College of Science

Bioinformatics

The University of Texas at El Paso

# Annual Poster Session 2011
# Titles and Presenters

1) **Venkata Alluri and Venkatesh Rajamanickam**
   Sequence and Structural Analysis of Bm86 Cattle Tick Proteins
2) **Kyle Boone**
   Cryo-EM Image Alignment Based on Nonuniform Fast Fourier Transform
3) **Ranjan Das**
   Global Gene Expression Profiling of *Plasmodium Falciparum* In Response to the Anti-Malarial Drug Pyronaridine
4) **Karla Garcia**
   Fractionation of Copper in Aquatic Foodweb
5) **Sankara Garimella**
   Influence of Solvents on Entrapment Efficiency and Drug Release Rate of Propranolol Hydrochloride from Ethyl Cellulose Microcapsules
6) **Rishabh Jain**
   Giant Virus: A Discovery That Can Change Our Perspective Towards Evolution
7) **Joe Knapka**
   Alternatives to Perl: When Biologists Are Not Programmers
8) **Vikram Kodimala**
   Predicting Secondary Structure of Long RNA Sequences
9) **Hector Lopez**
   Secondary Structure Prediction of p90, a Cancer-Associated Protein
10) **Divya Mandyam**
    An Extremophillic Origin to Life On Earth
11) **Javier Ordonez**
    Can Molecular Phylogenetics Uncover Cryptic Speciation in a Widespread Aquatic Invertebrate?
12) **Daniel Rodarte**
    Analysis of Phospholipid-transporting ATPases (Flippases)
13) **Erika Villa**
    Calcium Regulation in *B. subtilis*
14) **Meng Zhang**
    Use of Bioinformatics Approaches to Analyze How the Tumor-Associated Antigen P62/Imp2 Interacts With Other Proteins

## BUSN 329
## Friday, November 18, 2011, 3:00 – 5:00 PM

For more information, please contact the Program Coordinator, Ms. Oliana Alikaj, at oalikaj2@utep.edu or 915.747.8484.

## #1 Sequence and Structural Analysis of Bm86 Cattle Tick Proteins

Suman Alluri,[1*] Venkatesh Rajamanickam,[1*] Leo Saldivar,[1]
Ming-Ying Leung,[1,2] and Felix Guerrero[3]
[1]Bioinformatics program,[2] Department of Mathematical Sciences
The University of Texas at El Paso, El Paso, TX, and
[3]USDA-ARS Knipling-Bushland US Livestock Insects Research Laboratory, Kerrville, TX

(*These authors contribute equally to this project)

GAVAC is a version of the Bm86 tick protein used in vaccine against cattle ticks. It works well on the cattle in Cuba, but it does not work well against *Rhipicephalus (Boophilus) microplus* from Texas and Mexico. It is hypothesized that the Texas tick Bm86 protein, called Bm86Deutsch, differs from GAVAC in structure and epitope regions. In this study, we apply various bioinformatics tools to explore the similarity and differences in sequence and structure among GAVAC, Bm86Deutsch, and two other Bm86 proteins, Bm86TICKGARD (Australian Bm86 vaccine protein) and Bm86CampoGrande (Brazilian Bm86). Major differences among the four sequences are displayed in a ClustalW multiple sequence. The hydrophobicity profiles for the proteins are obtained using the Kyte-Doolittle and Hopp-Wood scales. For secondary and tertiary structure prediction, we use programs such as PORTER, I-Tasser, and the Prime module from the Schrödinger package. We are still in the process of obtaining the tertiary structure prediction, and interpreting the results from the sequence alignment, hydrophobicity profiles, and secondary structure predictions with the aim of identifying key differences that can help design a more effective vaccines against the Texas tick.

## #2 Cryo-EM Image Alignment Based on Nonuniform Fast Fourier Transform

Kyle Boone
Bioinformatics Program, The University of Texas at El Paso, El Paso, TX

This poster presentation is based on Yang and Penczek (*Ultramicroscopy* 108:959, 2008).

Single particle reconstruction alignment is a multiple-step process that could use improvements of computational time and accuracy. The current techniques of linear interpolation and quadratic interpolation produce fuzzy stacked alignments of images, even when some oversampling is used. A new gridding technique using nonuniform fast Fourier transforms can greatly improve the fuzziness of stacked alignments without oversampling when the signal to noise ratio is high.

