

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΓΕΩΠΟΝΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΓΕΩΠΟΝΙΑΣ ΙΧΘΥΟΛΟΓΙΑΣ
ΚΑΙ ΥΔΑΤΙΝΟΥ ΠΕΡΙΒΑΛΛΟΝΤΟΣ**

ΠΡΟΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**«Βιοπληροφορική ανάλυση της αναγεννητικής ικανότητας των
κεφαλόποδων μαλακίων»**

Μπουραντάς Νικόλαος

02859

ΒΟΛΟΣ 2025

**«Βιοπληροφορική ανάλυση της αναγεννητικής ικανότητας των κεφαλόποδων
μαλακίων»**

Τριμελής Εξεταστική Επιτροπή:

1. **Γεώργιος Γκάφας**, Αναπληρωτής Καθηγητής, Μοριακή Βιολογία της Διατήρησης Θαλάσσιων Θηλαστικών και Ιχθυοαποθεμάτων, Τμήμα Γεωπονίας Ιχθυολογίας και Υδάτινου Περιβάλλοντος, Σχολή Γεωπονικών Επιστημών, Πανεπιστήμιο Θεσσαλίας, **Επιβλέπων**.
2. **Εξαδάκτυλος Αθανάσιος**, Καθηγητής, Γενετική Υδρόβιων Ζωϊκών Οργανισμών, Τμήμα Γεωπονίας Ιχθυολογίας και Υδάτινου Περιβάλλοντος, Σχολή Γεωπονικών Επιστημών, Πανεπιστήμιο Θεσσαλίας, **Μέλος**.
3. **Βαφείδης Δημήτριος**, Καθηγητής, Βιοποικιλότητα των Θαλάσσιων Βενθικών Ασπονδύλων και άμεση – έμμεση χρηστικότητα τους, Τμήμα Γεωπονίας Ιχθυολογίας και Υδάτινου Περιβάλλοντος, Σχολή Γεωπονικών Επιστημών, Πανεπιστήμιο Θεσσαλίας, **Μέλος**.

ΠΕΡΙΛΗΨΗ

Η εξέλιξη της αναγεννητικής ικανότητας σε πολυκύτταρα ζώα αποτελεί ένα από τα πιο σύνθετα και ενδιαφέροντα προβλήματα στη βιολογία. Η αναγέννηση, μια διαδικασία που συνίσταται στην αναγέννηση κατεστραμμένων δομών και την λειτουργική τους αποκατάσταση, είναι ευρέως διαδεδομένη σε πολλά φύλα του ζωικού βασιλείου, από τα κατώτερα ασπόνδυλα έως τα θηλαστικά. Μεταξύ των ειδών που είναι ικανά για αναγέννηση, η πραγματική ικανότητα αποκατάστασης της πλήρους μορφής και λειτουργίας του τραυματισμένου ιστού ποικίλλει σημαντικά, από το να είναι σε θέση να υποβληθούν σε αναγέννηση ολόκληρου του σώματος και των εσωτερικών οργάνων, έως περιπτώσεις όπου αυτή η ικανότητα περιορίζεται σε λίγους ιστούς. Μεταξύ των ασπόνδυλων, τα κεφαλόποδα μαλάκια διατηρούν την ικανότητα να αναγεννούν διάφορες δομές (δηλαδή, μύες, νεύρα ή ολόκληρα τμήματα). Η παρούσα έρευνα εστιάζει κυρίως στα SNPs (μονονουκλεοτιδικοί πολυμορφισμοί), των κεφαλόποδων μαλακίων. Μελετήθηκαν 10 είδη εκ των οποίων τα 5 αποδεδειγμένα διαθέτουν ικανότητες αναγέννησης και τα άλλα 5 είναι υπό διερεύνηση, για το αν διαθέτουν την συγκεκριμένη ικανότητα. Πραγματοποιήθηκε βιοπληροφορική ανάλυση με στόχο να παρατηρήσουμε διαφορές και ομοιότητες μεταξύ των ειδών σε σχέση με το επιλεγόμενο γονιδίωμα αναφοράς. Οι αναλύσεις πραγματοποιήθηκαν σε λειτουργικό σύστημα Linux (Ubuntu 20.04.6 LTS, 64 -bit), με διάφορα ειδικά προγράμματα γενετικής ανάλυσης. Χρησιμοποιήθηκαν αλληλουχίες ολόκληρου του γονιδιώματος (Whole genome sequencing – WGS), διπλού διαβάσματος (paired) αλληλουχίες από γονιδιώματα από βάσεις δεδομένων όπως του NCBI (National Center for Biotechnology Information) και το ENA (European Nucleotide Archive) του Ευρωπαϊκού Ινστιτούτου Βιοπληροφορικής (EMBL-EBI), λήφθηκαν τα αρχεία από

πειράματα (SRA) σε μορφή FASTQ. Στη συνέχεια έγινε ευθυγράμμιση των αλληλουχιών, φιλτράρισμα του VCF αρχείου και GLM ανάλυση για τον προσδιορισμό των SNPs. Βρέθηκαν 33 μη συνώνυμα SNPs και σε αυτά επικεντρώθηκε περισσότερο η παρούσα μελέτη. Στα αποτελέσματα παρατηρήθηκε ο εμφανής διαχωρισμός των δύο ομάδων, όμως τίθεται ζήτημα πιο εντατικής έρευνας, για καλύτερα αποτελέσματα.

Περιεχόμενα

ΠΕΡΙΛΗΨΗ

1. ΕΙΣΑΓΩΓΗ

1.1 Αναγέννηση

1.2 Κεφαλόποδα

1.3 Κατηγορίες κεφαλόποδων

1.3.1 Χταπόδια

1.3.2 Σουπιές

1.3.3 Καλαμάρια

1.3.4 Ναυτίλοι

1.4 Μονονουκλεοτιδικοί Πολυμορφισμοί (SNPs)

1.5 Συγκλίνουσα εξέλιξη – Φυσική επιλογή

1.6 Σκοπός της εργασίας

2. ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ

2.2. Ευθυγράμμιση αλληλουχιών (Alignment)

2.3. Bcftools

2.4. Επεξεργασία VCF

2.5. Glm ανάλυση (General linear model)

2.6. Πρόβλεψη γενετικών παραλλαγών SnpEff

3. ΑΠΟΤΕΛΕΣΜΑΤΑ

4. ΣΥΖΗΤΗΣΗ

Βιβλιογραφία

Ελληνική,

Ξενόγλωσση

Ηλεκτρονική **Error! Bookmark not defined.**

ABSTRACT

ΠΑΡΑΡΤΗΜΑ

1.ΕΙΣΑΓΩΓΗ

1.1 Αναγέννηση

Η αναγέννηση, ένα από τα πιο συναρπαστικά ερωτήματα στην εξελικτική βιολογία, καθώς τίθενται ερωτήματα σχετικά με το πως ορισμένα ζώα μπορούν να αναγεννήσουν τραυματισμένες δομές ενώ άλλα όχι. Η αναγέννηση των αποφύσεων φαίνεται να είναι κοινή όταν εξετάζεται σε όλη τη φυλογένεση των μεταζώων, ωστόσο αυτή η ικανότητα έχει χαθεί σε πολλά taxa σε ποικίλους βαθμούς. Εντός των ειδών, η ικανότητα αναγέννησης μπορεί επίσης να ποικίλλει οντογενετικά μεταξύ των ατόμων. Η αναγέννηση των αποφύσεων κατά μήκος του δευτερεύοντος άξονα του σώματος μπορεί να περιορίζεται από θεμελιώδη χαρακτηριστικά όπως το μέγεθος του σώματος, η γήρανση, το στάδιο ζωής και το πρότυπο ανάπτυξης ([Seifert et al., 2012](#)). Σε αντίθεση με τα σπονδυλωτά, τα ασπόνδυλα μπορούν να επιδείξουν μια τεράστια ικανότητα αναγέννησης ολόκληρου του ζώου, η οποία είναι εξαιρετικά μεταβλητή ακόμη και μεταξύ μελών του ίδιου φύλου ([Echeverri & Zayas, 2018](#)).

1.2 Κεφαλόποδα

Τα κεφαλόποδα ανήκουν στην κατηγορία των μαλακίων, συμπεριλαμβανομένων των χταποδιών, των καλαμαριών, των σουπιών και των ναυτίλων. Υπάρχουν πάνω από 800 είδη που βρίσκονται στους ωκεανούς σε όλο τον κόσμο, που κυμαίνονται από ρηχά τροπικά νερά έως βαθιά νερά σε βάθος άνω των 5.000 μέτρων ([Hanlon και Messenger, 2018](#)). Το μέγεθος του σώματός τους μπορεί να κυμαίνεται από μόλις 1 cm έως πάνω από 18 m σε συνολικό μήκος και η αναλογία εγκεφάλου προς μάζα σώματος μπορεί να είναι υψηλότερη από αυτή ορισμένων σπονδυλωτών ([Packard, 1972](#)). Μπορούν να αλλάξουν γρήγορα το μοτίβο και το σχήμα του σώματός τους για να αποφύγουν την θήρευση και για επικοινωνία μεταξύ και εντός

των ειδών (Nakajima et al., 2018). Οι αναγεννητικές ικανότητες των κεφαλόποδων μαλακίων έχουν μελετηθεί σε διάφορες έρευνες και είναι γεγονός (Shaw et al., 2016). Ερωτήματα σχετικά με το πώς και γιατί συμβαίνει η αναγέννηση των ιστών έχουν τραβήξει την προσοχή αμέτρητων επιστημόνων, που συνεχίζουν πιο εντατικές έρευνες. Η αναγεννητική ικανότητα διαφέρει σημαντικά μεταξύ των οργάνων και των οργανισμών και έχει μελετηθεί μια σειρά από μοντέλα συστημάτων που χρησιμοποιούν διαφορετικές αναγεννητικές στρατηγικές και προσφέρουν διαφορετικά τεχνικά πλεονεκτήματα για την κατανόηση της αναγέννησης (Poss, 2010).

