



CIEE Global Institute - Yucatan

Course name:	Introduction to Bioinformatics
Course number:	(GI) BIOL 2007 MEME
Programs offering course:	Yucatan Open Campus Block: STEM and Society
Open Campus track:	STEM and Society
Language of instruction:	English
U.S. Semester Credits:	3
Contact Hours:	45
Term:	Fall 2019

Course Description

Bioinformatics studies how biological information and computers can be used. Specifically, it is the science of developing computer infrastructure to facilitate and expedite archiving, searching and analyzing heterogeneous biological data for large scale ecological research. In this course, students will learn the basic concepts and tools of bioinformatics, including biological databases, sequence alignment, gene and protein structure prediction, molecular phylogenetics, genomics and proteomics. They will apply these tools and concepts to ecology, biogeography, pharmacology and conservation. Upon successful completion, the students will understand key theoretical, methodological and practical issues that underpin the flow of environmental information from source to application, appreciating the contribution different scientific and social disciplines have in understanding and constructing environmental information pathways and their uses.

Learning Objectives

Upon completion of this course, students will:

- Define bioinformatics, including basic concepts and methodologies
- Analyze biological data, including but not limited to, gene sequences, gene expression, biological networks using existing tools using existing tools
- Evaluate bioinformatics tools and their application to biodiversity
- Interpret biological data phylogenetically
- Relate bioinformatics to ecology and conservation

Course Prerequisites

None

Methods of Instruction



This course is taught through the use of lectures (CIEE instructors and guest speakers), discussions, computer workshops, interviews, readings, and an internet based research project. There are co-curricular visits to local research centers and field sites. CIEE-led lectures, readings, computer workshops and guided internet and computer research with discussions supply foundational information, concepts, and terminology, and help students make necessary connections. Guest lectures and interviews with researchers and environmental professionals offer unusual opportunities to learn about “on-the-ground” application of bioinformatics to biological understanding and conservation.

Assessment and Final Grade

Independent research project written and oral report	20%
Weekly Quizzes	20%
Problem Sets	10%
Essays on Speakers/Site Visits	15%
Final exam	20%
Participation	20%

Course Requirements

Independent research project written and oral report (20%)

Students will undertake an internet-based project to use bioinformatics to construct a phylogeny or address a local conservation problem. Evaluation will be based on (1) quality of data collection and analysis; (2) an individual written report (formatted for a conservation peer-review journal); (3) an oral presentation (Powerpoint) for an audience of scientist peers.

Weekly Quizzes (15%) At the end of each week, students will complete a quiz covering content from that week. There will be three quizzes in total. Quizzes will include True/False, multiple choice, fill in the blank, short and long answer formats.

Problem Sets (10%)

Each week, students will complete a set of questions or address problems associated each topic. These may be related to concepts, computations, programming or application of topics explored in class.

Essays on Speakers/Site Visits (30%)

Students will write critical essays addressing topics from invited speakers and site visits. These essays will summarize major elements of the talk or visit, fully explain its link to bioinformatics, and extend learning outcomes by further researching associated applications from bioinformatics to biodiversity and conservation.

Final Exam (20%)



Students take a comprehensive exam on basic bioinformatic concepts, applications to biodiversity and conservation.

Participation

Participation is valued as meaningful contribution in the digital and tangible classroom, utilizing the resources and materials presented to students as part of the course. Meaningful contribution requires students to be prepared in advance of each class session and to have regular attendance. Students must clearly demonstrate they have engaged with the materials as directed, for example, through classroom discussions, online discussion boards, peer-to-peer feedback (after presentations), interaction with guest speakers, and attentiveness on co-curricular and outside-of-classroom activities.

Attendance Policy

Regular class attendance is required throughout the program, and all unexcused absences will result in a lower participation grade for any affected CIEE course. Due to the intensive schedules for Open Campus and Short Term programs, unexcused absences that constitute more than 10% of the total course will result in a written warning.

