

BRC bioinformatics

Faro, Portugal, Hosts 2004 Summer Institute

Sunny southern Portugal was the setting for the international edition of the 2004 Summer Institute in Statistical Genetics. Invited by the University of Algarve's Center for Marine Sciences (CCMAR), the Institute offered 10 three-day workshops at the

university's beautiful seaside Gambelas Campus in Faro, Portugal, July 19–28.

More than 130 students from 20 countries attended workshops on quantitative, population, and behavior genetics, gene mapping, bioinformatics, microarray analysis, and DNA data analysis, led by 18 faculty from NC State and other U.S. and international institutions.

Participants were housed in hotels near Faro's medieval center, which is surrounded by Roman walls and marked by a 13th-century cathedral and other historical buildings.

The CCMAR conducts research in the molecular biology, genetics, ecology, and conservation of marine plants and animals. The Summer Institute was pleased to be asked to participate in the educational mission of this vital and internationally distinguished research center.



Trudy Mackay, Robert Anholt, and Bruce Weir of NC State join Kent Holsinger of the University of Connecticut for a break on the terrace during the Summer Institute in Statistical Genetics in Faro, Portugal.

Since the Graduate Industrial Traineeship program began in 2000, the BRC has received more than \$1 million in internship funds.

From the Director

Summertime is travel time for many BRC members who are engaged in scientific and professional activities off campus. Several faculty are serving on NIH and NSF grant review panels, and many of our faculty and students are traveling to meetings to present their research. Zhao-Bang Zeng is wrapping up a productive sabbatical in Edinburgh, Scotland, and a number of our faculty and administrative staff are taking the Summer Institute in Statistical Genetics to Faro, Portugal.

Traveling on later this year is Stephane Aris-Brosou, who has been a post-doc with Jeff Thorne for two years. Congratulations to Stephane on his award-winning paper (see Recent Publications) and his faculty appointment at the University of Ottawa, Ontario, Canada. We will miss him, as well as the sight of his black Mustang in the parking lot.

Moving on to a career in pharmacogenetics is Xiang Yu, who graduates in August after completing his dissertation research at SAS Institute with Adjunct Professor Russell Wolfinger. Xiang is one of the students who have been supported by our Graduate Industrial Traineeship (GIT) program, which places students with our partners in the Research Triangle.

Since the GIT program began in 2000, we have received more than \$1 million in internship funds. This is in addition to the \$1 million from an NIEHS training grant and the bioinformatics share of the NSF IGERT \$2.5 million training grant in genomic sciences. Annual income from all grants and contracts to the BRC now exceeds \$3 million, underscoring the fund-raising prowess of our faculty, which has contributed significantly to the success of our graduate program and our center.

Bruce Weir
weir@stat.ncsu.edu

“The learning environment at the BRC has been very useful.... Without this level of support, I think it would be difficult for biologists to receive the statistical training required for bioinformatics analysis.”

Errol Strain,
Bioinformatics
Graduate Student

Project Screens 600 Genes for Clues to Pathogen Resistance

Errol Strain, a 4th-year bioinformatics grad student, is searching the DNA sequences of over 600 genes in *Arabidopsis*, a small plant used in genetics research, to identify coding regions that might be involved in pathogen resistance.

The 600 genes code for a family of proteins called receptor-like kinases (RLKs), which are vital for activating immune responses as well as for signal transduction and cell development.

Strain's DNA screening approach identifies sites in multiple genes where it appears that the DNA code has changed frequently and has altered protein composition. These sites, in contrast to more stable coding regions, are thought to be under the influence of positive natural selection and may be associated with resistance to new or rapidly evolving pathogens.

“When I started this study two years ago, there were very few positive selection studies on large gene families,” Strain said. “There are a lot of technical and analytical difficulties to overcome in extending the analysis from a small group of genes to a much larger group.”

Dr. Michael Purugganan of the Genetics Department helped Strain plan and set up the project. At the BRC, Dr. Spencer Muse and Dr. Jeffrey Thorne are helping with the analysis and interpretation of his results.

