Professional Science Master’s Program in Bioinformatics

Departments of Biology, Chemistry, Computer and Information Sciences

College of Science & Technology

Draft: September 17, 2014

Steering Committee:
- Dr. Jody Hey, Biology (CST)
- Dr. Ronald M. Levy, Chemistry (CST)
- Dr. Justin Y. Shi, Computer and Information Science (CST)
- Dr. Rob J. Kulathinal, Biology (CST)
- Dr. Vincent Voelz, Chemistry (CST)

Associated Faculty:
- Dr. Vincenzo Carnevale, Chemistry (CST)
- Dr. Nanjie Deng, Chemistry (CST)
- Dr. Roland Dunbrack, Fox Chase Cancer Center
- Dr. Ananias Escalante, (coming to Biology CST)
Dr. Giacomo Florin, Chemistry (CST)
Dr. Blair Hedges, Biology (CST)
Dr. Krishna Kant, Computer and Information Sciences (CST)
Dr. Michael Klein, Chemistry (CST)
Dr. Sudhir Kumar, Biology (CST)
Dr. David Liberles (coming to Biology, CST)
Dr. Anwar Mamat, Computer and Information Sciences (CST)
Dr. Spiridoula Matsika, Chemistry (CST)
Dr. Karl Morris, Computer and Information Sciences (CST)
Dr. Chris Schafmeister, Chemistry (CST)
Dr. Chiu Tan, Computer and Information Sciences (CST)
**Initial Outside Advisory Committee:**

Dr. Eric Ross,
Associate Research Professor
Assistant Vice President of Biometrics and Information Sciences
Director, Biostatistics and Bioinformatics Facility
Fox Chase Cancer Center

Dr. Jean-Pierre Issa,
Director, Fels Institute of Cancer Research and Molecular Biology
Professor of Medicine
Temple University School of Medicine

Dr. James Brown,
Director Computational Biology
Glaxo Smith Kline, Collegeville, PA

Dr. Kurtis Bachman,
Head of Computational and Systems Biology
Janssen Pharmaceuticals, Titusville, NJ
Executive Summary
Bioinformatics is a fast evolving field that integrates elements of biology, chemistry, computer science, and statistics, and has become an essential part of the biotechnology and pharmaceutical industries. Bioinformatics specialists are employed to process and analyze the high volume of genomic and proteomic data generated from large scale sequencing and related efforts that form the foundation of personalized medicine. Temple University has emerged as a leader in two key bioinformatics areas, with multiple established research programs in: i) evolutionary and genomic medicine, and ii) structural bioinformatics. The Professional Science Master’s (PSM) in Bioinformatics program at Temple University will provide the necessary training for students who complete the program to assume high level technical staff positions in both industrial and academic settings. The PSM-Bioinformatics program will emphasize practical computational skills at an advanced level to enable graduates to find rewarding careers in bio-pharmaceutical industries and at major medical centers. As bioinformatics specialists, they will become highly productive members of research teams working on such forefront problems as structure-based drug design and genomic medicine.

The PSM program in Bioinformatics will draw upon existing and specially designed graduate-level courses in Temple University's College of Science and Technology. A sampling of these courses includes: Computational Genomics, Introduction to Structural Bioinformatics, Molecular Modeling and Drug Design, Biostatistics, Scientific Programming, and Python Scripting.

To add value to the PSM-Bioinformatics program, an advisory board will be constituted with approximately eight experts in the field of bioinformatics representing industry, academia, and other research institutions in the Greater Philadelphia and Delaware Valley region. An independent research program (Capstone) will be an integral part of the training of PSM candidates. The Capstone will provide valuable experience outside the classroom for our PSM students, and will cover real world problems commonly experienced in the field of bioinformatics.

