

Programma del Corso di Biofisica Computazionale aa. 2016/2017 parte 1

I sistemi biologici

La cellula

Cellule Eucariotiche e procariotiche

Struttura cellulare

- Membrana cellulare
- Fosfolipidi
- Idrocarburi saturi e insaturi
- Genomi, Cromosomi e geni

Dogma centrale

Evoluzione

Energia libera, entalpia e entropia

Effetto idrofobico

Il materiale genetico

L'acido ribonucleico o RNA

- mRNA
- tRNA
- rRNA
- Small RNA
- Long non coding RNA

Differenza tra DNA e RNA

Ruoli dell'RNA

La struttura di un aminoacido

Il legame peptidico

I venti aminoacidi

Geni e genomi

La struttura di un gene

Lo splicing

Lo splicing alternativo

Il genoma umano

La replicazione del DNA

Polymerase chain reaction (PCR)

Metodi di sequenziamento del DNA

- Sequenziamento con il metodo di Sanger
- Next generation sequencing (NGS)

Qualità dei dati

Assemblaggio delle "reads"

- Grafi di de Bruijn
- I percorsi Euleriani

Ambiguità e errori di sequenza

Possibili problemi

Applicazioni dell'NGS

- Confronto tra trascrittomi (RNASeq)
- ChipSeq
- miRSeq

Quantificazione

Banche dati utili

- Gene Ontology
- KEGG (pathways)
- Interpro (famiglie)

Distribuzione ipergeometrica

Arricchimento funzionale dei geni up- down – regolati:

Ricerca di geni

Il modello biologico: gene procariota

Ricerca di geni in procarioti

- Identificazione di motivi specifici (per es. promotori e TATA box)
- Lunghezza dell'open reading frame
- Uso codoni
- Periodicità e proprietà statistiche

Il modello biologico: gene eucariota

Similarità con geni noti nello stesso o in altri organismi

Ricerca di segnali di sequenza (motivi)

- Espressioni regolari
- Matrici sito specifiche
- Modelli piu' complessi
- Entropia di Shannon
- Sequence logo
- Gibbs sampling
- MEME

Modelli statistici del gene o delle regioni che contengono i geni

- Modelli di Markov
- Hidden Markov Models
- Algoritmo di Viterbi

Misure di accuratezza

Curva ROC

Allineamento di sequenze

Distanza tra sequenze

Matrici di sostituzione

- Matrici di Dayhoff
- Matrici BLOSUM

Penalizzazione dei GAP

Algoritmo di allineamento globale: programmazione dinamica (Needleman and Wunsch)

Allineamento locale (Smith and Waterman)

Ricerca di omologhi

FASTA

BLAST1

BLAST2: (NCBI)

Statistica dei punteggi di similarità
Distribuzione dei valori estremi
Curva ROC

Allineamenti multipli (MSA)

Vantaggi degli MSA

Metodi per ottenere un MSA

Soluzioni euristiche:

- Allineamento iterativo
- Metodo Center Star
- Metodo Center Star
- MSA con T-coffee

Profili

PSI-Blast

Modelli di Markov

Modelli nascosti di Markov

Valutazione della significatività del punteggio

Visualizzazione di famiglie in Pfam

Alberi filogenetici

Paralogia e ortologia

Metodi di classificazione e clustering

Clustering

Alberi

- Etichettatura
- Topologia
- Alberi rooted e unrooted
- Costruzione di un albero

UPGMA

Metodi di costruzione e ricerca dell'albero

- Neighbour Joining
- Massima Parsimonia
- Branch and bound

Qualità di un albero: bootstrap

Altri metodi di classificazione

- Reti neurali
- Support vector machines
- Alberi decisionali
- Forest tree
- PCA

Gini index: valutazione impurità delle suddivisioni

Esempio: Compressione di immagini

Esempio: Riconoscimento di visi

La struttura delle proteine

Angoli ϕ e ψ

Grafico di Ramachandran

Struttura secondaria

Ponti disolfuro

Domini proteici

La banca dati PDB (Protein Data bank)

Classificazione strutturale di proteine