Weedall, GD

The Entamoeba lysine and glutamic acid rich protein (KERP1) virulence factor gene is present in the genomes of Entamoeba nuttalli, Entamoeba dispar and Entamoeba moshkovskii

The *Entamoeba* lysine and glutamic acid rich protein (KERP1) virulence factor gene is present in the genomes of *Entamoeba nuttalli*, *Entamoeba dispar* and *Entamoeba moshkovskii*.

Gareth D. Weedall

School of Biological and Environmental Sciences, Liverpool John Moores University, Liverpool, UK.

Correspondence address: G.D.Weedall@ljmu.ac.uk

Abstract

The lysine and glutamic acid rich protein KERP1 is a cell surface-expressed virulence factor in the human pathogen *Entamoeba histolytica*. It was originally suggested that the gene was absent from the related, avirulent human commensal *Entamoeba dispar*, an absence which would be relevant to the differential virulence of these species. Here, the gene is shown to be present in *E. dispar*, and its sequence is presented, as well as in a virulent parasite of macaques, *Entamoeba nuttalli*, and the primarily free living, opportunistically parasitic *Entameba moshkovskii*.

The lysine and glutamic acid rich protein KERP1 is a cell surface-expressed virulence factor in the human pathogen *Entamoeba histolytica* (Seigneur et al. 2005; Santi-Rocca et al. 2008; Perdomo et al. 2013; Faust et al. 2011; Perdomo et al. 2016). KERP1 exists as a trimer on the parasite surface (Perdomo et al. 2013) and can bind to human enterocytes (Seigneur et al. 2005) as well as playing a role in the development of amoebic liver abscesses (Santi-Rocca et al. 2008). This all suggests that KERP1 is among the set of key *Entamoeba* virulence factors (Wilson, Weedall, and Hall 2012).

In the original paper, that used a range of in-depth molecular and biochemical analyses to identify KERP1 in *E. histolytica*, it was suggested, based on sequence similarity searching and attempted PCR amplification, that the gene may be unique to *E. histolytica* and absent from the genome of its avirulent relative, the human commensal *Entamoeba dispar* (Seigneur et al. 2005). Loss of virulence factor genes from the *E. dispar* genome, or loss of their function, could explain the avirulence of this species and by doing so help us understand the molecular virulence processes in *E. histolytica*. Loss of function of another key virulence factor, cysteine proteinase 5, has been reported in *E. dispar* (Willhoeft, Hamann, and Tannich 1999).

However, proving the absence of a gene from an *Entamoeba* genome is not easy. Genome assemblies of *Entamoeba* species are highly fragmented due to several challenging features of their genomes, including extreme nucleotide composition bias and highly repetitive genomes (Weedall 2015; Weedall and Hall 2011). In such fragmented assemblies, genes can be partially represented or go unrepresented entirely. A chance match to part of the *E. dispar* genome in a BLAST sequence similarity search of KERP1 suggested some or all of the gene may in fact be present. This was explored further.

First, the protein sequence of the *Entamoeba histolytica* HM-1:IMSS KERP1 gene (accession number EHI_098210) was used to search the predicted proteomes of 4 species, in addition to *Entamoeba histolytica* (strain HM-1:IMSS): *Entamoeba nuttalli* (strain P19); *Entamoeba dispar* (strain SAW760); *Entamoeba moshkovskii* (strain Laredo); and *Entamoeba invadens* (strain IP1). The BLASTP (protein *vs*. protein BLAST) search was run with default parameters (in AmoebaDB on 2020-04-19). Only 3 matches were returned: *E. histolytica* EHI_098210 (100% self-match); *E. nuttalli* ENU1_189420 (97% amino acid identity over the whole protein); and *E. moshkovskii* EMO_099600 (45% amino acid identity over part of the protein). This confirmed that the *E. dispar* predicted proteome did not contain a KERP1 orthologue, confirming that a complete gene was not present in the genome annotation. All three putative KERP1 sequences were reciprocal best matches to one another, including the highly divergent *E. moshkovskii* protein (**Figure 1**). However, only the C-terminal part of this protein showed similarity to the other KERP1 proteins. The high level of divergence between *E. histolytica* and *E. moshkovskii* KERP1 suggests that the even more distantly related *E. invadens* may possess a KERP1 gene too highly divergent to be identified.

