

COMPUTATIONAL BIOLOGY (BIOCB)

BIOCB 2010 - Introduction to Computational Biology (3 Credits)

This class is designed to inspire students in the life sciences to see the power of Computational Biology in advancing fundamental research. The course will initially be team taught by experts in diverse applications of Computational Biology, with each sharing the excitement of making clear to students how the introduction of computational approaches has been transformative in their field. The course will also teach many fundamental skills in manipulating large data sets, including genome sequences, functional genomic data, images, protein structures, etc.. The course will consist of two lectures each week and one practical exercise session each week. In the practicals, students will learn the use of many of the latest software tools, and will develop some basic programming skills. Students will learn essential problem solving skills in several aspects of genome analysis, and modeling in neurobiology, ecology, physiology and evolution. This will prepare them for future courses in these more focused biological areas covering specific applications of computer programming as well as use of specialized software tools. Students will be able to explore their own specific interests in greater depth on a term project.

Distribution Requirements: (BSC-AG, DLG-AG, OPHLS-AG)

Last Four Terms Offered: Fall 2024, Fall 2023

Learning Outcomes:

- Formulate succinct and focused research questions that make effective use of computation in diverse problems in ecology, genomics, physiology and neuroscience.
- Organize complex data structures in formats appropriate for analysis.
- Write short computer programs that organize data into formats that are readable by a wide variety of software.
- Apply appropriate analysis to data collected in these diverse fields.
- Identify software resources for tackling novel problems that you have never encountered before.
- Implement those software tools.
- Validate results and be critical in assessing their validity and generality.

BIOCB 3620 - Dynamic Models and Data in Biology (4 Credits)

Life is dynamic and ever-changing. Because of this, a central tool used to study living systems are dynamic models. This course provides an introductory survey of the development, computer implementation, and applications of dynamic models in biology, as well as statistical and machine learning methods for building such models from biological data. Case-study format covering broad range of biological applications including gene regulation, neurobiology, physiology, behavior, epidemiology and ecology. Students learn mathematical methods for interpreting and building biological systems models, as well as computational methods for simulating models on the computer using a scripting and graphics environment.

Last Four Terms Offered: Spring 2025

Learning Outcomes:

- Students will be able to read a dynamic model, interpreting its equations as statements about underlying biological processes and the assumptions made about the rates and consequences of those processes.
- Students will be able to adapt existing models for applications to related systems or alternative scenarios.
- Students will be able to write computer programs (using R) to numerically solve low-dimensional matrix equations (deterministic and stochastic), difference equation, differential equation, and agent-based models for biological systems.
- Students will be able to write computer programs (using R) to estimate the parameters and structure of dynamic models from real data.
- Students will be able to locate equilibria, compute Jacobians, evaluate local stability through eigenvalue calculations and other methods, and interpret these results in terms of asymptotic system dynamics and bifurcations.
- Students will be able to read and understand biological research papers that use modeling as a primary methodology.
- Students will be able to formulate meaningful research questions about biological systems that can be addressed using dynamic models and data, and apply the skills learned in the course to answer those questions.

Schedule of Classes (<https://classes.cornell.edu/>)

Schedule of Classes (<https://classes.cornell.edu/>)

BIOCB 4381 - Biomedical Data Mining and Modeling (3 Credits)

A biomedical data science course using Python and available bioinformatics tools and techniques for the analysis of molecular biological data, including biosequences, microarrays, and networks. This course emphasizes practical skills rather than theory. Topics include advanced Python programming, R and Bioconductor, sequence alignment, MySQL database (DBI), web programming and services (CGI), genomics and proteomics data mining and analysis, machine learning, and methods for inferring and analyzing regulatory, protein-protein interaction, and metabolite networks.

Prerequisites: at least one introductory course in computer programming (any language) and one in statistical methods, or permission of the instructor.

Last Four Terms Offered: Fall 2024, Fall 2023, Fall 2022, Fall 2021

Learning Outcomes:

- Demonstrate familiarity with the basics of applied statistical methodology.
- Demonstrate familiarity with statistical software and a programming language.
- Demonstrate ability to perform complex data mining of biological datasets using a programming language.
- Demonstrate ability to effectively communicate the results of a statistical analysis to biologists.
- Demonstrate familiarity with statistical and computational tools for high throughput genomic data.
- Demonstrate ability to build stand-alone softwares, web tools, and databases for analyzing biological data.

