Esercitazione di bioinformatica: le sequenze di DNA sono stringhe di caratteri

Utilizziamo le sequenze di DNA per risolvere un caso di epidemia!

Cosa è successo? Guardiamo il seguente video su you tube: https://youtu.be/WoNso1JnL5Y. Il video è anche disponibile tra i materiali del modulo.
Nella città di New York si stanno verificando una serie di casi di tubercolosi. Alcuni malati vengono ricoverati nell'ospedale dove lavora un vostro amico medico, il quale tratta i pazienti con l'antibiotico Rifampicina. Mentre molti dei pazienti mostrano immediatamente segni di miglioramento, alcuni non sembrano invece rispondere al farmaco. Per capire la ragione del problema e riuscire a curare i pazienti che peggiorano, il vostro amico contatta il servizio di microbiologia dell'ospedale e gli consegna i campioni prelevati da 5 pazienti che non rispondono all'antibiotico e da 5 pazienti che invece stanno rispondendo alle cure. Nel laboratorio di microbiologia i batteri che infettano ciascuno dei pazienti vengono coltivati e testati per vedere se sopravvivono all'antibiotico. In effetti i batteri che infettano i 5 pazienti che non rispondono alle cure proliferano in coltura, nonostante la presenza dell'antibiotico, mentre quelli prelevati dai pazienti in via di miglioramento non sopravvivono in coltura, in seguito alla somministrazione dell'antibiotico.
Da 2 colture batteriche (una di un paziente che risponde all'antibiotico e una di un paziente che non risponde) viene estratto il DNA e ne viene determinata la sequenza al fine di indagare meglio questo comportamento.
Le 2 sequenze ottenute vengono inviate proprio a voi per vedere se, da quelle stringhe di caratteri, si riescono ad estrapolare informazioni utili.

Esercizio 1.

Cosa hanno di diverso le sequenze dei batteri che infettano i pazienti che non rispondono alle cure?

Iniziamo confrontando il risultato del sequenziamento per vedere se ci sono e quali sono le differenze.

DNA tubercolosi sensibile

>sequenza del batterio sensibile all'antibiotico
CGGGTATTTGAGGAATGCGCCGTCCTGCGCTATTGTTGGACGTTGCGCTGGCTACTTCCTGCCCACCTCACCCGCCACTT
GACACCGTGGTCTTAGTCTGAGCCCAGTTTGCGGCTCAGCGGTTTAGTTGCGTGCGTGAGATCCGGACAGATCGTTCGCC
GGCCGAAACCGACAAAATTATCGCGGTTGACAGGCCCGTGGGCACCGCTCCTATAATGGCTCTCGTTGGTCGCATGAAGT
GCAGGAGGGATGCATCTTGGCAGATTCCCGCCAGAGCAAAACAGCCGCTAGTCCTAGTCCGAGTCGCCCGCAAAGTTCCT
CGAATAACTCCGTACCCGGAGCGCCAAACCGGGTCTCCTTCGCTAAGCTGCGCGAACCACTTGAGGTTCCGGGACTCCTT
GACGTCGAGACGGATTCGTTCGAGTGGCTGATCGGTTGCGCCGCGTGGCGCGAATCCGGCGCGGGGGTGATGTCAA
CCCAGTGGGTGGCCTGGAAGAGGTGCTCTACGAGCTGTCTCCGATCGAGGACTTCTCCGGGTCGATGTCGTTGTCGTTCT
CTGACCCTCGTTTCGACGATGTCAAGGCACCCGTCGACGAGTGCAAAGACAAGGACATGACGTACGCGGCTCCACTGTTC
GTCACCGCCGAGTTCATCAACAACAACACCGGTGAGATCAAGAGTCAGACGGTGTTCATGGGTGACTTCCCGATGATGAC
CGAGAAGGGCACGTTCATCATCAACGGGACCGAGCGTGTGGTGGTCAGCCAGCTGGTGCGGTCGCCCGGGGTGTACTTCG
ACGAGACCATTGACAAGTCCACCGACAAGACGCTGCACAGCGTCAAGGTGATCCCGAGCCGCGGCGCGTGGCTCGAGTTT
GACGTCGACAAGCGCGACACCGTCGGCGTGCGCATCGACCGCAAACGCCGGCAACCGGTCACCGTGCTGCTCAAGGCGCT
GGGCTGGACGCCGAGCAGATTGTCGAGCGGTTCGCGGATCATGCGACGCTGAAGCAAGGACAACACCG
TCGGCACCGATGAGGCGCTGTTGGACATCTACCGCAAGCTGCGTCCGGCGAGCCCCCGACCAAAGAGTCAGCGCAGACG
CTGTTGGAAAACTTGTTCTTCAAGGAGAAGCGCTACGACCTGGCCCGCGTCGGTCGCTATAAGGTCAACAAGAAGCTCGG
GCTGCATGTCGGCGAGCCCATCACGTCGTCGACGCTGACCGAAGAAGACGTCGTGGCCACCATCGAATATCTGGTCCGCT
TGCACGAGGGTCAGACCACGATGACCGTTCCGGGCGGCGTCGAGGTGCCGGTGGAAACCGACGACATCGACCACTTCGGC
AACCGCCGCCTGCGTACGGTCGGCGAGCTGATCCAAAACCAGATCCGGGTCGGCATGTCGCGGATGGAGCGGGTGGTCCG
GGAGCGGATGACCACCCAGGACGTGGAGGCGATCACACCGCAGACGTTGATCAACATCCGGCCGGTGGTCGCCGCGATCA
AGGAGTTCTTCGGCACCAGCCAGCTGAGCGAATTCATGGACCACGAACAACCCGCTGTCGGGGTTGACCACAAGGCGCGA
CTGTCGGCGCTGGGGCCCGGCGGTCTGTCACGTGAGCGTCCGGGCTGGAGGTCCGCGACGTGACCCCGTCGCACTACG
GCCGGATGTGCCCGATCGAAACCCCTGAGGGGCCCAACATCGGTCTGATCGGCTCGCTGTCGGTGTACGCGCGGGTCAAC
C
CGTTCGGGTTCATCGAAACGCCGTACCGCAAGGTGGTCGACGGCGTGGTTAGCGACGAGATCGTGTACCTGACCGCCGAC
GAGGAGGACCGCCACGTGGTGGCACAGGCCAATTCGCCGATCGATGCGACGGTCGCTTCGTCGAGCCGCGCGTGCTGGT
CCGCCGCAAGGCGGGCGAGGTGGAGTACGTGCCCTCGTCTGAGGTGGACTACATGGACGTCTCGCCCCGCCAGATGGTGT
CGGTGGCCACCGCGATGATTCGCTTCCTGGAGCACGAGCACGCCAACCGTGCCCTCATGGGGGCAAACATGCAGCGGCAG
GCGGTGCCGCTGGTCCGTAGCGAGGCCCCCGCTGGTGGGCACCGGGATGGAGCTGCGCGCGGCGATCGACGCCGGCGACG
TCGTCGTCGCCGAAGAAAGCGGCGTCATCGAGGAGGTGTCGGCCGACTACATCACTGTGATGCACGACAACGGCACCCGGC
GTACCTACCGGATGCGCAAGTTTGCCCGGTCCAACCACGGCACTTGCGCCAACCAGTGCCCCATCGTGGACGCGGGCGACC
GAGTCGAGGCCGGTCAGGTGATCGCCGACGGTCCCTGTACTGACGACGGCGAGATGGCGCTGGGCAAGAACCTGCTGGTG
GCCATCATGCCGTGGGAGGGCCACAACTACGAGGACGCGATCATCCTGTCCAACCGCCTGGTCGAAGAGGACGTGCTCACC
TCGATCCACATCGAGGAGCATGAGATCGATGCTCGCGACACCAAGCTGGGTGCGGAGGAGATCACCCGCGACATCCCGAAC
ATCTCCGACGAGGTGCTCGCCGACCTGGATGAGCGGGGCATCGTGCGCATCGGTGCCGAGGTTCGCGACGGGGACATCCT
GGTCGGCAAGGTCACCCGAAGGGTGAGACCGAGCTGACGCCGGAGGAGCGGCTGCTGCGTGCCATCTTCGGTGAGAAGG
CCCGCGAGGTGCGCGACACTTCGCTGAAGGTGCCGCACGGCGAATCCGGCAAGGTGATCGGCATTCGGGTGTTTTCC
CGCGAGGACGAGGACGAGTTGCCGGCCGGTGTCAACGAGCTGGTGCGTGTGTATGTGGCTCAGAAACGCAAGATCTCCGA
CGGTGACAAGCTGGCCGGCCGGCACGGCAACAAGGGCGTGATCGGCAAGATCCTGCCGGTTGAGGACATGCCGTTCCTTG
CCGACGGCACCCCGGTGGACATTATTTTGAACACCCACGGCGTGCCGCGACGGATGAACATCGGCCAGATTTTGGAGACC
CACCTGGGTTGGTGTGCCCACAGCGGCTGGAAGGTCGACGCCGCCAAGGGGGTTCCGGACTGGGCCGCCAGGCTGCCCG
ACGAACTGCTCGAGGCGCAGCCGAACGCCATTGTGTCGACGCCGGTGTTCGACGGCGCCCAGGAGGCCGAGCTGCAGGGC
CTGTTGTCGTGGCAGCTGCCCAACCGGACGGCGACAGGGTGACGTGCTCTCGACGGGCG
CAGCGGCGAGCCGTTCCCGTACCCGGTCACGGTTGGCTACATGTACATCATGAAGCTGCACCACCTGGTGGACGACAAGAT
CCACGCCCGCTCCACCGGGCCGTACTCGATGATCACCCAGCAGCCGCTGGGCGGTAAGGCGCAGTTCGGTGGCCAGCGGT
TCGGGGAGATGGAGTGCTGGGCCATGCAGGCCTACGGTGCTGCCTACACCCTGCAGGAGCTGTTGACCATCAAGTCCGAT
GACACCGTCGGCCGCGTCAAGGTGTACGAGGCGATCGTCAAGGGTGAGAACATCCCGGAGCCGGGCATCCCCGAGTCGTT
CAAGGTGCTGCTCAAAGAACTGCAGTCGCTGTGCCTCAACGTCGAGGTGCTATCGAGTGACGGTGCGGCGATCGAACTGC
GCGAAGGTGAGGACGAGGACCTGGAGCGGGCCGGGCCAACCTGGGAATCAATCTGTCCGCAACGAATCCGCAAGTGTC
GAGGATCTTGCGTAAAGCTGTCGCAAAATTACTAAACCGGTTAGGGGAAAGGGAGTTACGTGCTCGACGTCAACTTCTTC
GATGAACTCCGCATCGGTCTTGCTACCGCGGAGGACATCAGGCAATGGTCCTATGGCGAGG

