Genomic Data Analysis

Genomische Datenanalyse (320.301/320.304)

2VO, 2PR = 4 Semester Stunden

Description

With the advent of the human genome project the available tools and resources has deeply impacted the field of biology. The aim of this course is to introduce students to the different tools and databases necessary for any research project in biology. The course has a laboratory module that will guide and train students with examples and a series of exercises leading students through the different steps and tools found in biological databases.

Goals

1. To provide an introduction to genomic databases with a focus on the National Center for Biotechnology Information (NCBI), UCSC, and EBI
2. To focus on the analysis of DNA and proteins
3. To introduce the student to the analysis of genomes
4. To combine theory and practice to help students solving common research problems in biology with the resources and information available in different online databases.

Syllabus

Part 1 – Introduction to genomics

1. Introduction to genomics-3 case studies
    - Definition of bioinformatics/genomics
    - The Human Genome Project-the start of genomics
    - Sequencing the human genome
    - Assembly: paired-end and shotgun sequencing
    - Main conclusions of the human genome project

2. Genomic variation
    - Genomic variation
    - From SNPs to copy number variants and their evolution
    - HapMap project
    - Uses of SNPs

3. Genome projects/ Comparative genomics
    - Methods to detect genomic variation

- Sequencing Projects
- Understanding a genome sequence
- Structural features of a genome

4. Emerging sequencing technologies
    - New sequencing technologies (NGS)
    - Principles of next generation sequencing technologies
    - Commercial platforms
    - Uses of NGS
    - Individual genomes

5. Application of genomics
    - 3 study cases—why is genomics important?
        Genetics perspective-cure diseases
        Synthetic biology: build your own genome
        Evolutionary biology: Where do we come from?
    - Commercializing genomics
    - Ethical aspects

Part 2 – Introduction to databases (lab based)

1. Accessing information about DNA and proteins
    - Overview of the NCBI website
    - Accessing information: accession numbers, RefSeq, FASTA sequence, genome assembly
    - NCBI databases: Gene, CCDS, UniGene, Taxonomy, Nucleotide, Protein
    - Database to get information about genetic diseases: OMIM

2. Literature search, polymorphisms and PCR
    - Literature search in PubMed
    - Definition of SNP, allele, genotype, haplotype
    - SNP database and Hardy-Weinberg Equilibrium
    - What is a Polymerase-Chain Reaction and how does it work?
    - Primer design with Primer3Plus
    - Design of restriction enzyme digests using NEBCutter

3. Genome Browser and sequence alignments
    - Two genome browser: UCSC, ENSEMBL
    - BLAST – Basic local alignment search tool
    - How to use BLAST
    - Definitions: homologs, paralogs, orthologs
    - Scoring Matrices
    - Pairwise sequence alignment
    - How to interpret BLAST results

- PrimerBLAST

4. Protein analysis
    - Protein structure
    - Protein Databases: UniProt, ExPASy
    - The protein data bank (PDB)
    - How to use JmolS

5. Revision and final report

Methods

The course will be taught in two parts. The first part will focus on the theoretical background of genomics including topics in genetics, molecular biology, and biochemistry. The second part will provide an introduction to the databases with step to step examples of how to retrieve different information. During the laboratory module students will solve a series of problems based on the taught material.

Grading

- 50% Final exam (short answer / multiple choice questions based on the material learned during the course and the lab modules)
- 10% Small assays (1-2 pages; single space; 11pts font)
- 20% Labwork (results from computer lab)
- 15% Final report of computer lab session 5, due at predefined date following the class
- 5% Attendance/Participation in class

Recommended textbooks

1. Bioinformatics and Functional Genomics by Jonathan Pevsner (Wiley-Blackwell, 2nd edition 2009).

2. A Primer of Genome Science by Greg Gibson (Spencer V. Muse Publisher: Sinauer Associates, 3rd Edition 2008)