

Bioinformatics Computing
CSE40532/60532
Homework #7

Problems: (due 12/9)

1. Run the ClustalW multiple sequence alignment tool on the large dataset from homework # 6 (web: <http://www.ebi.ac.uk/Tools/clustalw2/index.html>). Place a text version of the resulting “alignment file” in your drop box (5 points)
2. Obtain the sequence of the corn contig MAGI3.1_104273 from:
<http://magi.plantgenomics.iastate.edu/seqSearch.do?groupName=MAGI>
3. Run the FGENESH gene finder on this sequence using monocot plant settings (you’ll need to click a different button) found under “Gene finding in Eurkayota” on the left of <http://linux1.softberry.com/berry.phtml>. Place the resulting PDF file in your drop box (5 points)
4. Determine potential functions of these all genes on this contig by using a NCBI blastx of the large contig and of the individual genes from FGENESH against the nr database. Summarize the results in a small text file in your drop box. If the genes are of unknown function, try and also use Interpro scan (<http://www.ebi.ac.uk/Tools/InterProScan/>) look for functional motifs (5 points)
5. Develop your own GFF-formatted annotation file for the gene model and protein matches. Details about GFF can be found at:
http://www.sanger.ac.uk/Software/formats/GFF/GFF_Spec.shtml

To test if your GFF is working, use the “Upload annotation” function on the MAGI corn website and check to make sure your annotations show up correctly (GBrowse site: http://magi.plantgenomics.iastate.edu/cgi-bin/gbrowse/MAGI31/?name=MAGIv3.1_104273) (10 points)
6. Visit the short read archive at NCBI (<http://www.ncbi.nlm.nih.gov/Traces/home/>). Find *Melitaea cinxia* project data by searching in the “Run” page. Place in your drop box a small text file with the project accession, the sequence of read “SRR000670.1” and the the graphic/screenshot of the associated “flowgram” in your drop box (5 points)
7. Run MEME (<http://meme.sdsc.edu/meme4/cgi-bin/meme.cgi>) on the small set of samples on the course website. Place the results in your drop box (5 points)