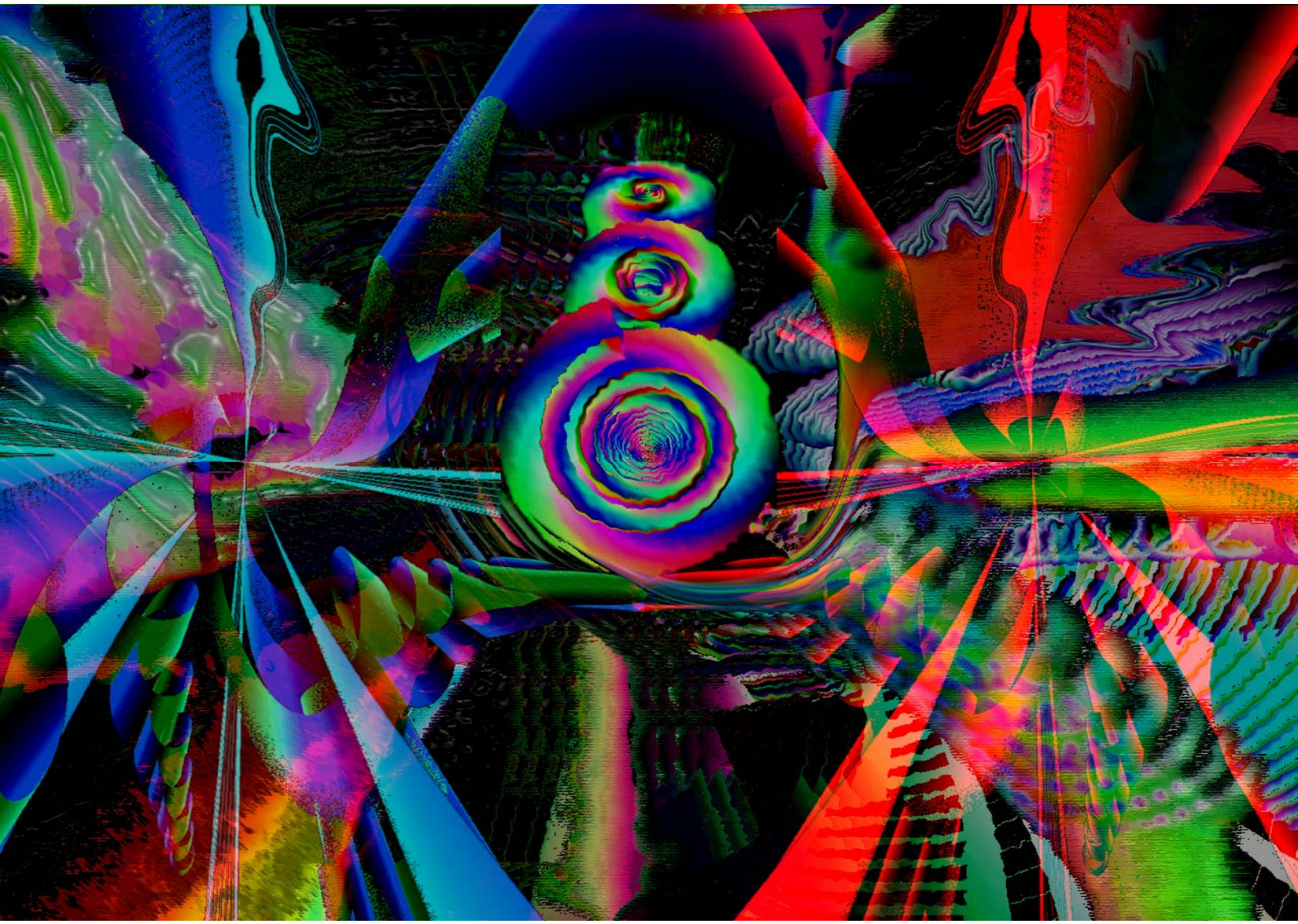
First Annual Systems Biology Workshop

The Pennsylvania State University
September 7, 2006

Sponsored by The Eberly College of Science and The Huck Institute



A Meeting of the Minds

1st Annual SYSTEMS BIOLOGY WORKSHOP

September 7th 2006, Penn Stater Conference Hotel, Room: 208

We would like to give a special thanks to Dr. Daniel Larson and Dr. Channa Reddy for making this meeting possible.

The **scope** of this workshop will be to highlight the strength of multidisciplinary approaches to Biology. Specifically we desire to demonstrate how the integration of mathematical-modeling, molecular-interaction networks, evolution, comparative genomics, bioinformatics, knowledge bases, and the physics of cellular structures can be applied to empirical experimental biology.

We expect to **learn** the types of research questions are applicable and are being pursued within quantitative bioscience groups. We also hope to **learn** what types of programs exist in this new discipline of biological sciences through interactions with the speakers which attend the workshop and how they were implemented at their respective institutions.

