

MidSouth
Bioinformatics Center

GEMS

Calendar of Events

September

- 4 BRIN External Advisory Council Meeting at UALR
- 5 Dr. Cynthia Gibas Colloquium Speaker
- 10 MidSouth Bioinformatics Center Board Meeting
- 16-19 Virtual Genomics and Bioinformatics Conference UALR (Access Grid)
- 19-20 Research Symposium UAF
- 29 Dr. Grier Page Seminar Speaker

October

- 6 Dr. James C. Fuscoe Seminar Speaker
- 14 Last day lodging guaranteed for the MCBIOS Conference
- 17 Dr. Robert L. Burnap Colloquium Speaker
- 24 Dr. Hubert Preißl Colloquium Speaker
- 31 MBC Open House

November

- 14 Dr. David Mount Colloquium Speaker
- 14-16 MCBIOS Conference UALR

December

- 5 Dr. Jaiwei Han Colloquium Speaker

Welcome



Welcome to the first issue of "MBC Gems"! We've chosen the name Gems due to its association with "mining" which in turn is a play on "data mining," an important technique in the field of bioinformatics. Maybe by the time we produce the next installment of the newsletter we'll have a nice logo to go along with the name.

The MidSouth Bioinformatics Center (MBC) has come a long way since it was just a dream of mine when I joined UALR in May of 2002. The financial support of the Arkansas BRIN (Biomedical Research Infrastructure Network) program funded by the NIH NCRR (National Institutes of Health's National Center for Research Resources) not only provided funds to hire me, but made possible the establishment of the Center. I want to especially thank

Drs. Larry Cornett and Helen Beneš (UAMS faculty who are the Director and Associate Director, respectively, of the Arkansas BRIN) and Drs. Mary Good and Roger Hawk (my Dean and Chair, respectively, at UALR) for their support and encouragement.

The MBC is intended to be a regional resource for faculty, students, and researchers and we encourage all to participate in any of the many programs that we intend to offer and to make use of our computing and reference facilities. In this newsletter we'll be describing some of the programs and facilities that we've already established — we have many more on the drawing board.

I want to also introduce our staff to you:

Lisa Henderson joined the MBC in July as our Managing Director. In this capacity she'll be managing the administrative side and developing and running many of the initiatives that we're undertaking (such as a "matching program" to help researchers find collaborators in the life science and/or

Welcome continued page 3



In this issue

MBC Announces Board of Directors	2
MCBIOS Conference Information	2
Interesting Links	3
UALR and UAMS Collaborate to Offer New Academic Programs	4
UALR to Host 2003 Research Symposium	6
Seminar Speakers Planned for Fall	7
Bioperl Workshops Offered this Summer	7
A Report from the Genetic and Evolutionary Computation Conference (GECCO)	8
Students and Positions Wanted	10
New Computing Center Installed	10

MBC Board

by: Steve Jennings

Through my volunteer involvement with a number of non-profits earlier in my career I came to appreciate the tremendous value that a Board of Directors can make to the running of an organization. Because of this, I've recruited a wonderful group of people to advise the MBC on policy, funding, organizational, and operational issues. I expect that the Board will eventually grow to include more representatives from outside the State of Arkansas, but I wanted to introduce those who have agreed to serve in this capacity initially:

- .. John Ahlen, President, Arkansas Science and Technology Authority
- .. Susan Davis Allen, Vice Chancellor for Research and Academic Affairs, Arkansas State University
- .. David Belcher, Provost, University of Arkansas at Little Rock
- .. Don Bobbitt, Dean, Fulbright College of Arts & Sciences, University of Arkansas at Fayetteville
- .. Dan Casiano, Director, National Center for Toxicological Research
- .. Larry Cornett, AR BRIN Director and Professor, Department of Physiology and Biophysics, University of Arkansas for Medical Sciences
- .. Mary Good, Dean, College of Information Science and Systems Engineering, University of Arkansas at Little Rock
- .. Tom Kieber-Emmons, Director of Basic Breast Cancer Research, University of Arkansas for Medical Sciences
- .. Michael Ledbetter, Dean, College of Science and Mathematics, University of Arkansas at Little Rock
- .. Sylvia Spengler, Program Director, Biological Databases and Informatics, National Science Foundation
- .. Charles Winter, Associate Dean for Research, College of Medicine, University of Arkansas for Medical Sciences

We'll plan on providing profiles of these Board members in future issues of our newsletter.

