43rd MAINE BIOLOGICAL AND MEDICAL SCIENCES SYMPOSIUM
hosted by
MDI Biological Laboratory
with support from the
Maine IDeA Network of Biomedical Research Excellence (INBRE)

MDI Biological Laboratory Conference Center
Salisbury Cove, Maine
April 29-30, 2016

Poster session A

Dahlgren Hall: Genetics, Bioinformatics, and Software tools

Listed alphabetically by presenting author

Arsenic induces changes in neurological gene expression in developing zebrafish
Alsaady, T.[1], Knowles, S.[1], Kuun S.[1], Roderka M.[1], Soohey R.[1], Hutchison K.[1,2], King B.[3]
[1] Honors College, University of Maine, Orono, ME 04469 [2] Molecular and Biomedical Sciences, University of Maine, Orono, ME 04469 [3] Mount Desert Island Biological Laboratory, Salisbury Cove, ME 04672
thoalfakar.alsaady@maine.edu, sarah.knowles@maine.edu, sierra.kuun@maine.edu, meredith.roderka@maine.edu, robert.soohey@maine.edu

Arsenic in drinking water has been associated with a variety of diseases including defects in the development of the central and peripheral nervous systems, cancer, and heart and vascular diseases. In Maine, the USGS estimates 16% of private wells are estimated to have concentrations of arsenic higher than the federal standard of 10ppb and up 23% of the wells within coastal Maine. To study the effect of arsenic at the molecular level in a model organism we used a non-biased approach to look at gene expression in zebrafish exposed to arsenic at 2ppb and 10ppb. We found that 69 of the differentially expressed genes had human homologs related to 81 neurological diseases. These included Parkinson's, diseases of the peripheral nervous system, and intellectual disabilities. This study has provided us with a short list of genes that need to be further investigated regarding their involvement in neurological diseases.

Modeling binding affinity of the multiple zinc finger protein PRDM9
Arat, S.[1], Fine, A.[1,2], Walker, M.[1], Billings, T.[1], Paigen, K.[1], Petkov, P.M.[1], Carter, G.W.[1,2]
[1]The Jackson Laboratory, Bar Harbor, ME [2]Tufts University, Boston, MA
Seda.Arat@jax.org

Mammalian genomes contain over 700 zinc finger proteins, but most have unknown functions. One example, PRDM9, recognizes and binds DNA to mediate the process of meiotic homologous recombination. To identify all binding sites of PRDM9, we used a novel, *in vitro* assay called Affinity-Seq and found over 39,000 significant binding sites for the PRDM9^{Dom2} allele in mouse

DNA. Quantification of the binding frequency at each sequence enabled estimation of binding affinity at each site in addition to standard nucleotide frequencies. To gauge the contribution of each nucleotide, we built a linear regression model that includes additive and interactive effects on the binding preference of PRDM9. The model provided insight into core nucleotides and nucleotide-nucleotide interactions that can drive binding. Our work yields a detailed view of the targeting mechanism of PRDM9 in meiosis and can be broadly applied to any zinc finger protein.

Anecdata: A Platform for Crowdsourced Data

Duncan Bailey and Jane Disney
MDI Biological Laboratory, Salisbury Cove, ME
dbailey@mdibl.org

The Internet has opened new opportunities for citizen science, but existing citizen science websites are limited in the scope of the data they can collect. Anecdata.org, developed at MDI Biological Laboratory's Community Environmental Health Laboratory, is a new web app that makes it easy to get started crowdsourcing a wide range of biotic and abiotic data. Users can include their device's GPS coordinates, upload photos, as well as run advanced searches and download custom reports in spreadsheet, Google Earth, and ArcGIS shapefile format. Potential uses range from species mapping to managing water quality monitoring data. In addition to offering these services to the environmental community, the Anecdata software stack can be used for custom data collection solutions or the collection of confidential data. These uses may include researchers collecting data for health research studies. These could include demographic information and tracking of exposures, self-reporting of symptoms, follow-up treatments and exposure outcomes.

