



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΙΑΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Ταυτοποίηση των πρωτεϊνικών υποδοχέων και των γονιδίων χημειοαντίληψης που καθοδηγούν τις διατροφικές και αναπαραγωγικές προτιμήσεις του δάκου της ελιάς



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΤΜΗΜΑ ΒΙΟΧΗΜΕΙΑΣ & ΒΙΟΤΕΧΝΟΛΟΓΙΑΣ
ΕΡΓΑΣΤΗΡΙΟ ΜΟΡΙΑΚΗΣ ΒΙΟΛΟΓΙΑΣ & ΓΟΝΙΔΙΩΜΑΤΙΚΗΣ

Ταυτοποίηση των πρωτεϊνικών υποδοχέων και των γονιδίων
χημειοαντίληψης που καθοδηγούν τις διατροφικές και αναπαραγωγικές
προτιμήσεις του δάκου της ελιάς

Identification of protein receptors and sensory genes guiding the dietary
and reproductive preferences of the olive fruit fly

Μπαϊλγκάμη Ελπίδα

Η παρούσα διπλωματική εργασία εκπονήθηκε στο εργαστήριο Μοριακής Βιολογίας και Γονιδιωματικής του τμήματος Βιοχημείας & Βιοτεχνολογίας του Πανεπιστημίου Θεσσαλίας υπό την επίβλεψη του καθηγητή Μοριακής Βιολογίας, κ. Μαθιόπουλου Κωνσταντίνου.

ΤΡΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ

Μαθιόπουλος Κωνσταντίνος: Καθηγητής Μοριακής Βιολογίας του Τμήματος Βιοχημείας & Βιοτεχνολογίας του Πανεπιστημίου Θεσσαλίας

Γιακουντής Αντώνιος: Επίκουρος Καθηγητής Μοριακής Βιολογίας του Τμήματος Βιοχημείας & Βιοτεχνολογίας του Πανεπιστημίου Θεσσαλίας

Τσουμάνη Κωνσταντίνα: Διδάσκουσα Π.Δ. 407 του τμήματος Βιοχημείας & Βιοτεχνολογίας του Πανεπιστημίου Θεσσαλίας

ΕΥΧΑΡΙΣΤΙΕΣ

Αρχικά, θα ήθελα να ευχαριστήσω τα μέλη της Τριμελούς Επιτροπής μου, τον κ. Ματθιόπουλο, τον κ. Γιακουνή και την κ. Τσουμάνη που ευγενικά δέχτηκαν να αξιολογήσουν την πτυχιακή μου.

Ιδιαίτερα, θα ήθελα να ευχαριστήσω τον καθηγητή κ. Ματθιόπουλο Κωνσταντίνο για την εμπιστοσύνη που μου έδειξε αναθέτοντάς μου το συγκεκριμένο θέμα. Θεωρώ ότι είναι ένας αξιόλογος καθηγητής και σπουδαίος άνθρωπος που εμπνέει τους φοιτητές. Η αμεσότητα του και ο σεβασμός απέναντι στους φοιτητές του, αποτέλεσε κίνητρο για να απευθυνθώ σε αυτόν για το σημαντικό αυτό κομμάτι των σπουδών μου.

Επίσης, θα ήθελα να ευχαριστήσω την κ. Τσουμάνη Κωνσταντίνα για τον πολύτιμο χρόνο που διέθεσε και για την υπομονή της. Την ευχαριστώ για τη σημαντική καθοδήγηση της και τις πολύτιμες συμβουλές που με κατεύθυναν στην επεξεργασία του θέματος αυτού. Επίσης, την ευχαριστώ για την αδιάκοπη συμπαράσταση και ενθάρρυνση της και για τη συμβολή της στην ολοκλήρωση της πτυχιακής μου.

Θερμές ευχαριστίες απευθύνω, επίσης, σε όλους τους καθηγητές και καθηγήτριες μου για τις πολύτιμες γνώσεις που μου πρόσφεραν.

Τέλος, ευχαριστώ μέσα από την καρδιά μου την οικογένεια μου, που ήταν δίπλα μου καθ' όλη τη διάρκεια των φοιτητικών μου χρόνων στηρίζοντας με σε κάθε δύσκολη στιγμή και πιστεύοντας σε μένα. Ιδιαίτερα θα ήθελα να ευχαριστήσω τον αδερφό μου για το υπέροχο εξώφυλλο που σχεδίασε για την εργασία μου.

ΠΕΡΙΛΗΨΗ

Ένας από τους βασικότερους εχθρούς για την αγροτική παραγωγή είναι τα ζημιογόνα έντομα. Η αντιμετώπιση των εντόμων αυτών, μέχρι σήμερα, γίνεται με τη χρήση εντομοκτόνων και παγίδων και συχνά δεν είναι επιτυχής. Ωστόσο, με την πρόοδο των τεχνολογιών ανάλυσης των γονιδιωμάτων παράγονται πλέον νέες πληροφορίες που μπορούν να συνεισφέρουν στην ανάπτυξη εναλλακτικών μεθόδων για την αντιμετώπιση τους. Πιο συγκεκριμένα, τα εργαλεία που παρέχουν οι τομείς της γονιδιωματικής βοηθούν στην ταυτοποίηση και βαθύτερη κατανόηση των βιολογικών συστημάτων που εμπλέκονται στην αναπαραγωγή και την επιβίωσή τους όπως αυτό της χημειοαντίληψης. Στην παραπάνω κατηγορία εντόμων ανήκει και ο δάκος της ελιάς (*Bactrocera oleae*), ένα μονοφάγο έντομο του οποίου το γονιδίωμα έχει πρόσφατα αλληλουχηθεί και πληροφορίες σχετικές με τις αλληλουχίες των γονιδίων τους βρίσκονται αποθηκευμένες σε προσιτές βάσεις δεδομένων. Ωστόσο, δεν έχουν ταυτοποιηθεί τα γονίδια που εμπλέκονται στα συστήματα χημειοαντίληψης και συμβάλλουν στην αναγνώριση της ελιάς. Στην παρούσα διπλωματική εργασία, εστίασαμε στην ταυτοποίηση και την χειροκίνητη επιμέλεια των γονιδίων χημειοαντίληψης του δάκου αξιοποιώντας γονιδιωματικά και μεταγραφωμικά δεδομένα, καθώς και εργαλεία βιοπληροφορικής. Επιπλέον, αναλύσαμε την τοπολογία των γευστικών και των ιονοτροπικών υποδοχέων ώστε να διαπιστωθεί ο αριθμός των διαμεμβρανικών τους περιοχών και ταυτοποιήσαμε τα συντηρημένα αμινοξικά μοτίβα κυστεϊνών για καθεμία από τις OBPs, ώστε να ταξινομηθούν στις τέσσερις γνωστές υπο-οικογένειες. Οι αμινοξικές αλληλουχίες των επιμελημένων γονιδιακών μοντέλων, στη συνέχεια, συγκρίθηκαν με ομόλογες των συγγενικών με το δάκο ειδών *Ceratitis capitata*, *Drosophila melanogaster* και *Musca domestica*. Από τη μελέτη των φυλογενετικών δέντρων που δημιουργήθηκαν ήταν εφικτή η κατηγοριοποίηση των γονιδίων του δάκου βάσει των λειτουργιών των ομόλογων στα υπόλοιπα τρία είδη και η ταυτοποίηση γονιδιακών επεκτάσεων στο δάκο. Τα διπλασιασμένα γονίδια που ταυτοποιήθηκαν πιθανότατα είναι και αυτά που κάνουν το δάκο να ξεχωρίζει από τα υπόλοιπα είδη όσον αφορά την αποκλειστική προτίμηση του προς τους καρπούς της ελιάς. Ωστόσο, η λειτουργία των γονιδίων αυτών δεν είναι ακόμα γνωστή και γι' αυτό κρίνεται απαραίτητη περαιτέρω έρευνα για τη διερεύνηση του ρόλου τους.

ABSTRACT

One of the major threats for agricultural production are insect pests. Commonly used methods for their confrontation are based on the use of insecticides and bait sprays which lack effectiveness. However, progress in genomics research technologies has produced a wide range of genomic data that can contribute to the development of alternative pest control methods. Particularly, the tools offered by the different fields of genomics can be utilized for the identification and deeper understanding of the biological systems that guide their reproduction and survival, such as the chemosensory system. Among these pests is included the olive fruit fly (*Bactrocera oleae*) a monophagous species, whose genome has been sequenced and the information related to its genomic sequences are stored in accessible databases. However, the genes implicated in its chemosensory system and the identification of the olive fruit are not yet known. In the present study, we focused on the identification and manual curation of the chemosensory genes' structures exploiting both genomics and transcriptomics data with the use of bioinformatics tools. Additionally, we investigated the topology of the gustatory and ionotropic receptors, to define the exact number of their transmembrane domains and we further identified the conserved cysteine motifs of the OBPs, to enable their classification into the four known OBP subfamilies. The amino acid sequences of the annotated *B.oleae* gene models were then compared with homologues retrieved from the closely related species, *Ceratitis capitata*, *Drosophila melanogaster* and *Musca domestica*. Further analysis of the generated phylogenetic trees enabled the classification of the olive fruit fly's genes based on the functional annotation of their homologues in the compared species, as well as the identification of putative gene expansions in the *Bactrocera oleae* species. The identified putative duplicated genes could be responsible for the unique preference of the fly for the olive fruit. However, the function of these genes is still unknown and for this reason further studies are required to investigate their role.

ΠΕΡΙΕΧΟΜΕΝΑ

1. ΕΙΣΑΓΩΓΗ	9
1.1. Ολιστικές τεχνολογίες ανάλυσης των γονιδιωμάτων	10
1.2. Συγκριτική γονιδιωματική	11
1.3. Βιοπληροφορική ανάλυση	12
1.4. Συνεισφορά της γονιδιωματικής στον τομέα της γεωργίας	13
1.5. Μοριακοί μηχανισμοί που καθοδηγούν τα έντομα – Χημειοαντίληψη	14
1.5.1. Οσμοδεσμευτικές πρωτεΐνες (Odorant-Binding Proteins, OBPs)	14
1.5.2. Οσφρητικοί υποδοχείς (Odorant Receptors, ORs)	15
1.5.3. Γευστικοί υποδοχείς (Gustatory Receptors, GRs)	16
1.5.4. Ιονοτροπικοί υποδοχείς (Ionotropic Receptors, IRs)	17
1.6. Εξέλιξη των γονιδίων χημειοαντίληψης στα έντομα	18
1.6.1. Εξελικτικό μοντέλο «Γέννησης και Θανάτου»	18
1.6.2. Διπλασιασμοί και απώλειες γονιδίων χημειοαντίληψης στα έντομα	19
1.7. <i>Bactrocera oleae</i>	21
1.8. Σκοπός εργασίας	24
2.1. Επιμέλεια συναινετικών γονιδιακών μοντέλων των γονιδίων χημειοαντίληψης (Annotation)	26
2.1.1. Βάση δεδομένων της NCBI (National Center For Biotechnology Information)	26
2.1.2. Blast (Basic Local Alignment Search Tool)	26
2.1.3. Apollo – i5k Initiative	27
2.2. Στοιχίση νουκλεοτιδικών και αμινοξικών αλληλουχιών	29
2.2.1. Clustal Omega	30
2.2.2. Muscle	31
2.3. Φυλογενετική ανάλυση	32
2.3.1. Mega-X	33
2.3.2. Figtree	35
2.4. Αναγνώριση διαμεμβρανικών περιοχών των χημειοϋποδοχέων	35
2.4.1. TMHMM	36
2.5. Ανάλυση αμινοξικών μοτίβων	36
2.5.1. MEME (Multiple Em For Motif Elicitation)	36
2.6. UniProt (Universal Protein Knowledgebase)	37
3.1. Ανάκτηση των αλληλουχιών των γονιδίων χημειοαντίληψης του δάκου και συγγενικών ειδών από την NCBI	39

3.2. Επεξεργασία και ανάλυση των συναινετικών γονιδιακών μοντέλων των GRs, IRs και OBPs του δάκου	40
3.2.1. Επεξεργασία και ανάλυση των συναινετικών γονιδιακών μοντέλων των GRs του δάκου 41	
Εντοπισμός των GRs στα ικρίωματα του γονιδιωματος.....	44
3.2.2. Επεξεργασία των συναινετικών γονιδιακών μοντέλων των IRs του δάκου	47
3.2.3. Επεξεργασία των συναινετικών γονιδιακών μοντέλων των OBPs του δάκου	50
3.3. Φυλογενετική ανάλυση των υπό μελέτη γονιδίων χημειοαντίληψης του δάκου και των συγγενικών ειδών.....	56
3.3.1. Φυλογενετική ανάλυση των GRs του δάκου και των συγγενικών ειδών	56
3.3.2. Φυλογενετική ανάλυση των IRs του δάκου και των συγγενικών ειδών.....	60
3.3.3. Φυλογενετική ανάλυση των OBPs του δάκου και των συγγενικών ειδών	64
3.4. <i>In silico</i> διερεύνηση της λειτουργίας του συνόλου των GRs, IRs και των OBPs στο δάκο της ελιάς	66

1. ΕΙΣΑΓΩΓΗ

1.1. Ολιστικές τεχνολογίες ανάλυσης των γονιδιωμάτων

Οι ολιστικές τεχνολογίες ανάλυσης των γονιδιωμάτων εστιάζουν στη μελέτη του συνόλου των γονιδίων ενός οργανισμού, αξιοποιώντας το συνολικό γενετικό του υλικό (Griffiths *et al*, 2000). Συνεπώς, διαφέρουν από τους κλάδους της γενετικής που εστιάζουν στη μελέτη ενός μεμονωμένου γονιδίου τη φορά (Chen X, 2010). Οι τεχνολογίες αυτές συμπεριλαμβάνουν ένα συνδυασμό μεθόδων DNA και RNA αλληλούχησης και βιοπληροφορικών εργαλείων για την αλληλούχηση, συναρμολόγηση και ανάλυση των υπό μελέτη γονιδιωμάτων. Έτσι, καθίσταται δυνατή η αποκάλυψη και μελέτη των αλληλουχιών των γονιδίων πλήθους οργανισμών, συμπεριλαμβάνοντας και αυτών που προσφέρουν πλεονεκτήματα ή προκαλούν προβλήματα για τον άνθρωπο.

Η διαθεσιμότητα των συνολικών DNA αλληλουχιών¹ από πλήθος οργανισμών είναι καίριας σημασίας για τις γονιδιωματικές μελέτες και έγινε δυνατή με τη συνεισφορά των σύγχρονων τεχνολογιών αλληλούχησης δεύτερης και τρίτης γενιάς που έχουν διαδεχθεί την ευρέως χρησιμοποιούμενη μέθοδο αλληλούχησης κατά Sanger (Kchouk *et al*, 2017). Στις τεχνολογίες αλληλούχησης τρίτης γενιάς ανήκουν η PacBio, η Nanopore, η Quantapore και η Stratos, καθεμία από τις οποίες βασίζεται σε διαφορετικές προσεγγίσεις για την αλληλούχηση ενός DNA μορίου. Έτσι, καθεμία από αυτές παρέχει διαφορετικά πλεονεκτήματα και επιλέγεται ανάλογα με τις ανάγκες κάθε βιολογικής μελέτης. Μερικά από τα πλεονεκτήματα που προσφέρουν συνοψίζονται στο χαμηλό κόστος, την εύκολη προετοιμασία του δείγματος, αλλά και την ακρίβεια (Kchouk *et al*, 2017), ενώ πραγματοποιείται μαζική αλληλούχηση εκατομμυρίων τμημάτων DNA ταυτόχρονα. Με την κάλυψη του συνολικού γονιδιώματος ενός οργανισμού μέσω αλληλούχησης είναι δυνατή η ανίχνευση γενετικής ποικιλομορφίας σε ατομικό επίπεδο και σε επίπεδο πληθυσμών και ειδών.

Οι τεχνολογίες αλληλούχησης είναι οι πλέον κυρίαρχες ακόμα και για τη μελέτη του μεταγραφώματος ενός οργανισμού, καθώς και του επιπέδου έκφρασης των γονιδίων του σε διαφορετικούς κυτταρικούς πληθυσμούς. Το μεταγράφημα ενός οργανισμού είναι το σύνολο των μορίων RNA των κυττάρων του και η αλληλούχηση του RNA, συνοψίζεται στην αντιγραφή των μορίων RNA σε cDNA (Su *et al*, 2014) Κατά την RNA αλληλούχηση ανιχνεύονται οι αλληλουχίες του ώριμου mRNA που αποτελούν τα εξόνια του γονιδίου, αλλά οι και αμετάφραστες περιοχές στα 3' και 5' άκρα κάθε γονιδίου. Έτσι, μετά από συναρμολόγηση του μεταγραφώματος και αντιστοίχισή του στις περιοχές του γονιδιώματος αναφοράς αποκαλύπτονται οι αλληλουχίες που αποτελούν τα γονίδια. Ανιχνεύονται οι νουκλεοτιδικές βάσεις του γονιδίου που αντιστοιχούν στα εξόνια, αλλά και οι μη κωδικοποιητικές περιοχές, τα ιντρόνια, που δε μεταγράφονται και δε συμπεριλαμβάνονται στο αποτέλεσμα της RNA

¹ Ο όρος αλληλούχηση αναφέρεται στον καθορισμό της ακριβούς σειράς των νουκλεοτιδικών βάσεων σε έναν κλώνο DNA, είτε πρόκειται για ένα μόνο γονίδιο, είτε για το συνολικό γονιδίωμα ενός οργανισμού (Fritz, 2011).

αλληλούχησης. Έτσι, αποκαλύπτονται τα όρια των εξονίων και των ιντρονίων, ενώ είναι δυνατή και η ανίχνευση εναλλακτικών μεταγραφών (Su *et al*, 2014). Με την ανάγκη για μεταγραφωμικές πληροφορίες, το μήκος των αναγνώσεων, καθώς και η ακρίβεια και η ευαισθησία της μεθόδου συνεχώς βελτιώνονται και πλέον είναι δυνατή ακόμα και η *in situ* RNA αλληλούχηση κατά την οποία το μεταγράψωμα ακόμα και ενός μόνο κυττάρου μπορεί να εξεταστεί σε δείγματα ιστών (Lee *et al*, 2014). Έτσι, είναι δυνατή η σύγκριση του μεταγραφώματος, όχι μόνο διαφορετικών ειδών, αλλά και διαφορετικών κυττάρων του ίδιου οργανισμού.

Οι νουκλεοτιδικές και αμινοξικές αλληλουχίες των οργανισμών που αποκωδικοποιούνται με τα εργαλεία αλληλούχησης, αποθηκεύονται σε ειδικές βάσεις δεδομένων, όπως αυτές της NCBI (NCBI, 1988). Οι βάσεις δεδομένων αυτές καλύπτουν σχεδόν κάθε είδους σχετική πληροφορία, από νουκλεοτιδικές, αμινοξικές αλληλουχίες και μοριακές δομές μέχρι φαινότυπο και βιοποικιλότητα. Επιπλέον, μερικές από τις βάσεις δεδομένων είναι εξειδικευμένες για ένα συγκεκριμένο είδος, ενώ άλλες παρέχουν τις γονιδιωματικές αλληλουχίες από πλήθος ειδών. Στην τελευταία κατηγορία ανήκουν και η Flybase (Thurmond *et al*, 2019) που ενσωματώνει πληροφορίες για έντομα που ανήκουν στην οικογένεια των δροσοφιλίδων και η Vectorbase για ασπόνδυλα που αποτελούν φορείς παθογόνων του ανθρώπου (Giraldo-Calderón *et al*, 2015). Έτσι, οι σχετικές πληροφορίες είναι διαθέσιμες οποιαδήποτε στιγμή για τους επιστήμονες για κάθε είδους γονιδιωματική, μεταγραφωμική και συγκριτική ανάλυση.

1.2. Συγκριτική γονιδιωματική

Με την αύξηση του αριθμού των διαθέσιμων γονιδιωματικών και μεταγραφωμικών αλληλουχιών η ανάγκη για αποκωδικοποίηση της δομής και της λειτουργίας τους έγινε πιο έντονη. Με τη σύγκριση των γονιδιωματικών πληροφοριών που διατίθενται για πλήθος ειδών είναι δυνατή η εύρεση των ομόλογων αλληλουχιών τους και η ανίχνευση κοινών λειτουργιών. Ο τομέας της συγκριτικής γονιδιωματικής παίζει σπουδαίο ρόλο στις συγκριτικές αυτές αναλύσεις (Zasha Weinberg, 2010). Μια ανάλυση συγκριτικής γονιδιωματικής μπορεί να επιτευχθεί σε ολόκληρα γονιδιώματα ή συντενικές² περιοχές διαφορετικών ειδών ή διαφορετικών πληθυσμών του ίδιου είδους και αποκαλύπτει τις σχέσεις τους, τις ομοιότητες και τις διαφορές τους, τις συντηρημένες και τις ειδο-ειδικές αλληλουχίες (Liping Wei, 2002), αφού δίνει τη δυνατότητα σύγκρισης των γονιδιωμάτων σε τρία επίπεδα. Σε επίπεδο νουκλεοτιδίων για τη μελέτη γενικών στατιστικών στοιχείων σχετικά με τη συχνότητα κάθε νουκλεοτιδίου, σε επίπεδο DNA για την ανάλυση της δομής του γονιδιώματος και την ταυτοποίηση συντηρημένων, συντενικών περιοχών και σε επίπεδο γονιδίων για την ανάλυση της δομής του μεταγραφώματος. Τελικά, με σύγκριση γονιδιωμάτων στα τρία αυτά επίπεδα παρέχονται πληροφορίες σχετικά με την οργάνωση, τη συντήρηση και την εξέλιξη των γονιδίων τους,).

2 Συντενικές είναι οι περιοχές στο γονιδίωμα διαφορετικών ειδών που εμφανίζουν μεγάλη ομοιότητα στην αλληλουχία τους και στη σειρά με την οποία ανιχνεύονται τα γονίδια στις περιοχές αυτές (Myers, 2008) και διευκολύνουν την ανίχνευση των συντηρημένων ομόλογων αλληλουχιών και συνεπώς και των ειδο-ειδικών (Liping *et al*, 2002).

Στις μελέτες συγκριτικής γονιδιωματικής χρησιμοποιούνται διάφορα υπολογιστικά εργαλεία όπως οι αλγόριθμοι πολλαπλής στοίχισης των αλληλουχιών που συγκρίνονται και οι στρατηγικές απεικόνισης των σχέσεων τους μέσα από φυλογενετικές αναλύσεις. Τα εργαλεία αυτά επιτρέπουν την ανάλυση της εξελικτικής πορείας κάθε γονιδίου (Frazer *et al*, 2004). Η στοίχιση των αλληλουχιών γίνεται με την τοποθέτηση τους τη μία κάτω από την άλλη ανάλογα με τα μοτίβα τους. Τα κοινά μοτίβα βρίσκονται το ένα κάτω από το άλλο στις συγκρινόμενες αλληλουχίες, ενώ οι περιοχές που δεν ανιχνεύονται σε κάποια από τις αλληλουχίες παρουσιάζονται ως κενά. Η στοίχιση των αλληλουχιών επιτρέπει την ανίχνευση των κοινών δομικών τους μοτίβων και των διαφορών τους (Dunn CW, 2013). Επιπλέον, έπειτα από μια φυλογενετική ανάλυση αποκαλύπτεται η φυλογενετική απόσταση μεταξύ τους. Η φυλογενετική απόσταση αντιστοιχεί στο βαθμό στον οποίο διαφέρουν οι αλληλουχίες ή τα είδη και στις διαφορές της εξελικτικής τους πορείας. Όσο μικρότερη η φυλογενετική απόσταση, τόσο μεγαλύτερη ομοιότητα θα ανιχνευθεί μεταξύ των αλληλουχιών ή ειδών (Touchman, 2010). Έτσι, τα ομόλογα³ γονίδια ανιχνεύονται σε μικρή απόσταση στο φυλογενετικό δέντρο και επιβεβαιώνονται πιθανές ομοιότητες στη λειτουργία τους.

Ωστόσο, για να προκύψει όσο το δυνατόν πιο ακριβές αποτέλεσμα μέσα από μια συγκριτική ανάλυση, είναι απαραίτητο οι νουκλεοτιδικές και αμινοξικές αλληλουχίες που συγκρίνονται να είναι πλήρως επιμελημένες. Γι' αυτό οι κλάδοι της γονιδιωματικής και της συγκριτικής γονιδιωματικής απασχολούν πλήθος βιοπληροφορικών εργαλείων για τη μελέτη και επιμέλεια της δομής των γονιδιωμάτων (Narayana P.S., 2016)

1.3. Βιοπληροφορική ανάλυση

Η βιοπληροφορική είναι ένας τομέας που χρησιμοποιεί την υπολογιστική ισχύ, μαθηματικούς υπολογισμούς και στατιστική στη μοριακή βιολογία για την απόκτηση και ανάλυση πλήθους βιολογικών δεδομένων (Abouelhoda *et al*, 2008). Χαρακτηριστικά παραδείγματα βιοπληροφορικών αναλύσεων αφορούν μεταξύ άλλων τη στοίχιση του γονιδιώματος με το αντίστοιχο μεταγράψωμα και συνεπώς τη χαρτογράφηση και πρόβλεψη των γονδιακών μοντέλων (Luscombe *et al*, 2001). Τα γονδιακά μοντέλα μπορούν να προβλεφθούν μέσα τις ανεπτυγμένες υπολογιστικές μεθόδους ανάλυσης αλληλουχιών. Οι μέθοδοι αυτοί είναι αυτοματοποιημένες γιατί τα περισσότερα γονιδιώματα είναι πολύ μεγάλα και είναι επιθυμητό να επιμεληθούν όσο γρηγορότερα και όσα περισσότερα είναι δυνατόν. Η πλέον επικρατούσα αυτοματοποιημένη μέθοδος ανάλυσης αλληλουχιών είναι οι αλγόριθμοι πρόβλεψης των γονδιακών μοντέλων (McDonnell E., 2018). Η πρόβλεψη των γονδιακών μοντέλων από τους αλγορίθμους πραγματοποιείται με την αναγνώριση των εκφυλισμένων περιοχών αρχής και τέλους, τα κωδικόνια έναρξης και λήξης (Narayana P.S., 2016). Ωστόσο, οι σύγχρονοι αλγόριθμοι είναι επιρρεπείς σε λάθη και πραγματοποιούν εσφαλμένες προβλέψεις. Για παράδειγμα, πολλές φορές παραλείπουν γονίδια, ενώ σε πολλές περιπτώσεις συνδέουν ή χωρίζουν γονδιακά μοντέλα. Έτσι, είναι απαραίτητη η χειροκίνητη επιμέλεια των γονδιακών μοντέλων ώστε ο λειτουργικός τους σχολιασμός που θα ακολουθήσει να είναι ακριβής.

³ Ομόλογα ονομάζονται τα γονίδια που μοιράζονται έναν κοινό πρόγονο και οι νουκλεοτιδικές ή αμινοξικές τους αλληλουχίες είναι σχεδόν όμοιες, υπονοώντας και όμοια λειτουργία.

- Χειροκίνητη επιμέλεια γονιδιακών μοντέλων

Με τη χειροκίνητη επιμέλεια των νέων γονιδιακών μοντέλων είναι δυνατή η παραγωγή πιο ακριβών και αξιόπιστων γονιδιακών μοντέλων. Χειροκίνητη επιμέλεια μπορεί να επιτευχθεί σε διαδικτυακές βιοπληροφορικές πλατφόρμες που έχουν δημιουργηθεί για αυτόν τον σκοπό. Οι πλατφόρμες αυτές επιτρέπουν την επιμέλεια των συναινετικών γονιδιακών μοντέλων των αλγορίθμων πρόβλεψης, προσφέροντας υπολογιστικά και πειραματικά αποδεικτικά δεδομένα ώστε η επιμέλεια να γίνεται με ακρίβεια. Επιπλέον, ενθαρρύνουν τη συλλογική προσπάθεια και επιτρέπουν την πρόσβαση σε κάθε εργαστήριο της επιστημονικής κοινότητας

Ένα παράδειγμα πλατφόρμας επιμέλειας γονιδιακών μοντέλων είναι και το Apollo με την ανάπτυξη της πρωτοβουλίας i5k (“5000 arthropod genomes initiative”). Στην πλατφόρμα του Apollo φιλοξενούνται 35 είδη αρθρόποδων, από τα οποία τα 29 ανήκουν στην κλάση των εντόμων. Για κάθε ένα από τα είδη που φιλοξενεί, παρέχεται το συνολικό γονιδίωμα και τουλάχιστον ένα ζευγάρι συναινετικών γονιδιακών μοντέλων σε συνδυασμό με πειραματικά και υπολογιστικά αποδεικτικά στοιχεία. Ένα από τα αποδεικτικά στοιχεία για τη διευκόλυνση της διαδικασίας της επιμέλειας είναι και τα δεδομένα RNA αλληλούχησης που είναι διαθέσιμα για κάθε είδος και προσφέρουν πληροφορίες σχετικά με την παρουσία ή μη γονιδίων στα διάφορα ικριώματα του γονιδιώματος.

Στόχος της πρωτοβουλίας i5k είναι ο συντονισμός της αλληλούχησης του γονιδιώματος 5000 εντόμων ή συγγενικών ειδών αρθρόποδων, ώστε τα γονιδιώματα τους να είναι διαθέσιμα και σωστά επιμελημένα για την επιστημονική κοινότητα (Poelchau M, 2015). Οι πληροφορίες αυτές, στη συνέχεια, θα μπορούν να χρησιμοποιηθούν σε μελλοντικές έρευνες για την βαθύτερη κατανόηση της βιολογίας των εντόμων υγειονομικής και οικονομικής σημασίας.

1.4. Συνεισφορά της γονιδιωματικής στον τομέα της γεωργίας

Η συνεισφορά της γονιδιωματικής είναι ιδιαίτερα σημαντική για τη βαθύτερη κατανόηση της βιολογίας των εντόμων υγειονομικής και οικονομικής σημασίας. Στην πρώτη περίπτωση περιλαμβάνονται έντομα-φορείς παρασίτων υπεύθυνων για την πρόκληση ασθενειών στον άνθρωπο, ενώ τα έντομα της δεύτερης περίπτωσης προκαλούν σοβαρές καταστροφές στην αγροτική παραγωγή. Μέχρι σήμερα, οι μέθοδοι που χρησιμοποιούνται για την αντιμετώπιση τους περιορίζονται στη χρήση εντομοκτόνων ή μαζικών συστημάτων παγίδευσης. Ωστόσο, οι μέθοδοι αυτοί συχνά επιβαρύνουν την υγεία του ανθρώπου και το περιβάλλον ή υστερούν σε αποτελεσματικότητα. Επιπρόσθετα, η έντονη χρήση των εντομοκτόνων έχει οδηγήσει, σε πολλές περιπτώσεις, σε ανθεκτικότητα των εντόμων προς αυτά, μειώνοντας έτσι την αποτελεσματικότητά τους ως προς τον έλεγχο των ζημιογόνων εντόμων.

Εκτός από τα ένζυμα που καθοδηγούν την ανθεκτικότητα στα έντομα, μέσω της αξιοποίησης της γονιδιωματικής είναι, επίσης, δυνατόν να ταυτοποιηθούν τα γονίδια που εμπλέκονται σε ζωτικά συστήματα επιβίωσης και αναπαραγωγής παρασιτικών εντόμων. Επιπλέον, με τη χρήση εργαλείων συγκριτικής γονιδιωματικής δίνεται η δυνατότητα διάκρισης των ειδο-ειδικών από τα συντηρημένα γονίδια. Τα ειδο-ειδικά γονίδια συχνά εμπλέκονται σε μοναδικούς μοριακούς μηχανισμούς που καθοδηγούν τα έντομα προς συγκεκριμένα υποστρώματα τροφής και ωοτοκίας και η αναγνώριση

και στόχευση τους μπορεί να προσφέρει σημαντικές πλεονεκτήματα για την αντιμετώπιση του εντόμου.

1.5. Μοριακοί μηχανισμοί που καθοδηγούν τα έντομα – Χημειοαντίληψη

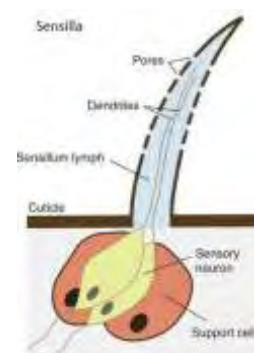
Τα συστήματα χημειοαντίληψης παίζουν κρίσιμο ρόλο στη συμπεριφορά των εντόμων, καθοδηγώντας τα προς την εύρεση συντρόφων, τις επιθυμητές πηγές τροφής, τους κατάλληλους ξενιστές και τα κατάλληλα υποστρώματα για ωοτοκία μέσω της αντίληψης χημικών ενώσεων, πτητικών και μη. Στα συστήματα χημειοαντίληψης ανήκουν το γευστικό και οσφρητικό σύστημα και οι πρωτεΐνες τους.

Στα έντομα, οι πρωτεΐνες που είναι υπεύθυνες για την ερμηνεία των οσφρητικών και γευστικών ερεθισμάτων χωρίζονται σε τέσσερις οικογένειες (Paranicolaou *et al*, 2016):

Τους οσφρητικούς υποδοχείς (Olfactory Receptors, ORs) που δεσμεύουν τις πτητικές ενώσεις.

1. Τις οσμοδεσμευτικές πρωτεΐνες που προσδένουν τις οσμές (Odorant Receptors, OBPs) και τις μεταφέρουν στους οσφρητικούς υποδοχείς.
2. Τους γευστικούς υποδοχείς (Gustatory Receptors, GRs) που αναγνωρίζουν τα σάκχαρα, τις πικρές ενώσεις, αλλά και το διοξειδίο του άνθρακα.
3. Τους ιονοτροπικούς υποδοχείς που συμμετέχουν κυρίως στην όσφρηση, με την απόκριση σε αμίνες, οξέα και άλλες ενώσεις, αλλά ένα ποσοστό αυτών σχετίζεται με τη γεύση.

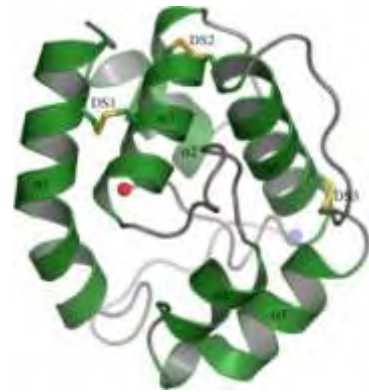
Οι χημειούποδοχείς εκφράζονται στην επιφάνεια των δενδριτών των νευρώνων που εδράζονται στα αισθητήρια τριχίδια (“sensilla”) των οσφρητικών και γευστικών οργάνων. Οι οσφρητικοί και οι ιονοτροπικοί υποδοχείς βρίσκονται σε διαφορετικούς τύπους τριχιδίων στις κεραίες και τις γναθικές προσακτρίδες που φέρουν πολλούς πόρους (Missbach *et al*, 2014). Αντίθετα, οι γευστικοί υποδοχείς ανιχνεύονται σε τριχίδια με έναν πόρο στις προσακτρίδες, τα πόδια, τα φτερά, ακόμα και τον ωοθέτη (Joseph *et al*, 2015), ενώ παρατηρούνται ακόμα και στο φάρυγγα και την κορυφή της προσακτρίδας (Vosshall *et al*, 2007). Η διέγερση των οσφρητικών και γευστικών νευρώνων από τις οσμές, τις φερομόνες και τις λοιπές ενώσεις θα μεταφερθεί στο κεντρικό τμήμα του εγκεφάλου όπου θα ερμηνευθεί. Κάθε μία από τις οικογένειες των παραπάνω πρωτεϊνών αναλύονται παρακάτω.



Εικόνα 1.1: Αισθητήριο τριχίδιο. Από τους πόρους εισέρχεται η ένωση και μεταφέρεται στον αντίστοιχο νευρώνα μέσω των χημειούποδοχέων.