Students who transfer from one CIEE class to another during the add/drop period will not be considered absent from the first session(s) of their new class, provided they were marked present for the first session(s) of their original class. Otherwise, the absence(s) from the original class carry over to the new class and count against the grade in that class.

For CIEE classes, excessively tardy (over 15 minutes late) students must be marked absent. Attendance policies also apply to any required co-curricular class excursion or event, as well as to Internship, Service Learning, or required field placement. Students who miss class for personal travel, including unforeseen delays that arise as a result of personal travel, will be marked as absent and unexcused. No make-up or re-sit opportunity will be provided.

Attendance policies also apply to any required class excursion, with the exception that some class excursions cannot accommodate any tardiness, and students risk being marked as absent if they fail to be present at the appointed time.

Unexcused absences will lead to the following penalties:

<i>Percentage of Total Course Hours Missed</i>	<i>Equivalent Number of Open Campus Semester classes</i>	<i>Minimum Penalty</i>
Up to 10%	1 content classes, or up to 2 language classes	Participation graded as per class requirements
10 – 20%	2 content classes, or 3-4 language classes	Participation graded as per class requirements; written warning
More than 20%	3 content classes, or 5 language classes	Automatic course failure , and possible expulsion



Weekly Schedule

NOTE: this schedule is subject to change at the discretion of the instructor to take advantage of current experiential learning opportunities.

Week 1 Introduction to Bioinformatics

Session 1.1 Introductory Concepts

Students begin with an overview of bioinformatics to review and assemble general principles of molecular biology necessary to understand genetic sequences, structures, interactions, metabolism and regulation. Students will appreciate data streams producing new data and how bioinformatics manages them. They will describe essential characteristics of a data base: coverage, organization and access routes for information retrieval. Students explore the role computer hardware and software plays in bioinformatics and how to create programs and develop expertise in their use. The instructor will demonstrate the use of PERL in Bioinformatics and will work with groups of students on how to write a simple Bioinformatics program. Students discuss applications and limitations of Bioinformatics.

Readings: Chapter 1 Introduction, and

Bik HM (2017) Let's rise up to unite taxonomy and technology. *PLoS Biol* 15(8): e2002231. <https://doi.org/10.1371/journal.pbio.2002231>

Ouzounis, C.A., 2012. Rise and demise of bioinformatics? Promise and Progress. *PLoS Computational Biology*, 8(4), p.e1002487

Vincent, A.T. and Charette, S.J., 2015. Who qualifies to be a bioinformatician?. *Frontiers in genetics*, 6, p.164

Week 2 Genomic Organization and Evolution

Session 2.1 Genome Organization and Evolution

Students investigate the basic sizes, contents and organizing principles of simple and complex genomes. They define genes, transcriptomes and proteomes. They see how genomes are analyzed and the relation of gene sequences to phenotypic features. Students recognize the importance and challenges of deriving amino acid sequences and from a genome sequence. Students go on to explore the genomes of prokaryotes and eukaryotes and how they differ.

Weekly Quiz 1

Due: Assigned problems from Chapter 1



Readings: Chapter 1 Genomic Organization and Evolution, pp. 59-86 and Weiss, K.M. and Lambert, B.W., 2014. What type of person are you? Old-fashioned thinking even in modern science. *Cold Spring Harbor perspectives in biology*, 6(1), p.a021238, Lewis, J., Bartlett, A. and Atkinson, P., 2016. Hidden in the middle: culture, value and reward in bioinformatics. *Minerva*, 54(4), pp.471-490

Session 2.2 The Human Genome

Students identify the basic structure and features of the human genome. They consider protein-coding genes, repeat sequences, ARN, single nucleotide polymorphisms and haplotypes. They link genes to phenotype using inheritable diseases in humans and how the complete human genome has changed medicine. Students discuss ethical, legal and social issues related to human genomic information. They explore the use of genetic information in anthropology and in personal identification. Students will begin working with bioinformatics data using R.