Next, the EHI_098210 protein sequence was used to search the *Entamoeba dispar* genome using TBLASTN (protein *vs*. translated nucleotide BLAST) with default parameters (search run in AmoebaDB on 2020-04-19). One highly similar match was found. This was to *E. dispar* scaffold DS550082. The match was of the C-terminal part of the *E. histolytica* protein (from amino acid 103 to the C-terminal end at 184) and matched an open reading frame running from nucleotide 3 to 284 in DS550082, indicating that the gene is in fact present in the *E. dispar* genome but is incompletely represented in the genome assembly and therefore unannotated. In support of this, the scaffold also contains a partial beta-amylase gene (EDI_095020) downstream of the putative KERP1, as is seen in *E. histolytica* (**Figure 1**).

By itself, this is too little evidence to claim that the gene is complete or functional. Using a unpublished *Entamoeba dispar* (SAW760) low coverage 454 read dataset, reads similar to *Eh*KERP1 (by BLAST sequence similarity searching of the raw reads; results shown in **Supplementary Data 1**) were assembled and used to extend the partial *Ed*KERP1 gene to reconstruct a full-length protein coding gene and flanking sequence (GenBank accession MT431639). It is shown in alignment (aligned using MUSCLE (Edgar 2004)) with *E. histolytica* and *E. nuttalli* (**Figure 2** and **Figure 3**; *Em*KERP1 is not shown due to its divergence from the other sequences). Four mismatches in the 3' portion of the primer binding site of one of the primers used to amplify KERP1 from *E. dispar* genomic DNA may explain the failure to amplify the gene reported previously (Seigneur et al. 2005).

The *Ed*KERP1 gene contained 45 observed (not corrected for multiple changes at the same site) nucleotide differences to *Eh*KERP1 and 43 to *En*KERP1. By contrast, *E. histolytica* and *E. nuttalli* were much more closely related, with only 10 observed nucleotide differences (**Figure 2**). Of these observed differences, roughly equal numbers were synonymous and nonsynonymous in all comparisons (*Ed vs*. *Eh* = 24 synonymous, 21 nonsynonymous, 20 amino acid mismatches; *Ed vs*. *En* = 22 synonymous, 21 nonsynonymous, 20 amino acid mismatches; *Eh vs*. *En* = 5 synonymous, 5 nonsynonymous, 5 amino acid mismatches). In addition to single nucleotide differences, a two-codon indel was observed as an insertion in *E. dispar* (**Figure 2** and **Figure 3**). This indel and six amino acid changes are within the predicted coiled-coil domain, and the indel and two of the amino acid changes within the predicted universal stress protein (USP) domain, of the protein (**Figure 3**) (Perdomo et al. 2013).

Evolutionary distance between *E. histolytica*, *E. nuttalli* and *E. dispar* KERP1 were estimated, accounting for the underestimation of true divergence due to multiple substitutions at the same site, using a maximum likelihood model (general time reversible with invariant sites, GTR+I) implemented in MEGA 7 (Perdomo et al. 2013; Kumar, Stecher, and Tamura 2016). The results showed the close relatedness of *E. histolytica* and *E. nuttalli* (0.018-0.019 nucleotide substitutions per site; **Figure 4A,B**) compared to *E. dispar* (0.095-0.100 nucleotide substitutions per site), five times the level of divergence between *E. histolytica* and *E. nuttalli*. However, this is dwarfed by the divergence of *E. moshkovskii* (1.398-1.410 nucleotide substitutions per site), fourteen times the level of divergence between *E. histolytica* and *E. dispar* (**Figure 4B**).

