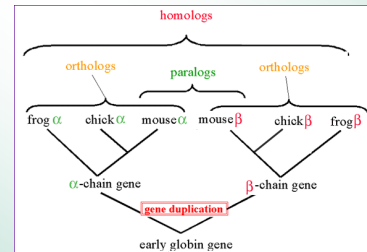




## ΣΤΟΙΧΙΣΗ ΑΚΟΛΟΥΘΙΩΝ ΑΝΑ ΖΕΥΓΗ

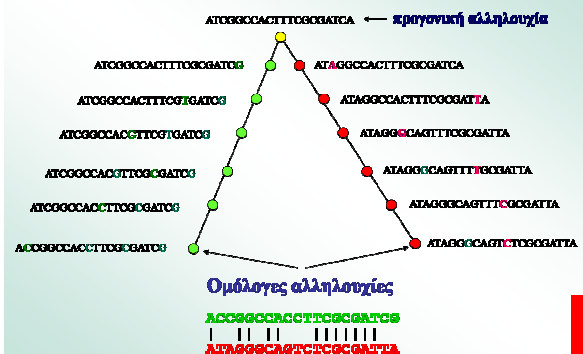
## Ομολογία



## Ομολογία

- **Ομολογία**
  - κοινή εξελικτική καταγωγή
- **Ορθόλογα γονίδια**
  - ειδογένεση
  - συνήθως, ίδια βιολογική λειτουργία
- **Παράλογα γονίδια**
  - γονιδιακός διπλασιασμός
  - εξελικτικό "εργαλείο" ανάπτυξης διαφοροποιημένων λειτουργιών

## Ομολογία



## Ομολογία / Ομοιότητα

- Ομολογία (homology)
  - κοινή εξελικτική καταγωγή
- **Ομοιότητα (similarity)**
  - παρατηρήσιμη ποσότητα που μπορεί να εκφραστεί ως το ποσοστό των ταυτώσεων καταλοίπων μεταξύ δύο ακολουθιών ή με τη χρήση κάποιου άλλου κατάλληλου μέτρου
- Η **υψηλή ομοιότητα** σε επίπεδο ακολουθίας αποτελεί **ένδειξη** κοινής εξελικτικής καταγωγής - **ομολογίας**.

## Ομοιότητα

- Η μελέτη των ομοιοτήτων σε επίπεδο ακολουθίας μπορεί να δώσει πληροφορίες:
  - για την **εξελικτική σχέση** των ακολουθιών των βιομακρομορίων που παρατηρούνται στους σύγχρονους οργανισμούς
  - για τις περιοχές των ακολουθιών οι οποίες είναι σημαντικές για τη **δομή** και τη **λειτουργία** τους

## Μέθοδοι σύγκρισης ακολουθιών

- Τεχνικές που βασίζονται στη **στοίχιση** των ακολουθιών
  - Η σύγκριση δύο ακολουθιών μπορεί να γίνει με τη στοίχισή τους και τον υπολογισμό κάποιου μέτρου ομοιότητας (ή διαφοράς) με βάση τη δεδομένη στοίχιση.

```
>P01922|HBA_HUMAN  GSAQVKGHGKKNVADALTNL-----VAHVDDMPNALSALSIDLHAKKL
                    + +++ H KV + A V V L L +H K
>P02240|LGB2_LUPLU  NNPELQAHAGKVFKLIVYEAAIQLVTVGVNVTDAITLKNLGSVHNSKG
```

- Τεχνικές που δεν βασίζονται στη στοίχιση των ακολουθιών

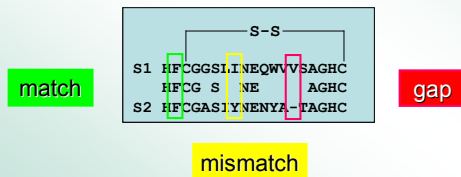
## Στοίχιση Ακολουθιών

- απόκλιση από κοινό προγονικό μόριο μέσω:
  - **αντικαταστάσεων** (substitutions), π.χ. G στη θέση του A
  - indels
    - **ενθέσεων** (insertions), π.χ. εισαγωγή του A μεταξύ των M και V
    - **διαγραφών** (deletions), π.χ. απαλοιφή του L μεταξύ των I και Y

```
AAVILYMV → προγονικό μόριο
AGVILYMV
AGVILYMAV
AGVIYMAV
```

