

BLAST

Πρόβλημα

— Άγνωστη αλληλουχία

Σύνολο γνωστών
αλληλουχιών

⋮ ⋮ ⋮

Η χρήση ενός υπολογιστή κι ενός αλγόριθμου είναι απαραίτητη για την ανακάλυψη της σχέσης μιας αλληλουχίας με τις γνωστές υπάρχουσες

Τι είναι το BLAST

- **Basic Local Alignment Search Tool**
(Altschul et al., J Mol Biol 215:403-410; 1990)
- Πρόγραμμα που χρησιμοποιεί μία σειρά αλγόριθμων
- Μπορεί να είναι εγκατεστημένο **τοπικά** ή σε **διακομιστή (server)**
- Γρήγορος αλγόριθμος και αρκετά ακριβής

Ορισμένες Χρήσεις

- Ταυτοποίηση ορθόλογων και παράλογων
- Ανακάλυψη νέων γονιδίων, πρωτεϊνων κλπ...
- Ανακάλυψη νέων μορφών γνωστών γονιδίων, πρωτεϊνων κλπ...
- Μελέτη της πρωτεϊνικής δομής και λειτουργείας

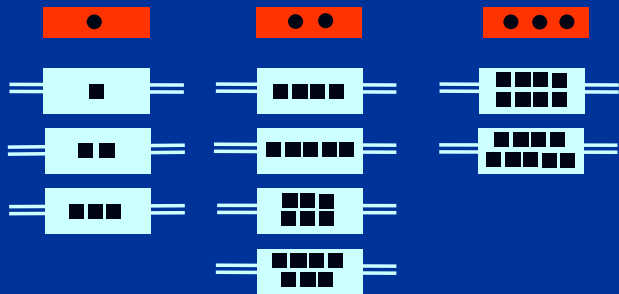
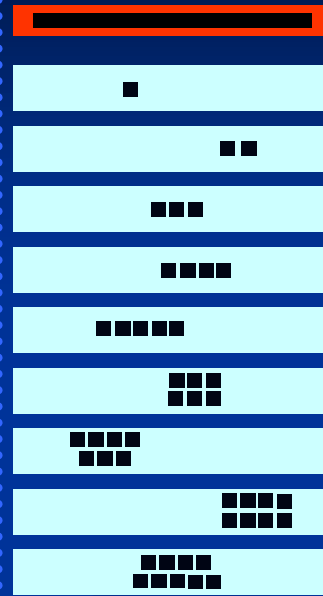
Τρόπος λειτουργίας

Άγνωστη

αλληλουχία

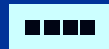


Γνωστές
αλληλουχίες



100%

Ποσοστό ομοιότητας



Τοπικό Blast vs Δικτυακό Blast

Τοπικό

- + Κατά βούληση χρήση
- + Γρήγορο
- + Επιτρέπει τη δημιουργία τοπικής βάσης δεδομένων
- Συντήρηση

Δικτυακό

- + Δε χρειάζεται συντήρηση
- + Πρόσβαση από παντού
- Αργό
- Εξαρτάται από το αν ο server είναι προσβάσιμος

Πίνακες PAM και BLOSUM

- Οι πίνακες αντικατάστασης μας δίνουν ένα σκορ για την ευθυγράμμιση ενός καταλοίπου α με ένα κατάλοιπο β , στη σύγκριση δύο αλληλουχιών.
- Εξαρτάται από την πιθανότητα αντικατάστασης του καταλοίπου α από το κατάλοιπο β .
- Οι πίνακες αντικατάστασης που χρησιμοποιούνται σήμερα είναι οι PAM και BLOSSUM.

