

1.) DATE: 09/10/2012	2.) COMMUNITY COLLEGE: Maricopa Co. Comm. College District
3.) COURSE PROPOSED: Prefix: BIO Number: 283 Title: Bioinformatics and Scientific Computing Credits: 3	
CROSS LISTED WITH: Prefix: CSC Number: 283 ; Prefix: Number: ; Prefix: Number: ; Prefix: Number: ; Prefix: Number: ; Prefix: Number: ;	
4.) COMMUNITY COLLEGE INITIATOR: SAGARIKA DASH PHONE: 623 845 3635 FAX:	
ELIGIBILITY; Courses must have a current Course Equivalency Guide (CEG) evaluation. Courses evaluated as NT (non-transferable are not eligible for the General Studies Program.	
MANDATORY REVIEW: <input type="checkbox"/> The above specified course is undergoing Mandatory Review for the following Core or Awareness Area (only one area is permitted; if a course meets more than one Core or Awareness Area, please submit a separate Mandatory Review Cover Form for each Area). POLICY: The General Studies Council (GSC-T) Policies and Procedures requires the review of previously approved community college courses every five years, to verify that they continue to meet the requirements of Core or Awareness Areas already assigned to these courses. This review is also necessary as the General Studies program evolves.	
AREA(S) PROPOSED COURSE WILL SERVE: A course may be proposed for more than one core or awareness area. Although a course may satisfy a core area requirement and an awareness area requirement concurrently, a course may not be used to satisfy requirements in two core or awareness areas simultaneously, even if approved for those areas. With departmental consent, an approved General Studies course may be counted toward both the General Studies requirements and the major program of study.	
5.) PLEASE SELECT EITHER A CORE AREA OR AN AWARENESS AREA: <u>Core Areas:</u> Computer/statistics/quantitative applications (CS) <u>Awareness Areas:</u> Select awareness area...	
6.) On a separate sheet, please provide a description of how the course meets the specific criteria in the area for which the course is being proposed.	
7.) DOCUMENTATION REQUIRED <input checked="" type="checkbox"/> Course Description <input checked="" type="checkbox"/> Course Syllabus <input checked="" type="checkbox"/> Criteria Checklist for the area <input checked="" type="checkbox"/> Table of Contents from the textbook required and/or list or required readings/books <input checked="" type="checkbox"/> Description of how course meets criteria as stated in item 6.	
8.) THIS COURSE CURRENTLY TRANSFERS TO ASU AS: <input type="checkbox"/> DEC prefix <input checked="" type="checkbox"/> Elective Current General Studies designation(s): None Effective date: 2013 Spring Course Equivalency Guide Is this a multi-section course? <input checked="" type="checkbox"/> yes <input type="checkbox"/> no Is it governed by a common syllabus? <input checked="" type="checkbox"/> yes <input type="checkbox"/> no District-wide course competencies/outline	
Chair/Director: PETER BROWN Chair/Director Signature: Emailed approval to J Ricker	

AGSC Action: Date action taken: Approved Disapproved
Effective Date:

Proposer: Please complete the following section and attach appropriate documentation.

ASU--[CS] CRITERIA			
A COMPUTER/STATISTICS/QUANTITATIVE APPLICATIONS [CS] COURSE MUST SATISFY ONE OF THE FOLLOWING CRITERIA: 1, 2, OR 3			
YES	NO		Identify Documentation Submitted
		1. Computer applications*: courses must satisfy both a and b:	
<input checked="" type="checkbox"/>	<input type="checkbox"/>	a. Course involves the use of computer programming languages or software programs for quantitative analysis, modeling, simulation, animation, or statistics.	Course Description, Course Competencies, Course Outline, Syllabus, Textbook Table of Contents (for specifics, please see the table on the last page of this checklist)
		b. Course requires students to analyze and implement procedures that are applicable to at least one of the following problem domains (check those applicable):	
<input checked="" type="checkbox"/>	<input type="checkbox"/>	i. Spreadsheet analysis, systems analysis and design, and decision support systems.	Course Description, Course Competencies, Course Outline, Syllabus, Textbook Table of Contents (for specifics, please see the table on the last page of this checklist).
<input checked="" type="checkbox"/>	<input type="checkbox"/>	ii. Graphic/artistic design using computers.	Course Description, Course Competencies, Course Outline, Syllabus, Textbook Table of Contents (for specifics, please see the table on the last page of this checklist).
<input type="checkbox"/>	<input checked="" type="checkbox"/>	iii. Music design using computer software.	
<input checked="" type="checkbox"/>	<input type="checkbox"/>	iv. Modeling, making extensive use of computer simulation.	Course Description, Course Competencies, Course Outline, Syllabus, Textbook Table of Contents (for specifics, please see the table on the last page of this checklist).

ASU--[CS] CRITERIA			
<input checked="" type="checkbox"/>	<input type="checkbox"/>	v. Statistics studies stressing the use of computer software.	Course Description, Course Competencies, Course Outline, Syllabus, Textbook Table of Contents (for specifics, please see the table on the last page of this checklist).
<p>*The computer applications requirement cannot be satisfied by a course, the content of which is restricted primarily to word processing or report preparation skills; learning a computer language or a computer software package; or the study of the social impact of computers. Courses that emphasize the use of a computer software package or the learning of a computer programming language are acceptable, provided that students are required to understand, at an appropriate level, the theoretical principles embodied in the operation of the software and are required to construct, test, and implement procedures that use the software to accomplish tasks in the applicable problem domains.</p>			
2. Statistical applications: courses must satisfy both a and b.			
<input type="checkbox"/>	<input type="checkbox"/>	a. Course has a minimum mathematical prerequisite of College Mathematics, College Algebra, or Precalculus, or a course already approved as satisfying the MA requirement.	
<input type="checkbox"/>	<input type="checkbox"/>	b. The course must be focused principally on developing knowledge in statistical inference and include coverage of all of the following:	
YES	NO		Identify Documentation Submitted
<input type="checkbox"/>	<input type="checkbox"/>	i. Design of a statistical study.	
<input type="checkbox"/>	<input type="checkbox"/>	ii. Summarization and interpretation of data.	
<input type="checkbox"/>	<input type="checkbox"/>	iii. Methods of sampling.	
<input type="checkbox"/>	<input type="checkbox"/>	iv. Standard probability models.	
<input type="checkbox"/>	<input type="checkbox"/>	v. Statistical estimation	
<input type="checkbox"/>	<input type="checkbox"/>	vi. Hypothesis testing.	

Course Prefix	Number	Title	Designation
BIO	283	BIOINFORMATICS AND SCIENTIFIC COMPUTING	CS

Explain in detail which student activities correspond to the specific designation criteria. Please use the following organizer to explain how the criteria are being met.

