Βιοπληροφορική

Ενότητα 3η: Πολλαπλή ευθυγράμμιση

Σ. Γκέλης
Τμήμα Βιολογίας

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Άδειες Χρήσης

- Το παρόν εκπαιδευτικό υλικό υπόκειται σε άδειες χρήσης Creative Commons.
- Για εκπαιδευτικό υλικό, όπως εικόνες, που υπόκειται σε άλλου τύπου άδειας χρήσης, η άδεια χρήσης αναφέρεται ρητώς.

Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στα πλαίσια του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «Ανοικτά Ακαδημαϊκά Μαθήματα στο Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης» έχει χρηματοδοτήσει μόνο τη αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Περιεχόμενα ενότητας

- Ευθυγράμμιση ακολουθιών
- Αλγόριθμοι
- Δομή προγραμμάτων
- Ευθυγράμμιση (ή στοίχιση)
- Πολλαπλή ευθυγράμμιση (π.ε.)
- Οπτικοποιώντας μια ευθυγράμμιση

Ευθυγράμμιση ακολουθιών: τι;

σύγκριση ακολουθιών (DNA/RNA - πρωτεΐνες)

που έχουμε εισαγωγή ή διαγραφή σε κάθε ακολουθία

καλύτερη ευθυγράμμιση:

- ✓ μέγιστος αριθμός ταύτισης
- ✓ ελάχιστος αριθμός κενών & μη όμοιων

Ευθυγράμμιση ακολουθιών: γιατί;

ομοιότητα (similarity) μεταξύ των ακολουθιών

μάλλον υπάρχει και κάποιας μορφής ομολογία (homology)

ομοιότητα (similarity): μοιάζουν με βάση κάποιο χαρακτηριστικό τους, π.χ. στην ακολουθία

ομολογία (homology): δύο ή περισσότερες ακολουθίες έχουν ένα κοινό πρόγονο (εξελικτική ιστορία)

[βασική υπόθεση: τα γονίδια και οι πρωτεΐνες σχετίζονται εξελικτικά]

Ευθυγράμμιση ακολουθιών: πώς; (1/2)

```
chite   ---ADKPKRPLSAYMLWLNSARESIKRENPDFK-VTEVAKKGGELWRGLKD
wheat   --DPNKPKRAPSAFFVFMGEFREEFKQKNPKNKSVAAVGKAAGERWKSLSE
trybr   KKDSNAPKRAMTSFMFFSSDFRS----KHSDLS-IVEMSKAAGAAWKELGP
mouse   -----KPKRPRSAYNIYVSESFQ----EAKDDS-AQGKLKLVNEAWKNLSP
              ***. ::: .: ..  .    :  . .      *  .  *: *

chite   AATAKQNYIRALQEYERNGG-
wheat   ANKLKGEYNKAIAAYNKGESA
trybr   AEKDKERYKREM---------
mouse   AKDDRIRYDNEMKSWEEQMAE
        *   : .* . :
```

γίνεται και «με το χέρι»;

Ευθυγράμμιση ακολουθιών: πώς; (2/2)

γιατί χρειάζεται η πληροφορική;

αποδοτικότερη διαχείριση & αποθήκευση της πληροφορίας

πολλές ακολουθίες [συνεχώς αλλάζουν & προστίθενται καινούργιες]

μείωση χρόνου υπολογισμών

υπολογισμός στατιστικών, πχ. e-value

διάθεση αποτελεσμάτων & βιβλιογραφίας στο διαδίκτυο

...ας πάμε λίγο πίσω...

[ιντερμέδιο πληροφορικής]

Αλγόριθμοι (1/2)

συγκεκριμένη ακολουθία βημάτων που οδηγεί στη λύση ενος προβλήματος, πχ. ευθυγράμμιση ακολουθιών πρέπει να:

1. βρίσκει την σωστή λύση σε αποδεκτά δεδομένα
2. υπολογίζει την λύση σε λογικό χρόνο

χαρακτηριστικά αλγόριθμου [στην πληροφορική]:

1. πολυπλοκότητα (χρονική, χωρική)
2. ολοκλήρωση (completeness)
3. βέλτιστος (optimal)

Αλγόριθμοι (2/2)

Χαρακτηριστικά αλγορίθμου:

1. πολυπλοκότητα – συμβολίζεται με O()
   χωρική: μνήμη που χρειάζεται
   χρονική: πόσος χρόνος χρειάζεται να “τρέχει”
   διαφοροποίηση ανάμεσα στην μέση πολυπλοκότητα

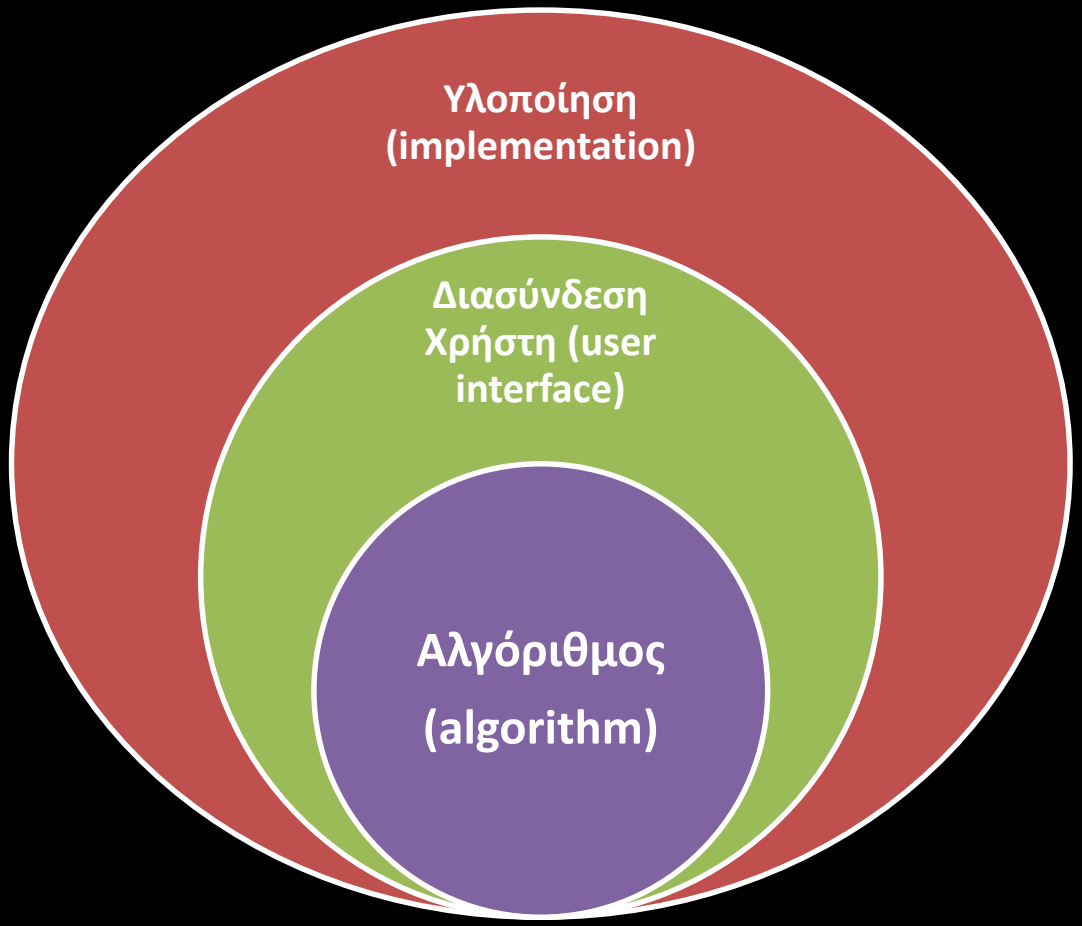
και

στη χειρότερη περίπτωση

1. ολοκλήρωση
   αν υπάρχει έστω μία λύση, εγγυάται ό,τι θα την βρει

1. βέλτιστος
   εγγυάται ό,τι θα βρει την καλύτερη δυνατή λύση

Δομή προγραμμάτων (1/2)



