



ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

**ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
"ΤΕΧΝΟΛΟΓΙΕΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΣΤΗΝ ΙΑΤΡΙΚΗ ΚΑΙ ΤΗ ΒΙΟΛΟΓΙΑ"**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**Εκτίμηση της μικροβιακής κατάστασης germ-free ποντικών
με τη χρήση RNA-Seq δεδομένων**

Φίλιππος Σ. Καρδαράς

Επιβλέπουσα: Άρτεμις Χατζηγεωργίου, Καθηγήτρια

ΑΘΗΝΑ

ΦΕΒΡΟΥΑΡΙΟΣ 2020

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Εκτίμηση της μικροβιακής κατάστασης germ-free ποντικών με τη χρήση RNA-Seq δεδομένων

Φίλιππος Σ. Καρδαράς

A.M.: ΠΙΒ0181

Επιβλέπουσα: **Άρτεμις Χατζηγεωργίου**, Καθηγήτρια Βιοπληροφορικής, Τμήμα Πληροφορικής με Εφαρμογές στη Βιοϊατρική, Πανεπιστήμιο Θεσσαλίας

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ: **Άρτεμις Χατζηγεωργίου**, Καθηγήτρια Βιοπληροφορικής, Τμήμα Πληροφορικής με Εφαρμογές στη Βιοϊατρική, Πανεπιστήμιο Θεσσαλίας

Martin Reczko, Ειδικός Λειτουργικός Επιστήμονας Α',
Ερευνητικό Κέντρο Βιοϊατρικών Επιστημών Αλέξανδρος
Φλέμιγκ

Δήμητρα Καραγκούνη, Μεταδιδακτορική Ερευνήτρια,
Ελληνικό Ινστιτούτο Παστέρ, Πανεπιστήμιο Θεσσαλίας

ΠΕΡΙΛΗΨΗ

Στην προσπάθεια διερεύνησης του ρόλου της φυσιολογικής μικροχλωρίδας του ανθρώπου στη φυσιολογία του, αλλά και των συσχετίσεών της με παθολογικές καταστάσεις, έχει βρει ευρεία εφαρμογή η χρήση ποντικών-μοντέλων με γνωστή μικροχλωρίδα (gnotobiotic mice) ή/και χωρίς καθόλου μικροχλωρίδα (germ-free mice), σε αυστηρά ελεγχόμενες συνθήκες.

Σκοπός της παρούσας εργασίας ήταν η εκτίμηση της μικροβιακής κατάστασης germ-free και gnotobiotic ποντικών χρησιμοποιώντας δεδομένα από πειράματα RNA Sequencing (RNA-Seq). Με τον τρόπο αυτό, εξετάζεται αν τα RNA-Seq δεδομένα γονιδιακής έκφρασης, από ιστούς των ποντικών, μπορούν να χρησιμοποιηθούν και για την διαπίστωση κάποιας μόλυνσης των germ-free ή gnotobiotic ποντικών από μη επιθυμητούς μικροοργανισμούς. Τα RNA-Seq δεδομένα που χρησιμοποιήθηκαν προέρχονται από τη βάση δεδομένων NCBI GEO και αναλύθηκαν με τρία προγράμματα μεταγονιδιωματικής ανάλυσης: Kraken 2, CLARK και AGAMEMNON.

Διαπιστώσαμε την ύπαρξη μικροβιακών αλληλουχιών σε αυτού του είδους τα δεδομένα και επιχειρήσαμε τον χαρακτηρισμό των μικροβίων και των γονιδίων τους που εκφράζονται. Προτείνουμε με αυτόν τον τρόπο, μια επιπλέον μεθοδολογία για τον έλεγχο ή την επιβεβαίωση της στειρότητας germ-free και gnotobiotic ποντικών.

Επιπρόσθετα, συλλέξαμε, από 101 επιστημονικές δημοσιεύσεις, 995 συσχετίσεις μικροβίων με ανθρώπινες ασθένειες, οι οποίες θα φανούν χρήσιμες στη δημιουργία αντίστοιχης βάσης δεδομένων.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Βιοπληροφορική

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: μικροβίωμα, μεταγονιδιωματική, germ-free, gnotobiotic, RNA-Seq

ABSTRACT

In the ongoing attempt to investigate the role of microbes in human physiology and their associations with pathological conditions, researchers have turned to using mouse models with known microbiota (gnotobiotic mice) or without microbiota (germ-free mice), maintained in well-controlled conditions.

The purpose of the present study was the evaluation of the microbial status of germ-free and gnotobiotic mice using data from RNA Sequencing (RNA-Seq) experiments. Thus, it is examined whether data from expression profiling with high throughput sequencing on mouse tissue samples can also be used to detect infections of germ-free or gnotobiotic mice with unwanted microorganisms. The RNA-Seq data were derived from the NCBI GEO database and were analyzed with three metagenomic programs: Kraken 2, CLARK and AGAMEMNON.

Eventually, we detected presence of microbial sequences in this type of data and attempted to identify the microbes and their expressed genes. In this way we propose an additional method for testing or confirming the sterility of germ-free and gnotobiotic mice.

In addition, we collected 995 experimentally supported microbe-disease associations, by manually curating 101 scientific papers, that will appear useful for the creation of a microbe-disease associations database.

SUBJECT AREA: Bioinformatics

KEYWORDS: microbiome, metagenomics, germ-free, gnotobiotic, RNA-Seq

ΕΥΧΑΡΙΣΤΙΕΣ

Για τη διεκπεραίωση της παρούσας διπλωματικής εργασίας, θα ήθελα να ευχαριστήσω την επιβλέπουσα καθηγήτρια Άρτεμις Χατζηγεωργίου για την ευκαιρία που μου έδωσε να ασχοληθώ με το παρόν αντικείμενο και την καθοδήγησή της. Επίσης, θα ήθελα να ευχαριστήσω τους Δρ. Martin Reczko και Δρ. Δήμητρα Καραγκούνη για τις γνώσεις που μου μετέδωσαν κατά τη διάρκεια των μεταπτυχιακών μου σπουδών και που δέχτηκαν να συμμετέχουν στην τριμελή εξεταστική επιτροπή. Τέλος, θα ήθελα να ευχαριστήσω τον υποψήφιο διδάκτορα Γιώργο Σκούφο για τη συνεργασία και την πολύτιμη συμβολή του στην ολοκλήρωση της.

ΠΕΡΙΕΧΟΜΕΝΑ

1. ΕΙΣΑΓΩΓΗ	17
1.1 Μικροοργανισμοί	17
1.1.1 Ορισμός.....	17
1.1.2 Ομάδες Μικροοργανισμών	17
1.2 Φυσιολογική Μικροχλωρίδα	19
1.2.1 Ορισμός.....	19
1.2.2 Μέγεθος.....	20
1.2.3 Διαμόρφωση της Ανθρώπινης Μικροχλωρίδας	21
1.2.4 Ρόλος Φυσιολογικής Μικροχλωρίδας	22
1.3 Συσχετίσεις με Ασθένειες	25
1.4 Χρήση Ποντικών στην Έρευνα του Μικροβιώματος	26
1.4.1 Διαφορές με τον Άνθρωπο.....	26
1.4.2 Gnotobiotic και Germ-Free ποντίκια	28
1.4.3 Χρησιμότητα GF ποντικών - Συμβολή στην κατανόηση της σχέσης μεταξύ φυσιολογικής μικροχλωρίδας και ξενιστή	29
1.4.4 Δημιουργία Germ-Free Ποντικών.....	31
1.4.5 Διατήρηση Germ-Free ποντικών και δυσκολίες	32
1.4.6 Έλεγχος πιθανών μολύνσεων GF ποντικών στο εργαστήριο	34
1.5 Μεταγονιδιωματική	35
1.6 Σκοπός της εργασίας	36
2. ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ	38
2.1 Αναζήτηση RNA-Seq Δεδομένων	38
2.2 Επιλογή Δειγμάτων	38
2.3 Κατέβασμα των αρχείων	38
2.4 Έλεγχος ποιότητας και προ-επεξεργασία αλληλουχιών	38
2.5 Στοιχισμός αλληλουχιών/reads στο γονιδίωμα του ποντικίου και εξαγωγή των χρήσιμων αλληλουχιών	39
2.6 Χρήση εργαλείων μεταγονιδιωματικής	39
2.6.1 KRAKEN2.....	40
2.6.2 CLARK	40

2.6.3	AGAMEMNON	41
2.7	Στοιχισή των αλληλουχιών σε γονιδιώματα επιλεγμένων μικροβίων και ταυτοποίηση των γονιδίων που εκφράζονται.....	42
2.8	Ροή Εργασίας (Workflow - Pipelines).....	42
3.	ΑΠΟΤΕΛΕΣΜΑΤΑ.....	45
3.1	Σειρά δειγμάτων GSE102543.....	45
3.1.1	Περιγραφή δειγμάτων, προ-επεξεργασία και στοίχιση στο γονιδίωμα του ποντικού.....	45
3.1.2	Μεταγονιδιωματική Ανάλυση.....	46
3.1.3	Στοιχισή στο γονιδίωμα της <i>E. coli</i>	49
3.2	Σειρά δειγμάτων GSE112571	53
3.2.1	Περιγραφή δειγμάτων, προ-επεξεργασία και στοίχιση στο γονιδίωμα του ποντικού.....	53
3.2.2	Μεταγονιδιωματική Ανάλυση.....	55
3.2.3	Στοιχισή στο γονιδίωμα του <i>Bacteroides thetaiotaomicron</i> και του HIV-1	58
3.3	Σειρά δειγμάτων GSE81117	62
3.3.1	Περιγραφή δειγμάτων, προ-επεξεργασία και στοίχιση στο γονιδίωμα του ποντικού.....	62
3.3.2	Μεταγονιδιωματική Ανάλυση.....	64
3.3.3	Στοιχισή στο γονιδίωμα της <i>E. coli</i>	66
3.4	Ανάλυση άλλων δειγμάτων	72
3.4.1	Σειρά δειγμάτων GSE115814	72
3.4.2	Σειρά δειγμάτων GSE122312	76
3.4.3	Σειρά δειγμάτων GSE77221	78
3.4.4	Σειρές δειγμάτων GSE108718 και GSE133809.....	81
4.	ΣΥΝΟΨΗ - ΣΥΖΗΤΗΣΗ.....	86
	ΠΙΝΑΚΑΣ ΟΡΟΛΟΓΙΑΣ	88
	ΣΥΝΤΜΗΣΕΙΣ – ΑΡΚΤΙΚΟΛΕΞΑ – ΑΚΡΩΝΥΜΙΑ	89
	ΠΑΡΑΡΤΗΜΑ Ι.....	90
	ΑΝΑΦΟΡΕΣ	94

ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

Εικόνα 1.1 Ανάπτυξη της μικροχλωρίδας της εντερικής οδού από τη γέννηση μέχρι τα γηρατειά. Αυτό το διάγραμμα δείχνει τα κυρίαρχα βακτηριακά taxa στα διαφορετικά στάδια της ζωής και πώς αλλάζουν υπό διαφορετικές συνθήκες υγείας και διατροφής. Η αρχική δομή της κοινότητας επηρεάζεται από τον τρόπο γέννησης και τον τύπο της διατροφής. Η ποικιλομορφία αυξάνεται καθώς το βρέφος μεγαλώνει και αρχίζει να τρώει στερεά τρόφιμα. Διαφορετικές καταστάσεις ασθένειας, δίαιτες και περιβαλλοντικές συνθήκες παίζουν ρόλο στη διαμόρφωση της μικροβιακής κοινότητας καθ 'όλη τη διάρκεια της ζωής [17].22

Εικόνα 1.2 Ολική ανατομία της γαστρεντερικής οδού του ανθρώπου και του ποντικού. (Α) Το ανθρώπινο κόλον διαιρείται σε διαφορετικά τμήματα (ανιόν, εγκάρσιο, κατión και σιγμοειδές) με την παρουσία κολικών ταινιών (*taenia coli*) και τη διαμερισματοποίηση σε κολικές κυψέλες (*haustra*), τα οποία απουσιάζουν στο παχύ έντερο του ποντικού. Το ανθρώπινο στομάχι (C) είναι επενδυμένο με έναν αδενικό βλεννογόνο που εκκρίνει γαστρικό οξύ, ενώ ο στόμαχος του ποντικού (B) διαιρείται σε δύο περιοχές - ένα αδενικό στομάχι και ένα μη αδενικό. Ο αδενικός στόμαχος του ποντικού είναι υπεύθυνος για την έκκριση του γαστρικού οξέος, ενώ ο μη αδενικός στόμαχος λειτουργεί ως προσωρινός χώρος αποθήκευσης και πέψης της τροφής. (Ε) Διατομή ανθρώπινου παχέος εντέρου, που έχει παχύτερο μυϊκό τοίχωμα και βλεννογόνο σε σύγκριση με το παχύ έντερο του ποντικού (D). M: βλεννογόνος, ME: *muscularis externa*; TC: κολική ταινία [51, 55].28

Εικόνα 1.3 Δημιουργία GF ποντικιών με καισαρική τομή. (Α) Αφαίρεση του αμνιακού σάκου αφαιρείται. (Β) Ο αμνιακός σάκος τοποθετείται σε γυάλινο βάζο που περιέχει απολυμαντικό. (Γ) Ο αμνιακός σάκος μεταφέρεται στον προστατευτικό θάλαμο, όπου ανοίγεται και τα νεογνά αφαιρούνται, καθαρίζονται και διεγείρονται προκειμένου να αναπνεύσουν. (Δ) Τα νεογνά παραδίδονται στη θετή GF μητέρα [62].32

Εικόνα 1.4 Αποστειρωμένη χειρουργική επέμβαση εμβρυομεταφοράς ποντικού. Όλα τα υλικά, συμπεριλαμβανομένου του μικροσκοπίου και της πιπέτας, πρέπει να είναι αποστειρωμένα [61].32

Εικόνα 1.5 a) Απομονωτής ανοξείδωτου χάλυβα με ατμό αποστείρωσης, Πανεπιστήμιο Notre Dame, 1946. (1. Τεχνικός, 2. Ηλεκτρική σύνδεση, 3. Έξοδος αέρα, 4. Βάση μεταφοράς, 5. Είσοδος / έξοδος αυτοκλείστου). b) Gustafsson απομονωτής από χάλυβα [62], c) Πλαστικός απομονωτής. Το φίλτρο αέρα και η πόρτα του αποστειρωμένου θαλάμου εισόδου βρίσκονται στο δεξιό άκρο του απομονωτή [63].33

Εικόνα 1.6 Ένας σύγχρονος απομονωτής [62].	34
Εικόνα 2.1 Ροή Εργασίας (workflow) που περιγράφει τα βήματα που ακολουθούμε για κάθε σειρά δειγμάτων.	44
Εικόνα 3.1 Τα 6 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 6 δείγματα (11 διαφορετικά είδη) και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, σύμφωνα με το πρόγραμμα Kraken 2, για τη σειρά δειγμάτων με id: GSE102543. Κάθε χρώμα αντιστοιχεί σε διαφορετικό δείγμα.	47
Εικόνα 3.2 Τα 6 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 6 δείγματα (9 διαφορετικά είδη) και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, σύμφωνα με το πρόγραμμα CLARK, για τη σειρά δειγμάτων με id: GSE102543. Κάθε χρώμα αντιστοιχεί σε διαφορετικό δείγμα.	48
Εικόνα 3.3 Τα 6 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 6 δείγματα (8 διαφορετικά είδη) και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, σύμφωνα με το πρόγραμμα AGAMEMNON, για τη σειρά δειγμάτων με id: GSE102543. Κάθε χρώμα αντιστοιχεί σε διαφορετικό δείγμα.	48
Εικόνα 3.4 Τα δέκα πιο συχνά γονίδια σε κάθε ένα από τα έξι δείγματα (δέκα διαφορετικά γονίδια, κοινά και στα έξι δείγματα) και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, για τη σειρά δειγμάτων με id: GSE102543. Κάθε χρώμα αντιστοιχεί σε διαφορετικό δείγμα. Και τα δέκα γονίδια είναι rRNA γονίδια.	50
Εικόνα 3.5 Ένα παράδειγμα της στοίχισης αλληλουχιών του δεύτερου δείγματος στο γονιδίωμα της <i>E. coli</i> , και συγκεκριμένα στα rRNA γονίδια rrsC, rrlC, όπως φαίνεται με το πρόγραμμα οπτικοποίησης στοιχίσεων IGV. Τα δύο αυτά γονίδια ανήκουν στο ίδιο οπερόνιο (βλ. Εικόνα 3.6).	51
Εικόνα 3.6 Οπερόνιο rrsC, στο γονιδίωμα της <i>E. coli</i> . Αποτελείται από τα γονίδια rrsC (16S rRNA), gltU (tRNA-Glu(UUC)), rrlC 23S rRNA και rrfC (5S rRNA) [78].	51
Εικόνα 3.7 Τα δέκα πιο συχνά γονίδια, εξαιρουμένων των rRNA γονιδίων, στα έξι δείγματα και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, για τη σειρά δειγμάτων με id: GSE102543. Κάθε χρώμα αντιστοιχεί σε διαφορετικό δείγμα. Στον ένθετο πίνακα αναγράφεται το προϊόν (product) κάθε γονιδίου. Ξεχωρίζει το γονίδιο ssaA που παράγει ένα RNA με δράση μεταφορικού-RNA και αγγελιαφόρου-RNA (tmRNA).	52

Εικόνα 3.8 Ένα παράδειγμα της στοίχισης αλληλουχιών του δεύτερου δείγματος στο γονιδίωμα της <i>E. coli</i> , και συγκεκριμένα στο γονίδιο <i>ssrA</i> , το οποίο ήταν το πιο συχνό και στα έξι δείγματα. Η ισχυρή παρουσία του γονιδίου αυτού οφείλεται σε πολλά αντίγραφα αλληλουχιών/reads που καλύπτουν όλη την έκταση του γονιδίου, μήκους 362 ζευγών βάσεων.	53
Εικόνα 3.9 Τα 6 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 12 δείγματα (24 διαφορετικά είδη) και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, σύμφωνα με το πρόγραμμα Kraken 2, για τη σειρά δειγμάτων με id: GSE112571. Κάθε χρώμα αντιστοιχεί σε διαφορετικό δείγμα.	56
Εικόνα 3.10 Τα 6 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 12 δείγματα (21 διαφορετικά είδη) και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, σύμφωνα με το πρόγραμμα CLARK, για τη σειρά δειγμάτων με id: GSE112571. Κάθε χρώμα αντιστοιχεί σε διαφορετικό δείγμα.	57
Εικόνα 3.11 Τα 6 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 12 δείγματα (16 διαφορετικά είδη) και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, σύμφωνα με το πρόγραμμα AGAMEMNON, για τη σειρά δειγμάτων με id: GSE112571. Κάθε χρώμα αντιστοιχεί σε διαφορετικό δείγμα.	57
Εικόνα 3.12 Τα πέντε πιο συχνά γονίδια σε κάθε ένα από τα δώδεκα δείγματα (έξι διαφορετικά γονίδια) και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, για τη σειρά δειγμάτων με id: GSE112571. Κάθε χρώμα αντιστοιχεί σε διαφορετικό δείγμα. Και τα έξι γονίδια είναι rRNA γονίδια.	59
Εικόνα 3.13 Ένα παράδειγμα της στοίχισης αλληλουχιών του δέκατου δείγματος στο γονιδίωμα του <i>Bacteroides thetaiotaomicron</i> , και συγκεκριμένα στα rRNA γονίδια BT_r02 (23S ribosomal) και BT_r03 (16S ribosomal), όπως φαίνεται με το πρόγραμμα οπτικοποίησης στοίχισης IGV.	60
Εικόνα 3.14 Τα πέντε πιο συχνά γονίδια, εξαιρουμένων των rRNA γονιδίων, στα έξι δείγματα και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, για τη σειρά δειγμάτων με id: GSE112571. Κάθε χρώμα αντιστοιχεί σε διαφορετικό δείγμα. Στον οριζόντιο άξονα αναφέρεται το προϊόν (product) κάθε γονιδίου.	60
Εικόνα 3.15 Ένα παράδειγμα της στοίχισης αλληλουχιών του δέκατου δείγματος στο γονιδίωμα του <i>Bacteroides thetaiotaomicron</i> και συγκεκριμένα στο γονίδιο BT_2510, που κωδικοποιεί το ένζυμο αφυδρογονάση μηλικού (Malate dehydrogenase). Η ισχυρή	

παρουσία του γονιδίου αυτού οφείλεται σε πολλά αντίγραφα αλληλουχιών που καλύπτουν μεγάλο μέρος του μήκους του γονιδίου.	61
Εικόνα 3.16 Στοίχιση των αλληλουχιών του τρίτου δείγματος (χωρισμένες σε δύο τεχνικά αντίγραφα με SRA ids: SRR6921776, SRR6921777) στο γονιδίωμα του ιού HIV-1. Οι αλληλουχίες του δείγματος καλύπτουν όλο σχεδόν το μήκος του γονιδιώματος του ιού.....	62
Εικόνα 3.17 Τα 6 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 15 δείγματα (17 διαφορετικά είδη) και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, σύμφωνα με το πρόγραμμα Kraken 2, για τη σειρά δειγμάτων με id: GSE81117. Κάθε χρώμα αντιστοιχεί σε διαφορετική κατηγορία δειγμάτων σύμφωνα με το υπόμνημα.....	65
Εικόνα 3.18 Τα 6 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 15 δείγματα (15 διαφορετικά είδη) και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, σύμφωνα με το πρόγραμμα CLARK, για τη σειρά δειγμάτων με id: GSE81117. Κάθε χρώμα αντιστοιχεί σε διαφορετική κατηγορία δειγμάτων σύμφωνα με το υπόμνημα.....	65
Εικόνα 3.19 Τα 6 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 15 δείγματα (7 διαφορετικά είδη) και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, σύμφωνα με το πρόγραμμα AGAMEMNON, για τη σειρά δειγμάτων με id: GSE81117. Κάθε χρώμα αντιστοιχεί σε διαφορετική κατηγορία δειγμάτων σύμφωνα με το υπόμνημα.	66
Εικόνα 3.20 Τα έξι πιο συχνά γονίδια σε κάθε ένα από τα δεκαπέντε δείγματα (εννέα διαφορετικά γονίδια) και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, για τη σειρά δειγμάτων με id: GSE81117. Κάθε χρώμα αντιστοιχεί σε διαφορετικό δείγμα και κάθε σχήμα σε διαφορετική κατηγορία ποντικών σύμφωνα με το υπόμνημα. Και τα εννέα γονίδια είναι rRNA γονίδια.	68
Εικόνα 3.21 Τα έξι πιο συχνά γονίδια, εξαιρουμένων των rRNA γονιδίων, σε κάθε ένα από τα δεκαπέντε δείγματα (δέκαπέντε διαφορετικά γονίδια) και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, για τη σειρά δειγμάτων με id: GSE81117. Κάθε χρώμα αντιστοιχεί σε διαφορετικό δείγμα και κάθε σχήμα σε διαφορετική κατηγορία ποντικών σύμφωνα με το υπόμνημα. Στον ένθετο πίνακα αναγράφεται το προϊόν (product) κάθε γονιδίου.....	68
Εικόνα 3.22 Ένα παράδειγμα της στοίχισης αλληλουχιών του πρώτου και του τρίτου δείγματος στο γονιδίωμα της <i>E. coli</i> , και συγκεκριμένα στο γονίδιο <i>ssrA</i> , το οποίο ήταν το δεύτερο πιο συχνό σε όλα τα δείγματα. Η ισχυρή παρουσία του γονιδίου αυτού	

οφείλεται σε πολλά αντίγραφα αλληλουχιών/reads που καλύπτουν όλη την έκταση του γονιδίου, μήκους 362 ζευγών βάσεων.....	69
Εικόνα 3.23 Ένα παράδειγμα της στοίχισης αλληλουχιών του πρώτου και του τρίτου δείγματος στο γονιδίωμα της <i>E. coli</i> , και συγκεκριμένα στο γονίδιο <i>gprB</i> , το οποίο ήταν το πιο συχνό σε όλα τα δείγματα. Η ισχυρή παρουσία του γονιδίου αυτού οφείλεται σε πολλά αντίγραφα αλληλουχιών/reads που καλύπτουν όλη την έκταση του γονιδίου.....	70
Εικόνα 3.24 Ένα παράδειγμα της στοίχισης αλληλουχιών του πρώτου και του τρίτου δείγματος στο γονιδίωμα της <i>E. coli</i> , και συγκεκριμένα στο γονίδιο <i>trmD</i> . Η παρουσία του γονιδίου αυτού οφείλεται σε λίγα αντίγραφα αλληλουχιών/reads που καλύπτουν μεγάλο μέρος της έκταση του γονιδίου. Αλληλουχίες/reads εντοπίζονται και στο γειτονικό γονίδιο <i>rimM</i> . Και τα δύο γονίδια βρίσκονται στο ίδιο οπερόνιο (βλ. Εικόνα 3.25).	71
Εικόνα 3.25 Οπερόνιο των γονιδίων <i>rpsP</i> , <i>rim</i> , <i>trmD</i> , <i>rplS</i> , στο γονιδίωμα της <i>E. coli</i> [78].	71
Εικόνα 3.26 Αποτελέσματα των προγραμμάτων Kraken 2 και CLARK για τη σειρά δειγμάτων GSE115814. Διαπιστώνεται η ύπαρξη του <i>Bacteroides thetaiotaomicron</i> . Το ποσοστό των reads (% Proportion) που αντιστοιχούν στο κάθε είδος μικροοργανισμού υπολογίστηκε ως προς τον αριθμό των reads που υπήρχαν σε κάθε δείγμα μετά την προ-επεξεργασία/φιλτράρισμα (βλ. Πίνακα 3.10). Κάθε χρώμα αντιστοιχεί σε διαφορετική κατηγορία δειγμάτων.....	73
Εικόνα 3.27 Αποτελέσματα των προγραμμάτων Kraken 2 και CLARK για τη σειρά δειγμάτων GSE115814, όπου φαίνονται τα 6 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 22 δείγματα. Διαπιστώνεται η ύπαρξη κάποιων μη αναμενόμενων μικροοργανισμών (<i>Blautia hansenii</i>) με πολλές αλληλουχίες/reads και με τα δύο προγράμματα.	75
Εικόνα 3.28 Αποτελέσματα των προγραμμάτων Kraken 2 και CLARK για τη σειρά δειγμάτων GSE122312, όπου φαίνονται τα 12 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 6 δείγματα. Διαπιστώνεται η ύπαρξη κάποιων μη αναμενόμενων μικροοργανισμών με πολλές αλληλουχίες/reads.	77
Εικόνα 3.29 Αποτελέσματα των προγραμμάτων Kraken 2 και CLARK για τη σειρά δειγμάτων GSE77221. Έχουν επιλεγεί μόνο οι 12 πιο συχνοί μικροοργανισμοί κάθε δείγματος. Το ποσοστό των reads (% Proportion) που αντιστοιχούν στο κάθε είδος μικροοργανισμού υπολογίστηκε ως προς τον αριθμό των reads που υπήρχαν σε κάθε	

δείγμα μετά την προ-επεξεργασία/φιλτράρισμα (βλ. Πίνακα 3.12). Μεγαλύτερο ποσοστό αλληλουχιών αποδίδεται σε μικρόβια στα δείγματα GF απ' ό τι στα δείγματα CTR..... 79

Εικόνα 3.30 Αποτελέσματα των προγραμμάτων Kraken 2 και CLARK για τη σειρά δειγμάτων GSE77221, όπου φαίνονται τα 12 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 4 δείγματα και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα. ...80

Εικόνα 3.31 Αποτελέσματα των προγραμμάτων Kraken 2 και CLARK για τη σειρά δειγμάτων GSE108718, όπου φαίνονται τα 6 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 14 δείγματα και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα. .82

Εικόνα 3.32 Αποτελέσματα των προγραμμάτων Kraken 2 και CLARK για τη σειρά δειγμάτων GSE108718. Έχουν επιλεγεί μόνο οι 6 πιο συχνοί μικροοργανισμοί κάθε δείγματος. Το ποσοστό των reads (% Proportion) που αντιστοιχούν στο κάθε είδος μικροοργανισμού υπολογίστηκε ως προς τον αριθμό των reads που υπήρχαν σε κάθε δείγμα μετά την προ-επεξεργασία/φιλτράρισμα (βλ. Πίνακα 3.13)..... 83

Εικόνα 3.33 Αποτελέσματα των προγραμμάτων Kraken 2 και CLARK για τη σειρά δειγμάτων GSE133809, όπου φαίνονται τα 6 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 14 δείγματα και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα. .85

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας 1.1 Μέγιστη τάξη μεγέθους του αριθμού των βακτηρίων σε διάφορα όργανα/μέρη του σώματος. Συμπεριλαμβάνονται η τυπική συγκέντρωση βακτηρίων ανά mL διαθέσιμου όγκου και τα mL του διαθέσιμου όγκου σε κάθε όργανο/μέρος του σώματος. (Για το δέρμα χρησιμοποιείται η βακτηριακή πυκνότητα της περιοχής και η ολική επιφάνεια του δέρματος, αντί της συγκέντρωσης και του όγκου, αντίστοιχα) [13].	21
Πίνακας 3.1 Χαρακτηριστικά δειγμάτων της σειράς με id GSE102543	45
Πίνακας 3.2 Ποσοστό αλληλουχιών που δεν στοιχίζονται στο γονιδίωμα του ποντικού (unmapped pairs%) στα έξι δείγματα. Επίσης, παρουσιάζεται ο NCBI GEO κωδικός του κάθε δείγματος (sample id), ο SRA κωδικός του (SRA id), ο χαρακτηρισμός του (tag , MC=monocolonized), ο αριθμός των αλληλουχιών μετά την προ-επεξεργασία/φιλτράρισμα (filtered sra pairs) και ο αριθμός των αλληλουχιών που δεν στοιχίζονται στο γονιδίωμα του ποντικού (unmapped pairs).....	46
Πίνακας 3.3 Στατιστικά στοίχισης στο γονιδίωμα της <i>E. coli</i> . Αριθμός αλληλουχιών που δε στοιχήθηκαν στο γονιδίωμα του ποντικού (unmapped reads). Ποσοστό των unmapped reads που στοιχήθηκαν στο γονιδίωμα της <i>E. coli</i> (mapped to <i>E. coli</i> %). Ποσοστό των unmapped reads που στοιχήθηκαν σε γονίδια της <i>E. coli</i> (mapped to <i>E. coli</i> genes %). Αριθμός γονιδίων με τουλάχιστον μια αλληλουχία/read στην περιοχή τους.	49
Πίνακας 3.4 Χαρακτηριστικά δειγμάτων της σειράς με id GSE112571	54
Πίνακας 3.5 Ποσοστό αλληλουχιών που δεν στοιχίζονται στο γονιδίωμα του ποντικού (unmapped pairs%) στα δώδεκα δείγματα. Επίσης, παρουσιάζεται ο NCBI GEO κωδικός του κάθε δείγματος (sample id), ο SRA κωδικός των δύο τεχνικών αντιγράφων του (SRA id), ο χαρακτηρισμός του (tag , MC=monocolonized), ο αριθμός των αλληλουχιών μετά την προ-επεξεργασία/φιλτράρισμα (filtered sra pairs) και ο αριθμός των αλληλουχιών που δεν στοιχίζονται στο γονιδίωμα του ποντικού (unmapped pairs).	55
Πίνακας 3.6 Στατιστικά στοίχισης στο γονιδίωμα του <i>B. thetaiotaomicron</i> . Αριθμός αλληλουχιών που δε στοιχήθηκαν στο γονιδίωμα του ποντικού (unmapped reads). Ποσοστό των unmapped reads που στοιχήθηκαν στο γονιδίωμα του <i>B. thetaiotaomicron</i> (mapped to <i>B. thetaiotaomicron</i> %). Ποσοστό των unmapped reads που στοιχήθηκαν	

σε γονίδια του <i>B. thetaiotaomicron</i> (mapped to <i>B. thetaiotaomicron</i> genes %). Αριθμός γονιδίων με τουλάχιστον μια αλληλουχία/read στην περιοχή τους.....	58
Πίνακας 3.7 Χαρακτηριστικά δειγμάτων της σειράς με id GSE81117	63
Πίνακας 3.8 Ποσοστό αλληλουχιών που δεν στοιχίζονται στο γονιδίωμα του ποντικού (unmapped reads%) στα δεκαπέντε δείγματα. Επίσης, παρουσιάζεται ο NCBI GEO κωδικός του κάθε δείγματος (sample id), ο SRA κωδικός του (SRA id), ο χαρακτηρισμός του (tag, GF=germ-free, ConVR=Conventionally raised, ConVD=conventionalized), ο αριθμός των αλληλουχιών μετά την προ-επεξεργασία/φιλτράρισμα (filtered sra reads) και ο αριθμός των αλληλουχιών που δεν στοιχίζονται στο γονιδίωμα του ποντικού (unmapped reads).	63
Πίνακας 3.9 Στατιστικά στοίχισης στο γονιδίωμα της <i>E. coli</i> . Αριθμός αλληλουχιών που δε στοιχήθηκαν στο γονιδίωμα του ποντικού (unmapped reads). Ποσοστό των unmapped reads που στοιχήθηκαν στο γονιδίωμα της <i>E. coli</i> (mapped to <i>E. coli</i> %). Ποσοστό των unmapped reads που στοιχήθηκαν σε γονίδια της <i>E. coli</i> (mapped to <i>E. coli</i> genes %). Αριθμός γονιδίων με τουλάχιστον μια αλληλουχία/read στην περιοχή τους.	67
Πίνακας 3.10 Αριθμός των αλληλουχιών μετά την προ-επεξεργασία/φιλτράρισμα των δειγμάτων (filtered sra pairs). Επίσης, παρουσιάζεται ο NCBI GEO κωδικός του κάθε δείγματος (sample id), ο SRA κωδικός του (SRA id) και ο χαρακτηρισμός του (tag, MC=επικοισμένο με <i>Bacteroides thetaiotaomicron</i> , GB 2= επικοισμένο με <i>Bacteroides thetaiotaomicron</i> και <i>Candidatus Arthromitus</i> sp. SFB-mouse, GB 3= επικοισμένο με <i>Bacteroides thetaiotaomicron</i> , <i>Candidatus Arthromitus</i> sp. SFB-mouse και <i>Clostridium</i> species, GB 4= επικοισμένο με <i>Bacteroides thetaiotaomicron</i> , <i>Candidatus Arthromitus</i> sp. SFB-mouse, <i>Clostridium</i> species και <i>Lactobacillus</i> species).	72
Πίνακας 3.11 Αριθμός των αλληλουχιών μετά την προ-επεξεργασία/φιλτράρισμα των δειγμάτων (filtered sra pairs). Επίσης, παρουσιάζεται ο NCBI GEO κωδικός του κάθε δείγματος (sample id), οι SRA κωδικοί του (SRA id) και ο χαρακτηρισμός του (tag).	76
Πίνακας 3.12 Αριθμός των αλληλουχιών μετά την προ-επεξεργασία/φιλτράρισμα των δειγμάτων (filtered sra pairs) της σειράς GSE77221. Επίσης, παρουσιάζεται ο NCBI GEO κωδικός του κάθε δείγματος (sample id), οι SRA κωδικοί του (SRA id) και ο χαρακτηρισμός του (tag).....	78
Πίνακας 3.13 Αριθμός των αλληλουχιών μετά την προ-επεξεργασία/φιλτράρισμα των δειγμάτων (filtered sra pairs) της σειράς GSE108718. Επίσης, παρουσιάζεται ο NCBI	

GEO κωδικός του κάθε δείγματος (sample id), οι SRA κωδικοί του (SRA id) και ο χαρακτηρισμός του (tag).....81

Πίνακας 3.14 Αριθμός των αλληλουχιών μετά την προ-επεξεργασία/φιλτράρισμα των δειγμάτων (filtered sra pairs) της σειράς GSE133809. Επίσης, παρουσιάζεται ο NCBI GEO κωδικός του κάθε δείγματος (sample id), οι SRA κωδικοί του (SRA id) και ο χαρακτηρισμός του (tag).....84

1. ΕΙΣΑΓΩΓΗ

1.1 Μικροοργανισμοί

1.1.1 Ορισμός

Μικροοργανισμοί (ή μικρόβια) είναι οι ζώντες οργανισμοί που βρίσκονται παντού γύρω μας και είναι τόσο μικροί ώστε να μην είναι ορατοί με γυμνό μάτι. Πολλοί από τους μικροοργανισμούς περνούν όλη τη ζωή τους στο φυσικό περιβάλλον: ζουν στο έδαφος, πάνω σε φυτά, στις λίμνες, στα ποτάμια και στους ωκεανούς. Άλλοι, προκειμένου να επιβιώσουν και να αναπαραχθούν, περνούν ένα μέρος ή ολόκληρη τη ζωή τους στο εσωτερικό κάποιου πολυκύτταρου ζωικού οργανισμού. Το ανθρώπινο σώμα φιλοξενεί εκατομμύρια τέτοιους μικροοργανισμούς [1].

1.1.2 Ομάδες Μικροοργανισμών

Υπάρχουν πέντε κύριες ομάδες μικροοργανισμών: τα βακτήρια (προκαρυωτικοί οργανισμοί), οι μύκητες, τα φύκη, τα πρωτόζωα (ευκαρυωτικοί οργανισμοί) και οι ιοί. Από αυτές, εμείς θα ασχοληθούμε με τα βακτήρια, τους μύκητες, τα πρωτόζωα και τους ιούς καθώς είναι αυτές που βρίσκονται (ή μπορεί να βρεθούν) σε έναν ανθρώπινο οργανισμό.

- **Βακτήρια**

Τα βακτήρια είναι μονοκύτταροι προκαρυωτικοί οργανισμοί. Ορισμένα χρειάζονται οξυγόνο για να επιβιώσουν (υποχρεωτικά αερόβια), όπως τα βακτήρια του γένους *Mycobacterium*, άλλα αναπτύσσονται με ταχύτερο ρυθμό παρουσία οξυγόνου (προαιρετικά αερόβια) και άλλα, όπως βακτήρια του γένους *Clostridium* επιβιώνουν μόνο απουσία οξυγόνου (υποχρεωτικά αναερόβια) [2].

Τα περισσότερα βακτήρια δεν είναι επικίνδυνα για τον άνθρωπο. Πολλά ζούνε πάνω ή μέσα στο σώμα μας και είναι απαραίτητα για την επιβίωσή μας. Για παράδειγμα το βακτήριο *Escherichia coli* στο έντερο συνθέτει τη βιταμίνη Κ την οποία εμείς δεν παράγουμε, ενώ άλλα βακτήρια συμβάλλουν στην άμυνα του οργανισμού καθώς ανταγωνίζονται τα παθογόνα βακτήρια.

Όσον αφορά τη δομή των βακτηρίων, το γενετικό τους υλικό (DNA) βρίσκεται κατά κανόνα σε μια συγκεκριμένη περιοχή, που ονομάζεται πυρηνική περιοχή (πυρηνοειδές), ενώ συχνά διαθέτουν, επιπλέον, μικρότερα μόρια γενετικού υλικού, τα πλασμίδια. Η πλασματική τους μεμβράνη περιβάλλεται από το κυτταρικό τοίχωμα, το οποίο προσδίδει σταθερότητα στο κύτταρο και είναι απαραίτητο για την αύξηση και διαίρεση

του. Διαφέρει σε πάχος και σύνθεση από βακτήριο σε βακτήριο, ώστε να χρησιμοποιείται ως γνώρισμα για τον προσδιορισμό και την ταξινόμηση των μικροοργανισμών. Το κυτταρικό τοίχωμα των θετικών κατά Gram βακτηρίων περιέχει περισσότερη πεπτιδογλυκάνη, περίπου 50% του ξηρού βάρους του κυτταρικού τοιχώματος, πολυσακχαρίτες και τειχικά οξέα τα οποία βοηθούν στη μεταφορά και αποθήκευση ιόντων φωσφόρου. Αντίθετα το κυτταρικό τοίχωμα των αρνητικών κατά Gram βακτηρίων είναι πολυπλοκότερο: μια εξωτερική μεμβράνη περιβάλλει την πεπτιδογλυκάνη, η οποία αποτελεί μόνο 5-10% του ξηρού βάρους του κυτταρικού τοιχώματος [2]. Ορισμένα βακτήρια διαθέτουν και ένα επιπλέον περίβλημα, την κάψα.