Criteria (from checkstreet)	How course meets spirit (contextualize specific examples in next column)	Please provide detailed evidence of how course meets criteria (i.e., where in syllabus)
<p>Criterion I.a.</p> <p>Course involves the use of computer programming languages or software programs for quantitative analysis, modeling, simulation, animation, or statistics.</p>	<p>Criterion I. a.</p> <p>Students enrolled in this course are required to learn Perl programming language and develop software for various bioinformatic applications.</p> <p>Students will learn programming in R for statistical analysis and graphic presentation of biological data.</p> <p>Students will learn to write programs to access (file handling, user input and output, accessing data from hash tables), retrieve, store data for analysis.</p>	<p>Criterion I. a.</p> <p>MCCCD Official Course Competencies:</p> <ol style="list-style-type: none"> 1. Identify major genetic databases and access tools and describe special features and strengths of each. (I) 2. Use software to perform basic statistical analyses and characterize sequences. (II) 3. Use software to locate open reading frames. (II) 4. Use available software to search genetic databases for similar sequences. (III) 5. Use available software to align sequences and determine similarities. (III, IV) <p>MCCCD official course outline.</p> <p>I. Introduction:</p> <ol style="list-style-type: none"> A. What are Bioinformatics and Scientific Computing? 2. Computer applications to molecular biology B. History C. Definitions and Concepts D. Uses of Bioinformatics E. Levels of Focus F. Major Genetic Databases and Access Tools <ol style="list-style-type: none"> d. BLAST searching programs for DNA and protein sequences e. Online Mendelian Inheritance in Man f. Other NCBI databases and tools 3. Other Protein Databases and Software <p>II. Analyzing One Sequence</p> <ol style="list-style-type: none"> A. Nucleotide Sequences <ol style="list-style-type: none"> 1. Detecting Vector Contamination 2. Restriction Maps 3. Analyzing DNA/RNA composition <ol style="list-style-type: none"> a. Basic statistics and probability b. Counting words and repeats

		<p>c. GC content: percentage of Guanine and Cytosine nucleotides e. Analysis software</p> <p>4. Finding Protein Coding Regions a. Open Reading Frames b. Exons c. Assembly B. Protein Sequences 1. Predicting physical properties 2. Predicting structure 3. Finding domains</p> <p>VII. Scientific Computing A. Text Editors 1. Creating, editing, saving, and retrieving data files 2. Converting data files to different file formats a. Operating system file formats b. FASTA (Fast-All) and other sequence file formats</p> <p>Course Syllabus:</p> <p>Module 1: Introduction to Bioinformatics and Biological Databases</p> <p>Lecture 1: Introduction to Bioinformatics Chapter 1. Analyzing DNA, RNA and Protein Sequences in Database (Bioinformatics and Functional Genomics by Pevsner)</p> <p>Assigned Reading from Fundamental Concepts of Bioinformatics by Krane and Raymer. Chapter 1. Molecular Biology and Biological Chemistry</p> <p>Lab Exercise 1: Exploring NCBI website: www.ncbi.nlm.nih.gov</p> <p>Access literature information on PubMed www.ncbi.nlm.nih.gov/entrez/ (Learn how to use the fields and limits to optimize the literature search for the assigned research topics).</p> <p>Lecture 2: Introduction to Biological Databases</p> <p>Chapter 2: Access to Sequence Data and Literature Information (Bioinformatics and Functional Genomics by Pevsner).</p> <p>Lab Exercise 2:</p> <p>Retrieve DNA sequences from GenBank www.ncbi.nlm.nih.gov/entrez/ (learn about EMBL and DDBJ, learn about FASTA and understand the various features of the GenBank record).</p> <p>Lab Exercise 3:</p>
--	--	--

	<p>Use of bioinformatics software for data retrieval and pairwise alignment.</p> <p>Students will install ActivePerl and EnginSite Perl Editor LE for writing and running Perl programs.</p> <p>Students will download and install R software and learn programming in R for statistical analysis and graphical presentation of bioinformatics data.</p> <p>Pairwise alignment software: BLAST FASTA PSI-BLAST</p>	<p>Retrieve Protein sequences using (http://www.expasy.org/)</p> <p>Store both DNA and protein sequence in FASTA format (*.txt) for further analysis.</p> <p>Module 2: Programming in Perl</p> <p>Chapters 1, 2 and 3: Beginning Perl for Bioinformatics by Jame Tisdall</p> <p>Chapter 1: Biology and Computer Science Chapter 2: Getting Started with Perl: 2.3 Installing Perl on Your Computer 2.4 How to Run Perl Programs 2.5 Text Editors</p> <p>Lecture:</p> <p>Programming in Perl: Installation of ActivePerl and EnginSite Perl Editor LE for writing Perl programs.</p> <p>Variables, functions, regular expressions, pattern matching, data structures, file handling.</p> <p>Module 3: Introduction to programming in R for statistical analysis of bioinformatics data.</p> <p>Pairwise Sequence Alignment and Multiple Sequence Alignment.</p> <p>Chapter 2. Data Searches and Pairwise Alignments. (Fundamental Concepts of Bioinformatics by Krane and Raymer)</p> <p>Lab Exercise 4:</p> <p>Run BLAST (Basic Local Alignment Search Tool) for comparison of nucleotide sequence with other nucleotide sequences and translated sequences.</p> <p>Learn to use BLAST to compare your protein sequence with other protein sequences.</p> <p>Hands- on practice with nucleotide and protein sequence comparisons using BLAST http://blast.ncbi.nlm.nih.gov/Blast.cgi</p> <p>Lab Exercise: 5</p> <p>Download and install R software. Create data tables in Excel and how to import data to R environment and use R functions for statistical analysis</p>
--	--	--

	<p>Multiple sequence alignment ClustalW2 - Multiple Sequence Alignment software (EBI) www.ebi.ac.uk/Tools/msa/clustalw2/ Geneious MUSCLE T-Coffee</p>	<p>of bioinformatics data and create graphs. Lab Exercise 6: Multiple Protein Sequence Alignment using ClustalW2 http://pir.georgetown.edu/pirwww/search/multialn.shtml Students use the sequence manipulation suite to analyze and manipulate their nucleotide and protein sequences. www.bioinformatics.org/sms2/</p>
<p>riterion 1.b.: Course requires students to analyze and implement procedures that are applicable to at least one of the following problem domains (check those applicable):</p>	<p>The course requires students to use a variety of genomics, protein sequence databases and data mining, data manipulation and analysis softwares. Analysis of Microarray Data Using Excel (using tutorial in Bioinformatics and Functional Genomics textbook by Pevsner).</p>	
<p>Criterion 1. b. i. Spreadsheet analysis, systems analysis and design, and decision support systems.</p>	<p>Criteion 1. b. i Select the parameters for pairwise (two sequence) local and global alignment for BLAST program. Interpret and evaluate the bit scores and E-values for nucleotide and protein sequence homology and identity analysis.</p>	<p>Criterion 1.b. i MCCCD Official Course Competencies 10. Use a spreadsheet to perform statistical analysis and create graphs of bioinformatic data. (VII) MCCCD Official Course Outline VII. Scientific Computing C. Spreadsheets 1. Bioinformatic data computation and statistical analysis 2. Editing rows and columns, entering calculation formulas 3. Converting microarray or other bioinformatics data to graphs III. Searching for Similar Sequences A. Homology, similarity, and identity concepts B. The BLAST family of programs for nucleotides and proteins 1. Choosing parameters 2. Masking 3. Bit scores and E-values 4. Formatting output</p>

