Christopher Barrett

cbarrett@vt.edu

PROFESSOR, *Virginia Bioinformatics Institute*
PROFESSOR, *Department of Computer Science, Virginia Tech*
DIRECTOR, *Network Dynamics and Simulation Science Laboratory at Virginia Bioinformatics Institute*

The Network Dynamics and Simulation Science Laboratory

In five years, the Network Dynamics and Simulation Science Laboratory (NDSSL) has developed the conceptual foundations, formal methods and software systems to represent and understand large co-evolving socially coupled systems. The approach emphasizes the novel use of distributed high-performance computing (HPC) systems and agent-oriented methods and the technology leverages progress in social computing. NDSSL has focused on contributions to policy and decision informatics for very large socially coupled systems. The resulting transdisciplinary and translational effort has resulted in funded programs of more than $30 million in the past five years with the National Institutes of Health, the Department of Transportation (through AECOM), the Centers for Disease Control and Prevention, the Department of Defense, the National Science Foundation and the Bill & Melinda Gates Foundation. NDSSL research centers on development of HPC synthetic information technologies in a usable platform, Simfrastructure, the design of which is driven by policy and decision informatics applications. NDSSL has continued with work on the Comprehensive National Incident Management System for the Department of Defense, developing an extensive Simfrastructure-based decision informatics system for analysts and decision makers.

NDSSL published 156 papers, gave 71 invited presentations, supported approximately 80 students, and undertook 12 analytical studies for our sponsors. A provisional patent was obtained and a patent filing made for the basic Simfrastructure synthetic information framework. International outreach and collaborations have been established in Europe, India, and Australia. NDSSL has established a presence in the National Capital Region and plays a leading role in a new Virginia Tech initiative in Policy Informatics for Complex Systems. NDSSL is pursuing new programs in wireless networks, commodity markets, computational economics, energy systems, sustainable interdependent infrastructure design and analysis, and HPC. ■

Scientific Achievements

The main achievements were:

- NDSSL fielded Simdemics (scalable to networks with 300 million agents) and Simfrastructure to support pandemic planning and response. Department of Homeland Security and Defense Threat Reduction Agency analysts recently used Simdemics, in an integrated cyber-environment, to respond to the recent H1N1 influenza outbreak. General Victor Eugene Renuart, Commander of the United States Northern Command and Commander of the North American Aerospace Defense Command, personally acknowledged the importance and significance of the work done for the Defense Threat Reduction Agency Comprehensive National Incident Management System program;
- In the area of economics of sustainable interdependent infrastructures, NDSSL has developed innovative simulation-based methodology to (i) assess the economic impact of system-wide failures and the resulting cascades in sociotechnical systems, (ii) model the architecture of emerging commodity markets, e.g. electricity and spectrum, and (iii) represent economically motivated agents that provide an in-silico laboratory to study network-based behavioral economics. Economic evidence from the group concerning school closures in response to the threat of epidemics such as H1N1 has been selected by the Community Guide Branch at the Centers for Disease Control and Prevention and will be included in their planned presentations to the United States Task Force on Community Preventive Services in June 2010;
- In the area of computational network science, NDSSL has developed new HPC-oriented rigorous techniques for the design, analysis and synthesis of large-scale multi-layered sociotechnical networks.

Keywords interaction-based computing; policy informatics; sustainable planning; public health; grid and cloud computing; service-oriented architectures; population dynamics; social, technological, information, biological networks; commodity markets; discrete dynamical systems.

Group Contributors Kofi Adasi, Ashwin Aji, Andrea Apolloni, Shankha Banerjee, Chris Barrett, Richard Beckman, Keith Bisset, Katelyn Bitley, Karthik Channakeshava, Jiangzhuo Chen, Suruchi Deodhar, Lisa Durbeck, Tridib Dutta, Stephen Eubank, Annette Feng, Xizhou Feng, Joshua Ferrier, Venkat Ganesan, Emily Gooding, Steve Harris, Dat Hoang, Lauren Hoops, Fei Huang, Rahul Kanna Narayanan Jayaraman, Maleq Khan, Sanjay Kishore, Chris Kuhlman, Nagarajan Kuppuswami, Jonathan Leidig, Bryan Lewis, Peng Lu, Yifei Ma, Achla Marathe, Madhav Marathe, Sharon Matchen, Gabriel Mateescu, Khristi Moore, Henning Mortveit, Kunal Mudgal, Kalyani Nagaraj, Ganesh Narayanaswamy, Elaine Nsoesie, Ashwin Palani, Zhengzheng Pan, Nidhi Parikh, Guanhong Pei, Tom Scogland, Paula Stretz, Samarth Swarup, Claudia Taylor, Vinh To, Vivek Venugopal, Anil Vullikanti, Sandra Wagener, Katherine Wendelsdorf, Ginger Williams, Mary Williams, Huadong Xia, Jae-Seung Yeom, Zhao Zhao.

Josep Bassaganya-Riera

jbassaga@vt.edu

ASSOCIATE PROFESSOR, *Virginia Bioinformatics Institute*

Nutritional Immunology and Molecular Nutrition

The Nutritional Immunology and Molecular Nutrition research group leads research programs related to the role of inflammation and immunity in human health. The central integrative theme of our research program is understanding the mechanisms of immune modulation that underlie various immune-mediated, infectious and chronic diseases, and developing novel therapeutic approaches for modulating inflammatory responses. In the reporting period, we strengthened our intellectual property portfolio in non-communicable chronic inflammatory diseases, improved our understanding of the immunoregulatory mechanisms in the gastrointestinal mucosa, and characterized the mechanisms of action of novel bioactive constituents.

We have developed novel therapeutic approaches for decreasing the infiltration of macrophages and T cells into the atherosclerotic plaque thereby decreasing inflammatory lesions and preventing hypertension; characterized the synergistic effects between abscisic acid and the diabetes drug rosiglitazone, a popular anti-diabetic drug, in the prevention of macrophage infiltration and insulin resistance; and discovered novel applications for abscisic acid in the treatment of inflammatory bowel disease.

In the gastrointestinal health area, we have dissected the cell specificity of peroxisome proliferator-activated receptor (PPAR) γ in experimental inflammatory bowel disease and demonstrated that adequate expression of PPAR γ in intestinal epithelial cells is required for the regulation of mucosal immune responses and prevention of experimental inflammatory bowel disease, possibly via modulation of lysosomal and antigen presentation pathways. Additionally, we have continued to develop a clinical trial in human patients with Crohn's disease to determine the tolerability and efficacy of a lipid immune modulator. We have also demonstrated that conjugated linoleic acid ameliorates colitis and prevents colorectal cancer in part through a PPAR γ-dependent mechanism.

In collaboration with the Network Dynamics and Simulation Science Laboratory at VBI, we have developed a preliminary model of gut inflammation. In this model (*EN*teric *I*mmunity *SI*mulator, *ENISI*), individual immune cells make contact and interact with dynamic populations of bacteria and cytokines as they migrate within and among three tissue sites: (i) the lumen/lamina propria border, where epithelial cells reside, (ii) the lamina propria, more generally termed the effector site of the mucosal immune system, and (iii) the mesenteric lymph node, the inductive site of the immune system. At the respiratory mucosa, we have developed models of infection with influenza viruses and *Francisella tularensis* and determined the role of PPAR γ in the regulation of respiratory immune responses. These efforts have received support from the National Institutes of Health, Cognis Nutrition and Health GmbH, VBI and the European Commission. ■

Scientific Achievements

The main achievement was:

- We have discovered an atypical set of host-interactive genes (*cagA* and *vacA*) in an Amerindian *Helicobacter pylori* strain (V225d). *H. pylori* is the dominant member of the gastric microbiota, and has been associated with increased risk of gastric cancer and peptic ulcers. The unusual V225d *cag* architecture was fully functional via conserved elements, but the natural deletion of 13 *cag* pathogenicity island genes and the truncation of CagA impaired its ability to induce inflammation. The atypical nature of CagA in *H. pylori* strains from Africa and Latin America Amerindians may contribute to the low incidence of *H. pylori*-associated gastric cancer.