- **Μύκητες**

Οι μύκητες αποτελούν ένα από τα 5 βασίλεια των ευκαρυωτικών οργανισμών και συναντώνται σε πολλά διαφορετικά ενδιαιτήματα. Πρόκειται για ετερότροφους οργανισμούς, είτε σαπροφυτικούς (παίρνουν τις οργανικές ενώσεις από νεκρή οργανική ύλη) είτε βιοτροφικούς (παράσιτα ή ξενιστές) είτε και τα δύο. Όπως και τα βακτήρια, μερικοί μύκητες ζουν φυσικά στο δέρμα ή στο σώμα μας ωστόσο μπορεί να προκαλέσουν και ασθένειες.

Οι ασθένειες που προκαλούνται από τους μύκητες ονομάζονται μυκητιάσεις, με χαρακτηριστικά παραδείγματα την πνευμονική καντιντίαση, την κολπίτιδα, τη στοματίτιδα και την προσβολή των μεσοδακτύλιων περιοχών από τα δερματόφυτα (ομάδα μυκήτων). Στην τελευταία περίπτωση, οι μύκητες μεγαλώνουν ανάμεσα στα κύτταρα του ξενιστή με τη χρήση σχηματισμών σαν ριζίδια που καλούνται μυζητήρες. Ο κνησμός που εμφανίζεται οφείλεται στον ερεθισμό αισθητήριων απολήξεων από τέτοιους μυζητήρες [3].

Σχετικά με τη δομή των μυκήτων, αυτοί είναι μονοκύτταροι (οπότε χαρακτηρίζονται ως ζύμες) ή κοινοκυτταρικοί οργανισμοί (δηλαδή διαθέτουν κυτταρόπλασμα με πολυάριθμους πυρήνες). Στην τελευταία περίπτωση, αυτοί αποτελούνται συνήθως από απλούστερες νηματοειδείς δομές, τις υφές (όπως η κοινή μούχλα). Το δίκτυο που σχηματίζουν οι υφές καλείται μυκήλιο και μπορεί είτε να εισχωρεί στο θρεπτικό υπόστρωμα είτε να είναι εναέριο. Επειδή δεν έχουν εσωτερικό σύστημα πέψης εκκρίνουν πεπτικά ένζυμα στο θρεπτικό μέσο και τα προϊόντα της πέψης απορροφούνται από τα κύτταρα του μύκητα, με σχηματισμούς σα ριζίδια, τα ριζοειδή [4].

- **Αρχαία**

Αν και μερικά είδη των Αρχαίων ζουν σε σχετικά ήπια περιβάλλοντα, όπως στο πεπτικό σύστημα των ομοιόθερμων ζώων, εν τούτοις το γνώρισμα που τα χαρακτηρίζει είναι η ικανότητα τους να ευδοκιμούν σε ακραία περιβάλλοντα. Τα κύτταρα των αρχαίων είναι προκαρυωτικά (δε διαθέτουν πυρήνα) και έχουν μοναδικές ιδιότητες που τα διαχωρίζουν από τις άλλες δύο επικράτειες των έμβιων όντων, τα Βακτήρια και τους Ευκαρυώτες. Η πλασματική τους μεμβράνη αποτελείται από μοναδικά λιπίδια που δε βρίσκονται σε άλλους οργανισμούς, ενώ ορισμένα διαθέτουν και κυτταρικό τοίχωμα, όχι όμως από μόρια πεπτιδογλυκάνης όπως στα βακτήρια. Τα αρχαία χωρίζονται περαιτέρω σε πολλές συνομοταξίες, αλλά η ταξινόμηση είναι δύσκολη επειδή τα περισσότερα δεν έχουν απομονωθεί στο εργαστήριο και έχουν μόνο εντοπιστεί μέσω της ανάλυσης των νουκλεϊκών οξέων σε δείγματα από το περιβάλλον.

- **Ιοί**

Οι ιοί, σε αντίθεση με τους μικροοργανισμούς που αναφέρθηκαν, δεν αποτελούνται από κύτταρα και αυτό συνεπάγεται ότι, αυστηρά μιλώντας, δεν μπορούμε να τους θεωρήσουμε οργανισμούς. Είναι υποχρεωτικά παράσιτα, αφού χρησιμοποιούν τους μεταβολικούς μηχανισμούς του κυττάρου ξενιστή για να αντιγραφούν, και τα κύρια ιδιαίτερα χαρακτηριστικά τους είναι το μέγεθος, το είδος του γενετικού υλικού και οι μεταβολικές τους ικανότητες.

Το μέγεθός τους είναι πολύ μικρό (20-300 nm) επομένως είναι ορατοί μόνο με το ηλεκτρονικό μικροσκόπιο. Μπορεί να περιέχουν είτε DNA είτε RNA ως γενετικό υλικό, το οποίο περιβάλλεται από ένα πρωτεϊνικό περίβλημα, το καψίδιο. Ορισμένοι ιοί διαθέτουν και ένα επιπλέον περίβλημα, το έλυτρο, το οποίο είναι λιποπρωτεϊνικής φύσης.

Πολλοί ιοί του ανθρώπου είναι υπεύθυνοι για ήπιες (όπως το κρυολόγημα) ή σοβαρές ασθένειες (όπως η πολιομυελίτιδα και η ηπατίτιδα). Στο σώμα μας, οι ιοί μολύνουν όχι μόνο κύτταρα μας, αλλά και μικρόβια που ζουν στο σώμα μας. Οι ιοί που μολύνουν τα βακτήρια ονομάζονται βακτηριοφάγοι.

1.2 Φυσιολογική Μικροχλωρίδα

1.2.1 Ορισμός

Η φυσιολογική μικροχλωρίδα του ανθρώπου (human microbiota) αποτελείται από 10-100 τρισεκατομμύρια μικροοργανισμούς που ζουν συμβιωτικά με τον άνθρωπο σε όλες σχεδόν τις σωματικές επιφάνειες και τους βλεννογόνους του, και στην πλειονότητά τους

πρόκειται για βακτήρια που βρίσκονται στο παχύ έντερο. Με τον όρο μικροβίωμα (microbiome) θα αναφερόμαστε στο σύνολο αυτών των μικροοργανισμών και των γονιδίων που φέρουν [5].

Μεγάλα project παγκοσμίως έχουν πραγματοποιηθεί με στόχο την ταυτοποίηση των μικροοργανισμών αυτών, την κατανόηση του ρόλου τους και των επιπτώσεών τους στην ανθρώπινη υγεία [6, 7]. Δυσκολίες σε αυτή την προσπάθεια αποτελούν οι μεταβολές της φυσιολογικής μικροχλωρίδας του κάθε ατόμου, ο υποκειμενικός ορισμός των OTUs (λειτουργικές ταξινομικές μονάδες, operational taxonomic units) που συνθέτουν τη μικροχλωρίδα, και η αμφισβήτηση του αν ένα άτομο έχει ένα μικροβίωμα ή πολλά. (OTU είναι η βασική μονάδα της αριθμητικής ταξινομικής και είναι ο όρος που δίνεται στην κατώτατη μελετώμενη ταξινομική μονάδα σε μία έρευνα).

Οι μέθοδοι χαρακτηρισμού μικροβίων που δεν απαιτούν την καλλιέργειά τους, μαζί με την μοριακή φυλογενετική προσέγγιση για τον προσδιορισμό της ποικιλομορφίας της ζωής, επέτρεψαν στους ερευνητές να συγκρίνουν διαφορετικές μικροβιακές κοινότητες με κοινά κριτήρια, απαλλαγμένα από την υποκειμενική μακροσκοπική παρατήρηση ή βιοχημική ανάλυση που χρησιμοποιούνταν παλαιότερα [8].

Αν και τα συμβιωτικά μικρόβια του ανθρώπου προήλθαν πιθανώς από το φυσικό περιβάλλον, η σύνθεση της μικροχλωρίδας των θηλαστικών, ειδικά στο έντερο, είναι εκπληκτικά διαφορετική από αυτή στα ελεύθερα οικοσυστήματα [9]. Αυτό υποδηλώνει ότι η συμβίωση σπονδυλωτών και μικροβίων για εκατοντάδες εκατομμύρια χρόνια οδήγησε στη φυσική επιλογή εξειδικευμένης ομάδας μικροοργανισμών που ευδοκίμουν στο ζεστό, ευτροφικό και σταθερό περιβάλλον του εντέρου [9]. Όπως αναφέρθηκε, στη φυσιολογική μικροχλωρίδα του ανθρώπου τα βακτήρια αποτελούν τον κύριο όγκο της βιομάζας και της ποικιλομορφίας, αν και τα αρχαία, οι μύκητες και οι ιοί είναι επίσης παρόντα σε μικρότερους αριθμούς και δεν πρέπει να παραβλέπονται [10-12].

1.2.2 Μέγεθος

Αν επικεντρωθούμε μόνο στα βακτήρια, οι αναφερόμενες τιμές στη βιβλιογραφία σχετικά με τον αριθμό τους στο ανθρώπινο σώμα διαφέρουν κατά τάξεις μεγέθους και πολύ σπάνια υποστηρίζονται από οποιεσδήποτε μετρήσεις ή υπολογισμούς. Σε μια πρόσφατη μελέτη, ο συνολικός αριθμός των βακτηρίων στον μέσο «άνθρωπο αναφοράς», βάρους 70 kg, εκτιμήθηκε ότι είναι $3,8 \cdot 10^{13}$ ενώ τα ανθρώπινα κύτταρα υπολογίστηκαν σε $3,0 \cdot 10^{13}$, εκ των οποίων περίπου το 90% είναι αιμοποιητικά κύτταρα. Έτσι, η ευρέως αναφερόμενη αναλογία «10 βακτήρια : 1 ανθρώπινο κύτταρο» μπορεί

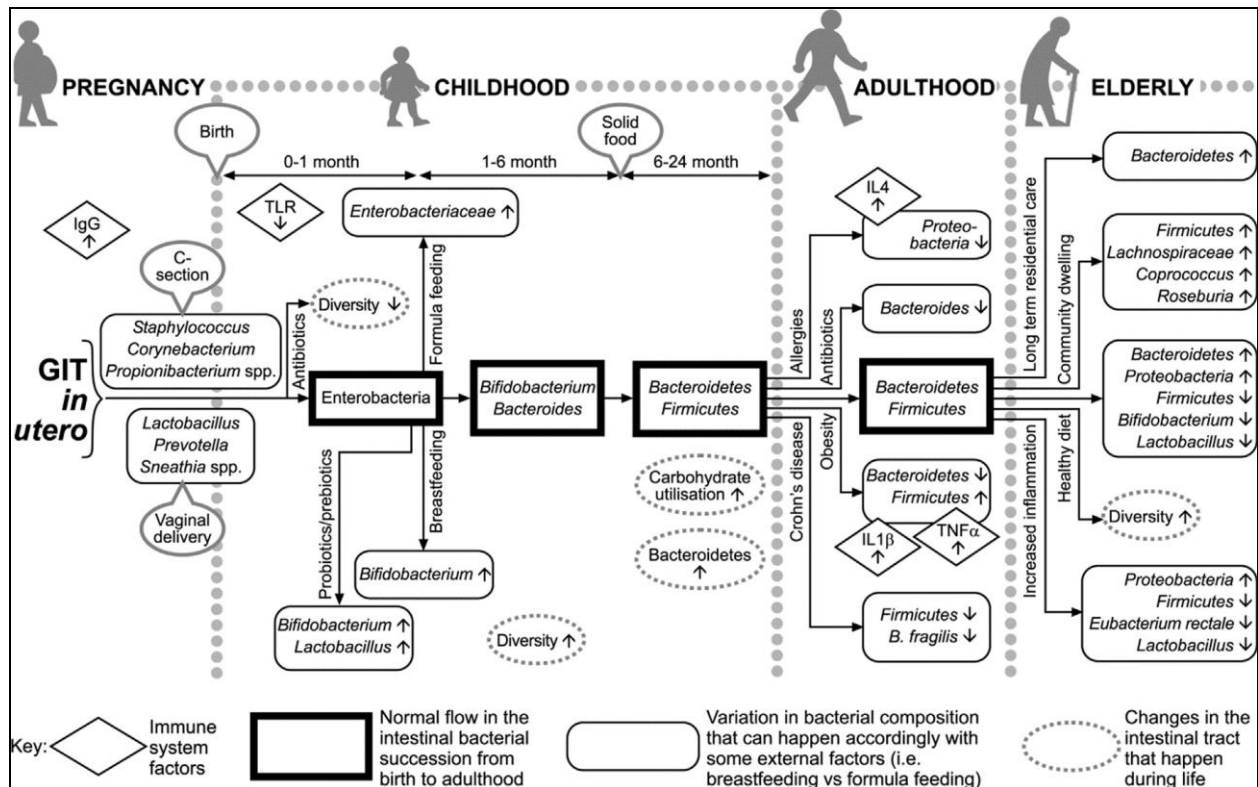
να αναθεωρηθεί σε 1:1, δείχνοντας ότι ο αριθμός των βακτηρίων στο σώμα είναι στην πραγματικότητα της ίδιας τάξης μεγέθους με τον αριθμό των ανθρώπινων κυττάρων, και η συνολική τους μάζα είναι περίπου 0,2 kg [13]. Στον **Πίνακα 1.1** βλέπουμε πως αυτά κατανέμονται στο ανθρώπινο σώμα.

Πίνακας 1.1 Μέγιστη τάξη μεγέθους του αριθμού των βακτηρίων σε διάφορα όργανα/μέρη του σώματος. Συμπεριλαμβάνονται η τυπική συγκέντρωση βακτηρίων ανά mL διαθέσιμου όγκου και τα mL του διαθέσιμου όγκου σε κάθε όργανο/μέρος του σώματος. (Για το δέρμα χρησιμοποιείται η βακτηριακή πυκνότητα της περιοχής και η ολική επιφάνεια του δέρματος, αντί της συγκέντρωσης και του όγκου, αντίστοιχα) [13].

Θέση	Τυπική συγκέντρωση βακτηρίων (αριθμός/ mL)	Όγκος (mL)	Ανώτατη τάξη μεγέθους του αριθμού των βακτηρίων
Παχύ έντερο	10^{11}	400	10^{14}
Οδοντική πλάκα	10^{11}	<10	10^{12}
Κάτω λεπτό έντερο	10^8	400	10^{11}
Σάλιο	10^9	<100	10^{11}
Δέρμα	$<10^{11}$ per m^2	$1.8 m^2$	10^{11}
Στομάχι	$10^3 - 10^4$	250 – 900	10^7
Άνω λεπτό έντερο	$10^3 - 10^4$	400	10^7

1.2.3 Διαμόρφωση της Ανθρώπινης Μικροχλωρίδας

Η μέθοδος του τοκετού είναι αυτή που επηρεάζει ποια είδη μικροβίων ένα βρέφος θα αποκτήσει αρχικά [14]. Είκοσι λεπτά μετά τη γέννηση, η μικροχλωρίδα των βρεφών που γεννήθηκαν από τον κόλπο της μητέρας είναι όμοια με την κολπική μικροχλωρίδα της μητέρας τους, ενώ τα βρέφη που ήρθαν στον κόσμο με καισαρική τομή έχουν μικροχλωρίδα που αποτελείται από μικρόβια που συναντώνται συνήθως στο ανθρώπινο δέρμα [15]. Η απόκτηση νέων μικροβίων συνεχίζεται τα πρώτα χρόνια ζωής, και η εντερική μικροχλωρίδα του βρέφους αρχίζει να μοιάζει με αυτή ενός ενήλικα ήδη από το 1ο έτος της ζωής του [14]. Είναι ενδιαφέρον ότι κάθε διαιτητική αλλαγή στην πορεία της ζωής του ανθρώπου, από το στάδιο του βρέφους ως την ενηλικίωση και μετά, συνοδεύεται από αλλαγές στη μικροχλωρίδα του [16] (**Εικόνα 1.1**).



Εικόνα 1.1 Ανάπτυξη της μικροχλωρίδας της εντερικής οδού από τη γέννηση μέχρι τα γηρατειά.

Αυτό το διάγραμμα δείχνει τα κυρίαρχα βακτηριακά taxa στα διαφορετικά στάδια της ζωής και πώς αλλάζουν υπό διαφορετικές συνθήκες υγείας και διατροφής. Η αρχική δομή της κοινότητας επηρεάζεται από τον τρόπο γέννησης και τον τύπο της διατροφής. Η ποικιλομορφία αυξάνεται καθώς το βρέφος μεγαλώνει και αρχίζει να τρώει στερεά τρόφιμα. Διαφορετικές καταστάσεις ασθένειας, δίαιτες και περιβαλλοντικές συνθήκες παίζουν ρόλο στη διαμόρφωση της μικροβιακής κοινότητας καθ 'όλη τη διάρκεια της ζωής [17].

1.2.4 Ρόλος Φυσιολογικής Μικροχλωρίδας

Η φυσιολογική μικροχλωρίδα του ανθρώπου θα μπορούσαμε να πούμε ότι έχει επίδραση σε 4 διαφορετικά επίπεδα στον άνθρωπο: στο μεταβολισμό των θρεπτικών, στο ανοσοποιητικό σύστημα, στη συμπεριφορά και στις ασθένειες.

- **Μεταβολισμός Θρεπτικών**

- ❖ **Υδατάνθρακες**

Τα μικρόβια του εντέρου αποκομίζουν σε μεγάλο βαθμό τα θρεπτικά τους από τους υδατάνθρακες της τροφής του ανθρώπου. Η ζύμωση των υδατανθράκων που διέφυγαν της χώνευσης στο δωδεκαδάκτυλο (το πρώτο τμήμα του λεπτού εντέρου) και των άπεπτων (από τους μικροοργανισμούς του παχέος εντέρου όπως τα *Bacteroides*, *Roseburia*, *Bifidobacterium*, *Fecalibacterium* και *Enterobacteria*) ολιγοσακχαριτών έχει ως αποτέλεσμα τη σύνθεση λιπαρών οξέων βραχείας αλυσίδας (SCFA) όπως βουτυρικό, προπιονικό και οξικό, που είναι πλούσιες πηγές ενέργειας για τον ξενιστή

[18, 19]. Επιπλέον, το βουτυρικό μπορεί να αποτρέψει τη συσσώρευση τοξικών μεταβολικών παραπροϊόντων όπως το D-γαλακτικό [20]. Είδη του γένους *Bacteroides*, που είναι οι κυρίαρχοι οργανισμοί που συμμετέχουν στον μεταβολισμό των υδατανθράκων, το επιτυγχάνουν με την έκφραση ενζύμων όπως οι γλυκοσυλικές τρανσφεράσες, οι υδρολάσες γλυκοσιδίων και οι λυάσες πολυσακχαριτών. Το καλύτερο παράδειγμα μεταξύ αυτών των οργανισμών είναι το *Bacteroides thetaiotaomicron* που εκφράζει πάνω από 260 υδρολάσες [21]. Επίσης, το οξαλικό που συντίθεται στο έντερο από τη ζύμωση των υδατανθράκων και το μεταβολισμό των βακτηρίων «καταπολεμάται» από μικροοργανισμούς (*Oxalobacter formigenes*, *Lactobacillus* και *Bifidobacterium*), μειώνοντας έτσι τον κίνδυνο σχηματισμού οξαλικής πέτρας στα νεφρά [22, 23].

❖ Λίπη

Τα μικρόβια του εντέρου έχει επίσης αποδειχθεί ότι έχουν θετική επίδραση στον μεταβολισμό των λιπιδίων, καταστέλλοντας την αναστολή της δραστηριότητας της λιποπρωτεϊνικής λιπάσης στα λιποκύτταρα. Επιπλέον, το *Bacteroides thetaiotaomicron* αποδεικνύεται ότι αυξάνει την αποτελεσματικότητα της λιπιδικής υδρόλυσης ρυθμίζοντας την έκφραση μιας co-λιπάσης (colipase), που απαιτείται από την παγκρεατική λιπάση για την πέψη των λιπιδίων [24].

❖ Πρωτεΐνες

Τα μικρόβια του εντέρου διαθέτουν επίσης ένα αποτελεσματικό μηχανισμό μεταβολισμού πρωτεϊνών που λειτουργεί μέσω των μικροβιακών πρωτεϊνών και πεπτιδασών παράλληλα με τις ανθρώπινες πρωτεϊνάσες. Αρκετοί μεταφορείς αμινοξέων στο βακτηριακό κυτταρικό τοίχωμα διευκολύνουν την είσοδο αμινοξέων από τον εντερικό σωλήνα στα βακτήρια, όπου πολλά γονιδιακά προϊόντα μετατρέπουν τα αμινοξέα είτε σε μικρά μόρια σηματοδότησης είτε σε αντιμικροβιακά πεπτίδια (βακτηριοκίνες) [25].

❖ Βιταμίνες και φλαβονοειδή

Η σύνθεση της βιταμίνης K και διάφορων μερών της βιταμίνης B είναι μια άλλη σημαντική μεταβολική λειτουργία των μικροβίων του εντέρου. Επίσης, είδη του γένους *Bacteroides* συνθέτουν συζευγμένο λινελαϊκό οξύ (conjugated linoleic acid, CLA) που είναι γνωστό ότι έχει δράση κατά του διαβήτη, της αθηρωματικής πλάκας, της παχυσαρκίας και της υπερλιπιδαιμίας και συμμετέχει στην ανοσορύθμιση [26, 27]. Πρόσφατες μελέτες έχουν δείξει ότι η φυσιολογική μικροχλωρίδα του εντέρου

εμπλέκεται επίσης στη διάσπαση διαφόρων πολυφαινολών (φαινολικών ενώσεων) που καταναλώνονται με την τροφή. Οι πολυφαινόλες υπάρχουν ως γλυκοζυλιωμένα παράγωγα συνδεδεμένα με σάκχαρα, όπως γλυκόζη, γαλακτόζη και ριβουλόζη, και μετατρέπονται σε δραστικές ενώσεις μετά την αφαίρεση του σακχάρου, χάρις στη συνδρομή μικροβίων του εντέρου. Τα τελικά ενεργά προϊόντα απορροφώνται από την πυλαία φλέβα και ταξιδεύουν σε άλλους ιστούς και όργανα, παρέχοντας έτσι αντιμικροβιακή και άλλη μεταβολική δράση [28].

- **Άμυνα**

Οι άνθρωποι αποκτούν τα πρώτα τους μικρόβια από την είσοδο του τραχήλου της μητέρας τους κατά την άφιξή τους στον κόσμο ή με άλλους τρόπους (επαφή με το σώμα άλλων ατόμων). Χωρίς αυτούς τους πρώτους μικροβιακούς επισκέπτες, δεν θα υπήρχε η προσαρμοστική (ή επίκτητη) ανοσία. Αυτός είναι ένας ζωτικός αμυντικός μηχανισμός που μαθαίνει στον ανθρώπινο οργανισμό πώς να ανταποκρίνεται στα μικρόβια μετά την είσοδο αυτών. Αυτό επιτρέπει μια ταχύτερη και αποτελεσματικότερη ανοσοαπόκριση σε μικροοργανισμούς που προκαλούν ασθένειες [29]. Ως αντιπαράδειγμα, τα ποντίκια που αναπτύσσονται ελεύθερα μικροβίων (germ-free mice, βλ. [Gnotobiotic και Germ-Free ποντίκια](#)) παρουσιάζουν υποανάπτυκτο ανοσοποιητικό σύστημα. Η μικροχλωρίδα επίσης σχετίζεται με αυτοάνοσες καταστάσεις και αλλεργίες, οι οποίες είναι πιθανότερο να αναπτυχθούν όταν η έκθεση σε μικρόβια διαταραχθεί στα αρχικά χρόνια ζωής [30, 31].

Τα συμβιωτικά βακτήρια προστατεύουν τον άνθρωπο από παθογόνα και με άμεσο τρόπο. Συγκεκριμένα, πολλά βακτήρια αναστέλλουν άμεσα την εποίκιση του εντέρου από παθογόνα καθώς είτε ανταγωνίζονται για τα ίδια θρεπτικά είτε προκαλούν την παραγωγή ανασταλτικών ουσιών. Για παράδειγμα το συμβιωτικό βακτήριο *Bacteroidetes thetaiotaomicron* καταναλώνει υδατάνθρακες που χρησιμοποιούνται από το παθογόνο *Citrobacter rodentium*, οδηγώντας έτσι σε ανταγωνιστικό αποκλεισμό του παθογόνου από το έντερο [32].

Μέσω της παραγωγής συγκεκριμένων μεταβολιτών, το εντερικό μικροβίωμα μπορεί επίσης να τροποποιήσει τις περιβαλλοντικές συνθήκες εντός του ξενιστή, και έτσι να εμποδίσει την εγκατάσταση και τον πολλαπλασιασμό παθογόνων. Το βουτυρικό, ένα λιπαρό οξύ βραχείας αλυσίδας (SCFA) που παράγεται από το εντερικό μικροβίωμα (βλ. Μεταβολισμός θρεπτικών), μπορεί να ρυθμίσει προς τα κάτω την έκφραση αρκετών γονιδίων πολλαπλασιασμού της *Salmonella enterica* serovar Enteritidis (*S. Enteritidis*) και *Typhimurium* (*S. Typhimurium*) [33] και έχει αποδειχθεί ότι αναστέλλει την ανάπτυξη

της εντεροαιμορραγικής *Escherichia coli* (EHEC) [34]. Ορισμένα στελέχη του γένους *Bifidobacterium* μπορούν επίσης να προστατεύσουν από την μόλυνση με EHEC μέσω της παραγωγής οξικού οξέος [34]. Η εντερική μικροβιακή κοινότητα είναι επίσης ικανή να παράγει ένα μεγάλο αριθμό βιοενεργών μικρών μορίων τα οποία δρουν προς άλλα μέλη της εντερικής μικροβιακής κοινότητας ή/και προς εντερικά παθογόνα. Οι βακτηριοκίνες, για παράδειγμα, είναι αντιμικροβιακά πεπτιδία που μπορούν να θανατώσουν επιλεκτικά ή να αναστείλουν την ανάπτυξη ανταγωνιστικών βακτηρίων [35].

• Συμπεριφορά

Η μικροχλωρίδα του εντέρου, το πεπτικό σύστημα και το κεντρικό νευρικό σύστημα αλληλεπιδρούν, σχηματίζοντας έτσι το λεγόμενο άξονα εντέρου-εγκεφάλου (gut-brain axis). Αυτό συμβαίνει μέσω προσαγωγών και απαγωγών νευρικών, ενδοκρινολογικών, θρεπτικών και ανοσολογικών σημάτων [36]. Για παράδειγμα, μερικά εντερικά μικρόβια προκαλούν συμπεριφορά παρόμοια με το άγχος και την κατάθλιψη καθώς ελέγχουν διάφορους τύπους υποδοχέων στον εγκέφαλο: α) GABA (γ-αμινοβουτιρικό οξύ) υποδοχείς, β) γλουταμινεργικοί υποδοχείς, γ) NMDA υποδοχείς και δ) 5-HT_{1A} σεροτονινεργικοί υποδοχείς. Αντίθετα, τα ελεύθερα μικροβίων ποντίκια παρουσιάζουν μειωμένη συμπεριφορά άγχους [37, 38]. Οι ερευνητές έχουν επίσης παρατηρήσει σχέσεις μεταξύ του μικροβιακού εντέρου και των ψυχιατρικών διαταραχών, όπως η κατάθλιψη και η διαταραχή του φάσματος του αυτισμού (ASD) (βλ. επόμενη ενότητα).

1.3 Συσχετίσεις με Ασθένειες

Οι μεταβολές στη σύνθεση και τη λειτουργία του μικροβιώματος του εντέρου έχουν συσχετιστεί με πολλές παθολογικές καταστάσεις που κυμαίνονται από μεταβολικές διαταραχές όπως η παχυσαρκία [39, 40] και ο διαβήτης τύπου 2 [41], μέχρι σύνθετες ασθένειες, όπως η νόσος του Crohn και η ελκώδης κολίτιδα, που είναι φλεγμονώδεις νόσοι του εντέρου [42], και αυτοάνοσες ασθένειες, όπως η ρευματοειδής αρθρίτιδα [43] και η αλλεργία [31]. Πιο πρόσφατα, η λειτουργία του άξονα εντέρου-εγκεφάλου, που αναφέρθηκε στην προηγούμενη παράγραφο έχει επίσης προσελκύσει προσοχή σε γαστρεντερικές διαταραχές, όπως το σύνδρομο ευερέθιστου εντέρου, αλλά και σε πιο απροσδόκητες παθολογίες, όπως ο αυτισμός [44].

Αρκετές μελέτες έχουν βρει στοιχεία γαστρεντερικής (GI, gastrointestinal) φλεγμονής στη σχιζοφρένεια. Τα αντισώματα έναντι του μύκητα *Saccharomyces cerevisiae*, που είναι δείκτης GI φλεγμονής, είναι αυξημένα σε άτομα με πρώτο επεισόδιο ψύχωσης,

σχιζοφρένεια και διπολική διαταραχή [45, 46]. Επιπλέον, οι δείκτες, στον ορό του αίματος, βακτηριακής μετατόπισης (ή αλλόθεσης) συσχετίστηκαν με τα επίπεδα CRP (C-Reactive Protein) του ορού, υποδηλώνοντας ότι η GI φλεγμονή συμβάλλει σε χαμηλού επιπέδου συστημική φλεγμονή [47]. Σε αυτήν, μπορεί να συμβάλει η αύξηση της διαπερατότητας του γαστρεντερικού σωλήνα, αφού μελέτες έχουν βρει και αυξημένα αντισώματα έναντι των τροφών σε άτομα με σχιζοφρένεια [48]. Αυτά τα ευρήματα υποδεικνύουν ότι τα άτομα με ψυχωσικές διαταραχές μπορεί να υποφέρουν από φλεγμονή του γαστρεντερικού σωλήνα και «διαρροή του εντέρου» (“leaky gut”), τα οποία μπορεί να συμβάλλουν σε ανοσολογικές μεταβολές στις ψυχωσικές διαταραχές.

1.4 Χρήση Ποντικών στην Έρευνα του Μικροβιώματος

Τα μοντέλα ποντικών έχουν χρησιμοποιηθεί ευρέως στη βιοϊατρική έρευνα. Οι εκτεταμένες ομοιότητες τους με τον ανθρώπινο οργανισμό, στην ανατομία, τη φυσιολογία και τη γενετική, επέτρεψαν να αντληθούν πολυάριθμα συμπεράσματα σχετικά με την ανθρώπινη βιολογία από τα πείραμα σε αυτά. Επιπλέον, το χαμηλό κόστος συντήρησής τους (σε σύγκριση με άλλα πειραματικά μοντέλα θηλαστικών), οι υψηλοί ρυθμοί αναπαραγωγής και ο μικρός κύκλος ζωής είναι σημαντικά πλεονεκτήματα του μοντέλου ποντικών. Η μελέτη της μικροχλωρίδας μοντέλων ποντικών στο εργαστήριο επιτρέπει την εξέταση αιτιωδών σχέσεων μεταξύ μικροοργανισμών και ασθενειών και την εξεύρεση πιθανών θεραπειών [49].

1.4.1 Διαφορές με τον Άνθρωπο

Παρόλο που τα αποτελέσματα από τέτοια πειράματα συνέβαλαν πολύ στην κατανόηση της δυναμικής και πολυπλοκότητας της σχέσης μεταξύ της μικροχλωρίδας του εντέρου και του ξενιστή, η μετάφραση τέτοιων αποτελεσμάτων από μοντέλα ποντικών στους ανθρώπους παραμένει δύσκολη λόγω της ύπαρξης ορισμένων βασικών διαφορών μεταξύ των δύο συστημάτων που χρειάζονται να ληφθούν υπόψη.

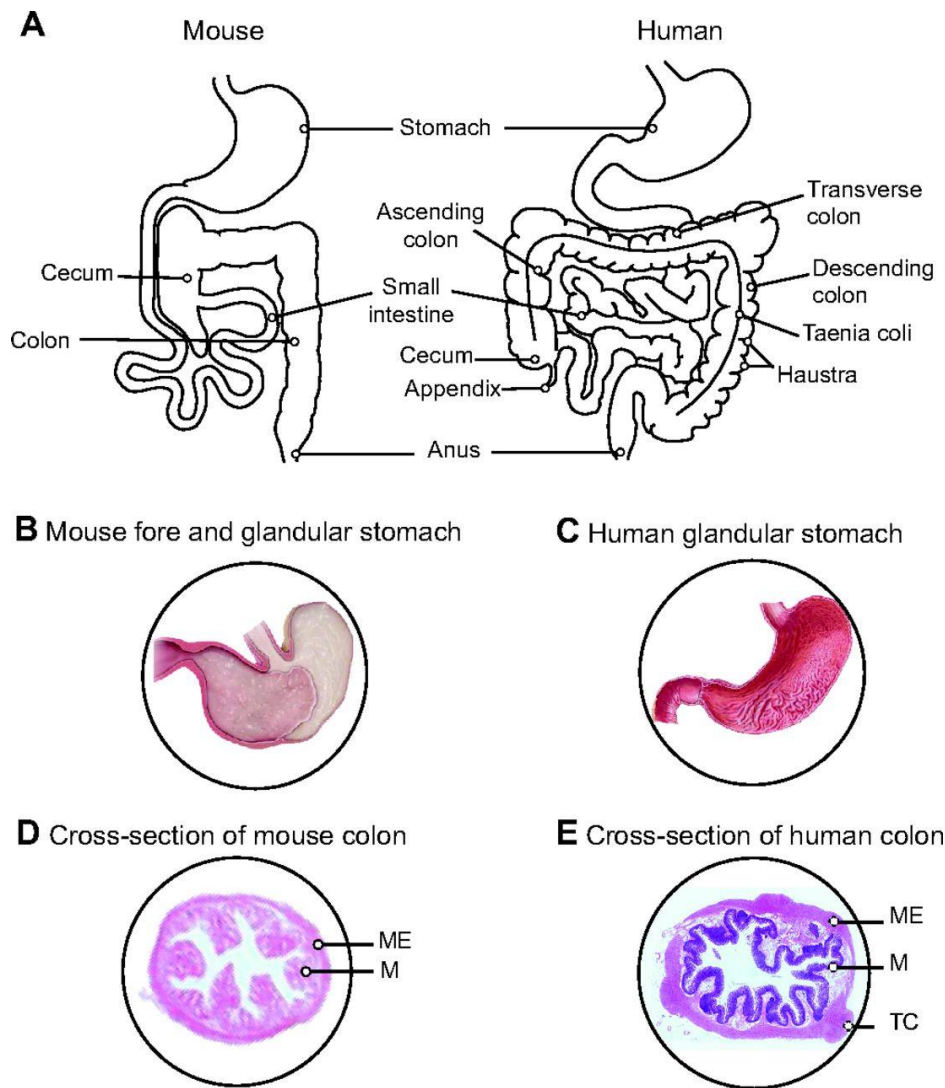
Το ποντίκι και ο άνθρωπος είναι αρκετά όμοια στη φυσιολογία και την ανατομία τους. Ειδικά, οι γαστρεντερικές οδοί και στα δύο είδη αποτελούνται από όργανα που είναι ανατομικά παρόμοια. Ωστόσο, η ανατομία του εντερικού σωλήνα στα δύο είδη έχει επίσης σημαντικές διαφορές (**Εικόνα 1.2**), οι οποίες μπορεί να έχουν διαμορφωθεί από τις διαφορετικές δίαιτες, τα πρότυπα διατροφής, τα μεγέθη του σώματος και τις μεταβολικές απαιτήσεις.

Αν και η μέση αναλογία επιφάνειας εντέρου:επιφάνειας σώματος είναι παρόμοια μεταξύ ποντικών και ανθρώπων [50], ο λόγος αυτός διαφέρει πολύ μεταξύ των δύο ειδών σε

συγκεκριμένα τμήματα/διαμερίσματα του εντέρου. Για παράδειγμα, ο λόγος «μήκος λεπτού εντέρου προς μήκος του κόλον» είναι 2,5 σε ποντικούς έναντι 7 σε ανθρώπους [51] και ο λόγος «επιφάνεια λεπτού εντέρου προς επιφάνεια του κόλον» είναι μόνο 18 στους ποντικούς σε σύγκριση με 400 στους ανθρώπους, κατά μέσο όρο [50].

Αυτές οι μορφολογικές διαφορές αντικατοπτρίζουν την προσαρμογή του ποντικού για αυξημένη χωρητικότητα στο κόλον και το τυφλό έντερο, επιτρέποντάς τους να απορροφούν θρεπτικά συστατικά από την τροφή τους η οποία περιέχει μεγαλύτερο ποσοστό από άπεπτα συστατικά, σε σχέση με την τροφή του ανθρώπου [49].

Σχετικά με τη σύσταση της μικροχλωρίδας, στον άνθρωπο και τον ποντικό κυριαρχούν δύο βακτηριακά φύλα, τα *Bacteroidetes* και *Firmicutes* [40, 52, 53]. Ωστόσο, στις κατώτερες ταξινομικές βαθμίδες, οι Ley *et al.* έδειξαν ότι το 85% από τα βακτηριακά γένη που βρίσκονται στην εντερική μικροχλωρίδα του ποντικού δεν υπάρχουν στον άνθρωπο [53]. Επιπροσθέτως, η διαβίωση των ποντικών στις ελεγχόμενες συνθήκες ενός εργαστηρίου για πολλές συνεχόμενες γενιές, μπορεί να οδηγήσει σε μείωση της βιοποικιλότητας της μικροχλωρίδας του εντέρου. Συγκεκριμένα, μια μελέτη των Linnenbrink *et al.* αποκαλύπτει ότι τα φυσικού τύπου (wild type) ποντίκια έχουν μεγαλύτερη βακτηριακή ποικιλότητα στο τυφλό έντερο απ' ό,τι τα εργαστηριακά ποντίκια που στεγάζονται σε εργαστηριακές SPF (specific pathogen free) συνθήκες [54].



Εικόνα 1.2 Ολική ανατομία της γαστρεντερικής οδού του ανθρώπου και του ποντικού. (A) Το ανθρώπινο κόλον διαιρείται σε διαφορετικά τμήματα (ανιόν, εγκάρσιο, κατίον και σιγμοειδές) με την παρουσία κολικών ταινιών (taenia coli) και τη διαμερισματοποίηση σε κολικές κυψέλες (haustra), τα οποία απουσιάζουν στο παχύ έντερο του ποντικού. Το ανθρώπινο στομάχι (C) είναι επενδυμένο με έναν αδενικό βλεννογόνο που εκκρίνει γαστρικό οξύ, ενώ ο στόμαχος του ποντικού (B) διαιρείται σε δύο περιοχές - ένα αδενικό στομάχι και ένα μη αδενικό. Ο αδενικός στόμαχος του ποντικού είναι υπεύθυνος για την έκκριση του γαστρικού οξέος, ενώ ο μη αδενικός στόμαχος λειτουργεί ως προσωρινός χώρος αποθήκευσης και πέψης της τροφής. (E) Διατομή ανθρώπινου παχέος εντέρου, που έχει παχύτερο μυϊκό τοίχωμα και βλεννογόνο σε σύγκριση με το παχύ έντερο του ποντικού (D). M: βλεννογόνος, ME: muscularis externa; TC: κολική ταινία [51, 55].

1.4.2 Gnotobiotic και Germ-Free ποντίκια

Ο όρος **germ-free** (GF, ελεύθερος μικροβίων) αναφέρεται σε ένα ζώο που αποδεδειγμένα είναι απαλλαγμένο από μικρόβια, συμπεριλαμβανομένων των βακτηρίων, ιών, μυκήτων, πρωτόζωων και παρασίτων, καθ' όλη τη διάρκεια ζωής του

[48, 49]. Τα GF ζώα που έχουν εποίκιστεί επιλεκτικά με ένα ή περισσότερα βακτηριακά είδη αναφέρονται ως **gnotobiotic** (δηλ. με γνωστό μικροβίωμα) [50].

