The annotation of *Drosophila littoralis* fosmid, XAAA122

Jenny Myoung
Washington University
Bio4031: Research Explorations in Genetics
May 5, 2004

I. Overview

The annotation of a 41 kilobase portion of the *Drosophila littoralis* genome, fosmid XAAA122, was done by comparative analysis to its distant relative, *Drosophila melanogaster* (Figure 1). The segment of DNA from XAAA122 is more similar to the 3R chromosome of *D. melanogaster* than to the 4th chromosome, and most likely comes from the 3R chromosome of *D. littoralis*. Four putative genes and one repeat were annotated for this fosmid. Multiple sequence alignments between the protein sequences of Best1 gene homologues in multiple organisms demonstrated the conservation of the Best1 gene throughout evolution.

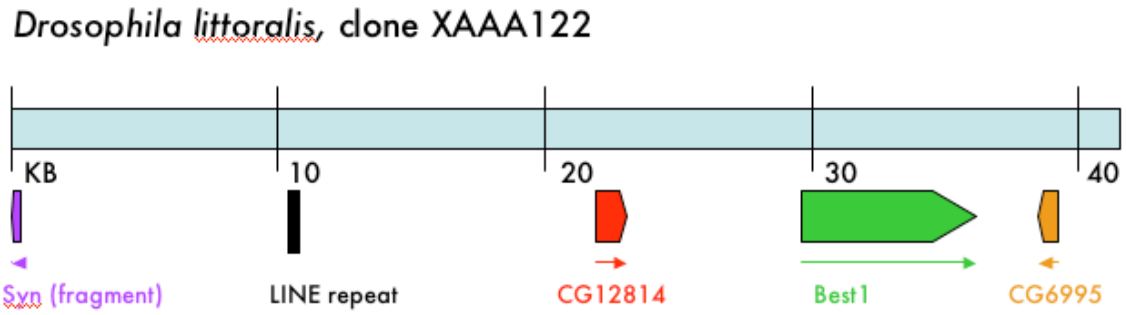


Figure 1. A schematic annotation of *D. littoralis* clone, XAAA122. The colored arrow blocks represent the position and 5' to 3' direction of the putative genes. The black box represents the single repeatable element present in this 41kb fosmid.

II. Genes

Four features were determined and annotated for XAAA122: a partial fragment of a gene at the end of the fosmid, possibly generated by the random cutting of fosmid ends in the production of the clones; two putative genes with a high degree of similarity to homologous genes in the *D. melanogaster* genome; and a final feature with a highly conserved RNA binding domain that may be a gene that was subject to rearrangement sometime during the evolution between *D. melanogaster* and *D. littoralis*. All four of these putative genes have the greatest similarity to sequences on the 3R chromosome of *D. melanogaster*, suggesting that XAAA122 is from the 3R chromosome of *D. littoralis* and not the Dot chromosome. A PCR probe from the *Best* gene on the *D. melanogaster* 4th chromosome was used to locate this clone.

The *Best* gene on the 4th chromosome of *D. melanogaster* (CG2165) was compared to XAAA122 using a Blast2 alignment. A TBlastN alignment between the protein sequence of CG2165 and the fosmid wielded no sequences with significant similarity, even with very low stringency parameters. The same alignment performed with a different *Best* gene present on the 3R chromosome of *D. melanogaster* (*Best1*, CG6264) against the fosmid with the same search parameters showed a strong similarity between the two sequences, with percent identity averaging at around 70-80% similarity.

The gene prediction program, GenScan, predicted six putative genes on the fosmid (Figure 2). BlastX searches were performed using each of the predicted mRNA sequences as a query against the NCBI protein database. Of the six predicted genes, only three (XAAA122.002, XAAA122.004, and XAAA122.005) showed promising hits to genes on *D. melanogaster* and the top hits of each search were further analyzed to determine the presence of homologous genes in *D. littoralis*.

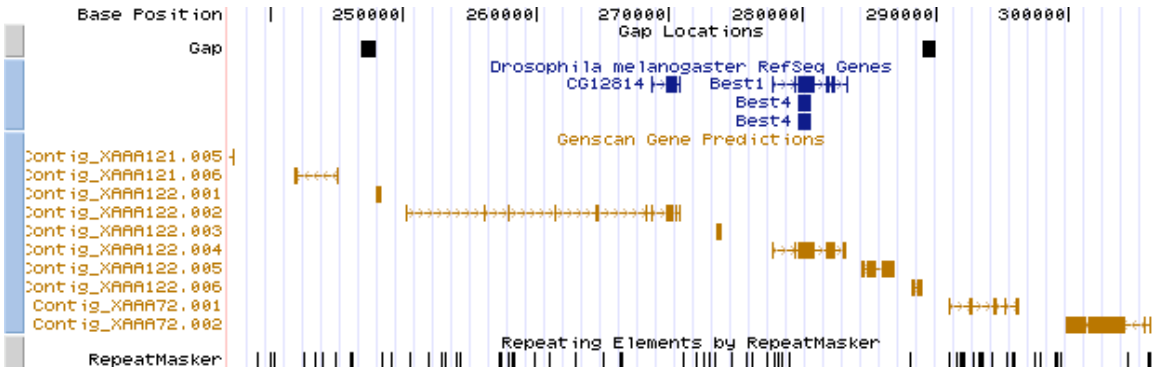


Figure 2. The GenScan gene prediction output produced by the UCSC Genome Browser. In this picture, two of the putative genes, XAAA122.002 and XAAA122.004, show similarity to RefSeq genes present in *D. melanogaster*, CG12814 and Best1 & 4.

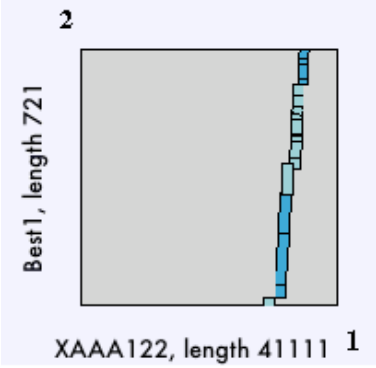
Bestrophin 1

Bestrophin 1 (Best1) is a protein that has a putative function in chloride ion transport across the membrane of the retinal pigment epithelium (Sun et al., 2002). It has multiple homologues in *D. melanogaster* and there are several Best genes present on different chromosomes, including 2L, 3L, 3R, and 4. A BlastX search querying the fosmid against the NCBI insect protein database resulted in high quality matches to multiple Best proteins. The highest match was to Best1 (CG6264, protein GI:21358115),

a 721 residue protein with its gene locus on the 3R chromosome at cytogenetic position 85F615.

On *D. littoralis*, the CG12814 coding region was determined to span fosmid XAAA122 at base position 29745-35449. A Blast2 alignment between the *D. melanogaster* protein and XAAA122 showed a match of 70-80% similarity along the entire length of the protein (Figure 3a). The boundaries of the coding sequences were found by extracting genomic sequences corresponding to predicted intron/exon boundaries based on BlastX outputs and then looking for splice sites specific to introns, with "GT" denoting the start of the intron and "AG" denoting the end (Figure 3b). My annotation is similar to the GenScan prediction for this gene, contig XAAA122. 004, which predicted seven exons with similar intron/exon boundaries.

(a)



(b)

Exon	Start	End
1	29745	29815
2	31543	31653
3	31714	32137
4	32227	32559
5	32622	32865
6	33808	33992
7	35094	35449

cds:(29745-29815;31543-31653;31714-32137;32227-32559;32622-32865;33808-33992;35094-35449)

Figure 3. Bestrophin 1, CG6264 (a) The Blast2 output of the alignment between the *D. melanogaster* Best1 peptide sequence and *D. littoralis* XAAA122 genomic sequence shows a match along the entire length of the protein. (b) The coding sequence boundaries of the putative *D. littoralis* gene for Best1. Lack of information on the untranscribed 5' and 3' ends of the mRNA transcript make it difficult to determine the exact exon boundaries of this gene.

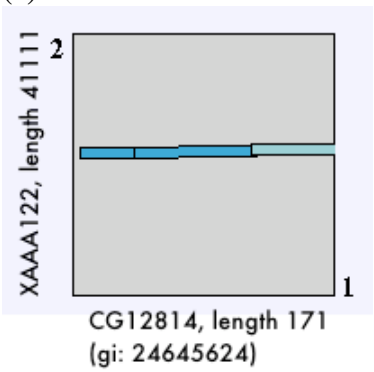
CG12814

The function of the *D. melanogaster* protein CG12814 (protein gi: 24645624) is unknown, but it is 171 amino acids long and is present on the 3R chromosome at the cytogenetic position, 85F16.

In *D. littoralis*, the CG12814 coding region was determined to span fosmid XAAA122 at base position 22192-22912. Blast searches revealed this gene as the top hit, but the first seven amino acids of the protein were not part of the output, even with searches of very low stringency parameters (Figure 4a). Using intron and exon length information provided by Ensembl, the nucleic acid sequence 200 bases upstream from

exon 2 was extracted using ExtractRegion and analyzed for potential start codons and intron splice sites. The translation start site was found about 90 bases upstream from the start of exon 2. The boundaries of the coding sequences are listed in Figure 4b and the protein was translated and found to be 172 amino acids long. Blast2 alignments were performed with sequences of the 5' and 3' UTR's from the *D. melanogaster* gene, but no significant similarity was found to the fosmid.

(a)



(b)

Exon	Start	End
1	22192	22210
2	22279	22473
3	22543	22685
4	22754	22912

cds:(22192-22210;22279-22473;22543-22685;22754-22912)

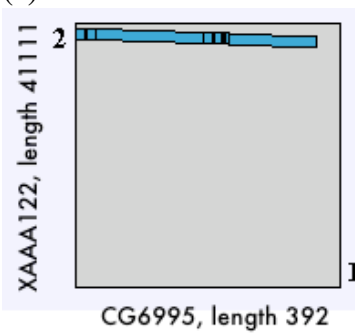
Figure 4. CG12814 (a) The Blast2 output of the alignment between the *D. melanogaster* CG12814 peptide sequence and *D. littoralis* XAAA122 genomic sequence shows a match along the length off the protein except for a seven amino acid region at the beginning of the protein. This coding region was later determined through analysis of upstream sequences of the gene. (b) The coding sequence boundaries of the putative *D. littoralis* gene for CG12814.

