Dr. Steven B. Roberts
Associate Professor

UNIVERSITY OF WASHINGTON

Dear Dr. Lakshmi Kumar Matukumalli,

Please accept this letter of intent to submit a proposal to Program #A1201. The requested information is provided below.

i. Project Director:
Steven Roberts
Associate Professor
School of Aquatic and Fishery Sciences
University of Washington
sr320@uw.edu

Collaborating Investigators:
Bill Howe
Director of Research, Scalable Data Analytics, eScience Institute
Affiliate Assistant Professor
Department of Computer Science and Engineering
University of Washington

Eric Peatman
Assistant Professor
School of Fisheries, Aquaculture and Aquatic Sciences
Auburn University

Mohamed Salem
Assistant Professor
Animal and Veterinary Sciences
West Virginia University

ii. Program Area and the Priority Area within that Program Area most closely addressed in the application: Animal Health and Production and Animal Products: Tools and Resources - Animal Breeding, Genetics and Genomics

Sincerely,

Steven Roberts

i. Title: A User-Friendly Web Interface for Genomic Analysis: Developing a Framework for Integrating Disparate Datasets to Promote Biological Discovery

ii. Rationale: There is an abundance of gene- and trait-based information for numerous agricultural animals that has not been thoroughly interrogated for biological relationships. This is because the data are in variety of formats, information is in disparate locations, and resources are not readily available to examine very large datasets. The ability to access information in a common location with custom, scalable tools via a simple and natural interface would contribute significantly to characterizing genotype-phenotype relationships.

iii. Overall Goal: To produce a web-based resource for 1) integrating existing molecular information and new gene regulatory information (*e.g.* Ag-ENCODE), 2) exposing biological relationships through visualization, 3) sharing primary and analyzed data. Our initial work will target three major US aquaculture species including oysters, catfish and trout. Aquaculture species represent an excellent system for development given the increasingly abundant genomic resources (*i.e.* draft genomes), diversity of genetic information, and agricultural importance.

iv. Specific Objectives
1) Aggregate genomic data into a web-based platform that facilitates annotation and exchange.
2) Specialize our prior visualization platforms to expose relationships in genomic data.
3) Engage with users and the general public on all aspects of resource development.

v. Approach: This project will bring together scientists studying fundamental aspects of genotypic relationships in diverse species with computer scientists and educators to ultimately improve biological discovery. Building on our prior work in the context of the UW eScience Institute on web-based services for data management, integration, and visualization, we will integrate multiple types of gene-based data (*e.g.* gene expression profiles, genotypes, epigenetic marks) into a web-based infrastructure that will allow for automated annotation and association. The interface will support direct server-side query to encourage collaboration and reuse during both exploratory and hypothesis-driven investigation. Visualization tools will be developed to assist in identifying non-obvious relationships and data will be interoperable with other relevant systems. This multi-species approach will not only facilitate biological discovery and advancement in important aquaculture species, but also generate a broad user base of scientists to ensure large community input.

vi. Potential impact and expected outcomes: This endeavor will provide a "blueprint" for integrating disparate molecular-based phenotype information for other agriculturally important species as their genome sequences become available. This project will be "open" and we will explicitly solicit feedback from the research community at all stages. The proposed project will also use the platform to advance genome-wide methods for the discovery and analysis of epigenetic diversity that will fill in gaps in our basic understanding of gene regulation and environment-phenotype interactions.