Schedule of Classes (<https://classes.cornell.edu/>)

BIOCB 4810 - Population Genetics (4 Credits)

Crosslisted with BIOMG 4810

Population genetics is the study of the transmission of genetic variation through time and space. This course will provide a comprehensive introduction to the fundamental concepts and methods in population genetics, with a focus on exploring how patterns of genetic variation are connected to the underlying evolutionary processes. Topics include genetic drift, mutation, coalescence theory, demography, population structure, selection, fitness, quantitative traits, selective sweeps, and adaptation at the molecular level. Emphasis is placed on the interplay between theory, computer simulations, and the analysis of genetic data from natural as well as experimental populations. We will also discuss efforts to connect genotype with phenotype and ultimately fitness. Specific case studies will include the evolution of drug resistance, genetic ancestry mapping, experimental evolution of microbes, and the genetic structure and demographic history of human populations.

Prerequisites: BIOMG 2800, BIOEE 1780, or equivalents.

Distribution Requirements: (BIO-AS), (OPHLS-AG)

Last Four Terms Offered: Fall 2023, Spring 2022, Fall 2020, Fall 2019

Learning Outcomes:

- Describe and interpret the fundamental evolutionary processes that shape patterns of genetic variation within and between populations.
- Develop simple computer simulations of fundamental evolutionary processes.
- Apply appropriate mathematical and computational analyses for inferring evolutionary parameters from population genomic data.
- Evaluate the power and limitations of inferring evolutionary parameters from population genomic data.
- Explain the application of population genetics to fields such as conservation biology, agriculture, and medicine.
- Critically assess current research findings in population genetics.
- Discuss the ethical and societal implications of population genetics research.

Schedule of Classes (<https://classes.cornell.edu/>)

BIOCB 4830 - Quantitative Genomics and Genetics (4 Credits)

A rigorous treatment of analysis techniques used to understand complex genetic systems. This course covers both the fundamentals and advances in statistical methodology used to analyze disease and agriculturally relevant and evolutionarily important phenotypes. Topics include mapping quantitative trait loci (QTLs), application of microarray and related genomic data to gene mapping, and evolutionary quantitative genetics. Analysis techniques include association mapping, interval mapping, and analysis of pedigrees for both single and multiple QTL models. Application of classical inference and Bayesian analysis approaches is covered and there is an emphasis on computational methods.

Prerequisites: BTRY 3080 and introductory statistics or equivalent.

Distribution Requirements: (OPHLS-AG)

Last Four Terms Offered: Spring 2025, Spring 2024, Spring 2023, Spring 2022

Learning Outcomes:

- Students will learn a statistical modeling strategy that is both basic and general, as well as how to apply this strategy to learn information about biological systems when analyzing genome-wide data. More specifically, students will learn the mathematics and interpretation of linear statistical models.
- Students will learn what these models can be used to infer when applied to genome-wide genetic and related data.
- Students will learn how to effectively and efficiently analyze large-scale genomic data and how to program in R for this purpose.
- Students will learn the limits of interpretation when applying these statistical models to genomic data when inferring information about a biological system.

Schedule of Classes (<https://classes.cornell.edu/>)

BIOCB 4840 - Computational Genetics and Genomics (4 Credits)

Crosslisted with CS 4775

Computational methods for analyzing genetic and genomic data. Topics include sequence alignment, hidden Markov Models for discovering sequence features, motif finding using Gibbs sampling, phylogenetic tree reconstruction, inferring haplotypes, and local and global ancestry inference. Prior knowledge of biology is not necessary to complete this course. Grad students must do a final project that involves original research and that in most circumstances will involve programming and real data. Undergrads will not need to do research but will do a final project that involves a sizeable amount of programming, comprehensive literature review of a topic, or similar.

Prerequisites: BTRY 3010 and CS 2110 or equivalents.

Distribution Requirements: (BIO-AS, SDS-AS), (OPHLS-AG)

Last Four Terms Offered: Fall 2024, Fall 2023, Fall 2022, Fall 2020

Learning Outcomes:

- Understand computational algorithms used for the analysis of genetic and genomic data
- Formulate computational approaches for solving problems in computational genomics
- Understand challenges and limitations in inference methods used in computational genetics and genomics

Schedule of Classes (<https://classes.cornell.edu/>)

BIOCB 4910 - Advanced Population Genetics (3 Credits)

This course covers the latest development and cutting-edge research topics in population genetics, aiming to enable students to perform research in population genetics. The first part will cover coalescent theory and inference involving complex demography. The second part will discuss natural selection and methods for inferring selection. We will allude to the complexity of demographic history and natural selection and their importance in explaining genomic patterns. The third part will introduce new data types and the challenges and opportunities with these data. We will dive into genotype likelihood and will emphasize the importance of simulation in population genetics. The course will be mostly delivered through lecturing, each interspersed with short conversations about reading assignments. Coursework involves reading literature, solving problem sets, and a course project.