Professional Science Masters (PSM)
This degree will be recognized as a PSM degree program in Bioinformatics. The primary objective of this PSM program is to develop a portfolio of knowledge and experiences for each student that will enable them to tackle problems in genomics, proteomics, and structural bioinformatics. Other program objectives include providing (a) a foundation for understanding the major theoretical and applied concepts in bioinformatics, (b) a range of practical skills in bioinformatics, and (c) business and work skills applicable to academia, industry, and government. The goal of this program is the successful placement of graduates into relevant jobs and to achieve career advancement. The program will monitor their progress in these jobs by assessing their ability to develop initiative and leadership roles in the community and beyond.

The program includes core requirements in current topics in bioinformatics, biotechnology, genomics and proteomics, as well as bioethics and policy. Hands-on training will take place in
both the Capstone project and in the core classes of the first year. All of these classes have large computational components and for each of them students will write their own scripts, and design and carry out their own analyses. Elective classes for further specialization are also components of this curriculum.

Program Rationale
The development of new technologies based on very large databases of DNA sequences and protein structures continues to transform research and development in diverse scientific fields. The need for effective bioinformatics approaches is underscored by the rapidly increasing wealth of information encoded in protein and gene networks. The fields of proteomics and genomics have grown at a vigorous pace. This Temple program will provide graduates with the necessary knowledge base and research experience to embark on careers that will tackle these pressing issues. The curriculum is designed for students with a BA or BS degree in science or engineering. Elective courses will serve to strengthen student expertise and laboratory research fundamentals, which will complement an independent thesis research project under the direction of faculty and industry mentors.

Many examples of the kinds of positions that will become available to graduates can be easily found online at job posting websites. A search for “Bioinformatics” typically returns hundreds of positions, many of which explicitly ask for Master’s level degree of training. And virtually all of these positions seek people capable of computer programming and working with large data sets (including biological, medical and biochemical data sets). For example, bioinformaticians are in great demand to help handle the flood of genome sequence data being generated in a many biomedical and biotech research areas. This kind of work involves writing programs to gather data as it comes from one machine, and then pass it through a series of analyses. Such “pipeline” development requires good scripting skills and a kind of flexible know-how for learning how to fit new software packages into an analyses stream. For another example, many positions require skills in working with large protein structure databases and large amounts of protein-protein interactions. Such data is increasingly generated at high rates with the latest technologies.

Potential Student Pool
Our pool of students is expected to be diverse, consisting of domestic and foreign applicants, and will include professionals in health and environmental occupations who intend to obtain an MS degree, as well as students with undergraduate degrees in science, technology, engineering and mathematics (STEM), or related fields. Admittance will be based on evidence of a strong academic record and preparation in the sciences, as well as a demonstrated strong desire to fully engage in the rapidly developing field of bioinformatics.
Demand for Professional Science Masters (PSM) Degree in Bioinformatics

According to several surveys, Philadelphia is one of the largest metropolitan areas for the pharmaceutical industry. One estimate is that four-fifths of the world’s largest pharmaceutical companies—including firms like GlaxoSmithKline, Merck, and Rhone-Poulenc Rorer—have a strong presence in the Philadelphia region. According to the United States Labor Department, the employment of biological scientists is expected to increase much faster than the average of all occupations, although there will continue to be stiff competition for basic research positions. Given the growth in the biotechnology and pharmaceutical industries, the demand for a highly trained workforce with a strong science background continues to increase. The fastest area of growth in higher education is in Master’s degree programs: from 1970 to 2004, the growth was ~150%, according to the National Academy of Sciences report on Science Professionals: Master’s Education for a Competitive World (2008). For many, the PSM degree provides access to a professional career requiring both technical skills and professional development training in areas related to business, policy and ethics. As time goes on, the program will continue to focus upon some of the world’s most important health related issues that could be tackled by students trained in this program.

