

MOLECULAR EVOLUTION OF A SUPERFAMILY OF CHEMORECEPTORS IN *MUSCA*  
*DOMESTICA* AND *CERATITIS CAPITATA*

BY

SARAH D. GIERS

THESIS

Submitted in partial fulfillment of the requirements  
for the degree of Master of Science in Entomology  
in the Graduate College of the  
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Advisers:

Professor Stewart H. Berlocher  
Professor Hugh M. Robertson

## ABSTRACT

The house fly (*Musca domestica*) and the Mediterranean fruit fly (*Ceratitis capitata*) are two important pest species. While both are generalists, the first is a major vector of a variety of human and livestock disease, while the second is a broadly polyphagous herbivore and important crop pest. Each must be able to navigate a complex set of chemical cues to gain access to resources. While it is relatively-well studied in *Drosophila* species, less is known about chemoreception in other dipterans. Generalist species of *Drosophila* retain higher diversity of chemoreceptor genes as compared to the rapid loss characteristic of specialists, but the same may not be true of other dipterans. Here the odorant receptor subfamily in *M. domestica* and the gustatory receptor subfamily in *C. capitata* were annotated, and phylogenetic analyses were performed.

Compared with *Drosophila melanogaster*, the *M. domestica* odor receptor repertoire is considerably larger with 84 models in the gene set, plus OrCo (59 in *D. melanogaster*), and contains expansion via complicated relationships of duplications in some lineages and losses in others. *C. capitata* shows a similar pattern of expansion within its gustatory receptor repertoire, with 81 protein models, which is intermediate between *D. melanogaster* (68 protein models) and *M. domestica* (103 protein models). The many species-specific expansions are the likely result of major changes in the chemosensory ability and ecology of each species. Similarity of ecologies between species appears to be a major driver for gene family evolution. Further exploration of these losses and expansions might ultimately lead to innovations in pest management and provide a base for which to further understand evolution of specialist versus generalist dipterans.

## **ACKNOWLEDGEMENTS**

I would like to thank my advisors, Stewart Berlocher and Hugh Robertson, for their help and guidance. I would like to thank Kim Walden for her advice on methods and minutia of genomics. I would like to thank my husband, Daniel Giers, for proofreading my thesis and being ever supportive throughout my research and writing. I would like to acknowledge my fellow graduate students for their advice, expertise, encouragement, and support. Finally, I would like to thank Kimberly Leigh for patiently answering my questions about various processes involved in depositing my thesis and defending.

## TABLE OF CONTENTS

INTRODUCTION .....	1
MATERIALS AND METHODS.....	7
RESULTS .....	11
DISCUSSION .....	23
TABLES .....	26
REFERENCES .....	32
APPENDIX.....	38

## INTRODUCTION

Among insects, chemoreception plays a strong role in locating and assessing mates, food, and oviposition sites. A large superfamily of seven-transmembrane receptors mediate response to a variety of volatile and soluble chemicals, to include odorants, tastants, and pheromones (Clyne et al., 1999; Hallem & Carlson, 2004; Leal 2013; Suh et al., 2014). This superfamily contains the distantly-related odor receptor (OR) and the gustatory receptor (GR) families (Robertson et al., 2003). In addition, the ionotropic receptor (IR) family, comprised of three-transmembrane domains, is also used in gustation and some aspects of olfaction (Benton, 2009). A fourth, odorant binding protein (OBP), family serves to bind and solubilize odorants then transport them through the sensillar lymph to the membrane-bound ORs (Leal, 2013). In adult flies, the antennae and maxillary palps are covered in sensilla (sensory hairs) and contain the olfactory receptor neuron (ORN) dendrites. These sensilla are well conserved in structure across insect species, consisting of a cuticular wall with multiple pores to allow odors through (Rospar, 1988). Each ORN expresses one to three ligand-binding ORs in with one being the most common for odor-specificity (Vosshall, 2000). In addition, *Orco*, which is necessary for olfaction (Vosshall, et al. 1999), is coexpressed with ORs in a heteromeric complex (Neuhaus et al. 2005). These complexes are thought to form ligand-gated cation channels (Leal, 2013). In *Drosophila*, a combinatorial code appears to contribute to olfactory sensitivity, as multiple ORs often detect the same compounds (Leal, 2013). Gustatory sensilla are distributed over the adult fly rather than being in only a few locations as are ORs (Stocker, 1994). As with ORs and ORNs, GRs are expressed in small subsets of gustatory receptor neurons (GRNs) (Smith, 2001), with some GRNs containing many coexpressed GRs (Thorne et al., 2004).

Conservation of chemoreceptor genes varies in flies and in other insects, with both *Orco* (well-conserved across most insects) (Hallem et al., 2006) and *Gr21* (well-conserved among Diptera and confers a CO<sub>2</sub> response in coexpression with *Gr63*) (Hill et al., 2002, Kwon et al. 2007) being examples of highly conserved genes. However, gene losses and duplications can be indicative of host changes, life history changes, and speciation (Wu and Ting, 2004). Rapid loss of chemoreceptor genes of specialists has been documented in *Drosophila* species (Wu & Ting, 2004; McBride & Arguello, 2007), whereas generalist species retain a higher diversity (Clark et al., 2007). This trend likely relates to the ecology of the species, where specialists may encounter fewer odorants, tastants, and toxic compounds than generalists.

The ecology of the model organism *Drosophila melanogaster*, as well as conspecifics, have been well studied and have genomes, including annotated chemoreceptors, available (Adams et al., 2000; Guo and Kim, 2007) for use in comparisons. Comparisons among species in the genus can provide some insight. However, few full chemoreceptor annotation studies have been completed outside of blood-feeders and *D. melanogaster* (NCBI, 2016). The genome annotations of the tsetse fly (*Glossina morsitans*) (Obiero et al., 2014) shows a continuation of the pattern of rapid gene loss in specialist feeders, but outside of *Drosophila*, little work has been done on patterns of chemoreceptor gene loss and duplication in generalist fly species. As genes which no longer serve a function (such as extraneous duplications) have no selective pressure against mutations and are likely to become pseudogenes (Li et al., 1981), and given the short generation time of flies, retained duplications are likely to be evolutionary important. To gain a more comprehensive understanding of patterns of dipteran gene loss and gain, further analysis at a deeper phylogenetic level is necessary. This study considers two generalist species within the suborder Brachycera (Wiegmann et al., 2011).

House flies, *Musca domestica*; (Diptera: Muscidae), are extreme generalists which are able to feed on a wide variety of decomposing matter and the bacteria it contains (Keiding, 1986; Watson & Martin, 1993). Larvae are unable to leave the substrate in which they develop. As decomposing matter, to include animal feces and carcasses, contains many toxins, pathogens and competitors (Zvereva, 1986a; Zvereva, 1986b) a variety of chemoreceptors likely play an important role in allowing the female to choose the correct oviposition sites. House flies are important vectors of human and animal diseases (Greenberg, 1973; Keiding, 1986, Scott & Lettig, 1962). Dangerous bacterial (including anthrax, typhoid fever, tuberculosis, cholera, and leprosy), fungal, and other pathogens are may be carried and transmitted by house flies. Due to this and their close association with humans and livestock, a comprehensive study of *M. domestica* chemoreception is of particular interest for their control.

The Mediterranean fruit fly, *Ceratitis capitata*; (Diptera: Tephritidae), hereafter referred to as the “medfly”, is a generalist plant feeder that originated in sub-Saharan Africa (Thomas et al., 2001). Its ability to shift host preferences based on availability has enabled the medfly to spread around the world (Malacrida et al., 1992). The medfly attacks over 260 documented fruits, flowers, vegetables, and nuts (Liquido et al., 1990) and is one of the most important pests of fruit crops worldwide. Combined annual costs of damage, control, and prevention in the US alone are in the millions of dollars (Szyniszewska & Tatem, 2014). Mating occurs primarily on the host plant (Kaspi & Yuval, 1999), and both males and females must distinguish between a wide range of chemical cues in order to locate appropriate host plants. As with *M. domestica*, fitness of medfly larvae is affected by competition (Dukas et al., 2001), and suitability of the potential oviposition location in regards to fruit ripeness and species must be determined by the female (Joachim-Bravo & Fernandes, 2001). Though many of the important pest species in

Tephritidae are specialists, chemoreceptor sequences of generalists such as the medfly can serve as models to search for and characterize other tephritid species chemoreceptor genes.

Annotations of the chemoreceptors in these two widespread and highly adaptive species may also offer insights into how flies adapt to new habitats and novel food sources. Such shifts can also result in sympatric speciation (Berlocher & Feder, 2002), allowing insight into processes of natural selection and evolution. Novel food sources such a non-indigenous hosts may serve as drivers for this selection, such as in the case of the *Rhagoletis pomonella* (apple maggot fly) complex. A relatively short (~150 years) shift from hawthorn (*Crataegus* spp.) to apple (*Malus pumila*) with associated loss of gene flow (a rate of 4-6% per generation) between host races underlines how quickly distinct populations can form (Dambroski et al., 2005).

As well as shedding light on speciation, understanding host shifts is especially critical to understand in a worldwide economy in which new and potentially invasive pests are easily transported from their country to origin across the world. Ecological traps—in which an organism is preferentially attracted to a low-quality habitat over a high-quality habitat, have long been studied (Dwernychuk & Boag, 1972), but the importance has only more recently been considered. While polarized light pollution has been considered one of the more well-documented sources of ecological traps for aquatic insects (Horváth et al., 2009), chemical cues cannot be overlooked. The mosquito *Culex pipiens* has shown preference for oviposition in aquatic habitats with non-indigenous blackberry (*Rubus allegheniensis*) detritus despite a low emergence rate in the same habitats versus those of more suitable habitats (Gardener et al., 2015). When presented with plant infusions (lacking in visual cues), the trend was retained (Gardner et al., 2014), suggesting a possibility for new methods of control.



Specific blends of attractive volatiles are already in use for selective trapping of the apple maggot fly (Zhang et al., 1999; Nojima et al., 2003). Conversely, non-host fruit volatiles can act as antagonists, repelling insects (Linn et al. 2005). Even the genotype of the host plant can affect host-finding and oviposition behaviors (Rull & Prokopy, 2004). Furthermore, duplications of chemoreceptors parallel those of resistance, such as the case of cytochrome P450s, which have been implicated in pesticide resistance a variety of insects (Wondji et al., 2009; Bass & Field, 2011). This underscores the usefulness of chemoreceptor annotations, as behavioral assays without ligand identification may not be as fine-tuned as would be ideal for such control methods as the production of better attractants and repellants and in developing insect-resistant crop varieties. The ability to stay one step ahead of resistance by implementing a variety of control methods, to include those based on chemoreception, may prove invaluable in integrative pest management systems. These methods are likely to become increasingly important in the light of pesticide resistance. Cross-resistance of one pesticide to another is documented in mosquitoes (Chandre et al., 1999) and beetles (Mota-Sanchez & Hollingworth, 2006) and is likely a common cause of resistance in insects, making development of new pesticides all the more difficult.

A better understanding of gene families thought to be involved may ultimately improve prevention and control methods for not only medflies and houseflies, but for many current pests, as well as invasive and potentially invasive insect species. With comparisons to the well-studied *D. melanogaster* genome, it is possible to make inferences about the evolution of chemoreceptor families in these two generalist species. While *Drosophila* is more closely related to *Musca*, it is expected that the more similar ecologies of *Drosophila* and *Ceratitis* will have resulted in similar expansions and deletions as compared to the filth fly, *Musca*, though they may be independent gene duplications due to the greater evolutionary distance between the species. Conversely,

genes involved in the detection of odorants and tastants associated with decay are likely to have been retained between *Drosophila* and *Musca*, as—though they feed on different substrates—both feed on decaying food sources.

## MATERIALS AND METHODS

### *Musca domestica* methods

Gene families were manually annotated and analyzed with the aid of corrected distance phylogenetic trees. BLAST searches were performed on the available Official Gene Set of proteins in REFSEQ at NCBI. All *D. melanogaster* and available *M. domestica* proteins were used as queries for TBLASTN searches. TextWrangler was used to manually assemble gene models. OBPs, gustatory receptors, and ionotropic receptors were not annotated by the author and the methods will not be discussed within but are instead available in the house fly genome paper (Scott et al., 2014). Gene nomenclature was done in accordance with Drosophila convention, with gene names italicized and proteins not.

Common problems involved absence of exons in gaps between contigs with scaffolds or off ends of scaffolds (denoted with suffices NTE [N-terminus missing sequence], CTE [C-terminus missing sequence], and INT [internal missing sequence] in the figures, tables, and proteins). Only a few of these gene models were able to be corrected using raw reads (suffix FIX in the figures, tables, and proteins) due to frequent large and complicated introns. Several of the gene models spanned scaffolds, and they had no support other than the agreement of the available exons in both scaffolds, and their appropriate relatedness to similar genes (suffix JOI in the figures, tables, and proteins). These problems are denoted in the Tables section. Genes on short scaffolds that were identical to ones on longer scaffolds were ignored as likely resulting from separate assembly of another haplotype. Extremely short fragments and highly degraded pseudogenes were also excluded. The *OrCo* gene, which is highly conserved across insects, had the last two exons duplicated 4kb downstream as well as a duplication of the first four exons at

the 5' end of another 231kb scaffold (originally modeled as XP\_005184813). These were ignored as likely assembly artifacts due to polymorphisms or, if real, as identical fragments.

Pseudogenes were translated as best possible in order to do phylogenetic analysis with intact proteins. Pseudogenes that were at least half the average length of related proteins were included (denoted with suffix PSE in the figures, tables and proteins). CLUSTALX v2.0 (Larkin et al., 2007) on default settings was used to align protein families with relevant *D. melanogaster* families, and these alignments were used to help further correct problematic gene models and of pseudogenes. Pseudogenes with in-frame deletions or insertions, crucial amino acid changes, promoter defects, or other unrecognizable changes were necessarily included as encoding functional proteins and may have artificially inflated the totals.

Poorly-aligned and variable length N-terminal and C-terminal regions were excluded for phylogenetic analysis, as was an internal region of the ORs which does not align with the OrCo proteins. Though potentially misleading, other regions of potentially problematic alignment were retained as they can provide important information for relationships within subfamilies. A combination of model-based corrections of distance between each pair of proteins and distance-based phylogenetic tree building was used for phylogenetic analysis. The BLOSUM62 amino acid exchange matrix within the maximum likelihood phylogenetic program, TREEPUZZLE v5.2 (Schmidt et al., 2002), was used to correct pairwise distances for multiple changes in the past. The resulting corrections were entered into PAUP\*v4.0b10 (Swofford, 2003) where a full heuristic distance search was conducted with tree-bisection-and-reconnection branch swapping to search for the shortest tree. Bootstrap analysis with 10,000 replications of neighbor-joining using uncorrected distances was performed to assess the confidence of branches, and they are shown above major branches in the figures. Adobe Illustrator® was used to manually color and

attach labels to lineages and subfamilies. Subfamilies were defined by shared evolutionary ancestry and function as described in *Drosophila*, where the degree of sequence similarity across all ORs is 26 percent or less (though still considered a family due to highly conserved sequences at certain locations throughout the protein) while the degree of subfamily similarity is upwards of 40 percent (Vosshall et al., 1999).

### ***Ceratitis capitata* methods**

The gustatory gene receptor family was manually annotated and analyzed using maximum likelihood phylogenetic trees. BLASTP searches were performed on both the JAMg Consensus Gene Set v1 (Poelchau et al., 2015) and high- and low-confidence protein sets from NCBI. *D. melanogaster* and *M. domestica* gene relatives were also used as queries for TBLASTN searches. Gene models were assessed for intactness and correctness with regards to length, introns (location, number, and boundaries), and conserved residues. Errors were repaired and named in the WebApollo (Lee et al., 2013) tool for manual gene annotations. Difficult gene models were manually assembled in TextWrangler prior to modification in WebApollo.

The genome assembly was generally straightforward with regards to this gene family. The most common corrections that needed to be made were truncation of exons in gaps between contigs within scaffolds or off of end of scaffolds (denoted with the suffix NTE in all figures, tables, and protein sequences). Raw reads were used to correct these gene models (suffix FIX, as in *M. domestica* results) where possible. Pseudogenes were translated as best possible for the purpose of aligning with intact proteins for phylogenetic analysis (suffix PSE). To qualify for analysis, a pseudogene had to be at least half the average length of related proteins. Several shorter fragments were not included. Protein alignment with *D. melanogaster* proteins and

refinements of gene models and pseudogenes were completed in the same manner as in *M. domestica*.

Alignments were trimmed for phylogenetic analysis using TRIMAI v1.4 (Capella-Gutiérrez et al., 2009), retaining only positions present in more than 80% of the sequences. Phylogenetic analysis was performed using maximum likelihood methods in PHYML v3.0 (Guindon et al., 2010) using default settings, and trees were prepared in FIGTREEv1.4 (Rambaut, 2014) and Adobe Illustrator®.

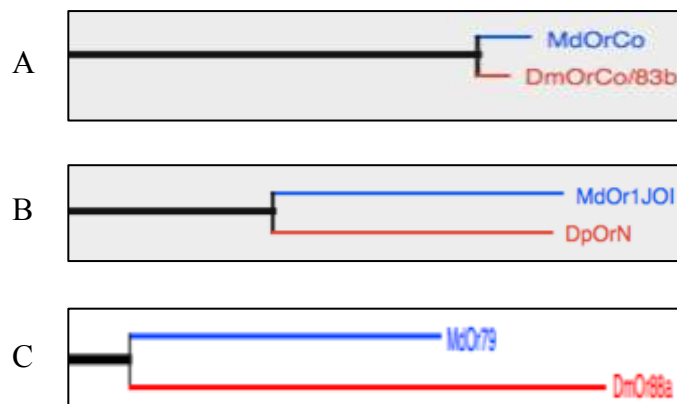
## RESULTS

### *Musca domestica* results

To avoid assumptions of orthology based on numbers, the numbering of MdOr genes starts with *MdOr1* as the ortholog of *DpORN*, a *D. pseudoobscura* gene which was lost from *D. melanogaster*, then roughly follows the *D. melanogaster* cytologically-named genes in order. The MdOr gene set consists of 84 models, plus the *OrCo* gene, compared to 59 models in *D. melanogaster*. While *D. melanogaster* has two sets of proteins encoded through alternative splicing of one gene (66aA/B and 69aA/B), the MdOr set only contains one such alternatively spliced gene (MdOr84A/B), and this model is questionable due to a large gap between them that might contain the C-terminal exons for *MdOr84A*. There are 7 apparent pseudogenes (8%), and 8 genes with missing parts could also be pseudogenes, with the result of 78 total apparently intact OR proteins. While approximately 12 gene fragments were so short and incomplete as to not be included, some may yet represent intact genes.

The automated gene modeling had access to all available insect ORs in GenBank for comparative information. The REFSEQ set used as the official gene set of the genome project was sufficient for building at least partial gene models for all but 2 of the current 78 intact genes. Unlike the case of many insect genome projects, more than half of these (44) models were precisely correct, presumably due to the relatively close relationship between *M. domestica* and *Drosophila*. The remaining 32 models required at least one change, and two new gene models were generated. Pseudogenes and models requiring joining across scaffolds were not included in this count.

As is expected, there was a single conserved ortholog (a) of the DmOr83b (now called Orco), sharing 87% amino acid identity (Vosshall, 2011). These were declared the out-group to root the tree (Figure 4). Fourteen simple 1:1 orthologous relationships appear in the set, such as the relationships of MdOr1JOI and DpOrN near the base of the tree (Figure 1b), sharing 44% amino acid identity, which allowed for confident building of the MdOr1 model across two scaffolds. Despite appearing to be simple orthologs, some cases such as in the pair of MdOr79 and DmOr88a (Figure 1c) (25% amino acid identity) are extremely divergent. While they are the best reciprocal BLAST matches and cluster together confidently in the tree and appear to be orthologous, they do not share microsyntenic neighbors in the two genomes and might instead have lost their true orthologs in each species and be inappropriate comparisons of paralogs.

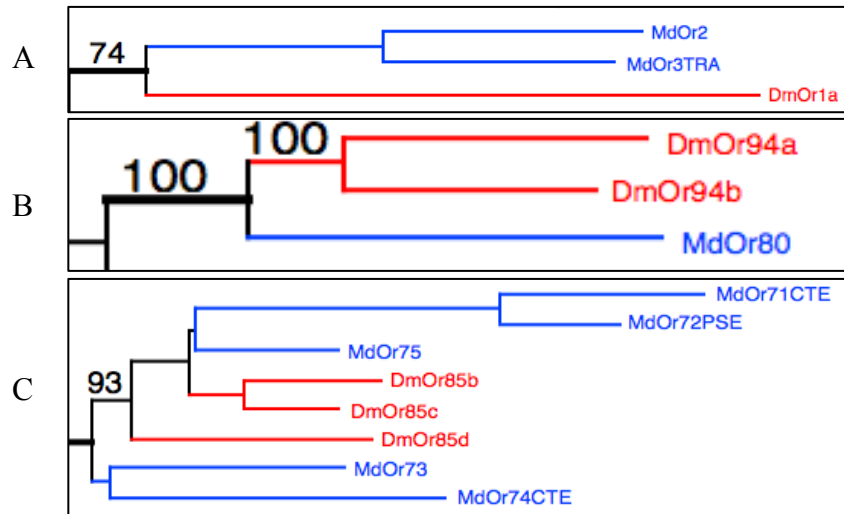


**Figure 1.** The 1:1 orthologous relationship between DmOrco/83b and MdOrCo (A) as well as DpOrN and MdOr1JOI (B). (C) Shows a more complicated relationship between the seemingly orthologous DmOr88a and MdOr79. The extreme sequence divergence may indicate a paralogous relationship. (Branch lengths across the three diagrams are not on the same scale. See Figure 4 for true lengths.)

A birth-and-death mode of evolution in this large family corresponds to its ecological relevance within a broadly generalist feeder. As well as the simple orthologs, there were a variety of more complicated relationships such as gene duplications in one or both species, large expansions in one species, and apparent gene losses. DmOr1a was duplicated as MdOr2/3 (Figure 2a), while the MdOr80 gene was duplicated as DmOr94a/b in *Drosophila* (Figure 2b).

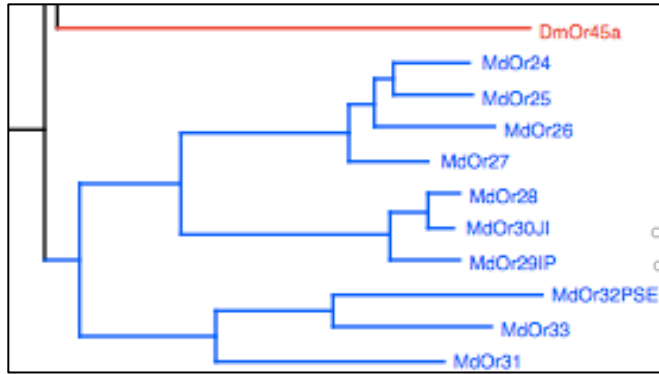


The set of DmOr85b-d and MdOr71-75 (Figure 2c) is considerably more complicated, and the orthology is less clear. This may be, in part, due to the missing C-termini in two of the *Musca* genes, as well as one being pseudogenized.



**Figure 2.** A duplication of DmOr1a in *M. domestica* (A) and the duplication of MdOr80 in *Drosophila* (B). (C) shows a complicated relationship between DmOr85b-d and MdOr71-75. (Branch lengths across the three diagrams are not on the same scale. See Figure 4 for true lengths.)

*D. melanogaster* contains nine genes related to MdOr22. There is an expansion of MdOr24-33 related to DmOr45a (Figure 3). DmOr67d has expanded to MdOr53-64 in *M. domestica*. Though some of the duplicated genes are in tandem arrays, these expansions are likely old events as each species appears to have undergone considerable gene movement since the separation of these two fly lineages. MdOr24-30 are in four different scaffolds but have the potential to be in a single tandem array save for their apparent move from the other three tandemly arrayed genes (MdOr31-33), which are in microsynteny with DmOr45a.



**Figure 3.** An expansion of a DmOr45a relative to 10 genes in in *M. domestica*. (Branch lengths across the three diagrams are not on the same scale. See Figure 4 for true lengths and full relationship.)

As there are several highly divergent genes and lineages in one species without any clear orthologous relative in the other it is indicative of the loss of several genes and lineages from each species (at least 8 from *M. domestica*, 12 from *Drosophila*). These losses, plus several expansions in *M. domestica* leads to the relatively larger size of the OR family in *M. domestica*.





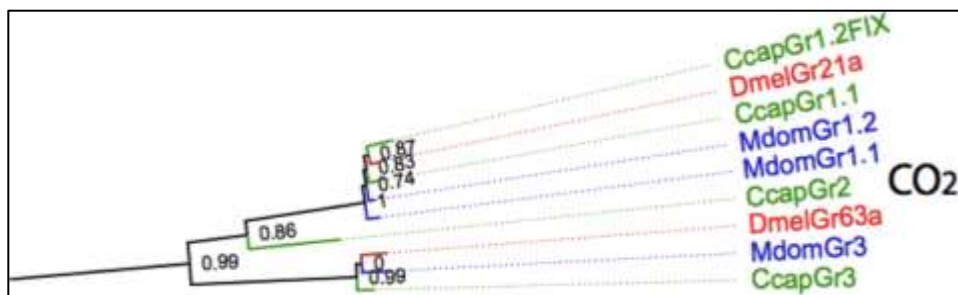
### ***Ceratitis capitata* results**

*C. capitata* has a GR family that is intermediate in size between these two species with 73 gene models encoding 81 potential proteins through alternative splicing (Table 2). In *D. melanogaster* the GR family consists of 60 genes encoding 68 proteins via alternative splicing of some genes (Robertson et al., 2003). *M. domestica* shows a considerable expansion to 79 genes encoding 103 proteins (with 13 pseudogenes) (41). *C. capitata* has two relatively small species-specific expansions, independent of two relatively small expansions in *D. melanogaster*, compared to three large species-specific subfamily expansions in *M. domestica* (Figure 9).

The CcapGr gene set contains a few fragments, but no long pseudogenes were found. As with both *Drosophila* and *Musca* GRs, some of the genes (five in *C. capitata*) appear to have an unusual form of alternative splicing in which multiple alternative long first exons are spliced into a shared set of C-terminal exons downstream of the last long first exon in these tandem arrays. The resultant proteins differ considerably in most of their sequence, and hence presumably bind different ligands. They are indicated with a lower case letter after the gene number (Table 2).

The carbon dioxide receptors are highly conserved within most of the holometabolous insects, except the Hymenoptera to date, with two proteins represented by DmGr21a/AgGr22 and DmGr63a/AgGr24 constituting the functional receptor (Robertson & Kent, 2009). *Drosophila* species have, however, lost a third member of this subfamily, first recognized as AgGr23, which is present in insects as divergent as *Tribolium*, *Bombyx*, mosquitoes, and *Glossina*, and known in those species as Gr2 (Robertson & Kent, 2009). This gene is an ancient paralog of the DmGr21 or Gr1 lineage. *Musca* does not have this Gr2 lineage, but *Ceratitis* does, so it was apparently lost between the divergence of the tephritid lineage from that leading to the *Musca/Drosophila* split. The importance of this protein is debated, with one study (Lu et al.,

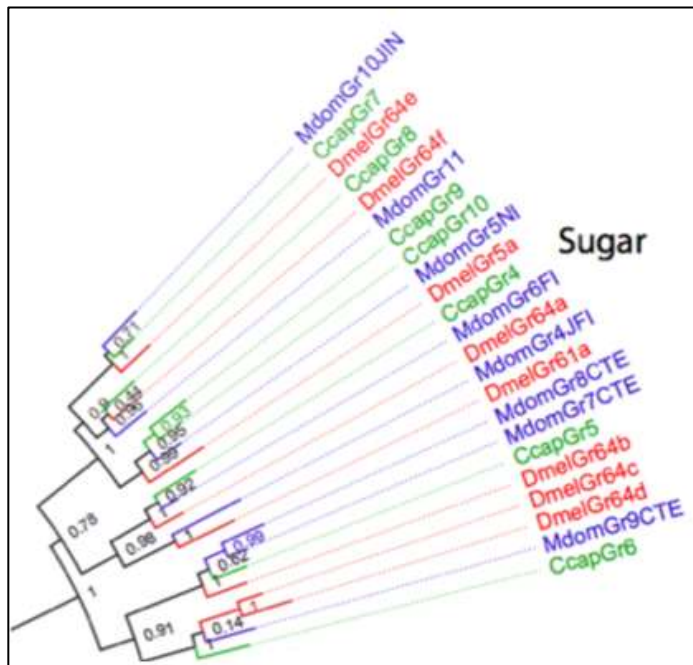
2007) finding that it enhanced perception of carbon dioxide, while another (Erdelyan et al., 2012) found that it did not. Like *Musca*, *Ceratitis* also has a recent duplication of the DmelGr21 or Gr1 lineage and, in an attempt to maintain the naming convention proposed by Robertson and Kent (Robertson & Kent, 2009), these are called CcapGr1.1 and 1.2, while the DmGr63a ortholog is called CcapGr3 (Table 2 and Figure 5). The tsetse fly contains the only other known recent duplication (Gr2 lineage) of a carbon dioxide receptor (Obiero et al., 2014).



**Figure 5.** The duplications of DmelGr21a in *C. capitata* and *M. domestica*.

The medfly has a slightly smaller set of seven sugar receptor subfamily genes (Figure 6) as compared to eight in *D. melanogaster* (Fujii et al., 2015). These genes have been found to represent four major lineages, which duplicated in basal Diptera (Kent & Robertson, 2009). AgGr16 represents one lineage, which was lost from *Drosophila* and *Musca*, and is not found in *Ceratitis*. The other three lineages are each represented by two or three paralogs in *Drosophila*, specifically DmGr61a and 64a, Gr64b/c/d, and 64e/f and 5a, all of which are proposed to have once been in a large tandem array, with the terminal Gr61a and Gr5a genes moving from that array (Kent & Robertson, 2009). These genes were poorly assembled in *Musca* but appeared to have a similar arrangement (Scott et al., 2014). *Ceratitis* has all seven genes in a single perfect tandem array, confirming this model (Table 2). There is a loss of the DmelGr61a/MdomGr4 ortholog in *Ceratitis*, though there is a duplication of its ortholog of DmelGr5a/MdomGr5

(Figure 9). The effect of these losses and duplications of these sugar receptors is unclear, as they are expressed in complicated combinations in subsets of sweet receptor gustatory neurons in *Drosophila* (Fujii et al., 2015), and even a single change could potentially have substantial effects.



