

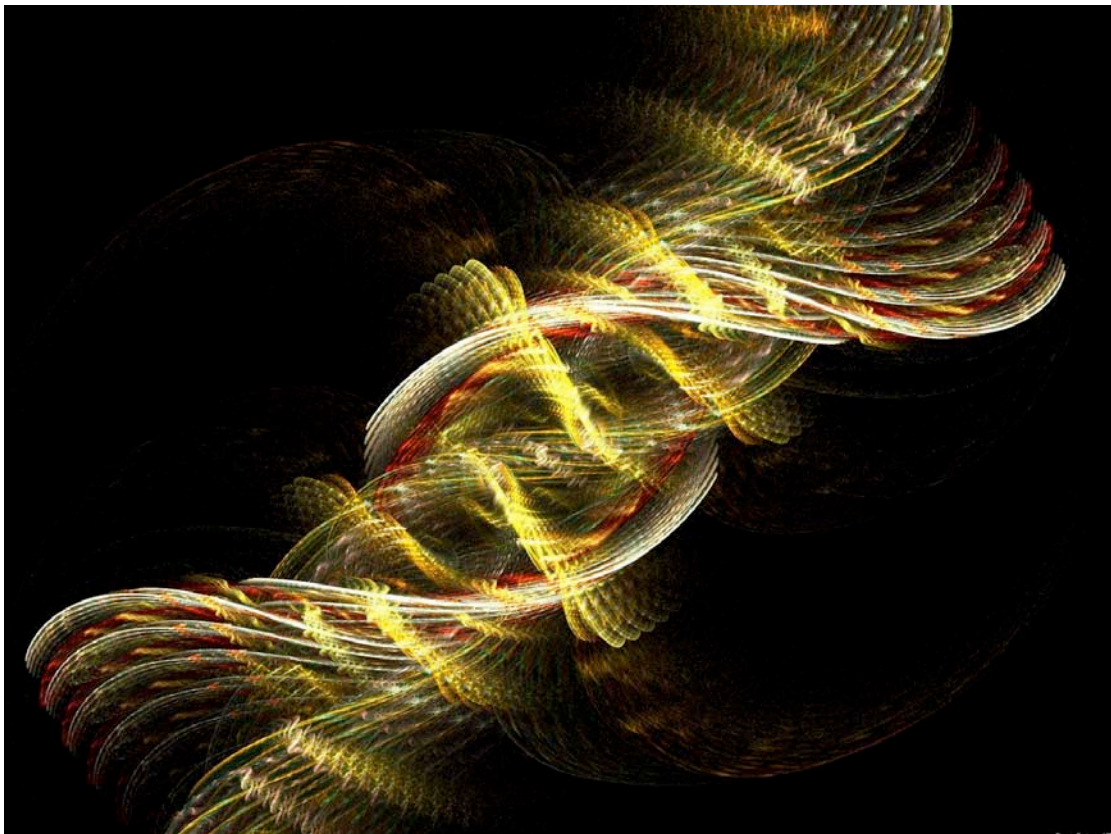
ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

**Πρόγραμμα Μεταπτυχιακών Σπουδών του Τμήματος Βιοχημείας και
Βιοτεχνολογίας**

**«ΕΦΑΡΜΟΓΕΣ ΜΟΡΙΑΚΗΣ ΒΙΟΛΟΓΙΑΣ-ΜΟΡΙΑΚΗ ΓΕΝΕΤΙΚΗ,
ΔΙΑΓΝΩΣΤΙΚΟΙ ΔΕΙΚΤΕΣ»**

ΔΟΞΑΡΑ ΑΛΕΞΙΑ

**ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ ΜΕΛΕΤΗ ΜΙΤΟΧΟΝΔΡΙΑΚΩΝ ΔΙΚΤΥΩΝ ΚΑΙ
ΣΥΝΑΦΩΝ ΑΣΘΕΝΕΙΩΝ ΑΠΟ ΔΕΔΟΜΕΝΑ ΓΟΝΙΔΙΩΜΑΤΙΚΗΣ ΚΑΙ
ΒΙΟΛΟΓΙΑΣ ΣΥΣΤΗΜΑΤΩΝ**



ΛΑΡΙΣΑ 2012

«Βιοπληροφορική μελέτη μιτοχονδριακών δικτύων και συναφών ασθενειών από δεδομένα γονιδιωματικής και βιολογίας συστημάτων»

«Bioinformatic study of mitochondrial networks and related diseases from genomics and systems biology data»

Τριμελής Συμβουλευτική Επιτροπή

- ❖ *Αμούτζιας Γρηγόριος, Λέκτορας Βιοπληροφορικής στη Γεωμική, Τμήμα Βιοχημείας & Βιοτεχνολογίας Πανεπιστημίου Θεσσαλίας*
- ❖ *Ζίφα Αιμιλία, Επίκουρος Καθηγήτρια Βιολογίας / Νευροβιολογίας, Τμήμα Βιοχημείας & Βιοτεχνολογίας Πανεπιστημίου Θεσσαλίας*
- ❖ *Μαμούρης Ζήσης, Καθηγητής Γενετικής Ζωικών Πληθυσμών, Τμήμα Βιοχημείας & Βιοτεχνολογίας Πανεπιστημίου Θεσσαλίας*

Ευχαριστίες

Η παρούσα διπλωματική εργασία εκπονήθηκε στο εργαστήριο Βιοπληροφορικής του τμήματος Βιοχημείας & Βιοτεχνολογίας του Πανεπιστημίου Θεσσαλίας στα πλαίσια του Προγράμματος Μεταπτυχιακών Σπουδών “ Εφαρμογές Μοριακής Βιολογίας – Μοριακή Γενετική – Διαγνωστικοί Δείκτες”.

Σε αυτό το σημείο θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες σε όλους όσους συνέβαλαν στην ολοκλήρωση της.

Ιδιαίτερα ευχαριστώ τον κ. Γρηγόρη Αμούτζια για την ανάθεση του θέματος, για την επιστημονική του καθοδήγηση καθ' όλη τη διάρκεια της προσπάθειας μου καθώς επίσης τον ευχαριστώ θερμά για την υπομονή και την κατανόηση που έδειξε.

Ευχαριστώ τον κ. Ζήση Μαμούρη και την κα Αιμιλία Ζίφα, για την συμμετοχή τους στην Τριμελή Συμβουλευτική Επιτροπή και τις υποδείξεις και συμβουλές τους.

Ένα μεγάλο ευχαριστώ οφείλω στον προπτυχιακό φοιτητή Ανάργυρο Γ. Χαλιώτη για τη βοήθεια που μου παρείχε κατά την εκτέλεση του πειραματικού μέρους.

Περίληψη

Το μιτοχόνδριο είναι το εργοστάσιο ενέργειας του κυττάρου και παίζει σημαντικότατο ρόλο στην λειτουργία του κυττάρου καθώς και στην εξέλιξη των ειδών. Τα γονίδια που κωδικοποιούνται στο *mt-DNA* (τα ονομάζουμε *mt_mt_genes*) εξελίσσονται γρηγορότερα από τα υπόλοιπα γονίδια. Σκοπός αυτής της εργασίας ήταν να μελετηθεί εάν τα γονίδια που κωδικοποιούνται στον πυρήνα αλλά συμμετέχουν στην δομή/λειτουργία του μιτοχονδρίου (τα ονομάζουμε *mt_genes*) ‘παρασύρονται’ από τα *mt_mt_genes* και έχουν κάποιες ιδιαιτερότητες σε σχέση με τα υπόλοιπα πυρηνικά γονίδια (τα ονομάζουμε *nuc_genes*), όσον αφορά την εξέλιξη τους, τις πρωτεϊνικές αλληλεπιδράσεις των προϊόντων τους και το κατά πόσο συμμετέχουν σε ασθένειες. Για το σκοπό αυτό συλλέχθηκαν δεδομένα λειτουργικής κατηγοριοποίησης, γονιδιωματικής, πρωτεϊνικών αλληλεπιδράσεων και ασθενειών από τις βάσεις δεδομένων *Gene Ontology*, *Ensembl*, *BioGrid* & *GeneCards* για τον άνθρωπο. Για τις εξελικτικές μελέτες συλλέχθηκαν δεδομένα (από την *Ensembl*) από τις συγκρίσεις ορθόλογων γονιδίων μεταξύ του ανθρώπου και πέντε άλλων θηλαστικών (*P. troglodytes*, *O. cuniculus*, *M. musculus*, *R. norvegicus*, *C. familiaris*). Όλα τα παραπάνω δεδομένα ενσωματώθηκαν & αναλύθηκαν στατιστικά με προγράμματα από τις γλώσσες προγραμματισμού *Perl* & *R*. Ένα από τα βασικά συμπεράσματα είναι ότι τα προϊόντα των *mt_genes* έχουν περισσότερες πρωτεϊνικές αλληλεπιδράσεις και εμπλέκονται πιο συχνά σε ασθένειες απ’ότι τα προϊόντα των *nuc_genes*. Όσον αφορά όμως την εξέλιξη των *mt_genes* (dN/dS , βαθμό συντήρησης της ακολουθίας), δεν εμφανίζουν πολύ αξιοσημείωτες διαφορές ως προς τα υπόλοιπα γονίδια. Νέες μελέτες με i) διαφορετικό τρόπο υπολογισμού του dN/dS , ii) διαφορετικά κριτήρια εμπλοκής ενός γονιδίου στο μιτοχόνδριο (με βάση την πρωτεωμική και όχι την Γονιδιακή Οντολογία) iii) καθώς επίσης και πιο αυστηρά κριτήρια επιλογής πρωτεϊνικών αλληλεπιδράσεων θα χρειαστούν για να αποσαφηνιστεί κατά πόσο τα γονίδια του *mt-DNA* συμπαρασύρουν τα γονίδια του πυρήνα.

Abstract

*The mitochondrion is the cell energy factory and has a great role in cell function and the evolution of species. The genes that are encoded in mt-DNA (named mt_mt genes) have a faster evolution than the others. The aim of this study is to search whether the genes encoded in the nucleus that participate in the structure/function of the mitochondrion (named mt_genes), are affected from their functional interaction with the mitochondrial genome and the possibility to have some particularities compared with the rest of the nuclear genes (we call them nuc_genes), regarding their evolution, the protein interactions of their products (their proteins) and whether they are involved in diseases. For this purpose, we collected data for humans from many databases such as Gene Ontology, Ensembl, BioGrid and GeneCards databases. For our evolutionary study we collected data from Ensembl, and we gathered some information from the comparison of orthologous genes between humans and five other mammals (*P. troglodytes*, *O. cuniculus*, *M. musculus*, *R. norvegicus*, *C. familiaris*). All these data were integrated and statistically analyzed using Perl and R programming languages. One of the main conclusions is that mt_gene products (proteins) have, on average, more protein interactions and tend to be involved in diseases more frequently than the nuc_gene products. However, concerning the evolution of mt_genes (dN/dS, and the degree of the sequence conservation,), we did not observe any substantial differences with the other genes. In order to investigate more thoroughly the hypothesis that the nuclear genome that is involved in mitochondrial functions is affected by the fast evolution of the mt genome, further analyses will be needed in the future. For example, one should use i) a different method of dN/dS calculation, ii) different criteria of search (assignment of mitochondrial role based on proteomics and not on Gene Ontology), iii) and finally one could use stricter criteria regarding protein interactions.*

ΠΕΡΙΕΧΟΜΕΝΑ

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή - Ο ρόλος και η σημασία του μιτοχονδρίου

<i>1.1 Ιστορική αναδρομή.....</i>	<i>9</i>
<i>1.2 Προέλευση των μιτοχονδρίων.....</i>	<i>10</i>
<i>1.3 Δομή των μιτοχονδρίων.....</i>	<i>13</i>
<i>1.4 Κατανομή, μέγεθος, σχήμα και αριθμός των μιτοχονδρίων.....</i>	<i>15</i>
<i>1.5 Λειτουργίες των μιτοχονδρίων.....</i>	<i>16</i>
<i>1.5.1 Ο κύκλος του Krebs.....</i>	<i>17</i>
<i>1.5.2 Οξειδωτική φωσφορυλίωση.....</i>	<i>17</i>
<i>1.5.3 Η β-οξείδωση των λιπαρών οξέων ή έλικα του Lyman.....</i>	<i>18</i>
<i>1.5.4 Η σύνθεση των λιπαρών οξέων.....</i>	<i>18</i>
<i>1.6 Γενετική ημιαυτονομία των μιτοχονδρίων.....</i>	<i>18</i>
<i>1.7 Παθογόνες και ουδέτερες μεταλλάξεις σε μιτοχονδριακά γονίδια tRNA.....</i>	<i>22</i>
<i>1.8 Μιτοχόνδρια και ασθένειες.....</i>	<i>24</i>

ΚΕΦΑΛΑΙΟ 2

Εξέλιξη

<i>2.1 Βασικές αρχές.....</i>	<i>27</i>
<i>2.2 Οι μεταλλάξεις στους πληθυσμούς και η θεωρία της ουδετερότητας.....</i>	<i>28</i>
<i>2.3 Εξελικτικές σχέσεις – Ορθόλογα και Παράλογα γονίδια.....</i>	<i>29</i>
<i>2.4 Ρυθμός εξέλιξης των αλληλουχιών – μιτοχονδρίων.....</i>	<i>30</i>
<i>2.5 Φυλογενετική ανάλυση.....</i>	<i>34</i>
<i>2.5.1 Πολλαπλή στοίχιση.....</i>	<i>34</i>
<i>2.5.2 Φυλογενετικά δένδρα.....</i>	<i>36</i>
<i>2.6 Πρωτεϊνικές αλληλεπιδράσεις και πρωτεϊνικά δίκτυα.....</i>	<i>38</i>
<i>Εισαγωγή.....</i>	<i>38</i>
<i>Η εξέλιξη των δικτύων των πρωτεϊνικών αλληλεπιδράσεων.....</i>	<i>39</i>
<i>Οι διάφοροι τύποι αλληλεπιδράσεων.....</i>	<i>40</i>

ΚΕΦΑΛΑΙΟ 3

Υλικά & μέθοδοι

<i>3.1 Γλώσσα προγραμματισμού Perl</i>	42
<i>3.2 Γονιδιακή οντολογία – Gene Ontology (GO)</i>	44
<i>3.2.1 Εργαλεία</i>	47
<i>3.2.2 Ιστορικά στοιχεία</i>	48
<i>3.3 Βάση δεδομένων Ensembl (Ensembl Genome Browser)</i>	48
<i>3.3.1 Ensembl Genome Browser</i>	48
<i>3.3.2 Πρόβλεψη ορθόλογων και παράλογων γονιδίων στη βάση δεδομένων Ensembl</i>	55
<i>3.4 Biomart</i>	56
<i>3.5 BioGRID (Biological Repository for Interaction Datasets)</i>	59
<i>3.6 Cytoscape</i>	62
<i>3.7 OMIM (Online Mendelian Inheritance in Man)</i>	63
<i>3.8 GeneCards</i>	65

ΚΕΦΑΛΑΙΟ 4

<i>Αποτελέσματα - Συζήτηση</i>	67
--------------------------------	----

<i>Συμπεράσματα και μελλοντικές αναλύσεις</i>	101
---	-----

<u><i>ΒΙΒΛΙΟΓΡΑΦΙΑ</i></u>	103
----------------------------	-----

ΚΕΦΑΛΑΙΟ 1

ΕΙΣΑΓΩΓΗ - Ο ΡΟΛΟΣ ΚΑΙ Η ΣΗΜΑΣΙΑ ΤΟΥ ΜΙΤΟΧΟΝΔΡΙΟΥ

Στη Βιολογία, η έννοια της **αναπνοής** έχει διπλή σημασία. Αφενός αναφέρεται στη φυσιολογική λειτουργία κατά την οποία γίνεται πρόσληψη οξυγόνου και αποβολή διοξειδίου του άνθρακα μέσω των πνευμόνων και αφετέρου στην οξείδωση των τροφών από τα κύτταρα, για την παραγωγή ενέργειας. Για τη διατήρηση της ζωής απαιτείται **ενέργεια** η οποία παράγεται και χρησιμοποιείται από τα κύτταρα με τη μορφή **ΑΤΡ**. Η διαδικασία δημιουργίας του ΑΤΡ παρουσία οξυγόνου καλείται **αερόβια αναπνοή** και είναι γνωστή ως «**αληθινή κυτταρική αναπνοή**». Βασικό ρόλο στη λειτουργία αυτή, στο ευκαρυωτικό κύτταρο, παίζουν τα **μιτοχόνδρια** (Μαργαρίτης *et al.*, 2004). Η μεγάλη σημασία και η σπουδαιότητα του συγκεκριμένου οργανιδίου αποτέλεσε αντικείμενο πολλών μελετών δεδομένου ότι το μιτοχόνδριο είναι ένα ημιαυτόνομο οργανίδιο, το οποίο διαθέτει έναν ανεξάρτητο κύκλο αναπαραγωγής, περιέχει δικό του γενετικό υλικό και υπόκειται σε πολλές μεταλλάξεις. Επιπλέον, το μιτοχονδριακό γονιδίωμα αλληλεπιδρά με το πυρηνικό, με αποτέλεσμα να δημιουργούνται πρωτεϊνικές αλληλεπιδράσεις και σύμπλοκα, σχηματίζοντας έτσι πολύπλοκα δίκτυα τα οποία μπορούν να εξελιχθούν είτε με μεταλλάξεις σε υπάρχοντα γονίδια, είτε με γονιδιακούς διπλασιασμούς/απώλειες ενώ δυσλειτουργίες στα γονίδια αυτά και κατ' επέκταση στα μιτοχόνδρια μπορούν να οδηγήσουν σε περισσότερες από 50 ασθένειες.

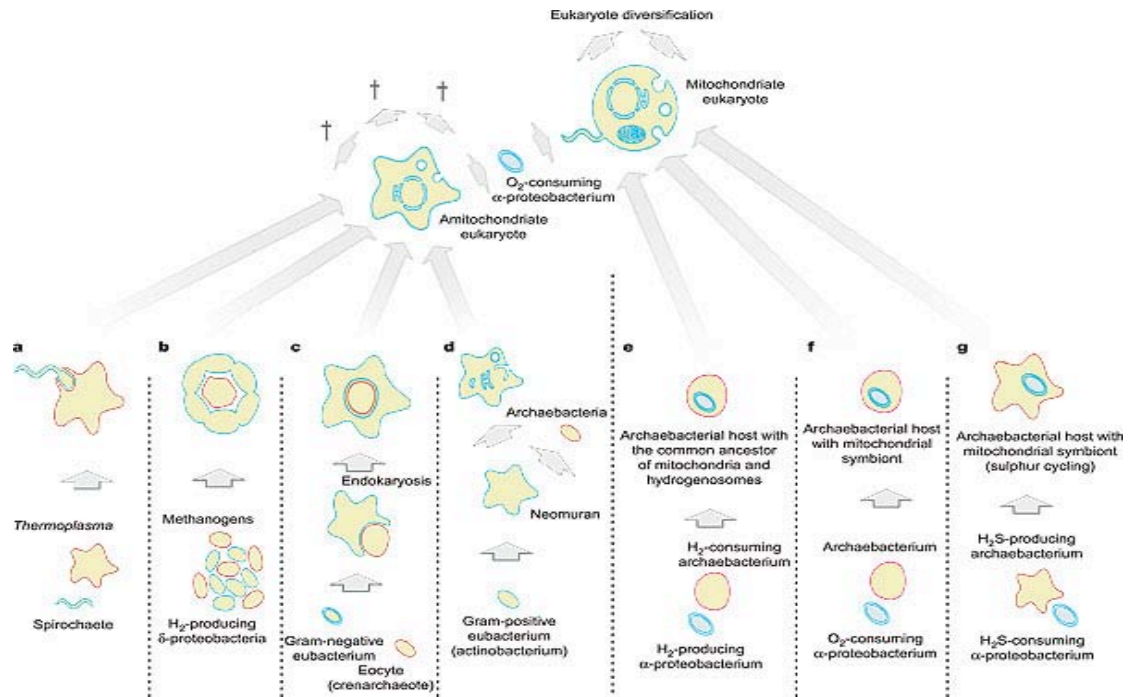
1.1 ΙΣΤΟΡΙΚΗ ΑΝΑΔΡΟΜΗ

Τα μιτοχόνδρια, ανακαλύφθηκαν ως συστατικό του κυττάρου το **1886** από τον **Altmann** ο οποίος χρησιμοποίησε μια ειδική χρωστική, το **πράσινο του Janus**, για φωτονικό μικροσκόπιο. Το **1898** ο **Benda** εισάγει τον όρο «μιτοχόνδριο», το **1904** ο **Meves** περιγράφει τα μιτοχόνδρια στα φυτικά κύτταρα και το **1934** οι **Bensley** και **Hoerr** απομονώνουν τα οργανίδια με υπερφυγοκέντρηση. Η λεπτομερειακή λεπτή δομή, όμως, έγινε γνωστή αρκετά χρόνια αργότερα, μετά την κατασκευή του

ηλεκτρονικού μικροσκοπίου και χάρη στις μελέτες των **G. Palade, K.R. Porter, F. Sjostrand** και άλλων, το **1952**, όπου επετράπη η κατανόηση των βιοχημικών μηχανισμών και η συσχέτιση δομής και λειτουργίας του σημαντικότερου κυτταρικού οργανιδίου (*Μαργαρίτης et al., 2004*).

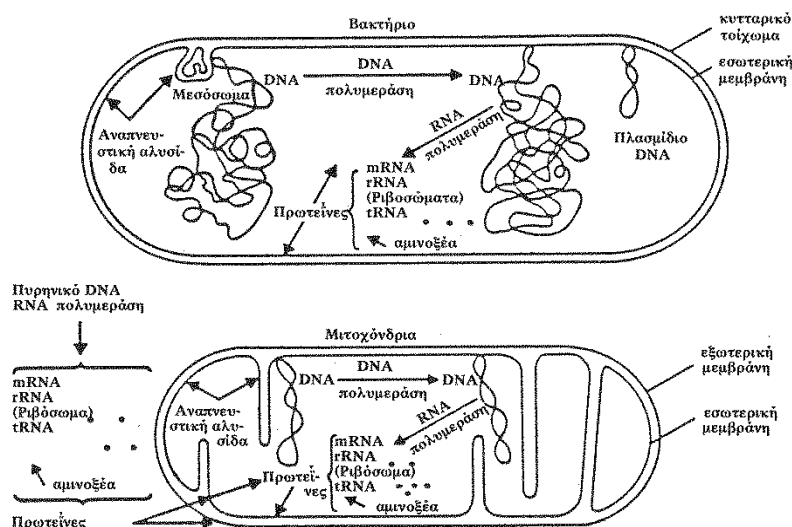
1.2 ΠΡΟΕΛΕΥΣΗ ΤΩΝ ΜΙΤΟΧΟΝΔΡΙΩΝ

Έχουν υπάρξει πολλές θεωρίες σχετικά με την προέλευση των μιτοχονδρίων. Η θεωρία όμως που είναι γενικά αποδεκτή και θεωρείται ότι προσεγγίζει περισσότερο τη προέλευση των μιτοχονδρίων είναι η **ενδοσυμβιωτική υπόθεση** (Martin & Mentel. 2010), η οποία υποστηρίζει ότι τα μιτοχόνδρια δημιουργήθηκαν από τη συμβίωση βακτηρίων και αρχέγονων κυττάρων. Σύμφωνα λοιπόν με αυτήν τα μιτοχόνδρια προέρχονται από την ανάπτυξη συμβιωτικών σχέσεων ανάμεσα σ' ένα βακτήριο και το κύτταρο ξενιστή που προσφέρει στη σχέση αυτή το πυρηνικό γονιδίωμα, ενώ ταυτόχρονα χρησιμοποιεί το σύστημα οξειδωτικής φωσφορυλίωσης του βακτηρίου για δική του χρήση. Σύμφωνα με τη θεωρία αυτή, το αρχικό προκαρυωτικό κύτταρο ξενιστής ήταν ένα αρχαιοβακτήριο και πιο συγκεκριμένα α-πρωτεοβακτήριο όμοιο με το σημερινό *Thermoplasma*, που είναι ανθεκτικό σε υψηλές θερμοκρασίες και όξινες συνθήκες. Το κύτταρο-ξενιστής δεν είχε φωτοσυνθετική ικανότητα ούτε ήταν ικανό να χρησιμοποιεί αποτελεσματικά το οξυγόνο. Είναι σημαντικό σ' αυτό το σημείο να αναφερθεί ότι τα μιτοχόνδρια παραμένουν ημιαυτόνομα από την άποψη ότι κρατούν το δικό τους γενετικό υλικό το οποίο αντιγράφεται και εκφράζεται αλλά δεν μπορεί να υπάρξει αυτόνομο. Έτσι, το μιτοχονδριακό γονιδίωμα θεωρείται το υπόλειμμα του βακτηριακού γενώματος της αρχικής ενδοσυμβιωτικής σχέσης. (Εικ.1.1)



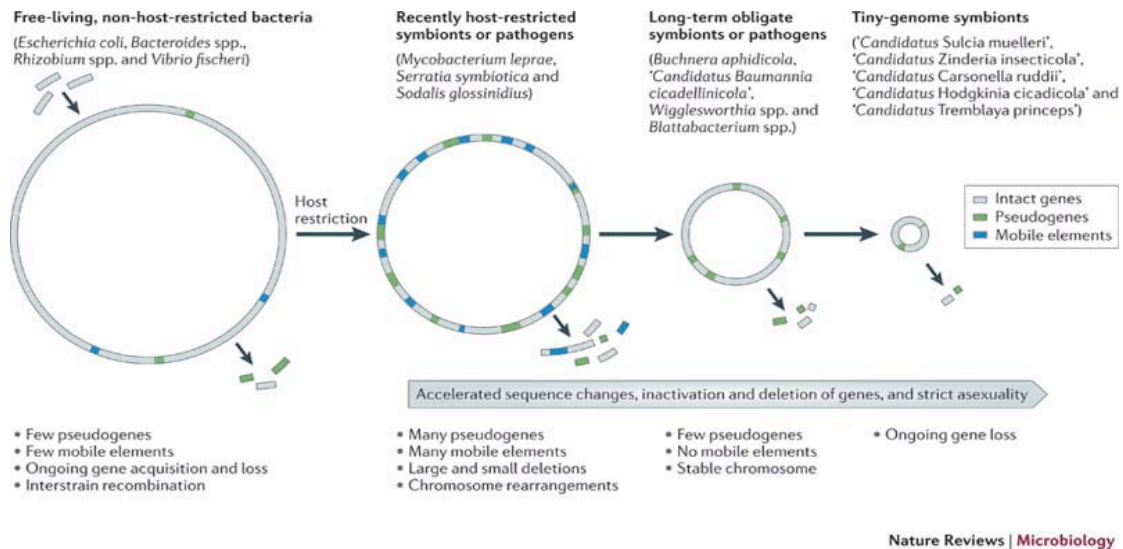
Εικόνα 1.1 Προέλευση των μιτοχονδρίων και μοντέλα οργανισμών που πιθανότατα συνέβαλλαν στην δημιουργία του μιτοχονδρίου (Martin & Mentel, 2010).

Η υπόθεση της ενδοσυμβιωτικής προέλευσης βασίζεται στην ομοιότητα που υπάρχει ανάμεσα στα μιτοχόνδρια και τα προκαρυωτικά κύτταρα (Εικ.1.2). Υπάρχουν ομοιότητες στη θέση της αναπνευστικής αλυσίδας, ενώ το βακτηριακό μεσόσωμα (προέκταση της πλασματικής μεμβράνης) μοιάζει με τις μιτοχονδριακές ακρολοφίες και περιέχει τα ένζυμα της αναπνευστικής αλυσίδας. Παράλληλα παρατηρούνται ομοιότητες ως προς την ύπαρξη κυκλικού DNA, τη δυνατότητα ανεξάρτητης πρωτεϊνικής σύνθεσης, τον τρόπο πολλαπλασιασμού και την παρουσία καρδιολιπίνης που αποτελεί κύριο μεμβρανικό φωσφολιπίδιο.



Εικόνα 1.2 Διάγραμμα στο οποίο φαίνονται οι ομοιότητες ανάμεσα σ' ένα βακτήριο και ένα μιτοχόνδριο (Θωμόπουλος, 1993).

Με αφορμή την συμβιωτική υπόθεση για την προέλευση των μιτοχονδρίων αλλά και των χλωροπλαστών και μετά από πολλές γενωμικές μελέτες που έχουν πραγματοποιηθεί έχουν αναφερθεί πάρα πολλές περιπτώσεις δημιουργίας συμβιωτικών βακτηρίων τα οποία περιέχουν εξαιρετικά μικρό γενετικό υλικό. Αυτοί οι οργανισμοί έχουν ανεξάρτητη καταγωγή από τις διάφορες βακτηριακές ομάδες ενώ ένα βασικό χαρακτηριστικό όλων είναι ο εξαιρετικά γρήγορος ρυθμός εξέλιξης των πρωτεϊνών τους όπως επίσης και ο αυξημένος αριθμός των chaperons. Το μικρό γενετικό υλικό αυτών των οργανιδίων κωδικοποιεί πολύ λιγότερες πρωτεΐνες σε σχέση με τα γονιδιώματα άλλων οργανιδίων και ιών, όμως η βασική διαφορά σε σχέση με αυτές τις οντότητες είναι η διατήρηση πολλών γονιδίων καθ' όλη τη διάρκεια του κύκλου ζωής που είναι πολύ σημαντικά για την επιβίωση του οργανισμού. Μελλοντικές εργασίες και μελέτες που αφορούν αυτούς τους οργανισμούς, δύναται να δώσουν πολλές πληροφορίες σχετιζόμενες με την εξέλιξη των κυττάρων, την φύση των οργανιδίων καθώς και το τι σημαίνει να είναι μια κυτταρική οντότητα αυτόνομη (Εικ.1.3) (Mc Cutcheon *et al.*, 2011).



Εικόνα 1.3 Η σταδιακή μείωση του γονιδιώματος σε έναν βακτήριο-ξενιστή και δημιουργία ενός συμβιωτικού βακτηρίου (Mc Cutcheon *et al.*, 2011).

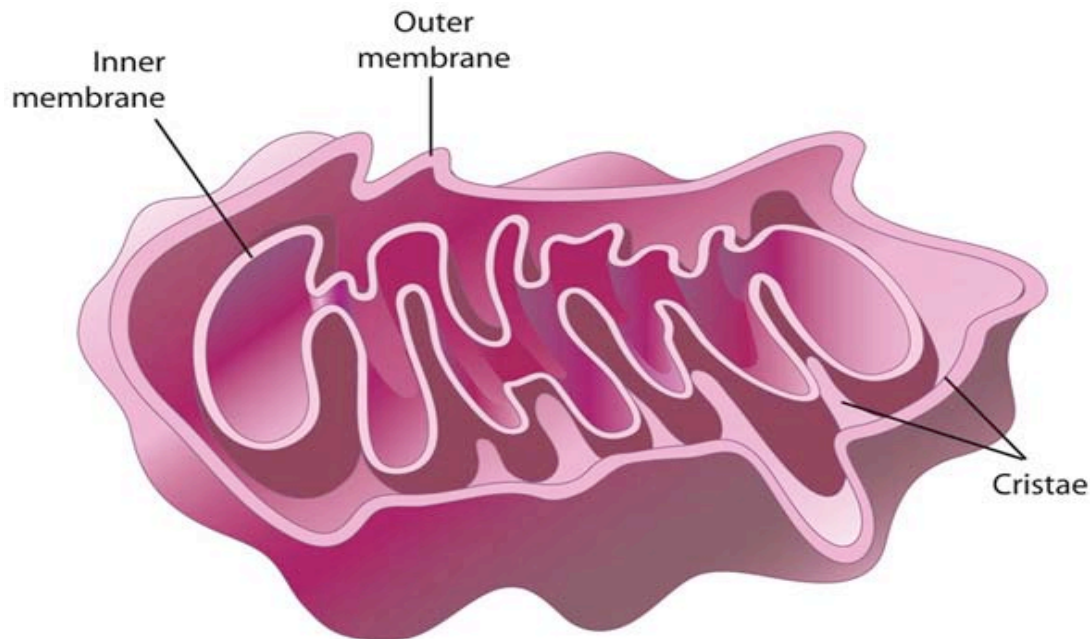
1.3 ΔΟΜΗ ΤΩΝ ΜΙΤΟΧΟΝΔΡΙΩΝ

Τα μιτοχόνδρια χαρακτηρίζονται ως διπλομεμβρανικά οργανίδια. Συνεπώς υπάρχουν δύο μεμβράνες, η εξωτερική και η εσωτερική. Οι δύο αυτές μεμβράνες είναι πολύ σημαντικές για τη διατήρηση και την ύπαρξη του μιτοχονδρίου δεδομένου ότι καθορίζουν το σκελετό του (Θωμόπουλος, 1993).

Η **εξωτερική μιτοχονδριακή μεμβράνη** (outer membrane) έχει πάχος περίπου 6nm και διαχωρίζει το εσωτερικό του μιτοχονδρίου από το υπόλοιπο κυτταρόπλασμα. Η μεμβράνη αυτή περιέχει μια μεγάλη πρωτεΐνη, την πορίνη, η οποία σχηματίζει υδρόφιλους διαύλους και είναι διαπερατή από όλα τα μόρια με MW μικρότερο από 5 kDa. Στην μεμβράνη αυτή υπάρχουν πρωτεΐνες που περιλαμβάνουν ένζυμα τα οποία σχετίζονται με τη σύνθεση των λιπιδίων, ενώ υπάρχουν και άλλα που μετατρέπουν τα λιπίδια σε παράγωγα, τα οποία στη συνέχεια μεταβολίζονται στην μιτοχονδριακή μήτρα.

Η **εσωτερική μιτοχονδριακή μεμβράνη** (inner membrane) περικλείει μία κοκκιώδη περιοχή που ονομάζεται **μιτοχονδριακός χυμός** ή **μήτρα** (matrix) και έχει και αυτή πάχος περίπου 6 nm. Η μεμβράνη αυτή είναι αδιαπέραστη από τα ιόντα και τα περισσότερα φορτισμένα μόρια, ενώ σχηματίζει πολυάριθμες αναδιπλώσεις προς το εσωτερικό του μιτοχονδρίου οι οποίες ονομάζονται **μιτοχονδριακές ακρολοφίες**. Οι

μιτοχονδριακές ακρολοφίες είναι ατελή διαφράγματα που δεν διακόπτουν τη συνέχεια του εσωτερικού χώρου (Εικ. 1.4), και περιέχουν πρωτεΐνες που εκτελούν τρεις διαφορετικές λειτουργίες: (α) τις πρωτεΐνες που διεκπεραιώνουν τις οξειδωτικές αντιδράσεις της αλυσίδας μεταφοράς των ηλεκτρονίων, (β) τη συνθετάση του ATP που παράγει ATP και (γ) μεταφορικές πρωτεΐνες (περμεάσες) που επιτρέπουν τη δίοδο μεταβολιτών προς και από τη μήτρα (Ζίφα *et al.*, 2010).



Εικόνα 1.4 Σχηματική απεικόνιση του μιτοχονδρίου (<http://www.nature.com/scitable/topicpage/mitochondria-14053590>).

Η **μήτρα (matrix)**, αποτελεί στην ουσία τον εσωτερικό χώρο του μιτοχονδρίου και περιέχει ένα πολύ πυκνό μίγμα εκατοντάδων ενζύμων, στα οποία περιλαμβάνονται τα ένζυμα που απαιτούνται για την οξείδωση του πυροσταφυλικού και των λιπαρών οξέων, καθώς και τα ένζυμα του κύκλου του Krebs. Η μήτρα περιέχει επίσης αρκετά αντίγραφα του μιτοχονδριακού DNA, ειδικά μιτοχονδριακά ριβοσώματα, μόρια t-RNA και ποικίλα ένζυμα απαραίτητα για την έκφραση των μιτοχονδριακών γονιδίων (Ζίφα *et al.*, 2010). Επίσης στη μιτοχονδριακή μήτρα παρατηρούνται και τα μιτοχονδριακά έγκλειστα, τα οποία μορφολογικά είναι κοκκία και αποτελούν τις θέσεις δέσμευσης δισθενών ιόντων, κυρίως Ca^{++} και Mg^{++} και ονομάζονται ενδομιτοχονδριακά κοκκία ενώ ο αριθμός των ενδομιτοχονδριακών κοκκίων διαφέρει από τον ένα κυτταρικό τύπο στον άλλο. Στην μιτοχονδριακή μήτρα έχουν επίσης παρατηρηθεί κοκκώδη σωμάτια τα οποία μετά από αναλύσεις αποδείχτηκε ότι είναι **μιτοχονδριακά ριβοσώματα**. Η μορφολογία τους είναι

ίδια με τη μορφολογία των κυτταροπλασματικών ριβοσωμάτων και λειτουργούν με τον ίδιο μηχανισμό για τη σύνθεση μιτοχονδριακών πρωτεϊνών.

Τέλος αξίζει να προσθέσουμε ότι το γονιδίωμα των μιτοχονδρίων είναι κυκλικό DNA και έχει μήκος 4,5-5,5 μm και θα γίνει αναλυτικότερη αναφορά σε επόμενη παράγραφο (Θωμόπουλος, 1993).

1.4 ΚΑΤΑΝΟΜΗ, ΜΕΓΕΘΟΣ, ΣΧΗΜΑ ΚΑΙ ΑΡΙΘΜΟΣ ΤΩΝ ΜΙΤΟΧΟΝΔΡΙΩΝ

Σε πολλά κύτταρα τα μιτοχόνδρια είναι ομοιόμορφα κατανεμημένα στο κυτταρόπλασμα, αλλά σε αρκετές περιπτώσεις συγκεντρώνονται σε συγκεκριμένες περιοχές. Η κατανομή των μιτοχονδρίων σε συγκεκριμένες κυτταρικές θέσεις διαφόρων τύπων κυττάρων σχετίζεται με την κύρια λειτουργία τους ως κέντρο παραγωγής ενέργειας. Πχ. σε ένα σπερματοζωάριο, τα μιτοχόνδρια βρίσκονται στην ουρά, γύρω από τον πυρήνα του κινητού μαστιγίου, που χρειάζεται ATP για την κίνηση του, ενώ σ' ένα μυοκαρδιακό κύτταρο, τα μιτοχόνδρια βρίσκονται κοντά στη συσταλτική συσκευή έτσι ώστε η υδρόλυση του ATP να παρέχει την απαιτούμενη ενέργεια για τη συστολή.

Όσον αφορά το μέγεθος των μιτοχονδρίων, αυτό είναι συνήθως διαφορετικό μεταξύ των κυτταρικών τύπων. Το πλάτος των μιτοχονδρίων κυμαίνεται συνήθως από **0,5-1 μm** ενώ το μήκος μπορεί να φτάσει και τα **10 μm** . Το μέγεθος των μιτοχονδρίων είναι δυνατόν να μεταβληθεί φυσιολογικά ή παθολογικά υπό συγκεκριμένες συνθήκες. Στα κύτταρα του εγγύς εσπειραμένου νεφρικού σωληναρίου π.χ., τα μιτοχόνδρια αυξάνουν φυσιολογικά τη μάζα τους για να ανταποκριθούν στις αυξημένες μεταβολικές απαιτήσεις κατά την προσρόφηση γλυκόζης, αμινοξέων, δικαρβονικών και φωσφορικών ιόντων. Μεταβολές στο μέγεθος των μιτοχονδρίων προκαλούνται και από ουσίες που επηρεάζουν την οξειδωτική φωσφορυλίωση όπως π.χ. κυανίδιο, δινιτροφαινόλη κτλ. Πολύ μεγάλα μιτοχόνδρια (μήκους μέχρι και 30 μm) παρατηρούνται σε περιπτώσεις αλκοολισμού και σε διάφορες παθολογικές καταστάσεις. Στο μυοκάρδιο π.χ. κάτω από συνθήκες ανοξίας, υποξίας ή μετά από επίδραση ορμονών ή φαρμάκων, σχηματίζονται πολύ μεγάλα μιτοχόνδρια που ονομάζονται **μεγαμιτοχόνδρια**. Αυτά δημιουργούνται από σύντηξη μικρότερων μιτοχονδρίων, έχουν κανονικές μιτοχονδριακές ακρολοφίες και αποικοδομούνται με αυτοφαγοκυττάρωση.

Το σχήμα των μιτοχονδρίων επίσης δεν είναι σταθερό αλλά συνήθως είναι επιμήκες, αλλαντοειδές ή σκωληκόμορφο.

