



Βιοπληροφορική

Ενότητα 13:

Μοντέλα Πολλαπλής Στοίχισης (1/2), 1.5ΔΩ

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

- παρουσίαση των μοντέλων πολλαπλής στοίχισης.
- κατανόηση των εφαρμογών και των περιορισμών τους.



Λέξεις Κλειδιά

- Λέξεις κλειδιά: Συναινετική ακολουθία, Μοτίβο.
- Key words: Consensus sequence, PROSITE Pattern, Position Specific Scoring Matrix (PSSM), Generalized Profile.

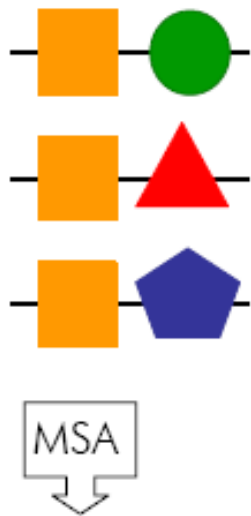


Μοντέλα Πολλαπλής Στοίχισης

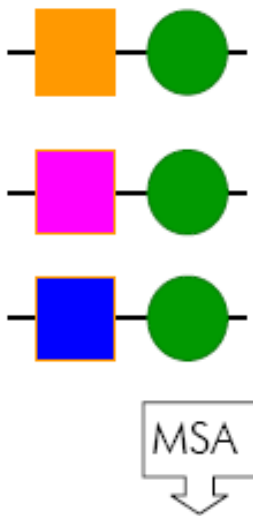
- Consensus sequences
 - Patterns and regular expressions
 - Position Specific Scoring Matrices (PSSMs)
 - Generalized Profiles
 - Hidden Markov Models (HMMs)
-
- Στοίχιση νέων ακολουθιών
 - Αναζήτηση σε βάσεις δεδομένων
 - Σχολιασμός νέων ακολουθιών



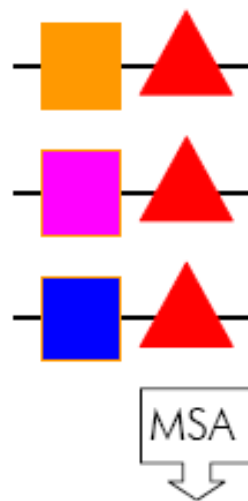
Από την Ακολουθία στη Λειτουργία



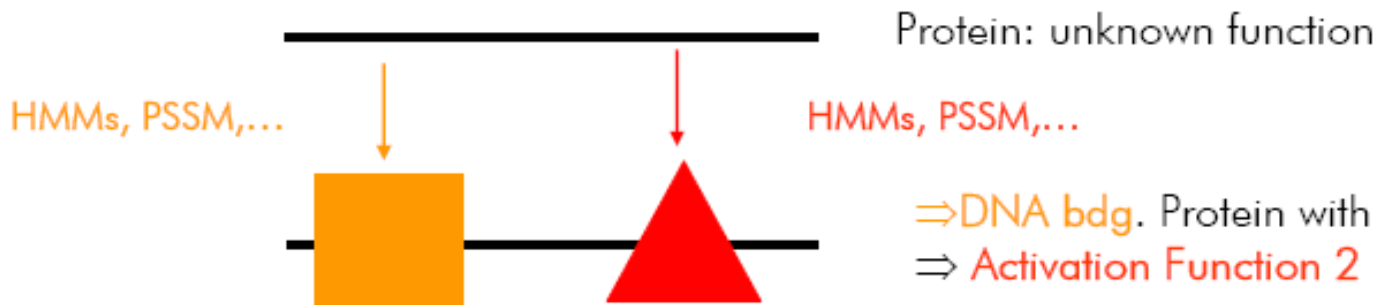
Model (HMM, PSSM,...) for DNA bdg. Function



Model for Activation Function 1



Model for Activation Function 2





Consensus sequences 1/2

- Συναινετική ακολουθία
- Majority wins

```

GHEGVGVKVVKIG
GHEKKGYFEDRG
GHEGYGGRSRGG
GHEFEGPKGCGA
GHELRGTTFMPA
  
