

BRC. bioinformatics

"With our approach, we can directly couple phenotype and genotype into evolutionary studies. This is a case where the development of statistical tools has substantial value to biology."

Dr. Jeffrey Thorne,
Depts. of Genetics
and Statistics

New Techniques Quantify Effects of Phenotype on Evolution of Genotype

An international research group led by Jeffrey Thorne, BRC faculty member and associate professor of genetics and statistics at NC State, has developed novel statistical procedures that couple the effects of phenotype (the observable physical or biochemical traits of an organism) and genotype (the DNA that codes for the traits) in the study of evolution. Their procedures may eventually help molecular scientists design better vaccines. They almost certainly will alter how scientists study the evolution of genes.

The relationship of phenotype and genotype is central to modern biological research, from the genetics of human disease and drought-resistance in crops to the evolution of viruses and color vision. Computational biologists are increasingly able to quantify this relationship and predict aspects of phenotype based solely on the genotype.

To better understand the reverse process—how phenotype affects the evolution of genotype—Thorne's group has developed novel statistical approaches and combined them with the tools of computational biologists to integrate phenotype and genotype into evolutionary studies.

"With the statistical procedures we've formulated, the methods devised by computational biologists for predicting phenotype from genotype can be incorporated into probabilistic models of DNA or amino acid sequence change. We can then draw statistical inferences about evolutionary history and evolutionary process. This is a case where the development of statistical tools has substantial value to biology," Thorne said.

Now, with the help of a recent grant from the National Institutes of Health, they will refine and expand the application of those procedures.

"We can adapt our procedures to use any *in silico* system for predicting phenotypic traits from genotype. We can then directly couple phenotype and genotype into evolutionary studies. Traditionally, the relationship between genotype and phenotype has been studied more indirectly by looking for unusual evolutionary patterns in genes known to affect phenotype or for phenotype effects of genes with unusual evolutionary patterns. Our approach is more direct and more informative," said Thorne.

Thorne's group intends to adapt the method to reconstruct DNA sequences that coded for ancestral viral proteins. "Improved ancestral sequence reconstruction

Continued on page 4

From the Director

2005 is a landmark year for the BRC as we approach five years since its authorization as a center by the Board of Governors of the UNC system in 2000. We are happy to say our NIEHS training grant has been renewed and expanded for another five years and now supports ten graduate students.

In these five years, the BRC has strengthened its international reputation as a preeminent center for graduate training and research in bioinformatics and statistical and population genetics. The Center owes its renown to our outstanding and innovative faculty and to our superior students and research associates. The work of two of our current students, both interns at Duke University's Center for Human Genetics, is described in this issue.

We are happy to say that our NIEHS training grant has been renewed and expanded for another five years and now supports ten graduate students.

Another 2005 landmark is the 10th anniversary of the annual Summer Institute in Statistical Genetics, which has been continually funded by NSF and NIH. In late July we are also teaching a week-long Institute at Yonsei University in Seoul, Korea.

We are pleased to welcome several post-doctoral associates and visiting research scientists to the Center. Please meet them in the article on page 2.

In the farewell department, we congratulate Dr. Garrick Skalski, a research associate in organismal and population biology, on his new faculty position with the department of zoology at the University of Kansas.

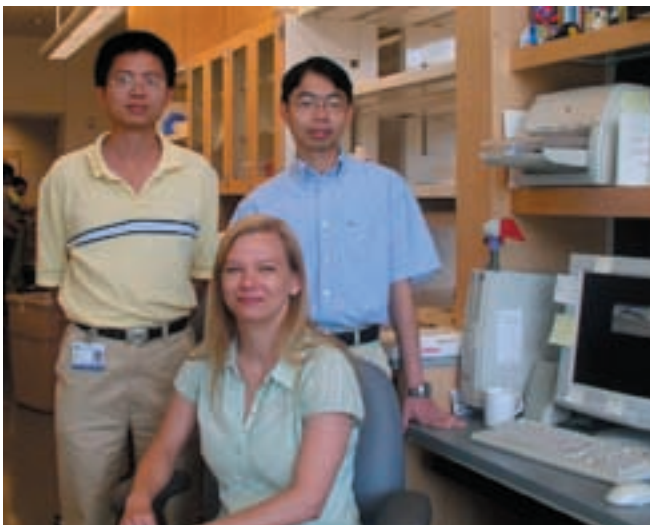
And another warm welcome - to the eight new graduate students who will join the Bioinformatics program this August.

