

课程详述

COURSE SPECIFICATION

以下课程信息可能根据实际授课需要或在课程检讨之后产生变动。如对课程有任何疑问，请联系授课教师。

The course information as follows may be subject to change, either during the session because of unforeseen circumstances, or following review of the course at the end of the session. Queries about the course should be directed to the course instructor.

1.	课程名称 Course Title	计算生物学/Computational Biology
2.	授课院系 Originating Department	生物系/ Department of Biology
3.	课程编号 Course Code	BIO309
4.	课程学分 Credit Value	3
5.	课程类别 Course Type	专业核心课 (生物信息专业) Major Core Courses (Bioinformatics Major) 专业选修课 (生物科学、生物技术专业、生物医学工程专业) Major Elective Courses (Bioscience、Biotechnology、Biomedical Engineering Majors)
6.	授课学期 Semester	秋季 Fall
7.	授课语言 Teaching Language	中英双语 English & Chinese
8.	授课教师、所属学系、联系方式 (如属团队授课, 请列明其他授课教师) Instructor(s), Affiliation & Contact (For team teaching, please list all instructors)	靳文菲, 生物系 Dr. Wenei JIN, Department of Biology jinwf@sustech.eud.cn 0755-88018478
9.	实验员/助教、所属学系、联系方式 Tutor/TA(s), Contact	待公布 To be announced
10.	选课人数限额(可不填) Maximum Enrolment (Optional)	

11. 授课方式 Delivery Method	讲授	习题/辅导/讨论	实验/实习	其它(请具体注明)	总学时
	Lectures	Tutorials	Lab/Practical	Other (Please specify)	Total
学时数 Credit Hours	32		32		64
12. 先修课程、其它学习要求 Pre-requisites or Other Academic Requirements					
13. 后续课程、其它学习规划 Courses for which this course is a pre-requisite	BIO306 生物信息学 /Bioinformatics				
14. 其它要求修读本课程的学系 Cross-listing Dept.	None 无				

教学大纲及教学日历 SYLLABUS

15. 教学目标 Course Objectives

计算生物学是利用数学，计算机科学和统计学的方法解决生物学问题的交叉学科。该学科科主要用生物数据开发模型来理解生物医学问题。它是研究遗传学、进化、基因组学、流行病学和系统生物学的一种重要实用工具。该课程是生物信息学专业的必修课，是其它生物学相关专业的选修课。该课程旨在帮助学生理解有关基因组、序列分析、生物数据库、疾病关联基因和进化分析的主要问题。将对现有各种方法进行批判性描述，讨论每种方法的优点和局限性，并使用相关工具进行实践练习。该课程将同时培养学生积极的科学精神、激发他们的科学好奇心。先修课程包括分子生物学入门课程或讲师许可。

Computational biology is an interdisciplinary subject that uses mathematical, computer science and statistical methods to solve biological problems. This subject mainly focuses on using biological data to develop models to understand biomedical issues. It is a practical, hands-on tool to study genetics, evolution, genomics, epidemiology and systems biology. It is a mandatory course for Bioinformatics major, and elective course for all other majors related to biology. It is designed to help students to understand the major issues concerning genomics, analysis of sequence, biological database, disease associated gene and evolution. Various existing methods will be critically described and the strengths and limitations of each will be discussed, with practical assignments utilizing the tools. It is to train students' vigorous Scientific Spirit and inspire their scientific curiosity. Prerequisites include an introductory molecular biology course or permission of the instructor.

16. 预达学习成果 Learning Outcomes

本课程完成后，学生将能够：

- (1) 熟悉计算生物学领域常用数据分析方法。掌握生物信息常用 linux 命令，序列比对原理和方法，数据降维和聚类的原理和方法，进化分析的原理和方法。
- (2) 能够独立对生物序列数据进行基本生物信息分析。
- (3) 对计算生物学产生更浓厚的兴趣。理解计算生物学交叉学科的特点，发展的动力和前景。

With the completion of this course, the students will

- (1) Understand the common data analytical approaches in computational biology. Master the Linux commands commonly used in bioinformatics, the principles and methods of sequence alignment, the principles and methods of data dimensionality reduction and clustering, and the principles and methods of evolutionary analysis.
- (2) Develop the capability of independently analyzing biological sequence data
- (3) Become more interested in computational biology. Understand the characteristics, driving forces and prospects of computational biology.