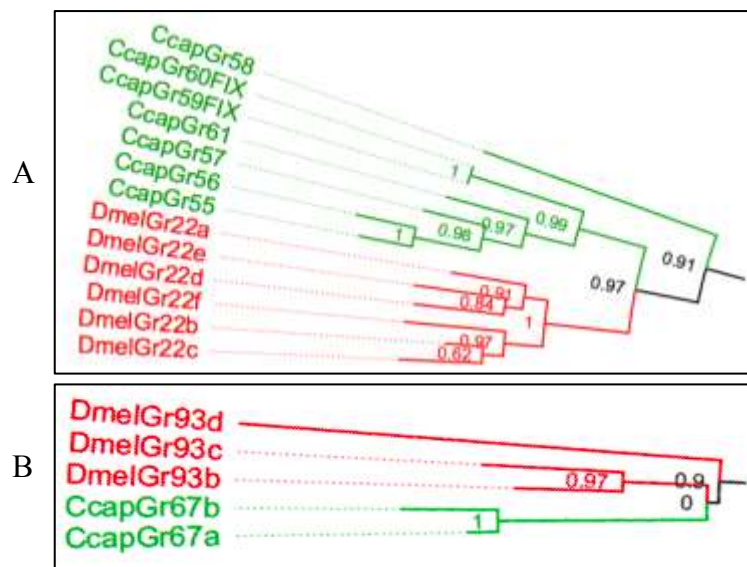




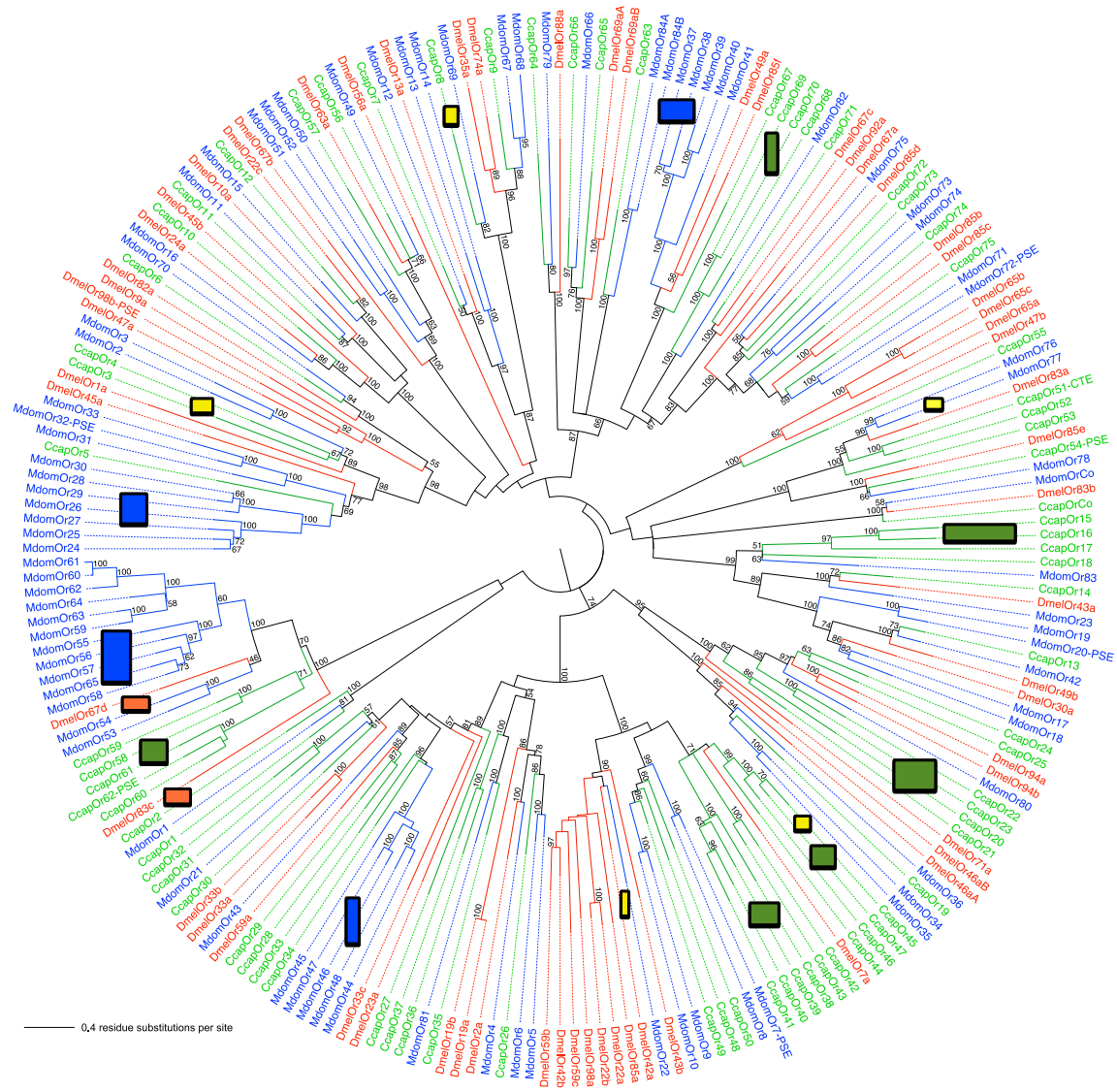
CcapGr43-52 are in a mostly tandem array (two genes are in inverted orientation) on a single scaffold (Table 2). Similarly, CcapGr21-34 are two arrays on the same scaffold (Table 2). The expansion of CcapGr55-61 are also in two arrays on the same scaffold as one another (Table 2).

In the tree (Figure 9), most of the bitter taste receptors cluster together on the bottom half and upper left quadrants, except for the Gr66a and 47b/57a lineages, which cluster with the Gr28a/b and Gr43a (fructose) receptors.

DmelGr22a-f was independently amplified to seven genes in a split array in a single scaffold in *Ceratitis* (Table 2, Figure 8a), and DmelGr93b-d is an alternatively-spliced locus in *Ceratitis* (CcapGr67a/b) (Table 2, Figure 8b). Both of these lineages were lost from *Musca*.



**Figure 8.** Independent amplification of DmelGr22a-f to seven genes in *C. capitata* (A) and the alternatively-spliced locus CcapGr67a/b as compared to Dmel93b-d (B). Neither lineage is found in *M. domestica*. (Branch lengths across the three diagrams are not on the same scale. See Figure 8 for true lengths.)



**Figure 9.** Phylogenetic tree of the *Ceratitis capitata*, *Drosophila melanogaster*, and *Musca domestica* GRs. This maximum likelihood tree was rooted by declaring the carbon dioxide and sugar receptor subfamilies as the outgroup. The *C. capitata*, *M. domestica* and *D. melanogaster* gene/protein names are highlighted in green, blue, and red, respectively, as are the branches leading to them to emphasize gene lineages. Clades discussed in the text are indicated on the outer edge. Suffixes after the gene/protein names are: NTE, amino terminus missing; CTE, carboxyl terminus missing; PSE, pseudogene; FIX, fixed assembly.

## DISCUSSION

While functional characterization of genes in both *Musca* and *Ceratitis* is necessary to be sure of their roles, it is possible to extrapolate potential roles based on the known ligands in *Drosophila*, especially in well-characterized subfamilies.

Several large species-specific expansions among the ORs are the likely result of major changes in the chemosensory ability and ecology of each species. As an example, DmOr45a mediates larval repulsion from aversive chemicals (Cobb and Dannel, 1994; Hallem et al., 2004; Kreher et al., 2005), and it is possible that MdOr24-30 proteins serve a similar function. The MdOr53-64 gene expansion may be involved in pheromone sensing in *M. domestica*, as the related DmOr67d receptor is known to sense the male-specific pheromone 11-cis-vaccenyl acetate (Kurtovic et al., 2007).

While the expansion of the DmGr43a/MdomGr12/13 fructose receptors in *Ceratitis* appears on the tree to only include one gene that is truly orthologous (Figure 4), separate phylogenetic analysis of this lineage in diverse insects supports these genes actually being a species-specific expansion. The hessian fly (*Mayetiola destructor*) (Anderson et al., 2014) and the silkworm (*Bombyx mori*) (Wanner & Robertson, 2008) have both duplicated this lineage, and *Tribolium castaneum* has 10 paralogs (Richard et al., 2008), suggests duplications in this lineage are not uncommon. This may be an example of subfunctionalization, though the specific role of these duplications is not clear.

The various splice forms of the DmelGr28b locus are expressed in gustatory cells as well as in the brain and elsewhere (Montell, 2013), and some have also been implicated in perception of light and of warmth (Montell, 2013; Ni et al., 2014). While there was not an expansion in *Ceratitis*, the expansion in *Musca*, as well as the unusual expression in *Drosophila*, highlights

this locus as one of potential interest for understanding ecological and behavioral differences in flies.

The duplication of the DmelGr66a protein, which is required for detection of caffeine and other bitter tastants (Lee et al., 2009), indicates a loss from the *Drosophila* lineage. The lost gene was presumably also involved in detection of bitter tastants in *Drosophila* and potentially is for the two other species, as well.

DmelGr32a is involved in courtship through expression in a small set of gustatory receptor neurons on the male foreleg and mediates rejection of non-conspecific females as targets of courtship (Miyamoto & Amrein, 2008). The related DmelGr68a gene is also a candidate pheromone receptor expressed on the foretarsi of males (Bray & Amrein, 2003), and the expansion of this lineage in *Ceratitis* (CcapGr16-19) indicates a more complex set of cuticular hydrocarbons and other pheromones involved in sex and species recognition in *Ceratitis* (Siciliano et al., 2014).

In *Drosophila*, DmelGr10a/b, DmelGr26a-c, DmelGr47, and Dmel59a-d are on four different chromosome arms. However in both *Musca* and *Ceratitis*, the relatives MdomGr42-64 (Scott et al., 2014) and CcapGr43-52 are in mostly tandem array on a single scaffold. This situation presumably reflects higher genomic flux in the *Drosophila* lineage after the separation from the *Musca* lineage. CcapGr21-34 are on two arrays on the same scaffold, but the related *Musca* expansion (Mdom16-19, 22-29, and 41) are on several scaffolds (Table 1). In *Drosophila*, the relatives (DmelGr22a-f and 85a) are on two different chromosomes.

The addition of *Ceratitis* GRs to the comparison of *Drosophila* and *Musca* corrected some of the several instances of possible gene loss in one or the other species amongst the candidate bitter taste receptors, showing them to be artifacts of the phylogenetic analysis of these

sometimes highly-divergent proteins. DmelGr39b and MdomGr19 now cluster in the tree. The addition of *Ceratitis* also lends support to and improved understanding of some of the other gene losses. Both *Musca* and *Ceratitis* appear to have lost DmelGr9a and 10b, which given the relationship of the three species to one another, indicates independent losses.

The loss of several DmelGr22a-f/CcapGr55-61 and DmelGr93b-d/CcapGr67a-b lineages in *Musca* may indicate they have a common role of fruit tasting, as *Musca* does not require this role. Similarly, *Musca* has lost the DmelOr7a lineage, a receptor for several fruit odors in *Drosophila* (Hallem & Carlson, 2006) which has been expanded to 10 genes in *Ceratitis* (Handler et al, 2016, in press in Genome Biology).

In summary, while maintaining a conserved set of receptors for carbon dioxide, sugars, most bitter tastants, and light and warmth, with an expansion of fructose receptors, *Ceratitis* is more similar to *Drosophila* than *Musca* in maintaining and expanding receptors that might mediate perception of fruits, while being differentiated from both in having expansions of receptors implicated in perception of courtship chemicals. This may be due to similar ecologies, despite *Ceratitis* being more basal phylogenetically than either *Musca* or *Drosophila* (Wiegmann et al., 2011).

As with *Drosophila* species, similar ecologies appear to be an important driver for gene family evolution. While *D. melanogaster*, *M. domestica*, and *C. capitata* are all generalists, *Musca* stands out in its role as a filth fly when compared to *Drosophila* and *Ceratitis*. Further exploration of losses and expansions via functional characterization and Ka/Ks ratios may offer important insight into the evolution of chemoreceptors across these species, as well as providing a base for which to further understand evolution of specialist versus generalist dipterans.

## TABLES

**Table 1.** Details of MdomOR family genes and proteins. Columns are: Gene – the gene and protein name we are assigning (suffixes are PSE – pseudogene; FIX – assembly was repaired; JOI – gene model spans scaffolds; NTE – N-terminus unidentified; CTE C-terminus unidentified; INT- internal regions unidentified; multiple suffixes are abbreviated to single letters); Ortholog – the *Drosophila melanogaster* ortholog, if relevant; OGS – the official gene number in the 17508 REFSEQ proteins (prefix is XP\_00); Scaffold – the genome assembly scaffold ID (amongst 20,487 scaffolds in assembly v2.0.2); Coordinates – the nucleotide range from the first position of the start codon to the last position of the stop codon in the scaffold; Strand – + is forward and - is reverse; Introns – number of introns in the coding region; AAs – number of encoded amino acids in the protein; Comments – comments on the OGS gene model, repairs to the genome assembly, and pseudogene status (numbers in parentheses are the number of obvious pseudogenizing mutations).

Gene	Ortholog	OGS	Scaffold	Coordinates	Strand	Introns	AAs	Comments
OrCo	OrCo/83b	5175278	1259	14618-39294	+	6	478	Fine as is
Or1JOI	DpOrN	-	1870	<1-704	-	5	410	Join across scaffolds
		5190411	3803	1138->12456	-			
Or2	Or1a	5191432	6947	296-4448	-	4	397	Adjust intron
Or3FIX	Or1a	5192136	926	7639->8205	+	4	394	Fix assembly
		5192135	926	<1-1303	+			
Or4	Or2a	5188267	2	1376434-1380567	-	2	403	Fine as is
Or5	-	5187291	19768	143389-145795	-	2	409	Fine as is
Or6	-	5187290	19768	138355-141817	-	2	392	Fine as is
Or7PSE	Or7a	5190908	469	31864-35096	-	2	301	Pseudogene (2)
Or8	Or7a	5190907	469	18390-27574	-	3	395	Fine as is
Or9	Or7a	5190906	469	1710-3196	-	3	435	Fine as is
Or10	Or7a	5189494	2052	6088-7581	+	3	437	Fine as is
Or11	Or10a	5178484	18678	341323-347596	+	4	404	Fine as is
Or12	Or13a	5191323	622	48571-53025	-	4	429	Fine as is
Or13	-	5189591	2206	16714-25745	-	5	458	Fine as is
Or14	-	5189590	2206	850-7860	-	5	461	Multiple changes
Or15JOI	Or22c	5187358	19788	164641->167811	+	5	398	Join across scaffolds
		5191818	840	<1-15841	+			
Or16	Or24a	5181476	18895	47341-53063	-	5	399	Fine as is
Or17	Or30a	5181900	18939	276808-280725	-	5	373	Fine as is
Or18	Or30a	5187208	19749	35029-43469	+	5	373	Fine as is
Or19	-	5190189	3153	9025-11972	+	4	390	Fine as is
Or20PSE	-		20052	61513-67014	-	4	378	Pseudogene (1)
Or21	Or33a-c	5185739	19448	85849-87055	+	1	372	Fine as is
Or22	multiple	5187966	19928	114212-116248	+	2	433	Fine as is
Or23	Or43a	5175768	172	5489-13041	-	6	375	Remove final exon

**Table 1.** Details of MdomOR family genes and proteins (cont'd).

Or24	Or45a	5184443	19219	250742-255417	+	4	409	Fine as is
Or25	Or45a	5191600	707	52120-53601	-	4	409	Fine as is
Or26	Or45a	5191599	707	45043-49510	-	4	410	Extend N-terminus
Or27	Or45a	5191598	707	28376-37151	-	4	410	Remove double exon
Or28	Or45a	5191597	707	21161-23884	-	4	401	Fine as is
Or29IP	Or45a	-	707	17156-18697	-	4	379	Pseudogene (1)
Or30JI	Or45a	-	6047	<1-630	-	4	392	Join across scaffolds
		-	11325	2175->2947	-			
Or31	Or45a	5176399	18582	264366-270628	+	4	412	Fine as is
Or32PSE	Or45a	5176400	18582	279858-283416	+	4	401	Pseudogene (1)
Or33	Or45a	-	18582	286433-293007	+	4	401	New gene model
Or34	Or46aA/B	5190201	3215	667-4827	+	2	390	Fine as is
Or35	Or46aA/B	5180017	18775	6136-10301	+	2	387	Change intron
Or36	Or46aA/B	5190949	4876	3163-7162	-	2	388	Extend N-terminus
Or37	Or49a/85f	5180070	18780	379848-399553	-	3	405	Remove an intron
Or38	Or49a/85f	5180069	18780	367541-370610	-	3	416	First half of model
Or39	Or49a/85f	5180069	18780	362526-365181	-	4	417	Second half of model
Or40	Or49a/85f	5180068	18780	355026-357699	-	3	405	Fine as is
Or41	Or49a/85f	5180067	18780	350440-347195	+	3	415	Extend N-terminus
Or42	Or49b	5179498	18733	464-7606	+	5	371	Fine as is
Or43	Or59a	5184787	19276	22286-28146	+	1	398	Fine as is
Or44	-	5182355	18985	218891-220146	+	1	381	Multiple changes
Or45	-	5182356 /7	18985	224017-225238	+	1	385	Join two models
Or46	-	5182358	18985	228629-231719	+	1	381	Extend N-terminus
Or47	-	5182359	18985	235409-238825	+	1	379	Fine as is
Or48	-	5186318	19571	23826-25032	+	1	381	Fine as is
Or49	Or63a	5178182	18661	230762-246942	-	9	415	Fine as is
Or50	-	5182424	1899	14153-21208	-	8	418	Fine as is
Or51JP	-	-	1899	24301->37688	+	9	374	Join across scaffolds
		5189964	282	<1-5485	+			Pseudogene (2)
Or52	-	5189964	282	9213-18804	+	9	415	Second part of model
Or53	Or67d	5191896	86	43081-47468	+	4	392	Fine as is
Or54	Or67d	5191897	86	59515-63895	+	4	391	Fine as is
Or55CTE	Or67d	5190382	371	<35334-39602	-	2	283	C-terminus missing
Or56	Or67d	5190383	371	45516-50769	+	3	389	Remove an exon

**Table 1.** Details of MdomOR family genes and proteins (cont'd).

Or57	Or67d	5190384	371	55524-57553	+	3	401	Fine as is
Or58	Or67d	5190385	371	60142-64096	+	3	392	Change N-terminus
Or59	Or67d	5184154	19194	200473-205423	+	3	401	Fine as is
Or60	Or67d	5187726	1986	5170-6520	-	3	393	Fine as is
Or61PSE	Or67d	5187725	1986	551-1914	-	3	393	Pseudogene (1)
Or62NTE	Or67d	5187292	19769	<155517-156592	+	3	315	N-terminus missing
Or63INT	Or67d	-	19329	84501-94471	+	3	340	Internal exon missing
Or64CTE	Or67d	5185047	19329	100221->101311	+	3	320	C-terminus missing
Or65	Or67d	5176290	18578	784217-787484	+	3	392	Add N-terminal exon
Or66	Or69aA/B	5180133	18785	70492-77903	+	4	414	Fine as is
Or67	Or35a/74a	5188535	20038	58824-61223	-	4	405	Fine as is
Or68CTE	Or35a/74a	5188534	20038	<52090-53182	-	3	325	C-terminus missing
Or69	Or35a/74a	5185292	19367	744621-746455	+	4	404	Fine as is
Or70	Or82a	5189776	254	14537-21410	-	5	382	Extend N-terminus
Or71CTE	Or85b-d	5186501	19612	<131090-132093	-	1	314	First part of model
Or72PSE	Or85b-d	5186501	19612	117647-123777	-	3	407	Pseudogene (1)
Or73	Or85b-d	-	19612	109012-113608	-	3	430	New gene model
Or74CTE	Or85b-d	5186500	19612	27601-33931	+	3	431	Last exon missing
Or75	Or85b-d	-	19158	246021-256220	-	3	422	New gene model
Or76	Or83a	5189143	20298	44324-57416	-	5	453	Fine as is
Or77	Or83a	5189142	20298	30730-38016	-	5	480	Fine as is
Or78	Or85e	5182622	19003	148795-151719	+	4	462	Fine as is
Or79	Or88a	5179127	18710	611653-616793	+	3	409	Fine as is
Or80	Or94a/b	5185000	19322	151735-155579	+	3	394	Fine as is
Or81	-	5180861	18837	332652-347139	-	2	393	Fine as is
Or82	-	5181716	18920	328080-337530	-	4	399	Fine as is
Or83	-	5186825	19662	78318-83939	+	6	381	Fine as is
Or84A	-	-	18719	42819-57451	+	3	354	Alternatively spliced?
Or84B	-	5179301	18719	54021-57451	+	3	354	Fine as is



**Table 2.** Details of CcapGR family genes and proteins. Columns are: Gene – the gene and protein name we are assigning (suffixes are PSE – pseudogene; FIX – assembly was repaired); Ortholog – the *Drosophila melanogaster* ortholog, if relevant; Scaffold – the genome assembly scaffold ID; Coordinates – the nucleotide range from the first position of the start codon to the last position of the stop codon in the scaffold; Strand – + is forward and - is reverse; Introns – number of introns in the coding region; AAs – number of encoded amino acids in the protein.

Gene	Scaffold	Coordinates	Strand	Introns	AAs
Gr1.1	NW_004522754.1	2538981-2541774	-	3	453
Gr1.2FIX	NW_004524179.1	616410-619268	-	3	457
Gr2	NW_004522955.1	907124-961389	+	6	418
Gr3	NW_004523769.1	2394633-2398158	+	6	489
Gr4	NW_004524024.1	276530-286395	+	7	422
Gr5	NW_004524024.1	289459-300345	+	8	416
Gr6	NW_004524024.1	308241-319444	+	8	420
Gr7	NW_004524024.1	322859-334335	+	8	433
Gr8	NW_004524024.1	345036-355741	+	8	506
Gr9	NW_004524024.1	359168-371927	+	7	464
Gr10	NW_004524024.1	376169-383933	+	6	414
Gr11	NW_004523199.1	807240-816067	-	11	421
Gr12	NW_004523199.1	824877-833605	+	11	395
Gr13	NW_004523199.1	837796-847745	-	11	402
Gr14	NW_004523199.1	851100-864679	-	11	417
Gr15	NW_004523845.1	205555-207051	-	3	437
Gr16	NW_004524079.1	1250919-1252448	+	3	443
Gr17a	NW_004524079.1	1254576-1258977	+	3	379
Gr17b	NW_004524079.1	1256140-1258977	+	3	429
Gr18	NW_004524079.1	1261932-1263539	+	3	446
Gr19	NW_004524079.1	1266078-1270355	+	3	443
Gr20	NW_004523642.1	688336-694829	+	3	409
Gr21	NW_004524356.1	2552687-2555918	+	2	424
Gr22	NW_004524356.1	2560080-2561917	+	2	375
Gr23	NW_004524356.1	2562016-2565166	-	2	438
Gr24	NW_004524356.1	2566717-2570944	-	4	417
Gr25	NW_004524356.1	2576218-2577563	-	2	403
Gr26	NW_004524356.1	2579304-2580819	+	2	396
Gr27	NW_004524356.1	2582966-2584281	-	2	399
Gr28	NW_004524356.1	2586065-2587412	+	2	404
Gr29	NW_004524356.1	2589835-2591633	+	2	403
Gr30	NW_004524356.1	2593164-2594467	-	2	397

**Table 2.** Details of CcapGR family genes and proteins (cont'd).

Gr31	NW_004524356.1	2597365-2598731	-	2	411
Gr32	NW_004524356.1	2601901-2603262	+	2	403
Gr33	NW_004524356.1	4729555-4732392	+	5	396
Gr34a	NW_004524356.1	4719310-4726698	-	4	390
Gr34b	NW_004524356.1	4719310-4723149	-	4	390
Gr35	NW_004524112.1	2125168-2127267	+	3	405
Gr36	NW_004524112.1	2129067-2132367	+	3	405
Gr37	NW_004524112.1	2134777-2136158	+	3	379
Gr38	NW_004523802.1	5867855-5871692	+	5	525
Gr39	NW_004523913.1	6812795-6816553	+	5	496
Gr40	NW_004523676.1	890106-897989	-	4	483
Gr41a	NW_004524023.1	9289156-9311740	-	3	441
Gr41b	NW_004524023.1	9289156-9306998	-	3	467
Gr41c	NW_004524023.1	9289156-9298844	-	3	472
Gr41d	NW_004524023.1	9289156-9294457	-	3	439
Gr41e	NW_004524023.1	9289156-9292399	-	3	448
Gr42	NW_004524023.1	9286477-9288048	-	3	451
Gr43	NW_004524212.1	2960679-2961953	+	1	402
Gr44	NW_004524212.1	2957329-2958615	-	0	428
Gr45	NW_004524212.1	2955262-2956508	-	1	394
Gr46	NW_004524212.1	2953066-2954355	-	1	392
Gr47	NW_004524212.1	2949693-2951060	-	1	425
Gr48	NW_004524212.1	2948143-2949535	+	1	436
Gr49	NW_004524212.1	2933559-2934826	-	1	402
Gr50	NW_004524212.1	2882872-2884177	-	1	408
Gr51	NW_004524212.1	2880687-2881968	-	1	404
Gr52	NW_004524212.1	2878439-2879951	-	1	408
Gr53	NW_004523199.1	4152001-4153685	-	1	469
Gr54	NW_004523199.1	4167641-4166385	-	1	400
Gr55	NW_004523680.1	206055-209471	-	1	394
Gr56	NW_004523680.1	203007-204368	-	1	430
Gr57	NW_004523680.1	199138-200373	-	1	388
Gr58	NW_004523680.1	748826-753133	+	1	434
Gr59FIX	NW_004523112.1	540-1307	+	2	256
	NW_004523680.1	771367-771936	+	NA	165
Gr60FIX	NW_004523680.1	776656-777901	+	2	347

**Table 2.** Details of CcapGR family genes and proteins (cont'd).

Gr61	NW_004523680.1	786169-787409	+	1	389
Gr62	NW_004523814.1	1342584-1345881	-	2	437
Gr63	NW_004523814.1	1350456-1352809	-	2	433
Gr64	NW_004523814.1	1356884-1360185	+	3	420
Gr65a	NW_004522866.1	5997294-6010015	+	1	486
Gr65b	NW_004522866.1	6003391-6010015	+	1	476
Gr66	NW_004523714.1	2082558-2083767	+	1	383
Gr67a	NW_004523968.1	1996565-2005973	+	1	410
Gr67b	NW_004523968.1	2001212-2005973	+	1	407
Gr68	NW_004523968.1	2043588-2046271	-	1	408
Gr69	NW_004524379.1	2209083-2210393	-	2	397
Gr70	NW_004522889.1	271092-281773	-	3	454
Gr71	NW_004523692.1	354117-355701	-	2	412
Gr72	NW_004523836.1	3169564-3171000	-	2	409

## REFERENCES

- Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., . . . Venter, J.C. (2000). The genome sequence of *Drosophila melanogaster*. *Science*, 287(5461), 2185-2195. doi:10.1126/science.287.5461.2185
- Andersson, M.N., Videvall, E., Walden, K. K., Harris, M. O., Robertson, H. M., & Löfstedt, C. (2014). Sex-and tissue-specific profiles of chemosensory gene expression in a herbivorous gall-inducing fly (Diptera: Cecidomyiidae). *BMC Genomics*, 15(50). doi:10.1186/1471-2164-15-50
- Bass, C., & Field, L.M. (2011). Gene amplification and insecticide resistance. *Pest Management Science*, 67(8), 886-90. doi:10.1002/ps.2189
- Berlocher, S., & Feder, J. (2002). Sympatric speciation in phytophagous insects: moving beyond controversy?. *Annual Review of Entomology*, 47(1), 773–815. doi:10.1146/annurev.ento.47.091201.145312
- Benton, R., Vannice, K.S., Gomez-Diaz, C., & Vosshall, L.B. (2009). Variant ionotropic glutamate receptors as chemosensory receptors in *Drosophila*. *Cell*, 136(1), 149-62. doi: 10.1016/j.cell.2008.12.001
- Bray, S., & Amrein, H. (2003). A putative *Drosophila* pheromone receptor expressed in male-specific taste neurons is required for efficient courtship. *Neuron*, 39(6), 1019-1029. doi:10.1016/S0896-6273(03)00542-7
- Cahill, G., Gardner, A., & Parker, A. (2014). Effects of blackberry leaf infusion on survival rate and outcomes of intraspecific competition of *Culex restuans* and *Aedes aegypti*. Retrieved from <https://www.ideals.illinois.edu/handle/2142/66085>
- Capella-Gutiérrez, S., & Silla-Martínez, J.M. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25 (15), 1972-1973. doi:10.1093/bioinformatics/btp348
- Chandre, F., Darrier, F., Manga, L., & Akogbeto, M. (1999). Status of pyrethroid resistance in *Anopheles gambiae* sensu lato. *Bulletin of the World Health Organization*, 8(11). doi: 10.1673/031.008.1101
- Clark, A.G., Eisen, M.B., Smith, D.R., Bergman, C.M., Oliver, B., Markow, T.A., . . . MacCallum, I. (2007). Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*, 450(7167), 203-18. doi:10.1038/nature06341
- Clyne, P.J., Warr, C.G., Freeman, M.R., Lessing, D., & Kim, J. (1999). A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila*. *Neuron*, 22(2), 327–338. doi:10.1016/S0896-6273(00)81093-4
- Dambroski, H.R., Linn, C., Berlocher, S.H., & Forbes, A.A. (2005). The genetic basis for fruit odor discrimination in *Rhagoletis* flies and its significance for sympatric host shifts. *Evolution*, 59(9), 1953-1964. doi:10.1111/j.0014-3820.2005.tb01065.x
- Dukas, R., Prokopy, R.J., & Duan, J.J. (2001). Effects of larval competition on survival and growth in Mediterranean fruit flies. *Ecological Entomology*, 26(6), 587-593. doi:10.1046/j.1365-2311.2001.00359.x
- Dwernychuk, L.W., & Boag, D.A. (1972). Ducks nesting in association with gulls—an ecological trap? *Canadian Journal of Zoology*, 50(5), 559-563. doi:10.1139/z72-076
- Erdelyan, C., Mahood, T.H., & Bader, T. (2012). Functional validation of the carbon dioxide receptor genes in *Aedes aegypti* mosquitoes using RNA interference. *Insect Molecular Biology*, 21(1), 119-127. doi:10.1111/j.1365-2583.2011.01120.x

