Scientists to Annotate Bacterial Genome at UA

By Susan McGinley

The entire genome of the important soil bacterium *Azotobacter vinelandii* was recently sequenced in a joint project between the University of Arizona and the Joint Genome Institute of the U.S. Department of Energy (DOE) in Walnut Creek, California.

The U.S. DOE and the UA's Arizona Research Laboratory sponsored the project as part of the DOE's Microbial Genome Program, a spin-off of the Human Genome Project. The goal is to sequence the genomes of microbes that have properties important in current scientific investigations that are relevant to energy and make them available in public databases.

In summer 2002, scientists from nine institutions in five countries convened at the UA to annotate the genome using open source software developed by the biotechnology computing facility.

This was the first whole genome annotation to be held at the UA. Microbiologist Christina Kennedy, from the UA Department of Plant Pathology, and Nirav Merchant, head of the UA Biotechnology Computing Facility, are the principal investigators on the project.

Academic and industrial scientists are using this data to make comparisons they were not able to do before among genes and with the genomes of other organisms. The results could help in solving problems pertaining to agriculture, medicine, industrial processes, and energy production and use.

Paul Rudnick, former UA graduate student, instructs microbiology undergraduate Rachel Hellman in use of annotation software.

UA Department of Plant Pathology

"By having the sequence of every gene in *A. vinelandii*, we know the sequence of every protein," Kennedy said. "We will then recognize the appearance of new genes we hadn't known before, with new functions related to nitrogen, carbon and energy metabolism."

Found in soils worldwide, *A. vinelandii* has been studied for its unique characteristics by scientists for more than 90 years. It can utilize nitrogen gas from the atmosphere and convert it to a form that serves as a nutrient for plant growth and development. Unlike other nitrogen-fixing bacteria, it can perform this process in the presence of oxygen. This attribute makes it more flexible in adapting to a wider range of environmental conditions.

In addition, *A. vinelandii* has the unusual ability to carry out nitrogen fixation by using any one of three different enzymes with different metal content—molybdenum, vanadium, or iron—all of which are encoded by different genes, according to Kennedy. It has a very high respiratory rate, and can grow on a wide variety of organic acids, alcohols, and carbohydrates, making it useful for applications in a wide variety of scientific studies.

"We still do not understand all of the genes in this organism. We already know the sequence of 250 because they were individually discovered and analyzed over the last 20 years," Kennedy says. "The total number of genes in *Azotobacter vinelandii* is 5,000, so there are 4,800 in this recently sequenced genome that we have not seen before." (There are about 10 times that number of genes in the human genome.)

Helping to steer the week-long annotation process was bioinformatics expert Paul Rudnick, who earned his undergraduate and doctoral degrees at the UA and is now working for SAIC, Inc. at the National Cancer Institute in Frederick, Md. Also on hand was Frank Larimer, head of annotation at the Oak Ridge National Laboratory in Tennessee. Scientists from the United States, England, Germany, Norway, Canada and Mexico were assisted by computer scientists, by DNA sequencers from the DOE, and also by five undergraduate students from the UA who have studied microbial genetics and have learned the process of gene annotation.

During the workshop, the team began to refine the computer annotation process by examining each gene in the database individually to correct errors. Back at their home institutions, they are continuing to annotate the remaining genes and are communicating their findings for inclusion in the public database via the Internet.