Practical Bioinformatics

Wagner Section

Exercise Block 4. Exploring the metabolism of *E. coli* more thoroughly and studying the effects of employing different metabolites as carbon sources

The analysis will be done with a specific model of the *E. coli* metabolism called iJO1366. The following exercises contain some guidelines of analyses you can do, but we encourage you to come up with questions on your own.

Exercise 4.1. How many metabolites, genes and reactions does the model of *E. coli* iJO1366 include? Maximize biomass production. How many reactions are active (carry a flux different from zero)? Look at the active external reactions: what is *E. coli* consuming and excreting?

Exercise 4.2. Find all essential reactions in the *E. coli* metabolic network for growth in a minimal glucose-containing environment. What fractions of the total reactions are essential?

Note: Thus far, you were able to find the essential reactions by looking for all the rows in the DataFrame where the "growth" value is below a certain threshold. In this model, you also want to look for the cases where the "status" is "infeasible". These are the deletions where no solution could be found. Check the COBRApy reference to find out how to use Boolean indexing with multiple conditions.

Exercise 4.3. Compute the fraction of reactions that are essential for *E. coli* biomass synthesis in an acetate minimal environment. Is it higher or lower than the number of essential reactions for biomass synthesis on glucose? Which reactions are essential in both environments?

Exercise 4.4. Modify the python program written for the previous exercise to find the reactions that are essential for growth on acetate, but not on glucose.

Exercise 4.5. Use the file iJO1366_reactionInfo.csv to find out how many essential reactions on glucose occur in each metabolic pathway (or subsystem). Which pathway is the one with the most essential reactions?

Understanding the metabolic pathway in which essential reactions occur can provide useful biological information, because pathways with many essential reactions are arguably more important for survival in a specific chemical environment. Additional

information about the model iJO1366.xml is given in a separate file (iJO1366_reactionInfo.csv). Explore the file. Where is the information about the subsystems given?

Exercise 4.6. In Exercise 4.2., you identified essential reactions for growth in a glucose minimal environment. Now compute these reactions for the rich environment specified by the file rich_environment.txt. For a gut bacterium such as *E. coli*, which of these environments may be more realistic? Why? Is the number of essential reactions for growth in these environments different? If so, is it smaller or larger, and why?

Hint: In the file rich_environment.txt, the uptake bounds are defined for the metabolite IDs, which are slightly different from the respective exchange reaction IDs.

Exercise 4.7. Search for the metabolites that have the potential to be carbon sources and calculate the maximum growth achievable on each.

Is it the same in aerobic and anaerobic conditions? Why or why not? Can you provide an example of a carbon source for which *E. coli* is viable only aerobically and a carbon source for which *E. coli* is viable only anaerobically?

Exercise 4.8. Focus on the aerobic condition. Plot the distribution of maximum biomass growth on the different carbon sources. Does biomass growth depend on the total number of carbon atoms in a carbon source? Why or why not? Does the number of active reactions change? Why or why not?

Hint: You can plot a histogram using matplotlib using:

```
import matplotlib.pyplot as plt
plt.hist(values, bins=15)
plt.show()
```

Exercise 4.9. In a previous exercise (Exercise 2.4.), you computed a measure of network flexibility that encapsulates how strongly flux values within a network can vary while achieving the same (optimal) biomass growth. Does network flexibility differ among different carbon sources? (Draw a histogram.) Can you speculate on the reasons why?

Hint: The code for this exercise will probably take quite some time to run, so run it for a small number of carbon sources first to see if the plotting works as intended.