

*ΣΥΓΚΡΙΣΗ ΤΩΝ ΕΞΕΛΙΚΤΙΚΩΝ ΡΥΘΜΩΝ ΤΩΝ ΓΟΝΙΔΙΩΝ ΠΟΥ ΚΩΔΙΚΟΠΟΙΟΥΝΤΑΙ ΑΠΟ ΜΙΤΟΧΟΝΔΡΙΑΚΟ ΚΑΙ ΠΥΡΗΝΙΚΟ DNA ΚΑΙ ΠΟΥ ΤΑ ΠΡΟΪΟΝΤΑ ΤΟΥΣ ΣΥΝΘΕΤΟΥΝ ΤΑ ΣΥΜΠΛΟΚΑ ΤΗΣ ΟΞΕΙΔΩΤΙΚΗΣ ΦΩΣΦΟΡΥΛΙΩΣΗΣ.*

*ΔΙΕΡΕΥΝΗΣΗ ΤΗΣ ΥΠΑΡΞΗΣ ΘΕΤΙΚΗΣ ΕΠΙΛΟΓΗΣ ΣΕ ΓΟΝΙΔΙΑ ΤΩΝ ΣΥΜΠΛΟΚΩΝ ΤΗΣ ΟΞΕΙΔΩΤΙΚΗΣ ΦΩΣΦΟΡΥΛΙΩΣΗΣ ΣΕ ΕΞΕΛΙΚΤΙΚΕΣ ΓΡΑΜΜΕΣ ΤΗΣ ΟΙΚΟΓΕΝΕΙΑΣ ΔΡΟΣΟΦΙΛΙΔΩΝ*

Χρήστος Ζουμαδάκης  
Μεταπτυχιακό Πρόγραμμα Μοριακής Βιολογίας – Βιοιατρικής

Μεταπτυχιακή Εργασία  
Οκτώβριος 2009 – Σεπτέμβριος 2010

Επιβλέπων καθηγητής:  
Μανόλης Λαδουκάκης



Πανεπιστήμιο Κρήτης  
Τμήμα Βιολογίας

*EVOLUTIONARY BEHAVIOUR ANALYSIS OF THE PROTEINS THAT PARTICIPATE IN THE OXIDATIVE PHOSPHORYLATION COMPLEXES, AND ARE ASSEMBLED BY EITHER NUCLEAR OR MITOCHONDRIAL DNA. A SCAN FOR POSITIVE SELECTION IN THE OXIDATIVE PHOSPHORYLATION COMPLEXES OF TEN DROSOPHILIDAE SPECIES*

Chris Zoumadakis  
Master's Programme in Molecular Biology - Biomedicine

Master Thesis  
October 2009 – September 2010

Supervisor:  
Manolis Ladoukakis



Biology Department  
University of Crete

**Acknowledgements:**

*I blame society*

## ΠΕΡΙΕΧΟΜΕΝΑ

### Περίληψη/Abstract

#### **Κεφάλαιο 1 – Μιτοχόνδρια και οξειδωτική φωσφορυλίωση**

- 1.1. Μιτοχόνδρια – Δομή και πρόλευση
- 1.2. Αερόβια αναπνοή και άλλες μιτοχονδριακές λειτουργίες
- 1.3. Μιτοχονδριακό DNA
- 1.4. Πυρηνικές πρωτεΐνες που στοχεύονται στο μιτοχόνδριο
- 1.5. Σύμπλοκα οξειδωτικής φωσφορυλίωσης

#### **Κεφάλαιο 2 – Σύγκριση των εξελικτικών ρυθμών των γονιδίων που κωδικοποιούνται από μιτοχονδριακό και πυρηνικό DNA, και που τα προϊόντα τους συνθέτουν τα σύμπλοκα της οξειδωτικής φωσφορυλίωσης**

- 2.1. Εισαγωγή
- 2.2. Σκοπός της μελέτης
- 2.3. Υλικά και μέθοδοι
- 2.4. Αποτελέσματα - Συζήτηση

#### **Κεφάλαιο 3 – Διερεύνηση της ύπαρξης θετικής επιλογής σε γονίδια των συμπλόκων της οξειδωτικής φωσφορυλίωσης σε εξελικτικές γραμμές της οικογένειας δροσοφιλιδών**

- 3.1. Εισαγωγή
- 3.2. Σκοπός της μελέτης
- 3.3. Υλικά και μέθοδοι
- 3.4. Αποτελέσματα
- 3.5. Συζήτηση

#### **Κεφάλαιο 4 – Παράρτημα, κώδικας**

### **Βιβλιογραφία**

## Περίληψη

Σκοποί της παρούσας εργασίας, ήταν:

α) η μελέτη του εξελικτικού προτύπου των πρωτεϊνών που συμμετέχουν στα σύμπλοκα της οξειδωτικής φωσφορυλίωσης, και που είτε κατασκευάζονται στο μιτοχόνδριο από μιτοχονδριακό DNA, είτε κατασκευάζονται στο κυτταρόπλασμα από πυρηνικό DNA και κατόπιν στοχεύονται στο μιτοχόνδριο. Επίσης, εξετάστηκε εάν συγκεκριμένα χαρακτηριστικά των πρωτεϊνών αυτών, όπως το μέγεθος, ο αριθμός των αλληλεπιδράσεων ή η ένταση της επιλογής που ασκείται πάνω τους, έχουν στατιστικά σημαντικές συσχετίσεις μεταξύ τους. Τα αποτελέσματα δείχνουν πως οι πυρηνικές πρωτεΐνες που αλληλεπιδρούν με μιτοχονδριακές στα σύμπλοκα της αναπνοής, έχουν κατά γενικό κανόνα παρόμοιους εξελικτικούς ρυθμούς με τις υπόλοιπες πυρηνικές πρωτεΐνες. Οι εξελικτικοί αυτοί ρυθμοί στα ζώα είναι ταχύτεροι από αυτούς των αντίστοιχων μιτοχονδριακών πρωτεϊνών ενώ στα φυτά είναι αργότεροι. Ωστόσο για να μπορούν οι πρωτεΐνες που προέρχονται από διαφορετικά γενετικά υπόβαθρα (μιτοχονδριακό και πυρηνικό) να συνεργαστούν για να φτιάξουν τα σύμπλοκα της αναπνοής σημαντικό ρόλο φαίνεται να παίζει η φυσική επιλογή η οποία στις πρωτεΐνες από μιτοχονδριακό υπόβαθρο είναι αυστηρότερη. Αντίθετο πρότυπο παρατηρείται στα φυτά. Αθροιστικά, τόσο στις μιτοχονδριακές, όσο και στις πυρηνικές πρωτεΐνες που αλληλεπιδρούν με μιτοχονδριακές, βρέθηκε θετική συσχέτιση ανάμεσα στο μήκος και στον αριθμό των αλληλεπιδράσεών τους, ενώ κάποια από τα σύμπλοκα οξ. φωσφορυλίωσης, βρέθηκε συσχέτιση ανάμεσα στην ένταση της επιλογής και είτε στο μήκος, είτε στον αριθμό των αλληλεπιδράσεων των πρωτεϊνών που συμμετέχουν σε αυτά.

β) η εφαρμογή διαφόρων τεχνικών μέγιστης πιθανοφάνειας ώστε να δοθεί μια ολοκληρωμένη εικόνα της ύπαρξης ή όχι θετικής επιλογής για όλα τα γονίδια που απαρτίζουν τα σύμπλοκα οξειδωτικής φωσφορυλίωσης (είτε αυτά κωδικοποιούνται από τον πυρήνα είτε κωδικοποιούνται από το μιτοχόνδριο) σε δέκα είδη της οικογένειας των δροσοφιλιδών, καθώς και η σύγκριση των ρυθμών εξέλιξης των πρωτεϊνών αυτών σε όλους τους κόμβους του φυλογενετικού δέντρου των ειδών αυτών. Τα αποτελέσματα δείχνουν την ύπαρξη διαφορών στην ένταση της επιλογής σε συγκεκριμένα γονίδια και συγκεκριμένους κόμβους του δέντρου. Επίσης, εντοπίστηκαν ενδείξεις θετικής επιλογής σε αρκετά γονίδια, οι περισσότερες από τις οποίες εντοπίζονται στα είδη *D.pseudoobscura* και *D.persimilis*, μετά τον διαχωρισμό τους από τον κοινό τους πρόγονο.

## Abstract

The purpose of this study was:

a) the analysis of the evolutionary behaviour of the proteins that participate in the oxidative phosphorylation pathway, and which are either assembled in the mitochondrion by mitochondrial DNA, or are assembled in the cell cytoplasm by nuclear DNA, and subsequently targeted to the mitochondrion. Also, it was researched whether certain characteristics of these proteins, such as their size, their number of protein-protein interactions, or the selection pressure that acts on them, have statistically significant correlations between each other. Results show that nuclear proteins that interact with mitochondrial ones in the respiratory pathway complexes, have similar evolutionary rates with the rest of the nuclear DNA proteins. These evolutionary rates are faster in animals than the respective mitochondrial rates, but slower in plants. However in order for the proteins from different genetic backgrounds (mitochondrial and nuclear) to co-interact in order to build the respiratory pathway complexes, natural selection plays an important role which seems to be stronger in mitochondrial proteins. Again, the opposite pattern can be observed in plants. In both types, positive correlation was found between their size and the number of protein interactions. In specific protein complexes, correlation between selection pressure, and either length, or number of interaction partners was shown.

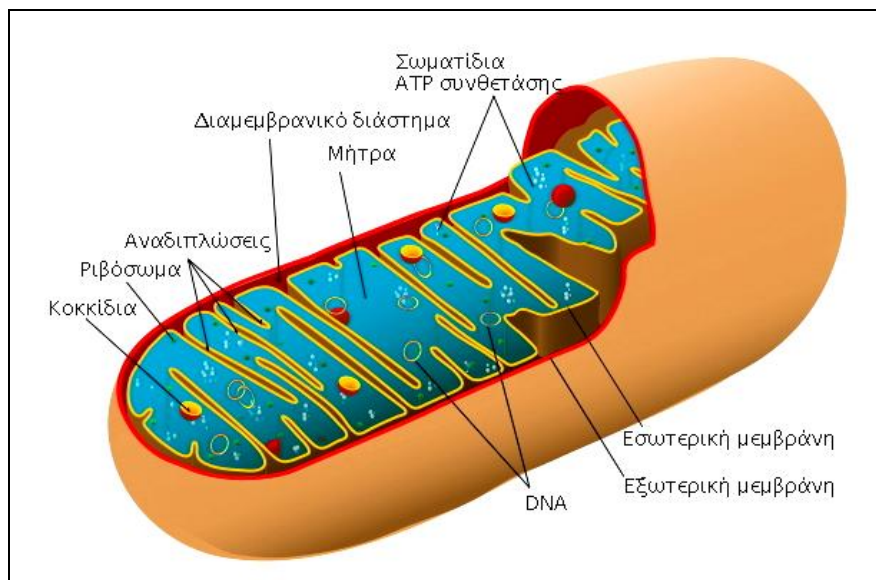
b) the application of specific maximum likelihood techniques, in order to quantify the selection pressure that acts in the oxidative phosphorylation genes of ten Drosophilidae species, as well as to compare these rates, among all the nodes of their phylogenetic tree. The results show that a number of genes have different selection pressures acting on them, in many of the tree nodes. Moreover, traces of positive selection were identified in a number of genes, with most of them detected in the *D.pseudoobscura* and *D.persimilis* species, after their divergence from their common ancestor.

## 1. ΕΙΣΑΓΩΓΗ

### 1.1. Μιτοχόνδρια – Δομή και προέλευση

Τα **μιτοχόνδρια** είναι ενδοκυτταρικά οργάνδια τα οποία απαντώνται στα περισσότερα ευκαρυωτικά κύτταρα. Η κύρια λειτουργία τους είναι η παραγωγή ενέργειας για τις ανάγκες του κυττάρου, μέσω του μονοπατιού της οξειδωτικής φωσφορυλίωσης<sup>1</sup>. Ο αριθμός των μιτοχονδρίων διαφέρει από κύτταρο σε κύτταρο, και μπορεί να κυμαίνεται από λίγα μέχρι αρκετές χιλιάδες. Τα μιτοχόνδρια διαθέτουν το δικό τους γονιδίωμα, το οποίο ακολουθεί διαφορετικό γενετικό κώδικα από το αντίστοιχο πυρηνικό.

Τα μιτοχόνδρια είναι εύκαμπτα σωματίδια, με μέγεθος που κυμαίνεται συνήθως από 0.5μm μέχρι 1μm, και σχήμα ράβδου ή περίπου σφαιρικό. Περιβάλλονται από δυο φωσφολιπιδικές μεμβράνες, την εξωτερική και την εσωτερική<sup>2</sup> (Εικόνα 1).



Εικόνα 1: Δομή του μιτοχονδρίου

Η **εξωτερική μεμβράνη**, η οποία περικλείει ολόκληρο το μιτοχόνδριο, έχει παραπλήσια δομή με άλλες διπλοστιβάδες του κυττάρου, καθώς και περίπου ίδια αναλογία φωσφολιπιδίων / πρωτεϊνών με την πλασματική μεμβράνη. Το κυριότερο χαρακτηριστικό της εξωτερικής μεμβράνης, είναι η παρουσία μιας οικογένειας διαμεμβρανικών πρωτεϊνών, οι οποίες ονομάζονται πορίνες. Η τεταρτοταγής δομή αυτών των πρωτεϊνών, τους επιτρέπει να σχηματίζουν κανάλια, μέσω των οποίων μπορούν να διαχυθούν μόρια μικρότερα των 5000Da, είτε προς το εσωτερικό είτε προς το εξωτερικό του μιτοχονδρίου<sup>3</sup>. Μεγαλύτερες πρωτεΐνες μπορούν μόνο να εισέλθουν, εφόσον διαθέτουν το σωστό πεπτιδίο στόχευσης, το οποίο αναγνωρίζεται από το πρωτεϊνικό σύμπλοκο TOM (Translocase of the Outer Membrane – Τρανσλοκάση της Εξωτερικής Μεμβράνης)<sup>4</sup>.

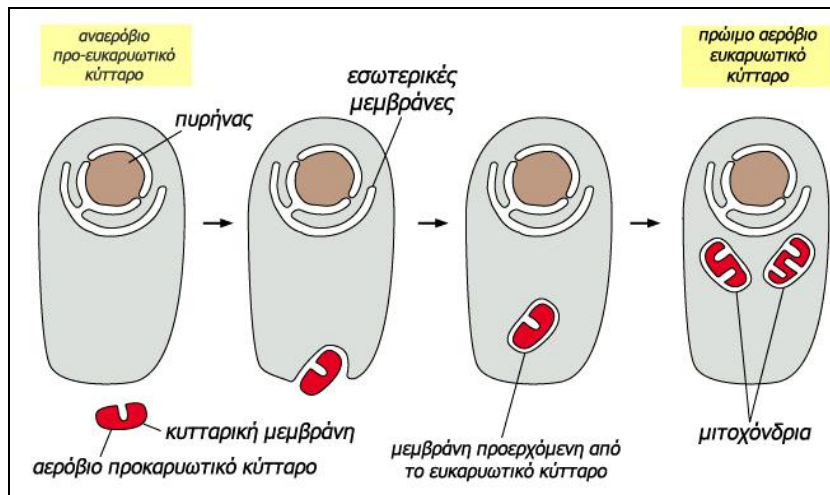
Το **διαμεμβρανικό διάστημα**, το ένα από τα δυο μιτοχονδριακά διαμερίσματα, βρίσκεται ανάμεσα στην εξωτερική και στην εσωτερική μεμβράνη. Λόγω της παρουσίας των πορινών στην εξωτερική μεμβράνη, το διαμεμβρανικό διάστημα περιέχει μικρά μόρια σε παρόμοια περίπου συγκέντρωση με αυτή του κυτταροπλάσματος.

Η **εσωτερική μεμβράνη** βρίσκεται στο εσωτερικό μέρος του διαμεμβρανικού διαστήματος, και περιέχει διαφορετική αναλογία φωσφολιπιδίων / πρωτεϊνών από τις άλλες μεμβράνες του κυττάρου. Η παρουσία ενός συγκεκριμένου φωσφολιπιδίου, της καρδιολιπίνης, έχει υποθεθεί ότι βοηθάει στη μείωση της διαπερατότητας της εσ. μεμβράνης από τα περισσότερα μικρομόρια που βρίσκονται στο διαμεμβρανικό διάστημα<sup>5</sup>, ενώ αυτά που είναι απαραίτητα για τις εσωτερικές διεργασίες του μιτοχονδρίου, μεταφέρονται από συγκεκριμένους διαμεμβρανικούς υποδοχείς. Όπως και στην εξωτερική μεμβράνη, μεγαλύτερες πρωτεΐνες μπορούν να εισέλθουν μόνο εφόσον διαθέτουν συγκεκριμένα πεπτίδια στόχευσης, τα οποία αναγνωρίζονται από τα πρωτεϊνικά σύμπλοκα TIM22 και TIM23 (Translocase of the Inner Membrane – Τρανσλοκάση της Εσωτερικής Μεμβράνης). Επειδή ένα σημαντικό μέρος των λειτουργιών του μιτοχονδρίου πραγματοποιούνται στην επιφάνεια της εσωτερικής μεμβράνης, έχει ευνοηθεί εξελικτικά η μεγιστοποίηση της επιφάνειάς της, κάτι το οποίο επιτυγχάνεται με την παρουσία πολλών αναδιπλώσεων. Οι αναδιπλώσεις αυτές, που ονομάζονται **ακρολόφια**, καλύπτονται από πλήθος πρωτεϊνών, και εκτείνουν την εσωτερική μεμβράνη στο εσωτερικό του μιτοχονδρίου.

Στο εσωτερικό της εσ. μεμβράνης, υπάρχει το δεύτερο μιτοχονδριακό διαμέρισμα, η **μήτρα**. Αυτή περιέχει ένα σημαντικό μέρος των ολικών πρωτεϊνών του μιτοχονδρίου, και σε συνδυασμό με τις πρωτεΐνες που βρίσκονται στην εσωτερική μεμβράνη, επιτελεί το μεγαλύτερο μέρος των λειτουργιών του. Στη μήτρα βρίσκονται επίσης πολλά αντίγραφα του μιτοχονδριακού γονιδιώματος, καθώς και οι μηχανισμοί μεταγραφής και μετάφρασής του, όπως ριβοσώματα και t-RNAs.

Η πιο ευρέως αποδεκτή θεωρία για την προέλευση των μιτοχονδρίων (αλλά και άλλων ενδοκυτταρικών οργανιδίων όπως οι χλωροπλάστες), είναι η **θεωρία της ενδοσυμβίωσης** (Εικόνα 2) . Σύμφωνα με αυτή, τα μιτοχόνδρια είναι απόγονοι βακτηρίων τα οποία παγιδεύτηκαν στο κυτταρόπλασμα αναερόβιων πρώιμων ευκαρυωτικών κυττάρων, και ανέπτυξαν ένα είδος ενδοσυμβίωσης, παρέχοντας στα κύτταρα ενέργεια, με αντάλλαγμα ένα πιο ασφαλές περιβάλλον και παροχή θρεπτικών ουσιών<sup>6</sup>. Η εξωτερική μεμβράνη των μιτοχονδρίων, θεωρείται πως προέρχεται από μέρος της εξωτερικής μεμβράνης του πρώιμου ευκαρυωτικού κυττάρου, ενώ η εσωτερική μεμβράνη τους θεωρείται πως είναι απόγονος της εξωτερικής μεμβράνης του παγιδευμένου βακτηρίου.





**Εικόνα 2:** Θεωρία της Ενδοσυμβίωσης για την προέλευση των μιτοχονδρίων

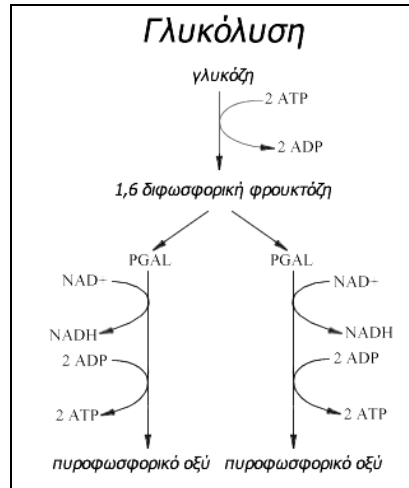
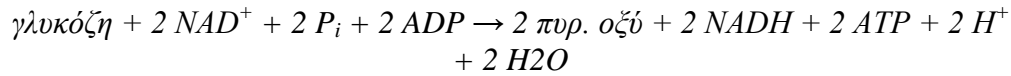
Κάποιες από τις σημαντικές ενδείξεις οι οποίες ισχυροποιούν την θεωρία της ενδοσυμβίωσης είναι οι εξής:

- Η σύσταση της εσωτερικής μεμβράνης των μιτοχονδρίων είναι παρόμοια με αυτή των προκαρυωτικών κυττάρων.
- Το μιτοχονδριακό DNA είναι κυκλικό σε μορφή πλασμιδίου όπως στα βακτήρια, σε αντίθεση με το πυρηνικό το οποίο είναι πακεταρισμένο σε χρωμοσώματα.
- Τα μιτοχόνδρια αναπαράγονται με διχοτόμηση, όπως και τα βακτήρια.
- Τα μιτοχονδριακά ριβοσώματα είναι παρόμοια με αυτά που βρίσκονται στα βακτήρια.
- Τα μιτοχόνδρια δεν ενώνονται με άλλα οργανίδια μέσω του πολύπλοκου ενδοκυτταρικού δικτύου.
- Αρκετές πρωτεΐνες οι οποίες κωδικοποιούνται από πυρηνικά γονίδια, στοχεύονται στο μιτοχόνδριο. Κάτι τέτοιο θα μπορούσε να σημαίνει πως στα πλαίσια της ενδοσυμβίωσης, τα παγιδευμένα βακτήρια άρχισαν να μεταφέρουν κομμάτια του γονιδιώματός τους στον πιο σταθερό πυρήνα του ευκαρυωτικού κυττάρου, προκειμένου να ελαχιστοποιήσουν τις ενεργειακές τους ανάγκες. Τα γονίδια άλλων πρωτεϊνών οι οποίες είναι αρκετά υδρόφοβες ώστε να μην επιτρέπεται η στόχευσή τους στο μιτοχόνδριο μέσω του κυτταροπλάσματος, αναγκαστικά παρέμειναν στο μιτοχονδριακό γονιδίωμα<sup>7</sup>. Αυτή είναι και η κρατούσα θεωρία που εξηγεί την ύπαρξη γενετικού υλικού στα μιτοχόνδρια.

## 1.2. Αερόβια αναπνοή και άλλες μιτοχονδριακές λειτουργίες

Όπως προαναφέρθηκε, η σημαντικότερη μιτοχονδριακή λειτουργία είναι η παραγωγή ενέργειας για το κύτταρο, με τη μορφή ATP. Η διαδικασία αυτή ονομάζεται **αερόβια αναπνοή**, αφού απαιτεί την παρουσία οξυγόνου, και είναι 13-19 φορές πιο αποδοτική από την αναερόβια αναπνοή, η οποία δεν απαιτεί οξυγόνο. Στο στάδιο της (αναερόβιας / αερόβιας) **γλυκόλυσης** (Εικόνα 3), η γλυκόζη διασπάται σε πυροφωσφορικό οξύ, και η ατελής αυτή

καύση απελευθερώνει μονάχα ένα μικρό ποσοστό της διαθέσιμης ενέργειας του μορίου σε μορφή ATP:

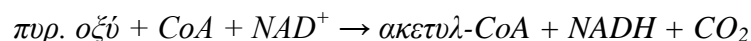


**Εικόνα 3:** Γλυκόλυση

Στους ευκαρυωτικούς οργανισμούς η γλυκόλυση λαμβάνει χώρα στο κυτταρόπλασμα.

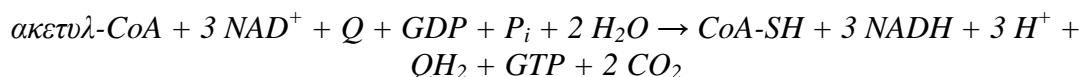
Κατά την αερόβια αναπνοή, το πυροφωσφορικό οξύ που έχει παραχθεί από τη γλυκόλυση οδηγείται στο μιτοχόνδριο, όπου οξειδώνεται πλήρως σε CO<sub>2</sub> και H<sub>2</sub>O, παράγοντας ATP. Η αερόβια αναπνοή απαιτεί την παρουσία μιας σειράς ενζύμων στο εσωτερικό του μιτοχονδρίου, τα οποία αποτελούν την **αναπνευστική αλυσίδα**.

Το πρώτο στάδιο της αερόβιας αναπνοής, είναι η **αποκαρβοξυλίωση του πυροσταφυλικού οξέος** στην εσωτερική μεμβράνη του μιτοχονδρίου, προς σχηματισμό ακετυλοσυνένζυμου A και NADH:



Το στάδιο αυτό καταλύεται από το σύμπλοκο της πυροσταφυλικής δεϋδρογονάσης.

Το επόμενο στάδιο είναι ο **κύκλος του κιτρικού οξέος** (ή κύκλος του Krebs), κατά τον οποίο, ένζυμα που βρίσκονται στη μήτρα του μιτοχονδρίου, διασπούν το ακετυλοσυνένζυμο A σε CO<sub>2</sub>, σε GTP, και σε άλλους συμπαραγόντες απαραίτητους για το επόμενο στάδιο της αναπνοής:



Το στάδιο αυτό καταλύεται από 8 διαφορετικά ένζυμα.

Οι συμπαραγόντες όπως το NADH που παράγονται στο προηγούμενο στάδιο, μεταφέρονται στην εσωτερική μεμβράνη του μιτοχονδρίου, όπου

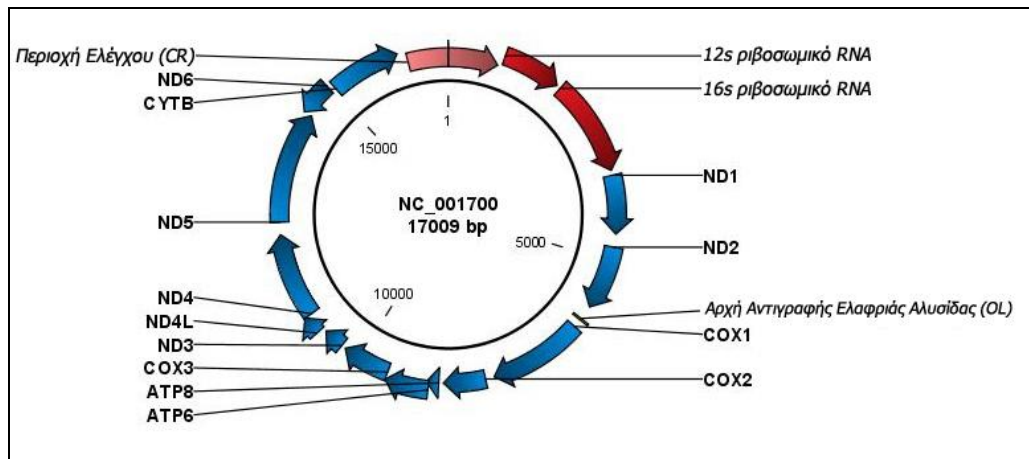
μπαινουν στην **αλυσίδα μεταφοράς ηλεκτρονίων**, μια σειρά πρωτεϊνικών συμπλόκων. Εκεί δρουν σαν ισχυροί δότες ηλεκτρονίων σε αντιδράσεις οξειδοαναγωγής, απελευθερώνοντας πρωτόνια στον διαμεμβρανικό χώρο. Η παρουσία των πρωτονίων αυτών, δημιουργεί ένα ηλεκτροχημικό δυναμικό, το οποίο είναι η κινητήρια δύναμη στην παραγωγή ATP. Το φαινόμενο της σύζευξης της οξείδωσης των συμπαραγόντων με τη φωσφορυλίωση ADP για την παραγωγή ATP, ονομάζεται **οξειδωτική φωσφορυλίωση**, και καταλύεται από 5 διαφορετικά πρωτεϊνικά σύμπλοκα (Παράγραφος 1.5).

Εκτός από την παραγωγή ATP για τις ενεργειακές ανάγκες του κυττάρου, τα μιτοχόνδρια συμμετέχουν σε μια πληθώρα άλλων λειτουργιών, όπως:

- Προγραμματισμένο κυτταρικό θάνατο. Παράγοντες στο εσωτερικό του μιτοχονδρίου, εάν διαχυθούν στο κυτταρόπλασμα, μπορούν να δράσουν σαν ενεργοποιητές κασπασών, και να οδηγήσουν το κύτταρο σε απόπτωση<sup>8</sup>.
- Σηματοδότηση ασβεστίου. Η παραγωγή ATP συγχρονίζεται με αυτόν τον τρόπο, με τις ανάγκες του κυττάρου<sup>9</sup>.
- Ρύθμιση του δυναμικού της μεμβράνης.
- Κυτταρική αύξηση.
- Αποθήκευση ιόντων ασβεστίου και σύνθεση στεροϊδών<sup>10</sup>.
- Παραγωγή θερμότητας.

### 1.3. Μιτοχόνδριακό DNA (mtDNA)

Το μιτοχονδριακό γονιδίωμα των ζώων έχει συνήθως τη μορφή ενός μικρού κυκλικού πλασμίδιου, μεγέθους περίπου 16.000 βάσεων (Εικόνα 4). Σε κάθε μιτοχόνδριο υπάρχουν 2-10 αντίγραφα του γονιδιώματος του. Το συμπαγές αυτό γονιδίωμα κωδικοποιεί τα ίδια 37 γονίδια στα περισσότερα μετάζωα: 2 γονίδια ριβοσωμικού RNA, 13 πρωτεϊνικά γονίδια, και 22 γονίδια tRNA. Τα 28 από αυτά βρίσκονται στη λεγόμενη **βαριά αλυσίδα**, δηλαδή την αλυσίδα του DNA που είναι πιο πλούσια σε γουανίνη (G), ενώ τα υπόλοιπα 9 βρίσκονται στην **ελαφριά αλυσίδα**, η οποία είναι πιο πλούσια σε κυτοσίνη (C). Στον παραπάνω κανόνα υπάρχουν σημαντικές εξαιρέσεις τόσο ως προς το μέγεθος του μιτοχονδριακού γονιδιώματος, όσο και ως προς περιεχόμενό του σε γονίδια. Σημαντική εξαίρεση στον κανόνα, αποτελούν τα μιτοχόνδρια των φυτών, το γονιδίωμα των οποίων σε μερικές περιπτώσεις μπορεί να φτάσει μέχρι και τις 2.500.000 βάσεις<sup>11</sup>. Τέλος, το mtDNA των ζώων δεν περιέχει ιντρόνια.



**Εικόνα 4:** Μιτοχονδριακό DNA του *Felis catus* (γάτα)

Τα γονίδια του μιτοχονδριακού DNA (mtDNA) κωδικοποιούν τα παρακάτω μακρομόρια:

ND1, ND2, ND3, ND4, ND4L, ND5, ND6: Κωδικοποιούν πρωτεΐνες οι οποίες είναι υπομονάδες της NADH δεϋδρογονάσης, του πρώτου πρωτεϊνικού συμπλόκου της αναπνευστικής αλυσίδας (Σύμπλοκο I, Εικόνα 7).

CYTb: Κωδικοποιεί μία από τις υπομονάδες της κυτοχρωμικής αναγωγής, του τρίτου πρωτεϊνικού συμπλόκου της αναπνευστικής αλυσίδας (Σύμπλοκο III, Εικόνα 7).

COX1, COX2, COX3: Κωδικοποιούν πρωτεΐνες οι οποίες είναι υπομονάδες της κυτοχρωμικής c οξειδάσης, το οποίο είναι το τέταρτο πρωτεϊνικό σύμπλοκο στην αναπνευστική αλυσίδα (Σύμπλοκο IV, Εικόνα 7).

ATP6, ATP8: Κωδικοποιούν υπομονάδες του συμπλόκου της ATP συνθάσης, του πέμπτου και τελευταίου συμπλόκου της αναπνευστικής αλυσίδας (Σύμπλοκο V, Εικόνα 7).

12sRNA, 16sRNA: Συμμετέχουν στις δυο αντίστοιχες υπομονάδες του μιτοχονδριακού ριβοσώματος.

22 tRNA γονίδια: Μεταγράφονται στα 22 μόρια tRNA που εμπλέκονται στη μετάφραση του μιτοχονδριακού γονιδιώματος.

Άλλες περιοχές του μιτοχονδριακού γονιδιώματος:

Περιοχή Ελέγχου (Displacement Loop): Η περιοχή ελέγχου είναι η τοποθεσία πάνω στο μιτοχονδριακό γονιδίωμα όπου ξεκινάει η αντιγραφή της βαριάς αλυσίδας. Το όνομα D-loop προέρχεται από το δόμημα τριών αλυσιδών DNA που δημιουργείται όταν η δημιουργούμενη βαριά αλυσίδα εκτοπίζει (displaces) την πατρική.

Αρχή Αντιγραφής Ελαφριάς Αλυσίδας (Origin of Light Strand Replication - OL): Η αντιγραφή της ελαφριάς αλυσίδας ξεκινάει από αυτή την περιοχή. Μόλις η διπλή αλυσίδα ανοίξει σε αυτή την περιοχή κατά τη διάρκεια της αντιγραφής της βαριάς αλυσίδας, ξεκινάει η σύνθεση της νέας ελαφριάς αλυσίδας προς την αντίθετη κατεύθυνση. Η περιοχή αυτή στα θηλαστικά, βρίσκεται συνήθως γύρω στις 11.000 βάσεις κάτω από την περιοχή ελέγχου.

**Μη Προσδιορισμένες Περιοχές (Unassigned Regions):** Είναι μη κωδικοποιούσες περιοχές του μιτοχονδριακού γονιδιώματος. Οι περιοχές αυτές δεν είναι παρούσες σε όλους τους οργανισμούς, και θεωρείται πως δεν βρίσκονται κάτω από την επίδραση κάποιου είδους επιλογής. Το συνολικό μήκος αυτών των περιοχών είναι σχετικά μικρό γι' αυτό και το mtDNA των ζώων είναι συμπαγές.

Η μετάφραση των μιτοχονδριακών γονιδίων δεν ακολουθεί τον παγκόσμιο γενετικό κώδικα, αλλά διαφορετικοί οργανισμοί χρησιμοποιούν ελαφρά διαφορετικούς κώδικες (Εικόνα 5)<sup>12</sup>. Έτσι, υπάρχει διαφορετικός μιτοχονδριακός κώδικας για τα θηλαστικά, για τα ασπόνδυλα, για τους πλατυέλμινθες, τη ζύμη, κ.ο.κ.

Οργανισμός	Κωδικόνιο	Παγκόσμιο	Μιτοχονδριακό
Θηλαστικά	AGA, AGG	Αργινίνη	Κωδικόνιο λήξης
	AUA	Ισολευκίνη	Μεθιονίνη
	UGA	Κωδικόνιο λήξης	Τρυπτοφάνη
Ασπόνδυλα	AGA, AGG	Αργινίνη	Σερίνη
	AUA	Ισολευκίνη	Μεθιονίνη
	UGA	Κωδικόνιο λήξης	Τρυπτοφάνη
Ζύμη	AUA	Ισολευκίνη	Μεθιονίνη
	UGA	Κωδικόνιο λήξης	Τρυπτοφάνη
	CUA	Λευκίνη	Θρεονίνη

**Εικόνα 5:** Παράδειγμα διαφορών ανάμεσα σε γενετικούς κώδικες

Η συνεχής παραγωγή ATP στα μιτοχόνδρια μέσω του μονοπατιού της οξειδωτικής φωσφορυλίωσης, έχει ως αποτέλεσμα την δημιουργία ελεύθερων ριζών οξυγόνου που μετατρέπουν το περιβάλλον σε ισχυρά οξειδωτικό, προκαλώντας ζημιές στο γονιδίωμα και στις πρωτεΐνες. Αν και το μιτοχόνδριο διαθέτει σύμπλοκα επισκευής DNA<sup>13</sup>, αυτά είναι λιγότερο αποτελεσματικά σε σχέση με τους μηχανισμούς επιδιόρθωσης του πυρηνικού DNA. Στη μειωμένη αποτελεσματικότητα των επιδιορθωτικών μηχανισμών και στο εξαιρετικά δραστήριο χημικά εσωτερικό περιβάλλον του μιτοχονδρίου αποδίδεται ο αυξημένος μεταλλακτικός ρυθμός του mtDNA (Παράγραφος 2.1). Ο μεταλλακτικός αυτός ρυθμός που έχει υπολογιστεί κοντά στο  $\sim 6 \times 10^{-8}$  ανά θέση, ανά γενεά στη *Drosophila melanogaster*<sup>14</sup>, και  $\sim 3 \times 10^{-5}$  στον *Homo sapiens*<sup>15</sup>. Ο αυξημένος μεταλλακτικός ρυθμός σε σχέση με το πυρηνικό DNA (ο οποίος για παράδειγμα στον *Homo sapiens* υπολογίζεται στο  $\sim 2.5 \times 10^{-8}$ )<sup>16</sup> αντικατοπτρίζεται στον αυξημένο ρυθμό εξέλιξης, ο οποίος είναι κατά κανόνα μια τάξη μεγέθους μεγαλύτερος σε σχέση με το μέσο ρυθμό εξέλιξης του αντίστοιχου πυρηνικού DNA<sup>17</sup>. Σε αντίθεση με τα ζωικά μιτοχόνδρια, στα φυτικά συμβαίνει το ακριβώς αντίθετο, δηλαδή ο ρυθμός εξέλιξής τους είναι μια τάξη μεγέθους μικρότερος από τον πυρηνικό. Αυτό εξηγείται από την ύπαρξη κάποιων πρωτεϊνικών συμπλόκων (4 NADH δευδρογονάσες και μία οξειδάση), τα οποία επιτρέπουν στα φυτικά μιτοχόνδρια να περιορίζουν τη λειτουργία τους κάτω από συνθήκες στρες, μειώνοντας έτσι την παραγωγή

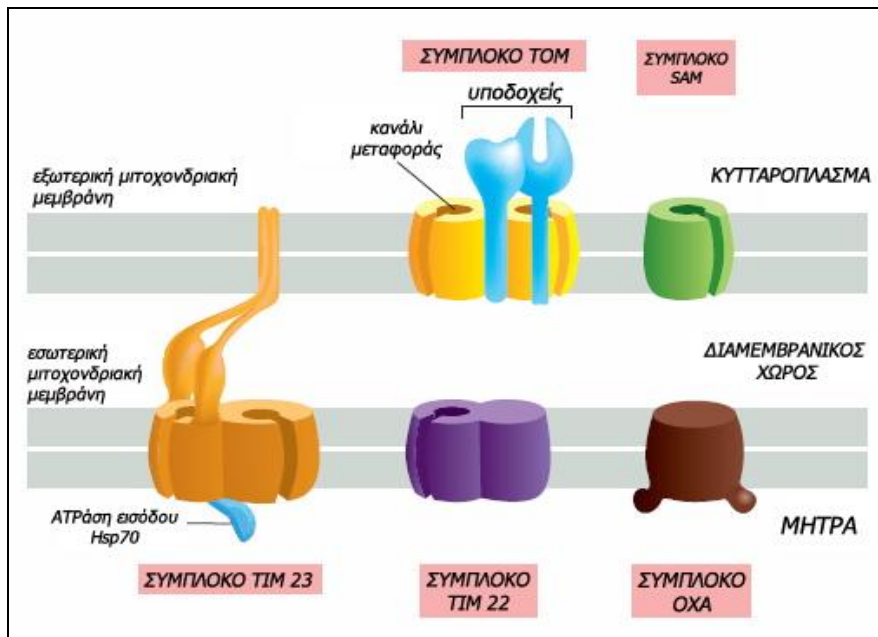
ελεύθερων ριζών και δίνοντας χρόνο στα σύμπλοκα επισκευής να διορθώσουν υπάρχουσες ζημιές<sup>18</sup>.