1.3 Κατηγορίες κεφαλόποδων

1.3.1 Χταπόδια

Τα χταπόδια εμφανίζονται σχεδόν σε όλα τα βενθικά θαλάσσια ενδιαιτήματα (Voight, 1993). Όλα τα είδη της τάξης Octopoda παρουσιάζουν ισχυρή ομοιότητα στη δομική μορφολογία και μεγάλη ποικιλομορφία στον χρωματισμό και τα μοτίβα του δέρματος, στη συμπεριφορά και στις στρατηγικές επιβίωσης, γεγονός που έχει παρεμποδίσει τη μελέτη των φυλογενετικών σχέσεων. (Acosta-Jofré et al., 2012). Τα χταπόδια χρησιμοποιούν τα πλοκάμια τους για ένα ευρύ φάσμα λειτουργιών, όπως η μετακίνηση, το κυνήγι, η άμυνα, η εξερεύνηση και η αναπαραγωγή. Ωστόσο, το φυσικό περιβάλλον ενέχει πολυάριθμες απειλές για την ακεραιότητα των πλοκαμιών τους, συμπεριλαμβανομένων των θηρευτών και των θηραμάτων κατά τη σύλληψη. Εντυπωσιακά, τα χταπόδια είναι σε θέση να κλείνουν ανοιχτές πληγές σε ένα υδάτινο περιβάλλον και να αναγεννούν πλήρως τα άκρα τους. Η φάση αναγέννησης των βραχιόνων των χταποδιών έχει περιγραφεί εκτενώς. Ωστόσο, υπάρχουν λίγες πληροφορίες σχετικά με την οξεία τοπική απόκριση που εμφανίζεται μετά από έναν

ακρωτηριασμό, συγκρίσιμη με αυτήν που συμβαίνει συχνά στην άγρια φύση.([Shaw et al., 2016](#)).

1.3.2 Σουπιές

Οι σουπιές βρίσκονται σε εύκρατα έως τροπικά νερά πάνω από ηπειρωτικές υφαλοκρηπίδες και πλαγιές, από τον δυτικό Ειρηνικό, τον Ινδικό Ωκεανό και τον ανατολικό Ατλαντικό, μέχρι βόρεια, τη Βόρεια Θάλασσα. Η ποικιλομορφία των σουπιών μπορεί να εξηγηθεί εν μέρει από τη βιολογία και τη βιογεωγραφία τους ([Allcock et al. 2014](#)).(Lupše et al., 2023). Όσον αφορά την αναγέννηση, έχουν πραγματοποιηθεί μελέτες της αναγέννησης των νεύρων και των ιστών σε σουπιές ([Sereni and Young, 1932](#))([Nakajima et al., 2018](#)) και έχει παρατηρηθεί αναγέννηση των άκρων έπειτα από ακρωτηριασμό. Η αποκατάσταση της λειτουργίας συμβαίνει λίγες ημέρες αργότερα, πολύ πριν από την πλήρη αναγέννηση του βραχίονα, η οποία αναφέρεται ότι απαιτεί περίπου 40 ημέρες ([Pamela Imperadore & Graziano Fiorito, 2018](#)).

1.3.3 Καλαμάρια

Τα καλαμάρια είναι κεφαλόποδα μαλάκια που χαρακτηρίζονται από δέκα περιφερειακά εξαρτήματα (ή βραχίονες), με το τέταρτο ζεύγος να έχει τροποποιηθεί σε πλοκάμια. Καταλαμβάνουν παράκτια και ωκεάνια νερά από την Αρκτική έως την Ανταρκτική. Η ποικιλομορφία στο βίοτοπο και τον τρόπο ζωής αντικατοπτρίζεται στη μορφολογία τους ([Lindgren, 2010](#)). Σχετικά με την αναγέννηση, η πλειονότητα των αναφορών μπορεί να μην σχετίζονται με τις αναγεννητικές ικανότητες των καλαμαριών, ωστόσο έχουν παρατηρηθεί αναγεννημένες δομές σε ορισμένα είδη καλαμαριών, ([Pamela Imperadore & Graziano Fiorito, 2018](#)).

1.3.4 Ναυτίλοι

Ο Ναυτίλος είναι το μοναδικό σωζόμενο κεφαλόποδο με εξωτερικό κέλυφος από την Παλαιοζωική Εποχή. Είναι μοναδικό στη γενεαλογία των κεφαλόποδων και κρίσιμο για την κατανόηση των εξελικτικών καινοτομιών των κεφαλόποδων ([Zhang et al., 2021](#)). Μετά την εκκόλαψη του, το κέλυφός του μεγαλώνει περίπου 139 mm για να φτάσει σε πλήρη ανάπτυξη και σχηματίζονται περίπου 28 διαφράγματα. Χρειάζεται να περάσουν 8 χρόνια περίπου για να φτάσει στην πλήρη ωριμότητα ([Westermann et al., 2004](#)). Δεν υπάρχουν αναφορές για αναγέννηση των άκρων στους Ναυτίλους, όμως έχει αποδειχθεί ότι οι Ναυτίλοι μπορούν να αναγεννήσουν κάποιο κατεστραμμένο τμήμα του κελύφους τους. ([Tsujino & Shigeta, 2012](#); [Y & Y, 2012](#)).

1.4 Μονονουκλεοτιδικοί Πολυμορφισμοί (SNPs)

Οι μονονουκλεοτιδικοί πολυμορφισμοί (SNPs) είναι υπεύθυνοι για την ποικιλομορφία των γονιδιωμάτων και για τις φαινοτυπικές διαφορές. ([Brookes, 1999](#)). Στην ουσία, ο μονονουκλεοτιδικός πολυμορφισμός (SNP) είναι η απλούστερη μορφή παραλλαγής του DNA μεταξύ των ατόμων. Ένας μονονουκλεοτιδικός πολυμορφισμός (SNP), είναι μια παραλλαγή σε ένα μόνο νουκλεοτίδιο που εμφανίζεται σε μια συγκεκριμένη θέση στο γονιδίωμα. Τα συνώνυμα SNPs δεν αλλάζουν την αλληλουχία αμινοξέων, ενώ τα μη συνώνυμα SNPs την αλλάζουν ([Shastry, 2007](#)). Επίσης μπορεί να επηρεάσουν τη δραστηριότητα του υποκινητή (γονιδιακή έκφραση), τη διαμόρφωση (σταθερότητα) του αγγελιοφόρου RNA (mRNA) και τον υποκυτταρικό εντοπισμό των mRNAs ή/και των πρωτεϊνών ([Shastry, 2009](#)).

1.5 Συγκλίνουσα εξέλιξη – Φυσική επιλογή

Η συγκλίνουσα εξέλιξη συχνά προκύπτει από παρόμοιες γενετικές αλλαγές, σε κατηγορίες οργανισμών που εμφανίζουν παρόμοια ή πανομοιότυπα χαρακτηριστικά σε ανεξάρτητες γενεαλογικές γραμμές, ([Stern, 2013](#)). Μερικές φορές, η συγκλίνουσα

εξέλιξη προκύπτει από την εξέλιξη μιας γενετικής αλλαγής που κληρονομείται από έναν προγονικό πληθυσμό ή από υβριδισμό μεταξύ ειδών([stern,2013](#)). Επιπλέον, η συγκλίνουσα εξέλιξη συχνά θεωρείται ως ένα αναπόφευκτο αποτέλεσμα της εξέλιξης μέσω φυσικής επιλογής([Stayton, 2015](#)).

Ένας καλά αναγνωρισμένος μηχανισμός με τον οποίο η εξέλιξη μπορεί να προχωρήσει για να παράγει βιολογικές καινοτομίες και να ενισχύσει την επιβίωση των ειδών είναι μέσω της φυσικής επιλογής. ([van Wijnen & Lewallen, 2024](#)). Το γεγονός ότι οι οργανισμοί είναι καλά προσαρμοσμένοι στο περιβάλλον τους αποτελεί ακρογωνιαίο λίθο της εξελικτικής βιολογίας([Vermeij, 2023](#)).. Από την εποχή του Δαρβίνου, ο ρόλος της φυσικής επιλογής στη διαμόρφωση των μορφολογικών και φυσιολογικών χαρακτηριστικών των ειδών και της προσαρμογής τους στο φυσικό περιβάλλον σε όλες τις γενιές, έχει κεντρική θέση στην κατανόηση της ζωής και της ποικιλομορφίας της([Lenski, 2017](#)).

Η αναγέννηση, δηλαδή η αντικατάσταση των χαμένων τμημάτων του σώματος, είναι ευρέως διαδεδομένη αλλά εξαιρετικά μεταβλητή μεταξύ των ζώων([Bely & Nyberg, 2010](#)). Μεταξύ των ασπόνδυλων, τα κεφαλόποδα μαλάκια διατηρούν την ικανότητα να αναγεννούν διάφορες δομές (δηλαδή, μύες, νεύρα ή ολόκληρα τμήματα) ([P. Imperadore & G. Fiorito, 2018](#)). Η ικανότητα αυτή αναπτύχθηκε στα κεφαλόποδα, καθώς οι δομές που αναγεννούν είναι χρήσιμες για την επιβίωσή τους, όπως συμβαίνει και σε κάθε ζώο που φέρει αυτή την ικανότητα ([Reichman, 1984](#)). Παρά τη μακρά ιστορία του θέματος, η εξέλιξη της αναγεννητικής ικανότητας στα ζώα απέχει πολύ από το να είναι πλήρως κατανοητή ([Bely και Nyberg, 2010](#))([Elchaninov et al., 2021](#)).

1.6 Σκοπός της εργασίας

Η αναγέννηση των ζώων αποτελεί αντικείμενο συνεχούς επιστημονικού ενδιαφέροντος. Οι πρώτες πειραματικές μελέτες σχετικά με την αναγέννηση πραγματοποιήθηκαν τον 18ο αιώνα ([Reaumur, 1712](#); [Tremblay, 1744](#)) ([Elchaninov et al., 2021](#)) και παρόλα αυτά, εξακολουθεί να παραμένει ένα θέμα το οποίο δεν έχει γίνει πλήρως κατανοητό ([Bely & Nyberg, 2010](#)). Η αναγέννηση είναι μία σημαντική και ενδιαφέρουσα ικανότητα ορισμένων ζώων, όσον αφορά την επιβίωσή τους ([Reichman, 1984](#)).

Παρά τη μεγάλη προσπάθεια, πολλές πτυχές της αναγέννησης ιστών στα κεφαλόποδα, συμπεριλαμβανομένου του σχετικού μοριακού και κυτταρικού μηχανισμού, παραμένουν σε μεγάλο βαθμό ανεξερεύνητες ([Pamela Imperadore & Graziano Fiorito, 2018](#)). Στην παρούσα εργασία γίνεται η προσπάθεια ανίχνευσης πολυμορφισμών μεταξύ διάφορων κεφαλόποδων με βάση την ανάπτυξη ή όχι της ικανότητας αναγέννησης, ως εξελικτικό χαρακτηριστικό, με την υπόθεση ότι οι πολυμορφισμοί αυτοί είναι μεταλλάξεις. Γίνεται έρευνα για τις μεταλλαγές στα ολόκληρα γονιδιώματα των ειδών, σε ποιες θέσεις εμφανίζονται μονονουκλεοτιδικοί πολυμορφισμοί, αν κωδικοποιούν πρωτεΐνες και αν αλλάζουν τα αμινοξέα ώστε να εκτιμηθεί και η πιθανή επίδραση τους στα γονίδια. Σκοπός είναι η διερεύνηση της εξελικτικής πορείας του μηχανισμού αλλά και η διεύρυνση της γονιδιωματικής βιβλιογραφικής πληροφορίας για τα γονίδια της αναγέννησης.

2. ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ

2.1. Επιλογή ειδών και βάση δεδομένων

Για την εύρεση των πολυμορφισμών χρησιμοποιήθηκε βιοπληροφορική ανάλυση με αλληλουχία ολόκληρου του γονιδιώματος (Whole genome sequencing –

WGS). Οι αναλύσεις πραγματοποιήθηκαν σε λειτουργικό σύστημα Linux (Ubuntu 20.04.6 LTS, 64 -bit), με διάφορα ειδικά προγράμματα γενετικής ανάλυσης.

Στην έρευνα χρησιμοποιήθηκαν διπλού διαβάσματος (paired) αλληλουχίες από γονιδιώματα, ενδεικτικά για 5 είδη με ικανότητες αναγέννησης και 5 αντίστοιχα συγγενικά χωρίς να έχει αποδειχθεί πως διαθέτουν τον μηχανισμό. Από βάσεις δεδομένων όπως του NCBI (National Center for Biotechnology Information) και το ENA (European Nucleotide Archive) του Ευρωπαϊκού Ινστιτούτου Βιοπληροφορικής (EMBL-EBI), λήφθηκαν τα αρχεία από πειράματα (SRA) σε μορφή FASTQ με τις ακατέργαστες πληροφορίες. Το SRA είναι μια ευρεία συλλογή πειραματικών αλληλουχιών DNA και RNA που αντιπροσωπεύουν την ποικιλομορφία του γονιδιώματος. Μέσω αυτής της βάσης δεδομένων, οι ερευνητές μπορούν να αναζητήσουν μεταδεδωμένα για αυτές τις αλληλουχίες για να εντοπίσουν τις αναγνώσεις της ακολουθίας για περαιτέρω αναλύσεις (Leinonen et al., 2011).

Από τα είδη που διαθέτουν ικανότητες αναγέννησης χρησιμοποιήθηκαν: Το *Enteroctopus dolphi* (SRX20019446), το *Octopus bamaculoides* (SRX23769060) το *Nautilus pompilius* (SRX1396872) το *Sepia officinalis* (SRX1396680) και το *Acanthosepion pharaonic* (ERX3454642)(Chavez Ramirez et al., 2023; Holst et al., 2022; Rohrbach & Schmidtberg, 2006; Tressler et al., 2014; Tsujino & Shigeta, 2012). Τα είδη που χρησιμοποιήθηκαν και δεν έχει μελετηθεί αν διαθέτουν ικανότητες αναγέννησης είναι τα εξής: Το *Todarodes sagittatus* (SRX10443654), το *Argonauta argo* (DRX187074), το *Sepioteuthis lessoniana* (SRX26121106), το *Uroteuthis edulis* (ERX14344795) και το *Doryteuthis pealeii* (SRX659582) (Yoshida et al., 2022; Albertin et al., 2022; Kitano et al., 2017; Hsiao et al., 2015 ; Ramirez et al., 2019).

2.2. Ευθυγράμμιση αλληλουχιών (Alignment)

Αρχικά πραγματοποιήθηκε ευθυγράμμιση (alignment) των αλληλουχιών των αρχείων FASTQ με το γονιδίωμα αναφοράς που έγινε μέσω του bwa mem αλγορίθμου. Το bwa mem είναι ένας από τους βασικούς αλγόριθμους του προγράμματος BWA (Burrows- Aligner, (Li and Durbin, 2009)Wheeler), που χρησιμοποιείται για την ευθυγράμμιση και χαρτογράφηση (mapping) αλληλουχιών DNA σε ένα γονιδίωμα αναφοράς. Συγκεκριμένα, το mem της BWA είναι ένας γρήγορος, ακριβής αλγόριθμος για τη χαρτογράφηση αλληλουχίας DNA που διαβάζει σε ένα γονιδίωμα αναφοράς, εκτελώντας τοπική στοίχιση, παράγοντας ευθυγράμμιση για διαφορετικά μέρη της αλληλουχίας προσπαθώντας να βρει τη θέση κάθε αλληλουχίας στο γονιδίωμα αναφοράς. Χρησιμοποιείται για τη χαρτογράφηση αλληλουχιών μικρού μήκους (short reads), όπως αυτές που προκύπτουν από τεχνολογίες αλληλούχισης επόμενης γενιάς, όπως το Illumina με αποτέλεσμα να προκύπτει ένα αρχείο SAM (Sequence Alignment/Map), το οποίο περιέχει τις ακριβείς θέσεις των αλληλουχιών που βρέθηκαν στο γονιδίωμα (Houtgast et al., 2018; Vasimuddin et al., 2019).

Ως γονιδίωμα αναφοράς (reference genome) χρησιμοποιήθηκε ολόκληρο το γονιδίωμα του χταποδιού (*Octopus vulgaris*) (SRX1396869) (Imperadore et al., 2017). Το πρώτο απαραίτητο βήμα ήταν η αντιστοίχιση του γονιδιώματος αναφοράς (reference.fa) που ήταν σε μορφή FASTA σε κατάλληλη μορφή αναγνώρισης (prefix) του προγράμματος (reference_indexed) με την εντολή:

```
$ bwa index -p reference_indexed -a bwtsv reference.fa
```

Η επιλογή -a bwtsv είναι η συνιστώμενη για μεγάλα γονιδιώματα, επειδή επιτρέπει στη BWA να χειριστεί γρήγορα και αποδοτικά τη μεγάλη ποσότητα δεδομένων που εμπεριέχεται στο γονιδίωμα (Lam et al., 2008). Η εντολή:

```
$ bwa mem -t 15 -aM reference_indexed 1.fastq.gz 2.fastq.gz | samtools view -h -b -q 30 @ 10 | samtools sort --threads 10 > file.above.mapQ30.bam
```

χρησιμοποιήθηκε με το bwa mem για την ευθυγράμμιση αλληλουχιών DNA (από αρχεία FASTQ σε ζευγάρια) σε ένα γονιδίωμα αναφοράς και στη συνέχεια επεξεργάστηκαν τα αποτελέσματα με το εργαλείο SAMtools για να δημιουργηθεί ένα αρχείο BAM (Binary Alignment Map) που περιέχει μόνο υψηλής ποιότητας ευθυγραμμίσεις ($\text{mapQ} \geq 30$). Η MAPQ είναι μια βαθμολογία που δείχνει πόσο αξιόπιστα μια ανάγνωση (read) έχει ευθυγραμμιστεί σε μια συγκεκριμένη θέση στο γονιδίωμα. Υψηλότερη τιμή MAPQ σημαίνει μεγαλύτερη εμπιστοσύνη στην ακρίβεια της ευθυγράμμισης. Η -t είναι η επιλογή που καθορίζει πόσα threads θα χρησιμοποιήσει το πρόγραμμα. Τα threads είναι υπολογιστικές μονάδες που επιτρέπουν την ταυτόχρονη εκτέλεση πολλαπλών εργασιών. Χρησιμοποιώντας περισσότερα threads, επιτυγχάνεται ταχύτερα η ευθυγράμμιση σε έναν υπολογιστή με πολλούς επεξεργαστές ή πυρήνες. Το bwa mem παράγει αρχεία σε μορφή SAM και το -aM συμβάλει στην αποφυγή λανθασμένων διπλών ευθυγραμμίσεων και στην διατήρηση των επικεφαλίδων. Η πρώτη εντολή του samtools μετέτρεψε τα αρχεία SAM σε δυαδικά αρχεία BAM (Binary Alignment Map), διατηρώντας τις επικεφαλίδες και φιλτράροντας ποιοτικά τις αξιόπιστες χαρτογραφήσεις με σκορ ≥ 30 , απορρίπτοντας δηλαδή τις χαμηλής ποιότητας ευθυγραμμίσεις. Το SAMtools είναι ένα πακέτο λογισμικού για την ανάλυση και το χειρισμό ευθυγραμμίσεων και είναι σε θέση να μετατρέπει σε άλλες μορφές, να ταξινομεί και να συγχωνεύει ευθυγραμμίσεις, να αφαιρεί διπλότυπα PCR, να δημιουργεί πληροφορίες ανά θέση σε μορφή pileup, να καλεί SNP και σύντομες παραλλαγές και να εμφανίζει τις ευθυγραμμίσεις σε κείμενο (Li et al., 2009). Η δεύτερη εντολή SAMtools ταξινόμησε τα αποτελέσματα ανά

χρωμόσωμα και θέση, προετοιμάζοντας τα αρχεία για περαιτέρω επεξεργασία, όταν τελειώσουν οι μετατροπές για όλα τα είδη.

