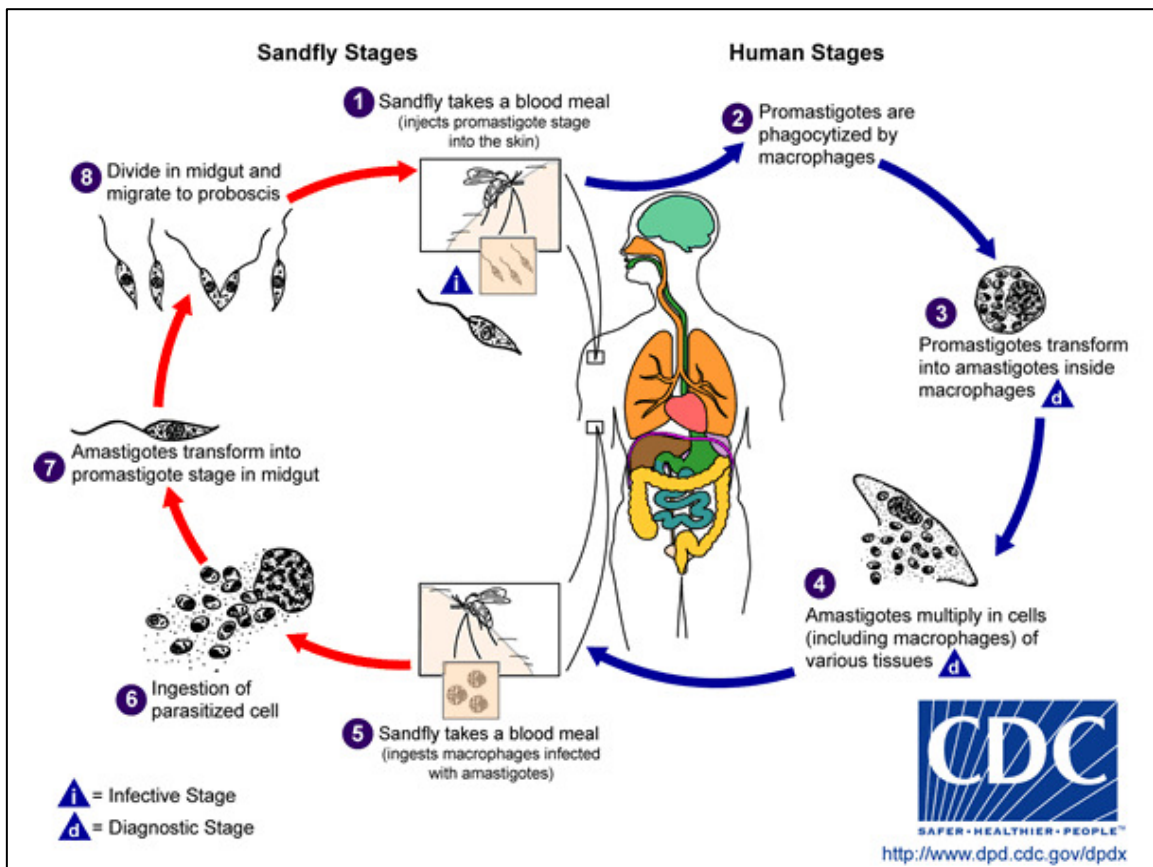


Background

Leishmania are intracellular protozoan parasites (Family: Trypanosomatidae. Order: Kinetoplastidae) which cause Leishmaniasis, a disease complex that appears in humans and other mammals in three forms: cutaneous, mucocutaneous and visceral. According to WHO data, Leishmaniasis is caused by 21 *Leishmania* species, in 88 countries worldwide [1]. Infection takes place during blood meals of female sand flies (genus: *Phlebotomus*) which serve as vectors of the disease. During infection the extracellular insect forms (promastigotes) undergo rapid differentiation into intracellular forms (amastigotes) that proliferate inside phagolysosomes of mammalian macrophages. Figure 1 describes the stages of the parasite's life cycle. Differentiation is a complex process, accompanied by a number of morphological and Biochemical changes. Recent studies employed microarray and iTRAQ analysis to investigate the time-course of changes in RNA and protein abundance of several hundred genes during differentiation [2,3].

Figure 1: The life cycle of *Leishmania* spp.



These studies revealed that some of these genes underwent an ordered progression of transient or permanent up- and down-regulation of mRNA abundance during differentiation. A significant number of genes showed a correlation between their mRNA abundance and protein expression patterns [Lahav et al., 2008 in press].

Regulation of gene expression in *Leishmania* is unusual because their protein-coding genes are transcribed as polycistronic RNAs with tens-to-hundreds of adjacent genes on the same DNA strand [5-7]. Mature mRNAs are subsequently obtained from coordinated polyadenylation and trans-splicing, which adds a 39-nt spliced leader (SL) sequence to the 5' end of all mRNAs [8]. As a consequence of this unusual gene organization, *Leishmania* gene expression appears to not be regulated at the level of transcription [9].

L. major and *L. infantum* currently have fully sequenced genomes available at GeneDB database, from where we retrieved our data. Although the differentiation iTRAQ and microarray analysis were performed on *L. donovani* we refer to the homologous genes in *L. infantum* for the convenience of bioinformatics research (*L. infantum* is the only species of the *L. donovani* complex for which genome data are available from the *Leishmania* genome projects).

Goals and Research plan

Research goals and questions

1. Discovery of possible regulatory elements that regulate transcription of polycistronic mRNA.
2. Do genes with similar mRNA abundance patterns tend to be located on the same polycistronic transcript?
3. For genes which share polycistronic location but differ in their mRNA abundance pattern can we find common mRNA regulatory elements?
4. Do genes with similar mRNA abundance pattern (correlated with polycistronic location) but different expression patterns share common functions or structures that might explain these patterns?

Research plan

1. Assigning all *L. infantum* genes to polycistrones.
2. Finding polycistronic transcription regulatory elements.
3. Looking for correlation between polycistronic location and mRNA abundance pattern.
4. Find common possible regulatory elements or secondary structures in genes which show common mRNA abundance patterns, that cannot be explained by Polycistronic location.
5. Find common possible regulatory elements in genes from previous step, that show a different protein abundance pattern than their mRNA (these genes are likely to be regulated on the translational/protein stability level).

Results

Assigning genes to polycistronic transcripts and retrieving sequences

In order to identify each gene's strand location we have created an automatic tool (in c#) that parses the information about the location of each *L. infantum* gene, according to the geneDB database. This includes predicted proteins as well.

For each of the chromosomes we receive a bitmap image of the genes (i.e. a static picture, see appendix 1). We then created an **image processing program** which assigned each gene to a strand.

The basic algorithm is as follows:

Find Lines:

Go to constant calculated shift of the bitmap (100). Go on the y-scale until you encounter a black dot. If you encounter a white dot after it - it is a line. Add it to the list of lines.

Find Genes:

We need to identify colorful rectangles. Go to an index of a line. Check for a pre-calculated shift of pixel up if we encounter a color different than black or white. If not- check for a pre-calculated shift of pixel down if you find a color of neither black or white. Mark finding a new rectangle in the direction found (up or down). Continue on the level of the found rectangle until you encounter a black pixel (the end of the rectangle). Continue procedure for all lines. Please see appendix for the code.

Strand detection

Additional properties of the gene were extracted from the HTML page describing the gene.

- Name (unique identifier of the gene), Chromosome of the gene,
- Strand -on what strand the gene is on (calculated with the above image processing program we have wrote)
- DNA Sequence of the gene
- Products - the function of the protein it creates.

All the above information was used for the analysis we will demonstrate in this paper.

Motif retrieval

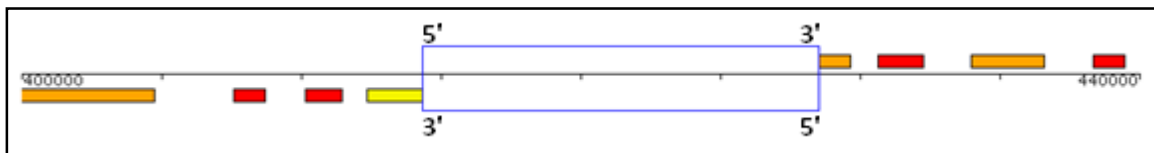
For each gene we retrieved the motif before it- the DNA sequence on the strand of the gene from the end position of the previous gene. To retrieve the sequences upstream of the genes, we retrieved the entire DNA sequence of the strand. Due to interface limitations, we reached the last gene and retrieved its final location. Only then could we extract the entire DNA sequence of the two strands.

Additionally, we separated the motifs that were located upstream of each polycistronic transcript we predicted ([manually](#)). All in all, we found a total of 158 polycistrones (see appendix 2), 8 of which had upstream sequences shorter than 8 nucleotides and thus couldn't be used as input for the next section.

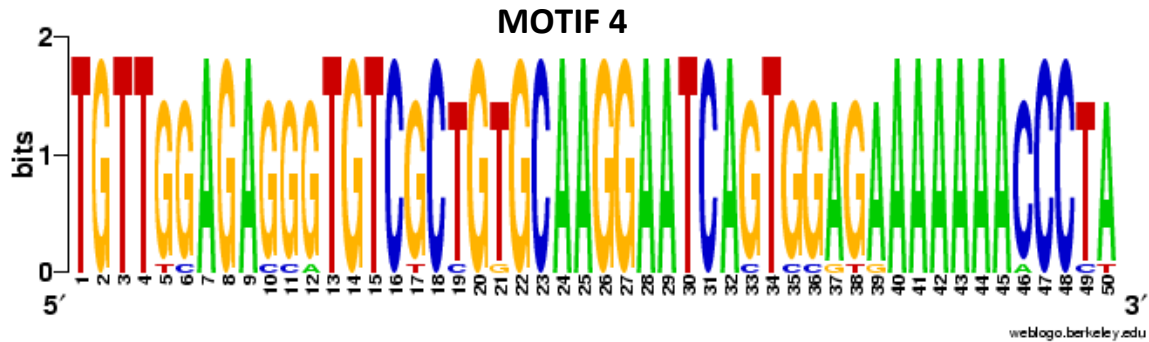
Finding polycistronic transcription regulatory elements

In order to identify motifs that might act as transcription regulatory elements we tried to discover common motifs in sequences located between the predicted polycistrones. We retrieved the sequence located upstream of each polycistron (in order to simplify our work process, we terminated these upstream sequences at the last nucleotide of the previous polycistron as described in figure 2. We decided it would suffice as a preliminary examination, especially since promoter regulatory sequences such as the TATA-box are known to be located within the first 100 nucleotides upstream of the gene in eukaryotes).

Figure 2: selection of sequences upstream of two predicted polycistrones in chromosome 8 (blue square)



These sequences were run in local MEME software for motif discovery (the web version of MEME has an input limit of 60,000 characters, so we preferred the downloadable software). The search was limited to 5 motifs, up to 50 nucleotides in length. Since some polycistrones share complimentary sequences in their upstream region (as shown in figure 2) we performed did not set the parameters to search in both strands. We received 5 motifs which were then run in HMMer package- HMMERsearch 2.3.2. HMMer searches for a given motif in the input sequences. In order to check if the frequency of these 5 motifs is significantly higher between polycistrones than within gene sequences or between genes we searched each motif in all genes sequences and in all sequences between genes. Results of MEME [10] and HMMer searches are shown in figure 3. Motif sequences are shown as Weblogo output (the overall height of the stack indicates the sequence conservation at that position, while the height of symbols within the stack indicates the relative frequency of each amino or nucleic acid at that position).

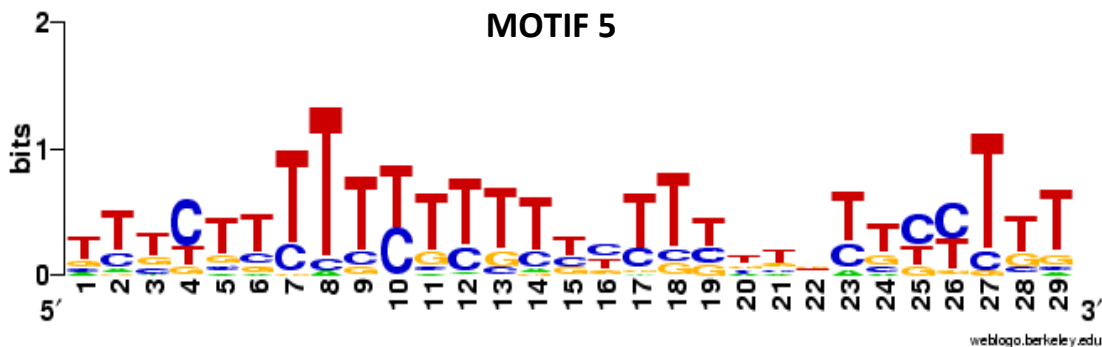


MEME search results for motif 4

E-value: 4.7e-232, width: 50 nt, polycistrones: 14/150.

HMMER search results for motif 4

Motif wasn't found within genes sequences or between genes.



MEME search results for motif 5

E-value: 5.4e-227, width: 29 nt, polycistrones: 149/150.

HMMER search results for motif 5

Discovered in 56/7948 between-genes sequences and not within genes. The cutoff point of the HMMER results was at 9.8e-005 (so that out of these 56 sequences the highest e-value was 9.8e-005).

Looking for correlation between polycistronic location and mRNA abundance pattern

For this section we used the microarray and iTRAQ results performed by Saxena et al. and Rosenzweig et al. [2,3]. This way we received information regarding the mRNA and protein expression pattern of genes identified in both of these analysis. The abundance patterns are divided into "up" "down" and "no change" according to the level of mRNA and protein found in amastigotes relative to promastigotes. We tried to establish whether there is a correlation between polycistronic transcript location of a gene (which we predicted earlier) and its abundance pattern in amastigotes vs. promastigotes. We hypothesized the answer to this would be positive for genes that are regulated on the transcription level. For each polycistron for which we had mRNA and protein abundance patterns information we calculated the p-

value of the mRNA abundance pattern. Low p-values indicate that there is a significant divergence from the 1:1:1 ratio between up, down and no change patterns within this polycistron. This 1:1:1 ratio would be expected when there is no correlation between polycistronic location and mRNA abundance pattern. Figure 4 shows all polycistrones with a p-value ≤ 0.06 . The polycistron name consists of the chromosome number, strand location and polycistron number within this chromosome. Interestingly enough the vast majority of genes located within these polycistrones belong to the "down" group.

