Tier 2 Canada Research Chair in Bioinformatics
Additional Information for Potential Applicants

St. Francis Xavier University (StFX)

Founded in 1853, St. Francis Xavier University (StFX) has a long and proud tradition as one of Canada's oldest universities. Today, StFX is one of Canada's premiere undergraduate universities. We are known for our strong spirit, our loyal alumni, our small class sizes, and our vibrant student body and residential campus experience. We have graduates in 140 countries around the globe, and this continues to strengthen the StFX experience and enrich the student experience at StFX.

Department of Mathematics, Statistics and Computer Science (MSCS)
The Department has very strong research and scholarly activities in a number of areas, and is keen to begin new collaborations. The Department seeks to support a CRC who will serve as a team builder for collaborative research opportunities involving members of the MSCS Department, other departments (Biology, Physics, etc.). The CRC, supported by the current strengths in MSCS, will be expected to take the lead on ambitious research proposals to the federal and provincial granting agencies. The Department has expertise in the broad topics that Bioinformatics deals with:

- Parallel and Distributed Computing

  Bioinformatics research relies heavily on large-scale computational resources as they need to manage Tbytes or Pbytes of data with large-scale structural and functional relationships, TFlops or PFlops of computing power for simulating highly complex models, or many-task processes and workflows for processing and analyzing data. The challenges of scale for computation, storage, and interpretation of petascale data are difficult to solve with single clusters. Parallel and distributed computing (including cloud computing) has the potential to help solve these problems by offering a utility model based on highly flexible computing and storage capabilities and abstraction layers that allows overcoming many of the constraints present in dedicated systems. It also may increase research productivity by allowing sharing applications, tools, and algorithms in an easy way.

- Resource Management and Behavior Recognition

  The heavy computational demands of Bioinformatics research require the efficient usage of computing resources. Thus, scheduling algorithms used for networked distributed systems under various system constraints, in particular, some optimization techniques, heuristic methods for energy efficient scheduling algorithms, may also be useful in the Bioinformatics domain. Similarly to Bioinformatics, learning based algorithms for

behavior recognition, which collect data through sensors, also require the storage, processing and analysis of large amounts of data. Therefore, some common problems like the system architecture and algorithms to store and process massive data are shared.

- Artificial Intelligence and Machine Learning

  Machine learning broadly refers to the changes in, e.g., structure, program, data, that a machine goes through as it "learns" in order to improve its expected future performance on a task, e.g., classification, recognition, diagnose, prediction, control, that is usually associated with Artificial Intelligence. Machine learning techniques are quickly becoming an indispensable tool in a wide range of Bioinformatics applications as it provides a variety of methods, e.g., supervised and unsupervised classification, models for knowledge discovery, deterministic and stochastic optimization methods, that are used to analyze data in the form of, e.g., sequences, images, which are critical for understanding diseases. Many techniques, e.g., artificial neural networks, support vector machines, Markov models, have been successful when handling large amounts of complex noisy data and in generalization.

- Statistical Learning and Data Mining

  Statistical learning and data mining tools (methods and algorithms) are used to analyze large data sets arising from drug discovery industry and public health. The data are usually very complicated not only in terms of the number of variables and the number of cases, but also in terms of data structures and relationships among variables and cases. Traditional statistical methods cannot deal with such complicated data, unless many assumptions are made. Statistical learning and data mining tools are definitely useful in Bioinformatics.

- Databases

  Databases and database management systems are designed to provide convenient and efficient access to large amounts of interrelated data. As Bioinformatics research needs to store and manage huge amounts of data, e.g., protein sequences, gene expression data, nucleotide sequences, macromolecular structures, the need for large databases and database management systems has grown tremendously.

- Modeling and Simulation

  Modeling and simulation has become indispensable as a way of analyzing complex systems and/or designing/testing products, e.g., a new drug. It is important to note that modeling and simulation are frequently very computationally-intensive tasks and thus rely heavily on large-scale computational resources and on the exploitation of parallel

and distributed computing methods which, as previously listed, is one of the areas in which the department has expertise.