		<p>5. Evaluating results</p> <p>V. Selection, Alignment and Comparison of Multiple Sequences</p> <p>A. Selecting Sequences to align</p> <ol style="list-style-type: none">1. How many to align at once2. How close in identity3. Evaluating quality of and interpreting alignments <p>B. Clustal family of alignment programs</p> <ol style="list-style-type: none">1. Progressive alignment is influenced by order of sequence entry2. Parameter refinements<ol style="list-style-type: none">a. Substitution matricesb. Gap-opening penaltiesc. Gap-extension penalties <p>C. Other alignment programs</p> <p>Module 4: Molecular Sequence Manipulation and Pattern Identification</p> <p>Lecture: Introduction to Genomics, Proteomics, Genetic Diseases and OMIM</p> <p>Chapter 2: Data Searches and Pairwise Alignments (Fundamental Concepts of Bioinformatics by Krane and Raymer).</p> <p>Dot Plots Simple Alignments Gaps: Simple gap penalties, Origination and length penalties</p> <p>Dynamic programming: The Needleman and Wunsch Algorithm</p> <p>Global and Local Alignments: The Smith-Waterman Algorithm</p> <p>Alignment scores and statistical significance of database searches</p> <p>Lab Exercise 7: Genomics exercise Programming in SequinR for statistical analysis and graph creation.</p> <p>•Exercise on HIV-1 genome www.ncbi.nlm.nih.gov/entrez/</p> <p>Lab Exercise 8: Genomics Exercise contd. •Genomics of microbes from the environment at the Department of Energy (DOE) http://img.jgi.doe.gov/ •Exploring the human genome www.ensembl.org or http://www.ensembl.org/index.html</p>
--	--	---

<p>ii. Graphic/artistic design using computers.</p>	<p>Criterion 1. b. ii Students will design and develop a website for presenting their term projects.</p>	<p>Criterion 1. b. ii MCCCD Official Course Competencies 9. Use a word processor to create bioinformatic reports and forms. (VII) MCCCD Official Course Outline VII. Scientific Computing B. Word Processing 1. Creating scientific reports a. Text layouts and headings b. Tables c. Charts or graphs d. Diagrams 2. Creating forms for scientific or lab use E. Web Page Development and Searching 1. Effective scientific web page design 2. Uploading web pages to a server 3. Using Internet search engines to locate scientific information 4. Evaluating integrity of websites</p>
<p>iv. Modeling, making extensive use of computer simulation.</p>	<p>Criterion 1. b.iv Students build Phylogenetic Trees (phylograms and cladograms) using PHYLIP, MEGA and ClustalW software Students use ClustalW2 program for multiple nucleotide and protein sequence alignments for finding conserved motifs and domains in various genes such as disease related genes, signal transduction genes, regulatory genes controlling gene expression.</p>	<p>Criterion 1. b. iv MCCCD Official Course Competencies 6. Select appropriate multiple sequences, align them, and evaluate the results. (V) 7. Construct various types of phylogenetic trees from multiple aligned sequences. (V) MCCCD Official Course Outline V. Selection, Alignment and Comparison of Multiple Sequences D. Phylogenetic Trees for Comparison of Multiple Sequences 1. Types of Trees a. Binary Trees b. Consensus Trees c. Phylograms 2. Construction Algorithms a. Maximum Parsimony b. Neighbor Joining c. Maximum Likelihood 3. Validation using Bootstrapping Module 5: Phylogenetics Lecture: Different methods of studying phylogenetics.</p>

	<p>Criterion 1. b. iv</p> <p>Students use the following sequence analysis and manipulation suite for characterizing nucleotide and protein sequences.</p> <p>www.bioinformatics.org/sms2/</p> <p>Biology Workbench 3.2 (San Diego Supercomputer Center)</p> <p>Primer3Plus software for Designing Primers for Polymerase Chain Reactions (PCR)</p> <p>OligoAnalyzer software for primer quality analysis.</p>	<p>Chapter 4: Distance Based Methods of Phylogenetics</p> <p>Chapter 5: Character Based Methods of Phylogenetics</p> <p>(Fundamental Concepts of Bioinformatics by Krane and Raymer)</p> <p>Lab Exercise 9:</p> <p>Building Phylogenetic Trees using PHYLIP, MEGA and ClustalW software http://evolution.genetics.washington.edu/phylip.html www.ebi.ac.uk/clustalw http://megasoftware.net/</p> <p>Criterion 1. b. iv</p> <p>MCCCD Official Course Competencies</p> <p>3. Use software to locate open reading frames. (II)</p> <p>MCCCD Official Course Outline</p> <p>II. Analyzing One Sequence</p> <p>A. Nucleotide Sequences</p> <ol style="list-style-type: none"> 1. Detecting Vector Contamination 2. Restriction Maps 3. Analyzing DNA/RNA composition <ol style="list-style-type: none"> a. Basic statistics and probability b. Counting words and repeats c. GC content: percentage of Guanine and Cytosine nucleotides d. Other analyses e. Analysis software 4. Finding Protein Coding Regions <ol style="list-style-type: none"> a. Open Reading Frames b. Exons c. Assembly <p>B. Protein Sequences</p> <ol style="list-style-type: none"> 1. Predicting physical properties 2. Predicting structure 3. Finding domains <p>Module 7: Genomics and Proteomics</p> <p>Lectures:</p> <p>Genomics: Genome Annotation and Gene Expression</p> <p>Proteomics: Predicting Primary and Secondary Structure and Function of the Protein Sequences</p> <p>Chapter 6. Genomics and Gene Regulation</p>
--	---	--

<p>v. Statistics studies stressing the use of computer software.</p>	<p>Criterion 1.b. v</p> <p>Download and install MAGIC Tool Software for DNA Microarray Data Analysis.</p> <p>MAGIC Tool online tutorial for Microarray gene expression analysis.</p>	<p>Chapter7. Protein and RNA Structure Prediction</p> <p>Fundamental Concepts of Bioinformatics by Krane and Raymer.</p> <p>Lab Exercise 10:</p> <p>Proteomics: Searching for domains in your protein sequence Predicting structure and physical properties of your protein sequence Predicting protease digestion patterns Primary Protein Structure Analysis Searching for Transmembrane Segments</p> <p>Lab Exercises 11 and 12:</p> <p>Proteomics Modeling Exercise: Predicting and Interpreting Post-Translational Modification Patterns of A Protein Predicting Secondary and 3D Structures of A Protein http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml http://bioinf.cs.ucl.ac.uk/psipred/ http://www.rcsb.org/pdb/home/home.do www.rasmol.org (RasMol) http://spdbv.vital-it.ch/index.html (DeepView, Swiss PdbViewer)</p> <p>Crterion 1. b. v MCCCD Official Course Competencies</p> <p>8. Perform normalization and statistical analysis on microarray gene expression data. (VI)</p> <p>MCCCD Official Course Outline</p> <p>VI. Microarrays for Gene Expression Analysis A. Target selection and experimental design B. Image analysis C. Data validation and normalization D. Statistical analysis</p> <p>Module 8: Gene Expression and Microarray Data Analysis</p> <p>Lecture:</p> <p>Chapter 6: Bioinformatic Approach to Gene Expression Chapter 7: Gene Expression and Microarray Data Analysis</p>
--	--	---

