# 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 | 生物信息学/Bioinformatics |
| 2. | 授课院系<br>**Originating Department** | 生物系 Department of Biology |
| 3. | 课程编号<br>**Course Code** | BIO306 |
| 4. | 课程学分 Credit Value | 4 |
| 5. | 课程类别<br>**Course Type** | 专业核心课 Major Core Courses（生物信息专业 Bioinformatics）<br>专业选修课 Major Elective Courses（生物科学、生物技术专业 Bioloical Sciences, Biotechnology） |
| 6. | 授课学期<br>**Semester** | 春季 Spring |
| 7. | 授课语言<br>**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）<br>**Instructor(s), Affiliation& Contact**<br>（**For team teaching, please list all instructors**） | 生物系 Department of Biology<br>翟继先 ZHAI Jixian, zhaijx@sustech.edu.cn |
| 9. | 实验员/助教、所属学系、联系方式<br>**Tutor/TA(s), Contact** | 待公布 To be announced |
| 10. | 选课人数限额(可不填)<br>**Maximum Enrolment**（**Optional**） | |

| 11. | 授课方式<br>**Delivery Method** | 讲授<br>**Lectures** | 习题/辅导/讨论<br>**Tutorials** | 实验/实习<br>**Lab/Practical** | 其它(请具体注明)<br>**Other**（**Please specify**） | 总学时<br>**Total** |
|---|---|---|---|---|---|---|
| | 学时数<br>**Credit Hours** | 32 | | 64 | | 96 |

| | | |
|---|---|---|
| **12.** | 先修课程、其它学习要求<br>**Pre-requisites or Other Academic Requirements** | BIO309 计算生物学 Computational Biology |
| **13.** | 后续课程、其它学习规划<br>**Courses for which this course is a pre-requisite** | None 无 |
| **14.** | 其它要求修读本课程的学系<br>**Cross-listing Dept.** | None 无 |

## 教学大纲及教学日历 SYLLABUS

15. 教学目标 **Course Objectives**

This is a practical course in Bioinformatics which will emphasize how to use the computer as a tool for biomedical research. Prerequisites include a thorough understanding of theoretical and practical aspects of molecular biology, and some University level mathematics and statistics, but no prior knowledge of computer programming or computer hardware is necessary.

这是一门实用的生物信息学课程，这门课将强调如何将计算机作为生物医学研究的工具。本课程的先修要求深入理解分子生物学的理论和实践，以及大学水平的数学和统计学知识，但该课程不需要具备计算机编程或计算机硬件方面的先修知识。

16. 预达学习成果 **Learning Outcomes**

1、处理高通量测序数据，包括 DNA 测序、RNA 测序。

2、Linux 编程

3、在庞大的数据库上进行复杂搜索并分析结果

4、基因组比对，在基因组浏览器中显示基因和较大的基因区域

5、对 DNA 测序数据进行序列比对、数据过滤、变异查找

6、进行基因表达分析

7、进行蛋白质组学数据分析

1.process high throughput sequencing data, including DNA sequencing, RNA sequencing.

2.programming in Linux

3.Perform sophisticated searches over enormous databases, interpret their results

4.Perform genomic comparisons, display genes and large genomic regions in Genome Browser

5.Perform sequence alignment, data filtering, variants calling in DNA sequencing data

6.Perform gene expression analysis

7.Perform proteomics data analysis

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.基因组学科研思维与技术策略(方晓东)

1.1 课程简介：课程目标、教学形式、考核方式与学习指南

1.2 基因组学与生物信息学研究进展

1.3 基因组学与生物信息学经典案例分享

1.4 科学思维与技术策略制定

7.Scientific research thinking and technological strategy of genomics (FANG Xiaodong)

1.1 introduction: objectives, teaching methods, assessments and learning guides

1.2 advances in genomics and bioinformatics

1.3 classic cases of genomics and bioinformatics

1.4 scientific thinking and formulation of technical strategies

2.计算机基础（1+5，linux+编程，实操）（王崇志）

2.1 Linux 系统与开源软件：github

2.2 Linux 基础操作与 shell 编程：集群与 qsub

2.3 perl 语言编程与字符串处理：正则表达式

2.4 python 语言编程

2.5 R 语言统计与绘图：统计检验与数据可视化工具

2.6 数据分析流程搭建：模块化与流程化，工程学思想

3.Computer basics (1+5, Linux + programming, practical operation) (WANG Chongzhi)