#### **1.4. Πυρηνικές πρωτεΐνες που στοχεύονται στο μιτοχόνδριο**

Αν και μόνο 13 πρωτεΐνες κωδικοποιούνται στα ζωικά μιτοχόνδρια, το πλήθος των πρωτεϊνών που βρίσκονται μέσα σε αυτά, μπορούν να ξεπερνούν τις 1000. Οι υπόλοιπες από αυτές συντίθενται έξω από αυτό, στο κυτταρόπλασμα, από πυρηνικά γονίδια. Προκειμένου να οδηγηθούν στο μιτοχόνδριο, απαιτείται ένας αριθμός διαφορετικών βημάτων και συμπλόκων (Εικόνα 6):

- Οι πρόδρομες πρωτεΐνες οι οποίες περιέχουν συγκεκριμένα πεπτίδια στόχευσης, συντίθενται στο κυτταρόπλασμα, χωρίς ωστόσο να αναδιπλώνονται εκεί, κάτι το οποίο επιτυγχάνεται με τη βοήθεια τσαπερονών.
- Το πεπτίδιο στόχευσης της πρωτεΐνης αλληλεπιδρά με το σύμπλοκο TOM, και μπαίνει στο κανάλι μεταφοράς.
- Το κατάλληλο πεπτίδιο στόχευσης, αφού φτάσει στον διαμεμβρανικό χώρο, μπορεί να αλληλεπιδράσει με το σύμπλοκο TIM23 και να μπει στο κανάλι μεταφοράς προς τη μήτρα. Εκεί κόβεται το πεπτίδιο στόχευσης από πρωτεάσες, και η πρωτεΐνη αναδιπλώνεται και μένει στη μήτρα.
- Εάν η πρωτεΐνη διαθέτει κατάλληλα εσωτερικά πεπτίδια στόχευσης, μπορεί να αλληλεπιδράσει με τη βοήθεια τσαπερονών με το σύμπλοκο TIM22, και να εισέλθει στην εσωτερική μεμβράνη του μιτοχονδρίου.
- Τέλος, οι πορίνες δεν αλληλεπιδρούν με κανένα σύμπλοκο TIM, αλλά με το σύμπλοκο SAM με τη βοήθεια τσαπερονών, και εισέρχονται στην εξωτερική μεμβράνη όπου αναδιπλώνονται και δρουν.
- Το σύμπλοκο OXA βρίσκεται στην εσωτερική μεμβράνη του μιτοχονδρίου, και κάνει παρόμοια δουλειά με το σύμπλοκο TIM22, με τη διαφορά ότι σε αυτό στοχεύονται συνήθως οι πρωτεΐνες που κατασκευάζονται στο μιτοχόνδριο.

Οι πρωτεΐνες που κωδικοποιούνται από το πυρηνικό γονιδίωμα αλλά στοχεύονται στο μιτοχόνδριο συμμετέχουν είτε στη δομή είτε στη λειτουργία του μιτοχονδρίου. Ειδικότερα, μερικές από αυτές συνεργάζονται με τις πρωτεΐνες που κωδικοποιούνται από το mtDNA για να φτιάξουν τα ενζυμικά σύμπλοκα της αναπνευστικής αλυσίδας (Εικόνες 7 και 8).



Εικόνα 6: Σύμπλοκα μεταφοράς πρωτεϊνών στο μιτοχόνδριο

### 1.5. Σύμπλοκα οξειδωτικής φωσφορυλίωσης.

Το μονοπάτι της οξειδωτικής φωσφορυλίωσης, αποτελείται από 5 πρωτεϊνικά σύμπλοκα, τα οποία βρίσκονται είτε στην μήτρα, είτε στην εσωτερική μεμβράνη του μιτοχονδρίου (Εικόνα 7). Με εξαίρεση το Σύμπλοκο II, τα υπόλοιπα αποτελούνται τόσο από πυρηνικές πρωτεΐνες (υπομονάδες) που στοχεύονται στο μιτοχόνδριο, όσο και από πρωτεΐνες που παράγονται εκεί (Εικόνα 8):

#### Σύμπλοκο I: NADH δεϋδρογονάση.

Το σύμπλοκο της δεϋδρογονάσης του νικοτιναμιδο-αδενινωδινουκλεοτίδιου (NADH) καταλύει τη μεταφορά ηλεκτρονίων από το NADH, στο συνένζυμο-Q (CoQ, ουβικινόνη). Με κάθε τέτοια αντίδραση, τέσσερα πρωτόνια μεταφέρονται έξω από την εσωτερική μεμβράνη, δημιουργώντας ένα δυναμικό πρωτονίων (το σύμπλοκο δηλαδή δρα σαν αντλία πρωτονίων). Αποτελείται από περίπου 45 υπομονάδες (ανάλογα με το είδος του οργανισμού), οι 7 εκ των οποίων είναι μιτοχονδριακές.

#### Σύμπλοκο II: Ηλεκτρικο-αναγωγή του συνένζυμου Q.

Το σύμπλοκο της ηλεκτρικο-αναγωγής του συνένζυμου Q οξειδώνει το ηλεκτρικό οξύ σε φουμαρικό οξύ. Τα ηλεκτρόνια που πηγαίνουν στο Q, χρησιμοποιούνται στο επόμενο βήμα της οξειδωτικής φωσφορυλίωσης. Αποτελείται από 4 υπομονάδες, όλες πυρηνικές.

#### Σύμπλοκο III: Κυτοχρωμική αναγωγή.

Το σύμπλοκο της κυτοχρωμικής αναγωγής αφαιρεί ηλεκτρόνια από το Q, και τα μεταφέρει στο κυτόχρωμα c. Η αναγωγή του Q δημιουργεί

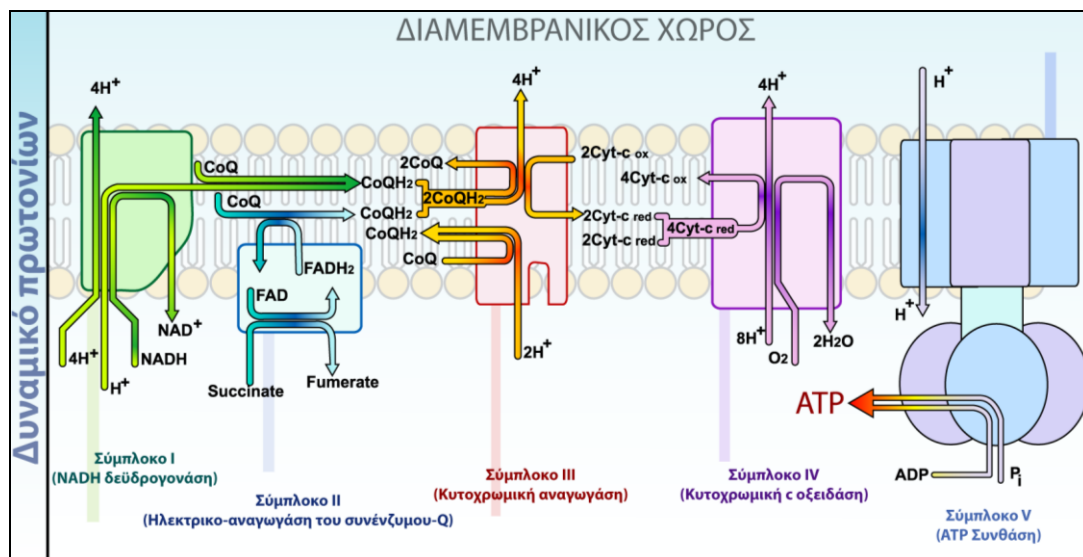
δυναμικό πρωτονίων. Το σύμπλοκο αυτό αποτελείται από 11 υπομονάδες, εκ των οποίων, μόνο η μία είναι μιτοχονδριακή.

Σύμπλοκο IV: Κυτοχρωμική οξειδάση.

Το σύμπλοκο της κυτοχρωμικής οξειδάσης αφαιρεί ηλεκτρόνια από το κυτόχρωμα c, και τα μεταφέρει στο οξυγόνο. Η αντίδραση αυτή παράγει τέσσερα πρωτόνια τα οποία ενισχύουν το υπάρχον δυναμικό. Το σύμπλοκο αυτό αποτελείται από 13 υπομονάδες, εκ των οποίων οι 3 είναι μιτοχονδριακές.

Σύμπλοκο V: ATP συνθάση.

Το τελικό σύμπλοκο στην αλυσίδα, η ATP συνθάση, χρησιμοποιεί το δυναμικό πρωτονίων που δημιουργήθηκε στα προηγούμενα βήματα, για να συνθέσει ATP από ADP και φωσφορικό άλας. Αποτελείται από 16 υπομονάδες, εκ των οποίων οι 2 είναι μιτοχονδριακές.



**Εικόνα 7: Τα 5 σύμπλοκα της οξειδωτικής φωσφορυλίωσης**



## 2. ΣΥΓΚΡΙΣΗ ΤΩΝ ΕΞΕΛΙΚΤΙΚΩΝ ΡΥΘΜΩΝ ΤΩΝ ΓΟΝΙΔΙΩΝ ΠΟΥ ΚΩΔΙΚΟΠΟΙΟΥΝΤΑΙ ΑΠΟ ΜΙΤΟΧΟΝΔΡΙΑΚΟ ΚΑΙ ΠΥΡΗΝΙΚΟ DNA ΚΑΙ ΠΟΥ ΤΑ ΠΡΟΪΟΝΤΑ ΤΟΥΣ ΣΥΝΘΕΤΟΥΝ ΤΑ ΣΥΜΠΛΟΚΑ ΤΗΣ ΟΞΕΙΔΩΤΙΚΗΣ ΦΩΣΦΟΡΥΛΙΩΣΗΣ.

### 2.1 Εισαγωγή

**Μεταλλάξεις** ονομάζονται οι αλλαγές στην νουκλεοτιδική αλληλουχία του DNA ενός οργανισμού (ή RNA εάν πρόκειται για κάποιους ιούς). Εάν πρόκειται για μεταλλάξεις που αλλάζουν τον αριθμό ή τη δομή ενός χρωμοσώματος (πχ. αναστροφή ενός μέρους του), τότε μιλάμε για χρωμοσωμικές μεταλλάξεις, ενώ εάν πρόκειται για μεταλλάξεις που επιδρούν σε ένα μόνο γονίδιο (πχ. προσθαφαίρεση νουκλεοτιδικών βάσεων, ανακατατάξεις εξωνίων), τότε μιλάμε για γονιδιακές μεταλλάξεις. Είναι προφανές ότι κάθε γονιδιακή μετάλλαξη είναι και χρωμοσωμική.

Όταν μια μετάλλαξη αφορά μία μόνο βάση DNA/RNA, τότε μιλάμε για **σημειακή μεταλλαγή**. Υπάρχουν τριών ειδών σημειακές μεταλλαγές:

- Μεταλλαγές αντικατάστασης, όταν μια βάση μετατραπεί σε μια άλλη βάση. Ονομάζονται μεταπτώσεις, όταν μια βάση πουρίνης (αδενίνη ή γουανίνη) ή πυριμιδίνης (θυμίνη, ή κυτοσίνη για DNA, ουρακίλη ή κυτοσίνη για RNA) μετατραπεί σε μια άλλη πουρίνη ή πυριμιδίνη αντίστοιχα, και μεταστροφές, όταν μια βάση πουρίνης μετατραπεί σε μία βάση πυριμιδίνης ή το αντίθετο. Επειδή κάθε βάση μπορεί να υποστεί δυο μεταστροφές αλλά μία μόνο μετάπτωση, εάν οι μεταλλάξεις γίνονται με τυχαίο τρόπο, τότε οι μεταστροφές είναι διπλάσιες από τις μεταπτώσεις. Στην πραγματικότητα ωστόσο, μεταπτώσεις παρατηρούνται αρκετά συχνότερες από ότι μεταστροφές<sup>19</sup>.
- Μεταλλαγές πρόσθεσης νουκλεοτιδίου, όταν δίπλα σε μια προϋπάρχουσα βάση, προστίθεται μια νέα βάση.
- Μεταλλαγές αφαίρεσης νουκλεοτιδίου, όταν αφαιρείται μια προϋπάρχουσα βάση. Οι μεταλλαγές πρόσθεσης και αφαίρεσης, εάν συμβούν σε εξώνιο γονιδίου, μπορούν να αλλάξουν το πλαίσιο ανάγνωσης του και να αλλάξουν ριζικά την αμινοξική μετάφρασή του (παρερμηνεύσιμες μεταλλάξεις), ή να δημιουργήσουν ένα πρόωρο κωδικόνιο λήξης (ανερμηνεύσιμες μεταλλάξεις).

Δεδομένης της ύπαρξης περισσότερων του ενός κωδικονίων για κάθε μεταφραζόμενο αμινοξύ (συνώνυμα κωδικώνια), οι μεταλλαγές αντικατάστασης που συμβαίνουν σε εξώνια γονιδίων μπορούν είτε να προκαλέσουν την αλλαγή του κωδικοποιούμενου αμινοξέος (αμινοξική μεταλλαγή ή μη συνώνυμη μεταλλαγή), είτε όχι (σιωπηλή μεταλλαγή ή συνώνυμη μεταλλαγή). Κάθε θέση ενός κωδικονίου μπορεί να χαρακτηριστεί σαν **συνώνυμη**, όταν οποιαδήποτε αντικατάσταση της βάσεως σε αυτή τη θέση δεν προκαλεί αμινοξική αντικατάσταση, και σαν **μη συνώνυμη**, όταν αντικατάσταση της βάσεως σε αυτή τη θέση μπορεί να έχει ως αποτέλεσμα αμινοξική μεταλλαγή. Για όλους τους γενετικούς κώδικες, οι πρώτες θέσεις

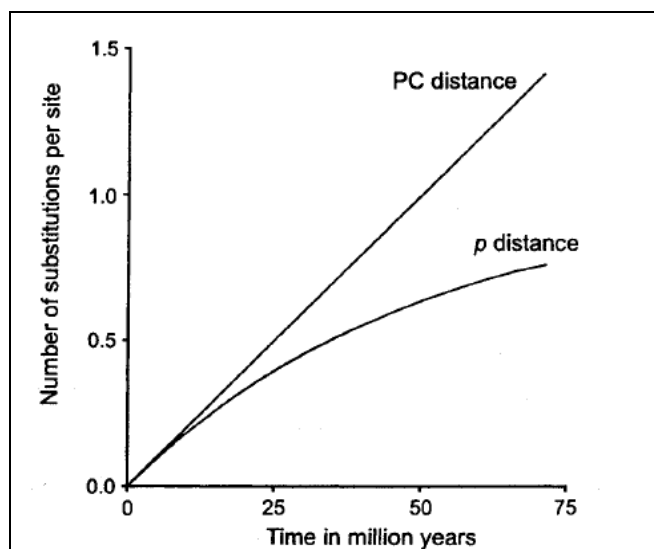
των κωδικονίων είναι συνήθως μη συνώνυμες αλλά με αρκετές εξαιρέσεις, οι δεύτερες θέσεις είναι πάντοτε μη συνώνυμες, ενώ τέλος οι τρίτες θέσεις είναι σχεδόν πάντα συνώνυμες. Οι μεταλλάξεις μπορούν να χαρακτηριστούν σαν ουδέτερες, όταν η φυσική επιλογή δεν επιδρά σε αυτές, σαν επιβλαβείς όταν ασκείται σε αυτές αρνητική επιλογή (για την ακρίβεια στα άτομα που τις φέρουν) ή τέλος σαν ευνοϊκές, όταν ασκείται σε αυτές θετική επιλογή (βλέπε παρακάτω).

Η πορεία των μεταλλάξεων στον πληθυσμό εξαρτάται από δύο σημαντικούς παράγοντες. Ο πρώτος είναι το μέγεθος του πληθυσμού σαν συνέπεια της **τυχαίας γενετικής παρέκκλισης**. Η πιθανότητα μια ουδέτερη μετάλλαξη να φτάσει σε συχνότητα πολύ κοντά στο 1 (επομένως να εγκαθιδρυθεί στον πληθυσμό) είναι αντιστρόφως ανάλογη με το διπλάσιο του μεγέθους του πληθυσμού, αν ο πληθυσμός αποτελείται από διπλοειδείς οργανισμούς. Με άλλα λόγια, σε ένα μεγάλο πληθυσμό η πιθανότητα εγκαθίδρυσης μιας ουδέτερης μετάλλαξης είναι μικρότερη σε σχέση με έναν μικρό πληθυσμό. Ο δεύτερος είναι ο συντελεστής επιλογής της μετάλλαξης. Μια επιβλαβής μετάλλαξη έχει μικρότερη πιθανότητα να εγκαθιδρυθεί στον πληθυσμό σε σχέση με μια ουδέτερη, ενώ αντίστοιχα, μια ευνοϊκή μετάλλαξη έχει μεγαλύτερη πιθανότητα να εγκαθιδρυθεί στον πληθυσμό σε σχέση με μια ουδέτερη. Όταν μια μετάλλαξη εγκαθιδρυθεί στον πληθυσμό, ονομάζεται **αντικατάσταση**.

Ως **εξελικτικός ρυθμός**, ορίζεται ο ρυθμός με τον οποίο μια νουκλεοτιδική ή πρωτεϊνική αλληλουχία αλλάζει με το πέρασμα του χρόνου. Ο ρυθμός αυτός είναι ο αριθμός αντικαταστάσεων ανά νουκλεοτιδική θέση, ανά γενεά, και αφορά μόνο τις μεταλλάξεις οι οποίες έχουν εγκαθιδρυθεί στον πληθυσμό που εξετάζουμε. Η συσσώρευση αντικαταστάσεων οδηγεί σε αύξηση της **γενετικής απόστασης** ανάμεσα σε δυο αλληλουχίες. Για παράδειγμα, δυο όμοιες αλληλουχίες των 100 νουκλεοτιδίων έχουν γενετική απόσταση 0%, ενώ εάν διαφέρουν σε μια νουκλεοτιδική θέση, έχουν απόσταση 1%.

Προκειμένου να γίνει συγκριτική μελέτη δυο ή παραπάνω ομόλογων αλληλουχιών, ο κλάδος της μοριακής εξέλιξης εξετάζει τον αριθμό των παρατηρούμενων αντικαταστάσεων, είτε ανάμεσα σε αυτές, είτε ανακατασκευάζοντας προγονικές τους αλληλουχίες και κάνοντας αντίστοιχες συγκρίσεις. Ένα συνηθισμένο πρόβλημα, το οποίο εντείνεται όσο μεγαλώνει ο εξελικτικός χρόνος (δηλαδή ο χρόνος που έχει περάσει από την εποχή που διαχωρίστηκαν από τον κοινό τους πρόγονο οι αλληλουχίες) μεταξύ των αλληλουχιών, είναι το πρόβλημα των **πολλαπλών αντικαταστάσεων** (multiple hits). Σύμφωνα με αυτό, όσο αυξάνεται ο εξελικτικός χρόνος ανάμεσα σε δυο αλληλουχίες, τόσο πιο πιθανό είναι να γίνουν περισσότερες από μια αντικαταστάσεις σε μια νουκλεοτιδική θέση, με αποτέλεσμα, ενώ θα έπρεπε η γενετική απόσταση μεταξύ τους να αυξάνεται, να μην παρατηρείται κάτι τέτοιο, ή ακόμα και να μειώνεται. Μείωση της απόστασης παρατηρείται στην περίπτωση που μια διπλή αντικατάσταση σε μια θέση οδηγήσει σε επαναφορά του αρχικού νουκλεοτιδίου, ή στην περίπτωση που η ίδια αντικατάσταση παρατηρηθεί και στις δυο αλληλουχίες, σε σχέση με την

προγονική (ομοπλασία). Δεδομένης της ύπαρξης τεσσάρων μόνο βάσεων, η γενετική απόσταση ανάμεσα σε δυο αλληλουχίες τείνει να μεγιστοποιηθεί στο 75%, άσχετα με τον μεταλλακτικό ρυθμό ή τον εξελικτικό χρόνο<sup>20</sup>. Προκειμένου να αντιμετωπισθεί αυτό το πρόβλημα, έχουν επινοηθεί διαφορετικά μοντέλα διόρθωσης, όπως το Jukes-Kantor (JK69)<sup>21</sup>, το Kimura's 2 Parameter (K2P81)<sup>22</sup>, και πιο πρόσφατα το Hasewaga, Kishino, Yano (HKY85)<sup>23</sup>. Οι διορθώσεις γίνονται λαμβάνοντας υπ' όψιν τη νουκλεοτιδική σύσταση της κάθε αλληλουχίας, και τους παρατηρούμενους ρυθμούς μεταπτώσεων-μεταστροφών.



**Εικόνα 8:** Παράδειγμα σύγκρισης του παρατηρούμενου ( $p$ ) ρυθμού αντικαταστάσεων ανά θέση, ανά εκατομμύριο χρόνια, και του ίδιου ρυθμού ύστερα από διόρθωση (Poisson Correction) σε μια πρωτεϊνική αλληλουχία

Ανάλογα με το πρόβλημα που καλούμαστε να εξετάσουμε μπορεί να χρησιμοποιήσουμε διαφορετικές μετρήσεις της διαφοράς δύο αλληλουχιών. Μπορούμε για παράδειγμα να χρησιμοποιήσουμε το σύνολο των διαφορών μεταξύ δύο αλληλουχιών. Όμως, αν η αλληλουχία είναι κωδικοποιούσα, μπορούμε να υπολογίσουμε χωριστά την απόσταση που οφείλεται στις μη συνώνυμες θέσεις και την απόσταση που οφείλεται στις συνώνυμες θέσεις. Η αποστάσεις αυτές αναμένεται να είναι διαφορετικές διότι οι αντίστοιχες θέσεις βρίσκονται κάτω από διαφορετική επίδραση της φυσικής επιλογής:

Ως **dN** ορίζεται ο αριθμός των μη συνώνυμων (**N**on synonyms) αντικαταστάσεων ανάμεσα σε δυο ομόλογες κωδικές αλληλουχίες, ανά μη συνώνυμη νουκλεοτιδική θέση. Για παράδειγμα, εάν οι δυο αλληλουχίες έχουν 100 μη συνώνυμες θέσεις και διαφέρουν στις 5 από αυτές, τότε ο ρυθμός dN ισούται με 0.05. Επειδή μια μη συνώνυμη μετάλλαξη θα έχει ως αποτέλεσμα την αλλαγή του κωδικοποιούμενου αμινοξέος, πάνω της μπορεί να ασκηθεί επιλογή. Όταν γίνεται σύγκριση με μια προγονική της αλληλουχία, η τιμή dN είναι ουσιαστικά ο ρυθμός εξέλιξης μιας πρωτεΐνης.

Ως **ds** ορίζεται το ο αριθμός των συνώνυμων (**S**ynonyms) αντικαταστάσεων ανάμεσα σε δυο ομόλογες κωδικές αλληλουχίες, ανά συνώνυμη νουκλεοτιδική θέση. Θεωρητικά, οι μεταλλάξεις αυτές είναι

ουδέτερες, δηλαδή δεν έχουν επίδραση πάνω στην αρμοστικότητα του πληθυσμού (την ικανότητα επιβίωσής του στο περιβάλλον του), και μπορούν να χρησιμοποιηθούν σαν δείκτης για τον μεταλλακτικό ρυθμό ενός γονιδίου. Αν και υπάρχουν και άλλες περιοχές του γονιδιώματος οι οποίες μπορούν να χρησιμοποιηθούν σαν τέτοιοι δείκτες, όπως τα ιντρόνια ή άλλες μη κωδικές περιοχές, ωστόσο οι συνώνυμες θέσεις είναι προτιμώτερες γιατί έχουν την ίδια εξελικτική ιστορία με τις μη συνώνυμες θέσεις του εκάστοτε γονιδίου, και επιτρέπουν περαιτέρω συγκρίσεις.

Ωστόσο, έχει αποδειχθεί πως υπάρχει προτίμηση χρήσης συγκεκριμένων κωδικονίων (**codon usage bias**) στα πυρηνικά γονιδιώματα πολλών οργανισμών, και χρήση μη προτιμώμενων κωδικονίων έχει επιπτώσεις στην αρμοστικότητά τους<sup>24 25</sup>. Το ίδιο ισχύει και για τα μιτοχονδριακά γονιδιώματα, όπου η ύπαρξη 22 μόνο tRNA γονιδίων προϋποθέτει ότι μερικά από αυτά πρέπει να αναγνωρίζουν έως και τέσσερα διαφορετικά κωδικόνια, ενώ υπάρχουν τάσεις «προτίμησης» για κάποια από αυτά<sup>26 27 28</sup>. Τέλος, υπάρχουν και οι περιπτώσεις άκρων εξωνίων, όπου μια μετάλλαξη, αν και συνώνυμη, μπορεί να αλλάξει την αλληλουχία των βάσεων που χρειάζονται για το μάτισμα του RNA<sup>29</sup>. Όμως, ακόμα και αν ασκείται επιλογή στις συνώνυμες βάσεις, αυτή είναι πολύ ασθενέστερη σε σχέση με την επιλογή που ασκείται στις μη συνώνυμες βάσεις. Ως ουδέτερο control για την ύπαρξη ή όχι επιλογής στις αλληλουχίες, μπορεί να χρησιμοποιηθούν άλλες περιοχές του γονιδιώματος που είναι "περισσότερο ουδέτερες" από τις συνώνυμες θέσεις. Τέτοιες περιοχές είναι συνήθως ιντρόνια (εκτός του πρώτου ιντρονίου και των 10 –περίπου- πρώτων θέσεων του κάθε ιντρονίου που πιθανόν να έχουν ρυθμιστικό ρόλο). Η χρησιμοποίηση ιντρονίων εξασφαλίζει μεγαλύτερη ουδετερότητα όμως έχει δύο σημαντικά μειονεκτήματα. Πρώτον, στις δικές μας αναλύσεις τα μιτοχονδριακά γονίδια δεν διαθέτουν ιντρόνια. Δεύτερον, διαφορετικές περιοχές του (πυρηνικού) γονιδιώματος μπορεί να έχουν διαφορετικές εξελικτικές ιστορίες λόγω της ύπαρξης ανασυνδυασμού. Αν υπάρχουν διαφορετικές εξελικτικές ιστορίες (π.χ. διαφορετικός μεταλλακτικός ρυθμός ή δημογραφικά φαινόμενα) επηρεάζεται το αποτέλεσμα των αναλύσεων. Αντίθετα οι συνώνυμες θέσεις έχουν την ίδια εξελικτική ιστορία με τις μη συνώνυμες διότι εξ ορισμού, συνυπάρχουν μέσα στο κάθε κωδικόνιο.

Ως **dN/dS ( $\omega$ )** ορίζεται το πηλίκο του ρυθμού των μη συνώνυμων προς τον ρυθμό των συνώνυμων αντικαταστάσεων σε ένα γονίδιο. Η τιμή αυτή μας δίνει μια εκτίμηση για την ένταση της επιλογής που ασκείται πάνω σε αυτό<sup>30</sup>. Έτσι, για τιμές:

- $<1$ , οι συνώνυμες αντικαταστάσεις είναι περισσότερες από τις μη συνώνυμες, και σε αυτό το γονίδιο ασκείται **αρνητική επιλογή** (εκκαθαριστική επιλογή), η οποία απομακρύνει ένα ποσοστό επιβλαβών μη συνώνυμων μεταλλάξεων. Με άλλα λόγια, εάν υποθεθεί ότι ο αριθμός των μεταλλάξεων είναι ίδιος τόσο στις συνώνυμες όσο και στις μη συνώνυμες θέσεις, η ύπαρξη περισσότερων αντικαταστάσεων στις μη συνώνυμες, υποδεικνύει ότι ένα μέρος των μη συνώνυμων μεταλλάξεων δεν εγκαθιδρύθηκε στον πληθυσμό, γιατί προκαλούσε μειωμένη αρμοστικότητα απέναντι στο περιβάλλον του.

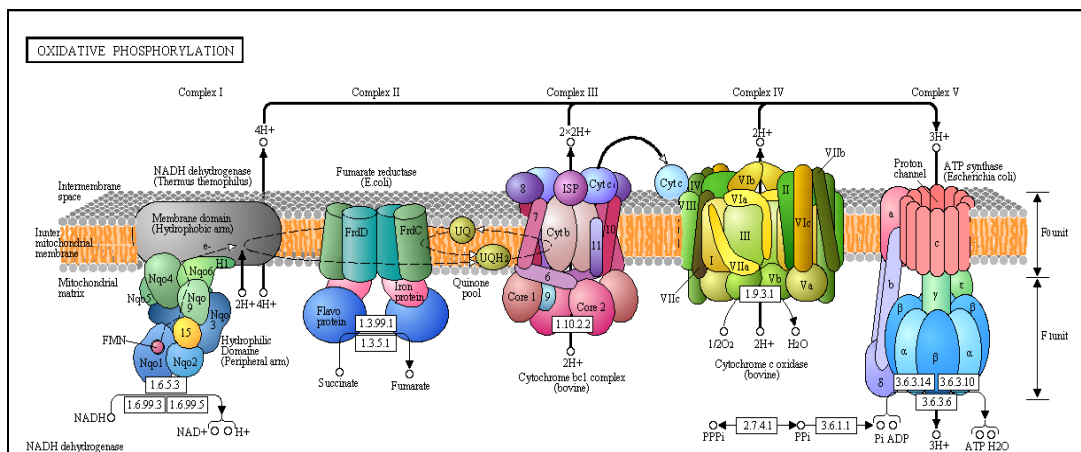
- $=1$ , οι συνώνυμες αντικαταστάσεις είναι περίπου ίσες με τις μη συνώνυμες, και σε αυτό το γονίδιο δεν ασκείται επιλογή, δηλαδή εξελίσσεται **ουδέτερα**. Με άλλα λόγια, οι μεταλλάξεις στις μη συνώνυμες θέσεις δεν εξαλείφθηκαν από την επιλογή, γιατί πιθανόν να μην είχαν καμία αρνητική επίδραση στην αρμοστικότητα του πληθυσμού απέναντι στο περιβάλλον του.
- $>1$ , οι συνώνυμες αντικαταστάσεις είναι λιγότερες από τις μη συνώνυμες, και είναι σε αυτό το γονίδιο ασκείται **θετική επιλογή**, η οποία εγκαθιδρύει ένα ποσοστό ωφέλιμων μη συνώνυμων μεταλλάξεων. Με άλλα λόγια, μεταλλάξεις σε μη συνώνυμες θέσεις είχαν θετική επίδραση στην αρμοστικότητα των ατόμων που τις διέθεταν, με αποτέλεσμα αυτές να εξαπλωθούν και να εγκαθιδρυθούν γρήγορα στον πληθυσμό.

Ένα πρόβλημα με τη χρήση του ρυθμού  $\omega$ , είναι ότι το είδος της επιλογής πολλές φορές δεν είναι ξεκάθαρο. Για παράδειγμα, διαφορετικά τμήματα ενός γονιδίου μπορεί να δέχονται διαφορετικές εξελικτικές πιέσεις. Έτσι, ένα τμήμα του γονιδίου (πχ. μία από τις λειτουργικές επικράτειες) μπορεί να βρίσκεται υπό την επίδραση θετικής επιλογής, ενώ το υπόλοιπο γονίδιο να βρίσκεται κάτω από αρνητική. Ο συνολικός ρυθμός  $\omega$  θα μπορούσε τότε να είναι  $<0$ , και τυχόν ανάλυση του γονιδίου να μην εντόπιζε τη δράση της θετικής επιλογής<sup>31</sup>. Ένας τρόπος που έχει επινοηθεί για την ανίχνευση θετικής επιλογής σε τέτοιες περιπτώσεις, είναι ο υπολογισμός διαφορετικής τιμής  $\omega$  για το κάθε κωδικόνιο (περισσότερα στο Κεφάλαιο 3).

Εδώ και αρκετά χρόνια, έχει δειχθεί πως στα μετázωα, ο ρυθμός εξέλιξης του μιτοχονδριακού γονιδιώματος είναι περίπου 10 φορές μεγαλύτερος από αυτόν του πυρηνικού. Πιο συγκεκριμένα, στα πρωτεύοντα θηλαστικά ο μιτοχονδριακός ρυθμός εξέλιξης είναι περίπου 0.02 αντικαταστάσεις ανά ζεύγος βάσεων, ανά εκατομμύριο χρόνια, ενώ ο πυρηνικός είναι μία τάξη μεγέθους μικρότερος<sup>15</sup>. Ο “ταχύς” αυτός ρυθμός εξέλιξης, έχει αποδοθεί στον μεγάλο αριθμό ζημιών που υπόκεινται τα μιτοχόνδρια, σαν αποτέλεσμα της παραγωγής ελεύθερων ριζών οξυγόνου κατά τη διαδικασία της οξειδωτικής φωσφορυλίωσης, στη μειωμένη αποτελεσματικότητα των επιδιορθωτικών μηχανισμών του mtDNA, και στον μεγαλύτερο ρυθμό αντιγραφής του σε σχέση με το πυρηνικό DNA. Αν και τα μιτοχόνδρια διαθέτουν πλήθος ενζύμων τα οποία αδρανοποιούν τις ρίζες αυτές, παραγωγή μεγάλης ποσότητας τέτοιων ελευθέρων ριζών οδηγεί σε οξειδωτικό στρες, με αποτέλεσμα (μεταξύ άλλων) βλάβες στο γονιδίωμα. Αδιόρθωτες βλάβες κατά την αντιγραφή, ή προβλήματα κατά την επισκευή του DNA, μπορούν να οδηγήσουν σε μεταλλάξεις. Φυσικά, μόνο μεταλλάξεις στα γαμετοκύτταρα μπορούν να περάσουν στις επόμενες γενιές.

Τα πρωτεϊνικά σύμπλοκα που εμπλέκονται στην οξειδωτική φωσφορυλίωση, είναι (με εξαίρεση το Σύμπλοκο II) προϊόντα αλληλεπίδρασης πρωτεϊνών που είτε κατασκευάζονται, είτε στοχεύονται στο μιτοχόνδριο. Πειράματα μεταφοράς μιτοχονδρίων μεταξύ συγγενών ειδών, και μέτρηση της ενεργότητας των συμπλόκων οξειδωτικής φωσφορυλίωσης, έχουν δείξει πως υπάρχει συνεξέλιξη μεταξύ μιτοχονδριακών και πυρηνικών πρωτεϊνών. Για

παράδειγμα, σε κύτταρα *Homo sapiens* με mtDNA *Pan troglodytes* ή *Gorilla gorilla*, παρατηρήθηκε μειωμένη απόδοση του συμπλόκου I της τάξης του 40%<sup>32</sup>. Αντίθετα, κύτταρα *Homo sapiens* με mtDNA *Pongo pygmaeus* (ο οποίος έχει πιο παλιό τελευταίο κοινό πρόγονο με τους ανθρώπους απ' ότι ο χιμπατζής ή ο γορίλλας), δεν κατάφεραν να μεγαλώσουν σε εργαστηριακές συνθήκες, υποδηλώνοντας ασυμβατότητα σε τουλάχιστον ένα από τα σύμπλοκα OXPHOS<sup>33</sup>. Παρόμοια πειράματα ανάμεσα σε κύτταρα *Mus musculus* και *Rattus norvegicus*, έδειξαν ασυμβατότητες σε παραπάνω από ένα σύμπλοκα OXPHOS<sup>34</sup>.



Εικόνα 9: Οι υπομονάδες των συμπλόκων OXPHOS

## 2.2 Σκοπός της μελέτης

Η παρούσα εργασία είχε ως στόχο να μελετήσει την εξελικτική συμπεριφορά των πρωτεϊνών που συμμετέχουν στο μονοπάτι της οξειδωτικής φωσφορυλίωσης, και που είτε κατασκευάζονται στο μιτοχόνδριο από μιτοχονδριακό DNA, είτε κατασκευάζονται στο κυτταρόπλασμα από πυρηνικό DNA, και κατόπιν στοχεύονται στο μιτοχόνδριο. Δεδομένης της συνεργασίας μεταξύ πυρηνικών και μιτοχονδριακών πρωτεϊνών στα σύμπλοκα της αναπνοής, αλλά και της διαφορετικής μεταλλακτικής πίεσης στα δύο γονιδιώματα που κωδικοποιούν αυτές τις πρωτεΐνες, τίθεται το ερώτημα με ποιον τρόπο επιτελείται η συνεργασία αυτή. Επίσης, τέθηκε το ερώτημα, αν συγκεκριμένα χαρακτηριστικά των πρωτεϊνών αυτών, όπως το μέγεθος, ο αριθμός των αλληλεπιδράσεων ή ο εξελικτικός τους ρυθμός, έχουν στατιστικά σημαντικές συσχετίσεις μεταξύ τους.

Στα πλαίσια της ανάλυσης αυτής, χρησιμοποιήθηκαν τρία ζεύγη οργανισμών:

- ***Homo sapiens* – *Pan troglodytes***
- ***Drosophila melanogaster* – *Drosophila simulans***
- ***Zea mays* – *Oryza sativa***

Τα δυο πρώτα ζεύγη οργανισμών που επιλέχθηκαν είναι αρκετά κοντά μεταξύ τους φυλογενετικά, έτσι ώστε να ελαχιστοποιηθούν οι πιθανότητες πολλαπλών αντικαταστάσεων ανάμεσα στις αλληλουχίες τους. Ο τελευταίος κοινός πρόγονος ανάμεσα στους *H.sapiens* και *P.troglodytes* έχει υπολογιστεί στα 4 με 6 εκατομμύρια χρόνια<sup>35,36</sup>, ενώ ο αντίστοιχος των *D.melanogaster* και *D.simulans* σε περίπου 3.4 εκατομμύρια χρόνια<sup>37</sup>. Είναι γνωστό πως στα φυτικά γονιδιώματα, αντίθετα με ό,τι ισχύει στα ζωικά, ο ρυθμός εξέλιξης των mtDNA τους είναι μια τάξη μεγέθους μικρότερος από τον nDNA. Έτσι, το τρίτο ζεύγος σύγκρισης πάρθηκε για να διαπιστωθεί εάν η συμπεριφορά των υπό μελέτη πρωτεϊνών επεκτείνεται και πέρα από το ζωικό βασίλειο. Αν και τα *Z.mays* και *O.sativa* είναι πιο απομακρυσμένα φυλογενετικά (παραπάνω από 50 εκατομμύρια χρόνια)<sup>38</sup> σε σχέση με τα αντίστοιχα ζεύγη ζωικών γονιδιωμάτων, κατά τη διάρκεια εκπόνησης της εργασίας ήταν από τα λίγα φυτικά είδη με αλληλουχημένα μιτοχονδριακά γονιδιώματα.