## Στοίχιση Ακολουθιών

- Σε μία "ιδανική" στοίχιση:
- Η αντιστοίχιση ανόμοιων καταλοίπων (**mismatch**) αντιπροσωπεύει αντικαταστάσεις.
- Η εισαγωγή κενών (**gap**) οφείλεται σε ενθέσεις ή διαγραφές.



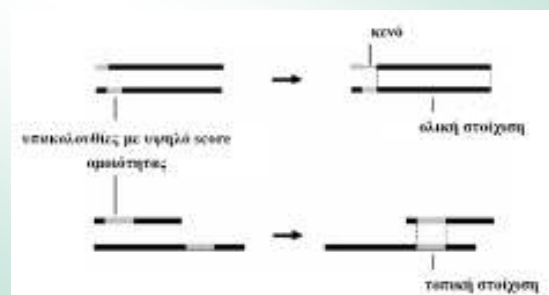
## Στοίχιση Ακολουθιών

- Δεδομένα
  - δύο ακολουθίες
  - ένα **σύστημα βαθμολόγησης** της στοίχισης:
    - όμοιων καταλοίπων (match score)
    - ανόμοιων καταλοίπων (mismatch score)
    - ενός καταλοίπου με κενό (ποινή για τα κενά, gap penalty)
- Ζητούμενο
  - εύρεση της **βέλτιστης στοίχισης** των δύο ακολουθιών που
    - διατηρεί τη σειρά των χαρακτήρων
    - εισάγει τα απαραίτητα κενά
    - **μεγιστοποιεί το συνολικό score ομοιότητας**

## Στοίχιση Ακολουθιών

- επιλογές
  - **είδος στοίχισης**
  - **σύστημα βαθμολόγησης** (scoring system)
  - **αλγόριθμος** για τον προσδιορισμό της στοίχισης
  - στατιστικές μέθοδοι για την αξιολόγηση της **σημαντικότητας της βαθμολογίας** (score) που προέκυψε κατά τη στοίχιση
    - π.χ. είναι η ομοιότητα αρκετά υψηλή ώστε να αποτελεί ένδειξη ομολογίας???

## Είδος στοίχισης

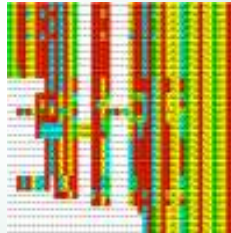




## Σύστημα βαθμολόγησης

- Πίνακες αντικατάστασης (Substitution matrices)

- αποδίδουν **διαφορετικά scores ομοιότητας** για την αντιστοίχιση διαφορετικών αμινοξέων
- εξάγονται από πολλαπλές στοιχίσεις
- $score \sim \log_2(\text{observed}/\text{expected})$
- observed
  - παρατηρηθείσα συχνότητα αντικατάστασης
- expected
  - αναμενόμενη συχνότητα αντικατάστασης



## Σύστημα βαθμολόγησης

- Πίνακες αντικατάστασης (Substitution matrices)

- **συμμετρικοί**
  - δεν είναι γνωστό πιο αμινοξύ προϋπήρχε
- $score > 0$ 
  - ένδειξη ότι η στοιχίση των καταλοίπων αποτελεί πραγματικό εξελικτικό γεγονός
- $score < 0$ 
  - ένδειξη ότι η στοιχίση των καταλοίπων είναι λιγότερο πιθανό να βρεθεί στην ευθυγράμμιση ομόλογων ακολουθιών