Ωστόσο, η παραγωγή και συντήρηση αυτών των ποντικών απαιτεί εξειδικευμένες εγκαταστάσεις και το κόστος, το εργατικό δυναμικό και οι δεξιότητες που απαιτούνται για τη διατήρησή τους μπορούν να καταστήσουν αυτά τα μοντέλα απρόσιτα για πολλούς ερευνητές.

1.4.3 Χρησιμότητα GF ποντικών - Συμβολή στην κατανόηση της σχέσης μεταξύ φυσιολογικής μικροχλωρίδας και ξενιστή

Την περασμένη δεκαετία, η ανάπτυξη των τεχνολογιών αλληλούχισης επόμενης γενιάς για τον χαρακτηρισμό του μικροβιώματος οδήγησε σε μια επανάσταση στην κατανόηση των συσχετισμών των μικροβίων με τις ασθένειες. Ωστόσο, αυτές οι μελέτες μπορούν να εντοπίσουν μόνο συσχετίσεις και δεν μπορούν να αποδείξουν αιτιώδη σχέση ή να αποκαλύψουν τους υποκείμενους βιολογικούς μηχανισμούς πίσω από αυτές τις σχέσεις. Τα GF ποντίκια είναι ένα σπουδαίο εργαλείο για την αποκάλυψη της αιτιώδους σχέσης μεταξύ των μικροβίων και των ασθενειών. Εντελώς απαλλαγμένα από όλους τους μικροοργανισμούς, τα GF ποντίκια μπορούν να χρησιμοποιηθούν όπως είναι, για τη μελέτη της πλήρους απουσίας μικροβίων, ή να εποίκιστούν επιλεκτικά με μικρόβια (gnotobiotic ποντίκια) για να ελεγχθούν συγκεκριμένες υποθέσεις. Στη συνέχεια θα αναφέρουμε χαρακτηριστικά παραδείγματα χρήσης GF ποντικών που ανέδειξαν τον ρόλο που παίζουν τα μικρόβια της φυσιολογικής μικροχλωρίδας στη ζωή μας.

• Επίδραση στη φυσιολογία

Μελέτες έχουν χαρακτηρίσει τις διαφορές στη φυσιολογία του ξενιστή μεταξύ GF και φυσιολογικών ποντικών, με την πιο εντυπωσιακή διαφορά να είναι το διευρυμένο τυφλό στα GF ποντίκια. *Ex vivo* παρασκευάσματα μύ από το τυφλό έντερο GF ποντικών παρουσιάζουν μειωμένο μυϊκό τόνο. Ωστόσο, εάν σε αυτά προστεθούν περιεχόμενα του εντέρου από συμβατικά ποντίκια (αυτά που περιέχουν φυσιολογική μικροχλωρίδα), διηθημένα προκειμένου να μην περιέχουν βακτήρια, τότε επαναφέρεται ο μυϊκός τόνος, σχεδόν στις φυσιολογικές τιμές [56].

Τα περιεχόμενα του τυφλού εντέρου είναι επίσης πιο υγρά σε ποντίκια χωρίς μικρόβια, συγκριτικά με τα συμβατικά ποντίκια, εν μέρει εξαιτίας μειωμένων συγκεντρώσεων ιόντων νατρίου και χλωρίου στο τυφλό και μειωμένης απορρόφησης νερού στο κόλον. Ωστόσο, ο ιστός του τυφλού GF ποντικών διατηρεί την ικανότητα να απορροφά νερό παρόμοια με τον συμβατικό ιστό, όπως παρατηρήθηκε σε μελέτες *in vitro* [56]. Αυτές οι

μελέτες δείχνουν ότι ενώ τα ποντίκια χωρίς μικρόβια έχουν τροποποιημένη φυσιολογία του γαστρεντερικού, αυτό είναι μια προσαρμογή στην απουσία μικροβίων και η εισαγωγή μικροβίων στο έντερο μπορεί να αντιστρέψει μερικές από τις αλλαγές.

- **Προστασία από παθογόνους μικροοργανισμούς**

Όπως αναφέραμε στην ενότητα 1.2.4, τα συμβιωτικά βακτήρια μας προστατεύουν από τα παθογόνα. Η αρχική απόδειξη ότι οι συμβιωτικοί μικροοργανισμοί διαδραματίζουν βασικό ρόλο στην πρόληψη της μόλυνσης από μικροοργανισμούς που προκαλούν ασθένειες περιελάμβανε τη χρήση GF ζωικών μοντέλων. Δείχτηκε ότι τα GF ζώα είναι 1000 φορές πιο ευαίσθητα στη μόλυνση και το θάνατο μετά από χορήγηση από το στόμα ενός παθογόνου, σε σύγκριση με τα συμβατικά ζώα. Αυτά τα πειράματα περιελάμβαναν μόλυνση GF ινδικών χοιριδίων με το παθογόνο βακτήριο *Shigella flexneri*, GF ποντικών με το παθογόνο *Vibrio cholerae* ή τον ιό της γρίπης A κ.α.

Περαιτέρω αποδείξεις, ότι η φυσιολογική μικροχλωρίδα προστατεύει τα ζώα από τη μόλυνση, προήλθαν από ένα πείραμα, στο οποίο GF ποντίκια αρχικά έλαβαν το βακτήριο *Bifidobacterium longum*, που συναντάται στην κανονική μικροχλωρίδα, και στη συνέχεια μολύνθηκαν με το παθογόνο *Salmonella typhimurium*. Οι ποντικοί που έλαβαν *Bifidobacterium* επιβίωσαν, ενώ η ομάδα ελέγχου (που υποβλήθηκε σε θεραπεία μόνο με *Salmonella*) πέθανε μέσα σε λίγες ημέρες [57].

- **Επίδραση στη συμπεριφορά**

Το μικροβίωμα επηρεάζει τη συμπεριφορά των ανθρώπων και των ζώων. Τα GF ποντίκια είναι πιο δραστικά και περνούν περισσότερο χρόνο περπατώντας γύρω από τα κλουβιά τους απ' ό,τι τα συμβατικά ποντίκια. Είναι επίσης λιγότερο αγχωμένα και είναι πιο πιθανό να αναλάβουν κινδύνους, όπως η παραμονή για μεγάλες χρονικές περιόδους σε έντονο φως ή ανοιχτούς χώρους. Όμως, η μετάγγιση μικροβίων από τα συμβατικά ποντίκια σε μωρά GF ποντίκια τα έκανε να συμπεριφέρονται με τον «κανονικό» επιφυλακτικό τρόπο. Παρ' όλα αυτά, εάν σε ενήλικα GF ποντίκια γίνει μετάγγιση βακτηρίων της φυσιολογικής μικροχλωρίδας, η συμπεριφορά τους δεν αλλάζει, κάτι που υποδηλώνει ότι το μικροβίωμα επηρεάζει την πρώιμη ανάπτυξη του εγκεφάλου και έτσι επηρεάζεται η συμπεριφορά και στην ενήλικη ζωή [58].

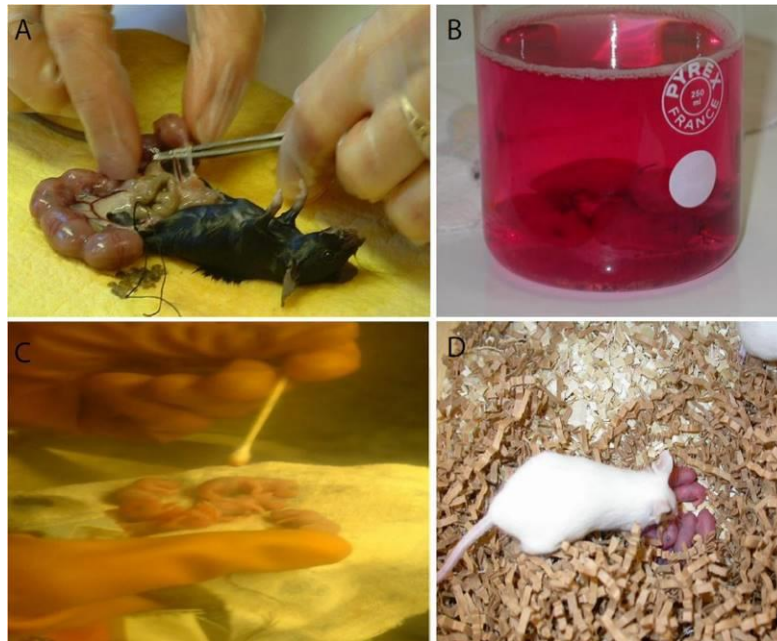
- **Το χαρακτηριστικό παράδειγμα της παχυσαρκίας**

Ένα έξυπνο πείραμα που πραγματοποιήθηκε στο εργαστήριο του Jeffrey Gordon στην Ιατρική Σχολή του Πανεπιστημίου της Ουάσινγκτον, στο Σαιντ Λούις, παρείχε ισχυρή υποστήριξη για τη βακτηριακή υπόθεση (bacterial hypothesis) της παχυσαρκίας [59].

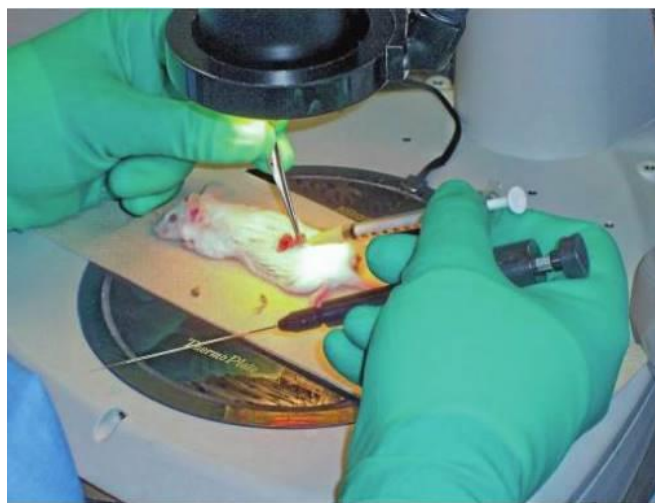
Μία ομάδα GF ποντικών εμβολιάστηκε με βακτήρια από παχύσαρκους ανθρώπους και μια δεύτερη ομάδα με βακτήρια από μη παχύσαρκους. Παρά το γεγονός ότι και οι δύο ομάδες ποντικών έλαβαν ακριβώς την ίδια ποσότητα τροφής, τα ποντίκια που έλαβαν «βακτήρια παχύσαρκων» πήραν πολύ περισσότερο βάρος από τα ποντίκια που έλαβαν «βακτήρια μη παχύσαρκων». Η ομάδα έδειξε επίσης ότι μια μικροβιακή κοινότητα χαρακτηριστική των παχύσαρκων ατόμων μπορεί να «εκτοπίσει» και να υπερισχύσει μιας κοινότητας χαρακτηριστικής των παχύσαρκων ατόμων, εμποδίζοντας τα ποντίκια να πάρουν βάρος, για όσο ακολουθούν υγιεινή διατροφή. Αυτά τα δεδομένα υποδεικνύουν μια σχέση μεταξύ της παχυσαρκίας και των μικροβίων του εντέρου και αφήνουν ανοιχτό το ενδεχόμενο εύρεσης θεραπείας της παχυσαρκίας μέσω χειρισμού της μικροχλωρίδας του εντέρου [59].

1.4.4 Δημιουργία Germ-Free Ποντικών

Ιστορικά, τα πρώτα GF ζώα προήλθαν από καισαρική τομή και ανατράφηκαν αποκλειστικά από τον άνθρωπο (hand raised), απουσία της μητέρας τους και σε ασηπτικές συνθήκες, προκειμένου να μην έρθουν σε επαφή με μικρόβια. Η παραγωγή των πρώτων σταθερών ομάδων/πληθυσμών GF ποντικών και αρουραίων επιτεύχθηκε και περιγράφηκε από τον James Reyniers στο Εργαστήριο Βακτηριολογίας του Πανεπιστημίου Notre Dame [60]. Σήμερα όμως, η προμήθεια GF τρωκτικών επιτυγχάνεται συνήθως από πληθυσμούς που εκτρέφονται απομονωμένοι και διατηρούνται με ενδογαμία, σε ασηπτικές συνθήκες. Εάν, ωστόσο, κάποιο εργαστήριο χρειάζεται να δημιουργήσει ένα καινούργιο GF στέλεχος (strains), πρέπει να καταστεί ελεύθερο μικροβίων ξεκινώντας από οργανισμούς του στελέχους αυτού που δεν είναι ελεύθεροι μικροβίων. Καινούργια στελέχη GF ποντικών μπορούν να δημιουργηθούν με δύο τρόπους: με αφαίρεση των νεογνών μέσω καισαρικής τομής από τη non-GF μητέρα και ανατροφή τους από «ανάδοχη» GF μητέρα (βλ. **Εικόνα 1.3**) ή με εμβρυομεταφορά από την non-GF μητέρα σε GF μητέρα (βλ. **Εικόνα 1.4**) [61]. Ωστόσο, δε συνιστάται η χρήση της πρώτης γενιάς GF ποντικών για πειράματα, καθώς η μητέρα τους δεν ήταν GF και ιοί, βακτήρια και βακτηριακοί μεταβολίτες μπορούν να μεταφερθούν μέσω του πλακούντα από τη μητέρα στο έμβρυο [62].



Εικόνα 1.3 Δημιουργία GF ποντικών με καισαρική τομή. (Α) Αφαίρεση του αμνιακού σάκου αφαιρείται. (Β) Ο αμνιακός σάκος τοποθετείται σε γυάλινο βάζο που περιέχει απολυμαντικό. (Γ) Ο αμνιακός σάκος μεταφέρεται στον προστατευτικό θάλαμο, όπου ανοίγεται και τα νεογνά αφαιρούνται, καθαρίζονται και διεγείρονται προκειμένου να αναπνεύσουν. (Δ) Τα νεογνά παραδίδονται στη θετή GF μητέρα [62].

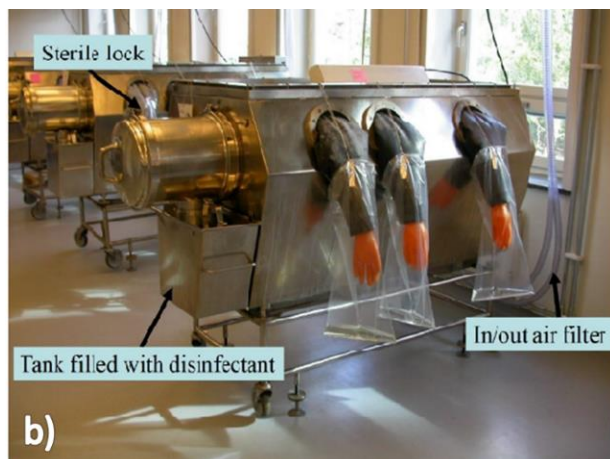
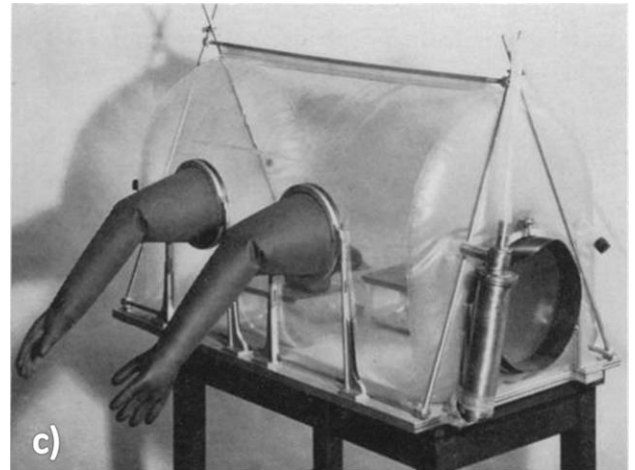
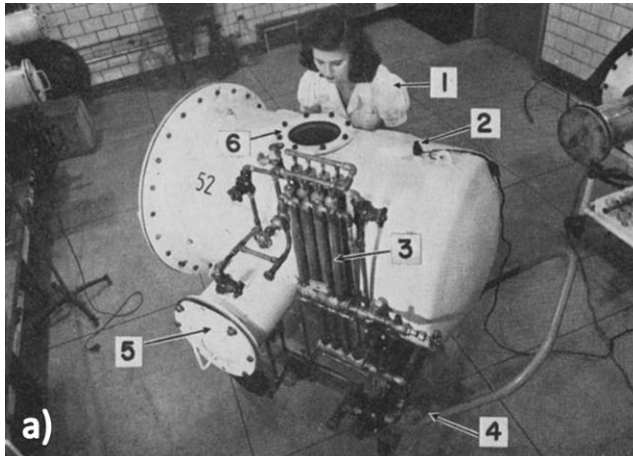


Εικόνα 1.4 Αποστειρωμένη χειρουργική επέμβαση εμβρυομεταφοράς ποντικίου. Όλα τα υλικά, συμπεριλαμβανομένου του μικροσκοπίου και της πιπέτας, πρέπει να είναι αποστειρωμένα [61].

1.4.5 Διατήρηση Germ-Free ποντικών και δυσκολίες

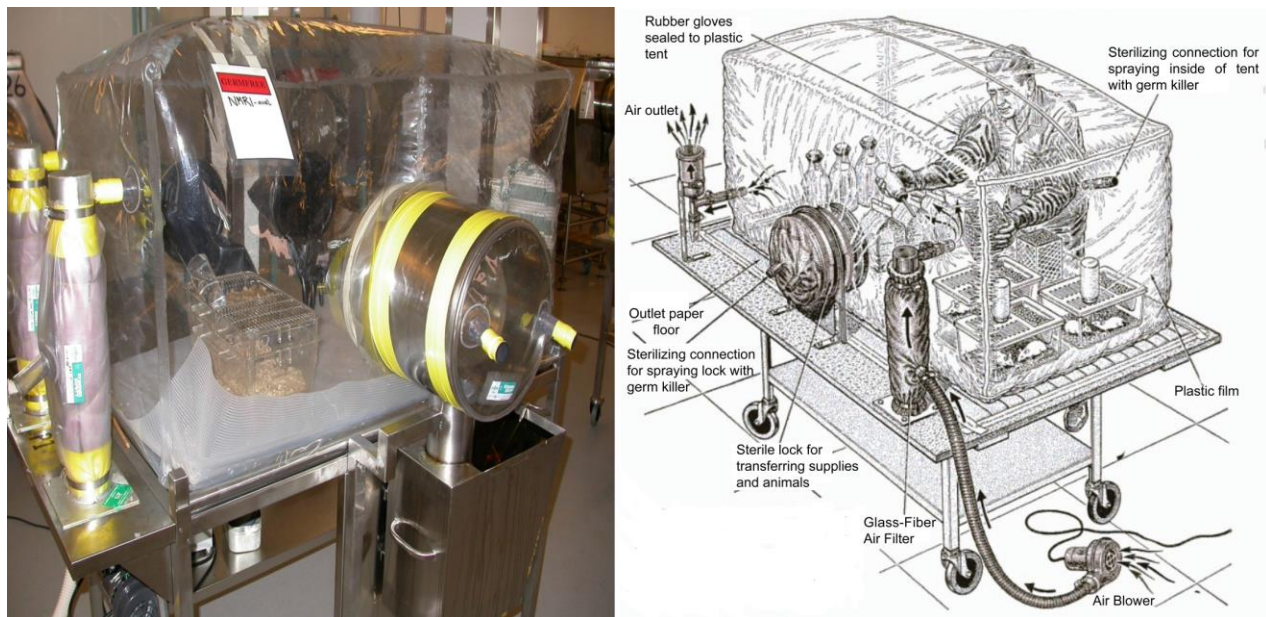
Η διατήρηση αυτών των ζώων απαιτεί να χρησιμοποιούνται στείρες τεχνικές συνεχώς, με περιοδική παρακολούθηση για πιθανή μόλυνση. Η μακροπρόθεσμη διαβίωση GF τρωκτικών είναι δυνατή σε απομονωτές (isolators) θετικής πίεσης. Πριν τα πλαστικά γίνουν ευρέως διαδεδομένα, οι απομονωτές ήταν κατασκευασμένοι από ανοξείδωτο χάλυβα και ήταν μεγάλοι, βαρείς και δύσκολοι στο χειρισμό. Οι πρώτοι απομονωτές ανοξείδωτου χάλυβα αναπτύχθηκαν από τους Reyniers και Trexler και σχεδιάστηκαν

ως θάλαμοι αυτοκλείστου (**Εικόνα 1.5a**). Απομονωτές χάλυβα με θάλαμο εισόδου όπου γινόταν αποστείρωση με ατμό αναπτύχθηκαν επίσης από τον Gustafsson και χρησιμοποιήθηκαν για δεκαετίες (**Εικόνα 1.5b**). Στα τέλη της δεκαετίας του 1950, ο Trexler σχεδίασε τον πρώτο διαφανή και εύκαμπτο πλαστικό απομονωτή από πολυβινυλοχλωρίδιο (PVC) (**Εικόνα 1.5c**) [63].



Εικόνα 1.5 a) Απομονωτής ανοξείδωτου χάλυβα με ατμό αποστείρωσης, Πανεπιστήμιο Notre Dame, 1946. (1. Τεχνικός, 2. Ηλεκτρική σύνδεση, 3. Έξοδος αέρα, 4. Βάση μεταφοράς, 5. Είσοδος / έξοδος αυτοκλείστου). **b)** Gustafsson απομονωτής από χάλυβα [62], **c)** Πλαστικός απομονωτής. Το φίλτρο αέρα και η πόρτα του αποστειρωμένου θαλάμου εισόδου βρίσκονται στο δεξί άκρο του απομονωτή [63].

Σήμερα χρησιμοποιείται κυρίως ο τελευταίος τύπος απομονωτών καθώς και ημιάκαμπτοι απομονωτές. Οι απομονωτές παρέχουν ένα φυσικό εμπόδιο επιτρέποντας τη δημιουργία ενός αποστειρωμένου περιβάλλοντος στο εσωτερικό τους. Οι συσκευές αυτές διαθέτουν παροχή αέρα, φίλτρα εισόδου και εξόδου αέρα, θυρίδα μεταφοράς, γάντια με μανίκια, καθώς και ειδική δεξαμενή γεμάτη με απολυμαντικό που χρησιμοποιείται για τη μεταφορά ποντικών μέσα και έξω (**Εικόνα 1.6**) [62].



Εικόνα 1.6 Ένας σύγχρονος απομονωτής [62].

Η ακρίβεια και η αναπαραγωγιμότητα των πειραματικών αποτελεσμάτων, όταν χρησιμοποιούνται GF ή gnotobiotic ποντίκια, εξαρτώνται από τη διατήρηση ακριβών μικροβιακών συνθηκών εντός κάθε απομονωτή, με αποκλεισμό ανεπιθύμητων εξωτερικών μικροοργανισμών. Αυτό επιτυγχάνεται με θύρες εισόδου διπλής όψεως, γάντια με μανίκια, συνεχή ροή φιλτραρισμένου αέρα και πλήρη αποστείρωση όλων των υλικών που εισέρχονται στον απομονωτή.

Ωστόσο, αυτό το σύστημα είναι απαιτητικό, από θέμα εργατωρών, και ευάλωτο σε μολύνσεις. Οποιαδήποτε διακοπή της ροής του αέρα, ελλιπής αποστείρωση τροφής, νερού ή κλινοστρωμάτων (εξαιτίας, παραδείγματος χάριν, μιας μη ανιχνευόμενης δυσλειτουργίας του αυτοκλείστου) ή παραβίαση της ακεραιότητας των τοιχωμάτων απομόνωσης, των θυρών εισόδου, των φίλτρων αέρα και των γαντιών που χρησιμοποιούνται για το χειρισμό των ζώων θα θέσουν σε κίνδυνο την αποστείρωση του απομονωτή [64].

1.4.6 Έλεγχος πιθανών μολύνσεων GF ποντικών στο εργαστήριο

Ένα σημαντικό κομμάτι της εργασίας με GF ποντίκια είναι η διασφάλιση ότι τα ποντίκια είναι όντως απαλλαγμένα από μικρόβια. Ο γρηγορότερος τρόπος ελέγχου είναι η καλλιέργεια δειγμάτων κοπράνων υπό αερόβιες και αναερόβιες συνθήκες, ή εναλλακτικά, η κατά Gram χρώση επιχρισμάτων από κόπρανα πάνω σε αντικειμενοφόρες πλάκες και παρατήρηση σε φωτονικό μικροσκόπιο. Άλλες πιθανές θέσεις δειγματοληψίας για την ανίχνευση βακτηριακής μόλυνσης είναι οι το τρίχωμα των ποντικών, τα μπουκάλια νερού, τα κλινοσκεπάσματα και επιφάνειες του απομονωτή.

Ωστόσο, και οι δύο τεχνικές έχουν περιορισμούς: δεν είναι εφικτό να καλλιεργηθούν τα περισσότερα βακτήρια έξω από το έντερο και οι περιορισμένου βαθμού μολύνσεις μπορεί να μην παρατηρηθούν σε επιχρίσματα. Έτσι, οι μοριακές τεχνικές, όπως η ενίσχυση με PCR (polymerase chain reaction amplification) του 16S rRNA γονιδίου ή άλλων συγκεκριμένων βακτηριακών γονιδίων, είναι μια καλή εναλλακτική λύση. Συνεπώς, συνιστάται να χρησιμοποιείται ένας συνδυασμός μεθόδων, για την εκτίμηση της απουσίας μικροβίων, και η καλλιέργεια κοπράνων να πραγματοποιείται εβδομαδιαίως/ανά δύο βδομάδες και η PCR τουλάχιστον μία φορά κάθε μήνα [65]. Σε ορισμένα εργαστήρια, η ανάλυση PCR παραλείπεται, καθώς υποστηρίζεται ότι μολύνσεις από βακτήρια που ανιχνεύονται από την PCR μπορεί να ανιχνευτούν και με καλλιέργεια ή χρώση κατά Gram [61].

Για την ανίχνευση ιών και κάποιων άλλων παθογόνων μικροοργανισμών απαιτείται αιματολογικός έλεγχος. Για το σκοπό αυτό, αφαιρείται και υποβάλλεται σε ευθανασία ένα ποντίκι από κάθε απομονωτή και ο ορός του ελέγχεται για ενδείξεις ιών (ιός της ηπατίτιδας των ποντικών, αδενοϊός ποντικών, παρβοϊός/ραϊνονίγους ποντικών, minute ιός, ιός ενδημικής διάρροιας βρεφών (EDIM), ιός ectromelia, ιός Sendai, ιός λεμφοκυτταρικής χοριομηνιγγίτιδας, ιός polyoma), του μύκητα pneumocystis, του βακτηρίου CAR (cilia-associated respiratory bacillus) και άλλων μικροοργανισμών αναλόγως το εργαστήριο και το πρωτόκολλο που ακολουθείται. Τα παραπάνω συνήθως πραγματοποιούνται ανά τρίμηνο [61].

1.5 Μεταγονιδιωματική

Η γονιδιωματική (Genomics) είναι τομέας της γενετικής και της βιοπληροφορικής, που εφαρμόζει διάφορες τεχνολογίες (τεχνολογία ανασυνδυασμένου DNA, μέθοδοι αλληλούχισης DNA κ.α.) για την αλληλούχιση, την οργάνωση, την ανάλυση της δομής και της λειτουργίας γονιδιωμάτων.

Μεταγονιδιωματική (Metagenomics) είναι η μελέτη μεταγονιδιωμάτων, δηλαδή γενετικού υλικού που προκύπτει από δείγματα που βρίσκονται στο περιβάλλον και άρα που η ακριβής του προέλευση δεν είναι γνωστή. Χρησιμοποιείται για την ανάλυση της γονιδιακής σύστασης μίας μικροβιακής κοινότητας και την ταυτοποίηση αυτής, με χρήση πειραματικών και υπολογιστικών μεθόδων, οι οποίες επιτρέπουν την αλληλούχιση του γονιδιώματος των μικροοργανισμών, χωρίς την ανάγκη καλλιέργειας τους [66].

Στη μέθοδο της μαζικής (shotgun) παράλληλης αλληλούχισης πραγματοποιείται τυχαία τμηματοποίηση του DNA, αλληλούχιση των DNA τμημάτων που προκύπτουν με τη

μέθοδο Sanger (παλαιότερα) ή με αλληλούχιση επόμενης γενιάς (next-generation sequencing). Στην τελευταία περίπτωση, παράγονται μικρότερες αλληλουχίες (reads) (μήκους 25-500bp) αλλά πολλές εκατοντάδες χιλιάδες ή εκατομμύρια σε αριθμό, σε σχετικά σύντομο χρονικό διάστημα. Βιοπληροφορικά προγράμματα, στη συνέχεια, «συναρμολογούν» (assembly) τις αλληλουχίες που αλληλεπικαλύπτονται διαβάζοντας τελικά μια συνεχόμενη αλληλουχία που αντιπροσωπεύει ένα μεγάλο τμήμα του γονιδιώματος.

Παραλείποντας το βήμα της συναρμολόγησης, έχουν δημιουργηθεί βιοπληροφορικά προγράμματα που διαβάζουν εκατομμύρια DNA αλληλουχίες (reads) και προσδιορίζουν τον οργανισμό από τον οποίο αυτές προέρχονται χρησιμοποιώντας, για παράδειγμα, k-mer ταξινομητές. Προϋπόθεση είναι το γονιδίωμα του οργανισμού, ή μέρος αυτού, να είναι ήδη γνωστό.

Όσον αφορά τη μελέτη της έκφρασης των γονιδίων στα κύτταρα, η αλληλούχιση επόμενης γενιάς έχει φέρει επανάσταση. Συγκεκριμένα, στην τεχνική «total RNA-Seq» το συνολικό συμπλήρωμα των RNA μορίων, από ένα δεδομένο δείγμα, απομονώνεται και προσδιορίζεται, ως προς την αλληλουχία, χρησιμοποιώντας τεχνολογίες υψηλής απόδοσης (high-throughput technologies). Έτσι, ταυτοποιούνται και ποσοτικοποιούνται τα RNA μόρια στο δείγμα στη δεδομένη χρονική στιγμή.

Με αυτόν τον τρόπο θα μπορούσαν να χρησιμοποιηθούν RNA-Seq δεδομένα γονιδιακής έκφρασης για τον προσδιορισμό και μικροβίων στο δείγμα.

1.6 Σκοπός της εργασίας

Σκοπός της παρούσας εργασίας ήταν η διερεύνηση της μικροβιακής κατάστασης GF και gnotobiotic ποντικών χρησιμοποιώντας RNA-Seq δεδομένα. Με τον τρόπο αυτό εξετάζεται αν τα RNA-Seq δεδομένα γονιδιακής έκφρασης μπορούν να χρησιμοποιηθούν και για την διαπίστωση κάποιας μόλυνσης των GF ή gnotobiotic ποντικών από μη επιθυμητούς μικροοργανισμούς. Τα RNA-Seq δεδομένα που χρησιμοποιήσαμε προέρχονται από τη βάση δεδομένων NCBI GEO [67] και είναι ελεύθερα διαθέσιμα.

Χρησιμοποιήθηκαν τρία εργαλεία, KRAKEN 2 [68], CLARK [69] και AGAMEMNON, ειδικά στον χαρακτηρισμό της μικροβιακής κοινότητας δειγμάτων, τα οποία παίρνουν ως είσοδο DNA-seq ή RNA-seq δεδομένα και υπολογίζουν το βαθμό αντιπροσώπευσης (αριθμός αλληλουχιών - reads) κάθε μικροβιακού taxon στο δείγμα. Η λειτουργία των τριών εργαλείων περιγράφεται στην επόμενη ενότητα.

Σε δεύτερο στάδιο, για τις ανάγκες του εργαστηρίου και ανεξάρτητα από το κύριο θέμα της παρούσας διπλωματικής εργασίας, συλλέχθηκαν, από 101 επιστημονικές δημοσιεύσεις, 995 συσχετίσεις μικροβίων με ανθρώπινες ασθένειες, οι οποίες και παρατίθενται στο Παράρτημα Ι. Αυτή η συλλογή θα φανεί χρήσιμη στη δημιουργία αντίστοιχης βάσης δεδομένων.

2. ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ

2.1 Αναζήτηση RNA-Seq Δεδομένων

Τα RNA-Seq datasets που χρησιμοποιήσαμε προέρχονται από τη βάση NCBI GEO [67]. Έγιναν δύο αναζητήσεις, μία με τον όρο «germ-free» και μία με τον όρο «gnotobiotic». Τα αποτελέσματα από κάθε μία αναζήτηση φιλτραρίστηκαν προκειμένου να περιλαμβάνουν μόνο τις εγγραφές που αφορούσαν το ποντίκι («*Mus musculus*» στο πεδίο «Organism») και την ποσοτικοποίηση της γονιδιακής έκφρασης σε αυτό («Expression profiling with high throughput sequencing» στο πεδίο «Study type»). Οι τελικές εγγραφές επομένως ήταν 91 και 121 αντίστοιχα (τελευταία ενημέρωση, 29/10/2019).

2.2 Επιλογή Δειγμάτων

Η εγκυρότητα κάθε εγγραφής – μελέτης ελέγχθηκε μη αυτόματα, διαβάζοντας τη δημοσίευση που τη συνόδευε. Επιλέχτηκαν τελικά για ανάλυση 8 μελέτες με 97 δείγματα.

2.3 Κατέβασμα των αρχείων

Για κάθε δείγμα κατεβάσαμε το αντίστοιχο sra αρχείο, που περιέχει όλες τις αλληλουχίες (reads) του δείγματος, και το μετατρέψαμε σε fastq, χρησιμοποιώντας τις μεθόδους esearch και efetch του NCBI Entrez-utilities [70], srapath και fasterq-dump του sra-toolkit [71] και την εντολή wget.

2.4 Έλεγχος ποιότητας και προ-επεξεργασία αλληλουχιών

Η εκτίμηση της ποιότητας των αλληλουχιών έγινε με το εργαλείο FastQC [72]. Στην περίπτωση παρουσίας «προσαρμογέων» (adapters) στις αλληλουχίες, χρησιμοποιήσαμε, για την αφαίρεσή του, το εργαλείο cutadapt [73]. Το ίδιο εργαλείο χρησιμοποιήθηκε και για την περικοπή των αλληλουχιών με score ποιότητας μικρότερο του 10 (στο 3' άκρο της αλληλουχίας) και το φιλτράρισμα αλληλουχιών μήκους μικρότερου από 80 βάσεις. Πραγματοποιήσαμε περαιτέρω επεξεργασία των αλληλουχιών, όποτε χρειαζόταν, σύμφωνα με τις οδηγίες που δίνονταν στην αντίστοιχη εγγραφή στη βάση GEO [67].

2.5 Στοιχίση αλληλουχιών/reads στο γονιδίωμα του ποντικού και εξαγωγή των χρήσιμων αλληλουχιών

Στη συνέχεια, στοιχίσαμε τις επεξεργασμένες αλληλουχίες, κάθε fastq αρχείου, στο γονιδίωμα του ποντικού (*Mus musculus*), προκειμένου όσες στοιχηθούν (και άρα μπορούν να χαρακτηριστούν ως αλληλουχίες του ποντικού) να αφαιρεθούν από το αρχείο. Επιλέχθηκε το γονιδίωμα έναντι του γονιδιώματος του ποντικού, καθότι παρατηρήσαμε μεγαλύτερο ποσοστό στοιχίσης των αλληλουχιών σε αυτό. Οι αλληλουχίες που απέτυχαν να στοιχηθούν στο γονιδίωμα του ποντικού μετά από δύο συνεχόμενες προσπάθειες, συλλέχθηκαν προκειμένου να διερευνηθεί αν προέρχονται από μικροοργανισμούς.

Η στοιχίση έγινε με το πρόγραμμα Hisat2 (v. 2.1.0) [74], όπου χρησιμοποιήθηκε το γονιδίωμα αναφοράς (reference genome) GRCm38 του ποντικού (*Mus musculus*) με το χαρακτηρισμό των γονιδίων της Ensembl [75], προκειμένου να γίνει η δεικτοδότηση (indexing). Η στοιχίση έγινε με τις προκαθορισμένες (default) παραμέτρους. Το αρχείο εξόδου του Hisat2 είναι ένα αρχείο SAM.

2.6 Χρήση εργαλείων μεταγονιδιωματικής

Τα τρία εργαλεία μεταγονιδιωματικής που χρησιμοποιήσαμε ήταν τα Kraken 2 (v. 2.0.7-beta) [68], CLARK (v. 1.2.6) [69] και AGAMEMNON, εκ των οποίων το τελευταίο έχει αναπτυχθεί από το εργαστήριό μας (DIANA-Lab).

Τα Kraken 2 και CLARK αρχικά δέχονται ένα σύνολο γονιδιωμάτων μικροοργανισμών προκειμένου να κατασκευάσουν μία λίστα με τα k-μερή, όπου το καθένα συναντάται μόνο σε μία ταξινομική ομάδα μικροοργανισμών. Κάθε k-μερές αποτελείται από τα γράμματα A, T, C, G που αντιστοιχούν στις 4 αζωτούχες βάσεις του DNA και το k αντιστοιχεί στο μήκος του που επιλέγεται από το χρήστη. Στη συνέχεια τα k-μερή της κάθε αλληλουχίας του δείγματος αναζητούνται στη λίστα αυτή, προκειμένου τελικά να χαρακτηριστεί η ταξινομική ομάδα μικροοργανισμών από την οποία προέρχεται αυτή η αλληλουχία. Το πλεονέκτημα των μεθόδων αυτών είναι ότι δεν στοιχίζεται ολόκληρη η αλληλουχία, αλλά τμηματικά τα k-μερή της, προς μία λίστα k-μερών που είναι κατά πολύ μικρότερη σε μήκος από το συνολικό μήκος των γνωστών γονιδιωμάτων, και μάλιστα λεξικογραφικά οργανωμένα.

Το πρόγραμμα AGAMEMNON χρησιμοποιεί τα γνωστά γονιδιώματα μικροοργανισμών που δίνει ο χρήστης προκειμένου να κατασκευάσει μία δομή (Suffix Array) η οποία περιέχει όλες τις δυνατές αλληλουχίες που υπάρχουν στα γονιδιώματα λεξικογραφικά

οργανωμένες. Έτσι, όταν επιχειρείται να στοιχηθεί μία αλληλουχία του δείγματος, το πρώτο k-μερές της εντοπίζεται σε μια περιοχή της δομής, με μία συνάρτηση/πίνακα κατακερματισμού, και επεκτείνεται κατάλληλα. Με τη μεθοδολογία αυτή, που ονομάζεται «quasi-mapping», κατά την επέκταση της στοίχισης είναι δυνατή η παράλειψη 40 – 60% των k-μερών. Στη συνέχεια ο AGAMEMNON προχωράει σε ποσοτικοποίηση των μικροβίων μέχρι και την ταξινομική βαθμίδα του υποείδους.

2.6.1 KRAKEN2

Για την κατασκευή της βάσης/index, με την εντολή build, χρησιμοποιήσαμε:

- Τα γονιδιώματα των βακτηρίων, αρχαίων, μυκήτων και ιών που χαρακτηρίζονται ως «Complete Genome», από το αποθετήριο RefSeq του NCBI και συγκεκριμένα από τον ιστότοπο <ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/>, όπως προτείνεται από τους δημιουργούς του Kraken 2.
- Το γονιδίωμα αναφοράς του ανθρώπου: Human reference genome GRCh38.p12.
- Το γονιδίωμα αναφοράς του ποντικιού: Mouse reference genome: GRCm38.p4

Επιλέχτηκε μήκος k-μερών: $k=31$. Επίσης, θέσαμε μήκος l -μερών: $l=27$. Το πρόγραμμα kraken 2 με την εντολή build και αυτές τις δύο παραμέτρους, για κάθε δυνατό 31-μερές, αποτελούμενο από τις 4 βάσεις A,T,C και G (4^{31} διαφορετικοί συνδυασμοί), βρίσκει ποιο είναι το λεξικογραφικά (ή αλφαβητικά) μικρότερο 27-μερές του, και στη συνέχεια βρίσκει ποια είναι η μικρότερη ταξινομική ομάδα, στα γονιδιώματα της οποίας υπάρχει αυτό το 27-μερές. Η αντιστοίχιση κάθε l -μερούς με μια ταξινομική ομάδα αποθηκεύεται σε έναν πίνακα/index.