Prerequisites: BIOMG 2800 or BIOEE 1780, BTRY 3080 or BTRY 3010, BTRY 4810, or equivalents.

Last Four Terms Offered: Spring 2025, Spring 2024, Spring 2023

Learning Outcomes:

- Apply population genetic theory to interpret patterns of genetic data.
- Explain how mutation, recombination, gene flow, and selection affect the coalescence process.
- Develop mathematical models to describe the genetic data-generating process.
- Construct the genotype likelihood under different models.
- Choose appropriate population genetic inference tools for analyzing different types of genetic data.
- Extract and use public domain genetic data and perform sensible tests on them.
- Develop population genetic simulations with popular simulators such as SLiM and msPrime.

Schedule of Classes (<https://classes.cornell.edu/>)

BIOCB 6010 - Introduction to Computational Biology (3 Credits)

This class is designed to inspire students in the life sciences to see the power of Computational Biology in advancing fundamental research. The course will initially be team taught by experts in diverse applications of Computational Biology, with each sharing the excitement of making clear to students how the introduction of computational approaches has been transformative in their field. The course will also teach many fundamental skills in manipulating large data sets, including genome sequences, functional genomic data, images, protein structures, etc.. The course will consist of two lectures each week and one practical exercise session each week. In the practicals, students will learn the use of many of the latest software tools, and will develop some basic programming skills. Students will learn essential problem solving skills in several aspects of genome analysis, and modeling in neurobiology, ecology, physiology and evolution. This will prepare them for future courses in these more focused biological areas covering specific applications of computer programming as well as use of specialized software tools. Students will be able to explore their own specific interests in greater depth on a term project.

Last Four Terms Offered: Fall 2024, Fall 2023

Learning Outcomes:

- Formulate succinct and focused research questions that make effective use of computation in diverse problems in ecology, genomics, physiology and neuroscience.
- Organize complex data structures in formats appropriate for analysis.
- Write short computer programs that organize data into formats that are readable by a wide variety of software.
- Apply appropriate analysis to data collected in these diverse fields.
- Identify software resources for tackling novel problems that you have never encountered before.
- Implement those software tools.
- Validate results and be critical in assessing their validity and generality.

Schedule of Classes (<https://classes.cornell.edu/>)

BIOCB 6381 - Biomedical Data Mining and Modeling (3 Credits)

A biomedical data science course using Python and available bioinformatics tools and techniques for the analysis of molecular biological data, including biosequences, microarrays, and networks. This course emphasizes practical skills rather than theory. Topics include advanced Python programming, R and Bioconductor, sequence alignment, MySQL database (DBI), web programming and services (CGI), genomics and proteomics data mining and analysis, machine learning, and methods for inferring and analyzing regulatory, protein-protein interaction, and metabolite networks.

Prerequisites: at least one introductory course in computer programming (any language) and one in statistical methods, or permission of the instructor.

Last Four Terms Offered: Fall 2024, Fall 2023, Fall 2022, Fall 2021

Learning Outcomes:

- Demonstrate familiarity with the basics of applied statistical methodology.
- Demonstrate familiarity with statistical software and a programming language.
- Demonstrate ability to perform complex data mining of biological datasets using a programming language.
- Demonstrate ability to effectively communicate the results of a statistical analysis to biologists.
- Demonstrate familiarity with statistical and computational tools for high throughput genomic data.
- Demonstrate ability to build stand-alone softwares, web tools, and databases for analyzing biological data.

Schedule of Classes (<https://classes.cornell.edu/>)

BIOCB 6620 - Dynamic Models and Data in Biology (4 Credits)

Life is dynamic and ever-changing. Because of this, a central tool used to study living systems are dynamic models. This course provides an introductory survey of the development, computer implementation, and applications of dynamic models in biology, as well as statistical and machine learning methods for building such models from biological data. Case-study format covering broad range of biological applications including gene regulation, neurobiology, physiology, behavior, epidemiology and ecology. Students learn mathematical methods for interpreting and building biological systems models, as well as computational methods for simulating models on the computer using a scripting and graphics environment.