17. 课程内容及教学日历（如授课语言以英文为主，则课程内容介绍可以用英文；如团队教学或模块教学，教学日历须注明主讲人）

Course Contents (in Parts/Chapters/Sections/Weeks. Please notify name of instructor for course section(s), if this is a team teaching or module course.)

1. Introduction of computational biology and course introduction 学时：2+2/Hours: 2+2

1. 计算生物学概要和课程概要

1.1 The emergence of computational biology/bioinformatics

1.1 计算生物学/生物信息学的出现

1.2 Development of computational biology

1.2 计算生物学的发展

1.3 Major topics in computational biology

1.3 计算生物学的主要话题

1.4 Computational biology and genomics

1.4 计算生物学和基因组学

1.5 Examples of Computational biology applications

1.5 计算生物学应用实例

1.6 Course Introduction: Goals, outline, evaluation/examination and learning guidelines

1.6 课程介绍：目标、大纲、评估/考试和学习建议

2. Basic computational skills (Linux + programming) 学时：2+2/Hours: 2+2

2. 计算机基础（Linux+编程）

2.1 Linux system and Open Source Software: GitHub

2.1 Linux 系统和开源软件：Github

2.2 Terminal and basic Linux operations

2.2 终端和基本 Linux 操作

2.3 Introduction of programming languages for bioinformatics

2.3 生物信息学编程语言简介

2.4 Programming language Perl

2.4 编程语言 Perl

2.5 Programming language Python

2.5 编程语言 python

2.6 R Language Statistics and Drawing

2.6 R 语言统计与绘图

3 Shell Programming and bioinformatics 学时：2+2/Hours: 2+2

3 Shell 编程和生物信息学

3.1 Basic shell commands

3.1 基本 shell 命令

3.2 File system

3.2 文件系统

3.3 Managing data

3.3 管理数据

3.4 Output redirection

3.4 输出重定向

3.5 Software for Linux

3.5 Linux 软件

3.6 Biological data analysis: Modularization and pipeline

3.6 生物数据分析：模块化和流水线化

4 Human Genome and Human genome project (HGP) 学时：2+2/Hours: 2+2

4 人类基因组与人类基因组计划 (HGP)

4.1 The basic information about human Genome

4.1 人类基因组的基本信息

4.2 HGP initiation and sequencing Strategies

4.2 人类基因组计划的启动和测序策略

4.3 Completion of HGP and its findings

4.3 人类基因组计划完成和产生的结果

4.4 HGP greatly promote sequencing technologies and genomic study

4.4 人类基因组计划极大的促进了测序技术和基因组学研究

4.5 Post HGP era: HapMap, ENCODE, 1000 genome, 3D/4D genome

4.5 后基因时代：人类单体型计划，ENCODE, 千人基因组，三维/四位基因组

4.6 Precision medicine and personalized medicine

4.6 精准医学和个性化医学

4.7 Access the human genome database

4.7 人类基因组数据库访问

5 Pairwise sequence alignments 学时：2+2/Hours: 2+2

5 成对序列比对

5.1 Genomes change over time

5.1 基因组随时间变化

5.2 Sequence comparisons

5.2 序列比较

5.3 Dynamic programming alignment

5.3 动态规划算法比对序列

5.3.1 Global alignment (Needleman-Wunsch)

5.3.1 全局序列比对 (Needleman-Wunsch)

5.3.2 Local alignment (Smith-Waterman)

5.3.2 局部序列比对 (Smith-Waterman)

