

Gene Function Discovery: New Computation Model Predicts Gene Function

ScienceDaily (Feb. 1, 2010) — Scientists have created a new computational model that can be used to predict gene function of uncharacterized plant genes with unprecedented speed and accuracy. The network, dubbed AraNet, has over 19,600 genes associated to each other by over 1 million links and can increase the discovery rate of new genes affiliated with a given trait tenfold. It is a huge boost to fundamental plant biology and agricultural research.

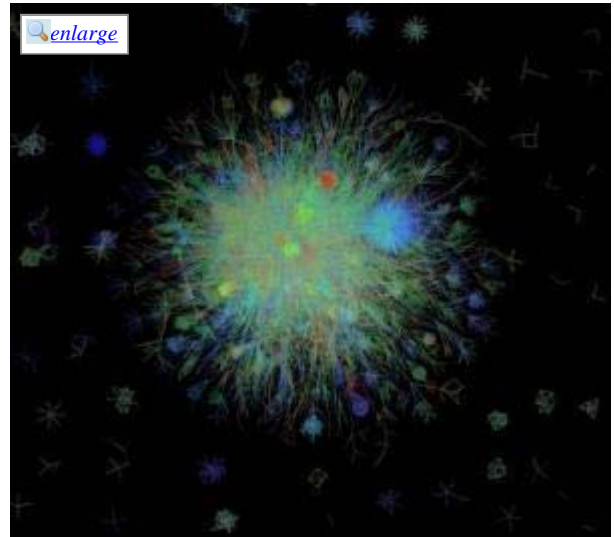
Despite immense progress in functional characterization of plant genomes, over 30% of the 30,000 *Arabidopsis* genes have not been functionally characterized yet. Another third has little evidence regarding their role in the plant.

"In essence, AraNet is based on the simple idea that genes that physically reside in the same neighborhood, or turn on in concert with one another are probably associated with similar traits," explained corresponding author Sue Rhee at the Carnegie Institution's Department of Plant Biology. "We call it guilt by association. Based on over 50 million scientific observations, AraNet contains over 1 million linkages of the 19,600 genes in the tiny, experimental mustard plant *Arabidopsis thaliana*. We made a map of the associations and demonstrated that we can use the network to propose that uncharacterized genes are linked to specific traits based on the strength of their associations with genes already known to be linked to those characteristics."

The network allows for two main types of testable hypotheses. The first uses a set of genes known to be involved in a biological process such as stress responses, as a "bait" to find new genes ("prey") involved in stress responses. The bait genes are linked to each other based on over 24 different types of experiments or computations. If they are linked to each other much more frequently or strongly than by chance, one can hypothesize that other genes that are as well linked to the bait genes have a high probability of being involved in the same process. The second testable hypothesis is to predict functions for uncharacterized genes. There are 4,479 uncharacterized genes in AraNet that have links to ones that have been characterized, so a significant portion of all the unknowns now get a new hint as to their function.

The scientists tested the accuracy of AraNet with computational validation tests and laboratory experiments on genes that the network predicted as related. The researchers selected three uncharacterized genes. Two of them exhibited phenotypes that AraNet predicted. One is a gene that regulates drought sensitivity, now named Drought sensitive 1 (Drs1). The other regulates lateral root development, called Lateral root stimulator 1 (Lrs1). The researchers found that the network is much stronger forecasting correct associations than previous small-scale networks of *Arabidopsis* genes.

"Plants, animals and other organisms share a surprising number of the same or similar genes -- particularly those that arose early in evolution and were retained as organisms differentiated over time," commented a lead and corresponding author Insuk Lee at Yonsei University of South Korea. "AraNet not only contains information from plant genes, it also incorporates data from other organisms. We wanted to know how much of the system's accuracy was a result of plant data versus non-plant derived data. We found that although the plant linkages



Each line of this AraNet network represents a functional link between two genes. The colors indicate the strength of the link using a red-blue heat map scheme. The image includes about 100,000 functional links made among about 10,000 Arabidopsis genes. (Credit: Image courtesy Sue Rhee)

provided most of the predictive power, the non-plant linkages were a significant contributor."

"AraNet has the potential to help realize the promise of genomics in plant engineering and personalized medicine," remarked Rhee. "A main bottleneck has been the huge portion of genes with unknown function, even in model organisms that have been studied intensively. We need innovative ways of discovering gene function and AraNet is a perfect example of such innovation.

"Food security is no longer taken for granted in the fast-paced milieu of the changing climate and globalized economy of the 21st century. Innovations in the basic understanding of plants and effective application of that knowledge in the field are essential to meet this challenge. Numerous genome-scale projects are underway for several plant species. However, new strategies to identify candidate genes for specific plant traits systematically by leveraging these high-throughput, genome-scale experimental data are lagging. AraNet integrates all such data and provides a rational, statistical assessment of the likelihood of genes functioning in particular traits, thereby assisting scientists to design experiments to discover gene function. AraNet will become an essential component of the next-generation plant research."

The research is published in the January 31st, advanced on-line *Nature Biotechnology* and was supported by the Carnegie Institution for Science, the National Research Foundation of Korea, Yonsei University, The National Science Foundation, the National Institutes of Health, and the Packard Foundation.

Email or share this story:

| [More](#)

Story Source:

Adapted from materials provided by [Carnegie Institution](#), via [EurekAlert!](#), a service of AAAS.

Need to cite this story in your essay, paper, or report? Use one of the following formats:

APA

MLA

Carnegie Institution (2010, February 1). Gene function discovery: New computation model predicts gene function. *ScienceDaily*. Retrieved February 3, 2010, from <http://www.sciencedaily.com/releases/2010/01/100131142436.htm>

Note: If no author is given, the source is cited instead.