

Testi del Syllabus

Resp. Did. **GERDOL MARCO** **Matricola: 014720**

Docente **GERDOL MARCO, 6 CFU**

Anno offerta: **2020/2021**

Insegnamento: **644SM - BIOINFORMATICA**

Corso di studio: **SM51 - SCIENZE E TECNOLOGIE BIOLOGICHE**

Anno regolamento: **2018**

CFU: **6**

Settore: **BIO/18**

Tipo Attività: **B - Caratterizzante**

Anno corso: **3**

Periodo: **Secondo Semestre**

Sede: **TRIESTE**



Testi in italiano

Lingua insegnamento ITALIANO

Contenuti (Dipl.Sup.) Introduzione alla Bioinformatica. Gestione ed analisi delle informazioni biologiche. Database campi e livelli di specializzazione. Information retrieval e interrogazioni complesse. Allineamento di sequenze di acidi nucleici e proteine. Allineamento dinamico ed euristico. Matrici di sostituzione e FASTA. Applicazioni principali di BLAST. Applicazioni speciali di BLAST. Allineamenti multipli, algoritmi iterativi e progressivi. Allineamento di profili (PSI-BLAST). Pattern e motivi funzionali nel DNA. Pattern e motivi funzionali in proteine. Modeling strutturale di sequenze proteiche. Basi di analisi filogenetica, metodi basati su matrici di distanze e caratteri.

Testi di riferimento Pascarella, Paiardini. Bioinformatica: dalla sequenza alla struttura delle proteine. Zanichelli (2011)
Valle, Helmer-Citterich, Attimonelli, Pesole. Introduzione alla Bioinformatica. Zanichelli
Lesk. Introduzione alla Bioinformatica. McGraw-Hill
Krane, Raymer. Fondamenti di Bioinformatica. Pearson
Mount D. Bioinformatics. Sequence and Genome Analysis. Second Edition. CSHL Press
Baxevanis, Oullette. Bioinformatics. A practical guide to the analysis of genes and proteins. Wiley

Obiettivi formativi Il corso si prefigge di permettere agli studenti di acquisire le abilità per utilizzare in modo razionale ed efficiente i metodi bioinformatici e per interpretare correttamente i risultati.

Conoscenza e comprensione
-conoscere i potenziali campi di applicazione della bioinformatica
-conoscere l'organizzazione dei principali database bioinformatici ed apprendere il tipo di contenuti che è possibile ritrovarvi

- apprendere quali sono gli strumenti più utili per ricercare una sequenza all'interno di un database a seconda dell'applicazione di interesse
- conoscere i principali genome browser e comprendere la loro struttura ed il loro utilizzo
- comprendere il significato e le implicazioni biologiche di un allineamento multiplo di sequenze
- conoscere quale è lo stato dell'arte dei metodi di predizione strutturale delle proteine
- apprendere come interpretare un albero filogenetico

Capacità di applicare conoscenza e comprensione

Gli studenti sapranno effettuare ricerche semplici e complesse di letteratura e di sequenza nei principali database bioinformatici. Acquisiranno competenze nell'utilizzo di BLAST per ricercare similarità tra sequenze in genomi, trascrittomi e vari database di sequenza, nella consultazione di genome browser e nella caratterizzazione funzionale di sequenze proteiche. Inoltre, saranno capaci di generare allineamenti multipli da analizzarsi per la ricerca di motivi conservati, siti sotto pressione selettiva e generazione di alberi filogenetici.

Autonomia di giudizio

L'autonomia di giudizio viene sviluppata tramite la preparazione all'esame, che necessita della rielaborazione e assimilazione individuale del materiale presentato in aula, oltre che tramite il coinvolgimento diretto degli studenti nelle dimostrazioni pratiche dell'utilizzo di alcuni tools bioinformatici descritti nelle lezioni precedenti.