6 Sequence Similarity Searching 学时：2+2/Hours: 2+2

6 序列相似性搜索

6.1 Approximate alignment is fast

6.1 近似比对速度快

6.2 FASTA Algorithm

6.2 FASTA 算法

6.3 BLAST Algorithm

6.3 BLAST 算法

6.4 Advanced Similarity search

6.4 高级相似性搜索

6.5 Interpretation of the BLAST output

6.5 BLAST 输出解释

7 Next generation sequencing and its applications 学时：2+2/Hours: 2+2

7 二代测序测序及其应用

7.1 Next generation sequencing technology

7.1 二代测序测序技术

7.2. Reads mapping (Burrows-Wheeler Algorithm)

7.2 读长比对 (Burrows-Wheeler 算法)

7.3 Genome sequencing

7.3 基因组测序

7.4 RNA sequencing

7.4 RNA 测序

7.5 Epigenomics (ChIP-seq, DNase-seq, MNase-seq, ATAC-seq, BS-seq)

7.5 表观基因组学 (ChIP-seq, DNase-seq, MNase-seq, ATAC-seq, BS-seq)

7.6 3D genome (Hi-C, CHIA-PET and TracLooping)

7.6 三维基因组 (Hi-C, CHIA-PET and TracLooping)

8 Multiple Sequence Alignment 学时：2+2/Hours: 2+2

8 多序列比对

8.1 Multiple alignment versus Pairwise Alignment

8.1 多序列比对和成对序列比对

8.2 Dynamic programming in 3-D

8.2 三维动态规划

8.3 Progressive Alignment (ClustalW)

8.3 渐进式对准 (CLUSTALW)

8.4 Scoring Multiple Alignments

8.4 多序列比对打分

9. Phylogenetics 学时：2+2/Hours: 2+2

9 系统发育

9.1 Basics of phylogeny: Characters, traits, nodes, branches, lineages

9.1 系统发育基础：特征、性状、节、枝、谱系

9.2 Molecular clock

9.2 分子钟

9.3 Modeling sequence evolution

9.3 建模序列演化

9.4 Distances and clustering algorithm: UPGMA and Neighbor Joining (NJ)

9.4 距离和聚类算法：UPGMA 和 NJ

9.5 From sequence alignments to trees: Parsimony methods

9.5 从序列比对到进化树：简约法

9.6 Probability based approach: Maximum likelihood methods

9.6 基于概率的方法：最大似然法

10 Population genetics 学时：2+2/Hours: 2+2

10 群体遗传学

10.1 Basic concepts in population genetics

10.1 群体遗传学的基本概念

10.2 Genetic drift

10.2 遗传漂变

10.2 Hardy-Weinberg Equilibrium (HWE)

10.2 哈代-温伯格平衡 (HWE)

10.4 Deviations of HWE (Wahlund effect)

10.4 HWE 偏差 (Wahlund 效应)

10.5 HWE in diseases prevalent and prediction

10.5 疾病中的 HWE 流行与预测中的应用

10.6 Genome-wide tests of HWE

10.6 全基因组检测哈代-温伯格平衡

11 Population stratification/structure 学时: 2+2/Hours: 2+2

11 群体分层/群体结构

11.1 The major forces shaping population

11.1. 塑造群体的主要因素

11.2 Population substructure

11.2 群体亚结构

11.3 Measure population structure (F-statistics)

11.3 度量人群结构 (F-统计量)

