



Πανεπιστήμιο Θεσσαλίας
Σχολή Επιστημών Υγείας
Τμήμα Βιοχημείας και Βιοτεχνολογίας

University of Thessaly
School of Health Sciences
Department of Biochemistry and
Biotechnology

Πτυχιακή εργασία . Τίτλος:

**« Μετα-ανάλυση μιτοχονδριακού γονιδιώματος σε
φυλές του είδους *Sus scrofa* »**

Bachelor's thesis. Title:

**«Meta-analysis of mitochondrial genome on breeds of
Sus scrofa species»**

Όνομα: Τσιτσικλή Μαγδαληνή

Έτος: 2016-2017

**Εργαστήριο: Γενετικής, Συγκριτικής
και Εξελικτικής Βιολογίας**



Μέλη τριμελούς επιτροπής

- **Z. Μαμούρης:** Καθηγητής Γενετικής Ζωικών Πληθυσμών
(Υπεύθυνος καθηγητής)
- **A. Μούτου:** Αναπληρώτρια Καθηγήτρια Βιολογίας Σπονδυλωτών
- **Θ. Σαραφίδου:** Επίκουρος Καθηγήτρια Μοριακής Γενετικής Ζωικών Οργανισμών



Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον κύριο Ζήση Μαμούρη, υπεύθυνο καθηγητή της εργασίας μου, για την ανάθεση της μελέτης αυτής, καθώς και τις καθηγήτριες Μούτου Αικατερίνη και Σαραφίδου Θεολογία οι οποίες είναι μέλη της τριμελούς επιτροπής μου.

Ευχαριστώ επίσης όλα τα μέλη του εργαστηρίου Γενετικής, Συγκριτικής και Εξελικτικής Βιολογίας για το ευχάριστο κλίμα συνεργασίας και πιο πολύ τον Γιαννούλη Θ. και Τσιπουρλιάνο Α. για την ιδιαίτερη βοήθειά τους



Περίληψη

Ο χοίρος είναι θηλαστικό και ανήκει στην τάξη των αρτιοδακτύλων. Η επιστημονική του ονομασία είναι *Sus scrofa* και στο είδος αυτό κατατάσσονται τόσο οι αγριόχοιροι (*Sus scrofa scrofa*) όσο και οι οικόσιτοι, εξημερωμένοι χοίροι (*Sus scrofa demesticus*). Οι οικόσιτοι χοίροι είναι μορφές ζώων οι οποίες έχουν σχηματιστεί μετά από εκατοντάδες, ή ακόμα και χιλιάδες, χρόνια εξημέρωσης και επιλεκτικής αναπαραγωγής. Τα εξημερωμένα πλέον ζώα κάτω από εξελικτική πίεση διαχωρίστηκαν σε ράτσες οι οποίες εμπεριέχουν άτομα που κατέχουν κοινά γενοτυπικά και φαινοτυπικά χαρακτηριστικά έναντι άλλων ατόμων. Ακόμη και σήμερα ο διαχωρισμός των διαφόρων ρατσών χοίρων γίνεται φαινοτυπικά βάσει μορφολογικών χαρακτηριστικών τους. Ο διαχωρισμός αυτός δεν είναι πάντοτε ακριβής και εγκυμονεί κινδύνους εσφαλμένων αποτελεσμάτων. Ιδανική θα ήταν η εύρεση γενοτυπικών χαρακτηριστικών τα οποία με μικρή ποσότητα δείγματος θα μπορούσαν να δώσουν τον χαρακτηρισμό μιας συγκεκριμένης ράτσας σε ένα άτομο. Ο λόγος αυτός έδωσε κίνητρο να πραγματοποιηθεί η παρούσα εργασία όπου γίνεται προσπάθεια εύρεσης γενοτυπικών διαφορών στο μιτοχονδριακό γονιδίωμα (*mtDNA*) των χοίρων. Ως δεδομένα χρησιμοποιήθηκαν 167 αλληλουχίες ατόμων οι οποίες βρέθηκαν στην ηλεκτρονική βάση δεδομένων *GenBank* του *National Center for Biotechnology Information (NCBI)*, στις οποίες εμπεριέχονται 77 ράτσες (συμπεριλαμβανομένων και εμπορικών ρατσών) . Με την βοήθεια εργαλείων βιοπληροφορικής οι αλληλουχίες αυτές ομαδοποιήθηκαν και μελετήθηκαν με τέτοιο τρόπο ώστε να βρεθούν απλότυποι ή μονονουκλεοτιδικοί πολυμορφισμοί (*SNPs*) οι οποίοι δύνανται να διαχωρίσουν τις ράτσες μεταξύ τους. Μελλοντικός στόχος είναι η περεταίρω μελέτη των γενοτυπικών αυτών χαρακτηριστικών σε ευρύτερη κλίμακα, δηλαδή κατέχοντας ως δεδομένα περισσότερα άτομα της ίδιας ράτσας, και μακροπρόθεσμα η καταχώρηση μοριακών δεικτών οι οποίοι θα είναι ικανοί να κατηγοριοποιήσουν τα διαθέσιμα δείγματα στις ράτσες στις οποίες ανήκουν.



Abstract

The pig is a mammal that belongs to the order Artiodactyla. Its scientific name is *Sus scrofa* and in that species are classified the wild boar (*Sus scrofa scrofa*) and the Domestic pig (*Sus scrofa demesticus*). The domestic pigs are animal forms that have formed after hundreds, or even thousands, of years of domestication and selective breeding. These domesticated animals, under evolutionary pressure, separated into breeds which include individuals, holding common genotypic and phenotypic characteristics over other breeds. Even nowadays, the separation of pig breeds is being phenotypically using the morphological characteristics of the animal. However the separation based on morphologically characteristics is not always accurate and might endanger false results. An ideal situation would be to find genotypic characteristics that could characterize an individual's breed, using a small amount of animal tissue. The Data for this research were 167 sequences of pig individuals which were found on the GenBank database of National Center for Biotechnology Information. (NCBI), in which there were contained 77 breeds of pigs. Bioinformatics and several computational tools helped to group the sequences and make them easy to be studied. This study's aim was the location of several SNPs (single nucleotide polymorphism) or combination of them which leads to an haplotype. Future goal of this study is to study these molecular markers extensively using as data more mitochondrial sequences of several breeds. Eventually pig breeds could be fully separated from each other genotypically , simply by owning a small amount of animal tissue and the right molecular markers.



Περιεχόμενα

Περίληψη.....	3
Abstract	4
1. Εισαγωγή.....	6
1.1 Χοίρος (<i>Sus scrofa</i>)	6
1.1.1 Ταξινόμηση του χοίρου	6
1.1.2 Εξημέρωση.....	8
1.2 Ταξινόμηση και μοριακή φυλογενετική.....	10
1.2.1 Μοριακοί δείκτες έναντι μορφολογικών χαρακτηριστικών	12
1.2.2 Μοριακοί δείκτες.....	12
1.3 Βιοπληροφορική.....	19
1.3.1 Γονιδιωματική	20
1.3.2 Μετα-ανάλυση.....	21
1.3.3 Βάσεις δεδομένων	22
1.3.4. Πως εμφανίζονται οι αλληλουχίες (<i>FORMAT</i>)	26
1.3.5 Περιήγηση στο <i>BLAST</i>	27
1.4 Σκοπός της εργασίας	29
2. Υλικά και Μέθοδοι.....	30
2.1 Αλληλουχία αναζήτησης (<i>query</i>)	30
2.2 Blast.....	30
2.3 BioEdit.....	34
2.4 Πρωτόκολλο εργασίας.....	35
3. Αποτελέσματα	37
4. Συζήτηση.....	48
5. Βιβλιογραφικές αναφορές	52



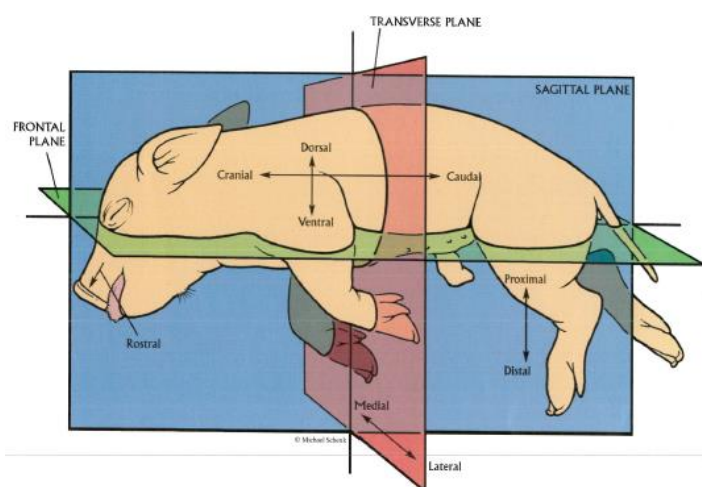
1.Εισαγωγή

1.1 Χοίρος (*Sus scrofa*)

Ο χοίρος είναι οπληφόρο θηλαστικό και ανήκει στην τάξη των αρτιοδακτύλων (*Artiodactyla*, ζώα με άρτιο αριθμό δακτύλων). Η επιστημονική του ονομασία είναι *Sus scrofa* και προέρχεται από την αρχαία ελληνική ονομασία του χοίρου Σϋς. Ανατρέφονται και αναπαράγονται σε όλο τον κόσμο και χρησιμοποιούνται από τους ανθρώπους για διάφορους σκοπούς, από κτηνοτροφία μέχρι και για συντροφιά ως κατοικίδιο ζώο. Είναι από τα πρώτα ζώα που εξημερώθηκαν και αυτό έχει ως αποτέλεσμα σήμερα να υπάρχουν πολλές διαφορετικές ράτσες χοίρου, με ποικιλία διαφόρων χαρακτηριστικών, που εξυπηρετεί ένα ευρύ φάσμα αναγκών, τόσο του ίδιου του ζώου όσο και του ανθρώπου, και επίσης μεγαλύτερη προσαρμοστικότητα στα διάφορα περιβάλλοντα.

1.1.1 Ταξινόμηση του χοίρου

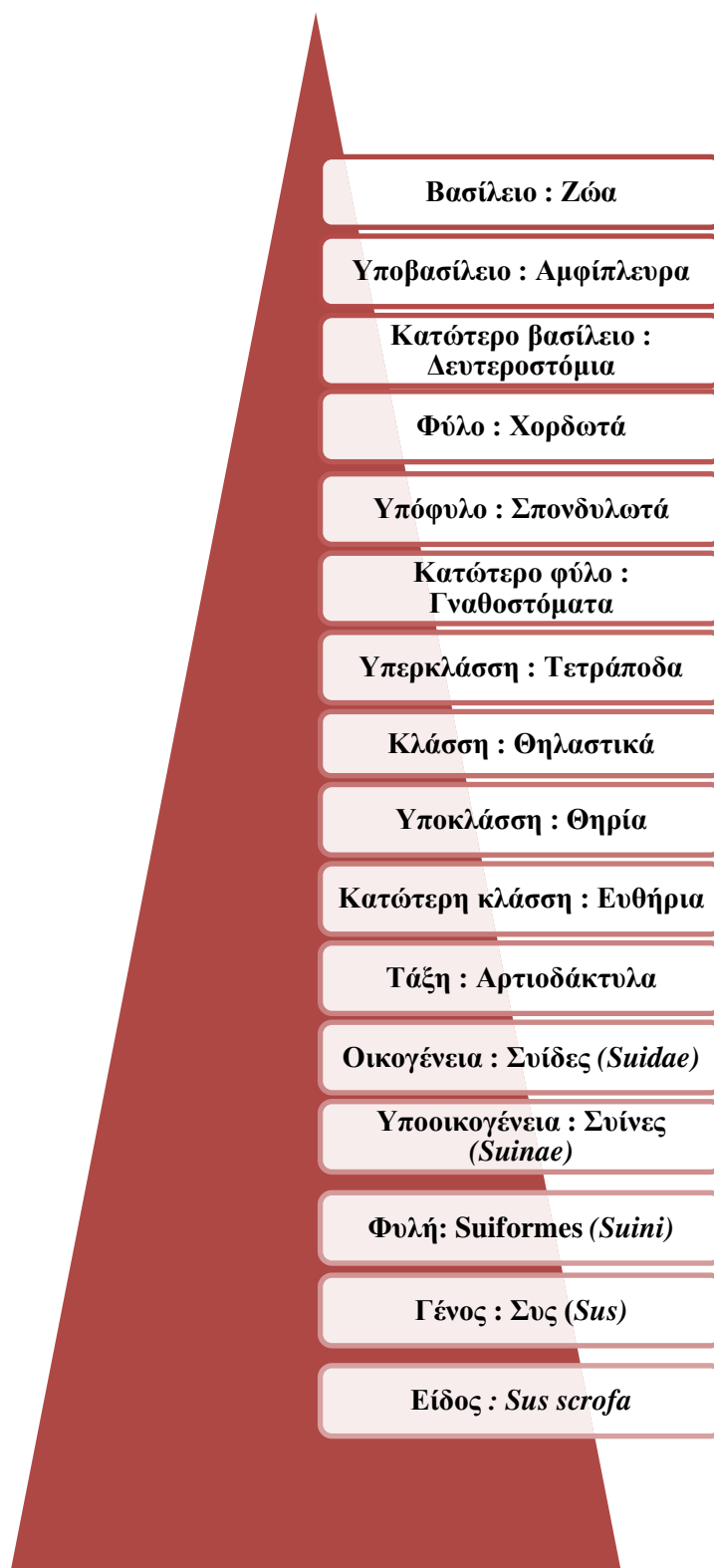
Οι χοίροι που συναντάμε στις φάρμες είναι η μια μορφή ζώου που έχει σχηματιστεί μετά από εκατοντάδες ή ακόμα και χιλιάδες χρόνια επιλεκτικής αναπαραγωγής και εξημέρωσης. Οι άνθρωποι πρώτα εξημέρωσαν τους αγριόχοιρους και έπειτα σχηματίστηκαν ράτσες με μεγάλη ποικιλία μορφής, μεγέθους, χρώματος κ.α. Τα ζώα αυτά επιλέχθηκαν για διάφορους σκοπούς και προσαρμόστηκαν ώστε να επιζούν και αναπαράγονται σε διάφορα περιβάλλοντα. Οι χοίροι είναι θηλαστικά με ογκώδες σώμα, επίπεδη μουσούδα, μικρά μάτια και μεγάλα αυτιά. Είναι ζώα έξυπνα, κοινωνικά και βρίσκονται σε όλο τον κόσμο.



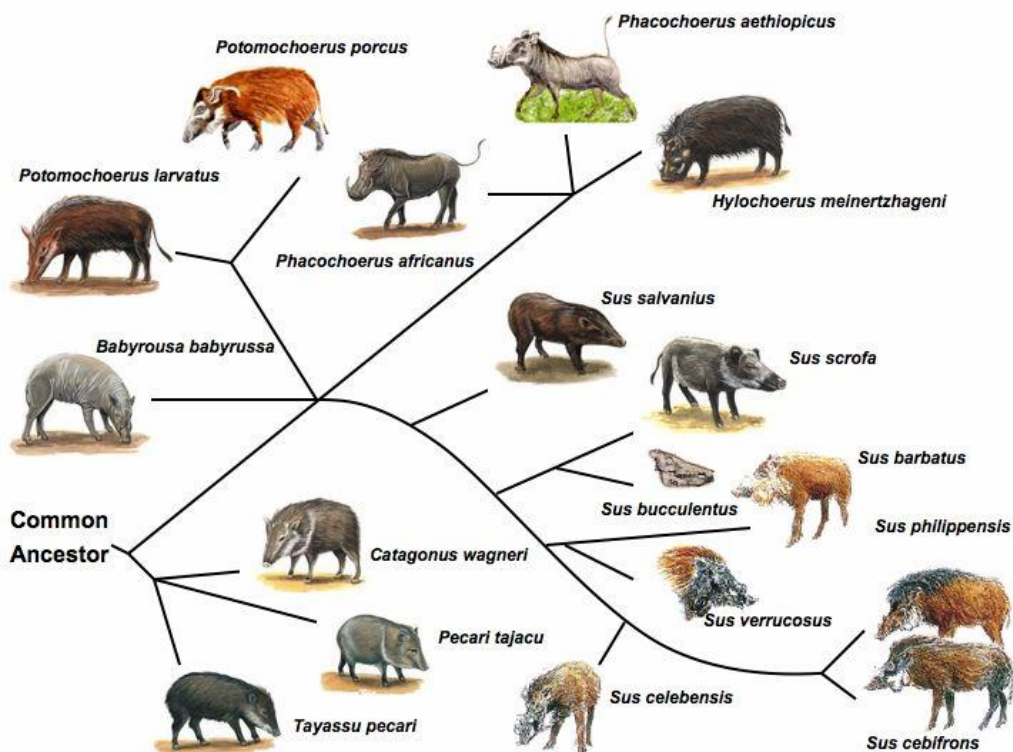
Εικόνα 1 Ανατομία του χοίρου



Ταξινομικά ανήκει στην οικογένεια *Suidae*, η οποία περικλείει οκτώ γένη και δεκαέξι είδη. Σύμφωνα με το *ITIS (International Taxonomic Information System)* η ταξινομική ιεραρχία του χοίρου έχει ως εξής :



Σχεδιάγραμμα 1 Ταξινόμηση του χοίρου



Εικόνα 2 Ποικιλομορφία και φυλογενετικές σχέσεις μεταξύ των *Suini* (*suiformes*)

1.1.2 Εξημέρωση

1.1.2.1 Η εξημέρωση από τον Δαρβίνο μέχρι σήμερα

Όπως είναι γνωστό ο Δαρβίνος επικεντρώθηκε πολύ στην εξημέρωση των οργανισμών, ως μια διαδικασία κατά την οποία προκύπτουν πολλές μεταβολές, αποκλίσεις και ποικιλομορφίες στα γενικότερα χαρακτηριστικά του οργανισμού. Είχε θέσει επανειλημμένως το ερώτημα «γιατί τα εξημερωμένα-οικόσιτα ζώα παρουσιάζουν τόση ποικιλομορφία». Στις αναλύσεις του περί αύξησης της ποικιλομορφίας λόγω της εξημέρωσης, ο Δαρβίνος υπέθεσε, και προσπάθησε να στηρίξει, ότι οι αλλαγές οφείλονται σχεδόν αποκλειστικά στις περιβαλλοντικές επιδράσεις.