Readings Chapter 2: Genome Organization and Evolution, and Huang, X., Hawkins, B.A., Lei, F., Miller, G.L., Favret, C., Zhang, R. and Qiao, G., 2012. Willing or unwilling to share primary biodiversity data: results and implications of an international survey. *Conservation Letters*, 5(5), pp.399-406

Due: Assigned exercises and problems from Chapter 2

Week 3 Scientific Publication and Archives

Session 3.1 Media, Content and Access

Students examine and differentiate general kinds of data describing the molecules and processes of life assembled in the the data banks supporting research and applications in biology and conservation. They consider the basic infrastructure of bioinformatics in terms of the sites and responsibilities of major archival projects. Students comprehend basic concepts of information retrieval, including how to frame inquiries. They gain facility with general online search engines and websites for bioinformatics. Students successfully search information on specific sequences, structures, metabolic pathways. They discuss their applications, including conservation, and how to launch analyses of data retrieved. Finally, students survey programming languages and tools most often used in bioinformatics.

Readings: Chapter 3 Scientific Publications and Archives: Media, Content and Access, and Cock, P.J., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B. and De Hoon,



M.J., 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11), pp.1422-1423

Weekly Quiz 2

Session 3.2 Site Visit to Genomics Institute

Students visit a local genomics institute, tour their facilities and speak with professionals in genomic research and bioinformatics. Students will talk to researchers about how bioinformatics are used, local copyright laws or customs. They will also discuss how bioinformatics professionals are trained and how their work benefits society.

Session 3.3 Archives and Information Retrieval. Students consider database indexing and specification of search terms. They then explore different bioinformatics databases, including nucleic acid sequence databases, genome databases and genome browsers, protein sequence databases, protein family databases and bibliographic databases. They discuss genetic information copyrights. Finally, they begin framing a bioinformatics inquiry to investigate.

Readings: Chapter 4 Archives and Information Retrieval and

Cock, P.J., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B. and De Hoon, M.J., 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11), pp.1422-1423

Doolittle, R.F., 1981. Similar amino acid sequences: chance or common ancestry?. *Science*, 214(4517), pp.149-159

Heckel, R., 2018. An archive written in DNA. *Nature biotechnology*, 36(3), p.236

Silvester, N., Alako, B., Amid, C., Cerdeño-Tarrága, A., Clarke, L., Cleland, I., Harrison, P.W., Jayathilaka, S., Kay, S., Keane, T. and Leinonen, R., 2017. The European nucleotide archive in 2017. *Nucleic acids research*, 46(D1), pp.D36-D40

Due: Problem Set, and Essay on Site Visit

Week 4

Phylogenetic Trees

Session 4.1 Alignments

Students define sequence alignment and the assignment of residue-residue correspondences. They construct and interpret dotplots and connect dotplots to alignments. Students also define the Hamming and Levenshtein distances as measures of dissimilarity of character strings. They distinguish scoring schemes



for string alignment and appreciate differences between global and local alignments.

Readings: Chapter 5 Alignments and Phylogenetic Trees, pp. 175-202 and, Booth, A., Mariscal, C. and Doolittle, W.F., 2016. The modern synthesis in the light of microbial genomics. *Annual review of microbiology*, 70, pp.279-297

Due: Problem Set

Weekly Quiz 3

Session 4.2 Phylogeny

They work with Z scores and interpret p and E values returned by database searches. Students interpret multiple alignments of amino acid sequences and make inferences from multiple sequence alignments about protein structure. They define homology, similarity, clustering and phylogeny. They use PSI-BLAST and related programs. In the end, students understand the contents and significance of phylogenetic trees and the methods available for deriving them.