1.5.1. Οσμοδεσμευτικές πρωτεΐνες (Odorant-Binding Proteins, OBPs)

Οι οσμοδεσμευτικές πρωτεΐνες (Odorant-Binding Proteins, OBPs) είναι μικρές, σφαιρικές και υδατοδιαλυτές πρωτεΐνες που ανιχνεύονται στους οσφρητικούς νευρώνες και προσδένουν εξειδικευμένα τις πτητικές ενώσεις (Isono *et al*, 2010). Οι περισσότερες από αυτές τις πρωτεΐνες εντοπίζονται στη λέμφο των τριχιδίων των κεραίων. Παρατηρούνται στο άκρο της προσακτρίδας, στα πόδια, στο έντερο, στα αναπαραγωγικά όργανα των αρσενικών, ακόμα και σε ιστούς που δεν ανήκουν στα οσφρητικά και γευστικά όργανα (Sun *et al*, 2018). Οι OBPs



Εικόνα 1.2.: Τρισδιάστατη δομή μιας οσμοδεσμευτικής πρωτεΐνης

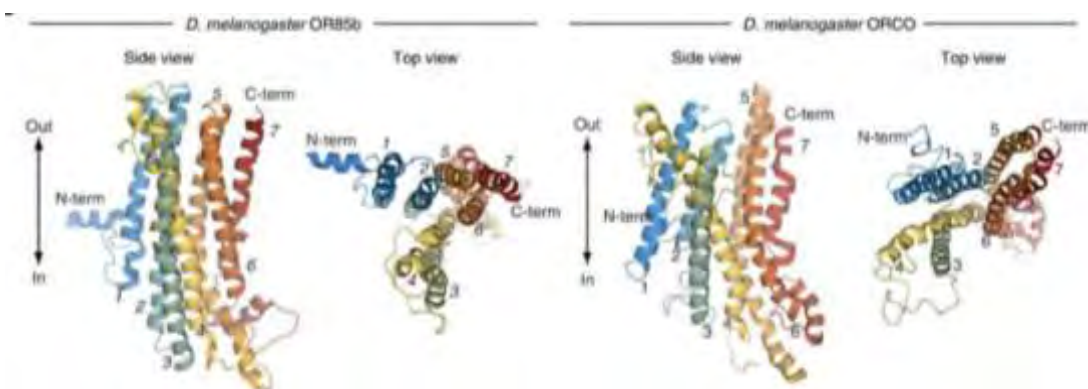
αποτελούνται από 120-150 αμινοξέα οργανωμένα σε έξι α-έλικες (Xu *et al*, 2009) και κατηγοριοποιούνται, σύμφωνα με τα μοτίβα κυστεϊνών τους, στις παρακάτω τέσσερις υπο-οικογένειες (Campanini *et al*, 2016)

1. *Classic OBPs*: Οι classic OBPs περιέχουν ένα μοτίβο έξι κυστεϊνών, το οποίο ανιχνεύεται ως εξής στα δίπτερα: C1-X₂₁₋₆₈-C2-X₃-C3-X₂₁₋₄₆-C4-X₈₋₂₈-C5-X₈₋₉-C6, όπου το X αντιστοιχεί σε οποιοδήποτε αμινοξύ και ο δείκτης του στον αριθμό των αμινοξέων που παρεμβάλλονται μέχρι την επόμενη κυστεΐνη (Xu *et al*, 2009).
2. *Plus-C OBPs*: Σε αντίθεση με τις classic OBPs, οι plus-C έχουν παραπάνω από έξι κυστεΐνες και μια προλίνη. Τα μοτίβα Plus-C έχουν ανιχνευτεί στο είδος *Bombyx Mori* ως: C1 X₈₋₄₁ C2 X₃ C3 X₃₉₋₄₇ C4 X₁₇₋₂₉ C4a X₉ C5 X₈ C6 X₉₋₁₁ C6a (Zhou *et al*, 2009). Αντίθετα, στο είδος *Drosophila melanogaster* εκτός από τις 6 συντηρημένες κυστεΐνες ανιχνεύονται και 3 επιπλέον κυστεΐνες και μια προλίνη ως εξής: C6-X₈-C6a-P-X₁₀₋₁₁-C6b-X₉-C6c-X₃₋₅₀ (Hekmat-Scafe *et al*, 2002)
3. *Minus-C OBPs*: Οι minus-C OBPs διαφέρουν από τις classic στο ότι περιέχουν μόνο τέσσερις συντηρημένες κυστεΐνες. Συγκεκριμένα, τα μοτίβα των OBPs της υπο-οικογένειας Minus-C έχουν ανιχνευτεί στη *Drosophila melanogaster* όπως και στις classic OBPs με τη διαφορά ότι στις περισσότερες απουσιάζουν η 2^η και η 5^η κυστεΐνη (C2 και C5) (Hekmat-Scafe *et al*, 2002).
4. *Dimer OBPs*: Τέλος, οι διμερείς OBPs αποτελούνται από δύο μοτίβα έξι κυστεϊνών, τα οποία πιθανότατα προέκυψαν από την ένωση δύο OBP της υπο-οικογένειας Classic (Jiang *et al*, 2017)

1.5.2. Οσφρητικοί υποδοχείς (Odorant Receptors, ORs)

Το οσφρητικό σύστημα των εντόμων ανιχνεύει τις χημικές ενώσεις στο περιβάλλον μέσω οσφρητικών υποδοχέων (Odorant Receptors, ORs) που φέρουν δομή επτά διαμεμβρανικών περιοχών. Ωστόσο δεν ανήκουν στην οικογένεια των GPCRs

(Υποδοχείς προσδεμένοι με G- πρωτεΐνη), όπως οι ORs των θηλαστικών. Επίσης, εκτός από την απουσία ομολογίας με τους GPCRs, παρουσιάζουν και διαφορετική τοπολογία, καθώς τα άκρα τους είναι ανεστραμμένα με το αμινοτελικό τους άκρο να βρίσκεται στο εσωτερικό της μεμβράνης και το καρβοξυτελικό στο εξωτερικό (Leal WS, 2012). Οι ORs εντοπίζονται στις κεραίες και τις γναθικές προσακτρίδες, πάντα ως ετεροδιμερή που συγκροτούνται από τις εξής πρωτεΐνες: α) τον συνυποδοχέα Orco (Odorant receptor-coreceptor) με λειτουργία πρωτεΐνης-συνοδού για τη σωστή αναδίπλωση των πρωτεϊνών και β) τον ειδικό υποδοχέα για τη συγκεκριμένη οσμή (Hopf *et al*, 2015)



Εικόνα 1.4.: Τρισδιάστατη δομή του συν-υποδοχέα Orco στη *Drosophila melanogaster* (Hopf *et al*, 2015). Διακρίνονται οι 7 διαμεμβρανικές περιοχές του, που εμφανίζουν μεγάλη συντήρηση από είδος σε είδος, καθώς και η τοπολογία του καρβοξυτελικού (C-term) και αμινοτελικού (N-term) άκρου, που είναι κοινή και για τους υπόλοιπους οσφρητικούς υποδοχείς.

Η οικογένεια των οσφρητικών πρωτεϊνών μπορεί να διαχωριστεί σε επιμέρους υπο-οικογένειες λόγω της σημαντικής ποικιλομορφίας τους στις αλληλουχίες των διαμεμβρανικών περιοχών. Οι διαμεμβρανικές περιοχές των οσφρητικών υποδοχέων εμφανίζουν μη συντηρημένες αλληλουχίες που παρατηρούνται κυρίως σε επικράτειες της τρίτης, τέταρτης και πέμπτης διαμεμβρανικής περιοχής. Η ποικιλομορφία στις περιοχές αυτές πιθανότατα αντικατοπτρίζει το ρόλο τους στην αναγνώριση διακριτών πτητικών ενώσεων (Buck *et al*, 1991)

1.5.3. Γευστικοί υποδοχείς (Gustatory Receptors, GRs)

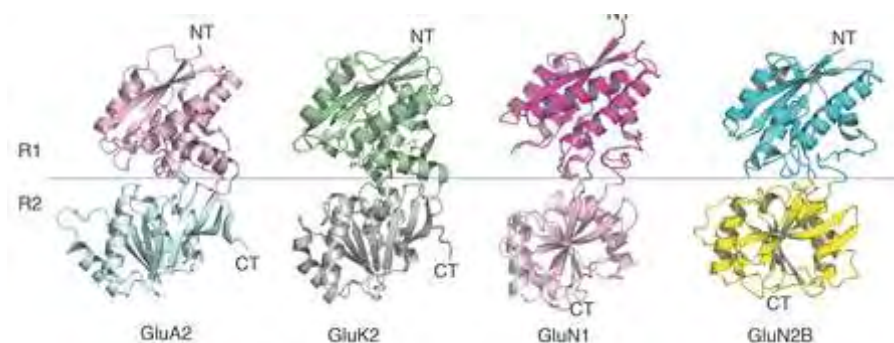
Οι γευστικοί υποδοχείς (Gustatory Receptors, GRs) είναι υπεύθυνοι για την ανίχνευση σακχάρων, πικρών ενώσεων και αλάτων και ο ρόλος τους περιορίζεται στις διατροφικές αποφάσεις. Στα έντομα το ρεπερτόριο των γευστικών υποδοχέων επεκτείνεται και στην πρόσδεση φερομονών και άλλων διεγερτικών ενώσεων, όπως τα οργανικά και ανόργανα οξέα και το διοξείδιο του άνθρακα, με ρόλο στην ωοτοκία, την εύρεση συντρόφου και τροφής, αλλά και τις διατροφικές αποφάσεις (Sánchez-Alcañiz *et al*, 2018). Δίνουν, έτσι, τη δυνατότητα στο έντομο να εντοπίζει ιδανικές πηγές τροφής και τα κατάλληλα υποστρώματα για ωοαπόθεση (Isono *et al*, 2010).

Όπως και οι οσφρητικοί υποδοχείς, οι γευστικοί υποδοχείς αποτελούνται από επτά διαμεμβρανικές περιοχές με την ίδια τοπολογία. (Isono *et al*, 2010). Ωστόσο, οι GRs δεσμεύουν μη πτητικές ενώσεις με τη μόνη εξαίρεση πτητικής ένωσης να είναι το

διοξειδίο του άνθρακα. Ο τρόπος με τον οποίο θα αποκριθεί το έντομο στη διέγερση από μια μη πτητική ένωση εξαρτάται από το συνδυασμό των GRs που διεγείρονται και από την περιοχή του σώματος του εντόμου στην οποία αυτοί βρίσκονται (Vosshall *et al*, 2007). Ωστόσο, η εξειδίκευση του κάθε γευστικού υποδοχέα παραμένει άγνωστη και οι γνώσεις σχετικά με αυτούς περιορίζονται στα μορφολογικά και φυσιολογικά χαρακτηριστικά τους χωρίς να επεκτείνονται σε μοριακό επίπεδο. Έτσι και τα μοτίβα των διαμεμβρανικών περιοχών που σηματοδοτούν τις διακριτές αποκρίσεις στις μη πτητικές ενώσεις παραμένουν άγνωστα, ενώ έχει παρατηρηθεί ότι πολλοί από τους γευστικούς υποδοχείς, όπως αυτοί του είδους *Bombyx mori*, δεν εμφανίζουν την αναμενόμενη τοπολογία των επτά διαμεμβρανικών περιοχών (Zhang *et al*, 2011)

1.5.4. Ιονοτροπικοί υποδοχείς (Ionotropic Receptors, IRs)

Οι ιονοτροπικοί υποδοχείς (Ionotropic Receptors, IRs) ανιχνεύουν τόσο πτητικές όσο και μη πτητικές ενώσεις, συμβάλλοντας στην όσφρηση αλλά και στη γεύση. Αποτελούνται από ένα προσδετο-εξαρτώμενο κανάλι τριών διαμεμβρανικών περιοχών και φέρουν ένα αμινοτελικό άκρο στο εξωτερικό της μεμβράνης και ένα βραχύ καρβοξυτελικό άκρο στο εσωτερικό (Gomez-Diaz *et al*, 2018).



Εικόνα 1.7.: Τρισδιάστατη δομή ιονοτροπικού υποδοχέα γλουταμινικού (iGluRs) που ομοιάζει με τους IRs (Furukawa, 2012). Διακρίνονται οι 3 διαμεμβρανικές περιοχές, καθώς και η τοπολογία του αμινοτελικού (NT) και του καρβοξυτελικού (CT) άκρου.

Αν και οι περισσότεροι ιονοτροπικοί υποδοχείς ανιχνεύονται στα οσφρητικά όργανα, εμφανίζουν σημαντικές διαφορές με τους οσφρητικούς υποδοχείς. Οι ιονοτροπικοί υποδοχείς είναι λιγότερο ευαίσθητοι από τους ORs και προσαρμόζονται στις οσμές καθυστερημένα. Η κυριότερη διαφορά, ωστόσο, είναι ότι αναγνωρίζουν και δεσμεύουν διαφορετικούς προσδέτες, με τους ιονοτροπικούς να εμφανίζουν μεγαλύτερη ευαισθησία προς αμίνες και οξέα (Rytz *et al*, 2013). Η οικογένεια των IRs χωρίζεται σε δύο υπο-οικογένειες:

α) Τους ιονοτροπικούς που βρίσκονται στις κεραίες (antennal ionotropic receptors) και είναι οι καλύτερα χαρακτηρισμένοι. Η υπο-οικογένεια αυτή συμπεριλαμβάνει ένα μικρό αριθμό των ιονοτροπικών υποδοχέων οι οποίοι είναι πολύ συντηρημένοι ανάμεσα στα διάφορα είδη εντόμων, είναι οσφρητικοί και εντοπίζονται μόνο στις κεραίες. Οι περισσότεροι έχουν χαρακτηριστεί ως συν-υποδοχείς, αφού συχνά συν-εκφράζονται μαζί με άλλους IRs (Sánchez-Alcañiz *et al*, 2018).

β) Τους ιονοτροπικούς που αποκλίνουν από είδος σε είδος (αποκλίνοντες ή «divergent» ιονοτροπικοί υποδοχείς). Στην κατηγορία αυτή ανήκει η πλειονότητα των ιονοτροπικών υποδοχέων, για τους οποίους δεν εντοπίζονται σαφείς ορθόλογες σχέσεις από είδος σε είδος και εμφανίζουν μεγάλη εξειδίκευση σε κάθε είδος (Paranicolaou *et al*, 2016). Σε αυτή την κατηγορία εντοπίζονται και ιονοτροπικοί υποδοχείς που δε σχετίζονται με την όσφρηση, αλλά ο ρόλος τους εστιάζεται στη γεύση (Croset *et al*, 2010)

1.6. Εξέλιξη των γονιδίων χημειοαντίληψης στα έντομα

Στα έντομα παρατηρούνται υψηλά επίπεδα ποικιλομορφίας στον αριθμό και τη δομή των γονιδίων χημειοαντίληψης. Μέχρι σήμερα έχουν προταθεί τρία διαφορετικά μοντέλα για την περιγραφή της εξέλιξης των οικογενειών γονιδίων, αλλά το πιο πρόσφατο και ολοκληρωμένο είναι το μοντέλο «γέννησης-θανάτου» (Nei *et al*, 2006).

1.6.1. Εξελικτικό μοντέλο «Γέννησης και Θανάτου»

Σύμφωνα με το μοντέλο «γέννησης και θανάτου» οι κύριοι παράγοντες που πιθανότατα συνεισφέρουν στην εξέλιξη των γονιδιακών οικογενειών είναι τα γεγονότα διπλασιασμών και διαγραφών (Zhao *et al*, 2015). Οι διπλασιασμοί γονιδιωματικών αλληλουχιών συμβαίνουν μέσω άνισου ανασυνδυασμού, κατά τον οποίο ομόλογες αλληλουχίες που εντοπίζονται σε διαφορετικές θέσεις σε ομόλογα χρωμοσώματα ανασυνδυάζονται τυχαία. Η διαδικασία αυτή πραγματοποιείται κατά την αντιγραφή και επιδιόρθωση του DNA και οδηγεί στο διπλασιασμό της αλληλουχίας στο ένα χρωμοσώμα (Hastings *et al*, 1998). Διπλασιασμός μπορεί, επίσης, να συμβεί μέσω ρετρομετάθεσης, κατά την οποία μεταθετά στοιχεία αντιγράφονται ή αποκόπτονται και μεταφέρονται σε άλλη περιοχή στο DNA (Magadum *et al*, 2013). Η περίπτωση αυτή του διπλασιασμού αναγνωρίζεται από την απουσία ιντρονίων (Lesk *et al*, 2017)

Διπλασιασμοί μπορούν να συμβούν και σε αλληλουχίες υπαρχόντων γονιδίων, δίνοντας «γέννηση» σε νέα γονίδια τα οποία μπορούν να ακολουθήσουν τρεις διαφορετικές πορείες (βλ. Εικόνα 1.9):

Πορεία 1: Τα διπλασιασμένα γονίδια αποκλίνουν και αποκτούν διαφορετική λειτουργία (παράλογα γονίδια).

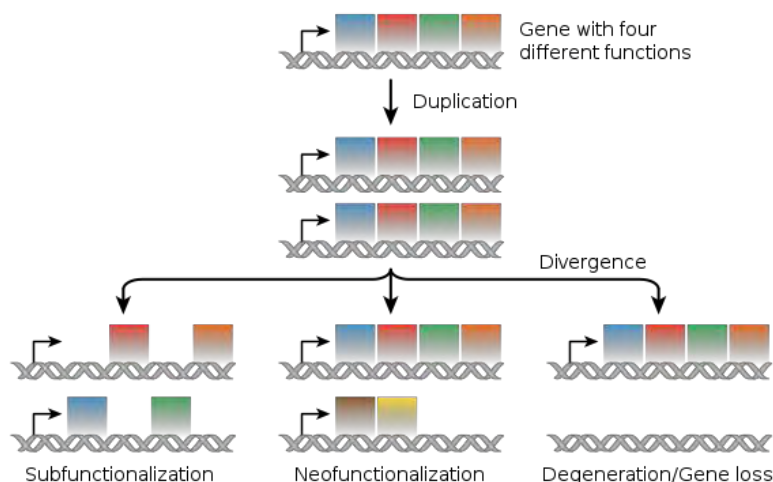
Πορεία 2: Το ένα από τα δύο γονίδια αποκτά καινούργια λειτουργία.

Πορεία 3: Παρατηρείται απώλεια του ενός από τα δύο γονίδια μέσω μετατροπής τους σε ψευδογονίδιο⁴ («θάνατος»).

Συνολικά, το μοντέλο «γέννησης και θανάτου» προβλέπει μια χαλάρωση της εξελικτικής πίεσης μετά το διπλασιασμό (Brand *et al*, 2017) που επιτρέπει τη

⁴ Τα ψευδογονίδια είναι εκφυλισμένα γονίδια που έχουν μεταλλαχθεί σε τόσο διαφορετικές μορφές σε σχέση με τις αρχικές αλληλουχίες, ώστε η πολυπεπτιδική αλληλουχία που κωδικοποιούν να μην είναι λειτουργική (Lesk, 2017).

διαφοροποίηση του ενός από τα δύο αντίγραφα του γονιδίου. Η διαφοροποίηση αυτή πραγματοποιείται υπό τη δράση της φυσικής επιλογής και μεταλλάξεων που πραγματοποιούνται τυχαία ή απαντούν σε ανάγκη για προσαρμογή (Suzuki *et al*, 2018).



Εικόνα 1.9: Εξελικτικό μοντέλο «γέννησης και θανάτου». Το γονίδιο που έχει υποστεί διπλασιασμό μπορεί να ακολουθήσει 3 διαφορετικές πορείες.

1.6.2. Διπλασιασμοί και απώλειες γονιδίων χημειοαντίληψης στα έντομα

Το μοντέλο «γέννησης και θανάτου» φαίνεται να είναι και ο κύριος εξελικτικός μηχανισμός των γονιδιακών οικογενειών στα έντομα, αφού πολλές από αυτές φαίνεται να προέκυψαν από διπλασιασμούς γονιδίων.

Σε έντομα που ανήκουν σε συγγενικά είδη, οι οικογένειες των γονιδίων χημειοαντίληψης χαρακτηρίζονται από παρόμοιο αριθμό γονιδίων και υπο-οικογενειών με ελάχιστες αποκλίσεις (Bhutkar *et al*, 2007). Ωστόσο, σε έντομα με μακρινή συγγένεια, μετά από σύγκριση (Sánchez-Gracia *et al*, 2009), βρέθηκε ότι το μέγεθος των γονιδιακών οικογενειών χημειοαντίληψης και ο αριθμός των συνολικών τους γονιδίων, διαφέρουν δραματικά. Ως αιτία έχει αναγνωριστεί το ότι τα διπλασιασμένα γονίδια εξελίσσονται ανεξάρτητα και οδηγούν σε μεγάλο ποσοστό γονιδιακών «γεννήσεων», απωλειών και ψευδογονιδίων. Οι απώλειες αυτές αποδεικνύονται από την ύπαρξη μεγάλου αριθμού μη λειτουργικών ψευδογονιδίων στα έντομα (Sánchez-Gracia *et al*, 2009).

- **Οικογένεια OBPs:** Συγκεκριμένα, για τις OBPs έχει παρατηρηθεί ότι γονίδια των υπο-οικογενειών classic, plus-C και minus-C σε έντομα της τάξης των διπτέρων και των κολεόπτερων έχουν υποστεί έναν αριθμό ειδο-ειδικών διπλασιασμών που οδήγησε σε αυξημένο αριθμό γονιδίων (Sánchez-Gracia *et al*, 2009). Ωστόσο, στα έντομα της τάξης των υμενόπτερων (*A.mellifera*) ανιχνεύονται μόνο λίγα μέλη γονιδίων στις υπο-οικογένειες classic και minus-

C, γεγονός που πιθανότατα οφείλεται σε απώλειες γονιδίων (Sánchez-Gracia *et al*, 2009).

- **Οικογένεια ORs:** Το ποσοστό των γεγονότων διπλασιασμών και απωλειών είναι ακόμα πιο υψηλό για τους χημειοϋποδοχείς των εντόμων σε σχέση με τις OBPs. Παραδείγματα διπλασιασμών σε γονίδια των ORs έχουν παρατηρηθεί στα είδη *Apis mellifera* (μέλισσα) (Robertson *et al*, 2006) και *Anopheles gambiae* (κουνούπι) (Brand *et al*, 2017). Αντίστοιχα, στη *D. Melanogaster* έχουν παρατηρηθεί 4 γεννήσεις και 12 απώλειες οσφρητικών υποδοχέων για τις οποίες οφείλονται διπλασιασμοί (McBride *et al*, 2009). Τα γεγονότα αυτά οδήγησαν σε σημαντική επέκταση της γονιδιακής οικογένειας των ORs και ο αριθμός των γονιδίων που ανήκουν σε αυτή ποικίλουν από είδος σε είδος. (Robertson *et al*, 2006).
- **Οικογένεια GRs:** Σε σχέση με τους οσφρητικούς υποδοχείς, η εξελικτική πίεση είναι σημαντικά πιο αδύναμη για τους γευστικούς υποδοχείς (GRs) και ειδικά για αυτούς που έχουν προέλθει από διπλασιασμούς. Έχει παρατηρηθεί ότι οι γευστικοί υποδοχείς που βρίσκονται κάτω από τη μικρότερη εξελικτική πίεση είναι των πικρών ενώσεων οι οποίοι εξελίσσονται γρηγορότερα σε σχέση με αυτούς των γλυκών ενώσεων και του διοξειδίου του άνθρακα (Amrein *et al*, 2005).
- **Οικογένεια IRs:** Γενικότερα, οι ιονοτροπικοί υποδοχείς εμφανίζουν υψηλή ποικιλομορφία από είδος σε είδος τόσο στον αριθμό των ενώσεων που ανιχνεύουν όσο και στο μέγεθος και τη δομή τους (Sánchez-Alcañiz *et al*, 2018). Ειδικότερα, οι IRs που συναντώνται στις κεραίες (antennal IRs) εμφανίζουν υψηλή συντήρηση ανάμεσα στα είδη και δεν έχει παρατηρηθεί κάποιος σημαντικός αριθμός διπλασιασμών ή απωλειών, ενώ η ανίχνευση ψευδογονιδίων είναι σπάνια. Στην οικογένεια των δροσοφιλίδων έχει ανιχνευθεί μόλις ένα ψευδογονίδιο και ένας διπλασιασμός (Croset *et al.*, 2010). Αντίθετα, στην υπο-οικογένεια των αποκλίνοντων IRs που δεν συναντώνται στις κεραίες (divergent IRs) η εξελικτική πίεση είναι αδύναμη και έχουν παρατηρηθεί πολλές περιπτώσεις απωλειών, ψευδογονιδίων και διπλασιασμών. Τα διπλασιασμένα γονίδια της υπο-οικογένειας αυτής συχνά δεν φέρουν κανένα ιντρόνιο και αποτελούνται από μόλις ένα εξόνιο, γεγονός που υποδηλώνει τη γέννηση του μέσω ρετρομετάθεσης (Croset *et al.*, 2010). Η ήπια εξελικτική πίεση που χαρακτηρίζει την υπο-οικογένεια των αποκλίνοντων IRs έχει ως αποτέλεσμα οι τελευταίοι να ποικίλλουν σημαντικά στην αλληλουχία τους και τον αριθμό τους από είδος σε είδος. Έτσι, δεν ανιχνεύονται ορθόλογες σχέσεις μεταξύ των διαφόρων ειδών. Η διαφοροποίηση τους στο κάθε είδος έχει υποτεθεί ότι παίζει ρόλο στην ανίχνευση με υψηλή εξειδίκευση συγκεκριμένων χημικών ενώσεων. Ωστόσο, για την πλειοψηφία των υποδοχέων αυτών δεν έχουν ανιχνευθεί οι συνδότες τους και η λειτουργία τους, ενώ οι μηχανισμοί που καθοδηγούν τη διαφοροποίηση τους παραμένουν άγνωστοι (Rytz R *et al*, 2013).

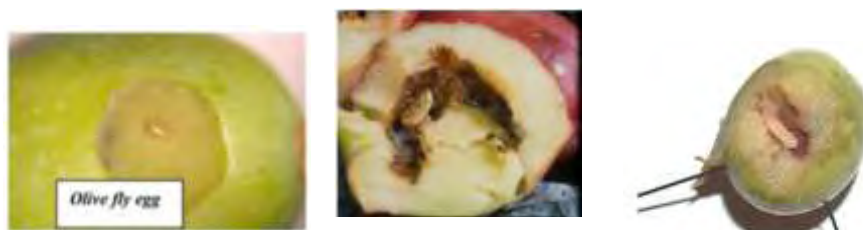
Από τα παραπάνω, προκύπτει ότι τα γονίδια χημειοαντίληψης διαφέρουν σημαντικά από είδος σε είδος, τόσο στην αλληλουχία τους, όσο και στον αριθμό τους (Wicher D., 2012). Οι διαφορές αυτές που προέκυψαν από διπλασιασμούς και απώλειες

γονιδίων βοηθούν στην προσαρμογή των εντόμων και καθοδηγούν τις μοναδικές προτιμήσεις του κάθε είδους. Για το λόγο αυτό οι διαφορές αυτές αποτελούν και το επίκεντρο γονιδιωματικών αναλύσεων που έχουν πραγματοποιηθεί γενικά σε διάφορες τάξεις εντόμων όπως στην ακρίδα (Wang *et al*, 2015), στο μυρμήγκι (Zhou *et al*, 2012), το κουνούπι-τίγρης (Zhou JJ *et al*, 2008), τη *D. melanogaster* (Ray A. *et al*, 2014), την *Bactrocera cucurbitae* (Elfekih *et al*, 2016), την *Bactrocera dorsalis* (Wu Z *et al*, 2015) και την μύγα της μεσογείου *Ceratitis capitata* (Papanicolaou *et al*, 2016).

Ιδιαίτερο ενδιαφέρον παρουσιάζει η ανάλυση και εξέλιξη των γονιδίων χημειοαντίληψης σε έντομα με εξειδικευμένες προτιμήσεις, όπως τα μονοφάγα έντομα. Στα έντομα αυτά η δράση της φυσικής επιλογής είναι εξαιρετικά αδύναμη και παρατηρούνται διακριτές διαφορές στη μορφολογία και τη συμπεριφορά τους σε σχέση με άλλα είδη εντόμων (Matsuo *et al*, 2007). Τα μονοφάγα έντομα επιλέγουν ως πηγή τροφής ή υπόστρωμα για ωοτοκία έναν μοναδικό καρπό και αποτελούν σημαντικούς εχθρούς της αγροτικής παραγωγής. Στην κατηγορία των μονοφάγων εντόμων ανήκει και ο δάκος της ελιάς, που επιλέγει ως μοναδικό υπόστρωμα τροφής και ωοτοκίας τους καρπούς της ελιάς, προκαλώντας μεγάλο ποσοστό καταστροφών. Ωστόσο, τα συστήματα που εμπλέκονται στη χημειοαντίληψη και τη γεύση δεν έχουν μελετηθεί επαρκώς. Σημαντική συνεισφορά αναμένεται να έχει προς την κατεύθυνση αυτή η ανάλυση του γονιδιώματος του και η ταυτοποίηση των γονιδίων που εμπλέκονται στις συγκεκριμένες βιολογικές λειτουργίες και καθοδηγούν τις μοναδικές προτιμήσεις του.

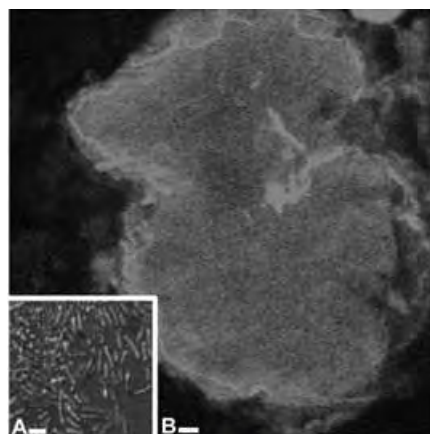
1.7. *Bactrocera oleae*

Ο δάκος της ελιάς ανήκει στα δίπτερα έντομα, στην οικογένεια των τεφριτίδων και στο είδος *Bactrocera oleae*. Κατά τη διάρκεια της ζωής του, σε διάστημα έξι μηνών, ο θηλυκός δάκος αφήνει 50-400 αυγά, συνήθως ένα σε κάθε καρπό (Rice, 2000). Με την εκκόλαψη του αυγού μέσα σε δύο με τρεις μέρες προκύπτει η προνύμφη (Nardi *et al*, 2003), η οποία θα τραφεί από τον καρπό, δημιουργώντας μια τρύπα σε αυτόν και καταστρέφοντας τον. Η μεταμόρφωση της προνύμφης πραγματοποιείται συνήθως στο εσωτερικό του καρπού (Rice, 2000).



Εικόνα 1.10.: Στάδια ανάπτυξης του δάκου της ελιάς. Από τα αριστερά προς τα δεξιά, φαίνονται στον καρπό της ελιάς στελέχη του δάκου στα στάδια ανάπτυξης του αυγού, της προνύμφης και της νύμφης.

Σε αντίθεση με άλλα δίπτερα έντομα, όπως η *C. capitata* ή η *D. melanogaster* τα οποία μπορούν να τρέφονται μόνο από υδρολυμένα συστατικά φρούτων, ο δάκος της ελιάς μπορεί να αξιοποιεί και τις μη υδρολυμένες πρωτεΐνες της ελιάς και άλλα θρεπτικά συστατικά του εξωκάρπιου (Pavlidí et al, 2017). Αυτό επιτυγχάνεται μέσω ενός συμβιώτη, του βακτηρίου *Candidatus dacicola erwinia* που εντοπίζεται στο έντερο, σε όλα τα αναπτυξιακά στάδια του δάκου και σε μεγαλύτερες συγκεντρώσεις στα θηλυκά κατά την ωοτοκία (Estes et al, 2018). Ο συμβιώτης *Ca. E. dacicola* συνεισφέρει στην αποτοξικοποίηση των φαινολικών ενώσεων της ελιάς που εμφανίζουν τοξικές επιδράσεις, όπως η ολεωρωπαΐνη που εντοπίζεται σε μεγάλες συγκεντρώσεις στις ελιές, ειδικότερα στις πράσινες που προτιμά ο δάκος (Andrews M. et al, 2003). Όταν η ολεωρωπαΐνη λαμβάνεται από το δάκο κατά το στάδιο της προνύμφης ο συμβιώτης εκφράζει αμινοξέα, όπως η γλυκίνη που αναστέλλουν τη δραστηριότητα της (Pavlidí et al, 2017).



Εικόνα: A: *Candidatus dacicola erwinia* μετά από FISH
B: Οισοφαγικός λοβός του δάκου της

Πρόσφατα ολοκληρώθηκε η αλληλούχηση του γονιδιώματος του δάκου της ελιάς εφαρμόζοντας τις τεχνολογίες σύγχρονες Illumina, PacBio και Nanopore από το εργαστήριο Μοριακής Βιολογίας και Γονιδιωματικής (Τμήμα βιοχημείας και Βιοτεχνολογίας, Πανεπιστήμιο Θεσσαλίας) και το McGill University and Genome Quebec Innovation Centre (Bayega et al, υπο δημοσίευση). Το γεγονός αυτό δίνει νέες προοπτικές για γονιδιωματικές αναλύσεις που θα αναδείξουν νέους στόχους για την αποτελεσματικότερη αντιμετώπισή του. Μέχρι σήμερα, έχει διεξαχθεί μικρός αριθμός γονιδιωματικών αναλύσεων για την ταυτοποίηση των γονιδίων που εμπλέκονται στην ανθεκτικότητα σε σύγχρονα εντομοκτόνα, όπως το *spinosad* (Sagri et al, 2014). Μέσα από παρόμοιες γονιδιωματικές μελέτες μπορούν να χαρακτηριστούν τα ένζυμα που είναι υπεύθυνα για την ανθεκτικότητα των εντόμων, ώστε να βελτιωθεί η κατανόηση των μηχανισμών που τα ρυθμίζουν και να αναπτυχθούν καλύτερα, πιο φιλικά προς το περιβάλλον και κυρίως, ειδο-ειδικά εντομοκτόνα.

Ωστόσο, καμία γονιδιωματική μελέτη δεν εστιάζει στην ανίχνευση των μοριακών μηχανισμών και των γονιδίων χημειοαντίληψης που καθοδηγούν το δάκο της ελιάς. Οι αλληλουχίες των γονιδίων αυτών δεν είναι καλά μελετημένες και τα γονιδιακά μοντέλα που προτείνονται έχουν προκύψει από επιρροές σε λάθη αλγορίθμους πρόβλεψης. Ως αποτέλεσμα οι δομές των γονιδίων, σε πολλές περιπτώσεις, είναι εσφαλμένες και οι μηχανισμοί στους οποίους εμπλέκονται παραμένουν άγνωστοι. Έτσι, η ταυτοποίηση των γονιδίων χημειοαντίληψης και το μοριακών μηχανισμών

που καθοδηγούν τις καταστροφικές του προτιμήσεις είναι μείζονος σημασίας για την αντιμετώπιση του.

1.8. Σκοπός εργασίας

Η παρούσα εργασία επικεντρώνεται στην επιμέλεια και την ανάλυση των γονιδίων χημειοαντίληψης του δάκου της ελιάς, τον σημαντικότερο εντομολογικό εχθρό των ελαιοκαλλιεργειών και συγκεκριμένα, στις οσμοδεσευτικές πρωτεΐνες (OBPs), τους γευστικούς υποδοχείς (GRs) και τους ιονοτροπικούς υποδοχείς (IRs) αξιοποιώντας τις γονιδιωματικές αλληλουχίες και μεταγραφωμικά δεδομένα. Οι πλήρως επιμελημένες αλληλουχίες χρησιμοποιήθηκαν για τον χαρακτηρισμό των συντηρημένων αμινοξικών τους μοτίβων και τον καθορισμό της τοπολογίας των διαμεμβρανικών τους περιοχών. Περαιτέρω συγκριτική ανάλυση των πρωτεϊνών που κωδικοποιούν τα υπό μελέτη γονίδια με ομόλογες των συγγενικών ειδών *C. capitata*, *D. melanogaster* και *M. domestica* αποσκοπούσε στην αποκάλυψη των εξελικτικών σχέσεων των γονιδίων αυτών και στην ταυτοποίηση δυνητικών γονιδιακών επεκτάσεων που πιθανότατα είναι ειδο-ειδικές και συνεπώς, η λειτουργία τους είναι άγνωστη. Τα γονίδια αυτά ίσως καθοδηγούν την προτίμηση του δάκου προς την ελιά και η ταυτοποίηση τους μπορεί να οδηγήσει στη δημιουργία νέων στόχων πιο ειδικών για την αντιμετώπιση του δάκου.

2. ΥΛΙΚΑ-ΜΕΘΟΔΟΙ

2.1. Επιμέλεια συναινετικών γονιδιακών μοντέλων των γονιδίων χημειοαντίληψης (Annotation)

2.1.1. Βάση δεδομένων της NCBI (National Center For Biotechnology Information)

Η βάση δεδομένων της NCBI δημιουργήθηκε το 1988 με την καθοδήγηση του David Lipman. Η ιστοσελίδα τους προσφέρει πληροφορίες συλλεγμένες από μια σειρά βάσεων δεδομένων σχετικές με βιοτεχνολογία και βιοϊατρική, ενώ ταυτόχρονα παρέχει σημαντικά εργαλεία βιοπληροφορικής για μελέτες σχετικά με τη δομή, τη λειτουργία και την εξέλιξη πρωτεϊνών.

Τα γονίδια του δάκου που αντιστοιχούν στις OBPs και στους γευστικούς και ιονοτροπικούς υποδοχείς και μελετώνται στην παρούσα εργασία λήφθηκαν από την NCBI.

2.1.2. Blast (Basic Local Alignment Search Tool)

Το BLAST (Altschul *et al*, 1990) αποτελεί έναν αλγόριθμο που αναπτύχθηκε από την NCBI και χρησιμοποιείται για τη σύγκριση βιολογικών αλληλουχιών, όπως οι αμινοξικές αλληλουχίες των πρωτεϊνών και οι νουκλεοτιδικές αλληλουχίες του DNA και του RNA. Μια αναζήτηση BLAST επιτρέπει την σύγκριση της επιθυμητής βιολογικής αλληλουχίας με τις αλληλουχίες μιας βιβλιοθήκης ή βάσης δεδομένων και την εύρεση των αλληλουχιών που φέρουν ένα συγκεκριμένο βαθμό ομοιότητας. Με αυτό τον τρόπο είναι δυνατός και ο εντοπισμός των ορθόλογων αλληλουχιών ανάμεσα σε διαφορετικά είδη ή των ομόλογων αλληλουχιών σε ένα είδος. Υπάρχουν τέσσερις διαφορετικοί τύποι BLAST. Η επιλογή του κατάλληλου από τους τέσσερις εξαρτάται από τον τύπο της επιθυμητής αλληλουχίας προς αναζήτηση και τον τύπο της αλληλουχίας που θα εξάγει τελικά ο αλγόριθμος. Στην περίπτωση που και η αρχική αλληλουχία και η αλληλουχία που επιθυμεί ο ερευνητής να εξαχθεί από το BLAST είναι νουκλεοτιδικές, τότε χρησιμοποιείται ο τύπος BLASTn. Αντίθετα, όταν η αρχική αλληλουχία και η αλληλουχία εξαγωγής είναι αμινοξικές επιλέγεται ο τύπος BLASTp του αλγόριθμου. Τέλος, αν η αρχική αλληλουχία είναι μεταφρασμένη νουκλεοτιδική και η αλληλουχία εξαγωγής είναι αμινοξική ή το αντίστροφο, επιλέγονται αντίστοιχα οι τύποι BLASTx ή tBLASTn.

