Overview of Computational Biology Concentration in Biological Sciences Major Cornell

Computational and quantitative analyses have become essential to biological research. Technology for collecting high-throughput data, such as genomic sequences and gene expression, mass spectrometry, MRI imaging, remote sensing, and geographic information systems, as well as the development of large-scale databases, such as those for genomes, epidemiology, animal and human population dynamics, factors of environmental concern (e.g. heavy metal concentrations in terrestrial or aquatic systems), “big data” for research, prediction and risk assessment have made available unprecedented amounts of detailed information that require computationally intensive methodologies to access and analyze. These data and the development of modeling approaches, bioinformatic tools, and computational methods for analyses are transforming almost all of biological research.

Problems investigated by computational biologists include topics as diverse as the genetics of disease susceptibility; comparing entire genomes to reveal the evolutionary history of life; predicting the structure, function and interactions of proteins; designing new therapeutic drugs; modeling the complex signaling mechanisms within cells or neural pathways in the brain; modeling animal and human population dynamics; predicting how ecosystems will respond to climate change; and designing recovery plans for endangered species.

The computational biologist must have skills in mathematics, statistics, machine learning, and the physical sciences as well as in biology. A key goal in training is to develop the ability to relate biological processes to computational and mathematical models.

Below is a list of “subareas” within the computational biology concentration in the Biological Sciences major at Cornell. Effort is made to match interests of students to faculty among these areas.

Bioinformatics -- an interdisciplinary field that both develops and applies methods and software tools for understanding biological data (including DNA, protein sequence/genomics and protein structure data). As an interdisciplinary field of science, bioinformatics combines biology, computer science, information engineering, mathematics and statistics to analyze and interpret biological data. Bioinformatics has been used for in silico analyses of biological queries using mathematical and statistical techniques. Common activities in bioinformatics include mapping and analyzing DNA and protein sequences, aligning DNA and protein sequences to compare them, and creating and viewing 3-D models of protein structures.

Computational and Statistical Genomics -- uses the latest approaches in genomics, quantitative genetics, computational sciences, bioinformatics and statistics to develop and apply computationally efficient and statistically robust methods to sort through increasingly rich and massive genome wide data sets to identify complex genetic patterns, gene functionalities and interactions, disease and phenotype associations involving the genomes of various organisms. This field is also often referred to simply as computational genomics.

Functional Genomics -- is a field of molecular biology that attempts to make use of the vast wealth of data given by genomic and transcriptomic projects (such as genome sequencing projects and RNA sequencing) to describe gene (and protein) functions and interactions. Functional genomics focuses on the dynamic aspects such as gene transcription, translation, regulation of gene expression and protein–protein interactions. Functional genomics attempts to answer questions about the function of DNA at the levels of genes, RNA transcripts, and protein products. A key characteristic of functional genomics studies is their genome-wide approach to these questions, generally involving high-throughput methods rather than a more traditional “gene-by-gene” approach. An important component of functional genomics is proteomics, the large-scale experimental and computational analysis of proteins and proteomes (e.g., the entire set of proteins that are produced or modified by an organism or system).

Modeling Biological Systems – plays a major role in systems biology, ecological modeling, epidemiology, neurobiology and mathematical biology. Sub areas include:

- Computational systems biology aims to develop and use efficient algorithms, data structures, visualization and communication tools with the goal of computer modelling of biological systems. It involves the use of computer simulations of biological systems (such as the networks of metabolites and enzymes which comprise metabolism, signal transduction pathways and gene regulatory networks), to both analyze and visualize the complex connections of these cellular processes. One goal is to create accurate real-time models of a system's response to environmental and internal stimuli.
- Ecosystem modeling includes mathematical representations of ecosystems, and, for example, simplify complex food webs down to their major components or trophic levels, and quantify these as either numbers of organisms, biomass or the inventory/concentration of some pertinent chemical element (for instance, carbon or a nutrient species such as nitrogen or phosphorus).
- Modelling infectious disease mathematically aims to discover parameters for various infections and to use those parameters to make useful calculations about epidemics, the spread of resistance, and the effects of mass vaccination programs.
- Computational neurobiology applies mathematical models and computational simulations to understand the principles that govern the development, structure, physiology and cognitive abilities of the nervous system.

Population, Quantitative and Comparative Genomics -- focuses on the comparison of genomes sampled from single species (population genomics) or different species (comparative genomics). The genomic features may include the DNA sequence, genes, gene order, regulatory sequences, and other genomic structural landmarks. Whole or large parts of genomes are compared to study basic biological similarities and differences as well as evolutionary relationships between organisms. Comparative genomic approaches start with making some form of alignment of genome sequences and looking for orthologous sequences (sequences that share a common ancestry) in the aligned genomes and checking to what extent those sequences are conserved or vary. Genome and molecular evolution are inferred and this may in turn be put in the context of, for example, phenotypic evolution or population genetics. Quantitative genomics focuses on the identification and analysis of the underlying genomic basis of quantitatively (continuously) varying phenotypes such as human height, crop yield, reproductive success, and behavior.

Questions? Contact Chip Aquadro (CFA1@cornell.edu), Director of Undergraduate Studies for the Computational Biology Concentration in the Biological Sciences major at Cornell

Sept. 25 2019 WA version