Δομή προγραμμάτων (2/2)

αλγόριθμος:

BLAST

CLUSTALW2

κλπ

υλοποίηση:

σε τι γλώσσα προγραμματισμού υλοποιήθηκε

ποιές δομές δεδομένων χρησιμοποιήθηκαν

τι δομές προγραμματισμού χρησιμοποιήθηκαν

διασύνδεση χρήστη

γραφικά (Graphical User Interface - GUI)

γραμμή εντολών (Command Line Interface - CLI)

1. ...επιστροφή στην ευθεία...

Ευθυγράμμιση (ή στοίχιση) (1/5)

ένα παράδειγμα

Ευθυγράμμιση (ή στοίχιση) (2/5)

Τύποι ευθυγράμμισης:

- δυαδική (pairwise): δύο ακολουθίες μεταξύ τους
- πολλαπλή (multiple): πολλές ακολουθίες μαζί

Ευθυγράμμιση (ή στοίχιση) (3/5)

Τύποι ευθυγράμμισης:

- καθολική (global - σε όλο το μήκος)
- τοπική (local - σε υποπεριοχές)

Ευθυγράμμιση (ή στοίχιση) (4/5)

Αλγόριθμοι δυαδικής ευθυγράμμισης:

- Needleman-Wunsch, 1970 καθολική
- Smith -Waterman, 1981 τοπική
- FASTA (Pearson 1990) τοπική
- BLAST (Basic Local Alignment Search Tool), Altschul *et al.* 1990,
  1997 τοπική

Ευθυγράμμιση (ή στοίχιση) (5/5)

Διαφορές αλγορίθμων ευθυγράμμισης:

- διαφέρουν ως προς την πολυπλοκότητά τους

  μικρή χωρική και χρονική πολυπλοκότητα, π.χ. BLAST

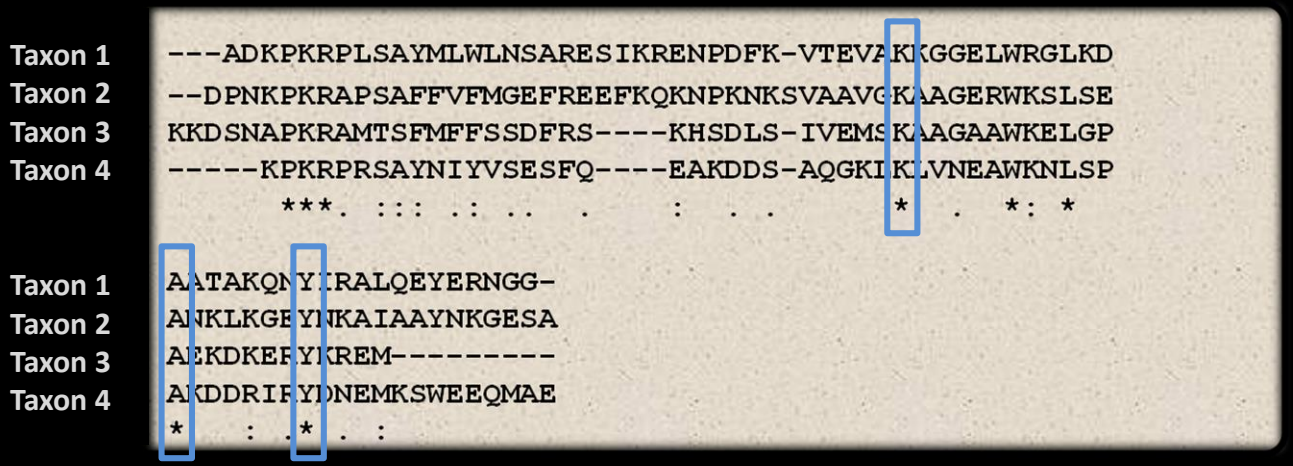
  μεγάλη χωρική και χρονική πολυπλοκότητα, π.χ. Smith-Waterman

- κάποιοι δεν είναι βέλτιστοι, π.χ.

  BLAST ➡ μη βέλτιστος αλγόριθμος

  Smith – Waterman ➡ βέλτιστος αλγόριθμος

Πολλαπλή ευθυγράμμιση (1/29)



- **δομικά κριτήρια**
  - τα στοιχεία τοποθετούνται έτσι ώστε αυτά που παίζουν παρόμοιο ρόλο να είναι στην ίδια στήλη.
- **εξελικτικά κριτήρια**
  - τα στοιχεία τοποθετούνται έτσι ώστε αυτά που έχουν τον ίδιο πρόγονο να βρίσκονται στην ίδια στήλη
- **κριτήρια ομοιότητας**
  - όσο περισσότερα παρόμοια στοιχεία γίνεται να βρίσκονται στην ίδια στήλη

Πολλαπλή ευθυγράμμιση (2/29)

- ευθυγράμμιση περισσότερων των δύο ακολουθιών
- ιδανική προσέγγιση: ευθυγράμμιση όλων των ακολουθιών μεταξύ τους ταυτόχρονα
- γρήγορη προσέγγιση: σταδιακή δυαδική ευθυγράμμιση
  (progressive pairwise alignment)

Πολλαπλή ευθυγράμμιση (3/29)

Τι περιέχει μια π.ε. ;

- ό,τι θέλεις εσύ να περιέχει
- μια π.ε. μπορεί να ιδωθεί ως:

  -μια εξελικτική καταγραφή

  -μια σύνοψη οικογένειας πρωτεϊνών

  -μια συλλογή πειραμάτων που έκανε για σένα η φύση

Πολλαπλή ευθυγράμμιση (4/29)

και σε τι χρησιμεύει;

Main applications of multiple sequence alignments

εφαρμογή	*Procedure*
δημιουργία οικογένειας πρωτεϊνών	A good multiple alignment can help convincing you that an uncharacterized sequence is really a member of a protein family.
δημιουργία φυλογενετικών δέντρων – εύρεση εξελικτικών σχέσεων	If you carefully chose the sequences to include in your multiple alignment, you can reconstruct the history of these proteins.
αναγνώριση συντηρητικών περιοχών	By discovering very conserved positions you can identify a region that is characteristic of a function (in proteins or in nucleic acid sequences).
σύγκριση ακολουθιών μεταξύ διαφορετικών οργανισμών	It is possible to turn a multiple sequence alignment into a profile that describes a protein family or a protein domain. You can use this profile to scan databases for new members of the family.
εύρεση ρυθμιστικών στοιχείων	You can turn a DNA multiple alignment of a binding site into a weight matrix and scan other DNA sequences for potential similar binding sites.
πρόβλεψη δευτεροταγούς – τριτοταγούς δομής πρωτεΐνης	A good multiple alignment can give you an almost perfect prediction of your protein secondary structure for proteins or RNA. Sometimes it can also help building a 3-D model.

