITNYHLLAGSHIDPGKHGYSQVAVKYTERILS-VKNFR-----FKAPFDYDN
51
52
53
54
55
56
57
58
59
60

1
2
3 YP--KSMIVDSVKVRLLEGRQS----TMQK-KDNKNVAIGKSRMMMSGCLDWLPSDPAYWP
4 SDKKISIRNLFINRMGPLLPSKSLHPKCYHSVQLPSILGGFNLGLK-----SEIYQA
5 YTLAPPPIRWVL---NKAAGVDVT-AELKILSLLNRNISDRGTDP----LQEYRA----
6 -----SIIDQL-----YSYPQLIGAITWKEILS--KFPP-----QNEN
7 VRWSIYAA---REAHY---LSIEEFAE-----WSTRGNTFVQLIM-----
8 -----
9
10 --GRGKNNVFNTRPYVRTFAR-VWTE-LEALQ--TDLYGDSPL--GEGEFNKTLRGLSRC
11 L-----YFDTSQSTTIDIGYYK-----PDDPDREEFEFVDATY-LEA
12 FDRMLPT--LVV-----GKKFIGFRN--
13 -----
14
15 >HACV01003050.1
16 -----AECMNFCGKLIQAQHLRDCVVGKRRNGSPWDKLL-PVLG
17 NPNGQQS-NTLLYYHMPSGSLLSQTfKGRSLRKQKFE-----KESKYFSPRTKY
18 GLFVLL-----TRDI--GVTkQIAKILVDRPVS-----HFKRIQEFINGLVDS
19 LWL---ANEQVFLECSQE--LLIVKKTIRKLFCVGTf--NLKDLVDQWKEWGNyLFHTV
20 ARTET-----IGPLKAVAKNNIFRTLNGI---PYISRMYHTN--EPDM-LLLQHL
21 SHLVS-SRQMPYMGFKTELDsrKKfKEVLLSSYE---PPK-----DF
22 LQKISMAARRIGGICRT---LRPKLNSGETHISVTSSGTITHSIEQGGQAAAVME----
23 AMKRILLVPEESFEEDTPFG-T---AVHKKG-----IPLWK--TLYRSSNDDM-----
24 -----SSYEFLEPY-NLIKEMS-----
25 -----GRFRGLDRVTGAQIMYVAWRELEATPILRAEVVPEMGNKARHVTL
26 SAYWLNVLQSPLCHLLIDSMKYHWRELEATPILRAEVVPEMGNKARHVTLSAYWLNVL
27 QS
28 PLCHLLIDSMKYHPSVFSSFNrQDQAFEAVK-----GMCLKKELQLLP-GH-----
29 AVLSSDLKDATNAQQWS-VTKAILKGYISGANL---SF-SDRYINLVLDL----IGPR
30 LV--KFCD-G-----TTVLSQVGIMMGEAIAKPSLTLLN-----LSIEELAFLEYNNAT
31 ELLT-----SNDPAP-YKQW-RFL--HIGG-----DDHLAI-----GP
32 VCyLNQITHNHMLAGSHIDPGKHGfSQICVRYTERLIN-IRNFQ-----YKQPFHRDD
33 YS--KSIIVDSVKVRLLEGRGLS----TEMK-RDNKNAAIGKSQTLGGTLEWLPIDDRFWT
34 ETKKESIRALFVERMGELLPRKAVNPRAYAAIHLPQYfGGYGLGfK-----REWGKY
35 LRDSPAPHRGLV---MLAHLGVDVR-HHLKVFRKLNSNLS-----
36 -----
37 -----
38 -----
39 -----
40 -----
41 -----
42 -----
43 -----
44 -----
45 -----
46
47 >GACM01002912.1
48 -----NFTGNLIQAHLlSECAGKDRKRSPWDRLL-PVLG
49 NPNKVDV--NRQFYLMPSGSLLSEtYKGSSLMRQKFH-----KRERFFNPRQKA
50 GLRIIL-----TRDI--GLSQRVAVQLVNRPIV-----HYKRIEEFISGFVDS
51 LWL---AEEQIFLFDGKW---IKLIRKLIRKIFCVGTy---NLGSLVDQWKEWGNnLFHTL
52 AESTT-----IGVLRKPAENNIFRRLDtI---DYISAVYRGD---KSMKLMQHv
53 SHLIS-SRQMPYMGSSTEEKAWDKfQSVLTSDfV---PSE-----KT
54 IVQLQMAARRVGGICRS---IRSRRIPDGVAHISVTSSGEYSYPIAKGAQAAAVKA----
55
56
57
58
59
60