Εκτός από τις επιδράσεις του περιβάλλοντος στην ποικιλομορφία, ο Δαρβίνος μελετούσε τα αποτελέσματα των διασταυρώσεων απόμων από φυλές, τόσο ίδιες όσο και από διαφορετικές, από γνωστές φυλές της εποχής του με την βοήθεια και την πληροφόρηση από διάφορους κτηνοτρόφους. Στηριζόμενος στις παρατηρήσεις του,



παρουσίασε ζωντανά παραδείγματα της επιλεκτικής αναπαραγωγής που είναι εμφανή σε αρκετά μικρό χρονικό διάστημα. Τελικά απέδωσε μεγάλο ρόλο στην μη επιλεκτική επιλογή, που δρα πάνω στους ζωικούς οργανισμούς εδώ και χιλιάδες χρόνια. Επειδή η εξημέρωση των ζώων ξεκίνησε σε απομακρυσμένους περιόδους, είναι άγνωστο πότε και σε τι βαθμό τα ζώα άρχισαν να παρεκκλίνουν μεταξύ τους, ο Δαρβίνος υποστήριξε ότι τα ζώα συνέχισαν να αλλάζουν τα χαρακτηριστικά τους ακόμη και πολλά χρόνια μετά την απαρχή της εξημέρωσής.

Εξημέρωση είναι η διαδικασία της γενετικής προσαρμογής ενός άγριου βιολογικού οργανισμού ώστε να ανταπεξέρχεται καλύτερα στις ανάγκες των ανθρώπων. Η εξημέρωση είναι αποτέλεσμα συνθηκών ζωής και εκτροφής που ελέγχονται από τον άνθρωπο για πολλαπλές γενιές (*C.Darwin*)

Οι παρατηρήσεις και τα συμπεράσματα του *Δαρβίνου* όσον αφορά την εξημέρωση των οργανισμών, υποστηρίζονται ακόμη και σήμερα. Πράγματι, το εύρος της διακύμανσης χαρακτηριστικών μέσα σε ένα ορισμένο είδος ξεπερνά σε μέγεθος τη διακύμανση χαρακτηριστικών σε επίπεδο οικογένειας ή ακόμα και τάξης. Παράδειγμα αποτελεί η παρακάτω εικόνα που δείχνει διακύμανση φαινοτυπικών χαρακτηριστικών στο είδος των γατών.



Εικόνα 3 Διακύμανση χαρακτηριστικών στο ίδιο είδος (μέγεθος-σχήμα σώματος, χρώμα-μήκος τριχώματος κ.α.)



1.1.2.2 Εξημέρωση του χοίρου

Οι χοίροι είναι ένα από τα πιο διεσπαρμένα είδη θηλαστικών στον πλανήτη. Η καθοδηγούμενη εξέλιξή τους βασίζεται στην φυσική επιλογή και στην εξημέρωσή τους. Οι εξημερωμένοι χοίροι βρίσκονται σε παγκόσμια κλίμακα σε ένα ευρύ φάσμα περιβαλλοντικών συνθηκών. Πολλά χαρακτηριστικά συμπεριλαμβανομένων των δοντιών, την μορφολογία του κρανίου, το μέγεθος, το τρίχωμα, το χρώμα, τους βιοχημικούς και μοριακούς πολυμορφισμούς, την συμπεριφορά κ.α χρησιμοποιούνται για τον χαρακτηρισμό και διαφοροποίηση κάθε φυλής από τις υπόλοιπες.

Η εξημέρωση των χοίρων έχει παίξει σημαντικό ρόλο στον κτηνοτροφικό κλάδο σε παγκόσμια κλίμακα. Οι γνώσεις σχετικά με την εξέλιξη και την εξάπλωση των εξημερωμένων χοίρων σε όλον τον πλανήτη μας επιτρέπουν στο να εμβαθύνουμε περισσότερο στην προέλευση και εξάπλωση της κτηνοτροφίας.

Οι παλαιότερη αναφορά εξημερωμένων χοίρων χρονολογείται στα 7000 χρόνια π.Χ., στο *Çayönü* της Ανατολίας. Βάσει υποθέσεων μέσα από ζωο-αρχαιολογικά έγγραφα, πιστεύεται ότι οι πρώτοι χοίροι εξημερώθηκαν στην Ανατολική Ευρώπη, και κατόπι εξαπλώθηκαν στην δυτική Ευρώπη και την ανατολική Κίνα. Ωστόσο πρόσφατες μελέτες, που βρίσκονται σε αρχικά στάδια, πάνω στο μιτοχονδριακό DNA, τόσο των Ευρασιατικών αγριόχοιρων όσο και πολλών φυλών οικόσιτων χοίρων, έχουν αποδείξει στοιχεία ότι η εξημέρωση τους έγινε ανεξάρτητα σε διαφορετικά μέρη και το γεγονός αυτό αναφέρεται με τον όρο “multiple and independent domestication hypothesis” (Kijas JM et al). Επίσης από πρόσφατες μελέτες ανάλυσης του μιτοχονδριακού DNA 685 ατόμων, συμπεριλαμβανομένων αγριόχοιρων και οικόσιτων χοίρων του Ευρασιατικού κλάδου, υποστηρίζεται η υπόθεση ότι η εξημέρωση των χοίρων συνέβη ανεξάρτητα σε διάφορες γεωγραφικές περιοχές της Ευρασίας (τρεις από την Άπω Ανατολή, μια από την Νοτιοανατολική Ασία και δυο από την Ευρώπη) (Larson.G. et.al 2005.)

1.2 Ταξινόμηση και μοριακή φυλογενετική

Εάν υποθέσουμε ότι τα γονιδιώματα των οργανισμών εξελίσσονται και διαφοροποιούνται βάσει της προοδευτικής συσσώρευσης μεταλλάξεων στον γενετικό τους κώδικα, τότε η διακύμανση στη διαφορά των νουκελεοτιδικών αλληλουχιών δυο γονιδιωμάτων πρέπει να υποδηλώνει το πόσο πρόσφατα αυτά τα δυο γονιδιώματα έχουν αρχίσει να αποκλίνουν από έναν κοινό πρόγονο. Για παράδειγμα, δυο γονιδιώματα που απέκλιναν στο σχετικά πρόσφατο παρελθόν, ενδέχεται να παρουσιάζουν λιγότερες διαφορές μεταξύ τους σε σχέση με δυο άλλα γονιδιώματα τα οποία απέκλιναν από τον κοινό τους πρόγονο πολλά χρόνια πριν. Η **μοριακή**

Μετα-ανάλυση mtDNA σε φυλές του είδους *Sus scrofa*



φυλογενετική ασχολείται με την σύγκριση γονιδιωμάτων των διαφόρων οργανισμών και τον προσδιορισμό των μεταξύ τους εξελικτικών σχέσεων.

Μοριακή φυλογενετική : είναι η μελέτη των εξελικτικών σχέσεων μεταξύ οργανισμών και γονιδίων βάσει συνδυασμού της μοριακής βιολογίας και στατιστικών τεχνικών

(Li, 1997 .*Molecular Evolution* , Sinauer Press)

Η μοριακή φυλογενετική υπήρχε σαν κλάδος πριν την αλληλούχιση του DNA, και προέρχεται από την παραδοσιακή **κατάταξη των οργανισμών**, με βάση τις ομοιότητες και τις διαφορές τους που εφαρμόστηκε διεξοδικά από τον *Carl Linnaeus* τον δέκατο όγδοο αιώνα. Στόχος του ήταν η λογική κατάταξη των μέχρι τότε γνωστών οργανισμών, και άθελά του έθεσε την απαρχή του πλαισίου ενασχόλησης με τα εξελικτικά μοντέλα, υποδιαιρώντας τους οργανισμούς σε μια σειρά ταξινομικών κατηγοριών βάσει ιεραρχίας. Το μοντέλο κατάταξης που επινόησε ο *Linnaeus* ερμηνεύτηκε ως **φυλογένεση**, και παρείχε πληροφορίες όχι μόνο για τις ομοιότητες που φέρουν τα είδη μεταξύ τους, αλλά και για τις μεταξύ τους εξελικτικές σχέσεις.

In natural science the principles of truth ought to be confirmed by observation.

Carolus Linnaeus

Όταν επιδιώκεται η ταξινόμηση των οργανισμών, αλλά και η μελέτη των φυλογενετικών σχέσεων μεταξύ τους, χρησιμοποιούνται ως δεδομένα **μεταβλητά χαρακτηριστικά** τους που δίνουν στον ερευνητή την δυνατότητα να ομαδοποιήσει ή να διαχωρίσει τους οργανισμούς μεταξύ τους. Τα παλαιότερα χρόνια, τα χαρακτηριστικά που χρησιμοποιούνταν ως “δεδομένα”, αποτελούσαν **μορφολογικά γνωρίσματα** του προς μελέτη οργανισμού, ωστόσο το 1904, ο *Nuttall* ξεκίνησε να εξετάζει διάφορα μοριακά δεδομένα χρησιμοποιώντας ανοσολογικές δοκιμές με σκοπό να ερευνήσει τις σχέσεις ανάμεσα σε διάφορα ζώα. Αν και μέσα από το έργο του *Nuttall* αποδείχτηκε ότι τα μοριακά δεδομένα μπορούν να χρησιμοποιηθούν στον διαχωρισμό και στην ομαδοποίηση οργανισμών, η προσέγγιση αυτή άρχισε να εφαρμόζεται μετά το 1950 μαζί με την εξέλιξη των μοριακών τεχνικών.



1.2.1 Μοριακοί δείκτες έναντι μορφολογικών χαρακτηριστικών

Οι αναλύσεις που στηρίζονται στα **μορφολογικά-φαινοτυπικά χαρακτηριστικά** των ατόμων έχουν αρκετά **μειονεκτήματα**, όπως για παράδειγμα περιπτώσεις στις οποίες οι μορφολογικοί χαρακτήρες είναι δύσκολο να διακριθούν οπότε χρειάζεται και η υποκειμενική γνώμη του ερευνητή, γεγονός που στερεί την αντικειμενική σκοπιά της έρευνας. Επίσης, ενδέχεται ορισμένα μορφολογικά χαρακτηριστικά να συγκλίνουν ή να συνεξελίσσονται, γεγονός που κάνει πιο δύσκολη την διευκρίνιση ομοπλασιών στους διάφορους οργανισμούς (*Moore and Willmer 1997, Conway 2003*). Ένα ακόμη σημαντικό μειονέκτημα χρήσης των μορφολογικών χαρακτηριστικών είναι ότι τα παραπάνω δεν ακολουθούν σαφή πρότυπα κληρονομιάς. Τα παραπάνω προβλήματα μπορούν να λυθούν με την χρήση μοριακών δεικτών, που εμπεριέχουν πληροφορίες μοναδικές για κάθε οργανισμό, συμβάλλοντας στην πιο αντικειμενική και ορθή διερεύνηση των εκάστοτε ερωτημάτων. Δυσκολία συναντάται όμως και σε αυτή την προσέγγιση αφού δεν είναι διαθέσιμοι μοριακοί δείκτες για όλους τους οργανισμούς. Η έρευνα για την εύρεση μοριακών δεικτών έχει προκαλέσει όλο και περισσότερους επιστήμονες τα τελευταία χρόνια.

1.2.2 Μοριακοί δείκτες

Εξαιτίας της μεγάλης προόδου που σημειώθηκε στην εξέλιξη της μοριακής βιολογίας κατά την διάρκεια των τελευταίων ετών, μπορούμε σήμερα να χρησιμοποιούμε διάφορα μοριακά δεδομένα και τεχνικές που μας διευκολύνουν να επιλέξουμε, μέσα από έναν πληθυσμό, συγκεκριμένα άτομα που φέρουν ορισμένους επιθυμητούς χαρακτήρες. Η επιλογή αυτή γίνεται σε μοριακό επίπεδο με την βοήθεια “δεδομένων” που ονομάζονται μοριακοί δείκτες. Οι δείκτες αυτοί μπορεί να είναι **είτε νουκλεοτιδικής είτε πρωτεϊνικής φύσης**, και ελέγχονται με διάφορες τεχνικές ανάλογα το είδος τους και σε ποια κατηγορία κατατάσσονται.

Θεωρητικά, όλες οι περιοχές του DNA που διαφοροποιούνται μπορούν να χρησιμοποιηθούν για μελέτες πληθυσμιακής γενετικής ή γενικά σε ενδοειδικές ή διαειδικές μελέτες (*Pleines et al., 2009*)



Δεν μπορεί να πει κάποιος με βεβαιότητα ότι υπάρχουν συγκεκριμένοι δείκτες που είναι κατάλληλοι για όλα τα είδη μελέτης υπάρχουν όμως κάποια **κριτήρια** που πρέπει να πληροί ο κάθε δείκτης για να χρησιμοποιηθεί σε φυλογενετικές ή εξελικτικές μελέτες. Κάποια από τα κριτήρια αυτά είναι τα εξής:

- Να είναι εύκολος στην απομόνωση και την ανάλυση
- Να έχει απλή δομή
- Να είναι πολυμορφικός
- Να έχει σαφή τρόπο κληρονόμησης
- Να έχει γρήγορο ή αργό ρυθμό εξέλιξης, ανάλογα με τα προς μελέτη είδη

1.2.2.1 DNA δείκτες και πρωτεϊνικοί δείκτες

Όπως προαναφέρθηκε οι **μοριακοί δείκτες** χωρίζονται σε δυο μεγάλες κατηγορίες, πρωτεϊνικής και νουκλεοτιδικής φύσης. Παρακάτω αναφέρονται επιγραμματικά οι δείκτες που χωρίζονται στις κατηγορίες αυτές.

Μοριακοί δείκτες πρωτεϊνικής φύσης:

- I. Ισοένζυμα
- II. Αλλοένζυμα

DNA δείκτες:

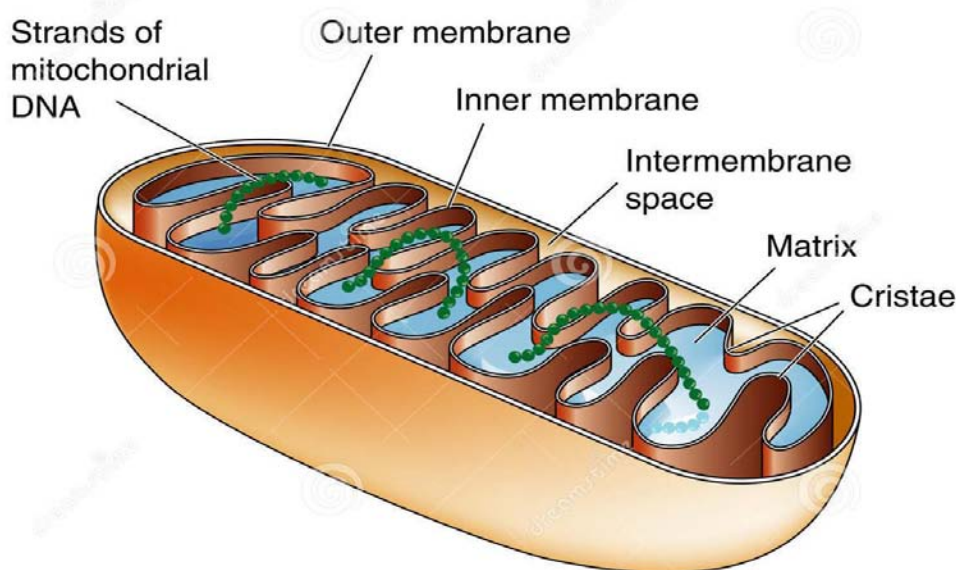
- I. Πολυμορφισμοί μεγέθους τμημάτων περιορισμού. **RFLPs** (*Restriction Fragment Length Polymorphisms*)
- II. Πολυμορφισμοί μεγέθους ενισχυμένων τμημάτων. **AFLPs** (*Amplified Fragment Length Polymorphism*)
- III. Ενίσχυση τυχαίων DNA πολυμορφισμών. **RAPDs** (*Randomly Amplified DNA Polymorphisms*)
- IV. **SSRs** (*Simple Sequence Repeats*) στους οποίους κατατάσσονται οι μίνι- και μικρο-δορυφόροι
- V. Και γενικά αλληλουχίες του πυρηνικού DNA (**nDNA**) και εξωπυρηνικού DNA, όπως είναι το μιτοχονδριακό DNA (**mtDNA**)

1.2.2.2 Μιτοχονδριακό DNA (mtDNA)

Το μιτοχονδριακό DNA περιέχεται σε **πολλά αντίγραφα** μέσα στα μιτοχόνδρια, και έχουν προταθεί **δύο θεωρίες** σχετικά με την **προέλευση** του



γονιδιώματος εντός των συγκεκριμένων οργανιδίων. Σύμφωνα με την **πρώτη**, το γονιδίωμα των οργανιδίων (μιτοχόνδρια και χλωροπλάστες) έχουν αυτογενή καταγωγή μέσα στα ευκαρυωτικά κύτταρα και προέρχονται από θραύσματα του πυρηνικού γονιδιώματος που εγκολπώθηκαν σε μεμβράνες και σχημάτισαν τα μιτοχόνδρια και τους χλωροπλάστες (Cavalier & Smith 1975, Uzell & Spolsky 1981). Η **δεύτερη** θεωρία υποστηρίζει ότι το γονιδίωμα των οργανιδίων αυτών έχει εξωγενή προέλευση, και συγκεκριμένα προέρχονται από βακτηριακούς προγόνους οι οποίοι εισέβαλαν (ή εγκολπώθηκαν) σε πρώτο-ευκαρυωτικά κύτταρα (Margulis, 1981).