Here it is shown that, in addition to the virulent human pathogen *Entamoeba histolytica*, the closely related virulent simian pathogen *Entamoeba nuttalli*, the avirulent human commensal *Entamoeba dispar* and the primarily free-living, opportunistic human pathogen *Entamoeba moshkovskii* all possess a gene encoding the lysine and glutamic acid rich protein KERP1. Extensive molecular and biochemical analyses first identified KERP1 and implicated it as an important virulence factor with roles in cell adhesion in the gut and in the development of extra-intestinal abscesses (Seigneur et al. 2005; Santi-Rocca et al. 2008; Perdomo et al. 2013; Faust et al. 2011; Perdomo et al. 2016). Further research to build upon this work and understand the biology of KERP1 is needed and comparative evolutionary analyses can be a part of this: for instance, given that the KERP1 gene is present, functional differences and differences in gene expression among species may be important in understanding their differing virulence phenotypes. It is hoped the data presented here can aid such studies of this important virulence factor.

References

Edgar RC. 2004. MUSCLE: A Multiple Sequence Alignment Method with Reduced Time and Space Complexity. BMC Bioinformatics 5 (August): 113.

Faust DM, Marquay Markiewicz J, Santi-Rocca J and Guillén N. 2011. New Insights into Host-Pathogen Interactions during *Entamoeba histolytica* Liver Infection. European Journal of Microbiology & Immunology 1 (1): 10–18.
Kumar S, Stecher G and Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution 33 (7): 1870–74.
Perdomo D, Baron B, Rojo-Domínguez A, Raynal B, England P and Guillén N. 2013. The α-Helical Regions of KERP1 Are Important in *Entamoeba histolytica* Adherence to Human Cells. Scientific Reports 3 (January): 1171.
Perdomo D, Manich M, Syan S, Olivo-Marin J, Dufour AC and Guillén N. 2016. Intracellular Traffic of the Lysine and Glutamic Acid Rich Protein KERP1 Reveals Features of Endomembrane Organization in *Entamoeba histolytica*. Cellular Microbiology 18 (8): 1134–52.
Santi-Rocca J, Weber C, Guigon G, Sismeiro O, Coppée J and Guillén N. 2008. The Lysine- and Glutamic Acid-Rich Protein KERP1 Plays a Role in *Entamoeba histolytica* Liver Abscess Pathogenesis. Cellular Microbiology 10 (1): 202–17.
Seigneur M, Mounier J, Prevost M, and Guillén N. 2005. A Lysine- and Glutamic Acid-Rich Protein, KERP1, from *Entamoeba histolytica* Binds to Human Enterocytes. Cellular Microbiology 7 (4): 569–79.
Weedall GD. 2015. The Genomics of *Entamoebae*: Insights and Challenges. Amebiasis. https://doi.org/10.1007/978-4-431-55200-0_3.
Weedall GD and Hall N. 2011. Evolutionary Genomics of *Entamoeba*. Research in Microbiology. https://doi.org/10.1016/j.resmic.2011.01.007.
Willhoeft U, Hamann L and Tannich E. 1999. A DNA Sequence Corresponding to the Gene Encoding Cysteine Proteinase 5 in *Entamoeba histolytica* Is Present and Positionally Conserved but Highly Degenerated in *Entamoeba Dispar*. Infection and Immunity 67 (11): 5925–29.
Wilson IW, Weedall GD and Hall N. 2012. Host-Parasite Interactions in *Entamoeba histolytica* and *Entamoeba dispar*: What Have We Learned from Their Genomes? Parasite Immunology 34 (2-3): 90–99.