Readings: Chapter 5 Alignments and Phylogenetic Trees, pp. 203-216 and, Puttick, M.N., O'Reilly, J.E., Tanner, A.R., Fleming, J.F., Clark, J., Holloway, L., Lozano-Fernandez, J., Parry, L.A., Tarver, J.E., Pisani, D. and Donoghue, P.C., 2017. Uncertain-tree: discriminating among competing approaches to the phylogenetic analysis of phenotype data. *Proc. R. Soc. B*, 284(1846), p.20162290

Session 4.3 Constructing Phylogenetic Trees

Students work in groups to construct multiple phylogenetic trees using different data sources and different taxa. They use maximum parsimony and maximum likelihood to distinguish different trees. They will also know the role and use of an outgroup in deriving phylogenies. Students also work on a proposal and begin collecting data for their independent research projects.

Reading: Termignoni-García, F., Jaramillo-Correa, J.P., Chablé-Santos, J., Liu, M., Shultz, A.J., Edwards, S.V. and Escalante-Pliego, P., 2017. Genomic footprints of adaptation in a cooperatively breeding tropical bird across a vegetation gradient. *Molecular ecology*, 26(17), pp.4483-4496

Assigned exercises problems from Chapter 5.

Week 5 Structural Bioinformatics, Drug Discovery and Systems Biology



Session 5.1: Structural Bioinformatics.

Students investigate protein stability and protein folding. They discuss applications of hydrophobicity. Students also consider superposition of structures and structure alignments. They explore the evolution of protein structures. Students will investigate applications of Bioinformatics to Biodiversity beyond Phylogeny. This will include a survey of how bioinformatics is applied to ecological, evolutionary and behavioral studies. Students will see how genetics are informing taxonomy, levels of biological organization, species identification, ecological studies on niche partitioning, biogeography, behavioral studies of inclusive fitness and many others. Students will explore recent bioinformatics literature to more fully appreciate the depth and breadth of bioinformatics to these disciplines, including the Barcode of Life Project. They will discuss if Bioinformatics will replace field taxonomic knowledge.

Readings: Chapter 6 Structural Bioinformatics and Drug Discovery, and Baptiste, E. and Huneman, P., 2018. Towards a Dynamic Interaction Network of Life to unify and expand the evolutionary theory. *BMC biology*, 16(1), p.56
Lanfear, R., Welch, J.J. and Bromham, L., 2010. Watching the clock: studying variation in rates of molecular evolution between species. *Trends in Ecology & Evolution*, 25(9), pp.495-503

Due: Problem Set

Weekly Quiz

Session 5.2 Protein Structure, Function and Drug Discovery/Development

Students study protein structure prediction and modeling. They discuss a priori and empirical methods, critical assessment of structure prediction, homology modeling, field recognition and use conformational energy calculations with molecular dynamics. They explore the assignment of protein structures to genomes and learn how to predict protein function from structure. Students then apply these concepts to drug discovery and development.

Read: Chapter 6 Structural Bioinformatics and Drug Discovery, pp. 241-276, and Peterson, A.T. and Vieglais, D.A., 2001. Predicting Species Invasions Using Ecological Niche Modeling: New Approaches from Bioinformatics Attack a Pressing Problem: A new approach to ecological niche modeling, based on new tools drawn from biodiversity informatics, is applied to the challenge of predicting potential species' invasions. *AIBS Bulletin*, 51(5), pp.363-371



Stuart, Y.E., Campbell, T.S., Hohenlohe, P.A., Reynolds, R.G., Revell, L.J. and Losos, J.B., 2014. Rapid evolution of a native species following invasion by a congener. *Science*, 346(6208), pp.463-466

Due: Essay on Site Visit

Session 5.3 System Biology

Students introduce themselves to Systems Biology and define it as the mathematical and computational modeling of biological systems. They investigate multiple examples of Systems Biology, including protein complexes, metabolic pathways, gene regulatory networks, species populations and entire ecosystems. Students use graphing techniques to explore networks. They generate and interpret indices of stability and robustness to explore complexity and chaos.