Εικόνα 4 Παρουσιάζεται το οργανίδιο μιτοχόνδριο και αναγράφονται χαρακτηριστικές δομές και συστατικά όπως : οι κλώνοι του DNA, η εξωτερική μεμβράνη, η εσωτερική μεμβράνη, ο διαμεμβρανικός χώρος, η μήτρα, οι αναδιπλώσεις

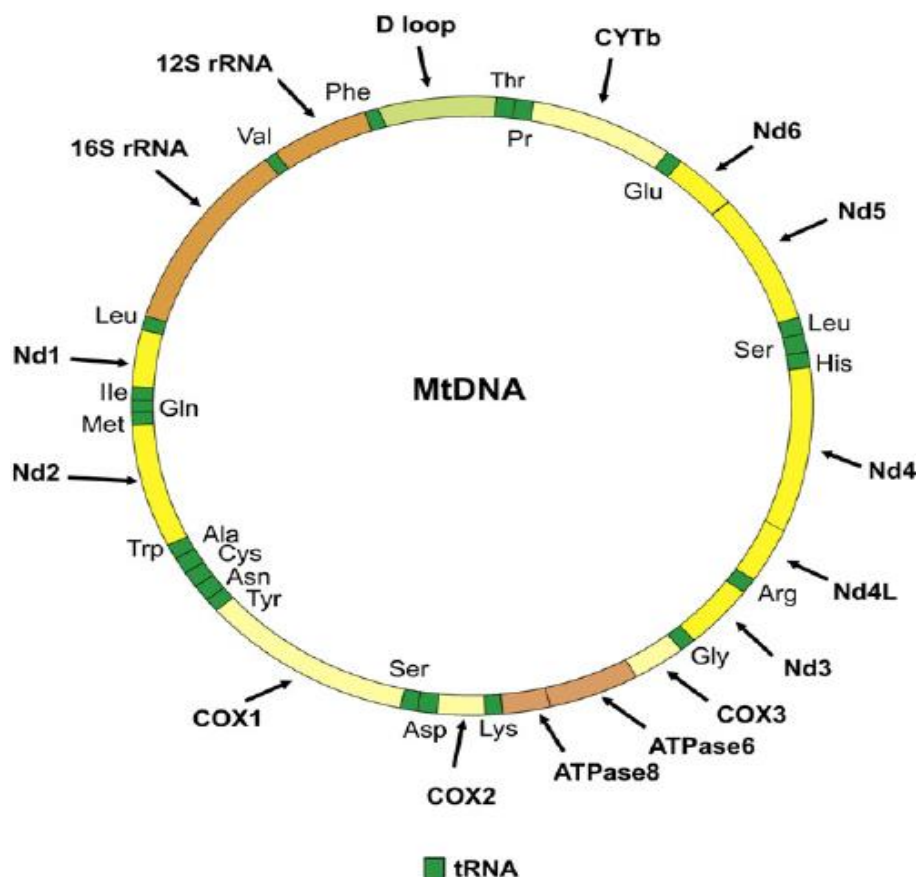
Το μιτοχονδριακό γονιδίωμα αποτελεί ένα κυκλικό δίκλωνο «χρωμόσωμα». Οι δύο αλυσίδες του *mtDNA* είναι η βαριά (**Heavy**), η οποία είναι και η κυρίως μεταφραζόμενη αλυσίδα, και η ελαφριά (**Light**) (Bibb *et al.*,1981). Το ζωικό *mtDNA* περιέχει συνήθως 37 γονίδια, από τα οποία :

- τα δύο κωδικοποιούν **ριβωσωμικό RNA**
- τα 22 κωδικοποιούν **tRNAs**
- τα υπόλοιπα 12 ή 13 κωδικοποιούν **υπομονάδες των πρωτεϊνών** της εσωτερικής μεμβράνης των μιτοχονδρίων, που συμμετέχουν στην αναπνευστική αλυσίδα και στη σύνθεση *ATP*.

Επιπλέον, υπάρχει μία μη κωδική περιοχή (**D-loop**) που ορίζεται ως ρυθμιστική περιοχή (*control region* -**CR**) εξαιτίας του ρόλου που έχει στην αντιγραφή και τη μεταγραφή των μιτοχονδριακών μορίων. Αυτή η περιοχή είναι ταχέως



εξελισσόμενη και χρησιμοποιείται περισσότερο για αναλύσεις σε ενδοειδικό επίπεδο (Awise, 1994). Η *D-loop* των σπονδυλωτών **υποδιαιρείται σε τρία τμήματα**, τα οποία διαφέρουν μεταξύ τους τόσο στην νουκλεοτιδική σύσταση όσο και στο ρυθμό εξέλιξής τους. Το **κεντρικό τμήμα**, που περιέχει το σημείο έναρξης αντιγραφής της βαριάς αλυσίδας, είναι αρκετά συντηρημένο. Αντίθετα, τα **δύο πλευρικά τμήματα** αυτής της περιοχής (τμήματα I και II) είναι συνήθως υπερμεταβλητά όσον αφορά στις νουκλεοτιδικές υποκαταστάσεις και στον αριθμό των μεταλλάξεων προσθήκης-απαλοιφής. Εξαιτίας του αυξημένου ρυθμού εξέλιξης των τμημάτων I και II, η *CR* θεωρείται κατάλληλη κυρίως για μελέτες στο επίπεδο του είδους.



Εικόνα 6 Αναπαράσταση του μιτοχονδριακού DNA

Στην **εικόνα 6** παρουσιάζονται οι διάφορες περιοχές του μιτοχονδριακού γονιδιώματος. Πιο συγκεκριμένα διακρίνονται τα γονίδια των 22 *tRNAs*, το γονίδιο του κυτοχρώματος Β (*CYTB*), τα γονίδια των υπομονάδων της κυτοχρωμικής οξειδάσης (*COX1-3*), τα γονίδια των υπομονάδων της αφυδρογονάσης του *NADH* (*ND1-6*), τα γονίδια των *ATP*ασών (*ATPase6* και *ATPase8*), τα γονίδια που κωδικοποιούν για ριβοσωμικό *RNA* (*16s rDNA* και *12s rDNA*) καθώς και η ρυθμιστική περιοχή *D-loop*.



Οι κωδικές περιοχές είναι συμπαγώς πακεταρισμένες, χωρίς ενδιάμεσα εσώνια. Το μιτοχονδριακό γονιδίωμα δεν περιέχει ιστόνες, έχει περιορισμένη ικανότητα επιδιόρθωσης και ως εκ τούτου ο ρυθμός εμφάνισης των μεταλλάξεων είναι υψηλός (*mutation fixation rate*), 5 -10 φορές υψηλότερος σε σχέση με το πυρηνικό γονιδίωμα, (Brown *et al.*, 1979). Στην κατάσταση αυτή συμβάλλει και το γεγονός ότι το μιτοχονδριακό γονιδίωμα βρίσκεται κοντά στην αναπνευστική αλυσίδα η οποία είναι η κύρια πηγή παραγωγής δραστικώς μορφών οξυγόνου (ROS). Οι δραστικές μορφές οξυγόνου είναι ικανές να προκαλέσουν οξειδωτικές βλάβες στις βάσεις του μιτοχονδριακού DNA (Inna Shokolenko *et al.*, 2009)

Γονίδια και φυλογενετικές αναλύσεις

- Το *12s rDNA* είναι πολύ συντηρημένο και έχει χρησιμοποιηθεί κυρίως στον προσδιορισμό της φυλογένειας σε ανώτερα ταξινομικά επίπεδα, όπως τα φύλα και τα υποφύλα (Wan *et al.*, 2004).
- Το *16s rDNA* χρησιμοποιείται συνήθως για φυλογενετικές μελέτες όσον αφορά σε μεσαία ταξινομικά επίπεδα, π.χ. οικογένειες και σπάνια γένη.
- Συγκριτικά με το *12s* και το *16S rDNAs*, τα **μιτοχονδριακά γονίδια που κωδικοποιούν πρωτεΐνες** εξελίσσονται πιο γρήγορα και είναι ισχυρότεροι δείκτες στον προσδιορισμό της εξελικτικής ιστορίας σε κατώτερα επίπεδα, όπως π.χ. οι οικογένειες, τα γένη και τα είδη (Wan *et al.*, 2004).

Όσον αφορά τα **γονίδια που κωδικοποιούν για πρωτεΐνες**, τα νουκλεοτίδια της τρίτης θέσης στις τριπλέτες έχουν αυξημένο ποσοστό υποκαταστάσεων, γεγονός που καθιστά το ρυθμό μοριακής εξέλιξης περίπου τρεις φορές υψηλότερο αυτού των *12s* ή *16s rDNA* (Knowlton & Weigt, 1998). Το σύνολο των υποκαταστάσεων αυτών χαρακτηρίζει την *υπόθεση ζευγαρώματος wobble* και παρατηρείται ζευγάρισμα ανάμεσα σε βάσεις το οποίο δεν ακολουθεί το μοντέλο ζευγαρώματος των Watson και Crick. Τα τέσσερα κύρια ήδη ζευγών είναι τα εξής: **G-U** (γουανίνη με ουρακίλη), **I-U** (υποξανθίνη με ουρακίλη), **I-A** (υποξανθίνη με αδερίνη) και **I-C** (υποξανθίνη με κυτοσίνη). Στην πραγματικότητα, η εξέλιξη αυτών των γονιδίων είναι αρκετά ταχεία ώστε να μπορεί να επιτευχθεί η διάκριση όχι μόνο κοντινών ειδών, αλλά και φυλογενετικών ομάδων εντός ενός είδους (Cox & Hebert, 2001, Wares & Cunningham, 2001).



1.2.2.3 Το μιτοχονδριακό *DNA* του χοίρου

Το 1998 οι *Bjorn M. Ursing* και *Ulfur Arnason* δημοσίευσαν στο *Journal of molecular evolution* την εργασία τους πάνω στην αλληλούχιση του μιτοχονδριακού γονιδιώματος του χοίρου (*S.scrofa*).

J Mol Evol (1998) 47:302–306

JOURNAL OF **MOLECULAR
EVOLUTION**
© Springer-Verlag New York Inc. 1998

The Complete Mitochondrial DNA Sequence of the Pig (*Sus scrofa*)

Björn M. Ursing, Ulfur Arnason

Department of Genetics, Division of Evolutionary Molecular Systematics, University of Lund, S-223 62 Lund, Sweden

Received: 2 December 1997 / Accepted: 20 February 1998

Μέσα από αυτή την εργασία μπορούμε πλέον να έχουμε στην διάθεσή μας χρήσιμα εργαλεία-γνώσεις που απαιτούνται για να χρησιμοποιηθεί το μιτοχονδριακό *DNA* του χοίρου ως δείκτης για διάφορες δια-ειδικές και ενδο-ειδικές μελέτες.

Το μιτοχονδριακό *DNA* που χρησιμοποιήθηκε στην εργασία αυτή, απομονώθηκε από το ήπαρ ενός αρσενικού χοίρου, και τα αποτελέσματα της αλληλούχισής του καταχωρήθηκαν στην βάση δεδομένων *EMBL*, με μοναδικό κωδικό (*Accession Number*) **AJ002189**.

Τα αποτελέσματα της εργασίας αυτής έδειξαν ότι, το **μήκος** του μιτοχονδριακού *DNA* του συγκεκριμένου χοίρου ήταν περίπου **16,679 νουκλεοτίδια** (αναφέρεται περίπου διότι ο αριθμός των νουκλεοτιδίων μπορεί να αλλάζει, λόγω ετεροπλασμιών που δημιουργούνται από αυξομειώσεις νουκλεοτιδίων, σε επαναλαμβανόμενα μοτίβα ρυθμιστικών περιοχών του μιτοχονδριακού *DNA*).



Feature	Position		Codon	
	From	To	Start	Stop
tRNA-Phe	1	70		
12S rRNA	71	1,032		
tRNA-Val	1,032	1,099		
16S rRNA	1,100	2,668		
tRNA-Leu (UUR)	2,668	2,742		
NADH1	2,745	3,701	ATG	TAG
tRNA-Ile	3,700	3,768		
tRNA-Gln	3,766	3,838	(L)	
tRNA-Met	3,840	3,909		
NADH2	3,910	4,953	ATT	TAG
tRNA-Trp	4,952	5,019		
tRNA-Ala	5,093	5,026	(L)	
tRNA-Asn	5,169	5,095	(L)	
Or. L-stand repl.	5,166	5,212		
tRNA-Cys	5,267	5,202	(L)	
tRNA-Tyr	5,332	5,267	(L)	
COI	5,334	6,878	ATG	TAA
tRNA-Ser (UCN)	6,950	6,882	(L)	
tRNA-Asp	6,958	7,025		
COII	7,026	7,713	ATG	T--
tRNA-Lys	7,714	7,780		
ATPase8	7,782	7,985	ATG	TAA
ATPase6	7,943	8,623	ATG	TAA
COIII	8,623	9,407	ATG	TA-
tRNA-Gly	9,407	9,475		
NADH3	9,476	9,822	ATA	TA-
tRNA-Arg	9,823	9,891		
NADH4L	9,892	10,188	GTG	TAA
NADH4	10,182	11,559	ATG	T--
tRNA-His	11,560	11,628		
tRNA-Ser (AGY)	11,629	11,687		
tRNA-Leu (CUN)	11,688	11,757		
NADH5	11,758	13,577	ATA	TAA
NADH6	14,088	13,559	(L) ATG	TAA
tRNA-Glu	14,157	14,089	(L)	
Cyt b	14,162	15,301	ATG	AGA
tRNA-Thr	15,302	15,369		
tRNA-Pro	15,433	15,369	(L)	
Control region	15,434	16,679		
CSB-1	16,109	16,134		
CSB-2	16,378	16,396		
CSB-3	16,430	16,448		

Εικόνα 6 Αναγράφονται οι διάφορες περιοχές του mtDNA του χοίρου καθώς και οι θέσεις στις οποίες βρίσκονται.

1.2.2.4 Μιτοχονδριακό DNA στις φυλογενετικές μελέτες

Σε συγκριτικές αναλύσεις των διαδικασιών προσαρμογής ή της μοριακής εξέλιξης, αλλά και σε μελέτες βιογεωγραφίας, απαιτείται ο προσδιορισμός των σχέσεων ανάμεσα στα είδη. Η διαλεύκανση αυτών των σχέσεων βασίζεται, σε μεγάλο βαθμό, σε γενετικούς δείκτες (Hurst & Jiggins, 2005). Θεωρητικά, όλες οι περιοχές του DNA που διαφοροποιούνται μπορούν να χρησιμοποιηθούν για μελέτες πληθυσμιακής γενετικής ή γενικά σε ενδοειδικές ή διαειδικές μελέτες (Pleines et al., 2009), ωστόσο ο πιο ευρέως χρησιμοποιούμενος δείκτης, σε μελέτες στο ζωικό βασίλειο είναι το μιτοχονδριακό DNA (mtDNA).



Η “επιτυχία” του *mtDNA* στις φυλογενετικές μελέτες οφείλεται σε πολλά δομικά χαρακτηριστικά, αλλά και σε άλλα, όπως είναι η **έλλειψη ανασυνδυασμού** (Clayton, 1992; Hayashi et al, 1985) αλλά και η **εύκολη απομόνωση** του (Michaels et al, 1982; Robin & Wong, 1988). Επιπρόσθετα, το *mtDNA* διαφόρων *taxa* μπορεί να **πολλαπλασιαστεί με ευκολία** τόσο εξαιτίας του μικρού του μεγέθους όσο και της ύπαρξης πληθώρας **συντηρημένων** (‘universal’) **εκκινήτων** για αρκετούς γενετικούς τόπους. Επιπλέον ο **υψηλός εξελικτικός ρυθμός** του, σε συνδυασμό με το **μικρό δραστικό μέγεθος** του, το οποίο είναι περίπου το ένα τέταρτο σε σχέση με αυτό των πυρηνικών δεικτών, δίνει τη δυνατότητα προσδιορισμού πρόσφατων ιστορικών γεγονότων, χωρίς να απαιτείται ιδιαίτερη προσπάθεια κατά την αλληλούχιση (Hurst & Jiggins, 2005).

Δεν είναι απαραίτητος ο προσδιορισμός της αλληλουχίας όλων των μεταφραζόμενων τμημάτων του *mtDNA* σε μία μελέτη ώστε να εξαχθούν αξιόπιστα αποτελέσματα. Έχειδειχθεί ότι 8.8 *kilobases* (*kb*), που αποτελούν σχεδόν το μισό της αλληλουχίας της μεταφραζόμενης περιοχής, έχει τον ίδιο βαθμό διαφοροποίησης συγκριτικά με ολόκληρη την περιοχή (Silva et al., 2002). Επιπλέον, η διαχείριση μεγάλου όγκου δεδομένων ενέχει κινδύνους για λάθη κατά τη διαδικασία προσδιορισμού της αλληλουχίας τους (Sun et al., 2006).

Η μελέτη περισσότερων γενετικών τόπων, έστω και αν το συνολικό μέγεθος της αλληλουχίας είναι μικρό, παρέχει περισσότερη πληροφορία σε σχέση με τη μελέτη λιγότερων τόπων με μεγαλύτερο συνολικό μέγεθος αλληλουχίας ή την προσθήκη δειγμάτων

Felsestein et al 2006

1.3 Βιοπληροφορική

Η βιοπληροφορική εφαρμόζει «**τεχνικές πληροφορικής**» οι οποίες προέρχονται από επιστημονικά πεδία όπως τα **εφαρμοσμένα μαθηματικά**, η **πληροφορική** και η **στατιστική**, για να κατανοήσει και να οργανώσει τις πληροφορίες που συνδέονται με τη βιολογία, σε μία ευρεία κλίμακα. Πιο συγκεκριμένα χρησιμοποιεί υπολογιστικές μεθόδους για να μελετήσει και να οργανώσει τα βιολογικά δεδομένα που υπάρχουν διαθέσιμα. Συγκεντρώνει, αποθηκεύει και αναλύει βιολογικά δεδομένα.

Βιοπληροφορική (Bioinformatics) είναι η έρευνα, ανάπτυξη, ή εφαρμογή υπολογιστικών εργαλείων και προσεγγίσεων για την επέκταση της χρήσης δεδομένων βιολογίας, ιατρικής, συμπεριφοράς ή υγείας, συμπεριλαμβανομένων εκείνων για την απόκτηση, αποθήκευση, οργάνωση, αρχειοθέτηση, ανάλυση, ή οπτικοποίηση αυτών των δεδομένων.