Keywords nutritional immunology; multiscale modeling of immunity; molecular modeling; infectious diseases; immunotherapeutics; peroxisome proliferator-activated receptors; virtual screening; influenza-related inflammation; inflammatory bowel disease; colon cancer; *Helicobacter pylori*; diabetes; atherosclerosis; obesity-related inflammation.

Group Contributors Josep Bassaganya-Riera, Adria Carbo, Montse Climent, Nick Evans, Amir J. Guri, Raquel Hontecillas, William T. Horne, Stephanie N. Lewis, Graciela Lopez, Pinyi Lu, Saroj K. Mohapatra, Jeff Skoneczka, Rong Song, Cristina Vives.

Allan Dickerman

dickerman@vt.edu

ASSISTANT PROFESSOR, *Virginia Bioinformatics Institute*

Genome-scale Phylogenomics and Comparative Genomics

The Phylogenomics Research Group applies phylogenetics and comparative genomics to large-scale biological data sets to advance our understanding of underlying processes. The phylogenetic principle holds that patterns in the genetic compositions of organisms, when compared across taxa, are produced by the tree-like historical pattern of species diversification from common ancestry. While there are very interesting exceptions, this rule is useful in predicting functions of genes across species and it is an important integrative principle in modern biology.

We have pushed bacterial phylogenomics analysis to new levels in our recent phylogeny of the gamma-proteobacteria, which includes more than 100 genomes and alignments of 350 to more than 1200 homologous single-copy proteins for various taxon subsets. This large volume of data was necessary to resolve some regions of this tree. Although we have provided the best evolutionary tree yet of this group, some relationships are unresolved and require further work.

Collaborations with other groups enabled us to make new discoveries in important eukaryotic groups. Working with Vladimir Shulaev of the Virginia Bioinformatics Institute and the international strawberry genome consortium, we provided a robust phylogeny placing the strawberry among the fully sequenced plant genomes. This evolutionary tree, which was based on 154 protein-coding genes, provided strong support for a topology conflicting with a recently published tree, based on chloroplast and ribosomal RNA genes, which was used to investigate the evolutionary relationships of poplar. We have also developed a mitochondrial phylogeny for all available mosquito data sets including four new mitochondrial genomes in collaboration with Dr. Jake Tu of the Department of Biochemistry, Virginia Tech.

In a project to detect protein interactions involved in wood formation in poplar, which is a collaboration with Eric Beers (Department of Horticulture, Virginia Tech) and Amy Brunner (Department of Forestry, Virginia Tech) , we are applying the comparative principle to the tangled web of protein functions in the face of complicating duplications and functional divergences. Our website (xylome.vbi.vt.edu) maintains lists of genes used in yeast two-hybrid experiments to test protein-protein interactions. The phylogenetic tree we produced suggests that model species in the legumes may provide less diverged gene comparisons.

We have developed the first thorough metagenomic dataset surveying the microbiome of the distal gut in the pig using the deep sequencing capabilities of the VBI Core Laboratory Facility. Bacterial DNA, which was isolated from fecal samples from antibiotic-free pigs from the Virginia Tech Swine Center, was purified and sequenced on the Roche GS-FLX™ genome sequencer in the Core Laboratory Facility, providing 400,000 reads (100 Mb). Interestingly, we found that a single known genome in the genus *Prevotella* seemed to dominate the population in contrast to our expectation of extreme diversity. ■

Scientific Achievements

The main achievements in phylogenetics, metagenomics, and functional analysis were:

- We published several papers on phylogenetics and two more are in preparation. The gamma-proteobacteria analysis was important for, among other things, it shows how the current taxonomy of this group needs to be re-organized. We found a mysterious ambiguity in the base of the enterobacterial group in the tree (*Escherichia coli* and relatives) where both halves of the genes favored distinct topologies at two different steps;
- The plant phylogenetic findings mentioned previously provide a novel picture of evolutionary relationships at odds with chloroplast-based phylogenies;
- Our recent results in metagenomic analysis provide a novel approach to discover the mobile component of the metagenome – that part which is actively moving between genomes.

Keywords phylogenomics; vertical inheritance; horizontal transfer; microbial diversity; human microbiome; environmental microbes; plant pathogen; host-microbe interaction; protein-protein interaction networks; systems biology; biomass production.

Group Contributors Eric Nordberg, Elena Shulaeva, Kelly Williams.

Harold R. "Skip" Garner

garner@vbi.vt.edu

PROFESSOR, *Virginia Bioinformatics Institute*
PROFESSOR, *Department of Biological Sciences, Virginia Tech*
PROFESSOR, *Department of Computer Science, Virginia Tech*
PROFESSOR, *Department of Basic Science, Virginia Tech Carilion School of Medicine*

Medical Informatics: From Microsatellite Genomics to Text Mining and Ethics

The Medical Informatics research group develops technologies, tools, databases, and methods for biomedical research to make discoveries in genetics, genomics, medicine, and new research areas such as ethics. The work integrates bioinformatics and wet lab research and proceeds in two directions. Computational research allows scientists to initiate new projects, generate new hypotheses, and pursue promising research areas that involve the use of clinical samples. Bioinformatics permits follow-on analysis and interpretation of laboratory data to form new knowledge.

We have focused on two primary research areas that complement the resources, research environment, and new collaborations with research faculty and clinical researchers at Virginia Tech: (1) The role of microsatellites (repetitive DNA sequences) as causative agents for phenotype, especially disease; (2) The development and exploitation of data mining techniques, especially text mining and analysis and interpretation of '-omic' data.

Microsatellites are repetitive DNA sequences, for example ATATATAT or CAGCAGCAG, which exist in about one million places in our genome. Microsatellite nucleotides are 10,000-times more likely to undergo mutation than nucleotides that exhibit single nucleotide polymorphisms, which explains the use of microsatellites in forensics and as paternity markers. Although they are known to cause many diseases and conditions, for example Fragile-X and Huntington's Disease, microsatellites are understudied largely because of the lack of methods to measure them collectively. We have overcome this limitation by developing and using new array-based and deep sequence-based techniques. This approach has enabled the group to develop genomic distance metrics, build phylogenetic trees that contribute to the 'tree of life', identify and characterize new genome-wide instability mechanisms linked to cancer, especially breast cancer, and produce new genomic predisposition markers that are ready for clinical testing.

Researchers depend heavily on accessing the scientific literature to find publication references, experimental techniques, and other information, and develop new ideas. Searching the literature is mainly done using keywords, but we have developed a text similarity approach as an alternative. Text similarity analysis at the sentence, paragraph or whole-paper level is provided as a free service using our tool, eTBLAST (etblast.org), and it is being expanded to encompass more collections of scientific and non-scientific literature. New post-processors are being created to extract more value out of the collection of similar findings. The eTBLAST similarity tool has been used to identify and measure departures from accepted publication ethics, e.g. duplicate publications and potentially plagiarized publications. We plan to continue this work to create educational materials that target different student and professional levels and which will raise awareness of quantitative publication ethics. ■

Scientific Achievements

The main achievements were:

- We pioneered the study of quantitative publication ethics by the thorough application of computer-initiated, human-verified text similarity analysis of the biomedical literature. We now provide on-line text-searching tools to intercept plagiarized material during review and catalog high-similarity text documents that have already appeared in peer-reviewed journals (etblast.org and the déjà vu database);
- We developed technologies and methods that for the first time enable the collective analysis of genomic microsatellite content. This approach has (1) identified new biomarkers that are predictive of susceptibility to breast, colon, prostate and other cancers, (2) indicated the existence of a new genomic destabilization mechanism, and (3) provided potential new targets for therapeutic development.

