Πολλαπλή αντιστοιχία αλληλουχίας

- Η πολλαπλή αντιστοιχία απεικονίζει τις σχέσεις μεταξύ δύο ή περισσοτέρων αλληλουχιών.

- Όταν οι αλληλουχίες που εμπλέκονται διαφέρουν, τα διατηρημένα (conserved) αμινοξέα και στις δύο αλληλουχίες συνήθως σχετίζονται με τη διατήρηση της σταθερότητας της δομής και της βιολογικής λειτουργίας της πρωτεΐνης.

- Αν ένα αμινοξύ διατηρείται σε μία οικογένεια αλληλουχιών, οι οποίες προέρχονται από διαφορετικά είδη, τότε αυτό δείχνει ότι το αμινοξύ παίζει ένα σημαντικό ρόλο στη δομή ή λειτουργία της πρωτεΐνης.

- Ένα τέτοιο αμινοξύ μπορεί να αναγνωριστεί με πολλαπλή αντιστοιχία.

Ορισμός

- Η πολλαπλή αντιστοιχία αλληλουχιών είναι ένας 2D πίνακας, όπου οι γραμμές αντιπροσωπεύουν τις διαφορετικές αλληλουχίες και οι στήλες τις θέσεις των αμινοξέων.

- Οι αλληλουχίες τοποθετούνται στο πίνακα έτσι ώστε α) τα πιο όμοια αμινοξέα να περιλαμβάνονται σε κάθετη εγγραφή, με τη χρήση κενών (-) και β) η τάξη των αμινοξέων σε κάθε αλληλουχία να διατηρείται.

- Για παράδειγμα, μια πολλαπλή αντιστοιχία για πέντε είδη αλληλουχιών (I-V) είναι η παρακάτω:

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
I	Y	D	G	G	A	V	-	E	A	L
II	Y	D	G	G	-	-	-	E	A	L
III	F	E	G	G	I	L	V	E	A	L
IV	F	D	-	G	I	L	V	Q	A	V
V	Y	E	G	G	A	V	V	Q	A	L

Κοινή-συνοπτική αλληλουχία

- Ο πίνακας αντιστοίχισης μπορεί να συνοψισθεί σε μια γραμμή (ψευδό-αλληλουχία [pseudo-sequence]) που προσθέτεται στο τέλος της αντιστοιχίας. Αυτή η ψευδό-αλληλουχία αποτελείται από σύμβολα, τα οποία συνοψίζουν τον χαρακτήρα της αντιστοιχίας σε κάθε κάθετη θέση, με τον παρακάτω τρόπο:

- 1) αν υπάρχει μόνο ένα σύμβολο αμινοξέως, τότε χρησιμοποιείται κεφαλαίο γράμμα,
- 2) αν η πλειοψηφία των συμβόλων είναι ένα γράμμα τότε χρησιμοποιείται μικρό γράμμα,
- 3) αν υπάρχει ίσος αριθμός διαφορετικών αμινοξέων τότε χρησιμοποιούνται όλα τα αμινοξέα.

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
I	Y	D	G	G	A	V	-	E	A	L
II	Y	D	G	G	-	-	-	E	A	L
III	F	E	G	G	I	L	V	E	A	L
IV	F	D	-	G	I	L	V	Q	A	V
V	Y	E	G	G	A	V	V	Q	A	L
	y	**d**	**G**	**G**	**A/I**	**V/L**	**V**	**e**	**A**	**l**

Διατηρημένα αμινοξέα – Παράδειγμα

- Ένα παράδειγμα πολλαπλής αντιστοιχίας μερικών αλληλουχιών της serine protease παρουσιάζεται παρακάτω.

- Αυτή η αντιστοιχία δείχνει ότι υπάρχον δύο κύριοι λόγοι για την διατήρηση αμινοξέων στις πρωτεΐνες:
  - 1) για τη διατήρηση της λειτουργίας και
  - 2) για τη διατήρηση της δομής.

```
SecStructure    ......................bBBBBb...----.bBBBBBb.....bBBb.aaa.bba
THRB_HUMAN      LESYIDGRIVEGSDAEIGMSPWQVMLFRKSP----QELLCGASLISDRWVLTAAHCLLYP
THRB_BOVIN      FESYIEGRIVEGQDAEVGLSPWQVMLFRKSP----QELLCGASLISDRWVLTAAHCLLYP
THRB_MOUSE      LDSYIDGRIVEGWDAEKGIAPWQVMLFRKSP----QELLCGASLISDRWVLTAAHCILYP
THRB_RAT        LDSYIDGRIVEGWDAEKGIAPWQVMLFRKSP----QELLCGASLISDRWVLTAAHCILYP
LFC_TACTR       SDSPRSPFIWNGNSTEIGQWPWQAGISRWLADHNMWFLQCGGSLLNEKWIVTAAHCVTYS
FA9_RAT         EPINDFTRVVGGENAKPGQIPWQVILNGEIE------AFCGGAIINEKWIVTAAHCLK--
FA9_RABIT       QSSDDFTRIVGGENAKPGQFPWQVLLNGKVE------AFCGGSIINEKWVVTAAHCIK--
FA9_PIG         QSSDDFIRIVGGENAKPGQFPWQVLLNGKID------AFCGGSIINEKWVVTAAHCIEP-
FA7_BOVIN       NGSKPQGRIVGGHVCPKGECPWQAMLKLNGA------LLCGGTLVGPAWVVSAAHCFER-
FA7_MOUSE       NSSSRQGRIVGGNVCPKGECPWQAVLKINGL------LLCGAVLLDARWIVTAAHCFDN-
FA7_RABIT       GASNPQGRIVGGKVCPKGECPWQAALMNGST------LLCGGSLLDTHWVVSAAHCFDK-
PRTC_HUMAN      QEDQVDPRLIDGKMTRRGDSPWQVVLLDSKK-----KLACGAVLIHPSWVLTAAHCMDE-
PRTC_RAT        EELELGPRIVNGTLTKQGDSPWQAILLDSKK-----KLACGGVLIHTSWVLTAAHCLES-
PRTC_MOUSE      DELEPDPRIVNGTLTKQGDSPWQAILLDSKK-----KLACGGVLIHTSWVLTAAHCVEG-
PSS8_HUMAN      CGVAPQARITGGSSAVAGQWPWQVSITYEGV------HVCGGSLVSEQWVLSAAHCFPS-
                       :  *        ***. :              *. ::   *:::****.
```

- Η ιστιδίνη (H) είναι απαραίτητη για τη λειτουργία αυτών των ενζύμων (πρωτεϊνών). Επειδή είναι απαραίτητη στη λειτουργία, διατηρείται σε κάθε μέλος της οικογενείας.
  - Η διατηρημένη ιστιδίνη είναι τοποθετημένη στην 6η θέση από τη δεξιά μεριά της αντιστοιχίας και όπως αναμένεται διατηρείται σε όλες τις αντιστοιχισμένες αλληλουχίες.
- Η κυστεϊνη (C) είναι απαραίτητη για τη διατήρηση μιας σταθερής τρισδιάστατης (3D) δομής.
  - Τα βασικά διατηρημένα στοιχεία της δομής είναι οι δύο διατηρημένες κυστεϊνες στην αντιστοιχία, οι οποίες είναι τοποθετημένες στην 5η και 20η θέση της αντιστοιχίας και διατηρούνται σε όλες τις αντιστοιχισμένες αλληλουχίες (δηλ. αυτές οι δυο συγκεκριμένες κυστεϊνες σε αυτές τις θέσεις είναι υπεύθυνες για τη διατήρηση της δομής της πρωτεΐνης).

Υπολογισμός – τεχνικές

- Οι τεχνικές πολλαπλής αντιστοιχίας αλληλουχίας χρειάζονται χρόνο επεξεργασίας και χώρο στη μνήμη ανάλογα με το μέγεθος των αλληλουχιών που συγκρίνονται: O(m1m2m3…ml), όπου το O είναι η τάξη του χρόνου που χρειάζεται ο αλγόριθμος και mi είναι το μήκος της αλληλουχίας i.
  - Οπότε, ο χρόνος που χρειάζεται για να υπολογιστεί μια αντιστοιχία αυξάνει εκθετικά με όσες πιο πολλές αλληλουχίες αντιστοιχίζονται.
- Έχουν δημιουργηθεί διάφορες τεχνικές για να μειωθεί ο χρόνος για την εύρεση καλών αντιστοιχιών, αυτές οι τεχνικές συμπεριλαμβάνουν:
- 1) αντιστοίχιση όλων των αλληλουχιών ανά ζεύγη,
- 2) αντιστοίχιση κάθε αλληλουχίας με μια συγκεκριμένη αλληλουχία, αντιστοίχιση αλληλουχιών σε αυθαίρετη τάξη, ή
- 3) αντιστοίχιση αλληλουχιών ακολουθώντας την τάξη διακλάδωσης ενός φυλογενετικού δέντρου.

Πρακτική αντιστοίχιση

- Η πολλαπλή αντιστοίχιση αλληλουχίας μπορεί να γίνει με τη χρήση πρακτικών μεθόδων.

- Τα πακέτα ανάλυσης αλληλουχιών προσφέρουν και οθόνες (editors) πρακτικής αντιστοίχισης. Για τον οπτικό εντοπισμό ομοιοτήτων μεταξύ των αλληλουχιών χρησιμοποιούνται διάφορα χρώματα.