A survey of programmatic offerings in bioinformatics by Universities/Colleges in the Philadelphia area find that none offer a degree comparable to the proposed PSM in Bioinformatics program at Temple, and that the only institution within Philadelphia that offers a Master’s degree in Bioinformatics is the University of the Sciences. This program is difficult to assess because of limited website content (www.gradschool.usciences.edu/bioinformatics/bioinformatics-program-overview), however it appears to have only two associated faculty members and to be quite limited in the scripting/programming component of the curriculum. Rowan University, approximately 20 miles from Philadelphia, has just begun offering both a thesis-based and non-thesis-based Master of Science degree in Bioinformatics (www.rowan.edu/colleges/csm/departments/bioinformatics/programs/). Both versions can be completed with 30 credits, The Rowan programs differ from that proposed here in not being a PSM program, in not providing a Capstone internship project, and in having fewer required hands-on computational courses. In the broader metropolitan region around Philadelphia, only the University of Delaware (UD) located, 45 miles from Philadelphia, offers a Professional Science Masters in Bioinformatics (bioinformatics.udel.edu/education/psm). Like the PSM proposed here, the UD program includes an internship. However the UD program requires a total of 42 credits, and has a flexible core curriculum with, apparently, little emphasis on actual scripting/programming skills.

Program Curriculum/Degree Requirements.
The curriculum will follow a professional work-force model, with teams assembled to meet learning goals in both the classroom and laboratory fundamentals course. This project-based approach will enhance interdisciplinary learning and workplace skills, while requiring that students put their technical knowledge into practice. Required core courses will cover the topics
listed in the table below, in addition to bioinformatics laboratory courses during the first year of
the program, and elective courses in various sub-disciplines. Students will complete a capstone
research project (internship) and will be encouraged to attend journal clubs offered across three
departments.

PSM programs are expected to train students in both a core of knowledge and skills in a highly
technical area, and to provide students with professional skills. Sometimes called the “Plus”
component of a PSM, the professional skills element will be accomplished in the proposed
program by requiring an Ethics in Bioinformatics course, a Professional Development course
(Biol 5505), as well as by participating in the Capstone project. The Ethics in Bioinformatics
course will address the social, legal, and privacy concerns of applying computational
approaches to large datasets, including human datasets. The PSM Professional Development
course will introduce students to best practices by interacting with professional bioinformaticians.

<table>
<thead>
<tr>
<th>Year 1</th>
<th>Fall</th>
<th>Year 1</th>
<th>Spring</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Structural Bioinformatics I* (3 cr)</td>
<td>Structural Bioinformatics II* (3 cr)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Genomics* (3 cr)</td>
<td>Intro to Computational Genomics* (3 cr)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scripting for Sciences and Business * (3 cr)</td>
<td>Elective (3 cr)</td>
<td></td>
</tr>
<tr>
<td><strong>Total:</strong></td>
<td><strong>9 cr</strong></td>
<td><strong>Total:</strong></td>
<td><strong>9 cr</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year 2</th>
<th>Fall</th>
<th>Year 2</th>
<th>Spring</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ethics in Bioinformatics (Biol 5xxx) (2 cr)</td>
<td>Elective course (3 cr)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PSM Professional Development (Biol 5506) (1 cr)</td>
<td>Independent Research* (Biol 9995 2 cr)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Biostatistics (Biol 5312)* (3 cr)</td>
<td>Independent Research*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Independent Research* (Biol 9995 1 cr)</td>
<td>(Biol 9995 1 cr)</td>
<td></td>
</tr>
<tr>
<td><strong>Total:</strong></td>
<td><strong>7 cr</strong></td>
<td><strong>Total:</strong></td>
<td><strong>5 cr</strong></td>
</tr>
</tbody>
</table>

**Total Credit Hours: 30**  
*core courses (24 credits)

**Capstone Research Project.**
A key component of PSM programs is an capstone project that will provide students with
experience in independent and original research using bioinformatics tools, and to do so in a
real work environment. For each student, the Capstone will be supervised and mentored by a
research scientist. The Capstone may be conducted with a Temple University scientist, or with
a member of the Advisory Committee, or with someone else working in a research capacity at a
company, as approved by the Steering Committee.
The Capstone Research Project requires a total of three credits. This includes a one credit instance of Biol 9995 (Independent Study) to typically be taken in the Fall after the first year, followed by a two credit instance of Biol 9995 taken in the Spring semester of the 2nd year.

During the first part of the Capstone, a student will identify a Research Advisor, write a research proposal for approval by the Steering Committee and begin the research project. The majority of the actual research will be conducted in the second semester of ear 2.

