

# Βιοπληροφορική

Blast/PSI-Blast

3ο εργαστήριο

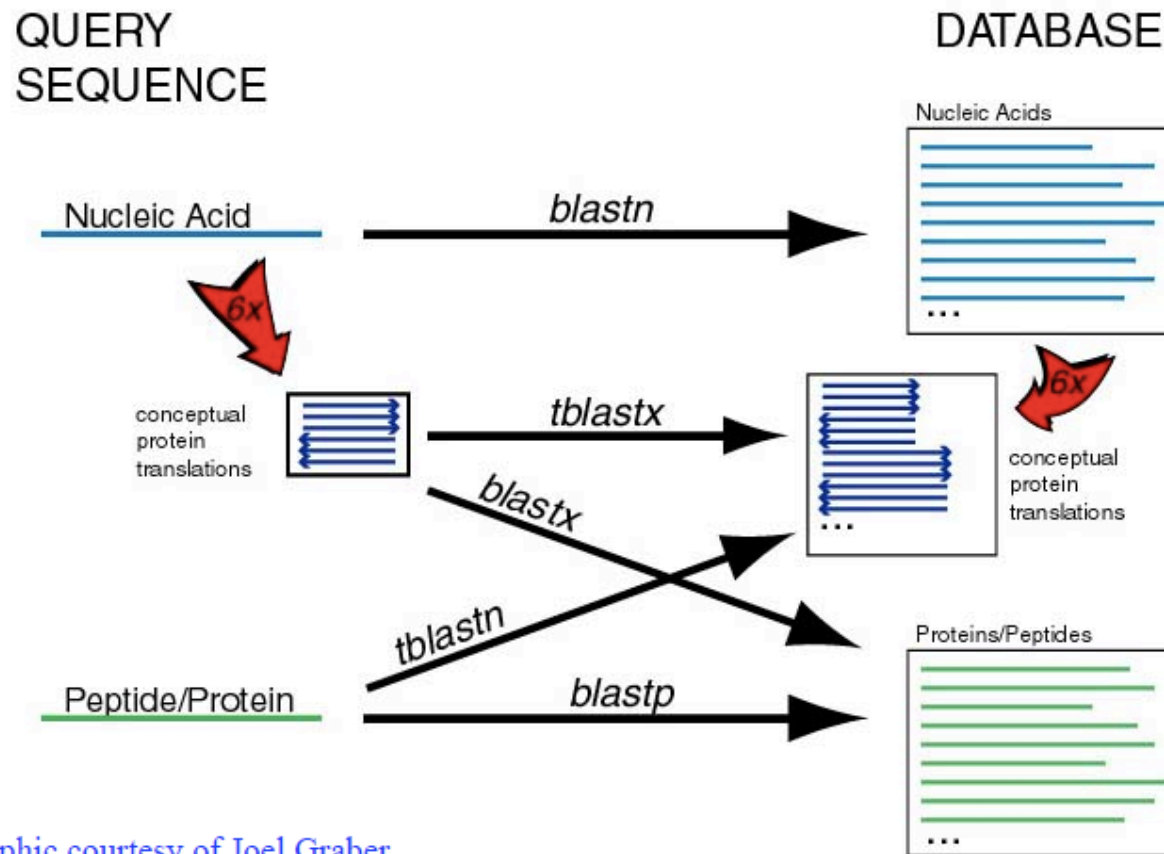
# Αναζήτηση ομόλογων ακολουθιών σε βάσεις δεδομένων (i)

- Ομόλογες ακολουθίες πιθανόν να έχουν παρόμοιες λειτουργίες.
- Ακολουθία επερώτησης (query sequence)
- Υποκείμενες ακολουθίες στην βάση δεδομένων (subject sequences).
- 1 ακολουθία X B.Δ
- N ακολουθίες X B.Δ
- Αναζήτηση με δυναμικό προγραμματισμό: Smith-Waterman, SSearch
- Ευρετικοί αλγόριθμοι για ανίχνευση ομόλογων ακολουθιών.
  - FASTA
  - BLAST
- 50 φορές γρηγορότεροι από δυναμικό προγραμματισμό, αλλά ενδέχεται:
  - να μην εντοπίσουν κάποιες 'απομακρυσμένες' ομόλογες ακολουθίες.
  - να μη γίνει η βέλτιστη στοίχιση

# Blast

Program	Database	Query	Typical uses
BLASTN	Nucleotide	Nucleotide	Mapping oligonucleotides, cDNAs, and PCR products to a genome; screening repetitive elements; cross-species sequence exploration; annotating genomic DNA; clustering sequencing reads; vector clipping
BLASTP	Protein	Protein	Identifying common regions between proteins; collecting related proteins for phylogenetic analyses
BLASTX	Protein	Nucleotide translated into protein	Finding protein-coding genes in genomic DNA; determining if a cDNA corresponds to a known protein
TBLASTN	Nucleotide translated into protein	Protein	Identifying transcripts, potentially from multiple organisms, similar to a given protein; mapping a protein to genomic DNA
TBLASTX	Nucleotide translated into protein	Nucleotide translated into protein	Cross-species gene prediction at the genome or transcript level; searching for genes missed by traditional methods or not yet in protein databases

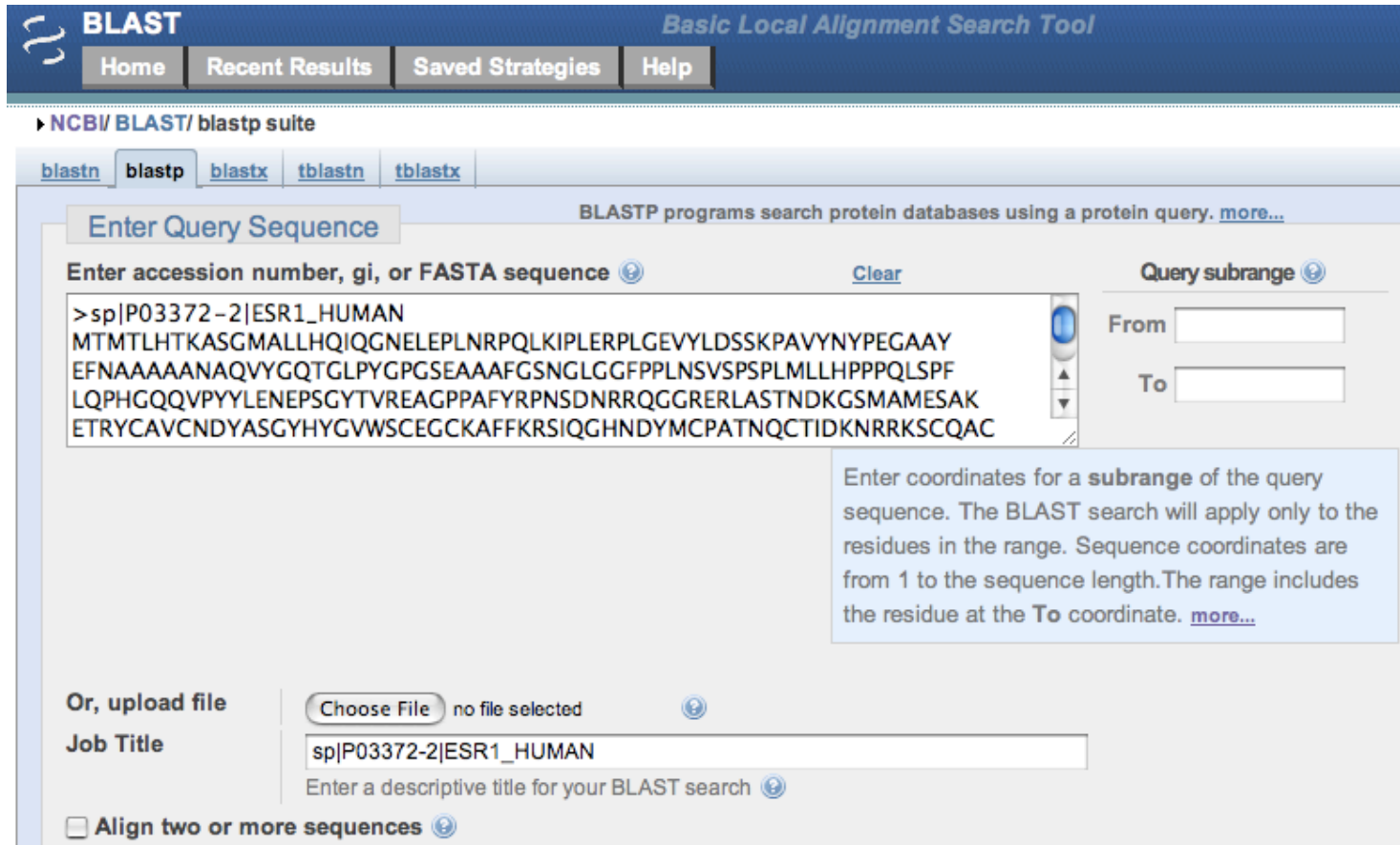
# Blast



Graphic courtesy of Joel Graber.

# Χρησιμοποιώντας το Blast (i)

- Επεξηγήσεις στο σύνδεσμο:
  - <http://www.ncbi.nlm.nih.gov/BLAST/blastcgihelp.shtml>
- Εισάγουμε την ακολουθία της πρωτεΐνης Estrogen receptor α σε μορφή FASTA



The screenshot shows the NCBI BLAST web interface. At the top, there is a navigation bar with the BLAST logo and the text "Basic Local Alignment Search Tool". Below this are tabs for "Home", "Recent Results", "Saved Strategies", and "Help". The main content area is titled "NCBI/ BLAST/ blastp suite" and includes sub-tabs for "blastn", "blastp", "blastx", "tblastn", and "tblastx". The "blastp" tab is selected. The main form is titled "Enter Query Sequence" and contains a text input field with the following FASTA sequence: 

```
>sp|P03372-2|ESR1_HUMAN
MTMTLHTKASGMALLHQIQGNELEPLNRPQLKIPLERPLGEVYLDSSKPAVYNYPEGAAY
EFNAAAAANAQVYGQTGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPLMLLHPPPQLSPF
LQPHGQQVPYYLENEPSGYTVREAGPPAFYRPNSDNRRQGRERLASTNDKGSMMAMESAK
ETRYCAVCNDYASGYHYGVWCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQAC
```

 To the right of the input field are "Clear" and "Query subrange" buttons. Below the input field is a text box with instructions: "Enter coordinates for a **subrange** of the query sequence. The BLAST search will apply only to the residues in the range. Sequence coordinates are from 1 to the sequence length. The range includes the residue at the **To** coordinate. [more...](#)". At the bottom left, there is a section for "Or, upload file" with a "Choose File" button and "no file selected" text. Below this is a "Job Title" field containing "sp|P03372-2|ESR1\_HUMAN" and a prompt "Enter a descriptive title for your BLAST search". At the bottom, there is a checkbox labeled "Align two or more sequences".

# Χρησιμοποιώντας το Blast (ii)

- Επιλέγοντας:
  - τη βάση δεδομένων που θα γίνει η αναζήτηση (Swissprot)
  - Τον οργανισμό που θα γίνει η αναζήτηση (Drosophila melanogaster)

The screenshot displays the NCBI BLAST search interface, divided into two main sections: "Choose Search Set" and "Program Selection".

**Choose Search Set**

- Database:** A dropdown menu is set to "Swissprot protein sequences(swissprot)".
- Organism (Optional):** A text input field contains "Drosophila melanogaster". To its right is an "Exclude" checkbox and a "+" button. Below the field is the text: "Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown."
- Exclude (Optional):** Two checkboxes are present: "Models (XM/XP)" and "Uncultured/environmental sample sequences", both of which are currently unchecked.
- Entrez Query (Optional):** An empty text input field is provided, with the instruction "Enter an Entrez query to limit search" below it.

