

Γονιδιωματική Εισαγωγή [2]

Τμήμα Γεωπονίας, Ιχθυολογίας
και Υδάτινου Περιβάλλοντος

Μεζίτη Αλεξάνδρα

Τεχνικές αλληλουχισής νέας γενιάς

High-throughput

1990-2003: Ανθρώπινο γονιδίωμα → ~ 3 δισ. \$

2016: 250 Gb δεδομένων την εβδομάδα, 25×10^9 bp/ημέρα → 1 ανθρώπινο γονιδίωμα με κάλυψη 8X.

****** Πολύ γρήγορη εξέλιξη της τεχνολογίας που επιτρέπει μεγαλύτερη ταχύτητα, καλή ποιότητα και χαμηλότερες τιμές ******

Επόμενοι στόχοι

- Αύξηση του read length (μήκος ανάγνωσης αλληλουχιών)
- Περαιτέρω μείωση της τιμής
- Εξέλιξη των λογισμικών (software) και των υπολογιστών για γρηγορότερη και πιο έγκυρη συναρμολόγηση (assembly) των αλληλουχιών σε γονιδιώματα

Ορισμοί της high-throughput αλληλούχισης

Fragment: Μικρό κομμάτι γενωμικού DNA, μερικές εκατοντάδες βάσεις.

Single-end read: Αλληλουχία που ξεκινάει από το ένα άκρο του fragment (Εικ.1).

Paired-end read: Αλληλουχία που ξεκινάει και από τα δύο άκρα του fragment (Εικ.1). Πολλές βάσεις στην ένωση των δύο αναγνώσεων δεν αναγνωρίζονται σωστά και είναι απροσδιόριστες.

Read length: Ο αριθμός των βάσεων που αναφέρονται από ένα fragment.

Assembly: Η συναρμολόγηση fragments χρησιμοποιώντας περιοχές που υπερκαλύπτονται.

Contig: Το κομμάτι που προκύπτει από τη μερική συναρμολόγηση των fragments.

Μεταβλητότητα (Variability): Η μεταβλητότητα των **Single nucleotide polymorphisms (SNPs)** ανάμεσα σε άτομα του ίδιου πληθυσμού.

RNAseq: Αλληλούχιση του συνόλου των RNAs στο κύτταρο, το μεταγράφημα (transcriptome).

Single-end read



Paired-end read



Introduction to Genomics, Lesk A.M.

Εικόνα 1

Αλληλούχιση γονιδιωμάτων

De novo αλληλούχιση

- Ο πιο φιλόδοξος τρόπος αλληλούχισης
 - 'Σπάσιμο' του DNA σε κομμάτια ~200 βάσεων
 - Αλληλούχιση σε single-end reads ή paired-end reads- περιορισμένος ή μηδενικός άγνωστος αριθμός βάσεων σε διάβασμα με paired-end reads.
 - Συναρμολόγηση γονιδιώματος (assembly) σε contigs από επικαλυπτόμενα κομμάτια.
 - **Απαιτείται ικανός αριθμός κομματιών (fragments) για την κάλυψη όλου του γονιδιώματος σε αρκετά αντίγραφα.**
 - **ανίχνευση λαθών**
 - **Ένα καινούριο γονιδίωμα απαιτεί κάλυψη 30-50X.**
- *** Κάλυψη (coverage): ο λόγος του συνολικού μήκους των αλληλουχιών προς το μήκος του γονιδιώματος (π.χ. Single end reads $6 \cdot 10^3$ bp για γονιδίωμα $6 \cdot 10^4$ bp → coverage 10X).**
- Αλληλούχιση ευκαρυωτικών γονιδιωμάτων σε κομμάτια.

Αλληλούχιση γονιδιωμάτων

Resequencing- Επαναλληλούχιση

- Ύπαρξη γονιδιώματος του συγκεκριμένου είδους.
- Απλή αντιστοίχιση αλληλουχιών στο ήδη γνωστό γονιδίωμα.
- Δυσκολία με τις επαναλαμβανόμενες αλληλουχίες.
- Το λάθος πρέπει να είναι χαμηλότερο από τη φυσική ποικιλότητα (παράδειγμα;)
- Σημασία σε καρκινικά κύτταρα για ανίχνευση αλλαγών.