## Πίνακες αντικατάστασης

- PAM (Point Accepted Mutation) (Dayhoff)

- 1PAM: μονάδα εξελικτικής απόκλισης που αντιστοιχεί σε αλλαγή 1% αμινοξέων
- 250PAM: απόκλιση ~ 80%
- Ένα αμινοξύ μπορεί να αλλάξει περισσότερες από μία φορές ή / και να επιστρέψει στον αρχικό του τύπο.
  - MATCG
  - MAGCG
  - MATGG
- Κάθε αμινοξύ αποκλίνει με διαφορετικό ρυθμό.

## Πίνακες αντικατάστασης

- PAM (Point Accepted Mutation) (Dayhoff)

- πίνακας **PAM1**
  - υπολογίσθηκε βάσει **ολικών στοιχίσεων** από 71 οικογένειες πρωτεϊνών με **>85% ομοιότητα**
- πίνακας **PAM250**
  - υπολογίζεται ως η **250-ιστή δύναμη** του πίνακα PAM1
- Με τον ίδιο τρόπο κατασκευάζονται πίνακες PAM με διαφορετικούς δείκτες.
- Οι πίνακες PAM με **μικρότερους δείκτες** χρησιμοποιούνται για τη στοιχίση **κοντινότερων εξελικτικά** ακολουθιών.

## Πίνακες αντικατάστασης

- PAM (Point Accepted Mutation) (Dayhoff)



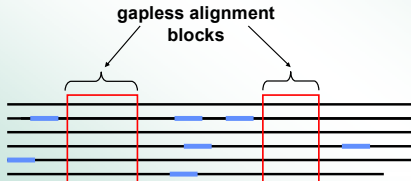
## Πίνακες αντικατάστασης

- PAM (Point Accepted Mutation) (Dayhoff)

- Θεωρεί ότι η πιθανότητα παρατήρησης μιας μετάλλαξης είναι ανεξάρτητη από
  - την περιοχή της πρωτεΐνης
  - προηγούμενες μεταλλάξεις στην ίδια θέση
- Στηρίζεται μόνο στις ιδιότητες σφαιρικών υδατοδιαλυτών πρωτεϊνών.

## Πίνακες αντικατάστασης

- BLOSUM (Blocks Amino Acid Substitution Matrix) (Henikoff)
- BLOCKS database
  - τοπική πολλαπλή στοίχιση χωρίς κενά, εξελικτικά απομακρυσμένων πρωτεϊνών

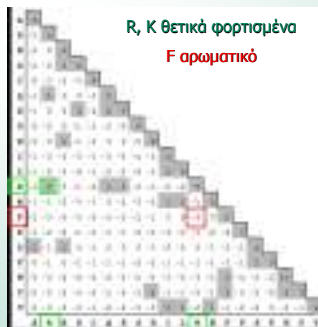


## Πίνακες αντικατάστασης

- BLOSUM (Blocks Amino Acid Substitution Matrix) (Henikoff)
- δείκτες πινάκων BLOSUM
  - ποσοστό ομοιότητας των ακολουθιών που χρησιμοποιήθηκαν για την κατασκευή του πίνακα
  - BLOSUM62: ακολουθίες με ομοιότητα >62% έχουν ομαδοποιηθεί
- Οι πίνακες BLOSUM με μικρότερους δείκτες χρησιμοποιούνται για τη σύγκριση περισσότερο απομακρυσμένων εξελικτικά ακολουθιών.
- Στηρίζονται σε πραγματικές στοιχίσεις και δεν προκύπτουν αναγωγικά όπως οι πίνακες PAM.

