

Esecitazione 1 - Caccia al batterio:

I condotti della rete idrica pubblica sono stati immersi in acqua per decenni e una comunità di microrganismi ha prosperato in tutto questo tempo sulla loro superficie.

Questi rivestimenti sono indicati come *biofilm* e la sua composizione batterica è generalmente sconosciuta perché gli scienziati non sono in grado di studiare la stragrande maggioranza di questi organismi in laboratorio. Nel 2003, Schmeisser e il suo team collezionarono e sequenzuarono il DNA dai batteri che crescevano sulle valvole dei tubi delle rete idrica potabile della Germania del Nord. Attraverso la ricerca di similarità nelle varie sequenze essi furono in grado di classificare un ampio numero di questi organismi come appartenenti a particolari gruppi. In questo processo essi identificarono molte nuove specie.

In questa esercitazione useremo BLAST per ripetere alcune delle loro analisi e identificare la composizione di questi biofilm.

Di seguito viene riportata una lista di 10 *sequence accession numbers* presi dai loro studi. E' possibile utilizzare NCBI BLASTN Web per ricercare le similarità e provare ad identificare i batteri che crescono all'interno di questi biofilm.

AY187314

AY187315

AY187316

AY187317

AY187318

AY187325

AY187326

AY187330

AY187332

AY187333

1. Recuperare ogni sequenza dalla NCBI GenBank e, basandosi sull'annotazione di ogni record, identificare quali geni sono stati usati nelle loro analisi.
2. Per ogni sequenza, convertire il formato del file in FASTA utilizzando le impostazioni di visualizzazione.
3. Navigare nella NCBI BLASTN Web form e incollare il formato FASTA di ogni sequenza di DNA nella apposita finestra.
4. Scegliere la "Nucleotide collection (nr/nt)" come database per la ricerca.
5. Per risparmiare tempo possiamo limitare la ricerca ai soli "bacteria (taxid:2)" nel campo "Organism"
6. Utilizzare "Somewhat similar sequences (BLASTN)" come programma per la ricerca.
7. Quando pronti, effettuare la ricerca cliccando sul pulsante "BLAST".
8. Aprire più finestre del browser e lanciare altre ricerche.
9. Dieci finestre di risultati saranno pronte in pochi minuti. Organizzate l'ambiente per registrare le vostre conclusioni per ogni accession number.

10. Per ogni ricerca, memorizzare i risultati grafici, tabelle e allineamenti in modo da assegnare ogni sequenza sconosciuta ad un organismo conosciuto. Notare che il primo *hit* potrebbe anche essere un organismo sconosciuto, quindi ci toccherà esaminare tutti gli hits prima di trarre qualunque conclusione relativa al tipo di batterio da cui proviene la sequenza.
11. Usando il NCBI PubMed database o altre risorse disponibili sul Web, provate a trovare informazioni di base circa il genere e/o la specie; per esempio, l'habitat dove i batteri comunemente vivono, e se essi sono associati con qualche patologia o forma di inquinamento.

Esercitazione 2: RuBisCO

Il gene della *ribulose bispfosfato carbossilasi* (ribulose biphosphate carboxylase), per gli amici RuBisCO, codifica per un enzima che è parte del ciclo di Calvin ed è l'enzima chiave nella incorporazione di carbonio da anidride carbonica in un organismo vivente. Esso è parte di un complesso enzimatico trovato nelle piante, terrestri o acquatiche, e probabilmente ha giocato un importante ruolo nello sviluppo della nostra atmosfera e della vita sulla terra.

L' *Arabidopsis thaliana*, un membro della famiglia delle Brassicaceae, è un importante modello per lo studio di piante superiori. E' facilmente coltivabile in laboratorio, in quanto si sviluppa molto in fretta, e produce un elevato numero di semi, rendendola di fatto un ottimo soggetto per gli studi genetici. Anche se non importante da un punto di vista dell'agricoltura, l' *Arabidopsis* ha fornito conoscenze fondamentali per la biologia ed il suo fu il primo genoma di pianta ad essere sequenziato (nel 2000).

In questa esercitazione, useremo BLASTN per identificare membri della famiglia del gene RuBisCO nelle Arabidopsis.

1. Recuperare l'mRNA di riferimento per l'Arabidopsis RuBisCO small chain subunit 1b, NM_123204, dal sito dell'NCBI.
2. Cambiare il formato in FASTA ed incollare la sequenza nella apposita form per le query.
3. Impostare il database su "Reference RNA sequences (refseq_rna)" e restringere l'organismo a "Arabidopsis thaliana (taxid:3202)."
4. Impostare il programma su "Somewhat similar sequences (BLASTN)" e cliccare sul pulsante BLAST per avviare la ricerca.
5. Una volta ottenuti i risultati, utilizzare i grafici, le tabelle e gli allineamenti per identificare i membri della famiglia.
6. Il database di riferimento *non dovrebbe* avere ridondanze, ma due membri della famiglia presentano sequenze alternate di mRNA. Comparare gli allineamenti accuratamente ed esaminare le annotazioni (specialmente le coordinate delle regioni codificate) di tutte le sequenze rilevanti per descrivere e capire le maggiori differenze fra queste trascrizioni di membri della famiglia.
7. Creare una tabella con una lista di nomi delle trascrizioni dei membri della famiglia e i loro accession number, la lunghezza dell' mRNA, le coordinate della coding region (CDS)