



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικόν και Καποδιστριακόν
Πανεπιστήμιον Αθηνών
—ΙΔΡΥΘΕΝ ΤΟ 1837—

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικό και Καποδιστριακό
Πανεπιστήμιο Αθηνών

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ»

Δ Ι Π Λ Ω Μ Α Τ Ι Κ Η Ε Ρ Γ Α Σ Ι Α

«Μεταγονιδιωματική ανάλυση του οικοσυστήματος Ελληνικού παραδοσιακού τυριού»



ΟΛΥΜΠΙΑ ΤΣΙΠΙΔΟΥ

Πτυχιούχος Δασολογίας και Διαχείρισης Περιβάλλοντος και Φυσικών Πόρων,
Δημοκρίτειο Πανεπιστήμιο Θράκης

ΑΘΗΝΑ (2019 - 2020)



HELLENIC REPUBLIC
National and Kapodistrian
University of Athens
—EST. 1837—

HELLENIC REPUBLIC
National and Kapodistrian
University of Athens

SCHOOL OF SCIENCE
DEPARTMENT OF BIOLOGY

MASTER IN «BIOINFORMATICS»

Master Diploma Thesis

«Metagenomic Analysis on the Ecosystem of Traditional Greek Cheese»



OLYMPIA TSIPIDOY

Department of Forestry and Environmental Management and Natural
Resources, Democritus University of Thrace

A T H E N S (2 0 1 9 - 2 0 2 0)



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικόν και Καποδιστριακόν
Πανεπιστήμιον Αθηνών
—ΙΔΡΥΘΕΝ ΤΟ 1837—

ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικό και Καποδιστριακό
Πανεπιστήμιο Αθηνών

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΒΙΟΛΟΓΙΑΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ»

Δ Ι Π Λ Ω Μ Α Τ Ι Κ Η Ε Ρ Γ Α Σ Ι Α

«Μεταγονιδιωματική ανάλυση του οικοσυστήματος Ελληνικού
παραδοσιακού τυριού»

Τριμελής εξεταστική επιτροπή

Επίκουρη Καθηγήτρια Βασιλική Οικονομίδου (Επιβλέπουσα)
*Τομέας Βιοφυσικής,
Τμήματος Βιολογίας, ΕΚΠΑ*

Επίκουρος Καθηγητής Κωνσταντίνος Παπαδημητρίου
*Τμήμα Επιστήμης και Τεχνολογίας Τροφίμων,
Πανεπιστήμιο Πελοποννήσου*

Ομότιμος Καθηγητής Σταύρος Ι. Χαμόδρακας
*Τομέας Βιοφυσικής & Υπολογιστικής Βιολογίας/Βιοπληροφορικής,
Τμήμα Βιολογίας, ΕΚΠΑ*

Ευχαριστίες

Θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες στον κ. Κωνσταντίνο Παπαδημητρίου, Επίκουρο Καθηγητή του τμήματος Επιστήμης και Τεχνολογίας Τροφίμων του Πανεπιστημίου Πελοποννήσου για την καθοδήγηση, την αμέριστη βοήθεια και την εμπιστοσύνη που έδειξε προς το πρόσωπο μου, καθ' όλη τη διάρκεια διεκπεραίωσης της διπλωματικής εργασίας. Επίσης, θα ήθελα να ευχαριστήσω την Επίκουρη Καθηγήτρια Βασιλική Οικονομίδου και τον Ομότιμο Καθηγητή Σταύρο Ι. Χαμόδρακα για τη συμμετοχή τους στην Τριμελή Επιτροπή εξέτασης. Συνεχίζοντας, θα ήθελα να απευθύνω ένα μεγάλο ευχαριστώ στην υποψήφια διδάκτορα Μαρκέλλα Τσιγκριμάνη και στον Δρ. Αναστάσιο Γεωργούλη, για τη συνεισφορά του στο τμήμα των εργαστηριακών αναλύσεων. Ολοκληρώνοντας, οφείλω να διατυπώσω ένα εξίσου μεγάλο ευχαριστώ στην οικογένεια μου και στους ανθρώπους που βρίσκονται κοντά μου όλο αυτό το διάστημα, για την πολύτιμη ψυχολογική και ηθική υποστήριξη.