Σκοπός της βιοπληροφορικής:

- Οργάνωση δεδομένων ώστε να είναι δυνατή η αποθήκευση, ανάκτηση και ενημέρωσή τους
- Ανάπτυξη εργαλείων που επιτρέπουν την ανάλυση βιολογικών δεδομένων
- Ανάπτυξη εργαλείων που επιτρέπουν την ερμηνεία των αποτελεσμάτων της εκάστοτε ανάλυσης

Πεδία στα οποία συνεισφέρει η βιοπληροφορική

Η βιοπληροφορική είναι πλέον αναπόσπαστο κομμάτι της έρευνας, της ανάλυσης, της διάγνωσης κ.α. και έχει συνεισφέρει στην ανάπτυξη και εξέλιξη αρκετών άλλων πεδίων όπως είναι τα παρακάτω:

- **Υπολογιστική βιολογία:** ασχολείται με την κλασσική βιολογία με κύριο αντικείμενο μελέτης την εξέλιξη των ειδών (πληθυσμιακή και θεωρητική βιολογία)
- **Ιατρική πληροφορική (*Medical Informatics*)** : έχει στόχο την κατανόηση και διαχείριση ιατρικών δεδομένων
- ***Cheminformatics*:** ασχολείται με την μελέτη και την εφαρμογή υπολογιστικών μεθόδων, σε συνδυασμό με χημικές και βιολογικές τεχνικές, για τον σχεδιασμό φαρμάκων.
- **Γονιδιωματική (*Genomics*):** ασχολείται με την ανάλυση και σύγκριση γονιδιωμάτων διαφόρων οργανισμών.
- **Πρωτεωμική (*Proteomics*):** ασχολείται με την μελέτη πρωτεϊνών που παράγονται από το γονιδίωμα και πιο συγκεκριμένα με την λειτουργία και τον τρόπο αλληλεπίδρασής με άλλα μόρια.

1.3.1 Γονιδιωματική

Η γονιδιωματική (*Genomics*) είναι τομέας της γενετικής που εφαρμόζει την τεχνολογία του ανασυνδυασμένου DNA, μεθόδους αλληλούχισης DNA (*DNA sequencing methods*) και βιοπληροφορικής (*bioinformatics*) για την αλληλούχιση, την οργάνωση και ανάλυση της δομής και της λειτουργίας γονιδιωμάτων.

(Pevsner et al., 2009).



Η γονιδιωματική χωρίζεται σε τέσσερα διαφορετικά πεδία:

1. **Λειτουργική γονιδιωματική (*Functional Genomics*):** ασχολείται με την έκφραση και λειτουργία γονιδίων και τις αλληλεπιδράσεις τους.
2. **Δομική Γονιδιωματική (*Structural Genomics*):** Καθορισμός της τρισδιάστατης δομής (*3D*) κάθε μιας πρωτεΐνης που κωδικοποιείται από ένα συγκεκριμένο γονιδίωμα.
3. **Επιγονιδιωματική (*Epigenomics*):** Ανάλυση των αναστρέψιμων επιγενετικών τροποποιήσεων του *DNA* (μεθυλίωση του *DNA*) ή των ιστονών (μεθυλίωση, ακετυλίωση, φωσφορυλίωση) ενός κυττάρου που αναφέρεται ως επιγένωμα ή επιγονιδίωμα, και οι οποίες επηρεάζουν την έκφραση γονιδίων, χωρίς να μεταβάλλουν την αλληλουχία του *DNA*.
4. **Μεταγονιδιωματική (*Metagenomics*):** Μελέτη μεταγονιδιωμάτων, δηλαδή γενετικού υλικού που προκύπτει από δείγματα που ανευρίσκονται στο περιβάλλον.

Αρχικά η γονιδιωματική είχε ως στόχο την καταγραφή πληροφοριών που σχετίζονται με το γονιδίωμα των οργανισμών. Όσο περνούσαν τα χρόνια όμως, το φάσμα των καταγεγραμμένων πληροφοριών πλήθαινε, δημιουργώντας ένα μεγάλο όγκο δεδομένων σε ηλεκτρονική μορφή. Τα δεδομένα αυτά από την εκάστοτε έρευνα μπορούν να αναλυθούν περεταίρω με την βοήθεια ενός σχετικά νέου πεδίου στον τομέα της έρευνας, το πεδίο της μετα-ανάλυσης.

1.3.2 Μετα-ανάλυση

Ως μετα-ανάλυση ορίζεται η ερευνητική **μεθοδολογία αναγνώρισης, αξιολόγησης και σύνθεσης των πρωτογενών μελετών** που έχουν πραγματοποιηθεί σε ένα συγκεκριμένο επιστημονικό πεδίο, η οποία στοχεύει στον υπολογισμό ενός συνολικού εκτιμητή μιας θεραπείας ή της σχέσης διαφόρων παραγόντων μιας έρευνας, και στη διερεύνηση της ετερογένειας των δημοσιευμένων μελετών που έχουν συμπεριληφθεί στην ανάλυση. (*Diana B. Petitti, 2000*)

Η μετα-ανάλυση δεν είναι μια απλή διαδικασία μέσω της οποίας συνδυάζονται αυθαίρετα τα δεδομένα των επιμέρους μελετών, αλλά **διέπεται από κανόνες** οι οποίοι πρέπει να εφαρμόζονται από της αρχική φάση του σχεδιασμού έως και το τελικό στάδιο της συγγραφής ενός άρθρου. Η αξιοπιστία των συμπερασμάτων που προκύπτουν εξαρτάται σε μεγάλο βαθμό από την πιστή τήρηση ενός πρωτοκόλλου.



Το 2002 οι *Nancy G. Berman* και *Robert A. Parker* στην εργασία τους *Meta-analysis: Neither quick nor easy* επιχείρησαν να δημιουργήσουν ένα **πρωτόκολλο** για τα αρχικά και κύρια βήματα σχεδιασμού μιας έρευνας/μετα-ανάλυσης όπως παρουσιάζεται παρακάτω.

1. Να οριστεί με σαφήνεια το **αντικείμενο της έρευνας**
2. Να συνταχθεί το **ερευνητικό πρωτόκολλο**, το οποίο αποτελεί τόσο ένα λεπτομερές προσχέδιο της έρευνας που θα πραγματοποιηθεί, όσο και ένα μόνιμο αρχείο των αρχικών στόχων, μεθόδων και διαδικασιών της μετα-ανάλυσης
3. Να εκτιμηθούν εξαρχής **οι απαραίτητοι πόροι** (υλικά και ανθρώπινο δυναμικό), δηλαδή οι ανάγκες α) ερευνητικού προσωπικού β) μέσων, όπως ηλεκτρονικοί υπολογιστές, λογισμικά και γραφική ύλη, και γ) προμήθειες δημοσιευμένων μελετών ή πρωτογενών δεδομένων
4. Να εξασφαλισθεί **έγκριση χρήσης** μη δημοσιευμένων ή πρωτογενών **στοιχείων**, στην περίπτωση διενέργειας μετα-ανάλυσης ατομικών δεδομένων (*individual level meta-analysis*)

Τα πρωτογενή δεδομένα που αναφέρθηκαν προηγουμένως είναι δημοσιευμένα στοιχεία που υπάρχουν στο διαδίκτυο από προηγούμενες μελέτες. Υπάρχει ένα ευρύ φάσμα τέτοιων δεδομένων σε ηλεκτρονική μορφή και μπορεί κάποιος με μια απλή αναζήτηση στο διαδίκτυο να βρει από αλληλουχίες γονιδίων μέχρι και τρισδιάστατες μορφές πρωτεϊνών (αναφορικά δυο παραδείγματα από τον τεράστιο όγκο δεδομένων).

1.3.3 Βάσεις δεδομένων

Μια βιολογική βάση δεδομένων είναι ένα μεγάλο **οργανωμένο σύστημα δεδομένων** (εγγράφων), που συνδέεται συνήθως με κατάλληλο λογισμικό για την ενημέρωση, αναζήτηση και ανάκτηση στοιχείων των δεδομένων που έχουν αποθηκευτεί. Τα δεδομένα που αποθηκεύονται στις βιολογικές βάσεις δεδομένων μπορεί να είναι δομές πρωτεϊνών, δεδομένα γονιδιακής έκφρασης και γενικότερα οποιαδήποτε σημαντικά δεδομένα προκύπτουν από έρευνες στον τομέα της Μοριακής Βιολογίας.

Οι βιολογικές τράπεζες δεδομένων μπορούν να κατηγοριοποιηθούν είτε ανάλογα με την προέλευση των δεδομένων που περιέχουν ή ανάλογα με τον τύπο των ακολουθιών που περιέχουν.

- **Πρωτογενείς** (*primary/archival*) ΒΔ, οι οποίες περιέχουν πειραματικά δεδομένα που αφορούν αλληλουχίες. Παραδείγματα γνωστών ΒΔ είναι η *GenBank* (*Benson et al., 2003*), η *EMBL-Bank* (*Stoesser et al., 2003*) και η *DDBJ* (*Miyazaki et al., 2003*).



- **Δευτερογενείς (secondary/curated)** ΒΔ, οι οποίες βασίζονται σε δεδομένα των πρωτογενών ΒΔ όμως περιέχουν επιπλέον πληροφορίες για μοτίβα ή πρότυπα, μεταλλάξεις ακόμα και εξελικτικές σχέσεις μεταξύ διαφόρων οργανισμών, οι οποίες δεν έχουν απαραίτητα πειραματική υποστήριξη. Παραδείγματα γνωστών δευτερογενών ΒΔ είναι η *Swiss-Prot*, η *PROSITE* και η *PRINT*.

Κατηγοριοποίηση με βάση τον τύπο των δεδομένων που περιέχουν:

- **Νουκλεοτιδικών αλληλουχιών** με κυριότερα παραδείγματα τις πρωτογενείς *NCBI*, *DDBJ* και *EBI*.
- **Πρωτεϊνικών αλληλουχιών** με κυριότερα παραδείγματα τις δευτερογενείς *PROSITE*, *PRINT*, *Swiss-Prot*, *UniProt*.
- **Γονιδιωματικές βάσεις δεδομένων** με κυριότερο παράδειγμα την *ensembl* καθώς και άλλες βάσεις δεδομένων που αναφέρονται αποκλειστικά σε κάποιους οργανισμούς (π.χ. *Flybase*, *MGI Mouse Genome* κ.ά.)
- **Δομικές βάσεις δεδομένων** που περιλαμβάνουν δομές πρωτεϊνών, νουκλεϊκών οξέων και υδατανθράκων με κυριότερα παραδείγματα τις *SCOP (Structural Classification of Proteins)* και *CATH (Protein Structure Classification)*.
- **Βάσεις δεδομένων με RNA** όπως η *Rfam*, η *snoRNADB* και η *mirBase*.
- **Βάσεις δεδομένων με microarray** όπως η *GPX* και η *ArrayExpress*.

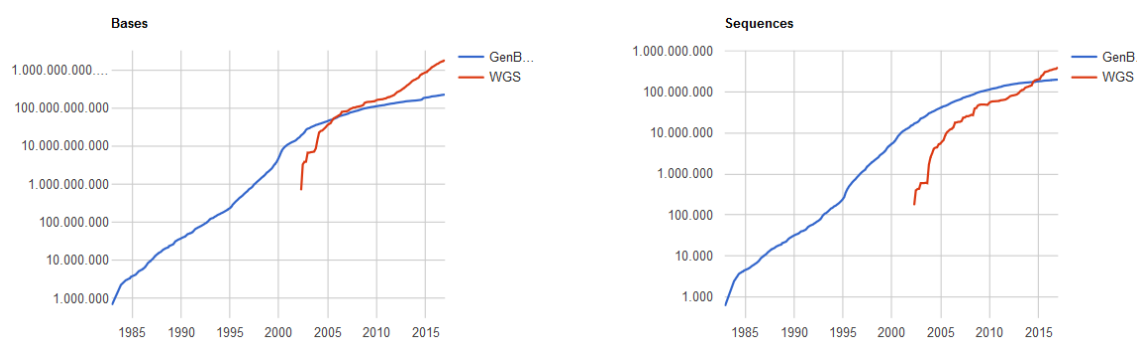
1.3.3.1 Βάση δεδομένων GenBank

Η πρώτη “αποθήκη” βιολογικών αλληλουχιών δημιουργήθηκε το 1979 (*Los Alamos Sequence Database*). Το 1982 η βάση αυτή ονομάστηκε *GenBank* και φιλοξενήθηκε στο Εθνικό Κέντρο Πληροφοριών Βιοτεχνολογίας, *NCBI*, των ΗΠΑ, όπου συνεχίζει να βρίσκεται σήμερα. Το 2014 η βάση δεδομένων *GenBank* (<https://www.ncbi.nlm.nih.gov/genbank/>) διαθέτε αλληλουχίες από περίπου 300.000 είδη. Ο χρόνος διπλασιασμού των αλληλουχιών που βρίσκονται στην *GenBank* υπολογίζεται κατά μέσο όρο στους 18 μήνες! Ο αριθμός μεμονωμένων αλληλουχιών (*GenBank sequences*) αλλά και ολόκληρα γονιδιώματα οργανισμών (*Whole.Genome.Shotguns*) που είχαν υποβληθεί μέχρι τον Δεκέμβριο του 2016 ήταν :



	GenBank		WGS	
Ημ/νία	Bases	Sequences	Bases	Sequences
12/ 2016	224.973.060.433	198.565.475	1.817.189.565.845	395.301.176

Πίνακας 1 Αριθμός μεμονωμένων αλληλουχιών και ολόκληρα γονιδιώματα οργανισμών που έχουν υποβληθεί στην GenBank μέχρι τον Δεκέμβριο του 2016



Εικόνα 7 Γραφική παράσταση της ταχύτητας συσσώρευσης δεδομένων στη βάση αλληλουχιών GenBank από το 1982 έως Δεκέμβριο του 2016. Στο αριστερό διάγραμμα απεικονίζεται ο αριθμός των βάσεων και στο δεξιό ο αριθμός των αλληλουχιών που υποβάλλονται. Με την κόκκινη γραμμή απεικονίζονται ξεχωριστά τα δεδομένα που συγκεντρώνονται από τα προγράμματα αλληλούχησης του συνόλου του γονιδιώματος διαφόρων ειδών, των Whole Genome Shotguns (WGSs).

Η GenBank συλλέγει πρωτογενή πειραματικά δεδομένα που υποβάλλονται από τους ερευνητές. Η υποβολή είναι τυποποιημένη και γίνεται με τη βοήθεια ειδικού λογισμικού (εφαρμογές *Bankit*, *Sequin* και *tbl2asn*). Οι αλληλουχίες υποβάλλονται σε έναν στοιχειώδη έλεγχο, κυρίως σε σχέση με το ελάχιστο της πληροφορίας που πρέπει να περιέχουν, και λαμβάνουν έναν μοναδικό αριθμό υποβολής (*accession number*). Οι εγγραφές έχουν τη μορφή αρχείων με τυποποιημένα σχόλια, όπως στο παράδειγμα της εικόνας 8 και 9. Η ευθύνη για την εγκυρότητα των εγγραφών που υποβάλλονται στην GenBank ανήκει στον ερευνητή που την υποβάλλει.



Sus scrofa breed Lanyu mitochondrion, complete genome

GenBank: DQ518915.2

[FASTA](#) [Graphics](#)

Go to:

LOCUS DQ518915 16747 bp DNA circular MAM 26-JUL-2016
DEFINITION Sus scrofa breed Lanyu mitochondrion, complete genome.
ACCESSION DQ518915
VERSION DQ518915.2
KEYWORDS .
SOURCE mitochondrion Sus scrofa (pig)
ORGANISM *Sus scrofa*
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
Sus.
REFERENCE 1 (bases 1 to 16747)
AUTHORS Chen,C.H., Huang,H.-L., Yang,H.-Y., Lai,S.-H., Yen,N.-T., Wu,M.-C.
and Huang,M.C.
TITLE Mitochondrial genome of Taiwan pig (Sus Scrofa)
JOURNAL Afr. J. Biotechnol. 10 (13), 2556-2561 (2011)
REMARK Publication Status: Online-Only
REFERENCE 2 (bases 1 to 16747)
AUTHORS Chen,C.H., Yang,H.Y., Yen,N.T. and Huang,M.C.
TITLE Direct Submission
JOURNAL Submitted (27-APR-2006) Breeding and Genetics Division, Livestork
Research Institute, COA, 112 Farm Road, Huinhua, Tainan 71246,
Taiwan
REFERENCE 3 (bases 1 to 16747)
AUTHORS Chen,C.H., Yang,H.Y., Yen,N.T. and Huang,M.C.
TITLE Direct Submission

Εικόνα 8 Μια τυπική καταχώρηση της GenBank. Η αλληλουχία αφορά το είδος *Sus scrofa* (χοίρος) και συγκεκριμένα το μιτοχονδριακό γονιδίωμα. Η αλληλουχία με μοναδικό αριθμό υποβολής (ACCESSION) DQ518915 έχει τροποποιηθεί 2 φορές, γι' αυτό αναφέρεται ως VERSION DQ518915.2 .

```
gene /product="tRNA-Glu"
15476..16615
/gene="cytb"
CDS 15476..16615
/gene="cytb"
/codon_start=1
/transl_table=2
/product="cytochrome b"
/protein_id="ABF49561.1"
/translation="MTNIRKSHFLMKIINNAFIDLPAFNSISWNPGLGICLILQ
ILTGLFLAMHYTSDTTTAFSSVTHICRDVNYGWVIRYLHANGASMPFICLFIVGRGL
YGSYMPLETWNIQVLLPTVMATAPMGVYVLPWQMSFWGATVITNLLSAIPYIGTDL
VENIWGGFSDKATLRRFFAFHFLPFIITALLAAVHLLFLHETGNSNPTGISSDMDKI
PFHPYTIKIDILGALFMMLLILLVLPSPDLLGDPDNYTPANPLNTPPHIKPEWYFLF
AYAILRSIPNKLGGVLAALVASILILMLPLHTSKQRSQMSFRFLSQCLFWMVLVADLT
LTWIGGQFVEHPFIIIGQLASILYFLIILVMPITSIIENLLKW"
tRNA 16616..16683
/product="tRNA-Thr"
tRNA complement(16684..16747)
/product="tRNA-Pro"
ORIGIN
1 caaccaaaac aagcattcca ttogtatgca aaccaaaacg ccaagtactt aattactatc
61 tttaaaacaa aaaaaccatc aaaaattgca cacaascata caaatatgcy accccaaaaa
121 ttaaacattt aaaaacaaa aattaatatc attatagccc tatgtacgct gtcgatatac
181 tgetagtccc catgcatata agcatgtaca tattattatt aatattacat agtacaatc
241 attattgata gtacatagca catatcatgt caaataatc cagtcaacat gctgatcact
301 gccactagat cacgagctta actaccatgc cgcgtgaaac cagcaaccoc cttggcaggg
361 atccctcttc tcgctccggg cccataaatt gtcggggggtt ctattgatga actttaacag
421 gcactcgggt cttacttcag gaccatctca cctaaaatcg cccactcttt ccccttaaat
481 aagacatctc gatggactaa tgactaatca gcccatgctc acacataact gagatttcat
541 gcatttggta ttttttaatt tttggggatg cttagactca gccatggcgc tcaaaaggcc
601 taacacagtc aaatcaatg tagctggact tcatggaact catgatccgg cacgataatc
661 caaacaaggt gctattcagt caatggttac aggacataac atgcgtacac gtacacatgc
721 atacacatcc atacacatcc atacacatcc atacacatcc atacacatcc atacacatcc
```

Εικόνα 9 Συνέχεια της καταχώρησης της εικόνας 8, εμφανίζονται όλες οι κωδικές και μη κωδικές περιοχές (γονίδια, μη κωδικές αλληλουχίες κλπ) και στο κάτω μέρος της σελίδας βρίσκεται ολόκληρη την αλληλουχία της καταχώρησης. Η GenBank δίνει την δυνατότητα αυτόματης μετατροπής της αλληλουχίας αυτής σε FASTA format επιλέγοντας με κλικ την λέξη FASTA στο πάνω μέρος της καταχώρησης.