2.1 Linux system and open source software: github

2.2 basic Linux operations and shell programming: clustering and qsub

2.3 perl and string processing: regular expressions

2.4 python programming

2.5 R based statistics and drawing: statistical inspection and data visualization tools

2.6 data analysis process building: modularization and routing, engineering ideas

3.常用软件和数据库介绍（3+3，软件 2+数据库 1+实操 3），包括常用文件格式（王崇志）

3.1 序列组装、比对及相关软件

3.2 实操 1：包括 fasta、fastq、bam、vcf 格式介绍

3.3 功能分析与进化分析软件

3.4 实操 2：包括 newick 格式介绍

3.5 常用生物数据库：核酸、蛋白、通路、变异、疾病、肿瘤、物种

3.6 实操 3：GenBank、KEGG、GO、dbSNP、OMIM、TCGA

3.Introduction to common software and database (3+3, software 2+ database 1+ practical operation 3), including common file format (WANG Chongzhi)

3.1 sequence assembly, comparison and related software

3.2 operation 1: introduction of fasta, fastq, bam and VCF formats

3.3 functional analysis and evolutionary analysis softwares

3.4 operation 2: introduction of newick format

3.5 commonly used biological databases: nucleic acid, protein, pathway, mutation, disease, tumor, species

3.6 operation 3: GenBank, KEGG, GO, dbSNP, OMIM, TCGA

4.生信常用算法：常用算法介绍，通过对软件数据调整加深对算法的理解，如何针对实际应用优化参数选择；比对：blast/soap/soap2; 算法原理、关键参数、如何优化、如何评价（王崇志）

4.1 生物问题与数学建模：同源与相似，种化与癌变，模式发现

4.2 问题求解与算法实现：分而治之、动态规划、马尔科夫模型

4.3 序列比对原理与关键参数 I：blast

4.4 序列比对原理与关键参数 II：SOAPaligner、bwa

4.5 建树与聚类算法

4.6 motif 识别算法

4. Common bioinformatics algorithms: introduction of common algorithms, to deepen the understanding of the algorithms through the adjustment of softwares, how to optimize parameters Alignment: blast/soap/soap2; Algorithm principle, key parameters, how to optimize, how to evaluate (WANG Chongzhi)

4.1 biological problems and mathematical modeling: homology and similarity, speciation and canceration, pattern discovery

4.2 problem solving and algorithm implementation: divide and conquer, dynamic programming, markov model

4.3 sequence alignment principle and key parameters I: blast

4.4 alignment principle and key parameters II: SOAPaligner, bwa

4.5 tree building and clustering algorithm

4.6 motif recognition algorithm

5.大数据和云计算（2+4）（王崇志）

5.1 生物大数据

5.2 生物云计算

5.3 Galaxy 系统安装

5.4 搭建你的第一个云流程

5.5 比较几款常见的生物信息云平台

5.6 优化或扩展你的云流程

5.Big data and cloud computing (2+4) (WANG Chongzhi)

5.1 biological big data

5.2 biological cloud computing

5.3 Galaxy system installation

5.4 build your first cloud process

5.5 comparison of several common biological cloud platforms

5.6 optimize or extend your cloud processes

6.De novo 基因组分析（1+5）（王崇志）

6.1 基因组组装问题建模

6.2 组装软件与应用场景

6.3 基因组组装：单个细菌基因组

6.4 组装评价：指标体系、QUAST 软件

6.5 基因组注释与比较基因组分析

6.6 宏基因组组装与分析

6.De novo genome analysis (1+5) (WANG Chongzhi)

6.1 modeling of genome assembly problems

6.2 assembly software and application scenarios

6.3 genome assembly: single bacterial genome

6.4 assembly evaluation: indicator system and QUAST software

6.5 genome annotation and comparative genome analysis

6.6 macronomic assembly and analysis

7.　　个人基因组学(张璐)