Program/Invited Speakers:

Session I: Chairs: Randen Patterson and Melik Demirel
8:00 Introduction by **Nina Fedoroff**
8:15 **Jayanth Banavar**, Penn State, From gene expression to genetic networks - a maximum entropy approach
9:05 **John Hogenesch**, U. Penn, Cell based screening approaches in the characterization of molecular and pathway function.
10:00 coffee break
10:15 **Harley McAdams**, Stanford, A systems perspective on the Caulobacter cell cycle control system
11:05 **Damian van Rossum**, Penn State, Towards decoding protein function from primary amino acid sequences
11:35 **Reka Albert**, Penn State, Dynamic modeling of signal transduction networks

12:05-2:00 Poster session and Lunch

Sesssion II: Chairs: Melik Demirel and Randen Patterson
2:00 **Ivet Bahar**, U. Pitt, Network Models for Protein Dynamics and Allostery
2:50 **Richard Rand**, Cornell, Mathematical models of circadian rhythms in eyes
3:45 **Arhur Lesk**, Penn State, Systems Biology of Protein Structures
4:15 Closing Remarks by **Dean Dan Larson**
5:00 Adjourn

Organizers: Randen Patterson, Biology, Melik Demirel, Engineering Science

Epochal evolution shapes the phylodynamics of influenza

Katia Koelle[1], Sarah Cobey[2], Bryan Grenfell[1], Mercedes Pascual[2]
1The Center for Infectious Disease, The Pennsylvania State University
2The University of Michigan, Ann Arbor

Influenza A H3N2 viruses are characterized genetically by their limited standing diversity and antigenically by clusters that emerge and replace each other within 2-8 years. By introducing an epidemiological model that allows for differences between influenza's genetic and antigenic properties, we show that these patterns can arise from cluster-specific immunity alone. Central to the formulation is a genotype-to-phenotype mapping, based on neutral networks, with antigenic phenotypes, not genotypes, determining the degree of strain cross-immunity. The model parsimoniously explains well known, as well as previously unremarked, features of interpandemic influenza dynamics and evolution. It captures the observed boom-and-bust pattern of viral evolution, with periods of antigenic stasis during which genetic diversity grows and episodic contraction of this diversity during cluster transitions.

VALIDATION OF PREDICTED ERYTHROID *CIS*-REGULATORY MODULES

Hao Wang, Yuepin Zhou, Yong Cheng, Ying Zhang, David C. King, James Taylor, Francesca Chiaromonte, Christine Dorman, Webb Miller, Louis C. Dore, Mitchell J. Weiss, Ross C. Hardison
Center for Comparative Genomics and Bioinformatics, Huck Institutes of Life Sciences and the Departments of Biochemistry and Molecular Biology, Biology, Statistics and Computer Science and Engineering, The Pennsylvania State University, University Park, PA and Children's Hospital of Philadelphia, Philadelphia, PA, USA

Multiple alignments of genome sequences are helpful guides to functional analysis, but predicting *cis*-regulatory modules (CRMs) accurately from such alignments remains an elusive goal. We predict CRMs for mammalian genes expressed in red blood cells by combining two properties gleaned from aligned, noncoding genome sequences: a positive regulatory potential (RP) score, which detects similarity to patterns in alignments distinctive for regulatory regions, and conservation of a binding site motif for the essential erythroid transcription factor GATA-1. Within eight target loci, we tested 75 noncoding segments by reporter gene assays in transiently transfected human K562 cells and/or after site-directed integration into murine erythroleukemia cells. Segments with a high RP score and a conserved exact match to the binding site consensus are validated at a good rate (50-100%, with rates increasing at higher RP), whereas segments with lower RP scores or nonconsensus binding motifs tend to be inactive. Active DNA segments were shown to be occupied by GATA-1 protein by chromatin immunoprecipitation, whereas sites predicted to be inactive were not occupied. We verify 4 previously known erythroid CRMs and identify 28 novel ones. Thus, high RP in combination with another feature of a CRM, such as a conserved transcription factor binding site, is a good predictor of functional CRMs. Genome-wide predictions based on RP and a large set of well-defined transcription factor binding sites are available through servers at http://www.bx.psu.edu/.

Predicting Essential Components of Signal Transduction Networks: A Dynamic Model of Guard Cell Abscisic Acid Signaling

Song Li[1], Sarah M. Assmann[1], Réka Albert[2*]
1 Biology Department, Pennsylvania State University, University Park, Pennsylvania, United States of America,
2 Physics Department, Pennsylvania State University, University Park, Pennsylvania, United States of America

Plants both lose water and take in carbon dioxide through microscopic stomatal pores, each of which is regulated by a surrounding pair of guard cells. During drought, the plant

hormone abscisic acid (ABA) inhibits stomatal opening and promotes stomatal closure, thereby promoting water conservation. Dozens of cellular components have been identified to function in ABA regulation of guard cell volume and thus in stomatal aperture, but a dynamic description is still not available for this complex process. Here we synthesize experimental results into a consistent guard cell signal transduction network for ABA-induced stomatal closure, and develop a dynamic model of this process. Our model captures the regulation of more than 40 identified network components, and accords well with previous experimental results at both the pathway and whole-cell physiological level. By simulating gene disruptions and pharmacological interventions we find that the network is robust against a significant fraction of possible perturbations. Our analysis reveals the novel predictions that the disruption of membrane depolarizability, anion efflux, actin cytoskeleton reorganization, cytosolic pH increase, the phosphatidic acid pathway, or K^+ efflux through slowly activating K^+ channels at the plasma membrane lead to the strongest reduction in ABA responsiveness. Initial experimental analysis assessing ABA-induced stomatal closure in the presence of cytosolic pH clamp imposed by the weak acid butyrate is consistent with model prediction. Simulations of stomatal response as derived from our model provide an efficient tool for the identification of candidate manipulations that have the best chance of conferring increased drought stress tolerance and for the prioritization of future wet bench analyses. Our method can be readily applied to other biological signaling networks to identify key regulatory components in systems where quantitative information is limited.