DNA tubercolosi resistente:

>sequenza del batterio resistente all'antibiotico
CGGGTATTTGAGGAATGCGCCGTCCTGCGCTATTGTTGGACGTTGCGCTGGCTACTTCCTGCCCACCTCACCCGCCACTT
GACACCGTGGTCTTAGTCTGAGCCCAGTTTGCGGCTCAGCGGTTTAGTTGCGTGCGTGAGATCCGGACAGATCGTTCGCC
GGCCGAAACCGACAAAATTATCGCGGTTGACAGGCCCGTGGGCACCGCTCCTATAATGGCTCTCGTTGGTCGCATGAAGT
GCAGGAGGGATGCATCTTGGCAGATTCCCGCCAGAGCAAAACAGCCGCTAGTCCTAGTCCGAGTCGCCCGCAAAGTTCCT
CGAATAACTCCGTACCCGGAGCGCCAAACCGGGTCTCCTTCGCTAAGCTGCGCGAACCACTTGAGGTTCCGGGACTCCTT
GACGTCGAGACGGATTCGTTCGAGTGGCTGATCGGTTGCGCCGCGTGGCGCGAATCCGGCGCGGGGGTGATGTCAA
CCCAGTGGGTGGCCTGGAAGAGGTGCTCTACGAGCTGTCTCCGATCGAGGACTTCTCCGGGTCGATGTCGTTGTCGTTCT
CTGACCCTCGTTTCGACGATGTCAAGGCACCCGTCGACGAGTGCAAAGACAAGGACATGACGTACGCGGCTCCACTGTTC
GTCACCGCCGAGTTCATCAACAACAACACCGGTGAGATCAAGAGTCAGACGGTGTTCATGGGTGACTTCCCGATGATGAC
CGAGAAGGGCACGTTCATCATCAACGGGACCGAGCGTGTGGTGGTCAGCCAGCTGGTGCGGTCGCCCGGGGTGTACTTCG
ACGAGACCATTGACAAGTCCACCGACAAGACGCTGCACAGCGTCAAGGTGATCCCGAGCCGCGGCGCGTGGCTCGAGTTT
GACGTCGACAAGCGCGACACCGTCGGCGTGCGCATCGACCGCAAACGCCGGCAACCGGTCACCGTGCTGCTCAAGGCGCT
GGGCTGGACGCCGAGCAGATTGTCGAGCGGTTCGCGGATCATGCGACGCTGAAGCAAGGACAACACCG
TCGGCACCGATGAGGCGCTGTTGGACATCTACCGCAAGCTGCGTCCGGCGAGCCCCCGACCAAAGAGTCAGCGCAGACG
CTGTTGGAAAACTTGTTCTTCAAGGAGAAGCGCTACGACCTGGCCCGCGTCGGTCGCTATAAGGTCAACAAGAAGCTCGG
GCTGCATGTCGGCGAGCCCATCACGTCGTCGACGCTGACCGAAGAAGACGTCGTGGCCACCATCGAATATCTGGTCCGCT
TGCACGAGGGTCAGACCACGATGACCGTTCCGGGCGGCGTCGAGGTGCCGGTGGAAACCGACGACATCGACCACTTCGGC
AACCGCCGCCTGCGTACGGTCGGCGAGCTGATCCAAAACCAGATCCGGGTCGGCATGTCGCGGATGGAGCGGGTGGTCCG
GGAGCGGATGACCACCCAGGACGTGGAGGCGATCACACCGCAGACGTTGATCAACATCCGGCCGGTGGTCGCCGCGATCA
AGGAGTTCTTCGGCACCAGCCAGCTGAGCGAATTCATGGACCACGAACAACCCGCTGTCGGGGTTGACCACAAGGCGCGA
CTGCCGGCGCTGGGGCCCGGCGGTCTGTCACGTGAGCGTCCGGGCTGGAGGTCCGCGACGTGACCCCGTCGCACTACG
GCCGGATGTGCCCGATCGAAACCCCTGAGGGGCCCAACATCGGTCTGATCGGCTCGCTGTCGGTGTACGCGCGGGTCAAC
C
CGTTCGGGTTCATCGAAACGCCGTACCGCAAGGTGGTCGACGGCGTGGTTAGCGACGAGATCGTGTACCTGACCGCCGAC
GAGGAGGACCGCCACGTGGTGGCACAGGCCAATTCGCCGATCGATGCGACGGTCGCTTCGTCGAGCCGCGCGTGCTGGT
CCGCCGCAAGGCGGGCGAGGTGGAGTACGTGCCCTCGTCTGAGGTGGACTACATGGACGTCTCGCCCCGCCAGATGGTGT
CGGTGGCCACCGCGATGATTCGCTTCCTGGAGCACGAGCACGCCAACCGTGCCCTCATGGGGGCAAACATGCAGCGGCAG
GCGGTGCCGCTGGTCCGTAGCGAGGCCCCCGCTGGTGGGCACCGGGATGGAGCTGCGCGCGGCGATCGACGCCGGCGACG
TCGTCGTCGCCGAAGAAAGCGGCGTCATCGAGGAGGTGTCGGCCGACTACATCACTGTGATGCACGACAACGGCACCCGGC
GTACCTACCGGATGCGCAAGTTTGCCCGGTCCAACCACGGCACTTGCGCCAACCAGTGCCCCATCGTGGACGCGGGCGACC
GAGTCGAGGCCGGTCAGGTGATCGCCGACGGTCCCTGTACTGACGACGGCGAGATGGCGCTGGGCAAGAACCTGCTGGTG
GCCATCATGCCGTGGGAGGGCCACAACTACGAGGACGCGATCATCCTGTCCAACCGCCTGGTCGAAGAGGACGTGCTCACC
TCGATCCACATCGAGGAGCATGAGATCGATGCTCGCGACACCAAGCTGGGTGCGGAGGAGATCACCCGCGACATCCCGAAC
ATCTCCGACGAGGTGCTCGCCGACCTGGATGAGCGGGGCATCGTGCGCATCGGTGCCGAGGTTCGCGACGGGGACATCCT
GGTCGGCAAGGTCACCCGAAGGGTGAGACCGAGCTGACGCCGGAGGAGCGGCTGCTGCGTGCCATCTTCGGTGAGAAGG
CCCGCGAGGTGCGCGACACTTCGCTGAAGGTGCCGCACGGCGAATCCGGCAAGGTGATCGGCATTCGGGTGTTTTCC
CGCGAGGACGAGGACGAGTTGCCGGCCGGTGTCAACGAGCTGGTGCGTGTGTATGTGGCTCAGAAACGCAAGATCTCCGA
CGGTGACAAGCTGGCCGGCCGGCACGGCAACAAGGGCGTGATCGGCAAGATCCTGCCGGTTGAGGACATGCCGTTCCTTG
CCGACGGCACCCCGGTGGACATTATTTTGAACACCCACGGCGTGCCGCGACGGATGAACATCGGCCAGATTTTGGAGACC
CACCTGGGTTGGTGTGCCCACAGCGGCTGGAAGGTCGACGCCGCCAAGGGGGTTCCGGACTGGGCCGCCAGGCTGCCCG
ACGAACTGCTCGAGGCGCAGCCGAACGCCATTGTGTCGACGCCGGTGTTCGACGGCGCCCAGGAGGCCGAGCTGCAGGGC
CTGTTGTCGTGGCAGCTGCCCAACCGGACGGCGACAGGGTGACGTGCTCTCGACGGGCG
CAGCGGCGAGCCGTTCCCGTACCCGGTCACGGTTGGCTACATGTACATCATGAAGCTGCACCACCTGGTGGACGACAAGAT
CCACGCCCGCTCCACCGGGCCGTACTCGATGATCACCCAGCAGCCGCTGGGCGGTAAGGCGCAGTTCGGTGGCCAGCGGT
TCGGGGAGATGGAGTGCTGGGCCATGCAGGCCTACGGTGCTGCCTACACCCTGCAGGAGCTGTTGACCATCAAGTCCGAT
GACACCGTCGGCCGCGTCAAGGTGTACGAGGCGATCGTCAAGGGTGAGAACATCCCGGAGCCGGGCATCCCCGAGTCGTT
CAAGGTGCTGCTCAAAGAACTGCAGTCGCTGTGCCTCAACGTCGAGGTGCTATCGAGTGACGGTGCGGCGATCGAACTGC
GCGAAGGTGAGGACGAGGACCTGGAGCGGGCCGGGCCAACCTGGGAATCAATCTGTCCGCAACGAATCCGCAAGTGTC
GAGGATCTTGCGTAAAGCTGTCGCAAAATTACTAAACCGGTTAGGGGAAAGGGAGTTACGTGCTCGACGTCAACTTCTTC
GATGAACTCCGCATCGGTCTTGCTACCGCGGAGGACATCAGGCAATGGTCCTATGGCGAGG