CG6995

The CG6995 gene in *D. melanogaster* encodes a protein (protein gi: 21355679) involved in RNA binding that contains a RNA recognition motif and SAP domain. It is 392 residues in length and resides on the 3R chromosome in the cytogenetic region 96A23.

In *D. littoralis*, the CG6995 coding region was determined to span fosmid XAAA122 at base position 37909-39075. BlastX searches revealed this gene as the top hit in this region, but a Blast2 sequence alignment between the *D. melanogaster* protein sequence and XAAA122 show a match along the length of the protein except at the last 36 residues (Figure 5a). The CG6995 gene in *D. littoralis* is also in the opposite direction and different position in the 3R chromosome, suggesting some sort of rearrangement occurred during the evolution between these two *Drosophila* species.

(a)



(b)

Exon	Start	End
1	39075	37909

cds:(39075-37909)

(c)

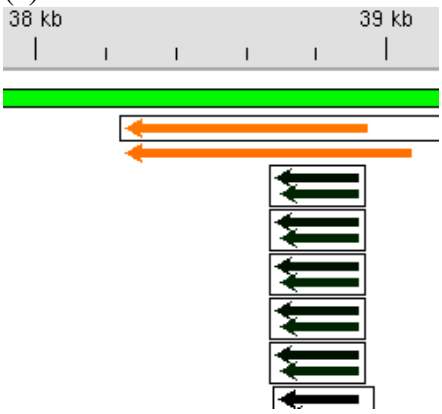


Figure 5. CG6995 (a) The Blast2 output of the alignment between the *D. melanogaster* CG6995 peptide sequence and *D. littoralis* XAAA122 genomic sequence shows similarity across the length of the protein, except at the last 36 amino acids. (b) The coding sequence boundaries of the putative *D. littoralis* gene for CG6995. (c) A Herne output of the top BlastX hits to this region. The *D. melanogaster* CG6995 hits are represented in orange while the black arrows represent hits to several RNA binding proteins in a wide range of species, suggesting the presence of a RNA binding motif.

The *D. littoralis* CG6995 protein is 389 amino acid residues in length, which is similar to the protein in *D. melanogaster* at 392 residues. The comparative length suggests that despite the 36 amino acid C-terminal discrepancy in sequence, the *D. littoralis* protein may still be functional. The region coding for the highly RNA binding motif, which shows similarity to organisms ranging from other insects to higher mammals, is not affected by this difference in amino acid sequences (Figure 5c).

The coding sequence boundaries were determined by analysis of splice sites, and only one coding region was annotated (Figure 5b). Alignments to 5' and 3' UTR's in the *D. melanogaster* sequence to the fosmid showed no significant similarity, making it difficult to determine these untranslated mRNA regions.

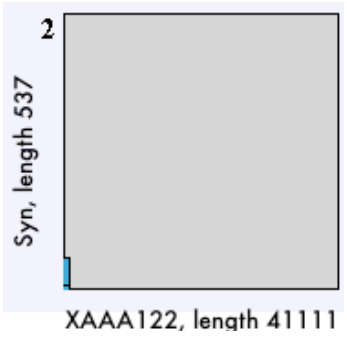
Synapsin

Synapsin (CG3985, protein gi: 24645628), Syn, is a protein in *D. melanogaster* involved in neurotransmitter secretion and synaptic vesicle exocytosis. It contains two conserved domains: a biotin carboxylase N-terminal domain and a glutathione synthetase ATP-binding domain. Its presence in the *D. littoralis* genome was predicted by GenScan

as contig XAAA122.001, but due to its small fragment size the top hits of a BlastX search of the predicted transcript sequence against the NCBI insect database were to gene sequences in the mosquito, *Anopheles gambiae*.

The Herne output of a BlastX search against the *D. melanogaster* protein database showed significant hits at positions 2-205 to Syn, with 91% identity across 64 amino acids until the end of the fosmid (Figure 6a). Previous annotation of the nearby genes for Best1 and CG12814, which are adjacent to Syn in *D. melanogaster* and show a similar orientation in *D. littoralis*, provided additional evidence that this sequence is a partial gene for Syn that was disrupted when the fosmid ends were cut and not an artifact. The full length of the protein is 537 amino acids. The gene for Syn in *D. melanogaster* is alternatively spliced and exon numbers range from nine to fourteen exons. The initiation of its coding region was determined by genomic analysis, but additional exon, coding sequence, and 5' UTR sequence information about the *D. littoralis* Syn gene could not be determined due to its partial sequence.

(a)



(b)

Exon	Start	End
1	205	2

Because this is a partial gene, the coding regions are unknown.

Figure 6. Synapsin, CG3985 (a) The Blast2 output of the alignment between the *D. melanogaster* Syn peptide sequence and *D. littoralis* XAAA122 genomic sequence shows similarity along the starting length of the protein until the fosmid end is reached. (b) The coding sequence boundaries of the putative partial *D. littoralis* gene for Syn. The end boundary of exon 1 reflects the end of the fosmid and not the end of the coding region.

Clustal Analysis

A multiple sequence alignment using the program, ClustalW, was performed on homologues of the Best1 gene found in *D. littoralis, D. melanogaster,* the mosquito *Anopheles gambiae, Caenorhabditis elegans,* and the mouse *Mus musculus*. Best1 is a member of a potentially novel family of chloride ion channel genes known as vitellinform macular dystrophy (VMD). This gene family is thought to encode proteins that form oligomeric chloride channels and has four predicted transmembrane domains

(Sun et al., 2002). A Clustal analysis of the peptide sequence of these homologous proteins should show conserved similarity between the proteins that make up the four predicted transmembrane domains.

Four regions of especially high similarity were seen at amino acid residues 1-55, 75-154, 183-260, and 273-357. This high conservation, as well as the high density of hydrophobic residues expected of a membrane-spanning protein, seems to indicate that these are the regions of the transmembrane domains (Figure 7).

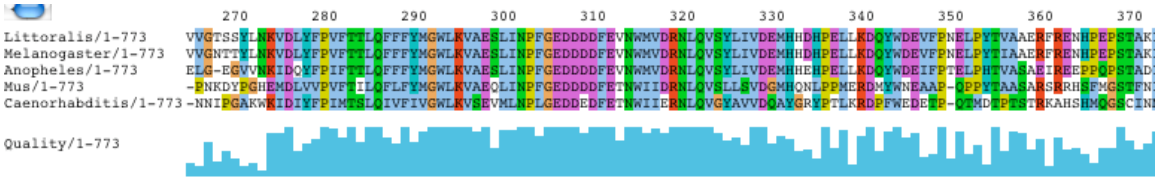


Figure 7. A JalView output of the ClustalW multiple sequence alignment for the protein, Best1, with sequence from *D. littoralis, D. melanogaster, A. gambiae, M. musculus,* and *C. elegans*. This region represents one of the four putative transmembrane domains. (Amino acid color key: Blue=small, hydrophobic, aromatic; Purple=acidic; Red=basic; Green=hydroxyl, amine side groups.)

A multiple sequence alignment was also performed on the genomic sequence regions 1-2 kb upstream of the start of translation of the same series of homologous genes to look for the presence of any conserved upstream regulatory elements. There did not seem to be any conservation of specific promoter sequences from this Clustal analysis, but this does not negate the presence of conserved upstream regulatory regions.

Repeat Analysis

The program RepeatMasker reported the presence of only one repetitious sequence within this fosmid. This 204 base length repeat belongs to the Penelope LINE family and falls on the contig at positions 10319-10522, a region that does not contain any features and is about 10 kb away from the nearest gene.

XAAA122 was compared to *D. littoralis* sequences generated by fellow student researchers to search for potentially novel intra-species repeats not masked by RepeatMasker. BlastN searches against a database of these masked contigs at varying stringency levels revealed little significant evidence supporting the presence of novel repeats. There were some regions that received several hits in the Blast output read, but

those regions matched to very small segments of only 20-30 bases of similar nucleotide sequences and were not significant enough of matches to be considered novel repeats.

The frequency of repeats in XAAA122 (one repeat in this 41.1 kb fosmid) was calculated to be 0.0243 repeats/kb, or 24.3 repeats/Mb. The final annotation of fosmid XAAA122 is shown in Figure 8.

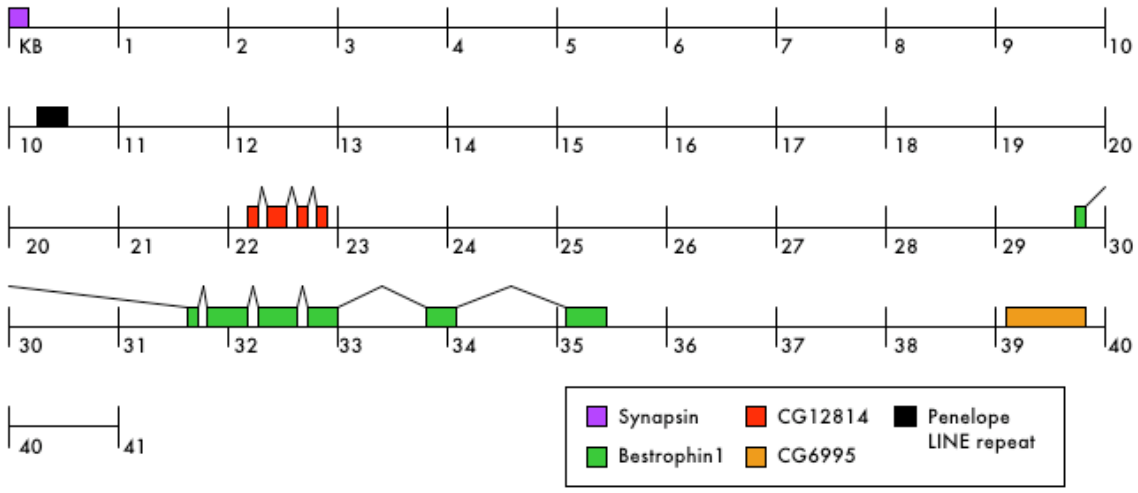


Figure 8. The final annotation of *D. littoralis* fosmid XAAA122.