ΠΕΡΙΛΗΨΗ

Η παρούσα διπλωματική εργασία πραγματοποιήθηκε στα πλαίσια του μεταπτυχιακού προγράμματος «Βιοπληροφορικής», με τίτλο «Μεταγονιδιωματική ανάλυση του οικοσυστήματος Ελληνικού παραδοσιακού τυριού». Η διαδικασία παρασκευής τροφίμων και συγκεκριμένα προϊόντων που έχουν υποστεί ζύμωση, περιλαμβάνει μια πολύπλοκη μικροχλωρίδα η οποία αποτελείται από βακτήρια, ζύμες και μύκητες. Η εκτενής ανάλυση της μικροβιακής σύστασης των περισσότερων τυροκομικών προϊόντων δεν έχει διερευνηθεί ως τώρα με τις τεχνικές που προϋπήρχαν. Μέσω της τεχνολογίας NGS μπορούν να παραχθούν χιλιάδες έως και εκατομμύρια αντίγραφα των ακολουθιών του εκάστοτε δείγματος την ίδια στιγμή, με υψηλή απόδοση. Το είδος αυτής της ανάλυσης ανήκει στον τομέα της μεταγονιδιωματικής καθώς αναλύει περιβαλλοντικά δείγματα που αποτελούνται από πλήθος μικροοργανισμών. Οι αλληλουχίες που εμπεριέχονται στα περιβαλλοντικά δείγματα μας επιτρέπουν την ακριβή αναγνώριση του μικροβιώματος και την ταξινόμησή του, ακόμα και σε περιπτώσεις πληθυσμών με μικρές συγκεντρώσεις.

Στη συγκεκριμένη μελέτη έγινε ταυτοποίηση της μικροχλωρίδας του Ελληνικού ΠΟΠ τυριού Σφέλα και του Κυπριακού τυριού Χαλίτζι. Πραγματοποιήθηκε η απομόνωση του DNA των δειγμάτων και η ανάλυση των δεδομένων αλληλούχισης NGS, αποδίδουν την μικροβιακή ποικιλότητα και τη δομή των πληθυσμών που εμπεριέχονται. Χρησιμοποιήθηκαν τρεις διαφορετικές μεταγονιδιωματικές προσεγγίσεις. Αρχικά εφαρμόστηκαν δύο τεχνικές Amplicon based των περιοχών 16s rDNA και ITS (Internal Transcribed Spacer) που επιτρέπουν την ταυτοποίηση βακτηρίων και ζυμών μέχρι την ταξινομική βαθμίδα του γένους. Επιπλέον, εφαρμόστηκε η τεχνική Shotgun (τυχαίας προσπέλασης) η οποία επιτρέπει την μερική συναρμολόγηση μεγάλων αλληλουχιών. Η τυφλή γονιδιωματική μπορεί να οδηγήσει σε ταυτοποίηση των μικροοργανισμών σε επίπεδο είδους. Χρησιμοποιήθηκαν για την ανάλυση των αρχικών δεδομένων οι πλατφόρμες βιοπληροφορικής CLC Microbial Genomic Module, KBase και MG-RAST Platform.

Ο σκοπός της έρευνας είναι ο καλύτερος χαρακτηρισμός της μικροχλωρίδας των δειγμάτων και η ακριβής ταξινόμηση των ειδών που περιέχονται. Τα αποτελέσματα της ανάλυσης έχουν ζωτική σημασία για την παρακολούθηση και τον χειρισμό των τυροκομικών προϊόντων όπως επίσης και για τη δημιουργία των προϋποθέσεων μελλοντικής έρευνας.

ABSTRACT

The present study, entitled “Metagenomic Analysis on the Ecosystem of Traditional Greek Cheese”, was part of the master program “Bioinformatics” by the Department of Biology at the National and Kapodistrian University of Athens.

