The mechanism of the *peel-1 zeel-1* toxin-antidote system

Lews Caro

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Michael Ailion, Chair

Robert H. Waterston

Adam Steinbrenner

Program Authorized to Offer Degree:

Molecular and Cellular Biology

University of Washington

Abstract

The mechanism of the *peel-1 zeel-1* toxin-antidote system

Lews Caro

Chair of the Supervisory Committee:
Michael Ailion
Department of Biochemistry

Part 1:

Toxin-antidote systems are selfish genetic elements composed of a linked toxin and antidote. The *peel-1 zeel-1* toxin-antidote system in *C. elegans* consists of a transmembrane toxin protein PEEL-1 which acts cell autonomously to kill cells. Here we investigate the molecular mechanism of PEEL-1 toxicity. We find that PEEL-1 requires a small membrane protein, PMPL-1, for toxicity. Together, PEEL-1 and PMPL-1 are sufficient for toxicity in a heterologous system, HEK293T cells, and cause cell swelling and increased cell permeability to monovalent cations. Using purified proteins, we show that PEEL-1 and PMPL-1 allow ion flux through lipid bilayers and generate currents which resemble ion channel gating. Our work shows that PEEL-1 kills cells by co-opting PMPL-1 and creating a cation channel.

Part 2:

Centromeric histones (CenH3s) are essential for chromosome inheritance during cell division in most eukaryotes. CenH3 genes have rapidly evolved and undergone repeated gene duplications

and diversification in many plant and animal species. In *Caenorhabditis* species, two independent duplications of *CenH3* (named *hcp-3* for *HoloCentric chromosome-binding Protein 3*) were previously identified in *C. elegans* and *C. remanei*. Using phylogenomic analyses in thirty-two *Caenorhabditis* species, we find strict retention of the ancestral *hcp-3* gene and ten independent duplications. Most *hcp-3L* (*hcp-3-like*) paralogs are only found in 1-2 species, are expressed in both males and females/ hermaphrodites, and encode histone fold domains with 69-100% identity to ancestral *hcp-3*. We identified novel N-terminal protein motifs, including putative kinetochore protein-interacting motifs and a potential separase cleavage site, which are well conserved across *Caenorhabditis* HCP-3 proteins. Other N-terminal motifs vary in their retention across paralogs or species, revealing potential sub-functionalization or functional loss following duplication. An N-terminal extension in the *hcp-3L* gene of *C. afra* revealed an unprecedented protein fusion, where *hcp-3L* fused to duplicated segments from *hcp-4* (nematode CENP-C). By extending our analyses beyond *CenH3*, we found gene duplications of six inner and outer kinetochore genes in *Caenorhabditis*, which appear to have been retained independent of *hcp-3* duplications. Our findings suggest that centromeric protein duplications occur frequently in *Caenorhabditis* nematodes, are selectively retained for short evolutionary periods, then degenerate or are lost entirely. We hypothesize that unique challenges associated with holocentricity in *Caenorhabditis* may lead to this rapid 'revolving door' of kinetochore protein paralogs.

Table of Contents

Chapter 1. INTRODUCTION *1*

1.1 ANIMAL TOXIN-ANTIDOTE SYSTEMS AND THEIR MECHANISMS **1**

1.2 REFERENCES **9**

Chapter 2. TOXIN MECHANISM OF THE PEEL-1 ZEEL-1 TOXIN-ANTIDOTE SYSTEM *18*

2.1 INTRODUCTION **18**

2.2 RESULTS **19**

2.2.1 PMPL-1 is required for PEEL-1 toxicity 19

2.2.2 PMPL-1 is a conserved membrane protein of unknown function 20

2.2.3 PEEL-1 and PMPL-1 are sufficient for toxicity in HEK293T cells 21

2.2.4 Plasma membrane localization of PEEL-1 and PMPL-1 is critical for toxicity . 22

2.2.5 PEEL-1 toxicity causes cell swelling in HEK293T cells 22

2.2.6 PEEL-1 has an amphipathic helix that is critical for toxicity 23

2.2.7 PEEL-1 and PMPL-1 create a non-specific monovalent cation channel 25

2.2.8 Predicted PEEL-1 pentamer structure has features of an ion channel 28

2.2.9 In vitro reconstitution of the PEEL-1/PMPL-1 ion channel 28

2.3 DISCUSSION **29**

2.4 MATERIALS AND METHODS **30**

2.4.1 Worm strains and maintenance 30

2.4.2 Isolation of suppressors of heat-shock induced PEEL-1 30

2.4.3 Identification of pmpl-1 31

2.4.4 Heat-shock PEEL-1 toxicity assay 32
2.4.5 Sperm-delivered PEEL-1 toxicity assay 32
2.4.6 Plasmid construction 33
2.4.7 Microscopy 33
2.4.8 Mammalian cell maintenance and generation of clonal cell lines 34
2.4.9 Cytotoxicity assay 35
2.4.10 SDS-PAGE and western blots 35
2.4.11 Electrophysiology 36
2.4.12 Protein purification 39
2.4.13 Proteo-liposomes 40
2.4.14 Synthetic planar lipid bilayers 40

2.5 ACKNOWLEDGMENTS 42

2.6 FIGURES 43

Fig. 2.1. PMPL-1 is necessary for PEEL-1 toxicity in *C. elegans*. 43
Fig. 2.2. PEEL-1 and PMPL-1 are sufficient for toxicity in HEK293T cells. 45
Fig. 2.3. PEEL-1 toxicity results in HEK293T cell swelling. 47
Fig. 2.4. PEEL-1's amphipathic helix is critical for toxicity. 49
Fig. 2.5. PEEL-1 and PMPL-1 create a monovalent cation channel. 51

2.7 SUPPLEMENTARY MATERIAL 53

Fig. 2.S1. *pmpl-1(yak103)* does not act through paternal-effect 53
Fig. 2.S2. *pmpl-1* expression pattern in C. elegans. 54
Fig. 2.S3. PMPL-1 is conserved in nematodes. 56
Fig. 2.S4. 15 PMP3-like proteins in C. elegans. 57

Fig. 2.S5. Structural predictions of PEEL-1 and PMPL-1. 58

Fig. 2.S6. PEEL-1 and PMPL-1 localization. 59

Fig. 2.S7. Toxicity causes ER swelling, ER fragmentation, and cell lysis. 60

Fig. 2.S8. Ectopic PEEL-1 toxicity is non-apoptotic. 61

Fig. 2.S9. Amphipathic helix properties of Actinoporin toxins. 62

Fig. 2.S10. Toxicity of PEEL-1 C-terminal truncations. 63

Fig. 2.S11. Raw traces of HEK293 electrophysiology experiments. 64

Fig. 2.S12. Tetracycline-induced toxicity in stable cell lines. 66

Fig. 2.S13. Ion permeabilities in uninduced experimental cells. 68

Fig. 2.S14. Predicted PEEL-1 pentameric structure. 69

Fig. 2.S15. PEEL-1 and PMPL-1 purification and liposomes. 70

Fig. 2.S16. PEEL-1 is toxic to bacteria and requires the amphipathic helix. 71

Fig. 2.S17. Purified PEEL-1 and PMPL-1 conduct ions through planar lipid bilayers. 72

Table 2.S1. *C. elegans* strains used in this study. 74

Table 2.S2. Constructs used in this study. 75

2.8 REFERENCES 76

Chapter 3. RECURRENT BUT SHORT-LIVED DUPLICATIONS OF CENTROMERIC PROTEINS IN HOLOCENTRIC CAENORHABDITIS SPECIES 82

3.1 INTRODUCTION 82

3.2 RESULTS 86

3.2.1 hcp-3 has duplicated at least ten independent times in Caenorhabditis 86

3.2.2 Motif retention and loss in the N-terminal region of HCP-3 and HCP-3L proteins 91

3.2.3 Selective constraints on hcp-3 orthologs and hcp-3L paralogs 95

3.2.4 Duplication of other centromere-localized proteins in Caenorhabditis species.. 98

3.3 DISCUSSION 101

3.4 METHODS 109

3.4.1 Strain maintenance 109

3.4.2 Strains used 109

3.4.3 Identification of hcp-3 and kinetochore protein homologs in sequenced genomes 110

3.4.4 Phylogenetic Analyses 111

3.4.5 Motif Analyses 112

3.4.6 Analysis of evolutionary selective pressures 113

3.4.7 C. elegans HCP-3 and CPAR-1 polymorphisms 114

3.4.8 RT-PCR 114

3.5 ACKNOWLEDGMENTS 115

3.6 FIGURES 117

Figure 3.1. Ten independent *hcp-3* duplications in Caenorhabditis species. 117

Figure 3.2. Phylogenetic analysis of *hcp-3* and *hcp-3L* genes in *Caenorhabditis* species. .119

Figure 3.3. *hcp-3L* genes are expressed in both sexes in Caenorhabditis species. 121

Figure 3.4. Differential retention of N-terminal tail motifs across HCP-3 and HCP-3L proteins encoded by Caenorhabditis species. 122

Figure 3.5. Two unusual Caenorhabditis *hcp-3L* paralogs arose by internal duplication or gene fusion. 124

Figure 3.6. Duplication of kinetochore proteins in Caenorhabditis species. 126

3.7 SUPPLEMENTARY MATERIALS 128

Supplementary Figure 3.S1. Exon-intron junctions are largely retained in *Caenorhabditis hcp-3* and *hcp-3L* genes except for intron gains and losses that are likely a result of partial retrotransposition. 128

Supplementary Figure 3.S2. Maximum-likelihood phylogenetic tree based on an amino acid alignment of HCP-3 and HCP-3L proteins encoded by Caenorhabditis species. 130

Supplementary Figure 3.S3. RT-PCR controls for expression analysis of *hcp-3* and *hcp-3L* genes. 132

Supplementary Figure 3.S4. Conserved motifs identified in the HCP-3 histone fold domains. 133

Supplementary Figure 3.S5. The ExxR putative separase cleavage site is retained in most Caenorhabditis HCP-3 and HCP-3L proteins. 134

Supplementary Figure 3.S6. A higher frequency of function-altering mutations is observed in *cpar-1* compared to *hcp-3* within natural C. elegans populations. 135

Supplementary Figure 3.S7. Other kinetochore genes also display patterns of partial intron loss in Caenorhabditis species. 136

Supplementary Data 3.S1: Pairwise alignments of Caenorhabditis HCP-3 HFD and N-tail used for Figure 1 138

Supplementary Data 3.S2: HCP-3 histone-fold domain codon based-alignments used for Figure 2 146

Supplementary Data 3.S3: All HCP-3 sequences used in this study 154

Supplementary Data 3.S4: HCP-3 protein motifs in C. elegans and C. sp54 found through manual alignments with sister species 162

Supplementary Data 3.S5: All HCP-4 sequences used in this study 163

Supplementary Data 3.S6: All KNL-2 sequences used in this study 176

Supplementary Data 3.S7: All KNL-1 sequences used in this study 189

Supplementary Data 3.S8: All ZWL-1 sequences used in this study 215

Supplementary Data 3.S9: All SPDL-1 sequences used in this study 224

Supplementary Data 3.S10: All NDC-80 sequences used in this study 232

Supplementary Data 3.S11: All HIM-10 sequences used in this study 250

3.8 REFERENCES 258

Chapter 4. FUTURE DIRECTIONS 273

4.1 MECHANISM OF PEEL-1 TOXICITY 273

4.1.1 Do PEEL-1 and PMPL-1 interact? 274

Table 4.1. Results and interpretations from mbSUS for PEEL-1 and PMPL-1 interaction 275

4.1.2 Does PEEL-1 oligomerize? 277

4.1.3 Does PMPL-1 oligomerize? 278

4.1.4 Which residues are critical for PMPL-1's role in toxicity? 279

4.1.5 An alternative approach for assaying the PEEL-1/PMPL-1 ion channel 280

Table 4.2. Interpretations and counter arguments explaining results from liposome flux assays. 281

4.1.6 Determine important residues in PEEL-1 and PMPL-1 282

4.1.7 Determine the pore-lining residues of PEEL-1 283

4.2 MECHANISM OF ZEEL-1 ANTIDOTE ACTIVITY 284

4.2.1 Do PEEL-1 and ZEEL-1 interact? 284

Table 4.3. Results and interpretations from mbSUS for ZEEL-1 interactions. 285

4.2.2 Does ZEEL-1 ubiquitylate PEEL-1 for degradation? 285

4.2.3 Deep mutational scan to determine interacting residues between PEEL-1 and ZEEL-1 286

4.3 EVOLUTION OF *PEEL-1 ZEEL-1*..... 286

4.3.1 Where did *peel-1 zeel-1* come from? 287

4.3.2 Why does PEEL-1 rely on another protein for toxicity? 287

4.4 METHODS 288

4.4.1 Mating-based split ubiquitin system (mbSUS) 288

4.4.2 Bimolecular fluorescence complementation (BiFC) 290

4.4.3 Mass photometry 290

4.4.4 Liposome flux assay 291

4.5 FIGURES..... 293

Figure 4.1. mbSUS assaying PEEL-1 and PMPL-1 interactions. 293

Figure 4.2. BiFC assaying PEEL-1 and PMPL-1 interactions. 295

Figure 4.3. mbSUS for homotypic PEEL-1 and PMPL-1 interactions. 296

Figure 4.4. PEEL-1 forms SDS-resistant oligomers. 297

Figure 4.5. PMPL-1 forms SDS-resistant oligomers. 299

Figure 4.6. Chimeric PMP3-like proteins in HEK293T cells. 300

Figure 4.7. Liposome flux assay to measure ion channel activity. 302

Figure 4.8. mbSUS assaying ZEEL-1 interactions with PEEL-1 and PMPL-1. 303

4.6 REFERENCES..... 304

ACKNOWLEDGEMENTS

Thank you to my wonderful family for their love and support. My mom, Bertha Lucia Caro, brightens up my days with her wit and humor. My dad, Guillermo Alexander Caro, is unmatched in his resourcefulness, positivity, and go-getter attitude. I try to emulate him in my approach to science. My older brother, Lennin Caro, has been both my teacher and my teammate throughout my academic career. My cousins, Kevin Amarillo, Sabrina Galeano, and Marthaluz Velez create a safe space for me and are always a source a comfort. I also thank my chosen families in Seattle, the Zoo Squad, Jordan Yerzy, the MUPS, and the many iterations of the MCB volleyball team.

Thank you to my mentor, Michael Ailion, for his unwavering confidence in me and his dedication to keeping science fun, accessible, and weird. I thank the many folks who also mentored me, including Irini Topalidou, Piero Lamelza, Janet M. Young, Pravrutha Raman, Aguan D. Wei, and Harmit S. Malik. Thank you to everyone in the University of Washington Department of Biochemistry and in the Malik lab at the Fred Hutch Cancer Center. It has been incredibly fun and challenging to navigate through many disciplines of biology, and I never felt judged for asking the basic questions. I also thank my committee members, Barbara Wakimoto, Robert H. Waterston, Maitreya J. Dunham, and Adam Steinbrenner for their contributions to my work and training.

DEDICATION

This thesis is dedicated to my parents, Bertha Lucia Caro and Guillermo Alexander Caro.

Los quiero mucho.

Chapter 1. INTRODUCTION

A major goal of my graduate studies was to gain a mechanistic understanding of evolutionarily unique proteins. This led to my primary project, investigating the mechanism of a selfish toxin-antidote system from *C. elegans*: *peel-1 zeel-1*. Here, I focused on the mechanism of PEEL-1 toxicity. I found that PEEL-1 toxicity requires a second protein, PMPL-1. By co-expressing PEEL-1 and PMPL-1 in a heterologous system, I demonstrated that these two proteins are sufficient for toxicity and act by creating a novel cation leak channel. This work is contextualized in Section 1.1, where I discuss the current mechanistic understanding of several animal toxin-antidote systems. I present my findings on the PEEL-1 mechanism in Chapter 2. More preliminary data on the mechanism and evolution of *peel-1 zeel-1* are in Chapter 4.

My second project was on centromeric histones, an essential protein which counterintuitively undergoes rapid evolution. I characterized the evolution of centromeric histones in holocentric nematodes and found multiple, short-lived duplications of these proteins. I identified novel motifs in nematode centromeric histones that act as putative sites of protein-protein interactions. This work is presented in Chapter 3.

1.1 ANIMAL TOXIN-ANTIDOTE SYSTEMS AND THEIR MECHANISMS

"Natural selection", the process of favorable traits increasing in frequency over generations, is a pillar of evolutionary biology *(Darwin 1859)*. On the surface, this implies that only genes which encode these favorable traits can spread in populations. However, selfish genetic elements (SGEs) spread in populations by biasing their inheritance into progeny, even at

the expense of host fitness *(Lindholm et al., 2016, Werren 2011)*. SGEs come in many shapes and sizes. Some of the smallest SGEs are mobile genetic elements like transposons, which over replicate in genomes by copy-paste or cut-paste mechanisms *(McClintock 1950, Shapiro 1969)*. Larger SGEs such as centromeres can consist of repetitive, non-coding DNA that manipulates asymmetrical female meiosis to favor their inheritance *(Pardo-Manuel de Villena and Sapienza 2001, Henikoff et al., 2001, Melters et al., 2013, Lampson and Black 2018)*. Regardless of their size or mechanism, SGEs are pieces of DNA that spread in populations without necessarily providing a fitness benefit to hosts.

Toxin-antidote (TA) systems are an aggressive SGE that often guarantee their inheritance by killing non-inheriting progeny. TAs consist of two linked genes: a toxin expressed in the parent and an antidote expressed in the progeny *(Burga et al., 2020)*. The parental toxin is cytoplasmically transmitted to progeny; in animals, the toxin is inherited through the sperm or egg. This toxin causes death or developmental delay of progeny unless they inherit the TA, which provides the antidote *(Burga et al., 2020)*. The tight linkage between the toxin and antidote genes ensure that they are almost always inherited together, acting as one genetic element. Overall, TA elements selectively poison non-inheritors in order to increase their own representation in the next generation.

TA systems were first discovered in bacteria from so-called plasmid addiction systems. Here, a plasmid ensures its inheritance by encoding a stable toxin and a less-stable antidote. This differential stability is key for bacterial TAs. The toxin's stability allows its cytoplasmic inheritance while the antidote is quickly degraded *(Aizenman et al., 1996, Van-Melderen et al., 1996)*. This differential stability ensures that only daughter cells lacking the TA plasmid are affected by the toxin. Interestingly, bacterial TA systems are not exclusively present in plasmids.

Many bacterial TA systems exist within the chromosomal DNA. These genomic TA systems are not selfish but instead are important for cell persistence in unfavorable environments *(Hayes 2003, Fineran et al., 2009, Labrie et al., 2010)*. Mechanisms of toxicity from both plasmid and chromosomal TAs include inhibition of gyrase *(Bernard and Couturier et al., 1992; Jiang et al., 2002)*, mRNA cleavage *(Christensen et al, 2003; Zhang et al., 2003; Pedersen et al., 2003)*, and loss of membrane potential *(Wilmaerts et al., 2018)*. Meanwhile, antidotes can act by directly binding the toxin protein *(Dao-Thi et al., 2002, Madl et al., 2006, Kamada et al., 2003)* or inhibiting expression of the toxin *(Gerdes et al., 1992)*. Bacterial TAs are by far the best understood TA systems; much less is known about eukaryotic TA systems.

Many fungal TA systems act during meiosis to kill non-inheriting spores *(Bravo Núñez et al., 2018)*. Some *wtf* TAs from *Schizosaccharomyces pombe* encode two isoforms of the same gene *(Zanders et al., 2014, Hu et al., 2017, Nuckolls et al., 2017)*. The resulting toxin and antidote proteins are largely similar, but the antidote has an extended N-terminus *(Nuckolls et al., 2017)*. The mechanism of toxicity is unknown, but the toxin either aggregates or localizes to the Golgi *(Nuckolls et al., 2020, Zheng et al., 2023)*. The antidote can interact with the toxin, re-localizing the protein and preventing its toxicity *(Nuckolls et al., 2020, Zheng et al., 2023)*. There are many *wtf* genes in *Schizosaccharomyces* species, however, and it is not known whether they all work through the same mechanism. Other fungal TAs such as *spok (Grognet et al., 2014)* and *spk-1 (Turner and Perkins, 1979, Svedberg et al., 2021)* are also encoded by single genes, but how these systems work is unclear.

TA elements were once thought to be rare in animals, but this idea is beginning to be challenged. About a dozen animal TAs have been discovered *(Beeman et al. 1992, Seidel et al., 2008, Ben-David et al., 2017, Ben-David et al. 2021, Noble et al., 2021)*, and almost all of these

are in hermaphroditic species of *Caenorhabditis* nematodes. All known animal TA systems appear to be segregating within species, partly because the approaches used to identify new TAs require hybridization, biasing for these segregating TAs. Nevertheless, the fact that animal TAs naturally segregate within species demonstrates that TAs are a source of intraspecific hybrid sub-fertility and can therefore form genetic barriers within species.

Some of the best understood animal TAs are the Medea element in the red flour beetle, *peel-1 zeel-1* in *C. elegans*, *sup-35 pha-1* in *C. elegans*, and *slow-1 grow-1* in *C. tropicalis*. I summarize below the current mechanistic understanding of these TAs.

Medea element from T. castaneum

The Medea element (maternal-effect dominant embryonic arrest) from the red flour beetle, *Tribolium castaneum*, was the first TA system discovered in animals. Medea was found through hybrid crosses between beetle strains from the United States and Singapore *(Beeman et al. 1992)*. While hybrid F1 progeny developed normally, crossing these hybrids together resulted in lethality in the F2. In backcross experiments, high lethality was only seen when hybrid F1 females were crossed to US males, but not Singapore males. Furthermore, this lethality was not seen when hybrid males were backcrossed to either parent. These crosses indicated the presence of a dominant, maternal-effect genetic system that acted by "self-selection" *(Beeman et al. 1992)*. Despite being the first animal TA discovered, the molecular components of this TA system are not known. The candidate region for the Medea genes has been mapped to a 100kb region containing a Tc1 transposon *(Lorenzen et al., 2008)*, pseudogenes, and at least three intact protein-coding genes *(Lorenzen et al., 2008)*. Among these genes is a gene encoding a protein domain previously only found in bacteria *(Knizewski et al., 2007)*. Another is a gene homologous

to *blot*. In fruit flies, *blot* has both maternal and zygotic functions *(Lorenzen et al., 2008, Johnson et al., 1999)*, making it an interesting candidate gene for this TA.

sup-35 pha-1 from C. elegans

The *sup-35 pha-1* TA system consists of the SUP-35 toxin and the PHA-1 antidote. This TA has a unique story of discovery because PHA-1 was thought to be a master regulator of development for several decades *(Schnabel and Schnabel 1990)* before it was recognized as a TA element *(Ben-David et al., 2017)*. *pha-1* mutants generated in the lab were found to have pharyngeal defects that result in death, so *pha-1* was thought to regulate the development of the pharynx *(Schnabel and Schnabel 1990)* and other tissues *(Kuzmanov et al., 2014)*. However, a *C. elegans* strain from Hawaii (DL238) lacked this gene entirely, casting doubt on the essential role of *pha-1* in development. This inconsistency led to the discovery that PHA-1 was an antidote for a maternal toxin, SUP-35 *(Ben-David et al., 2017)*. The *sup-35 pha-1* genes shared expected features of a toxin-antidote system: the genes are tightly linked, the toxin is cytoplasmically inherited through the mother, and zygotic expression of the antidote prevents toxicity. The genetics of the *sup-35 pha*-1 TA system have been well-studied, and two genes are required for toxicity: *sup-36* and *sup-37 (Schnabel et al., 1991)*. However, the precise molecular functions of the toxin and antidote are not yet understood.

slow-1 grow-1 from C. tropicalis

The recently described *slow-1 grow-1* TA system from *C. tropicalis (Ben-David et al. 2021)*, adds a new layer of complexity to animal TAs. Expression of the *slow-1* toxin is controlled by imprinting, also known as a parent-of-origin effect *(Pilota et al., 2024)*. Maternal

inheritance of *slow-1 grow-1* elicited expected selfish activity, resulting in overrepresentation of *slow-1 grow-1* in the F2 generation. However, paternal inheritance of the TA caused absence of toxin activity in the next generation. In other words, the activity of the TA is dependent on the genotype of the grandparents, parents, and progeny. Both *slow-1* mRNA and SLOW-1 protein are maternally loaded into embryos, but the parent-of-origin effect is specifically attributed to the maternal mRNA *(Pilota et al., 2024)*. Progeny affected by SLOW-1 do not die but instead experience a developmental delay *(Ben-David et al. 2021)*. SLOW-1 is homologous to a nuclear hormone receptor and contains two predicted transmembrane domains. Meanwhile, GROW-1 is evolutionarily novel *(Ben-David et al. 2021)*. The mechanism of toxin and antidote activity are unknown.

peel-1 zeel-1 from C. elegans

The *peel-1 zeel-1* TA is currently the best characterized animal TA system and is also the only known TA to act through a paternal, sperm-delivered toxin. *peel-1* codes for a toxin and *zeel-1* codes for an antidote. This TA was discovered in hybrid crosses between the *C. elegans* N2 lab strain and a Hawaiian strain (CB4856) *(Seidel et al., 2008)*. When F1 hybrid males are backcrossed to CB4856 hermaphrodites, about half of the progeny arrest during development. However, backcrosses to N2 did not result in lethality. Furthermore, this phenotype was specific to hybrid males since backcrossing hermaphrodite hybrids to parental males resulted in minimal F2 lethality. This inheritance pattern fits a model of paternal, cytoplasmic inheritance of a toxin that requires a linked, zygotic antidote for viability. The antidote gene, *zeel-1*, was identified through transgenic rescue of antidote activity. It took three years to molecularly identify the toxin gene, *peel-1*, partly because it was unannotated in the genome. Natural *C. elegans* strains that

had antidote activity but lacked toxin activity allowed identification of the *peel-1* gene *(Seidel et al, 2011)*.

The *peel-1* gene codes for an evolutionarily novel, toxic protein. PEEL-1 is 174 amino-acids and predicted to contain four transmembrane domains. PEEL-1 is exclusively expressed in the male germline and is packaged in sperm-specific organelles (called fibrous body membranous organelles) by an N-terminal packaging sequence. After fertilization, PEEL-1 is presumably delivered to the zygote plasma membrane during fusion between the sperm and egg, but the actual location of PEEL-1 within embryos has not been determined. Toxicity from sperm-delivered PEEL-1 manifests in the 2-fold stage of developing embryos, with visible defects in the muscle, epidermis, and excretory cell *(Seidel et al., 2011)*. Presumably, sperm-delivered PEEL-1 ends up in these tissues since ectopic expression in each tissue causes similar phenotypes. Ectopic expression in adults causes several cellular phenotypes, including necrotic vacuoles, swollen cells, and disintegrated tissues *(Seidel et al., 2011)*. Ectopic expression also showed that PEEL-1 toxicity is cell-autonomous, meaning that toxicity occurs in the cell containing PEEL-1. So far, all adult somatic tissues appear to be susceptible to PEEL-1 toxicity. PEEL-1 may therefore disrupt some fundamental cellular process, but the precise mechanism of PEEL-1 toxicity is not known. Possible toxin mechanisms include inhibition of the sodium-potassium pump, pore formation, ion channel formation, or physical disruption of the plasma membrane.

The antidote, ZEEL-1, prevents toxicity from PEEL-1. ZEEL-1 is a predicted six-pass transmembrane protein, with a large soluble domain *(Seidel et al., 2011)*. The transmembrane region is lineage-specific, containing no homology outside of *C. elegans*. Meanwhile, the soluble domain is homologous to ZYG-11, a conserved substrate-recognition subunit of an E3 ubiquitin

ligase *(Seidel et al., 2011)*. As expected, *zeel-1* is expressed before PEEL-1 toxicity. Antidote activity is also tissue-specific; for example, ZEEL-1 expression in the epidermis prevents PEEL-1 toxicity in this specific tissue *(Seidel et al., 2011)*. This expression pattern and tissue-specific antidote activity leads to a model of ZEEL-1 interacting with PEEL-1, causing PEEL-1 ubiquitylation and degradation. However, this hypothesis has not been directly tested.

Other animal toxin-antidote systems

In addition to the TAs described above, other animal TAs have been studied including *msft-1* from *C. briggsae (Widen et al., 2023)*, *mll-1 smll-1* from *C. elegans (Zdraljevic et al., 2024)*, and unnamed TAs from *C. tropicalis (Ben-David et al. 2021, Noble et al., 2021)*. Investigations into animal TAs have shed light on other aspects of biology, including providing insight into possible origins of imprinting *(Pilota et al., 2024)*, horizontal gene transfer in eukaryotes *(Widen et al., 2023)*, and paternal contributions to embryonic development *(Seidel et al., 2011)*. Like chromosomal TA systems in bacteria that benefit hosts, animal TAs have possibly become domesticated by host genomes. Recent work suggests that *peel-1(+)* worms have a slight fitness advantage when competed against *peel-1(-)* worms in laboratory conditions *(Long et al., 2023)*. Feasibly, TA systems could arise and spread via their selfish activity and then become domesticated by hosts, eliciting some fitness benefits. Since some animal TAs code for evolutionarily novel proteins, TA element domestication might be a path for novel cellular functions, as seen in bacteria *(Hayes 2003, Fineran et al., 2009, Labrie et al., 2010)*. Therefore, Studying the mechanisms of animal TA systems can therefore provide insights into new biology.

1.2 REFERENCES

Aizenman, E., Engelberg-Kulka, H., Glaser, G., 1996. An Escherichia coli chromosomal "addiction module" regulated by guanosine [corrected] 3',5'-bispyrophosphate: a model for programmed bacterial cell death. Proc. Natl. Acad. Sci. U. S. A. 93, 6059–6063. https://doi.org/10.1073/pnas.93.12.6059

Beeman, R.W., Friesen, K.S., Denell, R.E., 1992. Maternal-effect selfish genes in flour beetles. Science 256, 89–92. https://doi.org/10.1126/science.1566060

Ben-David, E., Burga, A., Kruglyak, L., 2017. A maternal-effect selfish genetic element in Caenorhabditis elegans. Science 356, 1051–1055. https://doi.org/10.1126/science.aan0621

Ben-David, E., Pliota, P., Widen, S.A., Koreshova, A., Lemus-Vergara, T., Verpukhovskiy, P., Mandali, S., Braendle, C., Burga, A., Kruglyak, L., 2021. Ubiquitous Selfish Toxin-Antidote Elements in Caenorhabditis Species. Curr. Biol. 31, 990-1001.e5. https://doi.org/10.1016/j.cub.2020.12.013

Bernard, P., Couturier, M., 1992. Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. J. Mol. Biol. 226, 735–745. https://doi.org/10.1016/0022-2836(92)90629-X

Burga, A., Ben-David, E., Kruglyak, L., 2020. Toxin-Antidote Elements Across the Tree of Life. Annu. Rev. Genet. 54, 387–415. https://doi.org/10.1146/annurev-genet-112618-043659

Christensen, S.K., Pedersen, K., Hansen, F.G., Gerdes, K., 2003. Toxin–antitoxin Loci as Stress-response-elements: ChpAK/MazF and ChpBK Cleave Translated RNAs and are Counteracted by tmRNA. J. Mol. Biol. 332, 809–819. https://doi.org/10.1016/S0022-2836(03)00922-7

Dao-Thi, M.-H., Charlier, D., Loris, R., Maes, D., Messens, J., Wyns, L., Backmann, J., 2002. Intricate Interactions within the ccd Plasmid Addiction System*. J. Biol. Chem. 277, 3733–3742. https://doi.org/10.1074/jbc.M105505200

Darwin, C., 1859. On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life. London : John Murray, 1859.

Fineran, P.C., Blower, T.R., Foulds, I.J., Humphreys, D.P., Lilley, K.S., Salmond, G.P.C., 2009. The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. Proc. Natl. Acad. Sci. U. S. A. 106, 894–899. https://doi.org/10.1073/pnas.0808832106

Gerdes, K., Nielsen, A., Thorsted, P., Wagner, E.G.H., 1992. Mechanism of killer gene activation. Antisense RNA-dependent RNase III cleavage ensures rapid turn-over of the stable Hok, SrnB and PndA effector messenger RNAs. J. Mol. Biol. 226, 637–649. https://doi.org/10.1016/0022-2836(92)90621-P

Grognet, P., Lalucque, H., Malagnac, F., Silar, P., 2014. Genes that bias Mendelian segregation. PLoS Genet. 10, e1004387. https://doi.org/10.1371/journal.pgen.1004387

Hayes, F., 2003. Toxins-Antitoxins: Plasmid Maintenance, Programmed Cell Death, and Cell Cycle Arrest. Science 301, 1496–1499. https://doi.org/10.1126/science.1088157

Henikoff, S., Ahmad, K., Malik, H.S., 2001. The centromere paradox: stable inheritance with rapidly evolving DNA. Science 293, 1098–1102. https://doi.org/10.1126/science.1062939

Hu, W., Jiang, Z.-D., Suo, F., Zheng, J.-X., He, W.-Z., Du, L.-L., 2017. A large gene family in fission yeast encodes spore killers that subvert Mendel's law. eLife 6, e26057. https://doi.org/10.7554/eLife.26057

Jiang, Y., Pogliano, J., Helinski, D.R., Konieczny, I., 2002. ParE toxin encoded by the broad-host-range plasmid RK2 is an inhibitor of Escherichia coli gyrase. Mol. Microbiol. 44, 971–979. https://doi.org/10.1046/j.1365-2958.2002.02921.x

Johnson, K., Knust, E., Skaer, H., 1999. bloated tubules (blot) Encodes a Drosophila Member of the Neurotransmitter Transporter Family Required for Organisation of the Apical Cytocortex. Dev. Biol. 212, 440–454. https://doi.org/10.1006/dbio.1999.9351

Kamada, K., Hanaoka, F., Burley, S.K., 2003. Crystal structure of the MazE/MazF complex: molecular bases of antidote-toxin recognition. Mol. Cell 11, 875–884. https://doi.org/10.1016/s1097-2765(03)00097-2

Knizewski, L., Kinch, L.N., Grishin, N.V., Rychlewski, L., Ginalski, K., 2007. Realm of PD-(D/E)XK nuclease superfamily revisited: detection of novel families with modified transitive meta profile searches. BMC Struct. Biol. 7, 40. https://doi.org/10.1186/1472-6807-7-40

Kuzmanov, A., Yochem, J., Fay, D.S., 2014. Analysis of PHA-1 Reveals a Limited Role in Pharyngeal Development and Novel Functions in Other Tissues. Genetics 198, 259–268. https://doi.org/10.1534/genetics.114.166876

Labrie, S.J., Samson, J.E., Moineau, S., 2010. Bacteriophage resistance mechanisms. Nat. Rev. Microbiol. 8, 317–327. https://doi.org/10.1038/nrmicro2315

Lampson, M.A., Black, B.E., 2017. Cellular and Molecular Mechanisms of Centromere Drive. Cold Spring Harb. Symp. Quant. Biol. 82, 249–257. https://doi.org/10.1101/sqb.2017.82.034298

Lindholm, A.K., Dyer, K.A., Firman, R.C., Fishman, L., Forstmeier, W., Holman, L., Johannesson, H., Knief, U., Kokko, H., Larracuente, A.M., Manser, A., Montchamp-Moreau, C., Petrosyan, V.G., Pomiankowski, A., Presgraves, D.C., Safronova, L.D., Sutter, A., Unckless, R.L., Verspoor, R.L., Wedell, N., Wilkinson, G.S., Price, T.A.R., 2016. The Ecology and Evolutionary Dynamics of Meiotic Drive. Trends Ecol. Evol. 31, 315–326. https://doi.org/10.1016/j.tree.2016.02.001

Long, L., Xu, W., Valencia, F., Paaby, A.B., McGrath, P.T., 2023. A toxin-antidote selfish element increases fitness of its host. eLife 12, e81640. https://doi.org/10.7554/eLife.81640

Lorenzen, M.D., Gnirke, A., Margolis, J., Garnes, J., Campbell, M., Stuart, J.J., Aggarwal, R., Richards, S., Park, Y., Beeman, R.W., 2008. The maternal-effect, selfish genetic element Medea is associated with a composite Tc1 transposon. Proc. Natl. Acad. Sci. U. S. A. 105, 10085–10089. https://doi.org/10.1073/pnas.0800444105

Madl, T., Van Melderen, L., Mine, N., Respondek, M., Oberer, M., Keller, W., Khatai, L., Zangger, K., 2006. Structural Basis for Nucleic Acid and Toxin Recognition of the Bacterial Antitoxin CcdA. J. Mol. Biol. 364, 170–185. https://doi.org/10.1016/j.jmb.2006.08.082

McClintock, B., 1950. The origin and behavior of mutable loci in maize. Proc. Natl. Acad. Sci. U. S. A. 36, 344–355. https://doi.org/10.1073/pnas.36.6.344

Noble, L.M., Yuen, J., Stevens, L., Moya, N., Persaud, R., Moscatelli, M., Jackson, J.L., Zhang, G., Chitrakar, R., Baugh, L.R., Braendle, C., Andersen, E.C., Seidel, H.S., Rockman, M.V., 2021. Selfing is the safest sex for Caenorhabditis tropicalis. eLife 10, e62587. https://doi.org/10.7554/eLife.62587

Nuckolls, N.L., Bravo Núñez, M.A., Eickbush, M.T., Young, J.M., Lange, J.J., Yu, J.S., Smith, G.R., Jaspersen, S.L., Malik, H.S., Zanders, S.E., 2017. wtf genes are prolific dual poison-antidote meiotic drivers. eLife 6, e26033. https://doi.org/10.7554/eLife.26033

Nuckolls, N.L., Mok, A.C., Lange, J.J., Yi, K., Kandola, T.S., Hunn, A.M., McCroskey, S., Snyder, J.L., Bravo Núñez, M.A., McClain, M., McKinney, S.A., Wood, C., Halfmann, R., Zanders, S.E., 2020. The wtf4 meiotic driver utilizes controlled protein aggregation to generate selective cell death. eLife 9, e55694. https://doi.org/10.7554/eLife.55694

Núñez, M.A.B., Nuckolls, N.L., Zanders, S.E., 2018. Genetic Villains: Killer Meiotic Drivers. Trends Genet. 34, 424–433. https://doi.org/10.1016/j.tig.2018.02.003

Pardo-Manuel de Villena, F., Sapienza, C., 2001. Nonrandom segregation during meiosis: the unfairness of females. Mamm. Genome Off. J. Int. Mamm. Genome Soc. 12, 331–339. https://doi.org/10.1007/s003350040003

Pedersen, K., Zavialov, A.V., Pavlov, M.Y., Elf, J., Gerdes, K., Ehrenberg, M., 2003. The bacterial toxin RelE displays codon-specific cleavage of mRNAs in the ribosomal A site. Cell 112, 131–140. https://doi.org/10.1016/s0092-8674(02)01248-5

Pilota, P., Marvanova H., Koreshov, A., Tikanova, P., Krogull, D., Hagmüller A., Widen, S.A., Handler, D., Gokcezade, J., Duchek, P., Brennecke, J., Ben-David, E., Burga, A., 2024. Nature 628, 122-129. https://doi.org/10.1038/s41586-024-07155-z

Schnabel, H., Bauer, G., Schnabel, R., 1991. Suppressors of the organ-specific differentiation gene pha-1 of Caenorhabditis elegans. Genetics 129, 69–77. https://doi.org/10.1093/genetics/129.1.69

Schnabel, H., Schnabel, R., 1990. An Organ-Specific Differentiation Gene, pha-1, from Caenorhabditis elegans. Science 250, 686–688. https://doi.org/10.1126/science.250.4981.686

Seidel, H.S., Ailion, M., Li, J., van Oudenaarden, A., Rockman, M.V., Kruglyak, L., 2011. A novel sperm-delivered toxin causes late-stage embryo lethality and transmission ratio distortion in C. elegans. PLoS Biol. 9, e1001115. https://doi.org/10.1371/journal.pbio.1001115

Seidel, H.S., Rockman, M.V., Kruglyak, L., 2008. Widespread genetic incompatibility in C. elegans maintained by balancing selection. Science 319, 589–594. https://doi.org/10.1126/science.1151107

Shapiro, J.A., 1969. Mutations caused by the insertion of genetic material into the galactose operon of Escherichia coli. J. Mol. Biol. 40, 93–105. https://doi.org/10.1016/0022-2836(69)90298-8

Svedberg, J., Vogan, A.A., Rhoades, N.A., Sarmarajeewa, D., Jacobson, D.J., Lascoux, M., Hammond, T.M., Johannesson, H., 2021. An introgressed gene causes meiotic drive in Neurospora sitophila. Proc. Natl. Acad. Sci. U. S. A. 118, e2026605118. https://doi.org/10.1073/pnas.2026605118

Turner, B.C., Perkins, D.D., 1979. Spore killer, a chromosomal factor in neurospora that kills meiotic products not containing it. Genetics 93, 587–606. https://doi.org/10.1093/genetics/93.3.587

Van Melderen, L., Thi, M.H., Lecchi, P., Gottesman, S., Couturier, M., Maurizi, M.R., 1996. ATP-dependent degradation of CcdA by Lon protease. Effects of secondary structure and heterologous subunit interactions. J. Biol. Chem. 271, 27730–27738. https://doi.org/10.1074/jbc.271.44.27730

Werren, J.H., 2011. Selfish genetic elements, genetic conflict, and evolutionary innovation. Proc. Natl. Acad. Sci. 108, 10863–10870. https://doi.org/10.1073/pnas.1102343108

Widen, S.A., Bes, I.C., Koreshova, A., Pliota, P., Krogull, D., Burga, A., 2023. Virus-like transposons cross the species barrier and drive the evolution of genetic incompatibilities. Science 380, eade0705. https://doi.org/10.1126/science.ade0705

Wilmaerts, D., Bayoumi, M., Dewachter, L., Knapen, W., Mika, J.T., Hofkens, J., Dedecker, P., Maglia, G., Verstraeten, N., Michiels, J., 2018. The Persistence-Inducing Toxin HokB Forms Dynamic Pores That Cause ATP Leakage. mBio 9, 10.1128/mbio.00744-18. https://doi.org/10.1128/mbio.00744-18

Zanders, S.E., Eickbush, M.T., Yu, J.S., Kang, J.-W., Fowler, K.R., Smith, G.R., Malik, H.S., 2014. Genome rearrangements and pervasive meiotic drive cause hybrid infertility in fission yeast. eLife 3, e02630. https://doi.org/10.7554/eLife.02630

Zdraljevic, S., Walter-McNeill, L., Bruni, G., Bloom, J.S., Leighton, D.H.W., Marquez, H., Kruglyak, L., 2024. A maternal-effect toxin-antidote element causes larval arrest in C. elegans. https://doi.org/10.1101/2024.04.26.591160

Zhang, Y., Zhang, J., Hoeflich, K.P., Ikura, M., Qing, G., Inouye, M., 2003. MazF cleaves cellular mRNAs specifically at ACA to block protein synthesis in Escherichia coli. Mol. Cell 12, 913–923. https://doi.org/10.1016/s1097-2765(03)00402-7

Zheng, J.-X., Du, T.-Y., Shao, G.-C., Ma, Z.-H., Jiang, Z.-D., Hu, W., Suo, F., He, W., Dong, M.-Q., Du, L.-L., 2023. Ubiquitination-mediated Golgi-to-endosome sorting determines the toxin-antidote duality of fission yeast wtf meiotic drivers. Nat. Commun. 14, 8334. https://doi.org/10.1038/s41467-023-44151-9

Chapter 2. TOXIN MECHANISM OF THE *PEEL-1 ZEEL-1* TOXIN-ANTIDOTE SYSTEM

The work in this chapter was done in collaboration with Aguan D. Wei (Seattle Children's Research Institute), Christopher A. Thomas (University of Washington), Galen Posch (University of Washington), Ahmet Uremis (University of Washington), Michaela C. Franzi (University of Washington), Sarah J. Abell (University of Washington), Andrew H. Laszlo (University of Washington), Jens H. Gundlach (University of Washington), and Jan-Marino Ramirez (Seattle Children's Research Institute).

2.1 INTRODUCTION

Selfish genetic elements ensure their inheritance, even at the expense of host fitness. Toxin-antidote (TA) systems are one class of selfish element, made up of genetically linked toxin and antidote genes. The toxin is cytoplasmically inherited across generations while the cognate antidote is expressed in progeny. In animal TA systems, the parental toxin is transmitted to progeny via the sperm or egg. Offspring that do not inherit the TA genetic element are affected by the toxin, typically resulting in death or developmental defects (*1*), whereas the offspring which inherit the TA genetic element express the cognate antidote to prevent toxicity. Therefore, TAs guarantee their presence in the next generation by killing non-inheriting offspring. Recent work suggests that TA elements are more common among animals than previously thought (*2*, *3*). However, the mechanisms of toxin and antidote activity remain largely unknown for animal TA systems.

One of the best characterized animal TA systems is *peel-1/zeel-1* in *C. elegans* (*4*). The *peel-1* toxin is expressed during sperm development but is not toxic to sperm. Mature sperm carry PEEL-1 protein and fertilization delivers the toxin, resulting in developmental arrest of embryos (*5*). However, offspring which inherit the TA element evade death by embryonic expression of the antidote, *zeel-1* (Fig. 2.1A). Expression of PEEL-1 in adult worms also causes death (*5*), suggesting toxicity is not specific to a developmental stage. Furthermore, PEEL-1 acts cell-autonomously in *C. elegans*; ectopic expression of PEEL-1 in specific tissues causes death of these cells, with no defects in neighboring cells (*5*). So far, no adult somatic cells have been found to be immune to PEEL-1, suggesting that toxicity works by disrupting a fundamental cellular process. In this study, we dissect the molecular mechanism of PEEL-1 toxicity. We find that PEEL-1 co-opts a conserved membrane protein of unknown function, PMPL-1. Together these proteins are sufficient to create a toxic, cation leak channel. Our findings therefore describe the molecular mechanism of toxicity of an animal toxin-antidote system.

2.2 RESULTS

2.2.1 PMPL-1 is required for PEEL-1 toxicity

To identify other genes required for PEEL-1 toxicity, we performed a large forward genetic screen for PEEL-1 suppressors in *C. elegans*. We mutagenized worms carrying transgenes for heat-shock inducible *peel-1* expression (*hsp-16.41p*::*peel-1*) (*5*). We isolated only two full-suppressors of heat-shock PEEL-1 toxicity, and both suppressors carry mutations in F47B7.1 (hereafter named *pmpl-1*) (Fig 2.1B). Endogenous, sperm-delivered PEEL-1 is also suppressed by *pmpl-1* mutants (Fig. 2.1C), but not via a paternal-effect (fig. 2.S1), suggesting that PMPL-1 acts in embryos to facilitate PEEL-1 toxicity and does not act in sperm. *pmpl-1*

codes for a 59 amino acid protein predicted to be an integral membrane protein (Fig. 2.1D). The *pmpl-1* mutants were identified as a missense allele (*yak52*, A47T) and a full-gene deletion allele (*yak103*) (see Materials and Methods).

The *pmpl-1* expression pattern is consistent with its role in PEEL-1 toxicity. The *pmpl-1* promoter drives expression in embryos prior to toxicity from sperm-delivered PEEL-1 (fig. 2.S2A) (*5*). *pmpl-1* is also widely expressed in several adult tissues (fig. 2.S2B), consistent with heat-shock PEEL-1 toxicity in adults (*5*). Publicly available RNA-seq data (*6*) indicates that *pmpl-1* expression levels are lowest in the male gonad compared to all other tissues, consistent with wild-type sperm being unaffected by PEEL-1 (fig. 2.S2C). These data suggest that PMPL-1 is required for cell susceptibility to PEEL-1 toxicity. We further confirmed that co-expression of PEEL-1 and PMPL-1 in the vulval muscle cells of *pmpl-1* mutants caused specific toxicity in this tissue (Fig. 2.1E and fig. 2.S2D). Defects were not observed in surrounding tissues, indicating that PEEL-1 and PMPL-1 act in the same cell to cause toxicity.

2.2.2 PMPL-1 is a conserved membrane protein of unknown function

PMPL-1 belongs to the Plasma Membrane Proteolipid 3 (PMP3) family of proteins, so we named it "PMP3-Like protein 1." PMP3 proteins are widely present in bacteria, plants, and fungi, and are found in some simple animals (*7*). The role of PMP3 proteins in animals is unknown, but PMP3 proteins in plants, fungi, and bacteria are important for cold-stress resistance, membrane protein trafficking, and ion homeostasis (*8–11*). *pmpl-1* mutant worms do not have any obvious phenotypes. However, PMPL-1 is highly conserved among nematodes (fig. 2.S3). We identified 15 PMP3-like proteins in *C. elegans* (fig. 2.S4) through BLAST searches. PMPL-2 (also known as TXT-9) is most similar to PMPL-1 (75% similarity) and was found in an

RNAi screen for defective transcellular chaperone signaling (*12*). However, there is no known molecular function for any *C. elegans* PMP3-like protein.

PMP3 proteins are thought to contain two transmembrane spanning domains (*13*, *14*), consistent with DeepTMHMM predictions for PMPL-1 (fig. 2.S5A) (*15*). However, AlphaFold2 predicts PMPL-1 as a monotopic protein, with four helices passing through only one leaflet of a lipid bilayer (Fig. 2.1D and fig. 2.S5B) (*16*, *17*). We favor the monotopic prediction of PMPL-1 because two recently solved structures of a bacterial photosynthetic complex showed a PMP3 protein within the complex having a similar monotopic structure (*18*, *19*).

In contrast to *pmpl-1*'s broad conservation in nematodes (fig. 2.S3), *peel-1* is found only in *C. elegans* and has no homology to any known protein (*5*). The dramatic difference in conservation of these genes suggests that PMPL-1 has biological roles other than supporting PEEL-1 toxicity. *pmpl-1* is on a different chromosome than *peel-1 zeel-1* (see Materials and Methods), making it genetically unlinked from the TA element. This suggests that PEEL-1 co-opts PMPL-1 for its own use.

2.2.3 PEEL-1 and PMPL-1 are sufficient for toxicity in HEK293T cells

Given that *pmpl-1* was the only full suppressor of PEEL-1 toxicity found in our screen, we hypothesized that PEEL-1 and PMPL-1 may be the only two components required for toxicity. We expressed these proteins in human embryonic kidney cells (HEK293T) and assayed for cytotoxicity, using lactate dehydrogenase (LDH) release into the culture media as a measure of plasma membrane rupture (*20*). We found that each protein alone was not toxic, but co-expression of PEEL-1 and PMPL-1 resulted in significant cytotoxicity (Fig. 2.2A). Other PMP3-

like proteins such as *C. elegans* PMPL-2 and yeast PMP3 were not toxic with PEEL-1 (Fig. 2.2B), suggesting that PMPL-1 has a specific role in toxicity that is not universal to the PMP3 family. The reconstitution of toxin activity in a heterologous system suggests that PEEL-1 and PMPL-1 are sufficient for toxicity and may act by disrupting an essential, conserved cellular process. Antidote activity could also be reconstituted in HEK293T cells by co-expression of ZEEL-1, resulting in reduced toxicity (Fig. 2.2A).

2.2.4 Plasma membrane localization of PEEL-1 and PMPL-1 is critical for toxicity

PEEL-1 and PMPL-1 localize to several membrane-bound compartments in HEK293T cells and *C. elegans*. Both proteins localize to the endoplasmic reticulum (ER), while PMPL-1 also localizes to the plasma membrane (PM) and to lipid droplets (Fig. 2.2C-D and fig. 2.S6A), consistent with the predicted monotopic topology. Although PEEL-1::eGFP is not easily detectable on the PM of transfected HEK293T cells, stable expression of PEEL-1::eGFP in both HEK293 cells and in *pmpl-1* knock-out worms shows clear PM localization (fig. 2.S6B-C), suggesting that PMPL-1 is not required for PEEL-1 to reach the plasma membrane.

To determine in which compartment these proteins perform their toxic roles, we prevented their movement to the PM by the addition of a GBR1 ER-retention tag (*21*). ER-retention tags on either protein resulted in more than an 80% drop in toxicity, and retention tags on both proteins completely suppressed toxicity (Fig. 2.2E), suggesting that PEEL-1 and PMPL-1 do not act in the ER and likely perform their toxic roles at the PM.

2.2.5 PEEL-1 toxicity causes cell swelling in HEK293T cells

LDH release is an end-point measure of plasma membrane rupture, but it was unclear whether plasma membrane rupture was a primary or secondary effect of toxicity. Imaging 48 hours after transfection showed that 92% of HEK293T cells co-expressing PEEL-1 and PMPL-1 were swollen, about double their typical size (Fig. 2.3A-B). Live imaging showed cells exhibiting abnormal phenotypes approximately 16 hours after transfection. Cells began with slight swelling of the nucleus and cell, followed by jetting out round protrusions of 5-10 μm, similar in size to the nucleus (Fig. 2.3C). These large protrusions are short-lived, typically existing for less than 10 minutes before being reabsorbed by the cell. Eventually, cells swell evenly in all directions (Fig. 2.3D). Swollen cells often lose the integrity of their plasma membrane, followed by ER fragmentation and ER swelling (fig. 2.S7). By 48 hours after transfection, we observe various outcomes for swollen cells: cell lysis, cell detachment from the plate, and swollen, intact cells attached to the plate. Our LDH assay likely captures only the first of these phenotypes and may underestimate the fraction of cells experiencing toxicity. The cell swelling phenotype is reminiscent of the necrotic vacuoles previously described in PEEL-1 affected embryos in *C. elegans* (*5*).

The cell swelling phenotype caused by PEEL-1 toxicity suggests that cell death is non-apoptotic (*22*). Indeed, *C. elegans* mutants defective in apoptosis (*ced-3*) and apoptotic cell engulfment (*ced-2* and *ced-5*) are still susceptible to PEEL-1 toxicity (fig. 2.S8) (*23–25*). Other mammalian cell death pathways such as pyroptosis and necroptosis appear to be absent from *C. elegans* (*26*). This suggests that the cytotoxic phenotypes we observe are a direct effect of PEEL-1 and PMPL-1 activity rather than indirect induction of programmed cell death.

2.2.6 PEEL-1 has an amphipathic helix that is critical for toxicity

PEEL-1 has no homology to known proteins (*5*), so we turned to structural predictions to identify important regions in PEEL-1. AlphaFold2 predicted a low-confidence structure with six alpha-helices (Fig. 2.4A) (*16*, *17*). The longest four helices matched DeepTMHMM's prediction of four transmembrane domains (fig. 2.S5C). Closer inspection revealed that the fourth predicted transmembrane domain is amphipathic (Fig. 2.4B). An amphipathic helix (AH) has opposing hydrophilic and hydrophobic faces and is a critical feature of known pore-forming toxins. For example, actinoporins are a diverse family of toxins which oligomerize their AHs to construct channels (*27*). The properties of the predicted PEEL-1 AH are similar to actinoporin AHs. The PEEL-1 AH is 22 amino acids in length (residues 111-132), making it long enough to span the lipid bilayer (*28*). The hydrophobic moment (μH), a measure of the degree of amphipathicity, is within the range of actinoporin channel-forming toxins (fig. 2.S9) (*29*). One notable difference is that the PEEL-1 AH is more hydrophobic than actinoporin AHs (fig. 2.S9), suggesting that the PEEL-1 AH may always exist within the membrane, unlike in actinoporins which have soluble AH conformations.

We hypothesized that the PEEL-1 AH may perform a similar role as it does in actinoporins, where AHs construct the lining of a toxic channel. To test this, we determined whether this helix was required for toxicity in HEK293T cells and *C. elegans*. We found that deleting the last two helices of PEEL-1 did not severely impair toxicity ("-28 aa" and "-39aa" in Fig. 2.4C-D). However, deleting the AH resulted in complete loss of toxicity in both mammalian cells and worms ("-65 aa" in Fig. 2.4C-D). By deleting one amino acid at a time, we determined that toxicity in HEK293T cells was eliminated after removing at least 40 amino acids (fig. 2.S10). This PEEL-1(-40aa) mutant contains only three residues following the AH, possibly destabilizing the AH.

We also tested whether the amphipathic property of the PEEL-1 AH was critical for toxicity by creating a series of missense mutants that modified the AH's hydrophobic moment. Six single-missense mutants were tested. Three of these mutants attenuated PEEL-1 toxicity (S124F, L122Q, L115Q) while three mutants were still fully toxic (L118Q, S124V, L126Q) (Fig. 2.4E). However, all pairwise combinations of these three fully toxic mutants resulted in complete loss of toxicity (Fig. 2.4E). Across the nine missense mutants we tested, toxicity correlated well with the amphipathicity of the AH (Fig. 2.4F). These data suggest that the amphipathic property of the PEEL-1 amphipathic helix is critical for toxicity, supporting its potential role as the lining of a channel. However, we cannot rule-out alternative explanations for the importance of the AH, including roles in protein-protein interaction or protein stability.

2.2.7 PEEL-1 and PMPL-1 create a non-specific monovalent cation channel

We hypothesized that PEEL-1 toxicity was caused by disruption of ionic gradients across the cell membrane, leading to osmotic imbalance and subsequent cell swelling. This could be through two mechanisms: an ion leak channel, where specific ions flow down their electrochemical gradient, or a non-selective pore, causing flux of all ions and smaller osmolytes across the plasma membrane. To test these possibilities, we performed whole-cell patch-clamp electrophysiology on HEK293 cells in physiological ion conditions (high internal K^+, high external Na^+). Currents from cells transfected with *peel-1* alone or *pmpl-1* appeared similar to currents from untransfected cells (Fig. 2.5A and fig. 2.S11A-B). Obtaining recordings of cells co-transfected with *peel-1* and *pmpl-1* was difficult, due to variable timing of toxicity, and was complicated by loss of cells due to swelling. Therefore, we generated tetracycline-inducible cell lines that more tightly regulate PEEL-1 and PMPL-1 expression. We generated a HEK293 cell

line with tetracycline-inducible PMPL-1::mCherry and then introduced stable expression of either eGFP ("Control" cell line) or PEEL-1::eGFP ("Experimental" cell line). Cells with and without tetracycline are hereafter referred to as induced and uninduced, respectively. Nearly all Experimental cells swell 24 hours after induction (fig. 2.S12A), with the first signs of swelling occurring 6 hours after induction (fig. 2.S12B).

We performed whole-cell patch clamp recordings of uninduced cells, and at 6 hours and 28 hours after induction (Fig. 2.5B-C). Similar to untransfected naïve HEK293 cells, uninduced Control cells exhibited small instantaneous negative (inward) currents at negative potentials without leak subtraction, and these currents did not increase after induction (Fig. 2.5B and fig. 2.S11C). In contrast, uninduced Experimental cells exhibited a larger instantaneous inward current compared to uninduced Control cells and naïve HEK293 cells, and the magnitude of these currents further increased after 28 hours of induction (Fig. 2.5C and fig. 2.S11C). When clamped at the most negative potential (-100 mV), inward currents from uninduced and 28-hour induced Experimental cells were 7-fold and 30-fold greater than naïve HEK293 cells, respectively. These data suggest that co-expression of PEEL-1 and PMPL-1 creates a constitutively active ionic channel.

The inward currents observed in uninduced Experimental cells (Fig. 2.5C and fig. 2.S11C) were instantaneous, exhibiting no voltage-dependent gating kinetics, consistent with a leak channel constitutively open at physiological resting membrane potentials. This ion leakage in uninduced Experimental cells was surprising since these cells appear healthy and do not have obvious growth defects. We detected some PMPL-1 expression in uninduced Control and Experimental cells (Fig. 2.5D and fig. 2.S12A), consistent with previous reports of leaky background expression from tetracycline-inducible promoters (*30–32*). Therefore, we reasoned

that uninduced Experimental cells may have sufficient background PMPL-1 expression to display an electrophysiological phenotype but not enough to cause cell swelling or cell death.

To characterize the ionic selectivity of channels formed by PEEL-1 and PMPL-1 co-expression, we performed ionic substitution recordings. We took advantage of the inward currents generated by uninduced Experimental cells since these cells are morphologically normal, reducing the possibility of contamination by endogenous HEK293 channels secondarily activated by cell swelling or cell death such as the LRRC8-associated volume-regulated anion channels (VRACs) (*33*, *34*). We substituted the physiological mixed cation bath solution (high Na^+/low K^+) with equivalent bath solutions containing single cations (Na^+, K^+, Cs^+, $NMDG^+$ or a combination of 20mM Ca^{2+}/120mM $NMDG^+$) and monitored instantaneous inward currents at negative potentials. We found that bath solutions with Na^+, K^+, or Cs^+ still yielded large inward currents, while bath solutions with $NMDG^+$ and Ca^{2+} abolished all inward currents (Fig. 2.5E and fig. 2.S13). These results suggest that PEEL-1 and PMPL-1 co-expression creates a channel that is permeable to monovalent cations (Na^+, K^+, and Cs^+) but impermeant to the divalent Ca^{2+} cation and the bulkier cation $NMDG^+$. We next tested for Cl^- permeability by recording with high internal Cl^- (140 mM KCl) and low external Cl^- (140mM $NMDG^+$). Any Cl^- permeability at negative potentials would cause Cl^- movement out of the cell and be recorded as a negative (inward) current. We observed no inward currents (Fig. 2.5E and fig. 2.S13), indicating that the channel is impermeable to Cl^-. Altogether, these results suggest that PEEL-1 and PMPL-1 create an ion channel rather than a non-selective pore. The PEEL-1 and PMPL-1 channel conducts monovalent cations and is impermeable to anions and divalent cations. An increase in PMPL-1 expression in induced Experimental cells may result in excessive influx of Na^+, sufficient to

overwhelm compensatory volume regulatory mechanisms, leading to osmotic dysregulation, cell swelling, and cell death.

2.2.8 Predicted PEEL-1 pentamer structure has features of an ion channel

Ion channels are typically made up of oligomeric complexes, so we used AlphaFold2 to determine possible oligomeric structures of PEEL-1. Structural predictions were of low-confidence due to PEEL-1's lack of homology to known proteins (*5*). Nevertheless, the predicted PEEL-1 pentamer has a striking resemblance to known cation channels (fig. 2.S14A-D): (a) the outside surface of the complex is hydrophobic, consistent with its location in lipid bilayers, (b) the complex has a central, uninterrupted hydrophilic pore, providing a potential path for ions through the complex, and (c) the opening of the pore region is surrounded by a ring of negatively charged residues. These features are strikingly similar to the structure of the ZAR1 cation channel, which is a toxic pentameric channel with a ring of acidic residues at the extracellular mouth of the pore that serve as a cation selectivity filter (*35*). In the predicted PEEL-1 pentamer, a similar ring of negative charges is formed by five D109 residues. Mutating this residue to an alanine (D109A) resulted in complete loss of PEEL-1 toxicity (fig. 2.S14E). This residue is two amino acids before the PEEL-1 AH. Five copies of the PEEL-1 AH meet in the middle of the complex, forming the pore-like region of the predicted structure (fig. 2.S14A-B), consistent with our model of PEEL-1 AHs constructing the lining of a channel.

2.2.9 In vitro reconstitution of the PEEL-1/PMPL-1 ion channel

We used an independent approach to test if PEEL-1 and PMPL-1 create an ion channel by assaying whether these purified proteins allow ions to flow through artificial planar lipid bilayers (Fig. 2.5F). PEEL-1 and PMPL-1 were individually expressed in *E. coli* with maltose-binding protein (MBP) tags, affinity purified in detergent, and incorporated into liposomes for delivery to planar lipid bilayers (fig. 2.S15 and Materials and Methods). Interestingly, PEEL-1 appears to be toxic to *E. coli* and removal of the PEEL-1 amphipathic helix attenuated this toxicity (fig. 2.S16). Truncation products of PEEL-1 were purified along with the full-length protein (fig. 2.S15). Only after adding PEEL-1 liposomes and PMPL-1 liposomes to the same bilayer did we observe discrete unitary conductance events similar to those created by typical ion channels (Fig. 2.5G and fig. 2.S17) (*35–38*). There was variation in single channel conductance and open duration (Fig. 2.5G and fig. 2.S17). Such variability could reflect heterogenous stoichiometries, or it could be due to technical limitations such as misfolded proteins, truncated PEEL-1 proteins in the sample, mixed orientation of proteins in the bilayer, or imprecise amounts of each protein in the bilayer. Nevertheless, the stepwise conductance *in vitro* and cation-selective property in cells suggest that PEEL-1 and PMPL-1 together construct a monovalent cation channel which causes osmotic dysregulation and cell death.

2.3 DISCUSSION

Our study provides unprecedented understanding of an animal TA system. We provide evidence that the PEEL-1 toxin co-opts a conserved membrane protein to create a cation channel which ultimately causes cell swelling and death. We show that PEEL-1 may be the primary structural component of the channel and hypothesize that PMPL-1 is required to gate the channel open. PEEL-1 may have evolved this co-option mechanism for temporal control over ion channel

activity, to avoid toxicity in sperm and early embryos. Prevention of off-target toxicity in the germline and before the maternal-to-zygotic transcriptional switch is likely an evolutionary hurdle faced by many animal TA systems. Therefore, co-option may be a common evolutionary route for toxicity in animal TA systems.

2.4 MATERIALS AND METHODS

2.4.1 Worm strains and maintenance

Worm strains were maintained using standard procedures (*39*). Strains used in this study are provided in Table 2.S1.

2.4.2 Isolation of suppressors of heat-shock induced PEEL-1

Worm strains XZ1047 and XZ1372 were mutagenized using ENU or EMS (*39*, *40*). These strains contained two copies of *hsp-16.41p*::*peel-1* in order to avoid isolating suppressors that have mutations in the transgene. F2 populations of worms were heat-shocked at 34°C for 2 hours and allowed to recover at room temperature overnight. Surviving worms were isolated. Approximately 200,000 haploid genomes were screened. *yak52* was isolated from EMS mutagenesis of XZ1047, while *yak103* was isolated from ENU mutagenesis of XZ1372.

We isolated a total of six independent suppressors of heat-shock PEEL-1 toxicity. Four of the mutations were partial suppressors while two were full suppressors. The causative genes were identified using genetic mapping and whole-genome sequencing. The four partial suppressors have mutations in mRNA export factors *nxf-1* and *nxt-1*, but act by affecting expression of *hsp-16.41p::peel-1* and do not suppress toxicity of endogenous, sperm-delivered PEEL-1 (*41*).

The two full suppressors *yak52* and *yak103* are recessive suppressors of PEEL-1 toxicity and have no other obvious phenotype. They were determined to be allelic by complementation testing and were mapped to the X chromosome by crossing to strains carrying visible or fluorescent markers on each chromosome (strains EG1000, EG1020, EG8040, EG8041). The closest linkage seen was to the visible marker *lon-2*.

2.4.3 Identification of pmpl-1

Strains XZ1177 and XZ1307 carrying 4X and 5X outcrossed pmpl-1(yak52) and pmpl-1(yak103) were subjected to whole-genome sequencing. DNA was purified according to the Hobert laboratory protocol (http://hobertlab.org/whole-genome-sequencing/). Illumina (XZ1177) or Ion Torrent sequencing (XZ1307) was performed at the University of Utah DNA Sequencing Core Facility. The dataset for XZ1177 contained approximately 17,000,000 reads of a mean read length of 36 bases, resulting in ~6X average coverage of the C. elegans genome. The dataset for XZ1307 contained approximately 12,800,000 reads of a mean read length of 147 bp, resulting in ~19X average coverage. The sequencing data were processed on the Galaxy server at usegalaxy.org (42). SNPs and indels were identified and annotated using the Unified Genotyper and SnpEff tools (43, 44). Although we found genes on the X chromosome in each strain containing nonsynonymous mutations, none of these genes were affected independently in both strains. Instead, to identify potential large deletion, we used BEDtools Genome Coverage tool to calculate sequencing coverage across the genome in intervals of consecutive bases with the same coverage (45). Intervals with zero sequencing coverage were then annotated using SnpEff, filtered to intervals on the X chromosome which overlapped protein-coding regions, and sorted by length. In XZ1307, the largest of these intervals on the X chromosome was found near where

yak103 and yak52 had been mapped and was subsequently confirmed to correspond to yak103, a 323-bp deletion spanning the F47B7.1 gene, from its 5'UTR to 3'UTR and deleting all the coding sequence. In XZ1177, we found two small regions of F47B7.1 lacking coverage in our whole genome sequencing. Via Sanger sequencing, we identified yak52 as a nonsynonymous mutation in one of these regions, with a G to A mutation that results in an A47T substitution.

2.4.4 Heat-shock PEEL-1 toxicity assay

Heat-shock PEEL-1 assays were performed with 50 gravid adults of each strain for each biological replicate. Adults were picked to clean NGM plates seeded with OP50 and heat-shocked (2 hours at 34°C). After recovering at room temperature for 2 hours, moving worms were counted. Non-moving worms were tested for response to touch stimulus and scored as dead if they did not move.

2.4.5 Sperm-delivered PEEL-1 toxicity assay

Sperm-delivered PEEL-1 toxicity in wild-type or *pmpl-1* mutant backgrounds was assayed by counting unhatched embryos. To make *pmpl-1*; *peel-1(+) zeel-1(+) / peel-1(-) zeel-1(-)* worms, AFS216 *peel-1(-) zeel-1(-)* males were crossed to *peel-1(+) zeel-1(+)* worms with *pmpl-1* mutations *yak103* (XZ2194), *yak52* (XZ2283), or wild-type *pmpl-1* (N2, control). F1 males were backcrossed to the parent strain (XZ2194, XZ2283, or N2) to obtain the desired genotype.

Embryonic lethality was assayed on individual, self-fertilizing hermaphrodites. Single worms were allowed to lay embryos for 16-24 hours before being removed from the plate. Plates were left for one more day to allow embryos to develop and hatch, and the progeny were scored.

Unhatched embryos and larval worms were counted. Unhatched embryos were scored as dead. Progeny were genotyped in bulk for *zeel-1(-)* by PCR using oGP65 (5'attctggagttcgtgaggtgc3') and oGP66 (5'ccctcctttcccacccaac3'). Only plates showing *zeel-1(-)* alleles were considered to have parents with the desired genotype, heterozygous *peel-1(+) zeel-1(+) / peel-1(-) zeel-1(-)*.

2.4.6 Plasmid construction

All plasmids used in this study are provided in Table 2.S2. Plasmids used for mammalian cell transfection were generated using Gibson assembly (*46*) using N1 vector backbone (CMV promoter). Constructs for tetracycline-inducible expression were cloned into pFTSH vector (gift from Nancy Maizels). Plasmids used for *C. elegans* transgenics were constructed using either Gibson assembly or Gateway cloning (Invitrogen). Typical Gateway cloning was performed using a 3-fragment approach (promoter, coding region, and fluorophore-UTR or UTR) into a destination vector, pCFJ150. Restriction digest was used to confirm the plasmid was correct. Sanger sequencing was used to confirm that constructs did not contain mutations (Table S2). Transgenic worms were made using microinjection. For each transgenic strain, we isolated at least two independent lines and confirmed similar results.

2.4.7 Microscopy

All microscopy was performed on a NikonTi2-E Crest X-light V2 spinning-disk confocal microscope. For live-cell imaging, mammalian cells were maintained in a humidified, heated chamber (37°C) in 5% carbon dioxide. For long-term imaging, the objective heater was also heated to 37°C. For imaging *C. elegans*, worms or embryos were picked onto a 2% agar pad,

immobilized in a humidified chamber for 10 minutes in 33mM sodium azide, and then immediately imaged.

2.4.8 Mammalian cell maintenance and generation of clonal cell lines

All mammalian cells were grown at 37°C and 5% carbon dioxide in DMEM with GlutaMAX (Thermo Fisher Scientific), 10% fetal bovine serum (RMBIO or Gibco), and 100 U/mL of penicillin-streptomycin (Gibco). All transfection-based cytotoxicity experiments were done in HEK293T cells (gift from Mary Claire-King). For electrophysiology experiments, HEK293 cells (CRL-1573; ATCC) were used. Tetracycline-inducible cells (tetON cells) were made in HEK293 Flp-In TReX cells (Invitrogen) (gift from Nancy Maizels). tetON cells were grown in media made with Tet system approved fetal bovine serum (Gibco).

Stable tetON lines were generated by co-transfection of FlpO (pOG44; Life Technologies) and PtetON::*pmpl-1*::mCherry (pLC79, in pFTSH backbone, a gift from Nancy Maizels). Two days after transfection, cells were plated at low density in 10cm plates with hygromycin B (150µg/mL). Two clonal populations were picked into media with hygromycin B (150µg/mL), grown to 80% confluency, and frozen in 10% DMSO at -80°C. One clonal population was used to generate the experimental and control cell line.

Experimental tetON and control tetON cell lines were generated parallel. One tetON::*pmpl-1*::mCherry clonal cell line was transfected with the respective plasmid, *peel-1*::eGFP (pGP9) or eGFP. Two days after transfection, cells were plated at low density in selective media containing both hygromycin B (150µg/mL) and G418 (400µg/mL). After about two weeks, multiple single colonies were picked and continued to be grown under selection. For the experimental cell line, more than half the picked clones which grew in selective media did

not have visible green fluorescence, presumably due to leaky *pmpl-1*::mCherry imposing selection against *peel-1*::eGFP-expressing cells. Therefore, only clones which had green fluorescence were maintained. This issue was not encountered during the generation of the control cell line. All experimental and control cell lines were screened following tetracycline treatment. 48 hours after tetracycline treatment, all cell lines had red fluorescence, all experimental cell lines had very few adhered cells (n= ~6), and all control cell lines had no obvious cellular phenotypes (n= ~10).

2.4.9 Cytotoxicity assay

Cytotoxicity in mammalian cells was measured using a colorimetric assay for LDH release (Promega CytoTox 96). 12-well plates were seeded with approximately $7x10^4$ HEK293T cells about 24 hours prior to transfection. Each well was transfected with a mixture of 1.5µg DNA and 3µg PEI in 200µL OptiMEM (Gibco). Combinations of three constructs were used (0.5µg each), using fluorophore-encoding vectors (mCherry or eGFP) as controls when less than three constructs are being tested. Supernatant was collected 43-45 hours after transfection and used in 96-well plates for LDH assays, following manufacturer protocol. Two technical replicates were measured for each well and absorption at 490nm was averaged between replicates. All experiments included wells for a non-killing control (fluorophore-encoding vectors), *peel-1*::eGFP alone, and *peel-1*::eGFP with *pmpl-1*::mCherry. The fold-change over *peel-1*::eGFP was calculated from transfections performed on the same day.

2.4.10 SDS-PAGE and western blots

For lysis of mammalian cells, cells were collected from 12-well plates and washed with PBS. Cell pellets were lysed in radioimmunoprecipitation assay (RIPA) lysis buffer (25mM Tris pH 7.4, 150mM NaCl, 0.1% SDS, 1% NP-40, 1% sodium deoxycholate) with DNase and Halt

Protease Inhibitor Cocktail (Thermo Fisher Scientific). After lysis (15min on ice), the sample was centrifuged (21,000 x *g*, 10min,4°C), and the supernatant was used for SDS-PAGE.

Following separation by SDS-PAGE, proteins were transferred onto a nitrocellulose membrane. After blocking in Intercept Blocking Buffer (LI-COR, 1hr room temperature) membranes were incubated with primary antibody and nutated overnight at 4°C. Mouse monoclonal anti-mCherry (a gift from Jihong Bai and the Fred Hutch Cancer Center antibody development shared resource center; 1:1000), mouse monoclonal anti-MBP (New England Biolabs, E8032; 1:10,000), rabbit polyclonal anti-GAPDH (Sigma Aldrich, G9545; 1:5,000). Appropriate secondary antibodies were used, either goat anti-mouse or donkey anti-rabbit conjugated to Alexa 680 or Alexa 790 (Invitrogen). Membranes were imaged on an Odyssey CLx (LI-COR Biosciences).

2.4.11 Electrophysiology

For electrophysiology on acutely transfected cells, plasmid constructs were transfected into HEK293 cells using Viafect reagent (E4981; Promega), following the manufacture's protocol. Cells were first prepared for transfection by plating onto 12-well tissue culture plates (Nunc 12-565-321; Thermo Fisher Scientific) at a density of ~0.5-2 x10^5 cells per well and grown to ~80-90% confluence with standard media, allowing for one confluent well per transfection condition. On the day of transfection, media were replaced with 0.5 mL fresh DMEM with 10% FCS, without P/S. Lipophilic/DNA transfection complexes were generated for each well, combining a total of ~1.0µg of plasmid DNAs with serum-free OptiMEM (Gibco 31985062) to a final volume of 100µL, then adding 3.0 µL Viafect with gentle trituration, allowing the mixture to assemble at 24°C for 30 minutes, and then added to each well, dropwise.

Transfected cells were incubated overnight at 37°C and visually monitored for transfection efficiency *in situ* using an inverted plate microscope equipped with fluorescence (Invitrogen EVOS M7000; Thermo Fisher Scientific). Transfection efficiencies were typically >70-80%. Specific amounts of plasmid DNAs (pDNAs) for acute transfections per well: a) pCMV::*peel-1*::eGFP (0.8 µg), b) pCMV::*pmpl-1*::mCherry (0.1µg) and pcDNA3 (0.8µg), c) pCMV::*peel-1*::eGFP (0.8 µg) and pCMV::*pmpl-1*::mCherry (0.8µg). Untransfected HEK293 cells served as controls.

Following overnight incubation, cells in transfected wells were dissociated with TrypLE (Gibco); and replated at low density onto 12 mm poly-*D*-lysine-coated glass coverslips (GG-12-pdl; NeuVitro) in 24-well tissue culture plates (FisherBrand FB012929; Thermo Fisher Scientific), for patch-clamp electrophysiology. Typically, ~10,000-15,000 cells were replated per well at sufficiently low density to isolated individual cells. This was necessary to prevent the formation of electrical junctions between contacting cells, which could preclude adequate space-clamp recording conditions. Recordings were performed from 0.5-3 days after replating at low densities. Control and Experimental stable HEK293 cell lines were similarly replated at low density on coverslips for patch-clamp recordings.

For patch-clamp recordings, coverslips containing adherent cells were transferred to a Zeiss AxoExaminer.A1 microscope, equipped with an 40X water immersion objective and epifluorescence capability. Pipettes were positioned with a Sutter MPC-325 micromanipulator (Novato). Whole-cell voltage-clamp recordings were acquired with an AxoClamp200B amplifier (Molecular Devices), using pClamp10.4. Composition of recording solutions are listed below:

Bath Solutions:

i. HEK293 bath (4 K^+, 145 Na^+) (in mM): 4.0 KCl, 145 NaCl, 2.0 $CaCl_2$, 2.0 $MgSO_4$, 10 HEPES, 10 glucose, pH to 7.4 with NaOH.

ii. *140 Na^+ (in mM):* 140 NaCl, 2.0 $CaCl_2$, 2.0 $MgCl_2$, 10 HEPES, pH to 7.4 with NaOH.

iii. *140 K^+ (in mM):* 140 KCl, 2.0 $CaCl_2$, 2.0 $MgCl_2$, 10 HEPES, pH to 7.4 with KOH.

iv. *140 Cs^+ (in mM):* 140 CsCl, 2.0 $CaCl_2$, 2.0 $MgCl_2$, 10 HEPES, pH to 7.4 with CsOH.

v. *140 $NMDG^+$ (in mM):* 140 N-methyl-*D*-glucamine ($NMDG^+$), 2.0 $CaCl_2$, 2.0 $MgCl_2$, 10 HEPES, pH to 7.4 with HCl.

vi. *20 Ca^{2+} (in mM):* 20 $CaCl_2$, 120 $NMDG^+$, 2.0 $MgCl_2$, 10 HEPES, pH to 7.4 with HCl.

Internal Pipette Solutions:

i. *140 K^+, low Cl^- (in mM)*: 140 K-*D*-gluconate, 1.0 $CaCl_2$, 2.0 $MgCl_2$, 10 HEPES, 2.4 EGTA, 4.0 Na_2ATP, 0.3 Na_2GTP, pH to 7.4 with KOH.

ii. *140 K^+, high Cl^- (in mM)*: 140 KCl, 1.0 $CaCl_2$, 2.0 $MgCl_2$, 10 HEPES, 2.4 EGTA, 4.0 Na_2ATP, 0.3 Na_2GTP, pH to 7.4 with KOH.

For all recordings of cationic currents, different bath solutions were used in combination with *140 K^+, low Cl^-* internal pipette solution. For recordings of Cl^- currents, *140 $NMDG^+$* bath solution was used in combination with *140 K^+, high Cl^-* internal pipette solution; under these conditions outward Cl^- conductance at negative potentials would be seen as inward currents by standard electrophysiological recording convention. Patch pipettes were pulled from borosilicate glass (1B120F-4; World Precision Instruments) on a P-97 Sutter Instruments puller (Novato), with resistances of 3.0-5.0 MΩ. Currents were allowed 3-5 mins to stabilize after achieving

whole-cell recording configuration, filtered at 5 kHz, and acquired at 10 kHz. Series resistance compensation was >80% for all recordings.

Mean currents normalized to cell capacitance were analyzed and plotted using pClamp10.4 (Molecular Devices) and Microsoft Excel. Leak subtraction correction was not applied to any of the data. Current-voltage (IV) data were plotted as means with standard errors (SE) and statistical calculations were performed in Prism (GraphPad).

2.4.12 Protein purification

MBP::PEEL-1::His_8 and MBP::PMPL-1 were purified from *E. coli* BL21(DE3) and C43(DE3) cells, respectively. An overnight culture (37°C, 220rpm, 100µg/mL ampicillin, 33µg/mL chloramphenicol) was used to inoculate 1L of LB media (100µg/mL ampicillin, 220rpm, baffled flask), grown at 37°C to OD_{600} of 0.4-0.7, and induced with IPTG (0.5mM). PMPL-1 cells were chilled on ice prior to induction, and then induced in an 18°C shaker overnight. PEEL-1 cells were induced at 37°C for 1 hour since longer induction at lower temperatures resulted in decreased yields. Cells were pelleted and either stored at -80°C or used immediately for purification.

For purification, cell pellets were resuspended in Buffer A (HEPES pH 7.4, 150mM NaCl, 5mM 2-mercaptoethanol, 10% glycerol). Cells were lysed using a Dounce homogenizer, followed by 30 minutes rocking at room temperature after addition of DNase, lysozyme, protease cocktail inhibitor (Pierce), and n-octyl-β-glucopyranoside detergent (β-OG, 2% final concentration) (Anatrace). Lysate was clarified by centrifugation (20 min, 20,000 x *g*, 4°C) before batch binding with amylose resin for 1-2 hours at 4°C. Resin was pelleted (5 min, 700 x *g*) and resuspended in wash buffer to load into a clean column. Four washes (5 column volumes) and four elutions (two column volumes) were performed by gravity flow. Wash buffer and

elution buffer contained 1% β-OG in Buffer A. Maltose (10mM) was included in elution buffer. Protein yield was estimated by absorbance at 260nm. Proteins were stored at 4°C for less than three days before incorporation into liposomes.

2.4.13 Proteo-liposomes

Liposomes containing MBP::PEEL-1::His_8 or MBP::PMPL-1 were made with 50% DOPC:POPS (1,2-dioleoyl-sn-slycero-3-phosphocholine: 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine). Lipid stocks were purchased in chloroform (Avanti Polar Lipids) and β-OG stock was made in methanol. Lipid films were made by mixing lipids with β-OG detergent (lipid:detergent molar ratio of 4:35) in glass vials, dried under a nitrogen stream (10-20 min), and further dried in a Speedvac evaporator (4-6 hours). Buffer A was added to the dried lipid-detergent mixture and resuspended via two or three rounds of bath sonication (5 min, room-temperature) with nutation (20 min, 4°C). Protein was added in an approximate protein:lipid molar ratio of 1:400 (PMPL-1) or 1:2,500 (PEEL-1). Detergent was removed by dialysis: 500µL of sample was dialyzed (20 kDa cutoff) in 250mL Buffer A with 0.5g Biobeads SM-2 (Bio-Rad) for 16-18 hours at 4°C. Dialyzed sample was floated through a density gradient (Histodenz 35%, 25%, 0%; Sigma-Aldrich) using ultracentrifugation (SW60Ti, 55krpm, 4°C, 30 min). Liposomes were collected from the top fraction (200µL, using a wide-bore pipette tip), aliquoted, flash frozen in liquid nitrogen, and kept at -80°C. Liposomes were thawed fresh on the day of each experiment. A second liposome prep was made from an independent protein purification with two adjustments: soybean lipids (Avanti Polar Lipids) with cholesterol were used for liposomes instead of DOPC and POPS, and 1mM EDTA was included in Buffer A.

2.4.14 Synthetic planar lipid bilayers

Two ~50µL wells (*cis* and *trans*) linked by a ~20µm PTFE aperture were filled with Buffer B (20mM HEPES, pH 7.4, 150mM NaCl, 10% glycerol, 1mM $CaCl_2$). A mixture of asolectin lipid (Sigma-Aldrich) and hexadecane was painted across the aperture to establish a planar bilayer membrane as previously described (*47*). To promote liposome fusions, an osmotic gradient across the bilayer was established by perfusion of Buffer C (20mM HEPES, pH 7.4, 600mM NaCl, 3.36% glycerol, 1mM CaCl2) into the *cis* well. A voltage of -180mV was applied across the membrane by two Ag/AgCl electrodes and the ion current was measured. Liposomes were added to the *cis* well and mixed via pipetting. Sharp transient spikes in currents were interpreted as successful liposome fusions. After observing channel activity Buffer D was perfused into the *cis* chamber (20mM HEPES, pH 7.4, 600mM NaCl, 10% glycerol) to remove free liposomes and a voltage of +180mV was applied since channel activity was more stable at +180mV than -180mV. Ion current was recorded at 50 kHz and downsampled to 50 Hz for analysis. The recorded ion current is divided by the applied voltage to give units of conductance. All-points histograms of conductance were constructed using a bin width of 2 pS and normalized based on probability density (area under the histogram equals 1).

We observed channels in 8 independent experiments and with two independent protein preps. A subset of experiments was run with controls, where we observed ion channels in 2 out of 3 trials after adding both PEEL-1 liposomes and PMPL-1 liposomes, but we did not see channel activity when adding PEEL-1 alone (n=4) nor PMPL-1 alone (n=3). A typical experiment ran as follows: 2uL of PEEL-1 liposomes and 2uL of PMPL-1 liposomes were added to the *cis* chamber and mixed via pipetting. Liposome fusions to the bilayer were confirmed by transient spikes in conductance, beginning approximately 5-20 minutes after addition and channel activity was seen approximately 5-60 minutes after fusions began. The amount of time

until channels were observed was used as a benchmark for control experiments, where 2uL of either PEEL-1 liposomes or PMPL-1 liposomes were added to the chamber and allowed to fuse for the same amount of time or longer than experimental runs.

2.5 ACKNOWLEDGMENTS

We thank Amy Clippinger, Chau Vuong, Irini Topalidou, Michael Crawford, Tyler Couch, Emma Mackey, Teresa Swanson, Andrew Oberst, Adam Steinbrenner, Sharona Gordon, Bertil Hille, Jihong Bai, Alexey Merz, Suzanne Hoppins, Harmit Malik, and members of the Malik lab (Fred Hutch) for their technical help and input throughout this project. We thank Aaron Severson for the AFS216 strain and Jérôme Cattin-Ortolá for isolating *yak52*. Some strains were provided by the CGC, which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). This work was supported by the following grants: National Institutes of Health Institutional Training Grant T32 GM007270 (LC), National Science Foundation CAREER Award MCB-1552101 (MA), National Science Foundation Grant MCB-2344838 (MA), National Institutes of Health Grant R01 HL144801 (J-MR), National Institutes of Health Grant R01 HL151389 (J-MR), National Institutes of Health Grant R01 HL126523 (J-MR), National Human Genome Research Institute R01HG005115 (AHL, JHG).

2.6 FIGURES

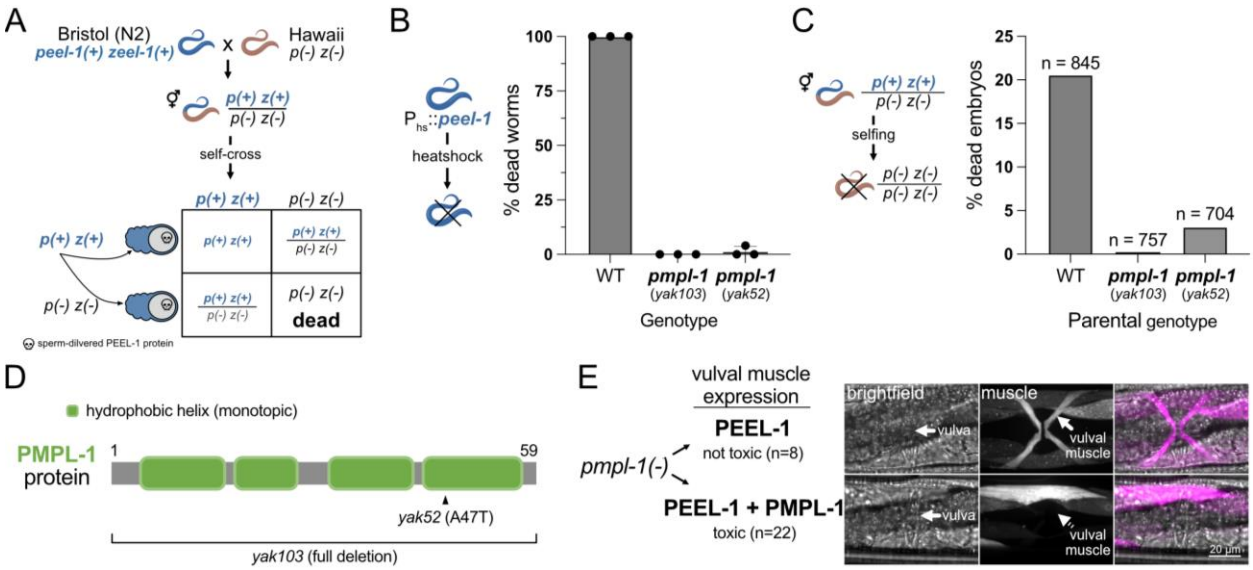


Fig. 2.1. PMPL-1 is necessary for PEEL-1 toxicity in *C. elegans*.

(A) Selfish activity of the *peel-1 zeel-1* toxin-antidote system in *C. elegans* shown in a genetic cross of strains from Bristol (which have the genetic element) and Hawaii (which lack the genetic element). Hermaphrodite worms heterozygous for the presence of *peel-1 zeel-1 (p(+) z(+) / p(-) z(-))* have 25% inviable progeny. This is due to sperm-delivered PEEL-1 toxicity causing developmental arrest of *zeel-1(-)* progeny. **(B)** Proportion of worms dying from ectopic, heat shock-PEEL-1 expression. Two *pmpl-1* mutant alleles (*yak103* and *yak52*) provide resistance to toxicity. **(C)** Proportion of dead, arrested embryos from self-fertilizing hermaphrodites heterozygous for *peel-1 zeel-1*. n = total progeny scored. **(D)** Predicted domain structure of the PMPL-1 protein with mutant alleles shown. The hydrophobic helices are predicted to be monotopic, passing through one leaflet of a lipid bilayer. **(E)** Body wall muscle (magenta) of *pmpl-1*(*yak103*) worms with vulval muscle-specific expression of PEEL-1 alone (top) or PEEL-1 and PMPL-1 (bottom). Vulval muscles appears normal with PEEL-1 alone but

are missing or atrophied when PEEL-1 and PMPL-1 are co-expressed in these cells. All worms expressing PEEL-1 and PMPL-1 in vulval muscle cells had missing or severely deformed vulval muscles (n=22). *peel-1* and *pmpl-1* are both GFP tagged. All channels are shown in fig. 2.S2D. Scale bar = 20μm.

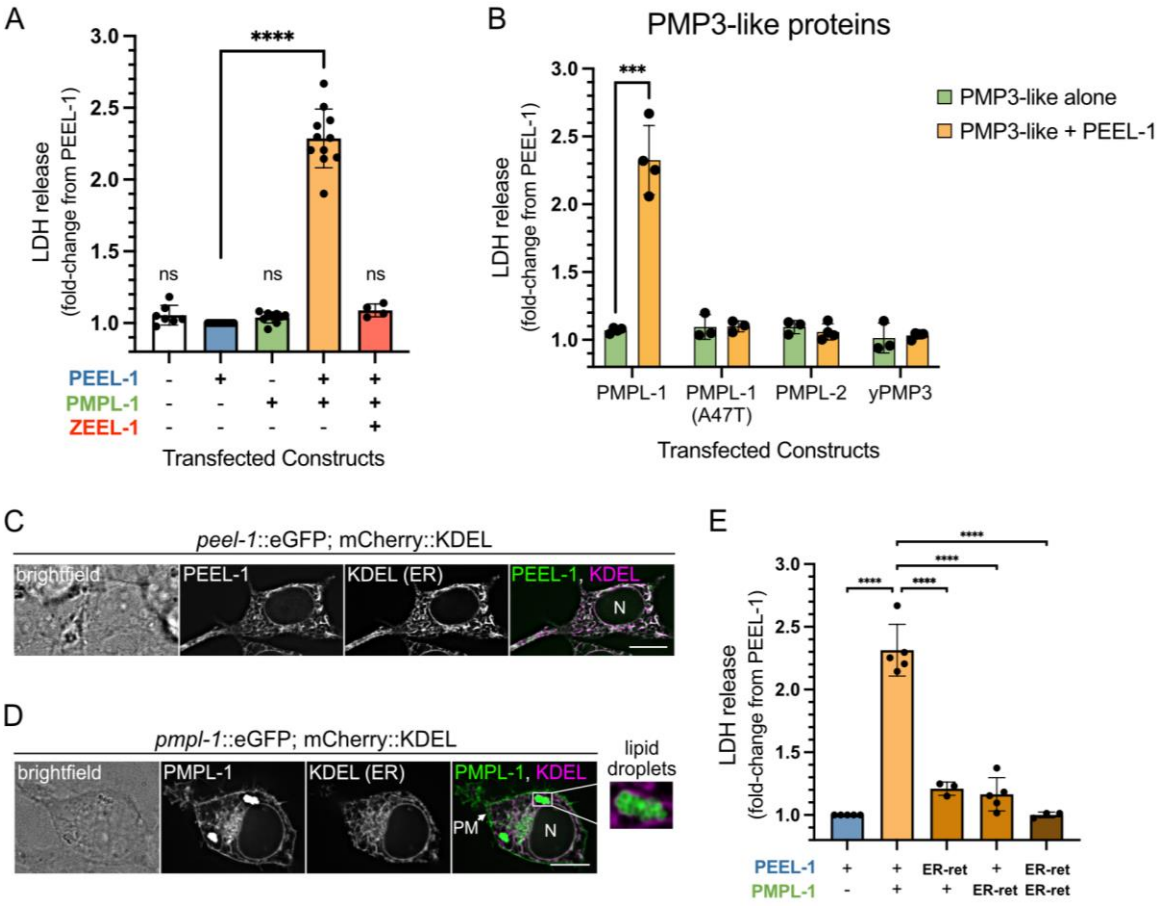


Fig. 2.2. PEEL-1 and PMPL-1 are sufficient for toxicity in HEK293T cells.

(A) Cytotoxicity (measured by LDH release) of combinations of three constructs transfected in HEK293T cells. Each data point is a biological replicate, normalized to LDH release from transfection with *peel-1*::eGFP in the same experiment. All plots show means with SD. Transfections combine constructs encoding for mCherry or eGFP (-) or a fluorescent-tagged protein (+): PEEL-1::eGFP (top), PMPL-1::mCherry (middle), or mCherry::ZEEL-1 (bottom). **(B)** Cytotoxicity of PMP3-like proteins alone or with PEEL-1::eGFP. The PMPL-1 *yak52* (A47T) mutant protein, *C. elegans* PMPL-2, and the yeast homolog yPMP3 are shown. **(C)** Live-cell imaging of a single cell transfected with an ER-marker (mCherry::KDEL) and *peel-1*::eGFP or **(D)** *pmpl-1*::eGFP. The cell nucleus is indicated (N). PMPL-1 is also seen on the plasma

membrane (PM) and on lipid droplets (inset). Scale bar = 10µm. **(E)** Cytotoxicity is suppressed by addition of the GBR1, ER-retention tag on the C-terminus of PEEL-1::eGFP or PMPL-1::mCherry. P-values in (A) and (E) calculated using one-way ANOVA with Dunnett's multiple comparisons test, comparing all samples to PEEL-1 alone in (A) and all samples to PEEL-1 with PMPL-1 in (E). In (B), multiple unpaired t-tests were used with Holm-Šídák test, comparing each PMP3-like protein alone to PMP3-like with PEEL-1 (***, $p < 0.001$; ****, $p < 0.0001$).

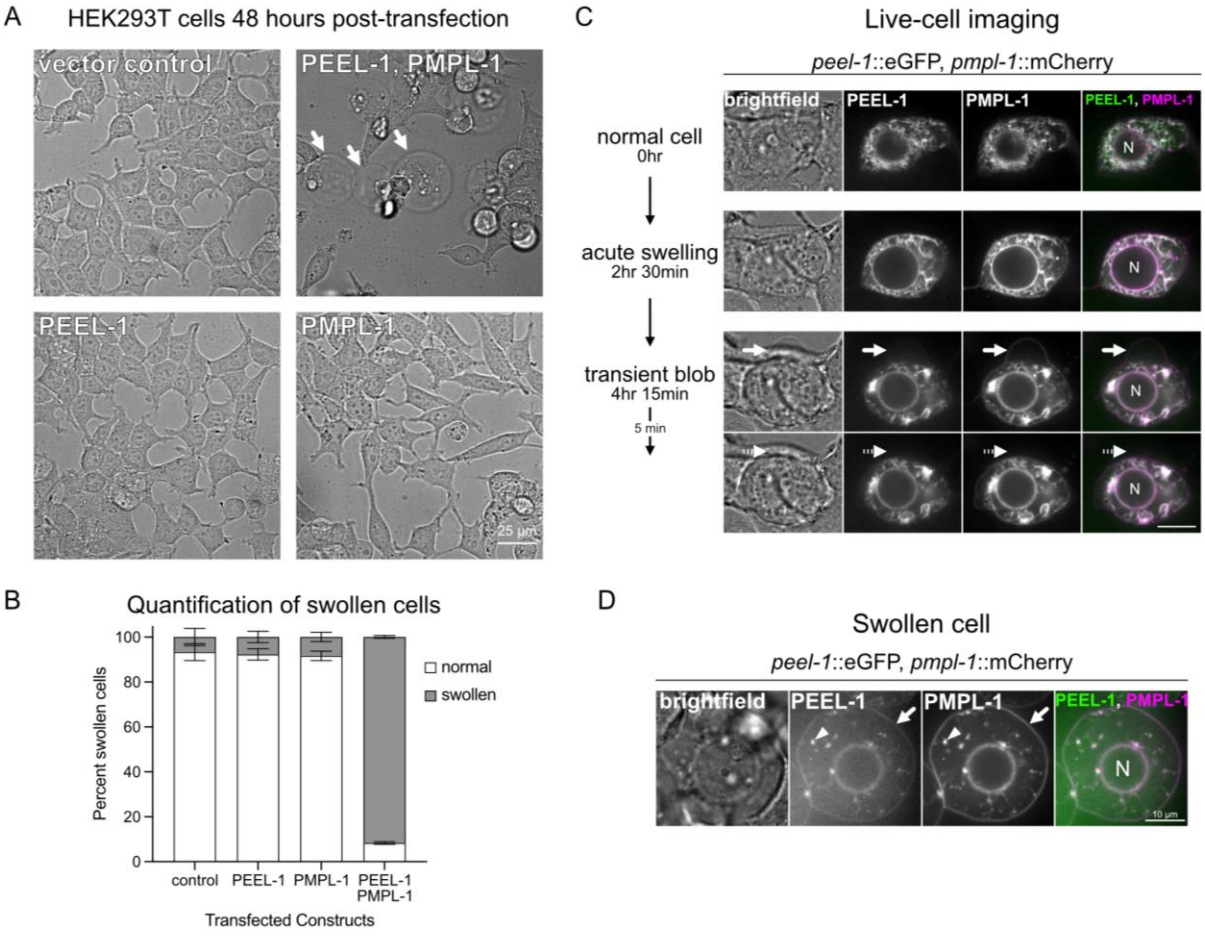


Fig. 2.3. PEEL-1 toxicity results in HEK293T cell swelling.

(A) Brightfield images of cells transfected with indicated constructs. Cells were imaged 48 hours after transfection. Scale bar = 25µm. Arrows point to swollen, transfected cells. **(B)** Percent of transfected cells that appear normal or swollen under brightfield. Mean with SD of three biological replicates is shown. 100 cells were scored for each condition in each biological replicate. **(C)** Selected frames of a live-cell imaging time-course experiment. A single cell is shown, co-expressing PEEL-1::eGFP and PMPL-1::mCherry. The nucleus is labeled ("N"). t=0 is 16 hours after transfection. Acute swelling can be seen (t=2hr 30min), followed by a transient blob or protrusion (arrow) jetted out by the cell (t=4hr 15min) and later reabsorbed (dotted arrow) (t=4hr 20min). Scale bar = 10µm. **(D)** A typical phenotype from a cell transfected with

PEEL-1::eGFP and PMPL-1::mCherry at 48 hours post-transfection. PEEL-1 and PMPL-1 can be seen on the plasma membrane (arrow) and in fragmented ER (arrowhead). Scale bar = 10μm.

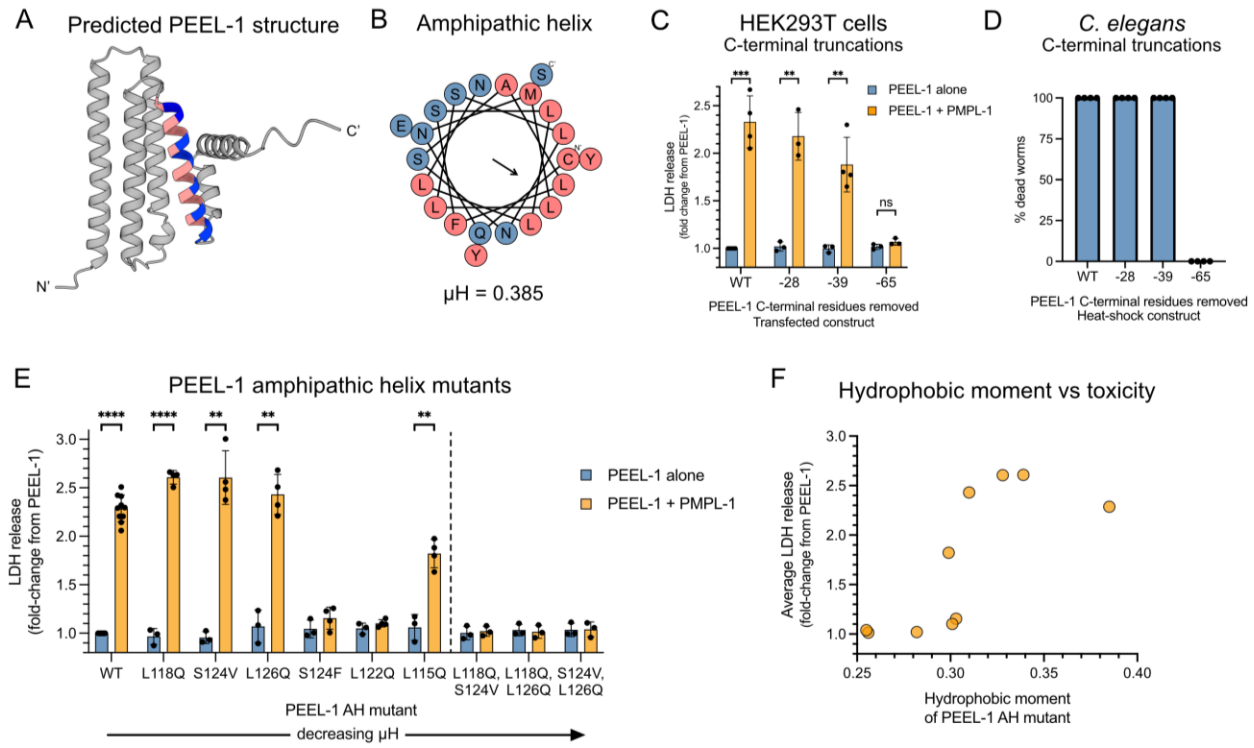


Fig. 2.4. PEEL-1's amphipathic helix is critical for toxicity.

(A) The AlphaFold2 predicted structure of PEEL-1 and **(B)** a helical wheel representation of the putative PEEL-1 amphipathic helix. Amphipathic helix residues are colored (pink = hydrophobic, blue = hydrophilic). **(C)** Cytotoxicity of a series of PEEL-1 C-terminal truncations expressed in HEK293T cells. The number of amino acids removed are indicated (ex. "-28" means the last 28 residues were removed). Each truncation removes an additional alpha helix. The "-65" truncation removes the amphipathic helix. **(D)** Percent dead worms after heat-shock PEEL-1 expression of the indicated truncation mutant. 50 worms were assayed for each data point, and two independent transgenic lines were tested or each construct. **(E)** Cytotoxicity of PEEL1 amphipathic helix missense mutants in HEK293T cells. Mutants are ordered by descending hydrophobic moment (μH). Six single mutants (left of dotted line) and three double mutants are shown (right of dotted line). All bar graphs show mean with SD. Statistics were

performed using multiple unpaired t-tests with Holm-Šídák test, comparing each PEEL-1 alone to PEEL-1 and PMPL-1 (**, $p < 0.01$; ***, $p < 0.001$; ****, $p < 0.0001$). **(F)** The average toxicity of missense mutants from (E) plotted against their hydrophobic moment.

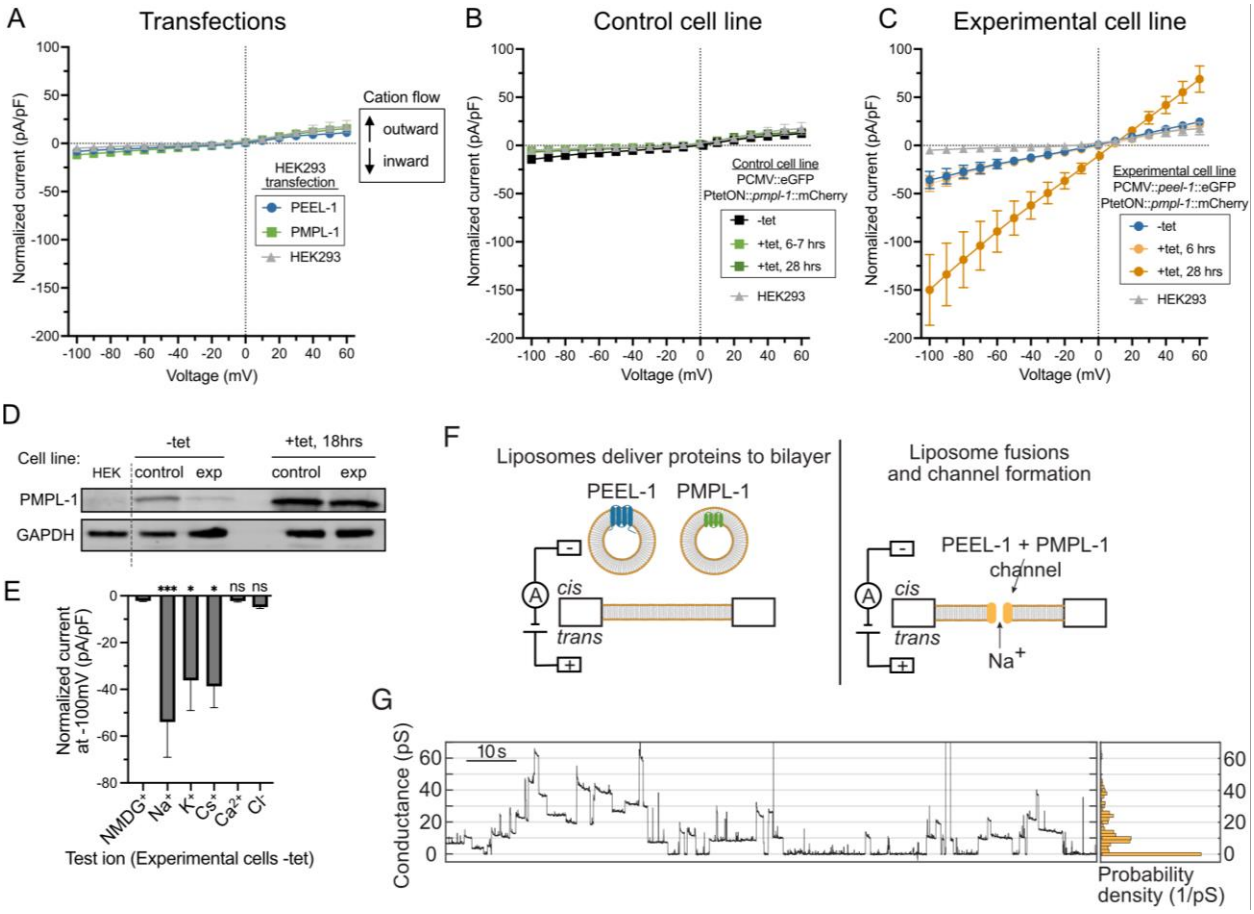


Fig. 2.5. PEEL-1 and PMPL-1 create a monovalent cation channel.

Current-voltage plots of whole-cell patch-clamp electrophysiology on **(A)** transfected cells, **(B)** Control cell line, and **(C)** Experimental cell line. High intracellular potassium (140mM K^+/ 8.6mM Na^+) and high extracellular sodium (145mM Na^+/ 4mM K^+) solutions are used. Currents elicited by a family of 0.5 second voltage steps from a -30mV holding potential, from -100mV to 60mV, in 10mV increments. Currents normalized to cell capacitance (pF). Negative and positive currents indicate cation flow into or out of the cell, respectively. **(A)** Plots from HEK293 cells acutely transfected with *peel-1*::eGFP or *pmpl-1*::mCherry. **(B)** Plots from Control cell line expressing constitutive eGFP and tetracycline-inducible *pmpl-1*::mCherry. **(C)** Plots from Experimental cell line expressing constitutive *peel-1*::eGFP and tetracycline-inducible *pmpl-*

1::mCherry. Control and Experimental cell lines are shown without tetracycline or with tetracycline at the indicated time after addition of tetracycline. All plots shown as mean with SEM. **(D)** Western blot for PMPL-1::mCherry in Control and Experimental cell lines without tetracycline (-tet, left) and 18 hours after addition of tetracycline (+tet, right). Leaky expression of PMPL-1 is seen in both cell lines in the absence of tetracycline. Less background PMPL-1 expression is seen in Experimental cells than in Control cells, likely because of selection against higher background PMPL-1 expression when in combination with PEEL-1 but not eGFP. GAPDH loading control shown. **(E)** Permeability of indicated ions was assayed in Experimental cells without tetracycline. Test ionic solutions substituted previous bath solution (145mM Na^+/ 4mM mM K^+) with 140mM pure cations (external) or 20mM anions (internal), except Ca^{2+} (20mM external, 1mM internal Ca^{2+} with 2.5mM EGTA). All plots show mean with SEM. Statistical tests compare all results to NMDG (treated as control) in one-way ANOVA with Dunnett's multiple comparisons test (*, $p < 0.05$; ***, $p < 0.001$). **(F)** Schematic of planar lipid bilayer experiment. Liposomes containing purified PEEL-1 or PMPL-1 are added to the *cis* well to deliver proteins to the lipid bilayer. **(G)** Conductance traces of one experiment (left) and a histogram of the trace (right, 2pS bin width, normalized based on probability density) at an applied voltage of +180mV. SDS-PAGE gels of purified proteins are shown in fig. 2.S15 and more example conductance traces and controls shown in fig. 2.S16.

2.7 SUPPLEMENTARY MATERIAL

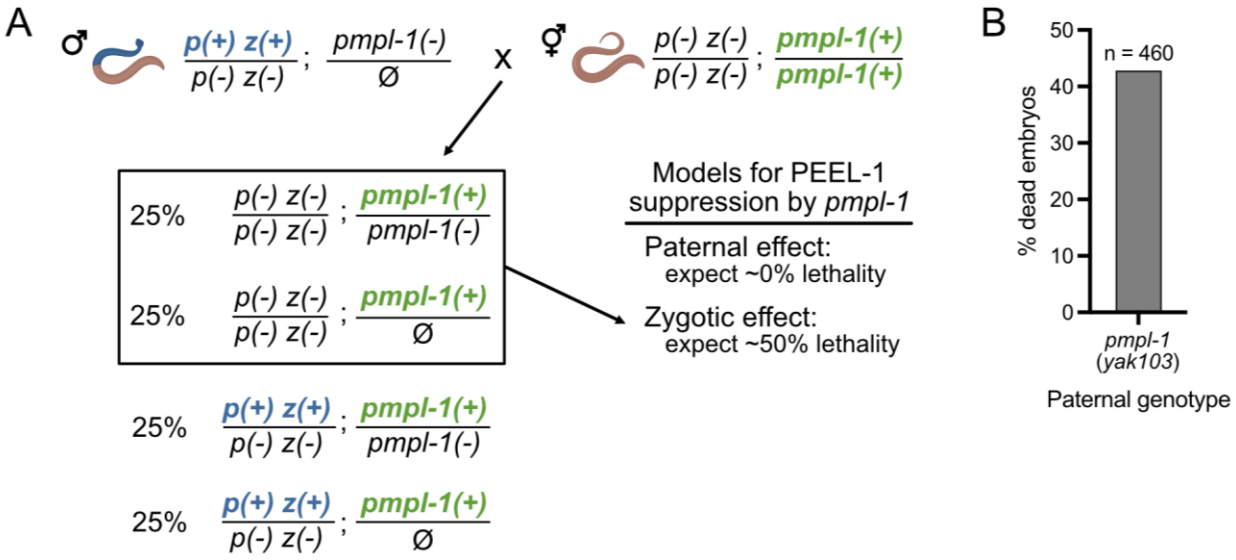


Fig. 2.S1. *pmpl-1(yak103)* does not act through paternal-effect

(A) Genetic cross to test for paternal versus zygotic effect of sperm-delivered PEEL-1 suppression by *pmpl-1(yak103)* deletion mutant (denoted *pmpl-1(-)*). Males heterozygous for the selfish element and carrying *pmpl-1(-)* are mated to *p(-) z(-); pmpl-1(+)* hermaphrodites. Suppression via paternal effect would result in ~0% embryonic lethality while suppression by a zygotic effect would result in ~50% lethality. **(B)** The percent of dead embryos seen from the cross shown in panel (A). Results are consistent with a model of *pmpl-1*(*yak103*) suppression of PEEL-1 through zygotic effect.

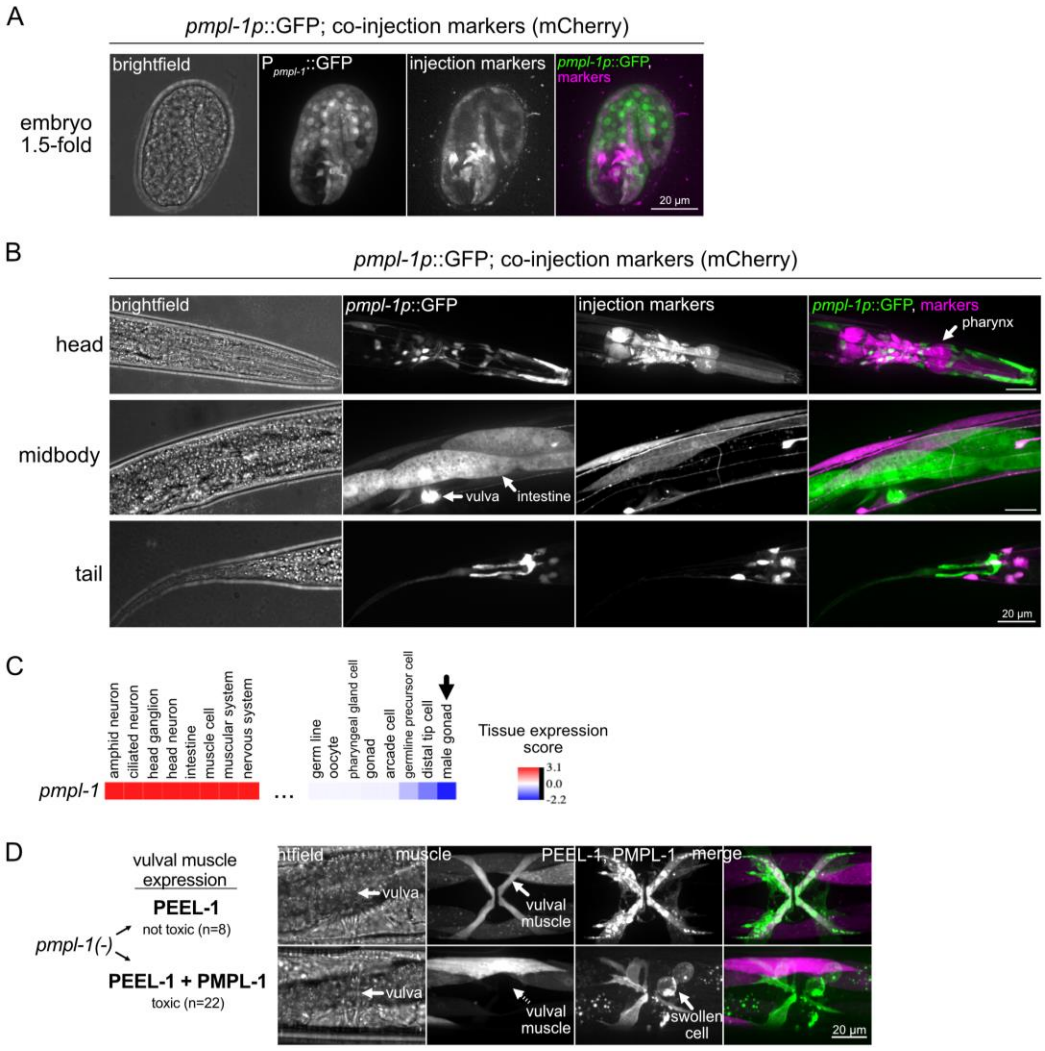


Fig. 2.S2. *pmpl-1* expression pattern in C. elegans.

(A) Maximum intensity projection of a *C. elegans*, 1.5-fold embryo with GFP driven by the *pmpl-1* promoter and co-injection markers *myo-2p*::mCherry, *rab-3p*::mCherry, and *myo-3p*::mCherry. Toxicity from sperm-delivered PEEL-1 occurs after this embryonic stage. **(B)** Maximum intensity projections of hermaphrodite adult head (top), midbody (middle), and tail (bottom) of the same strain shown in panel (A). **(C)** Heatmap of *pmpl-1* of tissue expression scores from RNA-seq dataset of Day 1 adult worms from Kaletsky et al., 2018. Tissues with the

highest *pmpl-1* expression (red) and lowest expression (blue) are shown. Only the 8 most highly expressed tissues (left) and the 8 most lowly expressed tissues (right) are shown. Expression of *pmpl-1* is lowest in the male gonad (arrow). **(D)** *pmpl-1(yak103)* worms with vulval muscle cell expression of *peel-1*::GFP alone (top) or with *pmpl*-1::GFP (bottom). Number of scored worms are indicated. Green channel brightness is increased in the bottom panel to show fluorescence in the vulval muscle, since toxicity in this cell likely caused decreased levels of fluorescent-tagged proteins.The vulval muscle appears swollen in the green channel when co-expressing *peel-1* and *pmpl-1*. This is the same worm as shown in Figure 2.1E.

Fig. 2.S3. PMPL-1 is conserved in nematodes.

Alignment of PMPL-1 amino acid sequences (right) between representative *Caenorhabditis* species, *Pristionchus pacificus*, and *Toxocara canis* shown in a species phylogeny (left). The *pmpl-1* (*yak52*) allele is mutated at conserved residue A47 (*).

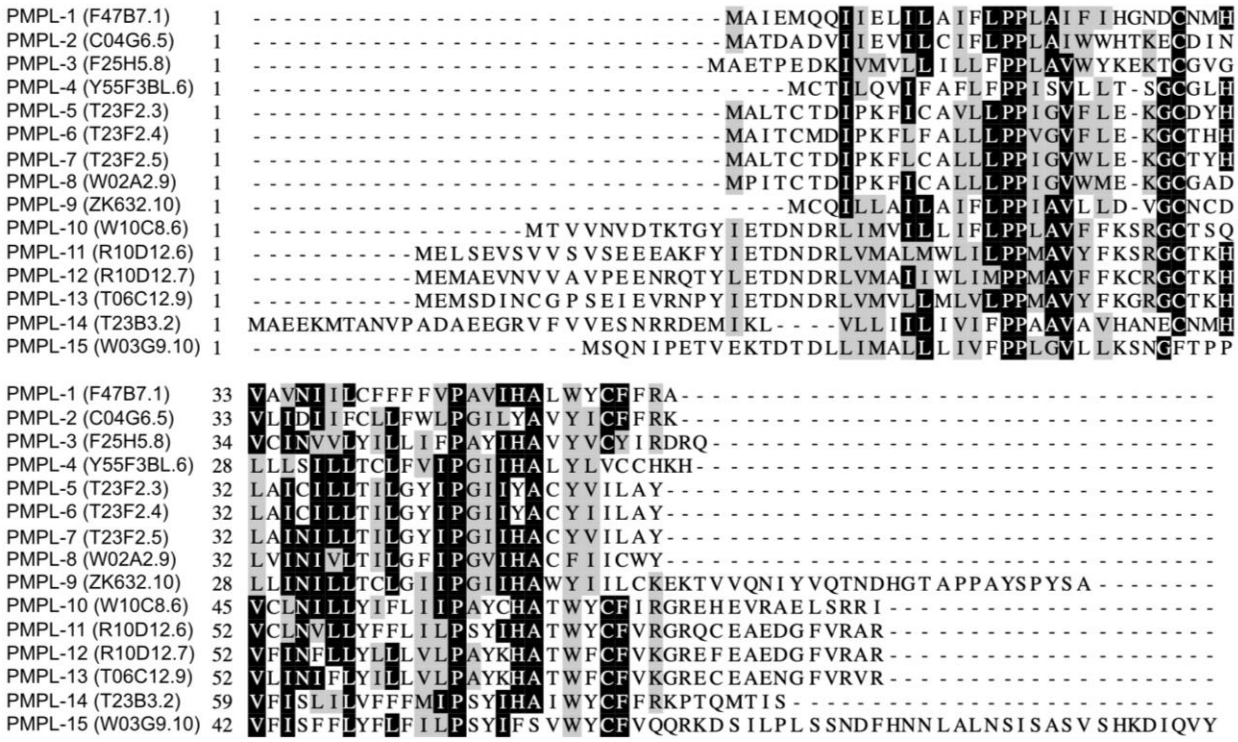


Fig. 2.S4. 15 PMP3-like proteins in C. elegans.