		<p>(From Bioinformatics and Functional Genomics by Jonathan Pevsner)</p> <p>Lab Exercise 13: DNA Microarray Data Analysis. MAGIC Tool online tutorial for Microarray gene expression analysis.</p> <p>Lab Exercise 14: Primer design laboratory exercise using Primer3Plus software. Students choose their own DNA sequence from NCBI Nucleotide Database.</p> <p>Primer quality analysis using OligoAnalyzer software.</p>

Official Course Description: MCCC CD Approval: 6-28-2011

BIO283 2012 Summer I - 9999

L+L 3.0 Credit(s) 4.0 Period(s) 3.7 Load
Acad

Bioinformatics and Scientific Computing

Introduction to Bioinformatics, including history, concepts, major genetic databases and access tools. Computer software and techniques for analyzing one nucleotide or protein sequence, searching for similar sequences, and aligning and comparing two or multiple sequences. Microarray analysis and phylogenetic trees. Application of standard software to bioinformatic computing tasks, including word processing of reports, and use of spreadsheets for statistical analysis and graphing. Text editors, Unix, Internet web site searching and construction, and ethics.

Prerequisites: [(BIO156 or BIO181) and (MAT120 or MAT121 or MAT122)], or permission of Instructor. Concurrent enrollment in, or previous completion of, BIO208 or BIO212AA is strongly suggested but not required.

Cross-References: [CSC283](#)

[Go to Competencies](#) [Go to Outline](#)

MCCC CD Official Course Competencies:

BIO283 2012 Summer I - 9999

Bioinformatics and Scientific Computing

1. Identify major genetic databases and access tools and describe special features and strengths of each. (I)
2. Use software to perform basic statistical analyses and characterize sequences. (II)
3. Use software to locate open reading frames. (II)
4. Use available software to search genetic databases for similar sequences. (III)
5. Use available software to align sequences and determine similarities. (III, IV)
6. Select appropriate multiple sequences, align them, and evaluate the results. (V)
7. Construct various types of phylogenetic trees from multiple aligned sequences. (V)
8. Perform normalization and statistical analysis on microarray gene expression data. (VI)
9. Use a word processor to create bioinformatic reports and forms. (VII)
10. Use a spreadsheet to perform statistical analysis and create graphs of bioinformatic data. (VII)

[Go to Description](#) [Go to top of Competencies](#)

MCCC CD Official Course Outline:

BIO283 2012 Summer I - 9999

Bioinformatics and Scientific Computing

I. Introduction

- A. What are Bioinformatics and Scientific Computing?
- B. History
- C. Definitions and Concepts
 - 1. Review of molecular biology
 - a. Frequencies and abbreviations for amino acids
 - b. Reading protein and DNA sequences correctly
 - 2. Computer applications to molecular biology
- D. Uses of Bioinformatics
- E. Levels of Focus
 - 1. Single sequences or genes
 - 2. Domains and families
 - 3. Metabolic pathways
 - 4. Genomes
- F. Major Genetic Databases and Access Tools
 - 1. National Center for Biotechnology Information (NCBI)
 - a. GenBank database
 - i. Organisms and types of data stored
 - ii. Organization of the data, accession numbers
 - iii. Understanding a GenBank entry
 - b. Entrez to search across databases
 - c. PubMed literature research
 - i. Basic searching
 - ii. Searching using fields
 - iii. Searching using limits
 - iv. MeSH terminology
 - d. BLAST searching programs for DNA and protein sequences
 - e. Online Mendelian Inheritance in Man
 - f. Other NCBI databases and tools
 - 2. Other Nucleotide Sequence Databases and Software
 - a. Additional NCBI databases and software
 - b. Internet sites around the world
 - 3. Other Protein Databases and Software
 - a. Additional NCBI databases and software
 - b. Internet sites around the world

II. Analyzing One Sequence

A. Nucleotide Sequences

- 1. Detecting Vector Contamination
- 2. Restriction Maps
- 3. Analyzing DNA/RNA composition
 - a. Basic statistics and probability
 - b. Counting words and repeats
 - c. GC content: percentage of Guanine and Cytosine nucleotides
 - d. Other analyses
 - e. Analysis software
- 4. Finding Protein Coding Regions
 - a. Open Reading Frames
 - b. Exons

- c. Assembly
- B. Protein Sequences
 - 1. Predicting physical properties
 - 2. Predicting structure
 - 3. Finding domains
- III. Searching for Similar Sequences
 - A. Homology, similarity, and identity concepts
 - B. The BLAST family of programs for nucleotides and proteins
 - 1. Choosing parameters
 - 2. Masking
 - 3. Bit scores and E-values
 - 4. Formatting output
 - 5. Evaluating results
- IV. Alignment and Comparison of Two Sequences
 - A. Issues
 - 1. Sequence lengths
 - 2. Scoring algorithms
 - 3. Manual evaluation and tweaking
 - B. Dot Plots
 - C. Local and Global Alignments
- V. Selection, Alignment and Comparison of Multiple Sequences
 - A. Selecting Sequences to align
 - 1. How many to align at once
 - 2. How close in identity
 - 3. Evaluating quality of and interpreting alignments
 - B. Clustal family of alignment programs
 - 1. Progressive alignment is influenced by order of sequence entry
 - 2. Parameter refinements
 - a. Substitution matrices
 - b. Gap-opening penalties
 - c. Gap-extension penalties
 - C. Other alignment programs
 - D. Phylogenetic Trees for Comparison of Multiple Sequences
 - 1. Types of Trees
 - a. Binary Trees
 - b. Consensus Trees
 - c. Phylograms
 - 2. Construction Algorithms
 - a. Maximum Parsimony
 - b. Neighbor Joining
 - c. Maximum Likelihood
 - 3. Validation using Bootstrapping
- VI. Microarrays for Gene Expression Analysis
 - A. Target selection and experimental design
 - B. Image analysis
 - C. Data validation and normalization
 - D. Statistical analysis

VII. Scientific Computing

A. Text Editors

1. Creating, editing, saving, and retrieving data files
2. Converting data files to different file formats
 - a. Operating system file formats
 - b. FASTA (Fast-All) and other sequence file formats

B. Word Processing

1. Creating scientific reports
 - a. Text layouts and headings
 - b. Tables
 - c. Charts or graphs
 - d. Diagrams
2. Creating forms for scientific or lab use

C. Spreadsheets

1. Bioinformatic data computation and statistical analysis
2. Editing rows and columns, entering calculation formulas
3. Converting microarray or other bioinformatics data to graphs

D. Databases

1. Searching and retrieval in major online databases
2. Organizing lab data for local storage and retrieval
3. Organizing lab data for submission to online databases

E. Web Page Development and Searching

1. Effective scientific web page design
2. Uploading web pages to a server
3. Using Internet search engines to locate scientific information
4. Evaluating integrity of websites

F. Electronic Mail

1. Effective use of email
2. Professional etiquette
3. Attaching files to emails

G. Privacy, Safety, and Security

1. Downloading files - size issues
2. Privacy and security issues
3. Virus dangers

H. Introduction to Unix/Linux

1. Unix file system commands
2. Uploading and downloading files
3. Running local and/or remote bioinformatics programs
4. Introduction to scripting

I. Ethics

1. Copyrights and infringement
2. Professional responsibility and integrity
3. Obligation to protect data

[Go to Description](#) [Go to top of Competencies](#) [Go to top of Outline](#)

BIOINFORMATICS AND SCIENTIFIC COMPUTING

BIO 283/CSC 283 Course Syllabus

Glendale Community College

Fall Semester 2012

COURSE SCHEDULE

BIO 283 Lecture and Lab Section: 10372

CSC 283 Lecture and Lab Section: 34148

Class Times: Tuesday and Thursday 1:00 PM -2:15 PM

Location: High Tech Center 2 Room 159 (HT2 159)

INSTRUCTOR

Sagarika Dash, Ph.D.

sagarika.dash@gccaz.edu

Phone: 623 845 3635

Office: Life Science Building Second Floor Room 223 (LS 223)

Office hours:

MW 12:50 – 1:50 PM

TTH 2:20 – 3:20 PM

Friday 10:30 – 11:30 AM or by appointment.