Computational identification of candidate regulatory modules for PDGFRβ

Congdon, C.B.[1,2,3], Lessard, C.R.[2,3], Karolak, M.J.[2], Oxburgh, L.H.[2]
[1]Bowdoin College, Brunswick, ME [2]Maine Medical Center Research Institute, Scarborough, ME
[3]University of Southern Maine, Portland, ME
congdon@bowdoin.edu

In this project, we use the computational tool GAMI (Genetic Algorithms for Motif Inference) to identify candidate regulatory regions for the PDGFRβ gene, which is expressed in the mesangium of the kidney and is of particular interest due to its role in renal scarring. To better understand the regulation of PDGFRβ in the kidney, candidate cis-regulatory modules (CRMs) for PDGFRβ were discovered using GAMI, a *de novo* motif inference system that predicts candidate regulatory regions by first identifying putative conserved elements. GAMI is able to search multiple long sequence lengths yielding hundreds of putative conserved motifs. Motifs are then assembled into putative CRMs. However, putative conservation is not sufficient to ensure function. Furthermore,

activity of conserved sequences is dependent on biological context. In order to identify context-dependent regulatory elements, we have incorporated tissue-specific epigenetic data from the ENCODE project into our workflow. This allowed for the discovery of tissue-specific regulatory elements.

DNA barcoding for medical Cannabis chemotype

Doty, J.[1], Horrocks, J.[1], Lackeyram-Owen, S.[1]
[1]Department of Natural Sciences, University of Maine at Farmington, Farmington, ME
jeandoty@maine.edu

Identifying conserved sequences in Cannabaceae provides a basis for molecular testing, especially for medical Cannabis. Cannabis chemotypes differ in their cannabinoid profile: Chemotype I has a THC/CBD ratio $>>1.0$, Chemotype II has a THC/CBD ratio close to 1.0, and Chemotype III has a THC/CBD ratio $<<1.0$. These chemotypes have different medicinal properties, making early growth stage testing for Cannabis chemotype desirable to both caregivers and patients. Our goal is to identify conserved sequences in the THC A and CBD A Synthase gene for use in DNA barcoding of medical Cannabis chemotypes, which will lead to more effective treatment for medical Cannabis patients.

Multiplexed SERS imaging in biological systems using biocompatible Raman active nanostars

Grant, J.[1], Allgeyer, E.[2], Browne, M.[3], Pongan, A.[3], Mason, M.[1,3]
[1]Graduate School of Biomedical Sciences, The University of Maine, Orono, ME 04469, USA
[2]Department of Physics and Astronomy, The University of Maine, Orono, ME 04469, USA
[3]Department of Chemical and Biological Engineering, The University of Maine, Orono, ME 04469, USA
jeremy.grant@umit.maine.edu

The objective of this study is to design a nanoparticle labeling system suited for enhanced Raman micro-spectroscopic imaging to directly probe the concentration, chemical dynamics, and spatial distribution of individual Raman active nanoprobes at nanometer length scales in biological systems. Surface-enhanced Raman spectroscopy (SERS), using specifically engineered spherical metallic nanoprobes, has received considerable attention recently. Considered a powerful alternative to fluorescence labeling, it offers several advantages over traditional methods, including increased photostability, narrower emission peaks, and ease of bioconjugation. Most existing nanoparticle geometries require mean diameters on the order of 50-80 nm to provide single particle signals comparable to those observed via competitive fluorescence methods. We propose to investigate a new nanoparticle geometry based on a star-shaped architecture which shows further enhancement over spherical particles. We further propose to create 10 spectrally distinct Raman active nanostars, overcoming the channel limitations encountered using standard fluorescence labeling methods.

Towards understanding host range and expanding the diversity of mycobacteriophage

Illingworth, E. , Lesko, D. , Zwirner, C. , Molloy, S.
University of Maine, Department of Molecular and Biomedical Sciences, Orono, ME.
emily.illingworth@maine.edu

Mycobacteriophage (phage) are viruses that infect mycobacteria. Phage are classified into clusters, A-X, based on nucleotide sequence similarity. Though nearly 7,000 phage have been isolated and characterized, all but 10 were isolated from rapidly growing, non-pathogenic *Mycobacterium smegmatis,* limiting what we know about phage host range and diversity. This research aims to determine the most favorable growth conditions for identifying phage infection in alternative mycobacterial species *M. salmoniphilum* and various strains of *M. chelonae,* both of which are pathogenic isolates from fish. The research also aims to suggest a protocol for phage isolation optimized for these untried species. By performing susceptibility tests on each bacterial species, we have established a better picture of the host range of phage representing nearly every cluster. Future research will assess whether phages that can infect these species also have the ability to form lysogens in them.