- Fujii, S., Yavuz, A., Slone, J., Jagge, C., Song, X., & Amrein, H. (2015). *Drosophila* sugar receptors in sweet taste perception, olfaction, and internal nutrient sensing. *Current Biology*, 5(25), 621–627. doi: 10.1016/j.cub.2014.12.058
- Gardner, A.M., Allan, B.F., Frisbie, L.A., & Muturi, E.J. (2015). Asymmetric effects of native and exotic invasive shrubs on ecology of the West Nile Virus vector *Culex pipiens* (Diptera: Culicidae). *Parasites & Vectors*, 8(329). doi: 10.1186/s13071-015-0941-z
- Greenberg, B. (1973). Flies and disease. Retrieved May 26, 2016, from <http://jid.oxfordjournals.org/content/123/3/local/advertising.pdf>
- Guindon, S., Dufayard, J.F., & Lefort, V. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Society of Systematic Biologists*, 59 (3), 307-321. doi:10.1093/sysbio/syq010
- Guo, S., & Kim, J. (2007). Molecular evolution of *Drosophila* odorant receptor genes. *Molecular Biology and Evolution*, 24(5), 1198–1207. doi:10.1093/molbev/msm038
- Handler, A.M., Papanicolaou, A., Schetelig, M.F., Arensburger, P., Atkinson, P.W., Benoit, J.B., . . . Richards, S. (In review). The whole genome sequence of the Mediterranean fruit fly, *Ceratitidis capitata* (Wiedemann): a highly invasive and destructive pests of fruits and vegetables throughout the world. *Genome Biology*.
- Hallem, E.A., & Carlson, J.R. (2004). The odor coding system of *Drosophila*. *TRENDS in Genetics*, 20(9), 453-459. doi:10.1016/j.tig.2004.06.015
- Hallem, E.A., & Carlson, J.R. (2006). Coding of odors by a receptor repertoire. *Cell*, 125(1), 143-60. doi: 10.1016/j.cell.2006.01.050
- Hallem, E.A., Dahanukar, A., & Carlson, J. R. (2006). Insect odor and taste receptors. *Annual Review of Entomology*, 51, 113doi:10.1146/annurev.ento.51.051705.113646
- Hill, C.A., Fox, A.N., Pitts, R.J., Kent, L.B., & Tan, P.L. (2002). G protein-coupled receptors in *Anopheles gambiae*. *Science*, 298(5591), 176–78. doi:10.1126/science.1076196
- Horváth, G., Kriska, G., & Malik, P. (2009). Polarized light pollution: a new kind of ecological photopollution. *Frontiers in Ecology and the Environment*, 7(6), 317-325. doi: 10.1890/080129
- Joachim-Bravo, I.S., & Fernandes, O.A. (2001). Oviposition behavior of *Ceratitidis capitata* Wiedemann (Diptera: Tephritidae): association between oviposition preference and larval performance in individual females. *Neotropical Entomology*, 30(4), 559-564. doi:10.1590/S1519-566X2001000400008
- Kaspi, R., & Yuval, B. (1999). Lek site selection by male Mediterranean fruit flies. *Journal of Insect Behavior*, 12(2), 267-276. doi:10.1023/A:1020975018826
- Keiding, J. (1986). The Housefly—biology and Control. Training and Information Guide (advanced Level) (Switzerland, World Health Organization, Division of Control of Tropical Diseases). Retrieved May 26, 2016, from [http://www.who.int/water\\_sanitation\\_health/resources/vector302to323.pdf](http://www.who.int/water_sanitation_health/resources/vector302to323.pdf)
- Kent, L.B., & Robertson, H.M. (2009). Evolution of the sugar receptors in insects. *BMC Evolutionary Biology*, 9(41). doi: 10.1186/1471-2148-9-41
- Kurtovic, A., Widmer, A., & Dickson, B.J. (2007). A single class of olfactory neurons mediates behavioural responses to a *Drosophila* sex pheromone. *Nature*, 449(26), 542-546. doi:10.1038/nature05672

- Kwon, J.Y., Dahanukar, A., & Weiss, L.A.. (2007). The molecular basis of CO<sub>2</sub> reception in *Drosophila*. *Proceedings of the National Academy of Sciences*, 104(9), 3574-3578. doi:10.1073/pnas.0700079104
- Larkin, M.A., Blackshields, G., Brown, N.P., & Chenna, R. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, 23 (21), 2947-2948. doi:10.1093/bioinformatics/btm404
- Leal, W.S. (2013). Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes. *Annual Review of Entomology*, 58, 373-391. doi:10.1146/annurev-ento-120811-153635
- Lee, E., Helt, G.A., Reese, J.T., & Munoz-Torres, M.C. (2013). Web Apollo: a web-based genomic annotation editing platform. *Genome Biology*, 14(R93). doi:10.1186/gb-2013-14-8-r93
- Lee, Y., Moon, S.J., & Montell, C. (2009). Multiple gustatory receptors required for the caffeine response in *Drosophila*. *Proceedings of the National Academy of Sciences*, 106(11), 4495–4500. doi:10.1073/pnas.0811744106
- Li, W.H., Gojobori, T., & Nei, M. (1981). Pseudogenes as a paradigm of neutral evolution. *Nature*, 292(5820), 237-239. doi:10.1038/292237a0
- Linn, C., Nojima, S., & Roelofs, W. (2005). Antagonist effects of non- host fruit volatiles on discrimination of host fruit by *Rhagoletis* flies infesting apple (*Malus pumila*), hawthorn (*Crataegus* spp.), and flowering dogwood (*Cornus florida*) *Entomologia experimentalis et Applicata*, 114(2), 97-105. doi:10.1111/j.1570-7458.2005.00222.x
- Liquido, N., Cunningham, R.T., & Nakagawa, S. (1990). Host Plants of Mediterranean Fruit Fly (Diptera: Tephritidae) on the Island of Hawaii (1949-1985 Survey). *Journal of Economic Entomology*, 83(5), 1863–1878. doi:10.1093/jee/83.5.1863
- Lu, T., Qiu, Y.T., Wang, G., Kwon, J.Y., Rutzler, M., & Kwon, H.W. (2007). Odor coding in the maxillary palp of the malaria vector mosquito *Anopheles gambiae*. *Current Biology*, 17(18), 1533-1544. doi:10.1016/j.cub.2007.07.062
- Malacrida, A.R., Guglielmino, C.R., Gasperi, G., & Baruffi, L. (1992). Spatial and temporal differentiation in colonizing populations of *Ceratitis capitata*. *Heredity*, 69, 101–111. doi:10.1038/hdy.1992.102
- McBride, C.S., & Arguello, J.R. (2007). Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics*, 177(3), 1395-1416. doi:10.1534/genetics.107.078683
- Miyamoto, T., & Amrein, H. (2008). Suppression of male courtship by a *Drosophila* pheromone receptor. *Nature Neuroscience*, 11(8), 874-876. doi:10.1038/nn.2161
- Miyamoto, T., Slone, J., Song, X., & Amrein, H. (2012). A fructose receptor functions as a nutrient sensor in the *Drosophila* brain. *Cell*, 151(5), 1113–1125. doi:10.1016/j.cell.2012.10.024
- Montell, C. (2009). A taste of the *Drosophila* gustatory receptors. *Current Opinion in Neurobiology*, 19(4), 345-353. doi:10.1016/j.conb.2009.07.001
- Montell, C. (2013). Gustatory receptors: not just for good taste. *Current Biology*, 23(20), R929-R932. doi:10.1016/j.cub.2013.09.026
- Mota-Sanchez, D., & Hollingworth, R.M. (2006). Resistance and cross- resistance to neonicotinoid insecticides and spinosad in the Colorado potato beetle, *Leptinotarsa decemlineata* (Say) (Coleoptera: Chrysomelidae). *Pest Management Science*, 62(1), 30-37. doi:10.1002/ps.1120

- NCBI. (2016). Genome information by organism. Retrieved from <http://www.ncbi.nlm.nih.gov/genome/browse/>
- Neuhaus, E.M., Gisselmann, G., Zhang, W., & Dooley, R. (2005). Odorant receptor heterodimerization in the olfactory system of *Drosophila melanogaster*. *Nature Neuroscience* 8, 15-17. doi:10.1038/nn1371
- Ni, P., Bronk, P., Chang, E.C., Lowell, A.M., Flam, J.O., Panzano, V.C., ... Garrity, P.A. (2014). A gustatory receptor paralog controls rapid warmth avoidance in *Drosophila*. *PubMed Central*, 500(7464), 580-584. doi:10.1038/nature12390
- Nojima, S., Jr, L.C., Morris, B., & Zhang, A. (2003). Identification of host fruit volatiles from hawthorn (*Crataegus* spp.) attractive to hawthorn-origin *Rhagoletis pomonella* flies. *Journal of Chemical Ecology*, 29(2), 321-336. doi: 10.1023/A:1022677827233
- Obiero, G., Mireji, P.O., & Nyanjom, S. (2014). Odorant and gustatory receptors in the tsetse fly *Glossina morsitans morsitans*. *PLoS Neglected Tropical Diseases*, 8(4). doi:10.1371/journal.pntd.0002663
- Poelchau, M., Childers, C., & Moore, G. (2015). The i5k Workspace@ NAL—enabling genomic data access, visualization and curation of arthropod genomes. *Nucleic Acids Research*, 43(D1), D714-D719. doi:10.1093/nar/gku983
- Rambaut, A. (2014) Figtree, a graphical viewer of phylogenetic trees. Retrieved from <http://tree.bio.ed.ac.uk/software/figtree>
- Tribolium Genome Sequencing Consortium. (2008). The genome of the model beetle and pest *Tribolium castaneum*. *Nature*, 452(7190), 949-955. doi:10.1038/nature06784
- Robertson, H.M., & Kent, L.B. (2009). Evolution of the gene lineage encoding the carbon dioxide receptor in insects. *Journal of Insect Science*, 9 (19). doi:10.1673/031.009.1901
- Robertson, H.M., & Warr, C.G. (2003). Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences*, 100(Suppl 2), 14537-14542. doi:10.1073/pnas.2335847100
- Rospars, J. (1988). Structure and development of the insect antennodeutocerebral system. *International Journal of Insect Morphology and Embryology*, 17(3), 243-294. doi:10.1016/0020-7322(88)90041-4
- Rull, J., Prokopy, R. (2004). Host-Finding and Ovipositional-Boring Responses of Apple Maggot (Diptera: Tephritidae) to Different Apple Genotypes. *Environmental Entomology*, 33(6), 1695–1702. doi:10.1603/0046-225X-33.6.1695
- Shim, J., Lee, Y., Jeong, Y.T., Kim, Y., Lee, M.G., Moncell, C., Moon, S.J. (2015). The full repertoire of *Drosophila* gustatory receptors for detecting an aversive compound. *Nature Communications*, 6, 8867. doi:10.1038/ncomms9867
- Schmidt, H.A., Strimmer, K., & Vingron, M. (2002). TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics*, 18(3), 502-504. doi:10.1093/bioinformatics/18.3.502
- Siciliano, P., He, X.L., Woodcock, C., Pickett, J.A., Field, L.M., Birkett, M.A., . . . Zhou, J.J. (2014) Identification of pheromone components and their binding affinity to the odorant binding protein CcapOBP83a-2 of the Mediterranean fruit fly, *Ceratitidis capitata*. *Insect Biochemistry and Molecular Biology*, 48, 51-62. doi: 10.1016/j.ibmb.2014.02.005
- Scott, H.G., & Lettig K.S. (1962). Flies of public importance and their control (USA, Communicable Disease Center). Retrieved May 26, 2016, from <https://stacks.cdc.gov/view/cdc/7611>

- Scott, J.G., Warren, W.C., Beukeboom, L.W., Bopp, D., Clark, A.G., Giers, S.D., . . . Liu, N. (2014). Genome of the house fly, *Musca domestica* L., a global vector of diseases with adaptations to a septic environment. *Genome Biology*, 15(10). doi:10.1186/s13059-014-0466-3
- Swofford, D.L. (2003). PAUP\*. Phylogenetic Analysis Using Parsimony (\*and Other Methods). Version 4. Retrieved from [http://people.sc.fsu.edu/~dswofford/paup\\_test/](http://people.sc.fsu.edu/~dswofford/paup_test/)
- Szyniszewska, A.M., & Tatem, A.J. (2014). Global assessment of seasonal potential distribution of Mediterranean fruit fly, *Ceratitis capitata* (Diptera: Tephritidae). *PLoS ONE*, 9(11): e111582. doi:10.1371/journal.pone.0111582
- Thomas, M.C., Heppner, J.B., Woodruff, R.E., & Weems, H.V. (2001). Mediterranean fruit fly, *Ceratitis capitata* (Wiedemann)(Insecta: Diptera: Tephritidae). Retrieved May 26, 2016, from <http://www.growables.org/information/documents/FruitFlyMediterranean.pdf>
- Thorne, N., Chromey, C., Bray, S., & Amrein, H. (2004). Taste perception and coding in *Drosophila*. *Current Biology*, 14(12), 1065-1079. doi:10.1016/j.cub.2004.05.019.
- Vieira, F.G., & Rozas, J. (2011). Comparative genomics of the odorant-binding and chemosensory protein gene families across the Arthropoda: origin and evolutionary history of the chemosensory system. *Genome Biology and Evolution*, 3, 476–490. doi:10.1093/gbe/evr033
- Vosshall, L.B., Amrein, H., Morozov, P.S., & Rzhetsky, A. (1999). A spatial map of olfactory receptor expression in the *Drosophila* antenna. *Cell*, 96 (5), 725–736. doi:10.1016/S0092-8674(00)80582-6
- Vosshall, L.B., & Hansson, B.S. (2011). A unified nomenclature system for the insect olfactory coreceptor. *Chemical Senses*, 36(6), 497-498. doi:10.1093/chemse/bjr022
- Vosshall, L. B., Wong, A. M., & Axel, R. (2000). An olfactory sensory map in the fly brain. *Cell*, 102(2), 147–159. doi: 10.1016/S0092-8674(00)00021-0
- Wanner, K.W., & Robertson, H.M. (2008). The gustatory receptor family in the silkworm moth *Bombyx mori* is characterized by a large expansion of a single lineage of putative bitter receptors. *Insect Molecular Biology*, 17(6), 621-629. doi:10.1111/j.1365-2583.2008.00836.x
- Watson, D.W., & Martin, P. (1993). Egg yolk and bacteria growth medium for *Musca domestica* (Diptera: Muscidae). *Journal of Medical Entomology*, 30(4), 820-823. doi:10.1093/jmedent/30.4.820
- Wiegmann, B.M., Trautwein, M.D., Winkler, I.S., Barr, N.B., Kim, J., Lambkin, C., . . . Yeates, D.K. (2011). Episodic radiations in the fly tree of life. *Proceedings of the National Academy of Sciences*, 108(14), 5690–5695. doi:10.1073/pnas.1012675108
- Wondji, C.S., Irving, H., Morgan, J., & Lobo, N.F. (2009). Two duplicated P450 genes are associated with pyrethroid resistance in *Anopheles funestus*, a major malaria vector. *Genome Research*, 19(3), 352-459. doi:10.1101/gr.087916.108
- Wu, C.I., & Ting, C.T. (2004). Genes and speciation. *Nature Reviews Genetics*, 5, 114-122. doi:10.1038/nrg1269
- Zhang, A., Jr, L.C., Wright, S., & Prokopy, R. (1999). Identification of a new blend of apple volatiles attractive to the apple maggot, *Rhagoletis pomonella*. *Journal of Chemical Ecology*, 25(6), 1221-1232. doi: 10.1023/A:1020910305873
- Zvereva, E.L. (1886a). The effect of ecological factors on competitive relations between house fly larvae (*Musca domestica* L., Diptera, Muscidae) and microscopic fungi. *Entomologicheskoe Obozrenie*, 86.



Zvereva, E.L. (1886b). Peculiarities of competitive interaction between larvae of the house fly *Musca domestica* and microscopic fungi. *Zoologicheskii Zhurnal*, 65.

## APPENDIX

### 87 MdOr proteins in FASTA format

>MdOrCo

MQANLQPTKYTGLVADLMPNIKLMKYSGLFMHAFTGGSALLKNVYSSIHVLIVLQFI  
FILVNMALENADEVNELSGNTITALFFTHCITKFVYIAVNQKNFYRTLNIWNQPNHPLF  
AESDARYHSIALAKMRKLFFLVMLTTVASAVAWITITFFGESVKFATDKETNSTITVPIPI  
RLPIKSFYFPWDASSGMFYMISFGYQAYYLLFSMVHSNLCDFVFCSWLIFACEQLQHLK  
GIMKPLMELSASLDTYRPNSAALFRSLANSKSELIQNEEKEPVNDLDMSGIYSTKADW  
GAQFRAPSTLQTFNGINGGNPNGLTKKQEMMVRSIAKYWVERHKKHVRLVAAIGDTY  
GAALLHMLTSTIKLTLLAYQATKITGVNVYAFTVIGYLGYALAQVFHFCIFGNRLIEE  
SSSVMEAAYSCHWYDGSEEAKTFVQIVCQQCQKAMSISGAKFFTVSLLDFASVLGAV  
VTYFMVLVQLK

>MdOr1JOI

MKDKFKTFMRDFFPSNVEKGEIGSVKLNWLQAQITGVPIIGLKDESSLIKNLILLYGIFTT  
TVVTFIYTGFEYDLYMNWHDLSLTQNTCLSLTHVSGAIKTVNIIFHLPRLEGVIRKL  
KHVTKTYIKSEKQLVVFYDGEVENKLVLSIYIGIVGFTGFMGMIMLYMPEAVAGKIFPY  
RVILPDWMPQQQLQLLYMGLSVIIFAIQIIAVDYLNVTIINQIRFQLNILNLAFFDDLIVETQ  
ANSRETKSLVLYKDDPVKRMDSIVEHHCLLGELRQETEDIFSQPILWQFMTSVIIFAMT  
GFQATVRSSGSSAAVLIYAYCGCIFCELFVYCWFGEVSEQSKTLGTSGFHSSWYHFD  
RRYGKSLILFTNAQRPFVFTAGGFMGLSLPSFTGILSKSYSYIALLRQIYGK

>MdOr2

MYNNVDGKTRQDLEFLDVQYRALIRVGLDIGAIRGKDFLNDRGKFLIYGIITTYLQYGL  
ILFAVHIFGVQIDKASAALSMFNQGSLLMLKVSILIFKSNRLLKLIWDMNLLATMANEP  
ERETWLSERNFSKVIGNIYSTACIASVILSISIPFMSYEHFKGLEVSLKLPFDGEFPYHL  
GIPIFILNYILSVIYVYTLLCWTIGIDTLFGWLIHAVSGHFRILRLKVEMAAKKIDEHGNH  
LDFVQDIGAIVRYHIKTLGFVDALNEIFGQIFWAEVAFSCLQMCFLIFTLNNGSDKRMIP  
FNAMVFTAISIQMMIYCFGGGEKIKSENEMFCFDIYSKFPWEKMYPSSEKRMMLPLQRS  
QQDAALRGLFFELDRNLLVYIYRTAFSYNTLLGAMKE

>MdOr3FIX

MSDTERQNLDYLPVQFGAFMVLGLDIGVTRRSALLKSGWTFLFNILCTVFMEYGFANF  
VINSITDIDAITSSLSMFNQGMLLTFKVLVMVFKGDEMLKLIWDMNRLARGANAKAW  
EIWISERNMGKWIALGYYYCCYIAATIMAVMPWLFMLYEVVQGRGVHLRLPFQLQFF  
FVSGNGFHISIFYIYTLLVVRWFNMSVGIDTLFGWYIFAVSGHFRILRHKIKETALKI  
DAYDNHRDFVSDVAAFVSYHNRTLKFTENLNRLYGEILWSEISMSCLQLCFLLYSLTN  
DENFANIPFHFFASAAITMQLMIYCFGGGEKLNENDMLCHDIYMAMPWEKMYPSSEK  
LMLLPLLRTQREISLKGLYFVINVNLLVFIFKTAFSFITLLGAMKEI

>MdOr4

MTNALTDNNKNIYSKLDTNVAFYHVKVWRWTGIKPPQDMNPQLYRLYAIVLNFLA  
TVLFPLSLIANVFFTQNLQQLCENLTITISDCQSNLKFINVFLVRHQLDRIKSILRLDRR  
VQDDKEFAVLKSAIATARSSFLIFFRLYSFGTTLSSVVKVALAESRSLFPWFVGNWDG

NLSTYVVVIVYQFFGLAVQALQNVANDSYPPAYLVILSAHMRALAIRVKAVGQFRQEG  
MQQPLTLSAEEQAKCLKEFNECIKDYLNLKLSHIIQRIISKACLAQFACSALVQCTVGL  
HFMYVVDAANYEAQLMSIIFVAVTLEAFVICYFGHMMSLQSSNLTYAFYSCGWLAQ  
SPEFKRNLITLMRTQRTSTIRAGSYIPVDLPTFVVLMMKYAYSVFTLLIRFK

>MdOr5

MALQPMASSSSSASNKIHTWQAFRNHWILWKFCGLHPPKRNSRWFPYLIYAIVLNVT  
TTLMPITLIVDLILSQNLTELCENLYVTITDVICSLKFINIFTVRHKLLEVRWILERLDVR  
ATTPEQRQELRHGIQTSHKWFMAFFRFYTCAVITSQLVVYLSKERVLMYPSWFPWDW  
KASKRNLFAHCYQVYTVSVQTVQNLGSDTYPQAYIVVLIAHIRALGLRIKALGEALS  
ATAAGDVSSPSSSSKKLSDDELYRELVNCVKDHQIVHELYLTIQECISKTCLAQFVATG  
LAQCTIGVYIIVYVGSDFSRLNSFMFFGAITIEILICYFGDLYCRANDFLIDAIYDCNWID  
KDERFKKALLLLQRSQQADCLKAGNLIPVRLPTFVKIMKTAYSFTVLNEVN

>MdOr6

MSVLFSPHPNTWEAFKYHWLLWKWCGLQPPSRDSKWFRPYLAYAIIFNLTILFPLSL  
VLDLTLSQNLTEIFQNLVTVTVVFSSSLKFVNVFLIRKLLLEVRFLERLDVRANTEEQ  
QQELKNGIAMAHKCFMIFLRLYVCAITTSQLVVYFSSSERVLMYPSWLPWDWRESKRY  
FLFAICFQIYAVSAQLSQNLGNDTYPQAYIVILIAHIRALALRIKHLGVVSTSVPAPEGKL  
SQEDFYRELRCVKDHEHVHELYLTIQECLSTTCLAQFIATGLAQCIIGVYILYVGDDFS  
RLNSLVFFGAVTIEILVLCYFGDLYCQANEFLIDAIYATNWMDRDGRFKKALLVLQ  
RAQVTNCLKAGNLTPVMLPTFVTIMKTAYSFTVLNKNV

>MdOr7PSE

AVKISKKVATKQALTNLICFRVVGIVHTKSNPHLYIVYAIVIHSLTTVFTPISTTSYFR  
KTDQDFNVGVFLTSIQAVINVYGCIAKILLIYYKTKLEAAEKLMDKMDQHCRAEDEI  
QELFNIRDLGRKIVLGYITAYWYTTMTYISALVSGVPSYSINLFFLDWKRSKREFYLAS  
FLEYVLVTWTCLQQVANDSYGTIYVCILRGHVRVLLLRIRKMGRKVDQADQNLEEL  
KSCIKDHKDLELYNIISPVISRTIFLQFSITAVILGITLIXIAKFSFSLYTLIKQMGIKERLG  
L

>MdOr8

MSKQTVKVIKKVATKQALTYLYGCFRVMGIHFTKSHTHLYLIYVIVIHSLTTVFTPIST  
TSYFRKTDENFNMGVFLTSVQAVINVYGCIVKIFFLVYYKKKLEAAEKLMDQMDQHC  
QADDEIQEIYNIRNLGRRRIIGYGIAYWIYTTMTYISALASGVPSYSLNLFIDWRRSKLE  
FYVASFIEYFLTSWTCFQQVANDSYGTIYVCILRGHVRILLRIRKMGRKVNRTADQNL  
EELKTCIKDHKELIELYNAISPVISRTIFLQFSITAAILGITLVNIAIFASSITAMAASAFYIV  
AVSVEIFPLCYANCLLYDSDTLATEIFHSAWIGQDRRYRKMLIFFIQRTQKSMELWAG  
KMFAINLNTFISIAKFSFSLYTLIKQMGIKERLGL

>MdOr9

MEIPNITTVLPQQVQEDEQEPSTSSNKTLSHANKSDTNDSSVQTRHGLRFLFIGFRL  
LGVYFPKRGRFLYFLWSLNVLYATIYLPTGLVVGIIHTRDVAIGDMLTSLQVAIDVVG  
CAIKIVLMYFLLPQLLQCDPVLRLDKRCTSPEEKDLVRRFISHGNRFVILFGMAYWSY  
ASSTCISAVLFHRLPYNLNPLLDATASKGSFVLGVFVEMMPMYLACSQQVDDSYA  
VIYTQILRTHLKALVFRLQHLNDDHRNENGVISPEAEERNIENLKLCCIIDHKNIIELYTRV

APVISITLHVQFTITASLLGVTLINILIFATNTASIVASCFYVLAVVVEIFPLCYAQCMLN  
ENDHLTEAIFHSNWIHQSKRYRQMLIFFMQRSQKSIEFTAGKLPITLSSFLSIKFSFSL  
YTLIKEMDIKTHYGLD

>MdOr10

MEHPDIGEQPALLPQQIQEEQPQPETKSNEIPKLNHENKWDLKAEPPLETRQGLRYLYN  
GFRFLGIYFPKRRKGLYLLWSIIVNLYVTIFLPTGFIMGIISVTDENVEIGNLLTSFQVAIN  
VVGCSIKIILMYFLLPQLLKCEPIFERLDGRCTSREEKDLIRQFVHDGNRLVVLFTVAYW  
SYSSSTCISAVLFGRLPYNIYNPFIDANASRGYFILAVFMMVPMIDIACFQQVVDSDYA  
VIYTQILRTHLQALLIRLQHLNDDDAADLDDEAQERNVEKLKLCIHDHKSIIELYNRVAP  
VISITIFVQFTITASLLGSTLINILIFATNTASIVASCFYVLAVVVEVFPLCYAQCMLDEN  
NRLTEAIFHSNWIYQNKRYRQMLIFFMQRSQKVIEFTAGKLPITLSSFLSIKFSFSLYT  
LIKEMDLKERYGLN

>MdOr11

MFLRFLSRSNPLKEYFYVPRICLQLMGFWPGSPRSRRILCWAVFNFIILLVGVVTELH  
AGFSYLNVDLEKGLDTLCPAGTSAVTVLKMILISYYRQDLEAVLKKMHQMLYGCNEK  
DMEHKAVYNRIIRQSSVMAARVNFAPFLAGFITCTAYNLKPLILVWIFWSKGGKDLMWL  
TPFNMTMPKFLLEGPLYPLAYIFTAYTGYVTIFTFGGSDALYFEYCTHIATLLKMLQTD  
VKLLFRKFEGKLTLPTEAAYVEEQILILIKRHNVIEMTDFFRKRYSIITLAHFVSASMVI  
GASIFEMLTYTGFGRFIYLGYTVAALSQAVCYGGTLVAENSIYLATVVFKCNWYIC  
DPKLRRILMIICRSQKSLNMSVPFFSPSMSTFASILQTSGSIIALASSFQ

>MdOr12

MFNPKPNNDLNYRIPGQCIWLKLNGSWPYNHQEANKDFYSSRYVWGWLYTVWSWY  
VWWSVGITIGFQTAFLINNLDIMMTTENCCTTFMGALNFVRLMHMRLNQRQFKVVIQ  
QFVEDIWINKKQHPHVAAVCSRNMRTFRIMTVLLSCLISMYCVLPLVVLFFDVGLDAD  
EKPFPYKMLFPFDAHHGWRYIVTYIFTSYAGMCVTTLFAEDSIFGFFVTYTCGKFQIL  
HERIDNLVFDAYESVANRQNELEIQECYVKLLNRIAYDHNKLIEFAGKLENFFNPILLV  
NFTISSILICMVGFLVTGKDMFIGDYVKFIVYISSSLSQLYVLCWNGDSLQHSLETAN  
HLYTCNWEGGQIRSYMPASKKFRQNLEIMIMCSQRPVKITALKFSTLSLQSFTAILSTSM  
SYFTLLKTVYDENQEDGPAN