Ο ακριβής αριθμός των μιτοχονδρίων ενός κυττάρου είναι δύσκολο να προσδιοριστεί με ακρίβεια, αλλά γενικά ο αριθμός αυτός μεταβάλλεται ανάλογα με τον τύπο και το λειτουργικό στάδιο του κυττάρου (Θωμόπουλος, 1993). Τα μιτοχόνδρια αποτελούν συνήθως το 18-20% του συνολικού κυτταροπλασματικού όγκου του ευκαρυωτικού κυττάρου και ο αριθμός τους κυμαίνεται από 1.700 σ' ένα ηπατοκύτταρο και 10.000 στη γιγάντια αμοιβάδα *Chaos chaos* μέχρι και 10.000.000 στα ωκύτταρα αμφιβίων (Μαργαρίτης *et al.*, 2004). Επιπλέον στα φυτικά κύτταρα υπάρχουν πολύ λιγότερα μιτοχόνδρια απ' ό τι στα ζωικά κύτταρα ενώ στα καρκινικά κύτταρα ο αριθμός τους ελαττώνεται σε σχέση με τα αντίστοιχα μη καρκινικά κύτταρα. Αυτό, οφείλεται πιθανότατα στη μείωση της οξειδωτικής και αύξηση της γλυκολυτικής δράσης των καρκινικών κυττάρων (Θωμόπουλος, 1993).

1.5 ΛΕΙΤΟΥΡΓΙΕΣ ΤΩΝ ΜΙΤΟΧΟΝΔΡΙΩΝ

Ο κύριος ρόλος των μιτοχονδρίων είναι η παραγωγή ενέργειας με μηχανισμούς που είναι γνωστοί ως **βιολογικές οξειδώσεις** και οι οποίοι περιλαμβάνουν τις αντιδράσεις παραγωγής CO₂ από τρι- και δικαρβονικά οξέα (κύκλος του Krebs), τη μεταφορά ηλεκτρονίων, την παραγωγή **ATP (τριφωσφορικής αδενοσίνης)**, την οξείδωση λιπαρών οξέων και τον καταβολισμό των αμινοξέων.

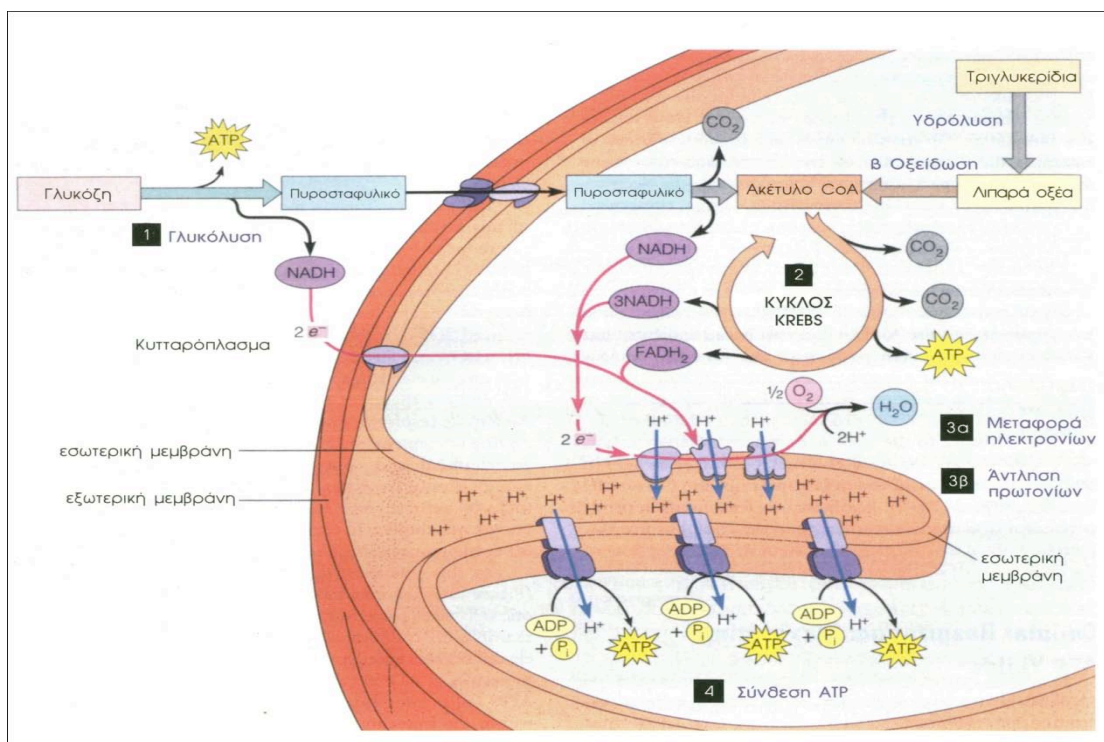
Κάθε τυπικό ευκαρυωτικό κύτταρο παράγει ~10.000.000.000 μόρια ATP καθ όλη τη διάρκεια της ζωής του, συμβάλλοντας έτσι στην επίτευξη **σημαντικών λειτουργιών** όπως:

- τη σύνθεση μακρομορίων (πρωτεϊνών DNA, RNA, πολυσακχαριτών, λιπιδίων κ.α)
- την κίνηση μαστιγίων και βλεφαρίδων
- τη σύσπαση μυϊκών κυττάρων
- τον πολυμερισμό ινιδίων κυτταρικού σκελετού
- την ενεργή μεταφορά ιόντων και ρυθμιστικών μορίων μέσω των μεμβρανών
- την ενδοκύτωση και έκκριση

- την ενζυμική κατάλυση κ.α. (Μαργαρίτης *et al.*, 2004).

1.5.1 Ο κύκλος του Krebs

Ο κύκλος του Krebs αποτελεί την κοινή καταβολική οδό οξειδωτικής αποικοδόμησης των λιπιδίων, υδατανθράκων και πρωτεϊνών σε όλα τα κύτταρα, σε αερόβιες συνθήκες και εξελίσσεται στο χώρο της μήτρας (20.000 κύκλοι Krebs πραγματοποιούνται κάθε στιγμή σ' ένα καρδιακό μιτοχόνδριο) Είναι μια οργανωμένη αλληλουχία ενζυμικών αντιδράσεων, οι οποίες έχουν σαν αρχικό και τελικό υπόστρωμα το οξαλοξικό οξύ (Εικ.1.5) (Ζίφα *et al.*, 2010).



Εικόνα 1.5 Τα μόρια της γλυκόζης διασπώνται σε πυροσταφυλικό (1), το οποίο μπαίνει στα μιτοχόνδρια, όπου και οξειδώνεται με τη βοήθεια της πυροσταφυλικής αφυδρογονάσης σε ακέτυλο-CoA. Άλλη πηγή ακέτυλο CoA είναι τα λιπαρά οξέα, μέσω της β-οξείδωσης. Στη συνέχεια το ακέτυλο CoA μπαίνει στο κύκλο του Krebs (2). Τα κύρια προϊόντα του κύκλου, NADH και FADH₂ δίνουν τα ηλεκτρόνια τους στην αναπνευστική αλυσίδα (3α), όπου με την αναγωγή του O₂ σε H₂O δημιουργούν τη διαφορά συγκέντρωσης πρωτονίων (3β) που ενεργοποιεί τη σύνθεση του ATP στην εσωτερική μεμβράνη για να παραχθεί ATP(4) (Ζίφα *et al.*, 2010).

1.5.2 Οξειδωτική φωσφορύλιωση

Οξειδωτική φωσφορύλιωση ονομάζεται η διαδικασία παραγωγής ATP από ADP και P_i και πραγματοποιείται στην εσωτερική μιτοχονδριακή μεμβράνη. Η

οξειδωτική φωσφορυλίωση είναι η κύρια πηγή ATP και αποτελεί το επιστέγασμα μιας σειράς ενεργειακών μετασχηματισμών, οι οποίοι ονομάζονται στο σύνολο τους **κυτταρική αναπνοή**. Γενικά, κατά την κυτταρική αναπνοή καταναλώνονται πυροσταφυλικό οξύ, λιπαρά οξέα, ADP, φωσφορικές ομάδες και οξυγόνο, ενώ παράγονται CO₂ και H₂O με τη χρήση ενζύμων της αναπνευστικής αλυσίδας (Θωμόπουλος, 1993).

1.5.3 Η β οξείδωση των λιπαρών οξέων ή έλικα του Lyman.

Τα λιπαρά οξέα οξειδώνονται (αποικοδομούνται) μόνο εάν ενεργοποιηθούν προηγουμένως από το συνδυασμό τους με το συνένζυμο A και με την προϋπόθεση ότι θα εισβάλλουν στη μιτοχονδριακή μήτρα. Το πέρασμα μέσω της μεμβράνης γίνεται με τη βοήθεια της καρνιτίνης. Η αποικοδόμηση εξελίσσεται μέσα στη μήτρα σε διαδοχικά στάδια. Η πλήρης οξείδωση των λιπαρών οξέων αποφέρει σημαντικά αποθέματα ATP στο κύτταρο. Για παράδειγμα η απόδοση σε ATP της αποικοδόμησης του παλμιτικού οξέως σε CO₂ και H₂O είναι 129 ATP (Ζίφα *et al.*, 2010).

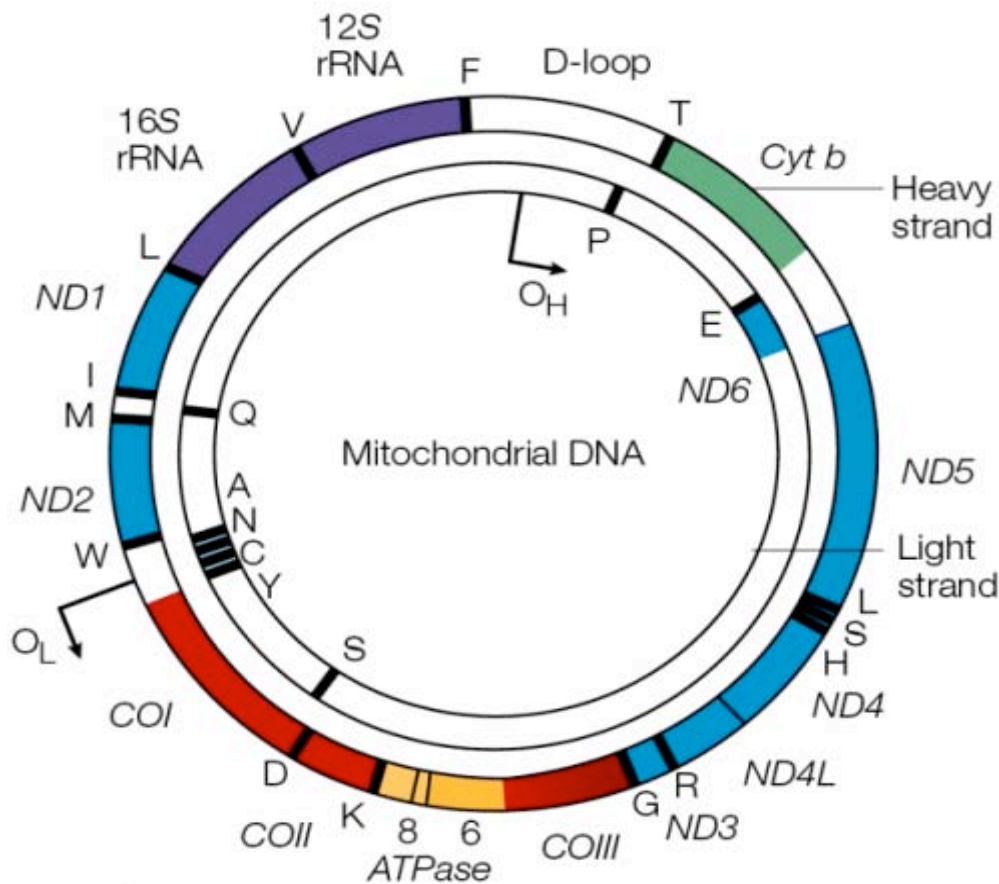
1.5.4 Η σύνθεση των λιπαρών οξέων.

Η σύνθεση των λιπαρών οξέων μπορεί να γίνει στο εσωτερικό των μιτοχονδρίων με τη βοήθεια ενός συστήματος, που είναι γνωστό με το όνομα μιτοχονδριακό σύστημα. Η σύνθεση ακολουθεί την αντίθετη πορεία από εκείνη που χρησιμοποιήθηκε για την αποικοδόμηση των λιπαρών οξέων. Το αποτέλεσμα είναι συνήθως η σύνθεση του στεατικού οξέως, αλλά επίσης και μικρών ποσοτήτων παλμιτικού, λαουρικού, μυριστικού και αραχιδονικού οξέως (Ζίφα *et al.*, 2010).

1.6 Η ΓΕΝΕΤΙΚΗ ΗΜΙΑΥΤΟΝΟΜΙΑ ΤΩΝ ΜΙΤΟΧΟΝΔΡΙΩΝ

Η ύπαρξη DNA στα μιτοχόνδρια επιβεβαιώθηκε από τους Nass και Nass το 1963. Υπάρχουν αρκετές διαφορές μεταξύ mtDNA και πυρηνικού. Το μιτοχονδριακό DNA (mtDNA) είναι ένα δίκλωνο κυκλικό μόριο ενώ το πυρηνικό είναι γραμμικό (Εικ.1.6). Το 1981 βρέθηκε ότι το μιτοχονδριακό γονιδίωμα αποτελείται από 16.569 ζεύγη βάσεων, ενώ το πυρηνικό από 3.3 δισεκατομμύρια ζεύγη βάσεων, είναι μη-ανασυνδυαζόμενο, γυμνό από ιστόνες και μοιάζει με το βακτηριακό DNA. Είναι πολύ σημαντικό να αναφερθεί ότι μόνο το 3% του γενώματος δεν κωδικοποιείται, ενώ στο

πυρηνικό γονιδίωμα το ποσοστό αυτό ανέρχεται στο 93% (Chial *et al.*, 2008). Στα πιο πολλά κύτταρα θηλαστικών το mtDNA έχει μήκος 5,5 μm, στα φυτά και στους μικροοργανισμούς είναι 3-5 φορές μεγαλύτερο ενώ σε σχέση με το βακτηριακό DNA (1000μm) το mtDNA είναι μικρότερο. Το ανθρώπινο mtDNA αποτελεί μοντέλο κυτταρικής οικονομίας, αφού χρησιμοποιούνται όλα τα ζεύγη βάσεων και δεν υπάρχουν ιντρόνια, ψευδογονίδια, επαναλαμβανόμενο DNA και μεγάλες ακολουθίες που να παρεμβάλλονται μεταξύ των γονιδίων, ενώ αυτά βρίσκονται πολύ κοντά το ένα στο άλλο.



Εικόνα 1.6 Χάρτης του ανθρώπινου μιτοχονδριακού γονιδιώματος. Το mtDNA περιλαμβάνει 37 γονίδια που κωδικοποιούν 13 πρωτεΐνες, 2 rRNAs και 22 tRNAs. Τα tRNAs είναι διάσπαρτα κατανεμημένα ως μονάδες ή σχηματίζοντας ομάδες, μεταξύ των άλλων RNAs. Τα tRNAs που κωδικοποιούνται από τη βαριά αλυσίδα υποδεικνύονται με γαλάζιο χρώμα, και τα tRNAs που κωδικοποιούνται από την ελαφριά αλυσίδα υποδεικνύονται με κίτρινο χρώμα (Chial *et al.*, 2008).

Στον παρακάτω πίνακα γίνεται αναφορά των διαφορών μεταξύ του μιτοχονδριακού και πυρηνικού DNA (Πιν.1)

Characteristic	Nuclear genome	Mitochondrial genome
Size	~3.3 x 10 ⁹ bp	16,569 bp
Number of DNA molecules per cell	23 in haploid cells; 46 in diploid cells	Several thousand copies per cell (polyploidy)
Number of genes encoded	~20,000–30,000	37 (13 polypeptides, 22 tRNAs and 2 rRNAs)
Gene density	~1 per 40,000 bp	1 per 450 bp
Introns	Frequently found in most genes	Absent
Percentage of coding DNA	~3%	~93%
Codon usage	The universal genetic code	AUA codes for methionine; TGA codes for tryptophan; AGA and AGG specify stop codons
Associated proteins	Nucleosome-associated histone proteins and non-histone proteins	No histones; but associated with several proteins (for example, TFAM) that form nucleoids
Mode of inheritance	Mendelian inheritance for autosomes and the X chromosome; paternal inheritance for the Y chromosome	Exclusively maternal
Replication	Strand-coupled mechanism that uses DNA polymerases α and δ	Strand-coupled and strand-displacement models; only uses DNA polymerase γ
Transcription	Most genes are transcribed individually	All genes on both strands are transcribed as large polycistrons
Recombination	Each pair of homologues recombines during the prophase of meiosis	There is evidence that recombination occurs at a cellular level but little evidence that it occurs at a population level

*Table modified from REE. 171 © (1999) John Wiley and Sons. TFAM, mitochondrial transcription factor A; rRNA, ribosomal RNA.

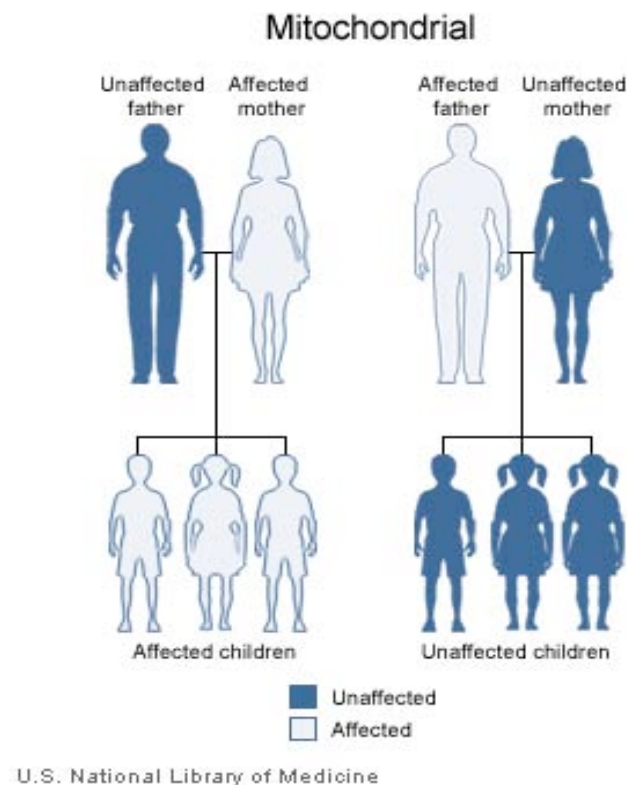
Πίνακας 1. Διαφορές μιτοχονδριακού και πυρηνικού DNA (Taylor *et al.*, 2005).

Εδώ αξίζει να αναφερθεί πως ο ρυθμός μετάλλαξης του mtDNA είναι περίπου 100 φορές μεγαλύτερος από αυτόν του πυρηνικού (Chial *et al.*, 2008), και συνεπώς το γεγονός αυτό συμβάλλει στην αυξημένη συσσώρευση πολυμορφισμών και τη δημιουργία μεγάλης ποικιλομορφίας μεταξύ των μιτοχονδρίων, όχι μόνο μεταξύ διαφορετικών ειδών, αλλά και ανάμεσα στο ίδιο είδος. Οι λόγοι που ευθύνονται για τον τόσο μεγάλο ρυθμό μεταλλακτικότητας θα συζητηθούν στο επόμενο κεφάλαιο. Ο αυξημένος ρυθμός των μεταλλάξεων των μιτοχονδρίων συμβάλλει μακροπρόθεσμα στην εξέλιξη των ειδών. Τέλος πρέπει να σημειωθεί ότι η πλειοψηφία των πρωτεϊνών που είναι απαραίτητες για τη μιτοχονδριακή λειτουργία κωδικοποιείται από το πυρηνικό και όχι από το μιτοχονδριακό γονιδίωμα (Κορώνας & Σαλγκάμης, 2011).

Μετά από αναλύσεις του χάρτη του ανθρώπινου μιτοχονδριακού γονιδιώματος βρέθηκε πως πέρα από την ύπαρξη του κυκλικού δίκλωνου μορίου του έχει αναφερθεί διαφορετικότητα των κλώνων, όσον αφορά το περιεχόμενο σε GC, με αποτέλεσμα ο ένας να είναι βαρύτερος του άλλου. Ο ένας κλώνος ονομάζεται H (heavy) και ο άλλος L (light). Το mtDNA κωδικοποιεί 37 γονίδια: 13 γονίδια των πρωτεϊνών που ανήκουν στα σύμπλοκα I, III, IV και V της αναπνευστικής αλυσίδας, 22 μιτοχονδριακά tRNAs και 2

rRNAs (12S και 16S). Ειδικότερα, ο κλώνος Η κωδικοποιεί 28 γονίδια και ο κλώνος L κωδικοποιεί 9 γονίδια, εκ των οποίων τα 8 κωδικοποιούν μιτοχονδριακά μόρια tRNAs. Στο μιτοχονδριακό γονιδίωμα έχει βρεθεί μια περιοχή πλούσια σε AT, η οποία αποτελεί το σημείο έναρξης της αντιγραφής και ονομάζεται replication origin ή D-loop (Εικ.1.6)

Λόγω της κυτταροπλασματικής του εντόπισης, το μιτοχονδριακό γονιδίωμα κληρονομείται μητρικά (τα ώριμα σπερματοζώαρια περιέχουν πολύ λίγο κυτταρόπλασμα με λίγα μιτοχόνδρια). Ως εκ τούτου, μόνο τα θηλυκά άτομα μπορούν να μεταβιβάσουν τις μεταλλάξεις σε όλους τους απογόνους ανεξαρτήτως φύλου (Εικ.1.7).



Εικόνα 1.7 Σχηματική απεικόνιση της μητρικής κληρονομής. Το μιτοχονδριακό γονιδίωμα μεταβιβάζεται μητρικά σε όλους τους απογόνους ανεξαρτήτως φύλου (<http://www.web-books.com/eLibrary/Medicine/Appendix/MitoInheritance.jpg>).

Εξάιρεση στην μέχρι τώρα παραδοχή ότι το mtDNA κληρονομείται αποκλειστικά μητρικά αποτελούν τα μύδια της οικογένειας *Mytilidae*, (και συγκεκριμένα το είδος *Mytilus galloprovincialis*) τα οποία εμφανίζουν ένα ξεχωριστό τρόπο κληρονομής του μιτοχονδριακού DNA. Τα θηλυκά άτομα μεταβιβάζουν το μιτοχονδριακό τους DNA (τύπος F) στους θηλυκούς και στους αρσενικούς απογόνους. Επιπλέον τα αρσενικά μύδια περιέχουν έναν ακόμη τύπου mtDNA (τύπος M), τον οποίο μεταβιβάζουν μόνο

στους αρσενικούς τους απογόνους (Διπλή Μονογονική Κληρονόμηση). Εδώ αξίζει να αναφέρουμε ότι σε ορισμένες περιπτώσεις παρατηρείται το φαινόμενο της αρρενοποίησης κατά το οποίο ένα γονιδίωμα τύπου F αντικαθιστά το Μ στην πατρική γραμμή κληρονόμησης, με αποτέλεσμα να υπάρχουν αρσενικά μύδια ετεροπλασμικά για δύο διαφορετικά τύπου F γονιδιώματα (Ladoukakis *et al.*, 2002).

Τα μιτοχονδριακά γονίδια χρησιμοποιούν τέλος έναν εναλλακτικό κώδικα κατά την μιτοχονδριακή πρωτεϊνοσύνθεση. Οι παραλλαγές στο γενετικό κώδικα των μιτοχονδρίων σε σχέση με τον «παγκόσμιο γενετικό κώδικα» συνοψίζονται στον παρακάτω πίνακα 2 (Θωμόπουλος, 1993).

Οργανισμός	Κωδικόνιο	Παγκόσμιος γενετικός κώδικας	Μιτοχόνδρια
Θηλαστικά	UGA	Κωδικόνιο Λήξης	Τρυπτοφάνη
	AUA	Ισολευκίνη	Μεθειονίνη
	AGA, AGG	Αργινίνη	Κωδικόνιο Λήξης

Πίνακας 2. Παραλλαγές στον γενετικό κώδικα των μιτοχονδρίων στα θηλαστικά (Κορώνας & Σαλγκάμης, 2011).

1.7 ΠΑΘΟΓΟΝΕΣ ΚΑΙ ΟΥΔΕΤΕΡΕΣ ΜΕΤΑΛΛΑΞΕΙΣ ΣΕ ΜΙΤΟΧΟΝΔΡΙΑΚΑ ΓΟΝΙΔΙΑ tRNA

Οι μεταλλάξεις διακρίνονται σε **παθογόνες** και σε **ουδέτερες**. Οι πρώτες επηρεάζουν τη δομή και τη λειτουργία του μιτοχονδριακού DNA με αποτέλεσμα την εμφάνιση ασθενειών στον οργανισμό. Οι δεύτερες δεν επηρεάζουν τη δομική ακεραιότητα και τη λειτουργικότητα του μιτοχονδρίου και συμβάλλουν στη δημιουργία γενετικής ποικιλότητας ενώ μπορούν να χρησιμοποιηθούν και για τη φυλογενετική ανάλυση των πληθυσμών.

Μεταλλάξεις έχουν βρεθεί σε όλο το μιτοχονδριακό γονιδίωμα, στα tRNAs, στα rRNAs και στις πρωτεΐνες της αναπνευστικής αλυσίδας. Ωστόσο, οι περισσότερες αφορούν mt-tRNAs. Η πλειοψηφία των mt-tRNA γονιδίων συγκεντρώνει τόσο **παθογόνες** όσο και **ουδέτερες** μεταλλάξεις, γεγονός αξιοσημείωτο καθώς οι tRNA αλληλουχίες αποτελούν μόνο το 10% του συνολικού μιτοχονδριακού γονιδιώματος (Zifa *et al.*, 2007).

Κατά την διάρκεια της τελευταίας δεκαετίας υπήρξε μια προοδευτική συσσώρευση αναφορών που συνδέουν τις μεταλλάξεις των mt-tRNA γονιδίων με σοβαρές διαταραχές στον άνθρωπο. Μέχρι στιγμής έχουν περιγραφεί 139 παθογόνες και 243 πολυμορφικές μιτοχονδριακές tRNA μεταλλάξεις. Δεδομένης της πολυπλοκότητας που παρουσιάζει γενετικά και βιοχημικά το μιτοχόνδριο, οι κλινικές εκδηλώσεις των μιτοχονδριακών παθήσεων είναι αρκετά ετερογενείς. Έτσι μπορούν να παρουσιαστούν βλάβες σε συγκεκριμένους ιστούς ή δομές ιστών και να προκληθούν πολύ σοβαρές ασθένειες όπως μυοπάθειες, καρδιομυοπάθειες, εγκεφαλομυοπάθειες ή ακόμη και πιο πολύπλοκα σύνδρομα (Zifa *et al.*, 2007). Σε πολλές από αυτές τις ασθένειες διαπιστώθηκε η ύπαρξη συνδυασμών μεταλλάξεων σε μιτοχονδριακά και πυρηνικά γονίδια με αποτέλεσμα την αναστολή της δράσης των συμπλόκων της αναπνευστικής αλυσίδας και κατ' επέκταση τη διακοπή της παραγωγής ενέργειας. Αξίζει να επισημανθεί πως μεταλλάξεις σε άλλες περιοχές του mtDNA πλην των tRNA μιτοχονδριακών γονιδίων, μπορούν να προκαλέσουν παθογένεση, ωστόσο οι πληροφορίες που έχουμε μέχρι σήμερα είναι αρκετά περιορισμένες (Zifa *et al.*, 2008).

Με δε δομένο ότι το μιτοχονδριακό γονιδίωμα διπλασιάζεται αυτόνομα και ανεξάρτητα από το πυρηνικό και ότι τα μιτοχόνδρια διαχωρίζονται στα θυγατρικά κύτταρα ανεξάρτητα από τα πυρηνικά χρωμοσώματα, το ποσοστό των μιτοχονδρίων που φέρουν στο γενετικό τους υλικό τη μετάλλαξη μπορεί να διαφέρει σημαντικά μεταξύ διαφορετικών σωματικών κυττάρων και ιστών. Αυτή η ετερογένεια ονομάζεται **ετεροπλασμία** και από το βαθμό αυτής της ετερογένειας εξαρτάται το αν μια παθογόνος μετάλλαξη μπορεί να οδηγήσει στο κλινικό φαινότυπο μιας ασθένειας. **Ετεροπλασμία** λοιπόν, είναι η συνύπαρξη φυσιολογικού και μεταλλαγμένου DNA στο κύτταρο, που δημιουργείται ύστερα από την κυτταρική διαίρεση όπου τα μιτοχόνδρια και τα γονιδιώματα τους κατανέμονται τυχαία στα θυγατρικά κύτταρα. Η αναλογία των μεταλλαγμένων έναντι των φυσιολογικών mtDNA καθορίζει την έκφραση και σοβαρότητα του νοσήματος.

Οι παθογόνες μεταλλάξεις των tRNA γονιδίων είναι συνήθως ετεροπλασμικές, ενώ οι πολυμορφισμοί είναι ομοπλασμικοί. Επιπρόσθετα, οι ιστοί διαφέρουν ως προς την εξάρτησή τους από την οξειδωτική φωσφορυλίωση, με την καρδιά, τους σκελετικούς μύες και το κεντρικό νευρικό σύστημα να είναι τα πιο εξαρτημένα. Γι' αυτόν λοιπόν το λόγο τα μιτοχονδριακά νοσήματα χαρακτηρίζονται συχνά από μυοπάθειες και εγκεφαλοπάθειες.

Συνοψίζοντας, θα μπορούσαμε να πούμε ότι η φαινοτυπική έκφραση μιας μετάλλαξης είναι το προϊόν της αλληλεπίδρασης μιας σειράς παραγόντων. Στους παράγοντες αυτούς περιλαμβάνονται το ποσοστό της ετεροπλασμίας, η κατανομή των μεταλλαγμένων γονιδιωμάτων στα διαφορετικά κύτταρα και ιστούς, ο μιτωτικός διαχωρισμός καθώς και η εξάρτηση που παρουσιάζουν οι διάφοροι ιστοί και όργανα από τη διαδικασία της οξειδωτικής φωσφορυλίωσης και την παραγωγή ενέργειας. Επιπλέον, η παρουσία και άλλων μεταλλάξεων στο μιτοχονδριακό αλλά και στο πυρηνικό γονιδίωμα και η επίδραση περιβαλλοντικών παραγόντων είναι πιθανό να εξηγούν τη διαφορετική έκφραση των μιτοχονδριακών μεταλλάξεων.

1.8 ΜΙΤΟΧΟΝΔΡΙΑ ΚΑΙ ΑΣΘΕΝΕΙΕΣ

Οι μιτοχονδριακές διαταραχές συχνά εμφανίζονται ως νευρολογικές διαταραχές, αλλά μπορούν να εμφανιστούν και ως **μυοπάθειες** οι οποίες προκαλούνται από σφάλματα της αναπνευστικής αλυσίδας. Οι μιτοχονδριακές μυοπάθειες χαρακτηρίζονται από υποτονία και αδυναμία των μυών ενώ πολύ συχνά επηρεάζονται ο εγκέφαλος, το κεντρικό νευρικό σύστημα, η καρδιά και τα νεφρά. Η οξέωση από γαλακτικό οξύ είναι το κοινό σύμπτωμα σ' όλες τις περιπτώσεις ενώ αξίζει να επισημανθεί ότι τα μιτοχόνδρια στις μιτοχονδριακές μυοπάθειες σχηματίζουν συχνά μεγάλα συσσωματώματα, έχουν μεγάλο μέγεθος, οι μιτοχονδριακές ακρολοφίες αποκτούν ασυνήθεις διαμορφώσεις και παρατηρούνται κρυσταλλικά έγκλειστα στη μιτοχονδριακή μήτρα (Taylor *et al.*, 2005).

Πιστεύεται ότι υπάρχει επίσης κάποια σχέση ανάμεσα στα μιτοχόνδρια και τον καρκίνο. Συνήθως στα καρκινικά κύτταρα παρατηρείται μειωμένος αριθμός μιτοχονδρίων, ενώ στην εσωτερική μιτοχονδριακή μεμβράνη βρίσκονται μεγάλα ποσά

χοληστερόλης. Τα μιτοχόνδρια είναι συνήθως μικρότερα στα καρκινικά κύτταρα, περιέχουν μικρότερο αριθμό μιτοχονδριακών ακρολοφιών και είναι κάπως διογκωμένα.

Επιπλέον, η σωστή λειτουργία των μιτοχονδρίων είναι απαραίτητη για τη φυσιολογική παραγωγή και έκκριση ινσουλίνης από τα παγκρεατικά κύτταρα. Συνεπώς, βλάβες στο mtDNA που κληρονομούνται μητρικά και διαταράσσουν τη φυσιολογική μιτοχονδριακή λειτουργία οδηγούν στην εμφάνιση διαβήτη.

Οι ασθένειες που οφείλονται σε μεταλλάξεις του mtDNA περιλαμβάνουν επιπλέον διάφορα σύνδρομα, όπως σύνδρομο Kearns-Sayre (KSS) (Schapira, 2012), το σύνδρομο MELAS (Wang *et al.*, 2010), το σύνδρομο Pearson, το σύνδρομο MERRF ή αλλιώς μυοκλονική επιληψία με τραχιές κόκκινες ίνες, την αταξία του Friedreich, την ασθένεια του Wilson και την κληρονομική οπτική νευροπάθεια Leber (Fruhman *et al.*, 2011) (Πίνακας 3).

Table 2 | **Clinical disorders that are caused by mutations in mitochondrial DNA**

Mitochondrial DNA disorder	Clinical phenotype	mtDNA genotype	Gene	Status	Inheritance	Reference
Kearns-Sayre syndrome	Progressive myopathy, ophthalmoplegia, cardiomyopathy	A single, large-scale deletion	Several deleted genes	Heteroplasmic	Usually sporadic	61,158
CPEO	Ophthalmoplegia	A single, large-scale deletion	Several deleted genes	Heteroplasmic	Usually sporadic	61, 64
Pearson syndrome	Pancytopenia, lactic acidosis	A single, large-scale deletion	Several deleted genes	Heteroplasmic	Usually sporadic	65
MELAS	Myopathy, encephalopathy, lactic acidosis, stroke-like episodes	3243A>G; 3271T>C Individual mutations	<i>TRNL1</i> <i>ND1 and ND5</i>	Heteroplasmic Heteroplasmic	Maternal Maternal	159 160, 161
MERRF	Myoclonic epilepsy, myopathy	8344A>G; 8356T>C	<i>TRNK</i>	Heteroplasmic	Maternal	162
NARP	Neuropathy, ataxia, retinitis pigmentosa	8993T>G	<i>ATP6</i>	Heteroplasmic	Maternal	163
MILS	Progressive brain-stem disorder	8993T>C	<i>ATP6</i>	Heteroplasmic	Maternal	67
MIDD	Diabetes, deafness	3243A>G	<i>TRNL1</i>	Heteroplasmic	Maternal	164
LHON	Optic neuropathy	3460G>A 11778G>A 14484T>C	<i>ND1</i> <i>ND4</i> <i>ND6</i>	Hetero- or homoplasmic Hetero- or homoplasmic Hetero- or homoplasmic	Maternal Maternal Maternal	165 62 166
Myopathy and diabetes	Myopathy, weakness, diabetes	14709T>C	<i>TRNE</i>	Hetero- or homoplasmic	Maternal	167,168
Sensorineural hearing loss	Deafness	1555A>G Individual mutations	<i>RNR1</i> <i>TRNS1</i>	Homoplasmic Hetero- or homoplasmic	Maternal Maternal	55 169,170
Exercise intolerance	Fatigue, muscle weakness	Individual mutations	<i>CYB</i>	Heteroplasmic	Sporadic	68
Fatal, infantile encephalopathy; Leigh/Leigh-like syndrome	Encephalopathy, lactic acidosis	10158T>C; 10191T>C	<i>ND3</i>	Heteroplasmic	Sporadic	66

ATP6, ATPase 6; *CPEO*, chronic progressive external ophthalmoplegia; *CYB*, cytochrome *b*; *LHON*, Leber hereditary optic neuropathy; *MELAS*, mitochondrial myopathy, encephalopathy, lactic acidosis and stroke-like episodes; *MERRF*, myoclonic epilepsy and ragged-red fibres; *MIDD*, maternally-inherited diabetes and deafness; *MILS*, maternally-inherited Leigh syndrome; *ND1, 3-6*, NADH dehydrogenase subunits 1, 3-6; *NARP*, neurogenic weakness, ataxia and retinitis pigmentosa; *RNR1*, 12S ribosomal RNA; *TRNE, TRNK, TRNL1, TRNS1*, mitochondrial tRNAs.

Πίνακας 3. Μιτοχονδριακές ασθένειες και σύνδρομα (Taylor *et al.*, 2005).

Άλλες παθολογικές καταστάσεις με αιτιολογία που εμπεριέχει μιτοχονδριακή δυσλειτουργία περιλαμβάνουν την σχιζοφρένεια, την διπολική διαταραχή, την άνοια, τη νόσο Alzheimer (Stewart *et al.*, 1992), τη νόσο του Parkinson (Sherer *et al.*, 1992) και

τα εγκεφαλικά επεισόδια όταν εμφανίζονται σε νεαρή ηλικία Ένα κοινό χαρακτηριστικό που θεωρείται ότι συνδέει αυτές τις κατά τα άλλα άσχετες μεταξύ τους διαταραχές είναι η κυτταρική βλάβη που προκαλεί το οξειδωτικό στρες.

Στη βιβλιογραφία συγκεντρώνονται όλο και περισσότερα στοιχεία που προκύπτουν ότι στον εγκέφαλο και στους περιφερικούς ιστούς ασθενών με Alzheimer παρουσιάζεται μιτοχονδριακή δυσλειτουργία και οξειδωτικό στρες (Mancuso *et al.*, 2006) και όλο και περισσότερα στοιχεία υποδεικνύουν ότι το οξειδωτικό στρες, το οποίο αυξάνεται στα μιτοχόνδρια που παρουσιάζουν βλάβη, είναι ένα πρώιμο γεγονός που σχετίζεται με τη νευροεκφύλιση. Όμως, παρά τα στοιχεία που αναφέρουν την ύπαρξη μορφολογικών, βιοχημικών και μοριακών ανωμαλιών των μιτοχονδρίων σε διάφορους ιστούς ασθενών με Alzheimer, ο ρόλος του μιτοχονδριακού γονιδιώματος και των απλότυπων του ως παράγοντες κινδύνου παραμένει αμφισβητήσιμος. Ακόμα υπάρχει αντιπαράθεση αν το οξειδωτικό στρες και η βλάβη των μιτοχονδρίων εμπλέκονται στην εκδήλωση και εξέλιξη των νευροεκφυλιστικών διαταραχών ή απλά είναι συνέπεια της νευροεκφύλισης (Mancuso *et al.*, 2008).

ΚΕΦΑΛΑΙΟ 2

ΕΞΕΛΙΞΗ

2.1 ΒΑΣΙΚΕΣ ΑΡΧΕΣ

Στην επιστήμη της βιολογίας, με τον όρο *εξέλιξη* εννοούμε την αλλαγή στις ιδιότητες/χαρακτηριστικά ενός πληθυσμού ατόμων του ίδιου είδους με το πέρασμα του χρόνου. Αν και τέτοιου είδους μεταβολές παρατηρούνται σε μικρή κλίμακα σε κάθε γενιά, μακροπρόθεσμα και αθροιστικά μπορούν να οδηγήσουν σε σημαντικές αλλαγές ώστε να δημιουργηθούν τελικά νέα άτομα που είναι πλέον αναπαραγωγικά απομονωμένα από τα τυπικά άτομα του είδους, δημιουργώντας έτσι ένα νέο είδος. Αυτό το φαινόμενο ονομάζεται *ειδογένεση*. Γενικά η εξέλιξη πραγματοποιείται με διαφορετικούς ρυθμούς στα διάφορα είδη ενώ πολλές ομοιότητες που παρουσιάζονται μεταξύ αυτών οφείλονται στο ότι όλα τα είδη προέρχονται από έναν κοινό πρόγονο. Η ποικιλότητα και οι φαινοτυπικές διαφοροποιήσεις μεταξύ των πληθυσμών οφείλονται σε διάφορα γεγονότα πχ. στις μεταλλάξεις, στον γενετικό ανασυνδυασμό καθώς και στην γονιδιακή ροή όπου παρατηρείται μεταφορά γονιδίων. Υπάρχουν δύο βασικοί μηχανισμοί: *η φυσική επιλογή* και *η γενετική παρέκκλιση*.

Η θεωρία της *φυσικής επιλογής* διατυπώθηκε από τον Κάρολο Δαρβίνο το 1858 ο οποίος υποστήριξε πως η φυσική επιλογή αποτελεί μια διαδικασία εξέλιξης μέσω της οποίας οι οργανισμοί που είναι καλύτερα προσαρμοσμένοι στο περιβάλλον αφήνουν περισσότερους απογόνους από εκείνους που είναι λιγότερο προσαρμοσμένοι. Αντίθετα η θεωρία της *γενετικής παρέκκλισης* υποστηρίζει την τυχαία αλλαγή της γονιδιακής συχνότητας, δηλαδή της συχνότητας με την οποία εμφανίζεται ένα αλληλόμορφο σ' ένα πληθυσμό. Όταν λέμε τυχαία εννοούμε την αλλαγή που δεν οφείλεται σε φυσική επιλογή. Η γενετική παρέκκλιση θεωρείται πολύ σημαντική εξελικτικά, δεδομένου ότι οδηγεί σε τυχαία αύξηση ή ελάττωση των συχνοτήτων των αλληλομόρφων από γενιά σε γενιά, στην τυχαία εγκαθίδρυση ή και απώλειά τους ενώ είναι εντονότερη στους μικρότερους πληθυσμούς (Κοσσίδα, 2008).