From molecules to metapopulations: systems biology of metabolic performance in the real world

James. H. Marden[1], Christopher W. Wheat[1,2], J. Cris Vera[1], Howard W. Fescemyer[1], and Ilkka Hanski[2]
(1) The Pennsylvania State University, Department of Biology, University Park, PA
(2) University of Helsinki, Department of Biological and Environmental Sciences

In the broadest sense, systems biology aims to examine how gene and protein networks affect organismal function in the real world. We are accomplishing this goal by examining how alleles, gene expression phenotypes, and alternative splice variants are associated with metabolism and flight of a butterfly that is a leading model system for population biology. Glanville fritillary butterflies in the Aland islands of Finland exist as a metapopulation in which scattered local populations frequently go extinct, but are replaced, on average, by the founding of new populations. New population founders are individual females that have dispersed away from their natal patch. Empirical studies that we began in 2004 have shown that the female offspring of new population founders have a higher flight metabolic capacity and increased fecundity compared to females in old populations (those that have persisted for > 5 years), particularly when the new populations are poorly connected within the population network. Variation in flight metabolism and fecundity are associated with allelic variation in the metabolic enzyme *phosphoglucose isomerase (Pgi)*. *Pgi* allele frequency varies with population age and connectivity, and has surprisingly strong effects on year-to-year population growth in a context-dependent manner wherein high frequency of one *Pgi* allele is favorable for population (deme) growth in small patches, whereas a high frequency of the other most common *Pgi* allele is favorable for population growth in large patches. Alternative splicing of a gene that encodes a muscle contractile protein, troponin-t is strongly associated with female fecundity and appears to interact with *Pgi* to affect muscle metabolism and flight performance. We are in the process of using the Penn State 454 sequencer to characterize the transcriptome of this species and determine how global gene expression phenotypes (using Agilent arrays) are related to *Pgi* genotype, quantitative variation in alternative splicing, and metapopulation biology. Ultimately we aim to achieve a new level of understanding of how metapopulation dynamics maintain genetic and phenotypic variation in nature. This work is supported by NSF Biocomplexity grant EF-0412651.

Density Dependence Explains Tree Species Abundance and Diversity in Tropical Forests

Igor Volkov(1,2), Jayanth R. Banavar(2), Fangliang He(3), Stephen P. Hubbell(4) and Amos Maritan(5)
1 Department of Physics, 104 Davey Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16802 USA
2 Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, University Park, PA 16802 USA
3 Department of Renewable Resources, University of Alberta,
Edmonton, AB T6G 2H1 Canada
4 Department of Plant Biology, The University of Georgia, Athens, GA
30602 USA and The Smithsonian Tropical Research Institute, Box 2072,
Balboa, Panama
5 Dipartimento di Fisica `G. Galilei', Universita di Padova, via
Marzolo 8, 35131 Padova, Italy

Explaining recurrent patterns in the commonness and rarity of species in ecological communities -relative species abundance (RSA)- has been a central goal of community ecology for more than half a century. Here we show that the framework of the current neutral theory in ecology can be easily generalized to incorporate symmetric density dependence. One can calculate precisely the strength of the rare species advantage that is needed for an explanation of a given RSA distribution. Previously, we demonstrated that a mechanism of dispersal limitation also fits RSA data well. Here we compare fits of the dispersal and density dependence mechanisms for the empirical RSA data on tree species in six New and Old World tropical forests and demonstrate that both mechanisms offer sufficient and independent explanations. We suggest that RSA data by themselves cannot be used to discriminate among these explanations of RSA patterns -- empirical studies will be required to determine whether RSA patterns are due to one or the other
mechanism, or to some combination of both.

Determination of Guard Cell-Specific Gene Regulation from Transcriptome Analysis

Sona Pandey, Song Li Zhixin Zhao, Laetitia Perfus-Barbeoch, Liza A. Wilson, Timothy E. Gookin and Sarah M. Assmann
Biology Department, Pennsylvania State University, University Park PA 16802

Stomatal complexes, consisting of a pair of multi-sensory guard cells surrounding a microscopic pore, are present in the epidermis of all aerial plant parts. By their reversible opening and closing movements stomata control gas exchange from plants. The unique location and physiology of guard cells makes them an important model system to study regulation of some of the most important plant processes including control of water loss, CO2 influx for efficient photosynthesis, interaction with a multitude of environmental signals such as light, humidity and temperature, and responses to various biotic and abiotic stresses.