COURSE DESCRIPTION

Introduction to Bioinformatics, including history, concepts, major genetic databases and access tools. Computer software and techniques for analyzing nucleotide or protein sequences, searching for similar sequences, and aligning and comparing two or multiple sequences.

Microarray analysis and phylogenetic trees. Application of standard software to bioinformatic computing tasks, including word processing of reports, and use of spreadsheets for statistical analysis and graphing. Text editors, UNIX, Internet web site searching and construction, and ethics. Prerequisites: [(BIO156 or BIO181) and (MAT120 or MAT121 or MAT122)], or permission of Instructor. Concurrent enrollment in, or previous completion of, BIO208 or BIO212AA is strongly suggested but not required.

COURSE OBJECTIVES

This is a hands-on course in bioinformatics for learning how to use computers and the web, as tools to analyze and represent large collections of biological sequence and structure data. The main topics include the following:

1. Exploration of biological databases, retrieval and analyses of nucleotide and protein sequences, alignments and comparisons nucleotide and protein sequences.
2. Programming in Perl language for developing customized bioinformatics tools.
3. Introduction to programming in R for statistical data analysis and graphic presentation.
4. Introduction to genomics and proteomics, gene annotation and prediction of protein structures and functions
5. Building phylogenetic trees.

TEXT

Fundamental Concepts of Bioinformatics, by Dan E. Krane and Michael Raymer, Published by Benjamin Cummings

Bioinformatics and Functional Genomics, 2nd Edition <http://www.bioinfbook.org/> , Jonathan Pevsner, Published by John Wiley and Sons Inc.

Recommended textbooks:

Beginning Perl for Bioinformatics, by James D. Tisdall, Published by O'Reilly Media, Inc.

Web based tutorials for introduction to programming languages such as Perl, Python and UNIX shell scripting will be posted on Blackboard. Additional readings will be assigned throughout the course.

Online tutorial for Programming in R by Thomas Girke, UC Riverside

EQUIPMENT

PC running Microsoft Windows, Internet Connection and browser, freely downloadable programming software are needed for the course. You may use your own computer or the computers provided for you on the eGCC Network in the High Tech Centers on campus.

The course materials, announcements, assignments and grades will be posted on Blackboard.

TESTING AND GRADING

You will be doing a series of homework assignments, quizzes and discussions based on class lectures, online tutorials and textbook exercises. The homework assignments and due dates will be posted in the assignment folder on Blackboard. You will complete a term project on a genetic-based human disease of your own selection. **The term project requires a write up (50 points) and in-class presentation (50 points).**

5 Homework Assignments (30%)	30 points each	150 points
5 In-class Quizzes (10%)	10 points each	50 points
Midterm exam (20%)	100 points	100 points
Term project (20%)	100 points	100 points
Final exam (20%)	100 points	100 points

Total Points		500 points
--------------	--	------------

The final exam for the course is on Tuesday, December 11, 2012 at 1:00 PM.

Final letter grades will be assigned based on the following grade scale:

A	450 - 500 points
B	400 - 449 points
C	350 - 399 points
D	300 - 349 points
F	<299 points

EXTRA CREDIT

Students can earn a **total of 20 extra credit points** in this course for developing a bioinformatics project, implementing web based bioinformatics tools and databases or writing program codes in Perl, Unix or Python programming languages to solve the problem.

POSTING OF EXAM SCORES

The points scored in exams, quizzes, homework and other assignments will be posted on Blackboard.

CHALLENGE PERIOD

Your graded exams, quizzes and assignments will be given back to you in a timely manner. If you have any questions or clarifications about your scores please discuss with me within a week of getting the graded materials. At end of this syllabus you have a spread sheet to record you score in exams, lab quizzes and assignments. Keep track of scores for individual items and grade carefully, record them on the spread sheet. If you notice any missing grade or any mistake in posting the correct score on Blackboard, let me know immediately. After a two-week period, the scores will not be reviewed further.

MAKE UP EXAM POLICY

If you miss an exam due to unavoidable circumstances, contact me within next two days via email or phone to arrange for a makeup exam. **I will leave the makeup exam at the testing center by the next class meeting time and the makeup exam should be taken within the specified deadline (one week from the date of the exam).**

ASSIGNMENT DUE DATES

All class assignments, including homework, have a definite due date. If assignments are turned in late, the points will be reduced by 25 percent for each class day late.

INCOMPLETE GRADE

Incomplete grade (I) is only given to a student doing acceptable work (C or better) who is unable to complete the course requirements due to illness or extenuating circumstances, and who only needs to make up the last lecture and/or lab exam to complete the requirements. An **“Incomplete grade contract”** must be filed with the Records office. It is the student’s responsibility to initiate and file the incomplete grade contract.

COURSE POLICIES

Attendance: Only students enrolled in the course are allowed to attend the lectures and laboratories for the course. I take roll at the beginning of each class. In case of official absence (a college sponsored activity) submit the college absence verification card to the instructor before the absence. **More than 3 absences will lead to a drop from the course.**

Withdrawal Policy: Withdrawal from the course is student’s responsibility. Please check GCC Fall 2012 class calendar for specific deadlines. You are guaranteed to receive a **grade of ‘W’ if you withdraw by the deadline** for student initiated withdrawal without instructor’s signature (see GCC calendar). After that deadline, you will receive a ‘W’ if you have passing grade on the date of withdrawal or a ‘Y’ if your total points in the course is <60%.

Disability: If you have any disability and need special accommodation, please inform the instructor and the Disability Services and Resources office (Phone: 623 845 3080).

Disruptive Behavior: Disruptive behavior will not be tolerated in class. The student causing the disturbance will be asked to leave the class (will be marked absent for that class period) and will be asked to talk to the Dean of Students before returning to class.

Electronic Devices: Cell phones and pagers must be put on vibration mode inside the classroom and lab. In case of an emergency call, please leave the classroom to use your phone.

Laptop computers with wireless internet are allowed in the classroom but must be used only for academic purposes relevant to the course content.

Academic Grievances: If you have any academic grievances first discuss with your instructor for a resolution of the matter. If you are not satisfied with the solution provided by your instructor, then speak with the department chair. Finally, if you feel the problem is still not resolved discuss it with the Dean of Instruction.