The production process of fermented foods, like cheese, include a complex interactive system of microorganisms. The extensive analysis of this composition on dairy products has not been researched in detail. NGS technology produces thousands to millions of sequence copies of the metagenomic DNA isolated from each environmental sample. These copies are generated in parallel with high performance. The metagenomic approach allows the investigation of microbial communities from environmental samples, without the need for cultivation, and they can provide important information. The sequences contained on samples can be used to identify accurately the microbiome including taxa identification and functional properties.

This study attempts to identify the microbiome of Greek cheeses, more specifically Greek PDO Sfela cheese and Cyprus Chalitzia cheese. After DNA sequencing, the information obtained can reveal the microbial diversity and the structure of the population in cheese products. We used three different NGS metagenomics techniques. Techniques called amplicon based like 16S rDNA and ITS (Internal Transcribed Spacer) are used to investigate the classification of microorganisms up to the taxonomic level of genus. 16S rDNA analysis was used to identify bacteria, while ITS analysis was used to identify fungi. Additionally, shotgun metagenomics is a method used for sequencing random DNA strands, which includes all the genomic material of the samples and allows the partial assembly of large sequences. Shotgun technique can provide information on taxonomic species level. We used three different platforms to analyse the raw data, CLC Microbial Genomic Module, KBase Platform and MG-RAST Platform.

The purpose of this study is to characterize the microbial ecosystem in two different white brined cheeses and the classification of the bacterial and yeast species. The results of the analysis are vital to monitor and handle dairy products as well as to extract information of the hidden world of cheese microbiota.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

1. ΕΙΣΑΓΩΓΗ.....	15
1.1 Ιστορική Εξέλιξη Τυροκομικών Προϊόντων	15
1.1.1. Τυροκομικά προϊόντα.....	15
1.1.2. Τυροκομία σε παγκόσμιο επίπεδο	17
1.1.3. Τυροκομία στην Ελλάδα.....	18
1.2. Στάδια Παραγωγής.....	21
1.2.1. Οργανικά και Ανόργανα Στοιχεία Γάλακτος - Τυριού	21
1.2.2. Παραγωγική Διαδικασία Τυριού	22
1.2.3. Είδη Τυριών	24
1.3 Μικροοργανισμοί στα Τυριά.....	26
1.3.1. Μικροβίωμα Τυριού	26
1.3.2. Κύριες καλλιέργειες εκκίνησης	27
1.3.3. Βακτήρια	28
1.3.3.1 Οξυγαλακτικά Βακτήρια	28
1.3.3.2. Αντιμικροβιακές ιδιότητες των οξυγαλακτικών βακτηρίων	32
1.3.3.3. Βακτηριοσίνες.....	32
1.3.3.4. Παθογόνα Βακτήρια στα Τυριά	33
1.3.4. Μύκητες.....	35
1.3.5. Ζύμες.....	36
1.4 Τυριά που θα αναλυθούν στην Μελέτη	37
1.4.1. Σφέλα.....	37
1.4.2. Χαλίτζι.....	38
1.5 Αλληλούχιση Επόμενης Γενιάς (Next Generation Sequence).....	40
1.5.1. Θεωρητικό Πλαίσιο Αλληλούχισης Επόμενης Γενιάς	40
1.5.2. Διαδικασία Αλληλούχισης Επόμενης Γενιάς	41
1.5.2.1. Βασικά βήματα αλληλούχισης.....	41
1.5.2.2. Sequencing by Synthesis (Illumina/Solexa)	43
1.6 Βιοπληροφορική ανάλυση Μεταγονιδιωματικών Δεδομένων	49
1.6.1 Μεταγονιδιωματική (Metagenomics)	49
1.6.2 Τεχνική Shotgun – Τυχαίας Προσπέλασης.....	50