### 2.3 Υλικά και μέθοδοι

Οι παρακάτω βάσεις δεδομένων χρησιμοποιήθηκαν για τη συλλογή των υπό μελέτη γονιδίων:

1. NCBI GenBank  
(<http://www.ncbi.nlm.nih.gov/Genbank>)
2. KEGG GENES  
(<http://www.genome.jp/kegg/genes.html>)
3. ORGANELLE  
(<http://www.shigen.nig.ac.jp/organelle/index.jsp>)

Στον Πίνακα 1 φαίνονται τα σύμπλοκα στα οποία ανήκουν τα γονίδια που συλλέχθηκαν. Εκτός από τα πέντε OXPHOS σύμπλοκα, συλλέχθηκαν επίσης γονίδια τα οποία εμπλέκονται στο μονοπάτι της γλυκόλυσης. Τα γονίδια της γλυκόλυσης χρησιμοποιήθηκαν σαν ένα είδος control, δεδομένου ότι και αυτά σχετίζονται λειτουργικά με την διαδικασία της αναπνοής, χωρίς ωστόσο να στοχεύονται στο μιτοχόνδριο, και χωρίς να αλληλεπιδρούν με μιτοχονδριακά γονίδια.

**Πίνακας 1:** Αριθμός γονιδίων ανά σύμπλοκο. Στα αριστερά φαίνεται το σύμπλοκο, και το εάν περιέχει πυρηνικά (n) ή μιτοχονδριακά (mt) γονίδια, και ο αριθμός τους για το κάθε ζεύγος οργανισμών.

Complex	<i>H.sapiens</i> - <i>P.troglodytes</i>	<i>D.melanogaster</i> - <i>D.simulans</i>	<i>Z.mays</i> - <i>O.sativa</i>
<b>OXPHOS C. I (n, mt)</b>	31 - 7	18 - 7	15 - 9
<b>OXPHOS C.III (n, mt)</b>	9 - 1	5 - 1	2 - 1
<b>OXPHOS C.IV (n, mt)</b>	12 - 3	10 - 3	4 - 7
<b>OXPHOS C.V (n, mt)</b>	15 - 2	8 - 2	5 - 2
<b>OXPHOS C.II (n, mt)</b>	4 - 0	3 - 0	2 - 0
<b>GLYCOLYSIS (n)</b>	54	20	13

Για κάθε ζεύγος οργανισμών, έγινε η στίχοιση των ομόλογων αλληλουχιών με το πρόγραμμα ClustalW<sup>39</sup> αφού έγινε η μετάφραση του κάθε γονιδίου στην αντίστοιχη πρωτεΐνη ώστε να στοιχηθούν σωστά τα κωδικόνια. Οι στοιχήσεις ελέγχθηκαν "με το μάτι" σε νουκλεοτιδικό επίπεδο, προκειμένου

να διορθωθούν προβλήματα όπως ασαφείς (ambiguous) βάσεις, RNA editing (στα φυτικά γονίδια), και να αφαιρεθούν κωδικωνία λήξης και περιοχές χαμηλής ταυτότητας. Τα γονίδια που δεν παρουσίαζαν ικανοποιητικό βαθμό στοίχισης, ή που είχαν λανθασμένα annotations στις βάσεις δεδομένων, αφαιρούνταν από την ανάλυση. Όλα τα παραπάνω έγιναν στη σουίτα προγραμμάτων γενετικής ανάλυσης MEGA4<sup>40</sup>. Στους Συμπληρωματικούς Πίνακες 1-3, περιγράφονται στη στήλη Comment οι διορθώσεις που έγιναν στην κάθε αλληλουχία. Για κάθε γονίδιο, υπολογίστηκαν τα dN, dS και dN/dS.

Για τον υπολογισμό των τιμών dN, dS, dN/dS, χρησιμοποιήθηκε το πρόγραμμα YN00, από τη σουίτα προγραμμάτων φυλογενετικής ανάλυσης PAML<sup>41</sup>, τροποποιημένο για διαδικτυακή χρήση, από την ιστοσελίδα διαδικτυακών εργαλείων μοριακής εξέλιξης Phylemon (<http://phylemon.bioinfo.cipf.es>)<sup>42</sup>. Το YN00 χρησιμοποιεί διαφορετικούς αλγόριθμους υπολογισμού των τιμών αυτών, όπως το μοντέλο διόρθωσης p-distance από Nei & Gojobori<sup>43</sup>, ή το μοντέλο διόρθωσης των Yang & Nielsen (2000)<sup>44</sup>. Το δεύτερο, θεωρείται πιο ακριβές, επειδή λαμβάνει υπόψιν τους διαφορετικούς ρυθμούς μεταθέσεων/μεταστροφών, τη σύσταση των αλληλουχιών, καθώς και το codon usage bias. Ο αλγόριθμος του μοντέλου υπολογίζει τις συχνότητες των κωδικονίων βάσει της συχνότητας κάθε νουκλεοτιδίου σε κάθε θέση του κωδικονίου ξεχωριστά, κατά μήκος της αλληλουχίας (αναφέρεται στην βιβλιογραφία σαν μοντέλο F3x4, δηλ. 3 διαφορετικές θέσεις, 4 διαφορετικά νουκλεοτίδια για την κάθε θέση)<sup>45</sup>. Ωστόσο, επειδή στα μιτοχονδριακά γονίδια, δεν παρατηρείται ο συνηθισμένος λόγος μεταπτώσεων/μεταστροφών που υπάρχει στο πυρηνικό DNA, τα αποτελέσματα αυτού του αλγορίθμου δεν ήταν αξιόπιστα, και τελικά χρησιμοποιήθηκε το απλό μοντέλο των N & G. Για τα πυρηνικά γονίδια, και τα δυο μοντέλα έδωσαν σχεδόν πανομοιότυπα αποτελέσματα.

Τέλος, υπολογίστηκαν οι μέσες τιμές dN, dS, dN/dS για κάθε σύμπλοκο, βάσει των επιμέρους γονιδίων από τα οποία αυτό αποτελείται.

Για τον προσδιορισμό του αριθμού των αλληλεπιδράσεων στις οποίες λαμβάνει μέρος κάθε πρωτεΐνη, χρησιμοποιήθηκε η βάση δεδομένων STRING 8.2 (<http://string-db.org>)<sup>46</sup>, η οποία είναι μια συγκεντρωτική βάση πολλών επιμέρους βάσεων Protein-Protein Interaction (PPI). Ο αλγόριθμος της βάσης ελέγχει κάποιες παραμέτρους, για να προσδιορίσει εάν δυο πρωτεΐνες αλληλεπιδρούν. Τέτοιες παράμετροι είναι για παράδειγμα, το εάν βρίσκονται τα γονίδια τους στον ίδιο γενετικό τόπο, εάν συνεκφράζονται στους ίδιους ιστούς, και φυσικά, εάν έχει αποδειχθεί πειραματικά η αλληλεπίδρασή τους. Από τις παραμέτρους αυτές, κατασκευάζεται ένα confidence score ( $0 \leq c.s. \leq 1$ ) για το πόσο πιθανό είναι οι δυο αυτές πρωτεΐνες να αλληλεπιδρούν). Στη μελέτη μας χρησιμοποιήθηκαν confidence scores  $\geq 0.9$ , και μόνο πειραματικά αποδεδειγμένες αλληλεπιδράσεις.



## 2.4 Αποτελέσματα - Συζήτηση

Στους Πίνακες 2, 3 και 4 παρουσιάζονται οι τιμές dN, dS και ο λόγος dN/dS για τα ζεύγη γονιδίων *H. sapiens* - *P. troglodytes*, *D. melanogaster* - *D. simulans* και *Z. mays* - *O. sativa* αντίστοιχα. Τα γονίδια χωρίζονται σε τρεις κατηγορίες. Πρώτον, τα γονίδια που κωδικοποιούνται από μιτοχονδριακό DNA και συμμετέχουν στα σύμπλοκα της οξειδωτικής φωσφορυλίωσης (στους πίνακες συμβολίζονται ως mt). Ο αριθμός των γονιδίων αυτών δεν είναι σταθερός για κάθε σύμπλοκο, για παράδειγμα στο Σύμπλοκο I συμμετέχουν 5 μιτοχονδριακά γονίδια, στο Σύμπλοκο III συμμετεχει μόνο ένα και στο Σύμπλοκο II κανένα μιτοχονδριακό γονίδιο (Πίνακας 1). Δεύτερον, γονίδια που συμμετέχουν στα σύμπλοκα της οξειδωτικής φωσφορυλίωσης αλλά κατασκευάζονται από πυρηνικό DNA (στους πίνακες συμβολίζονται ως (n)). Κι εδώ ο αριθμός των πυρηνικών γονιδίων που συμμετέχουν στα σύμπλοκα της αναπνοής διαφέρει από σύμπλοκο σε σύμπλοκο. Τρίτον, γονίδια τα οποία κωδικοποιούνται στο πυρηνικό DNA, συμμετέχουν σε σύμπλοκο της αναπνοής αλλά δεν έχουν επαφή με το μιτοχόνδριο. Αυτά είναι τα γονίδια του συμπλόκου της γλυκόλυσης. Όπως αναφέρθηκε στην εισαγωγή, τα γονίδια αυτά χρησιμοποιήθηκαν σαν control, επειδή, όπως και τα πυρηνικά γονίδια των συμπλόκων OXPHOS, κατασκευάζονται από το πυρηνικό γονιδίωμα, συμμετέχουν στη διαδικασία της αναπνοής, αλλά δεν αλληλεπιδρούν με γονίδια που κατασκευάζονται από μιτοχονδριακό γονιδίωμα, και ούτε εισέρχονται στο μιτοχόνδριο.

Ο λόγος dN/dS δείχνει την ύπαρξη ή όχι φυσικής επιλογής στα συγκεκριμένα γονίδια καθώς και την ένταση της φυσικής επιλογής. Οι μέσοι όροι του λόγου αυτού για τα γονίδια όλων των συμπλόκων που συμμετέχουν στην αναπνοή, μιτοχονδριακών και πυρηνικών είναι μικρότεροι της μονάδας (Πίνακες 2,3,4). Αυτό είναι ένδειξη αρνητικής επιλογής, η οποία είναι αναμενόμενη σε γονίδια γενικής έκφρασης<sup>47</sup>. Επειδή οι λειτουργίες που επιτελούν οι πρωτεΐνες που κωδικοποιούνται από αυτά τα γονίδια είναι βασικές σε όλους τους οργανισμούς, η αρνητική επιλογή «δεν επιτρέπει» εύκολα αλλαγές στη δομή αυτών των γονιδίων.

**Πίνακας 2:** Τιμές για *H.sapiens* - *P.troglodytes*

	<i>OXPHOS C. I (mt)</i>		<i>OXPHOS C.III (mt)</i>	<i>OXPHOS C.IV (mt)</i>	<i>OXPHOS C.V (mt)</i>
dN	0,0214		0,0360	0,0093	0,0276
dS	0,4159		0,4829	0,4592	0,3354
dN/dS	0,0514		0,0745	0,0203	0,0822
	<i>OXPHOS C. I (n)</i>	<i>OXPHOS C.II (n)</i>	<i>OXPHOS C.III (n)</i>	<i>OXPHOS C.IV (n)</i>	<i>OXPHOS C.V (n)</i>
dN	0,0021	0,0019	0,0084	0,0072	0,0027
dS	0,0117	0,0070	0,0252	0,0133	0,0178
dN/dS	0,1805	0,2714	0,3330	0,5440	0,1519
	<i>GLYCOLYSIS (n)</i>				
dN	0,0021				
dS	0,0164				
dN/dS	0,1269				

Για το ζεύγος οργανισμών *Homo sapiens* - *Pan troglodytes* (Πίνακας 2), το dS όλων των μιτοχονδριακών OXPHOS συμπλόκων είναι της ίδιας τάξης μεγέθους, περίπου  $10^{-1}$ . Η τιμή αυτή σύμφωνα με τη θεωρία, εκφράζει την πίεση που ασκείται στις αλληλουχίες από τη μετάλλαξη. Ο λόγος dN/dS (ο

οποίος αποτελεί και το μέτρο της έντασης της φυσικής επιλογής) είναι και αυτός της ίδιας τάξης μεγέθους,  $10^{-2}$  για τα μιτοχονδριακά γονίδια όλων των συμπλόκων. Σε όλα τα σύμπλοκα φαίνεται πολύ ισχυρή η ένταση της αρνητικής επιλογής. Τα πυρηνικά γονίδια των OXPPOS συμπλόκων έχουν μεταξύ τους παρόμοιες τιμές dS της τάξης του  $10^{-2}$ . Εξαιρέση αποτελεί το σύμπλοκο II, στο οποίο το dS είναι ελαφρά μικρότερο (0,007). Αντίστοιχα το dS για τα γονίδια της γλυκόλυσης είναι 0,0164 και βρίσκεται μέσα στο εύρος των τιμών του dS για τα πυρηνικά γονίδια (μέση τιμή για όλα τα πυρηνικά γονίδια της αναπνοής είναι 0.0149). Ο λόγος dN/dS κυμαίνεται μεταξύ του 0,1805 και 0,5440. Ο λόγος αυτός είναι μικρότερος της μονάδας που σημαίνει την επίδραση αρνητικής επιλογής στα γονίδια αυτά. Η μέση τιμή του λόγου dN/dS για τα γονίδια της γλυκόλυσης είναι 0,1269. Ο λόγος αυτός είναι ελαφρώς μικρότερος από (αλλά πάντως της ίδιας τάξης μεγέθους με) την κατώτερη τιμή που παρατηρείται στα πυρηνικά γονίδια των συμπλόκων της αναπνοής (η κατώτερη τιμή είναι 0,1805). Συγκρίνοντας τις μέσες τιμές dS των γονιδίων που κωδικοποιούνται από το mtDNA (0.1487) με τα γονίδια που κωδικοποιούνται από το πυρηνικό (συμπεριλαμβανομένων των γονιδίων των συμπλόκων της αναπνοής και αυτών της γλυκόλυσης) παρατηρείται ότι υπάρχει διαφορά μιας τάξης μεγέθους. Δηλαδή ότι τα μιτοχονδριακά γονίδια εξελίσσονται στις συνώνυμες θέσεις τους κατά 10 περίπου φορές ταχύτερα σε σχέση με τα αντίστοιχα πυρηνικά. Αυτή η διαφορά αντικατοπτρίζει τη διαφορά στο μεταλλακτικό ρυθμό των δύο γονιδιωμάτων η οποία υποστηρίζεται από πλήθος δεδομένων<sup>15,16,17</sup>. Αντίθετη είναι η εικόνα που δίνουν οι τιμές του λόγου dN/dS. Τα γονίδια που κωδικοποιούνται από το mtDNA εμφανίζουν συστηματικά μικρότερο λόγο σε σχέση με τα πυρηνικά γονίδια (συμπεριλαμβανομένων και αυτών της γλυκόλυσης). Αυτό φανερώνει πώς η φυσική επιλογή είτε είναι αυστηρότερη στα μιτοχονδριακά γονίδια, είτε είναι χαλαρότερη στα πυρηνικά είτε και τα δύο μαζί.

**Πίνακας 3:** Τιμές για *D.melanogaster* - *D.simulans*

	<b>OXPPOS C. I (mt)</b>		<b>OXPPOS C.III (mt)</b>	<b>OXPPOS C.IV (mt)</b>	<b>OXPPOS C.V (mt)</b>
dN	0,0158		0,0068	0,0073	0,0191
dS	0,1741		0,3039	0,2094	0,1634
dN/dS	0,0909		0,0224	0,0349	0,1166
	<b>OXPPOS C. I (n)</b>	<b>OXPPOS C.II (n)</b>	<b>OXPPOS C.III (n)</b>	<b>OXPPOS C.IV (n)</b>	<b>OXPPOS C.V (n)</b>
dN	0,0063	0,0043	0,0051	0,0106	0,0043
dS	0,1044	0,0860	0,0938	0,1038	0,0734
dN/dS	0,0605	0,0504	0,0541	0,1017	0,0584
			<b>GLYCOLYSIS (n)</b>		
dN			0,0062		
dS			0,1151		
dN/dS			0,0534		

Για το ζεύγος οργανισμών *D. melanogaster* – *D. simulans* (Πίνακας 3), τα μιτοχονδριακά γονίδια των συμπλόκων της αναπνοής έχουν dS από 0,1634 έως 0,3039. Ο λόγος dN/dS σε αυτά τα γονίδια κυμαίνεται από 0.0224 έως 0,1166. Στα πυρηνικά γονίδια (συμπεριλαμβανομένων και αυτών της γλυκόλυσης) οι τιμές dS είναι από 0,0734 έως 0,1044, ενώ ο λόγος dN/dS κυμαίνεται από 0,0504 έως 0,1017. Κατά μέσο όρο η τιμή dS στα γονίδια που κωδικοποιούνται από το mtDNA είναι σημαντικά μεγαλύτερη σε σχέση με τα αντίστοιχα γονίδια που κωδικοποιούνται από το πυρηνικό DNA (0.1906 και 0.0962 αντίστοιχα) πράγμα που υποδηλώνει την αυξημένη μεταλλακτική πίεση

στα μιτοχονδριακά γονίδια. Αυτή η παρατήρηση συμφωνεί με άμεσες και έμμεσες μελέτες που δείχνουν τον αυξημένο μεταλλακτικό ρυθμό στο mtDNA<sup>15,16,17</sup>. Αντίθετα κατά μέσο όρο ο λόγος dN/dS είναι παρόμοιος μεταξύ των γονιδίων του mtDNA και των πυρηνικών (0,0717 και 0,0690 αντίστοιχα).

**Πίνακας 4:** Τιμές για *Z.mays* - *O.sativa*

	<i>OXPHOS C. I (mt)</i>		<i>OXPHOS C.III (mt)</i>		<i>OXPHOS C.IV (mt)</i>	<i>OXPHOS C.V (mt)</i>
dN	0,0066		0,0045		0,0044	0,0148
dS	0,0213		0,0221		0,0153	0,0522
dN/dS	0,3103		0,2036		0,2892	0,2828
	<i>OXPHOS C. I (n)</i>	<i>OXPHOS C.II (n)</i>	<i>OXPHOS C.III (n)</i>	<i>OXPHOS C.IV (n)</i>	<i>OXPHOS C.V (n)</i>	
dN	0,0458	0,0275	0,0312	0,0431	0,0588	
dS	0,4934	0,3967	0,7925	0,5766	0,5132	
dN/dS	0,0928	0,0693	0,0394	0,0747	0,1145	
			<i>GLYCOLYSIS (n)</i>			
dN			0,0535			
dS			0,5289			
dN/dS			0,1012			

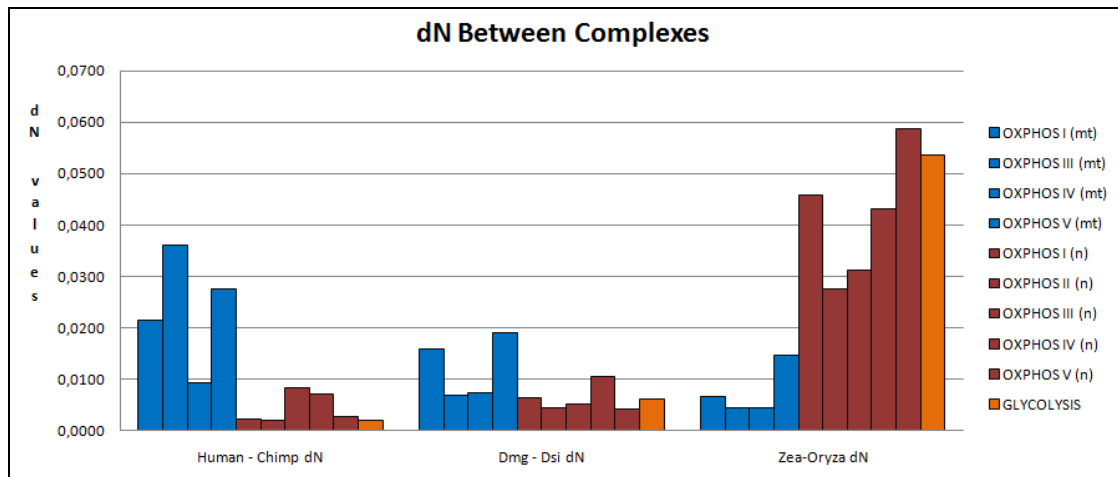
Για το ζεύγος οργανισμών *Zea mays* – *Oryza sativa* (Πίνακας 4), οι τιμές dS για τα γονίδια που κωδικοποιούνται από το mtDNA κυμαίνονται από 0,0153 έως 0,0522. Για τα αντίστοιχα πυρηνικά οι τιμές dS κυμαίνονται μεταξύ 0,3967 και 0,7925 (σε αυτό το εύρος συμπεριλαμβάνονται και τα γονίδια για τη γλυκόλυση). Οι λόγοι dN/dS είναι μεταξύ 0,2036 και 0,3103 για τα μιτοχονδριακά, ενώ για τα αντίστοιχα πυρηνικά (συμπεριλαμβανομένων και αυτών της γλυκόλυσης) είναι 0,0394 και 0,1145. Στο ζεύγος των φυτών οι μέσες τιμές dS για τα γονίδια που κωδικοποιούνται από το mtDNA είναι περίπου μια τάξη μεγέθους μικρότερες από τις αντίστοιχες τιμές των γονιδίων που κωδικοποιούνται από το πυρηνικό DNA (0,0224 και 0,5233 αντίστοιχα). Αυτή η παρατήρηση συμφωνεί με άλλες μελέτες που αναφέρουν βραδύτερη εξέλιξη των μιτοχονδριακών γονιδίων στα φυτά σε σχέση με τα ζώα<sup>18</sup>. Οι τιμές για το λόγο dN/dS στα πυρηνικά είναι μικρότερες κατά μια τάξη μεγέθους σε σχέση με τα μιτοχονδριακά γονίδια. Το πρότυπο αυτό των μικρότερων τιμών του λόγου dN/dS και των μεγαλύτερων τιμών dS στα μιτοχονδριακά σε σχέση με τα πυρηνικά είναι το αντίθετο με αυτό που παρατηρείται στο ζεύγος ανθρώπου-χιμπατζή.

Σύμφωνα με τη θεωρία της ουδετερότητας που εισήγαγε ο Kimura, στις ουδέτερες αλληλουχίες ο ρυθμός εξέλιξης (δηλαδή ο ρυθμός με τον οποίο αλλάζουν οι αλληλουχίες) εξαρτάται αποκλειστικά από τον μεταλλακτικό ρυθμό<sup>48</sup>. Αν θεωρήσουμε ότι οι συνώνυμες νουκλεοτιδικές θέσεις είναι ουδέτερες τότε μεταβολές στο ρυθμό αντικαταστάσεων στις συνώνυμες θέσεις εκφράζουν μεταβολές στο μεταλλακτικό ρυθμό των αλληλουχιών. Ενώ όμως ο ρυθμός εξέλιξης στις ουδέτερες θέσεις εξαρτάται από το μεταλλακτικό ρυθμό, η ποσότητα της διαφοράς που συσσωρεύεται μεταξύ δύο αλληλουχιών εξαρτάται τόσο από το ρυθμό εξέλιξης όσο και από το χρόνο που έχει μεσολαβήσει από τον διαχωρισμό των αλληλουχιών αυτών από τον κοινό τους πρόγονο. Αν οι μεταλλακτικοί ρυθμοί στα τρία ζεύγη γονιδιωμάτων ήταν ίδιοι, οι τιμές dS του ζεύγους των δροσοφιλών θα έπρεπε να είναι μικρότερες σε σχέση με αυτές του ζεύγους ανθρώπου-χιμπατζή και αυτές μικρότερες από τις τιμές του ζεύγους των φυτών διότι ο χρόνος που μεσολάβησε από τον διαχωρισμό των ζευγών αυτών από τον κοινό τους πρόγονο είναι μικρότερος

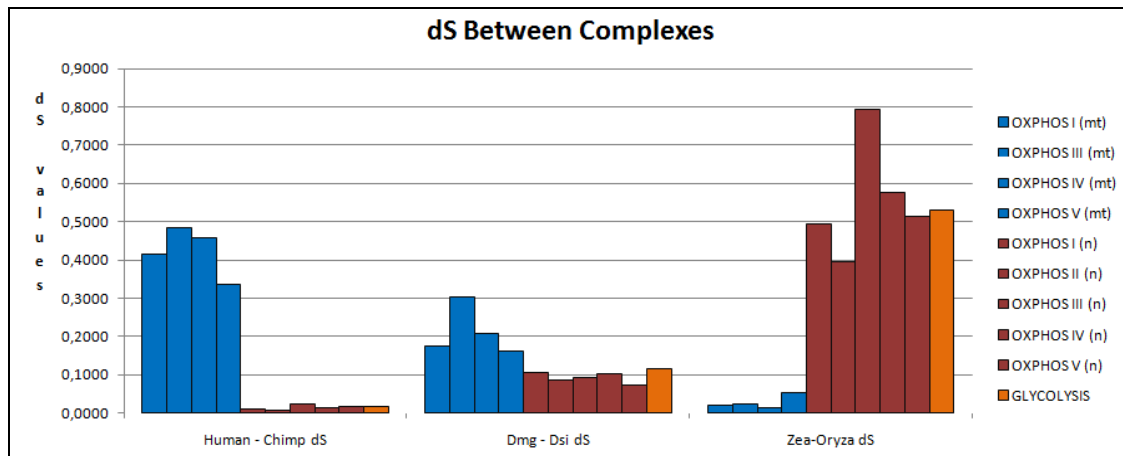
στις δροσόφιλες σε σχέση με αυτόν του ζεύγους ανθρώπου-χιμπατζή και αυτός μικρότερος με το χρόνο που μεσολάβησε από το διαχωρισμό των φυτών<sup>35,36,37,38</sup>. Στη Εικόνα 12 εμφανίζεται διαφορετικό πρότυπο για τις τιμές dS τόσο των μιτοχονδριακών όσο και των πυρηνικών γονιδίων. Μία πιθανή εξήγηση του μεγαλύτερου μεταλλακτικού ρυθμού στα πυρηνικά γονίδια της δροσόφιλας, είναι ο κατά πολύ μικρότερος χρόνος γενεάς τους σε σχέση με τα θηλαστικά. Τα γονιδιώματά τους να αντιγράφονται πολύ περισσότερες φορές και αυτό ίσως να οδηγεί σε αυξημένα ποσοστά μεταλλάξεων. Ωστόσο, εάν ίσχυε αυτό θα έπρεπε να παρατηρείται κάτι αντίστοιχο και στα μιτοχονδριακά γονίδια. Εναλλακτικά, είναι πιθανόν τα θηλαστικά να διαθέτουν πιο εξειδικευμένους μηχανισμούς αντιγραφής και διόρθωσης λαθών στον πυρήνα των κυττάρων τους (αλλά όχι στο μιτοχόνδριο) σε σχέση με τα αρθρόποδα, και για αυτόν τον λόγο οι πυρηνικοί μεταλλακτικοί τους ρυθμοί να είναι πιο χαμηλοί. Μια ακόμα εξήγηση, είναι ότι ο πυρηνικός μεταλλακτικός ρυθμός είναι ο αναμενόμενος εάν λάβουμε υπόψη τον χρόνο γενεάς, ενώ ο μιτοχονδριακός ο οποίος θα έπρεπε να είναι αρκετά μεγαλύτερος, δεν είναι. Προηγούμενες μελέτες έχουν επίσης επισημάνει χαμηλότερους από τους αναμενόμενους ρυθμούς αντικατάστασης στα μιτοχονδριακά γονιδιώματα των ψυχρόαιμων ζώων<sup>49</sup>. Τέλος, ο αρκετά χαμηλός μιτοχονδριακός μεταλλακτικός ρυθμός του ζεύγους των φυτών σε σχέση με τον χρόνο απόκλισής τους, είναι πιθανόν επίσης να οφείλεται στα εξειδικευμένα σύμπλοκα επισκευής DNA τα οποία υπάρχουν μόνο στα μιτοχόνδρια των φυτών (βλέπε Εισαγωγή).

Η φυσική επιλογή αναμένεται να δρα κυρίως στις μη συνώνυμες θέσεις (δες Εισαγωγή). Επομένως οι τιμές dN (Εικόνα 11) θα έπρεπε να εκφράζουν την ένταση της επιλογής. Όμως οι τιμές αυτές δεν δίνουν αντικειμενική εικόνα της φυσικής επιλογής διότι πέρα από αυτήν επηρεάζονται και από το μεταλλακτικό ρυθμό. Γι' αυτό το λόγο ως μέτρο της έντασης της φυσικής επιλογής χρησιμοποιείται ο λόγος dN/dS ο οποίος κανονικοποιεί (normalizes) την ένταση της επιλογής σε σχέση με το μεταλλακτικό ρυθμό κάθε αλληλουχίας. Στην εικόνα 13 παρουσιάζονται συγκριτικά οι τιμές του λόγου αυτού στα μιτοχονδριακά και πυρηνικά κωδικοποιούμενα γονίδια στα τρία ζεύγη οργανισμών. Στο ζεύγος των δροσοφιλών, η ένταση είναι περίπου η ίδια σε όλα τα σύμπλοκα, κάτι το οποίο θα μπορούσε να εξηγηθεί από την αλληλεπίδραση των μιτοχονδριακών-πυρηνικών υπομονάδων η οποία οδηγεί σε συνεξέλιξη, και από το ότι έχουν πολύ πρόσφατο τελευταίο κοινό πρόγονο.

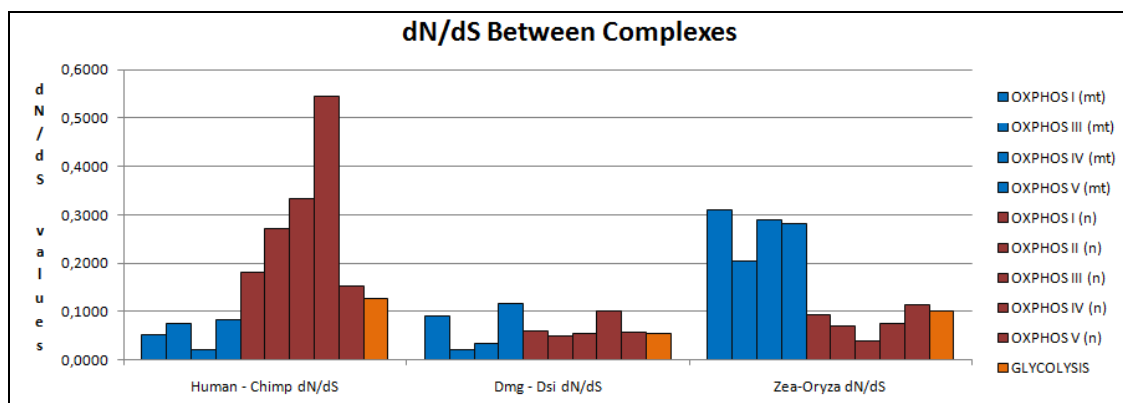
**Δεδομένου ότι στα πρωτεύοντα, η ένταση της επιλογής στα πυρηνικά γονίδια των περισσότερων OXPHOS συμπλόκων είναι πιο χαλαρή από την ένταση της επιλογής στα γονίδια της γλυκόλυσης, φαίνεται πως η χαλάρωση αυτή είναι και ο τρόπος με τον οποίο τα πυρηνικά γονίδια καταφέρνουν να συνεξελίσσονται με τα μιτοχονδριακά, τα οποία έχουν πολύ μεγαλύτερο ρυθμό αντικατάστασης. Αντίθετα, στα φυτά, η ένταση της επιλογής στα μιτοχονδριακά γονίδια είναι πιο χαλαρή από την ένταση της επιλογής στα πυρηνικά γονίδια (η οποία είναι παρόμοια με αυτή των γονιδίων της γλυκόλυσης), και έτσι τα μιτοχονδριακά καταφέρνουν να συνεξελίσσονται με τα πυρηνικά, τα οποία έχουν πολύ μεγαλύτερο ρυθμό αντικατάστασης.**



**Εικόνα 10:** Διαφορετικά dN ανά σύμπλοκο



**Εικόνα 11:** Διαφορετικά dS ανά σύμπλοκο



**Εικόνα 12:** Διαφορετικά ω ανά σύμπλοκο

Το αρχικό μας ερώτημα ήταν με ποιο τρόπο καταφέρνουν να συνεργάστουν πρωτεΐνες που κωδικοποιούνται από διαφορετικά γενετικά περιβάλλοντα για να επιτελέσουν μια βασική κυτταρική λειτουργία όπως είναι η αναπνοή. Τα δεδομένα μας, επιβεβαιώνοντας προηγούμενες παρατηρήσεις, δείχνουν πως οι ρυθμοί εξέλιξης στις συνώνυμες θέσεις είναι διαφορετικοί μεταξύ των γονιδίων που κωδικοποιούνται από πυρηνικά και μιτοχονδριακά

γονίδια. Για να μπορέσει να διατηρηθεί η συνεργασία μεταξύ των γονιδίων στα σύμπλοκα της αναπνοής θα πρέπει η φυσική επιλογή να «αναλάβει» να εξωμαλύνει τις διαφορές που προκαλούν οι διαφορετικοί μεταλλακτικοί ρυθμοί των διαφορετικών γενετικών υποβάθρων. Αυτό θα μπορούσε να γίνει με δύο τρόπους; είτε αυξάνοντας την ένταση της επιλογής στα γονιδιώματα που μεταλλάσσονται ταχύτερα είτε χαλαρώνοντας την ένταση της επιλογής στα γονιδιώματα που μεταλλάσσονται πιο αργά. Πού ασκείται όμως ισχυρότερη επιλογή για να επιτευχθεί η συνεργασία, στα πυρηνικά ή στα μιτοχονδριακά κωδικοποιούμενα γονίδια; Σαν έλεγχο για αυτή την υπόθεση χρησιμοποιήσαμε τα γονίδια του συμπλόκου της γλυκόλυσης. Τα γονίδια αυτά παρουσιάζουν κοινά χαρακτηριστικά με τα πυρηνικά γονίδια των συμπλόκων της αναπνοής. Πρώτον, κωδικοποιούνται από τον πυρήνα και παρόλο που οι μεταλλακτικές πιέσεις δεν είναι ομοιόμορφες στο πυρηνικό γονιδίωμα, οι τιμές dS των γονιδίων της γλυκόλυσης είναι πολύ κοντά σε αυτές των υπόλοιπων πυρηνικών γονιδίων των συμπλόκων της αναπνοής. Δεύτερον, οι πρωτεΐνες που κωδικοποιούν συμμετέχουν σε σύμπλοκα όπως και τα πυρηνικά γονίδια των συμπλόκων της αναπνοής. Τρίτον, συμμετέχουν και εκείνα στην αναπνοή. Το πλεονέκτημα είναι ότι παρόλο που έχουν τα παραπάνω κοινά χαρακτηριστικά, έχουν μια σημαντική διαφορά με αυτά διότι δεν έχουν επαφή με το μιτοχόνδριο (η γλυκόλυση είναι διαδικασία που επιτελείται στο κυτταρόπλασμα). Σε όλα τα ζεύγη οργανισμών που χρησιμοποιήσαμε φαίνεται πως τα πυρηνικά γονίδια των συμπλόκων της αναπνοής τόσο στις dS, όσο και στο λόγο dN/dS συμβαδίζουν με τις αντίστοιχες τιμές των γονιδίων της γλυκόλυσης και διαφέρουν από τα αντίστοιχα μιτοχονδριακά κωδικοποιούμενα γονίδια (οι διαφορές δεν είναι εμφανείς στο ζεύγος των δροσοφιλών- βλέπε παρακάτω). Αυτή η συμπίεση των πυρηνικών γονιδίων των συμπλόκων της αναπνοής με τα γονίδια της γλυκόλυσης είναι εμφανής και στα τρία ζεύγη των οργανισμών που χρησιμοποιήθηκαν (Πίνακες 11-13). Μάλιστα είναι περισσότερο εμφανής στο ζεύγος των θηλαστικών και στο ζεύγος των φυτών. Στο πρώτο, τα μιτοχονδριακά γονίδια έχουν ταχύτερους ρυθμούς εξέλιξης από τα πυρηνικά ενώ στο ζεύγος των φυτών συμβαίνει το αντίθετο. Όμως, και στις δύο περιπτώσεις τα πυρηνικά γονίδια της αναπνοής ακολουθούν τα υπόλοιπα πυρηνικά γονίδια του συμπλόκου της γλυκόλυσης. Επομένως με βάση αυτές τις παρατηρήσεις φαίνεται πως για να διατηρηθεί η συνεργασία πυρηνικών και μιτοχονδριακών γονιδίων στα σύμπλοκα της αναπνοής η φυσική επιλογή ασκείται κυρίως στα μιτοχονδριακά γονίδια όταν αυξάνεται η πίεση από τη μετάλλαξη όπως συμβαίνει στο mtDNA των ζώων (περισσότερες μεταλλάξεις και επομένως μεγαλύτερο dS) η επιλογή γίνεται αυστηρότερη και απομακρύνει τις επιβλαβείς μεταλλάξεις (μικρότερο dN/dS). Αντίθετα, όταν η πίεση από τη μετάλλαξη χαλαρώνει όπως συμβαίνει στο mtDNA των φυτών (λιγότερες μεταλλάξεις και επομένως μικρότερο dS), και η επιλογή χαλαρώνει (μεγαλύτερο dN/dS). Για πληρέστερο έλεγχο αυτών των δεδομένων θα ήταν χρήσιμη η σύγκριση των ρυθμών εξέλιξης γονιδίων που κωδικοποιούνται στο μιτοχόνδριο και στοχεύονται στον πυρήνα, τα οποία θα έπρεπε να ακολουθούν το μιτοχονδριακό ρυθμό εξέλιξης. Αυτό όμως είναι κάτι το οποίο δεν μπορεί να ελεγχθεί, στα μετάζωα τουλάχιστον, αφού αυτά δεν διαθέτουν τέτοιου είδους πρωτεΐνες.