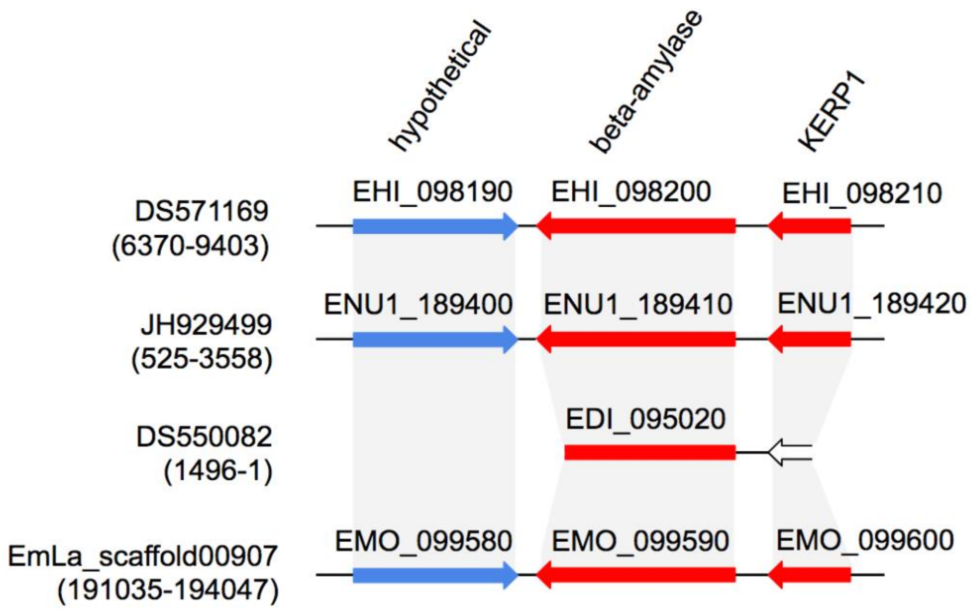


Figure 1. Orthology and synteny among (top to bottom) *Entamoeba histolytica*, *E. nuttalli*, *E. dispar* and *E. moshkovskii* near the KERP1 gene. Arrows are genes (blue = positive strand; red = negative strand; accessions shown above each; gene products labeled at the top). Assembly scaffold labels and positions are shown on the left (the order of the numbers indicates the orientation of the scaffold). Orthology (determined by reciprocal BLAST) is indicated with grey shading. The white, open-ended arrow is the unannotated partial *Ed*KERP1.