Στην παρούσα εργασία με τη χρήση του αλγόριθμου BLASTp έγινε αναζήτηση των αμινοξικών αλληλουχιών του δάκου της ελιάς και εντοπίστηκαν οι αντίστοιχες ορθόλογες πρωτεΐνες στα είδη *Drosophila melanogaster* και *Musca domestica*.

2.1.3. Apollo – i5k Initiative

Η πρωτοβουλία «5000 arthropod genomes initiative» ή «i5k» θεμελιώθηκε από τους δημιουργούς του Apollo (Poelchau *et al*, 2015) για την αλληλούχηση των γονιδιωμάτων 5000 διαφορετικών ειδών εντόμων, καθώς και άλλων αρθρόποδων. Η πλατφόρμα για την υλοποίηση της i5k φιλοξενείται από την Εθνική Γεωργική Βιβλιοθήκη (National Agricultural Library-NAL) όπου έχει ενσωματωθεί και το πρόγραμμα Apollo.

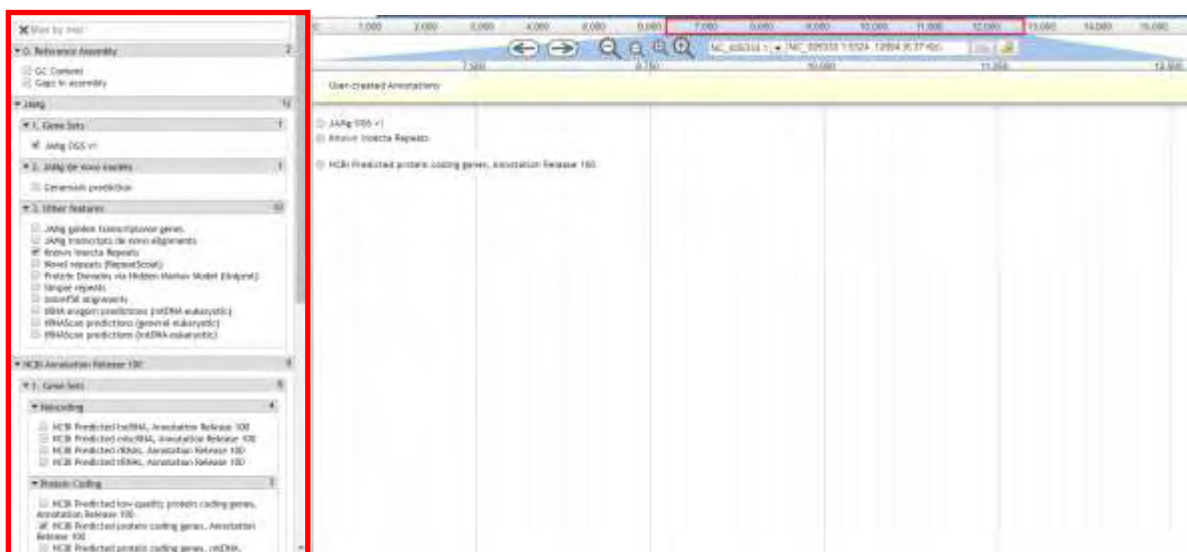
Το Apollo είναι μια διαδικτυακή πλατφόρμα που διαθέτει ενσωματωμένο τον φυλλομετρητή JBrowse. Δημιουργήθηκε με σκοπό τη χειροκίνητη επιμέλεια των σχεδιασμένων από αλγόριθμους (συναινετικών) γονιδιακών μοντέλων των εντόμων που φιλοξενούνται. Τα γονιδιακά μοντέλα που παράγονται από τους αλγόριθμους στηρίζονται σε πειραματικά δεδομένα και προβλέψεις ανάλογα με τη σειρά εντολών που τους έχει οριστεί. Οι αλγόριθμοι αυτοί, ωστόσο και η σειρά εντολών τους δεν είναι τέλεια σχεδιασμένοι και μπορεί να μειώσουν την ακρίβεια της γονιδιακής αναπαράστασης. Για το σκοπό αυτό το πρόγραμμα Apollo παρέχει υπολογιστικά και πειραματικά στοιχεία και επιτρέπει τη χειροκίνητη επιμέλεια των διαθέσιμων γονιδιακών μοντέλων. Πιο συγκεκριμένα, για τους οργανισμούς που φιλοξενεί, το Apollo παρέχει το σύνολο του γονιδιώματος τους σε συνδυασμό με:

- *Δεδομένα RNA αλληλούχησης*: Στο περιβάλλον του Apollo δίνεται η επιλογή για μελέτη των δεδομένων αλληλούχησης του συνολικού RNA από τους ιστούς και τα στάδια ανάπτυξης ενός οργανισμού, αλλά και δεδομένων για κάθε στάδιο και ιστό ξεχωριστά.
- *Συναινετικά γονιδιακά μοντέλα*: Επιπρόσθετα, παρέχονται συναινετικά γονιδιακά μοντέλα από βάσεις δεδομένων, όπως η NCBI και η JAMg τα οποία οι χρήστες μπορούν να μελετήσουν και να επεξεργαστούν.
- *«Known insecta repeats»*: Με το Apollo είναι δυνατή και η μελέτη των γνωστών επαναλήψεων στο σύνολο των εντόμων, με σκοπό την εξαγωγή πληροφοριών για τυχόν μεταθετά στοιχεία και ψευδογονίδια στην περιοχή, ή ικρίωμα, του γονιδιώματος (scaffold) που μελετάται.

Παράλληλα, ενσωματωμένα στο Apollo υπάρχουν και εργαλεία όπως το BLAST, το HMMER για την εύρεση ομόλογων πρωτεϊνών και πολλαπλή στοίχιση τους και το Clustal, που αναλύεται παρακάτω, σε συνδυασμό με άλλα εργαλεία εξάσκησης.

Στην παρούσα εργασία, το Apollo χρησιμοποιήθηκε γιατί ανάμεσα στους οργανισμούς που φιλοξενεί βρίσκονται και δέκα έντομα της τάξης των διπτέρων συμπεριλαμβανομένου και του δάκου της ελιάς. Στο περιβάλλον του Apollo πραγματοποιήθηκε η επιμέλεια και επεξεργασία των συναινετικών γονιδιακών μοντέλων των οσμοδεσμευτικών πρωτεϊνών, των γευστικών και των ιονοτροπικών υποδοχέων του δάκου της ελιάς. Για την επεξεργασία του επιθυμητού κάθε φορά γονιδίου στο περιβάλλον του Apollo, αρχικά, εντοπίζονταν η περιοχή του στο γονιδίωμα με βάση τον μοναδικό κωδικό “XM” του που λήφθηκε από την NCBI. Από τη λίστα «Available tracks» (βλ. Εικ. 2.1) επιλέγονταν οι επιθυμητές γονιδιακές προβλέψεις και αποδείξεις. Συγκεκριμένα, για όλα τα γονίδια που μελετήθηκαν στην παρούσα εργασία, από τα “available tracks” προστέθηκαν τα συναινετικά γονιδιακά

μοντέλα των βάσεων δεδομένων JAMg και NCBI, το σύνολο των δεδομένων της RNA αλληλούχησης (compiled RNA coverage) και τα γνωστά μεταθετά στοιχεία στα έντομα (known insecta repeats) (βλ. Εικ. 2.1).

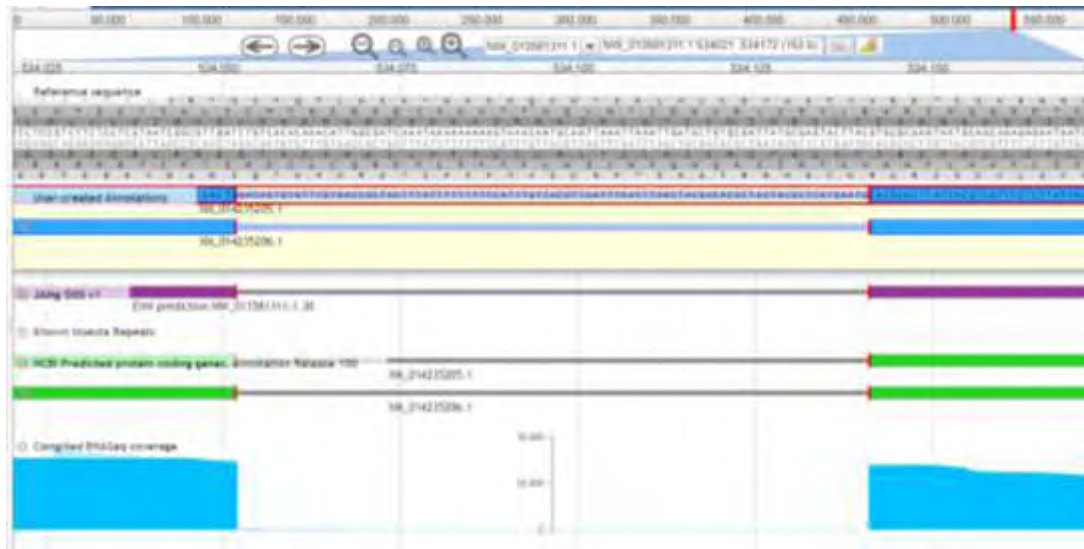


Εικόνα 2.1: Το περιβάλλον (User Interface) του Apollo i5k στο οποίο φαίνονται τα εργαλεία που χρησιμοποιήθηκαν για την επιμέλεια των επιθυμητών γονιδίων.

Τα συναινετικά γονιδιακά μοντέλα στοιχίζονται κάτω από τις αντίστοιχες νουκλεοτιδικές βάσεις από τις οποίες μεταγράφονται, σύμφωνα με την αλληλουχία τους, στα αντίστοιχα ικριώματα (scaffolds) στο γονιδίωμα. Τα εξόνια του γονιδίου που αποτελούν το ώριμο mRNA αντιστοιχούν στις περιοχές του γονιδιώματος για τις οποίες παρατηρείται υψηλός αριθμός αναγνώσεων μετά από RNA αλληλούχηση (RNAseq), ενώ τα ιντρόνια αντιστοιχούν στις περιοχές του γονιδιώματος που εμφανίζουν πολύ χαμηλό αριθμό αναγνώσεων.

Για κάθε γονίδιο επιλέγονταν το συναινετικό μοντέλο που συμφωνούσε περισσότερο με τα δεδομένα RNAseq και ακολουθούσε η επεξεργασία του με βάση και πάλι τις πληροφορίες αυτές ώστε να αποκτηθεί η πραγματική δομή τους. Η επεξεργασία του επιτυγχάνονταν στη μπάρα “User-created Annotations” (Εικόνα 2.1). Πιο συγκεκριμένα, η επιμέλεια και η επεξεργασία των συναινετικών γονιδιακών μοντέλων συμπεριλάμβανε:

1) Τον έλεγχο του αριθμού των προτεινόμενων εξονίων, αλλά και της δομής τους με τον έλεγχο των περιοχών ματίσματος. Σε μερικές περιπτώσεις τα δύο συναινετικά γονιδιακά μοντέλα δεν συμφωνούσαν ως προς στον αριθμό εξονίων ή τα όρια των εξονίων-ιντρονίων και στις περιπτώσεις αυτές το γονιδιακό μοντέλο που επιλέγονταν ως το πραγματικό ήταν και αυτό που συμφωνούσε με τα δεδομένα RNA αλληλούχησης και έφερε τις αποδεκτές νουκλεοτιδικές βάσεις στα όρια των ιντρονίων: “GT” στο 5’ άκρο και “AG” στο 3’ άκρο (βλ. Εικ. 2.2.).



Εικόνα 2.2.: Έλεγχος ορίων ιντρονίων- εξονίων στο περιβάλλον του Apollo και ύπαρξης των δύο διακριτών νουκλεοτιδίων GT και AG στα άκρα των ιντρονίων.

2) Μετά την επιλογή του πραγματικού από τα συναινετικά γονιδιακά μοντέλα και την περαιτέρω επιμέλειά του όπου χρειαζόταν, ελέγχονταν οι αμετάφραστες περιοχές τους, οι οποίες συνήθως επιμηκύνονταν στις περιπτώσεις ασυμφωνίας με τα αποδεικτικά στοιχεία από την RNA αλληλούχηση.

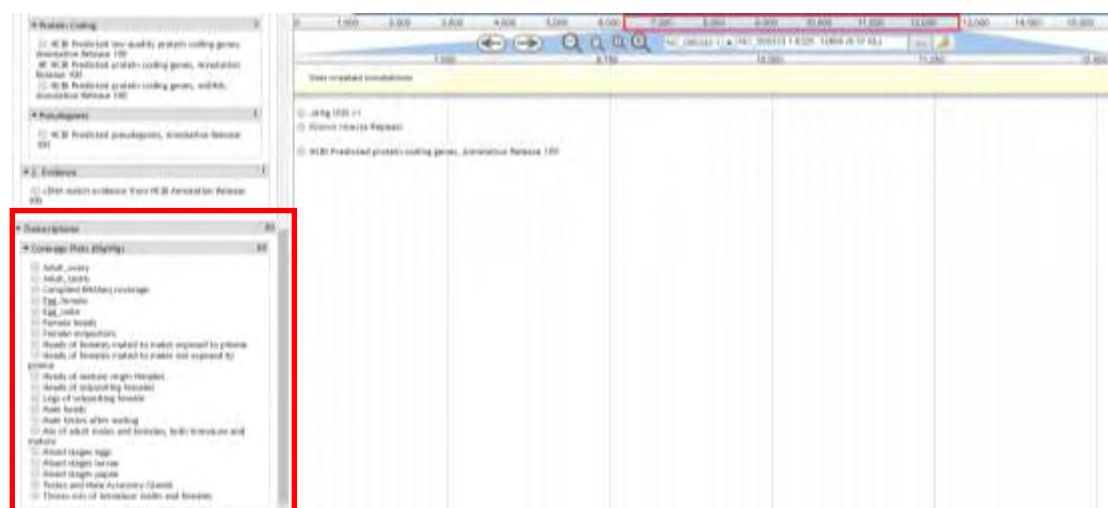
3α) Για να επιβεβαιώσουμε ότι η επιμέλεια των γονιδιακών μοντέλων κάθε πρωτεΐνης ήταν ακριβής, ακολουθούσε πολλαπλή στοίχιση της στο Clustal Omega με τις ορθόλογες στα είδη *C. capitata*, *D. melanogaster* και *M. domestica*. Οι πρωτεΐνες που φέρουν ορθόλογες σχέσεις διαθέτουν υψηλή ομολογία μεταξύ τους και αναμένουμε όμοια αμινοξικά μοτίβα στα είδη που τις φέρουν. Έτσι και για το δάκο, τα γονιδιακά μας μοντέλα θα διαθέτουν όμοια αμινοξικά μοτίβα με τα ορθόλογα των υπόλοιπων ειδών, τα οποία θα αναμένουμε να ανιχνεύονται μετά από πολλαπλή στοίχιση.

3β) Η πολλαπλή στοίχιση των πρωτεϊνών του δάκου με τις ορθόλογες για την ανίχνευση της πραγματικής τους δομής, πραγματοποιούνταν και στις περιπτώσεις όπου εμφανίζονταν πολύ χαμηλή έκφραση τους. Η χαμηλή έκφραση συνεπάγεται και χαμηλό αριθμό αναγνώσεων με αποτέλεσμα, στις περιπτώσεις αυτές, η έκφραση των εξονίων να μην είναι εμφανής με RNA αλληλούχηση και οι πληροφορίες για το διαχωρισμό των εξονίων από τα ιντρόνια να είναι ανεπαρκής.

4) Τέλος, ελέγχονταν αν υπήρχαν επιμελημένα γονίδια που να εντοπίζονται στο ίδιο γονιδιωματικό ικρίωμα (scaffold) στα δεδομένα αλληλούχησης και οι σχετικές τους συντεταγμένες καταγράφονταν. Τα γονίδια που ανιχνεύονται σε γειτονικές θέσεις στο ίδιο ικρίωμα μελετώνται περαιτέρω, καθώς πιθανότατα προέκυψαν από διπλασιασμούς και ενδέχεται είναι ειδο-ειδικές. Πληροφορίες, όπως οι περιοχές του γονιδιώματος και οι συντεταγμένες κάθε γονιδίου, το μήκος κάθε πρωτεΐνης και οι αλληλουχίες τους καταγράφηκαν και παρατίθενται σε πίνακες.

Επιπρόσθετα, το Apollo χρησιμοποιήθηκε για τη μελέτη των σταδίων ανάπτυξης και ιστών του δάκου όπου παρατηρείται έκφραση των επιθυμητών γονιδίων. Το Apollo παρέχει αποτελέσματα αλληλούχησης του RNA από: Ωοθέτες ενήλικων θηλυκών,

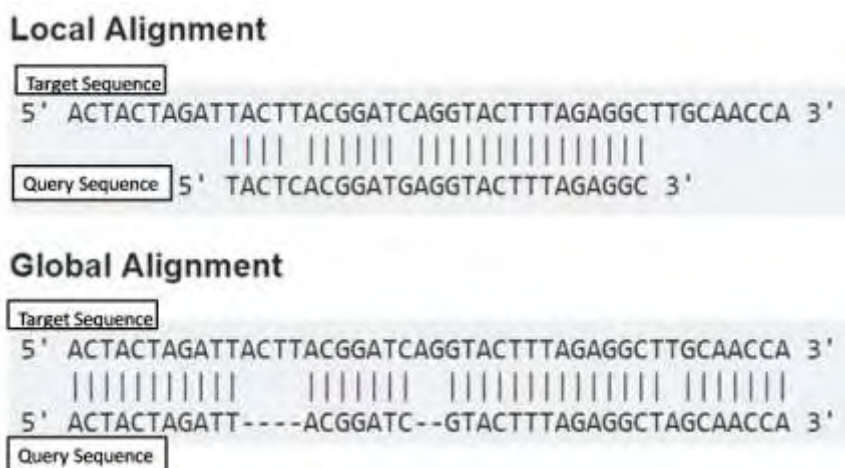
όρχεις ενήλικων αρσενικών, αυγά που φέρουν αρσενικά, αυγά που φέρουν θηλυκά, κεφάλια θηλυκών, ωοθήκες θηλυκών, κεφάλια θηλυκών που έχουν ζευγαρώσει με αρσενικά κατά την έκθεση τους ή μη σε πινένιο, κεφάλια ώριμων παρθένων θηλυκών, κεφάλια θηλυκών κατά την ωοτοκία, πόδια θηλυκών κατά την ωοτοκία, κεφάλια αρσενικών, όρχεις αρσενικών μετά το ζευγάρωμα, μίξη ενήλικων ανώριμων και ώριμων θηλυκών και αρσενικών, μίξη σταδίων αυγών, μίξη σταδίων προνύμφης, μίξη σταδίων νύμφης, όρχεις και αδένες αρσενικών και μίξη θωράκων από ανώριμα αρσενικά και θηλυκά (Βλ. Εικ. 2.3.)



Εικόνα 2.3.: Δεδομένα αλληλούχησης RNA από διαφορετικά στάδια και ιστούς ανάπτυξης που παρέχονται στην πλατφόρμα του Apollo.

2.2. Στοιχισή νουκλεοτιδικών και αμινοξικών αλληλουχιών

Η πιο συχνά χρησιμοποιούμενη μέθοδος για τη στοιχισή ενός ζεύγους νουκλεοτιδικών ή αμινοξικών αλληλουχιών είναι μέσω πολλαπλής στοιχισής. Αρχικά, στοιχίζονται οι ομάδες των αλληλουχιών που είναι οι πιο συγγενικές και σταδιακά οι συγγενικές αυτές ομάδες αλληλουχιών στοιχίζονται μεταξύ τους χωρίς να επηρεάζεται η αρχική στοιχισή. Έτσι, είναι δυνατή η διερεύνηση των ομοιοτήτων και των διαφορών πολλών αλληλουχιών ταυτόχρονα με μεγάλη ακρίβεια (Thompson *et al*, 1997). Η μέθοδος αυτή χρησιμοποιείται και για τη στοιχισή και σύγκριση ακόμα και μεγάλου μήκους πρωτεϊνών με τις ορθολογές τους, με σκοπό την εύρεση συντηρημένων μοτίβων ακόμα και σε μακρινά είδη. Ο πιο διαδεδομένος τρόπος πολλαπλής στοιχισής είναι η «pairwise» στοιχισή, μέσω της οποίας πραγματοποιείται είτε καθολική, είτε τοπική στοιχισή. Κατά την καθολική στοιχισή, γίνεται προσπάθεια στοιχισής ολόκληρης της αλληλουχίας, ενώ στην τοπική, στοιχίζονται μόνο οι αλληλουχίες με τη μέγιστη ομοιότητα (Βλ. Εικ. 2.4). Τα κενά (gaps) που δημιουργούνται από τη στοιχισή μεταξύ ομόλογων αλληλουχιών αντιστοιχούν σε μεταλλάξεις στην αλληλουχία, όπως οι προσθήκες και οι διαγραφές νουκλεοτιδίων (Brudno *et al*, 2003).



Εικόνα 2.4.: Παράδειγμα τοπικής και καθολικής στοίχισης

2.2.1. Clustal Omega

Το Clustal Omega (Sievers *et al*, 2011) αποτελεί πρόγραμμα πολλαπλής στοίχισης που χρησιμοποιεί δέντρα-οδηγούς και profile-profile κρυπτομαρκοβιανά μοντέλα (HMM profile-profile), επιτρέποντας τη στοίχιση και σύγκριση πολλαπλών αλληλουχιών από οποιαδήποτε είδη.

Το Clustal Omega αποτελεί την καινούργια έκδοση των παλιών ClustalX και ClustalW και διατηρεί τη βασική προοδευτική προσέγγιση πολλαπλής στοίχισης (MSA) (Madeira *et al*, 2019). Δηλαδή η σειρά των στοίχισεων καθορίζεται από ένα δέντρο-οδηγό, που με τη σειρά του δημιουργείται από τον υπολογισμό των pairwise αποστάσεων μεταξύ των αλληλουχιών. Η διαφορά του Clustal omega από τις προηγούμενες εκδόσεις του είναι ότι χρησιμοποιεί τον αλγόριθμο mBed για τη δημιουργία δέντρων-οδηγών οποιουδήποτε μεγέθους. Ο αλγόριθμος mBed του Clustal λειτουργεί σε τρία βήματα:

Αρχικά, επιλέγεται ένας αριθμός αλληλουχιών από αυτές που έχουν εισαχθεί και συγκρίνονται μεταξύ τους. Αν δύο από αυτές τις αλληλουχίες είναι μεταξύ τους πολύ όμοιες (πάνω από το αποδεκτό όριο απόστασης/ομοιότητας), τότε η μικρότερη από αυτές διαγράφεται γιατί θεωρείται περιττή. Το όριο της απόστασης γενικά τίθεται στο μηδέν, ώστε να διαγράφονται μόνο οι τελείως ίδιες αλληλουχίες.

Κατά το δεύτερο βήμα, οι αλληλουχίες που επιλέχθηκαν στο πρώτο βήμα αναλύονται και σε σχέση με τις πιο απομακρυσμένες και αν φέρουν διαφορές εισάγονται μαζί με τις τελικές αλληλουχίες. Όπως και στο πρώτο βήμα αυτές που είναι όμοιες διαγράφονται και κρατούνται μόνο οι αλληλουχίες που φέρουν διαφορές.

Τέλος, όλες οι τελικές αλληλουχίες που προέκυψαν από αυτή τη διαδικασία χρησιμοποιούνται για να συσχετιστούν όλες οι αλληλουχίες που είχαν εισαχθεί αρχικά, μέσω υπολογισμού των αποστάσεων τους με τις τελικές αλληλουχίες που προέκυψαν από τη διαδικασία (Blackshields *et al*, 2010)

Στη συγκεκριμένη εργασία το Clustal Omega χρησιμοποιήθηκε για την εύρεση του πραγματικού από τα συναινετικά γονιδιακά μοντέλα του Apollo όπου τα δεδομένα RNA αλληλούχησης δεν επαρκούσαν, αλλά και την επιβεβαίωση της αλληλουχίας τους μετά από την επιμέλεια τους. Συγκεκριμένα, στο Clustal Omega η κάθε αλληλουχία μας συγκρίθηκε με τις ορθόλογες της, στα είδη *Ceratitidis capitata*, *Drosophila melanogaster* και *Musca domestica* με τις οποίες μεγάλος αριθμός αλληλουχιών του δάκου εμφανίζει υψηλή ομολογία και συνεπώς, είναι αναμενόμενο να διαθέτουν πολλά κοινά μοτίβα. Ανάλογα, με τα μοτίβα αυτά στο δάκο και στα υπόλοιπα τρία είδη ελέγχονταν και επιδιορθώνονταν κάθε φορά τα σφάλματα που έφεραν τα συναινετικά γονιδιακά μοντέλα. Οι πληροφορίες που προκύπτουν από τη στοίχιση των αλληλουχιών χρησιμοποιούνται για την κατασκευή φυλογενετικών δέντρων.

2.2.2. Muscle

Το πρόγραμμα MUSCLE χρησιμοποιεί μετρήσεις δύο αποστάσεων (two-distance) για την πολλαπλή στοίχιση αλληλουχιών, την απόσταση kmer και την απόσταση kimura. Η απόσταση kmer είναι μια συνεχής ακολουθία μήκους k και προκύπτει από το κλάσμα των kmers που είναι κοινά. Οι συγγενείς αλληλουχίες συνήθως έχουν περισσότερα kmers κοινά από ότι θα ήταν αναμενόμενο και ο υπολογισμός των kmers αυτών είναι ταχύς και δεν προ-απαιτεί στοίχιση των αλληλουχιών αυτών. Για το λόγο αυτό η απόσταση kmer χρησιμοποιείται για μη στοιχισμένα ζεύγη αλληλουχιών. Για στοιχισμένα ζεύγη αλληλουχιών χρησιμοποιείται η απόσταση kimura που αποκαλύπτει την εξελικτική απόσταση των αλληλουχιών λαμβάνοντας ως δεδομένο ότι οι μεταπτώσεις και οι μεταστροφές συμβαίνουν με διαφορετικούς ρυθμούς.

Η πολλαπλή στοίχιση στο MUSCLE πραγματοποιείται σε τρία στάδια, με την απόσταση kmer να υπολογίζεται στο πρώτο στάδιο και την απόσταση Kimura στο δεύτερο στάδιο.

Στο πρώτο στάδιο, που ονομάζεται draft progressive, πραγματοποιείται μια πρώτη πολλαπλή στοίχιση με την απόσταση kmer ώστε να προκύψει ένα πρόχειρο, γρήγορο αποτέλεσμα στοίχισης το οποίο, ωστόσο, δεν είναι ακριβές.

Κατά το δεύτερο στάδιο, το στάδιο improved progressive, το αποτέλεσμα που προέκυψε με την απόσταση kmer εκτιμάται και πάλι, αλλά αυτή τη φορά με την απόσταση Kimura. Σκοπός είναι να προκύψει ένα πιο ακριβές αποτέλεσμα το οποίο όμως προϋποθέτει στοίχιση όλων των αλληλουχιών.

Τέλος, κατά το τρίτο στάδιο, το στάδιο refinement, πραγματοποιείται βελτίωση της τελικής πολλαπλής στοίχισης που προέκυψε από την απόσταση Kimura. Αυτό επιτυγχάνεται με τη δημιουργία νέων στοίχισεων από την τελική της Kimura μέχρι να προκύψει με βάση τις παραμέτρους που τέθηκαν στο πρόγραμμα η επιθυμητή στοίχιση (Edgar *et al*, 2004)

Στην παρούσα εργασία ο αλγόριθμος του MUSCLE χρησιμοποιήθηκε για την πολλαπλή στοίχιση των OBPs του δάκου μεταξύ τους και την ταξινόμησή τους στις τέσσερις υπο-οικογένειες που αναφέρθηκαν στην *Εισαγωγή* με βάση τα αμινοξικά

τους μοτίβα. Επιπρόσθετα, πραγματοποιήθηκε η πολλαπλή στοίχιση των πρωτεϊνικών αλληλουχιών στο δάκο με τα υπόλοιπα τρία είδη, ξεχωριστά για κάθε οικογένεια των πρωτεϊνών. Τα αποτελέσματα που προέκυψαν από την πολλαπλή στοίχιση, χρησιμοποιήθηκαν για την κατασκευή φυλογενετικών δέντρων για κάθε μία από τις τρεις οικογένειες πρωτεϊνών που μελετώνται.

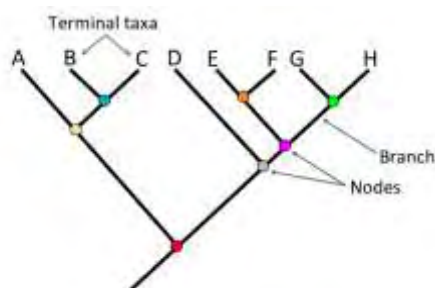
2.3. Φυλογενετική ανάλυση

Φυλογένεση είναι η μελέτη των εξελικτικών σχέσεων μεταξύ γονιδίων, πρωτεϊνών ή ειδών και τα φυλογενετικά δέντρα αποτελούν μια απεικόνιση των σχέσεων αυτών.

Οι μέθοδοι που χρησιμοποιούνται για την κατασκευή των φυλογενετικών δέντρων συνοψίζονται σε μεθόδους βασισμένες σε αποστάσεις που υπολογίζουν τις αποστάσεις κάθε ζεύγους αλληλουχιών και μεθόδους βασισμένες σε χαρακτήρες, όπου η θέση κάθε χαρακτήρα προσαρμόζεται ανάλογα με τις θέσεις των υπόλοιπων χαρακτήρων.

Τα φυλογενετικά δέντρα μπορεί να είναι άριζα ή έρριζα και όταν η ρίζα υπάρχει, αντιστοιχεί στον κοινό πρόγονο των ταξινομικών ομάδων (taxa) που μελετώνται. Οι ταξινομικές ομάδες οργανώνονται σε κλάδους και κάθε κλάδος αποτελείται από τις ταξινομικές ομάδες που φέρουν κοινό πρόγονο. Το μήκος του κάθε κλάδου αντιστοιχεί στις τροποποιήσεις που συνέβησαν μέχρι την επόμενη διάστασή του. Τα σημεία στα οποία ενώνονται οι κλάδοι ονομάζονται κόμβοι και αντιστοιχούν στο γεγονός που συνέβη και οδήγησε στο διαχωρισμό των κλάδων αυτών.

Ομάδες γονιδίων με συσχετισμένες εξελικτικές αλλαγές στην έκφραση τους έχουν κοινή δομή και λειτουργία και ανιχνεύονται στον ίδιο κλάδο στο φυλογενετικό δέντρο. Τα γονίδια αυτά χρησιμοποιούνται ως «δωμάτια» για την ανίχνευση των γονιδίων των οποίων η δομή και η λειτουργία αποκλίνει και συνεπώς, βρίσκονται σε διαφορετικό κλάδο στο φυλογενετικό δέντρο. Έτσι, αποκαλύπτονται και οι διπλασιασμοί ή οι απώλειες που έχουν υποστεί τα γονίδια διαφόρων ειδών και που είναι υπεύθυνα για τις διαφορές στη μορφολογία και λειτουργία τους (Dunn *et al*, 2013).

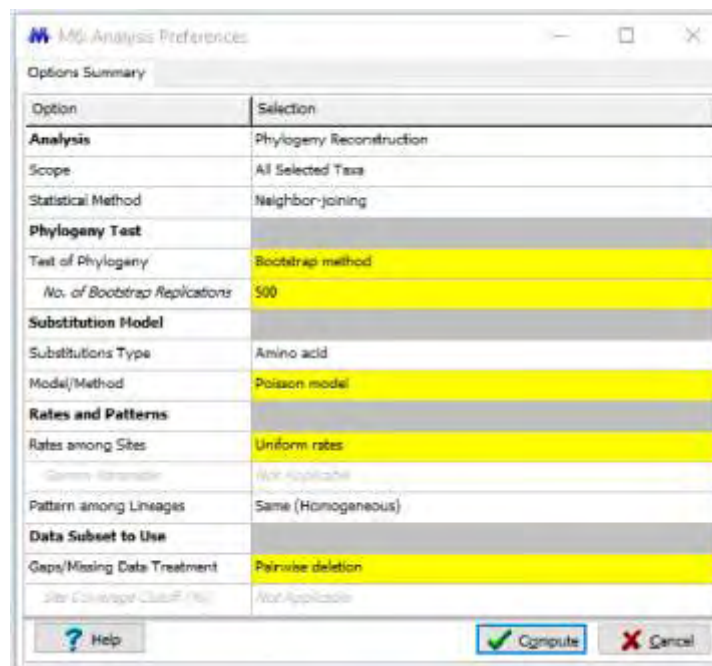


Εικόνα 2.3.: Έρριζο φυλογενετικό δέντρο, κλάδοι(branches), κόμβοι(nodes), και ταξινομικές ομάδες(taxa).

2.3.1. Mega-X

Το Mega-X αποτελεί ένα υπολογιστικό λογισμικό που αναπτύχθηκε πρώτη φορά το 1994 (Kumar *et al*, 1994) και αναβαθμίστηκε οκτώ φορές μέχρι σήμερα (Kumar *et al*, 2018). Χρησιμοποιείται για τη διεξαγωγή στατιστικής ανάλυσης μοριακής εξέλιξης και για την κατασκευή φυλογενετικών δέντρων, παρέχοντας στατιστικές μεθόδους, όπως αυτές των αποστάσεων και της μέγιστης φειδωλότητας. Το πρόγραμμα Mega-X υποστηρίζει παράλληλα την πολλαπλή στοίχιση αλληλουχιών με τη χρήση των αλγορίθμων ClustalW και MUSCLE που βρίσκονται ενσωματωμένοι στο λογισμικό του, ενώ ταυτόχρονα επιτρέπει την επιλογή των επιθυμητών παραμέτρων.

Μετά την πολλαπλή στοίχιση με τον αλγόριθμο MUSCLE των αλληλουχιών του δάκου με αυτές των υπόλοιπων 3 ειδών, χρησιμοποιήθηκαν οι μέθοδοι κατασκευής φυλογενετικού δέντρου που παρέχονται από το Mega-X με την επιλογή των επιθυμητών παραμέτρων και κατασκευάστηκε ένα πρώτο φυλογενετικό δέντρο για τις OBPs, τους γευστικούς και τους ιονοτροπικούς υποδοχείς του δάκου της ελιάς. Οι παράμετροι που επιλέχθηκαν για τη φυλογενετική ανάλυση φαίνονται στην εικόνα 2.2.a, και συνοψίζονται στη μέθοδο bootstrap για 500 επαναλήψεις με το μοντέλο του Poisson και με «pairwise» διαγραφή των ελλειπών δεδομένων. Το φυλογενετικό δέντρο δημιουργήθηκε με τη μέθοδο Neighbor-joining με την οποία είναι ορατές και οι αποστάσεις μεταξύ κάθε αμινοξικής αλληλουχίας ανάμεσα στα είδη.



Εικόνα 2.4.: Παράμετροι που τέθηκαν για τη φυλογενετική ανάλυση

NEIGHBOR JOINING METHOD: Με τη μέθοδο αυτή είναι δυνατή η εύρεση ζευγών λειτουργικών ταξινομικών μονάδων («γειτόνων») που ελαχιστοποιούν το συνολικό μήκος του κλάδου σε κάθε στάδιο της ομαδοποίησής τους. Η μέθοδος neighbor joining χρησιμοποιήθηκε ώστε να ληφθεί και το μήκος του κάθε κλάδου, καθώς και η τοπολογία ενός φειδωλού φυλογενετικού δέντρου από δεδομένα εξελικτικά μακρινά (Saitou *et al*, 1987).

ΕΚΤΙΜΗΣΗ BOOTSTRAP: Αποτελεί μια στατιστική μέθοδος για την εκτίμηση του πόσο μια συγκεκριμένη ομαδοποίηση υποστηρίζεται από όλα τα δεδομένα που χρησιμοποιήθηκαν για να οικοδομηθούν το δέντρο. Κάθε μία από τις συνολικές εκτιμήσεις bootstrap είναι ένας πίνακας δεδομένων που προκύπτει από τον προηγούμενο του με κάποιους χαρακτήρες να επαναλαμβάνονται πολλές φορές και κάποιοι καθόλου ανάλογα με το αν υποστηρίζονται. Στο τέλος, όλοι οι πίνακες αναλύονται ώστε να βρεθούν τα κλάσματα που ομαδοποιούνται όμοια με τον αρχικό πίνακα. Όσο πιο μεγάλο ένα κλάσμα τόσο περισσότερο υποστηρίζεται ένας κλάδος από τα δεδομένα.

