BaMBA 9

Biology and Mathematics in the Bay Area
UC Davis - November 22nd, 2014

Poster Session

Modeling pathways of DNA unlinking by site-specific recombination

Robert Stolz, Reuben Brasher, and Mariel Vazquez

Mathematics, San Francisco State University

In Escherichia Coli, replication of circular chromosomes yields topologically linked DNA molecules. Topo IV, one of the type-II topoisomerases in E. coli, plays a major role in the decatenation of the newly replicated chromosomes. It has been shown that in the absence of Topo IV, site-specific recombinases XerC/D, in cooperation with the translocase FtsK, can also unlink the replication links [1,2]. The goal of this research is to explore the possible pathways of unlinking by XerCD-FtsK. We use computational methods to exhibit the recombination pathways and assign transition probabilities to each recombination step. Our results give strong support to the stepwise unlinking pathway proposed in [1,2].

Representing Migration Patterns And Constructing Phylogenetic Trees Using A Percolation Model: A Case Study Of Family Camelidae

Olga Beketova, Sandra J. Carlson, Donald Turcotte

Earth and Planetary Sciences, UC Davis

Computational methods have received increased attention as powerful tools for morphological and molecular data processing and phylogenetic tree reconstruction in paleobiology and evolutionary biology. Numerous techniques from fractal geometry have been adapted to model network distributions and reconstruct branching patterns. Percolation theory is one of the theories related to fractal geometry, which focuses on phase transitions. It is widely incorporated in studies of groundwater flow, hydraulic fracturing ("fracking"), and the spread of epidemics. Patterns associated with such models mimic the geometry of trees. In this study, we have adapted invasion percolation methods to model both phylogenetic trees and patterns of biogeographic dispersal, using the Family Camelidae as an empirical case study.

We begin by generating a null model tree, which results in a network that forms at the initial point and growth outwards based on the randomly assigned connection probabilities. The nodes within the network diagram represent species in Camelidae. The next step is to modify our model to incorporate a pattern of migration and dispersal of Camelidae out of North America, relying on the relationship between speciation and geographic isolation. Allopatric speciation is considered to be a common mode of cladogenesis requiring geographic isolation. New species are formed by the migration of populations away from the center of origin, and further divergence prevents later interbreeding with the parent population. Therefore, we modify our null model by adjusting the probability field to reflect stratigraphic and geographic fossil occurrence data, both of which serve as controlling parameters to steer the network's progression through space and time. In the end, the resulting patters from null and modified models are reconfigured as branching diagrams, and are compared to a morphological cladogram of Camelidae obtained from the literature. Given the wide variety of processes that can generate similar types of branching patterns in nature, modeling these different topologies and comparing them quantitatively can reveal the similarities and differences among the empirical and simulated patterns. Such can allow paleontologists to better understand bursts of diversification in time and space.

BaMBA 9

Biology and Mathematics in the Bay Area
UC Davis - November 22nd, 2014

Sliding filament mechanism for anaphase B; evaluation of "slide-and-cluster" versus "slide-and-flux-or-elongate" models.

Ingrid Brust-Mascher, Gul Civelekoglu-Scholey, Jonathan M. Scholey

Molecular and Cellular Biology, UC Davis

The elongation of the mitotic spindle during anaphase B contributes to chromosome segregation in many cells. Here, we quantitatively test two models of anaphase B spindle elongation using experimental data from Drosophila embryos. In the "slide-and-flux-or-elongate" mechanism kinesin-5 motors persistently slide apart antiparallel interpolar (ip) microtubules (MT). During pre-anaphase B, this outward sliding of ipMTs is balanced by the depolymerization of their minus ends at the poles producing poleward flux, while the spindle maintains a constant length. Following cyclin B degradation, ipMT depolymerization ceases so the sliding ipMTs can push the poles apart. The competing "slide-and-cluster" model proposes that MTs nucleated at the equator are slid outward by the cooperative actions of kinesin-5 plus a minus-end directed motor which then pulls the sliding MTs inward and clusters them at the poles. The results show that with a narrow range of MT dynamic instability parameters both models can reproduce the steady state length and dynamics of pre-anaphase B spindles and the rate of anaphase B spindle elongation. However, only the slide-and-flux-or-elongate model reproduces the anaphase B change in MT dynamics. This highlights the importance of acquiring and quantitatively evaluating data on different dynamic aspects of spindle behavior in order to test competing models for mitosis.