AMERILTVVPLEDQEEDTPFG-L---VRHHKG-----IPIWK--TLFRTEPLD-----
-----TELPFLAPY-ALIKEQE-----
-----GRFAGLDRVTGKQIMYVAWKE-----
-----YRPLPVLRAEVVPEMGNKARLVTISDYWLNILQS
PLSHVLIDAMKFHPSVFSSFHRQDQAFEATK-----GLCSLKRKSLK-GE-----
AVLSSDLQDATNAQQWS-VTIAMLRGFIQGFGL---SF-RPEYVELVLST----IGPR
LV--CFRD-E----ISILSKVGIMMGEAIAKPSLTLLN-----LSIEELAFLRHCRAE
ELLY----SVEPAP-YRDW-RYI--HIGG-----DDHLVK-----GP
IPYLNLTQIHL DAGSHIDPGKHGFSRICVKYTERLLN-LSNLE-----HGCPFDPSD
YS--RSTIVDSVKVRLLEGRQS---TLLK-KDNKNVAIGKSTQLGGCLEWLPIDDRFYT
ETKKASIRALFIERMGSLLPRKAVNPRAFAAHLPTIVGGYGLGMS-----SELQKF
LEASPEPHKGLL---MKAFCGVNVK-EDLKIFRVLNTNTSDRGVEN---IQQFQQ---
-----KIIDQL-----SEYPQMVNAMDWWELKR--KFPDP-----SGNS
KRTIALAA---DAG-I---LSFEFAK-----RATRGNLFQALLM-----

--GRKDLKVFNTSPFVRTYKNIVWDE-AESRGLLTWSEGFNLN--NDE-IAIAIKNIAPQ
W-----YFDINQETAMDTGHWD-----PENPETETWDFKDSTY-IDK
YTEGFPS--FDV-----GFKVLGLRH--

>GFCJ01029165.1

SHLIS-TRHFPTGGKSVERQAVSDFVEVTTTLKN---VDS-----DF
LKKVRSASTIIGARLKE-----NSRPNLQTHFSLASGSLNKTISEGGRGQEVKE---
AIENLKFIPIEDKEIKILGWT----VNDIAG----RPRWQ--TWFRDPDWENPE---
-----PDVPFGELLKDTLLGYP-----
-----LYRQGFDEVLGKQIMALAANT-----FLEEV
S-----AGRPSVRVITVPEPGGKARIVTTGPWWLYILEQ
PFSHAMRSLLVHHPSAIAGLCRADQAWQYLR-----LLSNAKKYLK--SN-----
YCLSSDLKSATDAIPFE-ICEAVFEGFFDGIGQ-----SPTMSPVIAKLL----TCGR
EC--ILKS-G----FSYTTVRGIFMGEPMAKPILTLVN-----LVYEEIALRDFL---
-----KVKQKPISVPW-RCF--AVAG-----DDHIAV-----GP
LPYLKMITENHLKGGSIISPDKHGLSRVLVRYCEKLID-LTRLKT-DWTM-ESINKDTLG
YR--NGPFVDCVKVRLISPASK---GNES-FNEKNTAIGKGLTLGATLRWLNKDH--FP
TKWVNMVRDRFFNRMGTLLPSRESG--VFWHLLLPALGGLGLWIE-----DDVPLL
NNRLPQPSRALV---RDIDANHPDLKKNLKLKFELTSNSTFRGFKL---GDEPFE---
-----QCIDTA-----KEILQSTKC-TLRDLET--EFPVLK-----GLNP
RGQIAFLR---TKE-Y---VTEEELTK-----SIKRPFLFNQILA-----

--GTSKRKPFNTEALKHRYAK-IWES-LYKIE----PNFELD--EVV-VRKALKYREYQ
D-----LYSIQERELIPFRN-----SEVFNVA-IEE
ITLGLPA--FSV-----EW-----SFAGVA---
--SDTQ-----DQSTKNQPT-----