Guard cells (GC) have been long studied for elucidation of signal transduction pathways using electrophysiological, biochemical and molecular genetic methods, but these methods are limited only to a small number of genes, enzymes or pathways and fall short of elucidating the complex signal transduction networks that exist in these cells. To have a complete blueprint of guard cell function and to elucidate integrated signaling networks, we are taking a genomics approach, which has been limited so far because of unavailability of sufficient quantities of purified guard cells.

We have established methods to obtain GC RNA from high quality epidermal peels with negligible contamination of any other cell types and have used these samples to hybridize whole genome Affymetrix chips (ATH1). RNA isolated from total leaves was used as a comparison, along with extensive analysis of extant microarray databases. The transcriptional data obtained from these chips was used to address the following questions:

1. Are there genes that are guard cell-specific?
2. What determines this specificity?
    a. Chromosomal structure

b. Presence of specific regulatory motifs on the promoters of these genes
c. Involvement and role of specific miRNA targets and/or sequences in guard cell specific genes
d. Regulation of specificity by natural cis-acting elements

Our analyses indicate that about 1-5% of the genes can be classified as guard cell-specific, depending on the stringency of selection criteria. *In silico* data indicate that all the above listed factors are involved in determining this guard cell-specific expression. Biological verification of these results is underway.

Rapid and asymmetric divergence of duplicate genes in the human gene coexpression network

Wen-Yu Chung1,6, Reka Albert2, Istvan Albert5, Anton Nekrutenko3,5,6 , Kateryna D. Makova4,6§
Departments of 1Computer Science and Engineering, 2Physics, 3Biochemistry and Molecular Biology, 4Biology, 5Huck Institute for Life Sciences, and 6Center for Comparative Genomics and Bioinformatics, Penn State University, University Park, PA, 16802, USA

Background: While gene duplication is known to be one of the most common mechanisms of genome evolution, the fates of genes after duplication are still being debated. In particular, it is presently unknown whether most duplicate genes preserve (or subdivide) the functions of the parental gene or acquire new functions. One aspect of gene function, that is the expression profile in gene coexpression network, has been largely unexplored for duplicate genes.
Results: Here we build a human gene coexpression network using human tissuespecific microarray data and investigate the divergence of duplicate genes in it. The topology of this network is scale-free. Interestingly, our analysis indicates that duplicate genes rapidly lose shared coexpressed partners: after approximately 50 million years since duplication, the two duplicate genes in a pair have only slightly higher number of shared partners as compared with two random singletons. We also show that duplicate gene pairs quickly acquire new coexpressed partners: the average number of partners for a duplicate gene pair is significantly greater than that for a singleton (the latter number can be used as a proxy of the number of partners for a parental singleton gene before duplication). The divergence in gene expression between two duplicates in a pair occurs asymmetrically: one gene usually has more partners than the other one. The network is resilient to both random and degree-based *in silico* removal of either singletons or duplicate genes. In contrast, the network is especially vulnerable to the removal of highly connected genes when duplicate genes and singletons are considered together.
Conclusions: Duplicate genes rapidly diverge in their expression profiles in the network and play similar role in maintaining the network robustness as compared with singletons.

Inference and analysis of gene-regulatory networks in the bacterium *B.subtilis*

Claire Christensen[1], Anshuman Gupta[2], Costas Maranas[2], Reka Albert [1]
1. Department of Physics, The Pennsylvania State University
2. Department of Chemical Engineering, The Pennsylvania State University

We present the methods and results of a two-stage modeling process that generates candidate gene-regulatory networks of the bacterium *B.subtilis* from experimentally obtained, yet mathematically underdetermined microchip array data. By employing a computational, linear correlative procedure to generate these networks, and by analyzing the networks from a graph theoretical perspective, we are able to verify the biological viability of our simulated networks, and we demonstrate that our networks' graph theoretical properties are remarkably similar to those of other, more well-studied biological systems. In addition, by introducing artificial noise into the

experimental data prior to the implementation of the linear inference process, and by comparing the graph theoretical properties of the resulting perturbed networks to those of the original networks, we are able to identify trends in graph theoretical behavior that occur both in the original networks as well as in their perturbed counterparts. These commonalities in behavior at multiple levels of complexity allow us to ascertain the level of complexity to which our process is robust to noise.

Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns

Timothy R. Lezon*, Jayanth R. Banavar*, Marek Cieplak§, Amos Maritan†‡ and Nina Fedoroff¶
*Department of Physics, 104 Davey Laboratory, and ¶Department of Biology and Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA 16802 and the Santa Fe Institute, Santa Fe NM 87501; §Institute of Physics, Polish Academy of Science, Aleja Lotnikow 32/48, 02-668 Warszawa, Poland; †Dipartimento di Fisica, Universita di Padova, Via Marzolo 8, 35131 Padova, Italy.