Strain became interested in plant molecular genetics at Purdue University, where he majored in biochemistry and worked as a lab tech on sex determination in ferns.

“I became frustrated with not having the quantitative and technical background to deal with large amounts of DNA sequence data,” he said.

NC State's Bioinformatics Program has provided Strain the skills to analyze and interpret genomic data. “I was weak in statistics when I started the program. Not only the coursework, but the learning environment at the BRC has been very useful. Students and faculty with diverse backgrounds are able to interact on a daily basis. Without this level of support, I think it would be difficult for biologists to receive the statistical training required for bioinformatics analysis.”

Strain plans to return to lab work in the future, armed with new tools for probing the molecular evolution of plants.

Recent Publications

The Publisher's Award for Excellence, offered by the Journal Systemic Biology, went to BRC post-doc Stephane Aris-Brosou for his paper, “Least and most powerful phylogenetic tests to elucidate the origin of the seed plants in presence of conflicting signals under misspecified models,” which appeared in *Systematic Biology*, December 2003. The cash award is presented annually to the two best papers based on research conducted while a student. Congratulations, Stephane!

Most Frequently Downloaded: Pierre Bushel's 2002 paper, “Computational selection of distinct class- and subclass-specific gene expression signatures (*J Biomedical Informatics* 35:160-170), co-authored with HK Hamadeh, L Bennett, and others, was among the ten articles most frequently downloaded from the journal between January 2003 and December 2004. Pierre is manager of bioinformatics at the NIEHS MicroArray Center in Research Triangle Park and a graduate student at the BRC.

- Chu TM, Weir BS, and Wolfinger RD. 2004. Comparison of Li-Wong and loglinear mixed models for the statistical analysis of oligonucleotide arrays. *Bioinformatics* 20(4):500-506.
- Czika W, and Weir BS. 2004. Properties of the multiallelic trend test. *Biometrics* 60(1):69-74.
- Du CG, Buckler E, and Muse S. Development of a maize molecular evolutionary genomic database. *Comp Funct Genom* 4(2):246-249.
- Gibson G. 2003. Population genomics: Finding the variants of mass disruption. *Current Biology* 13(23):R901-3.
- Gibson G. 2003. Microarray analysis: Genome-scale hypothesis scanning. *PLoS Biol* 1(1):E15.
- Hasegawa M, Thorne JL, and Kishino H. 2003. Time scale of eutherian evolution estimated without a constant rate of molecular evolution. *Genes and Genetic Systems* 78:267-83.
- Haydon D, and Awadalla P. 2004. Recombination and linkage disequilibrium in the foot and mouth disease virus. *J General Virology* 85:1095-1100.
- Hill WG, and Weir BS. 2004. Moment estimation of population diversity and genetic distance from data on recessive markers. *Molecular Ecology* 13(4):895-908.
- Honeycutt E, and Gibson G. 2004. Use of regression methods to identify motifs that modulate germline transcription in *Drosophila melanogaster*. *Genetical Research* 83:177-188.
- Kishino H, Thorne JL, Seo TK, and Kajitani Y. 2003. Modeling of variable evolutionary rates to estimate divergence times and adaptive evolution. *Proceedings of the Conference on Science of Modeling: 30th Anniversary of Information Criterion (AIC)*. Pp. 297-306.
- Liu KJ, Goodman M, Muse S, et al. Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. *Genetics* 165(4):2117-2128.
- Ranz JM, Namgyal K, Gibson G, and Hartl DL. 2004. Anomalies in the expression profile of interspecific hybrids of *Drosophila melanogaster* and *Drosophila simulans*. *Genome Research* 14(3): 373-9.
- Seo TK, Kishino H, and Thorne JL. 2004. Estimating absolute rates of synonymous and nonsynonymous nucleotide substitution in order to characterize natural selection and date species divergences. *Molecular Biology and Evolution* 21:1201-1213.
- Skalski GT. 2004. The diffusive spread of alleles in heterogeneous populations. *Evolution* 58(3): 470-478.
- Thon MR, Martin SL, Goff S, Wing RA, and Dean RA. 2004. BAC end sequences and a physical map reveal transposable element content and clustering patterns in the genome of *Magnaporthe grisea*. *Fungal Genetics and Biology* 41(7):657-666.
- Thorne JL, and Goldman N. 2003. Probabilistic models of protein evolution. In *Handbook of Statistical Genetics*, 2nd ed. Vol 1. pp 209-226. DJ Balding, M Bishop, and C Cannings, eds. John Wiley & Sons.
- Weir BS. 2003. Forensics. In *Handbook of Statistical Genetics*, 2nd ed. Vol 2. pp 830-852. DJ Balding, M Bishop, and C Cannings, eds. John Wiley & Sons.
- Weir BS. 2004. The 2004 Genetics Society of America Medal-Trudy F.C. Mackay. *Genetics* 166(2):647-648.