```



1	2	3	4	5	6	7	8	9	10	11	12
G	H	E	G	V	G	K	V	V	K	I	G
			K	K		Y	F	E	D	R	A
			F	Y		G	R	S	R	G	
			L	E		P	K	G	C	P	
				R		T	T	F	M		
<hr/>											
Consensus:	G	H	E	.	.	G



Consensus sequences 2/2

- Πλεονεκτήματα
 - Γρήγορη και εύκολη υλοποίηση
- Περιορισμοί
 - Δεν έχουν πληροφορία σχετικά με τη μεταβλητότητα κάθε στήλης
 - Εξαρτώνται πολύ από τις αρχικές ακολουθίες (training set)
 - Διαδικό αποτέλεσμα (ΝΑΙ / ΌΧΙ)
- Χρησιμότητα
 - Αναζήτηση **ιδιαίτερα συντηρημένων "υπογραφών"** (signatures), όπως θέσεις περιορισμού (enzyme restriction sites) στο DNA



Patterns and Regular Expressions 1/9

● Μοτίβο

- Περιγράφει σε μία γραμμή μια ομάδα εναλλακτικών ακολουθιών
- Στην επιστήμη των υπολογιστών, τα μοτίβα ονομάζονται κανονικές εκφράσεις

ADLGAVFALCDRYFQ

SDVGPRSCFCERFYQ

ADLGR TQNRCDRYYQ

ADIGQPHSLCERYFQ

Regular Expression or Pattern
[AS]-D-[IVL]-G-x(4)-{PG}-C-[DE]-R-[FY]2-Q



Patterns and Regular Expressions 2/9

● Σύνταξη Prosite

- "x" οποιοδήποτε κατάλοιπο
- [] : επιτρεπτά κατάλοιπα
 - π.χ. [ALT] = Ala ή Leu ή Thr
- { } : μη επιτρεπτά κατάλοιπα
 - π.χ. {AM} οποιοδήποτε κατάλοιπο εκτός Ala και Met
- () : ποσοτικός τελεστής
 - π.χ. x(3) = x-x-x, x(2,4) = x-x ή x-x-x ή x-x-x-x
- "-" διαχωρίζει γειτονικά στοιχεία
- "<" στο αμινοτελικό άκρο
- ">" στο καρβοξυτελικό άκρο



Patterns and Regular Expressions 3/9

● Σύνταξη Prosite

– $\langle A-x-[ST](2)-x(0,1)-\{V\}$

- Ala στο αμινοτελικό άκτο
- Οποιοδήποτε αμινοξύ
- Ser ή Thr δύο φορές
- Κανένα ή ένα οποιοδήποτε αμινοξύ
- Οποιοδήποτε αμινοξύ εκτός από Val