Εικόνα 10 Η κοινοπραξία INSDC αποτελείται από το NCBI, το EMBL και την τράπεζα δεδομένων της Ιαπωνίας DDBJ

Το NCBI, το EMBL και η Τράπεζα Δεδομένων DNA της Ιαπωνίας (*DNA Data Bank of Japan (DDBJ)*), έχουν σχηματίσει μια διεθνή κοινοπραξία, την *International Nucleotide Database Collaboration (INSDC)*. Ανεξάρτητα σε ποιον από τους τρεις οργανισμούς υποβληθεί μια αλληλουχία, τα πληροφορικά συστήματα των τριών ιδρυμάτων ανταλλάσσουν τα δεδομένα τους και ενημερώνουν τις βάσεις για τις νέες αλληλουχίες που υποβλήθηκαν κάθε εικοσιτετράωρο, με αποτέλεσμα να περιέχουν ακριβώς τις ίδιες πληροφορίες. Σύμφωνα με την

πολιτική της INSDC, η πληροφορία παραμένει ελεύθερα διαθέσιμη για πάντα, χωρίς ειδικές αδειοδοτήσεις, με σκοπό το κοινό όφελος.

1.3.4. Πως εμφανίζονται οι αλληλουχίες (*FORMAT*)

Ένα πολύ διαδεδομένο *format* για δεδομένα νουκλεοτιδικών και πρωτεϊνικών ακολουθιών είναι το *FASTA format*. Μια ακολουθία σε *FASTA format* αρχίζει με μια γραμμή περιγραφής και ακολουθούν στις επόμενες γραμμές τα δεδομένα της ακολουθίας. Η γραμμή περιγραφής ξεκινά με το σύμβολο ">". (*Lipman & Pearson, 1985*)

Παράδειγμα πρωτεϊνικής ακολουθίας σε *FASTA format*:

```
>|cl|Query_21577:9827-10123 gi|11055671|gb|AF304203.1|AF304203 Sus scrofa breed  
Swedish wild boar mitochondrion, partial genome
```

```
GTGCCATTAGTATATATAAACATCATTATAGCATTTCGCGATCGCCCTTGCAGGGTT  
ACTTATATATATCGATCTCACTTAATATCTTCACTACTATGCCTAGAAGGAATGATACT  
ATCACTATTCATCATATCGACTCTAATTATCCTAAACACACACTTCACCCTAGCTAA  
CATAATACCCATTATTTTACTAGTGTFTTGCAGCCTGCGAAGCTGCACTAGGCCTGT  
CACTACTAGTAATAGTATCCAACACATACGGTACCGATTACGTCCAAAACCTTAAAC  
CTTTTACAATGCTAA
```



1.3.5 Περιήγηση στο *BLAST*

Μεταξύ των προσεγγίσεων **γρήγορων αναζητήσεων**, αυτή που ξεχωρίζει είναι η μέθοδος του *BLAST* (*Basic Local Alignment and Search Tool*). Το *BLAST* αναπτύχθηκε από τον *Stephen Altschul* και τους συνεργάτες του το 1990 (*Altschul et al. 1990*), τη στιγμή που ο αριθμός των αλληλουχιών στις βάσεις δεδομένων καθιστούσε ήδη επιτακτική την ύπαρξη αλγορίθμων για γρήγορες και αποτελεσματικές συγκρίσεις. Η **ευριστική προσέγγισή** του, που προχωρά μέσω του σταδιακού αποκλεισμού αλληλουχιών, κάνει τον αλγόριθμο εξαιρετικά γρήγορο, κάτι που επιτρέπει την ενσωμάτωσή του σε διαδικτυακούς εξυπηρετητές (*servers*) και την απευθείας εκτέλεσή του από οποιονδήποτε έχει πρόσβαση στο διαδίκτυο. Το γεγονός αυτό κάνει το *BLAST* την πιο ευρέως χρησιμοποιούμενη εφαρμογή βιοπληροφορικής.

Από πρακτικής πλευράς το *BLAST* αποσκοπεί να λύσει το παρακάτω πρόβλημα:

Δεδομένης μιας αλληλουχίας DNA/πρωτεΐνης (αναζήτηση, *query*) και μιας βάσης δεδομένων αλληλουχιών DNA/πρωτεϊνών (στόχος, *target*) να προσδιοριστούν οι αλληλουχίες του στόχου που έχουν σημαντική ομοιότητα με την αλληλουχία αναζήτησης πάνω από ένα συγκεκριμένο όριο.

1.3.5.1 Ο αλγόριθμος του *Blast*

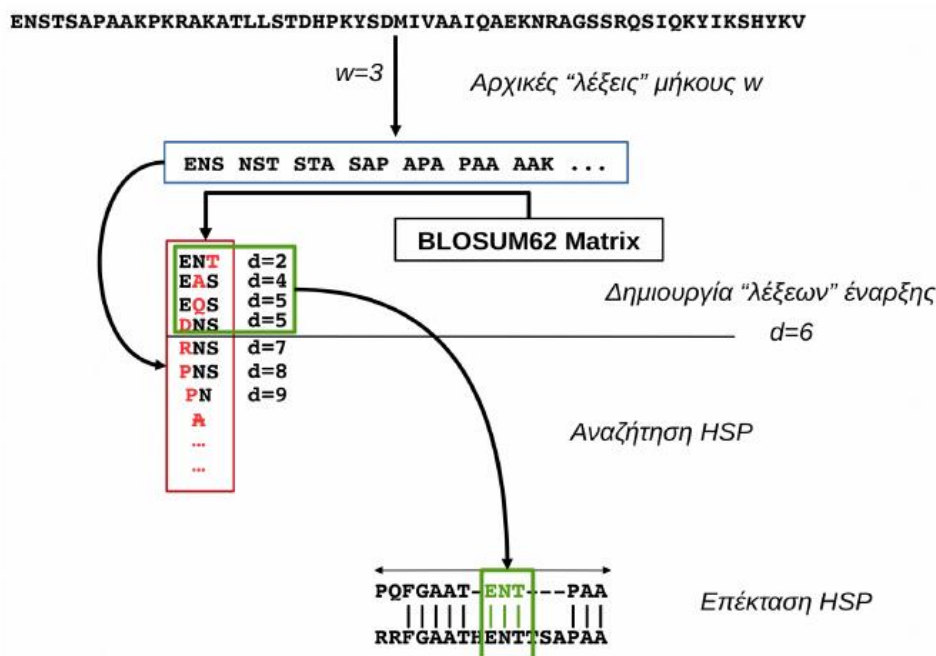
Το *BLAST* χρησιμοποιεί μια προσέγγιση μέσω της οποίας αποκλείει σταδιακά αλληλουχίες της βάσης δεδομένων-στόχου που δεν πληρούν κάποιες βασικές προϋποθέσεις. Η διαδικασία έτσι βαθμιαία μειώνει τις υποψήφιες προς στοίχιση αλληλουχίες, κάτι που κάνει την εφαρμογή εξαιρετικά γρήγορη.

Πιο αναλυτικά, ο **αλγόριθμος του *BLAST*** μπορεί να χωριστεί σε τρία βήματα:

- I. Στο **πρώτο βήμα** το *BLAST* δημιουργεί ένα σύνολο λέξεων που θα χρησιμοποιήσει στη συνέχεια της αναζήτησης. Η αλληλουχία αναζήτησης (*query*) διαιρείται σε υπο-αλληλουχίες συγκεκριμένου μήκους, το οποίο μπορεί να επιλέξει ο χρήστης σαν παράμετρο. Στη συνέχεια, για καθεμία από αυτές τις αλληλουχίες, το *BLAST* δημιουργεί μια ομάδα "γειτονικών" αλληλουχιών μέσω σημειακών αντικαταστάσεων. Από αυτές αποθηκεύει σε ένα σύνολο υπο-αλληλουχιών αυτές των οποίων η απόσταση από την αρχική υπο-αλληλουχία (όπως αυτή υπολογίζεται από έναν πίνακα αντικατάστασης) δεν ξεπερνά ένα όριο απόκλισης που επίσης δίνεται από το χρήστη. Στο τέλος αυτού του πρώτου σταδίου έχει δημιουργηθεί ένα σύνολο "λέξεων" που θα αποτελέσουν τα αρχικά σημεία της αναζήτησης.



- II. Στο **δεύτερο βήμα** το *BLAST* εντοπίζει όλες τις αλληλουχίες της βάσης δεδομένων του στόχου (*target*) που περιέχουν τουλάχιστο μία από τις "λέξεις" που δημιουργήθηκαν στο προηγούμενο στάδιο. Η διαδικασία αυτή μπορεί να είναι η πιο χρονοβόρα καθώς εξαρτάται από τις παραμέτρους μήκους και βαθμού απόκλισης των λέξεων, καθώς και από το πλήθος και τα μήκη των αλληλουχιών του στόχου. Είναι ωστόσο σε κάθε περίπτωση ταχύτερη από μια διαδικασία που θα περιλάμβανε στοιχίσεις καθώς απαιτείται ακριβής ταύτιση "λέξης"-στόχου και κάθε αλληλουχία που δεν περιέχει τουλάχιστο μια λέξη αυτόματα απορρίπτεται από τη συνέχεια. Το αποτέλεσμα αυτού του σταδίου είναι ένα σύνολο ζευγών "λέξεων"-στόχου που το *BLAST* ονομάζει ζεύγη υψηλής βαθμολογίας (*high-scoring pairs, HSP*) και που αποτελούν τα δεδομένα εισόδου για το τρίτο και τελευταίο βήμα.
- III. Στο **τελευταίο αυτό στάδιο**, το *BLAST* προεκτείνει τα *HSP* και προς τις δύο κατευθύνσεις και για όσο μήκος το αποτέλεσμα της στοίχισης δεν πέφτει κάτω από ένα όριο. Όταν η τελική επέκταση κάθε *HSP* ξεπερνά σε μήκος ένα όριο (που επίσης μπορεί να ρυθμιστεί ως παράμετρος του αλγορίθμου), το *BLAST* θεωρεί τη στοίχιση επιτυχή, την επιστρέφει ως αποτέλεσμα μαζί με ένα *score* ομοιότητας και μια στατιστική εκτίμηση της σημασίας.



Εικόνα 11 Αναπαράσταση του αλγορίθμου του BLAST στην στοίχιση μιας πρωτεϊνικής αλληλουχίας



1.4 Σκοπός της εργασίας

Σκοπός της παρούσας μελέτης είναι ο εντοπισμός πολυμορφικών θέσεων σε γνωστές ράτσες του είδους *Sus scrofa* οι οποίες θα μπορούσαν να χρησιμοποιηθούν για την ανάπτυξη μοριακών δεικτών για την ταυτοποίηση των ρατσών. Το μέσο για την πραγματοποίηση είναι η μετα-ανάλυση του συνόλου των διαθέσιμων μιτοχονδριακών γονιδιωμάτων του είδους *Sus scrofa* που βρίσκονται καταχωρημένα σε βάσεις δεδομένων με χρήση εργαλείων βιοπληροφορικής.



2. Υλικά και Μέθοδοι

2.1 Αλληλουχία αναζήτησης (*query*)



Εικόνα 12 Χοίρος ράτσας Landrace

Η αλληλουχία που χρησιμοποιήθηκε ως πρωταρχικό αντικείμενο αναζήτησης (αλληλουχία επερώτησης), ήταν η νουκλεοτιδική αλληλουχία ολόκληρου του μιτοχονδριακού *DNA* ενός χοίρου ράτσας *Landrace*. Η αλληλουχία εντοπίστηκε στην βάση δεδομένων της *GenBank* με μοναδικό αριθμό αναζήτησης (*Accession number*): *NC_000845*.

Sus scrofa mitochondrion, complete genome

NCBI Reference Sequence: NC_000845.1

Εικόνα 13 τίτλος καταχώρησης τα αλληλουχίας στο NCBI

Το μήκος της αλληλουχίας είναι 16613 bp. Επίσης η αλληλουχία διατίθεται σε *FASTA format* τόσο για όλο το μήκος της, όσο και για κάθε λειτουργική ή ρυθμιστική περιοχή ξεχωριστά.

2.2 Blast

Χρησιμοποιήθηκε η νουκλεοτιδική αλληλουχία του καθενός από τα παρακάτω γονίδια, έτσι ώστε να εντοπιστούν στην βάση δεδομένων τα ομόλογά τους σε όσο το δυνατόν περισσότερες ράτσες χοίρων.

Γονίδιο	Μήκος αλληλουχίας	Θέση στο mtDNA
tRNA-Phe	70 bp	1176..1245
12S ribosomal DNA	960 bp	1246..2205
tRNA-Val	68 bp	2206..2273



16S ribosomal RNA	1571 bp	2274..3844
tRNA-Leu	75 bp	3845..3919
ND1	955 bp	3922..4876
tRNA-Ile	69 bp	4877..4945
tRNA-Gln	73 bp	4943..5015
tRNA-Met	70 bp	5017..5086
ND2	1042 bp	5087..6128
tRNA-Trp	68 bp	6129..6196
tRNA-Ala	68 bp	6203..6270
tRNA-Asn	75 bp	6272..6346
tRNA-Cys	66 bp	6379..6444
tRNA-Tyr	66 bp	6444..6509
COX1	1545 bp	6511..8055
tRNA-Ser	71 bp	8059..8129
tRNA-Asp	68 bp	8135..8202
COX2	688 bp	8203..8890
tRNA-Lys	67 bp	8891..8957
ATP8	204 bp	8959..9162
ATP6	681 bp	9120..9800
COX3	784 bp	9800..10583
tRNA-Gly	69 bp	10584..10652
ND3	346 bp	10653..10998
tRNA-Arg	69 bp	11000..11068
ND4L	297 bp	11069..11365
ND4	1378 bp	11359..12736
tRNA-His	69 bp	12737..12805



tRNA-Ser	59 bp	12806..12864
tRNA-Leu	70 bp	12865..12934
ND5	1821 bp	12935..14755
ND6	528 bp	14739..15266
tRNA-Glu	69 bp	15267..15335
CYTB	1140 bp	15342..16481
tRNA-Thr	68 bp	16482..16549
tRNA-Pro	64 bp	16550..16613

Πίνακας 2 Τα γονίδια του mtDNA που χρησιμοποιήθηκαν στην έρευνα

Πιο αναλυτικά χρησιμοποιήθηκε η ιστοσελίδα *Blast* (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch), και συγκεκριμένα το *Blastn* το οποίο ενδείκνυται για αναζήτηση νουκλεοτιδικών αλληλουχιών χρησιμοποιώντας ως *Query* μια νουκλεοτιδική αλληλουχία .

Εικόνα 14 Αρχική σελίδα του Blast

Οι παράμετροι που χρησιμοποιήθηκαν είναι οι εξής:

- **Query Sequence** : Για κάθε γονίδιο ενσωματώθηκε η αντίστοιχη αλληλουχία
- **Database**: Επιλέχθηκε το *Nucleotide collection* (γίνεται ευρεία αναζήτηση σε νουκλεοτιδικές βάσεις δεδομένων)
- **Organism**: Επιλέχθηκε ο οργανισμός *Sus scrofa* (*taxid:9823*)



- **Program selection:** Ως αλγόριθμος επιλέχθηκε το *Highly similar sequences* (*megablast*) που χρησιμοποιείται, φέροντας καλύτερα αποτελέσματα, όταν αναζητάμε οργανισμούς που είναι ομόλογοι με ποσοστό πάνω από 95%.

The screenshot shows the NCBI BLAST search interface. The 'Enter Query Sequence' section contains a text area with a FASTA sequence for *Sus scrofa* mitochondrion. Below it, the 'Job Title' is set to 'NC_000845.1:8959-9162 Sus scrofa mitochondrion, ...'. The 'Choose Search Set' section has the following settings: Database: Nucleotide collection (nr/nt); Organism: *Sus scrofa* (taxid:9823); Exclude: Models (XM/XP), Uncultured/environmental sample sequences, Sequences from type material; Limit to Entrez Query: empty. The 'Program Selection' section has 'Optimize for' set to 'Highly similar sequences (megablast)'. A 'BLAST' button is visible at the bottom.