Ένα πολύ σημαντικό φαινόμενο που συμβάλλει στην εξέλιξη των ειδών τροποποιώντας τη γενετική σύσταση αυτών, είναι οι μεταλλάξεις. Οι μεταλλάξεις είναι

λάθη/τροποποιήσεις του DNA και προκαλούνται από διάφορες αιτίες . Το γενετικό υλικό δεν μεταβιβάζεται πιστά από τη μία γενιά στην άλλη ενώ είναι σημαντικό να αναφέρουμε ότι οι μεταλλάξεις είναι η πηγή των πολυμορφισμών στους πληθυσμούς. Ο ρυθμός μετάλλαξης διαφέρει από είδος σε είδος, ακόμα και σε διαφορετικές περιοχές του ίδιου γονιδιώματος, ενώ θ έσεις με αυξημένο ρυθμό μετάλλαξης ονομάζονται μεταλλακτικά hotspots, π.χ. οι θέσεις CpG στα θηλαστικά παρουσιάζουν αυξημένο μεταλλακτικό ρυθμό. Εκτός από τις μεταλλάξεις υπάρχουν και άλλα φαινόμενα που συμβάλλουν στην γενετική ποικιλότητα όπως ο ανασυνδυασμός, η μετανάστευση ή γονιδιακή ροή (εισροή νέων αλληλομόρφων σε ένα πληθυσμό) και τα *μεταθετά στοιχεία*. Τα τελευταία είναι γενετικά στοιχεία (γενετικές αλληλουχίες), που συναντούμε και στους προκαρυωτικούς και στους ευκαρυωτικούς οργανισμούς, και που έχουν την ικανότητα να αλλάξουν την θέση τους μέσα στο γονιδίωμα. Είναι ικανά να δημιουργήσουν νέα γονίδια και μπορούν να επηρεάσουν την έκφραση γονιδίων (Κοσσίδα, 2008).

2.2 ΟΙ ΜΕΤΑΛΛΑΞΕΙΣ ΣΤΟΥΣ ΠΛΗΘΥΣΜΟΥΣ ΚΑΙ Η ΘΕΩΡΙΑ ΤΗΣ ΟΥΔΕΤΕΡΟΤΗΤΑΣ

Οι μεταλλάξεις διαίρουνται σε τρεις γενικές κατηγορίες ανάλογα με το πώς τις αντιμετωπίζει η φυσική επιλογή: στις *ευνοϊκές* (*advantageous*), στις *ουδέτερες* (*neutral*), και στις *επιβλαβείς* (*deleterious*). *Ευνοϊκές* είναι οι μεταλλάξεις που προσφέρουν εξελικτικό πλεονέκτημα σ τους οργανισμούς που τις φέρουν και άρα η εξέλιξη χαρακτηρίζεται θετική. *Ουδέτερες* είναι οι μεταλλάξεις που δεν δίνουν ούτε πλεονέκτημα ούτε μειονέκτημα στον οργανισμό που τις φέρει. Είναι επιλεκτικά αδιάφορες (αφού η επίδραση τους στην επιβίωση και την αναπαραγωγή είναι αμελητέα). Τέλος, *επιβλαβείς* είναι οι μεταλλάξεις αυτές που κάνουν τον οργανισμό που τις φέρει ασθενέστερο έναντι των οργανισμών που δεν φέρουν τη μετάλλαξη (εντός ενός συγκεκριμένου περιβάλλοντος). Τις (πολύ) επιβλαβείς μεταλλάξεις δεν θα τις δούμε με μεγάλη συχνότητα σε πληθυσμούς, γιατί η φυσική επιλογή τις απομακρύνει από τους πληθυσμούς μαζί με τους φορείς τους.

Το 1968 ο Motoo Kimura διατύπωσε εμπειριστατωμένα την *Θεωρία της Ουδετερότητας* υποστηρίζοντας ότι οι περισσότερες μεταλλάξεις που συμβαίνουν δεν

έχουν κάποιο αποτέλεσμα στην προσαρμοστικότητα του ατόμου. Η σημασία της Θεωρίας της Ουδετερότητας στην εξέλιξη είναι σημαντική. Πρώτον, διότι έδωσε νέα διάσταση στην εξελικτική διαδικασία, τονίζοντας το ρόλο των ουδέτερων μεταλλάξεων. Δεύτερον, η εμφάνιση της Θεωρίας της Ουδετερότητας συνδέθηκε με την ανάπτυξη μαθηματικών μοντέλων και ελέγξιμων (άρα επιστημονικών) υποθέσεων γύρω από τη Θεωρία της εξέλιξης, καθώς επίσης έδωσε τη μηδενική υπόθεση (null hypothesis) για να ελεγχθεί η ύπαρξη ή όχι επιλογής στα μοριακά δεδομένα (Κοσσίδα, 2008).

2.3 ΕΞΕΛΙΚΤΙΚΕΣ ΣΧΕΣΕΙΣ – ΟΡΘΟΛΟΓΑ ΚΑΙ ΠΑΡΑΛΟΓΑ ΓΟΝΙΔΙΑ

Το πρώτο βήμα στη μελέτη των εξελικτικών σχέσεων των ακολουθιών είναι η αναζήτηση ομοιοτήτων με ακολουθίες που υπάρχουν ήδη σε βάσεις δεδομένων. Οι ακολουθίες που μοιράζονται ένα βαθμό ομοιότητας ενδέχεται να είναι *ομόλογες* (*homologs*). Οι ομόλογες ακολουθίες προέρχονται από μια κοινή προγονική ακολουθία και υπάρχει πιθανότητα να έχουν διατηρήσει μια κοινή προγονική βιοχημική λειτουργία. Η σύγκριση ομολόγων επιτρέπει τη διεξαγωγή συμπερασμάτων αναφορικά με την εξελικτική ιστορία και τη λειτουργικότητα των ακολουθιών. Συνεπώς, είναι δυνατό να αποδώσουμε πιθανές λειτουργίες στις νέες ακολουθίες βασιζόμενοι στην ομολογία τους με ακολουθίες πειραματικά αποδεδειγμένης βιοχημικής λειτουργίας. Υπάρχουν διάφορα είδη ομολόγων ακολουθιών:

- *Ορθόλογες* (*orthologs*) είναι οι ομόλογες ακολουθίες προερχόμενες από γεγονότα ειδογένεσης (*speciation*). Συνήθως έχουν κοινή λειτουργία.
- *Παράλογες* (*paralogs*) είναι οι ομόλογες ακολουθίες που προέρχονται από γονιδιακούς διπλασιασμούς (*gene duplication*). Συνήθως έχουν διαφορετικές λειτουργίες.
- *Ξενόλογες* (*xenologs*) είναι οι ομόλογες ακολουθίες που προκύπτουν από οριζόντια μεταφορά γενετικού υλικού (*horizontal gene transfer*) μεταξύ δύο ειδών (Κοσσίδα, 2008).

2.4 ΡΥΘΜΟΣ ΕΞΕΛΙΞΗΣ ΤΩΝ ΑΛΛΗΛΟΥΧΙΩΝ - ΜΙΤΟΧΟΝΔΡΙΩΝ

Όσον αφορά τον ρυθμό εξέλιξης των αλληλουχιών, είτε πρόκειται για αμινοξικές είτε για νουκλεοτιδικές αλληλουχίες, διαφέρει αρκετά από είδος σε είδος, ενώ έχει αποδειχθεί ότι μερικές αλληλουχίες DNA εξελίσσονται με ρυθμούς πολύ υψηλότερους από άλλες. Πχ οι αλληλουχίες των μη λειτουργικών ή σχεδόν μη λειτουργικών πολυπεπτιδίων, όπως είναι το πεπτιδίο C (που απομακρύνεται από την προ-προϊνσουλίνη όταν αυτή μετατρέπεται και σχηματίζει ινσουλίνη) εξελίσσονται με υψηλότερους ρυθμούς από τις πρωτεΐνες εκείνες που βρίσκονται κάτω από λειτουργικούς περιορισμούς. Κατά την διάρκεια μεγάλων χρονικών περιόδων, μπορεί να συμβούν πολλαπλές αντικαταστάσεις βάσεων στην ίδια θέση, με αποτέλεσμα ο αριθμός των διαφορών που παρατηρείται μεταξύ των ορθόλογων γονιδίων δύο ειδών που έχουν ένα μακρινό κοινό πρόγονο, να είναι μικρότερος από τον αριθμό των αντικαταστάσεων που έχουν στην πραγματικότητα συμβεί.

Για το μιτοχονδριακό DNA των θηλαστικών, η σχέση μεταξύ της διαφοράς των αλληλουχιών και του χρόνου διαχωρισμού είναι γραμμική για 5-10 εκατομμύρια χρόνια περίπου, γιατί μέσα σ' αυτό το χρονικό διάστημα οι περισσότερες αντικαταστάσεις συμβαίνουν σε διαφορετικές θέσεις. Σ' αυτό το σημείο είναι πολύ σημαντικό να αναφερθεί ότι ο ρυθμός μετάλλαξης του mtDNA των θηλαστικών είναι κατά πολύ μεγαλύτερος από αυτόν του πυρηνικού (100 φορές υψηλότερος ρυθμός) (*Chial et al., 2008*) και συνεπώς το mtDNA εξελίσσεται ταχύτερα από το πυρηνικό DNA. (*Nabholz et al., 2008*). Γιατί όμως συμβαίνει αυτό?

Γιατί άραγε το mtDNA υπόκειται σε τόσες πολλές μεταλλάξεις? Αυξημένη εμφάνιση μεταλλάξεων οφείλεται στη παραγωγή ελευθέρων ριζών οξυγόνου στα μιτοχόνδρια. Τα μιτοχόνδρια κατά τη διαδικασία της οξειδωτικής φωσφορυλίωσης προκειμένου να παράγουν ενέργεια με τη μορφή ATP υφίστανται βλάβη, μέσω της παραγωγής ελευθέρων ριζών οξυγόνου. Έτσι τα ίδια τα μιτοχόνδρια είναι πηγή παραγωγής ελευθέρων ριζών. Οι ελεύθερες ρίζες είναι ομάδες ατόμων τα οποία περιέχουν ένα μονήρες ηλεκτρόνιο και θεωρούνται μόρια καταστρεπτικά για το mtDNA.

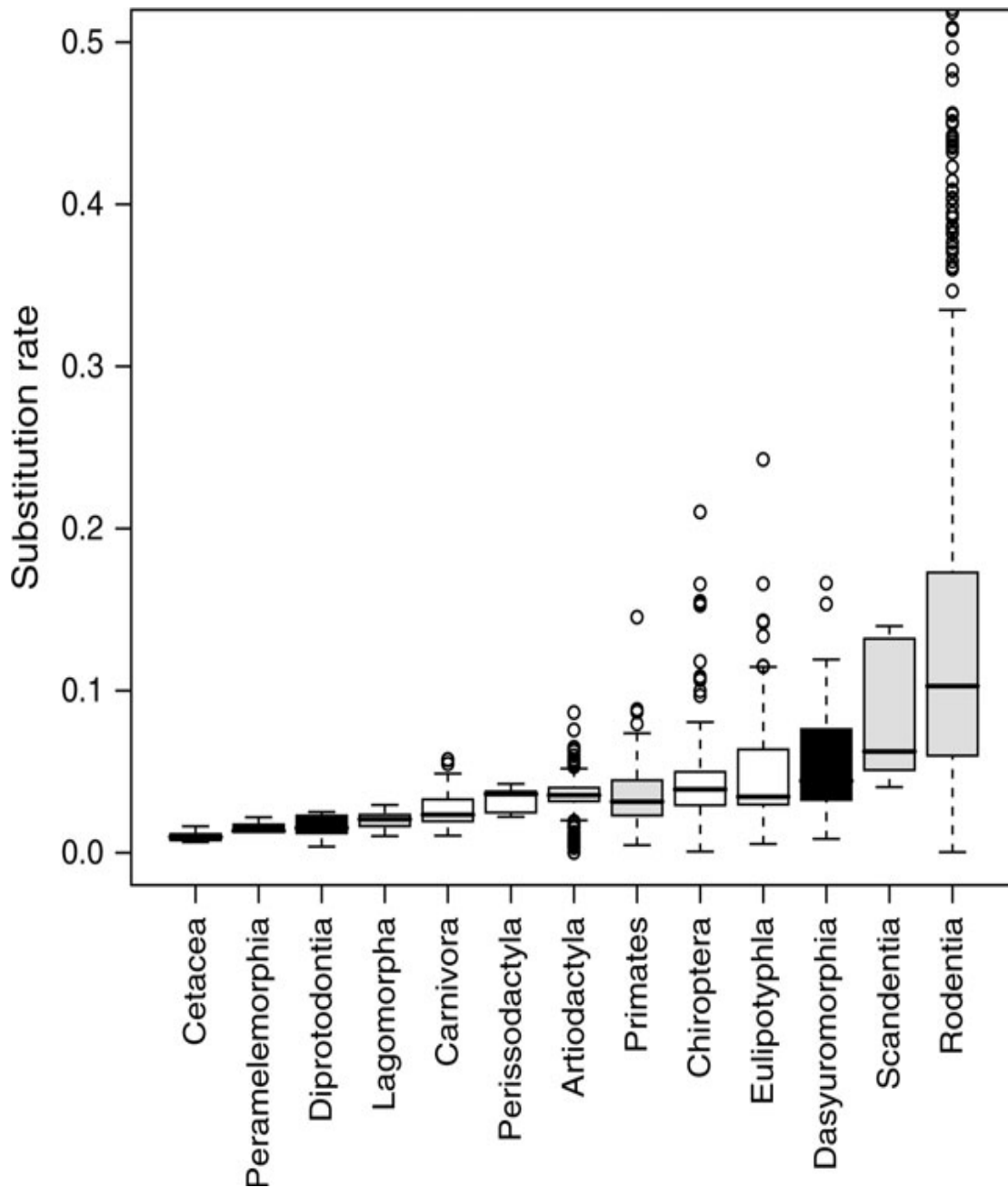
Πολύ πρόσφατα και μετά από πολλές μελέτες έχει αποδειχθεί η ύπαρξη ενός πυρηνικού γονιδίου το οποίο ονομάζεται γ-DNA πολυμεράση (POLG) το οποίο

κωδικοποιεί τη DNA πολυμεράση που είναι υπεύθυνη για την αντιγραφή του μιτοχονδριακού DNA. Η συγκεκριμένη πρωτεΐνη αποτελείται από δύο τμήματα: μία καταλυτική περιοχή με δράση ενδονουκλεάσης και μία δεύτερη περιοχή που εμπλέκεται στην αναγνώριση και απομάκρυνση των DNA-mismatches που συμβαίνουν κατά την διάρκεια της αντιγραφής του DNA και έχει δράση εξονουκλεάσης. Μία μελέτη έδειξε ότι στα μιτοχόνδρια μπορεί να παρατηρηθεί μια νουκλεοτιδική ανισσοροπία που οδηγεί στην αναστολή δράσης της γ -DNA πολυμεράσης, με αποτέλεσμα να μην γίνονται επιδιορθώσεις του γενώματος με άμεση συνέπεια την ύπαρξη πολλών λαθών στο mtDNA και εμφάνιση υψηλού ρυθμού μεταλλάξεων (Εικ.2.1) (Chial *et al.*, 2008).

dNTP pool used	Wild-type pol γ		Exonuclease-deficient pol γ	
	Mutant frequency $\times 10^{-4}$	Relative frequency	Mutant frequency $\times 10^{-4}$	Relative frequency
Equimolar dNTP pools, μ M				
1,000	11	1.4	62	1.4
100	ND	—	56	1.2
10	ND	—	48	1.1
1	7.8	1.0	45	1.0
Intramitochondrial dNTP pools				
Heart, SSM	23	2.9	160	3.6
Heart, IFM	21	2.7	170	3.8
Liver	≤ 22	—	42	0.9
Brain	ND	—	60	1.3
Skeletal muscle	13	1.7	270	6.0

Εικόνα 2.1 Απεικόνιση συχνότητας μεταλλάξεων σε περιπτώσεις όπου παρατηρείται λειτουργική ανωμαλία στη περιοχή της εξονουκλεάσης της POLG με αποτέλεσμα να μην πραγματοποιούνται επιδιορθώσεις και άρα να αυξάνεται ο μεταλλακτικός ρυθμός (Chial *et al.*, 2008).

Όπως αναφέρθηκε στην προηγούμενη παράγραφο ο ρυθμός μετάλλαξης του mtDNA των θηλαστικών είναι κατά πολύ ταχύτερος σε σχέση με το πυρηνικό γονιδίωμα (έως και 100 φορές) (Nabholz *et al.*, 2008). Συνεπώς το γεγονός αυτό συμβάλλει στην εξέλιξη των ειδών με την πάροδο του χρόνου. Εξελίσσονται όμως όλα τα θηλαστικά με τον ίδιο ρυθμό; Μετά από πολλές μελέτες και γενωμικές αναλύσεις οι ερευνητές κατέληξαν στο συμπέρασμα πως δεν έχουν τον ίδιο ρυθμό εξέλιξης όλα τα θηλαστικά. Έτσι λοιπόν υπάρχουν ισχυρές παραλλαγές του μιτοχονδριακού ρυθμού μετάλλαξης (Nabholz *et al.*, 2007) (Εικ.2.2)



Εικόνα 2.2 Διακυμάνσεις του ρυθμού μετάλλαξης ανάμεσα σε θηλαστικά. Στον κατακόρυφο άξονα αναφέρεται ο ρυθμός των υποκαταστάσεων ενώ στον οριζόντιο διάφορες οικογένειες θηλαστικών. Για παράδειγμα η οικογένεια Rodentia σε βάθος χρόνου εμφανίζει μεγαλύτερο ρυθμό μεταλλάξεων (Nabholz *et al.*, 2007).

Ο ρυθμός εξέλιξης μεταξύ δύο ορθόλογων γονιδίων φαίνεται από τον λόγο dN/dS. Η μελέτη του λόγου dN/dS αποτελεί ένα θεμελιώδη τρόπο μελέτης και παρατήρησης της φυσικής επιλογής. Οι νουκλεοτιδικές αντικαταστάσεις στα γονίδια που κωδικοποιούν πρωτεΐνες μπορεί να είναι είτε συνώνυμες (S)-Synonyms που σημαίνει πως οι αντικαταστάσεις αυτές δεν αλλάζουν το αμινοξύ, άρα και τη λειτουργία

της πρωτεΐνης (ονομάζονται και σιωπηλές μεταλλάξεις), είτε μη-συνώνυμες (N)-Non Synonyms που σημαίνει πως οι αντικαταστάσεις αυτές αλλάζουν το αμινοξύ και πιθανόν και τη λειτουργία της πρωτεΐνης. Ο λόγος που ερευνούμε τις συνώνυμες και μη συνώνυμες αντικαταστάσεις και κατ' επέκταση το λόγο dN/dS είναι για να παρατηρήσουμε το τύπο επιλογής που υφίσταται ένα γονίδιο που κωδικοποιεί μια πρωτεΐνη. Υπάρχουν τρεις περιπτώσεις πιθανών αποτελεσμάτων του λόγου dN/dS :

- $dN/dS < 1$
- $dN/dS = 1$ ή $dN/dS \sim 1$
- $dN/dS > 1$

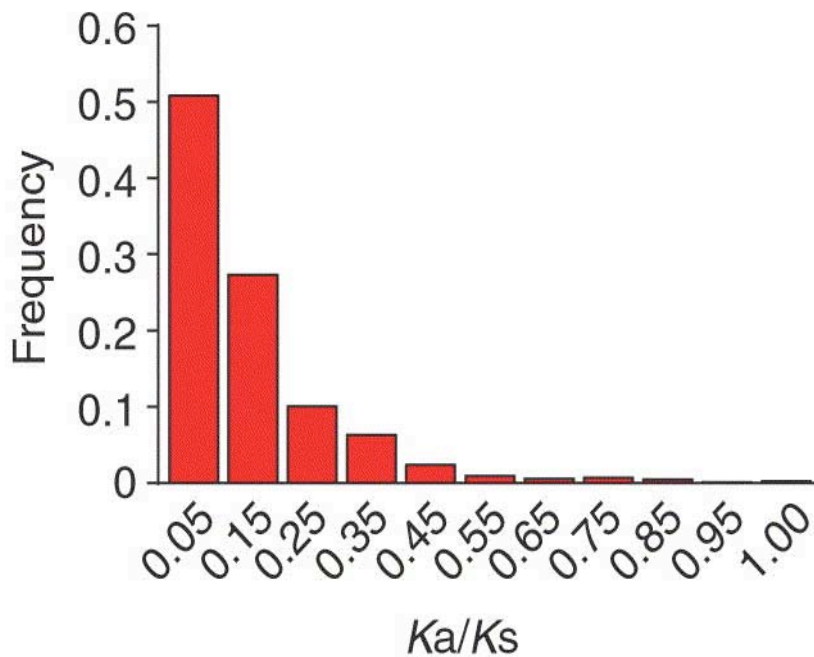
Τι φανερώνει όμως η κάθε περίπτωση?

Όταν το $dN/dS < 1$, σημαίνει πως οι μη συνώνυμες υποκαταστάσεις είναι συνήθως επιβλαβείς και απομακρύνονται από τον πληθυσμό, μέσω της φυσικής επιλογής (αρνητική επιλογή).

Όταν το $dN/dS = 1$ ή ~ 1 , δείχνει ότι οι μη συνώνυμες υποκαταστάσεις σε μεγάλο βαθμό είναι ουδέτερες. Ουδέτερες είναι οι μεταλλάξεις που δεν δίνουν ούτε πλεονέκτημα ούτε μειονέκτημα στον οργανισμό που τις φέρει. Είναι επιλεκτικά αδιάφορες (αφού η επίδραση τους στην επιβίωση και την αναπαραγωγή είναι αμελητέα).

Όταν το $dN/dS > 1$, αναφέρεται σε περιπτώσεις όπου οι μη συνώνυμες υποκαταστάσεις δίνουν ένα εξελικτικό πλεονέκτημα και αμέσως εγκαθιδρύονται στον πληθυσμό.(θετική επιλογή).

Σε πολύ λίγες μόνο περιπτώσεις η τιμή του dN είναι πολύ μεγαλύτερη αυτής του dS ($dN/dS \gg 1$) (Εικ 2.3) και αυτό το φαινόμενο αποτελεί ισχυρή ένδειξη ότι η φυσική επιλογή ευνοεί την αλλαγή της πρωτεΐνης (θετική επιλογή). Το φαινόμενο αυτό είναι πολύ σπάνιο και εμφανίζεται κυρίως σε περιπτώσεις όπως π.χ. τα γονίδια που συμμετέχουν στο ανοσοποιητικό σύστημα συν-εξελίσσονται με παράσιτα (Hurst, 2002).



TRENDS in Genetics

Εικόνα 2.3 Η κατανομή των διαφορετικών τιμών dN/dS 835 ζευγών ορθόλογων γονιδίων ποντικού – αρουραίου (Hurst, 2002).

2.5 ΦΥΛΟΓΕΝΕΤΙΚΗ ΑΝΑΛΥΣΗ

Η φυλογενετική είναι η εκτίμηση της εξελικτικής ιστορίας και η φυλογενετική ανάλυση είναι το μέσο μελέτης της. Η φυλογενετική ανάλυση βασίζεται στις πολλαπλές στοιχίσεις νουκλεϊνικών ή αμινοξικών ακολουθιών, πάνω στις οποίες βασίζονται τα προγράμματα κατασκευής φυλογενετικών δένδρων (Κοσσίδα, 2008).

2.5.1 Πολλαπλή στοιχίση ακολουθιών

Η φυλογενετική ανάλυση βασίζεται στη πολλαπλή στοιχίση ακολουθιών (multiple sequence alignment), που είναι η στοιχίση των ο μολογων θέσεων περισσότερων από δύο νουκλεϊνικών ή αμινοξικών ακολουθιών (Εικ.2.4) (Κοσσίδα, 2008).

```

Q5E940 BOVIN -----MPREDRATWKSNYFLKIITQLDDYPKCFIVGADNVGSKMQOIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE 76
RLA0 HUMAN -----MPREDRATWKSNYFLKIITQLDDYPKCFIVGADNVGSKMQOIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE 76
RLA0 MOUSE -----MPREDRATWKSNYFLKIITQLDDYPKCFIVGADNVGSKMQOIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE 76
RLA0 RAT -----MPREDRATWKSNYFLKIITQLDDYPKCFIVGADNVGSKMQOIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE 76
RLA0 CHICK -----MPREDRATWKSNYFLKIITQLDDYPKCFIVGADNVGSKMQOIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--SALE 76
RLA0 RANSY -----MPREDRATWKSNYFLKIITQLDDYPKCFIVGADNVGSKMQOIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE 76
Q7ZUG3 BRARE -----MPREDRATWKSNYFLKIITQLDDYPKCFIVGADNVGSKMQOIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE 76
RLA0 ICTPU -----MPREDRATWKSNYFLKIITQLDDYPKCFIVGADNVGSKMQOIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE 76
RLA0 DROME -----MYRENKAAKQYFIKVVYLFDEFPKCFIVGADNVGSKMQOIRMSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE 76
RLA0 DICDI -----MSGAE-SKRRKLEIEKATKLETTDKMIVAEADVFGSSELQIRKRSIRGI-GAVLMGKNTMIRKVIIRDADSK--PELD 75
Q54LP0 DICDI -----MSGAE-SKRRKLEIEKATKLETTDKMIVAEADVFGSSELQIRKRSIRGI-GAVLMGKNTMIRKVIIRDADSK--PELD 75
RLA0 BLRFB -----MKLSKQKQKQMYEIKESSELQQQSKLIVVIVDVMGSMASVYKSLRGK-AVVLMGKNTMMRKAIRGHLENN--PALE 76
RLA0 SULAC -----MILAVTTEKIKAKKVVDEVAELIKIKIKITITIANIEGFPADKLHEIRKRLGK-ADIKVYKNNLFIALKNAG--PDLE 79
RLA0 SULTO -----MRIMAVITQRKIAKWKIEEVKLELKRREYHTIIIANIEGFPADKLHEIRKRLGK-AEIKVYKNNLFIALKNAG--LDVS 80
RLA0 SULSO -----MKRLALALKQRKVASWLEEVKLELTKNSNTILGNLEGFPADKLHEIRKRLGK-AEIKVYKNNLFIALKNAG--LDIE 80
RLA0 AERPE MSVVSIVGQMYKREKPIPEWKTLMLELELFSKRVVLFADLTGPFVYVQRVKKLEWKK-PMVAKKRITLIRAKNAGLE--LDDN 86
RLA0 PYRAE -MMLAIGKRRYVTRQYVARKVKIVSEATELQKQYVYVFDLHGLSRILHEVRYRLRY-GLIKIKPFLFKIAFTKVVGG--TFAE 85
RLA0 METAC -----MAEERHTEHPQWKDEIENIKELIQSHKVFQMGVIEGILATKMKIRRDLDKV-AVLKVRNLLERALNQLG--ETIP 78
RLA0 METMA -----MAEERHTEHPQWKDEIENIKELIQSHKVFQMGVIEGILATKMKIRRDLDKV-AVLKVRNLLERALNQLG--ESID 78
RLA0 ARCFU -----MAAVRGS--PPEYKRAVEEIKRMISSEKVVVAIVSFRNVPAGOMKIRREFRGK-AEIKVYKNNLFIALKNAG--GDYL 75
RLA0 METKA MAVKAKGQPPSCYEPAVAEWRKREVKLELMDENYVGLVLELGIIPAPLOEIRAKLRERDITIRMSKNTLMRIRALEKLDER--PELE 88
RLA0 METH -----MAHVAEKKKEVQELHDLIKGVEVVGIANLADIPARLOKMMRQTLRDS-ALIRMSKNTLISLALKEKREL--ENVD 74
RLA0 METTL -----MITASEHKIAPWKIEEYNKLELLKNGQIVALVDMMEVVPARLDEIRDKIR-ETMFLKMSRNTLIEIRAKIEVAEETGNFEFA 82
RLA0 METVA -----MIDAKSEHKIAPWKIEEYNKLELLKSNYVIALIDMMEVVPARLDEIRDKIR-DQMLKMSRNTLIEIRAKIEVAEETGNFEFA 82
RLA0 METJA -----METKVAHVAEWEIEEYKTLKGLIKSNVVAIVDMMEVVPARLDEIRDKIR-DKVKLMSRNTLIEIRAKIEVAEENWPKLA 81
RLA0 PYRAB -----MAHVAEKKKEVEELAKLKSRYVIALVDVSSMPAYPLSQMRRLLIRENGGLLRVSRNLLIELAIKKVAEELKQFELE 77
RLA0 PYRHO -----MAHVAEKKKEVEELAKLKSRYVIALVDVSSMPAYPLSQMRRLLIRENGGLLRVSRNLLIELAIKKVAEELKQFELE 77
RLA0 PYRFU -----MAHVAEKKKEVEELAKLKSRYVIALVDVSSMPAYPLSQMRRLLIRENNGLLRVSRNLLIELAIKKVAEELKQFELE 77
RLA0 PYRKO -----MAHVAEKKKEVEELAKLKSRYVIALVDVAGVPAVYPLSKMRDKLS-GKALLRVSRNLLIELAIKKVAEELKQFELE 76
RLA0 HALMA -----MSAESERTETIPEWQEEVDVAIVMIESYVESVGVNIGVPIRQLDMMRDLHGCTAELRVSRNLLERALDQVD--DGLE 79
RLA0 HALVO -----MSSEVYRTEVIPQWKRREVDLELDFIESYVESVGVVAGVPIRQLDSMRRELHGCSAAVMSRNTLVNRRALDQVD--DGF 79
RLA0 HALSA -----MSSEEQRTTEVPEWKRQEAELVDLLETDSVGVVNVGTGIPKQLDMMRDLHGCTAALRVSRNLLERALEEAG--DGLD 79
RLA0 THEAC -----MKEVSOQKKELVNEITRIRKASRSVAIVDIEGIRSRQLDIDIRGKNRGK-INLKVYKNNLFIKALENLGD--EKLS 72
RLA0 THEVO -----MRKINPKKKEIVSELAQDITKSKAVAVIVDKGVRRLMDDIRAKNRDK-VKIKVYKNNLFIKALENLGD--EKLT 72
RLA0 PICTO -----MTEPRQWIDFVKNELEINSRKVAATVSIKGLRNNLFFKIRNSIRDK-ARIKVSARLLRLATENLQK--NNIV 72
RuLER 1.....10.....20.....30.....40.....50.....60.....70.....80.....90

```

Εικόνα 2.4 Πολλαπλή στοίχιση αμινοξικών ακολουθιών. (http://upload.wikimedia.org/wikipedia/commons/thumb/7/79/RPLP0_90_ClustalW_aln.gif/575px-RPLP0_90_ClustalW_aln.gif)

Από τους στοιχισμένους χαρακτήρες μιας πολλαπλής στοίχισης μπορούμε να εξάγουμε σημαντικά συμπεράσματα αναφορικά με την εξελικτική ιστορία των ακολουθιών, όπως για παράδειγμα για τις μεταλλάξεις που εμφανίστηκαν κατά την διάρκεια της εξέλιξης. Με τη βοήθεια των πολλαπλών στοίχισων των ακολουθιών μας δίνεται η δυνατότητα να προσδιορίσουμε τα αμινοξικά κατάλοιπα που είναι κρίσιμα για την ειδική λειτουργία των πρωτεϊνών. Παραδείγματος χάρη, εάν ένα αμινοξύ είναι συντηρημένο σε όλες τις στοιχισμένες πρωτεΐνες αυτό σημαίνει ότι το αμινοξύ είναι σημαντικό για τη βιοχημική λειτουργία, καθώς επίσης και για τη σταθεροποίηση της δευτεροταγούς και τριτοταγούς δομής των πρωτεϊνών (Κοσσίδα, 2008).

Ένας τρόπος για να αξιολογήσουμε εάν η στοίχιση είναι καλής ποιότητας και έχει πραγματικά βιολογική σημασία είναι η βαθμολόγηση της πολλαπλής στοίχισης. Ο βαθμός ή score είναι ενδεικτικός της σημαντικότητας της στοίχισης. Γενικά τα κενά βαθμολογούνται με αρνητικές τιμές ή ποινές κενών (gap penalty) και τα ταυτόσημα αμινοξικά κατάλοιπα βαθμολογούνται με υψηλότερη θετική τιμή από τα αμινοξέα με συντηρημένες φυσικοχημικές ιδιότητες. Τα sum-of-pairs (SP) είναι ένας συνηθισμένος τρόπος βαθμολόγησης πολλαπλής στοίχισης. Γενικά, ένα υψηλό SP-score υποδηλώνει ότι η στοίχιση είναι σωστή. Ένας εναλλακτικός τρόπος βαθμολόγησης είναι το Z-score. Για να προκύψει αυτός οι ακολουθίες στοιχίζονται και υπολογίζεται το αρχικό score. Εν συνεχεία, γίνεται τυχαία αναδιάταξη των ακολουθιών και υπολογίζεται ο μέσος όρος

και η τυπική απόκλιση των scores που προκύπτουν από τη σύγκριση των αναδιαταγμένων ακολουθιών. Η παραπάνω διαδικασία επαναλαμβάνεται αρκετές φορές (Κοσσίδα, 2008).

Ένα από τα πιο γνωστά και εύχρηστα προγράμματα πολλαπλής στοίχισης ακολουθιών, το οποίο διατίθεται ελεύθερα μέσω διαδικτύου, είναι το πρόγραμμα CLUSTALW, όπου το γράμμα «W» συμβολίζει τη στάθμιση των ακολουθιών (Κοσσίδα, 2008).

2.5.2 Φυλογενετικά δένδρα

Ο στόχος των φυλογενετικών δένδρων είναι να διευκρινίσει πως μια ομάδα γονιδίων ή οργανισμών, συνδέονται μεταξύ τους και να παρουσιάσει τις εξελικτικές τους σχέσεις. Η αναπαράσταση των φυλογενετικών σχέσεων που προκύπτουν από τη φυλογενετική ανάλυση γίνεται με τη χρήση ενός φυλογενετικού δένδρου το οποίο αποτελείται από τα εξής βασικά στοιχεία (Κοσσίδα, 2008) :

Κόμβος: ένα σημείο του κλαδιού σε ένα δένδρο (για υποτιθέμενη προγονική ταξινομική ομάδα)

Κλαδί ή βραχίονας: προσδιορίζει τη σχέση μεταξύ των ομάδων με όρους απογόνου και προγόνου

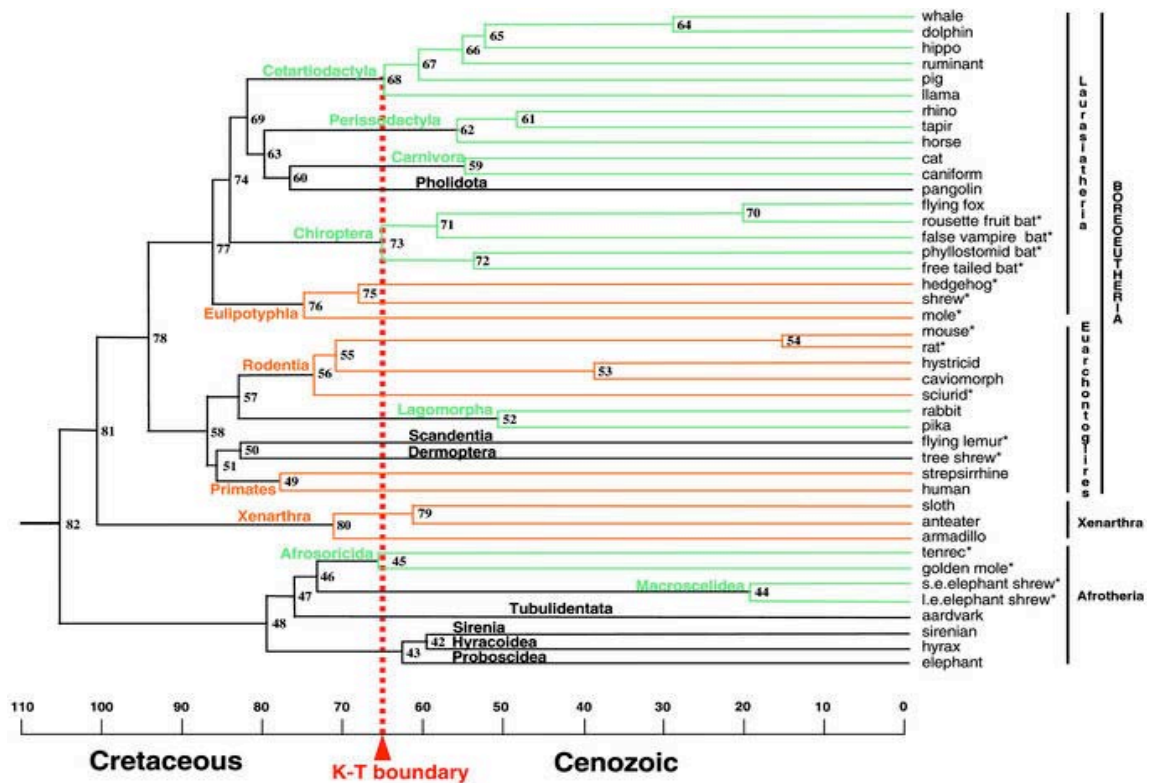
Τοπολογία: το πρότυπο διακλάδωσης του δένδρου

Μήκος κλαδιού: αντιπροσωπεύει τον αριθμό των αλλαγών που συνέβησαν στο κλαδί

Ρίζα: ο κοινός πρόγονος όλων των ομάδων του δένδρου

Κλάδος: μια ομάδα από δύο ή περισσότερες ομάδες ή DNA αλληλουχίες που περιλαμβάνει τον κοινό τους πρόγονο και όλους τους απογόνους τους

Λειτουργική Ταξινομική Ομάδα (ΛΤΟ): ταξινομικό επίπεδο του επιλεγμένου δείγματος που χρησιμοποιείται στη μελέτη, όπως άτομα, πληθυσμοί, είδη, γένη, ή βακτηριακά στελέχη (Εικ.2.5)



Εικόνα 2.5 Φυλογενετικό δένδρο που αναπαριστά τις εξελικτικές σχέσεις 42 πλακουντοφόρων θηλαστικών. Για τη συγκεκριμένη μελέτη έγινε πολλαπλή στοίχιση 16.397 ζευγών βάσεων και τέθηκαν προς ανάλυση 19 πυρηνικά και 3 mt γονίδια (Springer et al., 2003).

Υπάρχουν δύο κατηγορίες δένδρων. Τα δένδρα με ρίζα και αυτά χωρίς ρίζα. Στο δένδρο με ρίζα είναι σαφώς εμφανής ο πιο κοντινός κοινός πρόγονος από τον οποίο οι υπό εξέταση ταξινομικές μονάδες προήλθαν, ενώ στο δένδρο χωρίς ρίζα περιγράφεται η συσχέτιση μεταξύ των ταξινομικών μονάδων, χωρίς να γνωρίζουμε που βρίσκεται ο κοινός πρόγονος. Επίσης τα φυλογενετικά δένδρα διακρίνονται στα κλαδογράμματα και στα φυλογράμματα. Στο κλαδογράμμα φαίνεται η τοπολογία και όχι το μήκος των βραχιόνων, ενώ το φυλογράμμα ενσωματώνει και το μήκος των βραχιόνων όπου διαφαίνεται η εξελικτική διαφοροποίηση.