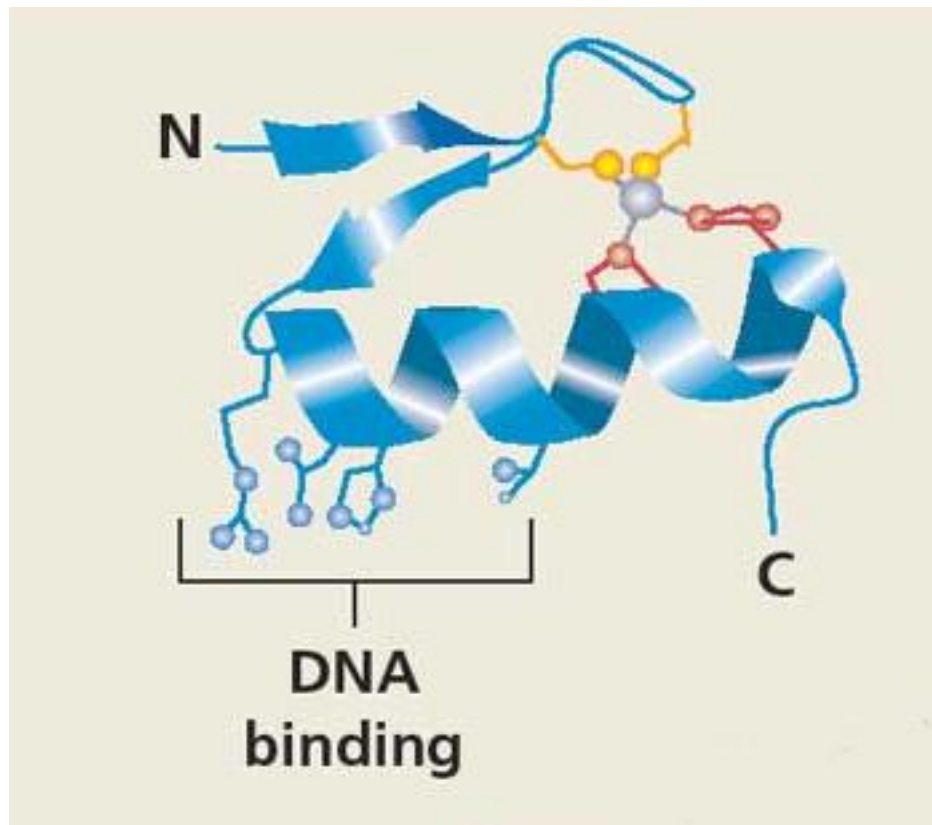
Patterns and Regular Expressions 4/9

- [AC]-x-V-x(4)-{ED}
- DEHSDVLPVLDVCSVKHVAEVFQALIYWIKAMNQQTTLDT
- Σε ποια (-ες) θέση(-εις) της ακολουθίας ταιριάζει;
- DEHSDVLPVLDVCSVKHVAEVFQALIYWIKA
MNQQTTLDT
- [AC]-x-V-x(4)-{ED}



Patterns and Regular Expressions 5/9

- ZINC_FINGER_C2H2_1
 - **C**-x(2,4)-**C**-x(3)-[LIVMFYWC]-x(8)-**H**-x(3,5)-**H**





Patterns and Regular Expressions 6/9

- Τα μοτίβα
 - σχεδιάζονται από επιστήμονες
 - εξάγονται αυτόματα από μη στοιχισμένες ακολουθίες με ειδικά προγράμματα
- Πλεονεκτήματα
 - Η αναγνώριση μοτίβων είναι γρήγορη και υλοποιείται εύκολα.
 - Τα μοτίβα σχεδιάζονται και κατανοούνται εύκολα.
- Χρησιμότητα
 - Αναζήτηση **μικρού μήκους "υπογραφών"** (signatures)



Patterns and Regular Expressions 7/9

- Περιορισμοί
 - Ανεπαρκής μοντελοποίηση κενών
 - Ανεπαρκής πρόβλεψη, τείνουν να αναγνωρίζουν μόνο τις ακολουθίες από τις οποίες κατασκευάστηκαν (training set)
 - Δυαδική απάντηση (ΝΑΙ/ΟΧΙ)



Patterns and Regular Expressions 8/9

ACA T AT**G**

TCA A ATC

ACA C A**G**C

A**G**A - ATC

AC**C** G ATC

[AT]-[CG]-[AC]-x(0,1)-A-[TG]-[GC]

- Είναι το ίδιο πιθανές οι ακολουθίες που προκύπτουν από το μοτίβο;