>MdOr13

MGFWYKPNCPFDEKFSFVSDFYVHLIINGCWPTDGDGPKSLSYRICNALYTLWSCQVIFS  
LNFTLYAECMYVYENSADLGKVVGNMCLIMIALMVSLRLLYFRGDISRMKRLTTFMA  
EKIWDSEAHPKAYERAVRRTKPTFYISLSLWICLVLYLLFPIIFNLQGKSPDSNDKPLP  
FPTVFPYDTQTHWAYIFTYIFLSYAGYIAVSLFYAMDAILAYFISFVAGQFEILHADIARL  
IPECHAEWLRRYGAGAAENGVKLNYLQEMYAKRLHGIKRHKDIIAFCKELEKFMSFP  
LFANYGTSTFLICFVGFMIAGLKSFGDFMRFFMFFMAVTGQLFIVCKLGNLLITQST  
DTAHYLFACNWECCGYLSKNSPLLLYPDIMELQELNRNLPLWKDLSYIPANRNFKLKL  
MFMIMRSNRPVQLSVMQFTVLSLQSFNKVVSNSLSYFALLKSFLDK

>MdOr14

MAILYKPRCGEDVNFVLPLKVRTFLMINGCWPMEDNANNTNGLWNRLKSLYQLWFI  
FGVVCLFYIVCVGWVYIVANFSDVKKVVEAISTSTIGINVLIRMIYLRCSRFSKFHVLEK

FTNKIWINKVTHPLIFKRCIKRTVPTFYLSITLWMVLFYICALPIFVLITTDQTIHSNDKTF  
PYPMIFPYDPQKPINYILTYMTSIYTGAITVTLFYATDAILAIFISFLCGQFEILHGNIARLI  
PECHAEFLANYRGESTGSKKNDIFLHNLYVKRLHELATAHDELIRFSMDLEKLFSFQL  
MVNVVTSTFQICTNLFQFIVAGRNSLSDFLRFFLFFFSVTGQLYVMCELGTILTRSTDT  
ANYLFSCNWEGGILSQHSPLLQVDYITLDSLNTKLPAWRTLEYYPPTNRDFRMKCLKLM  
IMRSQRPVHLTAMKFTVSSLESFTRILSTSMSYFTLLNSFLD

>MdOr15JOI

MFDFLKASMPIAKSFMLVPRACGRLCGVWPDPEYRWRNTLFIESTVVTFLFGGVGELS  
YGFTHLNDLVDALDAFCPAVTKIISFFKATIIFINRKKFYDIMQRLRTLIMREQHDSKKM  
KMOVQGFSSFGNICTFIIVSGGSSTNVFYNLRAIITNIIYHFQEEERKLEFPFKSLVPEFTTR  
FPYFPGMFLILTASGVMTVFSFSIVDGYYVCTTVFICSIFKIIQQDIGSIFDELKDCEHATD  
EQNHRIRQKLNIAIVERHNTIIDLSADFTASFTVIIMLHFMSAAIVVCSSLLDMLNTNSV  
GLFIYISYNIAAFAQLFVYCVGGTFVSDSSAAVADVLYNVIEWYKCDIKTRKIIILMILHRS  
QKATTISVPFFTPSLPAFSSIISTAGSYIALLKTF

>MdOr16

MVPNFLKNSYPLNKQYLLIPRFALRILGFYPESEWNVWLKSWAFFNISILAYGCIYAEY  
YGIYYLPIDIVMSLDALCPVASSIMSFIKIFFIWYREYQYKQLIEEVRRLTEDQNTLRKE  
KMKRWYFTIATRLTALVLFGLCCSTSYSIRAILTNTLLYLNGKDIVYETPFKMMFPEPL  
LAMPYIPITFLLVHWHGYITVLSFVAGDGLFLGFCFYFSTLLKALQQDLTEVLGVIDET  
KKYRKLTESEKVMSSLSKIIRRHNEIADLTMKLSSIMVEITLCHFITSVIGTSVIDLLLFA  
GGYGSIVYIVYTCAVLSEIFLYCLGGTAVIESSQELAVKAYTSNWyGQSVRIQKMVLLII  
VRSQRHFVVKVPFFTPSLPALTAILRFTGVSIALVKSMI

>MdOr17

MQIRSIEDVPLLSTNLSIMKFWSFLLRHDWRRYFALIPYLFINTTQFLDVYFSTEPIDAIV  
RNAYIAVLFFNTILRAVLLCVNRFEYEGFMEKIRLLYIELMNSDPALRKMLQECTVAS  
RFISKVNLLMGFTSCVGFNMYPFATSKVLPFGMYVPGVDKYESPYQICFLFQIIITPA  
GCCMYIPFTNLIVSFILFGILMCKVLQHKLRNLKDVSEKARTVIVWCIKYQLQLINFVD  
TINDLTFTFLFEFMAFGAMLCAMLFLLIIVETVAQMCIICIYIFMIFAQSVIMYYFANEL  
YDQSLKVAIAAYESNWFDVDFVSTQKTIKLFILRAQKPCAILVGKVYPMNLEMLQSLN  
ATYSYFTLLKRVYG

>MdOr18

MLIETIEDVPLYNNSLRIMKFWSFLLRHDWRRYLSLIPYIILTSSQFVDLFFSTEPMDAIR  
NAYLAVLFFNTTLRGIAVCIHQSRYEDFLERIRVLYIDMMESDQWVREELQAITLAAN  
NISRVNLVMGTCSVISFLIYPIFATTKVLPFGIYVPGVDKNISPYEICFIVQTVMAPIGCC  
MFIPFTNMIVAIMLFAILMCRMRQRKLRHLCHVTSEEARATIIWCIKYQTELIRYVNTIN  
DLITYTNLLEFLAFGAMLCAMMFTLVTVETVSQMCLICVYILMIFAQSTILYYYANKVF  
DESLNVGTAAYESEWFDVDVDTQRTLRLILRAQKPCAILVGRVYPMNLELLQSLNLT  
TYTYFTLLRNVYD

>MdOr19

MKFLTERKTNKITKYSAKIKRLEDVPMLWFNVRILKFWSVLIDNNWRQYFSYIPFFFLN  
IFQILDLYYTEKEINDKIHDTYMTMIIFNTFLRAIVMVTNRRKFSESLEYMKDLYAELIM

EYDFEIRQIIRKYSMDVLKVSKINLTMGILTGLGFSMFPIMAEEREFFIFGMYPYLYNEYQ  
TPWYEILLAVQSVLNLSGMCTFIPFAGMFVSFLVFAMAISKVLQYKLSKLSTEISSKLAE  
RQIIECIKLHLKLISFIDKVNELCSIISLVDCILFVVILCIMLLSFILVKTVIQKCVIVVYMIM  
VFTQTFLLYYFSNETYHESLEISTAAYNIDWFNYDVETQKVLQQLLLRSQKPCAILIAKA  
YPINLVRLQAMLRVTYSVFTLLDKFYG

>MdOr20PSE

MRISMRQKDQLISKYSNKKIKNIEDVPMLWFNVRILKFWSVLIDDNWRQYFTYIPFFVLN  
IFQLMDLCLTQKELNDKIHDIYMTMLMFNTFLRTVVMVTNRKKFCKLLEHIRQMYEEL  
MMERDAEICRIMEDHTAMVLKISKINLIMGMLTTXEFIFGIYVPYLEEYQSPWYEILLTG  
QSFLNLSAMCIFIATAMFISYFMFAIAISKVLQYKLSRTCTEVSSKIVEEKIVECIKLHLR  
LISFIEQINELCGFIALMDFLLFVIVLCIMLLSFVLVKTVTQKCIITVYISMVFTQAFLLYY  
FSNEVFYESLQISTAAYDINWFNYDVRTQEVKLKLLLLRSQKPCAILIVKSYPVNLQRLQ  
VLVKITYSVFTLLEKVYG

>MdOr21

MAFENFYQTNSVENFKMFWFLWRLLGFRGFQNKYANIVHNLVLHVAISFWYPMHLTL  
GLLSLPNQGEIFKNLSITITCIVCSMKQLFLRWKIRQMHDIEMLFLELDASVESRQEYHF  
FTNGPRKHAQWITKLYCTCYMGANVAITMVMLDSQRRLMYPAWFPFDWSSSSQVY  
WAVLMYQFMGVTTQIVQNLVNDAPAGVLLCLISGHVRLGMRVSRIGHDSKKTENEN  
LADLGKLFKLVEDTQSYVQLILYISGGLNICVAVVYLIFVESLTAYLYYSAFILAITIEIY  
PSYYYGSSCQQEFNDLSYAIFCSNWLEQPKRFHKNMRIFVESTLPKVTMTAGGIVRMQI  
ENFFAICKMAYSLFTLIRSIK

>MdOr22

MPSSTKHFFNSSLNTRFPVIYKVIFYFSIFCRRIFHLAHMDIDAPLPKTRDATVYIFRGLNII  
GYVPTETNKLAFYMWSGFVNFFVTVYLPVGFLMSFLLRLNTFSPSDDFTSLQIWVNCIG  
CSLKMVFVFFLHRRILIESRKFMMDRLDVRIDNDEDRLVIRKIVAFSNRSLTLYSSLYLSYA  
SSTFLVAVINSKPPYQVFNPFLLWKENVWKFTMQAGFEYMMIAFHCFQQALLDSYPVI  
FITIIRTHLHILTRISRLGSISTMTSDERYEALVQCVLDHKNIMGLYSIFCPVISGTMFVQ  
FLIIGLILGITTLLHIFLFADRLAIIASLFYVASILAEFTPCSFLANCLMDDSDRISLAIFHSA  
WHEEPRYKQMIFFLQHTQKTLILTAMKIFPITLNSNINVVKFAFSVYTMMKQMGLG  
QNLQNVVGKEL

>MdOr23

MMEENRMVSINIKIWKFFAIYPTSDKLWRLYSIQFVTILLNFMQFMFLIEMWGNLAPFI  
LNVFYVSATFDCLLRTGVIVYNRSKFEAFLAEFDSMYSEIEENGDDYAKGKLKEATEF  
CRKFSLFNVLASFLDLIGTMSHPILTGTTRTHPFGVALPGIDSAVSPYYEIFYFILQLHCPITL  
SVLYMPFVSIFVTFSSFGKTALQILQHRLKDIFEIYDDDETRLEALKECAHYYNRLTRFI  
KVFDDEMVTYVILGEFLLFGAIICSLFCINIIDTMAQFVSIIMYVGTMLYVLFACYYSAN  
EMLEESLKVSEAAYSIPWYEGTPQFRKTLLLFIQRTQKPLCLTVGNVYPMTLLIFQSLN  
MSYSYFTMLRGLKIQ

>MdOr24

MFSVPNPPDALPPQNSLKNFFLIQRICFSVIGLDPTSLKRTMYRPWLTFIPLLSLMGLLGP  
MGVYAFNYLKIDLGKAITALSPFWQSMLSTIKFFVFMNLNRKKIVGLVRKVWSWTLEAT

EEELKIIDEEIKGDARISLFYYSMVNITGVLAALAPLAISAIYTFHGRGFMETLDAPFKAE  
YFYDIRASYMGYILCYTWNVLGIHYILNGALSIDTLYSWIVHNIAAQFRILNLRQRQSE  
KIIAHQAAGNHNEKEFLKSVVECVNYHRRIIQMSERFSEVYQGLVFIKFLVSCMQLACL  
SFIPLGGEFADQSFNLSFLIAVTTQLMLYCHGGQKIQDMSTSVNLAIYEYFHWHDLSIK  
SQKLLMITMIRAQKPCDIRGIFFTADLSLFWVYRTAASFMTMLMSMQDK

>MdOr25

MFNVPKAPDALQPQTSIKKFFLIQKISFAAVGLDPTSIRRTIFRPWLTFIPLVSIIAVLGPM  
GIYAFNYLKIDLGKAVSALSPFWQALLSIVKFFVFMNLNRKKIVGLVRKVWLWTLEANE  
EELKIIAEENRGDAKVCTFYYSMVNITGVLATLAPVAVAAIYAWQGHDFWESLDAPFK  
AEYFIDIKASIVIYAACTFWNFIGIYYIVNGSLSIDTLYSWIVSNISAQFRILNLHYHQLSQ  
NIIAHKAMGNHNEEKFLKSIIDCVKYHRRIIQMSERFSEVYKVLVFFKFLVSCLQLACLS  
FIPLGGEIADQLFNLSFLMAVTTQLMLYCHGGQKIQDMSISVNWAIYESFHWHDLSIK  
SQKLLLLTMIRAQKPCIEIRGIFFKTDLSLFWVYRTAGSFMTMLMSMEDK

>MdOr26

MLKPPIAPDSLPSQTSIKNFVFIQRICFWAIGLDPTSIRRTIYRPWLTVIPLLAMIGLLGPM  
TAYAFNNLKMDLGKAISALSPFWQAILSIVKFFFFMVNRKKILQLLRDVWLWTLEATA  
EELEIIAEENKNDKICGFYFAMVNISGVLHLAPLAVASVYAWQGNGLNSLDAPLK  
AEYFFNIRQSYITYIVCYLWNVISIYFIIYGSFLIDTLYSWLVHNISAQFRILSLRYRKLSL  
MMVTHKSSEIQNDEIFMKSIVECIQYHLRILEISKRFSEAYQHLVLIKFLISCLQLACLSFI  
IPLGGEMADQLFNLSFLVAATTQLILYCHGGQKIKDMSTSVNWTIYESFHWHLNLSVKS  
QKLLLFVMMRTRKPCIEINCFRANLNLFWVYRTAASFVAMLSLQNKI

>MdOr27

MFKIPRAPDALPRQPSLRKFLYIQKICFAGIGFDPTSIVKRTIFSPWLTFIPLFSILGLLAPM  
GVYAFKYIKIDLAKTAAALSPFWQSLLSSVKFFVFMNLNRKKIVESVRKVWLWTLEANE  
EEVEIIAEENKYDARISKFYFASVYVTGVLAVLAPLAIASVYAWQGYGFLESLDAPLKA  
EYFFNIRGSYQAYIFCYVWNCIGIYYVLHGALSIDTLYSWFVHNISAQFRILNLRQRQSL  
ERTMMLRAIGEHNEEKFITAIIECVKYHRRIIQMAERFNDVYKGLVFIKFLISCLQLACL  
SFQIPSGGEIADLLFSLSFLISVTTQLMLYCHGGQKIQDMSTSVSLAIYEHFQWHDLSVK  
SKKLLLLTMLRAQKPCYVRGIFFTTDLSLFWVYRTAGSFMTMLITMDGKK

>MdOr28

MTSDDLPPLEGVKYYFVVQKFCFTAIGVDALSARRTIVNGFLFWIPNIVQFILSQPLTLY  
SLQHLEDMSLVTDAMAPVWQVLMANMKMALFLWHKKEMKKLVRDLWLWNLEATP  
DELKILEVENRKDTMTSFSFYMTVLTGILALTSPFFKAFYRYLKGDNYWDALETPLK  
GSYFIDPKETYMGYFIAYMWAFIAIYAVLNTTLAADSLFSWIVHNISAHFWILRERLKSI  
AATNREGSHGYGKFRKSGDCVRYHQRIIDTIDEFNKVFMTIVFVKFLISCIQIAFLAQF  
VRGGDFAGQVFHMLFLMSISIQMMLYCYGGQRIKDESASISVAIYEYFHWDLCPKSR  
KLLLLPLARSQKPCCLTGVFfiADLSLFLWVYKTAGSFVTLMMSVSDTSN

>MdOr29IP

MATPSADVLPPLLEGVKYYFVVQNFCFRAIGVDLLSMKRTMVSGLLFWLPNILELAICV  
PLARYALENLEDMSLVTDAMAPVWQVLMAILKMALFMWHKKDIKKLVWNLWLWN  
LEAKQEELEIIADENRXDTVKSFSFYMTVLTGILALTAPYYYVDPKGSYLGFTVHIWT

CIAIYAVLNTTLAADSLFSWIFHNISAHFAILRERLICVAFSETEGKQSYANLKKSLAEY  
VRYHQRILDTIDDFNEVFMMIVFVKFLISCIQIAFLAFQFVRGGDFAGQIFHMFLLTSISI  
QMMLYCYGGQRIKDESIASVVIYEHFQWEVLCPKSRKLLLLPFARAQKHSELTGFFFT  
ADLSLFLWVYKTAGSFVTLMMSVSDTSK

>MdOr30JI

MTSDDLPPLEGVKYYFVVQKFCFTAIGVDALSARRTIVNGFLFWIPNIVQFILSQPLTLY  
SLQHLEDMSLVTDAMAPVWQVLMANMKMALFLWHKKEMKKLVRDLWLWNLEATP  
DELKILEVENRKDTMTSFSFYMTVLTGTGILALTSPFFKAFYRYLKGDNYWDALETPLK  
GYLGYFTVHVWTCIAIYAVLNTTLAADSLFSWIFHNISAHFAILRERLISVASSETEGKQ  
SYANLKQSLAECVRYHQRILDTIDDFNEVFMMIVFVKFLISCIQIAFLAFQFVRGGDFAG  
QIFHMLFLTSISIQMVLYCYGGQRIKDESTSIAVAIYEHFQWEILCPKSRKLLLLPLARAQ  
KHSELNGVFFTANLSLFLWVYKTAGSFVTLMMSVSDTSK

>MdOr31

MTRILKRYFRLQRFI FSGGLDIAATPEKMKVCRPWLMMTPLVMSILLCIANGHYVLDN  
ASDYLEATDSLTLCLQSLISVWKVIMVIWKRKEFANMIARIELNVKAEGEELKIVRRE  
NTRDIIFSTTYFVLVLLTGAWSLVPIYFAVHVYVTTGEVDLPVPHKATYFWNHEHVK  
GYSLVYIWDVFIYFIACSAVSTESMFSWLVCNIIAQFRILMHRLEVASRQVMSTRPMTA  
SHHVDDDDDDNPLMGELDPQAGMVDIIACVKFHRRTLRLTQELNSLYGAIIFVKFIVSG  
TQICCLAFHLVRGNNSLNFVAYLCMFLSAAALQLILYCYNGQRLKDESLVTTKIYSIF  
PWSKMPVSTQRMILLIPMIRAQQFSELRGVFFTVDLSLYLWVFRTAGSLIAALKTLEEKE

>MdOr32PSE

MKIVKRYFGIQRRTLTAIGIDVNAFLPNGPERIAKHPLLLL VITVMPVLQYISIGHYAYK  
NSNNMVTATYSFSLSCQGVICLTKILIFLKRKRDIVKLVKMLQEDVFNAKSDELVITKEE  
NSRDVLHCTVYGSAVFSTGFFGILHRLLRPSSSTSNMGIWCWYHHILPXYLWDYSHLP  
GYSLVYIWNMMRMYTLAFASVAIDSLFSWLVCNIVAHFRILMLRFQRAAWLTPGLDR  
PEVSVSREQERLIFDCVRFHNRNLVQELNLVYGGIIFVKFVVS SVQICCSAFFLNSFG  
ASQSMAKL MYQFLLLSAVALQLMLYCYNGQRITDVSFQVATKVYSTFPWSKMPASTK  
RMLLPPMIRAQRFSSELRGVFFTVDLSLYLWVFKTAGSLIAALKTLEEDK

>MdOr33

MKILKRYFGMQKFAFAALGVEVESMSPAGSERIFRHPIRYAVLFILTVLQYISIGHYSYV  
YTSDIVSAAYSIALSCQGVICITKLVIFFFKRQGIVELVRMLQTDAFNAQSEELAIKEEN  
RKDIRICTLYCIVYGTTFFGMTLPFARTILGYLRNGYLVYVTPVASPSLWNYD TVHGY  
TLVYIITLLRLGTLCFTTIGIDTLYSWLMSNIVAQFRILTHR FQQA AWATTALDGSEISIS  
EEQHRLINDCIRFHNRTL DLVKELNRVYGAITFVKFVVS SIQICCSVFFVSSSDSKESAFN  
LFYQSIFLGAVSMQLATYCYNAQRITDESELVATKVYLIFPWSKLP IPTQRMLLLPMIRA  
QRSCEMRGVFFRIDLSLFWVFKTAGSLIAVLQTIDEAQ

>MdOr34

MNSREHRELLEIFYKKQSYVFRLLALWKL PDTVTERFRLLRHFYFYI LFFWVLSFDAS  
CMIQFIANITDLNEVIK VFFIFATSLAVFAKFATIKLKNHLYAELIETIHEPAYRPVNSRE  
VKIFRQTHRLCGTVRN FYLVISLCALNVVMLTQYIFDNSELPLSLYNPINIDTKLRYRLM  
YLYQYVAVSICCYMNIAFDSISASFMIIHKGQLDILCDRLEHLGMDQESRDEDITRQLK



NCVKYYGDIIHIVRIAENLISFPISIQIACSVLVLVANFYAMSFLSDPGDYANFIKFLIYQL  
CMLSQIYILCYFPSEVTAKSEEVPYYLYCSNWVYWNRNMNRKLTLLMMTRFDIPIRIRSI  
NPTYTFNLAAFTSIVNSSYSYFALLKRINS

>MdOr35

MNSLEHREAMKTFYKKQSFIFRIFAQLKLSDTVSDRFRLHRIYFYIILIGWVLSFDISC  
LIQFISNITDLNEVIKVFYIFATAMGVLAKFLAIKIKNNLYAELIEAMHEAKFRPTNSREL  
QLFRESQRLARTVRNFYTTISLCALNALLFTQYIIDTTQLPMSIYNPINTDTKLRFVLVYI  
YQYLAVSVCCYTNIAFDSISASFMIHAKGQLDILCDRLKHLGMDSETSDEEITAQLKNC  
VKYYGDIIHIVKIAEDLISFPISVQIACSVLVLVANFYAMSFLSDFANFIKFLIYQLCMLS  
QIYILLYFPSEVTSKSEEVPYHLYCSKWANWSASNRKLTLLMMTRFDIPIRIKSINPTYTF  
NLAAFTSIVNCSYSYFALLKRINS

>MdOr36

MFHHKRELIRTFYIRQYQLLKLFAWQLPEDASAYQRLGYRIFFWCFLIFWMLLLDCC  
MILQIATHLGDVDEVIKVFIIFATAFAVMGKYLYLKIYNRYFEQLFQMMHQPEYLPENP  
TEWQIYCQAIDLSRRVRNYYASLSVSALSALFLSQFLGDEQELPASIYYPFQLNTNWKY  
GLMYVYQCVSLAILCFVNVGFDSLTAFFINIKGQLDVLGMRLQTIGVGVRDQRRILKK  
LKDCIRNYQRILRMTHLMEELVRIPMSVQIGGSVFVLIANFYSMSMLSDNADMGIFAKL  
LLYQTCMLTQIFILCYFANEVSLKSSDISFNLYESNWYDWDKVNRLVLLMMIRFDTPI  
SIKSINRCYSFNLAAFTAIVNSSYSYFALLKRINS

>MdOr37

MAEVERYFEDFVNLPVLLKTLGYDFLEISRPWLARWLMKLYFFLTLICCLYCTYFVT  
DEIFADIVSGANNLPLLLRLINDFNNAIGILKSFYFFRNIKSKKELFRKFREIFPTSIEDRF  
AYRVNESYWPRWITTTLYLYFCATALILFSPLAESIIEYFVDLIKVGADAEFTYHKLYE  
EQSYVVDHRNPLGYMVIYSMEVMNSHYAIVFNICPDIWLIAYAIQLCMHFDYISRNLES  
YEPMEKRQQKDLKVMAELVRKHQVLLELADDLKEIFSLVLVMLFSTVATLFCAAVY  
VLTQGINKNVLGYMAFLPTSLGQYFMVCYYGQLIINKSLQIGEAAYSQTWYNGCQSY  
KKSILAILGRAQRQCEINAGGFQTTNLKGFESVMRMTFQLFTLWRTMMEPK

>MdOr38

MYSVFQQPLTVMATTERYFEDFVNMPCALLRTLIGIDFLNISRSLLAKCLMQLYFVLSLL  
SCFYCTYFVMEMAVREIHCGSGNLPLILRLVDDIFHSLNGLLKSYYFFRIWKSNNKSLFN  
RFCEIFPISMEDRREYRVNDYYWPRWITCMVYVQCGAIAVIFSPFAATLKDYFLAILKF  
GFSDAKFSYHILYEEHTYIVDHQRPTGYIFIYSVLAMGTQYAVIFNICPDIWLVAAYAIQL  
CMHFDYISRNLENYEPKEERSHKDLEVVAKLVKKHQILLDLANDLRKTFSILVLIMLFS  
TVVTLFGAAYVLTQGINSNVLGYLAFPTTLGQYFMVCYYGQLIINKSLRIGDAAYS  
QTWYNGCQSYKKSILAILGRSQSQCEINAGGFQTTNLKAFEGVIRMTFQLFAVWRTLM  
EPK

>MdOr39

MKVTAFASSSALKTAEKELYFDDFVKLPCVLVRTIGYDFIDKPRPLWLRALMLLYLVLC  
LIFCAWFTYFAWDFMMAEIAAGANDLALVLRLSVDVIYNVAAIVKSLFFFRNLKSLKS  
LLQRFDRDIFPISREDRLAYRVNDYYWPKWITTILYMQLFALSIIFLPFVEAVYEYFGAL  
LTVGYANAKFGYYRMYPETTYGINHYNPLGYIIVYTMDIMNGHYCTVWMMGPDVWL

VAFSIQLCMHFDYVSRTLENYKPSKERAQAQDLRVLAELVRKHQTVLELADDVQENFS  
VLILVMLFSTASILFGAAELVITQGITAHLVGLYLAHVPTGVGQFYMICYYGQLINKSLQ  
VSEAAYNQTWYNGCQSYKKSILTIMRRAQCHSEINAGGFQTTNLMAFESVMRMTYQL  
FAIWSTMTSSK

>MdOr40

MTTKERTFADLAKLPCVLLKTLGYDFLDQPRPRWLRMLLTLYFVLCLMCCSYFTYFA  
LDFAVAELAVGAKDLPLLLRLIDDIVHNVVGILKSYFFIRNSRSIKKLYKKFGDIFPISME  
DRLAYRVDEYYWPKWITTILYMQLCALTIILFVPFAESIFEYVGALISLGYGNAKFGYY  
RMYEETSYGFGHHNFLGYVVSYSLDVMNALYSAIWMICTDIWLVAFALQLCMHFDYI  
SRTLENYEPHKERSQDDEKVLAGLVRKHQTIELADELKINFSALILVMLFSTISMLFGA  
AELVLTQGITTHVIGYLAHVPTSVGQFFMVCYYGQLINKSLRVSEAAYSQTWYNGSQ  
SYKKSILTIMRRAQRHSEINAGGFQTTNLMAFESVMRMTYQLFAIWSTIMESK

>MdOr41

MSIVRVKKARVNFQRDFRDFCHLPNYLMRIYGRDFSERKRTKWQTLRLRYAVVTVS  
SHIYCFYFISQQVFLMFLSGVPNLEFLRLLSGFNYGLFAIMKYLAFAKNRITDAAAINRV  
LREIYPKAGRERILYRVNAFFWPKWMLTVIYFYFGAVAFIVLSPLLESVIVFVIGVGRLG  
WNEAQFGYIKLYDIPYSFDHRSPFAYVLTYSIELFHAQFVIICNVCGDIWLLCYAMQLC  
MHLDYLIKILEHYEPRVEHHLRDTQFIAGFSQKHQILLNIADDVNTVFGVQLLLILISTA  
ATICCAGIYTLTQGVGKELLEYYAFLPCVVGQYYLICFYGQRLVSSSENVGAAAYNHA  
WYNGSPSYKKSVLVIMTRSQRSMKLKAYGLSSVSLGSFRMVMSESYRFFAVLKHAVF  
DKKN

>MdOr42

MFEDIPLIYMNVKILKFWSLLYDHNWRRYVTLIPPTFLVFTQFYMFMTTEEGIDAIIRNS  
YMLVLWFNTILRAYILIKDRVEYQSLLQDLEAYFYDLDKSNDVYVRNLLSHVNSNGK  
VMARGNLFGLLTCIGFGLYPLLAERVLPGFSIIPGIDEYQSPFYECWYVFQMLITPVG  
CCMYIPYTSLIVSFIMFGIVMCKYLQRRLATLSRFKGQPEWIYDEVIECIKYQKKIIEYCE  
TVNRLTTFMFLLEFVAFGTLLCALLFLLIFVDSAAQAIIVCAYITMIFCQILALYWYANE  
LKEQNLSIAAAAYETEWFTYEIPVQKLILLMIMRAQKPKTIKVGNIYPMTLELFQALLN  
ASYSYFTLLKRVYG

>MdOr43

MAPSMEINSNEFFKINRTCWKLGLGMLMVEGHKTNGQRKMSTNLYMVWAIVINLM  
ATCCFPIHLFLGIFESENKTSFFDSISITITSIGASTKLLIIAIIKMKKILEMQSLLRTLDRIT  
HHEEVRHFRQDIRSRIMNIQRLYFVVYCGVGISVLGAFLFSKEQRLFYSGWFPFDWRSS  
LGNYAAAIISYQCIFFQMMQTFCNDSFSPIALCVLSAHIELLYMRVVRIGQDKNGKMR  
ETTTLQEDEEELNRCVLDQMNLYELYNTMQNIISWAMFIQFFVSVVNNCVAIVALLFF  
VTDVFERIYYVIYILAMGIQLFPTCYYGSDFVLLFEKLHYAVFSCNWIGQSKSFKRHMM  
IFTERSLRETVALAGGIFPIHLDTFFGTCKATYSLFAVVMTMK

>MdOr44

MTEEPNTKALFKTHFIAWRILGMSPPDNYRPLYWIYSILLNIFVTIGYPLHLIFGLFTSTT  
MYEIIQNVAINFTCSVCAMKTIAIWWRFNKVDVMFEIIQRQDQRFTSHEEIIAYLRKEVY  
PPVRRRIILLFSILCTFIGISGESAVLVTGLLGTWNLMYKAYFPFDVFASTKNYMAAHLYQ

