

BIO230: Εισαγωγή στην Υπολογιστική Βιολογία
Πρακτικό Εργαστήριο 2: Πολλαπλή Στοίχιση Αλληλουχιών
(Multiple Sequence Alignment)

Στέλλα Ταμανά, Βασίλης Προμπόνας
Λευκώσια, 2016

Εισαγωγικά στοιχεία

Κατά την διάρκεια αυτού του πρακτικού εργαστηρίου θα έχετε την δυνατότητα να εξοικειωθείτε με εργαλεία για την κατασκευή αλλά και την διαδραστική επεξεργασία και οπτικοποίηση πολλαπλών στοιχίσεων αλληλουχιών βιολογικών μακρομορίων.

Συγκεκριμένα, θα χρησιμοποιήσουμε μέσω του διαδικτύου την πρόσφατη τροποποίηση του αλγόριθμου CLUSTAL (CLUSTAL Omega; Sievers et al., 2011) που βρίσκεται διαθέσιμη μέσω διαδικτύου στο URL <http://www.ebi.ac.uk/Tools/msa/clustalo/> για τη δημιουργία πολλαπλών στοιχίσεων αλληλουχιών.

Για την επεξεργασία και οπτικοποίηση πολλαπλών στοιχίσεων θα χρησιμοποιηθεί η εφαρμογή Jalview (Waterhouse et al., 2009) η οποία βρίσκεται διαθέσιμη στο URL <http://www.jalview.org/>. Συγκεκριμένα, η εφαρμογή Jalview αναπτύχθηκε για τη διαδραστική επεξεργασία, ανάλυση και οπτικοποίηση των πολλαπλών στοιχίσεων αλληλουχιών, αλλά μπορεί επίσης να λειτουργήσει για το σχολιασμό αλληλουχιών (sequence annotation) με ανάκτηση σχολίων από τη βάση δεδομένων UniProt αλλά και δομικών πληροφοριών από 3D μοριακές δομές μέσω της βάσης δεδομένων Protein Data Bank (PDB). Επιπλέον, μπορεί να πραγματοποιήσει προγνώσεις δευτεροταγούς δομής πρωτεϊνών (secondary structure) και εγγενώς μη-δομημένων περιοχών (disorder prediction) και να κατασκευάσει φυλογενετικά δέντρα (phylogenetic trees). Περισσότερες πληροφορίες και διδακτικό υλικό που καλύπτει πολλές από τις δυνατότητες που έχει το Jalview μπορείτε να βρείτε στο εκπαιδευτικό υλικό που παρέχεται στην ιστοσελίδα του

λογισμικού (<http://www.jalview.org/training/Training-Videos>).

Στα πλαίσια αυτού του εργαστηρίου, θα πρέπει να είστε σε θέση να κατασκευάσετε μια πολλαπλή στοίχιση, να την επεξεργαστείτε και να παρουσιάσετε με κατάλληλο τρόπο τα αποτελέσματά σας.

Πρακτικό μέρος

Η αιμολυσίνη (hemolysin) είναι μια βακτηριακή τοξίνη η οποία έχει αιμολυτική δράση στα θηλαστικά, σχηματίζοντας πόρους στην κυτταρική μεμβράνη.

1. Εντοπίστε την πρωτεϊνική αλληλουχία της αιμολυσίνης από το εντεροβακτήριο *Escherichia coli* (στέλεχος K12).
2. Με βάση την αλληλουχία αυτή να εντοπίσετε τις ομόλογες (χωρίς πλεονασμό) στα εντεροβακτήρια [enterobacteria (taxid:91347)] εκτός του γένους *Escherichia*. Πόσες αλληλουχίες βρήκατε; Από πόσα διαφορετικά είδη/στελέχη;
3. Να σώσετε τις αλληλουχίες αυτές σε ένα αρχείο απλού κειμένου στον υπολογιστή σας.
4. Να εκτελέσετε πολλαπλή στοίχιση με το λογισμικό Clustal Omega και να σώσετε τη στοίχιση σας στον υπολογιστή σας. Πόσο βολική είναι η μορφή της στοίχισης που έχετε κατασκευάσει;
5. Χρησιμοποιήστε την εφαρμογή Jalview ώστε να οπτικοποιήσετε τη στοίχιση που σώσατε.
6. Μπορείτε να βρείτε τις ανά ζεύγος ομοιότητες των αλληλουχιών;
7. Μπορείτε να προσθέσετε στη στοίχιση την αλληλουχία της αιμολυσίνης από το *Escherichia coli* K12;
8. Να προβλέψετε τη δευτεροταγή δομή των πρωτεϊνών με τα εργαλεία που σας παρέχει το Jalview. Για την αιμολυσίνη από το *Escherichia coli* K12 είναι πειραματικά γνωστή η τρισδιάστατη δομή. Πώς συγκρίνεται η πρόβλεψη με το αποτέλεσμα του πειράματος;

Να ετοιμάσετε μια σύντομη αναφορά (μέχρι 1500 λέξεις) στην οποία θα περιγράψετε τον τρόπο με τον οποίο εκτελέσατε τα παραπάνω βήματα και τα αποτελέσματα τα οποία

πήρατε. Από τη σχετική βιβλιογραφία να περιγράψετε σε μια σύντομη παράγραφο ποιο πιστεύετε πως είναι το σπουδαιότερο νέο χαρακτηριστικό του λογισμικού Clustal Omega σε σχέση με τις παλιότερες εκδόσεις του Clustal.

Πηγές και Χρήσιμα Εργαλεία

- National Center of Biotechnology Information (NCBI) - <https://www.ncbi.nlm.nih.gov/>
- UniProtKB - <http://www.uniprot.org/>
- Protein Data Bank (PDB) - <http://www.rcsb.org/pdb/home/home.do>
- JalView - <http://www.jalview.org/>
- JalView Help page - <http://www.jalview.org/Help>

Βιβλιογραφία

Sievers F, Wilm A, Dineen DG, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* 7:539 doi:10.1038/msb.2011.75

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25: 1189-1191. doi:10.1093/bioinformatics/btp033