Abilità comunicative

Gli studenti saranno stimolati ad interagire con il docente durante le lezioni, al fine di sviluppare un corretto lessico scientifico e favorire un confronto tra gli studenti stessi. Le ultime lezioni saranno principalmente caratterizzate da dimostrazioni pratiche di quanto affrontato nelle lezioni precedenti, di modo da dare un ruolo di primo piano agli studenti stessi nella risoluzione di problemi.

Capacità di apprendimento

La capacità di apprendimento è stimolata dall'approfondimento delle conoscenze apprese durante le lezioni frontali tramite interazione con il docente e tra gli studenti stessi nelle ultime lezioni del corso, che prevedono l'applicazione pratica di concetti visti teoricamente in precedenza, che daranno modo di approfondire punti poco chiari e consentire una migliore comprensione di argomenti ostici per quali le conoscenze pregresse acquisite da altri corsi non dovessero essere sufficienti a garantire una completa comprensione

Autonomia di giudizio

L'autonomia di giudizio è sviluppata nel percorso individuale di preparazione all'esame mediante l'assimilazione e la rielaborazione dei contenuti trattati nel corso.

Abilità comunicative

Lo studente dovrà essere in grado di esprimere le proprie conoscenze sui contenuti del corso utilizzando termini appropriati ed un adeguato linguaggio bioinformatico. Il test scritto prevede delle domande aperte in cui

lo studente dovrà dimostrare capacità di rielaborazione e comunicazione delle conoscenze apprese.

Capacità di apprendimento

Le capacità di apprendimento saranno valutate durante lo svolgimento del corso coinvolgendo gli studenti nella discussione degli argomenti trattati

Prerequisiti

Aver sostenuto con profitto l'esame del corso di genetica

Metodi didattici

Lezione frontale coadiuvata da presentazioni ed esercitazioni.

Eventuali cambiamenti alle modalità qui descritte, che si rendessero necessari per garantire l'applicazione dei protocolli di sicurezza legati

all'emergenza COVID19, saranno comunicati nel sito web di Dipartimento, del Corso di Studio e dell'insegnamento.

Altre informazioni

Il materiale didattico verrà messo a disposizione sulla piattaforma moodle.

Modalità di verifica dell'apprendimento

Test scritto con domande a risposta multipla, vero/falso, a scelta mutualmente esclusiva e completamento. Il punteggio viene calcolato assegnando un punteggio +1 per ogni risposta corretta e 0 per ogni risposta errata (ad eccezione delle domande che prevedono soltanto due possibilità di risposta, alle quali in caso di errore viene assegnato un -1). Il punteggio totale viene poi convertito proporzionalmente in una votazione in trentesimi.

Eventuali cambiamenti alle modalità qui descritte, che si rendessero necessari per garantire l'applicazione dei protocolli di sicurezza legati all'emergenza COVID19, saranno comunicati nel sito web di Dipartimento, del Corso di Studio e dell'insegnamento.

