



Βιοπληροφορική

Ενότητα 10:

Αναζήτηση Ομοιοτήτων σε
ΒΔ Ακολουθιών - Blast,
(1/2) 1ΔΩ

Τμήμα: Βιοτεχνολογίας

Όνομα καθηγητή: Τ. Θηραίου



Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο



ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης





Μαθησιακοί Στόχοι

- Αναφορά στις παραλλαγές του BLAST.
- Εξοικείωση με τη διεπαφή του BLAST.



Λέξεις Κλειδιά

- Λέξεις κλειδιά: Βάσεις δεδομένων BLAST, Παραλλαγές BLAST, Ανταποδοτικό BLAST.
- Key words: BLAST databases, BLAST family of programs, PSI-BLAST, PHI-BLAST, Reciprocal BLAST, BLAST interface.



Βάσεις Δεδομένων BLAST 1/2

● Πρωτεϊνικές ΒΔ:

– nr:

- περιεκτική συλλογή αλληλουχιών χωρίς πλεονασμούς.

– month:

- nr με τις αλληλουχίες των τελευταίων 30 ημερών.

– swissprot:

- αλληλουχίες της SWISS-PROT.

– pdb:

- αλληλουχίες η δομή των οποίων είναι κατατεθειμένη στην Protein Data Bank.



Βάσεις Δεδομένων BLAST 2/2

● Νουκλεοτιδικές ΒΔ:

– nr:

- περιεκτική συλλογή αλληλουχιών χωρίς πλεονασμούς.

– month:

- nr με τις αλληλουχίες των τελευταίων 30 ημερών.

– refseq_mrna:

- αλληλουχίες mRNA από το NCBI Reference Sequence Project.

– refseq_genomic:

- γενωμικές αλληλουχίες από το NCBI Reference Sequence Project.



Παραλλαγές του BLAST 1/12

- Πρωτεϊνική ακολουθία:
 - **blastp**:
 - ακολουθία επερώτησης: πρωτεϊνική.
 - ΒΔ: πρωτεϊνική.
 - αναγνώριση κοινών περιοχών μεταξύ πρωτεϊνών.
 - εύρεση σχετιζόμενων ακολουθιών για φυλογενετική ανάλυση.
 - πρόβλεψη λειτουργίας.



Παραλλαγές του BLAST 2/12

- Πρωτεϊνική ακολουθία:

- **tblastn**:

- ακολουθία επερώτησης: πρωτεϊνική.
- ΒΔ: νουκλεοτιδική (μετάφραση στα 6 πλαίσια ανάγνωσης).
- εύρεση μη σχολιασμένων κωδικοποιουσών περιοχών σε Β.Δ.
- χαρτογράφηση πρωτεϊνών σε γενωμικό DNA.



Παραλλαγές του BLAST 3/12

- Νουκλεοτιδική ακολουθία:

- **blastn**:

- ακολουθία επερώτησης: νουκλεοτιδική.
- ΒΔ: νουκλεοτιδική.
- για ακολουθίες με μεγάλη ομοιότητα.
- χαρτογράφηση ολιγονουκλεοτιδίων, cDNAs και προϊόντων PCR σε ένα γονιδίωμα.
- σχολιασμός γενωμικού DNA.

- **MegaBlast**:

- 10× ταχύτερο από blastn.
- στοίχιση mRNA με γενωμικό DNA.



Παραλλαγές του BLAST 4/12

- Νουκλεοτιδική ακολουθία:

- **blastx**:

- ακολουθία επερώτησης: νουκλεοτιδική (μετάφραση στα 6 πλαίσια ανάγνωσης).
 - ΒΔ: πρωτεϊνική.
 - εύρεση γονιδίων που κωδικοποιούν πρωτεΐνες σε γενωμικό DNA.
 - χρήσιμο όταν δεν είναι γνωστό το πλαίσιο ανάγνωσης ή υπάρχουν σφάλματα στο πλαίσιο ανάγνωσης.



Παραλλαγές του BLAST 5/12

- Νουκλεοτιδική ακολουθία:

- **tblastx**:

- ακολουθία επερώτησης: νουκλεοτιδική (μετάφραση στα 6 πλαίσια ανάγνωσης).
- ΒΔ: νουκλεοτιδική (μετάφραση στα 6 πλαίσια ανάγνωσης).
- αναζήτηση γονιδίων που δεν έχουν αναγνωριστεί με τις συμβατικές μεθόδους.
- υψηλές υπολογιστικές απαιτήσεις (6×6 blastp).



