COMPUTATIONAL BIOLOGY (GRADUATE FIELD)

Program Website (https://cals.cornell.edu/computational-biology/)

Field Description

Computation has become essential to biological research. Genomic databases, protein databanks, MRI images of the human brain, and remote sensing data on landscapes contain unprecedented amounts of detailed information that are transforming almost all of biology. The computational biologist must have skills in mathematics and computation as well as in biology. A key goal in training is to develop the ability to relate biological processes to computational models.

The field provides interdisciplinary training and research opportunities in a range of subareas of computational biology involving topics such as DNA and protein databases, protein structure and function, computational neuroscience, biomechanics, population genetics, and management of natural and agricultural systems.

Students majoring in computational biology are expected to obtain a broad, interdisciplinary knowledge of fundamental principles in biology, computational science, and mathematics. But because the field covers a wide range of areas, it would be unrealistic to expect a student to master each facet in detail. Instead, students choose from specific subareas of study: they are expected to develop competence in at least one specific subdomain of biology (i.e., genetics, macromolecular biology, cellular biology, organismal biology, behavioral biology, or ecology) and in relevant subareas of computational science and mathematics.

Students are supervised by field faculty drawn from sixteen departments.

Data and Statistics

- Doctoral Program Statistics (https://gradschool.cornell.edu/about/program-metrics-assessments-and-outcomes/doctoral-program-statistics/?SelectGradField=4A)

Field Manual

- Manual (https://compbio.cornell.edu/academics/graduate/)

Subject and Degrees

Computational Biology

- Computational Biology (PhD) (https://catalog.cornell.edu/programs/computational-biology-phd/)

Concentrations by Subject

Computational Biology

- computational behavioral biology
- computational biology
- computational cell biology
- computational ecology
- computational genetics
- computational macromolecular biology
- computational organismal biology

Faculty

Daniel Nicolas Anstett (http://cals.cornell.edu/daniel-anstett/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology:* computational ecology; computational genetics
- **Research Interests:** My research focuses on rapid evolution across spatiotemporal scales with a focus on climate change and biotic interactions. My lab uses a combination of evolutionary ecology, chemical ecology, and population and landscape genomics to research geographical variation in plant trait evolution. I link these fields through an empirically driven research program that leverages computation and large geographic, phenotypic and genomic datasets to explore adaptation in 4-D (space, elevation, time). We seek to find solutions to the biodiversity crisis by leveraging genomics and spatial-based science to inform conservation management and restoration ecology. To achieve this ultimate goal, we will leverage population genetic theory and simulations (e.g. SLiM) in conjunction with herbarium and living (seed) collections to document contemporary evolution and model possible future adaptation.

Sumanta Basu (http://www.stat.cornell.edu/people/faculty/sumanta-basu/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational biology
- **Research Interests:** My primary field of work is in modeling and analysis of large-scale time series data sets. I have been collaborating with Professor Andrew Clark on developing data-driven methods for learning gene regulatory networks form time-course omics data sets (RNAseq, ChIP-seq etc.). I also collaborate with Professor Amy Kuceyeski on developing high-dimensional methods for learning functional connectivity patterns from brain signals measured using fMRI and EEG data sets.

Adam Bogdanove (http://cals.cornell.edu/adam-j-bogdanove/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational genetics
- **Research Interests:** functional genomics of plant interactions with plant pathogenic bacteria, especially rice with Xanthomonas oryzae; role of TAL effects in plant disease and applications for disease control and DNA targeting, including genome editing

James G. Booth (http://cals.cornell.edu/james-g-booth/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational genetics
- **Research Interests:** Linear Models; generalized linear models; random effects; Monte Carlo methods; bioinformatics

Adam Boyko (http://www2.vet.cornell.edu/research/faculty/adam-boyko-phd/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational biology; computational genetics
- **Research Interests:** genetics of adaptation and domestication; bioinformatics

Ilana Lauren Brito (http://www.bme.cornell.edu/faculty-directory/ilana-lauren-brito/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational biology; computational genetics
- **Research Interests:** population genetics; bacterial genomics;human microbiome;systems biology

Edward S. Buckler (http://cals.cornell.edu/edward-s-buckler/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational genetics; computational organismal biology
- **Research Interests:** plant breeding and genetics, statistical genetics, computational genetics

Casey Lu Cazer (http://www.vet.cornell.edu/research/faculty/casey-cazer-dvm-phd/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational biology
- **Research Interests:** Epidemiology of zoonotic diseases, evidence-based medicine, and One-Health; classical statistical methods, mathematical modeling, and machine learning

Andrew G Clark (http://cals.cornell.edu/andrew-g-clark/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational genetics
- **Research Interests:** Computational biology

Thomas A. Cleland (http://psychology.cornell.edu/thomas-cleland/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational biology
- **Research Interests:** computational neurobiology; perception; cognition; olfaction

Charles G. Danko (http://www.vet.cornell.edu/research/faculty/charles-danko-phd/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational biology; computational genetics
- **Research Interests:** transcriptional regulation, epigenetics, evolution of gene regulation