- Η κατάλληλη επιλογή χρώματος προσφέρει ένα γρήγορο τρόπο για την αποκάλυψη των περιοχών που είναι σημαντικές για τη δομή ή τη λειτουργία μιας πρωτεΐνης (δηλ. μπορούν να απεικονισθούν εύκολα σημαντικά διατηρημένα αμινοξέα, ή ομάδες αμινοξέων, και ασυνήθιστες μεταλλάξεις). Στα πακέτα 3D γραφικών για πρωτεΐνες χρησιμοποιούνται τα παρακάτω χρώματα:

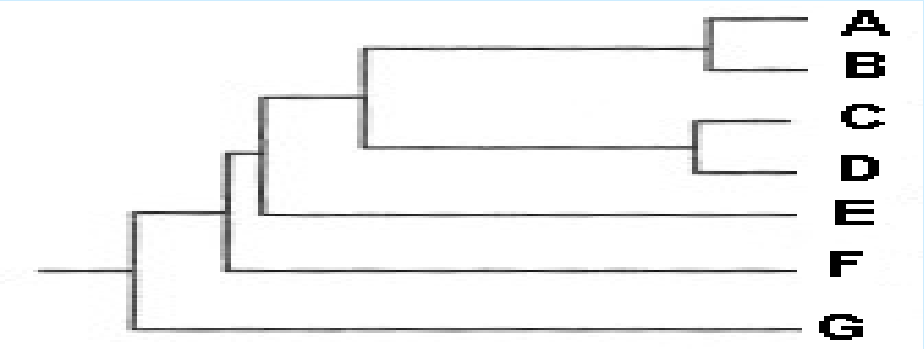
Residue	Property	Color
Asp, Glu	Acidic	red
His, Arg, Lys	Basic	blue
Ser, Thr, Asn, Gln	Polar neutral	green
Ala, Val, Leu, Ile, Met	Hydrophobic aliphatic	white
Phe, Tyr, Trp	Hydrophobic aromatic	purple
Pro, Gly	Special structural properties	brown
Cys	Disulphide bond former	yellow

Προοδευτική αντιστοιχία

- Η προοδευτική αντιστοιχία, πρώτα συμπεριλαμβάνει μια αρχική εκτίμηση για το πώς οι αλληλουχίες σχετίζονται (distance) χρησιμοποιώντας αντιστοιχίες ανά ζεύγη, και μετά δημιουργείται ένα καθοδηγητικό δέντρο.

- Με τη χρήση αυτού του καθοδηγητικού δέντρου, προσθέτονται αλληλουχίες προοδευτικά στην αντιστοιχία, ξεκινώντας με τις πιο σχετιζόμενες αλληλουχίες και τελειώνοντας με τις πιο απομακρυσμένες.

- Η διαδικασία είναι όπως παρακάτω:

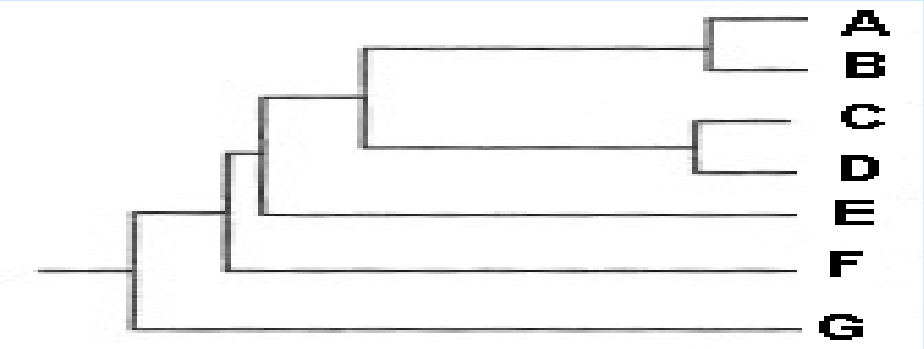


Αντίστοιχη αλληλουχία αντιστοιχίας
1. A με B ===> αντιστοιχία AB
2. C με D ===> αντιστοιχία CD
3. Αντιστοιχία AB με αντιστοιχία CD ===> ABCD
4. ABCD με E ===> ABCDE
5. ABCDE με F ===> ABCDEF
6. ABCDEF με G ===> ABCDEFG

•Αυτό το δέντρο μπορεί να θεωρηθεί ως ένας πιθανός τρόπος με τον οποίο οι αλληλουχίες εξελίχτηκαν (φυλογενετικό δέντρο- phylogenetic tree).

• Η A και B είναι πολύ σχετιζόμενες και έχουν απομακρυνθεί από έναν κοντινό κοινό πρόγονο, όπως συμβαίνει και με την C και D. Αυτές οι 4 αντιστοιχίες έχουν ένα κοινό πρόγονο και είναι πιο πολύ σχετιζόμενες μεταξύ τους απ' ότι με τις αλληλουχίες E, F και G.

- Η διαδικασία είναι όπως παρακάτω:



Αντίστοιχη αλληλουχία αντιστοιχίας
1. A με B ===> αντιστοιχία AB
2. C με D ===> αντιστοιχία CD
3. Αντιστοιχία AB με αντιστοιχία CD ===> ABCD
4. ABCD με E ===> ABCDE
5. ABCDE με F ===> ABCDEF
6. ABCDEF με G ===> ABCDEFG

Η αλληλουχία της αντιστοιχίας είναι η ακόλουθη:

-πρώτα, αντιστοιχίζονται οι πιο σχετιζόμενες αλληλουχίες A και B

-ακολουθούν οι αλληλουχίες C και D

-μετά αυτές οι δύο αντιστοιχίες αντιστοιχίζονται για να δημιουργηθεί μια αντιστοιχία από τέσσερις αλληλουχίες

-μετά οι αλληλουχίες E, F και G προσθέτονται διαδοχικά σ' αυτήν την αντιστοιχία

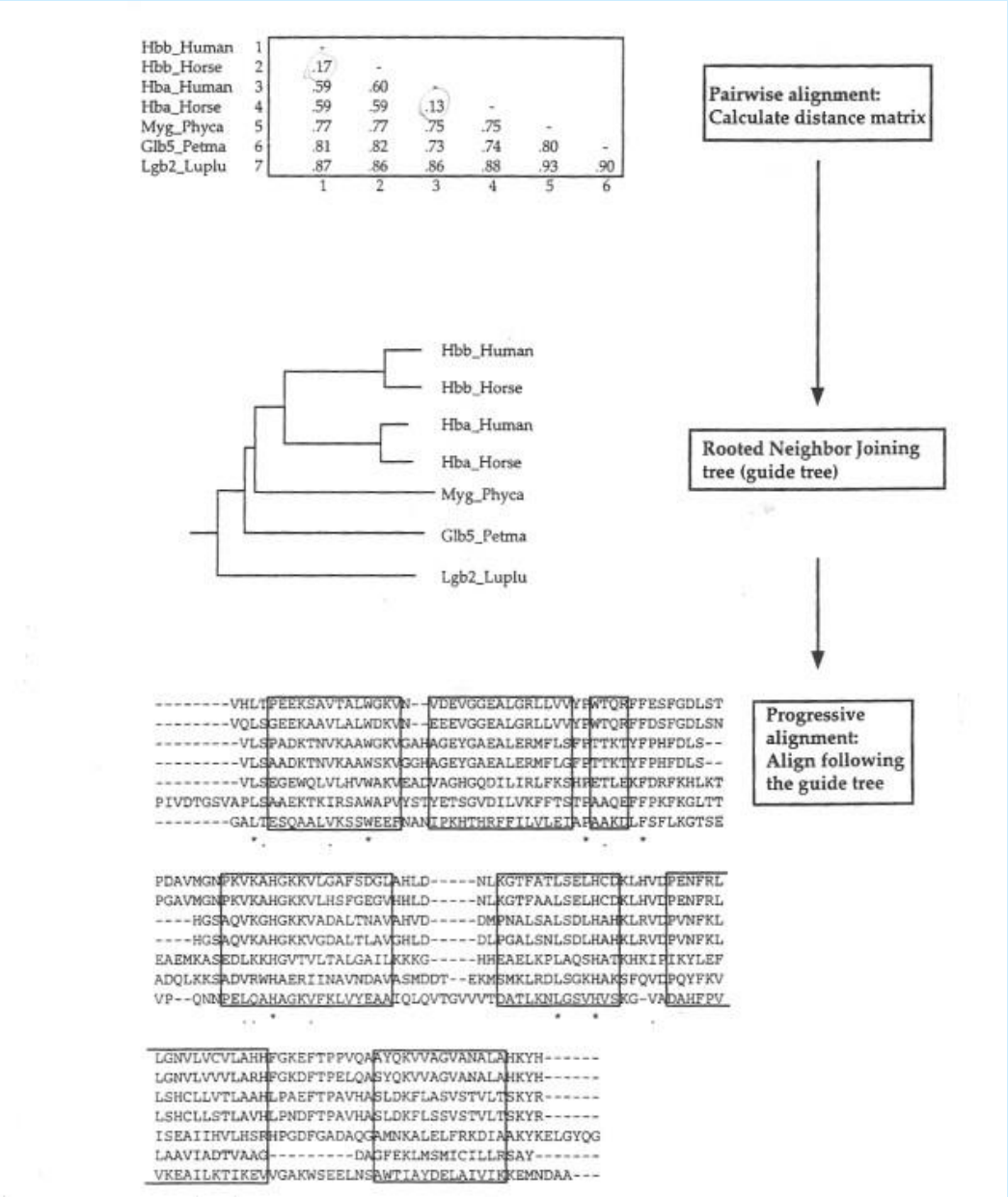
Όλες αυτές οι αντιστοιχίες πραγματοποιούνται με το δυναμικό προγραμματισμό.

ClustalW

- Το ClustalW είναι το πιο γνωστό πρόγραμμα υπολογιστή για την εκτέλεση μιας προοδευτικής αντιστοιχίας.

  Παρακάτω, παρουσιάζεται η πολλαπλή αντιστοίχιση επτά αλληλουχιών από globins, χρησιμοποιώντας το πρόγραμμα ClustalW.