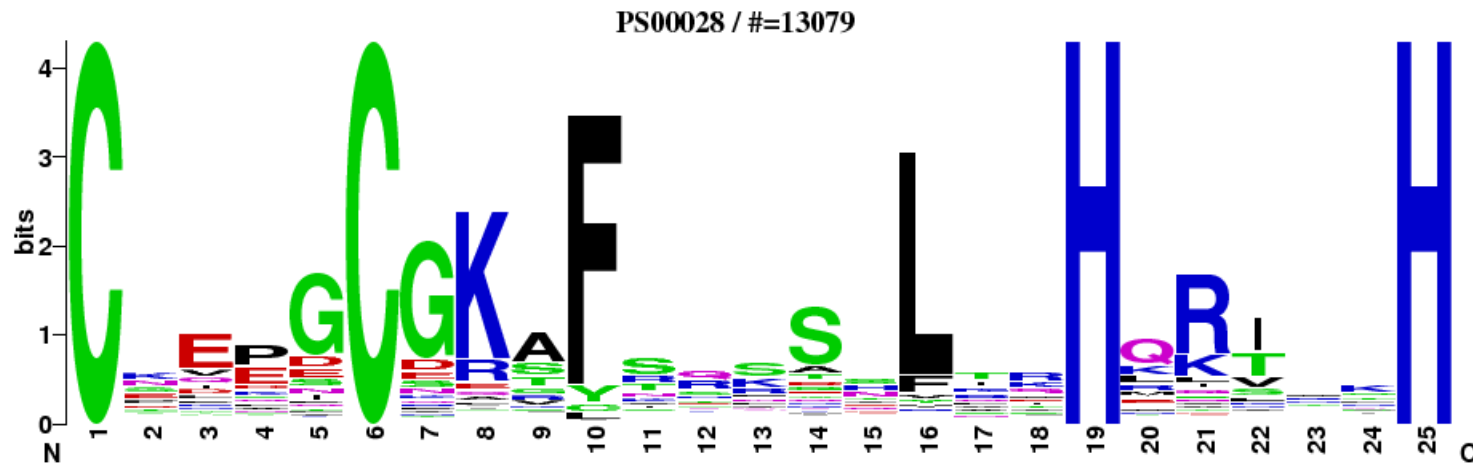
T G C T A G G

A C A C A T C



Patterns and regular expressions 9/9

- Prosite <http://prosite.expasy.org/>
 - 1308 patterns
 - Precision = true hits / (true hits + false positives)
 - Recall = true hits / (true hits + false negatives)
 - Sequence logo





Position Specific Scoring Matrix (PSSM) 1/10

- Πολλαπλή στοίχιση χωρίς κενά
- Υπολογισμός της **συχνότητας εμφάνισης** $f_{i,j}$ κάθε καταλοίπου i σε κάθε στήλη της στοίχισης j
 - π.χ. $f_{A,1}=0/5=0$, $f_{G,1}= 5/5=1$...
 - π.χ. $f_{A,2}=0/5=0$, $f_{H,1}= 5/5=1$...
- Ορισμένες συχνότητες αμινοξέων για κάποιες στήλες είναι ίσες με 0.
- Αποκλείονται αυτά τα αμινοξέα από τις συγκεκριμένες θέσεις.
- Χρήση pseudocounts

```

GHEGVGKVVKIG
GHEKKGYFEDRG
GHEGYGGRSRGG
GHEFEGPKGCGA
GHELRGTTFMPA

```



	1	2	3	4	5	6	7	8	9	10	11	12
A	0	0	0	0	0	0	0	0	0	0	0	2
C	0	0	0	0	0	0	0	0	0	1	0	0
D	0	0	0	0	0	0	0	0	0	1	0	0
E	0	0	5	0	1	0	0	0	1	0	0	0
F	0	0	0	1	0	0	0	1	1	0	0	0
G	5	0	0	2	0	5	1	0	1	0	2	3
H	0	5	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	0	0	1	0
K	0	0	0	1	1	0	1	1	0	1	0	0
L	0	0	0	1	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	1	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	1	0	0	0	1	0
Q	0	0	0	0	0	0	0	0	0	0	0	0
R	0	0	0	0	1	0	0	1	0	1	1	0
S	0	0	0	0	0	0	0	0	1	0	0	0
T	0	0	0	0	0	0	1	1	0	0	0	0
V	0	0	0	0	1	0	0	1	1	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0
Y	0	0	0	0	1	0	1	0	0	0	0	0