Keywords genetics; genomics; cancer; neurological disorders; microsatellites; deep genotyping; biomarker discovery; target discovery; data analysis and interpretation; text similarity searching; text data mining; hypothesis generation; data unification.
Group Contributors Cristi Galindo, Tara Long, Lauren McIver, Zhaohui Sun, Hongseok Tae, David Trusty.

Ina Hoeschele

inah@vbi.vt.edu

PROFESSOR, *Virginia Bioinformatics Institute*
PROFESSOR, *Department of Statistics, Virginia Tech*
ADJUNCT PROFESSOR, *Department of Cancer Biology, Biostatistics Core Member, Comprehensive Cancer Center, Wake Forest University School of Medicine, Winston-Salem, NC*
ADJUNCT PROFESSOR, *Virginia Tech Carilion School of Medicine*

The Statistical Genetics Group

Highly multivariate Quantitative Trait Locus (QTL) mapping in genetical systems biology. The goal of this project (NIH 1R01HG005254-01) is to implement, evaluate and compare several types of multivariate statistical analyses for integrative analysis of two or more high-dimensional groups of variables (e.g., genome-wide DNA variant genotypes, expression and methylation profiles).

SysGenSIM: A tool for the simulation of genetical systems biology data. In collaboration with Dr. de la Fuente, CRS4 Bioinformatica, Italy, we are creating a tool to simulate DNA variant data on inbred line crosses and human association studies, network topologies for several topology models and networks with thousands of nodes, and steady-state expression and (disease) phenotype data using (non)linear dynamical equations. A National Institutes of Health application is pending. We used our simulation tool to provide a challenge for the 5th international DREAM (Dialogue for Reverse Engineering and Methods) competition in 2010.

Inferring gene regulatory networks in genetical systems biology. In collaboration with Dr. de la Fuente, we are investigating an approximate method using local structural models and false discovery rate control.

Nonparametric Bayesian method for QTL mapping with epistasis. This project is led by Dr. Fei Zou, Department of Biostatistics, The University of North Carolina at Chapel Hill, in collaboration with the Hoeschele group at VBI and the Yi group in the Department of Biostatistics at the University of Alabama. We developed a statistical method capable of identifying complex disease genes that may act through any order of interaction with other genes or environment. A National Institutes of Health application is pending.

Epigenome-wide association study of DNA methylation and expression in the Multi-Ethnic Study of Atherosclerosis (MESA). In this National Institutes of Health-funded project (1R01HL101250-01; Principal Investigator: Yongmei Liu, Wake Forest University School of Medicine; subcontract Principal Investigator: Hoeschele), we perform integrative association analyses of genome-wide DNA methylation, gene expression and single nucleotide polymorphism genotypes, and atherosclerosis phenotypes.

Genome-wide association analysis in the Health, Aging and Body Composition study using copy number polymorphisms (CNPs). This project identifies CNPs and performs association analysis. The data were provided by our collaborator, Dr. Yongmei Liu, Wake Forest University School of Medicine. We have identified ~2500 CNPs in each race (black and white). We are performing association analyses using an extended quantitative genetic model for the genotypes at all single nucleotide polymorphisms within a CNP and accounting for genotype uncertainty.

Breast cancer biomarker discovery. In this project, we assisted the Lazar group at VBI in implementing a statistical method to find protein changes in breast cancer cells treated with hormones or cancer drugs. ■

Scientific Achievements

The main achievements were:

- We have implemented a global structural equation modeling method for inferring gene regulatory and causal networks in genetical systems biology, which is computationally feasible for hundreds of nodes, and an approximate method using local structural models that is applicable to thousands of nodes.
- We have collaborated on the development of a nonparametric Bayesian method for QTL mapping in inbred line crosses and human association studies that is capable of identifying a QTL interacting with any other genes or environment in any order.

Keywords quantitative, statistical and human genetics; linkage and association mapping; quantitative trait loci; epistasis; copy number polymorphisms; statistical design and analysis of systems genetics and microarray gene expression experiments; gene and causal network inference; Structural Equation Modeling; parametric and nonparametric Bayesian methods.
Group Contributors Charles Weeks, Yan Ling, Hui Li, Xu Yang, Elaine Nsoesie.
Principal Collaborators Alberto de la Fuente, Yongmei Liu, Fei Zou, Guimin Gao, Iuliana Lazar.

Reinhard Laubenbacher

reinhard@vbi.vt.edu

PROFESSOR, *Virginia Bioinformatics Institute*
PROFESSOR, *Department of Mathematics, Virginia Tech*

Computational Systems Biology

The Laubenbacher research group works in the area of mathematical systems biology, comprising theoretical and applied projects. The goal of the theoretical part of the work is the development of computational tools for modeling and simulation of complex biological networks. This includes the development of network inference methods, the development of optimal control methods for agent-based models of complex systems, and the understanding of the relationship between the structure of dynamical systems and the resulting dynamics, that is their "genotype-phenotype" relationship. The group focuses on discrete models that are broadly applicable to systems biology. The research is funded by grants from the National Science Foundation and the United States Army Research Laboratory's Army Research Office.

The applied part of the research in the group focuses on cancer biology, primarily the development of computational methods for the comparison of mass spectrometry data obtained from both cell culture and plasma samples. The goal is to obtain reliable metabolic markers for breast cancer.

We are also working on a systems biology approach to understand the role of intracellular iron metabolism in breast cancer pathogenesis. There is ample evidence that iron metabolism in malignant cells is substantially altered, but there is no systematic understanding of those changes and how they arise in the transformation to malignancy. Creating a network of the iron metabolism process allows researchers to investigate changes in the network under different conditions. The goal is to build on this network to create predictive models of iron metabolism in normal and cancer cells, which will provide more information about key nodal points in iron metabolism and how they become different during malignant change. Leveraging this knowledge could be used to advance therapeutic or diagnostic research.

In a third project, the group is looking at the clinical uses of systems biology techniques. Before the functional differences between a cancer cell and a normal cell can be understood, an assessment of the overall biochemical network, not just the individual molecular mechanisms involved, is needed. Moving the use of systems biology techniques from the laboratory to the clinic could result in the development of improved diagnostic tools and treatment options, as well as potential new drug targets to help combat many potentially fatal types of cancer. The work on cancer systems biology involves collaborations with colleagues at the Virginia Bioinformatics Institute and the Wake Forest University Comprehensive Cancer Center. ■

Scientific Achievements

The main achievements were:

- Polynomial algebra of discrete models in systems biology: Discrete models of molecular networks are popular because they allow the construction of mathematical models without needing detailed information such as rate constants of biochemical reactions. We have shown that the framework of polynomial dynamical systems is a general framework that can incorporate commonly used modeling approaches such as Rene Thomas' logical models as well as Petri nets;
- We have developed a web-based software package for modeling and simulation of molecular networks using discrete models, including algorithms for network inference from time series data;
- We have developed a new mathematical method to analyze metabolomics data, which assigns a global signature to a sample, and which avoids the need for peak comparison.

Keywords computational systems biology; network inference; cancer biology; modeling; simulation; dynamical systems; mathematics; optimal control for complex systems; cancer metabolomics; bioinformatics.

Group Contributors Greg Blekherman, Franziska Hinkelmann, Abdul Jarrah, Reinhard Laubenbacher, David Murragarra, Shamira Shalom, Alan Veliz-Cuba.