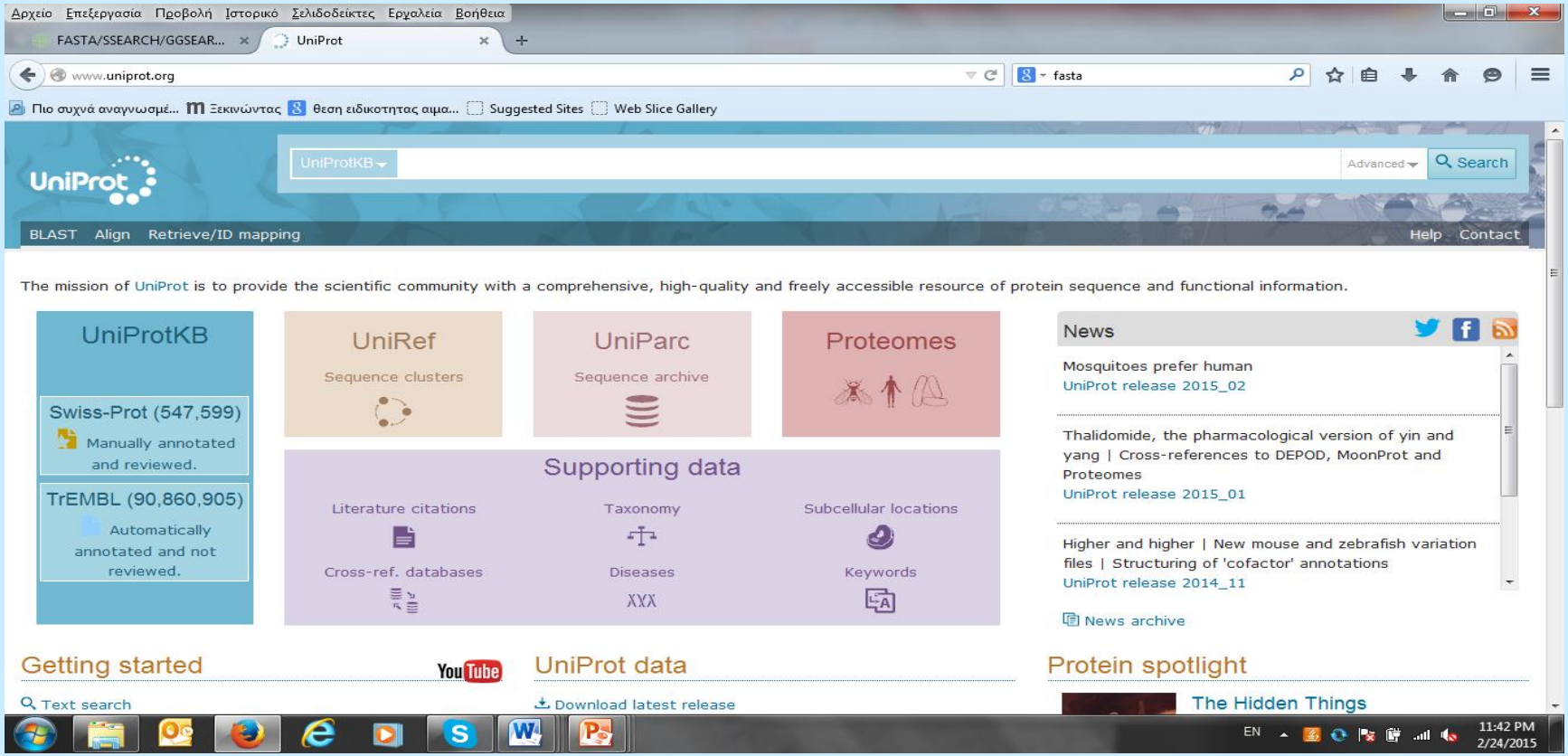
  Η ανάκτηση των αλληλουχιών έγινε από την SWISS-PROT, τη βάση δεδομένων με πρωτεΐνες.



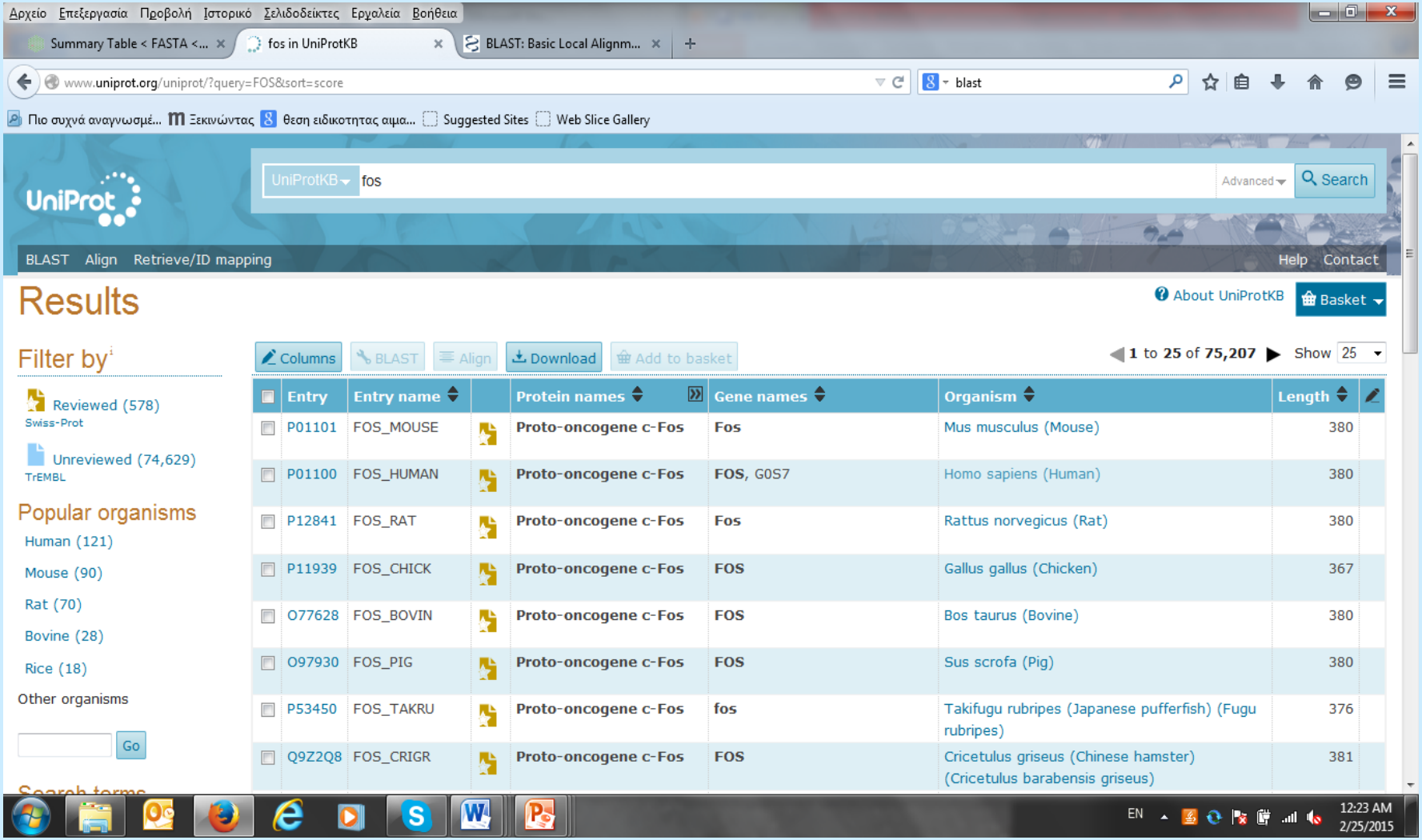
ΠΡΑΚΤΙΚΗ ΕΦΑΡΜΟΓΗ

Πολλαπλή αντιστοίχιση αλληλουχίας – ClustalW

- Θέλουμε να αντιστοιχίσουμε 5 πρωτεΐνες FOS από διαφορετικά είδη. Για να ανακτήσουμε τις αλληλουχίες των πρωτεϊνών, πρέπει να μπούμε στη βάση δεδομένων SWISS-PROT, οπότε ανοίγουμε το Internet Explorer και πληκτρολογούμε τη διεύθυνση: www.uniprot.org. Μετά πληκτρολογήστε FOS στο πλαίσιο παρακάτω και επιλέγουμε "Search".



- Τότε παρουσιάζεται μια λίστα των πρωτεϊνών FOS. Επιλέξτε την “FOS-HUMAN”.



Αρχείο Επεξεργασία Προβολή Ιστορικό Σελιδοδείκτες Εργαλεία Βοήθεια

Summary Table < FASTA <... | FOS - Proto-oncogene c-F... | BLAST: Basic Local Alignm...

www.uniprot.org/uniprot/P01100

blast

Πιο συχνά αναγνωσμέ... Ξεκινώντας Θεση ειδικοτητας αιμα... Suggested Sites Web Slice Gallery

UniProt

UniProtKB Advanced Search

BLAST Align Retrieve/ID mapping Help Contact

P01100 - FOS_HUMAN

Basket

Protein	**Proto-oncogene c-Fos**
Gene	**FOS**
Organism	*Homo sapiens (Human)*
Status	Reviewed - Annotation score: - Experimental evidence at protein level

Display None

- Function
- Names & Taxonomy
- Subcellular location
- Pathology & Biotech
- PTM / Processing
- Expression
- Interaction