Modeling Genetic Interactions Associated with Subtypes of Breast Cancer

Bo Ji [1], Anna L. Tyler[1], Guruprasad Ananda[2], Gregory W. Carter[1]
[1]The Jackson Laboratory, 600 Main St, Bar Harbor, ME, USA [2]The Jackson Laboratory for Genomic Medicine, 10 Discovery Drive, Farmington, CT, USA
Bo.ji@jax.org

Characterization of mRNA subtypes in breast cancer identifies processes that drive distinct molecular subtypes and allows inference of candidates for subtype-specific drug targeting. However, such therapies have limited efficacy, often due to unpredicted compensation in the network of mutations. We performed a multi-trait genetic interaction analysis of copy-number variation and gene expression from breast cancer samples in The Cancer Genome Atlas. Modules of co-expressed genes were derived and assessed for biological function and genetic association with somatic mutations in oncogenes and tumor suppressors. Module phenotypes were simultaneously analyzed to infer direct genetic effects and effects mediated by genetic interactions. We identified interacting mutations that combinatorially associate with distinct modules and subtypes. Our work demonstrates how integrative genetic and genomic analysis can be used to generate more precise hypotheses that may be used to prioritize therapeutic targets for robust tumor suppression.

Utilization of a sensitized genetic screen to identify modifiers of retinal dysplasia in *Nr2e3*[rd7/rd7] mice

 Kong, Y.[1, 2], Charette, J.[2], Hicks, W.[2], Naggert, J.[2], Zhao, L.[2], Nishina, P.[1, 2]
[1] The Graduate School of Biomedical Science and Engineering, University of Maine, Orono, ME, [2] The Jackson Laboratory, 600 Main Street, Bar Harbor, ME
yang.kong@maine.edu

The retinal dysplasia is usually a concomitant manifestation with degenerative changes. We aimed to identify and characterize genetic modifiers capable of altering the retinal dysplasia in *Nr2e3*[rd7] mutants, a model for human Enhanced S-Cone Syndrome (ESCS).

Specifically, *Nr2e3*[rd7] mice were chemically mutagenized and screened. *Tvrm222* mouse line was teased out for its altered pan retinal fundus spotting phenotype, a characteristic of *Nr2e3*[rd7/rd7] mice. Mapping analyses further identified *Frmd4b*[Tvrm222] allele as the potential modifier gene. Apart from fundus imaging, examinations on histological sections showed the formation of pseudorosettes in retina is remarkably diminished, and this biological effect might be achieved by FRMD4B-modulated integrity of retinal external limiting membrane.

In conclusion, retinal dysplasia in *Nr2e3*[rd7] mouse can be phenotypically altered by *Frmd4b*[Tvrm222]. It provides novel insights into the pathogenesis of retinal dysplasia in ESCS, and may become potential interfering target for clinical applications against this eye disease.

The Role of Neutrophil Cytosolic Factor 1 (ncf1) in Innate Immune Response to Influenza Virus Infection in Zebrafish

Longfellow, JR[1], Algeo, LD[1], Sullivan, C[1], Millard, PJ[2, 3], Kim, CH[1, 3]
[1]Department of Molecular and Biomedical Sciences, University of Maine, Orono, ME 04469, USA, [2]Department of Chemical and Biological Engineering, University of Maine, Orono, ME, 04469, USA, [3]Graduate School of Biomedical Sciences and Engineering, University of Maine, Orono, ME, 04469, USA
Jacob.longfellow@maine.edu

The Kim laboratory has recently developed a zebrafish model for human influenza infection, allowing direct examination of gene interactions and regulation of the host immune response to viral infection (Gabor et al., 2014). Neutrophil Cytosolic Factor 1 (*ncf1*) is a component of NADPH oxidase pathway, which is responsible for the production of reactive oxygen species (ROS). ROS are responsible for the destruction of invading pathogens like influenza. Previous literature has shown ncf1, in knock-down studies, to be partially responsible for lung damage in mice (Peiris et al., 2009). By employing the zebrafish as a model organism, further characterization of *ncf1* in response to influenza infection would be possible.