1.6.3 Amplicon Based - 16S rDNA.....	53
1.6.4 Amplicon Based - ITS (Internal Transcribed Spacer)	54
1.7 Σκοπός Διπλωματικής.....	56
2. ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ	58
2.1 Απομόνωση DNA.....	58
2.2 Μεταγονιδιωματική Ανάλυση - Analysis Workflow	60
2.2.1 Αρχεία FastQ.....	63
2.2.2. Quality Control - Trimming - Extraction of Host Genome	66
2.2.3 Assembly – Συναρμολόγηση των contigs.....	68
2.2.4 Binning	72
2.2.5 Gene Detection – Gene Annotation	75
2.3 Πλατφόρμες και Εργαλεία.....	77
2.3.1 CLC Platform.....	77
2.3.2. KBase Platform.....	82
2.3.2.1 Εργαλεία Assembly.....	84
2.3.2.2 Εργαλεία Binning.....	86
2.3.2.3 Εργαλείο για Βελτιστοποίηση και Φιλτράρισμα των Bins... 88	
2.3.2.4 Εργαλεία για Taxonomic Analysis	89
2.3.3. MG-Rast Platform.....	96
2.3.4 Άλλα εργαλεία που χρησιμοποιήθηκαν	101
2.3.4.1. BusyBee.....	101
3. ΑΠΟΤΕΛΕΣΜΑΤΑ	104
3.1 Αποτελέσματα Amplicon based Αναλύσεων	104
3.1.1 Αποτελέσματα Δεδομένων Ανάλυσης 16S rDNA	104
3.1.2. Αποτελέσματα Δεδομένων Ανάλυσης ITS	109
3.2. Αποτελέσματα Ανάλυσης Δεδομένων Shotgun	112
3.3. Αποτελέσματα Λειτουργικής Ανάλυσης Δεδομένων Shotgun	136
4. ΣΥΜΠΕΡΑΣΜΑΤΑ.....	142
4.1 Μεταγονιδιωματική Ανάλυση - Εργαλεία.....	142
4.2 Μικροβίωμα Λευκών Τυριών Άλμης.....	143
4.3 Μικροβίωμα Σφέλα - Χαλίτζι	145
4.4 Επέκταση Ανάλυσης – Μελλοντική Έρευνα	151
REFERENCES.....	154

ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ, ΔΙΑΓΡΑΜΜΑΤΩΝ, ΓΡΑΦΗΜΑΤΩΝ ΚΑΙ ΕΙΚΟΝΩΝ

ΠΙΝΑΚΕΣ

Πίνακας 1: Δεδομένα Παγκόσμιας Παραγωγής Τυροκομικών προϊόντων	18
Πίνακας 2: Π.Ο.Π. Τυριά παραγόμενα στην Ελλάδα	20
Πίνακας 3: Είδη του γένους <i>Lactobacillus</i> με βάση τα προϊόντα ζύμωσης [23]	30
Πίνακας 4: Πλεονεκτήματα και Μειονεκτήματα NGS	41
Πίνακας 5: Phred quality scores	66
Πίνακας 6: Επιμέρους στοιχεία του bin	128
Πίνακας 7: Επιμέρους στοιχεία του bin	129
Πίνακας 8: Επιμέρους στοιχεία του bin	130
Πίνακας 9: Επιμέρους στοιχεία του bin	131
Πίνακας 10: Επιμέρους στοιχεία του bin	132
Πίνακας 11: Επιμέρους στοιχεία του bin	134

ΔΙΑΓΡΑΜΜΑΤΑ

Διάγραμμα 1: Κατάταξη τυριών με βάση την υγρασία [10]	25
Διάγραμμα 2: Διάκρισης Οξυγαλακτικών βακτηρίων βάσει της μορφής του κυττάρου, της ζύμωσης της γλυκόζης και της θερμοκρασίας ανάπτυξής τους [7]	31
Διάγραμμα 3: Ροή επεξεργασίας τυριού Σφέλα [33]	38
Διάγραμμα 4: Ροή επεξεργασίας τυριού Χαλίτζι [34]	39
Διάγραμμα 5: Διαφορετικές τεχνικές αλληλούχισης επόμενης γενιάς	43
Διάγραμμα 6: Ροή αλληλούχισης με τεχνική Shotgun	52