παράδειγμα

Πολλαπλή ευθυγράμμιση (5/29)

Και που είναι το πρόβλημα;

Biology:

What is **a** good alignment?

Computation:

What is **the** good alignment?

Πολλαπλή ευθυγράμμιση (6/29)

υπάρχουν πολλές (πόσες;) πιθανές ευθυγραμμίσεις και τι κάνουμε;

διευθέτηση δύο ή περισσότερων αλληλουχιών (νουκλεοτιδικών ή πρωτεϊνικών σε ένα πλέγμα (μήτρα)

Στοιχεία (νουκλεοτίδια, αμινοξέα) της ίδιας σειράς προέρχονται από το ίδιο βιολογικό μακρομόριο (πρωτεΐνη ή νουκλεϊκό οξύ). Τα στοιχεία διευθετούνται με τη σειρά που εμφανίζονται στο μακρομόριο:

- από το N στο C άκρο στις πρωτεΐνες
- από το 5’ στο 3’ στα νουκλεϊκά οξέα

Πολλαπλή ευθυγράμμιση (7/29)

κάθε κελί περιλαμβάνει ένα μόνο στοιχείο [είτε ένα στοιχείο είτε ένα κενό (gap)]

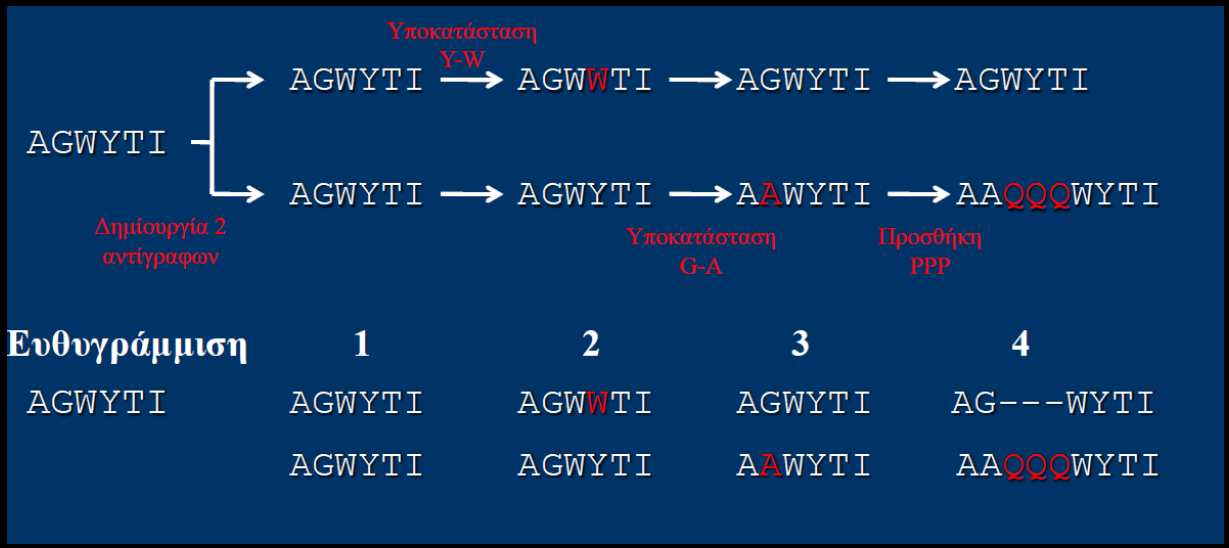
τα στοιχεία της ίδιας στήλης είναι

- δομικά ισοδύναμα
- εξελικτικά ισοδύναμα (ομόλογα)

Πολλαπλή ευθυγράμμιση (8/29)

εξελικτική ισοδυναμία=ομολογία

αναφερόμενοι στην ίδια στήλη, η ιστορία κάθε στοιχείου θα πρέπει να αναζητηθεί στο αντίστοιχο στοιχείο της προγονικής αλληλουχίας, όπου κάθε αλλαγή οφείλεται σε σημειακές αλλαγές



Πολλαπλή ευθυγράμμιση (9/29)

ένα ζεύγος αλληλουχιών μπορεί να ευθυγραμμιστεί γράφοντας την μία αλληλουχία κάτω από την άλλη με τέτοιο τρόπο ώστε να μεγιστοποιηθεί ο αριθμός των νολεοτιδίων που ταιριάζουν, βάζοντας κενά (gaps) στην μια ή στην άλλη αλληλουχία όταν απαιτείται

```
AF486227    TACGAAAACACCACCCAATCCTAAGAA
AF486228    TACGAAAACACGACCCAATCCTAAAAA
AF486223    TACGAAAACACCACCCTATCCTAAAAA
```

η ευθυγράμμιση γίνεται συνήθως με ειδικά υπολογιστικά πακέτα, που χρησιμοποιούν συγκεκριμένους αλγόριθμους. Οι περισσότεροι αλγόριθμοι αρχίζουν συγκρίνοντας την ομοιότητα των αλληλουχιών ανά ζεύγη,και ευθυγραμμίζοντας πρώτα τις δύο αλληλουχίες με τη μεγαλύτερη ομοιότητα. Οι άλλες αλληλουχίες, βάσει της σειράς ομοιότητας, προστίθενται σταδιακά.

Πολλαπλή ευθυγράμμιση (10/29)

Όταν σε μια ομάδα αλληλουχιών έχουν προστεθεί κάποια κενά, τότε το τελικό alignment συχνά βελτιώνεται από τον ίδιο τον ερευνητή με manual editing. Η απόκτηση μιας καλής ευθυγράμμισης είναι ίσως το πιο σημαντικό βήμα ώστε να εκτιμήσουμε ένα σωστό φυλογενετικό δέντρο

```
AF486227    TACGAA--AACACCACC---CAATCCTAAGAA
AF486228    TACGAA--AACACGACCGGGCAATCCTAAAAA
AF486223    TACGAATTAACACCACCGGGCTATCCTAAAAA
```

Είναι αναγκαίο να ορίσουμε τον αριθμό των gaps ώστε το τελικό αποτέλεσμα να έχει βιολογική υπόσταση

Για το λόγο αυτό χρησιμοποιείται ένα σύστημα σκοραρίσματος όπου τα ταιριάσματα παίρνουν ένα θετικό βαθμό και τα κενά ένα αρνητικό, που είναι γνωστό ως «gap penalty»

Πολλαπλή ευθυγράμμιση (11/29)