Position Specific Scoring Matrix (PSSM) 2/10

● pseudocounts

- Πρόσθεση ενός μικρού αριθμού κατά των υπολογισμό όλων των συχνοτήτων π.χ. 1

- π.χ. $f'_{A,1} = (0+1)/(5+20) = 0.04$

- π.χ. $f'_{G,1} = (5+1)/(5+20) = 0.24$

- Περισσότερο εξειδικευμένες προσεγγίσεις π.χ. Dirichlet mixtures

	1	2	3	4	5	6	7	8	9	10	11	12
A	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	2+1
C	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	1+1	0+1
D	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	1+1	0+1	0+1
E	0+1	0+1	5+1	0+1	1+1	0+1	0+1	0+1	1+1	0+1	0+1	0+1
F	0+1	0+1	0+1	1+1	0+1	0+1	0+1	1+1	1+1	0+1	0+1	0+1
G	5+1	0+1	0+1	2+1	0+1	5+1	1+1	0+1	1+1	0+1	2+1	3+1
H	0+1	5+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1
I	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	1+1	0+1
K	0+1	0+1	0+1	1+1	1+1	0+1	1+1	1+1	0+1	1+1	0+1	0+1
L	0+1	0+1	0+1	1+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1
M	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	1+1	0+1	0+1
N	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1
P	0+1	0+1	0+1	0+1	0+1	0+1	1+1	0+1	0+1	0+1	1+1	0+1
Q	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1
R	0+1	0+1	0+1	0+1	1+1	0+1	0+1	1+1	0+1	1+1	1+1	0+1
S	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	1+1	0+1	0+1	0+1
T	0+1	0+1	0+1	0+1	0+1	0+1	1+1	1+1	0+1	0+1	0+1	0+1
V	0+1	0+1	0+1	0+1	1+1	0+1	0+1	1+1	1+1	0+1	0+1	0+1
W	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1	0+1
Y	0+1	0+1	0+1	0+1	1+1	0+1	1+1	0+1	0+1	0+1	0+1	0+1



Position Specific Scoring Matrix (PSSM) 3/10

- Κανονικοποίηση των συχνοτήτων
 - Διαίρεση με την **αναμενόμενη συχνότητα εμφάνισης** q_i των καταλοίπων σε μια τυχαία ακολουθία
 - π.χ. ίδια συχνότητα για όλα τα αμινοξέα $q_i = 1/20 = 0.05$
 - π.χ. βάσει των συχνοτήτων εμφάνισης στην Uniprot
- Υπολογισμός του **score** S_{ij} του καταλοίπου i στη στήλη j

$$S_{ij} = \log_2 \frac{f'_{ij}}{q_i}$$

- π.χ. $S_{A,1} = \log_2(0.04/0.05) = -0.32$
- π.χ. $S_{G,1} = \log_2(0.24/0.05) = 2.26$



Position Specific Scoring Matrix (PSSM) 4/10

- **Στάθμιση των ακολουθιών** κατά των υπολογισμό των συχνοτήτων
 - μικρό στατιστικό βάρος για πολύ όμοιες ακολουθίες
 - μεγάλο στατιστικό βάρος για λιγότερο όμοιες ακολουθίες
- Το PSSM εφαρμόζεται ως **κυλιόμενο παράθυρο** σε μία ακολουθία
 - Σε κάθε θέση υπολογίζεται το score, αθροίζοντας τα επιμέρους scores.
 - Στο τέλος αναφέρεται **η θέση με το μεγαλύτερο score.**



Position Specific Scoring Matrix (PSSM) 5/10

	1	2	3	4	5	6	7	8	9	10	11	12
A	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	3.8
C	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	2.0	-1.0	-1.0
D	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
E	-1.0	-1.0	6.8	-1.0	2.0	-1.0	-1.0	-1.0	2.0	-1.0	-1.0	-1.0
F	-1.0	-1.0	-1.0	2.0	-1.0	-1.0	-1.0	-1.0	2.0	-1.0	-1.0	-1.0
G	6.8	-1.0	-1.0	3.8	-1.0	6.8	2.0	-1.0	2.0	-1.0	3.8	5.0
H	-1.0	6.8	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
I	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
K	-1.0	-1.0	-1.0	2.0	2.0	-1.0	2.0	2.0	-1.0	2.0	-1.0	-1.0
L	-1.0	-1.0	-1.0	2.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	2.0	-1.0
M	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	2.0	-1.0	-1.0
N	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
P	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	2.0	-1.0
Q	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
R	-1.0	-1.0	-1.0	-1.0	2.0	-1.0	-1.0	2.0	-1.0	2.0	2.0	-1.0
S	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	2.0	-1.0	-1.0	-1.0
T	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	2.0	2.0	-1.0	-1.0	-1.0	-1.0
V	-1.0	-1.0	-1.0	-1.0	2.0	-1.0	-1.0	2.0	2.0	-1.0	-1.0	-1.0
W	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
Y	-1.0	-1.0	-1.0	-1.0	2.0	-1.0	2.0	-1.0	-1.0	-1.0	-1.0	-1.0