2.3. Bcftools

Στη συνέχεια χρησιμοποιήθηκε το πρόγραμμα bcftools (Danecek et al., 2021) για την δημιουργία ενός αρχείου BCF (binary call format) που περιέχει τις γενετικές παραλλαγές (SNPs και τις μικρές εισαγωγές/διαγραφές - indels) από όλες τις ευθυγραμμίσεις των αναλύσεων:

```
$ bcftools mpileup -f reference_indexed -v listH | bcftools call -mv -Ob -o all_H.bcf
```

Το bcftools mpileup, διαβάζει τις ευθυγραμμίσεις από τα αρχεία BAM για κάθε θέση του γονιδιώματος και δημιουργεί μια συλλογή δεδομένων βάσει της ευθυγράμμισης για κάθε βάση στο γονιδίωμα. Στη συνέχεια υπολογίστηκαν οι πιθανότητες γονότυπου, αντιπροσωπεύοντας πόσο συνεπή είναι τα παρατηρούμενα δεδομένα με τους πιθανούς διπλοειδείς γονότυπους. Ο υπολογισμός λαμβάνει υπόψη τις ιδιότητες χαρτογράφησης των αναγνώσεων, τις ποιότητες των βάσεων και την πιθανότητα τοπικής εσφαλμένης ευθυγράμμισης. Το δεύτερο βήμα, bcftools call αξιολόγησε τον πιο πιθανό γονότυπο υπό την παραδοχή της ισορροπίας Hardy-Weinberg χρησιμοποιώντας συχνότητες αλληλόμορφων που υπολογίζονται από τα δεδομένα (Danecek et al., 2021). Το listH αναφέρεται σε αρχείο .txt με την μορφή λίστας όλων των ειδών (path file name), ενώ το -v χρησιμοποιείται για το φιλτράρισμα μόνο των θέσεων με παραλλαγές (SNPs και indels), απορρίπτοντας τις βάσεις που δεν παρουσιάζουν παραλλαγές (matches). Η επεξεργασία αυτή γίνεται στο BCF που είναι μικρότερο σε μέγεθος και έτσι βελτιώνεται η απόδοση της ανάλυσης σε σχέση με την VCF μορφή. Τέλος, εκτελέστηκε η κατάλληλη εντολή για την μετατροπή του αρχείου από BCF σε VCF (Variant Calling Format):

```
$ bcftools view all_H.bcf > all_H.vcf
```

2.4. Επεξεργασία VCF

Στη συνέχεια, το αρχείο VCF φιλτραρίστηκε αφαιρώντας όλες τις παραλλαγές που ήταν εισαγωγές και διαγραφές (indels), κρατώντας μόνο τους μονονουκλεοτιδικούς πολυμορφισμούς με ακριβή δεδομένα σε κάθε παραλλαγή (τουλάχιστον 80% πληροφoρίας) και έτσι δημιουργήθηκε το αρχείο all.biallelic.08.maf.recode.vcf:

```
$ vcftools --vcf all_H.vcf --remove-Indels --max-missing 0.8 --max 0.05 --recode --recode-INFO-all --out all.biallelic.08.maf
```

Στα αρχεία VCF μπορούμε να παρατηρήσουμε τα δεδομένα μας και να λάβουμε αποφάσεις σχετικά με το πώς θα τα επεξεργαστούμε, καθώς χρησιμοποιούνται από διάφορα λογισμικά προγράμματα βιοπληροφορικής (Zheng et al., 2017). Ο αριθμός των SNPs υπολογίζεται με την εντολή :

```
$ grep -cv "#" file.vcf
```

2.5. Glm ανάλυση (General linear model)

Δημιουργήθηκε ένα αρχείο .txt για την ομαδοποίηση των ειδών με βάση την χρήση ή όχι της ικανότητας αναγέννησης, χωρισμένο με tab (trait.txt) και στη συνέχεια χρησιμοποιήθηκε το πρόγραμμα TASSEL (Bradbury et al., 2007). Το TASSEL (Trait Analysis by aSSociation, Evolution, and Linkage) είναι ένα λογισμικό που χρησιμοποιείται κυρίως για την ανάλυση συσχετίσεων (association analysis), την ανάλυση εξελικτικών σχέσεων (evolutionary relationships) και τη γενετική σύνδεση (linkage) και αναπτύχθηκε για την διευκόλυνση της ανάλυσης γενετικών δεδομένων, όπως SNPs, και τον εντοπισμό συσχετίσεων μεταξύ γενετικών δεικτών και φαινοτύπων (Bradbury et al., 2007). Το TASSEL είναι ένα από τα πιο δημοφιλή εργαλεία για την πραγματοποίηση GWAS (Genome-Wide Association Studies), όπου ανιχνεύονται SNPs που σχετίζονται με φαινοτυπικές παραλλαγές (Leamy et al., 2017). Αρχικά

χρησιμοποιήθηκε εντολή για να ταξινομηθούν τα SNPs του VCF ανά θέση (POS ή χρωμόσωμα):

```
$ /tassel-5-standalone/run_pipeline.pl -Xmx50g -SortGenotypeFilePlugin inputFile
file.vcf -outputFile file_sort -fileType VCF
```

Η Γενικευμένη Γραμμική Μοντελοποίηση GLM χρησιμοποιείται συχνά σε μελέτες συσχέτισης σε όλο το γονιδίωμα για να βρεθούν οι παραλλαγές DNA που σχετίζονται με συγκεκριμένους φαινοτύπους (Clayton and Leung, 2007; Hilbe, 2011). Η γραμμική εντολή του tassel που ακολούθησε συνδύασε (combine3) τα γενετικά δεδομένα από το αρχείο VCF (input1) με τα φαινοτυπικά δεδομένα από το αρχείο trait.txt (input2), περιέχοντας μόνο τα κοινά δεδομένα (intersect), για την ανίχνευση αυτών των συσχετίσεων με γραμμικά μοντέλα (FixedEffectLMPlugin). Έκανε 1000 μεταθέσεις για να εκτιμήσει τη σημαντικότητα των αποτελεσμάτων (nperm 1000) και φίλτραρε τα αποτελέσματα με p-value μεγαλύτερο από 0.05:

```
$ /tassel-5-standalone/run_pipeline.pl -Xmx50g -fork1 -vcf file_sort.vcf -fork2 -
importGuess trait.txt -combine3 -input1 -input2 -intersect -FixedEffectLMPlugin -
permute true -nperm 1000 -maxP 0.05 -endPlugin -export file_sort_GLM.txt -runfork1
-runfork2
```

Μία ακόμα εντολή του TASSEL έγινε για το φιλτράρισμα του VCF αρχείου ώστε να περιέχει τα ID των μεταλλάξεων:

```
$ /tassel-5-standalone/run_pipeline.pl -Xmx50g -vcf file.vcf -includeSiteNamesInFile
outliersList -export file_outlier -exportType VCF
```

Πραγματοποιήθηκε και η διόρθωση bonferroni, που χρησιμοποιείται συχνά σε GWAS (έρευνες συσχέτισης φαινοτύπων με τεράστιο αριθμό SNPs) και συντελεί στη μείωση των ψευδώς θετικών αποτελεσμάτων (Type I errors) (Emerson, 2020; Gao et al., 2010). Ο συντελεστής bonferroni α υπολογίζεται διαιρώντας το επίπεδο

σημαντικότητας (με την εντολή ορίσαμε $p=0,05$) διά τα συνολικά SNPs που βρεθήκαν (Napierala, 2012). Η εντολή δημιούργησε ένα νέο αρχείο VCF περιέχοντας μόνο SNPs με επίπεδο σημαντικότητας μικρότερου του α και έγινε με αρχείο δέσμης εντολών (script) μέσω της γλώσσας προγραμματισμού Python 3.12 (Παρ. 7.1) (Ζουρνατζής, 2024).

2.6. Πρόβλεψη γενετικών παραλλαγών SnpEff

Το SnpEff είναι ένα βιοπληροφορικό πρόγραμμα για την ανάλυση και την πρόβλεψη των επιπτώσεων γενετικών παραλλαγών στα SNPs, σχολιάζοντας με βάση τις γονιδιωματικές τους θέσεις και τα προβλεπόμενα αποτελέσματα αφορούν γονίδια που κωδικοποιούν πρωτεΐνες, με τους αντίστοιχους γονότυπους να προσδιορίζονται, όποτε είναι διαθέσιμες οι πληροφορίες. (Cingolani et al., 2012).

Δημιουργήθηκε μια βάση δεδομένων με το γονιδίωμα αναφοράς του *Octopus vulgaris*:

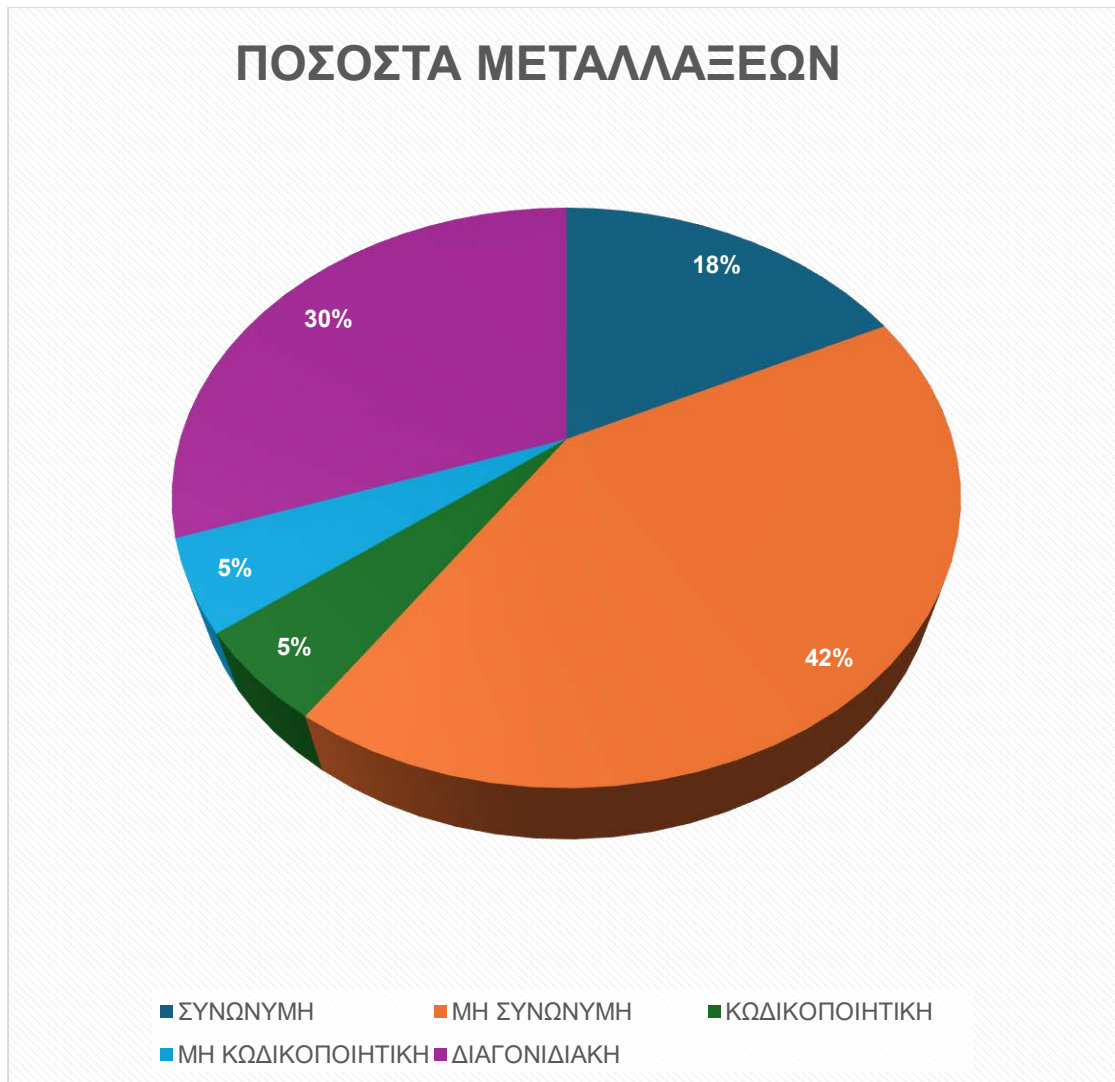
```
$ java -jar snpeff.jar build -gff3 -v reference
```