Alignment of all 15, PMP3-like proteins in C. elegans identified by BLAST.

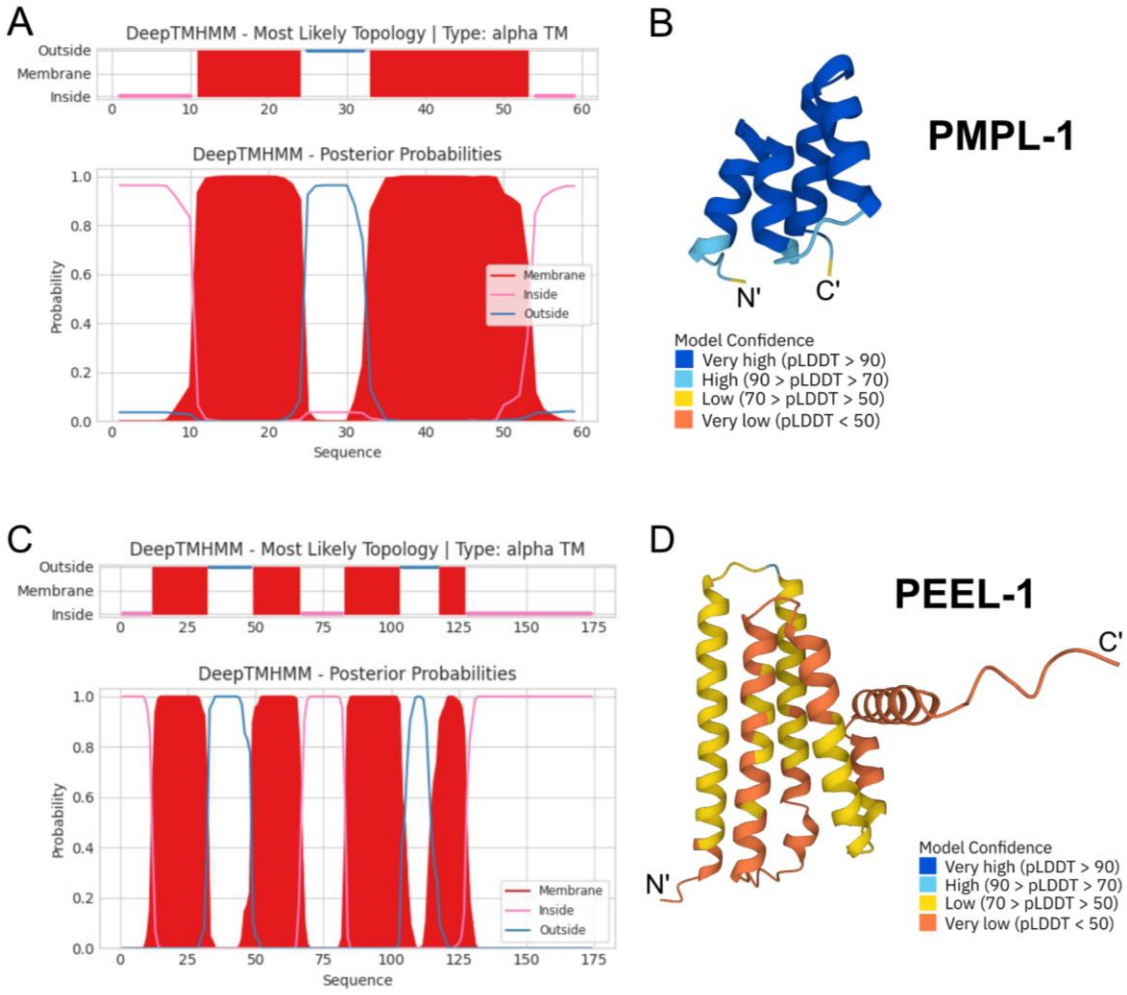


Fig. 2.S5. Structural predictions of PEEL-1 and PMPL-1.

DeepTMHMM predictions of **(A)** PMPL-1 and **(C)** PEEL-1. PMPL-1 is predicted to be a two-pass transmembrane protein. PEEL-1 is predicted to be a four-pass transmembrane protein. Both proteins are predicted to have their N- and C-termini facing the cytosol. Structural predictions from AlphaFold2 of **(B)** PMPL-1 and **(D)** PEEL-1. AlphaFold2 predicts a very high-confidence PMPL-1 structure that suggests it is a monotopic protein (passing through one leaflet of a lipid bilayer). The PEEL-1 structure is of low and very-low confidence, likely because there are no known homologs of this protein.

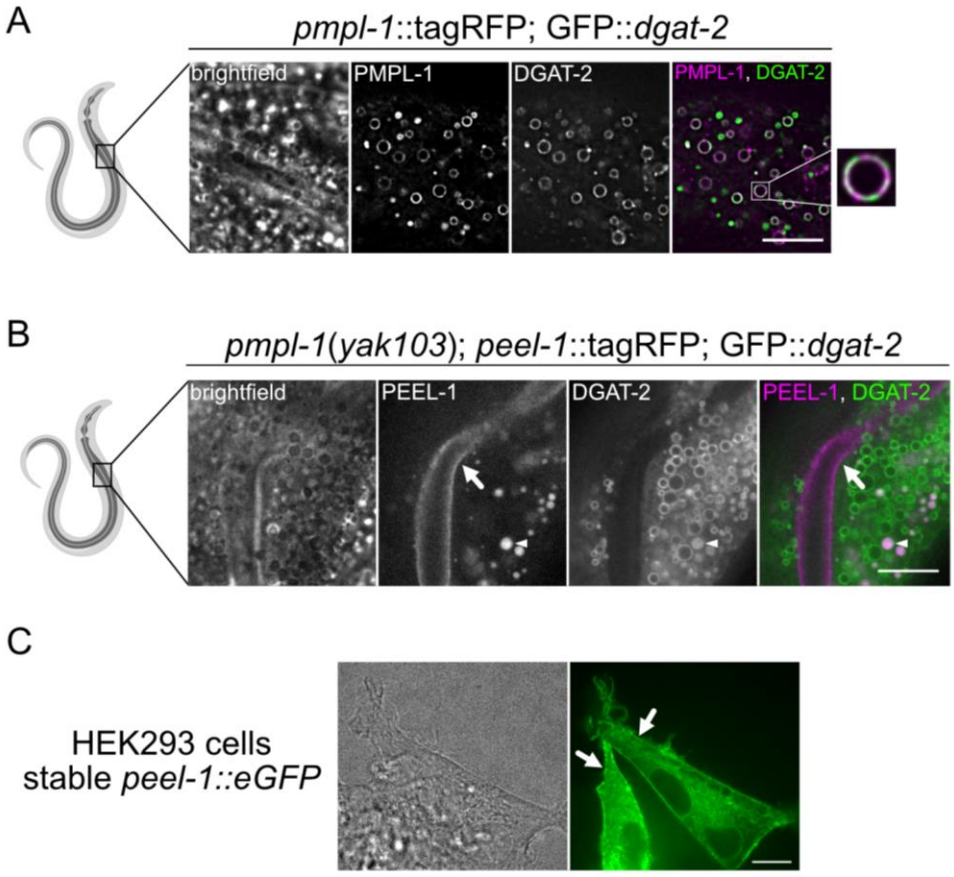


Fig. 2.S6. PEEL-1 and PMPL-1 localization.

(A-B) Intestinal cells of an adult *C. elegans* with the indicated constructs. **(A)** Wild-type worms expressing *pmpl-1*::tagRFP and GFP::*dgat-2*. DGAT-2 localizes to lipid droplet membranes, and PMPL-1::tagRFP co-localizes to these organelles. Inset shows one lipid droplet. **(B)** *pmpl-1*(*yak103*) expressing GFP::*dgat-2* and *peel-1*::tagRFP. PEEL-1 signal (arrow) appears on plasma membrane lining the intestinal lumen and does not co-localize with lipid droplets. Autofluorescence from gut granules appears as filled-in circles in both channels (arrowhead). **(C)** HEK293 cells stably expressing *peel-1*::eGFP. PEEL-1::eGFP localizes to the ER and plasma membrane (arrows). Scale bar = 10µm.

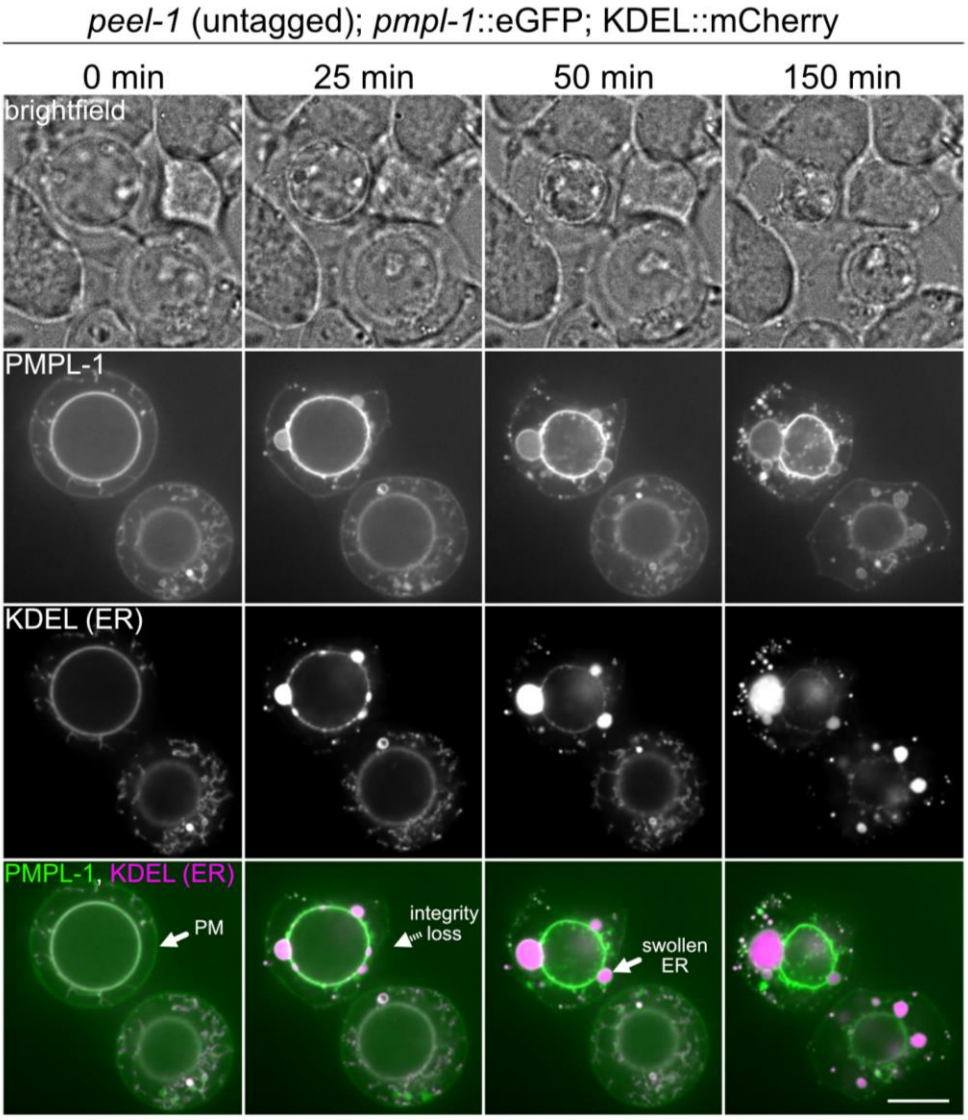


Fig. 2.S7. Toxicity causes ER swelling, ER fragmentation, and cell lysis.

Live-cell imaging time course of two HEK293T cells transfected with constructs coding for PEEL-1 (untagged), PMPL-1::eGFP, and mCherry::KDEL (ER marker). Imaging began at 20 hours post-transfection (t=0min).The top cell experiences a loss of plasma membrane (PM) integrity at 25 min, followed by ER swelling. The lower cell loses PM integrity at 150 min. Scale bar = 10μm.

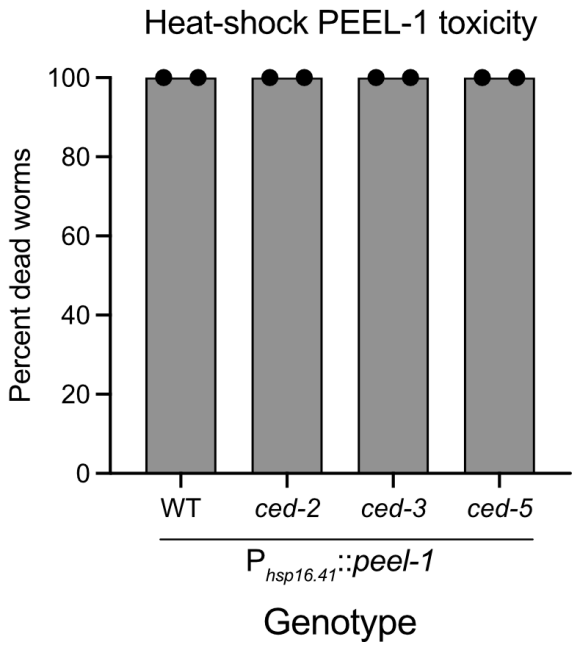


Fig. 2.S8. Ectopic PEEL-1 toxicity is non-apoptotic.

Percent dead worms after heat-shock induced expression of PEEL-1. Worms deficient in apoptosis (*ced-3*) and cell engulfment (*ced-2* and *ced-5*) were tested. Two independent experiments were done, with n=50 worms for each data point.

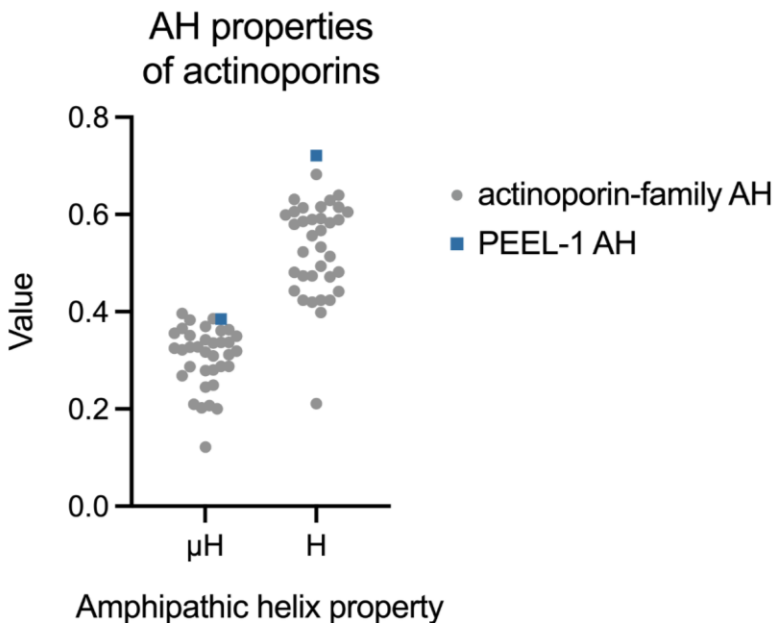


Fig. 2.S9. Amphipathic helix properties of Actinoporin toxins.

Hydrophobic moment (μH) and hydrophobicity (H) of the amphipathic helix of 35 Actinoporin toxin proteins (data acquired from Macrander and Marymegan et al., 2016) (gray circles) and the predicted PEEL-1 amphipathic helix (blue square).

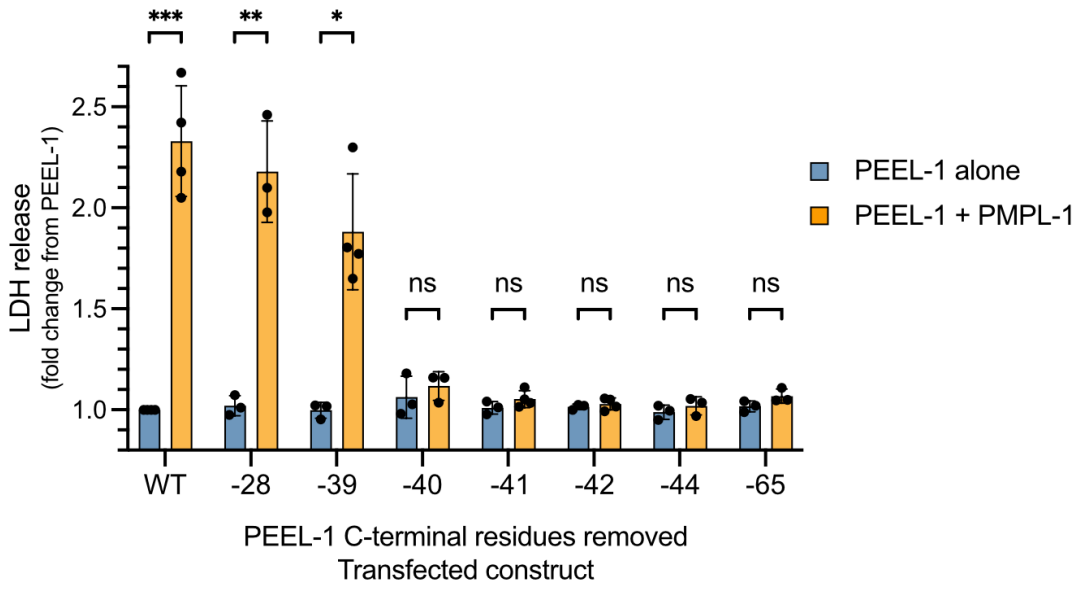


Fig. 2.S10. Toxicity of PEEL-1 C-terminal truncations.

Cytotoxicity of PEEL-1 C-terminal truncations assayed alone (blue bars) or with PMPL-1 (yellow bars) in HEK293T cell transfections. The number of amino acids removed from the C-terminus is indicated (ex. -28 means 28 amino acids were removed). Data from PEEL-1 WT, -28, -39, and -65 are from Figure 4C. Toxicity is lost upon removal of the -40 residue (Ala135). Statistical tests done using multiple unpaired t-tests with Holm-Šídák test, comparing each PEEL-1 alone to PEEL-1 and PMPL-1 (*, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$).

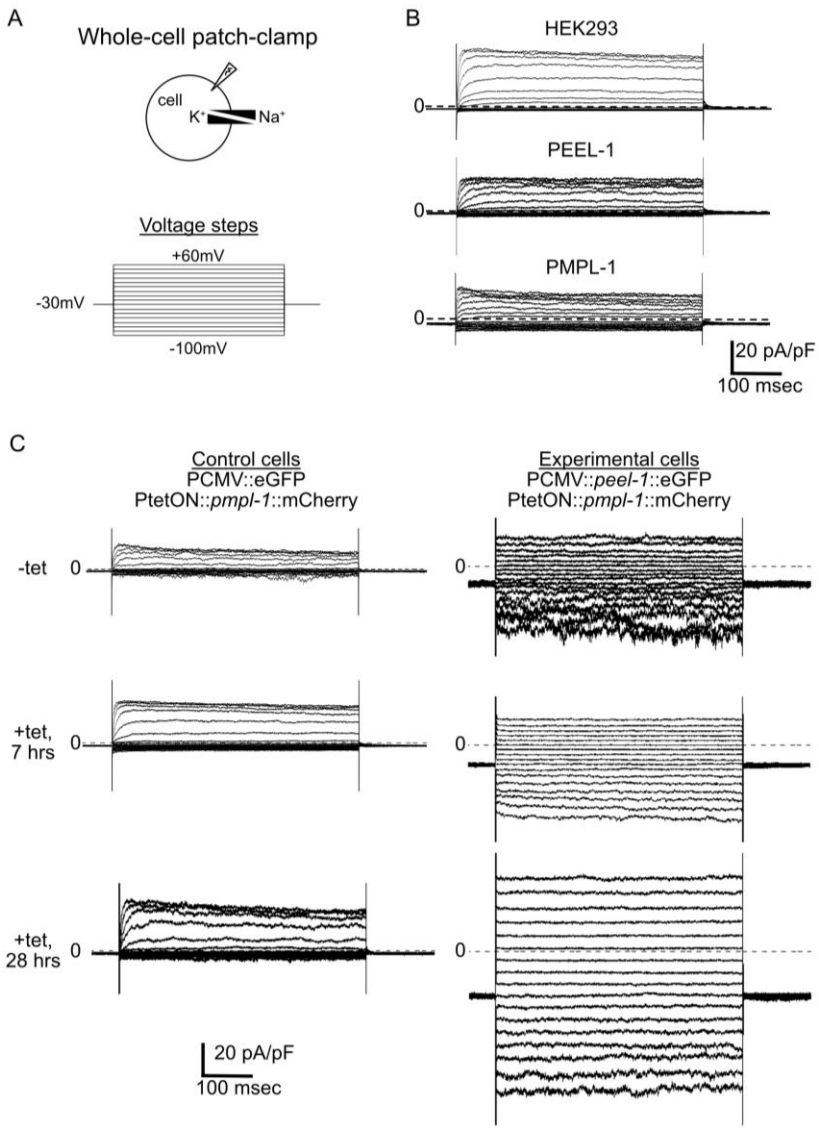


Fig. 2.S11. Raw traces of HEK293 electrophysiology experiments.

(A) Schematic of electrophysiology experiment. High intracellular potassium (140mM K^+/ 8.6mM Na^+) and high extracellular sodium (145mM Na^+/ 4mM K^+) solutions are used. Currents elicited by a family of 0.5 second voltage steps from a -30mV holding potential, from -100mV to 60mV, in 10mV increments. Current traces are normalized to capacitance (pF). **(B)** Representative traces of untransfected HEK293 cells (top) and HEK293 cells transfected with *peel-1*::eGFP (middle) or *pmpl-1*::mCherry (bottom). Scale bar shown (bottom-right). **(C)** Representative traces of tetracycline-inducible cells lines. Control cells (left) have stable

expression of eGFP and inducible expression of *pmpl-1*::mCherry. Experimental cells (right) have stable expression of *peel-1*::eGFP and inducible expression of *pmpl-1*::mCherry. Recordings of cell lines without induction (top), 7 hours after tetracycline addition (middle), and 28 hours after tetracycline addition (bottom) are shown.

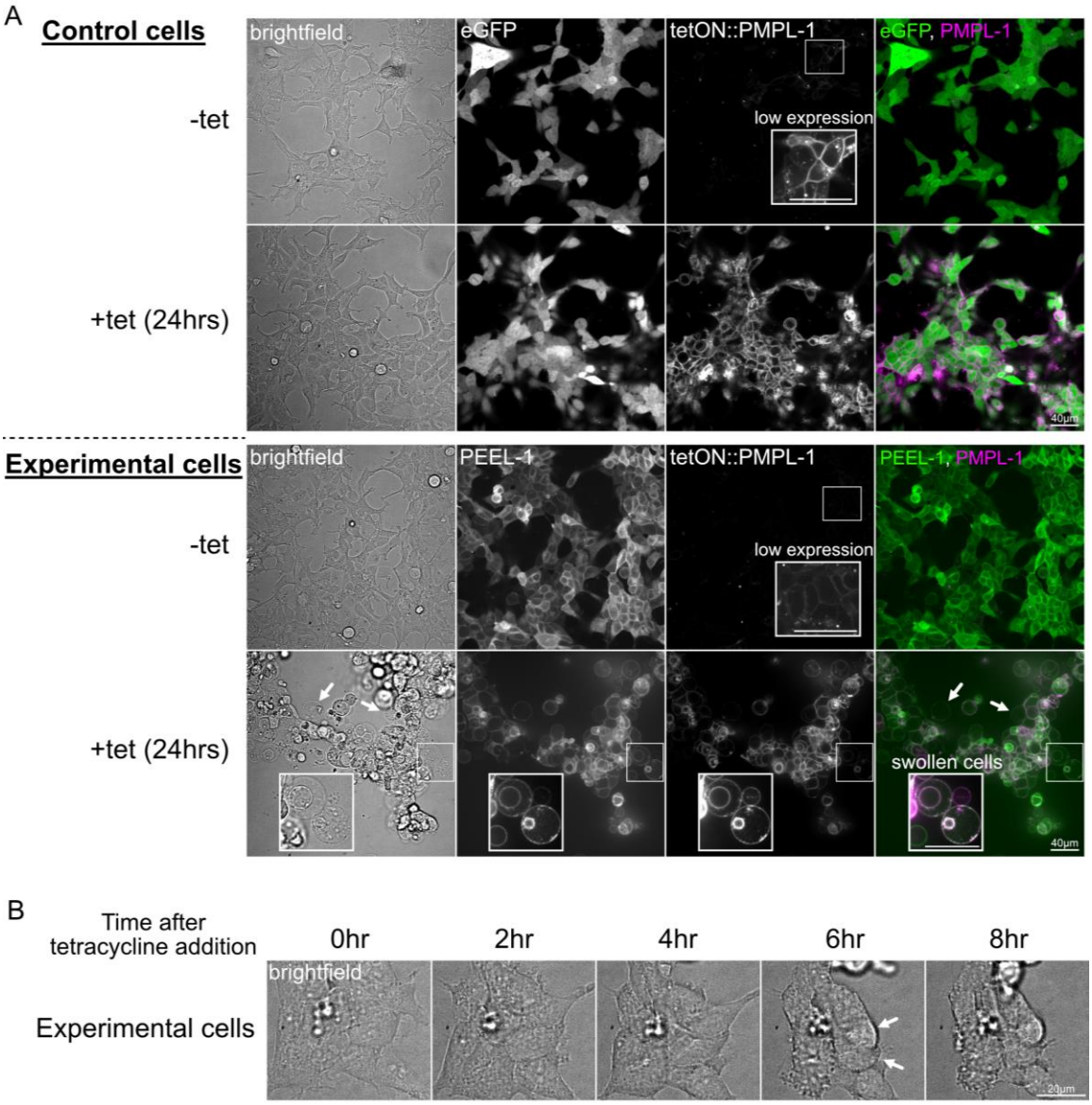


Fig. 2.S12. Tetracycline-induced toxicity in stable cell lines.

(A) Live-cell images of Control cells (CMV-driven eGFP; tetON::*pmpl-1*::mCherry) and Experimental cells (CMV-driven *peel-1*::eGFP; tetONp::*pmpl-1*::mCherry). Cell lines are shown without tetracycline (-tet) and 24 hours after tetracycline addition (+tet (24hrs)). Insets in -tet conditions show leaky expression of PMPL-1 in both cell lines (LUT adjusted within inset). Images show successful tetracycline-inducible expression of PMPL-1::mCherry and efficient tetracycline-induced killing in experimental cells but not control cells. Arrows and inset in

experimental cells +tet (24hrs) show examples of swollen cells. Exposure time in the green channel is different between cell lines. Scale bar = 40μm. **(B)** Time course of toxicity after addition of tetracycline to experimental cells. Noticeable cell swelling is seen after 6 hours (arrows). Some acute swelling may also be visible at 4 hours after addition of tetracycline. Scale bar = 20μm.

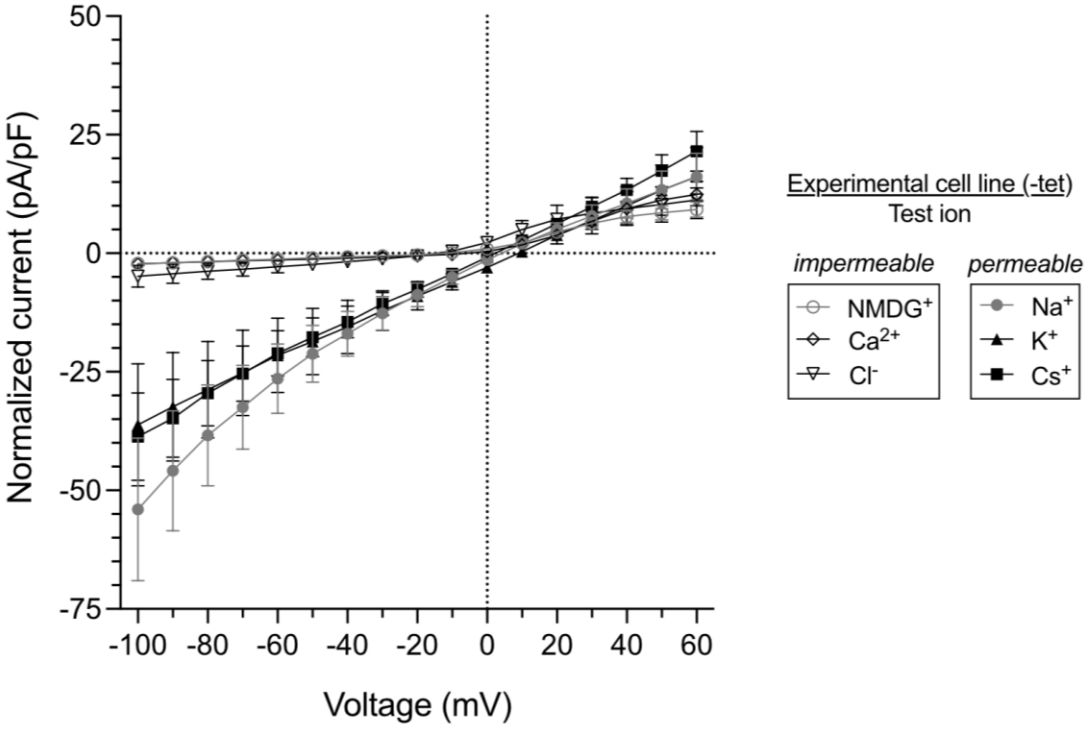


Fig. 2.S13. Ion permeabilities in uninduced experimental cells.

Current-voltage plots of experimental cells without tetracycline in different ionic conditions to test permeabilities of the indicated ions. Permeable ions have greater inward currents (Na^+, K^+, Cs^+; closed symbols) compared to impermeable ions ($NMDG^+$, Ca^{2+}, Cl^-; open symbols) at negative voltages. Mean with SEM is shown. These data are summarized in Fig. 2.5E.

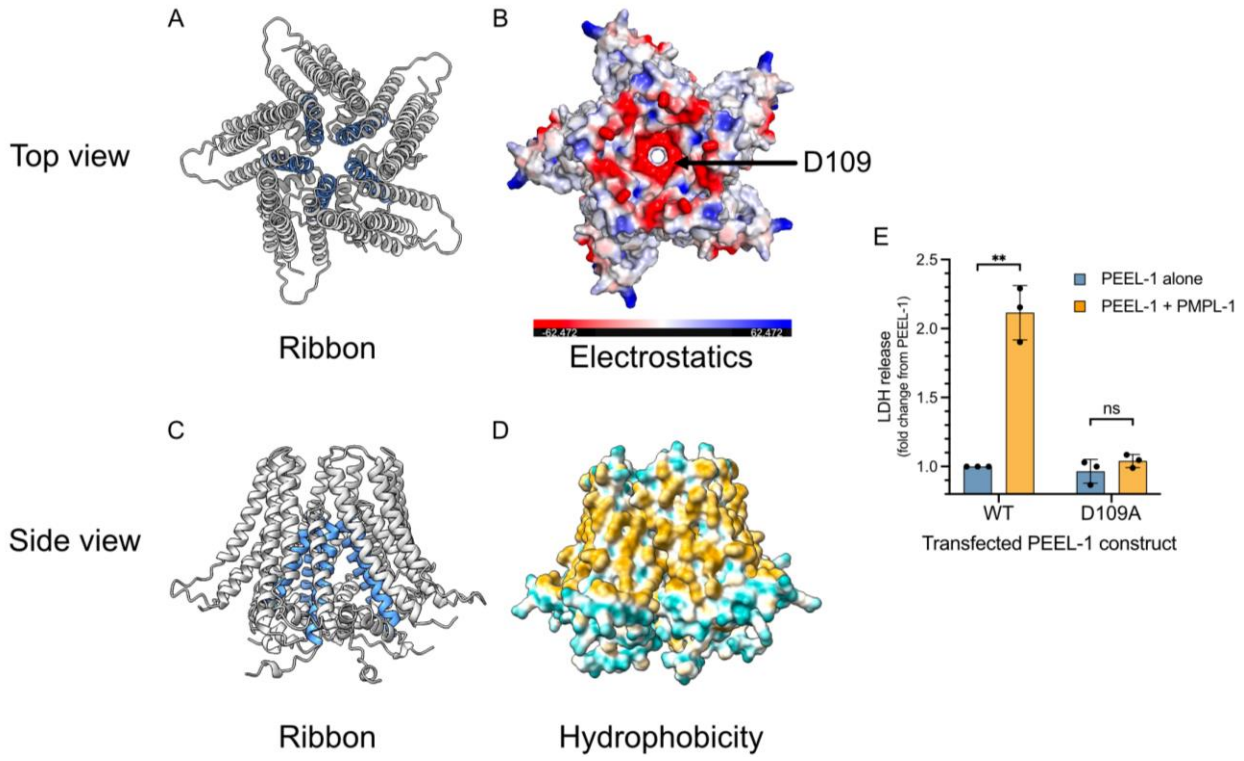


Fig. 2.S14. Predicted PEEL-1 pentameric structure.

AlphaFold2 prediction of the PEEL-1 pentameric structure is shown. Two angles are shown: **(A-B)** top view, showing the predicted extracellular face of the complex, and **(C-D)** side view, in the plane of the lipid bilayer. **(A and C)** Ribbon diagram with the amphipathic helix colored in blue creating the lining of a pore-like region. **(B)** Surface representation of electrostatic predictions (red = negative charge, blue = positive charge). An uninterrupted hole can be seen through the structure, with a ring of negative charge from five D109 residues. **(D)** Surface representation of hydrophobicity (yellow = hydrophobic, cyan = hydrophilic). **(E)** Cytotoxicity of mutants which eliminate the predicted ring of negative charge at the top of the complex via a D109A mutation. Statistics performed using multiple unpaired t-tests with Holm-Šídák test. All tested comparisons are shown.

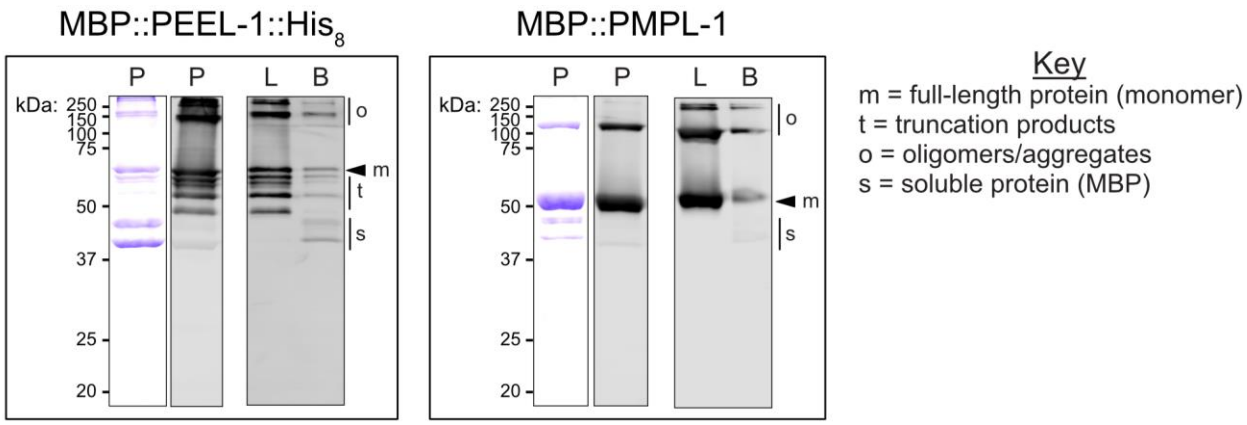


Fig. 2.S15. PEEL-1 and PMPL-1 purification and liposomes.

SDS-PAGE gels of indicated constructs after purification (P; Commassie Blue stain and anti-MBP western blot) and after incorporation into liposomes (L; anti-MBP western blots). Liposomes were collected after flotation through a density gradient before use in synthetic bilayer experiments. The top fraction contains liposomes (L) and the bottom fraction (B) contains unincorporated protein. Monomers (m), oligomers or aggregates (o), and truncation products of PEEL-1 (t) are indicated. The bottom bands seen in both MBP-tagged PEEL-1 and PMPL-1 purifications are soluble proteins (s) that did not get incorporated into liposomes, likely endogenous MBP (43kDa) or truncated MBP-tagged protein.

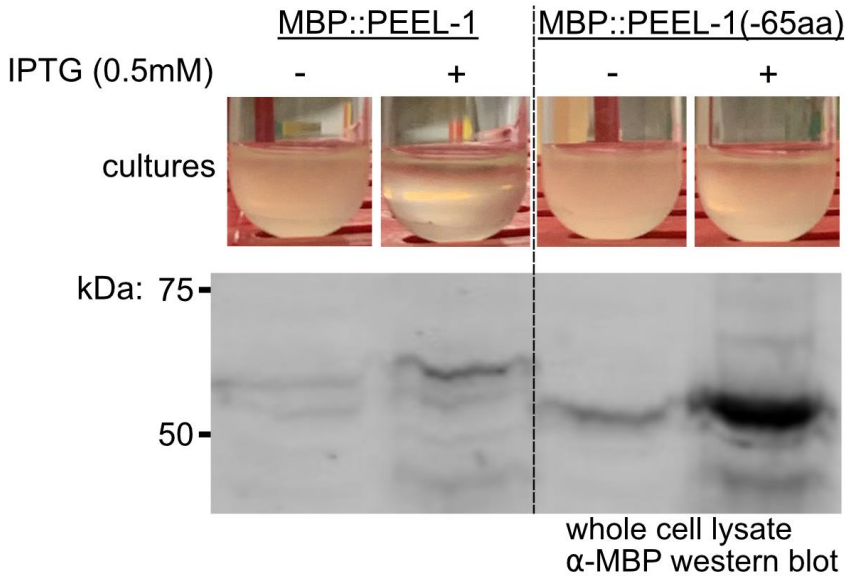


Fig. 2.S16. PEEL-1 is toxic to bacteria and requires the amphipathic helix.

E. coli C41(DE3) cells with constructs encoding IPTG-inducible expression of either MBP::PEEL-1 (left) or MBP::PEEL-1(-65aa) (right). Images of cultures are taken after shaking at 18°C overnight, with or without 0.5mM IPTG. Cultures appear lysed when full-length PEEL-1 is expressed but not PEEL-1(-65aa) which lacks the PEEL-1 AH. Western blot (anti-MBP) of corresponding whole-cell lysates are shown (bottom), confirming higher expression of MBP::PEEL-1(-65aa). This experiment was repeated four times and yielded similar results.

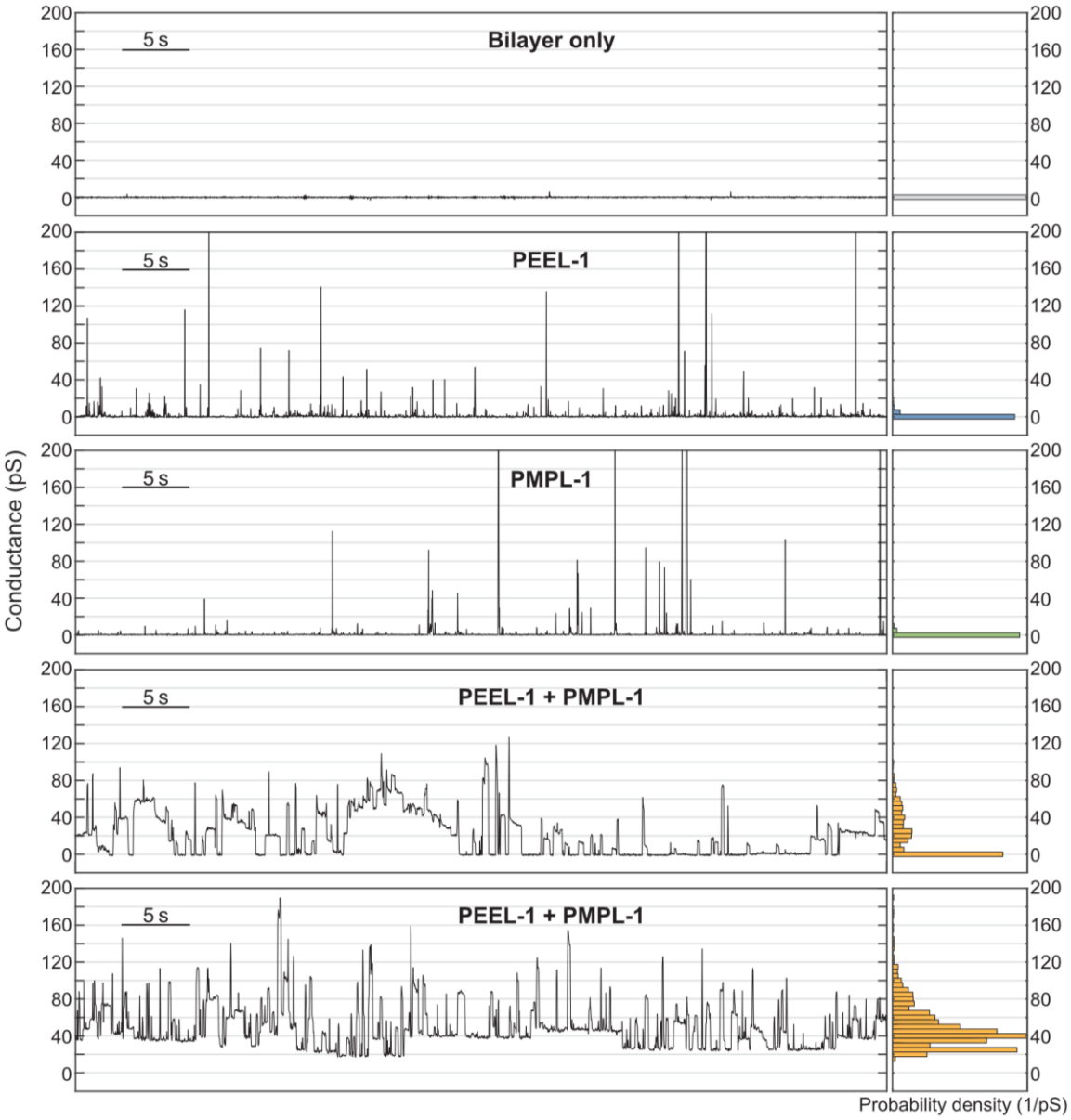


Fig. 2.S17. Purified PEEL-1 and PMPL-1 conduct ions through planar lipid bilayers.

Conductance traces through artificial planar lipid bilayers are shown. (top to bottom) Bilayer alone without addition of liposomes, bilayer with PEEL-1 liposomes added (transient spikes indicate successful liposome fusions), bilayer with PMPL-1 liposomes added, and two independent experiments of bilayers with PEEL-1 liposome and PMPL-1 liposomes added. An all-point histogram is shown (right, 2pS bin width, normalized based on probability density).

Channel activity after addition of PEEL-1 and PMPL-1 was observed in 8 independent experiments. A voltage of -180mV was applied to the bilayer for liposome fusions in all experiments. After observing channel activity, a voltage of +180mV was applied since channel activity was more stable at positive voltages. Bottom two panels show traces at +180mV. Scale bar = 5 seconds.

Table 2.S1. *C. elegans* strains used in this study.

Strain ID	Strain genotype	Source
N2	N2	CGC
XZ1372	yakTi4[hsp16.41p::eGFP::his-44 , NeoR] I ; oxSi507[hsp-16.41p::peel-1, Cb-unc-119] II ; oxSi280[hsp16.41p::peel-1, Cb-unc-119] IV	Crawford et al., 2023
XZ1047	oxSi507[hsp-16.41p::peel-1, Cb-unc-119] II ; unc-119(ed9) III ; oxSi280[hsp16.41p::peel-1, Cb-unc-119] IV ; him-5(e1490) V	Crawford et al., 2023
XZ103	oxSi507[hsp-16.41p::peel-1, Cb-unc-119] II ; oxSi280[hsp16.41p::peel-1, Cb-unc-119] IV ; pmpl-1(yak103) X	this study
XZ2283	oxSi507[Phsp-16.41::peel-1, Cb-unc-119] II ; oxSi280[Phsp16.41::peel-1, Cb-unc-119] IV ; F47B7.1(yak52) X	this study
AFS216	zeel-1(tm3419) I peel-1(cle6) I	Aaron Severson
XZ1177	oxSi507[hsp-16.41p::peel-1, Cb-unc-119] II ; unc-119(ed9) III ; oxSi280[hsp16.41p::peel-1, Cb-unc-119] IV ; him-5(e1490) V ; pmpl-1(yak52) X	this study
XZ1307	oxSi507[hsp-16.41p::peel-1, Cb-unc-119] II ; oxSi280[hsp16.41p::peel-1, Cb-unc-119] IV ; him-5(e1490) V ; pmpl-1(yak103) X	this study
EG1000	dpy-5(e61) I ; rol-6(e187) II ; lon-1(e1820) III	Erik M. Jorgensen
EG1020	bli-6(sc16) IV ; dpy-11(e224) V ; lon-2(e678) X	Erik M. Jorgensen
EG8040	oxTi302[Peft-3::mCherry cb-unc-119(+)] I ; oxTi75[Peft-3::GFP::H2B::tbb-2utr unc-18(+)] II ; oxTi411[Peft-3::TdTomato::H2B::unc-54 cb-unc-119(+)] III ; him-8(e1489) IV	Jorgensen lab
EG8041	oxTi76[Peft-3::GFP::H2B::tbb-2utr unc-18(+)] IV ; oxTi405[Peft-3::TdTomato::H2B::unc-54 cb-unc-119(+)] V him-5(e1490) V ; oxTi421[Peft-3::mCherry cb-unc-119(+)] X	Jorgensen lab
XZ2194	pmpl-1 (yak103) X	this study
XZ2103	oxSi507[hsp-16.41p::peel-1, Cb-unc-119] II ; ced-3(n717) IV ; him-5(e1490) V	this study
XZ2102	oxSi507[hsp-16.41p::peel-1, Cb-unc-119] II ; ced-5(n1812) IV	this study
XZ2096	oxSi507[hsp-16.41p::peel-1, Cb-unc-119] II ced-2(n1994) IV	this study
XZ2254	yakEx195[pmpl-1p::GFP; myo-2p::mcherry; myo-3p::mcherry; rab-3p::mCherry]	this study
XZ2276	pmpl-1(yak103) X ; yakEx203[exp-3p::peel-1::GFP, myo3p::mCherry]	this study
XZ2633	pmpl-1(yak103) X ; yakEx275[exp-3p::peel-1::GFP, exp-3p::pmpl-1::GFP,	this study
XZ2551	yakEx264[hsp16.41p::peel-1(-28aa), cc::GFP]	this study
XZ2634	yakEx276[hsp16.41p::peel-1(-39aa), cc::GFP]	this study
XZ2548	yakEx263[hsp16.41p::peel-1(-65aa), cc::GFP]	this study
XZ2454	pmpl-1(yak103) X ; yakEx275[exp-3p::peel-1::GFP, exp-3p::pmpl-1::GFP, myo3p::mCherry]	this study
XZ2452	hjSi56[Pvha-6::3xFLAG::TEV::GFP::dgat-2::let-858 3' UTR] IV ; yakEx242[vha-6p::pmpl-1::tagRFP, myo-2p::mCherry]	this study

Table 2.S2. Constructs used in this study.

Construct ID	Description	resistance	sequenced?	notes
pLC4	exp-3p::pmpl-1::tagRFP::tbb-2 3'UTR	carb		
pLC6	exp-3p::peel-1::GFP::tbb-2 3'UTR	carb		
pLC26	MBP::TEV::peel-1	carb	sequenced	
pLC28	MBP::TEV::pmpl-1	carb	sequenced	
pLC31	exp-3p::pmpl-1::GFP tbb-2 3'UTR	carb		
pLC37	PMP3(S. cerevisiae)::mCherry_N1	kan	sequenced	
pLC38	mCherry::zeel-1_N1	kan		
pLC54	pmpl-1::eGFP_N1	kan		
pLC65	MBP::TEV::peel-1	carb+chlor	sequenced	
pLC67	MBP::TEV::pmpl-1	carb+chlor	sequenced	
pLC79	tetON::pmpl-1::mCherry_pFTSH	carb	sequenced	
pLC84	exp-3p::pmpl-1::tagRFP::tbb-2 3'UTR	carb		
pLC103	vha-6p::pmpl-1::tagRFP::tb-2 3'UTR	carb		
pLC113	vha-6p::peel-1::tagRFP::tbb-2 3'UTR	carb		
pLC122	pmpl-2::mCherry_N1	kan	sequenced	
pLC123	pmpl-1(A47T)::mCherry_N1	kan	sequenced	
pLC124	peel-1(S124F)::eGFP_N1	kan	sequenced	
pLC172	hsp16.41p::peel-1(-28aa)::let-858 3'UTR	carb	sequenced	
pLC173	hsp16.41p::peel-1(-65aa)::let-858 3'UTR	carb	sequenced	
pLC174	peel-1(-28aa)_N1	kan	sequenced	
pLC175	peel-1(-65aa)_N1	kan	sequenced	
pLC226	peel-1(-39aa)_N1	kan	sequenced	
pLC227	MBP::TEV::peel-1(-65aa)	carb	sequenced	
pLC294	MBP::TEV::peel-1 (-65aa)	carb/chlor	sequenced	
pLC297	peel-1(-42aa)_N1	kan	sequenced	
pLC298	peel-1(-44aa)_N1	kan	sequenced	
pLC304	peel-1(-41aa)_N1	kan	sequenced	
pLC305	peel-1(-40aa)_N1	kan	sequenced	
pLC345	peel-1::eGFP::ER-ret(GBR1 C-tail)_N1	kan	sequenced	
pLC363	peel-1(S124V)::eGFP_N1	kan	sequenced	
pLC365	pmpl-1::mCherry::ER-ret(GBR1 C-tail)_N1	kan	sequenced	
pLC370	MBP::TEV::peel-1::8X His	carb	sequenced	
pLC376	MBP::TEV::peel-1::8X His	carb/chlor	sequenced	
pLC385	peel-1(D109A)::eGFP_N1	kan	sequenced	
pLC395	peel-1(L115Q)::eGFP_N1	kan	sequenced	
pLC396	peel-1(L118Q)::eGFP_N1	kan	sequenced	
pLC397	peel-1(L122Q)::eGFP_N1	kan	sequenced	
pLC398	peel-1(L126Q)::eGFP_N1	kan	sequenced	
pLC438	peel-1(L118Q,S124V)::eGFP_N1	kan	sequenced	
pLC439	peel-1(L118Q,L126Q)::eGFP_N1	kan	sequenced	
pLC440	peel-1(S124V,L126Q)::eGFP_N1	kan	sequenced	
pLC475	hsp16.41p::peel-1(-39aa)::let-858 3'UTR	carb	sequenced	
mCherry-KDEL	mCherry-KDEL	kan		gift from Suzanne Hoppins
pGP9	peel-1::eGFP_N1	kan	sequenced	
pGP10	pmpl-1::mCherry_N1	kan	sequenced	
pPD97/98	cc::GFP (unc-122p::GFP)	carb		gift from Piali Sengupta
pCFJ90	Pmyo-2::mCherry::unc-54 3'UTR	carb		
pCFJ104	Pmyo-3::mCherry::unc-54 3'UTR	carb		
pGH8	Prab-3::mCherry::unc-54 3'UTR	carb		
pBS_SK	pBluescript	carb		
pCFJ150	destination vector (4-1-2-3)	carb		
eGFP_N1	eGFP_N1	kan		gift from Suzanne Hoppins
mCherry_N1	mCherry_N1	kan		gift from Suzanne Hoppins

2.8 REFERENCES

1. A. Burga, E. Ben-David, L. Kruglyak, Toxin-Antidote Elements Across the Tree of Life. Annu. Rev. Genet. 54, 387–415 (2020).

2. E. Ben-David, P. Pliota, S. A. Widen, A. Koreshova, T. Lemus-Vergara, P. Verpukhovskiy, S. Mandali, C. Braendle, A. Burga, L. Kruglyak, Ubiquitous Selfish Toxin-Antidote Elements in Caenorhabditis Species. Curr. Biol. 31, 990-1001.e5 (2021).

3. L. M. Noble, J. Yuen, L. Stevens, N. Moya, R. Persaud, M. Moscatelli, J. L. Jackson, G. Zhang, R. Chitrakar, L. R. Baugh, C. Braendle, E. C. Andersen, H. S. Seidel, M. V. Rockman, Selfing is the safest sex for Caenorhabditis tropicalis. eLife 10, e62587 (2021).

4. H. S. Seidel, M. V. Rockman, L. Kruglyak, Widespread genetic incompatibility in C. elegans maintained by balancing selection. Science 319, 589–594 (2008).

5. H. S. Seidel, M. Ailion, J. Li, A. van Oudenaarden, M. V. Rockman, L. Kruglyak, A novel sperm-delivered toxin causes late-stage embryo lethality and transmission ratio distortion in C. elegans. PLoS Biol. 9, e1001115 (2011).

6. R. Kaletsky, V. Yao, A. Williams, A. M. Runnels, A. Tadych, S. Zhou, O. G. Troyanskaya, C. T. Murphy, Transcriptome analysis of adult Caenorhabditis elegans cells reveals tissue-specific gene and isoform expression. PLOS Genet. 14, e1007559 (2018).

7. J. Medina, M. L. Ballesteros, J. Salinas, Phylogenetic and functional analysis of Arabidopsis RCI2 genes. J. Exp. Bot. 58, 4333–4346 (2007).

8. C. Navarre, A. Goffeau, Membrane hyperpolarization and salt sensitivity induced by deletion of PMP3, a highly conserved small protein of yeast plasma membrane. EMBO J. 19, 2515–2524 (2000).

9. P. S. C. F. Rocha, Plant abiotic stress-related RCI2/PMP3s: multigenes for multiple roles. Planta 243, 1–12 (2016).

10. H.-S. Kim, W. Park, H.-G. Lim, S. Eom, J.-H. Lee, J. E. Carlson, S.-J. Ahn, NaCl-induced CsRCI2E and CsRCI2F interact with aquaporin CsPIP2;1 to reduce water transport in Camelina sativa L. Biochem. Biophys. Res. Commun. 513, 213–218 (2019).

11. A. C. M. Kwok, F. Zhang, Z. Ma, W. S. Chan, V. C. Yu, J. S. H. Tsang, J. T. Y. Wong, Functional responses between PMP3 small membrane proteins and membrane potential. Environ. Microbiol. 22, 3066–3080 (2020).

12. J. Miles, P. van Oosten-Hawle, Tissue-Specific RNAi Tools to Identify Components for Systemic Stress Signaling. JoVE J. Vis. Exp., e61357 (2020).

13. J. Capel, J. A. Jarillo, J. Salinas, J. M. Martínez-Zapater, Two homologous low-temperature-inducible genes from Arabidopsis encode highly hydrophobic proteins. Plant Physiol. 115, 569–576 (1997).

14. R. K. Yeshvekar, R. B. Nitnavare, T. Chakradhar, P. Bhatnagar-Mathur, M. K. Reddy, P. S. Reddy, Molecular characterization and expression analysis of pearl millet plasma membrane proteolipid 3 (Pmp3) genes in response to abiotic stress conditions. Plant Gene 10, 37–44 (2017).

15. J. Hallgren, K. D. Tsirigos, M. D. Pedersen, J. J. A. Armenteros, P. Marcatili, H. Nielsen, A. Krogh, O. Winther, DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. bioRxiv [Preprint] (2022). https://doi.org/10.1101/2022.04.08.487609.

16. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein,

D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021).

17. M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Žídek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis, S. Velankar, AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 50, D439–D444 (2021).

18. J.-H. Chen, W. Wang, C. Wang, T. Kuang, J.-R. Shen, X. Zhang, Cryo-electron microscopy structure of the intact photosynthetic light-harvesting antenna-reaction center complex from a green sulfur bacterium. J. Integr. Plant Biol. 65, 223–234 (2023).

19. R. Puskar, C. Du Truong, K. Swain, S. Chowdhury, K.-Y. Chan, S. Li, K.-W. Cheng, T. Y. Wang, Y.-P. Poh, Y. Mazor, H. Liu, T.-F. Chou, B. L. Nannenga, P.-L. Chiu, Molecular asymmetry of a photosynthetic supercomplex from green sulfur bacteria. Nat. Commun. 13, 5824 (2022).

20. P. Kumar, A. Nagarajan, P. D. Uchil, Analysis of Cell Viability by the Lactate Dehydrogenase Assay. Cold Spring Harb. Protoc. 2018 (2018).

21. S. Burmakina, Y. Geng, Y. Chen, Q. R. Fan, Heterodimeric coiled-coil interactions of human GABAB receptor. Proc. Natl. Acad. Sci. U. S. A. 111, 6958–6963 (2014).

22. E. Maeno, Y. Ishizaki, T. Kanaseki, A. Hazama, Y. Okada, Normotonic cell shrinkage because of disordered volume regulation is an early prerequisite to apoptosis. Proc. Natl. Acad. Sci. U. S. A. 97, 9487–9492 (2000).

23. E. M. Hedgecock, J. E. Sulston, J. N. Thomson, Mutations affecting programmed cell deaths in the nematode Caenorhabditis elegans. Science 220, 1277–1279 (1983).

24. H. Ellis, Genetic control of programmed cell death in the nematode C. elegans. Cell 44, 817–829 (1986).

25. R. E. Ellis, D. M. Jacobson, H. R. Horvitz, Genes Required for the Engulfment of Cell Corpses during Programmed Cell Death in Caenorhabditis Elegans. Genetics 129, 79–94 (1991).

26. M. A. Perez, J. L. Watts, Worms, Fat, and Death: Caenorhabditis elegans Lipid Metabolites Regulate Cell Death. Metabolites 11, 125 (2021).

27. P. Malovrh, G. Viero, M. D. Serra, Z. Podlesek, J. H. Lakey, P. Maček, G. Menestrina, G. Anderluh, A Novel Mechanism of Pore Formation: membrane penetration by the N-terminal amphipathic region of equinatoxin. J. Biol. Chem. 278, 22678–22685 (2003).

28. H. P. Werner, R. Preissner, C. Frömmel, Structural features of transmembrane helices. FEBS Lett. 559, 145–151 (2004).

29. J. Macrander, M. Daly, Evolution of the Cytolytic Pore-Forming Proteins (Actinoporins) in Sea Anemones. Toxins 8, 368 (2016).

30. D. Bohl, A. Salvetti, P. Moullier, J. M. Heard, Control of Erythropoietin Delivery by Doxycycline in Mice After Intramuscular Injection of Adeno-Associated Vector. Blood 92, 1512–1517 (1998).

31. G. Rizzuto, M. Cappelletti, D. Maione, R. Savino, D. Lazzaro, P. Costa, I. Mathiesen, R. Cortese, G. Ciliberto, R. Laufer, N. La Monica, E. Fattori, Efficient and regulated erythropoietin production by naked DNA injection and muscle electroporation. Proc. Natl. Acad. Sci. 96, 6417–6422 (1999).

32. A. T. Das, L. Tenenbaum, B. Berkhout, Tet-On Systems For Doxycycline-inducible Gene Expression. Curr. Gene Ther. 16, 156–167 (2016).

33. T. J. Jentsch, VRACs and other ion channels and transporters in the regulation of cell volume and beyond. Nat. Rev. Mol. Cell Biol. 17, 293–307 (2016).

34. Y. Okada, Physiology of the volume-sensitive/regulatory anion channel VSOR/VRAC. Part 1: from its discovery and phenotype characterization to the molecular entity identification. J. Physiol. Sci. JPS 74, 3 (2024).

35. G. Bi, M. Su, N. Li, Y. Liang, S. Dang, J. Xu, M. Hu, J. Wang, M. Zou, Y. Deng, Q. Li, S. Huang, J. Li, J. Chai, K. He, Y. Chen, J.-M. Zhou, The ZAR1 resistosome is a calcium-permeable channel triggering plant immune signaling. Cell 184, 3528-3541.e12 (2021).

36. R. Latorre, O. Alvarez, Voltage-dependent channels in planar lipid bilayer membranes. Physiol. Rev. 61, 77–150 (1981).

37. B. Xia, S. Fang, X. Chen, H. Hu, P. Chen, H. Wang, Z. Gao, MLKL forms cation channels. Cell Res. 26, 517–528 (2016).

38. S. B. Hladky, D. A. Haydon, Ion transfer across lipid membranes in the presence of gramicidin A. I. Studies of the unit conductance channel. Biochim. Biophys. Acta 274, 294–312 (1972).

39. S. Brenner, The genetics of Caenorhabditis elegans. Genetics 77, 71–94 (1974).

40. E. A. De Stasio, S. Dorman, Optimization of ENU mutagenesis of Caenorhabditis elegans. Mutat. Res. 495, 81–88 (2001).

41. M. W. Crawford, G. Posch, J. Cattin-Ortolá, I. Topalidou, M. Ailion, Mutations in the NXF-1:NXT-1 mRNA export complex affect gene-expression driven by the hsp-16.41 promoter. MicroPublication Biol., doi: 10.17912/micropub.biology.000918 (2023).

42. E. Afgan, D. Baker, M. van den Beek, D. Blankenberg, D. Bouvier, M. Čech, J. Chilton, D. Clements, N. Coraor, C. Eberhard, B. Grüning, A. Guerler, J. Hillman-Jackson, G. Von Kuster, E. Rasche, N. Soranzo, N. Turaga, J. Taylor, A. Nekrutenko, J. Goecks, The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res. 44, W3–W10 (2016).

43. M. A. DePristo, E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, M. J. Daly, A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498 (2011).

44. P. Cingolani, A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, D. M. Ruden, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6, 80–92 (2012).

45. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010).

46. D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison, H. O. Smith, Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods 6, 343–345 (2009).

47. A. H. Laszlo, I. M. Derrington, J. H. Gundlach, MspA nanopore as a single-molecule tool: From sequencing to SPRNT. Methods San Diego Calif 105, 75–89 (2016).

Chapter 3. RECURRENT BUT SHORT-LIVED DUPLICATIONS OF CENTROMERIC PROTEINS IN HOLOCENTRIC *CAENORHABDITIS* SPECIES

This chapter is closely adapted from Caro *et al.,* 2022. The work in this chapter was done in close collaboration with Pravrutha Raman (Fred Hutch Cancer Center), Florian Steiner (University of Geneva), and Harmit S. Malik (Fred Hutch Cancer Center).

3.1 INTRODUCTION

The faithful inheritance of genetic material is indispensable for all life. In most eukaryotes, faithful inheritance of chromosomes relies on the centromeric histone H3 variant (*CenH3*) to attach chromosomes to microtubules. CenH3 acts both as a structural component of the multi-subunit complex that links chromosomes to microtubules for segregation and as the epigenetic mark that defines and maintains the centromeric location(s) on chromosomes (Allshire and Karpen 2008; De Wulf and Earnshaw 2008; Fukagawa and Earnshaw 2014; McKinley and Cheeseman 2016; Ali-Ahmad and Sekulić 2020; Mellone and Fachinetti 2021). CenH3 is critical for chromosome segregation during mitosis and meiosis. Mutations or misregulation of CenH3 have severe consequences for fertility and viability in many species (Stoler, et al. 1995; Buchwitz, et al. 1999; Howman, et al. 2000; Blower and Karpen 2001). CenH3 would therefore be expected to be conserved across eukaryotes and expected to evolve under strong evolutionary constraints to maintain functionality.

Despite this expectation for strong conservation, *CenH3* genes have rapidly evolved in animal and plant species (Malik and Henikoff 2001; Talbert, et al. 2004; Schueler, et al. 2010).

This rapid evolution is hypothesized to result from a unique genetic conflict that stems from asymmetric female meiosis in animals and plants, in which only one of four meiotic products gets selected to be included in the oocyte nucleus. As a result of this bottleneck, chromosomes compete for inclusion into the egg in a process termed 'centromere drive' (Henikoff, et al. 2001; Malik 2009; Schueler, et al. 2010; Lampson and Black 2017). This competition favors changes in centromeric DNA that result in over-recruitment of centromeric proteins (Chmátal, et al. 2014; Akera, et al. 2017; Iwata-Otsubo, et al. 2017). Conversely, genes encoding centromeric proteins evolve rapidly to suppress the 'selfish advantage' of cheating centromeres to restore parity and ameliorate the deleterious effects of centromere-drive (Finseth, et al. 2021; Kumon, et al. 2021). Thus, in many animal and plant species, CenH3 proteins evolve rapidly despite being essential for faithful chromosome segregation.

CenH3 proteins can also function differently during meiotic and mitotic segregations. Some plant *CenH3* mutants only show defects during meiosis, but not mitosis (Lermontova, et al. 2011; Ravi, et al. 2011; Schubert, et al. 2014). Conflicting evolutionary selective pressures on *CenH3* between these functions (*e.g.*, mitotic versus meiotic, conserved versus rapidly evolving) could be resolved by gene duplication, which allows the duplicate (paralog) and ancestral genes to specialize for different functions (Hittinger and Carroll 2007; Des Marais and Rausher 2008; Gallach and Betrán 2011). Indeed, *CenH3* genes have also undergone repeated gene duplications not just in plants but also in several animal species including cows, fruit flies, mosquitoes, and nematodes (Li and Huang 2008; Zedek and Bureš 2016; Kursel and Malik 2017; Ishii, et al. 2020; Kursel, et al. 2020; Despot-Slade, et al. 2021; Elisafenko, et al. 2021; Kursel, et al. 2021).

Cytological evidence in *Drosophila virilis* suggests that divergent *CenH3* paralogs can acquire separate, tissue-specific functions (Kursel, et al. 2021).

Although *CenH3* has undergone duplication and diversification in *Drosophila* and mosquito species, four orders of insects have completely lost *CenH3* (Drinnenberg, et al. 2014). *CenH3* loss appears to correlate with transitions from monocentricity, in which centromeric determinants are concentrated in one genomic region, to holocentricity, in which centromeres are dispersed along the length of their chromosomes. Thus, holocentricity may impose unique selective pressures that shape the path of *CenH3* and kinetochore evolution (Marques and Pedrosa-Harand 2016; Cortes-Silva, et al. 2020; Senaratne, et al. 2022; Wang, et al. 2022).

In contrast to holocentric insects that have lost *CenH3*, *CenH3* homologs are present in other holocentric animal and plant species (Drinnenberg, et al. 2014). Moreover, several nematode clades encode duplications and diversification of *CenH3* genes (Despot-Slade, et al. 2021). Holocentric chromosome segregation in nematodes has been best studied in *C. elegans,* which encodes two *CenH3* paralogs. The first of these to be characterized was *hcp-3*, which encodes a protein required for recruiting all other kinetochore proteins and is essential for embryonic mitotic divisions in *C. elegans* (Buchwitz, et al. 1999; Oegema, et al. 2001). However, HCP-3 appears to be dispensable for oocyte meiotic segregation (Monen, et al. 2005). A second *CenH3* paralog in *C. elegans*, CPAR-1, shares high sequence similarity to HCP-3 in the histone fold domain but is diverged in the N-terminal domain (Monen, et al. 2015). Although CPAR-1 is enriched in meiotic chromosomes, it does not appear to localize to centromeres at all, and its precise function is not well understood (Gassmann, et al. 2012; Monen, et al. 2015). An

independent *hcp-3* duplication occurred in a related species, *C. remanei* (Monen, et al. 2015), but its function is also unknown. These previous studies left unclear whether *CenH3* duplications in *C. elegans* and *C. remanei* were unusually rare or typical of *Caenorhabditis* nematodes.

Faithful chromosome segregation in *C. elegans* relies not only on CenH3 alone but also on CenH3 interaction with HCP-4 (CENP-C in mammals) and KNL-2 to form the inner kinetochore. A predicted structured region of the HCP-3 N-terminal tail interacts with KNL-2 (de Groot, et al. 2021; Prosée, et al. 2021). This interaction is necessary for the establishment of centromeres in the hermaphrodite germline, prior to the first embryonic mitosis (Prosée, et al. 2021). Identifying which HCP-3 residues are important for protein interactions has been challenging, owing to low sequence identity of CenH3 among species (de Groot, et al. 2021; Prosée, et al. 2021). Despite high sequence divergence of CenH3 N-terminal tails, *CenH3* evolution is likely constrained to maintain important protein-protein interaction interfaces (Malik, et al. 2002; Maheshwari, et al. 2015). Identifying these constraints may reveal insights into the molecular architecture of such interactions. Thus, a phylogenetic study of CenH3 and kinetochore protein evolution and duplication in *Caenorhabditis* nematodes would not only yield insights into the cadence of gene duplication and retention but also reveal functional constraints that would inform the molecular interactions that underlie the important function of chromosome segregation.

The growing collection of *Caenorhabditis* species and their genome sequences (Stevens, et al. 2019) (unpublished genomes at http://caenorhabditis.org/) provides a rich dataset for identifying both the evolutionary trajectory and constraints of their *CenH3* genes. Taking

advantage of this resource, we performed detailed phylogenomic analyses to understand the evolution of *CenH3* genes in *Caenorhabditis*. Our studies reveal that thirteen out of thirty-two analyzed *Caenorhabditis* species encode two or more *CenH3* paralogs, which were the result of at least ten independent duplication events. We confirm these paralogs are expressed in both sexes in representative species. We identify novel, conserved protein motifs within the N-terminal domains of *Caenorhabditis* CenH3 proteins that are likely important for interactions with other kinetochore proteins and for centromere biology. Although some motifs are strictly retained, others display variable instances of loss and retention between ancestral and duplicate genes, revealing clues to their sub-functionalization. In a possible case of neofunctionalization, we find an unusual *CenH3* paralog in *C. afra* that encodes a CENP-C-CenH3 fusion protein. Extending our analyses beyond *CenH3*, we find independent duplications of other inner and outer kinetochore proteins, revealing a remarkable pace of diversification of the kinetochore within *Caenorhabditis* nematodes. Our analyses thus reveal an unusual 'revolving door' of CenH3 protein duplications, with retention only over short evolutionary periods. This pattern contrasts with the strict, long-lived retention of *CenH3* paralogs seen in *Drosophila*, mosquito, plant, and even other holocentric nematode species (Maheshwari, et al. 2015; Kursel and Malik 2017; Kursel, et al. 2020; Kursel, et al. 2021, Despot-Slade, et al. 2021). We hypothesize that this pattern may result from the unusual mechanisms of centromere establishment and inheritance in holocentric *Caenorhabditis* species.

3.2 RESULTS

3.2.1 hcp-3 has duplicated at least ten independent times in Caenorhabditis

Global efforts to isolate and sequence *Caenorhabditis* species have recently resulted in several well-assembled genomes from highly diverged species (Stevens, et al. 2019) (unpublished genomes at http://caenorhabditis.org/). We used this resource for phylogenomic analyses of *CenH3* evolution. We used *C. elegans* HCP-3 as a query for tBLASTn searches against genome sequences from 32 *Caenorhabditis* species (Altschul, et al. 1990; Altschul, et al. 1997; Stevens, et al. 2019) (http://caenorhabditis.org/) to identify all *hcp-3* homolog (*hcp-3*-like) genes (Supplementary Data S1) and their syntenic location (surrounding genes) (Figure 3.1). Core histone H3 and H3 variant genes were also obtained in these analyses but were easily distinguished from *hcp-3* homologs because of their high similarity to each other. Since our focus was on putative *hcp-3* orthologs and paralogs, we ignored both highly conserved core histone H3 and H3 variant proteins, as well as species-specific instances of highly diverged H3-like genes such as *F20D6.9* (also referred to as *D6H3*) from *C. elegans* (Henikoff, et al. 2000; Delaney, et al. 2018).

Unlike in holocentric insects (Drinnenberg, et al. 2014), we found that *hcp-3* orthologs are strictly retained in all *Caenorhabditis* species. In 28 of 32 species, they are found in shared syntenic locations, between genes homologous to *C. elegans hlh-11* and *F58A4.6* (Figure 3.1). In three of the four remaining species, at least partial synteny is maintained downstream of *hcp-3* (genes *F58A4.6*, *pri-1*, and *bbs-4)* whereas upstream synteny is either not maintained (in *C. tropicalis*) or cannot be discerned due to short genomic scaffolds (*C. waitukubuli* and *C. japonica*, Figure 3.1). Only *C. species 49* (*C. sp49*) lacks an *hcp-3* gene in this shared syntenic locus. Based on its presence in the ancestral locus in its sister species *C. sp25* and all other species, we infer that this movement of *hcp-3* is specific to *C. sp49*. *C. sp49* encodes two *CenH3*

paralogs, both found in new syntenic loci that are not shared with sister species. We arbitrarily assign one homolog as *hcp-3* and the other as *hcp-3L9* (further explained below).

In addition to *hcp-3* orthologs, we found that thirteen out of thirty-two examined species encode at least one additional *hcp-3*-like sequence. We refer to these paralogs as "*hcp-3L*" genes (for *hcp-3 Like*) (Figure 3.1). These *hcp-3L* genes include previously reported *hcp-3* duplications in *C. remanei* and *C. elegans* (Monen, et al. 2005; Monen, et al. 2015), which we refer to as *hcp-3L4* and *cpar-1* (as previously named, also referred to as *hcp-3L1* in Figures 3.1 and 3.2), respectively. We also identified one additional *hcp-3L* paralog in *C. tribulationis*, *C. sp41, C. sinica*, *C. latens*, *C. brenneri*, *C. doughertyi*, *C. sp54*, *C. panamensis*, *C. afra,* and *C. sp49*, and two, independent *hcp-3L* paralogs in *C. sp48*. In most cases, *hcp-3L* paralogs shared identical exon-intron structure as their orthologs. However, we also observed a few instances of intron losses and gains in *hcp-3* or *hcp-3L* genes (Supplementary Figure 3.S1). Such partial intron losses have been observed previously in plants (Roy and Penny 2007), fungi (Nielsen, et al. 2004) and in *Caenorhabditis* species (Robertson 1998; Cho, et al. 2004; Kiontke, et al. 2004) and are thought to result from partial retrotransposition, in which cDNA partially replaced the genomic locus.

All *hcp-3* and *hcp-3L* genes encode proteins with conserved Histone Fold Domains (HFD) (See Supplementary Data), which are between 69-100% identical to the HFD of HCP-3 from the same species (Figure 3.1). In contrast, their N-terminal domains show high divergence from HCP-3 orthologs (26-97% identical, Figure 3.1). This pattern is consistent with overall trends of *CenH3* evolution, where the HFDs are more evolutionarily constrained due to

interactions with other histones, whereas the N-terminal domains can be so divergent that they cannot even be reliably aligned across different lineages (Malik and Henikoff 2001).

We next used a combination of syntenic and phylogenetic analyses to determine whether *hcp-3L* paralogs were shared between different species, which would indicate their functional co-retention with *hcp-3* orthologs for long evolutionary periods. The highly divergent N-terminal tail sequences of *hcp-3* and their paralogs cannot be reliably aligned and could distort our interpretations, so our phylogenies are based on HFD alignments. We first used the amino acid sequences for a maximum likelihood phylogenetic analysis (Supplementary Figure 3.S2). We found that the protein-based phylogeny suffered from poor resolution, was unable to resolve most of the important branches and groupings of interest and was even incongruous with the well-accepted *Caenorhadbitis* phylogeny. Therefore, we built a maximum likelihood phylogenetic tree using a codon-based alignment of the conserved HFD cDNA sequence (Figure 3.2). This phylogeny is much better resolved especially at shallow nodes (both phylogenies suffer from lack of resolution at deeply branching nodes) and largely agrees with our findings from the shared synteny analyses. For example, both synteny and phylogenetic analyses suggest that the duplication that gave rise to *hcp-3L4* occurred prior to the common ancestor of *C. latens* and *C. remanei* (Figure 3.2). Similarly, we can infer that *hcp-3L5* duplicated in the common ancestor of *C. sp48* and *C. brenneri*. In contrast, the *hcp-3L* paralogs in *C. doughertyi*, *C. sp54*, *C. elegans*, *C. panamensis, C. afra*, *C. sp49*, and the additional *hcp-3L* paralog in *C. sp48* each arose via seven independent duplications (Figure 3.2). In each of these seven species, the *hcp-3L* paralogs are present in unique genomic locations (Figure 3.1) and typically group most closely with *hcp-3* orthologs from the same species (Figure 3.2).

The only discrepancy between the synteny and phylogenetic analyses was for *hcp-3L2* genes found in *C. tribulationis, C. sp41* and *C. sinica*. These species are part of a group, with *C. sinica* believed to be an outgroup to *C. tribulationis, C. sp41,* and *C. 90anzibari*. Different genomic locations of *hcp-3L* duplicates among *C. tribulationis, C. sp41 and C. sinica* (Figure 3.1) would suggest that the duplications are the result of independent duplication events although the small size of *C. sinica* genomic scaffolds leave its shared synteny status ambiguous. In contrast, our phylogenetic analyses group *hcp-3L* genes from these species together with a high degree of confidence (Figure 3.2), suggesting that *hcp-3L2* is the result of a single duplication event, followed by transposition of this gene to a new locus in *C. sinica*. We infer that the absence of *hcp-3L2* in *C. zanzibari* could be the result of gene loss although there is no evidence of *hcp-3L* loss in any other species. Another possibility is that *C. zanzibari* may be ancestral to *C. tribulationis* and *C. sinica* for the *hcp-3L* syntenic location, in contrast to the accepted species phylogeny, and may have never acquired a *hcp-3L* paralog. Recent studies have revealed widespread roles in diverse taxa for introgression and/or incomplete lineage sorting, leading to different genomic locations having vastly different evolutionary histories (Hobolth, et al. 2011; Mailund, et al. 2014; Ginsberg, et al. 2019; Suvorov, et al. 2022). Thus, it is formally possible that *C. zanzibari* never acquired *hcp-3L2.* However, based on the well-resolved species phylogeny of this quartet of species, we favor the first possibility that *C. zanzibari* acquired, then lost *hcp-3L2.* Therefore, our analyses reveal that *hcp-3* has duplicated at least ten independent times within *Caenorhabditis* species.

We examined the expression of *hcp-3* and *hcp-3L* genes across representative *Caenorhabditis* species. We used RT-PCR analyses using specific primers on template RNA collected from a mixed population of males and females or hermaphrodites at various larval stages (see Methods). All analyzed species expressed both ancestral and duplicate *hcp-3* genes (Figure 3.3, Supplementary Figure 3.S3). We investigated whether *Caenorhabditis hcp-3L* genes have sex-restricted expression as is seen in some *Drosophila CenH3* paralogs (Kursel and Malik 2017). We performed RT-PCR on RNA collected from L4/young adult males or from L4/young adult hermaphrodites or females (these developmental stages capture both female and male meiosis). Unlike *Drosophila CenH3* paralogs, we did not find sex-restricted expression of any *hcp-3L* genes (Figure 3.3, Supplementary Figure 3.S3); instead, they appear to be expressed in both sexes.

3.2.2 Motif retention and loss in the N-terminal region of HCP-3 and HCP-3L proteins

Although CenH3 proteins all have a relatively conserved HFD, their N-terminal tails are often so divergent that they cannot be aligned nor even be considered homologous across different lineages (Malik and Henikoff 2001). Nevertheless, conserved motifs have been identified in the N-terminal tails of CenH3 proteins from many other lineages including *Drosophila*, mosquitos, and plants using alignment-independent approaches (Maheshwari, et al. 2015; Kursel and Malik 2017; Kursel, et al. 2020). These N-terminal tail motifs are often highly conserved within a lineage, but not conserved across different lineages. Although no such studies have been previously performed for the *Caenorhabditis* HCP-3 proteins, recent studies show that the N-terminal tail of *C. elegans* HCP-3 interacts with the inner kinetochore protein KNL-2 via a predicted structured region (de Groot, et al. 2021; Prosée, et al. 2021). This interaction between

KNL-2 and HCP-3 is necessary for the establishment of centromeres in the hermaphrodite germline, prior to the first embryonic mitosis (Prosée, et al. 2021).

We took advantage of our comprehensive identification of HCP-3 and HCP-3L proteins to *de novo* identify conserved residues or motifs in their N-terminal tails using the MEME suite of software (Bailey, et al. 2015) as previously described (Kursel and Malik 2017) (see Methods). For this, we first identified motifs by analyzing all *Caenorhabditis* species encoding a single HCP-3 protein, which are more likely to have retained all motifs essential for their functions. Using this analysis, we identified 13 motifs within HCP-3 (Figure 3.4A, Supplementary Figure 3.S4), numbered sequentially from the N-terminus, with 11 motifs in the N-terminal tail and motifs 12 and 13 in the histone fold domain (HFD). Not all 13 motifs are universally present in species encoding a single *hcp-3* gene. For example, motif 2 is present in only a subset of species examined. Based on phylogenetic analyses, we infer that motif 2 was acquired in the ancestor of a clade of eight species which includes *C. sulstoni* and *C. becei* (Figure 3.4B).

In the second step, we investigated how well these 13 motifs are conserved in species containing *hcp-3L* genes. We found that these motifs varied in their evolutionary stability and conservation. N-terminal tail motifs 1-11 are more variably retained than HFD motifs 12-13 (Supplementary Figure 3.S4), which are present in all HCP-3 and HCP-3L proteins as expected, except for HCP-3L3 from *C. sp54,* which has a divergent HFD. Overall, the motifs we have defined account for 48% and 63% of the total N-tail sequence in *C. elegans* HCP-3 and CPAR-1, respectively.

Our initial unsupervised motif analysis found that motif 3 was universally conserved in all HCP-3 and HCP-3L proteins (Figure 3.4B). In contrast, motifs 1 and 4 were universally retained in at least one paralog in each species (often both) with only a few exceptions. Recognizing that apparent 'motif loss' might be the result of indels or divergence of a critical conserved residue, we manually re-examined the sequences missing either motif 1 or 4 to see if they were missed because they fell below the statistical threshold of the unsupervised motif analysis. Based on these analyses (Supplementary Data File S4), we were able to confirm the presence of motifs 1 and 4 in all species (Figure 3.4B). Thus, three motifs (1, 3, and 4) are present in at least one HCP-3 paralog in all species. Notably, these motifs have not been identified in previous analyses of the N-terminal tail, highlighting the value of alignment-independent methods. These motifs include residues that are almost universally conserved in *Caenorhabditis* species (asterisks in Figure 3.4A). We predict that mutation of these residues may reveal important insight about the various functions of the HCP-3 N-terminal tail, including its interactions with kinetochore proteins such as KNL-2.

Motifs 5, 6, 7, 9, and 11 were less conserved, being present in 78-94% of species. For example, motif 6 appears to be lost in both HCP-3 paralogs from *C. tribulationis* and *C. afra*, while motif 11 is not found in *C. wallacei*, *C. elegans* (both paralogs), and in sister species *C. sulstoni* and *C. afra* (both paralogs). Motif 5 includes a 4-amino acid segment, ExxR (Figure 3.4A, where x represents any amino acid) that constitutes a putative cleavage motif for the separase enzyme, which initiates anaphase by cleaving the kleisin subunit of cohesin (Monen, et al. 2015). Although this ExxR motif is found in both HCP-3 and CPAR-1 in *C. elegans,* only the latter is cleaved by separase (Monen, et al. 2015). This suggested that the ExxR motif is

necessary but not sufficient for efficient separase cleavage. Since CPAR-1 is not associated with centromeres (Gassmann, et al. 2012; Monen, et al. 2015), it is difficult to establish the significance of the ExxR motif, whose mutation led to no deleterious fitness consequences (Monen, et al. 2015). In cases where motif 5 was missing, individual alignments of HCP-3 sequences allowed us to identify the ExxR separase motif in all HCP-3 and HCP-3L proteins, except for HCP-3L proteins from *C. latens* and *C. sp48* (Supplementary Figure 3.S5). Thus, although it is unclear whether it is required for separase cleavage or some other function, the ExxR motif is nevertheless largely conserved in all HCP-3 proteins and most HCP-3L proteins. It is possible that the cleavage site mediates the removal of the N-terminal tail from certain HCP-3L proteins, thereby eliminating it from a role in germline re-establishment of centromere identity (Prosée, et al. 2021).

Even though motif 2 was only acquired in eight *Caenorhabditis* species (Figure 3.4B), it has been retained in at least one HCP-3 paralog of each of these species. The only instances of motif 2 loss are seen in *C. panamensis hcp-3L10* and in *C. afra hcp-3*. Our findings suggest that motif 2 is functionally important in these species despite not being universally present within all *Caenorhabditis* species. We hypothesize that some clade-specific HCP-3 protein-protein interactions or functions were acquired via motif 2 in the ancestor of these eight species.

In some instances, motif loss occurred in only one of the two HCP-3 paralogs from the same species. For example, most HCP-3L proteins lack motif 7 whereas ancestral HCP-3 in the same species usually contained this motif. Similarly, in species containing motif 5 and/or 6, the *hcp-3L* gene almost always lost these motifs whereas the ancestral *hcp-3* maintained them. In

sister species *C. brenneri* and *C. sp48,* the converse is seen, where motif 11 is maintained in the duplicate *hcp-3L* gene but lost in ancestral *hcp-3*. Overall, however, motif loss tends to occur more frequently in the *hcp-3L* paralog instead of the ancestral *hcp-3*. Thus, *hcp-3L* paralogs may be capable of performing only a subset of the functions of an ancestral *hcp-3*. This asymmetric pattern of motif loss may also explain why ancestral *hcp-3* has been universally retained in all *Caenorhabditis* species, whereas *hcp-3L* paralogs are rarely present in more than two species.

3.2.3 Selective constraints on hcp-3 orthologs and hcp-3L paralogs

Our study represents an opportunity to evaluate the selective pressures imposed on *CenH3* genes either due to holocentricity or due to their recurrent duplication. A previous analysis had concluded there was weak evidence of positive selection from an analysis of *hcp-3* sequences from six divergent *Caenorhabditis* species whose sequence was available at that time (Zedek and Bureš 2012). However, extremely large divergence and low number of sequences can result in false signals of positive selection. Therefore, we revisited this analysis using maximum likelihood methods (see Methods). We separately analyzed *hcp-3* sequences from the two deep lineages of *Caenorhabditis* species evaluated here, as well as two subsets of species from one of the lineages for which we had enough representation (Supplemental Table S1A). In every case, we found no evidence of positive selection acting on *hcp-3* genes.

Since the presence of a paralog within the genome may affect the selective constraint on the ancestral *hcp-3* gene, we repeated the analysis by intentionally excluding all species that encode one or more *hcp-3L* paralogs (Supplemental Table S1A). Once again, we found no evidence for positive selection. Thus, in contrast to the previous study (Zedek and Bureš 2012)

and in contrast to findings that *CenH3* genes from multiple other animal and plant taxa evolve under positive selection (Malik and Henikoff 2001; Talbert, et al. 2004; Schueler, et al. 2010; Finseth, et al. 2015), we find no evidence for positive selection acting on *CenH3* genes in *Caenorhabditis*. Our inability to detect positive selection may reflect a lack of statistical power, although we note that the tree lengths used in our analysis are highly appropriate for such analyses.

Based on their presence in few species, we infer that most of the *hcp-3L* genes we identified in *Caenorhabditis* species are relatively young. Our finding that *hcp-3L* genes bore the brunt of motif loss (Figure 3.4) raised the possibility that many *hcp-3L* genes are not functionally constrained. To address this possibility, we carried out three types of analyses. First, we examined selective constraints acting on *hcp-3* and *cpar-1* by investigating polymorphisms within natural isolates of *C. elegans* strains that have been previously sequenced (Cook, et al. 2017) (Supplementary Figure 3.S6). We found only three synonymous (amino acid preserving) and zero non-synonymous (amino acid altering) polymorphisms in *hcp-3*. In contrast, *cpar-1* contained two synonymous polymorphisms (including one commonly shared between more than 25 strains) and six non-synonymous polymorphisms, four of which are shared among more than seven *C. elegans* strains. Some of these polymorphisms arise in otherwise conserved positions in the N-terminal tail (Figure 3.4, Supplementary Figure 3.S6) or HFD, implying that they are likely deleterious for function. In addition to non-synonymous changes, we found at least two strains that may have disrupted *cpar-1*, via either a frameshift or a splice site mutation. Based on this comparison, we infer that *cpar-1* is evolving under lower functional constraints than *hcp-3* in *C. elegans*.

Second, we tested whether *hcp-3L* paralogs are generally evolving under fewer stringent functional constraints than *hcp-3* genes. For this, we calculated dN/dS values, which measure the ratio of the normalized rates of non-synonymous substitutions to synonymous substitutions. A lower dN/dS ratio is reflective of higher functional constraints, whereas a dN/dS ratio of close to 1 is reflective of lack of functional constraints for protein-coding function. We calculated dN/dS values in pairwise comparisons of the HFD of *hcp-3L* orthologs present in two distinct species: *hcp-3L4* in *C. latens* and *C. remanei, hcp-3L2* in *C. sinica* and *C. tribulationis,* and *hcp-3L5 in C. brenneri* and *C. sp48* (Supplemental Table 1B). We obtained dN/dS ratios of 0.02, 0.04, and 0.08, respectively. These values are considerably lower than 1, suggesting that all three paralogs have been retained under functional constraint for protein-coding function during the divergence of the respective *Caenorhabditis* species. Moreover, in all three cases, we found that dN/dS values for *hcp-3L* orthologs were comparable to or lower than corresponding *hcp-3* orthologs from the same species (Supplemental Table 1B). For comparison, the dN/dS values for pairwise comparisons of ancestral *hcp-3* from *C. latens/ C. remanei*, *C. sinica/ C. tribulationis* and *C. brenneri/ C. sp48* are 0.18, 0.02 and 0.03, respectively. Thus, unlike *cpar-1* in *C. elegans*, we find that *hcp-3L* paralogs have evolved under similar or even more stringent constraints than ancestral *hcp-3* genes at least in some *Caenorhabditis* species.

Given this finding, we revisited the age of the *hcp-3L* paralogs in *Caenorhabditis* species in a third analysis. Unlike dN or dN/dS values, dS values are relatively unaffected by selective constraints and provide a more reliable proxy for their divergence from *hcp-3* ancestors. We calculated the synonymous divergence (dS) between the HFD of *hcp-3L* paralogs whose closest

relatives are *hcp-3* orthologs from the same species (Figure 3.2). These dS values range from 0.15 (for *C. afra*) to 0.74 (for *C. doughertyi*) (Supplemental Table 1B). These dS values are considerably lower than seen between *Drosophila CenH3* paralogs in the same species (*e.g., D. virilis*). Although we lack reliable molecular clock-like estimates to convert these dS values to millions of years of divergence (Cutter 2008), the dS values are high enough to imply that a majority of these *hcp-3L* paralogs have been functionally retained for several million years, even though most of them have not been retained across multiple speciation events (Figure 3.1).

The overall selective pressure acting on *hcp-3L* paralogs is that of purifying selection or evolutionary constraint. However, our comparison of HFD between *hcp-3* and *hcp-3L3* from *C. sp54* revealed a dN/dS of 1.74 in a maximum likelihood test, although this is not statistically significantly different from the neutral expectation of dN/dS =1. Based on the phylogeny of *CenH3* HFD (Figure 3.2), we could infer that *C. sp44 hcp-3* is an outgroup to the two *C. sp54 hcp-3* genes. We compared *C. sp44 hcp-3* to either *hcp-3* or *hcp-3L3* from *C. sp54*. These analyses revealed a lower dN/dS in a comparison between the two ancestral *hcp-3* orthologs (dN/dS =0.09) than between *C. sp44 hcp-3* and *C. sp54 hcp-3L3* (dN/dS = 0.34). This implies that it is the unusual paralog, *hcp-3L3*, that has evolved more rapidly. This combined with our finding that HCP-3L3 contains duplications of the N-terminal tail motifs (Figure 3.5A) suggests the possibility of incipient neofunctionalization of the *hcp-3L3* paralog in *C. sp54*.

3.2.4 Duplication of other centromere-localized proteins in Caenorhabditis species

In most cases, the protein sequence of HCP-3 paralogs can be confidently aligned to the ancestral HCP-3, indicating clear homology. However, aligning *C. afra* HCP-3 and *C. afra* HCP-

3L8 revealed that the paralog contained an additional 198 amino acids on its N-terminus. This region was not homologous to HCP-3. To our surprise, we found that this segment was instead homologous to CENP-C (known as HCP-4 in *C. elegans*). HCP-4 and HCP-3 directly interact with each other in *C. elegans* (Oegema, et al. 2001) and in other eukaryotes. *C. afra hcp-3L8* contained two copies of *C. afra hcp-4* exons 1 and 2, followed by a partial copy of *hcp-4* exon 3. These *hcp-4* homologous segments are contiguous with *hcp-3*-homologous sequence to constitute the *hcp-3L8* coding sequence (Figure 3.5B). We used RT-PCR to confirm that *hcp-3L8* was transcribed as a single transcript containing homology to both *hcp-4* and *hcp-3* sequences (Figure 3.5C). Therefore, *C. afra hcp-3L8* is a chimera of *hcp-4* and *hcp-3*. In addition to this *hcp-4-hcp-3* fusion gene, *C. afra* also maintains its ancestral *hcp-3* and *hcp-4* genes. The functional roles of the HCP-4-like regions present within *hcp-3L8* are unknown. However, a conserved CENP-C motif is absent in this chimera. The conserved CENP-C motif, which mediates the interaction with the CenH3 nucleosome (Kato, et al. 2013), is present at the C-terminus of *C. elegans* HCP-4 (Moore and Roth 2001). Thus, loss of the CENP-C motif in HCP-3L8 is not unexpected since the HCP-4 and HCP-3 segments are already physically linked to each other in this chimeric protein.

Encouraged by this finding of *hcp-4* duplication and fusion with *hcp-3* in *C. afra*, we investigated whether other centromere-localized proteins have also duplicated and diversified like *hcp-3*. We performed similar paralog searches for proteins from the inner kinetochore (*hcp-4* and *knl-2*), middle kinetochore (*knl-1*), and outer kinetochore (*him-10*, *ndc-80*, *spdl-1*, and *zwl-1*). We found an intact copy of each ancestral gene in every species (Figure 3.6) except for two instances where we were unable to identify full-length intact *zwl-1* genes (in *C. kamaaina* and *C.*

tropicalis) ('#' in Figure 3.6). We found instances of duplications for all kinetochore proteins except *zwl-1*. These duplications either appear to be retained with an intact open reading frame (filled, gray arrows), or are interrupted (double lines), or show clear signs of pseudogenization (unfilled arrows) (Figure 3.6). In the 32 species examined, we found seven *hcp-4* duplicates, four *knl-2* duplicates (including a pseudogene in *C. brenneri*), eight *knl-1* duplicates, four *spdl-1* duplicates, five *ndc-80* duplicates (including two pseudogenes), and three *him-10* duplicates (including one pseudogene). Duplications of inner and middle kinetochore proteins were only marginally more prevalent than duplications of outer kinetochore proteins. Interestingly, we observed several instances of partial intron losses that occurred recurrently in genes encoding ancestral and paralog outer kinetochore proteins (Supplementary Figure 3.S7) like what we observed previously for *hcp-3* and *hcp-3L* genes (Supplementary Figure 3.S1). Overall, our analyses suggest that in addition to HCP-3, other kinetochore proteins are also undergoing duplication and diversification in *Caenorhabditis* species.

To understand the evolutionary constraints on *Caenorhabditis* kinetochore proteins, we analyzed these genes using maximum likelihood methods. We found no evidence of positive selection acting on ancestral *hcp-4*, *knl-1*, *knl-2*, *zwl-1*, *spdl-1*, *ndc-80,* and *him-10* genes in either the *C. elegans* or *C. afra* clades (Supplemental Table S1C). Next, we examined the evolutionary constraints acting on paralogs of kinetochore proteins by comparing the paralogs to the ancestral kinetochore genes from the same species. In all cases except two, we found strong evidence that that the duplicates are retained under strong purifying selection (Supplemental Table S1D). For *knl-2* in *C. inopinata* and *ndc-80* in *C. sp54*, we could not rule out the null hypothesis of neutral evolution.

We investigated whether any kinetochore protein paralogs have been co-retained with *hcp-3L* paralogs, which would suggest a concerted duplication and retention of multiple kinetochore proteins, consistent with significant specialization. We found that four of six independent *hcp-4* duplications coincided with retention of *hcp-3L* paralogs in the same species (Figure 3.6). These include an *hcp-4* paralog whose origin coincides with the *hcp-3L4* paralog in *C. latens* and *C. remanei*, two *hcp-4* paralogs that co-occur with *hcp-3L3* in *C. sp54*, and the *hcp-4-hcp-3* fusion gene in *C. afra* (*hcp-3L8*). Thus, 4 of 14 species containing an *hcp-3L* paralog also encode a (complete or partial) *hcp-4* paralog whereas 2 of 18 species lacking *hcp-3L* paralogs encode a *hcp-4* paralog: *C. sp44* and *C. kamaaina.* Thus, there is no statistically significant evidence of co-retention (p=0.36), indicating that the duplication or retention of *hcp-3* and *hcp-4* paralogs may be independent.

Other kinetochore proteins analyzed also largely reflect this pattern of independent duplication. Even though KNL-2 is required to deposit HCP-3 proteins at centromeres in *Caenorhabditis* species (Maddox, et al. 2007; de Groot, et al. 2021; Prosée, et al. 2021), there does not appear to be a significant pattern of co-retention with *hcp-3L* paralogs. The one exceptional species is *C. sp54*, which encodes an *hcp-3L3* paralog, two *hcp-4* paralogs, a *knl-2* paralog, a *knl-1* paralog, an *ndc-80* paralog, and a *him-10* paralog. If the proteins encoded by these paralogs exclusively interact with each other, this species may represent an intriguing case of incipient kinetochore specialization.

3.3 DISCUSSION

Our analyses reveal that *hcp-3* has duplicated at least ten independent times within *Caenorhabditis* species. In contrast to ancient co-retention of *CenH3* paralogs in plants, *Drosophila*, mosquito species (Maheshwari, et al. 2015; Kursel and Malik 2017; Kursel, et al. 2020; Kursel, et al. 2021), and even holocentric *Meloidogyne* nematode species (Despot-Slade, et al. 2021), we observed only a few cases of *hcp-3L* paralogs that are shared across two or three *Caenorhabditis* sister species, although this may partly reflect density of species sampling in these different taxonomic groups. Our findings suggest that most of the *hcp-3L* paralogs we have found are relatively young, assuming the relative ages of *Caenorhabditis* and *Drosophila* species analyzed are comparable (Cutter 2008).

Our comprehensive phylogenomic approach in *Caenorhabditis* nematodes uncovered two novel aspects of CenH3 evolution. First, we uncovered a detailed molecular architecture of the N-terminal tail of HCP-3 proteins (Figure 3.4). The HCP-3 N-terminal tail is dispensable for mitotic chromosome segregation and centromere maintenance during *C. elegans* development (Prosée, et al. 2021) but is essential in establishing a functional HCP-3 distribution in the germline, which is maintained in the subsequent generation throughout development. At least part of this functionality of the HCP-3 N-terminal tail stems from its interactions with kinetochore proteins like KNL-2 (de Groot, et al. 2021; Prosée, et al. 2021). Thus far, however, the molecular architecture of the interactions of HCP-3 with other kinetochore proteins like KNL-2 has been only crudely defined. Like in other eukaryotic lineages, the N-terminal tail of HCP-3 proteins is much more divergent than the histone fold domain (HFD). Thus, comparisons of functional domains in CenH3 N-terminal tails between taxonomic groups or even within *Caenorhabditis* are very difficult, exacerbating the difficulty in defining functional domains

within HCP-3's N-terminal tail. Our description of 11 motifs in HCP-3 N-terminal tails, including three that are nearly universally conserved, provides an important resource for the fine-scale dissection of the various protein-protein interactions mediated by the N-terminal tail and the functional role these interactions play in centromere biology. In particular, the three conserved motifs contain residues that are as well conserved as many HFD residues across *Caenorhabditis* species.

We propose that these N-terminal tail motifs are sites of previously proposed or novel protein-protein interactions, either with kinetochore proteins or with other chromatin factors that could intersect with holocentromere formation or maintenance. Consequently, motif gains or losses could indicate gains or losses of HCP-3 interactions with partner proteins. We observe one unambiguous case of motif gain in one clade of *Caenorhabditis* species. Motif 2 likely represents a novel protein-protein interaction module important for *CenH3* function at least in those species. We also observe several cases of motif degeneration or loss. Unlike in *Drosophila CenH3* paralogs (Kursel and Malik 2017), we see little evidence for motif redistribution between the paralog and ancestral *hcp-3* genes, which would suggest sub-functionalization; the only exception is motif 2 that appears to be present in either HCP-3 or HCP-3L proteins, but not both. Overall, we find that motif loss or degeneration preferentially occurs in *hcp-3L* paralogs rather than ancestral *hcp-3*, suggesting that the paralogs progressively lose ancestral functions and interactions. Since tail-less HCP-3 proteins can still function in mitosis (Prosée, et al. 2021), it is tempting to speculate that HCP-3L paralogs could still function in mitosis despite progressive loss of N-terminal motifs.

The remarkable example of a chimeric gene in *C. afra,* where an HCP-3L protein is fused to an inner kinetochore protein, HCP-4 (CENP-C in mammals; Figure 3.5) exemplifies an instance where previously conserved motifs could be lost. HCP-3 and HCP-4 physically interact in many eukaryotes to form the kinetochore complex during mitosis. The fusion of these two proteins in HCP-3L8 guarantees a protein-protein interaction, which is consistent with the loss of the CENP-C motif (Kato, et al. 2013) required for HCP-3L and HCP-4 interactions. This could also lead to loss of HCP-3L N-terminal tail motifs required for HCP-4 association.

The second major conclusion from our evolutionary analyses is the unusually rapid cadence of turnover of *hcp-3* paralogs in *Caenorhabditis* species. Nearly half of the species we analyzed contain an *hcp-3* paralog. Yet, in contrast to analyses in *Drosophila* and mosquito lineages, where duplicates were older and fewer in number (Kursel and Malik 2017; Kursel, et al. 2020), the *Caenorhabditis* paralogs were acquired through ten independent duplication events. Most paralogs have only been retained in a single species with only one *hcp-3L* paralog being present in more than two species. Previous analyses suggest that *C. elegans* have a higher gene duplication rate than other species including *D. melanogaster* (Lynch and Conery 2000; Pan and Zhang 2007; Lipinski, et al. 2011), potentially as high a duplication rate per gene as 0.02 every million years (Lynch and Conery 2000). This high rate of gene duplication may account for the higher number of *hcp-3* duplications we observe in *Caenorhabditis.* However, these analyses also suggest that the vast majority of gene duplications that arise in *C. elegans* are efficiently purged by natural selection (Lipinski, et al. 2011). In contrast, our findings suggest that many *hcp-3L* paralogs are retained under purifying selection for significant periods of time.

Our evolutionary analyses thus reveal an unusual 'revolving-door' of *hcp-3L* paralogs in *Caenorhabditis* species. Under this regime, gene duplication is frequent, *hcp-3L* paralogs are retained under purifying selection for a significant evolutionary period before eventually either degenerating (*e.g.*, possibly *cpar-1* in *C. elegans*) or being lost entirely (*e.g.*, possibly *hcp-3L2* in *C. zanzibari*), returning to the ancestral state of the genome encoding only a single *hcp-3* gene. This cadence is unprecedented among most other taxonomic groups where *CenH3* duplications have been investigated. Even the high number of *hcp-3* duplications we have observed is likely an under-estimate of the true number, since extant species represent only one evolutionary snapshot. Indeed, our study implies that many previously arising *hcp-3L* paralogs have been lost or degenerated beyond recognition during *Caenorhabditis* evolution. This is akin to the 'revolving door' of HP1-family proteins previously reported in *Drosophila* (Levine, et al. 2012). Although we have not evaluated all of them in the same level of detail, duplications of other kinetochore proteins in *Caenorhabditis* also appear to occur with a similar revolving-door dynamic.

What could account for this revolving-door, *i.e.*, the short-term evolutionary retention of *hcp-3L* paralogs and their long-term loss or degeneration? We consider several possibilities for the sources of transient selective pressure to retain *CenH3* paralogs. First, this pattern could result from specialization of kinetochore paralogs for functions that are unrelated to chromosome segregation, as has been recently shown in *Caenorhabditis* and *Drosophila* neurodevelopment (Cheerambathur, et al. 2019; Zhao, et al. 2019). A previous study showed that CPAR-1 localizes to chromosomes but not centromeres in *C. elegans* (Monen, et al. 2015), although it is unclear whether this is typical for other HCP-3L paralogs.

A second explanation for this pattern might be sub-functionalization of *CenH3* paralogs for tissue- or sex-specific or meiosis-specific functions, as is proposed in *Drosophila* species (Despot-Slade, et al. 2021; Kursel, et al. 2021). Unlike monocentric chromosomes, holocentric chromosomes experience inherent challenges during meiosis, which have been overcome in different taxa via different means (Melters, et al. 2012; Marques and Pedrosa-Harand 2016). A recent study In *Meloidogyne* nematode species found that an ancestral aCenH3 is deeply conserved for function in mitosis whereas more rapidly evolving CenH3 paralogs lost mitotic function (Despot-Slade, et al. 2021). However, we found no evidence of sex-specific expression of *CenH3* paralogs in *Caenorhabditis* species. Moreover, unlike in most eukaryotes, *C. elegans* chromosomes connect to the meiotic spindle by a CenH3-independent mechanism (Monen, et al. 2005). Therefore, at least in *C. elegans*, *hcp-3* is entirely dispensable for meiotic chromosome segregation (Monen, et al. 2005). This relaxes constraints to maintain meiotic functions on *hcp-3* genes but cannot explain the revolving-door pattern.

A third possible explanation for the transient retention of *hcp-3* paralogs is suppression of either 'centromere-drive' or 'holokinetic drive'. Currently, it is unclear whether centromere drive could occur in holocentric organisms (Zedek and Bureš 2012, 2016; Krátká, et al. 2021). Although a previous study reported weak evidence of positive selection using an analysis of *hcp-3* from six highly diverged *Caenorhabditis* species (Zedek and Bureš 2012), our comprehensive reanalysis of *hcp-3* evolution across a much more densely-sampled series of closely-related species revealed no evidence of positive selection (Supplemental Table S1B). Similarly, the *aCenH3* gene required for mitosis is deeply conserved and slowly evolving in *Meloidogyne*

nematodes although other *CenH3* paralogs appear to be rapidly evolving (Despot-Slade, et al. 2021). Asymmetric meiosis in nematode species could also lead to another form of drive, leading to preferential inheritance of larger or smaller holocentric chromosomes ('holokinetic drive'), which could explain the observed negative correlation between chromosome number and genome size in many holocentric lineages (Bureš and Zedek 2014). If either of these drive mechanisms occur in *Caenorhabditis* species, then *hcp-3L* paralogs could arise and be temporarily retained as drive-suppressors, but only while the driving elements were still present in the genome. This suppression might result in loss of these driving elements from the genome, rendering *hcp-3L* gene functions superfluous and resulting in subsequent loss of these paralogs. Given the uncertainty about the existence of centromere-drive or holokinetic drive in nematodes, or the role that *hcp-3L* paralogs might play in either process, we cannot elaborate further on this possibility.

We favor a fourth hypothesis, in which the holocentricity of *Caenorhabditis* species, with HCP-3 distributed along the length of the chromosomes, might itself lead to the revolving-door dynamics of centromeric proteins. CenH3 incorporation into nucleosomes at holocentromeres is more plastic than at monocentromeres. Since CenH3 does not have to associate with specific sequences or chromosomal regions, holocentric chromosomes more easily tolerate chromosome breakage, fusion, or rearrangements. Indeed, even prior to clear cytological evidence, holocentric organisms were observed to maintain fertility despite radiation-induced chromosome breaks (Schrader 1935; Melters, et al. 2012). Moreover, even completely foreign DNA can form mini-chromosomes that assemble centromeres and be stably propagated (Zhu, et al. 2018; Lin and Yuen 2020; Lin, et al. 2021). Nevertheless, centromere distribution in holocentric organisms is

not random. Although HCP-3 presence is partially linked to certain 'HOT (High Occupancy Target) sites' in *C. elegans* (Steiner and Henikoff 2014), the overall pattern of centromere establishment in *C. elegans* appears to be predominantly linked to transcriptionally repressed genomic regions in the germline. This pattern of centromere definition via transcriptional inactivity is seen in both *C. elegans* and in the CenH3-devoid *Bombyx mori* (Gassmann, et al. 2012; Steiner and Henikoff 2014; Senaratne, et al. 2021). In contrast, some holocentric species like *Meloidogyne* nematodes and *Rhynchospora* plants localize their CenH3 proteins to specific repeats found distributed over the genome (Marques, et al. 2015; Despot-Slade, et al. 2021; Hofstatter, et al. 2022).

Although a transcriptional quiescence-dependent mode of centromere definition is more tolerant of genomic rearrangements than monocentric organisms, it could also be subject to transient stress. This stress could be imposed by either chromosomal rearrangements or transposon invasion, which can quickly and dramatically alter the landscape of transcription and repression in the germline. In such circumstances, it might be advantageous to retain HCP-3L paralogs to temporarily increase the dosage of proteins required to correctly establish centromere identity, as has been proposed in some plant lineages (Evtushenko, et al. 2021). Alternatively, it may be advantageous to express HCP-3 proteins with slightly altered sequences and localization preferences, allowing restoration of optimal centromere distributions even after periods of such 'genomic stress'. Under either scenario, eventual amelioration of the genomic stressor (*e.g.*, decay or silencing of the invading transposable element) would render *hcp-3L* paralogs superfluous and these would be lost. Therefore, we hypothesize that holocentric species like *C.*

elegans, which rely on a transcriptional quiescence-dependent mode of centromere definition, may be prone to revolving-door dynamics of their kinetochore proteins.

Different *Caenorhabditis* species might represent different stages of the revolving-door process for kinetochore proteins. Species like *C. sp54*, which possess paralogs of five of seven kinetochore genes investigated, may be actively selecting for the retention and function of these paralogs. In contrast, species like *C. elegans*, with a possibly nonessential *cpar-1* and no other kinetochore paralogs, may have already overcome the need for such innovation. We therefore predict that functional consequences of kinetochore paralog loss in different *Caenorhabditis* species will differ based on their stage of genetic innovation. Our study underlines the need for the analysis of non-model organisms and the value of evolutionary comparisons to reveal novelties even in well-studied cellular pathways.

3.4 METHODS

3.4.1 Strain maintenance

All strains were cultured on Nematode Growth Medium (NGM) plates seeded with 200 µl OP50 at 20°C using standard methods (Brenner 1974).

3.4.2 Strains used

N2 *C. elegans*

DF5081 *C. japonica*

JU727 *C. sinica*

JU1333 *C. doughertyi*

JU2744 *C. tribulationis*

JU1199 *C. afra*

VX88 *C. latens*

QG702 *C. panamensis*

3.4.3 Identification of hcp-3 and kinetochore protein homologs in sequenced genomes

To identify *hcp-3* paralogs and orthologs we iteratively queried the assembled genomes of thirty-two *Caenorhabditis* species: *C. tribulationis*, *C. sp41*, *C. zanzibari*, *C. sinica*, *C. nigoni*, *C. briggsae*, *C. remanei*, *C. latens, C. sp51*, *C. sp44*, *C. sp48*, *C. brenneri*, *C. wallacei, C. tropicalis*, *C. doughteryi*, *C. sp54*, *C. inopinata*, *C. elegans, C. oiwi*, *C. kamaaina*, *C. waitukubuli*, *C. panamensis*, *C. nouraguensis*, *C. becei*, *C. yunquensis*, *C. macrosperma*, *C. sulstoni*, *C. afra, C. sp49*, *C. sp25*, *C. imperialis*, and *C. japonica* (Supplementary Data S2). We used tBLASTn (Altschul, et al. 1990; Altschul, et al. 1997) on each species' genome (Stevens, et al. 2019) to perform a homology-based search starting with *C. elegans* HCP-3 (WBGene00001831) as our query. We used a combination of gene predictions, publicly available RNA sequencing data, *hcp-3* alignments, and splice site predictions to annotate intron-exon regions of all *hcp-3* genes that were found. To ensure that we had not missed any *hcp-3* paralogs, we repeated our analyses querying each species' hits on their own genome using tBLASTn and did not retrieve additional hits. To identify paralogs and orthologs of kinetochore proteins (Figure 3.6), we repeated this same homology-search procedure starting with *C. elegans* HCP-4 (WBGene00001832), KNL-1 (WBGene00002231), KNL-2 (WBGene00019432), ZWL-1 (WBGene00021460), SPDL-1 (WBGene00015515), NDC-80 (WBGene00003576), and HIM-10 (WBGene00001869). We used http://blast.caenorhabditis.org/ to perform all tBLASTn analyses

using pre-set parameters and setting a e-value threshold of at least 10^{-1} to obtain all possible paralogs.

Synteny was used to determine *hcp-3* orthology across *Caenorhabditis* species. We identified annotated genes immediately upstream and downstream of *hcp-3* and *hcp-3L* genes. We then used these neighbouring genes as queries for tBLASTn searches of the *C. elegans* genome to identify the orthologous syntenic genes (Figure 3.1). Dissimilar flanking genes for different *hcp-3L* paralogs provide support for the phylogenetic inference that they were acquired through independent *hcp-3* duplication events. In some cases, *hcp-3 or hcp-3L* genes were found in small genomic scaffolds or at the end of scaffolds, reducing our ability to identify upstream or downstream syntenic genes. In the latter case, we analyzed additional genes in the direction (upstream or downstream) that had sufficient genomic information available on the same scaffold. Absence of *hcp-3* in the ancestral locus in *C. sp49 hcp-3* and of *hcp-3L2* in the duplicate locus in *C. zanzibari* was confirmed by using tBLASTn of each gene in the expected locus, resulting in no detectable homologous gene sequence.

3.4.4 Phylogenetic Analyses

All protein alignments were performed using the MUSCLE algorithm (Edgar 2004) in Geneious Prime 2019.2.3 (https://www.geneious.com). Codon-based nucleotide alignments were created using the MUSCLE (codon) feature in MEGAX (Kumar, et al. 2018). We used only the HFD for phylogenetic inference and used the maximum likelihood method implemented in MEGA11 (Stecher, et al. 2020; Tamura, et al. 2021). Our amino acid-based phylogeny used the JTT model (Jones, et al. 1992) and our nucleotide-based phylogeny used the General Time Reversible model (Nei and Kumar 2000). We inferred the bootstrap consensus tree from 100

replicates. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 0.8943)). The rate variation model allowed for some sites to be evolutionarily invariable ([+*I*], 26.05% sites). All positions with less than 95% site coverage were eliminated, i.e., fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). There was a total of 267 nucleotide positions in the final dataset between all HCP-3 and HCP-3L HFD amino acid sequences. Supplementary Tables 4C and 4D present the pairwise distances and number of differences, respectively, between all *hcp-3* and *hcp-3L* HFD coding sequences.

3.4.5 Motif Analyses

13 motifs were identified using MEME (Bailey, et al. 2015) on predicted, full-length HCP-3 protein sequences from species lacking *hcp-3* paralogs (*C. zanzibari, C. nigoni, C. briggsae, C. sp51, C. sp44, C. wallacei, C. tropicalis, C. inopinata, C. oiwi, C. kamaaina, C. waitukubuli, C. nouraguensis, C. becei, C. yunquensis, C. macrosperma, C. sulstoni, C. sp25, C. imperialis, C. japonica*). E-values of all 13 discovered motifs were below 10^{-5}. Motif logo plots were generated and downloaded from MEME. Presence or absence of these motifs in all HCP-3 and HCP-3L proteins was determined by using MAST (Bailey and Gribskov 1998). We considered a motif as present in a protein by using default parameters in MAST and a P-value below 10^{-4}. Since the N-terminal tails of HCP-3 and its paralogs are highly divergent, we were not able to identify the separase motif efficiently via motif analyses. To identify presence of the

ExxR separase motif, we separately aligned each HCP-3 or HCP-3L protein sequence with *C. elegans* HCP-3 and CPAR-1 (HCP-3L1) either individually or together. This alignment was used to generate the predicted separase motifs shown in Supplementary Figure 3.S5.

3.4.6 Analysis of evolutionary selective pressures

To analyze selective pressures on *CenH3* genes, we compared rates of synonymous (dS) to nonsynonymous (dN) substitutions among *hcp-3* and *hcp-3L* genes. dN and dS between all pairwise combinations of *CenH3* genes were determined using SNAP (Korber 2000) (www.hiv.lanl.gov) on a codon alignment of the histone fold domain (Supplemental Table S1A). dN/dS ratios were used to determine the selective pressures acting on *CenH3* genes.

For all other tests, we generated codon alignments using MUSCLE (Edgar 2004), and manually adjusted them to improve alignments if needed. We also trimmed sequences to remove alignment gaps and segments of the sequence that were unique to only one species. We found no evidence of recombination for any of these alignments using the GARD algorithm at datamonkey.org (Kosakovsky Pond, et al. 2006). We used the alignment to generate a tree using PhyML maximum-likelihood methods with the HKY85 substitution model (Guindon, et al. 2010).

We analyzed selective pressures on *Caenorhabditis hcp-3* and kinetochore proteins using the codeml algorithm from the PAML suite (Yang 1997) (Supplemental Table S1A). We generated codon alignments using MUSCLE (Edgar 2004) via Geneious's Translation Align tool which we manually adjusted if needed to improve alignments. These alignments were used to generate trees using PhyML maximum-likelihood methods with the HKY85 substitution model (Guindon, et al. 2010). To test whether any residues evolve under positive selection, we

compared likelihoods between model 8 (where there are ten classes of codons with dN/dS between 0 and 1, and an eleventh class with dN/dS > 1) and model 7 (which disallows codons with dN/dS > 1) or model 8a (where the eleventh class has dN/dS fixed at 1). To test whether duplicates were evolving under positive of purifying selection, we compared the likelihood of model 0 with dN/dS fixed at 1 (neutral) with that of model 0 with dN/dS estimated from the alignment. In both cases, to determine statistical significance, we performed likelihood-ratio tests between the two models to a χ^2 distribution with the degrees of freedom reflecting the difference in number of parameters between the models being compared (Yang 1997).

3.4.7 C. elegans HCP-3 and CPAR-1 polymorphisms

To determine natural variation in *C. elegans hcp-3* and *cpar-1* genes (Supplementary Figure 3.S6), we used the *Caenorhabditis elegans* Natural Diversity Resource (Cook, et al. 2017). The synonymous mutations in *hcp*-3, as well as the frameshift, synonymous, and non-synonymous mutations in *cpar-1* were identified by the CeNDR variant annotation feature. The *cpar-1* partial deletion was found manually by looking at whole-genome sequencing reads from *C. elegans* strain ECA740 mapped onto the N2 reference genome.

3.4.8 RT-PCR

Total RNA was isolated using TRIzol (Fisher Scientific) from 50-100 L4 or young adult males, females, or hermaphrodites or from a near starved plate of mixed-stage animals. RNA was extracted by chloroform extraction, precipitated using isopropanol, washed with ethanol, and resuspended in 20µl of nuclease-free water. Next, RNA was treated with DNase I (New England Biolabs, 2 units/µl) at 37°C for 60 minutes followed by heat inactivation at 75°C for 10 minutes.

DNase-treated RNA was purified using the RNA Clean and Concentrator-5 kit (Zymo Research) and converted to cDNA using SuperScript III Reverse Transcriptase (Invitrogen) using polydT primers as per manufacturer's recommendations. RNA concentrations used to make cDNA were not kept the same between whole plate, male, and female/hermaphrodite samples except for samples from *C. afra* (in Figure 3.5C), *C. remanei* and *C. sinica*. PCR was done on cDNA using Phusion High-Fidelity DNA Polymerase Kit (New England Biolabs) guidelines according to the manufacturer's recommendations using primers for *hcp-3, hcp-3L,* and *tbb-2*. All primer sequences used are listed in Supplemental Table S2.

3.5 ACKNOWLEDGMENTS

We thank Ching-Ho Chang, Ines Drinnenberg, Lisa Kursel, André Marques, Isabel Meija Natividad, Risa Takenaka, Paul Talbert, Toshio Tsukiyama, and Janet Young for comments and discussions on the manuscript, the Bai lab at Fred Hutchinson Cancer Center for nematode resources, Lewis Stevens at Wellcome Sanger Institute for providing species phylogenies, and Matthew Rockman at New York University for providing the *C. panamensis* strain. Some strains were provided by the *Caenorhabditis* Genetics Center (CGC), which is funded by NIH Office of Research Infrastructure Programs (P40 OD010440). This work was supported by grants from the National Institutes of Health (NIH) Institutional Training Grant (T32 GM007270 to L.C.), the Republic and Canton of Geneva (to F.A.S), the Swiss National Science Foundation grant 310030_197762 (to F.A.S.), the National Science Foundation (NSF) CAREER Award MCB-1552101 (to M.A.), the NIH National Institute of General Medical Sciences grant R01 GM074108 (to H.S.M.), and from the Howard Hughes Medical Institute (to H.S.M.). The funders played no role in study design, data collection and interpretation, or the decision to

publish this study. P.R. is a Washington Research Foundation Fellow and H.S.M. is an Investigator of the Howard Hughes Medical Institute.

