

MARCH 10, 2006

New Recipe for Discovery: An Online Blend of Worms, Flies, Yeast

Howard Hughes Medical Institute scientists have developed a new computational approach that integrates data about how genes interact with each other in roundworms, fruit flies and yeast. They hope their database will become a valuable resource for biologists, who might first do a quick online query to learn new clues to how one gene might influence another before doing experiments to determine a gene's function.

The researchers built the new database because understanding how genes suppress or enhance the activity of one another is among the great challenges in genomics, said HHMI associate Weiwei Zhong and HHMI investigator Paul Sternberg. Zhong and Sternberg, who are both at the California Institute of Technology, published an article describing their approach in the March 10, 2006, issue of *Science*.

"The overall lesson we learned □ which was very pleasing □ was that there are clearly good databases out there; and we in the research community really know more than we thought we did."

— Paul W. Sternberg

Genes are crucial for specifying all properties of organisms — how they develop, how they behave, how they respond to their environment, said Sternberg. But genes don't act alone; they work in combination with other genes. So to get a handle on how the genome controls an organism, we need to understand genetic interactions.

Such interactions can take place on different levels. For example, the products of a genetic blueprint — RNA or proteins — can affect the activity of other gene products directly. Or the effects might be indirect, such as when one gene controls a biological regulatory pathway that alters a different regulatory pathway involving another gene.

A major problem facing researchers is that deciphering the function of one's particular gene of interest might be difficult, such as when that gene is part of

a redundant network. In such cases, perturbing the activity of that gene — a typical approach to exploring gene function — may not reveal its true function because there are other factors that compensate for the loss of that gene. But knowing in advance how a given gene fits into an interacting network would prove invaluable to researchers in planning their experiments.

But attempting to analyze interactions between gene pairs experimentally is a gargantuan task. A typical multicellular organism has some twenty thousand genes, Sternberg said. Even more overwhelming, however, is the fact that these genes could potentially have two hundred million pairwise interactions. So, that alone would mean two hundred million experiments. And, in fact, all those interactions might not even be occurring in a given cell or process. So, experimentally determining all the possible gene interactions is far too large a task to do.

Zhong and Sternberg decided that it would be useful to incorporate the existing data on gene interactions — information which is spread throughout the scientific literature — into an integrated database that researchers could access easily online.

They first developed a statistical model for integrating such data and trained it on data from the scientific literature on genetic interactions in the roundworm, *Caenorhabditis elegans*, a widely used model organism for genetic studies. These data included positive hits of 1,816 pairs of genetic interactions and 2,878 physical interactions. For examples of negatives, the researchers also gleaned from WormBase, the major database of *C. elegans* genomic data, a collection of 3,295 genes that had a low probability of interaction.

The idea was to take all that information and apply the statistical model so the computer can compute what weight should be given to each type of data, said Sternberg.

Once they had refined their technique using the *C. elegans* data, Zhong and Sternberg then integrated genetic interaction data on corresponding, or homologous genes, from the fruit fly *Drosophila melanogaster* and the yeast *S. cerevisiae*. In this integration, they used criteria such as whether a pair of genes had identical anatomical expression, produced the same phenotype characteristics when perturbed, or showed similar expression levels when their activity was analyzed in DNA microarrays.

Although each of these organisms had a lot of experimental data, each is obviously not all complete, said Sternberg. So, the idea was that we could use information from different species to complement one another, to fill in missing information on interactions.

By drawing on data from the fruit fly and yeast, the researchers developed a genetic interaction network for *C. elegans* consisting of 2,254 genes and 18,183 interactions.

They then experimentally evaluated how well the database worked by testing whether the interactions it predicted could actually be confirmed by laboratory experiments. In these experiments, they used the database to develop lists of predicted interactions for two well-characterized genes in two different pathways in the worm. One gene controls a pathway that governs development of cells in a structure called the vulva. And the second gene controls the machinery by which the worm pumps its food through its pharynx.

In both cases, the researchers were able to use the database to correctly predict the genetic interactions that were revealed by experiments in which the researchers knocked out genes in the two biological pathways, said Sternberg.

In a commentary on the paper that was published in the same issue of *Science*, HHMI investigator Sean Eddy wrote that Zhong and Sternberg had taken a step toward total information awareness in genetic analysis. They have developed an integrated database system that predicts genetic interactions in the worm *Caenorhabditis elegans*, one of the best-studied models of how animals work.

Eddy added in an interview, One reason the work is significant is that they didn't just make a bunch of predictions. They did the necessary experiments to verify some of them. That's still a rare combination in computational biology papers. The Sternberg lab is one of the few worm labs with the skills to do both computational and experimental work at a very high level. They're not only the home of the community WormBase database, they're also top flight worm geneticists.

Zhong and Sternberg are making the database publicly available at [Predictions\(tenaya.caltech.edu:8000/predict](http://Predictions(tenaya.caltech.edu:8000/predict). They are also updating the data, which were based on the scientific literature as of 2005.

As a next step, they plan to explore whether the database can function in the other direction — using the gene interaction information on *C. elegans* to predict interactions in fruit flies and yeast. One of their future goals is to integrate data from other organisms, such as mice and even plants, and to determine whether the new data improve the predictive capability of their database.

The overall lesson we learned — which was very pleasing — was that there are clearly good databases out there; and we in the research community really know more than we thought we did, said Sternberg. But even though there is a lot of knowledge about genetic interactions, people can't read all the papers on them. So, we hope this approach will enable researchers to do a simple search to see the predicted interactions for their gene of interest. Then they could save time in choosing which interactions to test experimentally, getting more bang for their buck. And more generally, the database will help in retrieving the genetic interaction data that's publicly available out there.