Το τελικό VCF με τις πληροφορίες, τους σχολιασμούς για τα SNPs και τα γονίδια δημιουργήθηκε με την παρακάτω εντολή:

```
$ java -jar snpeff -jar ann reference file.vcf >
```

annotated.vcf 3. ΑΠΟΤΕΛΕΣΜΑΤΑ

Από το bcftools mpileup βγήκαν 2.046.961 SNPs και μετά το φιλτράρισμα του VCFtools βγήκαν 39.486 SNPs. Έπειτα από τη διόρθωση Bonferroni βγήκαν 79 SNPs, εκ των οποίων τα 33 ήταν μη συνώνυμα, τα 14 συνώνυμα, τα 24 διαγονιδιακά, τα 4 κωδικοποιητικά και τα άλλα 4 μη κωδικοποιητικά. Τα συνολικά SNPs παρουσιάζονται στον παρακάτω πίνακα ανά ποσοστό %:



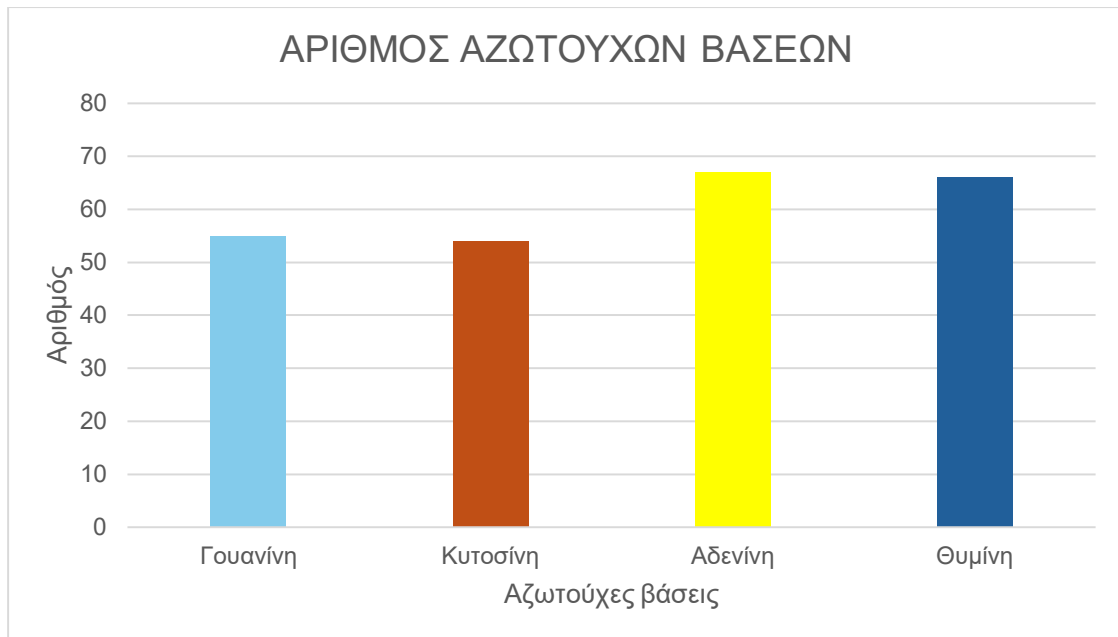
Διάγραμμα 1: Σύνολο μεταλλάξεων %.

Τα 33 μη συνώνυμα SNPs παρουσιάζονται στον πίνακα 1 μεταξύ των ειδών που εμφανίζουν ικανότητες αναγέννησης και αυτών που είναι υπό διερεύνηση. Στην πρώτη στήλη του πίνακα έχουμε το χρωμόσωμα δίπλα η θέση μετάλλαξης και στην τρίτη στήλη τα γονίδια στα οποία έχει γίνει η μετάλλαξη σε σχέση με το αρχικό γονίδιο του *O. Vulgaris*, που αποτελεί το γονιδίωμα αναφοράς και παρακάτω στο γράφημα 2 φαίνονται οι αναλογίες των αζωτούχων βάσεων μεταξύ των ειδών, στα μη συνώνυμα SNPs:

Πίνακας 1: Μη συνώνυμα SNPs μεταξύ των κεφαλόποδων που εμφανίζουν ικανότητες αναγέννησης (πράσινο χρώμα) και αυτών που είναι υπό διερεύνηση (κόκκινο χρώμα).

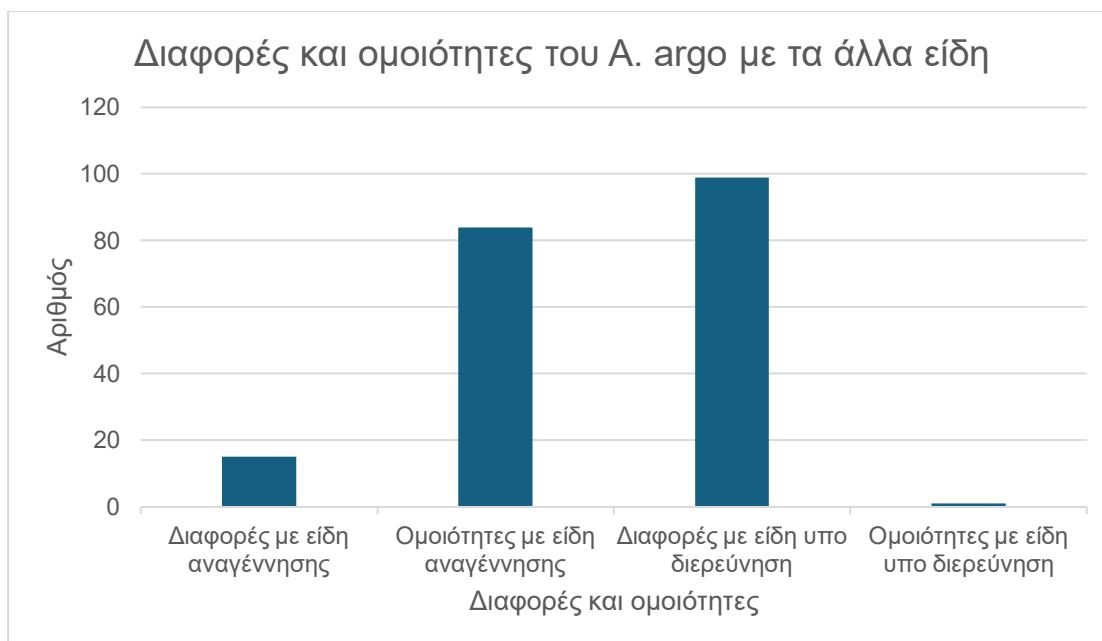
Οι αζωτούχες βάσεις παρουσιάζονται με τα εξής χρώματα: G (Γουανίνη) γαλάζιο χρώμα, A (Αδενίνη) κίτρινο χρώμα, T (Θυμίνη) μπλε χρώμα, C (Κυτοσίνη) καφέ χρώμα.

ΧΡΩΜΟΣΩΜΑ	ΘΕΣΗ	ΓΟΝΙΔΙΟ	ΓΟΝΙΔΙΩΜΑ ΑΝΑΦΟΡΑΣ (<i>Octopus vulgaris</i>)	<i>A. pharaonis</i>	<i>S. officinalis</i>	<i>E. dolfini</i>	<i>O. bimaculoi</i>	<i>T. sagittatus</i>	<i>U. edulis</i>	<i>S. lessoniana</i>	<i>D. pealeii</i>	<i>A. argo</i>
OX597814.1	7421147	OCTVUL_1B021034	G	G	G	G	G	N	A	A	A	G
OX597814.1	141866571	OCTVUL_1B027934	G	A	N	C	C	N	G	G	G	C
OX597814.1	159619212	OCTVUL_1B013700	G	A	N	A	T	G	T	G	G	A
OX597815.1	36759246	OCTVUL_1B015725	G	G	N	G	G	A	A	A	A	G
OX597815.1	66458077	OCTVUL_1B019845	G	G	G	G	G	A	A	A	A	N
OX597815.1	101682325	OCTVUL_1B020849	A	A	A	A	A	N	G	N	G	A
OX597815.1	113238336	OCTVUL_1B010976	T	T	T	T	T	N	C	C	C	T
OX597815.1	127646660	OCTVUL_1B008128	C	A	A	A	A	N	C	N	C	A
OX597816.1	167349273	OCTVUL_1B024134	A	G	G	N	G	A	A	A	A	N
OX597817.1	160380787	OCTVUL_1B025095	G	G	N	G	G	N	C	C	C	G
OX597817.1	173233759	OCTVUL_1B006565	T	N	N	G	G	T	T	T	T	G
OX597818.1	91768802	OCTVUL_1B030013	C	C	N	C	C	A	A	A	A	C
OX597819.1	70232576	OCTVUL_1B008596	T	T	T	T	T	N	C	C	C	T
OX597819.1	113836039	OCTVUL_1B011617	T	T	T	T	T	N	G	N	G	G
OX597820.1	105248155	OCTVUL_1B014941	G	A,T	G	G	G	A	A	A	A	N
OX597820.1	107133927	OCTVUL_1B004696	G	T	A/T	G	G	A	A	A	A	N
OX597822.1	95236236	OCTVUL_1B026355	G	T	T	T	N	G	G	G	G	T
OX597824.1	8208190	OCTVUL_1B030221	T	T	T	T	T	N	C	C	C	T
OX597827.1	3714354	OCTVUL_1B019154	T	C,A	N	T	T	C	C	A	C	T
OX597827.1	9372280	OCTVUL_1B027670	G	T	N	G	G	C,T	C,T	C,T	C	G
OX597827.1	33822165	OCTVUL_1B011480	T	T	T	C	T	A	A	A	A	N
OX597828.1	61348485	OCTVUL_1B000869	G	A	A	A	A	N	G	N	G	A
OX597828.1	75217919	OCTVUL_1B021862	T	T	T	T	T	N	C	C	C	T
OX597829.1	33622732	OCTVUL_1B006827	T	T	T	T	T	N	C	C	C	T
OX597830.1	40199582	OCTVUL_1B018675	A	T	N	A	A	T	T	T	T	T/G
OX597830.1	40199585	OCTVUL_1B018675	T	C	N	T	T	C	C	C	C	T
OX597834.1	40741783	OCTVUL_1B015815	A	G	A	A	A	C	N	C	C	G
OX597840.1	4791718	OCTVUL_1B022941	A	A	A	A	A	N	C	N	C	A
OX597840.1	10018293	OCTVUL_1B015072	A	C	N	C	C	N	A	A	A	C
OX597841.1	1960952	OCTVUL_1B011853	G	G	G	G	G	N	A	A	A	G
OX597841.1	11981367	OCTVUL_1B008964	C	T	T	T	T	N	C	C	C	T
OX597841.1	19019077	OCTVUL_1B024471	T	T	T	T	T	N	G	N	G	T
OX597843.1	5787960	OCTVUL_1B010311	C	C	C	C	C	N	T	T	T	C



Διάγραμμα 2: Αριθμός αζωτούχων βάσεων μεταξύ των ειδών, στα μη συνώνυμα SNPs. Γουανίνη(G), Κυτοσίνη(C), Αδερίνη(A), Θυμίνη(T).