BLAST Align Format Add to basket History

Feedback Help video

Function

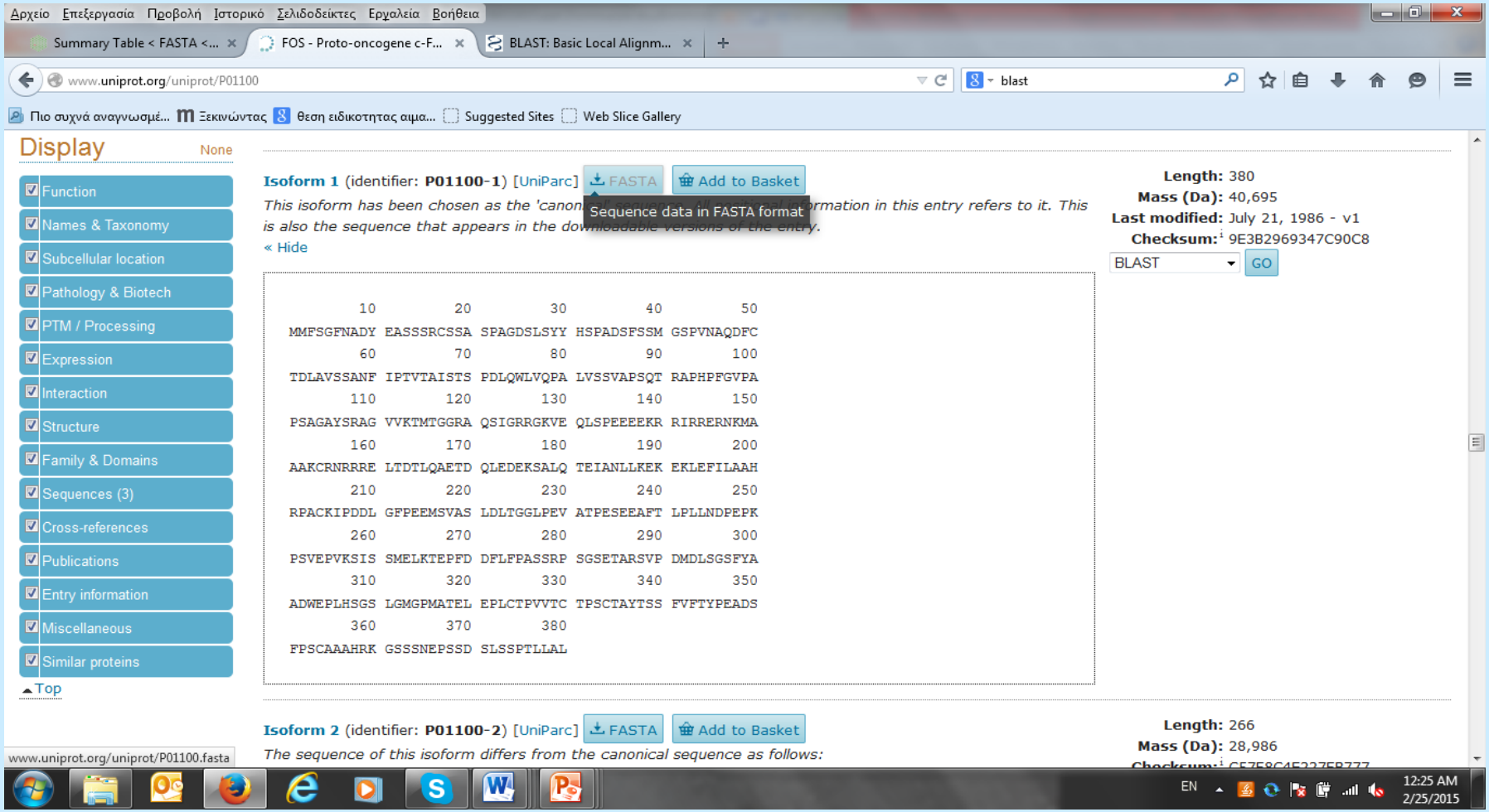
Nuclear phosphoprotein which forms a tight but non-covalently linked complex with the JUN/AP-1 transcription factor. In the heterodimer, FOS and JUN/AP-1 basic regions each seems to interact with symmetrical DNA half sites. On TGF-beta activation, forms a multimeric SMAD3/SMAD4/JUN/FOS complex at the AP1/SMAD-binding site to regulate TGF-beta-mediated signaling. Has a critical function in regulating the development of cells destined to form and maintain the skeleton. It is thought to have an important role in signal transduction, cell proliferation and differentiation. In growing cells, activates phospholipid synthesis, possibly by activating CDS1 and PI4K2A. This activity requires Tyr-dephosphorylation and association with the endoplasmic reticulum. 5 Publications

GO - Molecular function

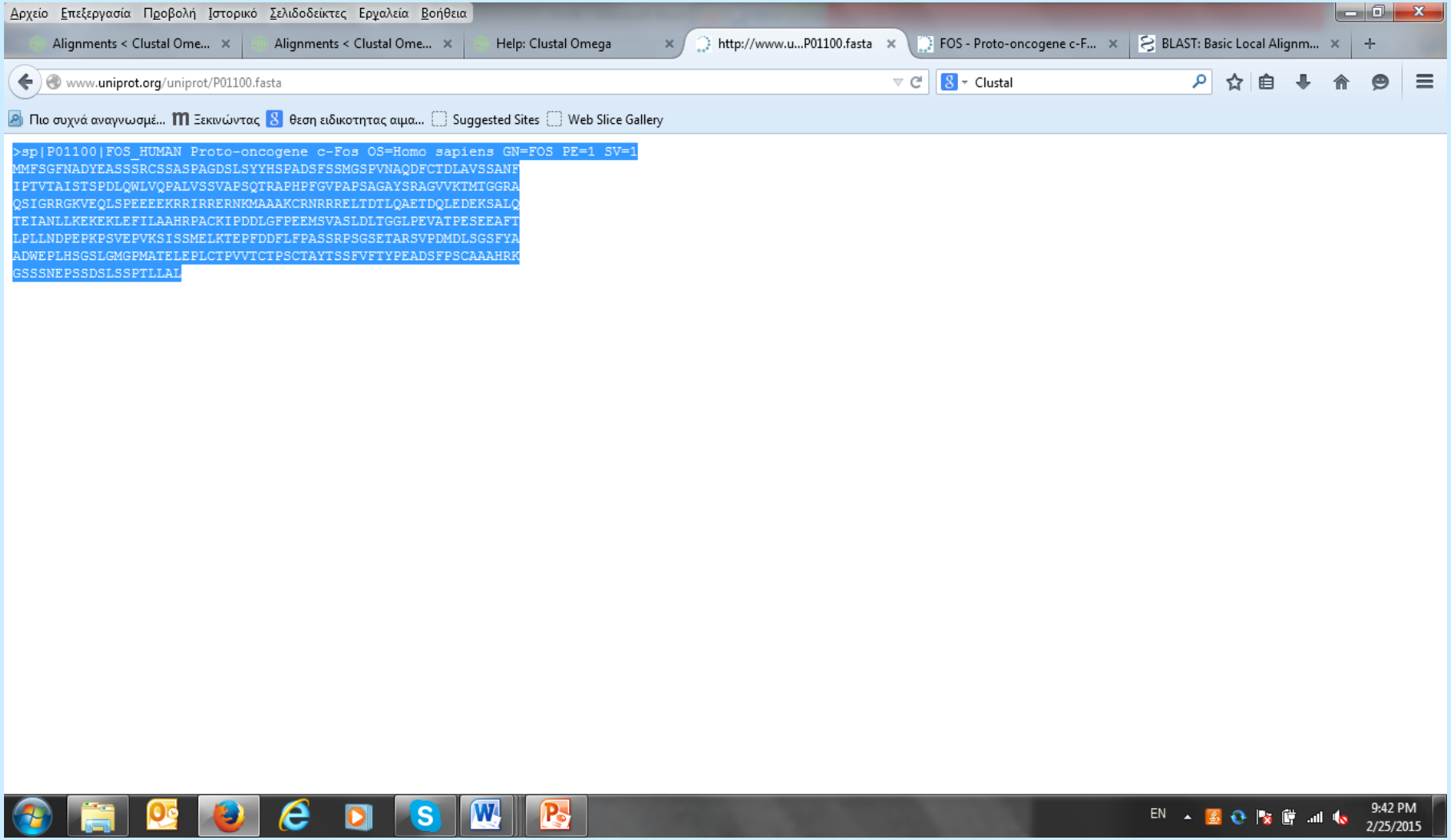
- double-stranded DNA binding Source: Ensembl
- RNA polymerase II core promoter sequence-specific DNA binding Source: Ensembl
- R-SMAD binding Source: BHF-UCL
- sequence-specific DNA binding transcription factor activity Source: BHF-UCL

EN 12:24 AM 2/25/2015

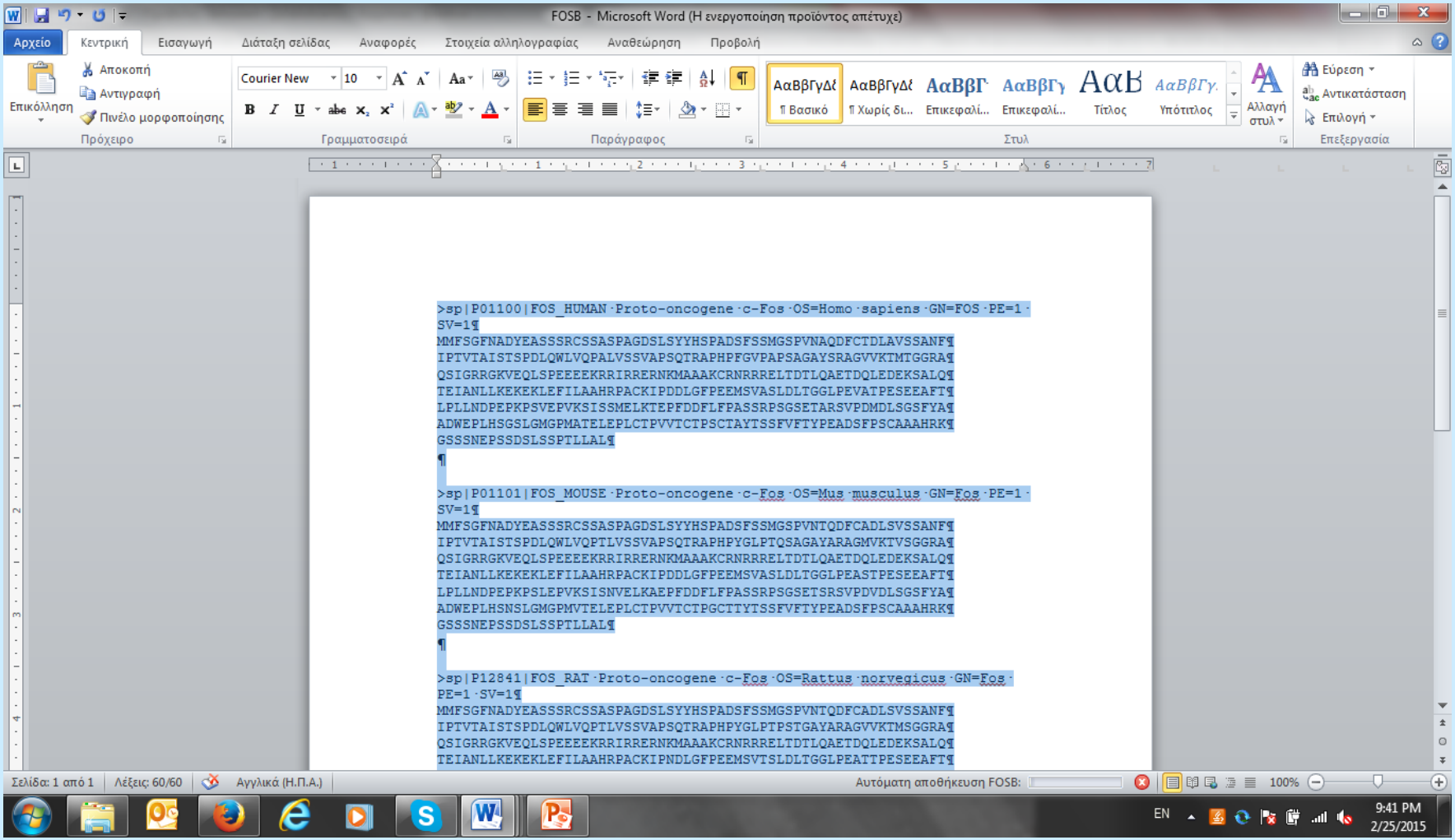
- Κάντε κλικ “fasta” για την ανάκτηση της αλληλουχίας.