υπάρχουν πολλές (πόσες;) πιθανές ευθυγραμμίσεις

για παράδειγμα συγκρίνετε

-GCGC-ATGGATTGAGCGA
TGCGCCATTGAT-GACC-A

με

------GCGCATGGATTGAGCGA
TGCGCC----ATTGATGACCA--

ποια είναι η καλύτερη;

παράδειγμα 2

Πολλαπλή ευθυγράμμιση (12/29)

και που είναι (πάλι) το πρόβλημα;

-ο αριθμός των πιθανών συνδυασμών ευθυγράμμισης αυξάνει ραγδαία σε σχέση με των αριθμό των ακολουθιών

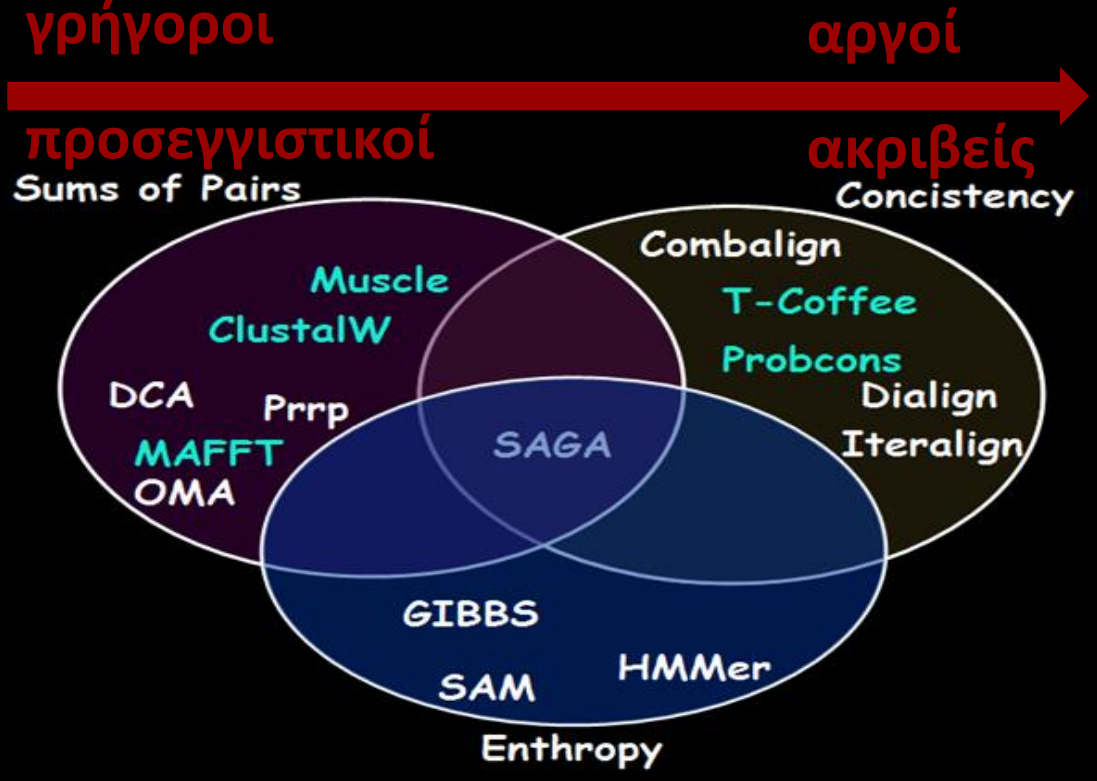
-χρονοβόρα διαδικασία

- συχνά δύσκολο να πούμε ποια ευθυγράμμιση είναι η καλύτερη

Πολλαπλή ευθυγράμμιση (13/29)

και ποιες είναι οι λύσεις;

μια κάποια ταξινόμηση των αλγορίθμων



Πολλαπλή ευθυγράμμιση (14/29)

τύποι αλγόριθμων π.ε

- σταδιακοί [progressive]: **ClustalW**
- επαναληπτικοί [iterative]: **Muscle**
- concistency based: **T-Coffee** και **Probcons**

Πολλαπλή ευθυγράμμιση (15/29)

Πολλοί αλγόριθμοι online:

http://www.ebi.ac.uk/Tools/msa/

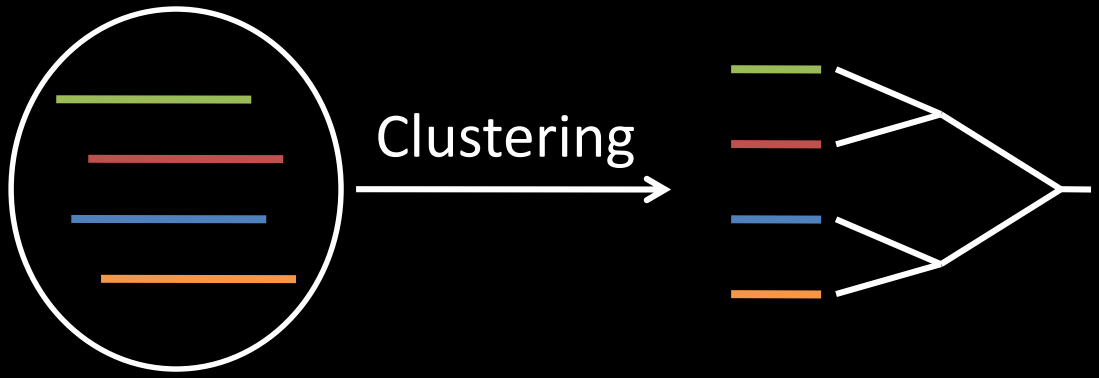
Πολλαπλή ευθυγράμμιση (16/29)

ClustalW2

- Global multiple sequence alignment
- δυνατότητα παραγωγής φυλογενετικών δέντρων
- ευθυγράμμιση είναι γρήγορη (progressive alignment)
- δυνατότητα επεξεργασίας με JalView της ευθυγράμμισης
- μέχρι 500 ακολουθίες ή 10MB αρχεία
- http://www.ebi.ac.uk/Tools/clustalw2

Πολλαπλή ευθυγράμμιση (17/29)

πως δουλεύει το ClustalW2;