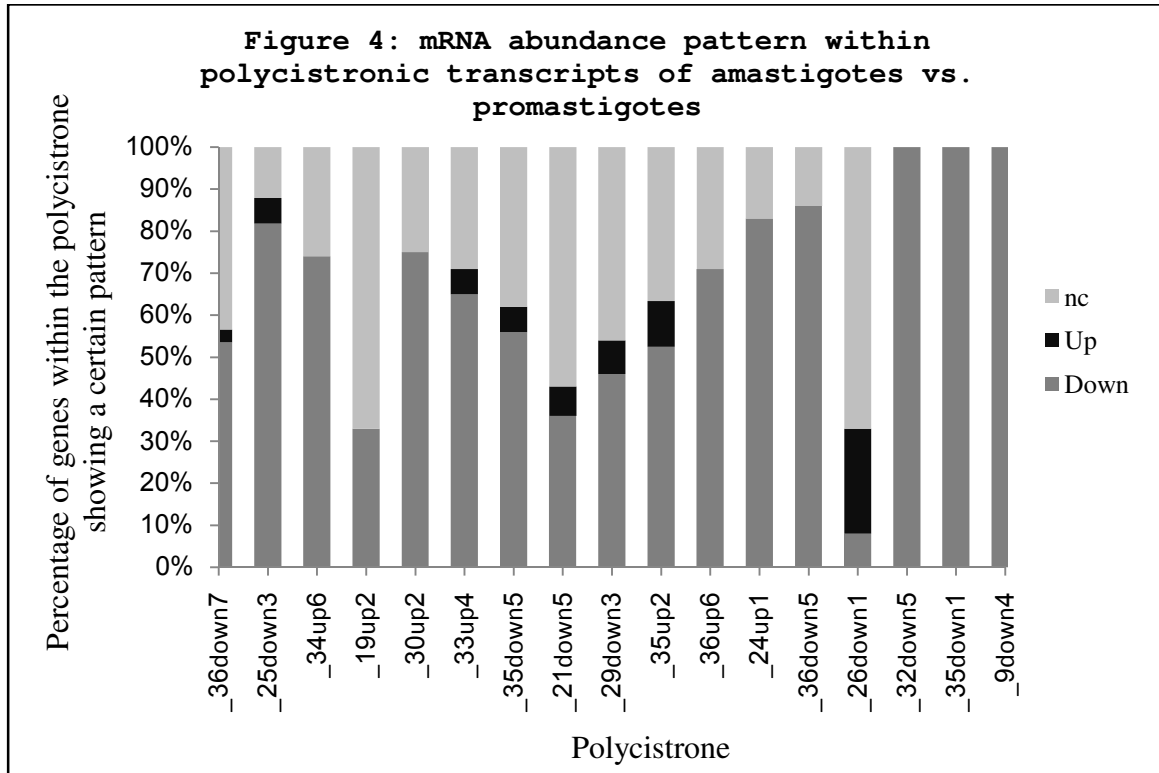
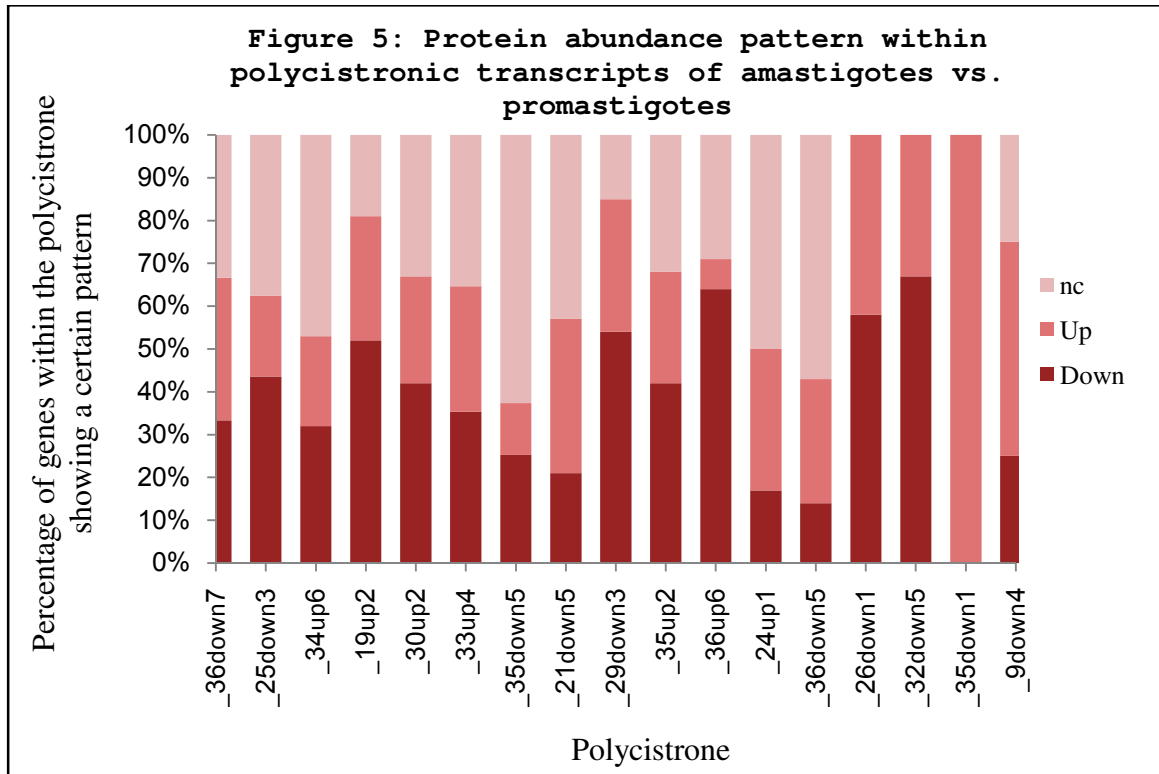


Figure 5 shows protein expression patterns of the genes that belong to the polycistrones presented in figure 4.



Seeking protein structures and RNA folding predictions that can explain abundance patterns

Due to the limitations of time and project requirements we decided to focus on two steps of mRNA and protein abundance patterns regulation.

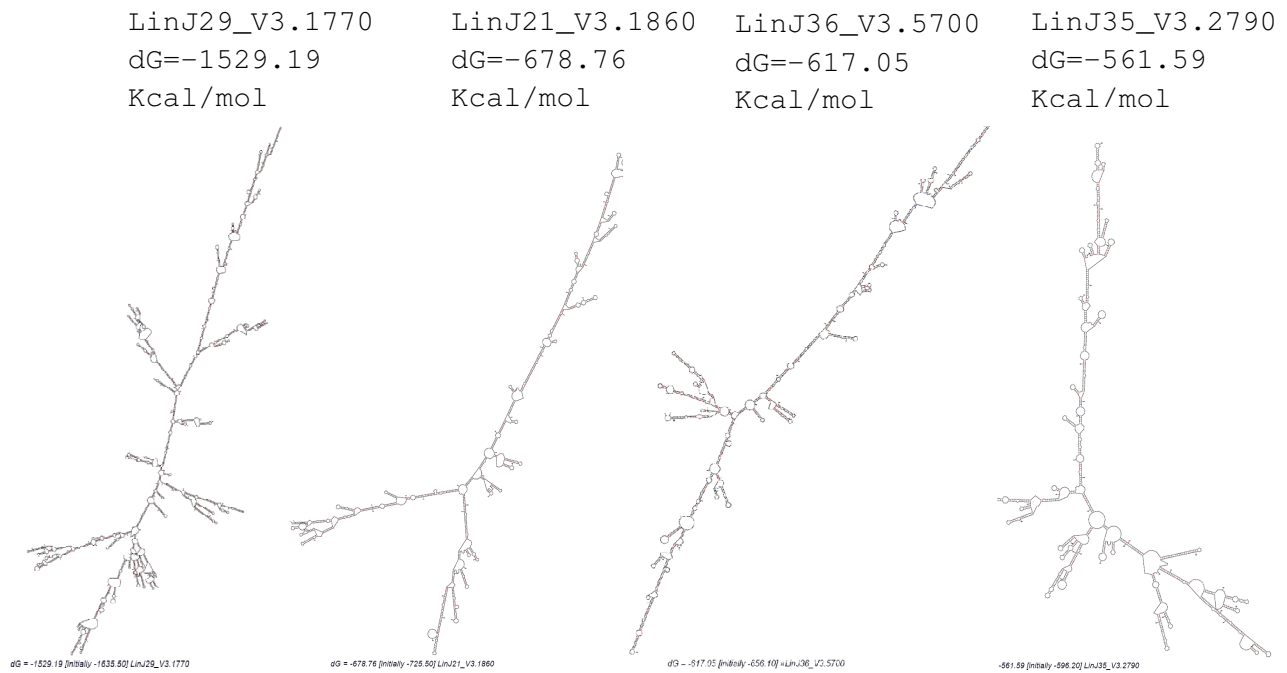
1. Looking for common RNA secondary structures and common sequences in genes located on the same polycistron, showing a different pattern than that of the majority of genes in this polycistron. We noticed the "up" pattern is under-represented as shown in figure 4, so these are the genes to be examined for common secondary structures within the mRNA (these might affect mRNA stability). Secondary structure prediction was carried out using mfold [11], and we also entered the sequences in ClustalW (however no similarity was found - data not shown).
2. Checking whether out of the genes examined in step 1 proteins that differ from their mRNA in the abundance pattern share similar structural domains both in their mRNA and protein structure (because it is possible that the initial regulation is common to genes located on the same polycistron, but translational and protein stability regulation allow different expression later on). This was carried out by checking RNA secondary structures (figure 6) and clustalW alignment on PSIPred results (figure 7), of genes with protein abundance patterns that

differ from "up" patterns (LinJ25_V3.1170, LinJ33_V3.3230, LinJ21_V3.1860, LinJ29_V3.1770, LinJ26_V3.0150).

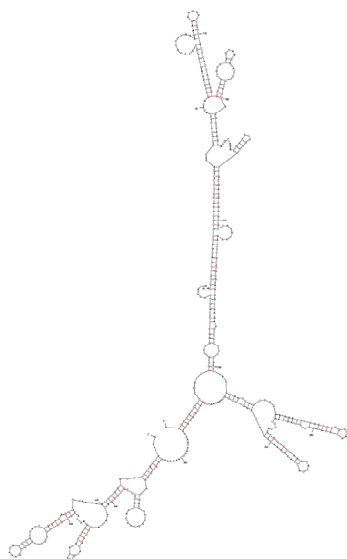
Table 1: mRNA and protein abundance patterns of genes with "up" mRNA pattern (located within the polycistronic transcripts found to have a significant correlation between mRNA abundance and polycistronic location)

Polycistronic transcript	Genes	mRNA pattern	Protein pattern
36_down7	LinJ36_V3.5700	Up	up
25_down3	LinJ25_V3.1170	Up	down
33_up4	LinJ33_V3.3230	Up	down
35_down5	LinJ35_V3.2790	Up	up
21_down5	LinJ21_V3.1860	Up	No change
29_down3	LinJ29_V3.1770	Up	No change
35_up2	LinJ35_V3.1140	Up	up
26_down1	LinJ26_V3.0150	Up	down
	LinJ26_V3.0630	Up	up
	LinJ26_V3.0780	up	up

Figure 6: Secondary structure prediction of "up" genes from fig. 4

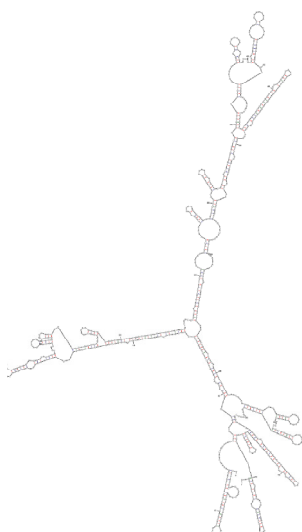


LinJ25_V3.1170
dG=-165.51
Kcal/mol



dG = -165.51 [initially -180.60] LinJ25_V3.1170

LinJ33_V3.3230
dG=-307.5
Kcal/mol



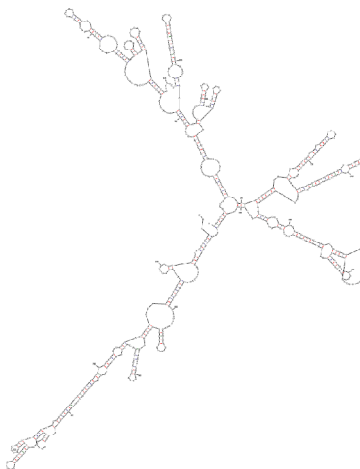
dG = -307.67 [initially -307.50] LinJ33_V3.3230

LinJ35_V3.1140
dG=-1035.42
Kcal/mol



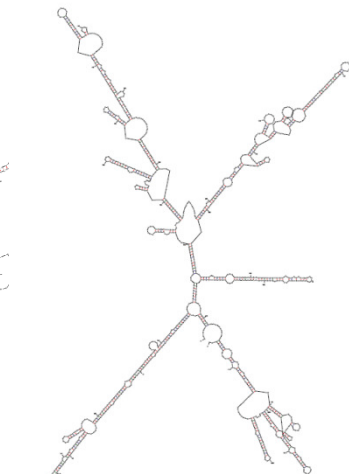
dG = -1035.42 [initially -1113.30] LinJ35_V3.1140

LinJ26_V3.0150
dG=-306.94
Kcal/mol



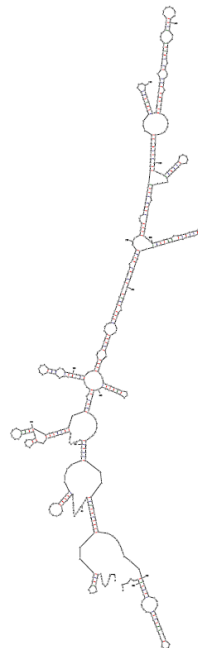
dG = -306.94 [initially -324.00] LinJ26_V3.0150