Synteny

Comparing the annotated map of XAAA122 to the *D. melanogaster* genome seems to indicate that in general, synteny has been preserved. The Syn gene fragment, CG12814, and Best1 loci are all in the same order and direction in the *D. littoralis* clone and the *D. melanogaster* 3R chromosome. However, the location of the CG6995 gene at the end of the fosmid does not follow this same trend. In *D. melanogaster*, the CG6995 locus is located about 15 megabases away from the other three genes on the same chromosome and is also in a different direction than its homologous gene in *D. littoralis*. Two genes directly downstream from Best1 in *D. melanogaster*, CG6254 and 1(3)IX-14, are not present in the *D. littoralis* fosmid (Figure 9). This suggests that its current location next to the Best1 locus may potentially be due to a recombination event or paracentric chromosomal inversion of the 3R chromosome that occurred sometime during the evolution of the two species.

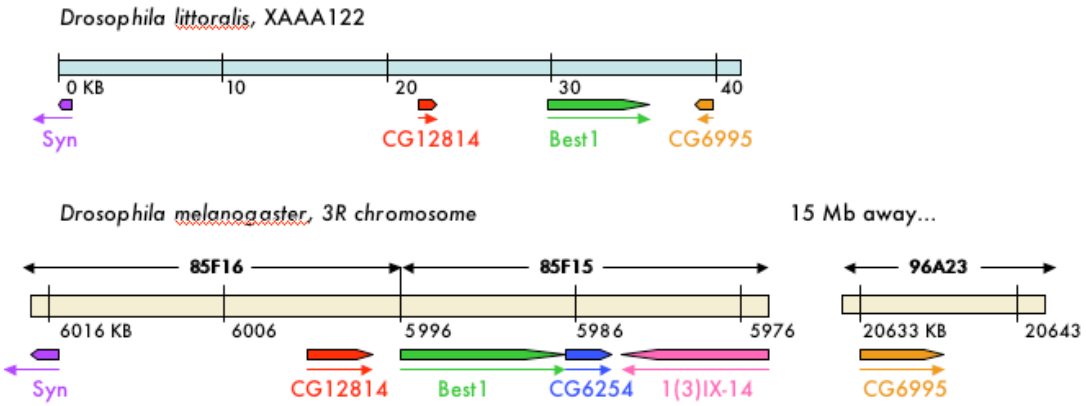


Figure 9. A comparison between XAAA122 and the homologous genomic region on *D. melanogaster* chromosome 3R. The gene boundaries annotated for *D. littoralis* do not include the flanking 5' and 3' untranslated regions, unlike the annotated *D. melanogaster* genes that include these untranslated regions.

The comparative 40 kb region in *D. melanogaster* (the region containing CG12814 and Best1) has the same gene density as the *D. littoralis* fosmid (four genes in a 40 kilobase portion), at 0.0973 genes/kb or 97.3 genes/Mb. The average gene density in the *D. melanogaster* genome is about one gene per 9 kb, but there is a wide variation in gene density that ranges from zero to almost 30 genes per 50 kb. The 3R arm of the third *D. melanogaster* chromosome is about 23% heterochromatic and 77% euchromatic (reviewed in Adams et al., 2000).

References

Sun, H., Tsunenari, T., Yau, K.W., and Nathans, J. (2002). The vitelliform macular dystrophy protein defines a new family of chloride channels. *Proc Natl Acad Sci USA* 99: 4008-13.

Adams, M.D. et al. (2000). The genome sequence of *Drosophila melanogaster*. *Science* 287: 2185-2195.

Appendix

What to put in the appendix for EACH protein:

1. *Fasta formatted file of the protein sequence*
2. *Fasta formatted file with nucleic acid sequence which codes for the protein (spliced mRNA transcript)*
3. *Fasta formatted file of the genomic region around the gene, including ~500 bases upstream and downstream of coding sequence (include in header the exact coordinates of the coding sequence found within the genomic sequence).*
4. *For new repeats and Clustal analysis, include any fasta files used in your analysis.*

Order in	Protein	Protein	Coding Sequence	Genomic Sequence

Appendix		Sequence		
1	Best1, CG6264	741 AA 81324 MW	Region in fosmid: 29745-35449 (35452, including stop)	Region in fosmid: 29250-36000 cds:(29745-29815; 31543-31653; 31714-32137; 32227-32559; 32622-32865; 33808-33992; 35094-35449)
2	CG12814	172 AA 19025 MW	Region in fosmid: 22192-22912 (22915, including stop)	Region in fosmid: 21700-23500 cds:(22192-22210; 22279-22473; 22543-22685; 22754-22912)
3	CG6995	389 AA 44254 MW	Region in fosmid: 37909 (37906, including stop) - 39075	Region in fosmid: 37400-39600 cds:(39075-37909)
4	Syn, CG3985 (partial protein)	68 AA 6634 MW	Region in fosmid: 2-205	Region in fosmid: 2-207 cds:(2-205)
5	Clustal Analysis: Protein Sequence Analysis of Best1 • D. littoralis, D. melanogaster, A. gambiae, C. elegans, M. musculus			
6	Clustal Analysis: Promoter Sequence Analysis of Best1 • D. littoralis, D. melanogaster, A. gambiae, C. elegans, M. musculus			

Bestrophin1 (Best1), CG6264 Fasta Files

```
> Best1     CG6264      protein sequence
SQ   SEQUENCE   741 AA;  81324 MW

MTITYTGEVA TCRGFGCFLK LLLRWRGSIY KLVWLDLLAF LTLYYLLNLV YRFALNESQK
ETFEAIVAYC NSYRDLIPLS FVLGFYVSIV MTRWWNQYTS IPWPDPIAVF VSSNVHGQDE
RGRVMRRTIM RYVCLCLTMV LANISPRVKK RFPGLNNLVE AGLLNENERT IIETMNKSFP
RPSKHWLPIV WAASIITRAR KEGRIRDDFA VKTIIDELNK FRGQCGLLIS YDTISVPLVY
TQVVTLAVYS YFLTCCMGQQ WTNGKVVGTS SYLNKVDLYF PVFTTLQFFF YMGWLKVAES
LINPFGEDDD DFEVNWMVDR NLQVSYLIVD EMHHDHPELL KDQYWDEVFP NELPYTVAAE
RFRENHPEPS TAKIDVPKNA AMTSGMSSVR IDEMGDDASG IHFSAGNGKS RLGLGSSPSL
VSVSGTLSRV NTVASALKRF LSRDDSRPSS ATPTEDPGKY RFPSSASSAS LTGTTGGPAA
VKPAGSMRIT QQVIEEVDEQ ATITSMRPND PRPNVKDIFA PATGQPQTAP LQPPPPATGA
ATPSRSQPVD IPRRSPQYER AHSQYDPTPS LFPPGGVDAL LSTSAPAGDS PLLMSAATAP
SSPVGDSSKS LFEQHKAASR ETVENLDIKT SNDILANTAA DAAPDVGDDF EKLKAAREKE
KLVRQQMNLA RTISTAPGVD PSGVVPLVPV VPVAMPMMAP TAPQPNTTTA LPAVASSADL
LAGGEPVTNS TTKSEDATPG S

> Best1     CG6264      coding sequence   29745-35452

atgacaataacatatacaggtgaagtggccacctgtcgtggctttggctgctttctcaaactgcttctcag
gtggcgcggaagtatttacaaattagtttggctggatctgctcgcttttcttacactatattatttgctga
atttggtctaccgctttgcgctcaatgaaagccaaaaagaaacgttcgaggcgatcgtggcctattgtaac
agctatcgtgatctcatcccgctgtcctttgtgctcggcttctatgtctcaattgtaatgacacgctggtg
gaaccagtacacatctattccttggccggatccaatcgctgtattcgtcagctccaatgtgcacggccagg
atgaacgcggtcgggttatgcgtcgcacaatcatgcgctatgtgtgcctgtgcctcacgatggttctggcg
aatatatcaccacgcgtcaagaaacgatttcctggcctaaacaatctcgtggaagcaggcttgctcaatga
gaacgagcgtacaatcattgagacaatgaacaaatcttttccgcgaccctccaagcactggctcccaattg
tgtgggcggccagcatcattacgcgggcaagaaaagagggccgcattagggacgatttcgccgtgaagacc
atcatcgatgaactaaacaagtttcgaggacagtgcggactgcttattagttatgatacgatcagtgtgcc
cctggtttacacacaggtggttactttggctgtttattcctatttcctcacctgctgcatggggcagcagt
ggacgaatggcaaagtcgtgggcaccagtagctatctgaacaaggtggatttgtattttccagtgtttacc
acgctgcagttctttttctacatgggctggctcaaggtggctgagtctctgatcaatccctttggcgagga
tgatgatgattttgaagttaattggatggtggatcgcaatctgcaagtctcctatttgattgtagatgaaa
tgcatcatgatcatccagaattacttaaggatcagtactgggacgaggtatttcccaatgagctaccctat
```

