

BRC bioinformatics

NC State's bioinformatics graduate program is one of the largest in the country, according to a national survey of 64 universities.

From the Director

Welcome to the first issue of BRC *bioinformatics*, the newsletter from the Bioinformatics Research Center at North Carolina State University. As the Center enters its fourth year, we hope you'll find the newsletter to be a lively and informative account of some of the BRC's current activities.

Our fourth year has begun very well. The staff has been joined by new faculty members Steffen Heber (who you will meet in the article below) and Jung-Ying Tzeng, post-doc Amy Anderson, industry liaison Josh Baker, and editor/writer Pat Westphal.

We also welcome 23 new graduate students to the bioinformatics program, which brings the total number of students in the program to more than 50. NC State's bioinformatics program is one of the largest in the country, according to a survey of 64 U.S. graduate programs in bioinformatics and related fields (bioinformatics newsletter, August 11, 2003, www.genomeweb.com).

In this issue you will also meet Christine Duarte, one of our graduate student industry interns, who is earning accolades for her bioinformatics programming in product development at SAS.

Our on-campus and industry partnership programs continue to grow, led by Dr. Dahlia Nielsen and Dr. Josh Baker, respectively. On campus we have established new relationships with Candace Haigler in Crop Science and Ron Sederoff in Forestry. Off campus we have new partnerships with CIIT, Dix Hospital, and Quintiles.

We welcome visitors and inquiries about technical support in bioinformatics or collaborative research. Please feel free to contact us by phone or email or, better yet, by coming to coffee with us any morning between 10:00 and 10:30 a.m.

Bruce Weir
weir@stat.ncsu.edu

Computer Scientist Seeks Clues to Gene Splicing

Dr. Steffen Heber, who joined the BRC faculty this fall with a joint appointment in Computer Science, is seeking to unravel the mysteries of alternative splicing in genes – the biological process by which genes are cut up and pieced together (spliced) to form multiple proteins.

Alternative splicing is thought to explain how a relatively small number of genes – only 36,000 in the human genome – can produce the hundreds of thousands of different proteins that regulate development, physiology, and disease.

"About 30 to 60 percent of all human genes are involved in alternative splicing, and about 15 percent of genetic diseases affect splicing," Heber said. But deciphering what bits of the original DNA code are re-assembled in multiple ways to

produce multiple proteins is a difficult biological and computational puzzle – one that Heber has been investigating for some time.

Heber completed his PhD at the German Cancer Research Center in Heidelberg and did post-doctoral work at the University of California San Diego, where he helped develop new computational methods for analyzing alternative splicing in the human genome.

"I was interested in coming to NC State because of the BRC's emphasis on statistics. It's an area I want to pursue further as a means of interpreting gene sequences," he said.

This spring, Heber will teach a "Special Topics in Bioinformatics" course on new computational approaches for DNA sequence analysis, database searching, and comparative genomics.

"About 30 to 60 percent of all human genes are involved in alternative splicing, and about 15 percent of genetic diseases affect splicing."

Steffen Heber,
BRC and Dept. of
Computer Science

SAS Intern Helps Zero in on Disease Genes

A team of developers at SAS, including NC State bioinformatics graduate student Christine Duarte, has collaborated with researchers from Purdue University and Duke medical Center to produce a new statistical procedure that will aid the search for genes associated with disease.

The procedure uses genetic markers to find locations (loci) on animal chromosomes that are associated with a disease or other “binary” trait (one that is either present or absent), said Duarte. The procedure, called Proc BTL (binary trait loci), also estimates genetic recombination rates between loci and the degree to which the trait is associated with a particular chromosome location.

“Once a locus is found, the biologists take over and try to find a gene in that region that controls the trait of interest,” said Duarte.

Working under the direction of Russ Wolfinger, head of the genomics department at SAS, Duarte was responsible for the successful programming and implementation of the statistical analyses. The

statistical approaches were developed by two NC State alumni, Lauren McIntyre (PhD Statistics and Genetics, 1996) of Purdue University, who is the principal investigator on the NSF-funded project, and Cynthia Coffman (PhD Biomathematics, 1997) of Durham Veterans Affairs Medical Center and Duke University Medical Center.

Duarte had virtually no knowledge of statistics when she began her internship last January. “The learning curve was steep,” said Duarte. She had to get up to speed on the theory behind BTL, refresh her knowledge of basic C programming, and learn the ropes of product development at SAS.

“The project is a nice mix of theoretical work and programming and it has introduced me to some important concepts in statistical genetics,” she said.

Duarte is currently in the master’s program but is considering pursuing a PhD. “Finding time for work, classes, and research is difficult, but at SAS I’ve jumped right into an important research area in bioinformatics, and I’m learning a lot.”

“At SAS I’ve jumped right into an important research area in bioinformatics.”