Exome sequencing- Αλληλούχιση εξωνίων

- Αλληλούχιση μόνο των εξωνίων ή των γονιδίων που μας ενδιαφέρουν.
- Τα εξώνια στο ανθρώπινο γονιδίωμα αποτελούν ~ 1% του γονιδιώματος.

Βάσεις δεδομένων

Στοχος: Αποθήκευση δεδομένων και παρουσίασή τους με χρήσιμο τρόπο.

*****Γονιδιωματική= Δεδομένα + Πληροφορική + Στατιστική*****

!!!Σημασία βιοπληροφορικής: Παραγωγή δεδομένων, συλλογή, αρχειοθέτηση, επεξεργασία, διανομή, επανάκτηση και ανάλυση.!!!

Πηγές Βιολογικών δεδομένων

1. Αλληλούχιση γονιδιωμάτων
2. Πρότυπα έκφρασης πρωτεϊνών
3. Μεταβολικά μονοπάτια
4. Πρότυπα αλληλεπίδρασης πρωτεϊνών και λειτουργικά δίκτυα
5. Επιστημονική βιβλιογραφία

***** Πλέον η κατάθεση των πρωτότυπων δεδομένων σε βάσεις είναι αναγκαστική για τη δημοσίευση *****

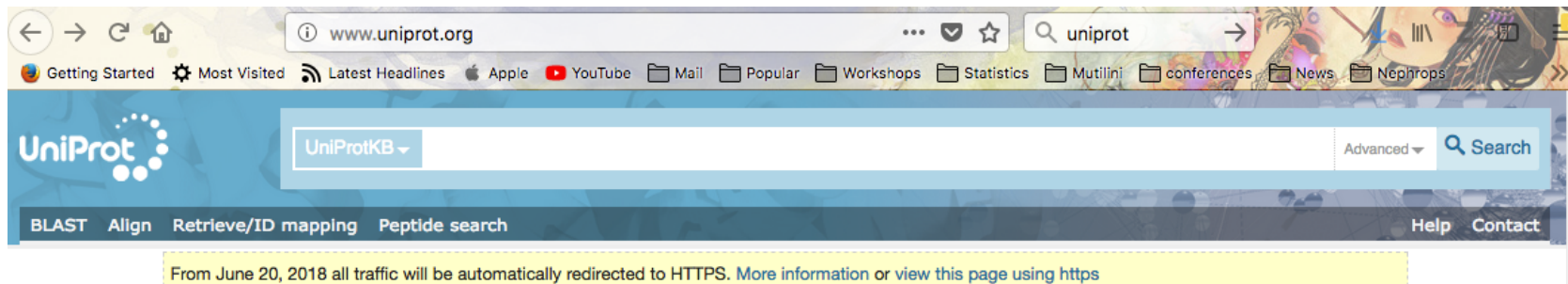
***** Η κατάθεση των δεδομένων γίνεται με συγκεκριμένη μορφοποίηση και με την χρήση συγκεκριμένων λογισμικών *****