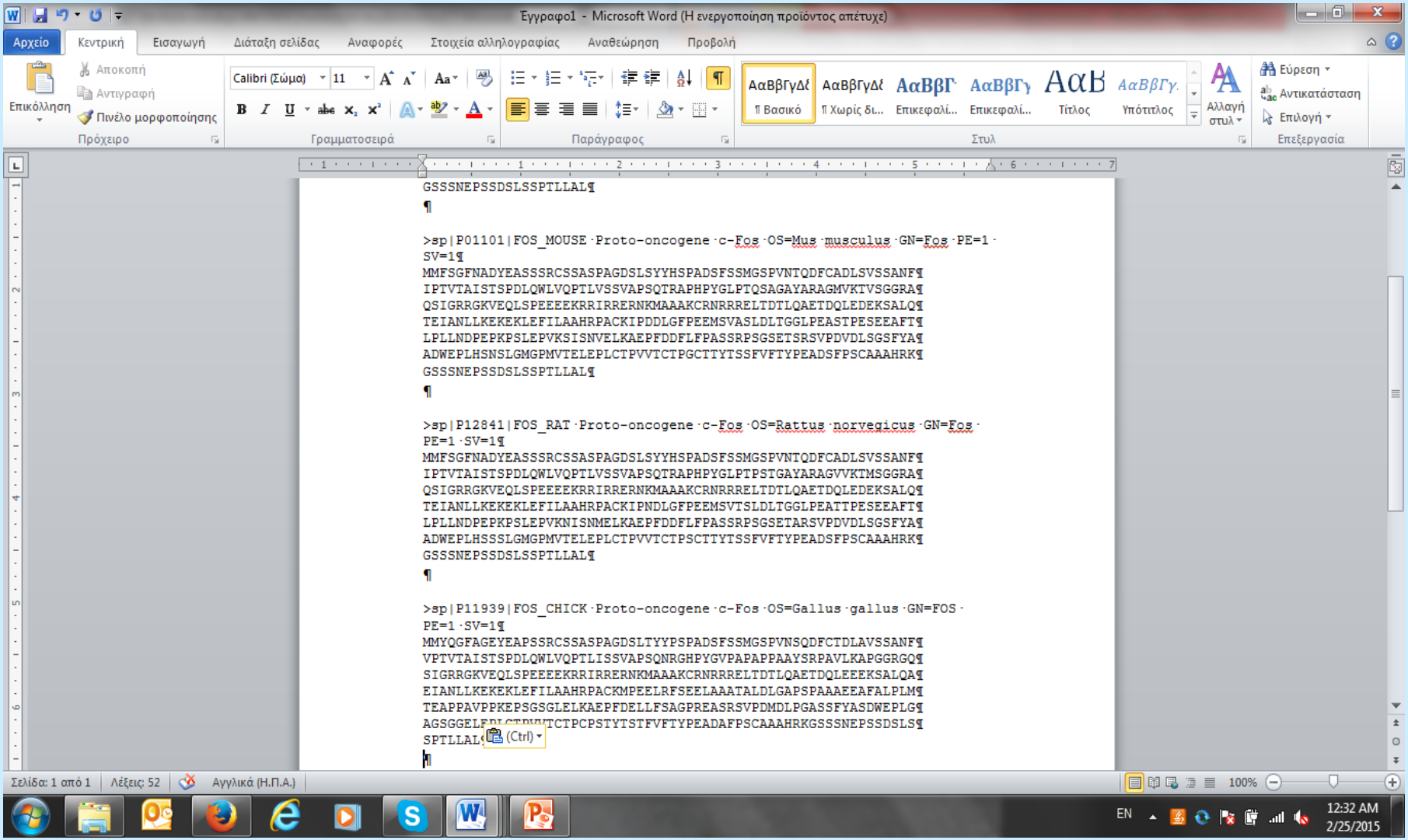
- Αντιγράψτε τη αλληλουχία .



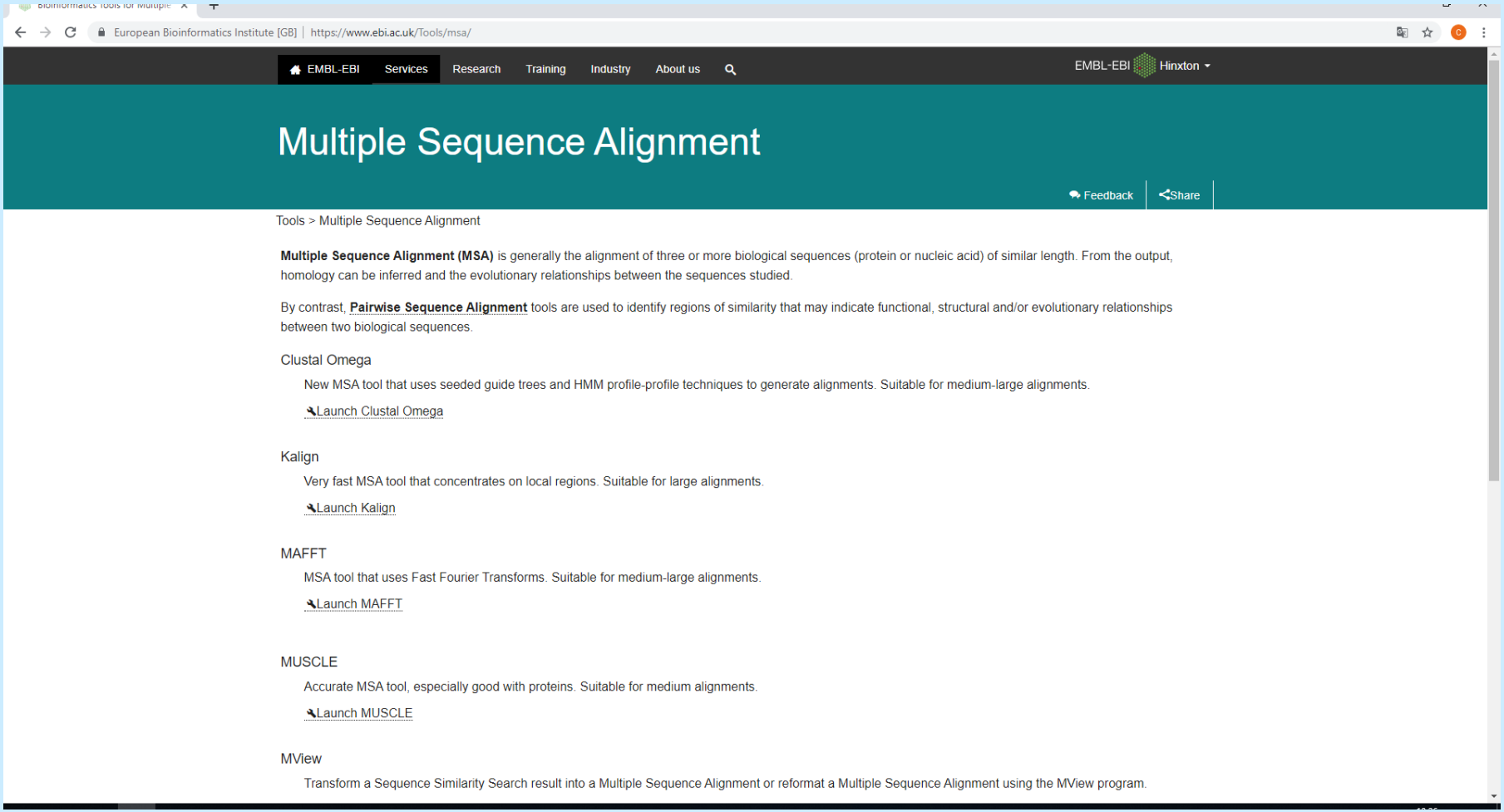
- Ανοίξτε στο Microsoft Word ή Notepad ένα αρχείο, επικολλήστε την αντιγραμμένη αλληλουχία και σώστε το αρχείο ως “FOSB”.