ΜΟΝΤΕΛΟ POISSON: Το μοντέλο Poisson εκφράζει την κατανομή της πιθανότητας μιας τυχαίας μεταβλητής και στο φυλογενετικό δέντρο αντιπροσωπεύει την πιθανότητα εμφάνισης ενός αριθμού γεγονότων σε ένα συγκεκριμένο χρονικό διάστημα. Χρησιμοποιείται για μια σειρά διακριτών γεγονότων όπου ο μέσος χρόνος μεταξύ των γεγονότων είναι γνωστός, αλλά η ακριβής στιγμή που συνέβη κάθε γεγονός είναι τυχαία. Στο μοντέλο αυτό η άφιξη ενός γεγονότος είναι ανεξάρτητη από το προηγούμενο γεγονός (Koehrsen *et al*, 2019)

«PAIRWISE» ΔΙΑΓΡΑΦΗ: Στην «pairwise» διαγραφή, χρησιμοποιούνται, πριν διαγραφούν, ακόμα και οι περιπτώσεις που φέρουν ελλιπή δεδομένα για την ανάλυση των υπόλοιπων μεταβλητών σε αντίθεση με τη listwise διαγραφή όπου οι «ελλιπείς» περιπτώσεις διαγράφονται άμεσα και δε χρησιμοποιούνται καθόλου για την ανάλυση των υπόλοιπων μεταβλητών. Χρησιμοποιήθηκε pairwise διαγραφή, λοιπόν, ώστε να ληφθούν υπόψη όλες οι μεταβλητές για τη δημιουργία του φυλογενετικού δέντρου.

2.3.2. Figtree

Το figtree (Rambaut, 2006) αποτελεί πρόγραμμα για την παρατήρηση φυλογενετικών δέντρων και την επεξεργασία τους. Μερικές από τις δυνατότητες που παρέχει είναι οι εξής:

- Διαπλατορμικό σύστημα γραφικής απεικόνισης δέντρων
- Επιλογή μεταξύ τριών διαφορετικών τύπων απεικόνισης του δέντρου
- Χρωματισμός των κλάδων και των ετικετών κάθε άκρου
- Κατάρρευση των κλάδων που εμφανίζονται με μικρή πιθανότητα σε τρίγωνα
- Απεικόνιση του ύψους των κόμβων, του μήκους των κλάδων και των τιμών bootstrap ή άλλων βοηθητικών τιμών.

Στην παρούσα εργασία, χρησιμοποιήθηκε για τη βελτίωση των φυλογενετικών δέντρων που προέκυψαν από το Mega-X. Με το πρόγραμμα αυτό έγινε η μεταβολή της θέσης της ρίζας, του σχήματος του δέντρου και προστέθηκαν διαφορετικά χρώματα σε κάθε κλάδο σε αντιστοιχία με το κάθε είδος, ώστε να είναι εύκολα διακριτές οι αλληλουχίες για κάθε είδος.

2.4. Χαρακτηρισμός διαμεμβρανικών περιοχών

Η πρόβλεψη των διαμεμβρανικών περιοχών των πρωτεϊνών αποτελούν πολύτιμο εργαλείο του τομέα της βιοπληροφορικής. Οι διαμεμβρανικές πρωτεΐνες ακολουθούν ένα γραμματικό μοτίβο στο οποίο οι κυτταροπλασματικές και οι εξωκυτταρικές περιοχές εναλλάσσονται. Σύμφωνα με τα μοτίβα αυτά προκύπτει και η πιθανή τοπολογία των πρωτεϊνών και συνεπώς, οι πιθανές διαμεμβρανικές περιοχές. Έτσι, μια βιοπληροφορική μέθοδος πρόβλεψης που λαμβάνει υπόψη τα μοτίβα αυτά μπορεί να παρέχει την ακριβή τοπολογία οποιασδήποτε πρωτεΐνης, ακόμα και αν αυτή φέρει μόνο μία διαμεμβρανική περιοχή (Krogh *et al*, 2001). Οι μέθοδοι πρόβλεψης των διαμεμβρανικών περιοχών και της πρωτεϊνικής τοπολογίας συγκαταλέγονται ανάμεσα στις πιο ακριβείς μεθόδους της βιοπληροφορικής (Punta *et al*, 2007). Οι πιο αποτελεσματικές από αυτές τις μεθόδους ενσωματώνουν αλγόριθμους που προβλέπουν τον αριθμό των διαμεμβρανικών περιοχών μιας πρωτεΐνης και τις εξωκυτταρικές και κυτταροπλασματικές περιοχές της στο σύνολο τους και όχι μεμονωμένα. Έτσι, διαχωρίζονται οι διαμεμβρανικές από τις υδατοδιαλυτές πρωτεΐνες και προβλέπεται η συνολική τοπολογία τους, ώστε να χρησιμοποιηθεί σε γονιδιωματικές αναλύσεις (Krogh *et al*, 2001).

2.4.1. TMHMM

Το TMHMM (Krogh *et al*, 2001) είναι ένα ευρέως χρησιμοποιούμενο εργαλείο βιοπληροφορικής που βασίζεται σε ένα κρυπτομακροβιανό μοντέλο (HMM) για την πρόβλεψη διαμεμβρανικών ελικών των πρωτεϊνών σύμφωνα με εσωτερικούς πιθανολογικούς κανόνες. χρησιμοποιείται, επιτυχώς, για την πρόβλεψη της τοπολογίας διαμεμβρανικών πρωτεϊνών και την ανίχνευση σηματοδοτικών πεπτιδίων (Lesk *et al*, 2017). Ο διακομιστής του TMHMM χρησιμοποιεί έναν αλγόριθμο εκπαιδευμένο με βάση μια σειρά από 160 επιβεβαιωμένες διαμεμβρανικές πρωτεΐνες για την επαλήθευση των προβλέψεων και μια σειρά από 645 πρωτεΐνες από τη βάση δεδομένων PDB που, επιβεβαιωμένα, δε διαθέτουν διαμεμβρανικές περιοχές ως αρνητικό μάρτυρα. Ο αλγόριθμος του TMHMM προβλέπει με ακρίβεια το 97-98% των διαμεμβρανικών πρωτεϊνικών ελικών και διαχωρίζει τις υδατοδιαλυτές και μεμβρανικές πρωτεΐνες με ευαισθησία και εξειδίκευση μεγαλύτερη από 99%. Ωστόσο, η ακρίβεια του μειώνεται στην περίπτωση των σηματοδοτικών πεπτιδίων. Με βάση το μοντέλο του TMHMM έχει εκτιμηθεί ότι το 20-30% όλων των γονιδίων στο γονιδίωμα των περισσότερων οργανισμών κωδικοποιούν διαμεμβρανικές πρωτεΐνες.

Το μοντέλο TMHMM χρησιμοποιήθηκε στην παρούσα μελέτη για την πρόβλεψη και ανάλυση των διαμεμβρανικών περιοχών στο σύνολο των γευστικών και ιονοτροπικών υποδοχέων στο δάκο, ώστε να διαπιστωθεί αν φέρουν τις αναμενόμενες 7 και 3 διαμεμβρανικές περιοχές αντίστοιχα.

2.5. Ανάλυση αμινοξικών μοτίβων

2.5.1. MEME (Multiple Em For Motif Elicitation)

Ο διακομιστής του MEME (Timothy *et al*, 1994) αποκαλύπτει ένα ή περισσότερα νουκλεοτιδικά ή αμινοξικά μοτίβα μέσα από μια συλλογή DNA ή πρωτεϊνικών αλληλουχιών. Ο αλγόριθμος που χρησιμοποιεί το πρόγραμμα MEME ονομάζεται «MM». Στον αλγόριθμο «MM» ορίζεται το σετ των δεδομένων από τις επιθυμητές αλληλουχίες ώστε να υπολογίσει τις παραμέτρους ενός πιθανού μοντέλου που θα μπορούσε να έχει παράγει το σετ δεδομένων. Το πιθανό μοντέλο αποτελείται από δύο δομικά στοιχεία. Το ένα δομικό στοιχείο περιγράφει τα αμινοξικά ή νουκλεοτιδικά μοτίβα, ενώ το άλλο περιγράφει όλες τις άλλες θέσεις στις αλληλουχίες.

Στην παρούσα ανάλυση, ο αλγόριθμος του προγράμματος MEME αξιοποιήθηκε για την πρόβλεψη των κοινών αμινοξικών μοτίβων μεταξύ των πρωτεϊνών του δάκου και των ειδών *C. cap*, *D. mel* και *M. dom*.

2.6. UniProt (Universal Protein Knowledgebase)

Η βάση δεδομένων UniProt (Bairoch *et al*, 1997) παρέχει πληροφορίες σχετικά με την αλληλουχία και τη λειτουργία καλά ταξινομημένων και επιμελημένων πρωτεϊνών με σκοπό τη στήριξη της επιστημονικής έρευνας. Δημιουργήθηκε από την ένωση των βάσεων δεδομένων Swiss-Prot/TrEMBL και PIR-PSD οι οποίες συνυπήρχαν με κάλυψη διαφορετικών αλληλουχιών. Η βάση δεδομένων της UniProt αποτελείται από τρία τμήματα:

- i) Το τμήμα UniParc (UniProt Archive) για την αρχειοθέτηση των πληροφοριών για κάθε πρωτεϊνική αλληλουχία.
- ii) Το UniProt (UniProt Knowledgebase) που αποτελεί την κεντρική βάση δεδομένων των πρωτεϊνών με ακριβή αλληλουχία και εκτενή επιμέλεια.
- iii) Το τμήμα UniRef (UniProt NREF databases) που παρέχει επιπλέον πληροφορίες με σκοπό την αξιοποίηση των αλληλουχιών της UniProt για διαφορετικές αναλύσεις.

Η ταξινόμηση των πρωτεϊνών από τη UniProt γίνεται σε οικογένειες και υπερ-οικογένειες μέσω της αναγνώρισης των επικρατειών τους, ενώ η επιμέλεια τους γίνεται μέσω αλγορίθμων, αλλά και χειροκίνητα, με προτεραιότητα στις πρωτεΐνες των οποίων η δομή και η λειτουργία έχουν αποκαλυφθεί.

Η UniProt στην παρούσα εργασία χρησιμοποιήθηκε για την απόκτηση πληροφοριών σχετικά με την γονιδιακή οντολογία των οσοδεσμευτικών πρωτεϊνών και των γευστικών και ιονοτροπικών υποδοχέων, ώστε να αποκαλυφθεί ο ρόλος τους στις κυτταρικές διεργασίες και η μοριακή τους λειτουργία.

3. ΑΠΟΤΕΛΕΣΜΑΤΑ

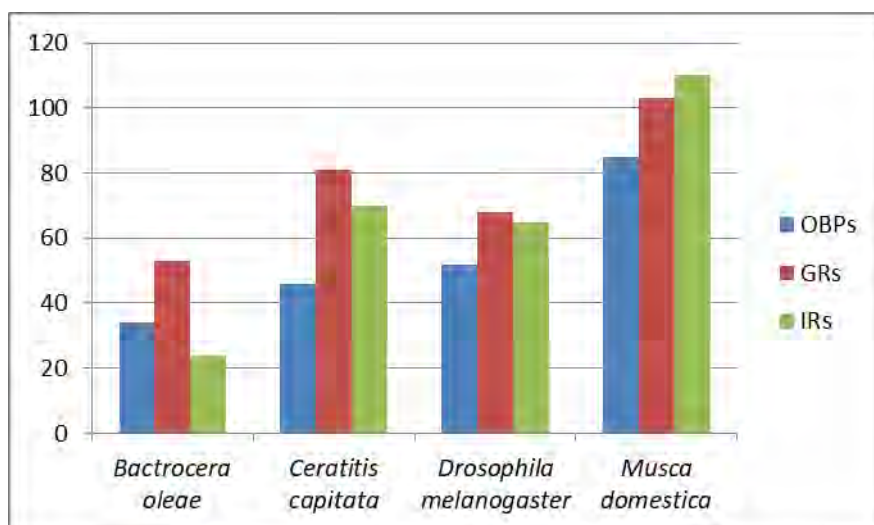
3.1. Ανάκτηση των αλληλουχιών των γονιδίων χημειοαντίληψης του δάκου και συγγενικών ειδών από την NCBI

Για τις ανάγκες της παρούσας εργασίας χρησιμοποιήθηκε το γονιδίωμα του δάκου όπως εντοπίζεται στην NCBI και το Apollo i5k. Οι δύο βάσεις δεδομένων προτείνουν διαφορετικά μοντέλα πρόγνωσης γονιδίων και από αυτές αναζητήθηκαν τα γονίδια χημειοαντίληψης του δάκου. Συνολικά, εντοπίστηκαν 37 αλληλουχίες γονιδίων OBPs, 51 αλληλουχίες γονιδίων GRs και 21 ιονοτροπικών υποδοχέων. Οι αλληλουχίες αυτές παρατίθενται στους πίνακες 5.1, 5.2 και 5.3 του Παραρτήματος με τους μοναδικούς XM κωδικούς που τις συνοδεύουν στην NCBI.

Για τις περαιτέρω συγκριτικές αναλύσεις επιλέχθηκαν τα συγγενικά είδη *Drosophila melanogaster*, *Ceratitis capitata* και *Musca domestica*. Ανακτήθηκαν οι αμινοξικές αλληλουχίες των ορθόλογων πρωτεϊνών της *D. melanogaster* όπως προκύπτουν από τις πλήρως επιμελημένες αλληλουχίες των αντίστοιχων γονιδίων της από τη βάση δεδομένων της NCBI, ενώ οι αλληλουχίες των *Ceratitis capitata* και *M. domestica* από τις μελέτες των Papanicolaou *et al* (2016) και Scott *et al* (2014) αντίστοιχα.

Σε αντίθεση με το δάκο, εντοπίζονται 68 GRs στη *D. melanogaster*, 81 στη *C. capitata*, και 110 στη *M. domestica*. Αντίστοιχα, για τους IRs ανιχνεύονται, 65 στη *D. melanogaster*, 110 στη *M. domestica* και 70 στη *C. capitata*. Τέλος, όσον αφορά τις OBPs, έχουν εντοπιστεί 52 OBPs στη *D. melanogaster*, 45 στη *C. capitata*, και 87 στη *M. domestica*.

Ο αριθμός των πρωτεϊνών αυτών συγκριτικά στα τέσσερα είδη φαίνεται στο ραβδόγραμμα της εικόνας 3.1. Στο δάκο παρατηρήθηκε ότι ο αριθμός των πρωτεϊνών αυτών και στις τρεις υπό εξέταση κατηγορίες σε σχέση με τα υπόλοιπα 3 είδη είναι αρκετά μικρότερος.



Εικόνα 3.1.: Παράθεση του πλήθους των OBPs, των γευστικών και των ιονοτροπικών υποδοχέων στο δάκο και στα τρία υπό σύγκριση είδη.

3.2. Επεξεργασία και ανάλυση των συναινετικών γονιδιακών μοντέλων των GRs, IRs και OBPs του δάκου

Αφού ανακτήθηκαν οι αλληλουχίες των γονιδίων των GRs, IRs και OBPs, ακολούθησε η επεξεργασία και χειροκίνητη επιμέλεια των συναινετικών γονιδιακών μοντέλων, που έχουν προβλεφθεί από τους αλγόριθμους της NCBI και της JAMg, στην πλατφόρμα Apollo. Τα συναινετικά γονιδιακά μοντέλα για κάθε ένα από τα παραπάνω γονίδια εντοπίστηκαν στο περιβάλλον του Apollo με αναζήτηση του μοναδικού XM κωδικού που τους αντιστοιχεί. Η επεξεργασία των συναινετικών γονιδιακών μοντέλων ολοκληρώθηκε με την ανάλυση της τοπολογίας των γευστικών και ιονοτροπικών υποδοχέων και των αμινοξικών μοτίβων των OBPs.

3.2.1. Επεξεργασία και ανάλυση των συναινετικών γονιδιακών μοντέλων των GRs του δάκου

- ΧΕΙΡΟΚΙΝΗΤΗ ΕΠΙΜΕΛΕΙΑ ΤΩΝ ΣΥΝΑΙΝΕΤΙΚΩΝ ΓΟΝΙΔΙΑΚΩΝ ΜΟΝΤΕΛΩΝ ΤΩΝ GRs ΣΤΟ APOLLO

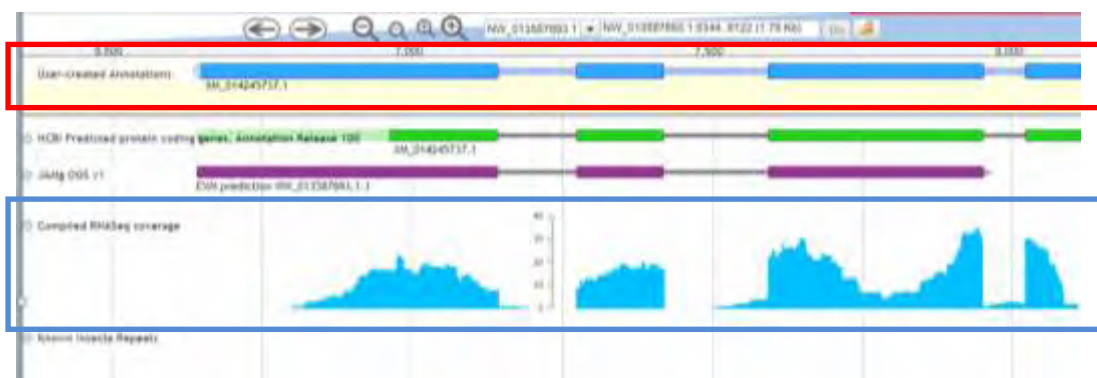
Συνολικά ανακτήθηκαν και επιμελήθηκαν χειροκίνητα 51 GRs. Από το σύνολο αυτό, παρατηρήθηκε συμφωνία μεταξύ των δύο μοντέλων για 8 γονίδια όσον αφορά την προτεινόμενη οργάνωσή τους (βλ. Παράρτημα, Πίνακας 5.2.). Στις περιπτώσεις αυτές μικρές διαφορές παρατηρήθηκαν όσον αφορά το μήκος των 5' και 3' αμετάφραστων περιοχών. Για τον ακριβή καθορισμό των ορίων των αμετάφραστων περιοχών, η επιμέλεια βασίστηκε στα δεδομένα της RNA αλληλούχησης. Επιπλέον, για 27 γονίδια τα δεδομένα αλληλούχησης δεν επαρκούσαν για την επιβεβαίωση των συναινετικών μοντέλων (βλ. Παράρτημα, Πίνακας 5.2.). Τέλος, εντοπίστηκαν 43 περιπτώσεις γονιδίων, στις οποίες παρατηρήθηκαν ασυμφωνίες μεταξύ των προτεινόμενων μοντέλων. Για τις τελευταίες περιπτώσεις, παρατίθενται αναλυτικά παραδείγματα παρακάτω.

● *Ασυμφωνία συναινετικών γονιδιακών μοντέλων*

Στις περιπτώσεις που παρατηρούνταν ασυμφωνία μεταξύ των συναινετικών γονιδιακών μοντέλων των NCBI και JAMg, ως πραγματικό επιλέγονταν αυτό που συμφωνούσε κάθε φορά με τα δεδομένα RNAseq. Η επιλογή του πραγματικού γονιδιακού μοντέλου επιβεβαιώνονταν μέσα από συγκριτική ανάλυση.

Ένα παράδειγμα φαίνεται στην εικόνα 3.2.a, όπου τα συναινετικά γονιδιακά μοντέλα ενός από τους γευστικούς υποδοχείς δεν συμφωνούν μεταξύ τους αναφορικά με την παρουσία ή μη του τελευταίου εξονίου. Ως το πραγματικό γονιδιακό μοντέλο επιλέχθηκε αυτό της NCBI (Εικ. 3.2.a, οργάνωση γονιδίου με πράσινο χρώμα), καθώς μετά από σύγκριση με τα δεδομένα αλληλούχησης του συνολικού RNA (RNAseq) προκύπτει ότι ο υψηλός αριθμός αναγνώσεων (Εικ. 3.2.a, μπλε πλαίσιο) υποστηρίζει την ύπαρξη εξονίου στη συγκεκριμένη περιοχή, αλλά και την ύπαρξη ενός ιντρονίου που προηγείται (χαμηλός αριθμός αναγνώσεων). Συνεπώς, το πραγματικό γονιδιακό μοντέλο για το συγκεκριμένο γονίδιο είναι αυτό που βασίστηκε

στην πρόγνωση του αλγορίθμου της NCBI και συμπεριλαμβάνει ένα επιπλέον εξόνιο λαμβάνοντας υπόψη και τα δεδομένα RNAseq.



Εικόνα 3.2.a.: Γευστικός υποδοχέας με κωδικό γονιδίου XM_014245737.1 του οποίου τα συναινετικά γονιδιακά μοντέλα βρίσκονται σε ασυμφωνία ως προς το τελευταίο εξόνιο. Ως πραγματικό, σύμφωνα με τα ολικά δεδομένα RNAseq (μπλε πλαίσιο), επιλέχθηκε το μοντέλο της NCBI. Η επιλογή αυτή εντοπίζεται στο κόκκινο πλαίσιο που αντιστοιχεί στα «User-created annotations»

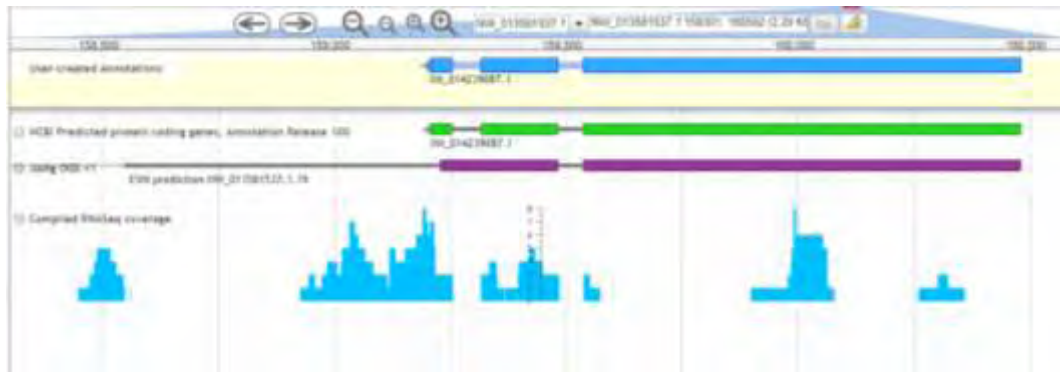
Για να επιβεβαιώσουμε ότι επιλέξαμε και επεξεργαστήκαμε σωστά το γονιδιακό μοντέλο του παραπάνω γευστικού υποδοχέα, ακολούθησε πολλαπλή στοίχιση της αμινοξικής του αλληλουχίας στο Clustal Omega με τις ορθόλογες των *C. capitata* και *M. domestica*. Από τη σύγκριση αυτή, στην οποία συμπεριλήφθηκαν και τα δυο γονιδιακά μοντέλα, προέκυψε ότι το πραγματικό γονιδιακό μοντέλο είναι αυτό της NCBI. Συγκεκριμένα, με βάση τις ορθόλογες σχέσεις των πρωτεϊνών αυτών αναμένουμε στην πρωτεΐνη του δάκου την ύπαρξη και των τελευταίων αμινοξέων που παρατηρούνται στις πρωτεΐνες των υπόλοιπων δύο ειδών (Εικ. 3.2.b, κόκκινο πλαίσιο). Το συναινετικό μοντέλο της NCBI συμπεριλαμβάνει τα αμινοξέα αυτά, ενώ απουσιάζουν από το μοντέλο της JAMg. Επιβεβαιώνεται, λοιπόν, ότι το πραγματικό γονιδιακό μοντέλο είναι αυτό που προτείνεται από τη βάση δεδομένων της NCBI.

Mdom	MAFNATVASR---EVASPRVMPALTPSQQFLHDELRYREKLNFLADMDOVHLSDYYVRK	57
CcapGr1_ZFIX	M5FVWKHADGSTOIVEKPFRIIPLFHPGQRFLEDEQRLRLQHEKRAL-DGGKSFDDYIRK	59
Boleae_NCBI-model	-----	0
Boleae_JAPg-model	-----	0
Mdom	EETVDDPELLEKDESEYHTTKSLULVQZNGVHPIDHNPFPKPLPRTQYSMTSKQVLSGH	117
CcapGr1_ZFIX	QSTLDDARLLEHDSFYKTKSLVLFQZNGZHPIDHNPFDHILPRTQYSMSRQVLVAI	119
Boleae_NCBI-model	-----KGIHPIDHNPFDHILPRTQYSMTSRQMLVAI	31
Boleae_JAPg-model	-----KGIHPIDHNPFDHILPRTQYSMTSRQMLVAI	31
Mdom	FVVVIQTTVVIFVLQERVVKFVTNSETRFDEADYVWIFISLLFTNIFLLPVASWHHGQVA	177
CcapGr1_ZFIX	CVFSLETIVVWVLRARVKSFEIQPKHFDVAIYNIIFISLLFTNIFLLPVASWHHGQVA	179
Boleae_NCBI-model	CIFSLETFIVWVLRARVTHFIEQPKHFDVAIYNIIFISLLFTNIFLLPVASWHHGPEVA	91
Boleae_JAPg-model	CIFSLETFIVWVLRARVTHFIEQPKHFDVAIYNIIFISLLFTNIFLLPVASWHHGPEVA	91
Mdom	IFKNWNTNYQKFKVVTGTPIVFPHLYPLTAGLCTFSWTLSEILINLSQVFLQPDFEFWYT	237
CcapGr1_ZFIX	IFKNWNTNYQKFMVVTGSPIVFPHLYALTAGLCTFSWTLSEIIVWVNSQVFLQPDFEFWYT	239
Boleae_NCBI-model	IFKNWNTNYQKFMVVTGSPIVFPHLYRLTAGLCTFSWALSVAWVNSQVFLQPDFEFWYT	151
Boleae_JAPg-model	IFKNWNTNYQKFMVVTGSPIVFPHLYRLTAGLCTFSWALSVAWVNSQVFLQPDFEFWYT	151
Mdom	FAYYPIAHLNCFCSLWYFNCHAFGTASRALSRSLQKTLRSEKPAEKLTEFHYLVVDLSH	297
CcapGr1_ZFIX	FAYYPIAHLNCFCSLWYFNCHAFGTASEALAKNELTLKSGKPAEKLTEFHYLVVDLSL	299
Boleae_NCBI-model	FAYYPIAHLNCFCSLWYFNCHAFGTASEDLAKHLAITLKSOKPAEKLTEFHYLVVDLSL	211
Boleae_JAPg-model	FAYYPIAHLNCFCSLWYFNCHAFGTASEDLAKHLAITLKSOKPAEKLTEFHYLVVDLSL	211
Mdom	HWQQLGRAYSNMVGYYCLVWFFTTIATYGSISEIDHGATYKEVGLFVIVFYCMSLLVI	357
CcapGr1_ZFIX	HWQQLGRAYSNMVGYYCLVWFFTTLTAAYGSISEIDHGATYKEIGLVIVHYCMSLLVI	359
Boleae_NCBI-model	HWQQLGRAYSNMVGYYCLVWFFTTLTAAYGSISEIDHGATYKEIGLVIVHYCMSLLVI	271
Boleae_JAPg-model	HWQQLGRAYSNMVGYYCLVWFFTTLTAAYGSISEIDHGATYKEIGLVIVHYCMSLLVI	271
Mdom	ICNEAHQASRKVGLDFQTKLINVHLIALDTASQREVQMFVAISKTPPEWILDOYANINR	417
CcapGr1_ZFIX	ICNEAHQASRKVGLDFQTKLINVHLIALDTASQREVQMFVAISKTPPEWILDOYANINR	419
Boleae_NCBI-model	ICNEAHQASRKVGLDFQTKLINVHLIALDTASQREVQMFVAISKTPPEWILDOYANINR	331
Boleae_JAPg-model	ICNEAHQASRKVGLDFQTKLINVHLIALDTASQREVQMFVAISKTPPEWILDOYANINR	331
Mdom	ELITSNVSFNATYLVNLLQFKITEQRGLRSQQATAMP	455
CcapGr1_ZFIX	ELFSSNLTFNATYLVNLLQFKITEQRGSHSQDFNTM	457
Boleae_NCBI-model	ELF5NLTFNATYLVNLLQFKITEQRGLRSHSQSH--	367
Boleae_JAPg-model	ELFSS-----	336
	**::L*	

Εικόνα 3.2.b.: Πολλαπλή στοίχιση του γευστικού υποδοχέα με κωδικό γονιδίου XM_014245737.1 με τις ορθόλογες των *C. capitata* και *M. domestica*.

- Μη επαρκή μεταγραφικά δεδομένα για τον καθορισμό του πραγματικού γονιδιακού μοντέλου

Η πολλαπλή στοίχιση των γευστικών υποδοχέων του δάκου με τους ορθόλογους για την ανίχνευση της πραγματικής τους δομής, πραγματοποιούνταν και στις περιπτώσεις όπου εμφανίζονταν πολύ χαμηλή έκφραση. Σε αυτές τις περιπτώσεις, η έκφραση των εξονίων δεν είναι εμφανής από τα δεδομένα RNAseq και οι πληροφορίες δεν ήταν επαρκείς για τον αποτελεσματικό καθορισμό των ορίων μεταξύ εξονίων και ιντρονίων. Ένα παράδειγμα τέτοιας περίπτωσης αποτελεί ο γευστικός υποδοχέας με κωδικό γονιδίου XM_014239087.1 που παρουσιάζεται στις εικόνες 3.2.c, 3.2.d. Λαμβάνοντας υπόψη τις ορθόλογες σχέσεις των πρωτεϊνών αυτών αναμένουμε υψηλή ομολογία των αλληλουχιών αυτών, όπως παρατηρείται και στο γονιδιακό μοντέλο της NCBI. Το γονιδιακό μοντέλο της JAMg συμπεριλαμβάνει αμινοξέα που δεν αντιστοιχούν στο συγκεκριμένο γευστικό υποδοχέα (Εικ. 3.2d, κόκκινο πλαίσιο). Έτσι, επιβεβαιώνεται ότι το πραγματικό γονιδιακό μοντέλο είναι αυτό που προτείνεται από τη βάση δεδομένων της NCBI.



Εικόνα 3.2.c.: Γευστικός υποδοχέας με κωδικό γονιδίου XM_014239087.1 που εμφανίζει χαμηλή έκφραση και τα δεδομένα RNA αλληλούχησης δεν επαρκούν για την επιμέλεια του.

h3dm	HEQDLKFLVNCETAFQEVNINQTEYSSRHHKICG-----TIYTSFLHNSL	44
Ccapd38	HEQDLQALRYGVYVDFEYVNAATPSTERSVSTELCHATQLYRKHPIYHRAVYVFLACIT	60
B.Oleae_JAMg-model	HEQDLQALRYCTFFQGVVSAHEVTDDEKSI---DGRQSAISRAVHLQRAVYVCLFLSI	58
B.Oleae_NCBI-model	HEQDLQALRYCTFFQGVVSAHEVTDDEKSI---DGRQSAISRAVHLQRAVYVCLFLSI	54
h3dm	SEIEVLDVYHSPLEEDFZIDWVSAFVYVSDQPSHLVYHRESLAKQKHTEFRLLEDEE	104
Ccapd38	CFAPALSLVNHITPPDQLNHAATCILPTSEYVYHSLLEALREKREDFPKLLVYDIE	120
B.Oleae_JAMg-model	CVYFACLSLCHITTPPGLLNHAAVSLFATFPITNAPILREAVKQQQHWPLKALQAEI	114
B.Oleae_NCBI-model	CVYFACLSLCHITTPPGLLNHAAVSLFATFPITNAPILREAVKQQQHWPLKALQAEI	116
h3dm	VAPFLKRTQDEIRDLAGLRRLEPFLAAEISELRLIIPZQIHTS-LNEDQQVYKVALPAIL	168
Ccapd38	ASLRIELKHWYETRQLAQTRDFLKPQLVSLFLCFFPIVPSLIVYVSDQYFQGLWYIC	180
B.Oleae_JAMg-model	ASLRIELKHWYETRQLAQTRDFLKPQLVSLFLCFFPIVPSLIVYVSDQYFQGLWYIC	176
B.Oleae_NCBI-model	ASLRIELKHWYETRQLAQTRDFLKPQLVSLFLCFFPIVPSLIVYVSDQYFQGLWYIC	178
h3dm	AVRHRFLQGMVYELVRYVLSLNRKIQGVVCLRETFAGLDVNVYQLETLEYLHHEK	223
Ccapd38	TARLRITLQQLVYVRELHRYLRDLCLIKRQHEFHMAPSCLLDIDYRDLTLCGLLWKE	240
B.Oleae_JAMg-model	DRHRYTQQLVYVRELHRYLRDLCLIQQLVAVYHRTQKCLLDIFRYRDLTLCGLLWKE	236
B.Oleae_NCBI-model	DRHRYTQQLVYVRELHRYLRDLCLIQQLVAVYHRTQKCLLDIFRYRDLTLCGLLWKE	238
h3dm	LVSSIVYAPHELHEPQASMPAVTASVLDCTQKIVVCLLALDALPSPAST-VLSSTETI	282
Ccapd38	TVTLHEATQLVYVAGLSLGGIISCVLDLSSAFVYVQLQSFDPHGRVYVLELWAFV	300
B.Oleae_JAMg-model	TVHLYEASELLNAPAGHSLFGQVITSVLDLSSAFVYVQLQSFDPHGRVYVLELWAFV	296
B.Oleae_NCBI-model	TVHLYEASELLNAPAGHSLFGQVITSVLDLSSAFVYVQLQSFDPHGRVYVLELWAFV	298
h3dm	PLSLHYRPFYTCQLVQ-----ECRLTALVTRILNVSQSNMCLFLQGVKSLVMD	336
Ccapd38	PVTVLVNEICVYCDQCHLLNWAYFQKTLTQLLSLQIYSS-----TSGEISRYVLELQ	355
B.Oleae_JAMg-model	PVVALICEVYLCDSQRL-----ICENVAPLSQLTRLS-----KASARQVQLVLDQ	344
B.Oleae_NCBI-model	PVVALICEVYLCDSQRL-----ICENVAPLSQLTRLS-----KASARQVQLVLDQ	346
h3dm	PSLQLLHQZIVVYVGRFFPDLQCFIHWELIVTHLIELIQFTKSDHSD--HWYQTEI	381
Ccapd38	PSHQZLQZIEVSAACNFKLSMRFVHSICTAPWHLVLSIQF-----	387
B.Oleae_JAMg-model	FANQLQZQZIEIATFFTLDFRFINVSLQIEYKFFVYSKMLYPLDARHNVAPALDI	404
B.Oleae_NCBI-model	FANQLQZQZIEIATFFTLDFRFINVSLQIEYKFFVYSKMLYPLDARHNVAPALDI	396
h3dm	QETPMV-----	397
Ccapd38	QTSANHPHQQTTELYSARTRHSAPSQTKALFARLQDNYVALLLWAFVYVNSLYC	387
B.Oleae_JAMg-model	QTSANHPHQQTTELYSARTRHSAPSQTKALFARLQDNYVALLLWAFVYVNSLYC	404
B.Oleae_NCBI-model	QTSANHPHQQTTELYSARTRHSAPSQTKALFARLQDNYVALLLWAFVYVNSLYC	386
h3dm	-----	397
Ccapd38	RRRIPSAPLTLVATIFPVAETVHLIIEATLWQDQHEAPLLQLQZAPAFKLRQF	387
B.Oleae_JAMg-model	RRRIPSAPLTLVATIFPVAETVHLIIEATLWQDQHEAPLLQLQZAPAFKLRQF	524
B.Oleae_NCBI-model	RRRIPSAPLTLVATIFPVAETVHLIIEATLWQDQHEAPLLQLQZAPAFKLRQF	386
h3dm	-----	397
Ccapd38	IEQQQFVILNRRHLYLLTSLNHLWFLITFHWLQVSDHVVYDGHFIIITHVNIQLLI	387
B.Oleae_JAMg-model	IEQQQFVILNRRHLYLLTSLNHLWFLITFHWLQVSDHVVYDGHFIIITHVNIQLLI	584
B.Oleae_NCBI-model	IEQQQFVILNRRHLYLLTSLNHLWFLITFHWLQVSDHVVYDGHFIIITHVNIQLLI	386
h3dm	-----	397
Ccapd38	VLRVLRHYRCLCKQLCQIVAVYVTAHNSQPLDYNGAKLLTFLYLLWKEIYALVYEAH	387
B.Oleae_JAMg-model	VLRVLRHYRCLCKQLCQIVAVYVTAHNSQPLDYNGAKLLTFLYLLWKEIYALVYEAH	644
B.Oleae_NCBI-model	VLRVLRHYRCLCKQLCQIVAVYVTAHNSQPLDYNGAKLLTFLYLLWKEIYALVYEAH	386

Εικόνα 3.2.d.: Πολλαπλή στοίχιση της αμινοξικής αλληλουχίας του γευστικού υποδοχέα με κωδικό γονιδίου XM_014239087.1 (βλέπε εικόνα 3.2.c) με τις ορθόλογες των *C. capitata* και *M. domestica*. Με κόκκινο πλαίσιο υποδεικνύεται η περιοχή διαφοροποίησης των μοντέλων JAMg και NCBI.

- ΕΝΤΟΠΙΣΜΟΣ ΤΩΝ GRs ΣΤΑ ΙΚΡΙΩΜΑΤΑ ΤΟΥ ΓΟΝΙΔΙΩΜΑΤΟΣ

Οι GRs εντοπίζονται συνολικά σε 45 ικρίωματα του γονιδιώματος. Συγκεκριμένα, στις 41 από τις 51 περιπτώσεις γευστικών υποδοχέων παρατηρήθηκε ένας υποδοχέας σε ένα ικρίωμα. Ωστόσο, για τους υπόλοιπους δέκα GRs, σε δύο περιπτώσεις ανιχνεύθηκαν δύο υποδοχείς σε ένα ικρίωμα και σε δύο περιπτώσεις ανιχνεύθηκαν τρεις υποδοχείς σε ένα ικρίωμα.