RNA-Seq Analysis of Senescent Fibroblast Cells

Robert Atwood

Buck Institute for Research on Aging

Senescent cells do not become replicative again due to the expression of cell cycle inhibitors but continue with their metabolic functions. In this study, cellular senescence is induced in fibroblast cells by exposure to high levels of ionizing radiation in vitro. Differential gene expression analysis was conducted between senescent and non-senescent fibroblast cells using RNA-Seq. Many of the genes differentially expressed in senescent cells are not related to arresting cell growth. Many of these differentially expressed genes code for secretory proteins that can change the tissue environment, which is known as the senescence-associated secretory phenotype. Accumulation of senescent cells in human tissues leads to chronic inflammation, which exacerbates age related diseases. On the other hand, cellular senescence has a beneficial effect in tumor suppression and wound healing. Cellular senescence may be an example of antagonistic pleiotrophy where it provides positive benefits earlier in life but deleterious effects later in life. The goal of this project is to better understand this complicated biological process and discover potential treatments to reduce the number and effect of senescent cells and increase a patient's healthspan.

BaMBA 9

Biology and Mathematics in the Bay Area
UC Davis - November 22nd, 2014

Towards a Better Understanding of the Unclassified Variants of the BRCA1 Gene

Santrupti Nerli, Vidya Ramprakash, Kiranmayee Dhavala, Anthony Bortolazzo, Sami Khuri

Computer Science

Recently, it has been recognized that many mutations are pathogenic because they impact the mRNA rather than the protein itself. Point mutations in the DNA or errors during transcription can activate a "cryptic splice site" in regions of the transcript that are not usually spliced. The activation of cryptic splice sites is often related to human hereditary diseases, such as familial breast cancer. Researchers are constantly aiming to improve upon existing screening methods for such deleterious alleles, but often these sequence variants can be ambiguous and difficult to screen for.

There exists a considerable amount of mutational heterogeneity in the BRCA1 gene, leading to familial breast cancer, that are of unknown biological and clinical relevance. These variants are often called unclassified variants (UCV or UV). Without an understanding of the functional implications on the mRNA and protein levels of expression, the effectiveness of cancer-risk estimation and clinical management is very challenging. Some UCV yield missense mutations, in-frame deletions or insertions, as well as aberrant pre-mRNA splicing events.

In this work, we characterize the genome-wide 5′ canonical and cryptic splice sites as well as those specific to tumor suppressor, BRCA1. Next, we will utilize these experimentally verified 5′ splice sites in the training of position weight matrices for their use in putative splice site prediction.

Protein-RNA Cross-Docking of Conformational Ensembles

Amélie Héliou, Rasmus Fonseca, Henry van den Bedem and Julie Bernauer

Computer Science department of Ecole Polytechnique (FRANCE)

Interactions between protein and RNA molecules are important in many cellular processes. The knowledge of the three dimensional structure of complexes is essential for understanding their function. Often, the conformation of a molecule in a complex (bound conformation) differs from its unbound conformation. Accounting for large conformational changes required for binding in computational docking experiments remains an enormous challenge, in particular for RNA. Here, we present a computational docking procedure that combines efficient probing of the conformational landscape with semi-rigid cross-docking.

For each partner in a protein-RNA complex, we first obtain a conformational distribution starting from the unbound conformation with an inverse-kinematics algorithm [1]. These conformations are then clustered, and cluster centers are selected for the cross-docking step. Finally, semi-rigid cross-docking experiments are performed with RosettaDock for each pair of protein-RNA cluster centers.

We show that center conformations lead to better results than the unbound conformation in the docking experiment. They provide complexes that are closed to those obtained from the bound conformation. Our procedure can significantly improve predictions of protein-RNA complexes.

1. Fonseca R, Pachov V Dimitar, Bernauer J, van den Bedem H (2014) Characterizing RNA ensembles from NMR data with kinematic models. Nucl Acids Res 42: 9562–72.

BaMBA 9

Biology and Mathematics in the Bay Area
UC Davis - November 22nd, 2014

Knots confined to tubes in the simple cubic lattice.