3.6 FIGURES

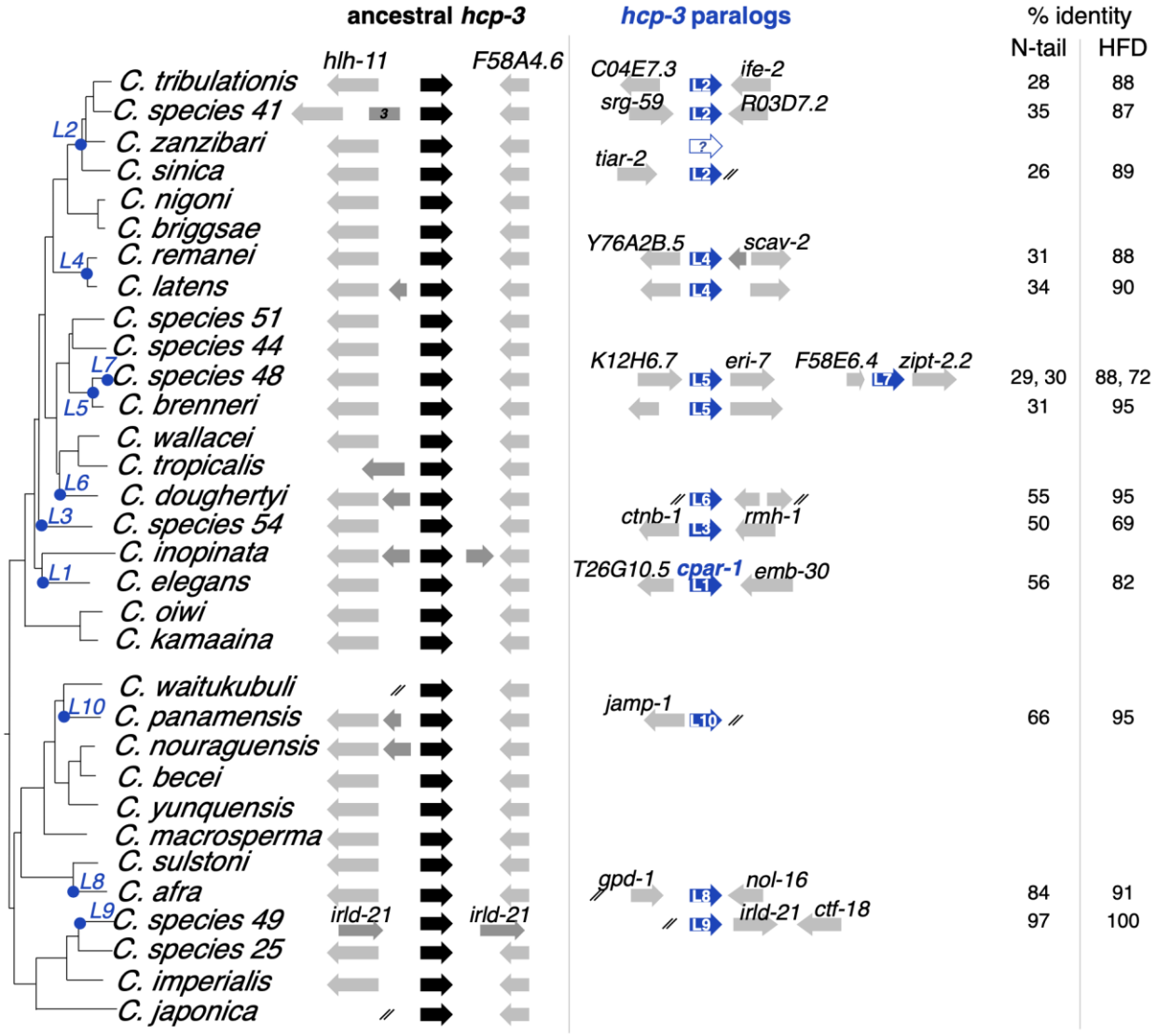


Figure 3.1. Ten independent *hcp-3* duplications in Caenorhabditis species.

A schematic representation of ancestral centromeric histone genes (*hcp-3*, black) and their duplicates (*hcp-3L*, blue) are shown alongside a *Caenorhabditis* species tree (adapted from http://caenorhabditis.org). *hcp-3* duplication events are represented on the species tree with a blue dot and numbered *L1* through *L10*, with paralogs arising from independent duplications assigned different numbers. Genes in the syntenic neighborhood near *hcp-3* and *hcp-3L* are

represented in grey and labelled with their orthologous gene names in *C. elegans*. In some cases, 1-3 genes were inserted between *hlh-11* and *F58A4.6* within the syntenic neighborhood of *hcp-3*. The white arrow with a question mark represents a possible loss of *hcp-3L2* in *C. zanzibari*. Ends of genomic scaffolds are denoted with two slashes. On the right, we show percent amino-acid identities between the paralog and ancestral *hcp-3* of each species (in the N-terminal tail or histone fold domain (HFD)).

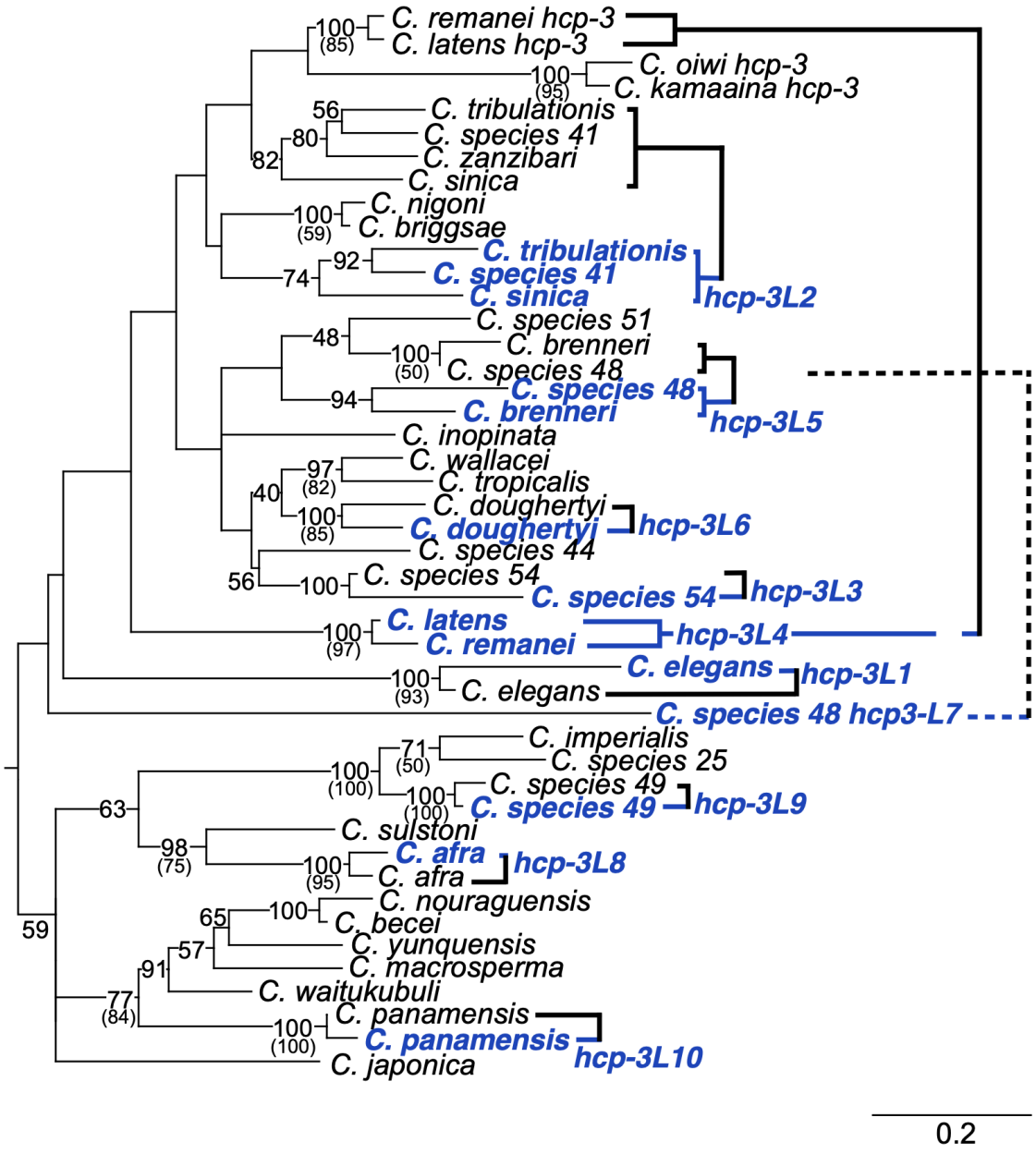


Figure 3.2. Phylogenetic analysis of *hcp-3* and *hcp-3L* genes in *Caenorhabditis* species. A maximum likelihood tree of a DNA, codon-based alignment of the HFD of ancestral *hcp-3* (black) and *hcp-3* paralogs (blue) is shown. Bootstrap values of 40 and above are indicated. Bootstrap values in parentheses are from corresponding nodes from a maximum likelihood tree based on an amino acid alignment of the HFD (see Supplementary Figure S2). In all except a few instances, the nucleotide and amino acid tree are in agreement, with higher bootstrap support

observed in the nucleotide tree. For the exceptions (nodes representing *hcp-3* or *hcp-3L2* in *C. tribulationis*, *C. sinica*, *C. sp41*, and *C. zanzibari*, and the node representing *hcp-3* in *C. nouraguensis*, *C. becei*, and *C. macrosperma*), bootstraps values were not included here since they were lower in the amino acid tree and because they do not alter conclusions from the nucleotide tree. A scale bar (branch lengths, substitutions per site) is shown at the bottom-right. On the right, thick lines show *hcp-3* paralogs from same species, the dashed line shows the second duplicate found in *C. species 48*.

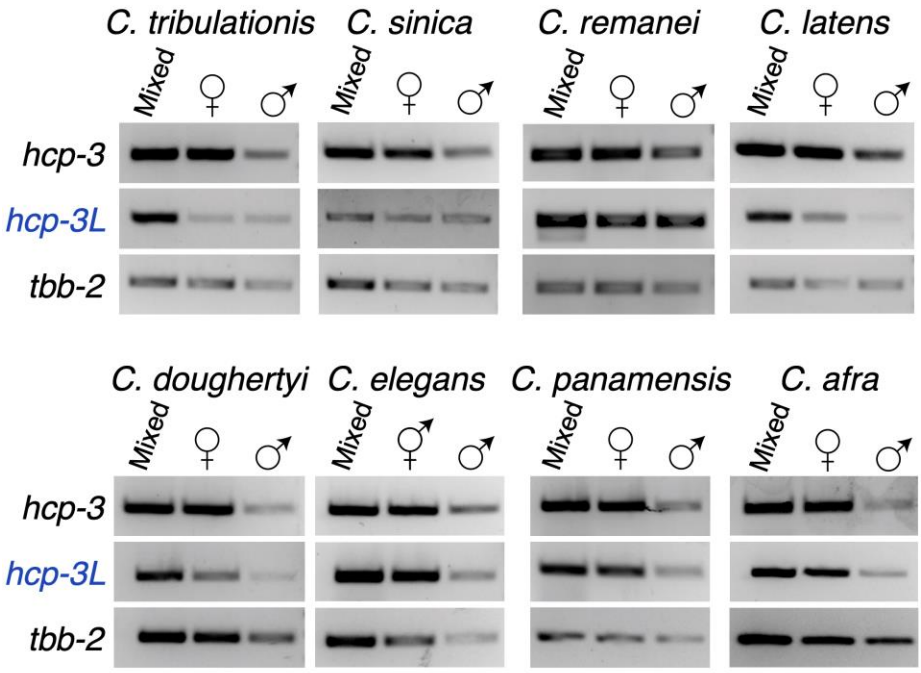


Figure 3.3. *hcp-3L* genes are expressed in both sexes in Caenorhabditis species.

RT-PCR of ancestral *hcp-3* (top), *hcp-3L* (middle), or *tbb-2* (bottom; loading control) in species with *hcp-3* duplicates. RNA from a mixed worm population of various larval stages, L4 or young adult females/hermaphrodites or L4 or young adult males were used.

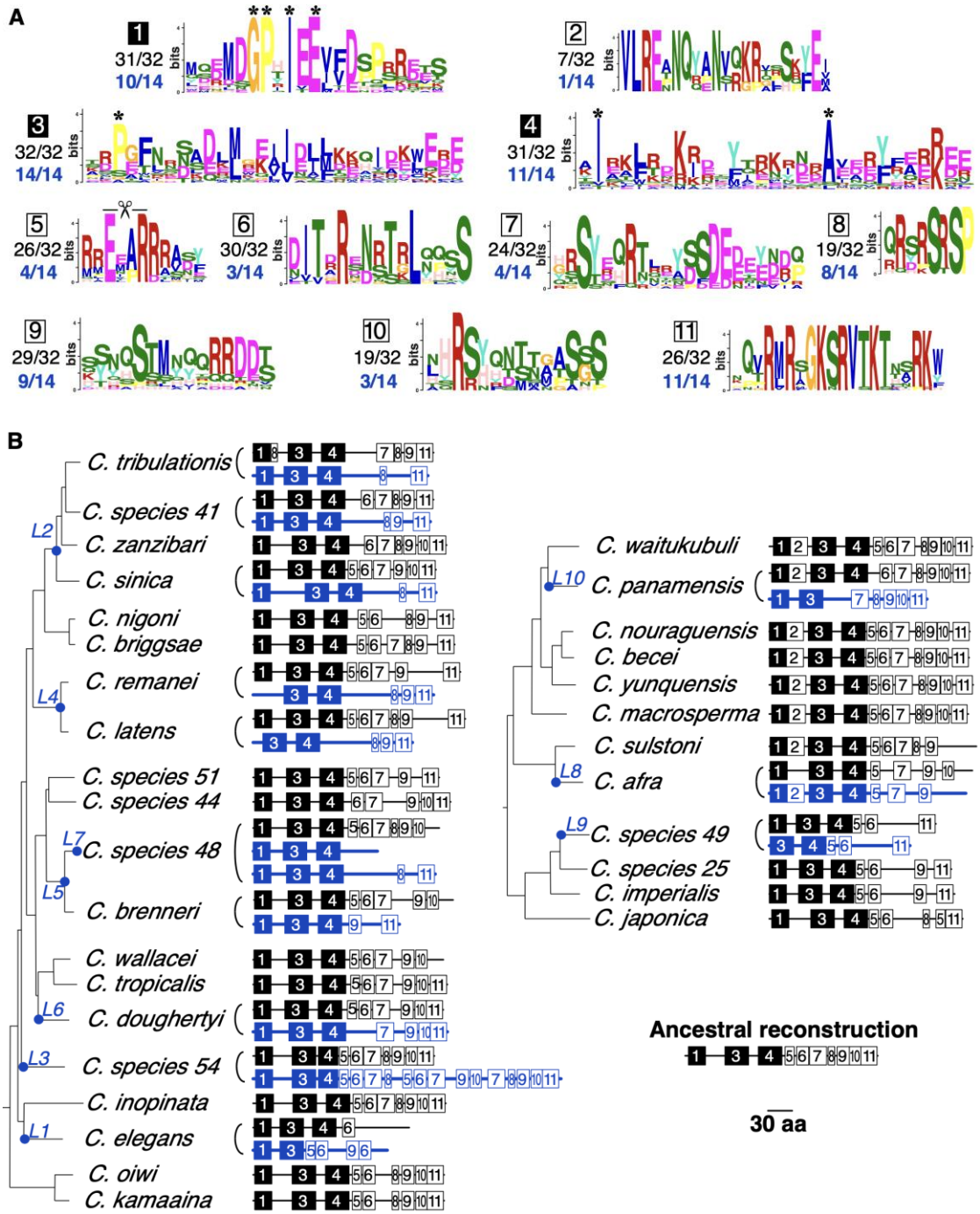


Figure 3.4. Differential retention of N-terminal tail motifs across HCP-3 and HCP-3L proteins encoded by Caenorhabditis species.

(A) Logo plots of eleven protein motifs within HCP-3 N-terminal tails discovered from an analysis of *Caenorhabditis* species without duplications. Motifs 12 and 13 are C-terminal motifs

(not shown, see Supplementary Figure S4) that reside within the HFD. E-values of all motifs were below 10^{-5}. Asterisks above logo plots for motifs 1, 3, and 4 indicate residues that are highly conserved within the motif. Proportion of all 32 ancestral HCP-3 proteins (black) or 14 HCP-3L duplicates (blue) that have retained the motifs are shown. (B) *Caenorhabditis* species tree with schematics of protein motifs that are present (numbered boxes) in ancestral HCP-3 (black) or HCP-3L (blue) in each species is shown. The presence of motif 1 in *C. elegans* and motif 4 in *C. sp54* was not detected by unsupervised MAST searches but was subsequently ascertained through manual alignments (see Supplementary Data File S4). All proteins contained a conserved, C-terminal HFD (not shown). Filled black boxes represent three motifs that show the highest retention in *Caenorhabditis* HCP-3 proteins. A structure of the N-terminal tail of HCP-3 in the last common ancestor of *Caenorhabditis* was inferred based on the retention and loss of motifs in the N-terminal tail. L1-L10 on the species tree indicate *hcp-3* duplication events as in Figure 3.1. A scale bar (number of residues) is shown on the bottom-right.

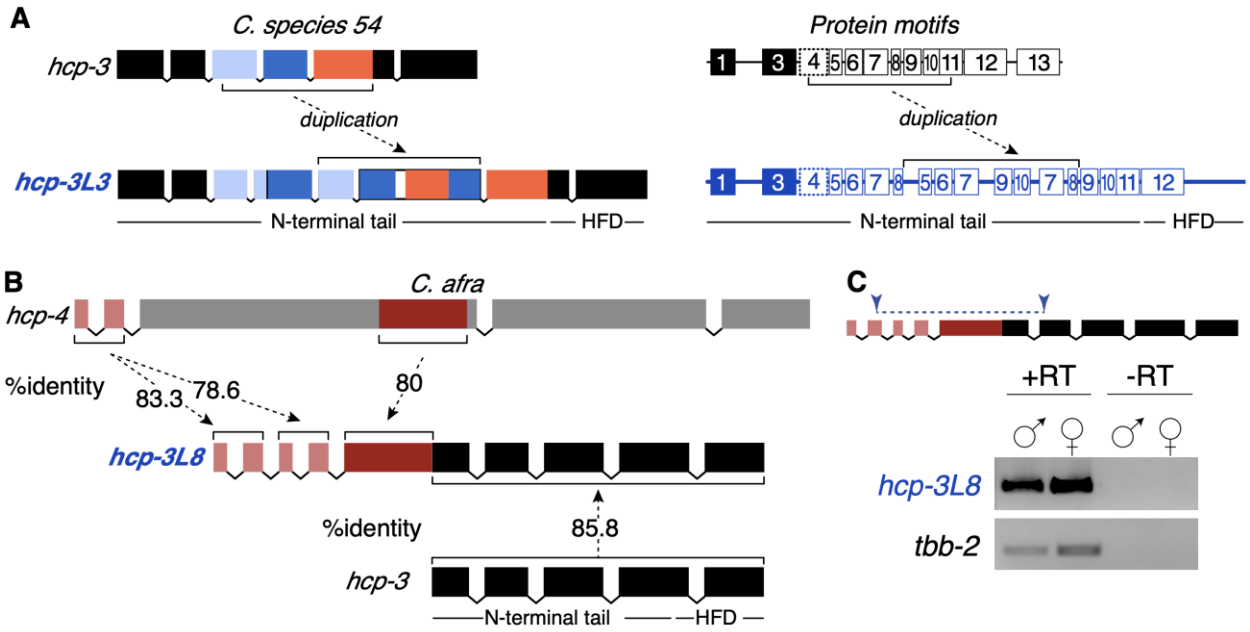


Figure 3.5. Two unusual Caenorhabditis *hcp-3L* paralogs arose by internal duplication or gene fusion.

(A) Schematic of the exon structure (left) and protein motif structure (right) of *C. sp54 hcp-3* (top) and *hcp-3L3* (bottom). Portions of *hcp-3* exon 3 (light blue), exon 4 (dark blue), and exon 5 (orange) are duplicated within the N-terminal tail of *hcp-3L3* (dashed arrow). Similarly, motifs 5-10 are duplicated within the N-terminal tail of HCP-3L3. Motif 13 resides within the HFD and is missing in HCP-3L3. The HFD is not within the duplicated region. (B) Schematic of the exon structure of *C. afra hcp-3L8* (middle) with homology to *C. afra hcp-4* (top) and *C. afra hcp-3* (bottom). The first five exons of *hcp-3L8* are homologous to *C. afra hcp-4* exons 1 and 2 (light red) as well as a portion of exon 3 (dark red). The last five exons of *hcp-3L8* are homologous to *C. afra hcp-3* (black). The HFD and the N-terminal tail of *hcp-3* are denoted. Percent amino acid identity between protein-coding exons are shown. (C) Primers designed to span exons that are homologous to *hcp-3* and *hcp-4* within *hcp-3L8* (top). Schematic of the gene shows primers used to amplify the *hcp-4-hcp-3* fusion region (top, blue) in RT-PCR of *C. afra hcp-3L8* and *tbb-2* in

males and females (bottom) to confirm expression of a chimeric transcript. +RT and -RT indicate cDNA preparation with or without reverse transcriptase enzyme, respectively.

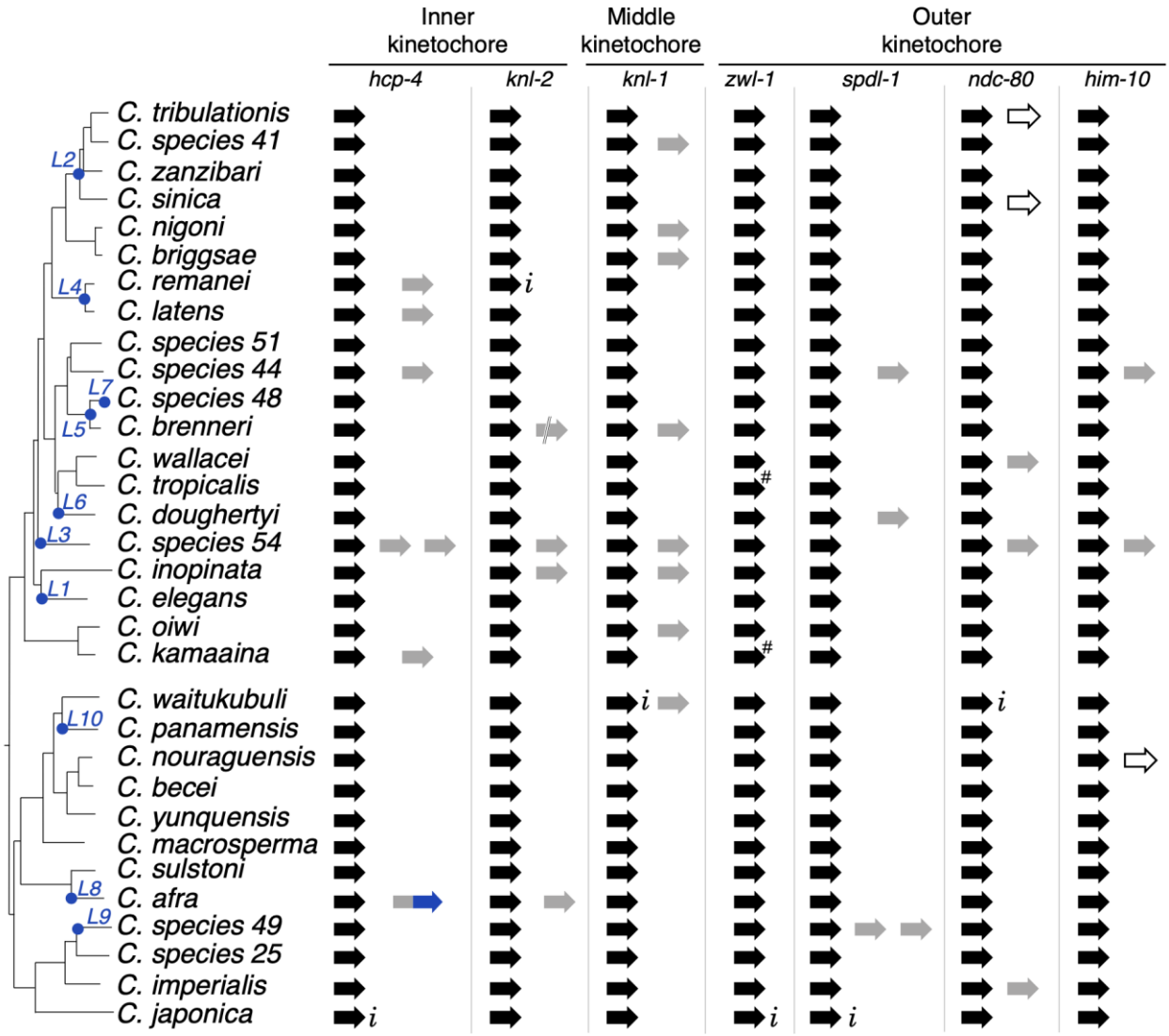
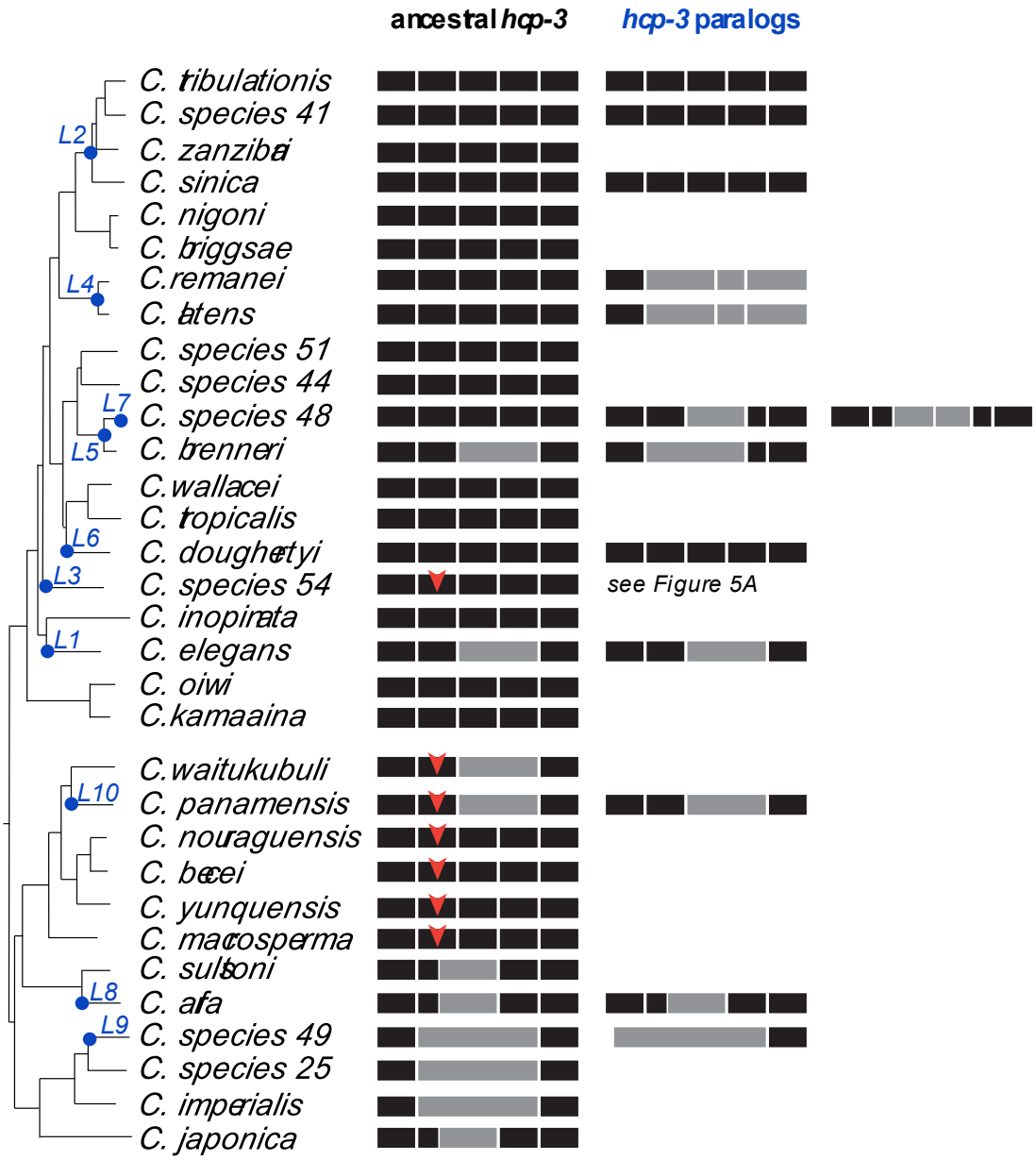


Figure 3.6. Duplication of kinetochore proteins in Caenorhabditis species.

A schematic representation of ancestral (black) and duplicate (grey) copies of seven kinetochore genes (*hcp-4*, *knl-2*, *knl-1*, *zwl-1*, *spdl-1*, *ndc-80* and *him-10*) shown alongside a *Caenorhabditis* species tree. *hcp-3* duplication events are denoted as a blue dot on the species tree, as in Figure 1. The unique fusion between *C. afra hcp-4* and *hcp-3* duplicates is shown in grey and blue. Incomplete sequence information in genomic scaffolds is denoted with *i* and apparent pseudogenes are denoted as unfilled arrows. Double slash in *C. brenneri knl-2* duplicate indicates

the sequence was split between two scaffolds. # indicates two potential pseudogenization events in *zwl-1* that are likely to represent sequencing errors.

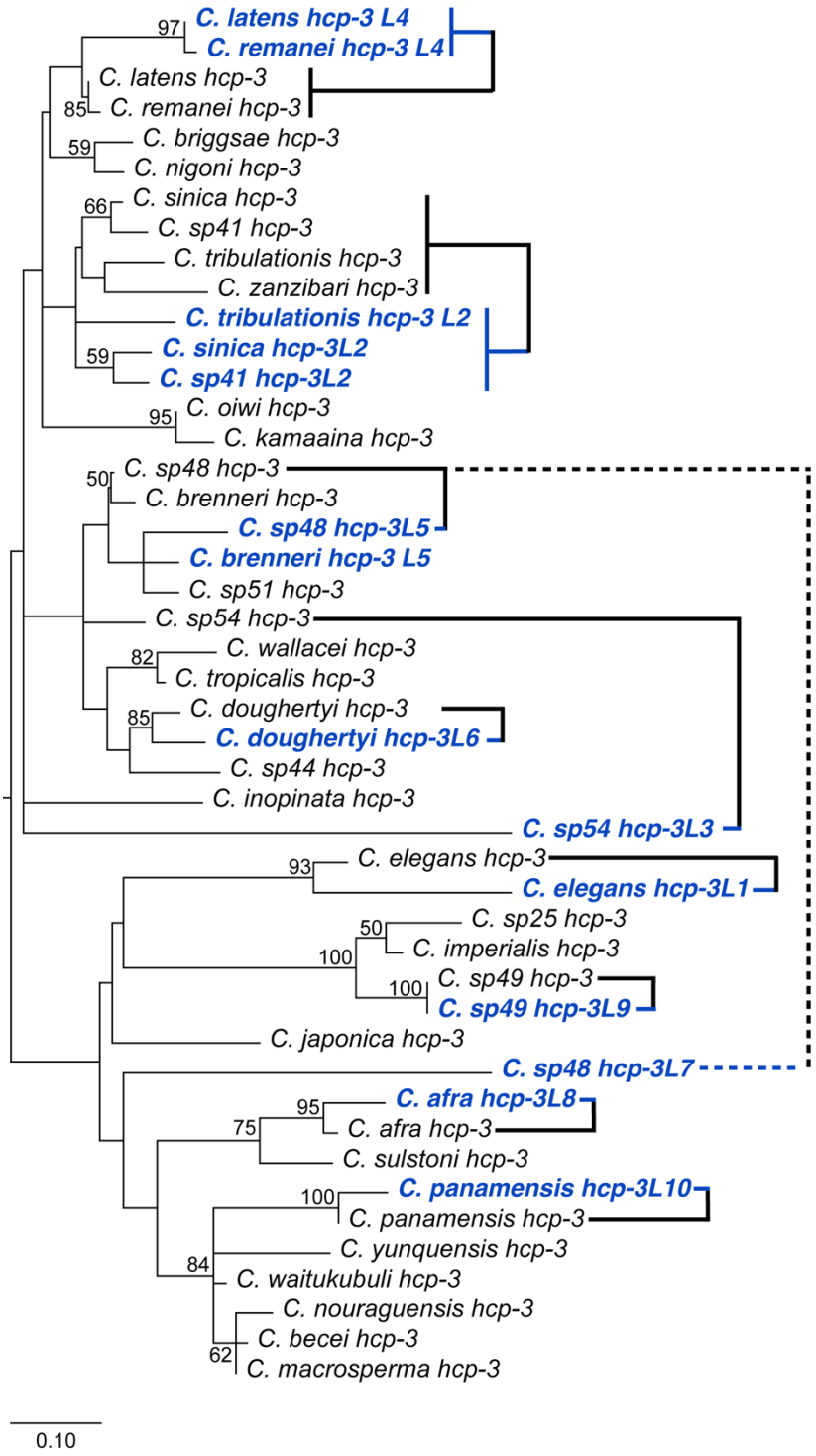
3.7 SUPPLEMENTARY MATERIALS



Supplementary Figure 3.S1. Exon-intron junctions are largely retained in *Caenorhabditis hcp-3* and *hcp-3L* genes except for intron gains and losses that are likely a result of partial retrotransposition.

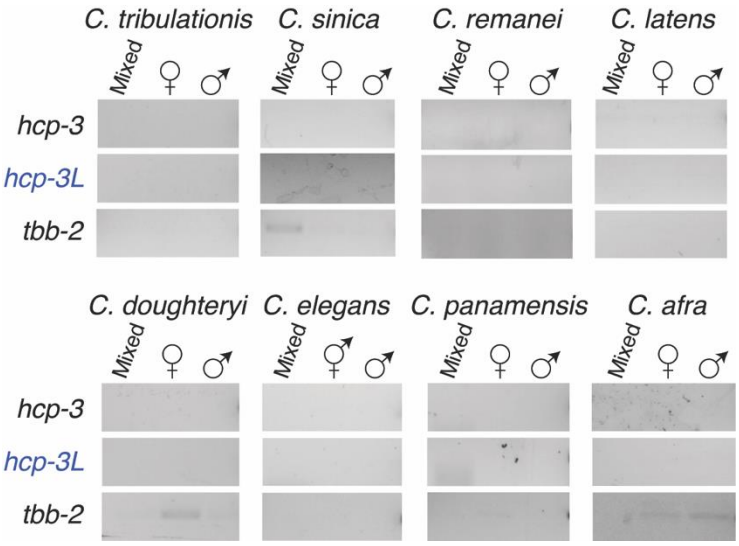
Schematic of the exons of *hcp-3* and *hcp-3L* genes in *Caenorhabditis* species. Each box represents an exon with black boxes showing ancestral exons. Grey boxes depict exon fusion

events that are a likely consequence of partial retrotransposition and overwriting of the genomic locus as has been previously observed (Robertson 1998; Cho, et al. 2004; Kiontke, et al. 2004). Red arrows indicate insertion events that likely create new introns. Exon duplications for *C. sp54 hcp-3* and *hcp-3L3* are discussed in more detail in Figure 5A in the main text. An N-terminal extension to *C. afra hcp-3L8* (homologous to *hcp-4*) is not shown but discussed in more detail in Figure 3.5B.



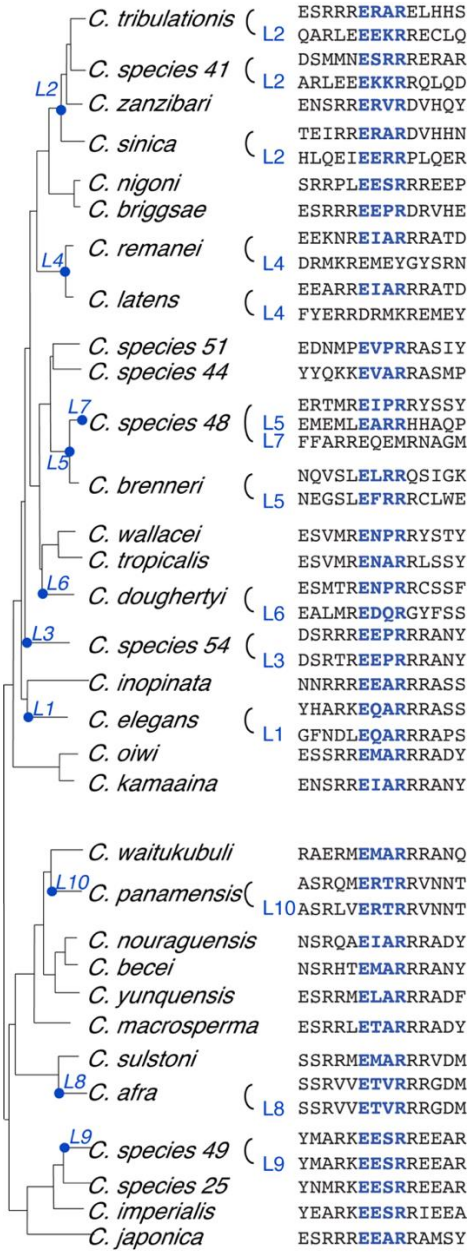
Supplementary Figure 3.S2. Maximum-likelihood phylogenetic tree based on an amino acid alignment of HCP-3 and HCP-3L proteins encoded by Caenorhabditis species.

A maximum likelihood tree based on an amino acid alignment of the histone fold domain (HFD) of ancestral *hcp-3* (black) and *hcp-3* paralogs (blue) is shown as a phylogram (branch lengths are scaled to evolutionary divergence indicated). Bootstrap values of 40 and above are indicated. Overall, this phylogeny is much more poorly resolved than one based on the nucleotide alignment (Figure 2) and does not fully recapitulate known relationships between *Caenorhabditis* species or relationships between *hcp-3* and *hcp-3L* genes from the same species. However, well-resolved nodes agree between both the protein and nucleotide phylogenies and are also indicated in Figure 2.



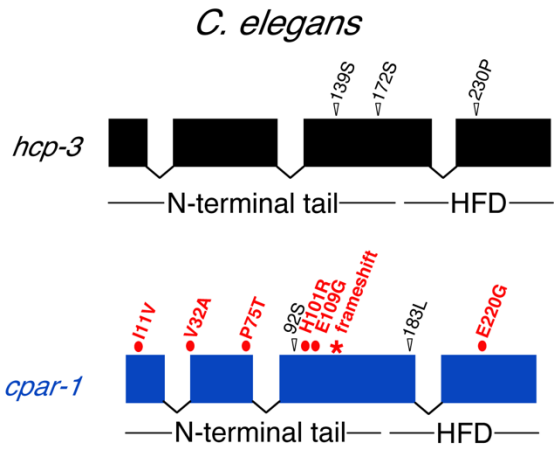
Supplementary Figure 3.S3. RT-PCR controls for expression analysis of *hcp-3* and *hcp-3L* genes. No reverse transcriptase (-RT) PCR control of ancestral *hcp-3* (top), *hcp-3L* (middle) paralogs, or *tbb-2* (bottom) loading control genes in selected species with *hcp-3L* duplicates. RNA from a mixed worm population of various larval stages, L4 or young adult females/hermaphrodites or L4 or young adult males were used. In some cases, we see very faint bands in the *tbb-2* -RT controls compared to +RT samples.

Supplementary Figure 3.S4. Conserved motifs identified in the HCP-3 histone fold domains. Logo plots of conserved motifs 12 and 13, which reside in the HCP-3 histone fold domain. Proportion of all 32 ancestral HCP-3 proteins (black) or 14 HCP-3L duplicates (blue) that have retained the motifs are shown. All HCP-3 and HCP-3L proteins contain motifs 12 and 13, except for *C. sp54* HCP-3L3 which has a poor statistical match to motif 13 owing to high divergence (see Figure 5A).



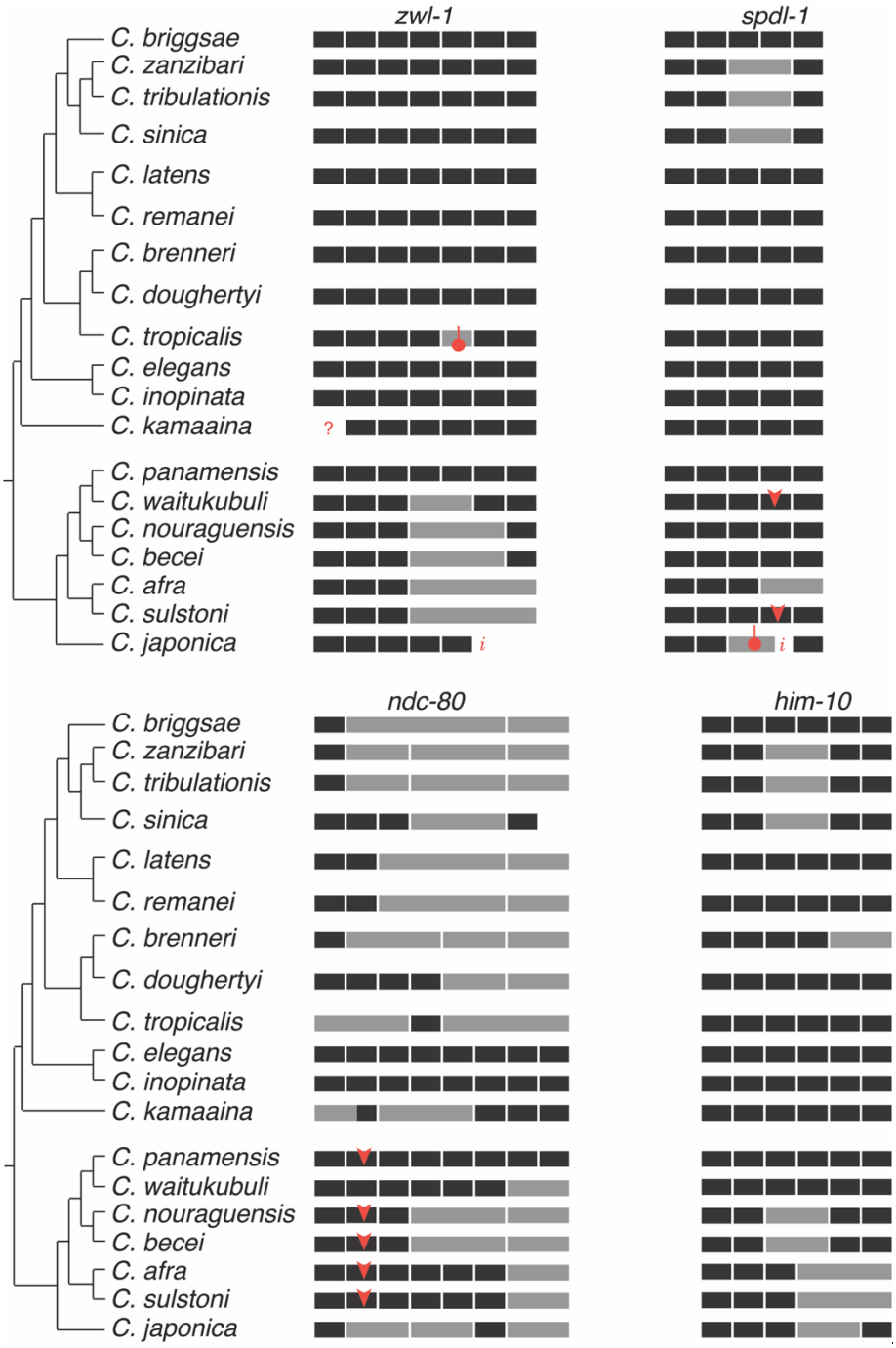
Supplementary Figure 3.S5. The ExxR putative separase cleavage site is retained in most Caenorhabditis HCP-3 and HCP-3L proteins.

Alignments of ExxR residues (blue colored) and flanking residues in HCP-3 and HCP-3L proteins are shown alongside a *Caenorhabditis* species tree.



Supplementary Figure 3.S6. A higher frequency of function-altering mutations is observed in *cpar-1* compared to *hcp-3* within natural C. elegans populations.

Schematic of exon-intron structure of *hcp-3* (top, black) and *cpar-1* (bottom, blue) coding regions in *C.* elegans, indicating the N-terminal domain and HFD. Natural variation found in *C. elegans* strains is indicated by arrowheads (black) and ovals (red) that represent synonymous and nonsynonymous mutations respectively. Three synonymous mutations and zero nonsynonymous mutations were found in *hcp-3*, whereas six nonsynonymous and two synonymous mutations were found in *cpar-1*. In addition, a single nucleotide insertion in *cpar-1* that causes a frameshift resulting in an early stop codon was found.



Supplementary Figure 3.S7. Other kinetochore genes also display patterns of partial intron loss in Caenorhabditis species.

Schematic of the exon arrangements of four kinetochore proteins (*zwl-1*, *spdl-1*, *ndc-80* and *him-10*) in a representative set of *Caenorhabditis* species. Each box represents an exon with black

boxes showing ancestral exons and grey boxes showing fusion events between exons that likely arose due to partial retrotransposition and overwriting of the genomic locus (Robertson 1998; Cho, et al. 2004; Kiontke, et al. 2004). Red arrows indicate insertion events that likely create new introns and red dots represent deletions in exons. Incomplete genomic sequence information is indicated with an *i.* Although we were unable to identify exon 1 of *C. kamaaina zwl-1* using homology, it is unlikely to have been pseudogenized, since *zwl-1* is an essential gene and is present as an intact gene in all related species.

Supplementary Data 3.S1: Pairwise alignments of Caenorhabditis HCP-3 HFD and N-tail used for Figure 1

***C. tribulationis* HCP-3/HCP-3L2 N-tail alignment**

>C_tribulationis_HCP-3_N-tail
---MYA--
HTGPIIEEVEEAATDGQTVHRDWSKDPDVLRIRRELQKLMILPGFNNNADLMQRAITILV
EQVDEWKLDQDIDGWNICRQEKIESFLPRIANFKNKRAQAIDQFYKERDSMINESRRRER
ARELHHSNDFGISGRELDHSSRLHQLSRRDSCASRVERYHSSDEDEENHPVPRYRSRSPG
PSSSYNQST-MRQRRDDVPQPV-RMRSGKSRVTKTRNAKWRPG
> C_tribulationis_HCP-3L2_N-tail
MEQMYENQHT-PIIEELFDCSS---VVERE------
VERVKHEIQALTSQSDFNKNYASMKEVINILTRQIAAWDADEDMGGSHPIRLRSIEKFHA
KRVLFTEKLEAAERAYYEKKQARLEEEKRRE---CLQDRGKIAGQNNQLCHR-
QGHRYERDDSSDDSSDEENQRQRSRACSPQRRNHPSTSSQQYRVNRHVDYHKKHQNVS
QKQQRLRAGINAVTKTKVRKFRPG

***C. tribulationis* HCP-3/HCP-3L2 HFD alignment**

>C_tribulationis_HCP-3_HFD
QKALREIRKYQKSTDMLIQKAPFARLVHEIIQETTLFSHDFRIRADALMALQEASEAFMV
EMFEGSFLICNHAKRVTLMPTDIQLYRRLCLR
>C_tribulationis_HCP-3L2_HFD
QKALAEIRRYQKSTDMLIQKAPFARLVHEIIQESTTLSRDFRIRSDALMALQEGAEAFMV
EMFEGSALICNHAKRVTLMPTDVQLYRRLCLR

***C. sp41* HCP-3/HCP-3L2 N-tail alignment**

>C_sp41_HCP-3_N-tail
---MYA-
HTGPIIEEVEDGPAEGHTVHRDWRQDTDVQRLGREIQKFISLPDFSKNADLMQRAIHVLE
KQVDEWKLNQELDGWDHHRQQKIELFQPKIAKFKEKREEAINHYYDVKDSMMNESRR
RERAREVLHNTDQNITGFGNSTRLYPNSRRQSFAPRKERYQSSDE--
DEENDPVPRRRSRSPGPSSSYSHSAMYQLRDD----------
SNAPQQQRMRSGKSRVTKTKNRKWRPG
>C_sp41_HCP-3L2_N-tail
MDQLYENHHTPVIEEILD---TEFVVERE------
VERVKHDIQLITSQPDFNKNYDSMKEVIDILARQILDWEADEEMSGSHPMRRRIIDKFQA
KKVLFTEKLEAAERAYYERKRARLEEEKKRRQLQDC-----GNIGGEQNR-----
HARFQFMGHRNGRDDSSDDSSDGENRQVQRRRSRSRSPQHR-
NHSSQQHRRDDRHADGHRNYQSTSKNQPRLRAGINGVTKTKVRKYRPG

***C. sp41* HCP-3/HCP-3L2 HFD alignment**

>C_sp41_HCP-3_HFD
QKALSEIRKYQKSTDMLIQKAPFVRLVNEIIQEATSFSKEFRIRADALMALQEASEAFMVE
MFEGSVLICNHAKRVTLMPTDIQLYRRLCLR
>C_sp41_HCP-3L2_HFD

QKALAEIRQYQKSTDMLIQKAPFARLVHEIIQDSTNFSRDFRIRADALMALQEAAEAFMV
EMFEGSTLICNHAKRVTLMPTDLQLYRRLCLR

***C. sinica* HCP-3/HCP-3L2 N-tail alignment**

>C_sinica_HCP-3_N-tail

MYGHTGPLIQEIDETPTEGSTAVRDWTTDEDVRRLGKEVQKLTCQEGFTKNANLMRRLI
ELLEKQVDEWKEDQDIHGHDHCRQQKIQCFEKRIADYKKKCERSIRRYCDQRDSRTEIR
RERARDVHHNSNYDITDRGDSIRLNQHYHRQSLAPREESYHSSDDDEENIPPRLYRSGRS
TMYQQRQDDSNVHYRSHHSTLGASSSQQVRMRSGKSRVTKTHNRKWRPG
>C_sinica_HCP-3L2_N-tail
MDHVPPPLTHQPPSLIAAFLPRILVFEVKMQAMQCTPIIEEIHEAPT---------
ALEMEIEVVKHNIKKLTSQPDFSKSYSLMEGAITILKKQIERWEYAERRDGPDGARQEHL
AKFRLKMEQFEEKLKAAERAYYEKKKAHLQEIEER-RPLQERQN----------
AINRQSFEQRRRERRNSSDESSDDETE---QFHHRSRSRSPRRNQLTTSNLQRHRQPT-------
QPRLRQGVDRISKTKARKWRPG

***C. sinica* HCP-3/HCP-3L2 HFD alignment**

>C_sinica_HCP-3_HFD
QKALAEIRKYQKSTDMLIQKAPFARLVHEIIQEATSFSKEYRIRADALMALQEASEAFMV
EMFEGSVLICNHAKRVTLMPTDIQLYRRLCLR
>C_sinica_HCP-3L2_N-tail
QKALAEIRQYQRTTEMLIQKAPFARLVHEIIQDATSFSRDFRIRADALMALQEAAEAFMV
EMFEGSVLICNHAKRVTLMPTDLQLYRRLCLR

***C. remanei* HCP-3/HCP-3L4 N-tail alignment**

>C_remanei_HCP-3_N-tail
MYQMHHNGPRIEEMVDPPSRST-
TNQLKNDTEYIKSEYRRISHLPDFNRDPELIQEVMNLTKRYIEKWLREERD-
EPNMERQGWIERFKTKLREWETKKETAEDEYYTRRDASSNEEKNREIARRRATDSQMNI
TGLHDSTRLNQQSYSRSYENRNRRYSSDEDDDENMAPQRRQRSRSPPSFAHHQRRDDTG
SYYRSHHTQNSSNQRTHNTDFSSHYRGQYGPSTSQNVGMPS--
NAQNVRMRSGKSRVTKTRSRKWRPG
>C_remanei_HCP-3L4_N-tail
---MAPLLIAVSSFVVCSAILIYLCSKKPKTIDLESEI-
GITSRKNFNRDSDSMQEVIDIMTRQINKWEQLEDDYGPDATRQRNIEAFQRKRDLWEEK
KEQAERAFYERRD-----RMKREM----------EYGYSRNQIRRERRMDDDSDSD-
VMEDDRRQ----PLGNLDYSRRDNYVRLREAKPMALLNRDRSRSRSPLLLPR-
QDHPSTSLQLRRPNIPSTPPVRVRPGKSRVTKSKNRKWRPG

***C. remanei* HCP-3/HCP-3L4 HFD alignment**

>C_remanei_HCP-3_HFD
QRALEEIRKYQKSTDMLIQKAPFARLVHEIMREATSESQDFRIRADALMALQEAAEAFM
VEMFEGSVLICNHAKRVTLMPTDIQLYRRLCLR
>C_remanei_HCP-3L4_HFD

QKALLEIRKYQKSTDMLIQKAPFARLVQEILRETTNESHDYRIRADALMALQEGAEAFM
VEMFEGSVLISNHAKRVTLMPTDVQLYRRLCLR

***C. latens* HCP-3/HCP-3L4 N-tail alignment**

>C_latens_HCP-3_N-tail
MYQVYHNGPRIEEMVDPPPKSTTNQLKIDTEYIKSEYGRISSHSDFNRNPDAIQEVIDLAR
RYIEKWQREERD-
EPNMDRRGWIERFKTKLREWETKKETAVDDYVTRRNASSNEEARREIARRRATDSQLNI
TGLQDSTRLNQQSYSRSYENRNRRYSSDEDDDENMAPQRRQRSRSPSSFAYHQNTLNHQ
RRDDTGSYYRSHHTQNSSHQRTHNTDISSHYRRQNGPSTSQNVVMPSNTQNVRMRSGK
SRVTKTRNRKWRPG
>C_latens_HCP-3L4_N-tail

MDQIKEEIEAITSRKNFNRDSDAMQEVIDIMTRQINKWEQLEDDYGPDATRQRNIEAFQR
KRDSWEEKKEQAERAFYERRD-----RMKREMEY----------
DNGYSRSQVRRERRMDDDSDSDV-MEDDGRQ-
PLGNLDYSKRNNFVRGREAMPMKVVNRDRSRSRSPLPRPHHPSTSLQLRRPNIPSSP-----
P-----VRVRPGKSRVTKSKNRKWRPG

***C. latens* HCP-3/HCP-3L4 HFD alignment**

>C_latens_HCP-3_HFD
QRALEEIRKYQKSTDMLIQKAPFARLVHEIMREATSESHDFRIRADALMALQEAAEAFM
VEMFEGSVLICNHAKRVTLMPTDIQLYRRLCLR
>C_latens_HCP-3L4_HFD
QKALLEIRKYQKSTDMLIQKAPFARLVQEILRETTNESHDYRIRADALMALQEGAEAFM
VEMFEGSVLISNHAKRVTLMPTDIQLYRRLCLR

***C. sp48* HCP-3/HCP-3L5 N-tail alignment**

>C_sp48_HCP-3_N-tail
MFRVTDGPTIEEVVETQLTEDTAEAEVRRDYDTILEELRAVLGVPGANRDQERLGRGLCI
LEKGIDKFEEDEDN-
QPLEIRRQYLGKLREKYSSCEAKLREAENAFHERKEREYEERTMREIPRRYSSYRDTDIT
RRNNTTGLYHHSQQSSSNYRQQGYSSDEEMENFPSSHRDRYRSPPRKFSHSTMLQQRRD
ISPVVNRSHQQSSASSQQVRMRSGKSRVTKTTRKYRPG
>C_sp48_HCP-3L5_N-tail
-MMDEDSPRIEEIVE----
EDEAEKDVKKEFDEYRREIEAVTSLPGFNRDSGKMTQVLRIMEKAIGKWEEDEENLGSI
QCRRRCLLEFKRRHDNYEKIIEKAEDDFYKRREMEMLE----------ARR-------
HHAQPESRQYPDGG----------AEQRQTIP----------
KMRAGKSSVTKKPKKFRPG

***C. sp48* HCP-3/HCP-3L5 HFD alignment**

>C_sp48_HCP-3_HFD
QKALAEIRQYQKSTDLLIQKAPFARLVHEIIREATSNSGDYRVRADALLALQEGAEAFMV
EMFEGSVLICNHAKRVTLMPTDIQLYRRLCLR

>C_sp48_HCP-3L5_HFD
EKALAEIRQYQRSTDLLIQKAPFARLVHEIVSEATSSSGDYRIRADALMALQEGAEAFMV
EMFEGSALICNHAKRVTLMASDVQLYRRLCLR

C. sp48 **HCP-3/HCP-3L7 N-tail alignment**
>C_sp48_HCP-3_N-tail
MFRVTDGPTIEEVVETQLTEDTAEAEVRRDYDTILEELRAVLGVPGANRDQERLGRGLCI
LEKGIDKFEEDEDNQ-PLEIRRQYLGKLREKYSSCEAKLREAENAFHERKEREYEE---
RTMREIPRRYSSYRDTDITRRNNTTGLYHHSQQSSSNYRQQGYSSDEEMENFPSSHRDRY
RSPPRKFSHSTMLQQRRDISPVVNRSHQQSSASSQQVRMRSGKSRVT-KTTRKYRPG
>C_sp48_HCP-3L7_N-tail
MLEDHDSPHIEELVDAD-
DEKNAEEAVKKEFFEFKNEIEKISCLPNFTKDSEKLRQVLAVIGKAIDKWEEDEDNEGSIE
VRRKYLKELKERYFKYDRIIEKAEQKFFARREQEMRNAGMKTTRYYPGHGSSNHQYPFS
REEDVV----NSRSPSRHRPNSGHPHQHAPKLSSCRQQHGYNYWDEKLCSRS---
PRRSNLPIRDQFHQ—SSTNAQRELKAGKSRVTKKVSHKYCSG

C. sp48 **HCP-3/HCP-3L7 HFD alignment**
>C_sp48_HCP-3_HFD
QKALAEIRQYQKSTDLLIQKAPFARLVHEIIREATSNSGDYRVRADALLALQEGAEAFMV
EMFEGSVLICNHAKRVTLMPTDIQLYRRLCLR
>C_sp48_HCP-3L7_HFD
QRAIAEIKHYQKTTELLIQKAPFARLVQEVVQEATSESSSYSIRTDALSALQEGAEAFLVE
MFEGSSMIANHAKRATLGSTDLKLYRRLCLR

C. brenneri **HCP-3/HCP-3L5 N-tail alignment**
>C_brenneri_HCP-3_N-tail
MFHLSDGPTIEELVDTQQLENTAEAEFKEELDVIKKELAAVLAIPDIHRNREALEKSIRILE
KAIDKWEEDEENQVSLELRRQSIGKFKEQRRSCKQKLRDAENAFHERREREYEERTMRE
IPRRYSSFRDTDITRRNNTTGLYHHSQQSSSNFRMQEYSSDEEIENIPSSHRDRYRLEKCLII
VFQNLYFSYPPKKISHSTMLQQRRDISPVVYRSQQQSSAGSQQERMRSGKSRVTKTT--
RKHRPG
>C_brenneri_HCP-3L5_N-tail
MMMMEDSPHIEEIVEVEE----
AEIEVKRQFEEYKKEIEEVARLPGFNKDSGKMNQVIRIMDKAIEKWEEDEENEGSLEFRR
RCLWEFKHRHNNYKKIIEKAEEDFYKRRE---EAMKMVSILNRTPQSHSYSQTR-
DGTAGSYR-TQATSSGPRHRAFESDEE--------RD----------
QRERMQVGRSTISKKTPKRKYRPG

C. brenneri **HCP-3/HCP-3L5 HFD alignment**
>C_brenneri_HCP-3_HFD
QKALAEIRKYQKSTDLLIQKAPFARLVHEIIREATTNSGDYRVRADALLALQEGAEAFMV
EMFEGSVLICNHAKRVTLMPTDIQLYRRLCLR
>C_brenneri_HCP-3L5_HFD
QKALAEIRKYQKSTDLLIQRAPFARLVHEIVREATASSGDYRVRADALMALQEGAEAFM
VEMFEGSVLICNHAKRVTLMPTDIQLYRRLCLR

***C. doughertyi* HCP-3/HCP-3L6 N-tail alignment**

>C_doughertyi_HCP-3_N-tail

---MYDDSDRPTIEEIDDTENSGYR-TAEAIFQERAEEVSKQIKTLLKNPTT---
DDLKQVIRIMSRSIDEWAEEEDNEGSIRARTTAIKTLTRKRNDYEMSLARKENEFLRKRG
QRHEEESMTRENPRRCSSFRDTDITRRTDRTGLNQSSYQGPSSNQQTHYYSSDEDYENRP
RSNRDPYRYNDSPQRSHQSSMYQQRRDASPPLHRSHTTLNGTSGSQQVRMRSGKSRVT
KTTARKYRPG

>C_doughertyi_HCP-3L6_N-tail

MMQMYDDG—
PTIEEILDNHHSGGPITAETIFKEDFDDVSKQIKTLTRDGSTKNADNLKQIISIMSRSIDKWA
EDEDMEGSIRMRKDAIEAFTKKRNEFREKITHAEDEYMRRKRQRMDEEALMREDQRGY
FSSRDNDMARHAGGTALDFNYQQGTSSNYRARYYSSDEDYENAPRSDRDRYRYPDSPQ
KSNQLATYHQRRDVSYSPSRSHQPMNGASSSTQVRMRAGKSRVTKKNSRKFRPG

***C. doughertyi* HCP-3/HCP-3L6 HFD alignment**

>C_doughertyi_HCP-3_HFD

QKALAEIRQYQKSTDLLIQKAPFARLVHEIIREESSQTDFRVRADALLALQEAAEAFMVE
MFEGSVLICNHAKRVTLMPADIQLYRRLCLR

>C_doughertyi_HCP-3L6_HFD

QKALAEIRQYQKSTDLLIQKAPFARLVHEIIREETAVDDFRVRADALLALQEAAEAFMVE
MFEGSVLICNHAKRVTLMPADVQLYRRLCLR

***C. sp54* HCP-3/HCP-3L3 N-tail alignment**

>C_sp54_HCP-3_N-tail

MEQIYDDMRGRIEEIVDPPSRNTTVFRDLTQDQDSDVDRIRNQMKLIMQKPNFNTSIHEM
QKVISILDDQINKWESEEELRDNLFEKMERAKADYYKRKEAEDDSRTREEPRRRANYTD
MDITDRDNATRLNHL----------
SYQRSYSHRNQIDNSDDDDIENMT----------
-----RSRRDRSRSPPSHSYYQSTMNQQRRDESYAHQRSHYSMSVPSSSHQVRMRSGKSR-
VTKTNSRKWRPG

>C_sp54_HCP-3L3_N-tail

MRQIYDDMRERIEEIVDPPSRNSTVFRDLTQDQDSDLERVRNQIKRIIEKPGSNTSIHELQK
VINILDDQIYKWESEEETRDNYLEKMERAKAEYYKRKEAQDDSRRREEPRRRANYTDM
DITDRYNATRLNHLSSYQRSYSHQNQIDNSDDDDIENMIPRSRRDRSWSPPSHYKRKVAQ
DDSRTREEPRRRANYTDMDITDRGNSTRLNHLSSYQRSYSHRNQIDNSDDDDIENMTRK
YNTIFKSPPSHGYHQSTMNQQRRDESYAHERSHYSKSAPSNSHQVRMSSYQRSYSHRNQ
SDNSDDDDIENMIPRSRCDRSRSPPSHGFHQSTMNQQRRDKSYAHQRSHYSMRAPSGSH
QERMRSSKSRRVTKKNYRKWRPG

***C. sp54* HCP-3/HCP-3L3 HFD alignment**

>C_sp54_HCP-3_HFD

HKALSEIRMYQKSTDLLIQKAPFARLVHEIIRDTTSNSQDYRVRADALLALQEAAEAFLV
EMFEGSVLICNHAKRVTLMPTDIQLYRRLCLR

>C_sp54_HCP-3L3_HFD

QKALYEIQKYQKSTDLLIPKAPFARIVHEMIHKATSTSHDLRVRANTFLPLQEAAEAFLVQ
MFHGSMKYCNSAKRVTLMQTDIQNYRSP---

***C. elegans* HCP-3/HCP-3L1 (CPAR-1) N-tail alignment**

>C_elegans_HCP-3_N-tail
MADDTPIIEEIAEQNESVTRIMQRLKHDMQRVTSVPGFNTSAAGVNDLIDILNQYKKELE
DDAANDYTEAHIHKIRLVTGKRNQYVLKLKQAEDEYHARKEQARRRASSMDFTVGRN
STNLVDYSHGRHHMPSYRRHDSSDEE-----NYSMDGTN-
GDGNRAGPSNPDRGNRTGPSSSDRVRMRAGRNRVTKTRRYRPG
>C_elegans_HCP-3L1_CPAR-1_N-tail
MADDGPIIEEIAEKNGRVARIMQRLQHDTQRVTSVPGFNTSATGYADLIALLDQYKNDLE
AVGFNDL----------
EQARRRAPSVDITVGSNSTNLVDYSHGRHDMPSHRRHDSSDEEITAANSHHQSPINVGNR
NDTDGTNGRNGSRAGSSSSDRVRMIAGRNRISKTRRYRPG

***C. elegans* HCP-3/HCP-3L1 (CPAR-1) HFD alignment**

>C_elegans_HCP-3_HFD
QKALEEIRKYQKTEDLLIQKAPFARLVREIMQTSTPFGADCRIRSDAISALQEAAEAFLVE
MFEGSSLISTHAKRVTLMTTDIQLYRRLCLR
>C_elegans_HCP-3L1_CPAR-1_HFD
QKALEEIRKYQESEDLLIPKAPFARLVREIMQTSTPFSSDLRIRSDAINALQEASEALLVQM
FDGSSLISAHSKRATLTTTDVQLYRRLCLP

***C. panamensis* HCP-3/HCP-3L10 N-tail alignment**

>C_panamensis_HCP-3_N-tail
MLQMEDLDGPRIEELPASPEREPAALRDNNRNGNVQRALPAHFENELRRLMSDPNFSKD
AELMTDAIELMKRQVNKMEDDQDMYGYEGGMAEMIRNLRIRIVSFTKKRDDAIARFRE
EREASRQMERTRRVNNTDFDMTDAENRTRLHPSMSSQRSYSYDRRDPYSSDENDDTYQ
PARHNAQQRMSRSPSPLMHQSHHSSINQRRDDSRQHRSYNNTATSSRPTTSNRSRMRVG
KNCVTKTKNRKWKPG
>C_panamensis_HCP-3L10_N-tail
MLQMEDLDGPRNEEMPASPEREPAA----------
EAQFENELRRLMNDPNFHRKADLMGKAIELMKRQVNRLEDDQDVY----------
DD---------DASRLVERTRRVNNTDLDMTDAENRTRLRSSVSSQ------
RRDPYSSDENNDTDQPARHNAQQRMSRSPSSLMHQGHGSTINQRRDDNRRQRSYNNTA
TSSRPTTSNQSRMRIGKNRVTKTKNRKWKPG

***C. panamensis* HCP-3/HCP-3L10 HFD alignment**

>C_panamensis_HCP-3_HFD
EKAMKEIRRYQKSTDLLIQKAPFVRLVHEIMADVTPRSSEYRIRAEALGALQEAAEAFLV
EMFEGSVLIANHAKRVTLMPTDIQLYRRLCLR
>C_panamensis_HCP-3L7_HFD
KKAMNEIRRYQKSNDLLIQKAPFVRLVHEIMADVTPRSSEYRIRAEALGALQEAAEAFLV
EMFEGSMLIANHAKRVTLTPTDIQLYRRLCLR

***C. afra* HCP-3/HCP-3L8 N-tail alignment**

>C_afra_HCP-3_N-tail

MLHQHVGPVITEMEEPASRDSSILRDSGRRHNVARSEPTAIERQIQDIFNQPRCNERPEAM
AEAIRLMRKQIDQWEREQDLYGPTEERTKNIRIWKNNRRKFIAQLEVAKERQERARRER
DESSRVVETVRRRGDMTETNVTAIHNSTRLQSSQRSYDQRTHFDSDEEE--
NGYTSRAPRPLPRSPPRAHQSYHSNLSHQSRRLDASSQDNRSRNDYGSNNVTSVTSSSH
QGKPKPRLRAGKSRVTKNMFRPR
>C_afra_HCP-3L8_N-tail
MNNSRNTRLKSDIVPGRRIIPDVIYRDAGIEEEPSYLAERTMDESVVNSTRNTRWNSNIVP
GRRIIPDVIYRDSGIRGEPSYLAERTMDESVKSKRIGLGEREKRAADPTRSTMSSMIEDISS
PGVQFFENNREKMRPTVATPRRSNLGANRLSTASDKEKTIDMLSIGGEPTIEANGSSYVEP
VSVNGSRVSPTREDMLHRHVGPVITEVEESPSRDPSVLRDSGRRHNVARSKSTAIERQIKL
IFIEPRFKERPEAMAEVIRLMRKQIDQWERDQDSYGPTEERTKNIRAWKDNRRRFIEQLE
DAKERQERARRERDESSRVVETVRRRGDMTGMNVTAIHNS-------
RSYDQRTHFDSEEEEKENGYTSRAPRPLPRSPPRAHQSYHSSLSHQSRRLDASSRDNRSR
NDYGSNNV---TSSSHQGKPKPRLRAGKSRVTKSMFRPR

***C. afra* HCP-3/HCP-3L8 HFD alignment**
>C_afra_HCP-3_HFD
RDRALLEIRQYQKSTNLLIQKAPFCRLVQEILREVTSSSDYRSSDYRIRADALSALQEAAE
AFLVEMFEGSQLIATHARRVTLMHSDIQLYRRLCLR
>C_afra_HCP-3L8_HFD
RDRALLEIRQYQKSTDLLIQKAPFCRLVQEILREEATTS---
SSDYRIRADALTALQEAAEAFLVEMFEGSQLIATHARRVTLMHSDIQLYRRLCLR

***C. sp49* HCP-3/HCP-3L9 N-tail alignment**
>C_sp49_HCP-3_N-tail
MEMLLILILALLPPSSSTCLFSVLIISSSSSFPSNCTRLEGRFRVDHSSDLADSQLVELFANV
RELHGVVQIWNSRLTTANFLANIRRITGDLLDNSISIVNNTLLTSINLTSLEYSDGKVEIQN
NPLLDLKPNCPALHKSFFNRRSISSLECGCQVTGTFSYPKNINFPENCVVLYGNFIINDVAP
PFELLYRLLSVRKLYGFLEVRNTNLETLGFLQNLEEIESGPDDEITVSIGSEMTWLGRVDLI
KLNSIKSRNPRKSVISISAKCLDPALVQLVAHSDIQDIKPEMQRMDGPIIEEVVDHDARQE
KRNRRINRIKDLMARPETKESKELLKKMLDLLQEHLNDLEDEQLEEGANHRDAIATLRH
KIDIFTPLYNGAVADYMARKEESRREEARRAMSFGSSQNNITGRDNRSKLHQSSQRTYSS
EDEENDEEYAGRNRQDARRYHHQSPEPQASSSARRDNTRRVDASSNNPRMRAGKSRVT
KTNSRRWRPG
>C_sp49_HCP-3L9_N-tail

MARPETKESKELLKKMLDLLQEHLNDLEDEQLNEGANHRDAIATLRHKIDIFTPLYNAA
VADYMARKEESRREEARRAMSFGSSQNNITGRDNRSKLHQSSQRTYSSEDEENDEEYAG
RNRQDARRYHHQSPEPQASSSARRDHTRRVENSSNNPRMRAGKSRVTKTNSRRWRPG

***C. sp49* HCP-3/HCP-3L9 HFD alignment**
>C_sp49_HCP-3_HFD
QKALSEIRKYQKSTDMLIQKAPFHRVVQEILCETSGFTNAHRIRADAISALQEAAEAFLVE
MFEGAMLLSNHAKRVTLMASDIQLYRRLCLR

>C_sp49_HCP-L9_HFD
QKALSEIRKYQKSTDMLIQKAPFHRVVQEILCETSGFTNAHRIRADAISALQEAAEAFLVE
MFEGAMLLSNHAKRVTLMASDIQLYRRLCLR

Supplementary Data 3.S2: HCP-3 histone-fold domain codon based-alignments used for Figure 2

>C_tribulationis_hcp3_HFD
GGGCAGAAAGCGTTAAGAGAAATTCGCAAGTATCAAAAGTCTACGGATATGCTTATTC
AGAAAGCACCCTTCGCTCGTCTCGTCCACGAAATCATACAGGAAACGACTTTGTTTA
GT----------
CATGACTTTCGTATTCGTGCCGACGCTCTGATGGCTCTTCAAGAAGCATCTGAAGCGT
TTATGGTGGAAATGTTCGAAGGATCCTTCTTGATCTGCAATCACGCCAAACGCGTCAC
CCTTATGCCGACTGATATTCAGTTGTACCGTCGTTTGTGTCTTCGA
>C_sp41_hcp3_HFD
GGACAAAAAGCACTATCTGAAATCCGCAAGTATCAAAAATCTACGGATATGCTTATCC
AGAAGGCGCCGTTCGTCCGCCTCGTCAACGAAATCATCCAGGAAGCGACGTCGTTCA
GT----------
AAAGAATTTCGTATTCGAGCCGACGCTTTGATGGCTCTACAAGAAGCCTCAGAAGCT
TTTATGGTGGAAATGTTCGAAGGATCCGTGTTGATCTGTAATCACGCTAAACGTGTCA
CCCTTATGCCAACTGATATTCAGTTATACCGTCGCTTGTGTCTCCGA
>C_zanzibari_hcp3_HFD
GGACAGAAAGCGCTCTCCGAAATCCGAAAATATCAGAAGTCTACGGATATGCTCATTC
ACAAAGCTCCATTTGCCCGTCTTGTTCACGAAATCATACAGGGATCGTCATTGGAGAG
T----------
AAAGACTTTCGTATTCGTGCGGACGCTTTGATGGCTCTTCAGGAAGCCGCGGAAGCG
TTTATGGTGGAAATGTTTGAAGGGTCCGCGTTGATCTGTAATCACGCTAAACGTGTCA
CCCTTATGCCGACCGATATTCAGCTATACCGTCGTTTGTGTCTACGA
>C_sinica_hcp3_HFD
GGACAGAAAGCTCTTGCTGAAATCCGGAAGTATCAGAAATCTACGGATATGCTCATTC
AAAAAGCTCCATTCGCCCGTCTCGTTCACGAGATCATCCAAGAAGCAACTTCGTTTAG
C----------
AAGGAGTATCGTATTCGTGCTGATGCCTTGATGGCCCTTCAAGAAGCGTCGGAAGCCT
TCATGGTGGAAATGTTTGAGGGATCTGTATTGATTTGTAATCACGCGAAACGAGTCAC
CCTAATGCCCACCGACATTCAGTTATATCGTCGTTTGTGCCTCCGA
>C_nigoni_hcp3_HFD
GGACAGAAAGCCTTAGCTGAAATTCGAAAGTATCAGAAGTCGACAGATATGCTGATC
CAGAAGGCTCCTTTTGCTCGTCTTGTTCATGAAATTGTTCGAGAACAAACCAACCAA
AGT----------
AAAGACTATCGTATTCGTGCCGATGCTTTGATGGCTCTACAGGAAGCAGCAGAAGCAT
TCATGGTTGAAATGTTCGAAGGATCCGTTCTGATTTGCAATCACGCTAAGCGTGTCAC
ACTCATGCCCACTGACATTCAGCTGTATCGTCGCTTGTGCCTCCGA
>C_briggsae_hcp3_HFD
GGACAGAAAGCCTTGGCTGAAATTCGAAAGTATCAGAAGTCGACAGATATGTTGATC
CAGAAGGCTCCTTTTGTTCGTCTTGTTCATGAAATTATTCGAGAACAAACCTACAAAA
GT----------
CAAGACTATCGTATTCGTGCGGATGCTTTGATGGCTCTACAGGAAGCAGCAGAAGCAT
TCATGGTTGAAATGTTCGAAGGATCCGTACTGATTTGCAATCACGCTAAGCGTGTCAC
ACTCATGCCCACTGACATTCAGCTGTATCGTCGCTTGTGCCTTCGA
>C_remanei_hcp3_HFD

GGACAGAGAGCGCTTGAGGAAATTCGAAAATACCAAAAGTCCACCGATATGCTGATT
CAGAAAGCTCCCTTTGCACGTCTTGTCCACGAAATTATGCGCGAAGCAACTTCGGAA
AGT----------
CAAGATTTTCGGATTCGTGCAGACGCTTTGATGGCTCTTCAAGAAGCGGCAGAAGCG
TTCATGGTGGAGATGTTCGAGGGATCCGTGTTGATTTGTAATCACGCGAAAAGAGTAA
CTCTCATGCCGACAGATATTCAATTATATCGTCGCTTATGTCTTCGG
>C_latens_hcp3_HFD
GGACAGAGAGCGCTCGAGGAAATTCGAAAATACCAAAAGTCCACCGATATGCTGATT
CAGAAAGctccgtttgctcgtCTTGTCCACGAAATTATGCGCGAAGCAACTTCGGAAAGT-------
-----
CATGACTTTCGGATTCGAGCAGATGCTTTGATGGCTCTTCAAGAAGCGGCAGAAGCG
TTCATGGTGGAGATGTTCGAGGGATCCGTGCTGATTTGTAATCACGCGAAAAGAGTAA
CTCTCATGCCGACAGATATTCAATTATATCGCCGCTTATGTCTTCGG
>C_sp51_hcp3_HFD
GGACAAAAAGCACTGGCTGAAATCAGACAATATCAGAAATCAACGGATTTTTTGATT
CAAAAAGCACCGTTTGCACGGTTGGTCCATGAAATAGTTCGTGAGGCTACTTCAAGC
AGT----------
GGTGATTTCCGAGTTCGCGCGGATGCTCTCATGGCCCTTCAAGAAGGTGCCGAAGCG
TTTATAGTAGAAATGTTTGAAGGATCTGTATTAATCTGCAATCATGCTAAGCGTGTCAC
CCTCATGCCAACTGATATCCAATTATACCGCCGATTGTGCCTCCGA
>C_sp44_hcp3_HFD
GGACATAAGGCCTTGGCAGAGATTCGACATTATCAGAAGACAACTGATCTTCTCATTC
AAAAGGCCCCATTCGCTCGTCTCGTCCATGAAATCATCCGCGAAATGACCCCTAACAA
T----------
GCCGACTATCGAGTCCGTGCCGACGCTCTTCTCGCCCTCCAAGAAGGCGCAGAAGCT
TTTATGGTAGAAATGTTCGAAGGATCCGCTCTGATTTGTAATCACGCAAAGCGTGTCA
CCCTTATGCCAGCGGATATTCAACTATATCGTCGATTGTGCCTTCGA
>C_sp48_hcp3_HFD
GGACAAAAAGCGTTGGCAGAGATCAGGCAATACCAGAAGTCAACTGATCTTTTGATT
CAAAAAGCTCCATTTGCACGACTTGTCCATGAAATTATTCGGGAAGCAACTTCAAAC
AGT----------
GGGGATTATCGCGTTCGCGCAGATGCTCTTCTAGCTCTCCAAGAAGGGGCAGAAGCG
TTTATGGTGGAAATGTTTGAAGGATCTGTATTAATTTGTAATCACGCAAAGCGTGTAAC
TCTTATGCCCACCGATATCCAATTATATCGACGTCTGTGCCTCCGA
>C_brenneri_hcp3_HFD
GGGCAAAAAGCGTTGGCAGAGATAAGGAAATACCAGAAGTCAACTGATCTTTTGATT
CAGAAAGCTCCATTTGCACGCCTTGTCCATGAAATTATCCGGGAAGCAACTACAAATA
GT----------
GGAGATTATCGCGTTCGTGCAGATGCTCTTCTAGCTCTCCAAGAAGGCGCTGAAGCAT
TTATGGTTGAAATGTTTGAAGGATCTGTATTAATTTGTAACCACGCGAAGCGCGTAACT
CTTATGCCCACAGATATTCAATTATATCGACGTCTGTGCCTCAGA
>C_wallacei_hcp3_HFD
GGACAGAAAGCGTTGTCCGAAATCCGACAATATCAAAAATCAACGGATCTTCTCATTC
AAAAAGCTCCGTTCGCACGTCTCGTTCATGAAATCATTCGTGAAGAAACCAAT---------
----
AAAGACTATCGAATCCGCGCTGATGCTATTCATGCCTTGCAGGAAGCGGCTGAAGCAT

TTATGGTTGAAATGTTCGAAGGATCTACATTAATTTCCAATCACGCAAAACGTGTCAC
CCTTATGCCAACTGATATCCAATTATATCGTCGCTTGTGTCTTCGA
>C_tropicalis_hcp3_HFD
GGACAAAGAGCTTTGGCCGAAATTCGGCAGTATCAAAAATCAACGGATCTTTTGATTC
AAAAAGCCCCCTTCGCACGTCTCGTCCATGAAATCGTTCGTGAAGAAACCAAC---------
----
CAAGATTATCGAGTCCGCGCTGATGCTATTCTTGCACTACAAGAAGCTGCCGAAGCGT
TTATGGTTGAAATGTTCGAAGGATCTACATTAATTTCCAATCACGCAAAGCGTGTTACT
TTGATGCCCACTGATATTCAACTATATCGGCGGTTGTGCCTCAGA
>C_doughertyi_hcp3_HFD
GGACAGAAAGCGTTAGCTGAAATTCGACAGTATCAAAAATCAACAGATCTTCTCATTC
AGAAAGCCCCGTTTGCACGTCTTGTTCATGAAATTATTCGCGAAGAATCATCACAG----

ACTGATTTTCGCGTTCGCGCAGATGCTCTTCTCGCCCTTCAGGAGGCTGCCGAAGCAT
TCATGGTGGAAATGTTCGAAGGGTCTGTGCTTATCTGCAACCACGCGAAACGTGTCA
CTCTCATGCCAGCAGATATTCAACTATATCGGCGGTTGTGTCTTCGA
>C_sp54_hcp3_HFD
GGACACAAAGCGTTGTCCGAAATTCGAATGTATCAAAAGTCTACTGATCTGCTTATTC
AGAAAGCTCCATTCGCTCGTCTTGTCCATGAAATCATTCGCGATACAACCTCCAACAG
T----------
CAAGATTATCGAGTCCGCGCGGATGCTCTTCTCGCCCTCCAAGAAGCAGCTGAAGCA
TTCTTAGTCGAAATGTTCGAGGGTTCCGTGCTGATTTGCAATCATGCCAAGCGTGTCA
CCCTTATGCCAACTGATATTCAACTGTATCGTCGCCTGTGCCTTCGA
>C_inopinata_hcp3_HFD
GGCCAAAAAGCGATGGCTGAGATTCGTAAATATCAAAAATCAACGGATCTACTTATTC
AAAAAGCTCCATTTGCTCGTCTCGTGCATGAAATTGTTCGTGAGGCAACTTCTCATAG
T----------
AATGATTTTCGAGTTCGTGCAGATGCCATTATGGCCCTCCAAGAAGCTGCCGAAGCTT
TTATAGTGAACATGTTTGAAGGATCTGCAATGATCTGCCATCATGCGAAACGTGTCAC
TCTTATGGAAACTGATATTCTGCTGTATCGTCGTTTGTGTCTTCCA
>C_elegans_hcp3_HFD
GGCCAGAAGGCATTGGAAGAGATCCGCAAGTACCAAAAAACTGAAGACCTTCTGATT
CAAAAGGCTCCGTTCGCACGCCTCGTCCGCGAAATTATGCAGACTTCCACTCCATTTG
GC----------
GCCGACTGCCGTATTCGTTCTGACGCCATCAGTGCTCTTCAAGAAGCGGCGGAAGCA
TTTTTGGTCGAAATGTTCGAAGGATCGTCTCTTATATCCACCCATGCGAAACGTGTCA
CACTCATGACAACGGATATTCAGTTATACAGACGTCTCTGCCTTCGA
>C_oiwi_hcp3_HFD
GGACAGAAAGCCTTAGCAGAAATCAGAAAATACCAAAGATCGACGGATATGCTCATT
CAAAAGGCTCCGTTTGCCCGCCTTGTACATGAAATCATCCGCCAAGAAACGACCCAG
GAT----------AGT---
TTCCGGATCAGAGCAGACGCTCTATGTGCTCTGCAAGAAGCTGCTGAAGCATTCCTTG
TCGAAATGTTCGAAGGATCCGTGTTGATCTGTAATCACGCCAAACGAGTCACTCTTAT
GCCCACGGATATTCAATTATATCGCCGACTATGTCTCCGG
>C_kamaaina_hcp3_HFD

GGACACAAAGCCTTGGCAGAAATCAGAAAGTACCAAAGATCGACGGATATGCTCATT
CAAAAGGCTCCGTTCGCCCGCCTCGTACACGAAATTGTTCGTCAAGAAACAACCAAA
GAT----------TGC---
TTCCGGATCAGAGCAGACGCTCTATGTGCTCTGCAAGAAGCTGCCGAAGCATTCCTT
GTTGAAATGTTCGAAGGATCCGTATTGATCTGTAATCACGCCAAAAGAGTCACTCTCA
TGCCCACGGATATTCAATTATATCGCCGTCTATGTCTCCGG
>C_waitukubuli_hcp3_HFD
GGAGAGAAAGCAATGCGCGAGATTCGGCAGTATCAGAAGTCCACTGATCTGCTCATC
CAAAAAGCTCCATTCTGCCGTCTGGTCCACGAAATCATGCAAGAAGTCACTTCGTTC
AGC----------
TCAGATTTCCGTATCCGAGCCGAAGCTCTCGGTGCCCTTCAAGAAGCGGCTGAAGCT
TTCCTAGTCGAAATGTTCGAAGGCTCCGTTCTGCTAGCCAATCACGCGAAACGTGTCA
CTCTGATGCCAACCGATATTCAACTATATCGCCGTCTCTGTCTTCGC
>C_panamensis_hcp3_HFD
GGAGAGAAGGCGATGAAGGAGATTCGGAGATACCAGAAGTCAACTGATCTGCTAATC
CAAAAAGCTCCATTCGTGAGACTCGTCCACGAAATCATGGCCGATGTGACGCCGCGC
AGC----------
TCTGAATACCGGATTCGTGCAGAGGCACTCGGCGCTCTCCAAGAAGCAGCCGAAGCA
TTCCTAGTCGAGATGTTCGAAGGATCAGTGCTCATCGCCAATCACGCCAAACGAGTC
ACCCTGATGCCCACGGATATCCAGCTGTACCGTCGTCTTTGTCTCCGT
>C_nouraguensis_hcp3_HFD
GGAGAGAAGGCTATGCGTGAGATCAGGCAGTACCAAAAGTCCACTGACATGCTCATC
CAAAAGGCACCGTTCTGTCGTCTGGTTCATGAAATCGTGCAAGACGTCACCTCATCC
AGT----------
TCCGGTTTCCGCATACGCGCCGAGGCTCTCGGTGCTCTGCAAGAAGCCGCTGAAGCA
TTTCTTGTCGAGATGTTCGAAGGATCAGTGCTCATCGCTAACCATGCCAAGCGAGTAA
CTCTCATGCCTACTGATATCCAATTGTACCGCCGTCTGTGCCTTCGT
>C_becei_hcp3_HFD
GGAGAGAAGGCGATGCGCGAGATCAGGCAGTACCAAAAGTCCACTGATATGCTCATC
CAAAAGGCACCGTTCTGTCGGCTGGTTCATGAAATCATGCAAGACGTCACCTCATCC
AGC----------
TCCGATTTCCGCATTCGCGCAGAGGCGCTTGGTGCTCTGCAAGAAGCCGCCGAAGCA
TTCCTTGTCGAGATGTTCGAAGGATCAGTCCTCATCGCCAACCATGCAAAGCGAGTAA
CTCTCATGCCCACTGATATCCAACTGTACCGCCGTCTGTGCCTTCGT
>C_yunquensis_hcp3_HFD
GGAGAGAAGGCGATGCGGGAGATCAGACAGTACCAGCAGTCCACCGATATGCTCATC
AAAAAAGCTCCATTCTGTCGACTGGTTCATGAAATTATGCAAGAAGTCACTGGGTTCA
GC----------
TCCGATTTCCGGATTCGGGCAGAAGCTCTCGCCGCCCTTCAAGAAGCCGCGGAAGCG
TTCCTTGTGGAAATGTTCGAAGGATCTGTCCTGATCGCGAGCCATGCCAAACGTGTCA
CTCTCATGACTTCTGATATCCGCCTGTACCGCCGTCTCTGCCTCCGT
>C_macrosperma_hcp3_HFD
GGAGAGAAAGCGATGCGAGAGATCAGGCAGTACCAGAAGTCCACCGACATGCTCAT
TCAAAAAGCTCCATTCTGTCGACTTGTCCACGAGATTATGCAAGATGTCACCTCCTTT
AGT----------
TCAGACTTCCGTATCCGAGCAGAGGCACTTGGTGCCCTTCAAGAAGCCGCAGAAGCG

TTCCTTGTCGAAATGTTCGAGGGTTCCGTTTTGATTGCGAACCACGCCAAGCGAGTCA
CTTTAATGCCAACAGATATACAACTCTATCGTCGTCTTTGCCTCCGT
>C_sulstoni_hcp3_HFD

GACCGCGCGCTTCAGGAGATTCGGCAGTACCAGAAGTCGACCGACCTGCTCATTCAA
AAAGCTCCATTCTGCCGTCTCGTTCAAGAAATCGTCCGGGAGTCGAGTTCGTCCACT--

TCAGATTTCCGTGTCAGAGCCGATGCTCTCTCGGCCCTTCAGGAAGCCGCCGAAGCG
TTCCTCGTCGAGATGTTCGAGGGCTCCCAATTGATTGCCGCGCACGCGAAACGTGTTA
CTCTGATGCCCTCCGATATCCAACTCTACCGTCGCTTGTGCCTGAGA
>C_afra_hcp3_HFD

GACCGCGCACTTCTGGAGATCCGACAGTACCAGAAGTCTACCAACCTGCTAATTCAG
AAGGCCCCATTCTGTCGTCTCGTTCAAGAAATCCTACGCGAGGTTACAAGTTCAAGC
GATTACCGTTCAAGCGATTACCGCATCAGAGCAGATGCTCTCTCCGCCCTTCAAGAGG
CCGCGGAAGCCTTCCTCGTTGAGATGTTCGAGGGCTCCCAGCTGATTGCCACTCACG
CAAGACGTGTCACTCTTATGCACAGCGATATTCAACTGTACCGCCGTCTCTGCCTTCG
G
>C_sp49_hcp3_HFD
GGACAGAAGGCGTTGAGCGAGATCAGAAAGTACCAGAAATCGACGGATATGTTGATT
CAAAAGGCCCCGTTCCACCGCGTCGTCCAGGAGATCCTCTGCGAGACGTCCGGCTTC
ACC----------
AACGCCCACCGCATCCGAGCCGACGCGATCTCCGCGCTCCAGGAGGCCGCCGAGGC
GTTCCTCGTCGAGATGTTCGAGGGCGCCATGCTGCTTTCGAATCACGCGAAACGGGT
CACTCTGATGGCGTCCGACATCCAGCTGTACCGTCGACTGTGCCTGCGG
>C_sp25_hcp3_HFD
GGACAGAAGGCGTTGGCCGAAATCAGAAAGTATCAGAAGACGAGTGATCTTTTGATA
CAAAAGGCGCCGTTCTACCGTGTCGTTCAAGAGATCCTTCGCGAAACGTCAGGCTTC
ACC----------
AACGACCACCGCATCCGAGCGGACGCGATCGCCGCTCTTCAGGAGGCTGCCGAGGC
GTTCCTCGTCGAGATGTTCGAGGGCTCAGCCCTTCTGTCATTGCACGCGAAGCGGGT
GACTCTTATGCCGTCGGACATTCAACTGTACCGCAGACTGTGTCTCAGA
>C_imperialis_hcp3_HFD
GGGCAGAAGGCGTTGAGCGAGATTAGAAAGTATCAGAAGTCGACTGATCTTCTTATT
CAAAAGGCTCCGTTCTACCGTGTGGTGCAAGAGATCCTCCGCGAGACGTCTGGCTTC
ACC----------
AACGACCACCGGATCCGTGCTGATGCCATCGCCGCACTTCAAGAAGCGGCCGAAGCG
TTCATCGTCGAGATGTTCGAGGGAGCGACGCTTCTGTCGACGCACGCGAAGCGAGTC
ACTCTGATGCCCTCCGACATCCAACTATACCGCAGATTGTGTCTCAGA
>C_japonica_hcp3_HFD
GGACAGAAGGCGTTGAGTGAGATTCGAAAATACCAAAATTCCACTGATTTGCTCATTC
AAAAAGCCCCCTTCCGTCGATTAGTTCACCAGATTATTCAAGAAGCGACCGGCTTCG
AT----------
TCCGGATTCCGCATTCGCGCCGACGCGATGTCTGCCCTACAAGAAGCCGCCGAGGCG
TTCATCGTCGAGATGTTCGAGGGATCTGTTCTCATCTCGAATCACGCAAAACGGGTCA
CTCTGATGACGGCCGACATTCAATTGTACCGTCGACTTTGCCTCCGA

>C_elegans_hcp3L1_cpar-1_HFD
GGCCAGAAGGCATTGGAAGAGATTCGCAAATACCAAGAATCTGAAGACCTTCTGATT
CCAAAGGCTCCGTTCGCACGCCTCGTCCGCGAAATTATGCAGACTTCCACTCCATTTT
CC----------
TCAGACCTTCGCATTCGTTCTGACGCCATCAATGCTCTTCAGGAAGCATCGGAAGCAC
TGTTGGTCCAAATGTTCGATGGTTCATCACTCATTTCCGCCCATTCCAAACGTGCCAC
ACTGACGACTACAGACGTGCAGCTTTACCGCCGTCTTTGCCTTCCA
>C_tribulationis_hcp3L2_HFD
GGACAAAAAGCGTTGGCTGAGATTCGCCGATACCAAAAATCTACGGATATGCTCATTC
AGAAGGCTCCGTTTGCTCGTCTCGTACATGAAATCATCCAAGAATCAACAACCTTGAG
T----------
CGCGACTTCCGTATTCGTTCTGATGCGTTAATGGCTCTCCAGGAGGGTGCCGAAGCAT
TCATGGTAGAGATGTTCGAAGGATCGGCTCTCATCTGCAATCACGCTAAGCGCGTTAC
TCTCATGCCAACTGATGTGCAGCTGTACCGTCGCCTGTGCCTTCGC
>C_sp41_hcp3L2_HFD
GGACAAAAAGCGTTGGCTGAGATTCGCCAATATCAGAAATCTACGGATATGTTGATCC
AGAAAGCTCCATTCGCTCGCCTTGTTCATGAAATCATCCAGGACTCCACAAACTTCAG
C----------
CGAGATTTCCGTATTCGTGCAGATGCCTTGATGGCTCTCCAGGAAGCGGCCGAAGCGT
TCATGGTTGAGATGTTCGAAGGATCGACTCTCATCTGTAATCACGCAAAGCGCGTCAC
TCTCATGCCCACGGATCTCCAGCTGTACCGTCGCCTGTGCCTTCGA
>C_sinica_hcp3L2_HFD
GGGCAAAAAGCATTGGCTGAAATTCGCCAATACCAAAGGACAACAGAAATGTTGATA
CAAAAAGCTCCCTTCGCTCGTCTCGTCCATGAAATCATCCAGGACGCCACTTCCTTCA
GC----------
CGTGATTTCCGTATCCGTGCTGATGCTCTAATGGCTCTCCAGGAGGCAGCCGAAGCGT
TTATGGTGGAAATGTTTGAGGGATCTGTGCTCATCTGTAACCACGCAAAACGTGTTAC
GCTCATGCCGACCGACCTTCAGCTCTACCGTCGCCTCTGCCTTCGC
>C_sp54_hcp3L3_HFD
GGACAAAAAGCGTTGTACGAGATTCAAAAGTATCAAAAGTCTACTGATCTGCTTATTC
CGAAAGCTCCATTCGCTCGTATTGTTCATGAAATGATTCACAAAGCAACCTCAACCAG
T----------
CACGATTTACGAGTCCGCGCCAATACTTTTCTCCCCCTCCAAGAAGCAGCGGAAGCAT
TCCTAGTCCAAATGTTCCACGGTTCCATGAAGTATTGCAATTCTGCCAAGCGTGTCAC
CCTTATGCAAACTGATATTCAAAATTATCGTAGCCCGTGA------
>C_remanei_hcp3L4_HFD
GGACAGAAGGCTCTTCTCGAAATCCGAAAATACCAAAAGTCCACCGACATGCTCATC
CAAAAAGCTCCATTCGCTCGTCTCGTCCAAGAAATCCTTCGAGAAACCACAAATGAG
AGT----------
CACGACTACCGTATCCGTGCAGACGCTCTCATGGCATTACAAGAGGGTGCCGAGGCG
TTTATGGTGGAAATGTTCGAGGGATCCGTGTTGATTTCGAACCATGCGAAACGGGTCA
CGTTGATGCCCACTGATGTTCAATTGTATAGACGTTTGTGTCTCAGA
>C_latens_hcp3L4_HFD
GGACAAAAGGCTCTTCTCGAAATTCGAAAATACCAAAAGTCCACTGACATGCTCATC
CAAAAAGCCCCATTCGCTCGTCTCGTCCAGGAAATCCTTCGGGAAACCACAAATGAG
AGT----------

CACGACTACCGTATCCGTGCAGACGCTCTCATGGCATTACAAGAGGGTGCCGAGGCG
TTTATGGTCGAGATGTTCGAGGGATCTGTATTGATTTCGAATCATGCGAAACGGGTCA
CCTTGATGCCCACTGATATTCAGCTGTATAGACGTTTGTGTCTCAGA
>C_sp48_hcp3L5_HFD
GGTGAGAAAGCCTTGGCCGAAATCAGACAGTACCAAAGGTCAACGGATCTTCTGATT
CAGAAAGCTCCATTCGCCCGTCTTGTCCATGAAATCGTGAGCGAAGCAACTTCTTCG
AGT----------
GGAGACTACAGGATTCGCGCCGATGCTCTCATGGCTCTGCAAGAAGGCGCCGAAGCA
TTCATGGTGGAAATGTTTGAGGGATCGGCGCTCATCTGCAACCACGCCAAACGAGTG
ACACTAATGGCATCGGACGTTCAGCTTTACAGGCGCCTTTGCCTGAGA
>C_brenneri_hcp3L5_HFD
GGACAGAAGGCCCTTGCGGAAATCCGCAAGTACCAAAAGTCCACGGATCTTCTGATT
CAGAGAGCACCATTCGCACGTCTTGTACATGAAATTGTCCGTGAAGCAACTGCTTCA
AGT----------
GGCGACTACCGGGTTCGCGCAGATGCTCTCATGGCTCTGCAAGAAGGAGCTGAGGCC
TTTATGGTGGAAATGTTTGAAGGATCGGTGCTCATTTGCAACCATGCCAAGCGTGTCA
CTCTCATGCCAACGGACATTCAGCTCTACCGCCGACTCTGCCTTCGT
>C_doughertyi_hcp3L6_HFD
GGACAAAAGGCATTGGCTGAAATCCGACAGTATCAAAAATCAACAGATCTTCTCATTC
AGAAGGCTCCATTCGCACGTCTCGTTCATGAAATCATTCGCGAAGAAACTGCAGTC---

GATGATTTCCGTGTCCGCGCAGATGCTCTTCTCGCCCTTCAGGAAGCTGCTGAAGCAT
TCATGGTGGAAATGTTCGAAGGATCTGTGCTTATCTGCAATCACGCAAAACGTGTCAC
TCTCATGCCAGCTGATGTTCAATTGTACCGACGTTTGTGCCTCCGT
>C_sp48_hcp3L7_HFD
GGCCAAAGAGCAATTGCGGAGATCAAACACTACCAGAAGACCACCGAACTTTTGATC
CAAAAAGCGCCATTTGCTCGACTCGTCCAAGAAGTGGTTCAAGAAGCCACCTCAGA
AAGT----------
AGCAGTTACAGCATCCGCACTGATGCACTATCAGCGCTCCAGGAGGGTGCCGAAGCG
TTTTTGGTGGAGATGTTCGAAGGCTCAAGCATGATTGCTAATCACGCCAAAAGAGCA
ACGTTGGGGTCGACCGATCTCAAATTGTATCGTCGCCTTTGCCTTAGG
>C_afra_hcp3L8_HFD

GATCGCGCACTTCTGGAGATCCGACAGTATCAGAAGTCTACCGACCTGCTAATTCAGA
AGGCCCCGTTCTGTCGTCTCGTTCAAGAGATCCTACGCGAGGAGGCTACTACAAGT---

TCAAGCGACTACCGCATCAGAGCGGATGCTCTCACCGCCCTTCAAGAGGCTGCTGAG
GCCTTCCTCGTTGAGATGTTCGAGGGCTCCCAGCTGATTGCCACTCACGCAAGACGT
GTCACTCTTATGCACAGTGATATTCAACTCTACCGCCGTCTTTGCCTTCGG
>C_sp49_hcp3L9_HFD
GGACAGAAGGCGTTGAGCGAGATCAGAAAGTACCAGAAGTCGACGGATATGTTGATT
CAAAAGGCCCCGTTCCACCGCGTCGTCCAGGAGATCCTCTGCGAGACCTCCGGCTTC
ACC----------
AACGCCCATCGCATCCGAGCCGACGCGATCTCCGCGCTCCAGGAGGCCGCCGAGGCG
TTTCTCGTCGAGATGTTCGAGGGCGCGATGTTGCTGTCGAACCATGCGAAACGGGTC
ACTCTGATGGCGTCCGACATCCAGCTGTACCGACGACTGTGCCTGCGG

>C_panamensis_hcp3L10_HFD
GGAAAGAAGGCGATGAATGAGATTCGGAGATACCAGAAGTCAAATGATCTGCTAATC
CAAAAAGCTCCATTCGTGAGACTCGTCCACGAAATCATGGCCGATGTGACGCCGCGC
AGC----------
TCTGAATACCGGATTCGTGCAGAGGCACTTGGCGCTCTCCAAGAAGCAGCCGAAGCA
TTCTTAGTGGAGATGTTCGAAGGATCGATGCTCATCGCCAATCACGCTAAACGAGTCA
CCCTGACGCCCACGGATATCCAGCTGTACCGTCGTCTTTGCCTCCGT

Supplementary Data 3.S3: All HCP-3 sequences used in this study

>C_tribulationis_HCP3
MYAHTGPIIEEVEEAATDGQTVHRDWSKDPDVLRIRRELQKLMILPGFNNNADLMQRAI
TILVEQVDEWKLDQDIDGWNICRQEKIESFLPRIANFKNKRAQAIDQFYKERDSMINESR
RRERARELHHSNDFGISGRELDHSSRLHQLSRRDSCASRVERYHSSDEDEENHPVPRYRS
RSPGPSSSYNQSTMRQRRDDVPQPVRMRSGKSRVTKTRNAKWRPGQKALREIRKYQKS
TDMLIQKAPFARLVHEIIQETTLFSHDFRIRADALMALQEASEAFMVEMFEGSFLICNHA
KRVTLMPTDIQLYRRLCLRNLS
>C_tribulationis_HCP3L2
MEQMYENQHTPIIEELFDCSSVVEREVERVKHEIQALTSQSDFNKNYASMKEVINILTRQI
AAWDADEDMGGSHPIRLRSIEKFHAKRVLFTEKLEAAERAYYEKKQARLEEEKRRECLQ
DRGKIAGQNNQLCHRQGHRYERDDSSDDSSDEENQRQRSRACSPQRRNHPSTSSQQYR
VNRHVDYHKKHQNVSQKQQRLRAGINAVTKTKVRKFRPGQKALAEIRRYQKSTDMLIQ
KAPFARLVHEIIQESTTLSRDFRIRSDALMALQEGAEAFMVEMFEGSALICNHAKRVTLM
PTDVQLYRRLCLRNM
>C_sp41_HCP3
MYAHTGPIIEEVEDGPAEGHTVHRDWRQDTDVQRLGREIQKFISLPDFSKNADLMQRAI
HVLEKQVDEWKLNQELDGWDHHRQQKIELFQPKIAKFKEKREEAINHYYDVKDSMMN
ESRRRERAREVLHNTDQNITGFGNSTRLYPNSRRQSFAPRKERYQSSDEDEENDPVPRRR
SRSPGPSSSYSHSAMYQLRDDSNAPQQQRMRSGKSRVTKTKNRKWRPGQKALSEIRKY
QKSTDMLIQKAPFVRLVNEIIQEATSFSKEFRIRADALMALQEASEAFMVEMFEGSVLICN
HAKRVTLMPTDIQLYRRLCLRNMS
>C_sp41_HCP3L2
MDQLYENHHTPVIEEILDTEFVVEREVERVKHDIQLITSQPDFNKNYDSMKEVIDILARQI
LDWEADEEMSGSHPMRRRIIDKFQAKKVLFTEKLEAAERAYYERKRARLEEEKKRRQL
QDCGNIGGEQNRHARFQFMGHRNGRDDSSDDSSDGENRQVQRRRSRSRSPQHRNHSSQ
QHRRDDRHADGHRNYQSTSKNQPRLRAGINGVTKTKVRKYRPGQKALAEIRQYQKST
DMLIQKAPFARLVHEIIQDSTNFSRDFRIRADALMALQEAAEAFMVEMFEGSTLICNHAK
RVTLMPTDLQLYRRLCLRNL
>C_zanzibari_HCP3
MYGHTGPVIQEVLENPNEGGQVVQRADQQNWRHDHDVLRLGREAQKLMSNPDFNTDP
DLMEEAISLFENKVAEWRADQGFEGWDHRRQEKIEFFQQRVVKFKEQRKRACDRYYDE
RDSKIENSRRERVRDVHQYTDNNISNFGNSTRLNQHSHRQSYTTREDRYHTSEDEENIPV
QHHRTRAPSSYSNYTQSTMQQRRDDSNVHYRSHQSAAGPSSSQQVRMRSGKSRVTKTR
SRKWRPGQKALSEIRKYQKSTDMLIHKAPFARLVHEIIQGSSLESKDFRIRADALMALQE
AAEAFMVEMFEGSALICNHAKRVTLMPTDIQLYRRLCLRNL
>C_sinica_HCP3
MYGHTGPLIQEIDETPTEGSTAVRDWTTDEDVRRLGKEVQKLTCQEGFTKNANLMRRLI
ELLEKQVDEWKEDQDIHGHDHCRQQKIQCFEKRIADYKKKCERSIRRYCDQRDSRTEIR
RERARDVHHNSNYDITDRGDSIRLNQHYHRQSLAPREESYHSSDDDEENIPPRLYRSGRS
TMYQQRQDDSNVHYRSHHSTLGASSSQQVRMRSGKSRVTKTHNRKWRPGQKALAEIR
KYQKSTDMLIQKAPFARLVHEIIQEATSFSKEYRIRADALMALQEASEAFMVEMFEGSVL
ICNHAKRVTLMPTDIQLYRRLCLRNLS
>C_sinica_HCP3L2

MDHVPPPLTHQPPSLIAAFLPRILVFEVKMQAMQCTPIIEEIHEAPTALEMEIEVVKHNIKK
LTSQPDFSKSYSLMEGAITILKKQIERWEYAERRDGPDGARQEHLAKFRLKMEQFEEKLK
AAERAYYEKKKAHLQEIEERRPLQERQNAINRQSFEQRRRERRNSSDESSDDETEQFHHR
SRSRSPRRNQLTTSNLQRHRQPTQPRLRQGVDRISKTKARKWRPGQKALAEIRQYQRTT
EMLIQKAPFARLVHEIIQDATSFSRDFRIRADALMALQEAAEAFMVEMFEGSVLICNHAK
RVTLMPTDLQLYRRLCLRNL
>C_nigoni_HCP3
MQRMYHHDSGPHIEEVFDPPPSQETMLREIASHPDVIALSKKVRKITKMPDSAFISSADR
LVEIIDAFSEQIEKWKEDETLDDPCPYLSLKIEFFTEKRNQYQRKNSSAVDRYYDGKDSR
DYSSRRPLEESRRREEPRDRGHETNIDITHRGDSTSLNHYSQRHYSQRQSQSSRFERDRES
EEKDENRHPRQQYRSRSPQHTHSYNQSTMHQRDDTNVYHRSHQSTSQPSQVRMRSGKS
RVTKTHNRKFRPGQKALAEIRKYQKSTDMLIQKAPFARLVHEIVREQTNQSKDYRIRAD
ALMALQEAAEAFMVEMFEGSVLICNHAKRVTLMPTDIQLYRRLCLRNLS
>C_briggsae_HCP3
MYHHDSGPHIEEVFDPPSRRTMMQEIETHPDVIAFGKKLRKIKNQPESTFLSSADRMEEII
DAFRDQIAKWEEEEELNEPCEYRQLKIEIFTQKKIEYQRKNNLAVDEFYKKRNLKNHSN
RKPLEESRRREEPRDRVHESNIDITHRGDSTSLNHYSRHHYSQRQSQSSRFERERESDEEE
ENSQPIQRYRSRSPKPSYSYNQSTMQQSQRDDTNVYHRSHQSTSQPPQVRMRSGKSRVT
KTHNRKFRPGQKALAEIRKYQKSTDMLIQKAPFVRLVHEIIREQTYKSQDYRIRADALM
ALQEAAEAFMVEMFEGSVLICNHAKRVTLMPTDIQLYRRLCLRNLS
>C_remanei_HCP3
MYQMHHNGPRIEEMVDPPSRSTTNQLKNDTEYIKSEYRRISHLPDFNRDPELIQEVMNLT
KRYIEKWLREERDEPNMERQGWIERFKTKLREWETKKETAEDEYYTRRDASSNEEKNR
EIARRRATDSQMNITGLHDSTRLNQQSYSRSYENRNRRYSSDEDDDENMAPQRRQRSRS
PPSFAHHQRRDDTGSYYRSHHTQNSSNQRTHNTDFSSHYRGQYGPSTSQNVGMPSNAQ
NVRMRSGKSRVTKTRSRKWRPGQRALEEIRKYQKSTDMLIQKAPFARLVHEIMREATSE
SQDFRIRADALMALQEAAEAFMVEMFEGSVLICNHAKRVTLMPTDIQLYRRLCLRNLS
>C_remanei_HCP3L4
MAPLLIAVSSFVVCSAILIYLCSKKPKTIDLESEIGITSRKNFNRDSDSMQEVIDIMTRQINK
WEQLEDDYGPDATRQRNIEAFQRKRDLWEEKKEQAERAFYERRDRMKREMEYGYSRN
QIRRERRMDDDSDSDVMEDDRRQPLGNLDYSRRDNYVRLREAKPMALLNRDRSRSRSP
LLLPRQDHPSTSLQLRRPNIPSTPPVRVRPGKSRVTKSKNRKWRPGQKALLEIRKYQKST
DMLIQKAPFARLVQEILRETTNESHDYRIRADALMALQEGAEAFMVEMFEGSVLISNHA
KRVTLMPTDVQLYRRLCLRNLS
>C_latens_HCP3
MYQVYHNGPRIEEMVDPPPKSTTNQLKIDTEYIKSEYGRISSHSDFNRNPDAIQEVIDLAR
RYIEKWQREERDEPNMDRRGWIERFKTKLREWETKKETAVDDYVTRRNASSNEEARREI
ARRRATDSQLNITGLQDSTRLNQQSYSRSYENRNRRYSSDEDDDENMAPQRRQRSRSPS
SFAYHQNTLNHQRRDDTGSYYRSHHTQNSSHQRTHNTDISSHYRRQNGPSTSQNVVMPS
NTQNVRMRSGKSRVTKTRNRKWRPGQRALEEIRKYQKSTDMLIQKApfarLVHEIMREAT
SESHDFRIRADALMALQEAAEAFMVEMFEGSVLICNHAKRVTLMPTDIQLYRRLCLRNL
S
>C_latens_HCP3L4
MDQIKEEIEAITSRKNFNRDSDAMQEVIDIMTRQINKWEQLEDDYGPDATRQRNIEAFQR
KRDSWEEKKEQAERAFYERRDRMKREMEYDNGYSRSQVRRERRMDDDSDSDVMEDD
GRQPLGNLDYSKRNNFVRGREAMPMKVVNRDRSRSRSPLPRPHHPSTSLQLRRPNIPSSP

PVRVRPGKSRVTKSKNRKWRPGQKALLEIRKYQKSTDMLIQKAPFARLVQEILRETTNES
HDYRIRADALMALQEGAEAFMVEMFEGSVLISNHAKRVTLMPTDIQLYRRLCLRNWS
>C_sp51_HCP3
MLQMDMDGPTIEELVDPQLDDTKQVIFKREVSDIEQECQKICSASNLPTNQQNRKILAVL
QKAIDKWREDEEIEGEYEPRIKAIEKFEKKRRLWLRNIENAGANSRERRERRYEDNMPEV
PRRASIYQETDITRRTNTTGLQQQSYQGSSSFRRNEYSSDEDTENIPYFRGSRYSSPPRKY
NQSSMYQQRRATSPIHQSQHTMNGTSNTQVRMRAGKSRVTKTAVRKHRPGQKALAEIR
QYQKSTDFLIQKAPFARLVHEIVREATSSSGDFRVRADALMALQEGAEAFIVEMFEGSVLI
CNHAKRVTLMPTDIQLYRRLCLRNL
>C_sp44_HCP3
MMLSDMDGPTIEELEDTQEDTRETIFNADIKKYENKMKAIAAVPGFLKDLDELRKIIDIC
DDAIEKWEAEEYRQGSYSRRREAIQQWGLKKRSFIDKLDRAEYVYYQKKEVARRASMP
RDVDITRRNDSTGLYHHSQQGTSYRHRTQEYSSDEEIENRPRSDRDRYRHLRAFSQNKIF
QYICFCRSPPRQYNQSSMYQQQRRDQSSMLQRSHQSTMNGSSTAQQVRMRSGKSRVTK
TTTRKYRPGHKALAEIRHYQKTTDLLIQKAPFARLVHEIIREMTPNNADYRVRADALLAL
QEGAEAFMVEMFEGSALICNHAKRVTLMPADIQLYRRLCLRNL
>C_sp48_HCP3
MFRVTDGPTIEEVVETQLTEDTAEAEVRRDYDTILEELRAVLGVPGANRDQERLGRGLCI
LEKGIDKFEEDEDNQPLEIRRQYLGKLREKYSSCEAKLREAENAFHERKEREYEERTMRE
IPRRYSSYRDTDITRRNNTTGLYHHSQQSSSNYRQQGYSSDEEMENFPSSHRDRYRSPPR
KFSHSTMLQQRRDISPVVNRSHQQSSASSQQVRMRSGKSRVTKTTRKYRPGQKALAEIR
QYQKSTDLLIQKAPFARLVHEIIREATSNSGDYRVRADALLALQEGAEAFMVEMFEGSVL
ICNHAKRVTLMPTDIQLYRRLCLRNL
>C_sp48_HCP3L5
MMDEDSPRIEEIVEEDEAEKDVKKEFDEYRREIEAVTSLPGFNRDSGKMTQVLRIMEKAI
GKWEEDEENLGSIQCRRRCLLEFKRRHDNYEKIIEKAEDDFYKRREMEMLEARRHHAQ
PESRQYPDGGAEQRQTIPKMRAGKSSVTKKPKKFRPGEKALAEIRQYQRSTDLLIQKAPF
ARLVHEIVSEATSSSGDYRIRADALMALQEGAEAFMVEMFEGSALICNHAKRVTLMASD
VQLYRRLCLRNL
>C_sp48_HCP3L7
MLEDHDSPHIEELVDADDEKNAEEAVKKEFFEFKNEIEKISCLPNFTKDSEKLRQVLAVIG
KAIDKWEEDEDNEGSIEVRRKYLKELKERYFKYDRIIEKAEQKFFARREQEMRNAGMKT
TRYYPGHGSSNHQYPFSREEDVVNSRSPSRHRPNSGHPHQHAPKLSSCRQQHGYNYWD
EKLCSRSPRRSNLPIRDQFHQSSTNAQRELKAGKSRVTKKVSHKYCSGQRAIAEIKHYQK
TTELLIQKAPFARLVQEVVQEATSESSSYSIRTDALSALQEGAEAFLVEMFEGSSMIANHA
KRATLGSTDLKLYRRLCLRNL
>C_brenneri_HCP3
MFHLSDGPTIEELVDTQQLENTAEAEFKEELDVIKKELAAVLAIPDIHRNREALEKSIRILE
KAIDKWEEDEENQVSLELRRQSIGKFKEQRRSCKQKLRDAENAFHERREREYEERTMRE
IPRRYSSFRDTDITRRNNTTGLYHHSQQSSSNFRMQEYSSDEEIENIPSSHRDRYRLEKCLII
VFQNLYFSYPPKKISHSTMLQQRRDISPVVYRSQQQSSAGSQQERMRSGKSRVTKTTRK
HRPGQKALAEIRKYQKSTDLLIQKAPFARLVHEIIREATTNSGDYRVRADALLALQEGAE
AFMVEMFEGSVLICNHAKRVTLMPTDIQLYRRLCLRNL
>C_brenneri_HCP3L5
MMMMEDSPHIEEIVEVEEAEIEVKRQFEEYKKEIEEVARLPGFNKDSGKMNQVIRIMDK
AIEKWEEDEENEGSLEFRRRCLWEFKHRHNNYKKIIEKAEEDFYKRREEAMKMVSILNR

TPQSHSYSQTRDGTAGSYRTQATSSGPRHRAFESDEERDQRERMQVGRSTISKKTPKRKY
RPGQKALAEIRKYQKSTDLLIQRAPFARLVHEIVREATASSGDYRVRADALMALQEGAE
AFMVEMFEGSVLICNHAKRVTLMPTDIQLYRRLCLRNL
>C_wallacei_HCP3
MHSNTDGPTIEEIVDNHLSDEQIAQEEAFKREFQLVREEMTQMTRGGANNDSQQLEKVV
NSLTRWIEKWEDDEDYGTKIKLRQDSIIKFRVRKDEYQRKIDRAEEEFFKRRQREQEESV
MRENPRRYSTYREDEITRRSDRTGLNHNSYEGSSTRHRTQAYDSDDDYENRPISSYDRYR
PSSKKSNQSTMYQQRRDESNSLHRSNFTMNGTSNSQQERMRSGKSRVTKTSRKYRPGQ
KALSEIRQYQKSTDLLIQKAPFARLVHEIIREETNKDYRIRADAIHALQEAAEAFMVEMFE
GSTLISNHAKRVTLMPTDIQLYRRLCLRNL
>C_tropicalis_HCP3
MHSDMDGPTIEEIVDTSFTDEQKSEEDAFRREIDRVKREMYNITSRPGWNNNSDELGKII
NIMTRYIDKWEEDEEYDSKINLRQDSIKKFREKRDDYQKKQERAEEEYFERRQRQHEES
VMRENARRLSSYRDDDITRRSNRTGLNQNSYEGSSTHHRTHGYDFDEDSENRPISHHDR
YRASSKKSNQSTMYHQRRDESSSLHRSHYTMNGTSHSQQQQQVRMRSGKSRVTKTNA
RKYRPGQRALAEIRQYQKSTDLLIQKAPFARLVHEIVREETNQDYRVRADAILALQEAAE
AFMVEMFEGSTLISNHAKRVTLMPTDIQLYRRLCLRNL
>C_doughertyi_HCP3
MYDDSDRPTIEEIDDTENSGYRTAEAIFQERAEEVSKQIKTLLKNPTTDDLKQVIRIMSRSI
DEWAEEEDNEGSIRARTTAIKTLTRKRNDYEMSLARKENEFLRKRGQRHEEESMTRENP
RRCSSFRDTDITRRTDRTGLNQSSYQGPSSNQQTHYYSSDEDYENRPRSNRDPYRYNDSP
QRSHQSSMYQQRRDASPPLHRSHTTLNGTSGSQQVRMRSGKSRVTKTTARKYRPGQKA
LAEIRQYQKSTDLLIQKAPFARLVHEIIREESSQTDFRVRADALLALQEAAEAFMVEMFE
GSVLICNHAKRVTLMPADIQLYRRLCLRNLY
>C_doughertyi_HCP3L6
MMQMYDDGPTIEEILDNHHSGGPITAETIFKEDFDDVSKQIKTLTRDGSTKNADNLKQIIS
IMSRSIDKWAEDEDMEGSIRMRKDAIEAFTKKRNEFREKITHAEDEYMRRKRQRMDEEA
LMREDQRGYFSSRDNDMARHAGGTALDFNYQQGTSSNYRARYYSSDEDYENAPRSDR
DRYRYPDSPQKSNQLATYHQRRDVSYSPSRSHQPMNGASSSTQVRMRAGKSRVTKKNS
RKFRPGQKALAEIRQYQKSTDLLIQKAPFARLVHEIIREETAVDDFRVRADALLALQEAAE
AFMVEMFEGSVLICNHAKRVTLMPADVQLYRRLCLRNL
>C_sp54_HCP3
MEQIYDDMRGRIEEIVDPPSRNTTVFRDLTQDQDSDVDRIRNQMKLIMQKPNFNTSIHEM
QKVISILDDQINKWESEEELRDNLFEKMERAKADYYKRKEAEDDSRTREEPRRRANYTD
MDITDRDNATRLNHLSYQRSYSHRNQIDNSDDDDIENMTRSRRDRSRSPPSHSYYQSTM
NQQRRDESYAHQRSHYSMSVPSSSHQVRMRSGKSRVTKTNSRKWRPGHKALSEIRMYQ
KSTDLLIQKAPFARLVHEIIRDTTSNSQDYRVRADALLALQEAAEAFLVEMFEGSVLICNH
AKRVTLMPTDIQLYRRLCLRNL
>C_sp54_HCP3L3
MRQIYDDMRERIEEIVDPPSRNSTVFRDLTQDQDSDLERVRNQIKRIIEKPGSNTSIHELQK
VINILDDQIYKWESEEETRDNYLEKMERAKAEYYKRKEAQDDSRRREEPRRRANYTDM
DITDRYNATRLNHLSSYQRSYSHQNQIDNSDDDDIENMIPRSRRDRSWSPPSHYKRKVAQ
DDSRTREEPRRRANYTDMDITDRGNSTRLNHLSSYQRSYSHRNQIDNSDDDDIENMTRK
YNTIFKSPPSHGYHQSTMNQQRRDESYAHERSHYSKSAPSNSHQVRMSSYQRSYSHRNQ
SDNSDDDDIENMIPRSRCDRSRSPPSHGFHQSTMNQQRRDKSYAHQRSHYSMRAPSGSH

