CU_CURR501
Page 1 of 4

THE CHINESE UNIVERSITY OF HONG KONG
Print Course Catalog Details

May 10, 2023
14:19:24 PM

Academic Org: Dept of Computer Sci & Engg – Subject: Computer Science

Course: CSCI3220   **Course ID:** 002593   **Eff Date:** 2022-07-01   **Crse Status: Active**   **Apprv. Status:** Approved   【Course Rev】

Algorithms for Bioinformatics 生物信息學中的算法

This course describes some algorithms commonly used in contemporary bioinformatics. After a brief introduction to basic molecular biology and genetics, four main topics will be covered with corresponding lists of algorithms:

1. Sequence alignment and assembly: dynamic programming for optimal sequence alignment, FASTA and BLAST for heuristic alignment of long sequences, tree-based methods for multiple sequence alignment, suffix-tree, suffix-array and Burrows-Wheeler Transform for short read alignment, and algorithms based on de Bruijn graphs for sequence assembly

2. Statistical modeling: forward, backward, Viterbi and Baum-Welch algorithms for hidden Markov models, Gibbs sampling for sequence motif discovery, and Bayesian classifiers, logistic regression and expectation-maximization for data classification and clustering

3. Phylogenetics: methods based on Jukes-Cantor and Kimura models for divergence time estimation, maximum parsimony, UPGMA, Neighbor-joining and maximum likelihood methods for phylogenetic tree reconstruction.

4. High-throughput data analysis: Hierarchical clustering and k-means for data clustering, and algorithms for selected problems such as signal peak calling, detection of gene fusion, haplotype phasing. Other topics such as RNA secondary structure prediction may also be covered depending on the available time.

本科介紹一些現代生物信息學中常用的算法。本科以基礎份子生物學及遺傳學簡介開始，接著圍繞四個主題討論相關的算法：
1. 序列對準及組裝：爲最佳化對準的動態規劃；爲對準長序列的FASTA和BLAST；爲多序列對準的樹狀方法；爲短序列的後綴樹、後綴數組和Burrows-Wheeler變換；以及爲序列組裝，基於de Bruijn圖形的算法
2. 統計模型：爲隱Markov模型的前向、反向、Viterbi和Baum-Welch算法；爲找尋序列主題的Gibbs取樣程序；以及爲數據分類和群組的Bayesian分類法、邏輯迴歸法和期望值最大化
3. 物種發生學：爲估測物種分離時間，基於Jukes-Cantor和Kimua模型的算法；以及爲重構物種樹的最大簡約法、UPGMA、鄰舍結合法和最大似然法
4. 高量數據分析：爲數據群組的等級群組法和k-means；以及特選課題的相關算法，例如搜尋訊息頂位、基因合併和單體型建構若時間許可，本科可能會加入諸如核糖核酸二級結構估計等其他題目。

Grade Descriptor:

A

EXCELLENT – exceptionally good performance and far exceeding expectation in all or most of the course learning outcomes; demonstration of superior understanding of the subject matter, the ability to analyze problems and apply extensive knowledge, and skillful use of concepts and materials to derive proper solutions.

有關等級說明的資料，請參閱英文版本。

B

GOOD – good performance in all course learning outcomes and exceeding expectation in some of them; demonstration of good understanding of the subject matter and the ability to use proper concepts and materials to solve most of the problems encountered.

有關等級說明的資料，請參閱英文版本。

CU_CURR501
Page 2 of 4

THE CHINESE UNIVERSITY OF HONG KONG
Print Course Catalog Details

May 10, 2023
14:19:24 PM

C

FAIR – adequate performance and meeting expectation in all course learning outcomes; demonstration of adequate understanding of the subject matter and the ability to solve simple problems.

有關等級說明的資料，請參閱英文版本。

D

MARGINAL – performance barely meets the expectation in the essential course learning outcomes; demonstration of partial understanding of the subject matter and the ability to solve simple problems.

有關等級說明的資料，請參閱英文版本。

F

FAILURE – performance does not meet the expectation in the essential course learning outcomes; demonstration of serious deficiencies and the need to retake the course.

有關等級說明的資料，請參閱英文版本。

Equivalent Offering:
Units: 3 (Min) / 3 (Max) / 3 (Acad Progress)
Grading Basis: Graded
Repeat for Credit: N
Multiple Enroll: N
Course Attributes:

Topics:

COURSE OUTCOMES

Learning Outcomes:

At the end of the course of studies, students will be able to
1. Explain the problems studied in both biological and computational terms.
2. Explain the details of the studied algorithms.

CU_CURR501
Page 3 of 4

THE CHINESE UNIVERSITY OF HONG KONG
Print Course Catalog Details

May 10, 2023
14:19:24 PM

3. Apply the learned algorithms to solving both problems in bioinformatics and other domains.
4. Study new algorithms proposed for the same problems.
5. Understand contents of other classes in computational biology and bioinformatics.

Course Syllabus:

This course describes some algorithms commonly used in contemporary bioinformatics. After a brief introduction to basic molecular biology and genetics, four main topics will be covered with corresponding lists of algorithms:
1. Sequence alignment and assembly: dynamic programming for optimal sequence alignment, FASTA and BLAST for heuristic alignment of long sequences, tree-based methods for multiple sequence alignment, suffix-tree, suffix-array and Burrows-Wheeler Transform for short read alignment, and algorithms based on de Bruijn graphs for sequence assembly
2. Statistical modeling: forward, backward, Viterbi and Baum-Welch algorithms for hidden Markov models, Gibbs sampling for sequence motif discovery, and Bayesian classifiers, logistic regression and expectation-maximization for data classification and clustering
3. Phylogenetics: methods based on Jukes-Cantor and Kimura models for divergence time estimation, maximum parsimony, UPGMA, Neighbor-joining and maximum likelihood methods for phylogenetic tree reconstruction.
4. High-throughput data analysis: Hierarchical clustering and k-means for data clustering, and algorithms for selected problems such as signal peak calling, detection of gene fusion, haplotype phasing.
Other topics such as RNA secondary structure prediction may also be covered depending on the available time.

Assessment Type:

Others	: 35%
Short answer test or exam	: 65%

Feedback for Evaluation:

1. Mid-term course evaluation
2. Term-end course evaluation
3. Students' performance in their homework, the midterm exam and final exam
4. Students' performance and participation in tutorial sessions

Required Readings:

-

Recommended Readings:

1. Wing-Kin Sung, Algorithms in Bioinformatics: A Practical Introduction, Chapman & Hall, 2009 (and corresponding online materials at http://www.comp.nus.edu.sg/~ksung/algo_in_bioinfo/)
2. Neil C. Jones and Pavel A. Pevzner, An Introduction to Bioinformatics Algorithms, MIT Press, 2004

CU_CURR501
Page 4 of 4

THE CHINESE UNIVERSITY OF HONG KONG
Print Course Catalog Details

May 10, 2023
14:19:24 PM

OFFERINGS

1. CSCI3220   Acad Organization=CSD; Acad Career=UG

COMPONENTS

LEC : Size=30; Final Exam=Y; Contact=3
TUT : Size=30; Final Exam=N; Contact=1

ENROLMENT REQUIREMENTS

1. CSCI3220   **Enrollment Requirement Group:**
Not for students who have taken ESTR3110.

New Enrollment Requirement(s):
Pre-requisite = no change
Co-requisite = no change
Exclusion = add "ESTR3110"

CAF

eLearning hrs for blended cls   0
No. of micro-modules   14
Research components (UG)   50%-74%

< E N D   O F   R E P O R T >