- Discrete Mathematics

  In computer science, which forms the theoretical foundation for a major part of Bioinformatics, the most heavily used subfield of mathematics is discrete mathematics. Discrete mathematics is thus frequently present in, e.g., underlying mechanisms of algorithmic solutions, bioinformatics problems.

Computing Infrastructure

In addition to the human capacity within the Department of MSCS, and other departments at StFX, there is also the computing infrastructure capacity. As a member of the Atlantic Computational Excellence Network (ACENet) High Performance Computing (HPC) Consortium, StFX faculty have access to HPC facilities located throughout the region, and through Compute Canada to HPC resources across Canada. Bioinformatics is a computationally intensive research area and the HPC resources are expected to be sufficient to support a Tier II CRC in Bioinformatics. For further information, visit http://www.ace-net.ca/wiki/ACEnet.

Profiles of Potential Collaborators from Departments of Biology and Physics

- Dr. Cory Bishop (Biology)

  My research program involves the study of genes and proteins in various organisms. I am a regular user of the databases to which I refer here and use such databases in my teaching. I am very interested in pursuing certain questions that pertain to my existing research program, such as transcriptomic analysis of algal symbionts. More importantly, however, when I renovated my laboratory, 1/3 of my space was dedicated to a common molecular biology facility. A CRC in Bioinformatics, and his or her students, would be welcome to make use of this space as a "wet-lab" extension of their *in silico* lab. I am open to initiatives for high-throughput sequencing capacity on campus.

  In 2012 I created a course in Bioinformatics (BIOL 452) to address the fact that large complex databases that hold biological (and of course many other kinds of) data are now commonplace and we would be remiss to not expose our senior students to the incredible progress that has been made along these fronts. However, bioinformatics is a highly technical and fast moving field and it is not my area of direct expertise. Therefore the teaching programs in Biology and Math/Statistics/Computer Science would benefit enormously from an authentic leader in the field of Bioinformatics. I would welcome a collaborative effort to build courses that are would augment and be complementary to the course I currently teach.

- Dr. Bill Marshall (Biology)

The key to human understanding of biodiversity of life on the planet is to deeply analyze the diversity of the genetic information inherent in gene sequences that are being revealed progressively by biologists of all types. Physiologists, developmental biologists and cell biologists seek to understand how genes are regulated. Evolutionists and population biologists seek to understand the phylogenetic relationships among genomes. Medical and veterinary researchers seek to connect the genome sequences to diseases and their cures in unique, patient-oriented genome-based health care. A bioinformatics specialist who can write efficient search code and meaningful human interfaces to those search algorithms will be invaluable to accessing the burgeoning genomic information that is coming online.   In my research, which is now linked to the completed genome of my model organism, *Fundulus heteroclitus*, I seek to connect the robust responses of this animal to various stresses to the genes, their expression, their regulation at all levels and their relationships to similar genes in other animals.   Furthermore, the computer programs that link between gene sequences, the predicted amino acid sequences and the predicted protein tertiary structures would also help me in our desire to understand the operation of ion and water transport proteins. Improvement in the programs surrounding genomic analysis is a crucial area of human endeavor that impacts biological science, biochemistry, food science, veterinary and human health.

- Dr. Karen Brebner (Biology)

The use of bioinformatic analysis in neuroscience has become necessary in recent years, due to an explosion of data derived from genomic and proteomic studies of genes and pathways that underlie many types of neurological diseases and disorders, including addiction. These studies have resulted in highly complex databases that are used to uncover the causes of neurological diseases at the cellular, molecular and genetic level, and to direct the development of investigative targets and tools that might be used to help fight or prevent such conditions. Many laboratories around the world are now using bioinformatic techniques to search for molecular factors/pathways that are associated with diverse neuropathologies and to develop new therapeutic targets for potential treatments which are then followed up using in vitro and in vivo techniques. A CRC in Bioinformatics would benefit my research program by helping to identify specific genes or common neural pathways that are implicated in various behavioural aspects of addiction that could then be examined further in rats that are trained to press a lever to receive an intravenous infusion of various addictive drugs. This approach is considered to be one of the best animal models of human addiction and can be used to pose questions about not just the genetic or molecular changes that result from exposure to drugs of abuse, but also factors that contribute to the acquisition, maintenance, and extinction of drug-reinforced responding, vulnerability to drug addiction, likelihood of relapse and the connections between addiction-like behaviours and normal processes of learning and memory.

