



Guía docente

270656 - BSG - Bioinformática y Genética Estadística

Última modificación: 12/07/2021

Unidad responsable: Facultad de Informática de Barcelona

Unidad que imparte: 715 - EIO - Departamento de Estadística e Investigación Operativa.

723 - CS - Departamento de Ciencias de la Computación.

Titulación: MÁSTER UNIVERSITARIO EN INNOVACIÓN E INVESTIGACIÓN EN INFORMÁTICA (Plan 2012). (Asignatura optativa).

Curso: 2021

Créditos ECTS: 6.0

Idiomas: Inglés

PROFESORADO

Profesorado responsable: GABRIEL ALEJANDRO VALIENTE FERUGLIO

Otros: Primer cuatrimestre:

JAN GRAFFELMAN - 10

GABRIEL ALEJANDRO VALIENTE FERUGLIO - 10

CAPACIDADES PREVIAS

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.

COMPETENCIAS DE LA TITULACIÓN A LAS QUE CONTRIBUYE LA ASIGNATURA

Específicas:

CEC1. Capacidad para aplicar el método científico en el estudio y análisis de fenómenos y sistemas en cualquier ámbito de la Informática, así como en la concepción, diseño e implantación de soluciones informáticas innovadoras y originales.

CEC2. 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.

CEC3. Capacidad para aplicar soluciones innovadoras y realizar avances en el conocimiento que exploten los nuevos paradigmas de la Informática, particularmente en entornos distribuidos.

Genéricas:

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.

Transversales:

CTR6. RAZONAMIENTO: Capacidad de razonamiento crítico, lógico y matemático. Capacidad para resolver problemas dentro de su área de estudio. Capacidad de abstracción: capacidad de crear y utilizar modelos que reflejen situaciones reales. Capacidad de diseñar y realizar experimentos sencillos, y analizar e interpretar sus resultados. Capacidad de análisis, síntesis y evaluación.

Básicas:

CB6. Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio.

CB7. Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios

CB9. Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.



METODOLOGÍAS DOCENTES

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.

OBJETIVOS DE APRENDIZAJE DE LA ASIGNATURA

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.

HORAS TOTALES DE DEDICACIÓN DEL ESTUDIANTADO

Tipo	Horas	Porcentaje
Horas grupo grande	54,0	36.00
Horas aprendizaje autónomo	96,0	64.00

Dedicación total: 150 h

CONTENIDOS

Introduction to bioinformatics

Descripción:

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

Descripción:

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

Agreement of phylogenetic trees

Descripción:

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

Phylogenetic reconstruction II

Descripción:

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



Agreement of phylogenetic networks

Descripción:

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

Phylogenetic reconstruction III

Descripción:

Mutation trees. Clonal trees. Clonal deconvolution.

Phylogenetic and taxonomic reconstruction

Descripción:

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

Introduction to statistical genetics

Descripción:

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

Hardy-Weinberg equilibrium

Descripción:

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

Descripción:

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

Descripción:

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

Population substructure

Descripción:

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

Descripción:

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

Descripción:

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

ACTIVIDADES

Development of syllabus topics

Objetivos específicos:

1, 2

Competencias relacionadas:

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.

CEC3. Capacidad para aplicar soluciones innovadoras y realizar avances en el conocimiento que exploten los nuevos paradigmas de la Informática, particularmente en entornos distribuidos.

CEC2. 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. Capacidad para aplicar el método científico en el estudio y análisis de fenómenos y sistemas en cualquier ámbito de la Informática, así como en la concepción, diseño e implantación de soluciones informáticas innovadoras y originales.

CTR6. RAZONAMIENTO: Capacidad de razonamiento crítico, lógico y matemático. Capacidad para resolver problemas dentro de su área de estudio. Capacidad de abstracción: capacidad de crear y utilizar modelos que reflejen situaciones reales. Capacidad de diseñar y realizar experimentos sencillos, y analizar e interpretar sus resultados. Capacidad de análisis, síntesis y evaluación.

CB6. Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio.

CB9. Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo.

CB7. Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios

Dedicación: 117h

Grupo grande/Teoría: 15h

Grupo pequeño/Laboratorio: 30h

Aprendizaje autónomo: 72h

Final exam

Dedicación: 33h

Grupo grande/Teoría: 3h

Aprendizaje autónomo: 30h



SISTEMA DE CALIFICACIÓN

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.

BIBLIOGRAFÍA

Básica:

- 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.

Complementaria:

- 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 línea]. 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 línea]. 2nd ed. Weinheim: Wiley-VCH, 2011 [Consulta: 16/07/2021]. Disponible a: <https://onlinelibrary-wiley-com/doi/book/10.1002/9783527633654>. ISBN 9783527633654.

RECURSOS

Enlace web:

- <http://rosalind.info/>
- <http://www.r-project.org/>