The role of dysregulated microRNAs during Influenza A virus infection

Miller, C.[1] , Millard, P.[3] , Kim, C.[1,2]
[1]Department of Molecular & Biomedical Sciences, University of Maine, Orono, ME 04469
[2]Graduate School of Biomedical Science & Engineering, University of Maine, Orono, ME 04469
[3]Department of Chemical and Biological Engineering, University of Maine, Orono, ME 04469
campbell.miller@umit.maine.edu

Small biomolecules known as micro-RNAs regulate gene expression in host cells. Their influence on cellular processes is far-reaching, as it has been estimated that microRNAs regulate more than 60% of all human protein-coding genes (Friedman et al., 2009). Not surprisingly, their dysregulation can cause a number of physiological problems. Amazingly, RNA viruses have been shown to intentionally dysregulate host microRNAs in an attempt to promote viral replication and survival.   One RNA virus, Influenza A virus (IAV), is of particular importance as it is a constant global health threat. Previous studies have shown that IAV causes the differential expression of host microRNAs in the infected mouse lung (Li et al., 2010). Using a zebrafish model for human IAV infection (Gabor et al., 2014), the dysregulation of two microRNAs, miR-223 and miR-21, will be investigated. Due to their roles in immunity, IAV may be utilizing these microRNAs to evade the immune system.

Cannabinoid profiling by Gas Chromatography/Mass Spectrometry

Morocco, T[1]., Briana, T.
[1]Department of Natural Sciences, University of Maine at Farmington, Farmington, ME
terrym@maine.edu

Cannabinoid compounds (*Cannabis* sp.) are known to have medicinal properties. Medicinal use requires more precise knowledge of cannabinoid profiles to allow for proper dosage, matched to the particular medical condition. To this end, we're developing a trimethylsilyl derivative method to analyze THC, CBD, THCA, CBDA, in whole cannabis or preparations, using Gas Chromatography/Mass Spectrometry (GC/MS).   Better analytical techniques for cannabinoid profiling will help improve medical Cannabis treatment.

Terpene profiles in Cannabaceae

Morocco, T[1]., Johnson, R.
[1]Department of Natural Sciences, University of Maine at Farmington, Farmington, ME
terrym@maine.edu

Terpene and terpenoid compounds are thought to work in concert with cannabinoid compounds (*Cannabis* sp.) resulting in medicinal qualities.   Medicinal use requires more precise knowledge of terpene profiles to allow for proper dosage, matched to the particular medical condition. To

this end, we have developed a Gas Chromatographic-Flame Ionization detection method for terpene profiling in whole cannabis or preparations using the Full Evaporation Technique (FET-GC-FID). Better analytical techniques for terpene profiling will help improve medical Cannabis treatment.

Population Genetic Diversity of Two Lichen Species on Rock Glacier Slopes in Debouille Public Reserved Land in Northern Maine

Corey, B. [1], Rochester, J. [2], Feinstein, L. [1], Nelson, P. [2], Roe, J.L. [1]
[1]University of Maine at Presque Isle, Presque Isle, ME 04769, [2]University of Maine at Ft. Kent, Ft. Kent, ME 04743
judith.roe@maine.edu

Several talus slopes in the Debouille Public Reserved Lands in northern Maine are rock glaciers which maintain subterranean ice year round. Unique plant communities inhabit these slopes. Lichen species were collected along elevation transects on each glacier to test the hypothesis of whether there was more genetic variation along the slope within a rock glacier or between rock glaciers. DNA was extracted from individuals from two species, *Cladonia stygia* and *Cladonia stellaris*. The ITS region was amplified from the fungal partner by PCR and sequenced. Both species exhibited high population genetic diversity and several haplotypes were identified in each, including length polymorphisms in the 18S region. Individuals with the same haplotype were found on two glaciers and at varying heights within a glacier suggesting that gene flow is occurring both within and between rock glaciers.

Congenital Muscular Dystrophy with Megaconial Myopathy (MDCMC): When big isn't better.

Sayed AA[1,2]., Cox GA [2], Sher RB[1].
[1]University of Maine, Orono, Maine. [2]Jackson Laboratory, Bar Harbor, Maine.