Last Four Terms Offered: Spring 2025

Learning Outcomes:

- Students will be able to read a dynamic model, interpreting its equations as statements about underlying biological processes and the assumptions made about the rates and consequences of those processes.
- Students will be able to adapt existing models for applications to related systems or alternative scenarios.
- Students will be able to write computer programs (using R) to numerically solve low-dimensional matrix equations (deterministic and stochastic), difference equation, differential equation, and agent-based models for biological systems.
- Students will be able to write computer programs (using R) to estimate the parameters and structure of dynamic models from real data.
- Students will be able to locate equilibria, compute Jacobians, evaluate local stability through eigenvalue calculations and other methods, and interpret these results in terms of asymptotic system dynamics and bifurcations.
- Students will be able to read and understand biological research papers that use modeling as a primary methodology.
- Students will be able to formulate meaningful research questions about biological systems that can be addressed using dynamic models and data, and apply the skills learned in the course to answer those questions.

Schedule of Classes (<https://classes.cornell.edu/>)

BIOCB 6810 - Population Genetics (4 Credits)

Crosslisted with BIOMG 6810

Population genetics is the study of the transmission of genetic variation through time and space. This course will provide a comprehensive introduction to the fundamental concepts and methods in population genetics, with a focus on exploring how patterns of genetic variation are connected to the underlying evolutionary processes. Topics include genetic drift, mutation, coalescence theory, demography, population structure, selection, fitness, quantitative traits, selective sweeps, and adaptation at the molecular level. Emphasis is placed on the interplay between theory, computer simulations, and the analysis of genetic data from natural as well as experimental populations. We will also discuss efforts to connect genotype with phenotype and ultimately fitness. Specific case studies will include the evolution of drug resistance, genetic ancestry mapping, experimental evolution of microbes, and the genetic structure and demographic history of human populations.

Prerequisites: BIOMG 2800, BIOEE 1780, or equivalents.

Last Four Terms Offered: Fall 2023

Learning Outcomes:

- Describe and interpret the fundamental evolutionary processes that shape patterns of genetic variation within and between populations.
- Develop simple computer simulations of fundamental evolutionary processes.
- Apply appropriate mathematical and computational analyses for inferring evolutionary parameters from population genomic data.
- Evaluate the power and limitations of inferring evolutionary parameters from population genomic data.
- Explain the application of population genetics to fields such as conservation biology, agriculture, and medicine.
- Critically assess current research findings in population genetics.
- Discuss the ethical and societal implications of population genetics research.

Schedule of Classes (<https://classes.cornell.edu/>)

BIOCB 6830 - Quantitative Genomics and Genetics (4 Credits)

A rigorous treatment of analysis techniques used to understand complex genetic systems. This course covers both the fundamentals and advances in statistical methodology used to analyze disease and agriculturally relevant and evolutionarily important phenotypes. Topics include mapping quantitative trait loci (QTLs), application of microarray and related genomic data to gene mapping, and evolutionary quantitative genetics. Analysis techniques include association mapping, interval mapping, and analysis of pedigrees for both single and multiple QTL models. Application of classical inference and Bayesian analysis approaches is covered and there is an emphasis on computational methods.

Prerequisites: BTRY 3080 and introductory statistics course or equivalent.

Last Four Terms Offered: Spring 2025, Spring 2024, Spring 2023, Spring 2022

Learning Outcomes:

- Students will learn a statistical modeling strategy that is both basic and general, as well as how to apply this strategy to learn information about biological systems when analyzing genome-wide data. More specifically, students will learn the mathematics and interpretation of linear statistical models.
- Students will learn what these models can be used to infer when applied to genome-wide genetic and related data.
- Students will learn how to effectively and efficiently analyze large-scale genomic data and how to program in R for this purpose.
- Students will learn the limits of interpretation when applying these statistical models to genomic data when inferring information about a biological system.

Schedule of Classes (<https://classes.cornell.edu/>)

BIOCB 6840 - Computational Genetics and Genomics (4 Credits)

Computational methods for analyzing genetic and genomic data. Topics include sequence alignment, hidden Markov Models for discovering sequence features, motif finding using Gibbs sampling, phylogenetic tree reconstruction, inferring haplotypes, and local and global ancestry inference. Prior knowledge of biology is not necessary to complete this course. Grad students must do a final project that involves original research and that in most circumstances will involve programming and real data.

Prerequisites: BTRY 3010 and CS 2110 or equivalents.