Διάγραμμα 7: Βασική Ροή Εργασίας Μεταγονιδιωματικής Ανάλυσης	61
Διάγραμμα 8: Βιοπληροφορική Ροή Εργασιών, Taxonomic Profiling και Functional Analysis	62
Διάγραμμα 9: Ροή βιοπληροφορικής ανάλυσης Amplicon τεχνικής.....	62
Διάγραμμα 10: Ροή εργασιών για De Novo Assembly	71
Διάγραμμα 11: Ροή εργασίας αναλύσεων [61].....	74

ΓΡΑΦΗΜΑΤΑ

Γράφημα 1: Χρήση εργαλείου FastQC. Στο πρώτο γράφημα αποτυπώνεται υψηλής ποιότητα αλληλούχισης και στο δεύτερο γράφημα χαμηλής ποιότητας αλληλούχιση	67
--	----

ΕΙΚΟΝΕΣ

Εικόνα 1: Γενεαλογικό δέντρο βασισμένο σε τεχνική που χρησιμοποιεί 16S rDNA και απεικονίζει της φυλογενετικές σχέσεις μεταξύ των LAB. Η γραμμή υποδεικνύει 10% αναμενόμενη γραμμική απόκλιση [24].....	30
Εικόνα 2: Πρόσδεση adapters στα άκρα των θραυσμάτων DNA	44
Εικόνα 3: Μέθοδος αλληλούχισης μέσω σύνθεσης Solexa/Illumina με ποσοτική ενίσχυση θραυσμάτων στερεάς φάσης (solid-phase amplification)	45
Εικόνα 4: Προσδιορισμός βάσεων στην αντίδραση επιμήκυνσης μέσω εμπλουτισμού dNTPs με φθορίζουσα χρώση.	47
Εικόνα 5: Ροή όλης της διαδικασίας αλληλούχισης μέσω σύνθεση.....	48
Εικόνα 6: Περιοχή ITS	55
Εικόνα 7: Μορφή Αρχείου FastQ.....	63
Εικόνα 8: Σύστημα κωδικοποίησης ASCII.....	65
Εικόνα 9: Τα reads τμηματοποιούνται σε k-mers και υπολογίζεται η διαδρομή για τα k-mers, de Bruijn [58].....	70

Εικόνα 10: Interface της πλατφόρμας CLC Genomic Workbench	78
Εικόνα 11: Ροή εργασιών CLC Genomic Workbench για Amplicon ανάλυση .	81
Εικόνα 12: Interface της πλατφόρμας KBase	82
Εικόνα 13: Ροή εργασία της Ανάλυσης	83
Εικόνα 14: Ροή εκτέλεσης υπολογισμών για δημιουργία του de Bruijn γράφου στο IDBA [63]	85
Εικόνα 15: Ροή εργασιών του αλγορίθμου που υλοποιείται από το εργαλείο KAIJU [67]	91
Εικόνα 16: Ροή αλγορίθμου με τα τρία βήματα που εκτελεί [68]	93
Εικόνα 17: Ροή αλγορίθμου για εκτέλεση ταξινομικής ανάλυσης πληθυσμών [69]	95
Εικόνα 18: Το Interface της πλατφόρμας MG-Rast	96
Εικόνα 19: Επιμέρους ροή εργασίας MG-Rast Ανάλυσης	97
Εικόνα 20: Γενική ροή εργασιών κατά την εκτέλεση του MG-Rast.....	99
Εικόνα 21: Οπτικοποίηση ταξινομικών αποτελεσμάτων που υλοποιεί το KRONA [87].....	100
Εικόνα 22: Αλγόριθμος Kraken [92]	102
Εικόνα 23: Δείγματα Σφέλα (α) και Χαλίτζι (β). Με μαύρο X συμβολίσθηκαν γένη που υπολογίσθηκαν ότι βρίσκονται και στα 3 δείγματα διαφορετικών λευκών τυριών άλμης	105
Εικόνα 24: Δείγματα Σφέλα 16S rDNA	106
Εικόνα 25: Δείγματα Χαλίτζι 16S rDNA	107
Εικόνα 26: Alpha-diversity δειγμάτων μέσω της καμπύλης rarefaction α)Σφέλα β)Χαλίτζι	108
Εικόνα 27: Stacked Bars για τα γένη μυκητών που υπολογίσθηκαν στα δείγματα 2 και 3 Χαλίτζι.....	109
Εικόνα 28: α) Απόδοση Alpha-diversity μέσω καμπύλης rarefaction, β) Απόδοση Beta-diversity μέσω γραφικής παράστασης PCoA.....	111