Readings: Chapter 7 Introduction to Systems Biology and Nicholson, J.K. and Lindon, J.C., 2008. Systems biology: metabonomics. *Nature*, 455(7216), p.1054, and Du, W. and Elemento, O., 2015. Cancer systems biology: embracing complexity to develop better anticancer therapeutic strategies. *Oncogene*, 34(25), p.3215.

Due: Problem Set

Week 6

Session 6.1 Metabolic Pathways.

Students will assign and classify protein function. They will learn the Gene Oncology Consortium protein function classification. Students will explore enzyme function, cofactors and kinetics. They investigate how proteins evolve new functions by controlling enzyme activity, as well as structural evolution of altered or novel protein functions. They consider protein evolution at the level of domain assembly. Students discover databases of metabolic pathways. They examine the evolution and alignment of metabolic pathways, and dynamics of metabolic networks.

Reading: Chapter 8 Metabolic Pathways

Due: Problem Sets

Weekly Quiz

Session 6.2 Gene expression and regulation



Students define and analyze DNA microassays, exploring how they are both quantitative but lack precision. They see how mass spectroscopy can identify components of a complex mixture, including protein and genome sequencing. They describe properties of protein-protein complexes. Students also investigate regulatory complexes, including in bacteriophages and yeast. Students begin presenting the results of their research using bioinformatics tools.

Reading: Chapter 9 Gene Expression and Regulation

Due: Problem Sets

Session 6.3 Bioinformatics and Conservation Biology.

Students will use bioinformatics to assess genetic diversity of endangered species populations. They then apply this to estimates of minimum viable population size, assessments of fragmentation and its impact on populations, identifying rare or endangered subspecies or genetically distinct populations and to measure genetic distances between geographically distinct populations. They will examine case studies of Cheetahs, Humpback Whales and Bornean and Sumatran Orangutans, among others. Students apply genetic tools to conservation forensics, to track illegal wildlife trade throughout the world. Students finish presenting final projects.

Readings: Drew, J., 2011. The role of natural history institutions and bioinformatics in conservation biology. *Conservation Biology*, 25(6), pp.1250-1252 Faith, D.P. and Baker, A.M., 2006. Phylogenetic diversity (PD) and biodiversity conservation: some bioinformatics challenges. *Evolutionary bioinformatics*, 2, p.117693430600200007 Ouborg, N.J., Pertoldi, C., Loeschcke, V., Bijlsma, R.K. and Hedrick, P.W., 2010. Conservation genetics in transition to conservation genomics. *Trends in genetics*, 26(4), pp.177-187 Pérez, G.D., Jiménez, I., Rey, I., Sánchez-Cañete, F.J.S., Luque, F., Muñoz, A.M., Gálvez, M., Galdós, J.S., Sánchez, A., Tham, T.E.R. and Sánchez, V.F.V., 2013. Genomics and proteomics in bioarchaeology. *Archaeobios*, (7), pp.4-17

Final Exam

Course Materials

Textbook

Lesk, A., 2014. *Introduction to Bioinformatics 4th edition*. Oxford University Press.