Βάσεις δεδομένων

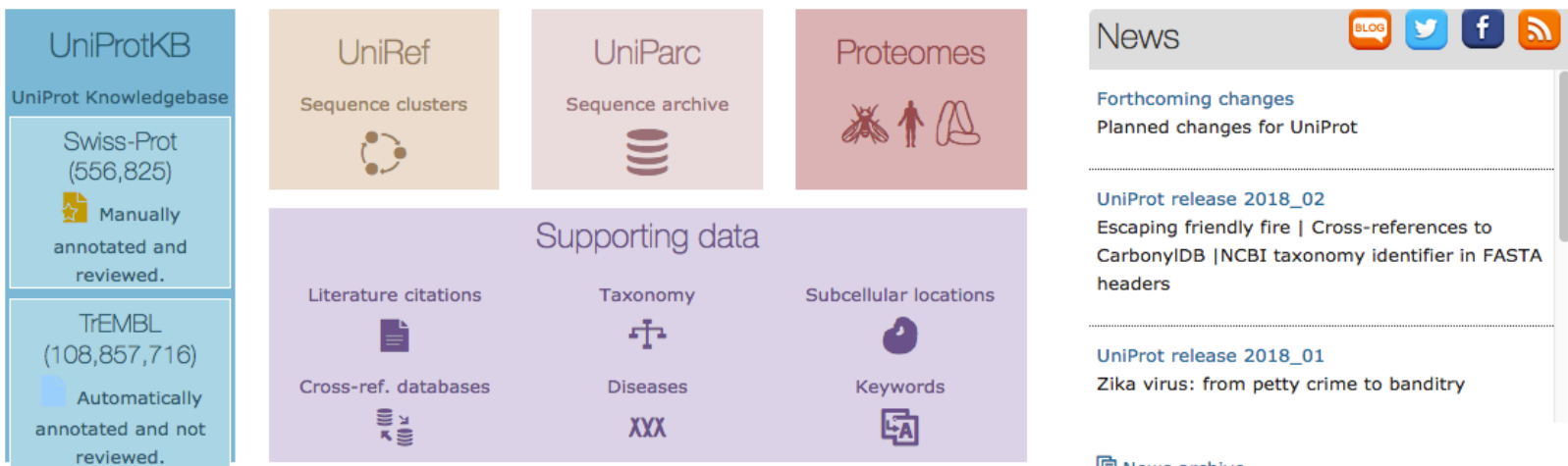
Βασεις ενο-devo

Συνδυαστική μελέτη των γενετικών κ αναπτυξιακών παραγόντων της φυσικής ποικιλότητας σε πληθυσμούς και τις εξελικτικές πιέσεις που τη δημιούργησαν.

→ Οι καλύτερες βάσεις είναι αυτές που έχουν τις καλύτερες παραπομπές σε άλλες βάσεις δεδομένων με συμπληρωματικά δεδομένα.



The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of protein sequence and functional information.



Βάσεις δεδομένων

Μηχανές αναζήτησης γονιδιωμάτων

- Οργάνωση και σχολιασμός (annotation) γονιδιωμάτων.

The screenshot shows the Ensembl genome browser website. At the top, there is a navigation bar with the Ensembl logo and links for BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, Blog, and Mirrors. A search bar is located on the right side of the navigation bar. Below the navigation bar, there is a main search area with a dropdown menu for 'All species' and a 'Go' button. Below the search area, there are several sections: 'Browse a Genome' with a description of the browser's capabilities and a list of 'Favourite genomes' including Human, Mouse, and Zebrafish; 'Find a Data Display' with a 'Try it now!' button; 'Variant Effect Predictor' with the 'Ve!P' logo; 'Gene expression in different tissues' with a microarray image; 'Retrieve gene sequence' with a DNA sequence; and 'Compare genes across species' with a phylogenetic tree. On the right side, there is a 'What's New in Ensembl Release 91 (December 2017)' section with a list of updates and a 'Go to Ensembl blog' link. Below that, there is a 'Tweets by @ensembl' section with a tweet from Ensembl about the Gene of the Week.




Search: for

e.g. **BRCA2** or **rat 5:62797383-63627669** or **rs699** or **coronary heart disease**

Browse a Genome

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

Favourite genomes

-  **Human**
GRCh38.p10
-  **Mouse**
GRCm38.p5
-  **Zebrafish**
GRCz10

[Edit favourites](#)

All genomes


-- Select a species --

[View full list of all Ensembl species](#)

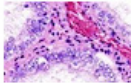
Find a Data Display

Not sure how to find the data visualisation you need? With our new [Find a Data Display](#) page, you can choose a gene, region or variant and then browse a selection of relevant visualisations

Variant Effect Predictor




Gene expression in different tissues



Retrieve gene sequence

```
GCCTGACTTCGGSTGG:
GGGCTTGTGGCGGAGC:
GGCGCTCTGCTGGCCT:
AGGGGACAGATTGTGA:
CACCTCTGGAGCGGTT:
CCCAGTCCAGCGTGGCC:
```

Compare genes across species



What's New in Ensembl Release 91 (December 2017)


- [New Primate Species](#)
- [Update of cat assembly and genebuild to Felis_catus_8.0](#)
- [Mouse: update to Ensembl-Havana Gencode gene set](#)
- [New and updated probe mapping data for primates](#)
- [Microarray Probe Mapping Update](#)

