

GongXin Yu Ph.D
1217 University City BLVD S-183
Blacksburg, VA 24060
Phone: (540) 953- 5567
E-mail: yu07_2000@yahoo.com

Nov. 26, 2005

Systems Biology/Microbiology Faculty Search
Department of Biology
Indiana University
Jordan Hall 142, 1001 E 3rd Street
Bloomington, IN 47405-7005

Dear Dr. Yves Brun,

I am applying for the faculty position in Bioinformatics at Department of Biology, Indiana University. The position seems to fit very well with my education, experience, and career interests.

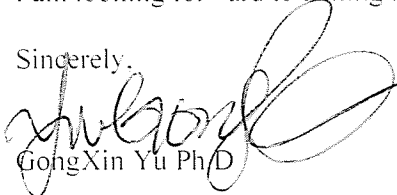
The advent of high-throughput technologies, such as genomics, DNA arrays, proteomics, mass spectrometry, and clinical data, has enabled the generation of biological information of increasing amounts and growing complexities. One of the main tasks is to extract knowledge from this massive biological data and apply the knowledge to meet challenges facing in agriculture, medics and biotech industries.

You will note from the enclosed resume that my research experience and training in both biology and computer science have well prepared me for this task. As an experimentally trained biologist, I have a deep understanding various levels of biological systems from individual molecular machines to biological processes and scientific problems associated with these systems. As an interdisciplinary scientist, I have learned how to approach these problems both analytically and programmatically. By this, I mean a carefully planned procedure to identify biological problems, collect and curate relevant but heterogeneous scientific data, develop strategies and then, computational algorithms for these problems. Furthermore, I understand that modern biological project is teamwork in nature, essentially involving multi-disciplinary, multi-institute and multi-scientist of course. I thus learned strong cooperative skills to work efficiently with other scientists of different disciplines. Because of these experiences, I am able to achieve successes in several challenging projects as witnessed by series of papers I have published from these projects. With such instances and more in mind, I am confident that I have the skills and abilities required to be a contributing member of your prestigious research team.

A letter and resume can tell you only so much about my motivation and qualification. I welcome the opportunity to discuss my background with you in person. Please let me know if you have any questions.

I am looking forward to talking to you.

Sincerely,



GongXin Yu Ph.D

Working toward systems-level understanding of biological processes: from molecular machines to systems biology

Research Interests

Biological Data Mining and Knowledge Discovering: Interested in developing computational algorithms for deriving information and new knowledge from large-scale biological data, including genomic sequence data, microarray gene expression data, proteomic data, etc. and apply such computational tools to solve real biological problems including 1) Inferring functions for hypothetical proteins, identifying critical pathogenic and immunogenic factors for host-pathogen interactions and key controlling factors for other important biochemical and biologic processes; 2) algorithms and server development for predicting protein function sites for substrate-binding specificities, gene regulatory and immunogenic control; 3) Developing database, and robust computational systems for biological data management and analysis.

Research Plans

The current computational algorithms have greatly contributed to our understanding of biological systems. However, there remains a challenge to understand, simulate, and, eventually, control molecular-scale and systems-level biological processes. To meet the challenge, I plan to, among others, 1) Develop an integrated knowledge system for highly specialized biological processes e.g. cellular signaling, immunology, cancer, and other animal and human disease; 2) An integrated genome analysis to understand pathogenesis and host immune response for Brucellosis (An example application of the knowledge base); 3) Develop an integrated system for predict protein functional sites.

1. Building knowledge databases for highly specialized biological processes e.g. cellular signaling, immunology, cancer and other animal and human diseases

Animal and human infectious agents employ a wide variety of strategies to bypass or overcome host defense mechanisms to colonize and thrive in the human and animal hosts, involving in very complex biological processes and networks. For instance, *Mycobacterium tuberculosis*, the pathogen for TB disease, devoted about 10 percent of its genes to making lipids—fatty substances in its outer coat. *Bacillus anthracis* employs poly-D-glutamate capsule to protect itself against cell lyses by cationic proteins in sera or in phagocytes. In addition, many non-infectious animal and human diseases are also involved in highly specialized biological and cellular processes and genetic networks. For example, cancer is associated with an increase in cell number, alterations in mechanisms regulating new cell birth, or cell proliferations. A systematical characterization of these highly specialized biological processes and the genetic networks, thus, would greatly facilitate our understanding biological processes in general, and animal and human diseases in particular.

In addition, much of such knowledge in most of the time exists in a piecemeal fashion and imbedded in heterogeneous data, which can impose an immense barrier to query and to gain necessary understandings. Therefore, systematically achieving in the format of knowledge database: data collecting, curating, classifying, and organizing, would be of a great benefit to understanding the diseases and the underlying biological mechanisms. Furthermore, the knowledge databases would also include *a variety of biological data* such as genomic sequence data, pathway data, protein interaction data, gene expression data, mutagenesis, and phenotypic data, and *analysis software*, such as one described in plan 3, to assist knowledge integration and applications. The databases will be organized in a way that the heterogeneous information can be easily peer-reviewed, queried, updated and refined to facilitate community research efforts.