**Program Selection**

- Algorithm:** Three radio buttons are shown:
  - blastp (protein-protein BLAST)
  - PSI-BLAST (Position-Specific Iterated BLAST)
  - PHI-BLAST (Pattern Hit Initiated BLAST)
- Below the radio buttons is the text: "Choose a BLAST algorithm".

# Χρησιμοποιώντας το Blast (iii)

- Παράμετροι του αλγόριθμου
- Expect threshold: ανάλογα με το τι αναζητούμε

Algorithm parameters Note: Parameter values that differ from the default are highlighted in yellow and marked with ♦ sign

### General Parameters

Max target sequences ♦ 50    
Select the maximum number of aligned sequences to display

Short queries  Automatically adjust parameters for short input sequences

Expect threshold ♦ 1e-10

Word size 3

Max matches in a query range 0

### Scoring Parameters

Matrix BLOSUM62

Gap Costs Existence: 11 Extension: 1

Compositional adjustments Conditional compositional score matrix adjustment

### Filters and Masking

Filter ♦  Low complexity regions

Mask  Mask for lookup table only   
 Mask lower case letters

**BLAST** Search database **Swissprot protein sequences(swissprot)** using **Blastp protein-protein BLAST**  
 Show results in a new window

# Χρησιμοποιώντας το Blast (iv)

- Αποτελέσματα για συντηρημένες επικράτειες (conserved domains)

**ⓘ** Your search is limited to records matching entrez query: *Drosophila melanogaster* [ORGN].

[Edit and Resubmit](#) [Save Search Strategies](#) [▶Formatting options](#) [▶Download](#)

**sp|P03372|ESR1\_HUMAN Estrogen receptor OS=Homo...**

<b>Query ID</b>	Id 47129	<b>Database Name</b>	swissprot
<b>Description</b>	sp P03372 ESR1_HUMAN Estrogen receptor OS=Homo sapiens GN=ESR1 PE=1 SV=2	<b>Description</b>	Non-redundant SwissProt sequences
<b>Molecule type</b>	amino acid	<b>Program</b>	BLASTP 2.2.24+ <a href="#">▶Citation</a>
<b>Query Length</b>	595		

Other reports: [▶Search Summary](#) [\[Taxonomy reports\]](#) [\[Distance tree of results\]](#) [\[Multiple alignment\]](#)

### ▼ Graphic Summary

▼ [Show Conserved Domains](#)

Putative conserved domains have been detected, click on the image below for detailed results.

Query seq. 1 100 200 300 400 500 595

zinc binding site DNA binding site dimer interface ligand binding site coactivator recognition site dimer interface

Specific hits

Superfamilies

Oest\_recep superfamily

NR\_DBD\_ER

NR\_DBD\_like super

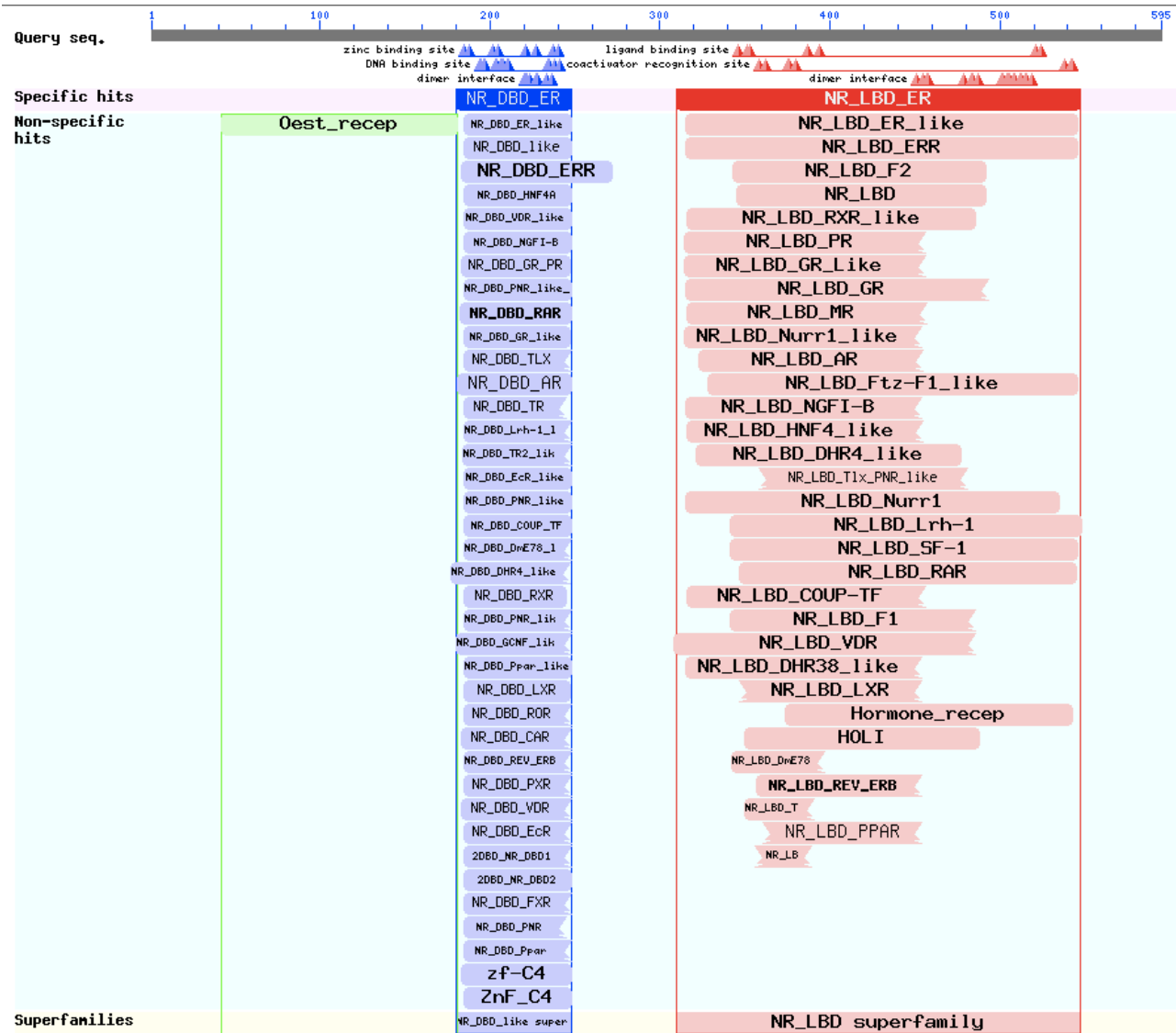
NR\_LBD\_ER

NR\_LBD superfamily



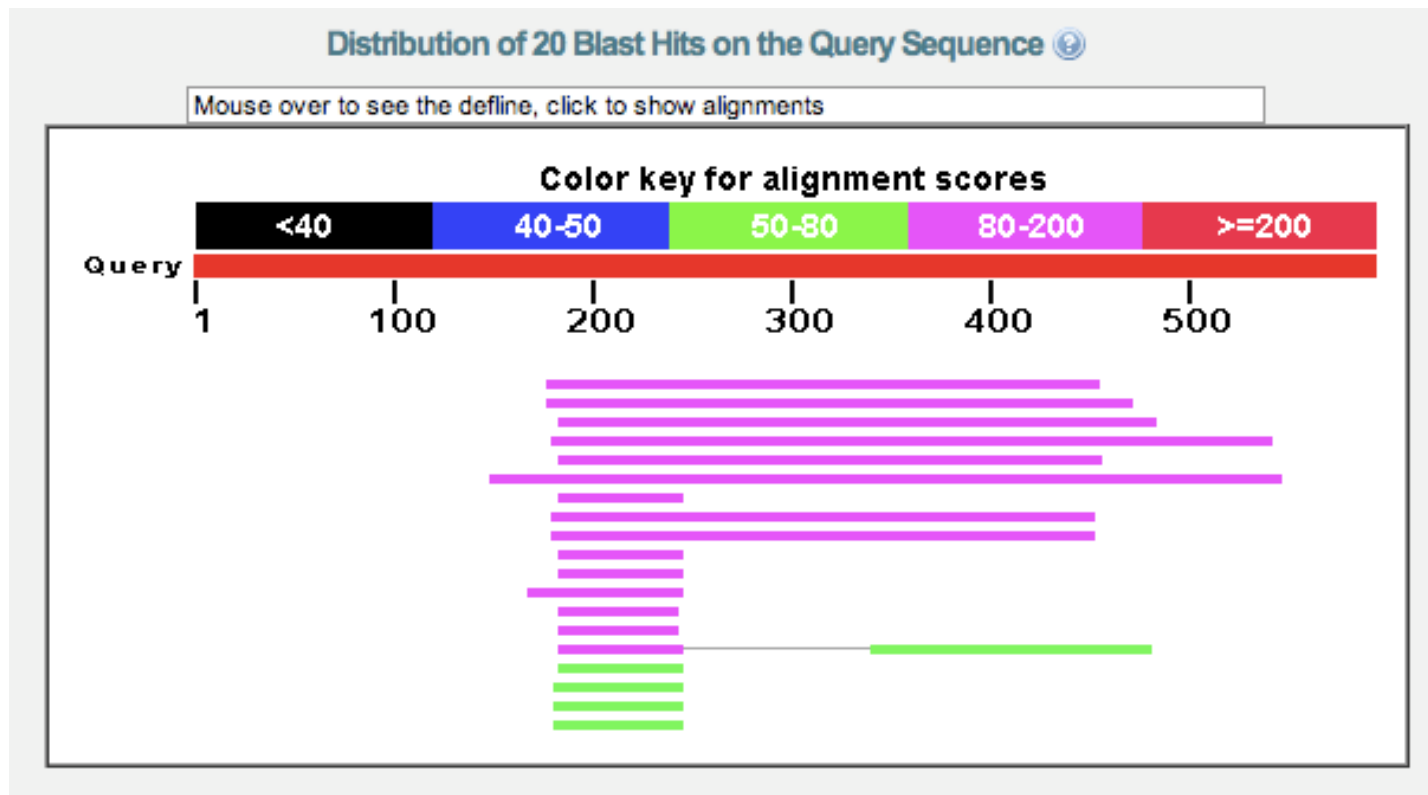
# Χρησιμοποιώντας το Blast (v)

- Ομάδες συντηρημένων επικρατειών



# Χρησιμοποιώντας το Blast (vi)

- Γράφημα των καλύτερων στοιχίσεων



# Χρησιμοποιώντας το Blast (vii)

- Περιγραφές των αποτελεσμάτων (με φίλτρο)

▼ **Descriptions**

Legend for links to other resources: [U](#) UniGene [E](#) GEO [G](#) Gene [S](#) Structure [M](#) Map Viewer [P](#) PubChem BioAssay