- Περιπτώσεις όπου δύο ή τρεις υποδοχείς εντοπίζονται σε ένα ικρίωμα

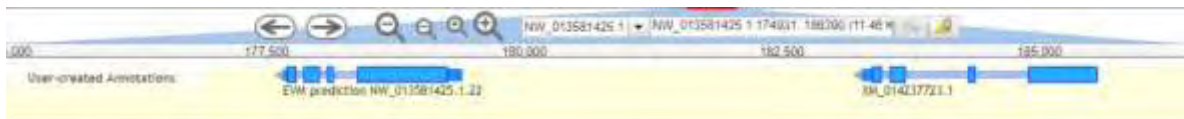
Στις τέσσερις αυτές περιπτώσεις, σύμφωνα με τις συντεταγμένες⁵ τους, τα γονίδια εντοπίστηκαν σε πολύ κοντινή απόσταση μεταξύ τους στο ίδιο ικρίωμα. Επιπλέον, παρατηρήθηκε ότι χαρακτηρίζονται από ίδιο αριθμό εξονίων και παρόμοιο αριθμό αμινοξέων. Από τη στοίχιση με Clustal Omega προκύπτει, επίσης, ότι τα γονίδια αυτά φέρουν πολλές ομοιότητες στην αλληλουχία τους, ενώ οι εννιά από τους δέκα φέρουν, επιπλέον, τον ίδιο προσανατολισμό. Τέλος, για δύο από τους δέκα υποδοχείς, αν και διαφέρουν ως προς τον αριθμό των εξονίων τους καθώς παρατηρήθηκε ότι αποτελούνται από ένα και τέσσερα εξόνια αντίστοιχα, οι αλληλουχίες τους αποδείχθηκε μετά από πολλαπλή στοίχιση ότι εμφανίζουν υψηλή ομολογία. Αναλυτικά τα δεδομένα αυτά παρουσιάζονται στον Πίνακα 3.1.

Οι περιπτώσεις των γονιδίων που εντοπίζονται σε γειτονικές περιοχές στο ίδιο ικρίωμα του γονιδιώματος, καθώς και οι σχετικές τους θέσεις στο γονιδίωμα φαίνονται παρακάτω:

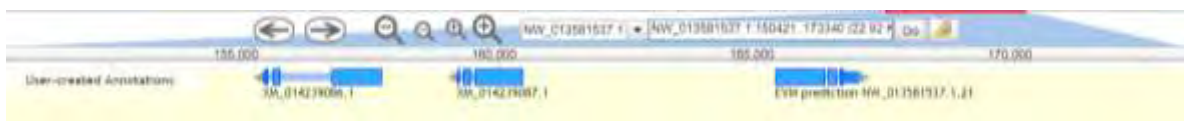
XM_014236435.1, XM_014236437.1, XM_014236438.1



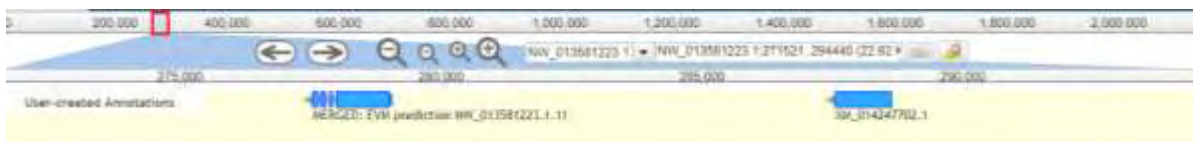
XM_014237722.1, XM_014237723.1



XM_014239086.1, XM_014239087.1, XM_014239088.1



XM_014247702.1, XM_014247700.1



⁵ Οι συντεταγμένες ενός γονιδίου αντιστοιχούν στη θέση των βάσεων που οριοθετούν το γονίδιο στη μελετούμενη γονιδιωματική περιοχή.

Τα γονίδια αυτά συγκεντρώθηκαν στον παρακάτω πίνακα με τα κοινά ικρίωματα του γονιδιώματος όμοια χρωματισμένα.

Πίνακας 3.1 Γονίδια GRs του δάκου που εντοπίζονται σε κοινά ικρίωματα

Γονίδια GRs στο δάκο της ελιάς που εντοπίζονται σε ίδια ικρίωματα του γονιδιώματος							
Όνομα NCBI	Κωδικός XM	Κωδικός XP	Περιοχή γονιδιώματος	Συντεταγμένες (bp)	aa	Αριθμός εξονίων	Αριθμός ιντρονίων
1. <i>B.oleae</i> GR 10a	XM_014236435.1	XP_014091910.1	NW_013581352.1	259501-260990	408	2	1
2. <i>B.oleae</i> GR 36b	XM_014236437.1	XP_014091912.1	NW_013581352.1	297795-299196	402	2	1
3. <i>B.oleae</i> GR 22e-like	XM_014236438.1	XP_014091913.1	NW_013581352.1	318361-320128	425	2	1
4. <i>B.oleae</i> GR 98b	XM_014237722.1	XP_014093197.1	NW_013581425.1	177620-179359	402	4	3
5. <i>B.oleae</i> GR 98b	XM_014237723.1	XP_014093198.1	NW_013581425.1	183237-185530	333	4	3
6. <i>B.oleae</i> GR 39b	XM_014239086.1	XP_014094561.1	NW_013581537.1	155435-157751	399	3	2
7. <i>B.oleae</i> GR 39b	XM_014239087.1	XP_014094562.1	NW_013581537.1	159204-160481	386	3	2
8. <i>B.oleae</i> GR 39b	XM_014239088.1	XP_014094563.1	NW_013581537.1	165365-167007	408	3	2

9. <i>B.oleae</i> GR 32a-like	XM_014247702 .1	XP_014103177 .1	NW_013581223.1	287475- 288557	360	1	0
10. <i>B.oleae</i> GR 32a-like	XM_014247700 .1	XP_014103175. 1	NW_013581223.1	277461- 278996	429	4	3

- **ΑΝΑΛΥΣΗ ΤΟΠΟΛΟΓΙΑΣ ΤΩΝ ΕΠΙΜΕΛΗΜΕΝΩΝ GRs ΜΕ ΤΜΗΜΜ**

Οι γευστικοί υποδοχείς των εντόμων αποτελούνται φυσιολογικά από επτά διαμεμβρανικές περιοχές, φέρουν το αμινοτελικό τους άκρο στο εσωτερικό της μεμβράνης και το καρβοξυτελικό τους στο εξωτερικό. Για να διαπιστώσουμε αν το παραπάνω ισχύει και για τους γευστικούς υποδοχείς του δάκου, οι επιμελημένες αλληλουχίες τους μελετήθηκαν με τη χρήση του προγράμματος ΤΜΗΜΜ. Τα αποτελέσματα της αναζήτησης των επιμελημένων GRs στο ΤΜΗΜΜ (βλ. Παράρτημα, Πίνακας 5.4) συνοψίζονται στο ραβδόγραμμα της εικόνας 3.2.ε.

Από τους 51 γευστικούς υποδοχείς του δάκου, οι 17 φέρουν τις αναμενόμενες επτά διαμεμβρανικές περιοχές. Από τους υπόλοιπους, έξι διαθέτουν δύο, τρεις, ή τέσσερις διαμεμβρανικές περιοχές, ενώ 18 αποτελούνται από πέντε ή έξι διαμεμβρανικές περιοχές. Τέλος, εννιά γευστικοί υποδοχείς αποτελούνται από οκτώ διαμεμβρανικές περιοχές, ενώ μόνο ένας γευστικός υποδοχέας φέρει εννιά.



Εικόνα 3.2.ε.: Αριθμός διαμεμβρανικών περιοχών του συνόλου των γευστικών υποδοχέων του δάκου

3.2.2. Επεξεργασία των συναινετικών γονιδιακών μοντέλων των IRs του δάκου

- ΧΕΙΡΟΚΙΝΗΤΗ ΕΠΙΜΕΛΕΙΑ ΤΩΝ ΣΥΝΑΙΝΕΤΙΚΩΝ ΓΟΝΙΔΙΑΚΩΝ ΜΟΝΤΕΛΩΝ ΤΩΝ IRs ΜΕ ΤΟ APOLLO

Η επιμέλεια των συναινετικών γονιδιακών μοντέλων των IRs πραγματοποιήθηκε με τον ίδιο τρόπο με τους γευστικούς υποδοχείς στο περιβάλλον του Apollo με την αξιοποίηση των δεδομένων RNAseq. Κάθε ιονοτροπικός υποδοχέας εντοπιζονταν με αναζήτηση του μοναδικού XM κωδικού που του αντιστοιχεί.

Συνολικά ανακτήθηκαν και επιμελήθηκαν χειροκίνητα 21 IRs. Συμφωνία μεταξύ των συναινετικών γονιδιακών μοντέλων παρατηρήθηκε για πέντε ιονοτροπικούς υποδοχείς (βλ. Παράρτημα, Πίνακας 5.3.), όσον αφορά την προτεινόμενη οργάνωσή τους στο γονιδίωμα. Στις περιπτώσεις αυτές μικρές διαφορές παρατηρήθηκαν στο μήκος των 5' και 3' αμετάφραστων περιοχών και η επιμέλεια τους πραγματοποιήθηκε με βάση τα δεδομένα RNAseq. Επιπλέον, για εννιά γονίδια τα δεδομένα αλληλούχησης δεν επαρκούσαν για την επιβεβαίωση των συναινετικών μοντέλων (βλ. Παράρτημα, Πίνακας 5.3.). Ένα παράδειγμα περίπτωσης ανεπαρκών δεδομένων RNAseq παρατίθεται αναλυτικά παρακάτω. Τέλος, σε 16 περιπτώσεις γονιδίων δε συμφωνούσαν τα συναινετικά γονιδιακά μοντέλα των NCBI και JAMg.

- *Ανεπαρκή μεταγραφικά δεδομένα για καθορισμό του πραγματικού γονιδιακού μοντέλου*

Όπως και με τους γευστικούς υποδοχείς, όταν η έκφραση των γονιδίων ήταν πολύ χαμηλή και τα δεδομένα RNAseq δεν επαρκούσαν, για την επιμέλεια της δομής των γονιδίων αυτών αξιοποιούνταν τα αποτελέσματα μετά από πολλαπλή στοίχιση των αμινοξικών τους αλληλουχιών στο Clustal Omega με τις αντίστοιχες ορθόλογες πρωτεΐνες των *C. capitata*, *D. melanogaster*, και *M. domestica*. Σε αντίθεση με τους γευστικούς υποδοχείς, η έκφραση ήταν χαμηλή για την πλειοψηφία των ιονοτροπικών υποδοχέων και για έναν μεγάλο αριθμό αυτών τα δεδομένα RNA αλληλούχησης δεν επαρκούσαν. Από την πολλαπλή στοίχιση, γενικά, παρατηρήθηκε ότι 12 από τους 21 ιονοτροπικούς υποδοχείς φέρουν ομόλογες αλληλουχίες με τη *C. capitata*, αλλά εμφανίζουν μεγάλες διακυμάνσεις στο μήκος τους, όπως φαίνεται και από το παράδειγμα της εικόνας 3.2.f.

B.Oleae_JAMg-model	-----	0
B.Oleae_NCBI-model	-----	0
Ccap	MNINIFERISEGKLFERPREWHFIYLDPRDRVFKSKKLVDDATKFTLNPKTLCRALRMKDT	240
B.Oleae_JAMg-model	-----	0
B.Oleae_NCBI-model	-----	0
Ccap	YCLKGFTFQGAILLEILRGLIEIKQRNVYMLQSYMMECNATSPNENATTGFDILEQFPLS	300
B.Oleae_JAMg-model	-----	0
B.Oleae_NCBI-model	-----	0
Ccap	DFIYFTTDSFPEDELELLPRLTYSPTININLYSSEHETATELAIWQYGLRKNQTLSP	360
B.Oleae_JAMg-model	-----	0
B.Oleae_NCBI-model	-----	0
Ccap	PRRFRFRIGTVEAMPWNYMKRDPNTDELVLDAYGNPIMEGFCIDSIKKLAERLNFYGLVLP	420
B.Oleae_JAMg-model	-----	0
B.Oleae_NCBI-model	-----	0
Ccap	PTTGEGFSYDKSTGKWDGIVGDLTTGETDFAVTALKMYAEREVIDYIAPYFEQTGISIV	480
B.Oleae_JAMg-model	-----	0
B.Oleae_NCBI-model	-----	0
Ccap	MRKPVQRQTSLFKFHTVLRVEWFSIIAALVGSFAFMWLLLEKISPYSYRNRAAYQYLCRE	540
B.Oleae_JAMg-model	-----	0
B.Oleae_NCBI-model	-----	19
Ccap	FTLRESFWFALTSFTPOGGGEAPKAVSGRVMVAAYWLVVLMLATFTANLAALFTVERMQ -----MLATFTANLAALFTVERMQ	600
B.Oleae_JAMg-model	-----MKFAEDTLYRMMKELTLNVTEDFQRYRIW	29
B.Oleae_NCBI-model	TPVQSLEQLARQSRINVTVEGSGTHQYFINMKFAEDTLYRMMKELTLNVTEDFQRYRIW	79
Ccap	TPVQSLEQLARQSRINVTVEGSGTHQYFINMKFAEDTLYRMMKELTLNVTDDFQRYRIW *****.*****	660
B.Oleae_JAMg-model	DYPIKEQYGTILLAINGSEPVKNAKEGFRKVNENADFAFIHDSSEIKYELTRNCNLTE	89
B.Oleae_NCBI-model	DYPIKEQYGTILLAINGSEPVKNAKEGFRKVNENADFAFIHDSSEIKYELTRNCNLTE	139
Ccap	DYPIKEQYGTILLAINGSEPVKDAKEGFRKVNENADFAFIHDSSEIKYELTRNCNLTE *****.*****	720
B.Oleae_JAMg-model	VGEVFAEQPYAIAIQGSHFADELSYALLELQKDRFFEDLKAKYMNIGTSKLVFSQPTD	149
B.Oleae_NCBI-model	VGEVFAEQPYAIAIQGSHFADELSYALLELQKDRFFEDLKAKYMN-----	185
Ccap	VGEVFAEQPYAIAIQGSHFADELSYALLELQKDRFFEDLKAKYMNISLIKACSVNEEQE *****	780
B.Oleae_JAMg-model	QCPSAGPSAFRHGGEFTRHANTH-----	172
B.Oleae_NCBI-model	-----	185
Ccap	-----GISLES LGGVFIATLFGGLAMVTLVLEIIYRRKYAIMHRFSEVIKVKPASASS	835
B.Oleae_JAMg-model	-----	172
B.Oleae_NCBI-model	-----	185
Ccap	TKQLVPMNPLKKRIATIWHTGDNSELPKTPPPAFDAIKFRGKIKPPSITLGGHEFKPRR	895
B.Oleae_JAMg-model	-----	172
B.Oleae_NCBI-model	-----	185
Ccap	LGARQPSSELDGAEYKRG LPPNRDELPPYTE	928

Εικόνα 3.2.f.: Στοιχίαση ιονοτροπικού υποδοχέα του δάκου (XM_014245834.1) με τον ορθόλογο υποδοχέα της *C. capitata*.

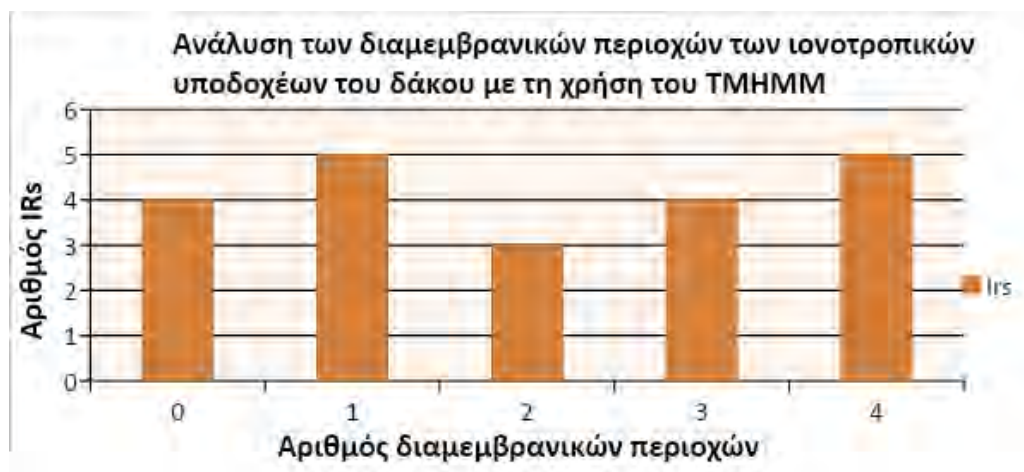
- ΕΝΤΟΠΙΣΜΟΣ ΤΩΝ IRS ΣΤΑ ΙΚΡΙΩΜΑΤΑ ΤΟΥ ΓΟΝΙΔΙΩΜΑΤΟΣ

Συνολικά, οι 21 IRs εντοπίζονται σε 21 ικρίωματα του γονιδιώματος. Επίσης παρατηρήθηκε ότι κανένας από τους ιονοτροπικούς υποδοχείς του δάκου δεν εντοπίζεται σε κοινό γονιδιωματικό ικρίωμα με άλλον (βλ. Παράρτημα, Πίνακας 5.3)

- ΑΝΑΛΥΣΗ ΤΟΠΟΛΟΓΙΑΣ ΤΩΝ IRS ΜΕ ΤΜΗΜΜ

Για την αποκάλυψη της τοπολογίας των IRs στο δάκο οι επιμελημένες αλληλουχίες αναζητήθηκαν στο ΤΜΗΜΜ. Από τους 21 ιονοτροπικούς υποδοχείς του δάκου μόνο οι τέσσερις φέρουν τις αναμενόμενες τρεις διαμεμβρανικές περιοχές. Από τους

υπόλοιπους, όπως φαίνεται και στο ραβδόγραμμα της εικόνας 3.2.g., οι δέκα αποτελούνται από μία ή τέσσερις διαμεμβρανικές περιοχές, τέσσερις δεν φέρουν καμία διαμεμβρανική περιοχή, ενώ οι υπόλοιποι τρεις διαθέτουν μόνο δύο διαμεμβρανικές περιοχές (βλ. Παράρτημα, Πίνακας 5.5).



Εικόνα 3.2.g.: Διαμεμβρανικές περιοχές του συνόλου των ιονοτροπικών υποδοχέων του δάκου

3.2.3. Επεξεργασία των συναινετικών γονιδιακών μοντέλων των OBP's του δάκου

- **ΧΕΙΡΟΚΙΝΗΤΗ ΕΠΙΜΕΛΕΙΑ ΤΩΝ ΣΥΝΑΙΝΕΤΙΚΩΝ ΓΟΝΙΔΙΑΚΩΝ ΜΟΝΤΕΛΩΝ ΤΩΝ OBP'S ΤΟΥ ΔΑΚΟΥ ΣΤΟ ΑΡΟΛΛΟ**

Ομοίως, χειροκίνητη επιμέλεια πραγματοποιήθηκε και για τα γονίδια των 37 OBP's τα οποία εντοπίστηκαν σε 26 scaffolds (βλ. Παράρτημα, Πίνακας 5.1). Συγκεκριμένα, στις 18 από τις 37 περιπτώσεις OBP's εντοπίστηκε ένας υποδοχέας σε ένα ικρίωμα. Από τους υπόλοιπους 19 OBP's, σε δύο περιπτώσεις εντοπίστηκαν σε δύο γονίδια σε κοντινή απόσταση μεταξύ τους σε ένα ικρίωμα και σε πέντε περιπτώσεις τρία γονίδια σε ένα ικρίωμα.

- *Περιπτώσεις όπου δύο ή τρεις υποδοχείς εντοπίζονται στο ίδιο ικρίωμα*

Με στοίχιση με Clustal Omega, με την ίδια διαδικασία που ακολουθήθηκε στους GRs και IRs, προέκυψε ότι φέρουν ομόλογες αλληλουχίες, ενώ ο αριθμός των εξονίων τους είναι, επίσης, ίδιος.

Τα γονίδια αυτά, καθώς και οι θέσεις τους στο γονιδίωμα παρουσιάζονται παρακάτω:

XM_014235205.1, XM_014235206.1, XM_014235214.1:



XM_014235204.1, XM_014235216.1 και XM_014235188.1:



XM_014247738.1, XM_014247739.1:



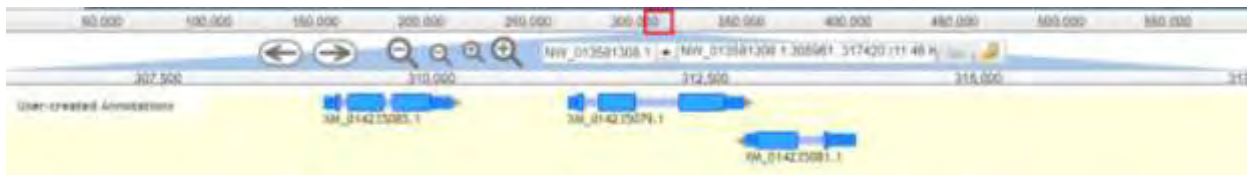
XM_014246634.1, XM_014246630.1:



XM_014246615.1, XM_014246664.1, XM_014246633.1:



XM_014235079.1, XM_014235081.1, XM_014235085.1:



XM_014234681.1, XM_014234656.1, XM_014234657.1:



Από τα παραπάνω 19 γονίδια, αυτά με κωδικό XM_014235205.1, XM_014235206.1 βρίσκονται στην ίδια ακριβώς θέση του ικρίωματος και σε αυτά αντιστοιχούν κοινές συντεταγμένες, σε αντίθεση με τα υπόλοιπα. Το σύνολο των 19 γονιδίων συγκεντρώνεται στον πίνακα που ακολουθεί με τα κοινά ικρίωματα όμοια χρωματισμένα.

Πίνακας 3.2. Γονίδια OBPs του δάκου σε κοινά ικρίωματα

Ομοδεσμευτικές πρωτεΐνες στο δάκο της ελιάς που εντοπίζονται σε κοινά ικρίωματα του γονιδιώματος							
Όνομα NCBI	Κωδικός XM	Κωδικός XP	ΙΚρίωμα γονιδιώματος	Συντεταγμένες (bp)	aa	Αριθμός εξονίων	Αριθμός ιντρονίων
BoleaeObp 99a-like	XM_014235205.1	XP_014090680.1	NW_013581311.1	533400-534221	159	2	1
BoleaeObp 99a-like	XM_014235206.1	XP_014090681.1	NW_013581311.1	533420-534280	159	2	2
BoleaeObp 99b	XM_014235214.1	XP_014090689.1	NW_013581311.1	540782-541457	154	2	1
BoleaeObp 99a-like	XM_014235204.1	XP_014090679.1	NW_013581311.1	418104-419087	149	2	1
BoleaeObp 99a-like	XM_014235216.1	XP_014090691.1	NW_013581311.1	410918-411628	147	2	1
BoleaeObp 99b-like	XM_014235188.1	XP_014090663.1	NW_013581311.1	404363-404898	127	2	1
BoleaeObp A5	XM_014247738.1	XP_014103213.1	NW_013581223.1	1491311-1493288	208	2	1
BoleaeObp A5	XM_014247739.1	XP_014103214.1	NW_013581223.1	1487103-1488201	204	2	1
BoleaeObp 56a-like	XM_014246630.1	XP_014102105.1	NW_013581217.1	454929-455622	165	1	0
BoleaeObp 56a-like	XM_014246634.1	XP_014102109.1	NW_013581217.1	457762-458356	165	1	0

BoleaeObp 56a	XM_014246615.1	XP_014102090.1	NW_013581217.1	409192-409745	138	2	1
BoleaeObp 56d-like	XM_014246533.1	XP_014102008.1	NW_013581217.1	408460-409155	132	2	1
BoleaeObp 56d-like	XM_014246564.1	XP_014102039.1	NW_013581217.1	403607-404297	138	2	1
uncharacte rized protein LOC106617 710	XM_014235079.1	XP_014090554.1	NW_013581308.1	311203-312821	272	3	2
BoleaeObp 99a	XM_014235081.1	XP_014090556.1	NW_013581308.1	312825-313841	142	2	1
uncharacte rized protein LOC106617 715	XM_014235085.1	XP_014090560.1	NW_013581308.1	308972-310144	240	3	2
BoleaeObp 19d-like	XM_014234681.1	XP_014090156.1	NW_013581300.1	586106-588485	157	5	4
BoleaeObp 28a-like	XM_014234656.1	XP_014090131.1	NW_013581300.1	580585-582550	141	5	4
BoleaeObp 19d-like	XM_014234657.1	XP_014090132.1	NW_013581300.1	576116-577968	143	5	4

- **ΑΝΑΛΥΣΗ ΑΜΙΝΟΞΙΚΩΝ ΜΟΤΙΒΩΝ OBPs**

Τα συντηρημένα αμινοξικά μοτίβα είναι σημαντικά στοιχεία των λειτουργικών περιοχών των πρωτεϊνών. Οι OBPs των εντόμων φέρουν συντηρημένα αμινοξικά μοτίβα κυστεϊνών και ταξινομούνται σε τέσσερις υπο-οικογένειες με βάση τον αριθμό των κυστεϊνικών μοτίβων: στην classic, plus-C, minus-C και dimers (βλ. Εισαγωγή). Για να διαπιστώσουμε αν οι OBPs στο δάκο διαθέτουν συντηρημένα αμινοξικά μοτίβα κυστεϊνών που εντοπίζονται και στα είδη *D. melanogaster*, *C. capitata* και *M. domestica* πραγματοποιήθηκε ανάλυση με τη χρήση του προγράμματος MEME.

Τα αποτελέσματα της αναζήτησης έδειξαν την ύπαρξη τριών αμινοξικών μοτίβων στο δάκο (εικόνα 3.2.h). Από τις 37 OBPs, τα μοτίβα 1 και 2 εντοπίζονται σε 34 και 32

πρωτεΐνες αντίστοιχα. Αντίθετα, το μοτίβο 3 εντοπίζεται μόνο σε εννιά πρωτεΐνες (βλ. Παράρτημα, Πίνακας 5.1). Στο δάκο εντοπίστηκαν μόλις δύο διατάξεις των αμινοξικών αυτών μοτίβων από τις οποίες επικρατεί η διάταξη 1-2 (3.2.i). Το μοτίβο 3 εντοπίζεται πάντα στο αμινοτελικό άκρο, ενώ το μοτίβο 1 στο καρβοξυτελικό.



Εικόνα 3.2.h.: Τα τρία αμινοξικά μοτίβα που εντοπίζονται στις OBPs του δάκου μετά από αναζήτηση στον σέρβερ του MEME.



Εικόνα 3.2.i.: Διατάξεις αμινοξικών μοτίβων που εντοπίστηκαν στο δάκο της ελιάς.

Ταξινόμηση των OBPs του δάκου στις 4 υπο-οικογένειες ανάλογα με τα αμινοξικά τους μοτίβα:

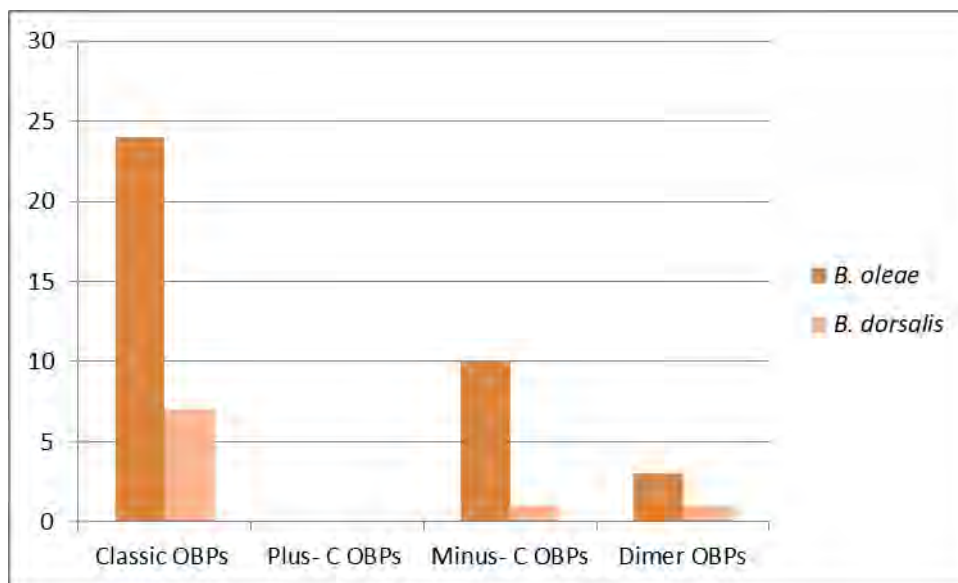
Για την ταξινόμηση των OBPs του δάκου στις τέσσερις υπο-οικογένειες πραγματοποιήθηκε πολλαπλή στοίχιση τους με τη χρήση του αλγόριθμου MUSCLE μεταξύ τους και με τις OBPs των ειδών *C. capitata*, *D. melanogaster* και *M. domestica*. Από την πολλαπλή στοίχιση διαπιστώθηκε ότι από τις 37 OBPs που βρέθηκαν στο δάκο, οι 24 ανήκουν στην υπο-οικογένεια Classic με έξι συντηρημένα μοτίβα κυστεϊνών. Από τις υπόλοιπες, δέκα ταξινομήθηκαν στη Minus-C, τρεις στην Dimer και καμία στην Plus-C υπο-οικογένεια. Οι πρωτεΐνες OBPs και οι υπο-οικογένειες στις οποίες ταξινομήθηκαν φαίνονται συγκεντρωτικά στον πίνακα 3.3.

Πίνακας 3.3. Ταξινόμηση OBPs του δάκου στις 4 υπο-οικογένειες

Υπο-οικογένειες των OBPs			
Classic	Plus-C	Minus-C	Dimer
1.BoleaeObp19d (XM_014244960.1)		1.BoleaeObp99a-like XM_014237824.1	1.XM_014235079.1 (uncharacterized)
2.BoleaeObp19a (XM_014241216.1)		2.BoleaeObp99a-like	2.XM_014235085.1

<i>B.oleae</i> OBPs			XM_014235216.1	(uncharacterized)
	3.BoleaeObp28a-like (XM_014239835.1)		3.BoleaeObp99a-like(3) XM_014235206.1	3. BoleaeObp68 XM_014239953.1
	4.BoleaeObp99a (XM_014239407.1)		4.BoleaeObp99a-like(4) XM_014235205.1	
	5.BoleaeObp99a (XM_014235081.1)		5.BoleaeObp99a-like(5) XM_014235204.1	
	6.BoleaeObp57c (XM_014235849.1)		6.BoleaeObpA10(XM_014231563.1)	
	7.BoleaeObp56a-like (XM_014246630.1)		7. BoleaeObpA5 (XM_014247739.1)	
	8.BoleaeObp99b (XM_014235214.1)		8. BoleaeObpA5 (XM_014247738.1)	
	9.BoleaeObp83a-like (XM_014235082.1)		9. BoleaeObp99b-like XM_014235188.1	
	10.BoleaeObp19d-like (XM_014234681.1)		10. XM_014241021.1(uncharacterized)	
	11.BoleaeObp19d-like(2) (XM_014234657.1)			
	12.BoleaeObp28a-like (XM_014234656.1)			
	13.BoleaeObplush (XM_014233795.1)			
	14.BoleaeObp69a (XM_014233356.1)			
	15.BoleaeObp56a-like (XM_014246634.1)			
	16.BoleaeObp56h-like (XM_014246628.1)			
	17.BoleaeObp56a (XM_014246615.1)			
	18.BoleaeObp56d-like (XM_014246564.1)			
	19.BoleaeObp56d-like(2) (XM_014246533.1)			
	20.BoleaeObp56h-like (XM_014246531.1)			
	21.BoleaeObp84a-like (XM_014242247.1)			
	22.BoleaeObp84a-like(2) (XM_014242235.1)			
	23.BoleaeObp56h-like (XM_014246623.1)			
	24. XM_014248081.1 (uncharacterized)			

Ο αριθμός των γονιδίων σε κάθε υπο-οικογένεια στο δάκο συγκρίνεται με του είδους *B. dorsalis* που ανήκει στο ίδιο γένος (εικόνα 3.3).



Εικόνα 3.3.: Συγκριτική απεικόνιση της κατανομής των γονιδίων των OBPs σε κάθε υπο-οικογένεια στο δάκο και την *B. dorsalis*.

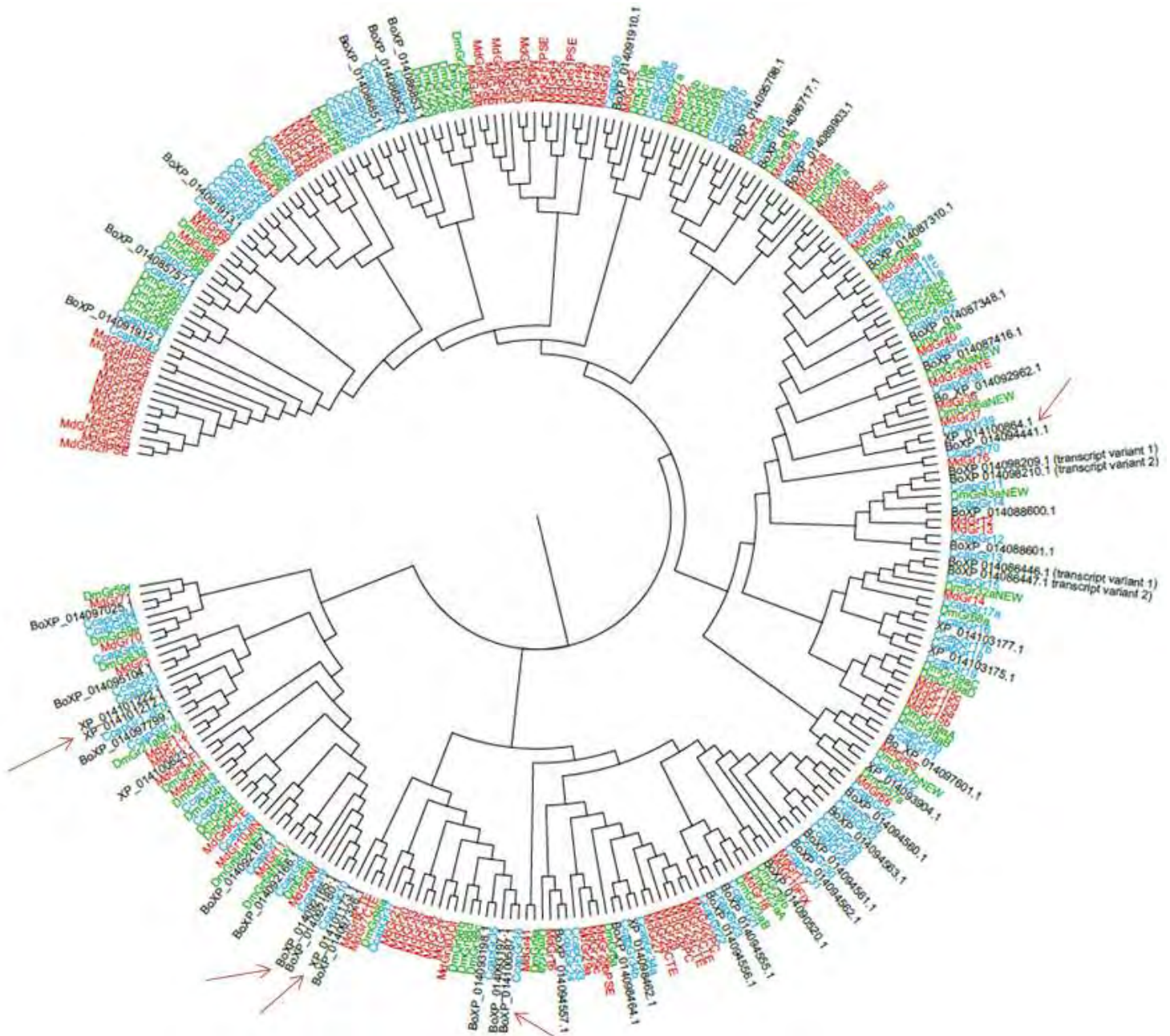
3.3. Φυλογενετική ανάλυση των υπό μελέτη γονιδίων χημειοαντίληψης του δάκου και των συγγενικών ειδών

Μετά την ανάκτηση και επιμέλεια των πραγματικών γονιδιακών μοντέλων των μελετούμενων γονιδίων, ακολούθησε η συγκριτική τους ανάλυση με άλλα συγγενικά είδη. Για τη σύγκριση των αλληλουχιών με αυτές των ειδών *C. capitata*, *D. melanogaster*, *M. domestica* αρχικά πραγματοποιήθηκε πολλαπλή στοίχιση τους στο περιβάλλον του Mega-X. Η στοίχιση έγινε με τον αλγόριθμο MUSCLE που φέρει ενσωματωμένο το Mega-X και τα αποτελέσματα αξιοποιήθηκαν για τη δημιουργία φυλογενετικού δέντρου και την ανάλυση των εξελικτικών σχέσεων των πρωτεϊνών των υπό μελέτη ειδών. Η δημιουργία του φυλογενετικού δέντρου πραγματοποιήθηκε με τη χρήση του προγράμματος Mega-X, ενώ η επεξεργασία και βελτίωση του έγινε με το πρόγραμμα Figtree.

3.3.1. Φυλογενετική ανάλυση των GRs του δάκου και των συγγενικών ειδών

Όσον αφορά τους GRs, συγκρίθηκαν 51 GRs του δάκου με 68 της *D. melanogaster*, 81 της *C. capitata* και 110 της *M. domestica*. Το φυλογενετικό δέντρο που προέκυψε μετά από την επεξεργασία με το Figtree για τους γευστικούς υποδοχείς φαίνεται παρακάτω (Εικ. 3.3.a).

