

Guia docent

270656 - BSG - Bioinformàtica i Genètica Estadística

Última modificació: 12/07/2021

Unitat responsable: Facultat d'Informàtica de Barcelona
Unitat que imparteix: 715 - EIO - Departament d'Estadística i Investigació Operativa.
723 - CS - Departament de Ciències de la Computació.

Titulació: MÀSTER UNIVERSITARI EN INNOVACIÓ I RECERCA EN INFORMÀTICA (Pla 2012). (Assignatura optativa).

Curs: 2021 **Crèdits ECTS:** 6.0 **Idiomes:** Anglès

PROFESSORAT

Professorat responsable: GABRIEL ALEJANDRO VALIENTE FERUGLIO

Altres: Primer quadrimestre:
JAN GRAFFELMAN - 10
GABRIEL ALEJANDRO VALIENTE FERUGLIO - 10

CAPACITATS PRÈVIES

Basic knowledge of algorithms and data structures.
Basic knowledge of statistics.
Basic knowledge of the Python programming language.
Basic knowledge of the R programming language.

COMPETÈNCIES DE LA TITULACIÓ A LES QUALS CONTRIBUEIX L'ASSIGNATURA

Específiques:

CEC1. Capacitat per aplicar el mètode científic en l'estudi i anàlisi de fenòmens i sistemes en qualsevol àmbit de la Informàtica, així com en la concepció, disseny i implantació de solucions informàtiques innovadores i originals.

CEC2. Capacitat per al modelatge matemàtic, càlcul i disseny experimental en centres tecnològics i d'enginyeria d'empresa, particularment en tasques de recerca i innovació en tots els àmbits de la Informàtica.

CEC3. Capacitat per aplicar solucions innovadores i realitzar avanços en el coneixement que explotin els nous paradigmes de la Informàtica, particularment en entorns distribuïts.

Genèriques:

CG3. Capacidad para el modelado matemático, cálculo y diseño experimental en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación e innovación en todos los ámbitos de la Informática.

Transversals:

CTR6. RAONAMENT: Capacitat de raonament crític, lògic i matemàtic. Capacitat de resoldre problemes en la seva àrea d'estudi. Capacitat d'abstracció: capacitat de crear i utilitzar models que reflecteixin situacions reals. Capacitat de dissenyar i realitzar experiments senzills, i analitzar-ne i interpretar-ne els resultats. Capacitat d'anàlisi, de síntesi i d'avaluació.

Bàsiques:

CB6. Que els estudiants sàpiguin aplicar els coneixements adquirits y la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

CB7. Que els estudiants siguin capaços d'integrar coneixements i enfrontar-se a la complexitat de formular judicis a partir d'una informació que, essent incompleta o limitada, inclogui reflexions sobre les responsabilitats socials i ètiques vinculades a l'aplicació dels seus coneixements i judicis.

CB9. Que els estudiants posseeixin les habilitats d'aprenentatge que els permetin continuar estudiant d'una manera que haurà de ser en gran mesura autodirigida o autònoma.

METODOLOGIES DOCENTS

All classes consist of a theoretical session (a lecture in which the professor introduces new concepts or techniques and detailed examples illustrating them) followed by a practical session (in which the students work on the examples and exercises proposed in the lecture). On the average, two hours a week are dedicated to theory and one hour a week to practice, and the professor allocates them according to the subject matter. Students are required to take an active part in class and to submit the exercises at the end of each class.

OBJECTIUS D'APRENTATGE DE L'ASSIGNATURA

- 1.Introduce the student to the algorithmic, computational, and statistical problems that arise in the analysis of biological data.
- 2.Reinforce the knowledge of discrete structures, algorithmic techniques, and statistical techniques that the student may have from previous courses.

HORES TOTS DE DEDICACIÓ DE L'ESTUDIANTAT

Tipus	Hores	Percentatge
Hores grup gran	54,0	36.00
Hores aprenentatge autònom	96,0	64.00

Dedicació total: 150 h

CONTINGUTS

Introduction to bioinformatics

Descripció:

Computational biology and bioinformatics. Algorithms in bioinformatics. Strings, sequences, trees, and graphs. Algorithms on strings and sequences. Representation of trees and graphs. Algorithms on trees and graphs.

Phylogenetic reconstruction I

Descripció:

Character-based phylogenetic reconstruction. Compatibility. Perfect phylogenies. Distance-based phylogenetic reconstruction. Additive trees. Ultrametric trees.

Agreement of phylogenetic trees

Descripció:

Partition distance. Triplets distance. Quartets distance. Transposition distance. Edit distance and alignment of phylogenetic trees.

Phylogenetic reconstruction II

Descripció:

Phylogenetic networks. Galled trees. Tree-child networks. Tree-sibling networks. Time consistency of phylogenetic networks.

Agreement of phylogenetic networks

Descripció:

Path multiplicity distance. Tripartition distance. Nodal distance. Triplets distance. Edit distance and alignment of phylogenetic networks.

Phylogenetic reconstruction III

Descripció:

Mutation trees. Clonal trees. Clonal deconvolution.

Phylogenetic and taxonomic reconstruction

Descripció:

Phylogenies and taxonomies. Classification of metagenomic samples. Agreement of classifications.

Introduction to statistical genetics

Descripció:

Basic genetic terminology. Population-based and family-based studies. Traits, markers and polymorphisms. Single nucleotide polymorphisms and microsatellites. R-package genetics.

Hardy-Weinberg equilibrium

Descripció:

Hardy-Weinberg law. Hardy-Weinberg assumptions. Multiple alleles. Statistical tests for Hardy-Weinberg equilibrium: chi-square, exact and likelihood-ratio tests. Graphical representations. Disequilibrium coefficients: the inbreeding coefficient, Weir's D. R-package HardyWeinberg.