**Elective Courses.**

The curriculum is designed so that most required courses can be taken in the first year and the elective courses taken in the second year. Students may choose to take elective courses within any one of the three departmental foundations for the program (i.e. Biology, Chemistry, and Computer and Information Sciences).

**Electives in Genetics/Genomics (Dept. Biology):**

- **Biol 3101 - Evolution (3 credits):** Topics covered include Darwinism and neo-Darwinian theory, including adaptation, natural selection, sexual selection, speciation, and techniques used to understand evolution of living and extinct organisms.
- **Biol 3211 - Human Evolution (3 credits):** This course covers the evolutionary history of humans, with an emphasis on the genetic/genomic aspects of the process.
- **Biol 5225 - Evolutionary Genetics/Genomics (3 credits):** This class covers fundamental principles of population and comparative genetics with special attention given to recent advances in genomics.
- **BIOL 5479 - Biotechnology (3 credits):** This course is designed to survey current issues in technologies including therapeutics and diagnostics, and to examine consequences of developments in this area.
- **BIOL 5505 - Ethics Regulation and Policy in Biotechnology (3 credits):** This course will provide an understanding of ethical decisions, governmental regulations and policies in biotechnology.
- **BIOL 5XXX - Advanced Topics in Computational Genomics (to be developed).**
Electives in Structural Bioinformatics (Dept Chemistry):

- **CHEM 3405** – Biophysical Chemistry (3 credits). Aspects of physical chemistry of use to biochemists in understanding the physical properties of biologically significant molecules and structures: thermodynamics, electrolytes, kinetics, quantum theory, spectroscopy, and x-ray crystallography.

- **CHEM 5301** - Quantum Chemistry (3 credits). An introduction to the principles of quantum mechanics and their application in chemistry.

- **CHEM 5302** – Statistical Thermodynamics (3 credits). An introduction to the basic concept of statistical mechanical ensembles and their application.

- **CHEM 8302** – Computational Chemistry (3 credits). An introduction to quantum mechanical and molecular mechanics methods for computing chemical structure and properties.

- **PHARMACEUTICS 8128** – Principles in Drug Discovery (3 credits). Important drug targets by class, in vitro screening methods, medicinal chemistry strategies in drug design, principles in pharmacokinetics and pharmacodynamics, animal models of disease states, clinical trial basics, and selected business aspects of the drug discovery process.

Electives in Computational Bioinformatics (Computer and Information Sciences):

- **CIS 5105** IT Process Management
- **CIS 5106** System Development Process
- **CIS 5107** Computer and Systems Security and Privacy
- **CIS 5108** Emerging Technologies
- **CIS 8108** Knowledge Management
- **CIS 8508** Enterprise Resource Plan SW
- **CIS 9603** Artificial Intelligence
- **CIS 8526** Machine Learning
- **CIS 8527** Social Network Data Analysis
- **CIS 8535** Probabilistic Graph Models
- **CIS 8537** Network and Information Security
- **CIS 9669** Distributed and Parallel Processing

**Program Outcomes.**
Graduates will attain the necessary skills for putting computers to work and to make sense of data in demanding research applications. They will have the depth of understanding and the skill sets necessary to identify the appropriate computational approach that is needed to
connect a particular data set to the question being asked of it. They will be able to read scientific papers across a diversity of advanced bioinformatics topics and to communicate effectively both with researchers who generate data and with investigators who deal strictly with models and analysis.

As a PSM program, the curriculum does not require a research thesis, and accordingly the training emphasis will not be to prepare graduates for careers as principal scientists. However graduates will be trained to work as part of research teams and to have sufficient breadth and depth to work as critical problem solvers in the research enterprise.