## #3 Global Gene Expression Profiling of *Plasmodium Falciparum* In Response to the Anti-Malarial Drug Pyronaridine

Ranjan Das
Bioinformatics Program, The University of Texas at El Paso, El Paso, TX

This presentation is based on Kritsiriwuthinan *et al*. (*Malaria Journal* 10:242, 2011).

Pyronaridine (PN) and chloroquine (CQ) are structurally related anti-malarial drugs with primarily the same mode of action. However, PN is effective against several multidrug-resistant lines of Plasmodium falciparum, including CQ resistant lines, suggestive of important operational differences between the two drugs. Synchronized trophozoite stage cultures of P. falciparum strain K1 (CQ resistant) were exposed to 50% inhibitory concentrations (IC50) of PN and CQ, and parasites were harvested from culture after 4 and 24 hours exposure. Global transcriptional changes effected by drug treatment were investigated using DNA microarrays. After a 4 h drug exposure, PN induced a greater degree of transcriptional perturbation (61 differentially expressed features) than CQ (10 features). More genes were found to respond to 24 h treatments with both drugs, and 461 features were found to be significantly responsive to one or both drugs across all treatment conditions. Filtering was employed to remove features unrelated to primary drug action, specifically features representing genes developmentally regulated, secondary stress/death related processes and sexual stage development. The only significant gene ontologies represented among the 46 remaining features after filtering relate to host exported proteins from multi-gene families.

The malaria parasite's molecular responses to PN and CQ treatment are similar in terms of the genes and pathways affected. However, PN appears to exert a more rapid response than CQ. The faster action of PN may explain why PN is more efficacious than CQ, particularly against CQ resistant isolates. In agreement with several other microarray studies of drug action on the parasite, it is not possible, however, to discern mechanism of drug action from the drug-responsive genes.

#4

# Fractionation of Copper in Aquatic Foodweb

Karla Garcia,[1] Jesica Navarrete,[2] Elizabeth Walsh[1,3]
David Borrok,[2] and Andrea Schwarzbach[4]
[1]Bioinformatics Program, Departments of [2]Geological Sciences and [3]Biological Sciences
The University of Texas at El Paso, El Paso, TX, and [4]Department of Biomedical Sciences
The University of Texas at Brownsville, Brownsville, TX

As an aquatic pollutant, copper has received considerable attention. The toxicity of copper has been demonstrated for a number of algal and rotifer species. Further it has been shown that exposing rotifers to copper in the presence of an algal food source decreases its toxicity. This affect has been detected both when algae are present during copper exposure and when rotifers are feed pre-exposed algae. To better understand the interaction between copper toxicity and trophic transfer, we investigate how copper (Cu) may be fractionated by cellular processes occurring is a simple food web consisting of a grazer, the rotifer *Brachionus plicatilis*, and its alga food source, *Tetraselmis suecica*. Algae were grown in Cu-citrate enriched media to simulate forms of organic bond Cu dissolved in natural waters and fed daily to rotifers for nine days. Algae were harvested daily while rotifers were collected 0, 5, and 8 days after algal inoculation and rinsed three times to remove any residue Cu. Samples were then digested with 70% nitric acid and ultrapure 30% hydrogen peroxide in preparation for ICP-OES analysis of Cu concentration and isotopic analysis. Preliminary results demonstrated that the algae assimilated the heavier $\delta^{65}$Cu isotope while the rotifer internalized the lighter $\delta^{63}$Cu isotope at day 0. However, as time increased rotifers assimilated the heavier isotope from the isotopically heavy algae. These findings, along with those from an earlier study investigating copper uptake by bacteria, may indicate that at lower trophic levels organisms preferentially uptake $\delta^{65}$Cu while primary consumers incorporate the heavier isotope through their diet. Cu mass balances are being calculated and additional experiments will be conducted with freshwater species to confirm these results. Our ultimate goal is to link Cu isotope changes with toxicological indicators such as reduced population growth rates and expression of metal regulating proteins.

## #5 Influence of Solvents on Entrapment Efficiency and Drug Release Rate of Propranolol Hydrochloride from Ethyl Cellulose Microcapsules

Sankara Garimella,[1,2] Ashok Reddy,[2] S. Satish,[2] and Sai Kishore[2]
[1]Bioinformatics Program, The University of Texas at El Paso, El Paso, TX, and
[2]Bapatla College of Pharmacy, Bapatla, India

The influence of solubility of drug in the dispersed phase (acetone:water mixtures) employed in the preparation on the entrapment efficiency and drug release from ethyl cellulose microcapsules was studied.