T S G H E L V G G V A F P A R C A S

T S G H E L V G G V A F P A R C A S

T S G H E L V G G V A F P A R C A S



T S G H E L V G G V A F P A R C A S



Position Specific Scoring Matrix (PSSM) 6/10

$$S_{T,1} + S_{S,2} + S_{G,3} + S_{H,4} + S_{E,5} + S_{L,6} + S_{V,7} + S_{G,8} + S_{G,9} + S_{V,10} + S_{A,11} + S_{F,12} = -1-1-1-1+2-1-1-1-1+2-1-1-1+2-1-1-1 = -6$$

	1	2	3	4	5	6	7	8	9	10	11	12
A	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	3.8
C	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	2.0	-1.0	-1.0
D	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
E	-1.0	-1.0	6.8	-1.0	2.0	-1.0	-1.0	-1.0	2.0	-1.0	-1.0	-1.0
F	-1.0	-1.0	-1.0	2.0	-1.0	-1.0	-1.0	-1.0	2.0	-1.0	-1.0	-1.0
G	6.8	-1.0	-1.0	3.8	-1.0	6.8	2.0	-1.0	2.0	-1.0	3.8	5.0
H	-1.0	6.8	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
I	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
K	-1.0	-1.0	-1.0	2.0	2.0	-1.0	2.0	2.0	-1.0	2.0	-1.0	-1.0
L	-1.0	-1.0	-1.0	2.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	2.0	-1.0
M	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	2.0	-1.0	-1.0
N	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
P	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	2.0	-1.0
Q	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
R	-1.0	-1.0	-1.0	-1.0	2.0	-1.0	-1.0	2.0	-1.0	2.0	2.0	-1.0
S	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	2.0	-1.0	-1.0	-1.0
T	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	2.0	2.0	-1.0	-1.0	-1.0	-1.0
V	-1.0	-1.0	-1.0	-1.0	2.0	-1.0	-1.0	2.0	2.0	-1.0	-1.0	-1.0
W	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
Y	-1.0	-1.0	-1.0	-1.0	2.0	-1.0	2.0	-1.0	-1.0	-1.0	-1.0	-1.0

Score = -6

T S G H E L V G G V A F P A R C A S



Position Specific Scoring Matrix (PSSM) 7/10

- $$S_{G,1} + S_{H,2} + S_{E,3} + S_{L,4} + S_{V,5} + S_{G,6} + S_{G,7} + S_{V,8} + S_{A,9} + S_{F,10} + S_{P,11} + S_{A,12} = 6.8 + 6.8 + 6.8 + 2 + 2 + 6.8 + 2 + 2 - 1 - 1 + 2 + 3.8 = 39$$

