**Fig. 1 The VirtualPlant Multinetwork.** The Arabidopsis multinetwork contains genes represented as nodes (A) that are connected by edges of many types (B) including metabolic, protein-DNA, protein-protein, microRNA-RNA, and edges derived from text mining [Katari et al 2010]. (C) shows a network neighborhood resulting from querying this multinetwork with microarray data, uncovering a regulatory hub (CCA1) involved in nitrogen signaling [Gutierrez et al 2008 PNAS].

**Fig. 2. Phylogenomic tree of 21 fully-sequenced plant genomes and expression datasets.** The total evidence tree shown here, creating using OrthologID [Chiu, JC, Lee, EK, Egan, MG, Sarkar, IN, Coruzzi, GM, and DeSalle, R, *OrthologID: automation of genome-scale ortholog identification within a parsimony framework.* Bioinformatics, 2006. **22**(6): p. 699-707.4] is the most parsimonious tree generated from simultaneous analysis matrix, using combination of drifting, rachet, and fusion in TNT [Goloboff, PA, Farris, JS, and Nixon, KC, *TNT, a free program for phylogenetic analysis.* Cladistics, 2008. **24**(5): p. 774-786.5], as used in the text of Aim 1 and in [Lee E, Katari M, Kolokotronis S, Cibrian A, Stamatakis A, Ott M, Little D, Stevenson D, McCombie WR, Chiu J, Martienssen R, Brenner E, Coruzzi G, DeSalle R (2011) “High resolution phylogeny of the seed plants: A functional phylogenomic view.” ***PLoS Genetics*** Dec;7(12):e1002411. Epub 2011 Dec 15]. Expression data for each species is shown as a pie chart, whose size is proportional to the data; Blue (Affymetrix data), Red (Next-Gen RNA-seq data).

**Fig. 3. Infernet: A machine-learning approach to inferring gene networks.** The “robin hood” approach to network inference consists of learning a regression model from each of several data-rich species to apply with a combining rule to data poor species.

**Fig. 4 The InferNet algorithm: Testing Precision and Recall:** When starting from a single data-rich species (e.g. Arabidopsis), we learn our model as a set of coefficients on orthology, correlation and p-value using another data rich species (Medicago) and then predict edges in Soy. For the sake of this preliminary study, we can measure precision and recall because Soy itself is data rich. (See Table I).

**Fig. 5. A workflow for trait-to-gene “weighted” networks.** The workflow for mining expression data associated with crop traits to drive “weighted” networks in the data-rich models for validation testing in Arabidopsis and Maize (see Aim 2B)**.**

**Fig 6. A prototype BUI (Biologist User Interface) for X-net**. The first row shows some options available to the plant biologist who wants to generate a predicted network for Species X. These options include selecting a “source” species, a “target” species, an orthology method, and a type of edge. The second row shows the different types of networks that can be created. A researcher who wishes to use InferNET must select at least one species for training, whereas this is not necessary for Interolog. Researchers can also upload their own experiments from which a correlation network will be created using the different options the researcher provides. The third row shows how a researcher can create a “weighted” network by combining different networks from different species. The text field near the different edge types allows the user to provide their own weight to the edges. The output here again is merged network which the user can visualize using Cytoscape (Shannon et al).

DENNIS- PLEASE PICK ONE VERSION OF TABLE 1 LEGEND (There were two in text) Done

**Table X**: Positive recall is the number of gene pairs in the target species correctly predicted to be positively correlated divided by the number of gene pairs that are positively correlated. Positive precision is the number of gene pairs correctly predicted to be positively correlated divided by the total number predicted to be positively correlated; similar for negative correlation. The coefficient of the percent identity score is 0.03, for the magnitude of the correlation is 1.2, and for the raw p-value (which is normally very small) of correlation is -0.14. The Interolog approach assumes that an edge in Soy that is orthologous to a positively (respectively, negatively) correlated edge in Arabidopsis will be positively (respectively, negatively) correlated.