Chris Lawrence

lawrence@vbi.vt.edu

ASSOCIATE PROFESSOR, *Virginia Bioinformatics Institute*
ASSOCIATE PROFESSOR, *Department of Biological Sciences, Virginia Tech*

Fungal PathoSystems Biology Laboratory

The Fungal PathoSystems Biology Laboratory works to further understanding of the interactions between airborne fungi and humans and plants at the biochemical and molecular levels. We primarily study fungal species found within the genera Alternaria and Aspergillus, which are some of the most ubiquitous, airborne moulds found in the environment and are clinically linked to human airway disorders such as allergies, severe allergic asthma, various forms of sinusitis, and deadly invasive diseases such as aspergillosis. These genera also include many important plant pathogenic species. This year we have made significant progress in all areas of research including our efforts to sequence and annotate additional Alternaria genomes, increase our knowledge of the interaction of fungi and fungal products with the mammalian and plant immune systems, and strengthen our expanding intellectual property portfolio in the area of therapeutic approaches for treatment of fungal-associated disorders of humans and plants.

We lead the Alternaria Genome Consortium, directing the efforts to sequence and annotate the world's first Alternaria genome, *A. brassicicola*, which infects a large number of crop plants found within the Brassicaceae (crucifer) plant family. This year we initiated the sequencing of approximately 20 additional Alternaria species genomes. We have continued to assess gene function in relation to pathogenicity in *A. brassicicola* using functional genomics approaches and have studied the effects of over 200 putative virulence factors mined from the *A. brassicicola* genome sequence. In collaboration with the Michigan State University-Department of Energy Plant Research Laboratory, we discovered the polyketide synthase gene cluster responsible for the biosynthesis of the small molecule histone deacetylase (HDAC) inhibitor depudecin and found this molecule was a minor virulence factor of plants. Working with the fungal pathogenesis laboratory at Montana State, we also discovered the novel gene tmpL and found that the corresponding protein is critical for intracellular redox homeostasis, oxidative stress tolerance, and infection of plants.

We have continued our collaborative efforts with Mayo Clinic, Allergic Diseases Laboratory, Rochester, MN, and initiated new collaborations with fungal researchers at the National Aspergillus Center, Department of Translational Medicine, University of Manchester, United Kingdom. We have formed a partnership with AlerGenetica SL, a European startup company focused on the development of novel therapeutics for treatment of mold allergies. We have discovered that Alternaria possess potent Th2 adjuvant activity and secrete proteases that interact with the G-protein-coupled receptor PAR-2, which triggers the release of proinflammtory molecules from epithelial cells. Finally, we have developed new mouse models of Alternaria-induced airway inflammation in order to screen potential therapeutics for inhibition of allergic disease phenotypes. Collectively this research is funded by the National Science Foundation (NSF), the Mayo Foundation for Medical Education and Research, and the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH). ■

Scientific Achievements

The main achievements were:

- Discovery and characterization of the TmpL protein unique to fungi and its importance in intracellular redox homeostasis and virulence in both plant and animal models of invasive disease;
- Completion of the *A. alternata* and *A. tenuissima* genomes using next-generation sequencing technologies, the initiation of sequencing a myriad of genomes from additional Alternaria species and the further development of a genome annotation database and visualization systems;
- Contribution to effector research in collaboration with Brett Tyler's research group at VBI by demonstrating that fungal and oomycete effectors bind to surface phosphatidylinositol 3-phosphate and enter human airway cells.

Keywords allergy; asthma; plant pathogens; fungal pathogens; virulence proteins; effector proteins; necrotroph; Alternaria; Aspergillus.

Group Contributors Mihaela Babiceanu, Ha X. Dang, Zhuoxin Jiang, Kwang-Hyung Kim, Prathyusha Kolconda, Sang-Wook Park, Amanda Cronin Rumore.

Collaborators Robert A. Cramer (Montana State University), Hirohito Kita (Mayo Clinic, Rochester, MN), Brett Tyler (VBI), Biswarup Mukhopadhyay (VBI, Virginia Tech), Andy Pereira (VBI, Virginia Tech), Paul Bowyer (University of Manchester, United Kingdom), Jonathan Walton (Michigan State University), Thomas Mitchell (Ohio State University), Barry Pryor (University of Arizona), Steve Rounsley (University of Arizona), Tobin Peever (Washington State University), Gillian Turgeon (Cornell University), Liwu Li (Biological Sciences, Virginia Tech), Thierry Rouxel (INRA, France), Richard Oliver (University of Murdoch, Australia), Reinhard Fischer (KIT, Karhlsruher, Germany), Yong-Hwan Lee (Seoul National University, South Korea).

Iuliana Lazar

lazar@vbi.vt.edu

ASSOCIATE PROFESSOR, *Virginia Bioinformatics Institute*
ASSOCIATE PROFESSOR, *Department of Biological Sciences, Virginia Tech*

Cancer Proteomics and Biomarker Discovery Laboratory

The research goal in the Lazar laboratory is to identify the proteins and pathways that are responsible for driving aberrant cells into division even when molecular checkpoints such as the G1/S restriction point are in place. The efforts have focused so far on profiling the G1 and S phases of the cell cycle of MCF7, an estrogen receptor positive mammary cancer cell line that serves as a valuable *in vitro* model for the study of breast cancer. By using mass spectrometry detection, vast lists of proteins that match pathways involved in cell proliferation and cell cycle regulation were generated.

The analysis of three biological replicates of MCF7 cells in the G1 and S phases of the cell cycle identified more than 2,500 proteins with a reproducibility of ~70 % relative to each replicate. Preliminary protein interaction network analysis with STRING has revealed that ~136 cell cycle pathways could be matched by nuclear-enriched MCF7 extracts. Other cancer-relevant pathways were involved in angiogenesis, metastasis and apoptosis. Future comparisons between MCF7 (cancer) and MCF10 (normal) cells will help identify cause-effect associations involved in fundamental mechanisms used by cancer cells to evade the G1/S restriction point and continue through the cell cycle.

Large-scale differential mapping of proteins that are essential to breast cancer cell-cycle progression will empower researchers with a broad range of versatile bioanalysis tools that will: (a) help develop a system-level model of the molecular mechanisms involved in cell cycle regulation and signaling networks; (b) provide novel insights into understanding and identifying fundamental causes of cancer cell proliferation; (c) provide a valuable source of information to support basic and clinical studies and facilitate the development of therapeutic agents that target cell-cycle regulatory components; (d) prove to be a priceless source of reference data that can be used as a control for various biological explorations involving targeted analysis of altered cell states; (e) ultimately, represent an ample contribution to the development of breast cancer protein databases and significantly enhance our capacity to intervene in disease detection, prevention and therapy.

The long-term objective of the research is focused on implementing this proteomic technology on disposable microfluidic platforms for reliable detection of protein co-expression patterns. Such microfluidic chips, used in tandem with electrospray ionization mass spectrometry detection, typically enable the detection of 50–100 proteins from 0.1–1 μg of complex protein extracts. Disposability, cost-effective analysis with minimal carry-over, and capability to detect protein co-expression patterns, will launch such devices as viable bioanalytical platforms for large-scale population screening applications.