The first Board of Directors meeting will be held from 9:00 a.m. to 4 p.m. at the UALR Reynolds Building in Room 460 on Wednesday, September 10, 2003 followed by a reception at the Jennings residence.

GEMS is published to share information about the scope of activities surrounding the MidSouth Bioinformatics Center located in the Donaghey College of Information Sciences and Systems Engineering (CyberCollege) at the University of Arkansas

MCBIOS Conference Announced

The newly-formed MidSouth Computational Biology and Bioinformatics Society (MCBIOS) announces its first annual conference to be held in Little Rock, AR on November 14-16, 2003. The conference will feature two days of technical presentations, posters, and panel discussions in addition to a business meeting and, on Sunday, a free NCBI workshop on GenBank and molecular biology tools. Dr. David Mount, author of "Bioinformatics: Sequence and Genome Analysis" will be the keynote speaker. A special dinner will be held at the Museum of Discovery in the beautiful Riverwalk area on Friday evening.



Tentative Conference schedule

Fri	7-8am	Poster setup
	8-8:30am	Welcome
	8:30-9:45am	Technical session
	9:45-10am	Break
	10-11:45am	Business meeting
	11:45am-1:30pm	Lunch break
	1:30-2:45pm	Panel discussion
	2:45-3pm	Refreshments
	3-4pm	Keynote presentation
	4-6pm	Poster session
	6:30-9pm	Special event: Museum of Discovery
Sat	8-9:15am	Technical session
	9:15-10:30am	Panel discussion
	10:30-10:45am	Break
	10:45am-Noon	Technical session
	Noon-1:45pm	Lunch break
	1:45-3pm	Panel discussion
	3-3:30pm	Closing remarks
	3:30-5pm	Planning for 2 nd Annual Conference
Sun	8-9:30am	NCBI Training
	9:30-9:45am	Break
	9:45-11:15am	NCBI Training cont.
	11:15-Noon	Lunch break
	Noon-2:00pm	NCBI Training Lab

Abstracts for papers or posters are due by Oct. 1st. Registration is required with an "early bird" discount offered for registrations received by Oct. 1st. Substantial registration discounts additionally are available for students through the Arkansas BRIN program funded by NIH's National Center for Research Resources. To register or obtain additional information, please see <http://www.MCBIOS.org> or call the MidSouth Bioinformatics Center at (501) 569-8501.



<i>Fee schedule</i>	<i>Regular</i>	<i>Students</i>
<i>Conference fee</i>	\$145	\$145
<i>Early registration discount (by 10/01/03)</i>	(\$25)	(\$25)
<i>AR BRIN special student discount</i>	--	(\$90)
Total	\$120	\$30

at Little Rock. For more information about the MidSouth Bioinformatics Center phone (501) 569-8501 / fax (501) 569-8020. You may also visit us in the ETAS Building, Room 507 on the UALR campus.

Welcome continued

computational fields and our first annual conference of the MidSouth Computational Biology and Bioinformatics Society [MCBIOS] which we're sponsoring). Lisa returned to Arkansas from Hawaii to take this position and has a significant amount of experience running academic centers in addition to her MBA degree. If you contact us, Lisa may be your first point of contact.

Roger Hall joined the MBC in April as our Technical Director; many of you have attended the first run of his bioPerl workshop series this summer. Roger is an expert in the areas of Internet technologies and system administration and has spent a lot of time establishing our Bioinformatics Computing Facility. Roger has had a long-time interest in bioinformatics and has contributed to the open-source bioPerl project and consulted for Genentech. Roger will begin pursuing his Masters degree in Bioinformatics next year.

Mildred Pearce has been providing me valuable administrative assistance this past year and just rejoined the MBC staff full-time. Mildred's long tenure with UALR gives her the contacts that have proven invaluable in getting things done!

Phil Williams and Vinay Ravindrakumar are Graduate Assistants just starting their second year of studies in the UALR Applied Science graduate program in applied biosciences and intend to switch to the graduate program in bioinformatics once it is approved. Both are making

significant contributions to our bioinformatics course development activities. Julia Sonka joined the MBC this summer as an undergraduate fellow while she completes her undergraduate degree in Biology and takes a few additional courses this fall in Information Science and Math and Statistics. She plans on pursuing a graduate degree in bioinformatics beginning in the spring. Julia has been helping us by maintaining our budding reference library and in preparing for the fall bioinformatics conference.