FIGISYLILQNVVNDTFGASHLCLLSQVRMLNIRVTKIGHDPKKSREENNQELLECIKV  
HKDLLEYRRQLEEIIISYMFQILIAALNMCVVLVFIILFVRDIFTLAYYVSYLTSMIFEIL  
PSCYYGTLLDEFEDLAYALFSCNWPQKQTLFCKNLRIVAEQAKRRIYVTAWLFRINN  
NAFLIACKNAYTLFALVMNMK

>MdOr45

MSETKLHTKSLFWAHFACWPILGMMTPPNVKYKALYWIYSFAVVITLMIGYPLHLILG  
LVSSSSLKELMQSLSTLTSTVCSIKTMAIWWRLNKVTDMMFTIIRQDERVRSTEEVDY  
MKNVVYPQVRVFIRLFYVICGFLSLFGELSLVVAGLLGNWRIQYKAYFPFDPYANTKN  
YVIAHVYQLLGVNFTLVQNIIVNDTFASSHLALLRGQVDMLARRVAKIGHDPQKTQRE  
NNQQLLECIRDHEDLLEYRQILEEIIISVYMFQILLCGLNMCVILVYMVIFVRNDVITLS  
YYSTHLIGVMCEILPSCYYGTLLDAFQDIAHALFSCNWMQDQDLEFKKNLRFIENSRR  
RIYVTAWLFRINNNAFIVACKNTYTLFALVMNLK

>MdOr46

MDKELNTKSLFRTHFKCWRILGMMPSKKYRLLYWIYSLIVNLLVTIGFPLHLILGLFQS  
TSLYEVIQNLAITLTSTVCSMKTFAIWWRFKDIERMFDIIRKQDEHTRHGEQLEYMKRK  
VYPPIRSLINLFYILCSMVALSAESSLIFNGLRGSWALMYQAYFPFDPFGSSGNYVVAHI  
YQFIGIITYTVTQNLVNDTFAGAHLSLLGGQVRLLGMRVAEIGHDPKKSLEANNKALLD  
CIHDHLDLLEYRRKVEDVISLYMFQILFSSMNMCVVLVFMLLFVKDTFTMSYYLFYF  
VGMIFEVLPSCYFGTILEDEFQELSYTLFKCNWADQNVVFCKNLRIFVEQASRIHVTA  
WLFRINNNSFVTAVKGSYSIFSLIMNTR

>MdOr47

MALLQNKLNKSLFNTHFMCWRILGMLPPQNYRPLYWVYSFIVNLMVTIGYPLHLILG  
LLTSTSMYEVQNLAITLTCTVCSMKTFAIWWRFQDVDRIFDIVNRQDEHTRYGEQSDY  
MREKVHPPIKWLIILFYILCSMVAISAEVSLVNGLRGSWLLMYQAYFPFDPFGSSMNY  
AVAHYQLIGLVYTVTQNLVNDTFAGANLSLLGGQVHLLGMRVANIGHDPNKSMEEN  
NKELLDCHDHLDLLEYRRKVEDVISLYMFQILFSSMNMCVVLVFMLLFVKDPFTMI  
YYMFYFVGMIFEVLPSCYGTILEEEFQDLAYSLFSCNWTEQDQVVFCKNLRIFVEQASR  
RIEVTAWLFRINNNTFLTAVKGSYSIFSL

>MdOr48

MADELNTKALFKTHFVAWRILGMLPPTKYRPLYWMYSVFLNLAVTIGYPLHLIVGLFT  
TTTAYEVVQNIAINLTCAFCAMKTIAIWWRFNKLDIMFEIIQRQDERVISEEGVAYVRN  
VVHPPVRRRIILAFILCSVIAASGESSVLFNGLLGNWTLMHKGYFPFDISNNTRNYAIAH  
LYQIIGLSYMILQNVVNDTFAASHMCLLRGQVQMLNVRIAKIGHDPKKSREQNNQEFL  
ECIKIHKDLLEYRRQLEEIIISVYMFQILVAAFNMCIILVFIILFVKDVFTLIYYILYFSAIV  
FEILPSCYYGTLLDEFQDFAYALFSCNWPQDQDVGFCKNLRIVAEFASRRIYVTAWLFR  
VNNNAFIIAVKNAYALFALVMKVK

>MdOr49

MMSEKEVQMLKKSNNYKIKELIRISFTLGVNLTSPSTLKDSLKIINIILVSSVISFYGHV  
CYTIESIKDIPKIAESVCTGFQTLISVIKMVYYLFIQRRLLYYLLYKAQTHEYIRKIDIFHKN  
FPMSERLQAKIDEILDASWKNINGQLIFYICCCAAIISNYFFMALFQNIYHTWKETPNYE  
FVLPFPSVYPSWKDKGMSFPYYHIQMFLGTCSCYISGMCAVSFDGVFIVLSVHGVGLV

KVLNMLIENSTSADVPERKERRVEYLRYCIYQYQRISDYTDELRKIYKHISLTQFLLSLLV  
WGIVLFQMSVGLSDMLTLVRMIMYISAAGYEIVLYCYNGQRLTSECEKIPYAFFSCD  
WFNESKEFQELTRMMILRSNRSFFMEISWFTTMTLPTLMAMIKTSGSYFLLLRNVAE

>MdOr50

MSQLLLDLLKEKQLENNKILNTFYRISFMTGVKIKYKTQFKDPVKLINLFLISVSLVGLC  
AQYCLVWNKRKEPFVESADAICTANQAWISIFKLIYLVFVQHKFYELLHTAINGSLLYD  
LGIFDLAIDCKQYLLQEINTILDSSWRHIKYQVNFFTFSCMMACGFYMFSCIAANYYYT  
NIQPQNFTLQLPMPALFPMWHDYGMTWPYYPIQYFIAGIENYICGMCAVCFDGIFIVIV  
VHCASLFEILHMLLEHVDDIPQSERVDYLLCCARLHVRIYNYAKINGMYKNPSLAQC  
VLSMLVLCVVMFMASIGLEEDITLFVKMLCFLCAAGLQIAIYCYNGQKIITQSEKSPDA  
WYNCCWYNESKQFKYIIDMMIMRTNRTLQVSGFTTMSHMTLLSIVQTSQSYFLLLK  
NLNGID

>MdOr51JIP

MYLALKETKANQILKYWKWIAFTSGCNIVYKTKFMKLFKLILNMSLAISAAIGCYGQA  
QFFWNHRHESFDVYLEAILIFFQIVISISKLMFLTLKQQQIFEIVQDVQNGEILNDLEIFEL  
NLINPSKILKDISAIMDQSWMSIKFQLNFFIGNVVVLCGVYLFKNLILNIHNFKNEGDRF  
QLAYAITFSGFLALISTHCRGLLRVLRITLITYSTTYHVLPEDRVLYLQGCILKQKIKYKI  
CNELNSLYRIPALAMFLVSLVICLLTFYATVDGGNDISTIVKVILFISGAYFEVAIFCFN  
GQHITTESEHLPLDIYGTXNGTKRANNLRRFMIQRSNITILMDVGGFTTMSFVTFLTIFR  
SSLSYFLFLQECM

>MdOr52

MDTILIDISDKGGRILNPLKWIGMFSGCNIKYKSKFLHPLKILNLFVTSILACYGQLYY  
VWERRHYTFEIIYIEAILIFFQSLISIWKLWMFTFSQDCLFDMMKSVENSETLQNL EIFQL  
ELIDSANIINDITQILNESWIDIKRQLLLLRFTVFGICSWYTGHSLSVSNIIYYLYISDENDKE  
KLEFPFPASFPVWYSNVNSLWHFYLEYFVVTMQIYLATVASITCSGLFSVISVHCLTML  
RVLRTLITYSTSEHVPSQHRTKYLEACVRLHQNLISFCSRLNRVYQKPSLGLFISCCLLI  
CLLTFKASVDLGKDISGSIKVCLYLLAAFYELLIFCLNGQRITSESERLPQAIYSSLWFDE  
NRNFKFMIQIMIMRTNQIRMDVGGFSSRMSLETLLTITRSSVSYFLFLRNCM

>MdOr53

MAGNIQLSPSERFAKFIKVIKLFAGFCGVNSLERDYRVTWVTWLVICVVTSSFFVCTFYT  
IYVGMAIQNNYSILLQSLCITGTGVQGYTKLLNAIFCGKHLRFAFEELTAIYEEYECKRL  
EYRDNLKENLEMVKRLIYGLLLINFILIAALFAVPLFYYYVRKEKIDVIPLMIPGINPSNN  
RIENYIYQFYHICCVIFSTFGNFASDTFMILIVVHVPMIKNIFKLKFDDMAETMKLHLRN  
RKKTEPLLRDIFQWHQKTILIIETMQKGFFWVIFVQIFTSMLNIIFTIVCIFLGWVPVAPV  
YLLYSFVILYIYCGIGNLVEISTDDITSIIYDFIWDLTVSEQKMILIMLRESQSPPTMTIG  
GVMPLSMNTALQLTKSIYTIAMLLNEFVN

>MdOr54

MATKCLKLTPSQRFNSFVRVVKIFAIVCGANIFRPDYRLNALTWVIGVIATFFIFTSYTM  
YVGVIDNDYTKILQLLCVTGSAIQGATKLVNGLYHASLIRSLIAEILTMYEEYECKDQ  
RYIKYLEHTLSLIKRAVFSLLNIYSIQTIGVLA VPLFYHLLLGQQIDIALLVPGIDKHTDF  
GFYTYQFYHFCVVGFAFGNFANDTLMVLLIVHVPLMKNILKLKFDALDELLKEFPRD

VDRTEPLLREIFQWHQKSTMFAQNCTDTFFWVIFVQIFASTLAIICIMVCQFLGVWPAA  
PVYMMYCFAIMYMFCGLGNLIEISNDDLTRIIYDCNWYELTVTEQKMILLMLRKSQQA  
PTMTVGGFMPLSMNTALQLTKTIYTAAMILNEFVN

>MdOr55CTE

MVISVVQRNEFIVRIIRLSSKYCGCDVLNPEWQMNLLTWTVITFINMFSILTCTVYVSI  
YLEGEWSHSLQALAMVGSGVHGYAKLLNAIRNKAYFRFLVDELHTIYKEYNEKKHSD  
YRAYLHKTMNRTVVGLKSMGIVYAIVVCSLITVVPFYRFFFNQRVFIMQFLLPGLDPK  
VERDFIIMNVVHFFSILFGGFGSFAADLCFVLLVFHVPQYKDILSCKFQEINEALELDEM  
ERSGELLRDIFEWHQRYMKFISIVKENYFWVILVEMATIFLCLALSLS

>MdOr56

MTVSVVDEYEGIVRLIKLCSGVCGANVVFVANYKVNVLTRIVVTFINLYFIFTGYTLYINI  
FIEKDWTHMLQVICFFGSALQGYCKLLNAIWNKDHLRYLVDDLREVYAEYAPKHDEY  
RDCLQKSINTAVKCIKLMAFFHVAITVGLIGVVPFFRFVNERIFVMQFQLPGVDGDTE  
YGYLIMNCMHSCIIIFGAFGNFAADLCFFTFVSHFPLFKGILSCKFHDLDNDVLEGSDDAK  
KAECKEMLKDIFRWHQKYMRYITTVKDNFYFWVLLVEMATIALSISSTLFCLLLTWPG  
GQTYLSYCFIMLYIYCGLGTVEVTNDSFTDLCYTQVIWYKLPAAERKMMLMMLMM  
AQKTGGLTIGAVIPLTVNTGLQLTKLIYTLTMMLINFLD

>MdOr57

MAKTVAQSYDKTILFIKISSAVCGANVLSPAYRMNILTWWIVVCINLYYVFTGYTLYVNI  
YVEKDWPVNLQVLCYLGSAMQGYCKLLNAIHNKESLRFLDQELREIYLEYDQKHADY  
RYCLKTTIDRANKFIKFMIIFQILISGSLIGVAPFYRLVFNQRIFVMQFLLPGVDPSTEYG  
YFVMNCMHCIIFGSFGNFAADLFFFVVVSHVPMFKDILTCKFHDLNDLLEEEVADNE  
NNNNNNRIKDVREDFRSLIDIFKWHQRYLRFIAIVKENYFWVLLVEMGTVALSLASTL  
FCLILGTWPGGQSYLAYCFIMLYIYCGLGTVEVTNDGFIDSCYTEIHWYKLPVSQRKM  
LRMMLMMAQNTDGLTIGSVIPLSMNTGLQLTKTIYTMTMMLINFLE

>MdOr58

MAKTLVQRYETIVRLIKICSAFCGANIFHPSYRMNILTWTVVIFINLYFAFTGYTLYVNI  
YIEKDWPVNLQVLCYLGSAMQGYCKLLNAIGNKDNIRYLTDELREIYRKYDLKHTDYR  
CCLQKSINTVNRFIKCMIIHFSITMSLIAVVPFHRVFNERIFVMQFLLPGIDPNTAYGY  
LMMNCMHCIIFGSFGNFAADLCFFTIVSHVPLFKDLLRCKCQDLNDILEEGKDVEQE  
GIGDCQILLKDIFQWHQKYMIIYITTVKDNFYFWVLIEMGTVALSLASTLFCILGTWPG  
GLTYLAYCLMMLYIYCGLGTLVEVTNDGFIDSGYTDVIWYKLPVERKMIQMMVMM  
AQNTGGLTIGSVVPLTMNTGLQLTKAIYTMTMMLINFLE

>MdOr59

MNLEDNRNANKLHRPSNRLRKIVRITRISYICGADVDPNYCVNIRTYFVLAVINFSIL  
LLSYTMYSGWVEEGDWAIVLQVLTIGGGTLLQGYCKLINSIRQKDKFRLLTEVYSIFE  
EYELKSCDYARHLKKGCHLLSYFMKLCVINVMMICGLILVAAAINVIFQKRDLIVYG  
DVIGIDPSTTSGFYVTFMVQACFLLVGGFGLYAGDMAFFTPISQVPTLKEILRCKFKDIN  
AAMEGDELQDSRHVSELLKDAVQFHQKYLRFLNTTQDTYYWVILTQISTYSVGIVCSM  
FCIFLGTWPGGYIYLLYCFVMMFVYCGVGTMDIANEGFIDACYNDILWYKLTAASDR  
KSLNMLILCQNTDGITIGSVLPLSMNTGLRVTKTIYSIAMMLINFFMD

>MdOr60

MAKTHTERLLKIVRITKFCSDICGVNIYEDDYRINYRTFFVIAVIGTSFSFSLSYTMYDGY  
GKEGDWTILVQVISLAGGTLLQGFFVLILFLTKQEKYRFLKECIILEYEYKMDSDYR  
VYLNKGIHLLANFMKVCAFINFMLVLGMTFTVTFYNYLIFGTNETLVYGYCPWVSLETT  
GGLWTTNMVQALLIAVGGFGLYSGDMSVLTPISQIPTFKGIIQCKFRELNDDLLDDDDHES  
EMAKKIKTLAALKDILQFHQTYLRFLDVSREAVYWSVFKVGTCTFIGIAFALFCILLGS  
WPAGYIYMLYCFVMMQVFCGMGTLDITNEEFHSCYNDVRWYDLTISEKKMLNIML  
MMAQNTEGLTIASIMPLSMNTGLQVTKTIYSLTMLLLTFVN

>MdOr61PSE

MAKKHSESLKIVRITKFCADICGLNIYADDYRINYRTIFVVLIISSFTFLTYTMYDGY  
GKEGDWTILVQVISLAGGTLLQGFFVLILFLTKQDKYRFLKECIILEYEYEXMDSDYR  
VYLSKGIQLLANFMKVCAFINFMLVLGMTFTVTHIYNLIFGTNETLVYGYCPWVSIETTG  
GLWATNMVQSLMIAVGGFGLYSGDMSVLTPISQIPTFKGIIQCKFRELNDDLLDDDDYDSK  
LEKESKTLAALKDILQFHQLYLRFLDVSREAVYWSVFKVGTCTFIGIAFALFCMLLGS  
WPAGYIYMLYCFVMMQVFCGMGTLDITNEEFHSCYNDVRWYDLTISEKKMLNIML  
MMAQNTEGLTIASIMPLSMNTGLQVTKTIYSLTMLLLTFVN

>MdOr62NTE

GFCVFLTFIKEQENLRFLLTECYDIYEKYERMDSYRVYLDRGVRLAKLMKLSAFIN  
AMLVFGMSSFTFLYNFIYGTKATIVYAFAPGLDVATPVGFWATNFIQAGFIAVGGFGL  
YSGDMSVLTPISQIPTFQGILQCKFREINQLDDDDYESAEERGIKTMAALKDILEFHQKY  
LIFLKVSREASYWSVFAKVGTCHIGIVGALFCIMLGSWPAGYIYMLYCFVMMQVFCVM  
GTLVQKTNDDFIHACYNDVRWYDLTIREKKMLNIMLIMTQNTKGLSVGSVIPLSMNTG  
LQVTKTIYSLTMLLMNFVIENEA

>MdOr63INT

MNLEDTRNINQVYRPSNRLRKIVQITRMCSADVCGADVFGHGYRVNIRTYMVLGIINFSI  
IFLSYTMYSWITEGDWTIILQVLTIGGGTLSQGYVKLVNSIRQQNNFRYLLGEVYSLFE  
EYELKSSDYAVYLLKKGCDLLSYFMKLCAVINVIMICGLILVAAAINVIFQKRQLIVYGQI  
FGIDPSTSTGFFVTFVSVQAGFLLVGGFGLYAGDMAFFTPISQISTLKEILRCKFKEINEAM  
QGDELSRPSNISELLKDAVQFHQRYLRNDGFIEACYEDILWYKLTASDRKSLLIMLILC  
QNTSGLTIGSVLPLSMNTGLRVTKTIYSIAMMLIRFLDKED

>MdOr64CTE

MTHRQSDRFKAIVRITKICADICGANVLEHDYRINVRTVLVVFVIIIITFVFMSYTIYDGFF  
VQGDWKIILQVLSIGAGTLVQGFVKLLNCIQQQENFRFLIGELYDIYEEYELKHTGYQR  
HLNKGIIHLLSYIMKLCAFIALLVIGMAAVTVVRSLVFDVNQVIVQCLIPGVDHTTPRG  
FFLTICIVQISFIAVGGFGFYAGDMAFFTPITQIVTFQGILRCKMFDLNEVLEKDGEENVK  
KSTEMLKEVIKFHQRVMVFLTVTQDITYFLVILVQIATYSTGIICTIFCVLLGAWPGGYV  
YMIYCFVMMYVYCGVGTLVEVT

>MdOr65

MAKTLVQRYETIVRLIRIFSGICGANIFNPAFKKNIITWIVIIIFIYQYFVFTGYTLYVKIYID  
KDRPSVLQVLCYLGSQVQGYCKLLNFWNKDDIRYLIYELRDIYEKYDLKHADYRCC

LEKNTNRVNRFIKFMATMHLVITITLIAVVPFYRVVFNERILIMQFLFPGVDPNTAYGYT  
IITTIHCICILFGSFGNFAADVCFNIVSHVPLFRDLLRCKCQDLNEILEEERASEEEGFAEI  
ELLLKDIFQWHQKYMRYITTVKENYFWVVLVEMGTVALSIASSTLFCLILGKWPGGLTY  
LTYCFIMLYMYCDLGTIVEITNDGFIDSCYTEIIWYRLSIHQRKMLQMMLMMTQNTG  
LTIGSVIPLTVNTGLQLTKSLYTMTMMLINFLE

>MdOr66

MNTHYRLQDFMVYPNIAFNLAMVQPFRLSGTLEEHQTANRCRGFMKSMLIKLWFVFG  
AVNLIYQNVGMLAYLLLPQLSEIFDDVEMVAKISETGGILGLTMVAVCKMFVLFWHG  
RRISILLQELEEIPDEKEQFAHPTLYRVRHFAQTSERLMGRRTTKFFIFAFCFYNSLP  
LYELLPLDQEIKYRYQSNTWYPWQTKDNARTWLNFIASYVCQVQSSLTGVGFIMAGE  
FMLCFFITQMOMHFDYLTNALRHLDAASVRANEKLYLIYHTKLLRYSKEINEIFNISF  
LVNFISSIAICMMACSMVMLSMAHTFKYSVGLLSFLVFTFFICYNGGEFTDASDAIMP  
SAFYNNWYEGDASYRRMILFFILRSCEPNVLTAYKFTTVSMPTFMAILKVSQYQLFTFLQ  
AMD

>MdOr67

MLYRPRLPDGRKVPLSWPIALFRLTNNICWPLEENASWLA VVFDRCWYLA FILFVITN  
DAEFYLRVNINNLDMLTGVPTYLV LIEIHLRAFTLGWRKQDFRRLLEKFYRQIYIESS  
LHPTIFKNIRSQLMPIFVLSSLYLSALISYVILPIYFLSIGSRELMYKMIPAFDYSPLWYLL  
CCLSNLWIGVIVATMMLGEATVLS TLVFHLNGRYLMMREKLMAKVDVVLEKKKRDN  
GNQHIAAEYNKILVETLQENVALNTFAQEIQREYSFRLFVIVAFMAASLCGLGFKVYTS  
PMTSIGYIFWAIGKIQEILAIGTMGSTIVTITNQISSMYYESNWELVVFQSEDSKSNARL  
MKLVQLAIATNSKPFCLTGLNFFTISTTTALAILQGAGSYFTCLTSFR

>MdOr68CTE

MLYTPRLVNGRPVALTWPMTFYRRFNIICWPLEDNAVWWTHIFTTVIYMVSFLIFVMH  
NDAEVRYLRVNFHNLDDMLTGIPTYFVLIEIHIRAFSAFEKKSFKWMLRKFYAEIFIEE  
SLRPDIHAGNLRSYYPVLA FSILYLCALLSYIVFTIYGLAVGEKPLPYKMIPFDYNSWYI  
YTPLVLSSLWVGFI VASTIVGESYALTMFVHNLDGRYQMMGERLNMGVENILKFSSND  
SEAIEKFHRILIALTKENIRLNKFAQEIQREFSFRIFIIFSFLAATLCVLVFKVYTS  
PVNSIPYVFWTIGKVQETIAFGQIGTTIISR

>MdOr69

MEFHRPLLNGEIAPLSWEIRLFFVNVSWPMKANAKLFTRIYDKATLVLGFLFFCYQNE  
AEMHYVVNNINDIGLALEGMATYLLVETHLRIYNKGLYKSSFREFLNEFYAKIYMEKS  
YNIETYLDIQRKLLPTKMCSYAYMLTLV TYFLVPVLGFFSNAHLVPFKTIFHYDLDIWY  
FYLPTLCLTLWIGVAVVSQLAAESNLLATIHLNARYLHLQSDLKELQTRLASDMKLS  
TDKVLGEYRREFIEIVKRNV EYNDFAQKFQNQYSFCIFVMMAFSAVLLCVLAFKAATL  
GMTTKNITFITWIIGKIVELLVFGTLGSQLIETTDKMSSCYMANWEDIILKSPKTTDNE  
LMKLIILSIELNQKPFSLTG YNYFSVSLATVV TILQGAGSYFTFLYAFR

>MdOr70

MLDLFAKQRQCCLLMGHNFVRDKSELLKKWHNIKYVSVLLLVSQAQWPIMNYTIYYI  
DDLQLATASMSISYTNVLT VVKITTF LFYKWRFAALMEKLES MYHELQEEESKAILKTS  
NRYAAILVNIYGNSVGLTGLYFMVAPILKIVWSKIRNTELQLELPMPMRPFDFESSPGY

EVCYIYTGLVTL SVM TYAIAIDGLFISFTINLVGHLKTLQHFIQSKSFEQNDEDVHKQISF  
YIRYHNLIHLHYQEVRQIYSPIVFGQLITSLQVCVIVYQMVTHINTFLVFVINCTFLLSIL  
LQLFIYSYGGEILKNESLMVGVSVQLSNWYNLKPRHRRMLWLLMLRSQRGAIRGGFY  
EASLANFMTILKAALS YITLIQSIE

>MdOr71CTE

MKKAATFDDFFKLASFFYRTIGIEPYDEPGVEVKKSFAENFIFYSGVINLNYVLIMEI  
VYVAVAFIRGENILEAIMCLSYIGFVIVGESKMFFVFRKKPILSKFVKRLVEIFPQEFELQ  
KTYNLSYLQRQSSRV TIGFALLYMILIWTYNLYAMTQYLLYEKWLGSRVVGGQQLPYT  
YAWWDWHDHWTYYLLYFIHAFAGYTSATGQIASDIMLCGFATQIIMHFHYISHVLTN  
YKVKVDEAKDKQAGRSQDITFLKDIIEYHNCLLELSEQLNSVFSPLLLNFSASSFVICF  
VGFQMTIGVEPDALIKLF

>MdOr72PSE

MSNAKFDDFFKLSRFFYTTIGVEPYNEPGVEVKKSFAANLIFYSGVINLNYLLSME  
MVYVAVAFVRGENILEAIMCLSYIGFVIVGESKMFFVFLKKPILSEFVKRLVSIFPQEVK  
LQKSCNLASYLRQYSRV TIFFALLYMILIWTYNLYAITQYVLYEKLLKSRVVGQQLPYY  
TYNWWDWQGHWSYLLYFMHAFAGYTSAAAGQIASDILLCGFVTQIIMHFHNISHVLT  
NYKVKIDQAKNRQVGLSKDMAFLKDIIVYHKCLLDLSEQLASVFSPLLLNFSASSFVI  
CFVGFQMTIGVEPDTLVKLFLFLFSSTAQVYLICHYSQMZMDASLNVADAVYNQNWISI  
ADVRYQKMLILMAERAQKPVQLRATTLVLSRGTMTLMQLSYKFFALLRTMYVKK

>MdOr73

MADHSNKVYFPRILDYVYFQTFLQLLTLLPWKMSKLISFEDFLSYANALNATIGLVAY  
EKPNTKPLKKLIFDVIFWLNFINLNLVLLGELVFVIESVNGRHEFLEMIMALSYIGFVAL  
GSFKTCIIMQKKSHLT TYARDMNQIFPNASIAVQRELNVRKYLKYSKFFSIMFSTMCLA  
MLVFFNFEAITEWLIATELRGDQNAAQHLPYFMYAPWDWTGNHWSYLLYGIQCWA  
GHTSVVAQFSSDLLLYAFIGQLIMHFEAITKDVSNYRLRSCTADMDFLRNIVFKHSILLE  
LSERINDLFLGLSLFVN FATSAVVMCFLGFQMSIGASFVNLLKLVLFLILMLTQGF LICHF  
GQLLTDASLSIAYAAFNQNWISSDVCCQKMLILITERAQKPVILKATTLPVPSRATMTQ  
LLQISYKFFALLRTMYVQ

>MdOr74CTE

MLDSNKLLPFNSFEIFEFLSKTTFTVQSSLRSENVTFSLIMFKLKNFEDFFIYVDFIYATFG  
IESWARRELRAPCKKYLKTIIFYINITNMNVVMLAEILNLFSTTDTVDIPDLLMSMSYI  
GFVINSSWKIYMIWKKRPLIESLICDFHDIFPTKLMLQQDYDVQVYLRKCHRKSKFMSL  
LFVFAIWFFNLLAILEFGISSRNHFHRSQQELPYFMYIPWNWQNHWSYLLYVMASM  
AGHTTAMGNVSNDMLLYSLISQLIMHFDFVANTMESYEIGSGSKGVAKMEGRENGND  
LEFLKIFIEYHSHLLGLSDRLNDIFGLLLFVHFASATFVICLLGFLMTIGTSFLSLFKLSLF  
LFTMLIQSAEICSYGQMLMDSSLRVSTAVLT TQWLKTEVRCQKMLILMSKRAQRPAQL  
KATYFIWISQGMNE

>MdOr75

MRQHQQRKSSKRNNQNIKTMAPKATSNGIGLNKFLQADILAKSIGLIPYDEENDKRSVR  
YEKLMKFIFILNMVNMNFVLFSEIMYVLLAMKNGNMFVEATMNL SYIGFVFGDIKIS  
VLRKKPVL TILMKEIEDIYPKDGRAQKAYQVREYVWRFNLISLGFVIVHEILIWFYNLYI



AVSYLIYEWLQWRVVPRTLPPYFWVPWQWQGHWSYVLYVSQNFAGHTCMSGQ  
LANDLLLCVAATQIIMHFEFLAKRLREYRPTGRHVDDLKFLREHIKYHQAVIHLSALM  
NEVFGVSLLVNFISSSFVMCFLGFQMTIGVEADTLVMLFMFLFCSLVQILMICNYGQQL  
IHKSEEIGHAVYSQEWLNSDLRYRKMLIGIARSQKPVILRATTFLNVSRSTMTLMQLS  
YKFFALLRTMYSK

>MdOr76

MEPIEARRDLFQFVRRTMYWAAMYPLHLDRRLPHYICGLGLFVECCFEMFLYLVSIIQIA  
ILYVCTIYLNYSGLLELLVNCMIQTIIVWTIVMKVYFRRVRPHHLEGMVDTINAEYR  
TRSAIGFTYVTMDQCLDMSNRWIKTYVYCCFIGTVFWLLPIAYGDRSLPLACWYPLD  
YKEPVIYETIYFLQSVGQIQVAAAFSASSGFHMLAILISGQYDSLFCSLKNILATVAIR  
MHSTKEELRKLYELQESTDSELNEFYCSEEITCDINMLVHINASPKQALMSSQEFYHF  
RHAFaecVHHHWYILDSLKSMEKFYSPIWFFKTGEVILLMCLVAFVSVKSTTANSSFM  
KVVSLGQYLMLVAWELLIICYFGEIIFINSQRCGDAILRSPWYLQREMKNDFLLFLN  
SYRPFKLTAGKMYPLNVERFRGVITTAFSFLTLLQKMDERV

