



Ciências
ULisboa

Biologia Computacional e Genómica

Código: 465195

ECTS: 6

Ano Letivo: 2015/16

Carga horária: T: 2:00 h; TP: 2:00 h; OT: 1:00 h;

Departamento: Biologia Animal

Área Científica: Biologia;

Objetivos da Unidade Curricular

Aquisição de conhecimentos específicos e operacionais de análise de dados genéticos e genómicos, concretamente a capacidade de autonomamente analisar, por vários métodos, e interpretar, resultados de sequências de DNA e SNPs. Construção e manipulação de matrizes de dados. Compreensão das aplicações das tecnologias genómicas e métodos de análise.

Pretende-se estimular a curiosidade dos alunos em relação à actividade científica e promover a sua capacidade de formulação de questões e a aprendizagem de técnicas de apresentação e divulgação de resultados.

Pretende-se ainda fornecer bases teóricas e práticas que permitam ao aluno vir a exercer uma actividade no âmbito da investigação científica. Procurando-se dar competências abrangentes relacionadas com a capacidade de crítica, análise e discussão de ideias numa área em permanente evolução, cultivando o espírito científico com consequente desenvolvimento da capacidade de identificar problemas pertinentes.

Pré-requisitos

Sem pré-requisitos

Conteúdos

Introdução às técnicas de análise de sequências de DNA. Conceitos teóricos e aplicações práticas de Filogenética e Filogenómica. Bases de dados genéticos, alinhamentos, construção e manipulação de matrizes de sequências de DNA. Métodos de inferência filogenética: parcimónia, distancias, máxima verosimilhança e bayesiana. Relógios moleculares.

Marcadores moleculares e Microarrays, STRs e SNPs. Propriedades e aplicações.

O genoma humano e os projectos e aplicações subsequentes. O projecto HapMap, Encode e 1000 Genomes.

O desenvolvimento das novas gerações de tecnologia de sequenciação o 454, Solexa/Illumina, SOLID, Heliscope, Pacific Bioscience, Ion Torrent e Nanopores. Resenha das aplicações abertas pelas novas tecnologias.

A abordagem genome wide association studies como aplicação do projecto HapMap ao estudo da doença Humana. Bases de dados de SNPs e sua utilização em estudos de estruturação das populações.

O estudo da detecção de selecção natural e adaptação ao nível molecular.

Descrição detalhada dos conteúdos programáticos

Componente Teórica

O curso teórico divide-se em três grandes partes. Na primeira é dada uma Introdução às técnicas de análise de sequências de DNA. Conceitos teóricos e aplicações práticas de Filogenética e Filogenómica. Na segunda parte são abordadas as novas tecnologias genómicas e suas aplicações em particular aos Humanos. Na terceira parte é dada ênfase aos marcadores moleculares e suas aplicações.

Na primeira parte são referidos os seguintes conceitos:

Breve história e Introdução à Filogenética.

Parsimonia e distâncias.

Modelos de evolução e Inferência de Máxima Verossimilhança

Inferência Bayesiana

i) Breve história e Introdução à Filogenética

O que é e para que serve a Filogenética. Com que dados se faz análise filogenética. Sequências de DNA como o material de eleição para fazer análise filogenética. Os nucleótidos sua composição e variações nas bases. O processo de sequenciação manual, o método de Sanger. A sequenciação automática em gel e em capilar. A Next generation of sequencing technology.

O DNA mitocondrial – características e limitações. Porque foi o marcador dominante na análise filogenética. Os genes mais utilizados de DNA mitocondrial, suas propriedades e critérios de selecção para estudos filogenéticos. Comparação DNA mitocondrial em animais e plantas e o cpDNA. Comparação com genes nucleares. As bases de dados de sequências de DNA.

Henning, Cavalli-Sforza e Edwards e a origem da filogenética nos anos 50 e 60. Noções básicas de filogenética: Tipos de árvores, ramos, nós, politomias, raízes, outgroups. Terminologia cladística: grupos monofiléticos, parafiléticos e polifiléticos. Homologia e homoplasia, plesiomorfias, apomorfias, simplesiomorfias, autopomorfias. Evolução convergente e paralela.

Árvores de genes e árvores de espécies.

Alinhamentos: racional e algoritmos e aplicações práticas. Needleman-Wunsch, Smith-Waterman. BLAST e Clustal como as aplicações mais usadas. O novo software de alinhamento. Exemplos numéricos. Penalidades de abertura e extensão de gaps.

ii) Parsimonia e distâncias.

Inferência filogenética. Parsimonia: racional e exemplos. Árvores de consenso: diferentes métodos e respectivos problemas. O método por majority rule.