Utilizziamo il programma Blast di NCBI per comparare 2 sequenze:

Accedete a Blast 2 sequences e incollate nelle due finestre la sequenza di DNA del batterio sensibile all'antibiotico e quella del batterio resistente all'antibiotico (Fig.1). Scegliete l'opzione "Optimize for Highly similar sequences (megablast)" e lanciate l'analisi (cliccando sul bottone blu in basso a sinistra "BLAST").

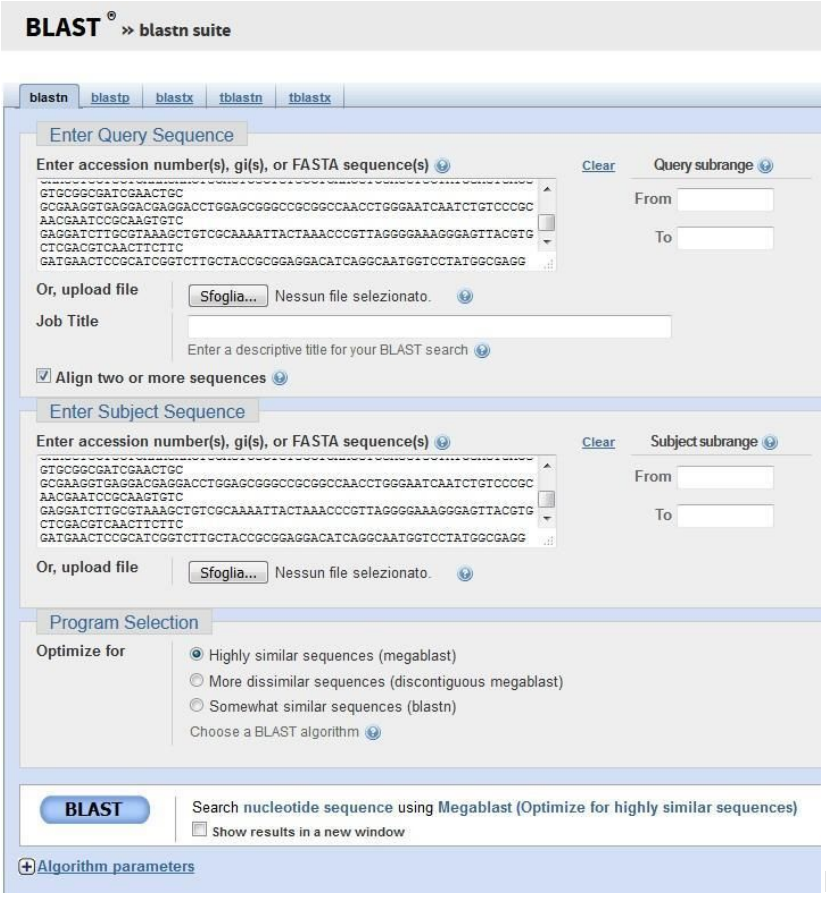

Fig.1

Ora analizziamo insieme il risultato. Ci sono differenze? Quali? Quali sono i possibili effetti di queste differenze?

Per capirlo osserviamo la pagina che ci si presenta: scorriamo verso il basso fino a mostrare la sezione *Descriptions* (Fig.2).

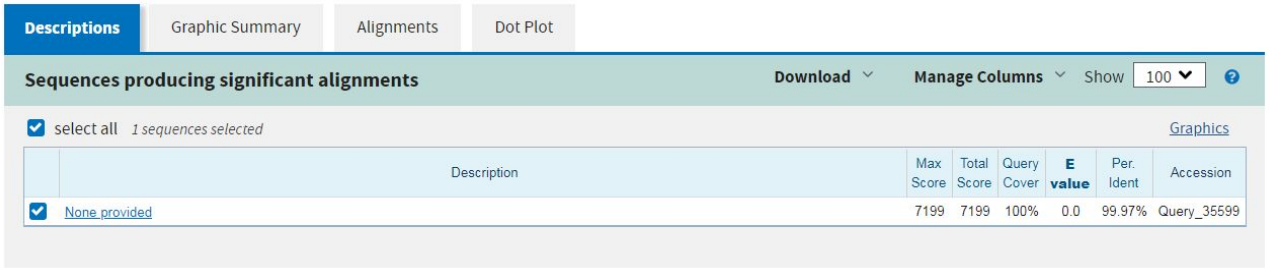

Fig.2

In *Descriptions* si trova il valore di Per. Ident. (Percent Identity), significa che le due sequenze hanno una corrispondenza di uguaglianza pari al 99,97%. Le differenze sono minime, ma dove si trovano?

Per capirlo vediamo nella sezione *Alignments,* dove sono comparate le due sequenze di DNA (Query e Subject): un trattino unisce le sequenze quando la base azotata è la stessa.

Non è facile trovare a occhio la differenza, pertanto clicchiamo su *Graphics* (Fig.3).

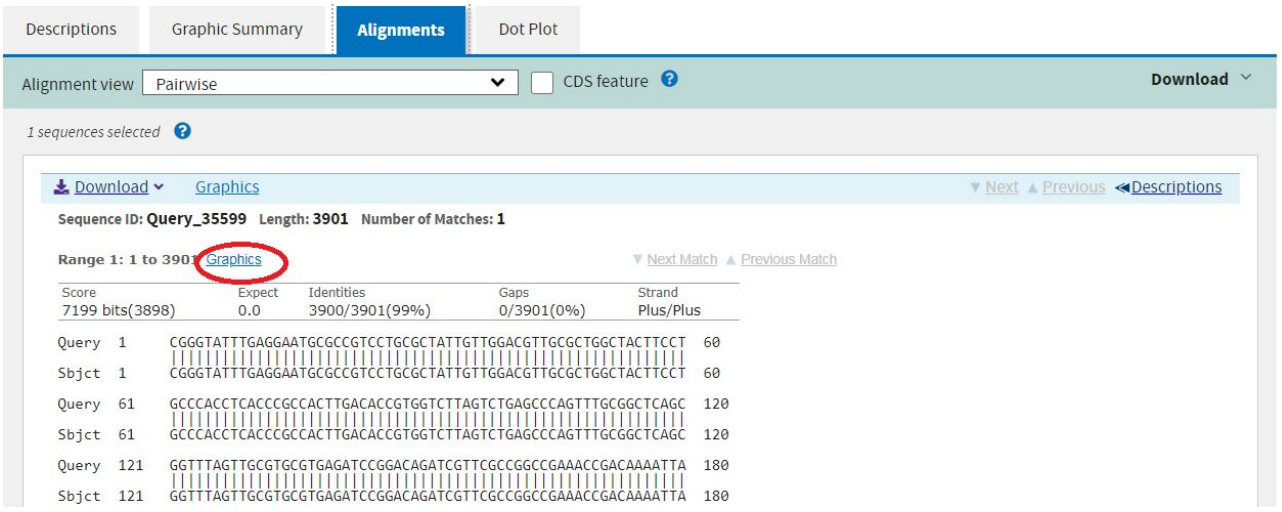

Fig.3

Qui possiamo trovare più facilmente le differenze, indicate da barrette rosse (Fig.4).