Propranolol hydrochloride was used as core and microcapsules were prepared by an emulsion solvent evaporation method. All the solvents gave discrete, large sized, free flowing spherical microcapsules. The microcapsules were evaluated for size analysis, drug content, microencapsulation efficiency, wall thickness, drug release characteristics, influence of solvent employed on entrapment efficiency and propranolol hydrochloride release from microcapsules, surface characteristics. Propranolol hydrochloride release from the microcapsules followed zero order kinetics and was influenced by the size of the microcapsules and the solvent employed in their preparation. The propranolol hydrochloride release rate from the microcapsules was found to be decreased with increased proportion of water in the dispersed phase. Among the solvents employed acetone:water mixture (92.5%:7.5%, v/v) was found to be more suitable for slow release of propranolol hydrochloride from ethyl cellulose microcapsules.

## #6 Giant Virus: A Discovery That Can Change Our Perspective Towards Evolution

Rishabh Jain[1] and Chuan Xiao[1,2]
[1]Bioinformatics Program and [2]Department of Chemistry
The University of Texas at El Paso, El Paso, TX

Giant viruses are a group of viruses which are discovered in the recent past and are the most complex viruses known till now. They have a much larger genome size with high amount of information content. They can code for hundreds of proteins which were not seen in any virus before. Their discovery has changed scientific views about the theory of evolution. It has also ignited a debate on the place of giant viruses in the "tree of life." The present study is a review of the prevailing theories about the evolution of these giant viruses. It also throws some light on the recently proposed new "tree of life."

## #7 Alternatives to Perl: When Biologists Are Not Programmers

Joe Knapka
Bioinformatics Program, The University of Texas at El Paso, El Paso, TX

Perl is a difficult language to learn well, largely due to its complex and often informally-defined semantics. It may therefore be a poor choice for scientists and students who are learning to program for the first time. This presentation evaluates Perl and some other modern, high-level programming languages, with respect to their suitability for basic tasks in bioinformatics. The task used for comparison is the Needleman-Wunsch global sequence alignment algorithm. Performance, code complexity, interoperability with existing tools, and pedagogical aspects are considered.

## #8 Predicting Secondary Structure of Long RNA Sequences

Vikram Kodimala,[1] Sameera Viswakula,[2] Rahul Vegesna,[3] and Ming-Ying Leung[1,4]
[1]Bioinformatics Program, The University of Texas at El Paso, El Paso, TX
[2]Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA
[3]The University of Texas M. D. Anderson Cancer Center, Houston, TX, and
[4]Department of Mathematical Sciences, The University of Texas at El Paso, El Paso, TX

Ribonucleic acid (RNA) is a sequence made up of 4 types of nucleotides (nt) A, U, G, and C. RNA plays an important role in all living organisms by participating in the process of transcription, translation etc. In some pathogenic viruses like HIV and Influenza, their RNA genomes control the pathogenic activities. These RNA molecules fold into complex secondary and tertiary structures. We need to predict structures of RNA molecules in order to help understand the process of pathogenicity. There are computational tools that predict secondary structure on the basis of minimal energy, but if the sequence length is long (say, 1000 or more nt), the computational time and memory requirements become excessive. To overcome this problem, we cut long sequences to smaller chunks based on the presence of inversions. Structures of chunks were first predicted using a popular secondary structure prediction algorithms pknotsRG and the results were joined together to give a predicted structure of the original sequence. To assess the retained prediction accuracy of the chunking method, we use 46 RNA sequences with lengths at least 125 nt and whose structures are contained present in the RFAM database (RFAM is a collection of RNA families with secondary structures). The structures obtained will be compared with the original structures in two ways. One is to compare the predicted structures base by base and other is to compare pair by pair.