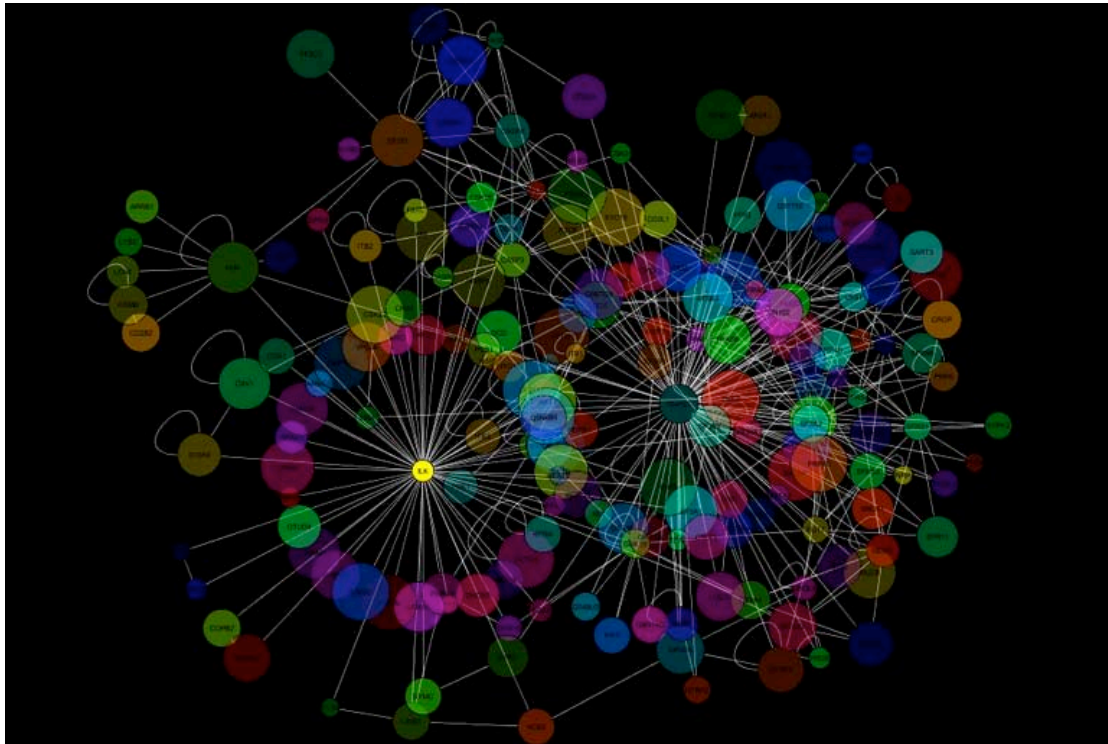
Οι μέθοδοι που χρησιμοποιούνται για την κατασκευή δένδρων μπορούν να κατηγοριοποιηθούν σε μεθόδους βασισμένες σε αποστάσεις και σε μεθόδους βασισμένες σε χαρακτήρες. Οι μέθοδοι απόστασης υπολογίζουν τις αποστάσεις κάθε ζεύγους ακολουθιών και δημιουργείται ένας πίνακας αποστάσεων. Στη συνέχεια, οι ακολουθίες (με βάση τον πίνακα αποστάσεων) ομαδοποιούνται και κατασκευάζεται ένα δένδρο. Οι μέθοδοι που βασίζονται σε χαρακτήρες αναζητούν δένδρα που βελτιστοποιούν την προσαρμογή των δεδομένων των χαρακτήρων σε ένα φυλογενετικό

δένδρο, χρησιμοποιούν τα δεδομένα των χαρακτήρων με τέτοιο τρόπο ώστε η θέση κάθε χαρακτήρα να εκτιμάται σε συνάρτηση με τις θέσεις των υπολοίπων χαρακτήρων. Για την αξιολόγηση της σταθερότητας ενός φυλογενετικού δένδρου και της αξιοπιστίας των κλάδων αυτού χρησιμοποιείται το *Bootstrap*. Ένα από τα ευρύτερα χρησιμοποιούμενα και περιεκτικότερα προγράμματα φυλογενετικής ανάλυσης είναι το PHYLIP (PHYLogeny Inference Package) (<http://evolution.genetics.washington.edu/phylip.html>), το οποίο κατασκευάζει δένδρα με μεθόδους απόστασης, μέγιστης οικονομίας και μέγιστης πιθανοφάνειας (Κοσσίδα, 2008).

2.6 Πρωτεϊνικές αλληλεπιδράσεις και πρωτεϊνικά δίκτυα

Εισαγωγή

Οι πρωτεΐνες δεν λειτουργούν ως αυτόνομες μονάδες μέσα στο κύτταρο, αλλά αλληλεπιδρούν με το μοριακό περιβάλλον τους, ώστε να επάγουν πληροφορίες στον πυρήνα ή να σχηματίσουν πρωτεϊνικά σύμπλοκα, τα οποία με τη σειρά τους θα λειτουργήσουν ως πολύπλοκες μοριακές μηχανές (Amoutzias and Van de Peer, 2010). Η λειτουργικότητα του κυττάρου βασίζεται στις επαφές και αλληλεπιδράσεις μεταξύ των πρωτεϊνών. Επομένως, είναι φυσικό επακόλουθο η μεγάλη προσπάθεια που συντελείται την τελευταία δεκαετία ώστε να χαρακτηριστεί το σύνολο των αλληλεπιδράσεων όλων των πρωτεϊνών ενός γονιδιώματος, το οποίο στη διεθνή ορολογία ονομάζεται αλληλεπίδρωμα (*interactome*). Το σύνολο των αλληλεπιδράσεων μεταξύ πρωτεϊνών (*protein-protein interactions: PPIs*) μελετάται με τη βοήθεια της θεωρίας γράφων (*graph theory*) ως ένα δίκτυο πρωτεϊνικών αλληλεπιδράσεων (*protein interaction network: PIN*). Σε αυτά τα δίκτυα κάθε πρωτεΐνη αναπαρίσταται ως κόμβος, ο οποίος συνδέεται με άλλους κ όμβους-πρωτεΐνες, όταν αλληλεπιδρά μαζί τους, ή όταν συμμετέχουν από κοινού στο σχηματισμό ενός πρωτεϊνικού συμπλόκου (Εικ.2.6)



Εικόνα 2.6 Παράδειγμα ενός δικτύου πρωτεϊνικών αλληλεπιδράσεων της εξοκινάσης (protein interaction network). Οι κόμβοι αναπαριστούν τις πρωτεΐνες και οι συνδέσεις ενώνουν πρωτεΐνες που αλληλεπιδρούν μεταξύ τους. (http://fiehnlab.ucdavis.edu/staff/kumar/Cytoscape_photo/variant/large)

Η εξέλιξη των δικτύων των πρωτεϊνικών

Γενικά τα PINs μπορούν να εξελιχθούν είτε με μεταλλάξεις σε υπάρχοντα γονίδια, είτε με μεταλλάξεις σε αντίγραφα γονιδίων που διπλασιάστηκαν (Amoutzias and Van de Peer, 2010). Στην πρώτη περίπτωση ο αριθμός των κόμβων-πρωτεϊνών στο PIN παραμένει σταθερός ενώ κάποιες συνδέσεις αναπτύσσονται ή χάνονται. Στην δεύτερη περίπτωση, ο διπλασιασμός οδηγεί σε ένα ζεύγος από πανομοιότυπα γονίδια, τα οποία μπορούν να ακολουθήσουν μια από τις τρεις παρακάτω πορείες: i) οι λειτουργίες/αλληλεπιδράσεις της πρωτεΐνης του προγονικού γονιδίου να μοιραστούν μεταξύ του ζεύγους (subfunctionalization), ii) το ένα από τα δύο γονίδια διατηρεί τις λειτουργίες του προγονικού γονιδίου, ενώ το δεύτερο αφήνεται ελεύθερο να αναπτύξει κάποιες νέες λειτουργίες (neo-functionalization), iii) ένα από τα γονίδια διατηρεί τις λειτουργίες του προγονικού γονιδίου, ενώ το δεύτερο χάνεται (non-functionalization). Η τρίτη περίπτωση είναι και η πιο συνηθισμένη. Στην πρώτη και δεύτερη περίπτωση, όλες οι αλληλεπιδράσεις του προγονικού γονιδίου κληρονομούνται αυτούσιες στο ζεύγος και ανάλογα με την έκταση των μεταλλάξεων που συμβαίνουν αργότερα, ένας αριθμός

αλληλεπιδράσεων είναι δυνατόν να χαθεί, ενώ ένας άλλος αριθμός αλληλεπιδράσεων είναι δυνατόν να εμφανιστούν.

Οι διάφοροι τύποι αλληλεπιδράσεων

Οι πρωτεϊνικές αλληλεπιδράσεις δεν είναι όλες ίδιες μεταξύ τους, αλλά χωρίζονται σε ξεχωριστούς τύπους (Nooren and Thornton. 2003). Ο πιο σημαντικός διαχωρισμός είναι μεταξύ αλληλεπιδράσεων που σχηματίζουν μόνιμα ή πρόσκαιρα σύμπλοκα. Στην πρώτη περίπτωση, οι αλληλεπιδράσεις είναι πολύ ισχυρές και το σύμπλοκο διαλύεται μόνο με πρωτεόλυση. Στην δεύτερη περίπτωση, οι αλληλεπιδράσεις είναι ασθενείς, αλλά εξακολουθούν να εμφανίζουν εξειδίκευση, δηλαδή αλληλεπιδρούν με συγκεκριμένα μόρια, και όχι με οποιοδήποτε. Ιδιαίτερο χαρακτηριστικό των αλληλεπιδράσεων μονίμων συμπλόκων είναι, ο μεγάλος βαθμός εξελικτικής συντήρησης τους, ακόμα και μεταξύ πολύ διαφορετικών οργανισμών, εν αντιθέσει με τις αλληλεπιδράσεις πρόσκαιρων συμπλόκων που είναι λιγότερο συντηρημένες. Αυτό το γεγονός αντικατοπτρίζεται και στη συντήρηση των αμινοξέων που υπάρχουν στην επιφάνεια της πρωτεΐνης που κάνει την επαφή με άλλες αλληλεπιδρώσες πρωτεΐνες. Επιπλέον, οι πρωτεΐνες που σχηματίζουν ένα σταθερό σύμπλοκο γενικά τείνουν να συν-εκφράζονται πολύ περισσότερο, απ' ότι οι πρωτεΐνες που σχηματίζουν ένα πρόσκαιρο σύμπλοκο (Brown and Jurisica. 2007). Είναι βέβαια αξιοσημείωτο το γεγονός ότι το υπάρχον ανθρώπινο αλληλεπιδρώμα κυριαρχείται από αλληλεπιδράσεις που σχηματίζουν πρόσκαιρα σύμπλοκα (Brown and Jurisica, 2007).

ΚΕΦΑΛΑΙΟ 3

ΥΛΙΚΑ & ΜΕΘΟΔΟΙ

Για να πραγματοποιηθεί η βιοπληροφορική ανάλυση αυτής της εργασίας, χρησιμοποιήθηκαν μια σειρά από Βάσεις Δεδομένων και εργαλεία Βιοπληροφορικής, για τα οποία θα δωθεί μια σύντομη εισαγωγή (για το κάθε ένα) παρακάτω και πώς ανακτήθηκαν τα δεδομένα.

Στην παρούσα εργασία μελετήθηκαν τέσσερις κύριες ομάδες γονιδίων. Για να διευκολυνθεί η συγγραφή των αποτελεσμάτων, χρησιμοποιήθηκαν συντομεύσεις για αυτές τις τέσσερις ομάδες (*mt_mt genes*, *mt_genes*, *nuc_genes*, *mt_mt interactors*). Συγκεκριμένα:

- *mt_mt genes*: Τα γονίδια που κωδικοποιούνται στο μιτοχονδριακό DNA (και κωδικοποιούν πρωτεΐνες)
- *mt_genes*: Γονίδια πυρηνικής προελεύσεως που εμπλέκονται στη δομή ή/και λειτουργία του μιτοχονδρίου, σύμφωνα με τη Γονιδιακή Οντολογία (Gene Ontology).
- *nuc_genes*: Γονίδια πυρηνικής προελεύσεως που ΔΕΝ φαίνονται να εμπλέκονται στην δομή ή/και λειτουργία του μιτοχονδρίου, με βάση τη Γονιδιακή Οντολογία (Gene Ontology).
- *mt_mt interactors*: Γονίδια πυρηνικής προελεύσεως που με βάση την BioGrid αλληλεπιδρούν άμεσα με κάποιο/α από τα *mt_mt_genes*.

Τα βασικά στάδια της μελέτης ήταν:

- Να βρεθούν τα ανθρώπινα γονίδια για την κάθε μία από τις παραπάνω κατηγορίες, με τη βοήθεια της γονιδιακής οντολογίας.
- Να βρεθούν οι πρωτεϊνικές αλληλεπιδράσεις των παραπάνω ανθρώπινων γονιδίων από την BioGrid.
- Να βρεθούν ποιά ανθρώπινα γονίδια εμπλέκονται σε ασθένειες από την GeneCards.
- Να βρεθούν τα ορθόλογα των παραπάνω ανθρώπινων γονιδίων σε πέντε άλλα θηλαστικά, από την Ensembl, με τη χρήση του Biomart, καθώς επίσης να

εξαχθούν οι τιμές των dN/dS και του βαθμού συντήρησης των ορθόλογων ακολουθιών (% Identity).

- Να ενσωματωθούν όλα τα παραπάνω δεδομένα με τη βοήθεια της γλώσσας προγραμματισμού Perl και να γίνουν στατιστικές αναλύσεις με τη βοήθεια της γλώσσας προγραμματισμού R.

3.1 Γλώσσα Προγραμματισμού Perl.

Η *Perl* είναι μια πολύ δημοφιλής γλώσσα προγραμματισμού η οποία σχεδιάστηκε από τον Larry Wall ενώ η πρώτη έκδοση της γλώσσας εμφανίστηκε το 1987 (Pierce, 2005).

Συνήθως ένα πρόγραμμα σε Perl εκτελείται χρησιμοποιώντας άμεσα ή έμμεσα το διερμηνέα της γλώσσας. Αυτό που διακρίνει την Perl από πολλές άλλες γλώσσες προγραμματισμού είναι το γεγονός ότι είναι διαθέσιμη για σχεδόν όλα τα λειτουργικά συστήματα. Ως γλώσσα προγραμματισμού η Perl έχει ένα *δυναμικό σύστημα τύπων*, αυτό σημαίνει ότι μία μεταβλητή αποκτά τύπο μόνο μετά από την ανάθεση μιας τιμής σε αυτή. Η Perl μπορεί να διαχειριστεί αριθμούς (δεκαδικούς και ακεραίους) και συμβολοσειρές. Επιπλέον, μπορούμε να έχουμε πίνακες απλούς (arrays) και συσχετιστικούς (associative arrays ή hash tables), οι οποίοι είναι μηχανισμοί οργάνωσης δεδομένων. Όσον αφορά τις μεταβλητές υπάρχει ένας διαχωρισμός αυτών όσον αφορά την ανάθεση απλών ή σύνθετων τιμών.

Ένα άλλο βασικό χαρακτηριστικό της γλώσσας είναι η εξαιρετική υποστήριξη των *κανονικών εκφράσεων*, δηλαδή παραστάσεων που μπορούν να περιγράψουν την γενική δομή μιας συμβολοσειράς και μπορούν να χρησιμοποιηθούν στη διαχείριση συμβολοσειρών. Επιπλέον παρέχει υποδομές για δικτυακό προγραμματισμό και για την εύκολη διασύνδεση με γνωστές βάσεις δεδομένων, ενώ τα τελευταία χρόνια έγινε ιδιαίτερη δημοφιλής λόγω της χρήσης της στη *βιοπληροφορική*.

```
test8.pl (~/Desktop/ergasia) - VIM
File Edit View Terminal Help
#!/usr/bin/perl -w
open (FIN, "< ./export_lar_seq_ratus.txt") or die "failed"
open (FOUT, "> ./largest_sequences_ratus.txt") or die "fail
while ($line=<FIN>) {
    chomp ($line);
    if ($line=~m/^(?<(\S+)\|(\S+)/) {
        $gene=" ";
        $gene=$1;
        $transcript=" ";
        $transcript=$2;
    }
    $line=<FIN>;
    chomp ($line);
    $seq=" ";
    @array=();
    $length = 0;
    $seq=$line;

    @array=split(/,/,$seq);
    $length=@array;
    $hash1{$gene}[0] = 0;
    $hash1{$gene}[1] = " ";

close(FIN);

open (FIN, "< ./export_lar_seq_ratus.txt") or die "failed"
while ($line=<FIN>) {
    chomp ($line);
    if ($line=~m/^(?<(\S+)\|(\S+)/) {
        $gene=" ";
        $gene=$1;
        $transcript=" ";
    }
}
```

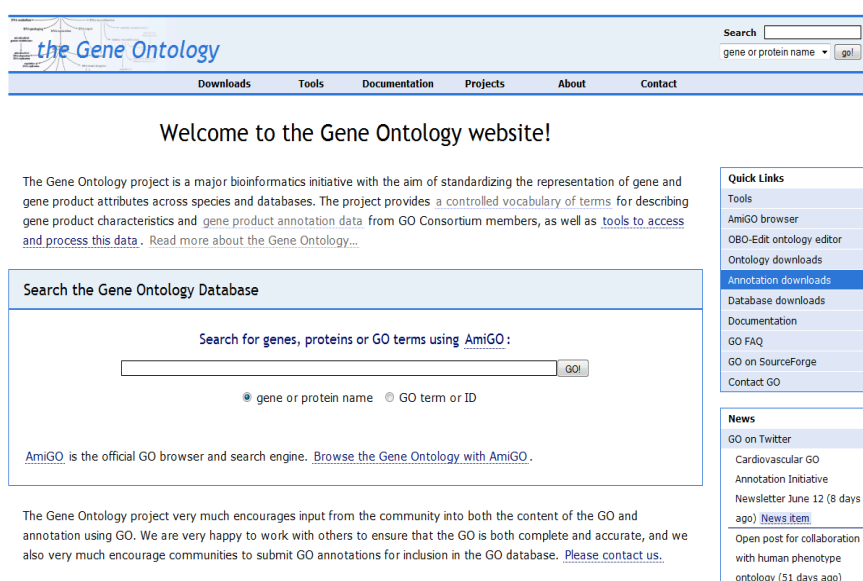
Εικόνα 3.1 Παράδειγμα ενός προγράμματος Perl που σχεδιάστηκε κατά την διάρκεια της εργασίας.

Η Perl χαρακτηρίζεται ως μια γλώσσα συγκόλλησης η οποία χρησιμοποιείται για να συνδέει πράγματα. Ένα αποτέλεσμα ότι η Perl είναι μια γλώσσα σχεδιασμένη για συγκόλληση πραγμάτων, είναι ότι η γλώσσα είναι πολύ προσαρμοστική. Εκτελείται μέσα σε τουλάχιστον 20 λειτουργικά συστήματα – και πιθανώς περισσότερα. Στη συγκεκριμένη εργασία χρησιμοποιήθηκαν τα UNIX. Το στιλ προγραμματισμού είναι πολύ ευέλικτο, οπότε μπορεί κάποιος προγραμματιστής να κάνει τα ίδια πράγματα κατά πολλούς τρόπους. Γενικώς η εγκατάσταση της Perl έχει σχεδιαστεί έτσι ώστε να είναι εύκολη και χωρίς σφάλματα. Τέλος αξίζει να σημειωθεί πως η Perl έχει χρησιμοποιηθεί σε κάθε κλάδο της βιομηχανίας, για σχεδόν οποιαδήποτε εργασία μπορούμε να φανταστούμε. Χρησιμοποιείται στην αγορά μετοχών, στις κατασκευές, στη σχεδίαση, στη βιοπληροφορική, στην υποστήριξη πελατών, στον έλεγχο ποιότητας, στον

προγραμματισμό συστημάτων, στην μισθοδοσία και στον έλεγχο αποθεμάτων. Φυσικά χρησιμοποιείτε και στο Web (Pierce, 2005).

3.2 Γονιδιακή Οντολογία(Gene Ontolog:GO)

Η Γονιδιακή Οντολογία ή gene ontology (GO) (<http://www.geneontology.org>) είναι μία σημαντική πρωτοβουλία της βιοπληροφορικής επιστήμης. Αποτελεί μία βάση δεδομένων (Β.Δ) η οποία περιλαμβάνει ένα ελεγχόμενο λεξιλόγιο. Αυτό το λεξιλόγιο έχει ως σκοπό την περιγραφή των γονιδίων και των χαρακτηριστικών των γονιδιακών προϊόντων (πρωτεϊνών) σε όλα τα είδη (Εικ.3.2)



Εικόνα 3.2 Αρχική σελίδα της Gene Ontology. (<http://www.geneontology.org>)

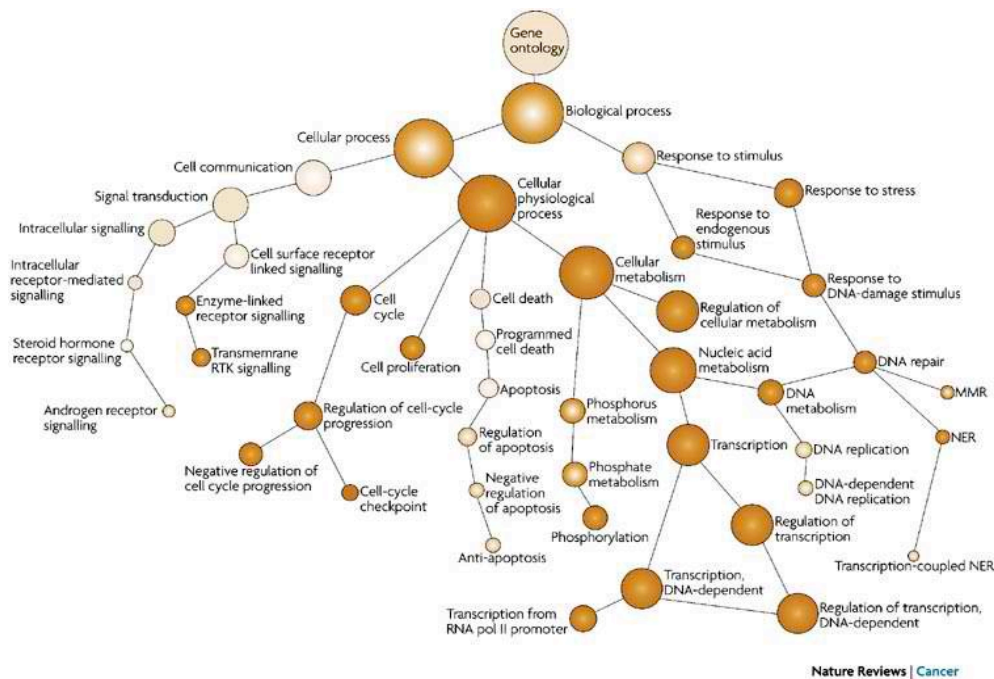
Πιο συγκεκριμένα αυτή η Β.Δ. στοχεύει:

- Στη διατήρηση και ανάπτυξη του λεξιλογίου που αφορά τα γονίδια και τα χαρακτηριστικά των γονιδιακών προϊόντων (πρωτεϊνών).
- Στο σχολιασμό των γονιδιακών δεδομένων και γονιδιακών προϊόντων, όπως επίσης και διάδοση αυτών.
- Στην παροχή εργαλείων για την εύκολη πρόσβαση σε όλες τις πτυχές των δεδομένων που παρέχονται από τη βάση.

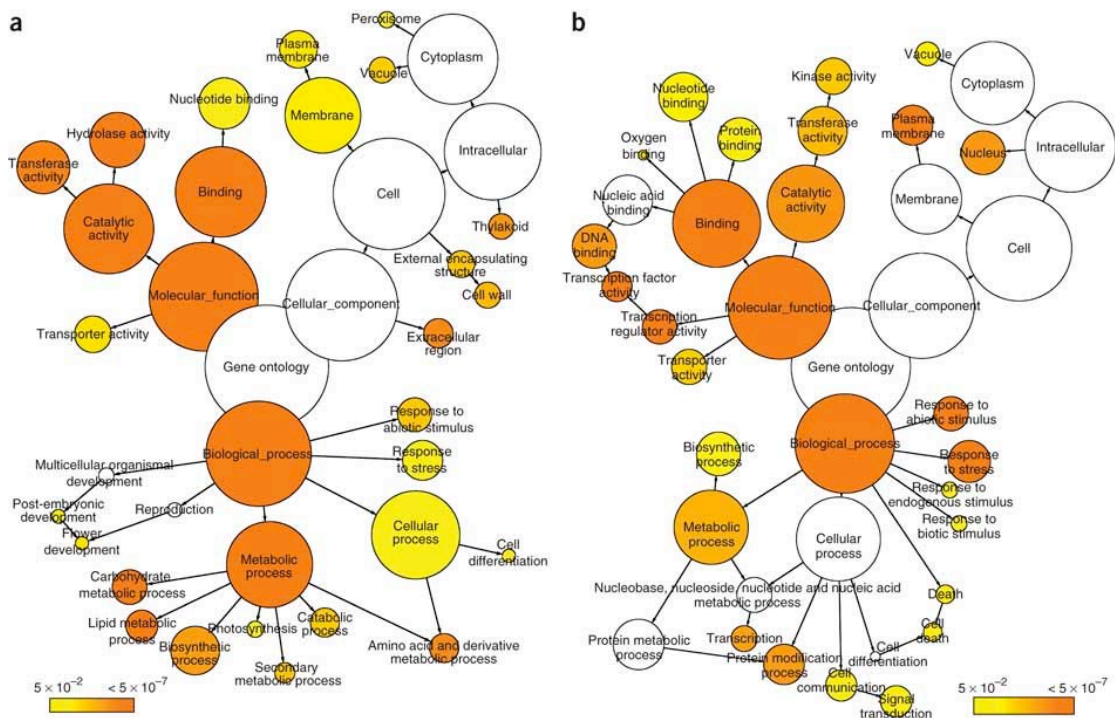
Η Β.Δ. της Γονιδιακής Οντολογίας παρέχει την οντολογία καθορισμένων όρων που αντιπροσωπεύουν ιδιότητες γονιδιακών προϊόντων. Η οντολογία καλύπτει τρεις τομείς:

- i. Την κυτταρική περιοχή (cellular component) που παρατηρείται ένα γονιδιακό προϊόν (βιομόριο). Π.χ μεμβράνες, οργανίδια κτλ.
- ii. Τη μοριακή λειτουργία του βιομορίου (molecular function). Αυτή περιλαμβάνει τις στοιχειώδεις δραστηριότητες και ιδιότητες των γονιδιακών προϊόντων σε μοριακό επίπεδο. Πχ .κατάλυση ή σύνθεση.
- iii. Τη βιολογική λειτουργία (biological process). Αυτή αφορά τις ολοκληρωμένες λειτουργίες του βιομορίου που συμβαίνουν σε μοριακό επίπεδο με καθορισμένη αρχή και τέλος και σχετίζονται με τις ολοκληρωμένες μονάδες διαβίωσης όπως είναι: τα κύτταρα, οι ιστοί, τα όργανα και οι οργανισμοί.

Όσον αφορά τη δομή της οντολογίας αυτή δείχνει τις σχέσεις μεταξύ των διαφορετικών όρων, ενώ ένας όρος μπορεί να αποτελεί μια πιο εξειδικευμένη περιγραφή ενός άλλου όρου.



Εικόνα 3.3 Γραφική αναπαράσταση Γονιδιακής Οντολογίας όπου διαφαίνονται οι σχολιασμοί λειτουργιών γνωστών καρκινικών γονιδίων. Το μέγεθος των κύκλων είναι ανάλογο με τη συνολική συμμετοχή του γονιδίου και η σκίαση δείχνει το βαθμό της στατιστικής σημαντικότητας. Αυτό σημαίνει ότι τα υπερεκφρασμένα γονίδια είναι εμπλουτισμένα για την «εμφάνιση του καρκίνου» (Hu *et al.*, 2007).



Εικόνα 3.4 Χαρτογράφηση Γονιδιακής Οντολογίας και λειτουργική περιγραφή των γονιδίων στη φράουλα. Παρατηρούνται κατηγορίες οντολογίας γονιδίων που εκφράζονται στα φρούτα (α) και ρίζα (β). Η ακτίνα του κάθε κύκλου δηλώνει τον αριθμό των γονιδίων σε κάθε κατηγορία. Η σκιά στους κύκλους είναι ανάλογη με το επίπεδο σημ. αντικότητας (κίτρινο, ποσοστό <math>< 0,05</math>). (<http://www.nature.com/ng/journal/v43/n2/images/ng.740-F3.jpg>).

Για παράδειγμα το γονιδιακό προϊόν κυτόχρωμα c, μπορεί να περιγραφεί, όσον αφορά τη μοριακή του λειτουργία με τον χαρακτηριστικό όρο : δραστηριότητα οξειδοοξειδωτικής, όσον αφορά τη βιολογική του λειτουργία: οξειδωτική φωσφορυλίωση, και όσον αφορά την κυτταρική περιοχή που παρατηρείται: μιτοχονδριακή μεμβράνη και εσωτερική μιτοχονδριακή μεμβράνη.

Κάθε όρος της Γονιδιακής Οντολογίας (gene ontology term) έχει ένα συγκεκριμένο όνομα εντός της οντολογίας το οποίο μπορεί να αφορά μια λέξη ή ένα χαρακτηρισμό, πχ: ένα μοναδικό αναγνωριστικό αριθμητικό στοιχείο, έναν ορισμό, ή ονόματα που υποδεικνύουν τον τομέα στον οποίο ανήκει. Επίσης, ένας όρος gene ontology μπορεί να αφορά κάποια παραπομπή σε αντίστοιχες έννοιες σε άλλες βάσεις δεδομένων. Η βάση της Γονιδιακής Οντολογίας περιλαμβάνει όρους που αφορούν ευκαρυωτικά και προκαρυωτικά κύτταρα καθώς και μονοκύτταρους και πολυκύτταρους οργανισμούς.

Η Γονιδιακή Οντολογία δεν είναι στατική. Παρατηρούνται προσθήκες, διορθώσεις και τροποποιήσεις που προτείνονται από μέλη των κοινοτήτων έρευνας και σχολιασμού καθώς και από όσους εμπλέκονται άμεσα στο έργο της gene ontology. Για παράδειγμα, ένας σχολιαστής μπορεί να ζητήσει έναν συγκεκριμένο όρο ώστε να αντιπροσωπεύεται μια μεταβολική οδός ή ένα τμήμα της Οντολογίας μπορεί να αναθεωρηθεί με τη βοήθεια των Εμπειρογνομώνων της Κοινότητας. Προτεινόμενες αλλαγές και προσθήκες από τους συντάκτες της Οντολογίας εφαρμόζονται ανάλογα με την περίπτωση.

Τέλος η Γονιδιακή Οντολογία είναι ελεύθερα διαθέσιμη από την ιστοσελίδα GO (<http://www.geneontology.org/>), ή μπορεί να υπάρξει πρόσβαση σε απευθείας σύνδεση με το πρόγραμμα GOAmigo.

```
id: GO:0000016
name: lactase activity
namespace: molecular_function
def: "Catalysis of the reaction: lactose + H2O = D-glucose + D-galactose." [EC:3.2.1.108]
synonym: "lactase-phlorizin hydrolase activity" BROAD [EC:3.2.1.108]
synonym: "lactose galactohydrolase activity" EXACT [EC:3.2.1.108]
xref: EC:3.2.1.108
xref: MetaCyc:LACTASE-RXN
xref: Reactome:20536
is_a: GO:0004553 ! hydrolase activity, hydrolyzing O-glycosyl compounds
```

Εικόνα 3.5 Παράδειγμα GO term (http://en.wikipedia.org/wiki/Gene_Ontology)

3.2.1 Εργαλεία

Υπάρχει ένας μεγάλος αριθμός διαθέσιμων εργαλείων online τα οποία χρησιμοποιούν τα δεδομένα που παρέχει η Γονιδιακή Οντολογία.. Η κοινοπραξία gene ontology αναπτύσσει και υποστηρίζει δύο εργαλεία, την AmiGO και OBO-Edit. Η AmiGO είναι μία διαδικτυακή εφαρμογή που επιτρέπει στους χρήστες την αναζήτηση, περιήγηση και οπτικοποίηση δεδομένων γονιδιακών προϊόντων. Επιπλέον, περιέχει το πρόγραμμα BLAST και εργαλεία που επιτρέπουν την ανάλυση μεγάλων δεδομένων. Το AmiGO μπορεί να χρησιμοποιηθεί online στην ιστοσελίδα (<http://www.geneontology.org/>), όπου υπάρχει πρόσβαση στα δεδομένα που παρέχονται από τη Γονιδιακή Οντολογία. Επιπλέον, μπορεί να εγκατασταθεί για τοπική χρήση σε οποιαδήποτε βάση δεδομένων που χρησιμοποιεί τη βάση της Γονιδιακής Οντολογίας. Παρέχεται δωρεάν και είναι διαθέσιμο ως μέρος του go-dev διανομής λογισμικού.

Το OBO-Edit είναι ένα εργαλείο το οποίο περιλαμβάνει μια ολοκληρωμένη αναζήτηση και ένα φίλτρο διασύνδεσης, με τη δυνατότητα να καταστήσει υποσύνολα των όρων ώστε να είναι διακριτά οπτικά. Αν και αναπτύχθηκε για την οντολογία που αφορά τη βιοϊατρική, το OBO-Edit μπορεί να χρησιμοποιηθεί για την προβολή, αναζήτηση και επεξεργασία κάθε οντολογίας. Είναι ελεύθερα διαθέσιμο στο διαδίκτυο (<http://www.geneontology.org>).

3.2.2 Ιστορικά στοιχεία

Η Γονιδιακή Οντολογία κατασκευάστηκε αρχικά το 1998 από κοινοπραξία ερευνητών που μελετούσαν το γονιδίωμα τριών οργανισμών μοντέλων: *Drosophila melanogaster* (μύγα των φρούτων), *Mus musculus* (ποντίκι), και *Saccharomyces cerevisiae* (κοινή μαγιά). Πολλές άλλες βάσεις δεδομένων έχουν προσχωρήσει στην κοινοπραξία της Γονιδιακής Οντολογίας, συμβάλλοντας όχι μόνο στο σχολιασμό των δεδομένων, αλλά και στην ανάπτυξη των οντολογιών και των εργαλείων ώστε να εφαρμοστούν τα δεδομένα. Μέχρι τώρα, οι περισσότερες μεγάλες βάσεις δεδομένων που αφορούν φυτά, ζώα και μικροοργανισμούς έχουν συμβάλει προς την κατεύθυνση αυτού του έργου. Από τον Ιανουάριο του 2008, η Gene Ontology περιέχει πάνω από 24.500 όρους που εφαρμόζονται σε μια ευρεία ποικιλία βιολογικών οργανισμών. Υπάρχει ένας σημαντικός όγκος βιβλιογραφίας σχετικά με την ανάπτυξη και τη χρήση του gene ontology, και αποτελεί ένα πρότυπο εργαλείο στο οπλοστάσιο της βιοπληροφορικής (<http://www.geneontology.org>).

3.3 Βάση Δεδομένων Ensembl (Ensembl Genome Browser)

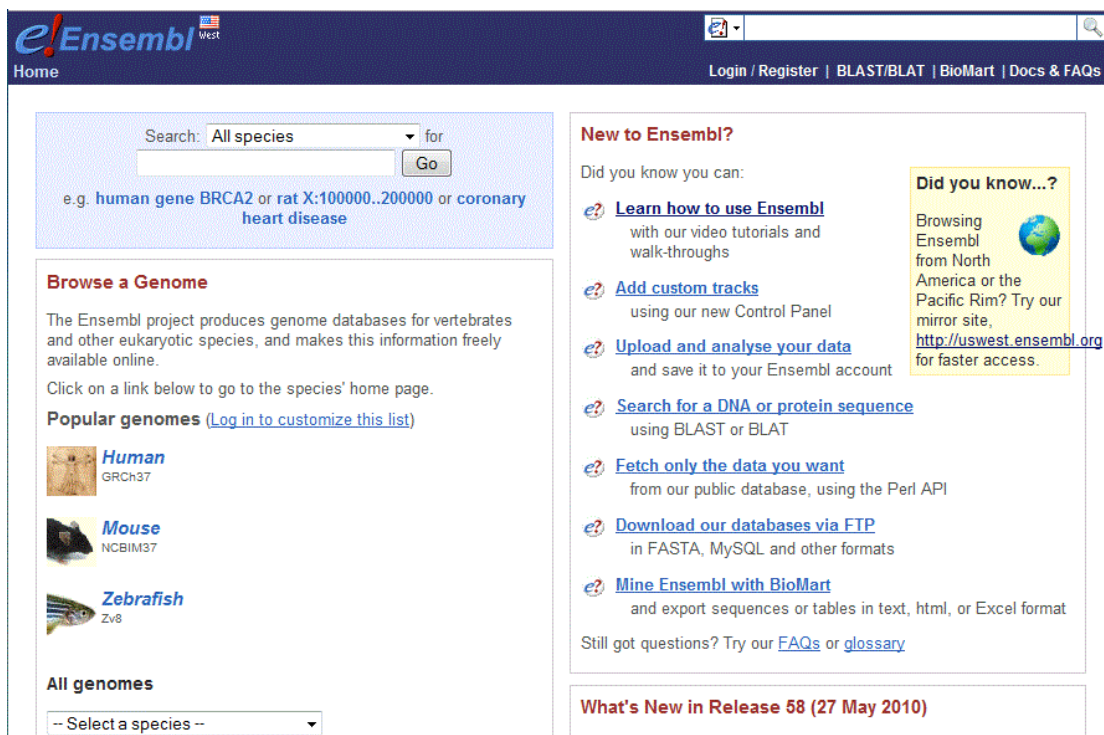
3.3.1 Ensembl Genome Browser (<http://www.ensembl.org>)

Ο αριθμός των βάσεων δεδομένων όσον αφορά το πεδίο της Μοριακής Βιολογίας έχει αυξηθεί ραγδαία τα τελευταία χρόνια. Υπάρχουν περισσότερες από 1000 διαφορετικές βάσεις δεδομένων όπου η κάθε μια παρέχει πληροφορίες σε διαφορετική μορφή. Πολύ συχνά η συλλογή πληροφοριών και η σύγκριση αυτών προερχόμενων από διαφορετικές πηγές προκαλεί σύγχυση σε πολλούς ερευνητές. Έτσι, λοιπόν οι φυλλομετρητές Γονιδιωμάτων (Genome Browsers) παρέχουν μεγάλη βοήθεια ως προς

την άντληση πληροφοριών και δεδομένων από διαφορετικές πηγές και παρουσίαση αυτών μ' ένα ολοκληρωμένο τρόπο.

Τρεις Genome Browsers χρησιμοποιούνται ευρέως στην επιστημονική κοινότητα: ο UCSC Genome Browser, ο NCBI Map Viewer και ο Ensembl Genome Browser. Οι βάσεις αυτές δεν παρέχουν μόνο πληροφορίες αλλά παράλληλα συνδέουν πληροφορίες από διάφορες πηγές και τις παρουσιάζουν ολοκληρωμένες.

Η Ensembl Genome Browser αποτελεί μια βάση δεδομένων που ξεκίνησε το 1999, λίγα χρόνια πριν ολοκληρωθεί η αποκρυπτογράφηση ολόκληρου του ανθρώπινου γονιδιώματος. Σκοπός του project ήταν η μελέτη ολόκληρου του ανθρώπινου γονιδιώματος (3 δισεκατομμυρίων ζευγών βάσεων) και η δημοσιοποίηση αυτής της μελέτης, ενώ από τον Ιούλιο του 2000 και μετά, πληθώρα γονιδιωμάτων διαφορετικών ειδών προστέθηκαν στη βάση αυτή. Η Ensembl Genome Browser περιλαμβάνει πάνω από 30 είδη, η πλειονότητα των οποίων είναι σπονδυλωτά, συμπεριλαμβανομένου του ανθρώπου, του ποντικού, του αρουραίου, του zebrafish καθώς και μια μεγάλη γκάμα οργανισμών που χρησιμοποιούνται ευρέως σε ερευνητικές μελέτες (πχ η *Drosophila Melanogaster* και ο *Caenorhabditis elegans*) (Εικ.3.6)



Εικόνα 3.6 Αρχική σελίδα του Ensembl Genome Browser (<http://www.ensembl.org/index.html>).