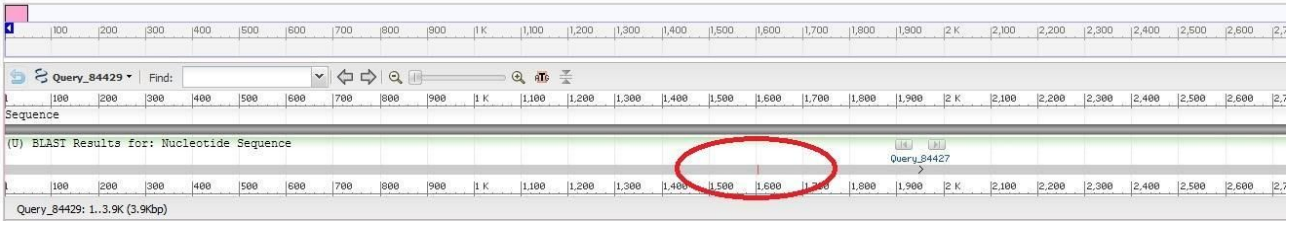

Fig.4

Come si può osservare è presente solo una piccolissima differenza: puntando il mouse sulla barra rossa per qualche secondo compare una finestra dove è segnato 'Mismatches:1' (Fig.5), significa che c'è una singola base azotata differente. Per capire quale base sia differente, bisogna ritornare su *Alignments* e confrontare le differenze in posizione 1604 (T oppure C nell'altra sequenza).

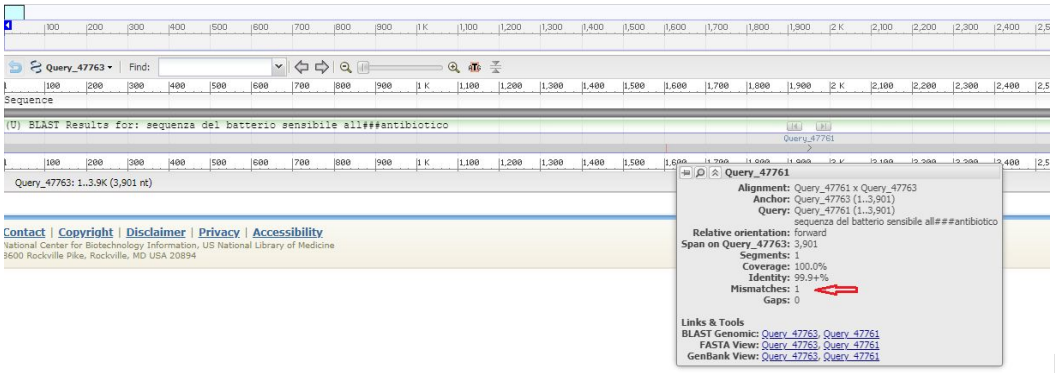

Fig.5

Esercizio 2.

Cerchiamo di associare alle sequenze di DNA possibili funzioni

Le sequenze di DNA potrebbero contenere geni che codificano per proteine e questi geni potrebbero essere preceduti da promotori. Per verificare l'eventuale presenza di geni nei tratti di DNA di interesse, proviamo a tradurre le due sequenze di DNA (sensibile e resistente) in proteine, partendo da tutte le posizioni possibili. Se troviamo una traduzione che ha almeno 50 amminoacidi uno di seguito all'altro senza stop (rappresentati dal programma da asterischi *) allora avremo trovato una possibile proteina.

Perchè la proteina deve avere almeno 50 amminoacidi? Esistono pochissime proteine più corte di 50 amminoacidi, la gran parte sono molto lunghe.

Incollate uno per volta il contenuto dei file di DNA tubercolosi sensibile e resistente utilizzati nell'esercizio 1 e sotto riportati, nella finestra di input del programma Sequence Manipulation Suit.

DNA tubercolosi sensibile

>sequenza del batterio sensibile all'antibiotico
CGGGTATTTGAGGAATGCGCCGTCCTGCGCTATTGTTGGACGTTGCGCTGGCTACTTCCTGCCCACCTCACCCGCCACTT
GACACCGTGGTCTTAGTCTGAGCCCAGTTTGCGGCTCAGCGGTTTAGTTGCGTGCGTGAGATCCGGACAGATCGTTCGCC
GGCCGAAACCGACAAAATTATCGCGGTTGACAGGCCCGTGGGCACCGCTCCTATAATGGCTCTCGTTGGTCGCATGAAGT
GCAGGAGGGATGCATCTTGGCAGATTCCCGCCAGAGCAAAACAGCCGCTAGTCCTAGTCCGAGTCGCCCGCAAAGTTCCT
CGAATAACTCCGTACCCGGAGCGCCAAACCGGGTCTCCTTCGCTAAGCTGCGCGAACCACTTGAGGTTCCGGGACTCCTT
GACGTCCAGACCGATTCGTTCGAGTGGCTGATCGGTTCGCCGCGCTGGCGCGAATCCGCCGCCGAGCGGGGTGATGTCAA
CCCAGTGGGTGGCCTGGAAGAGGTGGTCTACGAGCTGTCTCCGATCGAGGACTTCTCCGGGTCGATGTCGTTGTCGTTCT
CTGACCCTCGTTTCGACGATGTCAAGGCACCCGTCGACGAGTGCAAAGACAAGGACATGACGTACGCGGCTCCACTGTTC
GTCACCGCCGAGTTCATCAACAACAACACCGGTGAGATCAAGAGTCAGACGGTGTTCATGGGTGACTTCCCGATGATGAC
CGAGAAGGGCACGTTCATCATCAACGGGACCGAGCGTGTGGTGGTCAGCCAGCTGGTGCGGTCGCCCGGGGTGTACTTCG
ACGAGACCATTGACAAGTCCACCGACAAGACGCTGCACAGCGTCAAGGTGATCCCGAGCCGCGGCGCGTGGCTCGAGTTT
GACGTCGACAAGCGCGACACCGTCGGCGTGCGCATCGACCGCAAACGCCGGCAACCGGTCACCGTGCTGCTCAAGGCGCT
GGGCTGGACCAGCGAGCAGATTGTCGAGCGGTTCGGGTTCTCCGAGATCATGCGATCGACGCTGGAGAAGGACAACACCG
TCGGCACCGAACGAGGCGCTGTTGGACATCTACCGCAAGCTGCGTCCGGGCGAGCCCCGACCAAAGAGTCAGCGCAGACG
CTGTTGGAAAACTTGTTCTTCAAGGAGAAGCGCTACGACCTGGCCCGCGTCGGTCGCTATAAGGTCAACAAGAAGCTCGG
GCTGCATGTCGGCGAGCCCATCACGTCGTCGACGCTGACCGAAGAAGACGTCGTGGCCACCATCGAATATCTGGTCCGCT
TGCACGAGGGTCAGACCACGATGACCGTTCCGGGCGGCGTCGAGGTGCCGGTGGAAACCGACGACATCGACCACTTCGGC
AACCGCCGCCTGCGTACGGTCGGCGAGCTGATCCAAAACCAGATCCGGGTCGGCATGTCGCGGATGGAGCGGGTGGTCCG
GGAGCGGATGACCACCCAGGACGTGGAGGCGATCACACCGCAGACGTTGATCAACATCCGGCCGGTGGTCGCCGCGATCA
AGGAGTTCTTCGGCACCAGCCAGCTGAGCCAATTCATGGACCAGAACAACCCGCTGTCGGGGTTGACCCACAAGCGCCGA
CTGTCGGCGCTGGGGCCCGGCGGTCTGTCACGTGAGCGTGCCGGGCTGGAGGTCCGCGACGTGCACCCGTCGCACTACG
GCCGGATGTGCCCGATCGAAACCCCTGAGGGGCCCAACATCGGTCTGATCGGCTCGCTGTCGGTGTACGCGCGGGTCAAC
C
CGTTCGGGTTCATCGAAACGCCGTACCGCAAGGTGGTCGACGGCGTGGTTAGCGACGAGATCGTGTACCTGACCGCCGAC
GAGGAGGACCGCCACGTGGTGGCACAGGCCAATTCGCCGATCGATGCGGACGGTCGCTTCGTCGAGCCGCGCGTGCTGGT
CCGCCGCAAGGCGGGCGAGGTGGAGTACGTGCCCTCGTCTGAGGTGGACTACATGGACGTCTCGCCCCGCCAGATGGTGT
CGGTGGCCACCGCGATGATTCCCTTCCTGGAGCACGACGACGCCAACCGTGCCCTCATGGGGGCAAACATGCAGCGCCAG
GCGGTGCCGCTGGTCCTGAGCGAGGCCCGCTGGTGGGCACCCGGGATGGAGCTGCGCGGCGATCGACGCGCGGCGAGG
TCGTCGTCGGCCGAAGAAAGCGGCGTCATCGAGGAGGTGTCGGCCGACTACATCACTGTGATCACCGACGACCGCGACCCGGC
GTACCTACCGGATGCGCAAGTTTGCCCGGTCCAACCACGGCACTTGCGCCAACCAGTGCCCCATCGTGGACGCGGGCGACC
GAGTCGAGGCCGGTCAGGTGATCGCCGACGGTCCCTGTACTGACGACGGCGAGATGGCGCTGGGCAAGAACCTGCTGGTG
GCCATCATGCCGTGGGAGGGCCACAACTACGAGGACGCGATCATCCTGTCCAACCGCCTGGTCGAAGAGGACGTGCTCACC
TCGATCCACATCGAGGAGCATGAGATCGATGCTCGCGACACCAAGCTGGGTGCGGAGGAGATCACCCGCGACATCCCGAAC
ATCTCCGACGAGGTGCTCGCCGACCTGGATGAGCGGGGCATCGTGCGCATCGGTGCCGAGGTTCGCGACGGGGACATCCT
GGTCGGCAAGGTCACCCCGAAGGGTGACACCGAGCTGACGCGCGAGGAGCGGGTGCTGCGTGCCATCTTCGGTGAGAAGG
CCCGCGAGGTGCGCGACACTTCGCTGAAGGTGCCGCACGGCGAATCCGGCAAGGTGATCGGCATTCGGGTGTTTTCC
CGCGAGGACGAGGACGAGTTGCCGGCCGGTGTCAACGAGCTGGTGCGTGTGTATGTGGCTCAGAAACGCAAGATCTCCGA
CGGTGACAAGCTGGCCGGCCGGCACGGCAACAAGGGCGTGATCGGCAAGATCCTGCCGGTTGAGGACATGCCGTTCCTTG
CCGACGGCACCCCGGTGGACATTATTTTGAACACCCACGGCGTGCCGCGACGGATGAACATCGGCCAGATTTTGGAGACC
CACCTGGGTTGGTGTGCCCACAGCGGCTGGAAGGTCGACGCCGCCAAGGGGGTTCCGGACTGGGCCGCCAGGCTGCCCG
ACGAACTGCTCGAGGCGCAGCCGAACGCCATTGTGTCGACGCCGGTGTTCGACGGCGCCCAGGAGGCCGAGCTGCAGGGC
CTGTTGTCGTGCAGCGTCGCCCAACCGCGACGGTGACGTGCGTGGTCGACGCCGCGAAGGGCAAGGGCATGCTCTTCGACGGGCG
CAGCGGCGAGCCGTTCCCGTACCCGGTCACGGTTGCTACATGTACATCATGAAGCTGCACCACCTGGTGGACGACAAGAT
CCACGCCCGCTCCACCGGGCCGTACTCGATGATCACCCAGCAGCCGCTGGGCGGTAAGGCGCAGTTCGGTGGCCAGCGGT
TCGGGGAGATGGAGTGCTGGGCCATGCAGGCCTACGGTGCTGCCTACACCCTGCAGGAGCTGTTGACCATCAAGTCCGAT
GACACCGTCGGCCGCGTCAAGGTGTACGAGGCGATCGTCAAGGGTGAGAACATCCCGGAGCCGGGCATCCCCGAGTCGTT
CAAGGTGCTGCTCAAAGAACTGCAGTCGCTGTGCCTCAACGTCGAGGTGCTATCGAGTGACGGTGCGGCGATCGAACTGC
GCGAAGGTGAGGACGAGGACCTGGAGCGGGCCGCGGCCAACCTGGGAATCAATCTGTCCCGCAACGAATCCGCAAGTGTC
GAGGATCTTGCGTAAAGCTGTCGCAAAATTACTAAACCCGTTAGGGGAAAGGGAGTTACGTGCTCGACGTCAACTTCTTC
GATGAACTCCGCATCGGTCTTGCTACCGCGGAGGACATCAGGCAATGGTCCTATGGCGAGG