Iwijn De Vlaminck (http://www.bme.cornell.edu/faculty-directory/iwijn-de-vlaminck/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational biology; computational genetics
- **Research Interests:** precision medicine; single-cell genomics; circulating cell-free DNA; genomic medicine

Erik David Enbody (http://cals.cornell.edu/erik-enbody/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology:* computational biology
- **Research Interests:** Biodiversity genomics, population genetics, evolution, conservation genomics, behavioral ecology, ornithology

David James Field (http://psychology.cornell.edu/david-j-field/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational behavioral biology
- **Research Interests:** theories and models of sensory coding and visual processing; visual perception; relations between the structure of the natural environment and the representation of that environment by sensory systems

Katie Gold (http://cals.cornell.edu/kaitlin-gold/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational biology; computational ecology
- **Research Interests:** plant disease detection and epidemiology, pathology, remote sensing/imaging spectroscopy

Megan Ann Greischar (http://ecologyandevolution.cornell.edu/megan-greischar/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational biology; computational ecology
- **Research Interests:** My research investigates how subtle differences in ecology can lead to dramatic differences in evolved traits. I focus on time series--especially from infections--because those data offer unique opportunities to disentangle the interplay between ecology and complex traits. My lab uses a two-fold approach: (1) building dynamic models to quantify current understanding of ecology and determine when and why particular strategies (e.g., allocation to reproduction, the timing of development) would be favored; (2) using simulated data to assess methods for inferring traits that cannot be directly measured and developing novel approaches as needed. Incorporating even limited biological complexity quickly pushes models beyond the limits of analytical tractability, so my lab uses novel computational approaches to understand the ecological and evolutionary drivers of infection dynamics. By comparing ideal (and real) strategies, my research illuminates broader evolutionary patterns with relevance to pathogenic and free-living organisms.

Yrjo Tapio Grohn (http://www2.vet.cornell.edu/research/faculty/yrj-gr-hn/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational biology; computational organismal biology
- **Research Interests:** basic metabolism in ruminants and genetics to veterinary epidemiology, economic modeling and food safety

Giulia Guidi (http://gguidi.com/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational biology
- **Research Interests:** The use of massively parallel systems has been and continues to be critical for processing large volumes of data at an unprecedented speed and for scientific discoveries in simulation-based research areas. Today, these systems play a critical role in new and diverse areas of data science, such as machine learning, deep learning, and data analytics. Computational biology is a key

area where data processing is growing rapidly. The growing volume of data and increasing complexity have outstripped the processing capacity of single-node machines in these areas, making massively parallel systems an indispensable tool. The goal of my research is to enable new data-intensive scientific areas to achieve performance and scalability on large scale systems while maintaining productivity, through appropriate high-performance abstraction and resource and task scheduling in a structured heterogeneous programming model.

Andrew Hein (http://cals.cornell.edu/andrew-hein/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational biology; computational ecology; computational organismal biology
- **Research Interests:** My research combines computational models with state-of-the-art data collection systems to better understand the dynamics of ecological systems and the behavior of interacting organisms. I emphasize building theory that tightly links patterns in data to computational models that are useful for understanding how biological systems work. Problems I have worked on include understanding how predators and prey generate flexible yet precise sequences of movement decisions when interacting with one another, how information and misinformation spreads through biological populations, and how inter-individual behavioral interactions and emergent collective behavior can drive ecosystem states and tipping points.

Jaehee Kim (http://cals.cornell.edu/jaehee-kim/)

- **Campus:** Ithaca
- **Concentrations**: *Computational Biology*: computational biology; computational genetics
- **Research Interests:** population genetics, genetic epidemiology, stochastic process, dynamical system, Bayesian inference

Jon M Kleinberg (http://infosci.cornell.edu/content/kleinberg/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational behavioral biology
- **Research Interests:** combinatorial structure of networks and information, with applications to computational biology

Amy Frances Kuceyeski (http://stat.cornell.edu/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational behavioral biology; computational biology
- **Research Interests:** Computational neuroimaging, machine learning/ statistical modeling of brain connectivity networks, brain-behavior mapping

William Kai Ming Lai (http://cals.cornell.edu/william-lai/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational biology; computational cell biology
- **Research Interests:** Gene regulation, DNA sequencing, biochemistry, machine learning, next generation sequencing assays

Christiane Linster (http://nbb.cornell.edu/christiane-linster/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational organismal biology
- **Research Interests:** neural basis of sensory processing and memory

John T Lis (http://blogs.cornell.edu/johnlislab/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational biology; computational genetics
- **Research Interests:** gene expressions, gene regulation, molecular biology, molecular-genetics, RNA aptamer, RNA splicing, transcription

Philipp Messer (http://messerlab.org/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational genetics

Jason G. Mezey (http://mezeylab.cb.bscb.cornell.edu/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational genetics
- **Research Interests:** quantitative genetics/genomics; statistical genetics; computational biology; pathway modeling; molecular evolution

Gaurav Dilip Moghe (http://www.moghelab.org/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational biology
- **Research Interests:** plant biology, evolutionary genomics, metabolomics, natural product discovery