actgttgctgcagaacggtttcgcgagaatcatccggagccttccacagccaagatcgatgtaccgaagaa
tgcggccatgacctcagggatgtcatctgtgcgcatcgatgaaatgggcgacgacgccagcggcatacatt
tctcagcaggcaatggcaagtcgcggctggggctaggctcctcgcccagtctggttagtgtatcgggtaca
ctgtcccgggtaaatacggtggcgtcggcactgaaacgtttcctaagtcgcgacgacagccgtccaagctc
tgccacgcccaccgaagatccgggcaagtatcgcttcccgagcagcgccagctcggcaagtttaacgggca
ccactggcggcccagctgctgtgaagcccgcaggcagcatgcgcataacccaacaggtcattgaggaggtg
gacgagcaggccaccattacgtcgatgcgtcccaacgatccgcgtcccaatgtaaaggatatatttgcgcc
ggccactggtcagccgcaaacagcgccactgcagccgccaccgcctgcaacaggagcagcaacaccgtcac
gctcccagccagtggacatacccagacgctcgccgcagtacgagcgcgcccactcgcagtatgatccgacg
ccatcactctttccgcctggtggcgtggatgctctgctgagcacatcagccccagccggagacagtccctt
actcatgtcagcagccacggcgcccagttcgccggtgggcgacagcagcaagtcgctgttcgagcagcaca
aggccgccagccgtgagacggtcgaaaatcttgacatcaaaacatcgaacgatatactggccaacacagct
gccgatgctgctccggatgtcggcgatgattttgagaagctgaaagctgctcgggaaaaggagaagctggt
gcgccagcaaatgaacctggcccgtaccattagcactgccccgggtgtggatccctccggagttgttccgc
tggtaccagttgtgcccgtggccatgccaatgatggcgcccactgcgccccagccgaacacaacgaccgcg
ctcccggctgttgcttccagcgccgatttactcgccggcggcgaaccagtgaccaattcaaccactaaatc
cgaggatgccaccccaggaagttga

> Best1 CG6264 genomic sequence 29250-36000
cds:(29745-29815;31543-31653;31714-32137;32227-32559;32622-32865;33808-33992;35094-35449)

tgcaattaaattaatcagctcgtcactgttgcatttgaaatgttaaatagaatgcaatccatgcagttcaa
ttgtattccgcattttaatgtcgcatgcttacatgacttccagttgaaaatatcgatttttaaatttcaaa
tatattattatatttgttgttgttgctgccccggcttagctggtaattatctgcacattgcgcagattaca
cgtcagttacaataaaaatgaattattgactcattatccttattattattatcattattgatcaaatatgt
ttgtgttgtttctctaaatttaagaaatttatcacctcgtaaactgatgacattggctatgccccaattaa
aacaaagcaaaatgtacatttatctaaatcgcgaagcgtgataagaaaacataactgactttaagaacaaa
accaacattggttataacaaaaatctcaattgacaatataaaggagaaaattaaatacaacttgcgaaaat
gacaataacatatacaggtgaagtggccacctgtcgtggctttggctgctttctcaaactgcttctcaggt
aagtgcgagattttattctctatatattaatatatgtacttaacttatgtacagaactgtttgttctgacc
cagcccttatagcgaggaaaatgatgtgatgcaagtgcacgctaagacggaaaattttgaaaaatatacaa
ataaataatttttaaacagcggctatataaacctctttcgtataccactaacaggcctaaaacttaatttc
aacaagaaaattattttcattctcagaaattcacttttttgggggaaagctttagtttccttaaatcaga
aaggtacatttttcattttacaatataagtatatttatcgattccgaatatctactacggtagttggaacc
ggaaaatttgacttccggcaggcggaaaatatatatataatataataggtttcaatacttgtatatatata
catgtaggtgtatcaaaactatttttgcactttcaaattacagtgctaaatgcttgtgtatgaggagcgtg
taaaattgcaagtctggtgctaattaaagaactgacacagttcgacagagccagtaaattgagcgactctc
cgctggctacagctagtgatagtcgtcagacatcacacaattgcaactaatgcaagtaatagaattttcca
tgcgttccacacacacacacacacacacacacacacacacacactcaaattgttgtgcacactagattgtagt
tgagccgatatcagcaacaactccacccacacatgatactcgtgtcagcacatccacctcgtctctattgg
acctgtagccttatcaaagccgtttgggacgtcatgtgtttgacatctgtcatgcatgtaatttgaattat
aatttaatattaggtagccagtgcttctaacgatatctgtttgaatgtatgtatgagatatctgcaccccc
acttgtgtaaatgtagttcacacttatttgcatgatccgctgatagcgtcagctgcattgattgttacggc
taatcaatgaactggctgcccgagatttgatagtgtaactgaagttctgtgttattgctaatttaacacat
tatcatataaatttagtataggagagctaatgcacataagtagaagccaataagcagattgtactttgaaa
ctgtttccaaatttaaactaattttaataacattcaaaacaagaactgaggcacggcatctgtgcatataa
taaatttaattaaaattatcaacgtcggcaaaccttgaaatctttctggtaacgatacaaagtcaaattct
gtttgtcatttgtttatcataaatgattgtgtttgttttttttttttttttgctattttatcaaaaatt
ttggtgataagagcgtttttataattagtaggctttcatttgaaatgatggtgaggttattcacatatac
atacgtatgtatatgtgtgtgtgcttatttttatgacgtttcgtcagccacaagtggacggtggcatgatt
attattgttcaatatcaggatggtaaggctttgtgatgacgtcgccaattataggaatattgctgaacaca
tgccagacgcggtgccgcacattaataaaattgtccgtcgtatacgataagtttagggtatagcccaaacc
ataaacgatgatatgttagtcgatctactatattagggtttttcgtgtgcgttctatttataataatgatc
tgatatccaattccctttcaggtggcgcggaagtatttacaaattagtttggctggatctgctcgcttttc
ttacactatattatttgctgaatttggtctaccgctttgcgctcaatgaaagccaaaaagagtaagtgcat
atttataacgaatatagtccatattaaatgtttctctttttttcgctcagaacgttcgaggcgatcgtggc

ctattgtaacagctatcgtgatctcatcccgctgtcctttgtgctcggcttctatgtctcaattgtaatga
cacgctggtggaaccagtacacatctattccttggccggatccaatcgctgtattcgtcagctccaatgtg
cacggccaggatgaacgcggtcgggttatgcgtcgcacaatcatgcgctatgtgtgcctgtgcctcacgat
ggttctggcgaatatatcaccacgcgtcaagaaacgatttcctggcctaaacaatctcgtggaagcaggct
tgctcaatgagaacgagcgtacaatcattgagacaatgaacaaatcttttccgcgaccctccaagcactgg
ctcccaattgtgtgggcggccagcatcattacgcgggcaagaaaagaggtaaataatactaaatataaaca
agagcactattatgcggctttgttcctttgtttgggctacaatatgaaaagtacaaattttgacagggccg
cattagggacgatttcgccgtgaagaccatcatcgatgaactaaacaagtttcgaggacagtgcggactgc
ttattagttatgatacgatcagtgtgcccctggtttacacacaggtggttactttggctgtttattcctat
ttcctcacctgctgcatggggcagcagtggacgaatggcaaagtcgtgggcaccagtagctatctgaacaa
ggtggatttgtattttccagtgtttaccacgctgcagttctttttctacatgggctggctcaaggtggctg
agtctctgatcaatccctttggcgaggatgatgatgattttgaagtacgtctgaacaatggtagcttttaa
agtgttaaacttattcataagtgtctgcttcgtaggttaattggatggtggatcgcaatctgcaagtctcc
tatttgattgtagatgaaatgcatcatgatcatccagaattacttaaggatcagtactgggacgaggtatt
tcccaatgagctaccctatactgttgctgcagaacggtttcgcgagaatcatccggagccttccacagcca
agatcgatgtaccgaagaatgcggccatgacctcagggatgtcatctgtgcgcatcgatgaaatggttggt
aataatagcaccggcgcctctgggtctaagccgccttcagctgaacagccacaaggcgcacctatgcagac
agttacttatgacctttccaaagctgcgccagatgaggaggttagaatcaaaacgttggcagcattcacta
gcgatacgaagtattggcaccaaataatcggctaccttaactaaacgtacatccatgaacatcactcatgt
acatgtggcggaaaatcaaaatcaatcaaaaatcagatcaaaacgtattttctaaccttcttctatcctaa
actgaaaaacaattccatatgggattcagatctatcattggcatttcaaatcacaacggtgcttggtattt
atacctaaccttatatatatttatgatataaaatgaattggatgttgaacagtggattctaacataagcat
tggagatcgttctcatgcacctgtcaaagttgtcagctagacgtgaaaactacacatactttaacttctag
tcaattaacatcaatgttgacaagagcatgagaaaacagatctgaaccaatatctcttaatgtaatcggta
taacagagatcggtctactaaataatggacgtaaaataaacttttagaggctcgccgaaatgtacatata
gaaaccactgcgcattatcactaatctaactaaactaattatactgtactctggtgtttgtgtctctttaa
actataagactgcatactaacttttgattctaacaataatctgtttgcctacagcctctaactgcactacg
cactgcacggctttcagttgcttttgtgcaccccagaaaaaaaactgcacaaatcggtttcaagtaactta
tgctcactttgaactctccacatgaacagcagagtgcatttaactaatattctacattgtttctttcgtac
aatgcaactcgcagggcgacgacgccagcggcatacatttctcagcaggcaatggcaagtcgcggctgggg
ctaggctcctcgcccagtctggttagtgtatcgggtacactgtcccgggtaaatacggtggcgtcggcact
gaaacgtttcctaagtcgcgacgacagccgtccaagctctgccacgcccaccgaagatccgggcaagtatc
gcttcccgagcagcgccagctcggcaagtttaacgggcaccactggcggcccagctgctgtgaagcccgca
ggcagcatgcgcataacccaacaggtcattgaggaggtggacgagcaggccaccattacgtcgatgcgtcc
caacgatccgcgtcccaatgtaaaggatatatttgcgccggccactggtcagccgcaaacagcgccactgc
agccgccaccgcctgcaacaggagcagcaacaccgtcacgctcccagccagtggacatacccagacgctcg
ccgcagtacgagcgcgcccactcgcagtatgatccgacgccatcactctttccgcctggtggcgtggatgc
tctgctgagcacatcagccccagccggagacagtcccttactcatgtcagcagccacggcgcccagttcgc
cggtgggcgacagcagcaagtcgctgttcgagcagcacaaggccgccagccgtgagacgggtaagtggtca
gcacttgtgtggggaatggtaccaataatgccttcctcttcctgcttccgataatcacatttaacatttat
cttcatatacactcacatacatatacacattcgacgaatgcgcgtttaaaactgttaatacttaaagcgat
tgagagacccgatgcagcgcagattttcaggtggttcgtgacgcccggtaaagttcaatcattggccaaaa
taatttatttcgttcctcagcttattccacgtaaatgttgctagtcctcaaattgttattgttgcgtgatg
tcgtttaatattttaagttttcttaatatcataattcagttttttaatcgaaaaatatgttgacttgttt
attaaatactaaccagagcttagcaatttttttgaattggtagtttgtattgtgttgcataagacccgatc
caacagttgtgggttttcaaccaggtaatgccctcctaactccaaatttgtatttagagtttgcctagccg
taatccttctagccgttgcacgcatttcttaatcttgattaatcccaaatgccacatagttaaataagcta
acgcgactttctgcccttacagtcgaaaatcttgacatcaaaacatcgaacgatatactggccaacacagc
tgccgatgctgctccggatgtcggcgatgattttgagaagctgaaagctgctcgggaaaaggagaagctgg
tgcgccagcaaatgaacctggcccgtaccattagcactgccccgggtgtggatccctccggagttgttccg
ctggtaccagttgtgcccgtggccatgccaatgatggcgcccactgcgccccagccgaacacaacgaccgc
gctcccggctgttgcttccagcgccgatttactcgccggcggcgaaccagtgaccaattcaaccactaaat
ccgaggatgccaccccaggaagttgagctggctggcgtttagtattcaggcatacagtttaaagtatgcca
cttgtgttgtacaacttttcaaataagataagttaattatttagcgtattagttaatttatataagctact
gaacttaatgccttaacatgcagaatgacagaaacagattaagaatgctcataacgttaaaacgatcttga
gtgagaatgtacccatttaacgcagtttaaatacctattatttaaaatgatgtcaaaagtgatttgaatct
ttacaataatataagttttctaaattgccagaaaattgtgttttaattagagtaaaactactcttgggctt