Os problemas com a parsimonia, o long branch attraction. Decay e Ensemble indices.

O método das distâncias, racional e aplicações. Os algoritmos de UPGMA e Neighbour-Joining. Semelhanças versus relações evolutivas.

iii) Modelos de evolução e Inferência de Máxima Verossimilhança

Substituição de nucleótidos, modelos de substituição e selecção de modelos.

Substituição de nucleótidos, multiple hits, back substitutions. Saturação.

Tipos de substituições e velocidades diferentes de saturação. Matriz de probabilidade de substituição. Substituições assumidas como um processo em cadeias de Markov. Vectors de composição das bases em equilíbrio. Modelo de Jukes-Cantor, Kimura 2-parameters, Felsenstein 1981, HKY 1985 e GTR.

Pressupostos dos modelos. A distribuição gamma de substituição de nucleótidos.

A selecção de modelos. O software JModeltest. Métodos de selecção de modelos: Likelihood ratio test, Akaike information Criterion, Bayesian Information Criterion.

Inferência por Máxima Verossimilhança

A explicação mais provável para os dados observados. A probabilidade de obtenção das sequências dada uma certa hipótese filogenética. Condições para a realização de inferência filogenética: o modelo de evolução, as sequências e as árvores. Pressuposto da evolução independente entre sites e entre linhagens. O cálculo do likelihood, exemplo.

IV) Inferência Bayesiana

Princípios básicos de estatística bayesiana. Aplicação da estatística bayesiana à inferência filogenética. Os priors os likelihoods e as posterior probabilities. The acceptance ratio. A cadeias de Markov Monte Carlo. Importance sampling. Burn-in e a fase estacionária. Metropolis coupled. O MCMC ou MC3. Cadeias quentes e frias. Discussão entre a Inferência de máximo verosimilhança e a inferência bayesiana.

Na segunda parte são referidos os seguintes assuntos:

O genoma humano e os projectos e aplicações subsequentes. O projecto HapMap, Encode e 1000 Genomes.

O desenvolvimento das novas gerações de tecnologia de sequenciação o 454, Solexa/Illumina, SOLID, Heliscope, Pacific Bioscience, Ion Torrent e Nanopores.

Resenha das aplicações abertas pelas novas tecnologias.

Na terceira parte são referidos os seguintes temas:

Marcadores moleculares e Microarrays, STRs e SNPs. Propriedades e aplicações. O Linkage disequilibrium. Conceitos e medição. O Linkage disequilibrium e estudos de associação. Os Genome wide association studies. A abordagem genome wide association studies como aplicação do projecto HapMap ao estudo da doença Humana. Bases de dados de SNPs e sua utilização em estudos de estruturação das populações. O estudo da detecção de selecção natural e adaptação ao nível molecular.

Componente Teórica-Prática

Introdução ao programa da cadeira. Breve abordagem das técnicas de apresentação de seminários assim como da utilização de ferramentas de busca e gestão bibliográficas.

Introdução às bases de dados de armazenamento de sequências de DNA e aminoácidos.

Introdução à análise de dados filogenéticos. Pesquisa e downloads dos itens pretendidas. Formação de matrizes de análise.

Introdução à análise de dados com SNPs e com dados NGS. Análise de dados de acordo com os métodos e tópicos abordados nas aulas teóricas.

Componente Prática

não aplicável

Bibliografia

Recomendada

Balding DJ, Bishop M, Cannings C (2008). Handbook of statistical genetics. 3 edition edn.

Felsenstein J (2004). Inferring Phylogenies. Sinauer Associates, Inc., Sunderland.

Freeman S, Herron JC (2007). Evolutionary analysis. Fourth Edition edn. Prentice Hall Upper Saddle River, NJ,

Futuyma DJ (1998). Evolutionary Biology. Third edn. Sinauer Associates Inc., U.S.,

Outros elementos de estudo

Ficheiros pdf disponibilizados na página da cadeira, com síntese da matéria das aulas

Métodos de Avaliação

Um exame sobre a matéria teórica avalia sobretudo a capacidade de retenção e interpretação da informação ministradas nas aulas teóricas e constitui 30% da nota final do aluno. O exame prático consiste num exame individual realizado em computador e semelhante aos exercícios das aulas teórico-práticas, constituindo 20% da nota final. O seminário apresentado pelos alunos constitui 25% da nota final, a discussão de artigos científicos 15% e a participação geral na cadeira os restantes 10%.

Língua de ensino

A língua de ensino é o Português, contudo muitos dos slides apresentados nas aulas reproduzem artigos ou figuras cujas legendas estão em inglês.