>MdOr77

MSIAIRPQLLNRMHKRHHVRDNIIRIESRDKRHDLFQFIRRTMYWAAMPMSLEHLLP  
QRIRYLSSFIEVFYELFLHLVCIHIVILYLCIFYLNNSGDLELLVNCMMQTIIVWVIG  
MKLYFRRMNPRPLEELMKTMLNLYRTHSIKGFTYVTMEECLIMANKWIKTYVYSCFA  
GAVFWLIIPITYDDRSLPLSCWYPVDYKKPIIYEIIYFLQAVAQIQVAAAFSASSGLHMT  
LSILLSGQYDVLFCSLKNILANVALRMQSTEQQLRKLYKLHEITSHDTNEFYCSKEKTL  
DVERLFDAQQLFVETSQDFRHNFRNVFKECIVHHWFILDCLKSMERFYNPIWFLKTGQ  
AILLLCLVAFVSVKSTTTNSSFLKNLSLGQYLFLVAWEFLVICYFGEMIFYNSQRCGEAI  
LKSPWYLCMREIKSDLLLFLRSYRPFKLTAGRMFALNIDWYRWVITTAFSFLTLLQN  
MDQRDVNVST

>MdOr78

MTPSVYFYGGDSVDLYSEHDSGREDDVFKLQLLFMKFMGQVPMQLERRLPLGWKNV  
AGMFAKSYCFCVISNLHLAILYVKTTLDMLHNGELEETDALTMIIYSFSTFATCYWL  
FNAEALNSFIGDINANYRHHSMAGLTFVSAEHSIRLAYKVTLYWLIACCVGVVCWALA  
PLLLRSHTLPLRCWYPFDALKPVVYEVVYATQLWCQILMGCIFGNGSALFVSVVLIML  
GQFDVLYCSLKNVDYNAQLLAGGDLITLRNLQRDLPRPADDELNQYALLEEHLTDLT  
ALRVSKPNSRPSLKEALHSSLVECVLLHQFILKSCNTLEGLFNPYCLIKSLQITLQLCLLA  
FVGVAGERSTMRITNLVQYLALTLSSELLMFTYCGELLSSHISIRVGEAFWRSGWWLNG  
NLIKRDIFIFLANSKRVVVVTAGKFYRMDVQRLRSVITQAFSFLTLLQKLAENQ

>MdOr79

MEKHRLYTLDEFLLKLQPSQRYTRIIYLDFFRRENQNKPFRESLRLLYAALTLLIVDCAC  
NVLKIIFEIRAQRLSEAKQIGAVWSIAFLCLIRGIFVMFKHKSMLDLTNDLDKIFPRTRLL  
QNRMNCHKLARYLLIRHRFLFAYAVVGLSAFIGIPLLKYIVFYDPNSGEPLLEDEYHQHA  
SWFPFHLKENPTTYPYMYVSETILTLFGINCLFTWDHIYTVTVAQFIMHFEYVNTELAR  
LNAKDTMDVEKSKKFYDDLVEIIKYHQHVLRLGNKLRNTFNLPLFLTDLISGASICFHI  
YLIANTDDVIAITLFIFFCFVQVAFAFDNCYQGSRIENVTTNMSQVIFEQNWYDATLEY  
RKFFVHFLFFASRPFTLCGYNLFSIDMVHFRGTMMIAYRMFTFLQARGSKVE

>MdOr80

MNVRLHNDGGYDRTYAVRGILRVMKILGLWKWQTEADKETPRHILWLQYVQRLVCH  
GPFTFVFITLMWIEALRANGLDEMGDVLYMSLTEAALIVKILNIWQHSTKASTFLHALR  
HNAHFALHSGDEVTFWRNAQKKFRYIIMYSAGSVFTVISAFAGVLFVTEPQMAFAY  
WVPFEWQSNRRNYWLAYLYDFVSMVCTAGSNVCLDMMGCYMMFHVSLLYKVLSFR  
LQKLRAVKGEDVNEKFKKLILMHKSIRRM TRECEILSSKYVLSQIILSALILCFCCYRIK  
LDIVANFGQFLSMLQFLAVMIFEIFLPCYFGNEITLNSSEIMLDVYRTDWLEYSVANRK  
LIILFREFLKRPDKVTIGGYFEVGLPIFTKVVNNAYSFFALLMNVEK

>MdOr81

MQLRQPKDVGQQLNSVYGLKYLWWNFSIIGHPPAGVRTHPVWRFLYLVYAVVINFL  
AGFCLPATMLANLMLLKSLEEIIGNLSLSMTIAISMTKELAILYCRGGLKANHYLRLL  
DERCSAHPDRMTVM EAVRMCHWYYTVYISFYGFCAIGFAYIGWSNHTLVYSAWFP  
NIFANDQTNylaayIFQNLAQTFTVFQNGNNDMYPLCYITLMIYHVRALADRIQRVGG  
DAETSAEENVQELRNCIQDHKNVQSYFECIQPAISSTMFQQLWVAAFTLCLTAINLMAF  
ERTFAEKIFS VVYLG VIVIQIFPACLCVNFMMSSETS NLTTAMYKCNWIEQNRNFRMLII  
FMQRSQKVNVYIYAGGLAPVTLQTFVAIIKFSFSMYTILSQMKIQ

>MdOr82

MVLETDNSLILFDFIRLPLKFYSAVGIKIFQWDADDIMTTKEKCIFLLLGINFIGCFLAKS  
LFCVFGFVDTMQATQWILYFMFAMNGCCKTISVAIGRKKLYTVLKDIEGIFPATLKER  
QEFRLAHNYGYIMRHAKIMSIQHCSIAIMFIAFPLVQSTIEYLTSADSEFVTRTPYIMVYP  
FDATEGIGYVVG YFSQFLGGFTVSCYFVGSDMLLMCTIYLVIMQYDYICYRIENFKSRN  
YEEDMKELKIVLERHNLNIVAETVNEVFSISILLNYMISILIIVMISIQITKGSEFGLDMIK  
FVGFFTSASTQVYYICMFGNLLMDYSSRVSESIGQEWYWTDVRYQRMLVLAIARSQR  
PSHLTAFKFFTISMESYGNLMTTAYQFFTLRTTYNNN

>MdOr83

MYYNHPLFSFNVKMWKYLGFIEFKRINQALLILIIPCLINMCQVMNIAYNWNDMSVIAI  
GLFMTAILFNALVRITTVMRNQSKFIEFFEMIEQWYREIEMGPDDGAWDLLKHIPRTR  
LISILSFSFAAGAAVASATIPLFLEQRS LPYDMYIPFYDHLKSPMYEILYFMQGFISMPFC  
VLTYPFTNLFI AWLTFGISLLQILRYKLES LPHENDEEMLKQLIELIRFHHRIMNFGQTL  
ESLVSFVCLVELVLFTLMLCVLLASFLVMDNVMSKIATCIYIFCILYALFIPYWHANEFS  
WESTKIADAAYNIKWTRSNIKIRKCIAMLILRSQTPLKIKAGGIFPMTLEAFQALLNTTY  
TYFTMFKGMMGKEPNVHDRGQ

>MdOr84A

MTMEPRNFSKYLQITITLNQSIIVLQLMYNFTTQDEDVDVL TNMIYFNYIFVGLGKLL  
CMYYRRQTLAKVLET LQEIYPTQHIEEKYNLNKHFYYSRIEKFIWSFYRLVGPVYVAL  
PLLQSLKNIWTLGKFTLLLPLCLWKMGDPMDSNWWLT YLFYYLIGGSSSIFSGLTITGC  
DLCLYSLITQMCMHYDLLSQRILELQPASGEEIASKKLRGLTQQHWMITNVANEINIFS  
VMSSSFTLCLVAYQMLDDVSIFTIVKAFILLYESKQVIITCYIGQKLKECSSLVNASLY  
AHSWYDGSTRYRRRVLYMLLCTMQPFVLNFMGIADITVITLKEVYGNAYRLFTVFKS  
A

>MdOr84B

MTIEPRKFSKYLKITITLNQILIIVLQIIYNLTQDEGVDVLTNIIYINYNVVALGKLLSMY  
YRRQTLAKVLEILDGIYPTQRIEEKYNLSYFRYYSRIFTIWSFYRLVGPVYVTLPLVQ  
SLKSIWTLGKFTLILPLSLWKMGDPLDNDWWLTLYFYLLIGAFSSISSGMTITGCDLCL  
YSLITQLCMHYDLLSQRIMELQPAAGEENATKRLGILTRQHLIVTNVANEINIFSVMSSS  
FTLCLVAYQMLDDVSIFTIVKAFILLLYESKQVIITCYIGQKLKECSSLVNASLYAHSWY  
DGSTRYRRRVLYMLLCTMQPFVLNFMGIADITVITLKEVYGNAYRLFTVFKSA

## 81 CcapGr proteins in FASTA format

>CcapGr45

MRMHKQSYQSLYNYIAISLGMTSFRYNSITEKHKQMFWTQTTFALLANVCTIISLPLIFW  
NAIKRVTAIQWNPTMTYATYITCTIRALVVLTYILKRSKIDLMIHESLKKLRTLQRTYLA  
HFPRVSSIERRMNRLYYAKFISMCLLLTLTAIIYHQAAEFETWLNCFIFAMSLQTTAIL  
YITRFNFFWFLWSICSSRLRYVEWQLRHLLAQMNESLELGQRWRLTALATDLEGILEAH  
WQLAKLVILVSLNYSVLILAGLMEQITSLVRQLYYGIIFTFDHRATNSQLTIGTLYFCALI  
FDMYLSLLVCDMTVKTYASVEQLRCTNMVNAYCPALYEMLERWLLRIQCDGFRVD  
LYRTITIDKKRWLRIMQRVVFYESITLVQFHL IFNANKNFN

>CcapGr46

MSSQLRKINKLSSLAEVCFTHAAIMASVKSIVFYSWYYFLLATGATSLCFDFRRKRLTS  
TMRRRYYNLLPNALLGFLPFYLYKGMRYLDAYEEKPLLVTVGHFNIQMQTIVIAATC  
VIHYWHDVEYYIAKKLLYLERRYFERWSVYAQGPSYNWVFWLKFLSYFSQNASLIM  
GPMTVSAKRLLNWMDLLNLYFFIWNLYLNSIIFSCLTLLHIVRRYDVLMHLLRREL  
QRLSAEVLGMRSELSELRRVLQVHEELQVIVTGVNNLYRFHITWLLLIFFLSNTTAFYV  
GFIHLANWSTCHMFVLLSFSLFAIKSFDIFLINHCCELLVARHSVALELLRRQQHHLHA  
ATEYEQMCVNFGLNATVTKFSINIYGMFMDMDHRST FVALRNTVMHALILIQFD

>CcapGr1.1

MAYWAIATRKGESPPMKITPVLNPNQREFLEDELLYREKLEILAENNTLSTDLFVRKFE  
DIDDPVLLDKHDSFYHTTKSLLVLFQIMGVMPIHRNPQKYGLARTGYSWTSKQVFWAI  
FVFTVQTTVVVLVLRERVNNFLNDTDRRFDEAIYNVIFISLLFTNLLPVASWRHGPQV  
AIFKNMWTNYQLKFLKVTGAPIVFPNLYPLTWGLCFFSWGVSAINLSQYYLQPDFKL  
WYTFAYYPIIAMLNGFCSLWYINCTAFGTASRALSASLELTLMSDKPAKKLTEYRHLW  
VDLSHMMQQLGRAYSNMYGMYCLVVFFTTIATYGSLSLSEIMDHGATYKEVGLFVIVF  
YCMSLLYIICNEAHYATQSVGLDFQTKLLNVDLTAVDSATQKEVEMFLMAITKNPPIM  
NLDGYANINRELITSNVSFMATYLVVLLQFKITEQRNYSLKAAAML

>CcapGr1.2FIX

MSFWVKNADGSTDVEKPPRIMPLFNPGQREFLEDEQRLRLQMEKRALDGGKSFDYYI  
RKQSTLDDARLLDEHDSFYKTTKSLLVLFQIMGIMPIHRNPPDNNLPRTGYSWKSQV  
LYAICVFSLETYIVVMVLRARVKSFIQPDKFHDEAIYNIIFISLLFTHFLLPVASWRHGP  
QVAIFKNMWTNYQYKFWRVGTSPVFPNLYALTWGLCIFSWTLSIAVNVVSQYILQPDF  
EFWYTFAYYPIIAMLNCFCSLWYINCNAFGTVSEALAKNLELTLSGKPAEKLTEYRY  
LWVDLSLMMQQLGRAYSNMYGMYCLVVFFTTLTAAYGSISEIMDHGATYKEIGLFVI

MFYCMSLLYIICNEAHHASRKVGLDFQTKLLNVNLIALDTASQREVQMFLVAISKTPPI  
MNLDDYANINRELFSSNITFMATYLVVLLQFKITEQRGSHSNQDEFNTM

>CcapGr2

MKTGNLLFSEMLSKSPKSSGEKLSDHAMIGQLNSFLFDATDNDADVERHDQFYRDHKL  
LLVLFRLVLAVMPIQRTSPGRMSFSWKSSATIYAFIFWTVMTHIVLIVARERIQILYTTKQF  
DEYIYAVIFVIYLIPHFVIPFVGWGVAGEVAEYKTSWGKFQLIFYRVTGTSLQFPHLKST  
ILAISVGCLLCAFLFLFALSFFLEGYPLWHTLAYYHIIIMINMNCALWYINSRAIKTASTA  
LTTCFQKEIVSSHSSDVISKYRMLWLYLSELLQSLGNAYARTYSTYSIFMFINIVIAVYG  
AFAEIVDNTDVSRSYKEVGLIVDGLYCSTLLFIFCDCSHNATLGVAKGQKVLLEIDV  
RHIDRKAKSEIDLFIFATEMNPPIVSL KGYVNVNRELITSFIATITVYLLVLIQFKFTLN

>CcapGr3

MFNSYNRRKKHDTVFLNVKPTFNSQGGGLRKYSTGLLDKEENPFYNPNSSNASRASV  
GTITTLNENFRANIFYNNIAPIQWFLHLIGVLPITRRESGKAKFRINSVAFGYSLVFFILLS  
VFVAYVAKNRISIVTSLSGPFEEAVIAYLFLVNILPLLLIPLLWWEARKIARLWNDWDD  
FEILYYQISGHSIPLNIRRKTTLIAVILPLLSILSVVITHITMADFQLQVIPYCILDNLAM  
LGAWWFIICECLSMATANILGERFQRALRHIGPAAMVADYRALWLRLSKLTRDTGNAT  
CYTFTFINLYLFFVITLSVYGLMSQLSEGFGIKDIGLAITAIWNVFLLFYICDKAHYASFN  
VRTNFQKKLLMVELNWMNSDAQTEINMFIRATEMNPSPNINCGGFFDVNRNLFKGLLT  
TMVTYLVVLLQFQISIPTDSGKRMNISVAEFANDMLMDSAEIESTTTTTTTTTTPKPKIT  
TTKAPARGRKGRN

>CcapGr4

MSKCHICTFLLQHIRVSFTILIYYITSIVLIIGQCFSLMPVLGVGQPNPRHVRFSRLTRVRM  
ITLLFLVASTTLNLSMIKHLAKIGVNAKNLVGVMFFTCVQSSSILFFSLAPRWPRLRFW  
TRTEMIFIRKPYERPQPDLSRVRRAAATILSLSAVEHLLYLASAVVSQYRRAHLCAAL  
NNTTVHFTFEDYTYKNYDYVYELLPNTTAVGSFILVGNFFCTFVWNYMDLFIMMVGK  
GIAIRFEQIKMRINKLLDKEVPESIFMEIRDHYIKLIELLEYVDEDLSGIILLSCANNLYF  
VCYQLLNIFNKLWPINYYVFWFSLFLIGRTAFVFLTAAGINDEAKEALGVLRRVSDK  
TWCVEVERLIFQMATTVALSGKKFYF  
LTRRLFLGMAGTIVTYELVLLQFDEPNRSKGLPELCS

>CcapGr5

MSQKLQEQTFFHKSFSKVLLVSQFFGLLPVSNILADDVQRIHYSWCSVHVSYSVSGAIIILNT  
LEFCTVGYLVFKAGLNFHNSGTISLFFVCMLEHFYFWRLATQWPLIMRVWRRTEEIFL  
RPPYHIYATYNMKARIYALSSVMCSAMVEHSFLVFNSFHHSNIERVQCKYNVSFWES  
LYRRERPHLSRVLPFSIWYLPVVEWINLTAYPRSFTDAFIICVSVGLAARFHQLYLRIE  
AVHNKALPTLFWTEVREHFLELLHLMRLNDNIAPLILLACSNMNYFICFQLFNSFQNI  
GVDLIAVFAFWYSLFFAIFRTILTLLTASSVNDYSKILLTLRSVPSTSWCTEAQRFSEQL  
AFDLTAWSGCGFFFITRQLILAMASTIFT  
YEVMTVDVINKGAIQQVTNYCRPIEYDTQSE

>CcapGr6

MAKVNPVEVSHEAARQNTNTMHLALRPFMMISQLLATLPVTGTWQKSSLEHVHFSW  
CTLIAFLSLIMITFSIIDVVLSTKVVELGLKLYTVGPFSFSIISALSVSSFLQLARKWPNI  
KHMDRCEQIFLQKSYGNKESRNFNVRKFGGILLFGAALEHSVYIMTAIFNNDFAQIKQ  
CNLTVDLFWKNYYMRERLQIFSIFNYHAWLIPLVQWITISTTFAWNYVDIFLSMIFKCFI  
RFRQMHWRIKRHNKHMEDDFWHEVRNHFVVLVELLHLFDDGLSTLILVSCCNLY  
FICVQIFHSFNNDRTFIKEFYFWSSLLFVLLRILTMMLSASAVHDEANKIMSTMYEIPK  
FWCLELKRNLNEIIVHDLVAFSGKSFFFLTRR  
LIFAMAGTIVVYELVLIDQVEDKDVVTDFTSRNV

>CcapGr7

MDLKSWSAKPKIGKRHKRIIGVLRVRNIFRRGTKLDYQHSGSFLEAIGPVLLLAQIFAL  
MPVCGILSKSASELYFSWKCVRTLYAMIIIFCLGPASLCTIAFAFRESFNFDITIEAIVFYVS  
IFLIALAFFQLARKWPALMVTWESIESKLPLKTEMQKAALAHRIKMITLVATVCSVVE  
HLLSMLGIIYYVNGCPAMPGHPHQSFNWAQFFYFFEYTNLAGIFGKVLNVISTFAW  
NFNDIFVMAVSVALSARFRQLNEHMLRAAKQPTTEKFWMDNRINYRNMCKLCEATD  
DTISIITLLCFSNNLFFICGKILKSLQKKPSWSHVSFVFWFSLGFLMRTLMLSLSFAEIND  
ESKRPLVVFERSVPSKFWCPELKRFSSEVTTDVVALSGMKFFHLTRGLVLSVAGSIVTYE  
LVLLQFNKEDKVNDCEYEG

>CcapGr8

MKFTFKVGWKIGNVTADARLRVQQQQRKLKTQRLWRRQCERAKRRQSQSGSNAATK  
CDPLAQKHNRKALLIKRTCTEFIAQMELLHRQIDKPPTLPKSKSNFQQDGSFHQAVG  
KVLLFAEFFAFMPIKGVTTAHPRQLSFSKLNLRWTWYCLIFIMTTTIDLGLTMKVLPPIN  
FNGVEPLIFRLSIIVVCSSAIVLARKWPALMLDWHEVECDLPEYLTQIEKGRLAYKLM  
VTVMAMALSLAEHLLNITSHIYSNSCPQTNDRIIDFFILTNQHLFELFPYSIYLALWGKT  
QNILCTFIWNYMDVFMIVSIGLAAKFKQLNDNLYKFKGMRMPPEYWSTRKQYRNL  
CELCTRVDAASVITMISFSNNLYFICVQLLRSLNKMPSFAQAAYFYFSFFFLIGRTLAVS  
LYSASINDESRKPLRLRCVPKESWCTEVKRFSSEDISSDLVALSGMKFFYLTRKLVLSVA  
GTIVTYELVLIQFHQESKLIENSPFMRNGSRH

>CcapGr9

MERGKKSNSIIGHIKCVLGMKISLEAGYKRNQPQNQRKQPQITMDIQPSTHKQPTKQTR  
SKLGEKLEKGQEGNDSLRRKGKTLLVIAQWFGMPVRGILAVNAKSLKFSWVSFRTC  
YSVVYVILTSIDTGLTMNMVFRGAMEVKNISPLVFHVNALLVAICFLRLAVKWPQLML  
KWQRVERQLPPHQMWDRREALALRVHKVTFVLITLALTEHLLSVVSGIHFHAIHCS  
NSDPIKSFFIAVSPHTFLIFNYSTWLAWCGKLLNVLNTFGWSYMDVFLMIIGLGLSSLFG  
QVQNSLNRVKGKTMPPEAYWTRTRLQYRLICDLIEQVDGVVSGMIMLSFANNLFFVCIQ  
LLKSINKMPSISHFVYFYASLSFLLGRTLAVSLYLSEVNDRSREPLLVKHPVREGYCAEI  
ERFGHEVAVDKVALTGMQFFNITRGLVLAVAGTIVTYELVLIQFHEENLWACN

>CcapGr10

MFFLTCLTRKGGIRRATRADIHNGGFHEAVGPLLVIAQCFCLMPVRGILAATPKGLSF  
RWKCFRTWYCILYILATIADTGLTNMVMVKGVLDRNIEPLIFHANILLASFGFLRLAAK  
WPQLMRKWQRVERQLPPHQSWDRREALSVRVHKVTFVLITLSLTHELLSTISAIHFAN

YCPSRVDPIESYFMTVVSQIFFVFDYSTWLAWYGKILNVLNTFGWSYMDVFLMVIGVG  
LTSLEQVQKSLECVKGKVMPESYWTRTRLQYRLICELIEQVDASISGLTMLSFANNLY  
FVCIQLLKSMNTMPSVAHFVYFYASLSFLLGRTLAVSLYLSEVNDRSRKPLRIIKLIPKE  
GYHPEVDRLCHEIGMDTIALTGLQYFNVTRG  
LVLTVAGTIVTYELVLIQFHEDQNLWNCN

>CcapGr11

MEITEPTLCIFYVNKFLALAPFSVRRNSKGALDIRRSVMFSVYSGSLCLIMVILTYQGLL  
FDANSQVPVRMKSATSKVVTALDVSVVVLACSAGVGCGLLGYRATRELNTRLRKIDD  
SMHSYSNFKRDRTMAILMMVLPLTAITSILGLDLSTWLRFAIGVRTPQDDTELNVQWYI  
PFYSLYFILTGLQINFANTAFLGLRRFRRLNVMMLRNSFLKDNNQKDAPMKPPITTVKV  
SQHPLALHQSLAKLTNDTLQSGSKQKVNLLRLLEENHESLGKCMRLVSNHGFVLF  
LVSCLLHLVATSIFYLFLELLSKKDSGMVWLQVLWIFHALRLILVVEPCHLATVESKKT  
IQIVCEIERKIHDPILAEVKKFWQQLLVVDVEFSASGLCRVNRLLTSFSSAICTYLVILI  
QFQNTNG

>CcapGr12

MEISATSRSVFYLSKILGLAPYALKTNTKAEIEIKRSKFFTTYISIFLCLSNVVLHIMGFKEN  
CKVSLRANSTTLRVTTAIDVHIVALASIAGVGCIFGLRSTQELNQRLRKFDSSLQFLTIT  
KNERNKCGAMAITLAVCVGLLCFDYFVWMEILSGQHTAEQKLKATLMWYLPFYGRY  
IAVIGSQILYANNVLGFGRRFQRLNKILEYEFLSEQMSSSHHQLTAELRCKAKTRAHILR  
TLAYNYGSVSKCVEIFSNTFGFSTLCDLTSCLHLVITAHFILLALKNPIDRYLINLTMW  
MVLHIFRLLLVEVCQMATVQSEKTKQLVAEIKRTHYERLLAVELKSFLHQLLVFEVK  
FSVMGIFKINRNILTAFGSAIATYLI MVTQFQNSSG

>CcapGr13

MEINEKSLSVFYLSKFLALAPFSLKRNNKKGDIKRSKFFTLYSIGLCMAIVIVMISGFGE  
YYFGPLRKHSAPIKMVFALDVSIIILTLVGGVGCSLLSIQPTQELNSRLRKIDAFLLHLP  
HKNDHTITKILATLPIILTIGIFGFSCFLWMRYPLADFDKHKILKTAYHFYVPFYVSHLVII  
ALHIMYGNNALVLGGRFHRLNTILKYNFLYGKRYIPITWLINPGLGGGNELNTDAICKV  
SLLRSLADIYLSLSKCVDIFSNTFGFGLICGFTLCLVQLVLTSHFLILFITRAIWDIFRIFEL  
MW  
LSIHLFRLLIIVESCHKAAAESKKAVEIIEIRRKFDDDPSLVEELEKFSHVLRAFEVKFSV  
MAAFNIDRNILTTICS AIATYLMVTQFKNFKN

>CcapGr14

MEINESSFSVFYLSKFLALAPISMRRNIKGKIELKRSILFSLYSITLGVAMVILCYEGIIIFD  
ANSQVPLRMNSSTSKVVTALDVTVVVLSSVAGIGCGIFSVKPSRELNLRLRKFDAALHS  
FSNFKRDRIISRVMVIPMAISAALIGFDIWTWILHVELAPKKGSKNSIKWYLPFYLLYF  
VMIGFHILFANTAFLGLRRFRRLNGMLKCHYLADSKPYSTIKTQLNSIKITPDHMSLH  
ESLNTESLPKESLGKTRVMLLRSLADNHESLGKCVQIFSNTYGI AVLCLVSCLLHMVA  
TSYFLFLALLDRNVTGYVWGQVMWILHILRLLLVEPCHIATAESKKTIVTEIERKI  
HEPILAEIEKFWQQLLVVDVEFSAL GLCRINRNILTSFSSAIATYLVILIQFQKASG

>CcapGr15

MKHSNWFNNMPYNKVKPSAVQLHNIGYNEFLHRIPRDSSIFNDIRTTLFIKATGLLPF  
YEEISSYEVGPPTKPNIFYSYFIRGVVQALTIFNLVNLVTSGSAQLFSSYSDDNVNKWI  
EFLCMLAHSTTVIICARNSKSFLKIINEILKVEDDVDFRFGTAFENRCGFSLKFIYGICIC  
QWYLIILTVLEASDSLNMNSYIFIMFNAVQNGMAAIFIVFAAALLRIVKVRFAHLNSILS  
GYTYNQQRKQEHFAGRERFRNTMETFPEESLFTYRMHNKLLRIYRSINDCCSLILVAY  
MGYAFYTITTTTYKLFVQITTQSY PQISYNVLQTCFTWLAMHTCVLALLSRSCGQVTD  
EANWTSQVLARVYSKSKDHQNIQVDFLTKSVKQEVQFTAYGFFVIDNSTLKFISAVTT  
YLVILIQFKQLED SKTVNNSM

>CcapGr16

MHNCKRRTHPKSNTRDNEEQRAIHPTLAYLKWHL LLLKLYGIIPFYTILSLYEIGPPTRR  
VTRITRLMLLAKAVMNIVHVYIYLSSTILKILFLRSQTDGITNMLDVTFCMICDITITWTC  
VCNTTIIMEIINRFLKVDRLKQYADSPAERTSTNIFNRYLFFIYAYIFSLIISYLKRTYD  
LVSLYFWIYILFYQLENAISCAFFVFISALMHLLAERFRYINQLIAQYNRREIVRSSEYYT  
HDLMPASDFDGGKTTNRNEATLQLFATNSAIYGLYVDLLDILKMINNYAGLGLLAYL  
LYACYGLLSCAYNSLLFNWSPGEDLYMVWNFSWLPLYASILTILANNCATEEAN  
KTPILARVYGKGGKQYQTIIDKFLTKSLKQTVHFTAYHFFPINRATLFQIFSAVTTYLVIL  
MQFKELEKSKDKWNAEASKYNYE

>CcapGr17a

MKIINDMGSIIRLVGQIFGLIPFYECVRGELRISIFGKYTYTYIINTIFLLISLLVYLIVTKSGD  
EFNVFRNLRGADQSTEVLFSTVVIIVACASNTQRNCDIMREIGKLDL LLA KGFIHR  
YSCHILSTILALISFYMLGTACYYMQLSSHFHYHQLVLLNIYGLQLFISNLYALCLRVL  
LGNIRQRIHLVNAKLELVAKSDLDVEHNWRQMSLFIELLCKFRFITDKVNTSNGIALFS  
YMSFAFYMLTNQSFMAFMTIVKPQGYEEKYDILGVSLAWIFVELVTILVICSACDDLSS  
EANRTS  
QILARVYGKGGKEYQNIQVDFLTKSLKQDVQFTAYGFFAIDNTTLFKIFSAVTTYLVILIQF  
KQLED SRIDN

>CcapGr17b

MTNSNRTFINVKSAKKGKPGTQTILSRLKWNLIALKVIGLLPFYTMRTTRHEIATPTQCRI  
FIKRVILSKIIMCILHINALLSPSIWRMLFVQSKTNGLTATLDVTFCMLCDMTITSTCAF  
NMDKIIAILNRFLKIDQFISELPTSDAAPQKSKRTSGNKYIMWMFCYICLITIPVHRLFE  
TLSFEIVILAIAYQLENYTSYAFIVFISSLLHELTMRFYVNHQIEKYNRKFPKSSEQANIS  
SPRLEHKTTTQDILNMSRFAENSVVIYSLYNDLLDLLKMNQLADLALLTFLLYSCTG  
ILTC  
TYTSTLNFHKNESFSLMWNVSWIVNFSVAVILLALRCDRVTK EANRTSQILARVYG  
KGKEYQNIQVDFLTKSLKQDVQFTAYGFFAIDNTTLFKIFSAVTTYLVILIQFKQLED SRI  
DN