7.1 个人基因组学和精准医疗的历史

7.2 人类基因组测序数据分析

7.3 你来自哪里？

7.4 优先考虑导致疾病的突变因素

7.5 疾病预测中的多基因风险评分

7.6 生活方式与基因组学

7.Personal genomics （ZHANG Lu）

7.1 History of personal genomics and precision medicine

7.2 Data analysis for human genomic sequencing

7.3 Where are you come from?

7.4 Prioritize disease causal mutation

7.5 Polygenic Risk Score for disease prediction

7.6 Lifestyles and Genomics

8.　　　人类基因组学(张璐)

8.1 人类群体基因组学简介(HapMap, 1000 Genomes, UK Biobank)

8.2 人类基因组学的关键概念和算法

8.3 群体遗传结构

8.4 基于人群的复杂疾病关联研究

8.5 基因组+X (PheWAS, TWAS, EWAS)

8.6 著名软件介绍

8.Human Population Genomics（ZHANG Lu）

8.1 Introduction of human population genomics (HapMap, 1000 Genomes, UK Biobank)

8.2 Key concept and algorithms in human population genomics

8.3 Genetic structure of populations

8.4 Population based association study to complex diseases

8.5 Genomics+X (PheWAS, TWAS, EWAS)

8.6 Famous software introduction

9.表观遗传学（2+4）（王崇志）

9.1 表观遗传研究与高通量技术

9.2 表观遗传研究技术：碱基修饰（DNA、RNA）、组蛋白修饰（各种修饰）、蛋白质-核酸互作（TF，enhancer 等）及空间三维结构（2C，3C,HiC），BS-seq、ChIP-seq、HiC-seq、ATAC-seq

9.3 上机实操：数据质控与分析软件安装

9.4 上机实操：修饰位点注释与差异分析

9.5 上机实操：信息整合与统计绘图

9.6 综合实操：表观组学研究方案设计与讨论

10.Epigenetics (2+4) (WANG Chongzhi)

9.1 epigenetic research and high-throughput techniques

9.2 epigenetic research techniques: base modification (DNA, RNA), histone modification (various modifications), protein-nucleic acid interaction (TF, enhancer, etc.) and three-dimensional structure (2C, 3C,HiC), bs-seq, chip-seq, hec-seq, atac-seq

9.3 actual operation: data quality control and analysis software installation

9.4 computer operation: modification site annotation and difference analysis

9.5 computer operation: information integration and statistical drawing

9.6 comprehensive practice: design and discussion of the research scheme of apparent omics

10.转录组学（1+5）（王崇志）

10.1 转录组测序技术：mRNA、ncRNA、sRNA、降解组

10.2 转录组组装与结构分析：SNP 检测、SSR 预测、可变剪接

10.3 表达定量与差异分析：SEQC，ERCC，TPM，统计检验与多重校正

10.4 模式聚类与富集分析：GO、KEGG 富集、共表达网络

10.5 ncRNA、sRNA、降解组测序分析

10.6 转录组研究方案设计与讨论

10.Transcriptome (1+5) (WANG Chongzhi)

10.1 transcriptome sequencing technology: mRNA, ncRNA, sRNA, degradation group

10.2 transcriptome assembly and structure analysis: SNP detection, SSR prediction, variable splicing

10.3 expression quantitative and differential analysis: SEQC, ERCC, TPM, statistical test and multiple correction

10.4 pattern clustering and enrichment analysis: GO, KEGG enrichment, co-expression network

10.5 ncRNA, sRNA and degradation group sequencing analysis

10.6 design and discussion of transcriptome research scheme

11.宏基因组（1+5）（周勇）

11.1 宏基因组学的发展历程

11.2 数据质控与宏基因组组装

11.3 物种组成分析：MetaPhlAn2

11.4 微生态功能分析

11.5 环境因子关联分析

11.6 宏基因组分析结果可视化

11.Metagenome (1+5) (ZHOU Yong)

