Weller – Day 2 Session 1 Class Lab/ Exercises

Sample Calculations and finding sequences for use in the Blue Line.

Mostly we want to collect just enough sample material to measure the trait we are interested in, but not so much that we are destroying material for no purpose, since removal of tissue causes harm, at some level.

1. How many leaf cells do I have to collect in order to have 100,000 chromosomes?
   1. Note that plants can be diploid, tetraploid, hexaploid, octapolid, decaploid, and whatever the word for 300-ploid is.
2. What information do I need in order to find out how much leaf tissue is required to purify 10ng (nanograms, or 10-9 grams) of the RuBisCO protein?
3. If I am studying flower morphology, how many flowers can I sample in one day if it takes me 10 minutes to observe and score each flower? Assume an 8-hour working day.
4. How many copies of a chromosomal gene sequence are present in human
   1. Skin cells
   2. Red blood cells
   3. Reproductive cells
   4. Heart muscle cells
5. How many copies of a mitochondrial gene sequence are present in human
   1. Skin cells
   2. Hair follicles
6. Find an on-line document describing the Human Subjects Research Ethics Guidelines for the University of North Carolina at Charlotte. Find one for a college or University you are considering attending.
7. Carry out the sequence-finding lab as follows:

The remaining slides walk you through finding specific CP-rbcL bar code segment information for C. mollissima.

1. Google for ncbi. <http://www.ncbi.nlm.nih.gov/>
2. In the top selection box, drop down to Genome: <http://www.ncbi.nlm.nih.gov/genome/?term>=
3. On the left, under Using Genome, click on Browse by Organism, filter by Eukaryotes, go to page 5, find Castenea mollissima, click on GCA link – you will see the page on the slides – this is the assembly (ASM page for a genomic sequencing project).
   1. Links from Assembly: If you click the Genome link on the right, it takes you to the next page shown in the slide: [http://www.ncbi.nlm.nih.gov/genome/?term=txid60419[orgn](http://www.ncbi.nlm.nih.gov/genome/?term=txid60419%5borgn)]
4. On the Organism page you could also filter by Organelles. On page 14 you will find Castenea mollissima – if you click on the RefSeq number (NC\_014674.1) you will pull up the genomic sequence <http://www.ncbi.nlm.nih.gov/nuccore/NC_014674.1> of the chloroplast which is shown on the next slide.
   1. Since I am interested in just one of the genes, RubisCO – large subunit, I can scroll down through the annotations to try to find it.
   2. Or, I can use a different database that focuses on the individual proteins – UniProt is linked to ncbi. I I Google ‘UniProt Castenea mollissima rbcL’, this link is provided: <http://www.uniprot.org/uniprot/A0A068L542>, shown on the next slide.
5. If I scroll down the page, I come to the Sequence databases section. If I click on GenBank, I will get links – I want the genomic DNA sequence, not the translated (protein sequence). This takes me to this page: http://www.ncbi.nlm.nih.gov/nuccore/KF418893 , where I can see that the gene is 1388bp long.
   1. I want to locate this gene on the cp chromosome. If I copy the first 20 nucleotide (aagctggtgt taaagattat) into the complete cp genome I can use the ‘Find in this Sequence’ function on the right side of the cp genomic sequence page (shown on the next slide) and click find.
   2. The subsequence is highlighted (next slide) so you can see where your gene starts in the cp chromosome. (position 60,630).
6. I can look at a graphical view of this gene – go back to the top of the Chloroplast genome page and click on Graphics http://www.ncbi.nlm.nih.gov/nuccore/313183972?report=graph , shown in next slide.
   1. If you want to see the region of interest go to the Fin box (circled on the slide) and enter the start position noted above, as shown on the next slide. Note that rbcL is on the minus strand (in green), and the report shows it starting at position 60,590 (so there is a short untranslated region before the amino acids for the protein are coded).
   2. The rbcL bar code is 599bp long – the reference sequence is from the land plant Arabidopsis thaliana (you can retrieve this sequence by going back to the Genome/Organelle page, going to page 9 or <http://www.ncbi.nlm.nih.gov/nuccore/NC_000932.1> , go to position 54958-56397). To find the equivalent sequence in C mollissima cp genome use the start of that sequence (remember there may not be perfect matches so you may have to try a couple of times). I have as a sequence: atg tcaccacaaa cagagactaa agcaagtgtt.
   3. In this case I do get a perfect match, so it finds the sequence (if this did not work I could use BLAST, which will find less-perfect matches for me).