>GFIK01043228.1
-----RPFFCFNMSSFIACDLRAPAGRKHGRGYRLYI-PRLF
ENGDIYL-----PAKYDSVLNGCFL-----WKRKNLTNVGRG
LVIQVL-----CRDF--GFSKRFAEVFASRNIS-----EIRRIEQTIHAVLDC
LLL---YDTKLFIENRY---YSLLKHIVRRIMSVGTF---SIGLVISYWKDFTNDIWMTV
SEAE-----YQGEISIDKRNFFYEIERVLCLKSLKGEKLDKV-----QLQRF
AHLTS-TRGLPCGDKQSEKRAILKFKEHVTTPTQ---VSDYVVE-----DY
GRAAGLIARKCARRFEG-----FVRAAHISLATAGSLDTTVSEGGRGEEIRR----
AIFPMLTFIPKESSIVKTIPIF-D---LNEVAG-----VERWR--TWCRRPEPYMDF-----
-----PDKAFGTPNGETLGGCE-----
-----VFIQGFDEAIGEQLACAYFA-----YQQDK
G-----DQVEILTRVLTIIEPPGKSRIVTTAPWYVYVLEQ
PVAHIMRLVVSNHPSAESGLTRSDQAWAYLH-----LLERAKDHLQD--DF-----
KGLSSDLVASTDLIPKP-ILEVLYTEFNKGAHL-----VGPLFEISLDLL---RRQR
FC--LFRD-G----SNCLTARGIFMGEPLTKALLTLN-----LCTEEIAIRKYLS--
-----YDFIQPVKVPW-RCF--AVAG-----DDHIAV-----GP
VEYLRLITQTHLAAEAQLSAGKHAVSSVALRYCEKMLD-LRKVKS-DWKV-KTINSSLEN
YL--NSPFVDSVKVRLLSPCSK----TNES-FNDRNTAIGKGKSLARTLSWLNSTL--FP
TKWVSMVRDRFFSRMGPLPSRSSG--VYWHLLLPETLGGLGLGCP-----EDIPDL
LGKLPTLSKSFL---SEIAEGKITD--LYRAFRSLLSNSTYRGYEMATTEIELAKS---
-----LLFDSK-----WLFPCCKSMRELRE--SYPEIE-----EKPV
QNQLSYLM--KQG-W---MRYDRLED-----TFLRGFLFKEILT-----

--GNVVVKAFNSQKIKKRYSD-LWDR-FY-----DGDCVL--TEEQLKKAWKFRDFF
D-----FYQVKNEVKAWIRK-----KLVTCTA-LQE
LTYGLPD--LTIRPSLIAMRYHDEELRDEDKVLTEALKASDITTWEDAITHEVLKDL---
--KAKL-----DSESQVEPKEALTNIVSESCDKKEG-----
>KX883482.1
-----MQSTNLLTSKLIQTGCPRAPAGKLGRGYRIFI-PHLY
EN-----GDYLLPGRNHSL-----KGVVVRNRMVKP
NLLISK-----LTLM--GFGKR FATLVSLRPIS-----HIERAEQSVLGVIDS
LLL---YDSELFITDTG---YKLLKFIVKKTFFKVAIF---NCALVTKLWKEFSFYVWAHV
TETV-----TREKPVRKRENFFFSLVSEQDI-EVLKLTNLN-----KTLTRL
AHLTS-SRQFPPGERQQEIKSLKDFESHVTEPYI---GNP-----VF
LRRIRLASRVVGRKARR-----LGMKPLTDSHLSLAAAGSFYTSVKDGGRAEELN----
MLFKYLAYVPKESCEIKTPFY-T---LKDVEG-----VERWR--TWCRAVYDHF-----
-----PNVSFGNL---LPETL-----
-----MGFETYQGFDEILGMQILACSYLE-----QEEHL
K-----CNQNIPIVRLTITEPGSKARIVTTGPCWLYTLQQ
PVSHVLRGFLGQHPSAAAGLTMSDQAWQFLF-----LLEKAKSHFGD--DF-----
SCLSSDLTAATDVIPLC-ILKELYEGFLEGLRI-----EGPFLNIVGKMI---QLPR
LC--SVEKIN-----SYFLNSRGIFMGEPIAKVLLTLLN-----LSCEEIAIRDYLGV-
-----DFRASV-QVRW-RCF--SVAG-----DDHIAV-----GP
VGYLDKITETHIKAGSKISDLKHGHSRIA VRYCEKILD-VRNFKG-SWSK-YTINDSTEA
YL--ASPFVDSIKVRLLSPAK---NVLS-FNEKNTAFGKGKSLGRTLQWLNKDC--FD
SKWISLVRDWFFKRMSSLLPDRSSG--VYWHLLLPEHLGGLGLGTE-----RDFEDL