2. An integrated genome analysis to understand pathogenesis and host immune response for Brucellosis (An example application of the knowledge database)

Brucellae, a causal agent of brucellosis, is responsible for enormous economic losses and, a potential biological weapon, poses a significant threat to human life. The *Brucella* pathogenesis is involved in a multi-step, multi-faced, host-pathogen interaction. Several lines of evidence indicated that the main virulence factor of *Brucella* is the ability to live and replicate inside cells of the reticuloendothelial system. It is first up-taken by host phagocytic cells, a proton pump (probably of cellular origin) rapidly then acidifies the phagosome. The acidification helps to build up a stress for the bacteria and induce a signal that triggers virulence genes activation. However, we have very limited understanding at this most-likely

interference point (MIP) about what the pathogenesis factors are and how they regulate the host gene expression machinery for producing the proton pump or modifying the structures of existing proton pump for such purpose. *By the MIP, we mean any possible cellular steps or components within host biochemical processes, at which the pathogenic factors can interfere (either by gene expression regulation or structure modification and conformation editing) for the surviving and multiplication of the pathogens.* Understanding host protective immune response to this agent also presents as a big challenge including the identities and nature of specific antigens (pathogenesis, virulence and other *Brucella*-originated cell components that trigger immune protection from brucellosis.

The Knowledge database would facilitate the *Brucella* research in several different aspects including:

1. Genome comparative analysis, especially between virulence strains and attenuated vaccine strain will facilitate identification of potential targets for the pathogenesis factors.
2. Differential analysis of comparative expression of micro-array and proteomics data in the database will further facilitate such screening.
3. Collecting, curating and classification of *Brucella* and other pathogen triggered host protective immune response-related pathways would provide capability to infer potential host-pathogen interacting components and facilitate understanding the biological significance of identified targets and developing experimental testable hypothesis.
4. Identification of immunogenic *Brucella* proteins by database inference and knowledge application and using computational approaches as described as following (Plan 3) to identify immunogenic epitopes for vaccine targets.

3. Develop an integrated system (algorithm and servers) for predict protein functional sites

Protein functional site is one of the most influential factors to mediate protein interaction, molecular recognitions, protein trafficking, enzymatic activities, and substrate specificities. For instance, residue mutations in the functional sites frequently act as one of the leading mechanisms for antimicrobial resistance. Binding sites in *Ras* shape the binding strategies between the *Ras* and its diverse classes of partners. Residue composition and structural conformation of small peptide fragments (epitopes) in pathogenic protein dictate host protective immune response. Identification of these function sites, therefore, will play a critical role in understanding the molecular mechanism of protein actions. We have developed series of novel computation algorithms including Surface Patch Ranking (SPR) and Shuffling Algorithm (SA), which are the decision tree-based data mining algorithms for the identification of functional sites essential to ligand-binding specificity and other biologically important protein activities. Developing advanced computational algorithms and hosting an integrated system (web-based or application-based) established from these algorithms will greatly facilitate modeling- and site-directed mutagenesis-based protein function elucidations, protein engineering, and drug and vaccine development. In addition, development of such system will foster cooperation among scientists in different fields, at the same time, provides proper environment of undergraduate education and graduate research.

Summary

This plan attempts to build an integrated computational environment to analyze genomes and other highly specialized biological processes in a genome-scale and at a systems level. In particular, I illustrate the importance of building knowledge database in facilitating studying functionally important biochemical processes and understanding the dynamics of these processes. I will dedicate such system and related strategies to study biological systems such as cancer and other human disease-related biologically important cellular processes to gain a deep insight understanding about their related signaling mechanisms.

Statement of Teaching Interests

GongXin Yu Ph.D

1217 University City BLVD S-183, Blacksburg, VA 24060

Phone: (540) 953- 5567

E-mail: yu07_2000@yahoo.com

Teaching Experience

- Lecture for an upper level of undergraduate course of “Introduction of experimental design and statistics”, Huazhong Agri. University, Wuhan, China.
- Lecture for an undergraduate and graduate level of Experimental course in Rice genetics research. Huazhong Agri. University, Wuhan, China.
- Teaching assistant for an undergraduate and graduate level of Genetics (in English), Huazhong Agri. University, Wuhan, China.
- Supervisor for summer student in bioinformatics projects in Argonne National Lab. Argonne, IL USA and Virginia Tech.

Teaching Philosophy

As a teacher at all different levels such as teaching assistant, student supervisor and lecturer, I came to realize that how importance of communication between teacher and students. It is never a simple talk in front of students but the soul searching. You have to understand both the subject you are talking and the students you are teaching. Only by this way, you can find an efficient way to deliver what you understand deeply in your heart to your students who happen to know nothing about the subjects which you are teaching. In addition, you have to be creative enough to design course works that are both interesting and challenging enough for students to get involved in. Furthermore, you have to create ideal learning and working environment in that dependent thinking and innovation are encouraged and, meanwhile, a spirit of team or cooperative work are cultivated. For example, I discovered that students participate in the course work more actively and learn more effectively if they are given some challenging but clearly defined projects and let them work in a group. In summary, as teachers, you have to remember, always, you are teaching parts of knowledge to people who will benefit themselves from the learning and so do the society!

In addition, I believe in a standard of excellence in preparing well equipped students by:

- Emphasizing strong Biological background
- Encouraging to develop capability to discover and solve biological and biomedical problems
- Encouraging creative and analytical thinking
- Motivating to acquire skills for future biological professional career

Teaching Interests:

- Bioinformatics, and bioinformatics-related graduate courses
- Genetics
- Biology

Summary

Teaching is always the passion for my professional career in the field of biology and bioinformatics. It is from there where I find a great contentment whenever students enjoy their new knowledge and get interested in the scientific subjects I taught. It is there where I find motivation to rethink about give scientific problems and look for a new interpretation. It is there where I find new talents whom I can pull into scientific research. It is there where I find a real challenge for which I take advantage of my talents and develop skills so that complicated scientific knowledge can be simplified and students can easily understand. As a consequence, I always want to take the challenge.