DNA tubercolosi resistente:

>sequenza del batterio resistente all'antibiotico
CGGGTATTTGAGGAATGCGCCGTCCTGCGCTATTGTTGGACGTTGCGCTGGCTACTTCCTGCCCACCTCACCCGCCACTT
GACACCGTGGTCTTAGTCTGAGCCCAGTTTGCGGCTCAGCGGTTTAGTTGCGTGCGTGAGATCCGGACAGATCGTTCGCC
GGCCGAAACCGACAAAATTATCGCGGTTGACAGGCCCGTGGGCACCGCTCCTATAATGGCTCTCGTTGGTCGCATGAAGT
GCAGGAGGGATGCATCTTGGCAGATTCCCGCCAGAGCAAAACAGCCGCTAGTCCTAGTCCGAGTCGCCCGCAAAGTTCCT
CGAATAACTCCGTACCCGGAGCGCCAAACCGGGTCTCCTTCGCTAAGCTGCGCGAACCACTTGAGGTTCCGGGACTCCTT
GACGTCCAGACCGATTCGTTCGAGTGGCTGATCGGTTCGCCGCGCTGGCGCGAATCCGCCGCCGAGCGGGGTGATGTCAA
CCCAGTGGGTGGCCTGGAAGAGGTGGTCTACGAGCTGTCTCCGATCGAGGACTTCTCCGGGTCGATGTCGTTGTCGTTCT
CTGACCCTCGTTTCGACGATGTCAAGGCACCCGTCGACGAGTGCAAAGACAAGGACATGACGTACGCGGCTCCACTGTTC
GTCACCGCCGAGTTCATCAACAACAACACCGGTGAGATCAAGAGTCAGACGGTGTTCATGGGTGACTTCCCGATGATGAC
CGAGAAGGGCACGTTCATCATCAACGGGACCGAGCGTGTGGTGGTCAGCCAGCTGGTGCGGTCGCCCGGGGTGTACTTCG
ACGAGACCATTGACAAGTCCACCGACAAGACGCTGCACAGCGTCAAGGTGATCCCGAGCCGCGGCGCGTGGCTCGAGTTT
GACGTCGACAAGCGCGACACCGTCGGCGTGCGCATCGACCGCAAACGCCGGCAACCGGTCACCGTGCTGCTCAAGGCGCT
GGGCTGGACCAGCGAGCAGATTGTCGAGCGGTTCGGGTTCTCCGAGATCATGCGATCGACGCTGGAGAAGGACAACACCG
TCGGCACCGAACGAGGCGCTGTTGGACATCTACCGCAAGCTGCGTCCGGGCGAGCCCCGACCAAAGAGTCAGCGCAGACG
CTGTTGGAAAACTTGTTCTTCAAGGAGAAGCGCTACGACCTGGCCCGCGTCGGTCGCTATAAGGTCAACAAGAAGCTCGG
GCTGCATGTCGGCGAGCCCATCACGTCGTCGACGCTGACCGAAGAAGACGTCGTGGCCACCATCGAATATCTGGTCCGCT
TGCACGAGGGTCAGACCACGATGACCGTTCCGGGCGGCGTCGAGGTGCCGGTGGAAACCGACGACATCGACCACTTCGGC
AACCGCCGCCTGCGTACGGTCGGCGAGCTGATCCAAAACCAGATCCGGGTCGGCATGTCGCGGATGGAGCGGGTGGTCCG
GGAGCGGATGACCACCCAGGACGTGGAGGCGATCACACCGCAGACGTTGATCAACATCCGGCCGGTGGTCGCCGCGATCA
AGGAGTTCTTCGGCACCAGCCAGCTGAGCCAATTCATGGACCAGAACAACCCGCTGTCGGGGTTGACCCACAAGCGCCGA
CTGCCGGCGCTGGGGCCCGGCGGTCTGTCACGTGAGCGTGCCGGGCTGGAGGTCCGCGACGTGCACCCGTCGCACTACG
GCCGGATGTGCCCGATCGAAACCCCTGAGGGGCCCAACATCGGTCTGATCGGCTCGCTGTCGGTGTACGCGCGGGTCAAC
C
CGTTCGGGTTCATCGAAACGCCGTACCGCAAGGTGGTCGACGGCGTGGTTAGCGACGAGATCGTGTACCTGACCGCCGAC
GAGGAGGACCGCCACGTGGTGGCACAGGCCAATTCGCCGATCGATGCGGACGGTCGCTTCGTCGAGCCGCGCGTGCTGGT
CCGCCGCAAGGCGGGCGAGGTGGAGTACGTGCCCTCGTCTGAGGTGGACTACATGGACGTCTCGCCCCGCCAGATGGTGT
CGGTGGCACCGCGATGATTCCCTTCCTGGAGCACGACGACGCCAACCGTGCCCTCATGGGGGCAAACATGCAGCGCCAG
GCGGTGCCGCTGGTCCTGAGCGAGGCCCGCTGGTGGGCACCCGGGATGGAGCTGCGCGGCGATCGACGCGCGGCGAGG
TCGTCGTCGGCCGAAGAAAGCGGCGTCATCGAGGAGGTGTCGGCCGACTACATCACTGTGATCACCGACGACCGCGACCCGGC
GTACCTACCGGATGCGCAAGTTTGCCCGGTCCAACCACGGCACTTGCGCCAACCAGTGCCCCATCGTGGACGCGGGCGACC
GAGTCGAGGCCGGTCAGGTGATCGCCGACGGTCCCTGTACTGACGACGGCGAGATGGCGCTGGGCAAGAACCTGCTGGTG
GCCATCATGCCGTGGGAGGGCCACAACTACGAGGACGCGATCATCCTGTCCAACCGCCTGGTCGAAGAGGACGTGCTCACC
TCGATCCACATCGAGGAGCATGAGATCGATGCTCGCGACACCAAGCTGGGTGCGGAGGAGATCACCCGCGACATCCCGAAC
ATCTCCGACGAGGTGCTCGCCGACCTGGATGAGCGGGGCATCGTGCGCATCGGTGCCGAGGTTCGCGACGGGGACATCCT
GGTGGGCAAGGTCACCCCGAAGGGTGACACCGAGCTGACGCGCGAGGAGCGGGTGCTGCGTGCCATCTTCGGTGAGAAGG
CCCGCGAGGTGCGCGACACTTCGCTGAAGGTGCCGCACGGCGAATCCGGCAAGGTGATCGGCATTCGGGTGTTTTCC
CGCGAGGACGAGGACGAGTTGCCGGCCGGTGTCAACGAGCTGGTGCGTGTGTATGTGGCTCAGAAACGCAAGATCTCCGA
CGGTGACAAGCTGGCCGGCCGGCACGGCAACAAGGGCGTGATCGGCAAGATCCTGCCGGTTGAGGACATGCCGTTCCTTG
CCGACGGCACCCCGGTGGACATTATTTTGAACACCCACGGCGTGCCGCGACGGATGAACATCGGCCAGATTTTGGAGACC
CACCTGGGTTGGTGTGCCCACAGCGGCTGGAAGGTCGACGCCGCCAAGGGGGTTCCGGACTGGGCCGCCAGGCTGCCCG
ACGAACTGCTCGAGGCGCAGCCGAACGCCATTGTGTCGACGCCGGTGTTCGACGGCGCCCAGGAGGCCGAGCTGCAGGGC
CTGTTGTCGTGCAGCGTCGCCCAACCGCGACGGTGACGTGCGTGGTCGACGCCGCGAAGGGCAAGGGCATGCTCTTCGACGGGCG
CAGCGGCGAGCCGTTCCCGTACCCGGTCACGGTTGCTACATGTACATCATGAAGCTGCACCACCTGGTGGACGACAAGAT
CCACGCCCGCTCCACCGGGCCGTACTCGATGATCACCCAGCAGCCGCTGGGCGGTAAGGCGCAGTTCGGTGGCCAGCGGT
TCGGGGAGATGGAGTGCTGGGCCATGCAGGCCTACGGTGCTGCCTACACCCTGCAGGAGCTGTTGACCATCAAGTCCGAT
GACACCGTCGGCCGCGTCAAGGTGTACGAGGCGATCGTCAAGGGTGAGAACATCCCGGAGCCGGGCATCCCCGAGTCGTT
CAAGGTGCTGCTCAAAGAACTGCAGTCGCTGTGCCTCAACGTCGAGGTGCTATCGAGTGACGGTGCGGCGATCGAACTGC
GCGAAGGTGAGGACGAGGACCTGGAGCGGGCCGCGGCCAACCTGGGAATCAATCTGTCCCGCAACGAATCCGCAAGTGTC
GAGGATCTTGCGTAAAGCTGTCGCAAAATTACTAAACCCGTTAGGGGAAAGGGAGTTACGTGCTCGACGTCAACTTCTTC
GATGAACTCCGCATCGGTCTTGCTACCGCGGAGGACATCAGGCAATGGTCCTATGGCGAGG

