

Project: Ancestry Module Internship Project

We want you to make an ancestry script which inputs as arguments a file with an individual's genome, and outputs two files: a text with a vector/table of ancestries (person.txt/tab), and at minimum a (primitive) pie chart visualization.

Cohorts and Methods:

You're to follow [this tutorial](#). When all is done, you're to test it on genomes from the [HGDP](#).

Evaluation Criteria:

For accuracy, there's little one can do objectively evaluate ancestry - and it is of little health consequence this is somewhat subjective. I think that common sense is the only reasonable metric which need apply. The important thing will rather be consistency, and ready updatability.

Functionalization:

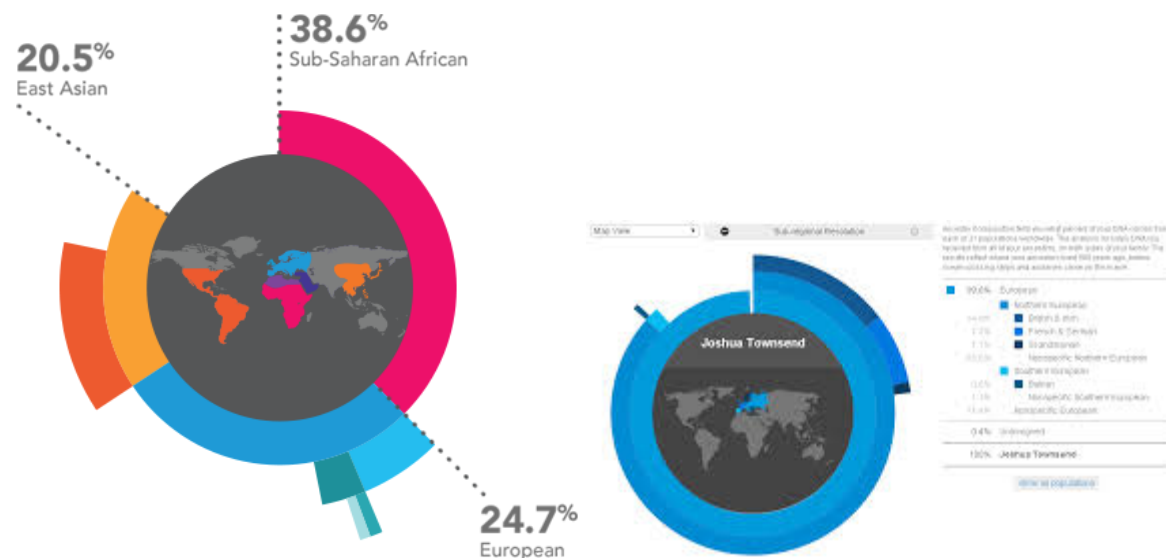
Function A: <ancestry-superpop> Input genome, output superpopulation fractions to a .tab file.

Function B: <ancestry-pop> Input genome, output subpopulation fractions to a .tab file.

Function C: <ancestry-piechart> Piechart visualization of each of the above ([R pie chart](#), fx).

Function D: <ancestry-popviz> Ring visualization of the above, in react:

Function C is pretty self explanatory: pie charts in pie charts. Function D is more subtle, and when you're getting close, we will hook you up with a React dev on our team, who will do omst of the coding in collaboration with you, and will evaluate your ability to work on a team. We want each region to be a clickable object, responsive to mouseover with a text popup, and you must supply much of the content and instruction for him. You must be able to tell him what to do.

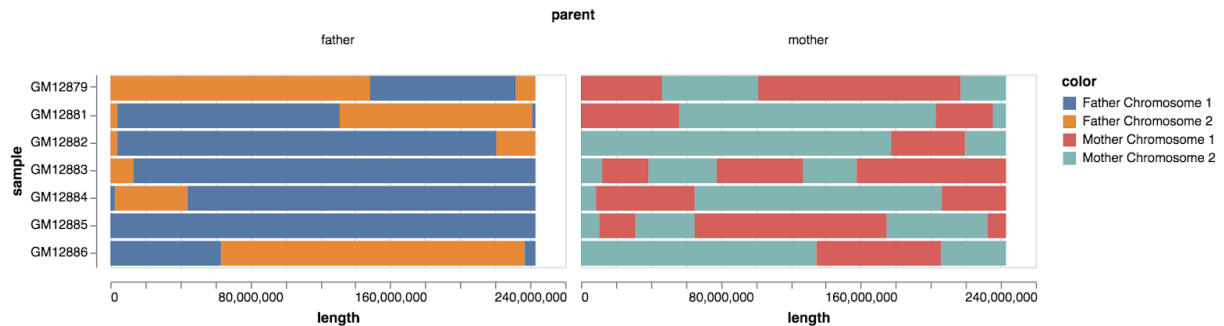


Function E: <ancestry-ancient> Neanderthal/Denisovan ancestry estimation components.

Later Functions: **Advanced Ancestry Module**

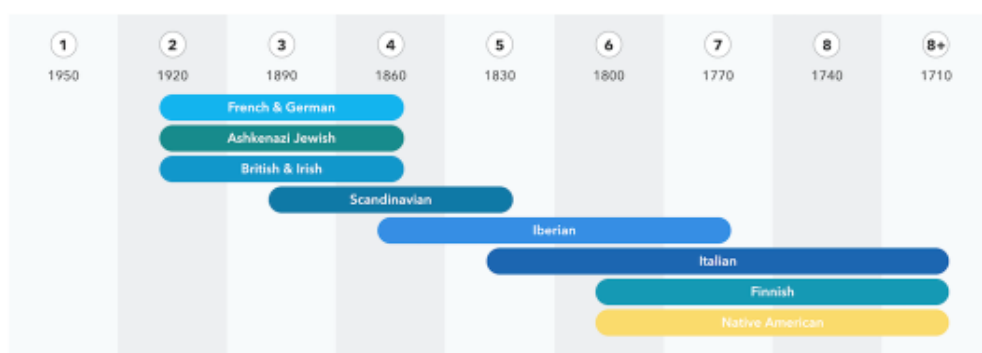
You needn't concern yourself with completing these during the internship period. We will want you to complete them after joining the team.

Function F: <ancestry-chrlevel> Input pregenome, output sub-chromosomal level ancestry. Then, visualize this in some simple way, using fx. R.



Function G: <ancestry-chrlevelviz> Visualization of the above, in react. Again, this is more subtle, and when you're getting close, we will hook you up with a React dev on our team, who will work with you, and provide his evaluation.

Function H: <ancestry-timeline> The most sophisticated kind of modern ancestry analysis allows us to use ML to estimate when admixture events took place. Needn't be done in react.



Internship Completion Criteria:

Criterion 1: Functions A-E are functional.

Criterion 2: You can have a script which takes a genome file as input, and converts it into a PDF.