Last Four Terms Offered: Fall 2024, Fall 2023, Fall 2022, Fall 2020

Learning Outcomes:

- Understand computational algorithms used for the analysis of genetic and genomic data
- Formulate computational approaches for solving problems in computational genomics
- Understand challenges and limitations in inference methods used in computational genetics and genomics

Schedule of Classes (<https://classes.cornell.edu/>)

BIOCB 6890 - Current Topics in Population Genomics (1 Credit)

Graduate seminar on current topics in population genetics. Readings are chosen primarily from current scientific literature. Participation in discussion and presentation of at least one paper required for course credit.

Prerequisites: BIOMG 4810, BIOCB 4810 or permission of instructor.

Last Four Terms Offered: Spring 2025, Fall 2024, Fall 2023, Fall 2022
Schedule of Classes (<https://classes.cornell.edu/>)

BIOCB 6910 - Advanced Population Genetics (3 Credits)

This course covers the latest development and cutting-edge research topics in population genetics, aiming to enable students to perform research in population genetics. The first part will cover coalescent theory and inference involving complex demography. The second part will discuss natural selection and methods for inferring selection. We will allude to the complexity of demographic history and natural selection and their importance in explaining genomic patterns. The third part will introduce new data types and the challenges and opportunities with these data. We will dive into genotype likelihood and will emphasize the importance of simulation in population genetics. The course will be mostly delivered through lecturing, each interspersed with short conversations about reading assignments. Coursework involves reading literature, solving problem sets, and a course project.

Prerequisites: BIOMG 2800 or BIOEE 1780, BTRY 3080 or BTRY 3010, BTRY 4810, or equivalents.

Last Four Terms Offered: Spring 2025, Spring 2024, Spring 2023

Learning Outcomes:

- Apply population genetic theory to interpret patterns of genetic data.
- Explain how mutation, recombination, gene flow, and selection affect the coalescence process.
- Develop mathematical models to describe the genetic data-generating process.
- Construct the genotype likelihood under different models.
- Choose appropriate population genetic inference tools for analyzing different types of genetic data.
- Extract and use public domain genetic data and perform sensible tests on them.
- Develop population genetic simulations with popular simulators such as SLiM and msPrime.
- Apply scientific writing and presentation skills.

Schedule of Classes (<https://classes.cornell.edu/>)

BIOCB 7200 - Statistical and Computational Genetics (1 Credit)

Weekly seminar series on recent advances in computational genomics. A selection of the latest papers in the field are read and discussed. Methods are stressed, but biological results and their significance are also addressed.

Prerequisites: BIOCB 4840 or BIOCB 6840 or CS 4775, or equivalent.

Last Four Terms Offered: Spring 2025, Spring 2024, Spring 2023, Spring 2021

Schedule of Classes (<https://classes.cornell.edu/>)

BIOCB 7210 - Topics in Quantitative Genomics (1 Credit)

Weekly seminar series on recent advances in quantitative genomics. A selection of the latest papers in the field is read and discussed. Methods are stressed, but biological results and their significance are also addressed.

Prerequisites: BIOCB 4830 or BIOCB 6830 or permission of instructor.

Last Four Terms Offered: Spring 2024, Spring 2023, Spring 2022, Spring 2021

Schedule of Classes (<https://classes.cornell.edu/>)

BIOCB 7600 - Data Driven Models in Biology (1 Credit)

Graduate seminar on methods for building models of biological systems using data, with an emphasis on recent methods including machine learning tools. Students will read and discuss recent literature in this area and, through group discussions, develop strategies for applying methods within their own research domains. Participation in discussion and presentation of at least one paper required for course credit.

Last Four Terms Offered: Spring 2025, Spring 2024

Learning Outcomes:

- Describe and explain methods for developing mathematical and machine learning models using biological data.
- Describe recent developments in data-driven modeling and how these methods can be applied to study biological systems.
- Present results of recent scientific studies to their peers (4) discuss applications of data-driven modeling methods within their own research domains.

Schedule of Classes (<https://classes.cornell.edu/>)

BIOCB 7700 - Topics in Statistical Genetics (1 Credit)

Weekly seminar series on recent advances in statistical genetics. A selection of the latest papers in the field are read and discussed. Methods are stressed, but biological results and their significance are also addressed.

Last Four Terms Offered: Fall 2024

Learning Outcomes:

- After completing this course, students should be able to critically read a statistical genetics paper.
- After completing this course, assess the strengths and weaknesses of a scientific paper.
- After completing this course, determine whether a paper's conclusions follow from the presented analyses.
- After completing this course, describe the current state of statistical genetics research.

Schedule of Classes (<https://classes.cornell.edu/>)