Reading Frame 1, 2 o 3 sono i 3 quadri di possibili letture in proteine della sequenza di mRNA ottenuta dalla questa sequenza di DNA copiata. *Direct e reverse strand* sono la sequenza di DNA inserito nel form (direct) o quello complementare (reverse) che possono essere ugualmente trascrivibili in 2 differenti mRNA.

Fate delle prove inserendo la reading frame 1, 2 o 3 e il direct o reverse strand. Cliccate sul bottone 'Submit' per aprire una finestra col risultato della sequenza di amminoacidi corrispondente alla lettura inserita.

Gli amminoacidi sono rappresentati da singole lettere secondo la seguente tabella:

Tabella degli aminoacidi

Sigla a tre lettere	Sigla a una lettera	Nome
Gly	G	Glicina
Ala	A	Alanina
Val	V	Valina
Leu	L	Leucina
Ile	I	Isoleucina
Met	M	Metionina
Cys	C	Cisteina
Pro	P	Prolina
Phe	F	Fenialalanina
Trp	W	Triptofano
Tyr	Y	Tirosina
Thr	T	Treonina
Ser	S	Serina
Asn	N	Asparagina
Gln	Q	Glutammina
Asp	D	Acido aspartico o Aspartato
Glu	E	Acido Glutamminico o Glutammato
His	H	Istidina
Lys	K	Lisina
Arg	R	Arginina

🔎 Focus sul quadro di lettura in proteine

La conversione da RNA ad amminoacidi (traduzione) avviene a triplette, cioè la sequenza di RNA viene letta lungo il filamento 3 basi alla volta, e ogni tripletta corrisponderà ad un amminoacido nella proteina finale (tabella sottostante) o può essere uno stop nella traduzione. A seconda della base da cui si parte sono possibili quindi tre quadri di lettura (Reading Frame) differenti che potrebbero portare a sequenze proteiche sufficientemente lunghe oppure che terminano dopo pochi amminoacidi.

			SECONDA		LETTERA				
		U	C		A		G		
PRIMA	U	UUU phe UUC phe UUA leu UUG leu	UCU ser UCC ser UCA ser UCG ser		UAU tyr UAC tyr UAA stop UAG stop		UGU cys UGC cys UGA stop UGG trp	U C A G	TERZA
LETT	C	CUU leu CUC leu CUA leu CUG leu	CCU pro CCC pro CCA pro CCG pro		CAU his CAC his CAA gln CAG gln		CGU arg CGC arg CGA arg CGG arg	U C A G	LETT
ERA	A	AUU ile AUC ile AUA ile AUG met	ACU thr ACC thr ACA thr ACG thr		AAU asn AAC asn AAA lys AAG lys		AGU ser AGC ser AGA arg AGG arg	U C A G	ERA
	G	GUU val GUC val GUA val GUG val	GCU ala GCC ala GCA ala GCG ala		GAU asp GAC asp GAA glu GAG glu		GGU gly GGC gly GGA gly GGG gly	U C A G	

ALA ⇒ alanina
ARG ⇒ arginina
ASP ⇒ acido aspartico
ASN ⇒ asparagina
CYS ⇒ cisteina
GLU ⇒ acido glutammico
GLN ⇒ glutammina
GLY ⇒ glicina
HIS ⇒ istidina
ILEU ⇒ isoleucina
LEU ⇒ leucina
LYS ⇒ lisina
MET ⇒ metionina
PHE ⇒ fenilalanina
PRO ⇒ prolina
SER ⇒ serina
THR ⇒ treonina
TRY ⇒ triptofano
TYR ⇒ tirosina
VAL ⇒ valina

Guardiamo insieme il risultato, ci sono possibili geni? Quale delle possibili traduzioni sceglierestе?

✔ Risposta corretta

La risposta corretta per entrambe le sequenze è: il ***reading frame 2*** *on the* ***direct*** *strand*. Perché permette di ottenere la proteina più lunga, cioè la sequenza di lettere più lunga tra due asterischi (Fig.6).

