**Biological Sciences Infrastructure Proposal for LSU IT Initiative**

A number of proposals are being submitted to the IT initiative for biological computing. Many of these proposals are spurred by the massive changes that are occurring in modern biology due to the influence of the Human Genome Project and the technical advances that have coincided with it. Primarily, researchers are being inundated by a flood of data that many are ill-equipped to analyze in a reasonable time frame. This is happening across biological disciplines, affecting genetics, gene expression, protein structure and function, evolution, and ecology. In response to this, it is necessary that biology departments develop expertise and capabilities in biological computing and its associated disciplines, bioinformatics, mathematical biology, database management, theoretical biology, and data mining. In addition, biology must absorb techniques from surrounding disciplines. These disciplines include not only mathematics, statistics and computer science, but also chemistry and physics, which have previously developed analytical methods that in some cases are suitable to biological application. Much as physicists are trained, biological scientists must also be trained in computational biology, mathematics and theoretical approaches to their discipline. It is critical that a new generation of biological scientists be comprehensively trained so that they are knowledgeable in biological fundamentals and have the computational skills to ask biologically motivated questions and forward biological progress with massive data sets.

The goal of attracting and retaining high quality faculty in computational biology would be enhanced by the development of a biological computation center in the Department of Biological Sciences. This center would act as a resource and focal point for biological computation at LSU. The core of this computation center will be a 200-node Beowulf cluster upon which will be installed a large variety of existing computer applications for bioinformatics, protein structure, molecular evolution, evolutionary genomics, and functional genomics analysis. This cluster would serve as a development platform devoted to biological computation, at which current and incoming faculty could expect to apply a reasonable amount of computational power to their respective analytical problems. The biological computation facility would also serve as a developmental platform for new programs and analytical approaches and for parallelization of algorithms and applications developed by faculty members. To make this facility fully functional, we propose that the IT Initiative also fund full-time computer technician(s) and graduate student fellowships for computer scientists to maintain the facility and work with biology faculty to run and parallelize their applications. To further enhance the utility of this computational center to potential faculty hires, we suggest that two large servers be purchased for maintaining databases accessible to the general scientific public. The addition of this computational infrastructure to the Department of Biological Sciences will have immediate positive impact for both pre-existing and new faculty at LSU. It is appropriate to note that this proposal was prepared by the Biological Sciences IT Committee: Mark Batzer (Chair), Richard Bruch, John Caprio, Mark Hafner, Steve Hand, Vince LiCata, Tom Moore and David Pollock.

**Relationship to Current Biological Computation at LSU**

To our knowledge, Drs. Pollock and Batzer are the only Biological Sciences faculty which have established small Beowulf clusters. In addition, a moderate-sized cluster has been established in the Department of Chemistry, and there are larger clusters in the Department of Physics and Astronomy that have the potential to be used for biological computing. We envision that the proposed cluster would form part of a larger network of clusters at LSU, including any larger university-wide clusters that may be established under the IT initiative. However, it is also important to have a dedicated cluster in Biological Sciences that will be reserved for biological computing. In order to recruit and retain the best faculty possible, it is necessary to have a set of nodes that Biological Sciences faculty know are dedicated for biological research. The cluster will be available and readily accessible to professors and their students involved in massive computation, so that they can perform the experiments necessary to develop new programs and parallelize established analytical programs. It will also be extremely useful to have the nodes and technician(s) on hand and dedicated to making commonly used computer-intensive biological programs readily available to researchers. This will also avoid the repetitive work of installing and establishing programs within individual research laboratories in a number of quantitative-intensive research areas such as comparative genomics, proteomics and cell signaling. The technician(s) will also establish and maintain libraries that are useful in bioinformatics, such as the Sanger Centre's BioPerl, NCBI's SEAL, NCSU's HiFi, BioPython, and newly developing XML standards and libraries.

Drs. Pollock and Batzer would oversee the development and maintenance of the cluster including the establishment of an oversight committee composed of faculty throughout the Department of Biological Sciences and the university who are interested in biological computing. We will also consult faculty that have expertise in relevant areas, such as Dr. Newcomer for structural visualization and Drs. Battisa, Hand and Noor for microarray and proteomic analysis. Both Dr. Battista and Dr. Pollock have colleagues at The Institute for Genomic Research (TIGR) who can be consulted on the functional genomic aspects of this cluster, while Dr. Batzer and Dr. Pollock also have strong connections with the National Laboratories at Livermore and Los Alamos that can act as an additional resource. The departmental computing cluster would be developed at the same time as individual laboratory clusters also expand and are linked together. Therefore, we anticipate a combined total of over 250 nodes to be available for biological computing.

**Summary**

The establishment of a biological sciences Beowulf cluster through the IT initiative will enhance LSU's reputation in biological computing and improve our chances of attracting top-quality faculty. The cluster will also be a key component in establishing LSU as a recognized academic center for biological computing. We request funds to establish a 200-node cluster along with two dedicated servers. The cluster will be

supported by at least one additional full-time technician and computer science graduate students with additional support provided as part of the support personnel for new faculty hires through the IT initiative.