Στο γράφημα που ακολουθεί βλέπουμε τις διαφορές και τις ομοιότητες του είδους *A. Argo* με τα υπόλοιπα είδη, όσον αφορά τις μη συνώνυμες μεταλλάξεις. Το *A. argo* είναι υπό διερεύνηση σχετικά με το αν εμφανίζει ικανότητες αναγέννησης. Παρατηρούμε πως έχει περισσότερες ομοιότητες με τα είδη που διαθέτουν ικανότητες αναγέννησης παρά με τα είδη που είναι υπό διερεύνηση.



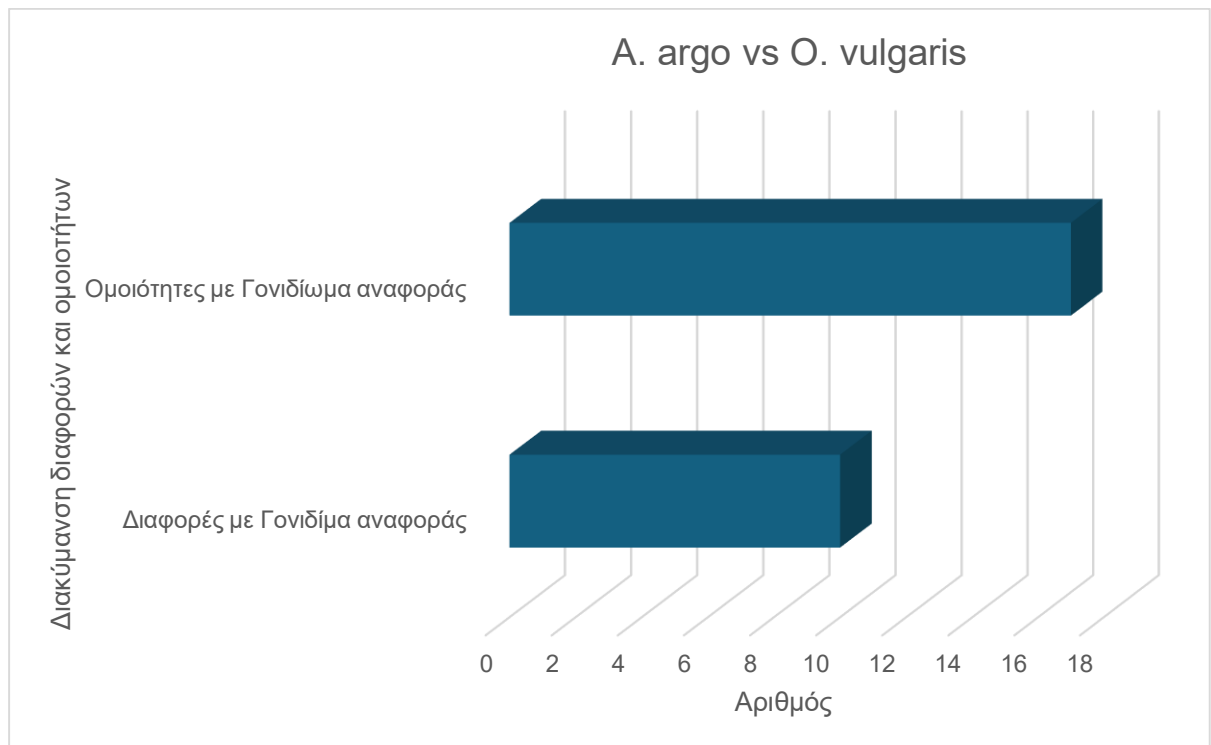
Διάγραμμα 3: Διαφορές και ομοιότητες του είδους A. Argo με τα υπόλοιπα είδη στα μη συνώνυμες SNPs.

Στον παρακάτω πίνακα παρουσιάζονται 4 κωδικοποιητικά SNPs, εκ των οποίων η μία είναι μετάλλαξη πρόωρου τερματισμού (stop gained) και οι άλλες τρεις είναι μεταλλάξεις απώλειας κωδικονίου λήξης (stop lost). Στην πρώτη στήλη του πίνακα φαίνεται το χρωμόσωμα, στη δεύτερη η θέση μετάλλαξης, στην τρίτη το είδος μετάλλαξης και στην τέταρτη το γονίδιο στο οποίο έχει γίνει η μετάλλαξη στα είδη που διερευνώνται, σε σχέση με το γονιδίωμα αναφοράς (*O. Vulgaris*).

Πίνακας 2: Τα κεφαλόποδα που εμφανίζουν ικανότητες αναγέννησης (πράσινο χρώμα) και αυτά που είναι υπό διερεύνηση (κόκκινο χρώμα). Οι αζωτούχες βάσεις παρουσιάζονται με τα εξής χρώματα: G (Γουανίνη) γαλάζιο χρώμα, A (Αδενίνη) κίτρινο χρώμα, T (Θυμίνη) μπλε χρώμα, C (Κυτοσίνη) Καφέ χρώμα.

ΧΡΩΜΟΣΩΜΑ	ΘΕΣΗ	ΕΙΔΟΣ ΜΕΤΑΜΟΡΦΩΣΗΣ	ΓΟΝΙΔΙΟ	ΓΟΝΙΔΙΩΜΑ ΑΝΑΦΟΡΑΣ (<i>O. vulgaris</i>)	<i>A. pharaonis</i>	<i>S. officinalis</i>	<i>E. dolferi</i>	<i>O. bimaculoides</i>	<i>T. sagittatus</i>	<i>U. edulis</i>	<i>S. lessoniana</i>	<i>D. pealeii</i>	<i>A. argo</i>
OX597814.1	66337296	Stop gained	OCTVUL_1B029167	G	G	G	G	N	A	A	A	G	G
OX597816.1	1,64E+08	Stop lost	OCTVUL_1B000980	T	N	T	T	C	C	C	C	T	T
OX597819.1	1,14E+08	Stop lost	OCTVUL_1B011617	T	T	T	T	N	G	N	C	T	T
OX597837.1	29030754	Stop lost	OCTVUL_1B014901	G	G	N	G	A	T	A	T	G	G

Στο γράφημα που ακολουθεί (Γράφημα 4) παρουσιάζονται οι διαφορές και οι ομοιότητες του *A. Argo* με το *O. vulgaris* (γονιδίωμα αναφοράς) όσον αφορά τα μη συνώνυμα SNPs. Το συγκεκριμένο γράφημα δημιουργήθηκε για να δείξουμε πως υπάρχουν αρκετές ομοιότητες μεταξύ τους όπως φαίνεται στον πίνακα 1 αλλά και στον πίνακα 2, παρόλο που το συγκεκριμένο είδος (*A. argo*) βρίσκεται υπό διερεύνηση, για το αν διαθέτει ικανότητες αναγέννησης.



Διάγραμμα 4: Διαφορές και ομοιότητες του *A. argo* με το *O. Vulgaris* (Γονιδίωμα αναφοράς) στα μη συνώνυμα SNPs.

4. ΣΥΖΗΤΗΣΗ

Η παρούσα έρευνα, έφερε στην επιφάνεια αρκετά στοιχεία σχετικά με τα γονίδια των διαφόρων ειδών κεφαλόποδων μαλακίων που εξετάστηκαν. Αρχικά βρέθηκαν 79 SNPs εκ των οποίων τα 33 ήταν μη συνώνυμα (missense), τα 14 ήταν συνώνυμα (synonym), τα 24 ήταν διαγονιδιακά (intergenic region), 4 μη κωδικοποιητικά (μία 3prime UTR variant, μία downstream gene variant και δύο

upstream gene variant) και 4 κωδικοποιητικά (μία stop gained και τρεις stop lost). Τα διαγονιδιακά (intergenic region) SNPs είναι αλλαγές που λαμβάνουν μέρος σε περιοχές μη κωδικοποιητών γονιδίων, υπάρχει όμως πιθανότητα να μεταβάλουν τη γονιδιακή έκφραση κάποιου γονιδίου, καθώς σε αυτές τις περιοχές υπάρχουν ρυθμιστικά στοιχεία που συμμετέχουν στην έκφραση γονιδίων (Chen & Tian, 2016; Lloyd et al., 2018). Οι συνώνυμες μεταλλάξεις θεωρούνται «σιωπηλές» καθώς δεν επηρεάζουν την αλληλουχία πρωτεϊνών, εφόσον δεν αλλάζουν το αμινοξύ της πρωτεΐνης που κωδικοποιεί το γονίδιο. Ωστόσο, διαφορετικά σιωπηλά κωδικόνια έχουν διαφορετική μεταφραστική αποτελεσματικότητα, γεγονός που εγείρει το ερώτημα σε ποιο βαθμό τέτοιες μεταλλάξεις είναι πραγματικά ουδέτερες (Waldman et al., 2011). Στην παρούσα έρευνα δεν εστιάσαμε τόσο στα συνώνυμα και στα διαγονιδιακά SNPs καθώς δώσαμε μεγαλύτερη έμφαση στα μη συνώνυμα και κωδικοποιητικά SNPs, τα οποία έχουν μεγαλύτερο αντίκτυπο στο γονιδίωμα αλλά και στο φαινότυπο, μαζί με τα SNPs σε ρυθμιστικές περιοχές (Ramensky et al., 2002).