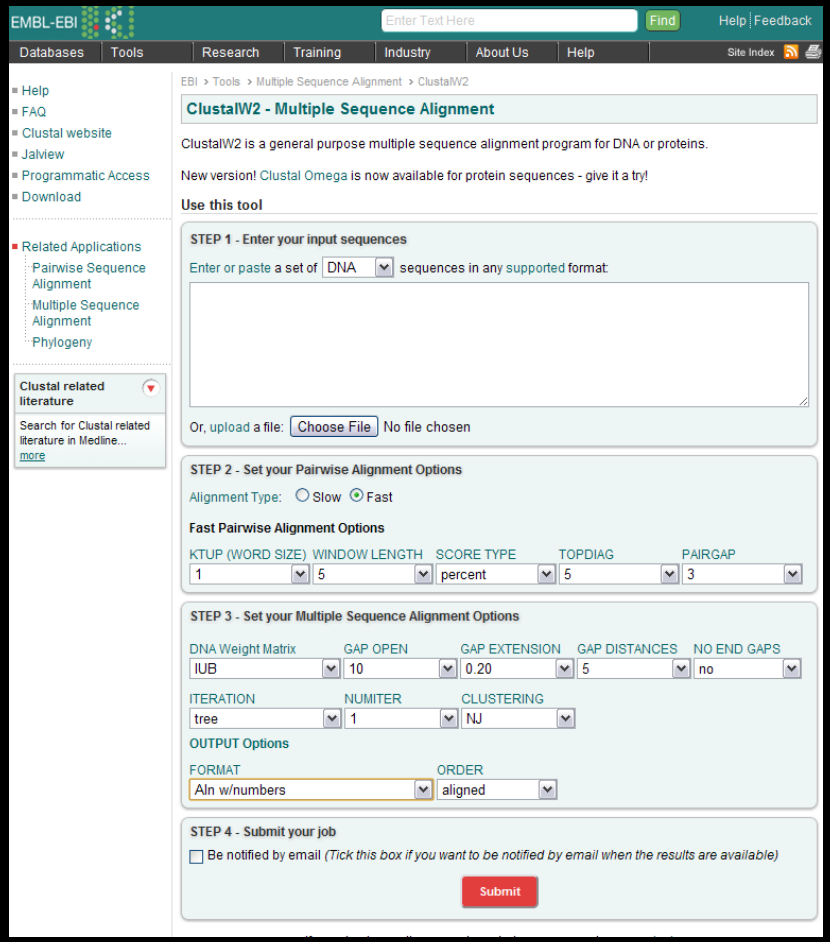
1. Εκτελεί δυαδική ευθυγράμμιση για όλα τα πιθανά ζεύγη των ακολουθιών – Οι ομοιότητες και τα αποτελέσματα των ευθυγραμμίσεων αποθηκεύονται σε ένα πίνακα

Πολλαπλή ευθυγράμμιση (18/29)

πως δουλεύει το ClustalW2;

2. Μετατρέπει τον πίνακα με τις ομοιότητες σε εξελικτική απόσταση μεταξύ των ακολουθιών
3. Φτιάχνει ένα βοηθητικό δέντρο (guide tree) που δείχνει πως θα συνδυαστούν οι προηγούμενες δυαδικές ευθυγραμμίσεις
4. Κάνει σταδιακή ευθυγράμμιση των ακολουθιών προσθέτοντας σταδιακά κάθε ακολουθία με βάση την σειρά που προέκυψε από το 2ο βήμα

Πολλαπλή ευθυγράμμιση (19/29)



Εικόνα 1: **το user interface του ClustalW2.**

- Alignment: χρήση γρήγορου αλγόριθμου (σταδιακή δυαδική ευθυγράμμιση) ή όχι
- Iteration: ο αριθμός των επαναλήψεων μέχρι να τερματίσει ο αλγόριθμος και να μας δώσει αποτελέσματα
- Fast Pairwise Alignment: επιλογές που σχετίζονται με γρήγορη ευθυγράμμιση KTUP, Window, Score, Topdiag, Pairgap
- Gaps: βαθμοί σε διάφορους τύπους κενών

http://www.ebi.ac.uk/Tools/msa/clustalw2/help/index.html#matrix

Πολλαπλή ευθυγράμμιση (20/29)

παράδειγμα - πρωτεΐνη fosb

εισαγωγή σε Fasta μορφοποίηση:

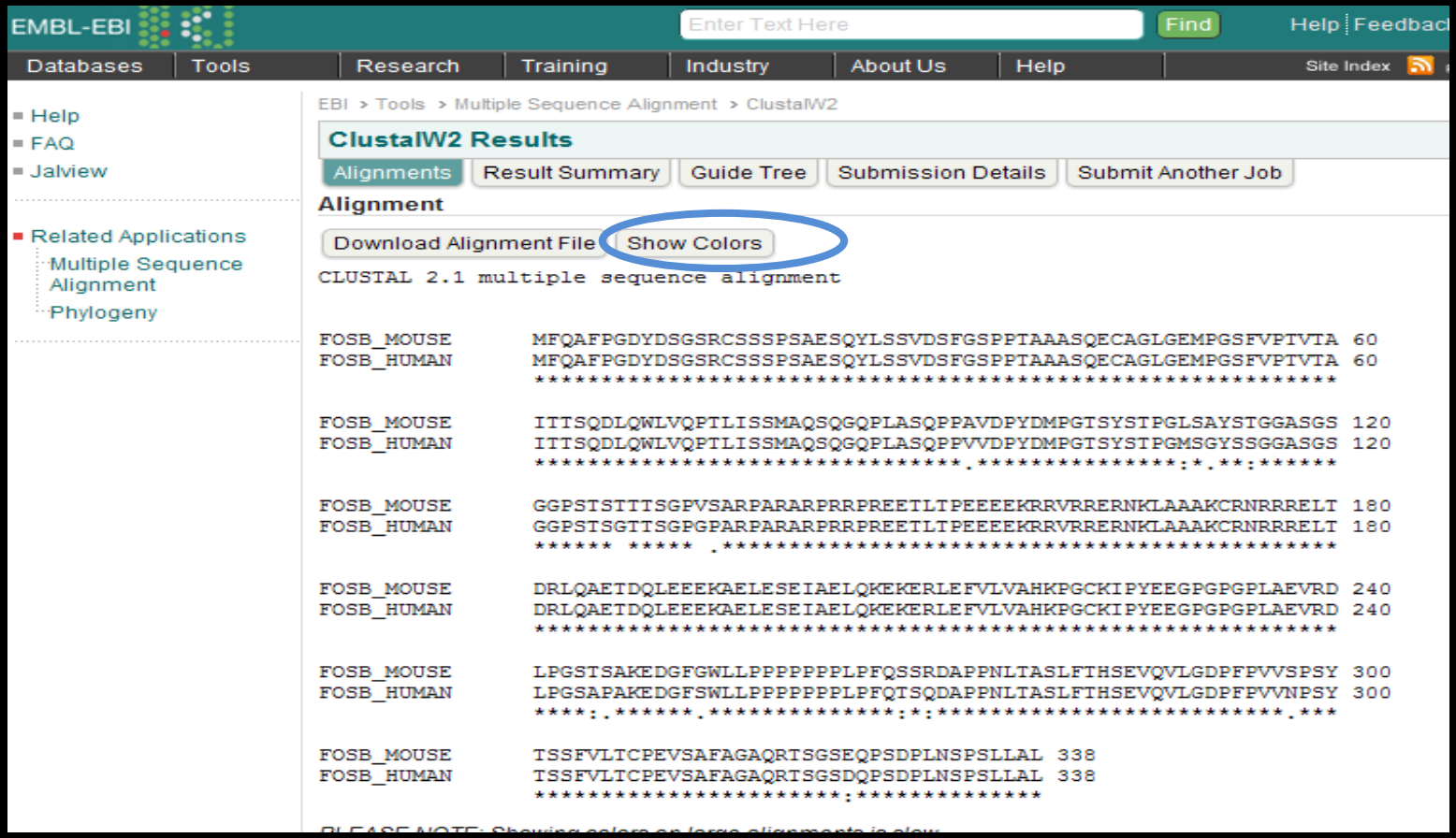
>FOSB_MOUSE Protein fosB

MFQAFPGDYD SGSRCSSSPS AESQYLSSVD SFGSPPTAAA SQECAGLGEM
PGSFVPTVTA ITTSQDLQWL VQPTLISSMA QSQGQPLASQ PPAVDPYDMP
GTSYSTPGLS AYSTGGASGS GGPSTSTTTS GPVSARPARARPRRPREETL
TPEEEEKRRV RRERNKLAAAKCRNRRRELT DRLQAETDQL EEEKAELESE
IAELQKEKER LEFVLVAHKP GCKIPYEEGP GPGPLAEVRD LPGSTSAKED
GFGWLLPPPP PPPLPFQSSR DAPPNLTASL FTHSEVQVLG DPFPVVSPSY
TSSFVLTCPE VSAFAGAQRT SGSEQPSDPL NSPSLLAL