New Models of Molecular Evolution Match Fossil Record

“Estimating species divergence times is difficult. But allowing the rates of molecular evolution to vary in time is critical for obtaining reasonable date estimates.”

Stephane Aris-Brosou,
Post-Doctoral Fellow

For biologists in fields as varied as anthropology to virology, estimating the dates at which species diverged on the evolutionary tree is an important task, but one fraught with uncertainties. One thing that is certain, though, is that estimates based on a molecular clock — the assumption that changes in DNA occur at a constant rate over time — are likely to be wrong.

“Estimates of species divergence times that are based on the molecular clock hypothesis contradict current evidence from the fossil record,” said Dr. Stephane Aris-Brosou, a post-doctoral fellow at the BRC whose research is in statistical models of molecular evolution.

For example, the available fossil record indicates that multicellular animals (metazoans) diverged from the single-celled animals about 575 million years ago, at the end of the Precambrian era. But molecular stud-

ies based on a molecular clock yield varying divergence dates that are all older than 700 million years ago.

Aris-Brosou and his colleague Dr. Ziheng Yang of the University College London (U.K.) have shown that allowing molecular evolutionary rates to vary according to two different models of episodic evolution yielded metazoan divergence times of roughly 582 million years ago. They analyzed 22 genes and all gave similar divergence time estimates that are consistent with the fossil record. Their work was published last December in *Molecular Biology and Evolution* (20[12]:1947-54).

The date that metazoans emerged will probably remain controversial because of uncertainties in the models and in the fossil record. “Estimating species divergence times is difficult. But allowing the rates of molecular evolution to vary in time is critical for obtaining reasonable date estimates,” Aris-Brosou said.

Symposium Honors Bruce Weir

Symposium speakers (below, left to right): Spencer Muse, NC State; Rebecca Doerge, Purdue Univ.; Jeff Thorne, NC State; John Storey, Univ. of Washington; William Hill, Univ. of Edinburgh; Norm Kaplan, NIEHS; John Buckleton, ESR Forensic, New Zealand; Bruce Weir; Paul Lewis, Univ. of Connecticut; Cathy Laurie, consultant in genetics; Brandon Gaut, Univ. of California, Irvine; Michael Clegg, Univ. of California, Riverside; Dmitri Zaykin, GlaxoSmithKline; Zhao-Bang Zeng, NC State.

More than 120 U.S. and international scientists gathered on June 5 at NC State for a symposium on statistical and population genetics held in honor of Dr. Bruce Weir, William Neal Reynolds Professor of Statistics and Genetics and Director of the Bioinformatics Research Center at NC State.

Organized by Weir’s former students and colleagues and sponsored by the College of Physical and Mathematical Sciences, the College of Agriculture and Life Sciences, and the Bioinformatics Research Center, the event was an occasion to honor Weir’s continuing contributions in research, education, and public service and to celebrate the enduring friendships generated within his wide circle of associates.