Research in the Lazar group is funded by: A National Science Foundation Career award (Principal Investigator) entitled "Development of Microfluidic Platforms with MS Detection for Proteomic Applications"; a National Institutes of Health Exploratory/Developmental Research Grant Award (R21) (Principal Investigator) entitled "Microfluidic MALDI-MS Device for High-Throughput Proteomics and Biomarker Discovery"; an Avon award (Co-Principal Investigator) entitled "Environmental Risk of Breast Cancer Development: Molecular Basis for Prevention". The following patent is pending: Microfluidic Devices and Methods Facilitating High-Throughput, On-Chip Detection of Separation Techniques. ■

Scientific Achievements

The main achievements were:

- Finalized a project aimed at developing mass spectrometry methods that enable the identification of a large number of proteins and candidate cancer biomarkers in complex cellular extracts;
- Evaluated the impact of peptide modifications on the accuracy of the iTRAQ (isobaric tags for relative and absolute quantitation) method;
- Finalized the analysis of MCF7 breast cancer cells at G1/S arrest in the cell cycle.

Keywords proteomics; mass spectrometry; microfluidics; cancer; biomarkers.
Group Contributors Jarod Kabulski, Xiuli Mao, Milagros Perez, Debby Reed, Yang Xu.

Pedro Mendes

mendes@vbi.vt.edu

PROFESSOR, *Virginia Bioinformatics Institute*
PROFESSOR, *School of Computer Science, The University of Manchester, United Kingdom*
ADJUNCT PROFESSOR, *Department of Biochemistry, Virginia Tech*
ADJUNCT PROFESSOR, *Department of Cancer Biology, Wake Forest University Medical School*

Biochemical Networks Modeling Group

The Mendes research group models biochemical networks, an activity that is at the core of systems biology and which is pursued in two areas in the group: (i) developing software for modeling and simulation, and (ii) combining data analysis methods and modeling to infer networks from experimental data. The research is funded by the National Institutes of Health.

The software activity continued with development and maintenance of COPASI (Complex Pathway Simulator) and related outreach activities. In the reporting period, the focus was on developing algorithms for combining discrete events with differential equation modeling methods simultaneously. These types of hybrid simulations are technically challenging but are needed for a certain set of models, such as the cell cycle models used in Dr. John Tyson's research group in the Department of Biological Sciences, Virginia Tech. They also found a new application for discrete events, which is to use them for monitoring and recording specific data of a time course simulation that can be further used to calculate high-level summaries of time series. This is now being explored to provide further functionality in the software.

The group is actively engaged in the definition of data standards for systems biology. Dr. Hoops was one of five elected members to edit the Systems Biology Markup Language (SBML). This activity includes collecting specifications from the systems biology research community and to frame these specifications in a coherent and scientifically valid framework that will later be adopted by various software applications (including their own). Dr. Mendes, in collaboration with colleagues in the University of Manchester, United Kingdom, developed a new standard for principled representation of systems biology data. While SBML has been very successful in allowing the exchange of models, there was no standard to communicate data associated with, for example, a systems biology model until now. The new standard, Systems Biology Results Markup Language (SBRML) allows one to associate data to a model. The data can have originated either from experiments or from simulations; the standard provides appropriate references to the techniques used to obtain the data, thus providing all the necessary metadata required for their interpretation.

Reverse engineering is an activity that infers a biochemical network directly from measurements of variables of that network. It is a very active area of systems biology where the Biochemical Networks Modeling Group has a history of research. Recently we have been investigating the use of Fourier transforms of time series data and have found that this reveals an association between molecules and pathways.

The group continues to collaborate on modeling methodologies and the software COPASI with colleagues at the University of Heidelberg, Germany, and the University of Manchester, United Kingdom, and on metabolomics data analysis and the modeling of iron metabolism with colleagues at the Virginia Bioinformatics Institute and Wake Forest University. ■

Scientific Achievements

The main achievements were:

- COPASI is now one of the main simulators used in systems biology in a diversity of applications such as drug design, cancer modeling, neurochemistry, immunology, metabolism, biotechnology, and many others. The software is also widely used in education. The software was downloaded more than 9,000 times in this reporting period;
- An accurate method for combining continuous and discrete event aspects in models has been developed, allowing for models to contain discontinuities and opening up a new set of modeling applications;
- The group contributed substantially to the specification of SBML Level 3 and the definition of a new standard for systems biology data (SBRML).

Keywords computational modeling; simulation; systems biology; dynamical systems; data standards; network inference; software development.

Group Contributors Hui Cheng, Stefan Hoops, Aejaaz Kamal, Pedro Mendes, Revonda Pokrzywa.

Biswarup Mukhopadhyay

biswarup@vt.edu

ASSOCIATE PROFESSOR, *Virginia Bioinformatics Institute*
ADJUNCT ASSOCIATE PROFESSOR, *Department of Biochemistry, Virginia Tech*
ADJUNCT ASSOCIATE PROFESSOR, *Department of Biology, Virginia Tech*

The Extreme Conditions Microbiology Group

The group studies the molecular mechanisms that allow certain microorganisms to thrive under conditions that are lethal to most life forms and investigates the evolutionary paths that led to the development of such traits. The resulting information is leveraged for research on energy production and new therapeutics for tuberculosis (TB).

TB is caused by *Mycobacterium tuberculosis*. Our work on TB is focused on the hypothesis that coenzyme F_{420} helps the mycobacteria to defend themselves against the human immune system. F_{420}, a deazaflavin, is found mostly in anaerobic methanogenic archaea and only in a few bacteria including the mycobacteria. In the past year, we have shown that reduced F_{420} neutralizes nitrogen dioxide (NO_2), a potent defense tool used by human macrophages against *M. tuberculosis*. We have now found that *M. tuberculosis* utilizes certain F_{420}-dependent enzymes in building cell wall components that help the organism to counter the human immune system. Therefore these enzymes are potential targets for the development of TB therapeutics.

A National Aeronautics and Space Administration (NASA)-funded study of the metabolism of a hyperthermophilic deep-sea hydrothermal vent dwelling methanogenic archaeon, *Methanocaldococcus jannaschii*, is directed towards understanding whether sulfate reduction and methanogenesis, two of the oldest respiratory metabolisms of earth that have been thought to be incompatible, evolved and existed together.

In collaboration with Altuda Energy Corporation (San Antonio, TX), with funds and genomics resources from the United States Department of Energy (DOE), and support from the natural gas industry, we are developing pathways and microorganisms that will convert unminable coal in deep underground formations into natural gas. We have assessed the metabolic potentials of the indigenous microbes of the coal bed, a nutritionally deprived location, through metagenomic studies and have isolated some of these organisms to provide a platform for a synthetic biology effort.

The direct microbial conversion of cellulose into H_2 at high temperatures will provide an economical avenue for biofuel production. Our work is based on *Desulfurococcus fermentans*, the first known cellulolytic archaeon that was isolated from a hot spring in the Uzon Caldera of Kamchatka Peninsula, Russia. It is also the most thermophilic cellulose degrader and produces H_2 from cellulose even in the presence of H_2. Its close relatives do not degrade cellulose and are inhibited by H_2. Inhibition of fermentation by H_2 is commonly observed with the H_2 producers and is a major stumbling block for H_2 production via microbial fermentation. Hence, *D. fermentans* could harbor new tools for cellulose degradation and hydrogen production. To identify these tools, our laboratory, in collaboration with the DOE Joint Genome Institute, Russian Academy of Sciences (Dr. Elizaveta Bonch-Osmolovskaya) and other United States laboratories, is carrying out comparative computational and functional genomics investigations that involve *D. fermentans* and four of its close relatives.

A small-size phosphoenolpyruvate carboxylase (PepcA) that is widespread in the archaea and found in a few bacteria, is being studied with the goal of developing designer PepcAs that will allow higher photosynthetic CO_2 fixation efficiency in C3 plants such as potato, wheat, soybean and rice, and improve yield in microbial processes for the industrial production of organic acids and amino acids. The gene for PepcA was identified first by our laboratory. In collaboration with Dr. Pete Dunten at the Stanford Synchrotron Radiation Lightsource (SSRL), we have solved the X-ray crystallographic structure of a PepcA, which sets the stage for engineering a PepcA. ■

Scientific Achievements

The main achievements were:

- The identification of two cell wall synthesizing enzyme systems in the mycobacteria that are potential targets for the development of TB drugs;
- Obtaining the first three-dimensional structure for PepcA, which sets the stage for detailed structure-function studies on this enzyme.