We devise a method based on the principle of entropy maximization to identify the gene interaction network with the highestÂ probability of giving rise to experimentally observed transcript profiles. In its simplest form, the method yields the pairwise gene interaction network, but it can also be extended to deduce higher order interactions. Analysis of microarray data from genes in SaccharomycesÂ cerevisiae chemostat cultures exhibiting energy metabolic oscillations identifies a gene interaction network that reflects the intracellular communication pathways that adjust cellular metabolic activity and cell division to the limiting nutrient conditions that trigger metabolic oscillations. The success of the present approach in extracting meaningful genetic connections suggests that the maximum entropy principle is a useful concept for understanding living systems, as it is for other complex, non-equilibrium systems.

A computational model for the formation of precise temporal sequences in nucleus HVC

Joseph K. Jun and Dezhe Z. Jin
Department of Physics, The Pennsylvania State University

During singing, projection neurons of the premotor nucleus HVC spike in temporally precise sequences. It is unknown how a neural circuit that performs such a task can develop; conventional theoretical models of learning have failed to predict a synaptic plasticity mechanism that accounts for this development. In this poster, we present a computational model aimed at describing how precise temporal sequences can form from random spiking activity. The model builds sequences by recruiting neurons into a synaptic chain using synaptic plasticity. Two kinds of synaptic plasticity are used: 1) a Hebbian mechanism which allows synapses to change strength based on activity and 2) axonal rewiring which allows synapses to deactivate and activate between neurons. Previous work focused mainly on the first kind of plasticity; we will explain why Hebbian plasticity by itself fails to produce stable temporal sequences. This will motivate the need for synaptic rewiring in the system. In particular, using Hebbian synaptic plasticity to modulate synaptic rewiring can guide neurons to make appropriate and remove inappropriate synaptic connections for the generation of temporal sequences; synaptic rewiring eliminates the inherent instabilities of using Hebbian plasticity alone. We will show that our model produces long, stable, reproducible, and temporally precise spike sequences that are consistent with the experimental data from recordings of HVC during song.

GABAergic control of adult hippocampal neurogenesis in relation to behavior indicative of chronic trait anxiety and depression states

John C. Earnheart[1], Claude Schweizer[1,5], Florence Crestani[3], Takuji Iwasato[4], Shigeyoshi Itohara[4], Hanns Mohler[3] & Bernhard Lüscher[1,2,6]

[1] Department of Biology, Department of Biochemistry and Molecular Biology and Penn State Neuroscience Institute, Penn State University, Life Sciences Building, University Park, PA, USA
[2] Department of Psychiatry, Penn State College of Medicine, Hershey, PA, USA.
[3] Institute of Pharmacology and Toxicology, University of Zurich and Swiss Federal Institute of Technology Zurich (ETH), Zurich, Switzerland
[4] Laboratory for Behavioral Genetics, RIKEN Brain Science Institute, 2-1 Hirosawa, Wako-shi, Saitama, Japan.
[5] Present address: Biologie Cellulaire de la Synapse N&P, INSERM U497, Ecole Normale Supérieure, 46 rue d'Ulm, 75005 Paris, France

Stressful experiences in early life are known risk factors for anxiety and depressive illnesses and inhibit hippocampal neurogenesis and the expression of γ-aminobutyric acid (GABA) type A receptors in adulthood. Conversely, deficits in GABAergic neurotransmission and reduced neurogenesis are implicated in the etiology of pathological anxiety and diverse mood disorders. Here we have used a conditional heterozygous gene knock out strategy in mice to show that induction of a modest deficit in γ2 subunit-containing $GABA_A$ receptors selectively in immature forebrain neurons of the embryonic and adult brain is sufficient to impair hippocampal neurogenesis and promote behavior indicative of heightened trait anxiety and depression in adults. Reduced neurogenesis is associated with normal proliferation of hippocampal cells, indicating a selective vulnerability of postmitotic neural cells to subtle deficits in $GABA_A$ receptors. By contrast, a similar forebrain-specific $GABA_A$ receptor deficit induced selectively in mature neurons during adolescence lacks neurogenic and behavioral consequences. These results suggest that reduced $GABA_A$ receptor expression in immature neurons can serve as a common molecular substrate for adult deficits in neurogenesis and for the manifestation of anxiety-driven depression.

Discovering gene regulatory network during anther development using co-expressed genes

Asela J. Wijeratne[1, 2, 3], Wei Zhang [2, 3, 6], Yujin Sun[2], Wenlei Liu[5], Dazhong Zhao[2, 3, 7],Reka Albert[4], Hong Ma [1, 2, 3]

[1] Intercollege Graduate Program in Plant Biology; [2] Department of Biology; [3] the Huck Institute for Life Sciences; [4]Department of Physics, The Pennsylvania State University, University Park, PA 16802; [5]Department of Health Evaluation Sciences Pennsylvania State College of Medicine Hershey, PA 17033; [6]Section of MCDB, Institute for Cellular and Molecular Biology, University of Texas at Austin, TX 78712; [7]Department of Biological Sciences, University of Wisconsin-Milwaukee, Milwaukee, WI 53211