**Specific science-related skills and knowledge sets.**
- Knowledge of the fundamental principles of inheritance and of evolution
- Knowledge of the details of genome architecture and the processes of gene expression and of the regulation of gene expression
- Understanding of advanced concepts used in modern genomic research and of how to represent those concepts computationally
- Knowledge of protein structures and the relationship between sequence and structure
- Knowledge of molecular modeling and advanced concepts used in structural bioinformatics
- Knowledge of the developing interface between genomic and structural bioinformatics
- Knowledge of algorithms and software tools from computer and information science used in bioinformatics

**Steering Committee.**
The program Steering Committee includes Presidential faculty from the departments of Biology, Chemistry and Computer and Information Sciences. The Steering Committee members are responsible for developing and administering the program, in consultation with their respective departments, the College of Science and Technology, and the Graduate School. The Steering Committee will form and work with a Scientific Advisory Committee made up of partners from local industries, hospitals and research centers.

**Program Resources, Estimated Costs, and Revenue Stream**

**Available Resources:**

**Faculty.** Course instructors will include faculty from the Departments of Biology, Chemistry and Computer and Information Sciences. Arrangements will be made with each department to determine appropriate teaching load and credit for teaching effort. Faculty research mentors are expected to have active, funded research programs that are relevant to the program. Grant funds will be used to support the research projects.
**Instructional and Research Space.** The Biology-Life Sciences and the adjacent new SERC buildings have multiple smart instructional classrooms that would be adequate for the program. Bio-Life and SERC also have computer labs which could be used for those PSM Bioinformatics courses which require access to specialized hardware or software that is used as part of the instructional program. The Tech Center also holds multiple rooms ideal for this program, and in particular, student-led group assignments.

**Costs and Budget (Preliminary estimates).**

The program is anticipated to be self-sufficient once a minimum of ten students are enrolled, and a self-sustaining admission rate of at least 5 students per year is attained. The program will require an initial, one-time outlay of resources and there will be recurrent costs. The recurring costs will be paid for by tuition revenue.

**One time (initial) Costs:**

1. Web page development and maintenance $6,000
2. New course development 30,000
3. Administrative staff support (1/3 time including fringe) 18,000
4. IT staff support 10,000
5. Computer hardware and software licenses for computer labs 15,000

**Estimated total initial costs:** $79,000

**Recurring Costs**

1. Administrative staff support (1/3 time including fringe) $18,000
2. Instructional costs (for lecture courses and computer labs) 20,000
3. Consumables (office supplies, equipment repair and upgrade) 15,000

**Estimated recurring costs:** $53,000

**Program Revenue.**

Tuition represents the main revenue stream, and is estimated as follows: (i) 5 students will be admitted each year, and tuition is assumed to be entirely at the in-state rate: $646/cr (out-of-state rate is $933). The numbers reflect a 3% yearly increase in tuition rate. Program year 1 includes 20 cr, and year 2 includes 14 cr).

- Year 1 tuition revenue (5 students): $61,370
- Year 2 tuition revenue (10 students): $113,050
- Year 3 tuition revenue (10 students): $116,450
At steady-state, with 10 students in the program, tuition revenue will exceed $110,000 per year. This will be in excess of the estimated recurrent costs of the program.

**Administrative Structure.**

- The Department of Biology will be the program home department. The Steering Committee is responsible for oversight of the program and for communicating and coordinating with all three departments to implement the curriculum.
- The Graduate Committees of the Biology, Chemistry, and Computer and Information Science Departments will oversee the admissions process and monitoring and reporting of student progress.
- The Program Co-Directors will be appointed from the ranks of the Biology, Chemistry and Computer and Information Science Departments, and will be active researchers and instructors with bioinformatics interests.
- As required for a PSM degree program, a Scientific Advisory Board will be formed that will include appropriate individuals from local industries and research centers, including hospitals.
Appendix I: Syllabus for Structural Bioinformatics I and II

Appendix II: Course Descriptions for core courses: Genetics and Introduction to Computational Genomics

Appendix III. Course Description for first year core course in CIS
Appendix I: Syllabus for Structural Bioinformatics I and II

STRUCTURAL BIOINFORMATICS I & II

Info

Course Number - Classification:
Class Schedule:
Instructors:

Description

This course will cover the basic concepts of structural bioinformatics and molecular modeling. A broad qualitative overview of macromolecular structure and protein folding will be provided which includes sequence alignment, secondary structure calculation, and tertiary structure prediction. An introduction to programming languages, data mining and algorithms used in Bioinformatics will be covered to provide competence in handling large and complex biological data. The course also offers practical training on the application of computational modelling in aspects of drug discovery.