Christine Duarte,
Grad Student Intern

Industry Partnering Initiatives Expand

With 11 industry and institutional partners in the Triangle area, the BRC is poised to expand its off-campus partnership program under the direction of industry liaison Dr. Joshua Baker.

Baker, a former CEO of a local biotechnology company, joined the BRC this past summer. His initiatives have already led to meetings with several well-established and start-up biotech firms to explore new partnering opportunities.

The BRC’s current off-campus partners include GlaxoSmithKline, Quintiles, SAS Institute, Paradigm Genetics, and Dix Hospital, among others. Partnerships may entail joint research with BRC faculty; consultations in bioinformatics and

statistical genetics; student internships that provide tuition, stipends, and research opportunities; or adjunct faculty appointments.

Academic-industry collaborations offer increased opportunities for students and faculty to participate in applied research and for industry partners to gain from BRC expertise in the management and analysis of genomic data.

The BRC welcomes opportunities to meet with potential institutional partners to discuss options for collaborative research and student internships. Please contact Dr. Josh Baker, Industry Liaison, at jsbaker@unity.ncsu.edu or Dr. Bruce Weir, Director, at weir@stat.ncsu.edu.

Entering Bioinformatics Class Is Largest Ever

This year 23 new students, culled from 235 applicants, were accepted into the bioinformatics graduate program, bringing the total number of current students to more than 50.

The number of entering students has increased dramatically each year of the program’s four-year history. Student qualifications have remained consistently high. The bioinformatics program

continues to attract outstanding students from prestigious universities who are drawn to NC State by the reputation of the university and faculty in bioinformatics and statistical genetics.

Most bioinformatics graduates remain in the area after graduation. To date, 13 of 19 graduates with master’s and PhD degrees work in industry or research institutions in the Triangle area.

Thirteen of 19 graduates with master’s and PhD degrees in bioinformatics reside and work in the Triangle area.

“Bioinformatics is coming to the rescue” in the search for genes related to function.

Dr. David Bird
Center for the Biology of
Nematode Parasitism

Fall BRC Retreat a Success

More than 40 faculty, staff, students, and family members enjoyed a weekend of research and recreation at the Bioinformatics Research Center’s first fall retreat, held September 12 – 14 at Lakeview Lodge in western North Carolina.

The retreat gave first- and second-year graduate students the opportunity to hear BRC faculty and students describe their current research and to consider the direction their own research might take. Entering students are encouraged to get involved in research their first summer, and by the second year students must identify a major advisor and dissertation research area.

With eighteen graduate students, post-docs, and faculty speaking over the course of two mornings, participants were exposed to a broad spectrum of applied and theoretical research in bioinformatics.

“Bioinformatics is coming to the rescue,” said Dr. David Bird, Co-Director of the Center for Nematode Biology and Parasitism, referring to the vital role of computational tools in understanding biology.

“There’s a lot of people here doing some really interesting work,” said Hermonta Godwin, a second-year student. “They’re bringing the theory into the lab, which is pretty cool.”

“The level of mastery of the third- and fourth-year students is impressive,” said David Aylor, another second-year grad student. “The retreat helped me see how far along people are at different stages of their program.”

Whitewater rafting, hiking, and the Saturday night barbecue contributed to the congeniality of the weekend. The response to the weekend was universally positive, and there’s already enthusiastic talk about next year’s retreat.

The entire BRC applauds Dr. Spencer Muse for organizing the retreat and doing so much to make it fun and successful.



Grad student Frank Mannino describes his research on rates of molecular evolution at the BRC fall retreat.