**Sequences producing significant alignments:**

Accession	Description	Max score	Total score	Query coverage	E value	Links
<a href="#">P16376.3</a>	RecName: Full=Steroid receptor seven-up, isoform A; AltName:	<a href="#">124</a>	124	46%	2e-29	<a href="#">G</a>
<a href="#">P16375.1</a>	RecName: Full=Steroid receptor seven-up, isoforms B/C; AltName:	<a href="#">124</a>	124	49%	3e-29	<a href="#">G</a>
<a href="#">P49866.2</a>	RecName: Full=Transcription factor HNF-4 homolog; Short=dH	<a href="#">117</a>	117	50%	3e-27	<a href="#">G</a>
<a href="#">P49869.3</a>	RecName: Full=Probable nuclear hormone receptor HR38; Sho	<a href="#">117</a>	117	61%	3e-27	<a href="#">G</a>
<a href="#">P20153.1</a>	RecName: Full=Protein ultraspiracle; AltName: Full=Chorion fa	<a href="#">113</a>	113	46%	6e-26	<a href="#">G</a>
<a href="#">P34021.1</a>	RecName: Full=Ecdysone receptor; AltName: Full=20-hydroxy	<a href="#">105</a>	105	67%	2e-23	<a href="#">G</a> <a href="#">M</a>
<a href="#">P18102.1</a>	RecName: Full=Protein tailless; AltName: Full=Nuclear recepto	<a href="#">91.3</a>	91.3	10%	2e-19	<a href="#">G</a>
<a href="#">P13055.2</a>	RecName: Full=Ecdysone-induced protein 75B, isoform B; AltN	<a href="#">90.1</a>	90.1	45%	6e-19	<a href="#">G</a>
<a href="#">P17671.2</a>	RecName: Full=Ecdysone-induced protein 75B, isoforms C/D; A	<a href="#">89.7</a>	89.7	45%	8e-19	<a href="#">G</a>
<a href="#">P33244.2</a>	RecName: Full=Nuclear hormone receptor FTZ-F1; AltName: F	<a href="#">88.2</a>	88.2	10%	2e-18	<a href="#">G</a>
<a href="#">Q9W539.4</a>	RecName: Full=Hormone receptor 4; Short=dHR4; AltName: F	<a href="#">86.7</a>	86.7	10%	7e-18	<a href="#">G</a>
<a href="#">P31396.1</a>	RecName: Full=Probable nuclear hormone receptor HR3; Short	<a href="#">85.5</a>	85.5	13%	1e-17	<a href="#">G</a>
<a href="#">P45447.3</a>	RecName: Full=Ecdysone-induced protein 78C; Short=DR-78; A	<a href="#">84.0</a>	84.0	10%	5e-17	<a href="#">G</a>
<a href="#">Q24142.2</a>	RecName: Full=Nuclear hormone receptor HR78; Short=dHR78	<a href="#">82.0</a>	82.0	10%	2e-16	<a href="#">G</a>
<a href="#">Q05192.3</a>	RecName: Full=Nuclear hormone receptor FTZ-F1 beta; AltName:	<a href="#">81.3</a>	149	34%	2e-16	<a href="#">G</a>
<a href="#">P10734.1</a>	RecName: Full=Zygotic gap protein knirps; AltName: Full=Nuc	<a href="#">77.4</a>	77.4	10%	4e-15	<a href="#">G</a>
<a href="#">Q24143.1</a>	RecName: Full=Nuclear hormone receptor HR96; Short=dHR96	<a href="#">76.3</a>	76.3	11%	9e-15	<a href="#">G</a>
<a href="#">P13054.1</a>	RecName: Full=Knirps-related protein; AltName: Full=Nuclear	<a href="#">73.9</a>	73.9	11%	5e-14	<a href="#">G</a>
<a href="#">P15370.2</a>	RecName: Full=Protein embryonic gonad; AltName: Full=Nucle	<a href="#">70.5</a>	70.5	11%	4e-13	<a href="#">G</a>

# Χρησιμοποιώντας το Blast (viii)

- Στοιχίσεις (με φίλτρο - μικρά γράμματα)
- Identities (επί του αριθμού θέσεων στην στοίχιση)
- Positives (επί του αριθμού θέσεων στην στοίχιση)

```
> sp|P16376.3|7UP2\_DROME G RecName: Full=Steroid receptor seven-up, isoform A; AltName: Full=Nuclear receptor subfamily 2 group F member 3, isoform A
Length=746

GENE ID: 41491 svp | seven up [Drosophila melanogaster] (Over 100 PubMed links)

Score = 124 bits (312), Expect = 2e-29, Method: Compositional matrix adjust.
Identities = 87/282 (31%), Positives = 133/282 (48%), Gaps = 25/282 (8%)

Query 179 AKETRYCAVCNDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMCPATNQCTIDKNRRKSCQ 238
      +K+  C VC D +SG HYG ++CEGCK+FFKRS++ + Y C + C ID++ R CQ
Sbjct 194 SKQNIIECVVCGDKSSGKHYGQFTCEGCKSFFKRSVRRNLTYSCRGRNCPIDQHHRNQSQ 253

Query 239 ACRLRKCYEVgmmkggirkdrrrggrmLKHKRQRDDGEGRGEVGSAGDMRAANLWPSPLMI 298
      CRL+KC ++GM + + +R R G G G + AN P+ I
Sbjct 254 YCRLKCLKMGMRRREAV-----QRGRVPPTQPGLAGMHGQYQIAN--GDPMGI 299

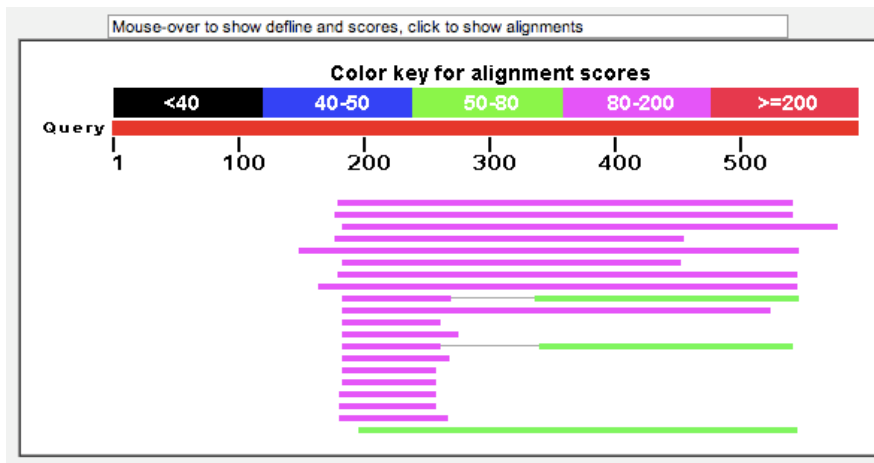
Query 299 KRSKKNLALSILTADQMVSALLDAEPPILYSEYDPTPRPFSEASMMGL--LTNLADRELVH 356
      +S S +S LL AEP Y + ++MG+ + LA R L
Sbjct 300 AGFNGHSYLSSY-----ISLLRAEP---YPTSRYGQCMQPNNIMGIDNICELARLLFS 351

Query 357 MINWAKRVPGFVDLTLHDQVHLLLECAWLEIILMIGLVWRSME-HPGKLLFAPNLLLDRNQG 415
      + WAK +P F +L + DQV LL W E+ ++ SM H LL A L
Sbjct 352 AVEWAKNIPFFPELQVTDQVALLRLVWSELVNLNASQCSMPLHVAPLLAAAGLHASPMAA 411

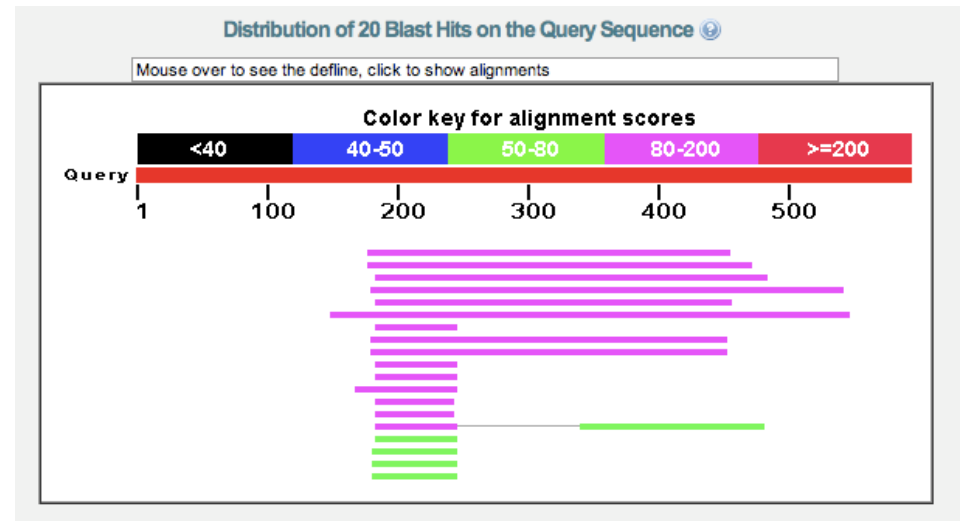
Query 416 KCVEGMVEIFDMLLATSSRFRMMNLQGEFVCLKSIILLNSG 457
      V ++ + + + + E+ CLK+I+L +G
Sbjct 412 DRVVAFMDHIRIFQEQVEKLKALHVDSAEYSCLKAIVLFTTG 453
```

# Χρησιμοποιώντας το Blast (ix)

χωρίς φίλτρο



με φίλτρο



Η χρήση φίλτρου αλλάζει το score  
Identities/Positives σταθερά

# Χρησιμοποιώντας το Blast (x)

## Χωρίς φίλτρο

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Links
<a href="#">P49869.3</a>	RecName: Full=Probable nuclear hormone receptor HR38; Sho	<a href="#">144</a>	144	61%	2e-35	
<a href="#">P16375.1</a>	RecName: Full=Steroid receptor seven-up, isoforms B/C; AltN	<a href="#">144</a>	144	61%	3e-35	<a href="#">G</a>
<a href="#">P49866.2</a>	RecName: Full=Transcription factor HNF-4 homolog; Short=dH	<a href="#">144</a>	144	66%	3e-35	<a href="#">G</a>
<a href="#">P16376.3</a>	RecName: Full=Steroid receptor seven-up, isoform A; AltName	<a href="#">132</a>	132	46%	1e-31	<a href="#">G</a>
<a href="#">P34021.1</a>	RecName: Full=Ecdysone receptor; AltName: Full=20-hydroxy	<a href="#">121</a>	121	67%	2e-28	<a href="#">G</a>
<a href="#">P18102.1</a>	RecName: Full=Protein tailless; AltName: Full=Nuclear recepto	<a href="#">121</a>	121	45%	2e-28	<a href="#">G</a>
<a href="#">P13055.2</a>	RecName: Full=Ecdysone-induced protein 75B, isoform B; AltN	<a href="#">110</a>	110	61%	6e-25	<a href="#">G</a>
<a href="#">P17671.2</a>	RecName: Full=Ecdysone-induced protein 75B, isoforms C/D; A	<a href="#">110</a>	110	64%	6e-25	<a href="#">G</a>
<a href="#">P33244.2</a>	RecName: Full=Nuclear hormone receptor FTZ-F1; AltName: F	<a href="#">103</a>	171	49%	4e-23	<a href="#">G</a>
<a href="#">P20153.1</a>	RecName: Full=Protein ultraspiracle; AltName: Full=Chorion fa	<a href="#">101</a>	101	57%	3e-22	<a href="#">G</a>
<a href="#">Q9W539.4</a>	RecName: Full=Hormone receptor 4; Short=dHR4; AltName: F	<a href="#">100</a>	100	13%	5e-22	<a href="#">G</a>
<a href="#">P31396.1</a>	RecName: Full=Probable nuclear hormone receptor HR3; Short	<a href="#">99.8</a>	99.8	15%	7e-22	<a href="#">G</a>
<a href="#">Q05192.3</a>	RecName: Full=Nuclear hormone receptor FTZ-F1 beta; AltNar	<a href="#">96.7</a>	166	47%	6e-21	<a href="#">G</a>
<a href="#">P45447.3</a>	RecName: Full=Ecdysone-induced protein 78C; Short=DR-78; A	<a href="#">93.2</a>	93.2	14%	7e-20	
<a href="#">P10734.1</a>	RecName: Full=Zygotic gap protein knirps; AltName: Full=Nuc	<a href="#">93.2</a>	93.2	12%	7e-20	<a href="#">G</a>
<a href="#">Q24142.2</a>	RecName: Full=Nuclear hormone receptor HR78; Short=dHR78	<a href="#">89.0</a>	89.0	12%	1e-18	<a href="#">G</a>
<a href="#">P13054.1</a>	RecName: Full=Knirps-related protein; AltName: Full=Nuclear	<a href="#">86.7</a>	86.7	13%	7e-18	<a href="#">G</a>
<a href="#">P15370.2</a>	RecName: Full=Protein embryonic gonad; AltName: Full=Nucle	<a href="#">83.6</a>	83.6	13%	6e-17	<a href="#">G</a>
<a href="#">Q24143.1</a>	RecName: Full=Nuclear hormone receptor HR96; Short=dHR96	<a href="#">82.0</a>	82.0	14%	2e-16	<a href="#">G</a>
<a href="#">P17672.2</a>	RecName: Full=Ecdysone-induced protein 75B, isoform A; Sho	<a href="#">68.9</a>	68.9	58%	2e-12	<a href="#">G</a>

