

Report progetto d'esame

PyLo3

PyLo3 è un software pensato e realizzato da Lorenzo D'Eusebio e Daniele Sponta.

È un software che propone l'analisi filogenetica di una o più sequenze genetiche.

Nasce dall'idea di analizzare le affinità fra il genoma SARS-CoV-2 del pipistrello con quello dei ceppi umani sviluppatosi poco dopo nel corso della pandemia.

"Il SARS-CoV-2 che ha colpito l'intera popolazione mondiale a partire da dicembre 2019 è stato isolato e sequenziato in Cina.

Nasce apparentemente già maturo e dotato di un'infettività sufficiente a diffondersi rapidamente, a circolare per lungo tempo.

Il virus è di origine zoonotica e in Cina è avvenuto il salto dall'animale, il pipistrello, all'uomo.

Il corpo umano è l'ambiente all'interno del quale il virus deve adattarsi e all'interno del quale deve sopravvivere; è quindi verosimile che nel diffondersi, il virus possa esser mutato andando, facendo sì che venissero selezionate nuove varietà virali." fonte: <https://www.scienzainrete.it/articolo/evoluzione-di-sars-cov-2-dallasia-alleuropa-mutazione-d614g/gaetano-di-chiara/2020-07-03>

A tal proposito nasce PyLo3.

Avviato il software l'utente si trova in presenza di una interfaccia grafica che lo pone di fronte ad una triplice scelta:

1. selezionare manualmente sequenze fasta dal proprio computer;
2. selezionare sequenze dal database di NCBI attraverso l'e-mail (la stesso con il quale si è iscritti in <https://www.ncbi.nlm.nih.gov/>) e l'accession number delle singole sequenze che si vogliono analizzare;
3. selezionare in maniera ibrida file dal proprio pc e dal database di NCBI.

Per l'operazione di allineamento multiplo si è scelto di usare MUSCLE, MULTiple Sequence Comparison by Log-Expectation, un software per l'allineamento multiplo di proteine e nucleotidi, pubblicato da Robert C. Edgar in due articoli nel 2004.

MUSCLE è spesso usato come sostituto di Clustal, ulteriore software per l'allineamento multiplo, poiché di solito (ma non sempre) fornisce allineamenti di sequenze migliori, a seconda delle opzioni scelte.

Inoltre, MUSCLE è significativamente più veloce di Clustal, soprattutto per allineamenti più ampi.

MUSCLE implementa un algoritmo di allineamento multiplo iterativo, calcolando una soluzione subottimale e successivamente modificandola fino a quando la soluzione converge verso quella ottimale.

Terminato l'allineamento si è scelto di convertire il file di output in un file formato Clustal nel quale le sequenze vengono incolonnate una sotto l'altra permettendo all'utente di visualizzare in maniera più efficace eventuali match, mismatch o gap.

Avviata l'operazione di analisi filogenetica PyLo3 genererà in output l'albero filogenetico costruito con il metodo UPGMA, Unweighted Pair Group Method with Arithmetic mean, generando all'utente di scaricare nel proprio pc i singoli file di output compreso il png dell'albero filogenetico attraverso un'interfaccia grafica generata in HTML tramite il proprio browser.

Al momento della stesura sono stati eseguiti gli esperimenti descritti nel file "Esperimenti.pdf".