

BRC. bioinformatics

New Faculty Member Aims to Predict Severity of Genetic Diseases



Dr. Eric Stone, assistant professor of statistics, has developed a method of scoring genetic variants to predict disease risk. Results were consistent with clinical evidence for three human diseases.

The geneticist's dream of predicting the likelihood and severity of disease by analyzing variations in the DNA code of human genes is a step closer to reality thanks to the work of Dr. Eric Stone, who joined the Department of Statistics faculty and the BRC this fall.

"Geneticists want to predict the health consequences of human coding variations," said Stone. To that end, Stone developed a method called MAPP (for Multivariate Analysis of Protein Polymorphism) that scores a nucleotide substitution at any location in protein-coding sequences of DNA based on how much that substitution alters the physicochemical properties of the protein. The further the score deviates from the range of tolerated substitutions observed in normally functioning genes, the greater the impairment of protein function and the greater the risk or severity of disease.

MAPP accurately predicted the degree of severity of anemia and of Hirschsprung disease (a congenital birth defect affecting the lower intestine) and the likelihood of tumor growth by scoring coding variations in genes associated with these diseases (see Stone and Sidow, 2005, and Kashuk et al., 2005, in Recent Publications).

"The MAPP score distributions were consistent with the clinical evidence," said Stone. "As better comparative mammalian sequence data and larger genotype collections become available, MAPP's accuracy is likely to improve."

Stone's other areas of research include using comparative genomics to detect highly conserved and presumably functional DNA sequences (see Stone et al. and Cooper et al. in Recent Publications), evaluating the robustness of phylogenetic methods and its effects on the outcome of evolutionary studies, and theoretical approaches to assessing the co-evolution of traits.

Stone earned his Ph.D. in statistics from Stanford University in 2004, working with geneticist Arend Sidow, with whom he also did a year of post-doctoral research.

"Here at NC State, I'm looking forward to collaborating with Jeffrey Thorne and Spencer Muse on developing molecular evolutionary strategies that reveal associations between genotype and phenotype," said Stone. This fall, Stone is also teaching Introduction to Statistics (ST 311), a class for non-majors, whom he hopes to engage with practical applications of statistical methods.

From the Director

This fall we celebrate the fifth anniversary of the founding of the BRC and our move into the Partners II building on Centennial Campus. We have accomplished a great deal in this short time, and the BRC is now an integral part of the intellectual climate at NC State University and in the Research Triangle. Our affiliated bioinformatics graduate program will recognize its 14th Ph.D. and 24th Master of Bioinformatics graduates at graduation on December 14. Our annual income from external sources has quadrupled to over \$4 million. Our faculty publish widely and are receiving international recognition for their work. Our staff have worked very effectively to create a wonderful working environment.

Our pride in past accomplishments, however, is accompanied by the need to adapt to some changes. Early

in the fall we said farewell to Debbie Hibbard, our executive assistant, who has taken a position with Professional Meeting Planners Network, where she will be employing her considerable client service and managerial skills. Our systems analyst, Chris Basten, is broadening his responsibilities substantially with a new position at Syngenta. He will be working in the areas of statistical genetics and bioinformatics at several locations throughout the world, although his home base will be in Durham. Our communications specialist, Pat Westphal, is logging off her computer after producing this, our 7th newsletter. We wish these three colleagues much happiness and success in their new endeavors.

My own situation is also about to change.

On January 1, 2006 I will assume the position of

Record Number of Student Research Talks at Fall Retreat

The sandy shore of North Carolina's coast was the setting of the BRC's third annual fall retreat, held at NC State's Center for Marine Science and Technology in Morehead City and the Seahawk Motor Lodge in Atlantic Beach on September 9 - 11.

About 60 people attended this year's retreat, and a record number of speakers (23 bioinformatics graduate students and 4 post-docs) described their current research on topics such as gene mapping methods, gene expression regulation and profiling, population genetics, and molecular evolution of genes and proteins.