Objectives

The objective of this course is to introduce students to the fundamental concepts and methods in structural bioinformatics. Students will be given training on a broad range of skill sets that is required for a sound understating of structural bioinformatics. Upon successful completion, students are expected to have a working knowledge of the Linux environment, parallel computing, scripting, the ability to analyze complex biological data and structure based drug design.

Organization

Some of the lectures will be held in the computer lab to keep a close connection between theoretical concepts and computational case studies. Most of the computational work will be performed in class; however specific tasks will be completed by the students and evaluated as homework assignments.
In addition to the following reference books, additional study materials will be made available in class.


Prerequisites

Requirements

In addition to homework assignments, each student will be required to complete an individual project before the end of the course. Students are strongly encouraged to propose their own project topics.

Evaluation

Syllabus **FIRST SEMESTER**
Introduction to protein structure

Weeks 1, 2

Structure and chemistry of amino acids

Basic structural features of polypeptides
  - primary structure
  - secondary structure
  - tertiary and quaternary structure

Protein main-chain conformation and Ramachandran plots

Sidechain conformation and rotamer libraries

Protein folding patterns and structural classification and structural superposition

Protein folding and design

Weeks 2, 3

Basic concepts: stability of the native state, kinetics of protein folding

Experimental characterization of events in protein folding

Thermodynamics of protein folding, hydrophobic collapse and molten globule

Impact of free energy landscapes on folding kinetics: folding funnel

Effect of denaturants on folding/unfolding equilibrium

Relationship between native structure and folding

The hierarchical model

Protein engineering and design

Bioalgorithms

Weeks 4, 6

Introduction to UNIX and the command line
  - Shell commands, file system
  - Process management
  - Text editing

Introduction to scripting with Python
  - The interpreter
  - Data types: Strings, Lists, tuples, and dictionaries
  - Looping and Control Flow
Writing python scripts
  - Functions, Classes, Modules
  - Scripting example: mc.py
- Making plots with matplotlib
- Scientific computing with python

Searching and sorting
- Graph theory
- Depth-first vs. Breadth-first searches
- Computational complexity

Data clustering and classification
- Distance metrics
- The RMSD
- Clustering
- k-centers, k-means clustering
- Hierarchical clustering
- Support Vector Machines, regression

Databases
  Weeks 7, 8
Repositories and information retrieval
- Nucleotide sequence databases
- Protein sequence databases
- Sequence motif databases
- Protein structure databases
- Small molecule databases

Protein structure repositories and visualization tools

Bioinformatics of protein sequence and structure
  Weeks 9, 12
Alignments and Phylogenetic trees
- Measures of sequence similarity
- Computing the alignment of two sequences (Smith-Waterman)
- The dynamic programming algorithm
• Statistical significance of alignments
• Multiple sequence alignments
• Structural inferences from multiple sequence alignments
• Phylogenetic trees

Markov chains and Hidden Markov Models
• Formal definition of HMMs
• Most probable state path: the Viterbi algorithm
• The forward algorithm
• Posterior decoding
• Parameter estimation for HMMs
• HMM model structure: choice of topology

Probabilistic modeling of sequence ensembles
• The direct problem
• Statistical models and observables
• Entropy and Kullback-Leibler divergence
• The inverse problem
• Statement of the inverse problem
• Bayesian formulation
• Maximum likelihood criteria

SECOND SEMESTER

Protein structure determination
• X-ray crystallography
• Nuclear Magnetic Resonance Spectroscopy
• Low temperature electron microscopy

Protein structure prediction
• Strategies: ab initio vs knowledge-based
• Secondary structure prediction
• Homology modeling
• Fold recognition
• Critical assessment of structure prediction