Congenital muscular dystrophies (CMDs) include over 30 types of muscle disorders, all having an onset at or just after birth. Most CMDs are autosomal recessive disorders with rapidly progressing symptoms that have clinical manifestations resulting in muscle weakness and resulting delayed/arrested motor abilities, respiratory, cardiac illnesses and delayed speech development. We plan to concentrate on Congenital Muscular Dystrophy with Megaconial Myopathy (MDCMC), a rare type of CMD, in which patients manifest muscular dystrophy in a rostral to caudal gradient with enlarged mitochondria localized at the cellular periphery and severe cognitive impairments. MDCMC has been shown to arise as a result of mutation in the Choline Kinase beta (*CHKB*) gene that leads to defective CHKB enzyme production. In this project we will be characterizing a mouse model for MDCMC as well as an engineered rescue mouse and will be testing if the rescues are a good model for cognitive impairment.

A newly discovered intact prophage in the *Mycobacterium chelonae* genome provides an opportunity to study host/prophage co-evolution

Sewell, E. , Hutchison, K.
The Honors College and Dept. of Molecular and Biomedical Sciences, The University of Maine
erica.sewell@maine.edu

Mycobacteriophage (phage), are viruses that infect the mycobacteria. All bacteria are infected by phage, and each bacterial species has unique phage that infect them, making phage prime candidates for studying viral diversity and evolution. Some phage integrate into the host genome upon infection (prophage), where they potentially may remain indefinitely, coevolving with the host, providing growth factors and other benefits to the host. The purpose of my research is to characterize a putative prophage within the genome of bacterial host *Mycobacterium chelonae* Bergey. The integrase cassette within the *M. chelonae* prophage is intact and molecular modeling suggests that the integrase protein would be functional if it is in fact expressed. The prophage would, therefore, be capable of excision and integration into other bacterial genomes. A putative repressor has also been identified. Experiments are underway to determine if these prophage genes are expressed.

Genotyping CYP2C19 and CFTR as a Model for Treatment and Prevention of Diseases

Thibault, E.[1] Averill, R.[1], Foley, J.[1], Gibula, A.[1], Wakeland, L.[1], Sato, D.[2], Stanton, B.[3]
[1]University of Maine Honors College, [2]Mount Desert Island Biological Laboratories, [3]Dartmouth Medical School
thibaulte14@gmail.com

Genotyping to determine the genetic variation from person to person can be useful in detecting diseases, genetic predispositions, and drug efficacies. There are many forms of genetic variation, including single nucleotide polymorphisms (SNPs) and mutations. SNPs occur roughly once every 1000 base pairs (bp). This study examines SNPs in exons 4 and 5 of gene CYP2C19 that inhibit metabolism of clopidogrel (Plavix), a blood thinner, and mutations in exon 10 of cystic fibrosis transmembrane conductance regulator (CFTR) that can cause and increase severity of cystic fibrosis (CF). These mutable loci were examined in fifteen anonymous individuals. Metabolism of Plavix was inhibited in one individual, partially inhibited in six, and uninhibited in four. No individuals examined are carriers of CF.

Undergraduate first year discovery through analysis of mycobacteriophage host range and the novel mycobacteriophage Greg

Waite, J. Bergdoll, A. Brooks, R. Doyle, J. Hashmi, F. Hernandez, R. Labbe , D. Lamoureaux, A. Lesko, D. Lilley, T. Lutes, J. Marcotte, A. Milashouskas, T, Monahan, B. Overturf, K. Perley,

A. Pollard, C. Riportella, M. Shoneye, M. Struba, A. Tanner, C. van der Schaaf, J. Vergara, C. Zwirner, C.
University of Maine Honors College, Orono, ME
jasmine.waite@maine.edu

Mycobacteriophage (phage) are viruses which infect bacteria of the genus *Mycobacterium*, including *M. smegmatis* and pathogenic *M. tuberculosis*. Phage research provides insight into the evolution of bacteriophage and their bacterial hosts. To better understand the diversity of mycobacteriophage, the 2015 HON/BMB150 class used non-pathogenic *M. smegmatis* as a host to isolate 22 novel phage from soil. The host range of the 22 phage was determined in *M. smegmatis* and in five alternative strains/species of *Mycobacterium* that are pathogenic in fish. One phage, Greg, was chosen for genome sequencing and further analysis. Greg is a cluster A1 mycobacteriophage capable of forming lysogens, meaning it can integrate its genome into the host genome. The genome of Greg has a length of 51,083 base pairs with 88 putative genes. Each gene was analyzed for potential function, and intergenic regions are being analyzed for the presence of regulatory sequences such as promoters and terminators.