Maxime Pouokam

San Francisco State University

[No Abstract Provided]

Genome accessibility is widely preserved and locally modulated during mitosis

Chris Morrissey, Chris C.-S. Hsiung, Maheshi Udugama, Christopher L. Frank, Cheryl Keller, Songjoon Baek, Belinda Giardine, Greg Crawford, Myong-Hee Sung, Ross Hardison, and Gerd A. Blobel

Bioinformatics

Mitosis entails global alterations to chromosome structure and nuclear architecture, concomitant with transient silencing of transcription. How cells transmit transcriptional states through mitosis remains incompletely understood. While many nuclear factors dissociate from mitotic chromosomes, the observation that certain nuclear factors and chromatin features remain associated with individual loci during mitosis originated the hypothesis that they could provide transcriptional memory through mitosis. To obtain the first genome-wide view of the dynamics of chromatin structure during mitosis, we compared the DNase sensitivity of interphase and mitotic chromatin at two stages of cellular maturation in a rapidly dividing murine erythroblast model. Despite global chromosome condensation visible during mitosis at the microscopic level, the chromatin accessibility landscape is largely unaltered. However, mitotic chromatin accessibility is locally dynamic, with individual loci maintaining none, some, or all of their interphase accessibility. Mitotic reduction in accessibility occurs primarily within narrow, highly hypersensitive sites that frequently coincide with transcription factor binding sites, whereas broader domains of moderate accessibility tend to be more stable. In mitosis, proximal promoters generally maintain their accessibility, whereas distal regulatory elements preferentially lose accessibility. Large domains of low DNA methylation are strongly associated with accessible chromatin marking a subset of promoters during mitosis and across many cell types in interphase. Erythroid transcription factor GATA1 exerts site-specific changes in interphase accessibility that are most pronounced at distal regulatory elements, but does not visibly influence mitotic accessibility. We conclude that features of open chromatin are remarkably stable through mitosis and are modulated at the level of individual genes and regulatory elements.

BaMBA 9

Biology and Mathematics in the Bay Area
UC Davis - November 22nd, 2014

Predicting Clinical Phenotypes of Papillary Thyroid Carcinoma Patients through RNA-sequencing

Kun-Hsing Yu, Wei Wang, Chung-Yu Wang, Michael Snyder

Biomedical Informatics

Unlike most cancers, thyroid cancer has an increasing incidence rate over recent years. Understanding the molecular basis of the disease may help us stratify patients with different clinical phenotypes and provide personalized treatments. We acquired clinical and RNA-sequencing data of papillary thyroid carcinoma patients (n=484) from The Cancer Genome Atlas, divided them into distinct training and test sets, and dichotomized patients with known survival time into two outcome groups according to their median survival time (2000 days). We leveraged supervised machine learning methods to predict stages and survival outcomes, selected the top features by forward feature selection, and evaluated prediction performance through the independent test set. Using the expression levels of three genes (TERT, CCDC60, and ACADSB), our best classifier predicted patient survival outcomes with area under receiver operating characteristic curve 0.9545 on the test set. We also predicted tumor stage with test accuracy around 70%, and discovered that genes involved in extracellular matrix structure are associated with stage progression. In summary, we identified novel genomic markers indicative of the clinical phenotypes of papillary thyroid carcinoma patients. We envision that personalized medicine based on predicted disease outcomes could increase the quality of care and reduce the cost of disease management.

Investigating The Role of the Motor In DNA Packaging in Bacteriophage Viruses

Brian Cruz, Javier Arsuaga, Mariel Vazquez

Mathematics

We use computer simulation to investigate the role of the DNA packaging motor of icosahedral bacteriophages in determining the final geometry of the packaged DNA. We simulate the motor using a kinetic Monte Carlo algorithm and combine it with a dynamics simulation of the DNA molecule within the capsid. The DNA is modeled as a Worm-Like Chain with torsional forces to account for supercoiling. We show that a motor that twists the DNA as it packages it into the capsid induces a writhe bias on the DNA.