## Με φίλτρο

▼ Descriptions

Legend for links to other resources: [U](#) UniGene [E](#) GEO [G](#) Gene [S](#) Structure [M](#) Map Viewer [P](#) PubChem BioAssay

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Links
<a href="#">P16376.3</a>	RecName: Full=Steroid receptor seven-up, isoform A; AltName	<a href="#">124</a>	124	46%	2e-29	<a href="#">G</a>
<a href="#">P16375.1</a>	RecName: Full=Steroid receptor seven-up, isoforms B/C; AltN	<a href="#">124</a>	124	49%	3e-29	<a href="#">G</a>
<a href="#">P49866.2</a>	RecName: Full=Transcription factor HNF-4 homolog; Short=dH	<a href="#">117</a>	117	50%	3e-27	<a href="#">G</a>
<a href="#">P49869.3</a>	RecName: Full=Probable nuclear hormone receptor HR38; Sho	<a href="#">117</a>	117	61%	3e-27	

Αλλάζει το score, E-value και η σειρά εμφάνισης

# Χρησιμοποιώντας το Blast (xi)

## Χωρίς φίλτρο

```
> sp|P16376.3|7UP2\_DROME G RecName: Full=Steroid receptor seven-up, isoform A; AltName: Full=Nuclear receptor subfamily 2 group F member 3, isoform A
Length=746
```

[GENE ID: 41491 svp](#) | seven up [Drosophila melanogaster] ([Over 100 PubMed links](#))

Score = 132 bits (331), Expect = 1e-31, Method: Compositional matrix adjust.  
Identities = 87/282 (31%), Positives = 133/282 (48%), Gaps = 25/282 (8%)

```
Query 179 AKETRYCAVCNDYASGYHYGVWSCEGCKAFFKRSIQGHNDYMC PATNQCTIDKNRRKSCQ 238
      +K+ C VC D +SG HYG ++CEGCK+FFKRS++ + Y C + C ID++ R CQ
Sbjct 194 SKQNIIECVVCGDKSSGKHYQFTCEGCKSFFKRSVRRNLTYSCRGSRNCPIDQHHRNQCQ 253

Query 239 ACRLRKCYEVGMMKGGIRKDRRGGRMLKHKRQRDDGEGRGEVGSAGDMRAANLWPSPLMI 298
      CRL+KC ++GM + + +R R G G G + AN P+ I
Sbjct 254 YCRLKCKLKMGRREAV-----QRGRVPPTQPGLAGMHGQYQIAN--GDPMGI 299

Query 299 KRSKKNLALSILTADQMVSALLDAEPPILYSEYDPTPRPFSEASMMGL--LTNLADREL VH 356
      +S S +S LL AEP Y + ++MG+ + LA R L
Sbjct 300 AGFNGHSYLSSY-----ISLLLRAEP---YPTSRYGQCMQPNNIMGIDNICELARLLFS 351

Query 357 MINWAKRVPGFVDLTLHDQVHLLCAWLEILMIGLVWRSME-HPGKLLFAPNLLDRNQG 415
      + WAK +P F +L + DQV LL W E+ ++ SM H LL A L
Sbjct 352 AVEWAKNIPFFPELQVTDQVALLRLVWSELVFLNASQCSMPLHVAPLLAAAGLHASPMAA 411

Query 416 KCVEGMVEIFDMLLATSSRFRMMNLQGEFVCLKSIILLNSG 457
      V ++ + + + + + E+ CLK+I+L +G
Sbjct 412 DRVVAFMDHIRIFQEQVEKALKALHVDSAEYSCLKAIVLFTTG 453
```

## Με φίλτρο

[GENE ID: 41491 svp](#) | seven up [Drosophila melanogaster] ([Over 100 PubMed links](#))

Score = 124 bits (312), Expect = 2e-29, Method: Compositional matrix adjust.  
Identities = 87/282 (31%), Positives = 133/282 (48%), Gaps = 25/282 (8%)

Identities & positives παραμένουν σταθερά



# Χρησιμοποιώντας το Blast (xi)

- Αλλαγή στον Πίνακα αντικατάστασης και στις ποινές για κενά
  - Blosum 45 13:3, χωρίς φίλτρο

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Links
<a href="#">P49869.3</a>	RecName: Full=Probable nuclear hormone receptor HR38; Sho	<a href="#">146</a>	146	61%	6e-36	
<a href="#">P16375.1</a>	RecName: Full=Steroid receptor seven-up, isoforms B/C; AltN	<a href="#">145</a>	145	61%	1e-35	<a href="#">G</a>
<a href="#">P49866.2</a>	RecName: Full=Transcription factor HNF-4 homolog; Short=dH	<a href="#">139</a>	139	66%	5e-34	<a href="#">G</a>
<a href="#">P16376.3</a>	RecName: Full=Steroid receptor seven-up, isoform A; AltName	<a href="#">138</a>	138	46%	2e-33	<a href="#">G</a>
<a href="#">P34021.1</a>	RecName: Full=Ecdysone receptor; AltName: Full=20-hydroxy	<a href="#">116</a>	116	67%	8e-27	<a href="#">G</a>
<a href="#">P17671.2</a>	RecName: Full=Ecdysone-induced protein 75B, isoforms C/D; A	<a href="#">112</a>	112	61%	7e-26	<a href="#">G</a>
<a href="#">P13055.2</a>	RecName: Full=Ecdysone-induced protein 75B, isoform B; AltN	<a href="#">112</a>	112	47%	1e-25	<a href="#">G</a>
<a href="#">P33244.2</a>	RecName: Full=Nuclear hormone receptor FTZ-F1; AltName: F	<a href="#">110</a>	177	49%	3e-25	<a href="#">G</a>
<a href="#">P18102.1</a>	RecName: Full=Protein tailless; AltName: Full=Nuclear recept	<a href="#">108</a>	108	16%	1e-24	<a href="#">G</a>
<a href="#">Q9W539.4</a>	RecName: Full=Hormone receptor 4; Short=dHR4; AltName: F	<a href="#">108</a>	108	13%	2e-24	<a href="#">G</a>
<a href="#">P31396.1</a>	RecName: Full=Probable nuclear hormone receptor HR3; Short	<a href="#">108</a>	108	18%	2e-24	<a href="#">G</a>
<a href="#">Q05192.3</a>	RecName: Full=Nuclear hormone receptor FTZ-F1 beta; AltNar	<a href="#">102</a>	172	37%	1e-22	<a href="#">G</a>
<a href="#">P10734.1</a>	RecName: Full=Zygotic gap protein knirps; AltName: Full=Nuc	<a href="#">99.8</a>	99.8	13%	7e-22	<a href="#">G</a>
<a href="#">P45447.3</a>	RecName: Full=Ecdysone-induced protein 78C; Short=DR-78; A	<a href="#">99.5</a>	99.5	12%	9e-22	
<a href="#">Q24142.2</a>	RecName: Full=Nuclear hormone receptor HR78; Short=dHR78	<a href="#">96.3</a>	96.3	13%	7e-21	<a href="#">G</a>
<a href="#">P13054.1</a>	RecName: Full=Knirps-related protein; AltName: Full=Nuclear	<a href="#">94.8</a>	94.8	13%	2e-20	<a href="#">G</a>
<a href="#">P15370.2</a>	RecName: Full=Protein embryonic gonad; AltName: Full=Nucle	<a href="#">90.7</a>	90.7	14%	3e-19	<a href="#">G</a>
<a href="#">Q24143.1</a>	RecName: Full=Nuclear hormone receptor HR96; Short=dHR96	<a href="#">89.5</a>	89.5	12%	9e-19	<a href="#">G</a>
<a href="#">P17672.2</a>	RecName: Full=Ecdysone-induced protein 75B, isoform A; Sho	<a href="#">67.3</a>	67.3	55%	4e-12	<a href="#">G</a>
<a href="#">P20153.1</a>	RecName: Full=Protein ultraspiracle; AltName: Full=Chorion fa	<a href="#">65.2</a>	65.2	31%	2e-11	<a href="#">G</a>

Blosum 62 11:1, χωρίς φίλτρο

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Links
<a href="#">P49869.3</a>	RecName: Full=Probable nuclear hormone receptor HR38; Sho	<a href="#">144</a>	144	61%	2e-35	
<a href="#">P16375.1</a>	RecName: Full=Steroid receptor seven-up, isoforms B/C; AltN	<a href="#">144</a>	144	61%	3e-35	<a href="#">G</a>
<a href="#">P49866.2</a>	RecName: Full=Transcription factor HNF-4 homolog; Short=dH	<a href="#">144</a>	144	66%	3e-35	<a href="#">G</a>
<a href="#">P16376.3</a>	RecName: Full=Steroid receptor seven-up, isoform A; AltName	<a href="#">132</a>	132	46%	1e-31	<a href="#">G</a>
<a href="#">P34021.1</a>	RecName: Full=Ecdysone receptor; AltName: Full=20-hydroxy	<a href="#">121</a>	121	67%	2e-28	<a href="#">G</a>