In later breaking news, we've just brought on Cory Brownfield as a new Graduate Assistant. Cory will be helping Roger on the technical side and in setting up some of the bioinformatics computer labs for our new courses.

I am truly grateful for the assistance that all these people are providing in helping us to build up the bioinformatics infrastructure within the state and region. This is very much an entrepreneurial effort! I LOVE doing this kind of thing, but it takes some time to start a number of programs and services totally from scratch. Please bear with us as we "get our act together." I'm sure that you'll be pleased with the end result!

Steve Jennings
MBC Executive Director

Interesting Links

1. Human genome project information:

http://www.ornl.gov/TechResources/Human_Genome/home.html

2. Genetic programming basics:

<http://www.genetic-programming.com/gpanimatedtutorial.html>

3. DNA Microarray Methodology - Flash Animation

<http://www.bio.davidson.edu/courses/genomics/chip/chip.html>

4. BioPerlTutorial - a tutorial for bioPerl:

<http://www.bioperl.org/Core/Latest/bptutorial.html>

5. The world's most powerful computer: Earth simulator:

<http://www.es.jamstec.go.jp/esc/eng/>

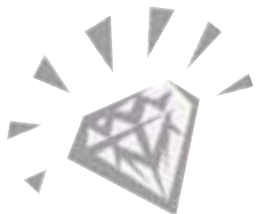
Japan already has the world's most powerful computer in the [Earth Simulator](#), which can theoretically perform 35 teraflops. Japan wants to create an even faster computer, a "[virtual supercomputer](#)" using grid

[technology](#). The aim of the project is to link 100 supercomputers in Japan over a high-speed network. The supercomputers will be linked over the [SINET network](#) (Science Information NETWORK), which operates at 10GB/second and connects Japanese research institutions both with one another and with the U.S. and Great Britain. The goal is to create a 100 teraflops virtual supercomputer that could be used for [biochemistry research](#), among other projects. Japan has recently formed a [grid consortium](#) to promote grid technology. Both the [National Institute of Informatics](#) and the [Institute for Molecular Science](#) will be participating in this endeavor.

6. DNA -O - Gram Generator:

<http://dna2z.com/DNA-o-gram/index.html>

Send a message in the language of life... Say it with DNA. Say it with DNA rather than flowers. The site allows you to generate a message in the language of life not love and reveal the hidden meaning of genetic missives you receive. DNA is not just about hidden codes. It can be used in bioremediation, alien communication, movie plots, forensics, jewelry, and apparently is critical to quagga repopulation.



New Bioinformatics Academic Programs Proposed

- Biology Bioinformatics Minor
- Information Science Bioinformatics Minor
- Professional Certificate in Bioinformatics
- Masters Degree in Bioinformatics
- Ph.D. Degree in Bioinformatics
- Graduate Certificate



Vinay Ravindrakumar, Graduate Assistant, works on bioinformatics new course development.

UALR and UAMS Unite to Offer

While it is premature to officially announce any new academic programs or courses in bioinformatics, many of you are aware of the plans that are being made to develop undergraduate minors in bioinformatics at UALR and a joint graduate program in bioinformatics between UALR and UAMS. This article will summarize those plans and their current status.

Undergraduate Minors

We are preparing proposals to start two new minors in bioinformatics at UALR which will be offered by the Biology and Information Science departments — if approved by these departments and the college and university curriculum committees. Both minors consist of eighteen credit hours and include two new bioinformatics courses. These bioinformatics courses (*BINF 2345:*

Introduction to Bioinformatics and *BINF 4445: Bioinformatics Theory and Applications*) are currently under development and will constitute seven of the eighteen hours; current plans are to offer the introductory course in Spring semesters and the advanced course in Fall semesters, both beginning in 2004. The Biology bioinformatics minor, intended primarily for majors in the Information Science and Computer Science departments, will also include *CHEM 2450: Organic Chemistry Short Course* and *BIOL 24xx: Introduction to Molecular Biology* (both sophomore-level courses; the latter is a new course proposed by Dr. John Bush who also teaches the senior-/graduate-level *Molecular Biology* course) and *BIOL 3300: Genetics*, a junior-level course. The Information Science bioinformatics minor, intended primarily for biology majors, will also include *IFSC 1202: Enterprise Information Science II* (programming with Visual Basic), *IFSC*

1305: Problem-Solving Techniques, *IFSC 2300: Object-Oriented Technology* (programming with Java), and *IFSC 3320: Database Concepts*.

