Name:_____ STR: DNA Analysis Lab

PRELAB ASSIGNMENT

1) Define STR:

2) What does PCR do and what are its steps?

3) If the statistical frequency for CSF1PO-12.0 is 0.3446 and the statistical frequency of D10S1248-14 is 0.2958, what are the chances of someone having both alleles?
   a) Decimal chance:_____
   b) 1 in _____

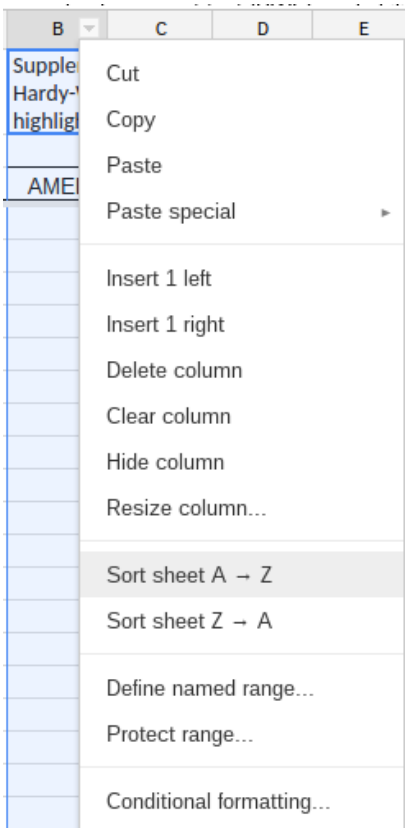
The Lab

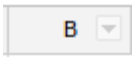
Go to this website **and read the description** for the 'STR Allele frequencies' spreadsheet.

- https://www.promega.com/products/pm/genetic-identity/population-statistics/allele-frequencies/

Then open up the google docs version of the spreadsheet here and mess with. **WORK WITH THIS ONE**

Working with a spreadsheet



Google sheets automatically makes everything a sortable table, meaning you can sort a column from smallest to largest (A-Z) or largest to smallest (Z-A) and it will rearrange the rows to keep the data together. This can be done by clicking the little drop down arrow next to the column label 

when your mouse hovers over it and selecting 'Sort sheet'. You will use this function to work with the data.

STR stands for Short Tandem Repeats and are regions of DNA where short sequences of DNA repeat a certain number of times. For example, in one region the sequence 'ACT' may repeat a specific number of times. These regions have no known function and vary greatly between individuals in how many repeats occur. For example; if STR gene 'X' had the sequence 'ACT' repeat 10 times, that allele would be called 'X-10'. Someone else may have 14 repeats of the X gene sequence 'ACT', and that allele would be called 'X-14' (more information on naming). It is these differences that are of forensic interest. In DNA profiling, 13 different STR genes on different chromosomes are typically analyzed to find which alleles a person/evidence has. The chart you are going to be working with has the

statistical data that you will be working with for how (un)common 29 of these STR genes are.

The vertical column titles are the names of the genes and the rows are the alleles, or versions of that gene. In the data table it will give you the decimal abundance (out of 1) for that allele. Remember that alleles are named 'the gene'-'the allele'.

Here is an example of what you may need to do:

What are the chances of someone having the D16S539-5 gene and F13A01-3.2 gene, in a decimal and '1 in _____' format?

- Find the abundances of each on the chart (0.0005 and 0.1390)
- Multiply the two values together. This will tell you the probability of both of them showing up together randomly:
    - .0005 X .1390 = .000695
- That is the decimal abundance, or how frequent it would show up out of 1. A more useful bit of data would be how often you might see this combination in a '1 in _____ people'. To get that data you simply take the reciprocal, or do: 1/'the decimal'
    - 1 / .000695 = 1439, or, 1 in 1439 people would be expected to have this combination.

<u>The Activity</u>

1) What is the most common allele for CSF1PO (sort column largest to smallest)?_____
2) What is the least common allele for CSF1PO (that actually occurs)?_____
3) Use the ctrl+f keyboard shortcut (that works in other programs as well) to find the loci and allele that has the frequency of 0.2336. What is it? _____-____
4) Select all of the cells <u>under</u> CSF1PO. Look at the bottom part of the window where the pages and 'Explore" can be found. It will show you some data such as count (how many cells have something in them), sum (the total of all the highlighted cell) and avg (the average of the highlighted cells). You can view all of this data by clicking on it when the cells have been selected.
    a. How many different alleles are there? (look at the count) _____
    b. What do they add up to be? (sum) _____
    c. Do the same thing for D5S818. What is the sum?_____
    d. Explain why the two sums come out to what they do

<u>Working with the data</u>

5) Examine the sheet for 'all data'. What are the chances that an individual would have CSF1PO-12.0 <u>and</u> D10S1248-14.0 alleles? (give decimal as well as 1:___ chances by taking the reciprocal of the decimal)

6) What about if it had D12S391-16.0 allele as well (decimal and 1:___) ?

7) And also D13S317-12.0 allele (decimal and 1:___)?

8) What happens to the probability of a random match as we use more alleles?

9) This data was collected for 'unrelated U.S. population samples'. Why do you think they chose unrelated samples?

10) Why is it important to specify that it came from the U.S. population?

Scenario: (based on an actual unsolved case)

The Schmiddletown Police Department has contracted you to be a part of their investigation. A 14 year old boy was found beaten, stabbed and slashed to death on the steps of Schmiddletown High School in 1986, before STR Analysis technology existed. The victim sneaked out of his house the night before and his body was found at 6AM when a custodian found him, with his body still warm. They believe that with the current advances in technology that this case is solvable, and they want you to help. You're job will be to perform a PCR-STR analysis of different possible DNA sources found on the victim in attempts to find a match in from a list of suspects.

Analysis

1) Now that you have run your DNA tests, check out your results on the Evidence tab in this spreadsheet!! You will use the results and the Student profiles tab in order to determine if there is a match, and if so, the significance of the match (what are the chances of the match occurring by chance?)

Sample: _____
Match: (YES / NO)
Matches with: _____
Significance: 1 in _____

2) You are called in to court to testify about your findings. Write out what you would plan on saying to the jury. Remember to keep your testimony scientific!