- Dr. Russell Wyeth (Biology)

  I look forward to pursuing potential collaborations with the Bioinformatics CRC chair. I am initiating a transcriptomics collaboration with international colleagues to study the olfactory system in gastropods. As I will still be new to the field, I expect local expertise will substantially broaden possibilities for my research program. I also anticipate undergraduate and graduate students in my lab working on projects that would benefit from training or collaboration with a bioinformatics chair (or the chair's students). Other possibly relevant research projects include breeding patterns in local invasive tunicates and analyzing reproductive isolation in nudibranch populations in the Bras d'Or lakes.

- Dr. Ricardo Scrosati (Biology)

  The Marine Ecology Lab investigates the drivers of natural biodiversity patterns on marine shores. A CRC in Bioinformatics would benefit my research program by allowing us to start investigating natural patterns of phylogenetic diversity and the factors that determine them.

- Dr. Moira Galway (Biology)

  I expect my own research and teaching to derive indirect benefits from this CRC. I use large genomics and proteomics databases to keep tabs on genes and proteins of interest to me as a plant cell and developmental biologist and teacher of genetics and cell biology. It would be beneficial for me to have a more sophisticated understanding of how large data sets are assembled and managed as databases, and to know more about database use, functions and limitations, in order to understand advances in my own areas of research interest, and to communicate and collaborate with colleagues who use bioinformatics in their own research. I would like to incorporate more about bioinformatics into my courses to improve student understanding of this topic, which is necessary to all educated citizens given the growing role of bioinformatics in our society (for example, in the health care system). These goals could be achieved through formal (seminars, workshops) and informal interactions with the CRC in Bioinformatics, members of his/her lab, as well as undergraduates who have completed course work or been involved in research and projects in this area.

- Dr. Peter Poole (Physics):

  We in the Physics Department have discussed the possibility of a CRC in Bioinformatics, and there is a general consensus that the presence of such a CRC at StFX will have positive benefits in the Physics Department.   The Physics Department has a long-standing expertise in biophysics, mainly through the work of Dr. David Pink, and more recently complemented by Dr. Shah Razul, who is an Adjunct Professor in Physics. Bioinformatics informs much recent research in biophysics, e.g. in the study of the statistical mechanics of protein folding. The computational facilities required for Bioinformatics are also complementary to those used for computational physics research. We thus expect that a Bioinformatics CRC will create collaborative opportunities both in research and in facilities development. We also anticipate that this CRC will create opportunities for Physics to participate in new academic courses and/or programs, such as in the area of biomolecular modeling.

Dr. H. Stanley and Doreen Alley Heaps Chair in Computing Science

An endowment fund of approximately $1 million created the Dr. H. Stanley and Doreen Alley Heaps Chair in Computing Science at St. Francis Xavier University (StFX). Please see the news webpage at: http://sites.stfx.ca/mscs/gift.

The principal objective of the Dr. H. Stanley and Doreen Alley Heaps Chair in Computing Science is to provide sustained funding for the support, exploration and advancement of computing science at StFX. Scholarships and summer internships are available to computing science students, and Funding for faculty or visitors is available for the scholarly pursuit of computing science. The funding for faculty or visitors leads to appointment as Alley Heaps Associate, or Alley Heaps Chair Professor.

Links

St. Francis Xavier University   http://www.stfx.ca
Mathematics, Statistics and Computer Science   http://sites.stfx.ca/mscs
Biology   http://sites.stfx.ca/biology/
Physics   http://sites.stfx.ca/physics/
Town of Antigonish   http://www.townofantigonish.ca/