LinJ26_V3.0630
dG=-494.1
Kcal/mol



dG = -466.09 [initially -494.10] LinJ26_V3.0630

LinJ26_V3.0780
dG=-187.90
Kcal/mol



dG = -170.37 [initially -187.90] LinJ26_V3.0780

Figure 7: ClustalW alignment of PSI-pred results for the genes that show an "up" mRNA pattern but a different protein abundance pattern (LinJ25_V3.1170, LinJ33_V3.3230, LinJ21_V3.1860, LinJ29_V3.1770, LinJ26_V3.0150)

```

LinJ26_V3.0150      -----CCCCC-----CCCHHHHHHHHH----- 1
LinJ33_V3.3230      CHHHCCCCCCCC-----CCCHHHHHHHHH----- 4
LinJ21_V3.1860      C---CCCCCCCCCEEEE-----BEECCCHHHHHHHHH----- 3
LinJ25_V3.1170      -----CCCCCCC-----CCCHHHHHHHHH----- 2
LinJ29_V3.1770      CCCCCCCCCCCCCCEEEBCCCCCCCCCEEECCCCCCCCCHHHHHHHHHHHCCCC 1
                      *****
                      *** *****

LinJ26_V3.0150      -CCC-----CCCCCCCCCCCC----- 3
LinJ33_V3.3230      -CCCHHHHCCHHCECCCCCCCCCCCCCCCCCCCC----- 7
LinJ21_V3.1860      -CCCCCEEEE-----BCCCCCCCCCE----- 6
LinJ25_V3.1170      -CCC----- 2
LinJ29_V3.1770      CCCCCCEEEEEE-----CCCCCCCCCCCCCEEEEEEEHHHHHHHHHH 1
                      ***

LinJ26_V3.0150      -----CCCHHHHHHH----- 4
LinJ33_V3.3230      -----CCCHHHHHHCCHHCCCHHHHHCCCC 1
LinJ21_V3.1860      -----BCCCHHHHHCCCCCCCCCCCC----- 8
LinJ25_V3.1170      -----CCHHHHHHH----- 3
LinJ29_V3.1770      HHHHHHHHHHHHHHHHEEEEEEEEEEEEEEEBCCCHHCCCHHHHHCCCCCHH----- 2
                      *****

LinJ26_V3.0150      -----HHHHHHHHHHHHHHHHHHHHHH-----HHH----- 7
LinJ33_V3.3230      CCCCCCEEEEEEEBCCCHHHHHHHHCCHHHHHHHHHHHHHHHHHHHHHCCCCCHHEEEEEEEBCC 1
LinJ21_V3.1860      -----BEECCCCCCCCCEEECCCCCECHHHHHHCCEEE----- 1
LinJ25_V3.1170      -----HHHCCCCCECCCHHHHHHH----- 4
LinJ29_V3.1770      -----HCCCCCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH----- 2
                      ..
                      .. *****

LinJ26_V3.0150      -----HHHHHHHH-----HHHHHHHH----- 5
LinJ33_V3.3230      EEEEEEECCCCCHHHHHHHHHCCCCCHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCC 2
LinJ21_V3.1860      -----CCCHHHHCCCCCCHHHHHHHHHHH----- 1
LinJ25_V3.1170      -----CCCHHHHH-----HHHHHHHH----- 6
LinJ29_V3.1770      -----HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH----- 2
                      *****
                      *****

LinJ26_V3.0150      -----HHHHHHHHCCCC-----CCCCCCCCCEEEEEE----- 1
LinJ33_V3.3230      HHHHCCHHHHHHHHHHHCCCCCHHHHHCCCCCCCCCEEEEEEEBCCCCCHHHHHHHHHHH 2
LinJ21_V3.1860      -----HHHHCCCC-----CCCCCCCCCEEE----- 1
LinJ25_V3.1170      -----HHHHCCCC----- 7
LinJ29_V3.1770      --HHHCCCCCHHHHHHHCCCCCHHCCCCCCCCCHHHHHHHHHHHHHHHCCCCCCCCCHHH 3
                      *****

LinJ26_V3.0150      -----BEEECCEEE-----CCCHHH 198
LinJ33_V3.3230      HHHHCCHHHHHHHHHHHHHCCCCCHHEEEEEEEBCCCCCHHH-----CCCCCHHH 626
LinJ21_V3.1860      -----BEEEEECCEEEEEEEBCCCCCHHH-----CCCCCHHH 285
LinJ25_V3.1170      -----BEEEC-----CCCHHH 118
LinJ29_V3.1770      -----HHHHHHHHHHHHHHHHHHHHCCCCBEEEEE-----CCCECCCCCHHH 548
                      ..
                      .. *****

LinJ26_V3.0150      HHH-----CCCCCCCCHH 213
LinJ33_V3.3230      HHHHCCHHHHHHHHHHHHHCCCCCCCCCCCCCCCCCCCC-----CCCCCCCCCHHH 679
LinJ21_V3.1860      HHH-----CCCCCCCC-----CCCCCCCCCE 309
LinJ25_V3.1170      HHH-----HHHHHCCHHH 134
LinJ29_V3.1770      HHCCCCCHHHHHHHHHHHHHHHHHHHHHHHHEEEBCCCCCEEEEEEEBCCCCCCCCCEEE 608
                      **
                      **

LinJ26_V3.0150      HHHHHCCCC----- 223
LinJ33_V3.3230      HHHHHCCCCCCCCCCCCCCCCCCCCCEEEEEEEBCCCCCHHHHHHHHHHHHHHHHHHHHH 739
LinJ21_V3.1860      EEEEECCCCCHHCCCHH----- 328
LinJ25_V3.1170      HHHHH----- 140
LinJ29_V3.1770      EEEEECCCCCHHHHHHHHH----- 632
                      .....

LinJ26_V3.0150      -----HHHHCCCC 233
LinJ33_V3.3230      HHCCCCHEEEEEEEBCEEEEEEECCCCCHHHHHHHHHHHHHCCCCCHHHHHHHHHHHCCCC 799
LinJ21_V3.1860      -----HHHHHHHHHHCCCC 344
LinJ25_V3.1170      -----HHHHHHHHCCCC 154
LinJ29_V3.1770      -----HHHHHHHHHHHHHHHHHHHHHHCCCC 661
                      *****

LinJ26_V3.0150      CCCCCCCCCCE----- 252
LinJ33_V3.3230      CCCCCCCCCCEEEEEEEBCCCCCHHHHHHHHHHHHHHHHHHHHHCCCCCCCCCCCCBEE 898
LinJ21_V3.1860      CCCCCCCCCCEEEEEEEBCCCCCEEE----- 407
LinJ25_V3.1170      CCCCCCCCC----- 164
LinJ29_V3.1770      CCCCCCCCCCHHHHHHHHHCCCCCHHHHHHHHH----- 755
                      *****

```

Conclusions

It seems that we managed to identify a few elements within the upstream regions of polycistronic transcripts that do not appear as frequently within the gene sequences. 4 of the 5 motifs we identified are at least 50 nucleotides long and appear in a relatively small number of sequences indicating they are less likely to perform as regulatory elements, both due to length and frequency. Nevertheless, the fact that they appear only upstream to predicted polycistrones is rather interesting. Another interesting result was received for motif 5, which appeared in the vast majority of sequences located upstream of polycistrones. It also appeared in 56 of 7948 (0.7%) sequences that according to our prediction are located within polycistrones (between genes). If this motif is indeed a good indication for polycistronic transcription, it is possible that these 56 motifs are also start points for polycistronic transcription.

While 23% of genes were previously found to have a significant correlation between mRNA and protein abundance (Lahav et al., 2008 in press), the correlation for transcriptional regulation of mRNA abundance is much lower. Our results showed that only 17 of 158 identified polycistrones exhibited significant correlation between mRNA abundance pattern and polycistronic location, supporting the notion that most gene expression regulation is post-transcriptional. However, this is true only for genes identified in both microarray and iTRAQ analysis (902 out of a total of over 7000 genes), which is probably not representative enough.

It seems that the 17 polycistrones we found might indicate there is partial transcriptional regulation. It was interesting to see that the majority of genes located within these 17 polycistrones showed a "down" pattern of mRNA abundance (and very few showed an "up" pattern), but the protein expression pattern was less clear. We tried to explain this by analyzing possible regulatory elements in two groups of genes:

1. Genes within the polycistrones we identified as significant, that exhibit the least common mRNA pattern ("up").
2. Genes from the previous group, that exhibit similar protein patterns that are different from their mRNA pattern.

A clustalW alignment between genes from the first group showed no significant similarity. It is easier to identify similar RNA secondary structure within shorter sequences, but since we didn't find any sequence similarities we couldn't narrow the structure prediction to smaller sequences and remained with complicated secondary structures where we couldn't really point out a common

structure to all examined genes. It could be argued that the tendency to form the structures presented in figure 6 could affect the translation capacity, as we see that out of the 10 examined genes 5 did not show an increase in the protein abundance.

We then examined the second group, looking at genes with protein abundance patterns that differ from "up" patterns (LinJ25_V3.1170, LinJ33_V3.3230, LinJ21_V3.1860, LinJ29_V3.1770, LinJ26_V3.0150). We couldn't point out any significant common RNA secondary structures or protein secondary structures that might be unique to these genes. Although our clustalW alignment performed here (figure 7) did find some common helixes and coils, it is difficult to determine them as significant or unique and further analysis is needed for this.

Needless to say there is a lot more to be done, and due to project limitations we couldn't examine all other groups with a wider selection of genes.

Future analysis should examine distances from ORF's of the motifs we discovered, and examine experimentally whether there is polycistronic transcription from the additional 56 sequences predicted by HMMER. Also, since our secondary structures analysis concentrated on a small number of genes, this needs to be examined further on larger groups of genes, and against random groups to make sure any common elements that were discovered are indeed unique to a certain pattern.

References

1. WHO/TDR (World Health Organization/Tropical Disease Research). 2005. Progress 2003-2004. URL: <http://www.who.int/tdr/publications/publications/pdf/pr17/pr1.pdf>
2. Saxena A, Lahav T, Holland N, Aggarwal G, Anupama A, Huang Y, Volpin H, Myler PJ, Zilberstein D. Analysis of the *Leishmania donovani* transcriptome reveals an ordered progression of transient and permanent changes in gene expression during differentiation. *Mol Biochem Parasitol* 2007; 152:53-65.
3. Rosenzweig D, Smith D, Opperdoes F, Stern S, Olafson RW, Zilberstein D. Retooling *Leishmania* metabolism: from sand fly gut to human macrophage. *FASEB J*. 2008;22:590-602.
4. Lahav et al., 2008 in press
5. Myler PJ, Audleman L, deVos T, et al. *Leishmania major* Friedlin chromosome 1 has an unusual distribution of protein-coding genes. *Proc Natl Acad Sci USA* 1999;96:2902-6.
6. Ivens AC, Peacock CS, Worthey EA, et al. The genome of the kinetoplastid parasite, *Leishmania major*. *Science* 2005;309:436-42.
7. Worthey E, Martinez-Calvillo S, Schnauffer A, et al. *Leishmania major* chromosome 3 contains two long "convergent" polycistronic gene clusters separated by *atrRNA* gene. *Nucl Acids Res* 2003;31:4201-10.
8. LeBowitz JH, Smith HQ, Rusche L, Beverley SM. Coupling of poly(A) site selection and trans-splicing in *Leishmania*. *Genes Dev* 1993;7:996-1007.
9. Clayton CE. 2002. Life without transcriptional control? From fly to man and back again. *EMBO J*. 21:1881-8.
10. Timothy L. Bailey and Charles Elkan, "Fitting a mixture model by expectation maximization to discover motifs in biopolymers", *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, pp. 28-36, AAAI Press, Menlo Park, California, 1994.
11. Zuker M, Markham N. 1995-2008. Mfold server. Rensselaer Polytechnic Institute. URL: <http://mfold.bioinfo.rpi.edu/>

Appendix

1. The image processing tool identifying the genes. The purple lines are the program identifications.



2. Summary table of polycistronic transcript assignment results.

Chromosome	Number of polycistronic transcripts	Names of polycistronic transcript
1	2	1_down1, 1_up2
2	2	2_down1, 2_up2
3	4	3_down1, 3_up2, 3_down3, 3_up4
4	2	4_up1, 4_down2
5	5	5_up1, 5_down2, 5_up3, 5_down4, 5_up5
6	4	6_down1, 6_up2, 6_down3, 6_up4
7	4	7_down1, 7_up2, 7_down3, 7_up4
8	3	8_up1, 8_down2, 8_up3
9	4	9_up1, 9_down2, 9_up3, 9_down4
10	4	10_down1, 10_up2, 10_down3, 10_up4
11	3	11_up1, 11_down2, 11_up3
12	7	No data available
13	4	13_down1, 13_up2, 13_down3, 13_up4, 13_down5, 13_up6, 13_down7
14	5	14_up1, 14_down2, 14_up3, 14_down4
15	4	15_down1, 15_up2, 15_down3, 15_up4,

		15_down5
16	2	16_down1, 16_up2, 16_down3, 16_up4
17	2	17_down1, 17_up2
18	2	18_down1, 18_up2
19	3	19_down1, 19_up2
20	6	20_down1, 20_up2, 20_down3
21	6	21_down1, 21_up2, 21_down3, 21_up4, 21_down5, 21_up6
22	3	22_up1, 22_down2, 22_up3, 22_down4, 22_up5, 22_down6
23	3	23_up1, 23_down2, 23_up3
24	5	24_up1, 24_down2, 24_up3
25	2	25_down1, 25_up2, 25_down3, 25_up4, 25_down5
26	10	26_down1, 26_up2
27	9	27_down1, 27_up2, 27_down3, 27_up4, 27_down5, 27_up6, 27_down7, 27_up8, 27_down9, 27_up10
28	5	28_up1, 28_down2, 28_up3, 28_down4, 28_up5, 28_down6, 28_up7, 28_down8, 28_up9
29	4	29_down1, 29_up2, 29_down3, 29_up4, 29_down5
30	5	30_down1, 30_up2, 30_down3, 30_up4
31	6	31_down1, 31_up2, 31_down3, 31_up4, 31_down5
32	4	32_down1, 32_up2, 32_down3, 32_up4, 32_down5, 32_up6
33	8	33_down1, 33_up2, 33_down3, 33_up4
34	9	34_down1, 34_up2, 34_down3, 34_up4, 34_down5, 34_up6, 34_down7, 34_up8
35	6	35_down1, 35_up2, 35_down3, 35_up4, 35_down5, 35_up6, 35_down7, 35_up8, 35_down9
36	6	36_down1, 36_up2, 36_down3, 36_up4, 36_down5, 36_up6, 36_down7