Γονιδιακές ακολουθίες μπορούν να εξαχθούν απευθείας από την Ensembl ή από ένα εργαλείο της που ονομάζεται Biomart. Με την βοήθεια της συγκεκριμένης βάσης δεδομένων μπορούμε να βρούμε και να μελετήσουμε γονίδια καθώς και να βρούμε πληροφορίες σχετικά με χρωμοσώματα και τα εναλλακτικά μετάγραφα ενός γονιδίου (alternative transcripts), είμαστε σε θέση να διερευνήσουμε ομολογίες και φυλογενετικά δένδρα σε περισσότερα από 50 είδη οργανισμών για κάθε γονίδιο, παράλληλα μπορούμε να συγκρίνουμε ολόκληρες γενωμικές ακολουθίες (genome alignments) και συντηρημένες περιοχές σε διάφορα είδη, όπως και να μελετήσουμε ακολουθίες που χρησιμοποιούνται στις μικροσυστοιχίες (microarray sequences), ανιχνευτές microarrays (microarray probes), mRNA και πρωτεΐνες για κάθε χρωμοσωμική περιοχή. Θα ήταν παράλειψη να μην αναφερθεί η δυνατότητα διερεύνησης πολυμορφισμών και πιο συγκεκριμένα των SNPs (Single Nucleotide Polymorphisms) ενός γονιδίου ή μιας χρωμοσωμικής περιοχής, ενώ μπορούμε να αναζητήσουμε ομόλογες ακολουθίες διαφόρων οργανισμών με τη βοήθεια του προγράμματος BLAST καθώς και εξαγωγή ακολουθιών και πληροφοριών σε διάφορες μορφές. Παρακάτω δίνονται κάποια παραδείγματα (Εικ.3.7-11)

Στην παρακάτω εικόνα δίνεται μια άποψη μετά από αναζήτηση ενός γονιδίου:

Gene: ACE (ENSG00000159640)
 angiotensin I converting enzyme (peptidyl-dipeptidase A) 1 [Source:HGNC Symbol;Acc:2707]
 Location: Chromosome 17: 61,554,432-61,599,209 forward strand.
 Transcripts: There are 6 transcripts in this gene

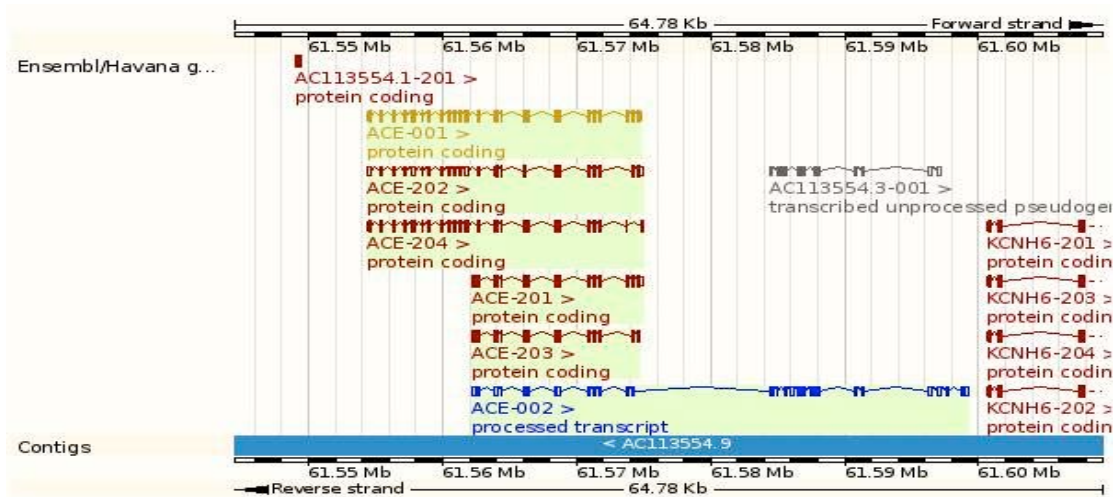
Name	Transcript ID	Length (bp)	Protein ID	Length (aa)	Biotype	CCDS
ACE-001	ENST00000290866	4022	ENSP00000290866	1306	Protein coding	CCDS11637
ACE-201	ENST00000290863	4022	ENSP00000290863	732	Protein coding	CCDS45755
ACE-202	ENST00000413513	3880	ENSP00000392247	511	Protein coding	-
ACE-203	ENST00000421982	2121	ENSP00000387760	691	Protein coding	-
ACE-204	ENST00000428043	3460	ENSP00000397593	1145	Protein coding	-
ACE-002	ENST00000490216	4862	No protein product	-	Processed transcript	-

Gene summary [help](#) [Splice variants >](#)

Name: [ACE](#) (HGNC Symbol)
 Synonyms: ACE1, CD143, DCP1 [To view all Ensembl genes linked to the name [click here](#)]
 CCDS: This gene is a member of the Human CCDS set: [CCDS11637](#), [CCDS45755](#)
 Gene type: Known protein coding
 Prediction Method: Gene containing both Ensembl genebuild transcripts and [Havana](#) manual curation, see [article](#).
 Alternative genes: This gene corresponds to the following database identifiers:
 Havana gene: [OTTHUMG00000154927](#) [view all locations]

Εικόνα 3.7 Αναζήτηση του γονιδίου ACE και πληροφορίες που μπορούμε να αντλήσουμε από την βάση Ensembl Genome Browser. Οι πληροφορίες αφορούν δεδομένα για το γονίδιο, όπως γενετικοί πολυμορφισμοί, ακολουθίες των μεταγράφων του καθώς γίνεται και αναφορά στο μέγεθος αυτών, γενωμικές στοιχίσεις, γονιδιακά δένδρα, ορθόλογα και παράλογα γονίδια σε

άλλους οργανισμούς, πρωτεϊνικές οικογένειες, καθώς και δυνατότητα εξαγωγής στοιχείων (<http://www.ensembl.org>).



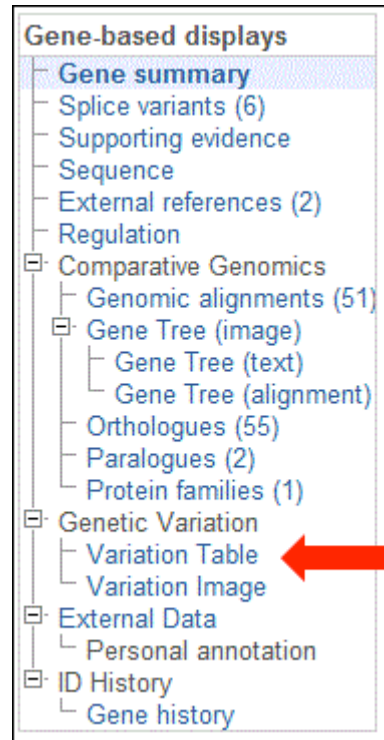
Εικόνα 3.8 Αναφορά των contig και των μεταγράφων του γονιδίου ACE (<http://www.ensembl.org>).

Location: 17:61,554,432-61,599,209 Gene: ACE Transcript: ACE-001

- Transcript-based displays**
- Transcript summary
 - Supporting evidence (8)
 - Sequence
 - Exons (25)
 - cDNA
 - Protein
 - External References
 - General identifiers (58)
 - Oligo probes (38)
 - Ontology
 - Ontology chart (37)
 - Ontology table (37)
 - Genetic Variation
 - Population comparison
 - Comparison image
 - Protein Information
 - Protein summary
 - Domains & features (17)
 - Variations (61)
 - External Data
 - Personal annotation
 - ID History
 - Transcript history
 - Protein history

Εικόνα 3.9 Πληροφορίες που μπορούμε να αντλήσουμε από το μετάγραφο ACE (<http://www.ensembl.org>)

Οι παρακάτω εικόνες αφορούν πληροφορίες που μπορούν να αντληθούν από γενωμικές περιοχές :

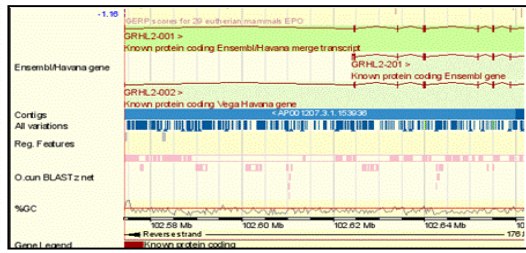


Εικόνα 3.10. Αφορά την περίληψη και τις πληροφορίες που μπορούμε να αντλήσουμε από ένα γονίδιο. Όπως το είδος του οργανισμού που θέλουμε να ανιχνεύσουμε ένα συγκεκριμένο γονίδιο, το όνομα του γονιδίου, την περιοχή του χρωμοσώματος που βρίσκεται το γονίδιο, το Id του μεταγράφου που επιθυμούμε να αναλύσουμε καθώς και τους πολυμορφισμούς (<http://www.ensembl.org>).

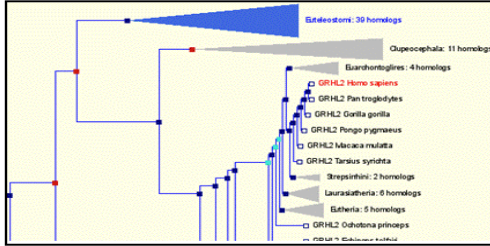
Transcript Sequence w/Variations

```
1 ATTGGATCAAACATGTCACAAAGATCGGACAAATAAAGACTAGTGGCTTAATGCCCC
.....ATGTCACAAAGATCGGACAAATAAATAAAGACTAGTGGCTTAATGCCCC
.....-M--S--Q--E--S--D--N--N--K--R--L--V--A--L--V--P--
61 ATGCCCAAGTACCCCTCCATTCAATACCCGAAAGAGCTACACCACTGAGGATGAAGCCTGG
49 ATGCCCAAGTACCCCTCCATTCAATACCCGAAAGAGCTACACCACTGAGGATGAAGCCTGG
17 -M--P--S--D--P--P--F--N--T--R--R--A--Y--T--S--E--D--E--A--M--
121 AAGTCATACCTTGGAGAATCCCGCTGACAGCAGCCACCAAGGCCATGATGAGCATTAAATGGT
109 AAGTCATACCTTGGAGAATCCCGCTGACAGCAGCCACCAAGGCCATGATGAGCATTAAATGGT
37 -K--S--Y--L--E--N--P--L--T--A--A--T--K--A--M--M--S--I--N--G--
181 GATGAGGACAGTGTCTGCTGCCCTCGGCCTGCTCTATGACTACTACAAAGTTCCTCGAGAC
169 GATGAGGACAGTGTCTGCTGCCCTCGGCCTGCTCTATGACTACTACAAAGTTCCTCGAGAC
57 -D--E--D--S--A--A--A--L--G--L--L--Y--D--Y--Y--K--V--P--R--D--
```

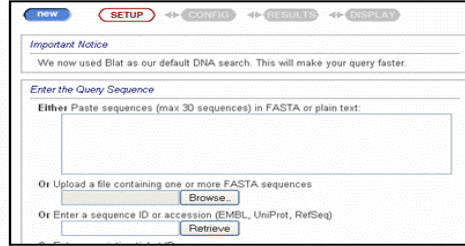
Genes, SNPs, and Conserved Regions



Homologues in Gene Trees



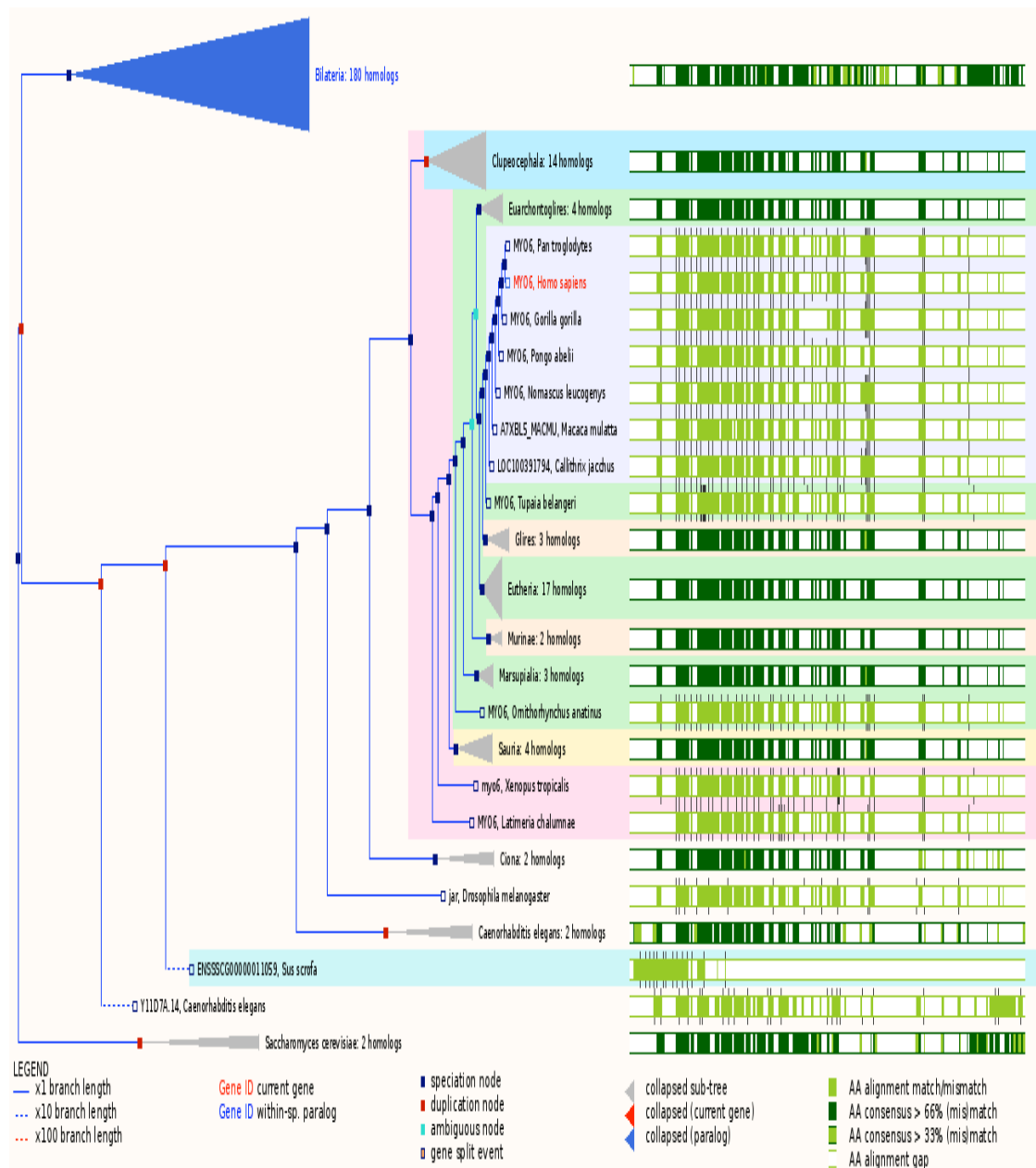
BLAST and BLAT aligners



Εικόνα 3.11 Πληροφορίες που αφορούν γονιδιακές αναλύσεις όπως ακολουθίες, πολυμορφισμοί, ομολογίες, φυλογενετικά δένδρα καθώς και εργαλεία αναζήτησης ομόλογων ακολουθιών π.χ. το Blast (<http://www.ensembl.org>).

3.3.2 Πρόβλεψη Ορθόλογων και Παράλογων γονιδίων στη βάση δεδομένων Ensembl

Η πρόβλεψη των ορθόλογων και παράλογων γονιδίων για όλα τα είδη στη βάση δεδομένων Ensembl πραγματοποιείται μέσω της κατασκευής φυλογενετικών δένδρων. Προκειμένου να κατασκευαστεί ένα φυλογενετικό δένδρο χρησιμοποιείται το μεγαλύτερο μετάγραφο (transcript) ενός Ensembl γονιδίου και μία άποψη αυτού φαίνεται την παρακάτω εικόνα:



Εικόνα 3.12 Παράδειγμα γονιδιακού δένδρου της ανθρώπινης μυοσίνης VI (MYO6) [κόκκινο χρώμα], του ENSG00000196586. Το συγκεκριμένο δένδρο έχει κατασκευαστεί χρησιμοποιώντας τα αντίστοιχα μεγαλύτερα μετάγραφα κάθε ορθόλογου και παράλογου γονιδίου σε διάφορους οργανισμούς (<http://www.ensembl.org/info/website/tutorials/coursebook.pdf>).

Οι κόμβοι αναπαριστούν είτε διπλασιασμούς (κόκκινο χρώμα), είτε γεγονότα ειδογένεσης (μπλε χρώμα) και μ' αυτόν τον τρόπο μπορούν να ανιχνευθούν τόσο τα αρχαία όσο και τα πιο πρόσφατα παράλογα. Έτσι με τη δημιουργία των φυλογενετικών δένδρων μπορούμε να αντλήσουμε σημαντικές πληροφορίες που αφορούν την εξέλιξη και ειδογένεση .

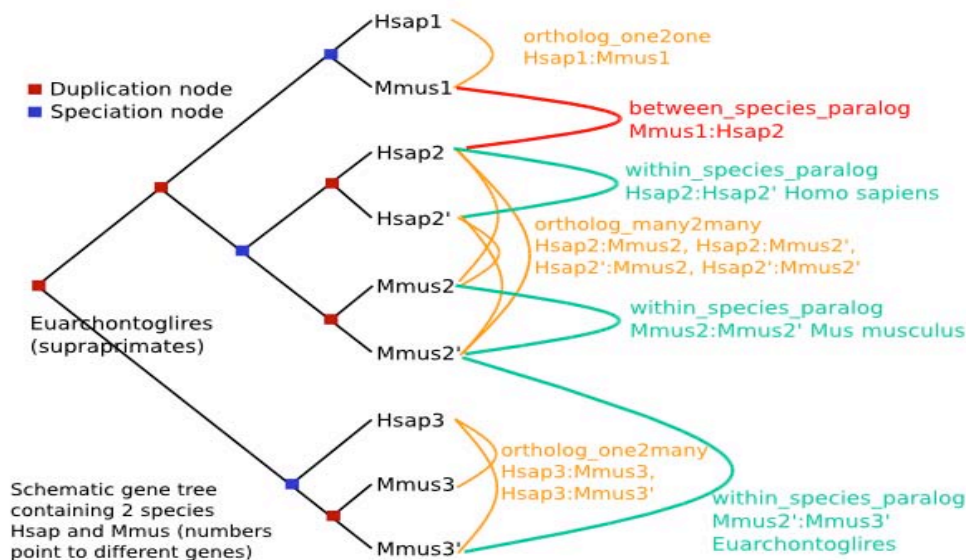
Χρησιμοποιώντας τα φυλογενετικά δένδρα μπορούν να εξαχθούν διάφορα συμπεράσματα που αφορούν τις ακόλουθες σχέσεις ζευγών ομόλογων γονιδίων (3.13-3.14) :

Ορθόλογες σχέσεις: Αφορούν ζεύγη ομόλογων γονιδίων όπου ο κοινός τους πρόγονος προήλθε από ειδογένεση. Έχουν προβλεφθεί αρκετές περιγραφές των ορθόλογων αυτών σχέσεων:

- i. Ορθόλογα ένα προς ένα (one2one)
- ii. Ορθόλογα ένα προς πολλά (one2many)
- iii. Ορθόλογα πολλά προς πολλά (many2many)
- iv. Προφανώς ορθόλογα ένα προς ένα (one2one)
- v. Πιθανότατα ορθόλογα

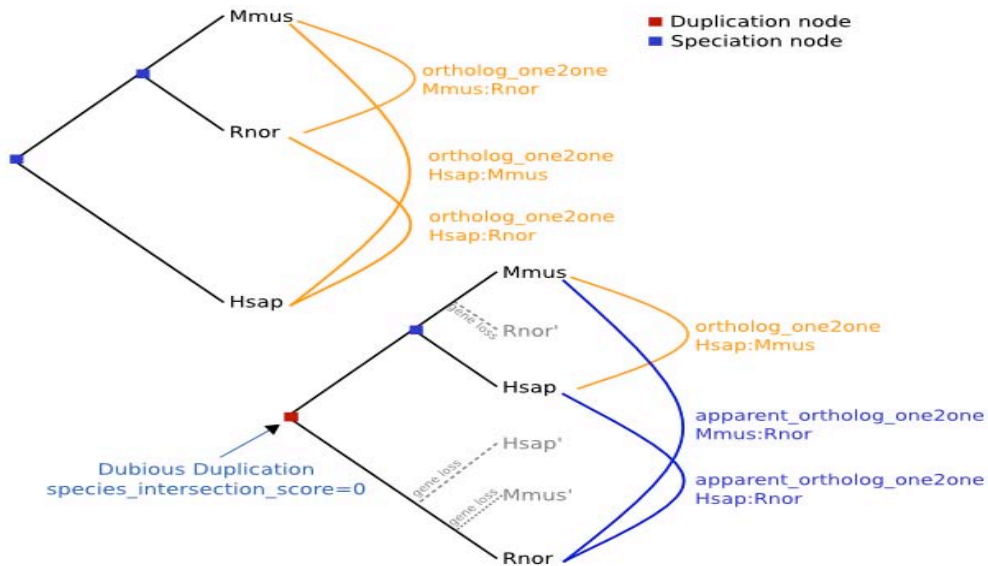
Παράλογες σχέσεις: Αφορούν ζεύγη ομόλογων γονιδίων που προήλθαν από γονιδιακό διπλασιασμό.

- i. Παράλογα εντός ενός είδους
- ii. Παράλογα μεταξύ των ειδών,
- iii. Άλλα παράλογα.
- iv. Παράλογα που προήλθαν από υποτιθέμενη διάσπαση του γονιδίου (http://www.ensembl.org/info/docs/compara/homology_method.html).



Εικόνα 3.13 Φυλογενετικό δένδρο που περιλαμβάνει δύο είδη, τον άνθρωπο και το ποντίκι. Φαίνονται οι σχέσεις των ορθόλογων γονιδίων τους

(http://www.ensembl.org/info/docs/compara/homology_method.html).



Εικόνα 3.14 Γονιδιακά δένδρα τριών διαφορετικών ειδών όπου διακρίνονται οι διαφορετικές σχέσεις των ορθόλογων γονιδίων ανάλογα με τη προέλευση του προγόνου. Στην πρώτη περίπτωση ο πρόγονος προήλθε από ειδογένεση (επάνω), ενώ στη δεύτερη περίπτωση ο πρόγονος προήλθε από διπλασιασμό (http://www.ensembl.org/info/docs/compara/homology_method.html).

3.4 Biomart

Η *Biomart* είναι ένα εργαλείο εξόρυξης δεδομένων το οποίο επιτρέπει την άμεση πρόσβαση σε γονίδια, ακολουθίες και γενικότερα μιας μεγάλης γκάμας πληροφοριών και υποστηρίζεται από την Ensembl Genome Browser, δίχως να απαιτείται η γνώση προγραμματισμού, και συμβάλλει στην ενοποίηση και σύγκριση των δεδομένων μας με άλλες βάσεις δεδομένων.

Με τη βοήθεια της *Biomart* μπορούν να εξαχθούν ακολουθίες σε μορφή FASTA, πίνακες σε μορφή κειμένου, html, είτε σε μορφή excel. Το συγκεκριμένο εργαλείο επιτρέπει στους χρήστες τη χρήση πολλών φίλτρων στα ήδη υπάρχοντα δεδομένα με σκοπό την άντληση πληροφοριών που αναζητούν οι ερευνητές της επιστημονικής κοινότητας. Με τη *Biomart* μπορούμε να μελετήσουμε και να απομονώσουμε(με τη βοήθεια φίλτρων) συγκεκριμένες περιοχές γονιδίων και χρωμοσωμάτων δίνοντας πληροφορίες που αφορούν συγκεκριμένες ζώνες έναρξης και λήξης (band start και band end), τον τύπο και περιοχή κωδικοποίησης και πολλές άλλες

χρήσιμες πληροφορίες. Επιπλέον, μπορούν να δωθούν Ids γονιδίων και μεταγράφων και να αντληθούν τα αντίστοιχα ορθόλογα σε διάφορα είδη και οργανισμούς. Θα ήταν παράλειψη να μην αναφερθεί η χρήση της Γονιδιακής Οντολογίας προκειμένου να αντληθούν και να ανευρεθούν γονίδια που ενδιαφέρουν τον εκάστοτε ερευνητή. Τέλος δίνεται η δυνατότητα αναζήτησης γονιδιακών χαρακτηριστικών και πρωτεωμικών πληροφοριών σε άλλες βάσεις δεδομένων χρησιμοποιώντας διάφορα ονόματα και Gene Ids (ανάλογα με τη βάση δεδομένων που δουλεύουμε) και ανεύρεση και αναζήτηση των συγκεκριμένων γονιδίων σε άλλες βάσεις και πηγές με διαφορετικά κάθε φορά ταυτότητα (Εικ.3.15-3.16)

A

Dataset: Mus musculus genes (NCBI37)

Filters: [None selected]

Attributes: Ensembl Gene ID, Ensembl Transcript ID

Dataset: [None Selected]

ENSEMBL 50 GENES (SANGER UK)

Mus musculus genes (NCBI37)

- CHODGE DATASET -
 - Drosophila melanogaster genes (DMM10.2)
 - Gallus gallus genes (WASR-LUC2)
 - Homo sapiens genes (NCBI36)
 - Mus musculus genes (NCBI37)
 - Rattus norvegicus genes (RGSC3.4)
 - Aedes aegypti genes (Aaeag1)
 - Anopheles gambiae genes (AgamP3)
 - Bos taurus genes (Btau_3.1)
 - Caenorhabditis elegans genes (WS1180)
 - Canis familiaris genes (BROAD02)
 - Civet parvulus genes (GJNEAP01)
 - Ciona intestinalis genes (LIC2)
 - Ciona savignyi genes (CSAV2.0)
 - Dasyatis novemmaculatus genes (ANIMA)
 - Drosophila melanogaster genes (DMM10.2)
 - Echinosps tetarti genes (TETREC)
 - Equus caballus genes (EqCab2)
 - Erhmannia europaeus genes (HEDEGH00)

B

Dataset 41 / 28329 Genes

Mus musculus genes (NCBI37)

Filters: Chromosome: 1, Gene Start (bp): 1, Gene End (bp): 10000000

Attributes: Ensembl Gene ID, Ensembl Transcript ID

Dataset: [None Selected]

Please restrict your query using criteria below

REGION:

Chromosome: 1

Base pair: 1 - 10000000

Band: 1p - 1p

C

Dataset 41 / 28329 Genes

Mus musculus genes (NCBI37)

Filters: Chromosome: 1, Gene Start (bp): 1, Gene End (bp): 10000000

Attributes: Ensembl Gene ID, Associated Gene Name, Chromosome Name, Gene Start (bp), Gene End (bp)

Dataset: [None Selected]

Please select columns to be included in the output and hit 'Results' when ready

Features: Structures Sequences

GENE:

Ensembl Attributes:

Ensembl Gene ID
 Ensembl Transcript ID
 Ensembl Protein ID
 Description
 Chromosome Name
 Gene Start (bp)
 Gene End (bp)
 Strand
 Band

Associated Gene Name
 Associated Transcript Name
 Associated Gene DB
 Associated Transcript DB
 Transcript count
 % GC content
 Biotype
 Source
 Status (gene)

D

Dataset 41 / 28329 Genes

Mus musculus genes (NCBI37)

Filters: Chromosome: 1, Gene Start (bp): 1, Gene End (bp): 10000000

Attributes: Ensembl Gene ID, Associated Gene Name, Chromosome Name, Gene Start (bp), Gene End (bp)

Dataset: [None Selected]

Export all results to: File

Email notification to: []

View: 10 rows as HTML

TSV
 HTML
 Mouse 36 assembly (HTML)
 CSV
 Mouse 36 assembly (CSV)
 TSV
 Mouse 36 assembly (TSV)
 XLS

Unique results only Go

Ensembl Gene ID	Associated Gene Name	Chromosome Name	Gene Start (bp)	Gene End (bp)
ENSMUGEN00000000001	Xist	1	32,210,000	32,212,000
ENSMUGEN00000000002	Rpl19	1	4,124,000	4,124,400
ENSMUGEN00000000003	Sirt1	1	11,252,000	11,252,400
ENSMUGEN00000000004	Mgfr1	1	11,252,000	11,252,400
ENSMUGEN00000000005	Lys1	1	11,252,000	11,252,400
ENSMUGEN00000000006	Lys1	1	11,252,000	11,252,400
ENSMUGEN00000000007	Rpl33	1	11,252,000	11,252,400
ENSMUGEN00000000008	Agpat1	1	11,252,000	11,252,400
ENSMUGEN00000000009	Tp53	1	11,252,000	11,252,400
ENSMUGEN00000000010	Rpl22t	1	11,252,000	11,252,400

E

Dataset 41 / 28329 Genes

Mus musculus genes (NCBI37)

Filters: Chromosome: 1, Gene Start (bp): 1, Gene End (bp): 10000000

Attributes: Ensembl Gene ID, Associated Gene Name, Chromosome Name, Gene Start (bp), Gene End (bp)

Dataset: Gene Targeting

Filters: Status: ES cells - Targeting confirmed

Attributes: Gene Symbol, EUCOMM, KOMP, NonCOMM, Status

Export all results to: File

Email notification to: []

View: 10 rows as HTML

TSV Unique results only Go

Ensembl Gene ID	Associated Gene Name	Chromosome Name	Gene Start (bp)	Gene End (bp)	Gene Symbol	EUCOMM	KOMP	NonCOMM	Status
ENSMUGEN00000000010	Tp53	1	64,721,200	64,721,200	Tp53	1			ES cells - Targeting confirmed

Εικόνα 3.15 Παράδειγμα αναζήτησης δεδομένων με τη χρήση της Biomart εξόρυξης αποτελεσμάτων (http://www.informatics.jax.org/userhelp/biomart_help.shtml#whatis).

A

Dataset 696 / 36582 Genes
Homo sapiens genes (NCBI36)

Filters
Affy hg u95av2 ID(s): [ID-list specified]

Attributes
Flank (Gene)
Upstream flank [1000]
Ensembl Gene ID
Chromosome Name
Gene Start (bp)
Gene End (bp)

Dataset
[None Selected]

Please select columns to be included in the output and hit 'Results' when ready

Features Homologs
Structures Sequences
Variations

SEQUENCES:
Sequences (max 1)

Unspliced (Transcript) 5' UTR
Unspliced (Gene) 3' UTR
Flank (Transcript) Exon sequences
Flank (Gene) cDNA sequences
Flank-coding region (Transcript) Coding sequence
Flank-coding region (Gene) Protein

Upstream flank
 Upstream flank [1000]

Downstream flank
 Downstream flank

Header Information
Gene Attributes
 Ensembl Gene ID Chromosome Name
 Description Gene Start (bp)
 Associated Gene Name Gene End (bp)
 Associated Gene DB

B

Dataset 696 / 36582 Genes
Homo sapiens genes (NCBI36)

Filters
Affy hg u95av2 ID(s): [ID-list specified]

Attributes
Flank (Gene)
Upstream flank [1000]
Ensembl Gene ID
Chromosome Name
Gene Start (bp)
Gene End (bp)

Dataset
[None Selected]

Export all results to File FASTA Unique results only

Email notification to

View 10 rows as FASTA Unique results only

```
>ENSG000000004487 | 1 | 23218533 | 23282771
TCACCCCTAATCAAGATATAAAACATTTCCATCACCACCAGCAAGTCTTTCTTGTCTTTT
TCAGTCAACCCCGACCCCATCTCAGGCAAGGCTGATTGGACCTATTACTGGAGAT
TAGTTTACCAGTTCTGAACTTCATGTAATGGAATAATTTTTGACAAATTCACCTCG
AAGAGACAGACTATTAAGAAGACAATGTTTGTGTAATAAGTACACGCATGCTAAAAGTA
CACGACCTGTTCCAGTTTTGGGTTAGCAAGGGATTCTCGAAAAATTAAGAATAAAC
GCTACTATCATTATTAGAAAGAATTAACCTGGAGTAGAAAGAGGTGTTATGTGGATGGG
ATGGGGGACATGAAGGATGGGGGACATGGGGGATGGGGGACATGGGGGATGGGAGAAGCG
CAAGACATGGAGTGCAGATGAGGTTGGGTCACGCCAGAAGGAAAGCATGAACAAGGC
TTGGAGACAAGAGCAATTCCTGCTGGACTCAAACCTGGATTTCTGTAACCTGTCTCCATT
TACTACCATCGAACTGCTGTCTCCAAAGTTTTGCCCCAGTCCAAGCAGGGCAATTA
TGAATAAGTAAGACGCTCTACCAACATATTCAGTGTGTTTCCCCATTAGACTGTAAGTT
CTTTGAAAGCTAGAATTCATAATGTAACCCCTGTTACGTAGTAGGCAAGCATTAATAA
TACTGTTGAAATTAACGAATAATGGGTGAGTGAATGAACGAACGCACGCATGCAAAACC
GAAAGTCCCTGGAGGAAATGGTCACCTTCGGAGGTTTGTCTGGCCGAGAAGCCCTAAGA
CCACGGACTGTGCCAGTCCCACTCCAAACCGCCGGGAGACGCTCTAGGCAAGCTACAGG
TTCTTTGCTGCGGTGCCACTCTAGCCGCGAGAACCGCGCTCTATGGCTCGGGGGAGGGG
CGGGCTCGTGGGTGCTCCGACCCCTTTTGTCCCGGCGC
>ENSG00000000938 | 1 | 27811390 | 27834314
TGTTAGCCAGGACGGTCTGATCTCTCTAACATCGTGATCCACCCGCTCGGGCTCCCA
AAGTCTGGGATTAACAGGCTGAGCCCTGCACCCGGCCCTTTCTTTCTTTTAAAAA
```

Εικόνα 3.16 Εξαγωγή ακολουθιών *Ensembl Gene Transcript* σε μορφή *FASTA* (http://www.informatics.jax.org/userhelp/biomart_help.shtml#whatis).

3.5 BioGRID (*Biological General Repository for Interaction Datasets*)

Η BioGRID (<http://thebiogrid.org/>) είναι μια επιμελημένη βιολογική βάση δεδομένων η οποία περιλαμβάνει πρωτεϊνικές και γενετικές αλληλεπιδράσεις και δημιουργήθηκε το 2003 από τους Mike Tyers, Bobby-Joe Breitkreutz και Chris Stark στο Ινστιτούτο Ερευνών «Samuel Lunenfeld» το οποίο εδρεύει στο Mount Sinai

Hospital. Στην ουσία αποτελεί μια ολοκληρωμένη πηγή πρωτεϊνικών και γενετικών αλληλεπιδράσεων των σημαντικότερων οργανισμών μοντέλων και πιο συγκεκριμένα 50 ειδών, συμπεριλαμβανομένου του ανθρώπου, του χιμπατζή και άλλων οργανισμών που είναι πολύ διαδεδομένοι σε μελέτες όπως η *Drosophila melanogaster*, ο *Caenorhabditis elegans*, ενώ τείνει να απομακρύνει τους πλεονασμούς έτσι ώστε να δημιουργήσει μια ενιαία χαρτογράφηση των αλληλεπιδράσεων. Πλήρης κάλυψη δεδομένων που αφορούν πρωτεϊνικές γενετικές αλληλεπιδράσεις έχει αναφερθεί σε τρεις οργανισμούς και πιο συγκεκριμένα στους: *Saccharomyces cerevisiae*, *Arabidopsis thaliana* και *Schizosaccharomyces pombe* ενώ γίνονται προσπάθειες εμπλουτισμού πληροφοριών και σε άλλα μετάζωα. Αντίθετα η πλήρης κάλυψη πληροφοριών που αφορούν τα θηλαστικά, συμπεριλαμβανομένου και του ανθρώπου απέχει αρκετά. Το BioGRID φιλοξενεί περισσότερα από 17 εκατ. ονόματα και ψευδώνυμα γονιδίων καθώς επίσης και αναγνωριστικά Ids τα οποία βοηθούν στην αναζήτηση και εξόρυξη δεδομένων (Stark *et al.*, 2011).

The image shows the BioGRID 3.1 website homepage. At the top, there is a navigation bar with links: home, help, wiki, tools, contribute, statistics, downloads, partners, about us. The main heading is "Welcome to the Biological General Repository for Interaction Datasets". Below this, there is a search box with the text "Search the BioGRID" and a search bar containing "Homo sapiens". There are buttons for "SUBMIT GENE SEARCH", "Advanced Search", "Search Tips", and "Featured Datasets". The page is divided into several sections: "AREAS OF INTEREST TO HELP YOU GET STARTED" with links to "Build and Download Interaction Datasets", "Link To Us or Submit Interactions", "Online Tools and Resources", and "View Our Interaction Statistics"; "BIOGRID FUNDING AND PARTNERS" with logos for various institutions like National Institutes of Health, CIHR, IRSC, BBSRC, Mount Sinai Hospital, Princeton University, University of Edinburgh, SGD, LAMHDI, and IMEx; and "LATEST NEWS" with a post about "BioGRID Version 3.1.92 Release (10,341 New Physical and Genetic Interactions Added)". There is also a section for "Get BioGRID Updates on Twitter".

Εικόνα 3.17 Αρχική σελίδα της BioGRID (<http://thebiogrid.org/>)

Η BioGRID είναι προσβάσιμη μέσω διαδικτύου (<http://thebiogrid.org/>) ενώ διατίθεται σε πολλές εκδόσεις. Η παρούσα έκδοση είναι η 3.1.93 η οποία αναζητά σε 39.996 δημοσιεύσεις, 557.934 πρωτεϊνικές και γενετικές αλληλεπιδράσεις των σημαντικότερων οργανισμών μοντέλων. Όλες οι πληροφορίες που παρέχονται είναι ελεύθερα προσβάσιμες και διατίθενται μέσω λήψεων (downloads) σε πολλές μορφές. Το BioGRID παρέχει πολλά εργαλεία και πρόσθετα (plugins) τα οποία συμβάλλουν στην απλοποίηση εισαγωγής δεδομένων από συγκεκριμένου τύπου αρχεία (bioGRID .tabfiles) και εισαγωγή αυτών στο Cytoscape με σκοπό την οπτικοποίηση των δεδομένων που παρέχονται από τη συγκεκριμένη βάση.

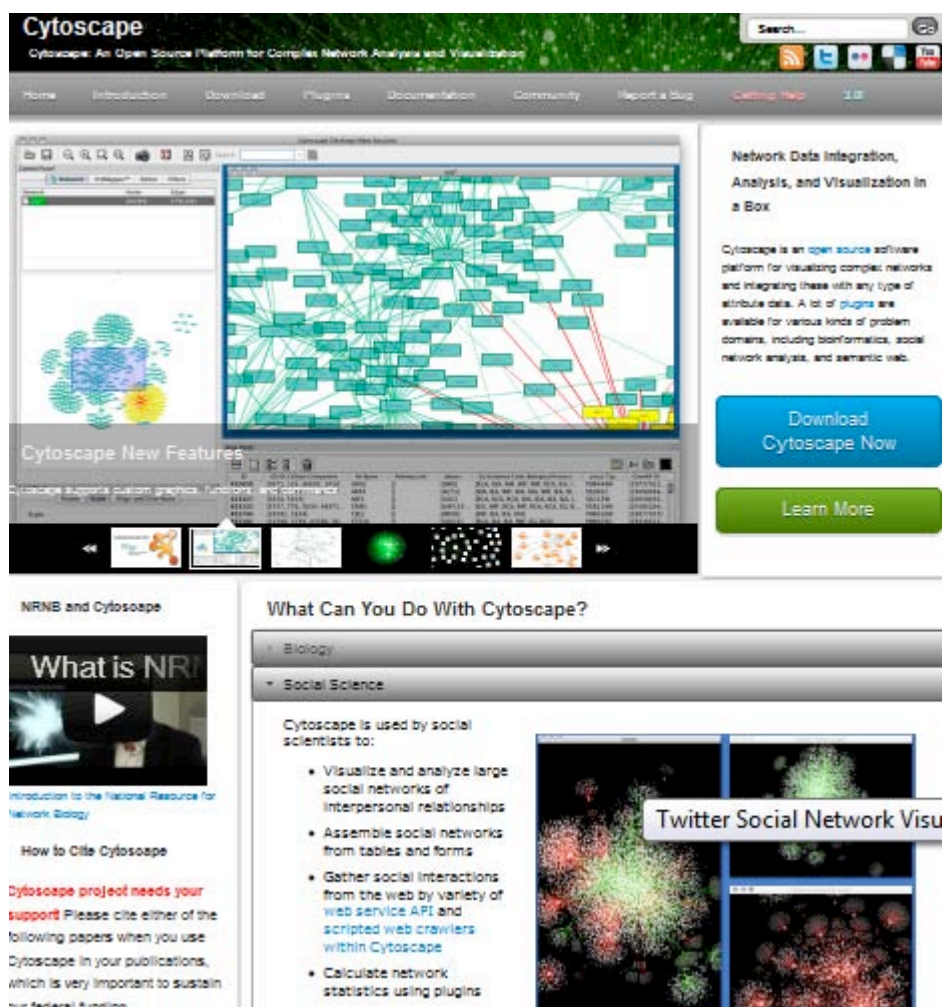
The screenshot displays the BioGRID 3.1 interface for the gene TRIM69 in Mus musculus. The search criteria are set to 'ALP' for the gene/identifier and 'Homo sapiens' for the organism. The page shows the following details:

- Gene/Identifier:** TRIM69 (Mus musculus)
- Accession:** 4921519C19Rik, Trif, Trimless, Rnf36, RP23-277B16.1
- Function:** RING finger B-box coiled-coil transcription factor
- GO Terms:** 0 Terms for Process, Function, and Component.
- External Database Linkouts:** VEGA, Ensembl, MGI, Entrez Gene, RefSeq, GenBank, UniprotKB.
- Download:** 36 Associations For This Protein.
- Stats & Filters:**
 - Current Statistics:** Publication: 1, High Throughput: 36 (100%), Low Throughput: 0 (0%), 36 Physical Interactions, 0 Genetic Interactions.
 - Search Filters:** No Filter: Show All Associations.
- Switch View:** Summary (selected), Sortable Table.
- Interactors:** Displaying 36 total unique interactors. Two examples are shown:
 - ALDH18A1:** PYCS, GSAS, P5CS, pyrroline-5-carboxylate synthetase (glutamate gamma-semialdehyde synthetase), Organism: Homo sapiens.
 - GPRC5A:** RAI3, GPCR5A, RAIG1, G protein-coupled receptor, family C, group 5, member A, Organism: Homo sapiens.

Εικόνα 3.18 Δεδομένα που μπορούμε να ανακτήσουμε από την αναζήτηση του γονιδίου ALP. Πιο συγκεκριμένα τα δεδομένα αφορούν περιγραφές γονιδίων και γονιδιακών λειτουργιών, σχολιασμούς από την Gene Ontology, στατιστικά αποτελέσματα φυσικών και γενετικών αλληλεπιδράσεων, χρήση φίλτρων που επιτρέπουν στο χρήστη την προσαρμογή και την απεικόνιση των αποτελεσμάτων καθώς και πληροφορίες που αφορούν την κάθε αλληλεπίδραση (<http://thebiogrid.org/>).