We plan on submitting the proposals, and hopefully winning approval, for these two minors and the three new undergraduate courses this fall.

Professional Certificate

Additionally, we intend on offering a “Professional Certificate in Bioinformatics” to those post-baccalaureate students who return and complete all of the aforementioned courses or their equivalent (with the exception of those courses that might have been part of their undergraduate degree); *BINF 4445* will need to be taken for graduate credit (*BINF 5445*) however. See below.

Graduate Programs

A large committee with representatives from UALR, UAMS, and NCTR and organized under the auspices of the AR BRIN program has been meeting for over a year now preparing a joint graduate program in bioinformatics. Our plans are to submit this proposal to the Arkansas Department of Higher Education (ADHE) during the Fall 2003 semester. This proposal, however, first has to be approved by the Graduate Councils of both UALR and UAMS and by the University of Arkansas Board of Trustees before making its way to the ADHE. If all goes well, we hope to have approval early enough in the Spring 2004 semester to undertake a full-scale recruiting effort for the 2004-2005 academic year. In the interim, there are a number of students taking courses that could apply to the proposed graduate degrees. The graduate program includes Masters, Doctorate, and Graduate Certificate degrees.

Masters Degree

The Masters Degree proposal is built around four core areas: bioinformatics, biostatistics, informa-

New Academic Programs

tion and computer sciences, and the life sciences. Students will be required to take a minimum of two graduate courses in each of the four areas with a minimum of nine courses for 34 credits overall. Additionally, students will take a one-credit-hour bioinformatics seminar each semester (for four credits) and participate in a lab rotation in each of the four core areas (1/2 credit hour each spread over their first two semesters for a total of two credits). While totalling forty credit hours, we anticipate that most undergraduate students will be able to waive at least one 4-hour course if their undergraduate degree was in one of the core areas (students entering with a graduate degree in a core area may be able to waive even more courses in a core). Currently only two graduate-level bioinformatics courses are under development, making them effectively required courses under this plan; the other cores have many more options. BINF 5445, to be offered in Fall semesters, is the same course as BINF 4445 course except that graduate students will be graded with a higher standard and have more responsibility on the class projects which they will supervise. *BINF 7645: Bioinformatics Project* (offered in Spring semesters) will consist of two parts: a survey of relevant, non-technical topics (e.g., project management, bioethics, grantsmanship, technical writing, and communication skills) and a significant, semester-long project under the guidance of a project mentor. This project is designed as a capstone project which will bring together the learning gained in the coursework of all four cores.

Ph.D. Degree

While the Masters Degree is breadth-focused, the Doctorate Degree is depth-focused and is simply explained as a minimum of 36 credit hours of research (and possibly advanced coursework) post-Masters; four of these credits are for attending

additional semesters of the bioinformatics seminar. Since completing a Masters Degree in Bioinformatics is required for admission to the Doctorate program, there are no “qualifying exams” typically found in such programs — the capstone project effectively takes its place. The “candidacy exam” is another matter: students will be required to prepare a complete grant proposal (NIH or NSF format) during their first year post-Masters, submit it in electronic format, and defend it orally to fulfill this requirement. While not required, we hope that many will submit it for actual funding consideration.

The focus on research is hopefully evident in this proposal. Students will gain their first exposure to research during their year-one lab rotations which will be followed by their capstone project in their second year. If a topic is chosen well at the Masters level, that will grow into a research topic during their third year (first year in the Doctorate program for those students who continue) which will be enhanced during the grant preparation process and expanded in their final year(s) as their dissertation and defended at their “orals.”

Graduate Certificate

The Graduate Certificate in Bioinformatics (GCB) proposal is approximately “half a Masters Degree” with the aforementioned Professional Certificate in Bioinformatics (PCB) as a prerequisite. The GCB consists of one course in each of the four cores (note that students will already have taken BINF 5445 as part of the PCB, so the only remaining option under the bioinformatics core is BINF 7645). Students will be advised to choose carefully courses that could become part of a well-planned Master Degree program should they decide to pursue that option at a later time (students will also be advised to participate in the

Survey Results

The Arkansas Department of Higher Education requires a “needs assessment” section in new program proposals. To complete this section in the Joint Graduate Program in Bioinformatics proposal we undertook a web-based survey in which 51 people responded. 98% of those who described themselves as potential students indicated that a graduate program in bioinformatics was important and 80% valued the availability of bioinformatics certificates. For those of us who will be teaching classes as part of this proposed program, it is relevant to note that 94% were interested in taking classes on a part-time basis and 85% stated that it was important to offer classes outside of normal work hours.