3. DNA and protein sequences of genes from figure 4 showing an "up" pattern of their mRNA abundance.

>LinJ36_V3.5700

```
ATGGCGTCTGTAGTTGGCAAAGTCTTCTCTGGCATGAGCCAGGTCGCCGTCGTGTGCGGGCTCGTCCG
CGCCGTCTACTACCGCCGCGAGTACGGCAACATGGCCACCGACGACGCCAGCACCAGTCACTCGCTCA
TCACCTTCTCCAGGGACGACCGCCGCGACATTTACGACGTCGCCATCGTCGGCGGTGGCATCGTCCGGC
GTCGCCACGGCACGCGAGATACGGCAAAAGTACCCCAGAAAGCGCGTCATCCTCATCGAGCGCGAGGC
CGATGTGGCGCAGCACCAGTCCGGCCACAACAGCGGCTGCCTCCATGCCGGGATGTTCTACCCGCCCG
GCTCCGCCATGGCGCGCTTGTGCCCCCGCGGCCACAGTCTGATCATCGACTACTGCAAGAAGAACAAG
CTTCCCTATGAGCTGTGCGGCAAGATTGTCGTCGCCACAGAGGACAGCCAGCGCCCAACCGTTTCAGCG
CATGTACGACTGGGGTGTGCGAACGGTGTCAAGGGGCTCGAGATCCTTGAAGGCGAGGAAGCTATCA
AAAAGAAGGAGCCACTGGTGACCGGTGTCTCTGCTTTGTGGTCCCCCGTTTCGGGCATCATTGACTTT
TCGGAAGTACGCGCTGCATGCTGCGTGAGCTGACGGCCCACGCCAAGAACAACCTTCGCCACACAGTT
CCAATTCGATGCTCAGGACTTTCGTGGGCGTGTGCGGTACGAAATCGAAGGGTGCGGGCACGGAGGAGA
TGGTCTCATTTCGCGGCCGCGAGAAGAACCACCTGGGGCCGAGAAGACCATCTTGGCAAAGAACGTT
ATCACTTGCTGTGGCCTCGACAGCGACGTGGTGGCGAAGCACTCTGGTGGCATTGTGGAGTGGCTGGG
TAAGCGTGTGATGCAGACATATGGCTTTTCGCGGGCGGTATTACCAGCTGACGCCGGAGCGCCGCGACA
TGGTGCATGCACGTGTACCCGTGTCCCGATAACCCGCAAAGGCTTGAGCGTCGGTGTGCATTTCACT
CCGACAGTGGATGTGCGCCGTGGACGCCAGGTGATTATTGGGCCAGGCTCTGCGCTGGCGCTGGATCG
CTATGGCTACACCCCTTACGCGATCGACCTTGAGTACTGCTTCAACTGCGCCTTTAGCAAGGGCGGCT
GGGTGAGTCTCGTCTCCAATTTTCGATGTCATCTTCCAGACGTACTACATGGACATCAGCAAGCGGCAG
TTCCTTCGCGAGGCACAAAAGCTCATCCCTCCATCGAGGGCGAAGGACATCGTGGACAGCTACTGCGG
TGTCATGGCTGTGCGCGTGGCGGAGGATGGCACTCTCAGCATGGATCTCGCGATGGAGTTTGACCGGC
CGCGTGTGACGGTACCGGCGACGATGGACAAGAAGATGCTGCTAGAGGCGATCAAAGATGCGCCGCAC
TCTGGCAAAGGGCCTTGAGGCGTCTGACAGCTCAAAGCCGCTCATCCTGAACGTTTCGCAACGCGCCGAG
TCCGGCCGCCACTGCGTCCATGGCGATTGCTGAAGACATTGTGAAGGCCGCCAGCTCGCGATTCCAAT
GGTAG
```

>LinJ36_V3.5700p

```
MASVVGKVFSGMSQVAVVCGLVGAVYYRREYGNMATDDASTAHSPLITFSRDDRRDIYDVAIVGGGIVG
VATAREIRQKYPRKRVILIEREADVAQHQS GHNSGCLHAGMFYPPGSAMARLCPRGHSLIIDYCKKNK
LPYELCGKIVVATEDSORPTVQRMVDWGVANGVKGLEILEGEEAIKKKEPLVTGVSALWSPVSGIIDF
SEVTRCMLRELTAHAKNNFATQFQFDAQDFVGVSVTKSKGAGTEEMVLIIRGREKNHLGPEKTI LAKNV
ITCCGLSDVVAKHSGGIVEWL GKRVMTYGFGRYRYQLTPERRDMVRMHVYPCPTRKGLSVGVHFT
PTVDVRRRQVVIIGPSALALDRYGYTPYAIDLEYCFNCAF SKGGWVSLVSNFDVIFQTYYMDISKRQ
FLREAQKLIP SIEAKDIVDSYCGVMAVGVAEDGTL SMDLAMEFARPRVTVPATMDKKMLLEAIKDAPH
SGKGLEASDSSKPLILNVRNAPSPAATASMAIAEDIVKAASSRFQW
```

>LinJ25_V3.1170

```
ATGCCGCGTCCGTTTCGGTGTGTTGGGCTCCGGCCACGACGCTTGCCGAGTATCGCGCACGCATCCCGAA
CCCTTTTCGCATACAGCTTCAAGTGGGTCTACGCCATGAAGAAGGAGGTCTTCTACCCACCAGAGGCGC
TGGCGCACTTCAAGGACCCCCAGCAGTTTAAAGGATGCAGTTGACACCGTCATCGCGATGCGTGTGTCG
GACAAGATCTTCGGCGAGGGCCAGAAGCTCCCCCGCCACCCGGTCATTCTCATGGTGTGTTGCTGGC
AACGCACGGTTTTCATCATTTTTAGCGGGCTAAAGTACTACTACGGAACGTATATGCCGGCAAACAACC
CGACGTGGCGCAAGACGGTCAACAAGGAGTGGGAGGAGGCCATCAACAACCTCCCCGTGGGACCACATG
TCGCACGTGTGGCAATACAGCGACGAATACGCGGCCGCGCTCGGCGATACGGCTGCTCCTGGCCAGCG
CAAGTTCTACATCCCCGCTTAA
```

>LinJ25_V3.1170p

```
MPRPFVWAPATTLAEYRARI PNPFAYSFKWVYAMKKEVFY PPEALAHFKDPQQFKDAVDTVIAMRVS
DKIFGEGQKLPRHPVILMVMVLATHGFIIF SGLKYYYGYMPANNPTWRKTVNKEWEEA INNSPWDHM
SHVWQYSDEYAAAALGD TAAPGQRKFIIPA
```

>LinJ33_V3.3230

```
ATGAAAGGCAAGCGCGTCAACCTTACGACACGGAAAGTTTCATGTCACTCGACGAATTCGACTCTGA
TAACAACGCTT CATAACAGCGGAAAACGCCAGTGACCACACAAAGCAAGACAATCTCAGCAGTACCG
CAGCCGCGCGGAACCTGTCAGAATCTTCCATGTTACTTGAGCCGCCGGGGAACACGCCTATATCGTAC
CTATCACATTCGAACTCCCCGAAGAGCTTTCGCGCCGCTGGAGGCGGCGGAGGAGGCCGGCGCGGTGCA
```

CGGTGAGGAGCAGGCGCCAGCGCTGCAGCGCGTGGCAGCGCCGGTGGAGCCGAGCCGCGCCTCGCGG
TCACGACCCACCGGGTTTCGCTGGACGGCGATTTGTGGGCGCGTGTCTGTCGACGAGTGGCCGGACCTG
CTCAAGCAGGAGTTTACAAGCGACGTGTGCGATGCCACGGCGCTGCCGCGGACGTTCGATGCAGCGGCT
GGTGCTCACAGCCGGCAGCCTCGTCGCCGACTTCCAGCTCTCGCACGGAGGCCTCGCGAAGAGGGAGC
TGAACAAGCAGCTTGGCAGCTCGCCCTTACCCCGCACCTGGGCACTGTACGAGCGCGTTCGCCGAGACG
AAAGAGACGCCGCCACCCGCGCCACGCCGCCGGCTGCCCTTCGCGCCTCGGAGAGCTTTCGCGTTCGCT
GGACCCGGTGGAGGAGGCCGAGTGTGCGTGCAGCAGGAGGAGCATGCGCCAGCGCTGCAGCGCG
TCGGCACGCCGGTGGAGCCGAGCCGCGCCTCGCGGTACGACACACCCGGTTTCGCTGGACGGCGAT
TTGTGGGCGCGTGTCTGTCGACGAGTGGCCGGACCTGCTCAAGCAGGAGTTTACAAGCGACGTGTGCGA
TGCCACGGCGCTGCCGCGGACGTTCGATGCAGCGGTGGTGTCTCACAGCCGGCAGCCTCGTCGCCGACT
TCCAGCTCTCGCAGGAGCCTCGCGAAGAGGGAGTGAACAAGCAGCTTCCAGCTCGCCCTTACC
CGCAGCTGGGCACTGTACGAGCGCGTTCGCCGAGACGAAAGAGACGCCGCCACCCGCGCCACCCGCGC
GGCTGCCCTTCGCGCCTCGGAGAGCTTTCGCGTTCGCTGGACCCGGTGGAGGAGGCCGAGTGTGCGTG
CGCAGCAGGAGGAGCATGCGCCAGCGCTGCAGCGCGTTCGGCACGCCGGTGGAGCCGAGCCGCGCCTC
GCGGTACGACACACCCGGTTTCGCTGGACGGCGATTTGTGGGCGCGTGTCTGTCGACGAGTGGCCGGA
CCTGCTCAAGCAGGAGTTTACAAGCGACGTGTGCGATGCCACGGCGCTGCCGCGGACGTTCGATGCAGC
GGCTGGTGTCTCACAGCCGGCAGCCTCGTCGCCGACTTCCAGCTCTCGCACGGAGGCCTCGCGAAGAGG
GAGCTGAACAAGCAGCTTGGCAGCTCGCCCTTACCCCGCACCTGGGCACTGTACGAGCGCGTTCGCCGA
GACGAAAGAGACGCCGCCACCCGCGCCACGCCGCCGGCTGCCCTTCGCGCCTCGGAGAGCTTTCGCGC
CGCTGGAGGCGGTGGAGGAGGCCGAGTGTGCGTGCAGCAGGAGGAGCATGCGCCAGCGCTGCAG
CGCGTTCGGCACGCCGGTGGAGCCGAGCCGCGCCTCGCGGTACGACACACCCGGTTTCGCTGGACGG
CGATTTGTGGGCGCGTGTCTGTCGACGAGTGGCCGGACCTGCTCAAGCAGGAGTTTACAAGCGACGTGT
GCGATGCCACGGCGCTGCCGCGGACGTTCGATGCAGCGGTGGTGTCTCACAGCCGGCAGCCTCGTCGCC
GACTTCCAGCTCTCGCACGGAGGCCTCGCGAAGAGGGAGTGAACAAGCAGCTTGGCAGCTCGCCCTT
CACCCGCACCTGGGCACTGTACGAGCGCGTTCGCCGAGACGAAAGAGACGCCGCCACCCGCGCCACGC
CGCCGGTTCGCCCTTCGCGCCTCGGAGAGCTTTCGCGTTCGCTGGACCCGGTGGAGGAGGCCGAGTGTG
CGTGCAGCAGGAGGAGCATGCGCCAGCGCTGCAGCGCGTTCGGCACGCCGGTGGAGCCGAGCCGCG
CCTCGCGGTACGACACACCCGGTTTCGCTGGACGGCGATTTGTGGGCGCGTGTCTGTCGACGAGTGGC
CGAGCTGTCTCAAGCAGGAGTTTACAAGCGACGTGTGCGATGCCACGGCGCTGCCGCGGACGTTCGATG
CAGCGGTGGTGTCTCACAGCCGGCAGCCTCGTCGCCGACTTCCAGCTCTCGCACGGAGGCCTCGCGAA
GAGGGAGCTGAACAAGCAGCTTGGCAGCTCGCCCTTACCCCGCACCTGGGCACTGTACGAGCGCGTTCG
CCGAGACGAAAGAGACGCCGCCACCCGCGCCACGCCGCCGGCTGCCCTTCGCGCCTCGGAGAGCTTA
GCGCCGCTGGAGGCGGCGGAGGAGGCTGGCGCGGCGCAGCGTGGAGGAGCAGGCGCCAGCGCTGCAGCG
CGTTCGGCACGCCGGTGGAGCCGAGCCGCGCCTCGCGGTACGACACACCCGGTTTCGCTGGACGGCG
ATTTGTGGGCGCGTGTCTGTCGACGAGTGGCCGGACCTGCTCAAGCAGGAGTTTACAAGCGACGTGTG
GATGCCACGGCGCTGCCGCGGACGTTCGATGCAGCGGTGGTGTCTCACAGCCGGCAGCCTCGTCGCCGA
CTTCCAGCTCTCGCACGGAGGCCTCGCGAAGAGGGAGTGAACAAGCAGCTTGGCAGCTCGCCCTTCA
CCCGCACCTGGGCACTGTACGAGCGCGTTCGCCGAGACGAAAGAGACGCCGCCACCCGCGCCACGCCG
CCGGTTCGCCCTTCGCGCCTCGGAGAGCTTTCGCGTTCGCTGGACCCGGTGGAGGAGGCCGAGTGTGCG
TGCGCAGCAGGAGGAGCATGCGCCAGCGCTGCAGCGCGTTCGGCACGCCGGTGGAGCCGAGCCGCGCC
TCGCGGTACGACACACCCGGTTTCGCTGGACGGCGATTTGTGGGCGCGTGTCTGTCGACGAGTGGCCG
GACCTGCTCAAGCAGGAGTTTACAAGCGACGTGTGCGATGCCACGGCGCTGCCGCGGACGTTCGATGCA
GCGGTGGTGTCTCACAGCCGGCAGCCTCGTCGCCGACTTCCAGCTCTCGCACGGAGGCCTCGCGAAGA
GGGAGCTGAACAAGCAGCTTGGCAGCTCGCCCTTACCCCGCACCTGGGCACTGTACGAGCGCGTTCGCC
GAGACGAAAGAGACGCCGCCACCCGCGCCACGCCGCCGGCTGCCCTTCGCGCCTCGGAGAGCTTTCG
GTCGCTGGACCCGGTGGAGGAGGCCGAGTGTGCGTGCAGCAGCAGGAGGAGCCACCCGGTGGAGG
CGATGACGGCGTTTCGAGGAGCATGCGCCAGCGCTGCAGCGCGTTCGGCACGCCGGTGGAGCCGATCGT
GGTGTCTTCGAGGGGGCTTCCCGGTAGACGCCCGTACCGCCACGCCGCCGGTTCGCCCTTCGCGCCTC
GGAGAGCTTTCGCGTTCGCTGGAGGCGGCGGAGGAGGCCGAGTGTGCGTGCAGCAGGAGGAGGAGC
CACCGGTGGAGGCGATGACGGCGTTTCGAGGAGCATGCGCCAGCGCTGCAGCGCGTTCGGCACGCCGGT
GAGCCGATCGTGGTGTCTTCGAGGGGGCTTCCCGGTAGACGCCCGTACCGCCACGCCGCCGGTTCG
CCTTCGCGCCTCGGAGAGCTTTCGCGTTCGCTGGACCCGGTGGAGGAGGCCGAGTGTGCGTGCAGC
AGGAGGAGGAGCCACCCGGTGGAGGCGATGACGGCGTTTCGAGGAGCATGCGCCAGCGCTGCAGCGCGT
GGCACGCCGGTGGAGCCGATCGTGGTGTCTTCGAGGGGGCTTCCCGGTAGACGCCCGTACCGCCAC
GCCGCCGGTTCGCCCTTCGCGCCTCGGAGAGCTTTCGCGTTCGCTGGACCCGGTGGAGGAGGCCTCAGTGC
TGCGTGCAGCAGGAGGAGGAGCCACCCGGTGGAGGCGATGACGGCGTTTCGAGGAGCATGCGCCAGCG