Παραλλαγές του BLAST 6/12

- **PSI-BLAST** (Position Specific Iterated BLAST):
 - **PSSM** (Position Specific Scoring Matrix).
 - Πολλαπλή στοίχιση ακολουθιών και υπολογισμός ενός score που σχετίζεται με τη συχνότητα εύρεσης κάθε καταλοίπου σε μία θέση της στοίχισης.

Sequence alignment

urt-71	A	T	T	T	A	G	T	A	T	C	A	A	A	A	A	T	A	A	C	A	A	T	T	C
glnA-71	G	T	T	C	T	G	T	A	A	C	A	A	A	G	A	C	T	A	C	A	A	A	A	C
nirA-71	A	T	T	T	T	G	T	A	G	C	T	A	C	T	T	A	T	A	C	T	A	T	T	T
ntcB-71	A	A	G	C	T	G	T	A	A	C	A	A	A	A	T	C	T	A	C	C	A	A	A	T
devBCA-71	C	A	T	T	T	G	T	A	C	A	G	T	C	T	G	T	T	A	C	C	T	T	T	A

Table of occurrences

A	3	2	0	0	1	0	0	5	2	1	3	4	3	2	2	1	1	5	0	2	4	2	2	1
C	1	0	0	2	0	0	0	0	1	4	0	0	2	0	0	2	0	0	5	2	0	0	0	2
G	1	0	1	0	0	5	0	0	1	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0
T	0	3	4	3	4	0	5	0	1	0	1	1	0	2	2	2	4	0	0	1	1	3	3	2

Position-specific scoring matrix: Log-odds form (B = 0.1)

A	0.2	0.4	2.2	2.2	0.7	2.2	2.2	0.0	0.4	0.7	0.2	0.1	0.2	0.4	0.4	0.7	0.7	0.0	2.2	0.4	0.1	0.4	0.4	0.7
C	0.7	2.5	2.5	0.4	2.5	2.5	2.5	2.5	0.7	0.1	2.5	2.5	0.4	2.5	2.5	0.4	2.5	2.5	0.0	0.4	2.5	2.5	2.5	0.4
G	0.7	2.5	0.7	2.5	2.5	0.0	2.5	2.5	0.7	2.5	0.7	2.5	2.5	0.7	0.7	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5
T	2.2	0.2	0.1	0.2	0.1	2.2	0.0	2.2	0.7	2.2	0.7	0.7	2.2	0.4	0.4	0.4	0.1	2.2	2.2	0.7	0.7	0.2	0.2	0.4



Παραλλαγές του BLAST 7/12

● **PSI-BLAST:**

1. Διεξαγωγή μιας **τυπικής αναζήτησης BLAST** με την ακολουθία επερώτησης και ένα πίνακα αντικατάστασης (π.χ. BLOSUM62).
2. Αυτόματη **δημιουργία ενός PSSM** από την πολλαπλή στοίχιση των καλύτερων hits της αρχικής αναζήτησης.
3. Το PSSM **αντικαθιστά** τον αρχικό πίνακα αντικατάστασης (π.χ. BLOSUM62) για την εκτέλεση μιας δεύτερης αναζήτησης BLAST.



Παραλλαγές του BLAST 8/12

● **PSI-BLAST:**

4. **Επανάληψη** των βημάτων 2 και 3 και δημιουργία νέων PSSM.
5. **Σύγκλιση του PSI-BLAST** όταν στην τελευταία επανάληψη δεν βρίσκονται νέες ακολουθίες.