Τα είδη που μελετήσαμε ήταν όλα κεφαλόποδα μαλάκια, μία κατηγορία οργανισμών που φημίζονται για τις αναγεννητικές τους ικανότητες (Pamela Imperadore & Graziano Fiorito, 2018), ωστόσο όπως φάνηκε και στον πίνακα 1 της εργασίας, τα χταπόδια και οι σουπιές έχουν περισσότερες ομοιότητες μεταξύ τους παρά με τα καλαμάρια. Το είδος ναυτίλου (*N. Pompilius*) που μελετήσαμε αφαιρέθηκε από τους πίνακες καθώς δεν βρήκαμε επαρκή στοιχεία, ώστε να κάνουμε τις απαιτούμενες συγκρίσεις. Τα χταπόδια και οι σουπιές έχει αποδειχθεί επανειλημμένα μέσα από έρευνες, πως διαθέτουν ικανότητες αναγέννησης σε σχέση με τα καλαμάρια που οι πληροφορίες είναι ελάχιστες (Nakajima et al., 2018). Τα 5 είδη που μελετήθηκαν και φέρουν ικανότητες αναγέννησης { Για το ναυτίλο (*N. pompilius*) δεν βρήκαμε επαρκεί στοιχεία (όπως προαναφέρθηκε), όμως ήταν μέσα στη λίστα των ειδών που φέρουν

ικανότητες αναγέννησης} έχει αποδειχθεί μέσα από έρευνες (Chavez Ramirez et al., 2023; Holst et al., 2022; Rohrbach & Schmidtberg, 2006; Tressler et al., 2014; Tsujino & Shigeta, 2012). Τα είδη αυτά ήταν 2 σουπιές (*S. officinalis*, *A. pharaonic*) και 2 χταπόδια (*E. dolfini*, *O. bimaculoides*). Όπως φάνηκε, έχουν πολλές ομοιότητες τόσο μεταξύ τους, όσο και με το γονιδίωμα αναφοράς (*O. vulgaris*), όσον αφορά τα μη συνώνυμα SNPs. Συνεπώς οι έρευνες σχετικά με την αναγεννητική τους ικανότητα επαληθεύονται κατά κάποιο τρόπο, καθώς δεν γνωρίζουμε εάν τα γονίδια που μελετήσαμε, ευθύνονται για την αναγέννηση, εφόσον δεν έχουν μελετηθεί.

Τα αμφισβητούμενα για ικανότητα αναγέννησης είδη που μελετήσαμε ήταν 3 καλαμάρια (*U. edulis*, *D. pealeii*, *S. lessoniana*), ένα θράψαλο (*T. Sagittatus*) και ένα χταπόδι (*A. argo*). Γενικά πολλά στοιχεία αποδεικνύουν ότι τα χταπόδια διαθέτουν ικανότητες αναγέννησης, ωστόσο στο συγκεκριμένο είδος δεν είχε αποδειχθεί ότι διαθέτει αυτή την ικανότητα.

Τα στοιχεία που έχουμε είναι ενθαρρυντικά, καθώς παρατηρούμε περισσότερες ομοιότητες του *A. argo* με τα είδη που διαθέτουν ικανότητες αναγέννησης όπως και με το *O. Vulgaris* (Γονιδίωμα αναφοράς) παρά με αυτά που μάλλον δεν έχουν (πίνακας 1). Ερωτήματα γεννιούνται σχετικά με τα αποτελέσματα που φαίνονται στον πίνακα 2, ο οποίος αφορά τα 4 κωδικοποιητικά SNPs (μία προώρου τερματισμού (stop gained) και τρεις απώλειας κωδικονίου λήξης (stop lost)). Το Stop gained SNP μετατρέπει ένα κωδικόνιο αμινοξέος σε κωδικόνιο λήξης, με αποτέλεσμα η παραγόμενη πρωτεΐνη συνήθως να χάνει τη λειτουργικότητά της (Savas et al., 2006). Το stop lost SNP αλλάζει ένα κωδικόνιο λήξης με αποτέλεσμα η μετάφραση της πρωτεΐνης αυτού του γονιδίου να συνεχίζεται (Degalez et al., 2021). Σύμφωνα με τα αποτελέσματα του πίνακα 2, δεν φαίνεται μόνο το *A. argo* να έχει πανομοιότυπα SNPs με τα είδη που διαθέτουν ικανότητες αναγέννησης, αλλά και το είδος *D. Paleii* που στον πίνακα 1 ένα βρέθηκε

να έχει μόνο μία ομοιότητα στα μη συνώνυμα SNPs με το *A. Argo*, η οποία ήτα και η μοναδική ομοιότητα του *A. Argo* με τα αμφισβητούμενα είδη προς αναγέννηση. Σίγουρα αυτή η πληροφορία είναι σημαντική, καθώς το *D. Paleii* μπορεί να αναπτύσσει κάποιον εξελικτικό μηχανισμό, ίσως και στην αναγέννηση, όμως δεν το γνωρίζουμε αυτό ακόμα. Σημαντικά ήταν και τα στοιχεία του ταυτίζουν το *A. Argo* με τα είδη που διαθέτουν ικανότητες αναγέννησης όμως χρειάζονται περισσότερες μελέτες και έρευνες να πραγματοποιηθούν στο μέλλον, ώστε να αποδειχθεί το οτιδήποτε.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Ελληνική,

Ζουρνατζής, Βασίλειος Χ. Ηχοεντοπισμός: μελέτη ανίχνευσης πολυμορφισμών με βιοπληροφορική ανάλυση. 2024. Master's Thesis.

Ξενόγλωσση

Acosta-Jofré, M. S., Sahade, R., Laudien, J., & Chiappero, M. B. (2012). A contribution to the understanding of phylogenetic relationships among species of the genus *Octopus* (Octopodidae: Cephalopoda). *Scientia Marina*, 76(2), 311-318. <https://doi.org/10.3989/scimar.03365.03B>

Albertin, C.B., Medina-Ruiz, S., Mitros, T. et al. Genome and transcriptome mechanisms driving cephalopod evolution. *Nat Commun* 13, 2427 (2022). <https://doi.org/10.1038/s41467-022-29748-w>

Bely, A. E., & Nyberg, K. G. (2010). Evolution of animal regeneration: re-emergence of a field. *Trends in Ecology & Evolution*, 25(3), 161-170. <https://doi.org/10.1016/j.tree.2009.08.005>

Brookes, A. J. (1999). The essence of SNPs. *Gene*, 234(2), 177-186. [https://doi.org/https://doi.org/10.1016/S0378-1119\(99\)00219-X](https://doi.org/https://doi.org/10.1016/S0378-1119(99)00219-X)

Chavez Ramirez, C., Khoo, M., Lopez, G. M., Ferguson, S., Walker, S., & Echeverri, K. (2023). Comparison of Blastema Formation after Injury in Two Cephalopod Species. *MicroPubl Biol*, 2023. <https://doi.org/10.17912/micropub.biology.000946>

Chen, J., & Tian, W. (2016). Explaining the disease phenotype of intergenic SNP through predicted long range regulation. *Nucleic Acids Res*, 44(18), 8641-8654. <https://doi.org/10.1093/nar/gkw519>

Degalez, F., Jehl, F., Muret, K., Bernard, M., Lecerf, F., Lagoutte, L., Désert, C., Pitel, F., Klopp, C., & Lagarrigue, S. (2021). Watch Out for a Second SNP: Focus on Multi-Nucleotide Variants in Coding Regions and Rescued Stop-Gained. *Front Genet*, 12, 659287. <https://doi.org/10.3389/fgene.2021.659287>

Echeverri, K., & Zayas, R. M. (2018). Regeneration: From cells to tissues to organisms. *Dev Biol*, 433(2), 109-110. <https://doi.org/10.1016/j.ydbio.2017.12.005>

Elchaninov, A., Sukhikh, G., & Fatkhudinov, T. (2021). Evolution of Regeneration in Animals: A Tangled Story [Review]. *Frontiers in Ecology and Evolution*, Volume 9 - 2021. <https://doi.org/10.3389/fevo.2021.621686>

Holst, M. M., Hauver, C. M., Stein, R. S., Milano, B. L., Levine, L. H., Zink, A. G., Watters, J. V., & Crook, R. J. (2022). Behavioral changes in senescent giant Pacific octopus (*Enteroctopus dofleini*) are associated with peripheral neural degeneration and loss of epithelial tissue. *Comp Biochem Physiol A Mol Integr Physiol*, 271, 111263. <https://doi.org/10.1016/j.cbpa.2022.111263>

Hsiao CD, Shen KN, Ching TY, Wang YH, Ye JJ, Tsai SY, Wu SC, Chen CH, Wang CH. The complete mitochondrial genome of the cryptic "lineage A" big-fin reef squid, *Sepioteuthis lessoniana* (Cephalopoda: Loliginidae) in Indo-West Pacific. *Mitochondrial DNA A DNA Mapp Seq Anal*. 2016 Jul;27(4):2433-4. doi: 10.3109/19401736.2015.1030630. Epub 2015 May 27. PMID: 26016882.

Imperadore, P., & Fiorito, G. (2018). Cephalopod Tissue Regeneration: Consolidating Over a Century of Knowledge [Review]. *Frontiers in Physiology*, Volume 9 - 2018. <https://doi.org/10.3389/fphys.2018.00593>

Imperadore, P., & Fiorito, G. (2018). Cephalopod Tissue Regeneration: Consolidating Over a Century of Knowledge. *Front Physiol*, 9, 593. <https://doi.org/10.3389/fphys.2018.00593>

Imperadore, P., Shah, S. B., Makarenkova, H. P., & Fiorito, G. (2017). Nerve degeneration and regeneration in the cephalopod mollusc *Octopus vulgaris*: the case of the pallial nerve. *Scientific Reports*, 7(1), 46564. <https://doi.org/10.1038/srep46564>

Kitano H, Nagano N, Sakaguchi K, Matsuyama M. Two vitellogenins in the loliginid squid *Uroteuthis edulis*: Identification and specific expression in ovarian follicles. *Mol Reprod Dev*. 2017 May;84(5):363-375. doi: 10.1002/mrd.22786. Epub 2017 Mar 13. PMID: 28218427.

Kröger, B., Vinther, J., & Fuchs, D. (2011). Cephalopod origin and evolution: A congruent picture emerging from fossils, development and molecules: Extant cephalopods are younger than previously realised and were under major selection to become agile, shell-less predators. *Bioessays*, 33(8), 602-613. <https://doi.org/10.1002/bies.201100001>

Lenski RE. What is adaptation by natural selection? Perspectives of an experimental microbiologist. *PLoS Genet*. 2017 Apr 20;13(4):e1006668. doi: 10.1371/journal.pgen.1006668. PMID: 28426692; PMCID: PMC5398481.