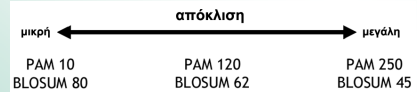
## Πίνακες αντικατάστασης

- BLOSUM (Blocks Amino Acid Substitution Matrix) (Henikoff)



## Πίνακες αντικατάστασης

- PAM
  - designed to track the evolutionary origins of proteins
  - στοίχιση συγγενών ακολουθιών
- BLOSUM
  - designed to find the conserved domains of proteins
  - στοίχιση απομακρυσμένων ακολουθιών
  - εύρεση τοπικών ομοιοτήτων



## Σύστημα βαθμολόγησης

- Μοντέλα νουκλεοτιδικών υποκαταστάσεων

	A	T	C	G	A	T	C	G
(A) Jukes-Cantor model	-	$\alpha$	$\alpha$	$\alpha$	-	$\beta\theta_T$	$\beta\theta_C$	$\alpha\theta_G$
T	$\alpha$	-	$\alpha$	$\alpha$	$\beta\theta_A$	-	$\alpha\theta_C$	$\beta\theta_G$
C	$\alpha$	$\alpha$	-	$\alpha$	$\alpha\theta_A$	$\alpha\theta_T$	-	$\beta\theta_G$
G	$\alpha$	$\alpha$	$\alpha$	-	$\alpha\theta_A$	$\beta\theta_T$	$\beta\theta_C$	-
(B) Kimura model	-	$\beta$	$\beta$	$\alpha$	-	$\beta\theta_T$	$\beta\theta_C$	$\alpha_1\theta_G$
T	$\beta$	-	$\alpha$	$\beta$	$\beta\theta_A$	-	$\alpha_2\theta_C$	$\beta\theta_G$
C	$\beta$	$\alpha$	-	$\beta$	$\beta\theta_A$	$\alpha_2\theta_T$	-	$\beta\theta_G$
G	$\alpha$	$\beta$	$\beta$	-	$\alpha_1\theta_A$	$\beta\theta_T$	$\beta\theta_C$	-
(C) Equal-input model	-	$\alpha\theta_T$	$\alpha\theta_C$	$\alpha\theta_G$	-	$\alpha\theta_T$	$b\theta_C$	$c\theta_G$
T	$\alpha\theta_A$	-	$\alpha\theta_C$	$\alpha\theta_G$	$\alpha\theta_A$	-	$d\theta_C$	$e\theta_G$
C	$\alpha\theta_A$	$\alpha\theta_T$	-	$\alpha\theta_G$	$b\theta_A$	$d\theta_T$	-	$f\theta_G$
G	$\alpha\theta_A$	$\alpha\theta_T$	$\alpha\theta_C$	-	$c\theta_A$	$e\theta_T$	$f\theta_C$	-
(D) Tamura model	-	$\beta\theta_1$	$\beta\theta_1$	$\alpha\theta_1$	-	$a_{12}$	$a_{13}$	$a_{14}$
T	$\beta\theta_2$	-	$\alpha\theta_1$	$\beta\theta_1$	$a_{21}$	-	$a_{23}$	$a_{24}$
C	$\beta\theta_2$	$\alpha\theta_2$	-	$\beta\theta_1$	$a_{31}$	$a_{32}$	-	$a_{34}$
G	$\alpha\theta_2$	$\beta\theta_2$	$\beta\theta_1$	-	$a_{41}$	$a_{42}$	$a_{43}$	-

Note: An element  $(\theta_{ij})$  of the above substitution matrices stands for the substitution rate from the nucleotide in the  $i$ th row to the nucleotide in the  $j$ th column.  $\theta_A, \theta_B, \theta_C,$  and  $\theta_G$  are the nucleotide frequencies.  $\theta_1 = \theta_C + \theta_G, \theta_2 = \theta_A + \theta_T, \theta_3 = \theta_A + \theta_C, \theta_4 = \theta_A + \theta_G, \theta_5 = \theta_T + \theta_C,$