QERMRSSKSRRVTKKNYRKWRPGQKALYEIQKYQKSTDLLIPKAPFARIVHEMIHKATST
SHDLRVRANTFLPLQEAAEAFLVQMFHGSMKYCNSAKRVTLMQTDIQNYRSP
>C_inopinata_HCP3
MHQMWNGMDGPQIEEIVDHPSRSSSVFRDVSQHMNRFFDNLKADLQRITSKPGFNKSF
DEMQKVIDLFDKNIDKLEHYEDMHGSSEELRRHIRLLREKRHNYDEKKEKARMEFFERR
ADENNRRREEARRRASSPDMDITDRNNGTRLNLHSYRQSYNQQYNINSSDEENYDVAR
NWRDRSPSRRTQSYNQSTLNVQHRDHSLQRSHHNMNDPSTSQPVRMRFGKNRVTKSLV
RKWRPGQKAMAEIRKYQKSTDLLIQKAPFARLVHEIVREATSHSNDFRVRADAIMALQE
AAEAFIVNMFEGSAMICHHAKRVTLMETDILLYRRLCLPHY
>C_elegans_HCP3
MADDTPIIEEIAEQNESVTRIMQRLKHDMQRVTSVPGFNTSAAGVNDLIDILNQYKKELE
DDAANDYTEAHIHKIRLVTGKRNQYVLKLKQAEDEYHARKEQARRRASSMDFTVGRN
STNLVDYSHGRHHMPSYRRHDSSDEENYSMDGTNGDGNRAGPSNPDRGNRTGPSSSDR
VRMRAGRNRVTKTRRYRPGQKALEEIRKYQKTEDLLIQKAPFARLVREIMQTSTPFGAD
CRIRSDAISALQEAAEAFLVEMFEGSSLISTHAKRVTLMTTDIQLYRRLCLRHL
>C_elegans_HCP3L1
MADDGPIIEEIAEKNGRVARIMQRLQHDTQRVTSVPGFNTSATGYADLIALLDQYKNDLE
AVGFNDLEQARRRAPSVDITVGSNSTNLVDYSHGRHDMPSHRRHDSSDEEITAANSHHQ
SPINVGNRNDTDGTNGRNGSRAGSSSSDRVRMIAGRNRISKTRRYRPGQKALEEIRKYQE
SEDLLIPKAPFARLVREIMQTSTPFSSDLRIRSDAINALQEASEALLVQMFDGSSLISAHSK
RATLTTTDVQLYRRLCLPNL
>C_oiwi_HCP3
MQQQDMDGPYIEEIIEPPSREVSVLRNPGRAEFERSMKQFELAVNGWRNDPELSNTSQK
MWGLVDIFNKQIDYCKFCCRQYGSSEDVEEAVKALTTKRDKMVKDINKSENRYFADKD
ERESSRREMARRRADYSELVVDRSNSTRLQQTSYRRPESHATRTHYSSDEENYHRRDRS
RSPTYSRSYDQSRTLQNRQDDNTTHSRRHHDTTGPSGSQQVRMRSGKSRVTKTTSRKYR
PGQKALAEIRKYQRSTDMLIQKAPFARLVHEIIRQETTQDSFRIRADALCALQEAAEAFLV
EMFEGSVLICNHAKRVTLMPTDIQLYRRLCLRNL
>C_kamaaina_HCP3
MQQQDMDGPYIEEVTESPSREVSIFRDAGRAEFASNMKSFELEVTRRKKDPEFVHNSDK
MWDLVGILDRQIEYCTRYSRENEYPDDVEEMTSAIRRKREKMAEEIKKLENRYFAEKDE
RENSRREIARRRANYSEVTVDRSNNTRLQQSSYRRPESHATRDHYSSDEENYHRRDKSR
SPTYSRSYDQSRTLQNRQADNSTHYRRHHDTNVPSGSQQVRMRSGKSRVTKTSSRKWR
PGHKALAEIRKYQRSTDMLIQKAPFARLVHEIVRQETTKDCFRIRADALCALQEAAEAFL
VEMFEGSVLICNHAKRVTLMPTDIQLYRRLCLRNL
>C_waitukubuli_HCP3
MQQMEEEEMDGPRIVELFDSPRRDTSVLRENNQYENSQGPQLSRYEVRIRSLMALPGFN
RDADLMGQAIELLKMHVAEMENEKRLNGRTPELDAKIEKTRNMAERCANKRNDALQR
FHEEREESRRAERMEMARRRANQSDLDITERENRTRLQPSSFHRSYEQRTVNYDSDERE
NEDRRAQSNSRRQRSRSRSPAYSRPPTMNQTRRDDTRNHRSYQNSTMATSSNQRKQTR
MRMGKSRVSKTHARKWKPGEKAMREIRQYQKSTDLLIQKAPFCRLVHEIMQEVTSFSSD
FRIRAEALGALQEAAEAFLVEMFEGSVLLANHAKRVTLMPTDIQLYRRLCLRPV
>C_panamensis_HCP3
MLQMEDLDGPRIEELPASPEREPAALRDNNRNGNVQRALPAHFENELRRLMSDPNFSKD
AELMTDAIELMKRQVNKMEDDQDMYGYEGGMAEMIRNLRIRIVSFTKKRDDAIARFRE
EREASRQMERTRRVNNTDFDMTDAENRTRLHPSMSSQRSYSYDRRDPYSSDENDDTYQ

PARHNAQQRMSRSPSPLMHQSHHSSINQRRDDSRQHRSYNNTATSSRPTTSNRSRMRVG
KNCVTKTKNRKWKPGEKAMKEIRRYQKSTDLLIQKAPFVRLVHEIMADVTPRSSEYRIR
AEALGALQEAAEAFLVEMFEGSVLIANHAKRVTLMPTDIQLYRRLCLRDK
>C_panamensis_HCP3L7
MLQMEDLDGPRNEEMPASPEREPAAEAQFENELRRLMNDPNFHRKADLMGKAIELMKR
QVNRLEDDQDVYDDDASRLVERTRRVNNTDLDMTDAENRTRLRSSVSSQRRDPYSSDE
NNDTDQPARHNAQQRMSRSPSSLMHQGHGSTINQRRDDNRRQRSYNNTATSSRPTTSN
QSRMRIGKNRVTKTKNRKWKPGKKAMNEIRRYQKSNDLLIQKAPFVRLVHEIMADVTP
RSSEYRIRAEALGALQEAAEAFLVEMFEGSMLIANHAKRVTLTPTDIQLYRRLCLRDK
>C_nouraguensis_HCP3
MDEMDGPRIEEVFDSPRRETSVLRETNQRANIRKRAPHPYEIQLKNLMKRPGFNQNAEL
MGEAIETMKKLCDVREREDETFGPTEGSTEFIRELKDRIRLYTTQRNEAIEQYRQEKENS
RQAEIARRRADYTNLDITDRENRTRLHPSSSSHRSYEQRNLRASSDEDEYDDQASRFRSQ
HHRSRSRSPAHSHTHHSTMNYQRQDDTRAHRSYQNSTAASGPNSRPKNQTRLRIGKSRV
TKTNARKWKPGEKAMREIRQYQKSTDMLIQKAPFCRLVHEIVQDVTSSSSGFRIRAEAL
GALQEAAEAFLVEMFEGSVLIANHAKRVTLMPTDIQLYRRLCLRNI
>C_becei_HCP3
MDEMDGPHIEEVFDSPRRETSVLREANQQANVRKRTRSQYEAQLMNLVKRPDFNRNAE
LMGEAIEIMKQQSDEMEREEEMFGPNEDSAYIRKLKDRIKAYTKKREDAIEQFRQERENS
RHTEMARRRANYTNLDITDRENRTRLQPSSSSYRSYEQRNLRASSDEDEFDDQTSRFRSQ
RQRSRSRSPAHSHTYHSTMNYQRQDDTRTHRSYQNSSAAAGSGSRAKHQTRLRIGKSRV
TKTNSRKWKPGEKAMREIRQYQKSTDMLIQKAPFCRLVHEIMQDVTSSSSDFRIRAEAL
GALQEAAEAFLVEMFEGSVLIANHAKRVTLMPTDIQLYRRLCLRNI
>C_yunquensis_HCP3
MLQMEEMDGPHIEEVFDSPRKEASVLREANQYANVQKRVSSKFEIQLKKLMTRPDFNR
DADLMSEAIELMKRQAAELEDNQLEYGSTAESSAYIRKLKDKIATYTRNRDEAMERYRE
EKEESRRMELARRRADFTDLDITERENRTRLQPASSSHRSYDQRTLRSPSDDEEYDDQPS
RYRSQRQRSPSTSPVRSQSYHSTMNYQRRDDSRNHRSYQNTAAASSSNTRNTNQGRLRI
GKSRVTKTKHPKWKPGEKAMREIRQYQQSTDMLIKKAPFCRLVHEIMQEVTGFSSDFRI
RAEALAALQEAAEAFLVEMFEGSVLIASHAKRVTLMTSDIRLYRRLCLRNL
>C_macrosperma_HCP3
MLQMDELDGPHIEEVFDSPRKEPSVLREPNQYPNVQKRGFSKFEMQLKTLVTSPGFNRD
ADKMTEAIELMKAHTRELEREERMYGRSEQSSAAIAKLRDKTESFIEKRDTALARFHAE
RDESRRLETARRRADYTDMDITARENRTRLPSSSYRSYEQRTRTYDSDEEEYEDQTSRFR
SQQRSRSRSPIQSQSYQSTMNIPRRDDTQNHRSYRNTTMASSSQPRQQTRVRIGKSRVTK
TTARKWKPGEKAMREIRQYQKSTDMLIQKAPFCRLVHEIMQDVTSFSSDFRIRAEALGA
LQEAAEAFLVEMFEGSVLIANHAKRVTLMPTDIQLYRRLCLRNM
>C_sulstoni_HCP3
MLQEHNGPIITEMVEPPSRDPSILRDSGRQYNVERNRPTAIEKQIKELFERPGFNDNLEAM
EQIISLMKKQIVEWQREQDLYGPTEERAKKIRAWQKNRETYIRNVNDANSRREELRRER
DESSRRMEMARRRVDMTQLDITARHNSTRLQQSYSQRSYDQRTRYESDEENDYVPQSQ
RHRSRSPPRAHQSYHSHSSQSQQARRPDTSRHVDRSQNDFSSAENPTASSSSQQQQRKPK
PRLRPGKSRVTKNLWRPRKDRALQEIRQYQKSTDLLIQKAPFCRLVQEIVRESSSSTSDFR
VRADALSALQEAAEAFLVEMFEGSQLIAAHAKRVTLMPSDIQLYRRLCLRNI
>C_afra_HCP3

MLHQHVGPVITEMEEPASRDSSILRDSGRRHNVARSEPTAIERQIQDIFNQPRCNERPEAM
AEAIRLMRKQIDQWEREQDLYGPTEERTKNIRIWKNNRRKFIAQLEVAKERQERARRER
DESSRVVETVRRRGDMTETNVTAIHNSTRLQSSQRSYDQRTHFDSDEEENGYTSRAPRPL
PRSPPRAHQSYHSNLSHQSRRLDASSQDNRSRNDYGSNNVTSVTSSSHQGKPKPRLRAG
KSRVTKNMFRPRRDRALLEIRQYQKSTNLLIQKAPFCRLVQEILREVTSSSDYRSSDYRIR
ADALSALQEAAEAFLVEMFEGSQLIATHARRVTLMHSDIQLYRRLCLRN
>C_afra_HCP3L8
MNNSRNTRLKSDIVPGRRIIPDVIYRDAGIEEEPSYLAERTMDESVVNSTRNTRWNSNIVP
GRRIIPDVIYRDSGIRGEPSYLAERTMDESVKSKRIGLGEREKRAADPTRSTMSSMIEDISS
PGVQFFENNREKMRPTVATPRRSNLGANRLSTASDKEKTIDMLSIGGEPTIEANGSSYVEP
VSVNGSRVSPTREDMLHRHVGPVITEVEESPSRDPSVLRDSGRRHNVARSKSTAIERQIKL
IFIEPRFKERPEAMAEVIRLMRKQIDQWERDQDSYGPTEERTKNIRAWKDNRRRFIEQLE
DAKERQERARRERDESSRVVETVRRRGDMTGMNVTAIHNSRSYDQRTHFDSEEEEKEN
GYTSRAPRPLPRSPPRAHQSYHSSLSHQSRRLDASSRDNRSRNDYGSNNVTSSSHQGKPK
PRLRAGKSRVTKSMFRPRRDRALLEIRQYQKSTDLLIQKAPFCRLVQEILREEATTSSSDY
RIRADALTALQEAAEAFLVEMFEGSQLIATHARRVTLMHSDIQLYRRLCLRK
>C_sp49_HCP3
MQRMDGPIIEEVVDHDARQEKRNRRINRIKDLMARPETKESKELLKKMLDLLQEHLND
LEDEQLEEGANHRDAIATLRHKIDIFTPLYNGAVADYMARKEESRREEARRAMSFGSSQN
NITGRDNRSKLHQSSQRTYSSEDEENDEEYAGRNRQDARRYHHQSPEPQASSSARRDNT
RRVDASSNNPRMRAGKSRVTKTNSRRWRPGQKALSEIRKYQKSTDMLIQKAPFHRVVQ
EILCETSGFTNAHRIRADAISALQEAAEAFLVEMFEGAMLLSNHAKRVTLMASDIQLYRR
LCLRKF
>C_sp49_HCP3L9
MARPETKESKELLKKMLDLLQEHLNDLEDEQLNEGANHRDAIATLRHKIDIFTPLYNAA
VADYMARKEESRREEARRAMSFGSSQNNITGRDNRSKLHQSSQRTYSSEDEENDEEYAG
RNRQDARRYHHQSPEPQASSSARRDHTRRVENSSNNPRMRAGKSRVTKTNSRRWRPGQ
KALSEIRKYQKSTDMLIQKAPFHRVVQEILCETSGFTNAHRIRADAISALQEAAEAFLVE
MFEGAMLLSNHAKRVTLMASDIQLYRRLCLRKF
>C_sp25_HCP3
MQRMDGPHIEEVFDSPRRSETRQSDRQWRIGRIQDLVHNPETGQSKDLLKEVLDLLKEH
RQDYEDEQDIDGVDHRQTIAKLTDKIDMYTPCYYRAVDDYNMRKEESRREEARRAVSF
GNSQHNITGRDNRSKLYQSSQRNYSSEDEENDEYVSRSHTRHADSRYESQDRDRDTRDY
QSSSSQAHHHHHHQRDDSRRHQTSSFPKQPRMRAGKSRVTKTNSRKWRPGQKALAEIR
KYQKTSDLLIQKAPFYRVVQEILRETSGFTNDHRIRADAIAALQEAAEAFLVEMFEGSAL
LSLHAKRVTLMPSDIQLYRRLCLRNF
>C_imperialis_HCP3
MQRMEGPYIEEVFDSPPRRHNSNRRELRVARLRELISDPATRESRDMLREVLDLLKEDKR
HFQDEQDNENIDHQKTISTLANKINQFERMYDRAVDDYEARKEESRRIEEARRTASFGGS
HHNVTARDNRSKLYQSSQRHYSSDEENEEYATRDRRQESRRDTRQDTRNYDSDGNDTR
DHQSSSSAYQRHDDTSRRQQSTNQQSSSSAPIRMRPGKSRVTKTNSRKWRPGQKALSEIR
KYQKSTDLLIQKAPFYRVVQEILRETSGFTNDHRIRADAIAALQEAAEAFIVEMFEGATLL
STHAKRVTLMPSDIQLYRRLCLRHL
>C_japonica_HCP3
MQRMIEMGGPHIEEIVDPPSPSNSVLQEADYRQNGPSRSRPKLIDQIRTLIRTPNFNKDAV
KMGQAIDLMELQIAEWVEEQIRYGFTQEREDAIYQYRRKVIVFKKSVEEAEERYFEEREE

SRRREEARRAMSYSRGDISARDNRSKLHQSHTQRNYGDSLDSDDENERENGYQSYRPPP
QRQQRLRSRSRSRSSSPMRSSYRHESPENSRRNASHQQTAQVRMRAGKNNVTKTKKWR
PGQKALSEIRKYQNSTDLLIQKAPFRRLVHQIIQEATGFDSGFRIRADAMSALQEAAEAFI
VEMFEGSVLISNHAKRVTLMTADIQLYRRLCLRNL

Supplementary Data 3.S4: HCP-3 protein motifs in C. elegans and C. sp54 found through manual alignments with sister species

***C. elegans* motif 1 in HCP-3/CPAR-1 aligned to *C. inopinata* motif 1**

>C_elegans_HCP-3_motif1_manuallyFound
-MADDTPIIEEIAEQNESVTR
>C_elegans_CPAR-1_motif1_manuallyFound
-MADDGPIIEEIAEKNGRVAR
>C_inopinata_HCP-3_motif1
WNGMDGPQIEEIVDHPSRSSS

***C. sp54* motif 4 in HCP-3/HCP-3L3 aligned to *C. tropicalis* motif 4**

>C_sp54_HCP-3_motif4_manuallyFound
-----EELRDNLFEKMERAKADYYKRKEA
>C_sp54_HCP-3L3_motif4_manuallyFound
-----EETRDNYLEKMERAKAEYYKRKEA
>C_tropicalis_HCP-3_motif4
SIKKFREKRDDYQKKQERAEEEYFERRQR

Supplementary Data 3.S5: All HCP-4 sequences used in this study

>C_tribulationis_HCP4
MNRKRGIRSSIVPGRRPITKIVALHEVGMDEDMSYMNEKTLLDNSSQMDDTVDAEERE
WRKKGLSENQIQRLFEARRQKQLKDAHDRLHKEKYGAKGFDEMLKLPQYASGRESED
DDDCHDENKSSSEIQVFAIPALPKHLNEKSIMGSPVAGSRGSGKAGLSCSTPKSGNDVSM
RSLRALDISHVVSTDHLDATKTTTHTKNIMIPVVESREESSLQKTHTIVKDASSELKRTFT
VVKDIEVSDADKEATKNKEDSSLQKTFNVIENSENDKTYVVVDSNEKEKESSLQATFVLS
KDWNDSLMPTEEKRKFESSKKDESSDHVQQVYVPGTSNQPTPNISVQADLLIVQNSTQG
GEIKKSNGVGAQKREKRGANMSTSLMTSMIEDVPSPGANYFKNPRKKLRNTAKSPSKSK
AIARLSTESDKEKTIDMLSIAEETSINAESVSSSYIEPISSDGTRVSPILEVEETMEVVTTTPK
SSRRGTAFSVRGSSMEKARGGIFCATPTAPTVDPTSPLVTPKLNYQKPTVSSMLKSKGPSD
LNHDLCATRGKNRPSVNNVVETSLSPSPDNNSTAKLDEQETDAQSKVIEESISDENAVRN
RTPGIEHQPSIGVDLPMDAMTIQSNNSNQHMNDFDMDYGDDVELHQYSEEKNRSGGGP
STRQRNRQRIGLLSDSIATINTPGVDRRPTRHLSRNETIPEGDWSDEEFEEMGRRRNGGK
HQRDVGLQLKKREIIQPDDTSDGVRRSQRTRVKPVRSWLGEQPVYINSPSGGKRLTGVT
DVIIKDKRLCKYKTADLRVATEREQKAKARRKEMAARRREQLARDHRRGRRLNESQEDI
HTDDDDDDEMS
>C_sp41_HCP4
MNRKRGIRSNIVPGRRPITQFAALHELGMAENMSYVNEKTILDESSLMNDTVDAEEREW
RKKGLSENQIQRLFEARKQKRLKDAQDRLNKEKYGARGFDEMLKLPQYYSGLESDDEC
HDENKNSSEIQVFAIPALPKHLSEKSIMGSPVAGSKGSGKAGLSCSTPKSGNDVSMRSLRS
LDISHVVSTDNLDATKTTTHTKNVMIPIVESQEESSLQKTHTIGKDASSEVKRVFTVAKDT
DISGSEKNMTKNKEDSSLQKTFTIAENCDNEEISVVVDSDQKDAESSLQGTFVLSRGWN
DSLMSTGEKRKNDQLKNIESADHVQQVGSPGTSNQTIPNAGTSVEVDLPIAQNSTQGGE
KRKSSGVGGEKREKRGANMSISMMSSMIEDVPSPGANFFKNPRKKLRNTAKSPSRTKAV
ARLSTESDKEKTIDMLSIAEEASIDAESVSSSYVDPISSNGTRVSPIPEAEETMGAVTTTPK
SSRRGAGFSVRGSMMEKARGGILCTTPTAPTFDAISPLVTPKHHYQKPTVSSMLKSKGPP
DIDQEMCITRAQNFTSVNNAVATPVRPLPSNNSSVDDEEQHARSKTAERDISEENEIRNRT
PDIVHQPSNDIDLPVEAMTIQSSDSNQRFDDFDMDYGDEIAPHQYSEERESNASGNGPST
RRNRHRVGLLSDSIATINTPGIDRRPTRHISRNESIPEGAWSDDEFEEAGRRRNGGRHPRDI
GLQLKKREIIQPEDTSDGVRRSQRTRVKPVRSWLGEQPVYVNSPSGGKRLTGVTDVIIKD
KRLCKYKTADLRVATEREQKAKARRKEMAAKRREQLARDHRRGQRLNESQEDIHTDDE
MS
>C_zanzibari_HCP4
MDRKRVIRSNIVPGRRPITEIAALHEVGIAQNMSYMNEKSLLDESSQVNDTVDVDEREW
KRQGFTENQIQRLYEDRKQKQLKDARDRLHKEKYGAKGFDEMLQLPQYSSGRESEDDD
EYHDENTNNIEIQVFAIPALPKHLSEKSIMGSPVAGSKGSGKAGLSCSTPKSGNDVSMRSL
RSLDISHVVSTDKLNPTRTTTHTKNIMLPILESREKPSMQNTHTIRKDVSSELEGTCIVAK
DDEVGGTRVESISQNKEISSLQKTHTIRKDASSELQGTFTVANDVEGAGSKTEADKNDPE
SSLQGTFVLSKDWSDCLISTQDKHGAELSKKKVDNVQLVDSPGTSQQTMSNVVVQVNP
SIVEVSMQGGEKGRFNGVGAQKREKRGANMSNSLMASMIEDIPSPGANYFKNPRKKLR
NTAKSPSKAKAINRLSTDSDKEKTIDMLSIAEEASIDTESVSSSYVDPVSINGTTQVSPIPEI
DEPMDVATTTPKSSNSRTAFSIRRGLMEKARSEILCATPPAPTENNTSSLVSPKLHYQKPTL
SSMLKNKGGLNSDVLLSATPRGKRAAVNNEIEVSQRPILNFDTSANLDHQETNDQTEPVE
ERAYNTNETRNSTFDVVNEPPNDIDLPIGEMTIQSNSNQHVNNNLDMDYDDDIEPNQGL

EENESNASRGGPSRGRRHRVGLLSDSIATVNTPGIDCRPTRHLPRNEEISDDSSVDDESEE
VGRRRNSGRHSREVGLLLKKREIIQPDDTSDGVRRSQRTRVKPVRSWLGEQPVYVNSPS
GGKRLTGVTDVIIKDKRLCRYKTADLRVATEREQKAKARRAAIKREQLARDHQRGRRLD
VSQDDIHTDEE
>C_sinica_HCP4
MKQXXXSGMDENLSYMNEKTVLDESSQMNDTTDVEEREWKKKGLSENQIQRLFEARK
QKQLKNAQERLHKEKYGAKGFDEMLRLPQYASGRESDEEEEYQDENKTITEIQVFAIPAL
PKHLSEKSIMGSPVAGSKGSGKAGLSCSTPKSASDVSMRSLRSLDISHVISTDQLDVTRAT
THTKNVMLPIVESREESSLQKTHTIKKDASSELQRTYTVAKDAEVVNTDTESASKNKDD
SSLQKTFTVAGNPESEETTSSLQKTFTVEAAEGKERESSLQNTFVRSKDWDNSFASTKEN
HKSENSKKIESTDKEQVVVSPKGAKQPMPNASIHVEVEPLIVESSIQGGERRKFNKTGAR
EKRGANNSMMASMIEDVPSPGANYFKNPRKKLRSTAKSPSKVKAISRLSTESEKEKTID
MLSIAEEASIDAESVSSSYVDPVSNNGTRVSPVPEVEEPMEVVTTTPKSSLRGTSYPIRGS
LMEKARGGLVCSSPTAPNLNNESLLGTPKLNYQKPTVSSILKSKGAPDCNDILGATRREH
RTPANNAISIPSRQSPVHNSPSNLKDDHEPSGLRKTPEKDISEVPEVRNRTSEIFHQPPNDID
LSMGAMTIQSADSNQRVNYDYDMDYGDDFEPYQDSDIREEHESESGGPSRPGKIRHRVG
LLSDSIATVNTPGVDRRPTRHVSRNETIPEDSWSEDELEEAGRRRNGGRRDRNVGLPLKP
RELIKPGNTSDGVRRSQRTRVKPVRSWLGEQAVYINSPSGGKRLTGVTDVIVKDKRLCRF
KTADLRVATEREQKAKARKKEMAARKREQLARDHRKGRRLNESQEDILTDDEMA
>C_nigoni_HCP4
MERKRAVRSTIVPGRKPITNIAALHEAGLSEDQSYLNEKTVLNETTQLNDTRDVEERKW
RKEGLDENDIFKRQEARKQKQMRDAEDRLNREKYGAKGFEEMLALPQYSSERESDEDD
HNEENQKAPEIPPIFAVPSLPKHMNEKSMMGSPVAGSRGSGKAGLSCSTPKSANDVSMRS
LRALDLSHVINTDHLDANKTIVQTKNVLLPIVENSEESTLQRTYTIHEDGNNERNLPTPEE
NEIADNEVRKFATPEDTKIVHNTMEGGEQTRKATKVGAQERERRGANLSMSLMHSMIE
DVPSPGANYFKHPRKKVRPETKSPQKPRVMGRLSTESDKEKTIEMGSIAEESSMTESIGSS
YVDPVPENTTAFSPVPEEEEPMEIANTTPKSSRRGSSFVTPHSLMERSRGATLFSTPVPPV
VDAITPKLNYQKQTVSSALKMKGAPNGCELLDVDRRCRSVPKKSANVARESPKDKGDG
NGKPEDIVIDDGKVNDEEMCDVTVEMPQNAEQAASGVELDMNGLTLNGNSANVSYNL
DHDGFDDFHGDPNLEDERNESEDAGPSTRRTARSRIGLLSDSIATVNSPGVDRRQTRRNF
RNDTIPEDSWCSDEEVVPRRRNDARNAVKIGLQLKKREIIQPDDAANGVRRSQRTRVKP
VRSWLGEKAVYVNSPSGGKRLTGVTDVIIKDKRLCKYKTGDLVLANEREQKTKARKKQ
LAAKRREQLSRDHQRGYRLNESQEDIVTDDELCDYT
>C_briggsae_HCP4
MSYLNEKSVLNETTQLNETGDVEEQVWRKEGLDENDIFKRQEARKQKQMRDVEDRLN
REKYGAKGFEDMLALPQYSSERESDEDDHNEENQKAPEIPIFAIPSLPKHMNEKSMLGSP
VAGSRGSGKAGLSCSTPKSANDVSMRSLRALDLSHVINTDHLDANKTNGEESDLQKTHT
IQKMASSEGSTLQKTHTIHEDGSNERNLPTPEENDIADTDVRNPSNGGEQKRKAKKVGA
QDRERRGANLSMSLMNSMMEDVPSSGANLLKHPLKKVRPETKSPQKPRVMGRLSTESD
KEKTIEMGSIAEESSIAESMGSSYVDPVPENATAFSPVPEEEEPMEIANTTPKSTRRGSSFV
TPHSLMERARGGTLFSTPVPPVVDAVTPKLNYQKQTVSSALKMKGAPNSCELLDVDRR
CRSGPKKSVNVARESPKDNDDGNGKPEDVVIDDSKLNDEEMCDVTVEMPQNVEQAAS
GVELDMNGLTLHSANVSYNLDHDGLDDFHGDPNFEDERNESEDAGSSTRRTTRSRIGLL
SDSIATVNSPGVDRRQTGKNYRNDTIPEDSWDSDEEVVSRRRNDVRNTVKIGLQLKKRE
IIQPDDATNGVRRSQRNRVKPVRSWLGEKPVYVNSPSGGKRLTGVTDVIIKDKRLCKYRT
GDSF

>C_remanei_HCP4_paralog1
MKQKHGSRSTIVPGRKPITQIAALHDAGVTEDMSYCNEKSVLDESSNMNDTHDNDIVDE
RSLIQQGIPDRDINKILQRRKLEKLLKVSQAQGRLERQMLGAKTFKELMEKHEYSERESG
DENNTAENISITNKPVFAIPALPKHFREKSMMGSPIAGSKGSGKAGLSCSTPKSVNDISMR
SLRILDISHVVNTDQLYYDKVVVHNKNVLIPIVENTEQSSVMGETFTVRDDQCDNKKHG
QVSANEASDSPIGSIKLDRNCISKELIDTTITENIVEGGKQRRRSNKVGVQERERRHADLN
SSLMNSMIEEVPSPGAKYFKNPRKKVRPIAEVPPKVFNMLSVESDKEKTIEMLSIVEEVSI
EAESNGPSFVDPLSVNGSHISPIPEVNELMNTANVTPKSSCPYIPILGNLMETVRVTQEND
VSSVITPKLNYLKPTISSLRKNINAPECDDILFGTRRERCTPGKNTTTAKRGVQDAPTIDK
TTVTVERITVKNDQRNIASNEDLSIGVPVETNRPSFNLELEMGEMSVRSSPKISNANLDSV
EPADFDPTEDRERVENQPGPFNQKSSRNRVALLSDSIATVNTPVYDYNPARCLEPDTNNG
NRRSTRTRVKPLRFFLGERAVYVNSPNGGKRLTGVTTVIIKDKRLCKYRTGDLKLATERE
QRAKAFKKKSAAGKRKRLLRDQQAGRRMDESSYDIHTDDEQ
>C_remanei_HCP4_paralog2
MERKHGWRSTIVPGRKAITQIAALHDAGVTEDMSYLNEKSVLDESSNLNDTHDDDEER
ALRKKGFSEREINKRLQGQKVEKLAKAHGRLKEQMLGAKNFKELMEKHEYSERESDDE
NDTAENITITNKPVFAIPALPKHLSEKSMMGSPVAGTKGSGKAGLSCSTPKSGKDVSMRS
LRVLDISHVVNTDQLDYDKVAVHNKNVLIPIVENTEQSSAKMGETFTVRDDQDDNQKH
GHVSANDASLQKTFTVDPRTEGDKSESSLQKTFTVPGGDQDSTRHSSLQNTFVKSGAND
SLLPNNGKRAQDDTKKIDESVNADQSGSVIDPSKLAGNCISNELIDTTITENTVEGGEQRI
RSNKVGVQERERRHADLNSSLMKSMIEEVPSPGANYFKNPRKKLRPTVEVPPKIISRLSV
ESDKEKTIEMLSMVEEVSMEAESNGPSFVDPLSVNGSRISPIPEVDKLTNPANVTPKSNRR
NLPIRGNSMETVRVTQANDVSSVITPKLNYLKPTISSLRKHVNEPECDDTLFGTRRERCTP
GKSATTAKTVVQDAPIIEKTSVTGEGVTVKNDQRNDASNRDLSTGMPEETNRPSFNLELE
MGDMSIRPSPKRLSANLDSVKPADFDLDLPTGRIENQAGPSNQRSSRNRAALLSDSIATV
NTPGYNRTARCRVMNDTNVEESWQSDEDDVILSRRNPGKNGKNVGLQLKKREIIQPDT
NNGNRRSARNRVKPLRSWLGEKAVYVNSPSGGRRLTSVSDVVIKDKRLCKYRTADLKL
ATEREQRAKAHKKELAARKREILLRDQQAGRRMDESHYDIHTDDEE
>C_latens_HCP4_paralog1
MKQKLGWRSTIVPGRRPITQIAALHDAGVTEDMSYLNEKSVLDESSNLNDTHDDDDERS
LIKKGIPERdinkelqrrkleklmkvqaqVRLEGQLLGAKTFKELMDKHEYSERESDDENYTAENI
SITSKPVFAIPALPKHLSEKSMRSPIAGTKGSGKAGLSCSTPKSGKDVSMRSLRVLDISHV
VNTDQSDYDKATVQNKNVRIPTVENSEHSSVARTKIGKTFTVRDDRDNNQKHGQVSAN
EASDSSIGSIKLAGNCISSELIDTTISGNTVEGGKQRRRSNKIGVQERERRHADLNSSLMK
SMIEEVPSPGANYFKNPRKKLRPIAEIPTKVINRLSVESDKEKTIEMLSMVEEVSMEAESN
GPSFVDPLSVNRSRISPIPEVNKLMNTANVTPKSRPPNLSIRANLMESVRVTQTNDVSSIIT
PKLNYLKPTISSLRKNVNAPECDDSLFGTRRERFTPGKIATTVKTVVQDVPSIEKTPVTVK
GITVKNDQRNIARNKNLSIGVPVETNRQSFNLELEMGEMSTQSSPKKLNANLDSVKPAD
FVWDLPAEDRERIENQPSSSNQRSSRNRAALLSDSIATVNTPEYDRISDRCYSINNKNVVI
PRRRNLEKNRKNVGLQLKKRVIIEPDDRNNGNRRSTRTRVKPVRSWLGEKAVYVNSPRG
GKRLTGVSDVFVKDKRLCKYRTADLKLATEREQRAKAHKKELAARKREQRLRDLQAG
RQMEESHYDIHTDDDE
>C_latens_HCP4_paralog2
MERKHGWRSTIVPGRKAITQIAALHDAGVTEDMSYLNEKSVLDESSNLNDTHDDDVER
SLRKRGFSERDINKEIQRQRNAKLAKAQGRLEDQMFGAKTFKELMEKHEYSERESDDE
NDIAENIPTTNEPVFAIPALPKHLSEKSMMGSPTAGTKGSGKAGLSCSTPKSGNDLSMRSL

RLLDISHVVNTEQLDYDKVTVHSKNVLIPIAENTEQSSAKMGETFTVRDDQDANQKHG
HVSAYEASLQKTFTVDPRKEGDKSESTLQKTFTVQQRAESDKSESTLQKTFTIPGGDEDS
TRHSSLQNTFVKSSSNDSLLPNNGKRAQDEKKKLDETVNAAQSDSEIDSSKLAGNCISNE
LIDTTITENTVEGGKQRRRSNKVGVQERERRHADLNSSLMKSMIEEIPSPGANYFKNPRK
KLRPTVEVPPKIINRLSVESDKEKTIEMLSMVEEVSMEAESNGPSFVDPLSVNGSRISPIPE
VDKLMNTANVTPKSNRPNRPIRGNSMETVRVTQANDVSSVITPKLNYLKPTISSLRKNV
NGPECDDILFGTRRERCTPGKSATTAKIVVQDTPIIEKTPVTEEGITVENDQRNVASNSDLS
TGMPEENNQPSFNLELEMGDMSIRPSPKRLNANLDSLEHVDFDGDLPTEDRERIENQPGP
SNQRSSRNRAALLSDSIATVNTPGYGRNRRRFMNDTNVEESWQSDEDDVVLGRRNPGK
NGKNVGLQLKKREIIQPDTNNGNRRSTRTRVKPVRSWLGEKAVYVNSPSGGKRLTGVTD
VVIKDKRLCKYRTADLKLATEREQRAKAHRKELAVRKREQLLRDQQEGRRMDESHYDI
HTDDEE
>C_sp51_HCP4
MENKRCFRSSIVPGRKMITNQVAYYQAGIIENMSYLAEKTIADDSSILNDTHDNEEREWR
KQGLTEKQIYHRLQDKKLQKLRNAENRLNKEKFGAKNFQELLNKHEYSQRESEGEEEA
TTEKFDVFAIPALPKYLSEKSMLGSPVSGSKGCGKPGLSCSTPKSANDVSMRSLRSLDISH
VVAVDQLDPKKVNIQSKTILIPIVQNSESSSLQKTHTSQKDQNSESLIGTKANSSLQESHTI
PNSPSQSQTYTSVQKTYIIPSSSPHKKQDDHQDQTEPLHEKDKETSLQESSNVDIERETSN
SDTLVVGTGLGNTLLDSVRKDQENEVQKLLNNNGDGLQEISSSNDVHIPVVQNLLQGG
KKKSATNKIGLQEREKRSNDSLMRSMIEVVDSPGAGFFKNARKKQRPIKTPTKMKVNRL
STESDKEKTIDMLSIAEEASMDNETNGSSFVDPVNGSKISPIREVNEPIDHSKSETTPKSRP
PTTPGSSIREKARAQRILEPLACNSFVVNNEIRLISTPKHHFQKPTFSSLVKNRNVDSNALL
ESSRRERINLPRRDITILKEAVPASLDEGEDQDDGTQNYFESTSDKTNRPDDDMSSKFDEI
SSHSTSDINDDLREMTVRDVPSMEIDYDPVDYVERDNQYDNSDVNTEHNERDSVNAGP
STKQKSFRKKIGLLSDSFASGSAPGIERRSTRHNFGNDTIVDDTWQPDEAPKRRNNKGRN
GKNMGMQLKKREIIEPSTPNDEVRRSKRTRVKPTRSWLGEKPVYVNSPSGGKRLTGVTD
VIINDKRLCKYRTADLKLATEREQKEKALKKMKAAERRRQLAADQKRGRRLNESQEDI
NTDDEYED
>c_sp44_HCP4_paralog1
MMKKHKERSKIAPGRRLITPQVAYHQAGIIENMSYLAEKTVVDDSSILNDTQDDEERQW
RKQGLSEKDIFRRLENRKLKRLKNAENRLQREKFGAKDFQELLNKHEYSERESDDEEEN
KNDGNPKPIEVFAIPALPKHLSDKSMMGSPVYGSRGSGKAGFSCSTPKSANDISMRSLRA
LDISHVVSVDQLDVNKVTIQNMKITIPITDKCESSSLQNAKTTAEDQNSTPQNGEGEDLTL
QKTHTVQSASLQKTYTIQDSSLQKTHTIQKPSLQETNTTSDSSLQETHTIQKPSPPKTIIITG
LSLQETPTIQEPSLQKTITISNSYLQKTHIIHEEDNIRILKKKNSDSIVQDEPATVKVPLVMN
QLQGGKKNQTTKGIGQQERGKRTANSSLMSSMIEDVQSPGAGLFKNTRKKLKPMVVTP
QRMKNNRLSTESDKEKTIDMLSMAEEASMDNETNGSSYVDPISVNGSRVSPVAEVDEPM
DVSKRSETPKTSRLTPRSSNIREKARAESIFATPSSDLKLNEDEVSLATPKHHYQKPTFSSI
VKSKNSLERNDLLEVTKQGRRNRSCRNNLIDEHPQQGIPDEVVDTQRENNTGKEKENSN
DSDVSGKESDNSSRSVDNIQAELEEMTVSGNPSTGFDYDSLYYNDQDNQVALCDPESPE
EEEDGGPGTSTRQKAPRRNVLLLSDSIATINTPGERRSTRRYSDTLAEETWVPDLVSKRSN
RGRNSKNTEMQLKKRVIIKPDTPADGVRRSTRTRVKPVRSWLGEQAVYANSPSGGKRLV
GVTDVIIKDKRFCKYKTADLALATEREQKERAWKRNKAAEKRRQLAADQSRGYRLNES
QEDIKTDDECDE
>c_sp44_HCP4_paralog2

MMGFPVYGSKGSGKAGLSCSTLKIANDTSMRSLKALNISHVVAGLDANKVTTQNRKITI
PIMEKWESSSKQKAKTTTEDQYSTLPNAKGEDLTRQKTRTVRSASIQKTNTIQDSSLQKT
RTIHEEDEMRILKKKDSDSIVRDEPATVKVPTVMNQLQGGKKNQTTKGNGRQEREKRTA
NSSLMSSMIEDIQSAGAGLLKNTRKRLKPMVVTSPGMKNSRLRTESDKEKTIDMLSMAE
EASMDNKTNSSSYVDQISVNGSRVPPVAEVDEPMDVSKRPEKPKTRRLTSRGSNIREKAR
AGSIFAIPSSDWKLNEDEVSLAIPKHHYQKPTFSSIVKRKSSLERNDLLEEMNQSRRNRSC
QNSLIDEYSQREIPDEEVNTRRENNITKEKENSNTTDVSGKESDNISRSLDNIQAELKEMT
VSGNPSTGLDYDSIDYNDQDNQVALCDPESPGEEDGRPESSTPQKAPSPRFIANSVSSKTE
HRAFVGEKKHREFVGEPMGDKEVTCIAGIGPTYGTKLTDAGFDKAYVLFEQYLFLKKDE
DLFVDWLKETAGVTAKHAKSAFNCLNEWAEQFL
>C_sp48_HCP4
MVNKRYVRSSIVPGRKMITTQVAYHQAGISENISYLAEKTVVEDSSILHESHDNEEREWK
KQGLSEREIYDRLRAKKLKKLRNAENRLNKEKFGAKNFLELLHKHEYSERESDDEENAT
VEMDVRNVEVFAIPALPKHLSEKSMFGSPVSGSKGSGKAGLSCSTPKSANDLSMRSLRSL
DISHVVAVDQLDAHKVTIQSKNILLPIAENAESSSLQKTHTIQSSSPAKTFTIPDSSLMKTH
TITRDLNTDSNNERKEESTLQKTYTIPSSSPQKSESIQDGSLMKTYTVQDSSLQKTFTVPD
SSVQKNAVVSNTSPSKNDNAQNNQSCQLQEEGEDMSLQKTFDVDADEQNTKCEGTFVV
QSGWRNTLLDSLRQDQEAQISKKGNGDLLSPSSSANIVLPVLENKLQGGEKKIGPKKTSL
QEREKRAANSSLMSSMIEDIPSPGAGLFKNTRKKLRPTNATPQRMKANRLSTESDKEKTI
DMLSIAEELSMENETNGSSFVDPVSINGSRVSPVIEVDEPIYISKNEATPKSRRHTLLGSSL
RERARAQSILASPASAVGVVRDIAPADITPKHHYQRPTFSSLVKNKTSVERNDLLEATKKD
RRNTPHRNTKTQEEAASLSGDEGVEDIHSETVPLSGEVDDCAEISSKPADISSHSITNIRLD
LDGMTVRDVPSMDIDFDAVDYVDHNHSDQFVAESPEESEGESDEAGPSSRKKSSRRKFG
TLNDSVTSLDARRSGSHSNFANDTVSNETWQPGETSRKNSRGGKNKPLEMQLKKREIIQ
PASPQGQVRRSERVRVKPIRSWLGEKAVYVNSPRGGKRLTGVTDVIIRDKRLCKYRTADL
KLATEREQKEKAYKKSVAAEKRRKLAADQRRGRRLNESQEDIITDSEDDNEK
>C_brenneri_HCP4_paralog1
MVNKRYVRSSIVPGRKMITTQVAFHQAGISENISYLAEKTIVEDSSILHESHDNEEREWK
KQGLSDKEIYDRLRDKKLKKLRNAENRLNKEKFGAKNFMELLNKHEYSERESDDEENA
TVEMDVRNVEVFAIPALPKHLSEKSMFGSPVSGSKGSGKAGLSCSTPKSANDLSMRSLRS
LDISHVVAVDQLDANKVTIQNKTILLPIAENVESSSLQKTHTIQSSSPPKTFPAQDSLLFMK
THTITRDSNTDSNNERDEESTLQKTHTIPSSSLQKSESIQDGSLQKTYTIQDASLQKTFTVP
DSTVHKNVIDSNTSPSKTDDAQDNQSCQLQEEGSDLSVQKTFDVDADEQNSTCEGTFVV
QSGWRNTLLDSLRRDQEAQSSNIGNGDLLSPSSTANNVLPVSQNHLQGGEKRTGSKKTS
LQERERRAVNSSLMSSMIEDIPSPGAGLFKNTRKKLRPTNVTPQRMKTNRLSTESDKEKT
IDMLSIAEELSMENETNGSSFVDPVSVNGSRVSPVIEVDEPMDISKNETTPKSRRHTLLNS
NLRERARALSILASPASAMKVVEDNAPTDITPKRHYLTPTFSSLVKKKTAVEINDLLEAKK
KDIRNTPHRNATIQEEEAVSISAGEGVDDIHSESVPLSQQVDDCAENSSKPVDISSHSITSIR
MELDEMTVNDGPSMDIDYDAVDYVDHNDRSDQSDAESLEESEGESDEAGPSSRKKSSR
RKIGTLSDSMTSMDARISGNHSNFANDTISNETWQPEETSRKNNRGGRNKTSEMQLKKR
EIIQPASPQGQVRRSERVRVKPVRSWLGEKAVYVNSPRGGKRLTGVTDVIIRDKRLCKYR
TADLKLATEREQKEKAYKKRVAAEKRKKLAADQRRGRRLNESQDDIFTDDDEQ
>C_brenneri_HCP4_paralog2
MVNKRYVRSSIVPGRKMITTQVAFHQAGISENISYLAEKTIVEDSSILHESHDNEEREWK
KQGLSDREIYDRLRDKKLKKLRNAENRLNKEKFGAKNFMELLNKHEYSERESDDEENA
TVEMDVRNVEVFAIPALPKHLSEKSMFGSPVSGSKGSGKAGLSCSTPKSANDLSMRSLRS

LDISHVVAVDQLDANKVTIQNKTILLPIAENVESSSLQKTHTIQSSSPPKTFTAQDSLLMKT
HTITRDSNTDSNNERDEESTLQKTHTIPSSSLQKSESIQDGSLQKTYTIQDASLQKTFTVPD
STVHKNVIDSNTSPSKTDDAQDNQSCQLQEEGSDLSVQKTFDVDADEQNSTCEGTFVVQ
SGWRNTLLDSLRRDQEAQSSNIGNGDLLSPSSTANNVPPVSQNQLQGGEKRTGPKKTSL
QERERRAVNSSLMSSMIEDIPSPGAGLFKNTRKKLRPTNVTPQRMKTNRLSTESDKEKTI
DMLSIAEELSMENETNGSSFVDPVSVNGSRVSPVIEVDEPMDISKNETTPKSRRHTLLNSN
LRERARALSISASPASAMKVVEDNAPTDITPKRHYLTPTFSSLVKKKTAVEINDLLEAKKK
DRRNTPHRNATIQEEEAVSLSAGEGVDDIHSESVPLSQQVDDCAEKSSKPVDISSHSITSIR
MELDEMTVNDGPSMVIDYDAVDYVDHNDRSDQFDAESLEESEGESDEAGPSSRKKSSR
RKIGTLSDSMTSMDARISGNHSNFANDTISNETWQPGETSRKNNRGGRNKTSEMQLKKR
EIIQPASPQGQVRRSERVRVKPVRSWLGEKAVYVNSPRGGKRLTGVTDVIIRDKRLCKYR
TADLKLATEREQKEKAYKKRVAAEKRKKLAADQRRGRRLNESQDDIFTDDDDQ
>C_wallacei_HCP4
MKLKGTARSTIVPGRKMITNQVAYHQAGIRENISYLAEKTIAETSSILNDTQDNEERDWR
KQGLTEKQIYLRLEDRRLKKIRNAENRLKKEMYGAKDFQELLEKHEYSERESDGEDVTV
TEDSQRKIEVFAVPSLPKHLEKSMFGSPAVGVRGSGKAGLSCSTPKSANDVSMGSLRSLD
ISHVVAIDQLDAHRVTVQSKNVLIPIVENGENSSLEKTFTDQKSQKSVEVSFHQNTPVIQN
DMEICEMSSLQKTYIIEENKDAELSSNRKEGSLLQKAFDKDDTDGSKRNSSLQNTFTVV
DNWKNTLLDSISEELENGHPKNTTDRAYPSSSSAGTKDATINDVVQPTDTSRIANDTLQG
GEKSRSSKKVGLQEREKRANVSIMSSMIEDIPSPGAGFFKNTRKKLRPTNTTPQRTKNNR
MSTESDKEKTIEMLSMAEEASVDLESIGPSFVDPISVNGSRVSPPRESEEPMDVTNAVTPK
STRRNNDNGMKAIRVQGTETSEAAESLKVRTPKYHYQTPTFSSLVKNKNANERNDLLES
TKRDRRHAPVRSVASVETEKDREKTQLAAQEEEEDLGTQKINDTRHASPAKVVDTSNNS
TRDINYELDVLSIRDQPELEDINEIPADREYENDNFDGFNDERRQTYAEEEDECSPSRVAL
MSDSIASVDFIDRRSTRRNMGNDTVVEDGWQPDEGSRRGNRGRNGSSSGLQLKKREIIQ
PATPTDEVRRSKRTRVKPVRSWLGEKPVYVNSPSGGKRLTGVTDVIIKDRRLCKYRTADL
KLATEREQKEKALRKNLAAERRRQLAADHKRGRRLNESQEDIHTDDEDE
>C_tropicalis_HCP4
MKFKSSSRSTIVPGRRMITNQVAYHQAGIKENTSYLAEKTIAESSSILNDTQDDEERDWR
KQGLTEKQIYLRLEDRRLKRIRNAENRLKKEMYGAKDFDELLEKHEYSERESDEEDVTI
VAEGHRNIEVFAIPSLPKHLVSDKSMFKSPAVGAKGSGKAGLSCSTPKSANDVSMRSLRS
LDISHVVAIDQLDVNRVNVHSKNVLIPIVEHAENSSLQKTFTIQKIQESIERSSTIQKDPETC
EISSIQKTIELEEKKTISLQVTQTRVYCLKKHQILIVIQKKLRKVHRFSLRLLSLLMLLCSSQ
TFRKSWKMLCKVAKKPEVRRKLEGMSMLSIAEEASVEIESNGPSFVDPISVNGSRVSPTG
EAEEATVEIESNSPSFVNPIPVNGSRVSPTREAEEATVEIESNGPLFSDPIPIGQAGESVDVS
NSITPKSSRHNQDRSLQVASNSSTGDINLELDVPEHQHDDFDGYEGVRRVDLVEEAEEEE
QIDHSPRSKQKSFKRRVALMSDSFASVDIPDLNRRPTRRYMGNETVEDDEWEPEKSEKK
KSNRNTGLQLKKRVIIEPATPTDGVRRSKRTRVKPVRSWLNEKAVYVNSPSGGKRLTGVT
DVIIKDKRLCKYRTADLKLATEREQKERAVKKKKAAEKRHQLALDHKRGKRLNESQEDI
LTDEEDGE
>C_doughertyi_HCP4
MVKENPKPVRSSIVPGRKMITKYVALHEGGMSENMSYLANKTVMEDSLDCTQDKEERE
WKRQGLSEEQIFRNLENRKLKKINNAEHRLMREKYGAKNFLELLDRPEYSELESGNEEP
SIAESKARDVVVFAIPALPKHLSEKSMFGSPIAGSKGSGKAGLSCSTPKSSNDESMHSLKA
LDFSHVIAVDTLEANKVTVHQKDIQIPMKDQANASSVEKTFIINVDHGIPETSSNSETVTV
QKDRTSDVANEEDESPCQKTFTVENESKNSSLQKTFTVAPGWKNSFLDSDCEHQKDGISE

PVDSTENVEEAGSSDKTRMDDSIMIARISTRSEITENDLQGGLKKNTSKLMGPQASKKRD
ANSTLMSSMIEDFPSPGAGLFKNTRKKMRPTNATQRKAINRLSNESDKERTIEMLSIVEEP
SVHNDSIGPSFVDPCSERLVNTSQVSPVQEKDEQIESSIVDTPKSTKRAVDQSSSMERARA
RSILSTPNPTILDNSILITPKHNYQKPTISSIVKSKNIDTALFEATKKDCRTRTPGRIDALLEV
VEEHTLTDDLEQGNHQDKKGRKNDKSGEDKIEFVPIADKTNSDLNEIRPEMDVSSVREQ
LEDFDFAADDRIERFGNEKVYDVNEFQEYHGEHERRHKTTTRRVALMSDSVAIENSLLTR
PNRRHVANDSTTEEGWQSDEEPVRYNRGRNNKEIGLQLKKREIIQPATPTNEVRRSTRTR
VKPVRSWLGEKPVYVNSPSGGKRLTGVTDVIINDKRQCKYRTADPRTATEREQKERAIK
KKAAADRRRQLALDQSRGNRMNESQDDIHTDDEE
>C_sp54_HCP4_paralog1
MITHVAACYEAGMMENMSYLMEKTIANDSSLDETQDLEEREWKRQGLSDIEIFKRLENR
KLKKLQNAEDRLKKEKYGAKDFDELLERPEYSECESFEKNDENVQNNYVFAIPSLPKHF
SEKSMMGSPAIGARGSGKAGLSCSTPKNGHNLSMGSLRGLDISHVCSIDRLDENKVTVH
NKNMLIPIIENYDSTCQKTHIIQKDQSSGLKGTFFVARDEEADNQKTNKIRKDTEESSLQD
TFLVAEDKNDDKRNSSLQRKSKKVDLQERERRYTNITMSLSNMSNITVGDISSPGSYRFK
QIPKKSKPTIKTPPRTKMMNRLSIESDKEKTIEMLSIVEEASIDADSNGPSFVDPLSVNGSC
VSPVPELDELMDVSNVVTPKSNRRICNTSIREKFMEMERACSPAAKPTIMNVESNNTLSV
TTPKHHYQKPTISSLVKSKNTIDINDLLEASRRDRRHTPSRTAIRTPNIYRTQTVKNATGSPI
YEQEEEEEEVVLEDETLNKRNDEKMNKMNDESISNAADASNRSLNNIEFELDSMTVND
QQYSDDFVYDKPQSFSDDRGIDEDTDSEENVSNLRSRQKSTRRRVELLSNSMASTPGIDC
RSTTRQGTSNETVTEDTWYPDIDYEVPQRKNKNRKNKVTVGLQLKKREIIKPERTSSNV
RRSGRNRVKPVRNWLGERPVYEFSPSGGERLIGVTDVIVTNKRFCKYRTADPKLANQRE
QKERAIKREIAARRQRDLALDHRRGRRFERFIREYSHERD
>C_sp54_HCP4_paralog2
MEIRKRIRSSIVPGRRMITHVAACYEAGMMENMSYLAEKTIANDSSLDETQDLEEREWK
RQGLSDIEIFKRLENRKLKRLQNAEDRLKKEKYGAKNFDELLERPEYSECESEFEKNDEN
VQKKFVFAIPSLPKHLSEKSMMGSPAVGARGSGKAGLSCSTPKSGHNVSMGSLRGLDISH
VVSIDRLDENKVTVHNKNVLIPIVENDDSTCQKTYTIQKDQNSGLQGTFIVARDEEADNQ
RNNDARKDTEESSLQKTFTVAEDKNDDKRNSSLQDTFVVAEDKNDDKRNSSLQDTFVV
AEDKNDDKRNSSLQETFTVAENKNGDKRNSSLQETFTVAEDKNDDKRNSSLEETYTVIS
VGDMPSPGSYRFKQLPKKSKPTIKTPPKTKMMNRLSTESDKEKTIEMLSIAEEASMEADS
NGPSFVDPLSVNGSCVSPVPELDEFMDVSSAVTPKSSRRICNTPIREKSMEMARARSLAA
TPTIKNVESNNTLSATTPKHHYQKPTISSLVKSKITIDSNDLLEASRRDRRHTPGRTANRTP
DRHRTQTVENATESPIDEQEEEEAALKDETLNKRNDEKMNEMNDESISNAAATSNRSLN
NIEFELDAMTVSDQQYSDDFDYDKPQSFSDDRGIDEDTESEESVTRLRPRQKSTRRRVEL
LSNSMASTPGIDRRSTRRQGTSNETVAEDTWYHGVDYEVPQMNNKNRKNKITVGLQLK
KREISK
>C_sp54_HCP4_paralog3
MEIEKKIRSSIVPGRRMITHVAACYEAGMMENMSYLMEKTIANDSSLDETQDLEERGWK
RQGLSDIEIFKRLENRKLKILQNAEDRLKKEKYGAKDFDELLERPEYSECESFEQNDENM
QKNNIFAIPSLQKHSNEIEKRIRSSIVPGRRMITHVAACYEAEMMENMSYLMEKTIANDSS
LDETQDLEERGWKRQGLSDIEIFKRLENRKLKKLQNSGKAGLSCSTPKNGHNVSMGSLR
GLNISHVCSIDRLDVNKVTIHNKNMLIPIIENDDYSCQKTHIIQKDHSSDLQGTFIVARDEE
ADNQKTNDVRKDTEESSLQDTFVVAEDKNDDKRNSSLQRKSKKVNLQERERRYANISM
SLSNITIGDMSSPGSYRFKQIPKKSKPTIKTPPRTKMMNRLSIESDKEKTIEMLSIVEEASM
EADSNSPSFVDPLSVNGSCVSPVPEFDELMDVSKAVTPK

>C_inopinata_HCP4
MHQRRNGRSTIVPGRRMITDQMAYHEAGLDVSNLSYLAEKTYMNESSVANDTVDQEER
EWKKQGLSEKEIFNILERRKLKKLRDARNRLEKEMFGAKNFQELLERPEYSEIKNTDEDL
NENAQKEPIFAVPAPPKHVSEKSMMSSPAAGSKGSGKAGLSCSTPKNSQDLSMRSLKLLD
ISHVVNIDRLDAKNVTVHNTKIVIPISKDTEKESTVLEQSSIDLQSTFTVETHKDPNAGKE
PSLQKTFIVENEEEENQAESSLERTYVVKEGWDNAQNFDSAFEKKGTVNSIICDPVCGEN
TIEGGEKKKKSNKIGVQKRENRNVDLTTSLISSMIEDIPSPGAHYFKTTRKKFRPDPKTQS
GTNKNNRLSTESDKEKTIEMLSIAEEASMEAELTGPSSIDPISVNEFQVSPIPEEIIAATTLRT
PNSNRRENLPNNGISDETMRYKRMTETTPTRFPFESEDAATTPKPNYLTPTFSSLVKNRKI
TIADTNELLQANRRQGYLRNTPSQNYNARTKMTVSDTPFEVNTDPIEPQNDNQVVSSGE
KDETTEVVNSSLDTALNALSVRERPDQPDIEDVDSDRYCQHDDDEDGFDNIGDLDYEEK
VEEIPVSPKPRSIRRRKELLSDSVATVNTPGLDRRPTTQDTTYGTFVEDTWCPDIEKPKKT
AGKNSTNEGMQLKKRELIMPGDAAGGVRRSSRVRVKPVRTWLGERAVYVNSPSGGKRL
TGVTDVIIKDKRLCRFRTADLKLATEREQREIAYKKEIALRKRQQLAHDHRKGRRLNESQ
DDIHTDDEDDV
>C_elegans_HCP4
MNRKPRTRSTIVPGRKMITEIVALREAGLDDTRPSYLEEPTVIVDESMMNDSANLEEREW
RKQGLSEKQMFAILEKRKQTQLNNARKRLEKEMYGAKTFKELIACREYSECESDSENTV
NQNVRSASVFAIPALPKHISEKSMMGSPIANSRGSGKAGLSCSTPKSSSDTSMRSLRSLDI
SHVVNTDRLDAERVTVHTKSVVIPTILEERESTLQKTLTIEKNHSSQLQKTFTVAEDAPEE
LQKTVTIEKNQSSEMQRTFTVAKDASKEQSQLQRTVTIEKNQSSQLQGNFTVAKDAPEEL
QKTVTIEKNQSSEMQGTVTVAKDAEHSSLQKTFTKPTEKDESSLQRTFNVANRDENNDS
TLQKTFIIEERDAYEQGTTSVGVKPPVLAQNTMEGGEKRRTSKVTSDERERRNASISNPL
HNSMLEKEEIQSPGANFFKHPRKKIRPVVQTPPRIKATSRLSTESNKERTIEMESVAEERT
MEADSIGSSYISPTFNASRVSPVPEVPEKVEPMHVSKAATPKSIRHTDNSIRNIGPTNNVEC
ARAALLSTPTRMDIVDSVNRVTPARRYEQPTFASLVKRMNMKDANRLLEETSRKNTPAK
TTATTSSAAVRMVLEDDEEDQATEVIEKRSENGGVIVDGEDEAADSSNRSLNIELNALTV
NEEPAHDISAIDFPEDDTNEMRSSSDEEEMEARPDPRKKSSRRLGLLSDSIALGLASSSRR
RPNDTFVDETWYPEPNSKGNRRPRTHRGMKLKEHQLMKPEDAPDGVRRSTRVRVKPVR
SWLGEQPVYVNSPISGCKRLTGVTAVVIKDPRLCYYRTADVRTATERELKDKANKRALA
QEKKQQRQNARSGRRHDSDDDEEDDM
>C_oiwi_HCP4
MHSASKIRRARSTIVPGRRMITAEVALHDCGMDVDQSYLNEKTIQSDESTIMNDTRDEEE
REMRRRGLSEQEILKIQENRKREQRKNAERRLHEHKFGAKNFKELLGKHELSACESDGE
EENRTVVIHRNQGEGGSKEIFAVPALPKHILGKSMLGSPVGARGSGKAGLSCSTPKGSKD
VSMSSLRALDISHVVNTESLDSIKVTVQNKNIVIPAIPEESSLQKTHTIHRSDLEGTFIIKER
SLDKTFTVESEKEMEQISEKNQENPSATEIVPAQEESSNSTLQKTFTAEKEWNDSLLDIVR
KKQAPGPVEEEVPVVEASIQGGSKKSKKVNLQEREKRMADLTSSLLTSMNDDIPSPGAH
CFRNTRKKIRPSLQTPPRAKVTSRLSTESDKEKTIEMLSIVEDGSMGGACSNGPSFVDAVS
NNATQVSPIQEVPEPQTPKSSSHPLHIANDTLRLESSRARQHLDASSAITPRRNYERPTISS
TIKSKNTAECESEVIVAIRSTERRHRRLPHDDDNGTLQAPEKPSEHVQLTEPEPEEEEMEQ
QSDVVETINQDSGESPRNQRPSVAAVEAKMDALTFEVTSPNMHHVMEFDYDNDQEEDD
EVPEEPANDTTPRLKLSRRRTQLISNSMTSDESGQGPSRRRGFTDETVAEETWHPNLDEPS
RRNRGGRKGKNDGLQLKKRELIVPDDEAPDGVRRSTRTRVKPVRSWLGEQPVYVHSPS
GGKRLKGVTDVVIKDKRLCKYRTADLKLATEREQKERAMKKERAAQRRAQLAKAHKR
GQRMNESQDDIHTDDEESD

>C_kamaaina_HCP4_paralog1
MHSASKIRRARSTIVPGRRMITSEVALHDCGMEVDQSYLNEKTIQSDDSSIMNSTRDEEE
REMRRRGVSEQEIFKFQENRRREQLKNSERRLQEQKFGAKNFEELLARHEFSACESDGE
EENRTIVIERNEEASKATFAVPALPKHILGKSMMSSPVGAKGSGKAGLSCSTPKGSKDVS
MSSLRALDISHVVNTESLDSAKVTVQNKNIIIPVIPEESSLQKTHTIQRSDLEGTFIIKERSL
ERTFTVESEKEKEQSVEENQEGPSAEEDSESSLQKTFTAEHQLSDNSLLNIVRNKQAAAP
VPVEAEEVPVAEASIQGGSKKPKKINLQDREKRIADLTSSLLTSMDDEIPSPGAHCFKNTR
KKIRPSLQTPPRAKAISRLSTESDKEKTIDMLSIVEDGSMGGACSSGPSFVDAVSVNATQV
SPIQEVPEPQTPKSSHHPLPSGNNTPCVESSRAKQHLDASAAITPRRNYERPTIASSMKNK
NAADNEALLASLSKDRRPRRLPLDEDDVLEKSSVNVEMTEPEPEEVVETSHQDSGESPR
HQRPSVAAVEVEMSALSFRGPSPNTNHVMDYDYHNDNGEEDADESEEPANANDTTPRL
KLTRRQTQLMSNSMASEGPSRRRGFTDETVAEETWNPSLDEEPSRRTRGGRKGKNDGLQ
LKKRELIVPDDDAPAGVRRSTRTRVKPVRSWLGEQPVYVHSPNGGKRLKGVTEVVIKDK
RLCKYRTADLKLATEREQKERAMKKERAAQKRADLAKYHKRGERMNESQDDIHTDDE
D
>C_kamaaina_HCP4_paralog2
MHSASEIRRARSTIVPGRRMITSEVALHDCGIEVDQSYLNEKTIQSDDSSAMNSTRDEEV
REMRRRGVSEQESSTPKGSKDVSKRSLLALDISNVVNAESFESAKVKTFTAGLQLSDNTL
RNIVRNKQAAAPVEVEEEEFPVAEASIRGGSKKPKKINLQIREKRIADLTSSHLTSMDDDI
LSPGAYCFQNTRKKIRPSLQTPPRAKAISRLSTESDKEKTIDMLSHAEDGSMRGACSSGPS
FADAVSVNAKQVSPIQEVPEPQTPKSPHHHLPSGNHTPCVESLRAKQHLDVSSPITSRRN
YKKPLIASLKIRVNTKKPEEIVETSNQDSEESPRLQRPSVAAVEVEMSALSFRFPSPTRNHV
MHFDYENDNEEEDADESEEPANANDTIPRLRLTRGQIQQMSNSMASEAPSRRHEFTDDT
VAEDEEPP
>C_waitukubuli_HCP4
MTRGSRKSNKIVPGRHMLSDQMQLHYLGMADTNHSYLAEKTADESSLMDCSREREER
MWRKQGLTEKQIFRIQEDRVLKKHADAEARLEREKYGAKTFEEMLALSQYAIGALDVG
EDDEPDVSKRENVFAIPALPKHVKPKSTMMSPAVRTTGKPGFGCSTPKGGKDLSMQALK
TLDISHVINTDDLQSKKVNVQEKNVRISLKRTSDHANEEVSLEQTHTIRRDEDSELQGTF
TVQKEGEGSLQGTFTVANDAEAEQSLQGTFVKEPWASMLCGTEDERDNGQSSGLQGTF
VKESAGIEDQSQNNGAAENRNESESLQGTFVKAVDAPADAQQTKLNNTLDLMMLISNE
SNAAAQMTVSEGTVIVENTMEGGKSRRKSKKTDLNEREKRMADMTSSLINSMIDDAPS
PGAALFKNTRKKIRPTVATPPRTRISRVSIESDKEKTIEMLSIAEEPTLEATGTSFVDPISVN
GSRISPIPEHKESHPPVITATPKSSRRTGSLREQSVEKSRATGPSMRLSVASPFVGDSVIDTP
VRTTPSRNYEKPTFASLAKSKNSKVVNLCETEKGHPEKQMSQKTLVRTPVLASKDFKDP
EDDQLLGHVSVNTSDIPPRSEVQQVQDDDDPTVFPVTSNISIRNIEGDLNALSVNERNTV
PVDYDMDFDDDDRQDPEMDRDEEEVEAEEEQVPARIGRQKPRSRRLGLLSDSIASVNTP
EIAENRRPTRTHRPTDDTVVDDGWNSDEEYEDVPRRSRRNDQVGLQLAKRKLIKPADN
DDGLRRSTRTRVKPVRSWLGEQAVYANSPSGGRRMTGVTDVYIKDKRLCKYRTADLGV
ATEREQRARAMKRARAEQKRRQLAHDQSMGRRLNESQDEIVTSEDEQ
>C_panamensis_HCP4
MEKKRSSIVPGRRMLSEQMALHQFGMADDNKSYLMEKTCDESTFMDCSQDREVRKWR
KDGFSETQIFRIQEERTLKKHTEAEARLEREKCGAKTFEELLALQEHARRNQTMDTDEEL
DADPQGSVFAVPALPKHIKAKSTEMSPAGKAKGKPGFGCSTPVKGKNASMRSLETLDISL
VVKPDALHPKNVNIQEKNIRVSLKRSTPHFHRIDDESERVEKDDTLQGTFTVAKDSASDS
SLQKTHTIHKPDEDPERTFTVQKDDSLQGTFTVAKDSVNDSSLQKTHTIHKENDEPGGTF

TVAKDASLQGTFTVAESSSQSLIKTFVRPPWDDEGLSAPSSQPAVVEDSAEDRSASLQGAF
VKETDEAEATGKADQLQNTLDVFESNERTIVIEDVHLAQNTVEGGPSRRKSNKMDLEER
EKRTADMTSSLLNSMVDDAPSPGQMFFKNANRKKIRPTTATPPRNKYVGRLSTDSDKEK
TIEMLSIAEDPSMESNGSSFVEPVSVNGSRFSPVPEEHDGGQSMDLSVTPKSTRRSTVPLR
QQSVEKDRAIGVSFQVSMCQSAVGSPCATPLRNYEKPTFSSLGKNKNATMFETEKAKQL
EKMNKKTSIDHTPTRASTALTLAGSNEEEDVVVAPVSIPNPNRSPSSPLAEEHDEDTPTVF
PVASNTSIGNVEADMNALAMNDNNDYGVNYDMDFDDDYRQDHRDVEEEDEQSPAPV
QRPKPRSRRGVALLSDSMVSVNTPGIAGRRTTRRAMNDSAVDYNSNSDEEEDEGTSRNS
RRNGNKKGLQLAKRKIIKPQANDDGLRRSTRTRVKPVRSWLGEQAVYVNSPSGGKRLK
GVTDVFITDKRLCKYRTADLGIATEREQRERARKRARAEERRRQLAHDQSMGRRLNESQ
EDINTSDEEQ
>C_nouraguensis_HCP4
MERNRRTNNRIVPGRRMLTTEMAMYYAGMPDENRSYLAERTADESSLMDCSQDREERL
LRKQGYSEKEILKMQEDRVQKKHADAEARLEREKYGAKNFEEMLNLPQYSMREPADD
DDDDAGPQENIFAVPALPKHVNETSVVMSPAGQTIVKLGFGCSTPKGGKDMSMHSLKTL
DLSHVVNTDKLHSQQINIQEKKVVISMKRTSNDDTESSLEKTHTIRNTFIVQNDDTLQHT
FVIVKDPEEGLQATFTKEVEKNALSDSSQSTCLQGTFVKENDGQDQESVKEAGENPAED
QGSPTSLQNTFIKTVDGNENNNQNTLQLFEFTAHKQNSLMEETVSEQTELVHNTTEGGK
NRRKSKKMDLGEREKRNADITSSLITRMIDDAPSPGARLFQNNRKKMRPTMPTPPRAKF
GSRLSTDSDKEKTIDMLSIAEEPTMESNGSSFVDPMSVNGSRVSPIPEKDETQSMGISATP
KSIRRTTGIRPISVEKSRASVLTPRISMATCLSDSVHIPAVTPTRNYEKPTLSSQLKNKNATE
ASVWEAEKEKEIGKMNRNTPKRTHATVAPVEASSDEEEEQVVVTPVNKNRSPVFERLDE
SPDEPTVYPGSSDTSIGNVESDLNALSVNDRFHAPAPEFDLGPDYDRSNQEEQSNGEEEE
VEHREPSPAQRANPTSSRRGIALLSESMVSVNTPGIAGRRPTRNRRQVDDTVVDDGWNS
EDDDSNRRGRRRNGKQEAGLQLAKRRIIEPKDAGDGLRRSTRTRVKPLRSWLGEQAVY
VNSPSGGKRLKGVTDVIIKDKRLCKYRTADLRVATEREQKEKARKRARAEQRRRQLALD
QSMGRRLNESQDDILTSDDEL
>C_becei_HCP4
MERNRRTKNSIVPGRRMLTTQVAMYNAGIHDENRSYLAEKTADESSLMDCSEDREERM
LRKQGYSEREILKMQEERVQKKHADAEARLEREKYGANNFEEMLNLPQYSMREPADDD
DDDVGPQENVFAVPALPKRVNETSIVLSPAGRVVGKPGFGCSTPKGGKDVSMRSLRTLDI
SHVVNTSKLHSERINIQGKNVVISMKRTSDADEESSLEKTHTIRNDGTFTVQKDDSVQRT
LIIIKDPEEDLQATFTKETAQDAQSNSSRSSSLQGTFVKRTDGQDQESVKDTEENPAENRD
SSASLQGTFIKPVDSDENKNQNTLQVFESSAQNQSSLMEETVTDQMEVVHNTTEGGKN
RRKSKKMDLGEREKRNADITSSLISSMIDDAPSPGARLFQNNRKKMRPTMPTPPRAKFG
SRLSTDSDKEKTIDMLSIAEDATLESNGSSFVDPISVNGSRVSPIPEKDETQPMEISATPKSS
RRTTGIRPASVEKSRGPLLSPRISMAPCLSDSVTVAAVTPTRNYEKPTASSLAKNKNPAEAI
RWEALKEKELGKINRQTPKKTPVAMPLAESSSDEEEEQAAVTPVNVNRSPAPDRLDESM
DEPTVYADPVNSSIGNVESDLNALSVNDRFDAPAPDFDLGPDYDREPQSDGEDEVEDRM
PSPVQRARPKSSRRGIALLSESMVTVNTPGIADRRPTRNRRQVDDTVVDDGWNSDDDDS
NRRGRRRNGKQETGLQLAKRKIIEPEDAGDGLRRSTRTRVKPLRSWLGEQAVYVNSPSG
GKRLKGVTDVIIKDKRLCKYRTADLRVATEREQKEKARKKARAEQRRRQLVLDQSMGR
RLNESQDDIHTSDDEL
>C_yunquensis_HCP4
MDRNRRTNNKIVPGRNMLSAEMQLYYMGVEDDNRSYLAERTVDESTLMNCSQDQEER
MWRKQGLSEKEIFRIQEERALKRHADAEARLEREKYGAVTFEDMLALPQYSVREPADD

DDNDDVSPQENVFAVPALPKHANEKSIVMSPAARGIGKPGFSCSTPKGGIDLSMRSLRTL
DISHVVNTDKLQSRKVNVQERNVVISMKRTSDRVDGTTSLQRTHTIRKDDEADLQGTFT
VQKSPDRTFTVVREAEESLQGTFVKETDNNAQSDNSRSSSLQGTFVKQNEKSSSLQGTF
VKETDNNAQSDNSRSSSLQGTFVKQNEKSSSLQGTFVKQTEESTSLQGTFVKEPSSDQSK
EPNLQSAFVQETVGAPTEKRDSCAVLQGTFVKHANDSGNDNQNTLQVLQFQTHDESSV
LVERVTEEVRVVHNTIEGGQNRRQSKKVDLDRREKRDADMTSSLITSMIDDAPSPGAQL
FKNNRKKMRPTMPTPPRAKFGSRLSTDSDKEKTIEMLSIAEEPTLESNGSTFVDPISVNGS
RVSPIPEKDETQMMDIASTPKSSRRTGSIRPSSVEKSRNNALSMRLSVAPSLTDSLIVTAITP
ARNYEKPTFSSLVKNKNSVEARIFLETQKDLGKMSKQTPKRSVAVHPVEPSSDEEKDQAV
QAPNNRSISPERRDESMQDEPTVYPGSSNTSIGNVEADLNALSVHDRYDAPLDLDLDDF
RPNDREAQSKGEDEVEERQAFPVRQAEPKSRRRVGLLSNSMITVNTPGIADCRQNRNRR
QVDDTVVDDGWNSDEDDSNRRNRRRNRKQEVEMQLAKRKIIVPEDAGDGLRRSARTR
VKPLRSWLGEQAVYVNSPSGGKRLKGVTDVVIKDKRLCKYRTADLALATAREQKERAR
KRARAEQKRQQLAHDQSMGRRLNESQDEIVTSDDEH
>C_macrosperma_HCP4
MKPDSRTSNKIVPGRKNITDLHGLHSAGMNDENRSYLAEKTADESAFMDCSQDHEERI
WRKQGLSEKEIFRIQEQRALKKHADAEKRLENEKFGARDFKEMCALPQYSLREDNEEE
DDDRFCSRENVFAVPALPKHVVEKSIVLSPAGRSLGKPGFSCSTPKGGKDLSMRSLKTLDI
SHVVNSEKLQVKKVNVQDKNVTITIKRTTEQAEKKKSIEKTHTVPEDETSELQGTFTVPK
DGAESLQGTFVVIPGSGESLQGTFVKETGNCSGEQSASLQGTFVKESDGNGSIGQSQSTS
LQRTFVKESDDAPNDDQEQSASLQGTFVKDVDASENDRTDSTLRVLQMNSHVQNELID
GTLTEELPTAQNTMEGGEKRRKSKKVDLGEREKRMANVSSFINSMIDDAPSPGAHFFKN
SRKKMRPTVATPPRTKFGSRLSTESDKEKTIEMLSIAEEATMESHGSSFADPISVNGSRVSP
IPEKDETQPMEICTTPKSSRRNNYSLRQSGTCEKSRASEVSMRLSVAPAASETLVSGIVTPS
RNYQKPTASSLAKSKNAAVFSQSEREKANERSKTNKQTPRISASHTPVFKEDSHRETVVT
PVCKDRSISPQGFEEDMQDEPTVYPGPSNDSIGNVEADLHALSVNDRYEAPFDLDFDAD
YGHDNAPESEEEGEEEQEEEEERRDLSPAQRPKPKSSRRRVALLSDSVASVNTPGLDRRS
KRNTRVTNDTVVDDDWDSEDEEDQMRSSRRRNGKNEVGLQLAKRKIIEPNSTADGLRR
STRTRVKPLRSWLGEQAVYVNSPSGGKRLKGVTDVIIKDKRLCKYRTADLKLATEREQK
EKARKKARAAQKREQLAADQSMGRRLDESQDEIVTSDDE
>C_sulstoni_HCP4
MENFRRPCVSRTIVPGRRIIPEVVYRDAGIAEDMSYLAEKTIDESIMNCTTDMEEKMWRK
QGLSEKEIFRRQEERTLKKHAQAKARLEEGKFGAKTFAEMCALSEYSQRESEDEENEVP
VQQKKKEDVFAVPAIPKHVNEKSILMSPSANTKGSGQPGLSCSTPKGNKDLSMRSLRTLD
ISHVVSTDRLDFRKITVQEKTVVIPVHLEEDSSSSLQKTHTIRREESSTQQGTFTVEKSGE
KESLQETFTKEKDHSGSLLGTFVKESNQPESLQGTFVVESGKEQMESLQGTFVKVIEEVP
QEVPQAVKEKTLVQVSEEVSTINETMLDGERRKSKKMGLKERERRVADLTSSMMSSMIE
DIPSPGAHFFKNNRKKMRPTVPTPPRSKVGTSRLSTESDKEKTIDMLSIAEELTIESSGSSF
VDPISVNGSRVSPIPEQNEAMEISVTPKSSRRPISSLRQDPASIEKSRRQGSTLRISMAPLAL
SIHSPMLEQTPRNNYLKPTFSSIAKHKTPAENKELLEAEKLNRRRTPCATPTKENGFVDFE
KQEEQEETEECSQRSAIVEGLEILEDPTLAPSNRSIENMEDQLDALSVTDHRQFDIPEVDY
GEDYPQYDDAPEPLMSEGEEEEPTQRPKKKEPKRRVGLLSDSMMTVNTPGLEDRRTTRS
RRGGRGASNAEGPSYDDSYEQESRGYGRRKKQEVGLQLAKREIIEPEQQTDGLRRSTRT
RVKKLRTWMGEKAIYVNSPSGGKRLVGVTDVFVKDKRLCKYRTADLNLAMERENRER
AQKRARAAERRRQLALDHSFGRRLNESQEDIVTSDDES
>C_afra_HCP4_paralog1

MNDLRNTRFQSTIVPGRRIIPEVIYRDAGIEGEPSYLAERTMDESVVNCTVDMEERMWR
KQGVPEAEIFRRQEERTLQKHAQARARLQKEKFGAKTFGEMCALSEYSARESEDEDDG
ARAQERKKKKENVFAIPTIPRHLNEKSILGPPSVHAKGSGKAGLSCSTPKGNKDLSMQSL
RILDISHVVNTDRLDNRKTTSQEKTVVIPVCVDDESSLQKTHTIQKEESQGGATFTVAKN
GEDGSLQGTFTLEKNGEEGSLQGTFTKGQNQSASLQGTFIKETDPTENQQGTFIVQSAKD
QTESMHGTSVKTTEEDPEVVTQLAEEKTLVQISDDVPTANETIQGGERRKSKKIGLGERE
KRVADLTSSLMSSMIEDIPSPGAHFFKNNRKKMRATVATPPRSKLGANRLSTESDKEKTID
MLSIAEELTIEANGSSYVEPVSVNGSRVSPIPEDNEAVPMEISVTPKSSRRPVSSLRQEAAS
SIEKGRRQASLMRISMNPMTSVHSPIAEQTPSKNYLKPTFSSIAKHKTPSESNELLEAEKM
NRRRTPVSTPTKNRMFEDIDVPEEFKAEEGIQRPAVIEYPEGPEDPTLQTSNRSLGNVADD
MNALSVTDNRPSASSDVQDHHHSCNPSDDYYPDDEAAAPMFEDEEEEEEDRTRKPKKK
QTRRRVALLSDSIATVNTPGLEDRRAARSRRYGGAQSSSVANSPSFDYSYEQEPRYGRKK
KQVEMMLAKRELIAPETPTDGIRRSKRTRVKPLRTWLGETAEYQNSPSGGKRLVGVKEV
IIKDKRLCKYRTADLNLAMQRESREKALKRARAAERRRQLAVEQSFGRRMNESQEDIVT
SDDE
>C_afra_HCP4_paralog2_HCP2L8
MNNSRNTRLKSDIVPGRRIIPDVIYRDAGIEEEPSYLAERTMDESVVNSTRNTRWNSNIVP
GRRIIPDVIYRDSGIRGEPSYLAERTMDESVKSKRIGLGEREKRAADPTRSTMSSMIEDISS
PGVQFFENNREKMRPTVATPRRSNLGANRLSTASDKEKTIDMLSIGGEPTIEANGSSYVEP
VSVNGSRVSPTREDMLHRHVGPVITEVEESPSRDPSVLRDSGRRHNVARSKSTAIERQIKL
IFIEPRFKERPEAMAEVIRLMRKQIDQWERDQDSYGPTEERTKNIRAWKDNRRRFIEQLE
DAKERQERARRERDESSRVVETVRRRGDMTGMNVTAIHNSRSYDQRTHFDSEEEEKEN
GYTSRAPRPLPRSPPRAHQSYHSSLSHQSRRLDASSRDNRSRNDYGSNNVTSSSHQGKPK
PRLRAGKSRVTKSMFRPRRDRALLEIRQYQKSTDLLIQKAPFCRLVQEILREEATTSSSDY
RIRADALTALQEAAEAFLVEMFEGSQLIATHARRVTLMHSDIQLYRRLCLRK
>C_sp49_HCP4
MSNGESSNRIKPGRRMISQEAALFAAGYHEEEASYRAEQTIMSANSTINDTMDLQERAW
KKQGVPDDEIFRRQEAKTLQKHTDHRTALDAGMDGANDFKELLMRPQYAVRRHFHDEE
DEENHHPPKGQKNNVFAIPALPKHMGTPKVNGKAGFGCSTPIGGDVSMRSLRNIDISHVI
NVDQLDANKTTTHSKNVVIDMVATSANSTLQKTHTIRKDSEEESLQKTFVKQSDQSLQG
TFVVAPATSDSTLQKTHTIRRDSETSTALQGTHVVEKDSTALQGIFVVEDKEPAAIQDTFV
VEKEPQKEPGYSESSSALQETFVKHSGQLEQFEQPNDATISHETTQGGRRVKKVDLKERE
NRMADLTTSIMGSMMDTPSPGAHHFKNPRKKLRETVATPPRLTTARMSIESDKERSIHME
SINNDSTRGRESVGSSYVDPVSVNASRMEIVVEEEEREEEKTHVPMDIATTPKSTRNTPRP
WSRATAGESARAGRPSNYADPISQEDTTNLVTPVRNFQKPTISSLSKNRRASARMILETM
SANRSTMKQSDDNTETSRKTSDAPVARKVQESPETADKVPDRQVSQNDSLGDVANHFR
HISFTDPPPAPQEDVDFGDYGQEESDPEEGTSGLPPAFQRDSQKDAESHRTKPSRKGRLG
LLSDSLTSTNTPGVDRRKNQVNDTAVDDWDSEEEEEEPTRRGKKAKQVGLQLAKRRIIE
PEAAPDGLRRSGRTRVKPVRNWLGEQPIYEHSPSGGRRLKGVTTVFVSDKRMCKYRTA
DVTLATEREVESKRKRAKRRAEERRRQLELDQSRGHRMDESQDDIHTSDED
>C_sp25_HCP4
MSRSLSNGIRPGRRMISEEAALFAAGYREEEASYRAEQTIVSANSTINDTIDMQERAWRK
QGISDEEIFRRLEARAVEKHTEHRQALRAAMDGAKNFEELLNRPQYAVRRHAEENEENQ
HSSGVQFAIPALPKHLGATPKAKGQPGFGCSTPIGSDVSMRSLRNIDISHVIHIEQLDPNAP
TTHTKNVVIEVESANSTLPKTHTIRKDGESLQGTFTVAQPAASTSESSLQKTFVKESQEPL
QGTFTVPPAAPSESTLQTTHTIQKDADETAIQGTFVVEKEKDKDDEQAAGDSEQSLQATF

VVEEPPRNHETSQGGNGARRARKVDLKEREGRMADLTSSMMGSLMDTPSPGAHHFKN
PRKKLRETAAPRLAATRLSTESSDQERSIQMLSINEDSTRGRESIGSSYVDPVSVNASQQM
ETVVEEEEAAIATRNTPRPLSPLTAAGESVRAAPRSSIIATTTTADTTGLLTPTRNFQRPTIS
SLSKSRNANAKEILETMSQKKSTAERKEAVEEEKEKEQVVPEVVVNSFRQISFHDPSPSPP
PRSSPPRSPSPVAALEDVDFGDDAADQSDPEEGTSGPPPAFLMLHSHSHRGEAKCQRKPR
IGLLSDSMTSINTPGIDKRRGRVNETVPDEWNSDEEEQEQEEMPRRRKPKQIGLQLAKRR
IIEPEQAPDGVRRSTRTRVKPVRNWLGEKPIYVHSPSGGRRLKGVTDVFVTDKRMCKYR
TADVRLCTEREVASRRKKAKRRAEEKRRQLEMDQSRGRRMDESQDDIHTSDEE
>C_imperialis_HCP4
MRDSDAASGIKPGRRMITQEAALYAAGYDEEASYRAEQTIVSANSTINDTMMMQEREW
RKQGISDAEIFKRQEEMARKKLSDHRAALTAAMDGATNFAELLSRPQYAVRRIPDEENKF
AIPALPKHLGTPKINGKAGFGCSTPIGSKDVSMRALRAIDLSHVIAVEQLDANKTTTHTM
NVVIDMAGTSSAESSLQKTHTIQKEAEDSLQKTHTIQKDSENSSALQGTFVVSPASDSSIQ
KTSEASTAIQGTFVKDSEDVTLQGTFVVEKEAVPAPEDVTIANNTTLGGRRGRKVNLKER
EGRMADLTTSMMGSMAETPSPGANHFKNPRKKLRETVAAMRMSIESDKERSIQMLSINE
ESTRGRESVGSSYVDPVSVNASQMETVVEEREQEVEEKEQEKEHGEAPMEITVTPKSSR
NTPRPSIFSPGEAARGARPSNFAAIAPADLTLLTPTHHFQKPTVSSLSKNRNAGAKEALEA
LTQKKLAKKVPEPAPESPTSPKSVEQPSRLHNNSFGDVTNGLRQISMHDVPPIEDVDLGD
HLERGEGTSGPPPEFQMTSPEIRSTRSSKSKKVGLLSDSMTSLNTPGVDRRRGPETTVVE
EWNSDEEPEDVPRRRNKKQVGLQLAKRRLIEPEAAHDGLRRSTRTRTKPVRNWLGEQPI
YEHSPSGGRRLKGVTTVFVTDKRMCRYRTADVKLATEREHEAMRKRAKRRAEERRRQL
ELDQSRGRRMDESQDDIHTSDEE
>C_japonica_HCP4
HQSRLVKGYRGAKNYHELLAAYNCLPDQPGPPDGDVEAEENVSNAQKKDDSVFAIPSM
PKKKNAGPSCSTPISGKDASMRTLKGLDISQVFSNVESSFTTVKCAGGGGEPSLQKTFTM
RNVEEEEQESIQKTFTVRNGEEESLQKTFVVAEADGGASSSSSSLQKTFTVRNEEEPLQG
TFVVEKEKEAEPETEHKPIPQPQEPPKTHNRPKKTNLEKREGRMADLTSSMMMGIDDTP
SPGARHFKPNARKKLRESLQTPPRGLSNRMSLDSDKDKTISMLSVAESSGRESMGSSYAD
PVSIHHQTSHISSIPEEHEEEQEEVEVEEEAETTVQNSEASVAIKTPPRAPRLTQSSIERRRA
NQSALRISTAPKAAEGTFLTTPTAHNYQKPTFSSLVKSKDRAECTELLTELNKNRTPRRTV
PPVEEPAQELGHVAEDLEALSISVGVLEKEQQEPRPLTEDVDFEDREEPVPDDESVKKML
RRVGLLSDSIATVNTPGHPRAAVHFNDTTAVDQWRSDEDDDEEEEEEGPSRRRQTRRGG
KKQEVGLQLAKRRIIEPEQAPDGIRRSSRVRVKPLRSWLGERLDYAFSPNGTRRLKGVND
VFIKDKRMCKYRTADCRLAMEREQREKARKRERAAKKRARLALDQSQGRRMDESQED
IVTSSDEE

Supplementary Data 3.S6: All KNL-2 sequences used in this study

>C_tribulationis_KNL2
MGDGDVVPARVQNVLDMDVVRLNLWSLQFNSSGVRLEGFVRSEDGNMMQKVHSGIIC
RRMSATMLFDVSGCFYELAGHIDREFQQKLGMPSRVIEEFMNGFPDNWAFLIKSCLPAEP
KSALRPIQAAPEPSKARGEPIVTMPDETIVDNQSAEKERKRKELQKKKEQEERVRIQREQ
RAAAEAAAERKRQEDEEFKRREEEEERKRQEEENDAANYTFRVPQSQNGDAITPIRFTR
GNGQKGANVRSIFEKTPRTKPAGPLASSTPQAPPQQHAPRISIAEAKQPDEPPHPHPKPQP
PTRESIRETQYCSDSEFAVPKLPAPRNNRHASASSQPAPLDFLDEMDALFETANVDHTPGR
GRKSRKPVSRSPSPGRFNSTARDRTGYDRYEPSRQSSSQRYYDDYNNVSRMSGRNHTLG
GHDMRRDESRNSRKRGYNNSPEDYSRRFDDRSRRRDNYESDSRYDSKRSRPRDQSSSSG
RSVRFEEDFPRNRRDESRDSRSYRHYEDHRNRESSGDREDKRKLDAIVRREKELVARLQ
NTQRSSSTLQRSGYSSEDDEMSDEWDRENQEMLDNSMMFGDGLSKKKGKSGGHKTMR
QAKTRYPPKPKPAQKPAQPKKKKKDVEYESDEMNDSIASNRPRRACATPSTPAPKRITWP
KRDLDRLKHTIELKKPTGAEADWAEVTRLLAKDGVDAEVVKQTAIAKLKWKEPSQKTI
QQEEEEKKRRRGATARVKEGVRMHEELREGGNNRAESSQSGVEAVDDYEPDDVAADQS
LLGLQTPIAVKKRGGTRASIMPQPVEDSPVVRGNNSTLNSPRLDQTKAKDVETTLKYVQ
HLSMMNARPSSRANSSYLNKSSSRGGDKKNTSLSVEQGARKALKIINRGRTIHEEDEDD
EDEDITID
>C_sp41_KNL2
MGDGDIVPARVQNVLDMDVVRMNLWSIRFNGSGVTLEGFVRSEDGNMMQKVHSGIICR
RMSATMLFDVSGCFYELAGHIDREFQLKIGMPSRVIDEFMNGFPDNWAFLIKSCSTSEPK
SALRPIQAAPVEPLKSRSEPIVTMPDETVLAESQTAEKERRRKELQKKKENEERLRIQREQ
QEAVKAAEKKSQEEEKDKLQEEERKRQEEEDAANYTFRAPKSQSGEAITPIRFTRGNGQ
KGANVRPIFDKTPVRTKPAGPLASSTPQAPPPQHPHRLPNIETTRPDASVPPVPQPPIRETQ
YCSDSEFFAVPKLPAPKNIRGTSAKPLGFLEEMDALFENVNVDQTPGRVRNPRKVSRSPS
PMRLNSSARNRDSGYDRYEPSRQSLSQRYYDDYTTSRMSGRDDTFGRNDTRRDESRNS
RKRGYNNSPDEYRNRWDDRSRRHDNFESDPRYDSKRSRPRDQSSSSGRSVRFEEDHPRS
RMDSRESMDSRNYRHYEDSRNRQSSGDREDKRKLNDILRREKELMARLQNSQRSSSTF
QRTAYSSEDDEMSDEWDRENQEMLDNSMMFGDGLSQKKRRSGGRQPAKKPNYLQQAK
TKRVPQPKPAPKPAQQKKKKQDSEDERDEMNDSIASNRPRRACVTPSTPAPKRITWPK
RDLDRLKHTIELKKPTGADADWAEVTRLLAKDGVDAEVVKQIAITKLKWKEPTQQTIQ
MEEEEQKRRRGATARVKEGVRMHEELRAGGNDRGNSSQGGVGAVEDYEPDDVAADQS
LLGLQTPIAIKKRGGTRASIMPEPVEDSPMVRGNNSTLNSPRLDQTKAKEVETTLKYVQ
HLSMMNARPNSRANTSYFNKSSSRDKTNTSLSVEQGARKALKIINRGRTIEEEDEDDEDG
DTTID
>C_zanzibari_KNL2
MGDGDVVPARVQNVLDMDIIRLNLWSIQFNGTGVKLEGFVRSEDGNMMQKVHSGIICR
RMSATMLFDVSGCFYELAGHIDREYQLKIGMPSRVIDEFMNGFPDNWAFLINSCSEPKSA
MRPIQAAPREPLRTRGEPIVTMPDETNIESQSSEKERKRKEREEQQKWEHEERLRARKEK
EQREAAEIAEKNRRREREIQKQREDEEKGRREEEEEAANYTFRAPKSQNGEAITPIRFTRG
NGQKGANIRPIFDSTPVRTKPAGPLASSTPQAPPPQNRHSNTENRQPDALPTAGPAPPVQK
PNRETQYCSDADLFAVPRLPAPRNNQNVSSSAQSAPLDFLDEMDALFESATVDKTPGRV
KKVVRVNRSPSPDRFSSRDRDSGYDRYDSSRYSHSQRYNDEHHMSRMSGRNDTFRRND
GWRDESRNSRKRGYNNSPEYTRGWDDRSRHRDNYESESRYDSKRSRPRDQSSSSGRSV
RFEDDHPGNHRDSRDPRNYRDYEDDRNRESSGDRADRRKLNDILRRERELEARLRNTQ

RSSTARRHAYTSEDDEMQDEWDRENQEMLDNSMLFGDGISQKKRRSGGRQPAKNSQSG
RSQQPKPARKPAQPKKKKERSDETDELNDSIASNRPRRACVTPSTPAPKRITWPKRDLDR
LKRTIDLKKPTGADADWDEVTRLLAKDGVDREIVKQTAIMKLKWKEPSQETVQQEEEE
KTRRRGAAAMVKEGVRMHQELREGGDNKRGDDLQGGVEAVDDYQPDDVAADQSLLA
LQTPIAVKKRGGTRASIMPQPVEDSPVVKGNNSTLNSPRLNQTKVKEVETTLKYVQHIS
MMNARPSSRANTSCLNKSSSRGGGSKNTSLSVEQGARKALKIINRGRTIQEEDEDDDDES
EAGDITIE
>C_sinica_KNL2
MGDGDVVPARVQNVLDMEIVRLNLWSIQFNGSGVKLEGFVRSEDGNMMQKVHSGTIC
RRMSATMLFDVSGCFYELAGHIDREYQLKIGMPSRVIDEFMNGFPDNWAFLIKSCSNPEP
KSALRPIKAAPKMPLRSRGEPIVTMPDETVMAESQAAEKDRKRKEHEEQQKRDDEKRF
RAEKEKEQREAADTAAKKRQEEEDRRRREEEDANYTLRAPESHNGEPVTPIRFIRGNGQ
KGANIRPIFEKTPVRTKPTGPLASSTPQAPPPPPTRRTSNIENKQPEASTSPKPQPPVQKPV
RETQYCSDTEFFAVPKLPASRNKHVPASDHAAPLAFLDEMDALFNGADVDHTPGRERKP
KKVSRSPSPERFDYSSRDRDSGYSRYDSSRYSHNQRYNDDYNMSRMSGRTDTSRRNDG
RRDESRMSRKRGYNNSPDEYSRRYDDRSRRQDDYDSGSRYDSKRSRPREQSSSSGRSVR
FEEDYSRHRQGSRDSRDPRDYRDYEDHRNRTSSGDREDTRKLNDILRRERELMARLQNS
QRSSSTVQRTAYSSDEDEMADESSLWDRENQEMLDNSMMFGDGLSQKKRRGGGRHPA
KQAKTRQPAQPKPARKPAQPKKKNRADETDDLNDSIASNRPRRACATPSTPAPKRITWPK
RDLDRLRHTIELKKPTGADADWAEVTRLLAKDGVDSEIVKQTAITKLKWKEPSQETIQM
EEEEKKRRKGAAARVKEGVRMHEELREGGDKRGDNSLGGVEAVEDYEPDDVAADQSL
LALQTPIAVKKRGGTRASIMPQPVEDSPVVRGGNNSTFNSPRLDQTKAKEVETTLKYVQ
HLSMMNARPSSRANTSYHNKSSSRGGGSKNTSLSVEQGARKALKIINRGRTIHEEEEDD
DEEGDTTIN
>C_nigoni_KNL2
MNDLFFVNYGSKFMNDYAQKKKASSHIEDYSLPEKKLRKVQEIVDKSPKCLFAKRAPPH
THFAANFSNLNSEGIKTVFLDIHLDFTISTPYFCYKASFCKKKMGDRNVVPARVQNVLD
MDIVRLNLWSIQFVDSGVILEGFVRSEDGNMMQKVRSGMICKRMTATMLFDFSGCFFEL
SGQIDREYQQKMGMPSRVIDEFTSGFPENWVSLIKSFLPVNPISAVKHIQPAPREPLRPISE
PIVTLPDETTLESEKDQKRREKEERLEKKRLEEERLEKERREQKLRAQKEKERREAVAAA
EAEKKRLEEEEDAANYTIRVPKSQSGEAITPIRFTRGNGQKGANVRPIFDKTPVREKTAGP
LASSTPQAPPPQQRLSNVEKKPASPEPQPRNPQRDREIQYASDSDLFAVPKLPAPKTTKPST
TSEASSGGVLDFLDEMDTLFDTVVVEQTPTRDRRPPRRYSRSPSPRRRQHSSSRDMGKG
YDNFESSRYSQRYNDNYNTSRMSRRDDTFRRNDERRDESRMSRKRVYNNSPDDFEYNN
RYDDRSRRPDYYDSDSRYDSKRSRPRETSSSSGRSVRFEDDYRRNHRDPRDRSDSKDYR
NYDESRRNPEDRREEKRKLNDILRREEELVTRLQNRKKPSASYRREPSSDEDDTADEWD
RENQEILDNSMMFGDGLSQKKRRSAGRPSKPSKKERQVQPKPVRKPPQPKKKQKSPPDE
LNDSIASNRPRRACVTPSTPAPKRIVWPKRDLDRLKHTIGLKKPTGSDADWAEVTRLLAK
EGVDAEVVKQVAITRLKWKEPAQDPETIQREEEEEKKRRRGVAARVKEGVKLHEELRQP
GVKRGDNSQTGVEAMEDYEPDDVAADQSLLALQTPVGAKRKGGTRASIMPEPVEDSPL
VRRNNSTFNSPRLDQTKAKEVETTLKYVQHLSMMNARPSSRANTSYYNKSSSRGGGSK
NTSMSLEQGTRKALKIINRGRTIHEEDEDEDDDDDDEDQYEDGVVY
>C_briggsae_KNL2
MGDRNVVPVRVQNVLDMDIVRLNLWSIQFVDSGVILEGFVRSEDGNMMQKVRSGMIC
RKRSETKKQEERLGKKRFEEERLEKEQREQKLRAEKEKERREAVAAAEAEKKRLEEEED
AANYTLRVPQSQSGEAITPIRFTRGNGQKGANVRPIFDKTPVRGKPVGPLASSTPQAPPPQ

QRLSNVEKKPASPEPRPRIPQRDRDVQYASDADLFAVPKLPAPKTTKPSTTSEPSTGGGFD
FLDEMDELFDTVVVEQTPTRNRRPPGWYSRSPSPRRRQHSPPRDSFESSRYSQRYNDNY
NTSRMSRRDDTFRRNDERRDEFRMNRKRVYVCDRGLVRKRKYVFQNNSPDDFEYSNR
YDDRSRRPDYYDSDSLYDSKRSRPRETSSSSGRSVRFEDDYRRNHCDPRNRSDSRDYRS
YDESRRKPEDHREDKRKLNDILRREQELVSRLQNRKKSSVSYRPEPSSDEDDMADEWDR
ENQEILDNSMMFGDGLSQKKRRSAGRPSKPSKKERQVQPKPVRKPPQPKKKKQKSPPDE
LNDSIASNRPRRACVTPSTPAPKRIVWPKRDLDRLKHTIGLKKPAGSDADWAEVTRLLAK
EGVDAEVVKQVAITRLKWKEPAQDPETIQREEEEEKKRRRGVAARVKEGVKLHEELRQP
GVKRADNSQTGVEAMEDYEPDDVAADQSLLALQTPVGAKRKGGTRASIMPEPVEDSPL
VRRNNSTLNSPRLDQTKAKEVETTLKYVQHLSMMNAHPSSRANTSYHNKSSSRGGGSK
NTSMSLDQGTRKALKIINRGRTIHEEDEDDEEDDEDQYEDAVVY
>C_remanei_KNL2
KRQEEEDADYTFRAPQSQNGEPITPIRFNRGNGQRQGVTRSVFERTPQRGQSGPLAASTP
QAPPPPPPQQRLSNIENREPPPRVVAPPSPVRQPPAPQPPIREPQFANDDDLFAVPKIPPPKIP
RGSAGNSGGNIDFLDEMDALFDTVYIDKTPKRDVKPKRPSSPSPERRRYSPMPRDREMG
YNDDFESSRRGGRYPSESSNMSRMSRRDETNRRNDGGMGRDESRMSRKRGYDQSPDD
MEYQRRREDHYRRPDYYPPRDSRQDSKRYRPRENSSSSGRSASVRFADDYQRNRGDSR
DPRDFRDPRDSFSRDPRFYYENNQRGESSKDRDTRKLNNILRQERELVARLQNIKSNTTT
NTNTTRRVTYSSEDDEMADEWERENQEIMDNSMMFGDGISKKGRRSGPGRPPQRKPKE
QPKPKSQSQPKRRNQNQSKPRRSQYDPVETDDLNDSIASNRPRRACVTPSPVVPKKIIWP
KKDLDRLKRTIELKKPTGSEADWTEIARLLAK
>C_latens_KNL2
MPSRVIDEFTNGFPENWGFLIKSCFGNESRSAMRPIQAAPREPLRQRNEPIVTLADETELT
NRKNSDNDSESENNRKRREREEEAERERQERYRADEMERERIAAEKKRQEEEDADYTF
RAPQSQNGEPITPIRFKRGNGQRQGVTRSVFEKTPQRGQSGPLAASTPQAPPPPPQQRLS
NIENREPPPRVAAPPSPIRQPPAPQPPIREPQFANDDDLFAVPKIPPPKIPRESVGNGGGSIDF
LDEMDALFDTVYIDKTPKRDVKPKRPLSPSPERRRYSPMPRDRELGYNDEFESSRRGGR
YESSNMSRMSRRDDTYRRNDGGMGRDESRMSRKRGHNYSPDDMEYERRREDHYHRP
DYYPDSRYDSKRYRPRENSSSSGRSASVRFADDYQRKRGELRDPRDFEDSFSRDPRFYHE
SNRKRESSQDREDTRKLNDILRREKELMARLQNNKTTTSKNTTTRRVTYSSEEDEMADE
WERENQEIMDNSMMFGDGISKKGRRSGPGRPPQRKPKEQSKPKSQSQPKPRNKNQSKP
RRSQHDPVETDDLNDSIASNRPRRACVTPSPVVPKKIIWPKKDLDRLKRTIELTKPTGSEA
DWTEVARLLAKNGCDGEAVKQTAITKLKWKEPVERDEETIQREEEEEEETKRRRGVNA
KVKEGVKLQEELRRGATHQKRAEDVRVEVTAEEIEPDDLAADQSLLAMATPVAVKKRG
GTRASIMPMPVEDSPIVRGSRNNSTFNSPTLNQTKAKEVETTLRYVHQLSMVQARPASR
NNTSYMNKSSTRGGGSKNTSMSLEQGVKKAMKIINRGTTIHEDDEEEEESESEEEDNSE
DLEEH
>C_sp51_KNL2
MGDTDILPTRVQNVLEMEIVRLNLWSIKFNALNIKLEGYVRSEDGNMMQKVCSEIICKR
MNSTMLFDVSGCFYELSGQIDREYQSRMGMPSRVIDEFVNGFPENWAFLIKSCLSMGQR
SALRPIQDTPREPIRTRVEPIVTLADETELVENERKKAESEKDKKRKEREEDRARENERNL
AAQREKERKDTVAAEERKRKKEEEEAHAVAERKRQEEEEAANYTLRAPQSQTGEAITPI
RFTRGNGGNGAHVKSIFQKTPVRGKSNGPIASSTPQQTRVLPTLEKPVETTVSKSTKSKSP
PPRSIRQTEYASDADLFAVPRNPAPKNNRKPAPTTSSDIGFDFLDQMDSLFDNPDFEQTPT
RDRKPRRQFSPSPEPRHRSYSRDRDRYAHLESTRYKQRYTDDYDMSRISRRDVPFRRND
GGRDESRMGRKRGYYGSPDGSEYGRRYDDRDRHMDYYEPDFKYDMKRSRPRETSSSS

GRSVRFENDYRRRREDSREPNSSRNHREYDDYRGRTDSEDNRRLEFLEKRENELMARL
QNHKKPSSSSSRRLTNSSDDEMADMWEQENQEILDNSMMFGDGISKKKRRSGGGRQPK
KQPTRKTQPKPTRKPAEPRKKTQRSPVQRDELNDSIASNRPRRACATPSTPAPKRITWPKR
DLDKLKHVIQLKKPTGADADWIEVARLLAKDGVDSELAKQTAVSKLKWVEPVQHEERI
DEEEEENKKRRRGAAAKIKEGVRMHEELMSGRQEDDIRGGGEIMEDYQPEDMAADQSL
LALGTPLAAKKKGGTRGSIMPKPVEDSPVTRGGNSTYNSPRLDRTKAKEVETTLRYVQH
LSNMQARPSSVSNTSTLNKSSGRTSKNTSMSLEQGTRKALKIINRGRTIHEEDEEESGDEE
SELSEEEEEGESFNY
>C_sp44_KNL2
MGDTDILPARVQNVLDMEIVRLNLWSIKFNASNIKLEGYVRSDDGNMMQKVCSEIICKR
MNATMLFDVSGRFFELSGQIDREYQAKMGMPNRVIDEFVNGFPENWAYLINSCLTISQRS
ALRPIQAKEPIRSKGGPIVTLADETELVENEQRNSDSEKEKKKKEGEQRASENKRNQEAQ
RERERKEALAAAERERKRHEEDAERKRRKEEEDAANYTFRVPESQLGEPLTPIRFTRGNG
GNGANIKPIFEQTPVRGKTTGPLASSTPQAPAPQQPRVLKDLEETVKPTAPQSSKAKPPPQ
KPQTEYASDSDLFAVPRPPADKNSRTTTSASSSDIGLGFLDQMDLLFDSAAVERTPTRHPK
PRRIAVSPPNPRYRSHSKDRNRYGDFNSTRYTQRPSDDYDMGHMSRRDATFRRNDGGRD
ESRMSRKRGYYGSPDESDYGRRRDDRDRRDNYYEHEYDMKRSRPRETSSSSGKSVRFE
DNYRRRREDPRETNYSRSYRDFDDFNTRGTSGDREDNRKLNDILRRENEVKARLQNHH
KSSFSRRHVQSSDEESDDMADEWDRENQELMDNSMLFGDGIPKKKRGSDGGRPAKKQP
VRKPQPKPTRKPAEPRKRAPRSPVDTDELNDSIASNRPRRACATPSTPAPKRITWPKRDLD
KLKHVIQLKKPTEAEADWVEVARLLAKEGVDSELVKQTAISKLKWKPPVQEERNDDED
DEDKRRRRGAAAKIKEGVRMHEELMSGRNPVDDIRSGVEFVDDYQPEDVAADQSLLA
MGTPLAAKTKGGTRGSIVPKPVEDSPIIRGRNSNYNSPRLDQTKAKEVETTLRYVQHLSN
IQARPSSRANTSTMKKTSSHASKNTSMTLEQGTRKAMKIMNRGRTIREEDEEEDESDRE
DESEQSGEEEEEDDIYY
>C_sp48_KNL2
MGDTDILPARVQNVLDMEIIRLNLWSIKFNASNIKLEGYVRSEDGNMMQKVCSEMICKR
MNATMLFDVSGRFYELSGQIDREFQAKMGMPGRVIDEFVNGFPENWAYLINSCLTVGQR
SALRPIQSAPREPIRSRAEPIVTLADETELVENERKNSEAEKEKKKNERDEQRARENERNQ
EAQRERERKEALAAAAETKRRKEEEEERAEAERKRREEEEAANYTFRAPESQQGEPITPI
RFKRGNGANGGYKSIFEKTPVRGKSNGPLASSTPQAPPPQQPRILSSLEREKKTEPDAPKS
PIADRTVQKPIRNTEYADDADLFAVPKLPPIRNNRPSAAAPSSEIGYDFLDQMDSLFDTVV
IDQTPNRNRMPRRAASSSMEERLRSSPPRNRDRYDDRESTRYSQRYGDDYNTSRMSRRD
ATFRRNDGGRDESRMSRKRGYYGSPDEFDHRRRDDRDRRGDYYEPDYKYDMKRSRPR
ENSSSSGRSVRFEDERFGDYRRHREDSRERKYSRNHREYDDNRGRRGSSGEDDRKLNDI
LRRENELMARLQNHRPSSSRRDASSSGEDEDDLANEWDRENQEILDNSMMFGDGLPKK
SRRSAGGKIGRQPKKQPTRNPPPKPARKPEPRKKAPKSPVETDDLNDSIASHRPRRACAT
PSTPAPKKITWPKRDLDKLKHVIQLKKPTADDADWAEVTRLLAKDGVDSEVVKQTAISK
LKWKEPVQEQERKEEEEEEERKRRRGAAAKIKEGVRMHEEMRRGRREEDTQMSADSM
EDYQPEDVAADQSLLALGTPLVAKKKGGTRGSIMPKPVEDSPITRGRNSTLNSPKLDQTR
VKEVETTLRYVQHLSNMQARPSSRANTSTVSKSGRGSKNTSMSLDQGARKAMKIINRG
RTIDEEDEDEESNEEEEDHSEDEAFDY
>C_brenneri_KNL2_paralog1
MGDTDILPARVQNVLDMEIIRLNLWSIQFNASNIKLEGYVRSEDGNMMQKMCSEMICKR
MNATMLFDVSGRFYELSGQIDREFQAKMGMPGRVIDEFVNGFPENWAYLINSCLTVGQR
SALRPIQNAPREPIRSRAEPIVTLADETELVENERKNSEAEKEKKKKELEEQRAKENERNL

EAQRERERKEALAAAERKRREEEEEAANYTLRAPESQPGQPITPIRFRRGNGGNGGHMK
PVFEKTPVRGKSNGPLASSTPQAPPPQQPRILSSLEKPKSPIAERSAQKPMRRTEYADDAD
LFAVPKLPPIRNSRPSASAPSSDMGFDFLDQMDSLFDTVVIDQTPTRNRMPRRAASSSME
SRIRSPPRDRERYDDRESTRYSQRYGDDFNMSRVSRRDATFRRNDGGRDESRMSRKRGY
VSTPDEFDHRRRDDRDRRGDYYEPDYKYDMKRSRPRENSSSSGRSVRFEDERFGDYRR
HREDSREPKYSRNHREYDNYRGGRGSSGEDDRKLNDILRRENELMARLQNHRPSSSRR
DAYSSDEDQDDLANEWDRENQEIMDDSMMFGDGLPKKQRRSGGKIGRPPKKQRTRKPP
PKPARKPEPRKKAPRSPVETDELNDSIASHRPRRACATPSTPAPKKITWPKRDLDKLKHVI
QLKKPTAAEADWAEVARLLAKDGVDSEIVKQAAISKLKWKEPVQEQEKKEEEEEEERK
QRRGAAAKIKEGVRMHEEMRKGRRENETQMSAESMEDYQPEDVAADQSLLALGTPLV
AKKKGGTRGSIMPKPVEDSPITRGRNSTFNSPRLDQTKVKEVETTLRYVQHLSNMQARP
NSRANTSTVSKSGRGSKNTSMSLEQGARKAMKIINRGRTIDEEDEDEESNGEEEEDHSED
EAFDY
>C_brenneri_KNL2_paralog2
IDREFQQNGMPVRSYDEFVNGFPETGLILLILASQEPIRSRAEPIVTLADETELVENERKNS
EAEKEKKKKELEEQRAKENERNLEAQRERERKEALAAAERRRRQEEEERADAERKRRE
EEEEAANYTLRAPESQPGQPITPIRFRRGNGGNGGHMKPVFEKTPVVSEGQ
>C_wallacei_KNL2
MGDNHVLPARVQNVLDMEIIRLNLWSIKFNASNIKLEGYVRSEDGCSMQKVCSEIICKR
MNATMLFDVSGRFFELAGQIDREFQLKMGMPSRVIDEFVNGFPENWGFLINSCLTMNQM
SVPRPIQAAPREPLRSRNEPIVTLADETELADSGRKTSESERDKKRREREEQRMKENEKR
LETEKEKERQKAQAAAAAEKKRQEEEEDAANYTLRGFQMESGDPITPIRFKRGQANGV
HAKPTFVQTPVRGKPSGPLASSTPQAPLPEHPRRSLDLGKRSTPLASPSPKVQQPVSKPAR
ETEYASDSELFAVPKLPAPKAARPSATSSAPSDTAFDFLDDMDVFFDTAVIDQTPVRARKQ
PRRVSPDVRRDRSSSRDPERFGDYDSTRNNQRYNNDYDMSRMSRRDATFRRNDERRDD
SRMNRKRGYYNSPDDYDNRRSDYYESDSRYDTKRSRHRETSSSSGRSVRFQDDHRRSE
NDDYRRREGFRDPEYQRNYRDFEERRVRGSSGDEDKKKLNDILRREKELMARLQNNQK
PSSSRRVSCSSDDDIDMADEWDRENQEIMDNSMMFGDGISKQKRRSGGSRQEKKKQPT
RTVQPKPKRVASKKKSPVPSDDLNDSIASNRPRRACVTPSTPAPKRITWPKRDLDKLQHV
IDLKKPSAEDADWVEVTRLLAKEGVESEVVKQIAISRLKWKEPVKNKNNEKKEEEEDET
RRRRGAVARVKEGVRMHEELIRGRDDEERDVQVESMEDYQPEDVAADQSLLALGTPLA
AKKKGGTRGSIMPKPVEDSPMARRQNSSFNSPRLDQTKAKEMETTLRYVQQISMMQAR
PSSRANKSKLNKTSARGSKNNSMSVEQGTRKALKIINRGRTIHEESEEEEEESDDDEFDE
REEGDNSIY
>C_tropicalis_KNL2
MEDNHILPARVQNVLDMEIVRLNLWSIKFNSSSIKLEGYVRSEDGCSMQKICSEIICKRMN
ATMLFDVSGRFFELAGQIDREFQSKMGMPSRIIDEFVNGFPENWGFLINSCLTMNQMSVP
RPIQAAPREPLRSRNEPIVTLADETELNDGGRKNSESEKDKKQKEREQQRMKENERKLQ
AEKDKKDKEAVAAAEKKRQEEEDAANYTLREFQTGSGDPITPIRFKRGTQSKGANNIKP
TFLQTPVRGKPNAPLASSTPQAPPPEHPRRSLDLGKPPIQLASPSPKVQKPETVRETQYAS
DSDLFAVPRLPAAKNNRPPPSSSSETVLDFFDDMDVFFDTAVIEKTPVVARRQRRLSPEAR
RDRSSSRDPERFGDYDSTRYTQRYNNDYDMSRMSRRDAVFRRNDERRDDSRMSRKRGY
YSSPENDFDRRRPEYYESDSRYDTKRPRPRENSSSSGRSVRFHEDHMGRGDEFRGSRYER
GYREFDDQRGRGSSADEDKRKLNAILIREKELMARLQNKHKPSASRRVSYSSDEDSDDM
AAEWERDNQEIMDNSMLFGDGIMQQKRRSARGHPAKKKQPIRTEQHKPAQKRAPPKKK
SPAPRDDLNDSIASNRPRRACATPSTPAPKRITWRKRDLDKLQHVISLKKPTAADVDWVE

VTRLLAKEGVEPEVVKQAAITKLKWKEPVRNDENMREEEDESRRRRGVVARVKEGVR
MHEELMNGGNEVEEDLQVQSVEDYQPEDVAADQSLLALGTPLLAKKKGGTRGSIIPKP
VEDSPIARGKNSSFNSPRLDQTKAKEMETTLKYVQQISMMQARPSFRANKSQMNNSTAR
GSKNTSMSVEQGTRKALKIINRGRAIQEESEEDEEDDDDDEFEEQEEGDNSIY
>C_doughertyi_KNL2
MGDTNILPSRVQNVLDMEIVRLNLWSINFNASNIKLEGYVRTDDGSMMRKVCSEIICKR
MNSTMLFDVSGRFYELAGQIDKEYQLKMGMPSRVIDEFSNGFPENWAFLIKSCLSLPPRS
TLRPIQDVPREPLRSRAEQPIVTLADETELVENERKNSDSERDKKRRDREEQRVRENEKR
VEDQREKEHQEAIAAAAAEKKRKEEEESAADYTFRVPGTDLGNPITPIRFTRGGQLNGHI
KSVFQQTPVRGKSSGPLASSTPQAPPPENPRRLSHLERASVPAAPLPEKDKTPAQRPVRST
DYASDSDLFAVPQLPPAKNIRPSVSSDSAFDFLDQMDSLFDSAVIEQTPGRDQKARKIAPP
SPDARRRFSSREPERRSDYESSRYINRYNDDYNMSRMSRRDEPWNRRHDEGRDESRMSR
KRGHFNSPPHDSDYGRRDNRGRRPEYYESNSKYDTKRYRPRETSSSSGRSVRFEDDYKR
RDDFREPSYSRSYRMYDDRRGRESSRDLEDKNKLDAILRREKELVDRLENTSKPSSSRRH
DLYSSDEDPIDLADEWDRENQEIMDNSMMFGDGISKKKRRSGAGRPQKKQPTGSTQPKP
ARKPAEPRKKARRSPPPADNLNDSIASNRPRRACVTPSTPAPKRITWPKRDLDKLKHVIEL
KKPTGSDADWAEVARLLAKEGVEAEVVKQAAITKLKWKEPVQIDNKEQEEEEEKKRRR
GVAARVKEGVKMREELMSGKQEEDRALRVDSMEDYQPEDVAADQSLLTMGTPLAAKK
KGGTRGSIMPKPVEDSPMSRGRNSTFNSPRLDQTKAKEVETTLKYVQHLSMMQGRPSS
RANKSYLDKSSRGSKNTSLSLEQGARKALKIINRGTTIHEDDEDDDEEEDDDDEFEDQEV
GDTSNY
>C_sp54_KNL2_paralog1
MGDTGIIPFRRKDVLERVEAVLNMEIVRLKLWSIKFNTSVIMLEGFVRSEEGTMIQKVCSE
NICKRISSTVLFDESGRFFELAGQIDREYQQKMGMPSRMIDEFINGFPENWADLINACVL
NQQRSALRPIQAAPKEPLRTRAEPIVTLADETELIGEESKKDMKRREREEQREREQEKKL
ATEKERKRQEAAAVEKKRREEEEAEEAANFTFRVPQSQSGEAITPIRFKRGNGKKGENTR
TIFDKTPVRQNSSGPLASSTPQAPPPQHPRRLSNIENNQPLRSPSQKTQLAASIPERETKFA
SDADLFAVPKLPPSKTTRPSTTSAVSSEGALDFFDEIDALFDTAIIEQTPTRDRKSRKPIATS
PSPDPRRRSLSRDRKSERYGSYESSRYTQRYNDDYDNSRMSRRDVTFKRHEGGRDDSR
MSRKRGYYRSPDDFEYSRRRDDREKDYYAYDSRHDTKRSRPRETSSSSGRSVRFEDDYR
RRGGEFRESKDSRDFRDYDDHRNRKNSGEHEDKKKLKDILRREKELLARLQKSQKSSSS
RRDSYSSEEETFDMADEWERENQEIMDNSMMFGDGLSKKKNRRSGGGGAAKKQLART
SQSKPHRKPAQPRKKTLRDPLETDDLNDSIASHRPRRACVTPSTPAPKKISWPRRDLDRL
KHVIELMKPTTTDKYWIEVTRLLAKEGVDPEAVKQIAVAKLKWKEPVQNEEVLKKDEEE
EKKRRRGAVARIKEGVKMHEELREGGDKKADDLESGVEAMEDYQPEDVAADQSLLAL
GTPIAAKKKGGTRASIMPKPVEDSPMAARGSNSTFNSPILDQTKAKEVETTLKYVQHLS
MMQARPSSRANKSYLNKSSTRGSRNTSMSLEQGTRKAMKIINRGRTIHEDDEEDDEEDE
DDEFDEQDGNTSIY
>C_sp54_KNL2_paralog2
MSEDPDEFLNLRSKVTDPWSVELIGLDAIQQQLDVMGENVAEVILKAEQSFNKKSLPFY
EKRKKFTSKIDNFWKKAFLNHHFLSKAIPEEQEDLIEVLRDLEVQEFDDLRSGFKIIMTFD
VNEFFENKVITKSYHLKSKPPKTQITEIQWKENQKPHMTSEDGDTTITFLEWLTHAVPPK
SDEIAHVIKDDLFINPLQYYMKPDIQENGQPIRRIPKINSSMMEVDDTAEGEDSDEKEEEA
TNCTYRVPQSQNGEAITPIRFKRGNGKKRENTRTIFDQIPMRGNSSGPLASSTPQAPPPQQ
PRRFSNIENNRPAGPIPERETQFASDADMFAVSKRAPSKTTSSTSTSAVLFEGEGAYHFLDE
VDDFFDTLVIETRERKPQRRRIATSPSPDPRRRRSSSRDRERYGSYESSSRYTRRYNDDDY