Linkage disequilibrium

Descripció:

Definition of linkage disequilibrium (LD). Measures for LD. Estimation of LD by maximum likelihood. Haplotypes. The HapMap project. Graphics for LD. The LD heatmap.

Phase estimation

Descripció:

Phase ambiguity for double heterozygotes. Phase estimation with the EM algorithm. Estimation of haplotype frequencies. R-package haplo.stats.

Population substructure

Descripció:

Definition of population substructure. Population substructure and Hardy-Weinberg equilibrium. Population substructure and LD. Statistical methods for detecting substructure. Multidimensional scaling. Metric and non-metric multidimensional scaling. Euclidean distance matrices. Stress. Graphical representations.



Genetic association analysis

Descripció:

Disease-marker association studies. Genetic models: dominant, co-dominant and recessive models. Testing models with chi-square tests. The alleles test and the Cochran-Armitage trend test. Genome-wide association tests.

Family relationships and allele sharing

Descripció:

Identity by state (IBS) and Identity by descent (IBD). Kinship coefficients. Allele sharing. Detection of family relationships. Graphical representations.

ACTIVITATS

Development of syllabus topics

Objectius específics:

1, 2

Competències relacionades:

CG3. Capacidad para el modelado matemático, cálculo y diseño experimental en centros tecnológicos y de ingeniería de empresa, particularmente en tareas de investigación e innovación en todos los ámbitos de la Informática.

CEC1. Capacitat per aplicar el mètode científic en l'estudi i anàlisi de fenòmens i sistemes en qualsevol àmbit de la Informàtica, així com en la concepció, disseny i implantació de solucions informàtiques innovadores i originals.

CEC2. Capacitat per al modelatge matemàtic, càlcul i disseny experimental en centres tecnològics i d'enginyeria d'empresa, particularment en tasques de recerca i innovació en tots els àmbits de la Informàtica.

CEC3. Capacitat per aplicar solucions innovadores i realitzar avanços en el coneixement que explotin els nous paradigmes de la Informàtica, particularment en entorns distribuïts.

CTR6. RAONAMENT: Capacitat de raonament crític, lògic i matemàtic. Capacitat de resoldre problemes en la seva àrea d'estudi. Capacitat d'abstracció: capacitat de crear i utilitzar models que reflecteixin situacions reals. Capacitat de dissenyar i realitzar experiments senzills, i analitzar-ne i interpretar-ne els resultats. Capacitat d'anàlisi, de síntesi i d'avaluació.

CB7. Que els estudiants siguin capaços d'integrar coneixements i enfrontar-se a la complexitat de formular judicis a partir d'una informació que, essent incompleta o limitada, inclogui reflexions sobre les responsabilitats socials i ètiques vinculades a l'aplicació dels seus coneixements i judicis.

CB9. Que els estudiants posseeixin les habilitats d'aprenentatge que els permetin continuar estudiant d'una manera que haurà de ser en gran mesura autodirigida o autònoma.

CB6. Que els estudiants sàpiguen aplicar els coneixements adquirits y la seva capacitat de resolució de problemes en entorns nous o poc coneguts dins de contextos més amplis (o multidisciplinaris) relacionats amb la seva àrea d'estudi.

Dedicació: 117h

Grup gran/Teoria: 15h

Grup petit/Laboratori: 30h

Aprenentatge autònom: 72h

Final exam

Dedicació: 33h

Grup gran/Teoria: 3h

Aprenentatge autònom: 30h

SISTEMA DE QUALIFICACIÓ

Students are evaluated during class, and in a final exam. Every student is required to submit one exercise each week, graded from 0 to 10, and the final grade consists of 50% for the exercises and 50% for the final exam, also graded from 0 to 10.

BIBLIOGRAFIA

Bàsica:

- Valiente, Gabriel. Algorithms on trees and graphs. 2nd ed. Springer Nature, 2021. ISBN 9783030818845.
- Valiente, Gabriel. Combinatorial pattern matching algorithms in computational biology using Perl and R. Boca Raton: Chapman and Hall/CRC, 2009. ISBN 9781420069730.
- Foulkes, Andrea S. Applied statistical genetics with R : for population-based association studies. New York: Springer, 2009. ISBN 9780387895536.
- Laird, Nan M.; Lange, Christoph. The Fundamentals of Modern Statistical Genetics. Springer, 2011. ISBN 9781441973375.

Complementària:

- Gusfield, Dan. Algorithms on strings, trees, and sequences : computer science and computational biology. Cambridge [England]: Cambridge University Press, 1997. ISBN 0521585198.
- Paradis, Emmanuel. Analysis of phylogenetics and evolution with R [en línia]. Second edition. New York: Springer, 2012 [Consulta: 16/07/2021]. Disponible a: <https://doi.org/10.1007/978-1-4614-1743-9>. ISBN 9781461417439.
- Weir, B.S. Genetic data analysis II: methods for discrete population genetic data. Sinauer Associates, 1996. ISBN 0878939024.
- Ziegler, Andreas; König, Inke R.. Statistical Approach to Genetic Epidemiology [en línia]. 2nd ed. Weinheim: Wiley-VCH, 2011 [Consulta: 16/07/2021]. Disponible a: <https://onlinelibrary-wiley-com/doi/book/10.1002/9783527633654>. ISBN 9783527633654.

RECURSOS

Enllaç web:

- <http://rosalind.info/>. Rosalind
- <http://www.r-project.org/>. The R Project for Statistical Computing