## Σύστημα βαθμολόγησης

- Μοντέλα βαθμολόγησης των κενών
  - κενό μήκους k καταλοίπων
  - Linear gap penalty
    - gap penalty a
    - $gp(k) = ak$
  - Affine gap penalty
    - gap-opening penalty b (ποινή για το άνοιγμα κενού)
    - gap-extension penalty a (ποινή για την επέκταση κενού)
    - $gp(k) = b + ak, (|b| > |a|)$
    - ευνοεί την εισαγωγή λίγων μεγάλων κενών σε σχέση με πολλά μικρότερα κενά



## Σύστημα βαθμολόγησης

- Το **ολικό score** της στοίχισης προκύπτει από το άθροισμα των scores των ζευγών καταλοίπων και των ποινών για τα κενά.
- **Εμπειρική επιλογή** τιμών ποινής για την εισαγωγή και την επέκταση κενών
  - <http://www.ebi.ac.uk/help/matrix.html>

Protein Group	Matrix	Open Gap	Extend Gap
0-200	BLOSUM50	-10	-3
05-200	BLOSUM62	-10	-3
08-200	BLOSUM62	-10	-4
1-200	PAM250	-10	-3
05-200	PAM120	-10	-4
10-200	BLOSUM	-10	-3
1-20	BLOSUM	-10	-4
1-30	BLOSUM	-10	-3

ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ

Σελίδα 31

Τ. Θηραίου

## Σύστημα βαθμολόγησης

- Blosum62
- gap-opening penalty =  $b = -10$
- gap-extension penalty =  $a = -1$
- score =  $S(K,K)+S(A,K)+S(H,H)+S(G,G)+S(K,V)+S(K,T)-(|b|+|a|\cdot 1) = 5 + (-1) + 8 + 6 + (-2) + (-1) - (10+1\cdot 2) = 3$

```

K L V A H G K K
K - - K H G V T
    
```

ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ

Σελίδα 32

Τ. Θηραίου

## Τεχνικές Στοίχισης Ακολουθιών

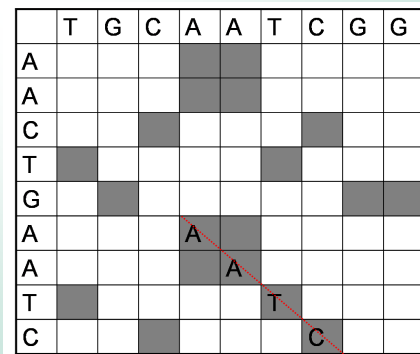
- Διαγράμματα Πινάκων Σημείων (Dot Matrix Plots)
- Αλγόριθμοι Δυναμικού Προγραμματισμού
  - Needleman-Wunsch algorithm (ολική στοίχιση)
  - Smith-Waterman algorithm (τοπική στοίχιση)
- Ευριστικοί Αλγόριθμοι
  - FASTA
  - BLAST

ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ

Σελίδα 33

Τ. Θηραίου

## Dot Plots



ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ

Σελίδα 34

Τ. Θηραίου

## Dot Plots

- κατασκευή δισδιάστατου ορθογώνιου διαγράμματος με τις υπό σύγκριση ακολουθίες στη θέση των αξόνων
- χρωματισμός των κελιών που αντιστοιχούν σε ταυτόσημα κατάλοιπα των δύο ακολουθιών
- περιοχές του πίνακα με χρωματισμένα κελιά **κατά μία διαγώνιο** υποδεικνύουν περιοχές ταύτισης των ακολουθιών
- εφαρμογή τεχνικών φιλτραρίσματος για τη μείωση του θορύβου (μεμονωμένες ή μικρού μήκους στοίχισεις)
- χρήση πινάκων αντικατάστασης για τη βαθμολόγηση ζευγών καταλοίπων
- **υποκειμενική ερμηνεία** των αποτελεσμάτων

ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ

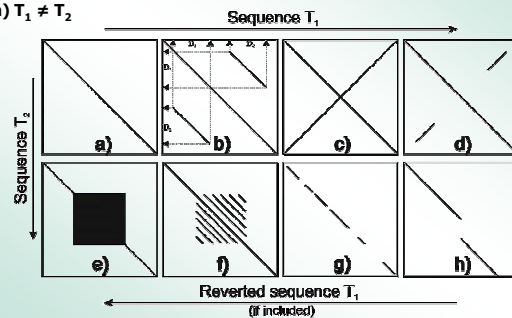
Σελίδα 35

Τ. Θηραίου

## Dot Plots

a-f)  $T_1 = T_2$

g-h)  $T_1 \neq T_2$



ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ

Σελίδα 36

Τ. Θηραίου

## Dot Plots

	A	C	G	T	A	C	G	T
A	X				X			
C		X				X		
G			X				X	
T				X				X
A	X				X			
C		X				X		
G			X				X	
T				X				X

	A	C	G	G	G	G	T
A	X						
C		X					
G			X	X	X	X	
G			X	X	X	X	
G			X	X	X	X	
G			X	X	X	X	
T			X	X	X	X	
T							X

	A	C	G	T	G	C	A
A	X						X
C		X					X
G			X			X	
T				X	X		
T				X	X		
G			X			X	
C		X					X
A	X						X

	E	F	G	H	I	K	L	N	P	Q
E	X									
F		X								
G			X							
N								X		
P									X	
Q										X

## Dot Plots

- **άμεση οπτικοποίηση**
  - επαναλήψεων
  - παλινδρομων ακολουθιών
  - περιοχών χαμηλής πολυπλοκότητας
  - ενθέσεων ή διαγραφών
- μελέτη γονιδιωμάτων
- δομή RNA

## Μέθοδοι Δυναμικού Προγραμματισμού

- Δημιουργία βέλτιστης στοίχισης χρησιμοποιώντας τις βέλτιστες στοίχισεις μικρότερων ακολουθιών.
  - **Ολική Στοίχιση (Needleman-Wunsch)**
  - **Τοπική Στοίχιση (Smith-Waterman)**
- 2 ακολουθίες x και y
- πίνακας **F(i,j)**: score της βέλτιστης στοίχισης μεταξύ του αρχικού τμήματος  $x_{1...i}$  της x μέχρι το κατάλοιπο  $x_i$  και του αρχικού τμήματος  $y_{1...j}$  της y μέχρι το κατάλοιπο  $y_j$
- "γέμισμα" του  $F(i,j)$ : επαναληπτική διαδικασία με  $F(0,0) = 0$

## Ολική Στοίχιση (Needleman-Wunsch)

- Η πρώτη γραμμή και η πρώτη στήλη του πίνακα συμπληρώνεται βάσει της σχέσης:

$$F_{i,0} = -i \times |gap|, F_{0,j} = -j \times |gap|$$

		A	G	T	A
	0	-1xd	-2xd	-3xd	-4xd
A	-1xd				
T	-2xd				
A	-3xd				

για απλότητα  
 $d = |gap|$

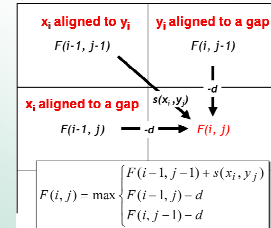
## Ολική Στοίχιση (Needleman-Wunsch)

- Οι τιμές των άλλων κελιών υπολογίζονται **κατά γραμμή**, ξεκινώντας από την επάνω αριστερά γωνία και προχωρώντας συνεχώς **προς τα δεξιά**.
- Όταν φτάνουμε στο τέλος μιας γραμμής ξεκινάμε από το πρώτο κελί της επόμενης.

