**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, 2009 #1}. (C) shows a network neighborhood resulting from querying this multinetwork with microarray data, uncovering a regulatory hub (CCA1) involved in nitrogen signaling {Gutierrez, 2008 #7}.

**Table I**. **Quantitative Information about the Edge Types of the Arabidopsis Multinetwork**. The multinetork is described further in {Katari, 2009 #1}.

**Fig. 2. The VirtualPlant Arabidopsis and Rice Home Pages.** TheVirtualPlant software platform ([www.virtualplant.org](http://www.virtualplant.org)) is designed to support multiple species {Katari, 2009 #1}. Shown are the two home pages for Arabidopsis and Rice. Each supports a common set of tools but is implemented on top of a separate database. An analysis within a species will not be slowed down by the addition of another species.

**Fig. 3. Network inference and Validation Pipelines for** **Cross Species Network Inference (CSNI).** The cross species network inference suite of Aim 3 will unify the techniques of Aims 1 and 2 into a full workflow for under-analyzed species: (Panel A) a CSNI inference pipeline, (Panel B) Time-series data collection, analysis and prediction and (Panel C) Validation. **Panel A**: The depicted CSNI pipeline will combine inference based on homology from a reference species and experimental data in a target species to obtain an *inferred* network in the target species. **Panel B**: Aim 2 provides a method to infer a network using closely spaced time series data by using State Space modeling.Validation is based on prediction on unsampled data. **Panel C**: When a validated network is available in the target species, the inferred network can be evaluated by using the pipeline from Aim 1 (Steps 5 & 6). Such validations will lead to improvements to homology parameters and inference rules as well as an inferred mulinetwork.

**Table II. Validation of Network Inference using Arabidopsis (reference) and Rice (target).** Inferring interaction relationships in Rice based on homology alone (to Arabidopsis) data (using Rice expression data) yields high precision relative to the “ground truth” network (of Rice). Combining the two (homology and correlation) gives even greater precision at some cost in recall. The use of the asterisk (\*) connotes statistically significant improvement in precision

**Fig. 4. State-Space modeling: A machine learning approach to network inference.** State-space modeling fueled by regulatory data (transcriptome depicted as heat map) at closely spaced time points seeks to explain the expression of a target gene X, as a function of the expression of one or more other regulatory genes (e.g. transcription factors, TFs) as a fixed relationship (*f*) between genes. Even though *f* is fixed, gene expressions can vary in value because their input genes (e.g. TFs) as well as signals to which they respond (e.g. nitrogen signals) can change over time. The function *f* is "simple" (or "regularized") in the sense that each target gene in the model is forced to depend on no more than three or four input TFs. Function *f* is computed through a cyclic series of steps of the form guess, compute error, and then refine f, using the time-series regulatory data {Krouk, 2010 #19;Mirowski, 2009 #56}

**Fig. 5. Prediction and Validation Using a Network Constructed from State-Space Modeling.** In the state space model depicted in (A), each node represents the values of all gene expressions at a particular time point. The set of all gene expressions at time t is modeled by a “latent” (i.e., hidden) variable (denoted Z(*t*)) from which noisy and sometimes missing observations Y(*t*) are made. An unknown *function* *f* (represented by red square) relates the values of latent variables Z(t) and Z(t+1) (for all *t*) corresponding to consecutive time measurements. In (B) Validation of the predictive modeling is tested by the ability to accurate predicting the direction of change of each gene at 20 min for state space modeling (B) vs. trend-forecast test (C) {Krouk, 2010 #19;Mirowski, 2009 #56}.

**Fig. 6. The Cross Species Network Inference (CSNI) User Interface.** A biologist user selects a target and reference species. Next the user selects a homology technique and its parameters. Finally, the biologist selects a “ground truth” data set from the reference species, as well as experimental data in the target species to aid in the prediction of inferred gene regulatory networks in the target (e.g. crop) species. See the workflow in Fig. 3, panel A.