Ένα πρόβλημα που παρουσιάστηκε κατά τον υπολογισμό των συνολικών τιμών για κάθε σύμπλοκο, ήταν οι ενίοτε μεγάλες διαφορές αποτελεσμάτων από γονίδιο σε γονίδιο. Για παράδειγμα, στο σύμπλοκο V Human-Chimp, το πυρηνικό γονίδιο *ATP5E* έχει τιμή  $dN=0.0279$ , ενώ τα περισσότερα άλλα έχουν τιμές κοντά στο 0. Ο μέσος όρος ανεβάζει το  $dN$  ratio του συμπλόκου σε 0.0029, ενώ απουσία αυτού του γονιδίου θα ήταν πολύ πιο κοντά στο 0. Τέτοια παραδείγματα υπάρχουν αρκετά στα δεδομένα μας (δες Παράρτημα).

Επειδή σε αρκετά από τα σύμπλοκα που μελετήθηκαν υπήρχε μικρός αριθμός πρωτεϊνικών αλληλουχιών, και μικρός αριθμός νουκλεοτιδικών αντικαταστάσεων, η τιμή  $dN/dS$  ήταν αρκετά ευαίσθητη σε μικρές αλλαγές τιμών, και έτσι διαφέρει αρκετά από πρωτεΐνη σε πρωτεΐνη και από σύμπλοκο σε σύμπλοκο. Σε περιπτώσεις όπου υπήρχαν μη συνώνυμες αλλά όχι συνώνυμες αντικαταστάσεις, η τιμή  $\omega$  δεν μπορεί να οριστεί (διαίρεση με 0), και όπου χρειάστηκε, βάλαμε απλά την τιμή " $>1$ ", ώστε να δείξουμε θετική επιλογή (κάποια προγράμματα δίνουν την τιμή 999 σε αυτή την περίπτωση, πχ PAML).

Όπως ήδη αναφέρθηκε, οι τιμές στους Πίνακες 2,3 και 4 αναφέρονται στους μέσους όρους των τιμών των γονιδίων. Από τη λίστα των γονιδίων που εξετάστηκαν έχει ενδιαφέρον να αναφερθεί μια ιδιαίτερη κατηγορία τα οποία παρουσιάζουν τιμές  $dN/dS$  μεγαλύτερες από 1 (Πίνακες 6 και 7) Φαίνεται ότι στα γονίδια αυτά έχει ασκηθεί θετική επιλογή. Τέτοια γονίδια βρέθηκαν μόνο στα ζεύγη *H.sapiens - P.troglodytes* και *Z.mays - O.sativa*. Όλα τα γονίδια της οξειδωτικής φωσφορυλίωσης είναι απολύτως αναγκαία για την αναπνοή, και ομολογικά τους βρίσκονται σε όλους τους ευκαρυωτικούς οργανισμούς. Και όπως φαίνεται (βλέπε Παράρτημα) στην πλειοψηφία τους είναι αρκετά συντηρημένα, και ασκείται ισχυρή φυσική επιλογή πάνω τους. Ωστόσο, μικρές αλλαγές στη λειτουργία και στην απόδοση των συμπλόκων, προκαλούν διαβαθμίσεις ανάμεσα στο ποσοστό παραγωγής ATP, και στο ποσοστό εκλυόμενης ενέργειας. Κάτι τέτοιο, είναι χρήσιμο στην προσαρμογή οργανισμών σε διαφορετικά περιβάλλοντα, όπου οι ανάγκες για παραγωγή και αποθήκευση ενέργειας μπορεί να διαφέρουν δραστικά<sup>50</sup>. Όμως δεν είμαστε σε θέση να εξηγήσουμε γιατί σε αυτά τα συγκεκριμένα γονίδια ασκήθηκε θετική επιλογή. Άλλες μελέτες έχουν ήδη βρει ίχνη θετικής επιλογής στα πρωτεΐοντα θηλαστικά για τα γονίδια *NDUFA2*<sup>51</sup>, *COX4*<sup>52</sup>, *COX5A*<sup>53</sup> και *UQCRH*<sup>54</sup>.

**Πίνακας 5:** Γονίδια *H.sapiens - P.troglodytes* με αυξημένο  $\omega$

<b>GENE NAME</b>	<b>dNS</b>	<b>dS</b>	<b>dNS/dS</b>	<b>LENGTH</b>	<b>NS POSITIONS</b>	<b>S POSITIONS</b>
<b>NDUFA2 (n)</b>	0,0131	0,0000	>1	207	160	47
<b>NDUFB2 (n)</b>	0,0084	0,0000	>1	315	229	86
<b>NDUFB9 (n)</b>	0,0024	0,0000	>1	537	380	157
<b>SDHD (n)</b>	0,0057	0,0000	>1	477	334	143
<b>COX4I1 (n)</b>	0,0076	0,0000	>1	513	418	95
<b>COX5B (n)</b>	0,0141	0,0000	>1	387	272	116
<b>COX17 (n)</b>	0,0208	0,0000	>1	189	136	53
<b>LDHAL6A</b>	0,0020	0,0000	>1	996	720	276
<b>UQCRH (n)</b>	0,0188	0,0176	1,0682	273	179	94
<b>COX5A (n)</b>	0,0090	0,0087	1,0345	450	309	141

**Πίνακας 6:** Γονίδια *Z.mays* - *O.sativa* με αυξημένο  $\omega$ 

<b>GENE NAME</b>	<b>dNS</b>	<b>dS</b>	<b>dNS/dS</b>	<b>LENGTH</b>	<b>NS POSITIONS</b>	<b>S POSITIONS</b>
<b>ccmB (mt)</b>	0,0087	0,0000	>1	618	434	184
<b>nad2 (mt)</b>	0,0054	0,0028	1,9286	1464	1026	438
<b>nad3 (mt)</b>	0,0150	0,0119	1,2605	354	263	91

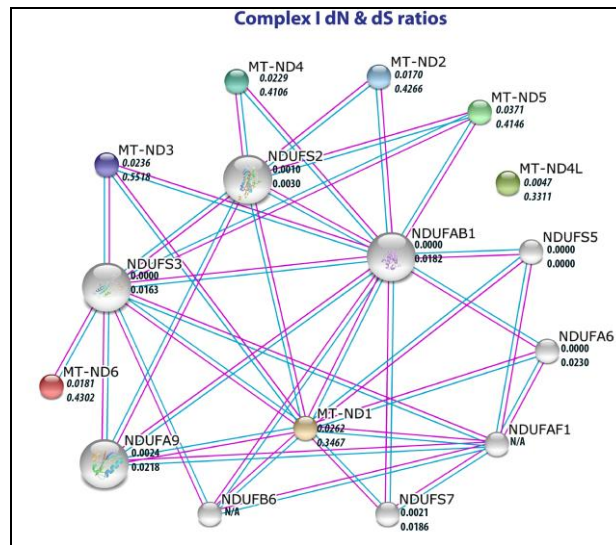
Τελειώνοντας με τη μελέτη των ρυθμών αντικατάστασης, υπήρξε η σκέψη εάν κάποιοι από αυτούς μπορούσαν να συσχετιστούν με τον αριθμό των πρωτεϊνικών αλληλεπιδράσεων στις οποίες συμμετέχει η αντίστοιχη πρωτεΐνη του κάθε γονιδίου. Μια θεωρία θα μπορούσε να είναι, πως όσο μεγαλύτερο αριθμό αλληλεπιδράσεων έχει μια πρωτεΐνη, τόσο πιο ισχυρή θα είναι και η ένταση της επιλογής επάνω της. Λόγω μεγαλύτερου εύρους διαθέσιμων δεδομένων, η μελέτη αυτή, επικεντρώθηκε στον *Homo sapiens*. Στον Πίνακα 5 φαίνεται ο αριθμός των πρωτεϊνικών αλληλεπιδράσεων για όλα τα μιτοχονδριακά γονίδια του Συμπλόκου I, καθώς και το μέγεθός τους (βλέπε Παράρτημα για όλα τα γονίδια).

**Πίνακας 7:** Μέγεθος γονιδίου και αριθμός PPI για τα μιτοχονδριακά γονίδια του Συμπλόκου I. Στη δεύτερη στήλη είναι το μήκος του γονιδίου, και στην Τρίτη, ο αριθμός των αλληλεπιδράσεων.

<b>OXPPOS C. I</b>	<b>LENGTH</b>	<b>PPI 90% confidence</b>
<b>ND1 (mt)</b>	954	8
<b>ND2 (mt)</b>	1041	2
<b>ND3 (mt)</b>	345	3
<b>ND4 (mt)</b>	1377	2
<b>ND4L (mt)</b>	294	0
<b>ND5 (mt)</b>	1809	3
<b>ND6 (mt)</b>	522	1

Στην Εικόνα 10 φαίνονται οι πρωτεϊνικές αλληλεπιδράσεις ανάμεσα στις υπομονάδες του Συμπλόκου I, σύμφωνα με τη βάση δεδομένων STRING. Κάτω από το όνομα της κάθε πρωτεΐνης, αναγράφονται οι τιμές dN, dS, και τα S.E. τους. Ειδικά για αυτό το σύμπλοκο, παρατηρήθηκε στατιστικά σημαντική συσχέτιση ανάμεσα στο μήκος της κάθε πρωτεΐνης, και στον αριθμό των αλληλεπιδράσεων της με άλλες πρωτεΐνες (βλέπε Παράρτημα για τα υπόλοιπα σύμπλοκα).





Εικόνα 13: Ενδεικτικά PPI και τιμές dN, dS για το Σύμπλοκο I.

Στον Πίνακα 8 φαίνονται οι συσχετίσεις ανάμεσα στα χαρακτηριστικά διαφορετικών ΟΧΡΗΟΣ συμπλόκων στον άνθρωπο, οι οποίες υπολογίστηκαν με βάση την τιμή Kendall's "τ", η οποία υπολογίζει τη στατιστική συσχέτιση ανάμεσα σε δυο ποσότητες:

**Πίνακας 8:** Συσχετίσεις χαρακτηριστικών ανά σύμπλοκο. Φαίνονται τα correlation coefficients στα στατιστικά σημαντικά αποτελέσματα, καθώς και το p-value (σε παρένθεση)

COMPLEX	No of genes	dS - ω	dS - length	dS - PPI	ω - length	ω - PPI	length - PPI
OXPHOS I (mit)	7	n/s	n/s	n/s	n/s	0,651 (0.046)	n/s
OXPHOS I (nucl)	31	n/s	n/s	n/s	0,338 (0.038)	n/s	0,393 (0.005)
OXPHOS II (nucl)	3	n/s	n/s	n/s	n/s	n/s	n/s
OXPHOS III (mit)	1	n/s	n/s	n/s	n/s	n/s	n/s
OXPHOS III (nucl)	9	n/s	n/s	n/s	n/s	n/s	n/s
OXPHOS IV (mit)	3	n/s	n/s	n/s	n/s	n/s	n/s
OXPHOS IV (nucl)	12	n/s	n/s	n/s	n/s	n/s	n/s
OXPHOS V (mit)	2	n/s	n/s	n/s	n/s	n/s	n/s
OXPHOS V (nucl)	15	n/s	n/s	n/s	n/s	n/s	0,644 (0.001)
All Mit.	13	n/s	n/s	n/s	n/s	n/s	0.416 (0.055)
All Nuclear	70	n/s	n/s	n/s	n/s	0.236 (0.035)	0.233 (0.008)

Στα πυρηνικά γονίδια των Συμπλόκων I και V, παρατηρείται στατιστικά σημαντική θετική συσχέτιση ανάμεσα στο μήκος των πρωτεϊνών, και στον αριθμό των άλλων αλληλεπιδρώντων πρωτεϊνών με αυτές. Αυτό είναι κάτι το

αναμενόμενο, δεδομένου ότι μεγαλύτερες πρωτεΐνες έχουν περισσότερη διαθέσιμη επιφάνεια για πρωτεϊνικές αλληλεπιδράσεις.

Στο Σύμπλοκο I, παρατηρείται επίσης θετική συσχέτιση ανάμεσα στον ρυθμό  $\omega$ , και στο μήκος της πρωτεΐνης. Αυτό σημαίνει, ότι όσο μεγαλώνει η πρωτεΐνη, μειώνεται η ένταση της επιλογής (αυξάνει το  $\omega$ ). Δεδομένου του ότι υπάρχει και συσχέτιση ανάμεσα στο μήκος των πρωτεϊνών και στον αριθμό των αλληλεπιδράσεών τους στο σύμπλοκο αυτό, αυτό θα μπορούσε να σημαίνει ότι μικρότερες πρωτεΐνες συμμετέχουν σε μεμονωμένες αλληλεπιδράσεις όπου η επιλογή είναι ισχυρή, ενώ οι μεγαλύτερες βρίσκονται σε πιο κεντρικά σημεία, οπότε κάποιες αμινοξικές αντικαταστάσεις δεν επιδρούν σημαντικά στην συνοχή του συμπλόκου, λόγω των αθροιστικών αλληλεπιδράσεων των γύρω πρωτεϊνών.

Στα μιτοχονδριακά γονίδια του Συμπλόκου I, φαίνεται θετική συσχέτιση ανάμεσα στην ένταση της επιλογής, και στον αριθμό πρωτεϊνικών αλληλεπιδράσεων. Εδώ φαίνεται πιο ξεκάθαρα η θεωρία του προηγούμενου παραδείγματος, όπου μεγαλύτερες πρωτεΐνες βρίσκονται σε πιο κεντρικά σημεία του συμπλόκου.

Στα σύμπλοκα με λίγες μιτοχονδριακές πρωτεΐνες, όπως τα Σύμπλοκα III, IV, και V, δεν ήταν δυνατόν να παρθούν στατιστικά σημαντικά αποτελέσματα, λόγω του μικρού αριθμού των γονιδίων. Ομοίως, και για τις πυρηνικές πρωτεΐνες του Συμπλόκου II που συμμετέχουν μόνο τέσσερις πρωτεΐνες.

Αθροιστικά, φαίνεται πως και στα μιτοχονδριακά, και στα πυρηνικά γονίδια, υπάρχει θετική συσχέτιση ανάμεσα στο μήκος των πρωτεϊνών και στον αριθμό των αλληλεπιδράσεών τους. Παρόμοιες μελέτες έχουν ήδη συσχετίσει το μήκος των πρωτεϊνών με τον αριθμό των πρωτεϊνικών αλληλεπιδράσεων στις οποίες συμμετέχουν, καθώς και με τον αριθμό των ενδοκυττάρων περιοχών στις οποίες οι πρωτεΐνες αυτές εντοπίζονται<sup>55</sup>.

### 3. ΔΙΕΡΕΥΝΗΣΗ ΤΗΣ ΥΠΑΡΞΗΣ ΘΕΤΙΚΗΣ ΕΠΙΛΟΓΗΣ ΣΕ ΓΟΝΙΔΙΑ ΤΩΝ ΣΥΜΠΛΟΚΩΝ ΤΗΣ ΟΞΕΙΔΩΤΙΚΗΣ ΦΩΣΦΟΥΡΙΩΣΗΣ ΣΕ ΕΞΕΛΙΚΤΙΚΕΣ ΓΡΑΜΜΕΣ ΤΗΣ ΟΙΚΟΓΕΝΕΙΑΣ ΔΡΟΣΟΦΙΛΙΔΩΝ

#### 3.1 Εισαγωγή

Οι περιοχές του γονιδιώματος στις οποίες δρα έντονα αρνητική επιλογή, χαρακτηρίζονται συνήθως από χαμηλούς ρυθμούς αμινοξικών αντικαταστάσεων. Οι περισσότερες μεταλλάξεις έχουν αρνητικές επιπτώσεις στις κωδικοποιούμενες πρωτεΐνες, και έτσι απαλείφονται γρήγορα από την αλληλική δεξαμενή. Ωστόσο, ένα μικρό ποσοστό των αμινοξικών μεταλλάξεων μπορούν να αυξήσουν την ικανότητα ενός οργανισμού να επιβιώσει στο περιβάλλον του (αρμοστικότητα), και υπό συγκεκριμένες συνθήκες, να εξαπλωθούν στον αντίστοιχο πληθυσμό και να εγκαθιδρυθούν. Τέτοιες αλλαγές στις αλληλουχίες είναι ιδιαίτερα ενδιαφέρουσες γιατί είναι εκείνες που διαφοροποιούν τα είδη μεταξύ τους και τους προσφέρουν τη δυνατότητα να προσαρμοστούν καλύτερα στο συγκεκριμένο περιβάλλον στο οποίο ζουν. Όπως προειπώθηκε, ένας από τους πιο συνηθισμένους τρόπους αναγνώρισης του ίχνους που αφήνει η επιλογή στα πρωτεϊνικά γονίδια, είναι μέσω του λόγου  $dN/dS$  ( $\omega$ ), δηλαδή του ρυθμού μη συνώνυμων αντικαταστάσεων κατά μήκος ενός γονιδίου, προς τον ρυθμό των συνώνυμων. Τιμές μεγαλύτερες της μονάδας είναι ενδείξεις θετικής επιλογής, τιμές κοντά σε αυτή δείχνουν ουδέτερες αλλαγές, ενώ μικρότερες τιμές είναι ενδείξεις αρνητικής επιλογής. Η τεχνική αυτή έχει χρησιμοποιηθεί για την μελέτη πληθώρας γονιδίων σε όλους τους οργανισμούς. Χαρακτηριστικό παράδειγμα με  $\omega > 1$  είναι κάποια γονίδια τα οποία κωδικοποιούν πρωτεΐνες του ανοσοποιητικού συστήματος<sup>56</sup>. Το φαινόμενο αυτό έχει υποθεθεί πως είναι αποτέλεσμα της "κούρσας εξοπλισμών" ανάμεσα σε παθογόνα και στους ξενιστές τους, και είναι γνωστό ως η υπόθεση της "Κόκκινης Βασίλισσας" (από τον ομώνυμο χαρακτήρα του Lewis Carroll: *"It takes all the running you can do, to keep in the same place"*). Στο παράδειγμά μας, κάποιες νέες μεταλλάξεις στα παθογόνα είναι ευνοϊκές επειδή τους επιτρέπουν να αποφύγουν το ανοσοποιητικό σύστημα του ξενιστή, του δίνουν πίσω το πλεονέκτημα απέναντι στα νέα παθογόνα, οπότε επαναλαμβάνονται τα ίδια βήματα. Ένα άλλο χαρακτηριστικό παράδειγμα με  $\omega > 1$ , είναι κάποια γονίδια οσφρητικών υποδοχέων στον άνθρωπο<sup>57</sup>. Ωστόσο, σε άλλα γονίδια οσφρητικών υποδοχέων, έχει σταματήσει να ασκείται επιλογή στον άνθρωπο επειδή έχουν μετατραπεί σε ψευδογονίδια, και έτσι έχουν  $\omega \approx 1$  (ωστόσο, εάν η μετατροπή σε ψευδογονίδιο είναι αρκετά πρόσφατη, είναι πιθανόν το  $\omega$  να είναι σχετικά μικρότερο του 1, λόγω της ύπαρξης προηγούμενων συνώνυμων αντικαταστάσεων)<sup>56</sup>. Επειδή τα γονίδια αυτά δεν μεταφράζονται, ή μεταγράφεται/μεταφράζεται ένα μόνο τμήμα τους το οποίο είναι συνήθως ανενεργό, δεν έχουν επίδραση στην αρμοστικότητα του οργανισμού, οπότε δεν βρίσκονται κάτω από την επίδραση επιλογής, και έτσι οι αντικαταστάσεις συσσωρεύονται στις συνώνυμες και στις μη συνώνυμες θέσεις με τον ίδιο ρυθμό. Η χρήση του ρυθμού  $\omega$  για την εξαγωγή συμπερασμάτων για τη μορφή και την ένταση της επιλογής παρουσιάζει το πλεονέκτημα ότι χρειάζεται μικρή ποσότητα δεδομένων (αρκούν δύο ομόλογες αλληλουχίες

από διαφορετικά είδη) ωστόσο, έχει αρκετά προβλήματα και το τεστ θεωρείται συντηρητικό. Για παράδειγμα, οι συνώνυμες μεταλλάξεις δεν είναι πάντα ουδέτερες (π.χ. σε περιπτώσεις codon usage bias ή άκρων εξωνίων)<sup>25</sup>. Επιπλέον, σε αρκετές περιπτώσεις, οι περισσότερες μη συνώνυμες μεταλλάξεις είναι δηλητηριώδεις, και λίγες μόνο αυξάνουν την αρμοστικότητα του οργανισμού (και ίσως εγκαθιδρύονται). Εάν το υπό εξέταση γονίδιο είναι αρκετά μεγάλο, τότε ο ρυθμός  $\omega$  ίσως να είναι δύσκολο να υπερβεί τη μονάδα, λόγω της πληθώρας συνώνυμων αντικαταστάσεων σε σχέση με τις μη συνώνυμες. Τέλος, ο ρυθμός  $\omega$  δεν υπολογίζεται μόνο ανάμεσα σε κοντινούς οργανισμούς, αλλά και ανάμεσα σε όλα τα κλαδιά ενός φυλογενετικού δέντρου, ενώ, όσο περισσότερο πίσω στον χρόνο και σε κοινούς προγόνους πηγαίνουμε, τόσο περισσότερο αντίκτυπο έχουν τυχόν λάθη στα αποτελέσματά μας. Αρκετές τεχνικές έχουν περιγραφεί προκειμένου να προσπεραστούν αυτά τα προβλήματα.

Μια λύση για την υπέρβαση των προβλημάτων της μεθόδου αυτής είναι η ανακατασκευή των προγονικών αλληλουχιών, χρησιμοποιώντας τα πραγματικά δεδομένα των διαθέσιμων αλληλουχιών. Οι κατασκευασμένες προγονικές αλληλουχίες μπορούν τότε να συγκριθούν με τις υπό μελέτη, και να προσδιοριστούν οι ρυθμοί αντικατάστασης. Υπάρχουν αρκετές διαφορετικές μέθοδοι για την κατασκευή των αλληλουχιών αυτών, όπως μέγιστη φειδωλότητα (maximum parsimony), μέγιστη πιθανοφάνεια (maximum likelihood), και Bayesian inference. Η κάθε μέθοδος έχει τα υπέρ και τα κατά της. Προβλήματα όπως τυχαία λάθη μπορούν να έχουν μεγάλο αντίκτυπο στα αποτελέσματα<sup>58</sup>.

Μια ακόμα τεχνική για τον προσδιορισμό του ρυθμού  $\omega$ , είναι η προσέγγιση πιθανοφάνειας (likelihood ratio), η οποία αποφεύγει να ανακατασκευάσει συγκεκριμένες προγονικές αλληλουχίες, κατασκευάζοντας αντ' αυτών έναν πίνακα πιθανοτήτων για όλες τις πιθανές προγονικές αλληλουχίες σε κάθε κόμβο του φυλογενετικού δέντρου (βλέπε παρακάτω). Η κάθε προγονική αλληλουχία έχει ανάλογο βάρος με την πιθανότητα να οδηγήσει στα παρατηρούμενα δεδομένα. Σε ένα τέτοιο μοντέλο, είναι εύκολο να ληφθούν υπ' όψιν ξεχωριστές παράμετροι, όπως συχνότητα μεταπτώσεων/μεταστροφών ή προτίμηση στη χρήση συγκεκριμένων κωδικονίων.

Προκειμένου να δειχθεί εάν κάποιο γονίδιο έχει υποστεί επιλογή μέσα σε μια ομάδα κοντινών μεταξύ τους οργανισμών, μία ή παραπάνω από τις τεχνικές αυτές για τον προσδιορισμό του  $\omega$ , εφαρμόζονται στο πλήθος των ορθολόγων του γονιδίου αυτού, με οδηγό το φυλογενετικό δέντρο των υπό μελέτη οργανισμών. Το φυλογενετικό δέντρο είναι ένα διάγραμμα το οποίο αναπαριστά την εξελικτική ιστορία ενός συνόλου οργανισμών. Μπορεί να κατασκευαστεί είτε από δεδομένα άλλων αναλύσεων οι οποίες συναινούν στην εξελικτική ιστορία των υπό μελέτη οργανισμών (**δέντρο βιβλιογραφίας**, Εικόνες 14, 15), είτε από συγκριτική ανάλυση ενός μόνο ομόλογου γονιδίου το οποίο υπάρχει σε καθέναν από τους οργανισμούς αυτούς (**δέντρο γονιδίου**). Για παράδειγμα, εάν τα περισσότερα δέντρα γονιδίου είναι πανομοιότυπα για το ίδιο σύνολο οργανισμών, το πιο πιθανό είναι αυτό το δέντρο να είναι και το πραγματικό, και τότε μπορεί να χαρακτηριστεί σαν δέντρο βιβλιογραφίας. Κάθε κλαδί του δέντρου αντιπροσωπεύει ένα είδος, το οποίο μπορεί είτε να

έχει εκλείψει, είτε όχι. Ανάλογα με το είδος του δέντρου, το μήκος του κάθε κλαδιού μπορεί να μετράει εξελικτικό χρόνο ή αριθμό νουκλεοτιδικών/αμινοξικών αλλαγών στο γονιδίωμα του κάθε οργανισμού. Στα έρριζα δέντρα, κάθε κόμβος αναπαριστά τον τελευταίο κοινό πρόγονο των κλαδιών/οργανισμών που διακλαδώνονται εκατέρωθεν αυτού.

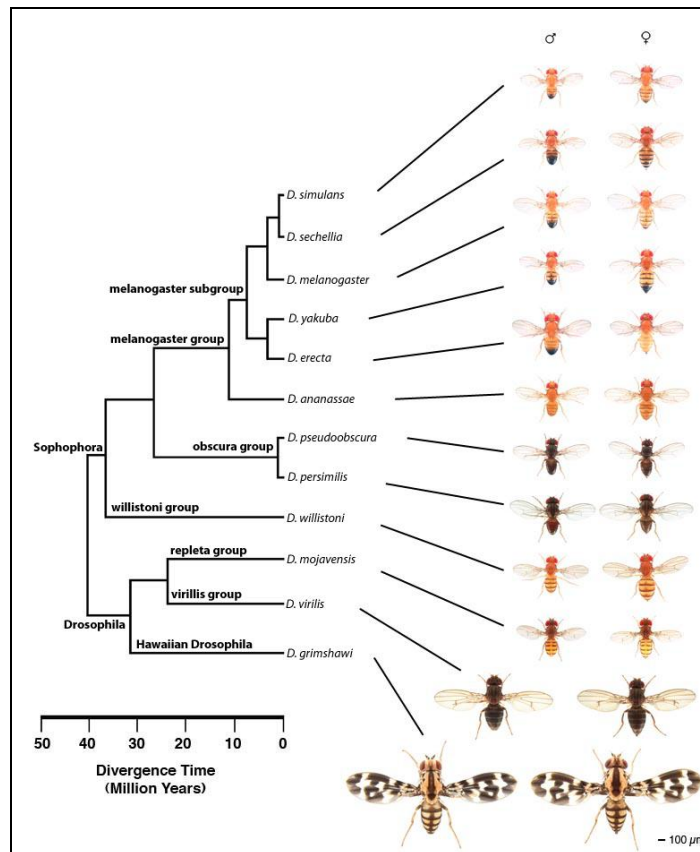
Για κάθε κλαδί ενός δέντρου γονιδίου, μπορεί να προσδιοριστεί ένας ρυθμός  $\omega$ , ανάλογα με την ένταση της επιλογής πάνω στο γονίδιο αυτό του εν λόγω οργανισμού. Ο ρυθμός αυτός μπορεί να υπολογιστεί χρησιμοποιώντας τις διαφορετικές προσεγγίσεις που περιγράφηκαν πιο πάνω (πχ. maximum parsimony, maximum likelihood). Όταν η ένταση της επιλογής δεν αλλάζει από οργανισμό σε οργανισμό, τότε όλα τα κλαδιά του δέντρου θα πρέπει να έχουν κοντινούς ρυθμούς  $\omega$ . Όταν υπάρχουν ένα ή παραπάνω κλαδιά με διαφορετικούς ρυθμούς, για κάποιον λόγο η ένταση της επιλογής πάνω σε αυτό το γονίδιο είναι πιθανόν έχει αλλάξει στους εν λόγω οργανισμούς.

### 3.2 Σκοπός της μελέτης

Η οικογένεια των δροσοφιλιδών αποτελείται από περίπου 65 γένη, με πάνω από 3.500 χαρακτηρισμένα είδη. Μετά το πέρας της αλληλούχισης του γονιδιώματος της *D. melanogaster* το 2000<sup>59</sup>, ξεκίνησε η αλληλούχιση διαφόρων άλλων συγγενών ειδών, προκειμένου να απαντηθούν νέα ερωτήματα σχετικά με τους μεταλλακτικούς και εξελικτικούς μηχανισμούς που οδήγησαν στην γένεση του κάθε είδους.

Όπως αναφέρθηκε και στο προηγούμενο κεφάλαιο, αμινοξικές αντικαταστάσεις στα σύμπλοκα της αναπνοής μπορούν να προκαλέσουν διαβαθμίσεις στην ενεργότητα των συμπλόκων φωσφορυλίωσης, και στα ποσοστά ενέργειας και θερμότητας που παράγονται. Οι διαβαθμίσεις αυτές, θεωρείται πως είναι εν μέρει υπεύθυνες για την προσαρμογή και επιβίωση οργανισμών σε διαφορετικά περιβάλλοντα θερμοκρασίας. Έτσι, μελέτες για πιθανά ενεργές μεταλλάξεις, καθώς και για αύξηση του εξελικτικού ρυθμού σε μερικά από τα γονίδια που απαρτίζουν τα σύμπλοκα αυτά, έχουν γίνει για τον άνθρωπο<sup>51</sup>, για άλλα πρωτεύοντα θηλαστικά<sup>47,48,49,50,60</sup>, για τη *D. melanogaster*<sup>61</sup>, για φυτά<sup>62</sup>, και για ψάρια<sup>63</sup>.

Η παρούσα εργασία είχε σκοπό να εφαρμόσει την likelihood approach για να δώσει μια ολοκληρωμένη εικόνα της ύπαρξης ή όχι θετικής επιλογής για όλα τα γονίδια που απαρτίζουν τα σύμπλοκα οξειδωτικής φωσφορυλίωσης (είτε αυτά κωδικοποιούνται από τον πυρήνα είτε κωδικοποιούνται από το μιτοχόνδριο) σε δέκα είδη της οικογένειας των δροσοφιλιδών (*Drosophilidae*), και να συγκρίνει τους ρυθμούς εξέλιξης των πρωτεϊνών αυτών σε όλους τους κόμβους του φυλογενετικού δέντρου. Σαν βάση χρησιμοποιήθηκαν δέκα είδη των οποίων τα γονιδιώματα αλληλουχήθηκαν πρόσφατα (Εικόνα 14)<sup>64</sup>.



Εικόνα 14: Φυλογενετικό δέντρο δροσοφιλιδών με αλληλουχημένα γονιδιώματα.

### 3.3 Υλικά και μέθοδοι

Την στιγμή εκπόνησης της εργασίας, υπήρχαν διαθέσιμα τα πλήρη πυρηνικά μιτοχονδριακά γονιδιώματα 12 ειδών (όπως φαίνεται στην Εικόνα 14), καθώς και τα μιτοχονδριακά γονιδιώματα των 10 από αυτών<sup>63</sup>. Τα δυο είδη για τα οποία δεν υπήρχαν διαθέσιμα τα μιτοχονδριακά γονιδιώματα, και δεν χρησιμοποιήθηκαν στη μελέτη, ήταν οι *D. virilis* και *D. willistoni*. Στον Πίνακα 9 φαίνονται τα είδη που χρησιμοποιήθηκαν στη μελέτη. Το *A. gambiae* χρησιμοποιήθηκε σαν το εξωμάδα (outgroup) προκειμένου να κατασκευαστούν τα φυλογενετικά δέντρα.

Πίνακας 9: Είδη από τα οποία συλλέχτησαν γονίδια

<i>Drosophila melanogaster</i> (DME)
<i>Drosophila simulans</i> (DSI)
<i>Drosophila yakuba</i> (DYA)
<i>Drosophila sechellia</i> (DSE)
<i>Drosophila mojavensis</i> (DMO)
<i>Drosophila ananassae</i> (DAN)
<i>Drosophila erecta</i> (DER)
<i>Drosophila persimilis</i> (DPE)
<i>Drosophila grimshawi</i> (DGR)
<i>Drosophila pseudoobscura</i> (DPO)
<i>Anopheles gambiae</i> (AGA)

Οι παρακάτω βάσεις δεδομένων χρησιμοποιήθηκαν για τη συλλογή των υπό μελέτη γονιδίων:

1. NCBI GenBank  
(<http://www.ncbi.nlm.nih.gov/Genbank>)
2. KEGG GENES  
(<http://www.genome.jp/kegg/genes.html>)

Συνολικά συλλέχθηκαν 39 γονίδια για το κάθε είδος, εκ των οποίων τα 26 ήταν πυρηνικά και τα 13 μιτοχονδριακά (βλέπε Παράρτημα). Αυτά ήταν όλα τα μιτοχονδριακά πρωτεϊνικά γονίδια, και όσα πυρηνικά είχαν αναγνωρισμένα ορθόλογα σε όλους τους υπό μελέτη τους οργανισμούς σύμφωνα με την βάση KEGG. Για το κάθε γονίδιο έγινε πολλαπλή στοίχιση με το πρόγραμμα ClustalW, καθώς και φυλογενετικά δέντρα με τα προγράμματα MEGA4 (για Neighbour Joining δέντρα), και DNAmI (για Maximum Likelihood δέντρα). Τα δέντρα Maximum Likelihood κατασκευάστηκαν μέσω του προγράμματος DNAmI, χρησιμοποιώντας σαν outgroup το *A.gambiae*, και 100 επαναλήψεις. Τα δέντρα Neighbour Joining, κατασκευάστηκαν μέσω του προγράμματος MEGA4, χρησιμοποιώντας τη μέθοδο Neighbor Joining, με 1000 επαναλήψεις bootstrap. Οι εξελικτικές αποστάσεις υπολογίστηκαν με τη μέθοδο Maximum Composite Likelihood.

Για τους υπολογισμούς, χρησιμοποιήθηκε το πρόγραμμα CODEML (πιο συγκεκριμένα, το CODONML), από τη σουίτα προγραμμάτων φυλογενετικής ανάλυσης PAML. Το CODEML μπορεί να επιτελέσει τις ακόλουθες λειτουργίες:

- Έλεγχο και σύγκριση φυλογενετικών δέντρων.
- Ελέγχους likelihood ratio για διαφορετικές εξελικτικές υποθέσεις, βάσει ενσωματωμένων μοντέλων.
- Ανακατασκευή προγονικών αλληλουχιών, βάσει της τεχνικής likelihood ratio, χρησιμοποιώντας διαφορετικά μοντέλα κωδικονίων.
- Υπολογισμό συνώνυμων και μη συνώνυμων ρυθμών αντικαταστάσεων, και εντοπισμό θετικής επιλογής σε πρωτεϊνικά γονίδια.

Το CODEML περιέχει τέσσερα μοντέλα ανίχνευσης επιλογής:

1. Το βασικό μοντέλο (**basic model**), το οποίο είναι το πιο απλοποιημένο μοντέλο μέτρησης dN/dS ratios. Μπορεί να πάρει μόνο δυο παραμέτρους, οι οποίες είναι ο παρατηρημένος ρυθμός μεταπτώσεων/μεταστροφών ( $k$ ), και ένα αρχικό dN/dS ratio ( $\omega$ ). Δίνει σαν αποτέλεσμα ένα και μόνο  $\omega$ , το οποίο είναι το μέσο  $\omega$  όλων των κλαδιών του δέντρου και των θέσεων των αλληλουχιών που χρησιμοποιήθηκαν. Όπως ειπώθηκε και στην εισαγωγή, ένα τέτοιο μοντέλο πολύ δύσκολα θα δώσει αποτελέσματα θετικής επιλογής, δεδομένου ότι τυχόν στατιστικά σημαντικά  $\omega$  θα χαθούν μέσα στον «θόρυβο» των γειτονικών κλαδιών/περιοχών.
2. Το μοντέλο κλαδιών (**branch model**) επιτρέπει στο  $\omega$  να διαφέρει ανάμεσα σε διαφορετικά κλαδιά του φυλογενετικού δέντρου, και χρησιμοποιείται για τον εντοπισμό θετικής επιλογής σε συγκεκριμένα

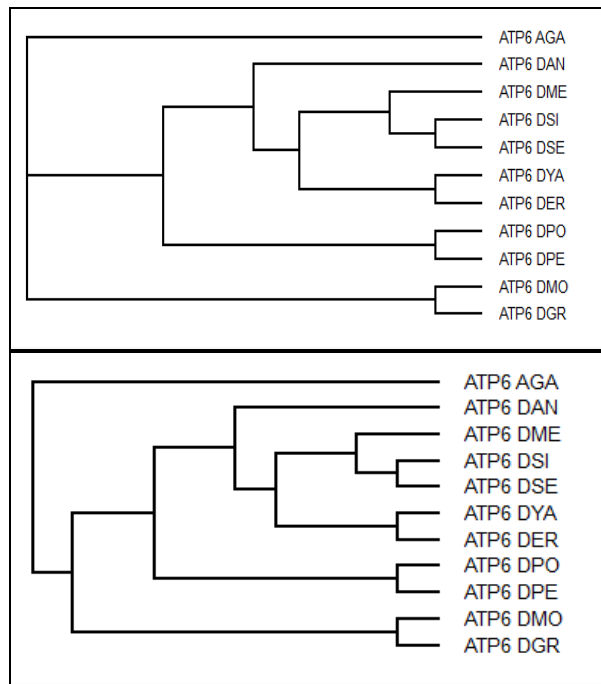
κλαδιά. Παραδείγματα επιμέρους μοντέλων τα οποία εφαρμόζονται στα δεδομένα και συγκρίνονται μεταξύ τους για το βέλτιστο, είναι:

- 2.1. One ratio (M0): Θεωρεί ότι ο ρυθμός  $\omega$  είναι ο ίδιος για όλα τα κλαδιά του δέντρου.
  - 2.2. Free ratio (M1): Θεωρεί ότι κάθε κλαδί έχει τον δικό του ρυθμό  $\omega$ .
  - 2.3. Two ratio: Θεωρεί ότι το συγκεκριμένο κλαδί που ορίζουμε εμείς, έχει διαφορετικό  $\omega$  από το υπόλοιπο δέντρο.
  - 2.4. Three ratio: Ομοίως με πριν, με δύο επιλεγμένα κλαδιά, και τρία διαφορετικά συνολικά  $\omega$ .
3. Το μοντέλο θέσεων (**site model**) επιτρέπει στο  $\omega$  να διαφέρει ανάμεσα σε διαφορετικές νουκλεοτιδικές θέσεις κατά μήκος των γονιδίων. Στην περίπτωση μας, σαν ξεχωριστή θέση ορίζεται κάθε κωδικώνιο, ενώ σαν τέτοιες μπορούν να θεωρηθούν επίσης μοναδικές νουκλεοτιδικές βάσεις, ή μοναδικά αμινοξέα (όταν εξετάζουμε αμινοξικές αλληλουχίες). Το μοντέλο αυτό ανιχνεύει συγκεκριμένες θέσεις που έχουν αποτελέσει στόχους θετικής επιλογή. Παραδείγματα επιμέρους μοντέλων τα οποία εφαρμόζονται στα δεδομένα και συγκρίνονται μεταξύ τους για το βέλτιστο, είναι:
- 3.1. Neutral (M1): Θεωρεί ότι υπάρχουν δυο κατηγορίες θέσεων μέσα στην αλληλουχία, με  $\omega$  είτε 0, είτε 1.
  - 3.2. Selection (M2): Προσθέτει και τρίτη κατηγορία θέσεων στο προηγούμενο μοντέλο, η οποία έχει το δικό της  $\omega$ , το οποίο και υπολογίζει από τα δεδομένα. Το μοντέλο παραδέχεται ότι το  $\omega$  έχει σταθερή τιμή και δεν ακολουθεί κάποια κατανομή.
  - 3.3. Beta (M7): Θεωρεί ότι το σύνολο των  $\omega$  μέσα για όλες τις θέσεις της αλληλουχίας, ακολουθεί τη beta κατανομή, με  $0 < \omega < 1$ . Εφόσον σύμφωνα με αυτό, δεν ασκείται θετική επιλογή σε καμία θέση, χρησιμοποιείται σαν μηδενική υπόθεση (null hypothesis).
  - 3.4. Beta &  $\omega$  (M8): Προσθέτει και μια τρίτη κατηγορία θέσεων στο προηγούμενο μοντέλο, η οποία έχει το δικό της  $\omega$ , το οποίο υπολογίζει από τα δεδομένα, και το οποίο μπορεί να πάρει τιμή  $> 1$ .
4. Το μοντέλο κλαδιών-θέσεων (**branch - site model**) επιτρέπει στο  $\omega$  να διαφέρει τόσο ανάμεσα σε διαφορετικές θέσεις, όσο και ανάμεσα στα κλάδιά του δέντρου. Είναι χρήσιμο στην ανίχνευση θετικής επιλογής όταν αυτή υφίσταται πάνω σε λίγες θέσεις, σε συγκεκριμένες γενεαλογίες. Στο μοντέλο αυτό, ορίζεται από τον χρήστη ένα συγκεκριμένο κλαδί (foreground branch), ο ρυθμός  $\omega$  του οποίου, θεωρείται ότι είναι διαφορετικός από αυτόν των άλλων κλαδιών (background branches), τα οποία έχουν όλα το ίδιο  $\omega$ . Παραδείγματα μοντέλων είναι:
- 4.1 Null model: Το foreground branch μπορεί να έχει διαφορετικό  $\omega$  από τα background branches, σε κανένα όμως δεν υπάρχουν θέσεις οι οποίες βρίσκονται κάτω από θετική επιλογή.