**Molecular Modeling**

Weks 4, 7

• Molecular Mechanics
• Potential Energy Surface and Optimization methods

**Structure Based Drug Design**

Weks 8, 12

Virtual screening for Drug Discovery

• Molecular Docking
• Pharmacophore based screening
• Shape Based Screening

Best practices in Virtual Screening

• Benchmarking virtual screening using various evaluation metrics

Binding Free energy methods in Structure Based Drug Design

• Statistical thermodynamics expressions for the computation of binding affinities
• Relative and Absolute binding Free energy methodologies

Success stories form structure based drug design

**Hands on practical sessions**

1) Basic model building, editing, visualization and MM based energy minimization using Maestro (Schrodinger)
2) Conformational analysis using Macromodel (Schrodinger)
3) MD simulations of model peptides in explicit solvent using Impact (Schrodinger)
4) Protein Modelling (Homology Modelling) using Prime (Schrodinger)
5) Computer-Aided Drug Design: Docking using Glide (Schrodinger)
6) Binding Free energy calculation using BEDAM workflow
Appendix II: Course Descriptions for core courses: Genetics and Introduction to Computational Genomics

Genetics

This is a required course for the Fall semester of Year 1. Genetics will be a thorough course in modern genetics and genomics. The course will be co-listed with an undergraduate genetics course (Biol 2296) but will have an added computational laboratory component.

<table>
<thead>
<tr>
<th>Tentative Topic List:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mendel’s law of segregation</td>
</tr>
<tr>
<td>Basic rules of probability. Mendel’s law of independent assortment</td>
</tr>
<tr>
<td>Chromosome structure and mitosis</td>
</tr>
<tr>
<td>Meiosis</td>
</tr>
<tr>
<td>Chromosome theory of inheritance</td>
</tr>
<tr>
<td>Gene Linkage and recombination</td>
</tr>
<tr>
<td>Gene mapping</td>
</tr>
<tr>
<td>DNA structure and replication</td>
</tr>
<tr>
<td>DNA recombination</td>
</tr>
<tr>
<td>Mutations and DNA repair</td>
</tr>
<tr>
<td>Mutation analysis reveals gene structure: complementation, deletion</td>
</tr>
<tr>
<td>Transcription</td>
</tr>
<tr>
<td>The genetic code, translation</td>
</tr>
<tr>
<td>Gene regulation</td>
</tr>
<tr>
<td>Effects of mutations on gene expression and function</td>
</tr>
<tr>
<td>Analyzing genomes: Physical and sequence maps</td>
</tr>
<tr>
<td>Genome Sequencing</td>
</tr>
<tr>
<td>Genome Architecture</td>
</tr>
<tr>
<td>Transposons and transposition</td>
</tr>
<tr>
<td>Genome Interrogation: Technologies, Data Mining and Sequence Acquisition</td>
</tr>
<tr>
<td>DNA polymorphisms</td>
</tr>
<tr>
<td>Structural Variation</td>
</tr>
<tr>
<td>Population genetics: Neutral models</td>
</tr>
<tr>
<td>Population genetics: Detecting natural selection</td>
</tr>
<tr>
<td>Genome-Wide Association Studies</td>
</tr>
<tr>
<td>Mammalian Epigenomics</td>
</tr>
<tr>
<td>Genomics of Cancer</td>
</tr>
<tr>
<td>Genome Evolution</td>
</tr>
</tbody>
</table>
Introduction to Computational Genomics

- This is a new graduate course to be developed and to be taught in the Spring semester. A genetics course will be prerequisite. The course will be a hands-on computational lab course, with students working on problems and assignments in class using their laptop computers.
- Probable Topics:
  - DNA sequence alignment
  - Genome Assembly
  - Genome Annotation
  - Molecular Evolution
  - Population Genetics
  - Mutation Models
  - Phylogenetic Reconstruction
  - Phylomedicine
  - Measuring Gene Expression
  - Biological Networks
Ethics in Bioinformatics  
Biology 5xxx