CTGCAGCGCTCGGCACGCCGGTGGAGCCGCATCGTGGTGCTCTCGAGGGGGCTTCCCCGGTAGACGC
CCGTACCGCCACGCCCGGGCTGCCCTTCGCGCCTCGGAGAGCTTCGCGTCGCTGGACCCGGTGGAGG
AGGCCACAGTGTGCGTGCAGCAGGAGGAGGAGCCACCGGTGGAGGCGATGACGGCGTTTCGAGGAG
CATGCGCCAGCGCTGCAGCGCTCGGCACGCCGGTGGAGCCGCATCGTGGTGCTCTCGAGGGGGCTTC
CCCGGTAGACGCCCGTACGGCCACGCCCGGGCTGCCCTTCGCGCCTCGGAGAGCTTCGCGTCGCTGG
ACCCGGTGGAGGAGGCCGAGTGTGCGTGCAGCAGAAAGGACGAGGCGCCGGCGGCGAGCGCGCC
GCGGAGCCGCAGGAGGCACCCATGGCTGCATCCACGGAGGTGACGCCGGCACGCGAGCTCTGGCAGTC
CGTCGCGGAGCCGGAGGGCTACTCGCGCGACGACCTCGGGTGTGCGGAGCCCGCTGCGCTGCATGCGC
TGGCGGAGGACGAGGCGCCGGCGGCGGAGCGCGCCGCGGAGCCGCAGGAGGCACCCATGGCTGCATCC
ACGGAGGTGCCGCCGGCACGCGAGCTCTGGCAGTCCGTGCGGAGCCGGAGGGCTACTCGCGCGACGA
CCTCGGGTGTGCGGAGCCGGCTGCGCTGCATGCGCTGGCGGAGGACAAGGCGCCGGCGGCGGAGCGC
CCGCGGAGCCGCAGGAGGCACCCATGGCTGCATCCACGGAGGTGCCGCCGGCACGCGAGCTCTGGCAG
TCCGTGCGGAGCCGGAGGGCTACTCGCGCGACGACCTCGGGTGTGCGGAGCCCGCTGCGCTGCATGC
GCTGGCGGAGGACAAGGCGCCGGCGGCGGAGCGCGCCGCGGAGCCGCAGGAGGCACCCATGGCTGCAT
CCACGGAGGTGCCGCCGGCACGCGAGCTCTGGCAGTCCGTGCGGAGCCGGAGGGCTACTCGCGCGAC
GACCTCGGGTGTGCGGAGCCCGCTGCGCTGCATGCGCTGGCGGAGGACGAGGCGCCGGCGGCGGAGCG
CGCCCGGAGCCGCAGGAGGCACCCATGGCTGCATCCACGGAGGTGCCGCCGGCACGCGAGCTCTGGC
AGTCCGTGCGGAGCCGGAGGGCTACTCGCGCGACGACCTCGGGTGTGCGGAGCCGGCTGCGCTGCAT
GCGCTGGCGGAGGACAAGGCGCCGGCGGCGGAGCGCGCCGCGGAGCCGCAGGAGGCACCCATGGCTGC
ATCCACGGAGGTGCCGCCGGCACGCGAGCTCTGGCAGTCCGTGCGGAGCCGGAGGGCTACTCGCGCG
ACGACCTCGGGTGTGCGGAGCCCGCTGCGCTGCATGCGCTGGCGGAGGACGAGGCGCCGGTGGCCGAG
CGCGCCCGGAGCCGCAGGAGGCACCCATGGCTGCATCCACGGAGGTGCCGCCGGCACGCGAGCTCTG
GCAGTCCGTGCGGAGCCGGAGGGCTACTCGCGCGACGACCTCGGGTGTGCGGAGCCGGCTGCGCTGC
ATGCGCTGGCGGAGGACGAGGCGCCGGTGGCCGAGCGCGCCGCGGAGCCGGAGGGCTACTCGCGGGGC
GACCTTGGGTGTGCGGAGCCCGCTGCGCTGCATGCGCTGGCGGAGGACGAGGCGCCGGTGGCCGAGCG
CGCCCGGAGCCGCAGGAGGCACCCATGGCTGCATCCACGGAGGTGCCGCCGGCACGCGAGCTCTGGC
AGTCCGTGCGGAGCCGGAGGGCTACTCGCGCGACGACCTCGGGTGTGCGGAGCCGGCTGCGCTGCAT
GCGCTGGCGGAGGACAAGGCGCCGGCGGCGGAGCCGCAGGAGGCACCCATGGCTGC
ATCCACGGAGGTGCCGCCGGCACGCGAGCTCTGGCAGTCCGTGCGGAGCCGGAGGGCTACTCGCGCG
ACGACCTCGGGTGTGCGGAGCCCGCTGCGCTGCATGCGCTGGCGGAGGACGAGGCGCCGGTGGCCGAG
CGCGCCCGGAGCCGCAGGAGGCACCCATGGCTGCATCCACGGAGGTGCCGCCGGCACGCGAGCTCTG
GCAGTCCGTGCGGAGCCGGAGGGCTACTCGCGCGACGACCTCGGGTGTGCGGAGCCGGCTGCGCTGC
ATGCGCTGGCGGAGGACGAGGCGCCGGTGGCCGAGCGCGCCGCGGAGCCGCACCTGCGCGAGGGCTAC
TCGCGGGGCGACCTTGGGTGTGCGGAGCCCGCTGCGCTGCATGCGCTGGCGGAGGACGAGGCGCCGGC
GGCCGAGCGCGCCGCGGAGCCGCAGGAGGCACCCATGGCTGCATCCACGGAGGTGCCGCCGGCACGCG
AGCTCTGGCAGTCCGTGCGGAGCCGGAGGGCTACTCGCGCGACGACCTCGGGTGTGCGGAGCCGGCT
GCGCTGCATGCGCTGGCGGAGGACAAGGCGCCGGCGGCGGAGCGCGCCGCGGAGCCGCAGGAGGCACC
CATGGCTGCATCCACGGAGGTGCCGCCGGCACGCGAGCTCTGGCAGTCCGTGCGGAGCCGGAGGGCT
ACTCGCGCGACGACCTCGGGTGTGCGGAGCCCGCTGCGCTGCATGCGCTGGCGGAGGACGAGGCGCCG
GTGGCCGAGCGCGCCGCGGAGCCGCAGGAGGCACCCATGGCTGCATCCACGGAGGTGCCGCCGGCACG
CGAGCTCTGGCAGTCCGTGCGGAGCCGGAGGGCTACTCGCGCGACGACCTCGGGTGTGCGGAGCCGG
CTGCGCTGCATGCGCTGGCGGAGGACGAGGCGCCGGTGGCCGAGCGCGCCGCGGAGCCGGAGGGCTAC
TCGCGGGGCGACCTTGGGTGTGCGGAGCCCGCTGCGCTGCATGCGCTGGCGGAGGACGAGGCGCCGGT
GGCCGAGCGCGCCGTGGAGCCGCAGGAGTCAAGAATTATCCGCATGAATGAGGTAAGGGAGGTGGCGT
GCCGTGAAGAAGATGCTGAGTTCACCTGTTTTCGCGACGTCTTTTTCTCCTGGGTGTTCTTGCCCGGGC
TCCTCTGGTTTTGTTGGAGTCTCCGTACGAAGCGTCGTGGGGATGCAGGTGGTGGATGGAAAGCCGTT
AAAGCCGCGAGTGACTCAACATAAGGTGAAGCTTCAGGGAAGGTACTGGGCCAAAGTGATTTTCGTTGC
CGTGTGCTGTTGAAGCATTTCGCCAGGATGTGCGCGACGCGCTTGATGTACGTGTGGCGGATATTACG

CACGTGGAGCTGGTAGCTGGGAGCTTGTGGGCACATTTGTTGTGGTGCACGCATTTGATCTACTTAC
GGCTACCGAGGCGGACGACAAGTTGCGCGCGTACCCTTTCCGCTTACTTTGGGAGCACTACCCGAAGA
CGCCCGACACGACGACGGGCCACTTCTTTGAGGCCGAGGCTCTGTCTACAATCCGGCGCACATCACGC
TGCATCAACTCAGCATCAAAGTCCCGATTACCACCCGCTGATATGGAGTTGTCCCGGCCACCCCGGC
TGCTCTGAAGCCGCACCCTCCACTGTGGGCTGCTCAAGAGGTGTCGCAGAGCACGTTATCGTCGGGTG
GTCGTAAGCTTGGCTCTGCGACGATTAGAACACCTGCGGTGCGAGAGAGCCTTCTGTGGGGCCTGCAGTG
CACGGAGCTCCTCCTCTCATTGAGAAAACGGACGTGGGCGAGCAACCTTCTTCTTGAAGGAACGCCA
TTACGCGGGCGACGATGGGAGTGAGTCATCGATGAAGTTCTTGCAGGAAACGGCGCTGGCGCCAGTAG
CAAAGGCCGTGAGAGCGCCGGCGGTTCTGAGACGGTGAGCACGAAGCACCGCGTCGGCTTTGTGGGG
AACCAGTGGTCCACAATCCTGAAGACGCGAGAGGGTACTTCTGTGGCCGCGTTCTGTGAAGGGCACTGG
CGAGAAGCTCGGTTATGAACCGGACTCGGTGACCAAGGTGCAGTGCAGGACGAGAGCACAGGCGACCA
TCGTGAGCTTTCTGTGTACCCACCCAAAGCGTACTGCCGAGAAAGCAGATCGACTTGGTACTCCGTAGT
GCTCCCTACGCCGATGTGTGGAAGCTGTACTACAAGAACACCCCGCCGGAAGAACAGGAGACGATTCA
CAATGACGTTACAACATTTTCATCGCGTCGGCTTTGTTGGGAGCAAGTGAAGGAAGTAAGGAACCGTG
GCATGGCTCGCTTTATCGAGGCGTTTCGCAGCTGACACGGCAACTGCCCTGAATGTAACCTCCGCAAGCG
GTGCGCATCGCCGACTACGCGGTAGCCGATGACATCGTGGTGGATTTCTATGTTACCCACCCGGGGAC
AGACACAGAGGAAGTAATCGATGCTAAGCTCGAGCAGTTTCCACTACCAGCGCGTTTGGGACCTCTACG
GAATCCCGAGCGAGGCGGACGAGCAGCAGCGCGTCTGTTCCGTGCGTTATGAAATGCAGCAACGCGTCC
CGCGGAACATCTCCTTCTCGACCTGCCGACTGCGAATGCCCGGCACGAACCTCTGTTTCCCAAGCCGA
CAGCGTCTGCCCAGGTGTCAACGGCAGTTCTCCTCGAAACAGGAGTTTCTTCAACCCGACGAAAGTA
GTGGCTGGAACCCACAGCCTATGCACAGTCACACTAGCTCCATCAGCGATGCAGTTCGCACCCGCGTG
ATGGAGGACAGCTACTCAGCGATTGCAAGCCATCGCAACACCCGCTCTTCCACACCTGGCGCACCG
GTACGCTCCACACCAAAGGTCAGAAGGGCAAACATCGCAGCAGCCGAGCGCGCATCAAACGCGTGGTA
TATCATTACATACGCGGCGCACAGATGACCCCGCGTGTGCCGTGGCACACTAGTGTCAACAACAGCCTC
ATTCCGACACCCCGCGACAGCGCTCCCTCAGCAACACTACACGGTATCGTCAACGTGCGGTCAATCT
ACGCGAACTGGAGAGCGAGCTGTGCGGTAACAACGTGAGCACAAGCACCAGGAACGACAGCAGCAGC
TCGATCGCGAGATGCGCCGCTCTCTGAACTGCAGTAAATTGTGATGTCCACCCGGCCCCACTATCAAC
GTTTCACGGAGCTTCTGCCACCCATTCCACCGTGTACACAAAGGTCCGCACAGGCCCCGAGCACAT
GAAGACAGGGATGGGCTCTCTGCTGGAGTAG