In flowering plants, the anthers contain highly specialized reproductive and somatic cells required for the male gametophytes development. Molecular geneticists have uncovered a number of genes that are important for anther development. However, little information is available regarding genes and gene interactions that are responsible for the specialization of functionally distinct anther cell types. In *Arabidopsis,* two previously isolated male sterile mutants are essential for the anther cell differentiation. The *sporocyteless (spl)*/*nozzle (nzz)* is defective in the differentiation of primary sporogenous cells into microsporocytes and does not properly form anther walls. The *excess microsporocytes1* (*ems1*) /*extrasporogenous cell* (*exs*) produces excess microsporocytes at the expense of tapetum, suggesting that *EMS1* plays a vital role in controlling tapetum differentiation. To gain further insights into how microsporocytes and tapetum differentiate, and also to find those genes that controls these events, mRNA expression profile of the wild-type anther (stage 4-6) was compared with that of the two mutant anthers, *ems1/exs* and *spl/nzz* using the Affimetrix ATH1 *Arabidopsis* genome chip. A significance analysis of microarray method (SAM) was used to identify genes with statistically significant changes among the gienotypes and 1954 genes were differentially expressed in the *ems1* and or *spl* anthers compared to the wild-type anther. These genes were grouped into 14 different co-expression clusters using K mean clustering method. To understand the biological significance of these co-expression clusters, they were broadly categorized into three groups based on the expression patterns; genes that were expressed in microsporocyte, genes that were expressed in tapetum

and genes that were down stream targets of *SPL*. To obtain clues about possible co-regulation within each co-expression clusters, we searched for the enriched *cis*-regulatory motifs in the putative promoter regions of the genes. The information that was generated from the above analyses was then used to develop a model gene regulatory network for tapetum differentiation and development. This information will be useful for the understanding of molecular mechanisms that are responsible for the tapetum differentiation and development in *Arabidopsis*.

Molecular Determinants of Flight Capacity in Migratory Lepidopterans

Howard W. Fescemyer[1], Suzanne P. MacFarland[1], Mara S. Kasputis[1], Robert L. Meagher[2], Rodney N. Nagoshi[2], Shelby J. Fleischer[3] and James. H. Marden[1],
(1) The Pennsylvania State University, Department of Biology, University Park, PA
(2) USDA-ARS, CMAVE, 1700 SW 23rd Drive, Gainesville, FL
(3) The Pennsylvania State University, Department of Entomology, University Park, PA

Alternative splicing of gene transcripts is an important mechanism by which a single gene can give rise to multiple distinct proteins and, in turn, phenotypic plasticity. Relative abundance of alternative splice forms of the calcium regulatory protein, troponin t, was previously shown to enable dragonflies to adjust performance of their flight muscles in order to cope with different levels of energy available from their habitat. We characterized four alternative exons for troponin-t in flight muscle from *Helicoverpa zea* (corn earworm) and *Spodoptera frugiperda* (fall armyworm), two migratory noctuid moths whose larval stage is familiar to people who shuck sweet corn for dinner. Then, we tested the hypothesis that variation in alternative splicing of troponin t in flight muscle underlies variation in the flight capacity of these moths. Flight capacity was measured in terms of flight metabolic rate, flight behavior, mass, wing area, wing loading, and lipid content, all before quantifying their troponin t splice forms. Relative abundance of the largest alternatively spliced exons of troponin t in both moth species is associated with increased flight metabolic rate, body mass and flight propensity (a behavioral variable). Variation in flight capacity has important consequences for the way animals disperse and procreate in nature, which in turn has important consequences for population dynamics, gene flow, and the spread of agricultural pests like these moths. Our findings suggest troponin t as a molecular marker for variation in migratory capability. We are also interested in pursuing the mechanism that regulates observed variation in troponin-t splicing between individuals, and how alternative splicing of troponin t is related to expression and processing of other genes that together determine organismal phenotypes.

Systems-level regulation of immune responses to Bordetella

Juilee Thakar, Mylisa Pilione, Girish Kirimanjeswara, Eric Harvill and Réka Albert
Department of Physics, The Pennsylvania State University

The generation of particular immune functions often involves signaling pathways with multiple steps, many of which have been shown to be manipulated by pathogens. We have examined a respiratory infection model system in which disruption of host immune functions or bacterial factors changes the dynamics of the course of infection. The disrupting both bacterial factors and the specific immune function(s) they target can, in some cases, restore the normal kinetics of infection. In order to gain a system-level understanding of bacterium-host interactions, we have synthesized a network of interactions between the host immune components and bacterial factors. The network organizes complex information sets in a form such that the components and their relationships are readily accessible and can be experimentally manipulated. We incorporate experimental information on the timing of immune regulatory events into a dynamic model that is validated by comparing the effects of node disruptions to experimental mutation studies. This model is then applied to a closely related pathogen that uses a different set of virulence factors that mediate pathogenesis. Although these factors interrupt immune pathways at different steps by distinct mechanisms, they have similar effects in their disruption of key immune functions. Our results offer predictions regarding cytokine regulation, the effects of perturbations in the immune system, as well as the clearance of secondary infections. Experimental infections of convalescent hosts validated the model's predictions. This type

of modeling provides new insights into the virulence, pathogenesis and host adaptation of disease causing microorganisms and allows analysis at the system level which is not always possible by using traditional methods. The observation that a range of virulence factors affects the same signaling pathways may reflect alternative strategies used by different pathogens to attain common goals in pathogenesis.