The retreat offers an opportunity for students at all levels of study to share some "R&R" (research and recreation) with their colleagues, many of whom work in labs and research centers outside the BRC.

"The BRC retreat is a great opportunity to network and socialize with other bioinformatics students that I don't see on a regular basis," said ClarLynda Williams-DeVane, a third-year student who is doing her Ph.D. research at EPA's National Center for Computational Toxicology in Research Triangle Park.

Hearing about others' research can spur new thinking about one's own work. "The comments after my talk gave me new perspectives and ideas to extend my research in other directions," Williams-DeVane said. "It was very helpful to me to see that although I have an unconventional bioinformatics research project,

there is other research going on at the BRC that is very much applicable to my work."

"I had a great time at the retreat, although I was a bit nervous about presenting my work because I'm just beginning my research," said second-year student Brian Howard, who spoke on prioritizing candidate disease genes with machine learning algorithms. "Fortunately, people were very polite and supportive. Afterwards, a few students and faculty offered their suggestions and insight into my research questions, and for me this was the most valuable aspect of the retreat."

"It was a fantastic retreat, maybe our best ever," said Zhao-Bang Zeng, professor of statistics and genetics and a mentor of several graduate students and post-docs at the BRC. "The student presentations were varied and excellent."

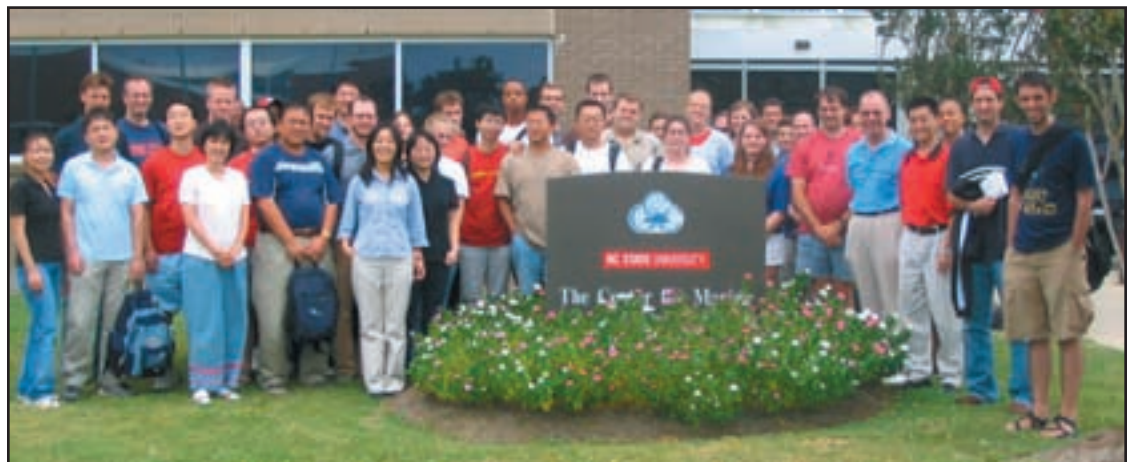
Beach soccer, shelling, and "sun" bathing (under skies made cloudy by off-shore hurricane Ophelia) filled the afternoons, while the Friday night welcome cookout (by grill chefs Josh Starmer and Errol Strain) and the catered barbecue dinner Saturday night marked the evening festivities.

The entire weekend was organized by the students, led by David Aylor, Jon Keebler, Jessica Maia, Kate McGee, Josh Starmer, and Errol Strain. Thanks also to Jubileth Briseno, the BRC's graduate student secretary, for suggesting the hotel and help with reservations.

"The BRC retreat is a great opportunity to network... with other bioinformatics students that I don't see on a regular basis. It was very helpful to me to see that although I have an unconventional bioinformatics research project, there is other research going on at the BRC that is very much applicable to my work."