Bruce Weir

Interns Probe Genetics of Autism, Alzheimer's Disease

As grad student interns at Duke University's Center for Human Genetics (CHG), NC State bioinformatics students Hao Mei and Ren-Hua Chung are helping create better tools for exploring the genetics of complex human diseases.

Hao Mei has helped develop new software, called EMDR 1.90 (for "extended multiple dimensionality reduction"), for identifying multiple gene interactions in autism and other diseases. The new program is able to identify genetic patterns (combinations of several gene locations) that appear to be associated with the disease even if each gene location analyzed alone does not.



Graduate student interns Hao Mei (left) and Ren-Hua Chung (right) with Dr. Eden Martin at the Duke University Center for Human Genetics.

Working under the direction of Dr. Allison Ashley-Koch and Dr. Eden Martin (Ph.D. 1997 NC State), Mei incorporated advanced test statistics and analytical methods into the program and designed efficient data structures and algorithms to optimize memory utilization and improve program efficiency.

When applied to the autism data, EMDR was able to detect interactions among three gamma aminobutyric acid (GABA) receptor subunit genes on chromosome 15. Mei is co-author on two papers that are accepted for publication and a third that is submitted. A new

project, MDR-Phenomics, aims to integrate clinical phenotype information with the genetic analyses.

More than 20 people from different fields – clinical medicine, biology, statistics, and computer science – attend weekly meetings on the autism project. "I am honored to be working with a diverse team and participating in several challenging projects that will be components of my Ph.D. thesis," Mei said. Prior to coming to NC State, Mei earned a degree in medicine from China Medical University in the People's Republic of China and an M.S. in biomedical informatics from the University of Washington.

Ren-Hua Chung, who trained as a computer scientist in Taiwan and at the University of California at Davis, is busy on two projects at CHG. As a data analyst on the Alzheimer's disease project, he checks for genotyping errors in family data using tools such as pedigree consistency and Hardy-Weinberg assumption testing. Then he performs linkage and association analyses using various software packages.

"We are excited to have found significant association of several genes and gene variants with Alzheimer's disease," said Chung, who is co-author on a paper to appear in the *Journal of Medical Genetics* and two conference presentations related to the project.

Chung also works with Eden Martin in developing statistical methods for family-based association analyses (finding genes associated with a trait such as a complex disease). Currently they are studying association in the presence of linkage (APL), which Chung implemented as a computer program.

"APL can handle missing parental genotypes and it works correctly even when the disease loci and the markers being tested are linked on the chromosome," said Chung. Scientists at CHG have used the APL program to help find a gene variant that is a major risk factor for age-related macular degeneration (Haines JL, et al. *Science*, April 15, 2005).

"The abundance of data at CHG is a great asset for exploring disease-causing genes," said Chung.

New Faces at the BRC

The BRC extends a warm welcome to post-doctoral research associate Dr. Asger Hobolth, from the Bioinformatics Research Centre at the University of Aarhus. Asger has joined Jeff Thorne's research group to work on developing statistical measures of the interaction of phenotype and genotype.

Post-doctoral fellows Dr. Augusto Garcia and Dr. Luciana Carlini-Garcia of Brazil have been working with Zhao-Bang Zeng since last fall on QTL mapping

and developing multi-QTL analytical methods to explore and compare the genetic basis of hybrid vigor of yields in maize and rice.

Dr. Diane Byers, on sabbatical from Illinois State University, is also working with Zhao-Bang on environmental effects on QTL expression in *Arabidopsis*. She returns to Illinois in late August.

It is a pleasure to have these talented scientists at the BRC.



Dr. Bruce S. Weir (right), founding director of the BRC, with NCSU Chancellor James L. Oblinger at the 2005 Holladay Medal presentation ceremony.

Weir Receives Prestigious Holladay Medal

Bruce Weir, founding director of the BRC and William Neal Reynolds Distinguished Professor of Genetics and Statistics, was one of five NC State faculty members to receive the Alexander Quarles Holladay Medal for Excellence, the highest honor bestowed on a faculty member by the university. The awards were presented by the chancellor and

board of trustees at ceremonies held on May 12.

Weir was honored for his pioneering research in statistical and population genetics, his continuing leadership in the field of bioinformatics and DNA forensics, and his many years of service to education at NC State and within the international scientific community.