3.6 Cytoscape

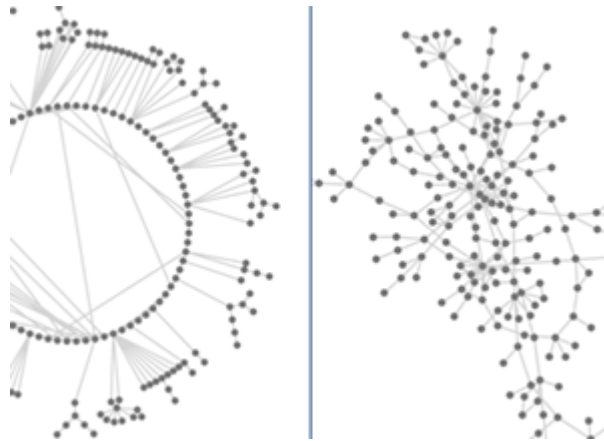
Το Cytoscape (www.cytoscape.org) είναι ένα πρόγραμμα βιοπληροφορικής το οποίο διατίθεται μέσω διαδικτύου και χρησιμοποιείται στην οπτικοποίηση των μοριακών αλληλεπιδράσεων, δικτύων και μοριακών μονοπατιών ενώ παράλληλα ενσωματώνει σχολιασμούς και πολλές άλλες πληροφορίες στα δίκτυα που δημιουργούνται (Εικ.3.19).



Εικόνα 3.19 Αρχική σελίδα του Cytoscape (www.cytoscape.org).

Το Cytoscap χρησιμεύει στην σύνθετη ανάλυση και οπτικοποίηση των δικτύων. Επιπλέον χαρακτηριστικά/λειτουργίες είναι διαθέσιμα ως πρόσθετα (plugins). Τα plugins είναι διαθέσιμα για μοριακές και δικτυακές αναλύσεις, νέες διατάξεις, επιπλέον υποστήριξη διαφόρων μορφών αρχείων, καθώς και συνδέσεις με άλλες βάσεις δεδομένων. Τα plugins μπορεί να δημιουργηθούν από οποιονδήποτε χρήστη του συγκεκριμένου προγράμματος, δεδομένου ότι το λογισμικό του είναι γραμμένο σε Java

ενώ είναι ελεύθερα διαθέσιμα στο διαδίκτυο. Το Cytoscape μπορεί να συνδεθεί απευθείας με εξωτερικές βάσεις δεδομένων και να εισάγει πολλές πληροφορίες και σχολιασμούς που αφορούν δεδομένα δικτύου. Παράλληλα δίνεται η δυνατότητα στους χρήστες να προσαρμόζουν τα δεδομένα τους χρησιμοποιώντας ισχυρά Visual Styles, δίνεται η δυνατότητα δικτύων σε δύο διαστάσεις (Εικ.3.20) ενώ υπάρχει και μια μεγάλη ποικιλία αλγορίθμων διάταξης πχ, κυκλική, διάταξη δένδρου κοκ.



Εικόνα 3.20 Διάταξη δεδομένων σε δύο διαστάσεις με τη χρήση του προγράμματος Cytoscape (www.cytoscape.org)

Επιπλέον, μπορεί να εφαρμοστεί zoom in και out στα δίκτυα, ενώ παρατηρείται εύκολη πλοήγηση και σε μεγάλα δίκτυα (>100.000 κόμβους) με τη βοήθεια ενός πολύ αποτελεσματικού προγράμματος απόδοσης. Τέλος παρέχει την δυνατότητα φιλτραρίσματος του δικτύου προκειμένου να επιλεγούν υποσύνολα των κόμβων ή/και των αλληλεπιδράσεων με βάση τα πρόσφατα δεδομένα μας.

3.7 OMIM (Online Mendelian Inheritance in Man)

Η OMIM (<http://www.ncbi.nlm.nih.gov/omim>) είναι μια ολοκληρωμένη, και έγκυρη βάση δεδομένων (Εικ. 3.21) που καταγράφει όλες τις γνωστές Μεντελικές διαταραχές και νοσήματα στον άνθρωπο και τις γενετικές συνιστώσες αυτών. Περιλαμβάνει επίσης περισσότερα από 12.000 γονίδια που εμπλέκονται σ' αυτές ενώ επιπλέον επικεντρώνεται στις σχέσεις μεταξύ φαινοτύπου και γονότυπου. Η συγκεκριμένη Βάση ενημερώνεται καθημερινά και οι καταχωρήσεις περιέχουν άφθονες συνδέσεις με άλλες πηγές γενετικής.

Αφού αναζητήσουμε μία ασθένεια η OMIM μας δίνει πολλές πληροφορίες όπως τα κλινικά γνωρίσματα της ασθένειας, τη χαρτογράφηση του γονιδίου, την κληρονομικότητα της ασθένειας, την επιδημιολογία, την παθογένεια κ.α. (Εικ.3.21-3.22).

The screenshot shows the OMIM search interface. The search term 'OMIM' is entered, and the results are filtered for 'PKU'. The results list 14 items, with the first seven visible:

- 1: #261600. PHENYLKETONURIA; PKU
HYPERPHENYLALANINEMIA, NON-PKU MILD, INCLUDED
Gene map locus [12q24.1](#)
- 2: *612349. PHENYLALANINE HYDROXYLASE; PAH
Gene map locus [12q24.1](#)
- 3: *608439. TOUSLED-LIKE KINASE 2; TLK2
Links
- 4: *608438. TOUSLED-LIKE KINASE 1; TLK1
Links
- 5: #261630. HYPERPHENYLALANINEMIA, BH4-DEFICIENT, C; HPABH4C
Gene map locus [4p13.31](#)
- 6: #261640. HYPERPHENYLALANINEMIA, BH4-DEFICIENT, A; HPABH4A
HYPERPHENYLALANINEMIA, BH4-DEFICIENT, DUE TO PARTIAL PTS DEFICIENCY, INCLUDED
Gene map locus [11q22.3-q23.3](#)
- 7: #222261. UNRESOLVED AT ANINEMIA, BH4 DEFICIENT, D; HPABH4D
GeneTests, Links

Εικόνα 3.21 Αναζήτηση δεδομένων από την OMIM (<http://www.ncbi.nlm.nih.gov/omim>).

The screenshot shows the OMIM entry for #261600, PHENYLKETONURIA; PKU. The page includes the following information:

- Alternative titles; symbols:** PHENYLALANINE HYDROXYLASE DEFICIENCY, PAH DEFICIENCY, OLIGOPHRENIA PHENYLPYRUVICA, FOLLING DISEASE
- Other entities represented in this entry:** HYPERPHENYLALANINEMIA, NON-PKU MILD, INCLUDED; HPA, NON-PKU MILD, INCLUDED; PHENYLKETONURIA, MATERNAL, INCLUDED
- Phenotype Gene Relationships:** A table showing the relationship between the phenotype and the gene/locus.

Location	Phenotype	Phenotype MIM number	Gene/Locus	Gene/Locus MIM number
12q23.2	Phenylketonuria	261600	PAH	612349
12q23.2	[Hyperphenylalaninemia, non-PKU mild]	261600	PAH	612349

Εικόνα 3.22 Αναζήτηση δεδομένων από την OMIM (<http://www.ncbi.nlm.nih.gov/omim>).

Το συγκεκριμένο project ξεκίνησε στις αρχές της δεκαετίας του 1960 και είναι πλέον διαθέσιμο σήμερα ως βιβλίο στη 12^η έκδοση του. Η online έκδοση της OMIM

είναι διαθέσιμη από το 1987. Όλες αυτές οι πληροφορίες που είναι καταχωρημένες στην OMIM έχουν συλλεγεί και επεξεργαστεί από τον Dr. 'Victor A. McKusick' στο Johns Hopkins University ενώ ήταν σημαντική η βοήθεια μίας μεγάλης ομάδας επιστημονικών συγγραφέων και εκδοτών.

Σε κάθε ασθένεια και γονίδιο αντιστοιχεί ένας εξαψήφιος αριθμός, όπου το αρχικό ψηφίο υποδηλώνει τον τρόπο κληρονόμησης. Αν ο αρχικός αριθμός είναι -1-, σημαίνει ότι το συγκεκριμένο γονίδιο κληρονομείται με αυτοσωμικό επικρατή τρόπο, αν είναι -2- με αυτοσωμικό υπολειπόμενο, αν -3- με X -συνδεδεμένο. Επίσης ο αστερίσκος (*) πριν το νούμερο υποδηλώνει ότι ο τρόπος κληρονόμησης είναι γνωστός ενώ η δίεση (#) ότι πολύ πιθανότητα ο φαινότυπος της νόσου οφείλεται σε μια μετάλλαξη ενός, δύο ή περισσότερων γονιδίων.

3.8 GeneCards

Η GeneCards (<http://www.genecards.org/>) είναι μία ολοκληρωμένη βάση δεδομένων η οποία περιλαμβάνει πολλές πληροφορίες που αφορούν γονίδια, όπως πληροφορίες γενωμικού, μεταγραφικού και πρωτεομικού ενδιαφέροντος. Επιπλέον περιλαμβάνει δεδομένα που αφορούν ορθόλογες σχέσεις γονιδίων, πολυμορφισμούς, SNPs, πληροφορίες που αφορούν την γονιδιακή έκφραση, καθώς και τη συσχέτιση γονιδίων με ασθένειες. Η GeneCards επίσης παρέχει στον χρήστη πληροφορίες που αφορούν εκτός από πρωτεΐνες, ψευδογονίδια, RNA γονίδια και συμπλέγματα γονιδίων (Εικ.3.23).

Search Results Sorted by Relevance Score

5268 GeneCards for pearson disease • Click + below for [hit context](#) • Click symbol for GeneCard showing 1-20 show 20

	Symbol ▲▼	Description	Category ▲▼	GIfts ▲▼	GC id	Score ▲▼
1	GBA	glucosidase, beta, acid	protein-coding	72	GC01M155204	1.06
2	NOD2	nucleotide-binding oligomerization domain containing 2	protein-coding	59	GC16P050731	0.87
3	STH	saitohin	protein-coding	39	GC17P044076	0.80
Disorders (3): <ul style="list-style-type: none"> • alzheimers disease / neurodegenerative diseases / parkinson disease • Saitohin, which is... neurodegenerative diseases, interacts with... peroxiredoxin 6; Gao L. 2005 • No evidence for an... Alzheimer's disease; Cook L. 2002 • TAU haplotype and... Tau in Alzheimer's disease and are not... Parkinson's disease; Johansson A. 2005 • The Saitohin 'Q7R'... Alzheimer disease and Parkinson's disease cohorts; Clark L.N. 2003 • Molecular evolution... Alzheimer's disease and argyrophilic... grain disease; Conrad C. 2004 						
4	APOE	apolipoprotein E	protein-coding	68	GC19P045408	0.69
5	NDP	Norrie disease (pseudoglioma)	protein-coding	60	GC0XM043808	0.68

Εικόνα 3.23 Εξόρυξη πληροφοριών με τη χρήση της GeneCards (<http://www.genecards.org/>).

ΚΕΦΑΛΑΙΟ 4

ΑΠΟΤΕΛΕΣΜΑΤΑ – ΣΥΖΗΤΗΣΗ

Οι βασικές ομάδες γονιδίων που μελετήθηκαν.

Στην παρούσα εργασία (όπως αναφέρθηκε και σε προηγούμενο κεφάλαιο) μελετήθηκαν τέσσερις κύριες ομάδες γονιδίων. Για να διευκολυνθεί η συγγραφή των αποτελεσμάτων, χρησιμοποιήθηκαν συντομεύσεις για αυτές τις τέσσερις ομάδες (mt_mt genes, mt_genes, nuc_genes, mt_mt interactors). Συγκεκριμένα:

- mt_mt_genes: Τα γονίδια που κωδικοποιούνται στο μιτοχονδριακό DNA(και κωδικοποιούν πρωτεΐνες)
- mt_genes: Γονίδια πυρηνικής προελεύσεως που εμπλέκονται στη δομή ή/και λειτουργία του μιτοχονδρίου, σύμφωνα με τη Γονιδιακή Οντολογία (Gene Ontology).
- nuc_genes: Γονίδια πυρηνικής προελεύσεως που ΔΕΝ φαίνονται να εμπλέκονται στην δομή ή/και λειτουργία του μιτοχονδρίου, με βάση τη Γονιδιακή Οντολογία (Gene Ontology).
- mt_mt interactors: Γονίδια πυρηνικής προελεύσεως που με βάση την BioGrid αλληλεπιδρούν άμεσα με κάποιο/α από τα mt_mt_genes.

Τα βασικά στάδια της μελέτης ήταν:

- Να βρεθούν τα ανθρώπινα γονίδια για την κάθε μία από τις παραπάνω κατηγορίες, με τη βοήθεια της γονιδιακής οντολογίας.
- Να βρεθούν οι πρωτεϊνικές αλληλεπιδράσεις των παραπάνω ανθρώπινων γονιδίων από την BioGrid.
- Να βρεθούν ποιά ανθρώπινα γονίδια εμπλέκονται σε ασθένειες από την GeneCards.
- Να βρεθούν τα ορθόλογα των παραπάνω ανθρώπινων γονιδίων σε πέντε άλλα θηλαστικά, από την Ensembl, με τη χρήση του Biomart, καθώς επίσης να εξαχθούν οι τιμές των dN/dS και του βαθμού συντήρησης των ορθόλογων ακολουθιών (% Identity).

- Να ενσωματωθούν όλα τα παραπάνω δεδομένα με τη βοήθεια της γλώσσας προγραμματισμού Perl και να γίνουν στατιστικές αναλύσεις με τη βοήθεια της γλώσσας προγραμματισμού R.

Αρχικά με τη βοήθεια της gene ontology (<http://www.geneontology.org>), (Εικ.4.1) κατεβάσαμε από το *annotation details and downloads* τα gene ontology terms (όρους γονιδιακής οντολογίας) που αφορούν όλες τις ανθρώπινες πρωτεΐνες. (Εικ.4.2).

Welcome to the Gene Ontology website!

The Gene Ontology project is a major bioinformatics initiative with the aim of standardizing the representation of gene and gene product attributes across species and databases. The project provides a controlled vocabulary of terms for describing gene product characteristics and gene product annotation data from GO Consortium members, as well as tools to access and process this data. Read more about the Gene Ontology...

Search the Gene Ontology Database

Search for genes, proteins or GO terms using AmiGO:

gene or protein name GO term or ID

AmiGO is the official GO browser and search engine. Browse the Gene Ontology with AmiGO.

The Gene Ontology project very much encourages input from the community into both the content of the GO and annotation using GO. We are very happy to work with others to ensure that the GO is both complete and accurate, and we also very much encourage communities to submit GO annotations for inclusion in the GO database. Please contact us.

Quick Links

- Tools
- AmiGO browser
- OBO-Edit ontology editor
- Ontology downloads
- Annotation downloads
- Database downloads
- Documentation
- GO FAQ
- GO on SourceForge
- Contact GO

News

- GO on Twitter
- Cardiovascular GO
- Annotation Initiative
- Newsletter June 12 (8 days ago) [News item](#)
- Open post for collaboration with human phenotype ontology (51 days ago)

Εικόνα 4.1 Αρχική σελίδα της Gene Ontology. (<http://www.geneontology.org>)

					README
<i>Drosophila melanogaster</i> FlyBase	13440	84535 (69503 non-IEA)	3/5/2012	annotations [1.4 mb]	README
<i>Escherichia coli</i> EcoCyc & EcoliHub	3291	10337 (9518 non-IEA)	5/22/2012	annotations [204.4 kb]	README
<i>Gallus gallus</i> GO Annotations @ EBI	16898	97272 (5809 non-IEA)	6/25/2012	annotations [1.1 mb]	README
<i>Homo sapiens</i> GO Annotations @ EBI	45259	342483 (172497 non-IEA)	6/25/2012	annotations [4.7 mb]	README
<i>Leishmania major</i> Sanger GeneDB	375	1035 (1035 non-IEA)	3/15/2012	annotations [22.6 kb]	README
<i>Magnaporthe grisea</i> PAMGO	11274	27627 (27627 non-IEA)	3/30/2012	annotations [339.2 kb]	README
<i>Mus musculus</i> MGI	25594	279648 (188130 non-IEA)	6/27/2012	annotations [3.4 mb]	README
Oomycetes PAMGO	30	126 (126 non-IEA)	5/31/2010	annotations [2.4 kb]	README

Εικόνα 4.2 Download των Gene Ontology terms που αφορούν τον άνθρωπο από το annotation download (<http://www.geneontology.org>).

Κατεβάσαμε λοιπόν το συγκεκριμένο αρχείο το βάλουμε σ' ένα excel και ένα μικρό υπόδειγμα των πληροφοριών που ανακτήσαμε φαίνεται στον παρακάτω πίνακα (Πιν.1).

<i>D B Object Id</i>	<i>DB Object Symbol</i>	<i>GO Id</i>	<i>DB:Refere nce</i>	<i>Eviden ce Code</i>	<i>A s p e c t</i>	<i>D B Object Name</i>	<i>D B Object Synony m</i>	<i>D B Object Type</i>	<i>Tax on</i>	<i>Dat e</i>	<i>Assigned By</i>	
U n i P r o t K B	P114 98	PC	GO: 000 573 9	PMID:167 29965	IDA	C	Pyruvate carboxyl ase, mitocho ndrial	PYC_H UMAN PC IPI0 0299402 Q16705	protein	taxo n:96 06	200 712 24	MGI
U n i P r o t K B	P188 87	XRC C1	GO: 000 001 2	GO_REF:0 000002	IEA	P	DNA Repair Protein XRCC1	XRCC1 _HUMA N XRC C1 IPI00 002564 Q6IBS4 Q9HCB 1	protein	taxo n:96 06	201 210 21	UniPro
U n i P r o t K B	G5E9 54	CCN C	GO: 001 659 1	GO_REF:0 000002	IEA	C	Cyclin C, isoform CRA_c	G5E954 _HUMA N hCG_ 1757197 CCNC	protein	taxo n:96 06	201 201 21	InterPro
U n i P r o t K B	P487 35	IDH2	GO: 000 573 9	PMID:180 29348	IDA	C	Isocitrat e dehydro genase [NADP] , mitocho ndrial	IDHP_H UMAN I DH2 IPI 0001110 7 B2R6 L6 Q96 GT3	protein	taxo n:96 06	200 906 16	HPA

U n 11	P494 TUF M	GO: 000 573 9	PMID:180 29348	IDA	C	Elongati on factor Tu, mitocho ndrial	EFTU_ HUMA N TUF M IPI00 027107 O15276	protein	taxo n:96 06	201 003 26	HPA
U n 59	P188 ATP5 J	GO: 000 573 9	PMID:121 10673	IC	C	ATP synthase - coupling factor 6, mitocho ndrial	ATP5J_ HUMA N ATP5 J ATP5 A ATP M IPI00 002521	protein	taxo n:96 06	201 201 21	GOC

Πίνακας 1. Πίνακας ενδεικτικών αποτελεσμάτων με τη βοήθεια της *gene ontology*. Η πρώτη στήλη αναφέρεται στη βάση δεδομένων όπου έχει αντληθεί το μοναδικό αναγνωριστικό Id της δεύτερης στήλης. Στην επόμενη στήλη αναγράφεται ο συμβολισμός του γονιδίου. Στην τέταρτη στήλη γίνεται αναφορά του αναγνωριστικού GO Id και στη συνέχεια παρατηρείται η μοναδική πηγή από την οποία έχει αντληθεί το αναγνωριστικό GO Id. Στην έκτη στήλη περιγράφονται οι μέθοδοι ανάλυσης όπου προήλθαν τα συγκεκριμένα γονιδιακά προϊόντα. Η έβδομη στήλη μας δίνει πληροφορίες σχετικά με όρους που αντιπροσωπεύουν ιδιότητες γονιδιακών προϊόντων και λειτουργιών, πχ το P (*biological process*) αναφέρεται σε βιολογικές λειτουργίες, το C (*cellular component*) στα κυτταρικά διαμερίσματα όπου παρατηρείται το βιομόριο και το F (*molecular function*) σε μοριακές λειτουργίες αυτού. Στην επόμενη στήλη αναγράφεται το όνομα του γονιδίου ή του γονιδιακού προϊόντος και στη συνέχεια παρατηρείται ο συμβολισμός του γονιδίου και συνώνυμα αυτών. Στη δέκατη στήλη παίρνουμε πληροφορίες σχετικά με το είδος του γονιδιακού προϊόντος (στη συγκεκριμένη εργασία ασχολούμαστε μόνο με πρωτεΐνες) και στην ενδέκατη στήλη αναγράφεται το αναγνωριστικό της ταξινόμησης. Στην δωδέκατη στήλη αναφέρεται η ημερομηνία όπου έγινε ο σχολιασμός και στην τελευταία στήλη αναγράφεται η βάση δεδομένων όπου έχει κάνει αυτόν τον σχολιασμό.

Ανακτήθηκαν 357.891 GO descriptions για όλα τα ανθρώπινα γονίδια. Ένα γονίδιο μπορεί να έχει περισσότερα του ενός GO descriptions.

Εν συνεχεία, με τη βοήθεια της perl και με τη χρήση των Regular Expressions γράψαμε ένα κώδικα (perl script) με σκοπό την ανάκτηση μόνο το GO Ids που έχουν σχέση με τα μιτοχόνδρια. Έτσι λοιπόν πήραμε μια λίστα με 289 μοναδικά GO Ids που σχετίζονται με το μιτοχόνδριο. Ένα μικρό παράδειγμα δίνεται στην παρακάτω εικόνα (Εικ 4.3)

GO ID NUMBER	DESCRIPTION OF FUNCTION	ASPECT
GO:0000001	mitochondrion inheritance	P
GO:0000002	mitochondrial genome maintenance	P
GO:0000066	mitochondrial ornithine transport	P
GO:0000262	mitochondrial chromosome	C
GO:0000266	mitochondrial fission	P
GO:0000274	mitochondrial proton-transporting ATP synthase, stator stalk	C
GO:0000275	mitochondrial proton-transporting ATP synthase complex, catalytic core F(1)	C
GO:0000276	mitochondrial proton-transporting ATP synthase complex, coupling factor F(o)	C
GO:0000422	mitochondrion degradation	P
GO:0000957	mitochondrial RNA catabolic process	P
GO:0000958	mitochondrial mRNA catabolic process	P
GO:0000959	mitochondrial RNA metabolic process	P
GO:0000960	regulation of mitochondrial RNA catabolic process	P
GO:0000961	negative regulation of mitochondrial RNA catabolic process	P
GO:0000962	positive regulation of mitochondrial RNA catabolic process	P
GO:0000963	mitochondrial RNA processing	P
GO:0000964	mitochondrial RNA 5'-end processing	P
GO:0000965	mitochondrial RNA 3'-end processing	P
GO:0000997	mitochondrial RNA polymerase core promoter sequence-specific DNA binding	F
GO:0001001	mitochondrial single-subunit type RNA polymerase binding	F
GO:0001018	mitochondrial RNA polymerase regulatory region DNA binding	F
GO:0001044	mitochondrial RNA polymerase regulatory region sequence-specific DNA binding	F
GO:0001045	mitochondrial RNA polymerase core promoter proximal region sequence-specific DNA binding	F

Εικόνα 4.3. Λίστα μοναδικών Go Ids που έχουν σχέση με τα μιτοχόνδρια στον άνθρωπο. Στην πρώτη στήλη φαίνεται το Go Id αναγνωριστικό, στη δεύτερη περιγράφεται η λειτουργία στην οποία εμπλέκεται το συγκεκριμένο Id number και στην τρίτη στήλη γίνεται αναφορά του τομέα όπου εμπλέκεται το συγκεκριμένο GO Id.

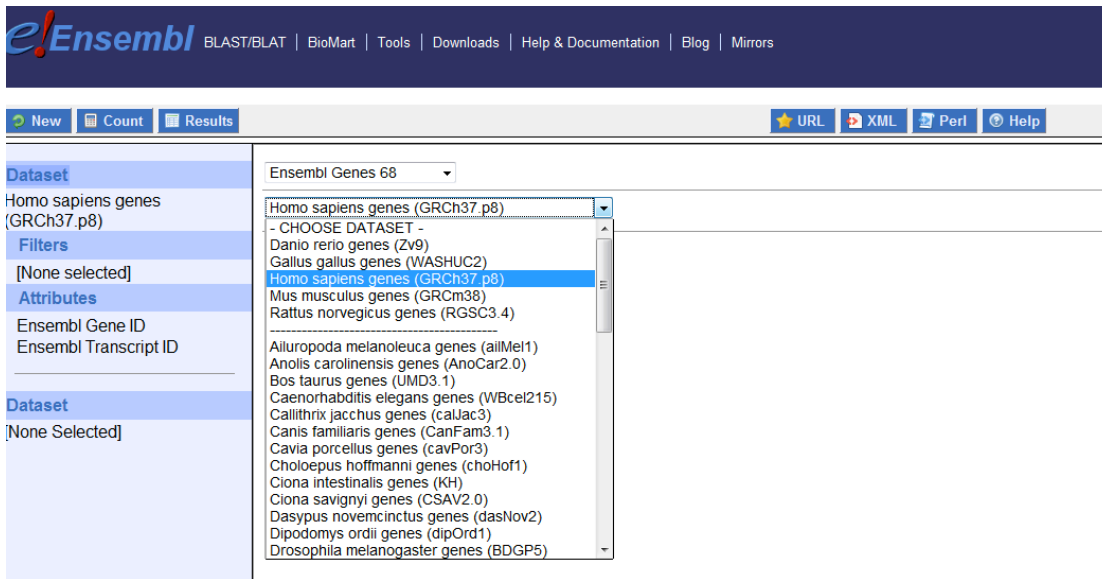
Κατόπιν με τη βοήθεια της Biomart και με τη χρήση των παραπάνω 289 μοναδικών μιτοχονδριακών GO Ids ανακτήσαμε όλα τα ανθρώπινα γονίδια που σχετίζονται με τα μιτοχόνδρια και συγκεκριμένα εξορύξαμε τα Ensembl Gene Ids, Ensembl Transcript Ids καθώς και τα Ensembl Protein Ids. Βρέθηκαν συνολικά 4.237 μιτοχονδριακά Ensembl Transcript Ids. Στη συνέχεια με τη χρήση της Biomart κατεβάσαμε και τις ακολουθίες αυτών των μεταγράφων και με τη βοήθεια των Perl scripts φιλτράραμε και πήραμε μόνο τα μεγαλύτερα μετάγραφα αυτών των γονιδίων. Έτσι λοιπόν βρέθηκαν συνολικά 1702 Ensembl Gene Ids (γονίδια) (και τα μεγαλύτερα μετάγραφα τους) που σχετίζονται με τα μιτοχόνδρια στον άνθρωπο. Παρακάτω φαίνεται αναλυτικά όλη η διαδικασία που έχει περιγραφεί πιο πάνω.

Η βάση δεδομένων που χρησιμοποιούμε προκειμένου να αντλήσουμε πληροφορίες είναι η Ensembl (Εικ.4.4)

The screenshot shows the Ensembl Genome Browser interface. At the top, there is a navigation bar with links for BLAST/BLAT, BioMart, Tools, Downloads, Help & Documentation, Blog, and Mirrors. A search bar is located in the top right corner. Below the navigation bar, there is a search area with a dropdown menu for 'All species' and a 'Go' button. The main content area is divided into several sections: 'Browse a Genome' with a description of the project, 'Popular genomes' featuring Human, Mouse, and Zebrafish, and 'All genomes' with a species selection dropdown. On the right side, there is a 'New to Ensembl?' section with several links for learning and data handling, a 'Did you know...?' box, and a 'What's New in Release 68' section.

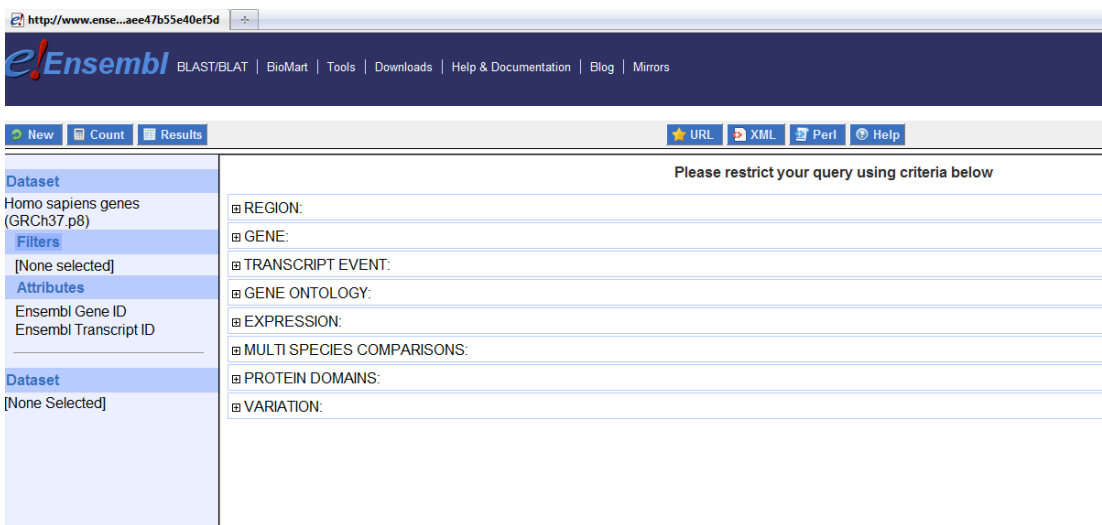
Εικόνα 4.4 Αρχική σελίδα της βάσης δεδομένων *Ensembl Genome Browser*.

Στη συνέχεια κάνουμε κλικ στη Biomart που όπως έχουμε ήδη αναφέρει αποτελεί ένα εργαλείο εξόρυξης γονιδίων, ακολουθιών και πολλών άλλων πληροφοριών και επιλέγουμε στη Database το Ensembl Genes 68 και στη Dataset το πεδίο που αναφέρετε στον άνθρωπο (*Homo Sapiens Genes*), δεδομένου ότι αναζητούμε ανθρώπινα γονίδια (Εικ.4.5)



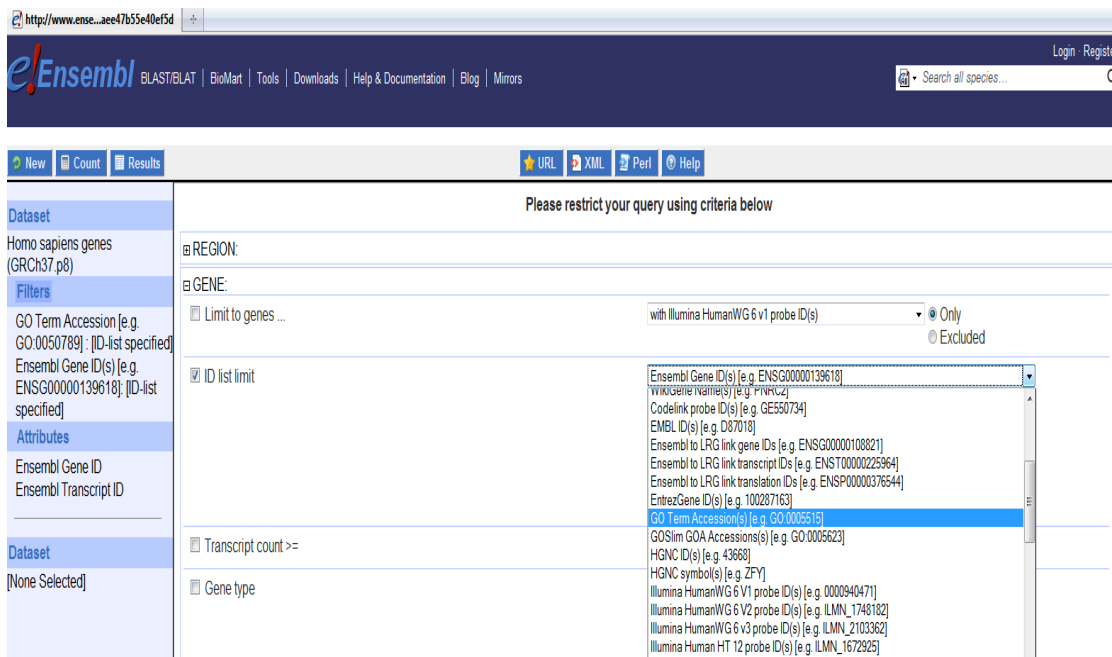
Εικόνα 4.5 Αναζήτηση δεδομένων στη Biomart με τη βοήθεια των φίλτρων.

Στη συνέχεια από τα filters μας ανοίγει ένα πεδίο με τη βοήθεια του οποίου εμείς μπορούμε να ορίσουμε τι ακριβώς θέλουμε να αναζητήσουμε (Εικ.4.6)



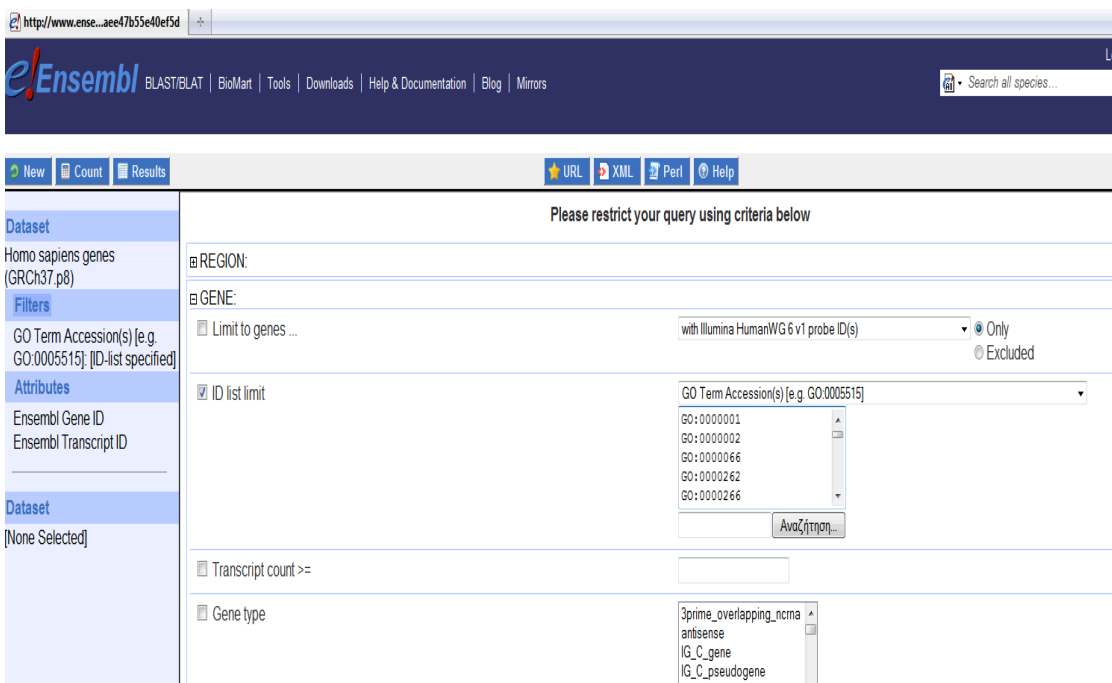
Εικόνα 4.6 Αναζήτηση δεδομένων στη Biomart με τη βοήθεια των φίλτρων.

Δεδομένου ότι θέλουμε να αναζητήσουμε γονίδια επιλέγουμε το πεδίο που αναφέρετε στα γονίδια και συγκεκριμένα εφόσον θέλουμε να εξορύξουμε δεδομένα χρησιμοποιώντας τα 289 μοναδικά μιτοχονδριακά GO Ids επιλέγουμε από τη λίστα ότι θέλουμε να αντλήσουμε δεδομένα από τα GO Ids. (Εικ.4.7)



Εικόνα 4.7 Αναζήτηση δεδομένων στη Biomart με τη βοήθεια των φίλτρων.

Έπειτα εισάγαμε τα 289 GO Ids στο συγκεκριμένο πεδίο και στη συνέχεια από τα attributes επιλέγουμε τα Ensembl Protein Ids και το πεδίο που αναφέρεται στη Description προκειμένου να μας τα αναρτήσει στα αποτελέσματα (Εικ.4.8-4.9)



Εικόνα 4.8 Αναζήτηση δεδομένων στη Biomart με τη βοήθεια των φίλτρων.

Εικόνα 4.9 Αναζήτηση δεδομένων στη Biomart με τη βοήθεια των φίλτρων.

Στη συνέχεια πατάμε τα Results προκειμένου να πάρουμε τα αποτελέσματα, τα οποία έχουν την παρακάτω μορφή (Εικ.4.10-4.11).

Εικόνα 4.10 Αναζήτηση δεδομένων στη Biomart με τη βοήθεια των φίλτρων. Αρχικά παίρνουμε τις 10 πρώτες γραμμές. Μπορούμε όμως αν επιλέξουμε το all να πάρουμε όλα τα αποτελέσματα.

Ensembl Gene ID	Ensembl Transcript ID	Ensembl Protein ID	Description
ENSG0000025708	ENST00000395680	ENSP00000379037	thymidine phosphorylase [Source:HGNC Symbol;Acc:3148]
ENSG0000025708	ENST00000252029	ENSP00000252029	thymidine phosphorylase [Source:HGNC Symbol;Acc:3148]
ENSG0000025708	ENST00000395678	ENSP00000379036	thymidine phosphorylase [Source:HGNC Symbol;Acc:3148]
ENSG00000068305	ENST00000560493	ENSP00000453478	myocyte enhancer factor 2A [Source:HGNC Symbol;Acc:6993]
ENSG00000068305	ENST00000354410	ENSP00000346389	myocyte enhancer factor 2A [Source:HGNC Symbol;Acc:6993]
ENSG00000068305	ENST00000558812	ENSP00000454120	myocyte enhancer factor 2A [Source:HGNC Symbol;Acc:6993]
ENSG00000068305	ENST00000557785	ENSP00000453441	myocyte enhancer factor 2A [Source:HGNC Symbol;Acc:6993]
ENSG00000068305	ENST00000558049	ENSP00000454023	myocyte enhancer factor 2A [Source:HGNC Symbol;Acc:6993]
ENSG00000068305	ENST00000557942	ENSP00000453095	myocyte enhancer factor 2A [Source:HGNC Symbol;Acc:6993]
ENSG00000068305	ENST00000559903	ENSP00000452989	myocyte enhancer factor 2A [Source:HGNC Symbol;Acc:6993]
ENSG00000068305	ENST00000453228	ENSP00000404110	myocyte enhancer factor 2A [Source:HGNC Symbol;Acc:6993]
ENSG00000068305	ENST00000338042	ENSP00000337202	myocyte enhancer factor 2A [Source:HGNC Symbol;Acc:6993]
ENSG00000081189	ENST00000340208	ENSP00000340874	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]
ENSG00000081189	ENST00000424173	ENSP00000389610	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]
ENSG00000081189	ENST00000514028	ENSP00000426665	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]
ENSG00000081189	ENST00000437473	ENSP00000396219	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]
ENSG00000081189	ENST00000510942	ENSP00000422390	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]
ENSG00000081189	ENST00000506554	ENSP00000425636	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]
ENSG00000081189	ENST00000508569	ENSP00000423597	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]
ENSG00000081189	ENST00000514015	ENSP00000424606	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]
ENSG00000081189	ENST00000513252	ENSP00000423826	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]

Εικόνα 4.11 Αναζήτηση δεδομένων στη Biomart με τη βοήθεια των φίλτρων. Ανάρτηση όλων των αποτελεσμάτων.

Κατόπιν εξάγουμε τα αποτελέσματα και τα αποθηκεύουμε σε μορφή excel (Εικ.4.12).

ENSEMBL GENE ID	ENSEMBL TRANSCRIPT ID	ENSEMBL PROTEIN ID	DESCRIPTION
ENSG00000115204	ENST00000233545	ENSP00000233545	MpV17 mitochondrial inner membrane protein [Source:HGNC Symbol;Acc:7224]
ENSG00000115204	ENST00000380044	ENSP00000380044	MpV17 mitochondrial inner membrane protein [Source:HGNC Symbol;Acc:7224]
ENSG0000025708	ENST00000395680	ENSP00000379037	thymidine phosphorylase [Source:HGNC Symbol;Acc:3148]
ENSG0000025708	ENST00000252029	ENSP00000252029	thymidine phosphorylase [Source:HGNC Symbol;Acc:3148]
ENSG0000025708	ENST00000395678	ENSP00000379036	thymidine phosphorylase [Source:HGNC Symbol;Acc:3148]
ENSG00000081189	ENST00000340208	ENSP00000340874	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]
ENSG00000081189	ENST00000424173	ENSP00000389610	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]
ENSG00000081189	ENST00000514028	ENSP00000426665	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]
ENSG00000081189	ENST00000437473	ENSP00000396219	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]
ENSG00000081189	ENST00000510942	ENSP00000422390	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]
ENSG00000081189	ENST00000506554	ENSP00000425636	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]
ENSG00000081189	ENST00000508569	ENSP00000423597	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]
ENSG00000081189	ENST00000514015	ENSP00000424606	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]
ENSG00000081189	ENST00000513252	ENSP00000423826	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]
ENSG00000081189	ENST00000506716	ENSP00000423656	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]
ENSG00000081189	ENST00000507984	ENSP00000424331	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]
ENSG00000081189	ENST00000502983	ENSP00000427163	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]
ENSG00000081189	ENST00000508610	ENSP00000426442	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]
ENSG00000081189	ENST00000502831	ENSP00000427286	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]
ENSG00000081189	ENST00000503075	ENSP00000426465	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]
ENSG00000081189	ENST00000509373	ENSP00000427309	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]
ENSG00000081189	ENST00000504921	ENSP00000421925	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]
ENSG00000081189	ENST00000539796	ENSP00000441153	myocyte enhancer factor 2C [Source:HGNC Symbol;Acc:6996]

Εικόνα 4.12 Εξόρυξη αποτελεσμάτων από τη Biomart και αποθήκευση σε μορφή excel.

