**RESEARCH PLAN**

**Aim 1. The nutrient-to-phenome matrix: correlation of NPK nutrient combinations to phenotypes.**

**Rationale**: In this aim, we will create and analyze a nutrient-to-phenome matrix to i) identify early morphometric seedling markers as biomass predictors and ii) to correlate NPK matrix combinations with low and high biomass yields. As for all plants, Arabidopsis biomass is sensitive to N-limitation (REFs), and we will exploit this feature to investigate the effect of a nutrient matrix of low, intermediate and high concentrations of N combined with low and high concentrations of the macronutrients potassium (K) and phosphorus (P), to select NPK matrix combinations that result in I) High-N:High biomass, II) Low-N:Low biomass and III) LowN:High biomass. To begin, Arabidopsis will be grown on all NPK matrix combinations and we will perform morphometric analysis on seedlings to identify associated traits using Principal Component Analysis (PCA). We will also measure biomass of adult plants (shoots and roots), as well as other N-use parameters (e.g. chlorophyll). The integrated analysis of these datasets (e.g. NPK matrix, morphometrics, and biomass) will allow us to identify NPK treatments and growth strategies that represent the adaptive response to N-use efficiency (Aim 2). In addition, we are going to test whether the growth strategy resulting from a particular NPK combination could be predicted early in seedling development by examining the correlation between plant allometry and biomass.

***Pilot Experiment:*** Arabidopsis (Col-0) seeds were germinated on N-free as well as MS media containing increasing concentration of N, to establish plant growth parameters on “low”, “intermediate” and “high” levels of inorganic N. Nitrate was used as the sole source of N, as it was previously shown that nitric nutrition is more effective than ammonium (NH4+) or ammonium/nitrate nutrition for sustaining Arabidopsis growth and biomass production (M’rah Helali, Nebli et al. 2010). Seedlings were grown for 6 weeks in short-day conditions (8 light-16 dark), to increase vegetative growth (biomass) and prevent flowering. The plants were grown on increasing concentrations of KNO3 (0.05–20mM) and fresh weight/N-available in the media was determined as an index of N-use efficiency (REF).

**Aim 1A**. **Optimizing NPK matrix of treatments**. To generate testable biological hypotheses on how N, P and K signaling interplays to regulate plant growth, we will test a complete matrix of NPK treatments that represent all combinations of low, intermediate and high N (as derived from our pilot experiment) with low versus high P and/or K while keeping the proper balance of the other mineral components in the proportions determined for MS medium (REF). The goal is to effectively span the widest range of the possible effects on Arabidopsis growth to select the NPK: phenotype states that will be analyzed in Aim 1B. Arabidopsis seedlings will be grown on vertical plates for 14 days for morphometric analysis of the root and shoot morphology (REF). In addition, we will measure root and shoot biomass production in hydroponically grown 6 week old plants as per (REF). The integration and analyses of these data (matrix, phenotype, biomass) will allow us to capture the variation due to different NPK treatments and distinguish alterative growth strategies at seedling and mature stages as driven by NPK signaling.

**Aim 1B. Quantifying and integrating the NPK matrix effect on plant phenotype and biomass.** To acquire a systems-wide view of the growth strategies that are driven by the matrix of NPK nutrient/signals, we will use morphometrics analysis and multivariant statistical methods. The quantitiative phenotype analysis of Arabidopsis seedlings will be carried out using the AAMT toolbox plugin for Matlab (REF). AAMT is a landmark-based geometric morphometric method in which primary and secondary landmarks are placed on the object to be analyzed at recognizable features. Mention shoot and flower paper as ref. This method - originally used for face recognition - has been applied to quantify changes in shoot architecture in Arabidopsis and it can capture the geometry of roots and shoots to then convert these morphometric measurements into Principal Components (PCs) [Please don’t use this acronym. PC = personal computer to everyone. Nobody in this area uses PC for principal component] (REF). In the case of Arabidopsis root morphometrics, 20 landmarks include 6 primary landmarks (e.g. the root-hypocotyl junction), and 14 secondary landmarks that are regularly spaced between the primary landmarks by the AAMT software (Fig. X) (REF for AAMT). Differences in root geometry between seedlings can be described by the difference in coordinates of corresponding landmarks. The resulting sets of coordinates are aligned (Procrustes Alignment), and subjected to principal component analysis by the AAMT software (REF). The first Principal Component (PC) accounts for as much of the variability in the data as possible, and each succeeding PC accounts for the remaining variability. The top-ranked PCs provides a quantitative measure of variation in root or shoot morphology across different NPK matrix conditions and will allow us to assess the quantitative allometry of root and/or shoot differences, and to synthesize the relationship between shoots and roots in a unified statistical model. Morphometric analysis will be carried out on 1- and 2-week old seedlings, which represent early stages of plant development. This approach has been successfully used in our laboratory to analyze the effect of combinations of N and plant hormones on Arabidopsis roots, and it is proving to be a valid method to capture most of the variability as the conventional measurement techniques and to identify new meaningful underlying variables of phenotypes (REF) Ulises paper.

To measure the effect of varying concentrations of NPK on biomass production at later developmental stages, we will measure the fresh weight of the rosettes and roots of 6 week-old plants to evaluate biomass partitioning for each NPK treatment. We will also measure soluble sugar and protein content, because of their close relationship with fresh weight. For the shoot, we will also measure chlorophyll content an index of nitrogen use (REF).

**Aim 1C. Correlation between biomass and morphometric** **analysis for early prediction of biomass production.** For eachNPK combination, the morphometric analysis will provide a quantitative measure of the effect of the matrix on the morphology of young seedlings while biomass will represent the result of prolonged growth in the same conditions. We will use stochastic gradient descent to establish relationships between the morphometric PCs as well as classic root/lateral root measurements (e.g. …) and the biomass measurements and identify early predictors of biomass. That is, stochastic gradient descent [refs] will identify the values of the coefficients (a1, a2, and a3 as well as b1 through b??) of a linear equation of the form  
biomass = a1\*PC1 + a2\*PC2 + a3\*PC3 + b1\*PrimaryRootLength + b2\*LateralRootLength + b3\* … We will use stochastic gradient descent with a regularization term (to reduce the tendency to overfit) and a training coefficient eta. To discover the proper values of eta, we will use 10-fold cross-validation with different parameter settings. In each “fold” of cross-validation, we withhold 10% of the plants from the training set and then evaluate our results on the left-out plants (the test plants). Different folds differ based on which plants are chosen as test and which as training. The regularization and training coefficients that offer the best results on cross-validation will then be used on all the data to determine the coefficient values. To evaluate the confidence interval of the coefficients, we will sample the plants with replacement to determine the range of values of each coefficient. The net result of all this analysis will be the identification of the features (whether principal component or classic root/lateral root measurement) that are the best early predictors of biomass. If we are interested in the coefficients for root and shoot biomass separately, we can reproduce the analyses for htem.

**Aim 1D.** We will also develop a method to classify the NPK combinations according to the amount of N in the medium and the biomass that is produced to determine the NPK:phenotype states that represent N-usage strategies (High-N:High biomass, Low-N:Low biomass, LowN:High biomass – should we include High-N:Low biomass as an example of poor NUE?)

Dennis- please fill in. Dennis does not believe this needs to be a separate subaim but rather aim 1b. All we are doing here is measuring average biomass in each NPK combination and taking the best of the high-Ns, the best of the low\_Ns, and the lowest of the low\_Ns.