24% of respondents were self-identified and affiliated with verifiable institutions. Their collective estimates were that as many as 65 bioinformatics graduates could be hired at their institutions over the next five years. 65% of these positions are expected to prefer that the person hired will have a graduate degree in bioinformatics, while another 21% were anticipated to be “entry level” bioinformatics positions.

Many thanks to Pat Pearce of UALR Computing Services and Roger Hall for their assistance in collecting and analyzing these results.

lab rotations and bioinformatics seminars as well).

We want to emphasize strongly that we are in the proposal stages only and that we have no sanction to offer the programs which have just been described. Hopefully this will change within the next 4-6 months!

UAF to Host 2003 Research Symposium

 Faculty and students at all levels are invited to participate in the 2003 Research Symposium – Networking Arkansas Institutions of Higher Learning to be held September 19-20, 2003 at the University of Arkansas, Fayetteville. This symposium will survey the latest research and instructional techniques in chemistry and biochemistry.



Featured speaker – The featured speaker is Bassam Z. Shakhashiri, William T. Evjue Distinguished Chair for the Wisconsin Idea and Professor of Chemistry, University of Wisconsin, Madison.

Poster presentations – All poster presentations are welcome. No abstract required—only the title, the authors and home institution needed.

Social event – A banquet is planned for Friday, September 19 at the UARK Bowl on Dickson Street. There is no

charge for the banquet but prior registration is required.

There are no registration fees and substantial travel and lodging subsidies are available. The subsidies should cover most of the travel and lodging expenses for faculty and students.

Register on-line http://www.uark.edu/depts/cheminfo/uarkchem/eventsinfo/research_symposium/index.htm

This symposium is being hosted by the Department of Chemistry and Biochemistry and sponsored by the University of Arkansas, UALR and UAMS with support from NIH and NSF.


For questions or comments email cheminfo@uark.edu.

If you prefer to speak to someone, contact Monica Mabie at (479) 575-4648.

Featured workshops

- **Determination of the Structures of Molecules via Single Crystal X-ray Diffraction**, Wally Cordes, University Professor Emeritus
- **From Protons to Proteins: A Multinuclear - Multidimensional NMR Experience**, James Hinton, University Professor
- **The Conceptual Consequences of Quantum Mechanics and their Application in the Science Curriculum**, Lothar Schäfer, Distinguished Professor
- **Protein Identification through Mass Fingerprinting**, Jack Lay, Director of the Statewide Mass Spectrometry Facility
- **Exploring the Nanoworld, Instructor** - Mike Condren, Professor of Chemistry, Christian Brothers University
- **Teaching Strategies for Science Courses**, Bob Gawley, Professor
- **Electronic Homework and Testing in the Chemical and Biological Sciences**, Professor Norb Pienta of the University of Iowa, Professor David Anderson of the University of Colorado at Colorado Springs and Professor Todor Gounev of the University of Missouri at Kansas City

MBC Home and Open House

 The MidSouth Bioinformatics Center (MBC) has established its home on the fifth floor of the ETAS building at the University of Arkansas at Little Rock. The space includes offices for the staff and graduate students, a guest computer workstation, a growing bioinformatics library, and a convenient study lounge.

An Open House will be held from 1:00 - 2:30 p.m. on Friday, October 31. Plan to join us for refreshments, a tour of the facilities, and demonstrations of the newly acquired computer equipment.

Workshop Series Offered

The MBC offered a perl and Bioperl workshop series this summer, giving attendees the opportunity to learn the programming skills required to create bioinformatics software and deliver it via the WWW. MBC Technical Director Roger Hall has been working with web technologies for eight years, and chose the topic based on his recent experiences working for Genentech and contributing to the Bioperl project.

Researchers and students from all organizations were invited to attend the workshop. Seventeen students attended in total, including researchers from NCTR, UAMS, and UALR and students from UAMS and UALR. The sessions were between four and six hours long, not including a meal break. Each session's meal was sponsored by MBC, which enabled some inter-agency mingling.