DDSRMSRRDIATFKRRHEGGRDNSRMSRKREYYRSPDDFEYRRRRNDDRDKDYYNAY
DSRHIAKRSRLGETSSSSGRSVRFEDDYRRSKDPRNFRDHRNRTISEKQLLARLQELRKPS
TSSHRDSYSSEEEETFNMADEWERENQEIMDNSMMFDDGLSKKKIKRRSGRGAVKKQL
SRTSQSKKPHRKPAEPRKKTIRDPLETDDLNDSIASHRPRRSCVTPSTNPVPKKISWPKRD
LDRLKHVIELMKPTTSDKYWIEVTRLLAKEGVDSEAVKQIAITKLKWEEHVQNEEMMK
MDEEEEKKRRRGAVARIKEGVKMHEELREGGDKKADDLKESGVEAMEDYQPEDVAAD
QSLLALGTPIAEKKKGGTRASIMPKPVEDSPMAARCSNSTFNSPRLDQTKAKEVETTLK
YVQHLSMMKARPSSSKTNKSYLNKSSTRGSRNTSMSLEQGTRRAMKIINRDRTIHEDDE
EDEADEDYEFEEQDGDTSIY
>C_inpoinata_KNL2_paralog1
MGDTNILPARVQNVLDMEIIRLNLWSIKFNSSNIKLEGFVKNEDGTAMQKVCSEIICKRM
NSVMLFDVSGRFFELVGQIDREYQQKMGMPSRIVDEFLNGFPENWPYLIKSCISVDSRSS
LRPIQAAPREPLRSRAEPIITLADETEVVVGDNKNSDTGNGRKHREQVERNSNDQVQMP
REANMKNNIQEEEDDIANYTMRAPVSQNGEAITPIRFTRGNGKKGANTRSIFESTPARGT
SSGPLQSSIASVPPPQQSRRLSTNQNSQLQSQCQKIQVAPSQNEKPPFREPQFSNDTDLFAV
PRLQENKNNRLPSLSAASSDGAHDFFDEMDTLFDTANIEHTPARDRNIIRRSRVASPSPPL
RRYRTSSRDCERDIYEPTRHSNRPDNGFNINRNRSDAFRKNEGRNDGSRMSRKRGYFYS
PDKYRDNFSRRIKNIDYEQMHDNKHSRERENSPYSGRSIQFEDVYRKRNYGDYDDFRKR
SRDEEEDRKLNEILRREKEMMARLQQHRKSVSSFPKSYLSDDDEFDMADQWERENQEL
LENSMMFGDGLPKKRKMSRDCKSERLKNPKESVKKPVQSQKSNRVKRLETDNANDSIA
LSRPRRLCATPSTPAPKKIVWPKRDLDRLKRIIELKKPTAANADWAEVARLLAKDGVELE
AAKQTAITKLKWKEAVQNEELLNCEEEDKKRRRGLAARVKEGVKMHEEIREGGENKGG
DLQNGVESMEEYRPEDMAADQSLLAQKTPIVVKRKGGTRASILPKPVEDSPLARGSNSA
LNSPKLDQTKAKEVETTLKYVHHLSQLHANPTCRGNKSFANKSLPGGRKNTSMSLEQG
TRKAMKIINRGQTFYEDEEDEDDECSENDEDTSMY
>C_inopinata_KNL2_paralog2
MGDTNIVPARVQNVLDMEIVRLNLWSIKFNSSNIKLEGFVKNEDGTAMQKVCSEIICRRM
NSVMLFDVSGRFFELAGQIDREHQQKLGMPRRIIDEFLNGFPENWPYLIKSCISVDSRSSL
RPIQAAPREPLRSRVEPIITLADETEVVVGDNKNSDTGNGRKHREQVERNSNDQVQMPR
EANMKNNIQEEEDDIANYTMRAPVSLNGEAITPLRFTRGNRKKGANARSTFESTPAEGTS
SGPLLSSITSGPPSQQSRQLSINQNSQLPSQCQKIRVAPSQNKKPPFREPQFSNQDLFDEMD
SLFDTANIEHTPVRERSINRSRVASPSPPRRRYRTSSRDCGRDIYEHTRYSSRSDNGFNISG
NRSDAFRKKEERNDESRISRKRGYFISPDKYRENFCRRIKNIDNDQMPDNTHSREGENSL
YSGRSARFEDVHHKRDYGEYDDFRKQSRDEGKDRKLNEILRREKELMARLEQSRKSAS
SFQKSYSSDDDELHMTDQWERENQELVENSMLGDDLSKKRKVSRDGKTKQSERLKNH
NESVKKPVQSEKRNRVKRLETDNANDCIALSRPRRLCATPSTPAPKKIVWPKRDLDRLKR
IIELKKPTAADADWAEVARLLAKNGVELEAAKQAAITKLKWKEDVQNEELLNCEEEDK
KRRRGLAARVKEGVKMHEEIREGGGLQNEVESMEEYRPEDQSLLALKAPTVVKRKGGT
RASVLPKPVEDSPLARGNNSALKSSKLDQAKAKELETSLKYVHHLSQMHANPSSRGNK
SHANKSLPCGKKNTSMPLEQGTRKSMKIINRGRTFCEDEKDDSTDNDEDT
>C_elegans_KNL2
MGDTEIVPLRVQNVLDSEIIRLNLWSMKFNATSFKLEGFVRNEEGTMMQKVCSEFICRRF
TSTLLFDVSGRFFDLVGQIDREYQQKMGMPSRIIDEFSNGIPENWADLIYSCMSANQRSA
LRPIQQAPKEPIRTRTEPIVTLADETELTGGCQKNSENEKERNRREREEQQTKERERRLEE
EKQRRDAEAEAERRRKEEEELEEANYTLRAPKSQNGEPITPIRFTRGHDNGGAKKVFIFE
QTPVRKQGPIASSTPQQKQRLADGANNQIPPTQKSQDSVQAVQPPPPRPAARNAQFASDA

DLFAVPKAPPSKSVRNLAASNVDIFADVDSVLDTFHFESTPGRVRKPGRRNVSSPSPEPRH
RSSSRDGYEQSRYSQRYEHDNSRWSRHNATYRRHEDESRMSRKRSIVRDDFEYSRRHDD
GARRRDYYDADIQGDSKRYRGRDASSSSGRSVRFEEEHRRHGDEYRDPRGPRDYNDYG
RRRNHANSRSGEDEEKLNAIVRREKELRNRLQKSQKASSSSYRHRSNSSDAEESLNEWD
IENQELLDNSMMFGDGIPKRSNARKDKFVKKQATRSKPANSTKSPAQARKKKRASLEDN
RDLNDSIACNRPRRSCVTPVAKKITWRKQDLDRLKRVIALKKPSASDADWTEVLRLLAK
EGVVEPEVVRQIAITRLKWVEPEQNEEVLKQVEEVEQKRRRGAVARVKENVKMHEELR
EGGNHRAEDLQSGVESMEDYQPEDVAADQSLLALRTPIVTKKRGGTRASIMPKPVEDSP
MSRGNNSTFNSPRLEQTKAKDIETNFKYVQHLSMMQARPSSRLKKSSSMNNSTYRGNK
NTSISLEKGTQKALKIINRGTTIHEDDENEDNDDDDDMREEDTSIY
>C_oiwi_KNL2
MGDVEILPQRVQNVLDMEIVQLKLWSIKFNSTNIRLEGFVRNEEGTMMQKVCSENLVKR
MNSTMLFDVSGRFYELSGQIDREYQQKLGMPSRIIEEFANGFPANYAILINSCLNSDQIRSI
RRPIEAAPRDPLRPIVTLANETENPKIPEKIPEKSILEEEKEENEADFTLKAPVSTDGNAITP
IRFTRGNGRALARTVFEQTPVRGKASTSGPLASSTPQPAPPPIPRRSSTTENRNPPEAPPPK
VQYATDSDLFAVPKLPAPRQQAPPISLGFLDDFDDIFADAVIQKTPKRGRPPKREEEDSRRF
YEDESRMSRKRGYYRSRSRSPDYDRRHRDRDYSDFYGREDSKRPRRRDESSSSGGGGK
SVRFEREDRHQRHHPHRRDRSNSRESQRHHHRRHHDDDYNVSRSRSHEKEEKRRLQEL
LRKEKELEARLQSFRRPPTSSSSSSEDSDEDEMAGEWERENQEIMDNSMMFGDGISKKR
RRQSGEKKRKQPSRPKPKSAAPAPSATRKPAKKNKRSPAETDSLNDSIASHRPRRACATPS
TPAPKKVTWPKRDLDRLRRVIELKKPSGREEEWVEVARLLQKEGVNPMEVKEAAIGRLK
WKEPVEKEHTEEEEEEEKKRRRGVAARIKEGVKMHEELREGRQQNREDALKNRVEEVE
EFQPEDVEADQSLLALTTPVALKKKGGTRASIMPKPVEDSPMSRGNNSTFADSPRFDQTK
AKNLETTAKYVHHLSILQGRPSSSAANRTQMNRSTTRGGGSKNTTLSLEQGTRKAMKMI
SEGRTIHEDSEEDDEDDEDMDDDVFN
>C_kamaaina_KNL2
MGDTEILPLRVQNVLDMEIVRLKLWSIKFNSTNIRLEGFVRNEEGTMMQKVCSENLVKR
MNSTMLFDVSGRFYELSGQIDREYQQKLGMPSRIIDEFANGFPENYAILINSCLNSDQIRSI
RRPIEAAPREPLRQKQEPIVTLANESEIQNPQKTKIVPSKSPKNSKILEKSIQEEEEEEEDPA
NFTLKVPISTNGDAITPIRFTRGNGRGMARTVFEQTPVRGKPSSSSGPLASSTPQPPPPPRR
SSTTENRNPKTEQAPAPKPPQNPQYATDSDLFAVPKLPAPKQQAPPISLGFLDDIDDIFADA
VINKTPRRGRIGRPTSPSKQDVTSDSRGRFYESRRDEMVDESRMSRKRGYYRSRSRSPDY
DRRRRDHRDDHRYSDYYERDSRSDSKRPRQREESSSSGKSVRFERREDRHPDRRRHRST
SRESHRHRRRDDDYDVSRSRSRESRKNEKRKLQKIMKREKELMSRLQNSRRHSPSSSSSS
SSSEGSDDEDEMAGEWERENQEIMDNSMMFGDGISRNRKRKSEGKKRKQPSRPKQIKS
EAPVPAKKSAPKRNKRTPAETDFLNDSIASNRPRRACATPSTPAPKKITWPRRDLDRLKRV
IELKKPSGKEEDWVEVARLLAKEDVKPMEVKEVAIGKLKWKEPVEKVQTEEEEEEEKK
RRRGAAARIKEGVKMHEELQGGNPLKREDALKNRVEDVEEFQPEDVEADQSLLALTTP
VALKKKGGTRASIMPKPVEDSPMSRGNNSTFANSPHFNQTKAKELETNYKYVQHLSMM
HARPSSSAANKSQMNRTTTRGGSKNTSMSLEKGARKAMKMISEGRTIQEDSEEDDDDD
KEEELEDDVFN
>C_waitukubuli_KNL2
MGDTDILPVRVQNVLDMEIIRLNLWSIKFTATNIKLEGYVRNEEGTMMQKVCSEVICKR
MNATMLFDVSGRFFELSGQIDREFQIKHGMPSRIVDEFTNGFPENWAFLISNCLTTEQRSA
IRPIQAAPVQPLRLREPIVTLADETELGDRTVVRKENEKDQRKKNREEEEKKHRLLEEKR
KKEQEDKIAAGKRKESREEAERKQKEEEDAANYTFRAPPSFGPDAITPIRFTRGTQGRGG

IRIFEDTPQRTTSGPIDSSTPKPPPVHEQLPIKVQQKPEVHQQPEPRQTEKAPQTFASDSDL
FAVPRLPARAPISSELPTGFDFYDEMDAIFDNAKIDKTPVMESRRKQIPRRFQYEPSPPPIL
HGQSSSSSQAYEPDFDSRSYYNGRYDHERSRMSRRDDRYVSERDDSRISRKRILYSPDHH
DDSDHSRRRAYGSRSRYEDDFPKRSRTRETSSSSGRSVRFEDDYRSNRYHNDREHRDRD
RTREQQERDMYESQKLKEIMRREKELEAKLRLKKTATRRVSYSSDDSADDTMNMTEEW
ERENQEILDNSIMGDRKTSRKRNNAPKREAVKPKKVDKPARAPAKRTKKEPRQAPDLND
SIASNRPRRACATPATPAPKRVTWPKRDLDKLRHVIDLKKPSASIEDWTEVARLLKKDGV
EPADVKLIAETKLKWKEPVQDEDVLLQEEEEEKKRRRGVAAKVKEGVRMREEMREGG
AKEDELRNRVESLEDYQPDDMDADQSLLALTTPVAAKKKGGTRASIMPQPVEDSPAVKG
NNTSSFMNSPKLDSTKVKEVETTLKYVQHLSTMQARPGSSMNKSYMNASSSRGRNASIS
VEQGTLTTHIVVSIFLRMQMTFHLTPSDTVLFESWKVTDGWTMMGACALVVVAGVFVE
AIKRYRKKINQDQMIREQLVYEPFSNRLFASML
>C_panamensis_KNL2
MGDTDILPVRVQNVLDMEIIRLNLWSIKFSATNIKLEGFVRNEEGTMMQKVCSEVICKR
MNSTMLFDVSGRFFELSGQIDREFQIKNGMPSRIVDEFMNGFPENWAFLISNCLTTEQRS
AIRPIRAAPVQPLRPREPIVTLADETELPGDRTMSKKETERERKNKERDEEERKQALLEEN
RQRDRERELQKEKEREAVERRKKEEREEAERKRKEDEDAANYTFRAPPSFGSDAITPIRF
TRGKKGAVVTRIFDDTPVRSTGQPLASSTPQRSLPVKEQPPQLPKQLKEDSKEQQPPPRQ
VERPMQSYASDADLFAVPRLPSRAPAPTGLSGGFDVLDEMDSLFDNAKIEKTPVMGSRR
KPIPRRYQYESSPPPFLRGQSSSQQAYERDLSSASYYSSRYDDNRSRMSRRDDTSRRYDS
GRDESRMSRKRGHYSPPDRRDDYAYNRRREDESRQRDSFRYDEDYNYDSKRSRPREASS
SSGRSVRFEDDYCSNRSRENKDRRNHERVRDQRDRNMYESRELKEIMRKEKELEAKLR
SKRSSVRRVSYSSDDSDNDTMADEWERENQEMLDNSMIMDRKPSNKRKNVPKKETSKP
KRVEKPTRAPAKKNKKDPVETDDLNDSIASNRPRRACVTPSTPAPKRITWPKRDLDKLR
HVIELKKPSGNLDDWAEVTRLLKKAGVEPTDVKQIAETKLKWKEPVQNEEALLLEEEEE
KKRRRGAAAKVREGVRMHEEMRNGGKKEDDLRSGVESMEDYQPDDMDADQSLLALA
TPVAAKKKGGTRASIMPQPVEDSPAVKGNNTSSFMNSPKLDSTKVKEVETTLKYVQHLS
TLQARPGSSMNKSHMNRSTSRGKNTSISVEQGARKAMKIINRGTTIHEGDEDDEDDDDE
VTEDDEENVIY
>C_nouraguensis_KNL2
MGDTDILPVRVQNVLDMEIVRLNLWSIKFTATNIKLEGFVRNDEGTMMQKVCSEVICKR
MNATMLFDVSGRFFELAGQIDREFQVKHGMPSRIVDEFINGFPENWAFLISNCLTTEQRS
AIRPIQAAPVHSLRPKEPIVTLADETELAGDRTMARKESEKDRRNREREEEKRKKLLEEK
RQREEELQRERDREEERSRQEEEEKRRQEEEEAEKLRKEEEANNTFRAPKSFGTDAITPIR
FTRGNSKKGLGAIKIFDNTPTPKRKNGGPLASSTPQPQRLPLVKEQSQKEKTPQPEKPPQ
QQQESSRPAERPKQTFANDADLFAVPRLPTKSTASSGFSSGFDMFDEMDTLFGTAKIDKTP
VMESKRKPIPRRYEYESSPPPFRGQSSSSQTYQRDHDSRFNFNDRYDDDRSRISRRDGTFS
RYDSGRDESRMSRKRGLYSSPDQRHDYEYSRRKEDESRYRDRSRYENRYDPKRSRARET
SSSSGRSVRFEDDYRSKKYREDRREHDRTRDHRERDMYEDRKLKEIMRREKELEAKLHS
QRTRRVSSSSDDSADETVNMADEWDRENQEMLDNSMMMDTKTSRKRKNVPKREFVKP
KSVEKKPARASAKKNKRDPIETVDLNESIASNRPRRSCVTTAPKRITWPKRDLDKLRHVI
DLKKPSADLDDWAEVTRLLKKDGVQPAYVKLIAETKLKWKEPTQDADVLRLEEEEETK
RRRGAAAKVKEGVRMRQEIRGGGEKEDDLRRGVEAVEDYQPDDMDADQSLLALATPV
AAKKKGGTRASIMPQPVEDSPLVKGNNTSSFMNSPKLDSTKVKEVETTLKYVQHLSTM
QARPGSSMNKSSLNRSSSRGKNTSISVEQGTRKAMKIINRGTTIHEDDEEEEEDENSEDD
EENAVY

>C_becei_KNL2
MGDTDILPVRVQNVLDMEIVRLNLWSIKFTATNIKLEGFVRNDEGTMMQKVCSEVICKR
MNATMLFDVSGRFFELAGQIDREFQIKHGMPSRIVDEFINGFPENWAFLISNCLTTEQRSA
IRPIQAAPVQPLRPKEPIVTLADETELAGDRTVVRKETEKDRRNREREEEKRKKLLEEKR
QREEEIQREKDREEEKRRQEEEAERKRKEEEEANNTFRAPKSFGTDAITPIRFTRGNAKK
GLGAIKIFDNTPTPKRINGGPLASSTPQRPPPVKEKTPQPEKPPQKQQEAPKPAEISKQTFA
NDADLFAVPRLPTKSTASSGFSSGFDMLDEMDTLFETAKIDKTPVVESRRKPIPRRFEYES
SPPPFRGQSSSSQAYQRDLDSRSYFNERYDDDRSRMSRRDGTFSRYDSGRDESRMSRKR
GLYSSPDQRHDYDYSRRREDESRDRDRSRYDYRYDPKRSRARETSSSSGRSVRFEDDYR
SKKHREDRRDREDRRREYDRTRDQRERDMYEDRKLKEIMRRERELEAKLQAQRKSARRV
SSSSDDSADETVNMADEWDRENQEMLDNSMMMDTKTTRKRKNVPKKEFVKPKSVAKP
ARASAKKNMREPLETVDLNESIASNRPRRSCVTAVPKRITWPKRDLDKLRHVIDLKKPSA
NIDDWVEVTRLLKKDGVQAAYVKLIAETKLKWKEPSQDADVLRLEEQEETKRRRGAA
AKVKEGVRMRQEIRGGGEKEDDLRRGVEAVEDYQPDDMDADQSLLALATPVAAKKKG
GTRASILPQPVEDSPLVRGNNTSSFMNSPKLDSTKVKEVETTLKYVQHLSTMQARPGSS
MNKSSLNRSSSRGKNTSISVEQGTRKAMKIINKGTTIHEDDEEEEDEDDENSGDEEENSIY
>C_yunquensis_KNL2
MGDTDILPVRVQNVLDMEIVRLNLWSIKFTATNIKLEGFVRNEDGTMMQKVCSEVICKR
MNATMLFDVSGRFFELAGQIDREYQIKHGMPSRIVDEFMNGFPENWAFLISNCLTTEQRS
VIRPIQAAPVQPLRPKEPIVTLADETELGDRTMAKKETEKERRNREREEERKQALLEEKR
QRDHELRQEKDREEERKRKEEEDAANNTFRAPKSFGADAITPIRFTRANGKKGHGAFKI
FDNTPKRNTNEPLASSTPQRPLPVKEPEQPPQKQPETSRPAEPSRTYASDADLFAVPKLPT
KSTGSSGISSGLDFFDEMDTLFETAKIDKTPVMESKRKQIPRRFEAASSPPAFRGQSSSSHA
SNRDYDSRSYLDDRFNDDRSRMSRRDGTFNRYDSGRDESRMSRKRGHYSPDQRRDYEY
SHRRDDESRYRDRSRYEDDYRYDPKRSRAREASSSSGRSVRFEDDYRPNYHREDHRQR
DRTRDGNDVYENRKLEEILRRERELEAKLRAKKKPARRVSCSSDDSEDENVDMADEWD
RENQEMLDNSMMLNRKTARKQKNVPKRDVAKPKRVEKPIRAPIKKKTRDPVETDDLND
SIASNRPRRACVTPSTPAPKKITWPKRDLDKLRHVIDLKKPSANIEDWVEVTRLLKKVGV
EPADVKGIAETKLKWKEPTQNAEVLRLEEEEETKRRRGVAAKIKEGVRMREEMRNGGE
KEDDLRSGVEALEDYQPDDMDADQSLLAFATPVAATKKKGGTRASIMPQPVEDSPLVNG
NNTSSFMNSPKLDSTKVKEVETTLKYVQHLSTMQARPGSSMNKSYMNNSSSRGKNTSIS
VEQGTRKAMKIINRGTTIHEDDEDEEEDDDEVSEDDEDNAIY
>C_macrosperma_KNL2
MEYIVMGDTILPVRVQNVIDMEIVRLNVWSIKFTATNIRLEGFVRNEEGTMLQKVCSENI
CKRMNATMLFDVSGKFYELAGQIDREFQIKHGMPGRVVDEFLNGFPENWAFLIGTCLTT
EQRSVLRPIQAAPVQPLRPKEPIVTLADETELTGDRTMAKMETEKERRAKQREEQRKLE
KEREEEAVAERKDKEKREREEAERKRKEEEDAANYTIRAPNPLGNEAITPIRFHRGKGGG
PRFFNVFDKTPKATSSAPLASSTPQARPPVRKEPEQPVSSKPVERPVAFASDADLFAVPRLP
AKGPTTSSSSAGAPLQDDFDFLDEMDTLFDTAKIEKTPVMDTRRKPIIRRFEHRSYSPQH
GQSSSSQAYEMDYNDRREPYYHDRYEDDRSRVSRRDFGGDASRMSRKRGYYQSPENRD
RDEYEDRRRREHESRYSQQIYSREFEMNGWPRRETSSSSGRSVRFDDDYHYSDRHRNPR
DKTRDERDREMYERQKLKEITRREKELEAKLKSRKVSRSSEESADDSMDMAEEWEREN
QEMLDNSMMMMPDRKSSRKQKVVPKREKPKKAAEKKPAARAPRKKNIREPLLEEEDD
PNDSIASSRPRRACATPSTLAPKRITWAKRDLDKLNHVIELKKPTANLDDWVEVTRLLKK
NGVEAADVKHAAETRLKWKEPVQNDERAQEDEIEEERKRRRGVAAKVKEGVRMREE
MREGGEREDEMETRVGAVEDYQPDDMDADQSLMALGTPAVKKKGGTRASILPQPVEDS

PIVRGNNTSSFMNSPRLDQTKAKEVETTLKYVQHLSTMQARPGSSMNKSHANKSYMNK
STSRGGKNTSISVEQGTRRAMKIVNRGTTIEEDEEEDEEENDDDEEDTSVY
>C_sulstoni_KNL2
MEENGILPTRLQNLADMDVIPLELWSIKFYGTRYKVEGIVRPEDFTTAQKFTSEFIIKRITS
TALVDASERFIELVGQIDRTFQIRNGMPEEIVGKFQNGFPRDWWELLKPCLAAGQKTSTM
RAIEAPPLKTSTTSSTVQNVQKTSSEPIVTFADETEVQSDREKERRHRQEEREVQKVLEK
DDDDDADDADYTIRQPRAIDGGENTPLRYKRGHSGGPQNRNRVFADTPQRTAIPLAAST
PNAPPKIAKPPVKFASDADLFAVPVPPTRRPAERKERSESPEGCDGLFDKISEVFSNVNFPV
TPLPAARVRAHFSPAERSEGSYYYRDDRYDEDRNRTSRRDRSRSREYDESRTSRKRGYY
RSPEGYEYSRRRDDRRRDRYEASSQRHGSSSSGRSVRFEDPYEYDVEQRRRRREDQKLR
EIERRREREIEARLERSKRSRRYSSDEQSEDMADEWERENRQILDDSMLFRSAAGKATNKK
RAGRPKKEVKTRSPPVQKPQKARKEPRKRRSEPAETVDLNESIASNRPRRACVTPVAPPP
PKRVTWPKRDLDKLRRVIEVKKPTGDEHDWVEVARLLAKDGVTPTDVRTIAETRLKWR
EPTRDPKVLEAAEEEETKRRNGLAARVKESIRMREEMREGRQEEDVELKPGGENYEPAD
VSADESLLAIGTPRVPPKKKGGTRASLIAPPVEDSPMTRRSGAAATVVASPKLDATEVAAT
QKHVHHLSMIQARPGSSMMNTSRANKSIRNGKNTSISVEQGARRAMKMVTRGRSTIAE
EDESEEDDDENEEDVTAEEDEIY
>C_afra_KNL2
MEENAILPTRLQNLAEMDVIPLELWSIKFYGTGYKLEGIVRPEDFTSAQKFSSELIIKRITQ
TALIDASERFIELVGQIDRTFQLRNGMPEEIVNKFQNGFPKNWYELLKPWASQNQKTSTW
RAIEAPPVAPPNPQIPPIPQKAPCEPIVTFDDETEVQSDKEKEEARRRRQLEVEEAEKQRKI
LEDLNDDADYTLRNPRSIDGVQNTPLRFKRGNPNFERNPRVFNDTPVRTGGGPPLAASTP
QAPKVSDPIPKAANYASEGDLFSFAVPTQPPSQRPTRTTDEDRKHQADRRRRSESLGGSD
DGGGGSGLDLFSQMSEVFSNAKFPITPLPAAPPRSSRFERERAERSDYYREDRDRERYRD
PSRESRERDESRMSRKRGYYRSPEGYEYSRRREDRHRDRYESSYQRYGESSSSGRSVRFD
DGYDDPYEREEEERRLRRHRDNQKLREIMRRREREVEARLERSKRSRRYSSEDDGPEEND
DDLADEWDRENREILDNSMSFARGGSGRAGKKRGAGRPKKTSAPKSRSPPTQQKPRKQ
QQAAKKRRSEPAETDDLNDSIASHRPRRACTTPKAAAPKRVTWPKRDLDKLRRIIEVKK
PTGDADDWVEVARLLAKDGVEPADVKTCAETRLKWKEPTRDPEVRQKEEEEEKKRRN
GLAARVKQGIRMREELREGYEEDVAPPPTSSSQEPDYEPDDVSADESLLAMGTPLPPPKK
KGGTRGSIVVLPVEDSPMVRRSGVAASPKFDASDLAATQRHVQHLSMIQARPGSSMNTS
RASKSILGGKNTSISIEQGARRAMKMMNRGRVAIAEEDEDSDSDDSDDDENDSNDNAEY
IKIKVVGQDSNEVHFRVKFGTSMAKLKKSYADRTGVSVNSLRFLFDGRRINDEDTPKSLE
MEDDDVIEVYQEQLGGGI
>C_afra_KNL2
MEENAILPTRLQNLAEMDVIPLELWSIKFHGTGYKLEGIVRSEDFTSVRKFSSELIIKRITQ
TALIDASERFIELVGQIDRTFQLRNGMPEEIVIKFQNGFPKEWYELLKPWASQNQKTSTW
RAIEKPPVVPPNPQIPPLTQKTPCEPIVTFDDETELRSDKEKEEARRRRRLEVEEAERQRK
LLEEGSDGAEKRGYNRSPAEKNNRERSERSRRCSSEDNRSDGELADEWERVNSEILDNS
MSRVAGRPKKMLAFGRALTQADVEKFAIRLEIAGARIPAVVASEKYVVTTLKNIPEASRK
LGFALSAFDHLEQRIELKIHGMHRHSDFVVLRRDEGEFSEAPVYRNEEVGDKYFVVKRSI
DGKKIVCHGKVTSVGATGFAFVSEPYEGFPVANGDGAFAFLDGKLLGVVKDTKEQGRV
LEFVSIYLVHFAIDIIFHHQPFSV
>C_sp49_KNL2
MTERTIVPVRVQNVMDMEILRMNVWSIKFNSSSVMVSGFVRTPDGNSMNLVTSEAICKR
MNSTLLFDISGRFYELLGQIDKSTQLKAGMSSRVMEEFLNGFPDNWVFLVGQCLTTEQR

SAIRPIQAAPREPIRPKSPIVTLADETELREQGKAQKEQKEDADATFRAPTAFFHGAVTPIR
FTRGNGATGRSQRVFESTPSREPSSSRPLASSTPQPQKLPVLPPVAPRDVPAAQPRTFATDD
DLFAVPKLPASSSGLDFLEDMEAIFNGAHINRTPSHAAPRARVPREFRSRSPSYEQRYQDR
RYQDESRVSRKRGMDRSPEYYEMERRAEESRRSHESKRARPRETSSSSGRSVRFEETYRR
DRHHQNDRSHRSHHYEQEKEHKRRLREIERREREVEERLKRNRKVSSDSNSSSGEEDGS
FNEWDVENREILDRSLMEASGRERRPKREPTSKKSTSKKPAAPKPKPTKNIRKKRDPIET
DDPLNTSIASSRPRRACATPATKKITWPRRDLDRLKRVLEMKAPTGRSEDWVEVARLLA
KDGVNSEEVRQIAESKLRWKEPEHHPEEHQEDPLNDSQLIEKKRRNGIAARVKESVKMA
EELQTRPPNSDEMRGRVEEVEEYQPADVSADESLLALSTPQVIKKRGGTRKSTLPEPIEDS
PMARGANSSVNSPRLNATKAKEQETTLRYVQHLTTMQTRPASRMNKSAMNKSTASAKN
TSMTVEQGTRHVQKMMKKAMAIDEEEEDDEESDENQEETSFY
>C_sp25_KNL2
MTEQTVVPVRVQNVMDMEILRMNVWSVKFNSSSVMVVGFVRTPDGNSMNLVTSEAIC
KRMNQTLLFDVSGRFYELLGQIDKDTQLKQGMTSRVMEEFINGFPENWVFLVGQCLTTE
QRSAIRPIQAAPRQPIRPKSPIVTLADETEVAAQAQAQREQKEQKENEQRRRREDKEKED
ADATFRAPASFANGAITPIRFTRGASVASRNQRVFDATPTRQAPLASSTPQQKVQKVPTIA
ADEDLFAVPKLPVKKGVAPASDLDFLDDMDALFTGARINRTPSHRVPPKPQQRRRSRTRS
RSPSIDRSRSRRYEEEFDSRMSRKRGYDRSPDYDDYRREESRRRFKESKRSRPRETSSSSG
RSVRFESDYHRQRHDDRRRHEDREELERQERRRLKEIERREREVEERLKRSRRKNYSSSS
SSDEDQDGSFNEWDAENREILDQSLLASTGGSSRSASGGNSRGRLPKREPSLKKAKPAA
AAPKPKPRKKKEVVDLNDSIASNRPRRACATPATPAVKKITWPRRDLDKLKRVIELKKPS
ASEEHWTEVTRLLAKEGVAWTDVKSIAETKLKWKEPRQEMVEVEEEEEEERERKRRRG
IAAMVKESVKMCEEIQTRNVQEDEMEGRVEEVEEYQPADVSADESLLAMSTPQITKKRG
GTRKSTLPEPIEDSPMAKGTNSSWNSPRLNATKAKEQETTLRYVQHLTTMQTRPGSSMN
KSYMNKSTASTRGKNTSLSVEEGTRKVQKMMKKAMAIEEEDEEEEDSDDEGETSLY
>C_imperialis_KNL2
MTESTVVPVRVQNVMDMEILQMNVWSIKFNSSSVMVVGFVRTPDGNSMNLVTSEAICKK
RMNSTLLFDVSGRFYELLGQIDKTSQLKQGMSARVLEEFLNGFPENWVFLVGQCLTSEQ
RSAIRPIQAAPRQPLRQKSPIVTLADETDLGDRHKREEVENEARKREERKMIERENEEKR
RAEERRQREEQEKEDANCTLRAPASFASGDLTPIRFTRGNGPCNRKNRVFEDTPSRETTS
RPLAASTPQQKPKPVDPPREVPRDVPARPKTNFATDDDLFAVPKLPANRAAKPSSDLDFL
DDMDALFAGAHINRTPSHAPKPVSRIRQSPPRRRSPSENRSRHEYSQRSNRYDDDYDVRF
NRNGKEESRLSRKRGYYRSPDEYYETRRDWGRDRSRDRSRQRYDSERYDSRYDSQVSR
PRETSSSSGRSVRFEDDYSQRNNRRYEHEVRERKQLKDIERREREVQERLQRSKKSAPAR
QATYSSSEDDDDASFNEWDAENREILDQSLLASTAGTSRRRSAPKREPKSKKPKPAPAPA
PKRKPSKPRKQRDPVETDDLNDSIASNRPRRACATPATPAPKKITWPRRDLDRLKRVIEM
KKPTASDGAWAEVARLLAKDGVTPADVKQIAETKLKWKEPTQDPVEIEEQEEVELKRRR
GIAAKVKESVAMREEMQRRAVDEDELNGRVEEVEDYQPADVSADESLLAMGTPQVTKK
RGGTRKSTLPEPIEDSPMARGNNSSFMNSPRLDATKAKEQETTLRYVQHLSTMQNRPGS
SMNKSQLNKSSTSRGKNTSMSVEQGTRKAQKMIRKAMAIEEEDEEDDEDDEDENREEE
ETSFY
>C_japonica_KNL2
MAETGVLPLCVQNVMDMEIVRLNLWSIKFYAESIKLEGFVRNSDGNMMQKVCSEAICK
RMTSVLLFDVSGRFFELVGQIDHEFQVKNGMPSRIVAEFMNGFPENWLFLIKQCLPPTDE
QIPQRQIQIQSAPAPPKEPIVTLADESEHVSDHPKKIVENGRIRDEESADCTFRVPHPVAID
AITPIRFTRGVRKENAQRIFDSSTPKPQVEKQKKPLAAFDADLFAAPAPARAPPPTRAPPP

NPPGPSSEFVLLDEMDELFNNAEFDKTPGRTVAQPRRINETSRMDRKTSRTYSEDCDRKY
DDRRVECYSRSARENVEPNRNEFRMNRKRPSVNNYADEMNGGYNYRSNKYGRREVFE
SNGYHSDEFRSREAPSSRSYADRDYYSRKSYHRGEEQRNEKMYGDILRREMALETRLK
NGRSQRHVSESSEDSMDDSYMVDEWDRENQRILDSSIVSSFGRFETSPKKSRIYVKKERT
AKKARPKPVPPRRKKNSPDSPEMDDLNDSIASNRPRRACVVPSTPVPKRVTWPKRDLDR
LSRVIELKKPSSSDSDWTEVARLLAKEGVEGADAMKIAMAKLKWKEPVKDCETVQEDL
EMVDKKRKRGFAAKVKESVQMMEEMRDGGEQSGEMGGQGVEQAEDYQPADLSADES
LLALATPCAGKKKGGTRRSVMPAPVEDSPVIKLNECSYANSPKLDVTKCKEVETTLRYV
HHLSTVQARPGSSSSRSYKNRSVCLGKSLSLEQGARKAQKLISMAVGIEEEDEEEEDDDD
AELSIF

Supplementary Data 3.S7: All KNL-1 sequences used in this study

>C_tribulationis_KNL2
MGDGDVVPARVQNVLDMDVVRLNLWSLQFNSSGVRLEGFVRSEDGNMMQKVHSGIIC
RRMSATMLFDVSGCFYELAGHIDREFQQKLGMPSRVIEEFMNGFPDNWAFLIKSCLPAEP
KSALRPIQAAPEPSKARGEPIVTMPDETIVDNQSAEKERKRKELQKKKEQEERVRIQREQ
RAAAEAAAERKRQEDEEFKRREEEEERKRQEEENDAANYTFRVPQSQNGDAITPIRFTR
GNGQKGANVRSIFEKTPRTKPAGPLASSTPQAPPQQHAPRISIAEAKQPDEPPHPHPKPQP
PTRESIRETQYCSDSEFAVPKLPAPRNNRHASASSQPAPLDFLDEMDALFETANVDHTPGR
GRKSRKPVSRSPSPGRFNSTARDRTGYDRYEPSRQSSSQRYYDDYNNVSRMSGRNHTLG
GHDMRRDESRNSRKRGYNNSPEDYSRRFDDRSRRRDNYESDSRYDSKRSRPRDQSSSSG
RSVRFEEDFPRNRRDESRDSRSYRHYEDHRNRESSGDREDKRKLDAIVRREKELVARLQ
NTQRSSSTLQRSGYSSEDDEMSDEWDRENQEMLDNSMMFGDGLSKKKGKSGGHKTMR
QAKTRYPPKPKPAQKPAQPKKKKKDVEYESDEMNDSIASNRPRRACATPSTPAPKRITWP
KRDLDRLKHTIELKKPTGAEADWAEVTRLLAKDGVDAEVVKQTAIAKLKWKEPSQKTI
QQEEEEKKRRRGATARVKEGVRMHEELREGGNNRAESSQSGVEAVDDYEPDDVAADQS
LLGLQTPIAVKKRGGTRASIMPQPVEDSPVVRGNNSTLNSPRLDQTKAKDVETTLKYVQ
HLSMMNARPSSRANSSYLNKSSSRGGDKKNTSLSVEQGARKALKIINRGRTIHEEDEDD
EDEDITID
>C_sp41_KNL2
MGDGDIVPARVQNVLDMDVVRMNLWSIRFNGSGVTLEGFVRSEDGNMMQKVHSGIICR
RMSATMLFDVSGCFYELAGHIDREFQLKIGMPSRVIDEFMNGFPDNWAFLIKSCSTSEPK
SALRPIQAAPVEPLKSRSEPIVTMPDETVLAESQTAEKERRRKELQKKKENEERLRIQREQ
QEAVKAAEKKSQEEEKDKLQEEERKRQEEEDAANYTFRAPKSQSGEAITPIRFTRGNGQ
KGANVRPIFDKTPVRTKPAGPLASSTPQAPPPQHPHRLPNIETTRPDASVPPVPQPPIRETQ
YCSDSEFFAVPKLPAPKNIRGTSAKPLGFLEEMDALFENVNVDQTPGRVRNPRKVSRSPS
PMRLNSSARNRDSGYDRYEPSRQSLSQRYYDDYTTSRMSGRDDTFGRNDTRRDESRNS
RKRGYNNSPDEYRNRWDDRSRRHDNFESDPRYDSKRSRPRDQSSSSGRSVRFEEDHPRS
RMDSRESMDSRNYRHYEDSRNRQSSGDREDKRKLNDILRREKELMARLQNSQRSSSTF
QRTAYSSEDDEMSDEWDRENQEMLDNSMMFGDGLSQKKRRSGGRQPAKKPNYLQQAK
TKRVPQPKPAPKPAQQKKKKKQDSEDERDEMNDSIASNRPRRACVTPSTPAPKRITWPK
RDLDRLKHTIELKKPTGADADWAEVTRLLAKDGVDAEVVKQIAITKLKWKEPTQQTIQ
MEEEEQKRRRGATARVKEGVRMHEELRAGGNDRGNSSQGGVGAVEDYEPDDVAADQS
LLGLQTPIAIKKRGGTRASIMPEPVEDSPMVRGNNSTLNSPRLDQTKAKEVETTLKYVQ
HLSMMNARPNSRANTSYFNKSSSRDKTNTSLSVEQGARKALKIINRGRTIEEEDEDDEDG
DTTID
>C_zanzibari_KNL2
MGDGDVVPARVQNVLDMDIIRLNLWSIQFNGTGVKLEGFVRSEDGNMMQKVHSGIICR
RMSATMLFDVSGCFYELAGHIDREYQLKIGMPSRVIDEFMNGFPDNWAFLINSCSEPKSA
MRPIQAAPREPLRTRGEPIVTMPDETNIESQSSEKERKRKEREEQQKWEHEERLRARKEK
EQREAAEIAEKNRRREREIQKQREDEEKGRREEEEEAANYTFRAPKSQNGEAITPIRFTRG
NGQKGANIRPIFDSTPVRTKPAGPLASSTPQAPPPQNRHSNTENRQPDALPTAGPAPPVQK
PNRETQYCSDADLFAVPRLPAPRNNQNVSSSAQSAPLDFLDEMDALFESATVDKTPGRV
KKVVRVNRSPSPDRFSSRDRDSGYDRYDSSRYSHSQRYNDEHHMSRMSGRNDTFRRND
GWRDESRNSRKRGYNNSPEYTRGWDDRSRHRDNYESESRYDSKRSRPRDQSSSSGRSV
RFEDDHPGNHRDSRDPRNYRDYEDDRNRESSGDRADRRKLNDILRRERELEARLRNTQ

RSSTARRHAYTSEDDEMQDEWDRENQEMLDNSMLFGDGISQKKRRSGGRQPAKNSQSG
RSQQPKPARKPAQPKKKKERSDETDELNDSIASNRPRRACVTPSTPAPKRITWPKRDLDR
LKRTIDLKKPTGADADWDEVTRLLAKDGVDREIVKQTAIMKLKWKEPSQETVQQEEEE
KTRRRGAAAMVKEGVRMHQELREGGDNKRGDDLQGGVEAVDDYQPDDVAADQSLLA
LQTPIAVKKRGGTRASIMPQPVEDSPVVKGNNSTLNSPRLNQTKVKEVETTLKYVQHIS
MMNARPSSRANTSCLNKSSSRGGGSKNTSLSVEQGARKALKIINRGRTIQEEDEDDDDES
EAGDITIE
>C_sinica_KNL2
MGDGDVVPARVQNVLDMEIVRLNLWSIQFNGSGVKLEGFVRSEDGNMMQKVHSGTIC
RRMSATMLFDVSGCFYELAGHIDREYQLKIGMPSRVIDEFMNGFPDNWAFLIKSCSNPEP
KSALRPIKAAPKMPLRSRGEPIVTMPDETVMAESQAAEKDRKRKEHEEQQKRDDEKRF
RAEKEKEQREAADTAAKKRQEEEDRRRREEEDANYTLRAPESHNGEPVTPIRFIRGNGQ
KGANIRPIFEKTPVRTKPTGPLASSTPQAPPPPPTRRTSNIENKQPEASTSPKPQPPVQKPV
RETQYCSDTEFFAVPKLPASRNKHVPASDHAAPLAFLDEMDALFNGADVDHTPGRERKP
KKVSRSPSPERFDYSSRDRDSGYSRYDSSRYSHNQRYNDDYNMSRMSGRTDTSRRNDG
RRDESRMSRKRGYNNSPDEYSRRYDDRSRRQDDYDSGSRYDSKRSRPREQSSSSGRSVR
FEEDYSRHRQGSRDSRDPRDYRDYEDHRNRTSSGDREDTRKLNDILRRERELMARLQNS
QRSSSTVQRTAYSSDEDEMADESSLWDRENQEMLDNSMMFGDGLSQKKRRGGGRHPA
KQAKTRQPAQPKPARKPAQPKKKNRADETDDLNDSIASNRPRRACATPSTPAPKRITWPK
RDLDRLRHTIELKKPTGADADWAEVTRLLAKDGVDSEIVKQTAITKLKWKEPSQETIQM
EEEEKKRRKGAAARVKEGVRMHEELREGGDKRGDNSLGGVEAVEDYEPDDVAADQSL
LALQTPIAVKKRGGTRASIMPQPVEDSPVVRGGNNSTFNSPRLDQTKAKEVETTLKYVQ
HLSMMNARPSSRANTSYHNKSSSRGGGSKNTSLSVEQGARKALKIINRGRTIHEEEEDD
DEEGDTTIN
>C_nigoni_KNL2
MNDLFFVNYGSKFMNDYAQKKKASSHIEDYSLPEKKLRKVQEIVDKSPKCLFAKRAPPH
THFAANFSNLNSEGIKTVFLDIHLDFTISTPYFCYKASFCKKKMGDRNVVPARVQNVLD
MDIVRLNLWSIQFVDSGVILEGFVRSEDGNMMQKVRSGMICKRMTATMLFDFSGCFFEL
SGQIDREYQQKMGMPSRVIDEFTSGFPENWVSLIKSFLPVNPISAVKHIQPAPREPLRPISE
PIVTLPDETTLESEKDQKRREKEERLEKKRLEEERLEKERREQKLRAQKEKERREAVAAA
EAEKKRLEEEEDAANYTIRVPKSQSGEAITPIRFTRGNGQKGANVRPIFDKTPVREKTAGP
LASSTPQAPPPQQRLSNVEKKPASPEPQPRNPQRDREIQYASDSDLFAVPKLPAPKTTKPST
TSEASSGGVLDFLDEMDTLFDTVVVEQTPTRDRRPPRRYSRSPSPRRRQHSSSRDMGKG
YDNFESSRYSQRYNDNYNTSRMSRRDDTFRRNDERRDESRMSRKRVYNNSPDDFEYNN
RYDDRSRRPDYYDSDSRYDSKRSRPRETSSSSGRSVRFEDDYRRNHRDPRDRSDSKDYR
NYDESRRNPEDRREEKRKLNDILRREEELVTRLQNRKKPSASYRREPSSDEDDTADEWD
RENQEILDNSMMFGDGLSQKKRRSAGRPSKPSKKERQVQPKPVRKPPQPKKKQKSPPDE
LNDSIASNRPRRACVTPSTPAPKRIVWPKRDLDRLKHTIGLKKPTGSDADWAEVTRLLAK
EGVDAEVVKQVAITRLKWKEPAQDPETIQREEEEEKKRRRGVAARVKEGVKLHEELRQP
GVKRGDNSQTGVEAMEDYEPDDVAADQSLLALQTPVGAKRKGGTRASIMPEPVEDSPL
VRRNNSTFNSPRLDQTKAKEVETTLKYVQHLSMMNARPSSRANTSYYNKSSSRGGGSK
NTSMSLEQGTRKALKIINRGRTIHEEDEDEDDDDDDEDQYEDGVVY
>C_briggsae_KNL2
MGDRNVVPVRVQNVLDMDIVRLNLWSIQFVDSGVILEGFVRSEDGNMMQKVRSGMIC
RKRSETKKQEERLGKKRFEEERLEKEQREQKLRAEKEKERREAVAAAEAEKKRLEEEED
AANYTLRVPQSQSGEAITPIRFTRGNGQKGANVRPIFDKTPVRGKPVGPLASSTPQAPPPQ

QRLSNVEKKPASPEPRPRIPQRDRDVQYASDADLFAVPKLPAPKTTKPSTTSEPSTGGGFD
FLDEMDELFDTVVVEQTPTRNRRPPGWYSRSPSPRRRQHSPPRDSFESSRYSQRYNDNY
NTSRMSRRDDTFRRNDERRDEFRMNRKRVYVCDRGLVRKRKYVFQNNSPDDFEYSNR
YDDRSRRPDYYDSDSLYDSKRSRPRETSSSSGRSVRFEDDYRRNHCDPRNRSDSRDYRS
YDESRRKPEDHREDKRKLNDILRREQELVSRLQNRKKSSVSYRPEPSSDEDDMADEWDR
ENQEILDNSMMFGDGLSQKKRRSAGRPSKPSKKERQVQPKPVRKPPQPKKKKQKSPPDE
LNDSIASNRPRRACVTPSTPAPKRIVWPKRDLDRLKHTIGLKKPAGSDADWAEVTRLLAK
EGVDAEVVKQVAITRLKWKEPAQDPETIQREEEEEKKRRRGVAARVKEGVKLHEELRQP
GVKRADNSQTGVEAMEDYEPDDVAADQSLLALQTPVGAKRKGGTRASIMPEPVEDSPL
VRRNNSTLNSPRLDQTKAKEVETTLKYVQHLSMMNAHPSSRANTSYHNKSSSRGGGSK
NTSMSLDQGTRKALKIINRGRTIHEEDEDDEEDDEDQYEDAVVY
>C_remanei_KNL2
KRQEEEDADYTFRAPQSQNGEPITPIRFNRGNGQRQGVTRSVFERTPQRGQSGPLAASTP
QAPPPPPPQQRLSNIENREPPPRVVAPPSPVRQPPAPQPPIREPQFANDDDLFAVPKIPPPKIP
RGSAGNSGGNIDFLDEMDALFDTVYIDKTPKRDVKPKRPSSPSPERRRYSPMPRDREMG
YNDDFESSRRGGRYPSESSNMSRMSRRDETNRRNDGGMGRDESRMSRKRGYDQSPDD
MEYQRRREDHYRRPDYYPPRDSRQDSKRYRPRENSSSSGRSASVRFADDYQRNRGDSR
DPRDFRDPRDSFSRDPRFYYENNQRGESSKDRDTRKLNNILRQERELVARLQNIKSNTTT
NTNTTRRVTYSSEDDEMADEWERENQEIMDNSMMFGDGISKKGRRSGPGRPPQRKPKE
QPKPKSQSQPKRRNQNQSKPRRSQYDPVETDDLNDSIASNRPRRACVTPSPVVPKKIIWP
KKDLDRLKRTIELKKPTGSEADWTEIARLLAK
>C_latens_KNL2
MPSRVIDEFTNGFPENWGFLIKSCFGNESRSAMRPIQAAPREPLRQRNEPIVTLADETELT
NRKNSDNDSESENNRKRREREEEAERERQERYRADEMERERIAAEKKRQEEEDADYTF
RAPQSQNGEPITPIRFKRGNGQRQGVTRSVFEKTPQRGQSGPLAASTPQAPPPPPQQRLS
NIENREPPPRVAAPPSPIRQPPAPQPPIREPQFANDDDLFAVPKIPPPKIPRESVGNGGGSIDF
LDEMDALFDTVYIDKTPKRDVKPKRPLSPSPERRRYSPMPRDRELGYNDEFESSRRGGR
YESSNMSRMSRRDDTYRRNDGGMGRDESRMSRKRGHNYSPDDMEYERRREDHYHRP
DYYPDSRYDSKRYRPRENSSSSGRSASVRFADDYQRKRGELRDPRDFEDSFSRDPRFYHE
SNRKRESSQDREDTRKLNDILRREKELMARLQNNKTTTSKNTTTRRVTYSSEEDEMADE
WERENQEIMDNSMMFGDGISKKGRRSGPGRPPQRKPKEQSKPKSQSQPKPRNKNQSKP
RRSQHDPVETDDLNDSIASNRPRRACVTPSPVVPKKIIWPKKDLDRLKRTIELTKPTGSEA
DWTEVARLLAKNGCDGEAVKQTAITKLKWKEPVERDEETIQREEEEEEETKRRRGVNA
KVKEGVKLQEELRRGATHQKRAEDVRVEVTAEEIEPDDLAADQSLLAMATPVAVKKRG
GTRASIMPMPVEDSPIVRGSRNNSTFNSPTLNQTKAKEVETTLRYVHQLSMVQARPASR
NNTSYMNKSSTRGGGSKNTSMSLEQGVKKAMKIINRGTTIHEDDEEEEESESEEEDNSE
DLEEH
>C_sp51_KNL2
MGDTDILPTRVQNVLEMEIVRLNLWSIKFNALNIKLEGYVRSEDGNMMQKVCSEIICKR
MNSTMLFDVSGCFYELSGQIDREYQSRMGMPSRVIDEFVNGFPENWAFLIKSCLSMGQR
SALRPIQDTPREPIRTRVEPIVTLADETELVENERKKAESEKDKKRKEREEDRARENERNL
AAQREKERKDTVAAEERKRKKEEEEAHAVAERKRQEEEEAANYTLRAPQSQTGEAITPI
RFTRGNGGNGAHVKSIFQKTPVRGKSNGPIASSTPQQTRVLPTLEKPVETTVSKSTKSKSP
PPRSIRQTEYASDADLFAVPRNPAPKNNRKPAPTTSSDIGFDFLDQMDSLFDNPDFEQTPT
RDRKPRRQFSPSPEPRHRSYSRDRDRYAHLESTRYKQRYTDDYDMSRISRRDVPFRRND
GGRDESRMGRKRGYYGSPDGSEYGRRYDDRDRHMDYYEPDFKYDMKRSRPRETSSSS

GRSVRFENDYRRRREDSREPNSSRNHREYDDYRGRTDSEDNRRLEFLEKRENELMARL
QNHKKPSSSSSRRLTNSSDDEMADMWEQENQEILDNSMMFGDGISKKKRRSGGGRQPK
KQPTRKTQPKPTRKPAEPRKKTQRSPVQRDELNDSIASNRPRRACATPSTPAPKRITWPKR
DLDKLKHVIQLKKPTGADADWIEVARLLAKDGVDSELAKQTAVSKLKWVEPVQHEERI
DEEEEENKKRRRGAAAKIKEGVRMHEELMSGRQEDDIRGGGEIMEDYQPEDMAADQSL
LALGTPLAAKKKGGTRGSIMPKPVEDSPVTRGGNSTYNSPRLDRTKAKEVETTLRYVQH
LSNMQARPSSVSNTSTLNKSSGRTSKNTSMSLEQGTRKALKIINRGRTIHEEDEEESGDEE
SELSEEEEEGESFNY
>C_sp44_KNL2
MGDTDILPARVQNVLDMEIVRLNLWSIKFNASNIKLEGYVRSDDGNMMQKVCSEIICKR
MNATMLFDVSGRFFELSGQIDREYQAKMGMPNRVIDEFVNGFPENWAYLINSCLTISQRS
ALRPIQAKEPIRSKGGPIVTLADETELVENEQRNSDSEKEKKKKEGEQRASENKRNQEAQ
RERERKEALAAAERERKRHEEDAERKRRKEEEDAANYTFRVPESQLGEPLTPIRFTRGNG
GNGANIKPIFEQTPVRGKTTGPLASSTPQAPAPQQPRVLKDLEETVKPTAPQSSKAKPPPQ
KPQTEYASDSDLFAVPRPPADKNSRTTTSASSSDIGLGFLDQMDLLFDSAAVERTPTRHPK
PRRIAVSPPNPRYRSHSKDRNRYGDFNSTRYTQRPSDDYDMGHMSRRDATFRRNDGGRD
ESRMSRKRGYYGSPDESDYGRRRDDRDRRDNYYEHEYDMKRSRPRETSSSSGKSVRFE
DNYRRRREDPRETNYSRSYRDFDDFNTRGTSGDREDNRKLNDILRRENEVKARLQNHH
KSSFSRRHVQSSDEESDDMADEWDRENQELMDNSMLFGDGIPKKKRGSDGGRPAKKQP
VRKPQPKPTRKPAEPRKRAPRSPVDTDELNDSIASNRPRRACATPSTPAPKRITWPKRDLD
KLKHVIQLKKPTEAEADWVEVARLLAKEGVDSELVKQTAISKLKWKPPVQEERNDDED
DEDKRRRRGAAAKIKEGVRMHEELMSGRNPVDDIRSGVEFVDDYQPEDVAADQSLLA
MGTPLAAKTKGGTRGSIVPKPVEDSPIIRGRNSNYNSPRLDQTKAKEVETTLRYVQHLSN
IQARPSSRANTSTMKKTSSHASKNTSMTLEQGTRKAMKIMNRGRTIREEDEEEDESDRE
DESEQSGEEEEEDDIYY
>C_sp48_KNL2
MGDTDILPARVQNVLDMEIIRLNLWSIKFNASNIKLEGYVRSEDGNMMQKVCSEMICKR
MNATMLFDVSGRFYELSGQIDREFQAKMGMPGRVIDEFVNGFPENWAYLINSCLTVGQR
SALRPIQSAPREPIRSRAEPIVTLADETELVENERKNSEAEKEKKKNERDEQRARENERNQ
EAQRERERKEALAAAAETKRRKEEEEERAEAERKRREEEEAANYTFRAPESQQGEPITPI
RFKRGNGANGGYKSIFEKTPVRGKSNGPLASSTPQAPPPQQPRILSSLEREKKTEPDAPKS
PIADRTVQKPIRNTEYADDADLFAVPKLPPIRNNRPSAAAPSSEIGYDFLDQMDSLFDTVV
IDQTPNRNRMPRRAASSSMEERLRSSPPRNRDRYDDRESTRYSQRYGDDYNTSRMSRRD
ATFRRNDGGRDESRMSRKRGYYGSPDEFDHRRRDDRDRRGDYYEPDYKYDMKRSRPR
ENSSSSGRSVRFEDERFGDYRRHREDSRERKYSRNHREYDDNRGRRGSSGEDDRKLNDI
LRRENELMARLQNHRPSSSRRDASSSGEDEDDLANEWDRENQEILDNSMMFGDGLPKK
SRRSAGGKIGRQPKKQPTRNPPPKPARKPEPRKKAPKSPVETDDLNDSIASHRPRRACAT
PSTPAPKKITWPKRDLDKLKHVIQLKKPTADDADWAEVTRLLAKDGVDSEVVKQTAISK
LKWKEPVQEQERKEEEEEEERKRRRGAAAKIKEGVRMHEEMRRGRREEDTQMSADSM
EDYQPEDVAADQSLLALGTPLVAKKKGGTRGSIMPKPVEDSPITRGRNSTLNSPKLDQTR
VKEVETTLRYVQHLSNMQARPSSRANTSTVSKSGRGSKNTSMSLDQGARKAMKIINRG
RTIDEEDEDEESNEEEEDHSEDEAFDY
>C_brenneri_KNL2_paralog1
MGDTDILPARVQNVLDMEIIRLNLWSIQFNASNIKLEGYVRSEDGNMMQKMCSEMICKR
MNATMLFDVSGRFYELSGQIDREFQAKMGMPGRVIDEFVNGFPENWAYLINSCLTVGQR
SALRPIQNAPREPIRSRAEPIVTLADETELVENERKNSEAEKEKKKKELEEQRAKENERNL

EAQRERERKEALAAAERKRREEEEEAANYTLRAPESQPGQPITPIRFRRGNGGNGGHMK
PVFEKTPVRGKSNGPLASSTPQAPPPQQPRILSSLEKPKSPIAERSAQKPMRRTEYADDAD
LFAVPKLPPIRNSRPSASAPSSDMGFDFLDQMDSLFDTVVIDQTPTRNRMPRRAASSSME
SRIRSPPRDRERYDDRESTRYSQRYGDDFNMSRVSRRDATFRRNDGGRDESRMSRKRGY
VSTPDEFDHRRRDDRDRRGDYYEPDYKYDMKRSRPRENSSSSGRSVRFEDERFGDYRR
HREDSREPKYSRNHREYDNYRGGRGSSGEDDRKLNDILRRENELMARLQNHRPSSSRR
DAYSSDEDQDDLANEWDRENQEIMDDSMMFGDGLPKKQRRSGGKIGRPPKKQRTRKPP
PKPARKPEPRKKAPRSPVETDELNDSIASHRPRRACATPSTPAPKKITWPKRDLDKLKHVI
QLKKPTAAEADWAEVARLLAKDGVDSEIVKQAAISKLKWKEPVQEQEKKEEEEEEERK
QRRGAAAKIKEGVRMHEEMRKGRRENETQMSAESMEDYQPEDVAADQSLLALGTPLV
AKKKGGTRGSIMPKPVEDSPITRGRNSTFNSPRLDQTKVKEVETTLRYVQHLSNMQARP
NSRANTSTVSKSGRGSKNTSMSLEQGARKAMKIINRGRTIDEEDEDEESNGEEEEDHSED
EAFDY
>C_brenneri_KNL2_paralog2
IDREFQQNGMPVRSYDEFVNGFPETGLILLILASQEPIRSRAEPIVTLADETELVENERKNS
EAEKEKKKKELEEQRAKENERNLEAQRERERKEALAAAERRRRQEEEERADAERKRRE
EEEEAANYTLRAPESQPGQPITPIRFRRGNGGNGGHMKPVFEKTPVVSEGQ
>C_wallacei_KNL2
MGDNHVLPARVQNVLDMEIIRLNLWSIKFNASNIKLEGYVRSEDGCSMQKVCSEIICKR
MNATMLFDVSGRFFELAGQIDREFQLKMGMPSRVIDEFVNGFPENWGFLINSCLTMNQM
SVPRPIQAAPREPLRSRNEPIVTLADETELADSGRKTSESERDKKRREREEQRMKENEKR
LETEKEKERQKAQAAAAAEKKRQEEEEDAANYTLRGFQMESGDPITPIRFKRGQANGV
HAKPTFVQTPVRGKPSGPLASSTPQAPLPEHPRRSLDLGKRSTPLASPSPKVQQPVSKPAR
ETEYASDSELFAVPKLPAPKAARPSATSSAPSDTAFDFLDDMDVFFDTAVIDQTPVRARKQ
PRRVSPDVRRDRSSSRDPERFGDYDSTRNNQRYNNDYDMSRMSRRDATFRRNDERRDD
SRMNRKRGYYNSPDDYDNRRSDYYESDSRYDTKRSRHRETSSSSGRSVRFQDDHRRSE
NDDYRRREGFRDPEYQRNYRDFEERRVRGSSGDEDKKKLNDILRREKELMARLQNNQK
PSSSRRVSCSSDDDIDMADEWDRENQEIMDNSMMFGDGISKQKRRSGGSRQEKKKQPT
RTVQPKPKRVASKKKSPVPSDDLNDSIASNRPRRACVTPSTPAPKRITWPKRDLDKLQHV
IDLKKPSAEDADWVEVTRLLAKEGVESEVVKQIAISRLKWKEPVKNKNNEKKEEEEDET
RRRRGAVARVKEGVRMHEELIRGRDDEERDVQVESMEDYQPEDVAADQSLLALGTPLA
AKKKGGTRGSIMPKPVEDSPMARRQNSSFNSPRLDQTKAKEMETTLRYVQQISMMQAR
PSSRANKSKLNKTSARGSKNNSMSVEQGTRKALKIINRGRTIHEESEEEEEESDDDEFDE
REEGDNSIY
>C_tropicalis_KNL2
MEDNHILPARVQNVLDMEIVRLNLWSIKFNSSSIKLEGYVRSEDGCSMQKICSEIICKRMN
ATMLFDVSGRFFELAGQIDREFQSKMGMPSRIIDEFVNGFPENWGFLINSCLTMNQMSVP
RPIQAAPREPLRSRNEPIVTLADETELNDGGRKNSESEKDKKQKEREQQRMKENERKLQ
AEKDKKDKEAVAAAEKKRQEEEDAANYTLREFQTGSGDPITPIRFKRGTQSKGANNIKP
TFLQTPVRGKPNAPLASSTPQAPPPEHPRRSLDLGKPPIQLASPSPKVQKPETVRETQYAS
DSDLFAVPRLPAAKNNRPPPSSSSETVLDFFDDMDVFFDTAVIEKTPVVARRQRRLSPEAR
RDRSSSRDPERFGDYDSTRYTQRYNNDYDMSRMSRRDAVFRRNDERRDDSRMSRKRGY
YSSPENDFDRRRPEYYESDSRYDTKRPRPRENSSSSGRSVRFHEDHMGRGDEFRGSRYER
GYREFDDQRGRGSSADEDKRKLNAILIREKELMARLQNKHKPSASRRVSYSSDEDSDDM
AAEWERDNQEIMDNSMLFGDGIMQQKRRSARGHPAKKKQPIRTEQHKPAQKRAPPKKK
SPAPRDDLNDSIASNRPRRACATPSTPAPKRITWRKRDLDKLQHVISLKKPTAADVDWVE

VTRLLAKEGVEPEVVKQAAITKLKWKEPVRNDENMREEEDESRRRRGVVARVKEGVR
MHEELMNGGNEVEEDLQVQSVEDYQPEDVAADQSLLALGTPLLAKKKGGTRGSIIPKP
VEDSPIARGKNSSFNSPRLDQTKAKEMETTLKYVQQISMMQARPSFRANKSQMNNSTAR
GSKNTSMSVEQGTRKALKIINRGRAIQEESEEDEEDDDDDEFEEQEEGDNSIY
>C_doughertyi_KNL2
MGDTNILPSRVQNVLDMEIVRLNLWSINFNASNIKLEGYVRTDDGSMMRKVCSEIICKR
MNSTMLFDVSGRFYELAGQIDKEYQLKMGMPSRVIDEFSNGFPENWAFLIKSCLSLPPRS
TLRPIQDVPREPLRSRAEQPIVTLADETELVENERKNSDSERDKKRRDREEQRVRENEKR
VEDQREKEHQEAIAAAAAAEKKRKEEEESAADYTFRVPGTDLGNPITPIRFTRGGQLNGHI
KSVFQQTPVRGKSSGPLASSTPQAPPPENPRRLSHLERASVPAAPLPEKDKTPAQRPVRST
DYASDSDLFAVPQLPPAKNIRPSVSSDSAFDFLDQMDSLFDSAVIEQTPGRDQKARKIAPP
SPDARRRFSSREPERRSDYESSRYINRYNDDYNMSRMSRRDEPWNRRHDEGRDESRMSR
KRGHFNSPPHDSDYGRRDNRGRRPEYYESNSKYDTKRYRPRETSSSSGRSVRFEDDYKR
RDDFREPSYSRSYRMYDDRRGRESSRDLEDKNKLDAILRREKELVDRLENTSKPSSSRRH
DLYSSDEDPIDLADEWDRENQEIMDNSMMFGDGISKKKRRSGAGRPQKKQPTGSTQPKP
ARKPAEPRKKARRSPPPADNLNDSIASNRPRRACVTPSTPAPKRITWPKRDLDKLKHVIEL
KKPTGSDADWAEVARLLAKEGVEAEVVKQAAITKLKWKEPVQIDNKEQEEEEEKKRRR
GVAARVKEGVKMREELMSGKQEEDRALRVDSMEDYQPEDVAADQSLLTMGTPLAAKK
KGGTRGSIMPKPVEDSPMSRGRNSTFNSPRLDQTKAKEVETTLKYVQHLSMMQGRPSS
RANKSYLDKSSRGSKNTSLSLEQGARKALKIINRGTTIHEDDEDDDEEEDDDDEFEDQEV
GDTSNY
>C_sp54_KNL2_paralog1
MGDTGIIPFRRKDVLERVEAVLNMEIVRLKLWSIKFNTSVIMLEGFVRSEEGTMIQKVCSE
NICKRISSTVLFDESGRFFELAGQIDREYQQKMGMPSRMIDEFINGFPENWADLINACVL
NQQRSALRPIQAAPKEPLRTRAEPIVTLADETELIGEESKKDMKRREREEQREREQEKKL
ATEKERKRQEAAAVEKKRREEEEAEEAANFTFRVPQSQSGEAITPIRFKRGNGKKGENTR
TIFDKTPVRQNSSGPLASSTPQAPPPQHPRRLSNIENNQPLRSPSQKTQLAASIPERETKFA
SDADLFAVPKLPPSKTTRPSTTSAVSSEGALDFFDEIDALFDTAIIEQTPTRDRKSRKPIATS
PSPDPRRRSLSRDRKSERYGSYESSRYTQRYNDDYDNSRMSRRDVTFKRHEGGRDDSR
MSRKRGYYRSPDDFEYSRRRDDREKDYYAYDSRHDTKRSRPRETSSSSGRSVRFEDDYR
RRGGEFRESKDSRDFRDYDDHRNRKNSGEHEDKKKLKDILRREKELLARLQKSQKSSSS
RRDSYSSEEETFDMADEWERENQEIMDNSMMFGDGLSKKKNRRSGGGGAAKKQLART
SQSKPHRKPAQPRKKTLRDPLETDDLNDSIASHRPRRACVTPSTPAPKKISWPRRDLDRL
KHVIELMKPTTTDKYWIEVTRLLAKEGVDPEAVKQIAVAKLKWKEPVQNEEVLKKDEEE
EKKRRRGAVARIKEGVKMHEELREGGDKKADDLESGVEAMEDYQPEDVAADQSLLAL
GTPIAAKKKGGTRASIMPKPVEDSPMAARGSNSTFNSPILDQTKAKEVETTLKYVQHLS
MMQARPSSRANKSYLNKSSTRGSRNTSMSLEQGTRKAMKIINRGRTIHEDDEEDDEEDE
DDEFDEQDGNTSIY
>C_sp54_KNL2_paralog2
MSEDPDEFLNLRSKVTDPWSVELIGLDAIQQQLDVMGENVAEVILKAEQSFNKKSLPFY
EKRKKFTSKIDNFWKKAFLNHHFLSKAIPEEQEDLIEVLRDLEVQEFDDLRSGFKIIMTFD
VNEFFENKVITKSYHLKSKPPKTQITEIQWKENQKPHMTSEDGDTTITFLEWLTHAVPPK
SDEIAHVIKDDLFINPLQYYMKPDIQENGQPIRRIPKINSSMMEVDDTAEGEDSDEKEEEA
TNCTYRVPQSQNGEAITPIRFKRGNGKKRENTRTIFDQIPMRGNSSGPLASSTPQAPPPQQ
PRRFSNIENNRPAGPIPERETQFASDADMFAVSKRAPSKTTSSTSTSAVLFEGEGAYHFLDE
VDDFFDTLVIETRERKPQRRRIATSPSPDPRRRRSSSRDRERYGSYESSSRYTRRYNDDDY

DDSRMSRRDIATFKRRHEGGRDNSRMSRKREYYRSPDDFEYRRRRNDDRDKDYYNAY
DSRHIAKRSRLGETSSSSGRSVRFEDDYRRSKDPRNFRDHRNRTISEKQLLARLQELRKPS
TSSHRDSYSSEEEETFNMADEWERENQEIMDNSMMFDDGLSKKKIKRRSGRGAVKKQL
SRTSQSKKPHRKPAEPRKKTIRDPLETDDLNDSIASHRPRRSCVTPSTNPVPKKISWPKRD
LDRLKHVIELMKPTTSDKYWIEVTRLLAKEGVDSEAVKQIAITKLKWEEHVQNEEMMK
MDEEEEKKRRRGAVARIKEGVKMHEELREGGDKKADDLKESGVEAMEDYQPEDVAAD
QSLLALGTPIAEKKKGGTRASIMPKPVEDSPMAARCSNSTFNSPRLDQTKAKEVETTLK
YVQHLSMMKARPSSSKTNKSYLNKSSTRGSRNTSMSLEQGTRRAMKIINRDRTIHEDDE
EDEADEDYEFEEQDGDTSIY
>C_inpoinata_KNL2_paralog1
MGDTNILPARVQNVLDMEIIRLNLWSIKFNSSNIKLEGFVKNEDGTAMQKVCSEIICKRM
NSVMLFDVSGRFFELVGQIDREYQQKMGMPSRIVDEFLNGFPENWPYLIKSCISVDSRSS
LRPIQAAPREPLRSRAEPIITLADETEVVVGDNKNSDTGNGRKHREQVERNSNDQVQMP
REANMKNNIQEEEDDIANYTMRAPVSQNGEAITPIRFTRGNGKKGANTRSIFESTPARGT
SSGPLQSSIASVPPPQQSRRLSTNQNSQLQSQCQKIQVAPSQNEKPPFREPQFSNDTDLFAV
PRLQENKNNRLPSLSAASSDGAHDFFDEMDTLFDTANIEHTPARDRNIIRRSRVASPSPPL
RRYRTSSRDCERDIYEPTRHSNRPDNGFNINRNRSDAFRKNEGRNDGSRMSRKRGYFYS
PDKYRDNFSRRIKNIDYEQMHDNKHSRERENSPYSGRSIQFEDVYRKRNYGDYDDFRKR
SRDEEEDRKLNEILRREKEMMARLQQHRKSVSSFPKSYLSDDDEFDMADQWERENQEL
LENSMMFGDGLPKKRKMSRDCKSERLKNPKESVKKPVQSQKSNRVKRLETDNANDSIA
LSRPRRLCATPSTPAPKKIVWPKRDLDRLKRIIELKKPTAANADWAEVARLLAKDGVELE
AAKQTAITKLKWKEAVQNEELLNCEEEDKKRRRGLAARVKEGVKMHEEIREGGENKGG
DLQNGVESMEEYRPEDMAADQSLLAQKTPIVVKRKGGTRASILPKPVEDSPLARGSNSA
LNSPKLDQTKAKEVETTLKYVHHLSQLHANPTCRGNKSFANKSLPGGRKNTSMSLEQG
TRKAMKIINRGQTFYEDEEDEDDECSENDEDTSMY
>C_inopinata_KNL2_paralog2
MGDTNIVPARVQNVLDMEIVRLNLWSIKFNSSNIKLEGFVKNEDGTAMQKVCSEIICRRM
NSVMLFDVSGRFFELAGQIDREHQQKLGMPRRIIDEFLNGFPENWPYLIKSCISVDSRSSL
RPIQAAPREPLRSRVEPIITLADETEVVVGDNKNSDTGNGRKHREQVERNSNDQVQMPR
EANMKNNIQEEEDDIANYTMRAPVSLNGEAITPLRFTRGNRKKGANARSTFESTPAEGTS
SGPLLSSITSGPPSQQSRQLSINQNSQLPSQCQKIRVAPSQNKKPPFREPQFSNQDLFDEMD
SLFDTANIEHTPVRERSINRSRVASPSPPRRRYRTSSRDCGRDIYEHTRYSSRSDNGFNISG
NRSDAFRKKEERNDESRISRKRGYFISPDKYRENFCRRIKNIDNDQMPDNTHSREGENSL
YSGRSARFEDVHHKRDYGEYDDFRKQSRDEGKDRKLNEILRREKELMARLEQSRKSAS
SFQKSYSSDDDELHMTDQWERENQELVENSMLGDDLSKKRKVSRDGKTKQSERLKNH
NESVKKPVQSEKRNRVKRLETDNANDCIALSRPRRLCATPSTPAPKKIVWPKRDLDRLKR
IIELKKPTAADADWAEVARLLAKNGVELEAAKQAAITKLKWKEDVQNEELLNCEEEDK
KRRRGLAARVKEGVKMHEEIREGGGLQNEVESMEEYRPEDQSLLALKAPTVVKRKGGT
RASVLPKPVEDSPLARGNNSALKSSKLDQAKAKELETSLKYVHHLSQMHANPSSRGNK
SHANKSLPCGKKNTSMPLEQGTRKSMKIINRGRTFCEDEKDDSTDNDEDT
>C_elegans_KNL2
MGDTEIVPLRVQNVLDSEIIRLNLWSMKFNATSFKLEGFVRNEEGTMMQKVCSEFICRRF
TSTLLFDVSGRFFDLVGQIDREYQQKMGMPSRIIDEFSNGIPENWADLIYSCMSANQRSA
LRPIQQAPKEPIRTRTEPIVTLADETELTGGCQKNSENEKERNRREREEQQTKERERRLEE
EKQRRDAEAEAERRRKEEEELEEANYTLRAPKSQNGEPITPIRFTRGHDNGGAKKVFIFE
QTPVRKQGPIASSTPQQKQRLADGANNQIPPTQKSQDSVQAVQPPPPRPAARNAQFASDA

DLFAVPKAPPSKSVRNLAASNVDIFADVDSVLDTFHFESTPGRVRKPGRRNVSSPSPEPRH
RSSSRDGYEQSRYSQRYEHDNSRWSRHNATYRRHEDESRMSRKRSIVRDDFEYSRRHDD
GARRRDYYDADIQGDSKRYRGRDASSSSGRSVRFEEEHRRHGDEYRDPRGPRDYNDYG
RRRNHANSRSGEDEEKLNAIVRREKELRNRLQKSQKASSSSYRHRSNSSDAEESLNEWD
IENQELLDNSMMFGDGIPKRSNARKDKFVKKQATRSKPANSTKSPAQARKKKRASLEDN
RDLNDSIACNRPRRSCVTPVAKKITWRKQDLDRLKRVIALKKPSASDADWTEVLRLLAK
EGVVEPEVVRQIAITRLKWVEPEQNEEVLKQVEEVEQKRRRGAVARVKENVKMHEELR
EGGNHRAEDLQSGVESMEDYQPEDVAADQSLLALRTPIVTKKRGGTRASIMPKPVEDSP
MSRGNNSTFNSPRLEQTKAKDIETNFKYVQHLSMMQARPSSRLKKSSSMNNSTYRGNK
NTSISLEKGTQKALKIINRGTTIHEDDENEDNDDDDDMREEDTSIY
>C_oiwi_KNL2
MGDVEILPQRVQNVLDMEIVQLKLWSIKFNSTNIRLEGFVRNEEGTMMQKVCSENLVKR
MNSTMLFDVSGRFYELSGQIDREYQQKLGMPSRIIEEFANGFPANYAILINSCLNSDQIRSI
RRPIEAAPRDPLRPIVTLANETENPKIPEKIPEKSILEEEKEENEADFTLKAPVSTDGNAITP
IRFTRGNGRALARTVFEQTPVRGKASTSGPLASSTPQPAPPPIPRRSSTTENRNPPEAPPPK
VQYATDSDLFAVPKLPAPRQQAPPISLGFLDDFDDIFADAVIQKTPKRGRPPKREEEDSRRF
YEDESRMSRKRGYYRSRSRSPDYDRRHRDRDYSDFYGREDSKRPRRRDESSSSGGGGK
SVRFEREDRHQRHHPHRRDRSNSRESQRHHHRRHHDDDYNVSRSRSHEKEEKRRLQEL
LRKEKELEARLQSFRRPPTSSSSSSEDSDEDEMAGEWERENQEIMDNSMMFGDGISKKR
RRQSGEKKRKQPSRPKPKSAAPAPSATRKPAKKNKRSPAETDSLNDSIASHRPRRACATPS
TPAPKKVTWPKRDLDRLRRVIELKKPSGREEEWVEVARLLQKEGVNPMEVKEAAIGRLK
WKEPVEKEHTEEEEEEEKKRRRGVAARIKEGVKMHEELREGRQQNREDALKNRVEEVE
EFQPEDVEADQSLLALTTPVALKKKGGTRASIMPKPVEDSPMSRGNNSTFADSPRFDQTK
AKNLETTAKYVHHLSILQGRPSSSAANRTQMNRSTTRGGGSKNTTLSLEQGTRKAMKMI
SEGRTIHEDSEEDDEDDEDMDDDVFN
>C_kamaaina_KNL2
MGDTEILPLRVQNVLDMEIVRLKLWSIKFNSTNIRLEGFVRNEEGTMMQKVCSENLVKR
MNSTMLFDVSGRFYELSGQIDREYQQKLGMPSRIIDEFANGFPENYAILINSCLNSDQIRSI
RRPIEAAPREPLRQKQEPIVTLANESEIQNPQKTKIVPSKSPKNSKILEKSIQEEEEEEEDPA
NFTLKVPISTNGDAITPIRFTRGNGRGMARTVFEQTPVRGKPSSSSGPLASSTPQPPPPPRR
SSTTENRNPKTEQAPAPKPPQNPQYATDSDLFAVPKLPAPKQQAPPISLGFLDDIDDIFADA
VINKTPRRGRIGRPTSPSKQDVTSDSRGRFYESRRDEMVDESRMSRKRGYYRSRSRSPDY
DRRRRDHRDDHRYSDYYERDSRSDSKRPRQREESSSSGKSVRFERREDRHPDRRRHRST
SRESHRHRRRDDDYDVSRSRSRESRKNEKRKLQKIMKREKELMSRLQNSRRHSPSSSSSS
SSSEGSDDEDEMAGEWERENQEIMDNSMMFGDGISRNRKRKSEGKKRKQPSRPKQIKS
EAPVPAKKSAPKRNKRTPAETDFLNDSIASNRPRRACATPSTPAPKKITWPRRDLDRLKRV
IELKKPSGKEEDWVEVARLLAKEDVKPMEVKEVAIGKLKWKEPVEKVQTEEEEEEEKK
RRRGAAARIKEGVKMHEELQGGNPLKREDALKNRVEDVEEFQPEDVEADQSLLALTTP
VALKKKGGTRASIMPKPVEDSPMSRGNNSTFANSPHFNQTKAKELETNYKYVQHLSMM
HARPSSSAANKSQMNRTTTRGGSKNTSMSLEKGARKAMKMISEGRTIQEDSEEDDDDD
KEEELEDDVFN
>C_waitukubuli_KNL2
MGDTDILPVRVQNVLDMEIIRLNLWSIKFTATNIKLEGYVRNEEGTMMQKVCSEVICKR
MNATMLFDVSGRFFELSGQIDREFQIKHGMPSRIVDEFTNGFPENWAFLISNCLTTEQRSA
IRPIQAAPVQPLRLREPIVTLADETELGDRTVVRKENEKDQRKKNREEEEKKHRLLEEKR
KKEQEDKIAAGKRKESREEAERKQKEEEDAANYTFRAPPSFGPDAITPIRFTRGTQGRGG

IRIFEDTPQRTTSGPIDSSTPKPPPVHEQLPIKVQQKPEVHQQPEPRQTEKAPQTFASDSDL
FAVPRLPARAPISSELPTGFDFYDEMDAIFDNAKIDKTPVMESRRKQIPRRFQYEPSPPPIL
HGQSSSSSQAYEPDFDSRSYYNGRYDHERSRMSRRDDRYVSERDDSRISRKRILYSPDHH
DDSDHSRRRAYGSRSRYEDDFPKRSRTRETSSSSGRSVRFEDDYRSNRYHNDREHRDRD
RTREQQERDMYESQKLKEIMRREKELEAKLRLKKTATRRVSYSSDDSADDTMNMTEEW
ERENQEILDNSIMGDRKTSRKRNNAPKREAVKPKKVDKPARAPAKRTKKEPRQAPDLND
SIASNRPRRACATPATPAPKRVTWPKRDLDKLRHVIDLKKPSASIEDWTEVARLLKKDGV
EPADVKLIAETKLKWKEPVQDEDVLLQEEEEEKKRRRGVAAKVKEGVRMREEMREGG
AKEDELRNRVESLEDYQPDDMDADQSLLALTTPVAAKKKGGTRASIMPQPVEDSPAVKG
NNTSSFMNSPKLDSTKVKEVETTLKYVQHLSTMQARPGSSMNKSYMNASSSRGRNASIS
VEQGTLTTHIVVSIFLRMQMTFHLTPSDTVLFESWKVTDGWTMMGACALVVVAGVFVE
AIKRYRKKINQDQMIREQLVYEPFSNRLFASML
>C_panamensis_KNL2
MGDTDILPVRVQNVLDMEIIRLNLWSIKFSATNIKLEGFVRNEEGTMMQKVCSEVICKR
MNSTMLFDVSGRFFELSGQIDREFQIKNGMPSRIVDEFMNGFPENWAFLISNCLTTEQRS
AIRPIRAAPVQPLRPREPIVTLADETELPGDRTMSKKETERERKNKERDEEERKQALLEEN
RQRDRERELQKEKEREAVERRKKEEREEAERKRKEDEDAANYTFRAPPSFGSDAITPIRF
TRGKKGAVVTRIFDDTPVRSTGQPLASSTPQRSLPVKEQPPQLPKQLKEDSKEQQPPPRQ
VERPMQSYASDADLFAVPRLPSRAPAPTGLSGGFDVLDEMDSLFDNAKIEKTPVMGSRR
KPIPRRYQYESSPPPFLRGQSSSQQAYERDLSSASYYSSRYDDNRSRMSRRDDTSRRYDS
GRDESRMSRKRGHYSPPDRRDDYAYNRRREDESRQRDSFRYDEDYNYDSKRSRPREASS
SSGRSVRFEDDYCSNRSRENKDRRNHERVRDQRDRNMYESRELKEIMRKEKELEAKLR
SKRSSVRRVSYSSDDSDNDTMADEWERENQEMLDNSMIMDRKPSNKRKNVPKKETSKP
KRVEKPTRAPAKKNKKDPVETDDLNDSIASNRPRRACVTPSTPAPKRITWPKRDLDKLR
HVIELKKPSGNLDDWAEVTRLLKKAGVEPTDVKQIAETKLKWKEPVQNEEALLLEEEEE
KKRRRGAAAKVREGVRMHEEMRNGGKKEDDLRSGVESMEDYQPDDMDADQSLLALA
TPVAAKKKGGTRASIMPQPVEDSPAVKGNNTSSFMNSPKLDSTKVKEVETTLKYVQHLS
TLQARPGSSMNKSHMNRSTSRGKNTSISVEQGARKAMKIINRGTTIHEGDEDDEDDDDE
VTEDDEENVIY
>C_nouraguensis_KNL2
MGDTDILPVRVQNVLDMEIVRLNLWSIKFTATNIKLEGFVRNDEGTMMQKVCSEVICKR
MNATMLFDVSGRFFELAGQIDREFQVKHGMPSRIVDEFINGFPENWAFLISNCLTTEQRS
AIRPIQAAPVHSLRPKEPIVTLADETELAGDRTMARKESEKDRRNREREEEKRKKLLEEK
RQREEELQRERDREEERSRQEEEEKRRQEEEEAEKLRKEEEANNTFRAPKSFGTDAITPIR
FTRGNSKKGLGAIKIFDNTPTPKRKNGGPLASSTPQPQRLPLVKEQSQKEKTPQPEKPPQ
QQQESSRPAERPKQTFANDADLFAVPRLPTKSTASSGFSSGFDMFDEMDTLFGTAKIDKTP
VMESKRKPIPRRYEYESSPPPFRGQSSSSQTYQRDHDSRFNFNDRYDDDRSRISRRDGTFS
RYDSGRDESRMSRKRGLYSSPDQRHDYEYSRRKEDESRYRDRSRYENRYDPKRSRARET
SSSSGRSVRFEDDYRSKKYREDRREHDRTRDHRERDMYEDRKLKEIMRREKELEAKLHS
QRTRRVSSSSDDSADETVNMADEWDRENQEMLDNSMMMDTKTSRKRKNVPKREFVKP
KSVEKKPARASAKKNKRDPIETVDLNESIASNRPRRSCVTTAPKRITWPKRDLDKLRHVI
DLKKPSADLDDWAEVTRLLKKDGVQPAYVKLIAETKLKWKEPTQDADVLRLEEEEETK
RRRGAAAKVKEGVRMRQEIRGGGEKEDDLRRGVEAVEDYQPDDMDADQSLLALATPV
AAKKKGGTRASIMPQPVEDSPLVKGNNTSSFMNSPKLDSTKVKEVETTLKYVQHLSTM
QARPGSSMNKSSLNRSSSRGKNTSISVEQGTRKAMKIINRGTTIHEDDEEEEEDENSEDD
EENAVY

>C_becei_KNL2
MGDTDILPVRVQNVLDMEIVRLNLWSIKFTATNIKLEGFVRNDEGTMMQKVCSEVICKR
MNATMLFDVSGRFFELAGQIDREFQIKHGMPSRIVDEFINGFPENWAFLISNCLTTEQRSA
IRPIQAAPVQPLRPKEPIVTLADETELAGDRTVVRKETEKDRRNREREEEKRKKLLEEKR
QREEEIQREKDREEEKRRQEEEAERKRKEEEEANNTFRAPKSFGTDAITPIRFTRGNAKK
GLGAIKIFDNTPTPKRINGGPLASSTPQRPPPVKEKTPQPEKPPQKQQEAPKPAEISKQTFA
NDADLFAVPRLPTKSTASSGFSSGFDMLDEMDTLFETAKIDKTPVVESRRKPIPRRFEYES
SPPPFRGQSSSSQAYQRDLDSRSYFNERYDDDRSRMSRRDGTFSRYDSGRDESRMSRKR
GLYSSPDQRHDYDYSRRREDESRDRDRSRYDYRYDPKRSRARETSSSSGRSVRFEDDYR
SKKHREDRRDREDRRREYDRTRDQRERDMYEDRKLKEIMRRERELEAKLQAQRKSARRV
SSSSDDSADETVNMADEWDRENQEMLDNSMMMDTKTTRKRKNVPKKEFVKPKSVAKP
ARASAKKNMREPLETVDLNESIASNRPRRSCVTAVPKRITWPKRDLDKLRHVIDLKKPSA
NIDDWVEVTRLLKKDGVQAAYVKLIAETKLKWKEPSQDADVLRLEEQEETKRRRGAA
AKVKEGVRMRQEIRGGGEKEDDLRRGVEAVEDYQPDDMDADQSLLALATPVAAKKKG
GTRASILPQPVEDSPLVRGNNTSSFMNSPKLDSTKVKEVETTLKYVQHLSTMQARPGSS
MNKSSLNRSSSRGKNTSISVEQGTRKAMKIINKGTTIHEDDEEEEDEDDENSGDEEENSIY
>C_yunquensis_KNL2
MGDTDILPVRVQNVLDMEIVRLNLWSIKFTATNIKLEGFVRNEDGTMMQKVCSEVICKR
MNATMLFDVSGRFFELAGQIDREYQIKHGMPSRIVDEFMNGFPENWAFLISNCLTTEQRS
VIRPIQAAPVQPLRPKEPIVTLADETELGDRTMAKKETEKERRNREREEERKQALLEEKR
QRDHELRQEKDREEERKRKEEEDAANNTFRAPKSFGADAITPIRFTRANGKKGHGAFKI
FDNTPKRNTNEPLASSTPQRPLPVKEPEQPPQKQPETSRPAEPSRTYASDADLFAVPKLPT
KSTGSSGISSGLDFFDEMDTLFETAKIDKTPVMESKRKQIPRRFEAASSPPAFRGQSSSSHA
SNRDYDSRSYLDDRFNDDRSRMSRRDGTFNRYDSGRDESRMSRKRGHYSPDQRRDYEY
SHRRDDESRYRDRSRYEDDYRYDPKRSRAREASSSSGRSVRFEDDYRPNYHREDHRQR
DRTRDGNDVYENRKLEEILRRERELEAKLRAKKKPARRVSCSSDDSEDENVDMADEWD
RENQEMLDNSMMLNRKTARKQKNVPKRDVAKPKRVEKPIRAPIKKKTRDPVETDDLND
SIASNRPRRACVTPSTPAPKKITWPKRDLDKLRHVIDLKKPSANIEDWVEVTRLLKKVGV
EPADVKGIAETKLKWKEPTQNAEVLRLEEEEETKRRRGVAAKIKEGVRMREEMRNGGE
KEDDLRSGVEALEDYQPDDMDADQSLLAFATPVAATKKKGGTRASIMPQPVEDSPLVNG
NNTSSFMNSPKLDSTKVKEVETTLKYVQHLSTMQARPGSSMNKSYMNNSSSRGKNTSIS
VEQGTRKAMKIINRGTTIHEDDEDEEEDDDEVSEDDEDNAIY
>C_macrosperma_KNL2
MEYIVMGDTILPVRVQNVIDMEIVRLNVWSIKFTATNIRLEGFVRNEEGTMLQKVCSENI
CKRMNATMLFDVSGKFYELAGQIDREFQIKHGMPGRVVDEFLNGFPENWAFLIGTCLTT
EQRSVLRPIQAAPVQPLRPKEPIVTLADETELTGDRTMAKMETEKERRAKQREEQRKLE
KEREEEAVAERKDKEKREREEAERKRKEEEDAANYTIRAPNPLGNEAITPIRFHRGKGGG
PRFFNVFDKTPKATSSAPLASSTPQARPPVRKEPEQPVSSKPVERPVAFASDADLFAVPRLP
AKGPTTSSSSAGAPLQDDFDFLDEMDTLFDTAKIEKTPVMDTRRKPIIRRFEHRSYSPQH
GQSSSSQAYEMDYNDRREPYYHDRYEDDRSRVSRRDFGGDASRMSRKRGYYQSPENRD
RDEYEDRRRREHESRYSQQIYSREFEMNGWPRRETSSSSGRSVRFDDDYHYSDRHRNPR
DKTRDERDREMYERQKLKEITRREKELEAKLKSRKVSRSSEESADDSMDMAEEWEREN
QEMLDNSMMMMPDRKSSRKQKVVPKREKPKKAAEKKPAARAPRKKNIREPLLEEEDD
PNDSIASSRPRRACATPSTLAPKRITWAKRDLDKLNHVIELKKPTANLDDWVEVTRLLKK
NGVEAADVKHAAETRLKWKEPVQNDERAQEDEIEEERKRRRGVAAKVKEGVRMREE
MREGGEREDEMETRVGAVEDYQPDDMDADQSLMALGTPAVKKKGGTRASILPQPVEDS

PIVRGNNTSSFMNSPRLDQTKAKEVETTLKYVQHLSTMQARPGSSMNKSHANKSYMNK
STSRGGKNTSISVEQGTRRAMKIVNRGTTIEEDEEEDEEENDDDEEDTSVY
>C_sulstoni_KNL2
MEENGILPTRLQNLADMDVIPLELWSIKFYGTRYKVEGIVRPEDFTTAQKFTSEFIIKRITS
TALVDASERFIELVGQIDRTFQIRNGMPEEIVGKFQNGFPRDWWELLKPCLAAGQKTSTM
RAIEAPPLKTSTTSSTVQNVQKTSSEPIVTFADETEVQSDREKERRHRQEEREVQKVLEK
DDDDDADDADYTIRQPRAIDGGENTPLRYKRGHSGGPQNRNRVFADTPQRTAIPLAAST
PNAPPKIAKPPVKFASDADLFAVPVPPTRRPAERKERSESPEGCDGLFDKISEVFSNVNFPV
TPLPAARVRAHFSPAERSEGSYYYRDDRYDEDRNRTSRRDRSRSREYDESRTSRKRGYY
RSPEGYEYSRRRDDRRRDRYEASSQRHGSSSSGRSVRFEDPYEYDVEQRRRRREDQKLR
EIERRREREIEARLERSKRSRRYSSDEQSEDMADEWERENRQILDDSMLFRSAAGKATNKK
RAGRPKKEVKTRSPPVQKPQKARKEPRKRRSEPAETVDLNESIASNRPRRACVTPVAPPP
PKRVTWPKRDLDKLRRVIEVKKPTGDEHDWVEVARLLAKDGVTPTDVRTIAETRLKWR
EPTRDPKVLEAAEEEETKRRNGLAARVKESIRMREEMREGRQEEDVELKPGGENYEPAD
VSADESLLAIGTPRVPPKKKGGTRASLIAPPVEDSPMTRRSGAAATVVASPKLDATEVAAT
QKHVHHLSMIQARPGSSMMNTSRANKSIRNGKNTSISVEQGARRAMKMVTRGRSTIAE
EDESEEDDDENEEDVTAEEDEIY
>C_afra_KNL2
MEENAILPTRLQNLAEMDVIPLELWSIKFYGTGYKLEGIVRPEDFTSAQKFSSELIIKRITQ
TALIDASERFIELVGQIDRTFQLRNGMPEEIVNKFQNGFPKNWYELLKPWASQNQKTSTW
RAIEAPPVAPPNPQIPPIPQKAPCEPIVTFDDETEVQSDKEKEEARRRRQLEVEEAEKQRKI
LEDLNDDADYTLRNPRSIDGVQNTPLRFKRGNPNFERNPRVFNDTPVRTGGGPPLAASTP
QAPKVSDPIPKAANYASEGDLFSFAVPTQPPSQRPTRTTDEDRKHQADRRRRSESLGGSD
DGGGGSGLDLFSQMSEVFSNAKFPITPLPAAPPRSSRFERERAERSDYYREDRDRERYRD
PSRESRERDESRMSRKRGYYRSPEGYEYSRRREDRHRDRYESSYQRYGESSSSGRSVRFD
DGYDDPYEREEEERRLRRHRDNQKLREIMRRREREVEARLERSKRSRRYSSEDDGPEEND
DDLADEWDRENREILDNSMSFARGGSGRAGKKRGAGRPKKTSAPKSRSPPTQQKPRKQ
QQAAKKRRSEPAETDDLNDSIASHRPRRACTTPKAAAPKRVTWPKRDLDKLRRIIEVKK
PTGDADDWVEVARLLAKDGVEPADVKTCAETRLKWKEPTRDPEVRQKEEEEEKKRRN
GLAARVKQGIRMREELREGYEEDVAPPPTSSSQEPDYEPDDVSADESLLAMGTPLPPPKK
KGGTRGSIVVLPVEDSPMVRRSGVAASPKFDASDLAATQRHVQHLSMIQARPGSSMNTS
RASKSILGGKNTSISIEQGARRAMKMMNRGRVAIAEEDEDSDSDDSDDDENDSNDNAEY
IKIKVVGQDSNEVHFRVKFGTSMAKLKKSYADRTGVSVNSLRFLFDGRRINDEDTPKSLE
MEDDDVIEVYQEQLGGGI
>C_afra_KNL2
MEENAILPTRLQNLAEMDVIPLELWSIKFHGTGYKLEGIVRSEDFTSVRKFSSELIIKRITQ
TALIDASERFIELVGQIDRTFQLRNGMPEEIVIKFQNGFPKEWYELLKPWASQNQKTSTW
RAIEKPPVVPPNPQIPPLTQKTPCEPIVTFDDETELRSDKEKEEARRRRRLEVEEAERQRK
LLEEGSDGAEKRGYNRSPAEKNNRERSERSRRCSSEDNRSDGELADEWERVNSEILDNS
MSRVAGRPKKMLAFGRALTQADVEKFAIRLEIAGARIPAVVASEKYVVTTLKNIPEASRK
LGFALSAFDHLEQRIELKIHGMHRHSDFVVLRRDEGEFSEAPVYRNEEVGDKYFVVKRSI
DGKKIVCHGKVTSVGATGFAFVSEPYEGFPVANGDGAFAFLDGKLLGVVKDTKEQGRV
LEFVSIYLVHFAIDIIFHHQPFSV
>C_sp49_KNL2
MTERTIVPVRVQNVMDMEILRMNVWSIKFNSSSVMVSGFVRTPDGNSMNLVTSEAICKR
MNSTLLFDISGRFYELLGQIDKSTQLKAGMSSRVMEEFLNGFPDNWVFLVGQCLTTEQR

SAIRPIQAAPREPIRPKSPIVTLADETELREQGKAQKEQKEDADATFRAPTAFFHGAVTPIR
FTRGNGATGRSQRVFESTPSREPSSSRPLASSTPQPQKLPVLPPVAPRDVPAAQPRTFATDD
DLFAVPKLPASSSGLDFLEDMEAIFNGAHINRTPSHAAPRARVPREFRSRSPSYEQRYQDR
RYQDESRVSRKRGMDRSPEYYEMERRAEESRRSHESKRARPRETSSSSGRSVRFEETYRR
DRHHQNDRSHRSHHYEQEKEHKRRLREIERREREVEERLKRNRKVSSDSNSSSGEEDGS
FNEWDVENREILDRSLMEASGRERRPKREPTSKKSTSKKPAAPKPKPTKNIRKKRDPIET
DDPLNTSIASSRPRRACATPATKKITWPRRDLDRLKRVLEMKAPTGRSEDWVEVARLLA
KDGVNSEEVRQIAESKLRWKEPEHHPEEHQEDPLNDSQLIEKKRRNGIAARVKESVKMA
EELQTRPPNSDEMRGRVEEVEEYQPADVSADESLLALSTPQVIKKRGGTRKSTLPEPIEDS
PMARGANSSVNSPRLNATKAKEQETTLRYVQHLTTMQTRPASRMNKSAMNKSTASAKN
TSMTVEQGTRHVQKMMKKAMAIDEEEEDDEESDENQEETSFY
>C_sp25_KNL2
MTEQTVVPVRVQNVMDMEILRMNVWSVKFNSSSVMVVGFVRTPDGNSMNLVTSEAIC
KRMNQTLLFDVSGRFYELLGQIDKDTQLKQGMTSRVMEEFINGFPENWVFLVGQCLTTE
QRSAIRPIQAAPRQPIRPKSPIVTLADETEVAAQAQAQREQKEQKENEQRRRREDKEKED
ADATFRAPASFANGAITPIRFTRGASVASRNQRVFDATPTRQAPLASSTPQQKVQKVPTIA
ADEDLFAVPKLPVKKGVAPASDLDFLDDMDALFTGARINRTPSHRVPPKPQQRRRSRTRS
RSPSIDRSRSRRYEEEFDSRMSRKRGYDRSPDYDDYRREESRRRFKESKRSRPRETSSSSG
RSVRFESDYHRQRHDDRRRHEDREELERQERRRLKEIERREREVEERLKRSRRKNYSSSS
SSDEDQDGSFNEWDAENREILDQSLLASTGGSSRSASGGNSRGRLPKREPSLKKAKPAA
AAPKPKPRKKKEVVDLNDSIASNRPRRACATPATPAVKKITWPRRDLDKLKRVIELKKPS
ASEEHWTEVTRLLAKEGVAWTDVKSIAETKLKWKEPRQEMVEVEEEEEEERERKRRRG
IAAMVKESVKMCEEIQTRNVQEDEMEGRVEEVEEYQPADVSADESLLAMSTPQITKKRG
GTRKSTLPEPIEDSPMAKGTNSSWNSPRLNATKAKEQETTLRYVQHLTTMQTRPGSSMN
KSYMNKSTASTRGKNTSLSVEEGTRKVQKMMKKAMAIEEEDEEEEDSDDEGETSLY
>C_imperialis_KNL2
MTESTVVPVRVQNVMDMEILQMNVWSIKFNSSSVMVVGFVRTPDGNSMNLVTSEAICKK
RMNSTLLFDVSGRFYELLGQIDKTSQLKQGMSARVLEEFLNGFPENWVFLVGQCLTSEQ
RSAIRPIQAAPRQPLRQKSPIVTLADETDLGDRHKREEVENEARKREERKMIERENEEKR
RAEERRQREEQEKEDANCTLRAPASFASGDLTPIRFTRGNGPCNRKNRVFEDTPSRETTS
RPLAASTPQQKPKPVDPPREVPRDVPARPKTNFATDDDLFAVPKLPANRAAKPSSDLDFL
DDMDALFAGAHINRTPSHAPKPVSRIRQSPPRRRSPSENRSRHEYSQRSNRYDDDYDVRF
NRNGKEESRLSRKRGYYRSPDEYYETRRDWGRDRSRDRSRQRYDSERYDSRYDSQVSR
PRETSSSSGRSVRFEDDYSQRNNRRYEHEVRERKQLKDIERREREVQERLQRSKKSAPAR
QATYSSSEDDDDASFNEWDAENREILDQSLLASTAGTSRRRSAPKREPKSKKPKPAPAPA
PKRKPSKPRKQRDPVETDDLNDSIASNRPRRACATPATPAPKKITWPRRDLDRLKRVIEM
KKPTASDGAWAEVARLLAKDGVTPADVKQIAETKLKWKEPTQDPVEIEEQEEVELKRRR
GIAAKVKESVAMREEMQRRAVDEDELNGRVEEVEDYQPADVSADESLLAMGTPQVTKK
RGGTRKSTLPEPIEDSPMARGNNSSFMNSPRLDATKAKEQETTLRYVQHLSTMQNRPGS
SMNKSQLNKSSTSRGKNTSMSVEQGTRKAQKMIRKAMAIEEEDEEDDEDDEDENREEE
ETSFY
>C_japonica_KNL2
MAETGVLPLCVQNVMDMEIVRLNLWSIKFYAESIKLEGFVRNSDGNMMQKVCSEAICK
RMTSVLLFDVSGRFFELVGQIDHEFQVKNGMPSRIVAEFMNGFPENWLFLIKQCLPPTDE
QIPQRQIQIQSAPAPPKEPIVTLADESEHVSDHPKKIVENGRIRDEESADCTFRVPHPVAID
AITPIRFTRGVRKENAQRIFDSSTPKPQVEKQKKPLAAFDADLFAAPAPARAPPPTRAPPP

NPPGPSSEFVLLDEMDELFNNAEFDKTPGRTVAQPRRINETSRMDRKTSRTYSEDCDRKY
DDRRVECYSRSARENVEPNRNEFRMNRKRPSVNNYADEMNGGYNYRSNKYGRREVFE
SNGYHSDEFRSREAPSSRSYADRDYYSRKSYHRGEEQRNEKMYGDILRREMALETRLK
NGRSQRHVSESSEDSMDDSYMVDEWDRENQRILDSSIVSSFGRFETSPKKSRIYVKKERT
AKKARPKPVPPRRKKNSPDSPEMDDLNDSIASNRPRRACVVPSTPVPKRVTWPKRDLDR
LSRVIELKKPSSSDSDWTEVARLLAKEGVEGADAMKIAMAKLKWKEPVKDCETVQEDL
EMVDKKRKRGFAAKVKESVQMMEEMRDGGEQSGEMGGQGVEQAEDYQPADLSADES
LLALATPCAGKKKGGTRRSVMPAPVEDSPVIKLNECSYANSPKLDVTKCKEVETTLRYV
HHLSTVQARPGSSSSRSYKNRSVCLGKSLSLEQGARKAQKLISMAVGIEEEDEEEEDDDD
AELSIF
>C_tribulationis_KNL1
MENNPRPKRNSILKVRQEKNLLEMLDETKAAAAPENATRRVSFHQMKHVKEYDRHLG
KIIDATPTKEKAFDTISSDGTATPHVTRTDMDITGETTTTSRLFENAQATPTSSAHFGHHHH
HQQQHDGTMDMSMDTTIENHRPIVQNEGSADETARLFDVTREKTVMYEEVTKTETQTT
QTRTLISVYHERAAPDDTMALFDMTNRDEVDMSMDGGGVAGVAAARVDDTMAVFNTT
NTEQIDMDITQDQKAPSVPLDDTMAVFRSPAPPRLQKASGIQKTSNENSESERVDMDITR
NSSVLAPTDSDDPMDITGIQKAPEDVEPGSEAMDITMAAPNDTMDVFATPKRIQKTSGH
QKTPREEDDDDDAMELETPMKVKTGSNETFDLLQSPARAKIQNPSHQILVVEEEEEHVD
MDITQREAVEPSDDTMAVFRSPAKTSHETTRQVFDNESMDIESTLRPEDVVPEESPEDVK
RSDSHTMMSMATETSIVVVEENRCVSTMLQMSPIGRSEDVEIQKMSTTTTTMASDSQMS
LSSSIHNSSVTLNVSKTGDVGEEDGGASERSQIQTMVQEDADVTNPLQNLTTKEEETEAD
SEDVTMKEESRISAMNVSSVSRRRRSLLMDPLCRESPRRLALGNSMLSMSTAAATTLQG
GGGALAEYRHSKRMLNESVGSVGSAGGNQTLANLTATGRDIFQLNTSVRSPAHRPIVTH
HHHLPNTTVTSPDATLQQQQHVVPPKSPESFRLPQFDAAIVNVIYLTPEDDAQEPIPEA FE
FQKLLAQQEAYAHREIEEAVNGDETLQKVVDEIGGDVVKIRGHLERDVIRIAIGQAEKKF
LELRGEFAEEHVAKSDALVVELEAENARLAQRVQDARNVETLRAELHELQRRPTREEAR
RIRREHHEAKMESMRMQLASTRARYEAEMEMRAQRQRVEDALEEKREQIARLEDARE
KIKTEMAGRVRAIVPEE
>C_sp41_KNL1_paralog1
MENNRKQKRNSILKVRQETNLIDETTVVAPNANRRVSFHQVKHVKEYDRDHGKIIDATPI
KEKAFDTMSSDGASTAHTTRIDMDVTGDTTMPSRLFEGGGSAHFAHHNGHNGTMDMS
MDQDHTMSLENNQNETARLFDVTREKTMVYEKKTTTTRTTTRITVVGGGSSSCSSSGVN
DTMAVFNMTNRDGVEMEIEGGGGRVDDTMTVFNTTNTDQIDMDITQQQQQQEQREEQ
KALDDTMAVFRSPAPMRGIQKTDIQKTSRRVAENSDMTKIADDTLSLFKSPQKKEDLMDI
TGIQKTSSSEEGAEPGSEAMDITVAPPDDTMEVFATPKRLQKAQNAEDMMLLETNETFD
LLQSPARAKIQGGKQHLEHQEHHDDMDITQKEVEPADCTMALFRSPEKIPAQNQTRQVF
DDDDENMDIESTLLRPSEDTPEDVKPSESSIMTMSMATEASMVVVEEERCTVSTVLQMS
PIEDVEIQNPLQNPLSMTSSSEMSLSSSIHNSSVTLNVSKTAESPLSSQIQTMILETDSTNTI
QNLTTTPIKEDSEDSEDVTMKEDESRVERVQEDSVMDVSKETNESRISNVSRRRRSLLLD
PSMCRESPRRLALENSMLSTVGAYPGAGGAGGGALAEYRQAKLNESVESSGGAGNRTL
LANLTATGRDIFKMNTSIRSPAHRAAATTMISTPEDTSSHIMSHAPKSQPPESASFHLPQFD
AAIVNVIYLTPEDVVAAQEEPIPEAFEFQKVLAQEEANAHRDIEKAVNGEESLQKMVDEI
GGDVVRMMGHLERDVIGIARGQAEQKFLEIRGRFAQEQVLKSNERIIELESENSKLAQRI
QDARNVNVLRAELQELQSRPTKEEASRIRKEHHELKMESMRMHMASIRARYELEVELR
EQRKKVEDALEEKRETLARLEEQEEKRREEMTTRVRAIVPGLQ
>C_sp41_KNL1_paralog2

MPENLNHSILKSPSAPDSHDENRHVHFDKMKEVQVYDRNTGSWMGQTPIKEKAFDTLS
SGGILTPQLSDMDISQSPESSSVSSINKSSYNFETSVNDTLSIFEVTDEPAPEEFKFPEFDKN
VANPTFLCPKIAQEFEKLLVAELTETQSEIQRLFEKSDGEYRKKLEFLKNSPAGEWAKTEQ
DVIVISSQLAESLLLDLRLKFFAEKNEKCEKEIEVLKKENSKISGSIEQKKAKRKNEDFHTI
QDDYYKMKKLEMKCRLDSVMKSYEEQLECEKERITLEMEIEELKVRLEFVAEKEKEAK
ESLEKLITSIDY
>C_zanzibari_KNL1
MENNTNTRQKRNSILKVRQEVNLIDETTVVTAPNANRRVSFHQVKHVKEYDRDHGKIID
ATPVKEKAFDTMSSDGASTAHTTRVDMDVTGESQMASHVFESAPVTPGPLEMSLENQN
LDETARLFDVTREKTLVYTETTKTKTTVTRITTTTEAAADDTMALFNMTNRDGVDMSIA
NHDDTINVFNTTNTDRIDMDITQQKGPLDDTMALFRSPMPPPPPPPSSRILQNPLVENSESI
DMDITTPQMAVVKDSEDMDITEAQNPEVEEEEEEPGSEAMDITVASAAARGGSAASDDT
MSVFLATPKKIQMTSEIQKNHEDMEMTLKNETLDLFQSPAVAAVARGGKTMLQEQEVD
MDITQREGGEPSDNTISVFRSPEKIPTRPVFNDNDSMDISSTFRPESEVAVEENRCVSTVLQ
MSAMEMASPEVSIQNSLPTTTSSATSHSSSINNSSVTLNVSKNDGDAGASERSQHATMM
MDVTNPVLEEDSEDVTMTKEDSRRSLNVSSVSRRRRSLLMDSARESPRRLALENSMLS
MTAVVGGGAGGGGGSEALAEYRQNKTMLNDSVASGGNQTFGNVTATGRDIFKMNTSIR
SPAHHRPTTTIVTSPESIPPPKSPESFRLPQFDAAIVNVIYLTPEDVTQEPIPEAFEFQKLLAQ
HEADAHRDIEKSVAGDESLQKMLDELSEDVVKMRGGHLERDVIGIARGLAEQRFLEIRE
GFAEEQVAKSEAKIVELEAENARIAQRIQDARNVDLLRAEIHELQRQPTREEASRIRREHH
EVKMEALRMQVASIRSRYDAELKLREQQKRMQDAMEEQRERIARLEEQEAMRKEEMV
KRVRAIVPGVL
>C_sinica_KNL1
MDDKNKKRNSILKQRQEVNLIDETVVTSSNNGNRRVSFHQVKHVKQYDRDHGKIIDATP
VKEKAFDTLSSDGTSTAHTTRIDMDVTGESAPATPRFPSHHHNGTMDMSMDFSTIENQN
RDETARLFDVTREKTTVVYEETTVEKTTKMTKIVTTTNDTMALFNVTNRDDVDMSVAA
DDTMTVFNSTNAGAVDMDITMPQQVQNPKNPNLLDDTMAVFRSPAPPRIQKTSSGIQNP
QNQILENSESVDMDITRNQNPIDNTMSVFDTPKKDSKSKDSESKNSEDVKPSDDVTKDS
EDMDITEEIQKASEDVEEEAEPGSEAMDITMTPKKIQKASDDTMQVFETPKKIPGIQKTLT
SSSGVLDNLEPVDMEMTLKNETFDLLQSPARIQKTSGRILENLEIQKTSGGILEQRDDMD
MDITQREPMQPGDDTMAVFRSPEKIPTKNTTRQVFTDDDDDDNMEIETTLRVSEEPSEDV
KPSESSMMMSMETTIIEENRCAIVGSAGNLEISPFGAFGAQEEEPEDDMEIEKTSMMTMT
MTSEVQNLNPPTSSASESQLSPEKTPLKSEQNRSPIATMMLQKSPQDSEDAPEDVTMVKE
QKEESRMSQDSENVTKDSEDAPEDVTIPDDSRISLNVSSVSRRRRSLLLLDPSLCRESPRR
LAMENSLLSMTAIQGAGAASEALAEYRMNKTLQMHQASFNETSGGGGNQTTLNNVSAT
GRDIFMMNTSVRSPAHRPITITNTITSPESHIANVTAPKSPESIQILLPEFDAAIFNVIYLTPE
DLFEEPIPEAFEFQKVLAQHEALAHQEIQKSVDGDERLQKIFDELNSNSEDSGANLEHFEP
KIAKNFGLLERDVIGIAHGLAEQKFLEIREKFALEQVGKSEAAIAELEAENAKMAQRIQN
CKNADVIRAELQELRAQPTREECSRIRREYHEMRMEKMRIQQVQLRRRYELEVAYREQR
EKMECVAKEQRERIARIEDAEKERKAEMARRVRAMVQGV
>C_nigoni_KNL1_paralog1
MDNQRKKRNSILKVRQETNLMDVLEDTTVATSSSATNRRVSFHQLKQVKNYDRAGGQII
DATPIKEKAYDTMSSDGNSTSHTTRLDMDITGLGTPVTPKTPAPFNGSMDMSMENQDET
ARLFDITRDKTICVYEKTVETTTKIVERVVRVPEGSSGANDDTLALFNETNRAEVDMSID
GALKVDDTISVFNQTNVEPVDMDITVQKPLDDTMGVFRSPAIPASSRIQKTFSDSQDMSM
DMDITSNETMAAFKSPKIMSSGLVDAVKDADGMDLTGLVNTRAEDVADDTMAVFRTPT

RAQTTIQKTSGEISESVDMEMTGIENSDVPDDTMAVFKTPTRAQMTIQKTSGEIPESVDM
EMTLQNETLALVQPVSAVKDNSDDVAMDITQQTLVGASDDTMAVFKSPAAEKKTPAKPL
FDDSMEIESTIVRPDDVTSSEIAQPKNPENHSSMMMMSMASEVSEDVVVQKTSESLQQS
SMMMSTTITEDVTASKNPESSMISEVQKIPEVVQETSNAVEDIQKAPETPEDVSMEITSEV
VEGERGTMYQMSMMDMDSLQKTSLASPKIQMTSFTDTSEKMGGASETSMIQTMMIEAS
ESMECSEAPEDVTVSPEGLTMAPEDVTVASNVSRIVSVSSISRRRRSQLQESLHRESPRRM
ALEKNLSMMSQMGGASEALADFRQNKLNNQTTLLNDSVNTTIGANTSGSIGRDIFKMN
TSIRSPAHRSSTAPSSMVSKTLPESPKFHVTPFDAAIVNVIYLTPEDAETQEPIPEAFEFEKV
LSAEESNVHKEIDTANWSISGAIKSNLDAEVMNIARGQAEMKFLELRGEFAKERNVEIAQ
KNQELESQNLELAEKMRDSQNLPELQKQIEELQKPQFSLEEAERIENEYHETKVALLRAQ
AASIRRQHELQMAIREERRRLIQEIEEKDELLARLEEEDRKKREEMVERVRDVMRA
>C_nigoni_KNL1_paralog2
MSSEHNHSILKNPDESFPHDENRHVHFDKMKEVQIYDPTTGSMTGATPTKEKAFETVSS
EGTPSSYMDVVEDSDDSLNKTKNVSHESIQEETLDIFDISEEFKLPTFDPEVVKTIFSSEEE
FQNLLKSKMAESQKEMEKTYEESDGEYHDRLKFLKSNSISELTKIEQKVIEVAYKQAEIR
FLELRLKFAAEQRVKSEETIVEMQKVQKSRDSIPEMFDRIQDEYFEAKKEIMRIRLEKIQK
SYEEQLDSERKRFQLRLEIEERRKKLERLEMKENESKERMEQLISEMEV
>C_briggsae_KNL1_paralog1
MDNQRKKRNSILKVRQETNLMDVLEDTTVATSSGATNRRVSFHQLKQVKNYDRAGGQII
DATPIKEKAYDTMSSDGNSTSHTTRLDMDITGLNSTPVTPKTQTPFNGSMDMSVENYDE
TARLFDITRDKTICVYEKTVETTTTKVVERVVRVPEGSSGANNDTLALFNMTDRAEVDM
SVDGGSEGALKVDDTISVFNQTNVEPVDMDITVQKPLDDTMGVFRSPAIPTSSRIQKTSA
STMDSQNMSMSMDMDITSNEMMAAFKSPKIMSSVLVAAVKDSDDMDLTGLVNTAAED
VADDTMAVFRTPTRAQTTIQKTSGEIPESVDMEMTGIGNSDAPDDTMAVFRTPTRAQQT
VQKTSGEIPESVDMEMTLLALLQPVSDVKDNFDDVAMDITQQTLVGVSDDTMAVFKNP
AAEKKTPGKPLFDESMEIESTIVCPDNVTFSETAQPENPAYHSSMLMSMASEVSEDVVVQ
KTSESLQQSSMRMSTTITEDVTASKNPESSTISEVQKIPEVVQKTSNGVEDIQNASETPED
VSMEITSEVVEGERGTMYQMSTMDVDSLQKTSLASPKIQMTSFTDTSEKMGGVSETSLI
QTMMIEASESMECSEAPEDVTASPEGLTMAPEDVTVASNVSGIVSVSSISRRRRSQLQESL
HRESPRRMALEKNLSMMSQMGGASEALAEFRQNKLKNQTTLLNDSVNTTIGANTSESI
GRDIFKMNTSIRSPAHRSSTAPSPMVSKTLPESPKFHVTPFDAAIVNVIYLTPEDAETQEPIP
EAFEFEKVLSAEESNVHKEIDTANWSISGAIKSNLDAEVMNIARGQAEMKFLELRGKFA
KESNVEIAQKIQELESQNLELAGKIRDSQNLRVLQKQIEELQKPQFSLEEAERIENEYHET
KVELLRAQAASIRRQHELLMTIREERRRLIHEIEEKDELLARLEEEDRKKKEEMVERVRG
VMRA
>C_briggsae_KNL1_paralog2
MFSEHNHSILKNPDDSFPHDENRHVHFDKMKEVQIYDPTTGSMIAATPTKEKAFETVSSE
GTPSSYMDVVEDSDDSLNKTKNVSHESIQEETLDIFDISEEFKLPTFDPEVVKTIFFAEEDF
QNLLKSEMAASQKEMEKSYAESDGAYHDRLKFLKSNSISELTKIEQKVIQVAYKKAEIRF
LELRLKFAAEQRVKSEKTIFELQKVQKSRNSIPEMFDRVQDEYFEAKKEIMRIRLEKIQKS
YKEQLDSERKRFQLRLEIEERRKKLERLEMKENESKERMEQLISEMEKINIGHLINHRQN
QLSQQNMKASKVFITFDDISKVEHPFQVIVVVIRIGGCCHA
>C_remanei_KNL1
MEGNRRRSILRQPRAMSEEITDENVVEGAQTAIVTNRRVSFHATKQVKEYDREYGKIING
TPVREKAFDTMSSDGGAMTPRVASVDMDVSESSTSTTPFKIFDTNRQDTDGSLMDMSLD
RTLNNETARLFDITREKTLVYEKTTEVTTRTTERILTVPTSGADGGGGQDDTMALFNQTD

RQEVDMSVDQGGANNETLNIFSTTNRDVTDMDITSDHRVAPQFKTPRLPADAKKAPITD
NMEPIDMDMDTTLNAANDTMAVFKSPARIKRNDSEPIDMDVTRDQLNNETLALFQTPE
NHRKNPMSVLQKTPEIGSEAMDISMAAGMTPKPAPAPVDMELSSDDTMALFKITTPVTK
KNFGGEEDMDITQRPASVADDTMALFKSPARVEAAPTAVQQQQLFEESMEMEDQSALSE
KVAEPEDVATSETPEAPSNHSSMMMETETSIVEEDRAAVQMSMMDVTAPLDDVMESML
LNQTNSEKTLTTTSTTEESPRPKSSVRDVTSSVHVSSVTLNVSTNRKENDSMIRTMQYTE
VDTTNTLQNTSQRMEDSDEEEIKEESTAKETTLIREESEDMTIQNPQEHQQEMSISSYNLS
VFNQTTNSTGSVPISRRRRSLLREVCESQRRHAMEKSMNISTAGGETALEEYRKEKMNA
SGVMNQSLDQSAQGRDIFTMNTSIRSPNVRLAQGTPPPKSPRFEMPLYDPAVVNIVYLTPE
NVNNPTPLPEAVEFQKVLAEESQRVQKEIQRKQKESGVPPEKLDWIQKNEMSKLTRDER
EVVTIAREEAEIRFLRLRLKFAKEQREKGDEMIEKMTAENEKLALKMQGVKNIPALREEV
EHLRNQPTLAECHRIEAEHHEMKRLMMQLSLDWIRFMIQKVMEQKELKKKIQMDIELQ
QERIYQLEEEEKKAMDEIKRRVEEMEIQ
>C_latens_KNL1
MDGNRRRSILRQPRATSEEVTDENVLEGAQTAIVTNRRVSFHATKQVKEYDREYGKIING
TPVREKAFDTMSSDGGAMTPRVASVDMDISESSTSTTPFKVFDTNRHVDTDASSMDMSL
DRTLNNETARLFDITREKTLVYEKTTEVTTKTTERILTVPTADADGGGGSGQDDTMALFN
QTDRQEVDMSVDQGGANNETLNIFNTTNRDVTDMDITSAPQFKTPRLPVDAKKAPLTD
NVEPIDMDMDTTLNAANDTMAVFKSPARIKKNELEPVDMDVTRDQINNETLALFQTPEN
RKKPMSVLQKTPEIGSEAMDISMAATPKVMTPKRAPAPVDMDMSMSSDDTLAVFKSTTP
ARKNFGGEEEMDITQRPASVADDTMALFKSPARVETAPTVIQPQQQLFEESMEMEEEDQS
TVAPESEKPADSEGVVTSEAPEAPEAPTDIRSNHSSMMMETETSIVEGERAAVQMSMMD
VTAPLDDVMESVFLNQTDSEKTLTTTSTTEESPRPKSSARDVTSSVHVSSVTLNVSTNRK
ENDSMIRTMQYTEVDTTNTLQNSSQRMEDSDEEEIKEETTARETTLIREESEDVTVQNPQ
DQNQQENQQEMSISSYNLSVFNQTTNSTGSVPISRRRRSLLREVCESQRRHAMMENSMN
ISTAGGETALAEYRKEKMNASGVMNQSLNQSAQGRDIFMMNTSIRSPNVRLAQETPPPK
SPRFELPIYDPAVVNIVYLTPEDVSNPTPLPEAVEFQKVLAEESQRVQMEIQRKMKESGVP
PEKLEWIQKNEMSKLTRDEQEVVTIAREEAEIRFLRLRLKFAKEQREKGDEMIEKMTAEN
EKLALQIQGVKNIPALREEVEHLRNQPTLAECGRIEAEHHEMKRLMMQLRLDSIRFMIQT
VMEHQELKKKIQMDIELQQERIYQLEEEERKAMEEIVRRVKEMETQ
>C_sp51_KNL1
MENKKGRRNSILKTRTTTVDVLSETVDNGAGPSGPVTNRRVSFHNMKQIQEYDRHHGQ
MIEGTPIKEKITDTLGSDGILTPRGGNMENTNILEHDATFQVFNGAQKEHGAADMSLETTI
GGNENNETARLFNVTRDKTVYYEEVIETTKTTKTVVKMFGATAGADDTMALFNQTNSE
EVDMSMDQRQNDTMAMFNQTNPDGIDMDLDATRTFLVPSAPAPRAVGGGNDRTMNMD
VVEEESRRANDTMNLFNMTNADNTDMDISQYPAPTTGPNDTISMFKSPAPPAAKKKSVF
QHVDMDITGEEPVEPGSEAMDISQAPPPTPKVNDLRNVTLDMDITNEHHEEKTKETSRM
DISAMKEEEETFSETMALFGSPARGKMPAPTTTATPNRGSKSPNETLQMFQSPARNEAKV
PVSAQKTRIFDESMEIEESTLRPLDPAPEDAPEDAPDDVVTENSQNSESMVLEEKSVLEEE
RGATVVLQNSMIMDPREDMEVEMDQTIQKYLQMSMDISMDQSLQRTFQKTPKVHSDFQ
MSMLHRPKSVQDASTIHNESITLNVSQGAKTSESMMIKTMQISTDAADVTHTLQDTEQD
DRTERDTERDEVTGEMDMEEDMEETLEEKEVSLRKKNDTTTSDTQEAEGESTKLQETFT
TKFHVNQSIEDVTIQKNEEENHVDMSISMSSAILNQSCSMIRNSSRRRRSLLRDASIRESP
RRVALENSMHISIHGGENGRMTALEEYRKNQSLLNATEQMNESGMNVSATSSSSNGARD
IFVMNTSIRSPSAAKFAITNMNTPNAPSPISIPSAPSPGFQMPDYDPAVTNVVYLTSEDGEP
LSEATEFQRLIMEEKKKVQKTLEQDSEIIEQNTVSKQDRQEIMKLAREEAEIKFLELRLQF