>CcapGr18

MQDET VSLNHLTKRQRYKLATAHTLVKDLKWILLYLKVLGLLPIFKITHITYQMAMPN  
NRSCAALYTHATHVFGFLLIISYIYIVCSPTANSVIFIYGEMDNATTAQILLSITCYAIVF

WKSANKAQPFLAIINQLLSVDQELQQYPRAPAALENKCDFYKGYLLMLLAHISTFWSM  
KWMNSHECNSSMICVCIFFIYLYMHAVINAYVVFVAILLHLLTMRFRYLNSFIGIYAAK  
RNAVGCAGVSTQCGDGTWAGCGGERTAEDVDFVVFVSQDMIFFYRLHNHLLVIFETLN  
DYVHIALLVFIGYFLYGSTIAVYYFYFAAILRDDIISVIWCALFLCLHGPTAAVLMRK  
SDAAAKEANATSCHLARIYGRGLDQQKLIDKFLTKSIKQNIKFTAYGFFVINKSTLFKIC  
SAITTYLVILIQFRQLEDSRAQNSKHRATDEL

>CcapGr19

MLRAKARSLKQALLATKRHCKAPRGKLTAYCIFSEIKWLVIYLTACGLLPFYEIISDSE  
IARKKGLFASLYTTIVRTLCLLVITTNSYTLFSLTTISYSFFGSVDLVNYYVLHMFNLNIFCF  
AVIIFSCAVKSRHFLRAANKLLLDVQLQRYANNTDCLNECAFQWKYFIYLGIFTVITL  
QHIDFHKMYKTSYFLLFLCYIETSICNFFIIFIALLCHELLFLFRYLNRFVKSFTNGGRN  
QKLKQQLRAKHSTENPIFYITLPEERVPRPKHEIQQESFLSVSSYIYRTHYELLKIYKQLNE  
YVGHA  
LLVYFIYVFYLTSLAMYRMIYRAKSSKLSKFPLYGFASLVHMSILVLVTRCCSKLTEE  
AQKIASIIANIYGRNEKSRKVTDKFLTTLNLKADLQFTACKVFTIDNSTLFLKCCAVATYL  
AILGQFCLLEKSRSEH

>CcapGr20

MFQSNTHFRRVLQIERSIKLYTKFVMSFLNIIRALRFYNMLLRHLGIMPIKYCMRENKYI  
LQKGWSFNITHWACQMVYLVIGVLVIRREEFFKVEYSAIENNYWNVLVSAAAFMAQF  
LINAWILALYRQHFNIISYCMYLMERLSFSAAQRKVGIFQSVMIAAVMGAVALNTYCM  
WQELAIPHSFLSICVYVTWQLYYIISILSIYLCIIQIITGHLNNIDTMNIVIVKNNHRQKR  
VTGIDLDNIQERLVAYDKLLLLCNEDISTSYGVVFLITFISTLDITFIVYILNMTTAVGY  
LQTIGSTVKTAFALPSIIFLCMMFLGSDIQEQANNTVRILAKLPRTGTGLDKMVDKFL  
KNMRKKPILTAYGFFQLDRSALFKLF TAIIMYIMILVQFSDIENSLKIN

>CcapGr21

MAIGNRENSRKSLLFHIFQFTVTLAGFCSAILACRCFPKSQHVVYKFFAGYWILHTTFVL  
MYSVFFEYSHYNFDVSGIMYFLLLITQTLAQLGVLVEGFLMTAKRALLARTFTDFEKR  
FIREFGTNEGTYTKFKREGWILTICFCTILVTIPHTIISWTLYRTRVLKTLLFWHFLPSTLT  
LQFCIFDLMLTITHLNEYAAILRNDLNEMGERNIRRVFGHYLRGLATVKGSIVRVTPQK  
LQPSEEDERRLRVYKKFYGDIYNMFKAISECFGWSLLFFIIMYFIYFVNSYYVIQSIIV  
QLDESSTVLSNLGFLMIVVTLLWILCWQSQNSHEQSRQIGCIMFKLVKPSGNKSYNDLV  
TDFSLQTLHQYVITAKEFFNLNLNL  
GTMVASIVTYLVILIQFMFSEETRVERNLMENNRNCTVE

>CcapGr22

MDQLEACVNYLLKTTFFVYLINPLPRAAPWLNIFLLSAWLLFFINYCHLPIKMLLFSDR  
ALVRWVRTCLSVGCAIISFCCMLETAVKWRQFAQIQQLRQQINYLISQQLKAHLRGRK  
RRSFGYKTRLLPLLLILIICESIKFSSVKNMTNTYFGTVTIVMGLRLRYFHALLMCEFN  
EYLCVLYETIDLLVTYNNTRPICEFNVWRPHNRWELEQLNLLRLLYGRLEFFQLINDC  
AGWSMTLIAQLALSEFVCYTYWCLTYKLVNLGIGAVIFNLATIISLGCLHYQWFGLA  
KINRKGQNI AALSTQLSKPLGSRRYNDLLLTFIQTLHQKLIVTAKDFIRIDLKSLSATM



HV FVN YLV FLL QFTYIKTKA

>CcapGr23

MFQCSRPPILQSHLGMHSDINRKDKCLKQSLPLNTRSMSNRNIHASFARYCFLLLMYT  
SLLQPIPRLLGRMLFILRLAWFVWLAISFTAVVYELLTNPYDSVGSIVGSALFIFNLLTNV  
VTFIESICKCEHYQELLRLEREVDTLVNKIDVQTHIEVRVWHMAFYVVLIQIYD VVQL  
VIAFTGYKSPVFYVYVVILFIQRARYAQITCAIARLNARSLCLITLLRVLVKANRPRHKY  
SSEVWQPYAVWEFEYLNLLRLIHGRCLCELRHRCVSNCFWSIVILLFSTFFT VSNLFWCI  
EIIRTNFQFGQFAYDAL TILRLDTLAAVLLFTA EQAKKHNM LLLGGLFNLAKPLGNKTY  
NDLVSKFSLQCLQQGFMINVKGFFSLNLSLLGDLIVLSSTYLVILMQFDLYQGSNTKVN  
VIYNSMANEESMKTEKFPTIN

>CcapGr24

MDLIDSLKPLHISMLIGGLSQYLVD RKTGKVYRSRLLESYAICTFIISLAVLLYGTFS DNE  
FFRKSDNDIGHTVDYIQMIGIRVSHIVTILEAFLQRKGQGT FVKQIQEIDRIFEC SLNVDV  
ENRSLRRKMLRRGAIMVFIYMSTELLILIA YFFT KDRHFSIYWLFYLLPFCICGVRYFQIF  
AAIMIIRQRDLKLVETLNE LNLLKAKPMQQMSEDSKTNIKFITATLPVYETTEYLKRKY  
NMDNPDMKRLLIIRDLYNRLWEVTVTVNNDFGVSILTNVGNDFISITSNCYWIFLNFKS  
YSGTLQDFLQIASSAIWSAPHLYNVLM LALLCERTIQKTTAIALGLHRMETNVWNDNH  
NTVIEQFSLQLLHQKLTFSAAGFFDI NCTLLYRIVGATTTYLIILIQFHMSADKL TTS

>CcapGr25

MEQQLRAWLRY SIVIGLYVSPAQETTHFSSCKVAKAPLAGQRQAWSTKKS NVQLIKRI  
YCYAFLFGLCVIFVNGLASQKSTPGMVSWTASTILFASQAASTMFIAWEGYWRQQEQE  
TFLRLLQEIEFSLKLRLKQDVRLDW FVSHMRWLFY YLLWLSIICYTCFVYYFTTAQYV  
GYFWHTTLYNITMRLRLIQLLIYVRVLQHYLECLSMKLRQIVAHRLAPCRQLLDVNYE  
KLQSLEYLVAIKEIYGLLFKATQLFNKFAGWSLFSIIVGYMLDYGCALCWSVLSWEGF  
LESHNYYPCLWWICPMTLILWHLCHLCNRCKQLDRLIANLSCRVIIMHSSHSLHTYRI  
VLHQFAIQLQLQRIEISARSFFTLD MRLIMSIFTSI ALYMVILIQFINIGNKYI

>CcapGr26

MDHHLLAWFRYAVILGIHTNDLDNYAIDLIELKTTLRRLFMHKLKLLICRCFN YLLLLC  
VFAVYINGLYWRKRISGIMLTWLVATILFTSQALTNLLIIETIWKEQEHEAFL LLLQRIE  
LTFKLRLRYATNKRLFAAKLRRYLYLYTISLAGLVFLVTTFWLQYIGYFWYGLWFII  
TMRIRIIQLLVYLRVLQHLYPCLALRLSGIVAYRMTPNQQILDVNNKRLASLDLL LAIRE  
IYALLFEAFQLLNEFAGWSLFSIITCYMLDITCNIYWSLLSLDGFARRRYYYISSIWWLV  
PMLAMVWYVCQLGESCKKLDRLLASLLRKIFVANSARFVCDYRLVLQQFTMQSKLQ  
HIEVCAKNFFILDTRLVMSIFTAVAAYLVI IRQFLNAQQIN

>CcapGr27

MEEQLCGWLRVCIVFGIYMRPPNDRVGDINGKASIRSTRQLLP PQQLYICLIALTLCAL  
YIHGLCWCGSLPTLQLTWVASLILFSFQVL TNFLILMETVWRHRQHAAFLQLLAQIEVA  
LKLRLRQCTHSAALQRNLCYIIVRLVVFSLVT LILFMITSIWLN YIGFFWNGLWSILTMR  
IRVIQLLLYVRMLQHYLECLCVKLQQVVEFHVSPQRQLLDIDYSRLTTVEYLLAIKEVY

TLIHDAFQLLNYCAGWSLFGIVVCFMFDVSCNVYWTVLSLDNWQNRRYYYYIAGPVAL  
LPLLVIICYLCVLCGKCKELGRRRIYLLSSIKVLNSSQSAALYRQVLHDFSQVLHQQQIE  
VTAQNFFVLELRLLRILFFVTATNFVILLQF LILEIEMVGLTEPN

>CcapGr28

MEQQIRAWLRYISIVLGLYVSPTEATPYIHNSNHCPANTSALRGERQMWLKQRQAFQR  
ERLKRIYLRLLMPLLCVLYVNGLLRRKPVPGTMLSWTVATIFFSSQALSNFLLAGAL  
WRQQEHETLLRLLHEIELSLKLRLRQDVRLDWLLPRVRWLLRYLLWTSVLFVFLVYN  
FVELQYIGYFWYSIWFAIAMRLRLIQLLIYVRVLQHYLECLSMKLSQLVAYRVAPSRQL  
MDVNYARLQSLEYLVAIKEIYDLLFKAFQQLNDFAGWSLFSIILCYMLDYGCTLYWSL  
LSWEGYLERRNYIACFWWLLPMSLITWHICHLCHSCRQLDRVIASMLCRMIIKRSSKS  
LRSYRILLQQFSTQLQLQQIEVS AKRFFTLDMRLIM SICTAIATYL VIIIQFLRI

>CcapGr29

MEQQLRVWLRYSICPLGVYLRENNENTAKIITIRNNKRDRDHKQEAWSQTRQQVTVNA  
RLQQHLVVSFLLCTICVLYTHGLYWRGYIPGLMLTWITSTVIFTSQICTNLLIIMETIWKRS  
QHETFLTLLDEIEVSLKLRLRQDVQRLALCRNMRAHLVHLLVISLAAMTLFIITSLWLN  
YIGYFWHGLWSIVTMRLRLIQIIMYVRILRHYLRCLCAKLCQIEAYRRAPDQQMLDINY  
ERLASLEYLLAVKEIYSLHRCFRLN NFAGWSLFSIITCYMFD FSCNVYWTLLSFDGF  
ARRRYYYYIAGPAAMFPLIALICHLCYLCDNCMKLAQRAAHLLCRIQIMSNAAPLRSHR  
LVLYQLSTQLQLQRMEVTGKHFFTLERFLMA IFTAIAMNLVILIQLKS

>CcapGr30

MEQQQLAWLRYCVYFGIYVNATPSTERVSSSTELCWATKQLYRKRNFYHPRAYIYFLA  
CITCIAFALSLVWHTYPPDLQLNWIATCILFISRYVTNSLILLEALRKRREHDEFLKLLVE  
IEASLRIRLHWQVESRQLLAQIKGFLKFQLVLSFLGLLPFMVFSLIVVEYGGYFWQGLW  
FICTARLRTLQLLVYLRILRHYLRGLCLKLRQIMEFHMAPSCRLLDIDYGR LGTLKCLL  
AVKETYTILHEAIQLLN YFAGLSLWGIIISCVLDLSSNFYWVLQSFDNFHGRRYYYYLIDL  
WWFVPVTVLVWELCYVCDDCMRLLNVA YFQGCTLTQLLSKIIVSSTSGIGRNYRIVLQ  
QFSMQMQLQRIEVSACNFFKLD MRFVMSICT AMAMHLVILIQF

>CcapGr31

MERQLRGWLRYCVYFGIYVGVEDTSATTISETGGDTNNKIALCATQRCHGQRPQWY  
RGQQNNKIFRQQRLYLWLLTLVLYTVYFLGLYTRSSYPGISWVAATILFTVHVLMTL  
MMREAWCKQAQHDLFLQILEKIETALRLRLRHN VQSAQLLCDLRR LIVYLSVISALGF  
ALFALTTMWLSYVDYFWFGFWLIIIRVRVIQLIVYLCILRHFLRCLCAQLRLIVAYHTA  
RGHTILDVDGVHLRSLKCLLAVKEIYELLHREFQLNEFAGWSLFTSISSYLLDAICTLY  
WMLLSLDG FVRRRNYE IAGIFVLLPLALFLWYLMFLCDNCKQLGRTVAHLLSKLILMS  
AKNPSRLHRLLLYQFSAQLQIQEIDVTAHNFFV LEMRFLIAICAVIVANLVILLQILKS

>CcapGr32

MAEYEYLRVWLRCTIFGIYVSSRSEHALREASQRQPTQRRAVWLSEPLRITFVLERM  
YVALLALIACTLYVHGLYARKIEDGLALTWL VATLVYTTQLLTHLAIFMEALWKRVE  
NETFLQLLAEIEVSLQLRLSCDTRQHALRQSLRRWLLGLVTL SLVG FVIFTSVSVLLNDI

GYLWHAAFSLLTLMRILQLLLYACILRHYMECLCFKLRQLVAYRTAPERRVLDVN  
YEKLESLTFLAFKENYALIFKAVGLFNNFAGWSLFGIIFGYMLDFTCHVYWSLLGLDG  
YGSRYTFVVGLPVLPFSVIVWQLCFVCDQCKELGDNISHLVSRLVTVTSPALKRYC  
AVVHQFSTQLQLQRIETAQHVFVLDLRLIMSICTAI ATNLVILIQFFKTEHAYD

>CcapGr33

MWQRKIKSMQPQLQSQSQSQSHSQLQPVSLDTFTLLQLRLFQLLGLCSVPLQSQADLA  
MANPRPPTTIIHLWTSVNHDSILYNSDSFGIFNDLLKFIA TIISHFIILSETIIQRQYMEQFL  
LAYTRLHNKWSSTHFRPEYAIYRKYVWRSSICIVIVIIIDLAYIQEISKNRSWLAFPIFVP  
SGLICHLRSLQIMYFMDMLRIEVMQLNRNVERLVTFSERHCVENSQRSEYFVRIVCAEL  
QILMECYQNIYELSMLLKKAVGISLTCNYIKEYVMILSECYWSYWMVYNGQDVIEYSL  
MTPSALTIFLLLITSRNCMRSTNFLAHNVHKIRHDVEDLNISTR LQSFALQISHQRIIMDG  
FGFFVLNCNMARDILGSIATYMIF FIQFMPKFKTF

>CcapGr34a

MSLQVPSVVRFHIRLFQLIGCFDVS LHTQPREQRLAEQRLVAWTVLILLIFTLTIVNTFV  
HPSAFLFTSESGYFVDALKVCMAHVAVTIIYLETIARRHALRNFWRQRFAMLNVELSK  
NEKYLPAEESDWRTQIRAYNLFYIFYGITVFDLILQIIYYNLREENEHL LQFLAMFTPYT  
YMVHLRNMEIIFHIAIIRHELEKLRKDVALLADYTRFSRRVAPFVGFE SFVRQKLA EKQ  
LTFQRIYEMYYYFQQSFGVSTIAVLLMTYTRLVVDAYFILYTYHQKKEPEFIENLLMLP  
AYLEIPSLLLTSQKCMNEVKFIAFELHNIRSSVDNSLISIQIQNFSLQILHQKIRIDGLGISA  
LDGKMLVSILGSITTYMVFFIQFMP KFKNL

>CcapGr34b

MSIQVPSVLRFHIRFFQVIGCFDASLSTQPTQQRIVEQRLVVWTFLLLLIFSMTTVNTILC  
PDAFLFTGNRFGYFNDMLKVCIAQITVFIIYTETVLGRYALRNFWQLYALLNQKAEVA  
EKLTPPKPHWPLQLCTYRRFLCIFYGITIFDIIIEILFHAMRPAGEDTMIVMRFWFMFTPY  
VYMAHWRNMQIILHIDVIRHELEKLRNDIAL LAEYTRFARRVVPFVGFE SFVRQKLA E  
KQLVFQHIYEMLYFFQRAFSVSTMAVLLMIYVRVVVDTYFMFFNDAVGWRFL ENLLM  
LPAYLEIPSLLLTSQKCMNEVKFIAFELHNIRSSVDNSLISIQIQNFSLQILHQKIRIDGLGI  
SALDGKMLVSILGSITTYMVFFIQFMP KFKNL

>CcapGr35

MDSARSTESALIVALTPYIRTFSLFAVSMPPSIILHTIPTNRRGLLFIIRTLFALYVIIHLLVS  
FWIIYVYIMLIGTFLENSLDLITCVFSVGIIIAQLLVQIAVYTQAVTGHYALRNILITGVQ  
LEKDIRKQFTIGCSLSSIRWRLGVRAGLWLLICSTFVPYLSYRLTPQKLYPLKRGIIVFFS  
CLIQIKGVEYCVGVQLVQELLLL VQQQLIHLRHKLLRCERSEARCHILYIELQANQQLLA  
RIWNLLNQVERYFCTPMLMLFFYNGFAIIQAIHWAYINFLDDLDLRLGRIVHTVMLIV  
ALL  
LPCYLSQCCIDEYNRFGTMLHKLKTVGIDENLSMRLQEYSLQLMHQRMLFTCGGFFDI  
NLKNFGAISLTITTYIVIL IQFKLQAETENKSALGARFE

>CcapGr36

MRPATYTSSALVSALTPYFWAFSLFAVVLPPCVILRTVPTQRQWSLQLFRAFFALYILA

QLATSCRVTYINSAMISRFVVRSSSDGITWLLSLGINVVQLIVQIVLYYQALTGHQLLTN  
MLFNVMELERDIRQHCKNECTLASIRWRFRRLRIGVWFLIICIVFPYLNYELATTSLTPLP  
RVLTVLFCITIVQLKGVEYCMNAQLVQELLHLVQKQLLHLRHELLRCEDVERRRIIYTD  
LQTNQNLLSRIWDLLNQVERYFCIPMLTLFFYNGFSITQTIHWGYINFELDDLNLRLCRI  
AFAVVIITTLFIPCYFSQCCIDEYNRFGTMLHKLKTVGIDELLAMRLQEYSLQLMHQQM  
MFTCGGLFDINLKNFGAAILTITTYVVIL IQFKLQAETERQSSFETRYE

>CcapGr37

MEKLESKIFDISLKPFRRLMVFLGIIPYTNYLTHKRGFCYKKVSKLSLHLLAVFTLTGFLI  
NEHVKRLQEVIYNLDDIFSSVMDGLGFISLIAVQLIVHLESTWKSQIYQNIFNKFEQMRT  
LLLQKFEMEFDPFKLKSCIKALRNHLAIYCTFLLFFVILLALQPIPTVRFLVCLHAEIVLKI  
KIFEFMYFTVVFIIVMLGDL CRAAKKQCRLLETARFENTFQQREYLLGFIALQDLHALL  
WENVQLLTDYFEWSLPDIVSSLIKYYKMLFIFVHAYFIHKDLIVWLIRLSTLTMACYLC  
TMCERRDRELRSYLRNCLKDRRNQPLMKCFHRISMLWQEPIKFDAGNFVTINMETLG  
KYFFTVMYIVILIQFRLSV

>CcapGr38

MAHRTVQPILVHFGTLFMFCKVLGLYPHDLQAFQRMHTLQGSKTGTAFVCATILAVV  
VAYNFLIYSFADEDRELKEAQSTLTFVIGIFLTYIGLGMMMLTDQWSALRNQSKIGELYE  
RIRAVDEQLLRENIVVDISKTSQRILIMLALTVICEVTILVSTYIILVDYTKWISLLWIFSC  
FPTFYNSLDKIWFANTLNALRQRFFMINTALDGMVESHEKLKRWANTGGDSLLSTRKP  
SIANVSIDASIEYLYKDFTHMEA VKKFNMPRNKITPIAHSNLFYGESFETPKKPYKLSIQP  
PSFNMVYESELNRDIEKVEENLNNLCQLHDEICEIGKQLNELWSYPILVLMAYGFLIFT  
AQLYFLYCATQKQNIPLFRSAKDPYITTIYLCYTAGKCIYLIYLSWKTSLESKRTGICL  
HKCGVVADNNLLYEIVNHL SLKLLNHSVDFSACGFFTLDMETLYGVSGGITSYLIILIQF  
NLAAQQA KDAIHNADESVSAMPTTDGGNTTALMELFTTTFLSSTHTTPY

>CcapGr39

MSEVNKQQSKPFLQSFNQIFYTCKILGIYPYDFHKFYQYGILQGSRTSCCVVIAVIFIIFV  
LFNITLTFGDQHVVAKYNHLNLIINVFTYMTPIAMFTDEITAMRNQKLLPKLFARID  
YVDEDLGQLGITVNNKR VYRRIWLMIVLTFVCEFYMFFAMMWLIVDEFWFNVLWIF  
TSIPTIYNSLDKIWFLGILLGLRDRFEAINTALDDITEEIEKNNLQKRKGLVYEQSVGVGE  
LMLHTSLSVGSERISDIKLERLVCNAFGELVCERESNKSAAVIVSYENSSYNFAALQERF  
ISLCQLHDSICRIAQQNLNQLWAYPLLVLMAFGFVFITSLLYFAYCATWNQSIPLIFSPAK  
NIYTSFIVIGYIAGKCVSLMFFSWQTTQASRRAGICLHKCGVAADTNEVYEIVNHL SLK  
LLNHDINFSACGFFKLDMGTLYAVCGAITSYLIILIQFNMAGEQVRISKELAAANESTAL  
TLAAENYTVMPLQTTMEVH

>CcapGr40

MDPRYTLFIEQQDFIFGLMPFNRTQVDSNIAVDYSLVLVMPLLYLFCYYTVNFSGFSVR  
HLPTCNSICSLSSKLLMHVGCFLYLTIIYFLTLLRQKGFYIEFEKRLTEIDELVGECHKQVAD  
TRNKLFVPKKKRLYYFTWIIILATFAFAIFYDINEMSQYYGPYCFISNMVFTFPYVAGSI  
VQGMFASYVSVISERFSTLNILFEKINHESDKRNLPIAVMDIENDA HKDNTYSGPAVISD  
LARRKNRFQKQKSVKSDDRSNNNDS DFAEDED SADEAYSFDEYDETRLSDGKTSEE

NLPSLFKLHDKILSLSVLVNAEFGAQCVPYMAACFVITIFSIFLETKVMFIVGGKNHVM  
DYVIYLYVIWSFTTMMVG YFVLRLCCNTNAHSKQSAMIVHEIMQKKPTFMLGNDVYY  
NKMKAFTLQFLHWEGYFQFNGIGL FALDYTFIFSTVSAATSYLIVLLQFDMTATLKNEG  
LTPKV FSTESN

>CcapGr41a

MPLLRRAFRNLLDADNYYKSIQPLFFCTQILGLTSYVVSTNKSGVKSLTMSALSHIVAL  
VQLVGFAYSSAYVVYFNATYIGQFLNTDITNGGEKSIVLASGATAVVVLITNFVRRRRL  
LWVLEYFPRIDKSFQRIGICWNYTKVLHCILSKIVFMICVLSIAFGIYVCCWRRLDVYPP  
LPIIFMSLNQQIGIYFTILLFSFVMLSTRVRMQALNKVLKNLAHQWDNSIVKPLPKQRS  
LQCLDSFSMYTIVTKDPCEIIQESMEIHHMICDAASTANKYFTYQLLTIISIAFLIIVFDAYY  
VLETLLGKSKRESKFKTGEFVTF FSCQMVLYLIAIISIVEGSNRAIKKSEKTGGIVHTLLN  
KAKNAEVKEKLQQFSMQLLHLKINFTAAGLFNIDRTLYFTISGALT TYLIILLQFTSSNT  
SDDHAAAHTHLNASSTAGIK

>CcapGr41b

MNSINSRTKLQLTTISALQPPPAQEMSYRVQRMRRYFISKEVFATLQPLFLFTYIYGLT  
PFRIVRRRNGTSEIRASCFGYCNTAAYVILYGICFLNSLLSAESVVG YFFRTNISIVGDTL  
QICNGIVTG FVIYATALTQRSKMRRRIEVFNMLDLNYANIGVRVKYSRIYRYALLLTISK  
TLIIVIYCAGVYLLLRSVHVRPSICVCVAFL LQH SVLLVAVCLFCFIARGFERRLVILNKV  
LKNLAHQWDNSIVKPLPKQRS LQCLDSFSMYTIVTKDPCEIIQESMEIHHMICDAASTA  
NKYFTYQLLTIISIAFLIIVFDAYYVLETLLGKSKRESKFKTGEFVTF FSCQMVLYLIAISI  
VEGSNRAIKKSEKTGGIVHTLLNKAKNAEVKEKLQQFSMQLLHLKINFTAAGLFNIDR  
TLYFTISGALT TYLIILLQFTSSNTSDDHAAAHTHLNASS TAGIK

>CcapGr41c

MSQSRRADEDANSIEAAAAAAVADQDVCATHPQRVSGLRRFFHAHQLYESVQPIFIITF  
WHGLTPFFIKSDGTGNKKLKESIFGYLNTLLHITIYVGCYVMTLMNNFETIAGYFFNTGI  
SRFADTMQIVSGLIGITVIYFTAILPKHRLQHSLRIMQDMDVLLH SVGVKILYSKLLRYS  
YLSLLVVFVAVDTVYSCGNFLLKSAELEPSTPLYIVFTLQHTVISIATALFQGI AKMLEL  
RLVMLNKVLKNLAHQWDNSIVKPLPKQRS LQCLDSFSMYTIVTKDPCEIIQESMEIHH  
MICDAASTANKYFTYQLLTIISIAFLIIVFDAYYVLETLLGKSKRESKFKTGEFVTF FSCQ  
MVLYLIAIISIVEGSNRAIKKSEKTGGIVHTLLNKAKNAEVKEKLQQFSMQLLHLKINFT  
AAGLFNIDRTLYFTISGALT TYLIILLQFTSSNTSDDHAAAHTH LNASSTAGIK

>CcapGr41d

MSFLKALFSPSDAYGAVQTVIWFNFLLGLTPFRLHNEPTSEQLRTLRI SQLGYLNTLLQ  
VIFMYCFIHSLADQASIVGFFFKSEISQIGDTLQKFIGLLGMLTLFGISSCECRVIVSLCD  
TVAYVDKRFLNIGVIFNYRYIMKLTHLKLFLVITLNASYMSSCFWMLFHNDIWPTFQA  
MITFFMPHVFLLSVVVLFSTFLFRLQQHFDLINKVLKNLAHQWDNSIVKPLPKQRS LQC  
LDSFSMYTIVTKDPCEIIQESMEIHHMICDAASTANKYFTYQLLTIISIAFLIIVFDAYYVL  
ETLLGKSKRESKFKTGEFVTF FSCQMVLYLIAIISIVEGSNRAIKKSEKTGGIVHTLLNKA  
KNAEVKEKLQQFSMQLLHLKINFTAAGLFNIDRTLYFTISGALT TYLIILLQFTSSNTSD  
DHAAAHTHLNASSTAGIK

>CcapGr41e

MFRRLSIGWRYLQRDLKHPRNIYASLRILFLLAHFYGLTPYHISVSPKTQRHSLHTSWF  
GYFNMATRCIFLHCNLMYSILHNHTVIDYFFITNVSHFTDSLQKFNGILGNLIAIAFCFFQ  
RRQLCTLMQRMYPADERFYNVGVVSFKQKFIWRTNRFSGFGLFALTICYFQLSMLGILA  
RNKVYVSLPAAVSFYAPHLLVCGIVVVFNGFIYKLTQCYRALNAVLKNLAHQWDNSI  
VKPLPKQRSLOCLDSFSMYTIVTKDPCEIIQESMEIHHMICDAASTANKYFTYQLLTIISI  
AFLIIVFDAYYVLETLLGKSKRESKFKTGEFVTFSCQMVLYLIAIISIVEGSNRAIKKSE  
KTGGIVHTLLNKAKNAEVKEKLQQFSMQLLHLKINFTAAGLFNIDRTLYFTISGALTTY  
LIILLQFTSSNTSDDHAAAHTHLNASSTAGIK

>CcapGr42

MAWQFLKRLTQSDSILQSLRPLAYISLIGLAPFRLSVNKEVRTSVFSLIAGIVHFIFYLLC  
FILSLYEGDSIISYFFQTSITRLGDLTSLSLSGIIAMVMIFSSILLKRNLLLHIIQNCLVDDIF  
LRLGLKLDYRKIMLYSFVTSGLLLFNFVYLLVSYMLLRS AEIWP SFVVFTTFALPHLNI  
SIMVFKFLCTTYLARSFRMMNEILQDILDSHIEDHDAVELSPMHSVIRMHNHNHNNPHN  
HTPIVIQPTGATRPRYSVTSLVRQNPERALKQVSNVHNLCDICYTIEDYFNYTMLAIIAI  
SFLFILFDDFYILEAIITPSRVDKFEADEFFAFFITQMLWYVVIILLIVEGSSKTIKESNKTA  
AIVHKILNISDDSAVQDRLLRLSLQLSHRRVAFTGAGLFKLDRTLIFTISGAATCYLIILIQ  
FRSTPSHWDRPSSSSVAGATSG