Όταν, στη συνέχεια, δίνουμε στο πρόγραμμα το fastq αρχείο με τις αγνώστου προέλευσης αλληλουχίες, για κάθε αλληλουχία, έστω μήκους 100 βάσεων, βρίσκει τα $100-30=70$ 31-μερή της, για κάθε ένα βρίσκει το λεξικογραφικά μικρότερο 27-μερές του και ανατρέχει στον πίνακα/index που δημιουργήθηκε στην αρχή και σημειώνει τις 70 ταξινομικές ομάδες (όχι απαραίτητα διαφορετικές) από τις οποίες «προέρχονται» τα 70 27-μερή. Τέλος, καταλήγει να απαντήσει από ποια ταξινομική ομάδα «προέρχεται» η αλληλουχία συγκρίνοντας τις 70 ταξινομικές ομάδες.

2.6.2 CLARK

Για την κατασκευή του λεξικού/index, χρησιμοποιήσαμε:

- Τα γονιδιώματα των βακτηρίων, αρχαίων, μυκήτων και ιών που χαρακτηρίζονται ως «Complete Genome» ή Chromosome, από το αποθετήριο RefSeq του NCBI και συγκεκριμένα από τον ιστότοπο <ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/>, όπως προτείνεται από τους δημιουργούς του CLARK.
- Το γονιδίωμα αναφοράς του ανθρώπου: Human reference genome GRCh38.p12.
- Το γονιδίωμα αναφοράς του ποντικίου: Mouse reference genome: GRCm38.p4

Επιλέχτηκε μήκος k-μερών: k=31 και ταξινομική βαθμίδα «Species». Με αυτές τις παραμέτρους, το πρόγραμμα CLARK δημιουργεί μία λίστα/index με εκείνα τα 31-μερή που εντοπίζονται μόνο σε γονιδιώματα ενός είδους (Species) και αποθηκεύει για κάθε 31-μερές το είδος αυτό.

Όταν, στη συνέχεια, δίνουμε στο πρόγραμμα το fastq αρχείο με τις αγνώστου προέλευσης αλληλουχίες, αναζητά σε αυτές αν περιέχουν 31-μερή που βρίσκονται στο λεξικό που δημιούργησε στο πρώτο βήμα και σημειώνει τα Species από τα οποία προέρχονται. Τέλος, καταλήγει σε ποιο Species πρέπει να αποδώσει την κάθε αλληλουχία, κοιτώντας από ποιο Species προέρχονται τα περισσότερα k-μερή.

Να σημειώσουμε ότι και στο Kraken 2 και στο Clark εκτός από τις μικροβιακές αλληλουχίες, χρησιμοποιούμε και το γονιδίωμα του ποντικού και του ανθρώπου για την κατασκευή του index. Επιδιώκουμε έτσι, αφού δώσουμε στα προγράμματα τις αγνώστου προέλευσης αλληλουχίες να αφαιρέσουμε από την μεταγονιδιωματική μελέτη μας εκείνες που με αυτή τη μεθοδολογία (k-mer ταξινόμηση) μπορούν να αποδοθούν στο ποντίκι ή τον άνθρωπο.

2.6.3 AGAMEMNON

Για την κατασκευή της βάσης/index, χρησιμοποιήσαμε τα γονιδιώματα των βακτηρίων, των αρχαίων, και των ιών που χαρακτηρίζονται ως «Complete Genome» από το αποθετήριο RefSeq του NCBI και συγκεκριμένα από τον ιστότοπο <ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/>. Δεν προσθέσαμε τα γονιδιώματα των μυκήτων, του ανθρώπου και του ποντικού καθώς δεν είναι εκπαιδευμένο το πρόγραμμα για στοίχιση (quasi-mapping) RNA-αλληλουχιών σε γονιδιώματα ευκαρυωτικών οργανισμών. Επιλέχτηκε μήκος k-μερών: k=31

2.7 Στοιχίση των αλληλουχιών σε γονιδιώματα επιλεγμένων μικροβίων και ταυτοποίηση των γονιδίων που εκφράζονται

Μετά τη μεταγονιδιωματική ανάλυση, για κάθε σειρά δειγμάτων παρατηρήσαμε ποιοι μικροοργανισμοί πιθανώς εντοπίζονται σε αυτά και κατεβάσαμε από τη βάση NCBI RefSeq (<ftp.ncbi.nlm.nih.gov/genomes/>) το γονιδιώμά τους. Με το πρόγραμμα Bowtie 2 (v. 2.3.5.1) [76] στοιχίσαμε σε αυτό τις αλληλουχίες των δειγμάτων, χρησιμοποιώντας τις προκαθορισμένες (default) τιμές για τις παραμέτρους.

Εκτός από το γονιδίωμα των μικροοργανισμών κατεβάσαμε και πληροφορίες για τη θέση και τη λειτουργία των γονιδίων στο γονιδίωμα (gene annotation) σε τύπο αρχείου GFF3 (Generic Feature Format Version 3). Συγκρίνοντας τις περιοχές του γονιδιώματος στις οποίες στοιχίζονται οι αλληλουχίες των δειγμάτων μας, με τις θέσεις των γονιδίων στο γονιδίωμα επιχειρήθηκε να εξάγουμε κάποιο συμπέρασμα για το ποια γονίδια πιθανώς εκφράζονται. Αυτή η σύγκριση (εύρεση της τομής) έγινε με τη μέθοδο intersect των Bedtools [77]. Κριτήριο για να θεωρήσουμε ότι μια αλληλουχία/read προέρχεται από ένα γονίδιο είναι τουλάχιστον το 10% του μήκους του read να βρίσκεται στην περιοχή του γονιδίου.

Επειδή αναμένουμε συνέκφραση γονιδίων που ανήκουν στο ίδιο οπερόνιο, επιχειρήσαμε να εντοπίσουμε τέτοιες περιπτώσεις συνέκφρασης, με τη βοήθεια του προγράμματος οπτικοποίησης στοιχίσεων IGV. Εδώ, φάνηκαν χρήσιμες βάσεις δεδομένων όπως η Ecocyc [78] και η RegulonDB [79], οι οποίες παρέχουν αναλυτικό χαρακτηρισμό των οπερονίων της *Escherichia coli*.

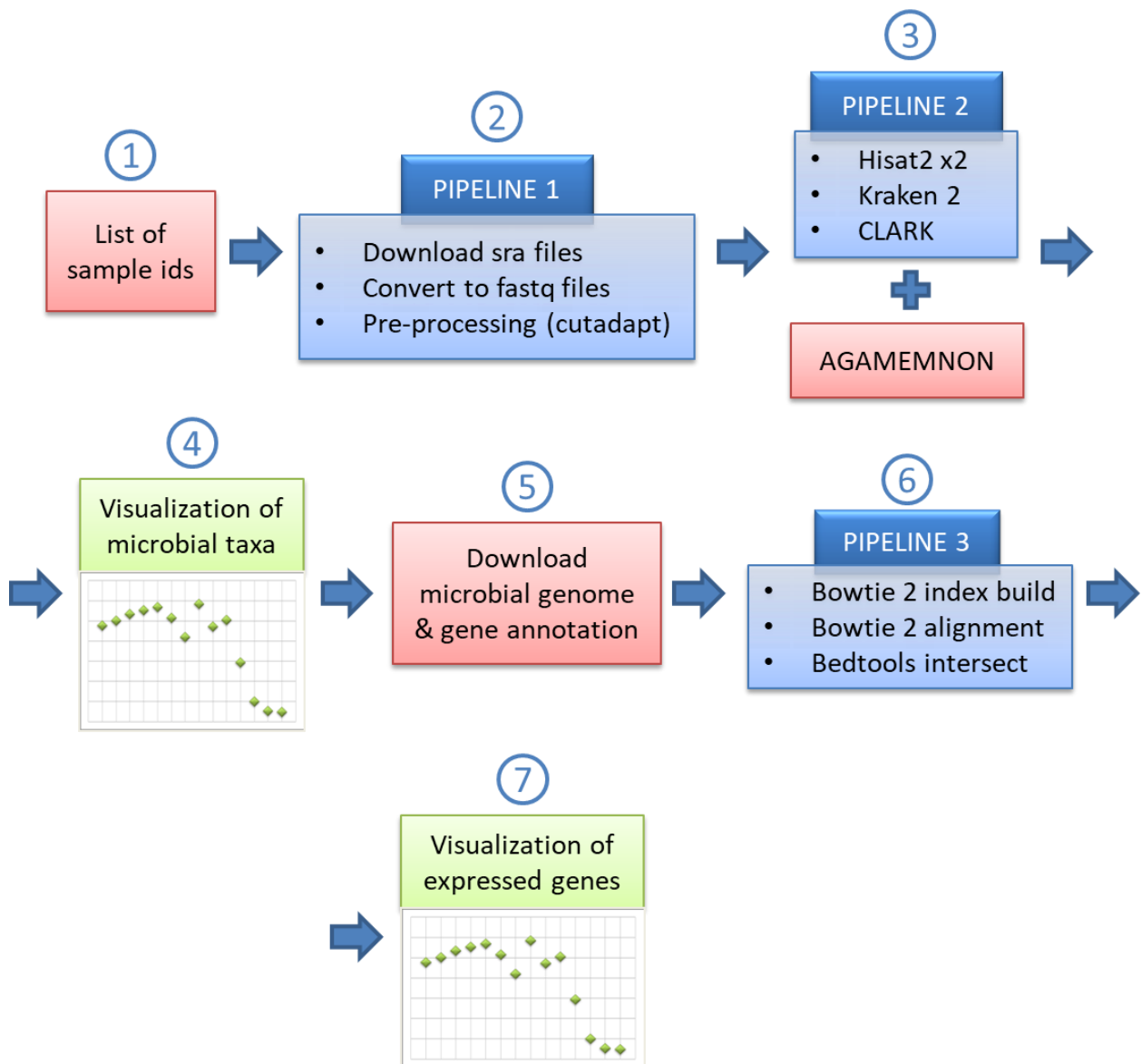
2.8 Ποή Εργασίας (Workflow - Pipelines)

Τα παραπάνω οργανώθηκαν σε τρία pipelines (σωλήνωση λογισμικού) και κάποια ενδιάμεσα βήματα, ως εξής (βλ. **Εικόνα 2.1**):

- Το 1^ο pipeline δέχεται ως είσοδο ένα text αρχείο με την λίστα των GSM ids των δειγμάτων που θέλουμε να μελετήσουμε και την παράμετρο --trim ή --no-trim αν από τις αλληλουχίες θέλουμε να αφαιρέσουμε τον Illumina universal adapter ή όχι, αντίστοιχα. Ολοκληρώνοντας τη λειτουργία του, έχει κατεβάσει τις αλληλουχίες των δειγμάτων από τη βάση SRA [71] σε τελική μορφή αρχείων fastq και έχει πραγματοποιήσει προ-επεξεργασία των αλληλουχιών.
- Το 2^ο pipeline δέχεται ως είσοδο το ίδιο text αρχείο και χρησιμοποιεί τα προ-επεξεργασμένα fastq αρχεία που παρήγαγε το 1^ο pipeline για να πραγματοποιήσει

στοίχιση των αλληλουχιών στο γονιδίωμα του ποντικού, δύο διαδοχικές φορές. Στη συνέχεια, οι αλληλουχίες που δε στοιχίζονται χρησιμοποιούνται για μεταγονιδιωματική ανάλυση από τα προγράμματα Kraken 2 και CLARK, οπότε στην έξοδό του το pipeline έχει παράγει για κάθε δείγμα ένα αρχείο με την αφθονία των μικροοργανισμών σε αυτό.

- Ενδιάμεσα βήματα I: Αφού αναλύσουμε τις ίδιες αλληλουχίες και με το πρόγραμμα AGAMEMNON, οπτικοποιούμε τα αποτελέσματα των τριών προγραμμάτων σε ιστογράμματα αφθονίας των μικροοργανισμών, χρησιμοποιώντας τρία αρχεία εντολών, γραμμένα σε γλώσσα R.
- Ενδιάμεσα βήματα II: Επιλέγουμε κάποιον μικροοργανισμό, με σκοπό να μελετήσουμε ποια γονίδια του εκφράζονται και κατεβάζουμε από τη βάση NCBI RefSeq (<ftp.ncbi.nlm.nih.gov/genomes/>) το γονιδίωμά του, μαζί με το αντίστοιχο αρχείο GFF, που περιέχει πληροφορίες για τη θέση και τη λειτουργία των γονιδίων.
- Το 3^ο pipeline δέχεται ως είσοδο το μονοπάτι (file path) στο οποίο βρίσκονται τα δύο παραπάνω αρχεία και το μονοπάτι όπου βρίσκονται τα δείγματά μας. Στη συνέχεια: i) κατασκευάζει το index του γονιδιώματος του μικροοργανισμού, ii) στοιχίζει τις αλληλουχίες των δειγμάτων (εκείνες που δε στοιχίζονται στο γονιδίωμα του ποντικού) στο index που κατασκεύασε, iii) εντοπίζει με το εργαλείο bedtools intersect πόσες αλληλουχίες στοιχίζονται σε περιοχές γονιδίων και ποια είναι αυτά. Κατά την έξοδό του, το pipeline αυτό έχει δημιουργήσει για κάθε δείγμα ένα αρχείο/πίνακα με τα γονίδια στα οποία αντιστοιχεί κάθε αλληλουχία/read.
- Τελικό βήμα: Επεξεργαζόμαστε τα αρχεία/πίνακες με ένα αρχείο εντολών, γραμμένο σε γλώσσα R, για να κατασκευάσουμε διαγράμματα με τα πιο συχνά γονίδια όλης της σειράς δειγμάτων. Με το πρόγραμμα IGV μπορούμε να πραγματοποιήσουμε οπτικοποίηση στοιχίσεων σε επιλεγμένα δείγματα και γονίδια.



Εικόνα 2.1 Ροή Εργασίας (workflow) που περιγράφει τα βήματα που ακολουθούμε για κάθε σειρά δειγμάτων.

3. ΑΠΟΤΕΛΕΣΜΑΤΑ

3.1 Σειρά δειγμάτων GSE102543

3.1.1 Περιγραφή δειγμάτων, προ-επεξεργασία και στοίχιση στο γονιδίωμα του ΠΟΝΤΙΚΟΥ

- **Δημοσίευση:** Uchimura Y *et al.*, "Antibodies Set Boundaries Limiting Microbial Metabolite Penetration and the Resultant Mammalian Host Response", *Immunity*, 2018 Sep 18
- **Ιστός προέλευσης:** λεπτό έντερο
- **Σχεδιασμός:** paired-end
- **Μόριο:** "total RNA", ολικό RNA
- **Δείγματα:** 6 δείγματα προερχόμενα από 6 gnotobiotic ποντίκια εποικισμένα μόνο με *Escherichia coli* (monocolonized, MC).

Πίνακας 3.1 Χαρακτηριστικά δειγμάτων της σειράς με id GSE102543

όνομα δείγματος	χαρακτηριστικά δείγματος
GSM2740422: Conditioned_1	1 ILLUMINA (Illumina HiSeq 2500) run: 69M spots , 20.3G bases, 7.2Gb downloads. Accession: SRX3088445
GSM2740423: Conditioned_2	1 ILLUMINA (Illumina HiSeq 2500) run: 66.1M spots , 19.9G bases, 7.1Gb downloads. Accession: SRX3088446
GSM2740424: Conditioned_3	1 ILLUMINA (Illumina HiSeq 2500) run: 58.1M spots , 17.5G bases, 6.1Gb downloads. Accession: SRX3088447
GSM2740425: Unconditioned_1	1 ILLUMINA (Illumina HiSeq 2500) run: 62.6M spots , 18.8G bases, 6.7Gb downloads. Accession: SRX3088448
GSM2740426: Unconditioned_2	1 ILLUMINA (Illumina HiSeq 2500) run: 64.6M spots , 19.4G bases, 7Gb downloads. Accession: SRX3088449
GSM2740427: Unconditioned_3	1 ILLUMINA (Illumina HiSeq 2500) run: 60.7M spots , 18.2G bases, 6.5Gb downloads. Accession: SRX3088450

Προ-επεξεργασία αλληλουχιών και στοίχιση

- Αρχικό μήκος αλληλουχιών: 151 βάσεις
- Δεν εντοπίστηκε αντάπτορας στις αλληλουχίες.
- Ύστερα από την περικοπή των αλληλουχιών με quality score <10 στο 3' άκρο τους, κρατήσαμε τις αλληλουχίες όπου και τα δύο mates είχαν μήκος ≥ 80 βάσεις.

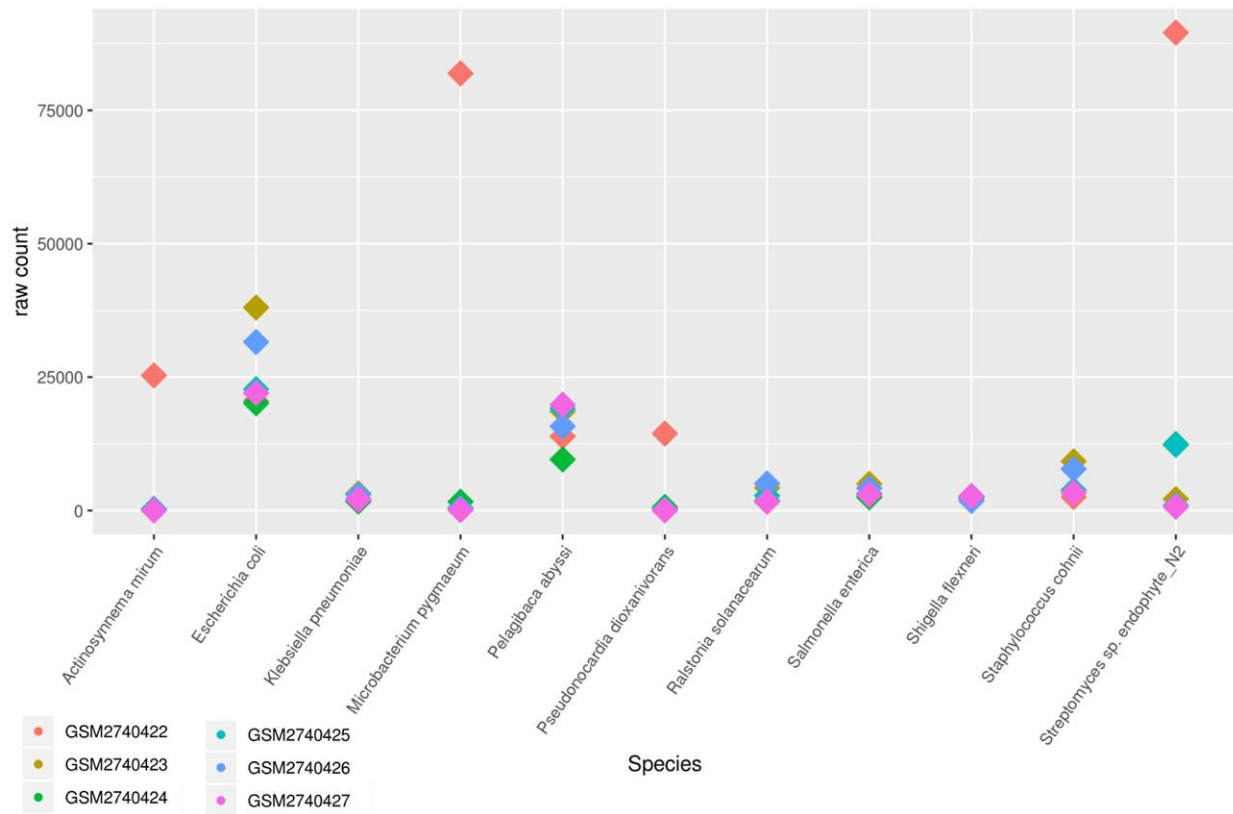
- Οι αλληλουχίες που δεν στοιχήθηκαν στο γονιδίωμα του ποντικού με το πρόγραμμα Hisat2, μετά από δύο προσπάθειες, συλλέχθηκαν (βλ. **Πίνακα 3.2**).

Πίνακας 3.2 Ποσοστό αλληλουχιών που δεν στοιχίζονται στο γονιδίωμα του ποντικού (unmapped pairs%) στα έξι δείγματα. Επίσης, παρουσιάζεται ο NCBI GEO κωδικός του κάθε δείγματος (sample id), ο SRA κωδικός του (SRA id), ο χαρακτηρισμός του (tag , MC=monocolonized), ο αριθμός των αλληλουχιών μετά την προ-επεξεργασία/φιλτράρισμα (filtered sra pairs) και ο αριθμός των αλληλουχιών που δεν στοιχίζονται στο γονιδίωμα του ποντικού (unmapped pairs).

sample id	SRA id	tag	filtered sra pairs	unmapped pairs	unmapped pairs%
GSM2740422	SRR5928123	MC	67,645,955	1,816,981	2.69
GSM2740423	SRR5928124	MC	66,059,011	932,694	1.41
GSM2740424	SRR5928125	MC	58,034,486	754,451	1.3
GSM2740425	SRR5928126	MC	62,520,497	865,499	1.38
GSM2740426	SRR5928127	MC	64,588,121	853,606	1.32
GSM2740427	SRR5928128	MC	60,629,658	722,630	1.19

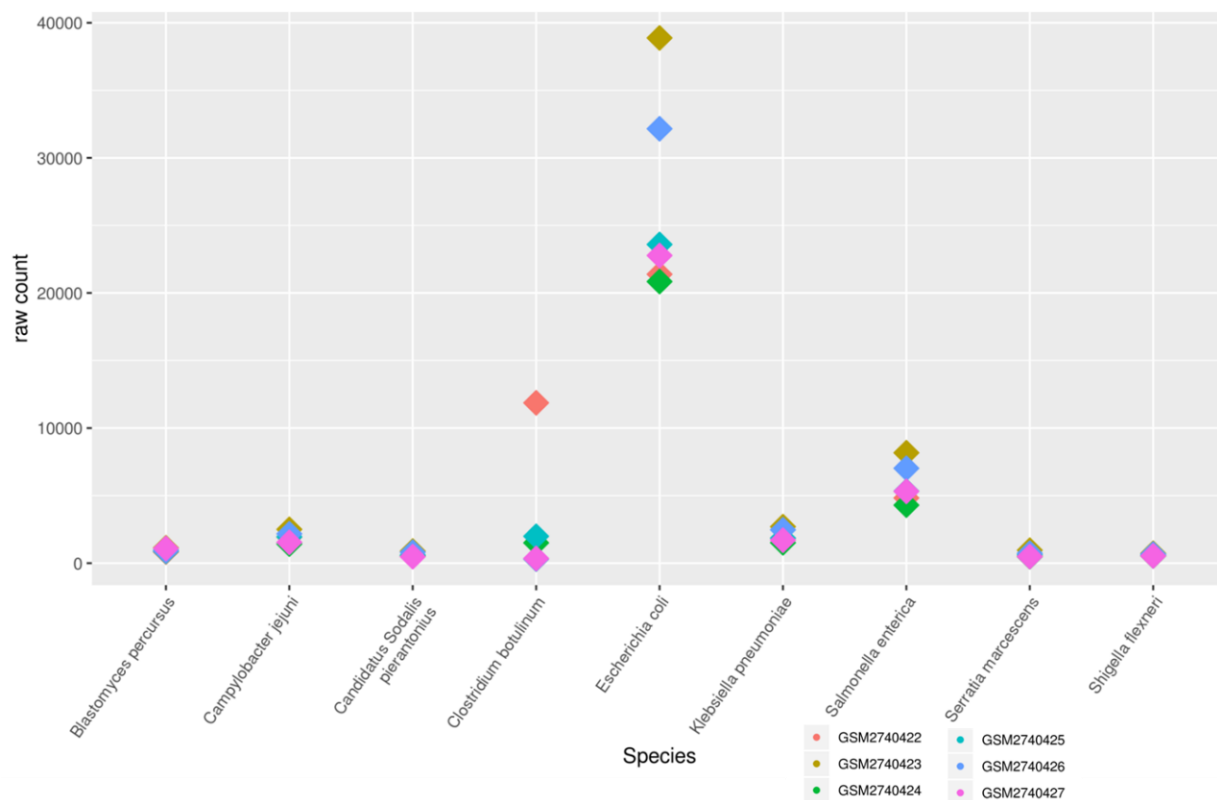
3.1.2 Μεταγονιδιωματική Ανάλυση

Χρησιμοποιώντας τα τρία προγράμματα μεταγονιδιωματικής για να αναλύσουμε τις αλληλουχίες που δε στοιχίζονται στο γονιδίωμα του ποντικού, επιβεβαιώνουμε την παρουσία της *Eschericia coli* και στα 6 δείγματα, όπως αναμενόταν. Η παρουσία άλλων μικροοργανισμών που παρατηρούμε σε μεμονωμένα δείγματα δεν επιβεβαιώνεται ταυτόχρονα και από τα τρία προγράμματα (βλ. **Εικόνες 3.1, 3.2, 3.3**).

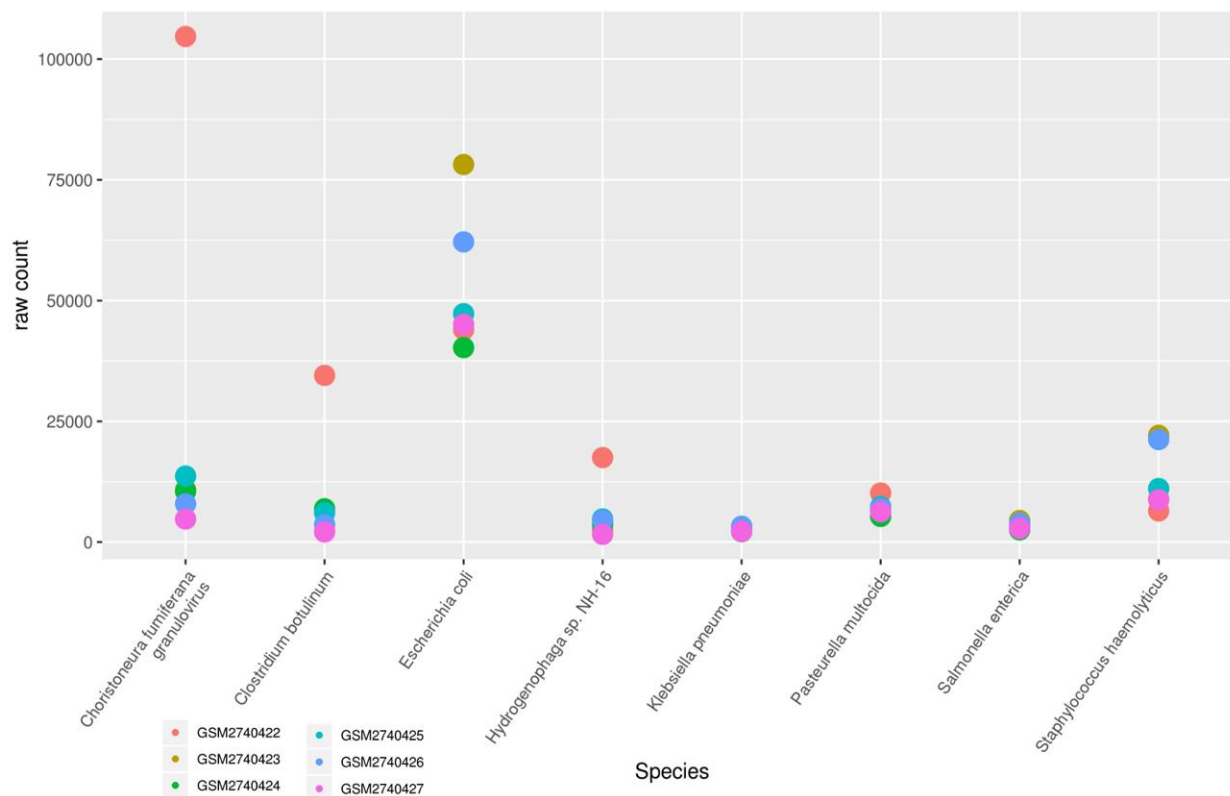


Εικόνα 3.1 Τα 6 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 6 δείγματα (11 διαφορετικά είδη) και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, σύμφωνα με το πρόγραμμα **Kraken 2**, για τη σειρά δειγμάτων με id: GSE102543. Κάθε χρώμα αντιστοιχεί σε διαφορετικό δείγμα.

Εκτίμηση της μικροβιακής κατάστασης germ-free ποντικών χρησιμοποιώντας RNA-Seq δεδομένα



Εικόνα 3.2 Τα 6 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 6 δείγματα (9 διαφορετικά είδη) και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, σύμφωνα με το πρόγραμμα **CLARK**, για τη σειρά δειγμάτων με id: GSE102543. Κάθε χρώμα αντιστοιχεί σε διαφορετικό δείγμα.



Εικόνα 3.3 Τα 6 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 6 δείγματα (8 διαφορετικά είδη) και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, σύμφωνα με το πρόγραμμα

AGAMEMNON, για τη σειρά δειγμάτων με id: **GSE102543**. Κάθε χρώμα αντιστοιχεί σε διαφορετικό δείγμα.

3.1.3 Στοίχιση στο γονιδίωμα της *E. coli*

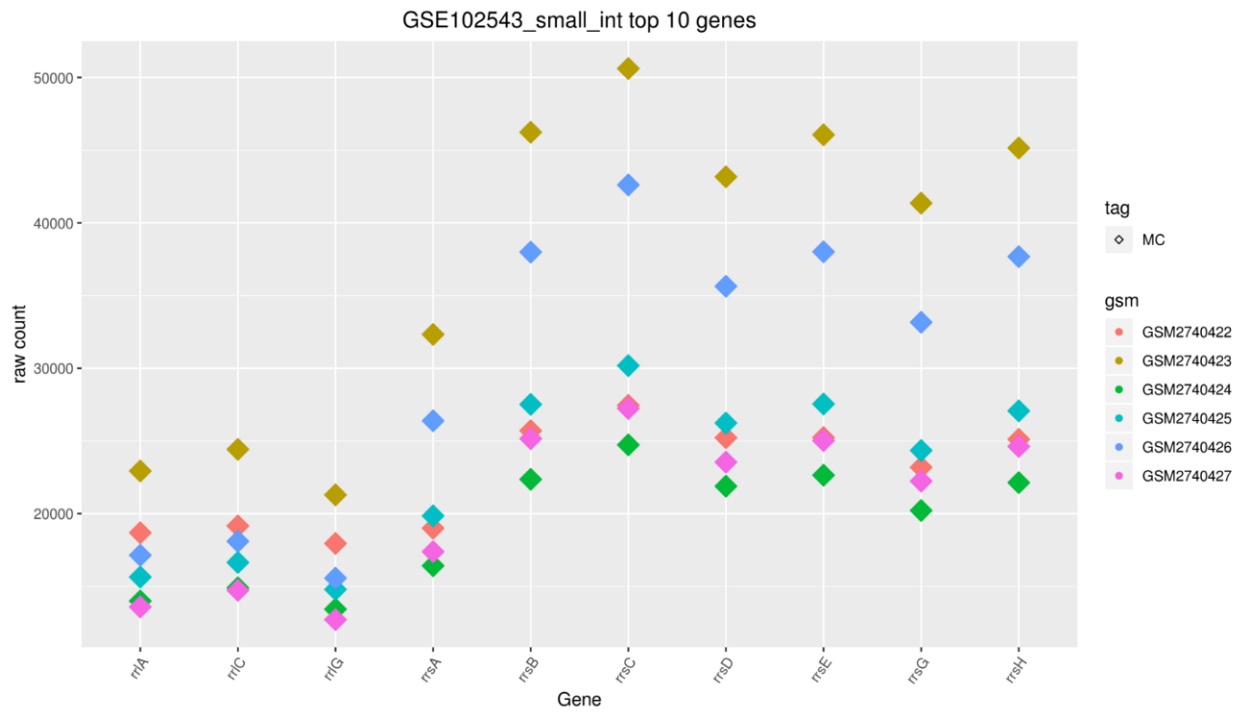
Στο σημείο αυτό, προχωρήσαμε σε στοίχιση των ίδιων αλληλουχιών στο γονιδίωμα της *E. coli* για να προσδιορίσουμε ποια γονίδια πιθανώς εκφράζονται. Χρησιμοποιήθηκε το γονιδίωμα με κωδικό GCF_000005845.2 (RefSeq assembly accession) της *E. coli* str, K-12 substr. MG1655. Το ποσοστό των αλληλουχιών που στοιχήθηκαν στο γονιδίωμα ήταν από 16% έως 49% (βλ. **Πίνακα 3.3**). Τα δέκα πιο συχνά γονίδια και στα έξι δείγματα ήταν rRNA γονίδια (τα *rrsA*, *rrsB*, *rrsC*, *rrsD*, *rrsE*, *rrsG*, *rrsH* που κωδικοποιούν 16S rRNAs και τα *rrlA*, *rrlC*, *rrlG* που κωδικοποιούν 23S rRNAs), γεγονός αναμενόμενο καθότι το 80% των RNA ενός τυπικού κυττάρου είναι rRNA. (βλ. **Εικόνα 3.4**). Όσον αφορά τα μη rRNA γονίδια, στα δέκα πιο συχνά συγκαταλέγονται τα *csrB*, *fusA*, *lacl*, *rimM*, *rnpB*, *rplJ*, *rplQ*, *rpoA*, *rpsD*, *rpsN*, *secY*, *ssrA*, *trmD* με προϊόντα που αναγράφονται στην **εικόνα 3.7**.

Με το πρόγραμμα IGV οπτικοποιήσαμε και παρουσιάζουμε τις αλληλουχίες που στοιχίζονται στα γονίδια *rrsC*, *rrlC* και *ssrA* (βλ. **Εικόνες 3.5, 3.8**), ενώ στην **Εικόνα 3.7** παρουσιάζουμε σε σχεδιάγραμμα τη δομή του οπερονίου των ριβοσωμικών γονιδίων *rrsC*, *rrlC*.

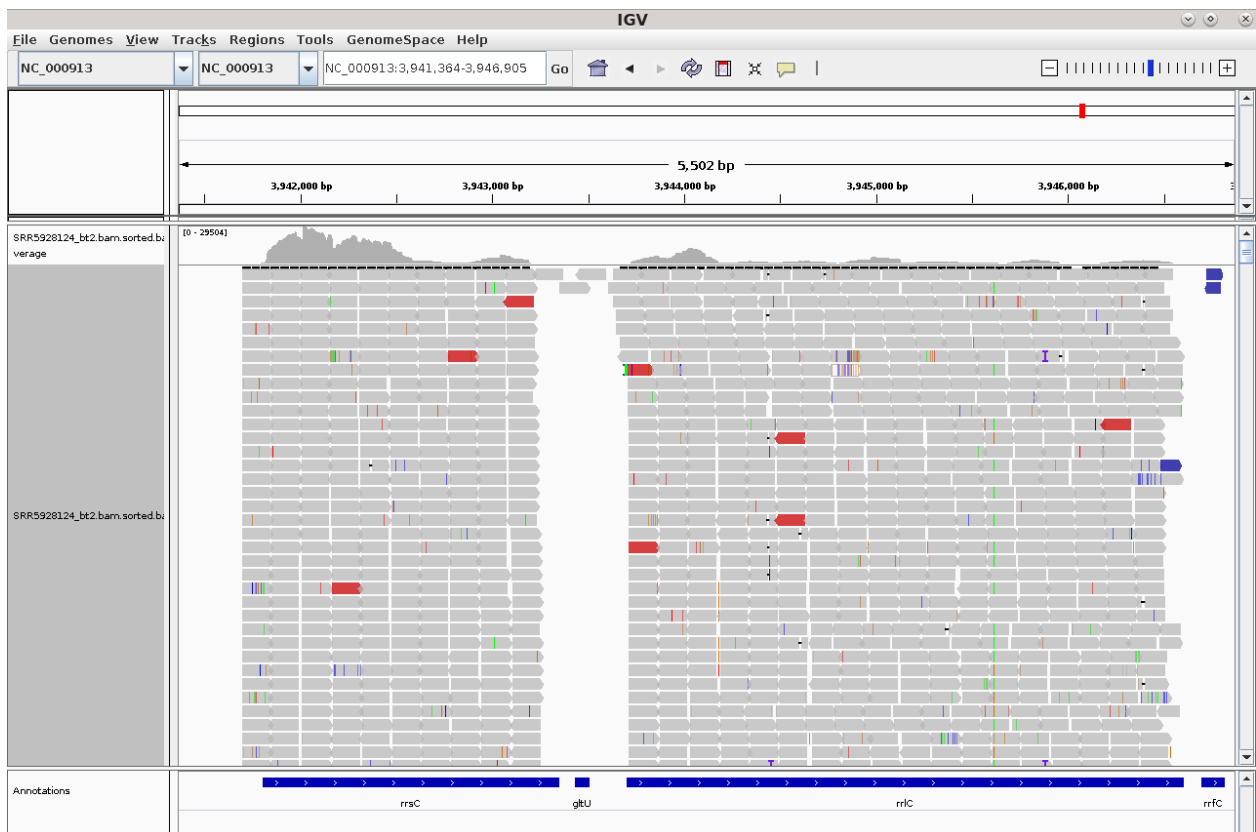
Πίνακας 3.3 Στατιστικά στοίχισης στο γονιδίωμα της *E. coli*. Αριθμός αλληλουχιών που δε στοιχήθηκαν στο γονιδίωμα του ποντικού (unmapped reads). Ποσοστό των unmapped reads που στοιχήθηκαν στο γονιδίωμα της *E. coli* (mapped to *E. coli* %). Ποσοστό των unmapped reads που στοιχήθηκαν σε γονίδια της *E. coli* (mapped to *E. coli* genes %). Αριθμός γονιδίων με τουλάχιστον μια αλληλουχία/read στην περιοχή τους.

sample id	SRA id	tag	unmapped reads	mapped to <i>E. coli</i> %	mapped to <i>E. coli</i> genes %	unique genes hit
GSM2740422	SRR5928123	MC	1,816,981	16.30	16.30	290
GSM2740423	SRR5928124	MC	932,694	48.91	48.91	323
GSM2740424	SRR5928125	MC	754,451	32.38	32.38	240
GSM2740425	SRR5928126	MC	865,499	33.10	33.10	258
GSM2740426	SRR5928127	MC	853,606	42.44	42.44	262
GSM2740427	SRR5928128	MC	722,630	35.42	35.42	257

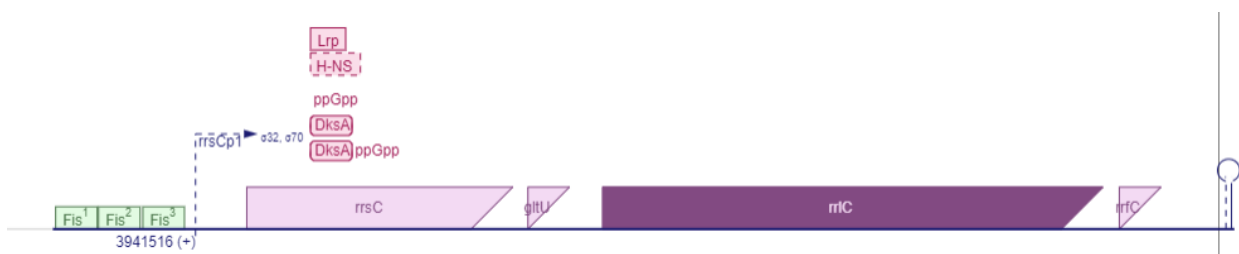
Εκτίμηση της μικροβιακής κατάστασης germ-free ποντικών χρησιμοποιώντας RNA-Seq δεδομένα



Εικόνα 3.4 Τα δέκα πιο συχνά γονίδια σε κάθε ένα από τα έξι δείγματα (δέκα διαφορετικά γονίδια, κοινά και στα έξι δείγματα) και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, για τη σειρά δειγμάτων με id: GSE102543. Κάθε χρώμα αντιστοιχεί σε διαφορετικό δείγμα. Και τα δέκα γονίδια είναι rRNA γονίδια.