# Χρησιμοποιώντας το Blast (xii)

## Blosum 45 13:3

## Blosum 62 11:1

<a href="#">sp P16376.3 7UP2_DROME</a> <b>G</b> RecName: Full=Steroid receptor seven-up, isoform A Full=Nuclear receptor subfamily 2 group F member 3, isoform A Length=746		<a href="#">sp P16376.3 7UP2_DROME</a> <b>G</b> RecName: Full=Steroid receptor seven-up, isoform A; Full=Nuclear receptor subfamily 2 group F member 3, isoform A Length=746	
<a href="#">GENE ID: 41491 svp</a>   seven up [Drosophila melanogaster] (Over 100 PubMed links)		<a href="#">GENE ID: 41491 svp</a>   seven up [Drosophila melanogaster] (Over 100 PubMed links)	
Score = 138 bits (456), Expect = 2e-33, Method: Compositional matrix adjust. Identities = 87/282 (31%), Positives = 137/282 (49%), Gaps = 25/282 (8%)		Score = 132 bits (331), Expect = 1e-31, Method: Compositional matrix adjust. Identities = 87/282 (31%), Positives = 133/282 (48%), Gaps = 25/282 (8%)	
Query 179 AKETRYCAVCNDYASGYHYGVWVSCGCKAFFKRSIQGHNDYMC PATNQCTIDKNRRKSCQ 238	Query 179 AKETRYCAVCNDYASGYHYGVWVSCGCKAFFKRSIQGHNDYMC PATNQCTIDKNRRKSCQ 238	Sbjct 194 SKQNIIECVVCGDKSSGKHYGQFTCEGCKSFFKRSVRRNLTYSCRGSRNCPIDQHHRNQCC 253	Sbjct 194 SKQNIIECVVCGDKSSGKHYGQFTCEGCKSFFKRSVRRNLTYSCRGSRNCPIDQHHRNQCC 253
Query 239 ACRLRKCYEVMGMMKGGIRKDRRGRMLKHKRQRDDGEGRGEVGSAGDMRAANLWPSPLMI 298	Query 239 ACRLRKCYEVMGMMKGGIRKDRRGRMLKHKRQRDDGEGRGEVGSAGDMRAANLWPSPLMI 298	Sbjct 254 YCRLKKCLKMGMRREAV-----QRGRVPPTQPGLAGMHGQYQIAN--GDPMGI 299	Sbjct 254 YCRLKKCLKMGMRREAV-----QRGRVPPTQPGLAGMHGQYQIAN--GDPMGI 299
Query 299 KRSKKNLALSLTADQMVSAALLDAEPPILYSEYDPTPRPFSEASMMGL--LTNLADRELVH 356	Query 299 KRSKKNLALSLTADQMVSAALLDAEPPILYSEYDPTPRPFSEASMMGL--LTNLADRELVH 356	Sbjct 300 AGFNHGSYLSS-----YISLLLRAEP---YPTSRYGQCMQPNINIMGIDNICELARLLFS 351	Sbjct 300 AGFNHGSYLSSY-----ISLLLRAEP---YPTSRYGQCMQPNINIMGIDNICELARLLFS 351
Query 357 MINWAKRVPGFVDLTLHDQVHLLCEAWLEILMIGLVVRSME--HPGKLLFAPNLLDRNQG 415	Query 357 MINWAKRVPGFVDLTLHDQVHLLCEAWLEILMIGLVVRSME--HPGKLLFAPNLLDRNQG 415	Sbjct 352 AVEWAKNIPFPPELQVTDQVALLRLVWSELVFLNASQCSMPLHVAPLLAAAGLHASPMAA 411	Sbjct 352 AVEWAKNIPFPPELQVTDQVALLRLVWSELVFLNASQCSMPLHVAPLLAAAGLHASPMAA 411
Query 416 KCVEGMVEIFDMLLATSSRFMMNLQGEFVCLKSIILLNSG 457	Query 416 KCVEGMVEIFDMLLATSSRFMMNLQGEFVCLKSIILLNSG 457	Sbjct 412 DRVVAFMHIRIFQEQVEKALKALHVDSAEYSCLKAIVLFTTG 453	Sbjct 412 DRVVAFMHIRIFQEQVEKALKALHVDSAEYSCLKAIVLFTTG 453

Μικρές διαφορές στη στοίχιση, στο score & E-value

# Χρησιμοποιώντας το Blast (xiii)

- Αν για το ίδιο γονίδιο ( ESR1\_Human) χρησιμοποιούσαμε το mRNA του (X03635.1 Homo sapiens mRNA for estrogen receptor α και όχι την πρωτεΐνη για την αναζήτηση στην Drosophila:
  - Blastn (nr database)

**ⓘ** Your search is limited to records matching entrez query: Drosophila melanogaster [ORGN].  
[Edit and Resubmit](#) [Save Search Strategies](#) [▶Formatting options](#) [▶Download](#)

**ESR1\_mRNA\_human**

<b>Query ID</b>	lc 45497	<b>Database Name</b>	nr
<b>Description</b>	ESR1_mRNA_human	<b>Description</b>	All GenBank+EMBL+DDBJ+PDB sequences (but no EST, STS, GSS,environmental samples or phase 0, 1 or 2 HTGS sequences)
<b>Molecule type</b>	nucleic acid	<b>Program</b>	BLASTN 2.2.24+ <a href="#">▶Citation</a>
<b>Query Length</b>	6450		

**ⓘ** No significant similarity found. For reasons why,[click here](#)

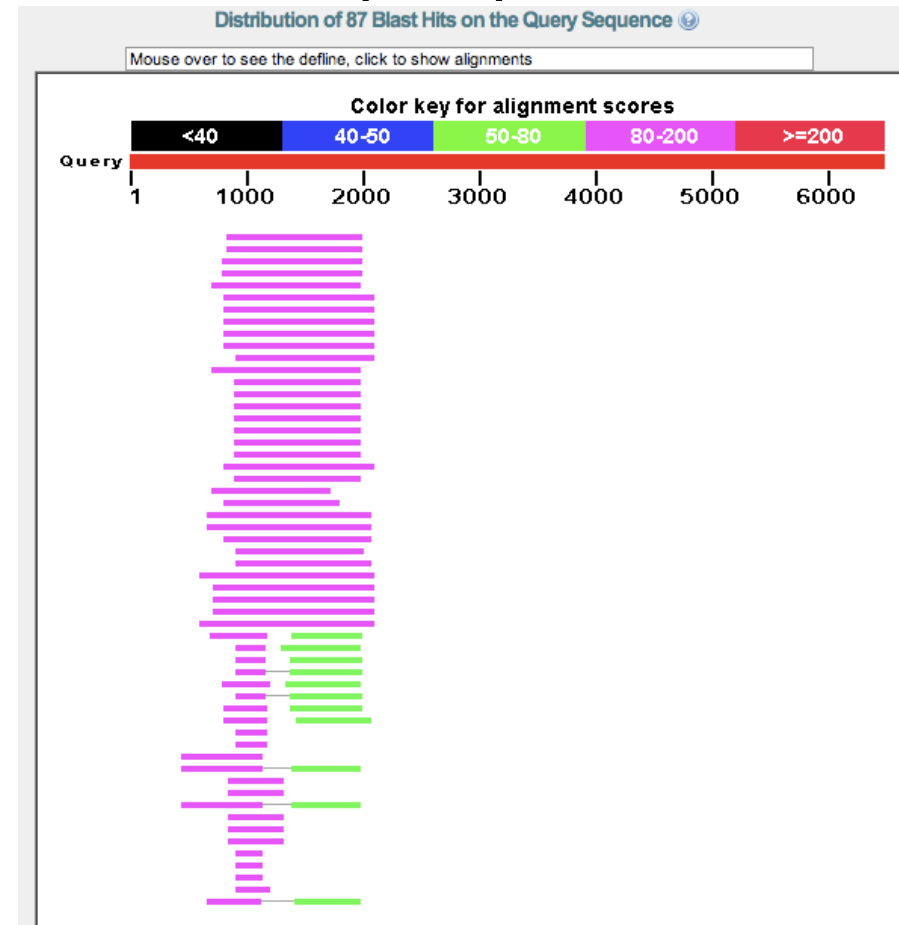
Other reports: [▶Search Summary](#)

Search Parameters	
Program	blastn
Word size	11
Expect value	1e-05
Hitlist size	100
Match/Mismatch scores	2,-3
Gapcosts	5,2
Low Complexity Filter	Yes
Filter string	L;m;
Genetic Code	1

# Χρησιμοποιώντας το Blast (xiv)

Για το ίδιο mRNA

- Blastx (nr database)



Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
<a href="#">NP_729340.1</a>	estrogen-related receptor, isoform A [Drosophila melanogaster]	<a href="#">198</a>	198	17%	5e-50	31%	<a href="#">UG</a>
<a href="#">NP_648183.3</a>	estrogen-related receptor, isoform B [Drosophila melanogaster]	<a href="#">198</a>	198	17%	5e-50	31%	<a href="#">UG</a>
<a href="#">CAA36827.1</a>	unnamed protein product [Drosophila melanogaster]	<a href="#">154</a>	154	18%	1e-36	26%	
<a href="#">NP_476781.1</a>	ultraspiracle [Drosophila melanogaster] >sp P20153.1 USP_	<a href="#">151</a>	151	18%	7e-36	25%	<a href="#">UG</a>
<a href="#">NP_524325.1</a>	seven up, isoform B [Drosophila melanogaster] >sp P16375	<a href="#">143</a>	143	19%	3e-33	27%	<a href="#">UG</a>
<a href="#">ACS68165.1</a>	FI04795p [Drosophila melanogaster]	<a href="#">140</a>	140	19%	2e-32	25%	
<a href="#">NP_001097126.1</a>	hepatocyte nuclear factor 4, isoform D [Drosophila melanogaster]	<a href="#">140</a>	140	19%	2e-32	25%	<a href="#">G</a>
<a href="#">NP_723413.1</a>	hepatocyte nuclear factor 4, isoform B [Drosophila melanogaster]	<a href="#">140</a>	140	19%	2e-32	25%	<a href="#">UG</a>
<a href="#">NP_476887.2</a>	hepatocyte nuclear factor 4, isoform A [Drosophila melanogaster]	<a href="#">140</a>	140	19%	2e-32	25%	<a href="#">G</a>
<a href="#">NP_723414.1</a>	hepatocyte nuclear factor 4, isoform C [Drosophila melanogaster]	<a href="#">140</a>	140	19%	2e-32	25%	<a href="#">UG</a>
<a href="#">AAB09592.1</a>	hepatocyte nuclear factor 4 homolog [Drosophila melanogaster]	<a href="#">139</a>	139	18%	4e-32	26%	
<a href="#">AAM76194.1</a>	RE08410p [Drosophila melanogaster]	<a href="#">139</a>	139	19%	5e-32	27%	

# PSI-Blast

# PSI-Blast

- PSI-Blast: Position-specific iterated Blast
- Position specific scoring matrices (PSSMs) (Πίνακες αντικατάστασης θέσης)
- Altschul et al., 1997
- <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC146917/pdf/253389.pdf>
- Η αναζήτηση μακρινών ομολόγων σε Β.Δ. είναι πιο ευαίσθητη με τη χρήση αυτών των πινάκων.
- Για ομόλογες ακολουθίες το PSI-Blast βρίσκει μέχρι και 3 φορές περισσότερες μακρινές ομόλογες ακολουθίες (ομοιότητα < 30%) σε σχέση με το Blastp.

# PSI-Blast

- Σε μια ακολουθία οι διάφορες θέσεις δεν είναι το ίδιο συντηρημένες/ευέλικτες λόγω δομικών/λειτουργικών περιορισμών.
- Χρησιμοποιώντας ομόλογες ακολουθίες από τον ίδιο ή άλλους οργανισμούς κατανοούμε την ευελιξία κάθε θέσης μιας ακολουθίας.
- Π.χ. Σε μια ακολουθία A, στην θέση 123 (ενεργό κέντρο ενζύμου) βλέπουμε ένα μόνο αμινοξύ.
- Σε μια πολλαπλή στοίχιση της A με ομόλογες ακολουθίες βλέπουμε για την ίδια θέση (123) ποιά άλλα αμινοξέα επιτρέπονται και σε τί συχνότητες.
- Το PSSM χρησιμοποιεί αυτή την πληροφορία για να αναζητήσει μακρινά ομόλογα σε μια Β.Δ.

# PSSM

- Αρχικά γίνεται πολλαπλή στοίχιση των ακολουθιών

## A. Sequence alignment<sup>a</sup>

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

- Στη συνέχεια, για ακολουθία μήκους L δημιουργείται πίνακας:
  - L X 4 (nucleotides)
  - L X 20 (proteins)

# PSSM

- Γίνεται καταμέτρηση των συχνοτήτων των χαρακτήρων για την κάθε θέση.