gtgcctgtcctttcaagaaatgcgtccgcccaatcgaggcaaaatagaaatgggttggtttaaataagcaa
catgtgttataggcaattaacaattatctagtttgagtgttcgactagaagatatgattttatgcctcttc
ttaatgcttatataccccttttgtagaaaatgctttcacacttagcttcggtttttactgtgtataaaatta
tttcta

CG12814 Fasta Files

> CG12814 protein sequence
SEQUENCE 172 AA; 19025 MW

MFKFPEGSEV HLQCDVVQCY GRCVEIDDCS EVALAGFTKG NGNGPRKFGP NEEGSSIAGT
TVFVLDPAEA RLISPNCEDG IRPSWLLWLA ITLGVLFLIM LLMNIFLCTA MSCSCANTEI
IEKEPSIIEE YDPYRSWHGS QYGSRYSLHG RDAHKGYTSG GSTIHSNRSD RY

> CG12814 coding sequence 22192-22915

atgttcaagtttcccgagggatccgaggttcatctccagtgtgacgtcgttcagtgctatgggcgctgcgt
cgaaattgacgattgcagcgaggtggccctggctggtttcacaaagggcaatggtaatggaccacgcaagt
ttggacccaatgaggaaggttccagcattgcgggaacaactgttttcgttttggatcccgccgaagcgcga
ctaatttcgccgaattgcgaggatggcatacgacccagctggctcctgtggctggccataaccttgggcgt
actgttcctcatcatgttgctaatgaacatttccttgcactgccatgagttgcagttgtgccaacacag
agatcattgagaaggagccctccatcatcgaggaatacgatccttaccgcagctggcatggcagccaatat
ggctcccgttactcgctacatgggcgtgatgcgcacaagggttataccagcggtggctcgacgatacactc
aaataggtctgatagatattaa

> CG12814 genomic sequence 21700-23500
cds:(22192-22210;22279-22473;22543-22685;22754-22912)

cttttaattagcaggtgcagctgcgcaatgggggaagtatgcctagaaaaccgttaactaatgtgccactt
ccattcccattcaaatttacagcgatgacaatgcccatgttgtgctcaagtttctggagaacgatcaccgc
gtgcgcgaaacagtatatggacacgagtataaaatacgcgctgagttctccaagccaaacggtgggtaatc
acaatgtcccttatatggagtgtgtgctaataaagattccctccaattatcagatacctatggtctacgtg
tgggcaactgctttgcctttgacaagaagaacatcaccctcaagctgacagatgatcgcgggtaagtgcga
ctgactcttcaataagctcctgagctgacagtcaactaacggtatatttcatgcacagctgcccctatgat
gccaatattatgagccgttttgtgcccactgccgatgggcgtgcagccgaggctacgctagcctccatgtt
caagtttcccgagggtaggtagtcgaaatcagcttatgctattcaaatcgaaaaccgaaataattccaaat
tgccttctcaggatccgaggttcatctccagtgtgacgtcgttcagtgctatgggcgctgcgtcgaaattg
acgattgcagcgaggtggccctggctggtttcacaaagggcaatggtaatggaccacgcaagtttggaccc
aatgaggaaggttccagcattgcgggaacaactgttttcgttttggatcccgccgaagcgcgacgtaagat
agaatagaaaaatttaaatctgaaatcaatcgaaatgacaatacaatttctcaactttgcagtaatttcgc
cgaattgcgaggatggcatacgacccagctggctcctgtggctggccataaccttgggcgtactgttcctc
atcatgttgctaatgaacattttcctttgcactgccatgagttgcagttgtgccaacacagaggtaattat
gctctaaaaacattacgaatgcaagtctgttatacgtgtgacctgtactactcattgcagatcattgagaa
ggagccctccatcatcgaggaatacgatccttaccgcagctggcatggcagccaatatggctcccgttact
cgctacatgggcgtgatgcgcacaagggttataccagcggtggctcgacgatacactcaaataggtctgat
agatattaagctgatctataaactacacacacaaacacacaagaacacaaaagctctactctacacttgat
atgcatctactctgctcttaaacatttttttctggtagctgtggagccaacacagacacacacaaacaca
ccgacatatatatatgtctatatatatatatatagatgtatctttatataacgcatgcccaaatagccttg
tatcgtatctatgtatcgtattttgttgtattcgtatctattgttaattctactgaagtgccctgaagccc
tgttggagaacaactacaacaaagcaacacaacttgtgttcacaaaaacaaaagacataaaagaatggaat
gaatttatttcgtgccttgtattgcaactgtaattaaccgttgataatttgcatttgttttaaatgatgac
taactttaaacctaaactaaacctatcgaatagaacgtacaattttctctggcccatacccgaaaacacta
acagtccactttaaagatttgctaactaaactgtcaaatccttttttttaaatacaccacatattataattt
gataatgattttgtatgaaacttatc

CG6995 Fasta Files

> CG6995 protein sequence
SEQUENCE 389 AA; 44254 MW

MELAEKTSTT AAGAASASGA AANKSGSGGG AVAAGASSSA VGKQALLSRN LWVSGLSTLT
RASDLKAIFS KFGKVIGAKV VTNTRTPGTR CYGYVTMSSS GDASRCIENL HRTELHGRII
SVERTKNEIG GSASSSKDAG NKGNNKANSK LNNESKKKDD DKKSGAKGSG DDKKGGETSG
EGKNSGGESS NKRDGKDKER NKPSRDDKDK PSTSSDKRHD RDHRSGTGHR SQNRDESNER
RSGRQPRDVD REIQQRQRER ERRQREMLSY QKIREERERQ RLRERERELR EEERRRREAR
ERQRIEEERL AEERRKLAIE RERLEREKAE LLRMERERQK LEREKIELER LELKRQQMKY
DKRKRVAHTT LKTTTPYLFF HILKRHLAV

> CG6995 coding sequence 37906-39075

ttacacggccaaatgccttttgagaatatgaaaaaataaatatggtgttgttgtttttagcgttgtgtgtg
caacgcgtttacgtttatcatacttcatctgttgacgcttcaattcgaggcgctccagctcgattttctca
cgctccagtttctggcgctcccgttccatgcgcagcagttcggccttctcacgctcaaggcgctcccgttc
gatggccagcttacgtcgctcctcggccagacgttcctcctcaatacgctgacgttcgcgtgcctcacgac
gtcgtcgctcctcctcgcgtagctcacgctcccgttcacgcaaacgctggcgctcccgctcctcacgaatc
ttctgataagatagcatctcacgctggcgtcgctcgcgctcacgctgacgctgctgaatctcacggtcaac
gtcgcgtggttggcgtcccgaacggcgctcgttggattcatcacggttttgagacctatgacccgtgccag
agcgatgatcacgatcgtggcgcttgtcactgctcgtcgacggcttgtccttgtcgtcgcgtgatggcttg
ttgcgctccttgtccttgccgtcccgcttgttgctggattcaccgcctgagtttttgccctccccgctagt
ttcacccccccttcttatcatctccactgcccttagcgccggacttttttatcgtcatccttcttcttcgact
cattgtttaacttgctgttcgccttgttgttgcccttgttaccagcatctttggatgaacttgcgctgccg
ccgatttcattcttggtgcgttccactgatataattcggccgtgcagctccgtacgatgtaaattctcaat
gcaccgcgaagcatccccagatgaggacatggtcacatagccgtaacagcgtgtgcctggggtacgggtgt
tggtgaccaccttggctccaatcaccttgccgaacttggagaaaatggccttcaagtcactggcgcgtgtc
aatgtagacaaaccagacacccaaagattacgtgacagcaacgcctgcttcccaacagcactagacgaagc
accagccgcaactgcgcctccgcccgagcccgacttgttggcagcggcacccgatgccgatgctgcgcccg
ctgctgttgtcgatgttttctctgcaagttccat