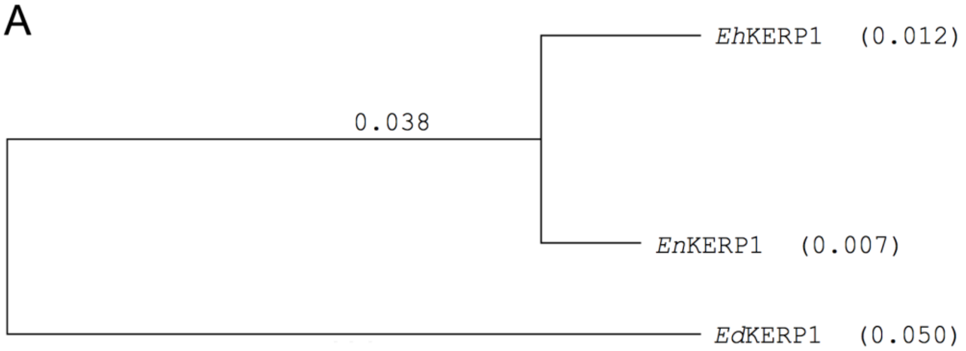
```
EhKERP1  ttgttcaaattgtcatcaaaatggctttataaaatataaaagaaatgagtttaacaaaac
EnKERP1  ttgttcaaattgtcatcaaaatggcttta aaaatataaaagaaatga tttaacaaaac
EdKERP1  ttgttcaaattgtcatcaaaatg ctttataaaat t aaagaaatga tttaacaaaac

EhKERP1  aaaggatttgatcttttcaagattcagtcaagttcagttaATGGAAAATATTATAAGCAC
EnKERP1   aag atttgatcttttcaagattcagtca  ttcagttaATGGAAAATATTATAAGCAC
EdKERP1  aaag atttgatcttttcaagatt agtca     a tt ATGGAAAATATTATAA CAC

EhKERP1  AACAAATACTATTCAAGGAAAAGCACAAGCTCTTCTCAAAAAAGAAGTATTAAATGAAAA
EnKERP1  AACAAATACTATTCAAGGAAAAGCACAAGCTCTTCTCAAAAAAGAAGTATTAAATGAAAA
EdKERP1  AAC AATACTATTCAAGGAAAAGCACAAG  CTTCTCAAAAAAGA   ATT AATGAAAA

EhKERP1  TGAAAAAGAGATAGTTGAAATGATTAATGAATTAGCTAATGCATTAAATAAAACTATCAC
EnKERP1  TGAAAAAGA ATAGTTGAAATGATTAA GAATTAGCTAATGCATTAAATAAAACTATCAC
EdKERP1  TGAAAAAGAGATAGTTGAAATGATTAA GAATTAGCTAATGCA TAAATAAAACTATCAC

EhKERP1  AATTCTTAATGCACAACCACCTTTAAAGACTGAATCAAAAACAAAAGAAGAATTAAAGAA
EnKERP1  AATTCTTAATGCACAACCACCT TAAAGAC GAATCAAAAACAAAAGAAGAATTAAAGAA
EdKERP1  AATTCTTAATGC CAACCACCTTTAAAGAC GAAT AAAAACAAAAGAAGAATTAAAGAA

EhKERP1  AGAAGAGAAAGAATTAAAGAAGCAAAAACAAATGGAAGAGAAGAAATTAAAAATGGAAAA
EnKERP1  AGAAGAGAAAGAATTAAAGAA CAAAAACAAATGGAAGAGAAGAAATTAAAAATGGAAAA
EdKERP1  AGAAGAGAAAGAATTAAAGAA CAAAA CAAAT GAAGAGAAGAAATTAAAAATGGAAAA

EhKERP1  GAAGGCA------GAAAAAGAAATTGTAAAAGAGAAGAAACCAAAGAAAAAACAAAGACT
EnKERP1  GAAGGCA------GAAAAAGAAATTGTAAAAGAGAAGAAACCAAAGAAAAAACAAAGACT
EdKERP1  GAAGGCT      GAAAAAGAAATTGT AAAGAGAAGAAACCAAA AAAAAACAAA ACT

EhKERP1  TAATGATGAAAATAATGATGAAGAAAAAGAAGTAAAAGATGATAAGAAAGTTAGTTCATT
EnKERP1  TAATGATGAAAATA TGATGAAGAAAAAGAAGTAAAAGATGATAAGAAAGTTAGTTCATT
EdKERP1  TAATGATGAAA TA TGA GAAGAAAAAGAAGTAAAAGATGATAA AAA T AGTTCATT

EhKERP1  GGAAGAAAATAAAATTTCAAAACAAACTAAAAACTACGGTAAAATTTTGCTTGAAGAAGA
EnKERP1   GAAGAAAATAAA TTTCAAA CAAACTAAAAACTACGGTAAAATTTTGCTTGAAGAAGA
EdKERP1    AAGAAAATAA  TTTCAAA CAAA TAAAAA TACGGTAAAATTTT CTTGAAGAAGA

EhKERP1  AGAAGGTGAAGCTCCTACTCCTAAAGAAGAAAAGAAAGAAAAATACAAAGAAACAAAGGC
EnKERP1  AGAAGGTGAAGCTCCTACTCCTAAAGAAGAAAAGAAAGAAAAATACAAAGAAA AAAGGC
EdKERP1  AGA GGTGA G TC TAC CCT AAGAAGAAAAGAAAGAAAA TACAAAGAAACAAAG C

EhKERP1  TGATGCCTTATTAGATAAAAAATCAAAGAAAGGAAAGAAAGATATTTTCTATGAAAATTA
EnKERP1  TGATGCCTTATTAGATAAAAAATCAAAGAAAGGAAAGAAAGATATTTTCTATGAAAATTA
EdKERP1  TGATGC TTATTAGA AAAAAATCAAA AAAGGAAAGAAAGATATTTTCTATGAAAATTA

EhKERP1  Acttatatttcaattaatttatcattacaaatatctcttattttaaataaaacataaact
EnKERP1  Acttatatttcaattaatttatcattacaaatatctcttattttaaataaaacataaact
EdKERP1  A ttatatttcaattaattta cattac aatatctcttattttaaataaaa ataaa t

EhKERP1  gaaatagaataaataatagaataaattattattaaaatgaa
EnKERP1  gaaatagaataaataatagaataaattattattaaaatgaa
EdKERP1   aaatagaataa taatagaataa t a tatta aa  a
```

