Data Overview

Tanya Strydom

2/7/2022

Goal

The purpose of this document is to give a bit of a more in-depth overview of what data we have for different combinations of sites, treatments, etc. This in turn can hopefully help in informing the game plan for moving forward with regards to what we want to do in the end with regards to the actual research question/aims.

High level issues

- The issue of needing to use samples from both the dry and wet season.
- Combining from different years (and imbalances regarding the number of samples we cave for site **x** treatment combos)

We also do not have any chemical traits data yet with no real time frame for when we could have them.

Potential directions to take

- Only look at species composition (don't work with traits data) and **possibly** spin this into a methods paper using the fuzzy phylogeny matching work that Eric has been doing (this will require re-working the context of the manuscript)
- Try and make the dataframe less Frankenstein-esque by using something like bootstrapping/simulations to try and even out the number of samples as well as bypass needing to use data from different season (possibly all years???)
- Try and make the best of a bad situation and soldier on - although we woudl still need to decide on waiting on the traits data or not.
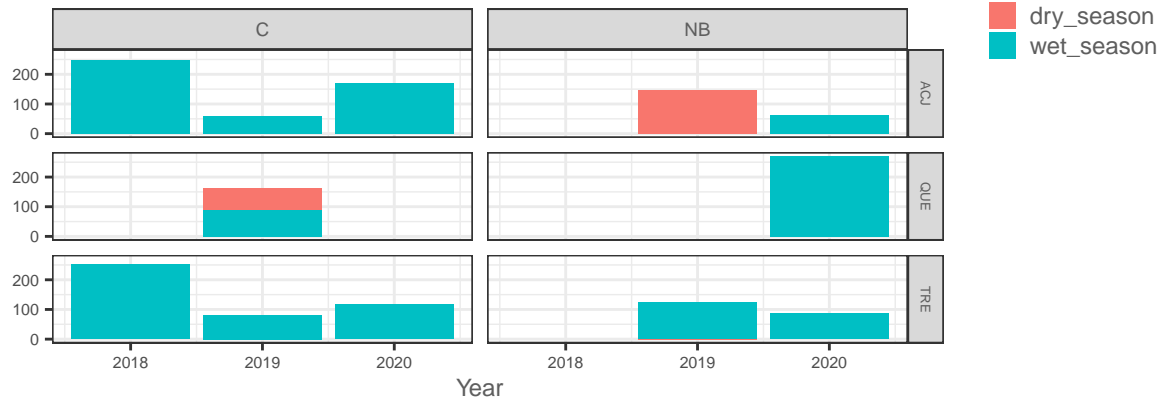
Data Summary

Below is the number of individuals sampled for *traits* across the different treatment, site, year and season combinations.

Number of samples/individuals
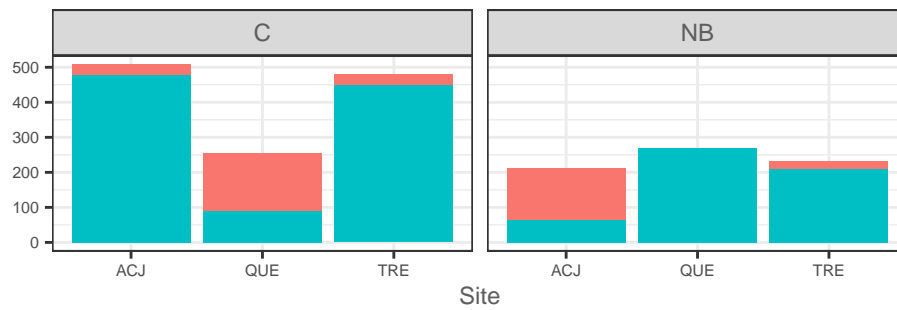
Number of samples by season and year

Coverage

All individuals are 'complete'. That is for the following traits: `plant_height_cm`, `leaf_area_cm2`, `sla_cm2_g`, `ldmc`, `leaf_thickness_mm` each individual (based in their id) has a recorded value.

Sampling Effort