Εικόνα 29: Δείγματα Σφέλας 1 – 2 – 3	112
Εικόνα 30: Δείγματα Χαλίτζι 1 – 2 – 3	113
Εικόνα 31: Δείγματα Σφέλας 1 – 2 – 3	114
Εικόνα 32: Δείγματα Χαλίτζι 1 – 2 – 3	116
Εικόνα 33: Δείγμα 1 Σφέλας, α)MG-Rast Platform β) CLC Microbial Genomic Module.....	117
Εικόνα 34: Δείγμα 2 Σφέλας, α)MG-Rast Platform β) CLC Microbial Genomic Module.....	118
Εικόνα 35: Δείγμα 3 Σφέλας, α)MG-Rast Platform β) CLC Microbial Genomic Module.....	119
Εικόνα 36: Δείγμα 1 Χαλίτζι, α)MG-Rast Platform β) CLC Microbial Genomic Module.....	120
Εικόνα 37: Δείγμα 2 Χαλίτζι, α) MG-Rast Platform β) CLC Microbial Genomic Module.....	121
Εικόνα 38: Δείγμα 3 Χαλίτζι, α)MG-Rast Platform β) CLC Microbial Genomic Module.....	122
Εικόνα 39: Δείγματα Σφέλας	123
Εικόνα 40: Heatmap για τα τρία δείγματα Σφέλας	124
Εικόνα 41: Δείγματα Χαλίτζι	125
Εικόνα 42: Heatmap για τα τρία δείγματα Χαλίτζι	126
Εικόνα 43: Σφέλα Δείγμα 1.....	128
Εικόνα 44: Σφέλας δείγμα 2	129
Εικόνα 45: Τρίτο δείγμα Σφέλας.....	130
Εικόνα 46: Δείγμα 1 Χαλίτζι.....	131
Εικόνα 47: Δείγμα 2 Χαλίτζι.....	132
Εικόνα 48: Δείγμα 3 Χαλίτζι.....	133

Εικόνα 49: Chart με τα αποτελέσματα λειτουργικής ανάλυσης για τα 3 δείγματα Σφέλας	138
Εικόνα 50: Heatmap με τα αποτελέσματα λειτουργικής ανάλυσης για τα 3 δείγματα Σφέλας	139
Εικόνα 51: Chart με τα αποτελέσματα λειτουργικής ανάλυσης για τα 3 δείγματα Χαλίτζι	140
Εικόνα 52: Heatmap με τα αποτελέσματα της λειτουργικής ανάλυσης για τα 3 δείγματα Χαλίτζι	141

1. ΕΙΣΑΓΩΓΗ

1.1 Ιστορική Εξέλιξη Τυροκομικών Προϊόντων

1.1.1. Τυροκομικά προϊόντα

Η εισαγωγή της γαλακτοκομίας στη ζωή του ανθρώπου ήταν ένα κρίσιμο και ζωτικής σημασίας βήμα στην πρώιμη γεωργία. Τα γαλακτοκομικά προϊόντα χρησιμοποιήθηκαν ευρέως μετά την επεξεργασία του γάλακτος από τον άνθρωπο, εξελίσσοντας την διαμόρφωση τους σε τυροκομικά προϊόντα που ικανοποιούσαν τις ανάγκες διατροφής σε μεγάλο βαθμό.