## A. Sequence alignment<sup>a</sup>

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

## B. Table of occurrences<sup>a</sup>

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



# PSSM

- Ακολουθεί μια σειρά μετασχηματισμών
  - Συντελεστής βαρύτητας της κάθε ακολουθίας με βάση την ομοιότητά της με άλλες.
  - Pseudocounts
  - Λαμβάνεται υπόψη η συχνότητα υποβάθρου του κάθε χαρακτήρα
    - Υπολογισμός των odds (παρατηρούμενη συχνότητα / συχνότητα υποβάθρου).
  - Log-odds
- Ο πίνακας αυτός χρησιμοποιείται για τοπική στοίχιση με ακολουθίες σε μια Β.Δ. (αντικαθιστά την ακολουθία επερώτησης).

F. Position-specific scoring matrix: Log-odds form ( $B = 0.1$ )<sup>c,d</sup>

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








# PSI-Blast

- Πρώτο στάδιο:
  - Blast με την ακολουθία επερώτησης σε μια Β.Δ. ( $E < 0.001$  default).
  - Οι τοπικές στοίχισεις που βρέθηκαν ( $E\text{-value} < \text{cutoff}$ ) χρησιμοποιούνται για τη δημιουργία μιας πολλαπλής στοίχισης  $M$  με σημείο αναφοράς την ακολουθία επερώτησης ( $L$  θέσεις).
    - Δεν επιτρέπονται κενά στην ακολουθία επερώτησης.
    - Αυτή η πολλαπλή στοίχιση (ακολουθία - σημείο αναφοράς) διαφέρει από τις τυπικές πολλαπλές στοίχισεις
  - Απαλοιφή ακολουθιών με πολύ μεγάλη ομοιότητα.
  - Δημιουργία PSSM.

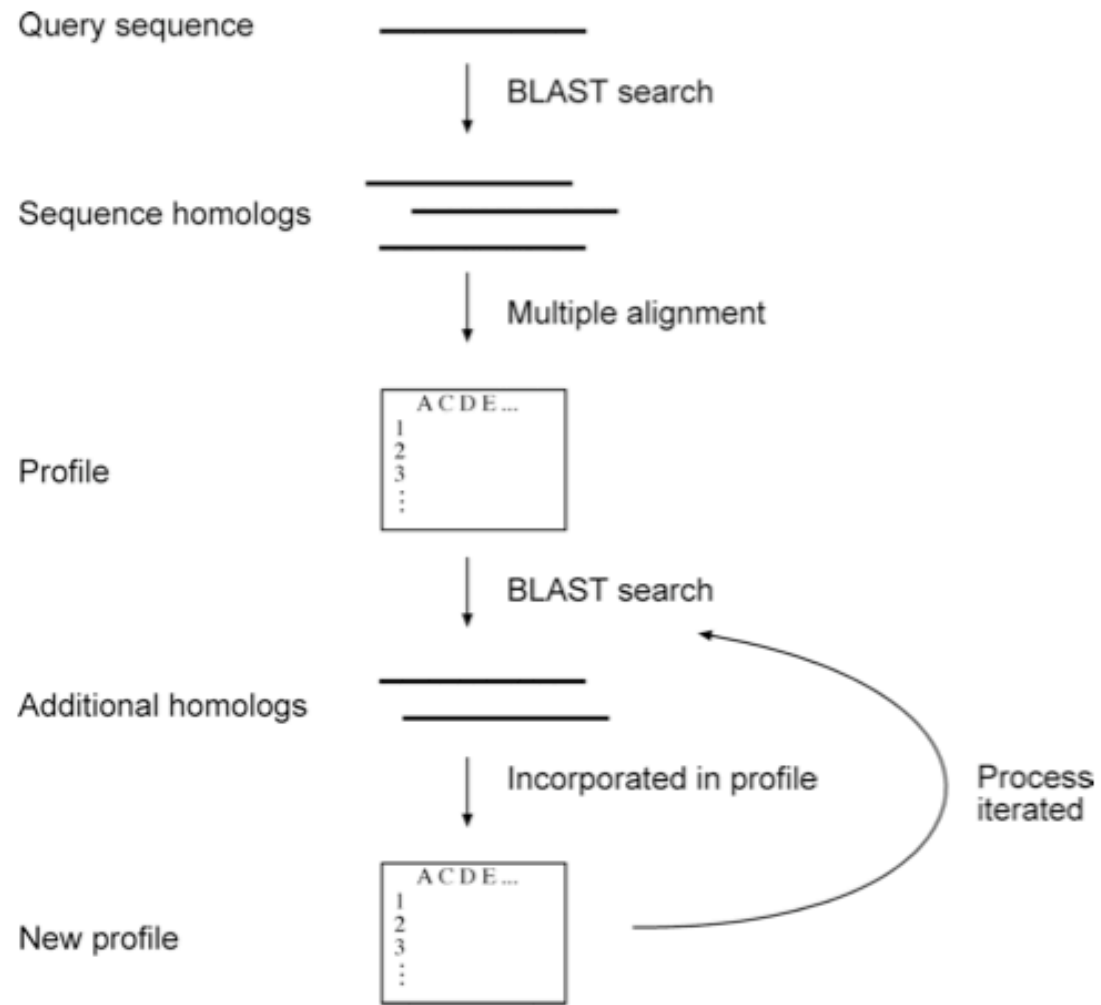
# PSI-Blast

- Δεύτερο στάδιο:
  - Νέα αναζήτηση στη Β.Δ. με το PSSM αντί της αρχικής ακολουθίας επερώτησης.
  - Οι νέες ακολουθίες που βρέθηκαν και ξεπερνούν το κατώφλι E-value ανανεώνουν την πολλαπλή στοίχιση και δημιουργείται ένα νέο PSSM.
- Η διαδικασία επαναλαμβάνεται μέχρι να μη βρεθούν νέες ακολουθίες με Evalue < τιμή κατωφλίου (convergence).
- Συνήθως, 3-5 κύκλοι αρκούν για να βρεθούν τα περισσότερα μακρινά ομόλογα.

# PSI-Blast

<b>a</b> <u>Accession</u>	<u>Alignment</u>	<u>E-value</u>
P49789		
P49779		8e-27
P49775		6e-18
Q11066		3e-07
Q09344		4e-05
P49378		0.001
P32084		0.002

# PSI-Blast



# PSI-Blast

- Πριν κάνουμε PSI-Blast πρέπει να ξέρουμε τι αναζητάμε!!!
  - αναζητούμε ομόλογες πρωτεΐνες με την ίδια αρχιτεκτονική επικρατειών (domain architecture);
  - Αναζητούμε πρωτεΐνες που να περιλαμβάνουν μια συγκεκριμένη περιοχή; Χρησιμοποιούμε μόνο αυτή την περιοχή στην αρχική αναζήτηση.
  - Αν η περιοχή αυτή είναι γνωστή επικράτεια που υπάρχει σε Β.Δ. Πρωτεϊνικών επικρατειών (π.χ. PFAM), τότε καλύτερα να χρησιμοποιήσουμε αυτές τις Β.Δ.
  - Κάποιες περιοχές/επικράτειες συναντώνται σε πολλές πρωτεΐνες.
    - Προσοχή στην αναζήτηση όταν υπάρχουν τέτοιες περιοχές
  - Αν ξεκινήσουμε με άλλη ομόλογη ακολουθία επερώτησης δεν είναι σίγουρο ότι θα φτάσουμε στο ίδιο αποτέλεσμα!
  - Προσοχή ποιές ακολουθίες συμπεριλαμβάνουμε στο PSSM. Αν εισέλθουν λάθος ακολουθίες, το λάθος θα ανατροφοδοτείται σε κάθε κύκλο (profile drift)

# Επικράτειες (Domains)

- Κάποιες επικράτειες συνδυάζονται πολύ συχνά με άλλες, στην ίδια πρωτεΐνη.
- <http://genome.cshlp.org/content/18/3/449.full>

## Evolution of protein domain promiscuity in eukaryotes

Click on image to view larger version.

Click on table to view larger version.

**Table 2.** The 10 most promiscuous domains in animals, fungi, and plants

Domain	Average promiscuity ( $\pi$ )	Most frequent bigram partner	No. of occurrences
Animals			
PH (smart00233)	972.18	SH3 (smart00326)	96
PDZ (smart00228)	675.6	SH3 (smart00326)	166
SH3 (smart00326)	556.45	GuKc (smart00072)	197
C1 (smart00109)	479.35	C2 (smart00239)	85
PHD (smart00249)	464.83	BROMO (smart00297)	123
RING (smart00184)	441.26	BBOX (smart00336)	128
TyrKc (smart00219)	413.74	FN3 (smart00060)	223
EGF_CA (smart00179)	397.07	CUB (smart00042)	55
SAM (smart00454)	371.45	TyrKc (smart00219)	138
EGF (smart00181)	353.07	LamG (smart00282)	155

# Επικράτειες και αναζήτηση σε Β.Δ.

**Family: zf-C2H2 (PF00096)**

238 architectures   31268 sequences   2 interactions   728 species   133 structures

**Summary**

- Domain organisation
- Alignments
- HMM logo
- Trees
- Curation & models
- Species
- Interactions
- Structures

**Jump to...**

### Domain organisation

Below is a listing of the unique domain organisations or architectures in which this domain is found. [More...](#)

**There are 3344 sequences with the following architecture: zf-C2H2**  
[ADR1\\_YEAST](#) [Saccharomyces cerevisiae (Baker's yeast)] Regulatory protein ADR1 (1323 residues)  
  
[Show](#) all sequences with this architecture.

**There are 1911 sequences with the following architecture: zf-C2H2 x 2**  
[AEF1\\_DROME](#) [Drosophila melanogaster (Fruit fly)] Adult enhancer factor 1 (308 residues)  
  
[Show](#) all sequences with this architecture.

**There are 638 sequences with the following architecture: zf-C2H2 x 3**  
[ODD\\_DROME](#) [Drosophila melanogaster (Fruit fly)] Protein odd-skipped (392 residues)  
  
[Show](#) all sequences with this architecture.

**There are 485 sequences with the following architecture: zf-AD, zf-C2H2**  
[ZN276\\_HUMAN](#) [Homo sapiens (Human)] Zinc finger protein 276 (614 residues)  
  
[Show](#) all sequences with this architecture.

**There are 388 sequences with the following architecture: zf-C2H2 x 4**  
[ESCA\\_DROME](#) [Drosophila melanogaster (Fruit fly)] Protein escargot (470 residues)  
  
[Show](#) all sequences with this architecture.

**There are 262 sequences with the following architecture: zf-C2H2 x 5**  
[CF2\\_DROME](#) [Drosophila melanogaster (Fruit fly)] Chorion transcription factor Cf2 (510 residues)  
  
[Show](#) all sequences with this architecture.

**There are 239 sequences with the following architecture: zf-C2H2 x 6**  
[Q9H7U2\\_HUMAN](#) [Homo sapiens (Human)] cDNA FLJ14260 fis, clone PLACE1001118, weakly similar to ZINC FINGER PROTEIN 135 (262 residues)  
  
[Show](#) all sequences with this architecture.