Academic Dishonesty: Cheating and plagiarism in any form will not be tolerated. According to **GCC Student Handbook** cheating or plagiarizing will lead to suspension or expulsion of the concerned student from GCC.

Tentative Schedule of Topics*:

Weeks	Lectures (Monday class periods)
Week 1 Aug. 21	Introduction to Bioinformatics and Practical Applications of Bioinformatics Databases and Analysis Tools
Week2 Aug. 28	Introduction to Biological Databases (NCBI, GenBank, Entrez, PubMed, MeSH)
Week3 Sept. 4	Introduction to Bioinformatics Algorithms and Perl Programming. Variables, functions, regular expressions, pattern matching, data structures, file handling
Week 4 Sept. 11	Introduction to Genomics, Genetic Diseases and OMIM Data Searches and Pairwise Alignments: Simple Alignments Gaps: Simple gap penalties, Origination and length penalties Dynamic programming: The Needleman and Wunsch Algorithm Global and Local Alignments: The Smith-Waterman Algorithm Alignment scores and statistical significance of database searches
Week 5 Sept. 18	Lecture on Wednesday: Introduction to programming in R. Accessing Literature Information from PubMed <ul style="list-style-type: none"> • How to Use Nucleotide and Protein Sequence Databases • Retrieve sequences, save the sequences in FASTA format • Pairwise Sequence Alignment using BLAST (Basic Local Alignment Search Tool)
Week 6 Sept. 25	Introduction to UniProtKB/Swiss-Prot of Swiss Institute of Bioinformatics (SIB) and European Bioinformatics Institute (EBI)
Week 7 Oct. 2	Multiple Sequence Alignment: ClustalW2, Geneious, T-Coffee and MUSCLE

	Midterm exam review.
Week 8 Oct. 9	Midterm exam on Tuesday. Term project outline due.
Week 9 Oct. 16	Building Phylogenetic Trees
Week 10 Oct. 23	Introduction to Genomics Genome annotation Bioinformatics Approaches to Gene Expression
Week 11 Oct. 30	Introduction to Proteomics
Week 12 Nov. 6	Analysis of Protein Sequences Predicting Primary and Secondary Structure and Function Proteins
Week 13 Nov. 13	DNA Microarray Data Analysis
Week 14 Nov. 20	Human Genomics and Diseases
Week 15 Nov. 27	Challenges and Future of Bioinformatics
Week 16 Dec. 4	Final exam review.
Week 17 Dec. 11	Final Exam (Tuesday, December 11, 2012 at 1 PM)

***Disclaimer: The lecture schedule is tentative and can be changed to meet the needs of the group.**

Laboratory Schedule*

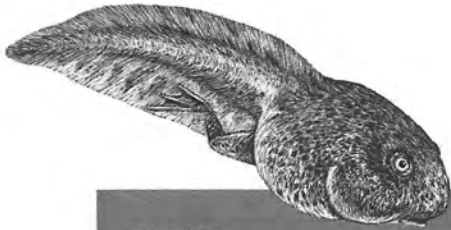
Weeks	Hands-on Laboratory Exercises (Thursday class periods)
Week 1	<p>Exploring NCBI website: www.ncbi.nlm.nih.gov</p> <p>Access literature information on PubMed www.ncbi.nlm.nih.gov/entrez/ (Learn how to use the fields and limits to optimize the literature search for your research interest)</p> <p>Retrieve DNA sequences from GenBank www.ncbi.nlm.nih.gov/entrez/ (learn about EMBL and DDBJ, learn about FASTA and understand the various features of the GenBank record).</p> <p>Install Perl and Introduce Programming in Perl</p>
Week 2	<p>Retrieve Protein sequences (click on http://www.expasy.org/).</p> <p>Exercise on advanced search for protein sequences in http://www.uniprot.org/uniprot/</p> <p>Select Swiss-Prot Box and Deselect TrEMBL box which is a database of</p>

	<p>unsupervised computer translations of new DNA sequences. Save the protein sequence in FASTA format in a word document (save as text only, filename must be *.txt)</p> <p>Retrieve the DNA sequence related to the protein by cross referencing to nucleotide databases (GenBank, EMBL, DDBJ and CoDingSequence link).</p> <p>Save your nucleotide and protein sequences in FASTA format (save as type option text only in a word document, filename should be *.txt) for future use.</p>
Week 3	<ul style="list-style-type: none"> • Introduction to programming and Perl syntax • Run BLAST (Basic Local Alignment Search Tool) for comparison of nucleotide sequence with other nucleotide sequences and translated sequences. • Learn to use BLAST to compare your protein sequence with other protein sequences. • Hands- on practice with nucleotide and protein sequence comparisons using BLAST http://blast.ncbi.nlm.nih.gov/Blast.cgi
Week 4	<p>Demos of Perl programming. Writing programs in Perl. Hands-on perl programming: User input and output, file handling.</p>
Week 5	<p>Download and install SequinR software package for DNA sequence statistics (1) tutorial.</p> <p>Create data tables in Excel and how to import data to R environment and use R functions for statistical analysis of bioinformatics data and create graphs.</p>
Week 6	<p>Multiple Protein Sequence Alignment using ClustalW2 http://pir.georgetown.edu/pirwww/search/multialn.shtml</p> <p>Students use the sequence manipulation suite to analyze and manipulate their nucleotide and protein sequences. www.bioinformatics.org/sms2/</p>
Week 7	<p>Genomics exercise Programming in SequinR software for statistical analysis and graph creation.</p> <ul style="list-style-type: none"> • Exercise on HIV-1 genome www.ncbi.nlm.nih.gov/entrez/
Week 8	<p>Genomics Exercise contd.</p> <ul style="list-style-type: none"> • Genomics of microbes from the environment at the Department of Energy (DOE) http://img.igi.doe.gov/

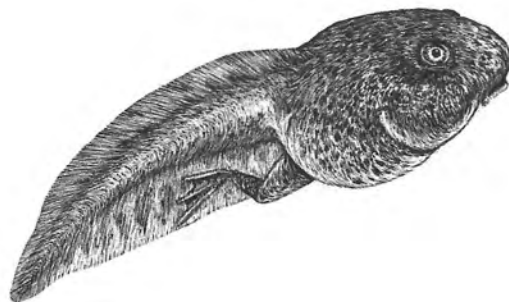
	<ul style="list-style-type: none"> Exploring the human genome www.ensembl.org or http://www.ensembl.org/index.html
Week 9	Building Phylogenetic Trees using PHYLIP, MEGA and ClustalW software http://evolution.genetics.washington.edu/phylip.html www.ebi.ac.uk/clustalw http://megasoftware.net/
Week 10	Proteomics Exercise Searching for domains in your protein sequence Predicting structure and physical properties of your protein sequence Predicting protease digestion patterns Primary Protein Structure Analysis Searching for Transmembrane Segments
Week 11	Proteomics Modeling Exercise: Predicting and Interpreting Post-Translational Modification Patterns of A Protein Predicting Secondary and 3D Structures of A Protein http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml http://bioinf.cs.ucl.ac.uk/psipred/ http://www.rcsb.org/pdb/home/home.do www.rasmol.org (RasMol) http://spdbv.vital-it.ch/index.html (DeepView, Swiss PdbViewer)
Week 12	Proteomics Modeling Exercise contd. Modeling secondary, tertiary and quaternary structures of proteins and predicting their functions.
Week 13	DNA Microarray Data Analysis. MAGIC Tool online tutorial for Microarray gene expression analysis.
Week 14	Primer design laboratory exercise using Primer3Plus software. Students choose their own DNA sequence from NCBI Nucleotide Database. Primer quality analysis using OligoAnalyzer software. Term project presentations.
Week 15	Term project presentations.