		A	G	T	A
	0	-1xd	-2xd	-3xd	-4xd
A	-1xd				
T	-2xd				
A	-3xd				

## Ολική Στοίχιση (Needleman-Wunsch)

- Παράλληλα με τον υπολογισμό της τιμής ενός κελιού, φυλάσσεται σ' ένα **πίνακα "ιχνηθέτη"** (trace-back) η διεύθυνση του κελιού βάσει του οποίου υπολογίστηκε η τιμή (από το διαγώνιο κελί, το αριστερό ή το κατακόρυφο;)



## Ολική Στοιχίση (Needleman-Wunsch)

- match = 1, mismatch = -1, gap = -1

F(i,j)		i = 0	1	2	3	4
			A	G	T	A
j = 0		0	-1	-2	-3	-4
1	A	-1	1	0	-1	-2
2	T	-2	0	0	1	0
3	A	-3	-1	-1	0	2

## Ολική Στοιχίση (Needleman-Wunsch)

- Η τιμή στο **κάτω δεξιά κελί** αποτελεί τη **βαθμολογία** της καλύτερης δυνατής στοιχίσης των δύο ακολουθιών με βάση το συγκεκριμένο σύστημα βαθμολόγησης.
- Ακολουθώντας τη **διαδρομή στον πίνακα ιχνηθέτη** από το κάτω δεξιά κελί παράγουμε τη **συγκεκριμένη στοιχίση**.
- Για κάθε θέση:
  - Αν κινηθούμε **διαγώνια**, τότε **στοιχίζουμε τα δύο κατάλοιπα** που αντιστοιχούν σ' εκείνη την θέση.
  - Αν κινηθούμε **οριζόντια ή κάθετα**, βάζουμε **κενό** στην ακολουθία που δείχνει το βέλος.

## Ολική Στοιχίση (Needleman-Wunsch)

F(i,j)		i = 0	1	2	3	4
			A	G	T	A
j = 0		0	-1	-2	-3	-4
1	A	-1	1	0	-1	-2
2	T	-2	0	0	1	0
3	A	-3	-1	-1	0	2

στοιχίση

A G T A

A - T A

score της στοιχίσης

score = 2

από διαγώνιο

κελί = 0 + 2

= 2

από αριστερό

κελί = 1 - 1

= 0

από κατακόρυφο

κελί = 1 - 1

= 0

max(2, 0, 0) = 2

match = 2

mismatch = 0

gap = -1

		T	G	C	A	A	T	C	G	G
A	-1	0	-1	-2	-1	-2	-3	-4	-5	-6
A	-2	-1	0	-1	0	1	0	-1	-2	-3
C	-3	-2	-1	2	1	0	1	2	1	0
T	-4	-1	-2	1	2	1	2	1	2	1
G	-5	-2	1	0	1	2	1	2	3	4
A	-6	-3	0	1	2	3	2	1	2	3
A	-7	-4	-1	0	3	4	3	2	1	2
T	-8	-5	-2	-1	2	3	6	5	4	3
C	-9	-6	-3	0	1	2	5	8	7	6

---TGCAATCGG

AACTG-AAATC--

← Βέλτιστη Στοιχίση

Score  
Βέλτιστης  
Στοιχίσης  
= 6

		T	G	C	A	A	T	C	G	G
A	-1	0	-1	-2	-1	-2	-3	-4	-5	-6
A	-2	-1	0	-1	0	1	0	-1	-2	-3
C	-3	-2	-1	2	1	0	1	2	1	0
T	-4	-1	-2	1	2	1	2	1	2	1
G	-5	-2	1	0	1	2	1	2	3	4
A	-6	-3	0	1	2	3	2	1	2	3
A	-7	-4	-1	0	3	4	3	2	1	2
T	-8	-5	-2	-1	2	3	6	5	4	3
C	-9	-6	-3	0	1	2	5	8	7	6

## Ολική Στοιχίση (Needleman-Wunsch)

- Στοιχίση των ακολουθιών AGTA και ATA με τη χρήση δυναμικού προγραμματισμού, του **πίνακα αντικατάστασης** Blosum62 και linear gap penalty = -1
- Αντί για match και mismatch score, χρησιμοποιούνται τα scores που είναι καταγεγραμμένα στον **πίνακα αντικατάστασης**.