Summer Institute Celebrates 10 Years, Record Enrollment

A record-high enrollment enlivened the 10th anniversary year of the Summer Institute in Statistical Genetics, held for three weeks in May and June on the NC State campus. The Summer Institute has offered workshops in genetics and the latest methods of statistical analysis of genetic data since 1996. That year, 7 modules were taught by 11 NC State faculty and attended by 95 participants.

By 2005, the Institute had grown nearly three-fold, offering up to 18 modules taught by 41 faculty from NC State and other U.S. and international research groups and attended by more than 270 participants from the U.S. and abroad.

International institutes, first offered in 2000, have

been held in New Zealand, Ireland, Australia, Portugal, and, in July 2005, South Korea.

"The growing success of the Summer Institute reflects the dedication and high quality of our instructors and their efforts to present new material each year. This year's module on Genetics in Clinical Trials is just one example of how the Institute focuses its training in new directions," said Institute director Bruce Weir.

Debra Hibbard, the BRC's executive assistant and the Summer Institute's able administrator since its inception, has managed the increase in the Institute's scope without a hitch. Trena Langhurst has been her energetic and indispensable aide since 2002.

Gene Expression, Evolution, and Sequencing Are Discussed at Joint NESCent/BRC Symposium

The evolutionary significance of microRNAs, interactions between gene expression and behavior, and the usefulness of genome sequencing projects were some of the topics discussed by an international group of scientists at the first Symposium on Comparative Genomics, held June 4 at NC State.

Sponsored by the National Evolutionary Synthesis Center (NESCent) in Durham and the BRC, the symposium was hosted by Dr. Greg Gibson, professor of genetics at NC State and NESCent's outreach and education director, and by BRC's director Dr. Bruce Weir.

"What a difference a point mutation makes" was the theme of two of the talks. Dr. Michel Georges of the University of Liège (Belgium) found that a point mutation in a microRNA controlled the unusual expression pattern of a mutant gene associated with enlarged muscle fibers in the hindquarters of sheep. "Bioinformatics predicts the existence of many microRNAs, and we're just beginning to investigate their evolutionary significance," Georges said.

Dr. John Doebley of the University of Wisconsin also found that a single regulatory element and a point mutation in a second gene accounted for many of the important differences between domestic corn and its wild ancestor, the grass teosinte.

Dr. Christina Grozinger of the Department of Entomology and the W.M. Keck Center for Behavioral Biology at NC State described her research on the genetic mechanisms underlying the complex behaviors of honeybees and how exposure to pheromones regulates gene expression in the brain.

Dr. Rytas Vilgalys of Duke University leads an international effort to develop a genome database for the entire kingdom of fungi. Biologists are beginning to identify previously unknown fungi using the genome database and can place new species on the tree of life. "The outline of the tree of life has entered the molecular era," he said.

Dr. Stephen Richards of Baylor College of Medicine, who is supervising the genome sequencing of several insects, noted that the ability to generate sequences has outpaced the ability of scientists to make experimental observations based on differences in gene sequence.

Dr. David Goldstein, a pharmacogeneticist at Duke, discussed the clinical relevance of identifying genetic variants associated with a disease or with dose-response to drugs. Using epilepsy as an example, he proposed that the discovery of genetic variants of enzymes in the metabolic pathway of a drug is much more likely to yield clinically relevant information for intervention in non-responsive patients.

"Bioinformatics predicts the existence of many microRNAs, and we're just beginning to investigate their evolutionary significance."

Michel Georges,
University of Liege

Evolutionary Geneticist to Give 2005 Cockerham Lecture

Geneticist Svante Pääbo, director of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, and a pioneering researcher who has shaped the study of the molecular evolution of humans, will give this year's C. Clark Cockerham Lecture on Friday, September 30 at 4 p.m. in 2215 Williams Hall.

Dr. Pääbo's groundbreaking research in the genetics of modern and ancient humans and apes has shed new light on human evolution and the emergence of genetic features that underlie distinctive human characteristics such as language and reasoning.

As a graduate student in Sweden in the 1980s, Pääbo became the first to isolate DNA from ancient humans (Egyptian mummies). Since then his research has addressed the DNA sequencing and phylogeny of numerous extinct and modern species, including Neanderthals; the use of mitochondrial DNA in de-

termining phylogeny; and the comparative genomics of apes and humans, including differences in genes related to speech and olfaction.

Dr. Pääbo received a Ph.D. in cell biology from the University of Uppsala in 1986 and did post-doctoral research with Allen Wilson at Berkeley. Director of the Max Planck Institute for Evolutionary Anthropology since its founding in 1997 and the recipient of numerous awards, Dr. Pääbo was elected a foreign member of the U.S. National Academy of Sciences in 2004.