***Disclaimer: This laboratory schedule is tentative and may be changed by the instructor to meet the needs of the group.**

An Introduction to Perl for Biologists



Beginning Perl for Bioinformatics



O'REILLY®

James Tisdall

Beginning Perl for Bioinformatics



With its highly developed capacity to detect patterns in data, Perl has become one of the most popular languages for biological data analysis. But many biologists have a difficult time learning how to apply the language to bioinformatics. The most popular Perl programming books are often too theoretical and too focused on computer science for a nonprogramming biologist who needs to solve very specific problems.

Beginning Perl for Bioinformatics is a practical introduction to Perl designed for biologists with little or no programming experience. It approaches programming as an important new laboratory skill and features many Perl programs and programming techniques that can be immediately useful in the lab. Select chapters focus on a problem or class of problems in bioinformatics and show how to use Perl to solve them. The problems, and the Perl, increase in difficulty as the chapters progress. Programming exercises are included for most chapters. By the end of the book you will have a solid understanding of Perl basics, as well as a collection of programs for tackling such tasks as parsing BLAST and GenBank.

If you've been wanting to learn how to solve programming problems in bioinformatics using Perl, this book has everything you need to get started.

The book covers:

- Programming basics and working with DNA sequences and strings
- Debugging your code
- Simulating gene mutations using random number generators
- Regular expressions and finding motifs in data
- Arrays, hashes, and relational databases
- Regular expressions and restriction maps
- Using Perl to parse PDB records, annotations in GenBank, and BLAST output

www.oreilly.com

US \$39.95

CAN \$59.95

ISBN-10: 0-596-00080-4

ISBN-13: 978-0-596-00080-6



9 780596 000806

Beginning Perl for Bioinformatics

James D. Tisdall

O'REILLY®
Beijing • Cambridge • Farnham • Köln • Sebastopol • Tokyo

Beginning Perl for Bioinformatics

by James D. Tisdall

Copyright © 2001 James D. Tisdall All rights reserved.
Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North,
Sebastopol, CA 95472.

O'Reilly Media, Inc. books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (*safari.oreilly.com*). For more information contact our corporate/institutional sales department: (800) 998-9938 or *corporate@oreilly.com*.

Editor: Lorrie LeJeune
Production Editor: Mary Anne Weeks Mayo
Cover Designer: Ellie Volckhausen
Interior Designer: Melanie Wang

Printing History:

October 2001: First Edition.

Nutshell Handbook, the Nutshell Handbook logo, and the O'Reilly logo are registered trademarks of O'Reilly Media, Inc. *Beginning Perl for Bioinformatics*, the image of a tadpole, and related trade dress are trademarks of O'Reilly Media, Inc. Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks. Where those designations appear in this book, and O'Reilly Media, Inc. was aware of a trademark claim, the designations have been printed in caps or initial caps.

While every precaution has been taken in the preparation of this book, the publisher assumes no responsibility for errors or omissions, or for damages resulting from the use of the information contained herein.

ISBN: 978-0-596-0080-6

[LSI]

[2011-05-2]

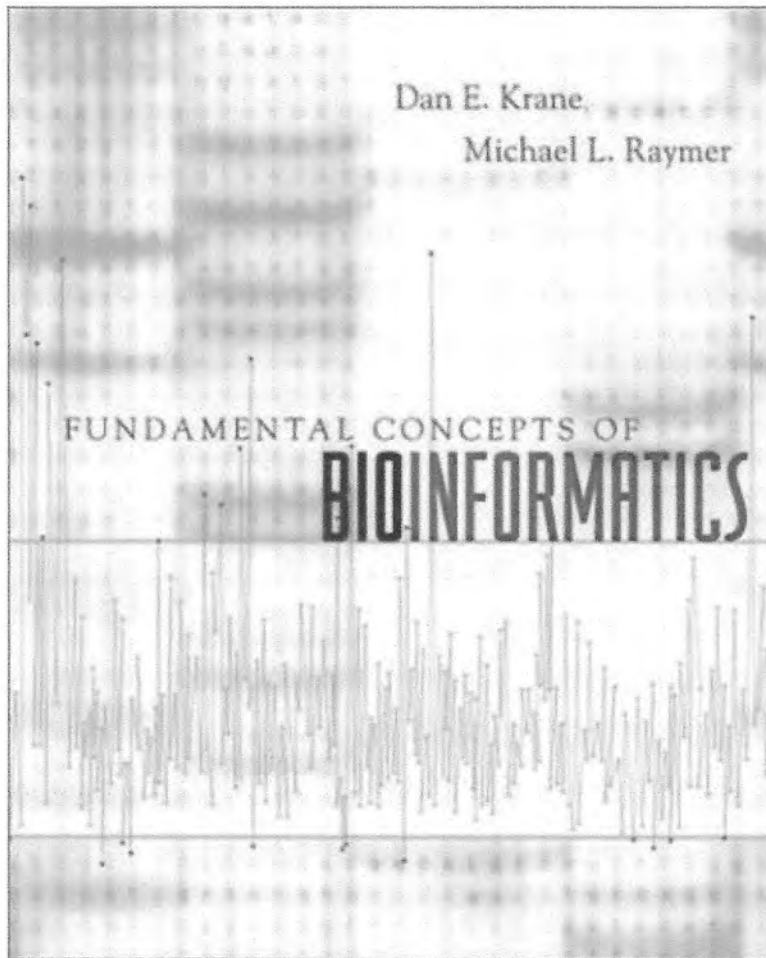
Table of Contents

Preface	vii
1. Biology and Computer Science	1
The Organization of DNA	2
The Organization of Proteins	3
In Silico	4
Limits to Computation	5
2. Getting Started with Perl	6
A Low and Long Learning Curve	6
Perl's Benefits	8
Installing Perl on Your Computer	10
How to Run Perl Programs	13
Text Editors	15
Finding Help	16
3. The Art of Programming	18
Individual Approaches to Programming	18
Edit—Run—Revise (and Save)	19
An Environment of Programs	21
Programming Strategies	22
The Programming Process	23
4. Sequences and Strings	29
Representing Sequence Data	29
A Program to Store a DNA Sequence	32
Concatenating DNA Fragments	36
Transcription: DNA to RNA	40
Using the Perl Documentation	42