>LinJ33_V3.3230p

MKGKRVNLHDTESFMSLDEFDSNNASYTAENASDHTKQDNLSSTAAARNSAESSMLLEPPGNTPI SY
LSHSNSPKSFAPLEAAEEAGAVQREEQAPALQRVGTPVEPQPRLAVTTHRVRDGDLDWARVVDEWPD
LKQEFSTDVCDATALPRTSMQRLVLTAGSLVADFQLSHGGLAKRELNKQLASSPFTRTWALYERVAET
KETPPTRATPPAALRASESFASLDPVEEA AVLRAQQEEHAPALQRVGTPVEPQPRLAVTTHRVRDGD
LWARVVDEWPDLLKQEFSTDVCDATALPRTSMQRLVLTAGSLVADFQLSHGGLAKRELNKQLASSPFT
RTWALYERVAETKETPPTRATPPAALRASESFASLDPVEEA AVLRAQQEEHAPALQRVGTPVEPQPRL
AVTTHRVRDGDLDWARVVDEWPDLLKQEFSTDVCDATALPRTSMQRLVLTAGSLVADFQLSHGGLAKR
ELNKQLASSPFTRTWALYERVAETKETPPTRATPPAALRASESFAPLEAVEEA AVLRAQQEEHAPALQ
RVGTPVEPQPRLAVTTHRVRDGDLDWARVVDEWPDLLKQEFSTDVCDATALPRTSMQRLVLTAGSLVA
DFQLSHGGLAKRELNKQLASSPFTRTWALYERVAETKETPPTRATPPAALRASESFASLDPVEEA AVL
RAQQEEHAPALQRVGTPVEPQPRLAVTTHRVRDGDLDWARVVDEWPDLLKQEFSTDVCDATALPRTSM
QRLVLTAGSLVADFQLSHGGLAKRELNKQLASSPFTRTWALYERVAETKETPPTRATPPAALRASESL
APLEAAEEAGAAQREEQAPALQRVGTPVEPQPRLAVTTHRVRDGDLDWARVVDEWPDLLKQEFSTDVC
DATALPRTSMQRLVLTAGSLVADFQLSHGGLAKRELNKQLASSPFTRTWALYERVAETKETPPTRATP
PAALRASESFASLDPVEEA AVLRAQQEEHAPALQRVGTPVEPQPRLAVTTHRVRDGDLDWARVVDEWP
DLLKQEFSTDVCDATALPRTSMQRLVLTAGSLVADFQLSHGGLAKRELNKQLASSPFTRTWALYERVA
ETKETPPTRATPPAALRASESFASLDPVEEA AVLRAQQEEPPVEAMTAFEHAPALQRVGTPVEPHR
GALEGASPDARTATPPAALRASESFASLEAAEEA AVLRAQQEEPPVEAMTAFEHAPALQRVGTPV
EPHRGALEGASPDARTATPPAALRASESFASLDPVEEA AVLRAQQEEPPVEAMTAFEHAPALQRV
GTPVEPHRGALEGASPDARTATPPAALRASESFASLDPVEEA AVLRAQQEEPPVEAMTAFEHAPA
LQRVGTPVEPHRGALEGASPDARTATPPAALRASESFASLDPVEEA AVLRAQQEEPPVEAMTAFEH
HAPALQRVGTPVEPHRGALEGASPDARTATPPAALRASESFASLDPVEEA AVLRAQQKDEAPAAERA
AEPQEAAPMAASTEVTPARELWQSVAEPEGYSRDDLGCAEPAALHALAEDAEPAAERAAPQEAAPMAAS
TEVPPARELWQSVAEPEGYSRDDLGCAEPAALHALAEDAEPAAERAAPQEAAPMAASTEVPPARELWQ
SVAEPEGYSRDDLGCAEPAALHALAEDAEPAAERAAPQEAAPMAASTEVPPARELWQSVAEPEGYSRD
DLGCAEPAALHALAEDAEPAAERAAPQEAAPMAASTEVPPARELWQSVAEPEGYSRDDLGCAEPAALH

ALAEDKAPAAERAAEPQEAPMAASTEVPARELWQSVAEPEGYSRDDLGCAEPAALHALAEDEAPVAE
RAAEPQEAPMAASTEVPARELWQSVAEPEGYSRDDLGCAEPAALHALAEDEAPVAERAAEPEGYSRG
DLGCAEPAALHALAEDEAPVAERAAEPQEAPMAASTEVPARELWQSVAEPEGYSRDDLGCAEPAALH
ALAEDEAPAAERAAEPQEAPMAASTEVPARELWQSVAEPEGYSRDDLGCAEPAALHALAEDEAPVAE
RAAEPPEGYSRDDLGAEPALHALAEDEAPAAERAAEPQEAPMAASTEVPARELWQSVAEPEGYSRD
DLGCAEPAALHALAEDEKAPAAERAAEPQEAPMAASTEVPARELWQSVAEPEGYSRDDLGCAEPAALH
ALAEDEAPVAERAAEPQEAPMAASTEVPARELWQSVAEPEGYSRDDLGCAEPAALHALAEDEAPVAE
RAAEPQEAPMAASTEVPARELWQSVAEPEGYSRDDLGCAEPAALHALAEDEAPVAERAAEPHLREGY
SRDDLGAEPALHALAEDEAPAAERAAEPQEAPMAASTEVPARELWQSVAEPEGYSRDDLGCAEPA
ALHALAEDEKAPAAERAAEPQEAPMAASTEVPARELWQSVAEPEGYSRDDLGCAEPAALHALAEDEAP
VAERAAEPQEAPMAASTEVPARELWQSVAEPEGYSRDDLGCAEPAALHALAEDEAPVAERAAEPEGY
SRDDLGAEPALHALAEDEAPVAERAVEPQESRIIRMNEVREVACREEDAFTCLRTSFSPGCSCPG
SSGLLESPSRSVVGMQVVDGKPLKPRVTQHKVKLQGRYWAKVISLPCVVEAFRQDVAHALDVRVADIQ
HVELVAGSLLGTFVVVHAFDLLLLTATEADDKLRAYHFPLTWEHYPKTPDTTTGHFFAEALSTIRRTSR
CINSASKSRLPPADMELSRPTPAALKPHPLWAAQEVSQSTLSSGGRTCSATIRTPAVRESLPVGPVAV
HGAPPLIEKRTWASNLPSKERHYAGDDGSESSMKFLQETALAPVAKAVRAPAVPETVSTKHRVGFVG
NQWSTILKTRDDFVAAFVKGTGEKLGYPEPDSVTKVQCDESTGDTIVSFRVTHPSVLPRKQIDLVLRS
APYADVWKLKYKNTPEEQETIHNDVTTFHRVGFVGSKWKEVRNRGMARFIEAFAADTATALNVTPQA
VRIADYAVADDIVVDFYVTHPGTDTEEVIDAKLEQFDYQRVWDLYGIPSEADEQQRVVPVCMKCSNAS
RGTSPSSTCLRMPGTNSVSQADSVCPRCQRQFSSKQEFLOPDESSGWNPPQMSHTSSISDAVRTAL
MEDSYSAIRKPSPTPLPLHLAHRYAPHQRSEGQTSQQPSAHQTRGISFIRGAQMTPRVPWHTSVNNSL
IPHPPRQRSLSNNTTRYRQRRVNLRELESELRSKQRQHKHQRQQQLDREMRRSLNCSKLSMSTGPTIN
VSRSLFPPIPCYTKVRTGPQHMKTMGMSLLE

>LinJ35_v3.2790

ATGCCCGCGGAGAGCTACCCGGACGAGCGGCTCGACAGCAGCCTTGCGACCTTGACGACCATCTTCTG
CGAGGAGTTCAAAGTCGAGAATGACGCGGATGTAGAGTGGCTAGTCTTACCTTCTCTCTGGCCGAG
TAAACCTGATTGGTGAGCACGTTGACTATATGGAGGGGTGGACTTGCCCGCGCGGCGGTGTTGGAGGGC
ACACACATCCTTGTGGGCGCGTGAAGCACTTCCAGAAGGAGGCGAAGCCGAAGGTTTCGCTTTTATGC
GACGTACACGAAGGAGCACTTCGACATGGACCACCTTGGCGGATGCGTGCACAATAAGGCGTGGACAA
CTTTTGTGCGCGGCCATGACGCTGCGACTCAACCGCCTCGGCGTTGCCATTGATGACGCGTCCCTG
AGTGGGTGTGCTAGTGTAGTGCACGCCACCTTCCCATTGGGAGCCGGCATGAGCGCGTCCGCTT
CGGCGTGCCTCATTACGCTATCAACTCTGTTGTACGAGGAGTTACAAGGATTGCCCCACCTCGA
GCGGCCCGCTACTCTATTGTACCGCGATGCCTAAGGAAGAGCTGATGGAGCTGGCCAAGGAAGCG
CGCCTTATCGAGACGGAATACTGCGGCGTGAATGTCGGCATCATGGATCAGTTCATCTCTGCCTTCGC
GGAGGTGGACAAGTTCATGTTTCTGGATTGCAAGCACCTCACCTTCGAGCCTATCGACATGACACCTC
TCCTGGGCCACGGCGAGTACTCGTGGATGCTGATCGACTCGATGATCAAGCACGACCTGCTTGGTGGC
ACGGCTCAGATGTATAACTCGGTTTCGCATCGATCAGGAGAGCTCGCAGAAGAAGATCGGTGAGCACCG
CTACCGTGGCAGGCCGTACTCCTTCTCCCTGATGGTGCACAACCCAGAGCAGTACGGCTTCGACGGTG
ATGTGCAAAAGTTCATGGCGGAGTTCAAGTCTCTCATGACACCCGGAGAGTTCGAGCGCGGCACCTAC
CAGATCATGGAGCAGCTGCGCACGTTGGAATTCGCAAGTTGAACGACCCGACGTTGCCGCTGTACAG
CGAAGAGCGCGTCAAAAAGCCGGCGAAATTTGAACGCCGGCCACGTCGGTATGCGCGAACTCATGA
AGATCACCACCCCGAGCTCGACTACATCCAGGAGCTCATCAACGAGGACAAGGATGTCGCCGGTGGT
CGTATGATGGGTGGTGGCTTTGGCGGCTGCATCATTATGTTGCTTCGCCGAGCGCTGAGGACCGTGT
CTTGGCGCACGTTTCGCAAACATTTCAAGGCCCGCTTCAACATGGAGAACTCCTGCTACAAGGTGAAGC
GCGCCGGCAGCGCGGTTTTGTCTTTTCTTGGAGGGAAAGAAGATGAGTGGTGTGGCAGCAAGCTC
TGA

>LinJ35_v3.2790p

MPAESYPDERLDSTLATLTTIFCEEFKVENDADVEWLVTTFSPGRVNLIGEHDVMEGWTCPAAVLEG
THILVGRVKHFQKEAKPKVRFYATYTKHEFDMHLGGCVHNKAWTTFVRGAMTLRLNRLGVAIDDASL
SGVCMVVHATLPMGAGMSASAFAFVALIHAINSVVTRSYKDCPTSSGRRYSIVPAMPKEELMELAKEA
RLIETEYCVGNVIMDQFI SAFAEVDKFMFLDCKHLTFEPIDMTPLLGHGEYSWMLIDSMIKHDLGG
TAQIMNSVRIHQESSQKKIGEHRVGRPYSLMVRNPEQYGFDDGVQKFMFAEFKSLMTPGEFERGTY
QIMEQLRTLEFRKLNPTLPLSREERVKKAGEILNAGHVMRELMKITTPELDYIQELINEDKDVAGG
RMMGGGFGGCIIMLLRRSAEDRVLAVRKHFKARFNMENSICYKVKRAGSGGFVVSLEGGKMSGAGSKL

>LinJ21_v3.1860

ATGCGTTCGCGGCGCCAAGTGCCTCGCCGGCGCGGCTGCGGACGTGGGCTCACTGAAGTTCGACGTCGT
CGTGGTTGGCGGCGGCGCCAGCGGCTCCCTTATTGCCGGCCGGCTGGCGATGGAGAACCTTAAGACGC
TCGTATCGAGGAAGGCGCCGATATCCGCAAGTGCCCGGAGTGGTACCACACGCTTCCCGCTAGCACC
ATCGCCACCCTGGCTCAGCGCGGCTTCGAGGCCAAGGACACTCTGACGACGCCGAGCAGCATCC
GGACGGGGTGACGAAGGCACCGGTGTGGATTCCGACGCCGCGGGTGTGGGCGGACATGGTGTAAATGG
GCAGTCGCACGTGGAACCTCGGCGATGAGCGTGACTGGGAGGGGAGCCCGTGGAGCTTCCGCGAAGAG
CTGCTGCCTCGCGTGCGCACCTTCGAAAACATGGAGGTGTTTGTGCCTCACCGCGGCAAGCGAGGTAA
GTTTCTCATTTGCCGCCGCTGAACTTCTCACCGTACTTTAAGGCGTTTTGCGAGGCAGCATCGCAGG
ATGTGCCGCTCATGAGCGAGTTCACGCGCAAGGAGTTCAGATCGACTGCGGCTGTGGGAGGCCCGAC
ACGTTTCGTAGACCAGTTCGCTGGGTGTCCCAACACCACCCTGCAGCGGTACCTGCTCGGCACCATCAA
GCTGAACCCGCTTTCGCGTGGAGTTCGCGTGCCAGGTGGTTCGATATCCGCGGAGCGGCGGCAACG
CGAGCACCGCTGCCGCGTGTGCTGCGGCGTGCAGATGGCCAGTTGGTGGATGTGGAAAGCACTCTG
GTGGTGGTGGTGGTAGTATCGGCAGCGCGCGGCTGCTGACGCGTTCGCGCGGCGAGCGTTCGATGT
GGACGCTTCAGTTGGCCAGCACTTCTGGGACATAACCGCAGGTGGTGTGTCAGTACCACGTGAAGAACC
GTGACAGCCACAACCTGCTATCTTGACCCGCTTGTGTGCAAGGTGCTGAGGGTGGATCTTTGGTACGGA
CGGCCACAGCTTAGCCTCCGCTCCAGCTGGGATGACCTCATCATGTACTGGTCTCTACCGGGTCTATC
GACGCCAGATGTGAGATCCAGTTTCAGCCATTTACGCTGACGAACGATGGCACGCAGCCGATGCCGA
GAGAGCACGGCTGCCAGTTCGTGCTGCGCCCCCTGCGGCCGCGCAGTTCGCGGGCAGGTGAGCGCGAAC
GGCAGCATCGACCCCAACTACTTTCAGCGATGCCGGCGACCTCGCCGCCCTGCAGAGGGGCGTTCGCGTA
TGTGAAGGACTGCCTCGTGAAGAAGACGCCTTTTGTACCGATCATTGGCGCCCTCGCTCACGAGCGCT
TTGAGTCTCCGGCATCAACGGCGGGAGCTGCGCCGGCGCCGTCGACCCGAGACCTGCCGCTCAAG
GGCGTGAGCAATGTGTACGTGTGCGATCACAGCATCCTGCCGAGCCCGCTAACTGGCAGTACGCTCCC
TTACACGTTGGCGCTGGCGGACCGATTTGTGGACAAGCTGCTCAAGAGGCGCGATGTGTAG