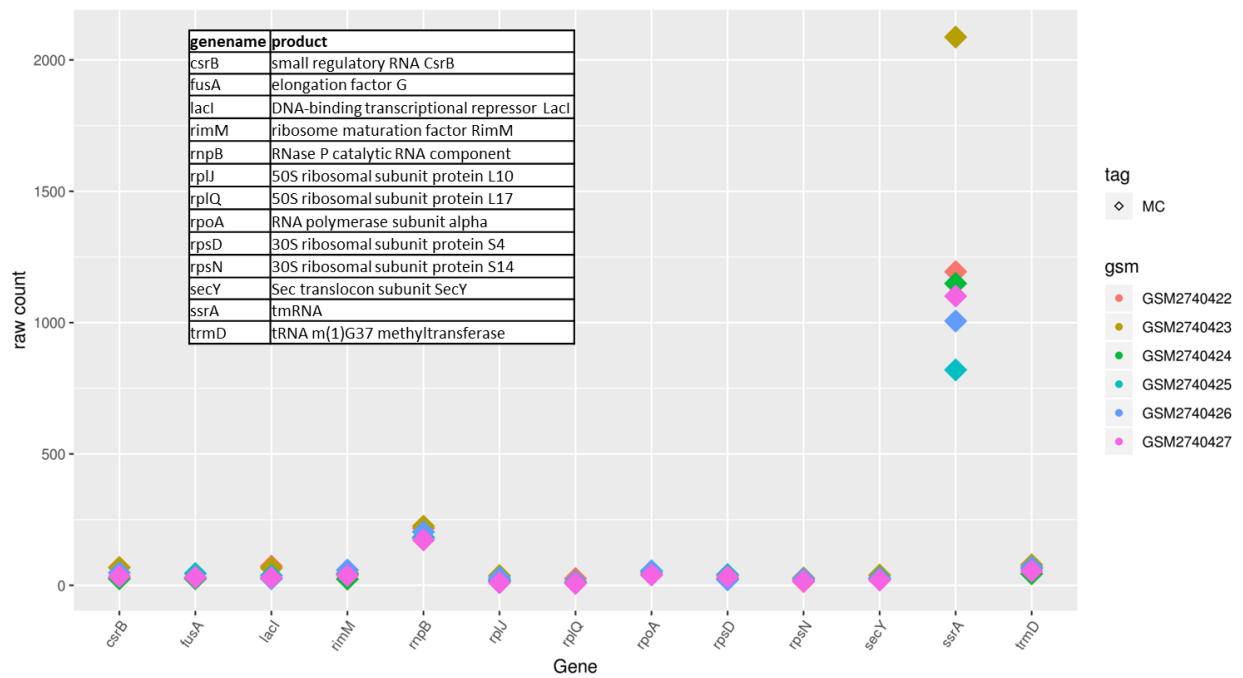


Εικόνα 3.5 Ένα παράδειγμα της στοίχισης αλληλουχιών του δεύτερου δείγματος στο γονιδίωμα της *E. coli*, και συγκεκριμένα στα rRNA γονίδια *rrsC*, *rriC*, όπως φαίνεται με το πρόγραμμα οπτικοποίησης στοίχισεων IGV. Τα δύο αυτά γονίδια ανήκουν στο ίδιο οπερόνιο (βλ. Εικόνα 3.6).



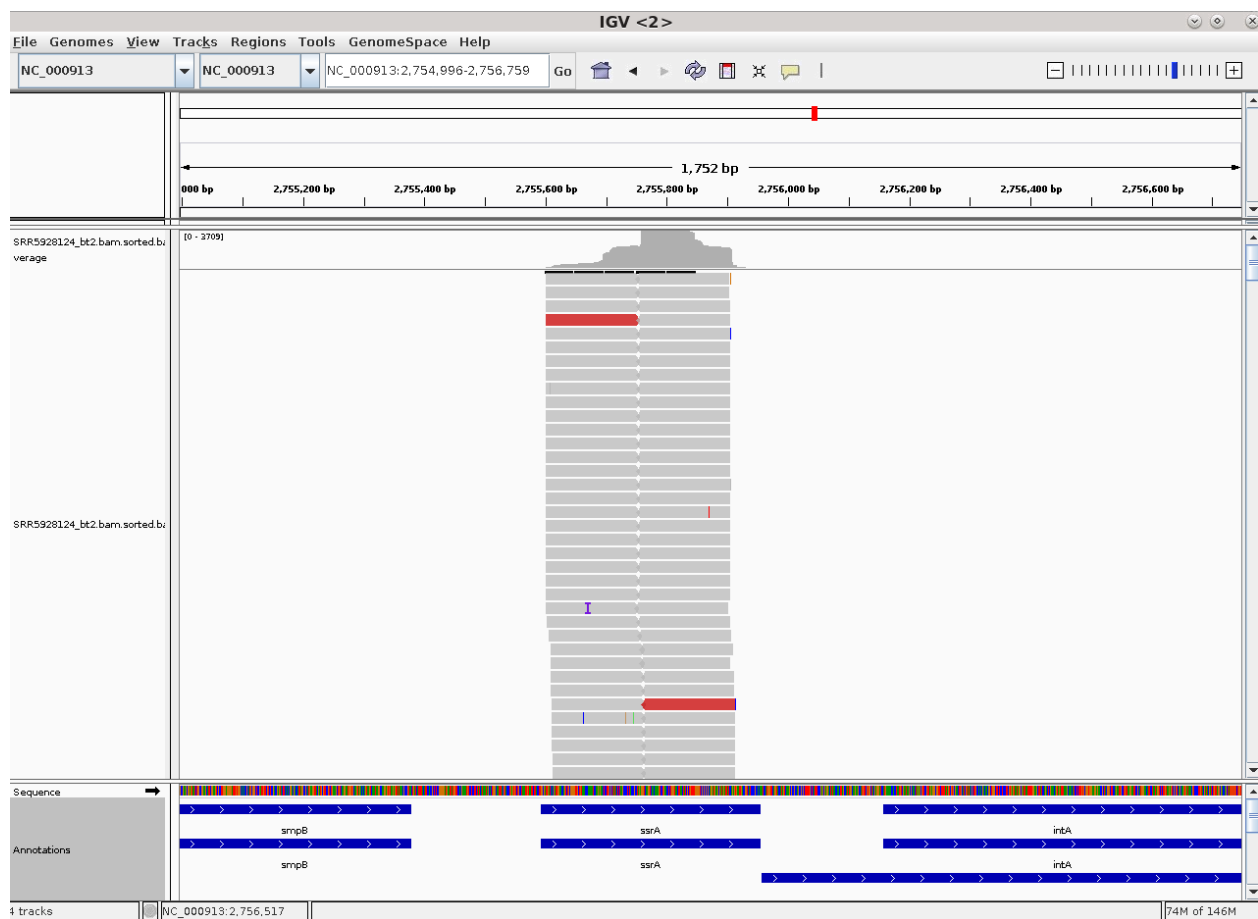
Εικόνα 3.6 Οπερόνιο *rrsC*, στο γονιδίωμα της *E. coli*. Αποτελείται από τα γονίδια *rrsC* (16S rRNA), *gltU* (tRNA-Glu(UUC)), *rriC* 23S rRNA και *rrfC* (5S rRNA) [78].

Εκτίμηση της μικροβιακής κατάστασης germ-free ποντικών χρησιμοποιώντας RNA-Seq δεδομένα



Εικόνα 3.7 Τα δέκα πιο συχνά γονίδια, εξαιρουμένων των rRNA γονιδίων, στα έξι δείγματα και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, για τη σειρά δειγμάτων με id: GSE102543.

Κάθε χρώμα αντιστοιχεί σε διαφορετικό δείγμα. Στον ένθετο πίνακα αναγράφεται το προϊόν (product) κάθε γονιδίου. Ξεχωρίζει το γονίδιο *ssrA* που παράγει ένα RNA με δράση μεταφορικού-RNA και αγγελιαφόρου-RNA (tm-RNA).



Εικόνα 3.8 Ένα παράδειγμα της στοίχισης αλληλουχιών του δεύτερου δείγματος στο γονιδίωμα της *E. coli*, και συγκεκριμένα στο γονίδιο *ssrA*, το οποίο ήταν το πιο συχνό και στα έξι δείγματα. Η ισχυρή παρουσία του γονιδίου αυτού οφείλεται σε πολλά αντίγραφα αλληλουχιών/reads που καλύπτουν όλη την έκταση του γονιδίου, μήκους 362 ζευγών βάσεων.

3.2 Σειρά δειγμάτων GSE112571

3.2.1 Περιγραφή δειγμάτων, προ-επεξεργασία και στοίχιση στο γονιδίωμα του ΠΟΝΤΙΚΟΥ

- **Δημοσίευση:** Yao L et al., "A selective gut bacterial bile salt hydrolase alters host metabolism.", *Elife*, 2018 Jul 17;7
- **Ιστός προέλευσης:** ειλεός (distal ileum)
- **Σχεδιασμός:** paired-end
- **Μόριο:** "total RNA", ολικό RNA
- **Δείγματα:** 12 δείγματα προερχόμενα από 12 gnotobiotic ποντίκια εποικισμένα μόνο με το βακτήριο *Bacteroides thetaiotaomicron* (monocolonized, MC). Για κάθε δείγμα διαθέτουμε δύο τεχνικά αντίγραφα (technical replicates).

Πίνακας 3.4 Χαρακτηριστικά δειγμάτων της σειράς με id GSE112571

όνομα δείγματος	χαρακτηριστικά δείγματος
GSM3073090: Bt_WT1_S1; Mus musculus; RNA-Seq	2 ILLUMINA (Illumina HiSeq 2500) runs: 18.5M spots , 3.7G bases, 1.6Gb
GSM3073091: Bt_WT2_S2; Mus musculus; RNA-Seq	2 ILLUMINA (Illumina HiSeq 2500) runs: 20.7M spots , 4.1G bases, 1.8Gb
GSM3073092: Bt_WT3_S3; Mus musculus; RNA-Seq	2 ILLUMINA (Illumina HiSeq 2500) runs: 19.1M spots , 3.8G bases, 1.7Gb
GSM3073093: Bt_WT4_S4; Mus musculus; RNA-Seq	2 ILLUMINA (Illumina HiSeq 2500) runs: 18.2M spots , 3.6G bases, 1.6Gb
GSM3073094: Bt_WT5_S5; Mus musculus; RNA-Seq	2 ILLUMINA (Illumina HiSeq 2500) runs: 24.2M spots , 4.8G bases, 2.1Gb
GSM3073095: Bt_WT6_S6; Mus musculus; RNA-Seq	2 ILLUMINA (Illumina HiSeq 2500) runs: 32.9M spots , 6.6G bases, 2.8Gb
GSM3073096: Bt_KO1_S7; Mus musculus; RNA-Seq	2 ILLUMINA (Illumina HiSeq 2500) runs: 23.1M spots , 4.6G bases, 2Gb
GSM3073097: Bt_KO2_S8; Mus musculus; RNA-Seq	2 ILLUMINA (Illumina HiSeq 2500) runs: 14.1M spots , 2.8G bases, 1.2Gb
GSM3073098: Bt_KO3_S9; Mus musculus; RNA-Seq	2 ILLUMINA (Illumina HiSeq 2500) runs: 16M spots , 3.2G bases, 1.4Gb
GSM3073099: Bt_KO4_S10; Mus musculus; RNA-Seq	2 ILLUMINA (Illumina HiSeq 2500) runs: 21.6M spots , 4.3G bases, 1.9Gb
GSM3073100: Bt_KO5_S11; Mus musculus; RNA-Seq	2 ILLUMINA (Illumina HiSeq 2500) runs: 23.3M spots , 4.7G bases, 2Gb
GSM3073101: Bt_KO6_S12; Mus musculus; RNA-Seq	2 ILLUMINA (Illumina HiSeq 2500) runs: 21M spots , 4.2G bases, 1.8Gb

Προ-επεξεργασία αλληλουχιών

- Αρχικό μήκος αλληλουχιών: 100 βάσεις
- Εντοπίστηκε ο universal Illumina «προσαρμογέας» στις αλληλουχίες, οπότε αφαιρέθηκε.
- Ύστερα από την περικοπή των αλληλουχιών με quality score <10 στο 3' άκρο τους, κρατήσαμε τις αλληλουχίες όπου και τα δύο mates είχαν μήκος ≥ 80 βάσεις.
- Οι αλληλουχίες που δεν στοιχήθηκαν στο γονιδίωμα του ποντικού με το πρόγραμμα Hisat2, μετά από δύο προσπάθειες, συλλέχθηκαν (βλ. Πίνακα 3.5).

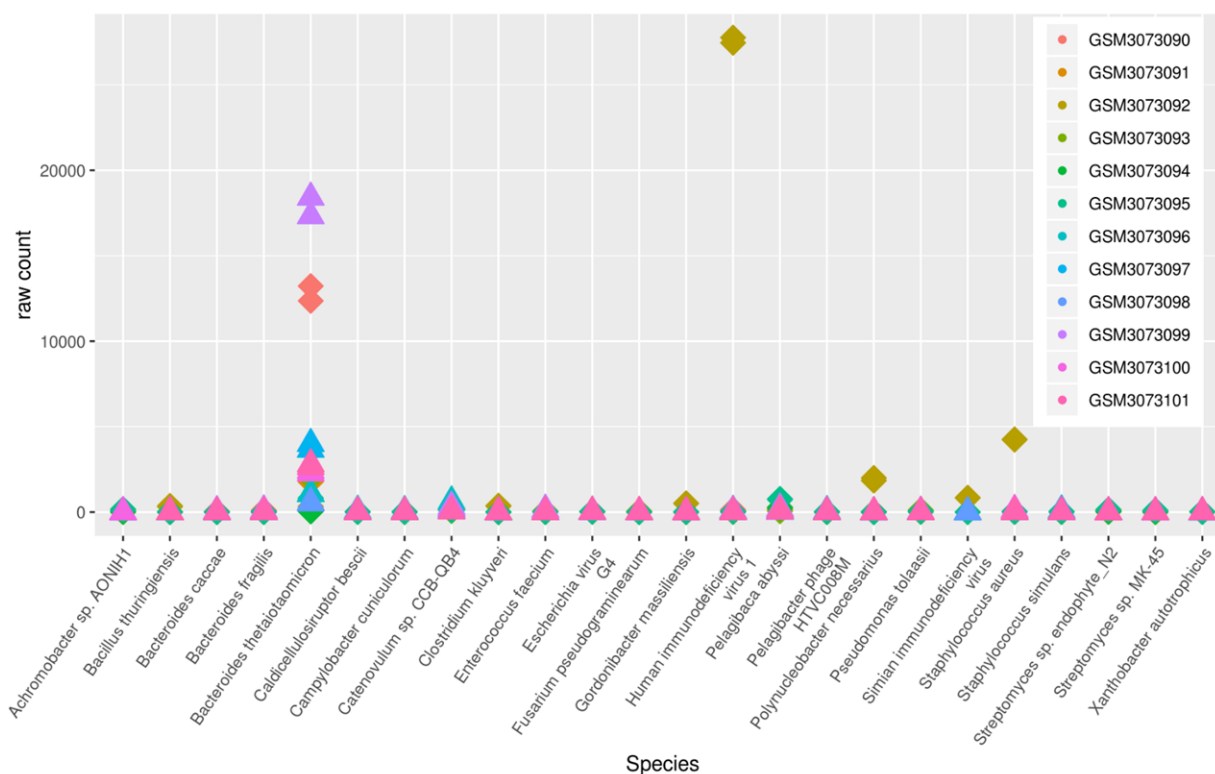
Πίνακας 3.5 Ποσοστό αλληλουχιών που δεν στοιχίζονται στο γονιδίωμα του ποντικού (unmapped pairs%) στα δώδεκα δείγματα. Επίσης, παρουσιάζεται ο NCBI GEO κωδικός του κάθε δείγματος (sample id), ο SRA κωδικός των δύο τεχνικών αντιγράφων του (SRA id), ο χαρακτηρισμός του (tag , MC=monocolonized), ο αριθμός των αλληλουχιών μετά την προ-επεξεργασία/φιλτράρισμα (filtered sra pairs) και ο αριθμός των αλληλουχιών που δεν στοιχίζονται στο γονιδίωμα του ποντικού (unmapped pairs).

sample id	SRA id	tag	filtered sra pairs	unmapped pairs	unmapped pairs%
GSM3073090	SRR6921773	MC	6,621,176	378,049	5.71
GSM3073090	SRR6921772	MC	6,206,105	352,240	5.68
GSM3073091	SRR6921775	MC	7,523,053	348,214	4.63
GSM3073091	SRR6921774	MC	7,017,820	322,075	4.59
GSM3073092	SRR6921777	MC	5,724,294	353,367	6.17
GSM3073092	SRR6921776	MC	5,379,162	332,759	6.19
GSM3073093	SRR6921779	MC	5,854,758	254,911	4.35
GSM3073093	SRR6921778	MC	5,464,656	236,648	4.33
GSM3073094	SRR6921781	MC	7,812,222	358,111	4.58
GSM3073094	SRR6921780	MC	7,280,689	330,539	4.54
GSM3073095	SRR6921783	MC	10,664,999	514,775	4.83
GSM3073095	SRR6921782	MC	9,953,125	477,721	4.8
GSM3073096	SRR6921785	MC_KO	7,658,451	469,230	6.13
GSM3073096	SRR6921784	MC_KO	7,117,496	434,797	6.11
GSM3073097	SRR6921787	MC_KO	4,165,143	244,272	5.86
GSM3073097	SRR6921786	MC_KO	3,875,713	224,834	5.8
GSM3073098	SRR6921789	MC_KO	4,756,637	293,182	6.16
GSM3073098	SRR6921788	MC_KO	4,466,900	273,825	6.13
GSM3073099	SRR6921791	MC_KO	6,171,045	434,532	7.04
GSM3073099	SRR6921790	MC_KO	5,752,258	402,082	6.99
GSM3073100	SRR6921793	MC_KO	7,483,594	420,560	5.62
GSM3073100	SRR6921792	MC_KO	7,020,611	393,988	5.61
GSM3073101	SRR6921794	MC_KO	5,844,077	325,303	5.57
GSM3073101	SRR6921795	MC_KO	6,257,707	351,015	5.61

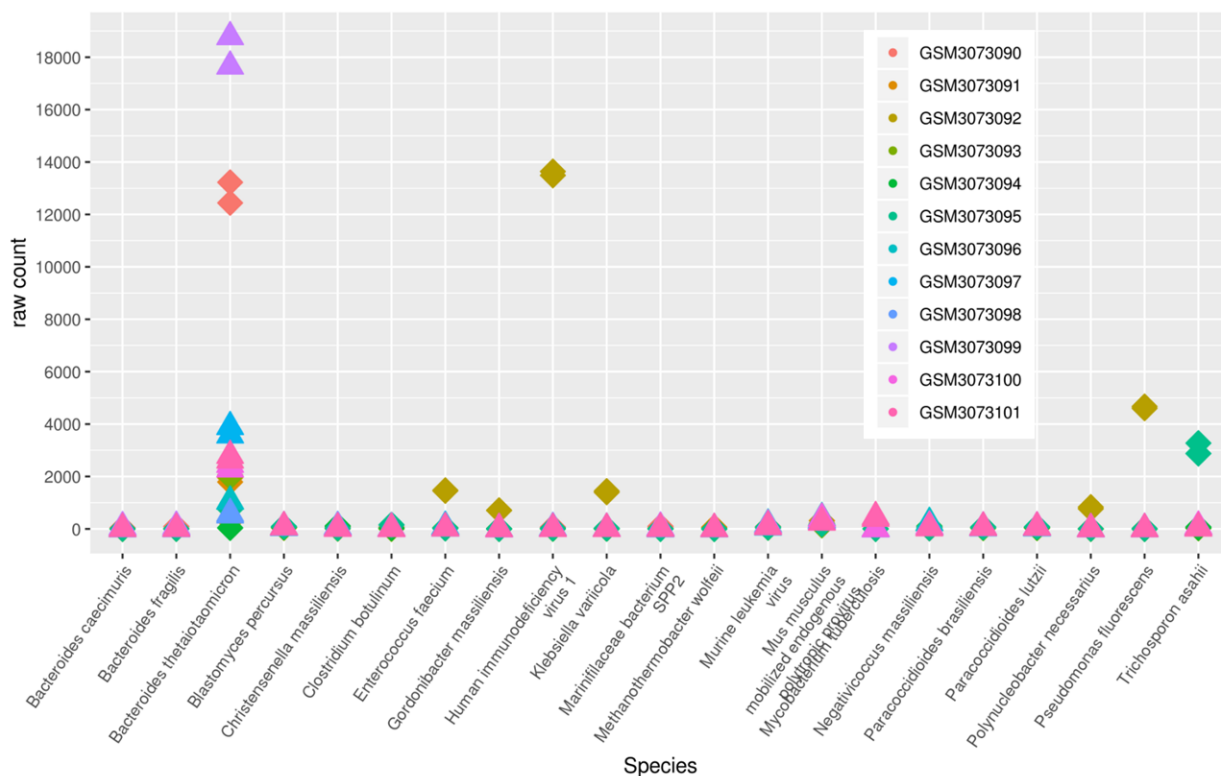
3.2.2 Μεταγονιδιωματική Ανάλυση

Χρησιμοποιώντας τα τρία προγράμματα μεταγονιδιωματικής ανάλυσης επιβιώνουμε την παρουσία του *Bacteroides thetaiotaomicron* και στα 12 δείγματα, από τον ειλεό 12 ποντικών επικοισμένων μόνο με *Bacteroides thetaiotaomicron* σε κατά τα άλλα germ-free συνθήκες. Η παρουσία άλλων μικροοργανισμών σε μεμονωμένα δείγματα δεν

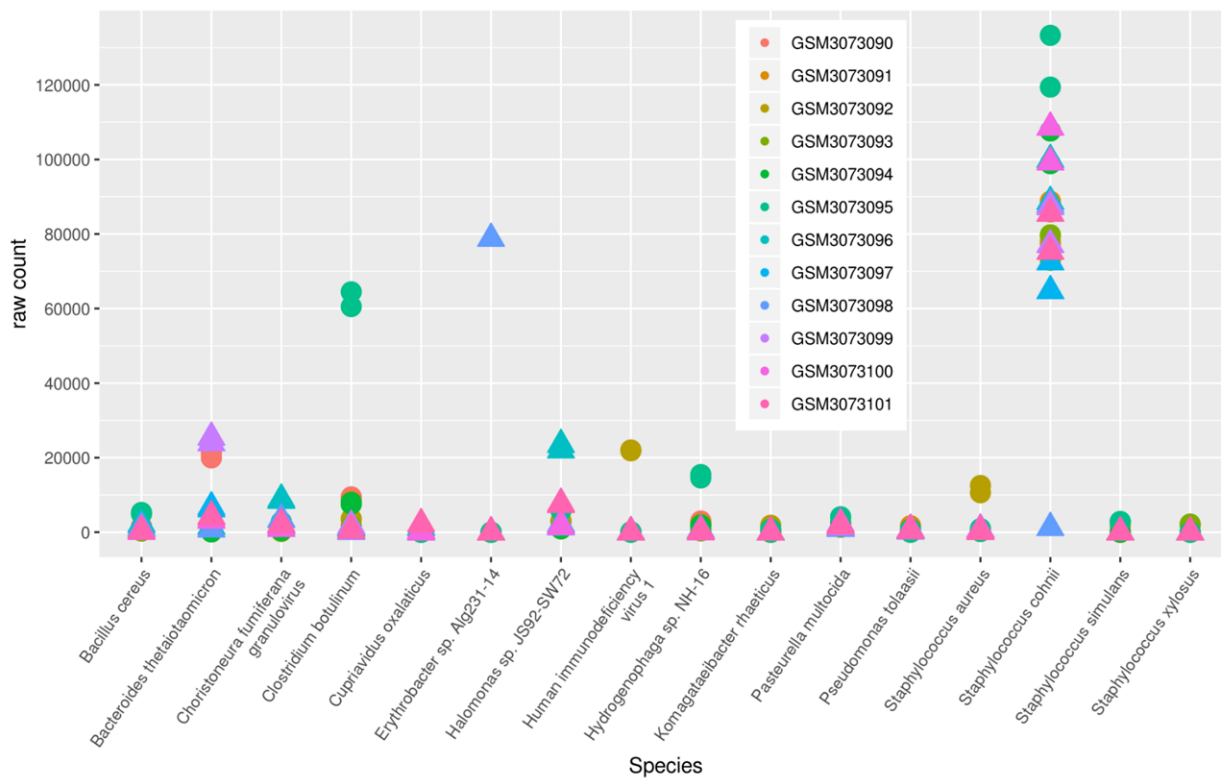
επιβεβαιώνεται ταυτόχρονα και από τα τρία προγράμματα (βλ. **Εικόνες 3.9, 3.10, 3.11**). Ενδιαφέρον έχει η εύρεση του ιού HIV-1 και με τα τρία προγράμματα μεταγονιδιωμικής ανάλυσης και στα δύο τεχνικά αντίγραφα του τρίτου δείγματος.



Εικόνα 3.9 Τα 6 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 12 δείγματα (24 διαφορετικά είδη) και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, σύμφωνα με το πρόγραμμα **Kraken 2**, για τη σειρά δειγμάτων με id: GSE112571. Κάθε χρώμα αντιστοιχεί σε διαφορετικό δείγμα.



Εικόνα 3.10 Τα 6 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 12 δείγματα (21 διαφορετικά είδη) και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, σύμφωνα με το πρόγραμμα CLARK, για τη σειρά δειγμάτων με id: GSE112571. Κάθε χρώμα αντιστοιχεί σε διαφορετικό δείγμα.



Εικόνα 3.11 Τα 6 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 12 δείγματα (16 διαφορετικά είδη) και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, σύμφωνα με το

πρόγραμμα **AGAMEMNON**, για τη σειρά δειγμάτων με id: GSE112571. Κάθε χρώμα αντιστοιχεί σε διαφορετικό δείγμα.

3.2.3 Στοιχίση στο γονιδίωμα του *Bacteroides thetaiotaomicron* και του HIV-1

Στο σημείο αυτό, προχωρήσαμε σε στοιχίση των ίδιων αλληλουχιών στο γονιδίωμα του *Bacteroides thetaiotaomicron* για να προσδιορίσουμε ποια γονίδια πιθανώς εκφράζονται. Χρησιμοποιήθηκε το γονιδίωμα με κωδικό GCF_000011065.1 (RefSeq assembly accession) του *Bacteroides thetaiotaomicron* strain VPI-5482. Το ποσοστό των αλληλουχιών που στοιχήθηκαν στο γονιδίωμα ήταν μικρότερο από 4% (βλ. Πίνακα 3.6).

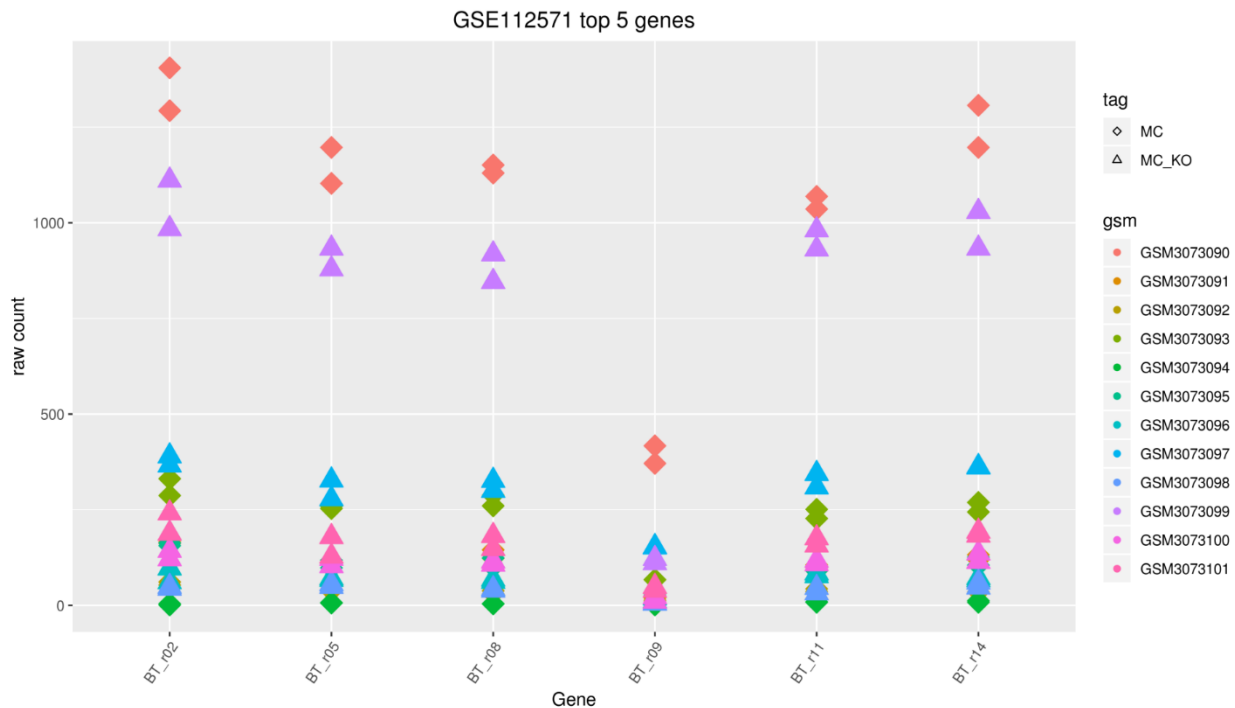
Πίνακας 3.6 Στατιστικά στοιχίσης στο γονιδίωμα του *B. thetaiotaomicron*. Αριθμός αλληλουχιών που δε στοιχήθηκαν στο γονιδίωμα του ποντικού (unmapped reads). Ποσοστό των unmapped reads που στοιχήθηκαν στο γονιδίωμα του *B. thetaiotaomicron* (mapped to *B. thetaiotaomicron* %). Ποσοστό των unmapped reads που στοιχήθηκαν σε γονίδια του *B. thetaiotaomicron* (mapped to *B. thetaiotaomicron* genes %). Αριθμός γονιδίων με τουλάχιστον μια αλληλουχία/read στην περιοχή τους.

sample id	SRA id	tag	unmapped reads	mapped to <i>B. thetaiotaomicron</i> %	mapped to <i>B. thetaiotaomicron</i> genes %	unique genes hit
GSM3073090	SRR6921773	MC	378049	3.64	3.26	1529
GSM3073090	SRR6921772	MC	352240	3.67	3.3	1470
GSM3073091	SRR6921775	MC	348214	0.48	0.41	508
GSM3073091	SRR6921774	MC	322075	0.49	0.41	462
GSM3073092	SRR6921777	MC	353367	0.11	0.1	77
GSM3073092	SRR6921776	MC	332759	0.1	0.09	67
GSM3073093	SRR6921779	MC	254911	0.97	0.91	433
GSM3073093	SRR6921778	MC	236648	0.99	0.94	408
GSM3073094	SRR6921781	MC	358111	0.01	0.01	19
GSM3073094	SRR6921780	MC	330539	0.01	0.01	13
GSM3073095	SRR6921783	MC	514775	0.19	0.18	146
GSM3073095	SRR6921782	MC	477721	0.19	0.18	135
GSM3073096	SRR6921785	MC_KO	469230	0.22	0.19	314
GSM3073096	SRR6921784	MC_KO	434797	0.23	0.19	345
GSM3073097	SRR6921787	MC_KO	244272	1.76	1.53	699
GSM3073097	SRR6921786	MC_KO	224834	1.77	1.55	669
GSM3073098	SRR6921789	MC_KO	293182	0.18	0.15	178
GSM3073098	SRR6921788	MC_KO	273825	0.18	0.16	185
GSM3073099	SRR6921791	MC_KO	434532	3.25	2.5	1587
GSM3073099	SRR6921790	MC_KO	402082	3.28	2.54	1552
GSM3073100	SRR6921793	MC_KO	420560	0.53	0.42	658

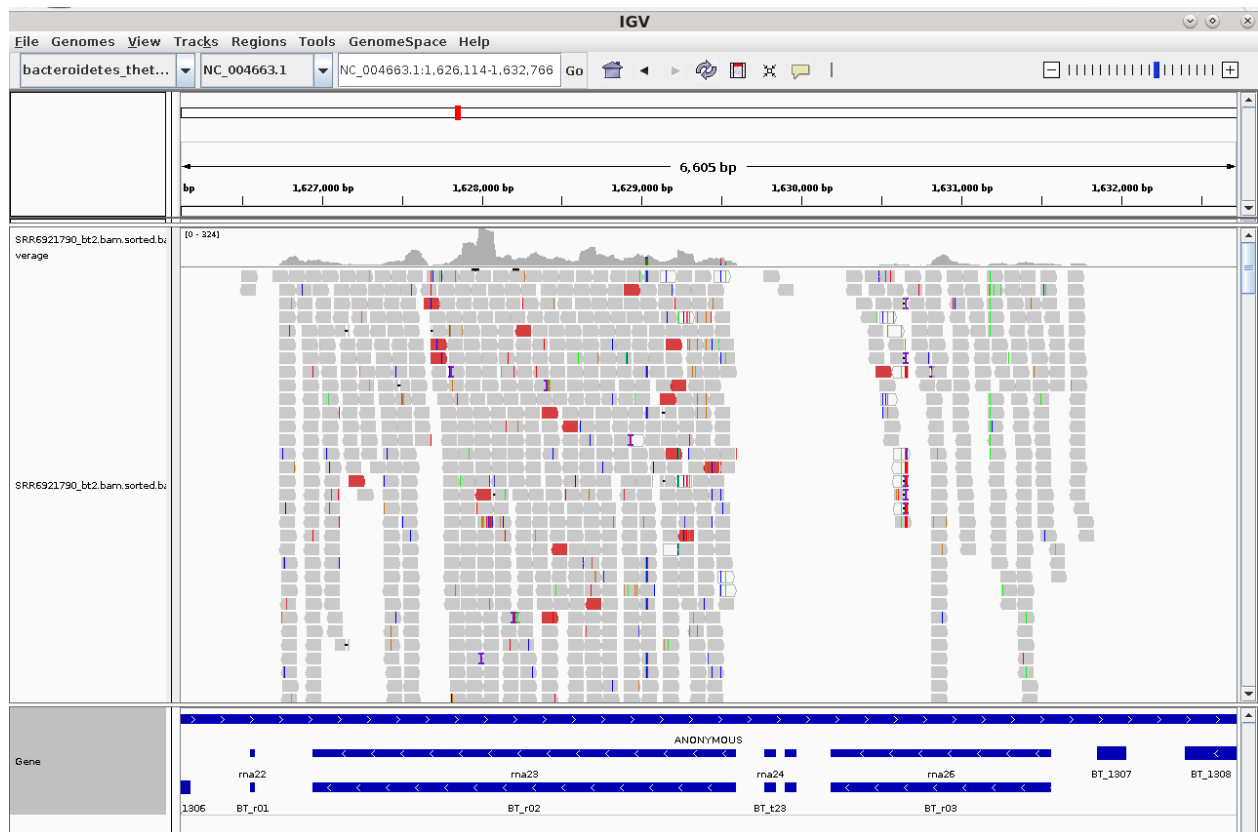
GSM3073100	SRR6921792	MC_KO	393988	0.52	0.4	598
GSM3073101	SRR6921794	MC_KO	325303	0.77	0.62	635
GSM3073101	SRR6921795	MC_KO	351015	0.76	0.61	656

Τα πέντε πιο συχνά γονίδια και στα δώδεκα δείγματα ήταν rRNA γονίδια, γεγονός αναμενόμενο καθότι το 80% των RNA ενός τυπικού κυττάρου είναι rRNA. (βλ. **Εικόνα 3.12**). Στην **Εικόνα 3.13** παρουσιάζουμε τη στοίχιση των αλληλουχιών στα ριβοσωμικά γονίδια BT_r02 (23S ribosomal) και BT_r03 (16S ribosomal).

Από τα 5 πιο συχνά γονίδια κάθε δείγματος, εξαιρουμένων των rRNA γονιδίων (**Εικόνα 3.14**), επιλέγουμε να παρουσιάσουμε, με τη βοήθεια του προγράμματος IGV, το γονίδιο που κωδικοποιεί το ένζυμο αφυδρογονάση μηλικού (Malate dehydrogenase). Οι αλληλουχίες των δειγμάτων καλύπτουν μεγάλο μέρος του μήκους του γονιδίου.



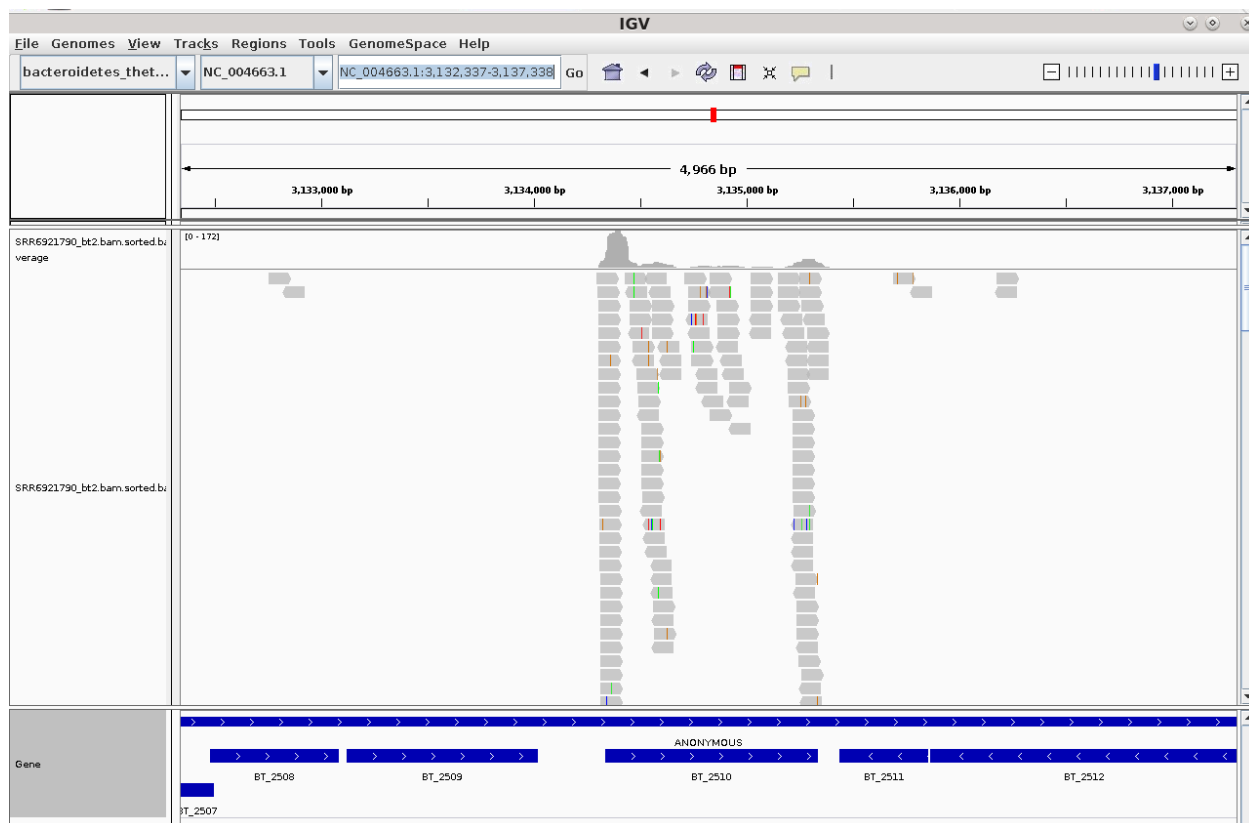
Εικόνα 3.12 Τα πέντε πιο συχνά γονίδια σε κάθε ένα από τα δώδεκα δείγματα (έξι διαφορετικά γονίδια) και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, για τη σειρά δειγμάτων με id: GSE112571. Κάθε χρώμα αντιστοιχεί σε διαφορετικό δείγμα. Και τα έξι γονίδια είναι rRNA γονίδια.