11.4 Approaches for analysis of population structure

11.4 分析群体结构的方法

11.5 Analysis of molecular variance (AMOVA)

11.5 分子生物学方差分析

11.6 Dimensionality reduction

11.6 降维分析

11.7 Model based approaches

11.7 模型化方法

12 Identifying disease associated variants 学时：2+2/Hours: 2+2

12 筛选疾病相关变异

12.1 Linkage analysis for rare disease

12.1 罕见病连锁分析

12.2 Association analysis

12.2 关联分析

12.2.1 Family based association study

12.2.1 基于家系的关联分析

12.2.2 Case control based association study

12.2.2 基于病例对照的关联分析

12.2.3 Association study based on next generation sequencing (NGS)

12.2.3 基于二代测序（NGS）的关联分析

12.3 Challenge in identifying disease associated variants

12.3 筛选疾病相关变异的面临的挑战

13 R language 学时：2+2/Hours: 2+2

13 R 语言

13.1 History of R

13.1 R 的历史

13.2 Basic principles and concepts

13.2 基本原理与概念

13.3 Data operation in R (Vectors, matrices, arrays, data frames)

13.3 R 的数据操作（向量、矩阵、数组、数据帧）

13.4 Plot figures

13.4 R 绘图

13.5 Statistical Analysis of R

13.5 R 的统计分析

13.6 Function definition and programing

13.6 函数定义和编程

13.7 packages

13.7 包

14 Clustering 学时: 2+2/Hours: 2+2

14 聚类

14.1 Why do clustering

14.1 为什么集群

14.2 Distance Metrics

14.2 距离度量

14.2.1 Euclidean distance

14.2.1 欧几里得距离

14.2.2 Pearson Linear Correlation

14.2.2 皮尔逊相关系数

14.3 Clustering algorithms

14.3 聚类算法

14.3.1 Hierarchical agglomerative clustering

14.3.1 层次集聚聚类

14.3.2 Partitioning methods

14.3.2 划分方法

15. Dimensionality reduction 学时: 2+2/Hours: 2+2

15 降维

15.1 Why dimensionality Reduction?

15.1 为什么降维?

15.2 Two approaches for dimensionality reduction (Feature Selection and feature extraction)



15.2 降维的两种方法（特征选择和特征提取）

15.3 Linear reduction

15.3 线性降维

15.3.1 Principal component analysis (PCA)

15.3.1 主成分分析（PCA）

15.3.2 Singular Value Decomposition (SVD)

15.3.2 奇异值分解（SVD）

15.3.3 Multi-Dimensional Scaling (MDS)

15.3.3 多维标度（MDS）

15.4 Non-linear reduction

15.4 非线性降维

15.4.1 t-distributed stochastic neighbor embedding（t-SNE）

15.4.1 t-分布随机邻域嵌入（t-SNE）

15.4.2 Uniform Manifold Approximation and Projection (UMAP)

15.4.2 均匀流形近似和投影（UMAP）

16 Theory of evolution

学时：2+2/Hours: 2+2

16 进化论

16.1 Evolution is a unifying theme in biology

16.1 进化是生物学的永恒主题

16.2 History of “evolutionary thought”

16.2“进化思想”的历史

16.3 Darwin's Four Postulates

16.3 达尔文的四个假设

16.3.1 Individuals within species vary.

16.3.1 物种内的个体不同

16.3.2 Some variations are heritable.

16.3.2 其中一些变异是可遗传的。

16.3.3 More offspring are produced than can survive

16.3.3 产生的后代多于存活的后代。

16.3.4 Survival and reproduction are nonrandom

16.3.4 存活和繁殖是非随机的

16.4 Modern evolutionary theory

16.4 现代进化论

18. 教材及其它参考资料 Textbook and Supplementary Readings

课程评估 ASSESSMENT

19. 评估形式 Type of Assessment	评估时间 Time	占考试总成绩百分比 % of final score	违纪处罚 Penalty	备注 Notes
出勤 Attendance		10		
课堂表现 Class Performance		20		I encourage you to ask questions during the class, and you will get credit by doing so. 鼓励在课堂上提问题
小测验 Quiz				
课程项目 Projects				
平时作业 Assignments		30		Completion of assignments 提交作业和作业完成情况
期中考试 Mid-Term Test				
期末考试 Final Exam				
期末报告 Final Presentation		40		Submit a paper at the end of the term. Do an oral presentation about the paper. 在学期结束时提交论文。做一个关于该论文的报告。
其它（可根据需要 改写以上评估方式） Others (The above may be modified as necessary)				

20. 记分方式 **GRADING SYSTEM**

- A. 十三级等级制 **Letter Grading**
 B. 二级记分制 (通过/不通过) **Pass/Fail Grading**

课程审批 **REVIEW AND APPROVAL**

21. 本课程设置已经过以下责任人/委员会审议通过

This Course has been approved by the following person or committee of authority

本课程经生物系本科教学指导委员会审议通过。

This Course has been approved by Undergraduate Teaching Steering Committee of Department of Biology.