#### 4.2 Model A: Το foreground branch μπορεί να έχει κάποιες θέσεις οι οποίες βρίσκονται κάτω από θετική επιλογή.

Στα μοντέλα κλαδιών ή θέσεων, το output κάθε επιμέρους μοντέλου είναι μια τιμή log likelihood ratio (LLR). Προκειμένου να συγκριθούν δυο επιμέρους μοντέλα μοντέλα (συνήθως συγκρίνεται η τιμή likelihood που δίνει ένα μοντέλο που παραδέχεται την ύπαρξη θετικής επιλογής με το αντίστοιχο μοντέλο που δεν κάνει τέτοια παραδοχή), υπολογίζεται το διπλασιο της διαφοράς των μοντέλων αυτών, και η τιμή αυτή συγκρίνεται με μια κατανομή  $\chi^2$ , με  $\alpha=0.05$ , και βαθμούς ελευθερίας (degrees of freedom) ίσους με τη διαφορά των ελεύθερων παραμέτρων (π.χ. αριθμός διαφορετικών  $\omega$ , πιθανότητα χρήσης κάθε κωδικωνίου) που χρησιμοποιήθηκαν σε κάθε μοντέλο και υπολογίζονται από τα δεδομένα. Εάν η τιμή αυτή είναι στατιστικά σημαντική, ( $p\text{-value} < 0.05$ ), τότε θεωρούμε πως το μοντέλο με το μεγαλύτερο log likelihood ratio εφαρμόζει καλύτερα στα δεδομένα μας από αυτό με το μικρότερο.



**Εικόνα 15:** Unrooted και Rooted δέντρα βιβλιογραφίας για τους υπό μελέτη οργανισμούς. Το rooted δέντρο, έχει το AGA σαν outgroup.

#### Μοντέλο κλαδιών

Η πρώτη μελέτη έγινε χρησιμοποιώντας το μοντέλο κλαδιών του CODEML. Προκειμένου να δούμε εάν ο ρυθμός  $\omega$  παραμένει ο ίδιος ανάμεσα στα διαφορετικά κλαδιά του δέντρου των δροσοφιλιδών, ή εάν αλλάζει σε συγκεκριμένα κλαδιά, συγκρίναμε τα μοντέλα M0 (σταθερό  $\omega$ ) και M1 (μεταβλητό  $\omega$ ). Εξετάσαμε δηλαδή, εάν δηλαδή υπήρχαν κάποια κλαδιά που έδειχναν αυξημένο  $\omega$ .

Η σύγκριση αυτή έγινε τόσο για τα πυρηνικά όσο και για τα μιτοχονδριακά γονίδια, και χρησιμοποιήθηκαν:

- Φυλογενετικά δέντρα των οργανισμών από τη βιβλιογραφία (Εικόνα 15), τα οποία ήταν είτε rooted είτε unrooted. Σύμφωνα με το εγχειρίδιο του PAML, τα rooted δέντρα χρησιμοποιούνται όταν θεωρείται ότι ο ρυθμός εξέλιξης ανάμεσα σε όλους τους οργανισμούς παραμένει σταθερός (υπάρχει δηλαδή μοριακό ρολόι). Σε αντίθετη περίπτωση χρησιμοποιούνται unrooted. Επειδή στη μελέτη μας δεν είχαμε πληροφορίες για την ύπαρξη ή όχι μοριακού ρολογιού, χρησιμοποιήσαμε και τα δυο είδη δέντρων.
- Για κάθε σετ στοιχισμένων αλληλουχιών, είτε λήφθηκαν υπόψιν ασαφή δεδομένα στις αλληλουχίες (π.χ. κενά στη στοίχιση ή ασαφή νουκλεοτίδια), είτε αυτά απορρίφθηκαν πριν την έναρξη του προγράμματος. Η μη χρήση ασαφών δεδομένων, αφαιρεί από την ανάλυση θέσεις οι οποίες περιέχουν αμφίσημες βάσεις ή κενά σε τουλάχιστον μια από τις αλληλουχίες (όπως αυτές εμφανίζονται στην πολλαπλή στοίχιση των αλληλουχιών) αυξάνουν όμως την αξιοπιστία των αποτελεσμάτων. Επειδή η αλληλουχία του *A. gambiae* ήταν σε κάποιες περιπτώσεις αρκετά διαφορετική από τις υπόλοιπες, εισάγονταν πολλά κενά στη στοίχιση των αλληλουχιών, και έτσι χανόταν αρκετή πληροφορία από τα διαφορετικά είδη των δροσοφιλιδών. Εάν δηλαδή υπήρχε κενό σε μια θέση στην αλληλουχία του *A. gambiae*, η μη χρήση ασαφών δεδομένων θα αφαιρούσε όλα τα αμινοξέα από αυτή την θέση της πολλαπλής στοίχισης, ακόμα και εάν οι αντίστοιχες θέσεις σε όλες τις δροσοφιλίδες δεν ήταν κενές. Για αυτόν τον λόγο, τα μοντέλα μας εφαρμόστηκαν τόσο παρουσία, όσο και απουσία των δεδομένων αυτών
- Το πρόγραμμα ουσιαστικά έτρεξε τέσσερις φορές για κάθε γονίδιο (δέντρο rooted/unrooted, χρήση/μη χρήση ασαφών δεδομένων).
- Για όλα τα γονίδια χρησιμοποιήθηκαν 19 βαθμοί ελευθερίας για το rooted δέντρο βιβλιογραφίας (Δηλαδή 20 πιθανά  $\omega$  για το M1, ένα ανά κλαδί, μείον 1  $\omega$  για το M0) ή αντίστοιχα 18 για το unrooted δέντρο (ένα λιγότερο κλαδί/ $\omega$ ).
- Για όλους τους υπολογισμούς, χρησιμοποιήθηκε αρχικός ρυθμός μεταπτώσεων/μεταστροφών = 2, ο οποίος είναι και ο συνηθέστερα παρατηρούμενος στη φύση<sup>19</sup>, επιτρεπόταν όμως στο πρόγραμμα να μεταβάλλει τον ρυθμό αυτό κατά βούληση, κατά την ανάλυση του κάθε σετ αλληλουχιών.

### Μοντέλο θέσεων

Η δεύτερη μελέτη έγινε χρησιμοποιώντας το μοντέλο θέσεων του CODEML. Προκειμένου να βρούμε ενδείξεις θετικής επιλογής σε συγκεκριμένες αμινοξικές θέσεις κάθε γονιδίου, συγκρίναμε τα μοντέλα M7 (beta κατανομή  $\omega$ , και  $\omega \leq 1$ ) και M8 (beta κατανομή  $\omega$ , και  $\omega > 1$ ).

Η σύγκριση αυτή έγινε τόσο για τα πυρηνικά όσο και για τα μιτοχονδριακά γονίδια, και χρησιμοποιήθηκαν:

- Rooted/Unrooted δέντρα βιβλιογραφίας, ασαφή/μη ασαφή δεδομένα (βλέπε Μοντέλο Θέσεων).
- Για όλα τα γονίδια χρησιμοποιήθηκαν 2 βαθμοί ελευθερίας για τη σύγκριση των μοντέλων, σύμφωνα με το εγχειρίδιο του PAML.
- Για όλους τους υπολογισμούς, χρησιμοποιήθηκε ρυθμός μεταπτώσεων/μεταστροφών = 2, επιτρεπόταν όμως στο πρόγραμμα να μεταβάλλει τον ρυθμό αυτό κατά βούληση, κατά την ανάλυση του κάθε σετ αλληλουχιών.

Εάν τα LLR είναι στατιστικά σημαντικά (δηλαδή υπάρχουν τέτοιες θέσεις μέσα στις αλληλουχίες), τότε χρησιμοποιείται ο αλγόριθμος Bayes Empirical Bayes (BEB) προκειμένου να εντοπιστούν οι θέσεις αυτές, και να υπολογιστούν οι αντίστοιχες πιθανότητες για την κάθε μια. Θέσεις με τιμές BEB *probability* > 95%, θεωρείται ότι βρίσκονται κάτω από την επίδραση θετικής επιλογής<sup>65</sup>. Οι τιμές BEB ωστόσο δεν είναι αξιόπιστες στην περίπτωση που τα LLR δεν έχουν στατιστικά σημαντικές τιμές.

### Z-tests of Selection

Για όλα τα γονίδια, πραγματοποιήθηκε επίσης το Z-test of Selection, με τη βοήθεια του MEGA4. Πρόκειται για ένα στατιστικό τεστ, το οποίο προσπαθεί να ανιχνεύσει ίχνη επιλογής λαμβάνοντας υπόψιν τους ρυθμούς συνώνυμων και μη συνώνυμων αντικαταστάσεων μέσα στις αλληλουχίες. Αυτό πραγματοποιείται συγκρίνοντας το null hypothesis test ( $dN=dS$ ) με είτε το positive selection test ( $dN>dS$ ), είτε με το negative selection test ( $dN<dS$ ). Στην περίπτωση μας, έγινε σύγκριση του null hypothesis ( $H_0: dN=dS$ ) με το positive selection hypothesis ( $H_1: dN>dS$ ), χρησιμοποιώντας το Z-test:  $Z=(dN-dS)/\text{SQRT}(\text{Var}(dS)+\text{Var}(dN))$ . Σαν επιμέρους παράμετροι, χρησιμοποιήθηκαν bootstrap 500, ο αλγόριθμος Nei-Gojobori p-distance, και οι ασαφείς θέσεις/κενά αφαιρέθηκαν πριν από κάθε pairwise ανάλυση.

### Μοντέλο κλαδιών - θέσεων

Προκειμένου να επεβεβαιώσουμε τα αποτελέσματα του MEGA4 μέσω του PAML, τρέξαμε και το μοντέλο κλαδιών-θέσεων στα γονίδια για τα οποία το Z-test ήταν θετικό. Επίσης, τρέξαμε το μοντέλο αυτό, σε γονίδια στα οποία το μοντέλο θέσεων έδωσε αυξημένο  $\omega$  σε συγκεκριμένα κλαδιά. Προκειμένου να βρούμε ενδείξεις θετικής επιλογής σε ένα συγκεκριμένο κλαδί (foreground), και σε συγκεκριμένες θέσεις αυτού, συγκρίναμε το null model, το οποίο επιτρέπει στο foreground κλαδί να έχει διαφορετικό  $\omega$  από το υπόλοιπο δέντρο, χωρίς κανένα από τα δυο όμως να περιέχει θέσεις κάτω από θετική επιλογή, με το model A, το οποίο επιτρέπει στο foreground κλαδί να έχει τέτοιες θέσεις. Για τις περιπτώσεις που χρησιμοποιήθηκε μοντέλο κλαδιών-θέσεων, χρησιμοποιήθηκαν μόνο unrooted δέντρα, όχι ασαφή δεδομένα, και ένας βαθμός ελευθερίας, σύμφωνα με το εγχειρίδιο του PAML.

Όλοι οι υπολογισμοί έγιναν σε τοπικό περιβάλλον Windows, χρησιμοποιώντας ένα αυτοματοποιημένο script γραμμένο σε Java (παράρτημα Κώδικα).

### 4.3 Αποτελέσματα

Στους Πίνακες 8-11, φαίνονται όλα τα γονίδια, πυρηνικά και μιτοχονδριακά, στα οποία το μοντέλο κλαδιών M1 ταιριάζει πιο καλά στα δεδομένα σε σχέση με το μοντέλο M0, με στατιστικά σημαντικό σκορ ( $p \leq 0.05$ ). Αυτό πρακτικά σημαίνει πως δεν ισχύει η ίδια ένταση επιλογής για όλα τα κλαδιά του δέντρου, κάτι το οποίο θα μπορούσε να αποτελέσει ένδειξη άλλου είδους επιλογής (ουδέτερη ή θετική) για κάποια από αυτά.

Ο κάθε πίνακας εξηγείται ως εξής: Στην πρώτη στήλη είναι το όνομα του κάθε γονιδίου για το οποίο το πρόγραμμα έχει θεωρήσει ότι η ένταση της επιλογής αλλάζει ανάμεσα στα κλαδιά του φυλογενετικού δέντρου. Γονίδια για τα οποία η ένταση της επιλογής είναι παρόμοια σε όλα τα κλαδιά, δεν εμφανίζονται. Η δεύτερη στήλη δείχνει το είδος του φυλογενετικού δέντρου βιβλιογραφίας (consensus) που χρησιμοποιήθηκε, rooted για έρριζα δέντρα, και unrooted για άρριζα. Η τρίτη στήλη δείχνει τους βαθμούς ελευθερίας που χρησιμοποιήθηκαν για τη σύγκριση των μοντέλων, που όπως ειπώθηκε, είναι ο αριθμός των ελεύθερων παραμέτρων που χρησιμοποιεί το μοντέλο. Η τέταρτη στήλη δείχνει εάν χρησιμοποιήθηκαν ή όχι, ασαφή δεδομένα στην ανάλυση. Οι επόμενες δυο στήλες, δείχνουν τα Log Likelihood Ratios για το κάθε μοντέλο. Η προτελευταία στήλη είναι το τετράγωνο της διαφοράς των δυο μοντέλων. Τέλος, η τελευταία δείχνει αν η τιμή  $\chi^2$  της προηγούμενης στήλης είναι στατιστικά σημαντική σε σχέση με τους βαθμούς ελευθερίας που χρησιμοποιήθηκαν. Εάν το p-value είναι  $<0.05$ , τότε θεωρούμε ότι το μοντέλο M1 εφαρμόζει καλύτερα στα δεδομένα μας, και άρα υπάρχει στατιστικά σημαντική ένδειξη αλλαγής στην ένταση επιλογής σε κάποια κλαδιά του δέντρου, για το συγκεκριμένο γονίδιο.

Η χρήση έρριζου δέντρου και ασαφών δεδομένων, έδωσε 9 τέτοια πυρηνικά και 8 μιτοχονδριακά γονίδια. Η μη χρήση ασαφών δεδομένων έδωσε τα ίδια αποτελέσματα, με ένα παραπάνω πυρηνικό γονίδιο, το *NDUFB8*. Η χρήση άρριζου δέντρου και ασαφών δεδομένων έδωσε 12 πυρηνικά και 7 μιτοχονδριακά γονίδια, ενώ η χρήση μη ασαφών δεδομένων, έδωσε παρόμοια αποτελέσματα, με εξαίρεση τα γονίδια *NDUFAB* και *NDUFV2*, ενώ έδωσε επιπλέον το γονίδιο *QCR2*.

**Πίνακας 8: Επιλεγμένα γονίδια για rooted δέντρα, ασαφή δεδομένα**

GENE	Tree	DF	Ambig. Data	M0	M1	$\chi^2$	$\rho$
COX10	Rooted Consensus	19	Yes	-7123,2344	-7104,1691	38,1	0,006
COX15	Rooted Consensus	19	Yes	-6569,9258	-6527,4674	84,9	0,000
COX6A	Rooted Consensus	19	Yes	-1293,9934	-1278,6868	30,6	0,044
NDUFA2	Rooted Consensus	19	Yes	-1778,0398	-1762,7762	30,5	0,045
NDUFA9	Rooted Consensus	19	Yes	-5520,3595	-5505,0684	30,6	0,045
NDUFS6	Rooted Consensus	19	Yes	-1973,6906	-1958,3877	30,6	0,044
NDUFV2	Rooted Consensus	19	Yes	-3015,7278	-2995,3785	40,7	0,003
SDHA	Rooted Consensus	19	Yes	-8381,8299	-8363,9678	35,7	0,011
SDHB	Rooted Consensus	19	Yes	-3346,2292	-3327,0685	38,3	0,005
ATP6	Rooted Consensus	19	Yes	-2350,8565	-2328,9476	43,8	0,001
ATP8	Rooted Consensus	19	Yes	-591,9530	-575,0773	33,8	0,020
COX1	Rooted Consensus	19	Yes	-5077,7170	-5051,4704	52,5	0,000
COX2	Rooted Consensus	19	Yes	-2373,4732	-2358,3903	30,2	0,049
COX3	Rooted Consensus	19	Yes	-2636,8840	-2620,1687	33,4	0,021
CYTB	Rooted Consensus	19	Yes	-4182,3212	-4148,8623	66,9	0,000
ND1	Rooted Consensus	19	Yes	-3099,7697	-3073,1955	53,1	0,000
ND5	Rooted Consensus	19	Yes	-6862,6740	-6842,6478	40,1	0,003

**Πίνακας 9: Επιλεγμένα γονίδια για rooted δέντρα, μη ασαφή δεδομένα**

GENE	Tree	DF	Ambig. Data	M0	M1	$\chi^2$	$\rho$
COX10	Rooted Consensus	19	No	-5786,2119	-5768,2628	35,9	0,011
COX15	Rooted Consensus	19	No	-6275,5807	-6233,4495	84,3	0,000
COX6A	Rooted Consensus	19	No	-1280,3382	-1264,9079	30,9	0,042
NDUFA2	Rooted Consensus	19	No	-1461,0891	-1445,7543	30,7	0,044
NDUFA9	Rooted Consensus	19	No	-4269,3331	-4254,2198	30,2	0,049
NDUFB8	Rooted Consensus	19	No	-2663,2879	-2647,5930	31,4	0,036
NDUFS6	Rooted Consensus	19	No	-1941,8795	-1926,6323	30,5	0,046
NDUFV2	Rooted Consensus	19	No	-2741,1856	-2725,6000	31,2	0,039
SDHA	Rooted Consensus	19	No	-7878,4576	-7861,4172	34,1	0,018
SDHB	Rooted Consensus	19	No	-3221,1967	-3203,2448	35,9	0,011
ATP6	Rooted Consensus	19	No	-2343,9006	-2321,9866	43,8	0,001
ATP8	Rooted Consensus	19	No	-591,9530	-575,5077	32,9	0,025
COX1	Rooted Consensus	19	No	-5071,9350	-5041,4025	61,1	0,000
COX2	Rooted Consensus	19	No	-2351,5970	-2336,1545	30,9	0,041
COX3	Rooted Consensus	19	No	-2592,8769	-2575,8586	34,0	0,018
CYTB	Rooted Consensus	19	No	-4152,9454	-4119,1651	67,6	0,000
ND1	Rooted Consensus	19	No	-3033,0023	-3006,2789	53,4	0,000
ND5	Rooted Consensus	19	No	-6607,2201	-6588,4847	37,5	0,007

**Πίνακας 10: Επιλεγμένα γονίδια για unrooted δέντρα, ασαφή δεδομένα**

GENE	Tree	DF	Ambig. Data	M0	M1	$\chi^2$	$\rho$
COX10	Unrooted Consensus	18	Yes	-7123,2344	-7104,4659	37,5	0,004
COX15	Unrooted Consensus	18	Yes	-6569,9258	-6528,1296	83,6	0,000
COX5B	Unrooted Consensus	18	Yes	-1199,4903	-1184,8649	29,3	0,045
COX6A	Unrooted Consensus	18	Yes	-1293,9934	-1278,9487	30,1	0,036
NDUFA2	Unrooted Consensus	18	Yes	-1778,0398	-1763,0889	29,9	0,038
NDUFA9	Unrooted Consensus	18	Yes	-5520,3595	-5505,3247	30,1	0,037
NDUFAB	Unrooted Consensus	18	Yes	-2501,9931	-2487,0707	29,8	0,039
NDUFB8	Unrooted Consensus	18	Yes	-2710,1126	-2695,4550	29,3	0,044
NDUFS6	Unrooted Consensus	18	Yes	-1973,6906	-1958,6005	30,2	0,036
NDUFV2	Unrooted Consensus	18	Yes	-3015,7278	-2997,1096	37,2	0,005
SDHA	Unrooted Consensus	18	Yes	-8381,8299	-8364,0456	35,6	0,008
SDHB	Unrooted Consensus	18	Yes	-3346,2292	-3327,1882	38,1	0,004
ATP6	Unrooted Consensus	18	Yes	-2350,8565	-2329,5859	42,5	0,001
ATP8	Unrooted Consensus	18	Yes	-591,9530	-575,5921	32,7	0,018
COX1	Unrooted Consensus	18	Yes	-5077,7170	-5050,5417	54,4	0,000
COX3	Unrooted Consensus	18	Yes	-2636,8840	-2621,0927	31,6	0,025
CYTB	Unrooted Consensus	18	Yes	-4182,3212	-4151,4276	61,8	0,000
ND1	Unrooted Consensus	18	Yes	-3099,7697	-3073,3773	52,8	0,000
ND5	Unrooted Consensus	18	Yes	-6862,6740	-6843,7673	37,8	0,004

**Πίνακας 11:** Επιλεγμένα γονίδια για unrooted δέντρα, μη ασαφή δεδομένα

GENE	Tree	DF	Ambig. Data	M0	M1	$\chi^2$	$\rho$
COX10	Unrooted Consensus	18	No	-5786,2119	-5768,4358	35,6	0,008
COX15	Unrooted Consensus	18	No	-6275,5807	-6233,5802	84,0	0,000
COX5B	Unrooted Consensus	18	No	-1194,5060	-1179,6551	29,7	0,040
COX6A	Unrooted Consensus	18	No	-1280,3382	-1264,9140	30,8	0,030
NDUFA2	Unrooted Consensus	18	No	-1461,0891	-1445,8773	30,4	0,033
NDUFA9	Unrooted Consensus	18	No	-4269,3331	-4254,2672	30,1	0,036
NDUFB8	Unrooted Consensus	18	No	-2663,2879	-2647,6003	31,4	0,026
NDUFS6	Unrooted Consensus	18	No	-1941,8795	-1926,7814	30,2	0,035
QCR2	Unrooted Consensus	18	No	-6349,2238	-6334,6138	29,2	0,046
SDHA	Unrooted Consensus	18	No	-7878,4576	-7861,7617	33,4	0,015
SDHB	Unrooted Consensus	18	No	-3221,1967	-3203,3785	35,6	0,008
ATP6	Unrooted Consensus	18	No	-2343,9006	-2322,6250	42,6	0,001
ATP8	Unrooted Consensus	18	No	-591,9530	-575,5921	32,7	0,018
COX1	Unrooted Consensus	18	No	-5071,9350	-5046,0068	51,9	0,000
COX3	Unrooted Consensus	18	No	-2592,8769	-2576,8308	32,1	0,021
CYTB	Unrooted Consensus	18	No	-4152,9454	-4120,2620	65,4	0,000
ND1	Unrooted Consensus	18	No	-3033,0023	-3006,6021	52,8	0,000
ND5	Unrooted Consensus	18	No	-6607,2201	-6589,8411	34,8	0,010

Στον Πίνακα 12, φαίνονται τα γονίδια που έδωσαν στατιστικά σημαντικά αποτελέσματα, και για τις τέσσερις διαφορετικές μεθόδους. Στην πρώτη στήλη είναι το όνομα του γονιδίου, ενώ στις στήλες 2-5 φαίνονται τα p-values από τους πίνακες 8-11 αντίστοιχα.

**Πίνακας 12:** Συγκεντρωτικός πίνακας των γονιδίων με ενδείξεις θετικής επιλογής.

GENE	$\rho$ (r,a)	$\rho$ (r,na)	$\rho$ (ur,a)	$\rho$ (ur,na)
COX10	0,006	0,011	0,004	0,008
COX15	0,000	0,000	0,000	0,000
COX6A	0,044	0,042	0,036	0,030
NDUFA2	0,045	0,044	0,038	0,033
NDUFA9	0,045	0,049	0,037	0,036
NDUFS6	0,044	0,046	0,036	0,035
SDHA	0,011	0,018	0,008	0,015
SDHB	0,005	0,011	0,004	0,008
ATP6	0,001	0,001	0,001	0,001
ATP8	0,020	0,025	0,018	0,018
COX1	0,000	0,000	0,000	0,000
COX3	0,021	0,018	0,025	0,021
CYTB	0,000	0,000	0,000	0,000
ND1	0,000	0,000	0,000	0,000
ND5	0,003	0,007	0,004	0,010

Στον Πίνακα 13, φαίνονται τα κλαδιά στα οποία το  $\omega$  είναι μεγαλύτερο της μονάδας. Τέτοια κλαδιά βρέθηκαν σε όλα τα γονίδια του Πίνακα 12, με εξαίρεση το *NDUFA9*, όπου απλά υπάρχουν διαφορές στους ρυθμούς  $\omega$  μεταξύ των κλαδιών, χωρίς όμως να φαίνεται κάπου θετική επιλογή. Επαναλαμβανόμενες ενδείξεις θετικής επιλογής φαίνεται να υπάρχουν ανάμεσα στις *D. pseudoobscura* και *D. persimilis* (*COX10*, *COX15*, *COX6A*, *NDUFS6*, *ATP6*). Για τα τρία COX γονίδια, η θετική επιλογή βρίσκεται στο κλαδί της *D. persimilis*, και δεδομένου ότι και τα τρία ανήκουν στο ίδιο σύμπλοκο, είναι πολύ πιθανόν να αλληλεπιδρούν, και η αλληλεπίδραση αυτή να ωθεί την συνεξέλιξη τους, και να εξισώνει την ένταση της επιλογής που δέχονται. Για 5 γονίδια υπάρχει θετική επιλογή μέσα ή γύρω από το melanogaster subgroup. Έτσι, για τα γονίδια *COX6A*, *NDUFA2* και *ND5* φαίνεται θετική επιλογή στο κλαδί του προγόνου του melanogaster group, μετά τον διαχωρισμό του από τον πρόγονο του obscura group, ενώ για τα γονίδια *NDUFA2* και *CYTB*, η επιλογή φαίνεται στο κλαδί όπου ο κοινός πρόγονος των *D. yakuba* και *D. erecta* διαχωρίστηκε από τον κοινό πρόγονο του melanogaster subgroup. Αρκετές ενδείξεις θετικής επιλογής υπήρχαν

στους κοινούς προγόνους των υπογενών *Drosophila* και *Sophophora* (*SDHA*, *SDHB*, *COX1*, *COX3*, *CYTB*, *ND1*, *ND5*) (βλέπε Εικόνα 14). Τέλος, για το γονίδιο *ATP8* βρέθηκαν ενδείξεις θετικής επιλογής για 7 διαφορετικά κλαδιά του φυλογενετικού δέντρου.

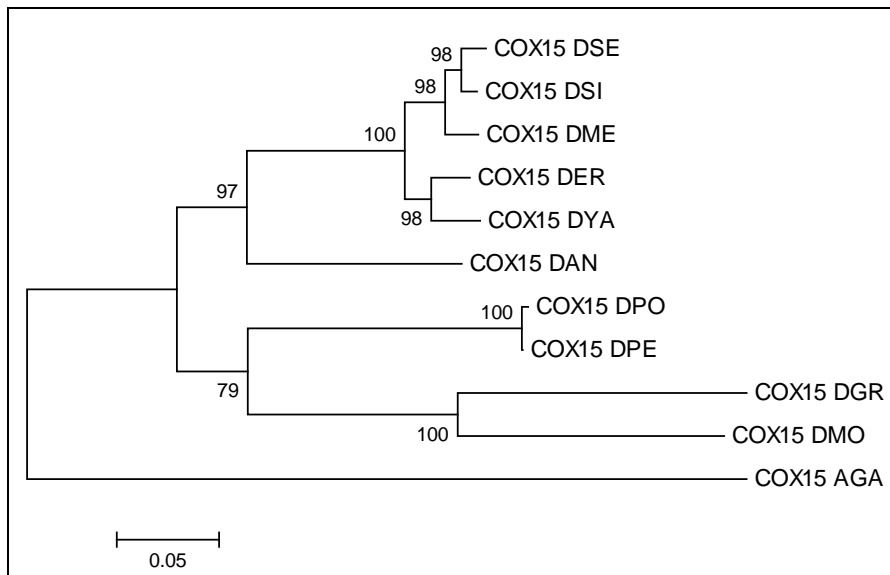
**Πίνακας 13:** Κλαδιά με ενδείξεις θετικής επιλογής για το κάθε γονίδιο (c.a. = common ancestor, ο τελευταίος κοινός πρόγονος)

GENE	POSITIVE SELECTION IN BRANCH
COX10	Between DPE and obscura group c.a.
COX15	Between DPE and obscura group c.a.
	Between DPO and obscura group c.a.
COX6A	Between melanogaster group c.a. and melanogaster - obscura c.a.
	Between DPE and obscura group c.a.
NDUFA2	Between (DYA,DER) c.a. and melanogaster subgroup c.a.
	Between melanogaster group c.a. and melanogaster - obscura c.a.
NDUFA9	No specific branch
NDUFS6	Between DPO and obscura group c.a.
	Between ((DSI,DSE),DME) c.a. and melanogaster subgroup c.a.
SDHA	Between Sophophora c.a. and Drosophila c.a.
SDHB	Between Sophophora c.a. and Drosophila c.a.
ATP6	Between DPO and obscura group c.a.
ATP8	Between 7 different lines
COX1	Between Sophophora c.a. and Drosophila c.a.
COX3	Between Sophophora c.a. and Drosophila c.a.
CYTB	Between (DYA,DER) c.a. and melanogaster subgroup c.a.
	Between Sophophora c.a. and Drosophila c.a.
ND1	Between Sophophora c.a. and Drosophila c.a.
ND5	Between melanogaster group c.a. and melanogaster - obscura c.a.
	Between Sophophora c.a. and Drosophila c.a.

Το μόνο γονίδιο που έδωσε στατιστικά σημαντικές ενδείξεις θετικής επιλογής ( $p$ -value = 0.017) με το Z-test of Positive Selection, ήταν το πυρηνικό γονίδιο COX15. Η θετική επιλογή, όπως φαίνεται στον Πίνακα 14, εντοπίζεται στα κλαδιά ανάμεσα στα είδη *D. pseudoobscura* και *D. persimilis*, και στον κοινό τους πρόγονο. Τα είδη αυτά βρίσκονται πιο κοντά φυλογενετικά μεταξύ τους, τόσο στη βιβλιογραφία, όσο και στο δέντρο του γονιδίου (Εικόνα 18).

**Πίνακας 14:** Ένδειξη θετικής επιλογής στο πυρηνικό γονίδιο COX15 (από το output του MEGA4)

	1	2	3	4	5	6	7	8	9	10	11
1. COX15 DME		20.74748	18.03009	-8.17742	20.83199	-5.62392	-5.36573	-8.78734	18.67804	17.76346	10.94582
2. COX15 DPO	1.00000		19.09305	21.43740	2.13316	21.40603	22.03781	20.91312	14.94994	14.95877	13.92942
3. COX15 DAN	1.00000	1.00000		18.42122	19.11284	17.08240	17.78522	18.38510	18.60890	18.11356	11.74894
4. COX15 DER	1.00000	1.00000	1.00000		21.60524	-8.08951	-7.46817	-6.62455	18.05921	16.76325	11.15333
5. COX15 DPE	1.00000	0.01747	1.00000	1.00000		21.43678	22.09483	21.01570	15.01229	15.12369	13.95866
6. COX15 DSE	1.00000	1.00000	1.00000	1.00000	1.00000		-3.96691	-8.43226	17.95449	18.18290	10.60844
7. COX15 DSI	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000		-8.14576	17.66162	17.68709	10.64444
8. COX15 DYA	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000		17.64122	17.38797	10.82960
9. COX15 DGR	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000		14.28156	14.70696
10. COX15 DMO	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000		15.59923
11. COX15 AGA	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	



**Εικόνα 18:** Bootstrap δέντρο NJ για το πυρηνικό γονίδιο *COX15* (διαφορετικό από το δέντρο βιβλιογραφίας)

Το μοντέλο κλαδιών έδωσε και αυτό ενδείξεις θετικής επιλογής για το συγκεκριμένο κλαδί (βλέπε Πίνακα 13). Προκειμένου να εξετάσουμε εάν η θετική επιλογή για το *COX15* επικεντρώνεται μόνο στα δυο τελικά κλαδιά ή εάν υπάρχει και στον κοινό τους πρόγονο (αν και δεν είχαμε τέτοιες ενδείξεις από το μοντέλο κλαδιών του PAML), τρέξαμε και το μοντέλο κλαδιών-θέσεων του PAML για αυτό το γονίδιο, χρησιμοποιώντας σαν foreground κλαδί τον κοινό πρόγονο (DPO,DPE). Ωστόσο, αν και ο BEB έδωσε residue με θετική επιλογή (37Q, *probability* = 0.927), το LLR της σύγκρισης των μοντέλων A - Null, δεν έδωσε στατιστικά σημαντική τιμή για την ύπαρξη θέσεων με θετική επιλογή ( $\chi^2=0$ , *p-value* = 1.0). Φαίνεται λοιπόν, πως η θετική επιλογή ξεκίνησε με τον διαχωρισμό των δυο ειδών, και όχι στον κοινό τους πρόγονο.

Η σύγκριση των μοντέλων θέσεων M7 - M8, έδωσε τέσσερα γονίδια, τρία πυρηνικά και ένα μιτοχονδριακό, όπου υπήρχε στατιστικά σημαντική ένδειξη ύπαρξης θέσεων υπό την επίδραση θετικής επιλογής (*Ndufa5*, *Ndufb9*, *Sdhb*, *Nd4*). Για αυτά τα γονίδια, χρησιμοποιήθηκε ο BEB προκειμένου να εντοπιστούν οι θέσεις αυτές. Στους Πίνακες 15-18, φαίνονται τα γονίδια, καθώς και οι θέσεις θετικής επιλογής από τον BEB. Και για τις τεσσέρις περιπτώσεις, βρέθηκαν τα ίδια γονίδια και οι ίδιες θέσεις (ο λόγος που κάποια αμινοξέα έχουν διαφορετική αρίθμηση, είναι λόγω της απαλοιφής ασαφών δεδομένων, π.χ. το *Nd4* 237S και το *Nd4* 236S στους Πίνακες 15,17 και 16,18 είναι το ίδιο αμινοξύ). Μόνο στο γονίδιο *Ndufb9* βρέθηκε τόσο ένδειξη θετικής επιλογής ( $p < 0.05$ ), όσο και συγκεκριμένη θέση (51R), με στατιστικά σημαντική πιθανότητα. Το ίδιο γονίδιο εντοπίστηκε και από τα LLR της σύγκρισης M0 - M1.