ClarLynda Williams-DeVane,
Grad Student Intern at EPA

BRC students and faculty attend the 2005 fall retreat, held at NC State's Center for Marine Science and Technology in Morehead City, NC.



From the Director, cont.

professor and chair of the Department of Biostatistics at the University of Washington. It is just over 40 years since I arrived at NC State as a graduate student, and nearly 30 years since I joined the faculty of the Department of Statistics. These have been rich and rewarding years for me, but the last five years at the BRC have been especially satisfying. It has been a wonderful experience to be part of the development of this enter-

prise and to feel that what we do here is important for our science and our university. My excitement at moving to Seattle is very much tempered by having to lose daily contact with the very fine group of people at the BRC and in the genomic sciences program at NC State.

My best wishes to the BRC and to Zhao-Bang Zeng as he takes responsibility for writing this column in the future.

Bruce Weir

Statistical Sleuth Sharpens Gene Mapping Tools

Scientists who are trying to pinpoint genetic differences that are associated with human disease face a staggering task. While human DNA from any two unrelated individuals differs by only 0.1 percent, that small percentage translates into 3 million pairs of nucleotides, the molecules that form the DNA chain. To genotype all of these differences (referred to as single nucleotide polymorphisms or SNPs) in thousands of people to identify disease-related variants is a daunting and expensive prospect.

Dr. Dahlia Nielsen, research assistant professor of statistics at the BRC, has taken a close look at human DNA sequences with an eye to improving current methods for identifying or “mapping” genetic variants associated with disease.

Geneticists hope to streamline association mapping studies by finding segments or “blocks” of DNA that contain a number SNP variants that frequently occur together and then selecting one or a few of the SNPs to represent the entire block. This would greatly reduce the number of SNPs in a DNA sample that required genotyping to identify variants associated with disease.

“Many investigators are taking for granted that we can reduce the number of SNPs used in association mapping, but they haven’t examined how that assumption might weaken their ability to identify disease-related variants,” said Nielsen. Nielsen hopes to increase the statistical validity and accuracy of these association mapping strategies.

“The assumption is that if two SNPs are correlated with one another, each SNP will be correlated with the disease at an equivalent level. But my analysis has shown that the associations between each SNP and the trait under study and the associations among the different SNPs are both important. In some cases, reducing the number of SNPs evaluated could result in valuable information being lost.”

Nielsen was invited to describe the results of her analyses at the American Society of Human Genetics in October.

Nielsen is also applying her statistical sleuthing talents to help Dr. Natasha Olby of the College of Veterinary Medicine search for genetic markers associated with a late-onset neurological disorder in dogs called cerebellar ataxia. Simulation studies performed by Nielsen confirmed that the number of DNA samples that Olby has available from a large family of dogs was suitable for valid statistical gene mapping studies.

Their research is currently funded by a grant from the American Kennel Club, but Olby and Nielsen hope to obtain funding from the National Institutes of Health for further research because the disease also occurs in humans. The gene responsible for the disorder in dogs may also be responsible for the human disease. “If we can find the gene in dogs, it brings us one step closer to understanding the human disease,” said Nielsen.

“Many investigators are taking for granted that we can reduce the number of SNPs used in association mapping, but they haven’t examined how that assumption might weaken their ability to identify disease-related variants.”

Dr. Dahlia Nielsen,
Dept. of Statistics

Three BRC Affiliates Earn Distinguished Professorships

The BRC is proud to congratulate our colleagues Zhao-Bang Zeng, Greg Gibson, and Michael Purugganan on their appointments as William Neal Reynolds Distinguished Professors, the highest honor open to faculty members in the College of Agriculture and Life Sciences. The appointments recognize their outstanding research programs in quantitative genetics and evolutionary biology and their international standing as leading scientists in their fields.

Zeng is a theoretical and statistical geneticist whose groundbreaking work in statistical methods for analyzing the genetic architecture of complex traits has had important applications in plant and animal breeding programs, human health research, and the study of evolution. Zeng joined the statistics faculty at NC State in 1990 and has held a joint appointment in the Department of Genetics since 2001.