Programma esteso

Definizione di bioinformatica e potenziali campi di applicazione, con particolare attenzione agli ambiti genomici. Database ed archivi di dati biologici; struttura e tipi, campi principali di una entry; EBI, NCBI e DDBJ. Approfondimento sui database di letteratura Pubmed e Scopus. Ricerche semplici ed avanzate nei database. Funzione e struttura dei genome browser; UCSC ed Ensembl, NCBI genomes. Analisi di sintenia. Human protein atlas. Concetti di similarità ed omologia, geni ortologi e paraloghi. Algoritmi di allineamento dinamico, allineamenti locali e globali. Algoritmo di Needleman-Wunsch a di Smith-Waterman. Matrici di sostituzione PAM e BLOSUM e loro utilizzo. Da FASTA a BLAST e valutazione della significatività di un allineamento. BLAST, funzionamento e parametri. BLASTn, BLASTp, tBLASTn, BLASTx e tBLASTx, applicazioni particolari. Interfaccia web ed interpretazione dei risultati. BLAST e BLAT locali ed integrati in genome browsers. Concetto di profilo proteico ed Hidden Markov Models: definizione e possibili utilizzi. PSI-BLAST, DELTA-BLAST. Ricerca di domini conservati con Interproscan, organizzazione di Interpro. Pattern matching/recognition e pattern discovery; applicazione nello studio di promotori. PHI-BLAST, ELM e MEME suite. Caratterizzazione funzionale di proteine: signalP, Phobius, TargetP, SecretomeP, proP, TMHMM. Predizione di struttura secondaria: metodi GOR, Chou-Fassman e machine-learning. Predizione di strutture 3D di proteine. Metodi ab initio e metodi knowledge-based. Homology-modeling e threading. ROSETTA, CASP e CAPRI. Predizione di struttura di RNA con ViennaRNA. Allineamento multiplo di sequenze; penalità dovute a gaps, metodi di allineamento progressivo: CLUSTALW e T-COFFEE. Metodi di allineamento iterativo, MUSCLE. Metodi basati su blocchi, metodi probabilistici e metodi consensus-based. Allineamenti profilo-profilo. Valutazione della qualità di un allineamento multiplo. Identificazione di siti sotto selezione positiva e negativa. Definizione di filogenetica e filogenesi, terminologia relativa ad alberi filogenetici. Concetti di monofilia, polifilia e parafilia. Tipi di albero, radicazione e scelta degli outgroup. Scelta dei marker molecolari e del modello di evoluzione molecolare. Metodi clustering-based: NJ ed UPGMA. Alberi optimality-based. Metodi basati sui caratteri: maximum parsimony e maximum likelihood. Metodi Bayesiani. Valutazione della solidità di un albero: bootstrap e jackknife. Alberi time-calibrated e BEAST. Cenni sugli approcci filogeneomici.



Testi in inglese

Italian

Introduction to Bioinformatics management and analysis of biological informations. Databases, fields and specialization levels. Information retrieval. Alignment of nucleic acid and protein sequences. Alignment dynamic and heuristic. Substitution matrices and FASTA. BLAST: principal applications. BLAST: advanced applications. Multiple alignments, iterative and progressive algorithms. Profiles alignment. DNA pattern and functional motifs. Protein pattern and functional motifs. Protein structural modeling. Phylogenetic analysis methods (distances and character matrices)

Pascarella, Paiardini. Bioinformatica: dalla sequenza alla struttura delle proteine. Zanichelli (2011)

Valle, Helmer-Citterich, Attimonelli, Pesole. Introduzione alla Bioinformatica. Zanichelli

Lesk. Introduzione alla Bioinformatica. McGraw-Hill

Krane, Raymer. Fondamenti di Bioinformatica. Pearson

Mount D. Bioinformatics. Sequence and Genome Analysis. Second Edition. CSHL Press

Baxevanis, Oullette. Bioinformatics. A practical guide to the analysis of genes and proteins. Wiley

The objective of the course is to allow students to acquire the skills necessary to rationally and proficiently use bioinformatics methods and to correctly interpret their outcomes. Knowledge and understanding-learn to know the potential fields of application of bioinformatics-know the organization scheme of the main bioinformatics databases and learn the type of contents stored-learn which are the most appropriate tools to search a sequence in a database based on the kind of application needed-know the main genome browsers and understand their use and structure-understand the meaning and the biological implications of a multiple sequence alignment-know the state of the art of the methods for structural protein prediction Abilities to apply knowledge and understanding The students are expected to be able to carry out simple and complex searches in the main sequence and literature bioinformatics databases. They will acquire skills in the use of BLAST, learning how to perform similarity-based searches in genomes, transcriptomes and various sequence databases, in the inspection of genome browsers and in the functional characterization of protein sequences. Moreover, they will be able to perform multiple sequence alignments to be analyzed looking for conserved motifs, sites under selective pressure and phylogenetic trees generation. Judgment autonomy judgment of autonomy will be developed by the preparation for the final exam, which requires the assimilation and re-elaboration of the material presented during frontal lessons, and by the direct involvement of students in practical demonstrations about the use of some previously described bioinformatics tools. Communication skills The students will be encouraged to interact with the teacher during lessons, in order to develop an appropriate scientific language, and to enable exchanges of opinions among students themselves. The latest frontal lessons will be characterized by practical demonstrations of some tools studied in the previous lessons, to allow the students to have a primary role in problem solving. Learning skills Learning skills will be encouraged by deepening the knowledge acquired during frontal lessons through the interaction between students and teacher during the latest lessons of the course, which include the practical application of previously studied theoretical concepts, which should allow to clarify specific points of the program, enabling a better comprehension of complex topics whenever the scientific background acquired from other studies will not be deemed sufficient to guarantee the full understanding of the topic.

