Lab 2: Finding and Reviewing Scientific Articles

Background

For this lab, we will do preliminary research to find genes associated with some phenotype of interest. We've discussed the essential genotype-phenotype relationship. We've also discussed how to access, sift, and organize the scientific literature, and obtain genetic data from various databases and organize it using regex. So, this lab will attempt to combine these various concepts in a single project. You may use Mendeley groups to collaborate and share literature with your classmates.

Here are the **goals**:

1. Research the phenotype of interest (listed below by first letter of your last name) enough to develop a reasonable introduction about why this phenotype is interesting and what might be controlling the phenotype from a genetic perspective.
2. Find genes associated with the phenotype of interest in the target organisms and develop a fasta formatted file of the associated data.

Phenotypes

Group 1: If your last name starts with **A-G**, your phenotype of interest is **allergies**.

Group 2: If your last name starts with **H-P**, your phenotype of interest is **the ability to live in toxic/polluted environments** (e.g. polychlorinated biphenyls in water/sediment).

Group 3: If your last name starts with **Q-Z**, your phenotype of interest is **longevity**.

Links to the corresponding Mendeley groups will be posted on Blackboard.

Tasks

1. Research the literature on your phenotype (feel free to narrow things down to a particular phenotype within these broad categories) and your target organism(s) - this means read the titles and abstracts and even the whole papers if they are relevant!
2. Set up an account in Mendeley and join the group (1, 2, or 3) relevant to your phenotype. Upload your papers into the group and look at the other literature your colleagues are finding for you.

3. Find genes associated with your phenotypes of interest via the literature and then by finding nucleotide sequence data at relevant data bases. Download the data and associated references. Include the references in Mendeley and organize your data into a FASTA formatted file. Include at least 10 entries (aka Genbank submissions, known to students as "sequences") per gene and at least two genes of interest.

4. Format your FASTA file to include ">accession#_genesymbol_taxon" as the name, using regular expressions when/if convenient. The gene symbol is usually a short abbreviation or acronym for the full gene name. There are often multiple variants, just pick one that is commonly used. For example, "ATP-binding cassette transporter sub-family C member 11 gene" = ABCC11, "breast cancer 1 gene" = BRCA1.

5. Write your lab report in the typical style (below) with the introduction including the literature reviewed in task 1 and a definition of your search strategy and regex (if any) for formatting your genetic data in your methods section. Your results section should describe the genes found and your FASTA file. The discussion should describe why you think the genes are associated with the phenotype of interest and the taxonomic diversity that you were able to find for these genes.

Guidelines

Begin the paper with an original title, followed by your name, the course name, and the date. All write-ups should be single-spaced and in 12 pt font. Your paper should have all of the following sections.

Sections

0. **Title:** A concise, but informative quick summary or description of the report.

1. **Introduction:** Introduce the general problem or issue you are addressing. Provide the background context for the reader. You should prepare the reader for the following sections of the report as well as convince the reader that you are tackling a problem worth addressing in your work.

2. **Materials and Methods:** Describe the methods used to obtained the data, analyze the data, and test hypotheses associated with the data. You should explicitly mention and cite any bioinformatic tools used. You may wish to provide the parameters (settings) of any software that you used, if relevant or important.

3. **Results:** Provide and describe the results of the data analysis and hypothesis testing. This may include presenting charts, tables, pictures, and other figures when appropriate for effectively communicating your results. This section is primarily for *only* presenting the results. The following section is for discussing them.

4. **Discussion:** Interpret your results and discuss their implications with the reader. This section should include a synthesis of ideas.
5. **Conclusion:** Wrap up and summarize the paper. You may also wish to discuss the future direction of similar research here, if relevant. The conclusion does not need to be its own section—you may choose to put a concluding section at the end of discussion. Your paper, however, should include some piece of writing the neatly wraps up your report.
6. **References:** List the relevant literature you have read and used to support your arguments/analyze your data. The literature cited should be in the format of the journal *Bioinformatics*. If you are unfamiliar with this citation style, look up the guidelines for it in the journal's "Instructions to Authors" guide. Failure to cite work used may be considered plagiarism, if the offense is serious.

Frequently Asked Questions