"I was really amazed at the range of background in the students," Mr. Hall said. "There were students from every science and computer background possible, and several students were just beginning to learn programming. It kept me on my toes. I really appreciate students that asked questions - especially when they quickly helped each other in both asking and explaining."

Workshop continued page 10

Fall Biosciences & Bioinformatics Seminars Scheduled

Make plans to attend the following seminars. Note: For location and times please visit <http://biosciences.ualr.edu/seminar> or contact Dr. Gary Thompson (501-371-7506)

September 5, 2003

Dr. Cynthia Gibas, Assistant Professor, Virginia Polytechnic Institute and State University and Fralin Biotechnology Center is giving a talk on "New Tools for Comparative Functional Genomics."

Dr. Gibas is the famous author of the book – "Developing Bioinformatics Computer Skills." Her research interests include:

- Whole genome comparison and comparative databases.
- DNA array design for comparative functional genomics and host-pathogen interaction experiments.
- Modeling and optimization of probe-target hybridization behavior on DNA arrays.
- Macromolecular structure, electrostatics and diffusion-controlled behavior.

For more information about Dr. Gibas please visit: <http://gibas.biotech.vt.edu/>

October 6, 2003

Dr. James C. Fuscoe, Director – Center for Functional Genomics, NCTR is giving a talk on "DNA Microarrays for Toxicogenomics."

Dr. Fuscoe is affiliated to the Division of Genetic and Reproductive Toxicology, which addresses the FDA's requirement that the potency of suspected mutagens and carcinogens be identified and measured. The Genetic Toxicology Laboratory conducts fundamental research aimed at defining the pathways from initial DNA damage to mutation.

October 17, 2003

Dr. Robert L. Burnap, Professor, Dept of Microbiology and Molecular Genetics, Oklahoma State University, is giving a talk on "Deciphering Photosynthetic Regulatory Networks Using DNA Microarrays in Conjunction with Perturbation Analysis."

Dr. Burnap's research interests include:

·Molecular genetics and biophysics of photosynthetic H₂O-oxidation.

·Biogenesis and adaptive mechanisms of the photosynthetic apparatus.

For more information about Dr. Burnap please visit:

http://okepscor.org/RIIProposal/genomics/Burnap_cv.htm

December 5, 2003

Dr. Jiawei Han, Professor, Dept of Computer Science, University of Illinois at Urbana-Champaign is giving a talk on "Mining Dynamics of Data Streams in Multidimensional Space."

Dr. Han is the famous author of the book "Data Mining Concepts and Techniques." His research interests include:

Database systems, knowledge-base systems, data mining (knowledge discovery in databases), data warehousing, spatial databases and spatial data mining, multimedia databases and multimedia data mining, Web database and Web mining, bioinformatics and DNA data mining, database applications, deductive and object-oriented databases, logic programming, and artificial intelligence.

For more information on Dr. Han please visit: <http://www.cs.sfu.ca/~han/>

This summer Roger Hall, MBC Technical Director, taught a workshop series on Bioperl.



Genetic and Evolutionary Computation Conference (GECCO)

by: Philip Williams

I recently attended the "Genetic and Evolutionary Computation Conference" (GECCO) held in Chicago July 12-16. There were several presentations that discussed the use of genetic algorithms specifically applied to bioinformatics problems.

Genetic algorithms (GA) and genetic programming (GP) have emerged as a method that can be applied to solving problems in various fields. Some of these include, but are not limited to, scheduling, mechanical design, evolutionary robotics, spam email control, evolvable hardware, grammar development, and optimization problems. Specifically in the bioinformatics field, work has been done on analysis of microarray data, phylogenetic tree construction, and multiple sequence alignment.

A brief description of how a genetic algorithm works is as follows:

First a random population of solutions is created and evaluated for fitness. Two individuals of high-fitness are either recombined to form two new individuals for fitness testing or mutation is performed on

these was the graduate student workshop. This workshop gave the students the opportunity to present their work and have it critiqued by major scientists in the field. One critique pointed out the need to explain concisely the relevance of the presented research. In another they ask the student why they had used a GA when traditional methods had already been proven to work well.

