**Εργαστηριακή άσκηση 4**

**Πολλαπλή στοίχιση ακολουθιών**

**Μάθημα: Βιοπληροφορική 1**

**Λάζαρος Κωνσταντίνος-Παναγιώτης**

**Α.Μ: 01639**

**Α)** Παρατηρούμε ότι με τις παραμέτρους που δώσαμε στο blastp, έχουμε πάρει 316 αποτελέσματα από τις 410,779 ακολουθίες που υπάρχουν συνολικά στην βάση δεδομένων. Επίσης παρατηρούμε ότι έχουμε ένα max score που είναι ίσο με 223 και e-value που είναι ίσο με 3e-75. Με δεδομένο λοιπόν ότι έχουμε ένα e-value με πολύ χαμηλή τιμή και ένα μεγάλο σκορ, θα μπορούσαμε να πούμε ότι έγινε μια καλή στοίχηση, κάτι το οποίο είναι αναμενόμενο καθώς οι αιμοσφαιρίνες είναι μια οικογένεια πρωτεινών καλά διατηρημένη μεταξύ των διαφόρων ειδών οργανισμών (η σύγκριση γίνεται μεταξύ ανθρώπινης αιμοσφαιρίνης και αιμοσφαιρινών από οργανισμούς που ΔΕΝ είναι θηλαστικά).

**Β)** Ακολουθεί πολλαπλή στοίχιση με χρήση του προγράμματος clustal omega. Η στοίχιση έγινε μεταξύ ακολουθιών που ανήκουν σε οργανισμούς οι οποίοι ανήκουν σε διαφορετικές ταξινομικές βαθμίδες με βάση το blast name που έχουν. Παρατηρούμε ότι υπάρχουν συνολικά 11 κενά. Η στοίχιση είναι καλή καθώς υπάρχουν διατηρημένες περιοχές (27), μερικώς διατηρημένες περιοχές (15) και περιοχές που αποτελούνται από παρόμοια κατάλοιπα (26). Η ύπαρξη κάποιον κενών καθώς και η ύπαρξη περιοχών χωρίς match θα μπορούσαμε να πούμε ότι είναι αναμενόμενη, καθώς η στοίχιση γίνεται μεταξύ ανθρώπινης αιμοσφαιρίνης β με αιμοσφαιρίνες β άλλων οργανισμων οι οποίοι δεν είναι θηλαστικά (πιθανώς αν γινόταν στοίχιση με αιμοσφαιρίνες β από άλλα θηλαστικά να προέκυπτε μια ακόμα καλύτερη στοίχιση). Παρατηρούμε ότι η πολλαπλή στοίχιση που προέκυψε με τη χρήση του προγράμματος tcoffee έχει ένα παραπάνω hit απ' ότι η στοίχιση που προέκυψε με χρήση του προγράμματος clustal, κάτι το οποίο όμως μπορεί να θεωρηθεί και ως στατιστικά μη σημαντικό.

Στην συνέχεια, γίνεται πολλαπλή στοίχιση με χρήση του προγράμματος tcoffee. Η στοίχιση έγινε μεταξύ ακολουθιών που ανήκουν σε οργανισμούς οι οποίοι ανήκουν σε διαφορετικές ταξινομικές βαθμίδες με βάση το blast name που έχουν. Παρατηρούμε ότι υπάρχουν συνολικά 11 κενά. Η στοίχιση είναι καλή καθώς υπάρχουν διατηρημένες περιοχές (27), μερικώς διατηρημένες περιοχές (16) και περιοχές που αποτελούνται από παρόμοια κατάλοιπα (26). Η ύπαρξη κάποιον κενών καθώς και η ύπαρξη περιοχών χωρίς match θα μπορούσαμε να πούμε ότι είναι αναμενόμενη, καθώς η στοίχιση γίνεται μεταξύ ανθρώπινης αιμοσφαιρίνης β με αιμοσφαιρίνες β άλλων οργανισμων οι οποίοι δεν είναι θηλαστικά (πιθανώς αν γινόταν στοίχιση με αιμοσφαιρίνες β από άλλα θηλαστικά να προέκυπτε μια ακόμα καλύτερη στοίχιση).

**Γ)** Ακολουθεί search μέσω του server του hmmer. Στο query χρησιμοποιήσαμε την ανθρώπινη πρωτείνη hemoglobin subunit beta (με accession number P68871). Η αναζήτηση έγινε έναντι της βάσης Swissprot αποκλείοντας όπως και πριν την ταξινομική βαθμίδα Mammalia (taxid:40674). Παρατηρούμε ότι από την αναζήτηση που έγινε προέκυψαν 391 significant query matches, γεγονός αναμενόμενο καθώς η οικογένεια των αιμοσφαιρινών είναι καλά διατηρημένη. Επίσης αξιοσημείωτο είναι το γεγονός ότι μέσω του phhmer πήραμε 75 παραπάνω αποτελέσματα απ'ότι με το blastp. Αυτό μπορεί να οφείλεται στο γεγονός ότι το hmmer είναι ένα εργαλείο το οποίο χρησιμοποιείται κυρίως για εύρεση δομών συγκεκριμένων τμημάτων (domain structure) μέσα σε μια πρωτεινική ακολουθία. Γενικά το hmmer χρησιμοποιείται περισσότερο όταν η μελέτη που κάνουμε έχει εξελικτική σημασία.

**Δ)** Τέλος, έχουμε αναζήτηση με το hmmsearch του hmmer χρησιμοποιώντας ως query την πολλαπλή στοίχιση που προέκυψε με χρήση του εργαλείου tcoffee. Η αναζήτηση έγινε έναντι της βάσης Uniprot/SwissProt και δεν λήφθηκε υπόψιν η τάξινομική βαθμίδα mammalia (με taxid 40674). Παρατηρούμε ότι από αυτή την αναζήτηση παίρνουμε συνολικά 415 αποτελέσματα. Αυτό συμβαίνει καθώς χρησιμοποιώντας ως query μια πολλαπλή στοίχιση έναντι μιας βάσης δεδομένων, ουσιαστικά ψάχνουμε να βρούμε μέλη της πρωτεινικής οικογένειας η οποία περιγράφεται απ'την στοίχιση που δώσαμε ως query ενώ στις άλλες αναζητήσεις που πραγματοποιήθηκαν χρησιμοποιώντας μόνο μια ακολουθία, ψάχνουμε να βρούμε πρωτεινικές ακολουθίες οι οποίες απλά μοιάζουν με την ακολουθία που είχαμε αρχικά. Επομένως στην συγκεκριμένη περίπτωση είναι αναμενόμενο το να πάρουμε περισσότερα αποτελέσματα ως output από την αναζήτηση.