[Full details](#) | [All web updates, by release](#) | [More news on our blog](#)

- 01 Mar 2018: [Ensembl sites down](#)
- 21 Feb 2018: [What's coming in Ensembl 92 and Ensembl Genomes 39](#)
- 09 Feb 2018: [Ensembl Plants Project Leader](#)

[Go to Ensembl blog](#)

Tweets by @ensembl

 **Ensembl** @ensembl
LDL1 (Lysine-specific histone demethylase 1 homolog 1) is associated with @news4go terms that implicate it in the #nucleus and H3-K4 #methylation status buff.ly/2GG2Dku #GeneoftheWeek

Επιλογή είδους

Search Human (*Homo sapiens*)

Search all categories ▾

Search Human...

Go

e.g. BRCA2 or 17:63992802-64038237 or rs1333049 or osteoarthritis

Genome assembly: GRCh38.p10 (GCA_000001405.25)

More information and statistics

Download DNA sequence (FASTA)

Convert your data to GRCh38 coordinates

Display your data in Ensembl

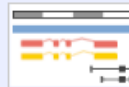
Other assemblies

GRCh37 Full Feb 2014 archive with BLAST, VEP and BioMart

Go



View karyotype



Example region

What's New in Human release 91

- Microarray Probe Mapping Update
- External database references update
- COSMIC data update

More news...

Gene annotation

What can I find? Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs.

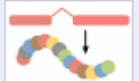
More about this genebuild

Download genes, cDNAs, ncRNA, proteins (FASTA)

Update your old Ensembl IDs



Example gene



Example transcript

Comparative genomics

What can I find? Homologues, gene trees, and whole genome alignments across multiple species.

More about comparative analysis

Download alignments (EMF)



Example gene tree

Regulation

What can I find? DNA methylation, transcription factor binding sites, histone modifications, and regulatory features such as enhancers and repressors, and microarray annotations.

More about the Ensembl regulatory build and microarray annotation

Experimental data sources

Download all regulatory features (GFF)



Example regulatory feature



ENCODE data in Ensembl

Variation

What can I find? Short sequence variants and longer structural variants; disease and other phenotypes

More about variation in Ensembl

Download all variants (GVF)