**Πίνακας 15:** Θέσεις γονιδίων με πιθανή θετική επιλογή (rooted, ασαφή δ.)

GENE	Tree	DF	Ambig. Data	M7	M8	$\chi^2$	$\rho$	<95%	>95%	>99%
NDUFA5	Rooted Consensus	2	Yes	-1723,8849	-1719,9690	7,8	0,020	25S, 66T		
NDUFB9	Rooted Consensus	2	Yes	-1934,6739	-1925,2234	18,9	0,000			51R
SDHD	Rooted Consensus	2	Yes	-3071,4510	-3067,2516	8,4	0,015	42V, 51I, 52V, 72S, 76L, 181E, 183A		
ND4	Rooted Consensus	2	Yes	-4996,4937	-4991,4732	10,0	0,007	237S		



**Πίνακας 16: Θέσεις γονιδίων με πιθανή θετική επιλογή (rooted, μη ασαφή δ.)**

GENE	Tree	DF	Ambig. Data	M7	M8	$\chi^2$	$\rho$	<95%	>95%	>99%
NDUFA5	Rooted Consensus	2	No	-1723,8849	-1719,9690	7,8	0,020	25S, 66T		
NDUFB9	Rooted Consensus	2	No	-1934,6739	-1925,2234	18,9	0,000			51R
SDHD	Rooted Consensus	2	No	-3071,4510	-3067,2516	8,4	0,015	39V, 48I, 49V, 67G, 71L		
ND4	Rooted Consensus	2	No	-4982,1183	-4977,1193	10,0	0,007	236S		

**Πίνακας 17: Θέσεις γονιδίων με πιθανή θετική επιλογή (unrooted, ασαφή δ.)**

GENE	Tree	DF	Ambig. Data	M7	M8	$\chi^2$	$\rho$	<95%	>95%	>99%
NDUFA5	Unrooted Consensus	2	Yes	-1723,8849	-1719,9690	7,8	0,020	25S, 66T		
NDUFB9	Unrooted Consensus	2	Yes	-1934,6739	-1925,2234	18,9	0,000			51R
SDHD	Unrooted Consensus	2	Yes	-3071,9626	-3067,2516	9,4	0,009	42V, 51I, 52V, 72S, 76L, 181E, 183A		
ND4	Unrooted Consensus	2	Yes	-4996,4937	-4991,4732	10,0	0,007	237S		

**Πίνακας 18: Θέσεις γονιδίων με πιθανή θετική επιλογή (unrooted, μη ασαφή δ.)**

GENE	Tree	DF	Ambig. Data	M7	M8	$\chi^2$	$\rho$	POSITIVELY SELECTED SITES	<95%	>95%	>99%
NDUFA5	Unrooted Consensus	2	No	-1714,5546	-1710,5991	7,9	0,019	25S, 66T			
NDUFB9	Unrooted Consensus	2	No	-1926,2747	-1916,8148	18,9	0,000				51R
SDHD	Unrooted Consensus	2	No	-2891,1295	-2886,9279	8,4	0,015	39V, 48I, 49V, 67G, 71L			
ND4	Unrooted Consensus	2	No	-4982,1183	-4977,1193	10,0	0,007	236S			

Η κατασκευή των δέντρων Maximum Likelihood και Neighbour Joining με βάση τις αλληλουχίες των γονιδίων (Πίνακας 18), έδωσε σε κάποιες περιπτώσεις δέντρο ίδιο με της βιβλιογραφίας, ενώ σε άλλες, το ανακατασκευασμένο δέντρο είχε διαφορετική διαρρύθμιση. Μόνο 4 Maximum Likelihood δέντρα γονιδίου από τα 39 συνολικά, είχαν την αποδεκτή από τη βιβλιογραφία τοπολογία, ενώ για τα δέντρα Neighbour Joining ο αριθμός αυτός αυξανόταν στα 11. Η μη ταύτιση ενός δέντρου γονιδίου με το αντίστοιχο δέντρο βιβλιογραφίας, εάν δεν προέρχεται από στοχαστικά λάθη των αλγορίθμων κατασκευής (για παράδειγμα, σε κάποιους οργανισμούς μπορεί να μην ισχύουν οι συνηθισμένοι ρυθμοί μεταπτώσεων/μεταστροφών), μπορεί να είναι ένδειξη διαφορετικών ρυθμών εξέλιξης για το γονίδιο του κάθε οργανισμού<sup>66</sup>. Ωστόσο, όπως προειπώθηκε, στις μελέτες με το PAML χρησιμοποιήθηκαν μόνο δέντρα βιβλιογραφίας, προκειμένου να υπολογιστούν τα αποτελέσματα με βάση τα σωστά κλαδιά.

**Πίνακας 19: Διαφορές μεταξύ ανακατασκευασμένων δέντρων και βιβλιογραφίας**

GENE	Max. Likelihood (unrooted)	Neighbour Joining (rooted)
ATP6	DAN classified close to obscura group	DAN classified close to obscura group, (DGR,DMO) classified close to melanogaster group
ATP8	DGR classified close to obscura group, other errors	DGR classified close to obscura group, other errors
COX1	DGR classified close to obscura group DAN classified as outgroup	DGR,DMO classified between melanogaster & obscura groups DAN classified close to root
COX2	(DGR,DMO) and (DPO,DPE) trees are switched	<b>Tree is correct</b>
COX3	DGR,DMO classified with Sophophora, DAN outgroup	DGR,DMO classified with Sophophora (DGR,DAN) close to obscura group
CYTB	DGR,DAN close to obscura group DYA,DER more distant common ancestor (DGR,DMO) and (DPO,DPE) trees are switched	DYA,DER more distant common ancestor
ND1	DAN close to (DGR,DMO)	DAN close to (DGR,DMO)
ND2	(DGR,DMO) and (DPO,DPE) trees are switched	<b>Tree is correct</b>
ND3	DMO closer to main tree than DGR	<b>Tree is correct</b>
ND4	DGR closer to main tree than DMO	DYA,DER more distant common ancestor DGR classified close to obscura group
ND4L	DYA,DER more distant common ancestor DGR classified with DAN	At least three errors
ND5	DGR closer to main tree than DMO	DGR closer to main tree than DMO DAN Sophophora outgroup
ND6	At least three errors	DYA,DER more distant common ancestor
COX4	(DPE,DPO) closer to (DMO,DGR)	(DPE,DPO) closer to (DMO,DGR)
COX5A	DYA,DER more distant common ancestor	At least three errors
COX5B	At least three errors	(DGR,DMO) and (DPO,DPE) branches are switched DYA,DER more distant common ancestor
COX6A	At least three errors	(DME,DSI) and (DER,DYA) branches are switched DAN classified close to obscura group DGR classified close to obscura group
COX6B	DGR closer to main tree than DMO	<b>Tree is correct</b>
COX10	At least three errors	<b>Tree is correct</b>
COX15	<b>Tree is correct</b>	(DPE,DPO) closer to (DMO,DGR)
COX17	(DPE,DPO) closer to (DMO,DGR) DAN closer to ((DPE,DPO),(DMO,DGR))	(DPE,DPO) closer to (DMO,DGR) DAN closer to ((DPE,DPO),(DMO,DGR))
NDUFA2	At least three errors	DAN most distant drosophila (DGR,DMO) and (DPO,DPE) branches are switched
NDUFA4	At least three errors	<b>Tree is correct</b>
NDUFA5	DAN most distant drosophila	DAN most distant drosophila DYA,DER more distant common ancestor
NDUFA8	DYA,DER more distant common ancestor	DYA,DER more distant common ancestor
NDUFA9	<b>Tree is correct</b>	<b>Tree is correct</b>
NDUFAB	DYA,DER more distant common ancestor	DYA,DER more distant common ancestor
NDUFB8	DYA,DER more distant common ancestor DMO closer to main tree than DGR	<b>Tree is correct</b>
NDUFB9	DME and (DYA,DER) branches are switched	DMO,DGR more distant common ancestor
NDUFS3	(DPE,DPO) closer to (DMO,DGR)	<b>Tree is correct</b>
NDUFS4	DYA,DER more distant common ancestor DMO closer to main tree than DGR	DMO,DGR more distant common ancestor
NDUFS6	DME and DSI branches are switched DYA,DER more distant common ancestor	DYA,DER more distant common ancestor
NDUFS8	DGR closer to main tree than DMO	<b>Tree is correct</b>
NDUFV2	<b>Tree is correct</b>	DYA,DER more distant common ancestor
QCR2	<b>Tree is correct</b>	<b>Tree is correct</b>
QCR8	DSE classified close to (DYA,DER)	DSE classified close to (DYA,DER)
SDHA	DYA,DER more distant common ancestor	<b>Tree is correct</b>
SDHB	(DPO,DPE) and DAN closer to (DMO,DGR)	(DPO,DPE) and DAN closer to (DMO,DGR)
SDHD	(DPE,DPO) closer to (DMO,DGR) DAN most distant drosophila	(DPE,DPO) closer to (DMO,DGR) DAN most distant drosophila
T.Genes: 39		Correct Trees: 11

#### 4.4 Συζήτηση

Οι μόνες ενδείξεις θετικής επιλογής οι οποίες παρατηρούνται σε αρτίγονα είδη δροσόφιλα σύμφωνα με τα αποτελέσματά μας, τόσο από το PAML όσο και από το MEGA4, είναι αυτές μεταξύ των ειδών *D. pseudoobscura* και *D. persimilis*. Αν και οι *D. pseudoobscura*, και *D. persimilis* είναι κοντινά μεταξύ τους είδη, ωστόσο είναι απομονωμένα περιβαλλοντικά μεταξύ τους, με τη *D. persimilis* να ζει σε μεγαλύτερα υψόμετρα όπου επικρατούν χαμηλότερες θερμοκρασίες. Στη δροσόφιλα, μεταλλάξεις σε γονίδια του Συμπλόκου II, έχουν εμπλακεί σε σύνδρομο υπερευαισθησίας στο οξυγόνο, και μείωσης του χρόνου ζωής<sup>67</sup>. Δεδομένης της γεωγραφικής κατανομής των διαφόρων ειδών δροσόφιλα, είναι πιθανόν αρκετά γονίδια της αναπνοής, να βρίσκονται υπό την επίδραση θετικής επιλογής, έτσι ώστε συγκεκριμένα είδη να είναι προσαρμοσμένα για συγκεκριμένα υψόμετρα. Μια πιθανή περαιτέρω

μελέτη θα μπορούσε να είναι η μέτρηση ενεργότητας των συμπλόκων ΟΧΡΗΟΣ σε διαφορετικά είδη. Οι υπόλοιπες ενδείξεις θετικής επιλογής που βρέθηκαν, αφορούσαν κομβικά σημεία του δέντρου, και κοινούς προγόνους σύγχρονων ειδών.

Το ορθόλογο του γονιδίου *Ndufa5* (*H.sapiens: NDUFA5*, *B.taurus: B13*), έχει χαρακτηριστεί ακόμα και σε πολύ μακρινά μεταξύ τους είδη, και είναι αρκετά συντηρημένο ανάμεσα σε κοντινά είδη. Ο ακριβής του ρόλος ωστόσο στο Σύμπλοκο I της αναπνευστικής αλυσίδας, αν και φαίνεται πως είναι αρκετά σημαντικός λόγω της συντήρησής του, δεν είναι απόλυτα ξεκαθαρισμένος<sup>68</sup>.

Το ορθόλογο του γονιδίου *Ndufb9* (*H.sapiens: NDUF9*, *B.taurus: B22*), εκτός από τον ρόλο του στο Σύμπλοκο I της αναπνευστικής αλυσίδας, είχε παλαιότερα εμπλακεί στο σύνδρομο BOR (Brancio-Oto-Renal syndrome), μία αυτοσωμική κυρίαρχη ασθένεια η οποία επηρεάζει τα νεφρά, τα αυτιά, και τον λαιμό<sup>69</sup>. Μια επόμενη μελέτη σε 9 οικογένειες με ιστορικό BOR, δεν κατάφερε να εντοπίσει ενοχοποιητικές μεταλλάξεις, χωρίς ωστόσο να αποκλείει εντελώς τη συμμετοχή του στην ασθένεια<sup>70</sup>.

Το ορθόλογο του γονιδίου *Sdhb* (*H.sapiens: SDHB*), κωδικοποιεί μια από τις δυο διαμεμβρανικές πρωτεΐνες του Συμπλόκου II. Εκτός από τον ρόλο του στο Σύμπλοκο II, μεταλλάξεις σε αυτό έχουν εμπλακεί στην δημιουργία παραγαγγλιωμάτων, τα οποία είναι συνήθως καλοήθεις όγκοι οι οποίοι εντοπίζονται στο νευροενδοκρινές σύστημα<sup>71</sup>.

Τέλος, κατά τη μελέτη των αποτελεσμάτων, φάνηκε πως ενώ το PAML έδωσε αρκετά γονίδια με ενδείξεις θετικής επιλογής, το MEGA4, έδωσε μόνο ένα. Αν και στα δυο χρησιμοποιήθηκαν μοντέλα διόρθωσης και παρόμοιες παράμετροι (όπως ρυθμός μεταπτώσεων/μεταστροφών), στο PAML χρησιμοποιήθηκε δέντρο βιβλιογραφίας, ενώ το MEGA4 κατασκεύασε δέντρα γονιδίου. Δείξαμε πως για τα περισσότερα γονίδια, το δέντρο βιβλιογραφίας ήταν διαφορετικό από το δέντρο γονιδίου, και ίσως για αυτόν τον λόγο, το MEGA4 να μην κατάφερε να ανιχνεύσει την θετική επιλογή σε περισσότερα γονίδια.

#### 4. ΠΑΡΑΡΤΗΜΑ - ΚΩΔΙΚΑΣ

Σ. πίνακας 1: Όλα τα γονίδια για το ζεύγος *H.sapiens P.troglodytes*

<i>OXPHOS C. I</i>	dNS	dS	dNS/dS	LENGTH	NS POSITIONS	S POSITIONS	PPI
ND1 (mt)	0,0262	0,3467	0,0756	954	717	237	8
ND2 (mt)	0,0170	0,4266	0,0398	1041	748	293	2
ND3 (mt)	0,0236	0,5518	0,0428	345	235	110	3
ND4 (mt)	0,0229	0,4106	0,0558	1377	999	378	2
ND4L (mt)	0,0047	0,3311	0,0142	294	199	95	0
ND5 (mt)	0,0371	0,4146	0,0895	1809	1344	465	3
ND6 (mt)	0,0181	0,4302	0,0421	522	363	159	1
AVERAGE	<b>0,0214</b>	<b>0,4159</b>	<b>0,0514</b>	906	658	248	
NDUFS1 (n)	0,0024	0,0174	0,1379	2181	1673	508	7
NDUFS2 (n)	0,0010	0,0030	0,3333	1389	986	403	12
NDUFS3 (n)	0,0116	0,0163	0,7117	792	587	205	12
NDUFS7 (n)	0,0021	0,0186	0,1129	639	530	109	2
NDUFS8 (n)	0,0000	0,0065	0,0000	630	514	116	7
NDUFV1 (n)	0,0010	0,0143	0,0699	1392	1054	338	7
NDUFV2 (n)	0,0035	0,0056	0,6250	747	562	185	8
NDUFA2 (n)	0,0131	0,0000	N/A	207	160	47	29
NDUFA4 (n)	0,0000	0,0000	N/A	243	169	74	1
NDUFA5 (n)	0,0000	0,0000	N/A	348	247	101	2
NDUFA6 (n)	0,0000	0,0230	0,0000	384	281	103	2
NDUFA7 (n)	0,0000	0,0340	0,0000	339	254	85	2
NDUFA8 (n)	0,0000	0,0000	N/A	516	365	151	2
NDUFA9 (n)	0,0024	0,0218	0,1101	1131	809	322	5
NDUFA10 (n)	0,0037	0,0249	0,1486	1065	758	307	2
NDUFA11 (n)	0,0000	0,0091	0,0000	423	316	107	1
NDUFAB1 (n)	0,0000	0,0182	0,0000	465	339	127	32
NDUFB1 (n)	0,0000	0,0133	0,0000	315	220	95	2
NDUFB2 (n)	0,0084	0,0000	N/A	315	229	86	1
NDUFB3 (n)	0,0000	0,0000	N/A	294	220	74	2
NDUFB4 (n)	0,0000	0,0000	N/A	387	269	118	1
NDUFB5 (n)	0,0000	0,0294	0,0000	567	409	158	2
NDUFB7 (n)	0,0032	0,0215	0,1488	408	358	51	2
NDUFB8 (n)	0,0023	0,0079	0,2911	458	415	143	2
NDUFB9 (n)	0,0024	0,0000	N/A	537	380	157	2
NDUFB10 (n)	0,0049	0,0190	0,2579	516	412	104	2
NDUFC2 (n)	0,0037	0,0115	0,3217	357	247	110	2
NDUFS4 (n)	0,0000	0,0251	0,0000	519	370	149	2
NDUFS5 (n)	0,0000	0,0000	N/A	318	230	88	2
NDUFS6 (n)	0,0000	0,0111	0,0000	372	276	96	2
NDUFV3 (n)	0,0000	0,0125	0,0000	324	222	103	2
AVERAGE	<b>0,0021</b>	<b>0,0117</b>	<b>0,1805</b>	603	447	155	
NDUFA3 (n)	0,1179	0,2578	0,4573	252	202	50	
<i>OXPHOS C.II</i>	dNS	dS	dNS/dS	LENGTH	NS POSITIONS	S POSITIONS	
SDHB (n)	0,0000	0,0210	0,0000	837	638	199	2
SDHC (n)	0,0000	0,0000	N/A	348	249	99	2
SDHD (n)	0,0057	0,0000	N/A	477	334	143	2
AVERAGE	<b>0,0019</b>	<b>0,0070</b>	<b>0,2714</b>	554	407	147	
SDHA (n)	0,0388	0,0906	0,4283	282	212	70	
<i>OXPHOS C.III</i>	dNS	dS	dNS/dS	LENGTH	NS POSITIONS	S POSITIONS	
CYTb (mt)	<b>0,0360</b>	<b>0,4829</b>	<b>0,0745</b>	1140	877	263	10
CYC1 (n)	0,0000	0,0239	0,0000	975	722	253	9
UCRC (n)	0,0000	0,0000	N/A	186	158	28	2
UQCR (n)	0,0000	0,0000	N/A	168	134	34	8
UQCRCB (n)	0,0511	0,1078	0,4740	333	256	77	8
UQCRC1 (n)	0,0000	0,0205	0,0000	1380	992	388	8
UQCRC2 (n)	0,0024	0,0072	0,3333	1125	801	324	8
UQCRCF1 (n)	0,0033	0,0143	0,2308	822	582	241	8
UQCRH (n)	0,0188	0,0176	1,0682	273	179	94	8
UQCRCQ (n)	0,0000	0,0357	0,0000	246	182	64	8
AVERAGE	<b>0,0084</b>	<b>0,0252</b>	<b>0,3330</b>	612	445	167	
<i>OXPHOS C.IV</i>	dNS	dS	dNS/dS	LENGTH	NS POSITIONS	S POSITIONS	
COX1 (mt)	0,0061	0,4309	0,0142	1539	1117	422	10
COX2 (mt)	0,0098	0,4827	0,0203	681	486	195	10
COX3 (mt)	0,0120	0,4640	0,0259	783	568	215	8
AVERAGE (mt)	<b>0,0093</b>	<b>0,4592</b>	<b>0,0203</b>	1001	724	277	
COX4I1 (n)	0,0076	0,0000	N/A	513	418	95	2
COX5A (n)	0,0090	0,0087	1,0345	450	309	141	8
COX5B (n)	0,0141	0,0000	N/A	387	272	116	8
COX6A1 (n)	0,0124	0,0242	0,5124	327	235	92	8
COX6A2 (n)	0,0000	0,0129	0,0000	291	232	59	0
COX6B1 (n)	0,0000	0,0193	0,0000	258	206	53	8
COX6C (n)	0,0058	0,0196	0,2959	225	165	60	8
COX7C (n)	0,0000	0,0000	N/A	189	131	58	8
COX8A (n)	0,0068	0,0340	0,2000	207	156	51	0
COX10 (n)	0,0000	0,0178	0,0000	1329	924	405	0
COX11 (n)	0,0101	0,0227	0,4449	534	389	145	0
COX17 (n)	0,0208	0,0000	N/A	189	136	53	0
AVERAGE (n)	<b>0,0072</b>	<b>0,0133</b>	<b>0,5440</b>	408	298	111	
<i>OXPHOS C.V</i>	dNS	dS	dNS/dS	LENGTH	NS POSITIONS	S POSITIONS	
ATP6 (mt)	0,0286	0,3231	0,0885	678	460	218	4
ATP8 (mt)	0,0265	0,3476	0,0762	204	140	64	0
AVERAGE	<b>0,0276</b>	<b>0,3354</b>	<b>0,0822</b>	441	300	141	
ATP5A1 (n)	0,0000	0,0143	0,0000	1659	1187	472	9
ATP5B (n)	0,0035	0,0176	0,1989	1554	1053	502	10
ATP5C1 (n)	0,0000	0,0096	0,0000	891	632	259	6
ATP5D (n)	0,0000	0,0293	0,0000	504	403	102	8
ATP5E (n)	0,0258	0,0294	0,8776	153	109	44	4
ATP5F1 (n)	0,0019	0,0057	0,3333	714	494	220	11
ATP5G1 (n)	0,0000	0,0102	0,0000	381	268	113	3
ATP5G2 (n)	0,0000	0,0206	0,0000	360	247	113	2
ATP5G3 (n)	0,0000	0,0000	N/A	426	294	132	3

ATP5H (n)	0,0000	0,0000	N/A	411	288	124	2
ATP5I (n)	0,0000	0,0198	0,0000	207	148	59	1
ATP5J2 (n)	0,0000	0,0379	0,0000	237	178	60	1
ATP5J (n)	0,0093	0,0471	0,1975	282	200	82	2
ATP5L (n)	0,0000	0,0000	N/A	309	224	85	1
ATP5O (n)	0,0000	0,0251	0,0000	639	447	192	6
AVERAGE	<b>0,0027</b>	<b>0,0178</b>	<b>0,1519</b>	582	411	171	
<b>GLYCOLYSIS</b>	<b>dNS</b>	<b>dS</b>	<b>dNS/dS</b>	<b>LENGTH</b>	<b>NS POSITIONS</b>	<b>S POSITIONS</b>	
ACSS1	0,0025	0,0184	0,1359	1704	1322	382	
ACSS2	0,0053	0,0124	0,4274	1326	955	371	
ADH1A	0,0037	0,0073	0,5068	1125	784	342	
ADH1B	0,0039	0,0164	0,2378	1125	784	341	
ADH1C	0,0026	0,0362	0,0718	1023	713	310	
ADH4	0,0059	0,0213	0,2770	1086	774	312	
ADH5	0,0042	0,0063	0,6667	1020	730	290	
ADH6	0,0010	0,0127	0,0787	828	585	243	
ADH7	0,0066	0,0157	0,4204	951	672	279	
AKR1A1	0,0015	0,0063	0,2381	855	614	241	
ALDH1A3	0,0010	0,0175	0,0571	1353	988	365	
ALDH1B1	0,0023	0,0231	0,0996	1551	1178	373	
ALDH2	0,0000	0,0561	0,0000	1296	995	302	
ALDH3A1	0,0000	0,0035	0,0000	1140	939	201	
ALDH3A2	0,0018	0,0536	0,0336	1245	906	339	
ALDH7A1	0,0047	0,0141	0,3333	1497	1135	362	
ALDOA	0,0010	0,0064	0,1563	1092	858	234	
ALDOB	0,0059	0,0150	0,3933	1008	714	294	
ALDOC	0,0024	0,0257	0,0934	1092	775	317	
BPGM	0,0026	0,0241	0,1079	777	541	236	
DLAT	0,0037	0,0191	0,1937	1941	1386	555	
DLD	0,0026	0,0041	0,6341	1230	950	280	
ENO1	0,0016	0,0051	0,3137	1302	955	348	
ENO3	0,0000	0,0263	N/A	1302	983	319	
FBP1	0,0016	0,0096	0,1667	1014	745	269	
FBP2	0,0010	0,0187	0,0535	1017	714	303	
G6PC2	0,0026	0,0106	0,2453	438	321	117	
GALM	0,0000	0,0130	0,0000	633	451	182	
GAPDH	0,0012	0,0148	0,0811	888	687	201	
GCK	0,0031	0,0211	0,1469	1353	1138	215	
GPI	0,0035	0,0166	0,2108	804	652	152	
HK1	0,0033	0,0173	0,1908	2691	2024	667	
HK2	0,0045	0,0152	0,2961	2667	2047	620	
HK3	0,0000	0,0101	0,0000	2376	1786	590	
LDHA	0,0000	0,0261	0,0000	822	592	230	
LDHAL6A	0,0020	0,0000	N/A	996	720	276	
LDHAL6B	0,0000	0,0142	0,0000	1143	816	327	
LDHB	0,0016	0,0106	0,1509	1002	716	286	
LDHC	0,0000	0,0044	0,0000	996	741	255	
PCK1	0,0023	0,0114	0,2018	1866	1436	430	
PCK2	0,0013	0,0043	0,3023	963	738	225	
PDHA1	0,0000	0,0043	0,0000	1170	832	338	
PDHA2	0,0028	0,0334	0,0838	1164	833	331	
PDHB	0,0042	0,0165	0,2545	882	634	248	
PFKL	0,0011	0,0296	0,0372	1989	1631	358	
PFKM	0,0011	0,0036	0,3056	2340	1682	658	
PFKP	0,0000	0,0290	0,0000	2238	1736	502	
PGAM1	0,0000	0,0256	0,0000	684	547	137	
PGAM2	0,0000	0,0056	0,0000	759	642	117	
PGAM4	0,0034	0,0170	0,2000	762	563	199	
PGK1	0,0000	0,0067	0,0000	1251	892	359	
PGK2	0,0021	0,0206	0,1019	1251	894	357	
PKLR	0,0026	0,0048	0,5417	1593	1205	388	
TPI1	0,0000	0,0222	0,0000	747	548	199	
AVERAGE	<b>0,0021</b>	<b>0,0164</b>	<b>0,1269</b>	1248	930	318	

Σ. πίνακας 2: Όλα τα γονίδια για το ζεύγος *D.melanogaster* – *D.simulans*

<b>OXPHOS C. I</b>	<b>dNS</b>	<b>dS</b>	<b>dNS/dS</b>	<b>LENGTH</b>	<b>NS POSITIONS</b>	<b>S POSITIONS</b>	
ND1 (mt)	0,0124	0,1700	0,0729	936	864	72	
ND2 (mt)	0,0199	0,1360	0,1463	1023	929	94	
ND3 (mt)	0,0108	0,2317	0,0466	351	307	44	
ND4 (mt)	0,0164	0,1694	0,0968	1338	1182	156	
ND4L (mt)	0,0087	0,1961	0,0444	288	257	31	
ND5 (mt)	0,0180	0,1965	0,0916	1713	1542	171	
ND6 (mt)	0,0246	0,1192	0,2064	522	462	60	
AVERAGE	<b>0,0158</b>	<b>0,1741</b>	<b>0,0909</b>	882	792	90	
NDUFA2 (n)	0,0199	0,1607	0,1238	273	204	69	
NDUFA4 (n)	0,0053	0,0530	0,1000	249	187	62	
NDUFA5 (n)	0,0035	0,1479	0,0237	372	294	78	
NDUFA6 (n)	0,0034	0,0617	0,0551	378	303	75	
NDUFA8 (n)	0,0000	0,1257	0,0000	525	414	111	
NDUFA9 (n)	0,0054	0,0898	0,0601	981	811	170	
NDUFAB1 (n)	0,0000	0,0755	0,0000	429	357	72	
NDUFB7 (n)	0,0149	0,0521	0,2860	351	317	35	
NDUFB8 (n)	0,0100	0,0975	0,1026	525	414	111	
NDUFB9 (n)	0,0122	0,1067	0,1143	432	342	90	
NDUFS2 (n)	0,0045	0,1530	0,0294	582	438	145	
NDUFS3 (n)	0,0034	0,1026	0,0331	795	613	182	
NDUFS4 (n)	0,0108	0,0934	0,1156	549	476	73	
NDUFS6 (n)	0,0034	0,0617	0,0551	378	303	75	
NDUFS7 (n)	0,0062	0,1848	0,0335	636	513	123	
NDUFS8 (n)	0,0050	0,0644	0,0776	651	526	125	
NDUFV1 (n)	0,0057	0,1502	0,0379	1386	1107	279,5	
NDUFV2 (n)	0,0000	0,0981	0,0000	726	593	133	
AVERAGE	<b>0,0063</b>	<b>0,1044</b>	<b>0,0605</b>	568	456	112	
<b>OXPHOS C. II</b>	<b>dNS</b>	<b>dS</b>	<b>dNS/dS</b>	<b>LENGTH</b>	<b>NS POSITIONS</b>	<b>S POSITIONS</b>	
SDHA (n)	0,0080	0,1300	0,0615	1830	1355	476	
SDHB (n)	0,0000	0,0865	0,0000	891	820	71	
SDHD (n)	0,0050	0,0415	0,1205	546	407	139	
AVERAGE	<b>0,0043</b>	<b>0,0860</b>	<b>0,0504</b>	1089	861	228	

<i>OXPHOS C. III</i>	dNS	dS	dNS/dS	LENGTH	NS POSITIONS	S POSITIONS
CYTB (mt)	<b>0,0068</b>	<b>0,3039</b>	<b>0,0224</b>	1134	984	150
CYC1 (n)	0,0142	0,1200	0,1183	1032	796	236
UQCRB (n)	0,0039	0,1304	0,0299	333	258	75
UQCRC2 (n)	0,0045	0,0625	0,0720	1221	967	254
UQCRCF51 (n)	0,0028	0,0782	0,0358	474	383	91
UQCRQ (n)	0,0000	0,0780	0,0000	267	227	41
AVERAGE	<b>0,0051</b>	<b>0,0938</b>	<b>0,0541</b>	665	526	139
<i>OXPHOS C. IV</i>	dNS	dS	dNS/dS	LENGTH	NS POSITIONS	S POSITIONS
COX1 (mt)	0,0017	0,2118	0,0080	1530	1366	164
COX2 (mt)	0,0094	0,2280	0,0412	684	593	91
COX3 (mt)	0,0108	0,1883	0,0574	786	689	98
AVERAGE	<b>0,0073</b>	<b>0,2094</b>	<b>0,0349</b>	1000	882	118
COX4 (n)	0,0249	0,0977	0,2549	528	416	112
COX5A (n)	0,0000	0,0969	0,0000	447	375	72
COX5B (n)	0,0000	0,0924	0,0000	360	282	78
COX6A (n)	0,0319	0,1343	0,2375	780	574	206
COX6B (n)	0,0000	0,0213	0,0000	231	220	11
COX7A (n)	0,0127	0,1448	0,0877	315	236	79
COX10 (n)	0,0139	0,1210	0,1149	1173	826	347
COX11 (n)	0,0109	0,1679	0,0649	723	540	183
COX15 (n)	0,0057	0,1060	0,0538	1173	905	268
COX17 (n)	0,0056	0,0556	0,1007	234	170	64
AVERAGE	<b>0,0106</b>	<b>0,1038</b>	<b>0,1017</b>	596	454	142
<i>OXPHOS C. V</i>	dNS	dS	dNS/dS	LENGTH	NS POSITIONS	S POSITIONS
ATP6 (mt)	0,0225	0,1784	0,1261	672	587	85
ATP8 (mt)	0,0156	0,1484	0,1051	159	147	12
AVERAGE	<b>0,0191</b>	<b>0,1634</b>	<b>0,1166</b>	416	367	49
ATP5B (n)	0,0010	0,0788	0,0127	1317	956	361
ATP5C1 (n)	0,0030	0,0565	0,0531	891	762	129
ATP5D (n)	0,0028	0,0708	0,0395	471	370	101
ATP5F1 (n)	0,0018	0,0521	0,0345	729	598	131
ATP5G (n)	0,0082	0,0472	0,1737	333	256	77
ATP5H (n)	0,0097	0,0974	0,0996	534	468	66
ATP5J (n)	0,0042	0,0968	0,0434	318	256	63
ATP5O (n)	0,0036	0,0875	0,0411	378	307	71
AVERAGE	<b>0,0043</b>	<b>0,0734</b>	<b>0,0584</b>	621	497	125
<i>GLYCOLYSIS</i>	dNS	dS	dNS/dS	LENGTH	NS POSITIONS	S POSITIONS
ACSS	0,0301	0,1601	0,1880	1665	1290	375
Ald	0,0034	0,0507	0,0671	1002	841	161
DLAT	0,0082	0,1456	0,0563	1488	1070	418
DLD	0,0018	0,0633	0,0284	1512	1153	359
ENO	0,0031	0,0863	0,0359	1263	1101	162
FBP	0,0027	0,1439	0,0188	984	765	219
Fdh	0,0012	0,1413	0,0085	1137	914	223
Hex-t1	0,0055	0,1607	0,0342	1194	931	263
Hex-t2	0,0136	0,1713	0,0794	1359	1053	306
LDHa	0,0027	0,0711	0,0380	996	826	170
LDHb	0,0160	0,1945	0,0823	1014	780	234
PDHA1	0,0000	0,0583	0,0000	1170	880	290
PDHA2	0,0082	0,1480	0,0554	1428	1151	277
PDHB	0,0012	0,0917	0,0131	1095	839	256
Pepck	0,0054	0,1036	0,0521	1941	1518	423
PFK	0,0008	0,1178	0,0068	1725	1362	363
PGAM	0,0060	0,0903	0,0664	864	705	159
PGK	0,0021	0,0865	0,0243	1245	1026	219
Pyk	0,0046	0,1520	0,0303	1713	1351	362
TPI	0,0064	0,0659	0,0971	1044	890	154
AVERAGE	<b>0,0062</b>	<b>0,1151</b>	<b>0,0534</b>	1292	1022	270

### Σ. πίνακας 3: Όλα τα γονίδια για το ζεύγος *Z.mays – O.sativa*

<i>OXPHOS C. I</i>	dNS	dS	dNS/dS	LENGTH	NS POSITIONS	S POSITIONS
nad1 (mt)	0,0056	0,0159	0,3522	975	731	244
nad2 (mt)	0,0054	0,0028	1,9286	1464	1026	438
nad3 (mt)	0,0150	0,0119	1,2605	354	263	91
nad4 (mt)	0,0054	0,0194	0,2784	1485	1070	415
nad4L (mt)	0,0000	0,0136	0,0000	300	230	70
nad5 (mt)	0,0013	0,0189	0,0688	2007	1473	534
nad6 (mt)	0,0233	0,0738	0,3157	579	410	169
nad7 (mt)	0,0011	0,0035	0,3143	1182	889	293
nad9 (mt)	0,0023	0,0316	0,0728	570	461	109
AVERAGE	<b>0,0066</b>	<b>0,0213</b>	<b>0,3103</b>	991	728	263
NDUFS1 (n)	0,0222	0,5366	0,0414	2184	1579	605
NDUFS4 (n)	0,0320	0,4582	0,0698	471	336	135
NDUFS5 (n)	0,0490	0,2461	0,1991	255	217	38
NDUFS6 (n)	0,0677	0,5611	0,1207	264	204	60
NDUFS7 (n)	0,0099	0,5306	0,0187	540	434	106
NDUFS8 (n)	0,0437	0,6865	0,0637	669	505	164
NDUFV1 (n)	0,0291	0,6186	0,0470	1512	1110	402
NDUFV2 (n)	0,0263	0,5606	0,0469	765	581	184
NDUFA5 (n)	0,0286	0,5382	0,0531	519	414	105
NDUFA6 (n)	0,0801	0,5234	0,1530	393	324	69
NDUFA8 (n)	0,0484	0,4455	0,1086	288	244	44
NDUFA9 (n)	0,0707	0,6472	0,1092	1221	918	303
NDUFAB1 (n)	0,0796	0,4014	0,1983	393	293	100
NDUFB7 (n)	0,0597	0,1748	0,3415	321	294	27
NDUFB9 (n)	0,0396	0,4722	0,0839	339	282	57
AVERAGE	<b>0,0458</b>	<b>0,4934</b>	<b>0,0928</b>	676	516	160
<i>OXPHOS C. II</i>	dNS	dS	dNS/dS	LENGTH	NS POSITIONS	S POSITIONS
SDHA (n)	0,0279	0,4991	0,0559	1845	1395	450
SDHB (n)	0,0271	0,2942	0,0921	840	775	65
AVERAGE	<b>0,0275</b>	<b>0,3967</b>	<b>0,0693</b>	1343	1085	258
<i>OXPHOS C. III</i>	dNS	dS	dNS/dS	LENGTH	NS POSITIONS	S POSITIONS
COB (mt)	<b>0,0045</b>	<b>0,0221</b>	<b>0,2036</b>	1164	891	273
ISP (n)	0,0367	0,8488	0,0432	750	557,3	193
CYT1 (n)	0,0257	0,7361	0,0349	867	650	217
AVERAGE	<b>0,0312</b>	<b>0,7925</b>	<b>0,0394</b>	809	604	205
<i>OXPHOS C. IV</i>	dNS	dS	dNS/dS	LENGTH	NS POSITIONS	S POSITIONS

ccmB (mt)	0,0087	0,0000	N/A	618	434	184
ccmC (mt)	0,0019	0,0163	0,1166	720	528	193
ccmFC (mt)	0,0062	0,0131	0,4733	1284	922	362
ccmFN (mt)	0,0100	0,0139	0,7194	1743	1282	461
cox1 (mt)	0,0008	0,0321	0,0249	1566	1214	352
cox2 (mt)	0,0017	0,0217	0,0783	780	589	192
cox3 (mt)	0,0017	0,0101	0,1683	795	582	213
AVERAGE	<b>0,0044</b>	<b>0,0153</b>	<b>0,2892</b>	1072	793	279
COX5B (n)	0,0422	0,7321	0,0576	312	244	68
COX6A (n)	0,0202	0,2500	0,0808	267	227	40
COX6B (n)	0,0505	0,6552	0,0771	426	281	145
COX15 (n)	0,0594	0,6689	0,0888	1122	807	315
AVERAGE	<b>0,0431</b>	<b>0,5766</b>	<b>0,0747</b>	532	390	142
<b>OXPHOS C. V</b>	<b>dNS</b>	<b>dS</b>	<b>dNS/dS</b>	<b>LENGTH</b>	<b>NS POSITIONS</b>	<b>S POSITIONS</b>
ATP6 (mt)	0,0171	0,0339	0,5044	624	445	179
ATP9 (mt)	0,0124	0,0704	0,1761	222	155	67
AVERAGE	<b>0,0148</b>	<b>0,0522</b>	<b>0,2828</b>	423	300	123
ATPb (n)	0,0244	0,5501	0,0444	1656	1248	408
OSCP (n)	0,0713	0,4721	0,1510	663	487	176
ATPd (n)	0,0573	0,4349	0,1318	573	415	158
ATPe (n)	0,0618	0,4621	0,1337	186	168	18
ATPdel (n)	0,0791	0,6467	0,1223	501	396	105
AVERAGE	<b>0,0588</b>	<b>0,5132</b>	<b>0,1145</b>	716	543	173
ATPg (n)						
<b>GLYCOLYSIS</b>	<b>dNS</b>	<b>dS</b>	<b>dNS/dS</b>	<b>LENGTH</b>	<b>NS POSITIONS</b>	<b>S POSITIONS</b>
5422a	0,0215	0,4877	0,0441	1746	1328	418
2611a	0,0697	0,8858	0,0787	1377	1001	376
2611b	0,0936	0,7625	0,1228	1359	993	366
5133a	0,1523	0,5192	0,2933	1011	846	165
5319a	0,0489	0,5794	0,0844	1701	1267	434
5319b	0,0431	0,5514	0,0782	1701	1258	443
31311a	0,0385	0,5013	0,0768	1017	800	218
31311b	0,0380	0,3022	0,1257	1047	945	102
41213a	0,0271	0,4601	0,0589	1131	830	301
41213b	0,0257	0,4836	0,0531	903	761	142
12112a	0,0169	0,5066	0,0334	1107	807	300
12112b	0,0446	0,4372	0,1020	1011	794	217
12119a	0,0288	0,5025	0,0573	1494	1086	408
5133b	0,1008	0,4253	0,2370	948	858	90
AVERAGE	<b>0,0535</b>	<b>0,5289</b>	<b>0,1012</b>	1277	978	299

Σ. πίνακας 5: Correlations για τα γονίδια του Συμπλόκου I (μιτοχ. – πυρηνικά)