>CcapGr43

MLVWSKMSRVVGC FYRLAYYYMLAFGLLRFEYDFRNNDVRRTRRVALYSLCVNLLM  
LSQIPQQFQMDISTYTQQSKARDIYQFLHSLIVLSRIFACLLILVFYWLRREYIIALVADF  
LRLRRAQSRYLSPSASALRRFHFALLIKFSFSITGDLVRFLMSWRSTNWQDRIIMMFTAI  
YIGLNNQVFNFHFVVSRLINLYLETINSELQHLVSAASELNSQRIQRRHILPWRSAQLMN  
KFNALMSLRQKLQRFALRVNRLG PQIAGVMLALYCYNVGVCYIIYSLMKPHEASLLD  
ISAWNMVVLVFGTVVYIADVTEFLVISQRLRQNFGLCMPITLQLELQQLEKNWDRSIE  
QFYLREACFTYRLMLLGFISLDNKMAFTQFE SVISKSIILVQYDYKNV

>CcapGr44

MSQLITLCLKYMNYSAILVGFTRLRIDYSNQRIIESKFTSYVYIILNIVTVVLLPAAHVISI  
KYITVNFNNLLVFTDLANLFINYAVVVSFVSRWRRECLYKELSQEIFKLDLYYFNKL  
SENLOIEKRANMVIFIKMCTVSLDALVPIFGVLNQAKHVDVFVFM LAMYSSLIGTILHA  
VLFLFFYMLWHVRRRIWRLNARLIELFHD LQKLQRNARDMLAPKSRTLQQLSGLAAE  
ELREISGVHARLTVMLLRLNRVYRWQVIAVLLTYLITNISYGYFILLSLLAPSNSAQSQT  
ASSIVVSIAAS FIVFVDINLLYWTYIMDAQMKTRDGS

>CcapGr47

MRTARTELRNCIIGYANIGAILGITCLRLNFPEQRIDRPRLRLYVFSVNLLTILLVPFA  
YVLEAAQLPQGYNHNLSMLTDGMTMLITFS AVIITVLLRRKHETLYVDIAEALFQLEK  
DYFSKLRIDNAVEQRSYLLFFKVVFVFAAHTIAPFYGALVFSPTASWLNMFASFYYGFV  
YNLLFGTLFVYFYLMWLLRKRFSLLNTKLSDLLHSIQIWLEADNGFMQLQKLGVC GAE  
EIMMISKIHERLSTLTERLNADFKLQILAGLLTQLLNCICYGYYGFLNEGFLKIKVSTP  
MGVLGVMFTVIMFVDGYLLYWITD TTARSYQRAAQLLSRFPLIQILCVEFEQQCELFAL

QLKLQNQKITIGGMFSLNLQASLAVLGFSFRHILILVQFEY EYRMLTAAESPVLEKIISLL  
RLAELGWK

>CcapGr48

MSRTQIAWLRLRIIGYFNVLVIFLGLTRFRIDFVAQRVVHSRLVAYYVICANIGFLFFLPI  
AYIYEVIDDSTYFGNNLLSLTGSVSRAINFAASAFSVFLRSLREAVHLEIAETIIVLDRRY  
FSKMPMLEAEEKVEAAQQCIWADRLLYFKMFALCMQILATIYYSLPTFDELDANEAW  
LDCYYVYTQSILASTHFMYCYQMWLLWKRFGLLNKVSALLRRLQRRTIKAGCGLG  
GSDLRQGRKQLALEAVDILRGIAQTHACLSHLLERLNEG YLLQILVVLLASVIESITFGY  
SLSLVFDGSLEWEFDFDTYLF CMFICIMFIDFNLVYWISDRTVYAHGQITEILRRFQLLPF  
LGVDFEQHCELFVLQLKQQHRNTKLAGTFHLNRSSALALWAFVLTQIIVLLQFDFGSR  
LKTATENAVMSRLNKLFRIQAV

>CcapGr49

MHRVTYWVLLIHHYYMLAVGIVRFRYDFVADKPHTTTTFITYYSVVANVLTICSLPWIY  
HTSFNVFLQHDRIGKLALYIFVGSRTLRLVAVVIAIVHNWLRREELLTVIREFNFLRWTF  
FSRFTVAAPAQRHFEIKIVRNILSGLMIDVSGFIMILFAYSEQASSEFILSSIMFTFVSNILY  
LVINQYYFALRTVILFGNLMNQQLTDIMKFTHEVASLQHKHRNHVQFKARCRLLSQEL  
DELSVVYWQLQTLAERITDIFKWQGLGVLTNMYIFNVSLIYTVVSMLWTGEFTVSHYD  
WVAVLMATTSVVYLLDVREFLGLQLEHMDIFNLKFLHLETYSSLPPLDKALERSVERF  
HMQLSINKYEMSLGLFKFNGEVGFMSMFAS MLLSSILVIQFEYQYFH

>CcapGr50

MSFWERHKHNIYKYGHIYATLYGLMVINYIPQRSTNTFAYRLANIYGHALSICLIVVLPI  
YFVRNISALTA AHDQRGQLQLVNFANTLLKYITVVVTYIANFAHYAAIRAVTRERQQ  
LEDDFDGSLRAATGYDEHPRRQFEAMLLFKFALINAMMVVQVGNILFQHFTGAHSVR  
VYIAIYTFVLWNYTENMADYFYFINSSALKFYQQLNQQLRQVVREAKLLHYFRLCGQ  
RRGTVPHF CGLLCDRLDTLVQRYQQINRLYQDSLTMHQFQILGMVFITLISNL TNSFILF  
NLFVKHSGASASPAIVLNALHAIIFYVDTYIVALVSENISLEMQNINQTLRQFNQFSMLD  
VRLQQTLEALT FVMNNKVEVRICGLFVLDRRL TYLTAATGLSYFITLVQFDLNL T

>CcapGr51

MQPHAAPFWQLYTY YFALLIGLTA FHVDIEQRRVYRCTFMQVYACTMNALVVAVVPF  
SLWTSLEFLDYLN LNALMKIVNEVNVKQIFFITGLTIIFRWQRERSIYEILTQLLQLERRY  
FEKCMPRSVKAHCRQYYSNLLWLKYFSVISQSLWTLYNMFFYLPDKSVISLFYCVYV  
ICIISAVLHMLVHYFLAMWYLWQRF CWLNARLQQILDVLRSL EAHKRCDQVRKR  
AYQRLGQELMEVVRVHYQLTGFAERITNCYRFQIITVLF SKVINNISITYLGLKYGNNSF  
LKSLGVTYHIFSTSFFIITLSDSYLLDILCERVVGASHEAAEVLKRFD ELD DAVEYA  
CSIFSQHLRQHKLKINIFGMMNVNKRLSFLVFAG FVKNVLLLLQWDLAKQFEV

>CcapGr52

MPAMKKAAAVPRWYLYTY YFASLIGLAGFRVDFKERRIRRCTLMQIYACTLNTLTVC  
FLPLAYNTALKFLDHLNDNHLMKIVNSVNTVQIFFITSLGLIFRWKRENTIFEILTQILDL  
ERRFFEKCVPQTVKAYCRQLHGNRLLWLKYFSVVTQIVYTLVNLVYFMPKISITWLLY

SAHMFLLISMLLHMILHYFLAMWYLWKRFSWLNAQLRRIFNFLRLLLEAHRLLDGDLA  
SDAQRRLAGELLDVTRVHYQLTQFAEHLTKCFRLEIMGVLFSKIINNISIGYLCLKYNN  
NSYLKSLSVVFQMFSLMAIITLSDSYLLDILCERVVGASHEAAEVLKRFDELDELDDAV  
EYACSIFSQHLRQHKLKINIFGMMNVNKRLSFL VFGGFVKNVLLLLQWDLAKQFEV

>CcapGr53

MSIKANNNNSCKYLHSGIAKGKPRNSKKYKMLRERLHRIVLNVLYYNSLLVGLLPAPL  
DRGTQAFRSSRFYLIHTVFAHSFCVLVTTYAGYYYYNHGFLTRDPVLQWTYSVTYLT  
NLLVVVLVKELWCKRKRTMAVYLDYWLQEMGLEASLAASRCIIAGERIVAGLNEEVT  
GNQTPLQHNRPQRHIENLIIFKFFLAYSLVVMNAYYFLNQHPGTDATYLPITCISFGLN  
TFVLTLSGLFYCVLSQLYRQFSQINIQLCLVLEQLRRQSSEAAEVQTALASRINSLGMLH  
LAGYRLTQEIFAIGELTLAALLRLLAANMRTVYGACYFLSKRETENLWSQAEEILFTT  
LFFSDTAMVMSMLNAVLSECNGAGKLLREYTEFMDVKGCDCRLKALLTVDLCSFFQL  
DAFSTHLGCHKLRYMVCGLFEFNKMASLQFFLTILVKVTVLVQFDMQNKLNN  
RNDINVV

>CcapGr54

MVHPTGHLLTRRLHYFTLRIAYYHGLCAGFMPSHYNWHNGELHQSQYLYIYSGCIQLG  
FALGTPYAYRKYLEKSSEIFGDEVKHWTDILTNSGRYLTFTMLALFVWMRRWEMFEL  
WSVFWRLERRYERILPTDDGNTTMHTRWLLLYYFWTQHIVTPLTSIYVLLQFPRQTNA  
HIVMTFVSWALSTVQVMLAYYTLCLHCHLLQKFEAINCELQRVLTLLRNLASMERTRL  
RLLTQRLSCLANAHFAYFRLTQRLLGFGQLPFLILIKCFIVNMTGMHTCVWHIFHLSSL  
YNLVALLLVVVLVYGLSLAPSIAGLTWLLSSCNRSQGILRDANVLVALRDNYELMAVLD  
RFAVYLRSHRLRYIVCGLLDFDLQQCFSLIVVVLVQV TVLIQFDMEGEINKV

>CcapGr55

MWSRCRATFSNCIINSSIWLSIALGLLPYGYCRRRKMFVQSNFGLVYSVIMDAVMISLA  
LHLWYKEQRFNFKRIWDWSVIDLLTQALYAVNIYGMLAVLWTNLREYKSLLEIFKKFS  
ALEQSYFGKYQYLSPKCVKFHNYIWKGLALFMNNFLYFTHVFTLLGRYQVLSVRVIF  
VQLYRMILLSTILLALLHFYRHVLLIYRYMRTLTEHLKYLANYGMSNAQAYTRWHEI  
NEIVRVYVRLQQLSREFTGIYGKQVFFGIVAITLDNAHNTIVLLLIWKSQTSVWYFTFII  
YVIVINCIDFWLIIVACELAVNAARDLSLVLRCFNDTTQLDVEAERELQLLALNCASNK  
PKFRVCGFMDLNFSTGLEISIIFILMLTYMMQKY YRGLQIFKN

>CcapGr56

MMEYRYDKVNKMFNCSENIFMSFKYRSAVALLSVVEFEMSSRWSRDELSCCIINSTV  
WLSIALGIFPYDYTGRRIVQSKCCLVYTLVVDVALAVFALHLWHEQQYLNLRWANW  
SLMEVFTQFLYAVNIYGILAILWTNSWEYTYVWKVFKEFAALERSYFGKHQSLSAKC  
MTFHNLLTCKGLIVLMNIFYLLIFTAMLIKARFIRTRIFVAVFSKFLNIVILVIMHFYTH  
VMVTYRYIWLKERLKHLSNFDTPHAQAQNLWREINEIVRIYMRLQQLSREFTRIYGK  
QVFFGIAALTAENAQNVLVLLFMLTPPAETWKIIFSAFFIAKNIVYFWLIICACELAVEA  
ARELGQLLRFCFNAMPQVDVEAERALQLLSFNCIINKPKFRVWGFVELSSSMGLEIILV  
LQIIYLLQSNYTKMSLFSN



>CcapGr57

MWAQCRRRAISDFTLNSAIYCSIALGILPYGYNRPRRGMVPSKAALVYSVCIDLLIICMA  
LYAWPSARLIHFERIRHWSLLELLSQFVILVNVSALLNIMWTNWSEYKCILGLLNEYSSI  
ERSYFIRHTSLMYECAAYEHYIILKGLAMLLKNLSFVYVALVLFKDVTWPIVAVHIMT  
MLLVNIISLVVMHFYIFILVTYRYIWIMRERLKCANDLGLPHSSNLQSIRYELSEITVIY  
MRLLRFCSEQFVRIYGKQMLLCISGVAGVNVLTLLLLFVWREVVHKLVAYSWSYVLV  
VNTLDFWLIISACELALGTAFDFVALQRSFSNYAPLSAALERELEMLALVCATSTPQFR  
LCGLLDLDYSTGLKVLMTTILYVIYLVQMNY KNL

>CcapGr58

MWLDTAKKLFCAARKSVAHTCVSTQLLLSIAFGLFPYKYNSKTRRLTTAKWLKYYW  
VLVNIVMVLITVYLNFFKRTPSATSFIIDKPLNKLLAYIHFLGLVVFVVISSANFRHRAE  
VLILHNAILQLQQQQQRWQQKWSKKSNPQISALFENYIIAKNVMTFLQAASNVHAKLG  
FNPNSFKYVCFTILTICVKNITLFTVSNFYLTLLYIYRLLQQLNWNFKEVVSFYSLRAR  
DVPDTPVIEDTTEIAFDLKFATPCKKTSVECTGVDVAAIAELCRQYVRICQLIRRVCKV  
YEWQVLLFLIILFGNVMSTFYFLVYLVGKMLPKELFSPTLFLLELFFINVLDLCCYMVIC  
ERSMMSAKETSFILHQLCRLEELPDEIRNEIEMLSIFMAGETIRIRFCGLLEWNFRTGASF  
MIATILYLIVLVQFDYYN

>CcapGr59FIX

MSFRLSKRISRAFVKTTLVCSVCFGLLPFYYSERQRRFISTKSITIYNICINVIIIVLMAFL  
LAVYASLDDAFLTGDPLAMILNRLAAYMTFSALLPMWVWNWLRQRQLQDLFNEFAA  
IEADFFYPYRHLLADCTTYDNYLLWKGLASLLQNLISIFYTTTVETPSILLIIFMCLLTILT  
NVVLLVATHFYLVVLHTYRLIWALNRRLEAIAADNIMPRLCQRRHLLSVEIDTMASIY  
ARLISLCERYTRMHQLHLLLVIGSVTACNIEVLFYVRLLWSGKIPERTAFNVFAVFQIFV  
VNLLDFWLTITICELALVESRKTSAILRGFSAKPKLTLNVERSLECFAIICSSTKLRFHICG  
LFDINHLTGLKVLSTMILYLIYLVQY YHDNL

>CcapGr60FIX

MSFRLSKRISRAFVKTTLVCSVCFGLLPFYYSERQRRFISTKSITIYNICINVIIIVLMAFL  
LAVYASLDDAFLTGDPLAMILNRLAAYMTFSALLPMWVWNWLRQRQLQDLFNEFAA  
IEADFFYPYRHLLADCTTYDNYLLWKGLASLLQNLISIFYTTTVETPSILLIIFMCLLTILT  
NVVLLVATHFYLVVLHTYRLIWALNRRLEAIAADNIMPRLCQRRHLLSVEIDTMASIY  
ARLISLCERYTRMHQLHLLLVIGSVTACNIEVLFYVRLLWSGKVPERTSFNVFATFQIFV  
VNLLDFWLTITICELALVESRKTSAILRGFSAKPKLTLNVERSLECFAIICSSTKLRFHICG  
LFDINHLTGLNVLSTMILYLIYLVQY YHDNL

>CcapGr61

MWSRWRTDMARFIYNATMWSSIAFGILPFGYNGKLQKISPSKYSLAYSVCVNVLQVV  
VSIPALPTEDFLNINVVKRNKWMDLLTRVITIANIYTLLVIVYINWRHYKSVLNIFNEFA  
DIECQYFAKYGLLAWKCSAYEKYIILKGLATLLRNLSFIYFLSGLAKVVTWNILLVLAL  
ALLLGNVILFVVLHFYNFVLTTRYRYMWILKRRLEYLSEHDVVHTHSRQLANEVYEIIGI  
YLRLIKISKCFGRYIGKQLLSFAIACGNIQALYYLGLLWADQFVGLSLLEIFNVLHGVS  
VNIFDFWLTITVCEQALGMPKEVTQLLRFTIVAELDVEFEKSLEILSVVSKTHKLEIRL

CGIIELNHL MGLKALLTMTLYLIYLVQFS YNDF

>CcapGr62

MSFKRMRYNQKNINTHSEPFLALISKLYLLCQLFATAPILDTPKNNRRILRYVRYVWC  
STLVITYITVNNISYMSRIHDPLVNIVKAFFAFELINNPLIVAFIVFTTYYTAEYRSIPSEL  
LRIYKVYQRLQANISAEKEFFRVLHMEVLVMAGVICSILSTFVVDYFRSERTLMRAILH  
ASVFSVANILISLHICQFWLSLRFISHL YRGFNEVLRQRLRRFRVIDLTALDECNNDLHA  
YGREISWYQREFQNTCNAMATDSFEILDMMRDIYFDLDTISTQLTNVFGITLLFNFLGS  
CVSLSVQFFAVFKFFDTENFLNDVQAELYNRILWIIHLGRIMTVLISNNAIIEEKCRTSC  
VLNELQVSSKEMERSINRFLRLMTHQPPEMLCGILQLDLLVLCGITGAVSNYFIFLAQI  
DLGTLPIQQKTTNSTA

>CcapGr63

MTPSNPRKPPNSKGHKQIRPSPLHRLMHVLT VFSQVFCMAPLPKASAGRNLNLRGFHY  
LYGLLILAHYILSWKQMTLTDFTILTINKIIFGDIIVNNLVSTLAVCSTFY YAGLYREV  
PQKCLQILQALKRLQGVQRSNKWNSLSVLYKELCCYCCCITLVTVILAFTYFSLTSEY  
YIFLLQLGSYHLPNILLNLKLGQYWFTRFAYHIYKRLNETLRERLKC FDLSGSTPSENE  
ELPMFVSLNSQPINDGEPASSYTRAYLSTHSCGHSDAEFLEHIRRLHTRVDALMAELAE  
VYGSIVLLNYLGSIFAVAINIFELYQG FYRPVLVLNIVLWQIQYMIRILPVLNASNAVVE  
EICRTTRILNALYITTKKMQLAVNRFLQLQVRKPAESAGGVVNL DMLITGFVSAIGQ  
CVTFLVQIDLSKVHYKY

>CcapGr64

MLPSIVRGSRIKRKSQKIRPFFFGPVGQHITVDVLETDLYRVVQHFLLLSVFFVALPYGA  
HQYMPTNRSSDHFLFLHIVWCICIYICLVFAVYGEYTQTLMSLPTIQKPLYFAEYLIYLI  
HIFHIAFMGYLARDKCRRFLRTIAEFDLALMSFGKKPKYQRLKSFIGAHLFLIFIFLITTA  
VVDYFYRDCVVASSIRSLTVYLLPNLILCVSLVQYYTLLY AIAQRSIRLNEILHEEISRQK  
CSRTLQRIRLLYSALQVFTKDVNSAFAPSMVLVYVGSFSNASVNLFLIYKYVDDIQGVA  
VWWILYSVIWTHITKMFLILFYNYDVQRQKNSTILIINEIGRTQNKEVEETVTHFVLQ  
LTINTRTNVVCGVAELNLSFITT LLMAMSTVFIFLLQYDITYEALTLTHSTGGPIRNN

>CcapGr65a

MQMHLNQRDLTSNSTVESVLKFAYRISACLGINAYVYHSAVPVDSWYNKRRQQRRSG  
NSQLRNRQRRRPDIGRPQLPAHISQVYFSRSLRLYSLAILLMVIIAIPFVAYTMYLNQLFL  
RNSRILSLVGFGRLTFLHICAVGTLLRQYCKQQELITCTNRMLRLQLNIRKL VQQLGRK  
TLAQPVGLRTHCDEQLTFMLWLMLIPLLLSPTWTVFIVKIENVLVRPAYLAGLLILYYC  
QLVLQLTLCNCFVCTLILAQQSNRLNMLLRAALHPTARHANVPQPLNIARRCHLNTLT  
YKIRQLQRLFEENSQINGLLFRLYGPQLVAFLGFVLTECTVQS FVLYYVSCSPHTALPY  
RYVGETCTIRWNLWAVIYVFGLLCNMCLVVGAAHQLSRVQYIRIVLAEGSARMPLV  
KAMCRHTYSEVQLSRTQPKKGRRYVYFMICGLFAFERRLLFVILQ TILNYLL  
ILIQYDKVTK

>CcapGr65b

MRLQKGKQLQRNTLQPRLIHRLEIIHYHMTTLMGITTFVRERTTKYREPRAAQPHALRL

NKQQWHRHNRVSLRSSHALLRWYSFILLILITAIPIFVLYTLYLDMSFLALTRLLAVVGN  
MRYVLIQGSIIIVMLINCCRQSRLRLFLNRLLRMQSNLLALTRHLRISGVTLADLQSF  
RSTCVLLCTKVLSAMVPLVWTVYVIYKELAWQRPRLASLLFNWFCTVFLQMMVVL  
YYFCLLRVVHYGKQLNLLLQNAISEARGHLDAVRPLRVTRNRVFCCLRALRRLQCL  
YKRNSLSGREL VHIFTPSLLCFSFFVVICVQLFVIYFLTC SHKAGKFRTDGTVTHRDR  
CSSPNFVGVIQVLT MFGDMCVVIWGTHCLQSRLSETRLMLAESSIELSRSRDVCGRGC  
EAQLERVQPKKGRRYVYFMICGLFAFERLLFVILQTILNYLLILIYDKVTK

>CcapGr66

MTQWADRRRRFYSSVAFFISIRSI LLTAQVMLLAPIIRSKQRGIYRTHRVL SYLSFFLLLT  
MCTAAPFLLRHIIASTYASVGVQYDALFVRIAMISQVSDISIVLITLSQIWQRQRLCGFLN  
RMQDTVQRVRSIYGRNFVRLRVLLPLWLQLGLTLYDVL MQLVFLVKLAIKIDPWQFA  
VNLASLYLQQCRATLQLLIMASILLTACYAQLALRLED FEQSKIVEFTYFEDLVALQK  
ELDELLARLQCVFQLPLLLLVAGEFINVLANLYAQLYYYVITKAWWFAFV FYYAKISA  
ELYLLIHTVYFCCVLRVITHLFLDREVD FEDAPETSRYELTHSDVMWPQPTQFFILGLF  
ELNNEFWLFLVTYSVNFIVIILQFGFFT

>CcapGr67a

MIKKWIGKGIQLIFFAIFYVRAFGVCTFTYRKGPPTANVQAFKNRGYSKRRPYAIE  
CTWFTWICTIFRILSACGYFYGSLMFWYTIDPFFNIITFLQSVTVSGAIVMGVGLLRSR  
TIFVNIVNYFIKLFQRTQRLAPHKSMCVQHFAFLLIILCMTRDIVDLLVLFDYTDYRER  
AVAFAKLYLEISLIFTIHIGCIGYMSVGALYEFVNKYMREEFQPRARCLDYVLKRHNHNI  
RSQRYQRKLIRLTRELNECACIYNNIYRAATSFHHHIRFHIFSLIFTFSFLT TVIYSVLYL  
YTFYHGFIWWALFVCLQVFTELLIMIAAAKSAVQGGFSVNKLSLDSIYMAGNAEWNQ  
SVEIFINRTNLYEFKPNLLGLFDISSN ILLMFLSASVTYLT YILQNTMLSQH

>CcapGr67b

MANKWICFAIEVLLNLFFHFTRFCGICSLNRHKERRSDNRLNIAPYSKGPIYTVTQCTW  
HKWLF GIFRILSAIGYLYGNLAMWYTADQFFNLINILQSTLSVSGSIALSVVLMRSDVRF  
ANNGNYFLHLYNRIQRLAPDKSMMGLHQLIFLLIIFVCMGNDLLGFCVLRDYKNVRE  
MAVVAAKLFLDMCLVISMHVASIGFICIGALYGF MNAYMREIFEPRARCLDYVLKQKH  
NYRLQMNQRKLIRLTRELNECTSIYNDIYSAANTFHHSIRYQIFFALLFEFSLFTTIMYSM  
LYVRTVLHDFHWDVAVVYSGQIFIGVLILILSVYS AVQSGLTTRNLSLDSVYMDGNAEW  
NRSVEIFINRTNLYEFKPNLLGLFDISSN ILLMFLSASVTYLT YILQNTMLSQH

>CcapGr68

MSSYSRSERFAYLTLVALYYYGRLLCVFDWKLEKSKMLVRPTKCTNRILLIWRIFITM  
IFISVMPSMMAFPQRLSIDSFLAFFANLQVITVTLFSVISFLIQELSERKIFKII EKIVKIYKR  
IYGKSHVRQILGRFTVFSVITKLLLSFLGLIYEVPLILEGDQSVKSLCGIYLWLGTIYTLD  
ACFLGLIIGQMYTAMAAHLGEMVQAMSDYESEEPSTLSKHERMKRLCVHSESIDDS  
NNIYFVLYGLTKQFHIFRWQLIYSIYYNFVILMVMHSFIWQYIYAGYVDFLALFSSLF  
KLWNLAMLILAAHG VVEKSQLPDLLNLDLVCSDIDARWDESVEAFICQRKVENLEIKV  
LGFFHLNNEFILVIISAIMSYLFIL IQFALT SKYKYMHTYIYICMYIT

>CcapGr69

MTDLKWVKICVGSVIVYLMLFGFLAHSVSLRKGRFKDSHLMRLYSFAIATTFTMSYGR  
NMLIDLLKNRFNLENIQLYSYMNIVAVIVNYLTALILSRTFVRFFNRVALYETIKFFDV  
NAKTLIVSVRLVLLKGLLIPVIYEVTLILQQIRNAPDKHLLWTFHTLLPMILSQMFNLF  
SALVICKTLMVSLNERLMEIVNEVNCLQNILQMSLQKPYRMRQFCELAADRLDLLAQK  
YDIICKQTTVYLHLLSAPLICSLLCNLCGITAGCFTQYLAIAETLVNNEPYDAFNALTNA  
VFLVISIVEVLLQGHICDDNRLKVQETGLILQRMNLTHADLRFKESVEAFSLLVLVTEY  
KIQPLGLLEINISLVQDVLSAVTSFLLIF VQSDTLTRFSLK

>CcapGr70

MSMFSWFWRTLHTYVTGNFHCSEFAYIAYPFGRYKLLGFTPIQLRRGEFLRNGCGSR  
MQWDTKAWMWTAVNVFVYTMGFFSGVYQLQSADYKRGDGISRLTGWAQVYTLFTL  
GLVAWITSWTHMEEQCRLCALIMKIDQRLRDIVEIDFDYKTLRKQLFWQLGLELLIAFP  
LSMVNCIIIQPDKSFMPMSACFWFICFAPISLLIFKQFQFCHLMRLLKTKFELINNKLCK  
FNRDTHFACGRRRIMTVRKMNSVSEVPLTATVATSAAVEAEELTEEPNVDILQKLLTIYS  
NVSDGVDLVLSIFGCHLLCLTAVSFGVITVQSYNLFSLVSHTLKIHVYEIVFIVAWILVQI  
MAIAVNVMICSRTSRAMENTVSVLHKMRVSSNEASNASLFYQILQIFSMEVMQRKRNF  
NAAGFFDMDYKLITSILASSTTYLVIIIQFHLTNIPDCHFPKE

>CcapGr71

MKVYTENIDSIDLCFKWIYNILYYGGCLSFQLRRKTLQLTKGNIVYTHFIRSALILSFVG  
SIALKNSSGYGSKAMLDQLSPVLKLILGFETFTSTVTYIAVTSAMHANRYKHLKLLFQF  
KELDEQMQUATYPKIKWNYHKTMRKFTPTLCGMFYYYAVSLIYVFNLNLCNCDYATT  
LMFALSYASITITPGCTFFLHLGMMDLQIRIRYRIQRLLRQEYCGLRKDASKQCKFKLR  
ITRLIDYKRYIELILQINDVFGVVCISLFDHFTVLTNMTFLMCQKATESGTRSEEYVFI  
FLFMLPRIYKVTIYAVYGYVTQMEQRNCAHEIRMSSKYFRSSLVMRNKLSAFLHWQM  
QKKYTFLVGRMTRCNLILLYTTVNSIASSVII  
LIQLQFQQNSITERMKNGRMLQDVELI

>CcapGr72

MSWRRLFYQPESTYEANGLFTAIQFITCCNGFIYRRGHFIVNFWTKLYTMLMPFITILSL  
GLGIQQLINDPVESARFEETDQLWLAVCVLEMIMATVAYVLIVYSMVKHTRDHVELY  
DRISALDRLLL RDFGVNLKYHKLMRKNLIEYVILSIVYCVALCWTLFMVKPNKMAHA  
CGMAYAYLAMTAGPHCSSYLQVNFAAMLIRFRLQKLLDEKFLLAKFPQSGLREVR  
LLKLVDVVRAFHELIDAINDVYRVTLTVGLAHDFTLVTHILYMLFGHSMGEAVDGVFF  
AFGGMWLAVPLHKFLTAPVYCNRAIEEGKRCLRLIEKIDICFPNFKSAKRIVTATMHW  
RLENKIQFTCGFNMIYNKTIITITAVVFNYLLILIQF  
RMTQLMGKQIEEQKNILQDWIGDL