

HOMOLOGENE

HomoloGene è un sistema automatizzato per l'identificazione di geni omologhi tra quelli annotati e completamente sequenziati dei genomi eucarioti.

FUNZIONAMENTO:

Il metodo adottato per l'identificazione di geni omologhi da una query, parte dal confronto delle sequenze amminoacidiche del prodotto proteico del gene in questione con quella dei geni contenuti nel database, tramite l'utilizzo di BLASTP.

Essendo la sequenza di amminoacidi maggiormente conservata in organismi diversi rispetto alla sequenza nucleotidica, basandosi la ricerca di HomoloGene sulla struttura primaria della proteina anziché sulla sequenza di DNA, è possibile rintracciare alcuni casi di omologie remote che altrimenti verrebbero ignorati.

In base all'identità di sequenza, i geni vengono suddivisi in gruppi usando un albero costruito tramite le similarità di sequenza così che gli organismi maggiormente correlati siano uniti per primi e successivamente gli altri.

La sequenza proteica viene quindi convertita in quella nucleotidica corrispondente così da identificarne la distanza molecolare usando, per esempio, come unità di riferimento il rapporto tra K_a e K_s (rispettivamente sostituzioni sinonimo e non sinonimo avvenute tra le due sequenze, corrette per tener conto di eventuali sostituzioni multiple tramite modelli matematici).

Quando possibile, nell'appaiamento delle sequenze viene anche tenuta in considerazione la sintenia tra i diversi geni.

Alcune tra le hits rimanenti sono unite tra loro attraverso un'analisi di similarità che tenda a massimizzare il punteggio di un allineamento globale tra le sequenze e che eviti che falsi ortologhi vengano raggruppati assieme.

Vengono identificati anche i paraloghi tramite la ricerca di sequenze che sono più vicine all'interno della stessa specie dell'organismo d'input rispetto le altre.

RICERCA

Nell'home-page viene subito mostrata una tabella indicante gli organismi presi in considerazione per la ricerca di omologie da parte di HomoloGene e per ognuno viene indicato il numero di geni identificati nel genoma, il numero di quelli raggruppati per omologia e il numero di gruppi stabiliti (fig.1.a).

Supponiamo di essere alla ricerca di geni omologhi al gene per l'alcool deidrogenasi. Nell'apposito spazio per la query (fig.1.b) è richiesto l'inserimento di una dei seguenti tipi di query:

- ✓ nome del gene;
- ✓ numero del gene;
- ✓ parola chiave.

Nel nostro caso inseriremo “*alcohol dehydrogenase*”

Il tool consente, inoltre, di filtrare i nostri risultati applicando dei limits al campo di ricerca (fig.2).

Cliccando, dunque, su “LIMITS” e sui rispettivi menù a tendina in basso, è possibile perfezionare la nostra ricerca limitandola ai seguenti campi:

- ✓ Ancestor (nome dell'ancestrale comune);
- ✓ Domain Name;
- ✓ Gene description;
- ✓ Gene ID;
- ✓ Gene name;
- ✓ Nucleotide accession (codice d'accesso per Nucleotide);
- ✓ Nucleotide UID (ID per Nucleotide);
- ✓ Organism;
- ✓ Properties;
- ✓ Protein accession: (codice d'accesso per Protein);
- ✓ Protein UID: (ID per Protein);
- ✓ Text word;
- ✓ Title;
- ✓ UniGene ID.

In figura 2 sono mostrate le altre limits possibili.

Avviando la ricerca (in questo caso senza utilizzare limits) quello che si ottiene è una prima serie di risultati che illustrano i diversi gruppi d'analogia identificati.

Supponendo di voler scegliere il primo, cliccando su di esso, appare la schermata contenente le informazioni sul gruppo.

Queste sono raggruppate in base alle seguenti categorie:

- ✓ Genes: (fig.3) .
- ✓ Proteins: (fig.4).
- ✓ Proteins Allineaments: (fig.5) .
- ✓ Phenotypes: fornisce informazioni sul fenotipo del gene query. Tali informazioni vengono ricavate da database di organismi modello.
- ✓ Conserved Domains: (fig.6).
- ✓ Related Homology Resources: (fig.7).
- ✓ Pubmed: (fig.8).
- ✓ UniGene: (fig 9).

Essendo HomoloGene collegato con tutti gli altri database Entrez (fig.12), links a questi sono disponibili dal menù a tendina in alto a destra della pagina dei risultati.

Infine, le sequenze ricavate, inoltre, possono essere scaricate in formato FASTA grazie al link "DOWNLOAD" in figura 10.

Da questo comando è possibile scegliere se scaricare la proteina, la sequenza mRNA o quella genomica e in questo ultimo caso è possibile specificare se includere le sequenze up/downstream del gene (fig.11).