> CG6995 genomic sequence 37400-39600
cds:(39075-37909)

cataatcatcatgcttcttgtcataagtactggtactgttaaaaatacagttgagtaaatatatgtctaat
acacattatgaattatgtggccaggccatcacttgaatagataacatacttgcgtgaactgaccaaagaag
catcaaagcgcggcggtggtggagcatcaaagacagcggaacgcttgcggtctgtaacatcggtatagcga
tcatcagtacgcttacttagcgaacgcttaacaggatcatcacgtgattccataatcctggagtaccatta
gtggaaagtgaggaaagatgaaaagagcaaagtttacaacattagcaacgtttaaaagataggcgacgtat
tcggtacatatatgttaaacatagtttaaactacgatgcagttcatttgcagttgcataaatatacatgaa
tgcaaaagcatttgcctagactgcccggcagtttttcgggttgaagaaacaacgtggctgctgcaacttaa
tacactaatttacacggccaaatgccttttgagaatatgaaaaaataaatatggtgttgttgtttttagcg
ttgtgtgtgcaacgcgtttacgtttatcatacttcatctgttgacgcttcaattcgaggcgctccagctcg
attttctcacgctccagtttctggcgctcccgttccatgcgcagcagttcggccttctcacgctcaaggcg
ctcccgttcgatggccagcttacgtcgctcctcggccagacgttcctcctcaatacgctgacgttcgcgtg
cctcacgacgtcgtcgctcctcctcgcgtagctcacgctcccgttcacgcaaacgctggcgctcccgctcc
tcacgaatcttctgataagatagcatctcacgctggcgtcgctcgcgctcacgctgacgctgctgaatctc
acggtcaacgtcgcgtggttggcgtcccgaacggcgctcgttggattcatcacggttttgagacctatgac
ccgtgccagagcgatgatcacgatcgtggcgcttgtcactgctcgtcgacggcttgtccttgtcgtcgcgt
gatggcttgttgcgctccttgtccttgccgtcccgcttgttgctggattcaccgcctgagtttttgccctc
cccgctagtttcaccccccttcttatcatctccactgcccttagcgccggacttttttatcgtcatccttct
tcttcgactcattgtttaacttgctgttcgccttgttgttgcccttgttaccagcatctttggatgaactt
gcgctgccgccgatttcattcttggtgcgttccactgatataattcggccgtgcagctccgtacgatgtaa

attctcaatgcaccgcgaagcatccccagatgaggacatggtcacatagccgtaacagcgtgtgcctgggg
tacgggtgttggtgaccaccttggctccaatcaccttgccgaacttggagaaaatggccttcaagtcactg
gcgcgtgtcaatgtagacaaaccagacacccaaagattacgtgacagcaacgcctgcttcccaacagcact
agacgaagcaccagccgcaactgcgcctccgcccgagcccgacttgttggcagcggcacccgatgccgatg
ctgcgcccgctgctgttgtcgatgttttctctgcaagttccatgtaaataataacaatattggaagagagt
tatggtcagagctgtgagtggctggaaaactgttttggtatttggtattgattgtgatcattgatcattaa
gggaagcactttaaacgcacactttgtagctgctgctgcttgctgttgtatcgtaatttttcgacataacg
aacctgtgactgattctacattataagtacatgtacatcattatttactgcgccctacgaaagctaagaga
gtgtgaatttgattgggttgctgctgatgctcccactgccgcttcttcgcatgttgcttccaccattaaaa
gggatctcgtacagtgtggccatgccaacaagtcacagtggcatacgttaagataccgtcagctgtcgtgt
atgaagggtcgtcttctggcgggaatcgtcttatccgtcatcatcgttgtggttccgctgaatattacaag
ctgcagctcaaacacgggtccgtccttcctagcttattgtgtatttcgggcgatgatatggtgtccgtaat

Synapsin (Syn) Fasta Files

> Synapsin CG3985   partial protein sequence
SEQUENCE   68 AA; 6634 MW

MPPPPAPGQL PPQPAGAAPE LSLSFGAGKA PASAAPAPPR GVSAPTSPAK SRESLLQRVQ
SLTGAARD

> Synapsin CG3985   partial coding sequence   2-205

atcacgcgctgcgcctgttaaactctgcacccgctgcaagagactctctcgcgacttggctgggctggtgg
gcgcgcttacgcctcgcgggggcgctggagccgccgaagccggtgccttgccagcgccgaagcttagcgac
agctccggagctgcgccggcaggctgcggcggcaattggcccggtgccggcggtggcggcatgctcggcgg
accgccggcaacatacggcttagttggctggtcctgtataggacgcgcagccggcggcaagctattgggat
caacatcatcgacctcggaagacaagtctcccgagctgaatctgcacgtgttgcgtggatgtggattagat
gtttgtgttgcgtgtatggtgtggttgttgcgtgggtttgttgtggttttcgccagggcgtttcaaatgcg
aaccgtggaacaagtagacaagtgaaaggtgcagcgcatttatgtgagagagcaaaagaaatagaggaaga
gagagagagagagagagagagagaaaagcgcaccatttgcgggggggggagagaaaatagacaaacatatga
aatctcaaattagtcttgtcctgaactgaagccccctccccgctaaggaaggggggcgtttccttttaaggt
ctttaccttctcttcaggaaattaacattggaagtgaagcttgatttaaacgatgagaaattcag

Clustal Analysis of Best1 Protein Sequence: Fasta Files

>Littoralis
MTITYTGEVA TCRGFGCFLK LLLRWRGSIY KLVWLDLLAF LTLYYLLNLV YRFALNESQK
ETFEAIVAYC NSYRDLIPLS FVLGFYVSIV MTRWWNQYTS IPWPDPIAVF VSSNVHGQDE
RGRVMRRTIM RYVCLCLTMV LANISPRVKK RFPGLNNLVE AGLLNENERT IIETMNKSFP
RPSKHWLPIV WAASIITRAR KEGRIRDDFA VKTIIDELNK FRGQCGLLIS YDTISVPLVY
TQVVTLAVYS YFLTCCMGQQ WTNGKVVGTS SYLNKVDLYF PVFTTLQFFF YMGWLKVAES
LINPFGEDDD DFEVNWMVDR NLQVSYLIVD EMHHDHPELL KDQYWDEVFP NELPYTVAAE
RFRENHPEPS TAKIDVPKNA AMTSGMSSVR IDEMGDDASG IHFSAGNGKS RLGLGSSPSL
VSVSGTLSRV NTVASALKRF LSRDDSRPSS ATPTEDPGKY RFPSSASSAS LTGTTGGPAA
VKPAGSMRIT QQVIEEVDEQ ATITSMRPND PRPNVKDIFA PATGQPQTAP LQPPPPATGA
ATPSRSQPVD IPRRSPQYER AHSQYDPTPS LFPPGGVDAL LSTSAPAGDS PLLMSAATAP

SSPVGDSSKS LFEQHKAASR ETVENLDIKT SNDILANTAA DAAPDVGDDF EKLKAAREKE
KLVRQQMNLA RTISTAPGVD PSGVVPLVPV VPVAMPMMAP TAPQPNTTTA LPAVASSADL
LAGGEPVTNS TTKSEDATPG S

>Melanogaster
MTITYTGEVA TCRGFGCFLK LLLRWRGSIY KLVWLDLLAF LTIYYAINMV YRFGLNPAQK
ETFEAIVQYC DSYRELIPLS FVLGFYVSIV MTRWWNQYTS IPWPDPIAVF VSSNVHGQDE
RGRMMRRTIM RYVCLCLTMV LANVSPRVKK RFPGLNNLVE AGLLNDNEKT IIETMNKAFP
RPSKHWLPIV WAASIITRAR KEGRIRDDFA VKTIIDELNK FRGQCGLLIS YDTISVPLVY
TQVVTLAVYS YFLTCCMGQQ WTDGKVVGNT TYLNKVDLYF PVFTTLQFFF YMGWLKVAES
LINPFGEDDD DFEVNWMVDR NLQVSYLIVD EMHHDHPELL KDQYWDEVFP NELPYTIAAE
RFRENHPEPS TAKIEVPKNA AMPSTMSSVR IDEMADDASG IHFSAGNGKM RLDSSPSLVS
VSGTLSRVNT VASALKRFLS RDDSRPGSAT PSQDQPYKFP ASASSASLSG AVVGSATSAG
KPAGSLRITQ QVIEEVDEQA TITSMRANDP RPNVMDIFAQ TSSGAGTSGP LQPPPAHSEP
VDIPSRPPSY NRAQSQYEPN LFPPGGVDAL LSTSAPAGGS PLLLSNAATA PSSPVGESSK
SLYDPQKGAS RETVESMDLR SSTDLLGDAA VQPEDEGDDF DKLKAEREKE KLMRQQKNLA
RTISTAPGME ATAVPMVPMV PVNVAVQQAQ LQPVASSADL LAGGDQFSNS TMKSEDAING
S