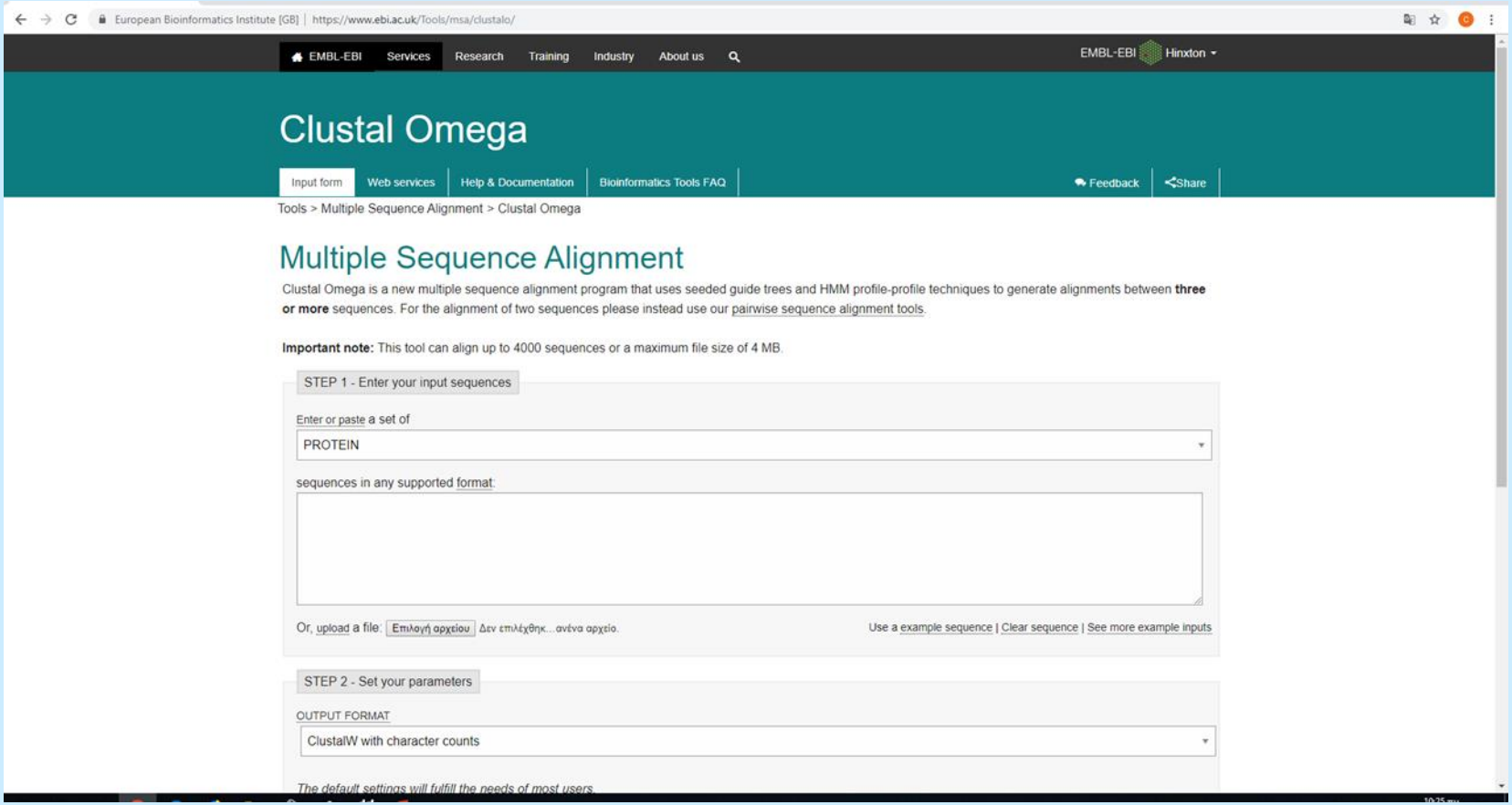
- Επαναλάβετε τα ίδια βήματα και για τις άλλες αλληλουχίες FOS_MOUSE, FOS_RAT, FOS_CHICK.



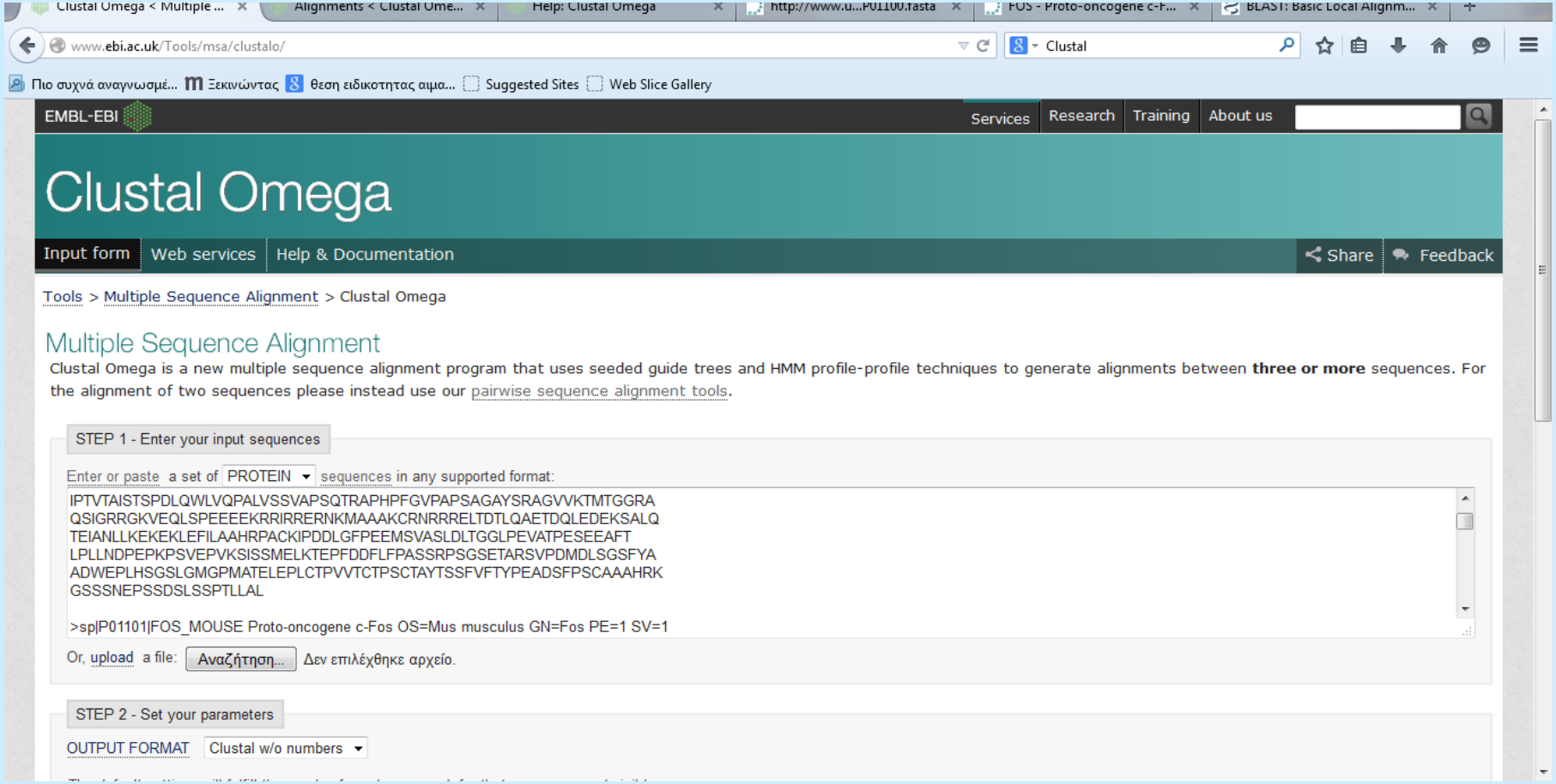
- Για την πρόσβαση στο πρόγραμμα Clustal, πληκτρολογήστε στο Internet Explorer τη διεύθυνση: http://www.ebi.ac.uk/Tools/msa/ και κάντε κλικ «Launch Clustal Omega».



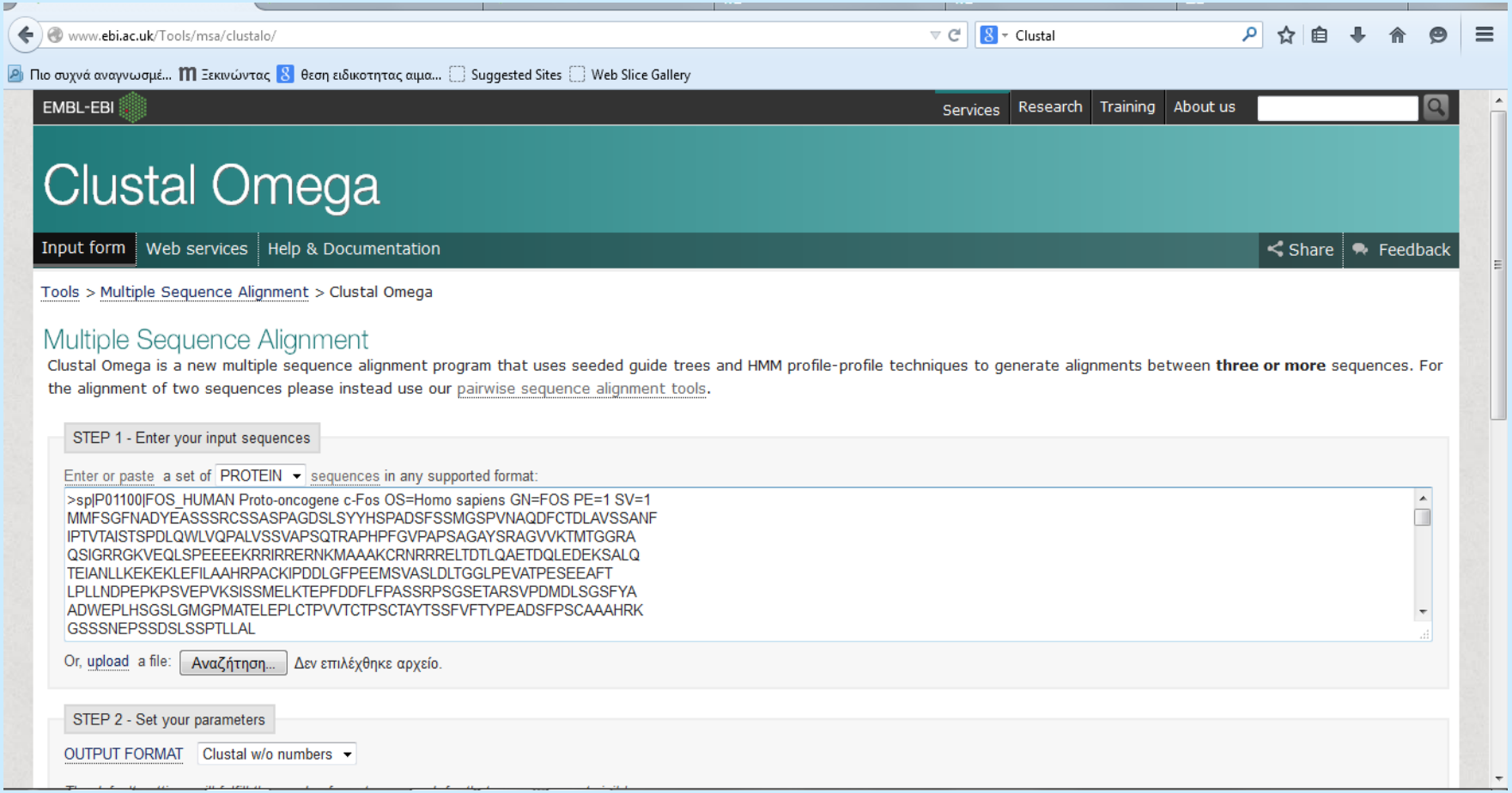
- Στο πλαίσιο παρακάτω βάλτε τις αλληλουχίες για πολλαπλή αντιστοίχιση.