Atomic Molecular Dynamics of Blood-Clotting Proteins

Coray M. Colina[1], Robert E. Duke[2,3], Lalith Perera[2], Tom Darden[2], and Lee Pedersen[2,3]. (1) Department of Materials Science and Engineering, Pennsylvania State University, 320 Steidle Building, University Park, PA 16802, (2) Laboratory of Quantitative and Computational Biology, National Institute of Environmental Health Science, Research Triangle Park, Raleigh, NC 27709, (3) Department of Chemistry, University of North Carolina at Chapel Hill, Campus Box 3290, Chapel Hill, NC 27599

Enzymatic cascades are often employed in biochemistry systems to achieve a rapid response. In a cascade, an initial signal institutes a series of steps, each of which is catalyzed by an enzyme. At each step, the signal is amplified. Blood clots are formed by a cascade of zymogen activations: the activated form of one clotting factor catalyses the activation of the next. This series of events is known as the blood coagulation cascade.

Our group has been working on predicting structure and dynamics of several blood coagulation proteins for almost 10 years now. With our refine methodology and the computational power available, we are now able to study relatively long molecular dynamics simulations of systems of interest as potential targets for anti-coagulant inhibitors design.

We are currently pursuing molecular dynamics studies of blood-clotting proteins in order to learn more about the behavior of these proteins and hopefully discover more about how entropy influences their function. The simulations were performed for five different systems raging from 76,000 to 150,000 atoms. The five models were subjected to aqueous-phase molecular dynamics (MD) simulations, where unconstrained dynamics were performed using the Cornell et al. force-field (using AMBER8/PMEMD8 programs) and the ff99 parameter set. The Particle Mesh Ewald (PME) method was used in all aqueous simulations to account for long-range interactions during dynamics. NPT simulations at 300 K, SHAKE algorithm for bond constraints and a 1.0 fs time step were selected. The solute, ions/counter-ions and crystal water molecules, with a total up to 150,000 atoms, were immersed in rectilinear periodic boxes of minimum 15Å each side. Several steps of energy minimization and dynamics were carried out before beginning the “production run”. Finally, the post-equilibration NPT dynamics at 300 K, or production run, was completed normally for 20 ns.

For the first time, changes in blood coagulation protein dynamics were observed within this time frame, which could help elucidate open questions in the literature. Specific examples for the dynamics of soluble tissue factor, factor VIIa and factor IXa as well as the TF-VIIa binary complex will be provided.

In each case, the simulations promoted the understanding of blood coagulation protein dynamics and in some cases helped guide in vivo experiments.

Ancient origin and rapid evolution of DANGER: a novel protein family functioning in cell differentiation

Nikolas Nikolaidis, Dimitra Chalkia, Damian B. van Rossum, Viktoriya Syrovatkina, Christine Wells, and Randen L. Patterson
Department of Biology, Pennsylvania State University

Cell differentiation is a process during which cells acquire distinct identities and specialized functions. Identification of novel proteins functioning in cell differentiation and characterization of their evolutionary history are of fundamental importance towards understanding the origin of organismal complexity and diversity. Here, we characterize a novel family of proteins (DANGER) that encode the MAB-21 domain and are involved in cellular differentiation. In addition, we present the initial functional characterization of DANGER1A, which

has neurotrophic activity. Our data show that DANGER family originated before the emergence of multicellular organisms and has been subjected to major expansion-contraction events during the metazoan evolution. DANGER proteins are evolving exceptional rapidly in contrast to the prototypic member of the family, the MAB-21 group, and other protein families involved in cell differentiation (i.e. WNT proteins). We demonstrate that in addition to amino acid substitutions, insertions and deletions of amino acids' stretches shaped the diversified repertoire of DANGER proteins. We propose that intron gain and loss together with transposition, retrotransposition and recombination events resulted in the highly divergent pattern of the MAB-21 domain sequence. Finally, we provide a scenario of the major steps during the evolution of the DANGER family.

Protein Interactions and Fluctuations in a Proteomic Network using a Network Model

Melik Demirel, Ozlem Keskin
College of Engineering, The Pennsylvania State University

A set of protein conformations are analyzed by normal mode analysis. An elastic network model is used to obtain fluctuation and cooperativity of residues with low amplitude fluctuations across different species. Slow modes that are associated with the function of proteins have common features among different protein structures. We show that the degree of flexibility of the protein is important for proteins to interact with other proteins and as the species gets more complex its proteins become more flexible. In the complex organism, higher cooperativity arises due to protein structure and connectivity.

SYSTEMS GENETIC ANALYSIS SHOWS COMMON REGULATORY QTL FOR BRAIN IRON, COPPER, AND ZINC IN BXD RECOMBINANT INBRED MICE.

Jones, L.C.[1], Beard, J.L.[1], and Jones, B.C[1].
[1]Neuroscience Graduate Program. The Pennsylvania State University, State College, PA 16801 USA.

