

## EMBER syllabus

### Bioinformatics I

#### Introduction

- What is bioinformatics?
- Definitions and concepts
- Genome projects
- Biological complexity
- The role of bioinformatics

#### Biological Databases

- Sequence databases (EMBL, GenBank, DDBJ, SWISS-PROT, PIR, TrEMBL)
- Protein family/domain databases (PROSITE, PRINTS, Pfam, SMART, etc)
- Cluster databases (ProDom, Systers, etc)
- Specialist databases (FlyBase, Kegg, etc)
- Database technologies (flat-file, relational, object)
- Search engines (SRS, Entrez)

#### Principles of Sequence Analysis I - Pairwise Comparison

- Sequence comparison algorithms
- Sequence comparison scoring systems
- Sequence database similarity searching algorithms (BLAST & FASTA family of programs)
- Similarity searching scores and their statistical interpretation

#### Principles of Sequence Analysis II - Motifs and Domains

- Algorithms for global multiple alignment
- Biological motifs (consensus, regular expressions, profiles, PSSMs, HMMs)
- The local multiple alignment problem and motif interference
- Applications for biological sequence similarity searching (PSI- & PHI-BLAST, motifs, patterns, databases)

#### Functional Genomics I - The Genome

- Data production and data flow (mapping, DNA sequencing, generation of scaffolds & contigs)
- Gene prediction (*ab initio* & similarity based)
- Genome annotation (pipelines, databases)
- Technology platforms

### Bioinformatics II

#### Functional Genomics II - The Transcriptome

- Strategies for generating ESTs and full length inserts
- EST clustering and assembly
- EST databases (DBEST, UNIGene, TGI, STACK, EGI)
- SAGE
- Statistical analysis of EST and EST data
- Microarrays (target selection/design, image analysis, data validation, stat. analysis)
- Data resources

#### Functional Genomics III - The Proteome

- 2D gel data (image analysis)
- Mass spec data (principles, analysis, integration of info & peptide sequence determination)

#### Molecular Evolution and Phylogeny

- Biological foundations and phylogenetic models
- Transformation of biological characteristics and associated probabilities
- Terminology (homology, homoplasy, orthology & paralogy)
- Methodologies (Cladistic MP, bootstrapping, branch & bound methods)

#### Ontologies in Bioinformatics

- The need for ontologies (Gene naming, functional classifications, references schemes)
- Gene ontology (GO model & tools)
- EcoCyc

**Principles of Protein Structure Prediction**

- Secondary structure analysis (secondary structure profiles)
- Protein structure and fold classification databases (PDB, CATH, SCOP)
- Principles of molecular dynamics
- *Ab initio* prediction
- Homology modelling
- Threading

**Basic Informatics**

- Introduction to information theory
- Basic statistics