# Χρησιμοποιώντας το PSI-Blast

blastn blastp **blastx** tblastn tblastx

BLASTP programs search protein databases using a protein query. [more...](#) [Reset page](#) [Bookmark](#)

### Enter Query Sequence

Enter accession number, gi, or FASTA sequence [Clear](#)

```
SKKNSLALSILTADQMVSALLDAEPPILYSEYDPTTRPFSEASMMGLLTNLADRELVHMINW
AKRVPGFVDLTLHDQVHLLECAWLEILMIGLVWRSMHPGKLLFAPNLLDRNQGKCEVG
MVEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLD
KITDTLIHLMAKAGLTQQQHQRQAQLLLSHIRHMSNKGMEHLYSMKCKNVVPLYDLL
LEM LDAHRLHAPTSRGGASVEETDQSHLATAGSTSSHSLQKYYITGEAEGFPATV
```

Query subrange [From](#)   
[To](#)

Or, upload file  no file selected

Job Title   
Enter a descriptive title for your BLAST search

Align two or more sequences

### Choose Search Set

Database

Organism Optional   Exclude   
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Exclude Optional  Models (XM/XP)  Uncultured/environmental sample sequences

Entrez Query Optional   
Enter an Entrez query to limit search

### Program Selection

Algorithm  blastp (protein-protein BLAST)  PSI-BLAST (Position-Specific Iterated BLAST)  PHI-BLAST (Pattern Hit Initiated BLAST)  
Choose a BLAST algorithm

Search database **Swissprot protein sequences(swissprot)** using **PSI-BLAST (Position-Specific Iterated BLAST)**  
 Show results in a new window

[Algorithm parameters](#) **Note: Parameter values that differ from the default are highlighted in yellow and marked with**

# Χρησιμοποιώντας το PSI-Blast

Algorithm parameters Note: Parameter values that differ from the default are highlighted in yellow and marked with ♦ sign

### General Parameters

Max target sequences: 500  
Select the maximum number of aligned sequences to display

Short queries:  Automatically adjust parameters for short input sequences

Expect threshold: ♦ 1e-3

Word size: 3

Max matches in a query range: 0

### Scoring Parameters

Matrix: BLOSUM62

Gap Costs: Existence: 11 Extension: 1

Compositional adjustments: Conditional compositional score matrix adjustment

### Filters and Masking

Filter: ♦  Low complexity regions

Mask:  Mask for lookup table only  
 Mask lower case letters

### PSI/PHI BLAST

Upload PSSM: Optional  no file selected

PSI-BLAST Threshold: 1e-3

Pseudocount: 0

**BLAST** Search database **Swissprot protein sequences(swissprot)** using **PSI-BLAST (Position-Specific Iterated BLAST)**  
 Show results in a new window

# Χρησιμοποιώντας το PSI-Blast

NCBI/BLAST/blastp suite/ Formatting Results - CX8ZUS47011

[Edit and Resubmit](#) [Save Search Strategies](#) [Formatting options](#) [Download](#)

PSI blast Iteration 1

sp|P03372|ESR1\_HUMAN Estrogen receptor OS=Homo...

<b>Query ID</b>	lc 74714	<b>Database Name</b>	swissprot
<b>Description</b>	sp P03372 ESR1_HUMAN Estrogen receptor OS=Homo sapiens GN=ESR1 PE=1 SV=2	<b>Description</b>	Non-redundant SwissProt sequences
<b>Molecule type</b>	amino acid	<b>Program</b>	BLASTP 2.2.24+ <a href="#">Citation</a>
<b>Query Length</b>	595		

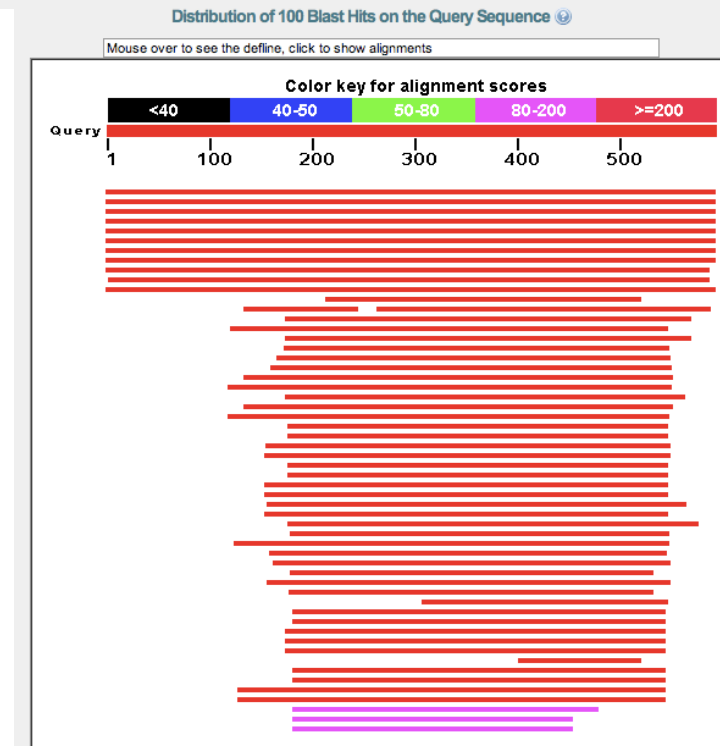
Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#) [Multiple alignment](#)

**Graphic Summary**

Show Conserved Domains

Putative conserved domains have been detected, click on the image below for detailed results.

Query seq.	1	100	200	300	400	500	595
Specific hits			NR_DBD_ER			NR_LBD_ER	
Superfamilies		Oest_recep superfamily	NR_DBD_like super			NR_LBD superfamily	



# Χρησιμοποιώντας το PSI-Blast

**Descriptions**

Legend for links to other resources: [U](#) UniGene [E](#) GEO [G](#) Gene [S](#) Structure [M](#) Map Viewer [P](#) PubChem BioAssay

**NEW** - alignment score below the threshold on the previous iteration

- alignment was checked on the previous iteration

Run PSI-Blast iteration 2 with max

**Sequences producing significant alignments with E-value BETTER than threshold**

	Accession	Description	Max score	Total score	Query coverage	E value	Links
<b>NEW</b>	<input checked="" type="checkbox"/> <a href="#">P03372.2</a>	RecName: Full=Estrogen receptor; Short=ER; AltName: Ful	<a href="#">1086</a>	1086	100%	0.0	<a href="#">G</a>
<b>NEW</b>	<input checked="" type="checkbox"/> <a href="#">Q53AD2.1</a>	RecName: Full=Estrogen receptor; Short=ER; AltName: Ful	<a href="#">1006</a>	1006	100%	0.0	<a href="#">G</a>
<b>NEW</b>	<input checked="" type="checkbox"/> <a href="#">P49884.3</a>	RecName: Full=Estrogen receptor; Short=ER; AltName: Ful	<a href="#">1003</a>	1003	100%	0.0	<a href="#">G</a>
<b>NEW</b>	<input checked="" type="checkbox"/> <a href="#">Q29040.1</a>	RecName: Full=Estrogen receptor; Short=ER; AltName: Ful	<a href="#">1001</a>	1001	100%	0.0	<a href="#">G</a>
<b>NEW</b>	<input checked="" type="checkbox"/> <a href="#">Q9TV98.1</a>	RecName: Full=Estrogen receptor; Short=ER; AltName: Ful	<a href="#">977</a>	977	100%	0.0	<a href="#">G</a>
<b>NEW</b>	<input checked="" type="checkbox"/> <a href="#">P19785.1</a>	RecName: Full=Estrogen receptor; Short=ER; AltName: Ful	<a href="#">967</a>	967	100%	0.0	<a href="#">G</a>
<b>NEW</b>	<input checked="" type="checkbox"/> <a href="#">P06211.1</a>	RecName: Full=Estrogen receptor; Short=ER; AltName: Ful	<a href="#">944</a>	944	100%	0.0	<a href="#">G</a>
<b>NEW</b>	<input checked="" type="checkbox"/> <a href="#">Q9QZJ5.1</a>	RecName: Full=Estrogen receptor; Short=ER; AltName: Ful	<a href="#">941</a>	941	100%	0.0	
<b>NEW</b>	<input checked="" type="checkbox"/> <a href="#">P06212.1</a>	RecName: Full=Estrogen receptor; Short=ER; AltName: Ful	<a href="#">860</a>	860	98%	0.0	<a href="#">G</a>
<b>NEW</b>	<input checked="" type="checkbox"/> <a href="#">Q91250.1</a>	RecName: Full=Estrogen receptor; Short=ER; AltName: Ful	<a href="#">852</a>	852	98%	0.0	<a href="#">G</a>

# Χρησιμοποιώντας το PSI-Blast

- Πράσινο σφαιρίδιο για ακολουθίες που είχαν βρεθεί σε προηγούμενο γύρο αναζήτησης.
- Μπορούμε να επιλέξουμε τον αποκλεισμό κάποιων ακολουθιών

<input checked="" type="checkbox"/>	<a href="#">O16662.3</a>	RecName: Full=Nuclear hormone receptor family member	<a href="#">95.2</a>	95.2	33%	1e-18		
<input checked="" type="checkbox"/>	<a href="#">P41933.1</a>	RecName: Full=Nuclear hormone receptor family member	<a href="#">93.7</a>	93.7	21%	4e-18		
NEW	<input checked="" type="checkbox"/>	<a href="#">O57568.1</a>	RecName: Full=Thyroid hormone receptor alpha; AltName:	<a href="#">85.2</a>	85.2	21%	1e-15	
NEW	<input checked="" type="checkbox"/>	<a href="#">O17573.1</a>	RecName: Full=Nuclear hormone receptor family member	<a href="#">80.9</a>	80.9	49%	2e-14	
NEW	<input checked="" type="checkbox"/>	<a href="#">P79404.1</a>	RecName: Full=Mineralocorticoid receptor; Short=MR; AltN	<a href="#">80.6</a>	80.6	19%	3e-14	
NEW	<input checked="" type="checkbox"/>	<a href="#">O15466.2</a>	RecName: Full=Nuclear receptor subfamily 0 group B mem	<a href="#">78.3</a>	78.3	43%	2e-13	
<input checked="" type="checkbox"/>	<a href="#">P35547.1</a>	RecName: Full=Glucocorticoid receptor; Short=GR; AltNam	<a href="#">76.7</a>	76.7	11%	5e-13		
NEW	<input checked="" type="checkbox"/>	<a href="#">O9TXJ1.2</a>	RecName: Full=Nuclear hormone receptor family member	<a href="#">74.8</a>	74.8	11%	2e-12	
NEW	<input checked="" type="checkbox"/>	<a href="#">P70503.1</a>	RecName: Full=Nuclear receptor subfamily 0 group B mem	<a href="#">73.2</a>	73.2	34%	4e-12	
NEW	<input checked="" type="checkbox"/>	<a href="#">O9BG94.1</a>	RecName: Full=Nuclear receptor subfamily 0 group B mem	<a href="#">70.9</a>	70.9	33%	2e-11	
NEW	<input checked="" type="checkbox"/>	<a href="#">O61066.1</a>	RecName: Full=Nuclear receptor subfamily 0 group B mem	<a href="#">70.9</a>	70.9	43%	2e-11	
NEW	<input checked="" type="checkbox"/>	<a href="#">P79386.2</a>	RecName: Full=Nuclear receptor subfamily 0 group B mem	<a href="#">70.5</a>	70.5	38%	3e-11	
NEW	<input checked="" type="checkbox"/>	<a href="#">O8QHI2.1</a>	RecName: Full=Mineralocorticoid receptor; Short=MR; AltN	<a href="#">70.2</a>	70.2	17%	4e-11	
NEW	<input checked="" type="checkbox"/>	<a href="#">O9BG97.1</a>	RecName: Full=Nuclear receptor subfamily 0 group B mem	<a href="#">69.8</a>	69.8	33%	5e-11	
NEW	<input checked="" type="checkbox"/>	<a href="#">P51843.2</a>	RecName: Full=Nuclear receptor subfamily 0 group B mem	<a href="#">69.8</a>	69.8	33%	6e-11	
NEW	<input checked="" type="checkbox"/>	<a href="#">O9BG93.1</a>	RecName: Full=Nuclear receptor subfamily 0 group B mem	<a href="#">69.8</a>	69.8	33%	6e-11	
NEW	<input checked="" type="checkbox"/>	<a href="#">O17025.2</a>	RecName: Full=Nuclear hormone receptor family member	<a href="#">69.4</a>	69.4	11%	7e-11	
NEW	<input checked="" type="checkbox"/>	<a href="#">P97947.1</a>	RecName: Full=Nuclear receptor subfamily 0 group B mem	<a href="#">69.4</a>	69.4	37%	7e-11	
NEW	<input checked="" type="checkbox"/>	<a href="#">O16360.3</a>	RecName: Full=Nuclear hormone receptor family member	<a href="#">69.4</a>	69.4	33%	7e-11	
NEW	<input checked="" type="checkbox"/>	<a href="#">O9BG96.1</a>	RecName: Full=Nuclear receptor subfamily 0 group B mem	<a href="#">67.5</a>	67.5	33%	3e-10	
NEW	<input checked="" type="checkbox"/>	<a href="#">O02305.2</a>	RecName: Full=Nuclear hormone receptor family member	<a href="#">66.3</a>	66.3	45%	6e-10	
NEW	<input checked="" type="checkbox"/>	<a href="#">O17934.1</a>	RecName: Full=Nuclear hormone receptor family member	<a href="#">65.9</a>	65.9	52%	8e-10	
NEW	<input checked="" type="checkbox"/>	<a href="#">O62227.1</a>	RecName: Full=Nuclear receptor subfamily 0 group B mem	<a href="#">65.2</a>	65.2	33%	1e-09	
NEW	<input checked="" type="checkbox"/>	<a href="#">O45907.1</a>	RecName: Full=Nuclear hormone receptor family member	<a href="#">55.9</a>	55.9	34%	7e-07	
NEW	<input type="checkbox"/>	<a href="#">P20659.4</a>	RecName: Full=Histone-lysine N-methyltransferase trithora	<a href="#">55.1</a>	55.1	18%	1e-06	
NEW	<input type="checkbox"/>	<a href="#">O24742.1</a>	RecName: Full=Histone-lysine N-methyltransferase trithora	<a href="#">55.1</a>	55.1	19%	1e-06	
NEW	<input checked="" type="checkbox"/>	<a href="#">O16354.1</a>	RecName: Full=Nuclear hormone receptor family member	<a href="#">53.6</a>	53.6	33%	4e-06	
NEW	<input checked="" type="checkbox"/>	<a href="#">O9PUA8.1</a>	RecName: Full=Thyroid hormone receptor alpha; AltName:	<a href="#">45.9</a>	45.9	3%	9e-04	