Readings



- Bapteste, E. and Huneman, P., 2018. Towards a Dynamic Interaction Network of Life to unify and expand the evolutionary theory. *BMC biology*, 16(1), p.56
- Bik HM 2017. Let's rise up to unite taxonomy and technology. *PLoS Biol* 15(8): e2002231. <https://doi.org/10.1371/journal.pbio.2002231>
- Cock, P.J., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B. and De Hoon, M.J., 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11), pp.1422-1423
- Doolittle, R.F., 1981. Similar amino acid sequences: chance or common ancestry?. *Science*, 214(4517), pp.149-159
- Du, W. and Elemento, O., 2015. Cancer systems biology: embracing complexity to develop better anticancer therapeutic strategies. *Oncogene*, 34(25), p.3215
- Heckel, R., 2018. An archive written in DNA. *Nature biotechnology*, 36(3), p.236
- Huang, X., Hawkins, B.A., Lei, F., Miller, G.L., Favret, C., Zhang, R. and Qiao, G., 2012. Willing or unwilling to share primary biodiversity data: results and implications of an international survey. *Conservation Letters*, 5(5), pp.399-406
- Lanfear, R., Welch, J.J. and Bromham, L., 2010. Watching the clock: studying variation in rates of molecular evolution between species. *Trends in Ecology & Evolution*, 25(9), pp.495-503
- Lewis, J., Bartlett, A. and Atkinson, P., 2016. Hidden in the middle: culture, value and reward in bioinformatics. *Minerva*, 54(4), pp.471-490
- Nicholson, J.K. and Lindon, J.C., 2008. Systems biology: metabonomics. *Nature*, 455(7216), p.1054
- Ouborg, N.J., Pertoldi, C., Loeschcke, V., Bijlsma, R.K. and Hedrick, P.W., 2010. Conservation genetics in transition to conservation genomics. *Trends in genetics*, 26(4), pp.177-187
- Ouzounis, C.A., 2012. Rise and demise of bioinformatics? Promise and Progress. *PLoS Computational Biology*, 8(4), p.e1002487
- Pérez, G.D., Jiménez, I., Rey, I., Sánchez-Cañete, F.J.S., Luque, F., Muñoz, A.M., Gálvez, M., Galdós, J.S., Sánchez, A., Tham, T.E.R. and Sánchez, V.F.V., 2013. Genomics and proteomics in bioarchaeology. *Archaeobios*, (7), pp.4-17
- Peterson, A.T. and Vieglais, D.A., 2001. Predicting Species Invasions Using Ecological Niche Modeling: New Approaches from Bioinformatics Attack a Pressing Problem: A new approach to ecological niche modeling, based on new tools drawn from biodiversity informatics, is applied to the challenge of predicting potential species' invasions. *AIBS Bulletin*, 51(5), pp.363-371
- Puttick, M.N., O'Reilly, J.E., Tanner, A.R., Fleming, J.F., Clark, J., Holloway, L., Lozano-Fernandez, J., Parry, L.A., Tarver, J.E., Pisani, D. and Donoghue, P.C., 2017. Uncertain-tree: discriminating among competing approaches to the phylogenetic analysis of phenotype data. *Proc. R. Soc. B*, 284(1846), p.20162290
- Silvester, N., Alako, B., Amid, C., Cerdeño-Tarrága, A., Clarke, L., Cleland, I., Harrison, P.W., Jayathilaka, S., Kay, S., Keane, T. and Leinonen, R., 2017. The European nucleotide archive in 2017. *Nucleic acids research*, 46(D1), pp.D36-D40



- Stuart, Y.E., Campbell, T.S., Hohenlohe, P.A., Reynolds, R.G., Revell, L.J. and Losos, J.B., 2014. Rapid evolution of a native species following invasion by a congener. *Science*, 346(6208), pp.463-466
- Termignoni-García, F., Jaramillo-Correa, J.P., Chablé-Santos, J., Liu, M., Shultz, A.J., Edwards, S.V. and Escalante-Pliego, P., 2017. Genomic footprints of adaptation in a cooperatively breeding tropical bird across a vegetation gradient. *Molecular ecology*, 26(17), pp.4483-4496
- Booth, A., Mariscal, C. and Doolittle, W.F., 2016. The modern synthesis in the light of microbial genomics. *Annual review of microbiology*, 70, pp.279-297
- Vincent, A.T. and Charette, S.J., 2015. Who qualifies to be a bioinformatician?. *Frontiers in genetics*, 6, p.164
- Weiss, K.M. and Lambert, B.W., 2014. What type of person are you? Old-fashioned thinking even in modern science. *Cold Spring Harbor perspectives in biology*, 6(1), p.a021238
- Faith, D.P. and Baker, A.M., 2006. Phylogenetic diversity (PD) and biodiversity conservation: some bioinformatics challenges. *Evolutionary bioinformatics*, 2, p.117693430600200007