Molecular Population Genetics (Binomial Distribution & Sample Sizes)

Lab Assignment 3 (due 1 October, 2014) – email me this sheet along with script

Your Name:

In this assignment you will explore the effect of varying sample sizes on the probability of detecting alleles. This is useful in experimental design, research proposals, and application permits.

The probability of detecting a specific allele is equivalent to 1 minus the probability of not detecting it.

Binomial equation:



n= number of trials (i.e. number of alleles sampled)

s = number of successes of event 1 (e.g. the allele of interest)

t = number of successes of event 2 (e.g. all other alleles in the population at that locus)

a = probability of event 1 occurring based on a single trail (e.g. popn. frequency of allele of interest)

b = probability of event 2 occurring based on a single trail (e.g. popn. freq. of other allleles in popn.)

1. Write a script that queries the user for the above information, and then calculates the probability of detecting a certain allele.

2. Explore for various numbers of alleles sampled and allele frequencies how the probability of detection changes. Summarize your findings using a graph (ideally, you will make the graph in python).

3. What sample size would you need to be 95% confident that you could detect an allele (i.e. record at least one detection) that has a frequency of 5% in the population? Explore how that probability changes as a function of sample size and summarize using a graph (ideally, you will make the graph in python, but R or excel are fine).

4. Read and work through Chapters 1 – 6 of Haddock and Dunn (Practical Computing for Biologists). Try this online tutorial http://gnosis.cx/publish/programming/regular\_expressions.html

Here are some practice problems tutorial http://regex.sketchengine.co.uk