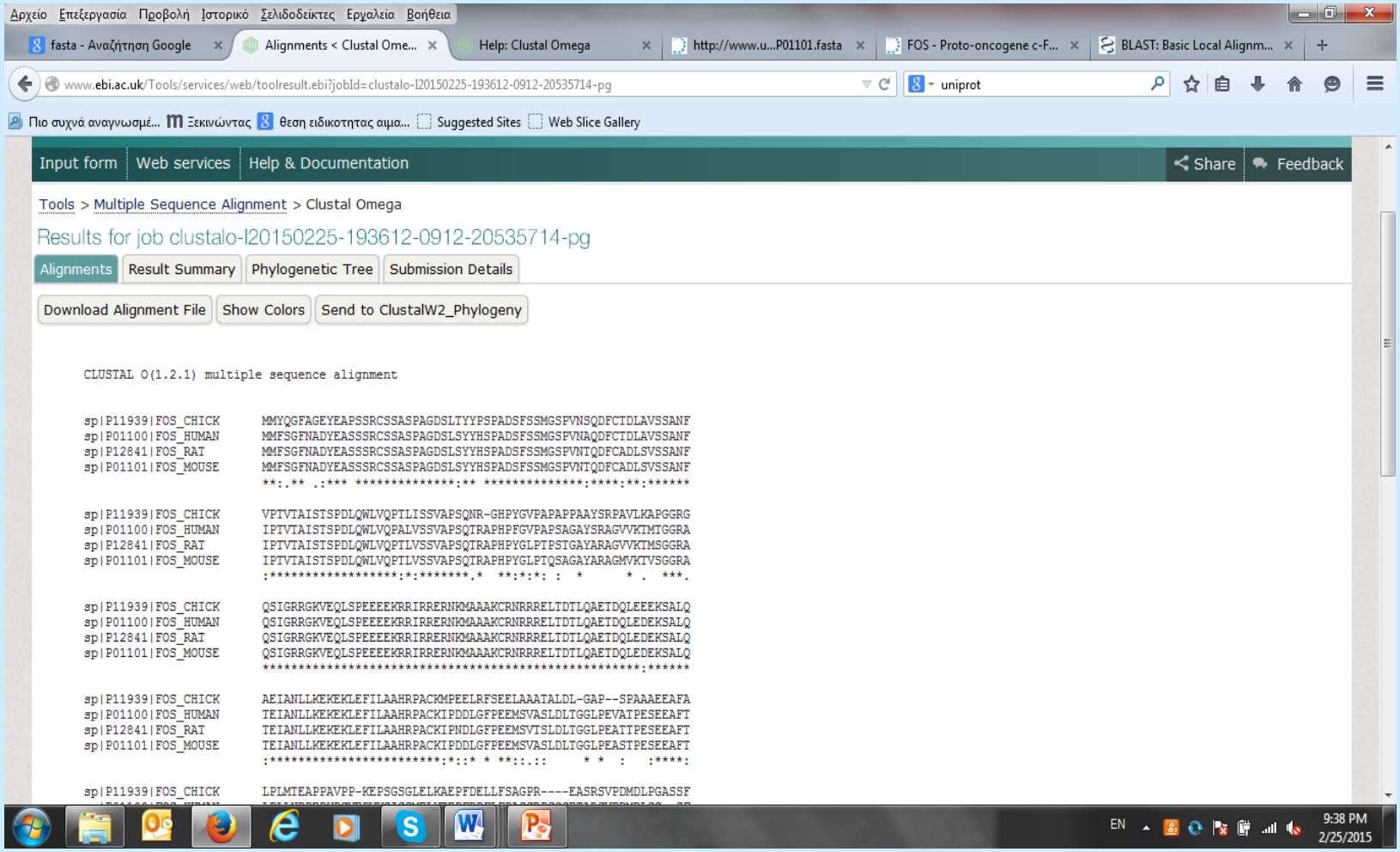
- Ανοίξτε το αρχείο FOSB στο Word ή Notepad, το οποίο περιέχει τις αλληλουχίες για αντιστοίχιση και επιλέξτε όλο το περιεχόμενο και αντιγράψτε το.



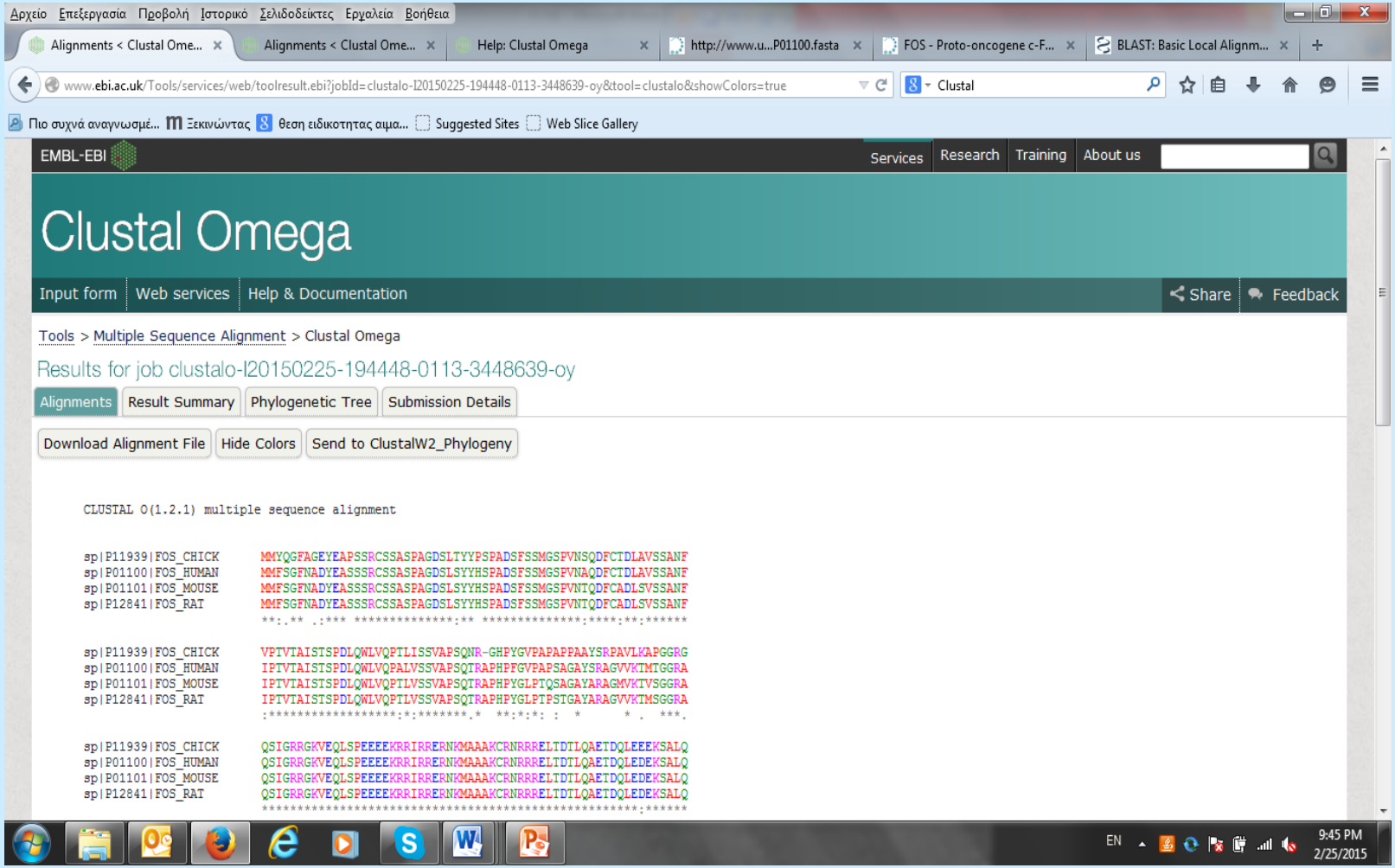
- Μετά την επικόλληση των αλληλουχιών στο πλαίσιο, προσδιορίστε τις παραμέτρους και επιλέξτε το “Submit”.



