**Εργαστηριακή άσκηση 5**

**Φυλογενετική ανάλυση**

**Μάθημα: Βιοπληροφορική 1**

**Λάζαρος Κωνσταντίνος-Παναγιώτης**

**Α.Μ: 01639**

**Α) Σύγκριση φυλογενετικών δέντρων**

**Γίνεται σύγκριση μεταξύ τον φυλογενετικών δέντρων που προκύπτουν χρησιμοποιώντας τις μεθόδους neighbor joining, maximum likelihood και upgma. Παρατηρούμε ότι υπάρχουν διαφορές τόσο στις αποστάσεις όσο και στην αρχιτεκτονική των δέντρων, κάτι το οποίο είναι αναμενόμενο εφόσον χρησιμοποιούνται διαφορετικοί αλγόριθμοι φυλογενετικής ανάλυσης σε κάθε περίπτωση (το upgma και το neighbor joining είναι αλγόριθμοι που κάνουν clustering ενώ το maximum likelihood είναι αλγόριθμος που κάνει στατιστική ανάλυση προκυμένου να δώσει φυλογενετικό δέντρο). Επίσης μια ακόμα αξιοσημείωτη διαφορά μεταξύ των αλγορίθμων είναι ότι με τον upgma πήραμε ένα δέντρο το οποίο έχει ρίζα ενώ με τους άλλους δύο αλγόριθμους παίρνουμε by default δέντρα χωρίς ρίζες. Ίσως θα μπορούσαμε να πούμε ότι το maximum likelihood δίνει πιό ακριβή αποτελέσματα εφόσον χρησιμοποιεί στατιστική ανάλυση προκυμένου να μας δώσει το τελικό δέντρο (επίσης το maximum likelihood προτιμάται από αρκετούς ερευνητές και reviewers). Παρ' όλα αυτά είναι άξιο παρατήρησης το γεγονός ότι ο maximum likelihood, δημιουργεί δέντρο στο οποίο τα νηματοειδή βρίσκονται εξελικτικά πολύ κοντά με τα βοωειδή, ενώ οι άλλοι δύο αλγόριθμοι δημιουργούν δέντρα όπου τα νηματοειδή αποτελούν μια ξεχωριστή κατηγορία η οποία βρίσκεται εξελικτικά πιο μακρυά από τους υπόλοιπους οργανισμούς.**

**Β) Σύγκριση φυλογενετικών δέντρων μετά από εφαρμογή αξιολόγησης bootstrap**

**Ακολουθεί σύγκριση των δέντρων μετά από εφαρμογή της μεθόδου αξιολόγησης bootstrap. Το bootstrap αξιολογεί την στατιστική ακρίβεια ενός φυλογενετικού δέντρου. Κάνει Ν επαναλήψεις (100 στην περίπτωση που παρατηρείται), όπου γίνεται shuffling των ακολουθιών 100 φορές και "παρατήρει" πόσες φορές εμφανίζονται κοινές αρχιτεκτονικές ανάμεσα στα 100 διαφορετικά δέντρα που έχουν προκύψει απ' τις επαναλήψεις. Στο τέλος θα μας δώσει ένα consensus tree το οποίο θα περιέχει τις κοινές αρχιτεκτονικές που είχαν τα δέντρα που προέκυψαν από τις επαναλήψεις. Στο δέντρο που προκύπτει με τον αλγόριθμο neighbor joining το μικρότερο bootstrap value είναι 66. Στο δέντρο που προκύπτει με τον αλγόριθμο maximum likelihood το μικρότερο bootstrap value έχει τιμή 51 και στο δέντρο που προκύπτει με τον αλγόριθμο upgma το μικρότερο bootstrap value έχει τιμή 93. Τα μικρά bootstrap values δείχνουν αρχιτεκτονικές κλάδων του δέντρου που δεν συναντώνται σε όλες τις επαναλήψεις που γίνονται (στην περίπτωση μας 100). Το γεγονός ότι μικρότερο bootstrap value είναι διαφορετικό στα τρία δέντρα είναι αναμενόμενο καθώς κάθε δέντρο έχει προκύψει από έναν διαφορετικό αλγόριθμο. Πάντως, είναι αξιοσημείωτο το γεγονός ότι και στις τρεις περιπτώσεις, τα μικρότερα bootstrap values, είναι στατιστικά σημαντικά γεγονός που ενισχύει την σκέψη ότι τα φυλογενετικά δέντρα που έχουν προκύψει είναι καλές προσεγγίσεις της πραγματικής εξελικτικής πορείας των διαφόρων ειδών.**