

Hints for final project

All supporting files are in:
/home/student/2005/eeb/final-project/

1. Look at the flow of logic before you proceed further (**flow.pdf**)
2. Visit the Ensembl site and read up on how to display your own data/annotations in the genome browser .

http://www.ensembl.org/info/data/external_data/index.html

Follow the link in the "URL-based Data Sets" section, for detailed instructions on how to create the output using GFF format.

3. You will be creating GFF format files and keeping them in your public_html directory and referencing them as

<http://amadeus.biosci.arizona.edu/~igertXX/gene.gff>

Where igertXX is your login name on amadeus

4. Open the **sample-data.xls** and look at the various sheets. They contain sample output from each of the teams which you will combine and use for prototyping your solution.
5. Lucky for you the teams have a format that is very close to GFF style. Take a look at **output.xls** a sample file and see how I have combined example data. Some of the columns need to be generated for some of the data sets. It will be easier for you to add those columns in Excel and then combine them to create a single sheet with data for the whole project (remember to export it as csv and delete the column names before loading into mysql)
6. Create the appropriate tables in Mysql using the layout of output.xls as hint !. Name your table "feature" and load data from step 4 (above) into it.
7. Now write a perl program to query the stored mysql data and create output
8. This program will be called mysql2gff.pl
 - This program needs to generate two flavors of GFF:
 - One with decoration information to show data as tracks in Ensembl browser
 - Second to create plain GFF to load data in tools like Artemis
 - It will take 2 arguments -s and -h:
 - -s will select given **source** name from the mysql database
 - -h for adding the decoration header
 - It will print output to STDOUT in GFF format (with or without header)

as decided by user say \$h==1 and \$h==0)

9. We have 4 types of tracks (*publication*, *gene*, *primers*, *SNP*). I have put the decorations for the tracks in a file called tracks.txt in /home/student/2005/eeb/final-project/ You can copy and paste the 4 lines into it into your perl program e.g.
\$snp_track = "track name=SNP description=\"Novel SNP\"
color=\"153,51,255\" "; Now you can print "\$snp_track\n" ; when you need it !
10. Your program will use 4 sql statements to pull appropriate data for a given -s value (i.e from one of the following: *publication*, *gene*, *primers*, *SNP*) then for each track depending on the value of -h option it will or will not print the track information followed by the GFF data
11. Test your program: .
./mysql2gff -s publication -h 0 > publication.gff
should output data and save it to file publication.gff without header/decoration
./mysql2gff -s gene -h 1 > gene.gff
should output data and save it to file gene.gff with header/decoration
12. Once you have created the 4 gff files (with header) copy them to your public_html folder. Now you are ready to add those tracks to Ensembl browser
13. Now fire up the web browser point it to (replace igertXX with your login) to check and see if you are able to access the 4 files.
<http://amadeus.biosci.arizona.edu/~igertXX/publication.gff>
14. If all the 4 files are visible, go ahead and add them as tracks to Ensembl genome browser. In Ensembl browser there is a "detailed view" section where you will find a drop down "DAS sources" and it has a option called "URL based data" you will provide the 4 URL's (one at a time)
15. Look at human chr 21 from 26690000 to 26890000. See your custom tracks .. click on a item in the pubmed track and follow the links ! (click on details)
- 16 . Now write a nice proposal , While using the Ensembl site click on the **view as PDF** link at the bottom right side (in Detail View) and save the output as pdf and print it for your report ! Feel free to turn tracks on/off.
19. Make sure the report includes ample instructions on how to use your product (load and view data etc. !). Also talk about its shortcomings and ways it can be improved once they pay the price !