Από τους 51 γευστικούς υποδοχείς του δάκου, οι δέκα φαίνονται να έχουν προκύψει από επεκτάσεις γονιδίων. Εντοπίζονται σε κοντινή απόσταση σε κοινούς κλάδους του φυλογενετικού δέντρου και μετά από πολλαπλή τους στοίχιση παρατηρήθηκε, επίσης, παρόμοια αλληλουχία. Επιπρόσθετα, δύο από αυτούς (υποδοχείς 5-6 του πίνακα 3.4) ανιχνεύονται στο ίδιο ικρίωμα του γονιδιώματος, χωρίς να συνεντοπίζονται (διαφορετικές συντεταγμένες). Οι 10 αυτοί υποδοχείς επισημαίνονται με κόκκινα βέλη στο φυλογενετικό δέντρο.



Εικόνα 3.3.a.: Φυλογενετικό δέντρο των GRs στα είδη *B.oleae*, *C.capitata*, *D.melanogaster* και *M. domestica*. Οι εξελικτικές σχέσεις προέκυψαν με τη μέθοδο Neighbor-joining και το συναინετικό φυλογενετικό δέντρο δημιουργήθηκε από 500 αντίγραφα με τη μέθοδο bootstrap. Οι κλάδοι που αναπαράγονταν σε λιγότερα από 50% των αντιγράφων bootstrap αφαιρέθηκαν από το φυλογενετικό δέντρο. Οι εξελικτικές αποστάσεις υπολογίστηκαν με τη μέθοδο Poisson με βάση τον αριθμό των αμινοξέων που έχουν αντικατασταθεί ανα περιοχή. Οι 51 γευστικοί υποδοχείς του δάκου φαίνονται με μαύρο χρώμα και με κόκκινα βέλη επισημαίνονται όσοι από αυτούς προέκυψαν από πιθανούς διπλασιασμούς.

Οι πρωτεΐνες αυτές συγκεντρώνονται στον πίνακα παρακάτω μαζί με τους μοναδικούς XP κωδικούς τους και τους κωδικούς XM των γονιδίων τους. Στον πίνακα φαίνονται επιπρόσθετα ο αριθμός των εξονίων και των ιντρονίων των γονιδίων, καθώς και των αμινοξέων (aa) των πρωτεϊνών. Επίσης, καταγράφονται οι περιοχές του γονιδιώματος στις οποίες ανιχνεύονται, αλλά και οι ακριβείς συντεταγμένες τους στις περιοχές αυτές.

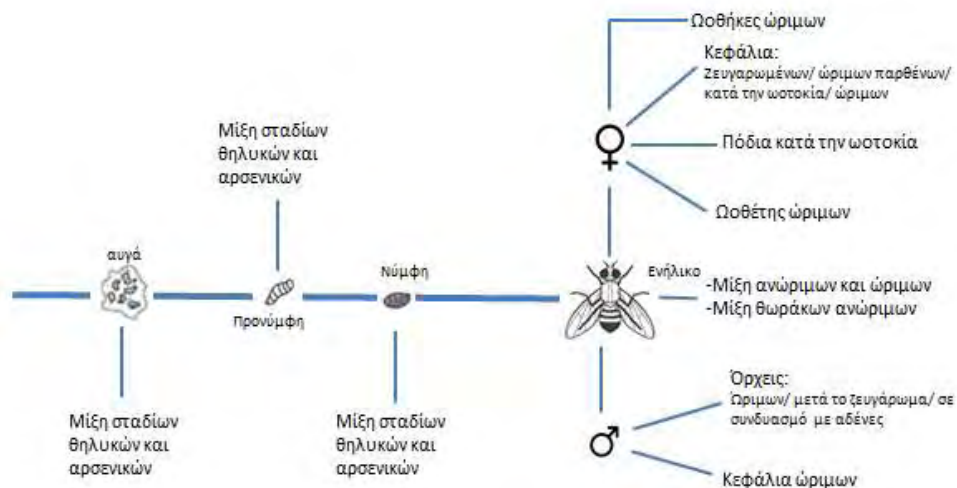
Πίνακας 3.4 Επεκτάσεις GRs στο δάκο της ελιάς

Επεκτάσεις GRs στο δάκο της ελιάς								
Όνομα NCBI	Κωδικός XM	Κωδικός XP	Ικρίωμα γονιδιώματος	Συντεταγμένες (bp)	aa	Αριθμός εξονίων	Αριθμός ιντρονίων	
1. <i>BoleaeGr</i> 21a-like	XM_014245747.1	XP_014101222.1	NW_013587886.1	42-7994	429	6	5	
2. <i>BoleaeGr</i> 21a-like	XM_014245737.1	XP_014101212.1	NW_013587693.1	6635-8114	367	4	3	
3. <i>BoleaeGr</i> 98b	XM_014237722.1	XP_014093197.1	NW_013581425.1	177620-179359	402	4	3	
4. <i>BoleaeGr</i> 98b	XM_014245112.1	XP_014100587.1	NW_013584271.1	6090-7742	377	4	3	
5. <i>BoleaeGr</i> 5a-for-trehalose	XM_014236690.1	XP_014092165.1	NW_013581359.1	36487-56316	441	8	7	
6. <i>BoleaeGr</i> 5a_trehalose-like	XM_014236705.1	XP_014092180.1	NW_013581359.1	19069-33140	484	9	8	
7. <i>BoleaeGr</i> 66a-like	XM_014238966.1	XP_014094441.1	NW_013581521.1	131336-139898	421	5	4	
8. <i>BoleaeGr</i> 66a-like (2)	XM_014245389.1	XP_014100864.1	NW_013585068.1	13767-14973	249	5	4	
9. <i>BoleaeGr</i> 21a-like	XM_014245698.1	XP_014101173.1	NW_013587184.1	7678-8672	278	2	1	

10.BoleaeGr22-like	XM_014241851.1	XP_014097326.1	NW_013581986.1	92308-101995	129	2	1	
--------------------	----------------	----------------	----------------	--------------	-----	---	---	--

Για να διερευνήσουμε σε ποια στάδια ανάπτυξης και σε ποιους ιστούς του δάκου παρατηρείται η έκφραση των δέκα πιθανών διπλασιασμένων υποδοχέων, αξιοποιήσαμε τα καταχωρημένα στο Apollo δεδομένα αλληλούχησης για το συνολικό μεταγράφημα του δάκου. Στόχος ήταν να συσχετιστούν πιθανά γονίδια με το φύλο ή συγκεκριμένες αναπτυξιακές διαδικασίες.

Συγκεκριμένα, στο Apollo υπάρχουν πληροφορίες που έχουν προκύψει από την RNA αλληλούχηση ιστών που αντιπροσωπεύουν όλα τα αναπτυξιακά στάδια καθώς και 16 διαφορετικών ιστών ενηλίκων που απομονώθηκαν υπό διαφορετικές συνθήκες (βλ. Εικ. 3.3.b).



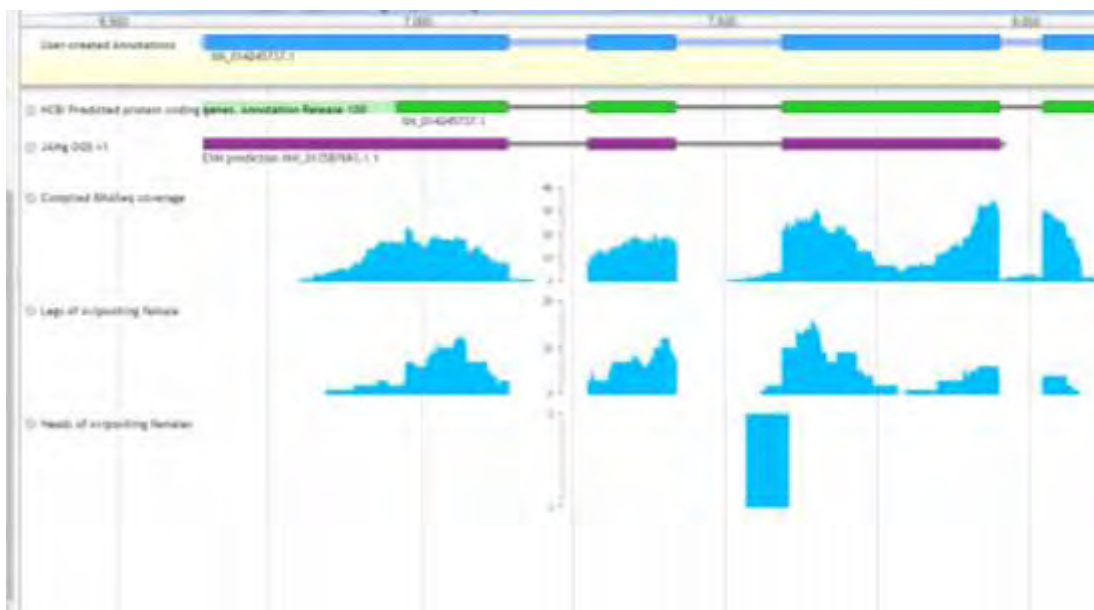
Εικόνα 3.3.b.: Στάδια ανάπτυξης και ιστοί του δάκου της ελιάς από τα οποία πραγματοποιήθηκε η συλλογή των δεδομένων RNA αλληλούχησης.

Όλοι οι γευστικοί υποδοχείς εμφανίζουν χαμηλά επίπεδα έκφρασης:

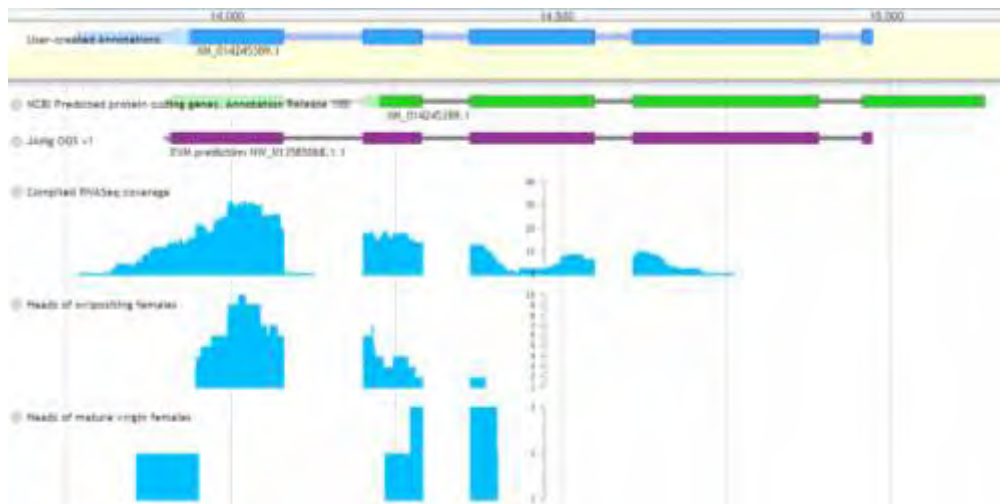
- Οι γευστικοί υποδοχείς με κωδικούς γονιδίων XM_014237722.1, XM_014236705.1, XM_014238966.1 και XM_014241851.1 εκφράζονται στα κεφάλια των ώριμων παρθένων θηλυκών. Ένα παράδειγμα (XM_014237722.1) φαίνεται παρακάτω:



- Οι γευστικοί υποδοχείς με κωδικούς γονιδίων XM_014245737.1, XM_014245747.1 και XM_014236690.1 εκφράζονται στα πόδια των θηλικών κατά την ωτοκία. Ένα παράδειγμα (XM_014245737.1) φαίνεται παρακάτω:



- Οι γευστικοί υποδοχείς με κωδικούς γονιδίων XM_014236705.1, XM_014245389.1, XM_014245698.1 και XM_014241851.1 εκφράζονται στα κεφάλια των θηλικών κατά την ωτοκία. Ένα παράδειγμα (XM_014245389.1) φαίνεται παρακάτω:



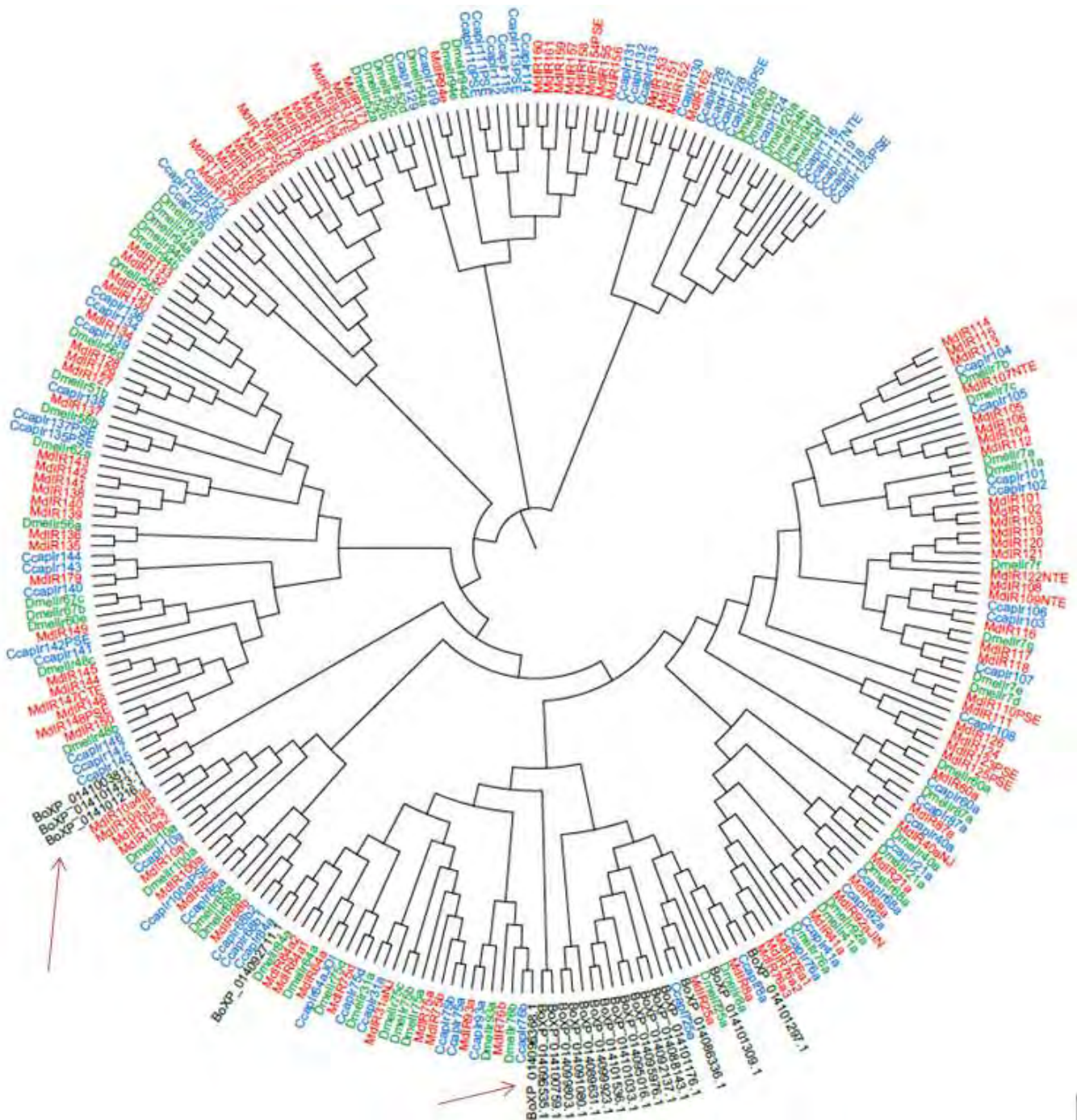
- Ο γευστικός υποδοχέας με κωδικό γονιδίου XM_014245112.1 εκφράζεται στα κεφάλια των θηλυκών:



3.3.2. Φυλογενετική ανάλυση των IRs του δάκου και των συγγενικών ειδών

Οι εξελικτικές σχέσεις των ιονοτροπικών υποδοχέων τόσο στο ίδιο είδος όσο και ανάμεσα στο δάκο και τα είδη *C. capitata*, *D. melanogaster* και *M. domestica* μελετήθηκαν με τη δημιουργία φυλογενετικού δέντρου. Για τη φυλογενετική ανάλυση του συγκρίθηκαν οι 21 IRs του δάκου με τους 65 της *D. melanogaster*, τους 110 της *M. domestica* και τους 70 της *C. capitata*. Η δημιουργία του φυλογενετικού δέντρου πραγματοποιήθηκε με τη χρήση του προγράμματος Mega-X, ύστερα από τη στοίχιση με το MUSCLE. Η βελτίωση του επιτεύχθηκε με το πρόγραμμα Figtree (εικόνα 3.4.d).

Από το φυλογενετικό δέντρο παρατηρήθηκε ότι οι 17 από τους 21 ιονοτροπικούς υποδοχείς του δάκου έχουν προκύψει από επεκτάσεις γονιδίων που πιθανότατα δεν εντοπίζονται στα υπόλοιπα τρία είδη (Πίνακας 3.5). Οι υποδοχείς αυτοί επισημαίνονται με κόκκινα βέλη στο φυλογενετικό δέντρο των ιονοτροπικών υποδοχέων που φαίνεται στην εικόνα 3.4.d.



Εικόνα 3.3.c.: Φυλογενετικό δέντρο των IRs στα είδη *B.oleae*, *C.capitata*, *D.melanogaster* και *M. domestica*. Οι εξελικτικές σχέσεις προέκυψαν με τη μέθοδο Neighbor-joining και το συναϊνικό φυλογενετικό δέντρο δημιουργήθηκε από 500 αντίγραφα με τη μέθοδο bootstr. Οι κλάδοι που αναπαράγονταν σε λιγότερα από 50% των αντιγράφων bootstr αφαιρέθηκαν από το φυλογενετικό

δέντρο. Οι εξελικτικές αποστάσεις υπολογίστηκαν με τη μέθοδο Poisson με βάση τον αριθμό των αμινοξέων που έχουν αντικατασταθεί ανα περιοχή. Οι 21 ιονοτροπικοί υποδοχείς του δάκου φαίνονται με μαύρο χρώμα και με κόκκινα βέλη επισημαίνονται όσοι από αυτούς φαίνεται να έχουν προκύψει από επεκτάσεις γονιδίων.

Πίνακας 3.5 Επεκτάσεις IRs στο δάκο της ελιάς

Ιονοτροπικοί υποδοχείς στο δάκο της ελιάς που προέκυψαν από επεκτάσεις γονιδίων							
Όνομα NCBI	Κωδικός XM	Κωδικός XP	Ικρίωμα	Συντεταγμένες (bp)	aa	Αριθμός εξονίων	Αριθμός ιντρονίων
1. <i>B.oleae</i> IR kainate 2	XM_014245558.1	XP_014101033.1	NW_013586217.1	2178-10444	904	14	13
2. <i>B.oleae</i> IR 2-like	XM_014246061.1	XP_014101536.1	NW_013594124.1	45-1935	201	3	2
3. <i>B.oleae</i> IR kainate 2	XM_014244448.1	XP_014099923.1	NW_013583280.1	3818-12084	108 5	15	14
4. <i>B.oleae</i> IR kainate 2	XM_014240501.1	XP_014095976.1	NW_013581697.1	82239-89746	729	11	10
5. <i>B.oleae</i> IR kainate 2	XM_014239541.1	XP_014095016.1	NW_013581582.1	60341-82224	101 2	12	11
6. <i>B.oleae</i> IR kainate 2-like	XM_014232668.1	XP_014088143.1	NW_013581264.1	23777-36385	901	15	15
7. <i>B.oleae</i> IR kainate 2	XM_014236662.1	XP_014092137.1	NW_013581357.1	365131-379578	921	15	14
8. <i>B.oleae</i> IR kainate 1-like	XM_014245701.1	XP_014101176.1	NW_013587227.	404-8621	176	4	3
9. <i>B.oleae</i> IR NMDA 2B	XM_014234156.1	XP_014089631.1	NW_013581290.1	104528-124352	732	9	8
10. <i>B.oleae</i> IR delta-2	XM_014240893.1	XP_014096368.1	NW_013581768.1	57618-63089	659	5	5
11. <i>B.oleae</i> IR kainate 1-like	XM_014244328.1	XP_014099803.1	NW_013583149.1	13335-28332	202	4	3
12. <i>B.oleae</i> IR kainate 2-like	XM_014241060.1	XP_014096535.1	NW_013581794.1	7052-72342	216	3	2
13. <i>B.oleae</i> glutamate [NMDA]	XM_014235605.1	XP_014091080.1	NW_013581321.1	154152-164509	985	13	12

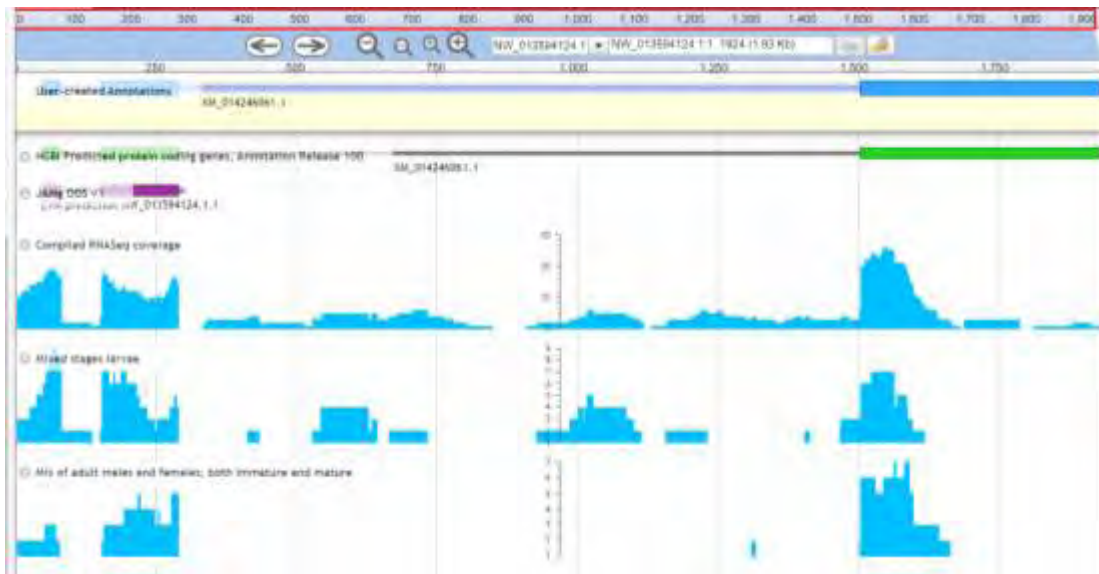
receptor subunit 1							
14. <i>B.oleae</i> IR kainate 2-like	XM_014245284.1	XP_014100759.1	NW_013584716.1	10-897	275	2	1
15. <i>B.oleae</i> IR kainate 5-like	XM_014245998.1	XP_014101473.1	NW_013592905.1	76-885	227	4	3
16. <i>B.oleae</i> IR kainate 1-like	XM_014245741.1	XP_014101216.1	NW_013587755.1	17-2700	311	5	4
17. <i>B.oleae</i> IR kainate 3-like	XM_014244906.1	XP_014100381.1	NW_013583866.1	1-4252	571	10	9

Οι υποδοχείς του παραπάνω πίνακα δεν εμφανίζουν ομολογία μεταξύ τους, ενώ ο αριθμός των εξονίων τους διαφέρει σημαντικά και κυμαίνεται από 3 έως 15. Με την αξιοποίηση των δεδομένων RNA αλληλούχησης που προσφέρονται από το Apollo για τα διάφορα στάδια ανάπτυξης και τους ιστούς του δάκου που αναφέρθηκαν παραπάνω, παρατηρήσαμε τα εξής:

- Ο ιονοτροπικός υποδοχέας με κωδικό γονιδίου XM_014245558.1 εμφανίζει μέτρια έκφραση που εντοπίζεται στους ωοθήτες και στα πόδια των θηλυκών κατά την ωοτοκία:



- Ο ιονοτροπικός υποδοχέας με κωδικό γονιδίου XM_014246061.1 εμφανίζει πολύ χαμηλή έκφραση στις προνύμφες και σε μια μίξη θηλυκών και αρσενικών ανώριμων και ώριμων:



- Ο ιονοτροπικός υποδοχέας με κωδικό γονιδίου XM_014244448.1 εμφανίζει πολύ υψηλή έκφραση, στα πόδια των θηλυκών κατά την ωοτοκία:



Ομοίως:

- Ιονοτροπικός υποδοχέας με κωδικό γονιδίου XM_014240501.1: Μέτρια έκφραση, στα κεφάλια των αρσενικών.
- Ιονοτροπικός υποδοχέας με κωδικό γονιδίου XM_014239541.1: Εμφανίζει μέτρια έκφραση που εντοπίζεται στα κεφάλια ώριμων παρθένων θηλυκών και τα κεφάλια θηλυκών κατά την ωοτοκία.
- Ιονοτροπικός υποδοχέας με κωδικό γονιδίου XM_014232668.1: Εκφράζεται σε μέτρια επίπεδα σε πολλά στάδια αλλά από τα πρώιμα στάδια του κύκλου

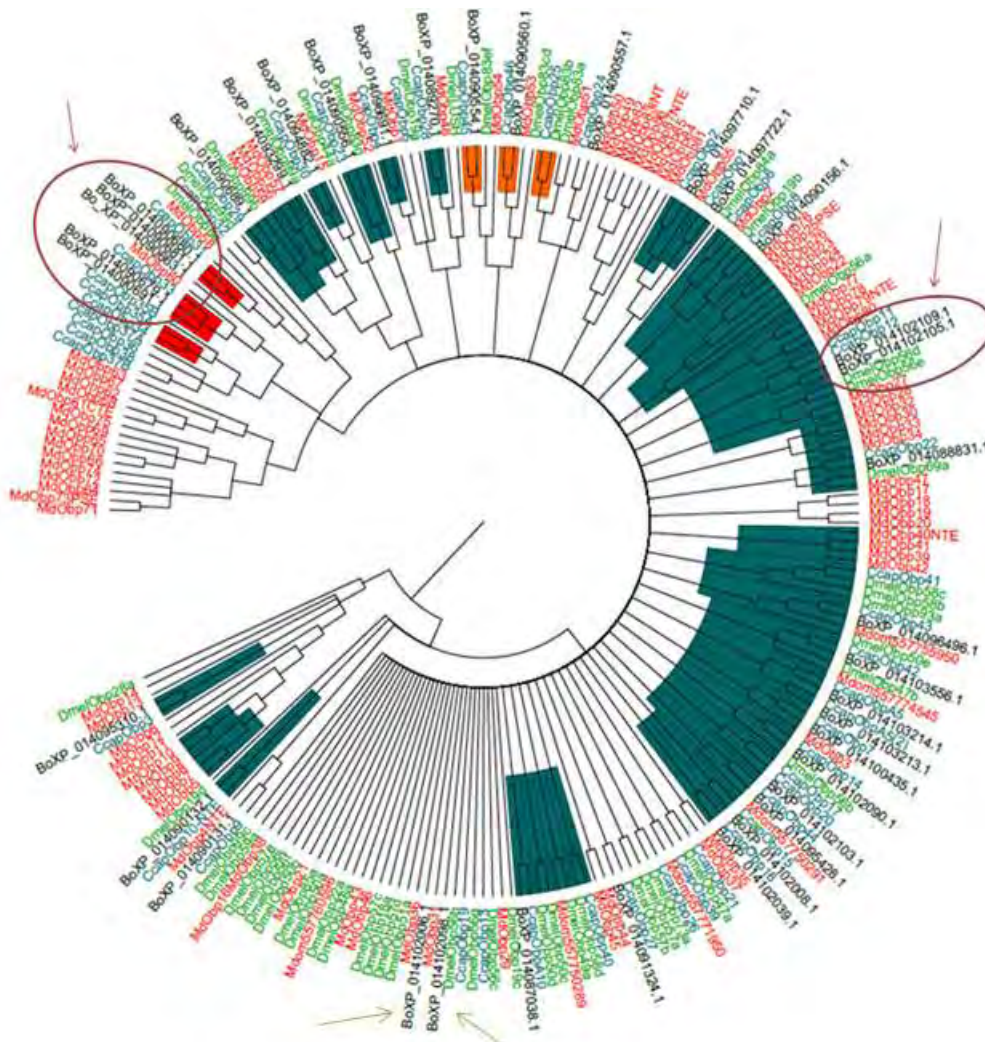
ζωής του δάκου μέχρι και τα ώριμα έντομα πριν από την ερωτοτροπία ή ωτοκία.

- Ιονοτροπικός υποδοχέας με κωδικό γονιδίου XM_014236662.1: Εμφανίζει πολύ υψηλή έκφραση. Εκφράζεται στα θηλυκά στα πόδια των θηλυκών κατά την ωτοκία.
- Ιονοτροπικός υποδοχέας με κωδικό γονιδίου: XM_014245701.1: Εμφανίζει μέτρια έκφραση στους ωοθέτες των θηλυκών και στα στάδια της προνύμφης.
- Ιονοτροπικός υποδοχέας με κωδικό γονιδίου XM_014245284.1: Εμφανίζει μέτρια έκφραση στα κεφάλια θηλυκών και αρσενικών και στους ωοθέτες των θηλυκών (Ίδια επίπεδα έκφρασης).
- Ιονοτροπικός υποδοχέας με κωδικό γονιδίου XM_014241060.1: Εμφανίζει υψηλή έκφραση που εντοπίζεται στα κεφάλια ώριμων θηλυκών και θηλυκών κατά την ωτοκία, αλλά και κεφάλια αρσενικών.
- Ιονοτροπικός υποδοχέας με κωδικό γονιδίου XM_014234156.1: Εμφανίζει χαμηλή έκφραση που εντοπίζεται σε αυγά που φέρουν αρσενικά.
- Ιονοτροπικός υποδοχέας με κωδικό γονιδίου XM_014240893.1: Χαμηλή έκφραση που εντοπίζεται στα κεφάλια ώριμων ζευγαρωμένων θηλυκών και ώριμαν παρθένων θηλυκών.
- Ιονοτροπικός υποδοχέας με κωδικό γονιδίου XM_014245998.1: Σχεδόν μηδαμινή έκφραση στην προνύμφη.
- Ιονοτροπικός υποδοχέας με κωδικό γονιδίου XM_014244906.1: Σχεδόν μηδαμινή έκφραση στους ωοθέτες των θηλυκών.
- Ιονοτροπικός υποδοχέας με κωδικό γονιδίου XM_014245741.1: Πολύ χαμηλή έκφραση κατά τα στάδια της προνύμφης και νύμφης.
- Ιονοτροπικός υποδοχέας με κωδικό γονιδίου XM_014235605.1: Μέτρια έκφραση, στα κεφάλια αρσενικών και θηλυκών, συμπεριλαμβανομένου και των θηλυκών που βρίσκονται κατά την ωτοκία.
- Ιονοτροπικός υποδοχέας με κωδικό γονιδίου XM_014244328.1: Υψηλή έκφραση που εντοπίζεται στα κεφάλια των αρσενικών και των ώριμων παρθένων θηλυκών.

3.3.3. Φυλογενετική ανάλυση των OBPs του δάκου και των συγγενικών ειδών

Ομοίως, το φυλογενετικό δέντρο για τη μελέτη των εξελικτικών σχέσεων των OBPs δημιουργήθηκε στο Mega-X και παρατίθεται στην εικόνα 3.4.e μετά από τη βελτίωση του με το Figtree. Κατά τη φυλογενετική ανάλυση συγκρίθηκαν οι 37 OBPs του δάκου

με 52 OBPs της *D. melanogaster*, 45 της *C. capitata*, και 87 της *M. domestica*. Από το φυλογενετικό δέντρο (εικόνα 3.3.d.) παρατηρήθηκε ότι οι δύο από τις OBPs εντοπίζονται σε πολύ μακρινή απόσταση σε σχέση με τα υπόλοιπα τρία υπό μελέτη είδη. Οι δύο αυτές πρωτεΐνες επισημαίνονται με πράσινα βέλη στο φυλογενετικό δέντρο και έχουν κωδικό XP_014102006.1 και XP_014102098.1. Τέλος, 7 από τις 37 OBPs του δάκου φαίνονται να έχουν προκύψει από επεκτάσεις γονιδίων (πίνακας 3.6) και επισημαίνονται στο φυλογενετικό δέντρο με κόκκινα βέλη. Στο φυλογενετικό δέντρο επισημαίνονται, επίσης, με μπλε, κόκκινο και πορτοκαλί οι OBPs του δάκου που ανήκουν στις υπο-οικογένειες Classic, Minus-C και Dimers αντίστοιχα.



Εικόνα 3.3.e.: Φυλογενετικό δέντρο των OBPs στα είδη *B. oleae*, *C. capitata*, *D. melanogaster* και *M. domestica*. Οι εξελικτικές σχέσεις προέκυψαν με τη μέθοδο Neighbor-joining και το συναϊετικό φυλογενετικό δέντρο δημιουργήθηκε από 500 αντιγράφα με τη μέθοδο bootstrap. Οι κλάδοι που αναπαράγονταν σε λιγότερα από 50% των αντιγράφων bootstrap αφαιρέθηκαν από το φυλογενετικό δέντρο. Οι εξελικτικές αποστάσεις υπολογίστηκαν με τη μέθοδο Poisson με βάση τον αριθμό των αμινοξέων που έχουν αντικατασταθεί ανα περιοχή. Από τις 232 αμινοξικές αλληλουχίες OBPs που συμπεριλήφθηκαν στο φυλογενετικό δέντρο, οι 37 του δάκου φαίνονται με μαύρο χρώμα.

Πίνακας 3.6. Επεκτάσεις OBPs στο δάκο της ελιάς

Επεκτάσεις OBPs στο δάκο της ελιάς							
Όνομα NCBI	Κωδικός XM	Κωδικός XP	Ικρίωμα	Συντεταγμένες (bp)	aa	Αριθμός εξονίων	Αριθμός ιντρονίων
1. <i>B. oleae</i> OBP 99a-like	XM_014235204.1	XP_014090679.1	NW_013581311.1	418104-419087	149	2	1
2. <i>B. oleae</i> OBP 99a-like	XM_014235216.1	XP_014090691.1	NW_013581311.1	410918-411628	147	2	1
3. <i>B. oleae</i> OBP 99b-like	XM_014235188.1	XP_014090663.1	NW_013581311.1	404363-404898	127	2	1
4. <i>B. oleae</i> OBP 99a-like	XM_014235205.1	XP_014090680.1	NW_013581311.1	533400-534221	159	2	1
5. <i>B. oleae</i> OBP 99a-like	XM_014235206.1	XP_014090681.1	NW_013581311.1	533400-534221	159	2	1
6. <i>B. oleae</i> OBP 56a-like	XM_014246634.1	XP_014102109.1	NW_013581217.1	457714-458449	526	1	0
7. <i>B. oleae</i> OBP 56a-like	XM_014246630.1	XP_014102105.1	NW_013581217.1	454810-455741	666	1	0

Οι πέντε πρώτες OBPs του πίνακα 3.6 εντοπίζονται διαδοχικά στο ίδιο ικρίωμα του γονιδιώματος, χωρίς όμως να συν-εντοπίζονται. Οι πρωτεΐνες αυτές κωδικοποιούνται από τον ίδιο αριθμό εξονίων και αποτελούνται από παρόμοιο αριθμό αμινοξέων. Επιπρόσθετα, μετά από στοίχιση τους με Clustal Omega παρατηρούνται πολλές ομοιότητες στην αλληλουχία τους και όπως είδαμε και από τη φυλογενετική ανάλυση (εικόνα 3.3.ε) εντοπίζονται στον ίδιο ή γειτονικούς κλάδους. Ομοίως και οι υπόλοιποι OBPs του πίνακα 3.6.

Μετά από ανάλυση της έκφρασης των γονιδίων του πίνακα 3.6, με βάση τα δεδομένα RNA αλληλούχησης, για κάθε ένα από τα στάδια ανάπτυξης και τους ιστούς του δάκου παρατηρήσαμε ότι γενικότερα τα γονίδια εκφράζονται στα ενήλικα έντομα αρσενικά και θηλυκά. Η OBP με κωδικό γονιδίου XM_014235216.1 εκφράζεται στα κεφάλια των ώριμων παρθένων θηλυκών και αυτών που βρίσκονται στο στάδιο της ωοτοκίας, αλλά και στα κεφάλια των αρσενικών. Το XM_014235204.1 εκφράζεται, επίσης και στα αρσενικά και στα θηλυκά σε περίπου ίδια επίπεδα, είτε πριν, είτε μετά την ερωτοτροπία και την ωοτοκία. Αντίστοιχα, η OBP με κωδικό γονιδίου XM_014235188.1 παρατηρήθηκε ότι εμφανίζει υπερβολικά υψηλή έκφραση που εντοπίζεται στα κεφάλια θηλυκών μετά από σύζευξη με αρσενικά και στα κεφάλια των ώριμων παρθένων θηλυκών. Η έκφραση των OBPs με κωδικούς γονιδίων XM_014235204.1 και XM_014235206.1 είναι επίσης υψηλή και παρατηρείται στα κεφάλια παρθένων θηλυκών και σε μεικτά στάδια νύμφης, ενώ των XM_014246634.1 και XM_014246630.1 στα κεφάλια ζευγαρωμένων θηλυκών.