Keywords coenzyme F_{420}; mycobacteria; tuberculosis; phosphoenolpyruvate carboxylase; archaea; methanogenesis and sulfate reduction; evolution of metabolism; coal bio-gasification; cellulosic hydrogen; hyperthermophiles; archaea-eukarya relationship.

Group Contributors Lakshmi Dharmarajan, Jennifer L. Downs, Eric F. Johnson, Philip Kohnke, Usha Loganathan, Lindsay C. Martin, Karla N. Piedl, Endang Purwantini, Jason Rodriguez, Bradley J. Rolfe, Matthew P. Smith, Dwi Susanti.

Jean Peccoud

peccoud@vt.edu

ASSOCIATE PROFESSOR, *Virginia Bioinformatics Institute*

Synthetic Biology Group

We made a theoretical breakthrough by proposing to use attribute grammars to represent the relations between structure and function for synthetic DNA sequences. Synthetic biology demands such a formalism and provides an ideal setting for testing hypotheses about relationships between DNA sequences and phenotypes beyond the gene-centric methods used in genetics. Attribute grammars are used in computer science to translate the text of a program source code into the computational operations it represents. By associating attributes with parts, modifying the value of these attributes using rules that describe the structure of DNA sequences, and using a multi-pass compilation process, it is possible to translate DNA sequences into molecular interaction network models. Attribute grammars represent a flexible framework connecting parts with models of biological function. They will be instrumental for building mathematical models of libraries of genetic constructs synthesized to characterize the function of genetic parts. This formalism is also expected to provide a solid foundation for the development of computer-assisted design applications for synthetic biology.

We have reported the release of the first version of GenoCAD (www.genocad.org), a web-based application to design protein expression vectors, artificial gene networks and other genetic constructs composed of multiple genetic parts. By capturing design strategies in grammatical models of DNA sequences, GenoCAD guides the user through the design process. By successively clicking on icons representing structural features or actual genetic parts, complex constructs composed of dozens of functional blocks can be designed in a matter of minutes. GenoCAD automatically derives the construct sequence from its comprehensive libraries of genetic parts. Upon completion of the design process, users can download the sequence for synthesis or further analysis. The source code of GenoCAD has been licensed through a partnership between Virginia Tech and the International Society for Computational Biology, which should facilitate the open source development of the software.

We reported the first use of GenoCAD to develop a genetic device. More specifically, we explored the application of the concept of co-design to biological engineering. Co-design is common in engineering, where it is necessary, for example, to determine the optimal partitioning between hardware and software for the implementation of the features of a system. We proposed to adapt co-design methodologies for synthetic biology. As a test case, we have designed an environmental sensing device that detects the presence of three chemicals, and returns an output only if at least two of the three chemicals are present. We showed that the logical operations can be implemented in three different design domains: (1) the transcriptional domain using synthetically designed hybrid promoters; (2) the protein domain using bimolecular fluorescence complementation; and (3) the fluorescence domain using spectral unmixing and relying on electronic processing. ■

Scientific Achievements

The main achievements were:

- The description of a compiler to translate DNA sequences into Systems Biology Markup Language (SBML) files, which was published in *PLoS Computational Biology*;
- The award of a $1.4 million grant from the National Science Foundation to develop GenoCAD;
- The licensing of GenoCAD through a partnership between Virginia Tech and the International Society for Systems Biology to facilitate the open source development of the software.

Keywords cell cycle; artificial gene networks; synthetic biology; systems biology; formal languages; compilation; imaging; microscopy.

Group Contributors Laura Adam, David Ball, Patrick Cai, Matt Lux, Julie Marchand, Sarah Zheng.

Andy Pereira

andy_pereira@vt.edu

PROFESSOR, *Virginia Bioinformatics Institute*
ADJUNCT PROFESSOR, *Department of Biological Sciences, Virginia Tech*
ADJUNCT PROFESSOR, *Department of Crop and Soil Environmental Sciences, Virginia Tech*

Systems Biology of Complex Plant Biological Processes

We have created gene interaction networks in the plant model *Arabidopsis*, integrating gene expression data, protein-protein/DNA interactions and other functional genomics data, to characterize and dissect complex biological processes at a systems biology level. These gene interaction networks can also be used to transfer gene functional information between species for gene function prediction and comparative studies. In one application in *Arabidopsis*, gene interaction networks were used to derive drought-response networks. Knockout mutants of drought-responsive transcription factors were used as genetic perturbations to the network, revealing subnetworks affected by the transcription factor mutations that characterize the molecular and physiological mechanisms involved in drought responses.

We identified the master regulator of lignin and cellulose biosynthesis using systems level analysis. Cellulose from plant biomass is the largest renewable energy resource of carbon fixed from the atmosphere, which can be converted into fermentable sugars for production into ethanol. However, the cellulose present as lignocellulosic biomass is embedded in a lignin matrix from which it needs to be extracted for efficient processing. We showed that expression of the master regulator in rice, a model for the grasses, causes a 34% increase in cellulose and a 45% reduction in lignin. We revealed the transcription factors coordinately regulating the induction of cellulose and repression of lignin biosynthesis using gene coexpression network analysis, as well as the downstream enzymes regulated by the cascade of transcription factors. The results thus support the development of non-food crop grasses and crop wastes with increased cellulose, and low lignin with good agronomic performance that would improve the economic viability of lignocellulosic crop utilization for biofuels.

Plant productivity and efficient water use under water-limited conditions are important traits for drought resistance in crops such as rice (*Oryza sativa*). From gene expression studies, we identified a drought-regulated transcription factor gene *HYR*, which on overexpression in rice exhibits enhanced shoot biomass, higher grain yield, and improved water use efficiency. The enhanced biomass is correlated with increased net photosynthesis under well-watered and drought-stressed conditions. The *HYR* rice plants also accumulate higher levels of soluble sugars (glucose, sucrose, and fructose) and maintain higher relative water content under drought stress. The increased grain yields are a function of more spikelets and larger grains, giving the *HYR* gene its name (*HIGHER YIELD RICE*). The results demonstrate the application of genetic engineering of a plant transcription factor for the improvement of plant productivity under water-limiting and normal conditions. ■

Scientific Achievements

The main achievements were:

- Systems level analysis of the cell wall synthesis pathway in rice, the model for grasses, revealed coordinate regulation of lignin and cellulose biosynthesis by expression of the master regulator SHN that controls the up-regulation of cellulose and down-regulation of lignin, providing a feedstock of high utility for biofuels;
- Analysis of a reproductive tissue-expressed rice transcription factor, HYR, showed that overexpression in rice conferred high water-use efficiency with an increase in photosynthesis, soluble sugars, biomass and grain yield;
- The *Botryococcus braunii* algal transcriptome was characterized by whole transcriptome shotgun sequencing (RNA-Seq), identifying around 30,000 genes including novel lipid biosynthesis genes.

Keywords abiotic stress; drought; network analysis; cellulose; cell wall biosynthesis; biofuels; alga; biodiesel.

Group Contributors Madana Ambavaram, Dragana Avirovik, Utlwang Batlang, Olivia Crasta, Ankit Gupta, Amal Harb, Arjun Krishnan, Abhi Loganathan, Sarah Misyak, Graciela Santopietro.