11.1 history of macronomics

11.2 data quality control and macronomic assembly

11.3 analysis of species composition: MetaPhlAn2

11.4 microecological function analysis

11.5 correlation analysis of environmental factors

11.6 visualization of metagenomic analysis results

12.肿瘤生物信息学（1+8）（周勇）

12.1 肿瘤基因组研究历程
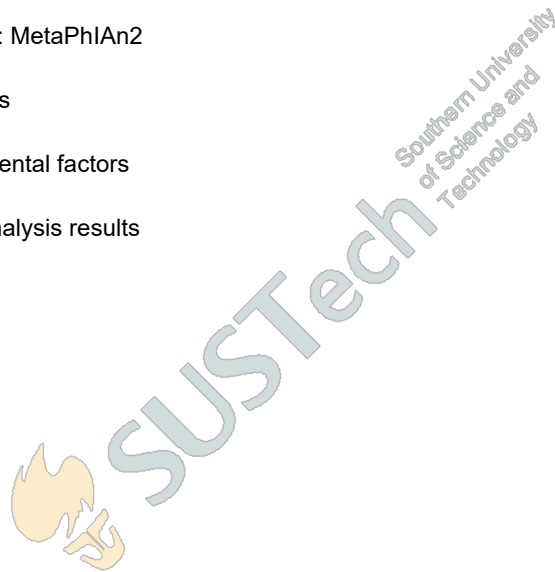
12.2 肿瘤基因组图谱

12.3 肿瘤标志物鉴定以及靶向治疗

12.4 肿瘤突变频谱以及机制

12.5 肿瘤进化

12.6 肿瘤单细胞基因组

12.7 肿瘤免疫微环境

12.8 肿瘤与肠道微生物

12.9 公共数据利用：研究主题与分析方案设计。

12. Tumor bioinformatics (1+8) (ZHOU Yong)

12.1 history of tumor genome research

12.2 tumor genome atlas

12.3 identification of tumor markers and targeted therapy

12.4 tumor mutation spectrum and its mechanism

12.5 tumor evolution

12.6 tumor single cell genome

12.7 tumor immune microenvironment

12.8 tumor and intestinal microorganism

12.9 public data: research topic and analysis scheme design.

13. 系统生物学网络分析和可视化(张璐)

13.1 生物网络的类型简介

13.2 复杂网络中的关键概念

13.3 蛋白-蛋白相互作用网络

13.4 基因调控网络

13.5 基因集富集与网络分析

13.6 可视化降维

13.Network analysis and visualization for system biology（ZHANG Lu）

13.1 Introduce types of biological networks

13.2 Key concept in complex network

13.3 Protein-Protein Interaction network

13.4 Gene Regulatory Network

13.5 Gene Set Enrichment and network analysis

13.6 Dimensionality reduction for visualization

14. 机器学习在基因组学中的应用(张璐)

14.1 机器学习的关键概念

14.2 机器学习在从头组装中的应用

14.3 机器学习在建立电子健康档案中的应用

14.4 基因组学中的深度神经网络

14.5 深度学习在生物网络中的应用

14.Machine learning in genomics（ZHANG Lu）

14.1 Key concept in machine learning

14.2 Machine learning in de novo assembly

14.3 Machine learning in electronic health record

14.4 Deep neural network in genomics

14.5 Deep learning for network biology

15.总结、讨论、答疑（方晓东 王崇志 张璐 周勇）

15.1 课程回顾与总结

15.2 小组分主题讨论

15.3 期末答疑及论文选题

16.Summary, discussion and Q&A (FANG Xiaodong, WANG Chongzhi, ZHANG Lu, ZHOU Yong)

15.1 course review and summary

15.2 group discussion on topics

15.3 final Q&A and thesis selection

16.学生报告展示

16.Students' presentation

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

Next-Generation DNA Sequencing Informatics, ISBN-13: 978-1936113873ISBN-10: 1936113872

## 课程评估 ASSESSMENT

**19.**

| 评估形式<br>Type of Assessment | 评估时间<br>Time | 占考试总成绩百分比<br>% of final score | 违纪处罚<br>Penalty | 备注<br>Notes |
|---|---|---|---|---|
| 出勤 Attendance | | | | |
| 课堂表现<br>Class Performance | | 10 | | |
| 小测验<br>Quiz | | | | |
| 课程项目 Projects | | | | |
| 平时作业<br>Assignments | | 40 | | |
| 期中考试 | | | | |

| | | | |
|---|---|---|---|
| **Mid-Term Test** | | | |
| 期末考试 **Final Exam** | 50 | | |
| 期末报告 **Final Presentation** | | | |
| 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.