Iron, copper, and zinc are vital for brain function yet highly toxic in excess. Thus, tight homeostatic regulation of these metals is required. Details of this regulation are still poorly described. In an effort to identify genetic regulators and coregulators of each metal, we performed QTL analyses of regional brain content of iron, copper, and zinc in 15 BXD recombinant inbred (RI) strains of mice (iron data previously published). Large interstrain variations were observed in the content of each metal. Interestingly, within-strain covariation among the three metals was also observed, suggesting a common QTL may coregulate the content of these metals. Consistent with this observation, our analyses revealed a QTL on chromosome 17 strongly associated with all three metals of interest. This QTL is also associated with ethanol acceptance, seizure susceptibility, and several measures of immune function, according to published data in the BXD phenotypes database of GeneNetwork. We report these and other phenotypic associations with this QTL, including gene expression clusters, and discuss candidate genes. This work contributes to understanding trace metal regulation as well as the relationship between iron, copper, zinc and many other aspects of behavior and disease. Future work will identify the gene(s) underlying this QTL.

This work was supported in part by USPHS grants USPHS grants NS 35088 and AG 21190 and by grants from the Restless Legs Foundation and from GlaxoSmithKline.

Gestalt Domain Detection Algorithm: A lipid binding tale

William Ivory, Kyung Dae Ko, Nikos Nikolaidis, Dimitra Chalkia, Yoojin Hong, Randen L. Patterson, Damian B. van Rossum
The Department of Biology, The Pennsylvania State University

The Gestalt Domain Detection Algorithm (GDDA) is an algorithm designed to identify highly diverged domain sequences in proteins not identified in rps-BLAST and Hidden Markov

Model (HMM) similarity searches. The innovation of the GDDA relies on the modification of the query sequence with a "seed" sequence. The algorithm modifies the original target sequence by inserting a proportion of the domain sequence at every amino acid position of the target sequence. The modified sequences are searched by rps-BLAST against a single domain database and the percentages of coverage are plotted against each amino acid position. The modification of the original sequence increases the sensitivity of rps-BLAST, since the "seed" sequences provides a "constant" initiation sequence allowing BLAST to extend (i.e. 'filling in the gaps' hence gestalt) the alignment even between highly divergent sequences. This algorithm has been successfully implemented in many biochemical studies to identify regions of functional importance. In order to test the accuracy and specificity of the GDDA we are currently testing a dataset of 124 proteins that have been experimentally shown to bind lipids yet rps-BLAST and HMM fails to identify lipid binding domains. These proteins are being searched against 38 lipid binding domains profiles. As a control we are running the same experiment using a kinase domain data set containing 37 kinase profiles. Using the GDDA on the 124 protein data set we were able to increase the coverage from 3% (4 domains identified using rpsBLAST) to 83% (103 domains identified using GDDA). In conclusion, these analyses are moving towards a functional specific scoring matrix for hitherto unrecognized functional domains.

Stochastic Processes are Key Determinants of Short-Term Evolution in Influenza A Virus

Martha I. Nelson[1], Lone Simonsen[2], Cecile Viboud[3], Mark A. Miller[3], Jill Taylor[4], Kirsten St. George[4], Sara B. Griesemer[4], Elodie Ghedin[5], Naomi A. Sengamalay[5], David J. Spiro[5], Igor Volkov[1], Bryan T. Grenfell[1,3], David J. Lipman[6], Jeffery K. Taubenberger[7] & Edward C. Holmes[1,3]*

[1]Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, University Park, PA 16802. USA.
[2]National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20818. USA.
[3]Fogarty International Center, National Institutes of Health, Bethesda, MD 20892. USA.
[4]Wadsworth Center, New York State Department of Health, Albany, NY 12201. USA.
[5]The Institute for Genomic Research, Rockville, MD 20850. USA.
[6]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894. USA.
[7]Department of Molecular Pathology, Armed Forces Institute of Pathology, Rockville, MD 20850. USA.

*Corresponding author: Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, University Park, PA 16802. USA.
Tel: 1 814 863 4689; Fax: 1 814 865 9131; e-mail: ech15@psu.edu

Understanding the evolutionary dynamics of influenza A virus is central to its surveillance and control. While immune-driven antigenic drift is a key determinant of viral evolution across epidemic seasons, the evolutionary processes shaping influenza virus diversity within seasons are less clear. Here we show with a phylogenetic analysis of 413 complete genomes of human H3N2 influenza A viruses collected between 1997 and 2005 from New York State, United States that genetic diversity is both abundant and largely generated through the seasonal importation of multiple and divergent clades of the same subtype. These clades co-circulated within New York State, allowing frequent reassortment and generating genome-wide diversity. However, positive selection and genetic diversity within and among clades were strikingly low at sites thought to be important in antigenic drift. These results indicate that adaptive evolution occurs only sporadically; rather, the stochastic process of strain migration, and subsequent reassortment, play a major role in shaping influenza virus evolution in the short-term. Thus, predicting future patterns of influenza virus evolution is inherently complex and requires intensive surveillance, whole-genome sequencing and phenotypic analysis.