BLOSUM 62 Amino Acid Similarity Matrix

Ala	4																				
Arg	-1	5																			
Asn	-2	0	6																		
Asp	-2	-2	1	6																	
Cys	0	-3	-3	-3	9																
Gln	-1	1	0	0	-3	5															
Glu	-1	0	0	2	-4	2	5														
Gly	0	-2	0	-1	-3	-2	-2	6													
His	-2	0	1	-1	-3	0	0	-2	8												
Ile	-1	-3	-3	-3	-1	-3	-3	-4	-3	4											
Leu	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4										
Lys	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5									
Met	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5								
Phe	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6							
Pro	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7						
Ser	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4					
Thr	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5				
Trp	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11			
Tyr	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7		
Val	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4	
	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	

Κενά (gaps)

- Για την εισαγωγή ενός κενού σε μια αλληλουχία το BLAST δίνει μία ποινή a και για κάθε κατάλοιπο που εισέρχεται μία ποινή b

- Ένα κενό x καταλοίπων δέχεται ποινή:

$$a + \beta \cdot x$$

- Το κενό ορίζεται εμπειρικά

Βαθμολόγηση ομολογίας (ομοιότητας) αλληλουχιών

Λαμβάνονται υπόψη:

1. τα κοινά αμινοξέα
2. οι πίνακες αντικατάστασης
3. τα κενά (gaps)
4. το μήκος της ακολουθίας...

Απλοποιημένο παράδειγμα: Φάση 1

Σύνθεση λιστών ζευγαριών λέξεων με **W** γράμματα πάνω από όριο **T** ($W=3$, $T=11$)

Παράδειγμα:

...F**SGTW**YA...

Λίστα λέξεων:

FSG SGT **GTW** TWY WYA

Απλοποιημένο παράδειγμα: Φάση 1

Επιτυχίες
> όριο

(T=11)

Επιτυχίες
< όριο

GTW	6, 5, 11	22
ASW	6, 1, 11	18
ATW	0, 5, 11	16
NTW	0, 5, 11	16
GTY	6, 5, 2	13
GNW		10
GAW		9

Απλοποιημένο παράδειγμα: Φάση 2

- Αναζήτηση στη βάση δεδομένων για εγγραφές που ταιριάζουν
- Με μικρό μέγεθος W η αναζήτηση γίνεται γρήγορα κι εύκολα

Απλοποιημένο παράδειγμα: Φάση 3

1. Επέκταση των επιτυχών ευθυγραμμίσεων προς τις δύο κατευθύνσεις
2. Συνεχής υπολογισμός του σκορ
3. Τερματισμός της επέκτασης όταν το σκορ πέσει κάτω ενός ορίου

