Capstone Two - Project Proposal

Background:

Nitrogen is one of the most important minerals in the plan’s life. It affects many processes and have a direct impact on plants growth and development. Low nitrogen supply will limit the plant’s growth, while high nitrogen supply will be toxic to the plant and will result in growth defects.

Glucosinolates (GSLs) are defense metabolites produced in Cruciferous plants. Plants can synthesize ~40 different structures of these molecules in different amounts, in response to different environmental ques. Among those ques are herbivores, drought, and nutrient availability.

Several lines of evidence suggest that there is a link between GSLs content and nitrogen availability:

1. All GSLs contain an atom of nitrogen as part of their basic structure. Hence, nitrogen is a limiting factor in the synthesis of GSLs.
2. Plants that are defected (mutated) in some GSLs processes present roots phenotypes that depend in nitrogen conditions.
3. A recent study showed that GSLs affect nitrogen homeostasis in the seeds.

Problem statement formation:

We want to understand the connection between these two elements – nitrogen and GSLs, how they affect each other, and how they are affected by different environmental parameters.

Context:

Understanding these relationships will enable us to understand how plants integrate between growth (as a result of nitrogen conditions) and defense (GSLs) and thrive under changing environments. The results of this analyzes may open the door for a variety of agriculture usages, among them is the development of crops that accumulate biomass by using their nitrogen source more efficiently and at the same time efficiently defend against herbivores, and all of that is happening under environmental changes.

Criteria for success:

Detecting correlations between parameters in the different elements (nitrogen and GSLs) and correlations between them and different environmental parameters will enable us to understand and describe the relationship between these elements, and how they are connected to each other, and affecting each other.

Scope of solution space:

There are three groups of elements that I will try to find correlations among them:

1. Developmental phenotypes: in my dataset I have ~10-15 developmental phenotypes (roots and shoots phenotypes) that were measures for each plant that grew under each of the nitrogen treatments (1100 plants, under 4 nitrogen treatment). These phenotypes are an indication to the effect of the nitrogen source on the plant’s development and growth.
2. Biochemical phenotypes: in my dataset I have ~20-25 biochemical phenotypes (GSLs measurements) that were measured for each plant that grew under each of the nitrogen treatments (1100 plants, under 4 nitrogen treatment). These phenotypes are an indication to the effect of the nitrogen source on defense molecules (GSLs).
3. Environmental conditions: I have a dataset of ~200 different environmental conditions (temp, humidity, rainfall, minerals etc.) under different locations around the world. The plants that I tested were collected from different locations, with different environmental conditions. Using this dataset, I will assign each plant a set of specific condition and will be able to test if and how these conditions affect the developmental and biochemical phenotypes.

Constraints:

There are a few constrains that I might face while doing this analyzes. The main one would be that we will not find correlations using a simple correlation analyzes based on each phenotype. There are a few ways to overcome this issue. One way would be to try different combinations of phenotypes/ conditions to test the correlations. Another option would be to test the correlations separately for each nitrogen condition.

Another constrain would be to understand the correlations that we will get, and to understand how we can apply it in a practical way.

Stakeholders:

We would eventually want to publish our conclusions in the scientific literature. Any conclusion that we might have will have to be backed up with a strong model, otherwise it will not be published.

Data sources:

1. Nitrogen conditions and phenotypes: this dataset contains data on 1100 plants with different genetic backgrounds, that grew on 4 different nitrogen conditions (2 sources, 2 concentrations). The data includes phenotypes for each seedling under each nitrogen condition, under two categories: developmental phenotypes (roots and shoots phenotypes) and biochemical phenotypes (GSLs measurements).
2. Environmental data: the different accessions (1100 accessions, with different genetic backgrounds) were collected from different locations around the world. This dataset contains information for different environmental parameters for each location (weather, geography, minerals etc.).
3. Seeds GSLs data: this dataset contains GSLs information on each one of the seeds of the 1100 accessions, and not the plants when they grow under the different nitrogen conditions. Might serve as a GSL reference.