Calculating the Reverse Complement in Perl	43
Proteins, Files, and Arrays	46
Reading Proteins in Files	47
Arrays	50
Scalar and List Context	54
Exercises	55
5. Motifs and Loops	56
Flow Control	56
Code Layout	62
Finding Motifs	63
Counting Nucleotides	70
Exploding Strings into Arrays	71
Operating on Strings	77
Writing to Files	81
Exercises	85
6. Subroutines and Bugs	87
Subroutines	87
Scoping and Subroutines	90
Command-Line Arguments and Arrays	96
Passing Data to Subroutines	98
Modules and Libraries of Subroutines	102
Fixing Bugs in Your Code	104
Exercises	116
7. Mutations and Randomization	118
Random Number Generators	119
A Program Using Randomization	120
A Program to Simulate DNA Mutation	126
Generating Random DNA	136
Analyzing DNA	141
Exercises	147
8. The Genetic Code	149
Hashes	149
Data Structures and Algorithms for Biology	150
The Genetic Code	155
Translating DNA into Proteins	163

Reading DNA from Files in FASTA Format	166
Reading Frames	175
Exercises	180
9. Restriction Maps and Regular Expressions	182
Regular Expressions	182
Restriction Maps and Restriction Enzymes	184
Perl Operations	197
Exercises	198
10. GenBank	199
GenBank Files	200
GenBank Libraries	203
Separating Sequence and Annotation	205
Parsing Annotations	212
Indexing GenBank with DBM	232
Exercises	236
11. Protein Data Bank	238
Files and Folders	240
PDB Files	248
Parsing PDB Files	257
Controlling Other Programs	267
Exercises	272
12. BLAST	274
Obtaining BLAST	275
String Matching and Homology	276
BLAST Output Files	277
Parsing BLAST Output	280
Presenting Data	290
Bioperl	294
Exercises	301
13. Further Topics	302
The Art of Program Design	302
Web Programming	302
Algorithms and Sequence Alignment	303
Object-Oriented Programming	303
Perl Modules	303

Complex Data Structures	304
Relational Databases	304
Microarrays and XML	305
Graphics Programming	305
Modeling Networks	305
DNA Computers	306
A. Resources	307
B. Perl Summary	315
Index	347



About Fundamental Concepts of Bioinformatics

Dan E. Krane, Wright State University

Michael L. Raymer, Wright State University

ISBN: [0-8053-4722-4](https://www.isbn-international.org/product/9780805347224)

Table of Contents

I. MOLECULAR BIOLOGY AND BIOLOGICAL CHEMISTRY

The genetic material

Nucleotides

Orientation

.....

Base pairing

The central dogma of molecular biology

Gene structure and information content

Promoter sequences

The genetic code

Open reading frames

Introns and exons

Protein structure and function

Primary structure

Secondary, tertiary and quaternary structure

The nature of chemical bonds

Anatomy of an atom

Valence

Electronegativity

Hydrophilicity and hydrophobicity

Molecular biology tools

Restriction enzymes

Gel electrophoresis

Blotting, hybridization and microarrays

Cloning

Polymerase chain reaction (PCR)

DNA sequencing

Genomic information content

C value paradox

Reassociation kinetics

II. DATA SEARCHES AND PAIRWISE ALIGNMENTS

Dot plots

Simple alignments

Scoring

Gaps

Simple gap penalties

Origination and length penalties

Scoring matrices

Dynamic programming: The Needleman and Wunsch algorithm

Local and global alignments

Global and Semi-global alignments

The Smith-Waterman algorithm

Database searches

BLAST and its relatives

Other algorithms

Multiple sequence alignments

III. SUBSTITUTION PATTERNS

Patterns of substitutions within genes

Mutation rates

Functional constraint

Synonymous vs. nonsynonymous changes

Indels and pseudogenes

Substitutions vs. mutations

Fixation

Estimating substitution numbers

Jukes/Cantor model

Transitions and transversions

Kimura's two-parameter model

Models with even more parameters

Substitutions between protein sequences

Variations in substitution rates between genes

Molecular clocks

Relative rate tests

Causes of rate variation in lineages

Evolution in organelles

IV. DISTANCE-BASED METHODS OF PHYLOGENETICS

History of molecular phylogenetics

Advantages to molecular phylogenies

Phylogenetic trees

Terminology of tree reconstruction

Rooted and unrooted trees

Gene vs. species trees

Character and distance data

Distance matrix methods

UPGMA

Estimation of branch lengths

Transformed distance method

Neighbor's relation method

Neighbor-joining methods

Maximum likelihood approaches

Multiple sequence alignments

V. CHARACTER-BASED APPROACHES TO PHYLOGENETICS

Parsimony

Informative and uninformative sites

Unweighted parsimony

Weighted parsimony

Inferred ancestral sequences

Strategies for faster searches

Branch and bound

Heuristic

Consensus trees

Tree confidence

Bootstrapping

Parametric tests

Comparison of phylogenetic methods

Molecular phylogenies

The tree of life

Human origins

VI. GENOMICS AND GENE RECOGNITION

Prokaryotic genomes

Prokaryotic gene structure

Promoter elements

Open reading frames

Conceptual translation

Termination sequences

GC-content

Prokaryotic gene density

Eukaryotic genomes

Eukaryotic gene structure

Promoter elements

Regulatory protein binding sites

Open reading frames

Introns and exons

Alternative splicing

CpG islands

GC-content

Isochores

Codon usage bias

Gene expression

cDNAs and ESTs

Serial analysis of gene expression (SAGE)

Microarrays

Transposition

Repetitive elements

Eukaryotic gene density

VII. PROTEIN FOLDING

Polypeptide composition

Amino acids

Backbone flexibility, phi and psi

Secondary structure

Accuracy of predictions

Chou-Fasman/GOR method

Tertiary and quaternary structure

Hydrophobicity

Disulfide bonds

Active structures vs. most stable structures

Protein folding

Lattice models

Off-lattice models

Energy functions and optimization

Structure prediction

Comparative modeling

Threading: Reverse protein folding

Predicting RNA secondary structures

VIII. PROTEOMICS

From genomes to proteomes

Protein classification

Enzyme nomenclature

Families and superfamilies

Folds

Experimental techniques

2D electrophoresis

Mass spectrometry

Protein microarrays

Inhibitors and drug design

Ligand screening

Docking

Database screening

X-ray crystal structures

Empirical methods and prediction techniques

Postranslational modification prediction

Protein sorting

Proteolytic cleavage

Glycosylation

Phosphorylation and sulfation

Appendix 1: A gentle introduction to programming and data structures

Introduction

The basics

Creating and compiling computer programs

Variables and values

Data typing

Basic operations

Program control

Statements and blocks

Conditional execution

Loops

Readability

Structured programming

Comments

Descriptive variable names

Data structures

Arrays

Pointers and dynamic memory allocation

Strings in PERL

Input and output

Appendix 2: Enzyme kinetics

Enzymes as biological catalysts

The Henri-Michaelis-Menten equation

V_{max} and K_m

Direct plot

Lineweaver-Burk reciprocal plot

Eadie-Hofstee plot

Simple inhibition systems

Competitive inhibition

Noncompetitive inhibition

Reversible and irreversible inhibition

Effects of pH and temperature

Appendix 3: Sample programs in PERL and worksets

Conceptual translation

Dot matrix

Relative rate test

UPGMA

Common ancestor

Splice junction recognition

Hydrophobicity calculator

DNA binding domains

Lineweaver-Burk plot