	σκορ
KENFDKARFS GTW YAMAKKDPEG	50
MKGLDIQKVA GTW YSLAMAASD.	44

επέκταση

επέκταση

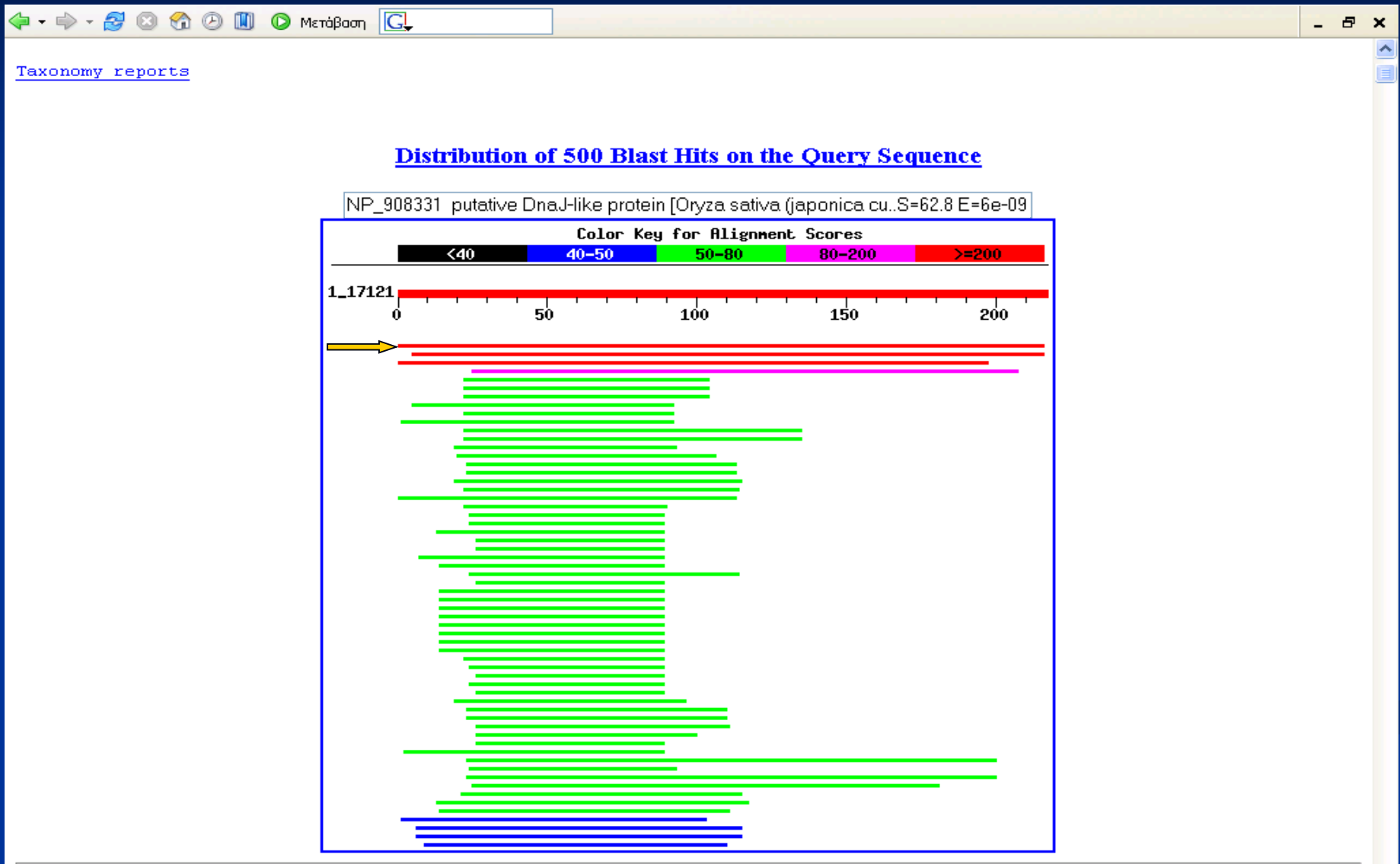


Απλοποιημένο παράδειγμα: Φάση 3

Στην αρχική εφαρμογή της BLAST (1990) κάθε μία επιτυχία επεκτείνεται προς τις 2 κατευθύνσεις

Ο αλγόριθμος βελτιώθηκε το 1997 έτσι ώστε δύο γειτονικές επιτυχίες είναι απαραίτητες. Με αυτό τον τρόπο μόνο $1/7$ των επεκτάσεων συμβαίνουν

Αποτελέσματα της Blast (2)



[Related Structures](#)