	1	2	3	4	5	6	7	8	9	10	11	12
A	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	3.8
C	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	2.0	-1.0	-1.0
D	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
E	-1.0	-1.0	6.8	-1.0	2.0	-1.0	-1.0	-1.0	2.0	-1.0	-1.0	-1.0
F	-1.0	-1.0	-1.0	2.0	-1.0	-1.0	-1.0	-1.0	2.0	-1.0	-1.0	-1.0
G	6.8	-1.0	-1.0	3.8	-1.0	6.8	2.0	-1.0	2.0	-1.0	3.8	5.0
H	-1.0	6.8	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
I	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
K	-1.0	-1.0	-1.0	2.0	2.0	-1.0	2.0	2.0	-1.0	2.0	-1.0	-1.0
L	-1.0	-1.0	-1.0	2.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	2.0	-1.0
M	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	2.0	-1.0	-1.0
N	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
P	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	2.0	-1.0
Q	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
R	-1.0	-1.0	-1.0	-1.0	2.0	-1.0	-1.0	2.0	-1.0	2.0	2.0	-1.0
S	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	2.0	-1.0	-1.0	-1.0
T	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	2.0	2.0	-1.0	-1.0	-1.0	-1.0
V	-1.0	-1.0	-1.0	-1.0	2.0	-1.0	-1.0	2.0	2.0	-1.0	-1.0	-1.0
W	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0	-1.0
Y	-1.0	-1.0	-1.0	-1.0	2.0	-1.0	2.0	-1.0	-1.0	-1.0	-1.0	-1.0

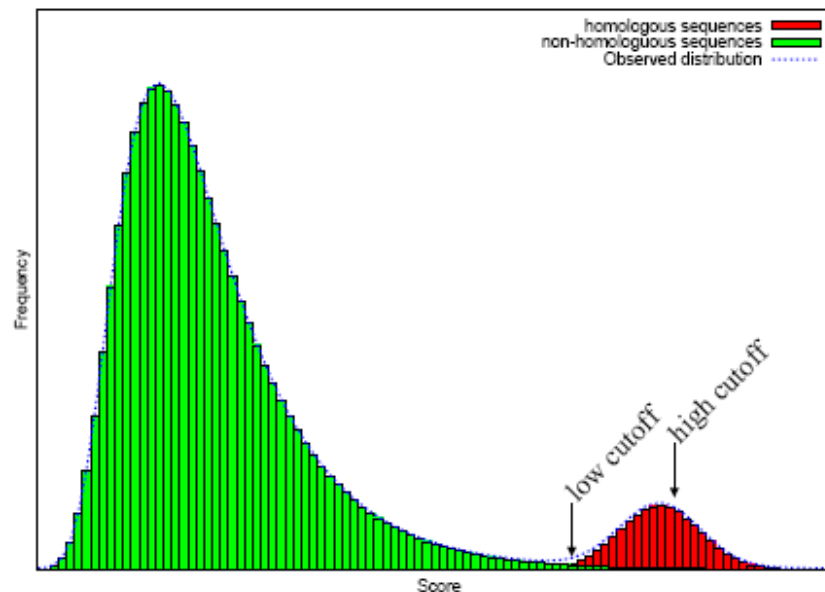
Score = 39





Position Specific Scoring Matrix (PSSM) 8/10

- Κατανομή των scores του PSSM σε τυχαίες ακολουθίες
- Εκτίμηση ενός **E-value**
 - αριθμός των τυχαίων ακολουθιών με PSSM score ίσο ή μεγαλύτερο του παρατηρηθέντος





Position Specific Scoring Matrix (PSSM) 9/10

- Εξάγονται αυτόματα από μη στοιχισμένες ακολουθίες
 - MEME <http://tools.genouest.org/tools/meme/intro.html>
- Πλεονεκτήματα
 - γρήγορη και εύκολη υλοποίηση
 - επιστρέφουν score
- Περιορισμοί
 - απαγορεύονται τα κενά, δεν μπορούν να μοντελοποιηθούν μεγάλες περιοχές
- Χρησιμότητα
 - για τη μοντελοποίηση **περιοχών υψηλής μεταβλητότητας αλλά σταθερού μήκους**