Variant Effect Predictor



ATCGAGCT
ATCCAGCT
ATCGAGAT

Example variant




Example phenotype

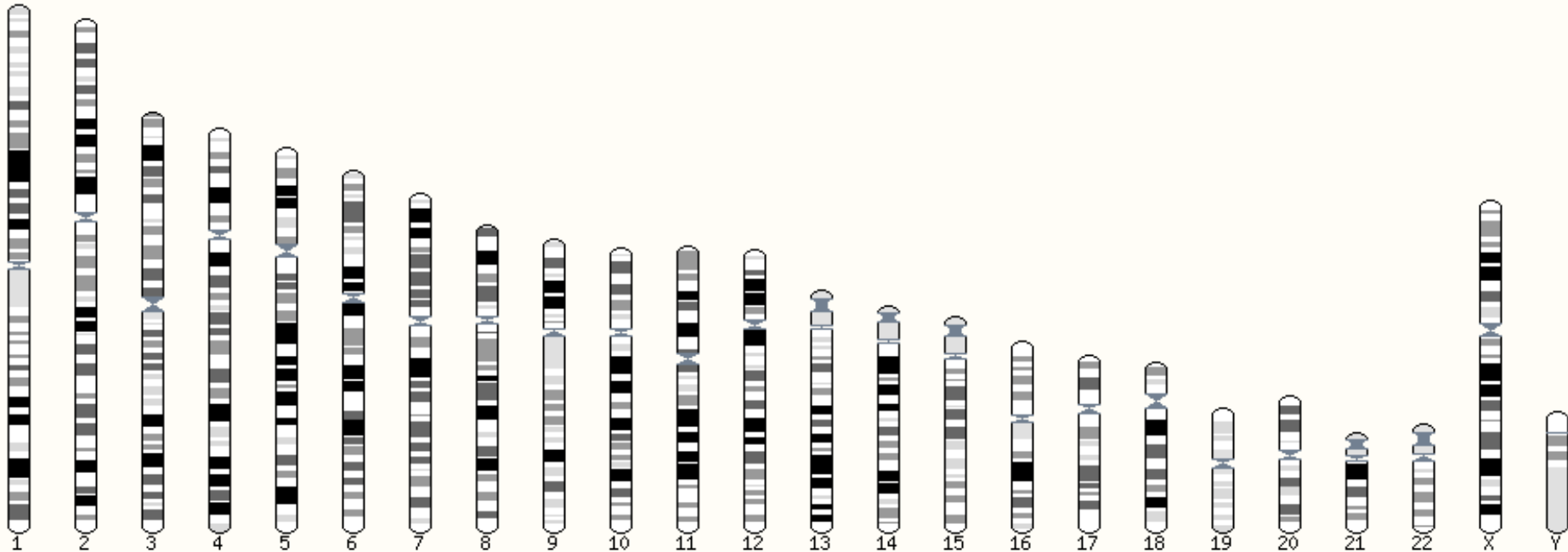


Example structural variant

Καρυότυπος

Whole genome 

 Add features



Click on the image above to jump to a chromosome, or click and drag to select a region

Γενικές πληροφορίες

Summary

Assembly	GRCh38.p10 (Genome Reference Consortium Human Build 38), INSDC Assembly GCA_000001405.25 , Dec 2013
Base Pairs	3,554,996,726
Golden Path Length	3,096,649,726
Annotation provider	Ensembl
Annotation method	Full genebuild
Genebuild started	Jan 2014
Genebuild released	Jul 2014
Genebuild last updated/patched	Jun 2017
Database version	91.38
Gencode version	GENCODE 27

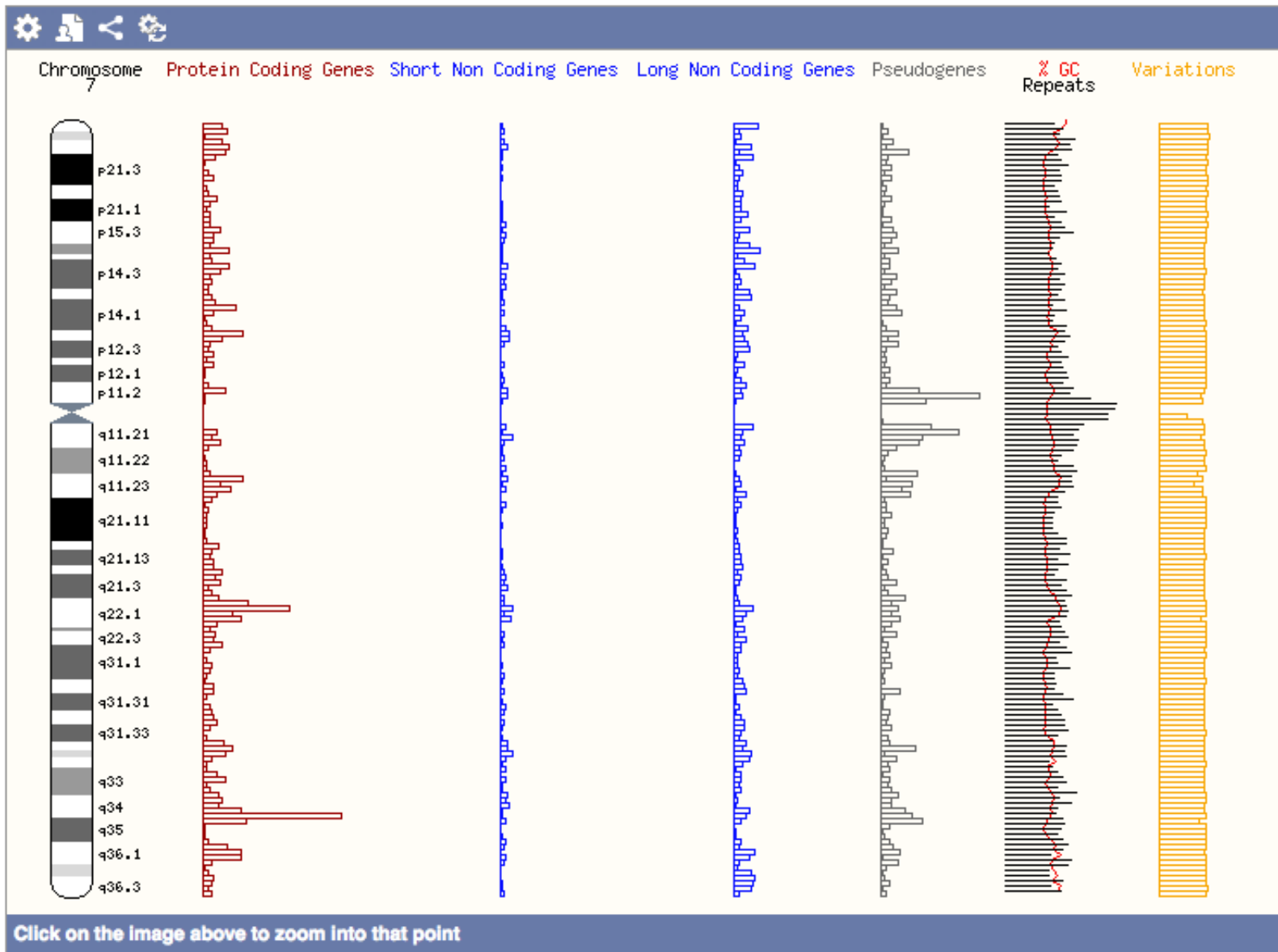
Gene counts (Primary assembly)