Έπειτα κατεβάσαμε και τις ακολουθίες των παραπάνω 4.237 μεταγράφων επίσης από τη Biomart και πήραμε την παρακάτω εικόνα (Εικ.4.13-4.14-4.15)

Εικόνα 4.13 Εξόρυξη ακολουθιών από τη Biomart.

Εικόνα 4.14 Εξόρυξη ακολουθιών από τη Biomart.

```
http://www.ense...582237b33834581 x Φόρτωση... x +
>ENSG00000013583|ENST00000014930
MLGMIKNSLFGSVETWPWQVLSKGDKEEVAYEERACEGGKFATVEVTDKPVDEALREAMP
KVAKYAGGTNDKGI GMGMTVPI SFAVFPNEDGSLQKKLKVWFRI PNQFQSDPPAPSDKSV
KIEEREGITVYSMQFGGYAKEADYVAQATRLRAALEGTATYRGDIYFCTGYDPPMKPYGR
RNEIWLLKT*
>ENSG00000008405|ENST00000008527
MGVNAVHWFRKGLRLHDNPALKECIQGADTIRC VYILD PWFAGSSNVGINRWRFL LQCLE
DL DANLRKLNSRL FVIRGQPADVFPRLFK EWNITKLSIEYDSE PFGKERDAAIKKLATEA
GVEVIVRISHTLYDLDKI IELNGGQPPLTYKRFQTLISKMEPLEIPVETITSEVIEKCTT
PLSDDHDEKYGVPSLEELGFDTDGLSSAVWPGGETEALTRLERHLERKAWVANFERPRMN
ANSL LASPTGLSPYLRFGLSCLRFYFKLTDLYKKVKKNSSPPLSLY GQLLWREFFYTAA
TNNPRFDKMEGNPICVQIPWDKNPEALAKWAEGR TGFPWIDAIMTQLRQEGWIHHLARHA
VACFLTRGDLWISWEEGMKVFEELLLDADWSINAGSWMWLS CSSFFQQFFHCYCPVGFGR
RIDPNGDYIRRYLPVLRGFP AKYIYDPWNAPEGIQKVAKCLIGVNYPKPMVNHA EASRLN
IERMKQIYQQLSRYRGLGLLASVPSNPNGNGGFMGYS AENIPGCSSSGSCSQSGILHYA
HGDSQQTHLLKQGRSSMGTGLSGGKRPSQEEDTQSIGPKVQRQSTN*
>ENSG00000004779|ENST00000007516
MASRVLSAYVSR LPAAFAPLPRVRMLAVARPLSTALCSAGTQTRLGLTLPALVLAQVPGR
VIQLCRQYSDMPPLTLEGIQDRVLYVLKLYDKIDPEKLSVNSHF MKDLGLDSDLQVEIIM
AMEDEFGFEIPDIDA EKLMCPQEIVDYIADKKDVYE*
>ENSG00000006625|ENST00000005374
MANS GCKDVTGPDEESFLYFAYGSNLLTERIHLRNP SAAFFCVARLQDFKLD FGN SQGKT
SQTWHGGIATIFQSPGDEVWGVVWKMKN SNLNSLDELFAWVQKQK M VCRWSIKRS*
>ENSG000000003509|ENST000000002125
MSVLLRSLGLPLCAVARAAI PFIWRGKYFSSGNEPAENPVT PMLRHLMYKIKSTGPITVA
EYMKEVLTNPAKGYVYRDMLGEKGFITSPEISQIFGELLGIWFI SEWMATGKSTAFQL
VELGPRGRTLVDILRVFTQLG SVLKNCDISVHLVEVSQKLSEIQALTLTKEKVPLERNA
GSPVYMKGVTKSGIPI SWYRDLHDV PKGYSFYLAHEFFDVL PVHKFQKTPQGWREVFVDI
DPQVSDKLRFLV LAPSATPAEAFIQHDETRDHVEVCPDAGVIEELSQR IALTGGAALVAD
YGHGDKTKTDTFRGFCDHKLHDVLIAPGTADLTADVDFSYLRRMAQ GKVASLGPIKQHTFL
KNMGIDVRLKVL LDKSNEPSVRQQLLQGYDMLMNP KKMGERFNFFALLPHQR LQGGRYQR
NARQSKPFASVVAGFSELA WQ*
>ENSG000000239672|ENST00000013034
MMLISTIGTVEQCFEPTISSCPTGTMANCFETETA LKEDGVORGIYCFI LKDFEOKGPT
```

Εικόνα 4.15 Εξόρυξη ακολουθιών από τη Biomart.

Στη συνέχεια με τη χρήση των Perl scripts και των Regular Expressions φιλτράραμε τα παραπάνω αποτελέσματα που αφορούν τις ακολουθίες, εξάγαμε τις ακολουθίες των μεγαλύτερων μεταγράφων και πήραμε τα αποτελέσματα σε μορφή excel (Εικ.4.16)

ENSEMBL GENE ID	ENSEMBL TRANSCRIPT ID	NUMBER OF AMINOACIDS
ENSG00000000938	ENST00000374003	530
ENSG00000001084	ENST00000229416	638
ENSG00000002330	ENST00000309032	169
ENSG00000002549	ENST00000226299	520
ENSG00000002822	ENST00000265854	719
ENSG00000003509	ENST00000002125	442
ENSG00000004139	ENST00000003834	691
ENSG00000004455	ENST00000354858	240
ENSG00000004779	ENST00000007516	157
ENSG00000004799	ENST00000005178	412
ENSG00000004864	ENST00000416240	677
ENSG00000004961	ENST00000321143	269
ENSG00000005022	ENST00000317881	299
ENSG00000005187	ENST00000289416	587
ENSG00000005249	ENST00000265717	419
ENSG00000005379	ENST00000343736	1858
ENSG00000005381	ENST00000340482	778
ENSG00000005469	ENST00000419147	641
ENSG00000005882	ENST00000503176	408
ENSG00000006530	ENST00000355413	423
ENSG00000006625	ENST00000275428	189
ENSG00000006695	ENST00000261643	444
ENSG00000006744	ENST00000338034	827

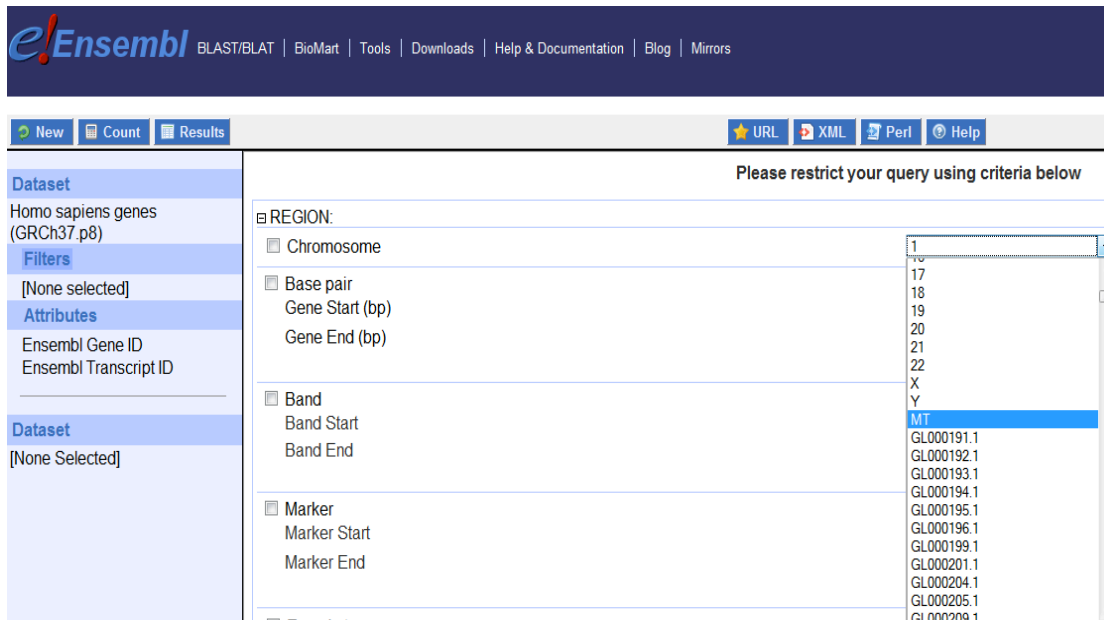
Εικόνα 4.16 Αποτελέσματα των μεγαλύτερων μεταγράφων ανθρώπινων γονιδίων που σχετίζονται με τα μιτοχόνδρια και αποθήκευση σε μορφή excel. Στην τρίτη στήλη φαίνεται ο αριθμός των αμινοξέων της πρωτεΐνης που κωδικοποιεί το κάθε μετάγραφο.

Έτσι λοιπόν όπως έχουμε ήδη αναφέρει και προηγουμένως βρέθηκαν συνολικά 1702 Ensembl Gene Ids (και τα μεγαλύτερα μετάγραφα του) που κωδικοποιούν πρωτεΐνες και που σχετίζονται με τα μιτοχόνδρια στον άνθρωπο.

Σ' αυτό το σημείο θα ήταν μεγάλη παράλειψη να μην αναφέρουμε ότι τα παραπάνω 1702 γονίδια που αντλήθηκαν αφορούν γονίδια που εμπλέκονται στην μιτοχονδριακή λειτουργία όμως δεν έχουν όλα αυτά καθαρά μιτοχονδριακή προέλευση. Για να είμαστε πιο συγκεκριμένοι τα περισσότερα έχουν πυρηνική προέλευση και μόνο 13 γονίδια έχουν αποκλειστικά μιτοχονδριακή προέλευση, δηλ. κωδικοποιούν πρωτεΐνες από το mtDNA και όχι από το πυρηνικό. Πιο αναλυτικά, το mt-DNA κωδικοποιεί 37 γονίδια, από τα οποία μόνο τα 13 δίνουν πρωτεϊνικό προϊόν.

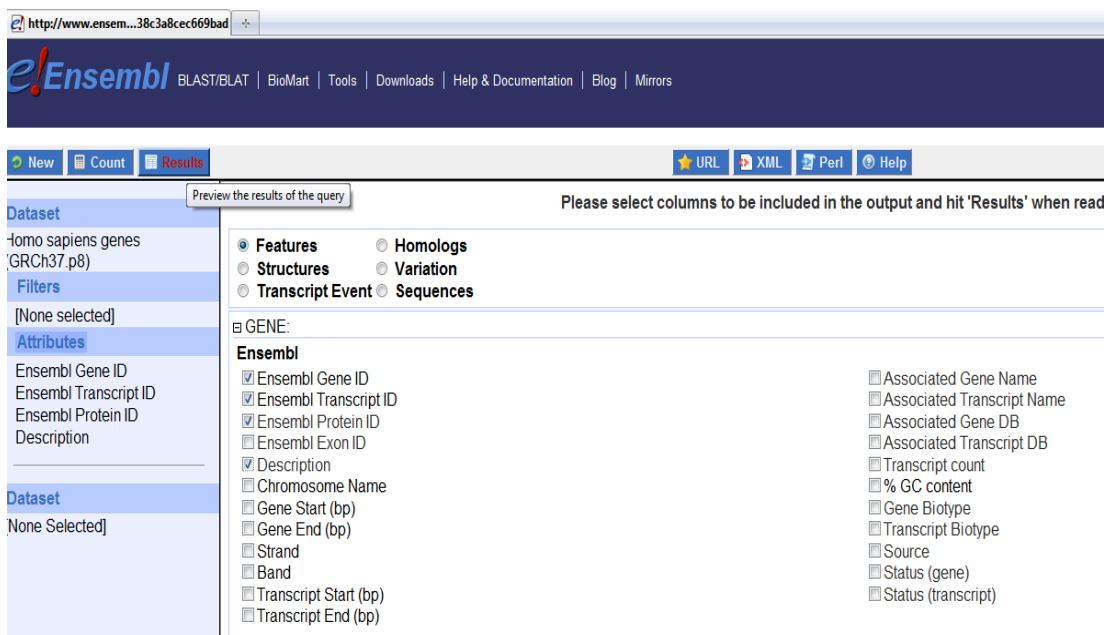
Τα παραπάνω αποτελέσματα τα εξορύξαμε με τη βοήθεια της Biomart. Αρχικά μπήκαμε στην αρχική σελίδα της Biomart και επιλέξαμε από το πεδίο που αναφέρεται στη Database το Ensembl Genes 68 και από το πεδίο που αναφέρεται στο Dataset το Homo Sapiens μιας και αναζητούμε τα γονίδια και τις πρωτεΐνες που κωδικοποιούν από

τον άνθρωπο. Στη συνέχεια επιλέξαμε από τα Filters το πεδίο που αναφέρεται στο Region όπου και επιλέξαμε το MT (μιτοχόνδριο). Είναι η περιοχή που ψάχνουμε να βρούμε τις ανθρώπινες πρωτεΐνες που κωδικοποιούνται από το μιτοχόνδριο και συνεπώς έχουν μιτοχονδριακή και όχι πυρηνική προέλευση (Εικ.4.17).



Εικόνα 4.17 Αναζήτηση δεδομένων στη Biomart με τη βοήθεια των φίλτρων.

Έπειτα επιλέξαμε από τα Attributes τις πληροφορίες που θέλουμε να εξαχθούν από τα αποτελέσματα και πατήσαμε το Results (Εικ.4.18-4.19).



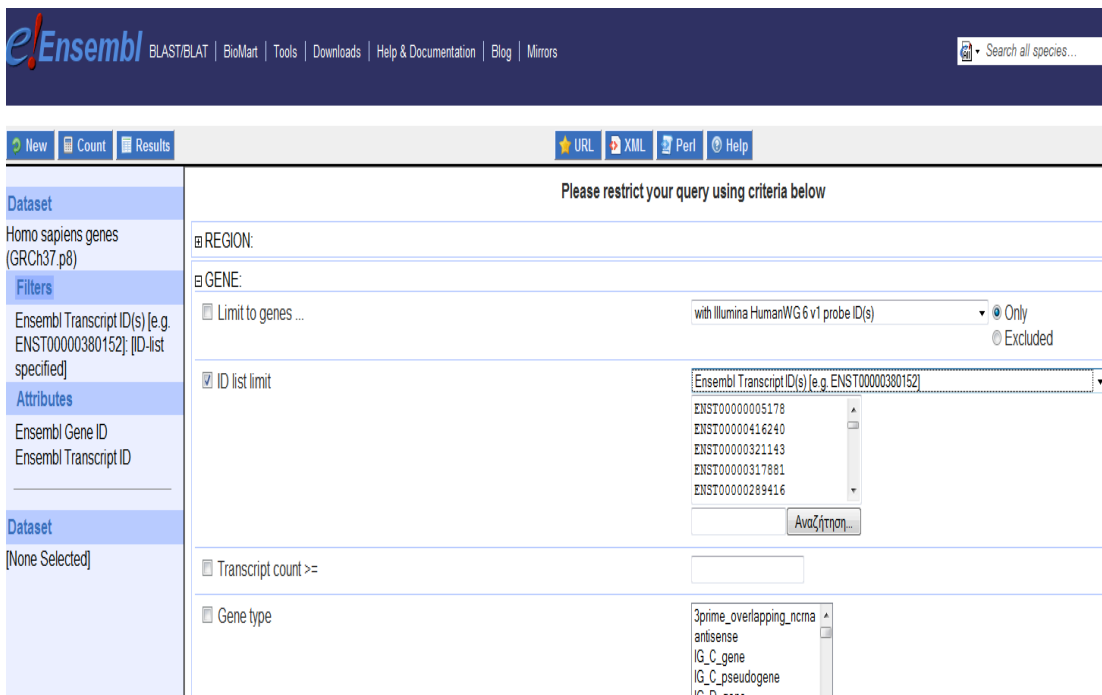
Εικόνα 4.18 Αναζήτηση δεδομένων στη Biomart με τη βοήθεια των φίλτρων.

The screenshot shows the Ensembl BioMart interface. The main content area displays search results for 'Homo sapiens genes (GRCh37.p8)'. The 'Export all results to' section is set to 'File' with 'TSV' as the format and 'Unique results only' checked. The 'View' section is set to '10 rows as HTML' with 'Unique results only' checked. The results table is as follows:

Ensembl Gene ID	Ensembl Transcript ID	Ensembl ID	Description
ENSG00000238042	ENST00000433576		
ENSG00000238042	ENST00000412445		
ENSG00000228437	ENST00000441160		
ENSG00000228437	ENST00000421147		
ENSG00000232679	ENST00000438158		
ENSG00000236230	ENST00000416510		
ENSG00000227613	ENST00000418582		glutamyl-tRNA synthase (glutamine-hydrolyzing)-like 1 pseudogene 2 [Source:HGNC Symbol;Acc:43668]
ENSG00000226643	ENST00000455363		
ENSG00000234419	ENST00000422015		capicua homolog (Drosophila) pseudogene 13 [Source:HGNC Symbol;Acc:37907]
ENSG00000225265	ENST00000413074		

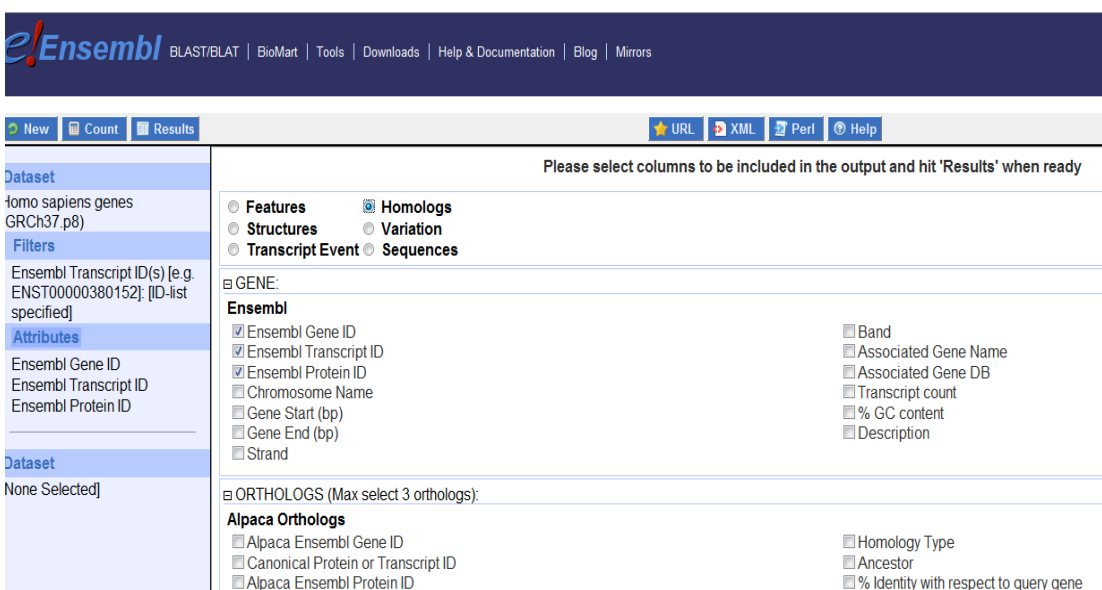
Εικόνα 4.19 Εξόρυξη αποτελεσμάτων από τη Biomart με τη βοήθεια των φίλτρων.

Στην συνέχεια προχωρήσαμε, με το να αναζητήσουμε τα αντίστοιχα ορθόλογα γονίδια των 1702 ανθρώπινων γονιδίων που εμπλέκονται στη μιτοχονδριακή λειτουργία, σε πέντε διαφορετικούς οργανισμούς. Οι οργανισμοί που μας ενδιαφέρουν είναι ο σκύλος (*Canis lupus familiaris*), το κουνέλι (*Oryctolagus cuniculus*), το ποντίκι (*Mus musculus*), ο αρουραίος (*Rattus norvegicus*) και ο χιμπατζής (*Pan troglodytes*). Η αναζήτηση των αντίστοιχων ορθόλογων γονιδίων έγινε με τη βοήθεια της Biomart όπου και αντλήσαμε όλες τις πληροφορίες που θέλαμε. Αρχικά, στην αρχική σελίδα της Biomart επιλέξαμε στο πεδίο που αναφέρετε η Database το Ensembl Genes 68 και στο Dataset το Homo sapiens. Στη συνέχεια, πήραμε τα Ensembl Transcripts Ids των 1702 ανθρώπινων γονιδίων που εμπλέκονται στη μιτοχονδριακή λειτουργία και αποτελούν και τα μεγαλύτερα μετάγραφα και τα θέσαμε ως κριτήρια ώστε να αναζητηθούν τα αντίστοιχα ορθόλογα γονίδια στους πέντε οργανισμούς που αναφέραμε πιο πάνω (Εικ.4.20).



Εικόνα 4.20 Αναζήτηση δεδομένων στη Biomart με τη βοήθεια των φίλτρων.

Κατόπιν επιλέξαμε από τα attributes το πεδίο που αναφέρεται στα ομόλογα και μας άνοιξε ένα παράθυρο που αναφέρεται σε ένα μεγάλο αριθμό οργανισμών όπου ανάλογα με τον οργανισμό που μας ενδιαφέρει μπορούμε να βρούμε τα ορθόλογα (Εικ.4.21).



Εικόνα 4.21 Αναζήτηση δεδομένων στη Biomart .

Έτσι λοιπόν αναζητήσαμε τα αντίστοιχα ορθόλογα στους πέντε διαφορετικούς οργανισμούς που αναφέραμε παραπάνω επιλέγοντας πάντα τις πληροφορίες που θέλουμε να εξορύξουμε και να πάρουμε ως αποτελέσματα. Η παρακάτω διαδικασία έγινε 5 φορές, προκειμένου να πάρουμε αποτελέσματα για τον κάθε οργανισμό ξεχωριστά (Εικ.4.22).

Cat Orthologs	
<input type="checkbox"/> Cat Ensembl Gene ID	<input type="checkbox"/> Homology Type
<input type="checkbox"/> Canonical Protein or Transcript ID	<input type="checkbox"/> Ancestor
<input type="checkbox"/> Cat Ensembl Protein ID	<input type="checkbox"/> % Identity with respect to query gene
<input type="checkbox"/> Cat Chromosome Name	<input type="checkbox"/> % Identity with respect to Cat gene
<input type="checkbox"/> Cat Chromosome Start (bp)	<input type="checkbox"/> dN
<input type="checkbox"/> Cat Chromosome End (bp)	<input type="checkbox"/> dS
Chicken Orthologs	
<input type="checkbox"/> Chicken Ensembl Gene ID	<input type="checkbox"/> Homology Type
<input type="checkbox"/> Canonical Protein or Transcript ID	<input type="checkbox"/> Ancestor
<input type="checkbox"/> Chicken Ensembl Protein ID	<input type="checkbox"/> % Identity with respect to query gene
<input type="checkbox"/> Chicken Chromosome Name	<input type="checkbox"/> % Identity with respect to Chicken gene
<input type="checkbox"/> Chicken Chromosome Start (bp)	<input type="checkbox"/> dN
<input type="checkbox"/> Chicken Chromosome End (bp)	<input type="checkbox"/> dS
Chimp Orthologs	
<input checked="" type="checkbox"/> Chimp Ensembl Gene ID	<input checked="" type="checkbox"/> Homology Type
<input checked="" type="checkbox"/> Canonical Protein or Transcript ID	<input checked="" type="checkbox"/> Ancestor
<input checked="" type="checkbox"/> Chimp Ensembl Protein ID	<input checked="" type="checkbox"/> % Identity with respect to query gene
<input type="checkbox"/> Chimp Chromosome Name	<input checked="" type="checkbox"/> % Identity with respect to Chimp gene
<input type="checkbox"/> Chimp Chromosome Start (bp)	<input checked="" type="checkbox"/> dN
<input type="checkbox"/> Chimp Chromosome End (bp)	<input checked="" type="checkbox"/> dS
Ciona intestinalis Orthologs	
<input type="checkbox"/> Ciona intestinalis Ensembl Gene ID	<input type="checkbox"/> Homology Type
<input type="checkbox"/> Canonical Protein or Transcript ID	<input type="checkbox"/> Ancestor
<input type="checkbox"/> Ciona intestinalis Ensembl Protein ID	<input type="checkbox"/> % Identity with respect to query gene
<input type="checkbox"/> Ciona intestinalis Chromosome Name	<input type="checkbox"/> % Identity with respect to Ciona intestinalis gene
<input type="checkbox"/> Ciona intestinalis Chromosome Start (bp)	<input type="checkbox"/> dN
<input type="checkbox"/> Ciona intestinalis Chromosome End (bp)	<input type="checkbox"/> dS

Εικόνα 4.22 Αναζήτηση ορθόλογων γονιδίων του χιμπατζή από την Biomart.

Έπειτα πατήσαμε τα Results και πήραμε τα αποτελέσματα που θέλαμε για κάθε ζεύγος ορθολόγων γονιδίων συσχετιζόμενων πάντα με τα ανθρώπινα μιτοχονδριακά γονίδια. Ένα μικρό υπόδειγμα μας δείχνει η παρακάτω εικόνα (Εικ.4.23).

Ensembl Gene ID	Ensembl Transcript ID	Chimp Ensembl Gene ID	Canonical Protein or Transcript ID	Chimp Ensembl Protein ID	Homology Type	Ancestor	% Identity with respect to q	% Identity with respect to p	dN	dS
ENSG00000003509	ENST00000002125	ENSPTRG00000011837	ENSP00000002125	ENSPTRP00000020328	ortholog_one2one	Homininae	100	100	0,0022	0,0106
ENSG00000004139	ENST00000003834	ENSPTRG00000008917	ENSP00000003834	ENSPTRP00000015218	ortholog_one2one	Eutheria	93	88	0,0456	0,1406
ENSG00000004799	ENST00000005178	ENSPTRG00000019421	ENSP00000005178	ENSPTRP00000003224	ortholog_one2one	Homininae	100	100	0,001	0,0159
ENSG00000004779	ENST00000007516	ENSPTRG00000007889	ENSP00000007516	ENSPTRP00000013471	ortholog_one2one	Homininae	100	100	0	0,0243
ENSG00000008405	ENST00000008527	ENSPTRG00000005396	ENSP00000008527	ENSPTRP00000009157	ortholog_one2one	Homininae	100	100	0,0009	0,0053
ENSG00000008394	ENST00000010404	ENSPTRG00000004737	ENSP000000379513	ENSPTRP00000008084	ortholog_one2one	Homininae	100	100	0	0,014
ENSG00000023962	ENST00000013034	ENSPTRG00000000945	ENSP000000337060	ENSPTRP00000015997	ortholog_one2one	Homininae	85	99	0,0029	0,0094
ENSG00000013583	ENST00000014930	ENSPTRG00000004712	ENSP00000014930	ENSPTRP00000008050	ortholog_one2one	Homininae	100	100	0	0,0053
ENSG00000014919	ENST00000016171	ENSPTRG00000002835	ENSP00000016171	ENSPTRP00000040337	ortholog_one2one	Homininae	100	100	0	0,0142
ENSG00000022374	ENST00000025399	ENSPTRG00000004734	ENSP00000025399	ENSPTRP00000008081	ortholog_one2one	Eutheria	96	100	0	0,0145
ENSG000000035141	ENST000000037869	ENSPTRG00000012030	ENSP000000037869	ENSPTRP00000020653	ortholog_one2one	Homininae	100	100	0	0,0104
ENSG000000036473	ENST000000039007	ENSPTRG000000021799	ENSP000000039007	ENSPTRP000000037399	ortholog_one2one	Homininae	80	99	0,0058	0,0105
ENSG000000048544	ENST000000053468	ENSPTRG00000018169	ENSP000000053468	ENSPTRP000000031041	ortholog_one2one	Homininae	99	99	0,0066	0,0216
ENSG000000030582	ENST000000053867	ENSPTRG000000009273	ENSP000000053867	ENSPTRP00000015777	ortholog_one2one	Homininae	84	99	0,0027	0,0107
ENSG000000051620	ENST000000058691	ENSPTRG00000018650	ENSP000000058691	ENSPTRP000000031833	ortholog_one2one	Homininae	99	99	0,0047	0,0058
ENSG000000133597	ENST000000072869	ENSPTRG000000019770	ENSP000000072869	ENSPTRP000000033867	ortholog_one2one	Homininae	99	99	0,0061	0,0077
ENSG000000061794	ENST000000081029	ENSPTRG000000004792	ENSP000000081029	ENSPTRP00000008183	ortholog_one2one	Homininae	97	97	0,0148	0,0328
ENSG000000013275	ENST000000157812	ENSPTRG000000010986	ENSP000000157812	ENSPTRP00000018835	ortholog_one2one	Eutheria	100	100	0	0,0137
ENSG000000068615	ENST000000165698	ENSPTRG00000012153	ENSP000000165698	ENSPTRP000000020845	ortholog_one2one	Homininae	100	100	0	0
ENSG000000186847	ENST000000167586	ENSPTRG000000009172	ENSP000000167586	ENSPTRP00000015622	ortholog_one2one	Homininae	99	99	0,0018	0,0369
ENSG000000072506	ENST000000168216	ENSPTRG000000021929	ENSP000000168216	ENSPTRP000000037656	ortholog_one2one	Homininae	100	100	0	0,0091
ENSG000000075336	ENST000000169551	ENSPTRG00000010110	ENSP000000169551	ENSPTRP00000017178	ortholog_one2one	Homininae	99	99	0,0036	0,0113
ENSG000000074071	ENST000000177742	ENSPTRG000000007606	ENSP0000000380531	ENSPTRP00000013000	ortholog_one2one	Homininae	100	100	0	0,0285
ENSG000000065518	ENST000000184266	ENSPTRG00000015270	ENSP000000184266	ENSPTRP000000026324	ortholog_one2one	Homininae	100	100	0	0
ENSG000000081791	ENST000000194118	ENSPTRG00000017354	ENSP0000000396225	ENSPTRP000000029662	ortholog_one2one	Homininae	99	99	0,004	0,0116
ENSG000000083720	ENST000000196371	ENSPTRG00000016829	ENSP000000196371	ENSPTRP000000028847	ortholog_one2one	Homininae	99	99	0,0035	0,0075
ENSG000000086504	ENST000000199706									
ENSG000000197208	ENST000000200652	ENSPTRG00000017215	ENSP000000200652	ENSPTRP000000029458	ortholog_one2one	Homininae	99	99	0,0042	0,0023

Εικόνα 4.23 Εξόρυξη αποτελεσμάτων ορθόλογων γονιδίων ανθρώπου/χιμπατζή και υπολογισμός dN/dS.

Σ' αυτό το σημείο θα ήταν πολύ σημαντικό να αναφερθούμε στις πληροφορίες που συλλέξαμε με τη βοήθεια της Biomart, προκειμένου να αναζητηθούν τα ορθόλογα γονίδια στους πέντε αυτούς διαφορετικούς οργανισμούς σε σχέση με τον άνθρωπο και πιο συγκεκριμένα με τα γονίδια που σχετίζονται με τα μιτοχόνδρια καθώς και με τις λειτουργίες αυτών. Πιο συγκεκριμένα λοιπόν στους παραπάνω πίνακες βλέπουμε τις εξής πληροφορίες: Στη πρώτη στήλη αναφέρεται το Ensembl Gene Id του ανθρώπου και στη δεύτερη στήλη το Ensembl Transcript Id του συγκεκριμένου γονιδίου, το οποίο αποτελεί και το μεγαλύτερο μετάγραφο του γονιδίου. Στη τρίτη, τέταρτη και πέμπτη στήλη γίνεται αναφορά των Ids που αντλήσαμε από τη βάση δεδομένων Ensembl για κάθε ένα οργανισμό και αποτελούν τα αντίστοιχα ορθόλογα γονίδια σε σχέση με τα γονίδια του ανθρώπου. Έτσι λοιπόν παίρνουμε πληροφορίες που αφορούν το Ensembl Gene Id του συγκεκριμένου γονιδίου, το Transcript Id (το οποίο αποτελεί και το μεγαλύτερο μετάγραφο) και το Ensembl Protein Id. Στην έκτη στήλη διαφαίνεται ο τύπος της ομολογίας που υπάρχει ανάμεσα στο συγκεκριμένο ζεύγος γονιδίων. Έτσι μπορεί ένα ζεύγος ορθόλογων γονιδίων να εμφανίζει ομολογία: ortholog one 2 one, ortholog one 2 many, ortholog many 2 many και apparent ortholog one 2 one. Πιο συγκεκριμένα πρέπει να αναφερθεί ότι 1-to-1 orthologues σημαίνει πως μόνο ένα

αντίγραφο του γονιδίου βρίσκεται σε κάθε είδος, 1-to-many orthologues σημαίνει πως σ' ένα από τα δύο είδη το ορθόλογο υπέστη γονιδιακούς διπλασιασμούς (μετά τον διαχωρισμό των δυο υπό μελέτη οργανισμών), ενώ many-to-many orthologues φανερώνει πως τα ορθόλογα γονίδια υπέστησαν γονιδιακούς διπλασιασμούς (μετά τον διαχωρισμό των δυο υπό μελέτη οργανισμών) και στα δύο είδη. Εδώ αξίζει να αναφέρουμε ότι οι ορθόλογες σχέσεις αφορούν ζεύγη ομόλογων γονιδίων όπου ο κοινός τους πρόγονος προήλθε από ειδογένεση και στη συγκεκριμένη περίπτωση ο κοινός αυτός πρόγονος αναφέρεται στην έβδομη στήλη. Στην συνέχεια, στην όγδοη και ένατη στήλη γίνεται αναφορά στην % ομοιότητα σε επίπεδο ακολουθιών όσον αφορά τα ορθόλογα γονίδια. Τέλος στις δύο τελευταίες στήλες υπολογίζονται τα dS και dN των συγκεκριμένων ορθολόγων.

Στην συνέχεια κατεβάσαμε τις ακολουθίες όλων αυτών των οργανισμών από τη Biomart και πήραμε τα μεγαλύτερα μεταγράφα των γονιδίων. Αυτή η διαδικασία έγινε ακριβώς όπως είχαμε πράξει προηγουμένως και για τα ανθρώπινα γονίδια. Ένα μικρό υπόδειγμα δίδεται στην παρακάτω εικόνα (Εικ.4.24)

ENSEMBL GENE ID	ENSEMBL TRANSCRIPT ID	NUMBER OF AMINOACIDS
ENSPTRG00000000027	ENSPTRT00000000057	200
ENSPTRG00000000029	ENSPTRT00000000060	84
ENSPTRG00000000062	ENSPTRT00000000142	458
ENSPTRG00000000084	ENSPTRT00000000186	381
ENSPTRG00000000094	ENSPTRT00000000222	559
ENSPTRG00000000102	ENSPTRT00000000234	190
ENSPTRG00000000106	ENSPTRT00000000243	1567
ENSPTRG00000000115	ENSPTRT00000000260	322
ENSPTRG00000000142	ENSPTRT00000000323	2503
ENSPTRG00000000144	ENSPTRT00000059661	339
ENSPTRG00000000159	ENSPTRT00000000354	758
ENSPTRG00000000202	ENSPTRT00000000426	353
ENSPTRG00000000206	ENSPTRT00000000438	305
ENSPTRG00000000214	ENSPTRT00000059601	264
ENSPTRG00000000242	ENSPTRT00000000501	281
ENSPTRG00000000255	ENSPTRT00000000524	564
ENSPTRG00000000269	ENSPTRT00000000542	79
ENSPTRG00000000288	ENSPTRT00000000569	353
ENSPTRG00000000291	ENSPTRT00000000572	582
ENSPTRG00000000304	ENSPTRT00000000599	1036
ENSPTRG00000000339	ENSPTRT00000000664	326
ENSPTRG00000000350	ENSPTRT00000000692	254

Εικόνα 4.24 Αποτελέσματα των μεγαλύτερων μεταγράφων γονιδίων χιμπατζή και αποθήκευση σε μορφή excel.

Πρωτεϊνικές αλληλεπιδράσεις

Η BioGrid χρησιμοποιεί ως identifiers τα Entrez gene IDs. Επομένως, έπρεπε να αντιστοιχισθούν τα Ensembl gene IDs με τα Entrez gene IDs ώστε να βρεθούν οι πρωτεϊνικές αλληλεπιδράσεις των προϊόντων (πρωτεϊνών) των ανθρώπινων γονιδίων. Αυτή η αντιστοίχιση έγινε με τη βοήθεια της Ensembl BioMart.

Εύρεση γονιδίων που εμπλέκονται σε ασθένειες

Με την βοήθεια της GeneCards, αναζητήσαμε και πήραμε όλα τα γονίδια τα οποία εμπλέκονται σε ασθένειες με σκοπό να εξορύξουμε τα Ensembl Gene Ids τους από τη Biomart. Ανακτήσαμε 2924 Ensembl Gene Ids που εμπλέκονται στην εμφάνιση γενετικών ασθενειών, με βάση την GeneCards.

Όλα τα παραπάνω δεδομένα ενσωματώθηκαν με τη βοήθεια της Perl και οι στατιστικές αναλύσεις έγιναν με τη βοήθεια της γλώσσας προγραμματισμού R σε δύο αρχεία excel (*ensembl_genes_22000* και *Five_species_dNdS*).

Μελέτη πρωτεϊνικών αλληλεπιδράσεων και ασθενειών στα ανθρώπινα γονίδια.

Ο παρακάτω πίνακας (Πίν.4.1) περιέχει στατιστικά στοιχεία για τα ανθρώπινα γονίδια που μελετήσαμε, σχετικά με τον αριθμό αλληλεπιδράσεων (<http://thebiogrid.org/>) και το αν εμπλέκονται σε ασθένειες ή όχι (<http://www.genecards.org/>).ενώ ο επόμενος πίνακας (Πίν.4.2) περιέχει τις στατιστικές αναλύσεις που έχουν γίνει.

ΠΕΡΙΓΡΑΦΗ	ΤΙΜΕΣ
Αριθμός mt_mt_genes	13
Αριθμός mt_genes	1702
Αριθμός nuc_genes	20343
Αριθμός mt_mt_interactors	20
Αριθμός mt_mt_interactors που εμπλέκονται σε ασθένειες (με βάση την	7

Genecards)	
Αριθμός nuc_genes που εμπλέκονται σε ασθένειες (με βάση την GeneCards)	411
Αριθμός mit_genes που εμπλέκονται σε ασθένειες (με βάση την GeneCards)	2468
Μέση τιμή αλληλεπιδράσεων των mt_genes (με βάση την BioGrid)	8.92
Μέση τιμή αλληλεπιδράσεων των nuc_genes (με βάση την BioGrid)	6.23
Μέση τιμή αλληλεπιδράσεων των mt_mt_interactors (με βάση την BioGrid)	498.4
Μέση τιμή αλληλεπιδράσεων (με βάση την BioGrid) των mt_genes που εμπλέκονται σε ασθένειες (με βάση την GeneCards).	13.18
Μέση τιμή αλληλεπιδράσεων (με βάση την BioGrid) των nuc_genes που εμπλέκονται σε ασθένειες (με βάση την GeneCards).	9.48
Μέση τιμή αλληλεπιδράσεων (με βάση την BioGrid) των mt_genes που ΔΕΝ εμπλέκονται σε ασθένειες (με βάση την GeneCards).	7.57
Μέση τιμή αλληλεπιδράσεων (με βάση την BioGrid) των nuc_genes που ΔΕΝ εμπλέκονται σε ασθένειες (με βάση την GeneCards).	5.78

Πίνακας 4.1 Στατιστικά στοιχεία για τα ανθρώπινα γονίδια που μελετήθηκαν, σχετικά με τον αριθμό των αλληλεπιδράσεων (<http://thebiogrid.org/>) και το αν εμπλέκονται σε ασθένειες ή όχι (<http://www.genecards.org/>).

ΠΕΡΙΓΡΑΦΗ	<i>T</i> ΙΜΕΣ
Wilcoxon test για έλεγχο διαφοράς αλληλεπιδράσεων μεταξύ mt_mt_interactors & mt_genes	5.74E-09
Wilcoxon test για έλεγχο διαφοράς αλληλεπιδράσεων μεταξύ mt_mt_interactors & mt_genes που εμπλέκονται σε ασθένειες	3.14E-07
Wilcoxon test για έλεγχο διαφοράς αλληλεπιδράσεων μεταξύ mt_genes που εμπλέκονται σε ασθένειες & mt_genes που ΔΕΝ εμπλέκονται σε ασθένειες	4.70E-07
Wilcoxon test για έλεγχο διαφοράς αλληλεπιδράσεων μεταξύ nuc_genes & nuc_genes	2.20E-16
Wilcoxon test για έλεγχο διαφοράς αλληλεπιδράσεων μεταξύ mt_genes που εμπλέκονται σε ασθένειες & nuc_genes που εμπλέκονται σε ασθένειες	0.1991
Υπεργεωμετρικό τεστ για έλεγχο διαφοράς σε ποσοστό γονιδίων που εμπλέκονται σε ασθένειες (με βάση την GeneCards) μεταξύ nuc_genes & mt_genes	1.88E-47

Πίνακας 4.2 Αποτελέσματα στατιστικών αναλύσεων για τα ανθρώπινα γονίδια που μελετήθηκαν, σχετικά με τον αριθμό αλληλεπιδράσεων (<http://thebiogrid.org/>) και το αν εμπλέκονται σε ασθένειες ή όχι (<http://www.genecards.org/>).

