

## SMALL RNAs

## Partners in time, partners in crime

**Phylogenetic profiling implicates diverse genes in small-RNA pathways.**

In his search for genetic culprits, Gary Ruvkun of Harvard University has resorted to criminal profiling. Faced with long lists of suspects involved in RNA interference (RNAi), he needed a way to prioritize candidates from noisy functional screens in the roundworm *Caenorhabditis elegans*. RNAi is used by many organisms to regulate gene expression or silence foreign RNA. While on the trail of the Argonaute proteins that bind RNA to trigger silencing, Ruvkun stumbled across a database showing their distribution in seven disparate species, and he realized that many hits from his screens shared the same evolutionary profile.

Phylogenetic profiling has been around for at least 15 years, since Eugene Koonin and his group at the US National Center for Biotechnology Information outlined the concept of clusters of orthologous groups (COGs). In Koonin's original approach, proteins that cluster by their shared presence or absence across evolutionary lineages are thought to do similar things because, for example, they act in the same protein complexes or pathways. If at least one member of a cluster has a known function, the same function can be inferred for the other members.

The approach has had success in bacteria, the first organisms to have their genomes sequenced. The clear loss and retention pattern of Argonautes actually led Koonin and his group, who developed the COG database that Ruvkun stumbled on, to predict their role in RNAi early on. In eukaryotes, profiling has been largely restricted to proteins related to the loss or retention of entire structures such as mitochondria.

Generalizing the approach fell to lead postdoc Yuval Tabach. Using his biostatistics savvy, Tabach found a way to represent protein sequence divergence and partial loss rather than only total loss or retention. Candidates are then ranked by how similar their profile is to that of a known worm RNAi protein. Tabach also included normalizations to make more meaningful evolutionary inferences. For instance, conservation between two proteins is normalized by the evolutionary distance between the two species.

Ruvkun's team generated profiles for worm RNAi components across 86 species, finding many new candidates. Tabach also developed a statistical method to combine gene expression and protein interaction data with the phylogenetic profile to predict function. The team experimentally evaluated a large fraction of the top candidates, including splicing factors that support an RNAi mechanism that monitors intron presence to distinguish invading viral genomes from host genes.

Ruvkun's excitement about the approach is captured in his favorite candidate: phosphoglycerate mutase, a sugar metabolism protein. "As a developmental biologist, if you were to ask me 20 years ago, 'Do you want to work on phosphoglycerate mutase?' I'd shout out, 'NO!'" says Ruvkun. "But here we are; its phylogenetic profile is gorgeous."

Worm phosphoglycerate mutase has the same profile as RNA-dependent RNA polymerase, the amplification engine of RNAi. "There's a 100% correlation—if you have the worm kind, you do RNAi like a race car, and if you're the human kind, you do RNAi like a jalopy," says Ruvkun, raising the provocative question of why glycolysis is so different between humans, who do not have these polymerases, and worms.

Unlike sequence similarity, phylogenetic profiling can be used to pin the same function on seemingly unrelated proteins, and Ruvkun believes that it should be used as a gene annotation tool. Only genes conserved outside their immediate evolutionary neighborhood are eligible, though, excluding roughly half of the worm genome.

The current work is limited to a gene- and organism-centric view of evolved functions. But the approach can be applied to other processes and may eventually be broadened to an all-against-all comparison. Phylogenetic profiling has found new culprits in a well-studied area, positioning it as a good way to bring diverse proteins and the people who study them together to the same crime scene.

**Tal Nawy****RESEARCH PAPERS**

Tabach, Y. *et al.* Identification of small RNA pathway genes using patterns of phylogenetic conservation and divergence. *Nature* advance online publication (23 December 2012).