Figure 2. Nucleotide alignment of the KERP1 gene from *Entamobea histolytica* (*Eh*KERP1; AmoebaDB accession EHI_098210), *Entamobea nuttalli* (*En*KERP1; ENU1_189420) and *Entamobea dispar* (*Ed*KERP1; GenBank accession MT431639). Capital letters denote the protein coding region and lower case letters are (100 bp) upstream and downstream flanking regions. Nucleotides mismatched with *E. histolytica* are highlighted in white on a black black background. Grey highlighted regions indicate binding sites for primers used to amplify the gene in (Seigneur et al. 2005), with underlining indicating regions mismatched in primers to introduce restriction sites.

```
EhKERP1  MENIISTTNTIQGKAQALLKKEVLNENEKEIVEMINELANALNKTITILNAQPPLKTESK
EnKERP1  MENIISTTNTIQGKAQALLKKEVLNENEKEIVEMINELANALNKTITILNAQPP KTESK
EdKERP1  MENII TTNTIQGKAQ LLKK  LNENEKEIVEMINELANALNKTITILNAQPPLKTE K
                                ......................................

EhKERP1  TKEELKKEEKELKKQKQMEEKKLKMEKKA--EKEIVKEKKPKKKQRLNDENNDEEKEVKD
EnKERP1  TKEELKKEEKELKKQKQMEEKKLKMEKKA--EKEIVKEKKPKKKQRLNDEN DEEKEVKD
EdKERP1  TKEELKKEEKELKKQKQ EEKKLKMEKKA  EKEIVKEKKPKKKQ LNDE  DEEKEVKD
         ............................................................

EhKERP1  DKKVSSLEENKISKQTKNYGKILLEEEEGEAPTPKEEKKEKYKETKADALLDKKSKKGKK
EnKERP1  DKKVSS EENK SKQTKNYGKILLEEEEGEAPTPKEEKKEKYKE KADALLDKKSKKGKK
EdKERP1  DKK SS  ENK SKQ K YGKILLEEEEGE  TP EEKKEKYKETK DALL KKSKKGKK
         ....

EhKERP1  DIFYEN*
EnKERP1  DIFYEN*
EdKERP1  DIFYEN*
```

Figure 3. Protein alignment of KERP1 from *Entamobea histolytica* (*Eh*KERP1; AmoebaDB accession EHI_098210), *Entamobea nuttalli* (*En*KERP1; ENU1_189420) and *Entamobea dispar* (*Ed*KERP1; GenBank accession MT431639). Amino acids mismatched with *E. histolytica* are highlighted. The dots below and line above the alignment indicate the predicted coiled-coil domain and universal stress protein (USP) domain, respectively (Perdomo et al. 2013).

A



B

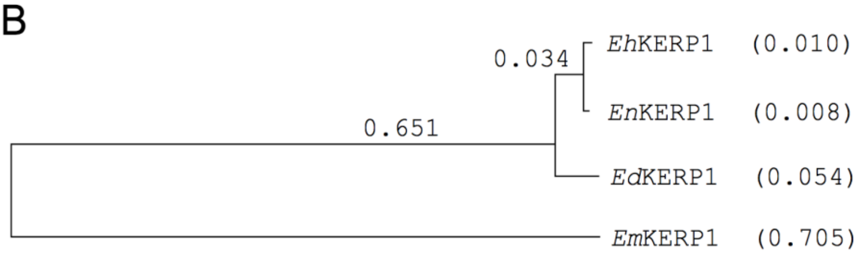


Figure 4. Phylogenies of KERP1 with branch lengths. Evolutionary history was inferred using the Maximum Likelihood method based on the General Time Reversible model with a set of invariant sites (GTR+I). Evolutionary rate differences among sites were modeled with a discrete Gamma distribution with 10 categories and an additional set of invariable sites. External branch lengths (number of substitutions per site) are shown in brackets, internal branch lengths beside the branches. Positions containing gaps were removed prior to analysis. (A) *E. histolytica*, *E. nuttalli* and *E. dispar* KERP1 phylogeny (555 ungapped positions analysed; log likelihood = -875.92; gamma parameter = 0.60; 0.001% sites invariant). (B) *E. histolytica*, *E. nuttalli*, *E. dispar* and *E. moshkovskii* KERP1 phylogeny (537 ungapped positions analysed; log likelihood = -1410.90; gamma parameter = 22.41; 28.87% sites invariant).