IVRLPSPSRTLL---KSIEDGNPNM-GHIRLFSGFTSNTTYRGYKIPETEKEIAAL----
-----FFQGI-----LEEKSDKFENLVKLYRYE-----DLIP
NTAIRLLR---TQG-W---VRKDYLE-----KILRPFLFKEILT-----

--GRAKVKAFNTEHLKSRYAR-LWDL-----TYRGEETI--SPQTLQVCFKPPAYH
L-----YYKMGEKLELPFRGRV-----LKVNL-LEE
GLLGMPD--LSI-----PWEVIGDISGP
LAPKRG-----NHARSRVIKSPRRRNS-----
>GFLH01017038.1
-----SIGTLMHGSQTAHLGTRVGSIHWTVVL-PQPH
DVGLKLR-----TSSKPFERISFKGEQI-----SSKKVFGEKGRS
GLRAYL-----CTTI--GIGVKHAEVIVSRKTS-----EIRRMFRQWESIIN
FLL---NPKFFELRF---QPAWKVKVLAFLGVAPF---SVQCTTRMYKEFVEFLKQEY
FEPL-----EGEKISLPKDNLFKPLV-----EVVRSVLNKD----RKT DYVHF
AHICT-TRNLAPGDAKEQDNSIREFIELTTTEFE---VSK-----LD
LSNLSLAMDRIVSICLN---SYESGGYLGAHMSISESAEYNVPQSKGGKAAAILA----
EAGEILNRIADKDQKVETPFR-S---YTHRQG---YPLWM---CI-----
-----ADVGVKDPV-EWLSPE-----
-----MTIVDRPPPREGLDKTTGEQLFVAAWIA-----F
N-----ESRSIKVRAVSVPEPGGKSRVVTVSQWWLPVIQQ
PLSHLLSGVLSKHPSAYSVFRRQDQAWQSLK-----LASRIKHTQTMS-GF-----
AWLHSDFTSATDAVPHK-VAYCIMERLIMRC-----FPNFKPVLRI----LGER
EV--EVKK-----DTFFTKRGIFMGEPFSKLVVGSI-----LAIEELCYSIEKSIS
L-----RRTFTK-PSNW-RFM--HVGG-----DDHRCY-----GP
VSYLKRLTRLCKQFGYIVSPTKHIMTRRAGMYTEKCIY-YHDRVINMKV-QQIND---N
VE--KSVWVDSIKLRLSPFTK---PLDG-RDDRNIAGKAKTLSRSIQWFPMGT--EL
ETIARLAIERYKLRFTMIPSSNRK-TMTAIIRLPDSVGGGLGLVYQ-----PEVED
FRNLPEIFWWAL---NVIHAGHPKNYKVRSVLQ--NTWQNFSGYEQ-----LKYLS---
-----DFKSQI-----EDYPNMVG-VTFNALKE--EFNEHF-----NGDN
TEFVDFLR---EKN-I---IQIHEFMK-----LIERPHKFYKLMT-----

---RSRIKHFNLTLDNRTRIAR-TWEA-LENLKSEITDPIPLIK--EMEVIQGAAKLARHD
W-----FIKLDEPSEAAFISTE-LSEKQLDDEEYFTFMLDGTISMRTIPF-SEH
ILYNAP--LKL-----
---HNM-----KIQN-----
>HACV01002003.1
ILNIETGIPRDPAGVQDCSEGSSTLRLGSPSESVRAPLGPVPSTKSRQWGRGLLV-PTLS
PDSVLQN-----PVRPIYEVKWGGVSLSG-----GITAYGSRKWQ
RLRKYL-----PTVW--GVSVRTAGILSGRSRS-----YLIQVFRNLDSLIDN
LML---FDRTLFDQNNAES-RSVITRFVRKVLTVSAF---SIGDCIAKWKDFVNHFVTVS
SDATV-----SSTVSLKGNCFKFLLNE---GKFKRLVND---KPSTREDYMDI
AHVIS-TRNLANGGVKAQQAASQFRKTTSTEFK---VSP-----EN
LKKLRLVASRLGRIVTS----IRGSNKLRGHCSLNGSGTLDPVPVQGGGRATDALL----
DLKKFLSHIPPSGESRTYPWG-E---VWFPSG---IPRWC--AFGKLENF-----
-----TDAEWMSLANREIIDRG-----
-----ALRNGCDSTTGRMVMAYAYEC-----YRPFE