3.4. *In silico* λειτουργικός χαρακτηρισμός του συνόλου των GRs, IRs και των OBPs στο δάκο της ελιάς

Μελέτες ομολογίας πρωτεϊνών της ίδιας λειτουργίας έχουν καταδείξει συντηρημένα μοτίβα στις αμινοξικές τους ακολουθίες (Patterns). Τα μοτίβα αυτά χαρακτηρίζουν τις πρωτεΐνες μιας οικογένειας και είναι ταξινομημένα στη βάση δεδομένων της UniProt. Με την αξιοποίηση της UniProt και του εργαλείου BLAST που προσφέρει, πραγματοποιήσαμε αναζήτηση ομολογιών των πρωτεϊνών χημειοαντίληψης του δάκου που έχουν προβλεφθεί υπολογιστικά. Με τον τρόπο αυτό ήταν εφικτό να καθοριστούν δομικές και λειτουργικές ομοιότητες των πρωτεϊνών αυτών στο δάκο με πειραματικά προσδιορισμένες πρωτεϊνικές δομές.

Από την αναζήτηση αυτή προέκυψε ότι από τους γευστικούς υποδοχείς, 34 πιθανότατα φέρουν λειτουργικές ομοιότητες με τους ομόλογους τους που εμπλέκονται στην αντίληψη πικρών ενώσεων, ενώ οκτώ από αυτούς με τους ομόλογους τους που εμπλέκονται στην αντίληψη γλυκών ενώσεων (βλ. Παράρτημα, Πίνακας 5.2.).

Για τους ιονοτροπικούς υποδοχείς βρέθηκε ότι στα είδη *D.melanogaster*, *C.capitata* και *M.domestica* παίζουν ρόλο στην αντίληψη της διέγερσης από πλήθος χημικών ενώσεων, όπως η πυρρολιδίνη, το βουτυρικό οξύ, το προπιονικό οξύ και άλλες χημικές ενώσεις με διαφορετική εξειδίκευση ο καθένας (βλ. Παράρτημα, Πίνακας 5.3).

Τέλος, για τις οσμοδεσμευτικές πρωτεΐνες παρατηρήθηκε ότι οι 21 από τις 37 συμμετέχουν σε διεργασίες του νευρικού συστήματος. Επιπρόσθετα, έξι OBPs εμπλέκονται σε αναπαραγωγικές διεργασίες, εννιά συνεισφέρουν σε αποκρίσεις σε ερεθίσματα και μία παίζει ρόλο στη σηματοδότηση. Τέλος, οκτώ από τις OBPs του δάκου εκτός από το ρόλο τους στην ανίχνευση οσμών, συμμετέχουν και στην ανίχνευση φερομόνων (βλ. Παράρτημα, Πίνακας 5.1.).

4. ΣΥΖΗΤΗΣΗ

Η πρόοδος των ολιστικών τεχνολογιών ανάλυσης των γονιδιωμάτων και των μεταγραφωμάτων έχουν αλλάξει τον τρόπο με τον οποίο διεξάγονται οι βιολογικές μελέτες. Με την ανάπτυξη των σύγχρονων μεθόδων αλληλούχησης που χαρακτηρίζονται από χαμηλότερο κόστος και μεγαλύτερη ακρίβεια δίνεται πλέον η δυνατότητα εφαρμογής των τεχνολογιών αυτών σε μεγάλο εύρος οργανισμών. Επιπλέον, τα δεδομένα που παράγονται μπορούν να χρησιμοποιηθούν σε συγκριτικές αναλύσεις ανάμεσα σε διαφορετικούς οργανισμούς, παρέχοντας σημαντικές πληροφορίες τόσο σε επίπεδο διερεύνησης εξελικτικών σχέσεων όσο και σε επίπεδο ταυτοποίησης ομόλογων ή μη αλληλουχιών μεταξύ των ειδών. Οι πληροφορίες που παράγονται αποθηκεύονται σε βάσεις δεδομένων, ώστε να είναι διαθέσιμες ανά πάσα στιγμή από την επιστημονική κοινότητα.

Ο χειρισμός των αλληλουχιών που παράγονται επιτυγχάνεται μέσω υπολογιστικών εργαλείων που επιτρέπουν την επιμέλεια και τον χαρακτηρισμό της δομής και της οργάνωσης των γονιδίων. Ανάμεσα στα εργαλεία αυτά βρίσκονται οι αλγόριθμοι πρόβλεψης γονιδιακών μοντέλων, οι οποίοι, όμως, συχνά, είναι επιρρεπείς σε λάθη και προβλέπουν εσφαλμένα γονιδιακά μοντέλα. Ωστόσο, για να ξεπεραστεί το παραπάνω πρόβλημα έχουν δημιουργηθεί εξειδικευμένες διαδικτυακές πλατφόρμες που επιτρέπουν τη χειροκίνητη επιμέλεια των γονιδιακών μοντέλων. Ένα παράδειγμα είναι η πλατφόρμα Apollo (i5k initiative) που παρέχει δεδομένα αλληλούχησης RNA (RNAseq) για πλήθος εντόμων σε συνδυασμό με τα συναινετικά γονιδιακά μοντέλα των αλγορίθμων των NCBI και JAMg και πειραματικά και υπολογιστικά αποδεικτικά στοιχεία για κάθε μοντέλο.

Μεταξύ των εντόμων που φιλοξενεί η πλατφόρμα Apollo βρίσκεται και ο δάκος της ελιάς, που αποτελεί μονοφάγο έντομο και είναι ο σημαντικότερος εχθρός των ελαιοκαλλιιεργειών, αφού κατά το στάδιο της προνύμφης τρέφεται αποκλειστικά από το εσωτερικό της ελιάς. Το γονιδίωμα του δάκου έχει αλληλουχηθεί, αλλά οι δομές των γονιδίων του δεν είναι επιμελημένες και οι μηχανισμοί που εμπλέκονται στην προτίμηση του προς την ελιά είναι άγνωστοι. Πιθανός μηχανισμός που καθοδηγεί τις προτιμήσεις του είναι η αντίληψη των ενώσεων των οσμών και φερομόνων μέσω του οσφρητικού και γευστικού συστήματος. Οι πρωτεΐνες που είναι υπεύθυνες για τη δέσμευση των παραπάνω χημικών ενώσεων και την επακόλουθη επιλογή συντρόφου ή την ανίχνευση υποστρωμάτων τροφής και ωτοκίας είναι οι οσμοδεσμευτικές πρωτεΐνες (OBPs), οι οσφρητικοί υποδοχείς (ORs), οι ιονοτροπικοί υποδοχείς (IRs) και οι γευστικοί υποδοχείς (GRs). Έτσι, στην παρούσα εργασία εστίασαμε στην επεξεργασία, επιμέλεια και ανάλυση των γονιδίων αυτών ώστε να εντοπιστούν αυτά που εμπλέκονται στους μηχανισμούς αναγνώρισης της ελιάς.

Για τη μελέτη των γονιδίων χημειοαντίληψης του δάκου, αρχικά, αναζητήθηκαν και ανακτήθηκαν οι μη επιμελημένες αλληλουχίες τους από την NCBI. Συνολικά ανακτήθηκαν 37 αλληλουχίες οσμοδεσμευτικών πρωτεϊνών, 51 αλληλουχίες γευστικών υποδοχέων και 21 ιονοτροπικών για τις οποίες πραγματοποιήσαμε επιμέλεια τους. Για τις OBPs διερευνήσαμε, επίσης, τα αμινοξικά τους μοτίβα, ενώ για τους GRs και IRs τις διαμεμβρανικές τους περιοχές. Με την ανάκτηση των επιμελημένων πρωτεϊνών πραγματοποιήσαμε, στη συνέχεια, συγκριτική ανάλυση με τα είδη *C. capitata*, *D. melanogaster* και *M. domestica*, ώστε να αποκαλυφθούν οι εξελικτικές τους σχέσεις και τυχόν επεκτάσεις γονιδίων που πιθανόν να είναι ειδο-ειδικοί στο δάκο. Τέλος, για τα γονίδια χημειοαντίληψης για τα οποία γνωρίζουμε τα

ομόλογα τους στα υπό μελέτη είδη διερευνήθηκε η λειτουργία τους μέσα από τις ομοιότητες τους με τα ομόλογα αυτά γονίδια. Η λειτουργία των διπλασιασμένων γονιδίων, ωστόσο, δεν είναι ξεκάθαρη και γι' αυτά τα γονίδια αναλύσαμε τα στάδια και τους ιστούς ανάπτυξης στα οποία εκφράζονται, ώστε να λάβουμε πληροφορίες σχετικά με τις λειτουργίες που πιθανόν να ρυθμίζουν.

Από το σύνολο των GRs για τους οποίους πραγματοποιήσαμε χειροκίνητη επιμέλεια, σε συμφωνία βρισκόταν τα συναινετικά γονιδιακά μοντέλα NCBI και JAMg μόνο για 15% αυτών, ενώ για τους υπόλοιπους GRs, τα δύο μοντέλα δεν ήταν σύμφωνα. Επιπρόσθετα, για το 50% των GRs τα δεδομένα RNAseq δεν επαρκούσαν για την επιμέλεια τους και την επιβεβαίωση των γονιδιακών μοντέλων. Όσον αφορά τον εντοπισμό τους στα scaffolds του γονιδιώματος, στο 80% των περιπτώσεων εντοπίσαμε ένα γονίδιο GR σε ένα scaffold, ενώ στις υπόλοιπες περιπτώσεις εντοπίσαμε δύο και τρία γονίδια σε ένα scaffold. Τα γονίδια που βρίσκονταν σε κοινά scaffolds, παρατηρήσαμε ότι έφεραν τον ίδιο αριθμό εξονίων και πολλές ομοιότητες στις αμινοξικές τους αλληλουχίες. Για το λόγο αυτό, υποθέτουμε ότι έχουν προκύψει από διπλασιασμούς. Εξαίρεση αποτελούν δύο υποδοχείς που αν και αποτελούνται από ένα και τέσσερα εξόνια αντίστοιχα, οι αλληλουχίες τους εμφανίζουν υψηλή ομολογία. Οι υποδοχείς αυτοί ίσως προέκυψαν από διπλασιασμούς που οφείλονται σε ρετρομετάθεση.

Οι παραπάνω διπλασιασμοί επιβεβαιώθηκαν με τη δημιουργία φυλογενετικού δέντρου στο οποίο συμπεριλήφθηκαν οι GRs του δάκου καθώς και των τριών ειδών με τα οποία συγκρίνεται. Μέσα από τη φυλογενετική αυτή ανάλυση παρατηρήσαμε επεκτάσεις σε περίπου 20% των γονιδίων GRs, που πιθανότατα οφείλονται σε αδύναμη εξελικτική πίεση, όπως έχει παρατηρηθεί και για τους GRs άλλων ειδών (Fujii *et al*, 2015). Συγκεκριμένα, παρόμοιες επεκτάσεις γονιδίων GRs έχουν παρατηρηθεί και στο συγγενικό με το δάκο είδος *Ceratitis capitata* και ειδικότερα σε γονίδια γευστικών υποδοχέων που σχετίζονται με την αντίληψη φρούτων (Paranicolaou, *et al*, 2016). Συνεπώς, η ύπαρξη παρόμοιων επεκτάσεων στο δάκο δεν προκαλεί έκπληξη, αφού οι γευστικοί υποδοχείς του εμπλέκονται, επίσης, στην αντίληψη φρούτων.

Αντίστοιχα, από τη χειροκίνητη επιμέλεια των IRs παρατηρήσαμε συμφωνία μεταξύ των συναινετικών γονιδιακών μοντέλων μόνο για 20% των περιπτώσεων, ενώ όλοι οι IRs εντοπίζονται σε διαφορετικά scaffolds. Επιπλέον, για το 40% των περιπτώσεων δεν επαρκούσαν τα δεδομένα RNAseq για την επιβεβαίωση των γονιδιακών μοντέλων. Στις περιπτώσεις αυτές, οι αντίστοιχοι IRs ίσως εκφράζονται σε άλλα συστήματα που δε σχετίζονται με τη χημειοαντίληψη. Η λειτουργία των περισσότερων IRs δεν είναι ακόμα γνωστή και έκφραση τους έχει παρατηρηθεί ακόμα και στο ακουστικό σύστημα στη *D. melanogaster* (Rytz *et al*, 2013).

Μέσα από φυλογενετική ανάλυση παρατηρήσαμε επεκτάσεις σε περίπου 75% των IRs, ποσοστό που πιθανότατα σχετίζεται με το ότι οι υποδοχείς αυτοί είναι υπεύθυνοι για την ανίχνευση ποικιλίας χημικών ενώσεων που διαφέρει για κάθε είδος.

Για τους IRs και τους GRs μελετήσαμε, επίσης, τον αριθμό των διαμεμβρανικών τους περιοχών. Από τους γευστικούς υποδοχείς παρατηρήσαμε ότι μόνο το 30% φέρει τις αναμενόμενες επτά διαμεμβρανικές περιοχές, ενώ από τους ιονοτροπικούς

υποδοχείς μόνο το 20%. Ωστόσο, παρόμοιες αποκλίσεις στον αριθμό των διαμεμβρανικών περιοχών έχουν παρατηρηθεί και σε άλλα είδη, τουλάχιστον όσον αφορά τους GRs. Για παράδειγμα, σε γευστικούς υποδοχείς του είδους *Bombyx Mori* έχουν παρατηρηθεί από τρεις έως εννιά διαμεμβρανικές περιοχές, ενώ πολλές φορές ασυμφωνία υπάρχει και στην τοπολογία του αμινοτελικού άκρου. Η τοπολογία αυτή των διαμεμβρανικών περιοχών πιθανότατα είναι γενική για όλους τους γευστικούς υποδοχείς των εντόμων. Ωστόσο, σε μερικές περιπτώσεις (βλ. Παράρτημα, Πίνακας 5.4, 5.5) εντοπίσαμε πρωτεΐνες που αποτελούνται μόνο από 100-200 aa σε αντίθεση με τα 300-600 aa που αποτελούν έναν GR του δάκου και τα 600-1000 aa που αποτελούν έναν IR του δάκου. Οι συγκεκριμένες αυτές μικρές πρωτεΐνες πιθανόν να κωδικοποιούνται από μερικές αλληλουχίες (partial sequences) και να συνιστούν τμήματα της πλήρους πρωτεΐνης. Στην περίπτωση αυτή, τα εξόνια που κωδικοποιούν το υπόλοιπο τμήμα της είτε δεν εντοπίστηκαν στο scaffold που μελετούσαμε, είτε παρεμβάλλεται κάποιο μεγάλο μήκος ιντρόνιο με αποτέλεσμα να μην υπολογίζονται από τους αλγορίθμους πρόβλεψης.

Όσον αφορά την ανάλυση των δομών των OBPs, στο 50% των περιπτώσεων εντοπίσαμε μία πρωτεΐνη σε ένα scaffold, ενώ στις υπόλοιπες περιπτώσεις σε ένα scaffold ανιχνεύθηκαν δύο ή τρεις πρωτεΐνες. Τα γονίδια που εντοπίσαμε στο ίδιο scaffold έφεραν ίδιο αριθμό εξονίων και όμοια αλληλουχία και γι' αυτό υποθέτουμε ότι έχουν προκύψει από διπλασιασμό μέσω άνισου ανασυνδυασμού. Από το φυλογενετικό δέντρο παρατηρήσαμε ότι επεκτάσεις γονιδίων έχουν συμβεί σε περίπου 15% των OBPs του δάκου. Μετά την επιμέλεια τους, ταξινομήσαμε τις OBPs στις τέσσερις υπο-οικογένειες που εντοπίζονται στα έντομα σύμφωνα με τα αμινοξικά τους μοτίβα. Το 60% των OBPs του δάκου ταξινομήθηκαν στην υπο-οικογένεια Classic, 25% στη Minus-C, 10% στη Dimers, ενώ καμία OBP δεν ανιχνεύθηκε στην Plus-C. Τα αποτελέσματα αυτά δε μας προκαλούν έκπληξη, αφού είναι συγκρίσιμα με του πολύ συγγενικού είδους *B. dorsalis* στο οποίο επίσης δεν έχει εντοπιστεί OBP στην υπο-οικογένεια Plus-C (Zheng *et al*, 2013).

Τέλος, χαρακτηρίσαμε τη λειτουργία των πρωτεϊνών που μελετάμε μέσα από αναζήτηση των ομόλογων πρωτεϊνών στη *C. capitata* και *D. melanogaster* ώστε με βάση τις δομικές ομοιότητες τους να καθοριστούν και οι λειτουργικές ομοιότητες. Συνολικά, περίπου το 70% των GRs σχετίζεται με την αντίληψη πικρών ενώσεων, ενώ μόνο το 15% αυτών εμπλέκονται στην αντίληψη των γλυκών ενώσεων. Άλλες GRs σχετίζονται με την αντίληψη της καφεΐνης, της φρουκτόζης ή παίζουν ρόλο σε σηματοδοτικά μονοπάτια. Από τις OBPs περίπου το 60% συμμετέχουν σε διεργασίες του νευρικού συστήματος, ενώ 20% εμπλέκονται στην αντίληψη φερομόνων. Ωστόσο, με την παραπάνω ανάλυση δεν ήταν δυνατή η διερεύνηση της λειτουργίας των γονιδίων του δάκου που προήλθαν από επεκτάσεις, καθώς πιθανότατα είναι ειδο-ειδικές. Επιπλέον, όσον αφορά τις OBPs, οι πρωτεΐνες με κωδικό XP_014102006.1 και XP_014102098.1 ανιχνεύονται σε πολύ μακρινή απόσταση στο φυλογενετικό δέντρο με άλλους και πιθανότατα είναι επίσης ειδο-ειδικές στο δάκο και η λειτουργία τους άγνωστη. Ωστόσο, μετά από ανάλυση της έκφρασης των γονιδίων αυτών στα διάφορα στάδια και ιστούς ανάπτυξης του δάκου παρατηρήσαμε ότι οι πρωτεΐνες με κωδικούς XP_014096691.1, XP_014090679.1, XP_014101212.1, XM_014236690.1, XP_014092180.1, XP_014100587.1, XP_014101212.1, XP_014101222.1, XP_014100864.1, XP_014100623.1, XP_014097326.1,

XP_014101033.1, XP_014099923.1, XP_014095016.1, XP_014088143.1, XP_014092137.1 και XP_014096535.1 εκφράζονται στα κεφάλια ή στα πόδια των θηλυκών κατά την ωοτοκία. Συνεπώς, οι πρωτεΐνες αυτές ίσως εμπλέκονται στους μοριακούς μηχανισμούς που καθοδηγούν το δάκο προς τους καρπούς της ελιάς.

Συμπερασματικά, με την αξιοποίηση των βιοπληροφορικών εργαλείων που παρέχονται για την ανάλυση γονιδιωματικών και μεταγραφόμενων αλληλουχιών, καταφέραμε στην παρούσα μελέτη να ταυτοποιήσουμε το σύνολο των τριών οικογενειών γονιδίων χημειοαντίληψης στο δάκο. Τα γονίδια τα οποία χαρακτηρίσαμε ως διπλασιασμένα πιθανότατα είναι ειδο-ειδικά στο δάκο και ίσως καθοδηγούν την προτίμηση του για την ελιά. Συνεπώς, τα δεδομένα που παράχθηκαν από την παρούσα μελέτη μπορούν να αποτελέσουν τη βάση για την ταυτοποίηση και τον χαρακτηρισμό νέων γονιδίων-στόχων μέσω μελλοντικών πειραματικών σχεδιασμών που θα συνεισφέρουν στην ανάπτυξη νέων εργαλείων για την εναλλακτική αντιμετώπιση του δάκου. Μελλοντικές προσεγγίσεις θα πρέπει επομένως να εστιάσουν στη διερεύνηση της λειτουργίας τέτοιων γονιδίων-στόχων, ώστε να απομονωθούν όσα εμπλέκονται στην προτίμηση του δάκου προς την ελιά ως αποκλειστική πηγή τροφής και ωοτοκίας και συνεπώς στην παρασιτική του δράση παρέχοντας νέες προοπτικές στις προσπάθειες δημιουργίας ειδο-ειδικών παγίδων και εντομοκτόνων φιλικών προς το περιβάλλον και τον άνθρωπο για την αντιμετώπισή του.

5. ΠΑΡΑΡΤΗΜΑ

Πίνακας 5.1. Οσμοδεσμευτικές πρωτεΐνες στο δάκο που ανακτήθηκαν από τη βάση δεδομένων της NCBI

Odorant binding proteins (OBPs) στο δάκο της ελιάς	XM code	XP code	scaffold	Καθόλου ορθόλογες σχέσεις ή πολύ κοντινή απόσταση με ένα είδος	Διάταξη αμινοξικών μοτίβων	Υπο-οικογένεια	Λειτουργία
1. PREDICTED: Bactrocera oleae general odorant-binding protein 19d	XM_014244960.1	XP_014100435.1	NW_013583958.1		1-2	Classic	Διαδικασίες νευρικού συστήματος
2. PREDICTED: Bactrocera oleae general odorant-binding protein 19a	XM_014241216.1	XP_014096691.1	NW_013581829.1		1-2	Classic	Αντίληψη φερομόνων, διαδικασίες νευρικού συστήματος και απόκριση σε ερεθίσματα
3. PREDICTED: Bactrocera oleae general odorant-binding protein 68	XM_014239953.1	XP_014095428.1	NW_013581625.1		1	Dimer	Διαδικασίες νευρικού συστήματος
4. PREDICTED: Bactrocera oleae general odorant-binding protein 68	XM_014239835.1	XP_014095310.1	NW_013581609.1	Πολύ κοντινή απόσταση με	1-2	Classic	Αντίληψη φερομόνων,

binding protein 28a-like				<i>C.cap.</i>			διαδικασίες νευρικού συστήματος
5. PREDICTED: Bactrocera oleae general odorant-binding protein 99a	XM_014239407. 1	XP_014094882. 1	NW_013581566. 1		3-1-2	Classic	
6. PREDICTED: Bactrocera oleae general odorant-binding protein 99a-like	XM_014237824. 1	XP_014093299. 1	NW_013581427. 1		3-1-2	Minus-C	Διαδικασίες νευρικού συστήματος και απόκριση σε ερεθίσματα
7. PREDICTED: Bactrocera oleae general odorant-binding protein 57c	XM_014235849. 1	XP_014091324. 1	NW_013581325. 1	Πολύ κοντινή απόσταση με <i>C.cap.</i>	1-2	Classic	Διαδικασίες νευρικού συστήματος και απόκριση σε ερεθίσματα
8. PREDICTED: Bactrocera oleae general odorant-binding protein 99a-like	XM_014235216. 1	XP_014090691. 1	NW_013581311. 1		3-1-2	Minus-C	Διαδικασίες νευρικού συστήματος και απόκριση σε ερεθίσματα
9. PREDICTED: Bactrocera oleae general odorant-	XM_014235214. 1	XP_014090689. 1	NW_013581311. 1		3-1-2	Minus-C	Διαδικασίες νευρικού

binding protein 99b	1	1	1			συστήματος και απόκριση σε ερεθίσματα
10. PREDICTED: Bactrocera oleae general odorant-binding protein 99a-like	XM_014235206. 1	XP_014090681. 1	NW_013581311. 1	3-1-2	Minus-C	-
11. PREDICTED: Bactrocera oleae general odorant-binding protein 99a-like	XM_014235205. 1	XP_014090680. 1	NW_013581311. 1	3-1-2	Minus-C	Διαδικασίες νευρικού συστήματος
12. PREDICTED: Bactrocera oleae general odorant-binding protein 99a-like	XM_014235204. 1	XP_014090679. 1	NW_013581311. 1	3-1-2	Minus-C	Διαδικασίες νευρικού συστήματος και απόκριση σε ερεθίσματα
13. PREDICTED: Bactrocera oleae general odorant-binding protein 99b-like	XM_014235188. 1	XP_014090663. 1	NW_013581311. 1	3-1-2		
14. PREDICTED: Bactrocera oleae general odorant-binding protein 83a-like	XM_014235082. 1	XP_014090557. 1	NW_013581308. 1	1-2	Classic	Ανίχνευση της φερομόνης (E,E)- α - farnesene
15. PREDICTED: Bactrocera oleae general odorant-	XM_014235081.	XP_014090556.	NW_013581308.	3-1-2	Classic	Διαδικασίες νευρικού

binding protein 99a	1	1	1				συστήματος
16. PREDICTED: Bactrocera oleae general odorant-binding protein 19d-like	XM_014234681.1	XP_014090156.1	NW_013581300.1	Πολύ κοντινή απόσταση με <i>C.carp.</i>	1-2	Classic	-
17. PREDICTED: Bactrocera oleae general odorant-binding protein 19d-like	XM_014234657.1	XP_014090132.1	NW_013581300.1		1-2	Classic	Αντίληψη φερομόνων, διαδικασίες νευρικού συστήματος
18. PREDICTED: Bactrocera oleae general odorant-binding protein 28a-like	XM_014234656.1	XP_014090131.1	NW_013581300.1	Πολύ κοντινή απόσταση με <i>C.carp.</i>	1-2	Classic	Αντίληψη φερομόνων, διαδικασίες νευρικού συστήματος
19. PREDICTED: Bactrocera oleae general odorant-binding protein lush	XM_014233795.1	XP_014089270.1	NW_013581283.1		1-2	Classic	Αντίληψη φερομόνων, απόκριση σε αλκοόλες, διαδικασίες νευρικού συστήματος, αναπαραγωγή και απόκριση σε ερεθίσματα

20. PREDICTED: Bactrocera oleae general odorant-binding protein 69a	XM_014233356. 1	XP_014088831. 1	NW_013581275. 1		1-2	Classic	Αντίληψη φερομόνων, διαδικασίες νευρικού συστήματος, αναπαραγωγή και απόκριση σε ερεθίσματα
21. PREDICTED: Bactrocera oleae putative odorant-binding protein A10	XM_014231563. 1	XP_014087038. 1	NW_013581247. 1	Πολύ κοντινή απόσταση με <i>C.carp.</i>		Minus-C	-
22. PREDICTED: Bactrocera oleae putative odorant-binding protein A5	XM_014247739. 1	XP_014103214. 1	NW_013581223. 1	Πολύ κοντινή απόσταση με <i>C.carp.</i>		Minus-C	-
23. PREDICTED: Bactrocera oleae putative odorant-binding protein A5	XM_014247738. 1	XP_014103213. 1	NW_013581223. 1	Πολύ κοντινή απόσταση με <i>C.carp.</i>		Minus-C	-
24. PREDICTED: Bactrocera oleae general odorant-binding protein 56a-like	XM_014246634. 1	XP_014102109. 1	NW_013581217. 1	Πολύ κοντινή απόσταση με <i>C.carp.</i>		Classic	-
25. PREDICTED: Bactrocera oleae general odorant-	XM_014246630. 1	XP_014102105. 1	NW_013581217. 1	Πολύ κοντινή απόσταση με	1-2	Classic	-

binding protein 56a-like				<i>C.car.</i>			
26. PREDICTED: Bactrocera oleae general odorant-binding protein 56h-like	XM_014246628. 1	XP_014102103. 1	NW_013581217. 1	Καθόλου ορθόλογες σχέσεις	1-2	Classic	Διαδικασίες νευρικού συστήματος και αναπαραγωγή
27. PREDICTED: Bactrocera oleae general odorant-binding protein 56a	XM_014246615. 1	XP_014102090. 1	NW_013581217. 1		1-2	Classic	Διαδικασίες νευρικού συστήματος
28. PREDICTED: Bactrocera oleae general odorant-binding protein 56d-like	XM_014246564. 1	XP_014102039. 1	NW_013581217. 1	Πολύ κοντινή απόσταση με <i>C.car.</i>	1-2	Classic	Διαδικασίες νευρικού συστήματος και αναπαραγωγή
29. PREDICTED: Bactrocera oleae general odorant-binding protein 56d-like	XM_014246533. 1	XP_014102008. 1	NW_013581217. 1	Πολύ κοντινή απόσταση με <i>C.car.</i>	1-2	Classic	Διαδικασίες νευρικού συστήματος και αναπαραγωγή
30. PREDICTED: Bactrocera oleae general odorant-binding protein 56h-like	XM_014246531. 1	XP_014102006. 1	NW_013581217. 1	Καθόλου ορθόλογες σχέσεις	1-2	Classic	
31. PREDICTED: Bactrocera oleae general odorant-	XM_014242247. 1	XP_014097722. 1	NW_013581214. 1		1-2	Classic	Αντίληψη φερομόνων, διαδικασίες

binding protein 84a-like							νευρικού συστήματος
32. PREDICTED: Bactrocera oleae general odorant-binding protein 84a-like	XM_014242235. 1	XP_014097710. 1	NW_013581214. 1		1-2	Classic	
33. PREDICTED: Bactrocera oleae general odorant-binding protein 56h-like	XM_014246623. 1	XP_014102098. 1	NW_013581217. 1	Πολύ κοντινή απόσταση με <i>C.carp.</i>	1-2	Classic	Διαδικασίες νευρικού συστήματος, αναπαραγωγή, απόκριση σε ερεθίσματα και σηματοδότηση
34. PREDICTED: uncharacterized protein LOC106617710	XM_014235079. 1	XP_014090554. 1	NW_013581308. 1		1-2	Dimer	Διαδικασίες νευρικού συστήματος
35. PREDICTED: uncharacterized protein LOC106617715	XM_014235085. 1	XP_014090560. 1	NW_013581308. 1		1-2	Dimer	Διαδικασίες νευρικού συστήματος
36. PREDICTED: uncharacterized protein	XM_014248081. 1	XP_014103556. 1	NW_013581226. 1		1-2	Classic	Διαδικασίες νευρικού

LOC106627806						συστήματος
37. PREDICTED:	XM_014241021.	XP_014096496.	NW_013581209.	3-1-2	Minus-C	Διαδικασίες
uncharacterized protein	1	1	1			νευρικού
LOC106621951						συστήματος

Πίνακας 5.2. Γευστικοί υποδοχείς στο δάκο που ανακτήθηκαν από τη βάση δεδομένων NCBI

Γευστικοί υποδοχείς στο δάκο της ελιάς	XM code	XP code	scaffold	Πραγματικό μοντέλο	Δεδομένα RNAseq	Ορθόλογες σχέσεις μόνο με ένα είδος	Λειτουργία
1. PREDICTED: <i>Bactrocera oleae</i> putative gustatory receptor 58a (LOC106614499), mRNA	XM_01423028.2.1	XP_014085757.1	NW_013581234.1	NCBI			CO ₂ , αντίληψη πικρού
2. PREDICTED: <i>Bactrocera oleae</i> gustatory and pheromone receptor 32a (LOC106614996), transcript variant X1, mRNA	XM_01423097.1.1	XP_014086446.1	NW_013581242.1	NCBI		Πολύ κοντινή απόσταση με <i>C.carp.</i>	CO ₂ , αντίληψη πικρού
3. PREDICTED: <i>Bactrocera oleae</i> gustatory and pheromone receptor 32a (LOC106614996), transcript variant X2, mRNA	XM_01423097.2.1	XP_014086447.1	NW_013581242.1	NCBI		Πολύ κοντινή απόσταση με <i>C.carp.</i>	CO ₂ , αντίληψη πικρού
4. PREDICTED: <i>Bactrocera oleae</i> putative gustatory receptor 89a (LOC106615154), mRNA	XM_01423124.2.1	XP_014086717.1	NW_013581245.1	NCBI			μετάδοση σήματος
5. PREDICTED: <i>Bactrocera oleae</i> putative gustatory receptor 22a (LOC106615238), mRNA	XM_01423137.6.1	XP_014086851.1	NW_013581246.1	-	Ανεπαρκή δεδομένα	Πολύ κοντινή απόσταση με <i>C.carp.</i>	CO ₂ , αντίληψη πικρού
6. PREDICTED: <i>Bactrocera oleae</i> putative gustatory receptor 22c (LOC106615239), mRNA	XM_01423137.7.1	XP_014086852.1	NW_013581246.1	NCBI	Ανεπαρκή δεδομένα	Πολύ κοντινή απόσταση με <i>C.carp.</i>	CO ₂ , αντίληψη πικρού
7. PREDICTED: <i>Bactrocera oleae</i> gustatory receptor for bitter taste	XM_01423137.8.1	XP_014086853.1	NW_013581246.1	NCBI, JAMg	Ανεπαρκή δεδομένα		CO ₂ , αντίληψη

22e-like (LOC106615240), mRNA				(Συμφωνία μοντέλων)			πικρού
8. PREDICTED: Bactrocera oleae putative gustatory receptor 28b (LOC106615562), mRNA	XM_01423183 5.1	XP_014087310.1	NW_013581251.1	NCBI	Ανεπαρκή δεδομένα		CO ₂ , αντίληψη πικρού
9. PREDICTED: Bactrocera oleae putative gustatory receptor 28a (LOC106615595), mRNA	XM_01423187 3.1	XP_014087348.1	NW_013581251.1	NCBI			CO ₂ , αντίληψη πικρού
19. PREDICTED: Bactrocera oleae gustatory and pheromone receptor 33a (LOC106615631), mRNA	XM_01423194 1.1	XP_014087416.1	NW_013581252.1	JAMg	Ανεπαρκή δεδομένα		CO ₂ , αντίληψη πικρού
11. PREDICTED: Bactrocera oleae gustatory receptor for sugar taste 43a-like (LOC106616461), mRNA	XM_01423312 5.1	XP_014088600.1	NW_013581272.1	NCBI	Ανεπαρκή δεδομένα	Πολύ κοντινή απόσταση με <i>C.car.</i>	CO ₂ , αντίληψη πικρού και γλυκού, απόκριση στη φρουκτόζη
12. PREDICTED: Bactrocera oleae gustatory receptor for sugar taste 43a-like (LOC106616462), partial mRNA	XM_01423312 6.1	XP_014088601.1	NW_013581272.1	JAMg		Πολύ κοντινή απόσταση με <i>C.car.</i>	CO ₂ , αντίληψη πικρού και γλυκού
13. PREDICTED: Bactrocera oleae putative gustatory receptor 94a (LOC106617318), mRNA	XM_01423442 8.1	XP_014089903.1	NW_013581294.1	NCBI	Ανεπαρκή δεδομένα		CO ₂ , αντίληψη πικρού
14. PREDICTED: Bactrocera oleae putative gustatory receptor 39b (LOC106617681), mRNA	XM_01423504 5.1	XP_014090520.1	NW_013581307.1	NCBI			-

15. PREDICTED: Bactrocera oleae gustatory receptor 10a (LOC106618632), mRNA	XM_01423643 5.1	XP_014091910.1	NW_013581352.1	NCBI	Ανεπαρκή δεδομένα		CO ₂ , αντίληψη πικρού
16. PREDICTED: Bactrocera oleae putative gustatory receptor 36b (LOC106618634), mRNA	XM_01423643 7.1	XP_014091912.1	NW_013581352.1	NCBI, JAMg (Συμφωνί α μοντέλων)		Πολύ κοντινή απόσταση με <i>C.car.</i>	CO ₂ , αντίληψη πικρού
17. PREDICTED: Bactrocera oleae gustatory receptor for bitter taste 22e-like (LOC106618635), mRNA	XM_01423643 8.1	XP_014091913.1	NW_013581352.1	-	Ανεπαρκή δεδομένα	Πολύ κοντινή απόσταση με <i>C.car.</i>	CO ₂ , αντίληψη πικρού
18. PREDICTED: Bactrocera oleae gustatory receptor 5a for trehalose (LOC106618803), mRNA	XM_01423669 0.1	XP_014092165.1	NW_013581359.1	NCBI, JAMg (Συμφωνί α μοντέλων)		Πολύ κοντινή απόσταση με <i>C.car.</i>	αντίληψη γλυκού, έκταση προσαρκτί δας
19. PREDICTED: Bactrocera oleae gustatory receptor for sugar taste 64f-like (LOC106618804), partial mRNA	XM_01423669 1.1	XP_014092166.1	NW_013581359.1	JAMg			αντίληψη γλυκού
20. PREDICTED: Bactrocera oleae gustatory receptor for sugar taste 64e (LOC106618805), mRNA	XM_01423669 2.1	XP_014092167.1	NW_013581359.1	NCBI, JAMg (Συμφωνί α μοντέλων)			αντίληψη γλυκού

21. PREDICTED: <i>Bactrocera oleae</i> gustatory receptor for sugar taste 64b-like (LOC106618806), mRNA	XM_01423669 3.1	XP_014092168.1	-	-	Καθόλου δεδομένα		αντίληψη γλυκού
22. PREDICTED: <i>Bactrocera oleae</i> gustatory receptor 5a for trehalose-like (LOC106618817), mRNA	XM_01423670 5.1	XP_014092180.1	NW_013581359.1	NCBI, JAMg (Συμφωνία μοντέλων)			CO ₂ , αντίληψη πικρού
23. PREDICTED: <i>Bactrocera oleae</i> gustatory receptor for sugar taste 64a (LOC106618820), mRNA	XM_01423670 7.1	XP_014092182.1	-	-	Καθόλου δεδομένα		αντίληψη γλυκού, επέκταση προσαρκτίδας, απόκριση σε μαλτόζη, σουκρόζη, γλυκόζη
24. PREDICTED: <i>Bactrocera oleae</i> gustatory receptor for bitter taste 66a (LOC106619412), mRNA	XM_01423748 7.1	XP_014092962.1	NW_013581411.1	NCBI, JAMg (Συμφωνία μοντέλων)			CO ₂ , αντίληψη πικρού, απόκριση στην καφεΐνη
25. PREDICTED: <i>Bactrocera oleae</i> putative gustatory receptor 98b (LOC106619572), mRNA	XM_01423772 2.1	XP_014093197.1	NW_013581425.1	JAMg		Πολύ κοντινή απόσταση με <i>C.carp.</i>	-