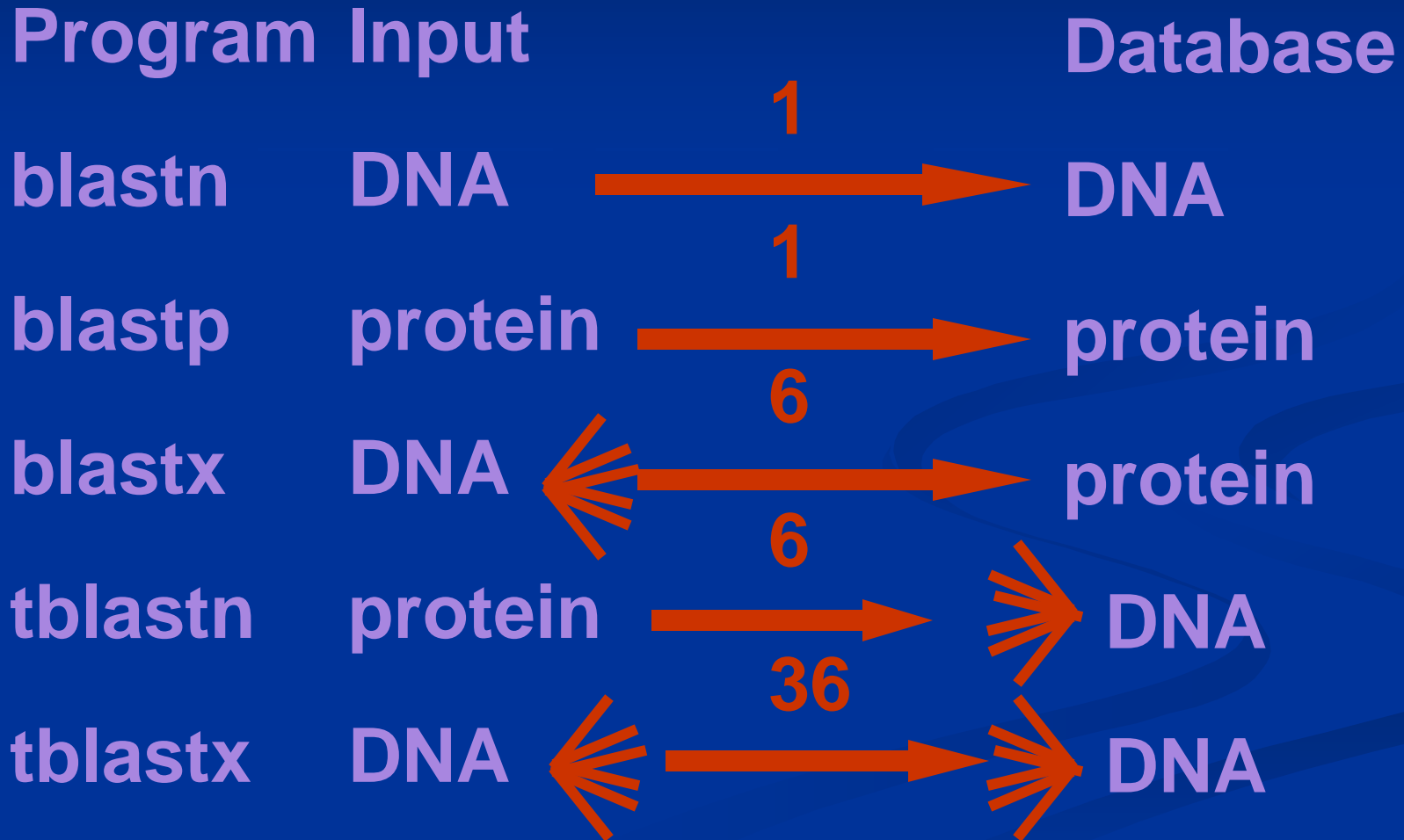
Σημαντικό

- Κοιτάμε τα scores
- Ελέγχουμε το μήκος των ακολουθιών
- Κοιτάμε τις εγγραφές σε βάσεις δεδομένων
- Παρατηρούμε τη συστοίχιση

Τύποι Blast (1)

- `blastp`: συγκρίνει μια πρωτεϊνική ακολουθία με πρωτεϊνικές βάσεις δεδομένων
- `blastn`: συγκρίνει μια νουκλεοτιδική ακολουθία με νουκλεοτιδικές βάσεις δεδομένων
- `blastx`: συγκρίνει μια νουκλεοτιδική ακολουθία μεταφρασμένη στα 6 πιθανά αναγνωστικά πλαίσια με πρωτεϊνικές βάσεις δεδομένων
- `tblastx`: συγκρίνει τις μεταφράσεις των 6 πιθανών αναγνωστικών πλαισίων μιας νουκλεοτιδικής ακολουθίας με τις μεταφράσεις των 6 πιθανών αναγνωστικών πλαισίων μιας νουκλεοτιδικής βάσης δεδομένων
- `tblastn`: συγκρίνει μια πρωτεϊνική ακολουθία με νουκλεοτιδικές βάσεις δεδομένων (οι οποίες μεταφράζονται στα 6 πιθανά αναγνωστικά πλαίσια)

Τύποι Blast (2)



Χρήσεις

- Θέλω να δω πιθανά λάθη στην ακολουθία που έχω αλληλουχήσει
Χρησιμοποιώ **blastx**
- Θέλω να δω ποιες πρωτεΐνες κωδικοποιεί η DNA αλληλουχία μου
Χρησιμοποιώ **blastx**
- Θέλω να ανακαλύψω καινούργιες πρωτεΐνες
Χρησιμοποιώ **tblastx**
- Θέλω να βρω πληροφορίες για τη λειτουργία της πρωτεΐνης που μελετώ
Χρησιμοποιώ **blastp**
- Με ενδιαφέρουν οι 5 ' αμετάφραστες περιοχές
Χρησιμοποιώ **blastn**
- Θέλω να βρω νέα γονίδια που κωδικοποιούν πρωτεΐνες σαν και αυτές που μελετάω
Χρησιμοποιώ **tblastn**