

Bioinformatics

Lab 10

Gene Prediction

Please read the overview about Gene Prediction in the following site:

http://en.wikipedia.org/wiki/Gene_prediction

List of gene prediction software

https://en.wikipedia.org/wiki/List_of_gene_prediction_software

Read the following papaya genome paper – to see you can understand the paper or not, and see how many genes you can predict – compare your results with the result in the paper.

http://proteomics.yzu.edu/courses/BIOL4800_6900/ExtraReadings/Papaya_XY.pdf

I have put genomic sequences of Papaya X and Y chromosome at

http://proteomics.yzu.edu/courses/BIOL4800_6900/LAB/papaya/

Use “save link as” to download the data – do not open it directly online

Please use these files (you can use part of the file, or you can use other genomic sequences from any eukaryotic species) to test the following programs for protein-coding gene prediction:

<http://hollywood.mit.edu/GENSCAN.html>

As the server has a limit in sequence length, thus, often stand-alone software tool is used in real experiments. If you have interest in learning how to use GenScan in Linux please see Dr. Min in the lab.

The tool is located at: /home/tools/genSCAN/

Read “readme” for learning how to run the program.

Go over the following sites to familiarize the format for genes/transcripts:

GTF – gene transfer format:

<http://mblab.wustl.edu/GTF22.html>

GFF – gene finding format (general feature format)

https://en.wikipedia.org/wiki/General_feature_format