Gibson, professor of genetics and assistant director of life sciences at the NC Agricultural Research Service, has made important advances in inves-

tigating the quantitative genetics of development and physiology, primarily using the model organism *Drosophila melanogaster* (fruit fly). He has been particularly interested in applying genomic approaches to address the nature of intra-species variation, focusing on the mapping of genotype and transcript abundance onto phenotypic variation. Gibson joined the NC State faculty in 1998.

Puruggan is a geneticist whose work in the molecular evolution and evolutionary ecology of wild and domesticated plants led to a multi-million-dollar NSF grant to study genetic variation in rice. Research interests include the molecular processes of adaptive selection in plants, the evolutionary genetics of plant domestication, and the dynamic interactions among genetics, evolution, and human culture. Purugganan joined the NC State faculty in 1995.

Other BRC-affiliated faculty who hold William Neal Reynolds Distinguished Professorships are William Atchley, Trudy Mackay, and Bruce Weir.

Congratulations to
William Neal Reynolds
Distinguished
Professors

Zhao-Bang Zeng
Greg Gibson
Michael Purugganan

New Statistical Method May Help Explain Gene Evolution



“Our method is much faster than previous methods [and has] great potential for increasing scientists’ understanding of the ... biological processes that shaped the genomes of present-day species.”

**Dr. Asger Hobolth,
BRC Post-doctoral
Research Fellow**

Dr. Asger Hobolth, a post-doctoral research associate who joined the BRC last spring, is working with Dr. Jeffrey Thorne to develop novel statistical methods to improve scientists’ ability to model the evolution of genes.

Until recently, genetic evolutionary models have assumed that nucleotides at each site in a DNA sequence change independently of neighboring sites, an assumption that is computationally manageable but biologically inaccurate. Thorne has been a leader in developing evolutionary models that incorporate interdependence among sites. But these new context-dependent models have had limited use because they are so time consuming to implement.

Now Hobolth and Thorne have developed and successfully tested a novel statistical procedure, called “exact path sampling,” for simulating context-dependent DNA sequence evolution.

“Our method is much faster than previous methods,” said Hobolth. “Efficient context-dependent evolutionary models have great potential for increasing scientists’ understanding of the structure and function of human genes and of the biological processes that shaped the genomes of present-day species,” he said.

The new procedure is also time-reversible, meaning that ancestral DNA sequences can be simulated from modern genes. This is an important feature for helping establish the times of divergence of species or for identifying conserved ancestral sequences in rapidly evolving species such as viruses.

Hobolth obtained his Ph.D. in theoretical statistics from Aarhus University in Denmark in 2002 and has done three years of post-doctoral research at the Bioinformatics Research Center in Aarhus.

“The main reason I’m here at NC State is to learn more about the biology of molecular evolution. Jeff Thorne is a very strong biologist,” said Hobolth.

Prior to working in DNA sequence analysis, Hobolth developed statistical models that characterized variability of cell shape in cancer tissue and that clarified the signals in diffusion tensor imaging of nerve connections in the brain. “The statistical methods applied in all these areas are surprisingly similar,” he said.

“I would like to thank the people at the BRC for being extremely friendly and helping me get settled when I arrived,” Hobolth said. You can’t take that for granted when you arrive in a new place.”