Eleven internationally recognized genomic scientists from the U.S., Scotland, and New Zealand presented their recent work in gene mapping of plants and hu-

mans, evolutionary genomics, forensic DNA analysis, and statistical interpretation of genomic data.

“I am happy and humbled,” Weir said of the symposium. “The level of science at the talks was very high, cutting edge even. So people who came to the symposium really learned something.”

“Bruce’s vision and initiative set him apart from other research scientists,” said symposium moderator Jeff Thorne. Weir established the graduate program in bioinformatics, the Summer Institute in Statistical Genetics, the C. Clark Cockerham lecture series, and the Bioinformatics Research Center. “These remarkable achievements are indicative of his initiative, dedication, and loyalty to the NC State community,” Thorne said.

Weir earned his Ph.D. in statistics at NC State in 1968 and has been on the faculty since 1976. In 2003, he received the O. Max Gardner Award, the University of North Carolina system’s highest faculty honor.

After dinner at the North Carolina Museum of Natural Sciences, William Hill, Professor Emeritus at the University of Edinburgh, spoke of his first encounters with Weir’s labyrinthine statistical equations and the long and fruitful association they have since enjoyed. Spencer Muse of NC State hailed Weir as a pre-eminent educator of current and future generations of genomic scientists.

“Bruce educates with a kindness that I believe is seldom encountered,” said Eden Martin, an investigator at the Duke University Center for Human Genetics. “His constant support really fosters students’ confidence as researchers,” she said.



Geneticist Explores Evolution of Malarial Parasite, Other Species

New Faculty Profile

“We’re interested in describing ... how genetic recombination and natural selection contribute to the evolution of pathogenicity, and, ultimately, which evolutionary factors contribute to speciation.”

**Philip Awadalla,
BRC and Dept. of Genetics**

Dr. Philip Awadalla, who joined the genetics faculty last January, explores the evolution of pathogens, plants, and humans using computer-based models of population genetics and molecular evolution.

“We’re interested in the rates at which genes mutate and recombine among different individuals and populations of the same species, how genetic recombination and natural selection contribute to the evolution of pathogenicity, and, ultimately which evolutionary factors contribute to speciation,” Awadalla said.

“We have a number of projects, but we’re generally looking for the same signals in the data,” he said. These signals shed light on the complex interactions of mutation, genetic recombination, and natural selection and their effects on the evolutionary history of a variety of species, including the parasite that causes malaria.

Awadalla has authored a review article on the evolu-

tionary genomics of pathogen recombination, which appeared in *Nature Reviews Genetics* 4 (Jan 2003).

A native of Ontario, Canada, Awadalla completed his Ph.D. at the University of Edinburgh, where he worked on the evolutionary genetics of plant mating systems. In post-doctoral work at the University of California at Davis, he developed a methodology for estimating recombination rates in genomes with complex mutational histories. He also helped identify a gene causing hybrid incompatibility in *Drosophila* and evaluated how natural selection shaped the evolution of the gene.

Like his colleague Dr. Greg Gibson, Awadalla shares time between the Genetics Department and the BRC. The BRC benefits greatly from these cooperative and fruitful arrangements.

Awadalla will be teaching a course on evolutionary genetics in Spring 2005.

Programming team contact:

Dahlia Nielsen at
dahlia@statgen.ncsu.edu.

Or visit [http://
bioinformatics.ncsu.edu/
brcwebsite/programming_
services.php](http://bioinformatics.ncsu.edu/brcwebsite/programming_services.php).

BRC Offers Programming Services

The BRC programming team can help you create data management systems, customized tools for data viewing and analysis, and web-based data display tools for your genomic research project. Projects to date include:

- Microarray data processing, data warehouse, and web-based data display tools for a \$3 million USDA-

IFAFS grant on juvenile pine wood and fiber quality.

- Data pipeline for processing rice sequence chromatogram files and tracking project progress.
- SAS code for formatting microarray data.
- Chromosome-based expression level viewer.
- Web page for discussion of hookworm publications.

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