Εικόνα 15 Παράδειγμα αναζήτησης στο Blast (συμπληρωμένοι παράμετροι)

The screenshot shows the 'General Parameters', 'Scoring Parameters', and 'Filters and Masking' sections of the NCBI BLAST search interface. The 'General Parameters' section has 'Max target sequences' set to 1000, 'Short queries' checked, 'Expect threshold' set to 1e-10, 'Word size' set to 28, and 'Max matches in a query range' set to 0. The 'Scoring Parameters' section has 'Match/Mismatch Scores' set to 1,-2 and 'Gap Costs' set to Linear. The 'Filters and Masking' section has 'Filter' checked for 'Low complexity regions' and 'Species-specific repeats for: Homo sapiens (Human)', and 'Mask' checked for 'Mask for lookup table only'. A 'BLAST' button is visible at the bottom.

Εικόνα 16 Παράδειγμα αναζήτησης στο Blast (συμπληρωμένοι παράμετροι)



2.3 BioEdit

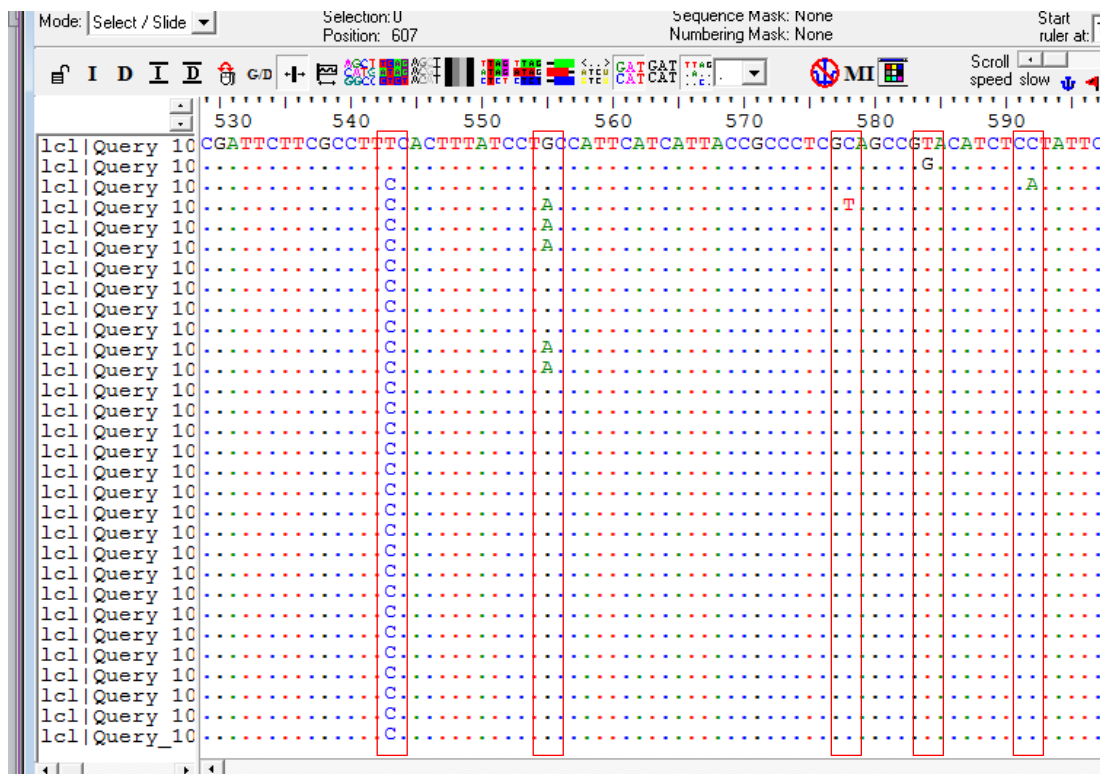
(Επίσημη ιστοσελίδα <http://www.mbio.ncsu.edu/BioEdit/bioedit.html>)



Εικόνα 17 Παράδειγμα στοίχισης αλληλουχιών του γονιδίου COX3 στο πρόγραμμα BioEdit

Το *BioEdit* είναι ένα χρήσιμο επιστημονικό εργαλείο το οποίο βοηθάει στον αυτοματοποιημένο, εύκολο χειρισμό των αλληλουχιών που μας ενδιαφέρουν. Μερικές από τις λειτουργίες που δύναται να πραγματοποιήσει είναι οι εξής:

- **Στοίχιση των αλληλουχιών** (πρωτεϊνικών ή νουκλεοτιδικών) οι οποίες φορτώνονται στο πρόγραμμα
- **Εμφάνιση πολυμορφικών θέσεων** και εύκολη αναγνώριση διαφορετικών αλληλομόρφων λόγω των διαφορετικών χρωμάτων που αποδίδονται σε κάθε μονάδα της αλληλουχίας (αμινοξύ ή βάση)
- **Αναζήτηση ανοιχτού αναγνωστικού πλαισίου (ORF)** σε κωδικές νουκλεοτιδικές αλληλουχίες, με την χρήση διαφόρων παραμέτρων που επιλέγονται από τον ερευνητή.



Εικόνα 18 Εμφάνιση πολυμορφικών θέσεων σε νουκλεοτιδική αλληλουχία

2.4 Πρωτόκολλο εργασίας

- I. Στο πρώτο βήμα πραγματοποιήθηκε *BLASTn* με ακολουθία επερώτησης ολόκληρη την αλληλουχία του μιτοχονδριακού γονιδιώματος του χοίρου (*NC_000845*). Τα αποτελέσματα της αναζήτησης που επιλέχθηκαν, αποθηκεύτηκαν σε *FASTA* format σε έγγραφο κειμένου. Με περαιτέρω επεξεργασία, τελικά επιλέχθηκαν οι αλληλουχίες για τις οποίες ήταν γνωστή η ράτσα προέλευσης. (167 αλληλουχίες, 77 ράτσες)
- II. Στο δεύτερο βήμα όπως προαναφέρθηκε πραγματοποιήθηκε *BLASTn*, ξεχωριστά για το καθένα από τα **37 γονίδια** (*tRNAs* και γονίδια που κωδικοποιούν για πρωτεΐνες) τα οποία αναφέρονται στον **πίνακα 2**. Τα αποτελέσματα του κάθε *BLAST* αποθηκεύτηκαν ξεχωριστά σε αρχεία κειμένου ώστε να χρησιμοποιηθούν στα επόμενα βήματα
- III. Στο τρίτο βήμα για κάθε γονίδιο ξεχωριστά ομαδοποιήσαμε τις αλληλουχίες βάσει κοινής ακολουθίας ώστε να είναι λιγότερες οι προς μελέτη αλληλουχίες και πιο εύκολος ο χειρισμός του κάθε γονιδίου. Η ομαδοποίηση των αλληλουχιών έγινε με το διαδικτυακό πρόγραμμα *BLASTClust* (<https://toolkit.tuebingen.mpg.de/blastclust/>)



- IV. Στο τέταρτο βήμα οι ομαδοποιημένες αλληλουχίες κάθε γονιδίου στοιχήθηκαν και επιλέχθηκαν μόνο οι πολυμορφικές θέσεις οι οποίες καταγράφηκαν σε διαφορετικό αρχείο.

Για τα επόμενα δυο βήματα χρησιμοποιήθηκαν μόνο οι ράτσες χοίρων για τις οποίες είχαμε στην διάθεσή μας δύο ή και παραπάνω αλληλουχίες από διαφορετικά άτομα.

- V. Στο πέμπτο βήμα για κάθε γονίδιο βρέθηκαν οι πολυμορφικές θέσεις που υπάρχουν για τα άτομα της κάθε ράτσας και καταγράφηκαν ξεχωριστά
- VI. Στο έκτο και τελευταίο βήμα για κάθε γονίδιο έγινε αναζήτηση μοναδικών θέσεων οι οποίες εμφανίζονται σε όλα τα άτομα μιας και μόνο ράτσας



3. Αποτελέσματα

Στο κεφάλαιο αυτό παρουσιάζονται τα αποτελέσματα τα παρούσας διπλωματικής εργασίας με την βοήθεια επεξηγηματικών πινάκων.

Στον **πίνακα 3** παρουσιάζεται το πλήθος των ρατσών που ταυτοποιήθηκαν στα αποτελέσματα του *BLASTn*. Οι ράτσες αυτές ανήκουν στο είδος *Sus scrofa* και έχει καταχωρηθεί στην *GeneBank* ολόκληρη η αλληλουχία του μιτοχονδριακού γονιδιώματός τους. Καταγράφεται επίσης ο αριθμός των ατόμων που εμπεριέχονται σε κάθε ράτσα και ο αριθμός καταχώρησής τους στην βάση δεδομένων (*Accession number*). Με έντονο χρώμα εμφανίζονται οι ράτσες για τις οποίες είναι διαθέσιμα παραπάνω από ένα άτομα.

Άτομα	Ράτσα	Accession number
1	Aba	EF545578.1
1	Bamei	EF545583.1
3	Berkshire	KC505410.1
		AY574045.1
		KP765602.1
2	Bihu	EF545590.1
		EF545591.1
1	Changbai mountains wild boar	KP765605.1
1	Chenghua	KP765603.1
1	Congjiang miniature	KM250424.1
1	Dabieshan black pig	KP294522.1
1	Dahe	GQ220329.1
2	Dahuabai	KM200762.1
		AF486870.1
1	Daweizi	KF472179.1
2	Diannan	KM044240.1
		AF486869.1
1	Ding Yuan	KJ737417.1
1	DLY	KF569218.1
4	Duroc	FJ236996.1
		AY337045.1
		FJ236997.1
		AF486858.1



1	Erhualian (Chinese)	AF486861.1
6	European wild boar	FJ237000.1
		FJ236998.1
		FJ237001.1
		FJ236999.1
		FJ237003.1
		FJ237002.1
1	Guanling	KM101042.1
1	Hampshire	AY574046.1
1	Hang	KJ737418.1
1	Huai	KJ737419.1
1	Huo shou black	KJ737420.1
2	Huzu	EF545587.1
		EF545588.1
6	Iberian	FJ236993.1
		FJ236994.1
		FJ236992.1
		FJ236991.1
		FJ236995.1
		EU117375.1
1	Italian wild boar	AF304201.1
4	Jeju native	DQ334860.1
		DQ334861.1
		DQ274110.1
		AY334492.2
1	Jeuma pig	KP223728.1
1	Jiangquhai (Chinese)	AF486872.1
2	Jinhua (Chinese)	AF486863.1
		KC469586.1
8	Korean wild boar	GQ351599.1
		DQ207755.1
		AY574047.1
		EU090703.1
		DQ207754.1
		DQ207753.1
		DQ268530.1
		EU090702.1
2	Lai wu	KJ737421.1
		KJ720205.1
3	Landrace	AF486866.1
		AF034253.1
		AF304202.1
2	Lantang	KC250274.1



		KM101043.1
4	Lanyu	GU147934.1
		EF375877.3
		DQ972936.3
		DQ518915.2
7	Large White	KF752550.1
		JN601074.1
		JN601075.1
		AP003428.1
		AY574048.1
		AF486874.1
		KC250275.1
1	Liangshan	KP257598.1
1	Lingao	KP681242.1
2	Long lin	KJ737422.1
		KM433673.1
3	Luchuan	KJ737423.1
		KM275217.1
		KP126954.1
1	Ma shen	KJ746662.1
3	Mangalica	KJ746666.1
		JN601066.1
		JN601068.1
5	Meishan	KM998967.1
		JN601070.1
		AF304200.1
		AF486855.1
		AB292606.1
4	Min pig	KF971862.1
		KF888634.1
		GQ220328.1
		AF486864.1
2	Nei jiang	KJ746663.1
		KC505406.1
2	Ningxiang	KF472178.1
		AF486857.1
1	Nuogu	DQ466081.2
1	Ohmini Pig	AB298688.1
2	Penzhou	KJ746664.1
		KC505407.1
1	Pietrain	KC469587.1
1	Qianshao Spotted pig	KF660222.1
3	Qingping	EF545581.1



		EF545582.1
		AF486865.1
1	Qingyu	KP257599.1
2	Rongchang	KM044239.1
		AF486860.1
2	Saba	EF545574.1
		EF545567.1
1	Sandu black	KM094194.1
1	Shanggao Mengshan	EF545575.1
1	Shaziling	KF472177.1
1	Swedish wild boar	AF304203.1
2	Taoyuan	DQ534707.2
		KF601700.1
7	Tibetan	KM073256.1
		KC493609.1
		KC493610.1
		KC493611.1
		KC493608.1
		KC493612.1
		KC493607.1
1	Tongcheng (Chinese)	AF486862.1
1	Turopolje	JN601072.1
1	Wannanhua (Chinese)	AF486873.1
1	Wei	EF545577.1
1	Wenchang	KP681244.1
14	Wild boar	EF545584.1
		KP681245.1
		EF545580.1
		EF545570.1
		EF545572.1
		EF545571.1
		KC505411.1
		EF545586.1
		EF545569.1
		EF545573.1
		EF545579.1
		EF545585.1
		EF545568.1
		EF545592.1
		EU333163.1
1	Wuhuang pig	KP765604.1
1	Wujin pig	KC505408.1
1	Wuyi Black	KM259826.1



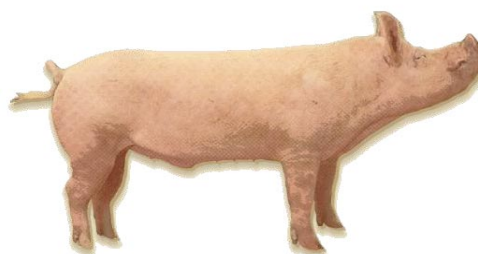
4	Wuzhishan	KJ909516.1
		KF767443.1
		AF486867.1
		KP681243.1
4	Xiang	KP126939.1
		KC250273.1
		EF545593.1
		AF486859.1
1	Ya cha	KJ746665.1
1	Yanan pig	KC505409.1
2	Yimenghei	EF545589.1
		AF486868.1
1	Yushanhei (Chinese)	AF486871.1
2	Zang	EF545576.1
		AF486856.1

Πίνακας3 Οι αλληλουχίες ατόμων και τα Accession numbers τους κατεναμημένες σε ράτσες

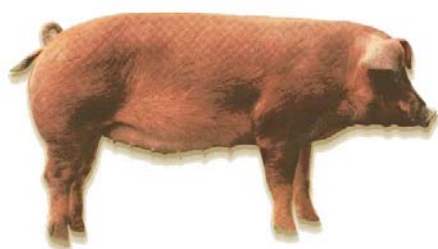
Στις παρακάτω εικόνες φαίνονται κάποιες από τις **εμπορικότερες ράτσες** χοίρων στον κόσμο οι οποίες λόγω των φαινοτυπικών και γενοτυπικών χαρακτηριστικών τους χρησιμοποιούνται για διασταυρώσεις με άλλες ράτσες χοίρων , στην κτηνοτροφία , στην βιομηχανία κλπ. Άτομα των ρατσών αυτών ενυπάρχουν στις αλληλουχίες-δεδομένα που χρησιμοποιούνται στην παρούσα διπλωματική εργασία.



Εικόνα 20 Berkshire (Βρετανική ράτσα γνωστή και ως Black Pork)



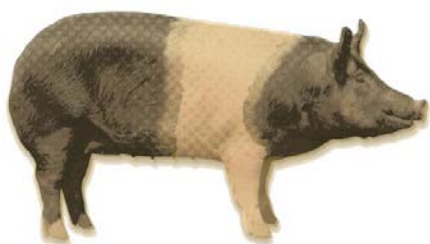
Εικόνα 21 Large white (η πιο εμπορική ράτσα σε πάνω από 60 χώρες, γνωστή και ως Yorkshire)



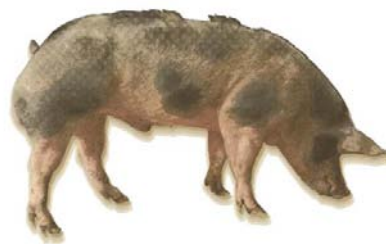
Εικόνα 22 Duroc (Βρετανική ράτσα)



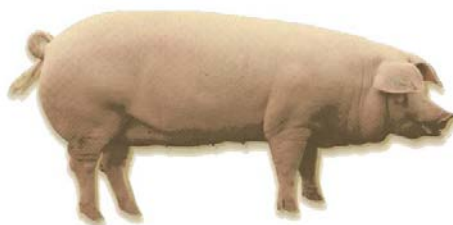
Εικόνα 23 Mangalica (γνωστό ως Woolly pig)



Εικόνα 24 Hampshire (Αμερικάνικη ράτσα)



Εικόνα 24 Piterain (Βέλγικη ράτσα η οποία χρησιμοποιείται για διασταυρώσεις στην Βρετανία)



Εικόνα 25 Landrace (διασταυρώσεις σε Δυτική Ευρώπη και Βόρεια Αμερική)

Στον **πίνακα 4** παρουσιάζεται το σύνολο των πολυμορφικών θέσεων για κάθε γονίδιο ξεχωριστά. Πιο συγκεκριμένα στην στήλη **variable** καταγράφονται οι θέσεις στις οποίες συναντάται έστω και ένας πολυμορφισμός σε μια μοναδική αλληλουχία ενός ατόμου, ενώ στην στήλη **parsimony** καταγράφονται οι θέσεις στις οποίες ένας συγκεκριμένος πολυμορφισμός συναντάται σε παραπάνω από δυο άτομα. Στις στήλες **variable/nucleotide** και **parsimony/nucleotide** υπολογίζεται το ποσοστό πολυμορφισμού (*variable*) και *parsimony* που αντιστοιχεί σε κάθε νουκλεοτίδιο της αλληλουχίας.

Τα μικρότερα ποσοστά εμφανίζονται στα *tRNAs* τα οποία είναι μικρού μήκους και περισσότερο συντηρημένα μέσα στο είδος και τα μεγαλύτερα στα γονίδια που κωδικοποιούν για τις υπομονάδες της αλυσίδας ηλεκτρονίων (αναπνευστική αλυσίδα) του μιτοχονδρίου, αλλά και τα γονίδια *12s* και *16s* που κωδικοποιούν για ριβοσωμικό *RNA*.