Run PSI-Blast iteration 3 with max

# Χρησιμοποιώντας το PSI-Blast

**PSI blast Iteration 3**

**sp|P03372|ESR1\_HUMAN Estrogen receptor OS=Homo...**

<b>Query ID</b>	lc 59255	<b>Database Name</b>	swissprot
<b>Description</b>	sp P03372 ESR1_HUMAN Estrogen receptor OS=Homo sapiens GN=ESR1 PE=1 SV=2	<b>Description</b>	Non-redundant SwissProt sequences
<b>Molecule type</b>	amino acid	<b>Program</b>	BLASTP 2.2.24+ <a href="#">▶ Citation</a>
<b>Query Length</b>	595		

 **No new sequences were found above the 0.001 threshold**

# Χρησιμοποιώντας το PSI-Blast

- Αν περιλαμβάνονταν οι 2 μεθυλ-τρανσφεράσες...

<input checked="" type="checkbox"/>	<a href="#">Q24742.1</a>	RecName: Full=Histone-lysine N-methyltransferase trithora	<a href="#">85.2</a>	85.2	19%	1e-15		
<input checked="" type="checkbox"/>	<a href="#">P79404.1</a>	RecName: Full=Mineralocorticoid receptor; Short=MR; AltN	<a href="#">84.4</a>	84.4	19%	2e-15		
<input checked="" type="checkbox"/>	<a href="#">O57568.1</a>	RecName: Full=Thyroid hormone receptor alpha; AltName:	<a href="#">84.4</a>	84.4	21%	2e-15		
<input checked="" type="checkbox"/>	<a href="#">P35547.1</a>	RecName: Full=Glucocorticoid receptor; Short=GR; AltNam	<a href="#">79.4</a>	79.4	11%	7e-14		
<input checked="" type="checkbox"/>	<a href="#">Q9TXJ1.2</a>	RecName: Full=Nuclear hormone receptor family member	<a href="#">79.4</a>	79.4	54%	7e-14		
<input checked="" type="checkbox"/>	<a href="#">O8QHI2.1</a>	RecName: Full=Mineralocorticoid receptor; Short=MR; AltN	<a href="#">76.0</a>	76.0	17%	8e-13		
<input checked="" type="checkbox"/>	<a href="#">P20659.4</a>	RecName: Full=Histone-lysine N-methyltransferase trithora	<a href="#">74.4</a>	74.4	18%	2e-12		
<input checked="" type="checkbox"/>	<a href="#">Q9PUA8.1</a>	RecName: Full=Thyroid hormone receptor alpha; AltName:	<a href="#">47.1</a>	47.1	11%	4e-04		
	<input checked="" type="checkbox"/>	<a href="#">P16356.2</a>	RecName: Full=DNA-directed RNA polymerase II subunit R	<a href="#">45.5</a>	45.5	19%	0.001	

Run PSI-Blast iteration 4 with max


<input checked="" type="checkbox"/>	<a href="#">P20659.4</a>	RecName: Full=Histone-lysine N-methyltransferase trithora	<a href="#">71.0</a>	71.0	18%	2e-11		
<input checked="" type="checkbox"/>	<a href="#">P16356.2</a>	RecName: Full=DNA-directed RNA polymerase II subunit R	<a href="#">66.7</a>	66.7	19%	4e-10		
	<input checked="" type="checkbox"/>	<a href="#">P35074.2</a>	RecName: Full=DNA-directed RNA polymerase II subunit R	<a href="#">61.7</a>	61.7	19%	1e-08	
	<input checked="" type="checkbox"/>	<a href="#">P24928.2</a>	RecName: Full=DNA-directed RNA polymerase II subunit R	<a href="#">50.6</a>	50.6	18%	3e-05	
	<input checked="" type="checkbox"/>	<a href="#">P08775.3</a>	RecName: Full=DNA-directed RNA polymerase II subunit R	<a href="#">50.6</a>	50.6	18%	3e-05	
<input checked="" type="checkbox"/>	<a href="#">Q9PUA8.1</a>	RecName: Full=Thyroid hormone receptor alpha; AltName:	<a href="#">46.3</a>	46.3	11%	6e-04		

Run PSI-Blast iteration 5 with max

# Χρησιμοποιώντας το PSI-Blast

- Αποθήκευση αποτελεσμάτων

[Edit and Resubmit](#) [Save Search Strategies](#) [▶ Formatting options](#) [▽ Download](#)

Download						
<b>Alignment</b>				<b>Search Strategies</b>	<b>Bloseq</b>	<b>PssmWithParameters</b>
<a href="#">Text</a>	<a href="#">XML</a>	<a href="#">ASN.1</a>	<a href="#">Hit Table(text)</a>	<a href="#">Hit Table(csv)</a>	<a href="#">ASN.1</a>	<a href="#">ASN.1</a> 



# Blast

- Βρείτε την ακολουθία του Estrogen receptor alpha (σε μορφή FASTA) ως:
  - mRNA από την EMBL bank (accession number: X03635).
  - ως πρωτεΐνη από την Uniprot (accession number: P03372).

# Blast

Τα προγράμματα του Blast θα τα βρείτε στο:

<http://blast.ncbi.nlm.nih.gov/Blast.cgi>

Θέλετε να βρείτε τις ομόλογες πρωτεΐνες του ανθρώπινου estrogen receptor alpha (πρωτεΐνη) στη μύγα *Drosophila melanogaster*, χρησιμοποιώντας τη ΒΔ Swissprot.

Ποιό πρόγραμμα του Blast πρέπει να χρησιμοποιήσετε;

Οι παράμετροι της αναζήτησης:

- ΒΔ Swissprot
- οργανισμός: *Drosophila melanogaster*
- Expect threshold: 1e-5
- Low-complexity filtering

# Blast

- Δείτε τα συντηρημένα domains. Ποιά είναι;
- Ποιό είναι το καλύτερο blast hit; με ποιό score & Evalue; Τι πρωτεΐνη είναι;
- Για το καλύτερο blast hit, δείτε στην τοπική στοίχιση:
  - Identities
  - Positives
  - Low complexity regions

# Blast

- Βρείτε την πρωτεϊνική ακολουθία (σε μορφή FASTA) του καλύτερου blast hit και με αυτή κάνετε την αντίστροφη διαδικασία.
- Δηλαδή, blast έναντι της ΒΔ Swissprot, για τον οργανισμό Homo sapiens, χρησιμοποιώντας ως ακολουθία επερώτησης (query sequence) το καλύτερο Blast hit. Όλες οι προηγούμενες παράμετροι του blast παραμένουν ίδιες.
- Βρίσκετε ως νέο καλύτερο blast hit το estrogen receptor alpha; Είναι ανταποδοτικό το blast; Τι σημαίνει αυτό για τις εξελικτικές σχέσεις μεταξύ των δύο ακολουθιών;

# Blast

- Χρησιμοποιώντας ως ακολουθία επερώτησης το mRNA του estrogen receptor alpha από τον άνθρωπο (EMBL-bank accession: X03635), βρείτε αν υπάρχουν ομόλογες νουκλεοτιδικές ακολουθίες στη *Drosophila melanogaster*, χρησιμοποιώντας τη νουκλεοτιδική ΒΔ nucleotide collection (nr/nt).
- Ποιό πρόγραμμα του Blast πρέπει να χρησιμοποιήσετε;
- Παράμετροι του blast που θα κάνετε:
  - νουκλεοτιδική ΒΔ nucleotide collection (nr/nt)
  - Οργανισμό *Drosophila melanogaster*
  - Optimize for somewhat similar sequences
  - Expect threshold  $1e-5$
  - Filter low-complexity regions
- Βρέθηκαν ομόλογες νουκλεοτιδικές ακολουθίες στη *Drosophila*;
- Γιατί;

# Blast

- Ποιό άλλο πρόγραμμα του Blast πρέπει να χρησιμοποιήσετε, για να δείτε αν υπάρχουν ομόλογες πρωτεΐνες για το mRNA σας, στη *Drosophila melanogaster*;
- Παράμετροι του Blast.
  - Genetic code standard
  - Database: non-redundant protein sequences (nr)
  - Οργανισμός: *Drosophila melanogaster*
  - Expectation threshold  $1e-5$
  - Low complexity regions filtering
- Τι βρίσκετε;

# PSI-Blast

- Χρησιμοποιώντας την πρωτεΐνη του ανθρώπινου estrogen receptor alpha (P03372), κάνετε PSI-blast για να βρείτε όλες τις ομόλογες πρωτεΐνες στον άνθρωπο.
- Παράμετροι του PSI-Blast:
  - ΒΔ Swissprot
  - Homo sapiens
  - Expect threshold 1e-3
  - Low complexity region filtering
  - PSI-blast threshold 1e-3
- Χρησιμοποιήστε όλα τα hits πρώτου κύκλου για να δημιουργηθεί το PSSM και συνεχίστε στον δεύτερο κύκλο (iteration 2).
- Βρέθηκαν καινούργιες ακολουθίες;
- Θα ενσωματωθούν όλες στο νέο κύκλο αναζήτησης (iteration3) ;