AQEISRKQAAAIVSIESENSRIAESIKDAQSIEAEIATLRAYSPPVQYTRGRRENCANGGGD
EEGERNDFGGVEGCYGAAMIFHIIDTSPMPYRLHHIIFLKIYFIIYFFFFVKFILRLQFHVRN
HQKDLSDNPPHTYSLDLGQLESPVLPPLQLTTMPSLKNSNPTEVHISSGRDLVQRTLVFRN
LTGKDFLLKLIASNESITFPTNVFRFPPLSHRVIQFRVATSKISQWDKSKLSIKGYVMPVYA
KNLKQFINQKTTAGTACQEAFSLAIKFTDQFSAPQTVINLPGGATCIEATDHPVDVEELDT
MTALNIERDVATAAPIGSMLGFVKEYQRGQKAKGCWLSNYICGGTAEKAEKEKPSMRSR
RSSSASVSGKSQSSCRTQSRKKNKDVNVCLEATPCGGSTIQA
>C_sp44_KNL1
MENKKGRRNSILKTRTTVDVLCDTVENNAVRPGGNRRVSFQNVKHVQEYDRHHGQMIE
GTPIKERITDTLDSDGILTPRGGNMDNTKVLEADRTFQLFNQREEHGAADMSIDTTLGGE
ADQTARLFGITRDKTLVYEEVVEETTKTTKRVTTKIFEPAGGEVAAGGNDTMALFNLTNT
EEVDMSMDPKHDDTMAIFNQTNTDEVDMELEGTFVVPRLPSRVKGVNKTINQTVAMDL
EDDQDQNMIRANETVNLFNITNTSCVDMDISRQPTSVCPNDTLAMFQSPNPKKKILTVVE
EHVEEPVEMGSEAMDISVAPPTPKRENQKNLTMALFDVTRDEDQDMDMDITKDATYLK
KDENHLDMDISVTVAAPTPISPPEDTHSETLKLFQSPARGGKSGNPQISKLFNESMEIENT
NIHPDPAPEDPKSSESLADPDTSVVEEKRAALQNSCMDVTMPREDEQEEVAPEETTIQKSI
HHTSMDMSMETTLRQNLFQNPEKNQNMGISTGVSQSVAQDASSIHNESVTLNVSSESKN
KEPTKAANDSMVKTMQISTDTADMTRTLQETEIEEIQEKTDDEMDIEKEEETLKDSNDSS
KIVEKTFNLSKVHIESTTSQEDVTFPNGQQVESTIHVETSHMDFFTSSHVNNKSGTPSTSIL
NQSCSMIRSSSRRRRSLLRDPSMRESPRRMALENSMHISMQMPNVKAGGSDGMMMTAL
EEYRQNQSLLNENSGMNESGLSVNNGTRDIFAMNTSLRSPRAAMANNRSFNLNSPRVPS
PISINYTAPPAASPGFSMPTYDPVVVNVVYLTSTDGSTLPEADGFQKMILQEKKVLEKDL
HDVDVPQANLSKRDEQEVMYIAREEAEIRFLELRLKFAENTNSSQIAEIEDLESENAKMA
EELRDAQNLPAIRKELEAMKAQYTKAEIREIRQEYKRCKQMQFDAQLAALKHNYNVGL
EIANKRSEIRMEMEEIDEIMARQDEKARKEREVLMKELKNIMEQ
>C_sp48_KNL1
MENKKGRRNSILKTRVTVDVLSETVDNNVPAGPAQNRRVSFHNMKQIQEYDRHHGQIID
GTPIKEKITDTLGSDGILTPRGGNLDATNILDPNSTFQVFNNGGGAADMSLETTENHDETA
RLFNVTRDKTIYYEEVVETTKTTKTKITKIAGSSGAAAAPGSSDDTMSLFNQSNTEEVDM
SVDRPNDTMALFNQTNPDEIDMDISGATFAVPKMPRHQNRTMNRTLPMDLEEEEDQTKK
ANETMNLFNVSNLENVSMDISQHPAQQSILHAPNDDTLSLFRSPAPKKQVFQEEMDISQV
QEPEEPGSEAMDISQAPPTPIHAKKNATMIALFEENDDMDITQNTTIVKETVEEKSCHMDI
STSVRPDVQSETMALFESGGVSPAREKMTVPTANTSPQDQEKTADATLKMFQSPAGKQA
ENEMKMPPSAQKTMVFEESMDIEESTIHVESATSPKAPGSSTLQDSESTLHKSLQESSLM
LEQSVVEEERAAVLQNSMMEVLPTDDQVEEEELGEIQKTSSFVRREVEEESMDVSMDIS
HQKTPQRSMHRPKSAQDASTIQNEASANLVDEEEPKNESTVKTMQISTDLDVTRTIQEEE
SEEDKTSGHPDGQMEELQETVQLFTSKLQVDSTELGEDVTIQMAMEESSIQVAQNNQSEI
SQFHSSLNESYSIRNTSRRRRSLLRDTMVRESPRRVALENSMHISMNMPNHYGKEDGRLT
ALAEYRQNKSLLNASQMNESGGSNVTMTASPGGNATRDLFAMNASIRSPQPTTINKSMN
LNTTPRVPSPISVDQSVVIPPPPESPFQPMPIYDPAVVNVLYLTADDGTAHPEAADFQKVL
MDEKQGLEEKDSEEAIQNSNLSKRDLRDMTLLAREEAEIRFLELRLKFATEHNEKQTRSI
EQLTSENAEMAEQIRDAQNLPELQKEIEKLKIEKTQYTKAEVRQIRQDYMYWRKKLFDA
QMTALKHNYEVAVAIADQRAELRMEVEELEERAGRSDSQDYGGMIFDTAIHHISLSPRH
HPHINLFFCKINPNYPDFLMSKTY
>C_brenneri_KNL1_paralog1

MENRNGRRNSILKTRVTVDVLSETVDNNVPAGPAQNRRVSFHNVKQIQEYDRHHGQMI
EGTPIKEKITDTLGSDGILTPCGGHMDVTNMLEHNSTFQVFNNGGADMSLETTVAGGEE
ENHHDQTAHLLNVTRDKTVYYEEVVETTKTTKSKITKICGSSAGGASDDTMTLFNQTNL
EEVDMSMDRPANDTMALFNQTNSDEVDMDMSQSATFAVPRMPRHQNKTLNRTLPMDL
EEEEEESQMKKANETMNLFNVSNLENTDMDISNHPAQNSILQAHNDTISMFRSPAAKKQ
VFQKQVEMDIYQAQEPEEIGSEAMDISQAPPTPSHVTKNATLAMFEDDDDNMDITQNTTI
IKETTVEKINENCHMDISSVTVTNPDVQSQTMALFESGASPAREKMAVPTAISSPKDEDQS
ADATLKMFQSPARQAGVEMKMPTSTQKTMVFEESMDVEESTIHVESPKSPASEEQLATE
TTESSESPGSPESSQIQDSESTFHKSLNENSIMLEQSLVEEERAAVLQNSMMEVLPRDEQE
EELVDVRKEEAMEQDSMDVSMDISHQKTPQRSMHRPKSVQDASTIQNETSANLVEEEER
KNESMVKTMQILTDTADVTHTIQEEETLKNRTSGHSDGQTEELQVKSTELGEEVTIQMA
MEESVAQNQSQMSFNESYSIRNTSRRRRSLLRDTLVRESPRRLALENSMHISIHMPNHYE
KEDGRLTALAEYRQNKSLLNASQEQMMNESGSNLSMTTGAAASPGNGTRDIFAMNASI
RSPRNTTINKSIDLNTPRVPSPISKDQSMVIPPESPFQPMPIYDPAVVNVLYLTADDGTAHPE
ATDFQKALMVEKLGLEKIQEVTIQNSNLTKRDVRDLMLLAREEAEIQFLEFRLKFAMEQS
QKQTRCIEELTSENSKMAEQIRDAQNLPEIKKEIEKLRIETTRYTKTEVHQIRQEYMYWR
KKMFDAQMAALKHNYEVGRAIADQRAELRMEVEQLEERAAQIKEKNRQERRELVDQI
LKIMEEW
>C_brenneri_KNL1_paralog2
MENKKGRRNSILKTRVTVDVLSETVDNNVPAGPAQNRRVSFHNMKQIQEYDRHHGQMI
EGTPIKEKITDTLGSDGILTPRGGNMDITNVLEHNSTFQVFNNGAADMSLETTIAGGEEE
NRHDQTARPFNVTRDKTVYYEEVVETTKTTKTKITKICGSLGAAPGSSDDTMALFNQSN
MEDVDMSMDRPANDTMALFNQTNSDEIDMDMSQSATFAVPRMPRQQNKTLNQTMLM
DLEEEEGQTKKANETMNLFNVSNLDNTDMDISNPAQNSILQAPNDTISMFRSPARKKQV
FQDQEPEEIGSEAMDISQAPATPSHVTKNATLAMFEDDDNMDITQNTTIIKETTVEKTNES
CHMDISSVTVTLANPDVQSKTMALFESGASPVSEKMAVPTAISLPKDQEQTADATLKMF
QSPARKTGEEVKLPMSTQKTMVFDESMDVEQSTIHVESPKYPAPEEPESSESAESPESSQI
QDSESTFQQSLHEHSVMLEQSVLEEERAAVLQNSMMEVLPSDEQEEEQLVDVGEIQKTS
SFVRREEAMEQDSMDVSMDISHQKTPQRSVHRPKSVQDAATIQNMTSANPVEEEERKD
ESMVKTMQISTDTTDVTHTIQEEEEETLKNKTSAHPDERTEELQQTVQMFTSTLQVDSTE
LGEDVTIQMAIEESTIHVAPQNQSQISQSSFSESYSIRNTSRRRRSLLRDTLVRESPRRVALE
NSMHISMHMPNHHGKVDGRLTALAEYRQNKSLLNASQDQMMNESGSNLTMATGAAAS
PGNGTRDIFAMNTSIRSPRNVTINKSINLNTPRVPSPISMDKSVVVPPESPFQTMPIYDPAV
VNVLYLTADDGTAHPEAADFQKALMVEKQGLEKIQESDEEGVIQNSNLPKQDVRDVML
LAREEAEIQFLELRLKFAMEHSQKQTRCIEQLTSENSKMAEQVRDAQNLPEIKKEIEKLKI
EKTRYTKTEVHQIRRDYMYWRKKMFDAQMAALEHNYEVALAIADQRAELRMEVEELE
ERATQIEEKNRQERRELVDQILKIMEEW
>C_wallacei_KNL1
MDRKKRNSILKIRQTVDVLSETVETGTTTNRRVSFHNVKHVKEYDRDHGKMVDATPVK
EKISETVGSDGILTPRGGNMDVTNVQLEDGTFRVFNGAGAADMSLESTIGGAGENKNLT
AGLFGITREETVMIEERINTNDTINIFNQTNTEAVDITVDGTVINDTMAIFNQTNSSEVDM
SMDQSVFAVPKLPSKNKMKAELENVEEEKEMDISKGNETMKLFNITNPDVIDMDMSQM
PTEARDETIAMFRSPAPPKKKEMDDVEMDITQTEQEEMGSEAMDISAAPVQQNSTLAMF
VENEIKEKDMSVESTMDITASKEQEMDISVASPPSPEASETLALFNSPARSKKDQTLKMF
QSPAQKTQIFDESMEMEEEEEDFVPQDPEVTPEVIPEVTLEKSTSQIESLKETLVVEEERAV
LQNSMMEVEAQGEEEALLKGEEDIQKTTPSFMDFTHNMTVETSEKTNNESIVKTMQLN

ETGTIREEEEDVEKTITKNESLSMEDTSRKMVESSDDVTIRMHTKIVEEERYTQMENTTE
SRMSISTFSQNMSISTSTTLNLPPISSVSKKRRSLIRDQSARESPRRMAIENSMLSMHNGG
GVALAEYRQNKSMMNTSVGQMNESGNMSIGSNHGRDIFIMNTSVRSPRHQTNSSINLNI
PRQPSPIAQKTHRFEMPDFDPAVTNVVYLTSDEQTIPEAVEFHKMISEEKQRIQEEIQNSDE
LIKEKIEKTDAKLTNDEQEAVMVARKEAELVFLQRRLQFATNQNQIQEKQIQNLESENRE
TAERIRDAQNLPIIKKELKEMKNHPTKSECRRIRQEYMKMKKEWFEIQMRSLKNTYEAA
VELADLRAEARREFDEREYEIKQLEEENKKKMSVFHGKVNEMIMQIEN
>C_tropicalis_KNL1
MERPKKRNSILKVRQTVDVLSETVETGTTTTRKGRVSFHNTKHVKEYDRDHGKMTDAT
PVKEKVSVSVGSDQALTPRGGNMDITNVQLEEGTFRVFGEADFGGAADMSLESTVIGEN
RNETANMFGITREEPMTFEERTTSGSRTNETLNIFNQTNTEVVDMTVDGEGIVNDTMALF
NQTNSSEMDMSIDQTRFVVPKLPSKSKVNKNMTVNKTIPMELEDDEEAPVKEKVSLSV
GSDQALTTSGSNMDITNAQLEDGTFRVFGEAVSGGAADMSLESTVIGENRNETANMFGI
TREEPMTFEEKTTVGSRTNETLNIFNQTNTEVVDMTVDGEGIVNDTMALFNQTNSSDVD
MSIDQTRFVGPKSPSKSKMDKTVNKTIPMDVEEDEEVNVHIGNDTMKMDVNEDTEKTL
TRNDSQLMEDTSRKMVNSSDDVSVRMHTKTVEEELYTQVENTITNQSRMSISTYSHNM
SITTSSTLNLPPISSVSKRRRSLLRDQSIRESPRRMALENSILSMHFPNGGEEALTEYSQNKI
NTSTGKLNESSNVSCGSNGGRDIFLMNTSIRSPLHQTNVSINLNTLPQQSPIAQESARSSF
HMPDFDRAVTEIVYLTSEDAEISLPEAVEFHKMISNEKSRIENEIQTKIQVSDVVLVEKME
NLKKDQVSRMTNDEQKAVMMARQEAELQLLKLRLQFATSQRQNQEQIIRKLESENMET
AEKIRDAQNLPSIQKELKEMKSHPTRTECRRIRQEYMETKKEWFNIQMKALKNTYDTAI
ELADFRAEARREFEERDYEMKRLEEENRKRMNIFRGKVNEIIQKVGE
>C_doughertyi_KNL1
MENRKKRNSILKVRQTVDVLAEIEDNSGPSGTTTTTNNRRVSFQNVKHVKQYDRDLGK
MCDATPIKEKITDTLGSDGILTPRGGNMDNTNVQLDEHGSFQVFGAGGGAHNTHGEAD
MSLESTQGGGAGGGGAPNDTARLFNVTREKTIVYERITTERIVVVPQDSGAPNVASGAS
GAPSGAPSGALNDTLSLFNQTNHEEVDMSMDHHQQRKDDTMAIFNQTNTEEIDMSMDE
THGFAVPKLMPRHHLTKNQMNKTMNETMPMDVVEEEEGLEIGKKGNETMSLFNVTNPE
AMDMDITQTAALAPTSANDTLALFRSPAPKKKTMTSSVAEDVEPIDMDITTPQRPPSELG
SQVMDISVAPTPRGNNSTLKLFESHLEPEKDTKEQDMSLEKTIQAPPPQSTEAIDMDISQL
PKSPEEHPQNQSIDMDITVEKAAASETMALFQQSPARSKTPSANETLKMFQSPARGNEK
MRMPGTIEKTKMFDDDSMDISIAPPTPKAASEDVSKSKTPKPAEKTKVFEDDSMDISVAP
PTPKAVSEPKTPGTTEKTKVFEDDGMDISVAPPTPKAVSKDVSEPKTPGTTEKTKMFEDD
GMNISVAPPTPKAVPEDVHEPKTPKSAEKTKVLEDDGMDISVQPPTPKSTPEDSEDVEDS
DDVEELQKTLHDNSSMMISTETSVLEAGRAVHQTSMMEVYQEDVVETEGLLKHQTSFV
VQNTSICRPSSGARDVSSLNNDSITLDVSAKNDSMLKTMQLTDDVDVTRTLNEDSDDVE
KMEEEEEEKTLTEKSEPSEVPEDVTQKIQEETVVEERTIGFDNMSSTRFSYNMSMSTTTQ
VLNQSASASRRRRSLLREPSLRESPRRLALENSMMSMLQQPSYGGDALSEYRQNKMNT
TLEHQQQMMLNESVQNTSTGSIGGGGRDIFAMNTSIRSPHHITQKNASFNMNTPRVPSPL
AFEDSPAPPHSTVFQMPDFDPAVVNVIYLTPEDPDAAPFPESLEFQKLVDVEASRMQKEIQ
TILENSEDVGGLLEKKLNALRAKDMSKLTSDERDVVMEAREEAELYFLELRLKFAEKTL
QNQRLQIEKREAENAEKAEQIRDAKNLPAIKKKLDEMKACPTRAERRQIRKEYMEVKK
AVFDLQMERLKKVQNTIFEIGNQREELQDELYGLGEQLAQLEEQKRREKAAIMELVMEI
YQQW
>C_sp54_KNL1_paralog1

MENRKKRNSILKVRQAVDVCDETVQTGGPSATTTITNRRVSFHNVKHVKHYDRDHGKL
MDATPAKEKISDTLDSDGILTPRGGNNMSINDAASFQLFDGAGSQDKSMNMSLETTTNE
NNETARLFDITREKTVVYEKTVETTTKITERIVSVPPPPPSVTAADDTLAMFNTTNTEEMD
MSINQSAVFAAPKLPRNNRTLNKTMPMDLDETTTANANETVNIFNMTNMDATEMDITLA
QNTSQFKHPAAINDTLAMFQSPAPKLQQQRQQKDDVEHVDMEITQPIFEEPASEAMDISA
VQIPMNASMNMSKVAIGNETMAMFQSPKKPEESFMDLEKTMENHLNSNSSSMDITQTPP
EVMEKVEKEEPEVKGNDSMDISIVPPMTGANDTMAMFQSPARNNAQNMTAAPPASLQK
KQQLFEESMDIEDEATLVAPEAPEAPEAPEDVPTVPEDVTPTPEEAIEDVTQESMMITETS
VLEEERAAVLQTPVMELTPSEEDPGTLLLSAYHQIQKASSSSLMQEQERPQSSTSRHSTIF
NDSSVGLINTERKLNDEDTMVKTMQFTETDATNTLQDTESISPMGKATEDVDVQMDME
EEEEETLQRNDTQQVEDVKYEEKLHETLQNSVNIQENINESIDTEVTIQKEMSIHQYTFN
QSINQSITSCSSKYCSPAISSKRRRSLLNASNHRESPRRVALENSMLSLHLTTEALSGYRQN
KMNLSSDSVANTSGGRNIFAMNTSVRSPHHLNTSMSSSTTPSPHHVTPIPTEVPPSPGFDM
PDFDPAVYNVIYLTPEDVTKEPLPEAVEFSSLVAEEKAKMQKEIQESIETSGVSMEKIDYLR
NNQISKLNSVERDAVIMAREQAEFRFLELRLKFATEQRKKNEQQIMEIQSENEKIAEQLR
DAENLPALKKELETLKSQPTQEECRVIRKEWRKMKQWAFDMQMKALSDTHETIMEIRE
QKTDLLMEIEEREEKLAQLEDLEKKKKATIVEKVMEVMQ
>C_sp54_KNL1_paralog2
MENRKRRTSILKKRQIVETGESSGRRVSFHNVKHVKQYDRDTSEAMDISADQIPMNSSM
NMSKVAIGNETMTIFQSPKKPEEKQEESFMDITQTPETLENMEKESEVKSNDSMDISIVPP
VTGANDTMAIFQSPARKNLQKMTAAPSASLQKKQQLLEESMDIEDEETLLAPAASEAPE
DVTQESMMITETSVFEEEAVLQTSMMELTPSEEDPGALLTSACRQIQKASSSSLMEALTSS
SPGFDMPNVIYLTPEDVTKEPLPEAVEFSSLVAEEIQKSIETSKNSEVSMEKIDEAEFRFLEL
RLKFATEQRLKNEQQIKEIQSENAKIAEQLALKKELEMLKSRPPPTREECRVILKEWREM
KQWAFDIQMKALSDIHETIMKIREQKTNLLMEIEEREEKLAQLEDWEKKKKATILEALED
FRQNVNFSDKLVFFGCGYAGRRPLFLTLRKTIFGSKPGLGRMTENRQAVKVRNKYLKRF
KEVKKWELDSKKSIIDSVNLPWLMIFYTLETLIDKLTIEGSCERVNTKMYVFSGFYLFMIL
WFIFAYRWLKDVNEEREQMDSGEHQMMNCDEDKFDENIQSVRESINSLNKKSIAVCILY
GTIVVLLLCVSSLS
>C_inopinata_KNL1_paralog1
MDNGRPGLPTTNRRVSFQNLKQVKQFDREHGKIIDGTPMKEKITDTLGSDGILTPRGGN
MDISDSQSACVTLQLFDGINPTKETVNMSLETITNENIVNGRMLDISKGKTLVYENTVET
TTKMTERIFMVPQSDETLKMFNSTICQHKDMDISLDGSAVSAAPKIIKHNVTLNETVPMD
LDHSKALDGSYNETLNMFNVSNLGNCAMDITLVQNASFASKPASFSHRRPNETLEMFGS
DPPSTLSQLNAIQLTDMKPGVPENPIENDYLMDISVQQGSPVAVNDTMAMFKSPSRKDM
SISFDRSAVSAAPEMLKHNVTLNETVPMDLDCSRAAAGSYHETVNMFNVSNLENSAMDI
TLVQNADFASKPAPFSHRRPNETLEMFESNPPSTLPQLNTIQLADINPGLPENPIENEGLM
DISVQQGSPVAVNDTMAMFKSPSKLNMKDKVKILLNPEIISKNETNRTVADDSMEIEETF
NLEDRKIVVKQLTLETTSVTEKFPNLQAPPNENSDTVRISRAYQTEEDLENKRAHLNSSA
MDVTQFGEDIDILNNLQELQLRPQFAASTEQSLAFHNSRVTVDVSGKLTENEETPKVMM
METMQISTTDPTNTLMEINADDNRTKMEGDNRKVIQVENSERTILEQRMNQTVTEIEKNI
QNVQNASVNLSRTAIERQEDIDVENEHSRMLTTSFNLSIHQGISKRRRSLLRESVISRESPR
RCIAEKSLLSMNLGTCTTPLNMDELSEQRQNAIKTSEDDDVSLNSSLNKSARDIFTLNIM
VRSPRNTSINGKSPATPYSLSHKSPLDSLRSQDSKSPEFELPDFDPSVFNVIYLTPNNGKIIP
EAAEFQRLVEDERKKVQLEIEDAVKSGEFFAEHWEILKNKQFSKLTYDEREVVRIAREMA
EFRFLELRLKFAKEQRVKSEQKILELEEEIAKMAELKRELENLPAIRAELDELKRRPTFAE

CQRIRQEYSEMTKLVFEQQMELMRRNYEIALQIREHRATLLRELDEREEILNQLEDKERQ
NKAEIEKVVKNVMAHMN
>C_inopinata_KNL1_paralog2
MDNGRPELSTTNRRVSFQNLKQVKEFDRDHGKIINGTPMKEKITDTLDTDGILTPRGGN
MDISDSQSACVTLQLFEGIDPAKETMDMSLETITNENNVTARMLDISAEKTLVNENTVEQ
TTKMTERIFSVPQSDETLRMFNSTICQHQDMDISVDGSTVSAAPKILKRNVTLNETVPMD
LDYSRAVAGRYNETVNMFNVSNLENPAMDITLVENANLASKPASDIKPGLPENPIENEDL
MDISVQQGSPVAVNDTMAMFKSPSRLNMKILPDPGIISKNETSGTIADESMEIEETLNLED
RKIVVKQLTLETTSVTERFQNLQDECKTLDENSDTVRISRTYQTEEDDVENERAHLNSSA
MDVTQLGEDIDILNNREERQLRPQFSVSTEQSLAFHNSSVTVDVSGKLTEHEKTPKVMR
MEAMQISAVDPTNTLMEINVDDNSTKMEGDNRKTIQVENSERTILENSMNQTATDIEKNI
KNVENASVNLSETAIERQESMDVENEHSRMLTTSFNLSIHQRISKRRRSLLRESAVSRESP
RRCIAEKSLLSMNIGTCTTPWNIDELSEQRQNAMKTSEDDVTLNSSLNKSARDIFTLNIAV
RSPRNTSINGKLPATPYSLSHKSPLDSLGSPASKSPEFELPDFDPSVFNVIYLTPNNGEIFPE
AAEFQKLVENEKKKMRLEIENGVKSGEFSAEQWEILKNNQLSKLTRDEREVVRIAREKA
ELRFLELRLKFAIEQRVKSEQQILELEEENAKMAELKRELENLPAIRAELDELKRRPTLAE
CKKIRQEYMEMTTLVWEQQMEAMRNTYKLALQIRDQRAALLREIEEKEEILNQLEEKE
QQNKAEIENVVKNVMAHMI
>C_elegans_KNL1
MSMEPRKKRNSILKVRQAVETIEETVMNSGPSSTTTNRRVSFHNVKHVKQYDRDHGKIL
DATPVKEKITDTIGSDGILTPRGGNMDISESPACTSSFQVFGGGNLDKTMDMSLETTINEN
NETARLFETTRDPTLLYEKIVETTTKVTERIVSMPLDDTLAMFNTTNQEDKDMSVDRSVL
FTIPKVPKHNATMNRTIPMDLDESKAAGGQCDETMNVFNFTNLEAAEMDTSKLDENNT
MNAIRIPINSNVMPVDMDITEHHTLIEEKKNDTFGPSQLMDISAPQVQVNDTLAIFNSPRD
ICNKGLGVPQNLINIASNVVPVDMDITDQAVLNAEKKNDQFETSQLMDISIPKVLVNDTM
AMFNSPKHVSKSSMDLEKTIEAADKSTKYPSIADEVEDLDMDMDITEQQPCEAGNQQN
DGLQLQKEDLMDISVIRDSPAVNDTMAVFQSPARVKIGANNSIIDSQKSIVFGDEMSIDET
QNDGTLTLPKSNVEVTTTNDVYTSLERQEENASENVSMINESSVHSEIDKKSFMLIEEER
AFMHSSMIDVAQKLEDDGSSKTPVILASQSASLATKEPSALHNSSATLNNSMELDNNTLL
KTMQITTCEDISMVHESIAVELNSNKEQEQFGDETLQKNDTSNTGANFTFQGHNETSQIM
NNVDSEAVNTSKISTYSAFNLSINQSISKRRRSLLNSARESPRRVALENSIMSMNGQTMEA
LTEYRQNKTMQTSQDSMPSMSLNDSGRDILAMNTSVRSPHLNSSKTAAPGTPSLMSQNV
QLPPPSPQFEMPDFDPAVVNVVYLTSEDPSTEQHPEALKFQRIVENEKMKVQHEIDSLNS
TNQLSAEKIDMLKTKELLKFSHDEREAIMIARKDAEIKFLELRLKFALEKKIESDQEIAEL
EQGNSKMAEQLRGLDKMAVVQKELEKLRSLPPSREESGKIRKEWMEMKQWEFDQKMK
ALRNVRSNMIALRSEKNALEMKVAEEHEKFAQRNDLKKSRMLVFSKAVKKIVNF
>C.oiwi_KNL1_paralog1
MENRNNRRNSILKVRQHVELLDETSADGPQASSTTTNINRRVSFHNVKHVKEYDREHG
KIIDATPQREKINETVASDGILTPRGTNMMIKEESTIPENVNGTVDMSVENNETLKMFETI
VYEKTVETTTRSTVNAPAPADDTLAMFNVTTSEEQDMSIDQSRVFAVPELPGARAVSVPA
AGEQDMSIDQSALLEAQKTMNQTVPMDLSGSKAANDTMSMFNATSNDTFDMDITAPNP
PPTPTMTRRLPKNVFSGEEDMEIEETVVNISAPQNDTMSIFQSQSPARNKMPVVPEDPRE
EEEEVREPMDISMVPGNETIGIFQSPARQKIGNVSEMIPMAQVSSNDMDITNVTVSEEKN
SEEVDMEIEDPTVQKIQNDTLSMFQSQSPARNRIVPVVPELQDESMDISAMPTIEPSDHQI
PEDSTLILDKKDTSQDLLESPARQNMSEELADRSEKDLEESKNTSAIEEDRAVLLHSMME
VTAVQEEPDLDQKMPERRPQSAASTEQDSTINNASITLATSASAKSMMTSPQKTMQIEED

ATHTLQETSLVQDQKDQEDQEEMILEKNESRSLIPEVQDPSDQEEMIVEKNESRSLIPEVP
EARDPQEMIVVESRMSISYHEHLRIRTPSSSATRRRRSLLHSTDSPRRLALENAMAEKTAL
EEYRHQKSLHTSHNLSNLNVTAAANNMSISMSMSAAGRDIFDMNTSVRSPKCMVNSPSP
ANVFPSEAPPSPVFQMPDFDPAVTNVVYLTSESGTTHPEASEFRQMIISEAARLEREIQNT
VGSGGVDSGKLEALRSGHFSRLSHDERDVVMKARDEAEIRFLKLRLKFAQEQRAKQDR
EILEVERKNEQLAEQLKNAENVEVLRAEIEELKRIPTRAEYEQIMKEWREMKKLEIDVQI
QTIREAHHKLLQIQYETAQRTMELEGADESRRDERNCLIIPISHLQPK
>C_oiwi_KNL1_paralog2
MFLNQQEAPIVNEHCEYVKKILNHDVVLKARGEAEIRFLELRLKFAQQQRAKQDREISEI
ERKNEQLAEQLKNAENVETLRAEIEELKRIPTRAEYEQIMKEWREMKKLEIDVQIQTIRD
AHHKLLQIQYETAQRTMELEGADESRRDERNCLIIPISHLQPK
>C_kamaaina_KNL1
MDNRNNRRNSILKVRQHLEVLDETSADGPQATTTNVNRRVSFHNVKHVKQYDREHGKI
IDATPQREKINETVASDGILTPRGTNMMIKEESTIPHNGTMDMSMESAAANENNETMKM
FDTTRETVVYEKTVETTTRTTEKIVIAPADDTLAMFNMTNSEEQDMSIDQPRVFAVPAVPS
RAVPTTGGVQDMSIDQSAILEAPKNDKTVNQTVPMDLSSSKAANDT
>C_waitukubuli_KNL1_paralog1
RRSSILKVRQHVDIDIDETTVAGSGAPGASTTTTNRRVSFHNVKHVKQYDRDHGKLIDAS
PLREKITDTMDSDGILTPRGRNLDASEVAGNQTIHVFGNANNANQTADMSLETVNENVG
SETMQLFNFTREQTMVFEKTTTTTTVTERRVEGPVMDDTMALFNMTNTEEKDMTVDQT
AFAVPVSRAARPNATVNQSVLMDFDESTIVPQPNNDTMSLFNMSSVSVKDMDISNATVN
ETFKVPKSPSRSVMPRVFEENEDEVDMELTNVEGEKSIVANESMDISVMPAATPKKVNA
DQSDMKSIVANESMDISVMPAATPKKVNADQSDMVLEKTNLIQKTRPVLNESMDISVLQ
TESPKNFVGPNDTLALFQSPARQANQSYQHQEQPKNIVFRDQSVDMDITQTKPPAAAIKL
NDSMDISEMPSEAHQSPARVQQSFVAPVSAPEPFDDNSMEVEEDVVREGTAQVIPTAEQT
EMMVEEPSIVEQHEEERAMAAVSMMDVSGIQESLDYQQKLSTSVHASPALGSSPKESEV
HIMDDESSVSILKDTTIASNAESTISPQKTMQIETDMTATPREDESMQMEDVENQEKTFN
EFVEQETMTMTMSFSSAKKSTSHLSRRRSLLRESVLESPRRVAIENSMITMNMAEMTALS
EHRRRKSIQLNQQNLTNTSALNDSANMSKGRDIFALNTSVRSPHCSSTPRSGNVSAAAEI
QIPAPESFQMPDFDPAVTNIVWLTSDNGDPIPEAEEFAKLVNAEKIKAEQENRMRLLDQNI
DKEKAEAVRNGQWSKLSHDEKEVVLIARQDAEIRFLELRLAFAEKTREKYEKEEKELTE
ANARFAKQLEDVESMNALRQEIEELKSQPTEEDFERIEKEWQEAKNEIFKWKRAALQHN
LDILLQMSEAKRQLEMEIEVREEKLAQLEAMETAEKEELMKIVATL
>C_waitukubuli_KNL1_paralog2
dettvagsgapgastttMNRRVSFHNVKHVKQYDRDHGKLIDASPLREKITDTMDSDGILTPRGR
NLDASEVAGNQTIHVFGNVNNANQTADMSLETVNENIGSETMQLFNFTREQTMVFEKTI
KTTTVTERRVEGTVMDDTMALFNMTNTEEKDMTVDQTAFAVPVSRAARPNATVNQSVI
MDLDESTIVPQANNDTMSLFNMSSVSVKDMDISNATVNETFKVPKSPSRSVMPRVFEEN
EDEVDMELTNVEGEKSIVANESMDISVMPAATPKKVNADQSDMVL
>C_panamensis_KNL1
MENRNRRSSILKVRQHVDIDIDETTVAGNGLPGASTTTMNRRVSFHNVKHVKQYDRDH
GKLIDASPLREKITDTMDSDGILTPRGRNLDASETANNQTIRVFGNVDNGNQSVDMSLET
VNENVGGETMQLFNITREPTMVFEKIVTTTTTVTERRMGPQGVLQSQAAPLDDTMALF
NVTNAEEKDMTVDQTRMFAAPEKRLVHSQTTMNDSMDISVAPSDTPIRVNDTMALFRSP
ATRANESDMVMEKTNVFEQEEGKVLHKQTMPPVMNDSMDISVMPDDSPKNLRALLPN
ATLNQSVLMDLDESNAAPANNDTMALFNISTASVRDMDISNATIHETFKVPKSPARSVLP

LVREENDDDVDMELTSMEGEKSIVQNQKTVALNDSMDISVAPVNQFLPVFPVVEQEEAN
QQNTLAPIVTDDAMDISVMPSDLPKNIDSNDTLALFQSPARQVNQSSQQPVHNVFNEPK
DESVDMDITQTKEVATPQRMNVANKSVVVAPTVLKGPKDDSMEIEEAGAGEGTFVVPTT
ESPETSPEDILIPNIEQANMGGDLCDKIIVEEPSTVEKHEENRAVAEVSMMDVSGIEESPSH
QKFPNLQASPTEESKSNSTLHESSVSHNDSSPKDTSTRGNESTLSPQKTMLVETDMITTLH
DMSLVQKENDSEEMEEQAEVEEPEKTVTEIVEQNTTTTAQMIMSSLNTSSNRKPTSHASR
RRSLLRESVLESPRRIALENSMITMNVAEMTALSEHRRRKSLQLNQQSVLNASNISMLNE
SLNMSKGRDIFALNTSVRSPHCHETTQSGNMSIVESGAAPETTSFSMPNFDPAVTKIVWLT
PEVGEDSIPEAEDFARLVAAEKIKAKQHAQKTSENLDKEKAEAVRTGQWSKLSHDEKDA
VLIARQDAEIRFLELRLTFAEKNRLKYEHEAKELLESNARIAEQLEDIENMESLQKEIEEL
KNRPTEADYERIGEEWRVAKEELFQLKHALLQQTLHIIMQMGAQKQQLKSEIEDRQKNL
ASLEAKETSEMQEMLKMMETL
>C_nouraguensis_KNL1
MTDETTVADNGLPGASTTTNRRVSFHNVKHVKQYDRDHGKLIDASPLREKITDTMDSD
GILTPRGRTLDASAAAANQSIHVFGNVEHGAADMSLETINENAGSETMQLFNITREQTVV
VEKSITRTVTERVVEQHVARGSNDTMSLFNMTDCDEKDMSVDQSAVFAVPRVLQQNTTI
NQSVLMDLDESSVETGTAPGPSNNDTMSLFNMSTASAKDMDISNATMNETFKEPKSPAR
AFMPKVIEEDNEEIDMELTQMEVEESIVQNQNTIAFNKSMNQMAAKSAVFAVPCVLKQN
ATINQSMLMDLDESSVGIGPAAVPSNNDTMSLFNMSTASAKDMDISNATINETFKEPKSP
ARVFMSEVIEEDNEEVDMELTQMEVEESSVQNQKTIAFNSNMSTASAKDMDISNATMNE
TFKESKSPSHVFMPNVVEEDNEEIARDLTQMEAEESIVQNQNTIPLNESIDQMTATSAVFT
VPCVLNENATINQSMLMDLNESSVGTEPAPVPSNNDTMSLFNMSTASVKEKVISNVTMN
QTLKEQKSPARALMPKVIEEDNEQIDSDLTQLGAEESIVQDQKTIPLNFDMSTASVKDKDI
SNTTMNETFKEPKSPARAFVPKVIEKDNDEINMELTQMEAGESIVQDQKTIPLNFDMSTA
SVKDKDISNATMDETFKEARSPAPVFIPKVIEEANEQIDRDLTKMETEESIVQDQKAIPLNS
NMSTASVEEKDISNATTNETFKEPRSPAHVFIPKVIEEDNEQIDRDLTQMKAEESIVQNQN
AIAFSELLDQMAATLLQMKELKTEVQMQIEEREEMLARLQTLEDSTTEMKRKTEEIRDSI
RDY
>C_becei_KNL1
MDRNRRSSILKVRQHIDIDETTVAGNGLPGASTTTNRRVSFHNVKHVKQYDRDHGKLID
ASPLREKITDTMDSDGILTPRGRTLDASVAAANQSIHVFDNIDQGTVDMSLETVNENAGC
ETMQLFNITREQTVVIEKSVTRTVTERVVERHVAPGSNDTMSLFNMTDCDEKDMTVDQS
AVFAVPRALKQNATINKSMLMDLDESNFGSGPAPGPSNNDTMSLFNMSTASAKDMDITN
ATMNETFKEPKSPARVFMPRVVEEDNEEIDMELTQMEAEKSIVQNQKTMALNESMDISV
APSPRRPNANDTMALFRSPARADQSSMVLETKKTIFEAEPEDVDMEITQMDEQKPMEQQ
PAMNETMDISVAQNAIVNNTMALFQSPANQANQSYQQMKIVAEDDNGVEMEIMQTEEP
ERMSQTMDISVAPTGSPALPGNNDTMAMFQSPATVNQSAVVIPVNTPKPFDVSMEQEEE
GAGEGTFIAQNDISLDDQSPEAVTTPITQPTNFEENKNEFIIAVEPSSIEKEEERAVGQASM
MDISNMQDSTLDQKSLILQASPAQESGNHSILHESHVSETNNESVMSPQKTMQLETDMT
NTIHEMTPIENEDVSMQVEEQVEIQEKLVDDVVEKETTMTSQLTMSFTSANKSKSHLSRR
RSLLRESILESPRRIAIENSMITMNIGEMTALSEHRRRKSLQKSQQNILNSSAASLLNESLN
VSKGRDIFALNTSVRSPRVTGTPQSGNDSTIAAIPETSSFHMPNFDPAITNMVWLTPEEGE
EPLPEAAEFASLIAAETLKAEQEVQRTLQNENLDKEKAEAVRTGQWAKLSHDEKDAVLL
ARQDAEIRFLEMRLAFASKTRLKYEKVVQEQLEKNTRMAEQLNDIEDMDVLRQEIEELK
KRPTEADFRRAKNELKATRAALVERQAHHLAQMATGVMQMRNQITELRMEIEEREEM
MARIETTVDSEMEKIRNQTEDILESMRDY

>C_yunquensis_KNL1
MDNRNRRSSILKVRQHIDITDETTVAGNGLPGASTTTNRRVSFHNVKHVKQYDRDHGKL
MDASPLREKITDTMDSDGILTPRGRNLDASETAVNQSIHVFNNIDHGAADMSMETTVNE
NAGSDTMRLFDITREQTVVFEKSITTTVTERVVERHVAPASNDTMSLFNMTDCEEKDMS
VNQSAMFAVPRISKQNATINQSMLMDLNESNVGSGPAPGPSNNDTMSIFNMSTASANDM
DISKAINETVKEPKSPSRSFMPRVVEEDNEEVDMELTQMEAEKSIVPNQKTMALNVSMD
FSLAPSPRQIGANDTMSLFRSPARADQSNMILETTKKVFEDEDKTVYMDITQAEEEKTLV
QPAIEESMDSSMPQNVSISKTLALFQSPALESNQTHQPTQAVSDVVKMEVEMEEQEKMS
QSMDISVAPMDSTDLLGSNDSLALFQSPARINQSNVTVTIDKPKLFDESVEMEQEAGAGE
GTFLVQNELSSDGRSPETLPAPVAAQITHLANMETDQSEMKVAVEPSASEREEERAVTHT
SMMDVSGMHDSPTLDQKSFVLPASPAQESASQSILQQSHVSETNDESVMTPQKTMLLET
DLTNTIHESTPIDKENLSMKMEEQVEIHEKTVDDVIEKETTMTRELTMSFMNTSANRSKS
HLSRRRSLLRDSIVESPRRIAIENSMITMNAAEMTALSEHRRRKSLQITQQNLLNSSTASM
LNESLNASKSRDIFALNTSVRSPHFPGTPQSAHLSKDATIEVVPEASSFHMPNFDPAITNIV
WLTPEEGEGPVPEAAEFASLIAAETAKAEQEIQRTLQNENLNKEKAEAVRTGQWAKLSH
DEKDAVLIARQDAEIRFLELRLVFAEKTRQKYEKQAQEQLEANARLAEQLEDIENMDAL
REEIEELKNRPTEADYQRVKAEWRAAKREYYRLKNAEIVQILVTLQQIKEQKAQFKMEI
EEREEKLARLQALEDSELETMKHIVLDIHESLRDN
>C_macrosperma_KNL1
MDKRNRRSSILKVRQHIDITDETIAGNGMPGASTTTNRRVSFHNVKHVKQYDRDHGKLI
DASPLREKITDTMDSDGILTPRGRNLDASEMAANQSIRIFGNVENGNQTADMSLETVNEN
EGSETLKLFDITREQTVVFEKTVKTTTLVTERVTERVMAAGSLGDDTMALFNMTDCEER
DMSVDQSAVFAAPRAPRNNLTINQTVPMDFDESRVGPANDTMSLFNMSIASAKDMDISN
ATVNETLAKSPARSVMAKVVEENDEDVDMELTQMEAEKTLVQNQKTIGNESMNISVMP
TEPFDSPKRAAVNDTMSLFRSPARVNQSDMGMETTKKVFEQEEFVAEDVNMEITNVEEP
KTSEPLVLNETMDISQMRSESPKTTVVNETLALFQSPARGVNQSLQQVDSVKKVFHQEE
QHHVDMEVTQTEEQEEVPMSQPMDISIAPSDSPKLAKNNDTLALFQSPARLNQSALIASL
KKKELFDDSMEVEEEQGAGEGTFILPTESPASVPPEDVVIPFNQAINAEDHQSEMMLTEE
ASALLVREEERAMAHTSMMEVSGVQESVDINQKSLALPSDSAQETTDTSTLQEKDVSVD
KVRDVSINETTSHEQESAMSPQKTMQIEADLSTLHEMTPVQNEAVSMHLEEDGKMEVE
VHEKTSGEFVETETTISSHMTMSFSGSSRKPAAKSHLSRRRSLLRESVLESPRRIALENSMI
TMNVGEMTALSEHRRRKSLQLNQNNTTNTSVLNESANLSVARDIFALNTSVRSPHFPSAP
QSTNDSTVDAHSTTESTSFQMANFDPAVTNIVWLTAENEKPVPEAQEFARLIDDEKIRME
QEIEKSLLDDTVDQGKSEAVRDGQWSKLSHDEKEAVLIARQAAEIRFLELRHSFAQTHRQ
KYEREAQELLQSNAQMAGKLSEIENMETLQAEIDELKNRPTEADYHRIKAEWKAAKAE
SFKRKLAALGQIHDTLLLMKEQKSALKMEIEEREEKLTQLQITENTEKEAMKTIVEDIMK
SLKN
>C_sulstoni_KNL1
MEHQKKRLSILKVRQPLETLDETNALGQQGAPGESVTTTTRDRRVSFHNVKHVRQYHRD
HGNLIDASPLREKITDTMDSDGILTPRGRNLDASETGENATINVFGKQNGTMDMSLETEN
GGNDTMKMFEVTTHEKTITTVTERIVERHVVGIPKQNTAADDTMVLFNTTNNEEKEMTI
DESAIFAVPRLPNPNATSNQTVRMDLDESQVPQSNETMNMFNVSTASVHEMDISNAQNS
TNAPKLEDVDMELTQVEQPEEATSRVMNQSTMNQSMDISIAPGAAVGGVNETLAMFQSP
ARGALINESMDISMAPEQKSPAREVVAPAREVLAPTANNETLALFQSPASVRRNEHSTMI
LERTQRVFEDDPNDVEMEITQVEQPEESPRAINQSIMNQSMDISVAPAVNETLKMFESPAR
NRTIPIELFNESVEMNEETAIPIDDSVPMVEQVSEQSPLVAEEERAQVQSSMMMDISSVRE

SLVDQSMQKSSQETEETLTEKSVQETLEESMQASSPQQKTMQIEKDATITPMPEVFEETM
ENQKTFQENTFTKDISITMNSSSIRNTTLPMSNRRRSLIRECSSRAAVLESPRRVALENSMI
SMNVGNMTALVEHRRRKSIQMNQSQKELTNLNESIANMSAGRDIFALNTSIRSPKPMGTP
SDNVSLIHIAPESPPSFQMPNFDRAVTNIVWLTPDEHEDVAPPEAAEFARLIAMEKLKAEE
EVRMTLLDRKLDPKKVVSGGKVSSKLSNDEKEVVLIAREEAEIRFLELRREFAAQTRRK
YEEQVQCLETENARMAEQLRDVEQMEALKEEIEELRNRPTHADYERTRAEWKTAKREA
FEMQMEALRAAQETLIALREERLSLLIDIEERQERLAQAEEREEQEKTEMKAILEGLMSG
I
>C_afra_KNL1
MENQKKRLSILKTRQPLEILDETNALGQAGPGDSVTTTTRDRRVSFHNVKHVRQYHRDH
GNLIDASPLREKITDTMDSDGILTPRGRQLDASQLGENATINVFEKQNGTMDMSLASENA
GNETMKMFEVTTHEKTVTTVTERIVERHVFGIPKNTAADDTLALFNMTNNEEKDMTID
QSTVFAVPRLPEPKSASNQSMQMDVDESHVPQDSEKSPARGVLHQSMDISVASVQKSPA
RQRVPETNNETLAMFASPARGVLNESMDISMVPESKSPAREAILGPTGINETLALFQSPAP
VRQNQTASMILDTPKRVFEEDPSDVDMEITQVGCQEEEELPRVVNQSMLNQSMDISIVPP
VNETMKMFESPARIKTIPVEPFNESVEMETDDVVPANDPEPEQGSEQSPLMAEEERAQVQ
SSMMMDMSSIQESLVDGSALKTFQETDDTVTEKSLHESLGESMQIQSPQQKTMQIEQDT
TITRMSEPLEETMENVKSFQEETFVKDTSVSMNSSSVRNTTVPMSARRRSLIRDCSSRGA
VLESPRRVAIENSIVSMNAANMTALVEHRRRKSIQLNSNQKDLSNLNESVANMSAGRDIF
ALNTSVRSPMPSAASTDPAFSFHIAPESPATFLMPNFDPAVTNVVWLTPAENEDVPPPEAA
EFARLVAAEKLKAQNEIEKALLDGKLDPKKIVSRGKMSSKLTDDEKEVVLIARQDAEIRF
LELRRDFAAQTRQKYDLKIRELETENARIAEQIREFEQLEELKKEVEAMRNRPTYADYEK
IRAEWKAAKKEMFEMRMEALRSAQETLVSLEEEKMALLVDIEERQERLAQMEDRAQRE
KAEMKKILEGMMSSF
>C_sp49_KNL1
MDKKARRSSILKVRQPVDVLDETTAPSAATTTMNRRVSFHNVKHVKQYDREHGKIIEES
PLREKITDTLDSDGILTPRGRTDVPQYNESFSVFGQGKSNNVTMDMSLETVKTEDETIGM
FEGLQTICEAPEKENATIHGERDMSIDQSAMFAVPHQPNATINQSVLMDIDESDRTLKAN
DTMNMFAPRNNSPGRQNVTKVFSDDDMEMSQLEKTHMEEKNQEKELDMDISEAPTNK
NDTMQMFADKKTATFAEDKKVFGEEEMELTLVETETPEGVHESMDISMTPRKTIVREND
TIGMFHSLEQRQEHVDMDISQSFRADDTRQMFKSPGAKKGGDDMELTLQESPARATKNI
SMSKQKVFEEEGAGEEVPEENVSPGEVDRDQSSLHITLDKKEEERAEMQTCLMEMDGT
NTTFNHSQKTDLPSNESISPGKTLCEDTLASPQKTMQIASDLTNTFHETSPIREIVEEVFVE
HQSFHNDSCDMSVDMEVTSRLSVSYNRSGVLNRSSSSARKRRSLVRESPRRMAIENSML
SGAAEMSALGEYRRRKSLQTSQQHLLNSTNPTNTSLGRDIFAMNTSIRSPGRPGMLKKV
DDSPAPEKFSFELPEFDAAVANIVWLTSENGHPIPEAETFANLLNVEKSKTESEIQKTLSHV
DGAKFQAVRSGELSQLSHDEKEVVLFARGHAETRFLHLRQDFAQEHSQKYEEKISNIQLE
NAQIASQLLDLEDEEPLRHQLEEKRKHEVLMEADKIREEWRERERLETERRSERLAASA
ALILKMHQEKLAMQKEKEELDERMARLEDEERAVTEQISEEIKMVLAAAAAAH
>C_sp25KNL1
MDQKARRSSILKIRQPVDVLDETAAPSATTIPANRRVSFHNVKHVKQYDREHGKIIEESPL
REKITDTLDSDGILTPRGRTDGPQYDHSFSVFGQGTSSNATMDMSVETVHAEDDTIRMFD
GAKTFCAPSEKENVTIHEERDMSIDQSVVFAVPHPMNATINQSVLMELDESERTLRACDT
MDLFAPLTSTNNDNSVAPNDTLTMFQSPARKNATKVFSDDDMELTHVENQEHVDMDISE
APNNKTDKNDTLQMFSEVKQEKVVAETKVFEDEMELTLVEKAGDDSMDISMTPCKKTI
LPGDDTIGMFKSPEQKNNLQHDDMDISQSAEIHQMCKSPETKGGDDMEITLQESPARAV

KEVSMCSMQKVFDEEPKEKTLPEEIRDHSSTTIDGREKERSMLQTSLMEVVAAETTLNNS
EKTEIHSVSPGITLGSPQKTMQIASNLTSTFNETSPLQEPIQEVVEHIVEHQSFHNESYDMS
LDMEVTSRLSVSMNRSLTSRRRRSMLRESPRRVALENSMISFGAAEVTALEEYRRRKSLQ
KSQQNLLNSTNQADLSLGRNIFTMNTSIRSPKTIGKLTIPEEPAPVPFELPEFDAAVANIVW
LTSEDGNPIPEAKQFTNKALDLIDGAKFEAVRNGKLSQLSSDEKEVVVFARGHAESKFLH
LRHAFAEEHLQKYGEKIHELQMENAQLAGQLLDLEDEELLRREVEVKRKLVVLADADK
IRAEWRENQRLETERRIEQLTTSAALMLELHQERLAMLKEKEQLEEQIARLNIEQQAFDD
QLLENLEMINAKALE
>C_imperialis_KNL1
MDKKSRRSSILKVRQPVDVLDETAAAPSTTNAVMNRRVSFHNVKHVKQYDREHGQLIE
ESPLREKIADTLDSDGILTPRGRSVNVSYHNETLVDMSLETVVAENDTIRIFDTIHENELEH
KENGHQFAVPQLPNATINQSVLMELDESERASDTMNMFNPPANVSMDISSYTQNATKVF
ADDDMDVTSSRIQHEDTMKMFETPKKVNESMELTLVEEAPIQKTSDDMDISMTPSRIQK
APATQSASDDTVMMFQSPIQKANDDMDISITPSRILKTPKAQSAPEDTMVMFQSPIQSQK
SSDEMDISMTPSRTLQRTPRAQNAQNAPDDMDISMTPIQKTPRVQKTSDDVDMEITLLEK
TLDQNDTRQLFKSPARTQKMFVEEPEHAVAVNAPESPIAEAVAEIDRDESSMMITEQPSA
MDLTDDMDQKTLDENQKTSEFQKTLTPQKTMRIDTEMSDIQNTLHETSDIQKTPQKTSM
CNDVTSDHQNPAQKTLTPNQAVESPQKTMRIESETSELQKTFNETCEMSISRSHLSISFNT
SITSIRRRRSIQKRESPRRLAMENSMITMSMAPEQMSALGEYRRRELLQTSQQHLLNSTN
LADTSRDIFKMNTSIRSSPHTFSKQKPNEPAEPLKLPEFDAAIANVIWLTSDDAGALVPEV
EHFQLLLEQEAQKTKALIHNQDSPNSLSNQDEQEVLLVSRELAEQRFLDLRLKFAQEQSH
ELQIQKIQLENAQMASNLLDLEDEHQLLGELEQLEKQKVLANADQIRAEWREKKKAES
EQRVAAISGAVELMLQLHQKKMEMKKEMEELAERIAQLEEAEAKFKEQFIEQINTNVPY
>C_japonica_KNL1
MENKKKRLSILKVRQHVDLLDEATEVKAPGATTPTMNRRVSFHNMKQVRQYDRDHGK
MIEESPLREKITDTMDSDGVLTPRGRTDAGNSTINVFGNVNKIDQLANGTMNMSLETVN
VNDDTLKMFSGLRDRTISESTMMTEGVVLQKQNMTIREVRDMSIVDESEIFAVPQMPKP
NATINKSVFMDLDESQSHHGQASSQNDTMALFLDATNNANISMDISIGSARQNVTKTFD
DDDMEITQMDEKTLIRKEVDETVDMFESPARNHIAKKESDDVDMEITLVQMTPKSSTSK
KEEDVNETVGLFKSPARNQQQVFKEVEEENNDMDITQPQMTPKNKSIQQINTVSPFDSP
KIIEPNKSSEDVDMEITLPHNQMTPKLTTRQRTGEINETFDMFKSPAPKQIDTSKFLQNDE
NVDMEITQEPANVTSDSRSEKSSYDVVETNSEESLAILEKSHSLVEVAEQDRAMGNCSV
MNISDVQETLNLQNNSPKLHGFADISQESMNRSTTSMSSHKLPQDDTDSSPQKTMQIEST
TTLHEISERIEEVIEEKVENVEIEIEVEVEVEQKTFQTENYDVTVTESTTKKQIHALFNHSL
NGSQNRTHSARRRRSLLRSSVLESPRRIALENSMVSTSVARERTALEEYRRRKSLQTSQQS
ILDSSNLHDVSASARDIFALNMSVRSPHQPDTPKSKIAQEVVSPAPQFVPFQMANFDPAVT
NIVWLTPDEDSATSIPEALEFSNILAAEKTAVGDEIQKALEYGQIDEVKWEAVRNDLMPQ
MSQDEKEAVLIAREEAEIRFLQLRLRFATEHHQNYELKVNEIQTENKTISDKVMDLQNID
SLSAAINEFERREVVREKEKIVAEWREAKQMSWDRVSKKMNAAMKLLLEINQQREADR
KELEEIEEKIARLDDASKKERAEIMDTISRIKMAAQAGINVS

Supplementary Data 3.S8: All ZWL-1 sequences used in this study

>C_tribulationis_ZWL1
MSIKIEDLQNYNQAVSARGEKEQAPTPVLLLNKYRVRLIPISSLPFVSNYSNVSQLGLNSD
DALVIDSPIESSEKQKKSSLVPRRQAARVNEDEDIMETESVEEITEDKENSVTENDRGPLV
TSFLTLEKLEKNGDVEIVGLDCEVEFFDANPVPFEDGVSLQRLRLESSNHFSAAVNKYPI
WISTFGSKFPSLCWLAAGRTAKNVQIGGATRVLGYFKDDSEKLVKQLNEACGVAQINRY
RAVFDEIRRVATPTRDSPGEVIVDMRWTTKSSLVLLEQPDNAADCTIKIDLGWRDNRFFID
DAIFEQLFFVLNLADVLANPEKEVEFPPEFDKFDYLVQEMNDLVEACSREDNVFASNEFR
SEEVTDKVWNIVRQCGDVKRATMLFKNFLQALTYGKIKSHVQEGNKSHLASLIRASKTC
DFRMPILERLSTIEMMMEIGVESLRRIINKFSNTLQFPSDELTFILKTCENDLSSGEGAINAS
VISLLPITMALATVYQIFGLLNVKDHVILPDLARRVLTKFTSSMVEKAKRGETETDYTFET
TLPLLRINKEIFMDKRPRVWTCENANTVGANVQTRIMTALELDVSLEHVSRLVNASRPV
RLIDEENRKPTAEELKADYTVSHTIFSYLPKL
>C_sp41_ZWL1
MSIKLEDLQSYNEAGLARGEGDDNPIPVLLLNKYRVRLMPLSSLPFVSNYSNISQLGLNS
DEALVIDSPIESSEKQKKSSSLFAKRETVTVKEDDEKMETDSVEEDKENSVTEEHNGPLV
TSFLTLDKLEKNGANDVEIVGLDCEVEFFDANPIPFEDGVSLQRFLRLESSKHFSAAVNKL
PIWISTFSSKFPSLCWLAAGRTNKNVQVAGATRVLGYFKDDSEKLVKQLNEACGAAQVN
RYRAVYDGIRRIATPTRDAPGEVIVDMRWTTKSSLVLLEQPDNAADCTIKIDLGWRDNRF
FIDDAIFEQLFFVLNLADVLANPEKDVVFPPERDNYDNLVQEMKDLVEACSHEDNVFAS
NERSEEVTDKVWNIVRKCGDVKHATMLFKNFLQALTYGKIKSHVQEGNKSHLASLIRAS
KTCDFRMPILERLSTIEMMMEIGVESLRGRIINKFSDTLQFPSDELTFILKTCENDLSSGEV
ALNASVVSLLPITMALATVHQIFGLLNAKDHVILPDLARRVLTKFTSSMVEKAKRGETET
DYTFETTLPLLRMNKDTFMDKRPRIWTCENTNTVGANVQARIMTALELESSLEHVNRLV
NASRPVRPVEEENKKPTTEELNADYTVSHTILSYLPKL
>C_zanzibari_ZWL1
MSIKLEDLQSYNEAVSARGEEDEAPTPVLLLDKYRVRLIALSSLPFVSNFSNVSQLGFNSE
DVLVIDSSIESSEKQKKANIFVKRETVELDDEEMETESIEEDKENTTNEDEGGPLVTSFLTL
DKLERNNDVEIVGLDCEVEFFDTNPIPYEDGISLQRLRLESSKHFSAAVDKLPIWISTFGSN
FPSICWLAAGRTNKNAQFAGATRVLGYFKDDSEKLIKQLNESCGAAQVNRYRAVYDGIR
RIVTPTRPSPGEVIVDMRWNTKSSLVLLEQPDNAADCTIKIDLGWRDNRFFTDDAIFEQLF
FVLNLADVLANPEKEVIFPVEYDKFDSLVQEMKDLVEACSHEDNVFASNEFRNEEVTDK
VWNIVRKCGDVQHVTLLFKNFLQALTYGKIKSHVQEGNKSHLASLIRDSKTCDFRMPIL
ERLSTISMMMEIGVESLRRRIINKFSDTLQFPTDELTFILKTCENDLSSIEGALNTSVVSLLP
ITMALATVHQVFGLLNVKDHVMLPDLARRVLTKFTSSMVEKAKKGETETDYTFETTLPL
LRMNKDTFMYKRPRIWTCENTNTVGANVQARMMTSLELESSLEHVNRLVNASRPIRPID
EENRKPTAQELNADYTVSHTVFSYLPKM
>C_sinica_ZWL1
MSLKLEDLQSYNEAVLNREEEDEAPAPVLLLDKYRVRLIALSSLPFVSNYSNVSQLGLNS
EDVLVIDSPIQSSEKQKKSSLFTKRATVPVKEDDEKMETESIEEDKENTAAENEGPLVTSF
LTLEKLEKNDVEIVGLDFEVEFFDANPIPYEDGISLQRLRLESSKHFSPAVANLPIWISTIGS
NFPSVCWLAAGRTNKNMQFVGATRVLGYFKDDSEKLVKQLNEACGAAQINRYRAVYD
GIRRIATPSRDSPGEVIVDMRWNTKSSFALLEQPDNAADCTIKIDLGWRDNRFFIDDAIFE
QLFFVLNLADVLVNPEKEVVFPTEYDKFDHLVQEMKDLIENCSQEDNVFASNEFRSEEV

TDKVWNIVRKCGDVKHATMLFKNFLQALTYGKIKSHVQEGNKSHLASLIRASKTCDFR
MPILERLSTIEMMMEIGVESLRRRIINKFSDTLQFPSDELTFILKTCENELSSGEGALNASA
VSLLPITMALATVHQIFGLLNVKDHVILPDLARRVLTKFTSGMVEKAKRGETETDYTFET
TLPLLRMNKDSFMDKRPRIWTCENTNTVGANVQARIMTAMELESSLEHVSRLVNASRPV
RPIDEESRKPTAEELNADYTVSHTIFSYLPKL
>C_nigoni_ZWL1
MSIKLEDLHSFNEAVLSKGEEDEAPAPVLLLDKYRIRLVPITELPIVSNYSNTSQLGLNSEE
VLVIDSPVESAEKQKTSSLLNRRECKKTIKSEKEDEPMDMETTEGDKENTVSEIGGGPLV
TSFLTLDKLEKDGVNDVEIVGLDCEIQFFDANPIPFEDGISLQRFLRLEGSKHFSAAVDKL
PIWISTIGHHFPSVCWLAAGRTNKNVQVSGATRILGYFNENSQKLVKQLNEACGAAQVN
RYRAVYDGIRKVATPTREAPGEVIIDMRWNTKSSLVLLEQPDNAADCTIKIDLGWGDNRF
FIDETIFEQLFFVLNLADVLANPEKEVVFRSESDKFDDLVQEMKQLVEACSHEDNVFASN
EKSDQVTDKVWNIVRKCSDVKQATMLFKNFLQALTYGKIKSHVQEGNKSHLASLIRAS
KTCDFRMPILERLSTIEMMMEIGVESLRGRIINKFSDTLQFPSDELTFILKTCENDLSTLEG
TLHSSVVSLLPITMAMATIHQIFGLLNAKDLVVLPDLARRVLTKYTTSMVEKAKRGETET
DYMFETTLPLLRMNKEAFMHKRPRIWTCENTNTVGANVQTRAMTTLELEPSLEHVSRL
VNASRPVRPIDEDNRKPTEEERNADYTVSHTIFSYLPKL
>C_briggsae_ZWL1
MSIKLEDLHSFNEAVLFKGEEDEAPAPVLLLDKYRVRLVPITELPLVSNYSNTSQLGLNSE
EVLVIDSPIESAEKQKTSSLLNRRENKKTIKSEKEDESMDMETAEGDKENTVSETGGGPL
VTSFLTLDKLEKNDVEIVGLDCEIQFFDANPIPFEDGISLQRLRLESSKNFSAAVDKLPIWI
STIGHHFPSVCWLAAGRTNKNVQVSGATRILGYFNENSQKLVKQLNEACGAAQVNRYR
AVYDVIRKIATPTREAPGEVIIDMRWNTKSSLVLLEQPDNAADCTIKIDLGWGDNRFFIDE
TIFEQLFFVLNLADVLANPEKEVVFRSESDKFDDLVQEMKQLVEACSHEDNVFASNEKF
RSEQVTDKVWNIVRKCSDVKQATMLFKNFLQALTYGKIKSHVQEGNKSHLASLIRASKT
CDFRMPILERLSTIEMMMEIGVESLRRRIINKFSDTLQFPSDELTFILKTCENDLSTLEGTL
HSSVVSLLPITMAMATIHQIFGFLNVKDLVVLPDLARRVLTKYTTGMVEKAKRGETETD
YVFETTLPLLRMNKEAFMHKRPRIWTCENTNTVGANVQTRAMTTLELEPSLEHVCRLV
NASRPVRPIDEDNRKPTEEERNADYTVSHTIFSYLPKL
>C_remanei_ZWL1
MALKLEDLHNYNTLVATKGEEDEEPTPVLLLNKYRIRLLPLSSLPFVQNYSNISQLSLNSD
DVLVIDSPIGATEKYSKKSIENSKKSMIKEEDESMEMEEDKENEGGEIGPLVTSFLTLEKL
QKGVNDIEIVGLDCEVQFFDANPIPFEDGVSLQRLRLESSAHFSPAVDKLPIWISTISQNIPS
VCWIASGRTQKNVQFSGVTRVIGHFNEINQKLIKQLNEAAGAAQVNRYRAVYDGIRKIA
TPTRESPGEVIIDMRWNTKSSLVLLEQPDNAADCIIKIDLGWRDNRFFIDESIFEQLFFVLN
LADVLANPEKEVIFPVEYYVKFGDLVKEMDEIVEACSHEDNVFASNEKFRNEEVTDKVW
NIVRKCGDIKHATMLFKNFLQALTYGKIKSHVQEGNKSHLASLIRASKTCDFRMPILERL
STIEMMMEIGVESLRRIINKFSSTLQFPADELTFILKTCENELTTGEGALNSSVVSLLPITMA
LATVNQIFGLLNEKDYVLLPELARRVLTKFTSGMIEKAKRGETETDYTFETTLPLLRMSK
EKFMDKRPRIWTCENMSTVGAHCQTRIMTSLELESSMEHVSRIVNAGRPIREIVETEETQ
RKPTIEELNADYTVSHTVFSYLPKM
>C_latens_ZWL1
MALKLEDLHNYNTLVAEKGEEDEEPTPVLLLNKYRIRLLPLSSLPFVQNYSNISQLSLNSD
DVLVIDSPIGSSEKYSKKSLEISKKSAIKQEDESMEMEEDKENEGGEDGPLVTSFLTLEKL
QKGNDVEIVGLDCEVQFFDANPIPFEDGVSLQRLRLESSAHFSPAVDKLPIWISTISSNIPS
VCWIASGRTQKNIQFSGVTRVIGHFNESNQKLIKQLNEAAGAAQVNRYRAVYDGIRKIAT

PTREAPGEVIIDMRWNTKSSLVLLEQPDNAADCIIKIDLGWRDNRFFIDESIFEQLFFVLNL
ADVLANPEKEVVFPMEYVKFGDLVKEMDEIVEACSHEDNVFASNEKRNEEVTDKVWNI
VRKCGDIKHATMLFKNFLQALTYGKIKSHVQEGNKSHLASLIRASKTCDFRMPILERLSTI
EMMMEIGVESLRRRIVNKFSSTLQFPADELTFILKTCENELTTGEGSLNSSVVSLLPITMAL
ATVNQIFGLLNEKDYVLLPELARRVLTKFTSGMIEKAKRGETETDYTFETTLPLLRMSKE
KFMDKRPRIWTCENMSTVGAHCQTRIMTSLELESSMEHVSRIVNAGRPIREIETDAETQR
KPTIEELNADYTVSHTVFSYLPKM
>C_sp51_ZWL1
MPIQLEDLQNYNESVKDWVEGENEPTPILLLNKYRVRLMSVSSLPFVSNYSNISQLGFSS
DDVIVIDSAIETFQKEKRIHSLPEASKKQSCAAINEENVSMETEDFEHDKENTAVDEGPLV
TSFLTLDKLQKNAVDDVEIVGLDCEIQFFDVNPVPFADGISIQRYLRLESSKHFSTDVDKL
PVWISTMSSHFPSVCWLASGRTNKNVQVSGVTRILGYFNESSQKIVKQINEACGAAQVN
RYRAVYDGIRKIETPNRNSPGEVIVDMRWNTKSSLVLLEQPDNAADCTIKIDLGWKDKR
FFIDEAIFEQLFFVLNLADVLANPEREVVFPSEWAKFDDLVHEMNELVDSCSQEDNVFVS
NEKSEEVTDKVWNIVRKCGDVKHATMLFKNFLQALTYGKIKSHVQEGNKSHLATLIRAS
KTCDFRMPILERLSTIEMMMEIGVESLRGRIINKFSSTLQFPSDELAFILKTCENDLSTGEE
ALNASVVSLLPITMALATVSQIFGLLNEKDHVILPDLARRVLTKFTSSMVEKAKRGETET
EYLFETTLPLLRISKDKFMDKRPRIWTCENTNTVGANVQTRIMTTLELESPVEHVCRLVN
ESRPIRETKEENEKPSVDELNADYTVSHTVFSYLPKM
>C_sp44_ZWL1
MSLKLEELQSYNDAVKSRKEDEAEPAPLLLLNKYRVRLMPISSLPFISNFSNISQLGFNSD
DIIVIDSPIGIQKEKKISFVDSAAKQNDMADSMETEEAHEDHDKENTVTHGGPLETSFLTL
TKLQKNAVDDIEIVGLDCEVQFFDVNPIPFADGVSVQRYLRIESSNHFSPTVDRLPVWIST
STSHFPSVCWLASGRTNKNIQVSGATRILGYFDENSQKIVKQMNEACGAAQVNRYRAVY
DGIRKIVTETRKYPGEVIVDMRWNTKSSLVLLEQPDNAADCTIKIDLGWKDKRFFIDDAI
FEQLFFVLNLADVLANANTEKPVEFPLEWAKFDDLVQEMNDLVDACSQEDNVFVANEK
SEEVTDKVWNIVRKCGDVKHATMLLKNFLQALTYGKIKSHVHEGNKSHLASLIRASKTC
DFRMPILERLSTIEMMMEIGVESLRGRIINKFSSTLQFPSDELSFILKTCEKDLSTGEEALN
ASVVSLLPIAMALATVCQIFGLLNEKDHVILPDLARRVLTKFTGSMVEKAKKGELETNYT
FETTLPLLRMSKDKFMDKRPRIWICENTNTVGANVQTRTMTALELESSVEHVNRLVNAS
RPIREMKEENEKPTAEELNADYTVSHTVFSYLPKM
>C_sp48_ZWL1
MSLKLNDLQSYNEQVESWVDGEEKPTPLLLLNKYRVRLMPVSSLPFVSNYSNISQLGFN
SDDVLVVDSTVELFQKEKRSSVLNDATKHNIITIDDSMETEDFEQNKENSTVDGGPLVTS
FLTLDKLQKNAVDDVEIVGLDCEVEFFDVNPIPFADGVSIQRYLRLESSKHFSPAVDKLPV
WISTISSHFPSVCWLAAGRTTKNIQVSGVTRVLGYFNENSQKVVKQMNEACGAAQVNR
YRAVYEGIRKIATPTRDAPGEVIVDMRWNTKSSLVLLEQPDNAADCTIRIDLGWRDKRFF
IDEAIFEQLFFVLNLADVLANPEKEVVFPSESAKFDDLVQDMNELVDSCSQEDNVFVSNE
KSEEVTDKVWNIVRKCGDVKHATMLFKNFLQALTYGKIKSHVQEGNKSHLASLIRASKT
CDFRMPILERLSTIEMMMEIGVESLRGRIINKFSSTLQFPNDELSFILKTCENDLSTVEESL
NASVVSLLPIAMALATVSQIFGLLNEKDHVILPDLARRVLTKFTSGMVEKAKKGETETEY
TFETTLPLLRMSKENFMDKRPRIWTCENTNTVGANVQTRIMTALELESSVEHVSRIVNAS
RPVREMKEENEKPTAEELNADYTVSHTVFSYLPKM
>C_brenneri_ZWL1
MSLKLNDLQSYNEQVKSWVDGEEKPTPLLLLNKYRVRLMSVSSLPFVSNYSNISQLGFN
SDDILVVDSTVELFQKEKRSSILDNAEKRSNIKIDDSMETEDFEQDKENNTVDGGPLVTSF

LTLDKLQKNDVEIVGLDCEVQFFDVNPIPFADGISIQRYLRLESSKHFSPAVDKLPVWISTI
SSHFPSVCWLAAGRTTKNIQVSGVTRVLGYFNENSQKIVKQMNEACGAAQVNRYRAVY
EGIRKIATPTRDAPGEVIVDMRWNTKSSLVLLEQPDNAADCTIKIDLGWRDKRFFIDEAIF
EQLFFVLNLADVLANPEKEVVFPSESAKFDDLVQDMNELVDSCSQEDNVFVSNEKRSEE
VTDKVWNIVRKCGDVKHATMLFKNFLQALTYGKIKSHVQEGNKSHLASLIRASKTCDF
RMPILERLSTIEMMMEIGVESLRRRIINKFSSTLQFPTDELSFILKTCENDLSTVEDALNAS
VVSLLPIAMALATVSQIFGLLNEKDHVILPDLARRVLTKFTSGMVEKAKKGETETEYTFE
TTLPLLRMSKENFMDKRPRIWTCENTNTVGANVQTRIMTALELESSVEHVSRIVNASRP
VREMKEENEKPTAEELDADYTVSHTVFSYLPKM
>C_wallacei_ZWL1
MAIKLEELQNYNEAVNAWEEGKEIPTPILLLNKYRLRLMSISSLPFVSNYSNVSQLGFNSE
DVLVIDSPIDSPEKRKTSTLFRNSSKQSDAETKNNDENMETEETEEDDKENSVTHGGPLA
NDIEIEGLDCEFNFFDANPIPFADGISLQRFLRIESSNNFSPAVDKLPVWISTISPTFPSVCWL
ASGKTNKNIQVSGATRILGYFDENSQRTVKQIHEACGAAQVNRYRAVYDVIRKISTPNRS
APGEVIVDMRWNTKSSLVLLEQPDNAADCTIKIDLGWRDNRFFIDDAIFEQLFFVLNLAD
VLANPEKEVVFPSEWAKFEDLVQEMNELVDACSKEDNVFASNEKSEQVTDKVWNIVRK
CADVKHATMLLKNFLQALTYGKIKSHVQEGNKSHLASLIRASKTGDFRMRILERLSTIEM
MMEIGVESLRERIINKFSSTLQFPRDELIYILESCENDLSSTGEDALNTSAVSLLPITMALAT
VTQIFRLLNEKDEVILPDLARRVLTKFTSSIVEKAKKGETETDYIFETTLPLLRMSKDTFM
YKRPRIWTCENTNTVGANVQTRMMSTLELESSVEHINRMVNASRPHRELKEENAKPTA
EELKADYTVSHTVFSYLPKL
>C_doughertyi_ZWL1
MPLKLEELQSYNEAVDAWRDGDDFPTPVLLLEKYRLRVLPLSSLPFVSNYSNISQLGLNS
EDVLVIDSPVESPEKQKRSSRSETDVKPSPMIKQESLSMETDDSGEDKENTVADGGPLVT
SFLTLEKLQQNNDVNIVGLDCEVEFFDINPIPFEDGVSLQRLRLESVKYFAPTVEKLPVWI
STMSSNFPSVCWLAAGRTNRNVQFAGVTRILGYFDETSQRTVKQMNEACGAAQVNRYR
AVFDGIRRIATPTRSSPGEVIVDMRWNTKSSLVLLEQPDNAADCTIKIDLGWRDNRFFIDE
AIFEQLFFVLNLADVLANPEKEVVFPSEWAKFDDLVQEMKLLVDICSKEDNVFASNEKFR
SEEVTDKVWNIVRKCGDVKHATMLFKNFLQALTYGKIKSHVQEGNKSHLAALIRASKT
GEFRMPILERLSTIEMMMEIGVESLRSRIINKFSSTLQFPSDEVAFILKSCEAELSNGDGAL
NSNVISLLPITMALATVNQIFGLLNEKDHVILPDLARRVLTKFTGSMIEKAKKGETETEYT
FETTLPLLRMSKDKFMDKRPRIWTCENTNTVGAHVQTRMMTSLELESSVEHVNLMVNA
SRPIRDVKEGNEKPTAEELKADYTVSHTVFSYLPKM
>C_sp54_ZWL1
MSVKLEELQSYNEAVIAKGEDEKAPAPILLLDKYRLRLVTLSSIPFVSNYSNISQLGLNSD
EVLVIDSSTELMGKEKKSSIFEKMPIKREEEHMETDDIEDDKENAATEDGPLVTSFLTLDK
LQKDAVNDVEIVGLDCQVQFFDVNPIPFEDGVSLQRYLRLKSSEHFSATVDKLPVWISTIS
SNFPSVCWLSSGRTNKNVQISGVTRILGHFNENSHNRYRAVYDGIRKIATPTREAPGEVII
DMRWNTKSSLVLLEQPDNAADCTIKIDLGWRDKRFFIDEAIFEQLFFVLNLADVLANPEK
EVVFPSEWPKFDDLVQEMNDLVEACSHEDNVFASNEKSEEVTDKVWNIVRKCGDVKHA
TMLFKNFLQALTYGKIKSHVQEGNKSHLASLIRASKTCDFRMPILERLSTIEMMMEIGVE
SLRGRIINKFSSTLQFPSDELTFILKTCENELSSGEGLLNSTVISLLPITMALATVNQIFGLLN
EKDHVILPELARRVLTKYTSSMVEKSKRGETETDYTFETTLPLLRMSKDKFMDKRPRIW
TCENANTVGANVQTRMMTALELESSLEHVNRIVNASRPVRDIKENEKPTVGELNADYT
VSHTVFSYLPKI
>C_inopinata_ZWL1

MITLEQLQKYNEETISKKEDDEPAPVLLLNANRVRLMQLSSLPIVSNYSNISQLGLNSDNV
IVIDLPIGSAEKTKTKASALENRKLPEQHEQMEMEEDKENTATEGGPLITSFLTLDKLQDI
EIVGMDCEEQFFEGNPIPFEDGISLQRLRLNASKHFSPSIDKLPVWILTTSSHFPSVCWLSS
GRTNKNVQFTGATRVLGHFDEKIHKLVKQMNEACGASHFNKYRAVYDEIRKVATPTRVS
PGEVFIDVRWNTKSSLVLLDHPDNSADCTIKIDLGWKDKRFFTDEAIFEQLFFVLNLADV
LVNPEKEVVFPSQYAIFDDLVQEMNVLIEASSHEDNVFASNEKFRNEEITDKVWNIVRKC
GNLNHATMLLKNFLQALTYGKIKSHAQERNKSHLALLIRASKTCEFRMPILERLGTIAM
MMEIGVESLKRRIINKFSNTLQFPSDELAFILKSCERDLANGEITLNSSTVSLLPITMAFAT
VNQIFALLNEKDHVILPDLARRVLTKYTSSMIEKAKRGETETEYSFETTLPLLRMCKERFS
EKRPRIWTCENTCTVGANVQARTMTLLELESSLEHVTSLVNAHRPVREMKKEDEKASRE
DLEADYTVSHTVFSYLPKM
>C_elegans_ZWL1
MPLTIEQLKQYNEKMAAKGEEDDAPAPVLLldkyrvrvlpltsipfvmnhssvsqlslssedvlvvdfpskgtgs
gagkaekkmifkkkiepmedsfedkenvneggplmtsfltleklrkgmVgdveiigyecetsffdanpiplsdgvslqrYirlncse
hfspsistlpvwisttsskfpsicwlaagrtnrnvqfaaatrvlghfnnenservvkqlnqacgtsqlnkyravyeeirkiatenrpapgev
tidvrwstkstlvllehpdnaadctikidlgwgdkrffvddeifeqlffvlnladvlanpdnevifpmappanfddlvqemdalveassr
ednvfvsneNfrggditdkvwnivrkcndvkqvtllfrnflqalaygkikshaqernkshlasliriskssefkipvlerlstidmmmei
gveslrRrvidifssqllypsdelefilqtcendlpagngamnsaaisllpitmalatanriyellnekdhvilpdltrrilqkytasmieksrr
geteteytfettlpllrmykegfmskrpciwtcensntvganvqarvltslelqpslehinrlvndsrpvwtateekpamikdlnadytvv
htifsylpkm
>C_oiwi_ZWL1
MLLALEDLHNYNRGIERMSEDENEEEPEPLLLLNKYRLRLIPIDSLPIVSNHSNINHLGLN
SEDALVIDLPVGPSKKETKLSVSESFSTKENQTVKMEGVEENKENASRNGGPLETSFLTL
DKLQKGSGNDDEVEIVGFDSEVDFFATNPIPFSDGVSLQRYLRIASKEHFSPLVDKLPVWI
STIGSDFSSVCWLSSGRTIKDGRFSAITRVIGHLNEHTEKLVKQLNEACGTAQSNRYRCIY
DGIRRISNSTREAPGEVIIDMHWNKKSSLVLLEQPDNAADCTIKIDLGWRDNRFYIDEAVF
DQLFFVLNLADVLANPEKEVIFPTELVKFDDLVKEMNELVEACSHEDKVFASNEKDDEV
TDKVWNIVRKCGDVKHATMLLKNFLQALTYGKIKSHVQEGNKSHLASLIRASKTSDFR
MPILERLSTIEMMMEIGVGYLRGRVINKFSSTLHFPSDELKFILTKCENDLSKDEGSLNSS
VVSLLPITMALATVNQIFGLLNEKDHVVLPELTRRVLAKYTSSMVEKAKKGETETDYTFE
VTLPLLRISKERLMDKRPRIWTCENTNTVGGNAQTRVMIALELESSLEHVNTLINASRPV
REAKENDEKPNKEEMNADYTVSHTVFSYLPKM
>C_kamaaina_ZWL1
LNKYRLRLIPIDSLPIISNHSNINHLGLNSEDALVIDLPVGLSKKEKKLAVSESFSPKENQS
MKMEGVEENKENTSGNGGPLETSFLTLEKLQKGEVEIVGFDSEVEFFATNPIPFSDAVSLQ
RYLRMASKEHFSPLVDKLPVWISTIASDFPSVCWLSSGRSSKDGRFSGITRVIGNLNEQTE
KLVKQLNEACGTAQSNRYRCIYDGIRRISNSTREAPGEVIIDMHWNKKSSLVLLEQPDNA
ADCTIKIDLGWRDNRFYIDEAVFDQLFFVLNLADVLANPEKEVIFPTEFVKFDDLVKEMN
ELVEACSHEDKVFASNEKFRDDEVTDKVWNIVRKCGDVKHATMLLKNFLQALTYGKIK
SHVQEGNKSHLASLIRASKTSDFRMPILERLSTIEMMMEIGVGYLRRRVINKFSSTLHFPS
DELKFILNKCENDLSKDEGSLNPSVVSLLPITMALATVNQIFGLLNEKDHVVLPELTRRVL
AKYTSSMVEKAKKGETETDYTFEVTLPLLRISKERLMDKRPRIWTCENTNTVGGNAQTR
VMIALELESSLEHVNTLINASRPMREVKENDEKPNKEEMNADYTVSHTVFSYLPKL
>C_waitukubuli_ZWL1
MSVSVEQLQEYSRAVEEAAEEDAAPTPVLLLDKFRLRLFKITDLPIISNYSNIPQLGLSSD
DVLLIDLPTGTEKKAKNVIKSTIPSSVESESEEVMEGIEEDEEDDKENTGGPLVTSFLTLEK

LQKGLVDLEIAGFDFEAPKFDVNPVQMSDGISIQRLRLQSSDHFTAALSKLPVWIATSLPS
VPSVCWLATGRTAKNLQFSGATRVLGYLNESTNKFVNQLNEACGAASVNRYRAVYDNI
RRIATQTRENPGEVIVDMRWHTKNSLVLLEQPDNAAECTIKIDLGWRDKRFYIDESIFEQ
LFFVLNLADVLANPEKEVVFPNEFEKFDDLVELMNKLVETCSDEENVFVSKDFRSEEVT
DKVWNIVRRCGDIKQATLLFKNFLQALTYGKIKSHVLEDNKSHLAALIRASKTCEFRMPI
LERLSTIEMMMEIGVQSLRRRIVDKFSNTLKFPSDELSLVLKTCENDLKSVECVMNVSVV
SLLPITMALATVHQIFGLLNDKDHVILPDLARRVLMKYTSSMVEKSKKGETELNYSFETT
LPLLRINKDRFMDKRPRIWTCENSKHVGANVHARMMTCLELETSLENVSSLVNSIRPVR
ETKEDEKPTVEELNAEYTVSHTVFTYFPKI
>C_panamensis_ZWL1
MLVGIEHLQEYNKAVESKTEEDEAPVPVLLLDKFRLRLIPITDIPIILNYSNIPQLGFSSDDV
LLIDSPTGSEKASKITKTLKSVKVDAEEEEQMEVEEEEGDKENTATEGGPLQTSFLSLNKL
QKRAEMEIDGFDFQAPKFDVNPIPLGDGISIQRSLRLESSNHFSAALSKLPVWIATTDSNL
PSACWLATGRTGKNVQFSGATRILGYYNETTDKFVNQLNEAGGGALVNRYRAVYDRIRR
IATKDRENPGEVMVDMRWNTKNSLVLLEQPVNAAECIIKIDLGWKDKRFYIDESIFNQLF
FVLNLADVLANPDNEVVFPNESVKFDNLVKLMNELVELCSDEENVFVAKDFRSEEVTDK
VWNIVRCCGDIKQATMLFKNFLQALTYGKIKSHVLEDNKSHLAALIRASKTCEFRMPILE
RLSTIEMMMEIGVQSLRRQIVDKFSKTLKFPSDELALVLKTCENDLKSVDCNLNASVVSL
LPITMALATVNQIYGILNEKDHVILPDLARRVLMKYTSSMVEKCKRGETEVNYSFETTLP
LLRINKEKFMDNRPRIWTCENSKHVGANVQTRMMTSLELETALDHVCSLVNSVRPTSEA
KDNETRTAAELNADYTVSHTIFTFSPKV
>C_nouraguensis_ZWL1
MLIGIEQLKQYNAAVESKAEDDEAPAPVTLLDKFRLRLMPITDLPIISNYSNIAQLGFYSD
DILLIDAPTGSAKKAKIAHMVLPTRQDNSEAMEVGDEEEDKENSAMEGGPLQTSFLTLE
KLQRDVEIAGFNFEAPKFDVNPIALFDGISIQRLRLESTSIFSAAVSKLPIWIATTQSDFPSM
CWLAAGRTNKNLQFSGATRVLGYYNETTDKFVNQLNEACGGALVNRYRAVYDGFRRIA
TQTRQHPGEVMMDMRWNTKNSLVLLEQPVNAAECTIKIDLGWKDERFYIDESIFEQLFF
VLNLAEVLANPEEEVKFPNKWEKFDDLVKLMNELVLKCSDEENVFASKDTSEEVTDKV
WNIVRRCRDIKEATLLFKNFLQALTYGKIKSHVLEDNKSHLAALIRASKTGEFRMPILERL
STIEMMMEIGVQRLRIVDKFSKTLKFPSDELTLMLKSCENDLKSVECSMNASVVSLLPIT
MALATVHQIFGILNEKDHVILPDLARRVLMKYTSSMIEKSKRGETELNYSFETTLPLLRIN
KDMFVDKRPRIWTCENSKHVGANVQARMMTSLELEMSLEHVSDLVNSIRPLRETKQDK
IPTTEELNAEYTVSHTVFTYYSKV
>C_becei_ZWL1
MLVGIEQLQQYNTKVESKTEDDEAPAPVLLLEKFRLRLMPITDLPIISNYSNIAQLGFTTD
NVLLIDAPARSAKKAKIAQVPQLTKRDDVEVMEVGDEEEEDKENTAMEGGPLETSFLTL
EKLQKSSVDVEIAGFNFEAPKFDVNPVALCDGISIQRYLRLESSAVFTAAVSKLPVWIATT
QSDFPSMCWLSAGRANKNLQFSGATRILGYYNEATDKFVNQLNEACGGALVNRYRAVY
DGFRRVATQSRQHPGEVIMDMRWNTKNSLVLLEQPVNAAECTIKIDLGWKDERFYIDESI
FEQLFFVLNLAEVLANPEEEVKFPKKWEKFDELVKLMNELVQKCSDEENVFASKDTSEE
VTDKVWNIVRRCSDIKEATMLLKNFLQALTYGKIKSHVLENNKSHLAALIRASKTCEFR
MPILERLSTIEMMMEIGVQRLRGRIVDMFSKTLKFPSDELTLMLKSCENDLKSVECSMNA
SVVSLLPITMALATVHQIFGILNEKDHVILPDLARRVLMKYTSSMIEKSKRGETELNYSFE
TTLPLLRINKDMFVDKRPRIWTCENSKHVGANVQARMMTSLELEMPLEHVSELVNSIRP
VRETKEDTAPTTEELNAEYTVSHTVFTYYSKI
>C_yunquensis_ZWL1

MLVGLEHLQQYNTAVKSKTEEDEEPTPVLLLGKFRLRLMPIADLPIVSNYSNIAQLGFSSD
DILLIDSPNGSEKKAKVQQVPFSIKQENGEAMEVGEDEDDKENAAMEDGPLETSFLTLE
KLQRSSGNEKDIEIAGFDFEVPMFDVNPIPLCDGISIQRYLRLESSSLFTPAVSKLPVWISTT
QSEFPSMCWLAAGRTNKNLQFSGATRILGYYNETTDKIVNQLNEACGGALVNRYRAVY
DGICRIATQSRANPGEVIVDMRWNTKNSLVLLEQPVNAAECTIKIDLGWKDKRFYIDESM
FEQLFFVLNLADVLANPEKEVVFPDKCEKFDDLVKLLNELVETCSDEDNVFVSKDSSEE
VTDKVWNIVRCCGDIKQATLLFKNFLQALTYGKIKSHVLEDNKSHLAALIRASKTSEFR
MPILERLSTIEMMMEIGVQSLRGRIVDKFSNTLKFPSDELTLILKSCENDLKANECAMNA
SVVSLLPITMALATVHQIFGILNEKDHVILPDLARRVLMKYTSNIIEKGKRGETELNYSFE
TTLPLLRIRKDEFMDRRPRIWTCENSKHVGANVHARMMTSLELETSLEHVSGLVNSIRPV
IDVEDNKKLTAEELNAKYTVSHTVFTYFSKF
>C_macrosperma_ZWL1
MSVSVEQLHEYNRALETKAEDDEAPTPVLLLNKFRLRLVPITDLPIVSNYSNIPQLGFSSD
DVILIDSPTGSEKKQKVSLEPQVVAETDDAMEMGENEEDKENATGGGPLETSFLTLAKLQ
RSSGKEVNVEIAGYDFEAPTFDVNPIPLGDGISIQRYLRLQSSDVFTAAVAKYPVWIATTEP
SLPSMCWLAAGRTGKNLRFSGATRILGYYNESTDKLVNQLNEACGGALVNRYRAVYDGI
RRIATPTRENPGEVMVDMRWNTKNSLVLLEQPVNAAECTIKIDLGWRDKRFYIDESIFEQ
LFFVLNLADVLANPEKEVVFPNEWEKFDDLVKLMNELVEMCSDEENVFVSKDTSEEVT
DKVWNIVRRCGDIKQATLLFKNFLQALTYGKIKSHVLEDNKSHLAALIRASKTCEFRMPI
LERLSTIEMMMEIGVQSLRGRIVEKFSKTLKFPGDELTLMLKTCENDLKSGECAMNASV
VSLLPITMALATVHQIFGILNEKDHVILPDLARRVLMKYTSSMIEKSKRGETELNYSFETT
LPLLRINKDKFMDKRPRIWTCENSKHVGANVQARMMTSLELETSLDHVCSLVNSIRPIRE
TAEDESPSAAELNAEYTVSHTVFTYFPK
>C_sulstoni_ZWL1
MAICLEDLKSYQNAIRTKEENDEAPPPVLFQYRLRLIPISELPIVSNYSNFTQLGLNSDEVL
VIDSSIESEKKETFSKVKKSAKIQAEQSDSMEVGGDEEEEADKENDTSATGPLVTSFLTLD
KLQKVDMEIGDFEFEMPSFDANPIPPEDGISLLRLRLSSPNHFSDAVAKLPIWISTSSDSFPS
TCWLSSGRTQKNVQFSGVTHVLGFFNEHSEKLVNQLNETCGGAQVNRYRAIYDGIRRIV
TENQKTPGEVIVDMRWNTKNSLVLLEQPVNAAECTIKIDLGWKDKRFFVDETIFEQLFFV
LNLADVLANPEKEVVFPSEWEKFDDLVKQMNELIEACSNEDNVFASNDSSEEVTDKVW
NIVRRCGDVKQATLLFKNFMQALTYGKIKSHVLENNKSYLAALIRSSKTCDFRMPILERL
STIEMMMEIGVESLRSRIIHEFETILKFPADELSLILKTCETDLINGEGSMNASVISLLPIAM
SLATVNQIFGLLNEKDHVILPDLARRVLTKYTNSMVEKSKRGETESHYTFETTLPLLRINK
EKFMNKRPRVWTCENSKTVGANVQCRMMTVLELESSLEHVTRVVNAIRPLREENSLDI
NSDYTVSHTVFSYLPKM
>C_afra_ZWL1
MSICLENLKKYNDAVKAKGEDDEAPAPVLLMSQYRLRVIPITDLPIISNYSNIAQLGLNSD
EVLVIDSPVEPEKKEKLLRVRNEERIETGSTESMEIEEEEEDKENTSATGPLVTSFLTLDKL
QKHSERMKWKSPDLTLKRPVSMRIRSPRKTEFLFSDNHLSDSLAKLPIWISTSSASFPATC
WLSSGRTQKSVQYSAVTRVLGFFDEHSEKLVNQLNEACGGAQVNRYRAIYDGIRRIATE
QRETPGEVVVDMRWNTKNSLVLLEQPVNAAECTIKIDLGWKDKRFFIDESIFEQLFFVLN
LADVLANPEKEVVFSCEYEKFDDLVKQMNELIEASSNEENVFTSNDTSEEVTDKVWNIV
RRCGDVKQATLLFKNFMQALTYGKIKSHVLENNKSYLAALIRSSKTCDFRMPILERLSTI
QMMMEIGVESLRSRIINKFESTLKFPADELSLILKTCETDLINGEGSMNASVISLLPIAMSL
ATVNQIFGLLNDKDHVILPDLARRVLTKYTGSMVEKSKRGETETNYTFETTLPLLRISKE

KFMNKRPRIWSCENSKTVGANVQCRMMTVLELESSLEHVARAVNAIRPMRDGQDEETS
HETNADYTVSHTVFSYLPKM
>C_sp49_ZWL1
MSITLEQLKAFNDAVDSKKEDDAAPTPVLLLNKYRVRLLPLSDLPIITHHSNLSQLALNSE
SVLVIDSPVNSEKPECRKLPTKTQNVCSDEKMDIEIIEKDENDKENVGGPLETTFLTLDKL
QKGVAVVGDVEIEGFDIAAATFDVNPIGYSEAGVLHRLIRHNPTTHFSSPIHQYPIWISTFT
DSLPHICWMSAGITQKNRQVTGITRVLGYLDENSEKLVNQLNASCGAGTVSRYRAVYDE
IRKIRTAKLPCPGEVVLEMRWNSNGSVVLLDKPVNAAECTIRIDLGLQDSRFFADKSLYS
ELFFVLNLANVLAHPENEVLFPSCTEKFDVLVKEINSLVEKCSEDDKVLISGDTEEVTHH
VWQIVRRCSDIKQATMLLKNFMQALTYKKIKSHVQESNKSHLASLIRSSKTGDFRMPILE
RLSTIEMMMEIGAESLRGRIIRNFSEKVAFPSEELKLILEACENDLKNSDVAVNAHAISLLP
VAMGLATVSHIHSLLNEKDQVILPDVTRRVLQKYTKGMVEAMKRGENENEYIYETTLPV
LRIKKEAFMAKRPRVWSCENMRTVGANVLSRMMTALELECPLSRVVDAVNSIRPHREEG
DQNQEKALDLNSDYTVSHTVFSFMPKM
>C_sp25_ZWL1
MPITLEELNTYNEATEAKKEEDEPPVPVLLLALNSESVIVIDSPIRLEKNEARRVAPAPQKT
VSDEAMDTEVTENDKENVGGPLETSFLTLDKLQKSGMIVGDIDIKGFDFAAPSFDANPIG
YSEAGVLHTVLRHDPTTYFSAPIDQYPIWISTFTESFSHICWMAAGRTQKNRQFTSITRVL
GFLDEHTEKLANQLNVACGGGTVSRYRAVYDEIRKIRTAKRPYPGEVVLEMRWNSNSSM
VLLEKPVNAADCTIRVDLGWQDSRFFVDESLYTELFFVLSLANVLANPENEVQFQSKSE
KFDELVKEINLLVAECADDNKVLVSNETEEVTHHVWQIVRRCGDIKQATLLLKNFMQAL
TYKKIKSHVQENNKSHLASLIRSSKTGDFRMPILERLSTIEMMMEIGVESLRGRIIKNYAK
TVAFPSDELDLVFKNCESDLKNCGVTVNSHAISLLPVAMGLATVSQIHSLLNEKDHVILPD
LTRRVLQKYTKEMVDATKRGEEDNEYVFETTLPVLRIKKEAFMTKRARVWSCEKVRTV
GANVLSRMMTLLELECSLPRVVDAVNKIRPYEQDEDPNQEKISDSNSDYTVSHTAFSFMP
VM
>C_imperialis_ZWL1
MPITLEQLKTYNEAVKTKGEEDEAPAPVLLLNKYRVRLLPLAELPIIENHSNLAQLALNSE
EVLVIDSPSAPAALEKQPVKIVTKPVTSDEQMDVIEAGDDVEDKENVGGPLETTFLTLEK
LQKGKKIVADVEIEGFDYALPTFDANPIAYTEADALHRILRLQPTAHFCSTIAHFPIWTSTF
SSSFPHICWKSSNLTQKGQQVTGVTRVVGFYDEHSEKLVNQLNAACGGSAVSRYRAVYD
QIRKISTAKRPFPGEIVLDMRWNTNGSVVLLEKPVNAAECTIRIDLGWQDRRFFADESLY
TELFFVLNLANVLANPELDVHFPSSSDKFDALVKEMNALVENCSEDDKVLASSETEEVT
NNVWHIVRRCGDIKQATLLLKNFMQALTYKKIKSHVQENNKSHLAALIRSSKTGDFRMP
ILERLSTVEMMMEIGVESLRGRIVKNFHESVRFASEELELVLKTCETELKNCNASINAHAV
SLLPVAMGLATVSQIHALLNVKDHVILPDLTRRVLQKYTKSMVEAAKRGESESGYVFET
TLPVLRIKKESFMTKRPRVWTCENVKMGGASVLSRMMTVLELESSLPRVTDAVNSIRPY
REEMEESPENMKKQLDVNSDYTVSHTTFSYLPKIQ
>C_kamaaina_ZWL1_Exon1NotFound
LNKYRLRLIPIDSLPIISNHSNINHLGLNSEDALVIDLPVGLSKKEKKLAVSESFSPKENQS
MKMEGVEENKENTSGNGGPLETSFLTLEKLQKGEVEIVGFDSEVEFFATNPIPFSDAVSLQ
RYLRMASKEHFSPLVDKLPVWISTIASDFPSVCWLSSGRSSKDGRFSGITRVIGNLNEQTE
KLVKQLNEACGTAQSNRYRCIYDGIRRISNSTREAPGEVIIDMHWNKKSSLVLLEQPDNA
ADCTIKIDLGWRDNRFYIDEAVFDQLFFVLNLADVLANPEKEVIFPTEFVKFDDLVKEMN
ELVEACSHEDKVFASNEKFRDDEVTDKVWNIVRKCGDVKHATMLLKNFLQALTYGKIK
SHVQEGNKSHLASLIRASKTSDFRMPILERLSTIEMMMEIGVGYLRRRVINKFSSTLHFPS

DELKFILNKCENDLSKDEGSLNPSVVSLLPITMALATVNQIFGLLNEKDHVVLPELTRRVL
AKYTSSMVEKAKKGETETDYTFEVTLPLLRISKERLMDKRPRIWTCENTNTVGGNAQTR
VMIALELESSLEHVNTLINASRPMREVKENDEKPNKEEMNADYTVSHTVFSYLPKL
>C_tropicalis_zwl1_cDNAWithFrameshift
ATGTCTCTGAAAATTGAAGACTTGCAAACGTATAATTTAGCTGTCAATAAAGCGAAAG
AAGGAGAAGAGAAACCAGCTCCTGTCCTATTGCTCAACAAGTATCGCATTCGGCTGA
TCCCAATCTCATCACTACCATTTGTGTCAAACTACTCGAATATTTCTCAGCTCGGGTTC
AACTCAGATGATGTTATCGTCATTGACTCACCAAACGAATTAGCTGAAAAACAGAAG
ACATCTAACCTTTTCTGCAATACCTCGAAGAAAAAATCTACTTCTGTGGAGCAAGAAA
CTGAGCAGATGGAAACCGATGAACCTGAAGAAGACAAAGAGAATATTGATGCCCATG
ACGGGCCATTGGCAACATCTTTTCTTACTTTAGACAAATTACAAAAGGACATGGAGAT
TGCTGGGTTGGACTGTGAGGTCCATTTTTTCGATGTGAATCCACTCCTATTTACAGAC
GGTGTTTCTCTCCAGAGACTTCGAATTGAAAGCTCTAACTATTTTTCACCACCGGTCG
ACAAACTTCCTGTTTGGATTTCCACTATCAGTCCGAACTTTCCATTAGTTTGCTGGTTA
GCTTGTGGAAAAACCAGCAAAAATGCACAAGTTTCGGGGACAACTCGAATTCTCGG
ATTTTTCGATGAAAGTAGTCAGAAAACTGTGAAACAAATACAGGAAGCATGTGGTGC
TGCTCAAGTAAACCGGTACCGTTCTGTTTACGATGTTATTCGAAAAATTCCCACAACA
ACTCGCCCTGCACCCGGAGAAGTTATAGTAGACATGCGTTGGAACACAAAAAGTAGT
CTCGTTCTTTTGGAACAACCAGATAATGCTGCCGATTGTACCATCAGAATTGACCTTG
GTTGGAAGGACAACCGTTTTTTCATTGATGATGCTATCTTTGAACAGCTTTTTTTGTGC
TTAATCTGGCCGAAGTTCTGGCTATTCCCGATAATGAAGTTACCTTCCCTTCGGAGTGG
ACAAAGTTCGAAGATCTTGTTCAAGAAATGAATGAACTCGTTGACGCTTGCTCCAAG
GAAGACAACGTGTTTGCCTCAAATGAAAAGTTCAGAAGTGAACAAGTCACCGACAA
GGTATGGAATATCGTTCGCAAGTGCGCGGATGTTAAACACGCGACAATGATTCTCAAA
AACTTCTTACAAGCGTTGACTTATGGAAAGATCAAATCGCATGTACAGGAAGGAAAC
AAAAGCCACCTGGCCTCATTGATTCGTGCTTCGAAAACTGGTGATTTTAGAATGCCTA
TCCTCGAACGATTGAGTACTATAGAAATGATGATGGAAATCGGTGTAGAAAGTCTGAG
ACGTATCGTCAACAAATTTACGAGCACTCTCCAGTTTCCACGTGATGAACTTATTTAC
ATCCTGGAAACTTGTCAAAACGATTTGTCGTCTACTGGTGAAGAAGCTTTGAATACAA
GCGCTGTATCTCTTCTTCCAATTACCATGGCATTGGCAACTGTCACTCAAATTTTCCGA
CTTCTCAATGAAAAGGATCAAGTTATTCTACCAGATTTGGCAAGACGCGTTCTTACTA
AATTCACAAGCGCCATGGTCGAAAAAGCAAAGAAAGGTGAAACTGAAACCGACTAC
ATATTCGAGACAACTCTTCCTCTTCTTCGTATGAGTAAAGATACATTTATGTACAAAAG
ACCACGAATTTGGACTTGTGAAAATACAAATACTGTTGGAGCCAACGTACAAACTCG
TATGATGACAACACTCGAGTTGGAGCCGTCGGTTGAACATATCAGTCGTATGGTTAAT
GCGAATCGACCACTTCGTGAGCTGAAAGAAGAGAATGTGAAGCCAACTGAAGAAGA
ACTCAGAGCGGACTACACAGTATCGCATACAGTTTTCTCCTATTTACCAAAGCCGTGA

Supplementary Data 3.S9: All SPDL-1 sequences used in this study

>C_tribulationis_SPDL1
MPDDEEKLQLRADVERFKKAIRQKDEMIEEMEHELNNIGKTPQSDGRAEARERELAGTI
RDLQFEMDGKDATIHGQTDIIASMREEIDKLEKKNRELINRPDCSEIDESNSFVESEMLRIS
EECEKFKELANTLGEENLDLKRAALELKEEYESACEHVKCLESHSKTKEEEIMRLEGEVF
DLKNSTAGKFSNTGNSIFAEAMEAEKKLEEDLKVLYREKQSLMGMVKRLNLEKEDAEQ
RARSNMNRGLVVRNAINHIEVQELNRLNTRLRQLETERFQFWEKMFIKMKSVPKRELGS
MFQGYFESFKCSITNMQSGYDELMKKNEQHITTIRGLQQEIETLRVKNEQLTFDVELLER
KVRSAGNCELDNPQLRAPLKPMNNARPSFFVKPKKADPAPPLETSLSNMMMTPQKPSPQ
PMCRSTAKKEEASEWAERKLKAKAEKKSATPAPRYNFVKMSAPVSSTKFKPAILQMPST
PSAIPELEN
>C_sp41_SPDL1
MPDDEEKLQLRADVERFKRAIRQKDEMIEEMEHELNNLGKVPQSDGRAEAREKELAGQ
IRDLQFEMDGKDATIHNQTDLIASMRLEIDRLEKNLELINRPDCSDVDESNSFVESEMIRIS
EECEKFKELVNTLGEENLELKKSAIELKEEYENACDHVKSLESHSKTKEEEIMRLEGEVF
DLKNSATGKFSSTGNSIFEEAMEAERKLEEDLKVLFKEKQSLMGMVKRLNIEKEEAEER
ARSNMNRGLVVRNAINHIDVQELNRLNTRVRQLETERFQFWEKMFIKIKSIPKKELGSMF
LGYFESFKCSIANMKGGFEELMKKNEQQIITIRGLQQEIETLRSKNEQLTFDFEILERKVRS
AENRELENPQLRAPLKPMNNARPSFFVKPKKADPAPSLEASMSNMMMTPQKPSPQPTCR
STAKKEETSEWAERKLKAKAEKKAATPAPRYNFVKLSAPISNTKFKPAILQMPSTPSAPSE
LEN
>C_zanzibari_SPDL1
MPDDEEKLQLRADVERLKRAVRQKDEMIEEMEHELNNYGKIPESDGRAEAREKELAGQ
IRDLQFEMDGKDAAIHDQTDIITSLRVEIDKLEKTNRELVNRTDCSEGDDSSSFVESEMLR
ISEECEKFKELANMLGEENLELKREAVELKENYESACDHVKSLESHSKTKEEEIMRLEGE
VFDLKNSTTGKFASTGNSIFAEAMDAERKLEEDLKVLYREKQALMGMVKRLTIEKEEIEE
RARAQMNRGLVVRNAINHIDVQELNRLNTRIRKLETERFQFWEKMFIKIRSVSKKELGA
MFLGYFESFKCSITNMKDGFDELMKKNEQHITTIRGQQQEIETLRMKNEQLTFDVECLER
KVRSAGNCELENPQLREPLKPMNNARPSLFVKPKKTDSVPPLETSMASMMMTPQKPSP
QPTCRSTAKKEETSEWAERKMKSKAEKKSATPAPRYNYVKMSVPVSNNRFKPATLTMPS
TPSTHPELEN
>C_sinica_SPDL1
MPDDDEKLQLRADVERYKRAIQQKDEMIEEMEHELSSYGKAPESNGKAEAREKELAVK
IRDLQYEMDGKDATIHDQTDLISSLRVEIDKLEQNRELINRHDITESDESNSFVESEMLRIS
DECEKFKELANMLGDENLELKKAAVELKEEYESACDHVKILESHSKTKEEEIMRLEGEIF
DLKNSNTGKFSNSGNSIFAEAMEAEQKLEEDLKVLFREKQALMGMVKRLTIEKEEAEER
ARSHMNRGLVVRNAINHIDVQELNRLNTRVRKLETERFEFWQKMFIKTKSIPKKELGSMI
VGYFESFKCSIASMTGGYEELMKKNEQHITTIRGLQQENETLRVKVEQLTFDVEILERKV
RSAGNCELDNPQLREPLKPMNNARPSFFVKPKKADPAPPLETSMSNMMMTPQKPSPQL
ACRSTAKKEETSEFADRWMKAKSEKKAAAPSRYNFVKMKVPVPANKFKPPVLQMPSTP
SADPELEN
>C_nigoni_SPDL1
MPDDEEKLQLRADVERFKRAIRQKDEMIEEMENELNSYGKQPVSNGKAEAREKELAGQ
MRDLQFEMDGKDATIHEQTELISSLKVEIDKLEKTNRELMNRSVCVDSDESNLLDESQM

LRISEECEKLKELASALGEENLELKKDAVTLREDYESAVEHVKSLESHSRTKEEEITRLEG
ELFDLRNSCTGKFANTGNSIFAEAMDAERKLEEDLKTLYREKQSLMGMVKRLTMEKEE
AEERARSHMNRGLVVRNAINHIDVQELNRLNKRVRQLETEKSHFWEKIFIKMRSIPKKEL
GALIVGHFGAFKCSIENMSEGYDELMKKNEQHVTTIRGLQQEIDTQRVKIEQLTFDVECL
ERKLRSSAADSEPEDPQLRAPLQPITNAARPSFFVKPKKTNPEPSLEMSMSNMMMTPQK
PSAQITCRSTVKKEDDELSEWAERRLKAKAEKKSATPAAKYNFVKLTAPAPSNKFKPAVL
QMPSIQSENTDEHEQ
>C_briggsae_SPDL1
MPDDEEKLQLRAdverfkrairqkdemieemenelnsygkqpvsngkaearekelagqmrdlqfemdgkdatiheqteli
sslkveidklekTnrelinrsvcvdsdesnlfdesqilriseeceklkelasalgeenlelkkdavtlrenyesavehvksleshsrtkeeeit
rlegelfdlrnsstgkfatagnsifaevmdaerqleedlktlyrekqslmsmvkrltmekeeaeerarshmnrglvvrnainhidvqeln
rlnkrvrqletekshfwekifikmrsiskkelgalivghfgafkcsienmsegydelmkknehhvttirglqqevdthqvkieqlkfdv
eclerklrssaansepenpqlraplqpitnatrpsffvkpkkadpepsleismssmmmtpqkpsaqitfrstvkkeddelsewaerrm
kakaekksatpaakynfvkltapapsnkfkpavlqmpsiqsenteehdnkcqifsiyefl
>C_remanei_SPDL1
MPDDDEKLQLRADVERYRKAIRQKDDMIEEMENELNHYGKPAISDGKAEAREKELNGR
IRDLQFEMDEKDAAIHDQTDLISSLRSEIEKLEKTNRELINRSECNESDESNSFVESEMIRIS
EECEKFKEVASTLYEQNRELKKEAVELREEHDSAMGHVKNLESHIKTNEEEIARLEGEVF
DLKNSNQGKHASTGNSIFAEVMEAEQKLEEDLKVLFREKQSLMSMVKRLTMEKDEAEE
RARSFMNRGLIVRNAINHIDVEEMRRLSARNRELETERTHFWERMFIKMKTVPKKEIGAI
IVGYFESFKCSIASIKGGFDDLMKKNEQNLTIIRGQNQDLENQRVKIEQLQFDMECLERKL
RSAVKAESEDSQLRAPLKPMNNSRPSFFVKPKKTDPVPPLETSMSNMMMTPQKPSPVITA
RSTAKKEDSSEWAERRLKAKAEKKSATPTPRYNYVTMSAPVPKFKAAVLQMPSTLPESQ
EN
>C_latens_SPDL1
MPDDEEKLQLRADVERYRKAIRQKDDMIEEMENELNHYGKPAISDGKAEAREKELNGR
IRDLQFEMDEKDAAIHDQTDLISSLRSEIEKLEKTNRDLINRSDCNESDESNSFVESEMIRI
SEECEKFKEVASTLYEQNLELKKEAVELREEHDSAMGHVKNLESHIKTNEEEIARLEGEV
FDLKNSNQGKHASTGNSIFAEVMEAEQKLEEDLKVLFREKQGLMSMVKRLTMEKDEAE
ERARSFMNRGLIVRNAINHIDVEEMRRLSARNRELETERTHFWERMFIKMKTVPKREIGP
IIVGYFESFKCSIASIKGGFDELMKKNEQNLTIIRGQNQDLENQRVKIEQMQFDMECLERK
LRSAAKAEIEDSQLRAPLKPMNNSRPSFFVKPKKTDSVPPLETSMSNMMMTPQKPSPAIT
ARSTAKKEDSSEWAERRLKAKAEKKSATPTPRYNYVTMSAPVPKFKAAVLQMPSTLPES
QEN
>C_sp51_SPDL1
MLDDEEKLQLKADVERFKKVIRQKDEMIEEMENELNSGKTTETDGKAEAREKGLNEQI
RNMQFEIDGKDARIHEQTGLISSLRHEIDELEKSNREFINRSEFNESDESNSSVETEMLRIS
EECEKFKEVAATLYEQNSDLRKEVVELKEDHETAIEHVKSLEIHIKTKEEEIMRLEGELFD
LKNSNHGKHASSGNSIFAEAMEAERKLEADLKVLFREKQSLLVTVKRLTVEKEDAEERA
RAFMNRGLVVRNAINHIDVEEMRRLRARVQHLESERIHLWERLFIKIKSVPKREIGSIMA
GYFDSFKCSIASVTGGFEELMKKNEKYVTTIRGLQQENENQRVKIEQLQFDMECAERKM
RSALNGDSENPQLRAPLKPLDNSRPSFFVKPKKSETVAQLETSMSSMLMTPQKPSVAMT
ASSTAKEEDASDWTERRLKAKAEKKSATPAVRYNFITLKAPAPSSKFKAAVLPMPSTPTEP
KEN
>C_sp44_SPDL1_paralog1

MLDEEKLQLRADVERYKKAIRQKDEMIEEMENELNSHGKGATDEKAEAKEKELNGQIR
DLQFEIDGKEATIHEQTDLISELRNEIDKLEKSNRELINRSDFIIESDESNSFGETEMLRLSE
ECEKFKEESTNLYEENLKLKKEVVELKEDHNSAMEHIRSLESHIKTKEEEIAMLEGEVFD
LKNSNHGKHASSGNSIYLEAMEAERKLEEDLKVIFREKQSLMAMVKRLTLEKEDAEER
ARAFVSRGLVVRNAINHIDVAEMQRLRTRVNELETERIHLWERFFIKMKTVPKREIGSMM
AGYFESFKCSIASVTGGFEDLMKRNEQYVTTIRGLQQETENQRVKIEQLQFDIECAERKM
RSSSNADSEHPQLRAPLKAVDNSRPSFFVKPKKVEPVPQLETSMSNMLMTPQRSSEPIAE
TSTAKKEDMSDWAERRLKAKAEKKSATPAARYNYIKLTAPASSSKFKAAVLPMPSSSSDS
KEN
>C_sp44_SPDL1_paralog2
MNSSSISEKPGRAKELTDAIITEIQSYEGKLKQTPFMKQVCAKLLALEMPETVGEQWEKK
LANRTKNEIQWIRDTNLIPQMLKDEQISNDNKLKLGTIFGTAVIQSETPEDGHGPTDDSFH
EQTESDESNSFGETEMLRLREECEKFKEESAKLYEENSKLKKEVVELKEDHNSAMEHIRS
LESHIKTKGEEIAMLEGEVFDLKNSNHGKDASSGNSIYLEAMEAERKLEEDLKVIFREKQ
SLMAMVKRLTLEKEDAEERARAFVSRGVVVRTAINHIDVAEMQRLRTRVNELETERIHL
WERLFIKMKTVPKREIGSIMAAYCESFKCSTASVTGGFEDLMKRNEQYVTTIRGLQQETE
NQRVKIEQLQFDIECAERKMRSSSKADSEHPQLPAPLKAVDNSRPSFFVKPKKVEPVPQL
ETSISNMLVTPQRSSEPIAETSTAKKEDMSDWAERRMKAEAASSSEFKADVSPMPSSSSG
SKEH
>C_sp48_SPDL1
MSTKMLDDDEKLQLRADVERFKKAIRQKDEMIEEMENELNNHGKASESDVKAEAREK
ELNVKIRDLQFEMDGKDVTIHEQTDLISSLRNEIDKLEKSNRELANRSDTNESDESNSFVE
TEMLRISEECEKFKEVAATLYEQNLELKKEAIELKEEHDAAVEHVKSLESHIKTKEDEIAR
LEGEVFDLKNSNHGKHASSGNSIFAEAMEAERKLEEDLKTIFREKQALMAMVKRLTVEK
EDAEERARAFMTRGLVVRNAINHIDVEEMRRLRARVQQLETERTHFWERLFIKMKTVPK
REIGSIMAGYFESFKCSIASVTGGFEEIMKKNEQYVTTIRGLQQENENQRVRIEQLQFDIE
CAERKLRSAVNSDSEHPQLRAPLKPMDNSRPSFFVKPKKAEPVTQLDISMSNMLMTPQK
PAAMITSNSTAKKEDASDWAERRLKAKAEKKSATPAARYNYITLKAPSAATKFKAAVLP
MPSSDSDLKEN
>C_brenneri_SPDL1
MLDDDEKLQLRADVERLKKAIRQKDEMIEEMENELNNHGKASESDGKAEARERELNG
KIRDLQFEIDGKDVTIHEQTDLISSLRNEIDKLEKSNRELANRSDTNESDESSSFVETEMLR
ISEECEKFKEVAGTLYEQNLELKKEAIELKEEHDAAVDHVKSLESHIKTKEDEIARLEGEV
FDLKNSNHGKHASSGNSIFAEVQAMEAERKLEEDLKKIFREKQGLMAMVKRLTLEKED
AEERARAFMARGLVVRNAINHIDVEEMRRLRARVQQLETERTHFWERLFIKMKTVPKRE
IGSIMAGYFESFKCSIASVTGGFEEIMKKNEQYVTTIRGLQQEIENQRVKIEQLQFDIECAE
RKMRSVLSSDSENPQLRAPLKPMDNSRPSFFVKPKKAEPVTQLDISMSNMLMTPQKPAA
MITSNSTAKKEDASDWAERRLKAKAEKKSATPAARYNYITLKAPSAATKFKAAVLPMPS
SDSDLKEN
>C_wallacei_SPDL1
MPEDEEKLQLRADVERFKRAIRQKDEMIEEMENELNCRGKSPVSNSKAEAREKELNGQI
RDLQFEMDGKETTIHEQTDLINSLRDEIDKLEKTNRELVNRSDCTESDESNSFVETEMLRI
SEECEKFKEVAASLYEQNVELKKEAVELREEHDAAVEHLKSLESHIKTKEEEITRLEGEVF
DLKNSSQGKHATSGNSIFAEAMEAERKLEQDLKVLFREKQSLMAMVKRLTLEKEDAED
RARNFIDRGRAVRQAMSHIDVEEMRRLRTRVQQLESERSHLWERLFIRMKNVPKRELGS
LVASYFESFRCSMASVTGGFEELMKKNEQYVTTIRGLQQEIENYRVKVEQLEFDNECSQR