Όνομα γονιδίου	gene length	variable	parsimony	variable/nucleotide	parsimony/nucleotide
ATP6	681	88	27	0,1292	0,039647577
ATP8	204	27	8	0,1323	0,039215686
COX1	1545	162	39	0,1048	0,025242718
COX2	688	74	23	0,1075	0,033430233
COX3	784	70	22	0,0893	0,028061224
CYTB	1140	122	49	0,1070	0,042982456
ND1	958	100	30	0,1044	0,03131524
ND2	1042	122	50	0,1171	0,047984645
ND3	346	39	11	0,1127	0,031791908
ND4	1378	157	50	0,1139	0,03628447
ND4L	297	42	18	0,1414	0,060606061
ND5	1828	204	53	0,1116	0,028993435
ND6	528	42	17	0,0795	0,03219697
tRNA Ala	68	6	4	0,0882	0,058823529
tRNA Arg	69	5	-	0,0725	-
tRNA Asn	75	4	-	0,0533	-
tRNA Asp	68	9	1	0,1323	0,014705882
tRNA Cys	66	2	1	0,0303	0,015151515
tRNA Gln	75	3	1	0,0400	0,013333333
tRNA Glu	69	4	1	0,0580	0,014492754
tRNA Gly	69	6	1	0,0869	0,014492754
tRNA His	69	5	1	0,0725	0,014492754
tRNA Ile	70	5	1	0,0714	0,014285714
tRNA Leu	75	5	1	0,0667	0,013333333
tRNA Leu (2)	70	6	4	0,0857	0,057142857
tRNA Lys	67	3	-	0,0448	-
tRNA Met	70	8	2	0,1143	0,028571429
tRNA Phe	71	13	4	0,1831	0,056338028
tRNA Pro	65	5	2	0,0769	0,030769231
tRNA Ser	72	5	-	0,0694	-
tRNA Ser (2)	59	4	1	0,0678	0,016949153
tRNA Thr	69	6	2	0,0869	0,028985507
tRNA Trp	68	10	1	0,1470	0,014705882
tRNA Tyr	66	3	1	0,0454	0,015151515
tRNA Val	68	7	2	0,1029	0,029411765
16s rDNA	1582	137	28	0,0866	0,017699115
12 s rDNA	984	112	13	0,1138	0,013211382

Πίνακας 4 Νουκλεοτιδικοί πολυμορφισμοί για την αλληλουχία κάθε γονιδίου



Στον **πίνακα 5** παρουσιάζονται τα ποσοστά *variable/nucleotide* και *parsimony/nucleotide* συνολικά για τα σύμπλοκα της αναπνευστικής αλυσίδας. Το *Complex I* περιέχει τα γονίδια *ND1-6*, το *Complex III* το γονίδιο *CYTB*, το *Complex IV* τα γονίδια *COX1-3* και το *Complex V* τα γονίδια *ATP6* και *ATP8*. Το υψηλότερο ποσοστό *variable/nucleotide* παρατηρείται στο *Complex V* ενώ το υψηλότερο ποσοστό *parsimony/nucleotide* παρατηρείται στο *Complex III*.

variable/complexe(nucl)	parsimony/complex(nucl)
<i>Complex I</i>	<i>Complex I</i>
0,110710365	0,035910303
<i>Complex III</i>	<i>Complex III</i>
0,107017544	0,042982456
<i>Complex IV</i>	<i>Complex IV</i>
0,101425257	0,027842227
<i>Complex V</i>	<i>Complex V</i>
0,129943503	0,039548023

Πίνακας 5 Στατιστικά Νουκλεοτιδικών πολυμορφισμών στα σύμπλοκα της αναπνευστικής αλυσίδας

Όσον αφορά την περαιτέρω ανάλυση των πολυμορφικών θέσεων κάθε γονιδίου που αναφέρθηκαν στον **πίνακα 4** έγινε προσπάθεια καταγραφής και χρήσης μόνο των πολυμορφικών θέσεων οι οποίες δίνουν δεδομένα για τον σχετικό διαχωρισμό των ρατσών μεταξύ τους. Πιο συγκεκριμένα χρησιμοποιήθηκαν μόνο οι θέσεις στις οποίες ένας πολυμορφισμός εμφανίζεται σε όλα τα άτομα μιας ή και περισσότερων ρατσών. Στον **πίνακα 6** παρουσιάζονται το πλήθος των θέσεων αυτών σε κάθε γονίδιο εκτός από τα γονίδια των *tRNAs* τα οποία δεν εμφάνισαν θέσεις στις οποίες οι πολυμορφισμοί να ανήκουν σε όλα τα άτομα κάποιας ράτσας.

Γονίδιο	Πολυμορφικές θέσεις
ATP6	8
ATP8	5
COX1	15
COX2	8
COX3	7
CYTB	15
ND1	9
ND2	20
ND3	6
ND4	15
ND4L	8
ND5	18
ND6	9
12s rDNA	5
16s rDNA	8

Πίνακας 6 Αριθμός θέσεων στις οποίες υπάρχει πολυμορφισμός σε όλα τα άτομα μιας ή και περισσότερων ρατσών



Αν μπορούμε να αναφέρουμε αυτές τις πολυμορφικές θέσεις των γονιδίων ως έναν απλότυπο για κάθε ράτσα, τότε βάσει των αποτελεσμάτων της ανάλυσης γίνεται διαχωρισμός τεσσάρων ρατσών οι οποίες παρουσιάζουν κοινό απλότυπο που διαφέρει από αυτούς των υπόλοιπων ρατσών. Οι τέσσερις αυτές ράτσες είναι οι εξής : *Duroc*, *European wild boar*, *Iberian* και *Mangalica*. Στην **εικόνα 26** εμφανίζονται ως παράδειγμα τα γονίδια *ATP6*, *ATP8* και *COX1* προς οπτικοποίηση των παραπάνω λεγόμενων . Με **X** συμβολίζονται οι πολυμορφισμοί στα άτομα τα ίδιας ράτσας

(λόγω όγκου δεδομένων δεν μπορούν να οπτικοποιηθούν όλες οι πολυμορφικές θέσεις για όλα τα γονίδια , γι' αυτό τον λόγο είναι διαθέσιμα σε ξεχωριστό αρχείο.)

Ράτσες	Άτομα	ATP6							ATP8					COX1																
		27	36	237	355	407	501	554	591	21	119	120	188	197	336	363	381	420	433	519	732	750	858	897	924	1080	1161	1248	1428	
Duroc	4	T	C	C	T	T	G	A	T	C	T	T	T	C	T	A	T	G	T	C	C	A	T	A	T	T	C	C	C	
European wild boar	6	T	C	C	T	X	X	A	T	C	T	T	T	C	T	A	T	G	T	C	C	A	T	A	T	T	C	C	C	
Iberian	6	T	C	C	T	X	G	A	T	C	T	T	T	C	T	A	T	G	T	C	C	A	T	A	T	T	C	C	C	
Mangalitsa	3	T	C	C	T	T	G	A	T	C	T	T	T	C	T	A	T	G	T	C	C	A	T	A	T	T	C	C	C	
Berkshire	3	C	T	T	C	C	G	G	C	T	C	C	C	T	A	G	C	A	X	T	T	G	C	G	C	C	T	T	T	
Bihu	2	C	T	T	C	C	G	G	C	T	C	C	C	T	A	G	C	A	C	T	T	G	C	G	C	C	T	T	T	
Dahubai	2	C	T	T	C	C	G	G	C	T	C	C	C	T	A	G	C	A	C	T	T	G	C	G	C	C	T	T	T	
Diannan	2	C	T	T	C	C	G	G	C	T	C	C	C	T	A	G	C	A	C	T	T	G	C	G	C	C	T	T	T	
Huzu	2	C	T	T	C	C	X	G	C	T	T	C	C	T	A	G	C	A	C	T	T	G	C	G	C	C	T	T	T	
Jeju nAtive	4	C	T	T	C	C	G	G	C	T	C	C	C	T	A	G	C	A	C	T	T	G	C	G	C	C	T	T	T	
Jinhua	2	X	X	T	C	C	G	G	C	T	X	X	X	X	A	G	C	A	C	T	T	G	C	G	C	C	T	T	X	
KoreAn wild boAr	7	C	T	T	C	C	G	G	C	T	C	C	C	T	A	G	C	A	C	T	T	G	C	G	C	C	T	T	T	
Laiwu	2	C	T	T	C	C	G	G	C	T	T	C	C	T	A	G	C	A	C	T	T	G	C	G	C	C	T	T	T	
Landrace	3	X	X	X	X	X	G	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Lantang	2	C	T	T	C	C	G	G	C	T	C	C	C	T	A	G	C	A	C	T	T	G	C	G	C	C	T	T	T	
Lanyu	4	C	T	X	C	C	X	G	C	T	X	C	C	T	X	X	X	A	C	T	T	X	C	G	X	X	T	X	X	
Large white/Yorkshire	7	X	X	X	X	X	G	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Longlin	2	C	T	T	C	C	G	G	C	T	C	C	C	T	A	G	C	A	C	T	T	G	C	G	C	C	T	T	T	
Luchuan	3	C	T	T	C	C	G	G	C	T	T	C	C	T	A	G	C	A	C	T	T	G	C	G	C	C	T	T	T	
MeishAn piG	5	C	T	T	C	C	G	G	C	T	T	C	C	T	A	G	C	A	C	T	T	G	C	G	C	C	T	T	T	
Min	4	C	T	T	C	C	G	G	C	T	T	C	C	T	A	G	C	A	C	T	T	G	C	G	C	C	T	T	T	
Niejiang	2	C	T	T	C	C	G	G	C	T	T	C	C	T	A	G	C	A	C	T	T	G	C	G	C	C	T	T	T	
NingXiang	2	C	T	T	C	C	G	G	C	T	T	C	C	T	A	G	C	A	C	T	T	G	C	G	C	C	T	T	T	
Penzhou	2	C	T	T	C	C	G	G	C	T	T	C	C	T	A	G	C	A	C	T	T	X	G	C	C	C	T	T	T	
Qingping	3	C	T	T	C	C	G	G	C	T	T	C	C	T	A	G	C	A	C	T	T	G	C	G	C	C	T	T	T	
Rongchang	2	C	T	T	C	C	G	G	C	T	T	C	C	T	A	G	C	A	C	T	T	G	C	G	C	C	T	T	T	
Saba	2	C	T	T	C	C	G	G	C	T	T	C	C	T	A	G	C	A	C	T	T	G	C	G	C	C	T	T	T	
Taoyuan	2	C	T	T	C	C	G	G	C	T	T	C	C	T	A	G	C	A	C	T	T	G	C	G	C	C	T	T	T	
Tibetan	7	C	T	T	C	C	G	G	C	T	T	C	C	T	A	G	C	A	C	T	T	G	C	G	X	C	C	T	T	T
wild boar	14	C	T	X	C	C	X	G	X	T	X	X	C	T	A	X	X	X	X	X	X	X	X	G	X	X	T	X	X	
Wuzhishan	4	C	T	T	C	C	G	G	C	T	T	C	C	T	A	G	C	A	C	T	T	G	C	G	C	C	T	T	T	
Xiang	4	C	T	T	X	X	G	G	C	T	T	C	C	T	A	G	C	A	C	T	T	G	C	G	C	C	T	T	T	
Yimenghei	2	C	T	T	C	C	A	G	C	T	T	C	C	T	A	G	C	A	C	T	T	G	C	G	C	C	T	T	T	
Zang	2	C	T	T	C	C	G	G	C	T	T	C	C	T	A	G	C	A	C	T	T	G	C	G	C	C	T	T	T	

Εικόνα 26 Πολυμορφισμοί των γονιδίων ATP6, ATP8 και COX1

Εκτός από την ύπαρξη κοινού απλοτύπου των τεσσάρων αυτών ρατσών έναντι των υπολοίπων, βρέθηκαν πολυμορφισμοί σε συγκεκριμένες θέσεις γονιδίων οι οποίοι εκ πρώτης όψεως δεν διαχωρίζουν κάποια ράτσα από τις υπόλοιπες, αλλά με περισσότερη μελέτη ίσως μπορέσουν να χρησιμοποιηθούν ως δευτερεύοντα δεδομένα. Στον **πίνακα 7** και στην **εικόνα 27** παρουσιάζονται οι πολυμορφισμοί αυτοί.



Γονίδιο	Ράτσα	Βάση-Θέση-Πολυμορφισμός
ATP6	Yimenghei	G-501-A
COX3	Jeju native	C-222-T
	Koren wild boar	C-222-T
	Jeju native	T-626-C
ND3	NingXiang	C-283-A
ND4	Laiwu	A-228-G
	Min	A-228-G
ND4L	Jeju native	G-219-A
	Korean wild boar	G-219-A
	Jeju native	A-225-G
	Korean wild boar	A-225-G
ND5	Laiwu	T-1392-C
	Min	T-1392-C
ND6	Jeju native	A-403-G
	Laiwu	A-403-G
	Min	A-403-G
	Longlin	A-403-G

Πίνακας 7 Πολυμορφισμοί που βοηθούν στον διαχωρισμό των ρατσών

Στην **εικόνα 27** φαίνεται πως μόνο η ράτσα *Jeju native* διαχωρίζεται από τις υπόλοιπες λόγω του πολυμορφισμού **T-626-C** στο γονίδιο *COX3*. Στις υπόλοιπες θέσεις που εμφανίζεται ένας πολυμορφισμός σε όλα τα άτομα μιας ράτσας (**G-501-A** στο γονίδιο *ATP6* και **C-283-A** στο γονίδιο *ND3*) δεν μπορεί να υπάρξει σίγουρος διαχωρισμός ράτσας διότι εμφανίζεται ο ίδιος πολυμορφισμός και σε μεμονωμένα άτομα διαφόρων άλλων ρατσών.

Έγινε προσπάθεια συνδυασμού των πολυμορφικών θέσεων μεταξύ τους ώστε να βρεθούν διάφοροι απλότυποι οι οποίοι ίσως διαχώριζαν και άλλες ράτσες, τελικά όμως δεν ήταν εφικτή η εύρεση τέτοιων συνδυαστικών θέσεων.



		ATP6	COX3	ND3	ND4	ND4L	ND5	ND6
		501	626 222	283	228	219 225	1392	403
Duroc	4	G	T T	C	A	A G	T	A
European wild boar	6	A/G	T T	C	A	A G	T	A
Iberian	6	G	T T	C	A	A G	T	A
Mangalitsa	3	G	T T	C	A	A G	T	A
Berkshire	3	G	T X	C	A	X X	T	X
Bihu	2	G	T C	C	-	G A	-	v
Dahuabai	2	G	T C	A/C	A	G A	T	A
Diannan	2	G	T C	C	A	G A	T	A
Huzu	2	G/A	T C	C	A	G A	T	A
Jeju nATive	4	G	C T	C	A	A G	T	G
Jinhua	2	G	T C	C	A	G A	T	X
KoreAn wild boAr	7	G	T T	C	A	A G	T	A
Laiwu	2	G	T C	C	G	G A	C	G
Landrace	3	G	T X	C/A	A	X X	T	A
Lantang	2	G	T C	C/A	A	G A	T	A
Lanyu	4	G/A	T X	C	A	G A	T	A
Large white	7	G	T X	C	A	X X	T	X
Longlin	2	G	T C	C	X	G A	X	G
Luchuan	3	G	T C	C/A	A	G A	T	A
MeishAn piG	5	G	T C	C	A	G A	T	X
Min	4	G	T C	C	G	G A	C	G
Niejiang	2	G	T C	C	A	G A	T	A
NingXiang	2	G	T C	A	A	G A	T	A
Penzhou	2	G	T C	C	A	G A	T	A
Qingping	3	G	T C	C	A	G A	T	X
Rongchang	2	G	T C	C	A	G A	T	X
Saba	2	G	T C	C	A	G A	T	A
Taoyuan	2	G	T C	C	A	X A	T	A
Tibetan	7	G	T C	C	A	G A	T	X
wild boar	14	G/A	T X	C/A	A	X A	T	X
Wuzhishan	4	G	T C	C	A	G A	T	A
Xiang	4	G	T C	C/A	A	G A	T	X
Yimenghei	2	A	T C	C	A	G A	T	A
Zang	2	G	T C	C	A	G A	T	A

Εικόνα 27 Πολυμορφισμοί που βοηθούν στον διαχωρισμό των ρατσών



4. Συζήτηση

Ο χοίρος είναι από τα πιο διεσπαρμένα είδη θηλαστικών στον πλανήτη, βρίσκεται σε όλες τις ηπείρους και έχει βρεθεί να ζει και να αναπαράγεται σε μεγάλο εύρος καιρικών συνθηκών. Διαφορετικά γενοτυπικά και φαινοτυπικά χαρακτηριστικά δίνουν το πλεονέκτημα στα άτομα να έχουν την δυνατότητα να ανταπεξέρχονται στις εκάστοτε συνθήκες. Τα διαφορετικά αυτά χαρακτηριστικά έχουν εδραιώσει τον χαρακτηρισμό της ράτσας στα άτομα με κοινό γενότυπο και φαινότυπο με την βοήθεια της εξέλιξης. Πρώτο εξελικτικό βήμα για τον οικόσιτο χοίρο (*Sus scrofa domestica*) ήταν η εξημέρωση του από τον πρόγονό του, τον αγριόχοιρο (*Sus scrofa scrofa*), ο οποίος συνυπάρχει εδώ και χιλιάδες χρόνια με τον οικόσιτο χοίρο διατηρώντας τα άγρια χαρακτηριστικά του. Στη συνέχεια υπό εξελικτική πίεση, λόγω επιλεκτικής αναπαραγωγής από τον άνθρωπο, δημιουργήθηκαν, και δημιουργούνται μέχρι και σήμερα, νέες ράτσες χοίρων οι οποίες χαρακτηρίζονται από γενετικά και μορφολογικά χαρακτηριστικά χρήσιμα για την εδραίωσή τους σε κάποιο περιβάλλον αλλά και για την χρήση τους από τον άνθρωπο. Φυσικό και επόμενο είναι κάθε μελέτη με αντικείμενο τον χοίρο να έχει μεγάλη σημασία για τον άνθρωπο λόγω του ρόλου που κατέχει η εκτροφή και το εμπόριό του στην παγκόσμια οικονομία.