BaMBA 9

Biology and Mathematics in the Bay Area
UC Davis - November 22nd, 2014

Fast and accurate mapping of phenotypic space with X-shift

Nikolay Samusik, Zinaida Good, Matt Spitzer, Garry Nolan

M&I, Stanford Medical School

Highly parameterized mass cytometry allows routine measures of up to 45 protein markers from single cells and can provide a system-level overview of various immune processes. However, comprehensive manual analysis of cell populations in multi-dimensional data by means of sequential gating can be an intractable task. We created a new density-gradient clustering method (X-shift) that is based on k-nearest neighbour density estimate. Our simulations show that KNN-DE is able to accurately estimate the density of normal as well as long-tailed distributions in multi-dimensional space (upward of 100 dimensions). We developed a new fast and exact KNN search strategy with sub-quadratic complexity that enables computation of density estimate on extra-large datasets (several millions of data points) in reasonable time. Testing on CyTOF mouse bone marrow data shows that X-shift is able to accurately resolve hand-gated cell types better than state-of the art flow clustering methods. It is also identifies novel, previously unreported types, such as IgM+MHCII- B-cells.

Live Cell Imaging of Chromosome Dynamics During Meiosis in Saccharomyces cerevisiae Reveals Chromosomes Are Undergoing Subdiffusion

James McGehee, Brian Bui, Sean Burgess

Molecular and Cellular Biology, UC Davis

Homologous chromosome pairing, recombination, and crossover formation are important processes that occur during meiosis that help insure proper segregation of chromosomes. Improper segregation can lead to aneuploidies, which result in premature abortions and diseases such as Down syndrome in humans. To better understand how chromosomes pair, we observe chromosome motion using live cell microscopy in budding yeast. To observe the chromosomes, we have integrated repeated tetracycline operator (tetO) sequences at a chromosomal locus, which bind tetracycline repressors (tetR) fused to green fluorescent protein (GFP). Chromosome movement can be quantified by calculating the mean squared change in distance (MSCD) and by counting the number of 3D pixels sampled to get an idea of volume sampled. From our results we can infer that chromosomes undergo subdiffusion.

BaMBA 9

Biology and Mathematics in the Bay Area
UC Davis - November 22nd, 2014

A Logistic Regression Approach for Integrating Tertiary Structure and Disordered Residue Propensities as Predictors of Solvent Accessible Protein Elements

G. Bhogal, T. Kamath, T. Poelman, A. Nedunuri, J. Bruce, G. Omictin, J. Oribello, S. Bros, K. Kantardjieff, A. Gottlieb and B. Lustig

San Jose State University

Logistic regression models, including both sequence-based qualitative predictors e.g. amino-acid type, as well as quantitative predictors e.g. the Lobanov-Galzitskaya disorder propensity (LGDP) are novel approaches for predicting solvent accessible residues. Our fitted logistic models can be used to evaluate whether a particular residue is solvent accessible and available for interaction with other protein chains. We initially utilized logistic regression models with a small set of homology-based quantitative predictors, including 20-term and 6-term sequence entropy, and qualitative descriptors specifically of the query residue type. The logistic regression models are fit using dichotomous responses indicating buried or accessible. These classifications are obtained from the RSA values calculated from X-ray crystal structure. Various models were fitted using a domain-complete 1363-protein training set and then utilized in regards to test sets, including separate ones for oligomers and non-oligomers. Accuracies typically exceed 75%, where non-oligomers are associated with slightly higher accuracy. Notably, a conservation signal specific to oligomers was indicated for those models that include homology-based predictors. More recently we have begun to explore the inclusion of a simplified quantitative predictor for tertiary contact propensity. This is consistent with our approach of developing parsimonious models using physically-based descriptors, allowing large-scale searching for potential protein-protein interfaces.

Identification of Copy Number Aberrations in Breast Cancer Subtypes using Persistence Topology

Georgina Gonzalez, Javier Arsuaga

Department of Molecular and Cell Biology and Department of Mathematics UC Davis

Chromosome aberrations are a hallmark of cancer initiation and progression. DNA copy number aberrations (CNAs), such as copy number gains and losses, are of particular interest because they may harbor oncogenes or tumor suppressor genes (driver aberrations). Genomewide experimental detection of copy number aberrations across the genome is achieved through microarray and DNA sequencing technologies. However, the identification of driver CNAs remains a challenge. Supervised methods address this problem by detecting CNAs that are common and specific to a given category (such as cancer subtype) or a cancer with specific clinical characteristics. Here we propose a complementary supervised method that identifies CNAs based on the topological properties of the CGH profile. We call this method Topological Analysis of aCGH (denoted by TAaCGH).