Course Instructors: Professors Rob Kulathinal and Jody Hey  
Office: BioLife Building, Rm 214, Rm 206  
Phone: (215) 204-0620, (215) 204-8569  
Email: robkulathinal@temple.edu, hey@temple.edu

Guest Instructors: TBA (from Steering and Advisory Committees, as well as NTT and postdoc pool)

Course Description: This course will examine the social, legal, and privacy issues of applying computational approaches to large datasets including those from personal genome projects. The class will expose students to variation-based approaches in genomics, policies and strategies to share genomic data, database management and security, open-access and open-source philosophies, the ethics of collecting, storing, and disseminating human data, and HIPAA, FDA, and IRB regulatory policies for health care professionals and bioinformaticians. Students will be given the opportunity to discuss contemporary case studies, in addition to NIH-sanctioned online training modules (Responsible Conduct in Research). After successful completion of this course, students will be able to:

- Identify potential ethical concerns and procedural solutions in the Information Sciences
- Understand the issues related to intellectual freedom, intellectual property, and copyright laws as they relate to electronic publishing
- Interpret the impact of confidentiality and privacy laws on information security policies
- Identify case studies of professional misconduct
- Understand the areas most impacted by ethical decisions and gain insight in making such decisions
- Appreciate how rapid changes in technology might affect ethical issues and changing norms of behavior
- Understand the ethical issues associated with gathering, storing, and accessing genetic information in public and private databases
- Recognize differences in the ethical codes of conduct among different cultures and countries
- Understand the ethical issues that arise from findings in genomic and bioinformatics analyses
Prerequisites: None except enrollment in the PSM-Bioinformatics program.

Credit Hours: 2.0 (CRN xxxx)

Class Location: BioLife Sciences Building, Rm xxx
Main Campus, Temple University

Required Textbook: Bioinformatics Ethics. First Edition. CRC Press, (Taylor & Francis). Students are required to read the appropriate chapters before each lecture.

Class Format: One two-hour lecture per week. Attendance is mandatory and active participation is expected during both lectures and recitation.

Grading: Based on weekly discussions (students will lead the discussions each week) and a final project.
Appendix III. Course Descriptions for first year courses in CIS

CIS 5xx1: Scripting for Sciences and Business

Introduction
Computing has become an indispensable tool for scientific research and for businesses. The ability to integrate numerical computation with experimental results becomes a necessary condition for a productive researcher and business operators. The proposed course aims to bridge the gaps between current CS (computer science) graduate programs and the emerging scientific computing needs. In particular, this course focuses on scripting languages for scientific computing and for businesses. Historically, scripting languages have been the “second-class citizens” amongst CS curriculum designers. Formal languages are preferred. Today, the paradigm has shifted. Driven by increasingly sophisticated data needs, scripting languages have emerged from obscurity to prominence for their versatility and ease of use. For practical scientific computing projects, scripting has become critical for integrating computational results with scientific research formulations and for delivering market insights for businesses. This course is a revised version of CIS1051 Python Programming. We focus on
   a) Differences between a compiler and an interpreter and what to choose and when?
   b) Basic controls
   c) Input and parsing
   d) Output formats
   e) Error handling
   f) Execution environments and integration with other systems

Tentative Textbooks:
   a) Programming Python, 3rd Edition by Mark Lutz
   b) Bash Cookbook Solutions and Examples for Bash Users, O'Reily, Cameron/Vossen Newham
   c) The AWK Programming Language: 1st Ed, Algred V. Aho, Brian W. Kernihan, Peter J. Weinberger

Syllabus:
- 1 week: Introduction to Linux and How to make it work for you
- 1 week: Python Syntax exercises
- 1 week: Loops and Conditionals
- 2 weeks: Input and parsing
- 1 week: output formats and integration points
- Midterm
- 1 week: Linux File System
- 1 week: File manipulation using scripts
- 2 weeks: Awk
- 1 week: Linux Process Management
- 1 week: Bash
- 1 week: Error Handling and Webification
- Final Exam

Staffing: Justin Shi, Anwar Mamat, Karl Morris, Chiu Tan + guest lecturers for Awk and Bash
Expected Enrollment: 20