Η ιστορική εξέλιξη των γαλακτοκομικών προϊόντων είναι αλληλένδετη με την εξημέρωση των πρώτων μηρυκαστικών, η οποία τοποθετείται στα οροπέδια του σημερινού Ιράν – Αφγανιστάν το 9.000 π.Χ. [1]. Στην κατηγορία των μηρυκαστικών ανήκαν αρχικά οι αίγες και τα πρόβατα. Ακολούθησε η εισαγωγή της άγριας αγελάδας, 2.000 χρόνια αργότερα, με δύο διαφορετικές εκδοχές για την περιοχή εξημέρωσης της, τη Μέση Ανατολή σε τμήματα της Αφρικής ή στην περιοχή της σημερινής Βόρειας Ευρώπης [2]. Σε αρκετές περιπτώσεις, η εξέλιξη της τυροκομίας, ταυτίζεται με την σταδιακή εξέλιξη της Ευρώπης όπως τη γνωρίζουμε σήμερα.

Έρευνα που διεξήχθη στο πανεπιστήμιο του Bristol το 2012, αποκαλύπτει ότι οι ρίζες της τυροκομίας ενδέχεται να βρίσκονται στη Νεολιθική εποχή, το 7.000 π.Χ. [3]. Το συμπέρασμα της χρονικής τοποθέτησης, στην εισαγωγή των τυροκομικών προϊόντων, πηγάζει από την εύρεση κατάλοιπων λίπους, παραγόμενο από γάλα, σε κεραμικά σκεύη σε βορειοδυτικές περιοχές της Ανατολής. Το γεγονός αυτό, υποδεικνύει την αρχή της επεξεργασίας του γάλακτος και την παρασκευή τυροκομικών προϊόντων. Η επεξεργασία του γάλακτος και ιδίως η παραγωγή τυροκομικών προϊόντων επέτρεψε τη μεταποίηση του γάλακτος σε μορφή που επέτρεπε την εύκολη μεταφορά και τη μεγαλύτερη χρονική διάρκεια αξιοποίησης τους, καθώς διατηρούνται για μεγαλύτερο διάστημα από ότι στην αρχική, ρευστή μορφή. Πολλές από τις ποικιλίες των τυριών που γνωρίζουμε έως σήμερα πρωτοεμφανίστηκαν το Μεσαίωνα, με την μαζική παραγωγή να υπολογίζεται το 17^ο αιώνα. Στην περιοχή της Ελλάδας, η πρώτη αναφορά σε τυροκομικά προϊόντα βρίσκεται

στην Οδύσσεια του Ομήρου (8^ο αιώνα π.Χ.), και συγκεκριμένα στον τυροκόμο Πολύφημο [4].

Επίσης είναι πλέον γνωστό ότι οι ενήλικες *Homo sapiens* δεν παρήγαγαν αρχικά το ένζυμο της λακτάσης, το οποίο συμβάλει στη διάσπαση της λακτόζης που περιέχει το γάλα. Αλλαγή προκλήθηκε μέσω μετάλλαξης, που επέτρεπε την πέψη του γάλακτος, η οποία εμφανίστηκε γεωγραφικά σε σημεία της σημερινής Ευρώπης. Σε περιοχές της Αφρικής και της Ασίας, αυτό το εξελικτικό πλεονέκτημα δεν έλαβε χώρα την ίδια χρονική περίοδο. Η μετάλλαξη αυτή σταδιακά μεταφέρθηκε στη Μέση Ανατολή μέσω αποικιακών εξαπλώσεων και συνδυαστικά με την εξημέρωση των μηρυκαστικών, ξεκίνησε η ανάπτυξη των τυροκομικών προϊόντων.

Σύμφωνα με το Μπαμπινιώτη η λέξη τυρί είναι υποκείμενο της αρχαίας λέξης «*τυρός*», όπως μαρτυρείται από τα Μυκηναϊκά χρόνια. Θεωρείται ότι παράγεται από το ρήμα τυρεύω δηλαδή πήζω, διαδικασία που λαμβάνει χώρα για την στερεοποίηση του γάλακτος [5].

Ο ορισμός για το τυρί βάσει τον Codex Alimentarius [6], είναι ο ακόλουθος:

«Τυρί είναι το νωπό ή ώριμο προϊόν που προέρχεται από στράγγιση, ύστερα από πήξη του πλήρους ή μερικώς αποβουτυρωμένου ή άπαχου γάλακτος ή βουτυρογάλακτος ή μίγματος ορισμένων ή όλων αυτών των προϊόντων.»