Making judgements

The autonomy of judgment is developed in the individual preparation for the examination through the assimilation and re-elaboration of the contents covered in the course.

Communication skills

The student must be able to express his / her knowledge about

the course contents using appropriate terms and an appropriate bioinformatics language. The written test includes open questions in which the student must demonstrate the ability to rework and communicate the knowledge learned.

Learning skills
Learning skills will be assessed during the course by involving students in the discussion of the topics covered

Having successfully passed the genetics exam

Frontal lessons, together with presentations and practical demonstrations.

Possible changes that may be required to ensure the application of the security protocols linked with the COVID-19 emergency will be communicated online.

The material for the preparation do the exam will be made available on the Moodle platform

Written test, questions with multiple answer, double (mutually exclusive) choices, true/false questions and simple question that require the completion of a sentence with keywords. The total number of points is calculated as follows: +1 for each correct answer, 0 points for each wrong answer, with the exception of questions with just two possible answers, where -1 points are given for each wrong answer. The total number of points will be then proportionally converted in a 30/30 scale.

Possible changes that may be required to ensure the application of the security protocols linked with the COVID-19 emergency will be communicated online.

Definition of bioinformatics and potential fields of application, with particular focus on genomics. Databases and biological data archives; types and structure, main fields of an entry; EBI, NCBI and DDBJ. Detailed description of literature databases, Pubmed and Scopus. Simple and complex searches in databases. Genome browser structure and functionality; USCS and ENSEMBL, NCBI genomes. Syntheny analyses. Human protein atlas. Basic concepts of similarity and homology, definition of orthologous and paralogous genes. Dynamic programming alignment algorithms. Needleman-Wunsch and Smith-Waterman algorithms. Substitution matrices, PAM and BLOSUM and their use. From FASTA to BLAST, evaluation of the significance of an alignment. BLAST, functionality and parameters. BLASTn, BLASTp, tBLASTn, BLASTx and tBLASTx, special applications. Web interface and results interpretation. BLAST and BLAST, local and integrated in genome browsers. Protein profiles and Hidden Markov Models: definition and possible uses. PSI-BLAST, DELTA-BLAST. Search of conserved domains with Interproscan, organization of Interpro. Pattern matching/recognition and pattern discovery; application in the study of promoters. PHI-BLAST, ELM and MEME suite. Functional characterization of proteins: signal, Phobius, TargetP, SecretomeP, prop, TMHMM. Secondary structure prediction: GOR, Chou-Fassman and machine-learning methods. 3D protein structural prediction. Ab initio and knowledge-based methods. Homology-modelins and threading. ROSETTA, CASP and CAPRI. RNA structure prediction using ViennaRNA. Multiple sequence alignment: gap penalties, progressive alignment methods: CLUSTALW and T-COFFEE. Iterative alignment methods: MUSCLE. Block-based methods, probabilistic and consensus-based methods. Profile-profile alignments. Evaluation of the quality of a multiple sequence alignment. Identification of sites under positive and negative selection. Definition of phylogeny and phylogenetics, tree-based phylogeny. Monophyly, polyphyly and paraphyly. Tree types, rooting and outgroup choice. Molecular marker and molecular model of evolution selection. Clustering-based methods: NJ and UPGMA. Optimality-based trees. Character-based methods:

maximum parsimony and maximum likelihood. Bayesian methods. Evaluation of the goodness of a tree: bootstrap and jackknife. Time-calibrated trees, BEAST. Basics of phylogenomics approaches.