## #9 Secondary Structure Prediction of p90, a Cancer-Associated Protein

Hector Lopez,[1] Rahul Vegesna,[2] and Ming-Ying Leung,[1,3] and Jianying Zhang,[1,4]
[1]Bioinformatics Program, The University of Texas at El Paso, El Paso, TX
[2]The University of Texas M. D. Anderson Cancer Center, Houston, TX, and
Departments of [3]Mathematical Sciences and [4]Biological Sciences
The University of Texas at El Paso, El Paso, TX

The p90 gene encodes 905 amino acids with a predicted molecular mass of 102 kD protein and contains the replicase and helicase motifs. This non-structural protein has several enzymatic activities. The localization of the non-structural protein, p90, is less well studied than that of structural proteins. This is because the non-structural proteins are expressed in relatively low concentrations in infected cells and furthermore, good antibodies have not been generated against these antigens. Structures for proteins with similar amino acids sequences to p90 are not known, but it is likely that they are replication complexes derived from endocytic vacuoles. Two host-cell encoded cycle regulatory proteins, retinoblastoma-tumor suppressor protein, and calreticulin have both been shown to interact with p90. The significance of these interactions remains to be firmly established, but limited evidence suggests that interactions between p90 and retinoblastoma protein may be important for cancer association. From these observations, it has been hypothesized that p90 expression may be an important facet of induced tumorigenesis and cancer. On the other hand, the secondary structure of p90 is currently not available; p90 was recently characterized as an inhibitor of the tumor suppressor PP2A (protein phosphatase 2A), and an autoantibody to p90 appears in high frequency in prostate cancer. Getting the secondary structure of this protein may help to determine the function of p90. In this project, four different secondary structure prediction servers, (PSIPRED, PROT, PROF, and PORTER) were used to predict the local secondary structures of p90, and the results were then converted into one common format and a consensus of these predictions is formed. Approximate start- and end-points of each helix or strand were determined based on the consensus. The results of this study will be used to help understand the possible structure and potential function of p90.

## #10 An Extremophillic Origin to Life On Earth

Divya Mandyam[1,2] and Tanya Noel[2]
[1]Bioinformatics Program, The University of Texas at El Paso, El Paso, TX, and
[2]Biology Department, York University, Toronto, Canada

Environmental conditions of the early earth resemble those tolerated by extremophillic microorganisms. This suggests that life of on earth began with a bacterium like organism having characteristics similar to extremophiles. However, controversies exist on whether the last universal common ancestor (LUCA) was a hyperthermophile, mesophile, or psychrophile. Evidence from prebiotic chemistry and thermodynamic studies suggest a mesophilic origin of life which later gave rise to hyperthermophily. However, models on an impact frustration event, rRNA based phylogenetic trees, and the relation between genomic GC content and optimal growth temperatures predict a hyperthermophillic LUCA. Both of these views are disputed from analyses of rRNA based phylogenetic trees via the slow-fast method. According to findings from this approach, LUCA may have been a psycrophile. I will present research studies which support each view and discuss future investigations which can further narrow the nature of LUCA.

## #11 Can Molecular Phylogenetics Uncover Cryptic Speciation in a Widespread Aquatic Invertebrate?

Javier Ordonez[1] and Elizabeth Walsh[1,2]
[1]Bioinformatics Program and [2]Department of Biological Sciences
The University of Texas at El Paso, El Paso, TX

Zooplankton plays predominant roles on trophic transfer of energy and nutrient in freshwater communities. Despite their importance in contributing to freshwater ecosystem services, many researchers do not distinguish among species and thereby lose resolution in their understanding of the dynamics of these systems. Compounding this problem, evidence for widespread cryptic speciation in zooplankton is accumulating with the application of DNA sequence analyses in phylogeny construction. Our objective is to build on our previous study in which we determined levels of genetic variation within the species *Lecane bulla*. Here we increase the number of populations surveyed and the sophistication of phylogenetic analyses. Using standard protocols, COI and ITS gene sequences were generated. Additional *L. bulla* sequences for these regions were obtained from Genbank. Outgroups consisted of *L. elsa*, *L. leontina*, and *L. luna*. Sequences were aligned using ClustalW and manually corrected. PAUP4.0* was used for genetic distances and to investigate phylogenetic relationship based on distance (NJ) approach. MrBayes3.1 was also used to generate phylogenies using an optimal model of nucleotide substitution determined using jModelTest software. Our results confirm and extend previous findings of high genetic divergence among Chihuahuan desert populations. Additionally, this information will help resource managers to protect and conserve the scarce water habitats in the Chihuahuan Desert.