### Τοπική Στοιχίση (Smith-Waterman)

- $F(i, 0) = F(0, j) = 0$
- Στοιχίσεις με αρνητική βαθμολογία δεν παρουσιάζουν ενδιαφέρον.
- Μια βέλτιστη τοπική στοιχίση κατασκευάζεται **Ξεκινώντας** από το **κελί με τη μεγαλύτερη τιμή** (όπου κι αν βρίσκεται) και **τερματίζεται** μόλις συναντήσουμε για πρώτη φορά **μηδενική τιμή**.
- Το **score** της τοπικής στοιχίσης ισούται με τη **μεγαλύτερη τιμή** του πίνακα.

### Ολική Στοιχίση

### Τοπική Στοιχίση

		$X_1$	$X_2$	$X_3$	$X_4$
	0	$-1 \times d$	$-2 \times d$	$-3 \times d$	$-4 \times d$
$Y_1$	$-1 \times d$				
$Y_2$	$-2 \times d$				
$Y_3$	$-3 \times d$				

		$X_1$	$X_2$	$X_3$	$X_4$
	0	0	0	0	0
$Y_1$	0				
$Y_2$	0				
$Y_3$	0				

$$d = |\text{gap}|$$

### Ολική Στοιχίση

$$F(i, j) = \max \begin{cases} F(i-1, j-1) + s(x_i, y_j) \\ F(i-1, j) - |\text{gap}| \\ F(i, j-1) - |\text{gap}| \end{cases}$$

### Τοπική Στοιχίση

$$F(i, j) = \max \begin{cases} F(i-1, j-1) + s(x_i, y_j) \\ F(i-1, j) - |\text{gap}| \\ F(i, j-1) - |\text{gap}| \\ 0 \end{cases}$$

### Ολική Στοιχίση

### Τοπική Στοιχίση

		$X_1$	$X_2$	$X_3$	$X_4$
	0	-1	-2	-3	-4
$Y_1$	-1	3	0	-1	-5
$Y_2$	-2	0	2	6	0
$Y_3$	-3	-5	-2	0	4

		$X_1$	$X_2$	$X_3$	$X_4$
	0	0	0	0	0
$Y_1$	0	1	3	0	1
$Y_2$	0	0	0	7	0
$Y_3$	0	1	0	0	5

από διαγώνιο  
κελί =  $3 + 2$   
= 5  
από αριστερό  
κελί =  $4 - 1$   
= 3  
από κατακόρυφο  
κελί =  $2 - 1$   
= 1  
0

$$\max(5, 3, 1, 0) = 5$$

match = 2  
mismatch = 0  
gap = -1

		T	G	C	A	A	T	C	G	G
		0	0	0	0	0	0	0	0	0
A	0	0	0	0	2	2	1	0	0	0
A	0	0	0	0	2	4	3	2	1	0
C	0	0	0	2	1	3	4	5	4	3
T	0	2	1	1	2	2	5	4	5	4
G	0	1	4	3	2	2	4	5	6	7
A	0	0	3	4	5	4	3	4	5	6
A	0	0	2	3	6	7	6	5	4	5
T	0	2	1	2	5	6	9	8	7	6
C	0	1	2	3	4	5	8	11	10	9

TGCAATC

TG-AATC

← Βέλτιστη Στοιχίση

Score  
Βέλτιστης  
Στοιχίσης  
= 11

		T	G	C	A	A	T	C	G	G
		0	0	0	0	0	0	0	0	0
A	0	0	0	0	2	2	1	0	0	0
A	0	0	0	0	2	4	3	2	1	0
C	0	0	0	2	1	3	4	5	4	3
T	0	2	1	1	2	2	5	4	5	4
G	0	1	4	3	2	2	4	5	6	7
A	0	0	3	4	5	4	3	4	5	6
A	0	0	2	3	6	7	6	5	4	5
T	0	2	1	2	5	6	9	8	7	6
C	0	1	2	3	4	5	8	11	10	9