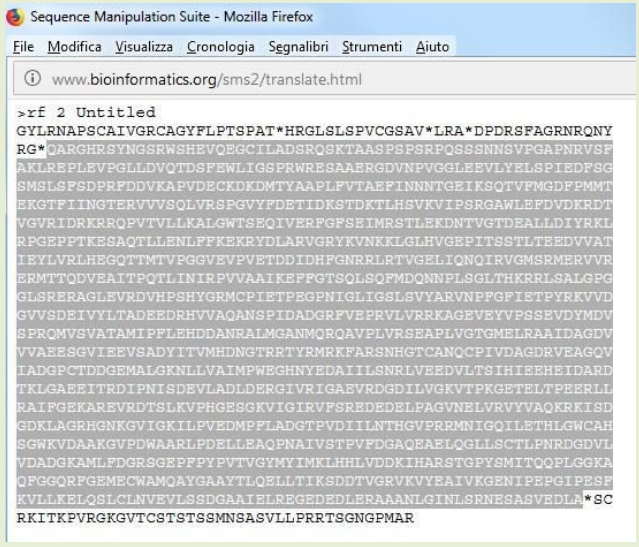

Fig.6 evidenzia la sequenza di aa tra 2 Stop (*)

Si tratta di funzioni note?

Avendo trovato una sequenza codificante per una proteina in entrambi i casi durante il passaggio precedente, vale la pena controllare se quelle proteine sono già note o assomigliano a proteine note a cui è stata già associata una funzione. Per farlo useremo di nuovo BLAST, il programma per comparare le sequenze fra di loro, ma questa volta con una diversa modalità.

Cliccate su BlastP per accedere alla versione del programma BLAST che serve per comparare le sequenze di proteine fra loro e, in particolare, che consente di comparare la sequenza di una proteina di interesse con le sequenze di tutte le proteine presenti nella banca dati delle sequenze proteiche.

Incollate nella finestra le sequenze proteiche che dovreste aver ottenuto traducendo il DNA tubercolosi sensibile e resistente in amminoacidi, scegliendo il risultato ottenuto per il **reading frame 2** on the **direct** strand.

Per comodità riportiamo di seguito le due sequenze:

Sequenza di aa da DNA tubercolosi sensibile	Sequenza di aa da DNA tubercolosi resistente
GYLRNAPSCAIVGRCAGYFLPTSPAT*HRGLSLSPVCGSAV*LRA*DPDRSFAGRNRQ NYRG*QARGHRSYNGSRWSHEVQEGCILADSRQSKTAASPSPSRPQSSSNNSVPG APNRVSFAKLREPLEVPGLLDVQTDSFEWLIGSPRWRESAAERGDVNPVGGLEEVL YELSPIEDFSGSMSLSFSDPRFDDVKAPVDECKDKDMTYAAPLFVTAEFINNNTGEIK SQTVFMGDFPMMTEKGTFIINGTERVVVSQLVRSPGVYFDETIDKSTDKTLHSVKVIP SRGAWLEFDVDKRDTVGVRIDRKRRQPVTVLLKALGWTSEQIVERFGFSEIMRSTLE KDNTVGTDEALLDIYRKLRPGEPPTKESAQTLLENLFFKEKRYDLARVGRYKVNKKLG LHVGEPITSSTLTEEDVVATIEYLVRLHEGQTTMTVPGGVEVPVETDDIDHFGNRRLR TVGELIQNQIRVGMSRMERVVRERMTTQDVEAITPQTLINIRPVVAAIKEFFGTSQLSQ FMDQNNPLSGLTHKRRLSALGPGGLSRERAGLEVRDVHPSHYGRMCPIETPEGPNI GLIGSLSVYARVNPFGFIETPYRKVVDGVVSDEIVYLTADEEDRHVVAQANSPIDADG RFVEPRVLVRRKAGEVEYVPSSEVDYMDVSPRQMVSVATAMIPFLEHDDANRALMG ANMQRQAVPLVRSEAPLVGTGMELRAAIDAGDVVVAEESGVIEEVSADYITVMHDNG TRRTYRMRKFARSNHGTCANQCPIVDAGDRVEAGQVIADGPCTDDGEMALGKNLLV AIMPWEGHNYEDAIILSNRLVEEDVLTSIHIEEHEIDARDTKLGAEEITRDIPNISDEVLA DLDERGIVRIGAEVRDGDILVGKVTPKGETELTPEERLLRAIFGEKAREVRDTSLKVPH GESGKVIGIRVFSREDEDELPAGVNELVRVYVAQKRKISDGDKLAGRHGNKGVIGKIL PVEDMPFLADGTPVDIILNTHGVPRRMNIGQILETHLGWCAHSGWKVDAAKGVPDW AARLPDELLEAQPNAIVSTPVFDGAQEAELQGLLSCTLPNRDGDVLVDADGKAMLFD GRSGEPFPYPVTVGYMYIMKLHHLVDDKIHARSTGPYSMITQQPLGGKAQFGGQRF GEMECWAMQAYGAAYTLQELLTIKSDDTVGRVKVYEAIVKGENIPEPGIPESFKVLLK ELQSLCLNVEVLSSDGAAIELREGEDEDLERAAANLGINLSRNESASVEDLA*SC RKITKPVRGKGVTCSTSTSSMNSASVLLPRRTSGNGPMAR	GYLRNAPSCAIVGRCAGYFLPTSPAT*HRGLSLSPVCGSAV*LRA*DPDRSFAGRNRQ NYRG*QARGHRSYNGSRWSHEVQEGCILADSRQSKTAASPSPSRPQSSSNNSVPG APNRVSFAKLREPLEVPGLLDVQTDSFEWLIGSPRWRESAAERGDVNPVGGLEEVL YELSPIEDFSGSMSLSFSDPRFDDVKAPVDECKDKDMTYAAPLFVTAEFINNNTGEIK SQTVFMGDFPMMTEKGTFIINGTERVVVSQLVRSPGVYFDETIDKSTDKTLHSVKVIP SRGAWLEFDVDKRDTVGVRIDRKRRQPVTVLLKALGWTSEQIVERFGFSEIMRSTLE KDNTVGTDEALLDIYRKLRPGEPPTKESAQTLLENLFFKEKRYDLARVGRYKVNKKLG LHVGEPITSSTLTEEDVVATIEYLVRLHEGQTTMTVPGGVEVPVETDDIDHFGNRRLR TVGELIQNQIRVGMSRMERVVRERMTTQDVEAITPQTLINIRPVVAAIKEFFGTSQLSQ FMDQNNPLSGLTHKRRLPALGPGGLSRERAGLEVRDVHPSHYGRMCPIETPEGPNI GLIGSLSVYARVNPFGFIETPYRKVVDGVVSDEIVYLTADEEDRHVVAQANSPIDADG RFVEPRVLVRRKAGEVEYVPSSEVDYMDVSPRQMVSVATAMIPFLEHDDANRALMG ANMQRQAVPLVRSEAPLVGTGMELRAAIDAGDVVVAEESGVIEEVSADYITVMHDNG TRRTYRMRKFARSNHGTCANQCPIVDAGDRVEAGQVIADGPCTDDGEMALGKNLLV AIMPWEGHNYEDAIILSNRLVEEDVLTSIHIEEHEIDARDTKLGAEEITRDIPNISDEVLA DLDERGIVRIGAEVRDGDILVGKVTPKGETELTPEERLLRAIFGEKAREVRDTSLKVPH GESGKVIGIRVFSREDEDELPAGVNELVRVYVAQKRKISDGDKLAGRHGNKGVIGKIL PVEDMPFLADGTPVDIILNTHGVPRRMNIGQILETHLGWCAHSGWKVDAAKGVPDW AARLPDELLEAQPNAIVSTPVFDGAQEAELQGLLSCTLPNRDGDVLVDADGKAMLFD GRSGEPFPYPVTVGYMYIMKLHHLVDDKIHARSTGPYSMITQQPLGGKAQFGGQRF GEMECWAMQAYGAAYTLQELLTIKSDDTVGRVKVYEAIVKGENIPEPGIPESFKVLLK ELQSLCLNVEVLSSDGAAIELREGEDEDLERAAANLGINLSRNESASVEDLA*SC RKITKPVRGKGVTCSTSTSSMNSASVLLPRRTSGNGPMAR

Lasciate tutte le impostazioni così come le trovate e cliccate sul bottone "BLAST" in fondo a sinistra (Fig.7). La ricerca potrà richiedere alcuni minuti.

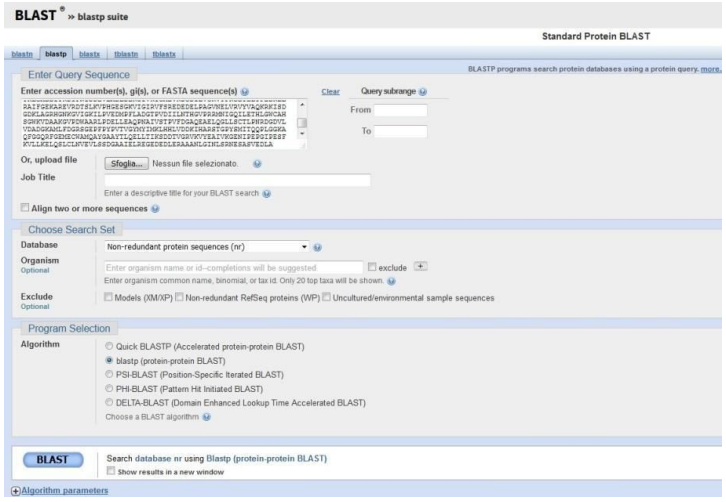

Fig.7

Guardiamo di nuovo i risultati. La ricerca identifica qualche proteina nota? Qual'è la funzione di questa proteina? Ritenete che sia una funzione importante per la sopravvivenza delle cellule batteriche?
Provate a fare lo stesso esercizio anche con la sequenza resistente alla tubercolosi.