## #12 Analysis of Phospholipid-transporting ATPases (Flippases)

Daniel Rodarte,[1] Tavis Mendez,[2] Siddhartha Das,[1,2] and Hugh Nicholas[3]
[1]Bioinformatics Program and [2]Department of Biological Sciences
The University of Texas at El Paso, El Paso, TX, and
[3]Pittsburgh Supercomputing Center, Pittsburgh, PA

ATPases are membrane-bound enzyme complexes/ion transporters that combine ATP synthesis and/or hydrolysis with the transport of protons across a membrane. ATPases have evolved different forms and functions over time, and they have been classified as F-, V-, A-, P- and E-type ATPases. Flippases belong to the P4 subfamily of the P-type ATPases, and their function is to transport phospholipids across a membrane. ATPases are found in all domains of life, and they are major pharmacological targets of drugs such as Digitalin and Omeprazole. In this work, we have conducted an analysis of phospholipid-transporting ATPases. Representative taxa from human parasites have been included in these analyses with the aim of investigating their evolution and improving known conserved patterns.

## #13 Calcium Regulation in *B. subtilis*

Erika Villa,[1] J. Valencia, C. Lema,[2] F. Cervantes-Gomez,[2] and Delfina Dominguez[1,2]
[1]Bioinformatics Program and [2]Clinical Laboratory Science Program
The University of Texas at El Paso, El Paso, TX

The importance of calcium ($Ca^{2+}$) as a cell regulator is well established in eukaryotes. But a similar role for $Ca^{2+}$ in prokaryotes is still not defined. $Ca^{2+}$ ions have been implicated in various bacterial physiological processes such as spore formation, chemotaxis, heterocyst differentiation, transport, and virulence. Several reports showed that bacteria are capable of maintaining intracellular $Ca^{2+}$ homeostasis, and $Ca^{2+}$ transients are produced in response to adaptation to nitrogen starvation, environmental stress, and metabolites of carbohydrate metabolism. These findings suggest a regulatory role for $Ca^{2+}$ in bacteria. We hypothesize that $Ca^{2+}$ ions are involved in the regulation of several intracellular processes in bacterial cells.

Using microarray data we found that 12 genes were up-regulated as $Ca^{2+}$ concentrations increased (2.5, 5 and 10 mM $CaCl_2$). Nineteen genes were down-regulated when extracellular $Ca^{2+}$ was increased. Q-rtPCR was used to verify the expression levels of certain genes regulated by calcium. Two genes were verified as having the same expression patterns. Some of the genes were identified as putative transporters, a protein kinase and enzymes of metabolism. Those genes involved in transport systems were transformed in pMutin4 which disrupts the expression of the gene. Growth measurements were taken for mutants and control cells which resulted in similar growth patterns.

Calcium ions appear to have a significant effect in the transcriptome of *B. subtilis*. Several genes appear to be specifically regulated by $Ca^{2+}$. Data suggest that $Ca^{2+}$ has a physiological role in *B. subtilis*. Future experiments involve mutating groups of genes to resolve for compensation of other transport mechanisms when a mutation is inserted.

## #14 Use of Bioinformatics Approaches to Analyze How the Tumor-Associated Antigen P62/Imp2 Interacts With Other Proteins

Meng Zhang,[1] Janeth Rodriguez,[1,2] Ming-Ying Leung,[1,3] and Jianying Zhang,[1,2]
[1]Bioinformatics Program and Departments of [2]Biological Sciences and [3]Mathematical Sciences
The University of Texas at El Paso, El Paso, TX

The protein p62 was first identified as a tumor-associated antigen (TAA) in liver cancer [Zhang *et al*. 1999]. Later, it has been demonstrated that p62 belongs to a family of proteins that bind to the 5'-UTR of the insulin-like growth factor 2 (IGF2) mRNAs, which suggests that p62 may regulate the translation of target mRNAs. In the current project, we will use bioinformatics approaches to identify the potential protein interactors from the in-vitro experiments results that are stored in databases like HPRD (Human Protein Reference Data Bank), BioGRID (Database of Protein and Genetic Interactions), DIPs (Database of Interacting Proteins), and IntAct (Protein Interaction Database and Analysis System). In the subsequent analysis, we try to evaluate these identified protein interactors by comparing the sequence similarity and structural information with the confirmed p62 interactors. However, the preliminary results did not retrieve any confirmed interactors, or known structure protein complex in the databases, and only 15 potential interactors were found. In further studies, we will focus on extracting and exploiting motifs from p62 protein sequence. Then, protein motifs will be analyzed using web-based tools such as STRING (http://string-db.org), a "one-stop shop" providing all information links between proteins, and 3DID (http://3did.irbbarcelona.org) for identifying domain-base interactions of known 3-dimensional structures. After getting the supplementary interaction information from the predictions, an algorithm can be developed to weigh the overlapping results and find more reliable interactors of p62.