KLRLSANIEQENPQIRAPLKPMNNSRPSFFVKPKKIDPVSTLETSMSSMLMTPQKPSVPTT
ASSTAKKEASDWAERREKSKAEKKSATPAQHFNFVKLAAPVPKFKAAVLQMPSSPTELN
EN
>C_tropicalis_SPDL1
MPDDEEKLQLRADVERFKKTIRQKDEMIEEMENELNNRGKSPVSNGKAEAKEKELSSQI
RDLQFEMDGKDATIHEQTDLINTLRDEIDKLEKTNRELINRSYCAESDESSSFVETELLRIS
EECEKFKEVASSLFEQNVELKKEAVELREEHDAAVDHVRSLESHIKTKEEEITRLEGEVFD
LRNSSQGKHASSGNSIFAEMEAERKLEEDLKLLFREKQSLMTMVKRLTLEKEDAEDRAR
NCIDRGRVVRQAMSHIDVEEMRRLRTRVHQLESERTHLWERLFIKMKSIPKRELGSLIAS
YFDSFKCSIASMKGGFEELLKKNEQYVTTIRGLQQELENQRGKVEQLEFDVECAERKLR
SAANVELENPQLRAPLKPMNTARPSFFVKPKKIDSTTSLETSMSSMLMTPQKPYNPIAAS
STAKTEASDWAERRQKAKAEKKSATPAQHFNYVKLAAPVPKFKAAVLQMPSSPTDSNP
Q
>C_doughertyi_SPDL1_paralog1
MPDDEEKLQLQADVERYKKAIRQKDEMIEEMENELNSHGKLPVSDSKAEEREKALNGQ
IRDLRFEIDGKDATINEQTDIINNLRDEMEKLEKKNRELVNRSDCTEIDESNSFVESEMIRI
SEECEKFKEMASSLFEQNMELKKEAVELREEYDSAIELMKNLESHIKTKEEEIARLEGEV
FDLKSSSQGKHASAGNSIFAEAMEAERKLEEDLKTLFREKQALMSMVKRLTLEKDEAEE
RARNFMNRGLIFRNSINHIDVEEMRRLRTRVQQLETERTNFWERLFIKMKTVPEREIGSII
AGYFESFKCSIANVTGGFEELMKKNEQYVTTIRGLQQEIENQRVKIEQMQFDIECSERKM
RSALSLESETPSLRAPLKPMDNSRPSFFVKPKKADPVPVPNLEVSLSNILLTPQKPSAEVT
TQSTTKKDDASEWTERRLKAKAEKKLATPALRYNYVQLKGPAPKFKPAVLPMPRTSPNS
KEN
>C_doughertyi_SPDL1_paralog2
MRDDEEKLQLQADVERYKKAIRQKDEMIEEMENELSSYGKLPMSDSKAEEREKALNGQ
IRDLRFEIDGKDATIKEQTDLINDLRKESNSLEESNSLESEMIRISEECEKFKEMASSLFEQ
NMELNKEAVELREDYDSAIELMRNLESHIETKEEEIARLEGEVFDLKSSSQRKHASAGNS
IFAEAMEAERKLEEDLKTIFREKQALLSMVKRLTLEKDEAEERARNFMNRGLIFRNSINHI
DVEEMRRLRTRVQQLETERTNFWERLFIKMKTVPEREIGSIIAGYFESFKCSIANVTGGFE
ELMKKNEQYVTTIRGLQQEIENQRVKIEQMQFDIECSERKMRSALSSESETPSLRAPLKP
MDNSRPSFFVKPKKADPVPTLTQKVAKEDDASEWTERRLKAKAEKKLATPALRYNYVQ
LKGPAPKLQPAVLPMPRASPNSKEN
>C_sp54_SPDL1
MPDDEEKLQLLADVERMKKTIRRKDDLIEEMENELNNFGKTSRSDGRAEAREKELNGQI
RDLQTEMDGKDTTIREQNDLIITLRGDLVKLENSNRELINRSECVDSDESNSFVETEMLRI
SEECEKFKELASTLYEQNLDLKKEALELREEHDTVVEHVRSLQSHIKTQEEEIARLEGDIF
DLKNSNQGKHAASGNSIFAEAMEAERKLEEDLKVLFREKQALMGMVKRLTMEKDEAE
ERARTFMNRGLIVRNAMNHIDVEEMRRLRARVYELETERTHFWERLFIKMRTIPKRDIG
QLFAGYLDSFKCSIASVKGGFEELLKKNEQYVLAIRGMQQEIENQRVKIEQLEFDIECSER
KMRSSANTDLDHPQLRAPLKPIDNSRPSFFIKPKKAYHVPQLEVSMSKMLMTPQKPSGE
TTAHSTVKKESTSEWAERRIKAKNEKKSSPAPKFKAAILQMPTALSESKEN
>C_inopinata_SPDL1
MVEDEEKLQLRADLERCRKMVRQKDEIIEEMEHDLNRRGKEYIFDKKLEAREKELNAQI
RQLQVEVDGKDATIQEQRDLVEDLRRDICTLEKSRELLNRTSSPKSDESTSFEKSEKIRISE
ECEAFKEVASTLYEQNAELKKNAVELKEELESAMEHIRSLQSHIKTLEEESSRLEGELFDL
KNSNNGKIASSGNSIFAEAMEAERKLEEDLKVLFREKQSLACRVKRLTMEKEEAEERAR

SFISRGLIIRNAINHIDVEEMRRLRTRAHELETERIHLWEKLFTKMKTTSKREMGAIIAGYF
ESFKCSIESVKGGFDDLMKKNENYVNMIRVQQKEILDLKEKIEQQKFDIECYERKMHSV
ANANLENAQLLEPLKAIDNARPSFFVKPKKVESINQLETSMSNTLITPQKEVLSEWAEKR
MKAKAEKNSATPGPRYNYITLTGPTPKFKSAILQMPSTPSEPKND
>C_elegans_SPDL1
MPDDEEKLQLLADVERLKKILRQKDEMLEEMEDDLKNQGKPCSSKLSLEERAQELSEQ
LRDLHVEMDGKNATILDRDALIDSLRSEIDKLEKINKEFANGSVIPEHDDSNSFGESEMLR
ISEDCQKYKETATALYERNAELEKEAVNLKDEIESMMDHIRDLKNHMETRDEEIARLEGE
LFDERNSHEGKLAARGNSMFSEVIDAERKVEEDLKVLHGENRALKGMVKRLRMEVEE
VEERLRSSTKRFNVTRMTTSDIDVKEMRRLRDRVCHLETERVHLWERMFIKMRSIPKRE
VGALITGYFKSFELSIASVKGGFDGLMKDNEKYVTIIRGLQQEVENLKADIVQLQFDNKC
AHRKAAPVVNKDFEHPLLAAPLKTLNNGRPSFFIKPKNVEPMPQLGHSLSSIAVTPQKPA
AKFTTRSSIKDDTSEWAERRMKAQAEKKLATPTPRYNYIKLSEPVPKFKPAVLQMPSTSE
TKEN
>C_oiwi_SPDL1
MPDDEEKLQLRADLERTKKMLHQKDDMIEELEQELNNKAQVSDWKSEAREKQLNGQI
RDLQMEMDGKDLTIQEQNGLIGELRVAVDRLEKRNRELANRSILPESDDSCSFDESGLIK
ASEECEKLRQNVVTLQEHNMELQKELIEKKEEHEGAVEHIRNLESHMKTQYEEIARLEG
EIFDMKNSAQGKHASCGNSIFAEAMEAERQLEEDLKKLFREKQALTSMVKRLTIEKEEA
EERARAYMNRGLIVRNAINHIDVEEMRRLRARVHELETERTHLWERLFIKMKSVPKREIG
AIIVGYFESFKCSIASITGGQSDLLKKNEQYVTTIRGLQQEVENQRVKIEQLQFDVECAER
KMRSAVTSNSEGENPQLRAPLKSVDNSRPSFFVKPKKKELAPPPSLELSMANMEMTPQK
PSLITNRSSTAKKEDTSEWAARQLKARADKKATPAPKYNYVSLKQPVPKFKTAVLQMPS
SPTESKEN
>C_kamaaina_SPDL1
MPEDEEKLQLRADVDRLKKMVHQKDEMIEEMEHELNNKAPVSDWKAEAREKQLNEQI
RDLQVEIDGKDVTIQEQQGLIGELRVEIDRLEKNRELANRSILPESDDSCSFDESGLIKASE
ECEKLRQNVVTLQEHNMELQKELIEKKEEHEGAVEHVRNLESHMKTQYEEIARLEGEVF
DMKNSANGKHASCGNSIFAEMEAERQLEEDLKKLFREKQALTSMVKRLTLEKEEAEER
ARAYMNRGLIVRNAINHIDVEEMRRLRARVHELETERTHLWERLFIKMKSVPKREIGAII
VGYFESFKCSIASITGGQSELLKRNEQYVTTIRGLQQEVENQRVKIEQLQFDVECAERKM
RSAVTSNSEGEIPQLRAPLKSIDNSRPSFFVKPKKKEPAPPPSLELSMANMEMTPQVTNRS
STAKKEETSEWAARHLKARADKKATPAPRYNFVELKQPVPKFKAAVLQMPSSPTERKEN
>C_waitukubuli_SPDL1
MPDDEEKLQLQADVERYKRMIRQKDEMIEDLEHELSNHGKTPVSDGKTEARERELNGQ
IRDLQLEIDGKDATIVQRDGVIDELREEIEKLEKSSRELANRSEYQDSDESNSFLEHEMVRI
SEELEKYKEATGNLFEQNSELKKNVLQLQEENENAMEHMKSLESHLKTREEQIASLESE
LFDLKNSSRGKHASSGNSIFAEVMEAERKLEGDLRNVFTQNQALSAAVRRLTLEKEDAE
ERARSAMNRGLVVRTAINHIDIEEMRRLRERVHELELERTHCWERLFIKMKTVPKRELGA
LVAGYFESFKCSIVSVKSGQAELLKKNEQYVTTIRGLQQENENQHVKIEQLQFDIESLTRK
LKSSVNDSEPAQLRAPLKPIDNSRPSFFVRPKPVEVPVPTASMSNMHVTPQKPANLPNTC
STAESKNENTSYWADRQKVKAEKKAATPAARYNFVTLTEPKPKFKPATLMMPST
>C_panamensis_SPDL1
MPVDEEKLQLRADVERYKRMIRQKDEMIEEFEHELNSHGKAPVSDGRAEAREKELNGQ
IRELQAEIDEKDATIVGKDGVIGELREEISKLEKQSRELANRSECQDSDESSSFLEQEITRIS
EELHKYKEATETLFEQNSELKKEAVQLKEEHESAMEHVRNLESHLKTREEEVARLESELF

DLKNSNQGKHASKGNSIFHEVMEAERKLEEDLRTVFAENQKLKALARRLNVEKEDAEE
RARSAMNRGFVVRTAINDVDVYEMNRLRERNRQLELERTSLWEKLFVKMRTIPKKELGF
LVAGYFESFKLSIISVKSGHEELLKKNEQYVTMIRGLQQQNENDSVKIEQLKFDIEGLNRK
LINATRDNDPARVPLKPVDNSRPCFAKKKPVETAAPVVSMSTMQITPQENTSYWEQRQKI
KAARAAKTPVAPVSRYNFVQLIEPKPLFKPTTLVVPSTPKEE
>C_nouraguensis_SPDL1
MPDDEEKLQLRADVERYKKMIRQKDEMIEEMEHELNSRGKTPTSDGRAGAREKELGVQ
IRELQLEIDGKDATIEMKDNVIEELRGEIDKLEKSRDLANRSEYPESDESSSFLEHEMVRIS
DELEKYKEATANLFEQNTEFKKDNLQLKDEHENAMEHVKSLESHLKTREEEVVRLESEL
FDLKNSSNGKHASTGNSIFAEAMEAERKLEEDLRKLFSEKQALASMVRRLDMEKNEAE
ERARTAMNRIVVRTAINNVDIEEMRRLRERVRELELERTHLWERLFIRMKSIPKRELGSLI
TGYFESFKCSIVSMKSGQDEVMKKNEKYVTTIRGLQQENENLRVKVEQLEFDIDSLNRK
LKSAATASSKFFSRILTNLYVHPDFESEPAQLRAPLKPFDNSRPSFFVKPKAAEKPTAMPN
VSNMYVTPQSTNEALKENTSDWAERRQKTKAEKKQAATPATRYNFIQLTAPQPKFKPATL
MMPSSDKED
>C_becei_SPDL1
MPEDEEKLQLRADVERYKKMIRQKDEMIEEMEHELSSRGKTPTSDGRAEAREKELGGQI
RELQLEIDGRDATIETKDGVIGELRDEIERLEKQSRDLANRSEYPESDESSSFLEHEMVRIS
EELEKFKEATANLFEQNAELKKVELQLKEDHENAMEHVKSLESHLKTREEEIVRLESELF
DLKNSSNGKHASTGNSIFAEVMEAERKLEGDLRTLFAEKQALTAMVRRLDMEKNEAEE
RARAAMNRIVVRTAINNVDVEEMRRLRERVRELELERTHLWERLFIKMKSVPKRELGSLI
VGYFESFKCSIVSMKSGQDELMKKNEKYVSTIRGLQQENENLRGKIEQLEFDIDSLNRKL
KSSANATSECSSCVLNIFFVHLDIETEPAQLRVPLKPIDNSRPSFFVRPKDVETPAAMPNMS
SMYITPQQKPAKMPSASNTVEAPNEDTSEWAERRQKAKAEKKQAATPATRYNFIQLTAP
KPKFKPATLMMPSTDKEN
>C_yunquensis_SPDL1
MPNDEETLQLRADVERYKKMIRQKDEMIEEMEHELNNHGKATVSDGKADVREKELGA
QIRDLQLEIDGKDATIVMKDGVIEELRDEIKKLEKQSQELANRSDYPESDESGSFLENEM
ARISDELEKYKEATANLFEQNSVLKKEGLQLKEEYENAMDHVKSLESHLETRNGEIARL
ESELFDLKSSNQGKHASSGNSIFAEAMEAERKLEGDLRKVFHEKLALTNMVRRLNMEK
DDAEERARNAMSRGRVVRTAINEIDIAEMRRLQERVRELELERTQLWEKLFIKMKSMPK
RELPALITGYFESFRCSIVSMKSGQDAVMKNNEKYVTTIRGLQQQCEMDRAKIEQLEFDV
ESLTRKLKSAAKTSDDQSGAALSQDSLKPTVNSRPSLFAKPKPVESSASVPAQLRAPLEPI
NNSRPSFYVRPKPMESSAAKPVLGNIHSTPQKSVDSASTCSTVERPNESTSEWAERREKA
KAERKQAATPAARYNFIKLSAPEPKFKPAKLMMPSSDQEN
>C_macrosperma_SPDL1
MPDDEEKLQLRADVDRYKRMIRQKDEMIEEMEHELSNHGKPATSDGRAEAREKELSGRI
RDLQIEMDGKDTTIAIKDGLIDELREEIMKLEKNSRDLANRSEYQDSDESSSFLEHEMVRI
SEELEKYKEATANLFEQNSELKKEGLQLKEEHEAAVEHVKSLESHLRTREEEIVRLESELF
DLKNSSQGKHASSGNSIFAEAMEAERKLEEDLRKVFREKQALTVMVRRLTLEKDDAEER
ARSAMSRGLVVRTAINHIDVEEMRRLRERVRELELERTHLWERLFIRMKTVPKRELGALI
SGYFESFKCSIVSIQSGQDEVMKKNEQYVTTIRGLQQENENQRVKIEQLQFDIECLNRKL
KSAVIASNADSEPAQLRAPLKLIDNSRPSFFLKPKPVESTAAPIFEMSSMHVTPQKPANIPS
ACSTVESKKENTSEWAERREKVKAEKKQAATPATRYNFITLAAPQPKFKPPTLMMPSTPK
EE
>C_sulstoni_SPDL1

MEDEEKLQLDVERLKRQIRQKDDMIEEMEQELTSGRHAHGPVSDGKAEAREKELNAQI
RDLQFEMDGKTTTIGEKEELIEKLRDEIDRLEKNRELANRSVVSESDESSSFHEVEMVKL
GEQLEECRHHIASLEEQNLQLRKEALQLKEEHEMVMDHVKNLESHIKTGEEEIARLEGE
VFDLRHSNQGKHATTGNSIFAEAMEAEQKLEQDLKTCYAENQALKKKIRRVILEKEDAE
ELARSAMGRGVVVRTAINHIDVLEMNRLRAKVHELEKERTLFWLHLFEKMKAAKITRR
ELGGLIAGYFESFKCSIASVTGGQEEILKKNEQFVTTIRGLQQENENLRVKTEQLQFDIGC
LNRKLRSAINLRQPLKPVDNSRPSFFHKRAPVVDPAVQSETSMSSSQVPEQLRQPLKPFD
NSRPSFFQRRAPVVESAVPLDSSMSSMHMTPQKPAFPQTTCSSQKENTSEWAERRQKAK
AERKQAATPAAYNYVTLLAPQPKFKLPTLKMPSSPTTPRVHEE
>C_afra_SPDL1
MEDEEKLQLRADVERLKRQIRQKDDMIEEMEHELTSGRHVPASDGRAEAREKELNTQIR
DLQIEVDGKNEAIEKREELIEELRDEIDKLEKRNRDLANRSVCPEAEDSSSFHEMEILKMG
EEVEECRRQMASLEEQNLELRKQALQLKEEHEMAVDHVRSLESHIKTGEEEIARLEGEV
FDLKNSSQGKHASAGNSIFAEAMEAEQQLEQDLKKCFAENQALKKLLRRVTDEKEDAE
ERARSAMSRGLIVRTGVNHIDVMEMNRLRAKVQELETERVHFWHHLFQKMKAAKLTR
RELGSLIVGYFESFKCSIASVKGGQEEILKKNEQYVTTIRGLQQEIENLNAKVETLQYDIE
CANRKLRRANNIGSKSGHSRSVLNSSSDSEADVPPQLRQPLKPVDNSRPSFFRRPAPAAEP
SAPLETSLSFMNVTPQKPAFLQNASSTQKENTSEWTERRQKAKAEKKAATPATRYNYVT
LSAPKPKFKQGTLQMPSTPASLVQEEAEHE
>C_sp49_SPDL1_paralog1
MVDDEEKLQLRADVERFKRMIRQKDEMIEEMEHELSRPKTPGDSLRTAARERDLATKIR
DLEIEIDAKNVTITQKDGQIDELRDEVDKLEKSNRDLAYRPESPSQDSSSSFLENELARLS
EELDTFRESTSSLMHQNLEYEKELLQLRSDLESATEHVKSLETHLKTREEELARLEGEMF
ELKASGHGKHASAGNSIFAEAMEHEKKLEEDLKFLFAQNQCLLKKVRRLQIEKEDAEQR
AQSAMRRKYTVGTAINHQGFAEMDRLRQKVRELEVERTHLWERLFVKLRGVNKREFGP
LIVGYHESFKLSITSVTGNQEKVLKENEELAKKVQGLETINADYQEQVEQLKYDLETAQ
RKAESAVDLQDSQYPLPKLTRPILTRKPKPEDDEEFTSGKSHPLLNAPLKPMSSAVRNSFF
VKPNRDPIDKGIMNMSIENSLRMMPGTPRAQDQENVSDFTLHHQRAKAERKAAATPAR
PNYNFVQLLQPTRTTSKFKPAILQMPKPTDE
>C_sp49_SPDL1_paralog2
MVHDEEKLQLRADVERYKRVIRQKDEMLEEMEYELSRPKTPADSMLTAARERDLVTKIR
DLEIEIDAKNVTITHKDGQIDELRDEVDKLEKSNRDLVYHSESPSQDSSSSFLENELARLS
EELDTFRETTSSLINQNLEYEKELLQLRSDLESATEHVKNLENHLQTREEELARLEGEIVE
LKALGHGKFATAGNSIFAEAMEHEKKLEEDLKFLFAQNQRLLKKVRFLQIEKEDAEQRV
QSAMRRKYTVTTAISYQGFAEMDRLRQKVRELEVERTHLWERLFVKMRCVSKREFGPV
VVAHYESFKLSIASVTGNQENVLKENEELSKQVHRLENINAEYEEQVKQLKYDLETAER
KAKSAVEEPKDLQDSQVPPPICTRPIFFSRKPKPEVAQNPEDEFTSGKSHPLLNAPLKPMS
SAVRNSFFVKPSRDPIDRGIKNMSVENSLQMRASTLRVQDQEKLSDFTLHHQRAKAERK
AAAPPAELKLNYIQLIQPTTTTSKLKPAILWMSKSTDK
>C_sp49_SPDL1_paralog3
MVDDKEKLQLRADVERFKRMIRQKDEMIEEMEHELSRPKTPGDSLRTAARERNLATKIR
DLEIEIDAKNVTITQKDGQIDELRDEVDKLEKSNRDLAYRSESPSQDSSSSFLENELARLS
EELDTFRETTSSLMHQNLEYEKELLQLRSDLESATEHVKSLETHLKTREEELARLEGEMF
ELKASGHGKHASAGNSIFAEAMEHEKKLEEDLKFLFAQNQCLLKKVRRLQIEKEDAEQR
LKSATRRKYTVSTAINHQGFAEMDRLRQKVRELEVERTHLWERFFVKLKRVSKREFVPLI
VGYHESFKLSIISVTGNQEKVLRENEELAKKVQGLETINADYEVQVAQLKYDLETAQRK

AESAVDLQDSQYPLPKLTRPILTRKPKPEDDEEFTSGKSHPLLNAPLKPMSSAVRNSFFVK
PNKDSIERRITNMSIQNSIRKAEMEAEAPPTTIKTNFVQLLQPTRTTSKFKPAILQMPKSTD
E
>C_sp25_SPDL1
MVEDEEKLQLRADVERFRRMIRQKDEMIEEMEHELSRPKTPGDKLRLEARERDLASKIR
DLELEIDAKNVTIHKKEGQIEELRDEVDKLEKSNRDLAYRPESPQQDTSSSFMENELARL
SEELDKCRETNSALMQQNLELEKENIQLKSDHESAMEHVKSLEVHLKTREEELGRLESE
NFQMKAAGHGSHASAGNSIFAEAMEAERKLEEDLKSLFAQNQCLIKKVRRLQIDKDDA
EQRVQSAMQRKYVVSSAINTQGYEEMVRLRQKVRELEVERTRLWEKLFVKLRRVNRKE
LGPLVLGYHESFKLSIASVTGNQENVLKENEELATKIQGLEKMNAEKEEKIEQLKYELAT
AQREKKVAAALESATEDFASEKDQNIPLKQSVLLNYSRKPKQPEIPENPDEEQFTSGRDH
PLLNAPLKPMNSGVRPSFFVKPKKEPVERTMMNMSIDPQSSFQTIPQTPTSGNDRENISDF
TFRYQKAKAERKAAATPAKLSFNYVQLIQPTTTSTSKFKPATLQMPKIDD
>C_imperialis_SPDL1
MVDDEEKLQLRADVERFKRMIRQKDEMIEEMEHELSRPKTPGDRVRMEARE REFAGKI
RDLQIELDARNVTIQKKDAVIDELRDEVDKLEKNNRELAYRPDSPTHDSTSSFLENEMSR
MSEELDKFRETTSSLMQQNMEYEKEVLQLRGDLESSMEHVRSLESHLKTREEELERLEA
ELFELKASGHGKAASAGNSIFAEAMEAERKLEEDLKSLFAQNQCLLKKVRRLHIEKEDA
EERAQSAMSRKYTVRSAINTQGHAEMDRLRQKVRELETERTHLWERLFVKLRNVRRSE
LGALVVGYHESFKLSIASVKGNHEQVLRENEELAVKIQGLERENADYEEKIEQLKYDLAT
AQRKEKAAIALEADEDFSSGKELMDAPLKGPVGRPTFFVKPKQVEFSEIPEDEEFTSGAS
HPLLSAPLKVQNSAARPSFFVKPKSEHIERGMMNMSMTPKNASRLPPTTPSSLDQENMS
DFALRHQRAKAERKAAATPAKPQYNFVQLVQPTAVSRFKPATLLPPKPAE
>C_japonica_SPDL1_IncompleteSequenceInfoInExon4
MVEDEEKLQLDIERLRRQIRQKDEMIEEMESELSREKTFIANDKAEIRERELNGQIRGLQI
EIDGKDGTISQKDGVIDELRVEIEKLNRELANRSDYPESDESSSFLETEMMRISDELQKFK
ETTSSLLEKNLEYEKECLQLKSDYESAIDHVNSLESHLKTREEEISRLESEVFDLKCSAHG
KIANAGNSIFAEAMDAEQKLEEDLKRVYHEKQYLMERLKRIQLEKEEAEERAQANLRR
NCTVRTAVSHVDIEEMRRLRARVHELEVERTCLWERLFIKMRTVPKREMGGLFAGYLES
FKPASLKTLFTDTKENTTSEYSERQQRIKAERKQAATPASRYNFVQLAVPSTVPKFKPPTL
NMPST

Supplementary Data 3.S10: All NDC-80 sequences used in this study

>C_tribulationis_NDC80_paralog1
MFGGRRTGGPGFNAGRLSTAVTPTKRFTDFGIGSTRKAGRLSMSQGPRPSFFTKGSTVPP
RDVKSLQAANVQKIYNFLVEQDGSDAPSESSIRSPGGGKDFKMIFESMYQHLSKDYEFPL
NARIEDEVMPIFKGLGYPFSLKTSFFQPMGGPHGWPHLLDALAWLVDVIRMNLAVSVDT
QNILFGDFLDHQKVQEKALSYMWYSTLFRDYTNDRKAAEDKDGVFWKEAKANLRQYF
ENSNEYEEMITNLRNVLQQLRFDCDEIEAEKGQEQTYVEDIARMKDDIRKAVEYLESTQ
RVKENKEAEVTSIKQELDMKNAEMEKCLGMVYELKERIEHQKQVHGCSGKEVRQMNL
ENSKDKETVSELQAELDEISKETWRLKNDDSFKEQKAKFVQLVENITKLLAGLDVKLNL
DPLSVPSDEKELKAGWETLNSVWVPEISRQMHQRKLELETDKARFVDKFAAAEERIQIE
NEMLCEAKKKEGRDERIQRIEREEWKVGRQQLEKRYDELENEREVLTKKMQMDGSLEK
EIKEEKDKMAKMESEAEAKSQEVQAAIREKLEEMLVEIAEIGQEKYMFHVEATEVSRLIS
GMCSLDS
>C_tribulationis_NDC80_paralog2_pseudogene
RNVFLLPIFSPAPRLFDQITFPNDINIKRTFYHFMGNTLVTKTIDEAKRIDQRYGGRYLITTF
EGAIIDQSGNLTGGGTPLTGRMNVTGASSSRFNNDIERKNHIFKTKTRMQQTEREINSIEA
MLNAEMTKMQTDKASAAALENEYNQLNETLKNLRSHVEKLHDQSMACDHRLAQIAPI
EEITASIAEITQELEVLRERQKRQVAQNQEAQHIVSSLSSKI
>C_sp41_NDC80
MFGGRRTGGPGFNAGRLSTAVTPTKRFTDYGIGSTRKSEAAGRLSMSQGHRPSLFQKGS
AVPPRDVKSIQAANVQKIYNFLVEHDGSEAPAESIIRTPRGKKDFEAIFESMYQHLSKDYE
FPTQGRIEDEVTQIFKGLGYPYPLKNSFFQPMGASHGWPHLLDALAWLVDVIRMNLAVS
VDTQNILFGDFLEQQKVQEKALSYMWYSTLFRDYTNDRKAAEDKDGNFWKETKANLR
QYFENSNEYEEMVTNLRNVLQQLRFDCDEIEAEKGQEQTYVEDIARMKDDIRKAMEYL
DSTQRVKENKESEVTSVKQELEMKTAEMEKAIGMVNELKERIEQQKLVHGCSGKEVRQ
MNLENSKDKETVSELQAELDEISKETWRLKNDDSFKEQKAKFVQLVENITKLLAGLDVQ
LKLEPLSVPIDEKQLKAGWETLNSVWVPEISRQMHQRKLELETEKARFVDKFAAAEERI
QIENEMLCEAKKKEGRDERIQRIEREEWKVARQQLEKRYDELENEREVLTKKMQMDGS
LEKEIKEEKDRMAKLESEAEAKSQDVQMAIREKMEQMVVEIAEIGQEKTMFHGESTDV
ARVIGGMCSLDC
>C_zanzibari_NDC80
MFGDRRKTGGPSFNGGRLSTAVTPTKRFTDARLSMSQGHRPSLFQKGSAVPPRDVKTLK
AANVSKIFNFLVESDGSEAPSESTIRSPPGKNDFIAIFESMYQHLSKDYEFPASARMEEEVS
SIFKGLGYPFPLKNSYFQPMGGAHGWPHLLDALAWLVDVVKMNQAVSRDTQNILFGDF
MDQGKVQEKALSYMWYSKVFRDYTNDRKAAEDKDGEFWTNSGAELRRYFENSNDNE
EMMTNLQNVLQQLHFDCDEIEAEKGQEQTYVEDIARMKDDIRKAAEYLESTQRVKELK
QAEFTAVKQDLDSRKAELEKAVGMVIELKERIEQQKRIHGCSGKEVRQMNLENSKDKE
MLSELQAELDEISKETWRLKNDDSFKDQTKKFVQVLVNIRKMLANLNIQLNLDLLQVPK
DEQELKVCWETLNGVWVPEISRQMHQRKLELETEKARFVDKFAAAEERIQIENEKLCEA
KKKEGRDERVQRIEREEWKVARQQLEKRYDELENEKEVLTKKMHLDGSLEKEIKEEKD
KMTKMEQVAEAKRQDLEMAIREKLEKMVVEISEIGQEKIMFHAESTDVARVINEKCLLD
C
>C_sinica_NDC80
MFGNDRRKTGGNFNTGRLSTAVTPTKRFTDFGMSSVRARHSISQSRPSLFTKGSAVAPRD
VKSIQAANVQKIRNFLVETDGPEAPDEATIRSPRGKNDFIAIFESMYQHLSKDYEFPVQAR

VEEEVTQIFKGLGYPYPLKNSFFQPMGASHGWPHLLDALAWLIDVIRMNQRVSSDTQGI
MFGDTFDQQKVQETALGYMWYSSLFRDYTNDRKAAEIKDGEFWVETKQKLRNFFENT
NEFEDMAVNLKNVLQQLCFDCDEIEAERGQEQTYVEDIARMKDDIRKAVEYLESTQRLK
ELKETDVTSVKQELEFKKSEIEKVHGMVNELKERIEQQKRIHGCSGKEVRQMNQENGK
DKETVSELQSELDELSKETWRLKNDDSFKEQKAKFVHIVENIMKILNGLKVEWNLAPLP
VPSDERQLKTCWETLNGVWVPEISRQMHQKKLELETERARFAHKFAAAERIQIESEMLR
EAKKKEGRDERIQRIEREEWKVTRQQQEKRYDELENEKEMLKKNMHLDGSLEKEIKEE
TEKMIKIETESEAKSQDLQATLRAKMEQILVEIA
>C_nigoni_NDC80
MFGGERRKTGGMNLNGRQSLAITPTKRFTDHGLSSVRKTDARPSMPLHGQSRPSLFQKG
STIPPRDVRSLQAANVQKIYNFFVETDGAEAPSERSIRAPGSRREFVVLFESMYQHLSKD
YEYPEPARLEDEVTQIFKGLGYPYPLKNSYYQPMGASHGWPHLLDALSWLVDVIKMNT
TVAANTQGILFGDILEQSKVQEKVLNYSWFASIYKDYTNDRKGTEDKDSQFWKDAKNK
LRQHFENSNEYDDFASNAKNVLQQLIFDCDEIESERGQEQTYVEDIARMRDDIRKAAEY
LQSVELVKEHKDAEMVKVKGELDSKVAEKEKLLRMVNELKDRIEQQKIIHGCSGKEVR
QMNLENSKDKEMVAELQAELDEVSKEMWRMKNDDSFKEQKAKFLQIVENITKLLSGL
NVQLNLDPMPVPADEKQLKTCWETLNSVWVTEISRQMHQRKLDLDTEKSRSLDRFAAA
QERIQIENEMLCEAKKKEGRDERTRRAERDEWKAARQQQEKRYDELENEKEVLMKKL
HLDGSLEQEIKEEKDKMAKIEKEAEEKTQYLRSAIRQKVEAVEMELAEIGQEKTMFHAE
CVAVEKLVQGTCGATH
>C_briggsae_NDC80
MFGGERRKTGGMNLNGRQSLAITPTKRFTDHGLSSVRKRTDARPSMSLHGQSRPSLFQK
GSTIPPRDVRSLQAANVQKIYNFFVETDGAEAPSERSIRAPGSRREFIVLFESMYQHLSKD
YEYPEPARLEDEVTQIFKGLGYPYPLKNSYYQPMGASHGWPHLLDALSWLVDVIKMNT
TVAANTQGILFGDFLEQSKVQEKVLNYSWFASIYKDYTNDRKGTEDKDSQFWKDAKNK
LRQHFENSNEYEDVASNAKNVLQQLLFDCDEIESERGQEQTYVEDIARMRDDIRKAAEY
LDSVERVKEHKDAEMVKVKGELESKVLEKEKLLRMVNELKDRIEQQKMIHGCSGKEVR
QMNLENSKDKEMVAELQAELDEVSKEMWRMKNDDSFKEQKAKFLQIIENITKLLSGLN
VQLNLDPMPVPADEKQLKVCWETLNTVWVTEISRQMHQRKLDLDTEKSRSADKFAAA
QERIQIENEMLCEAKKKEGRDERTRRAERDEWKAARQQQEKRYDELENEKEVLMKKL
HLDGSLEQEIKEEKDKMAKIEKEAEEKTQYLRSAIRQKVEAVEMEIAEIGQEKTMFHAEC
VAVEKLV
>C_remanei_NDC80
MFGTERRKTGGVNLNGRSSIAITPTKRFTDFGATSVRRTDGRPSLSVGQLGQQSSRPSIFK
TSVHAPRDVKSFQAANAHKIFNFLVESDGADAPAESVIRTPRGKNDFIVIFESMYQHLSK
DYEFPPLNQARIEEEVSSIFKGLGYPFHLKNSYFQPMGASHGWPHLLDALAWLVDFIKIN
KSVSADTQHIIFGDFLEPAKVQEKALSYAWFSTTFRDYTNDRKSAESSDSEFWVETKNKL
RKYFEDSNEYEDMTANAQSALQQLRFDCDEIESERGQEQTYVEDIARLKDDIRKAMDYF
ESVQHLKERKENETAVIKEELEAKVAENEKIQMAVNELKERIEQQKRVHGLNGKEVRKM
NLENSKDKDTVSDLQSEQEMLSKQLWRLRDENPFKDQKMKVIQIAENVTKILSGLNMQ
FELESLRPPENEKELRACWEILIGSWLPEINRQLHQRKLDVETEKSRFHQKFAAAERIQIE
NEVLDEVKKKEAREERIHRDERDEWKEARQKQEKRYDELENEKEMLKRKMQMDGSLE
KEIKAEKEKKEKIKKDLEEMHQGLEAVFRKKLEKIASETAEIGNEKTMFRIEAIEVEKLID
RKCSNQKV
>C_latens_NDC80

MFGTERRKTGGVNLNGRSSIAITPTKRFTDFGTTSVRKRTDGRPSLSVGQLGQQSSRPSIF
KTSVHAPRDVKSFQAANAHKIFNFLVESDGADAPAESVIRTPRGKNDFIVIFESMYQHLS
KDYEFPPLNQARIEEEVQSIFKGLGYPFQLKNSYFQPMGASHGWPHLLDALAWLVDFIKI
NKSVSADTQHIIFGDFLEPAKVQEKALSYAWFSTTFRDYTNDRKSAEYSDSEFWIETKNK
LRKYFEDSNEYEDMTANAQSALQQLRFDCDEIESERGQEQTYVEDIARLKDDIRKAMDY
FESVQHLKERKENEMAVIKDELEAKVAENEKIQMAVNELKERIEQQKRVHGLSGKEVRK
MNLENSKDKDTVSDLQSEQEMLSKQLWRLRDENPFKDQKMKVIQIAENVTKILSGLNM
QFELETLRPPENEKELRACWEILIGSWLPEINRQLHQRKLDVETEKSRFHQKFAAAEERIQ
IETEVLDEIKKKEAREERIRRDERDEWKEARQKQEKRYDELENEKEMLKRKMHMDGSL
EKEIKAEKEKKEKIKKDLEEIHQGLEAVFQKKLEQIASETAEIGNEKTIFRIEAIEVEKLIDR
KCSNQKV
>C_sp51_NDC80
MFGERRKTGGMSLGGRQSIAITPTKRFTDLGIASVRKTDARSSMSVNQPRMSLFQNKGS
SLPPREVKSIQQANVQKIHQFLVESDGDDAPAEVVVRTPRGKKDFTMVFEIIYQHLSKDY
EYPHPVRLEDEIPQIFNGLGYPYPLKNSYFQPLGAGHGWPHLLDALGWLVDVVKMSKA
VVPNTQSIIFGDFLEQSKVQEKTLNYAWTTNTYRDYTSDRKVIENREHPFWEQAKSQLR
NYFESSNEYEDMESNAKNALEQLMFDCEEIESERGQEQHLQEEIAKMKDDIRKAEDYLL
QTENLKSHKEKDLGKIRETLVEKQSELEKIQQLVNELKERIEKQKINHGLSGKEVRKLNL
ENFKDKETVSEVQGELEKLSKEIWTLNNDDRFKEQKSKFVQLTDNIAKLTTGLNIQLSLE
PLKAPTDERQLKTHWETLSNIWLPEINRQLHQRKLELETELSRFADSFAAAEERIQIESET
LCEAKKREAREERIRRNERDEWKTSRQELEKRYDELENEKEVLEKQMQIGGSLDKEIDE
ERARMLKTEQDLRAKEADLMAAIRQKLDGLFAGIVELDNEKMMILKEFNDLEMVMNV
KCANY
>C_sp44_NDC80
MFGNDRRKTGGMNLGGRASIAITPTKRFTDFGTGSVRKTDARLSVNQPRPSLFQKGSSL
PPREVKSIQQANAQKIYQFLVENDGDDAPSEASIRTPSGKKEFGAIFEAIYQHLSKDYEFN
HQARAEDEVTPIFKGLGYPYPLKNSYFQPMGSPHGWPHMLDALGWLVDFVKINKEVNK
DTQNIIFGDFLEQNKVQEKTLNYAWNTNTYREYTLDRKVLDDQSHPFWEQTRIQLREYF
ESSSDYGDMATNIKNALDQLMFDCEEIESERGQEQQLQEEIAKMKDDIRKAEDYLLQTE
NIKSHKEKDLEKTLQCLVERQSELEKVQIAVNELKERIEKQKINHGLSGKEVRKLNMENQ
KDKETVSEIQGELDKLSKKTWTLKNDDCFKEQKSKFVQLTDQISKLMSGLSVQLNLQPL
KAPSDEQELKTHWETLSNIWLPEINRQLHQRKLELDTELSRFGDKFSAAEERIQIESEMLC
EAKKREAREERIRRNERDEWKTSRQEMEKRYDELENEKEVLKKQIQIGGNLDKEIEEEK
TKMVKIEEALRTKEADLIAKLRQKLEEIVVGLAEIDQEKMRIHKEFNDLEIVMNKKCAH
Y
>C_sp48_NDC80
MFGNDRRKTGGMNLGGGGRASIAITPTKRFTDFGASSVRKVDPRLSVNQPRPSLFQKGS
SLAPREVKSLQAANIQKIYQFLVESDGDDAPSEQTIRAPRGKTDFVMIFETIYQHLSKDYE
FPQQTRLEDEVVLIFKGLGYPFPLKSSHFQPMGTGHGWPHLLDALGWLVDFVKISKGVS
TARQNILFGDFLEQDKVQEKALSYAWNTNTYRDFTSDRKVLDNRDHPFWEQTKSQLRT
FFESSNEYEDMANTGKNALEQLMFDCEEIESERGQEQHLQEEITKMKDDIRKAEDYLLQ
TENLKVRKEKDLMTMRETLAEKKAMLEKVLQAVNELKERIERQKIDHGLSGKEVRQLN
LENMKDKETVSEVQGELDALSKLSWTFKNDDCFREQKAKFVKLTVDIAKLMAGLDIQL
NLEPLKAPRDEPELKTHWETLSTVWLPEINRQMHQRKLELETELSRFADKFAAAEERIQI
ETEMLCEAKKREAREERVRRNERDEWKIARQELEKRYDALENEKEVLKRQMQIGGNLD

KEIDEEREKMRKIEEALRAKEAELETEIRLKLDEMFTQIAKIDAEKMMIFKQFTDLEAIM
NANCARY
>C_brenneri_NDC80
MFGNDRRKTGGMNLGGGRASIAITPTKRFTDFGGASSIRKVDPRLSVNQPRPSLFQKGST
LPPREVKSLQAANVQKIYQFMVEADGDDAPSEQTIRAPHGKTDFVMIFETIYQHLSKDY
EFPQQTRLEDEVIQIFKGLGYPFPVKPSYFQPMGVGHGWPHLLDALGWLVDFVKINKAV
NKDTQNIIFGDFLEQQKVQEKTLNYAWNTNTYREYTSDRKVLDNRDHPFWEQTRSQLR
AYFESSNEYEDMASNVKNALEQLMFDCEEIESERGQEQHLQEEITKMKDDIRKAEDYLL
QTENLKAHKEKDLVKVRENLAEKQAVLEKVLQAVGELKERIERQKIDHGLSGKEVRQL
NLENMKDKETVSEVQGELDALSKLSWTFKNDDCFREQKAKFVKLSVDITKLMAGLNIQ
LNLEPLKAPRDEPELKTHWETLSTVWLPEINRQMHQRKLELETELSRFANKFAAAEERIQ
IESETLCEAKKREAREERVRRNERDEWKTARQELEKRYDALENEKEVLKRQMQIGGNLV
KEIDVEREKMAKIEEALRAKEAELQTEIRLKLDEMFAQIAKIDAEKMMIFKQFTDLEAIM
NANCAHY
>C_wallacei_NDC80_paralog1
MYGNERRKTGAPNFSGRASLAITPSKRFTDVDMTSIRRSEVRPSLMGGQSRPSIFGKGST
LPPRDVKSMLYVNVQKIYHFLVEHDQGDAPSEQVIKSPQGKKDFIAIFESIYQHMSKDYE
FPDGKRLEEEIIPLFKALGYPYPLKDSYFKPIGAGHGWPHLLDALGWLVDLVRIMLSVKN
DTQNIMFGDFMEANKVQEKTLSYSWHTNIFREFTNDRKVIESREHPFWEEQKTNLRNVF
DSSNEYGDMTNNYKNALEQLKFDCDEIESERGQETNLQEEIARMSDDIRKANDYREQTE
QVRTLRETDFQKIQEALVEKKEENIRMQNTVNELKERIEQQKINHGLSGKQVREMNLEN
SKDKEHVSEIQGELDKLSKEMWTLNNENCFREQKAKFTEIIDHIQKTINGLDISIQLPHLK
PPTDEKQLKSGWETLTNNWIPEINRHLHQKKLELETEQSRFADRFAATQERIQIEIEMLQE
SKKRETREDRIRQNERSEWKSARQELEKRYDEQENRLEVLKKQLKMDGSLEKEIEEEKE
KMRKIEEEMTAKQEELVRILNGRLDEIVAMTAKIENEKLGIHKEFEEIERVILANQ
>C_wallacei_NDC80_paralog2
MYGNERRKTNFDGRPSNAITPNKRHTEGISSLRRSEVRPSIGGHPRQSIFGRASTIPSRDV
KALAHGNIDKIYNFFVEKGEEPPCKSIIQSPTGKDFAVIFETIYQHLSEGFVCSYQHPEEDV
IPIFKALNYPYTLKDSYFKPVGTSHGWPHLLDALGWLIDLVNISLSVNNNAQNMLFGDIL
ERNQVEEKTLSYAWYTSIYRDYTNDRKVIENRNIPFWEKQRDALRNAFNGFDGNMDVI
GRMKNELEQLKFDCDEIESETGQENQLKEEISRIKDDIRKADEYFEQTEQLREIREKNFQK
ILEELQEKQKENERIQKKVNEVKERIEEQIINHGIDGKQVREMNLENSKDKELVSEMRGE
LDKLSKEMWTLKNEDLFKEQKGKFQEVIDHIYKTINGLEIQITLTPMKSPSDERELKIGW
DTLTNIWIPEINRHLHQRKLELETEQSRFSDKFAASQERIQIELEMLEEAKKREAREERQR
QNERDEWKVSRQNLEKHYDELENEMEVLKKKLKMDGSLEKEIEEERVKIMKIEEELSSK
HNELINEITRKLDEIFAATAEIENEKLLIQKEVNELENAMKSVGLQ
>C_tropicalis_NDC80
MYGSERRKTGAMNLGGRTSMAITPSRRFTDVESSSVRRSEVRPSLMGSQARPSIFGKGST
IPPRDVKSMIGANVEKIWHFLVEHDQGDAPSDSVIRNPQGKKDFVAIFESIYQHMSKDYE
FPNGGRLEEEITPLFKALGYPYPLKDSYFKPIGATHGWPHLLDALGWLVDLVRIMLSVKD
DTQNIMFGDFPEATEVHERALSYSWHTSIYREYTNDRKVIESKGHPFWEKQRAELRNVF
DSSNEYGDMTSNYKNALEQLKFECDEIESERGQEANLLDEIARMKDDIRKAAAYQEQTE
NVKMLREADLEKLHAVLREKMEENQKMQKAVNELKERIEKQRVDHGLNGKQVRQMN
LENSKDKEMVSEIQAELDKVSKETWMLNNENGFRDQKLKFLEIHEHIIKTIHGLNLPVNL
PPLKPPSDELELKSGWETLTTVWIPDITRHLHQKKLELENEQSRFADRFAAAEERLQIENE
MLEEAKKREQREERTRQNERSEWKTARQRLEKESDELENQLEVLKKQMQIGGNLEMEI

EEERRKMKAIEEGMERKNEELIACLQKKLDEIFAATVEIENEKLAMQKEFEELEAVMQT
NCH
>C_doughertyi_NDC80
MFGGERRKTGGMSLTGRASLAITPTKRFTDFGISSSRTNPRPSLNIGPRQSLFGSKSASVAP
RDVKSIQQGNVQKIYRFFVETDGDDAPPESIIRAPGGKADFITIFEQIYQHLSKDYEFPQG
ARLEDEIAQIFKGLGYPFPLKKSFFQPMGCAHGWPHLLDALGWLVDLVRISIRVSSDTQN
ILFGDFMEESKLTSKFLNYAWHTKVYKDYTNDRKVIDNPHSQFWLEAKSQLRDHFETAN
EYDDWMDTAKNNLEQLKFDCDEIESERGQEQHLLDEIAKMKDDIRKADDYLQQTANLK
VHKETALQEVQQTLLEKQGELEKIQKVVNELKERIEQQKVKHGLSGKEVRQMNLENQK
DKETVSEIQGVLDQLSKEAWSLKNDDSFKEQKAKFVQLAEHITKIMSGLDIEINLARLSV
PSDERQLKANWETLTNVWMPEVNRQLHQKKLDLETEISRFTDKFAAAEERNKIESEMLC
EAKKKDAREERLYLNERKEWQEARQQLQKRCDELENEKEVLKKEMKMDGSLEKEIEE
EKTKMAKEVKELRAKQTELEAVLRQKLDQIVAEVTDIQNEKLMVYKEFNDLGMIFESKL
PNY
>C_sp54_NDC80_paralog1
MFGADRRKTGGGMSLGPRQSVAITPTKRFTDYGPSSMRKPDGRASLSHNIGQPRPSLFGT
KNSAVAPRDVKGLVTANVSKIYSFLIEFEGSEAPSEKDIKSPRGKNDFTFVFELIYQHLSKD
YEYPQSARIEDEIPQIFKGLQYPVTLKNSYFQPMGSSHGYPHLLDALAWLVDLVRINSCV
SADTQNIIFGDFMEQAIVQEKTLNYAWMSTTFRDYTNDRKAAENPETPFWDETKCRLRK
YFEESNEFDDMAKTASNGLEQLRFECEEIESERGNEQSLLEEIAKIKDDIRKATEFFESTEQ
VKELKEKELARVKLELEAKMAEHERAQRMVNELKARIEMQKTNHGLSGKEVRQLNIEN
NRDKETVAEIQAELDSLSKEAWRLNKDDSFKEQKAKFVQLYQKTMKIIAGLDIQLNLDPI
KAPADERQLKTCWEALNGIWLPEINRQLHQKKLELETKNSRFADNFAHTEQRIQMESEM
LCEAKKKEAREERIRRNERDEWKEARQQMEKRYDELENEKEVLKKQMQMDGSLEKEI
EAERVKMTKIEKELEEKKKYLETSIRQKLDQIVAETAEIENEKIMFHVECTELEKTVKAM
CLN
>C_sp54_NDC80_paralog2
MNSTFILNHFSLLRHVKKMFSGGPRDVKSLVTANVKKIYSFLIEFEGSEAPSEKDIKSPRG
MNDFISIFESIFHHLNKDYKYPQSARVEEEISQIFKGLQYPFPLKVSFFQPFGSSHGYPHLL
DALAWLVDLVRMNSCTNDTNEKEMLKMQMTKIEKELEEKKKNLETSIRQKLDQIVAET
AEIENAKMMLLEMKCTELEKTVKAMCLN
>C_inopinata_NDC80
MYGDRRRTGGISLNGRASLAITPTKRYTDFRTETRPSLGQSSGQHRLSIFHKTPSVAPRDV
KSIFTANVSKVHNFLAEHEGLEAPLESTIRNPRGRNDFICCFELIYQHLSKDYEFRKDERIE
DEISQIFKGLGYPFALKNSYFQPMGSSHGYPHLLDALAWLVDIIRVNAAVSADTQNIIFGY
FMEQSKAQERTLHYAWMSSIFREYTNERKVAENSESPFWGEVKNKLRKFFEDSGEFEDM
TNSAASALEMLRFECDEIESEKDNEQSILEEITRIKDDIRKASAYLETTQQVLQHKEIELAG
VQSEVADKIAENEKVQKMVNDLKTRIENQKNHHGLTGKEVRELNLENNRDKEMVSEIQ
AELNNLSKETWRLKDEDHFKEQRSKFVQLFDSVTKMIAGLNIQMNLKQISSPNNERELK
VCWETLDNVWLPEINRQLHQRKLDLETEQSRFSSKFVAAAERIQMKSEILSEVKKNEIRE
ERLRQNERDEWKENRKQKEKRYDELENEKEVLKTQMQMDGSLEKEIEAERISMAKIEK
ELEEKCKYLQSVIKKKLDQIVTETTAIESQKTIFHLQCTEIEKAVN
>C_elegans_NDC80
MFGDRRKTGGLNLNGRASIAITPTKRFTDYTGSTSVRKTDARPSLSQPRVSLFNTKNSSV
APRDVKSLVSLNGSKIYNFLVEYESSDAPSEQLIMKPRGKNDFIACFELIYQHLSKDYEFP
RHERIEEEVSQIFKGLGYPYPLKNSYYQPMGSSHGYPHLLDALSWLIDIIRINSAVSEDTQ

NILFGDFMEQGKAQEKTLNYAWMTSTFRDYTNDRKAAENPSSSYWDDTKHRLRKYFE
QSNEFEDMTKTAASALEMLNYECDEIEADKGNEASLKEEISRIRDDIRKAKDYLEQNLH
VKQHMEKELAMVKSEQEEKISENEKVQKMVDDLKNKIELQKQIHGLTGKEVRQMNLD
NNKDKEVVLEIQSELDRLSKETWKLKDEDFFKEQKSKFIHLAEQIMKILSGLNIQMNLEP
LRAPTNERDLKDYWETLNKIWVPEISRQLHQRKLELETEQSRFSNKAVTAEERIQIQSETL
CEAKKNEAREERIRRNERDSWKDARKHIEQRYEQLLNEKEVLLKQMKLDGSLEKEIEDE
TARMSATGEEHIQKRSQLEAGIRQILDLMVVEIAEIENKKIGFHVQCAGIEKAVL
>C_oiwi_NDC80
MFGDRRKTGGLNLNGPRASLAITPTKRFTDLGVNSVRKSTTRMSLSQTSGQPRMSLFQS
KSIAPRDVKALVSSNVTKIYNFLVEHEATEAPSEKDIRSPRGKNDFVLFFELIYGHLSKDY
EFPQDTRLEDEIPLIFKGLGYPYPLKNSYFQPMGSSHGYPHLLDALSWLIDLAKISSSVRS
DTKNIIFGDFMDQNKALEKILSYSWLSTVYKDFTNDRKAAETDQFWVETKGTLRKHFE
HNNEYDDMIETAKSALEQLRFECDEIESDKGNEQSLQEEIARIRDDIRKASDYLEQTEML
KERTQEELNEINTQLQVKSAENEKVQGMVNELKARLEQQKLHQGLTGKEVRQMNLEN
NKDKETVAELQAQLEELSKQAWKLKNDESFKEQKTKFVHLVENALKVLSGIGINTNSDA
FRVPTNEKELQASWETMNNVWLPEINRQLHQRKLELETEKLQFADKFASAEERIQMETE
MLREAKKKEVREDRLRRTEREEWKESRQQLEKRLDELENEKEVLKKQLQIDGSLEKEIE
TEKVRMAKIEKEMEAQKERLEAAIRKKMDEIVEETAGIENSKIMFHKSCVEVEKTVKTA
LKVSKK
>C_kamaaina_NDC80
MFGDRRKTGGLNLNGPRASLAITPTKRFTDLGVNSMRKSTARVSLSQTSGQPRMSLFQS
KSTAPRDVKALVSSNVTKIYNFLVEHEATEAPSEKDIRSPRGKNDFVVFFELIYGHLSKDY
EFPQDTRLEDEVKKIFKGLGYPYPLKNSYFQPMGSSHGYPHLLDALSWLIDLAKISSSVR
SDTKNIIFGDFMDQNKALEKILSYSWLSTVYKDFTNDRKAAETDQFWVETKETLRKHFE
HNNEYDDMIETAKSALEQLRFECDEIESEKGNEQSLQEEIARIRDDIRKASDYLEQTEML
KERTQEELNEIKTQLDTKSVESEKMQGMVNELKARIEQQKLHQGLTGKEVRQMNLENN
KDKETVAELQAQLEELSKQAWKLKNDESFKEQKTKFVHLVENALKVLSGIGINTNSDAF
KVPANEKELQASWEAINNVWLPEINRQLHQRKLELETEKLQFADKFASAEERIQMETEM
LREAKKKEMREDRLRRTEREEWKESRQQLEKRLDELENEKEVLKKQLQIDGSLEKEIET
ERVRMAKIEKEMEAQKQRLEAAIRKKMDEIVEETSGIENSKMMFHKSCVEVEKTVKAL
KVSNNQ
>C_waitukubuli_NDC80
MFGDRRKTGGPGFGPRLSTAITPTKRFTDVGPTSVRKADSRLSMGQGGGGQPRASLFQRS
SAVASRDVKSLMAANVNKIYNFFIESEGSEAPSEAQIRNPKGRNDFITFFELLYQHLSCDY
EFPQGARVEDEVSALFKNLGYPTALKNSFFQPMGSSHGYPHLIDALAWLVDLIRVNMCV
SNDTQNIIFGDFLEQSKVQERTLSYCWMSSTYKEYTNDRKAAETRDGPFWDSTKIKLRK
YFEDMNETGDMAASAKTSLEQLRFECEEIEADKGTEQNLIEEIAKIKDDFRKASDYCERL
EAVERQREVDLDQVKMEVDVKKKEMSEVQEEVNSLKQRIEEQKQNHGLTGKQVRQLN
SDNSRDKAIVSEIQSELERLSKELWLRKNEENFKEKKTMFLQLAEKLAKIGAGINVDLGL
TSLRSPQNEELKIECEKLVGVWLPEVTRHLNQKKLELEMEKSRFNDKFASAACANMEQE
TLCEAQKKAAREERIRRNEREEWTESRLAKEKRYDELENELDVLKKNMQMGDSLVKEI
EMERAKKVKLCEEQEKQKKTIEASIRQKMDQIMTETAEIDNEKMMIHVECNELEKMV
>C_panamensis_NDC80
MFGDRRKTGGPGFGPRLSTAITPTKRFTDVGAAASVRKVDNRLSMGGQHRASLFQRSSA
IAPRDVKSLMSANVSKIYNFFLEFEGPDAPSESQIKNPKGRNDFISFFELLYQHLSCDYEFP
ANVRVEDEVQISTIFKNLAYPTPLKNSFFQPMGSSHGYPHLIDALAWLVDLIRNKCVSDD

TQNIIFGDFLEPAKVQEKTLSYAWMSSTYREYTNDRKAAETKDGPFWEVTKNKLRKYFE
DMSETQDLAASVKNSLEQIRFECEEIEADKGTEQSLIEEIAKIKDDLRKATDYAEKLEAVE
RQRLMDLDKVKADLDVKKKENSEVQTEVDGLKQRIEEQKQNHGLTGKEVRQLNSDNA
RDKAVSNELQSEMEKVAKDLWRNNNEENFKEQKRTSFSRLVERIEKILAGINVTLRLEPL
RTPQTERDLKAGLDELMSVWLPEITRQLNQKKLELNMEESRFTDKFVAVQARVQLEKET
LDEAKKKEAREERMRRNEREEWMESRLATEKRYDEMENEQDVLKKQLQMDGSLEREI
EMEKARKVKISEEQENKKNKIDARIRQKMDEVMKETAEIDNGNLLFHVACNELEKMVL
KTQNLI
>C_nouraguensis_NDC80
MFGERRKTGGQGFGPRISTAITPTKRFTDVGAAASVRRTDSRLSMGQSGGQHRASLFQR
NSAAAPRDVKAMLGPNVNKIYNFFIEFEGSDAPAEAQIKCPKGRNDFINFFELLYQHLSC
DYEFPPNARVEDEVLWIQIFKCIGYPVSLKNSYFQPMGSSHGYPHLVDALAWLVDFIRNK
CVSDDTQNIIFGDFLEPAKVQEKTLSYAWMSSIFREYTNDRKAAETRDGPFWTITKERLR
DYFEKMNEAQGLAVSLKNALEQIRFECEEIEADKGQGQTMLEEINKMKDDMRKAVDYN
EALQTAQRQREIDLEKAMTDLEVRMKEDSEVQADVNELKKRIEEQKEKHGLTGKEVRQ
LNLENTRDKAVISEIQAELDKLSKELWLKKNEENFKDKRASFVRLAENIGKIVAEVYIDL
GLEPLRSPQNEKELKSELEKMNSVWLPEVTRQLRQKQLNLNKEKSFFRDKERVQIERDT
LCEAEKKMAREERIRRTERDEWKEIRLANEKRYDELENEQDVLKKQMQKDGSLDKEIE
LEMARKLKIDEESSKKKSALEAVIRRKLDHIMAETSKIDNEKLMFHTDCTEFEKQILKTR
NMK
>C_becei_NDC80
MFGERRKTGGPGFGPRISTAITPTKRFTDVGAAASVRKTDSRLSMGQPGGQHRASLFQR
NSAAAPRDVKSMLGANVSKIYNFFIEFEGSDAPAEAQIRNPKGRNDFINFFELLYQHLSC
DYEFPPNARVEDEVIHIFKCIGYPVSLKNSYFQPMGSSHGYPHLVDALAWLVDFIRNKCV
SDDTQNIIFGDFLEPAKVQEKTLSYAWMSSIFREYTNDRKAAETRDGPFWTVTKEKLKDF
FEKMNEAQGLAASLKNALEQIRFECEEIEADEEINKMKDDMRKAVDYNDALQTAQRQR
EIDLEKAMTDLEIRMKENSEVQADVNELKQQIEEQKEKHGLTGKEVRQLNLENTRDKAV
ISEIQAELDKLSKELWLKKNEENFKDKRAIFVQLAEKIGKIVAEVHIDLGLESLRSPQNER
ELKSEQEKMNSVWLPEVTRQLRQKQLNLNKEKSFFRDKFAAAEERVQIKKVTLCEDEK
EMAREERLRRNEREEWKEIRLVNEKRYDELENEQDVLKKQMQRDGSLDKEIELEMARK
LKIDEESSKKKSALEAKIRQKLDQILVETSKINNEKLMFHSDCQEFEKQILKTRNMK
>C_yunquensis_NDC80
MFGDRRKTGGPGFGQRNSTAITPTKRFTDVGAASSVRKTDNRLSMGQTGGQHRASLFQ
RNSAAAPRDVKSMMGANVSKIYNFFLEFEGPADAPSEGNIRSPKGRNDFINFFELLYQHL
SKDYEFPANARVEDEIIHIFKFLGYPISLKNSYFQPMGSSHGYPHLVDALAWLVDFIRCSIS
VSGDTQTIYYGDFLDQAKVQERTLSYAWMTTIFREYTNDRKAAETRDGPFWTMTKEKL
RTYFEKMNEDQDLTASVKNALDQIRFECEEIEADKGKGQEISEEIVKVKDDIVKAIDYGE
ALEAAQKQQESDLEKATAELRERVDENSEVQANVNELKQRIEDQKKKHGLTGKEVRQL
NSENTRDKAMITEIQAELDKLSKELWLKKNEENFKDKRESFVQLVERIGKIVAEVSIDLG
LEPLRSPRNERELKTELEKLNGIWLPEVTRQLRQKQLNLDKEKAFFRDKFAAAEERVQIE
RDTLCEAEKKMAREERVRRHERDEWKESRLVIEKRYDELENEQDVLKKQMQMDGSLD
KEIELEKAKKVKIDEEIAKKKAALEAAMRQKLDKIMVETAKIDNEKILFHVECTEFEKQI
LKTRNLN
>C_macrosperma_NDC80
MFGDRRKTGGPGFGPRISTAITPTKRFTDVGAAASARKVDSRLSMGQPGGQHRASLFQR
SSAVAPRDVKSMLAANVNKIYNFLVEYEGTDAPSEAQIRCPKGKNDFINFFELLYQHLSC

DYEFPANGRLEEEISNTFRHLGYLPHLKNSYFQPMGSSHGYPHLVDALAWMVDIIRCNK
CVSDDTQNIIFGDFLEPAKVQEKTLSYAWMSSTFREYTNDRKAAEAINGPFWNSTKNKL
RKFFEDMNDTQDLTASAKNALEQLRFECEEIEADKGTEQSLLEEITRIKDDIRKAICHYED
TKRVLNQRESDLEKVKAELDTRIKENNEVQAEVNLLKNRIEEQKEKHGLTGKEVRQLNS
NNTRDKAVVSEIQTELDKLSKELWLMKSEDTFKEKKTCFVHLTERIGKIVAGIEIDLGLEP
LRSPQNERELKVEWEKLNNVWLPEITRQLHQKKLELEMEKSRFSDKFAAIEERVQMERE
TLCEAKKKESREERLRRNEREEWKESRLLKEKRYDELENELDVLKRQMQMDGSLDREI
EIEKAKKVKNDEELKKKKAAIEASIRQKLDQIVAETAKIENEKTLFHVECIEFEKLILNTQ
NLV
>C_sulstoni_NDC80
MYGGDRRKTGGFGFGNPRTSAITPTKRFTDLGTTSSVRKDHGRLSMSQTSGHHRVSLFG
GASATGPKDVRANFDANAKKIYNFLVEHEGSEAPSENVIRCPAGKSDFVTVFELMYQHL
SKDYEFPQNVRVEEEVSNIFKALGYAPPLKNSYFQPMGSAMGYPHLVEALAWLVDLIRIN
ASVSADTQPIIFGDVMDQEKVHEKTLSYAWMSTTFRDYTNDRKAAESGTGPFWEETKS
KLRDYFEKSDEIEDVATNYKSALEQLRFECEEIEAERGNELTLLEEIAKIKDDVRKALEYD
ETTARVQKHKEDELKAVKGTLEQKMAENSKVQAEVAELKERIEVQKQMHGLTGKEVRQ
LNSDNNRDKETVTELQAELDEISKQMWRLKNEDTFKQQKTNFVRLVESVEKIVSGLDIE
MGLEPLRAPSNESELKSRWDELNSVWIPEINRQLHQKKLEFETEQSRFADRFAAAERVQ
MERELLCEAKKKEAREDRVRRNERDEWKENRLQREKKYDELENEKDVLTKKMQMDG
TLDKEMEEERTKAWKLEQELQKEADAVENAIRKNLDLIVNEVAQIENDKMLFHVECCEF
EKLIVKTQNMF
>C_afra_NDC80
MYGGDRRKTGGFGFGNPRTSAITPTKRFTDLGTTSSVRRVSLFGGGSAAGPKDVKGNFE
ANVKKIHSFLIEHEGNEAPADTVIRSPAGRNDFLTVFELMYQHLSNNYEFQQGVRIEDEV
SNLINASVGADTQTILFGDVMDQAKVQEMTLRYSWMSTTFRDYTIDRRAAESGTGPFW
DETKNKLREYFEKSDEIEDLVTSYKTALEQSAFECQEIEADKGNEQHLLEEISKMKDDVR
KAMEYAESTARVQKHKEEEMKTVKATLETKIAENSKVQAEVAELKERIEVQKQLHGLTG
KEVRQLNSDNNRDRETVTELQAELDEVSRQMWRLKSEDTFKQQKANFIRLVESVEKIVS
GLELQLGLDAILAPSDESELKVRWDELNNVWIPEVNRRLHQKKLEFETEQSIFVDKFAAA
ERVQMERELLCEAKKREAREDRVRRNERDEWKENRLQREKKYDELENEKYVLKKKMQ
MDGSLDKEVEEEQAKARKIEEELQKEADAIENSIRKNLDQIVSETAQIENDKMLFHVECC
EFEKLI
>C_sp49_NDC80
MMFGGPRRKTGGPNFGATGRTSTAITPTKRTTDLGAAHSVRKNDNRMSFGQGPSNSRPS
LFQRTGGPSKDFKAAVPGNAAKVYNFFVDTEGANAPSEKTIRTPAGKGDFVTMFELLYQ
HLSKDYEFPSNGVRLEDEFTNIMKALGYPHALKPSFFQPIGSSHGYPHLLDALAWLVEAV
EINEKVRLDTQNIIMGDFMEPDQVQDKVLSYSWLSKTFFEYSNNRKVLEDKADPFWATT
KCDLRNFFENNDESADMITSSKNMLEQLRFDCEEIEADKGNEQGLLDEIARIRDDIRKAQ
LYLDETMRVVAQSNHEYQTVVDACEAKLAELEKVKTEVAELKERIEEQKRLHGLSGKE
VRQMNAENSKDKEMLSDVNAEIERATKMMWRLRDAADYKEPSTRFYRLISHMHQILD
DADVHPQLEPIASPSSELDLKAGWDAINNLWLPEVNRQLQHKKLELETENARFLNKFEA
IEERATMEQEMLNEATKKEDREERVRRNEREEWKAARFQKEKRCDQLENEKDILVKQM
HLDGSLDRDLQEETEKAAKLSETLEQKQAAFEQKVREKIDEMMASTAEMEQEKLEFHK
ECVELEMVVKNS
>C_sp25_NDC80

MFGGPRRKTGGPGFSTTGRTSTAITPTKRNTDLGTAQSVRKNDGRMSFGANPSGSRPSLF
QRTVAPSKDVKAAVPANVARIYNFFVETEGPNAPSEKMIRSPAGKGDLITMFELIYQHISK
DYEFPSVGVRLEDEFTNIMKALGYPNALKQSFFQPIGSSHGYPHLLDALAWLVEAVEINE
KIRLVTQNILVGDFMEPDQVQDKVLSYSWFSKTFLEYTNNRKVLEDRSDPFWANTKREL
RDFFERNDESADMIATSKNMLEQLMYDCEEIEADKGNEQGLIDEIARIRDDVRKAQAYL
EETSRVVAQSIHEFQSVSDAFEAKRTEFEAVKTEIAELKERIEEQKRNHGLCGKEVRRLNL
ENIKDKETVAELQAELEKTSKIIWNLRDAGSFKEQKARFDRLITHMKMILDDTDIHFQLE
SITSPRTELELKAGWDTVNNVWLPEVNRQLQRKLLELETENARFASKFGAKEEHVTMEQ
EMLNEATKKEDREERVRRNEREEWKAARFQKEKRFDQLENEKDILMKQLHLDGSLDRE
LEEETEKSIKLTEALEQKKAEFEQKVREKIDLMMVSTAEMEQERLDVHKEFVELETIINSS
>C_imperialis_NDC80_paralog1
MFGGPRRKTGGPNFSATGRTSTAITPTKRNTDLGAVQSVRKTDTRMSFGQGPSAPRASLF
QRAGLPSRDVKACFASNVSRIYNFFVETEGSSAPSEKTIRSPNGKGDFISMFELIYQHLSK
DYEFPTQNRLEEEFSNILKSLGYPYPLKNSFFQPIGSSHGYPHLVDALAWLVEVCEVNEK
VRLATQNILLGDFMEPEQMKDKFISYSWFSKVFLEFTNQRKAAEDKTDPFWESTRRELR
EFFEQNDESVEMIASLKNMLEQLRFDCEEVEEDKGNEQSLLDEISRLRDDVRKAQSYLE
ETARVVAQSHHEFATVKEAVDAKMEEFERVKAEVAELKARIEQQKELHGLTGKEVRLMN
VENNKDKETVAEIQNELDRTSKIIWRLRDAGSFKEQKSRFERLVEHMLKIVDDADIHLTL
DPINSPSTELELRAGWEAVNTVWLPEVNRLLQRKKLELEGENARFSSKFGAVEEKAILEK
ELLNEATKQEAREERVRRNERDEWKESRLQKERRCDQLENEKDVLVKQLHLDGSLDVEI
REKADEMEKLVVELAGKKKTLEGEVRKRIDEMMESTVQMEHEKLLFHRECVELEALVK
SQ
>C_imperialis_NDC80_paralog2
MFGNARRKTGGPNFATGRPSTAITPTKRITDLGAAQSVRKNDSRMSFGPSAHRTSLFQRA
GPHSKDMKACFAANVAKIYHFFVETEGSEAPPEKMIRSPTKNEFITIFELVYQHLSKDYEF
HMQNRLEEEFTSILKALGYPYPLKNSFFQPIGSSHGYPNLVDALAWLVEVVEVNSAVSRV
TQNILIGDFMEPELAEDKVLSYSWFSKTFLEFTNNRKALEDKSDPFWESTRLGLREYFER
NDETAEMIASCKNTLEQMRFDCEEIEEDKGNEQSLLDEISRLRDDVRKAEAYLLETSRAL
EQATQEHESVKEAVETKGAEFEQVKMEVARVKELIEQQKEQHGLTGKEVRAMNVECQR
DKETVSEIQMELDKVSKVIWRLRDAGSFKEQKSRFDRLIEHMHTILVDADIHLELEPICSP
STEIELRAGWEAINGVWLPEVNRQLQHKKLELEAENARFSSKFGVVEEKAILEQELLNE
AIKQEARDERVRRNEREEWKASRLQKEQRCDELENEQDVLVKQLNLDGNLDAEIREAA
QKAERLFLELQAKKEQFEGAVRAKIDEMMESTVAMEQEKLMLHKECSELESSVKSTL
>C_japonica_NDC80
MFGERRKTGGANFGFAGGRPSTAITPTKRFTDLGTAQSTRRVDNRLSMGQNGGRLSVFQ
RGSAVAPRDVKSAQASNVTKIYNFLLEHEQSGAPPEKIIRQPTGKNDFITMFEQLYQHLSK
DYEFPQGARVEDEFTGIMKGLGYPFPLKNSFFQPMGSSHGYPHLLDALAWLIDVISLNQS
VSSVTQNILLGDFMEAAEAQDKIISYSFYSSTFREYTYDRKAIESKDAPFWAETKERLRD
YFEKNDEITDLVTSTKQVFEQLRFECEEIEADKGDEQGLLEEIARIKDDVRKAHEYLESTE
RVQKQTGVELGIVQESLEAKKAEFEAVQAEVNELKRRIEVQRQQHGLTGKEVRQLNVE
NNRDKEAVHEIQTELDNVSKTLWRMRDEDTFREQKANFVLVIENMEKILLDANVKIGLD
ALRPPQNERDLKVGWDALNNQWLPEVNRQLQHKKLELDDEKTQFSSRFAAIEERVQMQ
QELLREANKKEAREERVRRNERDEWKTDRLQREKRLDELENEKDVLKNQMQTGGSLD
REIEEEKAKSAKLVEALEQKKASLAEGIRQKLDETMREITKIDLENATFHAECTDIEKLVL
KTRNMNQ

>C_tribulationis_NDC80_paralog1
MFGGRRTGGPGFNAGRLSTAVTPTKRFTDFGIGSTRKAGRLSMSQGPRPSFFTKGSTVPP
RDVKSLQAANVQKIYNFLVEQDGSDAPSESSIRSPGGGKDFKMIFESMYQHLSKDYEFPL
NARIEDEVMPIFKGLGYPFSLKTSFFQPMGGPHGWPHLLDALAWLVDVIRMNLAVSVDT
QNILFGDFLDHQKVQEKALSYMWYSTLFRDYTNDRKAAEDKDGVFWKEAKANLRQYF
ENSNEYEEMITNLRNVLQQLRFDCDEIEAEKGQEQTYVEDIARMKDDIRKAVEYLESTQ
RVKENKEAEVTSIKQELDMKNAEMEKCLGMVYELKERIEHQKQVHGCSGKEVRQMNL
ENSKDKETVSELQAELDEISKETWRLKNDDSFKEQKAKFVQLVENITKLLAGLDVKLNL
DPLSVPSDEKELKAGWETLNSVWVPEISRQMHQRKLELETDKARFVDKFAAAEERIQIE
NEMLCEAKKKEGRDERIQRIEREEWKVGRQQLEKRYDELENEREVLTKKMQMDGSLEK
EIKEEKDKMAKMESEAEAKSQEVQAAIREKLEEMLVEIAEIGQEKYMFHVEATEVSRLIS
GMCSLDS
>C_tribulationis_NDC80_paralog2_pseudogene
RNVFLLPIFSPAPRLFDQITFPNDINIKRTFYHFMGNTLVTKTIDEAKRIDQRYGGRYLITTF
EGAIIDQSGNLTGGGTPLTGRMNVTGASSSRFNNDIERKNHIFKTKTRMQQTEREINSIEA
MLNAEMTKMQTDKASAAALENEYNQLNETLKNLRSHVEKLHDQSMACDHRLAQIAPI
EEITASIAEITQELEVLRERQKRQVAQNQEAQHIVSSLSSKI
>C_sp41_NDC80
MFGGRRTGGPGFNAGRLSTAVTPTKRFTDYGIGSTRKSEAAGRLSMSQGHRPSLFQKGS
AVPPRDVKSIQAANVQKIYNFLVEHDGSEAPAESIIRTPRGKKDFEAIFESMYQHLSKDYE
FPTQGRIEDEVTQIFKGLGYPYPLKNSFFQPMGASHGWPHLLDALAWLVDVIRMNLAVS
VDTQNILFGDFLEQQKVQEKALSYMWYSTLFRDYTNDRKAAEDKDGNFWKETKANLR
QYFENSNEYEEMVTNLRNVLQQLRFDCDEIEAEKGQEQTYVEDIARMKDDIRKAMEYL
DSTQRVKENKESEVTSVKQELEMKTAEMEKAIGMVNELKERIEQQKLVHGCSGKEVRQ
MNLENSKDKETVSELQAELDEISKETWRLKNDDSFKEQKAKFVQLVENITKLLAGLDVQ
LKLEPLSVPIDEKQLKAGWETLNSVWVPEISRQMHQRKLELETEKARFVDKFAAAEERI
QIENEMLCEAKKKEGRDERIQRIEREEWKVARQQLEKRYDELENEREVLTKKMQMDGS
LEKEIKEEKDRMAKLESEAEAKSQDVQMAIREKMEQMVVEIAEIGQEKTMFHGESTDV
ARVIGGMCSLDC
>C_zanzibari_NDC80
MFGDRRKTGGPSFNGGRLSTAVTPTKRFTDARLSMSQGHRPSLFQKGSAVPPRDVKTLK
AANVSKIFNFLVESDGSEAPSESTIRSPPGKNDFIAIFESMYQHLSKDYEFPASARMEEEVS
SIFKGLGYPFPLKNSYFQPMGGAHGWPHLLDALAWLVDVVKMNQAVSRDTQNILFGDF
MDQGKVQEKALSYMWYSKVFRDYTNDRKAAEDKDGEFWTNSGAELRRYFENSNDNE
EMMTNLQNVLQQLHFDCDEIEAEKGQEQTYVEDIARMKDDIRKAAEYLESTQRVKELK
QAEFTAVKQDLDSRKAELEKAVGMVIELKERIEQQKRIHGCSGKEVRQMNLENSKDKE
MLSELQAELDEISKETWRLKNDDSFKDQTKKFVQVLVNIRKMLANLNIQLNLDLLQVPK
DEQELKVCWETLNGVWVPEISRQMHQRKLELETEKARFVDKFAAAEERIQIENEKLCEA
KKKEGRDERVQRIEREEWKVARQQLEKRYDELENEKEVLTKKMHLDGSLEKEIKEEKD
KMTKMEQVAEAKRQDLEMAIREKLEKMVVEISEIGQEKIMFHAESTDVARVINEKCLLD
C
>C_sinica_NDC80
MFGNDRRKTGGNFNTGRLSTAVTPTKRFTDFGMSSVRARHSISQSRPSLFTKGSAVAPRD
VKSIQAANVQKIRNFLVETDGPEAPDEATIRSPRGKNDFIAIFESMYQHLSKDYEFPVQAR
VEEEVTQIFKGLGYPYPLKNSFFQPMGASHGWPHLLDALAWLIDVIRMNQRVSSDTQGI

MFGDTFDQQKVQETALGYMWYSSLFRDYTNDRKAAEIKDGEFWVETKQKLRNFFENT
NEFEDMAVNLKNVLQQLCFDCDEIEAERGQEQTYVEDIARMKDDIRKAVEYLESTQRLK
ELKETDVTSVKQELEFKKSEIEKVHGMVNELKERIEQQKRIHGCSGKEVRQMNQENGK
DKETVSELQSELDELSKETWRLKNDDSFKEQKAKFVHIVENIMKILNGLKVEWNLAPLP
VPSDERQLKTCWETLNGVWVPEISRQMHQKKLELETERARFAHKFAAAERIQIESEMLR
EAKKKEGRDERIQRIEREEWKVTRQQQEKRYDELENEKEMLKKNMHLDGSLEKEIKEE
TEKMIKIETESEAKSQDLQATLRAKMEQILVEIA
>C_nigoni_NDC80
MFGGERRKTGGMNLNGRQSLAITPTKRFTDHGLSSVRKTDARPSMPLHGQSRPSLFQKG
STIPPRDVRSLQAANVQKIYNFFVETDGAEAPSERSIRAPGSRREFVVLFESMYQHLSKD
YEYPEPARLEDEVTQIFKGLGYPYPLKNSYYQPMGASHGWPHLLDALSWLVDVIKMNT
TVAANTQGILFGDILEQSKVQEKVLNYSWFASIYKDYTNDRKGTEDKDSQFWKDAKNK
LRQHFENSNEYDDFASNAKNVLQQLIFDCDEIESERGQEQTYVEDIARMRDDIRKAAEY
LQSVELVKEHKDAEMVKVKGELDSKVAEKEKLLRMVNELKDRIEQQKIIHGCSGKEVR
QMNLENSKDKEMVAELQAELDEVSKEMWRMKNDDSFKEQKAKFLQIVENITKLLSGL
NVQLNLDPMPVPADEKQLKTCWETLNSVWVTEISRQMHQRKLDLDTEKSRSLDRFAAA
QERIQIENEMLCEAKKKEGRDERTRRAERDEWKAARQQQEKRYDELENEKEVLMKKL
HLDGSLEQEIKEEKDKMAKIEKEAEEKTQYLRSAIRQKVEAVEMELAEIGQEKTMFHAE
CVAVEKLVQGTCGATH
>C_briggsae_NDC80
MFGGERRKTGGMNLNGRQSLAITPTKRFTDHGLSSVRKRTDARPSMSLHGQSRPSLFQK
GSTIPPRDVRSLQAANVQKIYNFFVETDGAEAPSERSIRAPGSRREFIVLFESMYQHLSKD
YEYPEPARLEDEVTQIFKGLGYPYPLKNSYYQPMGASHGWPHLLDALSWLVDVIKMNT
TVAANTQGILFGDFLEQSKVQEKVLNYSWFASIYKDYTNDRKGTEDKDSQFWKDAKNK
LRQHFENSNEYEDVASNAKNVLQQLLFDCDEIESERGQEQTYVEDIARMRDDIRKAAEY
LDSVERVKEHKDAEMVKVKGELESKVLEKEKLLRMVNELKDRIEQQKMIHGCSGKEVR
QMNLENSKDKEMVAELQAELDEVSKEMWRMKNDDSFKEQKAKFLQIIENITKLLSGLN
VQLNLDPMPVPADEKQLKVCWETLNTVWVTEISRQMHQRKLDLDTEKSRSADKFAAA
QERIQIENEMLCEAKKKEGRDERTRRAERDEWKAARQQQEKRYDELENEKEVLMKKL
HLDGSLEQEIKEEKDKMAKIEKEAEEKTQYLRSAIRQKVEAVEMEIAEIGQEKTMFHAEC
VAVEKLV
>C_remanei_NDC80
MFGTERRKTGGVNLNGRSSIAITPTKRFTDFGATSVRRTDGRPSLSVGQLGQQSSRPSIFK
TSVHAPRDVKSFQAANAHKIFNFLVESDGADAPAESVIRTPRGKNDFIVIFESMYQHLSK
DYEFPPLNQARIEEEVSSIFKGLGYPFHLKNSYFQPMGASHGWPHLLDALAWLVDFIKIN
KSVSADTQHIIFGDFLEPAKVQEKALSYAWFSTTFRDYTNDRKSAESSDSEFWVETKNKL
RKYFEDSNEYEDMTANAQSALQQLRFDCDEIESERGQEQTYVEDIARLKDDIRKAMDYF
ESVQHLKERKENETAVIKEELEAKVAENEKIQMAVNELKERIEQQKRVHGLNGKEVRKM
NLENSKDKDTVSDLQSEQEMLSKQLWRLRDENPFKDQKMKVIQIAENVTKILSGLNMQ
FELESLRPPENEKELRACWEILIGSWLPEINRQLHQRKLDVETEKSRFHQKFAAAERIQIE
NEVLDEVKKKEAREERIHRDERDEWKEARQKQEKRYDELENEKEMLKRKMQMDGSLE
KEIKAEKEKKEKIKKDLEEMHQGLEAVFRKKLEKIASETAEIGNEKTMFRIEAIEVEKLID
RKCSNQKV
>C_latens_NDC80
MFGTERRKTGGVNLNGRSSIAITPTKRFTDFGTTSVRKRTDGRPSLSVGQLGQQSSRPSIF
KTSVHAPRDVKSFQAANAHKIFNFLVESDGADAPAESVIRTPRGKNDFIVIFESMYQHLS

KDYEFPPLNQARIEEEVQSIFKGLGYPFQLKNSYFQPMGASHGWPHLLDALAWLVDFIKI
NKSVSADTQHIIFGDFLEPAKVQEKALSYAWFSTTFRDYTNDRKSAEYSDSEFWIETKNK
LRKYFEDSNEYEDMTANAQSALQQLRFDCDEIESERGQEQTYVEDIARLKDDIRKAMDY
FESVQHLKERKENEMAVIKDELEAKVAENEKIQMAVNELKERIEQQKRVHGLSGKEVRK
MNLENSKDKDTVSDLQSEQEMLSKQLWRLRDENPFKDQKMKVIQIAENVTKILSGLNM
QFELETLRPPENEKELRACWEILIGSWLPEINRQLHQRKLDVETEKSRFHQKFAAAEERIQ
IETEVLDEIKKKEAREERIRRDERDEWKEARQKQEKRYDELENEKEMLKRKMHMDGSL
EKEIKAEKEKKEKIKKDLEEIHQGLEAVFQKKLEQIASETAEIGNEKTIFRIEAIEVEKLIDR
KCSNQKV
>C_sp51_NDC80
MFGERRKTGGMSLGGRQSIAITPTKRFTDLGIASVRKTDARSSMSVNQPRMSLFQNKGS
SLPPREVKSIQQANVQKIHQFLVESDGDDAPAEVVVRTPRGKKDFTMVFEIIYQHLSKDY
EYPHPVRLEDEIPQIFNGLGYPYPLKNSYFQPLGAGHGWPHLLDALGWLVDVVKMSKA
VVPNTQSIIFGDFLEQSKVQEKTLNYAWTTNTYRDYTSDRKVIENREHPFWEQAKSQLR
NYFESSNEYEDMESNAKNALEQLMFDCEEIESERGQEQHLQEEIAKMKDDIRKAEDYLL
QTENLKSHKEKDLGKIRETLVEKQSELEKIQQLVNELKERIEKQKINHGLSGKEVRKLNL
ENFKDKETVSEVQGELEKLSKEIWTLNNDDRFKEQKSKFVQLTDNIAKLTTGLNIQLSLE
PLKAPTDERQLKTHWETLSNIWLPEINRQLHQRKLELETELSRFADSFAAAEERIQIESET
LCEAKKREAREERIRRNERDEWKTSRQELEKRYDELENEKEVLEKQMQIGGSLDKEIDE
ERARMLKTEQDLRAKEADLMAAIRQKLDGLFAGIVELDNEKMMILKEFNDLEMVMNV
KCANY
>C_sp44_NDC80
MFGNDRRKTGGMNLGGRASIAITPTKRFTDFGTGSVRKTDARLSVNQPRPSLFQKGSSL
PPREVKSIQQANAQKIYQFLVENDGDDAPSEASIRTPSGKKEFGAIFEAIYQHLSKDYEFN
HQARAEDEVTPIFKGLGYPYPLKNSYFQPMGSPHGWPHMLDALGWLVDFVKINKEVNK
DTQNIIFGDFLEQNKVQEKTLNYAWNTNTYREYTLDRKVLDDQSHPFWEQTRIQLREYF
ESSSDYGDMATNIKNALDQLMFDCEEIESERGQEQQLQEEIAKMKDDIRKAEDYLLQTE
NIKSHKEKDLEKTLQCLVERQSELEKVQIAVNELKERIEKQKINHGLSGKEVRKLNMENQ
KDKETVSEIQGELDKLSKKTWTLKNDDCFKEQKSKFVQLTDQISKLMSGLSVQLNLQPL
KAPSDEQELKTHWETLSNIWLPEINRQLHQRKLELDTELSRFGDKFSAAEERIQIESEMLC
EAKKREAREERIRRNERDEWKTSRQEMEKRYDELENEKEVLKKQIQIGGNLDKEIEEEK
TKMVKIEEALRTKEADLIAKLRQKLEEIVVGLAEIDQEKMRIHKEFNDLEIVMNKKCAH
Y
>C_sp48_NDC80
MFGNDRRKTGGMNLGGGGRASIAITPTKRFTDFGASSVRKVDPRLSVNQPRPSLFQKGS
SLAPREVKSLQAANIQKIYQFLVESDGDDAPSEQTIRAPRGKTDFVMIFETIYQHLSKDYE
FPQQTRLEDEVVLIFKGLGYPFPLKSSHFQPMGTGHGWPHLLDALGWLVDFVKISKGVS
TARQNILFGDFLEQDKVQEKALSYAWNTNTYRDFTSDRKVLDNRDHPFWEQTKSQLRT
FFESSNEYEDMANTGKNALEQLMFDCEEIESERGQEQHLQEEITKMKDDIRKAEDYLLQ
TENLKVRKEKDLMTMRETLAEKKAMLEKVLQAVNELKERIERQKIDHGLSGKEVRQLN
LENMKDKETVSEVQGELDALSKLSWTFKNDDCFREQKAKFVKLTVDIAKLMAGLDIQL
NLEPLKAPRDEPELKTHWETLSTVWLPEINRQMHQRKLELETELSRFADKFAAAEERIQI
ETEMLCEAKKREAREERVRRNERDEWKIARQELEKRYDALENEKEVLKRQMQIGGNLD
KEIDEEREKMRKIEEALRAKEAELETEIRLKLDEMFTQIAKIDAEKMMIFKQFTDLEAIM
NANCARY
>C_brenneri_NDC80

MFGNDRRKTGGMNLGGGRASIAITPTKRFTDFGGASSIRKVDPRLSVNQPRPSLFQKGST
LPPREVKSLQAANVQKIYQFMVEADGDDAPSEQTIRAPHGKTDFVMIFETIYQHLSKDY
EFPQQTRLEDEVIQIFKGLGYPFPVKPSYFQPMGVGHGWPHLLDALGWLVDFVKINKAV
NKDTQNIIFGDFLEQQKVQEKTLNYAWNTNTYREYTSDRKVLDNRDHPFWEQTRSQLR
AYFESSNEYEDMASNVKNALEQLMFDCEEIESERGQEQHLQEEITKMKDDIRKAEDYLL
QTENLKAHKEKDLVKVRENLAEKQAVLEKVLQAVGELKERIERQKIDHGLSGKEVRQL
NLENMKDKETVSEVQGELDALSKLSWTFKNDDCFREQKAKFVKLSVDITKLMAGLNIQ
LNLEPLKAPRDEPELKTHWETLSTVWLPEINRQMHQRKLELETELSRFANKFAAAEERIQ
IESETLCEAKKREAREERVRRNERDEWKTARQELEKRYDALENEKEVLKRQMQIGGNLV
KEIDVEREKMAKIEEALRAKEAELQTEIRLKLDEMFAQIAKIDAEKMMIFKQFTDLEAIM
NANCAHY
>C_wallacei_NDC80_paralog1
MYGNERRKTGAPNFSGRASLAITPSKRFTDVDMTSIRRSEVRPSLMGGQSRPSIFGKGST
LPPRDVKSMLYVNVQKIYHFLVEHDQGDAPSEQVIKSPQGKKDFIAIFESIYQHMSKDYE
FPDGKRLEEEIIPLFKALGYPYPLKDSYFKPIGAGHGWPHLLDALGWLVDLVRIMLSVKN
DTQNIMFGDFMEANKVQEKTLSYSWHTNIFREFTNDRKVIESREHPFWEEQKTNLRNVF
DSSNEYGDMTNNYKNALEQLKFDCDEIESERGQETNLQEEIARMSDDIRKANDYREQTE
QVRTLRETDFQKIQEALVEKKEENIRMQNTVNELKERIEQQKINHGLSGKQVREMNLEN
SKDKEHVSEIQGELDKLSKEMWTLNNENCFREQKAKFTEIIDHIQKTINGLDISIQLPHLK
PPTDEKQLKSGWETLTNNWIPEINRHLHQKKLELETEQSRFADRFAATQERIQIEIEMLQE
SKKRETREDRIRQNERSEWKSARQELEKRYDEQENRLEVLKKQLKMDGSLEKEIEEEKE
KMRKIEEEMTAKQEELVRILNGRLDEIVAMTAKIENEKLGIHKEFEEIERVILANQ
>C_wallacei_NDC80_paralog2
MYGNERRKTNFDGRPSNAITPNKRHTEGISSLRRSEVRPSIGGHPRQSIFGRASTIPSRDV
KALAHGNIDKIYNFFVEKGEEPPCKSIIQSPTGKDFAVIFETIYQHLSEGFVCSYQHPEEDV
IPIFKALNYPYTLKDSYFKPVGTSHGWPHLLDALGWLIDLVNISLSVNNNAQNMLFGDIL
ERNQVEEKTLSYAWYTSIYRDYTNDRKVIENRNIPFWEKQRDALRNAFNGFDGNMDVI
GRMKNELEQLKFDCDEIESETGQENQLKEEISRIKDDIRKADEYFEQTEQLREIREKNFQK
ILEELQEKQKENERIQKKVNEVKERIEEQIINHGIDGKQVREMNLENSKDKELVSEMRGE
LDKLSKEMWTLKNEDLFKEQKGKFQEVIDHIYKTINGLEIQITLTPMKSPSDERELKIGW
DTLTNIWIPEINRHLHQRKLELETEQSRFSDKFAASQERIQIELEMLEEAKKREAREERQR
QNERDEWKVSRQNLEKHYDELENEMEVLKKKLKMDGSLEKEIEEERVKIMKIEEELSSK
HNELINEITRKLDEIFAATAEIENEKLLIQKEVNELENAMKSVGLQ
>C_tropicalis_NDC80
MYGSERRKTGAMNLGGRTSMAITPSRRFTDVESSSVRRSEVRPSLMGSQARPSIFGKGST
IPPRDVKSMIGANVEKIWHFLVEHDQGDAPSDSVIRNPQGKKDFVAIFESIYQHMSKDYE
FPNGGRLEEEITPLFKALGYPYPLKDSYFKPIGATHGWPHLLDALGWLVDLVRIMLSVKD
DTQNIMFGDFPEATEVHERALSYSWHTSIYREYTNDRKVIESKGHPFWEKQRAELRNVF
DSSNEYGDMTSNYKNALEQLKFECDEIESERGQEANLLDEIARMKDDIRKAAAYQEQTE
NVKMLREADLEKLHAVLREKMEENQKMQKAVNELKERIEKQRVDHGLNGKQVRQMN
LENSKDKEMVSEIQAELDKVSKETWMLNNENGFRDQKLKFLEIHEHIIKTIHGLNLPVNL
PPLKPPSDELELKSGWETLTTVWIPDITRHLHQKKLELENEQSRFADRFAAAEERLQIENE
MLEEAKKREQREERTRQNERSEWKTARQRLEKESDELENQLEVLKKQMQIGGNLEMEI
EEERRKMKAIEEGMERKNEELIACLQKKLDEIFAATVEIENEKLAMQKEFEELEAVMQT
NCH
>C_doughertyi_NDC80

MFGGERRKTGGMSLTGRASLAITPTKRFTDFGISSSRTNPRPSLNIGPRQSLFGSKSASVAP
RDVKSIQQGNVQKIYRFFVETDGDDAPPESIIRAPGGKADFITIFEQIYQHLSKDYEFPQG
ARLEDEIAQIFKGLGYPFPLKKSFFQPMGCAHGWPHLLDALGWLVDLVRISIRVSSDTQN
ILFGDFMEESKLTSKFLNYAWHTKVYKDYTNDRKVIDNPHSQFWLEAKSQLRDHFETAN
EYDDWMDTAKNNLEQLKFDCDEIESERGQEQHLLDEIAKMKDDIRKADDYLQQTANLK
VHKETALQEVQQTLLEKQGELEKIQKVVNELKERIEQQKVKHGLSGKEVRQMNLENQK
DKETVSEIQGVLDQLSKEAWSLKNDDSFKEQKAKFVQLAEHITKIMSGLDIEINLARLSV
PSDERQLKANWETLTNVWMPEVNRQLHQKKLDLETEISRFTDKFAAAEERNKIESEMLC
EAKKKDAREERLYLNERKEWQEARQQLQKRCDELENEKEVLKKEMKMDGSLEKEIEE
EKTKMAKEVKELRAKQTELEAVLRQKLDQIVAEVTDIQNEKLMVYKEFNDLGMIFESKL
PNY
>C_sp54_NDC80_paralog1
MFGADRRKTGGGMSLGPRQSVAITPTKRFTDYGPSSMRKPDGRASLSHNIGQPRPSLFGT
KNSAVAPRDVKGLVTANVSKIYSFLIEFEGSEAPSEKDIKSPRGKNDFTFVFELIYQHLSKD
YEYPQSARIEDEIPQIFKGLQYPVTLKNSYFQPMGSSHGYPHLLDALAWLVDLVRINSCV
SADTQNIIFGDFMEQAIVQEKTLNYAWMSTTFRDYTNDRKAAENPETPFWDETKCRLRK
YFEESNEFDDMAKTASNGLEQLRFECEEIESERGNEQSLLEEIAKIKDDIRKATEFFESTEQ
VKELKEKELARVKLELEAKMAEHERAQRMVNELKARIEMQKTNHGLSGKEVRQLNIEN
NRDKETVAEIQAELDSLSKEAWRLNKDDSFKEQKAKFVQLYQKTMKIIAGLDIQLNLDPI
KAPADERQLKTCWEALNGIWLPEINRQLHQKKLELETKNSRFADNFAHTEQRIQMESEM
LCEAKKKEAREERIRRNERDEWKEARQQMEKRYDELENEKEVLKKQMQMDGSLEKEI
EAERVKMTKIEKELEEKKKYLETSIRQKLDQIVAETAEIENEKIMFHVECTELEKTVKAM
CLN
>C_sp54_NDC80_paralog2
MNSTFILNHFSLLRHVKKMFSGGPRDVKSLVTANVKKIYSFLIEFEGSEAPSEKDIKSPRG
MNDFISIFESIFHHLNKDYKYPQSARVEEEISQIFKGLQYPFPLKVSFFQPFGSSHGYPHLL
DALAWLVDLVRMNSCTNDTNEKEMLKMQMTKIEKELEEKKKNLETSIRQKLDQIVAET
AEIENAKMMLLEMKCTELEKTVKAMCLN
>C_inopinata_NDC80
MYGDRRRTGGISLNGRASLAITPTKRYTDFRTETRPSLGQSSGQHRLSIFHKTPSVAPRDV
KSIFTANVSKVHNFLAEHEGLEAPLESTIRNPRGRNDFICCFELIYQHLSKDYEFRKDERIE
DEISQIFKGLGYPFALKNSYFQPMGSSHGYPHLLDALAWLVDIIRVNAAVSADTQNIIFGY
FMEQSKAQERTLHYAWMSSIFREYTNERKVAENSESPFWGEVKNKLRKFFEDSGEFEDM
TNSAASALEMLRFECDEIESEKDNEQSILEEITRIKDDIRKASAYLETTQQVLQHKEIELAG
VQSEVADKIAENEKVQKMVNDLKTRIENQKNHHGLTGKEVRELNLENNRDKEMVSEIQ
AELNNLSKETWRLKDEDHFKEQRSKFVQLFDSVTKMIAGLNIQMNLKQISSPNNERELK
VCWETLDNVWLPEINRQLHQRKLDLETEQSRFSSKFVAAAERIQMKSEILSEVKKNEIRE
ERLRQNERDEWKENRKQKEKRYDELENEKEVLKTQMQMDGSLEKEIEAERISMAKIEK
ELEEKCKYLQSVIKKKLDQIVTETTAIESQKTIFHLQCTEIEKAVN
>C_elegans_NDC80
MFGDRRKTGGLNLNGRASIAITPTKRFTDYTGSTSVRKTDARPSLSQPRVSLFNTKNSSV
APRDVKSLVSLNGSKIYNFLVEYESSDAPSEQLIMKPRGKNDFIACFELIYQHLSKDYEFP
RHERIEEEVSQIFKGLGYPYPLKNSYYQPMGSSHGYPHLLDALSWLIDIIRINSAVSEDTQ
NILFGDFMEQGKAQEKTLNYAWMTSTFRDYTNDRKAAENPSSSYWDDTKHRLRKYFE
QSNEFEDMTKTAASALEMLNYECDEIEADKGNEASLKEEISRIRDDIRKAKDYLEQNLH
VKQHMEKELAMVKSEQEEKISENEKVQKMVDDLKNKIELQKQIHGLTGKEVRQMNLD

NNKDKEVVLEIQSELDRLSKETWKLKDEDFFKEQKSKFIHLAEQIMKILSGLNIQMNLEP
LRAPTNERDLKDYWETLNKIWVPEISRQLHQRKLELETEQSRFSNKAVTAEERIQIQSETL
CEAKKNEAREERIRRNERDSWKDARKHIEQRYEQLLNEKEVLLKQMKLDGSLEKEIEDE
TARMSATGEEHIQKRSQLEAGIRQILDLMVVEIAEIENKKIGFHVQCAGIEKAVL
>C_oiwi_NDC80
MFGDRRKTGGLNLNGPRASLAITPTKRFTDLGVNSVRKSTTRMSLSQTSGQPRMSLFQS
KSIAPRDVKALVSSNVTKIYNFLVEHEATEAPSEKDIRSPRGKNDFVLFFELIYGHLSKDY
EFPQDTRLEDEIPLIFKGLGYPYPLKNSYFQPMGSSHGYPHLLDALSWLIDLAKISSSVRS
DTKNIIFGDFMDQNKALEKILSYSWLSTVYKDFTNDRKAAETDQFWVETKGTLRKHFE
HNNEYDDMIETAKSALEQLRFECDEIESDKGNEQSLQEEIARIRDDIRKASDYLEQTEML
KERTQEELNEINTQLQVKSAENEKVQGMVNELKARLEQQKLHQGLTGKEVRQMNLEN
NKDKETVAELQAQLEELSKQAWKLKNDESFKEQKTKFVHLVENALKVLSGIGINTNSDA
FRVPTNEKELQASWETMNNVWLPEINRQLHQRKLELETEKLQFADKFASAEERIQMETE
MLREAKKKEVREDRLRRTEREEWKESRQQLEKRLDELENEKEVLKKQLQIDGSLEKEIE
TEKVRMAKIEKEMEAQKERLEAAIRKKMDEIVEETAGIENSKIMFHKSCVEVEKTVKTA
LKVSKK
>C_kamaaina_NDC80
MFGDRRKTGGLNLNGPRASLAITPTKRFTDLGVNSMRKSTARVSLSQTSGQPRMSLFQS
KSTAPRDVKALVSSNVTKIYNFLVEHEATEAPSEKDIRSPRGKNDFVVFFELIYGHLSKDY
EFPQDTRLEDEVKKIFKGLGYPYPLKNSYFQPMGSSHGYPHLLDALSWLIDLAKISSSVR
SDTKNIIFGDFMDQNKALEKILSYSWLSTVYKDFTNDRKAAETDQFWVETKETLRKHFE
HNNEYDDMIETAKSALEQLRFECDEIESEKGNEQSLQEEIARIRDDIRKASDYLEQTEML
KERTQEELNEIKTQLDTKSVESEKMQGMVNELKARIEQQKLHQGLTGKEVRQMNLENN
KDKETVAELQAQLEELSKQAWKLKNDESFKEQKTKFVHLVENALKVLSGIGINTNSDAF
KVPANEKELQASWEAINNVWLPEINRQLHQRKLELETEKLQFADKFASAEERIQMETEM
LREAKKKEMREDRLRRTEREEWKESRQQLEKRLDELENEKEVLKKQLQIDGSLEKEIET
ERVRMAKIEKEMEAQKQRLEAAIRKKMDEIVEETSGIENSKMMFHKSCVEVEKTVKAL
KVSNNQ
>C_waitukubuli_NDC80
MFGDRRKTGGPGFGPRLSTAITPTKRFTDVGPTSVRKADSRLSMGQGGGGQPRASLFQRS
SAVASRDVKSLMAANVNKIYNFFIESEGSEAPSEAQIRNPKGRNDFITFFELLYQHLSCDY
EFPQGARVEDEVSALFKNLGYPTALKNSFFQPMGSSHGYPHLIDALAWLVDLIRVNMCV
SNDTQNIIFGDFLEQSKVQERTLSYCWMSSTYKEYTNDRKAAETRDGPFWDSTKIKLRK
YFEDMNETGDMAASAKTSLEQLRFECEEIEADKGTEQNLIEEIAKIKDDFRKASDYCERL
EAVERQREVDLDQVKMEVDVKKKEMSEVQEEVNSLKQRIEEQKQNHGLTGKQVRQLN
SDNSRDKAIVSEIQSELERLSKELWLRKNEENFKEKKTMFLQLAEKLAKIGAGINVDLGL
TSLRSPQNEELKIECEKLVGVWLPEVTRHLNQKKLELEMEKSRFNDKFASAACANMEQE
TLCEAQKKAAREERIRRNEREEWTESRLAKEKRYDELENELDVLKKNMQMGDSLVKEI
EMERAKKVKLCEEQEKQKKTIEASIRQKMDQIMTETAEIDNEKMMIHVECNELEKMV
>C_panamensis_NDC80
MFGDRRKTGGPGFGPRLSTAITPTKRFTDVGAAASVRKVDNRLSMGGQHRASLFQRSSA
IAPRDVKSLMSANVSKIYNFFLEFEGPDAPSESQIKNPKGRNDFISFFELLYQHLSCDYEFP
ANVRVEDEVQISTIFKNLAYPTPLKNSFFQPMGSSHGYPHLIDALAWLVDLIRNKCVSDD
TQNIIFGDFLEPAKVQEKTLSYAWMSSTYREYTNDRKAAETKDGPFWEVTKNKLRKYFE
DMSETQDLAASVKNSLEQIRFECEEIEADKGTEQSLIEEIAKIKDDLRKATDYAEKLEAVE
RQRLMDLDKVKADLDVKKKENSEVQTEVDGLKQRIEEQKQNHGLTGKEVRQLNSDNA

RDKAVSNELQSEMEKVAKDLWRNNNEENFKEQKRTSFSRLVERIEKILAGINVTLRLEPL
RTPQTERDLKAGLDELMSVWLPEITRQLNQKKLELNMEESRFTDKFVAVQARVQLEKET
LDEAKKKEAREERMRRNEREEWMESRLATEKRYDEMENEQDVLKKQLQMDGSLEREI
EMEKARKVKISEEQENKKNKIDARIRQKMDEVMKETAEIDNGNLLFHVACNELEKMVL
KTQNLI
>C_nouraguensis_NDC80
MFGERRKTGGQGFGPRISTAITPTKRFTDVGAAASVRRTDSRLSMGQSGGQHRASLFQR
NSAAAPRDVKAMLGPNVNKIYNFFIEFEGSDAPAEAQIKCPKGRNDFINFFELLYQHLSC
DYEFPPNARVEDEVLWIQIFKCIGYPVSLKNSYFQPMGSSHGYPHLVDALAWLVDFIRNK
CVSDDTQNIIFGDFLEPAKVQEKTLSYAWMSSIFREYTNDRKAAETRDGPFWTITKERLR
DYFEKMNEAQGLAVSLKNALEQIRFECEEIEADKGQGQTMLEEINKMKDDMRKAVDYN
EALQTAQRQREIDLEKAMTDLEVRMKEDSEVQADVNELKKRIEEQKEKHGLTGKEVRQ
LNLENTRDKAVISEIQAELDKLSKELWLKKNEENFKDKRASFVRLAENIGKIVAEVYIDL
GLEPLRSPQNEKELKSELEKMNSVWLPEVTRQLRQKQLNLNKEKSFFRDKERVQIERDT
LCEAEKKMAREERIRRTERDEWKEIRLANEKRYDELENEQDVLKKQMQKDGSLDKEIE
LEMARKLKIDEESSKKKSALEAVIRRKLDHIMAETSKIDNEKLMFHTDCTEFEKQILKTR
NMK
>C_becei_NDC80
MFGERRKTGGPGFGPRISTAITPTKRFTDVGAAASVRKTDSRLSMGQPGGQHRASLFQR
NSAAAPRDVKSMLGANVSKIYNFFIEFEGSDAPAEAQIRNPKGRNDFINFFELLYQHLSC
DYEFPPNARVEDEVIHIFKCIGYPVSLKNSYFQPMGSSHGYPHLVDALAWLVDFIRNKCV
SDDTQNIIFGDFLEPAKVQEKTLSYAWMSSIFREYTNDRKAAETRDGPFWTVTKEKLKDF
FEKMNEAQGLAASLKNALEQIRFECEEIEADEEINKMKDDMRKAVDYNDALQTAQRQR
EIDLEKAMTDLEIRMKENSEVQADVNELKQQIEEQKEKHGLTGKEVRQLNLENTRDKAV
ISEIQAELDKLSKELWLKKNEENFKDKRAIFVQLAEKIGKIVAEVHIDLGLESLRSPQNER
ELKSEQEKMNSVWLPEVTRQLRQKQLNLNKEKSFFRDKFAAAEERVQIKKVTLCEDEK
EMAREERLRRNEREEWKEIRLVNEKRYDELENEQDVLKKQMQRDGSLDKEIELEMARK
LKIDEESSKKKSALEAKIRQKLDQILVETSKINNEKLMFHSDCQEFEKQILKTRNMK
>C_yunquensis_NDC80
MFGDRRKTGGPGFGQRNSTAITPTKRFTDVGAASSVRKTDNRLSMGQTGGQHRASLFQ
RNSAAAPRDVKSMMGANVSKIYNFFLEFEGPADAPSEGNIRSPKGRNDFINFFELLYQHL
SKDYEFPANARVEDEIIHIFKFLGYPISLKNSYFQPMGSSHGYPHLVDALAWLVDFIRCSIS
VSGDTQTIYYGDFLDQAKVQERTLSYAWMTTIFREYTNDRKAAETRDGPFWTMTKEKL
RTYFEKMNEDQDLTASVKNALDQIRFECEEIEADKGKGQEISEEIVKVKDDIVKAIDYGE
ALEAAQKQQESDLEKATAELRERVDENSEVQANVNELKQRIEDQKKKHGLTGKEVRQL
NSENTRDKAMITEIQAELDKLSKELWLKKNEENFKDKRESFVQLVERIGKIVAEVSIDLG
LEPLRSPRNERELKTELEKLNGIWLPEVTRQLRQKQLNLDKEKAFFRDKFAAAEERVQIE
RDTLCEAEKKMAREERVRRHERDEWKESRLVIEKRYDELENEQDVLKKQMQMDGSLD
KEIELEKAKKVKIDEEIAKKKAALEAAMRQKLDKIMVETAKIDNEKILFHVECTEFEKQI
LKTRNLN
>C_macrosperma_NDC80
MFGDRRKTGGPGFGPRISTAITPTKRFTDVGAAASARKVDSRLSMGQPGGQHRASLFQR
SSAVAPRDVKSMLAANVNKIYNFLVEYEGTDAPSEAQIRCPKGKNDFINFFELLYQHLSC
DYEFPANGRLEEEISNTFRHLGYLPHLKNSYFQPMGSSHGYPHLVDALAWMVDIIRCNK
CVSDDTQNIIFGDFLEPAKVQEKTLSYAWMSSTFREYTNDRKAAEAINGPFWNSTKNKL
RKFFEDMNDTQDLTASAKNALEQLRFECEEIEADKGTEQSLLEEITRIKDDIRKAICHYED

TKRVLNQRESDLEKVKAELDTRIKENNEVQAEVNLLKNRIEEQKEKHGLTGKEVRQLNS
NNTRDKAVVSEIQTELDKLSKELWLMKSEDTFKEKKTCFVHLTERIGKIVAGIEIDLGLEP
LRSPQNERELKVEWEKLNNVWLPEITRQLHQKKLELEMEKSRFSDKFAAIEERVQMERE
TLCEAKKKESREERLRRNEREEWKESRLLKEKRYDELENELDVLKRQMQMDGSLDREI
EIEKAKKVKNDEELKKKKAAIEASIRQKLDQIVAETAKIENEKTLFHVECIEFEKLILNTQ
NLV
>C_sulstoni_NDC80
MYGGDRRKTGGFGFGNPRTSAITPTKRFTDLGTTSSVRKDHGRLSMSQTSGHHRVSLFG
GASATGPKDVRANFDANAKKIYNFLVEHEGSEAPSENVIRCPAGKSDFVTVFELMYQHL
SKDYEFPQNVRVEEEVSNIFKALGYAPPLKNSYFQPMGSAMGYPHLVEALAWLVDLIRIN
ASVSADTQPIIFGDVMDQEKVHEKTLSYAWMSTTFRDYTNDRKAAESGTGPFWEETKS
KLRDYFEKSDEIEDVATNYKSALEQLRFECEEIEAERGNELTLLEEIAKIKDDVRKALEYD
ETTARVQKHKEDELKAVKGTLEQKMAENSKVQAEVAELKERIEVQKQMHGLTGKEVRQ
LNSDNNRDKETVTELQAELDEISKQMWRLKNEDTFKQQKTNFVRLVESVEKIVSGLDIE
MGLEPLRAPSNESELKSRWDELNSVWIPEINRQLHQKKLEFETEQSRFADRFAAAERVQ
MERELLCEAKKKEAREDRVRRNERDEWKENRLQREKKYDELENEKDVLTKKMQMDG
TLDKEMEEERTKAWKLEQELQKEADAVENAIRKNLDLIVNEVAQIENDKMLFHVECCEF
EKLIVKTQNMF
>C_afra_NDC80
MYGGDRRKTGGFGFGNPRTSAITPTKRFTDLGTTSSVRRVSLFGGGSAAGPKDVKGNFE
ANVKKIHSFLIEHEGNEAPADTVIRSPAGRNDFLTVFELMYQHLSNNYEFQQGVRIEDEV
SNLINASVGADTQTILFGDVMDQAKVQEMTLRYSWMSTTFRDYTIDRRAAESGTGPFW
DETKNKLREYFEKSDEIEDLVTSYKTALEQSAFECQEIEADKGNEQHLLEEISKMKDDVR
KAMEYAESTARVQKHKEEEMKTVKATLETKIAENSKVQAEVAELKERIEVQKQLHGLTG
KEVRQLNSDNNRDRETVTELQAELDEVSRQMWRLKSEDTFKQQKANFIRLVESVEKIVS
GLELQLGLDAILAPSDESELKVRWDELNNVWIPEVNRRLHQKKLEFETEQSIFVDKFAAA
ERVQMERELLCEAKKREAREDRVRRNERDEWKENRLQREKKYDELENEKYVLKKKMQ
MDGSLDKEVEEEQAKARKIEEELQKEADAIENSIRKNLDQIVSETAQIENDKMLFHVECC
EFEKLI
>C_sp49_NDC80
MMFGGPRRKTGGPNFGATGRTSTAITPTKRTTDLGAAHSVRKNDNRMSFGQGPSNSRPS
LFQRTGGPSKDFKAAVPGNAAKVYNFFVDTEGANAPSEKTIRTPAGKGDFVTMFELLYQ
HLSKDYEFPSNGVRLEDEFTNIMKALGYPHALKPSFFQPIGSSHGYPHLLDALAWLVEAV
EINEKVRLDTQNIIMGDFMEPDQVQDKVLSYSWLSKTFFEYSNNRKVLEDKADPFWATT
KCDLRNFFENNDESADMITSSKNMLEQLRFDCEEIEADKGNEQGLLDEIARIRDDIRKAQ
LYLDETMRVVAQSNHEYQTVVDACEAKLAELEKVKTEVAELKERIEEQKRLHGLSGKE
VRQMNAENSKDKEMLSDVNAEIERATKMMWRLRDAADYKEPSTRFYRLISHMHQILD
DADVHPQLEPIASPSSELDLKAGWDAINNLWLPEVNRQLQHKKLELETENARFLNKFEA
IEERATMEQEMLNEATKKEDREERVRRNEREEWKAARFQKEKRCDQLENEKDILVKQM
HLDGSLDRDLQEETEKAAKLSETLEQKQAAFEQKVREKIDEMMASTAEMEQEKLEFHK
ECVELEMVVKNS
>C_sp25_NDC80
MFGGPRRKTGGPGFSTTGRTSTAITPTKRNTDLGTAQSVRKNDGRMSFGANPSGSRPSLF
QRTVAPSKDVKAAVPANVARIYNFFVETEGPNAPSEKMIRSPAGKGDLITMFELIYQHISK
DYEFPSVGVRLEDEFTNIMKALGYPNALKQSFFQPIGSSHGYPHLLDALAWLVEAVEINE
KIRLVTQNILVGDFMEPDQVQDKVLSYSWFSKTFLEYTNNRKVLEDRSDPFWANTKREL

RDFFERNDESADMIATSKNMLEQLMYDCEEIEADKGNEQGLIDEIARIRDDVRKAQAYL
EETSRVVAQSIHEFQSVSDAFEAKRTEFEAVKTEIAELKERIEEQKRNHGLCGKEVRRLNL
ENIKDKETVAELQAELEKTSKIIWNLRDAGSFKEQKARFDRLITHMKMILDDTDIHFQLE
SITSPRTELELKAGWDTVNNVWLPEVNRQLQRKLLELETENARFASKFGAKEEHVTMEQ
EMLNEATKKEDREERVRRNEREEWKAARFQKEKRFDQLENEKDILMKQLHLDGSLDRE
LEEETEKSIKLTEALEQKKAEFEQKVREKIDLMMVSTAEMEQERLDVHKEFVELETIINSS
>C_imperialis_NDC80_paralog1
MFGGPRRKTGGPNFSATGRTSTAITPTKRNTDLGAVQSVRKTDTRMSFGQGPSAPRASLF
QRAGLPSRDVKACFASNVSRIYNFFVETEGSSAPSEKTIRSPNGKGDFISMFELIYQHLSK
DYEFPTQNRLEEEFSNILKSLGYPYPLKNSFFQPIGSSHGYPHLVDALAWLVEVCEVNEK
VRLATQNILLGDFMEPEQMKDKFISYSWFSKVFLEFTNQRKAAEDKTDPFWESTRRELR
EFFEQNDESVEMIASLKNMLEQLRFDCEEVEEDKGNEQSLLDEISRLRDDVRKAQSYLE
ETARVVAQSHHEFATVKEAVDAKMEEFERVKAEVAELKARIEQQKELHGLTGKEVRLMN
VENNKDKETVAEIQNELDRTSKIIWRLRDAGSFKEQKSRFERLVEHMLKIVDDADIHLTL
DPINSPSTELELRAGWEAVNTVWLPEVNRLLQRKKLELEGENARFSSKFGAVEEKAILEK
ELLNEATKQEAREERVRRNERDEWKESRLQKERRCDQLENEKDVLVKQLHLDGSLDVEI
REKADEMEKLVVELAGKKKTLEGEVRKRIDEMMESTVQMEHEKLLFHRECVELEALVK
SQ
>C_imperialis_NDC80_paralog2
MFGNARRKTGGPNFATGRPSTAITPTKRITDLGAAQSVRKNDSRMSFGPSAHRTSLFQRA
GPHSKDMKACFAANVAKIYHFFVETEGSEAPPEKMIRSPTKNEFITIFELVYQHLSKDYEF
HMQNRLEEEFTSILKALGYPYPLKNSFFQPIGSSHGYPNLVDALAWLVEVVEVNSAVSRV
TQNILIGDFMEPELAEDKVLSYSWFSKTFLEFTNNRKALEDKSDPFWESTRLGLREYFER
NDETAEMIASCKNTLEQMRFDCEEIEEDKGNEQSLLDEISRLRDDVRKAEAYLLETSRAL
EQATQEHESVKEAVETKGAEFEQVKMEVARVKELIEQQKEQHGLTGKEVRAMNVECQR
DKETVSEIQMELDKVSKVIWRLRDAGSFKEQKSRFDRLIEHMHTILVDADIHLELEPICSP
STEIELRAGWEAINGVWLPEVNRQLQHKKLELEAENARFSSKFGVVEEKAILEQELLNE
AIKQEARDERVRRNEREEWKASRLQKEQRCDELENEQDVLVKQLNLDGNLDAEIREAA
QKAERLFLELQAKKEQFEGAVRAKIDEMMESTVAMEQEKLMLHKECSELESSVKSTL
>C_japonica_NDC80
MFGERRKTGGANFGFAGGRPSTAITPTKRFTDLGTAQSTRRVDNRLSMGQNGGRLSVFQ
RGSAVAPRDVKSAQASNVTKIYNFLLEHEQSGAPPEKIIRQPTGKNDFITMFEQLYQHLSK
DYEFPQGARVEDEFTGIMKGLGYPFPLKNSFFQPMGSSHGYPHLLDALAWLIDVISLNQS
VSSVTQNILLGDFMEAAEAQDKIISYSFYSSTFREYTYDRKAIESKDAPFWAETKERLRD
YFEKNDEITDLVTSTKQVFEQLRFECEEIEADKGDEQGLLEEIARIKDDVRKAHEYLESTE
RVQKQTGVELGIVQESLEAKKAEFEAVQAEVNELKRRIEVQRQQHGLTGKEVRQLNVE
NNRDKEAVHEIQTELDNVSKTLWRMRDEDTFREQKANFVLVIENMEKILLDANVKIGLD
ALRPPQNERDLKVGWDALNNQWLPEVNRQLQHKKLELDDEKTQFSSRFAAIEERVQMQ
QELLREANKKEAREERVRRNERDEWKTDRLQREKRLDELENEKDVLKNQMQTGGSLD
REIEEEKAKSAKLVEALEQKKASLAEGIRQKLDETMREITKIDLENATFHAECTDIEKLVL
KTRNMNQ