Εικόνα 3.13 Ένα παράδειγμα της στοιχίσης αλληλουχιών του δέκατου δείγματος στο γονιδίωμα του *Bacteroides thetaiotaomicron*, και συγκεκριμένα στα rRNA γονίδια BT_r02 (23S ribosomal) και BT_r03 (16S ribosomal), όπως φαίνεται με το πρόγραμμα οπτικοποίησης στοιχίσεων IGV.



Εικόνα 3.14 Τα πέντε πιο συχνά γονίδια, εξαιρουμένων των rRNA γονιδίων, στα έξι δείγματα και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, για τη σειρά δειγμάτων με id: GSE112571. Κάθε χρώμα αντιστοιχεί σε διαφορετικό δείγμα. Στον οριζόντιο άξονα αναφέρεται το προϊόν (product) κάθε γονιδίου.



Εικόνα 3.15 Ένα παράδειγμα της στοίχισης αλληλουχιών του δέκατου δείγματος στο γονιδίωμα του *Bacteroides thetaiotaomicron* και συγκεκριμένα στο γονίδιο BT_2510, που κωδικοποιεί το ένζυμο αφυδρογονάση μηλικού (Malate dehydrogenase). Η ισχυρή παρουσία του γονιδίου αυτού οφείλεται σε πολλά αντίγραφα αλληλουχιών που καλύπτουν μεγάλο μέρος του μήκους του γονιδίου.

Στη συνέχεια, προχωρήσαμε σε στοίχιση των αλληλουχιών του τρίτου δείγματος στο γονιδίωμα του ιού HIV-1, με κωδικό GCF_000864765.1 (RefSeq assembly accession). Διαπιστώσαμε ότι η ύπαρξη του ιού στα αποτελέσματα της μεταγονιδιωμικής ανάλυσης δεν οφείλεται στην ύπαρξη μίας μόνο συγκεκριμένης αλληλουχίας με πολλά αντίγραφα, αλλά σε πολλές διαφορετικές αλληλουχίες που καλύπτουν μεγάλο ποσοστό του συνολικού μήκους του γονιδιώματος του ιού (βλ. **Εικόνα 3.16**). Οι αλληλουχίες αυτές ωστόσο, περιέχουν αντικαταστάσεις, προσθήκες ή/και ελλείψεις νουκλεοτιδίων, όπως φαίνεται στην **Εικόνα 3.16**.



Εικόνα 3.16 Στοίχιση των αλληλουχιών του τρίτου δείγματος (χωρισμένες σε δύο τεχνικά αντίγραφα με SRA ids: SRR6921776, SRR6921777) στο γονιδίωμα του ιού HIV-1. Οι αλληλουχίες του δείγματος καλύπτουν όλο σχεδόν το μήκος του γονιδιώματος του ιού.

3.3 Σειρά δειγμάτων GSE81117

3.3.1 Περιγραφή δειγμάτων, προ-επεξεργασία και στοίχιση στο γονιδίωμα του ποντικού

- **Δημοσίευση:** Krautkramer KA, Kreznar JH, Romano KA, Vivas EI et al. Diet-Microbiota Interactions Mediate Global Epigenetic Programming in Multiple Host Tissues. Mol Cell 2016 Dec 1;64(5):982-992. PMID: 27889451
- **Ιστός προέλευσης:** συκώτι
- **Σχεδιασμός:** single-end
- **Μόριο:** “total RNA”, ολικό RNA
- **Δείγματα:** 15 δείγματα προερχόμενα από 15 ποντίκια εκ των οποίων:
 - ❖ 3 είναι germ-free με τυπική δίαιτα (GF_chow)
 - ❖ 3 είναι germ-free με δίαιτα πλούσια σε λίπη και σάκχαρα (GF_HFHS)
 - ❖ 3 έχουν φυσιολογική μικροχλωρίδα (Conventionally raised) και τυπική δίαιτα (ConvR_chow)
 - ❖ 3 έχουν φυσιολογική μικροχλωρίδα (Conventionally raised) και δίαιτα πλούσια σε λίπη και σάκχαρα (ConvR_HFHS)

- ❖ 3 είναι germ-free με τυπική δίαιτα που τους χορηγήθηκαν οι μικροοργανισμοί των ConvR ποντικών, (ConvD_chow)

Πίνακας 3.7 Χαρακτηριστικά δειγμάτων της σειράς με id GSE81117

όνομα δείγματος	χαρακτηριστικά δείγματος
GSM2143152: GF1_chow	1 ILLUMINA (Illumina HiSeq 2500) run: 18M spots , 1.8G bases, 860.1Mb
GSM2143153: GF2_chow	1 ILLUMINA (Illumina HiSeq 2500) run: 23.1M spots , 2.3G bases, 1.1Gb
GSM2143154: GF3_chow	1 ILLUMINA (Illumina HiSeq 2500) run: 28.4M spots , 2.9G bases, 1.3Gb
GSM2143155: ConvR1_chow	1 ILLUMINA (Illumina HiSeq 2500) run: 25.1M spots , 2.5G bases, 1.2Gb
GSM2143156: ConvR2_chow	1 ILLUMINA (Illumina HiSeq 2500) run: 22.9M spots , 2.3G bases, 1.1Gb
GSM2143157: ConvR3_chow	1 ILLUMINA (Illumina HiSeq 2500) run: 21.7M spots , 2.2G bases, 1Gb
GSM2143158: ConvD1_chow	1 ILLUMINA (Illumina HiSeq 2500) run: 24.2M spots , 2.4G bases, 777.2Mb
GSM2143159: ConvD2_chow	1 ILLUMINA (Illumina HiSeq 2500) run: 28.7M spots , 2.9G bases, 927.1Mb
GSM2143160: ConvD3_chow	1 ILLUMINA (Illumina HiSeq 2500) run: 23.2M spots , 2.3G bases, 747.5Mb
GSM2143161: GF1_HFHS	1 ILLUMINA (Illumina HiSeq 2500) run: 26.7M spots , 2.7G bases, 864Mb
GSM2143162: GF2_HFHS	1 ILLUMINA (Illumina HiSeq 2500) run: 26.2M spots , 2.6G bases, 842.9Mb
GSM2143163: GF3_HFHS	1 ILLUMINA (Illumina HiSeq 2500) run: 23.5M spots , 2.4G bases, 758.5Mb
GSM2143164: ConvR1_HFHS	1 ILLUMINA (Illumina HiSeq 2500) run: 28.1M spots , 2.8G bases, 903.5Mb
GSM2143165: ConvR2_HFHS	1 ILLUMINA (Illumina HiSeq 2500) run: 23.9M spots , 2.4G bases, 775.9Mb
GSM2143166: ConvR3_HFHS	1 ILLUMINA (Illumina HiSeq 2500) run: 27.4M spots , 2.8G bases, 877.2Mb

Προ-επεξεργασία αλληλουχιών

- Αρχικό μήκος αλληλουχιών: 101 βάσεις
- Εντοπίστηκε ο universal Illumina «προσαρμογέας» στις αλληλουχίες οπότε αφαιρέθηκε.
- Ύστερα από την περικοπή των αλληλουχιών με quality score <10 στο 3' άκρο τους, κρατήσαμε τις αλληλουχίες που είχαν μήκος ≥ 80 βάσεις.
- Οι αλληλουχίες που δεν στοιχήθηκαν στο γονιδίωμα του ποντικού με το πρόγραμμα Hisat2, μετά από δύο προσπάθειες, συλλέχθηκαν (βλ. **Πίνακα 3.8**).

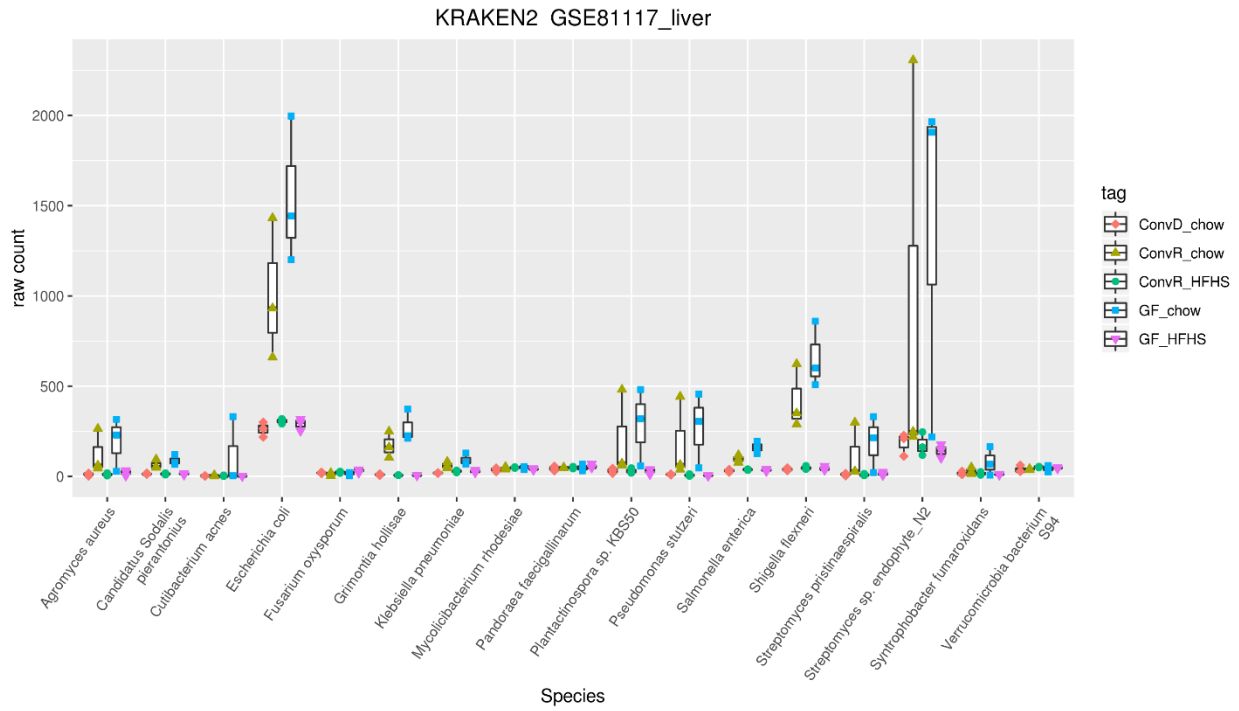
Πίνακας 3.8 Ποσοστό αλληλουχιών που δεν στοιχίζονται στο γονιδίωμα του ποντικού (unmapped reads%) στα δεκαπέντε δείγματα. Επίσης, παρουσιάζεται ο NCBI GEO κωδικός του κάθε δείγματος (sample id), ο SRA κωδικός του (SRA id), ο χαρακτηρισμός του (tag, GF=germ-free,

ConvR=Conventionally raised, ConvD=conventionalized), ο αριθμός των αλληλουχιών μετά την προ-επεξεργασία/φιλτράρισμα (filtered sra reads) και ο αριθμός των αλληλουχιών που δεν στοιχίζονται στο γονιδίωμα του ποντικού (unmapped reads).

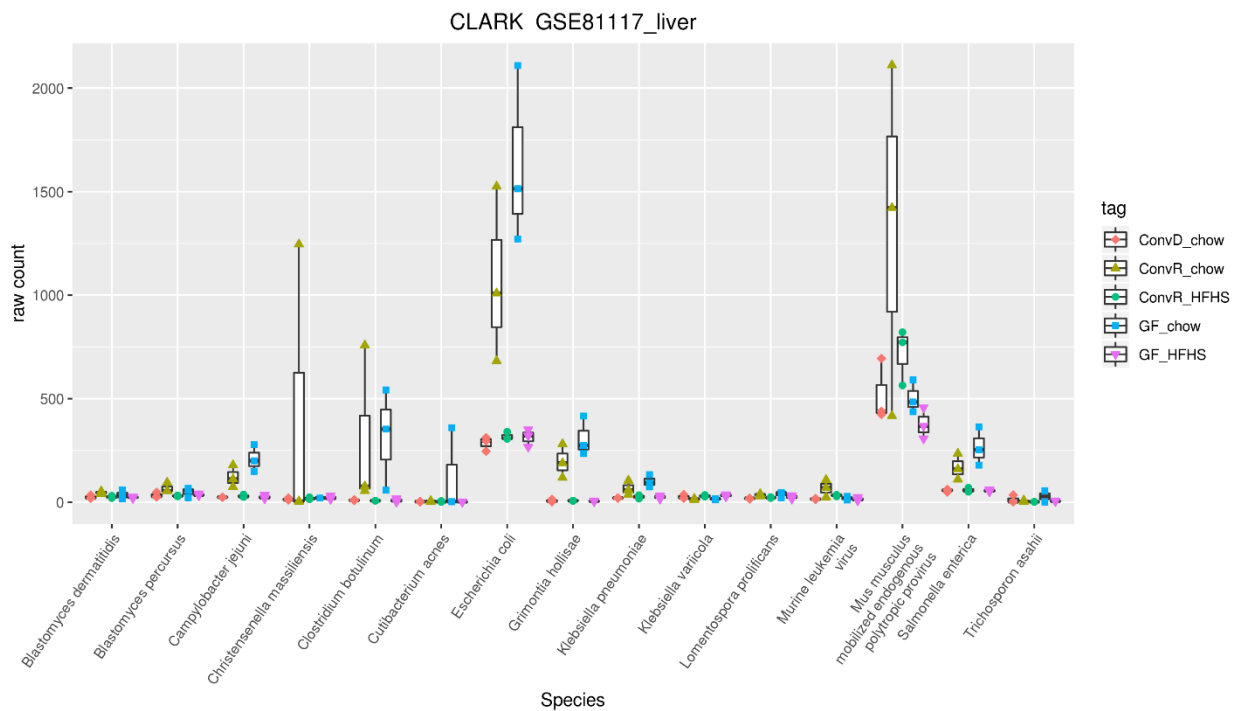
sample id	SRA id	tag	filtered sra reads	unmapped reads	unmapped reads%
GSM2143152	SRR3476414	GF_chow	16,916,618	399,017	2.36
GSM2143153	SRR3476415	GF_chow	22,462,738	296,859	1.32
GSM2143154	SRR3476416	GF_chow	27,282,915	478,082	1.75
GSM2143155	SRR3476417	ConvR_chow	24,550,114	460,182	1.87
GSM2143156	SRR3476418	ConvR_chow	22,204,225	345,385	1.56
GSM2143157	SRR3476419	ConvR_chow	21,179,805	286,230	1.35
GSM2143158	SRR3476420	ConvD_chow	23,350,910	185,711	0.8
GSM2143159	SRR3476421	ConvD_chow	27,995,429	212,452	0.76
GSM2143160	SRR3476422	ConvD_chow	22,507,975	193,154	0.86
GSM2143161	SRR3476423	GF_HFHS	26,134,447	230,603	0.88
GSM2143162	SRR3476424	GF_HFHS	25,618,974	194,250	0.76
GSM2143163	SRR3476425	GF_HFHS	23,015,664	189,272	0.82
GSM2143164	SRR3476426	ConvR_HFHS	27,078,884	200,465	0.74
GSM2143165	SRR3476427	ConvR_HFHS	23,462,430	152,896	0.65
GSM2143166	SRR3476428	ConvR_HFHS	26,534,890	240,577	0.91

3.3.2 Μεταγονιδιωματική Ανάλυση

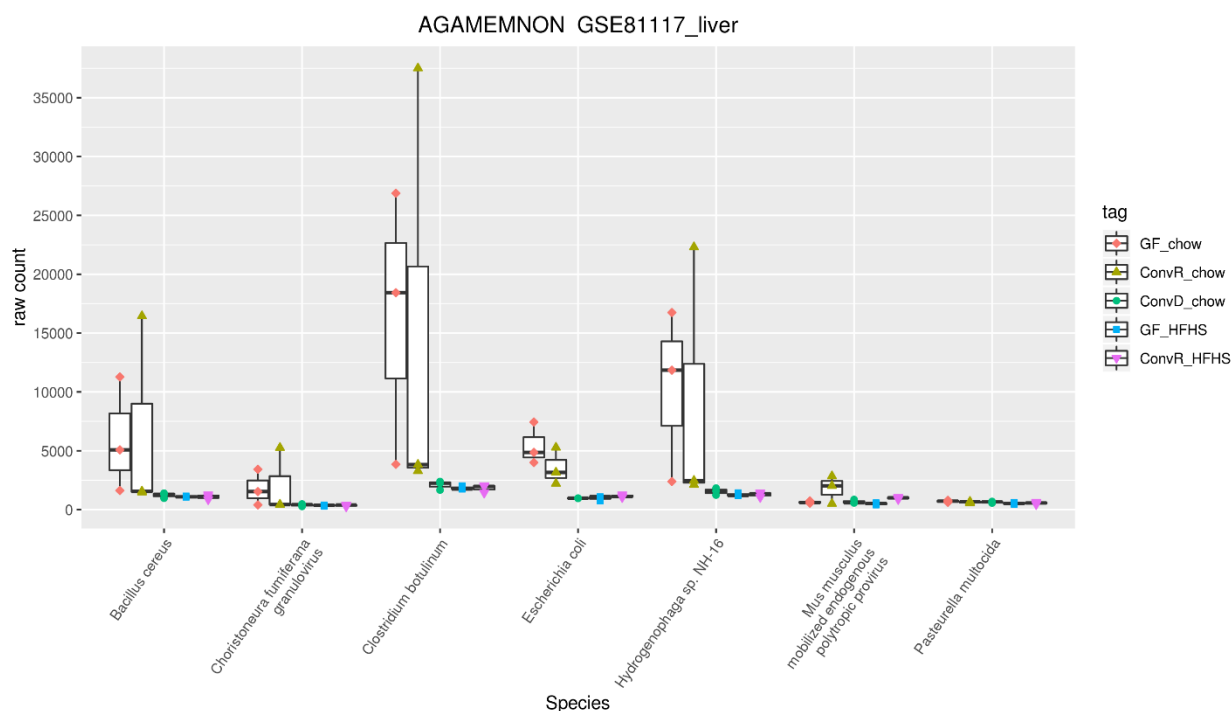
Με τη χρήση των τριών προγραμμάτων μεταγονιδιωματικής ανάλυσης διαπιστώνουμε ένδειξη ύπαρξης του μικροοργανισμού *E. coli* στα 3 δείγματα GF_chow ποντικών (βλ. **Εικόνες 3.17 – 3.19**). Η παρουσία άλλων μικροοργανισμών που παρατηρούμε στα δείγματα δεν επιβεβαιώνεται ταυτόχρονα και από τα τρία προγράμματα.



Εικόνα 3.17 Τα 6 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 15 δείγματα (17 διαφορετικά είδη) και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, σύμφωνα με το πρόγραμμα Kraken 2, για τη σειρά δειγμάτων με id: GSE81117. Κάθε χρώμα αντιστοιχεί σε διαφορετική κατηγορία δειγμάτων σύμφωνα με το υπόμνημα.



Εικόνα 3.18 Τα 6 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 15 δείγματα (15 διαφορετικά είδη) και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, σύμφωνα με το πρόγραμμα CLARK, για τη σειρά δειγμάτων με id: GSE81117. Κάθε χρώμα αντιστοιχεί σε διαφορετική κατηγορία δειγμάτων σύμφωνα με το υπόμνημα.



Εικόνα 3.19 Τα 6 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 15 δείγματα (7 διαφορετικά είδη) και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, σύμφωνα με το πρόγραμμα AGAMEMNON, για τη σειρά δειγμάτων με id: GSE81117. Κάθε χρώμα αντιστοιχεί σε διαφορετική κατηγορία δειγμάτων σύμφωνα με το υπόμνημα.

3.3.3 Στοιχισμός στο γονιδίωμα της *E. coli*

Στο σημείο αυτό, προχωρήσαμε σε στοιχισμό των ίδιων αλληλουχιών στο γονιδίωμα της *E. coli* για να προσδιορίσουμε ποια γονίδια πιθανώς εκφράζονται. Χρησιμοποιήθηκε το γονιδίωμα με κωδικό GCF_000005845.2 (RefSeq assembly accession) της *E. coli* str, K-12 substr. MG1655. Το ποσοστό των αλληλουχιών που στοιχήθηκαν στο γονιδίωμα ήταν από 11% έως 48% (βλ. Πίνακα 3.9). Τα έξι πιο συχνά γονίδια και στα δεκαπέντε δείγματα ήταν rRNA γονίδια (τα *rrsA*, *rrsB*, *rrsC*, *rrsD*, *rrsE*, *rrsG*, *rrsH* που κωδικοποιούν 16S rRNAs και τα *rrlC*, *rrlD*, *rrlG* που κωδικοποιούν 23S rRNAs), γεγονός αναμενόμενο καθότι το 80% των RNA ενός τυπικού κυττάρου είναι rRNA. (βλ. Εικόνα 3.20). Όσον αφορά τα μη rRNA γονίδια, στα έξι πιο συχνά κάθε δείματος συγκαταλέγονται τα *csrC*, *fusA*, *lacI*, *rimM*, *rnpB*, *rplA*, *rplB*, *rplF*, *rplO*, *rpoA*, *rpsD*, *rpsN*, *secY*, *ssrA*, *trmD* με προϊόντα που αναγράφονται στην εικόνα 3.21.

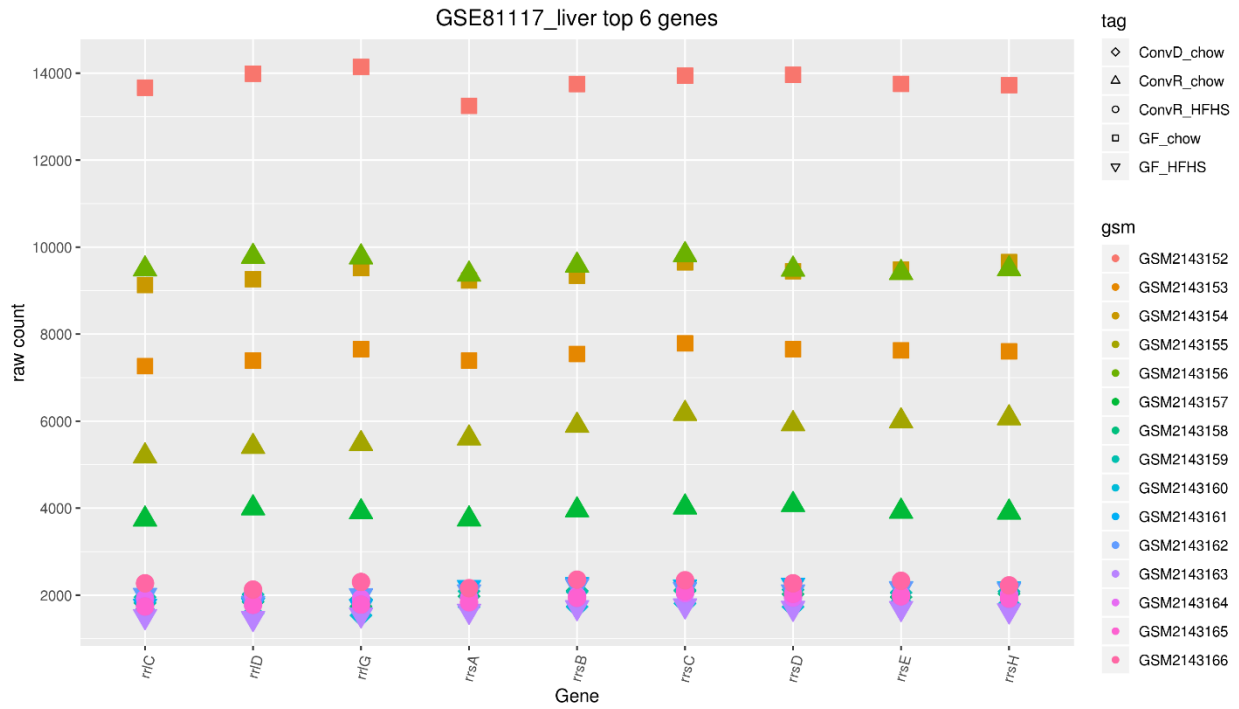
Παρατηρούμε και εδώ την ισχυρή παρουσία των δύο γονιδίων *ssrA*, *rnpB* όπως και στην πρώτη σειρά δειγμάτων. Η στοιχισμός των αλληλουχιών/reads σε αυτά τα γονίδια παρουσιάζεται στις Εικόνες 3.22, 3.23, όπου διαπιστώνουμε κάλυψη μεγάλου μέρους του μήκους κάθε γονιδίου από τα reads.

Η μόνη ένδειξη έκφρασης κάποιου σπερονίου φάνηκε για τα ριβοσωμικά γονίδια, όπως στην πρώτη σειρά δειγμάτων, και για τα γονίδια *trmD*, *rimM* (βλ. **Εικόνες 3.24, 3.25**).

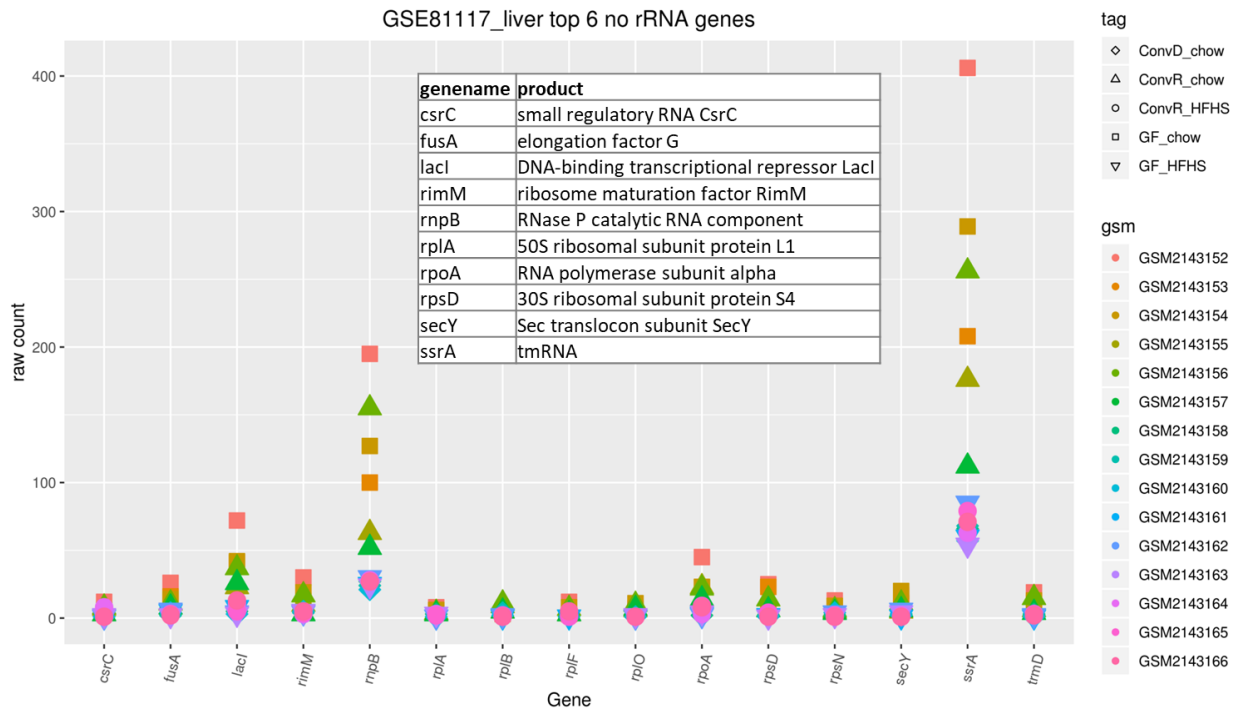
Πίνακας 3.9 Στατιστικά στοιχεία στο γονιδίωμα της *E. coli*. Αριθμός αλληλουχιών που δε στοιχήθηκαν στο γονιδίωμα του ποντικού (unmapped reads). Ποσοστό των unmapped reads που στοιχήθηκαν στο γονιδίωμα της *E. coli* (mapped to *E. coli* %). Ποσοστό των unmapped reads που στοιχήθηκαν σε γονίδια της *E. coli* (mapped to *E. coli* genes %). Αριθμός γονιδίων με τουλάχιστον μια αλληλουχία/read στην περιοχή τους.

sample id	SRA id	tag	unmapped reads	mapped to <i>E. coli</i> %	mapped to <i>E. coli</i> genes %	unique genes hit
GSM2143152	SRR3476414	GF_chow	399017	47.72	47.72	255
GSM2143153	SRR3476415	GF_chow	296859	34.68	34.68	194
GSM2143154	SRR3476416	GF_chow	478082	26.93	26.92	223
GSM2143155	SRR3476417	ConvR_chow	460182	16.81	16.81	165
GSM2143156	SRR3476418	ConvR_chow	345385	38.16	38.15	237
GSM2143157	SRR3476419	ConvR_chow	286230	18.66	18.66	147
GSM2143158	SRR3476420	ConvD_chow	185711	14.24	14.24	76
GSM2143159	SRR3476421	ConvD_chow	212452	13.16	13.15	95
GSM2143160	SRR3476422	ConvD_chow	193154	12.02	12.02	85
GSM2143161	SRR3476423	GF_HFHS	230603	11.85	11.85	97
GSM2143162	SRR3476424	GF_HFHS	194250	14.53	14.53	99
GSM2143163	SRR3476425	GF_HFHS	189272	11.71	11.71	73
GSM2143164	SRR3476426	ConvR_HFHS	200465	13.57	13.57	93
GSM2143165	SRR3476427	ConvR_HFHS	152896	16.94	16.94	87
GSM2143166	SRR3476428	ConvR_HFHS	240577	13.02	13.01	88

Εκτίμηση της μικροβιακής κατάστασης germ-free ποντικών χρησιμοποιώντας RNA-Seq δεδομένα

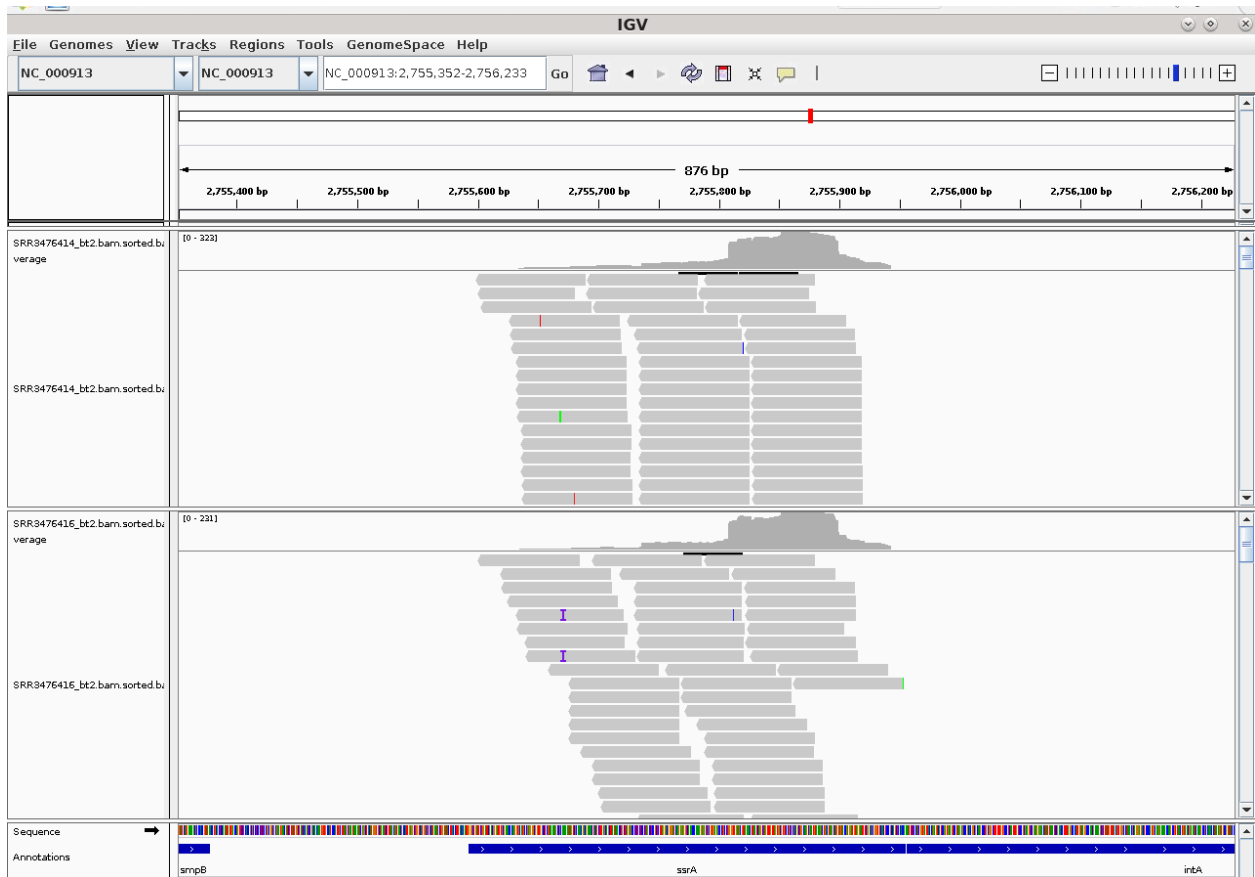


Εικόνα 3.20 Τα έξι πιο συχνά γονίδια σε κάθε ένα από τα δεκαπέντε δείγματα (εννέα διαφορετικά γονίδια) και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, για τη σειρά δειγμάτων με id: GSE81117. Κάθε χρώμα αντιστοιχεί σε διαφορετικό δείγμα και κάθε σχήμα σε διαφορετική κατηγορία ποντικών σύμφωνα με το υπόμνημα. Και τα εννέα γονίδια είναι rRNA γονίδια.



Εικόνα 3.21 Τα έξι πιο συχνά γονίδια, εξαιρουμένων των rRNA γονιδίων, σε κάθε ένα από τα δεκαπέντε δείγματα (δέκαπέντε διαφορετικά γονίδια) και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα, για τη σειρά δειγμάτων με id: GSE81117. Κάθε χρώμα αντιστοιχεί σε διαφορετικό

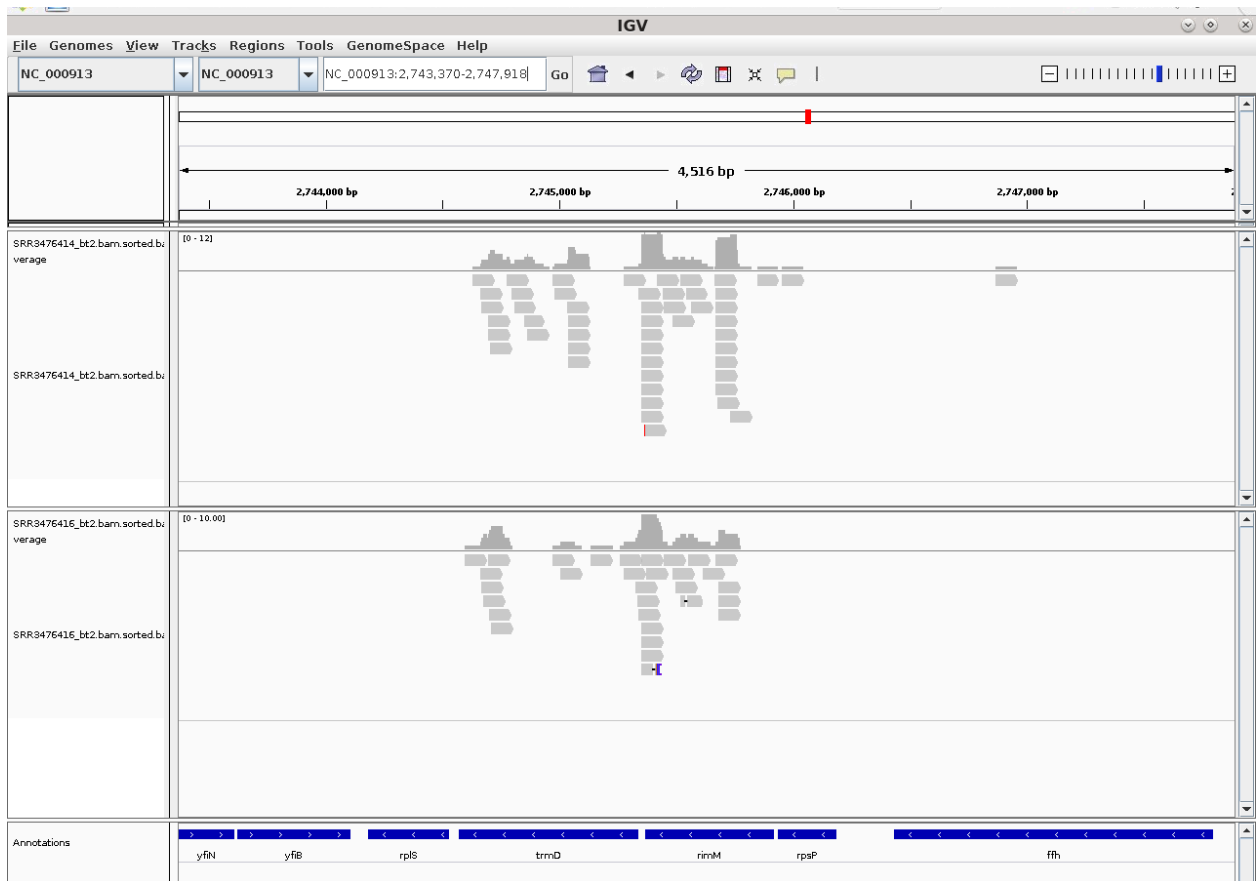
Δείγμα και κάθε σχήμα σε διαφορετική κατηγορία ποντικών σύμφωνα με το υπόμνημα. Στον ένθετο πίνακα αναγράφεται το προϊόν (product) κάθε γονιδίου.



Εικόνα 3.22 Ένα παράδειγμα της στοίχισης αλληλουχιών του πρώτου και του τρίτου δείγματος στο γονιδίωμα της *E. coli*, και συγκεκριμένα στο γονίδιο *ssrA*, το οποίο ήταν το δεύτερο πιο συχνό σε όλα τα δείγματα. Η ισχυρή παρουσία του γονιδίου αυτού οφείλεται σε πολλά αντίγραφα αλληλουχιών/reads που καλύπτουν όλη την έκταση του γονιδίου, μήκους 362 ζευγών βάσεων.



Εικόνα 3.23 Ένα παράδειγμα της στοίχισης αλληλουχιών του πρώτου και του τρίτου δείγματος στο γονιδίωμα της *E. coli*, και συγκεκριμένα στο γονίδιο rnpB, το οποίο ήταν το πιο συχνό σε όλα τα δείγματα. Η ισχυρή παρουσία του γονιδίου αυτού οφείλεται σε πολλά αντίγραφα αλληλουχιών/reads που καλύπτουν όλη την έκταση του γονιδίου.



Εικόνα 3.24 Ένα παράδειγμα της στοίχισης αλληλουχιών του πρώτου και του τρίτου δείγματος στο γονιδίωμα της *E. coli*, και συγκεκριμένα στο γονίδιο *trmD*. Η παρουσία του γονιδίου αυτού οφείλεται σε λίγα αντίγραφα αλληλουχιών/reads που καλύπτουν μεγάλο μέρος της έκτασης του γονιδίου. Αλληλουχίες/reads εντοπίζονται και στο γειτονικό γονίδιο *rimM*. Και τα δύο γονίδια βρίσκονται στο ίδιο οπερόνιο (βλ. Εικόνα 3.25).



Εικόνα 3.25 Οπερόνιο των γονιδίων *rpsP*, *rim*, *trmD*, *rplS*, στο γονιδίωμα της *E. coli* [78].