Coding genes	20,338 (incl 562 readthrough)
Non coding genes	22,521
Small non coding genes	5,363
Long non coding genes	14,720 (incl 238 readthrough)
Misc non coding genes	2,222
Pseudogenes	14,638 (incl 6 readthrough)
Gene transcripts	200,310

Gene counts (Alternative sequence)

Coding genes	2,750 (incl 37 readthrough)
Non coding genes	1,288
Small non coding genes	242
Long non coding genes	877 (incl 33 readthrough)
Misc non coding genes	169
Pseudogenes	1,600

Επιλογή χρωμοσώματος



Άσκηση 1: Βάσεις δεδομένων

→ Βρείτε γονίδια σχετικά με κάποιο ψάρι που σας ενδιαφέρει.

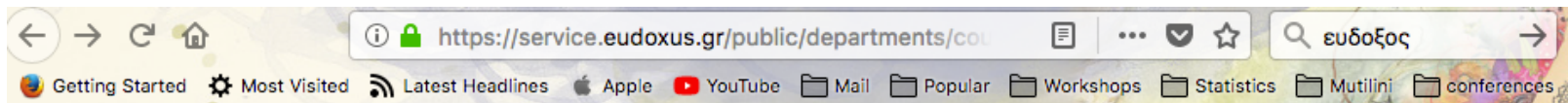
→ Έχετε μία αλληλουχία από γονίδιο του κυπρίνου (*Cyprinus carpio*)

1. Βρείτε ποια είναι η πιθανη λειτουργία της.
2. Μπορείτε να βρείτε αν υπάρχει δημοσίευση;
3. Υπάρχει το γονιδίωμα; Τι στοιχεία μπορείτε να βρείτε (μέγεθος, γονίδια κτλ.)

[Genbank](#)

[Uniprot](#)

[Ensembl](#)



Πρόγραμμα Σπουδών
ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΓΕΩΠΟΝΙΑΣ ΙΧΘΥΟΛΟΓΙΑΣ ΚΑΙ ΥΔΑΤΙΝΟΥ ΠΕΡΙΒΑΛΛΟΝΤΟΣ

Καταστάσεις Μαθημάτων/Συγγραμμάτων για το ακαδημαϊκό έτος 2017-2018

Μάθημα [ΒΠ1500]: ΓΟΝΙΔΙΩΜΑΤΙΚΗ

Εξάμηνο 0 - Εαρινό
Επιλογές Συγγραμμάτων:

1. Βιβλίο [59380291]: Εισαγωγή στη Γονιδιωματική, Arthur M. Lesk [Λεπτομέρειες](#)