Shaila Ahsan Musharoff (http://cals.cornell.edu/shaila-musharoff/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational biology
- **Research Interests:** I am a statistical and population geneticist focused on addressing the statistical challenges of, and leveraging the unique insights contained in, admixed populations. Though human genetic studies have generally focused on people of European descent, this excludes a substantial portion of the global population. The findings from European populations do not necessarily generalize to other populations, which has contributed to the increasing health disparity in the United States wherein individuals of diverse ancestries have a greater health burden due to social and systemic factors, including being underserved by the medical system. One way to reduce health disparities is to better understand population-specific genetic and environmental factors that contribute to traits and disease, which can be identified by genetic studies of diverse populations. A major challenge for such studies is accounting for ancestry, both as a confounder and as a source of trait-relevant information. This challenge is particularly complex for admixed populations, whose ancestry is from multiple continents, and make up 41% of the US's population (13% are African-American and 19% are Hispanic-Latinx). In my research, I develop and apply statistical methods to model ancestry, especially in admixed populations, with the goal of reducing health disparities and learning about novel trait and disease biology.

Christopher R. Myers (http://physics.cornell.edu/christopher-myers/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational biology
- **Research Interests:** systems biology; dynamic modeling

Matt Pennell (http://cals.cornell.edu/matt-pennell/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology:* computational biology
- **Research Interests:** Immunogenetics; Phylogenetics; Molecular Evolution; Functional Genomics; Evolutionary Systems Biology

Benjamin F. Pugh (http://cals.cornell.edu/franklin-pugh/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational biology; computational genetics
- **Research Interests:** biochemical, genetic, genomic and computational approaches toward deciphering how gene regulatory proteins assemble onto a gene and regulate its expression

Hudson Kern Reeve (http://nbb.cornell.edu/hudson-kern-reeve/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational behavioral biology
- **Research Interests:** behavioral ecology; plasticity in insect social behavior

Kelly R Robbins (http://cals.cornell.edu/kelly-r-robbins/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational biology; computational genetics
- **Research Interests:** quantitative genetics, genomics and computational science to improve the efficiency of crop breeding programs and increase understanding of complex traits

Frank C. Schroeder (http://chemistry.cornell.edu/frank-c-schroeder/)

- **Campus:** Ithaca - (Minor Member)
- **Concentrations:** *Computational Biology*: computational biology

Praveen Sethupathy (http://www2.vet.cornell.edu/research/faculty/praveen-sethupathy-phd/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational genetics
- **Research Interests:** Regulation of gene expression, Genomics of microRNAs, Quantitative genetics

David B Shmoys (http://infosci.cornell.edu/content/shmoys/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational genetics
- **Research Interests:** efficient algorithms for discrete optimization with applications in computational biology

Adam C. Siepel (http://cals.cornell.edu/adam-siepel/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational biology; computational genetics
- **Research Interests:** genetics, genomics, computational biology, computational genetics

Ayshwarya Subramanian (http://cals.cornell.edu/ayshwarya-subramanian-0/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational biology; computational cell biology
- **Research Interests:** I am interested in the design, development, and applications of methods in computational systems biology for the understanding of human health, and the prevention and treatment of human disease. My research focus is understanding the principles governing cellular heterogeneity, crosstalk, and evolution in the context of complex adult human tissues and disease. I work at the interface of high-throughput datasets, applied statistics and machine learning, and experimental design. My research leverages the power of single-cell and spatial measurements to uncover novel transcriptional and regulatory programs underlying health and disease using mouse models, patient tissue, and patient-derived organoid models.

Fei Wang (http://weill.cornell.edu/faculty-highlight/fei-wang-phd/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational biology; computational cell biology
- **Research Interests:** Machine learning and data mining, Computational multi-omics analysis, Integrative computational biomedicine

April Wei (http://cals.cornell.edu/april-wei/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational biology; computational genetics
- **Research Interests:** population genetics, statistical genetics, human evolution, inference, scalable algorithm

Xiaomu Wei (http://cals.cornell.edu/xiaomu-wei/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational genetics
- **Research Interests:** cancer genetics/genomics

Martin Timothy Wells (http://www.stat.cornell.edu/~wells/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational genetics
- **Research Interests:** applications of statistics in biology

Xiangtao Xu (http://ecologyandevolution.cornell.edu/xiangtao-xu/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology:* computational ecology; computational organismal biology
- **Research Interests:** The overarching goal of my research program is to enhance our mechanistic understanding of forest ecological processes, reduce model process and parameter uncertainties, and ultimately advance our predictive capacity of vegetation dynamics. In particular, my lab combines trait-based numerical modeling and machine learning approaches to understand ecological patterns from organism to landscape scales emerging big data in global forest ecosystems. In addition, my lab develops novel computational

methods to extract ecologically relevant and useful information from multi-platform remote sensing products. Current research projects focus on understanding the resilience of tropical forests under climate change and the ecology of woody dynamics and carbon cycling in temperate forests.

Haiyuan Yu (http://www.bme.cornell.edu/faculty-directory/haiyuan-yu/)

- **Campus:** Ithaca
- **Concentrations:** *Computational Biology*: computational genetics