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.

New Techniques, *cont. from page 1*

of retroviruses like HIV-1 is one of our top priorities, because scientists have suggested that ancestral viral sequences may lead to better vaccines," said Thorne. Tae-Kun Seo, a former post-doc with Thorne who is now on the faculty of the University of Tokyo, plans to work with Thorne to examine how viral evolution is shaped by the host immune system or antiviral agents.

The project originated with the thesis research of Douglas Robinson, who is now in the proteomics group at Bristol-Myers Squibb. Robinson was intrigued by the observation that a protein's three-dimensional structure (i.e., phenotype) tends to evolve much more slowly over time than does the DNA sequence that codes for the protein (i.e., genotype). In other words,

there seem to be constraints on the evolution of the gene that serve to maintain protein structure. Robinson, Thorne, and their collaborators developed statistical techniques for investigating the evolution of DNA sequences that incorporate the interdependence of sequence changes and protein structure.

BRC grad student Jiaye Yu is applying similar statistical techniques to look at the relationship between RNA secondary structure and evolution. Sang Chul Choi, another grad student, is extending Robinson's work to study evolutionary patterns in a large protein database. Other collaborators are Hirohisa Kishino, University of Tokyo; Nick Goldman, European Bioinformatics Institute; and David Jones, University College London.

"Improved ancestral sequence reconstruction of retroviruses like HIV-1 is one of our top priorities."

Dr. Jeffrey Thorne,
Depts. of Genetics
and Statistics

Recent Publications

Atchley WR, and Fernandes AD. 2005. Sequence signatures and the probabilistic identification of proteins in the Myc-Max-Mad network. *PNAS USA* 102:6401-6406.

Atchley WR, Zhao J, Fernandes AD, and Druke T. 2005. Solving the protein sequence metric problem. *PNAS USA* 102:6395-6400.

Dworkin I, Palsson A, and Gibson G. 2005. Replication of an Egfr-wing shape association in a wild-caught cohort of *Drosophila melanogaster*. *Genetics* 169: 2115-2125.

Goering LM, and Gibson G. 2005. Genetic variation for dorsal-ventral patterning of the *Drosophila melanogaster* eggshell. *Evol and Devel* 7(2): 81-88.

Gibson G. 2005. The synthesis and evolution of a supermodel. *Science* 307:1890-1891.

Gibson G. 2005. Mutation accumulation of the transcriptome. *Nature Genetics* 37(5): 458-460.

Halverson J, and Basten C. 2005. A PCR multiplex and database for forensic DNA identification of dogs. *J Forensic Science* 50(2):352-363.

Psarros M, Heber S, Sick M, Thoppae G, Harshman K, and Sick B. 2005. RACE: Remote analysis computation for gene expression data. *Nucl Acids Res* 33: W638-W643; doi:10.1093/nar/gki490.

Seo T-K, Kishino H, and Thorne JL. 2005. Incorporating gene-specific variation when inferring and evaluating optimal evolutionary tree topologies from multi-locus sequence data. *PNAS* 102(12): 4436-4441.

Stafford-Smith M, Podgoreanu M, Swaminathan M, Phillips-Bute B, Mathew JP, Hauser EH, Winn M, Milano C, Nielsen DM, et al. 2005. Association of genetic polymorphisms with risk of renal injury after coronary bypass graft surgery. *Am J of Kidney Disease* 45(3):519-531.

Starmer J, Stomp A, Vouk M, and Bitzer D. 2005. Predicting fidelity of protein synthesis in *E. coli*. *Proc. IEEE International Workshop on Genomic Signal Processing and Statistics (GENSIPS 2005)*, May 22-24, 2005 Newport, Rhode Island.

Tzeng, J-Y. 2005. Evolutionary-based grouping of haplotypes in association analysis. *Genetic Epidemiology* 28:220-231.

Welsby IJ, Podgoreanu MV, Phillips-Bute B, White WD, Nielsen DM, et al. 2005. Genetic factors contribute to bleeding after cardiac surgery. *J Thromb Haemost* 3(6):1206-1212.

Zeng Z-B. 2005. QTL mapping and the genetic basis of adaptation: recent developments. *Genetica* 123:25-37.

Zeng Z-B, Wang T, and Zou W. 2005. Modeling quantitative trait loci and interpretation of models. *Genetics* 169:1711-1725.