Supplementary Data 1. Sequence data used to reconstruct the *Entamoeba dipsar* KERP1 gene and flanking sequences.

```
# Raw 454 reads (from E. dispar SAW760) covering the KERP1 gene
# Reads were identified by BLAST search using EhKERP1 as the query
# Reads were used to reconstruct the EdKERP1 gene
>GV2P2VP01C54QP|length=438
ACAATTCAGACTAAAAGTAATAAAGGTTTGTCTGTTAATAAGAAATAAAATAAAATACTC
ATAACCTATATTAAATTTGAGTTAATTAGAGAATATTTAAATAAAATAGTTAATGAAATA
GCTATTGTTGAAGTTAAAATAACTAAAACAAAGTCTAAAAATGTTATTTTATGTATTAAA
GTAAATTTAGTAGATTTTGTTCAAATTGTCATCAAAATGACTTTATAAAATGTGAAAGAA
ATGAATTTAACAAAACAAAGTATTTGATCTTTTTCAAGATTAAGTCACAACTAATTTCAT
GGAAAATATTATAAACACAACTAATACTATTCAAGGAAAAGCACAAGTCCTTCTCAAAAA
AGATACATTGAATGAAAATGAAAAAGAGATAGTTGAAATGATTAACGAATTAGCTAATGC
ACTAAATAAAACTATCAC
>GV2P2VP01B7ZEY|length=594
ACAATTCAGACTAAAAGTAATAAAGGTTTGTCTGTTAATAAGAAATAAAATAAAATACTC
ATAACCTATATTAAATTTGAGTTAATTAGAGAATATTTAAATAAAATAGTTAATGAAATA
GCTATTGTTGAAGTTAAAATAACTAAAACAAAGTCTAAAAATGTTATTTTATGTATTAAA
GTAAATTTAGTAGATTTTGTTCAAATTGTCATCAAAATGACTTTATAAAATGTGAAAGAA
ATGAATTTAACAAAACAAAGTATTTGATCTTTTCAAGATTAAGTCACAACTAATTCATGG
AAAATATTATAAACACAACTAATACTATTCAAGGAAAAGCACAAGTCCTTCTCAAAAAAG
ATACATTGAATGAAAATGAAAAAGAGATAGTTGAAATGATTAACGAATTAGCTAATGCAC
TAAATAAAACTATCACAATTCTTAATGCGCAACCACCTTTAAAGACAGAATTAAAAACAA
AGAAGAATTAAAGAAAGAAGAAGAAAGAATTAAAGAAACAAAAGCAAATAGAAGAGAAGA
AATTAAAAANGGAAAAAGAAGGCTAGAAGAAAGAAAAAGAATTAGTTAAAGAAG
>GV2P2VP01BHT7U|length=393
GACTTTATAAAATGTGAAAGAAATGAATTTAACAAAACAAAGTATTTGATCTTTTCAAGA
TTAAGTCACAACTAATTCATGGAAAATATTATAAACACAACTAATACTATTCAAGGAAAA
GCACAAGTCCTTCTCAAAAAAGATACATTGAATGAAAATGAAAAAGAGATAGTTGAAATG
ATTAACGAATTAGCTAATGCACTAAATAAAACTATCACAATTCTTAATGCGCAACCACCT
TTAAAGACAGAATTAAAAACAAAAGAAGAATTAAAGAAAGAAGAGAAAGAATTAAAGAAA
CAAAAGCAAATAGAAGAGAAGAAATTAAAAATGGAAAAGAAGGACTGAAGAAAGAAAAAG