Recent Publications

- Aris-Brosou S. 2003. How Bayes tests of molecular phylogenies compare with frequentist approaches. *Bioinformatics* 19(5): 618–624.
- Barrier M, Bustamante CD, Yu J, and Purugganan MD. 2003. Selection on rapidly evolving proteins in the Arabidopsis genome. *Genetics* 163: 723–733.
- Brenner CH and Weir BS. 2003. Issues and strategies in the DNA identification of World Trade Center victims. *Theoretical Population Biology* 63: 173–178.
- Bushel PR, et al. 2003. Computational selection of distinct class- and subclass-specific gene expression signatures. *J. Biomedical Informatics* 35: 160–170.
- Chhabra SR, Shockley KR, Conners SB, Scott KL, Wolfinger RD, and Kelly RM. 2003. Carbohydrate-induced differential gene expression patterns in the hyperthermophilic bacterium *Thermotoga maritima*. *J. Biological Chemistry* 278(9): 7540–7552.
- Chu TM, Weir BS, and Wolfinger R. (in press) Comparison of Li-Wong and loglinear mixed models for the statistical analysis of oligonucleotide arrays. *Bioinformatics*.
- Gibson G. 2003. Population genomics: celebrating individual expression. *Heredity* 90(1):1–2.
- Heber S and Stoye J. 2003. The European Conference on Computational Biology. *Drug Discovery Today* 8(3): 113–114.
- Lalush DS. 2003. Characterization, modeling and simulation of mouse microarray data. In *Methods of Microarray Data Analysis III*, S. Lin and K. Johnson, eds. Kluwer Academic Publishers.
- Laurie C and Weir BS. 2003. Dependency effects in multi-locus match probabilities. *Theoretical Population Biology* 63: 207–219.
- Law B, Buckleton JS, Triggs CM, and Weir BS. 2003. Effects of population structure and admixture on exact tests for association between loci. *Genetics* 164: 381–387.
- Purugganan MD and Gibson G. 2003. Merging ecology, molecular evolution, and functional genetics. *Molecular Ecology* 12(5): 1109–1112.
- Riley RM and Gibson G. 2003. Contrasting selection pressures on components of the Ras-mediated signal transduction pathway in *Drosophila*. *Molecular Ecology* 12(5): 1315–1323.
- Robinson DM, Jones DT, Kishino H, Goldman N, and Thorne JL. 2003. Protein evolution with dependence among codons due to tertiary structure. *Molecular Biology and Evolution* 20(10): 1692–1704.
- Scholl EH, Thorne JL, McCarter JP, and Bird DMck. 2003. Horizontally transferred genes in plant-parasitic nematodes: a high-throughput genomic approach. *Genome Biology* 4:R39.
- Skalski GT and Gilliam JF. 2002. Feeding under predation hazard: testing models of adaptive behavior with stream fish. *Am Naturalist* 160: 158–172.
- Skalski GT and Gilliam JF. 2003. A diffusion-based theory of organism dispersal in heterogeneous populations. *American Naturalist* 161(3): 441–458.
- Tao Y, Zeng ZB, Li J, Hartl DL, and Laurie CC. 2003. Genetic dissection of hybrid incompatibilities between *Drosophila simulans* and *D. mauritiana*. II. Mapping hybrid male sterility loci on the third chromosome. *Genetics* 164: 1399–1418.

Dr. Ohta's once-controversial theories are now established as part of the mainstream of evolutionary theory.

NC State Alumna Gives 2003 Cockerham Lecture

Dr. Tomoka Ohta, Professor Emeritus at the National Institute of Genetics, Japan, and the originator of the "nearly neutral" theory of genetic evolution, gave this year's C. Clark Cockerham Lecture on October 17 at Witherspoon Auditorium. The annual lecture brings as speakers outstanding scientists in the field of statistical genetics.

Dr. Ohta, with her colleague Motoo Kimura, generated considerable controversy in the late 1960s and the 1970s by suggesting that evolution was a random process arising from small random changes in DNA. Ohta went on to develop both this neutral theory and a "nearly neutral" theory, which incorporates natural selection as a force in evolution. Her

work is now regarded as part of the mainstream of evolutionary theory.

Dr. Ohta received a PhD in genetics from NC State in 1966. Her advisor was Ken-ichi Kojima. Dr. Ohta was honored last year with an award from the Emperor of Japan, and she was also elected to the U.S. National Academy of Sciences.

The Cockerham lecture is jointly sponsored by the departments of Statistics and Genetics and the BRC. It honors Dr. C. Clark Cockerham, a Distinguished University Professor who taught in the statistics department from 1953 to 1990 and was internationally recognized as a profound and original thinker in the field of population and quantitative genetics.

On-Campus Collaborations Increase

The BRC assists other researchers on campus with statistical analyses of genomic data and with bioinformatics computing and data management. Collaborations may range from short conversations to more formal consultations and joint research.

Senior programmer Chris Smith is designing database management systems and query tools to support large-scale studies of the loblolly pine and rice genomes being conducted by the Forest Biotechnolo-

gy Group and by Dr. Michael Purugganan, respectively. Dr. Jeffrey Thorne, a specialist in molecular evolution, is working with Dr. Brian Wiegmann of Entomology and with Dr. David Bird of the Center for Biology of Parasitism in Nematodes. Other campus partners include the Fungal Genomics Laboratory, the Center for Computational Biology, and Crop Science.

Inquiries about possible collaborations can be directed to Dr. Dahlia Nielsen (dahlia@stat.ncsu.edu).

BRC
bioinformatics
is published by the
Bioinformatics Research
Center three times per year.

Director: Bruce Weir
Editor/Writer: Pat Westphal

Address correspondence and
suggestions to the editor
(pawestph@unity.ncsu.edu).

Bioinformatics Research Center
North Carolina State University
Campus Box 7566
Raleigh, NC 27695-7566

NON-PROFIT ORG.
U.S. Postage
PAID
Raleigh, NC
Permit No. 2353