>FOSB_HUMAN Protein fosB

MFQAFPGDYD SGSRCSSSPS AESQYLSSVD SFGSPPTAAA SQECAGLGEM
PGSFVPTVTA ITTSQDLQWL VQPTLISSMA QSQGQPLASQ PPVVDPYDMP
GTSYSTPGMS GYSSGGASGS GGPSTSGTTS GPGPARPARARPRRPREETL
TPEEEEKRRV RRERNKLAAAKCRNRRRELT DRLQAETDQL EEEKAELESE
IAELQKEKER LEFVLVAHKP GCKIPYEEGP GPGPLAEVRD LPGSAPAKED
GFSWLLPPPP PPPLPFQTSQ DAPPNLTASL FTHSEVQVLG DPFPVVNPSY
TSSFVLTCPE VSAFAGAQRT SGSDQPSDPL NSPSLLAL

Πολλαπλή ευθυγράμμιση (21/29)



Εικόνα 2: Αποτελέσματα ευθυγράμμισης του ClustalW2.

Πολλαπλή ευθυγράμμιση (22/29)

Clustal Ω: νέο Clustal

http://www.clustal.org/

http://www.ebi.ac.uk/Tools/msa/clustalo/help/index.html

http://www.ebi.ac.uk/Tools/msa/clustalo/help/faq.html#11

Πολλαπλή ευθυγράμμιση (23/29)

Προβλήματα ClustalW2

- όταν οι ακολουθίες μοιάζουν σε ορισμένα σημεία αλλά όχι σ’ όλο τους το μήκος
- όταν μία από τις ακολουθίες έχει ένα μεγάλο τμήμα ενσωματωμένο σε σχέση με τις άλλες
- αν μια ακολουθία έχει ένα επαναλαμβανόμενο τμήμα πολλές φορές

Πολλαπλή ευθυγράμμιση (24/29)

T-Coffee [Tree based Consistency Objective Function For alignmEnt Evaluation]

- **Πραγματοποιεί πολλαπλή ευθυγράμμιση ακολουθιών**
- **Μπορεί να ευθυγραμμίσει ακολουθίες από πρωτεΐνες, DNA και RNA**

Πολλαπλή ευθυγράμμιση (25/29)

πως δουλεύει το T-Coffee;

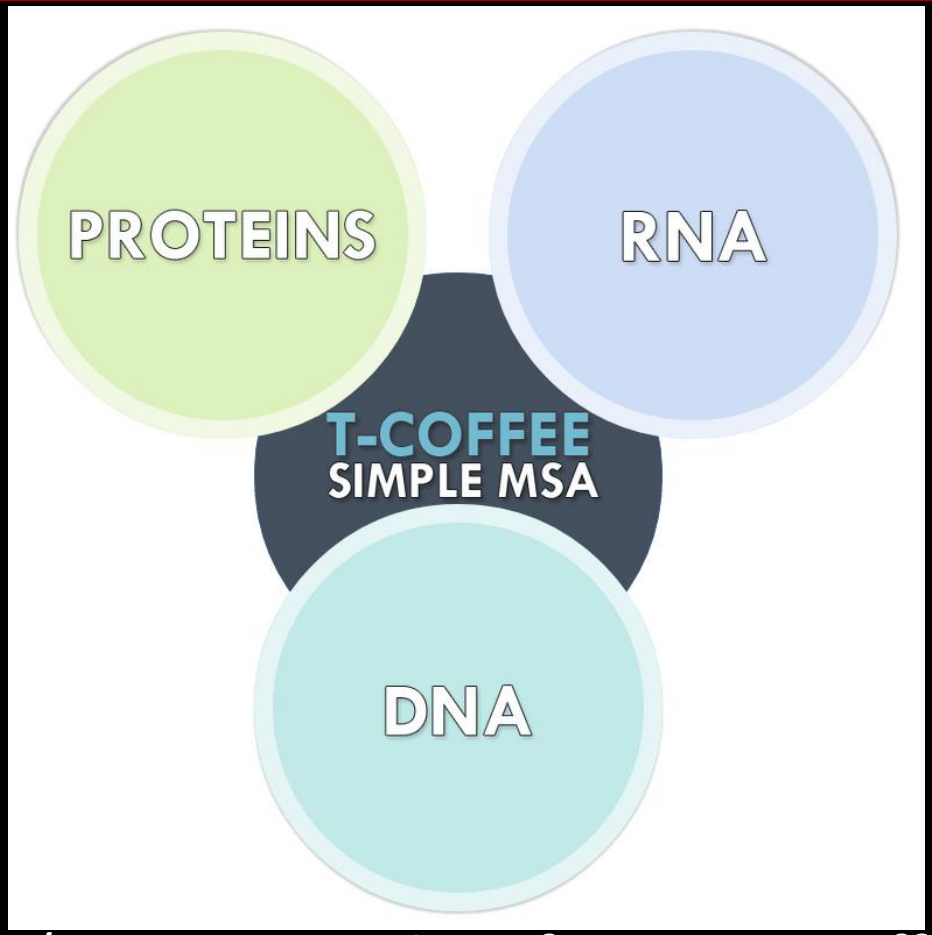
συνδυάζει αποτελέσματα από διαφορετικές μεθόδους ευθυγράμμισης: πχ. ClustalW2, δομική ευθυγράμμιση (structural alignment)

αν δώσουμε απλά κάποιες ακολουθίες τότε το T-Coffee θα τις συγκρίνει ανά δυο (global & local alignments)

θα συνδυάσει τις ευθυγραμμίσεις σε μία πολλαπλή ευθυγράμμιση

αντί της χρήσης πίνακα δημιουργείται μια βιβλιοθήκη που προκύπτει από την τοπική και ολική στοίχιση όλων των ζευγών ακολουθιών. Η βιβλιοθήκη αυτή, που τροποποιείται καθώς ο αλγόριθμος προχωρά, χρησιμοποιείται για την βαθμολόγηση

