

'Tree of Life' explored with ITR awards

By Steve Carr

UNM will collaborate with a number of institutions on two separate Information Technology Research (ITR) "large" (more than \$5 million) awards announced by the National Science Foundation recently. The grants total more than \$24 million and are two of only eight awarded from an initial field of 70 proposals.

This is the second consecutive year UNM is the lead institution on a large ITR grant. Last year it was the SEEK project led by Biology Professor William Michener. UNM joins Carnegie Mellon University, MIT, Cal-Berkeley, Cal-San Diego and the University of Florida as one of the six institutions ever to be the lead institution on more than one large ITR grant in the four-year history of the program.

UNM leads on the \$11.6 million, 13-institution effort to develop computational tools to explore evolutionary relationships among all species of living organisms forming the "Tree of Life." Spearheaded by Project Director Bernard Moret, professor of computer science in the School of Engineering (SOE), the main collaborative institutions also include Florida State University, UC Berkeley, UC San Diego and the University of Texas-Austin.

"This is an ambitious project to assemble an evolutionary Tree of Life that includes all known plants and animals," said Terry Yates, UNM Vice Provost for Research. "It will provide a predictive and comparative framework for all fundamental and applied biology. This will basically provide the infrastructure to allow us to pursue a variety of projects that benefit society such as new drug discoveries,

See TREE on PAGE 2

...Tree of Life project funded by ITR awards

From PAGE 1

identify merging diseases and predict outbreaks, to discover new life forms, to improve global agriculture and many other things we couldn't do previously because we didn't know how these organisms were related. Developing a comprehensive understanding of life's history will advance all biology and provide enormous benefits to society.

"Assembly of a comprehensive Tree of Life is like putting a man on the moon in terms of the scope of the project...there's a lot of computational challenges to handle in the assembly of roughly 1.7 million organisms."

Constructing the Tree of Life poses one of the most complex biological problems and represents challenges much greater than sequencing the human genome.

Almost two million species of organisms have been discovered and described, yet it is estimated that tens of millions remain to be discovered. Some 60 to 70 thousand species have been studied in some detail, but the resulting data are far from complete, so relatively little is known about phylogenetic relationships of Earth's species or among the major branches of the Tree.

"Reconstructing the Tree of Life is extremely important — we will get a better picture of how life has evolved on earth, a better understanding of where we come from as humans, and a sense of where life may be headed, on a very long time scale," said Moret. "Among the many consequences of obtaining an accurate reconstruction of the Tree, our understanding of the relationships between the genetic code and cell functions will expand enormously, thus accelerating the pace of biomedical discoveries."

The relationships in the Tree of Life can be determined by comparing DNA sequences; the encoded blueprint determining the characteristics of each organism. The relative similarities between DNA sequences among different

incubator to promote the development of new ideas for this enormously challenging computational task and to create a forum where experimentalists, computational biologists and computer scientists share data, compare methods and analyze

"Assembly of a comprehensive Tree of Life is like putting a man on the moon in terms of the scope of the project...there's a lot of computational challenges to handle in the assembly of roughly 1.7 million organisms."

Terry Yates, UNM Vice Provost for Research

organisms allow scientists to predict the relationships of these organisms to their common ancestors.

The end result is a map that describes species by their relationships to their close common ancestors and to their more distant relations, much like a family tree. The map will depict the evolutionary relationships of Earth's taxonomic diversity — including living and extinct forms — over the past 3.5 billion years of its existence. Developing this map has long been a high priority for biologists, but doing so requires an extraordinary computational effort.

"The computational problem is extremely difficult," said David A. Bader, co-investigator on the project and UNM professor of computer science. "Even with entirely novel solutions methods, an enormous amount of computational power will be required to construct the first version of such a tree."

The focus of the initiative is to establish a national resource to move the research community closer to realization of the Tree of Life. This resource will serve as an

results, thereby speeding up tool development while also sustaining current biological research projects.

"In order to assemble a Tree of Life we are going to need two different things. One is a lot of data on existing species," said Moret.

"We don't have nearly enough yet. Then, we're going to need computational methods and computational power to take the data and make sense out of it."

"Thus the goal of our ITR project is to provide the computational infrastructure — including algorithms, software, machines and databases — to support the analysis once more data have been collected," added Moret.

"We will do analyses all along the way, of course, but a full-scale attempt at reconstructing the Tree of Life will not take place for many years yet: just coming up with methods and platforms to operate at that scale will take us at least five years, not to mention that collecting enough data to support the reconstruction will require the efforts of teams of biologists all over the world for many years."

Forrest funded for ITR data mining project

An additional large ITR award funded this year benefits the work of UNM Computer Science Professor Stephanie Forrest. She is co-principal investigator on a \$12.5 million ITR award titled, "Sensitive Information in a Wired World," led by Stanford University Professor Dan Boneh.

The goal is to develop methods for data mining that respect and protect individual rights, but allow law enforcement and legitimate users to mine massive data sets. The research team will also develop database tools that enforce privacy policies while managing sensitive data and release tools for end-users to prevent identity theft via spoofed or malicious websites.

"The research we'll be conducting at UNM focuses on two aspects," Forrest said. "In the past we have studied biologically inspired methods for computer security. My role in the project is to think about biologically inspired methods for protecting data. In particular, one of the projects we'll be working on involves privacy enhancing databases. The idea is to protect the privacy of personal information stored in databases, while still allowing legitimate activities such as epidemiological studies or searches for potential terrorists."