AAATTGTTAAAGAGAAGAAACCAAAAAAAAAAC
>GVPMWNX02IG2FH|length=298
ATTTCTTCTCTTCTATTTGCTTTTGTTTCTTTAATTCTTTCTCTTCTTTCTTTAATTCTT
CTTTTGTTTTTAATTCTGTCTTTAAAGGTGGTTGCGCATTAAGAATTGTGATAGTTTTAT
TTAGTGCATTAGCTAATTCGTTAATCATTTCAACTATCTCTTTTTCATTTTCATTCAATG
TATCTTTTTTGAGAAGGACTTGTGCTTTTCCTTGAATAGTATTAGTTGTGTTTATAATAT
TTTCCATGAATTAGTTGTGACTTAATCTTGAAAAGATCAAATACTTTGTTTTGTTAAA
>GV2P2VP01CR42F|length=486
ATATTTTATTTAAAAATAAGAGATATTAGTAATGTTAAATTAATTGAAATATAATTTAAT
TTTCATAGAAAAATATCTTTCTTTCCTTTTTTTGATTTTTTTTCTAATAAAGCATCAGTC
TTTGTTTCTTTGTACTTTTCTTTCTTTTCTTCTTCAGGGGTAAGACCCTCACCCTCTTCT
TCTTCAAGTAAAATTTTACCGTACTTTTTACTTTGCTTTGAAACCTTATTTTCTTTAAAT
GAACTGATTTTTTATCATCTTTTACTTCTTTTCTTCGTCAATACTTTCATCATTAAGTTT
TTGTTTTTTTTTGGTTTCTTCTCTTTAACAATTTCTTTTTCTTTCTTCAGCCTTCTTTTC
CATTTTTAATTTCTTCTCTTCTATTTGCTTTTGTTTCTTTAATTCTTTCTCTTCTTCTTT
AATTCTTCTTTTGTTTTAATTCTGTCTTTAAAGGTGGTTGCGCATTAAGAATTGTGATTA
GTTTTA
>GVPMWNX02GQ8BB|length=353
TTTTCTTTCTTTTCTTCTTCAGGGGTAAGACCCTCACCCTCTTCTTCTTCAAGTAAAATT
TTACCGTACTTTTTACTTTGCTTTGTAAACCTTATTTTCTTTAAATGAACTGATTTTTTT
ATCATCTTTTACTTCTTTTTCTTCGTCAATACTTTCATCATTAAGTTTTTTGTTTTTTTT
TGGTTTCTTCTCTTTAACAATTTCTTTTTCTTTCTCAGCCTTCTTTTCCATTTTTAATTT
CTTCTCTTCTATTTGCTTTTGTTTCTTTAATTCTTTCTCTTCTTTCTTTAATTCTTCTTT
TGTTTTAATTCTGTCTTTAAAGGTGGTTGCGCATTAAGATTGTGATAGTTTTA
>GV2P2VP01B15VU|length=478
ATATTTTATTTAAAATAAGAGATATTAGTAATGTTAAATTAATTGAAATATAATTTAATT
TTCATAGAAAATATCTTTCTTTCCTTTTTTGATTTTTTTCTAATAAAGCATCAGTCTTTG
TTTCTTTGTACTTTTCTTTCTTTTCTTCTTCAGGGGTAAGACCCTCACCCTCTTCTTCTT
CAAGTAAAATTTTACCGTACTTTTTACTTTGCTTTGAAACCTTATTTTCTTAAATGAACT
GATTTTTTTANCATCTTTTACTTCTTTTTCTTCGTCAATACTTTCATCATTAAGTTTTTG
TTTTTTTTTGGTTCTTCTCTTTAACAATTTCTTTTTCTTTCTCAGCCTTCTTTTCCATTT
TTAATTTCTTCTCTTCTATTTGCTTTTTGTTTCTTTAATTCTTTCTCTTCTTTCTTTAAT
TCTTCTTTTGTTTTAATTCTGTCTTTAAAGGTGGTTGCGCATTAAGAATTGTGATTAG
```