Lindgren, A. R. (2010). Molecular inference of phylogenetic relationships among Decapodiformes (Mollusca: Cephalopoda) with special focus on the squid Order Oegopsida. *Molecular Phylogenetics and Evolution*, 56(1), 77-90. <https://doi.org/https://doi.org/10.1016/j.ympev.2010.03.025>

Lloyd, J. P., Tsai, Z. T.-Y., Sowers, R. P., Panchy, N. L., & Shiu, S.-H. (2018). A Model-Based Approach for Identifying Functional Intergenic Transcribed Regions and

Noncoding RNAs. *Molecular Biology and Evolution*, 35(6), 1422-1436.
<https://doi.org/10.1093/molbev/msy035>

Lupše, N., Reid, A., Taite, M., Kubodera, T., & Allcock, A. L. (2023). Cuttlefishes (Cephalopoda, Sepiidae): the bare bones—an hypothesis of relationships. *Marine Biology*, 170(8), 93. <https://doi.org/10.1007/s00227-023-04195-3>

Nakajima, R., Shigeno, S., Zullo, L., De Sio, F., & Schmidt, M. R. (2018). Cephalopods Between Science, Art, and Engineering: A Contemporary Synthesis [Review]. *Frontiers in Communication*, Volume 3 - 2018. <https://doi.org/10.3389/fcomm.2018.00020>

Poss, K. D. (2010). Advances in understanding tissue regenerative capacity and mechanisms in animals. *Nature Reviews Genetics*, 11(10), 710-722.
<https://doi.org/10.1038/nrg2879>

Ramensky, V., Bork, P., & Sunyaev, S. (2002). Human non-synonymous SNPs: server and survey. *Nucleic Acids Res*, 30(17), 3894-3900. <https://doi.org/10.1093/nar/gkf493>

Ramírez AS, Vega-Orellana OM, Viver T, Poveda JB, Rosales RS, Poveda CG, Spargser J, Szostak MP, Caballero MJ, Ressel L, Bradbury JM, Mar Tavío M, Karthikeyan S, Amann R, Konstantinidis KT, Rossello-Mora R. First description of two moderately halophilic and psychrotolerant *Mycoplasma* species isolated from cephalopods and proposal of *Mycoplasma marinum* sp. nov. and *Mycoplasma todarodis* sp. nov. *Syst Appl Microbiol*. 2019 Jul;42(4):457-467. doi: 10.1016/j.syapm.2019.04.003. Epub 2019 Apr 12. Erratum in: *Syst Appl Microbiol*. 2020 Jan;43(1):126027. doi: 10.1016/j.syapm.2019.126027. PMID: 31072660.

Reichman, O. J. (1984). Evolution of Regeneration Capabilities. *The American Naturalist*, 123(6), 752-763. <https://doi.org/10.1086/284237>

Rohrbach, B., & Schmidtberg, H. (2006). Sepia arms and tentacles: Model systems for studying the regeneration of brachial appendages. *Vie et Milieu*, 56, 175-190.

Savas, S., Tuzmen, S., & Ozcelik, H. (2006). Human SNPs resulting in premature stop codons and protein truncation. *Hum Genomics*, 2(5), 274-286. <https://doi.org/10.1186/1479-7364-2-5-274>

Seifert, A. W., Monaghan, J. R., Smith, M. D., Pasch, B., Stier, A. C., Michonneau, F., & Maden, M. (2012). The influence of fundamental traits on mechanisms controlling appendage regeneration. *Biological Reviews*, 87(2), 330-345. <https://doi.org/https://doi.org/10.1111/j.1469-185X.2011.00199.x>

Shastri, B.S. SNPs in disease gene mapping, medicinal drug development and evolution. *J Hum Genet* **52**, 871–880 (2007). <https://doi.org/10.1007/s10038-007-0200-z>

Shastri, B. S. (2009). SNPs: Impact on Gene Function and Phenotype. In A. A. Komar (Ed.), *Single Nucleotide Polymorphisms: Methods and Protocols* (pp. 3-22). Humana Press. https://doi.org/10.1007/978-1-60327-411-1_1

Shaw, T. J., Osborne, M., Ponte, G., Fiorito, G., & Andrews, P. L. R. (2016). Mechanisms of wound closure following acute arm injury in *Octopus vulgaris*. *Zoological Letters*, 2(1), 8. <https://doi.org/10.1186/s40851-016-0044-5>

Stayton, C. T. (2015). What does convergent evolution mean? The interpretation of convergence and its implications in the search for limits to evolution. *Interface Focus*, 5(6), 20150039. <https://doi.org/10.1098/rsfs.2015.0039>

Stern, D. L. (2013). The genetic causes of convergent evolution. *Nature Reviews Genetics*, 14(11), 751-764. <https://doi.org/10.1038/nrg3483>

- Tressler, J., Maddox, F., Goodwin, E., Zhang, Z., & Tublitz, N. J. (2014). Arm regeneration in two species of cuttlefish *Sepia officinalis* and *Sepia pharaonis*. *Invert Neurosci*, 14(1), 37-49. <https://doi.org/10.1007/s10158-013-0159-8>
- Tsujino, Y., & Shigeta, Y. (2012). Biological response to experimental damage of the phragmocone and siphuncle in *Nautilus pompilius* Linnaeus. *Lethaia*, 45, 443–449. <https://doi.org/10.1111/j.1502-3931.2012.00306.x>
- van Wijnen, A. J., & Lewallen, E. A. (2024). Natural selection and evolution: evolving concepts. *Acad Biol*, 2(2). <https://doi.org/10.20935/AcadBiol6245>
- Vermeij, G. J. (2023). When does natural selection take place? *Evolution*, 77(8), 1739-1743. <https://doi.org/10.1093/evolut/qpad108>
- Voight, J. R. (1993). The association between distribution and octopodid morphology: implications for classification. *Zoological Journal of the Linnean Society*, 108(3), 209-223. <https://doi.org/10.1111/j.1096-3642.1993.tb00296.x>
- Waldman, Y. Y., Tuller, T., Keinan, A., & Ruppin, E. (2011). Selection for translation efficiency on synonymous polymorphisms in recent human evolution. *Genome Biol Evol*, 3, 749-761. <https://doi.org/10.1093/gbe/evr076>
- Westermann, B., Beck-Schildwächter, I., Beuerlein, K., Kaleta, E. F., & Schipp, R. (2004). Shell growth and chamber formation of aquarium-reared *Nautilus pompilius* (Mollusca, Cephalopoda) by X-ray analysis. *Journal of Experimental Zoology Part A: Comparative Experimental Biology*, 301A(12), 930-937. <https://doi.org/https://doi.org/10.1002/jez.a.116>
- Y, T., & Y, S. (2012). Biological response to experimental damage of the

phragmocone and siphuncle in *Nautilus pompilius* Linnaeus. *MicroPubl Biol*, 2012.

<https://doi.org/10.17912/micropub.biology.000946>

Yoshida MA, Hirota K, Imoto J, Okuno M, Tanaka H, Kajitani R, Toyoda A, Itoh T, Ikeo K, Sasaki T, Setiamarga DHE. Gene Recruitments and Dismissals in the Argonaut Genome Provide Insights into Pelagic Lifestyle Adaptation and Shell-like Eggcase Reacquisition. *Genome Biol Evol.* 2022 Nov 4;14(11):evac140. doi: 10.1093/gbe/evac140. PMID: 36283693; PMCID: PMC9635652.

Zhang, Y., Mao, F., Mu, H., Huang, M., Bao, Y., Wang, L., Wong, N.-K., Xiao, S., Dai, H., Xiang, Z., Ma, M., Xiong, Y., Zhang, Z., Zhang, L., Song, X., Wang, F., Mu, X., Li, J., Ma, H., . . Yu, Z. (2021). The genome of *Nautilus pompilius* illuminates eye evolution and biomineralization. *Nature Ecology & Evolution*, 5(7), 927-938. <https://doi.org/10.1038/s41559-021-01448-6>

ABSTRACT

The evolution of regenerative capacity in multicellular animals is one of the most complex and interesting problems in biology. Regeneration, a process that consists of the regeneration of damaged structures and their functional restoration, is widespread in many lineages of the animal kingdom, from lower invertebrates to mammals. Among species capable of regeneration, the actual ability to restore the full form and function of injured tissue varies considerably, from being able to undergo regeneration of the entire body and internal organs, to cases where this ability is limited to a few tissues. Among invertebrates, cephalopod mollusks retain the ability to regenerate various structures (i.e., muscles, nerves, or entire segments). The present research focuses mainly on SNPs (single nucleotide polymorphisms) of cephalopod molluscs. 10 species were studied, of which 5 have proven regenerative abilities and the other 5 are under investigation, as to whether they possess this specific ability. A bioinformatic analysis was performed. A bioinformatic analysis was carried out with the aim of observe differences and similarities between species in relation to the selected reference genome. The analyses were performed on a Linux operating system (Ubuntu 20.04.6 LTS, 64 -bit), with various special genetic analysis programs, Whole genome sequencing (WGS) and paired-end sequences were used from genomes from databases such as NCBI (National Center for Biotechnology Information) and ENA (European Nucleotide Archive) of the European Bioinformatics Institute (EMBL-EBI), the experimental sequence alignment (SRA) files were downloaded in FASTQ format. Then the sequences were aligned sequences, filtering of the VCF file and GLM analysis to identify SNPs. 33 non-synonymous SNPs were found and these were the focus of the present study. The results showed a clear separation of the two groups, but more intensive research is needed for better results.

ΠΑΡΑΡΤΗΜΑ

7.1. Python script για την εύρεση των SNPs με το ζητούμενο p

```
threshold = 0.05 / 20252446
```

```
input_file = ballelic_sort_GLM1.txt'
```

```
output_file = filtered_output.txt'
```

```
with open(input_file, 'r') as infile, open(output_file, 'w') as outfile:
```

```
    header = infile.readline()
```

```
    outfile.write(header)
```

```
    header_columns = header.strip().split('\t')
```

```
    p_value_index = header_columns.index('p')
```

```
    for line in infile:
```

```
        columns = line.strip().split('\t')
```

```
        if len(columns) == len(header_columns):
```

```
            p_value = float(columns[p_value_index])
```

```
            if p_value < threshold:
```

```
                outfile.write(line)
```

```
print(f"Filtered data has been written to {output_file}")
```