OXPHOS I (mit) - 7 genes							OXPHOS I (nucl) - 31 genes							
		dN	dS	ω	Length	PPI			dN	dS	ω	Length	PPI	
Kendall's tau_b	dN	Correlation Coefficient	1,000	,143	<b>,905**</b>	,429	<b>,751*</b>	dN	Correlation Coefficient	1,000	,045	<b>,786**</b>	,217	,262
		Sig. (2-tailed)	.	,652	<b>,004</b>	,176	<b>,021</b>		Sig. (2-tailed)	.	,752	<b>,000</b>	,114	,084
		N	7	7	7	7	7		N	31	31	22	31	31
	dS	Correlation Coefficient		1,000	,048	-,048	,050	dS	Correlation Coefficient		1,000	-,222	,254	,103
		Sig. (2-tailed)		.	,881	,881	,878		Sig. (2-tailed)		.	,172	,052	,475
		N		7	7	7	7		N		31	22	31	31
	ω	Correlation Coefficient			1,000	,524	<b>,651*</b>	ω	Correlation Coefficient			1,000	<b>,338*</b>	,294
		Sig. (2-tailed)			.	,099	<b>,046</b>		Sig. (2-tailed)			.	<b>,038</b>	,101
		N			7	7	7		N			22	22	22
	Length	Correlation Coefficient				1,000	,350	Length	Correlation Coefficient				1,000	<b>,393**</b>
		Sig. (2-tailed)				.	,282		Sig. (2-tailed)				.	,005
		N				7	7		N				31	31
PPI	Correlation Coefficient					1,000	PPI	Correlation Coefficient					1,000	
	Sig. (2-tailed)					.		Sig. (2-tailed)					.	
	N					7		N					31	

**Σ. πίνακας 6: Correlations για τα πυρηνικά γονίδια των Συμπλόκων II-III**

OXPHOS II (nucl) - 3 genes

		dN	dS	ω	Length	PPI
Kendall's tau_b	dN	Correlation Coefficient	1,000	-.500	.	,000
		Sig. (2-tailed)	.	,480	.	1,000
		N	3	3	1	3
	dS	Correlation Coefficient		1,000	.	,816
		Sig. (2-tailed)		.	.	,221
		N		3	1	3
	ω	Correlation Coefficient			.	.
		Sig. (2-tailed)			.	.
		N		1	1	1
	Length	Correlation Coefficient			1,000	.
		Sig. (2-tailed)			.	.
		N			3	3
	PPI	Correlation Coefficient				.
		Sig. (2-tailed)				.
		N				3

OXPHOS III (nucl) - 9 genes

		dN	dS	ω	Length	PPI
Kendall's tau_b	dN	Correlation Coefficient	1,000	,265	<b>,778*</b>	,065
		Sig. (2-tailed)	.	,354	<b>,022</b>	,818
		N	9	9	<b>7</b>	9
	dS	Correlation Coefficient		1,000	-.206	,197
		Sig. (2-tailed)		.	,530	,463
		N		9	7	9
	ω	Correlation Coefficient			1,000	-.385
		Sig. (2-tailed)			.	,347
		N			7	7
	Length	Correlation Coefficient				1,000
		Sig. (2-tailed)			.	,387
		N			9	9
	PPI	Correlation Coefficient				1,000
		Sig. (2-tailed)			.	.
		N				9

**Σ. πίνακας 7: Correlations για τα γονίδια του Συμπλόκου IV (μιτοχ. – πυρηνικά)**

OXPHOS IV (mit) - 3 genes

		dN	dS	ω	Length	PPI
Kendall's tau_b	dN	Correlation Coefficient	1,000	,333	1,000	-.333
		Sig. (2-tailed)	.	,602	.	,602
		N	3	3	3	3
	dS	Correlation Coefficient		1,000	,333	-1,000
		Sig. (2-tailed)		.	,602	.
		N		3	3	3
	ω	Correlation Coefficient			1,000	-.333
		Sig. (2-tailed)			.	,602
		N			3	3
	Length	Correlation Coefficient				1,000
		Sig. (2-tailed)			.	,000
		N			3	3
	PPI	Correlation Coefficient				1,000
		Sig. (2-tailed)			.	.
		N				3

OXPHOS IV (nucl) - 12 genes

		dN	dS	ω	Length	PPI
Kendall's tau_b	dN	Correlation Coefficient	1,000	-.033	<b>,760*</b>	,016
		Sig. (2-tailed)	.	,886	<b>,013</b>	,944
		N	12	12	<b>8</b>	12
	dS	Correlation Coefficient		1,000	,189	,000
		Sig. (2-tailed)		.	,524	1,000
		N		12	8	12
	ω	Correlation Coefficient			1,000	,113
		Sig. (2-tailed)			.	,702
		N			8	8
	Length	Correlation Coefficient				1,000
		Sig. (2-tailed)			.	-.155
		N			12	12
	PPI	Correlation Coefficient				1,000
		Sig. (2-tailed)			.	.
		N				12



**Σ. πίνακας 8: Correlations για τα γονίδια του Συμπλόκου V (μιτοχ. – πυρηνικά)**

		dN	dS	ω	Length	PPI
Kendall's tau_b	dN	1,000	-1,000	1,000	1,000	1,000
		Correlation Coefficient				
		Sig. (2-tailed)				
		N	2	2	2	2
	dS		1,000	-1,000	-1,000	-1,000
	Correlation Coefficient					
	Sig. (2-tailed)					
	N	2	2	2	2	
ω				1,000	1,000	1,000
	Correlation Coefficient					
	Sig. (2-tailed)					
	N			2	2	2
Length					1,000	1,000
	Correlation Coefficient					
	Sig. (2-tailed)					
	N				2	2
PPI						1,000
	Correlation Coefficient					
	Sig. (2-tailed)					
	N					2

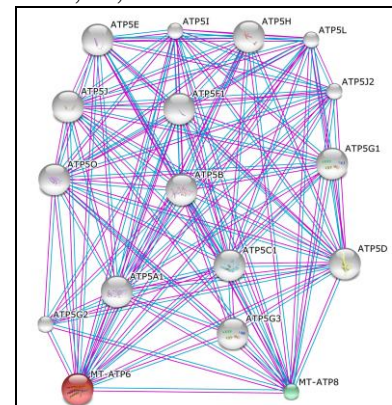
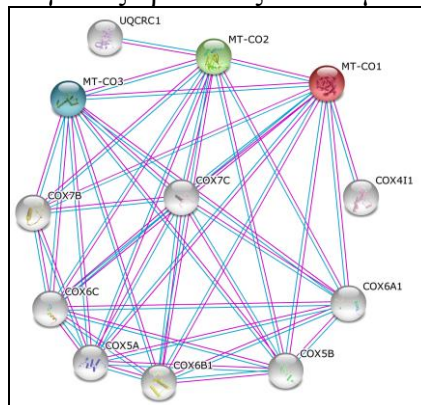
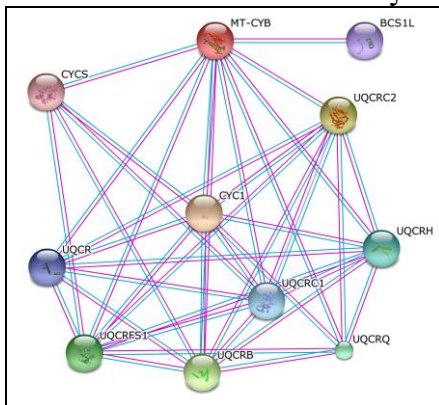
		dN	dS	ω	Length	PPI
Kendall's tau_b	dN	1,000	,280	,842**	-,083	,258
		Correlation Coefficient				
		Sig. (2-tailed)		,198	,002	,701
		N	15	15	12	15
	dS			1,000	,040	-,232
	Correlation Coefficient					
	Sig. (2-tailed)			,869	,233	,763
	N		15	12	15	
ω				1,000	-,080	,307
	Correlation Coefficient					
	Sig. (2-tailed)				,742	,214
	N			12	12	
Length					1,000	,644**
	Correlation Coefficient					
	Sig. (2-tailed)					,001
	N				15	15
PPI						1,000
	Correlation Coefficient					
	Sig. (2-tailed)					
	N					15

**Σ. πίνακας 9: Correlations για το σύνολο των γονιδίων (μιτοχ. – πυρηνικά)**

		dN	dS	ω	Length	PPI
Kendall's tau_b	dN	1,000	-,128	,390**	,128	,013
		Correlation Coefficient				
		Sig. (2-tailed)		,542	,000	,542
		N	13	13	13	13
	dS			1,000	-,245	,077
	Correlation Coefficient					
	Sig. (2-tailed)			,246	,714	,095
	N		13	13	13	
ω				1,000	,065	-,108
	Correlation Coefficient					
	Sig. (2-tailed)				,760	,620
	N			13	13	
Length					1,000	,417
	Correlation Coefficient					
	Sig. (2-tailed)					,055
	N				13	13
PPI						1,000
	Correlation Coefficient					
	Sig. (2-tailed)					
	N					13

		dN	dS	ω	Length	PPI
Kendall's tau_b	dN	1,000	,070	,817**	-,010	,143
		Correlation Coefficient				
		Sig. (2-tailed)		,442	,000	,907
		N	70	70	50	70
	dS			1,000	-,109	,061
	Correlation Coefficient					
	Sig. (2-tailed)			,298	,466	,603
	N		70	50	70	
ω				1,000	,004	,236*
	Correlation Coefficient					
	Sig. (2-tailed)				,971	,035
	N			50	50	
Length					1,000	,233**
	Correlation Coefficient					
	Sig. (2-tailed)					,008
	N				70	70
PPI						1,000
	Correlation Coefficient					
	Sig. (2-tailed)					
	N					70

**Σ. πίνακας 10: PPI για τις πρωτεΐνες των Συμπλόκων III, IV, V**



Σ. πίνακας 11: Σύγκριση M0-M1, πυρηνικά γονίδια, έρριζο δέντρο, ασαφή δεδ.

GENE	Tree	DF	Ambig. Data	M0	M1	χ <sup>2</sup>	ρ
COX10	Rooted Consensus	19	Yes	-7123,2344	-7104,1691	38,1	0,006
COX15	Rooted Consensus	19	Yes	-6569,9258	-6527,4674	84,9	0,000
COX17	Rooted Consensus	19	Yes	-1268,9099	-1259,7003	18,4	0,497
COX4	Rooted Consensus	19	Yes	-2313,8947	-2300,2305	27,3	0,097
COX5A	Rooted Consensus	19	Yes	-1983,2259	-1971,1968	24,1	0,194
COX5B	Rooted Consensus	19	Yes	-1199,4903	-1184,8134	29,4	0,061
COX6A	Rooted Consensus	19	Yes	-1293,9934	-1278,6868	30,6	0,044
COX6B	Rooted Consensus	19	Yes	-811,6109	-807,4478	8,3	0,983
NDUFA2	Rooted Consensus	19	Yes	-1778,0398	-1762,7762	30,5	0,045
NDUFA4	Rooted Consensus	19	Yes	-1095,6108	-1087,6677	15,9	0,666
NDUFA5	Rooted Consensus	19	Yes	-1761,3168	-1751,2952	20,0	0,391
NDUFA8	Rooted Consensus	19	Yes	-2360,0430	-2349,1117	21,9	0,291
NDUFA9	Rooted Consensus	19	Yes	-5520,3595	-5505,0684	30,6	0,045
NDUFAB	Rooted Consensus	19	Yes	-2501,9931	-2486,9858	30,0	0,052
NDUFB8	Rooted Consensus	19	Yes	-2710,1126	-2695,3909	29,4	0,059
NDUFB9	Rooted Consensus	19	Yes	-1957,9467	-1946,6153	22,7	0,252
NDUFS3	Rooted Consensus	19	Yes	-3478,3251	-3465,1414	26,4	0,120
NDUFS4	Rooted Consensus	19	Yes	-2722,5738	-2714,0823	17,0	0,592
NDUFS6	Rooted Consensus	19	Yes	-1973,6906	-1958,3877	30,6	0,044
NDUFS8	Rooted Consensus	19	Yes	-2662,2123	-2653,3152	17,8	0,538
NDUFV2	Rooted Consensus	19	Yes	-3015,7278	-2995,3785	40,7	0,003
QCR2	Rooted Consensus	19	Yes	-7052,1432	-7038,6383	27,0	0,105
QCR8	Rooted Consensus	19	Yes	-1187,9224	-1179,8677	16,1	0,651
SDHA	Rooted Consensus	19	Yes	-8381,8299	-8363,9678	35,7	0,011
SDHB	Rooted Consensus	19	Yes	-3346,2292	-3327,0685	38,3	0,005
SDHD	Rooted Consensus	19	Yes	-3131,6023	-3117,4053	28,4	0,076

Σ. πίνακας 12: Σύγκριση M0-M1, πυρηνικά γονίδια, έρριζο δέντρο, όχι ασαφή δεδ.

GENE	Tree	DF	Ambig. Data	M0	M1	χ <sup>2</sup>	ρ
COX10	Rooted Consensus	19	No	-5786,2119	-5768,2628	35,9	0,011
COX15	Rooted Consensus	19	No	-6275,5807	-6233,4495	84,3	0,000
COX17	Rooted Consensus	19	No	-990,2271	-982,3899	15,7	0,680
COX4	Rooted Consensus	19	No	-2306,0664	-2292,3520	27,4	0,095
COX5A	Rooted Consensus	19	No	-1939,7547	-1926,8408	25,8	0,135
COX5B	Rooted Consensus	19	No	-1194,5060	-1179,8628	29,3	0,062
COX6A	Rooted Consensus	19	No	-1280,3382	-1264,9079	30,9	0,042
COX6B	Rooted Consensus	19	No	-729,4767	-725,4361	8,1	0,986
NDUFA2	Rooted Consensus	19	No	-1461,0891	-1445,7543	30,7	0,044
NDUFA4	Rooted Consensus	19	No	-1095,6108	-1087,6677	15,9	0,666
NDUFA5	Rooted Consensus	19	No	-1751,8622	-1741,5697	20,6	0,360
NDUFA8	Rooted Consensus	19	No	-2360,0430	-2349,1117	21,9	0,291
NDUFA9	Rooted Consensus	19	No	-4269,3331	-4254,2198	30,2	0,049
NDUFAB	Rooted Consensus	19	No	-1689,1152	-1677,6611	22,9	0,242
NDUFB8	Rooted Consensus	19	No	-2663,2879	-2647,5930	31,4	0,036
NDUFB9	Rooted Consensus	19	No	-1949,5584	-1938,2276	22,7	0,252
NDUFS3	Rooted Consensus	19	No	-3284,1853	-3272,1216	24,1	0,192
NDUFS4	Rooted Consensus	19	No	-2544,0425	-2533,6247	20,8	0,345
NDUFS6	Rooted Consensus	19	No	-1941,8795	-1926,6323	30,5	0,046
NDUFS8	Rooted Consensus	19	No	-2623,6203	-2614,5896	18,1	0,520
NDUFV2	Rooted Consensus	19	No	-2741,1856	-2725,6000	31,2	0,039
QCR2	Rooted Consensus	19	No	-6349,2238	-6334,6095	29,2	0,063
QCR8	Rooted Consensus	19	No	-1177,9921	-1170,0499	15,9	0,666
SDHA	Rooted Consensus	19	No	-7878,4576	-7861,4172	34,1	0,018
SDHB	Rooted Consensus	19	No	-3221,1967	-3203,2448	35,9	0,011
SDHD	Rooted Consensus	19	No	-2936,2057	-2924,7571	22,9	0,243

Σ. πίνακας 13: Σύγκριση M0-M1, πυρηνικά γονίδια, άριζο δέντρο, ασαφή δεδ.

GENE	Tree	DF	Ambig. Data	M0	M1	χ <sup>2</sup>	ρ
COX10	Unrooted Consensus	18	Yes	-7123,2344	-7104,4659	37,5	0,004
COX15	Unrooted Consensus	18	Yes	-6569,9258	-6528,1296	83,6	0,000
COX17	Unrooted Consensus	18	Yes	-1268,9099	-1259,9786	17,9	0,466
COX4	Unrooted Consensus	18	Yes	-2313,8947	-2300,3284	27,1	0,077
COX5A	Unrooted Consensus	18	Yes	-1983,2259	-1971,2404	24,0	0,156
COX5B	Unrooted Consensus	18	Yes	-1199,4903	-1184,8649	29,3	0,045
COX6A	Unrooted Consensus	18	Yes	-1293,9934	-1278,9487	30,1	0,036
COX6B	Unrooted Consensus	18	Yes	-811,6109	-807,4572	8,3	0,974
NDUFA2	Unrooted Consensus	18	Yes	-1778,0398	-1763,0889	29,9	0,038
NDUFA4	Unrooted Consensus	18	Yes	-1095,6108	-1087,8227	15,6	0,623
NDUFA5	Unrooted Consensus	18	Yes	-1761,3168	-1751,8031	19,0	0,390
NDUFA8	Unrooted Consensus	18	Yes	-2360,0430	-2349,9829	20,1	0,326
NDUFA9	Unrooted Consensus	18	Yes	-5520,3595	-5505,3247	30,1	0,037
NDUFAB	Unrooted Consensus	18	Yes	-2501,9931	-2487,0707	29,8	0,039
NDUFB8	Unrooted Consensus	18	Yes	-2710,1126	-2695,4550	29,3	0,044
NDUFB9	Unrooted Consensus	18	Yes	-1957,9467	-1946,8217	22,3	0,222
NDUFS3	Unrooted Consensus	18	Yes	-3478,3251	-3465,5538	25,5	0,111
NDUFS4	Unrooted Consensus	18	Yes	-2722,5738	-2715,7564	13,6	0,753
NDUFS6	Unrooted Consensus	18	Yes	-1973,6906	-1958,6005	30,2	0,036
NDUFS8	Unrooted Consensus	18	Yes	-2662,2123	-2653,1851	18,1	0,451
NDUFV2	Unrooted Consensus	18	Yes	-3015,7278	-2997,1096	37,2	0,005
QCR2	Unrooted Consensus	18	Yes	-7052,1432	-7038,6384	27,0	0,079
QCR8	Unrooted Consensus	18	Yes	-1187,9224	-1180,6700	14,5	0,696
SDHA	Unrooted Consensus	18	Yes	-8381,8299	-8364,0456	35,6	0,008
SDHB	Unrooted Consensus	18	Yes	-3346,2292	-3327,1882	38,1	0,004
SDHD	Unrooted Consensus	18	Yes	-3131,6023	-3117,6815	27,8	0,065

Σ. πίνακας 14: Σύγκριση M0-M1, πυρηνικά γονίδια, άριζο δέντρο, όχι ασαφή δεδ.

GENE	Tree	DF	Ambig. Data	M0	M1	χ <sup>2</sup>	p
COX10	Unrooted Consensus	18	No	-5786,2119	-5768,4358	35,6	0,008
COX15	Unrooted Consensus	18	No	-6275,5807	-6233,5802	84,0	0,000
COX17	Unrooted Consensus	18	No	-990,2271	-982,7312	15,0	0,663
COX4	Unrooted Consensus	18	No	-2306,0664	-2292,4508	27,2	0,075
COX5A	Unrooted Consensus	18	No	-1939,7547	-1927,6979	24,1	0,152
COX5B	Unrooted Consensus	18	No	-1194,5060	-1179,6551	29,7	0,040
COX6A	Unrooted Consensus	18	No	-1280,3382	-1264,9140	30,8	0,030
COX6B	Unrooted Consensus	18	No	-729,6616	-725,4706	8,4	0,972
NDUFA2	Unrooted Consensus	18	No	-1461,0891	-1445,8773	30,4	0,033
NDUFA4	Unrooted Consensus	18	No	-1095,6108	-1087,8225	15,6	0,623
NDUFA5	Unrooted Consensus	18	No	-1751,8622	-1742,0910	19,5	0,359
NDUFA8	Unrooted Consensus	18	No	-2360,0430	-2349,9829	20,1	0,326
NDUFA9	Unrooted Consensus	18	No	-4269,3331	-4254,2672	30,1	0,036
NDUFAB	Unrooted Consensus	18	No	-1689,1152	-1677,8299	22,6	0,208
NDUFB8	Unrooted Consensus	18	No	-2663,2879	-2647,6003	31,4	0,026
NDUFB9	Unrooted Consensus	18	No	-1949,5584	-1938,4306	22,3	0,222
NDUFS3	Unrooted Consensus	18	No	-3284,1853	-3272,5387	23,3	0,180
NDUFS4	Unrooted Consensus	18	No	-2544,0425	-2534,5709	18,9	0,395
NDUFS6	Unrooted Consensus	18	No	-1941,8795	-1926,7814	30,2	0,035
NDUFS8	Unrooted Consensus	18	No	-2623,6203	-2615,2680	16,7	0,545
NDUFV2	Unrooted Consensus	18	No	-2741,1856	-2728,1920	26,0	0,100
QCR2	Unrooted Consensus	18	No	-6349,2238	-6334,6138	29,2	0,046
QCR8	Unrooted Consensus	18	No	-1177,9921	-1170,3184	15,3	0,639
SDHA	Unrooted Consensus	18	No	-7878,4576	-7861,7617	33,4	0,015
SDHB	Unrooted Consensus	18	No	-3221,1967	-3203,3785	35,6	0,008
SDHD	Unrooted Consensus	18	No	-2936,2057	-2925,6614	21,1	0,275

Σ. πίνακας 15: Σύγκριση M0-M1, μιτοχονδρ. γονίδια, έρριζο δέντρο, ασαφή δεδ.

GENE	Tree	DF	Ambig. Data	M0	M1	χ <sup>2</sup>	p
ATP6	Rooted Consensus	19	Yes	-2350,8565	-2328,9476	43,8	0,001
ATP8	Rooted Consensus	19	Yes	-591,9530	-575,0773	33,8	0,020
COX1	Rooted Consensus	19	Yes	-5077,7170	-5051,4704	52,5	0,000
COX2	Rooted Consensus	19	Yes	-2373,4732	-2358,3903	30,2	0,049
COX3	Rooted Consensus	19	Yes	-2636,8840	-2620,1687	33,4	0,021
CYTB	Rooted Consensus	19	Yes	-4182,3212	-4148,8623	66,9	0,000
ND1	Rooted Consensus	19	Yes	-3099,7697	-3073,1955	53,1	0,000
ND2	Rooted Consensus	19	Yes	-4120,3531	-4105,7651	29,2	0,063
ND3	Rooted Consensus	19	Yes	-1341,2860	-1331,9725	18,6	0,482
ND4	Rooted Consensus	19	Yes	-5106,6390	-5093,5358	26,2	0,125
ND4L	Rooted Consensus	19	Yes	-883,7143	-877,9423	11,5	0,904
ND5	Rooted Consensus	19	Yes	-6862,6740	-6842,6478	40,1	0,003
ND6	Rooted Consensus	19	Yes	-2335,7206	-2322,4876	26,5	0,118

Σ. πίνακας 16: Σύγκριση M0-M1, μιτοχονδρ. γονίδια, έρριζο δέντρο, όχι ασαφή δεδ.

GENE	Tree	DF	Ambig. Data	M0	M1	χ <sup>2</sup>	p
ATP6	Rooted Consensus	19	No	-2343,9006	-2321,9866	43,8	0,001
ATP8	Rooted Consensus	19	No	-591,9530	-575,5077	32,9	0,025
COX1	Rooted Consensus	19	No	-5071,9350	-5041,4025	61,1	0,000
COX2	Rooted Consensus	19	No	-2351,5970	-2336,1545	30,9	0,041
COX3	Rooted Consensus	19	No	-2592,8769	-2575,8586	34,0	0,018
CYTB	Rooted Consensus	19	No	-4152,9454	-4119,1651	67,6	0,000
ND1	Rooted Consensus	19	No	-3033,0023	-3006,2789	53,4	0,000
ND2	Rooted Consensus	19	No	-4113,2414	-4098,6583	29,2	0,063
ND3	Rooted Consensus	19	No	-1341,2860	-1331,9725	18,6	0,482
ND4	Rooted Consensus	19	No	-5091,7997	-5078,7640	26,1	0,128
ND4L	Rooted Consensus	19	No	-842,8105	-836,9645	11,7	0,899
ND5	Rooted Consensus	19	No	-6607,2201	-6588,4847	37,5	0,007
ND6	Rooted Consensus	19	No	-1536,6864	-1523,8747	25,6	0,141

Σ. πίνακας 17: Σύγκριση M0-M1, μιτοχονδρ. γονίδια, άριζο δέντρο, ασαφή δεδ.

GENE	Tree	DF	Ambig. Data	M0	M1	χ <sup>2</sup>	p
ATP6	Unrooted Consensus	18	Yes	-2350,8565	-2329,5859	42,5	0,001
ATP8	Unrooted Consensus	18	Yes	-591,9530	-575,5921	32,7	0,018
COX1	Unrooted Consensus	18	Yes	-5077,7170	-5050,5417	54,4	0,000
COX2	Unrooted Consensus	18	Yes	-2373,4732	-2359,4765	28,0	0,062
COX3	Unrooted Consensus	18	Yes	-2636,8840	-2621,0927	31,6	0,025
CYTB	Unrooted Consensus	18	Yes	-4182,3212	-4151,4276	61,8	0,000
ND1	Unrooted Consensus	18	Yes	-3099,7697	-3073,3773	52,8	0,000
ND2	Unrooted Consensus	18	Yes	-4120,3531	-4107,0284	26,6	0,086
ND3	Unrooted Consensus	18	Yes	-1341,2860	-1332,1736	18,2	0,440
ND4	Unrooted Consensus	18	Yes	-5106,6390	-5093,6284	26,0	0,099
ND4L	Unrooted Consensus	18	Yes	-883,7150	-878,4045	10,6	0,910
ND5	Unrooted Consensus	18	Yes	-6862,6740	-6843,7673	37,8	0,004
ND6	Unrooted Consensus	18	Yes	-2335,7206	-2323,8114	23,8	0,161

Σ. πίνακας 18: Σύγκριση M0-M1, μιτοχονδρ. γονίδια, άριζο δέντρο, μη ασαφή δεδ.

GENE	Tree	DF	Ambig. Data	M0	M1	χ <sup>2</sup>	ρ
ATP6	Unrooted Consensus	18	No	-2343,9006	-2322,6250	42,6	0,001
ATP8	Unrooted Consensus	18	No	-591,9530	-575,5921	32,7	0,018
COX1	Unrooted Consensus	18	No	-5071,9350	-5046,0068	51,9	0,000
COX2	Unrooted Consensus	18	No	-2351,5970	-2337,3866	28,4	0,056
COX3	Unrooted Consensus	18	No	-2592,8769	-2576,8308	32,1	0,021
CYTB	Unrooted Consensus	18	No	-4152,9454	-4120,2620	65,4	0,000
ND1	Unrooted Consensus	18	No	-3033,0023	-3006,6021	52,8	0,000
ND2	Unrooted Consensus	18	No	-4113,2414	-4099,9276	26,6	0,086
ND3	Unrooted Consensus	18	No	-1341,2860	-1332,1732	18,2	0,440
ND4	Unrooted Consensus	18	No	-5091,7997	-5078,7894	26,0	0,099
ND4L	Unrooted Consensus	18	No	-842,8105	-837,3419	10,9	0,897
ND5	Unrooted Consensus	18	No	-6607,2201	-6589,8411	34,8	0,010
ND6	Unrooted Consensus	18	No	-1536,6864	-1524,7260	23,9	0,158

Σ. πίνακας 19: Σύγκριση M7-M8, πυρηνικά γονίδια, έρριζο δέντρο, όχι ασαφή δεδ.

GENE	Tree	DF	Ambig. Data	M7	M8	χ <sup>2</sup>	ρ	POSITIVELY SELECTED SITES <95%	>95%	>99%
COX10	Rooted Consensus	2	No	-6912,5959	-6912,2845	0,6	0,732	6S, 325S		
COX15	Rooted Consensus	2	No	-6414,6197	-6414,6209	0,0	1,000	15V, 100N		
COX17	Rooted Consensus	2	No	-1214,0637	-1213,6902	0,7	0,688	18H		
COX4	Rooted Consensus	2	No	-2278,3771	-2278,3779	0,0	1,000			
COX5A	Rooted Consensus	2	No	-1949,8517	-1949,0566	1,6	0,452	148A		
COX5B	Rooted Consensus	2	No	-1190,0295	-1187,2587	5,5	0,063	45L		
COX6A	Rooted Consensus	2	No	-1261,2651	-1261,2661	0,0	1,000	90Q		
COX6B	Rooted Consensus	2	No	-797,8059	-797,8067	0,0	1,000			
NDUFA2	Rooted Consensus	2	No	-1756,0593	-1756,0602	0,0	1,000			
NDUFA4	Rooted Consensus	2	No	-1082,5280	-1082,5281	0,0	1,000			
NDUFA5	Rooted Consensus	2	No	-1723,8849	-1719,9690	7,8	0,020	25S, 66T		
NDUFA8	Rooted Consensus	2	No	-2305,9715	-2304,6346	2,7	0,263	40Q, 101S		
NDUFA9	Rooted Consensus	2	No	-5441,9585	-5441,9611	0,0	1,000			
NDUFAB	Rooted Consensus	2	No	-2437,9743	-2435,4208	5,1	0,078	16R, 20S, 21I, 28S, 42G, 43V		
NDUFB8	Rooted Consensus	2	No	-2645,5154	-2645,5161	0,0	1,000			
NDUFB9	Rooted Consensus	2	No	-1934,6739	-1925,2234	18,9	0,000			51R
NDUFS3	Rooted Consensus	2	No	-3396,9276	-3396,6927	0,5	0,791	54H		
NDUFS4	Rooted Consensus	2	No	-2663,7215	-2663,7224	0,0	1,000			
NDUFS6	Rooted Consensus	2	No	-1916,4463	-1916,4465	0,0	1,000			
NDUFS8	Rooted Consensus	2	No	-2595,0373	-2594,9933	0,1	0,957			
NDUFV2	Rooted Consensus	2	No	-2983,9063	-2981,5853	4,6	0,098			
QCR2	Rooted Consensus	2	No	-6969,6137	-6969,6177	0,0	1,000	82G, 127T, 150S		
QCR8	Rooted Consensus	2	No	-1166,1119	-1166,1122	0,0	1,000			
SDHA	Rooted Consensus	2	No	-8212,3059	-8210,1508	4,3	0,116	472Q, 566V, 581Q, 605T, 614T, 621S		
SDHB	Rooted Consensus	2	No	-3297,3905	-3297,3934	0,0	1,000			
SDHD	Rooted Consensus	2	No	-3071,4510	-3067,2516	8,4	0,015	39V, 48I, 49V, 67G, 71L		

Σ. πίνακας 20: Σύγκριση M7-M8, πυρηνικά γονίδια, έρριζο δέντρο, ασαφή δεδ.

GENE	Tree	DF	Ambig. Data	M7	M8	χ <sup>2</sup>	ρ	POSITIVELY SELECTED SITES <95%	>95%	>99%
COX10	Rooted Consensus	2	Yes	-6912,5959	-6912,2845	0,6	0,732	61S, 69T, 70T, 75S, 77A		
COX15	Rooted Consensus	2	Yes	-6414,6197	-6414,6209	0,0	1,000	29V, 141N		
COX17	Rooted Consensus	2	Yes	-1214,0637	-1213,6902	0,7	0,688	18H		
COX4	Rooted Consensus	2	Yes	-2278,3771	-2278,3779	0,0	1,000			
COX5A	Rooted Consensus	2	Yes	-1949,8517	-1949,0566	1,6	0,452	152A		
COX5B	Rooted Consensus	2	Yes	-1190,0295	-1187,2587	5,5	0,063	46L		
COX6A	Rooted Consensus	2	Yes	-1261,2651	-1261,2339	0,1	0,969	93Q		
COX6B	Rooted Consensus	2	Yes	-797,8059	-797,8065	0,0	1,000			
NDUFA2	Rooted Consensus	2	Yes	-1756,0593	-1756,0601	0,0	1,000			
NDUFA4	Rooted Consensus	2	Yes	-1082,5280	-1082,5281	0,0	1,000			
NDUFA5	Rooted Consensus	2	Yes	-1723,8849	-1719,9690	7,8	0,020	25S, 66T		
NDUFA8	Rooted Consensus	2	Yes	-2305,9715	-2304,6346	2,7	0,263	40Q, 101S		
NDUFA9	Rooted Consensus	2	Yes	-5441,9585	-5441,9611	0,0	1,000			
NDUFAB	Rooted Consensus	2	Yes	-2437,9743	-2435,4208	5,1	0,078	20R, 24S, 26I, 33S, 53K, 55G, 56V, 89*		
NDUFB8	Rooted Consensus	2	Yes	-2645,5154	-2645,5161	0,0	1,000			
NDUFB9	Rooted Consensus	2	Yes	-1934,6739	-1925,2234	18,9	0,000			51R
NDUFS3	Rooted Consensus	2	Yes	-3396,9276	-3396,6927	0,5	0,791	60H, 267*		
NDUFS4	Rooted Consensus	2	Yes	-2663,7215	-2663,7224	0,0	1,000	35T		
NDUFS6	Rooted Consensus	2	Yes	-1916,4463	-1916,4476	0,0	1,000			
NDUFS8	Rooted Consensus	2	Yes	-2595,0373	-2594,9933	0,1	0,957			
NDUFV2	Rooted Consensus	2	Yes	-2983,9063	-2981,5853	4,6	0,098	250*, 253*, 254*, 255*		
QCR2	Rooted Consensus	2	Yes	-6969,6137	-6969,6177	0,0	1,000	85G, 164T, 187S		
QCR8	Rooted Consensus	2	Yes	-1166,1119	-1166,1122	0,0	1,000			
SDHA	Rooted Consensus	2	Yes	-8212,3059	-8210,1508	4,3	0,116	30*, 31*, 509Q, 603V, 618Q, 642T, 651T, 658S		
SDHB	Rooted Consensus	2	Yes	-3297,3905	-3297,3934	0,0	1,000			
SDHD	Rooted Consensus	2	Yes	-3071,4510	-3067,2516	8,4	0,015	42V, 51I, 52V, 72S, 76L, 181E, 183A		

Σ. πίνακας 21: Σύγκριση M7-M8, πυρηνικά γονίδια, άριζο δέντρο, όχι ασαφή δεδ.

GENE	Tree	DF	Ambig. Data	M7	M8	χ <sup>2</sup>	ρ	POSITIVELY SELECTED SITES <95%	>95%	>99%
COX10	Unrooted Consensus	2	No	-5634,7726	-5634,7731	0,0	1,000	6S, 325S		
COX15	Unrooted Consensus	2	No	-6125,9978	-6126,0017	0,0	1,000	15V, 100N		
COX17	Unrooted Consensus	2	No	-947,1880	-946,0964	2,2	0,336	18H		
COX4	Unrooted Consensus	2	No	-2270,7272	-2270,7280	0,0	1,000			
COX5A	Unrooted Consensus	2	No	-1910,8852	-1910,5070	0,8	0,685	148A		
COX5B	Unrooted Consensus	2	No	-1185,0094	-1182,2425	5,5	0,063	45L		
COX6A	Unrooted Consensus	2	No	-1247,8712	-1247,8712	0,0	1,000	90Q		
COX6B	Unrooted Consensus	2	No	-715,4097	-715,6362	-0,5	1,000			
NDUFA2	Unrooted Consensus	2	No	-1443,5161	-1443,5166	0,0	1,000			
NDUFA4	Unrooted Consensus	2	No	-1082,5280	-1082,5281	0,0	1,000			
NDUFA5	Unrooted Consensus	2	No	-1714,5546	-1710,5991	7,9	0,019	25S, 66T		
NDUFA8	Unrooted Consensus	2	No	-2305,9715	-2304,6346	2,7	0,263	40Q, 101S		
NDUFA9	Unrooted Consensus	2	No	-4207,3448	-4207,3471	0,0	1,000			
NDUFAB	Unrooted Consensus	2	No	-1631,2588	-1629,6755	3,2	0,205	16R, 20S, 21I, 28S, 42G, 43V		
NDUFB8	Unrooted Consensus	2	No	-2601,4459	-2601,4466	0,0	1,000			
NDUFB9	Unrooted Consensus	2	No	-1926,2747	-1916,8148	18,9	0,000			51R
NDUFS3	Unrooted Consensus	2	No	-3203,7958	-3203,3747	0,8	0,656	54H		
NDUFS4	Unrooted Consensus	2	No	-2491,5454	-2491,5468	0,0	1,000			
NDUFS6	Unrooted Consensus	2	No	-1885,7235	-1885,7236	0,0	1,000			
NDUFS8	Unrooted Consensus	2	No	-2555,4513	-2555,3448	0,2	0,899			
NDUFV2	Unrooted Consensus	2	No	-2712,0309	-2712,0317	0,0	1,000			
QCR2	Unrooted Consensus	2	No	-6272,1001	-6272,1037	0,0	1,000	82G, 127T, 150S		
QCR8	Unrooted Consensus	2	No	-1156,1963	-1156,1967	0,0	1,000			
SDHA	Unrooted Consensus	2	No	-7726,3821	-7723,6421	5,5	0,065	472Q, 566V, 581Q, 605T, 614T, 621S		
SDHB	Unrooted Consensus	2	No	-3172,4252	-3172,4281	0,0	1,000			
SDHD	Unrooted Consensus	2	No	-2891,1295	-2886,9279	8,4	0,015	39V, 48I, 49V, 67G, 71L		

Σ. πίνακας 22: Σύγκριση M7-M8, πυρηνικά γονίδια, άριζο δέντρο, ασαφή δεδ.

