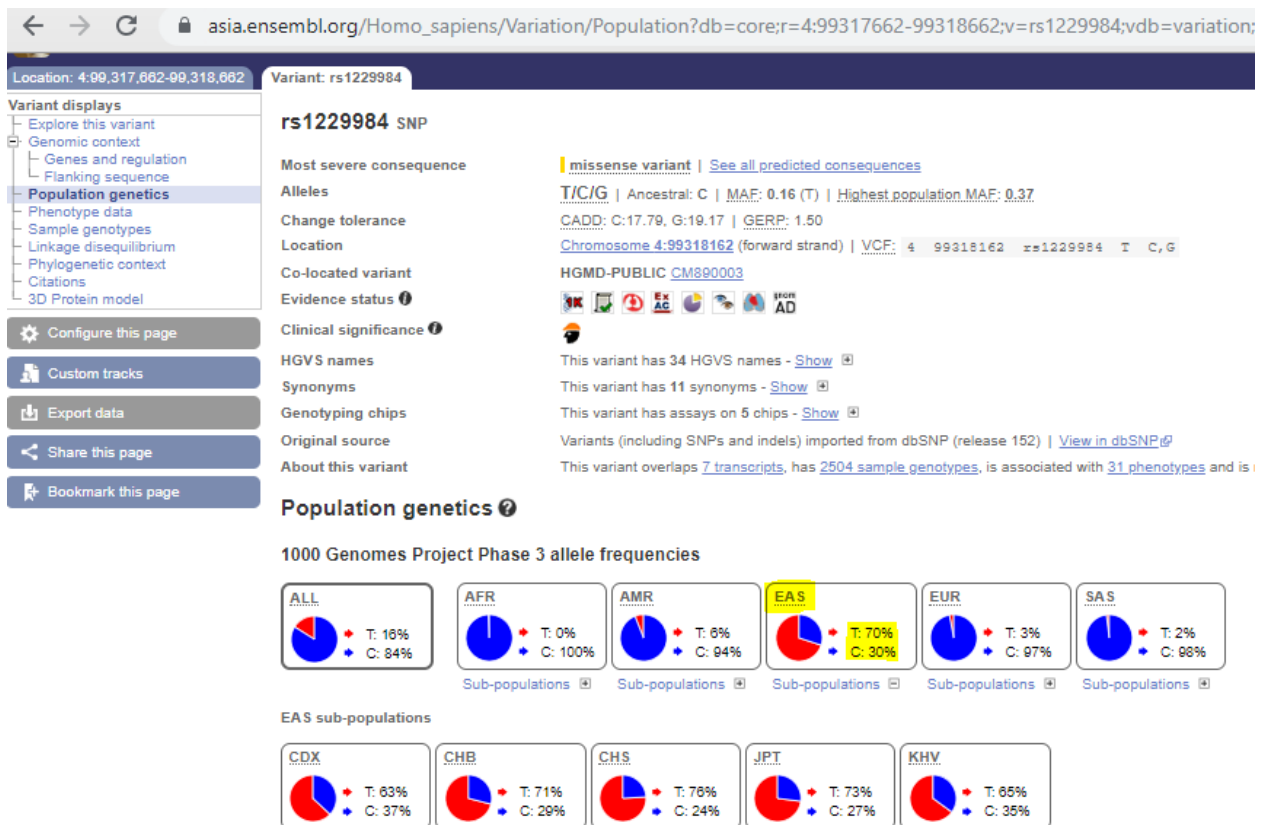


R Workshop Session 2 Exercise

Exercise 1:

The human genome is diploid, meaning that it carries two copies of an allele (one from each parent). The two alleles might be identical (homozygous genotype) or not (heterozygous).

rs1229984 is a missense variant in the ADH1B gene that has been associated with Esophageal cancer in Japanese (PMID:19698717) and with oral cavity and pharyngeal cancer in Europeans (PMID:21437268 and PMID:27749845).



Looking this SNP up on Ensembl (see above) shows the frequency of T-allele is 70% and C-allele is 30% in East Asians. If you were to screen 100 East Asians randomly, how many people do you expect to carry the:

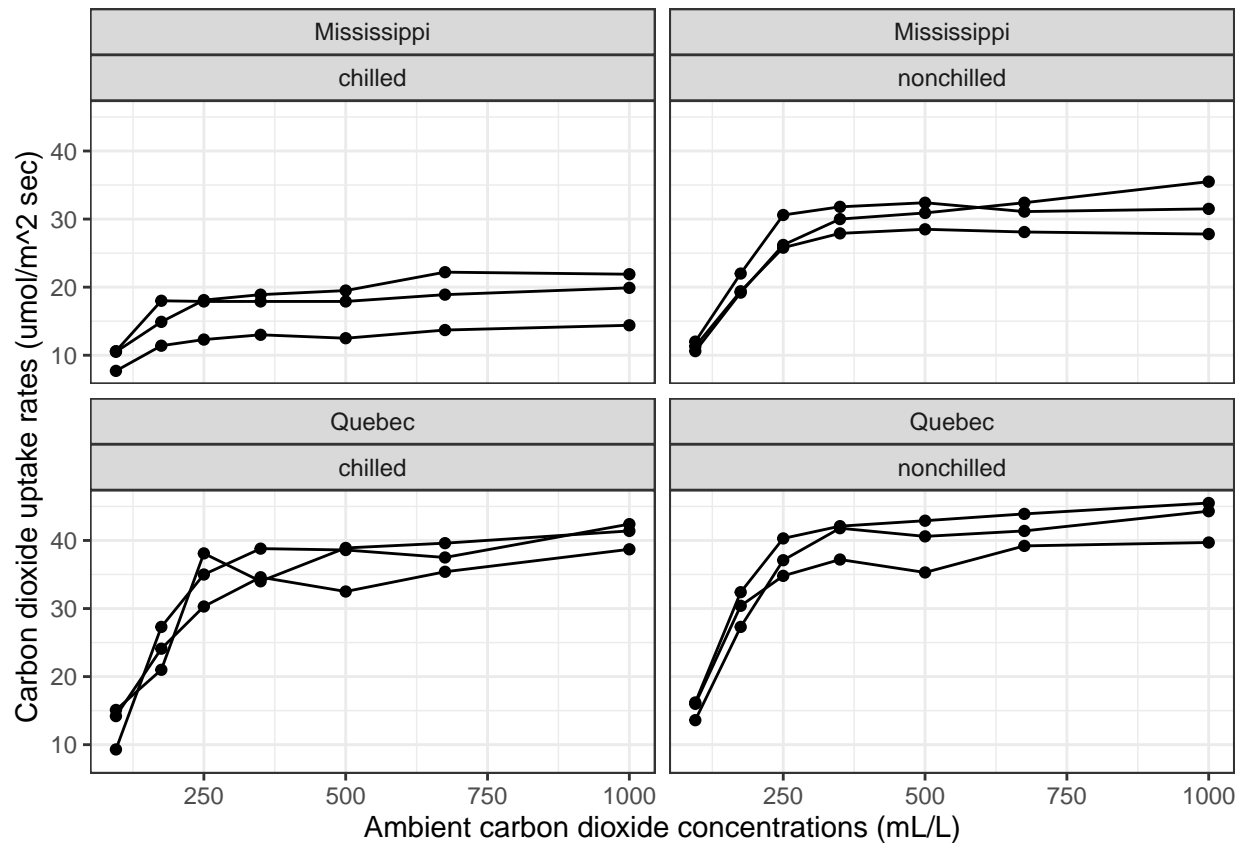
- TT genotype
- CT genotype
- CC genotype

Exercise 2:

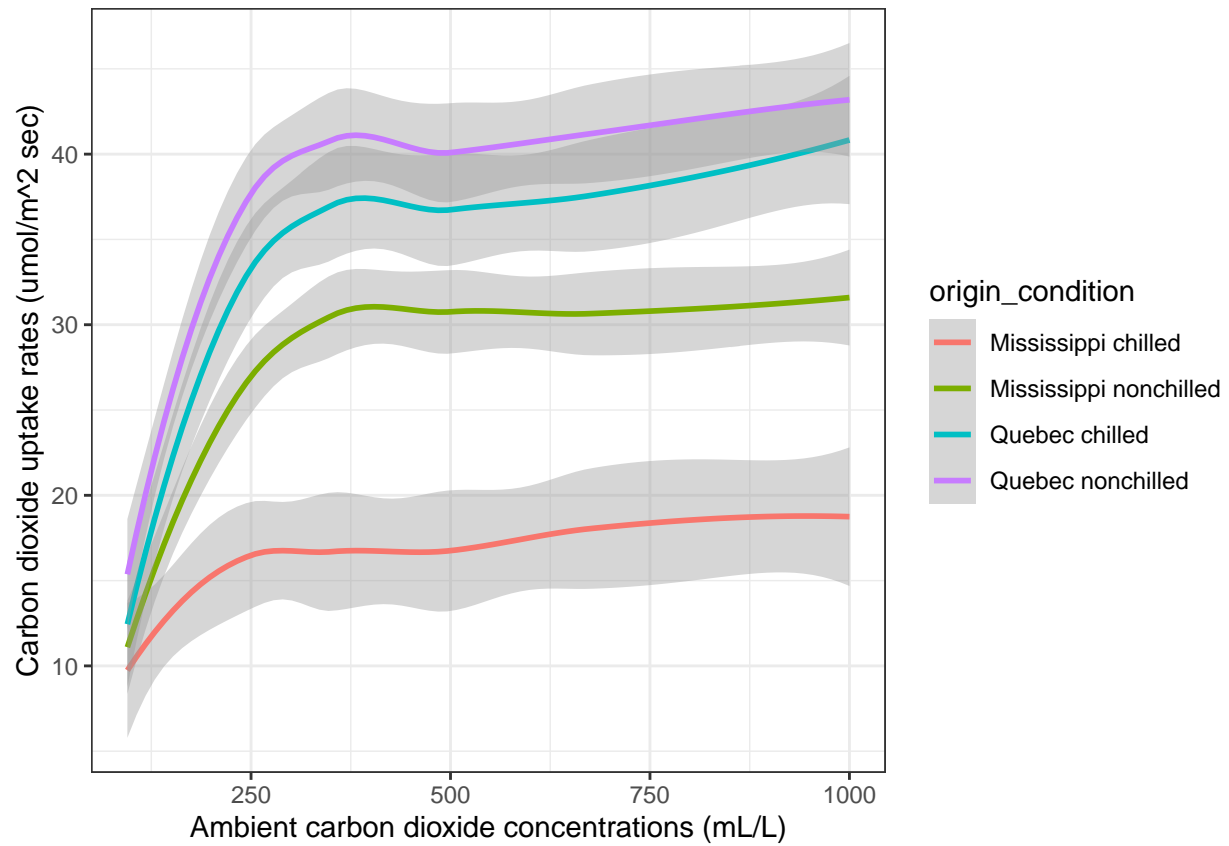
Background:

The data in CO2_uptake.csv comes from an experiment where carbon dioxide uptake in grass plants from two origins was measured with and without cold exposure.

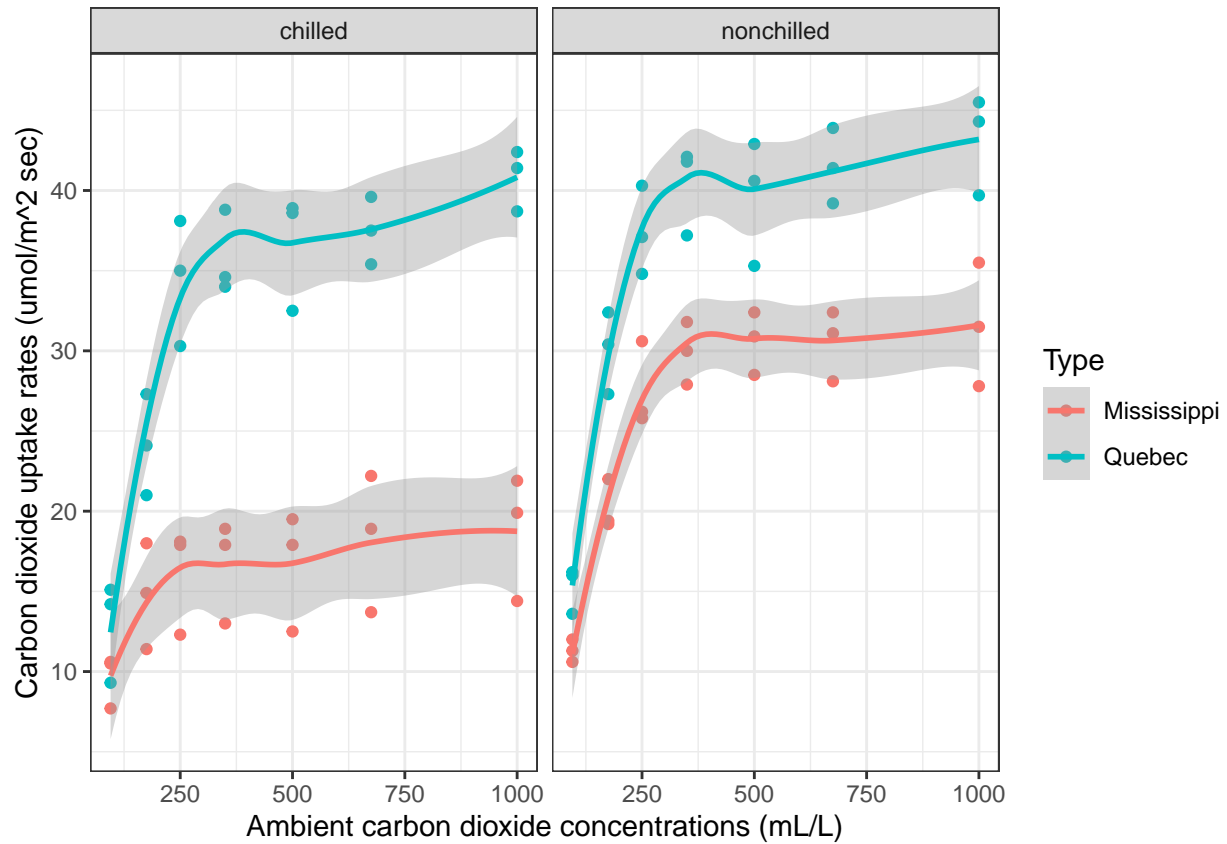
- Set your working directory, load your packages and clean your workspace. Then read in your data
- Let's visualize the data. Please reproduce the following graph:



- c. There seems to be very little variation between the plants in each condition. Let us simplify the graph to show the smoothed trendline representing the average and standard errors. Please reproduce the following graph:



d. Let's try another view of the same data. Please reproduce the following graph:



Exercise 3:

Background: The Puromycin data shows the reaction velocity versus substrate concentration in an enzymatic reaction involving untreated cells or cells treated with Puromycin.

- Read in the *puromycin_reaction.csv*
- Plot the reaction rate against the substrate concentration
- Repeat step b but with the concentration on a log10 scale