Identification of Mycobacteriophage Ukulele integration site, *attP* and characterization of Integrase in lysogeny regulation
Whitaker, E. , Beacham, G. , Hutchison, K. , Molloy, S.
University of Maine Honors College, Orono, ME, University of Maine Department of Molecular and Biomedical Sciences, Orono, ME
emily.whitaker@maine.edu

Mycobacteriophage (phage) are a group of viruses that infect bacteria in the genus *Mycobacterium*. Many phage have lysogenic lifestyles. During this lifecycle, a phage encoded integrase facilitates integration at sites *attP* in the phage genome and *attB* in the host to form a lysogen. The cluster E mycobacteriophage integration system is poorly understood. Ukulele, a lysogenic cluster E phage, is being used to identify the Cluster E *attP* and characterize lysogeny regulation. A putative *attP* containing sequence was identified in the Ukulele genome by computational analysis. To confirm the presence of *attP*, this sequence will be inserted into a plasmid and transferred into integrase expressing *M. smegmatis* (pST-KT-int). Cells will be screened for plasmid integrated into the genome. To characterize the role of the integrase in lysogeny regulation, we will determine the impact of integrase expression levels on induction event frequency in *M. smegmatis* (pST-KT-int) – Ukulele lysogens.

A new, quick, and accurate method for measuring visual color acuities in humans.
Wilson A.[1], Murphy M.[2,4], Emanetoglu N.[3,4], McNulty D.[4], Pelletier J.[1,4], Atkins L.[1,4], Kass L.[1,4]
[1]School of Biology and Ecology, University of Maine, Orono, ME, [2]John Bapst High School, Bangor, ME, [3]Department of Computer and Electrical Engineering, University of Maine, Orono, ME, [4]Sensory Cyber Systems, Ltd, Orono, ME, USA

andrew.b.wilson1@maine.edu

Visual acuity (VA) examinations are one of the most commonly performed test for visual function within the medical field, however there has been no reports on how different color combinations affect a person's acuity. We have created a VISION computer program that presents a box with an opening on one side that human subjects must correctly identify. This program has been automated to give the subjects their acuity scores quickly and accurately. Through comparisons with traditional VA charts (i.e. Landolt C), results show that the VISION computer program gives similar results to the traditional VA charts. Pilot studies have also identified color-specific differences in acuity scores (i.e. yellow box on gray background vs black box on blue background). These results show VA color combinations that people can distinguish between easily, while others are far more difficult. Current experiments are designed toward optimizing the speed, accuracy, and ease of acquiring VAs.

Characterization of Mycobacteriophage IHOP Genome and Immunity Patterns
Zwirner, C.[1], Lesko, D.[1], Milashouskas, T.[1], Tanner, C.[1], Lilley, T.[1], Doyle, J.[1], Riportella, M.[1], Molloy, S.[1]
[1]University of Maine, Orono, Maine
christian.zwirner@maine.edu

Mycobacteriophage (phage), viruses that infect mycobacteria, provide an important medium for studying genetic relationships due to their abundance, high diversity and rapid evolution. To better understand this diversity, the 2015 HON/BMB150 class isolated and characterized 22 novel phage. The mycobacteriophage, IHOP, was selected for genome sequencing. IHOP has a genome structure type (cluster) of E and like other E phage forms stable lysogens, cells with phage genome integrated into the host genome. To better understand IHOP's relationship to other types of phage, immunity assays were performed on IHOP lysogens with phage representing each cluster (A–X) of mycobacteriophage. The genome of IHOP is 75,668 base pairs in length with 140 predicted genes and two tRNAs. There are two genes with potential programmed ribosomal frameshifts. Predicted gene products were analyzed for functions and the location of promoters and terminators will be identified.