26. PREDICTED: <i>Bactrocera oleae</i> putative gustatory receptor 98b (LOC106619573), mRNA	XM_01423772 3.1	XP_014093198.1	NW_013581425.1	-	Ανεπαρκή δεδομένα	Πολύ κοντινή απόσταση με <i>C.carp.</i>	
27. PREDICTED: <i>Bactrocera oleae</i> putative gustatory receptor 57a (LOC106619776), mRNA	XM_01423842 9.1	XP_014093904.1	NW_013581209.1	JAMg	Ανεπαρκή δεδομένα		CO ₂ , αντίληψη πικρού
28. PREDICTED: <i>Bactrocera oleae</i> gustatory receptor for bitter taste 66a-like (LOC106620458), mRNA	XM_01423896 6.1	XP_014094441.1	NW_013581521.1	JAMg	Ανεπαρκή δεδομένα	Πολύ κοντινή απόσταση με <i>C.carp.</i>	-
29. PREDICTED: <i>Bactrocera oleae</i> putative gustatory receptor 23a, isoform B (LOC106620553), mRNA	XM_01423908 0.1	XP_014094555.1	NW_013581537.1	JAMg	Ανεπαρκή δεδομένα	Πολύ κοντινή απόσταση με <i>C.carp.</i>	CO ₂ , αντίληψη πικρού
30. PREDICTED: <i>Bactrocera oleae</i> putative gustatory receptor 23a, isoforms A/C (LOC106620555), mRNA	XM_01423908 1.1	XP_014094556.1	NW_013581537.1	JAMg	Ανεπαρκή δεδομένα	Πολύ κοντινή απόσταση με <i>C.carp.</i>	CO ₂ , αντίληψη πικρού
31. PREDICTED: <i>Bactrocera oleae</i> putative gustatory receptor 2a (LOC106620556), mRNA	XM_01423908 2.1	XP_014094557.1	NW_013581537.1	NCBI	Ανεπαρκή δεδομένα		CO ₂ , αντίληψη πικρού
32. PREDICTED: <i>Bactrocera oleae</i> putative gustatory receptor 39b (LOC106620558), mRNA	XM_01423908 5.1	XP_014094560.1	NW_013581537.1	NCBI		Πολύ κοντινή απόσταση με <i>C.carp.</i>	CO ₂ , αντίληψη πικρού
33. PREDICTED: <i>Bactrocera oleae</i> putative gustatory receptor 39b (LOC106620559), mRNA	XM_01423908 6.1	XP_014094561.1	NW_013581537.1	NCBI	Ανεπαρκή δεδομένα	Πολύ κοντινή απόσταση με <i>C.carp.</i>	CO ₂ , αντίληψη πικρού

34. PREDICTED: <i>Bactrocera oleae</i> putative gustatory receptor 39b (LOC106620560), mRNA	XM_01423908 7.1	XP_014094562.1	NW_013581537.1	NCBI	Ανεπαρκή δεδομένα	Πολύ κοντινή απόσταση με <i>C.carp.</i>	CO ₂ , αντίληψη πικρού
35. PREDICTED: <i>Bactrocera oleae</i> putative gustatory receptor 39b (LOC106620561), mRNA	XM_01423908 8.1	XP_014094563.1	NW_013581537.1	NCBI, JAMg (Συμφωνία μοντέλων)	-	Πολύ κοντινή απόσταση με <i>C.carp.</i>	CO ₂ , αντίληψη πικρού
36. PREDICTED: <i>Bactrocera oleae</i> gustatory and odorant receptor 63a (LOC106620951), mRNA	XM_01423962 9.1	XP_014095104.1	NW_013581213.1	NCBI			CO ₂ , αντίληψη πικρού
37. PREDICTED: <i>Bactrocera oleae</i> gustatory receptor for bitter taste 93a (LOC106621446), mRNA	XM_01424032 3.1	XP_014095798.1	NW_013581668.1	-	Ανεπαρκή δεδομένα		CO ₂ , αντίληψη πικρού, απόκριση στην καφεΐνη
38. PREDICTED: <i>Bactrocera oleae</i> putative gustatory receptor 59f (LOC106622393), mRNA	XM_01424155 0.1	XP_014097025.1	NW_013581896.1	-	Ανεπαρκή δεδομένα		CO ₂ , αντίληψη πικρού
39. PREDICTED: <i>Bactrocera oleae</i> gustatory and odorant receptor 22-like (LOC106622618), mRNA	XM_01424185 1.1	XP_014097326.1	NW_013581986.1	-		Πολύ κοντινή απόσταση με <i>M.dom.</i>	-
40. PREDICTED: <i>Bactrocera oleae</i> putative gustatory receptor 47b (LOC106622820), mRNA	XM_01424212 6.1	XP_014097601.1	NW_013582060.1	NCBI			CO ₂ , αντίληψη πικρού

41. PREDICTED: <i>Bactrocera oleae</i> gustatory and odorant receptor 21a (LOC106622974), mRNA	XM_01424232 4.1	XP_014097799.1	NW_013582108.1	NCBI, JAMg (Συμφωνί α μοντέλων)			CO ₂ , αντίληψη πικρού
42. PREDICTED: <i>Bactrocera oleae</i> gustatory receptor for sugar taste 43a-like (LOC106623283), transcript variant X1, mRNA	XM_01424273 4.1	XP_014098209.1	NW_013582234.1	NCBI		Πολύ κοντινή απόσταση με <i>C.car.</i>	
43. PREDICTED: <i>Bactrocera oleae</i> gustatory receptor for sugar taste 43a-like (LOC106623283), transcript variant X2, mRNA	XM_01424273 5.1	XP_014098210.1	NW_013582234.1	NCBI		Πολύ κοντινή απόσταση με <i>C.car.</i>	CO ₂ , αντίληψη πικρού και γλυκού, ρύθμιση της όρεξης, απόκριση σε φρουκτόζη
44. PREDICTED: <i>Bactrocera oleae</i> gustatory receptor 8a (LOC106623457), mRNA	XM_01424298 7.1	XP_014098462.1	NW_013582327.1	-	Ανεπαρκή δεδομένα	Πολύ κοντινή απόσταση με <i>C.car.</i>	
45. PREDICTED: <i>Bactrocera oleae</i> gustatory receptor 8a-like (LOC106623458), mRNA	XM_01424298 9.1	XP_014098464.1	NW_013582327.1	NCBI		Πολύ κοντινή απόσταση με <i>C.car.</i>	CO ₂ , αντίληψη πικρού
46. PREDICTED: <i>Bactrocera oleae</i> putative gustatory receptor 98b (LOC106625288), partial mRNA	XM_01424511 2.1	XP_014100587.1	NW_013584271.1	JAMg	Ανεπαρκή δεδομένα	Πολύ κοντινή απόσταση	-

						με <i>C.cap.</i>	
47. PREDICTED: Bactrocera oleae gustatory and odorant receptor 21a-like (LOC106625323), partial mRNA	XM_01424514 8.1	XP_014100623.1	NW_013584380.1	NCBI	Ανεπαρκή δεδομένα		-
48. PREDICTED: Bactrocera oleae gustatory receptor for bitter taste 66a-like (LOC106625559), partial mRNA	XM_01424538 9.1	XP_014100864.1	NW_013585068.1	JAMg	Ανεπαρκή δεδομένα	Πολύ κοντινή απόσταση με <i>C.cap.</i>	-
49. PREDICTED: Bactrocera oleae gustatory and odorant receptor 21a-like (LOC106625878), partial mRNA	XM_01424569 8.1	XP_014101173.1	NW_013587184.1	NCBI	Ανεπαρκή δεδομένα	Πολύ κοντινή απόσταση με <i>M.dom.</i>	αντίληψη της θερμοκρασίας
50. PREDICTED: Bactrocera oleae gustatory and odorant receptor 21a-like (LOC106625916), partial mRNA	XM_01424573 7.1	XP_014101212.1	NW_013587693.1	NCBI		Πολύ κοντινή απόσταση με <i>C.cap.</i>	-
51. PREDICTED: Bactrocera oleae gustatory and odorant receptor 21a-like (LOC106625926), mRNA	XM_01424574 7.1	XP_014101222.1	NW_013587886.1	JAMg	Ανεπαρκή δεδομένα	Πολύ κοντινή απόσταση με <i>C.cap.</i>	-
52. PREDICTED: Bactrocera oleae gustatory and pheromone receptor 32a-like (LOC106627548), mRNA	XM_01424770 0.1	XP_014103175.1	NW_013581223.1	-	Ανεπαρκή δεδομένα	Πολύ κοντινή απόσταση με <i>C.cap.</i>	
53. PREDICTED: Bactrocera oleae gustatory and pheromone receptor 32a-like (LOC106627551), mRNA	XM_01424770 2.1	XP_014103177.1	NW_013581223.1	-	Ανεπαρκή δεδομένα	Πολύ κοντινή απόσταση με <i>C.cap.</i>	

Πίνακας 5.3. Ιονοτροπικοί υποδοχείς στο δάκο που ανακτήθηκαν από τη βάση δεδομένων NCBI

Ιονοτροπικοί υποδοχείς στο δάκο της ελιάς	XM code	XP code	Scaffold	Καλύτερο συναινετικό μοντέλο	Δεδομένα RNAseq	Λειτουργία
1.PREDICTED: Bactrocera oleae glutamate [NMDA] receptor subunit 1	XM_014235605.1	XP_014091080.1	NW_013581321.1	NCBI		
2.PREDICTED: Bactrocera oleae glutamate receptor ionotropic, NMDA 2B	XM_014234156.1	XP_014089631.1	NW_013581290.1	JAMg	Ανεπαρκή δεδομένα	
3.PREDICTED: Bactrocera oleae glutamate receptor ionotropic, kainate 2-like	XM_014232668.1	XP_014088143.1	NW_013581264.1	NCBI, JAMg (Συμφωνία μοντέλων)		
4.PREDICTED: Bactrocera oleae glutamate receptor ionotropic, kainate 2	XM_014232588.1	XP_014088063.1	-	-	Καθόλου δεδομένα	
5.PREDICTED: Bactrocera oleae glutamate receptor ionotropic, kainate 2	XM_014230861.1	XP_014086336.1	NW_013581242.1	NCBI	Ανεπαρκή δεδομένα	Συν-υποδοχέας για την αντίληψη αμίνων
6.PREDICTED: Bactrocera oleae glutamate receptor ionotropic, kainate 2-like	XM_014246061.1	XP_014101536.1	NW_013594124.1	JAMg	Ανεπαρκή δεδομένα	
7.PREDICTED: Bactrocera oleae glutamate receptor ionotropic, kainate 5-like	XM_014245998.1	XP_014101473.1	NW_013592905.1	NCBI	Ανεπαρκή δεδομένα	
8.PREDICTED: Bactrocera oleae glutamate receptor ionotropic, kainate 2-like	XM_014245834.1	XP_014101309.1	NW_013590145.1	NCBI		
9.PREDICTED: Bactrocera oleae glutamate receptor ionotropic, kainate 3-like	XM_014245822.1	XP_014101297.1	NW_013589804.1	NCBI	Ανεπαρκή δεδομένα	
10.PREDICTED: Bactrocera oleae glutamate receptor ionotropic, kainate 1-like	XM_014245741.1	XP_014101216.1	NW_013587755.1	JAMg		
11.PREDICTED: Bactrocera oleae glutamate	XM_014245701.1	XP_014101176.1	NW_013587227.1	JAMg	Ανεπαρκή δεδομένα	

receptor ionotropic, kainate 1-like						
12.PREDICTED: Bactrocera oleae glutamate receptor ionotropic, kainate 2	XM_014245558.1	XP_014101033.1	NW_013586217.1	JAMg	Ανεπαρκή δεδομένα	2-οξοπεντανικό οξύ
13.PREDICTED: Bactrocera oleae glutamate receptor ionotropic, kainate 2-like	XM_014245284.1	XP_014100759.1	NW_013584716.1	NCBI	Ανεπαρκή δεδομένα	
14.PREDICTED: Bactrocera oleae glutamate receptor ionotropic, kainate 3-like	XM_014244906.1	XP_014101033.1	NW_013583866.1	NCBI	Ανεπαρκή δεδομένα	
15.PREDICTED: Bactrocera oleae glutamate receptor ionotropic, kainate 2	XM_014244448.1	XP_014099923.1	NW_013583280.1	NCBI, JAMg (Συμφωνία μοντέλων)		
16.PREDICTED: Bactrocera oleae glutamate receptor ionotropic, kainate 1-like	XM_014244328.1	XP_014099803.1	NW_013583149.1	JAMg		
17.PREDICTED: Bactrocera oleae glutamate receptor ionotropic, kainate 1-like	XM_014244010.1	XP_014099485.1	-	-	Καθόλου δεδομένα	
18.PREDICTED: Bactrocera oleae glutamate receptor ionotropic, kainate 2-like	XM_014241060.1	XP_014096535.1	NW_013581794.1	JAMg		
19.PREDICTED: Bactrocera oleae glutamate receptor ionotropic, delta-1	XM_014240893.1	XP_014096368.1	NW_013581768.1	NCBI, JAMg (Συμφωνία μοντέλων)		Απόκριση στη φαινυλαιθυλαμίνη
20.PREDICTED: Bactrocera oleae glutamate receptor ionotropic, kainate 2	XM_014240501.1	XP_014095976.1	NW_013581697.1	NCBI, JAMg (Συμφωνία μοντέλων)		
21.PREDICTED: Bactrocera oleae glutamate receptor ionotropic, kainate 2	XM_014239541.1	XP_014095016.1	NW_013581582.1	JAMg		
22.PREDICTED: Bactrocera oleae glutamate receptor ionotropic, delta-2	XM_014237236.1	XP_014092711.1	NW_013581387.1	JAMg		Υποδοχέας φαινυλακετοξικού/φαιν

υλακεταλδεϋδης, ρόλος
στην ερωτροτροπία σε
περιοχές πλούσιες σε
θρεπτικά

23.PREDICTED: Bactrocera oleae glutamate
receptor ionotropic, kainate 2

XM_014236662.1

XP_014092137.1

NW_013581357.1

NCBI, JAMg
(Συμφωνία
μοντέλων)

Πίνακας 5.4.: Αριθμός διαμεμβρανικών περιοχών του συνόλου των γευστικών υποδοχέων του δάκου

Κωδικός ΧΡ	Μήκος	Αριθμός ελικών	TM1- start	TM1- stop	TM2- start	TM2- stop	TM3- start	TM3- stop	TM4- start	TM4- stop	TM5- start	TM5- stop	TM6- start-	TM6- stop	TM7- start	TM7- stop	TM8- start	TM8- stop	TM9- start	TM9- stop
XP_014085757.1	411	6	7	29	42	64	149	171	181	203	264	286	306	411						
XP_014086446.1	425	5	161	183	195	217	285	307	314	336	388	410								
XP_014086447.1	425	5	161	183	195	217	285	307	314	336	388	410								
XP_014086717.1	384	8	13	35	50	72	85	107	144	166	195	212	249	271	284	306	361	383		
XP_014086851.1	388	6	10	32	44	66	81	103	155	177	182	204	361	383						
XP_014086852.1	362	3	43	65	1118	140	147	169												
XP_014086853.1	431	7	51	73	88	110	149	168	188	210	296	318	333	351	404	426				
XP_014087310.1	471	8	34	56	76	98	116	135	170	192	207	229	306	325	340	362	421	440		
XP_014087348.1	452	8	15	37	44	66	84	106	140	162	172	194	300	322	335	357	407	429		
XP_014087416.1	482	6	26	48	129	151	166	188	320	342	352	374	431	453						
XP_014088600.1	420	7	39	61	81	103	134	156	176	198	286	308	318	337	397	416				
XP_014088601.1	402	5	39	58	78	100	132	154	169	191	262	284								
XP_014089903.1	397	5	7	29	39	61	82	104	124	146	159	181								
XP_014090520.1	339	7	20	37	49	66	76	98	118	140	144	166	219	241	2511	274				

XP_014091910.1	408	8	15	30	42	64	79	101	145	167	177	199	274	296	306	328	384	403		
XP_014091912.1	402	8	4	26	38	60	75	97	136	158	162	184	265	287	292	311	372	394		
XP_014091913.1	426	7	4	26	38	60	73	95	134	153	168	190	264	286	296	318				
XP_014092165.1	441	9	56	78	91	113	123	145	182	204	219	241	248	270	306	328	340	362	407	429
XP_014092166.1	510	8	148	165	180	197	233	255	270	292	297	319	360	382	394	416	461	483		
XP_014092167.1	439	8	46	68	88	110	122	139	177	199	212	234	239	261	338	360	403	425		
XP_014092180.1	439	8	46	68	88	110	122	139	177	199	212	234	239	261	338	360	403	425		
XP_014092962.1	524	6	9	31	46	68	81	98	142	164	338	360	380	402						
XP_014093197.1	402	7	10	32	68	81	82	104	150	172	251	270	291	313	363	385				
XP_014093198.1	333	5	17	39	76	98	186	208	223	242	294	316								
XP_014093904.1	409	7	13	35	45	64	85	107	176	198	255	277	287	304	361	383				
XP_014094441.1	421	5	28	50	65	87	124	146	156	178	295	317								
XP_014094555.1	380	6	27	49	94	116	131	153	2225	247	257	278	334	356						
XP_014094556.1	391	6	13	32	67	89	126	148	238	260	273	292	349	371						
XP_014094557.1	417	5	36	58	73	95	135	157	167	189	387	409								
XP_014094560.1	394	7	53	75	85	107	151	173	178	200	256	278	298	320	363	385				

XP_014094561.1	399	8	58	75	85	107	150	172	177	199	235	257	261	283	296	315	371	393		
XP_014094562.1	386	7	47	69	73	95	140	162	167	189	247	266	286	308	361	383				
XP_014094563.1	408	7	60	82	86	108	151	173	178	195	265	287	297	319	379	398				
XP_014095104.1	485	7	103	125	145	164	200	222	227	249	296	318	333	350	410	432				
XP_014095798.1	393	7	51	70	80	102	133	155	170	192	262	284	294	313	369	391				
XP_014097025.1	422	7	73	95	110	132	165	187	202	224	273	295	305	327	380	402				
XP_014097326.1	129	2	13	35	55	77														
XP_014097601.1	414	7	45	67	82	104	147	169	179	201	258	280	295	317	371	393				
XP_014097799.1	456	6	113	135	150	172	197	219	234	256	305	327	342	359						
XP_014098209.1	427	7	39	61	83	105	134	156	176	198	294	316	326	344	404	423				
XP_014098210.1	423	7	38	60	80	102	130	152	172	194	290	312	322	340	400	419				
XP_014098462.1	316	5	39	61	71	93	137	159	164	181	247	269								
XP_014098464.1	385	6	40	59	74	93	137	155	165	184	246	268	360	379						
XP_014100587.1	377	7	13	35	48	70	90	112	152	171	181	203	258	280	295	314				
XP_014100623.1	200	2	115	137	150	172														
XP_014100864.1	249	3	71	93	114	133	184	206												

XP_014101173.1	278	4	113	135	150	172	197	219	234	256									
XP_014101212.1	367	5	27	49	62	84	148	170	219	241	256	273							
XP_014101222.1	445	4	184	206	221	243	292	314	329	346									
XP_014103175.1	429	7	32	49	62	84	104	121	182	204	208	225	237	259	279	301			
XP_014103177.1	429	7	32	49	62	84	104	121	182	204	208	225	237	259	279	301			

Πίνακας 5.5.: Αριθμός διαμεμβρανικών περιοχών του συνόλου των ιονοτροπικών υποδοχέων του δάκου

Κωδικός XP	Μήκος	Αριθμός κελιών		TM1-start	TM1-stop	TM2-start	TM2-stop	TM3-start	TM3-stop	TM4-start	TM4-stop
XP_014091080.1	985	3	575		594	650	672	833	855		
XP_014089631.1	732	3	314		336	391	413	577	599		
XP_014088143.1	901	3	557		576	633	655	819	841		
XP_014086336.1	941	3	559		581	638	660	860	882		
XP_014101536.1	201	0									
XP_014101473.1	194	0									

XP_014101309.1	185	0								
XP_014101297.1	250	1	56	78						
XP_014101216.1	311	1	288	310						
XP_014101176.1	176	2	86	108	158	175				
XP_014101033.1	904	4	547	569	590	607	622	644	810	832
XP_014100759.1	275	1	136	158						
XP_014100381.1	571	1	549	568						
XP_014099923.1	1085	2	550	569	625	647				
XP_014099803.1	202	0								
XP_014096535.1	216	1	121	143						
XP_014096368.1	659	4	203	225	238	257	272	294	470	492
XP_014095976.1	729	2	498	520	685	707				
XP_014095016.1	1012	4	553	575	595	617	632	654	818	837
XP_014092711.1	698	4	13	35	395	414	460	482	651	673
XP_014092137.1	921	4	13	35	550	572	627	649	814	836

6. ΒΙΒΛΙΟΓΡΑΦΙΑ

- Altschul, S. G. (1990). (1990). *Basic local alignment search tool*. *J. Mol. Biol.* 215:403-410. PubMed.
- Amrein H, T. N. (2005). (2005). *Gustatory perception and behavior in Drosophila melanogaster*. *Curr Bi*
- Andrews M.-J., B. J. (2003). (2003). *Sensory properties of virgin olive oil polyphenols: identification of deacetoxy-ligstroside agglycon as a key contributor topungency*. *Journal of agricultural and food chemistry* 51, 1415–14.
- Bairoch A, A. R. (1997). (1997). *The Swiss-Prot protein sequence database: its relevance to human molecular medical research* *J. Mol. Med.* 75:312-316.
- Bhutkar A, R. S. (2007). (2007). *Genome-scale analysis of positionally relocated genes*. *Genome Res.* 2007;17(12):1880–1887. doi:10.1101/gr.7062307.
- Blackshields G, S. F. (2010). (2010). *Sequence embedding for fast construction of guide trees for multiple sequence alignment*. *Algorithms Mol Biol.* 2010;5:21. Published 2010 May 14. doi:10.1186/1748-7188-5-21.
- Brand P, R. S. (2017). (2017). *The Evolutionary Dynamics of the Odorant Receptor Gene Family in Corbiculate Bees*. *Genome Biol Evol.* ;9(8):2023–2036. doi:10.1093/gbe/evx149.
- Brudno M, M. S. (2003). (2003). *Glocal alignment: finding rearrangements during alignment*. *Bioinformatics.* 2003;19 Suppl 1:i54-62. PubMed PMID: 12855437.
- Buck L., A. R. (1991). *A novel multigene family may encode odorant receptors: a molecular basis for odor recognition*.
- Campanini EB, d. B. (2016). (2016). *Molecular evolution of Odorant-binding proteins gene family in two closely related Anastrepha fruit flies*. *BMC Evol Biol.* 2016;16(1):198. Published 2016 Oct 7. doi:10.1186/s12862-016-0775-0.
- Chen X, J. E. (2010). *New tools for functional genomic analysis*. *Drug Discov Today.* 2009;14(15-16):754–760. doi:10.1016/j.drudis.2009.05.005.
- Consortium, S.-I. (2014). . *A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium*. *Nat Biotechnol.* 2014;32(9):903–914. doi:10.1038/nbt.2957.
- Croset V, R. R. (2010). (2010). *Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction*. *PLoS Genet.* 2010;6(8):e1001064. Published 2010 Aug 19. doi:10.1371/journal.pgen.1001064.

- D., W. (2012). *Functional and evolutionary aspects of chemoreceptors*. *Front Cell Neurosci*. 2012;6:48. Published 2012 Oct 26. doi:10.3389/fncel.2012.00048.
- Dunn CW, L. X. (2013). (2013). *Phylogenetic analysis of gene expression*. *Integr Comp Biol*. 2013;53(5):847–856. doi:10.1093/icb/ict068.
- Edgar RC., E. (2004). *MUSCLE: multiple sequence alignment with high accuracy and high throughput*. *Nucleic Acids Res*. 2004;32(5):1792–1797. Published 2004 Mar 19. doi:10.1093/nar/gkh340.
- Elfekih S, C. C. (2016). (2016). *Identification and preliminary characterization of chemosensory perception-associated proteins in the melon fly *Bactrocera cucurbitae* using RNA-seq*. *Sci Rep*. 2016;6:19112. Published 2016 Jan 11. doi:10.1038.
- Elkan, T. L. (1994). *Fitting a mixture model by expectation maximization to discover motifs in biopolymers*, *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, pp. 28-36, AAAI Press, Menlo Par.
- Estes AM, H. D. (2018). . (2018). *Comparative genomics of the *Erwinia* and *Enterobacter* olive fly endosymbionts*. *Sci Rep*. 2018;8(1):15936. Published 2018 Oct 29. doi:10.1038/s41598-018-33809-w.
- Frazer KA, P. L. (2004). (2004). *VISTA: computational tools for comparative genomics*. *Nucleic Acids Res*. 2004;32(Web Server issue):W273–W279. doi:10.1093/nar/gkh458.
- Fritz, M. H.-Y. (2011). (2011). *Exploiting high throughput DNA sequencing*.
- Fujii S, Y. A. (2015). *Drosophila sugar receptors in sweet taste perception, olfaction, and internal nutrient sensing*. *Curr Biol*. 2015;25(5):621–627. doi:10.1016/j.cub.2014.12.058.
- Giraldo-Calderón GI, E. S. (2015). *VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases*. *Nucleic Acids Res*. 2015;43(Database issue):D707–D713. doi:10.1093/nar/gku1117.
- Gomez-Diaz C, M. F.-F. (2018). (2018). *The Two Main Olfactory Receptor Families in *Drosophila*, ORs and IRs: A Comparative Approach*. *Front Cell Neurosci*. 2018;12:253. Published 2018 Aug 30. doi:10.3389/fncel.2018.00253.
- Griffiths AJF, M. J. (2000). *Introduction*. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK21851/>.
- Hastings, P. J. (1998). (1998). *Gene Conversion*. *Encyclopedia of Immunology*, 969–973. doi:10.1006/rwei.1999.0252.

- Hekmat-Scafe DS, S. C. (2002). (2002). *Genome-wide analysis of the odorant-binding protein gene family in Drosophila melanogaster*. *Genome Res.* 2002;12(9):1357–1369. doi:10.1101/gr.239402.
- Hopf TA, M. S. (2015). (2015). *Amino acid coevolution reveals three-dimensional structure and functional domains of insect odorant receptors*. *Nat Commun.* 2015;6:6077. Published 2015 Jan 13. doi:10.1038/ncomms7077.
- Hui-Jie Zhang, A. R.-R.-H.-Y. (2011). (2011). *Topological and Functional Characterization of an Insect Gustatory Receptor*.
- Isono K, M. H. (2010). (2010). *Molecular and cellular designs of insect taste receptor system*. *Front Cell Neurosci.* 2010;4:20. Published 2010 Jun 18. doi:10.3389/fncel.2010.00020.
- Jiang X, K. J. (2017). (2017). *Distinct Subfamilies of Odorant Binding Proteins in Locust (Orthoptera, Acrididae): Molecular Evolution, Structural Variation, and Sensilla-Specific Expression*. *Front Physiol.* 2017;8:734. Published 2017 Sep 26. do.
- Joseph RM, C. J. (2015). (2015). *Drosophila Chemoreceptors: A Molecular Interface Between the Chemical World and the Brain*. *Trends Genet.* 2015;31(12):683–695. doi:10.1016/j.tig.2015.09.005.
- Juan Antonio Sánchez-Alcañiz, A. F.-M. (2018). (2018). *An expression atlas of variant ionotropic glutamate receptors identifies a molecular basis of carbonation sensing*.
- Koehrsen, W. (2019). (2019). *A straightforward walk-through of a useful statistical concept*.
- Krogh, A. L. (2001). (2001). *Predicting transmembrane protein topology with a hidden markov model: application to complete genomes*¹¹Edited by F. Cohen. *Journal of Molecular Biology*, 305(3), 567–580.
- Kumar S, S. G. (2018). (2018). *MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms* *Molecular Biology and Evolution* 35:1547-1549.
- Kumar S, T. K. (1994). (1994). *MEGA: molecular Evolutionary*.
- Lee JH, D. E. (2014). (2014). *Highly multiplexed subcellular RNA sequencing in situ*. *Science.* 2014;343(6177):1360–1363. doi:10.1126/science.1250212 .
- Lesk, A. M. (2017). (2017). *Introduction to Genomics*.
- Liping Wei, Y. L. (20020). (2002). *Comparative genomics approaches to study organism similarities and differences*.
- Luscombe NM, G. D. (2001). (2001). *What is bioinformatics? A proposed definition and overview of the field*. *Methods Inf Med.* 2001;40(4):346-58. Review. PubMed PMID: 11552348.

- Madeira F, P. Y. (2019). (2019). *The EMBL-EBI search and sequence analysis tools APIs in 2019*.
- Magadum S, B. U. (2013). (2013). *Gene duplication as a major force in evolution*. *J Genet*. 2013 Apr;92(1):155-61. Review. PubMed PMID: 23640422.
- Matsuo T, S. S. (2007). (2007). *Odorant-binding proteins OBP57d and OBP57e affect taste perception and host-plant preference in Drosophila sechellia*. *PLoS Biol*. 2007;5(5):e118. doi:10.1371/journal.pbio.0050118.
- McBride CS, A. J. (2009). . *Five Drosophila genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily [published correction appears in Genetics*. 2009 Oct;183(2):755]. *Genetics*. 2007;177(3):1395–.
- McDonnell E, S. K. (2018). (2018). *Manual Gene Curation and Functional Annotation*. *Methods Mol Biol*. 2018;1775:185-208. doi: 10.1007/978-1-4939-7804-5_16. PubMed PMID: 29876819.
- Mehdi Kchouk, J.-F. G. (2017). *Generations of Sequencing Technologies: From First to Next Generation*.
- Missbach C, D. H. (2014). (2014). *Evolution of insect olfactory receptors [published correction appears in Elife*. 2014;3:e05087]. *Elife*. 2014;3:e02115. Published 2014 Mar 26. doi:10.7554/eLife.02115.
- Mohamed I Abouelhoda, S. K. (2008). (2008). *CoCoNUT: an efficient system for the comparison and analysis of genomes*.
- Nardi F, C. A. (2003). (2003). *The mitochondrial genome of the olive fly Bactrocera oleae: two haplotypes from distant geographical locations*. *Insect Molecular Biology* 12: 605-611.
- National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – [cited 2017 Apr 06]. Available from: <https://www.ncbi.nlm.nih.gov/>. (1988).
- P.S. Narayana, D. V. (2016). (2016). *Research Methodology in Plant Science*.
- Papanicolaou, A. S. (2016). (2016). *The whole genome sequence of the Mediterranean fruit fly, Ceratitis capitata (Wiedemann), reveals insights into the biology and adapt*.
- Pavlidis N, G. A. (2017). *Transcriptomic responses of the olive fruit fly Bactrocera oleae and its symbiont Candidatus Erwinia dacicola to olive feeding*. *Sci Rep*. 2017;7:42633. Published 2017 Feb 22. doi:10.1038/srep42633.
- Poelchau M, C. C. (2015). 2015). *The i5k Workspace@NAL--enabling genomic data access, visualization and curation of arthropod genomes*. *Nucleic Acids Res*. 2015;43(Database issue):D714–D719. doi:10.1093/nar/gku983.

- Punta M, F. L. (2007). (2007). *Membrane protein prediction methods. Methods.* 2007;41(4):460–474. doi:10.1016/j.ymeth.2006.07.026.
- Rambaut, A. (2006). (2006). *Molecular evolution, phylogenetics and epidemiology.*
- Ray A, v. N. (2014). . (2014). *Molecular determinants of odorant receptor function in insects. J Biosci.* 2014;39(4):555–563. doi:10.1007/s12038-014-9447-7.
- Rice, R. (2000). (2000). *Bionomics of the olive fruit fly Bactrocera (Dacus) oleae. University of California Plant Protection Quarterly.* 10. 1-5. . .
- Robertson HM, W. K. (2006). (2006). *The chemoreceptor superfamily in the honey bee, Apis mellifera: expansion of the odorant, but not gustatory, receptor family. Genome Res.* 2006;16(11):1395–1403. doi:10.1101/gr.5057506.
- Rytz R, C. V. (2013). (2013). *Ionotropic receptors (IRs): chemosensory ionotropic glutamate receptors in Drosophila and beyond. Insect Biochem Mol Biol.* 2013 Sep;43(9):888-97. doi: 10.1016/j.ibmb.2013.02.007. Epub 2013 Mar 1. Review. PubMed PMID: 23459169.
- Rytz, R. C. (2013). *Ionotropic Receptors (IRs): Chemosensory ionotropic glutamate receptors in Drosophila and beyond. Insect Biochemistry and Molecular Biology,* 43(9), 888–897. doi:10.1016/j.ibmb.2013.02.007 .
- Sagri E, R. M. (2014). (2014). *Olive fly transcriptomics analysis implicates energy metabolism genes in spinosad resistance. BMC Genomics.* 2014;15:714. Published 2014 Aug 25. doi:10.1186/1471-2164-15-714.
- Saitou N, N. M. (1987). (1987). *The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol.* 1987 Jul;4(4):406-25. PubMed PMID: 3447015.
- Sánchez-Gracia A, V. F. (2009). (2009). *Molecular evolution of the major chemosensory gene families in insects. Heredity (Edinb).* 2009 Sep;103(3):208-16. doi: 10.1038/hdy.2009.55. Epub 2009 May 13. Review. PubMed PMID: 19436326.
- Scott JG, W. W. (2014). *Genome of the house fly, Musca domestica L., a global vector of diseases with adaptations to a septic environment. Genome Biol.* 2014;15(10):466. doi:10.1186/s13059-014-0466-3.
- SEQC/MAQC-III Consortium. *A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. Nat Biotechnol.* 2014;32(9):903–914. doi:10.1038/nbt.2957. (2014).
- Sievers F, W. A. (2011). 2011). *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology* 7:539 doi:10.1038/msb.2011.75.

- Su Z, F. H. (2014). (2014). *An investigation of biomarkers derived from legacy microarray data for their utility in the RNA-seq era*. *Genome Biol.* 2014;15(12):523. Published 2014 Dec 3. doi:10.1186/s13059-014-0523-y. .
- Su, Z. Ł.-M.-M. (2014). *A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium*. *Nature Biotechnology*, 32(9),903.
- Sun JS, X. S. (2018). (2018). *The diverse small proteins called odorant-binding proteins*. *Open Biol.* ;8(12):180208. doi:10.1098/rsob.180208.
- Suzuki HC, O. K. (2018). (2018). *Evolution of Gustatory Receptor Gene Family Provides Insights into Adaptation to Diverse Host Plants in Nymphalid Butterflies [published correction appears in Genome Biol Evol. 2018 Sep 1;10(9)]*.
- Thompson JD, G. T. (1997). (1997). *The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools*. *Nucleic Acids Res.* 1997;25(24):4876–4882. doi:10.1093/nar/25.24.4876.
- Thurmond J, G. J. (2019). *FlyBase 2.0: the next generation*. *Nucleic Acids Res.* 2019;47(D1):D759–D765. doi:10.1093/nar/gky1003.
- Touchman, J. (2010). (2010). *Comparative Genomics*. (School of Life Sciences, Arizona State University) © 2010 Nature Education .
- Vosshall LB, S. R. (2007). (2007). *Molecular architecture of smell and taste in Drosophila*. *Annu Rev Neurosci.* 2007;30:505-33. Review. PubMed PMID: 17506643.
- Wang Z, Y. P. (2015). (2015). *Identification and functional analysis of olfactory receptor family reveal unusual characteristics of the olfactory system in the migratory locust*. *Cell Mol Life Sci.* 2015;72(22):4429–4443. doi:10.1007/s00018-015-2009-9.
- WS, L. (2012). *Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes*.
- Wu Z, Z. H. (2015). *Discovery of Chemosensory Genes in the Oriental Fruit Fly, Bactrocera dorsalis*. *PLoS One.* 2015;10(6):e0129794. Published 2015 Jun 12. doi:10.1371/journal.pone.0129794.
- Xu YL, H. P. (2009). (2009). *Large-scale identification of odorant-binding proteins and chemosensory proteins from expressed sequence tags in insects*. *BMC Genomics.* 2009;10:632. Published 2009 Dec 25. doi:10.1186/1471-2164-10-632.
- Zasha Weinberg, J. X. (2010). (2010). *Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes*.

- Zhao J, T. A. (2015). *A generalized birth and death process for modeling the fates of gene duplication*. *BMC Evol Biol.* 2015;15:275. Published 2015 Dec 8. doi:10.1186/s12862-015-0539-2.
- Zheng W, P. W. (2013). *Identification and expression profile analysis of odorant binding proteins in the oriental fruit fly Bactrocera dorsalis*. *Int J Mol Sci.* 2013;14(7):14936–14949. Published 2013 Jul 17. doi:10.3390/ijms14.01.2005 Sep 6;15(17):R673-84. Review. PubMed PMID: 16139201.
- Zhou JJ, H. X. (2008). (2008). *Identification of odorant-binding proteins of the yellow fever mosquito Aedes aegypti: genome annotation and comparative analyses*. *Insect Mol Biol.* 2008 Apr;17(2):147-63. doi: 10.1111/j.1365-2583.2007.00789.x. Erratum in: *Insect Mol.*
- Zhou JJ, R. G. (2009). (2009). *Characterisation of Bombyx mori Odorant-binding proteins reveals that a general odorant-binding protein discriminates between sex pheromone components*. *Mol Biol.* 2009 Jun 12;.
- Zhou X, S. J. (2012). (2012). *Phylogenetic and transcriptomic analysis of chemosensory receptors in a pair of divergent ant species reveals sex-specific signatures of odor coding*. *PLoS Genet.* 2012;8(8):e1002930. doi: 10.1371/journal.pgen.1002930. Epub 2012 Au