E-----SGTPIRQAAVSEPGGKVRMVTGTPWWLAVLQQ
PLCHGLREILAYHPSAHSCLMRCDQAWQSLH-----VLERLGISKLED-GM-----
AVLSSDLKEATDAIPHS-VGKVLLSGFLEGLGR-----EDHDWITNF-----IGER
TV--FCED-G-----EFTLRRGVMMGEPLSKICLVLLG-----LAMEELAFSEHEGLS
L-----HRRDTP-RCHW-RAF--HLGG-----DDHLAV-----GP
VTYLRITSNHLFGSVISPVKHRMSSIVVYTEKVLRL-IKGAVL-NLKA-SQVDGNPG-
----NCTFIDSVKVRLLSPFSK----ATDS-QNDKNIAIGKVRGIAATLKYPSES----
--IKRTVFDRVAYRFNHYIGSERHR-TIRAVMSLPSSELGGLGFAFD-----DRY
LGRLPPIFNRAV---RTVINGGGVGFRVRRILGQIFANRCPRGLVSNALMVDDYVE----
-----QMIEQF-----LDYPGSVDMVSSGDLWA--RLDENG-----LMSF
REFLEKAS---HEG-W---ISLHDIPR-----IAEKPFLMRKILE-----

--GSEVPEYFRTASNAQRIAE-CWNS-LEAIDD-VLDPTGGEL--SPLEIREASRRAKHT
V-----FVNLNEMTCAALVDPT-----SEAYDPGDPWISADFDLSM-REL
LTYGEPSTVPL-----

>GFBP01062652.1
-----LTLRRYMPPTQLTAPPGRFGEGGVHSPGLLV-PNLT
PDHVVV-----NPVRPIWEHRFGGVCFASLGSK-----QKWRT
SFARTL-----RVEL--DMGRRASETLANRGLS-----YIKRVLKNLDSLIDN
LIL---FSPSMFDRSTPGVIYDTLKRILRVTLKVSTY---SVEDQISQWKDFVNHFFTLV
SKSQV-----EKQICLRHNCYKWMLEQ---DPFKSLLSKI--DMTT-KDWTSTI
AHLIS-TRNFAPGGESAKLKAIRAFKETTTSGFS---VSE-----EN
LRKLKLATLRVTGISLH---LSEEKRLSGGHVSLNSSGTLDVSVAKGGRAADALL---
DLKQFLDEEPVQGETRVYPWG-E---LYYPPG-----LRRWR--AFGPLQD-DDG-----
-----PFMSKS-R--YDLI-----
-----DREAMLNGCGEYTGVMIFTVAYEC-----YRVYL
N-----SGELIPRQATVSEPGGKARIVTTGPWWLAVLQQ
PLCHGLRELIGYHPSAHSCLLRADQAWQSLH-----VWAKCSDNCIPE-GE-----
AVLSSDLKEATDAIPHS-VGQVLIRGCLEAIGA-----TKWLWMADL-----IGPR
LV--LAED-K-----DVFTLRRGVMMGEPLSKICLILLG-----LTMEELAFSEYSSLS
L-----KRKFTP-AQG-W-RAF--HLGG-----DDHLAK-----GP
MRYLKSITSYHRLFGSIVSPYKHRISVRAVITYTEKILV-FENRIL-NMPV-EQVNSNID-
----KSIFIDSIKVRLLSPFTK----ATDT-MNDKNVAIGKCIGIAGTLQWMHDS----
--LKRTLIDRIHYRFRDFIAGPHHR-TLSAIQSLPTTLGGLGLSVD-----TKY
LENLPEPYNRAI---REIKGGVLAAYKARMVMNSTFVNKNLRGIKS-----DDYVT----
-----ELITQW-----VEYPSSVNAMNFKEAIA--VVDPEG-----NNSF
RHNLWLLK---KRN-I---VSMNDIPQ-----IAESAYVMRRLLE-----

--QSQSGKYFRTERLRDRVSK-TWEG-LTKLDLPESGDVLSSED--ELKTGTRLAKQALFI
D-----LTTMGTIADVADSTNPD-----YLPDDPWLADEFIDIPF-GQS
IKYGPQN-----MGIDLH
PFPEEK-----DLAC-----