>LinJ21_v3.1860p

MRRGAKCLAGAAADVGLKFDVVVVGGGASGLIAGRLAMENLKTLLVIEEGADIRKCPWEYHTLPAST
IAHRLAQRGFQKADTLTTPQQHPDGVTKAPVWIPTPRVLGGHGVMSRTWNLGDERDWEWSPWFSFREE
LLPRVRTFENMEVFVPHRGKRGKFLICRPLNFSFYKAFCEAASQDVPLMSEFTRKEFQIDCGCRPD
TFVDQSLGVANTTLQRYLLGTIKLNRPVVVECGAKVVGIRGSGGNASTAAGVSLRRADQLVDVESTL
VVVSAGSISARLLTASRGSVDVDASVQHFWDIPQVVLQYHVKNRDSHNCYLDPLVCKVLRVDLWYG
RPQLSLRSSWDDLIMYWSSTGSSTPDVEIQFQPFLLTNDGTQPMPREHGCQFVVRPLRPRSRGQVSAN
GSIDPNYFSDAGDLAALQRGVAYVKDCLVKKTPFVPIIGALAHERFESSGINGGSCAGAVDPQTCLRK
GVSNNYVCDHSILPSPLTGSTLPHYTLALADRFDKLLKRRDV

>LinJ29_v3.1770

ATGACAGGACTCCCTCGTTCGAACGGCGAGGAAAACCTCGCGGGAGGCAACTGGGGAGCCTCTCACTAC
CGTGTATCCACCTCGACGGCCACAAAAGATGAGGATGCACAAAGGAGTGGGCTGGCTGAGAATGACC
TGTCAGGCGCGCTTCCGCAAACCTCTGCAGCTGCTGCCTCCCCCTTCCCTGGGTCGAGCCGCAACTCC
ACATCGGGGTCCCTTGTGAGTGTGGATCTGCTTCGACCCACCAGCACCCCTTGAACCCTATTGT
CACCCACCGCTGCCGCTGCCGCGGCTCTGTGCGATATGTTTCATCAAAAGTGCATCGATCGCTGGGTCA
TTCAGCAGCGCAACCGCGCTTCCGCGAGCTGCGGCGCGTCTACCAGCTGGTGCCTCCGAATATCCC
CCCGGTGCAAACCTTCCGCTGAGGCCGCGCAGAGCGCCGCGTGTTTTTGTTGAAGTTTTTGTATTAAGCC
GCTCCTCCTCGAATCCGCCGAGACGCTCTGCTGCGTCTTCTGCGCTTTTTTGTCTCCTCCCGCTCCTGC
TCGGGCTCGTCTACAGTTTTTCATCGCTGGCCGCTGCTGTGGCGGCCATGTGCACCTTTGCAGCATCA
CCGGAGCGGAACCTTAGCTACACCGCCTCAGTCGGTGTCCCCGGCAGCCAGGAGGTGCTGGAGAATGA
GGACGGCACGTACACCTTACCGGAGTCTTCTTGGCGTGGGCCGACGCAATTCTTTGGCCTCGTCC
TGTGCACCGTGATGAACGCGGTGGCTGTGGGGTGGGAGAAGTGAACCACTACTTCCGCGCTGCCCGG
GAGCGACTCGCCCGGAGACGGCAAGGCGAGCTGAGGAAGGAGGGTTGGCACGAGCGGCAGCGATGCC
TCCGGTAGGCCCTCTCGACTTTTCATGAGGAGAACGGCCCGCAGAGGGAGTTGAGGAGCGGGCGGAGG
CCACGGAGCAGGGGGGGAACAATGATTCTCCACAGGCGCAGCTGCCGCATCATGAGGAAGCGTCCGCA
GAGGGTTTTATCGTCGAGGGGTGGAGACCGGAATGGGGACCTCTGCCGGCGGCGGAGCAGGAACCTCT
GCCGCGCAGGATGCGCTGACGCGCGCAACCTGGGGCGGAGCAGAGGCGAGGCGAGTGGCAGGCG
AGCAGCATCGAGACGACAGCAGCAGCAGCATAATGACGACTCAGAAAGCGGCAGCAGCACGCCCTAC
CGACTTTCCTTCGTCGACGGGGTCTTTGACACGATCGACTGGGTGAAGTCGGGCGCCGGTGGCGGCGG
CATGAAGCGGCTCCTCCCTTCGTCGATGCGGCAGCTCTGCTACGGTCTTGCACTGACGCTGCTGCTGC
GCACAATCTGGGGCAGGGTAGTCGTAGCGCTTGCCTTTGCAGGCTTCATCGTGAGTTGGCGCTACCGC
CACACGCGAAACGCTCCTCACGACACAGAAGCGGCGCTTATACGAGGTTGCAGAGCGCGTGGCGCTGCT
GTCGAGGAGGCGGTGAAGACACTCTTTGGGGTGTACCTCGTGGATGCTTTTTTCTTTTACGGTGC
TACCCGAGCTGGGAGGGATGATGCTGCACTACGCCGTGCCCCGTACGTGGATATTGGCTTTGATCGC

GGCTTCCTGGCCTTTTTTACGGACTCACCGTGTGAAAGTGACGCTCTACTGGGTGGTGGGCGCGGT
GCTGGTGATGCTGCTCACCGGATGGAGTTGACGGTTGTGGGCCCTTGTGGTCCAAACGGCGTCGATC
TGTTCTTTGTGCGCAGCTTCGATGCCCGTTGGGACTCCGTTCTCGGGTACTGGCGCTGTGTGCTGACT
CAGGTATTCGACTCCGACCCTCCCCGCATCGTTTCGCGGCTTCTCCGCGTGGCGGTGGTGGAGCTGTT
GGTGCTGCTTGTCTTTGTGCGGGTGCCGTTTTGGGGAATGCTCGGGTGCCGTGACTTGTGTGGGGTG
ATGGCACAGCTGCGCAGACAGGGCTTCCGCTGAAAATGTCTCTTTACTGAGGGTGACGACCGGGTAC
GACATCAGCGCGGACGCCAGCCTTGACGAGTGGCGCCGCGCCGAGGTTCTGCGCGTTCTGGCGGAGCG
GGTGCTGCTCCCCTTTGGTGCCATGTACGCAGCCGAGTACGAGAGCAACTGAGGCCAATGCTTGATT
CAATCCCTCTCGACACCATCGCCCCCTCGAGCTGACTGCCGCGAGCTCCACTTCCATGACACCCGCC
GCCGAGAGATGTGGGAGCAAGTCTGGGGCAACATGTGAGATCGCTCGGTGTTCTTTTTACCTCAT
CGACCCCGAATCCGCCAAGAGGGTTCGACGCGGCTCGGAGGGGTTTCGCCACACGCTGGAGCGCGTTC
AGGACCCGATAGAGTGGCTGCGCGTAGCAACATCCGAGGATGGCGTGTGCAACACGCTGGAGCGCTTG
TCCTCCTCAGCAGAGTGGTTAGGGACGCTGCAGCGCGAGCCGCCAGTCGCAGATTTGGAGGCCATTGT
GGCGCCGCATTACGGTTTTAGAGCAGCAGAGCACCTCGACCGCGCTGCGCATCAGATGGTGCTCA
ACGCGTGGAGAAACACGTACCAGGATGCCATGAGCCGATTGCCAGTGTCTGCACCCCGAGGCGGTG
CCGACGCCGCGGTGCTTGTACCCAGAGCAGCACCGGTGCAGGAAGAGCCTCACACCGTGAGCGCGAT
CAAGTGTACATGGCCGCACTGCACCTTGTCTGCTACCTTACATGGGCATGGGCGCTGGCGCGCCCTCG
CGTATGAGTGGTGGTCACTGCTGCTGCTGCGCAACACATTCGCGAAGTACGTCGTCGACGTCCTCAAG
TACGTGGCAGCGATATCAACGGTCTCTCTGCTTCGCCGTGTACCCCTGCAGCGAGCGCAGCTACG
CATTCTCTTTCCGGTGGTTTCGCTGGATTGGTCAGAGTGTGTCGACATGGAGGACTACCTGTTTGACC
CAGAGCAGCTTCGCGCGGTGGAGGGATTTGTGGAGCAGGAGGGCGAGGACGAAGTGTCTGCCTCCG
ATGGTCAACCGCTGGGGTTCGACCGCGCGAGGAGTATGTGGAAGAGGCGGCTATCCCGTCTACCT
CCTGCTGCGCGTGTGGTGGTGGTGTCTCTCTTCTCGTGTGCGTCCGTCATGGTTTGGATCGTTC
CCGACTGTGCGGCACTCTCCTCTTGTGCATCACGAGGAATGCGCTACCGGTTCTTGTGGGGGCGGCC
TCCTTGAGCTTCTTCTGCTGGAACCCACATTGTTTCGCCCGCTCCGTCATCTTCGGCACAGCACTCAC
CATCGCCATTGTGTTGCTGCTGCCGATCGTCGGCATCATCTACATCGGACCTTTTGGCGAGCTGTTCT
GGAGCAGCTACCCACCCTGGTGGAGGAGACATTTGAGCGCCAGTACGACGTGCACCAGATGGTTGGC
CCTTACACGGGCTCCGCCACAGGCGGCGATGGCAGCGCGGACGGCGACTCCGATGACGACTGGGAAAG
CGTCGACTCCCGCAGCTGGTCCAAGGAAATGCCGAGGCCCTGCACCAGGATAGAGGGGCAGAGCCAG
ACGTGCGGGGACGGGCCGACGGCGACGTGCAAGCCGTAGCGGCTCACTAG

>LinJ29_V3.1770p

MQELPRRTARKTRREATGEPLTTVSSSTSTATKDEDAQRSGLAENDLSGALPQTSAAAAAPFLGSSRNS
TSGSLVECWICFDPTSTPLNP IVTHRCRCRGSVGVYVQKCIDRWVIQQRNRACRSCGASYQLVHSEYP
PGANLPLRPHERRVFLKFLIKPLLLESAETLCCVLLRFFVLP LLLGLVYSFHRWPLLWRPMS TFAAS
PERNF SYTASVGVPGSQEVLENEDGYTLTESFLAWADAILFGLVLCTVMNAVAVGWKWNHYFRAAR
ERLARETARRAEEGGLARAAAMPVPGPLDFHEENGPAEGVEERAEATEQGGNNDSPQAQLPHHEEASP
EGFIVEGWRPEWGPLPAAEQELLRPEDAPDARQPGARAEARAVAGEQHSSSDSDSDNDSESGSDTPY
RLSFVDGVFDTIDWVKSGAGGGGMRKLLPSSMRQLCYGLALTL LLLRTIWGRVVVALAFAGFIVSWRYR
HTRNVLTTQKRRLYEVAERVPLLSQEAVKTLFGVYLVDAFFFYGALPELGGMMLHYAVAPYVDIGFDR
GFLAFFHGLTVLKVTLYWVVGAVLVMLLTAMELTVVGPLFANGVDLFFVRSFDARWDSVLGYWRCVVT
QVFDSDPPRIVRGLRVAVVELLVLVFRVVPFWGMLGCRDLLWGDGTAAQTGLPLKMSLSLRVTTGY
DISADASLDEWRRAEVLRVLAERVLLPFGAMYAARVREQLRPMLDSIPLDTIAPLELTAASSTSMTPA
ARELWEQVWGNMSDRSVFLFHLIDPESAKRVAASEGFRHTLERVQEP IEWL RVATSEDGVCNTVMVL
SSSAEWLGT LQRQPPVADLEAIVAPPFTVSEQQSTSTARAHQMV LNAWRNTYQDAMSRIASVLHPEAV
PTPRSLYPEHDTVQEEPHTVSAIKCYMAALHLSYPYMGMRWRALAYEWWSL LLLRNTFANYVVDVLK
YVAAISTVLSCFAVYPLQRAQLRILFPVVRWIGQSVVHMEDYLFDPQLRAVEGFVEQEGEDELVLP
MVEPLGFDRREEYVEEAAIPSYLLLRVVAVLFLVVASVMVWIVPVL CGTLLLCITRNALPVLVGA
SLSFFCWNPTLFARSVIFGTAL TIAIVFVVP IVGIIYIGPFAQLFWSSYPTLVEETFERQYDVHQMVG
PYTGSATGGDGSADGSDDDWESVDSRDVVQGNAEALHQDRGAEPDVGDPHGDVQAVAAH

>LinJ35_V3.1140

ATGCCCGCCAAGAATCAACACAAAGGGGGCGGAGACGGCAACCCCGATCCTACCTCCACACCCGAAGC
GGCGTCACCAATGTGACAAGCACAAACGACGGTGCCGCCGTCGATTCTTCCGTGCCACCGTCCGGCG
AGACATGACTCTTTTATTGCCGCGCCGCCCGTACTCGAAGCTATCGTACGCCTTCAAAGGTATCATG
GCCGTCCTGATTCTCTGCGCCCTTCGCTCGGCGTACCAGGTTTCGCTGCTCTCCGTTT CAGATTTACGG
ATACCTGATCCACGAGTTCGACCCGTGGTTCAACTACCGCGCTGCCGAGTACATGTCCACGCACGGCT
GGTCCGCTTCTT CAGCTGGTTCGACTACATGAGCTGGTACCCGCTGGGCCGCCCTGTTGGCTCCACC