An excellent tutorial was given by Steffan Christensen and Mark Wineberg entitled "Using Appropriate Statistics." They discussed calculating nonparametric statistics using a spreadsheet by sorting the data and applying ranks. They also discussed the need to apply a Bonferoni correction for some types of data generated by GA and GP techniques. In critiquing work by others, these authors point out that using the median rather than the average may be more appropriate in determining success for some GA/GP methods being investigated.

All three presentations in the first Bioinformatics section involved the

analysis of microarray data. One approach used an artificial immune system to

classify cancer cells. Another used a genetic programming-based strategy called symbolic discriminant analysis (SDA) for the automatic selection of gene expression variables. The third in the section presented a technique where a hybrid genetic algorithm is used to visualize the gene order in each cluster from the microarray data.

The moderator for the second section of Bioinformatics presentations was our friend Dr. James Foster from the University of Idaho who gave a colloquium talk at UALR last year.

Bioinformatics 1 Chair:
Jason Moore

Artificial Immune System for Classification of Gene Expression Data
Shin Ando, Hitoshi Iba

Complex Function Sets Improve Symbolic Discriminant Analysis of Microarray Data
David Reif, Bill White, Nancy Olsen, Thomas Aune, Jason Moore

Finding the Optimal Gene Order in Displaying Microarray Data
Seung-Kyu Lee, Yong-Hyuk Kim, Byung-Ro Moon

The first talk in the second series proposed a genetic algorithm for optimizing the order of taxon addition in fastDNAmI. One drawback of using the maximum likelihood (ml) method of phylogenetic tree construction is that it can produce different taxonomic trees based on the order in which taxon are added. Here researchers have shown improved results based on an algorithm for determining the order of taxon addition.

The second presentation was about evolving the consensus sequence for multiple sequence alignment. The researchers developed an encoding scheme where the number of generations needed to find the optimal solution was approximately the same regardless of the number of sequences. For this system the number of generations depended only on the length of the template and the similarity between the sequences. They found that their system scaled well in comparison testing with the clustalW sequence alignment program. The final speaker in the second Bioinformatics session presented work on the subject of molecular docking. They used an evolutionary approach to solve problems of flexible docking between ligands and proteins.

The full publication for these papers can be found at:

<http://gal4.ge.uiuc.edu:8080/GECCO-2003/>

an individual and it is retested. There is usually a combination of "selection" and "chance" involved in these systems. The selection is based on level of fitness. The higher the fitness, the higher the chance is of being picked for a crossover or mutation operations. In most cases the location where the crossover or mutation operation takes place is also chosen by random. As the system continues to run through successive generations, the average overall fitness increases. This is because the population is adapting to better fit the solution space.

Several workshops were held on the first day of the conference. One of



Philip Williams, Graduate Assistant, attended the GECCO conference.

space of local minima instead of the whole space of feasible conformations. In contrast to existing hybrid evolutionary computing algorithms where local minimization algorithms have been used to “fine-tune” the solutions, these researchers focus primarily on constructing local minima from previ-

ously explored minima and only use genetic operators to assist in diversification. As a consequence, their total number of generations (iterations) was demonstrated to be low (~ 50). In contrast, standard genetic algorithms and Monte Carlo methods require a large number of generations. In some cases from 150,000 to nearly 20,000,000 generations for these methods converge and produce the best solution.

The full publication for each paper can be found at:

<http://gal4.ge.uiuc.edu:8080/GECCO-2003/>

Recombination from crossover, muta-

Bioinformatics 2 Chair:
James Foster

Optimizing the Order of Taxon Addition in Phylogenetic Tree Construction Using Genetic Algorithm
Yong-Hyuk Kim, Seung-Kyu Lee, Byung-Ro Moon

Evolving Consensus Sequence for Multiple Sequence Alignment with a Genetic Algorithm
Conrad Shyu, James A. Foster

An Evolutionary Approach for Molecular Docking
Jinn-Moon Yang

phylogenetic trees. This work has focused on using lineages with known phylogenies as well as computer simulated genes to test tree construction accuracy. They have shown improved results for neighbor-joining and maximum parsimony methods. The Avida system was used for this research. Avida, in some ways similar to Tom Ray’s Tierra, is a system that uses a population of digital organisms evolving within the computer environment.

“Effective Search of the Energy Landscape for Protein Folding”, by Eugene Santos Jr., Keum Joo Kim, and Eunice E. Santos.