João C. Setubal

setubal@vbi.vt.edu

ASSOCIATE PROFESSOR, *Virginia Bioinformatics Institute*
ASSOCIATE PROFESSOR, *Department of Computer Science, Virginia Tech*

Microbial Genomics and Bioinformatics

The Setubal research group works primarily on computational analysis tools and cyberinfrastructure for microbial genomics and metagenomics. In the past year, we led genomics efforts that resulted in the publication of the genome of *Azotobacter vinelandii,* a free-living γ-proteobacterium aerobe species that has nitrogen-fixation capabilities and which is a widely used model for biochemistry studies, and the publication of the genomes of two strains of *Xanthomonas fuscans* subsp. *aurantifolii*, which are bacteria that cause certain kinds of citrus canker, a serious disease in citrus agriculture. For the latter, we discovered a genomic region shared by all citrus canker pathogens sequenced to date that contains two effector genes that are candidates for being major factors in the process that results in citrus canker. This was possible due to the use of several comparative genomics tools, some of which were developed in the group.

The group also contributed to the design, implementation, and publication of PAMDB, a multilocus sequence typing and analysis database and website for plant-associated microbes (effort led by Boris Vinatzer, Department of Plant Pathology, Physiology and Weed Science, Virginia Tech).

In a purely computational collaboration with Wu Feng (Department of Computer Science, Virginia Tech), we used high-performance computing to detect 380 previously unknown putative gene families in prokaryotic genome sequences available from public databases.

The group continues to work on the Red Sea metagenomics project, in collaboration with Hamza El-Dorry from the American University in Cairo. This project has collected biological samples from several locations and depths in the Red Sea, with a focus on the brine pools, regions 2200 m deep with high salinity and temperatures of about 70°C. We are currently developing computational tools to improve the phylogenetic classification of DNA sequence data obtained in the project. Another project, which has just started, focuses on bacterial transcriptomics. Next-generation sequencing technologies have enabled a revolution in gene expression studies, by allowing messenger RNAs to be sequenced at an unprecedented volume (RNA-seq). In the case of bacteria, this is opening up a new window of available data on the way bacterial cells function. We plan to apply RNA-seq to *Azotobacter vinelandii*, to gain insights into nitrogen fixation processes, and to *Pseudomonas syringae*, a tomato pathogen, to investigate genes that play a major role in the plant infection process. The analysis of such data will require the development of new computational tools, and will provide exciting new links to previously available genomic data. ■

Scientific Achievements

The main achievements were:

- Leading the effort that resulted in the publication of the genome and annotation of *Azotobacter vinelandii*;
- Leading the international effort that resulted in the publication of the genomes and annotations of two important strains of citrus canker pathogens;
- The discovery and publication of 380 previously undetected gene families in prokaryotic genomes using high-performance computing techniques.

Keywords microbial genomics; phylogenomics; metagenomics; microbial transcriptomics; cyberinfrastructure.

Group Contributors Nalvo Almeida Jr, Elaine Batista, Ulisses Dias, Andrew Warren; Kuan Yang.

Bruno Sobral

sobral@vbi.vt.edu

PROFESSOR, *Virginia Bioinformatics Institute*
PROFESSOR, *Department of Plant Pathology, Physiology and Weed Science, College of Agriculture and Life Sciences*
DIRECTOR, *Cyberinfrastructure Division, Virginia Bioinformatics Institute*

Cyberinfrastructure Division

The CyberInfrastructure Division (CID) has been pursuing integrated approaches for the analysis and determination of countermeasures for infectious diseases. The award of the Bacterial Bioinformatics Resource Center, PATRIC 2.0, and the Pathogen Portal to CID as a competitive renewal of the original PATRIC project, provides the foundation for the evolution of CID's methodologies and distributed computational infrastructure. These projects drove our change to a JBoss portal website development platform, inclusion of the RAST genome annotation system (through a subcontract to the University of Chicago and Argonne National Laboratory), and a production-level website and database with failover capability that leverages the Virginia Tech Data Center's hybrid IBM petabyte storage system. PATRIC 2.0 also involves collaboration with the National Center for Text Mining at the University of Manchester, to provide deep semantic parsing of relevant literature for integration and dissemination via the Pathogen Portal. The Pathogen Portal serves as a gateway into all the Bioinformatics Resource Centers or BRCs (viral, bacterial, eukaryotic pathogens, and vectors) and provides programmatic access to proteomics, gene expression, and other data types for the BRCs, which is coordinated through an Interoperability Working Group led by CID.

For diagnostics, an overwhelming majority of genome identification implementations rely on using signature oligonucleotide sequences, i.e. sequences that are believed to be unique to a single genome. We have developed a group testing approach that uses sequences that exist in many different genomes with several theoretical advantages. Among the potential virtues of the group testing approach are the ability to identify genomes that completely lack any unique signatures, the ability to produce a complete or even redundant classification system using fewer sequences than genomes to be detected, and the ability of a system designed for a starting group of genomes to be expanded to a larger group of genomes without any changes in the sequences employed. This last ability is perhaps the most intriguing because of its obsolescence-proof potential.

Other significant achievements in the reporting period were:

- The revelation of a single ancestral acquirement of the *rvh* (Rickettsiales *vir* homolog) Type IV secretion system (T4SS) from a non-Alphaproteobacterial origin by phylogenomic analysis of more than 30 rickettsiae genomes;
- The completion of a new phylogenetic tree for the Gammaproteobacteria;
- The first construction of a deletion mutant of ChvI in *Sinorhizobium meliloti* by CID's Molecular Genetics Laboratory. ■

Scientific and Engineering Achievements

The main achievements in science and engineering were:

- Data and other components from six information resources funded in BRC programmatic phase 1 (2004-09) and the proteomics resource (2004-09) were successfully acquired and integrated into PATRIC 2.0 and Pathogen Portal within the allotted 90 days given by the National Institutes of Health;
- Successful transition of the Proteomics Resource Centers information system to the Pathogen Portal team and its full integration with PATRIC 1.0 using database federation technologies;
- Development of a new algorithm based on group testing theory to support entirely novel approaches to diagnostic detection of bacterial pathogens of epidemiological surveillance interest.

Keywords pathogen; infectious disease; alphaproteobacteria; gammaproteobacteria; pathosystem; interoperability; distributed information systems; text mining; immunology; modeling; simulation; informatics; databases.

Group Contributors Cory Byrd, Stephen Cammer, Oswald Crasta, Isabel Da Fonseca, Oral Dalay, Timothy Driscoll, Joe Gabbard, James Gardner, Joseph Gillespie, Debby Hix, Ron Kenyon, Yared Kidane, Stephen Klagholz, Eric Nordberg, Christine Lee-Urcia, Dan Liu, Shrinivasrao Mane, Chunhong Mao, Thero Modise, Derren Rosbach, Julie Schulman, Mark Scott, Joshua Shallom, Bruce Sharp, Maulik Shukla, Nebiyu Shukur, Eric Snyder, Bruno Sobral, James Stoll, Dan Sullivan, Nirali Vaghela, Chunxia Wang, Rebecca Wattam, Rebecca Will, Kelly Williams, Tian Xue, Hyunseung Yoo, Chengdong Zhang, Yan Zhang.

Brett Tyler

bmtyler@vt.edu

PROFESSOR, *Virginia Bioinformatics Institute*
PROFESSOR, *Department of Plant Pathology, Physiology and Weed Science*

Systems Biology of Infectious Disease

The Tyler research group focuses on understanding host–microbe interactions at a systems level, with a principal focus on infection of plants by oomycete pathogens. The genetic and biochemical networks that control the respective physiologies of microbes and their hosts are closely connected via exchanged signals, nutrients, toxins, and other factors. Thus understanding the infection process requires understanding this joined network. Our approaches include experimental molecular biology, comparative and functional genomics of hosts and microbes, including genome sequencing and transcriptomics, bioinformatics, and mathematical modeling.