- Τότε εμφανίζονται τα αποτελέσματα από την πολλαπλή αντιστοίχιση



Πατήστε “show colors”



Αρχείο Επεξεργασία Προβολή Ιστορικό Σελιδοδείκτες Εργαλεία Βοήθεια

Result Summary < Clustal ...

www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobId=clustalo-I20150225-200510-0268-61257505-pg&analysis=summary

Πιο συχνά αναγνωσμέ... Ξεκινώντας θεση ειδικοτητας αιμα... Suggested Sites Web Slice Gallery

EMBL-EBI

Services | Research | Training | About us

Clustal Omega

Input form | Web services | Help & Documentation | Share | Feedback

Tools > Multiple Sequence Alignment > Clustal Omega

Results for job clustalo-I20150225-200510-0268-61257505-pg

Alignments | Result Summary | Phylogenetic Tree | Submission Details

Input Sequences

clustalo-I20150225-200510-0268-61257505-pg.input

Tool Output

clustalo-I20150225-200510-0268-61257505-pg.output

Alignment in CLUSTAL format

clustalo-I20150225-200510-0268-61257505-pg.clustal

Phylogenetic Tree

clustalo-I20150225-200510-0268-61257505-pg.ph

Percent Identity Matrix

clustalo-I20150225-200510-0268-61257505-pg.pim

Jalview

Start Jalview

EN 10:08 PM 2/25/2015

Αρχείο Επεξεργασία Προβολή Ιστορικό Σελιδοδείκτες Εργαλεία Βοήθεια

Result Summary < Clustal ...

www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobId=clustalo-I20150225-200510-0268-61257505-pg&analysis=summary

Πιο συχνά αναγνωσμέ... Ξεκινώντας θεση ειδικοτητας αιμα... Suggested Sites Web Slice Gallery

EMBL-EBI Services Research Training About us

Clustal Omega

Input form Web servi Share Feedback

Tools > Multiple Sequenc

Results for job clust

Alignments Result Summ

Input Sequences

clustalo-I20150225-20

Tool Output

clustalo-I20150225-20

Alignment in CLUST

clustalo-I20150225-20

Phylogenetic Tree

clustalo-I20150225-20

Percent Identity Mat

clustalo-I20150225-20

/Tools/services/rest/clustalo/result/clustalo-I20150225-200510-0268-61257505-pg/aln-clustal

File Edit Select View Format Colour Calculate Help

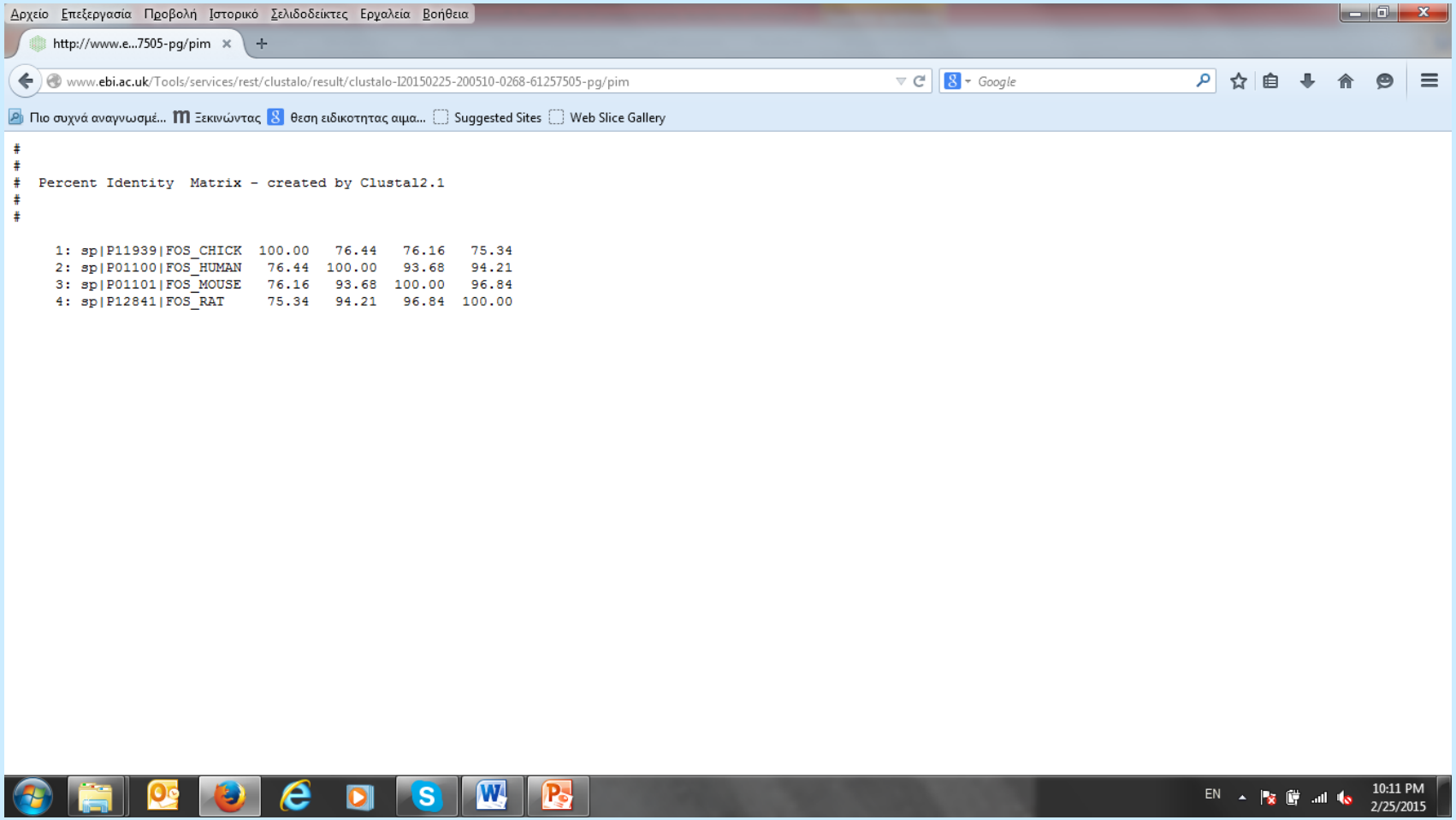
```
                                   10        20        30        40        50        60        70        80
sp|P11939|FOS_CHICK/1-367          MMYQGFAGEYEAPSSRCSSASPAGDSLTYYPSPADSFSSMGSPVNSQDFCTDLAVSSANFVPTVTAISTSPDLQWLVQPT
sp|P01100|FOS_HUMAN/1-380          MMFSGFNADYEASSSRCSSASPAGDSLSYYHSPADSFSSMGSPVNAQDFCTDLAVSSANFIPTVTAISTSPDLQWLVQPA
sp|P01101|FOS_MOUSE/1-380          MMFSGFNADYEASSSRCSSASPAGDSLSYYHSPADSFSSMGSPVNTQDFCADLSVSSANFIPTVTAISTSPDLQWLVQPT
sp|P12841|FOS_RAT/1-380            MMFSGFNADYEASSSRCSSASPAGDSLSYYHSPADSFSSMGSPVNTQDFCADLSVSSANFIPTVTAISTSPDLQWLVQPT

Conservation
Quality
Consensus
                                   MMFSGFNADYEASSSRCSSASPAGDSLSYYHSPADSFSSMGSPVNTQDFC+DL+VSSANFIPTVTAISTSPDLQWLVQPTL
```

Sequence position 37 F 100%

EN 10:09 PM 2/25/2015

Πατήστε “ Percent Identity Matrix”



Αρχείο Επεξεργασία Προβολή Ιστορικό Σελιδοδείκτες Εργαλεία Βοήθεια

Phylogenetic Tree < Clusta...

www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobId=clustalo-I20150225-200510-0268-61257505-pg&analysis=phylotree

Πιο συχνά αναγνωσμέ... Ξεκινώντας θεση ειδικοτητας αιμα... Suggested Sites Web Slice Gallery

Results for job clustalo-I20150225-200510-0268-61257505-pg

Alignments | Result Summary | Phylogenetic Tree | Submission Details

Phylogenetic Tree

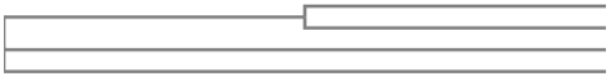
This is a Neighbour-joining tree without distance corrections.

Download Phylogenetic Tree File

```
(
(
sp|P11939|FOS_CHICK:0.20878,
sp|P01100|FOS_HUMAN:0.02684)
:0.01790,
sp|P01101|FOS_MOUSE:0.01505,
sp|P12841|FOS_RAT:0.01653);
```

Phylogram

Branch length: Cladogram Real



EMBL-EBI

News

Services: By topic

Research: Overview

Training: Overview

Industry: Overview

About us: Overview

EN 10:12 PM 2/25/2015

Επαναλάβετε την παραπάνω διαδικασία για 5 είδη της «aldose reductase»

Αρχείο Επεξεργασία Προβολή Ιστορικό Σελιδοδείκτες Εργαλεία Βοήθεια

Phylogenetic Tree < Clusta... | aldose reductase in UniPro...

www.uniprot.org/uniprot/?query=aldose+reductase&sort=score

Πιο συχνά αναγνωσμέ... Ξεκινώντας θεση ειδικοτητας αιμα... Suggested Sites Web Slice Gallery

UniProt

UniProtKB aldose reductase Advanced Search

BLAST Align Retrieve/ID mapping Help Contact

Results

Quote terms: "aldose reductase"

About UniProtKB Basket

Filter by

Reviewed (28)
Swiss-Prot

Unreviewed (437)
TrEMBL

Popular organisms

Human (7)

Rice (6)

Mouse (5)

Rat (5)

A. thaliana (4)

Other organisms

Go

Columns BLAST Align Download Add to basket

1 to 25 of 465 Show 25

Entry	Entry name	Protein names	Gene names	Organism	Length
Q91W30	Q91W30_RAT	Aldose reductase-like protein	**Akr1b8**, aldose reductase-like	Rattus norvegicus (Rat)	316
O60218	AK1BA_HUMAN	Aldo-keto reductase family 1 member...	**AKR1B10**, AKR1B11	Homo sapiens (Human)	316
Q5RJP0	ALD1_RAT	Aldose reductase-related protein 1	**Akr1b7**, Akr1b14	Rattus norvegicus (Rat)	316
P15121	ALDR_HUMAN	Aldose reductase	**AKR1B1**, ALDR1	Homo sapiens (Human)	316
P38715	GRE3_YEAST	NADPH-dependent aldose reductase GR...	**GRE3**, YHR104W	Saccharomyces cerevisiae (strain ATCC 204508 / S288c) (Baker's yeast)	327
P80276	ALDR_PIG	Aldose reductase	**AKR1B1**, ALR2	Sus scrofa (Pig)	316

EN 10:14 PM 2/25/2015