ACGTACCCGGGCTGCAGCTCACTGCCGTGCCATTACCCGCGCGCTGGCGGCTGCCGGCATGCCGAT
GTCTCTCAACAACGTGTGCGTGCTGATGCCGGCGTGGTTTTGGCGCCATCGCCACCGCTACTCTGGCTC
TCATAGCATTTCGAAGTGAGCGAATCCATCTGTATGGCGGCGTGGGCCGCACTCTCCTTCTCTATCATC
CCGGCCACCTGATGCGGTCCATGGCGGGTGAGTTTCGACAACGAGTGCATTGCCGTGCGAGCCATGCT
CCTGACCTTCTACTGCTGGGTGCGCTCGCTGCGCACGCGGTCTCGTGGCCCATCGGTGTCCTCACCG
GTGTGCGCTACGGCTACATGGTGGCGGCGTGGGGCGGCTACATTTTCGTGCTCAACATGGTTGCCATG
CATGCCGGCATATCATCGATGGTGGACTGGGCCCGCAACACGTACAACCCGTGCTGCTGCGTGCATA
CACGCTGTTCTACGTGCTCGGCACCGCCATCGCCGTGTGCGTGCCGCCAGTGGGGATGTCGCCCTTCA
AGTTCGCTGGAGCAGCTGGGTGCACTGCTGGTGTCTTCTTCTGTGTGGACTGCAGGCGTGCAGAGTG
TTTTGCGCACGCGCCGGTGTGAGGTTTCGCTCTCGCGCAACTTCAAGATCCGCGTGCAGGCTTTCAG
CGTGATGGCTGGCGTGGCTGCGCTTTCGATCGCGGTGCTGGCACCGACGGGGTACTTTCGGGCCCTTT
CGGTCCGTGTGCGTGGCTGTTTCGTGGAGCACACGCGCACTGGCAATCCGCTGGTGCAGTCCGTCGCC
GAGCACCATCCTGCCGACGCGCTCGCGTATCTGAACATTTGCACATCGTTTATTTTATGTGGATATT
CAGCTTCCCGGTGCAGCTCATCCTGCCGAGCCGAAACCAGTACGCGGTTCTCTTTGTCTTTGTCTACA
GCTTCATGGCCTACTACTTTCAGCACCCGCATGGTGCCTTGTCTATTCTGGCTGGCCCGGCGGCGTGC
CTCGGCGCAAGTGAGGTAGGTGGAACGCTGATGGAGTGGTGTCTTCAACAGCTGTTCTGGGACGACGG
CATGCCGACCGCCGATATGGTAGCAGCCGGTGACATGCCTTACCAAAAACAAGACCATGCCAGTAGAG
GTGCAGGCGCCCGACAGAAGCAGCAGAAACAGAAGCCGCGCCAGGTTTTTCGCGAGGGACTCCAGCACT
AGCAGCGAGGAGCGTCTTACAGGACACTGATCCCCGTGCACTTCCGCGAGGACGCGCAAATGAACCG
CTGGTCAGCCGAAAGACAAACGCCGCCCTCATCGTGGCTCTCACGATCGGTGTTCTTTTACCGATTG
CGTTTGTCTTCCACTTCTCGTGCCTCAGCTCAGCGTACTCCTTTGCTGGCCCGCGTATCGTGTCCAG
ACGCAGCTGCGCACCGGCGAGCAAGTGATAGTGAAGGACTACCTCGAGGCCTACGAGTGGCTGCGCGA
CAACACGCCAGAGGACGCGCGCATTTTGGCCTGGTGGGACTACGGCTACCAGATCACAGGCATCGGCA
ACCGCACCTCGCTGGCCGATGGCAACACCTGGAACCACGAGCACATCGCCACCATCGGCAAGATGCTG
ACGTGCGCCCGTGGCGGAGGCGCACTCGCTGGTGCGCCACATGGCCGACTACGTCCTAATCTGGGCTGG
GCAGAGCGGCGACCTGATGAAGTACCAGCACATGGCGCGCATCGGCAACAGTGTGTACCACGACATCT
GCCCCACGACCCGCTGTGCCAGCAATTTGGCTTTTACAGAAATGATTACAGTGCCCCAACACCGATG
ATGCGGGCGCTCGCTGTGTACAACCTGCACGAGGTGCGGAAAACAAGGGCGTGAAGGTGGACCCGCT
TCTCTTTTTCAGGAGGTGACTCGTCCAAGTACGGCCCTGGTGCAGCTTCAAGGTATGAAGTGCAGCG
AGGAGGACAAGAAGTGGGTTGCTGACCCGGCAAACCGCGTGTGCCACCCGCTGGGTGCTGGATCTGC
CCCGGGCAGTACCCGCGGCGAAGGAGATCCAGGAGATGCTGGCACACCGCGTCCCCTTCGATCAGGT
GGGCAAGGACAAGAAGGACAAGGAGGCGTACCACAAGGCGTACATGGAACGCAGCAGAACGCTGGGTG
AAGTTTAG

>LinJ35_V3.1140p

MPAKNQHKGGDGNPDPTSTPEAASTNVTSTNDGAAVDSSVPPSGETYLFHCRAAPYSKLSYAFKIM
AVLILCALRSAYQVRLLSVQIYGYLIHEFDPWFNYRAAEYMSTHGWSAFFSWFDYMSWYPLGRPVGST
TYPGLQLTAVAIHRALAAAGMPMSLNNVCVLPWFAGAIATATLALIAFEVSESI CMAAWAALSFSII
PAHLMRSMAGEFDNECIAVAAMLLTFYCWVRSRLRTRSSWP IGVLTGVAYGYMVAAWGGYIFVLNMVAM
HAGISSMVDWARNTYNP SLLRAYTLFYVVGTAIAVCVPPVGMSPFKSLEQLGALLVLVFLCGLQACEV
FRARAGVEVRSRANFKIRVRVFSVMAGVAALAI AVLAPTGYFGPLSVRVRALFVEHTRTGNPLVDSVA
EHHPADALAYLNYLHIVYFMWIF SFPVQLILPSRNQYAVLFFVYVYSFMAYYFSTRMVRLILLAGPAAC
LGASEVGGTLMEWCFQQLFWDDGMRTADMVAAGDMPYQKQDHASRGAGARQKQKQKPRQVFARDSST
SSEERPYRTLIPVDFRRDAQMNRWSAGKTNAALIVALTI GVLLPIAFVVFHFSCVSSAYSFAGPRIVFQ
TQLRTGEQVIVKDYLEAYEWRDNTPEDARILAWWDYGYQITGIGNRTSLADGNTWNHEHIATIGKML
TSPVAEAHSLVRHMADYVLIWAGQSGDLMKSPHMARIGNSVYHDI CPHDPLCQQFGFYRNDYSRPTPM
MRASLLYNLHEVGKTKGVKVDPSLFQEVYSSKYGLVRVFKVMNVSEESKKWVADPANRVCHPPGSWIC
PGQYPPAKEIQEMLAHRVFPDQVKGDKKDKKEAYHKAYMERSRTLGEV

>LinJ26_V3.0150

ATGCGCGAACTCACCAGCAACGCCACTCCACACGCGCTCTTCGAAAAACGTCTGACCATGAC
CACACTCAGTTTACGGCAACGCATCCGACATGCCCGCTGTCCCTGCCCTGAGTCCGCGA
TCAAGCGCGCTGCGTTCAAGCAGCAGCAGACGGAGAACTTCAAGAAGGCTGTGGTGGCCAGA
AAGGCCGCCAAGGCTGCCCTGAAGAAGACCGCTACTTGCCTGCCCGCAAATACTCCCGTGA
GTACCGCGGTGTGGAGAAGAAGCTGGTGCAGCTGCGCCGCCAGGCCGCTCTCACGGTAACT
ACTACCTGGAGGCGAAACCGAAGGTGCGCCGTGGTGACCCGCATCCGCGGTATCGCCAAGGTG
AACCCGAAGCAGCGCAAGATCCTCCAGCTGCTGCGCCTGCGCCAGATCTTCAACACGGTGT
CGTGAAGATGAACAAGCCGATGGAGAATATGCTGCGTGCAGTGGAGCCCTACATCGCGTACG

GCTACCCGTCCTGGCCACCGTCCGTGCGATGGTGTACAAGCGCGGCTACCTGAAGATCAAC
GGCCAGCGCGTGAAGATCACGGACAACCAGATGATCAAGGACAAGTACAACAACGCGGACAT
CGTGTGCGCGGAGGATATGGTGAACCAGATCTACACCTGTGGCAAGCACTTCCGCACGGTGA
CGCACGGCATGTGGCCCTTCAAGCTGGCCCTCCGGCCGGCGGCATGCGCCAGAAGCGCCGT
CACTTCGTGGAAGGTGGCGACTATGGTAACCGCGACACCTTGATCAACCGCTTCTCGCCCCG
CATGATCTGA

>LinJ26_v3.0150p

MRELTSNATPHAVFEKRLTMTTHSVYGNASDMPAVPAPESAIAKRAAFKQQQTENFKKAVVARKA
LKKTAYLRARKYSREYRGVEKKLVTLRRQAASHGNYYLEAKPKVAVVTRIRGIKVNPKQRKILQLLR
LRQIFNTVVFVKNMKNMENMLRAVEPIAYGYPSLATVRAMVYKRGYKINGQRVKITDNQMIKDKYNN
ADIVCAEDMVNQIYTCGKHFRVTVTHGMWPFKLAPPAGGMQRKRHFVEGGDYGNRDTLINRFLARMI

>LinJ26_v3.0630

ATGAATCGCCGGCTTAGTGTTGTGCTGGCGCTCGCGCTCGTGGTCTTCGTGCTCGCGGGGTCTGTCAG
CAGCGAAGACCCGGGTGCTGTGATGCCGGGTATCGTGCAGATGAGTAAGGACAACCTTCGACCAGCTCG
TGGGCAAGGACAAGGCGGCATTGGTGGAGTTTTACGCGCCGTGGTGTGGCCACTGCAAGAGCATGGCC
CCCGAGTACGCGGCTCTGGGCGCGGCCTACGAAGCGAGCACGAACGCTAAGGACCTGCTGTTAATTGG
CAAGGTGGACGCCACAGAGGACAGCGACCTCGGCAAGCGCTTCGGCGTACGCGGGTTCGCGACAATTC
TGTACTTCGCTTCTGGCTCATTGGAGCCGGAGAAGTACAAGGGCGGCCGCACGGCCGAGGACTTTGCC
AAGTACCTCTCGAGTGCAGTGGCGGGCTGAGGCTGACCATCCCGAATGAGCCACAGTTTGCAGTGA
GCTGGTGCACACGAATTTTGTGATGCAGTCGTCAAGGACCCGTGCAAGGCGGTGCTCGTTATGTTCTACG
CCCCGTGGTGCAGCCACTGCAAGGCGCTGAAGCCGATTTACAACAAGCTTGCCAAGGTGTTCTCAAAC
GATAAGGATGTTGTGATCGCCCGCATCAACGCTGACGATGCCGCCAACCAGGAAAATTGCAACCGAGTA
CGCTGTTTCCGGCTTCCGACGCTGTACTTCTTCCCAAAGGCGCCGATGAGAAGCCGGTTCGAGTACA
AGAATGGCCCAACCTGGAGGACTTCCTCACCTTCGTGAACGAGAACCGCGGCAAGCAGCCGCTCGCT
AACGGAGACCTCTCATGGGAGTACGGCGTCATCGAGAGCTCGCGGAGGCGGTTCGCGGCTGGCGAG
GAGCTCCGGCGAGACTCCAAGGCAGCAGTAGAGGCGGTGAAGGCGGTGCTTCCACGCTGACTGAGA
GCGAGGACGCCGCTACTACATCAAGGCAGCGGAACGCATTGCGGCGAAGGGCCCCGCTACGTGGAG
AGTGAATCGGCTCGCCTCAAGCGCACCCCTCGGAGGCCCTGTGGCGGGCGATCGCCGTGATAACATGGT
GATGCGCCTGAATATTCTGACCTCGATCCAGAAGCACATGAAGTAG

>LinJ26_v3.0630p

MNRRLSVVLALALVVFVLAGSCSSDPGAVMPGIVQMSKDNFDQLVKGDKAALVEFYAPWCGHCKSMA
PEYAALGAAYEASTNAKDLLLIGKVDATEDSDLGKRFVGTGFPTILYFASGSLEPEKYKGRTRAEDFA
KYLSSAVAGLRLTIPNEPQFAMELVHTNFDVAVKDP SKAVLVMFYAPWCGHCKALKPIYNKLAKVFSN
DKDVVIARINADDAANRKIATEYAVSGFPTLYFFPKGADEKPVYKNGRNLEDFLTFVNNAGKHRLA
NGDLSWEYGVIAELAEAVARVATSSGESSKAAVEAVKAVASTLSEDAAYYIKAAERIAAKGPAYVE
SESARLKRTRLGGPVAGDRRDNMVMRLNILTSIQKHMK

>LinJ26_v3.0780

ATGTCTATCTACGACTTCAAGGTCAACGGCAGCGACCACCAGCCATACGACCTTGGCCAGCACAAAGG
CCATCCGCTTCTCATCTACAACGTAGCCAGCAAGTGCAGGCTTACCAAGGGCGGCTACGAGACAGCCA
CGACGCTATAACAACAAGTACAAGCATCTCGGCTTACGGTACTGGCGTTCCCGTGAACCAAGTTCGCC
GGTCAGGAACCTGGAACAGAGGAGGAGGTGAAGGAGTTCGCCTGCACGCGTTTCAAGGCTGAGTTCCC
CATTATGGAGAAGGTCTGTGTCAACGGCGAGCACGAGCACCCGCTGTACCACTACTTGAAGAGCAAGT
GCAAGGGTATCCTCGGCACGACGTTTTGTGAAGTGGAACTTACGGCGTTTCTGGTGGACAAGGACGGC
CATCCCGTGTGCCGCTTCGCGCCCGGTGCGACGGTGTCCGAAATCGAGAAGAAGTTGCTGCCGCTACT
GCAGGCTGACAGTGACGGTGCAAGGTGTTGTCAAGCACTCAGTCGTGA

>LinJ26_v3.0780p

MSIYDFKVNQSDHPYDLGQHKHPLLIYNVASKCGFTKGGYETATTLYNKYKHLGFTVLAFCNQFA
GQEPGTEEEVKEFACTRFKAEFP IMEKVCVNGEHEHPLYHYLKSCKGILGTTTFVKWNFTAFVLVDKDG
HPVCRFAPGATVSEIEKLLPLLQADSDGAKVLSSTQS

Table of contents

Background	1
Goals and Research plan	3
Results	4
Conclusions	13
References	15
Appendix	16

Introduction to bioinformatics (236523)

Lecturer: Dr. Yael Mandel-Gutfreund
Teaching assistants: Martin Akerman
Sivan Bercovici

Characterization of gene expression patterns in
the differentiation process of the parasitic
protozoan *Leishmania donovani*

Submitted by: Kira Radinsky 306564303
Adele Goldman 015659477

June 2008