Supplementary Data 3.S11: All HIM-10 sequences used in this study

>C_tribulationis_HIM10
MNNAKQVILVNHDPKSIITKLNSKLHLGLTPENVTHPAETALTIFMSFVRYVLNVSDHSLT
TLPLSASGEHDPELHKRSVQLTLVYQSMKAFITDNSGQKLDMRMCDLVTPAKEPLRFRK
LLAFLVEFIKLHEISAPIFNEISDEFSEQKHAMESLQEDIATAERKKNELLSKQSLRKRREN
ELMDDHSKIKNELNGIVNQYNDNLAMTNDMEKQKAELIQQIEDIEREILTAKKTVEHLTE
EVLESPEELKKEMKERKRRIEEFREALAASRQTLKSKVEAREICANSEKNLPVIQQRIQS
WKEVREDILDLMDEIQEKLRKLDEMQEQLAFTTDKKNNSGKRMIEQAEMHEQLRKEHL
QRSENLNKNIEEIRQIASLGKNQPDVSRDIEKKRQELLAAKNAHSELMSRIMNSTKDAFV
KFRKIDAHFKATQNVALEKHCAMDRAKNRLSNSYKSRLPSDYTFSASSINESDAENLDP
QSPVFENFSVFKN
>C_sp41_HIM10
MNNQKQVIVMNYDARSIITKLNPKLHLGLSPDNITNPAETSLTIFMSFVRYVLNVSDHSLT
TLPLSASGEHDPELHKRSVQLTLVYQCMKAFIKDNSGQKMDMRMCDLVTPGKEPARFK
KLLCFLVDFIKLHEMAAPIFNEISDEFSEQKHAMEALQEDIAIAERKKNELLSKQNLRKRR
ENELMDDHSKIKNELNGVVNQYNDNLVITNEMEKQKVELIQQIEDIEREIMTAKKTVEH
LTEEVLESPEELKREMKERKKQIEEFRDSLAASRQTLKTKLEAREICANSEKNLPVIQQRI
QSWKEVREDILDLMDEIEEKLRKLNEMQEQLAFTTDKKNSSGKRMIERAEMHEQLKRE
HLQRSEDLNKNIEEILGPHSERISQMCREILRSRDKNSSPPKTPTANECLAS*TRRRTRSPN
FEKSTLASRRPNAWQSRKGVQSIVPRIASATRSNPAFPATTRSVLPVSTTRNRRILTHSRRR
SRTFLYLTI
>C_zanzibari_HIM10
MNNGKQVILFNHDAKTIVTKLNAKLHLGLSTDNIINPTAETALTVFMSFVRYILNVADQS
LTTLPLSASSEHDPEFHKRSVQLTIVYQSMKAFITDNSGQKLDMRMCDLVTPAKDATRFK
RLLSFLVEFIKLHEMASPIFNEISEDFSEQKNMVDRLQEDIADAERRKNELLSRQSLRKRR
ENELMDDHSKIKNELNGVVNQYTENLAITEDIEKQKVELFQQIEDIEREIMTAKKTADHL
TEEVLESPEELKNEMRERKKQIEEFRESLAASRNTLRLKLEARDICANSEKNLPVIEQRIE
AWKEVREDILDLMDEIEEKFRKLNEMQEQLAFTADKKETSGKRMIEQAEMHEQLKKEH
LQRSEKLNKNIEEIVQIASLGKNQPDVSRDIEKKQRQELLAAKNAHSERMSRIMNSTKDA
FAKFRKIDAHFKETQRVAMEKQCAMDRAKNRLANSFKSRLPSDYTFSASTINEQDSENF
DPESPVFENFSVFKN
>C_sinica_HIM10
MNNGKQVILVPNDVRTIVAVLNKKLRLGLSQDSINIPTETAFLVYMNFVRVVLDVNDIHL
ATLPMSANTDHDPELHRKSVPLIIVFQCMKAFISDNSGQKLDMRMCDFVIPGRDIPRFKK
LMSFLVEFIKLHEMAAPVFNEISGEFSDQKQEIEALQREINDTEKKKTELLSRQSLRKRRE
NELMDDHSKIKNELNGIVCQYSDNLAATEEMEKQKAELIQQIEDINREILTTKKTAEHLSE
EVLESPEELKREMTERKKQIEELRECLAASRQTLKAKLDARDICANSEKNVPVMQQRLE
TWNTVREDILTLMDVIEAELRKLAEMEEQLAFTTDKKNNSGKRMIEQAEMHEQLKREH
LQRNEKLSKNIEEIQIAALGKNQPEVSRDIEKKRQELLAAKNAHSERMSQIVRSTKEACG
KFQKIDAHFKDLHRVAQEKRCAMDRAKNRLCNSFKSRLPSDYTFSASSINEEASENCDP
QSPVFESFSVFKN
>C_nigoni_HIM10
MSNAKSVILIQYDTRLIAKVLQQKLGLGLTADSILNPDVSSFPSIAEVAQSIFSNFIRYILNV
SEQSLLTLPLSADSGHDPELHKKTIPLVIIYQCMKAFIVDNSDRKLELTMCDLVTPEKQPN

RFRRLTSFLVDFIKLHEMSAPIFNEISDEFSEQKQEMERMQDEIIQAEKRKNDLISKQSLRK
RRENELMNEHSKIKSELAGVVSQYTDVLERTEEIEKQEKELIQQIEEIEREIMTAKKTVEH
LTEEVLASPEELKQEMANRKKQIEELKESLAASRQTLKSKLEARNICANAEKNVPLINQR
IQAWSEVREDILDMMDEVEEKLRKLNEIQEQLAFAADKKTSSEKRMIEQTEMHEQLRRE
HLQRSAALGTSQPDVSRDIENKQRQELLAVKNAHSEKMALINNAQKDALAKHRRIDDH
FKETLRVAVEKRNAMERIKNRVSNAYTGRLPSDYTFSASSINESENCDPQSPVFENFSVFN
N
>C_briggsae_HIM10
MSNAKSVILIQYDTRLIAKVLQQKLGLGLTADSILNPDAEVAQSIFSNFIRYILNVSEQSLL
TLPLSADSGHDPELHKKTIPLVIIYQCRLTSFLVDFIKLHEMSAPMFNEISDEFSEQKQEME
RMQDEIIQAEKRKNDLISKQSLRKRRENELMNEHSKCKSELAGVVNQYTDVVERTEDIE
KQKTELIQQIEEIEREIMTAKKTVEHLTEEVLASPEELKQEMANRKKQIEELKESLVVSRQ
TLKSKLEARNICANAEKNVPLINQRIQAWSEVREDILDMMDEVEEKLRKLNEIQEQLAFA
ADKTTSSEKRMIEQTEMHEQLRKEHLQRSEKLYKNIDEIARQIAALGTSQPDVSRDIENK
RQELLAAKNAHSEKMALINNAQKDALAKHRKIDDCFKETQRVAVEKRNAMERIKNRVS
NSYTGRLPSDYTFSASSINESENCDPQSPLSNLGPRSLKHVQSAASSSSSVVLLHKSYTES
SAPPSTNRKLPYKTLRLASTQQVTEFIQNGELPTFKKQGSRSTNNPWRRHGALQKLCET
GGFAPNSGSKKERKNRGSLLAVKQNVKVPPAPTSSQMTLNKVKEQGQEIKLALKAPKN
VAVDVAAQNLAEQVKAFVLRYGSELVEDERTAQSK
>C_remanei_HIM10
MSNAKSVVLVMFDPRKISTTLNQKLQVGVTPDNILNPTAEIVQQIYLNFVRVVINISENSL
HTLPLNADSDFDQELHKKSIPLAIVYQSMKAFIKDNSGGKLDLTMCDLVTPGKNPQRFR
KLSSFLADFIKLDEIAAPIFNEISEEFSDQKVEMEALQEEIVAAEKRKDELVARQSQRRRRE
NELMDDHNKKKSELAGIINQYTEIGVKTEELEKQKNELIRQIEETEKESITAKKTVELLNE
EVLASPEELRQEMTERKKQIEDLKESIITAKQALQDKLEARDICANADKNVPVIEQKIQA
WAEEREDILDLMDEVDENLRKLSEMEEQLTFTTDKKSNHGKRMIEQAEMHEQLRREHL
QRSEELNKNIEEITGQIAALGKNQPSVSRDIEEKRQELLALKNAYSEQLAKYRNSSRDSFN
KFRKINALFNEVQRVSLEKKNAMDRAKNRLQNMLIGRLPSDYTFSTSSINDSENCDPISPI
ESDFSVFKN
>C_latens_HIM10
MSNAKSVVLVMFDPRKISTTLNQKLQVGVTPDNILNPTAEIVQQIYLNFVRVVINISENSL
HTLPLNADSDFDQELHKKSIPLAIVYQSMKAFIKDNSGGKLDLTMCDLVTPGKNPQRFR
KLSSFLADFIKLDEIAAPIFNEISEEFSGQKLEMEALQEEIVAAEKRKDELVARQSQRRRRE
NELMDDHNKKKSELAGIINQYTEIGVKTEELEKQRNELIRQIEEIEKESITAKKTVELLHE
EVLASPEELRQEMAERKKQIVDLKESIITAKRQALQDKLEARDICANAEKNVPVIEQKIQ
AWAEEREDILDLMDEVDENLRKLSEMEEQLTFTTDKKSNHGKRMIEQAEMHEQLRREH
MQRSEELNKNIDEITGQIAALGKNQPRVSRDIEEKRQELLALKNAYSEQLAKYRNSSRDS
FNKFRKINAHFNEVQRVALEKKNAMDRAKNRLQNMLIGRLPSEYTFSTSSINDSENCDPI
SPIDSDFSVFKN
>C_sp51_HIM10
MAHQKQVMLCTYDARIIAKSLSQKLQLGLTADNIINPTAENAQQIFSQFARIILNVSEHSL
TTLPLSADVDHDHELHRKSVPLTIVYQSMKAFIDDNSGGKLELTMCDLTTPAKNEIRFRK
LTSFLHDFIKLHEVASPVFNEICDEFSDRKLDMERIQEEVNIAEKKKEELLAKQASRKRRE
NELMNDHNKLKTELNNVVNQYMKNSEITSDIDKQTEEAFRQIESVERETVTGKKTVEHL
TEEVLTSPEELKHEMIQRKKHIEELKECLKSSKQSLQVQLEARDICINSEKSVPVVVEKIR
VWSEVRDDILDLIDSVEENLRKLNEKQEHLAFTADKRTKIAERVIEQAQMHDQLRKEHL

QRTEELQANIEKIAALGKNQPDVSKEIAQKQREELLSVKNAFSETIAKINNSCRDAISKFK
KIDVQFRETQRTSAEKRNAVQRAKDRLRTACIGRLPSDYTFSTSSINDSENCDPLSPIRLSN
FNVFK
>C_sp44_HIM10_paralog1
MLHQKQVVLIGHDHRIIARSLSQRLQLGLTPDNILNPTAEVSQQIFTHFARLILNVSEHSLT
TLPLSADIDHDHEMHRKSIPLIIVYQSMKAFIDDNSGGKLELTMCDLVTPGKNPQKFKRL
TSFLHDFIKFHEVATPIFNEISDEFSDRKLEMDQLQEELRDAEKKKDELLGKQASRKRREN
ELMKDHNTIKTELSSVVDKYMKNSELTNNIDKQSEEAIRQIEEVERETLTGKKTVEHLTE
EVLNSPEELKQEMEQRRAHIEELKECLKASQNLQAKLEAREICQNSEKNVPVVLEKIGV
WIEVRDEILDLIDSVEKNHRNLNEKNEQLSFTSNKKTNTNERMIEQAEMHQQLRKEHLQ
RMKELQANIEDIQRQIANLGINQPDVSKENAEKREALISVKNAHSETVSKIFSSCQEAVSK
YEKIVARFKETQNKSMEKKVAFDRAKDRLRAACVGRLPSDYTFSTSTLNDTENYDPLSPI
APSDVNVFK
>C_sp44_HIM10_paralog2
MQHQKQVVLKDYDVRTLARSLGQRLQLGLTPEDFINPTAECSQQIFTNFARLILNISEHSL
STLPLSASEIDINHEMHRKSIPLIIVFQSMKAFVHDNSGGKLDLSMCDLVTPGRNPQRFKK
LTSLLYDFIKLHEAAAPIFDEIAEEFSDRKIEMDQLQEELRAAEKKKDELLGKQASRKRRE
NELMNHHNKFKTELSSVVDQYTKNAELSNNNDKQSEEAFRQIEEVEREIITGKKTIEHLT
EEILDSPEELKQEMEQRRAHIEELKECLKASRQNLQDKLEARDICINSEKSGPVVHEKLG
AWKAVREEILALIELIEQNQRDLNDEYEKLTFIANKKTSVNERMVEQAEMYEQLRKEHL
QRMHNLQANIEDIQRQIASLGINQPEVSKENAAKREELISAKNQHSETIAKILSSCKEATA
KYEKIRADYIQTQRKAIEQRVAAARAKDRLRAACVGPLPSDYTFSTSTLTETENTEPLSPI
APSDFNVFK
>C_sp48_HIM10
MQHQKQVVLVNYNPRDIAKSLSQKLQLGLTGESITNPTGEVSQQIFSQFARIILNVSENSL
QQLPLTAGCDHDHELHRKSIPLIIVYQCMKAFIEDNSGGKLSFSMCDLVNPQRDATKFKR
LTSFLHDFIRLHEFAAPIFNEICDEFSDRKQEMELIQEELRAAEKRKDDLVAKQASRKRRE
NELMNDHNKLKTELNNVVNQYMKNTEMSNEIDKQAEEALRQVEEVERETITGKKTIEH
LTEEVLSSPEELKQEMAQRKKHIEELKECLKVSRQALQTKQEARDICTNAEKNVPVVAE
KIEVWAEVRDDILDLIDSVEENVRKLNEMQEDLALTANKKTKANELMVEQSQMHEQLR
NEHLQRTQKLQANIEEITKIAGLGKNQPEVSRENSKKQQELIVVKNAHSQTVARIVNSIQ
DSVSKFQKLELQFRETQKSALEKRNAVQRANDRLRAACVGRLPSDYTFSTSSLCDSENH
DPLSPIAPADFNVFK
>C_brenneri_HIM10
MQHQKQVILISYDQRTIARSLSLKLQLGLTGESITSPTAETSQQIFTQFARIILNVPEHSLTT
LPMSAGADHDNDLHRKSIPLIIVYQSMKAFIEDNSGGKLSLSMCDLVNPSKDPQKFKRLT
SFLHDFIRLHEYASPIFNEICEEFSDQKQEMELIKEELAAAEKRKNDLVAKQASRKRRENE
LMKNHNELKTELNNVVNQYMKNSELSNEIDKQTEEACRQVEEVERETITGKKTIEYLTE
EVLSSPEELKQEMAQRKKHIEELKECLKISSRQALQLKQEARDICINAEKNVPVVTEKIEV
WAEVRDDILDLMDSVEENVRKLNEMQEDLALTANKKAKANEQMVEQSQMHEQLRKE
HLQRTQELQANIEEITRKIAGLGKNQPEVSRENSKKQQELIAVKNAHSETVAKIINSIQDS
VSKFQKLEQQFRETQKSALEKRNAVQRANDRLRAACVGRLPSDYTFSTSSLCDSENHDP
LSPIAPADFNVFK
>C_wallacei_HIM10
MSNQPQVVLTLYDARLLAKSLSQKLQLGLTAENFLHPTAEVAQAVLTQFARIILNVPEHSL
STLPLSSNCDFDPELQRKSIPVVLVYLSKAFIKDNSGGKLELTMCDLTMPSKGNNNRFRK

LASFLHDFIKLHEFASPIFNEICEEFSDRKLEMEEVQEELIAAEKKKKDLLAKQASRKRRE
NELMNDHNKLKTELNNIVNQYMKNTELTSDIDKQSEEAIRQIEEIERETLTGKKTVEHLN
EEVLSSPEELKQEMDKRKKHIEELKECLKVSRQNLQTKLEARDICATAEKTLPVVVEKLE
AWSEVRDDILDLMDAVDGNLRKLNEMNEQLTFTANKKITVGERLVEQSQMQEQLRKEH
LQRTEELEANIEEISRIAALGKNQPDVSRDIANKRQELIAVKNAHSETIAKLTNSSHDAVSK
FRRIDAQFRETQRVSLEKRNAVQRAKDRFRNACVGRLPSDYTFSTSSINDSENCDPQTPM
ESDFNVFN
>C_tropicalis_HIM10
MSNQMQVVLTMFDAKIVAKALSQKLQLGLTGENITNPTEVAQNVLSQFARIILNVPEHSL
STLPLSSNCDFDHELQRKGIPVILVYLSMKAFIRDNSGGKLELTMCDLTMPAKTPNRFRK
LASFLHDFIRLHEFATPFFSEICEEFSDRKLEMEEVQEELMAAEKKKNDLLAKQASRKRH
ENELMNDHNKLKTELNNIVQQYTKNTEITKEVDKQTEETMRQIEEVERETLTGKKTVEH
LTEEVLTSPEELKQEMDTRKKHIEELKECLKVSKRQSLQSRLQARDICTTAEKNLPVAVE
KLQVWSDVRDDILDLIDAVDLNFRKLNELTDDLTINTDKKRNLGERLNEQSQMQEQLRR
EHMQRTEELQANIEEIKKISSLGSNQPDVSRDILKKKEELIAIKNAHSETVAKLTNSSVDA
MSKFSRIDAQFRETQRVSLEKRNAVHIAKSRVRNACIGRLTSDYTFSTSSMIDSENCDPLS
PVESDFSVFN
>C_doughertyi_HIM10
MSNQRNAVLTMFDSKNVSKLLNQKLQLGLTPDNITTPTAEIAHQVFSQFARMILNVSEHS
LSTLPLSVDSSDHDQEMHRKSIPIVIVYQSMKAFIKDNAGIDLTMCDLTTPAKVPNRFRRI
ASFLYDFIRLHEFASPIFNEISEEFADQKLEMASIQEELVVAEKRKNDLLSKQALRKRRENE
LMNDHNKLKTELNNVVNQYMKNSESSNDIDKQTEETTRQIESVEMETITGKKTIEHLNE
EVLSSPDELKQEMFERKKHIEELKECLKVSRQNLQAKREARDICIAAGKNVPIVIEKTEV
WSEVRDEIVDWMDMVDENRRKLSEMQEQLAFTTDKKAKAEQRIEEQKQVHEQLRQEH
LQRSQKLQADIEEITRITALGKNQPEVSRDIAKKREELIAVKNAYSETVAEITNSCQSAISK
FHKIDSQFKDTQRKAMEKQNSVHRAKDRFRNSFVGQLPSEYTFSTSSINDSENFDPQSPM
ESDFNVFN
>C_sp54_HIM10_paralog1
MANARPVVLIMYDARLIAKQLSQKLQVALTAESILTPTAEIAQQIYYNFVRLFLNVSEHSL
TTLPLSANSDHDQELHRKSISLVIVYQSMKAFIKDNSGEKLDLTMCDLVTPGKIPQRFRKL
TSFLVDFMKLHEMASPIFNEISEEFSDRKLEMEAIQEELLAAERRKNDLLSRQSLRKRREH
ELINDHNKVKGELNNIVSQFQKNVDDSTELDKQRNEAKQQINAFEKEIDTGKKTVEHLN
EEVLASPEELRQEMAERKKQIEELKDCLKSSRENLQTKLEARDICINSEKNLPVINEKIKM
WAEVRENIIDLIDIVNENLRKLNEMEEQLVFTTDKKKNTGERMVEQAEMHQQLRKEHL
QRIQELQNNIEDITRQITAMGKNQPDVSRGIEKKQRQELLATKNAHSATVANASNACQDA
LAKFRKVDSLFRDTQRIALEKKTAGDRAMGRLRNSFIGRLPSDYTFSTSSINDSENCDPCS
PVDSEFSVFK
>C_sp54_HIM10_paralog2
MANARPVVLIMYDARLIAKQLSQKLQVALTAENILTPTAEIAQQIYYNFVRLFLNVSDHS
LTTLPLSADSDHDQELHRKSISLVIVYQSMKAFIKDNSDKKLDLTMCDLVTPGKIPQRFRK
LTSFLVDFMKLHQMASPIFNEISEEFSDRKLEMEAIQEELLAAKKRKNDLVSRQSLRKRR
EHELMNDHNKVKEELSNIVSQYMENKDYSTALDEQKDEAEQQIEAFKKEIIAGKKTVEL
LNEEILDSPEELRQEMAERKKQIEELKDCLKSSRENLQTMLEARDICINSEKNVPVINEKI
KMWTEVRENIIDLIDIVNENRRKLKEIKEQFVFTADKKENTGERMVEQAEMHQQLRKEH
LQRIQELQNNIEDITRQITAMGKNQPDISRDIEKKRQELLATKNAHSVTVAKTSNTCQDAL

AKFRKVDSLFRDTQRIALEKKTAGDRAMGRVRNSFIGPLRNDYTFSTSSINDSDENCDPC
SPVDSEFSVFK
>C_inopinata_HIM10
MSTSKTVVLVMYDPRMIAKYLNQKLQVGLTPDDILAPTAEISQLVFTNFVRHVLGVSEQS
LNTLPLAVDFGPDQEMQRNSIPIIIVYQCMKAFIIDNSDKKLDLTMCDLVAPAKIPSRFRKL
TSFLVDFLKFNDLATPVFNEISDEFSDRKLEMEALQEELVAVEKRKNELISRQNLRKRRKH
ELINEHNKLKEELNKMVSEYTENQSNTVLLNKKKEDAKQQIEYLEKEVLTGKKTIDHLT
EEVLESPEELKQEMSERKKQIEELNDCLKCSRNLKSKLEDLEICVNAEKNVPVVVDKITT
WSVLREEILDLIDVENENLRKLKEMEEQLSFTSNKTEAANKRILEQAELHEQLRKQHLER
SDEWQKKIEEITRQISAMKFNQPDVSREIEKKRSDLLAAKNAHSEAISRMTNSCKETMSK
FRKIEAVFKDTQRTSAEKKTAGDRAFDRLRNACVGCLPSDYTFSTSSISHSENCDPIETEF
TVFK
>C_elegans_HIM10
MSNVVLIVYDPRMISKYLGQKLHMGLVADDIIKPTAEIAQQIFANFVRLVLNVSESSLTTL
PLSANCDYDPELHKKSIPIIILFQCMKAFIKDNSGNKLDLTMCDLVTPAKHEHRFRKLTSF
LVDFLKLHELATPAFNEISEEFSDRKFEMEKIREELLEAEKKKNDLLAKQSIRKRHEHELI
NEQSNAKAELKNVVNEYTETRQINEELDKQKEEAILHIQALEKEMLTGKKTIEHLNEEVL
TSPEQLKQEMEERKRHIEELRDCLESSKKGLQAKLEAREICINSEKNVPVIIEKIHQWTEV
REVIIDLIDVESENLRKLKEMEEQLDFMMKEMETAQKRLVEQSETHEQLRIEHTQKSEER
QRRIEEITEQIANLKTSQPDVSQEIAKKKQELLALKNAHSETISQITNSCQDAVAKFAKLN
AMFKETQKVAFEKNTAAAREMERLKSSLTGRLLSDYTFGSSTIDAGENTENCDPQPNDSS
FSVFK
>C_oiwi_HIM10
MTSARPAVLITYDARLVAKFLSQKLQVGLTAENILTPTAEIAQQVMVNFVRLIIGVGEHSL
STLPLSADCDHDPELFRNSIPLVIVYQFMKAFMMDYSGEKLDLTMCDLVTPAKFPQRFRK
LTSFLMDFIKLHERAAPLFDEISEEFRDRKIEMESLQEDLINEEKRKNDLISRQNLRRRREH
ELINDHNKVKGELNSIVNQYTENADISADVDKKKKDAKAQIEDFEREIITGKKQLEYLTE
EVLDSPEELRKEMEQRKHQIAELRECLESSRNLQAKMEALDICSNSDKNVPVVNERIRIW
SEKREAILDLIDSVEEDHRKLESLEEKLIFKQDEKNNVANQLLKCADSHEELRKEHLKRM
EEFNGKIADITQQIATLGKNQPEASRDIEKKKHELISLKNMHSQTVAQLSNTCKDTLAKF
RKVDALFKETQRVGLEKKTAGDRAMERLKSACVGRLPSDYTFSTSSINDSENFEPHSPFD
PESSVFK
>C_kamaaina_HIM10
MTSARPAVLITYDARLVAKFLSQKLQVGLTAENILTPTAEIAQQVMVNFVRLIIGVGEHSL
STLPLSADCDHDPELFRNSIPLVIVYQFMKAFMMDYSGEKLDLTMCDLVTPAKFPQRFRK
LTSFLMDFIKLHERAAPLFDEISEEFRDRKIEMESLQEDLINEEKRKNDLISRQNLRRRREH
ELINDHNKVKGELNSIVNQYTKNAEISADLDKKKKDAEAQIADFEREIITGKKTFEFLTEE
VLDSPEELKKEMEQRKHQIAELRECLESSRTNLQVKMEALDICSNSEKNVPVVNERIRIW
SEKREEILDLIDAVEEDHRKLESLEEKLIFKQDEKNNAAKQLINCADSHEELRKEHLKRM
EEFNGNISDITRQIATLGKNQPEASRDIEKKKHELISMKNIHSQTVAKLSNTCKDTLAKFH
KVDALFKEIQRVGLEKKTAGDRAMERLKSACVGRLPSDYTFSTSSINDSENCEPHSPFDP
ESSVFK
>C_waitukubuli_HIM10
MATQKPVALVPMDKVSICKILNQKLQIGLTQDTLITPTVEIAQLMYMNFVRSLLNVSESC
LSTLPLSATCDHDPELHRRSIPIIIAYQCMKAFIKDHFGDKLDFQMCDLVYPQKTPGRFKRI
AGFLADYIRFHEKGLPVFNEVSEEFSYQKQEVELLQEELLEEEKRKNALLAQQNQRKRR

EHELINEHNKANAELNGKIAQYEASTANAEVLEKEKLEAMDAVDRMENEIISCRKMVD
HLKEEVLSSPEELKREMALRKKQIEELKECLNGAKWSLAERLEAIEICTSFEKYKPTVDE
KMRQFAAMKEEIIQLFDAVHENQRVLSDLEDEKKFTEEKRKNVIEVLGENAHNHAELRE
KHLQRIEELNRKIEEIMQQIADLGKNQPDVSRDIGKKQRQKLLSVKNATSELVAMFEQEI
RETLTKFQKVLAMFQSVSRGADEKRVAFDRAKSRVVNSCNGQLRTDYTFSTSSIADDEN
TAPDAIVDGDFEVFK
>C_panamensis_HIM10
MAMQKPVVLMPMEKAPLIKLLNQKLQLGITLETLTTPTADLAQKMYRTFVRQILNVSES
CLSTLPLSADCDHDPELHRNSIPIIIVYQCMKAFIKDHSGDKLDLTMCDLIQPHRVPGRFK
KMATFLADYMRFHEIGNPVFNEISEEFSYQKQEVEMLHIELQNEESRKNGLLANQNLRK
RREHELINEHNKVKTEFGNMVAQYEASHAAAEALNKEKEEALEETDRMEMEIISGKKM
VDHLKEEVLSSPEELKVEMANRKKLIEELKDNLKATKKSYTERMEAIEICASFEKNKLMI
EEKFRQFAMVKDEIMELLDAENENQRKLDDMEVEHRYMVEQRKNMHELLEEKALNHA
QLRKEHLQRNEELNKKIEEITKKIAALGKNQPDVSRDIEKKRQELLAVKNCNSEIVAKAE
HETREKLAKFQKVQTAFLKVHRNAEEKKTACERRISRVVNASVGRFSDHTFSTSSINEDS
ENTAPDSLQFNVFK
>C_nouraguensis_HIM10_paralog1
MAQKPVVLVALDKAVICKILKPKLHLGQLTPEDINNPSSEIAQQIFSNFVRYTLNVSESCM
STLPLSATSDHDPELHRRSIPIIIVFQCLKAFIKDHSGDKLDLTMCDFVNPQRINGRFKKITS
FLADYIRFHENAQPIFNEVSEEFSYQRQEEQQLNEELQEEEKRKEMLISQQNARKRKDNE
LCNEHIKLKEELMNFMVACDEKKAFVEALFTEKEATEEKTESIENEILSGKKMVDHLKEE
ILSSPEELKREMAARKKQIEELKECLAGSKLALAERMEAIEICSSAEKNVPAIQEKINQFA
MMKEEILELLDAVNEDNRKLSDLEDELKFTEEKKIKTRELMVDSAQLHEQVRKEHLQR
NEQLNEKIKEITQISQMGTNQPDVSRDIEKKQELRAAKNATSKAVAKVVEETRETMAKY
EKVLAMFQKVQRDAVEKQVAADRATSRLVNACVGPLISDYTFSTSSCCAEDDENTAPGV
GTNFNVFPK
>C_nouraguensis_HIM10_paralog2_pseudogene
DKEAEMELRKKVLEERANKSVDDKETRALIRDMMENEAELKNARNEHEKLRLKLQQM
EKKLIVGGENLLEKVEEQAKLLEISNREMESSKQSEERLRSQLEEKTAWKVEIEERYSSL
QEESAAKTRKTKRVTNELREVRMELKDVEEEHQRQLEAMLEDARQLRK
>C_becei_HIM10
MAQKPVVLVTLDKAVICKILKPKLQLGHLTPEDINNPTEIAQQIFSNFVRYVLNVSESCMS
TLPLSASCDYDPELHRRSIPIIIVFQCLKAFIKDHSGDKLDLTMCDFVNPHRINGRFKKITS
FLADYIRFHENAQPVFNEISEEFSYQRQEEEQLNEELQEEEKRKEMLISQQSARKRKENE
LYNEQIKLKEEFSAIVGKCEEKKVFSETLFAEKVAAEEKTEGIENEILSGKKMVDHLKEEI
LSSPEELKQEMTARKKQIEELKECLAGSKLALAERMEAIEICTNVEKNVPAIQEKINQFVI
MKEEIIELMDAVNEDNRKLNDLEDELKFTEEKKIKTRELMVDSAQLHEQLRKEHLQRNE
QLNEKIKEITQISAMGKNQPDVSRDIEKKQELRAAKNATSEAVAKVVEETRETMAKYEK
VLAMFQKVQRDAVEKQVAADRATSRLVNACIGPLISDYTFSTSSCCAEDDENTVPGGGT
DFNVFPK
>C_yunquensis_HIM10
MAAQKPVVLVTLDKAVICKILKPKLHLGSLTPEDIMSPSAEIAQLVFSNFVRHTLNVSETC
MTTLPLSATCDHDPELHRRSVPIIIVLQYMNAFIRDHSGGKLDMTMCDLIHPERIPGRFKK
LTSFLADYIRFHENAQPVFNEISEEFSYQKLEEEQLRKELEEEEKRKASLTSQQNLRKRRE
NELRNELAKVKSEMFDIVATCESKKATFATLLSEKNAAVEETARIESDILSGRKMVDHLKE
EVLSSPEELKLEMAARKKQIEELKECLKGSKQALVERIEAIEICASAEKNEPVILEKINQFA

AMKEDIIELLDAVNEDNRKLSDLEDELKFTVEKKNNVHELMGEKAQLHAQLRNEHLQR
TEQLNEKINEITKQIAAMGTNQPDVSREIEKKSQELRDAKNANSEAVAAVIQETRETMAK
YNRVLAKFQQLQKDAAEKQVAAERATSRLVNACVGKLLTDYTFSTSSICTEEDENAAPD
GQSFNVFQK
>C_macrosperma_HIM10
MASQKPVVLVPMDKVPICKILNQALQLGLTPDNITNPTAEIAQQVYINFVRLILNVSESCL
STLPLSANCNYDPELHRRSIPIIIVYQCKAFIKDHSGDKLDLTMCDLVQPQRVQGRFKKLA
TFLADYIKFHENGQPVFNEISEEFSYRKLEVEQLQLAIEDEERRKNDLLSQQNLRKRREH
ELINEHNKVKSEFSGVVGQYEANKITAGELLKQNEEAVEQIEQVENEILRSRKMAEHLKE
ELLSSPEELRLEMAARKKQIEELNECLKGSKVALAERMEAIDICINVEKNTPAVNEKLKM
WAAVKEEIIMLFDAVNEDHRKLSDLENEQRFTAEKKKNVHELIGEQAQRHAQLQKEHLQ
RNVELNNKIDEITKQIAALGKNQPDVSRDIEKKRQELLAVKNANSESVAKVAQECREMF
TKYQNVLARFQEVRRDANEKQVAVDRTTSRVVNACDGRLPTDYTFSTSSCCTEDGENIT
PDTQLRTDFNVFK
>C_sulstoni_HIM10
MASVRPVVMIKHDFKSISRNLNAKLHLNTRPEDISNPTVAELAQNVYMNFVRLILNVPD
HSLSILPINSNVDYDPDQHRNSVRLIIVYQCMKAFVTDHSGRAFDLSMCDLVLPGKIPGR
FQKLMSFLVDFMKLYQLAKPIFSEISEEFSYRKQEVEELKQALYEEEKRKKEMLAQQSLR
RRREHELIDEHAKVTHELNGIVQQYTASTTRATELEKQKEEAIQSIERLESETLSGKKMVE
HLNEEFLASPEELKREMAERKKQIEELTECRNSSKAKLAALEICRHIEKNFPASAEKIKIFR
NVRAEILKLFDAVNENLRQLTDLEQELKFTTEKTKKSHEMMEEQAEMHKQLRNEHLQR
SRELDSRIEEISREIAAMVKNQPDLSRDIENKRQELLVFKNEHSQTVSRILRHCEDLLVKY
RKVHAMFEETQRTAQEKKTAGERAKGRVRTACFGRLPTDYTFNTSSLNEEKDENCRPDE
NFTVFK
>C_afra_HIM10
MASVRQIVLIKYDVKQISKVLNAKFQLGTKPDDIIKPSAELAQNIYMNFARLVLCIPDHSL
TTLPISAYTDFDQDQHRNSVRLSLVYQCKAFIVDMSLGALSLSMCDLVVPDRTPGRFQKL
MSFLVDFMKFHQVAEPTFSEISEEFSHRKKEVEELKQLLYEEERRKSELIAQQSLRKRREH
ELIDEHTRVNNELSGIIQQYTANTTTAGELDKQKEEALLTIERLEMETISGKKMVEHLNEE
FLTSPDELRREMAERKRQIEELTECRNSAKEICRNIEKNFPASSEKIKTFQNVRSEIVKLFD
AVNGNLRQLEDMEQELNFTTEKTRKSHEMMEEQAEMHKQLRKEHLQRSDELDARIEEI
TREIAAMGKNQPDLSREIEKKRQELLAFKNAHSQTVARILRHCEDLLAKYRKVHAMFEE
TRRNAEEKRIAGERAKGRVLNACSGRLPTDYTFNTSSVNEDLDENCRPDENFTVFN
>C_sp49_HIM10
MASEKEKRTVVLTMYDPNTIAKVLSSKLKLGLTPDDIINPTVRPTAFLVFQCFVRHVLDV
SDASMSSLPLSAQSDEMDHDSHKRTIQLAIVYQCLKAFIADNSGNKIILSMCDIVQPAAV
QNRFKRITSFLVDFIRLHDTALPIWDEIREEFSDRKHEVQSLQSDLITEEKRKNALLTQQSQ
RKRREHDLINDHNKVKQELTTTIDQYQANLAELEELKKKEKETNEETERLGTEILSGRK
MVEHLSEELLTDPEELKKEMKTRKKEIEEMRGRLLNAKNILQKREEEIKICAEADRNEIL
FNEKLDVWEQIKADIVSLREEISDNVRTLSELEEKLKLTCEKRKLVAERMKEQAETQNQL
RQQNSDRNEELQNKIDEVTEEIAALGRNQPDVSRIIEKKRQELLAVKNAMSSKEAECANS
CKETLLKLRRMEAMFEELHRVSLEKRTAADRARYRVKTACVGNSMADYTFTSESIDENN
PPNP
>C_sp25_HIM10
MTSEKEKRSVVLTIYDPNTIAKILNAKLKLGLTPDDIINPTHATAFQVFQSFVRHVLGVSD
AAMNSLPLAAQSDDFDHDSHRKTIQLGIVYQCLKAFIADNSGQKIILSMCDIVQPAAIQH

RFKKVTSFLADFIKLREVALPIWDEIREEFSDRNHEVQSLQGDLIAAEKRKNALLSQQSQ
RKRREHELINEHNKVNQELTNIVEQYTANMKEVEDRKKKKEETLNEIERLVSEILSGRKM
VEHLGEEVLTSPEDLKNEMAGRKKQIEELKEHLIQARQSVQEKEEAIKICAEAERNEGVF
NDKLDAWEKVKSDIVSIREEINENLRAFSELEEKLKLLSEKRKKVAERMQEQAETQNQL
RQQNSERNDELQNNIDAVTEQIAALGKSQPDVSRIIEKKRQDLLAVKNAMSANESECAN
SCKETLFKLRKMEAMFDELHRVSLEKRTAADRARCRIKTACVGNRMADYTFSTESIDEN
NPPNSSFNIFK
>C_imperialis_HIM10
VHLYMYDARLIAKVLSSKLKLGLTPDDITNPTSEVAIQVFTNFVRFVLDVSETSLTSLPLTA
QVDDLDYESHRKTIPLVIVYQCLKAFVSDNSGKKLILTMCDFVNPAAIQNRFKKVTSFLV
DFIKLHGHALPIWDDIRDEFSDRKHEVQSLQNDLVTEEKRKNALLSQQSQRKRREHELIN
EHNKVNTELTNVVGQYTANMEEVEERKKKKEEAYDKIERLTNEIISGRKMVEHLGEEVL
SSPEELKNEMAMRKKQIEELREHLAQARKALQEKDEAVKICTEAERNEVVFNEKLVMW
DQVREDIVTIREEINENLRALSEYEEKLKLTIEKRKIVGERMREQSEQQEEMREQHSERN
KQLQLKIDAVTEQIAALGKNQPDVSRDIEKKRQELLAVKNMMSEKEAECANSCRETLLK
LRKMETMYEELFRVALEKRTAADRAACRVKNACAGVSMADYTFNTESIDENNPPPHTSF
KVFN
>C_japonica_HIM10
MASGRPVLLTILDMRTILRVLNGKLHLGLTQENILTPTAEVAQQVFYNFVRYVLSVPESSL
TTLPLTADVDVDNEMNRKSIPLVIVYQCMKAFIKDNTGGKLDLTMCDFVTPAKIQNRFK
KLTSFLADFIRLHDMAMPLWNEISDEFGYRKHELESLQSEVMAVEKRKDDLLAQQSLRK
RREHELINEHNKVKSELNKIIGQYNSNKSAAEERSKQKEEAIELIEKVENDVISGKKMVE
HLSGEVLSSPEELKAEMEARRKQIEELRDCLRHSRKTLQNKEEALKICAEAEKNVPVLID
KINSWSELQEEIAELIDVINDNMRKLAELEENLQLTIEKKKKVGERMDEQAKLQTQLRR
QHFQRNEDLQHKIEEITAEISALGKNQPDVSRDIERKRQELLLVKNALSEDIAELTNWCHE
SMSKFRKVQELFGETHRIALEKQTAGKRAKHRVRNAIFGPLPTEYTFDYTKTLSMDEND
VGGNGSSVDFKVFK

All supplementary data and supplementary tables are available in Caro L et al., 2022.

Caro, L., Raman, P., Steiner, F.A., Ailion, M., Malik, H.S., 2022. Recurrent but Short-Lived Duplications of Centromeric Proteins in Holocentric Caenorhabditis Species. Mol. Biol. Evol. 39, msac206. https://doi.org/10.1093/molbev/msac206

3.8 REFERENCES

Akera T, Chmátal L, Trimm E, Yang K, Aonbangkhen C, Chenoweth DM, Janke C, Schultz RM, Lampson MA. 2017. Spindle asymmetry drives non-Mendelian chromosome segregation. Science 358:668-672.

Ali-Ahmad A, Sekulić N. 2020. CENP-A nucleosome-a chromatin-embedded pedestal for the centromere: lessons learned from structural biology. Essays Biochem 64:205-221.

Allshire RC, Karpen GH. 2008. Epigenetic regulation of centromeric chromatin: old dogs, new tricks? Nat Rev Genet 9:923-937.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403-410.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-3402.

Bailey TL, Gribskov M. 1998. Combining evidence using p-values: application to sequence homology searches. Bioinformatics 14:48-54.

Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME Suite. Nucleic Acids Res 43:W39-49.

Blower MD, Karpen GH. 2001. The role of Drosophila CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. Nat Cell Biol 3:730-739.

Brenner S. 1974. The genetics of Caenorhabditis elegans. Genetics 77:71-94.

Buchwitz BJ, Ahmad K, Moore LL, Roth MB, Henikoff S. 1999. A histone-H3-like protein in C. elegans. Nature 401:547-548.

Bureš P, Zedek F. 2014. Holokinetic drive: centromere drive in chromosomes without centromeres. Evolution 68:2412-2420.

Cheerambathur DK, Prevo B, Chow TL, Hattersley N, Wang S, Zhao Z, Kim T, Gerson-Gurwitz A, Oegema K, Green R, et al. 2019. The Kinetochore-Microtubule Coupling Machinery Is Repurposed in Sensory Nervous System Morphogenesis. Dev Cell 48:864-872 e867.

Chmátal L, Gabriel SI, Mitsainas GP, Martínez-Vargas J, Ventura J, Searle JB, Schultz RM, Lampson MA. 2014. Centromere strength provides the cell biological basis for meiotic drive and karyotype evolution in mice. Curr Biol 24:2295-2300.

Cho S, Jin SW, Cohen A, Ellis RE. 2004. A phylogeny of caenorhabditis reveals frequent loss of introns during nematode evolution. Genome Res 14:1207-1220.

Cook DE, Zdraljevic S, Roberts JP, Andersen EC. 2017. CeNDR, the Caenorhabditis elegans natural diversity resource. Nucleic Acids Res 45:D650-D657.

Cortes-Silva N, Ulmer J, Kiuchi T, Hsieh E, Cornilleau G, Ladid I, Dingli F, Loew D, Katsuma S, Drinnenberg IA. 2020. CenH3-Independent Kinetochore Assembly in Lepidoptera Requires CCAN, Including CENP-T. Curr Biol 30:561-572.e510.

Cutter AD. 2008. Divergence times in Caenorhabditis and Drosophila inferred from direct estimates of the neutral mutation rate. Mol Biol Evol 25:778-786.

de Groot C, Houston J, Davis B, Gerson-Gurwitz A, Monen J, Lara-Gonzalez P, Oegema K, Shiau AK, Desai A. 2021. The N-terminal tail of C. elegans CENP-A interacts with KNL-2 and is essential for centromeric chromatin assembly. Mol Biol Cell 32:1193-1201.

De Wulf P, Earnshaw WC. 2008. *The Kinetochore: from Molecular Discoveries to Cancer Therapy*: Springer New York.

Delaney K, Mailler J, Wenda JM, Gabus C, Steiner FA. 2018. Differential Expression of Histone H3.3 Genes and Their Role in Modulating Temperature Stress Response in. Genetics 209:551-565.

Des Marais DL, Rausher MD. 2008. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. Nature 454:762-765.

Despot-Slade E, Mravinac B, Širca S, Castagnone-Sereno P, Plohl M, Meštrović N. 2021. The Centromere Histone Is Conserved and Associated with Tandem Repeats Sharing a Conserved 19-bp Box in the Holocentromere of Meloidogyne Nematodes. Mol Biol Evol 38:1943-1965.

Drinnenberg IA, deYoung D, Henikoff S, Malik HS. 2014. Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. Elife 3:e03676.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792-1797.

Elisafenko EA, Evtushenko EV, Vershinin AV. 2021. The origin and evolution of a two-component system of paralogous genes encoding the centromeric histone CENH3 in cereals. BMC Plant Biol 21:541.

Evtushenko EV, Elisafenko EA, Gatzkaya SS, Schubert V, Houben A, Vershinin AV. 2021. Expression of Two Rye CENH3 Variants and Their Loading into Centromeres. Plants (Basel) 10.

Finseth FR, Dong Y, Saunders A, Fishman L. 2015. Duplication and Adaptive Evolution of a Key Centromeric Protein in Mimulus, a Genus with Female Meiotic Drive. Mol Biol Evol 32:2694-2706.

Finseth FR, Nelson TC, Fishman L. 2021. Selfish chromosomal drive shapes recent centromeric histone evolution in monkeyflowers. PLoS Genet 17:e1009418.

Fukagawa T, Earnshaw WC. 2014. The centromere: chromatin foundation for the kinetochore machinery. Dev Cell 30:496-508.

Gallach M, Betrán E. 2011. Intralocus sexual conflict resolved through gene duplication. Trends Ecol Evol 26:222-228.

Gassmann R, Rechtsteiner A, Yuen KW, Muroyama A, Egelhofer T, Gaydos L, Barron F, Maddox P, Essex A, Monen J, et al. 2012. An inverse relationship to germline transcription defines centromeric chromatin in C. elegans. Nature 484:534-537.

Ginsberg PS, Humphreys DP, Dyer KA. 2019. Ongoing hybridization obscures phylogenetic relationships in the Drosophila subquinaria species complex. J Evol Biol 32:1093-1105.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307-321.

Henikoff S, Ahmad K, Malik HS. 2001. The centromere paradox: stable inheritance with rapidly evolving DNA. Science 293:1098-1102.

Henikoff S, Ahmad K, Platero JS, van Steensel B. 2000. Heterochromatic deposition of centromeric histone H3-like proteins. Proc Natl Acad Sci U S A 97:716-721.

Hittinger CT, Carroll SB. 2007. Gene duplication and the adaptive evolution of a classic genetic switch. Nature 449:677-681.

Hobolth A, Dutheil JY, Hawks J, Schierup MH, Mailund T. 2011. Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. Genome Res 21:349-356.

Hofstatter PG, Thangavel G, Lux T, Neumann P, Vondrak T, Novak P, Zhang M, Costa L, Castellani M, Scott A, et al. 2022. Repeat-based holocentromeres influence genome architecture and karyotype evolution. Cell.

Howman EV, Fowler KJ, Newson AJ, Redward S, MacDonald AC, Kalitsis P, Choo KH. 2000. Early disruption of centromeric chromatin organization in centromere protein A (Cenpa) null mice. Proc Natl Acad Sci U S A 97:1148-1153.

Ishii T, Juranić M, Maheshwari S, Bustamante FO, Vogt M, Salinas-Gamboa R, Dreissig S, Gursanscky N, How T, Demidov D, et al. 2020. Unequal contribution of two paralogous CENH3 variants in cowpea centromere function. Commun Biol 3:775.

Iwata-Otsubo A, Dawicki-McKenna JM, Akera T, Falk SJ, Chmátal L, Yang K, Sullivan BA, Schultz RM, Lampson MA, Black BE. 2017. Expanded Satellite Repeats Amplify a Discrete CENP-A Nucleosome Assembly Site on Chromosomes that Drive in Female Meiosis. Curr Biol 27:2365-2373.e2368.

Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275-282.

Kato H, Jiang J, Zhou BR, Rozendaal M, Feng H, Ghirlando R, Xiao TS, Straight AF, Bai Y. 2013. A conserved mechanism for centromeric nucleosome recognition by centromere protein CENP-C. Science 340:1110-1113.

Kiontke K, Gavin NP, Raynes Y, Roehrig C, Piano F, Fitch DH. 2004. Caenorhabditis phylogeny predicts convergence of hermaphroditism and extensive intron loss. Proc Natl Acad Sci U S A 101:9003-9008.

Korber B. 2000. HIV signature and sequence variation analysis. In: Computational and Evolutionary Analysis of HIV Molecular Sequences.: Kluwer, Dordrecht.

Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol 23:1891-1901.

Krátká M, Šmerda J, Lojdová K, Bureš P, Zedek F. 2021. Holocentric Chromosomes Probably Do Not Prevent Centromere Drive in Cyperaceae. Front Plant Sci 12:642661.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 35:1547-1549.

Kumon T, Ma J, Akins RB, Stefanik D, Nordgren CE, Kim J, Levine MT, Lampson MA. 2021. Parallel pathways for recruiting effector proteins determine centromere drive and suppression. Cell 184:4904-4918.e4911.

Kursel LE, Malik HS. 2017. Recurrent Gene Duplication Leads to Diverse Repertoires of Centromeric Histones in Drosophila Species. Mol Biol Evol 34:1445-1462.

Kursel LE, McConnell H, de la Cruz AFA, Malik HS. 2021. Gametic specialization of centromeric histone paralogs in Drosophila virilis. Life Sci Alliance 4:e202000992.

Kursel LE, Welsh FC, Malik HS. 2020. Ancient Coretention of Paralogs of Cid Centromeric Histones and Cal1 Chaperones in Mosquito Species. Mol Biol Evol 37:1949-1963.

Lampson MA, Black BE. 2017. Cellular and Molecular Mechanisms of Centromere Drive. Cold Spring Harb Symp Quant Biol 82:249-257.

Lermontova I, Koroleva O, Rutten T, Fuchs J, Schubert V, Moraes I, Koszegi D, Schubert I. 2011. Knockdown of CENH3 in Arabidopsis reduces mitotic divisions and causes sterility by disturbed meiotic chromosome segregation. Plant J 68:40-50.

Levine MT, McCoy C, Vermaak D, Lee YC, Hiatt MA, Matsen FA, Malik HS. 2012. Phylogenomic analysis reveals dynamic evolutionary history of the Drosophila heterochromatin protein 1 (HP1) gene family. PLoS Genet 8:e1002729.

Li Y, Huang JF. 2008. Identification and molecular evolution of cow CENP-A gene family. Mamm Genome 19:139-143.

Lin Z, Xie Y, Nong W, Ren X, Li R, Zhao Z, Hui JHL, Yuen KWY. 2021. Formation of artificial chromosomes in Caenorhabditis elegans and analyses of their segregation in mitosis, DNA sequence composition and holocentromere organization. Nucleic Acids Res 49:9174-9193.

Lin Z, Yuen KWY. 2020. Construction and analysis of artificial chromosomes with de novo holocentromeres in Caenorhabditis elegans. Essays Biochem 64:233-249.

Lipinski KJ, Farslow JC, Fitzpatrick KA, Lynch M, Katju V, Bergthorsson U. 2011. High spontaneous rate of gene duplication in Caenorhabditis elegans. Curr Biol 21:306-310.

Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. Science 290:1151-1155.

Maddox PS, Hyndman F, Monen J, Oegema K, Desai A. 2007. Functional genomics identifies a Myb domain-containing protein family required for assembly of CENP-A chromatin. J Cell Biol 176:757-763.

Maheshwari S, Tan EH, West A, Franklin FC, Comai L, Chan SW. 2015. Naturally occurring differences in CENH3 affect chromosome segregation in zygotic mitosis of hybrids. PLoS Genet 11:e1004970.

Mailund T, Munch K, Schierup MH. 2014. Lineage sorting in apes. Annu Rev Genet 48:519-535.

Malik HS. 2009. The centromere-drive hypothesis: a simple basis for centromere complexity. Prog Mol Subcell Biol 48:33-52.

Malik HS, Henikoff S. 2001. Adaptive evolution of Cid, a centromere-specific histone in Drosophila. Genetics 157:1293-1298.

Malik HS, Vermaak D, Henikoff S. 2002. Recurrent evolution of DNA-binding motifs in the Drosophila centromeric histone. Proc Natl Acad Sci U S A 99:1449-1454.

Marques A, Pedrosa-Harand A. 2016. Holocentromere identity: from the typical mitotic linear structure to the great plasticity of meiotic holocentromeres. Chromosoma 125:669-681.

Marques A, Ribeiro T, Neumann P, Macas J, Novák P, Schubert V, Pellino M, Fuchs J, Ma W, Kuhlmann M, et al. 2015. Holocentromeres in Rhynchospora are associated with genome-wide centromere-specific repeat arrays interspersed among euchromatin. Proc Natl Acad Sci U S A 112:13633-13638.

McKinley KL, Cheeseman IM. 2016. The molecular basis for centromere identity and function. Nat Rev Mol Cell Biol 17:16-29.

Mellone BG, Fachinetti D. 2021. Diverse mechanisms of centromere specification. Curr Biol 31:R1491-R1504.

Melters DP, Paliulis LV, Korf IF, Chan SW. 2012. Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis. Chromosome Res 20:579-593.

Monen J, Hattersley N, Muroyama A, Stevens D, Oegema K, Desai A. 2015. Separase Cleaves the N-Tail of the CENP-A Related Protein CPAR-1 at the Meiosis I Metaphase-Anaphase Transition in C. elegans. PLoS One 10:e0125382.

Monen J, Maddox PS, Hyndman F, Oegema K, Desai A. 2005. Differential role of CENP-A in the segregation of holocentric C. elegans chromosomes during meiosis and mitosis. Nat Cell Biol 7:1248-1255.

Moore LL, Roth MB. 2001. HCP-4, a CENP-C-like protein in Caenorhabditis elegans, is required for resolution of sister centromeres. J Cell Biol 153:1199-1208.

Nei M, Kumar S. 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.

Nielsen CB, Friedman B, Birren B, Burge CB, Galagan JE. 2004. Patterns of intron gain and loss in fungi. PLoS Biol 2:e422.

Oegema K, Desai A, Rybina S, Kirkham M, Hyman AA. 2001. Functional analysis of kinetochore assembly in Caenorhabditis elegans. J Cell Biol 153:1209-1226.

Pan D, Zhang L. 2007. Quantifying the major mechanisms of recent gene duplications in the human and mouse genomes: a novel strategy to estimate gene duplication rates. Genome Biol 8:R158.

Prosée RF, Wenda JM, Özdemir I, Gabus C, Delaney K, Schwager F, Gotta M, Steiner FA. 2021. Transgenerational inheritance of centromere identity requires the CENP-A N-terminal tail in the C. elegans maternal germ line. PLoS Biol 19:e3000968.

Ravi M, Shibata F, Ramahi JS, Nagaki K, Chen C, Murata M, Chan SW. 2011. Meiosis-specific loading of the centromere-specific histone CENH3 in Arabidopsis thaliana. PLoS Genet 7:e1002121.

Robertson HM. 1998. Two large families of chemoreceptor genes in the nematodes Caenorhabditis elegans and Caenorhabditis briggsae reveal extensive gene duplication, diversification, movement, and intron loss. Genome Res 8:449-463.

Roy SW, Penny D. 2007. Patterns of intron loss and gain in plants: intron loss-dominated evolution and genome-wide comparison of O. sativa and A. thaliana. Mol Biol Evol 24:171-181.

Schrader F. 1935. Notes on the mitotic behavior of long chromosomes. Cytologia 6:422–430.

Schubert V, Lermontova I, Schubert I. 2014. Loading of the centromeric histone H3 variant during meiosis-how does it differ from mitosis? Chromosoma 123:491-497.

Schueler MG, Swanson W, Thomas PJ, Green ED, Program NCS. 2010. Adaptive evolution of foundation kinetochore proteins in primates. Mol Biol Evol 27:1585-1597.

Senaratne AP, Cortes-Silva N, Drinnenberg IA. 2022. Evolution of holocentric chromosomes: Drivers, diversity, and deterrents. Semin Cell Dev Biol.

Senaratne AP, Muller H, Fryer KA, Kawamoto M, Katsuma S, Drinnenberg IA. 2021. Formation of the CenH3-Deficient Holocentromere in Lepidoptera Avoids Active Chromatin. Curr Biol 31:173-181.e177.

Stecher G, Tamura K, Kumar S. 2020. Molecular Evolutionary Genetics Analysis (MEGA) for macOS. Mol Biol Evol 37:1237-1239.

Steiner FA, Henikoff S. 2014. Holocentromeres are dispersed point centromeres localized at transcription factor hotspots. Elife 3:e02025.

Stevens L, Félix MA, Beltran T, Braendle C, Caurcel C, Fausett S, Fitch D, Frézal L, Gosse C, Kaur T, et al. 2019. Comparative genomics of 10 new Caenorhabditis species. Evol Lett 3:217-236.

Stoler S, Keith KC, Curnick KE, Fitzgerald-Hayes M. 1995. A mutation in CSE4, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis. Genes Dev 9:573-586.

Suvorov A, Kim BY, Wang J, Armstrong EE, Peede D, D'Agostino ERR, Price DK, Waddell P, Lang M, Courtier-Orgogozo V, et al. 2022. Widespread introgression across a phylogeny of 155 Drosophila genomes. Curr Biol 32:111-123.e115.

Talbert PB, Bryson TD, Henikoff S. 2004. Adaptive evolution of centromere proteins in plants and animals. J Biol 3:18.

Tamura K, Stecher G, Kumar S. 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol 38:3022-3027.

Wang Y, Wu L, Yuen KWY. 2022. The roles of transcription, chromatin organisation and chromosomal processes in holocentromere establishment and maintenance. Semin Cell Dev Biol.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13:555-556.

Zedek F, Bureš P. 2016. Absence of positive selection on CenH3 in Luzula suggests that holokinetic chromosomes may suppress centromere drive. Ann Bot 118:1347-1352.

Zedek F, Bureš P. 2012. Evidence for centromere drive in the holocentric chromosomes of Caenorhabditis. PLoS One 7:e30496.

Zhao G, Oztan A, Ye Y, Schwarz TL. 2019. Kinetochore Proteins Have a Post-Mitotic Function in Neurodevelopment. Dev Cell 48:873-882.e874.

Zhu J, Cheng KCL, Yuen KWY. 2018. Histone H3K9 and H4 Acetylations and Transcription Facilitate the Initial CENP-A. Epigenetics Chromatin 11:16.

Chapter 4. FUTURE DIRECTIONS

This chapter addresses lingering questions about the mechanism of *peel-1 zeel-1*. Section 4.1 focuses on the toxin mechanism while Section 4.2 focuses on the antidote mechanism. Both sections highlight multiple unanswered questions and experiments probing these questions. The data shown here are preliminary, and therefore, results should be cautiously interpreted. Section 4.3 explores evolutionary questions regarding *peel-1 zeel-1*, including its possible origins.

My model of how the *peel-1 zeel-1* element works is the following. PEEL-1 is loaded into sperm and does not cause toxicity in these cells because they are missing PMPL-1. After fertilization, PEEL-1 protein is delivered to the embryo. PEEL-1 on its own forms an oligomer, where amphipathic helices construct the lining of the channel. However, the PEEL-1 channel is in a closed conformation when PMPL-1 is not present. The PEEL-1 channel remains in a closed and inactive conformation during early embryonic development because of missing (or low levels) of PMPL-1. Soon after the maternal-to-zygotic transition, *zeel-1(+)* embryos express the antidote, resulting in interactions between the antidote (ZEEL-1), PEEL-1, and E3 ubiquitin ligase machinery. As a result of these interactions, PEEL-1 is ubiquitylated and degraded. On the other hand, *zeel-1(-)* embryos do not degrade PEEL-1 and the toxin is maintained in the muscle and epidermal tissues. In the two-fold stage of embryonic development, PMPL-1 levels are high enough in the muscle and epidermis to cause toxicity in these tissues. PMPL-1 interacts with the PEEL-1 cation channel, gating the channel open, causing influx of sodium, osmotic stress, cell swelling, cell death, and finally, death of the embryo.

4.1 MECHANISM OF PEEL-1 TOXICITY

In my model, I hypothesize that oligomers of PEEL-1 amphipathic helices are the primary structural component of the toxic cation channel. I hypothesize that PMPL-1 interacts with a PEEL-1 oligomer, causing a conformational change in the channel to gate it open, allowing passage of monovalent cations through the membrane.

4.1.1 Do PEEL-1 and PMPL-1 interact?

To test the hypothesis that PEEL-1 toxicity involves a direct protein-protein interaction between PEEL-1 and PMPL-1, I have carried out several investigations, but results have varied. In some cases, I have been unable to detect an interaction between PEEL-1 and PMPL-1. These negative results may be because of technical challenges, namely, PEEL-1 and PMPL-1 co-expression causes cell death and the assays I have used are prone to false-positive results. Meanwhile, in other cases, I have detected an interaction between PEEL-1 and PMPL-1, but the results are difficult to interpret without stringent negative controls. Here I present data from assays probing PEEL-1 and PMPL-1 interactions in yeast, mammalian cells, worms, *E. coli*, and with purified proteins *in vitro*.

Mating-based split ubiquitin system (mbSUS) is an assay in yeast used to test for protein-protein interactions *(Horaruang & Zhang, 2017)*. Ubiquitin is split into two fragments (named Nub and Cub) and fused to proteins of interest. The Cub fragment is also fused to a transcription factor, LexA. Haploid strains of either Nub or Cub are generated, mated together, and assayed for growth on -his medium. This assay works because an interaction between the fused proteins results in reconstitution of the ubiquitin moiety, release of the LexA transcription factor, and expression of the gene HIS3, required for growth in -his medium. Therefore, growth on -his is indicative of a protein-protein interaction.

mbSUS strains expressing Cub-tagged PEEL-1 and Nub-tagged PMPL-1 did not grow on -his media, indicating these proteins did not interact (Fig. 4.1A and 4.1C). However, the fact that I could easily generate this strain and that it grows normally suggests that PEEL-1 and PMPL-1 do not kill yeast. This lack of killing activity raises the question of whether the reason these proteins do not kill yeast is the same reason why they do not interact in mbSUS, possibly due to yeast missing a co-factor for interaction and toxicity (e.g., a specific lipid). Alternatively, the yeast cell wall may suppress the cell swelling induced by toxicity or the tags may disrupt the functions of PEEL-1 and PMPL-1. Major results and interpretations are outlined below (Table 4.1).

Table 4.1. Results and interpretations from mbSUS for PEEL-1 and PMPL-1 interaction

Strain #	Cub construct	Nub construct	Growth on -his	Interpretation
1810	PEEL-1::Cub	NubG::PMPL-1	No	No interaction
1814		PMPL-1::NubG	No	No interaction
1851	PEEL-1(-94aa)::Cub	NubG	No	Not self-activating
1920		NubG::PMPL-1	Yes	True interaction?*
1846		NubG::PMPL-2	Yes	True interactor? Not a good negative control.
1728	PMPL-1::Cub	NubG	Yes	Self-activating Cub
1726		NubG::PEEL-1	Yes, but less than NubG	Ambiguous. Better negative control necessary
1731		PEEL-1::NubG	Yes, but less than NubG	Ambiguous. Better negative control necessary

*PEEL-1(-94aa)::Cub interacts with other constructs that wild-type PEEL-1::Cub does not interact with, including PMPL-1, PMPL-2, GADR-5(TM), and ZEEL-1(4TM) (see section on ZEEL-1 PEEL-1 mbSUS). This interaction suggests that removing over half of the PEEL-1 protein allows it to interact with more proteins than the wild-type PEEL-1. While this counterintuitive result be biologically relevant, it introduces some caution to interpreting data from this construct.

Bimolecular fluorescence complementation (BiFC) experiments can also test for interactions between proteins. The Venus fluorophore is split in two pieces (Venus-N and Venus-C) and fused to proteins of interest. The split-Venus fragments are not fluorescent on their own, but when in close proximity, a fluorescent, bimolecular complex is formed, with fluorescence read-out of protein-protein interaction. BiFC experiments have been performed mostly in HEK293T cells and complicates results when working with full-length proteins, since PMPL-1 and PEEL-1 kill these cells. To test for interaction between these proteins in cells that do not die, I have used non-toxic mutants of PEEL-1 and PMPL-1.

Since BiFC is known to have false-positive interactions, I attempted to control interactions between PEEL-1 and PMPL-1 by testing PEEL-1 with PMPL-1 paralogs (PMPL-2 and PMPL-7). A PEEL-1 loss-of-function mutant (D109A) interacts more with PMPL-1 than PMPL-2 (Fig. 4.2A). However, the difference in fluorescence is marginal, making interpretation of this result difficult. I also found that wild-type PEEL-1 interacts more with a truncated PMPL-1 (containing only C-terminal residues 28-59) than corresponding PMPL-2 and PMPL-7 truncations (based on alignments in Chapter 2) (Fig. 4.2B). This C-terminal PMPL-1 interaction with PEEL-1 is notable, since LDH assays using PMP3-like chimeras suggest that this same C-terminal region of PMPL-1 is sufficient for toxicity, as long as it is fused to any PMP3-like N-terminal half (discussed later) (Fig. 4.6B-C). A major caveat is that I have not tested whether PMP3-like::Venus-C constructs express at similar levels. Varied expression levels of PMP3-like constructs could explain these data.

I attempted co-immunoprecipitation experiments in HEK293T cells to assay PEEL-1 and PMPL-1 interaction but did not get a positive result. Here, an antibody targeting one protein allows isolation, i.e., immunoprecipitation (IP), of the target protein and co-immunoprecipitation

(co-IP) of the proteins it interacts with. Using an anti-GFP nanobody, I could successfully IP PEEL-1::eGFP. However, I was unable to detect co-IP of PMPL-1::mCherry (data not shown). This result may suggest PEEL-1 and PMPL-1 do not interact. However, co-IP experiments often must be optimized for membrane proteins. I used Nonidet P40 detergent to lyse cells, but this detergent may disrupt the very interactions I was hoping to detect. Alternatively, the interaction between PEEL-1 and PMPL-1 may be short-lived, so the use of crosslinkers prior to co-IP may be necessary. The cell death caused by PEEL-1 and PMPL-1 could also make it difficult to co-IP these proteins from HEK293T cells, so using loss-of-function mutants for co-IP experiments should be considered. The mutants used must be carefully chosen to ensure that the hypothesized protein-protein interactions are preserved.

4.1.2 Does PEEL-1 oligomerize?

My model of toxicity involves homotypic interactions among PEEL-1 proteins. mbSUS experiments did not support a homotypic PEEL-1 interaction (Fig. 4.3A). However, western blots of whole-cell lysate from bacteria expressing PEEL-1 with a maltose-binding protein tag (MBP) showed a high molecular weight band at about the expected size for an MBP::PEEL-1 dimer (Fig. 4.4A). Since whole-cell lysates have all the endogenous bacterial proteins, it seems unlikely the dimer band is due to PEEL-1 nonspecifically aggregating with proteins. The defined dimer band in whole cell lysate samples suggests a specific homotypic PEEL-1 interaction. After purifying MBP::PEEL-1, these dimer bands can still be seen (Fig. 4.4A-B). Interestingly, mass photometry of PEEL-1 in beta-octylglucoside detergent (β-OG, used for purification), does not show a dimer (Fig. 4.4C). If I instead dilute this same PEEL-1 in SDS, mass photometry shows that PEEL-1 oligomerizes (Fig. 4.4D). Therefore, SDS may induce the oligomerization of PEEL-

1. Since SDS is negatively charged, this SDS-induced oligomerization may suggest that negatively charged lipids permit or induce PEEL-1 oligomerization in cell membrane. However, whether these oligomers exist *in vivo* is still unknown.

One possible approach to testing oligomerization in HEK293T cells is through total internal reflection fluorescence (TIRF) microscopy, combined with single molecule photobleaching. By counting the number of photobleaching steps of a PEEL-1 complex at the plasma membrane, the oligomeric state of this protein in cells could be determined *(Senning and Gordon, 2015)*. However, this experiment may be difficult to interpret if PEEL-1 does not have a defined oligomeric state.

4.1.3 Does PMPL-1 oligomerize?

The molecular functions of PMPL-1 remain unknown, but its roles may involve homotypic interactions. I found that western blots of bacteria expressing MBP::PMPL-1 result in a band at the expected size of a homotypic dimer in both bacterial whole-cell lysate (Fig. 4.5A) and purified protein (Fig 4.5B). Mass photometry of purified MBP::PMPL-1 in beta-octylglucoside detergent shows multiple higher-order oligomers (Fig. 4.5C). Similar oligomers are seen with other detergents (lauryl maltose neopentyl glycol, n-decyl-beta-maltoside, and n-dodecyl-beta-maltoside), although to varying degrees (Fig. 4.5D). This detergent-dependent oligomeric state could suggest that PMPL-1 oligomerization is dependent on its nearby lipid and protein environment. Alternatively, this could be nonspecific aggregation *in vitro*, especially since mass photometry requires diluted samples which also dilute the detergent. It is still unclear whether these oligomers exist *in vivo* or whether it is important for PMPL-1's role in PEEL-1 toxicity.

mbSUS experiments for PMPL-1 homotypic interactions suggest that PMPL-1 interacts with itself, as well as PMPL-2 (Fig. 4.3B). I am cautious in interpreting this result as PMPL-1 interacting with itself and with PMP3 proteins broadly, since I have seen PMPL-1::Cub interacts with many membrane proteins, shown in Fig. 4.8C.

4.1.4 Which residues are critical for PMPL-1's role in toxicity?

Although PMPL-2 is 75% similar to PMPL-1 (Fig. 4.6A), it does not kill HEK293T cells when co-expressed with PEEL-1. Chimeras between these closely related PMP3-like proteins can therefore be used to determine which residues are important for toxicity. I found that the C-terminal half of PMPL-1 (PMPL-1(aa27-59)) is most important for toxicity (Fig. 4.6B). Expression of the C-terminal region alone is not sufficient for toxicity with PEEL-1. However, fusing this PMPL-1(aa27-59) to the N-terminal halves of either PMPL-2, PMPL-7, or yeast PMP3 allows for toxicity with PEEL-1 (Fig. 4.6C). These data suggest a model where the C-terminal end of PMPL-1 allows for specific interaction with PEEL-1, and the N-terminal end provides some important structural feature to the protein. This model is corroborated by preliminary BiFC data presented earlier (Fig. 4.2B), where PEEL-1 interacts more with PMPL-1(aa29-59) than with C-terminal truncations of other PMP3-like proteins.

Future directions

I have already outlined some future directions which test the main mechanistic hypotheses (see above). Below I discuss possible directions to further probe ion channel activity, protein interactions, and pore-lining residues.

4.1.5 An alternative approach for assaying the PEEL-1/PMPL-1 ion channel

My previous experiments used planar lipid bilayers to test for PEEL-1 and PMPL-1 ion channel activity (see Chapter 2). Planar lipid bilayers are a powerful tool to study ion channels since single channel activity can be measured. However, planar lipid bilayer experiments are time consuming, are variable day-to-day, and require extensive technical expertise. Finding an alternative assay that is higher throughput can help validate previous experiments and address new biochemical questions like: What is the ion selectivity series of the channel? Does mutating the putative selectivity filter (D109) change ion selectivity? Are there specific lipids that promote ion channel activity?

A liposome flux assay is a possible alternative approach to *in vitro* PEEL-1 and PMPL-1 experiments *(Su et al., 2016)*. This approach sacrifices the single channel resolution provided by planar lipid bilayer. Therefore, the liposome flux assay can average out variability, between single proteins or channels, making results easier to interpret. Liposome flux assays have been used to test for activity of channels conducting potassium, chloride, and protons *(Su et al., 2016, Cabanos et al., 2017, Dickson et al., 2014, Liu et al., 2023)*. Liposome flux assays rely on an intraliposomal, pH sensitive fluorophore impermeable to lipid bilayers, AMCA. The experiment works in four steps: (i) equilibration of ion channel-containing liposome in a buffer with a non-permeable cation, (ii) addition of a proton ionophore (CCCP) to allow for the passage of protons into the liposome, (iii) movement of ions from the lumen of the liposome into the extra-liposomal buffer, and (iv) countermovement of protons into the liposome, thus acidifying the intraliposomal solution and quenching of the fluorophore. Therefore, measuring fluorescence over time provides a proxy for ion movement. Addition of monensin at the end of the assay

allows a path for sodium to travel through the liposome, thus providing a maximum fluorescence quenching for normalization of fluorescence data.

I attempted to use the liposome flux assay to test for ion channel activity of PEEL-1 and PMPL-1, but I faced many technical challenges, possibly due to the many moving pieces in this assay. I created liposomes containing PEEL-1 alone, PMPL-1 alone, both proteins together, and no proteins. I used Na^+ as the permeable cation inside the liposomes and diluted liposomes in buffer containing $NMDG^+$, the impermeable cation. All protein-containing liposomes caused significant fluorescence quenching, suggesting that both PEEL-1 alone and PMPL-1 alone allow for Na^+ to leave the liposomes (Fig. 4.7A-B). Furthermore, even before the addition of CCCP, some experiments showed significant quenching (Fig. 4.7B), suggesting that these proteins allow H^+ conduction through the liposomes. Liposomes containing both proteins did often quench more than each protein alone, but this result is difficult to interpret since the intended negative controls do not behave as expected (Fig. 4.7A-B). The amount of quenching from aliquots of the same liposome prep was variable depending on the day and the material of the 96-well plate used (for example, polystyrene or polypropylene). Some possible interpretations of the data from these experiments are described in Table 4.2.

Table 4.2. Interpretations and counter arguments explaining results from liposome flux assays.

Possible interpretation	Counterargument
PMPL-1 weakens the liposome, allowing ions through	Acidification of a compartment is required for fluorescence quenching. Since we see quenching, it suggests there is an isolated compartment in this assay.
PEEL-1 creates a channel on its own	None. But means that additional controls may be needed, such as PEEL-1(-65aa).
PMPL-1 creates a channel on its own	None. But means that additional controls may be needed, such as PMPL-2.

The lipid composition of the liposome is sub-optimal (50% DOPC/POPS)	I repeated the experiment once, using soybean lipid extract with cholesterol, and saw similar results in the liposome assay.
Technical issue. I have seen variability depending on the material of the plate	None. Need a good positive and negative control for this experiment, with proteins known to/not conduct ions through a liposome.
The experiment is measuring something other than ion movement through the liposome.	Diluting liposomes in Na^+ or K^+ instead of the impermeable cation $NMDG^+$ result in minimal quenching. Therefore, quenching only occurs when the extraliposomal buffer contains an impermeable cation.

Other approaches are possible to test for ion flux through liposomes *in vitro*. One approach is incorporating membrane-impermeable, fluorescent ion indicators into liposomes. This fluorescence read-out provides a more direct measure of ion flux through liposomes compared to my current flux assay. Similar approaches have been used for Zn^{2+} channels *(Gati et al., 2017)*. A different approach uses radioactivity to measure ion movement into liposomes, using either $^{22}Na^+$ for sodium channels (Villegas et al., 1980) or $^{86}Rb^+$ for potassium channels *(Nimigean 2006)*. Radioactivity may be the most direct approach for measuring ion movement into liposomes. However, a disadvantage to this approach is that it requires running samples through a cation exchange column in order to specifically measure intraliposomal radioactivity, making this an endpoint rather than a kinetic assay. In personal communications with Jakub Sliwinski (TraceLab, UW Oceanography), one possible approach is by using non-radioactive Rb and inductively coupled mass spectrometry (ICP-MS) as a readout for Rb concentration in liposomes.

4.1.6 Determine important residues in PEEL-1 and PMPL-1

Deep-mutational scans (DMS) can be used to identify important residues in a protein. This approach is versatile since selection on the DMS library can be modified to test different aspects of a model. For example, to determine which PEEL-1 residues are important for toxicity, HEK293T cells expressing PEEL-1 DMS variants can be grown in bulk. By inducing PMPL-1 expression, functional variants cause cell death and are lost from the population over time. By sequencing pools of cells after several generations, you would expect an enrichment of PEEL-1 mutants which have lost their toxic activity. This would be a great system to thoroughly test whether the amphipathic property of the PEEL-1 AH correlates with toxicity. DMS can also be used to test for interactions between PEEL-1 and PMPL-1. Here, DMS can be coupled with BiFC, and fluorescence can be used as a selective force by fluorescence-activated cell sorting. This would allow us to find mutants which increase or decrease protein-protein interactions. Similar approaches can be used to test for homotypic protein interactions.

4.1.7 Determine the pore-lining residues of PEEL-1

I have data consistent with the PEEL-1 amphipathic helix constructing the lining of the cation channel, but this hypothesis can be tested more rigorously. The substituted cysteine accessibility method (SCAM) can be used to determine the pore-lining residues of an ion channel *(Liapakis et al., 2001)*. It relies on the principle that adding a bulky residue to the lining of the channel will change a specific channel property. Importantly, the bulky residue cannot abolish channel activity entirely, nor just symmetrically reduce its current. Either the channel rectification or ion selectivity must be altered to support the model. Bulky residues can be reversibly introduced by adding a thiol reagent such as 2-[(methylsulfonyl)thio]-ethanesulfonic acid (MTSES) which will covalently bind to cystines that are accessible from the extracellular

solution. A down-side to these experiments is that they would likely need to be performed using electrophysiology, which can be technically challenging because of the cell swelling and death caused by PEEL-1 and PMPL-1.

4.2 MECHANISM OF ZEEL-1 ANTIDOTE ACTIVITY

My current model of antidote activity is that ZEEL-1 and PEEL-1 interact through their transmembrane domains, resulting in ubiquitylation and degradation of PEEL-1. Although I have reconstituted antidote activity in HEK293T cells (see Chapter 2), the precise mechanism of antidote activity remains unknown.

4.2.1 Do PEEL-1 and ZEEL-1 interact?

My model of ZEEL-1 antidote activity involves a direct protein-protein interaction between PEEL-1 and ZEEL-1. mbSUS experiments support this model, showing that PEEL-1 interacts with both full-length ZEEL-1 and a truncated ZEEL-1 with only contains the transmembrane domains (ZEEL-1(TM)) (Fig. 4.8A). PEEL-1 interaction with ZEEL-1 is specific, since PEEL-1 does not interact with GADR-5(TM) in this assay (the most similar ZEEL-1 paralog) (*Seidel et al., 2011*) (Fig. 4.8A). Removing the last two ZEEL-1 TM domains eliminates its interaction with PEEL-1 (Fig. 4.8B), illustrating their requirement for this interaction. I hypothesize that the interaction between PEEL-1 and ZEEL-1 is required for antidote activity, but I have not tested this.

In addition to ZEEL-1 interaction with PEEL-1, it is possible that it also interacts with PMPL-1 as part of its antidote mechanism. However, mbSUS experiments that test ZEEL-1 interactions with PMPL-1 are difficult to interpret. PMPL-1::Cub interacts with both ZEEL-

1(TM) and GADR-5(TM) (Fig. 4.8C), as well as most other membrane proteins I have tested. This broad interaction might suggest that PMPL-1::Cub is prone to false-positive interactions. Alternatively, if these results are biologically meaningful, PMPL-1 may indeed interact broadly with membrane proteins. Differentiating between these two possibilities requires more experiments and possibly different assays. A summary of results is shown in Table 4.3.

Table 4.3. Results and interpretations from mbSUS for ZEEL-1 interactions.

Strain #	Cub construct	Nub construct	Growth on -his	Interpretation
1785	PEEL-1::Cub	NubG	No	No self-activation of Cub
1806		NubG::ZEEL-1	Yes	Interaction
1827		NubG::ZEEL-1(TM)	Yes	Interaction in TM domains
1887		NubG::GADR-5(TM)	No	No interaction. PEEL-1 interaction is specific to ZEEL-1
1914		NubG::ZEEL-1(4TM)	No	Last two TM domains of ZEEL-1 are required for interaction
1728	PMPL-1::Cub	NubG	Yes	Self-activating Cub
1828		NubG::ZEEL-1(TM)	Yes (more than NubG)	Interaction
1888		NubG::GADR-5(TM)	Yes (more than NubG)	Interaction. Not a good negative control?

4.2.2 Does ZEEL-1 ubiquitylate PEEL-1 for degradation?

My current model of antidote activity involves ZEEL-1-mediated ubiquitylation and degradation of PEEL-1. I have not tested this hypothesis, but there are several avenues to do so. I have already shown that mCherry::ZEEL-1 suppresses toxicity from PEEL-1::eGFP in HEK293T cells. Flow cytometry or western blots can be used to measure levels of PEEL-1::eGFP when transfected alone or in combination with mCherry::ZEEL-1. Any effect from co-

transfection can be controlled for by using mCherry::GADR-5, which does not suppress toxicity. The addition of a proteasome inhibitor like MG132 should prevent any proteasome-mediated degradation. Furthermore, finding that E3 ubiquitin ligase machinery can co-IP with ZEEL-1 would suggest that ubiquitylation is involved in the antidote mechanism. Similar co-IP experiments have been successful with GADR-5 in HEK293T cells *(Sawyer et al., 2011)*. An alternative hypothesis of the antidote mechanism may be that ZEEL-1 acts by disrupting homotypic PEEL-1 interactions or interactions between PEEL-1 and PMPL-1.

4.2.3 Deep mutational scan to determine interacting residues between PEEL-1 and ZEEL-1

mbSUS can be a platform for further probing the interaction between PEEL-1 and ZEEL-1. Mutants in either protein which reduce their interaction should result in lack of growth on assay plates (-his). In theory, these mbSUS experiments can be scaled-up and mutants can be assayed in bulk, using sequencing to assay for relative growth of mutants. This approach may find missense mutations that eliminate the interaction between ZEEL-1 and PEEL-1. Determining whether these same mutants result in loss of antidote activity in HEK293T cells or in *C. elegans* would allow us to more rigorously test whether interactions between PEEL-1 and ZEEL-1 are required for antidote activity.

4.3 EVOLUTION OF *PEEL-1 ZEEL-1*

Although we understand how many selfish genetic elements spread in populations, we often do not know how they first evolved. I hypothesize that animal TA systems face many hurdles during their evolution, requiring more complexity than fungal or prokaryotic TA systems.

These hurdles include evolving germline-specific expression of the toxin, as well as preventing toxicity in off-target cells.

4.3.1 Where did *peel-1 zeel-1* come from?

The evolutionary origin of *peel-1 zeel-1* remains a mystery. The *peel-1 zeel-1* TA has only been found in *C. elegans* and is absent in many strains within this species *(Seidel et al., 2008, Seidel et al., 2011)*. Both PEEL-1 and ZEEL-1(TM) are evolutionarily novel, and their evolution within *C. elegans* cannot be obviously traced. How did these genes evolve? An intriguing idea is that they were horizontally transferred from another organism. Horizontal gene transfer events are rare and often likely lost by genetic drift, but selfish activity can overcome this evolutionary hurdle. Recent work also provides a precedent for horizontal gene transfer of a toxin-antidote system in animals *(Widen et al., 2023)*. Finding transposable elements which flank the *peel-1 zeel-1* genomic locus may lend support to this hypothesis, but this has not been found. Without a DNA sample from another organism which shows homology to *peel-1 zeel-1*, it would be difficult to confidently determine whether these genes evolved *de novo* in *C. elegans* or whether they were horizontally transferred from another organism.

4.3.2 Why does PEEL-1 rely on another protein for toxicity?

One of the simplest mechanisms for a population to gain resistance to the *peel-1 zeel-1* selfish element might be by mutating *pmpl-1*. *pmpl-1* mutant worms are perfectly viable in laboratory conditions, so this may be a feasible evolutionary route to suppressing the TA (although the high conservation of *pmpl-1* suggests it is important in nature). But might there be some benefit for *peel-1 zeel-1* to require a second protein for its selfish activity? Although PEEL-

1 toxicity is critical for the success of the selfish element, the precise timing of toxicity is important. I hypothesize that relying on PMPL-1 allows for precise developmental timing of PEEL-1 toxicity.

There are two critical times where PEEL-1 protein should not kill: the sperm and the early embryo. Sperm carry the toxin to the progeny, so sperm viability is important to the success of the selfish element. Early embryonic development is mostly guided by maternally deposited transcripts. PEEL-1 toxicity must occur after the maternal-to-zygotic transition to allow time for zygotic expression of ZEEL-1. A route to prevent off-target killing in both cases is possible by requiring a second component that is only present in the target developmental stage or tissue. All of this means that PMPL-1 may not have been co-opted by PEEL-1 because of its endogenous functions, but instead, because of the developmental timing and tissues in which it is (and is not) expressed. Just PEEL-1's ability to detect the presence of PMPL-1 can serve as a trigger for PEEL-1 to turn from dormant to toxic. The *sup-35 pha-1* toxin-antidote system in *C. elegans* also requires genes that are expressed in specific tissues *(Ben-David et al., 2017)*. Tight control of the timing of toxicity may be a hurdle that TA systems must evolve in order to be successful genetic parasites. Co-option of endogenous proteins may be a common and evolutionarily accessible path for preventing off-target toxicity in animal toxin-antidote systems. Therefore, studying the mechanisms of toxin-antidote systems can provide insights into the origins and evolution of toxin-antidote systems.

4.4 METHODS

4.4.1 Mating-based split ubiquitin system (mbSUS)

mbSUS strains (gift from Dana Miller) were generated in haploid yeast, strain THY.AP4 (for Cub constructs) or strain THY.AP5 (for Nub constructs). Strains were streaked onto YPD plates, grown at 30°C for two days, and scraped off plates for transformation in PLATE solution (40% PEG, 100 mM LiOAc, 10mM Tris-HCl 7.5, 1mM EDTA). 10µL of 10µg/mL boiled salmon sperm DNA was added to cells, along with ~600ng of the target construct, and cells were vortexed. After 1-6 hours at room temperature, cells were heat shocked in a 42°C water bath for 15 minutes. Cells were pelleted (1,000 x *g*, 1 min), washed with sterile water, and plated on selective agar media, either -leucine (-L) for Cub constructs or -tryptophan (-W) for Nub constructs. Plates were placed at 30°C for 2-3 days. Multiple colonies were pooled together, grown in appropriate selective liquid medium (overnight, 30°C), and frozen at -80°C for later use.

To generate the desired diploid strain to test for interactions between Nub-tagged and Cub-tagged proteins, haploid Nub strains were mated with haploid Cub strains. 1mL of overnight Nub or Cub cultures (30°C) were pelleted (1,000 x *g*, 5 min) and resuspended in 100µL YPD. 10µL of Cub culture was mixed with 10µL of Nub culture, spotted onto YPD agar plates, and incubated at 30°C for 6-8 hours. Cells were then scraped off the plate and used to inoculate 4-5mL of selective liquid media (-WL). Cultures were place in a 30°C shaker overnight prior to the spot assay.

To assay diploid strains for growth on different media, a spot assay was used. The density of overnight liquid cultures (30°C, shaking) was determined by measuring the OD_{600} of a 1:10 dilution of the culture in water. In parallel, cells were pelleted (2,000 x *g*, 2 min) and resuspended in water to a final $OD_{600} = 1.0$. Serial dilutions (10-fold) were made and spotted (5µL) onto multiple plates. Plates lacking tryptophan and leucine (-WL) were control plates and

plates also lacking histidine (-WLH) were assay plates. Different methionine concentrations were also tested since increasing methionine concentrations reduces expression of the Cub construct. Plates were incubated at 30°C for 2-3 days before growth was recorded and imaged.

4.4.2 Bimolecular fluorescence complementation (BiFC)

BiFC experiments were performed using transient transfection in HEK293T cells, followed by flow cytometry. Approximately $7x10^4$ cells were plated in each well of a 12-well plate ~24 hours prior to transfection. Transfections were performed using combinations of two BiFC constructs (one Venus-N tagged, one Venus-C tagged, 750ng each, total 1.5µg DNA with 3µg PEI). 40-48 hours after transfection, culture supernatant was removed, and adhered cells were collected for flow cytometry. Cells were resuspended by 5mM EDTA in 1X PBS (rocked for 10 min), pelleted in microcentrifuge tubes (500g for 5 min), and washed with 1X PBS. After resuspension in 1X PBS, cells were subjected to flow cytometry (Symphony S6). Forward scatter, side scatter, and YFP fluorescence was measured for 10,000 events. Gating was performed based on forward scatter and side scatter, although raw, ungated datasets showed similar results to gated datasets. Histograms were generated at floreada.io.

All constructs were made in an mCherry_N1 backbone with stop codons after the BiFC construct, followed by mCherry. This downstream mCherry is expressed at low levels. This was not intentional. mCherry_N1 alone was tested for fluorescence in this experiment, and I confirmed it did not emit in the YFP channel in flow cytometry.

4.4.3 Mass photometry

Mass photometry experiments were performed on a Refyn Two MP. MBP::PEEL-1::His_8 and MBP::PMPL-1 were purified as described in Chapter 2, resulting in purified protein in Buffer A (HEPES pH 7.4, 150mM NaCl, 5mM 2-mercaptoethanol, 10% glycerol) with beta-octylglucoside detergent (1%) and maltose (10mM). Prior to each experiment, glass slides were washed with water and isopropanol, and dried under a nitrogen stream. Oil (Immersol 518F) was spotted onto the lens and the glass slide was added on the oiled lens. 18µL of 0.2-µm filtered Buffer A (without detergent) was spotted onto the glass slide and put into focus using the "drop dilution" mode on the Acquire MP software. A 1-minute movie was recorded to confirm low background signal. ~2µL of protein was added directly to the buffer on the slide, mixed by pipetting, and the movie was immediately recorded. A target final protein concentration of approximately 50nM was used. Serial dilutions were performed in Buffer A when necessary. Discover MP software was used to generate histograms and Gaussian fits to the data. Molecular weights were determined based on a standard curve generated by beta-amylase (BAM) protein.

4.4.4 Liposome flux assay

Proteo-liposomes were made with MBP::PEEL-1::His_8, MBP::PMPL-1, or both. Proteo-liposomes were made as described in Chapter 2, but an increased salt concentration was used during purification and liposome formation (400mM NaCl). Proteo-liposome lipid compositions were either 50% DOPC:POPS (1,2-dioleoyl-sn-slycero-3-phosphocholine: 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine) or soybean lipids (Avanti Polar Lipids) with cholesterol. After liposome formation, flotation, and collection (see Chapter 2), liposomes were snap frozen in liquid nitrogen and kept a -80°C. Liposomes were thawed fresh before each experiment. Outside flux buffer was made by mixing Buffer B (20mM HEPES pH 7.4, 400mM N-methyl-D-

glucamine, 1mM EDTA) mixed with ACMA (in DMSO) (final concentration 2μM ACMA) before each experiment. 190μL of Outside flux buffer was added to a well of a black 96-well plate (black, 96-well, conical bottom, Nunc catalog #249945). Fluorescence was measured (excitation 419nm, emission 490nm, on a Spectra Max Gemini XPS) for 1 minute in 20-second (or 30-second) intervals, and 10uL of proteo-liposomes were added and mixed by pipetting. Fluorescence was measured in 20-second intervals until fluorescence levels stabilized. 1μL of 0.4mM CCCP (in DMSO) was added and mixed by pipetting. Fluorescence was again measured in 20-second intervals until fluorescence levels stabilized. 1μL of 5mM monensin (in DMSO) was added and mixed. Fluorescence was measured for 1 minute in 20-second intervals. Each fluorescence data point was normalized using the following calculation: normalized fluorescence = (fluorescence at data point – final fluorescence) / (initial fluorescence – final fluorescence). Two time points could be used for initial fluorescence, either fluorescence prior to liposome addition or fluorescence prior to CCCP addition. Data points after monensin addition were averaged to get a value for final fluorescence.

4.5 FIGURES

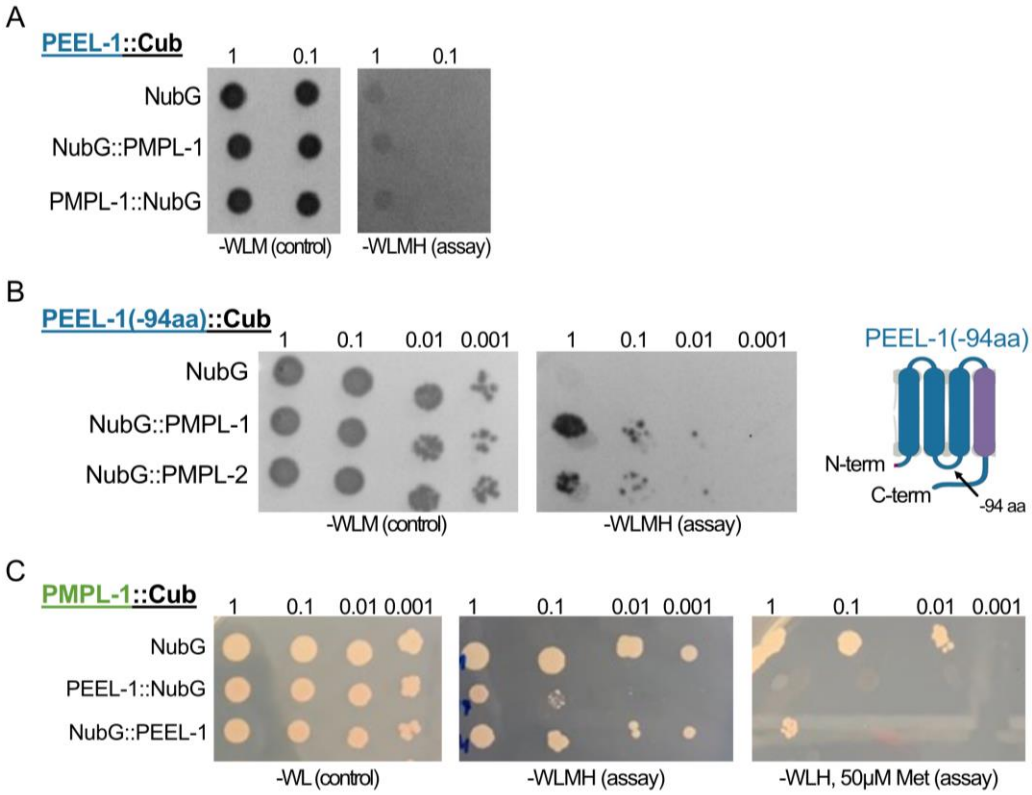


Figure 4.1. mbSUS assaying PEEL-1 and PMPL-1 interactions.

(A-C) Three experiments assaying PEEL-1 and PMPL-1 interaction using mbSUS in yeast. Yeast with indicated Cub construct were mated to yeast with indicated Nub construct (NubG is an untagged control). Resulting diploids were spotted on agar media with the indicated amino acids absent from the media (bottom; "-", W = tryptophan, L = leucine, M = methionine, H = histidine). 10-fold serial dilutions (ex. 0.1 indicates a 1 in 10 dilution) were spotted and an image of their growth is shown after 2-3 days at 30C. Left panels: -WL or -WLM = control (selects for Cub and Nub plasmids). Right panels: -WLMH = assay plates. All experiments were repeated at least three times and had similar results. **(B)** A truncated, PEEL-1(-94aa) Cub construct is used. Diagram on right shows the predicted structure of PEEL-1. indicated the -94aa. The -94aa truncation removed the PEEL-1 amphipathic helix (purple) and the third transmembrane domain.

(C) PMPL-1::Cub assayed with PEEL-1, showing growth in negative control (NubG). Increased methionine (M) concentrations reduce expression levels of Cub constructs. Two assay plates are shown, -WLMH and -WLH(50μM Met).

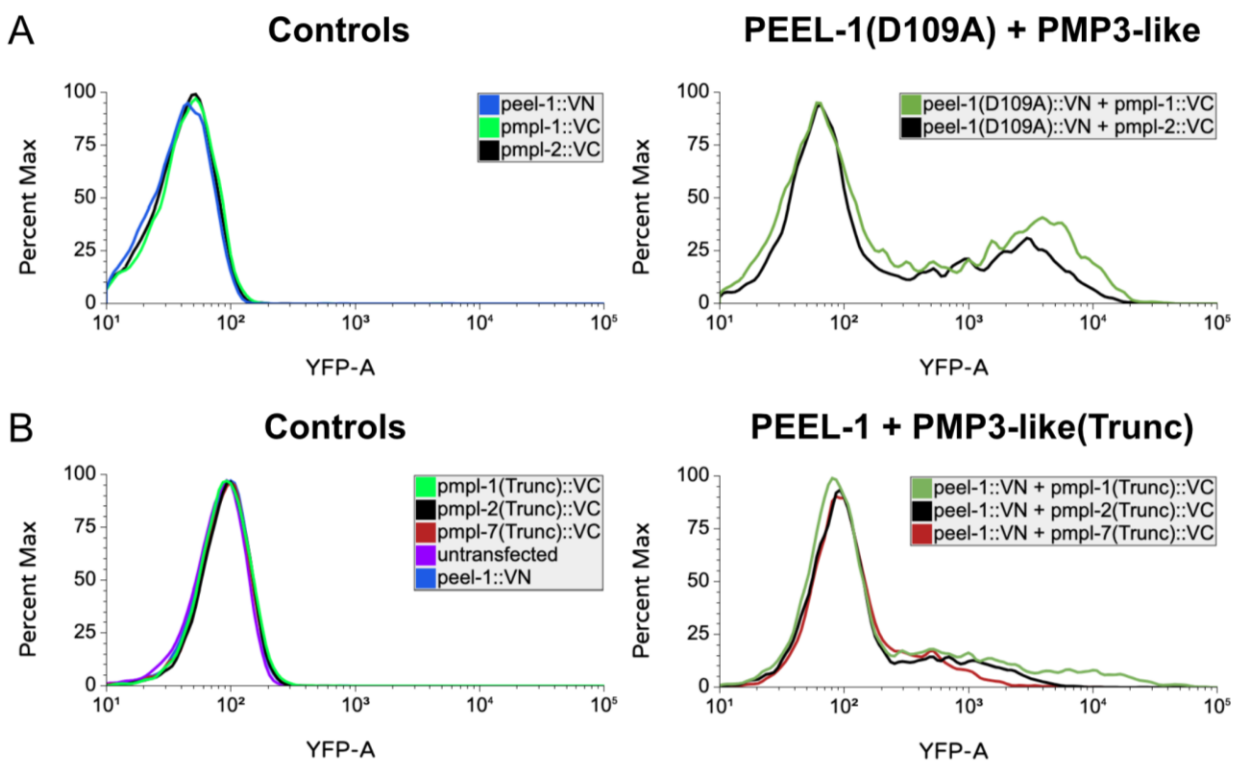


Figure 4.2. BiFC assaying PEEL-1 and PMPL-1 interactions.

Two examples of BiFC experiments performed in HEK293T cells to test for PEEL-1 and PMPL-1 interaction. Bimolecular fluorescence of Venus was assayed using a flow cytometry about 41 hours after transfection. Histograms of yellow fluorescence (YFP-A, x-axis) are shown as a percent max of the largest peak within each dataset (Percent Max, y-axis). Each construct is shown transfected alone (left, control) or together (right). **(A)** PEEL-1(D109A)::Venus-N tested for interaction with indicated PMP3-like::Venus-C. **(B)** PEEL-1::Venus-N interaction tested with indicated PMP3-like(Trunc)::Venus-C. Truncated, C-terminal-only of PMP3-like protein corresponds to PMPL-1(aa27-59) in alignments.

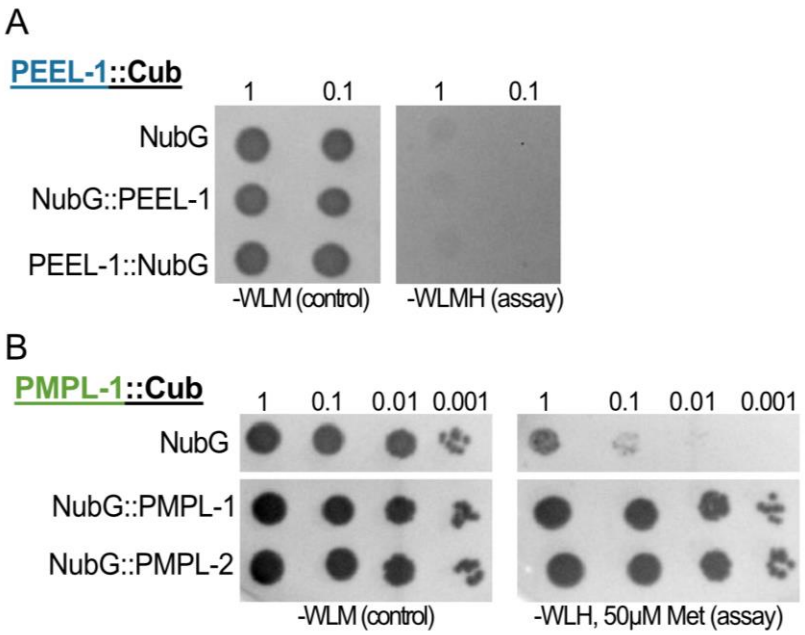


Figure 4.3. mbSUS for homotypic PEEL-1 and PMPL-1 interactions.

mbSUS experiments assaying homotypic interactions in **(A)** PEEL-1 and **(B)** PMPL-1. Yeast with indicated Cub construct were mated to yeast with indicated Nub construct and serial dilutions were plated. Control plates (-WLM) and assay plate (-WLH, with either -M or 50µM Met) are shown. **(A)** No growth is seen in assay plates of these diploids. **(B)** PMPL-1::Cub interacts with NubG::PMPL-1 and NubG::PMPL-2. All experiments were repeated at least three times.

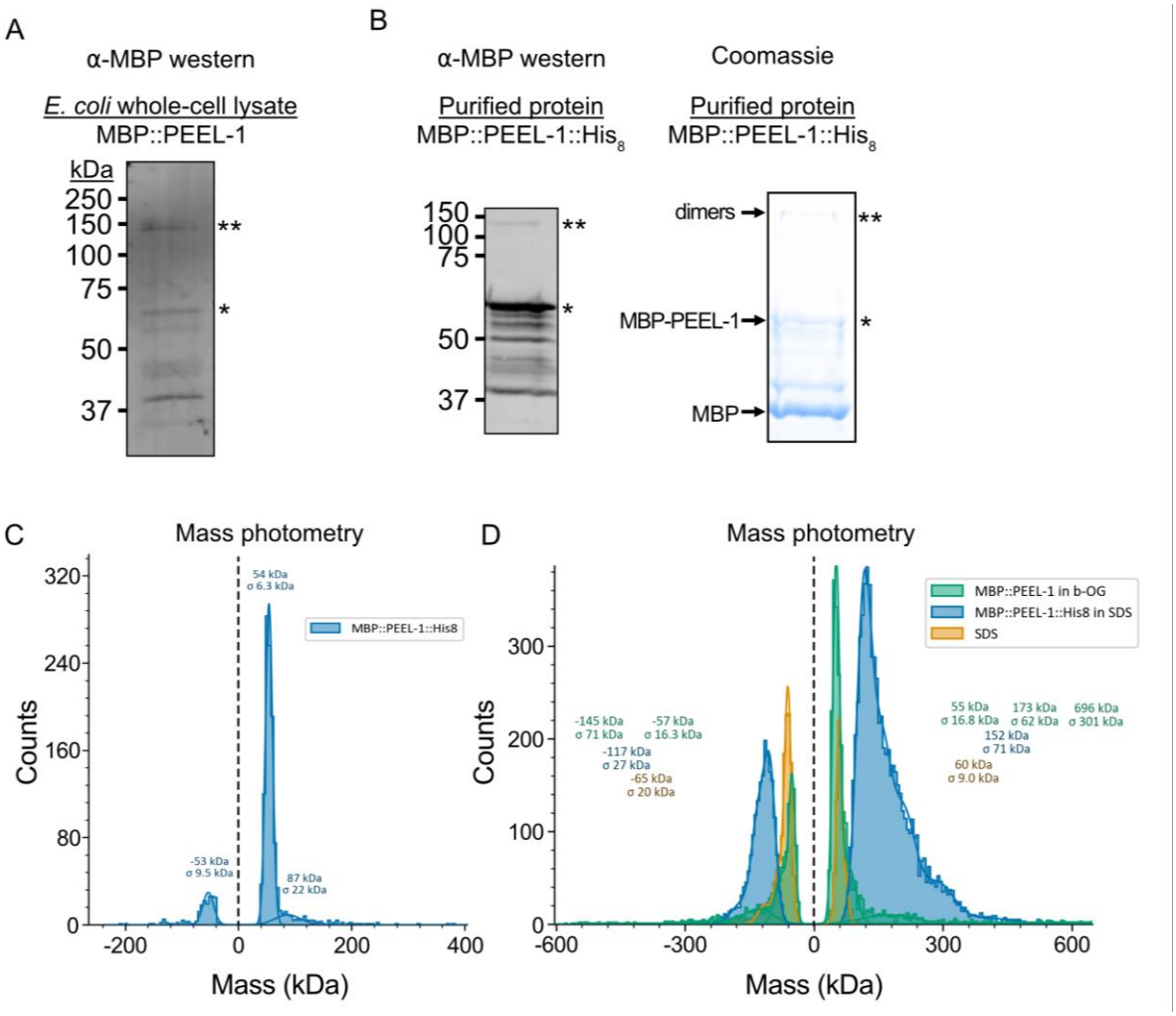


Figure 4.4. PEEL-1 forms SDS-resistant oligomers.

(A-B) SDS-PAGE gels with indicated PEEL-1 samples. Monomers (*) and putative SDS-resistant, dimers (**) are indicated. Monomers of MBP::PEEL-1 are predicted to be 64 kDa. **(A)** Whole-cell lysate anti-MBP western blot of *E. coli* C41(DE3) cells with MBP::PEEL-1 expression induced with 0.5mM IPTG overnight at 18C. **(B)** Western blot (left, anti-MBP) and coomassie stain (right) of purified MBP::PEEL-1::His_8. **(C-D)** Mass photometry of purified MBP::PEEL-1::His_8. Histogram of masses are shown with Gaussian fits. Predicted molecular weight (kDa) and standard deviation (σ) of each peak is shown. Negative masses can be ignored. MBP::PEEL-1::His_8 in beta-octylglucoside is shown in (C). MBP::PEEL-1::His_8 in either beta-

octylglucoside (green) or SDS (blue) is shown in (D). Addition of SDS increases the molecular weight of PEEL-1 protein in solution, possibly through oligomerization or aggregation. Masses of SDS micelles alone, without protein, are also shown (control, yellow).

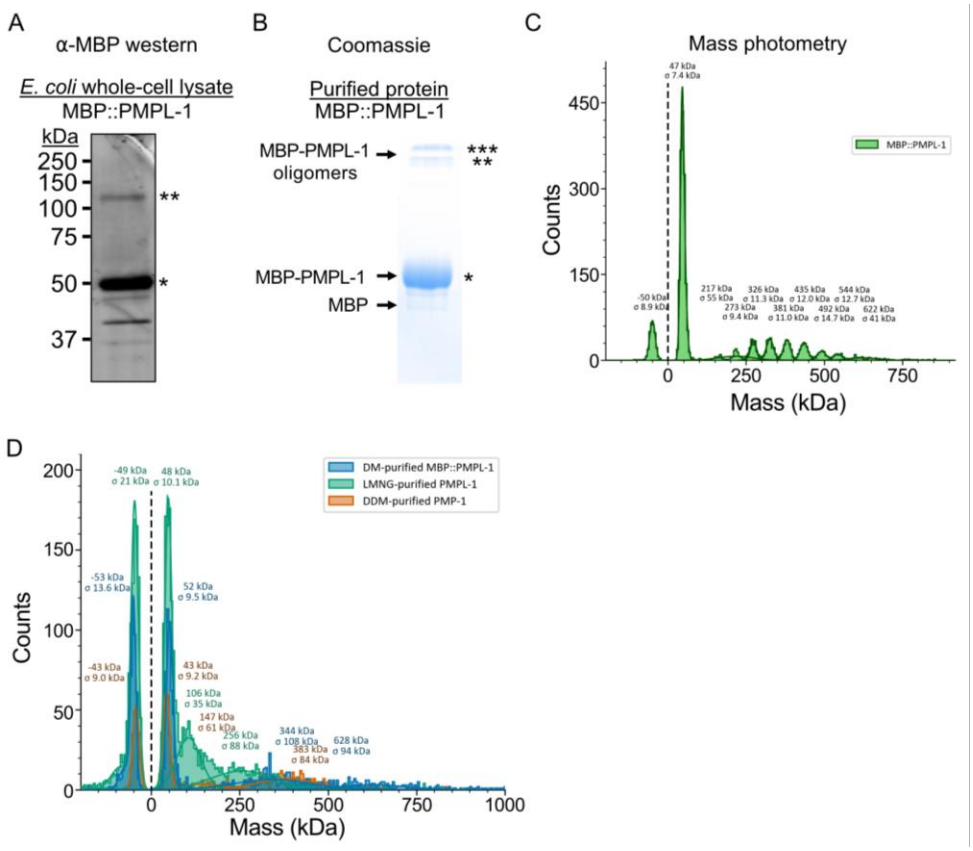


Figure 4.5. PMPL-1 forms SDS-resistant oligomers.

(A-B) SDS-PAGE gels with indicated PMPL-1 samples. Monomers (*) and putative SDS-resistant, dimers (**) and higher-order oligomers (***) are indicated. Monomers of MBP::PMPL-1 are predicted to be 50 kDa. **(A)** Whole-cell lysate anti-MBP western blot of *E. coli* C41(DE3) cells with MBP::PMPL-1 expression induced with 0.5mM IPTG. **(B)** Coomassie stain of purified MBP::PEEL-1. **(C)** Mass photometry of purified MBP::PMPL-1 in beta-octylglucoside (dilution required, resulting in ~0.01% detergent concentration). Histogram of masses are shown with Gaussian fits. Predicted molecular weight and standard deviation of each peak is shown. Negative masses can be ignored. **(D)** Mass photometry of MBP::PMPL-1 purified in n-decyl-beta-maltosome (DM, blue), lauryl maltose neopentyl glycol (LMNG, green), or n-dodecyl-beta-maltoside (DDM, orange).

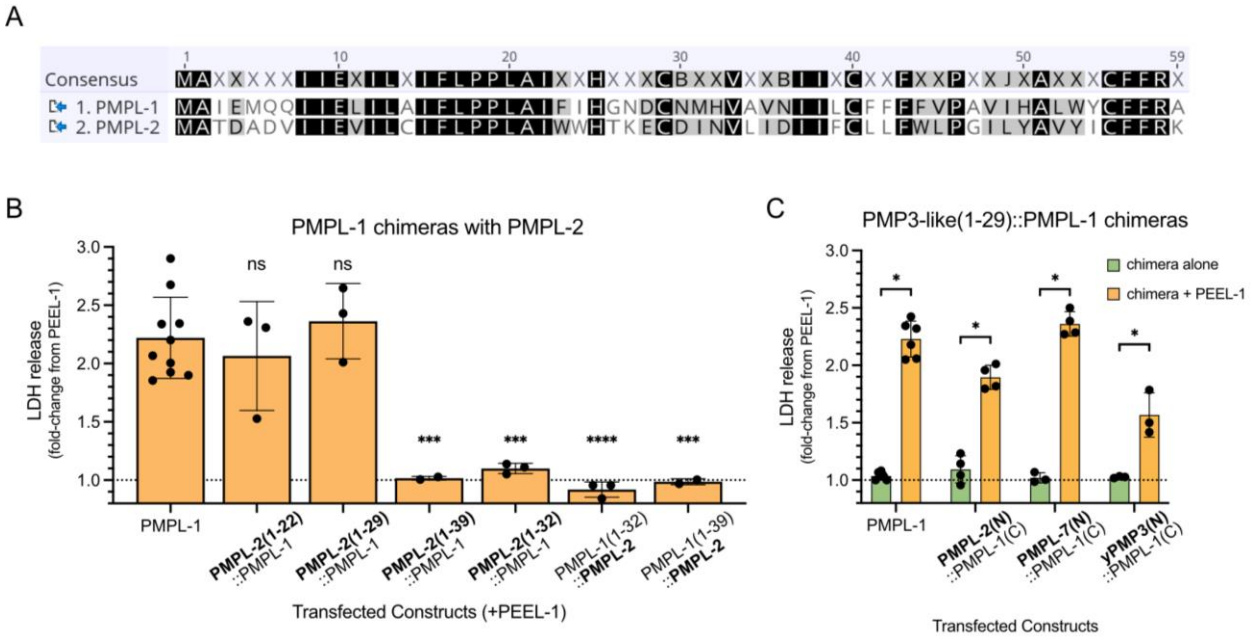


Figure 4.6. Chimeric PMP3-like proteins in HEK293T cells.

(A) Alignment of PMPL-1 and PMPL-2 amino acid sequence from Geneious. Proteins are 47.5% identical and 74.6 % similar. Black boxes indicate identical residues, and gray boxes indicate similar residues. Consensus sequence shown at top. **(B)** LDH assays of PMPL-1 and PMPL-2 chimeras co-transfected with PEEL-1 in HEK293T cells. Individual data points, mean, and SD are shown. Dotted line indicates no killing above PEEL-1 control. Numbers in parentheses corresponds to the amino acids of the indicated PMP3-like protein N-terminus (ex. PMPL-2(1-22)::PMPL-1 indicates a construct coding for the first 22 residues of PMPL-2, fused to residues 23-59 of PMPL-1). **(C)** PMP3-like chimeric constructs transfected alone (green bars) or with PEEL-1 (yellow bars). All chimeras test PMP3-like(N), corresponding to PMPL-1(1-29), fused to PMPL-1(aa30-59) (labeled "PMPL-1(C)"). PMP3-like(N) regions were determined from alignments which match PMPL-1 amino acids 1-29 (ends at a conserved Cys). Statistics in **(B)**: one-way ANOVA with Dunnett's multiple comparisons test. Statistics in **(C)**: multiple unpaired t-tests with Holm-Šídák test, comparing each chimera alone to the same chimera with PEEL-1. PMPL-1 was mCherry tagged or untagged, but all chimeras were untagged. The

yPMP3(N)::PMPL-1 construct had an unintended, two amino-acid insertion in the N-terminus (in-frame, occurred in two independent clones).

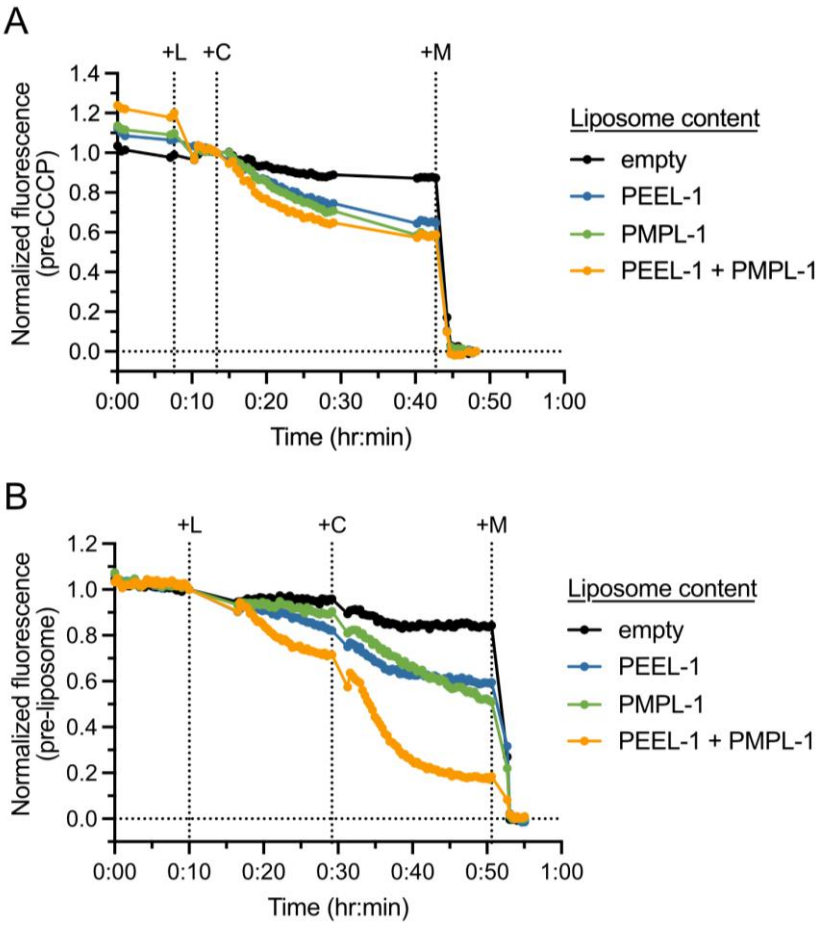


Figure 4.7. Liposome flux assay to measure ion channel activity.

Two examples of the liposome flux assay, measuring fluorescence levels over time in a 96-well plate. Timepoints at which liposomes were added (+L), CCCP (+C), and monensin (+M) are indicated. One fluorescence measurement was recorded every 20 seconds with some gaps occurring due to addition of reagent or instrument error. Normalized fluorescence was calculated by comparing fluorescence after monensin addition to fluorescence before CCCP addition in **(A)** or before liposome addition in **(B)**. The protein content of liposomes are indicated (right). A single experiment is shown in **(A)** and an average of two technical replicates is shown in **(B)**. Technical replicates in (B) were similar.

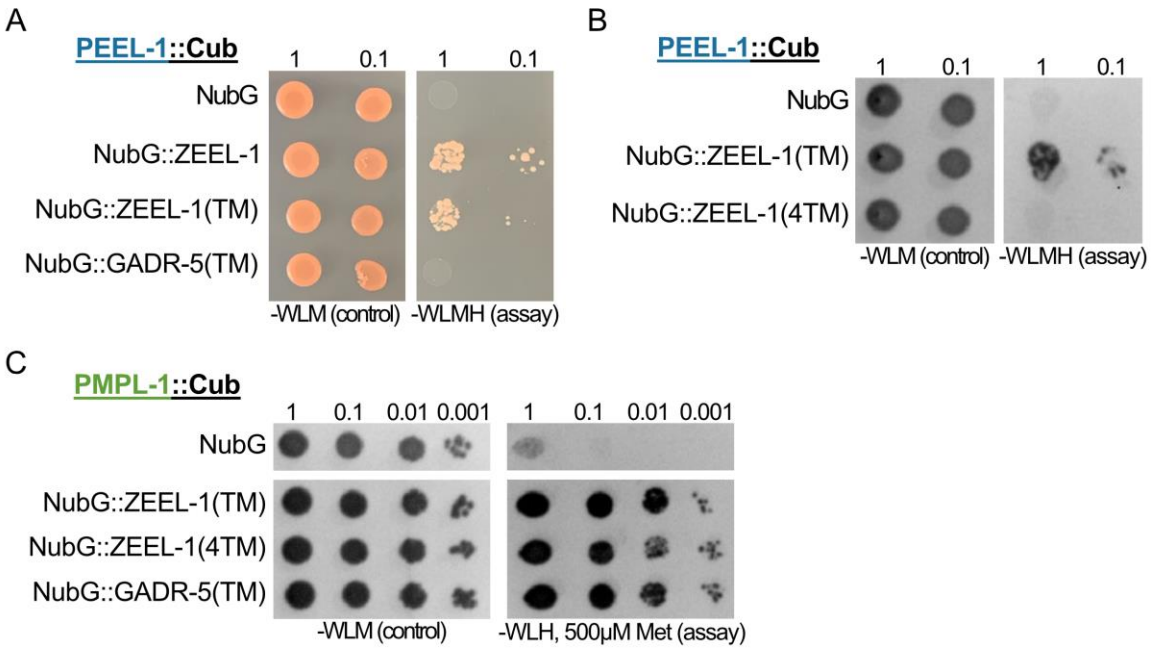


Figure 4.8. mbSUS assaying ZEEL-1 interactions with PEEL-1 and PMPL-1.

(A-B) mbSUS of yeast with PEEL-1::Cub with indicated Nub constructs. NubG is a negative control. Yeast were grown on -WLM media (control, left) or -WLMH media (assay plate, right) for 2 days at 30C. PEEL-1::Cub experiments were repeated at least three times and had similar results. **(A)** PEEL-1::Cub assayed with full-length ZEEL-1, or just the transmembrane domains (TM) of ZEEL-1 or GADR-5 (the closest ZEEL-1 paralog). **(B)** PEEL-1::Cub assayed with a ZEEL-1(4TM) which only contains 4 of the 6 transmembrane domains of ZEEL-1. **(C)** PMPL-1::Cub assayed with indicated constructs. 500µM methionine was used in assay plate to decrease PMPL-1::Cub expression. PMPL-1::Cub experiments were only performed once.

4.6 REFERENCES

Ben-David, Eyal, Alejandro Burga, and Leonid Kruglyak. 2017. "A Maternal-Effect Selfish Genetic Element in Caenorhabditis Elegans." Science (New York, N.Y.) 356 (6342): 1051–55. https://doi.org/10.1126/science.aan0621.

Cabanos, Cerrone, Miao Wang, Xianlin Han, and Scott B. Hansen. 2017. "A Soluble Fluorescent Binding Assay Reveals PIP2 Antagonism of TREK-1 Channels." Cell Reports 20 (6): 1287–94. https://doi.org/10.1016/j.celrep.2017.07.034.

Gati, Cornelius, Artem Stetsenko, Dirk J. Slotboom, Sjors H. W. Scheres, and Albert Guskov. 2017. "The Structural Basis of Proton Driven Zinc Transport by ZntB." Nature Communications 8 (1): 1313. https://doi.org/10.1038/s41467-017-01483-7.

Horaruang, Wijitra, and Ben Zhang. 2017. "Mating Based Split-Ubiquitin Assay for Detection of Protein Interactions." Bio-Protocol 7 (9): e2258. https://doi.org/10.21769/BioProtoc.2258.

Kane Dickson, Veronica, Leanne Pedi, and Stephen B. Long. 2014. "Structure and Insights into the Function of a Ca2+-Activated Cl− Channel." Nature 516 (7530): 213–18. https://doi.org/10.1038/nature13913.

Liapakis, G., M. M. Simpson, and J. A. Javitch. 2001. "The Substituted-Cysteine Accessibility Method (SCAM) to Elucidate Membrane Protein Structure." Current Protocols in Neuroscience Chapter 4 (May):Unit 4.15. https://doi.org/10.1002/0471142301.ns0415s08.

Liu, Bingxu, Rebecca J. Carlson, Ivan S. Pires, Matteo Gentili, Ellie Feng, Quentin Hellier, Marc A. Schwartz, Paul C. Blainey, Darrell J. Irvine, and Nir Hacohen. 2023. "Human STING Is a Proton Channel." Science 381 (6657): 508–14. https://doi.org/10.1126/science.adf8974.

Nimigean, Crina M. 2006. "A Radioactive Uptake Assay to Measure Ion Transport across Ion Channel-Containing Liposomes." Nature Protocols 1 (3): 1207–12. https://doi.org/10.1038/nprot.2006.166.

Sawyer, Jacob M, Stephanie Glass, Trudy Li, Gidi Shemer, Noor D White, Natalia G Starostina, Edward T Kipreos, Corbin D Jones, and Bob Goldstein. 2011. "Overcoming Redundancy: An RNAi Enhancer Screen for Morphogenesis Genes in Caenorhabditis Elegans." Genetics 188 (3): 549–64. https://doi.org/10.1534/genetics.111.129486.

Seidel, Hannah S., Michael Ailion, Jialing Li, Alexander van Oudenaarden, Matthew V. Rockman, and Leonid Kruglyak. 2011. "A Novel Sperm-Delivered Toxin Causes Late-Stage Embryo Lethality and Transmission Ratio Distortion in C. Elegans." PLoS Biology 9 (7): e1001115. https://doi.org/10.1371/journal.pbio.1001115.

Seidel, Hannah S., Matthew V. Rockman, and Leonid Kruglyak. 2008. "Widespread Genetic Incompatibility in C. Elegans Maintained by Balancing Selection." Science (New York, N.Y.) 319 (5863): 589–94. https://doi.org/10.1126/science.1151107.

Senning, Eric N, and Sharona E Gordon. 2015. “Activity and Ca2+ Regulate the Mobility of TRPV1 Channels in the Plasma Membrane of Sensory Neurons.” Edited by Richard Aldrich. eLife 4 (January):e03819. https://doi.org/10.7554/eLife.03819.

Su, Zhenwei, Emily C. Brown, Weiwei Wang, and Roderick MacKinnon. 2016. “Novel Cell-Free High-Throughput Screening Method for Pharmacological Tools Targeting K+ Channels.” Proceedings of the National Academy of Sciences 113 (20): 5748–53. https://doi.org/10.1073/pnas.1602815113.

Villegas, Raimundo, Gloria M. Villegas, Madalina Condrescu-Guidi, and Zadila Sárez-Mata. 1980. “Characterization of the Nerve Membrane Sodium Channel Incorporated into Soybean Liposomes: A Sodium Channel Active Particle.” Annals of the New York Academy of Sciences 358 (1): 183–203. https://doi.org/10.1111/j.1749-6632.1980.tb15396.x.

Widen, Sonya A., Israel Campo Bes, Alevtina Koreshova, Pinelopi Pliota, Daniel Krogull, and Alejandro Burga. 2023. “Virus-like Transposons Cross the Species Barrier and Drive the Evolution of Genetic Incompatibilities.” Science 380 (6652): eade0705. https://doi.org/10.1126/science.ade0705.