Researchers dig through the gene bank to uncover the roots of the evolutionary tree

July 26 2012

Ever since Darwin first published The Origin of the Species, scientists have been striving to identify a last universal common ancestor of all living species. Paleontological, biochemical, and genomic studies have produced conflicting versions of the evolutionary tree. Now a team of researchers, led by a professor at the State University of New York at Buffalo and including area high school students, has developed a novel method to search the vast archives of known gene sequences to identify and compare similar proteins across the many kingdoms of life. Using the comparisons to quantify the evolutionary closeness of different species, the researchers have identified Actinobacteria, a group of single membrane bacteria that include common soil and water life forms, as the base of the evolutionary tree.

They will present their findings at the annual meeting of the American Crystallographic Association (ACA), held July 28 – Aug. 1 in Boston, Mass.

"Today the gene banks are enormous. They contain more than 600,000 genes from the genomes of more than 6,000 species," says William Duax, a physical chemist and lead researcher on the team. However, many of the gene sequences, and the proteins they encode, are not systematically identified. Proteins that are structurally similar and perform the same function could be labeled with different numbers that obscure the fact that they belong to the same protein family. "Our first challenge is to make sure that we are comparing apples to apples and oranges to oranges," says Duax.

Duax and his team have developed efficient ways to search through the gene banks looking for all copies of the same family of protein. They concentrated their efforts on proteins that are found on the surface of cell components called ribosomes. The ribosomal proteins are among the most accurately identified proteins, and because they are not transferred between individuals independent of reproduction, are good candidates for tracing the evolution of all species.

Ribosomal proteins in the same family twist into the same shape. The sequence of amino acids in a protein determines what 3D structure it folds into and Duax and his colleagues identified patterns that marked specific types of turns. They used these marker sequences to identify and almost perfectly align the proteins, similar to the way you could use five points to identify the shape of a star and align its orientation to match other star shapes.

Structurally aligning the proteins allowed the researchers to easily spot small differences that indicate organisms belong on different branches of the evolutionary tree. For example, a single amino acid difference in one ribosomal protein separates bacteria with one cell membrane from those with two.

At the ACA meeting, the researchers will present the results from the analysis of two different ribosomal protein families, called S19 and S13. Duax will present the analysis of protein S19, while high school student Alexander Merriman will present analysis of protein S13. Merriman joined Duax's lab through a scientific mentorship program designed to give teenagers hands-on experience with cutting-edge research. "They are enthusiastic researchers and do great work," Duax says of the students he welcomes into his lab each Friday.

Both analyses point to Actinobacteria as the last universal common ancestor. This agrees with previous work done by the group on proteins

named S9 and S12.

The researchers will continue to search for more evidence to add to their developing picture of the evolutionary tree. The group plans to analyze additional proteins, as well as DNA and RNA. "We are applying a systematic approach to make sense of a sometimes messy gene bank," says Duax.

Provided by American Institute of Physics