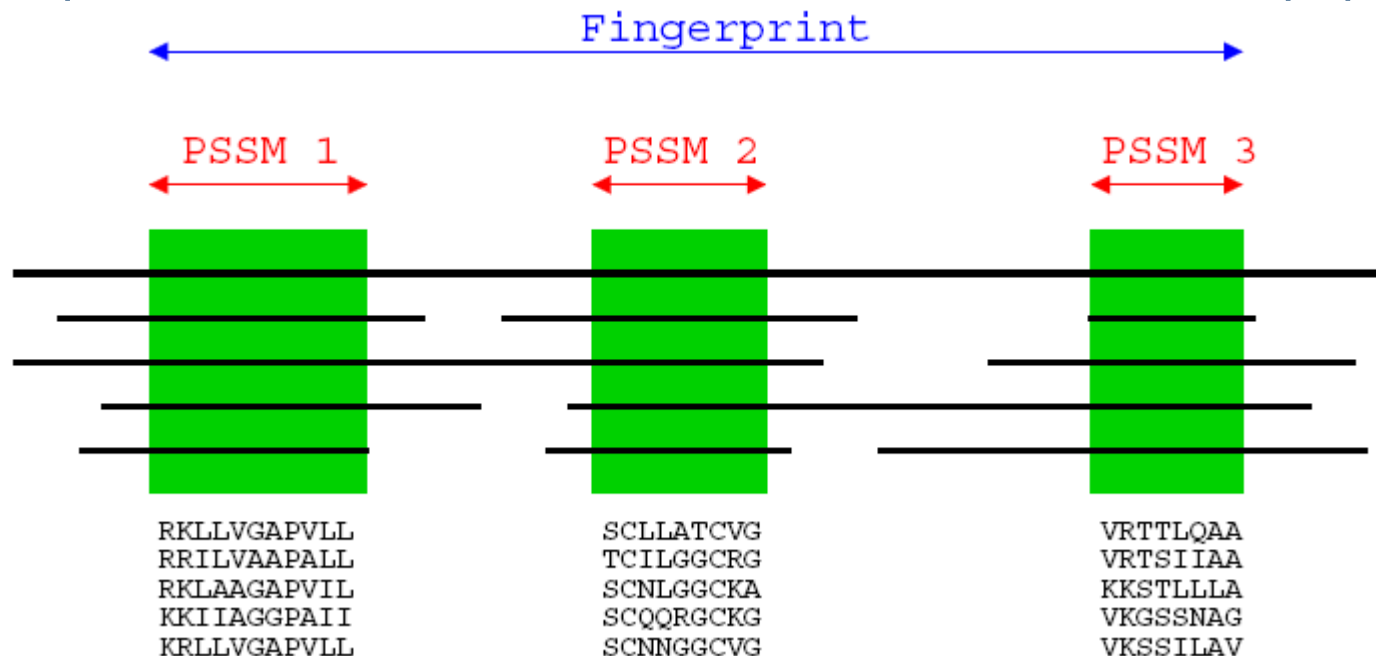


Position Specific Scoring Matrix (PSSM) 10/10

● fingerprints

- συνδυασμός δύο ή περισσότερων PSSMs για την περιγραφή μιας μεγάλου μήκους περιοχής
- PRINTS

<http://www.bioinf.man.ac.uk/dbbrowser/PRINTS/index.php>





Generalized Profiles

- Επέκταση των PSSMs με τη χρήση **ποιών για διαγραφές και προσθήκες** που εξαρτώνται από τη θέση
 - Prosite <http://prosite.expasy.org/>
 - MyHits <http://myhits.isb-sib.ch/>
 - Pftools <ftp://lausanne.isb-sib.ch/pub/software/unix/pftools/>
- Ιδιαίτερα ευαίσθητη αναζήτηση απομακρυσμένων ομόλογων ακολουθιών
- Προγράμματα για τη δημιουργία και τη βαθμονόμηση του profile
- Υψηλές υπολογιστικές απαιτήσεις



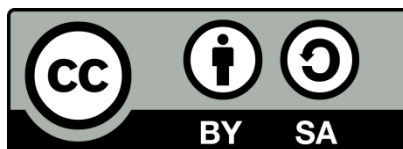
Βιβλιογραφία

- 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



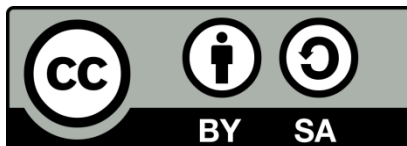
Σημείωμα Αναφοράς

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.