>Anopheles gambiae
MTVTYTAEVA TCRGFGCFLK LLVRWRGSIY KLVWLDLTCF LLLYYVLNIT YRFGLSEDQK
RIFEEIVKYC ATYSNLIPLS FVLGFYVTIV MTRWWNQYTS IPWPDPIAVF VSANIHGQDE
RGRVMRRTIM RYVCLCLTMV LTNISPRVKK RFPTIEHLVE AGLLNENEHN IMSHLNQKFP
RHSKHWLPIV WAASIVTRAR KEGRCRDDFA SKTIIDELNK FRGQCGLLIS YDTISVPLVY
TQVVTLAVYS YFLTSVMGQQ WVETKELGEG VVNKIDQYFP IFTTLQFFFY MGWLKVAESL
INPFGEDDDD FEVNWMVDRN LQVSYLIVDE MHHEHPELLK DQYWDEIFPT ELPHTVASAE
IREEPPQPST ADIVVKKREA EVIPVVPTAV RIDEMVGKQN QPVEYNFPAN PNDDEMDDNV
SGIHFNTSDK PIYRGGSSAT SLASVAGTMS RVNTVTSALK RFFSKDESSR PNSATPDAAG
SVPFKFPGSA SSNNLAGGRI PDRAMGSMRI TDQVIEEVDE QLTITSMQGN NDSRPTAASL
FEHGPPKPSE PVDVPQQAYN RTQSSVAHGF HHSQLTPPCG GDAMLSASAP ETGRFEHNYV
PLSPTDPDAV PMVASSTALQ GMMDETDRVT TTFNSEETEK EQLERKTRLL ARSISKQAVL
GLIDTDMLLD ELTRSTGDLE GSE

>Caenorhabditis elegans
MTVAYSLDVA TSGLFTQIKV LLRWKGSVWK SIWSELLIWL LCYSILSVIY RILLNKAQRE
VFEQLCTFFD TFSVFIPVTF MLGFYVSIVY NRWTKVFDNV GWIDTSALTI AQYIRGTSEK
ARLIRRNCVR YMIVAQVMVF RDVSPAIRRR FPTIKHLIGA GLLTEDELTE FDAITSPQSK
YWQPIQWLFS LVTVAKDEGL IADSYLYVDL IDKMREFRTK ILNLVIFDMV PIPLVYTQVV
NLAVRTYFVL ALFGRQFLEN NNNIPGAKWK IDIYFPIMTS LQIVFIVGWL KVSEVMLNPL
GEDDEDFETN WIIERNLQVG YAVVDQAYGR YPTLKRDPFW EDETPQTMDT PTSTRKAHSH
MQGSCINMNE ADLDNGLISY VRRRSRSFGD DDASSFSTEN TRTQSNASML PRHMWQQPRN
KISAAIAKFY KSEAEPRRAS VCVAALNLNH RRPSTDLREI EPIENTDDVV INVPALVQER
IRIAKQQAVK IEIPEEPLKI TASAEGKSAI SPSANNVKWF VEEMPVIEEE EERLHRKTPR
NLSASLCASL ASAYIFS

>Mus musculus
MTITYTNKVA NARLGSFSSL LLCWRGSIYK LLYGEFLVFI FLYYSIRGLY RMVLSSDQQL
LFEKLALYCD SYIQLIPISF VLGFYVTLVV SRWWSQYENL PWPDRLMIQV SSFVEGKDEE
GRLLRRTLIR YAILGQVLIL RSISTSVYKR FPTLHHLVLA GFMTHGEHKQ LQKLGLPHNT
FWVPWVWFAN LSMKAYLGGR IRDTVLLQSL MNEVCTLRTQ CGQLYAYDWI SIPLVYTQVV
TVAVYSFFLA CLIGRQFLNP NKDYPGHEMD LVVPVFTILQ FLFYMGWLKV AEQLINPFGE
DDDDFETNWI IDRNLQVSLL SVDGMHQNLP PMERDMYWNE AAPQPPYTAA SARSRRHSFM
GSTFNISLKK EDLELWSKEE ADTDKKESGY SSTIGCFLGL QPKNYHLPLK DLKTKLLCSK
NPLLEGQCKD ANQKNQKDVW KFKGLDFLKC VPRFKRRGSH CGPQAPSSHP TEQSAPSSSD
TGDGPSTDYQ EICHMKKKTV EFNLNIPESP TEHLQQRRLD QMSTNIQALM KEHAESYPYR
DEAGTKPVLY E

Clustal Analysis of Best1 Promoter Sequence: Fasta Files

> Littoralis

ctttctacatttctatttagaaataaactaggaaaaaaaagaggaaaatataagaatatagctatcgatt
tattatttaattatattatttataaggcatattaaatagggcatatttattcagcattgtgataaatctcg
tgtcgcaattcatcttcttggaagtctctttgtttacttattacgaaacggctcaagcacacagaagaccc
aaatgactcatttgcttatcgaatacatattcattatactcgttcgcttaaaattcgtacaactaaatact
ctaaatactcactaattcacagttagttgatcacgctgccaaatcgaactgcacgcacgtcgtgatccagt
taaacacatatcagaagccttgccccgtaatcagcagccaaattcagcatataatcagcaagctacgtcaa
gattttgtttttgtatgatttacgattttaagtgctaaaaaaaaaaaaaacaaaaaaaaaacatggctgta
ttggtaaaggtaaacaggcatatatacatatgaaaatactaatgtatgtatacaaaaagtctaccatttt
tgtgcggccaatttcatctgtaaaacaagccaagaaatgatgtcatgatccagacccaaaagcctggctca
cactcacacacataccaaacgtgcacacataccacgaatactgtagacaattttatttgtgctctgagctc
tgtttgtattgtgtatttcgatcattagctcaccggctcattcgccgaatgccatttatttatttattctc
tgctatacggaaattcatcaacgtactgctcattgaatacagacttttcccactcgaaaggggagacacat
gagctcatctttactcattattaaatacactatttatgcatttctctttacatacactctacataacatac
tcatgtgtccaagtacataaataccaataaaaataagagtaatggtcttttagtccggagtgtcgccttgc
taataccctctacccagaacccagatgccgtagtaataattggaatttcatgtttgttgttgttgttcttt
taacgtatgagaaatgcatgttagaataacttcggtatttcaatctcattgaaagctgtttgattgaaaga
aactagatccgcactggctctatatgtaattctctgaggtcaacagacaaacagtccagaaatatatccaa
cttatgagttccacttgcatattaaacgataatgcccttaaaaactaacgactttattttacgagacttga
agtgactttgtgaaatttgtttgctgaactaattaatgtgaactcaaatgcggctgaccctcagtgacacc
tccagctggaactgttgcagccttaatggcatcgcccacacgcgatctgccagcgcttcgatttgcgtcaa
ttcctactaaagcaacacacatacattgatatatttggcttgagctcggagcttctctaaagtttccataa
atttttaaattgctgcaattaaattaatcagctcgtcactgttgcatttgaaatgttaaatagaatgcaa
tccatgcagttcaattgtattccgcattttaatgtcgcatgcttacatgacttccagttgaaaatatcgat
ttttaaatttcaaatatattattatatttgttgttgttgctgccccggcttagctggtaattatctgcaca
ttgcgcagattacacgtcagttacaataaaaatgaattattgactcattatccttattattattatcatta
ttgatcaaatatgtttgtgttgtttctctaaatttaagaaatttatcacctcgtaaactgatgacattggc
tatgccccaattaaaacaaagcaaaatgtacatttatctaaatcgcgaagcgtgataagaaaacataactg
actttaagaacaaaaccaacattggttataacaaaaatctcaattgacaatataaaggagaaaattaaata
caacttgcgaaaatgacaataacatatacaggtgaagt

> Melanogaster

atttagccacgcgcctttctggcccactgacaactctcgccagccaagtgttcgtcgaaaaagtcattaac
ctcggtgatttggattcgaatgttgtcgtcagtctagaatcagtgtaacaaatattgaatgttatgctaat
gggaatgataataatggtggccaagtccgcaaacatgaccgatatatttgactttgtaggtcaacgatgtt
gacgctgaaattctgatctgcaaatattgcaatccttatataaaacccagagaagaggtcagttatcgcga
agccagcgcatcgaacagtttgctcaataaacaattgtctgcgacaaaggaatcagcaaaatgtcactgcg
ttgatatatagggttgatattagggttgatatatatccaagaagggttcggcagggcagtaccctgtattg
aagaggaagggtaaaccttcgaacacaacgatttgaaataagcaaaaatggaaacaataaataatgatccg
ccattgttgctttaggcatcagtaaattatttactgcgatcatttcttatagctgcggtcaagatattagc
tacaaatgcgtttgaatacatgaaaaaaacagtgtcgataaaatattttcccttaacaatactcaaacaaa
aagggttaactctttatttatttaacgtacacaactaatcttttgtccgcttctgtttacttgcttagttc
gtttatactcgtatatatacgataaattagcttatttggtttatacctactatatatacagttgaaagtct
cggtcctatcgcatttctctttgtttacccagaatttaaaacgcttgctatttcagcgacaagcgatcaca
agggagtgataaatcaatagactttagaaaatcttatttaaaatataataaaataaacccttaaaaacgat
gagttaccttgactggactaatcaaacccgaactagtagttagcaaactacgtcactttattgttttttta
ttttttttttgaaggtttatactcatgagagaaataaatcaaatcgaacacgagcatacacagagaagaa
tatattatatgttttatattacgataaatataatgatcttaactttattgggattaatattgcaaatacat
tacattattctatattgtattgctgcaatactaaaatactatgattattatataatttgctcttgtgtata
attttacgatcgctctgcaaacctatagcgataccctctagagtaaacagaccgattcatgatgtcatgtt
cgccgcaccagaaaggccgggatcacatggcacatggtcggaatggtcttggtaatataatttaagctcgg
agccgagttcgcggcgtgattttagatcattagctcaccggttagcagaaattctcgcatttttcagtctc
tcttttaggacggaaattcattattttatatgtatatggagaaaagctaactaaatgagctcatcttggta

tcaatatataatgtacttccccgaagaattcaatatctaaacggcatgtgtaactttgagcttgtttgctt
gggtaattgatatctaatcgttggaagctgacccccagagacacctctcgaaatcaaatggagaccgccca
aacactcaaacgctgactttgacatcgcccacacgcaaagttgagtgtattttagctcagcgcgttttcag
atacattttatatacacagaagtatctgaacatgcatcggggtttatgtggcatcgtccatataaattgcg
gcaattaatataatctacccgaagctgttgcatttctgttaaatgttaaatgcatttcaattgttttccgt
gccgcgagtttccttcaccggtttcctcgcatgtctaattgtga