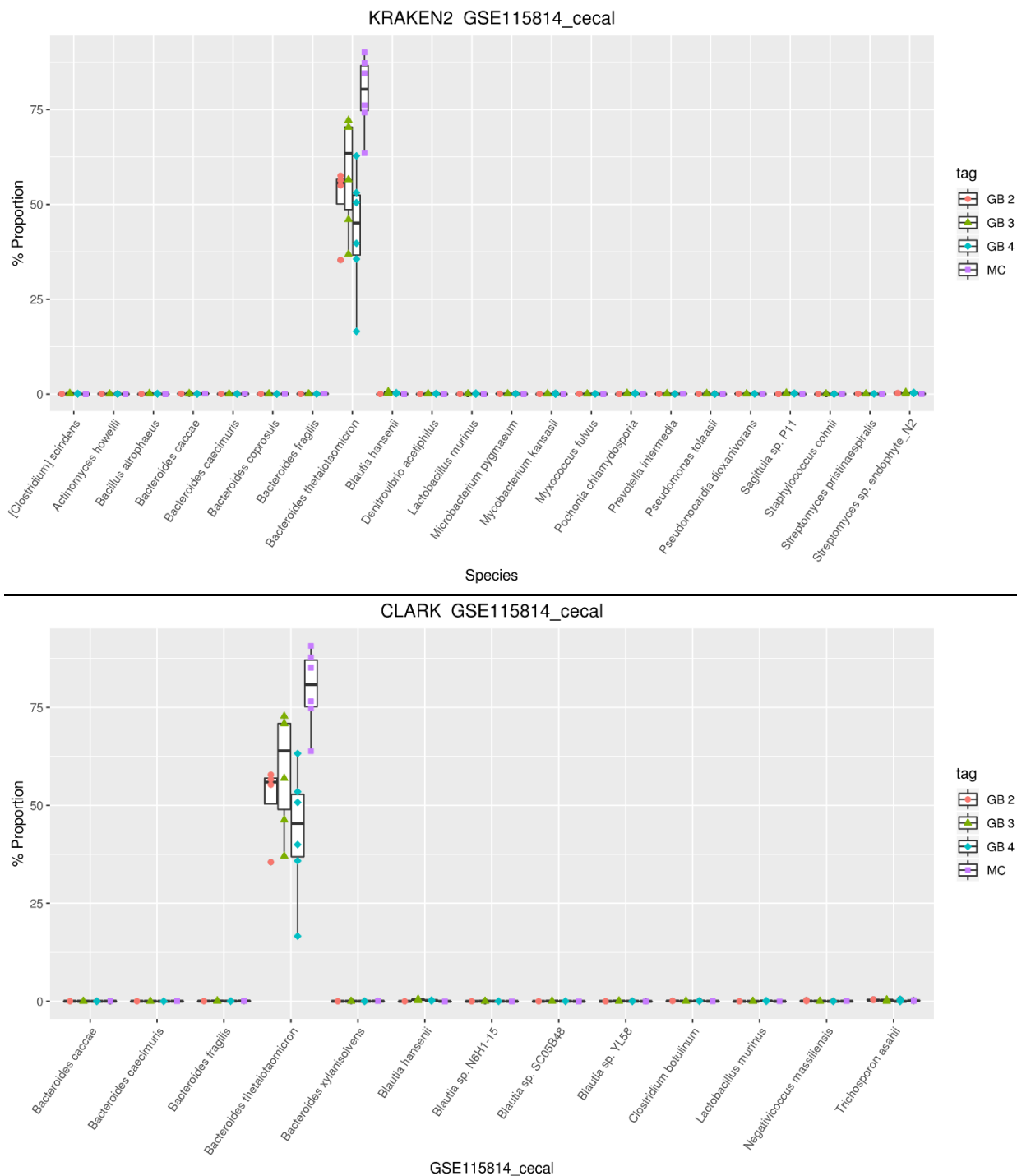
3.4 Ανάλυση άλλων δειγμάτων

3.4.1 Σειρά δειγμάτων GSE115814

Το πείραμα από το οποίο προέρχονται τα δείγματα αυτά στόχευε στην αλληλούχιση των μικροβιακών RNA σε δείγματα περιεχομένων του εντέρου, 22 gnotobiotic ποντικών επικοισμένων με το βακτήριο *Bacteroides thetaiotaomicron*. Από αυτά τα ποντίκια, τα 16 ήταν επικοισμένα και με άλλους μικροοργανισμούς, όπως περιγράφεται στον **Πίνακα 3.10**. Με τα προγράμματα Kraken 2 και CLARK επιβεβαιώσαμε την ύπαρξη του *Bacteroides thetaiotaomicron* στα δείγματα (βλ. **Εικόνα 3.26**).

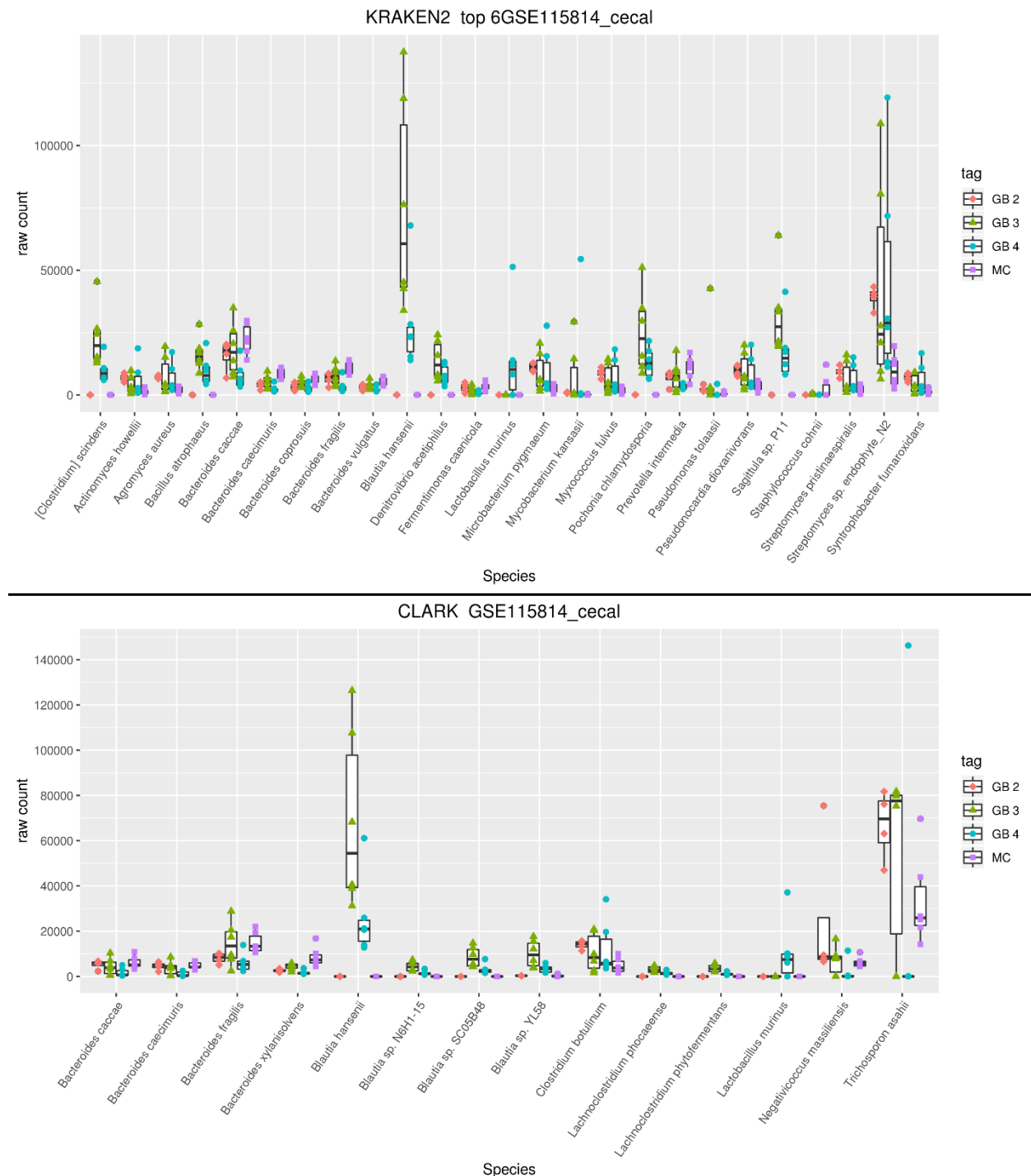
Πίνακας 3.10 Αριθμός των αλληλουχιών μετά την προ-επεξεργασία/φιλτράρισμα των δειγμάτων (filtered sra pairs). Επίσης, παρουσιάζεται ο NCBI GEO κωδικός του κάθε δείγματος (sample id), ο SRA κωδικός του (SRA id) και ο χαρακτηρισμός του (tag, MC=επικοισμένο με *Bacteroides thetaiotaomicron*, GB 2= επικοισμένο με *Bacteroides thetaiotaomicron* και *Candidatus Arthromitus* sp. SFB-mouse, GB 3= επικοισμένο με *Bacteroides thetaiotaomicron*, *Candidatus Arthromitus* sp. SFB-mouse και *Clostridium* species, GB 4= επικοισμένο με *Bacteroides thetaiotaomicron*, *Candidatus Arthromitus* sp. SFB-mouse, *Clostridium* species και *Lactobacillus* species).

sample id	SRA id	tag	filtered sra reads
GSM3190379	SRR7347422	MC	22552057
GSM3190380	SRR7347423	MC	14203337
GSM3190381	SRR7347424	MC	16994775
GSM3190382	SRR7347425	MC	26298450
GSM3190383	SRR7347426	MC	14593482
GSM3190384	SRR7347427	MC	22256096
GSM3190385	SRR7347428	GB 2	23182333
GSM3190386	SRR7347429	GB 2	15034281
GSM3190387	SRR7347430	GB 2	21481098
GSM3190388	SRR7347431	GB 2	13636197
GSM3190389	SRR7347432	GB 3	8491072
GSM3190390	SRR7347433	GB 3	25620044
GSM3190391	SRR7347434	GB 3	19188486
GSM3190392	SRR7347435	GB 3	8399498
GSM3190393	SRR7347436	GB 3	22486246
GSM3190394	SRR7347437	GB 3	23186566
GSM3190395	SRR7347438	GB 4	8207115
GSM3190396	SRR7347439	GB 4	8222678
GSM3190397	SRR7347440	GB 4	27460989
GSM3190398	SRR7347441	GB 4	8125303
GSM3190399	SRR7347442	GB 4	7998088
GSM3190400	SRR7347443	GB 4	22892582



Εικόνα 3.26 Αποτελέσματα των προγραμμάτων Kraken 2 και CLARK για τη σειρά δειγμάτων GSE115814. Διαπιστώνεται η ύπαρξη του *Bacteroides thetaiotaomicron*. Το ποσοστό των reads (% Proportion) που αντιστοιχούν στο κάθε είδος μικροοργανισμού υπολογίστηκε ως προς τον αριθμό των reads που υπήρχαν σε κάθε δείγμα μετά την προ-επεξεργασία/φιλτράρισμα (βλ. Πίνακα 3.10). Κάθε χρώμα αντιστοιχεί σε διαφορετική κατηγορία δειγμάτων

Παρόλο που οι αλληλουχίες του *Bacteroides thetaiotaomicron* αποτελούν το μεγαλύτερο ποσοστό των αλληλουχιών στα δείγματα, εντοπίζουμε σε αυτά και αλληλουχίες άλλων μικροοργανισμών (βλ. **Εικόνα 3.27**). Δεν εντοπίζουμε αλληλουχίες του *Candidatus Arthromitus* sp. SFB-mouse στις κατηγορίες δειγμάτων GB 2, GB 3 και GB 4 όπως θα αναμέναμε. Εντοπίζουμε αλληλουχίες από είδος του γένους *Clostridium* μόνο με το πρόγραμμα CLARK στα δείγματα GB 3, GB 4 όπως αναμέναμε, αλλά και στα δείγματα MC, GB που δεν αναμέναμε. Τέλος, εντοπίζουμε αλληλουχίες από είδος του γένους *Lactobacillus* κυρίως στα δείγματα GB 4, όπως αναμέναμε. Η ύπαρξη αλληλουχιών του μικροοργανισμού *Blautia hansenii* στα δείγματα των κατηγοριών GB 3 και GB 4 επιβεβαιώνεται και από τα δύο προγράμματα ενώ δεν αναμενόταν η ύπαρξή του σε κανένα δείγμα.



Εικόνα 3.27 Αποτελέσματα των προγραμμάτων Kraken 2 και CLARK για τη σειρά δειγμάτων GSE115814, όπου φαίνονται τα 6 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 22 δείγματα. Διαπιστώνεται η ύπαρξη κάποιων μη αναμενόμενων μικροοργανισμών (*Blautia hanseni*) με πολλές αλληλουχίες/reads και με τα δύο προγράμματα.

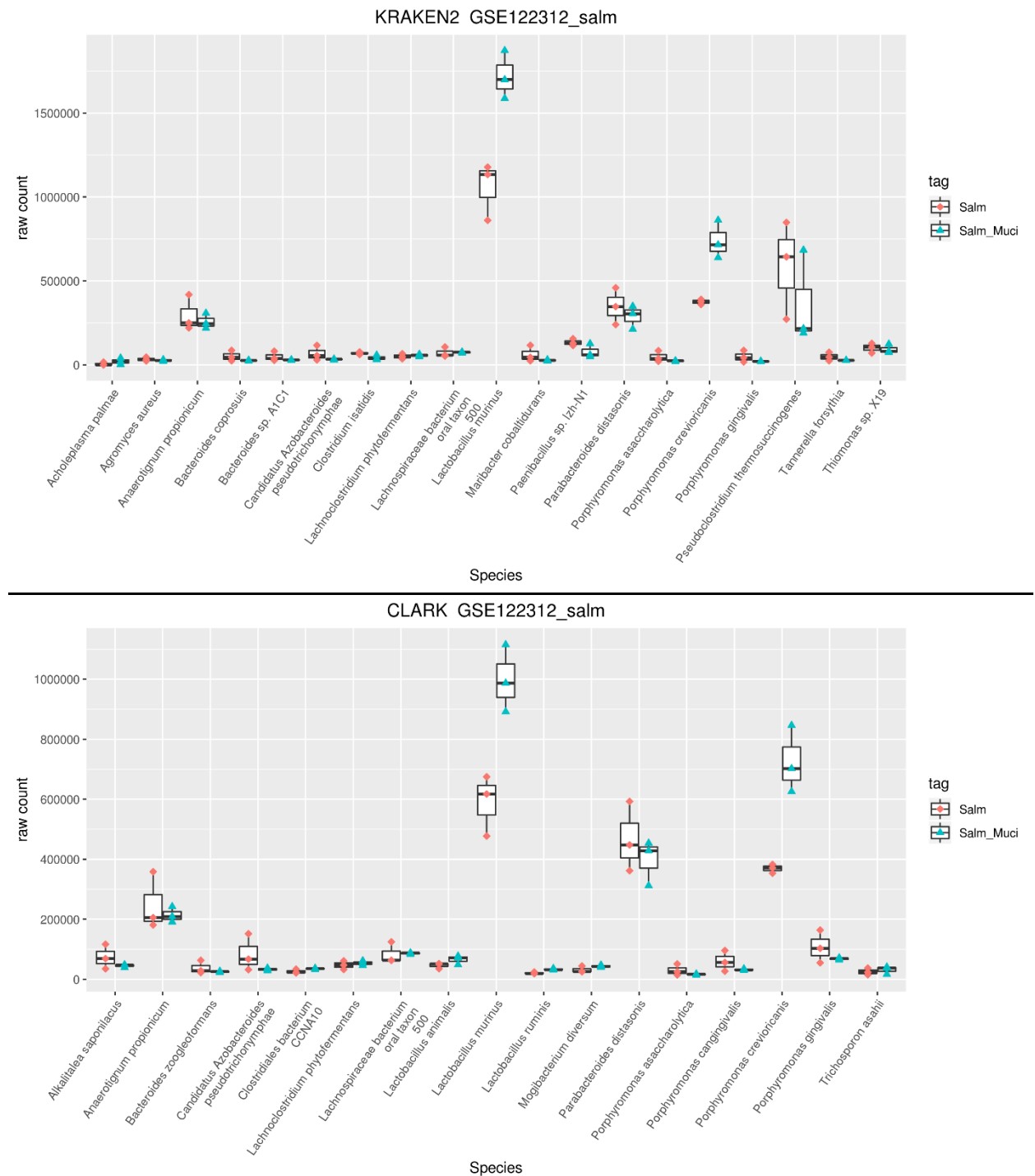
3.4.2 Σειρά δειγμάτων GSE122312

Η σειρά δειγμάτων με id GSE122312 προέρχεται από μια μελέτη της έκφρασης των γονιδίων του βακτηρίου *Salmonella Typhimurium* παρουσία και απουσία του βακτηρίου *Mucispirillum schaedleri* ASF 45, στο έντερο gnotobiotic ποντικών. Τρία δείγματα προέρχονται από τρία ποντίκια εποικισμένα μόνο με *Salmonella Typhimurium* (Salm) και τρία δείγματα από τρία ποντίκια εποικισμένα με *Salmonella Typhimurium* και *Mucispirillum schaedleri* ASF 45 (Salm_Muci) (βλ. **Πίνακα 3.11**).

Πίνακας 3.11 Αριθμός των αλληλουχιών μετά την προ-επεξεργασία/φιλτράρισμα των δειγμάτων (filtered sra pairs). Επίσης, παρουσιάζεται ο NCBI GEO κωδικός του κάθε δείγματος (sample id), οι SRA κωδικοί του (SRA id) και ο χαρακτηρισμός του (tag).

sample id	SRA id	tag	filtered sra reads
GSM3463835	SRR8175246	Salm	11036584
GSM3463836	SRR8175247	Salm	12759838
GSM3463837	SRR8175248	Salm	10074378
GSM3463838	SRR8175249	Salm_Muci	10901772
GSM3463839	SRR8175250	Salm_Muci	10295007
GSM3463840	SRR8175251	Salm_Muci	8767109

Μετά από την ανάλυση με τα προγράμματα Kraken 2 και CLARK παρατηρούμε ότι τα δύο προαναφερόμενα είδη μικροοργανισμών δεν συναντώνται στα είδη με τη μεγαλύτερη «αφθονία» (βλ. **Εικόνα 3.28**). Αντίθετα, εντοπίζουμε αρκετά μικροβιακά taxa με πολλές αλληλουχίες/reads, η προέλευση των οποίων θα πρέπει να διερευνηθεί.



Εικόνα 3.28 Αποτελέσματα των προγραμμάτων Kraken 2 και CLARK για τη σειρά δειγμάτων GSE122312, όπου φαίνονται τα 12 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 6 δείγματα. Διαπιστώνεται η ύπαρξη κάποιων μη αναμενόμενων μικροοργανισμών με πολλές αλληλουχίες/reads.

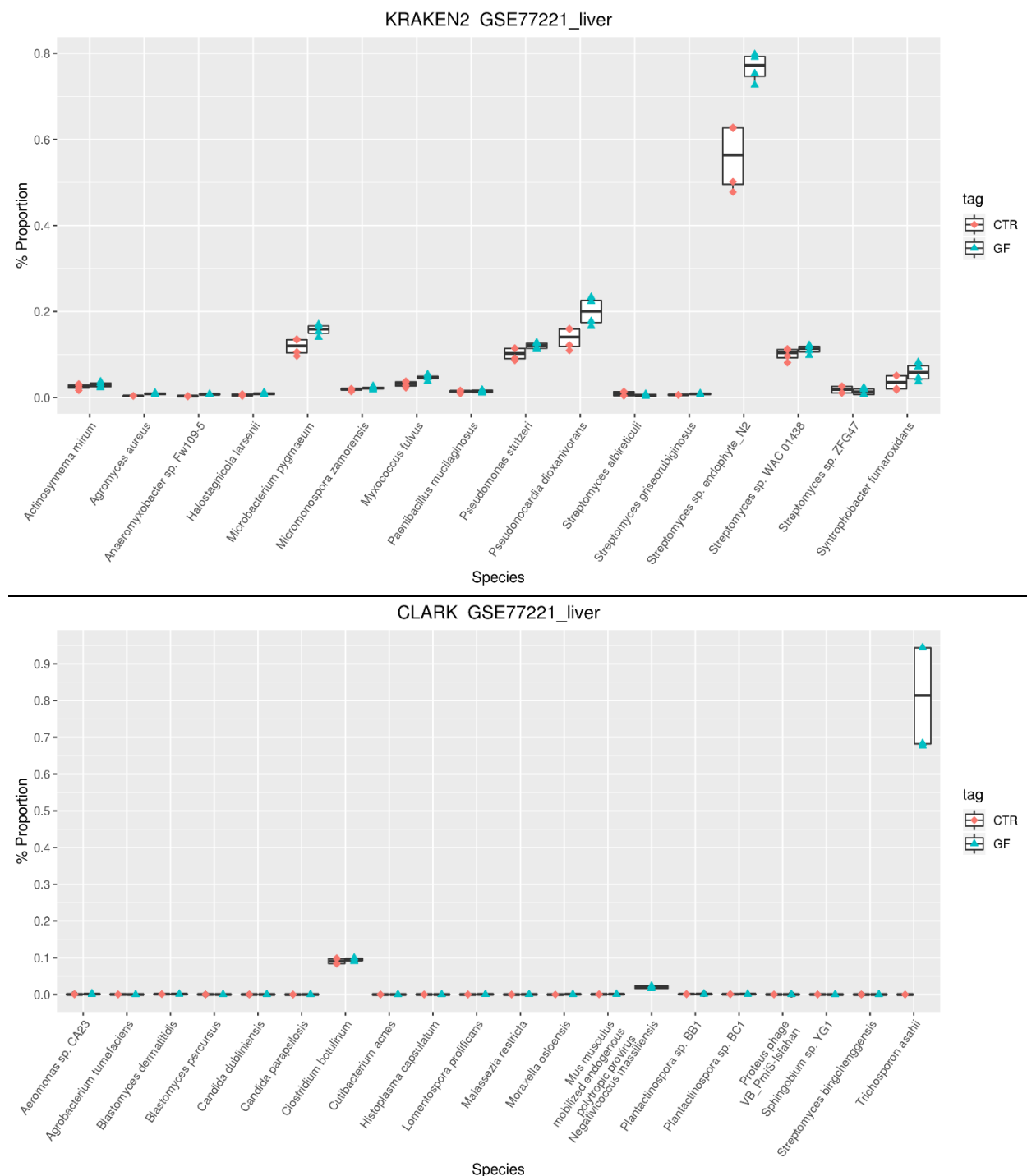
3.4.3 Σειρά δειγμάτων GSE77221

Από τη σειρά δειγμάτων με id GSE77221, επιλέξαμε δύο δείγματα από το συκώτι δύο ποντικών με φυσιολογική μικροχλωρίδα (CTR) και δύο δείγματα από το συκώτι GF ποντικών (GF) (βλ. **Πίνακα 3.12**).

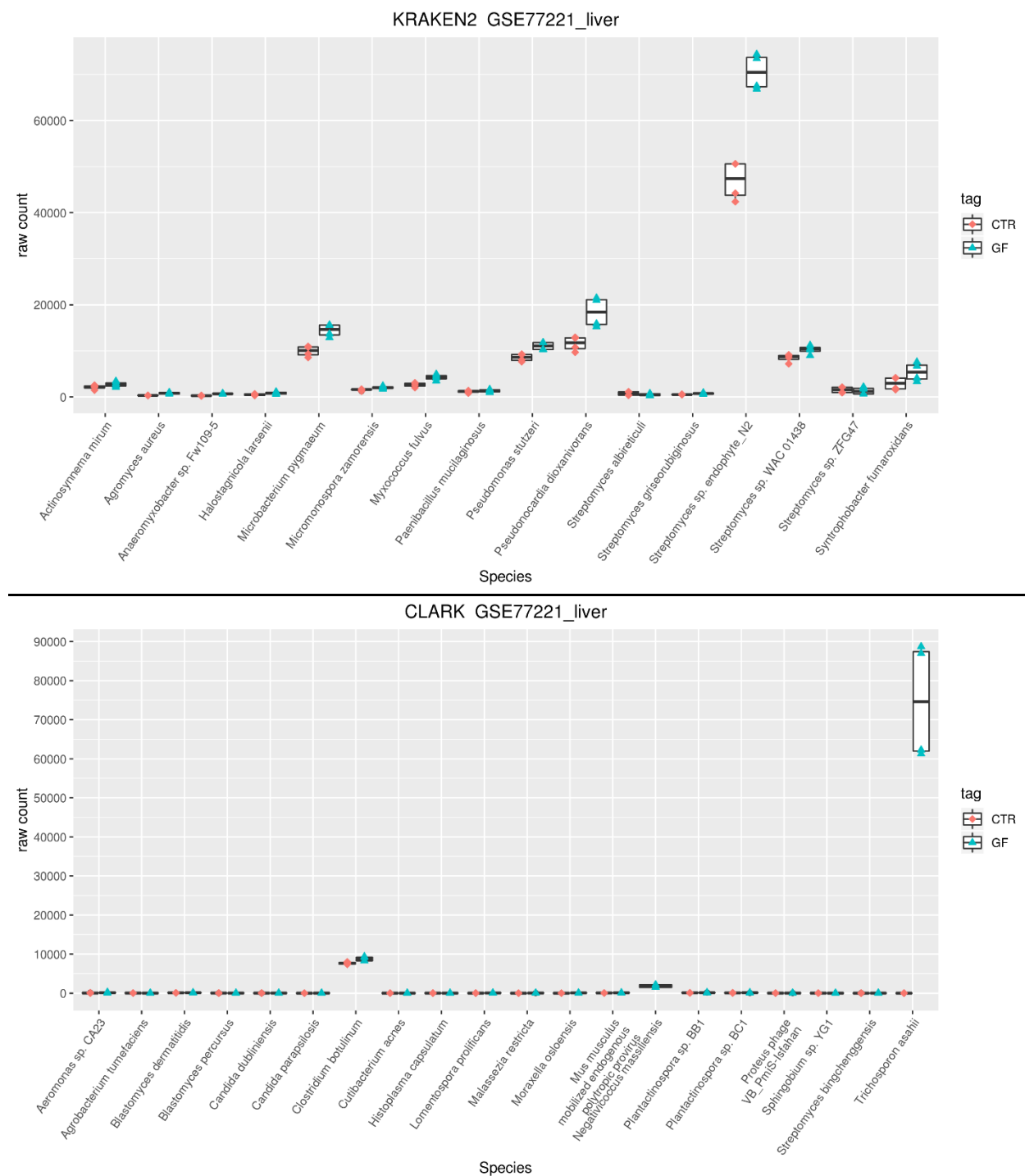
Πίνακας 3.12 Αριθμός των αλληλουχιών μετά την προ-επεξεργασία/φιλτράρισμα των δειγμάτων (filtered sra pairs) της σειράς GSE77221. Επίσης, παρουσιάζεται ο NCBI GEO κωδικός του κάθε δείγματος (sample id), οι SRA κωδικοί του (SRA id) και ο χαρακτηρισμός του (tag).

sample id	SRA id	tag	filtered sra reads
GSM2046160	SRR7161696	CTR	8870311
GSM2046160	SRR7161695	CTR	8813313
GSM2046184	SRR7161739	CTR	8076726
GSM2046184	SRR7161738	CTR	8060474
GSM2046207	SRR7161792	GF	9202432
GSM2046207	SRR7161791	GF	8955586
GSM2046228	SRR7161835	GF	9207817
GSM2046228	SRR7161833	GF	9403848

Με τη μεταγονιδιωματική ανάλυση των αλληλουχιών που δεν στοιχίζονται στο γονιδίωμα του ποντικού, εντοπίζουμε κάποιες αλληλουχίες που αποδίδονται σε μικρόβια, και στην κατηγορία δειγμάτων CTR και στην κατηγορία GF. Ενδιαφέρον έχει ότι και τα δύο προγράμματα που χρησιμοποιήθηκαν, Kraken 2 και CLARK, απέδωσαν περισσότερες αλληλουχίες σε μικρόβια στα δείγματα GF απ' ότι στα δείγματα CTR. Ωστόσο, στο ποια είναι αυτά τα μικρόβια τα δύο προγράμματα δε συμφωνούν (βλ. **Εικόνες 3.29, 3.30**).



Εικόνα 3.29 Αποτελέσματα των προγραμμάτων Kraken 2 και CLARK για τη σειρά δειγμάτων GSE77221. Έχουν επιλεγεί μόνο οι 12 πιο συχνόι μικροοργανισμοί κάθε δείγματος. Το ποσοστό των reads (% Proportion) που αντιστοιχούν στο κάθε είδος μικροοργανισμού υπολογίστηκε ως προς τον αριθμό των reads που υπήρχαν σε κάθε δείγμα μετά την προ-επεξεργασία/φιλτράρισμα (βλ. Πίνακα 3.12). Μεγαλύτερο ποσοστό αλληλουχιών αποδίδεται σε μικρόβια στα δείγματα GF απ’ ότι στα δείγματα CTR.



Εικόνα 3.30 Αποτελέσματα των προγραμμάτων Kraken 2 και CLARK για τη σειρά δειγμάτων GSE77221, όπου φαίνονται τα 12 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 4 δείγματα και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα.

3.4.4 Σειρές δειγμάτων GSE108718 και GSE133809

Συνοπτικά παρουσιάζουμε τη μελέτη δύο σειρών RNA-Seq δειγμάτων με μικρό μήκος αλληλουχιών/reads.

- **GSE108718**

Σε αυτή τη μελέτη έχουμε 14 δείγματα RNA-Seq από κύτταρα του δέρματος σε:

- ❖ 6 GF ποντίκια (GF)
- ❖ 3 ποντίκια εποικισμένα μόνο με το βακτήριο *Staphylococcus aureus* (MC)
- ❖ 3 ποντίκια με φυσιολογική μικροχλωρίδα (Control)
- ❖ 2 ποντίκια με φυσιολογική μικροχλωρίδα αλλά με δίαιτα χωρίς βιταμίνη A (VAD)

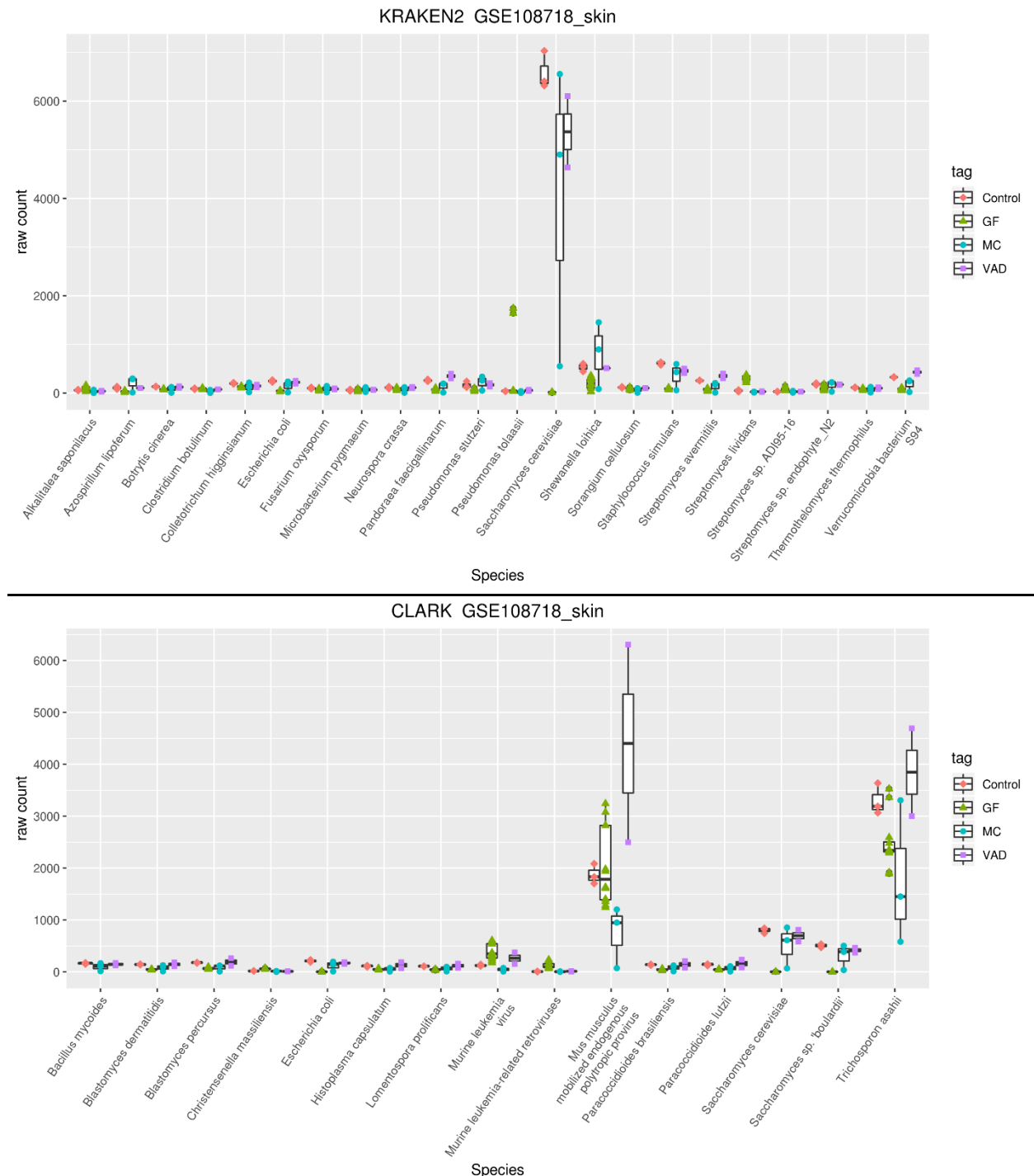
Επειδή το αρχικό μήκος των αλληλουχιών/reads ήταν 51 βάσεις, η προ-επεξεργασία έγινε όπως αναφέρεται στην **ενότητα 2.4**, αλλά τέθηκε ελάχιστο τελικό μήκος αλληλουχιών ίσο με 50 βάσεις.

Πίνακας 3.13 Αριθμός των αλληλουχιών μετά την προ-επεξεργασία/φιλτράρισμα των δειγμάτων (filtered sra pairs) της σειράς GSE108718. Επίσης, παρουσιάζεται ο NCBI GEO κωδικός του κάθε δείγματος (sample id), οι SRA κωδικοί του (SRA id) και ο χαρακτηρισμός του (tag).

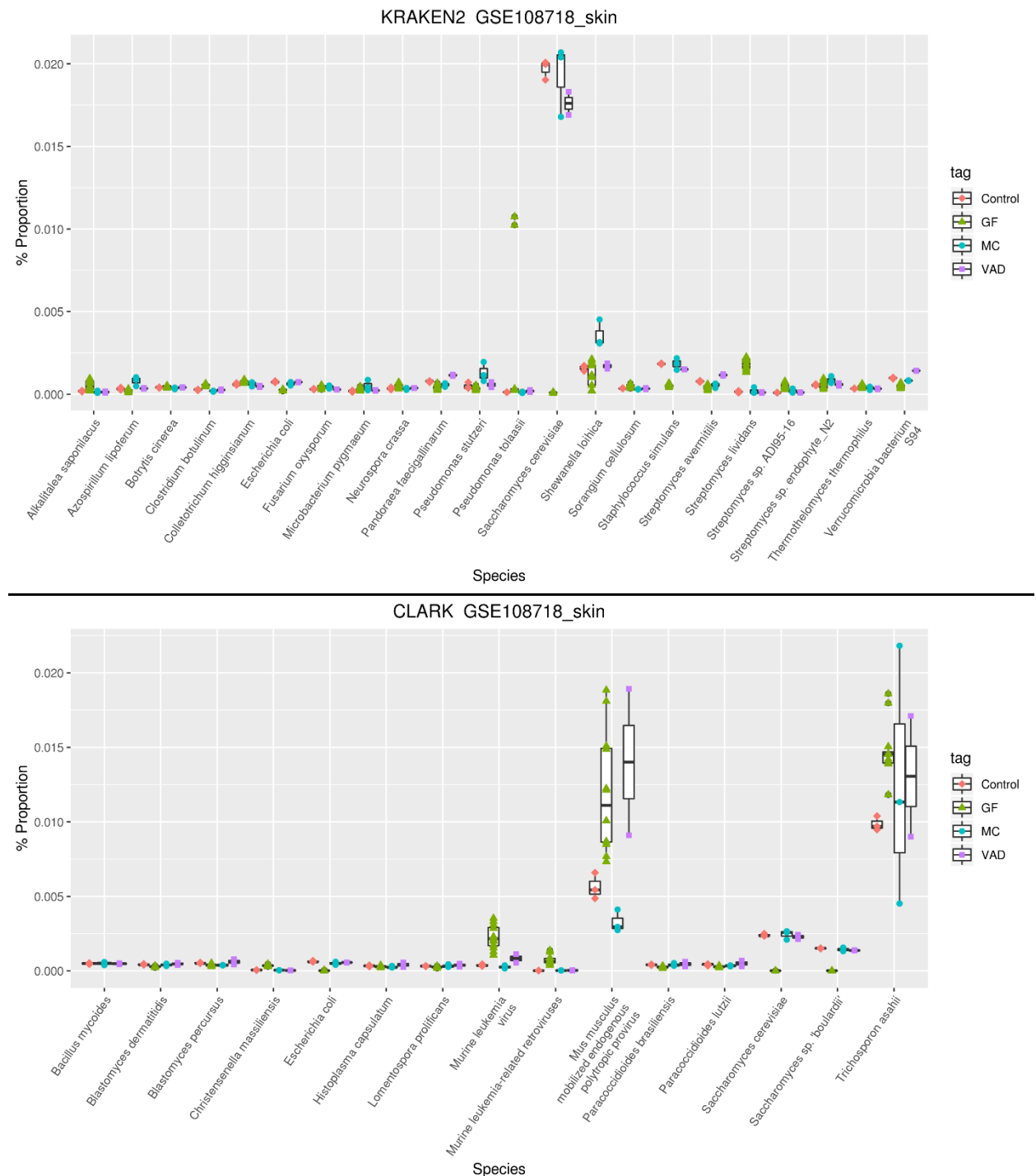
sample id	SRA id	tag	filtered sra reads
GSM2912489	SRR6493398	GF	16219361
GSM2912489	SRR6493399	GF	16007116
GSM2912490	SRR6493400	GF	17215586
GSM2912490	SRR6493401	GF	16998586
GSM2912491	SRR6493402	GF	18979996
GSM2912491	SRR6493403	GF	18720598
GSM2912492	SRR6493404	GF	16183990
GSM2912492	SRR6493405	GF	15981790
GSM2912493	SRR6493406	GF	16929701
GSM2912493	SRR6493407	GF	16699641
GSM2912494	SRR6493408	GF	16161322
GSM2912494	SRR6493409	GF	15949528
GSM2912495	SRR6493410	MC	2667549
GSM2912496	SRR6493411	MC	29193033
GSM2912497	SRR6493412	MC	32138873
GSM2912498	SRR6493413	Control	31636830
GSM2912499	SRR6493414	Control	35013583
GSM2912500	SRR6493415	Control	33656881
GSM2912501	SRR6493416	VAD	27431121
GSM2912502	SRR6493417	VAD	33328655

Εκτίμηση της μικροβιακής κατάστασης germ-free ποντικών χρησιμοποιώντας RNA-Seq δεδομένα

Και με τα δύο προγράμματα μεταγονιδιωματικής ανάλυσης που χρησιμοποιήθηκαν εντοπίσαμε αλληλουχίες που μπορούν να αποδοθούν σε μικροοργανισμούς, και στις τέσσερις κατηγορίες δειγμάτων (βλ. **Εικόνες 3.31, 3.32**). Τα δύο προγράμματα, ωστόσο, δε συμφωνούν στο ποιοι είναι οι μικροοργανισμοί αυτοί.



Εικόνα 3.31 Αποτελέσματα των προγραμμάτων Kraken 2 και CLARK για τη σειρά δειγμάτων GSE108718, όπου φαίνονται τα 6 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 14 δείγματα και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα.