Εδώ και χρόνια έχουν αρχίσει να πραγματοποιούνται μελέτες με αντικείμενο τόσο τον οικόσιτο όσο και τον άγριο χοίρο, με σκοπό την ολοκληρωμένη φαινοτυπική, γενετική, εξελικτική αλλά και φυλλογενετική γνώση. Μερικές από τις πιο σπουδαίες έρευνες που έχουν πραγματοποιηθεί σε αυτόν τον τομέα είναι η **αλληλούχηση του γονιδιώματος του χοίρου**, τόσο του πυρηνικού όσο και του μιτοχονδριακού, οι **εξελικτικές σχέσεις βάσει εξημέρωσης** του οικόσιτου χοίρου από τον πρόγονό του τον αγριόχοιρο, η **διαειδική φυλογένεση** η οποία μελετά τις σχέσεις του είδους (συμπεριλαμβανομένων τόσο των οικόσιτων όσο και των άγριων χοίρων) με άλλα πλησιέστερα ή και πιο απομακρυσμένα είδη αλλά και η **ενδοειδική φυλογένεση** η οποία μελετά τις σχέσεις των ρατσών μεταξύ τους ανάλογα με τον τόπο προέλευσής τους.

Το **1998** οι **Bjorn M. Ursing** και **Ulfur Arnason** δημοσίευσαν για πρώτη φορά ολόκληρη την αλληλουχία του μιτοχονδριακού γονιδιώματος του χοίρου καθώς και τις ανάλογες θέσεις στις οποίες βρίσκονται τα γονίδια και οι μη κωδικές περιοχές του.

Το **2005** οι **Larson et al** μελέτησαν χοίρους του Ευρασιατικού κλάδου και κατέληξαν στο συμπέρασμα ότι η εξημέρωση του οικόσιτου χοίρου από τον πρόγονό του πραγματοποιήθηκε την ίδια χρονική περίοδο, γεωγραφικά ανεξάρτητα όμως στην Άπω Ανατολή, στην Νοτιοανατολική Ασία και στην Ευρώπη.

Το **2012** οι **Groenen et al** δημοσίευσαν μια φυλογενετική μελέτη με αντικείμενο μελέτης Ευρωπαϊκούς και Ασιατικούς χοίρους (οικόσιτους και άγριους)



όπου έφτασαν στο συμπέρασμα ότι οι αγριόχοιροι εμφανίστηκαν πριν 1 εκατομμύριο χρόνια στην Ασία και εξαπλώθηκαν στην συνέχεια στην Ευρώπη.

Το 2013 οι *Guanghui Yu et al* πραγματοποίησαν φυλλογενετική μελέτη στην οποία επικεντρώθηκαν στις κωδικές αλληλουχίες (CDS) και στη ρυθμιστική περιοχή D-loop του μιτοχondριακού γονιδιώματος Ευρωπαϊκών και Ασιατικών χοίρων. Αποτέλεσμα της έρευνας αυτής ήταν ο διαχωρισμός των διαφόρων ρατσών σε Ευρωπαϊκό και Ασιατικό κλάδο ανάλογα με την αλληλουχία της *D-loop* ή το γονίδιο που μελετούσαν κάθε φορά.

Δεν είναι σημαντική μόνο η ομαδοποίηση ρατσών σε γεωγραφικούς κλάδους αλλά και η ικανότητα διαχωρισμού των διαφόρων ρατσών μεταξύ τους βάσει γενοτυπικών χαρακτηριστικών. **Μέχρι σήμερα η ταυτοποίηση και η κατηγοριοποίηση των χοίρων σε ράτσες γίνεται βασιζόμενη σε μορφολογικά χαρακτηριστικά** τα οποία τις περισσότερες φορές είναι δυσδιάκριτα και οδηγούν σε εσφαλμένα αποτελέσματα. Πρόκληση αποτελεί για την επιστημονική κοινότητα η εύρεση μοριακών δεικτών οι οποίοι θα μπορέσουν να αντικαταστήσουν την μέθοδο αναγνώρισης ρατσών βάσει μορφολογικών χαρακτηριστικών, δίνοντας την δυνατότητα ταυτοποίησης και κατηγοριοποίησης ατόμων του είδους σε κάποια ράτσα με μεγαλύτερη ακρίβεια και αξιοπιστία.

Η παρούσα μελέτη αποσκοπεί στον εντοπισμό τέτοιων γενοτυπικών χαρακτηριστικών. Πιο συγκεκριμένα η μελέτη πραγματοποιήθηκε *in silico*, δηλαδή με την βοήθεια του ηλεκτρονικού υπολογιστή, άτομα χοίρων με απώτερο σκοπό την εύρεση μιτοχondριακών μοριακών δεικτών οι οποίοι είναι κοινοί στα άτομα μιας ράτσας αλλά έχουν την ικανότητα να τα διαχωρίζουν από τις υπόλοιπες. Πιο συγκεκριμένα ως δεδομένα συλλέχθηκαν και χρησιμοποιήθηκαν 167 αλληλουχίες ατόμων, από την τράπεζα αλληλουχιών *GenBank* του *NCBI*, οι οποίες κατατάσσονταν σε 77 ράτσες όπως φαίνεται στον **πίνακα 3**. Με την βοήθεια της βιοπληροφορικής και διαφόρων εργαλείων της πραγματοποιήθηκε κατηγοριοποίηση, στοίχιση και ανάλυση των αλληλουχιών αυτών. Η μελέτη επικεντρώθηκε στα γονίδια της αναπνευστικής αλυσίδας του μιτοχondρίου (*COXI-3*, *NDI-6*, *ATP6* και *ATP8* και *CYTB*), των ριβοσωμικών rRNA (*12s rDNA* και *16s rDNA*) καθώς και των tRNAs ενώ απορρίφθηκε η D-loop λόγω δυσαναλογίας στο μέγεθος σε κάθε άτομο και άρα δύσκολη διαχείρισή της. Καταγράφηκε το πλήθος των πολυμορφικών θέσεων για κάθε γονίδιο ξεχωριστά όπως φαίνεται στον **πίνακα 4** και συνολικά για κάθε σύμπλοκο της αναπνευστικής αλυσίδας όπως φαίνεται στον **πίνακα 5**. Στην περειαίρω ανάλυση των θέσεων αυτών επιλέχθηκαν και αξιοποιήθηκαν οι θέσεις αυτές στις οποίες εμφανίζεται ένας πολυμορφισμός σε όλα τα άτομα μιας ή και περισσότερων ρατσών, το πλήθος των θέσεων αυτών εμφανίζεται στον **πίνακα 6**. Ο σημαντικός περιορισμός των πολυμορφικών θέσεων έδωσε την δυνατότητα ευκολότερης μελέτης τους και εύρεση πολυμορφισμών οι οποίοι μπορούν να βοηθήσουν στον διαχωρισμό των ρατσών. Πιο συγκεκριμένα τα αποτελέσματα της



παρούσας έρευνας είναι τα εξής και φαίνονται στο **επιπρόσθετο αρχείο** καθώς και στον **πίνακα 7** και **εικόνα 27**:

- Σε όλα τα γονίδια, για τις πολυμορφικές θέσεις που φαίνονται στο επιπρόσθετο αρχείο, οι ράτσες **Duroc**, **European wild boar**, **Iberian** και **Mangalica** βρέθηκαν να παρουσιάζουν κοινό μοτίβο πολυμορφισμών μεταξύ τους, δημιουργώντας έναν **απλότυπο** αλλά διαφορετικό σε σχέση με τις υπόλοιπες ράτσες.
- Η ράτσα **Jeju native** φαίνεται να διαχωρίζεται από τις υπόλοιπες ράτσες λόγω παρουσίας μονονουκλεοτιδικού πολυμορφισμού (**SNP**) **T626C** στο γονίδιο **COX3**. Όλα τα άτομα της ράτσας αυτής εμφανίζουν C στην θέση αυτή έναντι του T που παρουσιάζουν όλα τα άτομα των υπολοίπων ρατσών.
- Η ράτσα **Yimenghei** παρουσιάζει μονονουκλεοτιδικό πολυμορφισμό (**SNP**) **G501A** στο γονίδιο **ATP6**. Το **SNP** αυτό δεν μπορεί να διαχωρίσει πλήρως την ράτσα αυτή από τις υπόλοιπες διότι ο πολυμορφισμός αυτός εμφανίζεται και σε μεμονωμένα άτομα άλλων ρατσών.
- Η ράτσα **NingXiang** παρουσιάζει μονονουκλεοτιδικό πολυμορφισμό **SNP C283A** στο γονίδιο **ND3**.
- Οι ράτσες **Laiwu** και **Min** παρουσιάζουν κοινούς πολυμορφισμούς **A228G** στο γονίδιο **ND4** και **T1392C** στο γονίδιο **ND5**.
- Οι ράτσες **Korean wild boar** και **Jeju native** παρουσιάζουν κοινούς πολυμορφισμούς **C222T** στο γονίδιο **COX3** και **G219A, A225G** στο γονίδιο **ND4L**.
- Οι ράτσες **Jeju native, Longlin, Min** και **Laiwu** παρουσιάζουν κοινό πολυμορφισμό **A403G** στο γονίδιο **ND6**.

Εκτός από τον πολυμορφισμό **T626C** στο γονίδιο **COX3** που μπορεί να διαχωρίσει πλήρως την ράτσα **Jeju native** από τις υπόλοιπες, οι υπόλοιποι πολυμορφισμοί δύνανται να γίνουν χρήσιμο εργαλείο για περαιτέρω ανάλυση του μιτοχονδριακού γονιδιώματος του χοίρου και εύρεση περισσότερων μοριακών δεικτών οι οποίοι θα έχουν την ικανότητα να διαχωρίσουν περισσότερες ράτσες.

Η παρούσα μελέτη πραγματοποιήθηκε κατέχοντας μόνο έναν μικρό αριθμό ατόμων που αντιπροσωπεύουν κάθε ράτσα, γι' αυτόν ακριβώς τον λόγο δεν μπορούμε να κατοχυρώσουμε με μεγάλη ακρίβεια πως τα αποτελέσματα είναι φερέγγυα. Μπορούν όμως να θεωρηθούν όμως ως απαρχή μιας έρευνας η οποία έχει ως απώτερο σκοπό την εύρεση αξιόπιστων μοριακών δεικτών οι οποίοι θα πληρούν τις προδιαγραφές για τον ορθό διαχωρισμό ρατσών.



Ως μελλοντικός στόχος τίθενται τα εξής ζητήματα:

- Συλλογή περισσότερων αλληλουχιών από μιτοχονδριακό γονιδίωμα ατόμων της κάθε ράτσας και μελέτη των προαναφερθέντων γονιδίων με σκοπό την πιο ολοκληρωμένη και ορθή άποψη για την αξιοπιστία των υπαρχόντων μοριακών δεικτών.
- Στοχευμένη μελέτη των πολυμορφικών θέσεων που εντοπίστηκαν σε μεγαλύτερο αριθμό ατόμων μέσω τεχνικών βασισμένων στην PCR.
- Μελέτη του γενωμικού *DNA* του χοίρου ώστε να εντοπιστούν περισσότεροι δείκτες οι οποίοι θα είναι ικανοί να διαχωρίσουν περισσότερες ράτσες χοίρων οι οποίες δεν ήταν εφικτό να διαχωριστούν βάσει των μιτοχονδριακών αλληλουχιών.

Συμπερασματικά ο εντοπισμός και καταχώρηση μοριακών δεικτών από την επιστημονική κοινότητα, ενδέχεται να βοηθήσει στην γρήγορη, οικονομική και αξιόπιστη κατηγοριοποίηση ενός ατόμου χοίρου σε μια συγκεκριμένη ράτσα με την χρήση μόνο ενός μικρού δείγματος από τον ζώντα οργανισμό.



5. Βιβλιογραφικές αναφορές

- ❖ Εικόνα 1: Mammalian Anatomy & Physiology Pig Dissection and Histology, www.studyblue.com
- ❖ Εικόνα 2: Animal Diversity Web , www.animaldiversity.org
- ❖ Εικόνα 6: Ursing,U Arnason (1998) The complete mitochondrial DNA sequence of the pig (*Sus scrofa*), *J Mol Evol.* 1998 Sep;47(3):302-6.
- ❖ Εικόνα 7, Πίνακας 1 : GenBank and WGS Statistics, NCBI, www.ncbi.nlm.nih.gov/genbank/statistics/
- ❖ Εικόνα 8, Εικόνα 9: *Sus scrofa* breed Lanyu mitochondrion, complete genome, www.ncbi.nlm.nih.gov/nucore/DQ518915.2
- ❖ Galtier N, Nabholz B, Glémin S, Hurst GD (2009), Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol Ecol.* 2009 Nov;18(22):4541-50
- ❖ Guanghui Yu, Hai Xiang, Jikun Wang and Xingbo Zhao (2013), The phylogenetic status of typical Chinese native pigs: analyzed by Asian and European pig mitochondrial genome sequences . *J Anim Sci Biotechnol.* 2013 Mar 8;4(1):9.
- ❖ Kijas JM, Andersson L. (2001) , A phylogenetic study of the origin of the domestic pig estimated from the near-complete mtDNA genome. *J Mol Evol.* 2001 Mar;52(3):302-8.
- ❖ Trut L, Oskina I, Kharlamova A. (2009), Animal evolution during domestication: the domesticated fox as a model. *Bioessays.* 2009 Mar;31(3):349-60
- ❖ Dominic Wright. (2015), The Genetic Architecture of Domestication in Animals. *Bioinform Biol Insights.* 2015; 9(Suppl 4): 11–20
- ❖ Nancy G, Robert A Parker (2002), Meta-analysis: Neither quick nor easy. *BMC Medical Research Methodology* 2002 2:10
- ❖ Wu GS *et al* (2007), Population phylogenomic analysis of mitochondrial DNA in wild boars and domestic pigs revealed multiple domestication events in East Asia. *Genome Biol.* 2007;8(11)
- ❖ Grossi SF, Lui JF, Garcia JE, Meirelles FV. Genetic diversity in wild (*Sus scrofa scrofa*) and domestic (*Sus scrofa domestica*) pigs and their hybrids based on polymorphism of a fragment of the D-loop region in the mitochondrial DNA. *Genet Mol Res.* 2006 Oct 31;5(4):564-8.
- ❖ Paulo C. Alves. (2010). Genetic diversity of wild boar populations and domestic pig breeds (*Sus scrofa*) in South-western Europe. *Biological Journal of the Linnean Society*, 2010, 101, 797–822.
- ❖ Chen K *et al*, (2007) Genetic resources, genome mapping and evolutionary genomics of the pig (*Sus scrofa*). *Int J Biol Sci.* 2007 Feb 10;3(3):153-65.



- ❖ Giuffra E, Kijas JM, Amarger V, Carlborg O, Jeon JT, Andersson L. (2000) The origin of the domestic pig: Independent domestication and subsequent introgression. *Genetics*. 2000; 154(4):1785-1791.
- ❖ Larson G, Dobney K, Albarella U, *et al.* (2005) Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science*. 2005; 307(5715):1618-1621.
- ❖ Reed CA. The pattern of animal domestication in the prehistoric Near East. In: Ucko PJ, Dimbleby GW, eds. *The Domestication and Exploitation of Plants and Animals*. London, UK: Duckworth, 1969:361-380.
- ❖ Darwin C. *On the Origins of the Species by Means of Natural Selection, Or the Preservation of Favoured Races in the Struggle for Life*. UK: John Murray, 1859.
- ❖ Shokolenko I. *et al.* (2009). Oxidative stress induces degradation of mitochondrial DNA. *Nucleic Acids Res*. 2009 May;37(8):2539-48
- ❖ Groenen MA *et al.* (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*. 2012 Nov 15;491(7424):393-8
- ❖ Abbott RJ, Ritchie MG, Hollingsworth PM. (2008). Introduction. *Speciation in plants and animals: pattern and process*. *Philos Trans R Soc Lond B Biol Sci*. 2008 Sep 27;363(1506):2965-9
- ❖ Louis Ollivier (2005). An assessment of European pig diversity using molecular markers: Partitioning of diversity among breeds. *Conservation Genetics* (2005) 6:729–741
- ❖ Greger Larson, Joachim Burger. (2013). A population genetics view of animal domestication. *Trends Genet*. 2013 Apr;29(4):197-205
- ❖ Ursing, U Arnason (1998) The complete mitochondrial DNA sequence of the pig (*Sus scrofa*), *J Mol Evol*. 1998 Sep;47(3):302-6.
- ❖ Massimo Scandura *et al.* (2011). Genetic diversity in the European wild boar *Sus scrofa*: phylogeography, population structure and wild x domestic hybridization. *Mammal Rev*. 2011, Volume 41, No. 2, 125–137.
- ❖ T. Cavalier-Smith (2002) The phagotrophic origin of eukaryotes and phylogenetic classification of Protozoa. *International Journal of Systematic and Evolutionary Microbiology* (2002),52, 297–354
- ❖ Margulis L(1981). *Symbiosis in cell evolution*. New York: W.H. Freeman, (1981) 1993. 452 p
- ❖ T. Uzzell, C. Spolsky. (1981). *Origins and Evolution of Eukaryotic Intracellular Organelles* Pages 481–499
- ❖ Hurst GD, Jiggins FM. (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proc Biol Sci*. 2005 Aug 7;272(1572):1525-34
- ❖ Pleines *et al.*(2009) Application of non- coding DNA regions in intraspecific analyses. *Plant Systematics and Evolution* 282 : 281 – 294 .



- ❖ Sun N, et al. (2006) Bayesian error analysis model for reconstructing transcriptional regulatory networks. *Proc Natl Acad Sci U S A* 103(21):7988-93
- ❖ Pevsner J. (2009). *Bioinformatics and functional genomics*, 2nd publication, John Wiley and Sons (951) p. 211-227.
- ❖ Felsenstein J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* 17: 368–376.
- ❖ Stephen F. Altschul *et al.* (1990) Basic local alignment search tool. *Journal of Molecular Biology* Volume 215, Issue 3, 5 October 1990, Pages 403–410