**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 regulatory hubs (e.g. CCA1) involved in nitrogen signaling [Gutierrez et al 2008 PNAS].

**Fig. 2 Phylogenomic tree of 21 sequenced plant genomes and expression datasets.** The total evidence tree shown here, was created as described in the text of Aim 1, using methods 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 (used in Aim 1) 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). Numbers indicate number of experiments for Microarray Data/Next Gen Data. Data rich species: Arabidopsis, Poplar, Medicago, Soybean, Rice, and Maize.

**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 using Soy experimental data, 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 (Aim 2A), for validation testing in model and crop (Aim 2B)**.**

**Fig. 6 A prototype Interface for “X-net: A network learning platform”**. 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 step 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 a weight to the edges. The output here again is a merged network which the user can visualize using Cytoscape [ Shannon 2003: Cytoscape: a software environment for integrated models of biomolecular interaction networks.  Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T.  [Genome Research 2003 Nov; 13(11):2498-504](http://www.genome.org/cgi/content/full/13/11/2498) ].

(Dennis- Please check my edits of table 1 especially)

**Table I**: **Validation testing of predicted networks in Soy (Glycine max) using InferNet vs Interolog (see Aim 1)**. 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 validated to be positively correlated based on experimental data. Positive precision is the number of gene pairs correctly predicted to be positively correlated divided by the total number predicted to be positively correlated; similarly, negative recall and precision pertain to negatively correlated edges. In this pilot study, the coefficient of the percent identity score used for orthology 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 [Yu (2004) Genome Research,Annotation Transfer Between Genomes: Protein–Protein Interologs and Protein–DNA Regulogs], assumes that an edge in Soy that is orthologous to a positively (respectively, negatively) correlated edge in Arabidopsis will be positively (respectively, negatively) correlated.