Παραλλαγές του BLAST 9/12

● **PSI-BLAST:**

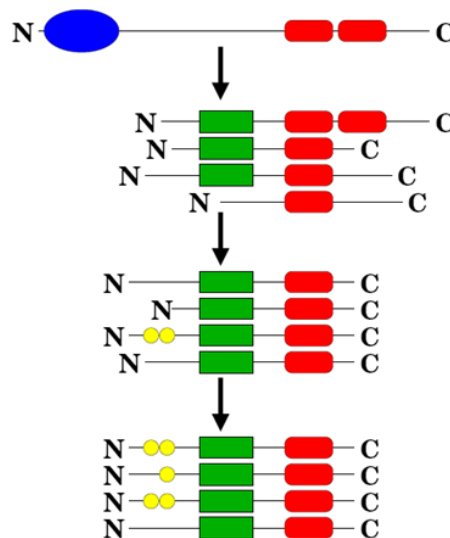
- PSI-BLAST επιτρέπει την εύρεση **περισσότερο απομακρυσμένων ομόλογων ακολουθιών** σε σχέση με το τυπικό BLAST.
- χρησιμοποιεί δύο τιμές κατωφλίου:
 - **threshold E-value.**
 - για την αρχική αναζήτηση BLAST (τυπικά ίσο με 10).
 - **inclusion E-value.**
 - για τη συμμετοχή των ακολουθιών στη δημιουργία του PSSM (τυπικά ίσο με 0.001).



Παραλλαγές του BLAST 10/12

● PSI-BLAST:

- Αν χρησιμοποιηθούν εσφαλμένες ακολουθίες για τη δημιουργία του PSSM, το σφάλμα θα ανατροφοδοτείται σε κάθε κύκλο.
- Το E-value δεν αντικατοπτρίζει την σημαντικότητα της στοίχισης με την αρχική ακολουθία.





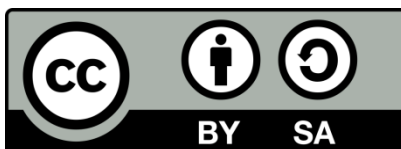
Βιβλιογραφία

- David Mount, "Bioinformatics: Sequence and Genome Analysis", Cold Spring Harbor Laboratory Press; 2nd edition (March 12, 2013).
- Jonathan Pevsner, "Bioinformatics and Functional Genomics", Wiley-Blackwell; 2nd edition (May 4, 2009).
- Andreas D. Baxevanis, B. F. Francis Ouellette, "Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins", Wiley-Interscience; 3rd edition (October 29, 2004).



Άδειες Χρήσης

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





Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα Γεωπονικού Πανεπιστημίου Αθηνών**» έχει χρηματοδοτήσει μόνο την 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 ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ

Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ΕΣΠΑ
2007-2013
πρόγραμμα για την ανάπτυξη
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ



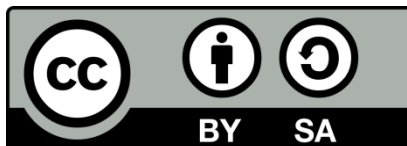
Σημείωμα Αναφοράς

Copyright Γεωπονικό Πανεπιστήμιο Αθηνών 2015. Τμήμα Βιοτεχνολογίας, Θηραίου Τριάς. «Βιοπληροφορική». Έκδοση: 1.0. Αθήνα 2015. Διαθέσιμο από τη δικτυακή διεύθυνση:
<https://mediasrv.aua.gr/eclass/courses/OCDB100/>



Σημείωμα Αδειοδότησης

Το παρόν υλικό διατίθεται με τους όρους της άδειας χρήσης Creative Commons Αναφορά, Παρόμοια Διανομή 4.0 [1] ή μεταγενέστερη, Διεθνής Έκδοση. Εξαιρούνται τα αυτοτελή έργα τρίτων, π.χ. φωτογραφίες, διαγράμματα κ.λ.π., τα οποία εμπεριέχονται σε αυτό και τα οποία αναφέρονται μαζί με τους όρους χρήσης τους στο «Σημείωμα Χρήσης Έργων Τρίτων».



Η άδεια αυτή ανήκει στις άδειες που ακολουθούν τις προδιαγραφές του Ορισμού Ανοικτής Γνώσης [2], είναι ανοικτό πολιτιστικό έργο [3] και για το λόγο αυτό αποτελεί ανοικτό περιεχόμενο [4].

[1] <http://creativecommons.org/licenses/by-sa/4.0/>

[2] <http://opendefinition.org/okd/ellinika/>

[3] <http://freedomdefined.org/Definition/EI>

[4] <http://opendefinition.org/buttons/>



Διατήρηση Σημειωμάτων

Οποιαδήποτε αναπαραγωγή ή διασκευή του υλικού θα πρέπει να συμπεριλαμβάνει:

- το Σημείωμα Αναφοράς
 - το Σημείωμα Αδειοδότησης
 - τη δήλωση Διατήρησης Σημειωμάτων
 - το Σημείωμα Χρήσης Έργων Τρίτων (εφόσον υπάρχει)
- μαζί με τους συνοδευόμενους υπερσυνδέσμους.