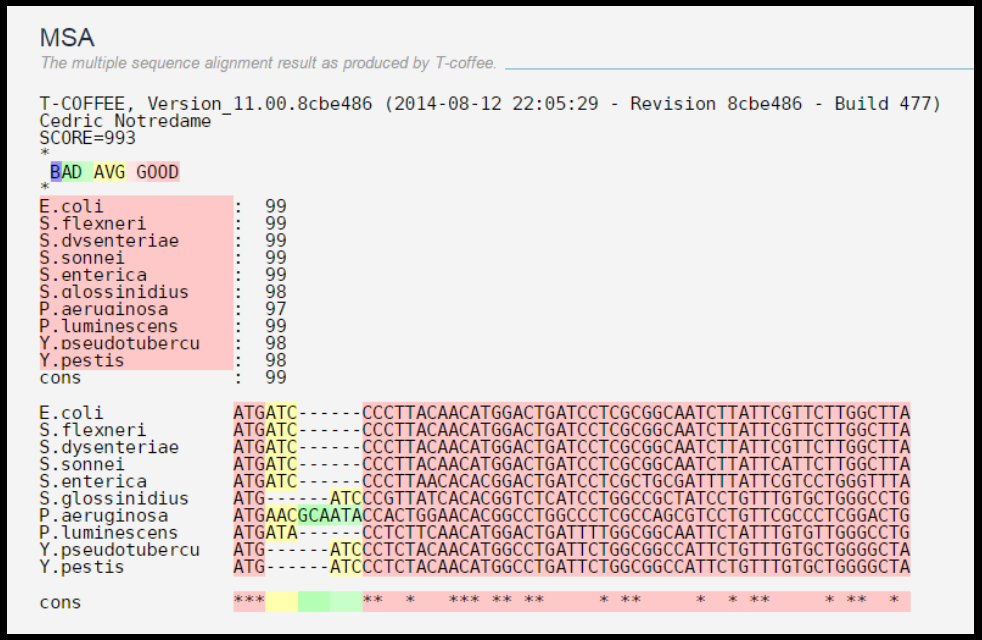
Πολλαπλή ευθυγράμμιση (26/29)



Εικόνα 3: **το user interface του T-Coffee**
http://tcoffee.crg.cat/apps/tcoffee/index.html

Πολλαπλή ευθυγράμμιση (27/29)

Εισάγουμε τις ακολουθίες που θέλουμε να ευθυγραμμίσουμε (http://tcoffee.crg.cat/apps/tcoffee/do:regular) και το T-coffee κάνει την ευθυγράμμιση



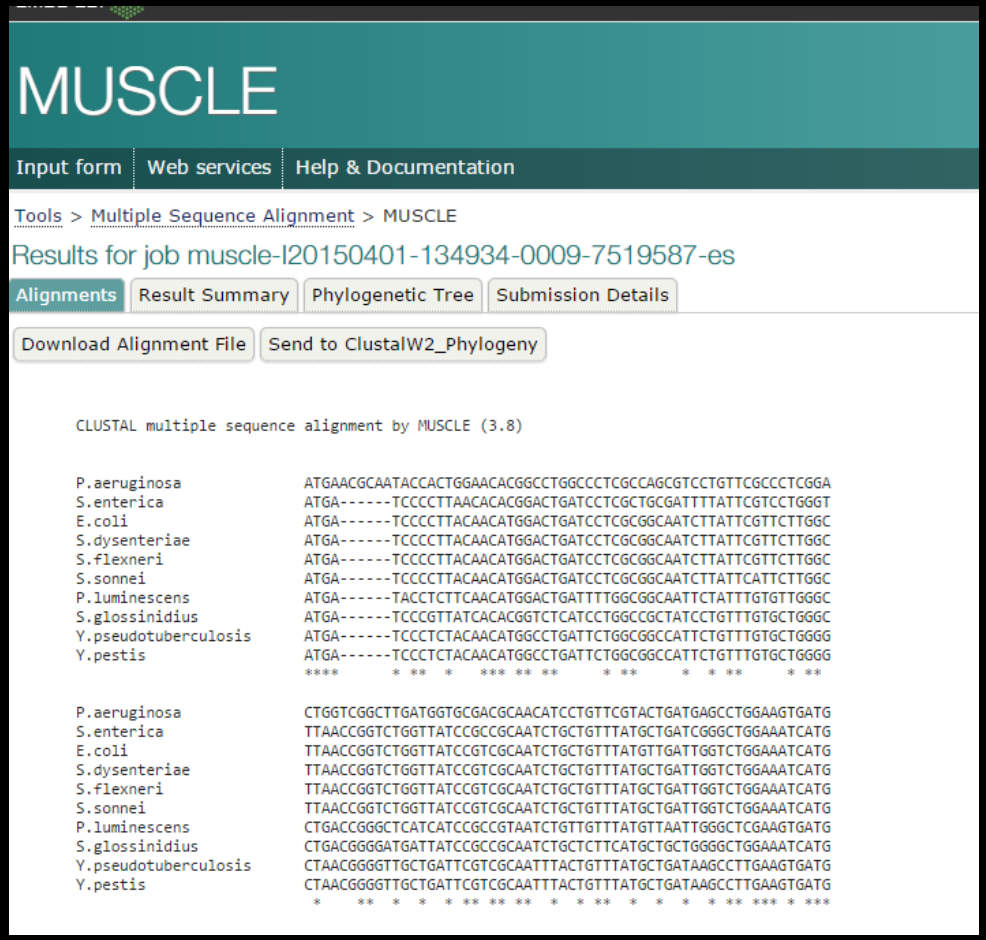
Εικόνα 4: Αποτελέσματα ευθυγράμμισης του T-coffee.

Πολλαπλή ευθυγράμμιση (28/29)

MUSCLE [**MU**ltiple **S**equence **C**omparison by **L**og-**E**xpectation]

- ακρίβεια που σε περιπτώσεις ξεπερνάει ClustalW & T-Coffee
- πολλαπλή ευθυγράμμιση πρωτεϊνών
- αρκετά γρήγορος αλγόριθμος
- 2 - 3 φορές πιο γρήγορος από CLUSTALW2 κατά μέσο όρο
- μπορεί να ευθυγραμμίσει 1000 πρωτεΐνες μέσου μήκους 282 αμινοξέων σε 21 sec. σε ένα μέσης υπολογιστικής ισχύος μηχάνημα

Πολλαπλή ευθυγράμμιση (29/29)



Εικόνα 5: Αποτελέσματα ευθυγράμμισης του MUSCLE.