TAaCGH focuses on the relationships between multiple genomic regions by mapping overlapping fragments of aCGH profiles into a 2D point cloud using a sliding window method. We then use the theory of computational algebraic homology to find patterns and associations within the data with $\beta 0$, the number of connected components of a simplicial. As a result, $\beta 0$ provides us with a measure of genomic instability that help us to identify aberrant regions.

We test TAaCGH by analyzing a set of 68 breast cancer patients with different subtypes that has been previously published. Our results confirm most of the regions found in the original publication.

BaMBA 9

Biology and Mathematics in the Bay Area
UC Davis - November 22nd, 2014

Characterizing Rigidity in Biomolecules with Geometric Tools

Dominik Budday, Sigrid Leyendecker, and Henry van den Bedem

Friedrich-Alexander University Erlangen-Nuremberg

Proteins perform their function by exchanging between conformational substates on a wide range of spatio-temporal scales. The value of kinematic tools to characterize these exchanges and closely related rigidity has long been recognized. Here, we model a protein as a kinematic linkage, with groups of atoms as rigid bodies and covalent, rotatable bonds as links. Hydrogen bonds are encoded as additional constraints, leading to nested, interdependent cycles that require coordinated changes of the torsional degrees of freedom. Admissible changes in the degrees of freedom lie in the tangent space TqQ to the constraint manifold Q at the current configuration q, which coincides with the nullspace of the constraint Jacobian matrix. We identify the set of rigid clusters in the protein directly and exactly from analysis of the tangent space. In contrast to methods based on combinatorial constraint counting, we obtain valid results for both generic and singular configurations. In addition, our geometric approach provides an explicit basis for motions along floppy modes, resulting in an efficient procedure to probe conformational space. Our rigidity analysis and conformational exploration can provide high-level insights into dynamic processes beyond the reach of MD simulations, with broad implications for drug design and protein engineering.

Classification of Breast Cancer Subtypes Using Signaling Pathways and Persistent Homology

Michael Steiner, Niles Baas, and Javier Arsuga

San Francisco State University

Motivation: In 2010, 206,900 women and 2,039 men in the United States were diagnosed with breast cancer, and 40,990 women and 439 men died from breast cancer [3]. Each breast cancer subtype has a different prognosis and responds differently to chemotherapy. A computational method that can quickly determine different breast cancer subtypes based on microarray data will be valuable for both cancer treatment and cancer research.

Hypothesis: We propose that different breast cancer subtypes have different signaling pathways. We will test this hypothesis by associating a point cloud with a breast cancer subtype and a signaling pathway, and compute the homology of the point cloud. Our preliminary results indicate that by our topological analysis we can distinguish between some of the breast cancer subtypes.

Methods: We used the ErbB and the PI3K-akt signaling pathway maps from the Kyoto Encyclopedia of Genes and Genomes along with microarray breast cancer data to determine if breast cancer subtypes differ topologically by associating a point cloud with a breast cancer subtype and a signaling pathway, and then computing and comparing the homology of the point clouds.

Results: We have been able to differentiate between some breast cancer subtypes. This method we have developed to define the homology of a network is novel, although our work is an extension of Arsuaga et al. (2012).

BaMBA 9

Biology and Mathematics in the Bay Area
UC Davis - November 22nd, 2014

Improving Rosetta's Energy Scoring Function for Membrane Protein Prediction

Diane Nguyen, Vladimir Yarov-Yarovoy

Department of Physiology and Membrane Biology

Alpha helical transmembrane proteins (TMP) are crucial for cell signaling and ion transport. Understanding their structures can help us deduce their functionality and lead to novel methods of drug design and specific target binding. The Rosetta Membrane method (RMM) is an adaptation of the Rosetta algorithm, a computational approach designed to predict the three-dimensional structure of a protein using de novo and homology modeling approaches. The RMM works to model nature's design of amino acid packing in membrane proteins by using statistically derived data and an energy function from known membrane protein structures. My aim is to improve the accuracy of de novo protein prediction by refining the environment, pair and density terms in the energy scoring function of the RMM. I compiled a dataset of TMP from the Protein Data Bank of Transmembrane Proteins (PDBTM). Rosetta scores each protein and finds the position where the protein will be most favorable, at its lowest energy score. My findings suggest trends in amino acid frequencies in the environment and pairwise interactions. My next step is to do benchmark runs on known protein structures to gauge how accurately the RMM can predict protein structures in relation to their native structure.