Εικόνα 3.32 Αποτελέσματα των προγραμμάτων Kraken 2 και CLARK για τη σειρά δειγμάτων GSE108718. Έχουν επιλεγεί μόνο οι 6 πιο συχνόι μικροοργανισμοί κάθε δείγματος. Το ποσοστό των reads (% Proportion) που αντιστοιχούν στο κάθε είδος μικροοργανισμού υπολογίστηκε ως προς τον αριθμό των reads που υπήρχαν σε κάθε δείγμα μετά την προ-επεξεργασία/φιλτράρισμα (βλ. Πίνακα 3.13)

- **GSE133809**

Σε αυτή τη μελέτη έχουμε 14 δείγματα RNA-Seq από ιστό κόλον εντέρου:

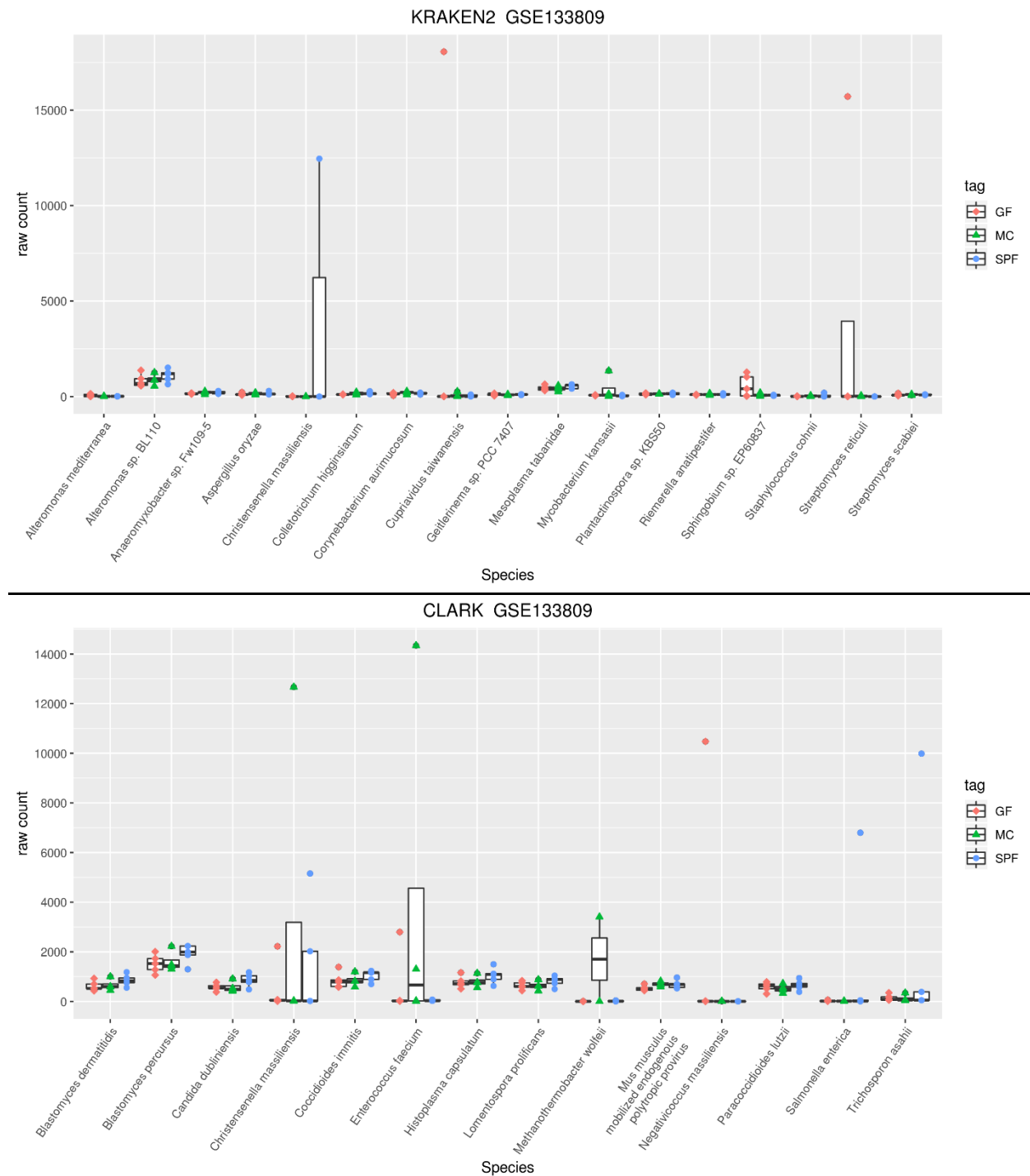
- ❖ 5 GF ποντικών (GF)
- ❖ 4 ποντικών επικοισμένων μόνο με το βακτήριο *Turicibacter sanguinis* MOL361 (MC)
- ❖ 5 ποντικών με φυσιολογική μικροχλωρίδα, δίχως γνωστούς παθογόνους μικροοργανισμούς (SPF) (βλ. **Πίνακα 3.14**).

Επειδή το αρχικό μήκος των αλληλουχιών/reads ήταν 65 βάσεις, η προ-επεξεργασία έγινε όπως αναφέρεται στην **ενότητα 2.4**, αλλά τέθηκε ελάχιστο τελικό μήκος αλληλουχιών ίσο με 50 βάσεις.

Πίνακας 3.14 Αριθμός των αλληλουχιών μετά την προ-επεξεργασία/φιλτράρισμα των δειγμάτων (filtered sra pairs) της σειράς GSE133809. Επίσης, παρουσιάζεται ο NCBI GEO κωδικός του κάθε δείγματος (sample id), οι SRA κωδικοί του (SRA id) και ο χαρακτηρισμός του (tag).

sample id	SRA id	tag	filtered sra reads
GSM3927586	SRR9641556	GF	9557305
GSM3927587	SRR9641557	GF	10028685
GSM3927588	SRR9641558	GF	10685201
GSM3927589	SRR9641559	GF	11065365
GSM3927590	SRR9641560	GF	11372087
GSM3927591	SRR9641561	MC	10308648
GSM3927592	SRR9641562	MC	8395305
GSM3927593	SRR9641563	MC	11676434
GSM3927594	SRR9641564	MC	9457414
GSM3927595	SRR9641565	SPF	10102219
GSM3927596	SRR9641566	SPF	10868830
GSM3927597	SRR9641567	SPF	9886983
GSM3927598	SRR9641568	SPF	10574655
GSM3927599	SRR9641569	SPF	8845627

Ύστερα από την ανάλυση των δειγμάτων με τα προγράμματα Kraken 2 και CLARK, δεν εντοπίσαμε τον μικροοργανισμό *Turicibacter sanguinis* στην κατηγορία δειγμάτων MC, όπως αναμενόταν. Η παρουσία άλλων μικροοργανισμών που παρατηρούμε σε μεμονωμένα δείγματα δεν επιβεβαιώνεται ταυτόχρονα και από τα δύο προγράμματα, με εξαίρεση το βακτήριο *Christensenella massiliensis* (βλ. **Εικόνα 3.33**).



Εικόνα 3.33 Αποτελέσματα των προγραμμάτων Kraken 2 και CLARK για τη σειρά δειγμάτων GSE133809, όπου φαίνονται τα 6 πιο συχνά είδη μικροοργανισμών σε κάθε ένα από τα 14 δείγματα και ο αριθμός των αλληλουχιών/reads τους στο κάθε δείγμα.

4. Σύνοψη - Συζήτηση

Στην παρούσα εργασία εξετάζεται η δυνατότητα χρήσης RNA-Seq δεδομένων γονιδιακής έκφρασης (expression profiling with high throughput sequencing) για την ποιοτική εκτίμηση της μικροβιακής κατάστασης gnotobiotic και germ-free ποντικών και περιγράφεται η μεθοδολογία που μπορεί να ακολουθηθεί, χρησιμοποιώντας 3 εργαλεία μεταγονιδιωματικής ανάλυσης, τα Kraken 2, CLARK και AGAMEMNON.

Διαπιστώσαμε, στην περίπτωση μοναδικά εποικισμένων ποντικών (monocolonized), την ύπαρξη αλληλουχιών από το μικροβιακό οργανισμό στα RNA-Seq δεδομένα γονιδιακής έκφρασης, παρόλο που τα δεδομένα αυτά προέρχονται από πειράματα που δεν στοχεύουν στην αλληλούχιση μικροβίων (βλ. **Ενότητες 3.1, 3.2**).

Η παρουσία μικροβιακών αλληλουχιών σε δείγματα GF ποντικών στα ίδια επίπεδα όπως σε control ποντίκια προβληματίζει για την στειρότητα των ποντικών αυτών και απαιτεί περαιτέρω διερεύνηση (βλ. **Ενότητες 3.3, 3.4.3**).

Σε δείγματα μεταγονιδιωματικής, από περιεχόμενα εντέρου (cecal content), όπου ο σκοπός των ερευνητών ήταν η μελέτη της έκφρασης γονιδίων του βακτηρίου *Bacteroides thetaiotaomicron* (βλ. **Ενότητα 3.4.1**) και του *Salmonella Typhimurium* (βλ. **Ενότητα 3.4.2**) εντοπίσαμε αλληλουχίες και άλλων μικροοργανισμών που δεν αναμέναμε και που απαιτούν περαιτέρω διερεύνηση της ύπαρξής τους.

Όσον αφορά την προσπάθεια αναγνώρισης των μικροβιακών γονιδίων που εκφράζονται, συναντήσαμε πληθώρα αλληλουχιών από rRNA γονίδια.

Συμπερασματικά, για τη μελέτη αλληλουχιών μικροβίων σε RNA-Seq δεδομένα δειγμάτων ιστών του ξενιστή τους προτείνουμε:

- α) Στοίχιση των αλληλουχιών των δειγμάτων στο γονιδίωμα του ξενιστή δύο φορές και συγκέντρωση των αλληλουχιών που δεν στοιχίζονται σε καμία από τις δύο προσπάθειες.
- β) Στην περίπτωση paired-end δεδομένων θεωρούμε κάποια αλληλουχία/read ότι δεν στοιχίζεται στο γονιδίωμα του ξενιστή αν και τα δύο mates του read δε στοιχίζονται.
- γ) Συμπερίληψη του γονιδιώματος του ξενιστή μαζί με τα γονιδιώματα των μικροοργανισμών που θα χρησιμοποιηθούν για τη δημιουργία της βάσης/index από τα προγράμματα μεταγονιδιωματικής ανάλυσης.
- δ) Χρησιμοποίηση παραπάνω του ενός προγραμμάτων μεταγονιδιωματικής για εξαγωγή συναινετικών παρατηρήσεων· στις **Ενότητες 3.1.2, 3.2.2 και 3.4.4**

συναντήσαμε περιπτώσεις όπου εντοπίζονται σε μεμονωμένα δείγματα αλληλουχίες μικροβίων από ένα πρόγραμμα αλλά δεν επιβεβαιώνονται από όλα τα προγράμματα.

Οι περιορισμοί της παραπάνω μεθοδολογίας συνοψίζονται στην αδυναμία ποσοτικοποίησης της αφθονίας των μικροβίων ή των εκφραζόμενων γονιδίων τους, καθώς χρησιμοποιούνται δεδομένα από πειράματα που δεν στοχεύουν στην αλληλούχιση μικροβίων. Επίσης, όπως σε κάθε υπολογιστική μεταγονιδιωματική μελέτη, η ταυτοποίηση μικροοργανισμών στα δείγματα δεν εξασφαλίζει ότι αυτά πράγματι βρίσκονταν εκεί πριν τη δειγματοληψία και ότι δεν επιμόλυναν τα δείγματα κατά τη δημιουργία ή/και επεξεργασία τους.

Μία τέτοια μεθοδολογία, ωστόσο, μπορεί να χρησιμοποιηθεί για την ανάδειξη συσχετίσεων μικροβίων με ασθένειες, μελετώντας RNA-Seq δεδομένα που έχουμε ήδη στη διάθεση μας και δεν είχαν δημιουργηθεί για το σκοπό αυτό.

ΠΙΝΑΚΑΣ ΟΡΟΛΟΓΙΑΣ

Ξενόγλωσσος όρος	Ελληνικός Όρος
metagenomics	μεταγονιδιωματική
microbiota	μικροχλωρίδα
microbiome	μικροβίωμα
monocolonized with	επικοισμένο μόνο από
germ-free	ελεύθερος μικροβίων
gnotobiotic	με γνωστό μικροβίωμα
gut-brain axis	άξονας εντέρου - εγκεφάλου
taxon/taxa	ταξινομική ομάδα/ομάδες
read	αλληλουχία πειράματος NGS
rRNA	ριβοσωμικό γονίδιο
RNA-Sequencing	αλληλούχιση του RNA
SPF, specific pathogen free	δίχως συγκεκριμένα παθογόνα μικρόβια

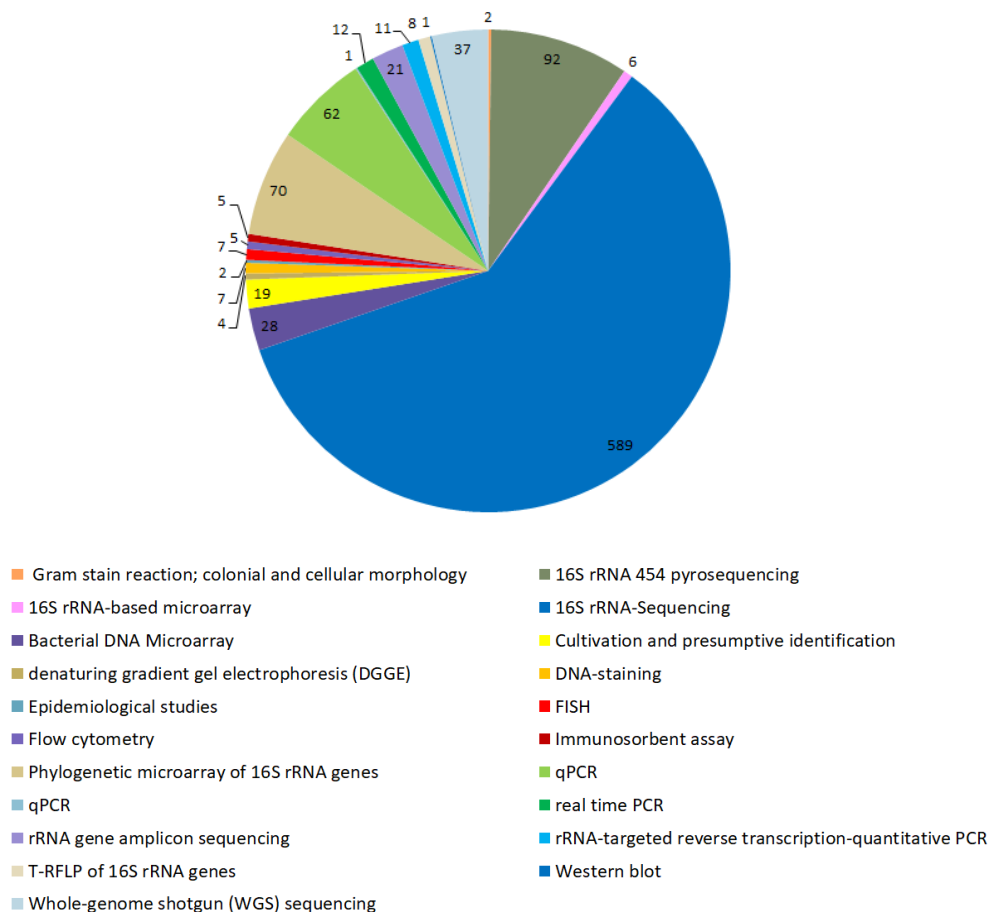
ΣΥΝΤΜΗΣΕΙΣ – ΑΡΚΤΙΚΟΛΕΞΑ – ΑΚΡΩΝΥΜΙΑ

DNA	Deoxyribonucleic Acid, δεοξυριβονουκλεϊκό οξύ
GABA	gamma-aminobutyric acid, γ- αμινοβουτιρικό οξύ
GEO	Gene Expression Omnibus, συλλογή της έκφρασης των γονιδίων
GF	Germ-Free, ελεύθερος μικροβίων
GI	Gastrointestinal, γαστρεντερικός
MC	Monocolonized, επικοισμένο μόνο από έναν μικροοργανισμό
NGS	Next Generation Sequencing, αλληλούχιση επόμενης γενιάς
OTU	Operational Taxonomic Unit, πληθυντικός OTUs
PCR	Polymerase Chain Reaction, αλυσιδωτή αντίδραση πολυμεράσης
RNA	Ribonucleic acid, ριβονουκλεϊκό οξύ
RNA-Seq	RNA Sequencing, αλληλούχιση του RNA
rRNA	ribosomal RNA, ριβοσωμικό RNA
SPF	Specific Pathogen Free, δίχως συγκεκριμένα παθογόνα μικρόβια

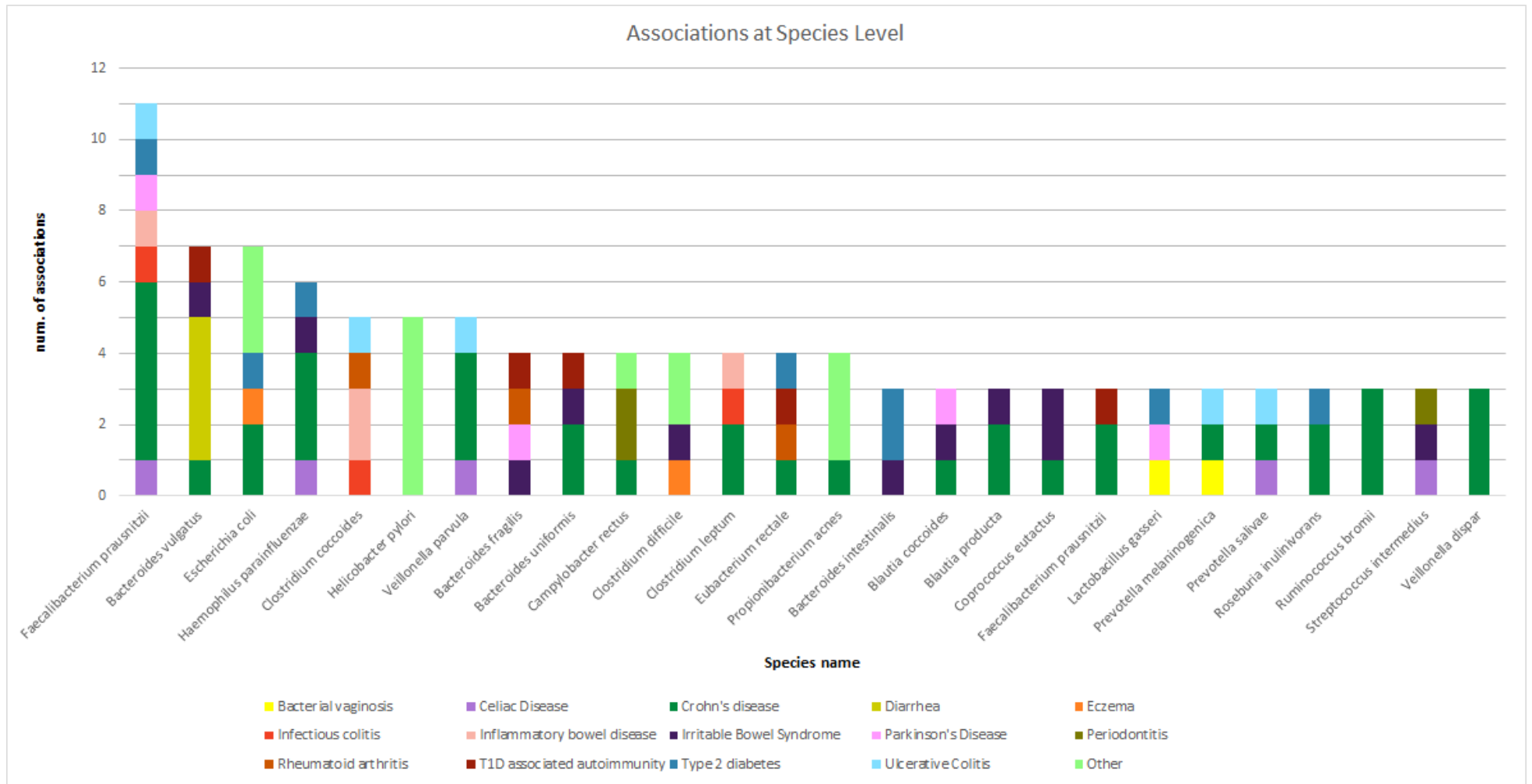
ΠΑΡΑΡΤΗΜΑ Ι

Για την κάλυψη της ανάγκης συλλογής, διαχείρισης και συγκριτικής μελέτης των νέων συσχετίσεων μεταξύ μικροβίων και ασθενειών του ανθρώπου, που ανακαλύπτονται συνεχώς από την επιστημονική κοινότητα, δημιουργήσαμε μια λίστα που συγκεντρώνει τέτοιες συσχετίσεις, με σκοπό τη δημιουργία αντίστοιχης βάσης δεδομένων. Συγκεκριμένα, από 101 επιστημονικές δημοσιεύσεις, συλλέξαμε 995 συσχετίσεις. Σε αυτές τις δημοσιεύσεις οι πειραματικές μέθοδοι που χρησιμοποιούνται περιλαμβάνουν καλλιέργεια κυττάρων, ανοσοϊστοχημεία, αλληλούχιση 16S rRNA, PCR, qPCR, Shotgun αλληλούχιση κ.ά (βλ. **Γράφημα 1**). Εμείς λάβαμε υπ' όψιν μόνο συσχετίσεις στατιστικά σημαντικές και για κάθε μία σημειώσαμε μετα-πληροφορίες (meta-data), όπως: ιστός προέλευσης των δειγμάτων, αριθμός συμμετεχόντων, τα δημογραφικά χαρακτηριστικά τους, κ.ά.

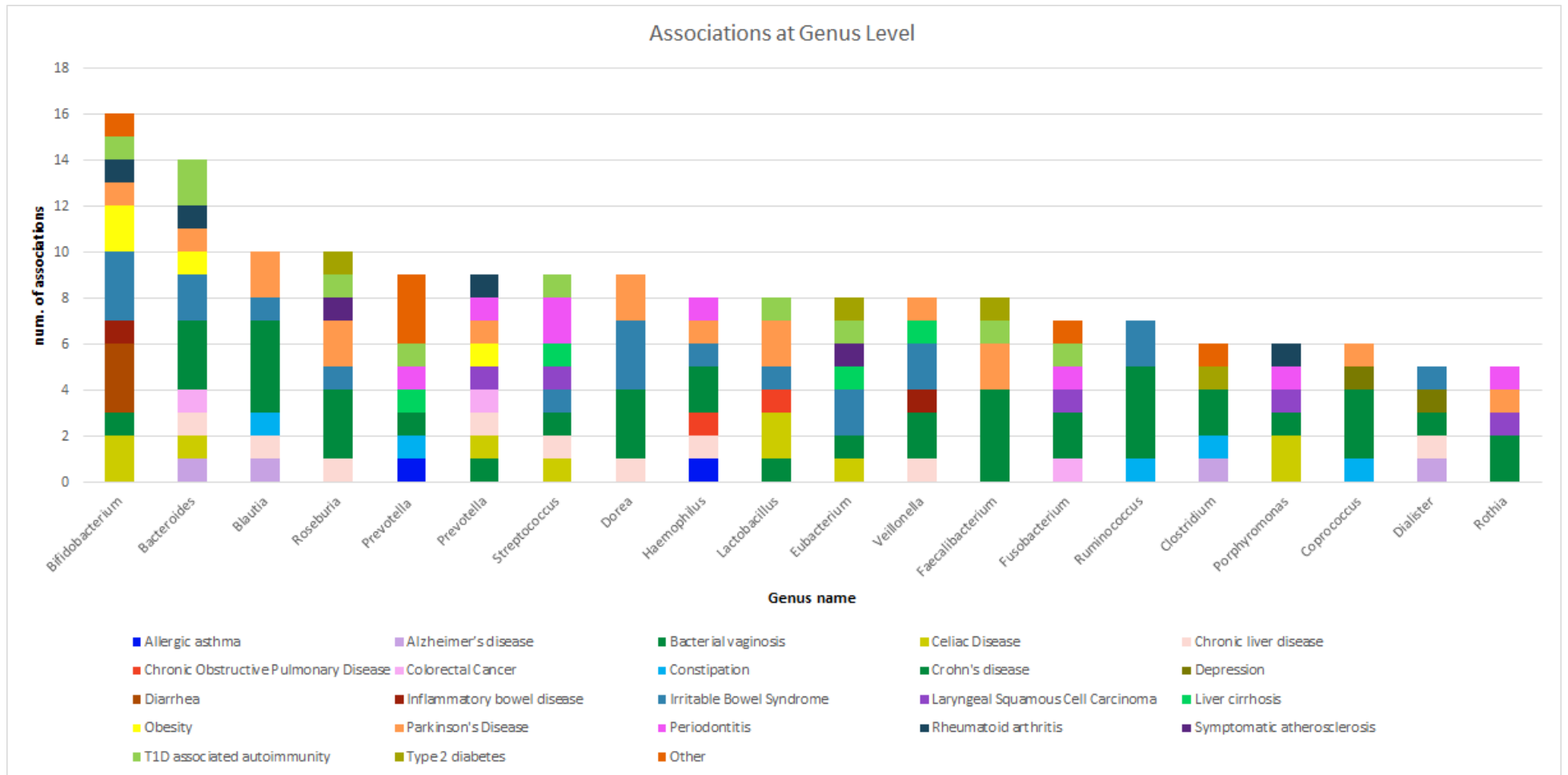
Στο **Γράφημα 2** δείχνουμε ενδεικτικά εκείνα τα είδη μικροοργανισμών που έχουν συσχετιστεί τουλάχιστον 3 φορές με ασθένειες, ενώ στο **Γράφημα 3** το ίδιο, αλλά σε επίπεδο γένους.



Γράφημα 1 Κυκλόγραμμα (πίτα) με τον αριθμό των συσχετίσεων που έχουν επιβεβαιωθεί χρησιμοποιώντας μία από τις μεθόδους του υπομνήματος.



Γράφημα 2 Τα είδη μικροοργανισμών με τουλάχιστον 3 συσχετίσεις με ασθένειες. Στον κάθετο άξονα μετράται ο αριθμός των συσχετίσεων και κάθε χρώμα αντιστοιχεί σε διαφορετική ασθένεια, σύμφωνα με το υπόμνημα.



Γράφημα 3 Τα γένη μικροοργανισμών με τουλάχιστον 3 συσχετίσεις με ασθένειες. Στον κάθετο άξονα μετράται ο αριθμός των συσχετίσεων και κάθε χρώμα αντιστοιχεί σε διαφορετική ασθένεια, σύμφωνα με το υπόμνημα.

ΑΝΑΦΟΡΕΣ

- [1] *InformedHealth.org [Internet]. Cologne, Germany: Institute for Quality and Efficiency in Health Care (IQWiG); 2006-. What are microbes? 2010 Oct 6 [Updated 2019 Aug 29]. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK279387/>.*
- [2] Κύρτσου, Α.Δ.Κ.-. *Γενική Μικροβιολογία*. 2012: Εκδόσεις Αθ. Σταμούλης.
- [3] White, T.C., et al., *Fungi on the skin: dermatophytes and Malassezia*. Cold Spring Harbor perspectives in medicine, 2014. **4**(8): p. a019802.
- [4] Stahl, M.T.M.J.M.M.K.S.B.D.N.D.A., *Brock Βιολογία Των Μικροοργανισμών*. Vol. Κεφάλαιο 2. 2018: Πανεπιστημιακές Εκδόσεις Κρήτης.
- [5] Turnbaugh, P.J., et al., *The human microbiome project*. Nature, 2007. **449**(7164): p. 804-10.
- [6] Qin, J., et al., *A human gut microbial gene catalogue established by metagenomic sequencing*. Nature, 2010. **464**(7285): p. 59-65.
- [7] Peterson, J., et al., *The NIH Human Microbiome Project*. Genome Res, 2009. **19**(12): p. 2317-23.
- [8] Pace, N.R., *A molecular view of microbial diversity and the biosphere*. Science, 1997. **276**(5313): p. 734-40.
- [9] Ley, R.E., et al., *Worlds within worlds: evolution of the vertebrate gut microbiota*. Nat Rev Microbiol, 2008. **6**(10): p. 776-88.
- [10] Marchesi, J.R., *Prokaryotic and eukaryotic diversity of the human gut*. Adv Appl Microbiol, 2010. **72**: p. 43-62.
- [11] Breitbart, M., et al., *Viral diversity and dynamics in an infant gut*. Res Microbiol, 2008. **159**(5): p. 367-73.
- [12] Ursell, L.K., et al., *Defining the human microbiome*. Nutr Rev, 2012. **70 Suppl 1**: p. S38-44.
- [13] Sender, R., S. Fuchs, and R. Milo, *Revised Estimates for the Number of Human and Bacteria Cells in the Body*. PLoS Biol, 2016. **14**(8): p. e1002533.
- [14] Dominguez-Bello, M.G., et al., *Development of the human gastrointestinal microbiota and insights from high-throughput sequencing*. Gastroenterology, 2011. **140**(6): p. 1713-9.
- [15] Dominguez-Bello, M.G., et al., *Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns*. Proc Natl Acad Sci U S A, 2010. **107**(26): p. 11971-5.
- [16] Koenig, J.E., et al., *Succession of microbial consortia in the developing infant gut microbiome*. Proc Natl Acad Sci U S A, 2011. **108 Suppl 1**: p. 4578-85.
- [17] Roy, N.C., et al., *Chapter 18 - The Importance of Microbiota and Host Interactions Throughout Life*, in *Food Structures, Digestion and Health*, M. Boland, M. Golding, and H. Singh, Editors. 2014, Academic Press: San Diego. p. 489-511.
- [18] Macfarlane, S. and G.T. Macfarlane, *Regulation of short-chain fatty acid production*. Proc Nutr Soc, 2003. **62**(1): p. 67-72.
- [19] Sartor, R.B., *Microbial influences in inflammatory bowel diseases*. Gastroenterology, 2008. **134**(2): p. 577-94.
- [20] Bourriaud, C., et al., *Butyrate production from lactate by human colonic microflora*. Reprod Nutr Dev, 2002. **42**: p. S55.
- [21] Cantarel, B.L., V. Lombard, and B. Henrissat, *Complex carbohydrate utilization by the healthy human microbiome*. PLoS One, 2012. **7**(6): p. e28742.
- [22] Sidhu, H., et al., *Absence of Oxalobacter formigenes in cystic fibrosis patients: a risk factor for hyperoxaluria*. Lancet, 1998. **352**(9133): p. 1026-9.
- [23] Magwira, C.A., et al., *Diversity of faecal oxalate-degrading bacteria in black and white South African study groups: insights into understanding the rarity of urolithiasis in the black group*. J Appl Microbiol, 2012. **113**(2): p. 418-28.
- [24] Hooper, L.V., et al., *Molecular analysis of commensal host-microbial relationships in the intestine*. Science, 2001. **291**(5505): p. 881-4.

- [25] De Biase, D. and E. Pennacchietti, *Glutamate decarboxylase-dependent acid resistance in orally acquired bacteria: function, distribution and biomedical implications of the gadBC operon*. Mol Microbiol, 2012. **86**(4): p. 770-86.
- [26] Baddini Feitoza, A., et al., *Conjugated linoleic acid (CLA): effect modulation of body composition and lipid profile*. Nutr Hosp, 2009. **24**(4): p. 422-8.
- [27] Devillard, E., et al., *Metabolism of linoleic acid by human gut bacteria: different routes for biosynthesis of conjugated linoleic acid*. J Bacteriol, 2007. **189**(6): p. 2566-70.
- [28] Marin, L., et al., *Bioavailability of dietary polyphenols and gut microbiota metabolism: antimicrobial properties*. 2015. **2015**: p. 905215.
- [29] Chung, H., et al., *Gut immune maturation depends on colonization with a host-specific microbiota*. Cell, 2012. **149**(7): p. 1578-93.
- [30] Bisgaard, H., et al., *Reduced diversity of the intestinal microbiota during infancy is associated with increased risk of allergic disease at school age*. J Allergy Clin Immunol, 2011. **128**(3): p. 646-52.e1-5.
- [31] Russell, S.L., et al., *Early life antibiotic-driven changes in microbiota enhance susceptibility to allergic asthma*. EMBO reports, 2012. **13**(5): p. 440-447.
- [32] Kamada, N., et al., *Regulated virulence controls the ability of a pathogen to compete with the gut microbiota*. Science, 2012. **336**(6086): p. 1325-9.
- [33] Gantois, I., et al., *Butyrate specifically down-regulates salmonella pathogenicity island 1 gene expression*. Appl Environ Microbiol, 2006. **72**(1): p. 946-9.
- [34] Shin, R., M. Suzuki, and Y. Morishita, *Influence of intestinal anaerobes and organic acids on the growth of enterohaemorrhagic Escherichia coli O157:H7*. J Med Microbiol, 2002. **51**(3): p. 201-6.
- [35] Kommineni, S., et al., *Bacteriocin production augments niche competition by enterococci in the mammalian gastrointestinal tract*. Nature, 2015. **526**(7575): p. 719-22.
- [36] Wang, Y. and L.H. Kasper, *The role of microbiome in central nervous system disorders*. Brain Behav Immun, 2014. **38**: p. 1-12.
- [37] Diaz Heijtz, R., et al., *Normal gut microbiota modulates brain development and behavior*. Proc Natl Acad Sci U S A, 2011. **108**(7): p. 3047-52.
- [38] Neufeld, K.M., et al., *Reduced anxiety-like behavior and central neurochemical change in germ-free mice*. Neurogastroenterol Motil, 2011. **23**(3): p. 255-64, e119.
- [39] Le Chatelier, E., et al., *Richness of human gut microbiome correlates with metabolic markers*. Nature, 2013. **500**(7464): p. 541-6.
- [40] Ley, R.E., et al., *Microbial ecology: human gut microbes associated with obesity*. Nature, 2006. **444**(7122): p. 1022-3.
- [41] Qin, J., et al., *A metagenome-wide association study of gut microbiota in type 2 diabetes*. Nature, 2012. **490**(7418): p. 55-60.
- [42] Manichanh, C., et al., *The gut microbiota in IBD*. Nat Rev Gastroenterol Hepatol, 2012. **9**(10): p. 599-608.
- [43] Vaahtovuori, J., et al., *Fecal Microbiota in Early Rheumatoid Arthritis*. The Journal of Rheumatology, 2008. **35**(8): p. 1500.
- [44] De Angelis, M., et al., *Fecal microbiota and metabolome of children with autism and pervasive developmental disorder not otherwise specified*. PLoS One, 2013. **8**(10): p. e76993.
- [45] Severance, E.G., et al., *Gastrointestinal inflammation and associated immune activation in schizophrenia*. Schizophrenia Research, 2012. **138**(1): p. 48-53.
- [46] Severance, E.G., et al., *Seroreactive marker for inflammatory bowel disease and associations with antibodies to dietary proteins in bipolar disorder*. Bipolar Disorders, 2014. **16**(3): p. 230-240.
- [47] Severance, E.G., et al., *Discordant patterns of bacterial translocation markers and implications for innate immune imbalances in schizophrenia*. Schizophrenia Research, 2013. **148**(1): p. 130-137.

- [48] Samaroo, D., et al., *Novel immune response to gluten in individuals with schizophrenia*. Schizophrenia Research, 2010. **118**(1): p. 248-255.
- [49] Nguyen, T.L.A., et al., *How informative is the mouse for human gut microbiota research?* Disease Models & Mechanisms, 2015. **8**(1): p. 1.
- [50] Casteleyn, C., et al., *Surface area assessment of the murine intestinal tract as a prerequisite for oral dose translation from mouse to man*. Laboratory Animals, 2010. **44**(3): p. 176-183.
- [51] Treuting, P.M., M.J. Arends, and S.M. Dintzis, *12 - Lower Gastrointestinal Tract*, in *Comparative Anatomy and Histology (Second Edition)*, P.M. Treuting, S.M. Dintzis, and K.S. Montine, Editors. 2018, Academic Press: San Diego. p. 213-228.
- [52] Eckburg, P.B., et al., *Diversity of the Human Intestinal Microbial Flora*. Science, 2005. **308**(5728): p. 1635.
- [53] Ley, R.E., et al., *Obesity alters gut microbial ecology*. Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(31): p. 11070.
- [54] Linnenbrink, M., et al., *The role of biogeography in shaping diversity of the intestinal microbiota in house mice*. Molecular Ecology, 2013. **22**(7): p. 1904-1916.
- [55] Available from: <http://theses.ulaval.ca/archimede/fichiers/24866/ch07.html>.
- [56] Gordon, H.A. and L. Pesti, *The gnotobiotic animal as a tool in the study of host microbial relationships*. Bacteriol Rev, 1971. **35**(4): p. 390-429.
- [57] Silva, A., et al., *Effect of Bifidobacterium longum ingestion on experimental salmonellosis in mice*. Journal of applied microbiology, 2004. **97**: p. 29-37.
- [58] Rosenberg, E., *Chapter 13 - Contribution of Microbes to the Health of Humans, Animals, and Plants*, in *It's in Your DNA*, E. Rosenberg, Editor. 2017, Academic Press. p. 115-128.
- [59] Ridaura, V.K., et al., *Gut Microbiota from Twins Discordant for Obesity Modulate Metabolism in Mice*. Science, 2013. **341**(6150): p. 1241214.
- [60] Reyniers, J.A., P.C. Trexler, and R.F. Ervin, *Rearing germ-free albino rats*. Lobund reports, 1946(1): p. 1-84.
- [61] Chriss J. Vowles, N.E.A., Kathryn A. Eaton, *Gnotobiotic Mouse Technology: An Illustrated Guide*. 2015: CRC Press.
- [62] Al-Asmakh, M. and F. Zadjali, *Use of Germ-Free Animal Models in Microbiota-Related Research*. J Microbiol Biotechnol, 2015. **25**(10): p. 1583-8.
- [63] Trexler, P.C. and L.I. Reynolds, *Flexible film apparatus for the rearing and use of germfree animals*. Applied microbiology, 1957. **5**(6): p. 406-412.
- [64] Packey, C.D., et al., *Molecular detection of bacterial contamination in gnotobiotic rodent units*. Gut Microbes, 2013. **4**(5): p. 361-70.
- [65] Arvidsson, C., A. Hallén, and F. Bäckhed, *Generating and Analyzing Germ-Free Mice*. Current Protocols in Mouse Biology, 2012. **2**(4): p. 307-316.
- [66] F. Gypas, A.-F.A.M., *Τεχνικές μελέτης των φυσιολογικών μικροχλωρίδων του ανθρώπου - μεταγονιδιωματική*. Acta Microbiologica Hellenica, 2014. **59**(2).
- [67] Barrett, T., et al., *NCBI GEO: archive for functional genomics data sets--update*. Nucleic Acids Res, 2013. **41**(Database issue): p. D991-5.
- [68] Wood, D.E., J. Lu, and B. Langmead, *Improved metagenomic analysis with Kraken 2*. Genome biology, 2019. **20**(1): p. 257-257.
- [69] Ounit, R., et al., *CLARK: fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers*. BMC Genomics, 2015. **16**: p. 236.
- [70] <https://eutils.ncbi.nlm.nih.gov/entrez/eutils/>.
- [71] *Sequence Read Archive Submissions Staff. Using the SRA Toolkit to convert .sra files into other formats*. In: *SRA Knowledge Base [Internet]*. . 2011; Available from: <https://www.ncbi.nlm.nih.gov/books/NBK158900/>, <http://ncbi.github.io/sra-tools/>.
- [72] Andrews S. (2010). *FastQC: a quality control tool for high throughput sequence data*. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.

- [73] Martin, M., *Cutadapt removes adapter sequences from high-throughput sequencing reads*. 2011, 2011. **17**(1): p. 3.
- [74] Kim, D., et al., *Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype*. *Nature Biotechnology*, 2019. **37**(8): p. 907-915.
- [75] Aken, B.L., et al., *The Ensembl gene annotation system*. *Database*, 2016. **2016**.
- [76] Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2*. *Nat Methods*, 2012. **9**(4): p. 357-9.
- [77] Quinlan, A.R., *BEDTools: The Swiss-Army Tool for Genome Feature Analysis*. *Curr Protoc Bioinformatics*, 2014. **47**: p. 11 12 1-34.
- [78] Karp, P.D., et al., *The EcoCyc Database*. *EcoSal Plus*, 2018. **8**(1).
- [79] Santos-Zavaleta, A., et al., *RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in E. coli K-12*. *Nucleic Acids Res*, 2019. **47**(D1): p. D212-d220.