Οπτικοποιώντας μια ευθυγράμμιση (1/4)

JalView

- γραφική αναπαράσταση της ευθυγράμμισης
- δυνατότητα επεξεργασίας της ευθυγράμμισης

http://www.jalview.org/

Οπτικοποιώντας μια ευθυγράμμιση (2/4)

JalView

πολλές λειτουργίες

τα λογισμικά τείνουν να κάνουν
το ένα τη «δουλειά» του άλλου

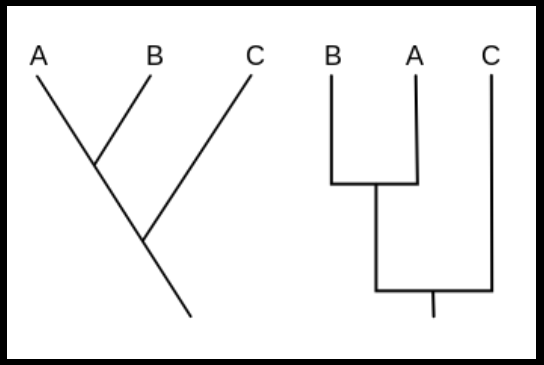
Οπτικοποιώντας μια ευθυγράμμιση (3/4)

JalView

- επεξεργασία αλληλουχιών
- πλοήγηση στις αλληλουχίες
- “conservation colouring”
- απεικόνιση τρισδιάστατων δομών πρωτεϊνών

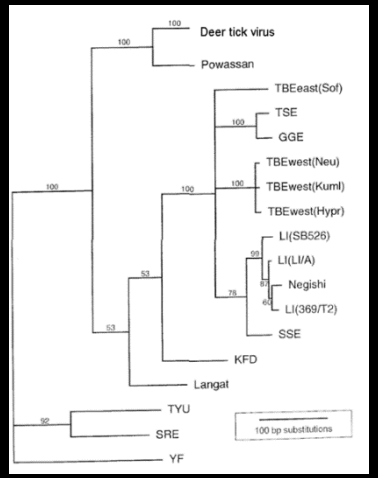
Οπτικοποιώντας μια ευθυγράμμιση (4/4)

λίγο πριν απ' τα δέντρα



Εικόνα 6: κλαδόγραμμα : όλοι οι κλάδοι έχουν το ίδιο μήκος,

http://commons.wikimedia.org/wiki/File:Identical_cladograms.svg
by Alexei Kouprianov, CC-BY-SA-3.0,
https://creativecommons.org/licenses/by-sa/3.0/deed.en .



Εικόνα 7: φυλόγραμμα : οι κλάδοι έχουν διαφορετικό μήκος (δείχνει και χρονική μεταβολή),

http://commons.wikimedia.org/wiki/File:Tick-borne_encephalitis_phylogram_simplified.gif
by Telford et al.

Ανακεφαλαίωση (1/2)

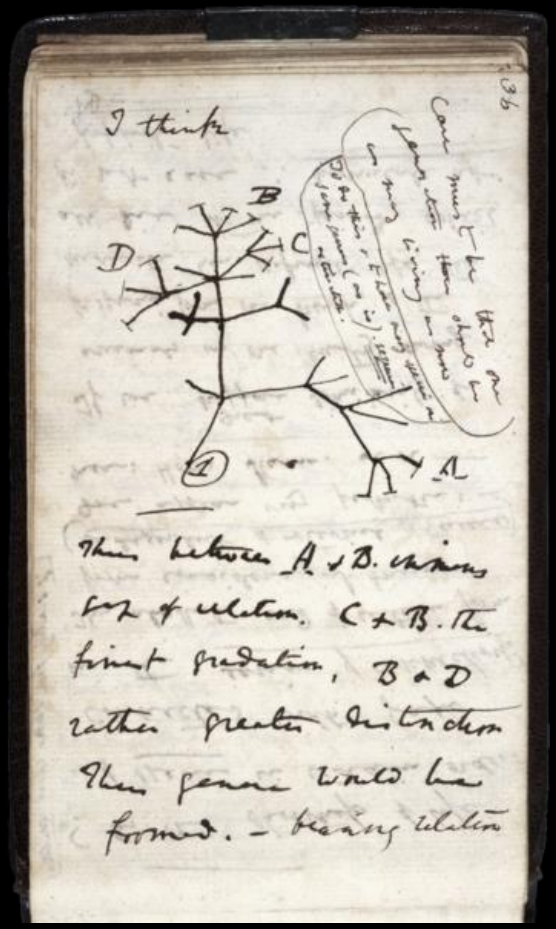
να θυμάστε:

- ✓ κάθε πρόγραμμα = υλοποίηση συγκεκριμένου αλγορίθμου
- ✓ οι αλγόριθμοι έχουν συγκεκριμένα χαρακτηριστικά, π.χ. πολυπλοκότητα, ολοκλήρωση, κτλ.
- ✓ πολλαπλή ευθυγράμμιση = ευθυγράμμιση πολλών ακολουθιών (>2) ταυτόχρονα
- ✓ κυριότερα προγράμματα για πολλαπλή ευθυγράμμιση: ClustalW2, T-Coffee, MUSCLE

Ανακεφαλαίωση (2/2)

συνήθη βήματα στην π.ε.:

- ✓ ανάκτηση αλληλουχιών
- ✓ ClustalW2 ή T-coffee ή MUSCLE
- ✓ JalView ή αντίστοιχο εργαλείο
- ✓ εξαγωγή συμπερασμάτων
- ✓ φυλογενετικό δέντρο



είστε έτοιμοι για το πρώτο φυλογενετικό σας δέντρο!

αν και το πρώτο φυλογενετικό δέντρο φτιάχτηκε πριν εφευρεθούν οι υπολογιστές και πριν ανακαλυφθούν τα γονίδια

Εικόνα 8: Το δέντρο του Δαρβίνου,

http://commons.wikimedia.org/wiki/File:Darwin_Tree_1837.png

Βιβλιογραφία - Πηγές

1. Αρχεία Βοήθειας (Help files) των προγραμμάτων που παρουσιάστηκαν
2. Lecture notes by Per Kraulis, KTH Bioinformatics Oct-Dec 2001, http://www.avatar.se/molbioinfo2001/index.html
3. An Introduction to Bioinformatics Algorithms (Computational Molecular Biology) by Neil C. Jones, Pavel A. Pevzner, ISBN: 978-0262101066
4. *Robert C Edgar, «MUSCLE: a multiple sequence alignment* method with reduced time and space complexity», BMC Bioinformatics 2004 vol. 5, pp. 113
5. Notredame C., Higgins D., Heringa J., «T-Coffee: A novel method for multiple sequence alignments». *Journal of Molecular Biology 2000, 302: 205-217*
6. L. Bromham. 2008. Reading the story in DNA. Oxford University Press
7. Cédric Notredame, T-Coffee: What's New in The Grinder *http://www.google.gr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CDgQFjAA&url=http%3A%2F%2Fwww.tcoffee.org%2FPresentations%2Fmcoffee.ppt&ei=fAhzT67zBtDZ4QSyloGHDw&usg=AFQjCNFWjw-WddSAqN4TF81VLzQguHzIKg*
8. *ClustalW and ClustalX version 2 Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ and Higgins DG Bioinformatics 2007 23(21): 2947-2948*

Σημείωμα Αναφοράς

Σημείωμα Αδειοδότησης

Σας ευχαριστώ!

Σπύρος Γκέλης

Τμήμα Βιολογίας Α.Π.Θ.

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο

ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Τέλος ενότητας

Επεξεργασία: Στυλιανή Μηνούδη

Θεσσαλονίκη, Εαρινό εξάμηνο 2014

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
- το Σημείωμα Αδειοδότησης
- τη δήλωση Διατήρησης Σημειωμάτων

μαζί με τους συνοδευόμενους υπερσυνδέσμους.