Η μέση τιμή πρωτεϊνικών αλληλεπιδράσεων των mt_genes ήταν 43% υψηλότερη από την αντίστοιχη των nuc_genes (8.9 έναντι 6.2; Wilcoxon test $p < 2.2 \text{ E-}16$). Επιπλέον, οι 20 mt_mt_interactors έχουν 55 φορές υψηλότερη μέση τιμή αλληλεπιδράσεων από τα mt_genes (498 έναντι 8.9; Wilcoxon test $p < 5.7\text{E-}9$). Αυτή η πολύ υψηλή τιμή οφείλεται εν μέρει στην ubiquitin C, η οποία έχει 8580 αλληλεπιδράσεις. Ακόμη όμως και εάν αποκλειστεί η ubiquitin C, τα mt_mt_interactors πάλι εμφανίζουν υψηλή μέση τιμή αλληλεπιδράσεων (73 έναντι 8.9; Wilcoxon test $p < 2.4\text{E-}8$).

Τα mt_genes είναι 2 φορές πιο εμπλουτισμένα σε γονίδια με ασθένειες απ'ότι τα nuc_genes (24% έναντι 12%; hypergeometric test $p < 2E-47$). Τα mt_mt_interactors είναι πιο εμπλουτισμένα σε γονίδια με ασθένειες απ'ότι τα mt_genes (35% έναντι 24%), όμως η διαφορά αυτή δεν εμφανίζεται στατιστικά σημαντική (hypergeometric test $p = 0.11$), ίσως λόγω του μικρού αριθμού των mt_mt_interactors.

Τα γονίδια που εμπλέκονται σε ασθένειες έχουν περισσότερες αλληλεπιδράσεις απ'ότι τα γονίδια που δεν φαίνονται να εμπλέκονται σε ασθένειες (με βάση την GeneCards). Αυτό ισχύει και για την κατηγορία των mt_genes (13.2 έναντι 7.6 αλληλεπιδράσεις; Wilcoxon test $p < 5E-7$) και για την κατηγορία των nuc_genes (9.5 έναντι 5.8 αλληλεπιδράσεις; Wilcoxon test $p < 3E-16$). Όταν όμως συγκρίνουμε τις αλληλεπιδράσεις των γονιδίων που εμπλέκονται σε ασθένειες, μεταξύ των mt_genes & nuc_genes, τότε δεν εμφανίζεται η διαφορά να είναι στατιστικά σημαντική (13.2 έναντι 9.5 αλληλεπιδράσεις; Wilcoxon test $p = 0.19$).

Είναι πιθανόν ότι τα γονίδια που εμπλέκονται σε ασθένειες έχουν μελετηθεί σε μεγαλύτερη έκταση απ'ότι τα γονίδια που δεν φαίνονται να εμπλέκονται σε ασθένειες, γ'αυτό και παρατηρήθηκαν αυτές οι διαφορές στις πρωτεϊνικές αλληλεπιδράσεις, δίχως όμως να υπάρχει μια βιολογική βάση για αυτή τη διαφορά. Προκειμένου να αποκλειστεί αυτή η περίπτωση, θα πρέπει σε μια μελλοντική μελέτη να χρησιμοποιηθούν μόνο αλληλεπιδράσεις από πειράματα μεγάλης κλίμακας (high-throughput) και να αποκλειστούν αλληλεπιδράσεις που προέρχονται από στοχευμένα πειράματα σε ένα γονίδιο (low-throughput).

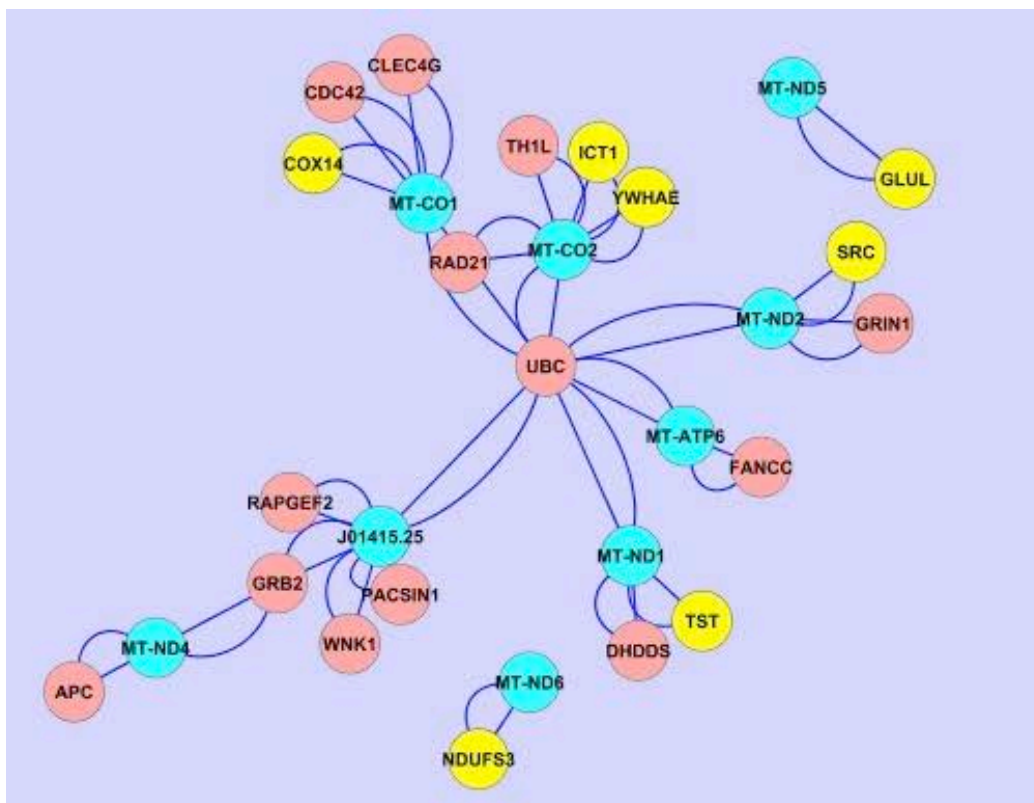
Mt_mt_interactors

Σχετικά με τα mt_mt_interactors, όπως φαίνεται και από τον πίνακα 4.1, μόνο τα 7 από τα 20 αυτά γονίδια έχουν καταγραφεί από την Γονιδιακή Οντολογία να εμπλέκονται στην δομή/λειτουργία του μιτοχονδρίου. Το γεγονός αυτό εγείρει ερωτηματικά είτε σχετικά με την πληρότητα της Γονιδιακής Οντολογίας είτε σχετικά με την ποιότητα των αποτελεσμάτων πρωτεϊνικών αλληλεπιδράσεων που αποθηκεύονται στη BioGrid.

Ensembl Gene Id	Type	HGNC	Entrez Gene ID	Biogrid interactions	Ασθένειες
ENSG00000060237	nuc	WNK1	65125	25	Genecards
ENSG00000070831	nuc	CDC42	998	114	X
ENSG00000101158	nuc	TH1L	51497	32	X
ENSG00000108953	mt	YWHAE	7531	94	X
ENSG00000109756	nuc	RAPGEF2	9693	12	X
ENSG00000117682	nuc	DHDDS	79947	4	X
ENSG00000124507	nuc	PACSIN1	29993	33	X
ENSG00000128311	mt	TST	7263	2	X
ENSG00000134982	nuc	APC	324	113	Genecards
ENSG00000135821	mt	GLUL	2752	14	Genecards
ENSG00000150991	nuc	UBC	7316	8580	X
ENSG00000158169	nuc	FANCC	2176	34	Genecards
ENSG00000164754	nuc	RAD21	5885	219	X
ENSG00000167862	mt	ICT1	3396	206	X
ENSG00000176884	nuc	GRIN1	2902	5	X
ENSG00000177885	nuc	GRB2	2885	267	X
ENSG00000178449	mt	COX14	84987	3	X
ENSG00000182566	nuc	CLEC4G	339390	20	X
ENSG00000197122	mt	SRC	6714	184	Genecards

ENSG00000213619	mt	NDUFS3	4722	7	Genecards
-----------------	----	--------	------	---	-----------

Πίνακας 4.3 Πληροφορίες που αφορούν τους *mt_mt_interactors* (πρωτεΐνες πυρηνικής προέλευσης που αλληλεπιδρούν άμεσα με τις πρωτεΐνες των *mt_mt_genes*).



Εικόνα 4.25 Οπτικοποίηση πρωτεϊνικών αλληλεπιδράσεων και σχηματισμός δικτύων των *mt_mt_interactors* με τη χρήση του Cytoscape. Με το κίτρινο χρώμα διακρίνονται τα *mt_genes*, με το κόκκινο τα *nuc_genes* και με το γαλάζιο τα *mt_mt_genes*.

Εξελικτικές μελέτες

Προκειμένου να πραγματοποιηθούν εξελικτικές μελέτες (δηλαδή ανάλυση dN/dS, % identity) στα υπό διερεύνηση γονίδια, ανακτήθηκαν τα ορθόλογα γονίδια με τη βοήθεια της BioMart από 5 διαφορετικά θηλαστικά, το *Pan troglodytes* (χιμπατζής), το *Oryctolagus cuniculus* (κουνέλι), *Mus musculus* (ποντίκι) *Rattus norvegicus* (αρουραίος), *Canis familiaris* (σκύλος). Για να εξασφαλιστεί η ορθή ανάθεση των ορθόλογων γονιδίων θέσαμε ως κριτήριο ώστε η ανθρώπινη πρωτεΐνη και η πρωτεΐνη από το άλλο είδος να έχουν τουλάχιστον 30% ταύτιση (% identity) στην μεταξύ τους στοίχιση κατά ζεύγη (και προς τις δύο κατευθύνσεις). Επιπλέον, σε αυτές τις εξελικτικές

μελέτες χρησιμοποιήθηκαν τα γονίδια που είχαν $dN/dS \geq 0$ (γονίδια με $dS=0$ αποκλείστηκαν από τις στατιστικές μελέτες).

Λόγω γονιδιακών διπλασιασμών, μετά τον διαχωρισμό δύο ειδών είναι δυνατό να υπάρχει π.χ ένα ορθόλογο στον άνθρωπο και δύο στο ποντίκι (one 2 many orthologs) ή αν οι διπλασιασμοί συνέβησαν και στις δύο εξελικτικές γραμμές (από τότε που διαχωρίστηκαν), να έχουμε π.χ. δύο ορθόλογα στον άνθρωπο και δύο στο ποντίκι (many 2 many orthologs). Επομένως, ο αριθμός των ορθόλογων γονιδίων που βρέθηκαν σε κάθε έναν από τους πέντε οργανισμούς δεν αντικατοπτρίζει αναγκαστικά και τον χρόνο που μεσολάβησε από τον διαχωρισμό του υπό μελέτη είδους και του ανθρώπου.

Στον πίνακα 4.4 φαίνονται τα στατιστικά στοιχεία για τα ορθό λογα των ανθρώπινων γονιδίων που μελετήσαμε στα άλλα πέντε θηλαστικά, σχετικά με το dN/dS και τον βαθμό συντήρησης της ακολουθίας, ενώ στον πίνακα 4.5 φαίνονται οι στατιστικές αναλύσεις που έγιναν.

Περιγραφή	<i>P.troglodytes</i>	<i>O.cuniculus</i>	<i>M.musculus</i>	<i>R.norvegicus</i>	<i>C.familiaris</i>
Αριθμός nuc_genes	14609	14560	18768	16344	15333
Αριθμός mt_genes	1285	1456	1546	1548	1538
Αριθμός nuc_genes που εμπλέκονται σε ασθένειες (με βάση την GeneCards)	1911	1941	2202	2096	2035
Αριθμός nuc_genes που ΔΕΝ εμπλέκονται σε ασθένειες (με βάση την GeneCards)	12698	12619	16566	14248	13298
Αριθμός mt_genes που εμπλέκονται σε ασθένειες (με	327	362	390	389	374

βάση την GeneCards)					
Αριθμός mt_genes που ΔΕΝ εμπλέκονται σε ασθένειες (με βάση την GeneCards)	958	1094	1156	1159	1164
μέση τιμή dS των nuc_genes	0.04425648	0.797797	0.868846	0.8912821	0.590838
μέση τιμή dS των mt_genes	0.03991805	0.636035	0.675411	0.8649115	0.475801
μέση τιμή dN/dS των nuc_genes	0.5245792	0.176193	0.197998	0.1749982	0.178987
μέση τιμή dN/dS των mt_genes	0.4029305	0.160432	0.147219	0.1586179	0.181466
Μέση τιμή %identity των nuc_genes	95.10966	80.15598	74.7728	76.61454	82.25279
Μέση τιμή %identity των mt_genes	95.97043	82.98558	82.91979	80.13178	83.27893
μέση τιμή %identity των mt_mt_interactors	95.5625	86.41176	89	84.88235	83.7619
Μέση τιμή %identity των nuc_genes που εμπλέκονται σε ασθένειες (με βάση την GeneCards)	95.46991	82.93251	82.45095	80.50477	84.55577
Μέση τιμή %identity των nuc_genes που ΔΕΝ εμπλέκονται σε ασθένειες (με	95.05544	79.7289	73.7522	76.04225	81.90036

βάση την GeneCards)					
Μέση τιμή % identity των mt_genes που εμπλέκονται σε ασθένειες (με βάση την GeneCards)	95.28746	84.35912	83.82564	81.10283	85.76203
Μέση τιμή % identity των mt_genes που ΔΕΝ εμπλέκονται σε ασθένειες (με βάση την GeneCards)	96.20355	82.53108	82.61419	79.80587	82.4811
μέση τιμή dN/dS των nuc_genes που εμπλέκονται σε ασθένειες (με βάση την GeneCards)	0.4035858	0.150149	0.146387	0.1463281	0.157533
μέση τιμή dN/dS των nuc_genes που ΔΕΝ εμπλέκονται σε ασθένειες (με βάση την GeneCards)	0.5427882	0.180199	0.204858	0.1792158	0.182271
μέση τιμή dN/dS των mt_genes που εμπλέκονται σε ασθένειες (με βάση την GeneCards)	0.6691584	0.142838	0.131454	0.140323	0.157943
μέση τιμή dN/dS των mt_genes που ΔΕΝ εμπλέκονται σε ασθένειες (με	0.3120573	0.166254	0.152538	0.1647583	0.189024

βάση την GeneCards)					
μέση τιμή dN/dS των mt_mt_interactors	0.1299341	0.144253	0.107891	0.1304985	0.112912
% nuc_genes που εμπλέκονται σε ασθένειες (με βάση την GeneCards)	13.08098	13.33104	11.73274	12.82428	13.27203
% mt_genes που εμπλέκονται σε ασθένειες (με βάση την GeneCards)	25.44747	24.86264	25.22639	25.1292	24.3173

Πίνακας 4.4 Στατιστικά στοιχεία για τα ορθόλογα των ανθρώπινων γονιδίων που μελετήθηκαν στα άλλα πέντε θηλαστικά, σχετικά με το dN/dS, τον βαθμό συντήρησης της ακολουθίας και τις ασθένειες στις οποίες εμπλέκονται (<http://www.genecards.org/>).

Περιγραφή	<i>P.troglodytes</i>	<i>O.cuniculus</i>	<i>M.musculus</i>	<i>R.norvegicus</i>	<i>C.familiaris</i>
Υπεργεωμετρικό τεστ για έλεγχο διαφοράς σε ποσοστό γονιδίων που εμπλέκονται σε ασθένειες (με βάση την GeneCards) μεταξύ nuc_genes & mt_genes	1.19E-36	1.31E-36	3.77E-54	2.81E-44	7.45E-36
Wilcoxon test για έλεγχο διαφοράς του dS	0.5152206	5.00E-06	1.44E-38	2.83E-15	7.31E-08

<i>μεταξύ mt_genes & nuc_genes</i>					
<i>Wilcoxon test για έλεγχο διαφο- ράς του dN/dS μεταξύ mt_genes &nuc_genes</i>	<i>1.75E-09</i>	<i>0.891913</i>	<i>2.34E-13</i>	<i>0.6783917</i>	<i>3.52E-05</i>
<i>Wilcoxon test για έλεγχο διαφοράς του %identity μεταξύ mt_genes & nuc_genes</i>	<i>5.91E-10</i>	<i>0.000111</i>	<i>1.62E-28</i>	<i>6.30E-05</i>	<i>0.9966137</i>
<i>Wilcoxon test για έλεγχο διαφοράς του %identity μεταξύ mt_genes & mt_mt_int eractors</i>	<i>0.6861609</i>	<i>0.147481</i>	<i>0.010947</i>	<i>0.04294347</i>	<i>0.4452345</i>
<i>Wilcoxon test για έλεγχο διαφοράς του %identity μεταξύ nuc_genes &mt_mt_int eractors</i>	<i>0.2981892</i>	<i>0.085419</i>	<i>0.001078</i>	<i>0.0239191</i>	<i>0.4365548</i>
<i>Wilcoxon test για έλεγχο διαφοράς του %identity μεταξύ nuc_genes που εμπλέκονται</i>	<i>0.00518091</i>	<i>1.71E-09</i>	<i>2.25E-51</i>	<i>3.97E-17</i>	<i>2.10E-06</i>

σε ασθένειες και των <i>nuc_genes</i> που ΔΕΝ εμπλέκονται σε ασθένειες(με βάση την GeneCards)					
<i>Wilcoxon test</i> για έλεγχο διαφοράς του %identity μεταξύ <i>mt_genes</i> που εμπλέκονται σε ασθένειες και των <i>mt_genes</i> που ΔΕΝ εμπλέκονται σε ασθένειες(με βάση την GeneCards)	0.09039482	0.021611	0.01829	0.05055916	2.87E-05
<i>Wilcoxon test</i> για έλεγχο διαφοράς του %identity μεταξύ <i>mt_genes</i> που εμπλέκονται σε ασθένειες και των <i>nuc_genes</i> που εμπλέκονται σε ασθένειες(με βάση την GeneCards)	0.5275299	0.203310	0.192651	0.9322514	0.1753727
<i>Wilcoxon test</i> για έλεγχο διαφοράς του <i>dN/dS</i>	7.25E-14	6.95E-07	1.29E-37	1.29E-12	1.89E-06

μεταξύ <i>nuc_genes</i> που εμπλέκονται σε ασθένειες και των <i>nuc_genes</i> που ΔΕΝ εμπλέκονται σε ασθένειες(με βάση την GeneCards					
<i>Wilcoxon</i> <i>test</i> για έλεγχο διαφοράς του <i>dN/dS</i> μεταξύ <i>mt_genes</i> που εμπλέκονται σε ασθένειες και των <i>mt_genes</i> που ΔΕΝ εμπλέκονται σε ασθένειες(με βάση την GeneCards	0.2974982	0.004807	0.005873	0.00050521	0.0002838
<i>Wilcoxon</i> <i>test</i> για έλεγχο διαφοράς του <i>dN/dS</i> μεταξύ <i>nuc_genes</i> που εμπλέκονται σε ασθένειες και των <i>mt_genes</i> που εμπλέκονται σε ασθένειες(με βάση την GeneCards	0.2484463	0.947212	0.752875	0.4975621	0.161978
<i>Wilcoxon</i> <i>test</i> για έλεγχο	0.02248317	0.076372	0.027239	0.08233521	0.0036302

διαφοράς του dN/dS μεταξύ <i>mt_mt_inter</i> <i>actors</i> &<i>mt_genes</i>					
---	--	--	--	--	--

Πίνακας 4.5 Αποτελέσματα στατιστικών αναλύσεων για τα ορθόλογα των ανθρώπινων γονιδίων που μελετήθηκαν στα άλλα πέντε θηλαστικά, σχετικά με το dN/dS , τον βαθμό συντήρησης της ακολουθίας και τις ασθένειες στις οποίες εμπλέκονται (<http://www.genecards.org/>).

Σχετικά με τα γονίδια που υφίστανται θετική επιλογή (θέσαμε ως όριο την τιμή $dN/dS > 1.1$), στην σύγκριση ανθρώπου – χιμπατζή βρέθηκαν 47 *mt_genes* που υφίστανται θετική επιλογή, ωστόσο στις άλλες τέσσερις συγκρίσεις δεν βρέθηκε κανένα γονίδιο της κατηγορίας *mt_genes* που να υφίσταται θετική επιλογή. Επιπλέον, σε καμία από τις 5 συγκρίσεις δεν βρέθηκαν *mt_mt_interactors* που να υφίστανται θετική επιλογή.

Σε τέσσερις από τις πέντε συγκρίσεις το dN/dS των *mt_genes* ήταν μικρότερο από το dN/dS των *nuc_genes* κατά ~17%, ενώ η διαφορά αυτή ήταν στατιστικά σημαντική μόνο στις δύο από τις τέσσερις περιπτώσεις. Επιπλέον, και στις πέντε συγκρίσεις τα *mt_genes* εμφανίζονται να έχουν ~4% υψηλότερο βαθμό συντήρησης (% identity) σε σχέση με τα *nuc_genes*, ενώ η διαφορά αυτή είναι στατιστικά σημαντική για τις τέσσερις συγκρίσεις. Επομένως, από τα υπάρχοντα δεδομένα δεν φαίνεται τα *mt_genes* να βρίσκονται υπό σημαντικά εντονότερη αρνητική επιλογή σε σχέση με τα *nuc_genes*. Μια πιθανή εξήγηση είναι ότι ίσως πολλά από τα *mt_genes* να εμπλέκονται και σε άλλες λειτουργίες, άσχετες με τα μιτοχόνδρια. Σε μια μελλοντική μελέτη θα πρέπει να χρησιμοποιηθούν ως *mt_genes* μόνο εκείνα των οποίων οι πρωτεΐνες έχουν εντοπιστεί αποκλειστικά στο μιτοχονδριακό διαμέρισμα.

Από τη σύγκριση μεταξύ *mt_mt_interactors* & *mt_genes*, όσον αφορά το dN/dS , αυτό εμφανίζεται ~32% μικρότερο στα *mt_mt_interactors* (και στις πέντε συγκρίσεις), ωστόσο η διαφορά αυτή είναι στατιστικά σημαντική στις τρεις από τις πέντε συγκρίσεις.

Όσον αφορά το % identity, αυτό εμφανίζεται ελαφρά πιο αυξημένο στους *mt_mt_interactors* (κατά ~4.5%) στις τέσσερις από τις πέντε συγκρίσεις, ωστόσο η διαφορά είναι στατιστικά σημαντική μόνο στις δύο συγκρίσεις.

Τα γονίδια που εμπλέκονται σε ασθένειες έχουν ελαφρά χαμηλότερο dN/dS σε σχέση με τα γονίδια που δεν εμπλέκονται σε ασθένειες. Αυτό ισχύει για την κατηγορία των mt_genes (14% χαμηλότερο; στατιστικά σημαντική διαφορά στις τέσσερις συγκρίσεις) και για την κατηγορία των nuc_genes (20% χαμηλότερο; στατιστικά σημαντική διαφορά και στις πέντε συγκρίσεις).

Όμως, δεν υπάρχει στατιστικά σημαντική διαφορά στο dN/dS μεταξύ mt_genes & nuc_genes που εμπλέκονται σε ασθένειες, σε καμία από τις πέντε συγκρίσεις

Τα γονίδια που εμπλέκονται σε ασθένειες έχουν ελαφρά υψηλότερο βαθμό συντήρησης (% identity) σε σχέση με τα γονίδια που δεν εμπλέκονται σε ασθένειες. Αυτό ισχύει για την κατηγορία των mt_genes (κατά 1% υψηλότερο; στατιστικά σημαντική διαφορά στις τέσσερις συγκρίσεις) και για την κατηγορία των nuc_genes (κατά 5% υψηλότερο; στατιστικά σημαντική διαφορά και στις πέντε συγκρίσεις). Όμως δεν υπάρχει στατιστικά σημαντική διαφορά στο βαθμό συντήρησης (% identity) μεταξύ mt_genes & nuc_genes που εμπλέκονται σε ασθένειες, σε καμία από τις πέντε συγκρίσεις.

Συμπεράσματα & μελλοντικές αναλύσεις

Στον άνθρωπο, τα προϊόντα των γονιδίων (τα ονομάζουμε *mt_genes*) πυρηνικής προέλευσης που εμπλέκονται στην δομή/λειτουργία του μιτοχονδρίου έχουν περισσότερες πρωτεϊνικές αλληλεπιδράσεις και εμπλέκονται πιο συχνά σε ασθένειες απ'ότι τα προϊόντα των γονιδίων (τα ονομάζουμε *nuc_genes*) που δεν έχουν καμία σχέση με το μιτοχόνδριο. Μάλιστα, αυτές οι παρατηρήσεις είναι πιο έντονες για τις πρωτεΐνες που αλληλεπιδρούν άμεσα (*mt_mt_interactors*) με κάποια από τις 13 πρωτεΐνες που κωδικοποιούνται στο mt-DNA (τα γονιδιά τους τα ονομάζουμε *mt_mt_genes*).

Όσον αφορά την εξέλιξη των *mt_genes* (*dN/dS*, βαθμό συντήρησης της ακολουθίας), δεν εμφανίζουν πολύ αξιοσημείωτες διαφορές ως προς τα υπόλοιπα γονίδια. Τα *mt_genes* έχουν κατά μέσο όρο ~17% μικρότερο *dN/dS* και ~4% υψηλότερο βαθμό συντήρησης της ακολουθίας (% identity) σε σχέση με τα *nuc_genes*. Μία πιθανή εξήγηση είναι ότι ίσως πολλά από τα *mt_genes* (όπως τα ορίσαμε) να εμπλέκονται και σε πολλές άλλες λειτουργίες. Σε μια μελλοντική μελέτη θα πρέπει να χρησιμοποιηθούν ως *mt_genes* μόνο εκείνα των οποίων οι πρωτεΐνες έχουν εντοπιστεί αποκλειστικά στο μιτοχονδριακό διαμέρισμα, με μεθόδους πρωτεωμικής.

Τα *mt_mt_interactors* έχουν κατά μέσο όρο ~32% μικρότερο *dN/dS* και ~4.5% υψηλότερο βαθμό συντήρησης της ακολουθίας (% identity) σε σχέση με τα *mt_genes*, όμως οι διαφορές αυτές δεν είναι πάντα στατιστικά σημαντικές στις πέντες συγκρίσεις που έγιναν (άνθρωπος – χιμπατζής, άνθρωπος – κουνέλι, άνθρωπος – ποντίκι, άνθρωπος – αρουραίος, άνθρωπος – σκύλος).

Τα γονίδια που εμπλέκονται σε ασθένειες (με βάση την GeneCards) έχουν γενικά περισσότερες αλληλεπιδράσεις (με βάση την BioGrid) απ'ότι τα γονίδια που δεν φαίνονται να εμπλέκονται σε ασθένειες. Αυτό ισχύει και για την κατηγορία των *mt_genes* και για την κατηγορία των *nuc_genes*. Όταν όμως συγκρίνουμε τις αλληλεπιδράσεις των γονιδίων που εμπλέκονται σε ασθένειες, μεταξύ των *mt_genes* & *nuc_genes*, τότε δεν εμφανίζεται η διαφορά να είναι στατιστικά σημαντική. Είναι πιθανό ότι τα γονίδια που εμπλέκονται σε ασθένειες έχουν μελετηθεί σε μεγαλύτερη έκταση απ'ότι τα γονίδια που δεν φαίνονται να εμπλέκονται σε ασθένειες γι'αυτό και παρατηρήθηκαν αυτές οι διαφορές στις πρωτεϊνικές αλληλεπιδράσεις, δίχως όμως να

υπάρχει μια βιολογική βάση για αυτή τη διαφορά. Προκειμένου να αποκλειστεί αυτή η περίπτωση, θα πρέπει να χρησιμοποιηθούν μόνο αλληλεπιδράσεις από πειράματα μεγάλης κλίμακας (high-throughput) και να αποκλειστούν αλληλεπιδράσεις που προέρχονται από στοχευμένα πειράματα σε ένα γονίδιο (low-throughput).

Τα γονίδια που εμπλέκονται σε ασθένειες έχουν χαμηλότερο dN/dS σε σχέση με τα γονίδια που δεν εμπλέκονται σε ασθένειες. Αυτό ισχύει για την κατηγορία των mt_genes (~14% χαμηλότερο) και για την κατηγορία των nuc_genes (~20% χαμηλότερο). Όμως, δεν υπάρχει στατιστικά σημαντική διαφορά στο dN/dS μεταξύ mt_genes & nuc_genes που εμπλέκονται σε ασθένειες, σε καμία από τις πέντε συγκρίσεις.

Με βάση τα αυστηρά κριτήρια που θέσαμε για την ανάθεση ορθόλογων ακολουθιών, το πως ορίστηκαν τα μιτοχονδριακά γονίδια πυρηνικής προέλευσης αλλά και το πως υπολογίστηκε το dN/dS από την Ensembl, δεν παρατηρήθηκε θετική επιλογή (positive selection) στα mt_genes (με εξαίρεση την σύγκριση ανθρώπου - χιμπατζή). Εάν σε ένα γονίδιο ένα μέρος της ακολουθίας του υφίσταται θετική επιλογή (dN/dS > 1) και ένα άλλο μέρος του υφίσταται αρνητική επιλογή (dN/dS < 1), είναι δυνατόν το γονίδιο συνολικά να εμφανίζει dN/dS ~ 1 και να μην εντοπιστεί η θετική επιλογή. Επομένως, άλλες εναλλακτικές μέθοδοι υπολογισμού του dN/dS για κάθε κωδικόνιο ξεχωριστά θα πρέπει να επιλεγούν σε μια μελλοντική ανάλυση (π.χ. από μία πολλαπλή στοίχιση με αρκετές ορθόλογες ακολουθίες).

Οι πρωτεϊνικές αλληλεπιδράσεις ανέδειξαν ένα αριθμό γονιδίων των οποίων τα προϊόντα αλληλεπιδρούν άμεσα με κάποια από τις 13 πρωτεΐνες που κωδικοποιούνται στο mt-DNA, όμως δεν χαρακτηρίζονται ως μιτοχονδριακά εμπλεκόμενα στην Γονιδιακή Οντολογία. Αυτό εγείρει ερωτηματικά είτε ως προς την πληρότητα της Γονιδιακής Οντολογίας είτε ως προς την ποιότητα των δεδομένων πρωτεϊνικών αλληλεπιδράσεων. Σε μια μελλοντική μελέτη θα πρέπει να χρησιμοποιηθούν εκείνες οι αλληλεπιδράσεις που έχουν ανιχνευθεί περισσότερες από μια φορές, δεδομένου ότι κάποιες τεχνολογίες εμφανίζουν υψηλό αριθμό ψευδώς θετικών αποτελεσμάτων (π.χ. yeast two hybrid).

ΒΙΒΛΙΟΓΡΑΦΙΑ

- *Alberts B, Bray D, Johnson A, Lewis J, Raff M, Walter P. Βασικές Αρχές Κυτταρικής Βιολογίας, Εισαγωγή στη Μοριακή Βιολογία του Κυττάρου. Ιατρικές Εκδόσεις Πασχαλίδη. 2000.*
- *Amoutzias, G and Van de Peer, Y. Single-Gene and Whole-Genome Duplications and the Evolution of Protein-Protein Interaction Networks. In: Evolutionary genomics and systems biology. 2010; pg. 413-429.*
- *Baloyannis SJ. Mitochondrial alterations in Alzheimer's disease. J Alzheimers Dis. 2006 Jul;9(2):119-26.*
- *Baxevanis AD, Ouellette BFF (Επιμέλεια Ελληνικής Έκδοσης: Μουδριανάκης Ε.Ν., Χαμόδρακας Σ.Ι.). Βιοπληροφορική, Ένας Πρακτικός Οδηγός για την Ανάλυση Γονιδίων και Πρωτεϊνών. Επιστημονικές Εκδόσεις Παρισιάνου. 2004.*
- *Βήχου Α, Βαιτσοπούλου Χ. Ανίχνευση μεταλλάξεων σε μιτοχονδριακά tRNA γονίδια σε ασθενείς με διαβήτη (διπλωματική εργασία). 2011.*
- *Brinckmann A, Weiss C, Wilbert F, von Moers A, Zwirner A, Stoltenburg-Didinger G, Wilichowski E, Schuelke M. Regionalized pathology correlates with augmentation of mtDNA copy numbers in a patient with myoclonic epilepsy with ragged-red fibers (MERRF-syndrome). PLoS One. 2010 Oct 20;5(10):e13513.*
- *Brown KR and Jurisica I. Unequal evolutionary conservation of human protein interactions in interlogous networks. Genome Biol. 2007, 8:R95.*
- *Chial, H. & Craig, J. mtDNA and mitochondrial diseases. Nature Education 2008 1(1).*
- *Chinnery PF, Howell N, Andrews RM, Turnbull DM. Clinical mitochondrial genetics. J Med Genet. 1999 Jun;36(6):425-36.*
- *Chinnery PF, Schon EA. Mitochondria. J Neurol Neurosurg Psychiatry. 2003 Sep;74(9):1188-99.*
- *Enns GM, Kinsman SL, Perlman SL, Spicer KM, Abdenur JE, Cohen BH, Amagata A, Barnes A, Kheifets V, Shrader WD, Thoolen M, Blankenberg F, Miller G. Initial experience in the treatment of inherited mitochondrial disease with EPI-743. Mol Genet Metab. 2012 Jan;105(1):91-102.*

- Finsterer J. Parkinson syndrome as a manifestation of mitochondriopathy. *Acta Neurol Scand.* 2002 May;105(5):384-9.
- Fruhman G, Landsverk ML, Lotze TE, Hunter JV, Wangler MF, Adesina AM, Wong LJ, Scaglia F. Atypical presentation of Leigh Syndrome associated with a Leber hereditary optic neuropathy primary mitochondrial DNA. *Mol Genet Metab.* 2011 Jun;103(2):153-60.
- Futuyma DJ. Εξελικτική βιολογία (Επιμέλεια: Λευτέρης Ζούρος). Πανεπιστημιακές Εκδόσεις Κρήτης. 1995.
- Hu P, Bader G, Wigle DA, Emili A. Computational prediction of cancer gene function. *Nat Rev Cancer.* 2007 Jan;7(1):23-34.
- Hurst LD. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet.* 2002 Sep; 18(9):486.
- Κορώνας Ε, Σαλγκάμης Β. Μελέτη πολυμορφισμών μιτοχονδριακού DNA στο φυσιολογικό πληθυσμό (διπλωματική εργασία). 2011.
- Κοσσίδα Σ. Βιοπληροφορική – Δυνατότητες και Προοπτικές. Εκδόσεις: Ίδρυμα Ιατροβιολογικών Ερευνών της Ακαδημίας Αθηνών. 2008.
- Kryazhimskiy S, Plotkin JB. The Population Genetics of dN/dS . *PLoS Genet.* 2008 Dec;4(12):e1000304.
- Ladoukakis ED, Saavedra C, Magoulas A, Zouros E. Mitochondrial DNA variation in a species with two mitochondrial genomes: the case of *Mytilus galloprovincialis* from the Atlantic, the Mediterranean and the Black Sea. *Mol Ecol.* 2002 Apr;11(4):755-69.
- Lee YS, Kim WY, Ji M, Kim JH, Bhak J. MitoVariome: a variome database of human mitochondrial DNA. *BMC Genomics.* 2009 Dec 3;10 Suppl 3:S12.
- Μαργαρίτης Α.Χ, Γαλανόπουλος Β, Κεραμάρης Κ, Μαρίνος Ε, Παπασιδέρη Ι, Τρουγκάκος Π, Στραβοπόδης Δ. Βιολογία του Κυ τάρου, Τέταρτη Έκδοση. Εκδόσεις Λίτσας. 2004.
- Martin W. & Mentel, M. The origin of mitochondria. *Nature Education* 2010 3(9):58.
- Martin W, Müller M. The hydrogen hypothesis for the first eukaryote. *Nature.* 1998 Mar 5;392(6671):37-41.
- McCutcheon JP, Moran NA. Extreme genome reduction in symbiotic bacteria. *Nat Rev Microbiol.* 2011 Nov 8;10(1):13-26.

- Nabholz B, Glémin S, Galtier N. *Strong Variations of Mitochondrial Mutation Rate across Mammals – the Longevity Hypothesis*. *Mol Biol Evol*. 2008 Jan;25(1):120-30.
- Nooren IM, Thornton JM. *Diversity of protein-protein interactions*. *EMBO J*. 2003 Jul 15;22(14):3486-92.
- Omoto CK, Lurquin PF. *Genes and DNA: A Beginner's Guide to Genetics and its Applications*. Columbia University Press. 2004.
- Pieczenik SR, Neustadt J. *Mitochondrial dysfunction and molecular pathways of disease*. *Exp Mol Pathol*. 2007 Aug;83(1):84-92.
- Pierce C. *Οδηγός της Perl*, Εκδόσεις: Μ. Γκιούρδας. 2005.
- Pruess M, Kersey P, Apweiler R. *The Intergr8 Project – A Resource for Genomic and Proteomic Data*. *In Silico Biol*. 2005;5(2):179-85.
- Raff RA, Mahler HR. *The non symbiotic origin of mitochondria*. *Science*. 1972 Aug 18;177(4049):575-82.
- Schapira AH. *Mitochondrial disease*. *Lancet*. 2012 May 12;379(9828):1825-34.
- Sherer TB, Betarbet R, Greenamyre JT. *Environment, mitochondria and Parkinson's disease*. *Neuroscientist*. 2002 Jun;8(3):192-7.
- Springer MS, Murphy WJ, Eizirik E, O'Brien SJ. *Placental mammal diversification and the Cretaceous – Tertiary boundary*. *Proc Natl Acad Sci U S A*. 2003 Feb 4;100(3):1056-61.
- Spudich GM, Fernández-Suárez XM. *Touring Ensembl: A practical guide to Genome browsing*. *BMC Genomics*. 2010 May 11;11:295.
- Spudich G, Fernández-Suárez XM, Birney E. *Genome Browsing with Ensembl: A practical Overview*. *Brief Funct Genomic Proteomic*. 2007 Sep;6(3):202-19.
- Stark C, Breitkreutz BJ, Chatr-Aryamontri A, Boucher L, Oughtred R, Livstone MS, Nixon J, Van Auken K, Wang X, Shi X, Reguly T, Rust JM, Winter A, Dolinski K, Tyers M. *The BioGRID Interaction Database*. *Nucleic Acids Res*. 2011 Jan;39 (Database issue):D698-704.
- Stewart PA, Hayakawa K, Akers MA, Vinters HV. *A morphometric study of the blood-brain barrier in Alzheimer's disease*. *Lab Invest*. 1992 Dec;67(6):734-42.
- Taylor RW, Turnbull DM. *Mitochondrial DNA mutations in human disease*. *Nat Rev Genet*. 2005 May;6(5):389-402.

- *Turnbull HE, Lax NZ, Diodato D, Ansorge O, Turnbull DM. The mitochondrial brain: From mitochondrial genome to neurodegeneration. Biochim Biophys Acta. 2010 Jan;1802(1):111-21.*
- *Weiner WJ and Factor SA. Parkinsons Disease: Diagnosis and clinical Management (Second Edition). Demos Medical Publishing. 2002.*
- *Θωμόπουλος Γ. Βιολογία του κυττάρου. University Studio Press. 1993.*
- *Wang Z, Qi XK, Yao S, Chen B, Luan X, Zhang W, Han M, Yuan Y. Phenotypic patterns of MELAS/LS overlap syndrome associated with m.13513G>A mutation, and neuropathological findings in one autopsy case. Neuropathology. 2010 Dec;30(6):606-14.*
- *Zeviani M, Di Donato S. Mitochondrial disorders 2. Brain. 2004 Oct;127(Pt 10):2153-72.*
- *Zifa E, Giannouli S, Theotokis P, Stamatis C, Mamuris Z, Stathopoulos C. Mitochondrial tRNA mutations: clinical and functional perturbations. RNA Biol. 2007 Jan-Mar;4(1):38-66.*
- *Zifa E, Theotokis P, Kaminari A, Maridaki H, Leze H, Petsiava E, Mamuris Z, Stathopoulos C. A novel G3337A mitochondrial NDI mutation related to cardiomyopathy co-segregates with tRNA^{Leu(CUN)} A12308G and Trna^{Thr} C 15946T mutations. Mitochondrion. 2008 Jun;8(3):229-36.*
- *Ζίφα Α, Μαμούρης Ζ, Μούτου Κ . Βιολογία. Πανεπιστημιακές Εκδόσεις Θεσσαλίας. 2010.*