This year we have had four ongoing oomycete genome sequence projects — *Phytophthora sojae*, *Hyaloperonospora arabidopsidis*, *Saprolegnia parasitica* and *Phytophthora parasitica*. Several of these projects also include deep transcriptome sequencing. Our interest in these sequences is focused on identifying effector proteins, that is proteins that can enter the cytoplasm of host cells to suppress immune responses. In conjunction with genome sequencing, we have been developing Gene Ontology terms for annotating microbial genes involved in colonizing plant and animal hosts. This project, carried out as part of the six-institution Plant Associated Microbe Gene Ontology (PAMGO) consortium, has contributed nearly 900 new terms to the Gene Ontology, including remodeling a large part of the Biological Process ontology to accommodate terms for describing inter-organismal interactions. We also are developing new computational tools for analyzing gene expression data and predicting gene functions.

A major focus of our research is currently on effector proteins produced by oomycete and fungal plant pathogens. Since these proteins are secreted by pathogens and then enter the cytoplasm of host cells, they are key to understanding how the physiology of microbe and host are interlinked. Oomycete pathogens such as the soybean pathogen *P. sojae* have around 400 genes encoding potential effector proteins, based on bioinformatic predictions. One ongoing project, in collaboration with Nanjing Agricultural University, has been to systematically characterize the functions, expression patterns, and evolutionary change of these 400 genes. A second focus has been on the mechanism by which oomycete effectors enter host cells. We have discovered that an amino acid sequence motif, RXLR, conserved among oomycete effectors, is responsible for binding the effectors to phosphatidylinositol 3-phosphate on the surface of plant cells, which then enables the effectors to enter the cells via lipid-raft-mediated endocytosis. Surprisingly, the same mechanism was found in fungal and insect effectors, and the mechanism also enabled effectors to enter human cells, suggesting that effectors play a role in fungal infection of humans and animals. ■

Scientific Achievements

The main achievements were:

- Completion of a project carried out by the PAMGO consortium, which we lead, that added nearly 900 terms to the Gene Ontology for describing functions of host-associated microbes;
- Completion of the genome sequence of the *Arabidopsis* downy mildew pathogen *H. arabidopsidis*;
- Demonstration that fungal and oomycete effector proteins bind cell surface phosphatidylinositol 3-phosphate in order to enter plant and animal host cells via lipid-raft-mediated endocytosis, and identification of two methods to block this process that could lead to new disease interventions against infection in medicine and agriculture.

Keywords oomycetes; *Phytophthora*; soybean; *Arabidopsis*; plant pathogens; fungal pathogens; virulence proteins; effector proteins; cell penetrating peptides; Gene Ontology.

Group Contributors Vincenzo Antignani, Felipe Arredondo, Todd Brengel, Danielle Choi, Daolong Dou, Lee Falin, Emily Feldman, Regina Hanlon, Shelby Hughes, Shiv Kale, Kwang Hyung Kim, Sara Marsico, Trudy Torto-Alalibo, Sucheta Tripathy, Lachelle Waller, Lecong Zhou.

Collaborators Jim Beynon (Warwick University, United Kingdom), Madan Bhattacharyya (Iowa State University), Jeff Boore (Genome Project Solutions, Inc.), Robin Buell (Michigan State University), Sandra Clifton (Washington University), Daniel Capelluto (Biological Sciences, Virginia Tech), Alan Collmer (Cornell University), Candace Collmer (Wells College), Ralph Dean (North Carolina State University), Anne Dorrance (Ohio State), Daolong Dou (Nanjing Agricultural University, China), Mark Gijzen (Agriculture Canada), Jeremy Glasner (University of Wisconsin), Ina Hoeschele (VBI, Virginia Tech), Jonathan Jones (The Sainsbury Laboratory, United Kingdom), Sophien Kamoun (The Sainsbury Laboratory, United Kingdom), Chris Lawrence (VBI, Virginia Tech), John McDowell (Plant Pathology, Physiology and Weed Science, Virginia Tech), Saghai Maroof (Crop & Soil Environmental Sciences, Virginia Tech), T.M. Murali (Computer Science, Virginia Tech), Chad Nusbaum (Broad Institute), Franck Panabières (Institut National de la Recherche Agronomique, France), Jean Peccoud (VBI, Virginia Tech), Nicole Perna (University of Wisconsin, Madison), Thierry Rouxel (Institut National de la Recherche Agronomique, France), Carsten Russ (Broad Institute), Jeremy Schmutz (Hudson Alpha Institute for Biotechnology), Weixing Shan (Northwest Agriculture and Forestry University, China), Pieter van West (University of Aberdeen, United Kingdom), Yuanchao Wang (Nanjing Agricultural University, China).

John Tyson

tyson@vt.edu

COLLEGE OF SCIENCE FELLOW, *Virginia Bioinformatics Institute*
UNIVERSITY DISTINGUISHED PROFESSOR, *Department of Biological Sciences, Virginia Tech*

Simulation and Analysis of Molecular Regulatory Systems in Cell Biology

The Tyson group builds mathematical models of molecular regulatory systems that control important aspects of cell physiology, in particular cell growth and division in yeast, signal processing and programmed cell death in mammalian cell cultures, differentiation of immune cells in response to infection, and the spatial distribution of proteins during the asymmetric division process of α-proteobacteria. The group uses a variety of modeling approaches: nonlinear differential equations; Monte-Carlo simulations; and hybrid methods (continuous-discrete, deterministic-stochastic).

The models are carefully formulated to account in quantitative detail for comprehensive collections of published experimental data. The team also collaborates with several experimental groups to model unpublished data and to design and carry out experimental tests of model predictions. Of special interest in this regard are collaborations with Dr. Jean Peccoud (Virginia Bioinformatics Institute) on single-cell measurements of budding yeast cells as they progress through the cell cycle, with Professor Liwu Li (Department of Biological Sciences, Virginia Tech) on T-cell and macrophage differentiation in response to pathogen challenge, and with Dr. Robert Clarke (Lombardi Comprehensive Cancer Center, Georgetown University) on the development of resistance to endocrine therapy of breast cancer cells.

The research group is also committed to building effective computer software in support of deterministic and stochastic modeling. Dr. Tyson is assisted in this research program by other faculty at Virginia Tech: Dr. William Baumann (Bradley Department of Electrical & Computer Engineering), Kathy Chen and Jianhua Xing (Department of Biological Sciences), Yang Cao and Cliff Shaffer (Department of Computer Science), and Mark Paul (Department of Mechanical Engineering). ■

Scientific Achievements

The main achievements were:

- Constructed a stochastic model of the budding yeast cell cycle that accurately accounts for single-cell measurements of progression through G1 phase and commitment to DNA synthesis;
- Participated in the establishment of a new, National Cancer Institute-funded Center for Cancer Systems Biology, in collaboration with the breast cancer research group of Dr. Robert Clarke at the Georgetown University School of Medicine;
- Wrote a major review article on functional motifs in biochemical reaction networks for the *Annual Reviews of Physical Chemistry*.*

---

* Tyson JJ, Novak B (2010) Functional motifs in biochemical reaction networks. *Annual Review of Physical Chemistry* **61**: 219-240.

Keywords network dynamics; cell division cycle; bifurcation analysis.

Group Contributors Debashis Barik, William Baumann, Chun Chen, Kathy Chen, Yan Fu, Baris Hancioglu, Tian Hong, Sandip Kar, Teeraphan Laomettachit, Umma Juka Mobassera, Janani Ravi, Rajat Singhania, Kartik Subramaniam, Iman Tavassoly, Anael Verdugo, Jianhua Xing.