Other papers of interest to the bioinformatics community include:

“Evolutionary Computation Method for Promoter Site Prediction in DNA”, by Daniel Howard and Karl Benson. In this paper the goal of the authors was to find short consensus sequences that are promoters in the eukaryotic genomes. This has the potential of discovering genes that are regulated together. The method they used was a combination of Genetic Programming and finite state automata.

“The Effect of Natural Selection on Phylogeny Reconstruction Algorithms”, by Dehua Hang, Charles Ofria, Thomas M. Schmidt, and Eric Torng

Here the authors point out the need for the construction of accurate

How a Genetic Algorithm Works

First a random population of solutions is created and evaluated for fitness. Two individuals of high-fitness are either recombined to form two new individuals for fitness testing or mutation is performed on an individual and it is retested. There is usually a combination of “selection” and “chance” involved in these systems. The selection is based on level of fitness, the higher the fitness the higher the chance of being picked for a crossover or mutation operations. In most cases the location where the crossover or mutation operation takes place is also chosen by random. As the system continues to run through successive generations the average overall fitness increases. This is because the population is adapting better to fit the solution space.

The authors have proposed a new algorithmic approach for global optimization in protein folding. They are using information found in various local minima to direct the search for the global minimum. This explores the energy landscape more efficiently by considering only the

tions, and survival of the fittest are all concepts that originated in the field of biology. Increasingly these concepts are used in genetic algorithms to evolve solutions in information science. Today evolution has gone beyond theory and is now being applied successfully in the discovery of solutions to bioinformatics problems.

Work Opportunities Needed and Available for Students

There are a number of Computer and Information Science students at UALR who are looking for work assignments or graduate assistantships. This may be done through a simple referral whereby you may hire the student directly, or, UALR can set up a contractual arrangement with another institution to cover a graduate student's stipend and tuition if that is preferred. Contact Steve Jennings at 569-8216 if you have an opening.

The MidSouth Bioinformatics Center (MBC) is working with the Cooperative Education (Co-op) department to create opportunities for students to gain bioinformatics related experience prior to graduation. Interested students who qualify for the cooperative education program should contact the Center at 569-8058.

Workshop continued

The MBC is planning to sponsor the workshop again this Fall at an affiliated campus, and offer it to local students again next Spring. "We are looking at various topics for an on-campus seminar this Fall, and so far it seems that one of the biggest perceived needs locally is for more biostatistics training," Mr. Hall said.

Perl is a popular open source language available for nearly every platform. It was credited with 'saving the Human Genome Project' when project programmers used perl's

short development cycle and built-in "string manipulation" capabilities to reorganize seemingly endless fragments, eventually prevailing before commercial interests locked in invaluable intellectual property.

Bioperl is an open source software package designed and built by scientists and programmers from all over the world, including researchers from the NCBI, EMBL, and several major universities. It is used daily in production at those facilities.

New Bioinformatics Computing Center Installed



Recently, a Bioinformatics Computing Center featuring three Sun Microsystems servers purchased in conjunction with a Sun Academic Equipment grant valued at approximately \$350K was installed.

- o The "Production" server is a Sun Fire V880 with eight 900Mhz UltraSPARC-III processors, 32GB memory, and 438GB of disk space which will be used to host project databases and services, including supporting the Center's "web services" paradigm. Planned software will include Oracle databases and Spotfire microarray analysis software packages and a variety of open-software packages.

- o The "Web" server is a Sun Fire 280R with dual 1.2Ghz UltraSPARC-III processors and 8GB memory with an add-on StorEdge 3510 Fiber Channel Array with 365GB of disk space and will be used to host websites, a Project Manage-

ment Workbench under development by Roger Hall to provide service discovery and workflow automation, and access to general internet resources.

- o The "Development" server is also a Sun Fire 280R with dual 1.2Ghz UltraSPARC-III processors, 8GB memory, 146GB of disk space and an add-on StorEdge L8 8-cartridge SCSI tape drive (800Mb capacity) which will be used to host eight SunRay thin-clients.

- o We are currently identifying the software packages that we would like to support (mostly "free, open-source" such as EMBOSS) and trying to identify funding sources for the commercial packages we would like to buy. Additionally, we would like to work on negotiating multi-institutional shared software licenses and/or a key-distribution system on a regional basis.



MidSouth Bioinformatics Center

UALR/ETAS 575

2801 S. University Avenue

Little Rock, AR 72204

Tel: 501.569.8501 | Fax: 501.569.8020