GENE	Tree	DF	Ambig. Data	M7	M8	χ <sup>2</sup>	ρ	POSITIVELY SELECTED SITES <95%	>95%	>99%
COX10	Unrooted Consensus	2	Yes	-6912,5959	-6912,2847	0,6	0,733	61S, 69T, 70T, 75S, 77A		
COX15	Unrooted Consensus	2	Yes	-6414,6197	-6414,6236	0,0	1,000	29V, 141N		
COX17	Unrooted Consensus	2	Yes	-1214,0637	-1213,6902	0,7	0,688	18H		
COX4	Unrooted Consensus	2	Yes	-2278,3771	-2278,3779	0,0	1,000			
COX5A	Unrooted Consensus	2	Yes	-1949,8517	-1949,0566	1,6	0,452	152A		
COX5B	Unrooted Consensus	2	Yes	-1190,0295	-1187,2587	5,5	0,063	46L		
COX6A	Unrooted Consensus	2	Yes	-1261,2651	-1261,2348	0,1	0,970	93Q		
COX6B	Unrooted Consensus	2	Yes	-797,8059	-797,8067	0,0	1,000			
NDUFA2	Unrooted Consensus	2	Yes	-1756,0593	-1756,0597	0,0	1,000			
NDUFA4	Unrooted Consensus	2	Yes	-1082,5280	-1082,5281	0,0	1,000			
NDUFA5	Unrooted Consensus	2	Yes	-1723,8849	-1719,9690	7,8	0,020	25S, 66T		
NDUFA8	Unrooted Consensus	2	Yes	-2305,9715	-2304,6346	2,7	0,263	40Q, 101S		
NDUFA9	Unrooted Consensus	2	Yes	-5441,9585	-5441,9611	0,0	1,000			
NDUFAB	Unrooted Consensus	2	Yes	-2437,9743	-2435,4208	5,1	0,078	20R, 24S, 26I, 33S, 53K, 55G, 56V, 89*		
NDUFB8	Unrooted Consensus	2	Yes	-2645,5154	-2645,5161	0,0	1,000			
NDUFB9	Unrooted Consensus	2	Yes	-1934,6739	-1925,2234	18,9	0,000			51R
NDUFS3	Unrooted Consensus	2	Yes	-3396,9276	-3396,6927	0,5	0,791	60H, 267*		
NDUFS4	Unrooted Consensus	2	Yes	-2663,7215	-2663,7224	0,0	1,000	35T		
NDUFS6	Unrooted Consensus	2	Yes	-1916,4463	-1916,4465	0,0	1,000			
NDUFS8	Unrooted Consensus	2	Yes	-2595,0373	-2594,9933	0,1	0,957			
NDUFV2	Unrooted Consensus	2	Yes	-2983,9063	-2981,5853	4,6	0,098	250*, 253*, 254*, 255*		
QCR2	Unrooted Consensus	2	Yes	-6969,6137	-6969,6175	0,0	1,000	85G, 164T, 187S		
QCR8	Unrooted Consensus	2	Yes	-1166,1119	-1166,1122	0,0	1,000			
SDHA	Unrooted Consensus	2	Yes	-8212,3059	-8210,1508	4,3	0,116	30*, 31*, 509Q, 603V, 618Q, 642T, 651T, 658S		
SDHB	Unrooted Consensus	2	Yes	-3297,3905	-3297,3934	0,0	1,000			
SDHD	Unrooted Consensus	2	Yes	-3071,9626	-3067,2516	9,4	0,009	42V, 51I, 52V, 72S, 76L, 181E, 183A		

Σ. πίνακας 23: Σύγκριση M7-M8, μιτοχονδρ. γονίδια, έρριζο δέντρο, όχι ασαφή δεδ.

GENE	Tree	DF	Ambig. Data	M7	M8	χ <sup>2</sup>	ρ	POSITIVELY SELECTED SITES <95%	>95%	>99%
ATP6	Rooted Consensus	2	No	-2313,8499	-2312,3563	3,0	0,225			
ATP8	Rooted Consensus	2	No	-586,4296	-586,4301	0,0	1,000			
COX1	Rooted Consensus	2	No	-5029,3065	-5035,6431	-12,7	1,000			
COX2	Rooted Consensus	2	No	-2310,4894	-2310,4916	0,0	1,000			
COX3	Rooted Consensus	2	No	-2563,5406	-2563,5427	0,0	1,000			
CYTB	Rooted Consensus	2	No	-4088,8683	-4088,8717	0,0	1,000			
ND1	Rooted Consensus	2	No	-3001,3150	-3001,3164	0,0	1,000			
ND2	Rooted Consensus	2	No	-3999,7401	-3997,8442	3,8	0,150	316Y		
ND3	Rooted Consensus	2	No	-1317,3200	-1317,3211	0,0	1,000			
ND4	Rooted Consensus	2	No	-4982,1183	-4977,1193	10,0	0,007	236S		
ND4L	Rooted Consensus	2	No	-817,2405	-817,2873	-0,1	1,000			
ND5	Rooted Consensus	2	No	-6383,3517	-6382,1813	2,3	0,310			
ND6	Rooted Consensus	2	No	-1493,6719	-1493,5562	0,2	0,891			

Σ. πίνακας 24: Σύγκριση M7-M8, μιτοχονδρ. γονίδια, έρριζο δέντρο, ασαφή δεδ.

GENE	Tree	DF	Ambig. Data	M7	M8	χ <sup>2</sup>	ρ	POSITIVELY SELECTED SITES <95%	>95%	>99%
ATP6	Rooted Consensus	2	Yes	-2320,8008	-2319,3057	3,0	0,224			
ATP8	Rooted Consensus	2	Yes	-586,4296	-586,4301	0,0	1,000			
COX1	Rooted Consensus	2	Yes	-5035,0919	-5041,4287	-12,7	1,000			
COX2	Rooted Consensus	2	Yes	-2330,1617	-2330,1638	0,0	1,000			
COX3	Rooted Consensus	2	Yes	-2601,9649	-2601,9671	0,0	1,000			
CYTB	Rooted Consensus	2	Yes	-4117,4396	-4117,4431	0,0	1,000			
ND1	Rooted Consensus	2	Yes	-3066,2656	-3066,2687	0,0	1,000			
ND2	Rooted Consensus	2	Yes	-4006,5317	-4004,6134	3,8	0,147	318Y		
ND3	Rooted Consensus	2	Yes	-1317,3200	-1317,3211	0,0	1,000			
ND4	Rooted Consensus	2	Yes	-4996,4937	-4991,4732	10,0	0,007	237S		
ND4L	Rooted Consensus	2	Yes	-860,5026	-860,5035	0,0	1,000			
ND5	Rooted Consensus	2	Yes	-6623,9772	-6623,9809	0,0	1,000			
ND6	Rooted Consensus	2	Yes	-2253,7187	-2253,7187	0,0	1,000			

## Σ. πίνακας 25: Σύγκριση M7-M8, μιτοχονδρ. γονίδια, άριζο δέντρο, όχι ασαφή δεδ.

GENE	Tree	DF	Ambig. Data	M7	M8	$\chi^2$	$\rho$	POSITIVELY SELECTED SITES <95%	>95%	>99%
ATP6	Unrooted Consensus	2	No	-2313,8499	-2312,3563	3,0	0,225			
ATP8	Unrooted Consensus	2	No	-586,4296	-586,4301	0,0	1,000			
COX1	Unrooted Consensus	2	No	-5029,3065	-5029,2534	0,1	0,948			
COX2	Unrooted Consensus	2	No	-2310,4894	-2310,4915	0,0	1,000			
COX3	Unrooted Consensus	2	No	-2563,5406	-2563,5427	0,0	1,000			
CYTB	Unrooted Consensus	2	No	-4088,8683	-4088,8717	0,0	1,000			
ND1	Unrooted Consensus	2	No	-3001,3150	-3001,3181	0,0	1,000			
ND2	Unrooted Consensus	2	No	-3999,7401	-3997,8442	3,8	0,150		316Y	
ND3	Unrooted Consensus	2	No	-1317,3200	-1317,3211	0,0	1,000			
ND4	Unrooted Consensus	2	No	-4982,1183	-4977,1193	10,0	0,007		236S	
ND4L	Unrooted Consensus	2	No	-817,2405	-817,2873	-0,1	1,000			
ND5	Unrooted Consensus	2	No	-6383,3517	-6382,1813	2,3	0,310			
ND6	Unrooted Consensus	2	No	-1493,6719	-1493,5562	0,2	0,891			

## Σ. πίνακας 26: Σύγκριση M7-M8, μιτοχονδρ. γονίδια, άριζο δέντρο, ασαφή δεδ.

GENE	Tree	DF	Ambig. Data	M7	M8	$\chi^2$	$\rho$	POSITIVELY SELECTED SITES <95%	>95%	>99%
ATP6	Unrooted Consensus	2	Yes	-2320,8008	-2319,3057	3,0	0,224			
ATP8	Unrooted Consensus	2	Yes	-586,4296	-586,4301	0,0	1,000			
COX1	Unrooted Consensus	2	Yes	-5035,0919	-5035,0392	0,1	0,949			
COX2	Unrooted Consensus	2	Yes	-2330,1617	-2330,1640	0,0	1,000			
COX3	Unrooted Consensus	2	Yes	-2601,9649	-2601,9671	0,0	1,000			
CYTB	Unrooted Consensus	2	Yes	-4117,4396	-4117,4431	0,0	1,000			
ND1	Unrooted Consensus	2	Yes	-3066,2656	-3066,2669	0,0	1,000			
ND2	Unrooted Consensus	2	Yes	-4006,5318	-4004,6134	3,8	0,147		318Y	
ND3	Unrooted Consensus	2	Yes	-1317,3200	-1317,3211	0,0	1,000			
ND4	Unrooted Consensus	2	Yes	-4996,4937	-4991,4732	10,0	0,007		237S	
ND4L	Unrooted Consensus	2	Yes	-860,5026	-860,5035	0,0	1,000			
ND5	Unrooted Consensus	2	Yes	-6623,9772	-6623,0516	1,9	0,396			
ND6	Unrooted Consensus	2	Yes	-2253,7187	-2253,7187	0,0	1,000			

## Κώδικας 1: Πρόγραμμα σε Java το οποίο αυτοματοποιεί το «τρέξιμο» του PAML για όλα τα γονίδια, με παραμέτρους που ορίζονται από τον χρήστη

```

//THIS PROGRAM PARSSES A DIRECTORY, CREATES PAML CONTROL FILES FOR EACH SUBDIRECTORY,
//RUNS CODEML FOR EACH DIR, AND SAVES THE OUTPUT FILES.
//USER MUST MANUALLY SET THE PUBLIC VARIABLES, DEPENDING ON MODEL-PARAMETERS USED.

import java.io.*;
import java.lang.Runtime;
importumontreal.iro.lecuyer.probdist.*;

public class BatchPAMLa
{
    //This is the path to the parent dir.
    public static String path = "C:\\Documents and Settings\\PervyPirate\\Desktop\\Molecular Biology @
Uoc\\Master\\New\\ALL GENOMES\\BS\\MIT";
    //This is a temporary file.
    public static String tempfile = (path+"\\tempfile.tmp");
    //This is the outfile with all the goodies.
    public static String finalfile = (path+"\\BS_results_M3.out");
    //Degrees of freedom.
    //public static int DF = 19; or DF-1 if DNAML (unrooted)
    //remember to change the filename in main, add or remove +DNAML+
    //remember to change DF, 2 for M7-M8
    //remember to change the genetic code.
    public static int DF = 1;
    public static String Codeml1 = "MODEL_NULL_Amb_UCons.out";
    public static String Codeml2 = "MODEL_A_NoAmb_UCons.out";
    //Control file values
    //0: ambiguous data, 1: no ambiguous data.
    public static int Cleandata = 1;
    //models. for M0 is 0, for M1 is 1 ...Others are usually 0.
    public static int Model1 = 2;
    public static int Model2 = 2;
    //0: one class, 7:beta 8:beta & w
    //for models 7,8, two types of sites
    public static int NSSites1 = 2;
    public static int NSSites2 = 2;
    // 0: universal, 4: invertebrate mit.
    public static int ICode = 4;
    public static int Noisy = 0;
    public static int Verbose = 0;
    public static int Runmode = 0;
    public static int SeqType = 1;
    public static int CodonFreq = 2;
    public static int Ndata = 10;
    public static int Clock = 0;
    //public static int Model = 0;
    //public static int NSSites = 0;
    public static int MGene = 0;
    public static int FixKappa = 0;
    public static double Kappa = 2;
    public static int FixOmega1 = 1;
    public static int FixOmega2 = 0;
    public static double Omega1 = 1;
    public static double Omega2 = 1;

    public static void main(String[] args) throws Exception
    {
        File dir = new File(path);
        String[] children = dir.list();
        String temp = "";
        File Output = new File(finalfile);
        PrintWriter PW = new PrintWriter( new FileWriter(Output));
    }
}

```

```

PW.write("GENE\tM0\tM1\tx^2\t\r\n");
PW.close();
    if (children == null)
    {
        //Either dir does not exist or is not a directory.
    }
    else
    {
        for (int i=0; i<children.length; i++)
        {
            //Get filename of file or directory.
            String filename = children[i];
            temp = path+"\\"+filename;
            //initialising output file.
            ChiSquareDist Arr = new ChiSquareDist(DF);
            double a = 0;
            double b = 0;
            double x = 0;
            File IsDir = new File(temp);
            if (IsDir.isDirectory())
            {
                PW = new PrintWriter( new FileWriter(Output, true));
                //correcting alignment file attribute.
                CorrectPhyFile(temp+"\\"+filename+"_ali.ph3");
                //Careful to specify the correct .out file each time.
                //Specify Model,NS
                //This is for M0. (or M7)
                CreateControlFile(temp+"\\"+filename+".ctl", filename+"_ali.ph3",
filename+"UTreeBS.tree", Codeml1, Model1, NSSites1, FixOmega1, Omega1);
                CreateBatchFile(filename);
                RunPAML(filename);
                //This is for M1. (or M8)
                CreateControlFile(temp+"\\"+filename+".ctl", filename+"_ali.ph3",
filename+"UTreeBS.tree", Codeml2, Model2, NSSites2, FixOmega2, Omega2);
                RunPAML(filename);
                a = ExtractLnL(temp+"\\"+Codeml1);
                b = ExtractLnL(temp+"\\"+Codeml2);
                x = 2*(b-a);

                System.out.println(filename+"\tM0:"+a+"\tM1:"+b+"\tx^2:"+x+"\t\r\n"+Arr.barF(DF,2,x)+"\n");
                PW.write(filename+"\t"+a+"\t"+b+"\t"+x+"\t"+Arr.barF(DF,2,x)+"\n");
            }
            PW.close();
        }
    }
}

//This function parses an output file and returns the lnL.
public static double ExtractLnL(String filename)
{
    String line = null;
    try
    {
        BufferedReader BR = new BufferedReader( new FileReader(filename));
        while ((line = BR.readLine()) != null)
        {
            if (line.startsWith("lnL"))
            {
                line=line.substring(line.indexOf("-"),line.lastIndexOf(" "));
                line=line.trim();
                break;
            }
            //System.out.println(line);
        }
    }
    catch (Exception x)
    {
        System.out.println("Some bad IO stuff happened in function 'ExtractLnL.'+x);
    }
    //double res = Double.parseDouble(line);
    return Double.parseDouble(line);
}

//This function runs codeml (and more)
public static void RunPAML(String filename)
{
    String line;
    String pathtobat = "\"" + path + "\" + filename + "\" + filename + ".bat\"";
    System.out.println(pathtobat);
    try
    {
        //Process p = Runtime.getRuntime().exec("c:/windows/system32/notepad.exe");
        //String[] cmd = { "myProgram.exe", "-o=This is an option" };
        //String[] cmd = { "codeml", "C:\\Documents and
Settings\\PervyPirate\\Desktop\\Molecular Biology @ UoC\\Master\\New\\Drosophilidae\\NDUFA2\\codeml.ctl"};

        //Process p = Runtime.getRuntime().exec("C:/Documents and
Settings/PervyPirate/Desktop/Molecular Biology @ UoC/Master/New/Drosophilidae/NDUFA2/runme.bat");
        Process p = Runtime.getRuntime().exec(pathtobat);
        BufferedReader input = new BufferedReader(new InputStreamReader(p.getInputStream()));
        while ((line = input.readLine()) != null) {System.out.println(line);}
        System.out.println("Exit Value = " + p.waitFor());
        input.close();
        //p.waitFor();
        //p.destroy();
        return;
    }
    catch (Exception x)
    {
        System.out.println("Exception in Runtime processing."+x);
    }
}

```

```

    }

    //Bioedit saves Phylip v3 files in Sequential format, but specifies them as Interleaved. This function
    corrects the file attributes.
    public static void CorrectPhyFile(String Filename)
    {
        String line = null;
        int linenumber = 0;
        File InFile = new File(Filename);
        try
        {
            BufferedReader BR = new BufferedReader( new FileReader(Filename));
            System.out.println(Filename);
            File TempFile = new File(tempfile);
            PrintWriter PW = new PrintWriter(new FileWriter(TempFile));

            while ((line = BR.readLine()) != null)
            {
                if (linenumber==0)
                {
                    //System.out.println(line);
                    line = line.replaceFirst("I","S");
                    //System.out.println(line);
                    PW.println(line);
                }
                else
                {
                    PW.println(line);
                    linenumber++;
                }
            }
            PW.close();
            BR.close();

            //Delete the original file
            if (!InFile.delete()) {
                System.out.println("Could not delete file");
                return;
            }

            //Rename the new file to the filename the original file had.
            if (!TempFile.renameTo(InFile))
                System.out.println("Could not rename file");
        }
        catch (FileNotFoundException NoFilename)
        {
            System.out.println("No .ph3 file was found in dir "+ Filename+".");
            return;
        }
        catch (IOException Exc)
        {
            System.out.println("Some bad IO stuff happened in function 'CorrectPhyFile'.");
            return;
        }
    }

    //This function creates a control file to be loaded in every PAML iteration.
    public static void CreateControlFile(String Filename, String InputFile, String TreeFile, String OutFile,
    int Model, int NSSites, int FixOmega, double Omega)
    {
        try
        {
            File Control = new File(Filename);
            PrintWriter PW = new PrintWriter(new FileWriter(Control));

            PW.write("seqfile = "+InputFile+" * sequence data filename.\n");
            PW.write("treefile = "+TreeFile+" * tree structure filename.\n");
            PW.write("outfile = "+OutFile+" * main result filename.\n\n");
            PW.write("noisy = "+Noisy+" * 0,1,2,3,9:how much rubbish on the screen\n");
            PW.write("verbose = "+Verbose+" * 0:concise; 1:detailed, 2:too much\n");
            PW.write("runmode = "+Runmode+" * 0:user tree; 1:semi-automatic; 2:automatic;
3:StepwiseAddition; (4,5):PerturbationNNI; -2:pairwise\n\n");
            PW.write("segtype = "+SegType+" * 1:codons; 2:AAs; 3:codons-->AAs\n");
            PW.write("CodonFreq = "+CodonFreq+" * 0:1/61 each, 1:F1X4, 2:F3X4, 3:codon
table\n\n");
            PW.write("ndata = "+Ndata+"\n");
            PW.write("clock = "+Clock+" * 0:no clock, 1:clock; 2:local clock;
3:CombinedAnalysis\n");
            PW.write("model = "+Model+" * models for codons: 0:one, 1:one/branch, 2:2 or more
dN/dS ratios for branches\n");
            PW.write("NSSites = "+NSSites+" * 0:one w; 1:neutral; 2:selection; 3:discrete;
4:freqs; 5:gamma; 6:2gamma; 7:beta; 8:beta&w; 9:beta&gamma; 10:beta&gamma+1; 11:beta&normal>1; 12:0&2normal>1;
13:3normal>0\n");
            PW.write("icode = "+ICode+" * 0:universal code; 1:mammalian mt; 4:invertebrate
mt...\n");
            PW.write("Mgene = "+MGene+" * codon: 0:rates, 1:separate; 2:diff pi, 3:diff kapa,
4:all diff\n");
            PW.write("fix kappa = "+FixKappa+" * 1:kappa fixed, 0:kappa to be estimated\n");
            PW.write("kappa = "+Kappa+" * initial or fixed kappa\n");
            PW.write("fix_omega = "+FixOmega+" * 1:omega or omega_l fixed, 0:estimate \n");
            PW.write("omega = "+Omega+" * initial or fixed omega, for codons or codon-based
AAs\n\n");
            PW.write("fix_alpha = 1 * 0: estimate gamma shape parameter; 1: fix it at alpha\n");
            PW.write("alpha = 0. * initial or fixed alpha, 0:infinity (constant rate)\n");
            PW.write("alpha = 0 * different alphas for genes\n");
            PW.write("ncatG = 8 * # of categories in dG of NSites models\n");
            PW.write("getSE = 0 * 0: don't want them, 1: want S.E.s of estimates\n");
            PW.write("RateAncestor = 1 * (0,1,2): rates (alpha>0) or ancestral states (1 or
2)\n\n");
            PW.write("Small Diff = .5e-6\n");
            PW.write("cleandata = "+Cleandata+" * remove sites with ambiguity data (1:yes,
0:no)\n");
            PW.write("**fix_blength = -1 * 0:ignore, -1:random, 1:initial, 2:fixed\n");
        }
    }

```



```

        PW.write("method = 0 * Optimization method 0: simultaneous; 1: one branch a
time\n");
        PW.close();
    }
    catch (IOException Exc)
    {
        System.out.println("Some bad IO stuff happened in function 'CreateControlFile'.");
        return;
    }
}

//This function creates a batch file to run codeml.
public static void CreateBatchFile(String filename)
{
    File batch = new File(path+"\\ "+filename+"\\ "+filename+".bat");
    try
    {
        PrintWriter PW = new PrintWriter(new FileWriter(batch));
        PW.write("cd /d "+path+"\\ "+filename+"\n");
        PW.write("codeml "+filename+".ctl\n");
        PW.write("exit\n");
        PW.close();
    }
    catch (IOException Exc)
    {
        System.out.println("Some bad IO stuff happened in function 'CreateBatchFile'.");
        return;
    }
}
}
}

```

## Κώδικας 2: Απλό script σε Java για τη δημιουργία φυλογενετικών δέντρων για κάθε γονίδιο, για χρήση με το PAML

```

//THIS PROGRAM PARSES A DIRECTORY, AND CREATES A TREE FILE IN EACH
//SUB-DIR, USING THE NAME OF THE DIR FOR THE TREE BRANCHES.
import java.io.*;
public class GeneNamesToTree
{
    public static String path = "C:\\Documents and Settings\\PervyPirate\\Desktop\\Molecular Biology @
UoC\\Master\\New\\ALL GENOMES\\BS\\MIT";

    public static void main(String[] args) throws Exception
    {
        File dir = new File(path);
        String[] children = dir.list();
        String temp = "";
        if (children == null) {
            // Either dir does not exist or is not a directory
        } else {
            for (int i=0; i<children.length; i++) {
                // Get filename of file or directory
                String filename = children[i];
                temp = path+"\\ "+filename;
                File IsDir = new File(temp);
                if (IsDir.isDirectory())
                {
                    //make tree in dir.UTree is unrooted,Tree is rooted
                    temp = temp + "\\ "+filename+"UTreeBS.tree";
                    File NewTree = new File(temp);
                    BufferedWriter out1 = new BufferedWriter(new FileWriter(NewTree));
                    //ROOTED

                    //out1.write("((((("+filename+"_DME, (" +filename+"_DSI, "+filename+"_DSE)), (" +filename+"_DYA, "+filename+"_
DER))), "+filename+"_DAN), (" +filename+"_DPO, "+filename+"_DPE)), (" +filename+"_DMO, "+filename+"_DGR)), "+filename+"_AGA
");
                    //UNROOTED

                    out1.write("((((("+filename+"_DME, (" +filename+"_DSI, "+filename+"_DSE)), (" +filename+"_DYA, "+filename+"_DER
)), "+filename+"_DAN), (" +filename+"_DPO, "+filename+"_DPE)#1), (" +filename+"_DMO, "+filename+"_DGR), "+filename+"_AGA);"
);

                    System.out.println(temp);
                    out1.close();
                    /*
                    temp = path+"\\ "+filename;
                    temp = temp + "\\ "+filename+"Tree_labeled.tree";
                    NewTree = new File(temp);
                    BufferedWriter out2 = new BufferedWriter(new FileWriter(NewTree));
                    out2.write("((((("+filename+"_DME, (" +filename+"_DSI, "+filename+"_DSE
#1)#5, (" +filename+"_DYA, "+filename+"_DER)#2)#6, "+filename+"_DAN)#7, (" +filename+"_DPO, "+filename+"_DPE)#3)#8, (" +file
name+"_DMO, "+filename+"_DGR)#4)#9;");
                    System.out.println(temp);
                    out2.close();
                    temp = path+"\\ "+filename;
                    temp = temp + "\\ "+filename+"Tree_unrooted.tree";
                    NewTree = new File(temp);
                    BufferedWriter out3 = new BufferedWriter(new FileWriter(NewTree));

                    out3.write("(((("+filename+"_DME, (" +filename+"_DSI, "+filename+"_DSE)), (" +filename+"_DYA, "+filename+"_DER
)), "+filename+"_DAN), (" +filename+"_DPO, "+filename+"_DPE)), (" +filename+"_DMO, "+filename+"_DGR));");
                    System.out.println(temp);
                    out3.close();
                    */
                }
            }
        }
    }
}
}

```

```
}  
}
```

### Κώδικας 3: Απλό script σε Java για μετατροπή του format γονιδίων της Organelle DB, σε fasta.

```
//THIS PROGRAM PARSES FILES WITH THE FORMAT USED BY THE ORGANELLE DB IN JAPAN, AND EXPORTS THEM TO FASTA.  
  
import java.io.*;  
  
public class OrgToFasta  
{  
    //Input file.  
    public static String InputFile = "C:\\Documents and Settings\\PervyPirate\\Desktop\\Molecular Biology @  
UoC\\Master\\temp.fas";  
  
    public static void main(String[] args) throws Exception  
    {  
        String Line = null;  
        BufferedReader Input = new BufferedReader( new FileReader(InputFile) );  
  
        while ( ((Line = Input.readLine()) != null) && Line.length()>5)  
        {  
            Line = Line.substring(5);  
            Line = Line.replace(" ", "");  
            System.out.println(Line);  
        }  
    }  
}
```

### Κώδικας 4: Απλό script σε Java για τον υπολογισμό των p-values για chi-square tests. Χρησιμοποιεί το εξωτερικό πακέτο umontreal.iro.lecuyer (βιβλιοθήκη με εργαλεία στοχαστικής ανάλυσης)

```
//TEST PROGRAM FOR CHI SQUARE TEST  
  
import java.lang.Object;  
import umontreal.iro.lecuyer.probdist.*;  
  
public class ChiSquare  
{  
    //second parameter is decimal spaces returned  
    public static double ComputeChiSquare(int df, int decimals , double x)  
    {  
        ChiSquareDist Arr = new ChiSquareDist(18);  
        double p = Arr.barF(df,decimals,x);  
        return p;  
    }  
    public static void main(String[] args)  
    {  
        double a = -4269.333144;  
        double b = -4256.827336;  
        double x = 2*(b-a);  
        //System.out.println(x);  
        //degrees of freedom.  
        int df = 18;  
        //second parameter is decimal spaces returned  
        System.out.println(ComputeChiSquare(df,5,x));  
    }  
}
```

## Βιβλιογραφία

---

- <sup>1</sup> McBride, H.M., Neuspiel, M. & Wasiak, S., 2006. Mitochondria: More Than Just a Powerhouse. *Current Biology*, 16(14), R551-R560.
- <sup>2</sup> Alberts, Bruce; Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, Peter Walter (1994). *Molecular Biology of the Cell*. New York: Garland Publishing.
- <sup>3</sup> Schirmer, T., 1998. General and Specific Porins from Bacterial Outer Membranes. *Journal of Structural Biology*, 121(2), 101-109.
- <sup>4</sup> Herrmann, J.M. & Neupert, W., 2000. Protein transport into mitochondria. *Current Opinion in Microbiology*, 3(2), 210-214.
- <sup>5</sup> McMillin, J.B. & Dowhan, W., 2002. Cardiolipin and apoptosis. *Biochimica Et Biophysica Acta*, 1585(2-3), 97-107.
- <sup>6</sup> Sagan, L., 1967. On the origin of mitosing cells. *Journal of Theoretical Biology*, 14(3), 255-274.
- <sup>7</sup> Oca-Cossio, J. et al., 2003. Limitations of Allotopic Expression of Mitochondrial Genes in Mammalian Cells. *Genetics*, 165(2), 707-720.
- <sup>8</sup> Green, D.R., 1998. Apoptotic pathways: the roads to ruin. *Cell*, 94(6), 695-698.
- <sup>9</sup> Hajnóczky, G. et al., 2006. Mitochondrial calcium signalling and cell death: approaches for assessing the role of mitochondrial Ca<sup>2+</sup> uptake in apoptosis. *Cell Calcium*, 40(5-6), 553-560.
- <sup>10</sup> Rossier, M.F., 2006. T channels and steroid biosynthesis: in search of a link with mitochondria. *Cell Calcium*, 40(2), 155-164.
- <sup>11</sup> Logan, D.C., Plant mitochondria. *Annual Plant Reviews*, 31.
- <sup>12</sup> Jukes, T.H. & Osawa, S., 1990. The genetic code in mitochondria and chloroplasts. *Cellular and Molecular Life Sciences*, 46(11), 1117-1126.
- <sup>13</sup> Bohr, V. & Anson, R., 1999. Mitochondrial DNA Repair Pathways. *Journal of Bioenergetics and Biomembranes*, 31(4), 391-398.
- <sup>14</sup> Haag-Liautard, C. et al., 2008. Direct Estimation of the Mitochondrial DNA Mutation Rate in *Drosophila melanogaster*. *PLoS Biol*, 6(8), e204.
- <sup>15</sup> Parsons, T.J. et al., 1997. A high observed substitution rate in the human mitochondrial DNA control region. *Nature Genetics*, 15(4), 363-368.
- <sup>16</sup> Nachman, M.W. & Crowell, S.L., 2000. Estimate of the Mutation Rate per Nucleotide in Humans. *Genetics*, 156(1), 297-304.
- <sup>17</sup> Brown, W.M., George, M. & Wilson, A.C., 1979. Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 76(4), 1967-1971.
- <sup>18</sup> Palmer, J.D. & Herbon, L.A., 1988. Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *Journal of Molecular Evolution*, 28(1-2), 87-97.
- <sup>19</sup> Fitch, W.M., 1967. Evidence suggesting a non-random character to nucleotide replacements in naturally occurring mutations. *Journal of Molecular Biology*, 26(3), 499-507.

- 
- <sup>20</sup> Ridley, M., 2004. *Evolution*, Wiley-Blackwell.
- <sup>21</sup> Jukes, T. & Cantor, C., 1969. *Evolution of Protein Molecules*, Academy Press.
- <sup>22</sup> Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, 16(2), 111-120.
- <sup>23</sup> Hasegawa, M., Kishino, H. & Yano, T., 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, 22(2), 160-174.
- <sup>24</sup> Carlini, D.B. & Stephan, W., 2003. In vivo introduction of unpreferred synonymous codons into the *Drosophila Adh* gene results in reduced levels of ADH protein. *Genetics*, 163(1), 239-243.
- <sup>25</sup> Pagani, F., Raponi, M. & Baralle, F.E., 2005. Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 102(18), 6368-6372.
- <sup>26</sup> Xia, X., 2005. Mutation and selection on the anticodon of tRNA genes in vertebrate mitochondrial genomes. *Gene*, 345(1), 13-20.
- <sup>27</sup> Ikemura, T., 1985. Codon usage and tRNA content in unicellular and multicellular organisms. *Molecular Biology and Evolution*, 2(1), 13-34.
- <sup>28</sup> Moriyama, E.N. & Powell, J.R., 1998. Gene length and codon usage bias in *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Escherichia coli*. *Nucleic Acids Research*, 26(13), 3188-3193.
- <sup>29</sup> Pagani, F., Raponi, M. & Baralle, F.E., 2005. Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 102(18), 6368 -6372.
- <sup>30</sup> Miyata, T., Miyazawa, S. & Yasunaga, T., 1979. Two types of amino acid substitutions in protein evolution. *Journal of Molecular Evolution*, 12(3), 219-236.
- <sup>31</sup> Hurst, L.D., 2002. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends in Genetics: TIG*, 18(9), 486.
- <sup>32</sup> Barrientos, A., Kenyon, L. & Moraes, C.T., 1998. Human Xenomitochondrial Cybrids. *Journal of Biological Chemistry*, 273(23), 14210-14217.
- <sup>33</sup> Kenyon, L. & Moraes, C.T., 1997. Expanding the functional human mitochondrial DNA database by the establishment of primate xenomitochondrial cybrids. *Proceedings of the National Academy of Sciences of the United States of America*, 94(17), 9131-9135.
- <sup>34</sup> McKenzie, M. et al., 2003. Functional Respiratory Chain Analyses in Murid Xenomitochondrial Cybrids Expose Coevolutionary Constraints of Cytochrome b and Nuclear Subunits of Complex III. *Mol Biol Evol*, 20(7), 1117-1124.
- <sup>35</sup> Hobolth, A. et al., 2007. Genomic Relationships and Speciation Times of Human, Chimpanzee, and Gorilla Inferred from a Coalescent Hidden Markov Model. *PLoS Genetics*, 3(2), e7.
- <sup>36</sup> Patterson, N. et al., 2006. Genetic evidence for complex speciation of humans and chimpanzees. *Nature*, 441(7097), 1103-1108.
- <sup>37</sup> Capy, P. & Gibert, P., 2004. *Drosophila melanogaster*, *Drosophila simulans*: so similar yet so different. *Genetica*, 120(1-3), 5-16.

- 
- <sup>38</sup> 2005. Sequence, annotation, and analysis of synteny between rice chromosome 3 and diverged grass species. *Genome Research*, 15(9), 1284 -1291.
- <sup>39</sup> Chenna, R. et al., 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research*, 31(13), 3497-3500.
- <sup>40</sup> Tamura, K. et al., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, 24(8), 1596-1599.
- <sup>41</sup> Yang, Z., 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol Biol Evol*, 24(8), 1586-1591.
- <sup>42</sup> Tarraga, J. et al., 2007. Phylemon: a suite of web tools for molecular evolution, phylogenetics and phylogenomics. *Nucl. Acids Res.*, gkm224.
- <sup>43</sup> Nei, M. & Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, 3(5), 418-426.
- <sup>44</sup> Yang, Z. & Nielsen, R., 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Molecular Biology and Evolution*, 17(1), 32-43.
- <sup>45</sup> Yang & Bielawski, 2000. Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution (Personal Edition)*, 15(12), 496-503.
- <sup>46</sup> Jensen, L.J. et al., 2009. STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucl. Acids Res.*, 37(suppl\_1), D412-416.
- <sup>47</sup> Zhang, L. & Li, W., 2004. Mammalian Housekeeping Genes Evolve More Slowly than Tissue-Specific Genes. *Molecular Biology and Evolution*, 21(2), 236 -239.
- <sup>48</sup> Gu, X. & Li, W., 1998. Estimation of evolutionary distances under stationary and nonstationary models of nucleotide substitution. *Proceedings of the National Academy of Sciences of the United States of America*, 95(11), 5899 -5905.
- <sup>49</sup> Rand, D.M., 1994. Thermal habit, metabolic rate and the evolution of mitochondrial DNA. *Trends in Ecology & Evolution*, 9(4), 125-131.
- <sup>50</sup> Mishmar, D. et al., 2003. Natural selection shaped regional mtDNA variation in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 100(1), 171 -176.
- <sup>51</sup> Mishmar, D. et al., 2006. Adaptive selection of mitochondrial complex I subunits during primate radiation. *Gene*, 378, 11-18.
- <sup>52</sup> Wu, W. et al., 1997. Molecular Evolution of Cytochrome c Oxidase Subunit IV: Evidence for Positive Selection in Simian Primates. *Journal of Molecular Evolution*, 44(5), 477-491.
- <sup>53</sup> Schmidt, T.R., Goodman, M. & Grossman, L.I., 2002. Amino acid replacement is rapid in primates for the mature polypeptides of COX subunits, but not for their targeting presequences. *Gene*, 286(1), 13-19.
- <sup>54</sup> Osada, N. et al., 2002. Search for Genes Positively Selected during Primate Evolution by 5'-End-Sequence Screening of Cynomolgus Monkey cDNAs. *Genomics*, 79(5), 657-662.
- <sup>55</sup> Tan, T. et al., 2005. Length, protein-protein interactions, and complexity. *Physica A: Statistical Mechanics and its Applications*, 350(1), 52-62.

- 
- <sup>56</sup> Hughes, A.L., 1997. Rapid evolution of immunoglobulin superfamily C2 domains expressed in immune system cells. *Molecular Biology and Evolution*, 14(1), 1-5.
- <sup>57</sup> Gilad, Y. et al., 2003. Natural Selection on the Olfactory Receptor Gene Family in Humans and Chimpanzees. *American Journal of Human Genetics*, 73(3), 489-501.
- <sup>58</sup> Williams, P.D. et al., 2006. Assessing the Accuracy of Ancestral Protein Reconstruction Methods. *PLoS Comput Biol*, 2(6), e69.
- <sup>59</sup> Adams, M.D. et al., 2000. The Genome Sequence of *Drosophila melanogaster*. *Science*, 287(5461), 2185-2195.
- <sup>60</sup> Grossman, L.I. et al., 2004. Accelerated evolution of the electron transport chain in anthropoid primates. *Trends in Genetics: TIG*, 20(11), 578-585.
- <sup>61</sup> Rand, D.M., Dorfsman, M. & Kann, L.M., 1994. Neutral and non-neutral evolution of *Drosophila* mitochondrial DNA. *Genetics*, 138(3), 741-756.
- <sup>62</sup> Jobson, R.W. et al., 2004. Adaptive evolution of cytochrome c oxidase: Infrastructure for a carnivorous plant radiation. *Proceedings of the National Academy of Sciences of the United States of America*, 101(52), 18064-18068.
- <sup>63</sup> Dalziel, A.C. et al., 2006. Molecular evolution of cytochrome c oxidase in high-performance fish (teleostei: Scombroidei). *Journal of Molecular Evolution*, 62(3), 319-331.
- <sup>64</sup> Tweedie, S. et al., 2009. FlyBase: enhancing *Drosophila* Gene Ontology annotations. *Nucleic Acids Research*, 37(Database), D555-D559.
- <sup>65</sup> Yang, Z., Wong, W.S. & Nielsen, R., 2005. Bayes Empirical Bayes Inference of Amino Acid Sites under Positive Selection. *Mol Biol Evol*, msi097.
- <sup>66</sup> Cao, Y. et al., 1998. Conflict among individual mitochondrial proteins in resolving the phylogeny of eutherian orders. *Journal of Molecular Evolution*, 47(3), 307-322.
- <sup>67</sup> Walker, D.W. et al., 2006. Hypersensitivity to oxygen and shortened lifespan in a *Drosophila* mitochondrial complex II mutant. *Proceedings of the National Academy of Sciences*, 103(44), 16382 - 16387.
- <sup>68</sup> Russell, M.W. et al., 1997. Cloning of the human NADH: ubiquinone oxidoreductase subunit B13: localization to chromosome 7q32 and identification of a pseudogene on 11p15. *Mammalian Genome: Official Journal of the International Mammalian Genome Society*, 8(1), 60-61.
- <sup>69</sup> Gu, J.Z., Lin, X. & Wells, D.E., 1996. The Human B22 Subunit of the NADH-Ubiquinone Oxidoreductase Maps to the Region of Chromosome 8 Involved in Branchio-Oto-Renal Syndrome. *Genomics*, 35(1), 6-10.
- <sup>70</sup> Lin, X. et al., 1999. Human NDUFB9 gene: genomic organization and a possible candidate gene associated with deafness disorder mapped to chromosome 8q13. *Human Heredity*, 49(2), 75-80.
- <sup>71</sup> Baysal, B.E. et al., 2000. Mutations in SDHD, a mitochondrial complex II gene, in hereditary paraganglioma. *Science* (New York, N.Y.), 287(5454), 848-851.