> Anopheles gambiae
ctccgagccctccagatcgcccgtgctgcgggtcagctcgtctagcagcatgtcggtatcgatcagcccca
gcaccgcttgcttgctgatcgagcgtgccagcagccgcgttttgcgctccagctgctccttctccgtctcc
tcgctgttgaacgtcgtcgtgacgcggtccgtttcatccatcatgccctgcagcgccgtcgaggacgctac
catcggcacggcgtccggatcggtcggactcagcgggacgtagttgtgctcgaagcgtcccgtttcgggcg
cggaagcggaaagcatcgcgtccccaccacacggtggcgttagctgcgaatgatggaacccgtgggcgacg
gacgattgcgtacgattgtacgcctgctgcggcacatcgaccggttcactcggtttcggcggtccgtgctc
gaacagactggctgcggttgggcgggaatcgttgttgccctgcatcgacgtaatggtcagctgctcgtcca
cctcctcgatcacctggtcggtgatgcgcatgctgcccatcgcccggtccggtatgcgtccaccggccaga
ttgttcgagctggccgagccggggaacttgaatgggaccgaaccggccgcgtccggggtcgcactgttcgg
gcggctcgattcgtccttgctgaagaaccgcttcagcgccgaggtgaccgtgttgacgcgggacatcgtac
ccgccacgctggccaggctggtggccgagctgccaccgcggtagatcggtttatcggacgtgttaaagtgg
atgccgctcacgttatcatccatctgggagttgtaagaaacgtttgtaagtaaaatcgatcgagtcaaatg
ttaatgtagaatcagaaagtgaatctgttcggaatggttttagctccagcatcgcaacaatgtctgaagga
gaatacttcgttaaaatagtataaaagaatagtacccctcgatgtatgcaatttttctatcaattaaccat
gaaagttatgcaatttgcattacacgttacaccatttagtggtttttacagcttattagtccctttctag
acaaaattgcttagaggaggcttttcagaaaccattccagaacgtgtccactctctgaatgttataggtta
agacaagattagtagtgcagcaggatagtatgcatttaagggaatttaaaggaaccaaaggtgagtagga
tagatcgacagaaagaaccttatgtcacacaccgatgttaagtagttgcttgacgttatgttaagaaagaa
ataaacagtccgcatgtactaaaaaggcctatgctgtaacttattgaaaggaaggctatcggtaagattgg
aaagacattctaaagtatacaatttctatggtaagtatagaaaaatattgtacaagaagctctaaaactgg
cggctcttgatgacctacaaatttttcacaaaatttttggtgacatttttcgcctgtcaattcttctttaa
tttagcacgaaacaatagctaatttcaagcgtagcatcgcaaaaagcgagcattgattttatgttaaaaaa
aaaacccctaaaaatccgctacaagcacggttgttctagagtttgtactaaaatgaaatgtttttagtaca
aagatttcgcagaaaatttgacgcgaaaattagtgaattgtctagattcagtacaaattgacgaattttta
gcttgagtttttagtacaatgagcctgttatatcctctaagtaagc

> Caenorhabditis elegans
aggcaagtctggaaagcgggagagaaagaagagcgagtgatttttgaacagtatgataaaatatttttttgt
ttctcttttcactctaaactgaagatccctttcatttcatttttacatatttattatatttttaaatttcaa
attgcttaattaattttctatttttttaataaacaattgtgtaaatatatatatttttttaatacagtgtggg
aaagttctataggacccccccctaatttgaaggtttgaggaacttccgaaaatttttttgaaaaactgctaa
tgccattcgtttttaaattgaaaaaaacctatatacatttttttccagaagtttatctcaaaaactgaggt
cgcgctggaaaaaacgtcaaaatccagtgtgaaacttctataggacccccgtttttttttcacgattttta
ctaaaatcaacagattttggaatttttgacaaagctcaaatcaagtttgagttagaaatgagttcatataa
gcagttttgactttaaaaattaatacgaaatgttctcgtgggatctccagactggttctgattcttccgat
ctttgatgttcaagtctgtttcaagcttcctggtgctctcggtaatgccaaaacttgataaactctcttta
acaagttcctactaaaattcctagcacacacaccataaacatttttacgccatccccaagaaaccagtcag
aaacagcgtattaacaagttgcagttatttttgatcaacaacagaacattcatatactaaaatcaagaaag
gatcaatagttaatcgggtttccttgtgtgcggatgatctcaaacagtctgtcctccattgatctgaccaa
acttttcagctggttgtccggaatagacttccaagcgtcgagaattccttgcttcaacgatgcaactgttg
ggtaagtcttgttctgagcatacacgatacggacaagaatcccccacaaattttcgattggattgagatca
ggacttcgagctggccaatcaagaaggttgatcttcttgagcttgaaatagtcgcgggttgagttgctcac
atggattgtcgcattatcctgctgaaatctaaagtttttttctggagtagtgacgaagatatttggagagct
ccagttccaagacgttctgatagtcagtgctgttcatcttgctactgacgaactgtatctcaagcttcttc
ttctccgtgaacgctccccaaaccatcaccgttcctcctccaaaattacgtctcgaaaaaaccattggttc
cttgcgcaaatcgcgccaatagtagcggcaaccgtcaggcccatcgagattgaatttcttttcatcggaga
agacaacctaaaacaatgatcctaattattcactcttgctttttttaaattctcactttactccaattcgtt
cccatattgttcttagcaaattccaatcgcttgagtttatggtctgcagagagtaacggagcagggcgaag
tttctgacgaacgattacaccagatcgtttgatgacattgaggatggtccttttgaagcagacaattgaa

gctcattgcgaatatctcttgccgtcttacaggagttggaggcagcacgaatcacatttcgttcgtcacgc
acggagagagctttgcgacgaggagctcttttagatgtaccgtagctcaccggatccttcagatacacgcg
aatacagtgtcgagaacgggaaatttcctactcatttcatgcagggacacattgagcaatttcataacat
ccagctgagcgcgttcagtgtccgaaagggcagatcctcgaggcattgcaagttagactgctttcgaagta
agctttccagcctctatatgtgtgccacaacacatgccacaattccacatttaataattcacgcaaaaaat
agtaaataacatctgtgagggacaatttaacttgaaatattggtcccatggaaccttgtaatcaaagaaaa
acgatttgattcctgataagccttccattgtttcctgctgcatattttgccaaatcagcttgactacacag
tcgaaacatctaaagtgcgtgctaggaattttagtaggaacttgttaaagagagtttatcaagttttggca
ttaccgagagcaccaggaagcttgaaacagacttgaacatcaaagatcggaagaatcagaaccagtctgga
gatcccacgagaacatttcgtattaatttttaaagtcaaaactgcttatatgaactcatttctaactcaaa
cttgatttgagctttgtcaaaaattccaaaatctgttgattttagtaaaaatcgtgaaaaaaaacgggggg
tcctatagaagtttcacactggattttgacgtttttccagcgcgacctcagttttgagataaact

> Mus musculus
gagtaggagggtggattcactttaaggctcaggatcataatgtaaattgggactcccagaagcgtggttcc
caagaccacccaaaccacagacgcctacagcacttgataaaaacgcaacgtttccaaggacacagcatcag
gctcctaaggtctggagaagctgcgtatttggtaactgtcttggcaactgtctcagtcagtgaagacttgc
tgtccatgctcttcatgtacagtggcttgttggcaaactctcatcttagcctggtgctcagaagaacctac
agaaaggactggcacactaaggttgcctgagctgctgagatgccacccagctccctctggatgcttgggca
tgtgtctttgtttttctcaggctgtctgggattaggccgtttaaatagcatcaagccccgttgccttcctg
ctccccttttcttttttctgactttgaagtacctactagtaggccctgccaactaaacaagccaccacaacc
tatgatgtcaccagcggccaatcagagcttctcaccaatatatgcattgatgttacttttaactgtgctga
cattccagagcaacacatccttttcagacgaggacgctgtggctcagtcacagctagcacacaggtcgcta
gaagccagtggcatgtgtcaattgttccagagcggtccccagagacctggacaaagtctccttcttcccct
gagctcttgataccttagatggagctcaatcttgccccagatcttctctgaaagacttcaagggtcaggac
gcagtggcggcgacacagatttatttatttatttatttatttatttatttatttatttatttattta
tcgagacagggtttctctgtgtagccctggctgtcctggaactcactttgtagaccaggctggccttgaac
tcagaaattcacctgcctctgcctcccaagcgctaggattaaggacgtgcggcaccaccacctggcacaga
tttcttgatctaaacttagagccatggaagaggggctatggcacctaattccagcctaataagtcaacact
ccccaggaaagaattagaggaacgctggccaggctgtgctggctgctgattttgagcagatgacaggaggg
gctgagacaatggcttctttgtagaacctttgcttagaaagtcaggtgataactcagcactatctaggtg
ggcagtgccagcctccaagtatgggcaagggtacctgccccagggtctcagaaagtcccatcaacatagtg
gccagagcttttatggggttacctgaactgctggaaccctgtgtgagcaggggctgccccaggtggacagg
gccggggctttgggggcttttaggagtggtcacctgtgcttatgtatgcagttgtgatgggcacacatcgg
gagcacatgtgggagatgtgggctctgtatgtctgtgcaaagaccccgaggtaggaatgaactttgtgtgc
aggcacagagacaacagctggtttgtggtggcactggaatggtgggaatgggttagttgatcactggttta
atccgtttttattgagcaactatgctgagaacacactaagagcttttgggggggagaagagaacaaagggta
agcaggagtgcccggtggcacacggattcaagtagaaacaatgcaccatgtagaaggatcaagcacacggg
gaagatgctgctggtgacaactgaggcatgacttccagtaggatggctggggacctttttcctaggagatg
acctaatgtggaggggggccatggactgacctcaggaaaagcatc