✔ Risposta

Clicchiamo sul primo risultato della sequenza sensibile nella sezione *Descriptions* (Per.Ident. 100%, Fig.8). Questo risultato corrisponde ad una RNA polimerasi subunità beta appartenente proprio al *Mycobacterium tuberculosis*.

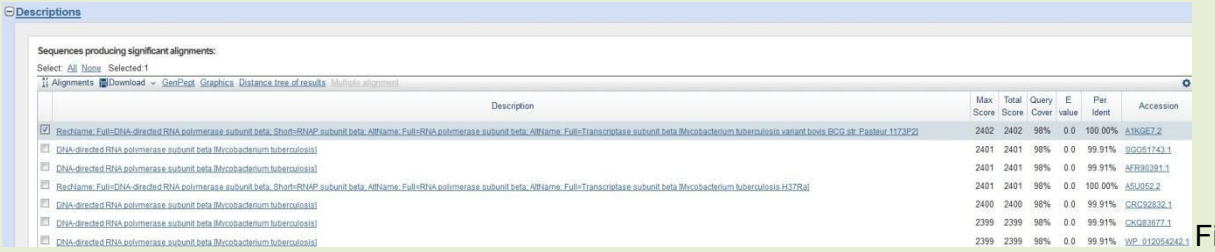

Fig.8

Facendo la stessa procedura con la sequenza resistente otteniamo come risultato una RNA polimerasi subunità beta appartenente proprio al *Mycobacterium tuberculosis con una Per. Ident 99.91 %*. Quest'ultimo risultato inferiore al 100% è dovuto alla base differente.

Fig.8

Esercizio 3.

Le differenze fra le sequenze di DNA si traducono in qualche effetto sulle sequenze delle proteine codificate?

Per rispondere a questa domanda ricorriamo di nuovo all'aiuto di BLAST. Lo useremo per confrontare due sequenze fra loro, ma questa volta saranno sequenze di aminoacidi invece che sequenze di nucleotidi.
Cliccate per accedere al programma "Blast 2 sequences" che questa volta è impostato per comparare sequenze di proteine e incollate nelle due finestre rispettivamente la proteina che avete

tradotto a partire dal DNA tubercolosi sensibile e quella che avete tradotto a partire dal DNA tubercolosi resistente (si veda esercizio 2). Cliccate nuovamente su "Blast".

Ci sono differenze fra le due proteine?
Anche in questo caso seguiamo la procedura effettuata per il Blast delle sequenze di DNA: osserviamo dalla sezione *Descriptions* una Per. Ident. tra le 2 proteine del 99,92%, quindi nella sezione *Alignments* clicchiamo su *Graphics* (Fig.9).

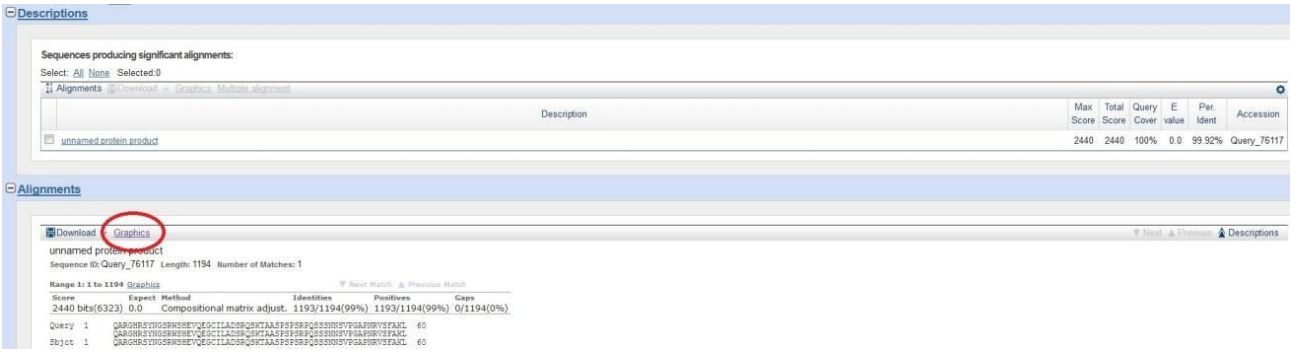

Fig.9

La differenza tra le due proteine è evidenziata dal tratto rosso (Fig.10).

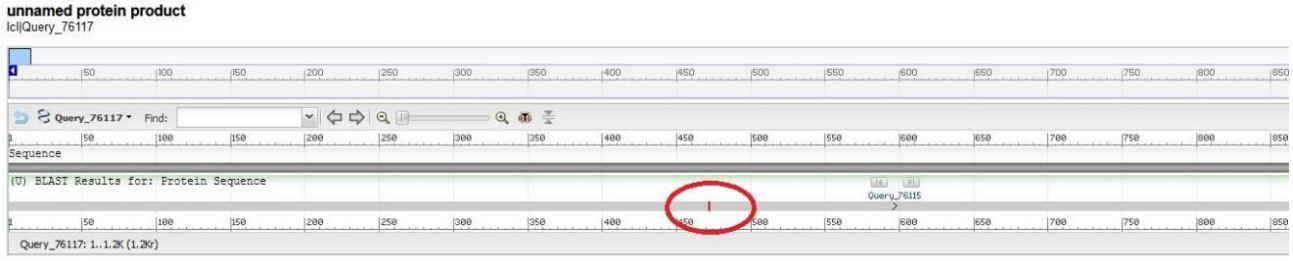

Fig.10

Puntando il mouse sulla barra rossa per qualche secondo compare una finestra dove è segnato 'Mismatches:1', significa che c'è un singolo amminoacido differente. Potete ipotizzare qualche interpretazione di questo dato? Perchè un batterio viene ucciso dall'antibiotico e l'altro no?

Con i dati che avete ottenuto provate a fare una ipotesi su cosa sia successo e poi leggete di seguito le conclusioni.

CONCLUSIONI

La resistenza agli antibiotici, un problema reale

E' proprio utilizzando un metodo molto simile a quello che abbiamo descritto in questa esercitazione che si sono scoperte le basi molecolari di alcuni casi comuni di resistenza batterica agli antibiotici. Nel sito linkato trovate una tabella che ne elenca alcuni, divisi per tipologia.
La resistenza di *Mycobacterium tuberculosis* agli antibiotici è causa di una nuova ondata di vittime della tubercolosi.
Il caso che abbiamo esaminato è un caso reale di resistenza di *Mycobacterium tuberculosis* all'antibiotico Rifampicina dovuto ad una mutazione puntiforme (cioè di una singola base del DNA)

nel gene che codifica per una proteina fondamentale per la sopravvivenza del batterio: la subunità beta della RNA Polimerasi.
L'antibiotico Rifampicina si lega a questa proteina e ne blocca l'attività, così i batteri, senza una delle funzioni fondamentali per la sopravvivenza, muoiono. Questa singola mutazione (ma ce ne sono altre note) impedisce alla Rifampicina di legarsi alla subunità beta della RNA Polimerasi ma non impedisce alla RNA Polimerasi di svolgere la sua funzione, quindi i batteri continuano a proliferare anche in presenza dell'antibiotico.
Per fare approfondimenti sulla tubercolosi e per capire perchè nella storia presentata nell'introduzione è stata scelta proprio New York, potete consultare la pagina di wikipedia dedicata a questa malattia.
Un video interessante sulla resistenza agli antibiotici in inglese sottotitolato è quello proposto da TED Ed.

Le sequenze analizzate

La sequenza del gene per la "DNA-directed RNA polymerase subunit beta" di *Mycobacterium tuberculosis* H37Rv (un ceppo virulento sensibile alla Rifampicina) che avete utilizzato in questa esercitazione è la vera sequenza ottenuta da studi sperimentali e depositata nel database pubblico dell'NCBI. La differenza che avete trovato è una mutazione puntiforme realmente osservata in batteri che non rispondono all'antibiotico. Si tratta di una delle mutazioni sequenziate frequentemente nei soggetti resistenti ed è ampiamente studiata, nota come mutazione in posizione 531. Il residuo in posizione 531 è normalmente una serina, ma nei mutanti viene sostituita da un altro aminoacido. Il residuo 531 corrisponde al 450-esimo residuo amminoacidico della sequenza che vi è stata fornita e la vostra mutazione scambia una serina con una prolina.