Recent Publications

- Chou JW, Paules RS, and Bushel PR. 2005. Systematic variation normalization in microarray data to get gene expression comparison unbiased. *J Bioinform Comput Biol* 3(2): 225–41.
- Coffman CJ, Doerge RW, Simonsen KL, Nichols KM, Duarte CK, Wolfinger RD, and McIntyre LM. 2005. Model selection in binary trait locus mapping. *Genetics* 170: 1281–97.
- Cooper GM, Stone EA, Asimenos G, NISC Comparative Sequencing Program, Green ED, Batzoglou S, and Sidow A. 2005. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res* 15: 901–913.
- Gardner RJ, Hobolth A, Jensen EBV, and Sorensen FB. 2005. Shape discrimination by total curvature, with a view to cancer diagnostics. *J Microscopy* 217: 49–59.
- Gibson G. 2005. The synthesis and evolution of a supermodel. *Science* 307: 1890–91.
- Gibson G. 2005. Q&A: Greg Gibson. *Current Biology* 15: R531–532.
- Hobolth A, and Jensen JL. 2005. Statistical inference in evolutionary models of DNA sequences via the EM algorithm. *Statistical Applications in Genetics Molecular Biology* 4(1): article 18.
- Hobolth A, and Jensen JL. 2005. Applications of hidden Markov models for characterization of homologous DNA sequences with a common gene. *J Computational Biology* 12: 186–203.
- International HapMap Consortium (BS Weir is among the many authors). 2005. A haplotype map of the human genome. *Nature* 437: 1299–1320.
- Jorgensen FG, Hobolth A, Hornshoj H, Bendixen C, Fredholm M, and Schierup MH. 2005. Comparative analysis of protein coding sequences from human, mouse and the domesticated pig. *BMC Biology* 3:2.
- Kashuk CS, Stone EA, Grice EA, Portnoy ME, Green ED, Sidow A, Chakravarti A, and McCallion AS. 2005. Phenotype-genotype correlation in Hirschsprung disease is illuminated by comparative analysis of the RET protein sequence. *PNAS* 102(25): 8949–54.
- Liu Y, and Zeng Z-B. 2005. Mixture model equations for marker-assisted genetic evaluation. *J Anim Breed Genet* 122: 229–239.
- Martin ER, Bronson P, Li Y-J, Wall N, Chung R-H, Pericak-Vance M, et al. 2005. Interaction between the alpha-T catenin gene (VR22) and APOE in Alzheimer’s disease. *J Med Genet* 42: 787–792.
- Moser JM, Freitas T, Arasu P, and Gibson G. 2005. Gene expression profiles associated with the transition to parasitism in *Ancylostoma caninum* larvae. *Mol Biochem Parasitol* 143: 39–48.
- Mu J, Awadalla P, Duan J, McGee KM, Joy DA, McVean GAT, and Su X. 2005. Recombination hotspots and population structure in *Plasmodium falciparum*. *Plos Biol* 3(10): e335.
- Palsson A, Dodgson J, Dworkin I, and Gibson G. 2005. Tests for the replication of an association between Egfr and natural variation in *Drosophila melanogaster* wing morphology. *BMC Genet* 6: 44.
- Shockley KR, Scott KL, Pysz MA, Conners SB, Johnson MR, Montero CI, Wolfinger RD, and Kelly, RM. 2005. Genome-wide transcriptional variation within and between steady states for continuous growth of the hyperthermophile *Thermotoga maritima*. *App Environ Microbiol* (Sept 2005): 5572–76.
- Stone EA, Cooper GM, and Sidow A. 2005. Trade-offs in detecting evolutionarily constrained sequence by comparative genomics. *Annu Rev Genomics Hum Genet* 6: 143–164.
- Stone EA, and Sidow A. 2005. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Res* 15: 978–986.
- Strain E, and Muse S. 2005. Positively selected sites in the Arabidopsis receptor-like kinase gene family. *J Mol Evol* 61: 325–332.
- Thomson S, Kennerly E, Olby N, Mickelson J, Hoffman D, Dickinson P, Gibson G, and Breen M. 2005. Microarray analysis of differentially expressed genes of primary tumors in the canine central nervous system. *Veterinary Pathology* 42(5): 550–558.
- Weir BS, Cardon LR, Anderson AD, Nielsen DM, and Hill WG. 2005. Measures of human population structure show heterogeneity among genomic regions. *Genome Res* 15: 1468–76.