Σύμφωνα με τον Ελληνικό Κώδικα Τροφίμων και Ποτών (Άρθρο 83: Τυροκομικά προϊόντα 3 Απρ. 2014):

«Τα τυριά είναι τα προϊόντα ωρίμανσης του πήγματος (στάλπης) που είναι απαλλαγμένο από το τυρόγαλα στον επιθυμητό κάθε φορά βαθμό και τα οποία παρασκευάστηκαν, με την επενέργεια πυτιάς ή άλλων ενζύμων που δρουν ανάλογα σε γάλα (νωπό ή παστεριωμένο, αγελάδος, προβάτου, κατσίκας, βουβάλου και μίγματα αυτών) ή σε μερικώς αποβουτυρωμένο γάλα ή σε μίγμα αυτών ή/και σε μίγματα αυτών με κρέμα γάλακτος (αφρόγαλα).»

1.1.2. Τυροκομία σε παγκόσμιο επίπεδο

Η αγορά και η ζήτηση τυροκομικών προϊόντων σε παγκόσμιο επίπεδο καταλαμβάνει έναν από τους πιο δυναμικούς τομείς τροφίμων τα τελευταία 20 έτη. Το 1/3 της παραγωγής προέρχεται από αγελαδινό γάλα, ξεπερνώντας σε ποσότητα τους 20 μεγατόνους στο πέρας της πρώτης δεκαετίας του 21^{ου} αιώνα. Οικονομικά στατιστικά στοιχεία αποδίδουν το 30% των παραγόμενων τυροκομικών προϊόντων να προέρχεται από τις Η.Π.Α., καταδεικνύοντας τους ως το μεγαλύτερο παραγωγό. Ακολουθεί η Γερμανία και η Γαλλία. Την πρώτη θέση στον τομέα εξαγωγής τυριού καταλαμβάνει η Γαλλία, ενώ στον τομέα εισαγωγής, στην πρώτη θέση βρίσκεται η Γερμανία, ακολουθούμενη από το Ηνωμένο Βασίλειο και την Ιταλία. Τέλος η Ελλάδα βρίσκεται στην πρώτη θέση στην κατά κεφαλήν κατανάλωση τυριού με 27,3 κιλά κατ' άτομο, και ειδικότερα η φέτα να αποτελεί τα 3/4 της κατανάλωσης.

Η παγκόσμια παραγωγή γάλακτος άρχισε να αυξάνεται ραγδαία από το 1970, με στατιστικά στοιχεία του 2012, να δείχνουν ότι η παραγωγή αυξήθηκε από 480 μεγατόνους σε 790 μεγατόνους. Συγκεκριμένα, η παγκόσμια παραγωγή τυριών το 2011 έφτασε το 20 εκατομμύρια μεγατόνους ακολουθώντας ανοδική πορεία. Με την Ε.Ε. και τις Η.Π.Α. να παράγουν το 75% της παγκόσμια αγοράς.

Πίνακας 1: Δεδομένα Παγκόσμιας Παραγωγής Τυροκομικών προϊόντων¹

World cheese production 2000-2020

1,000 MT	2000	2012	2012/2000	2020 (prognosis)	2020/2012
EU-28	7,709	9,333	+21 %	10,606	+14%
Other Europe	266	291	+9%	338	+16%
CIS	448	866	+93%	1,072	+24%
North America	4,227	5,618	+22%	6,720	+20%
Oceania	665	700	+5%	930	+33%
South America	1,118	1,625	+45%	2,067	+27%
Asia	293	456	+55%	1,288	+182%
Middle East + Africa	744	1.512	+103%	2,054	+36%
Total World	15,470	20,401	+32%	25,075	+23%

1.1.3. Τυροκομία στην Ελλάδα

Στα μέσα του 20^{ου} αιώνα, παρουσιάζεται αύξηση της ίδρυσης σύγχρονων τυροκομείων στην Ελλάδα. Κύριο γνώρισμα του συγκεκριμένου κλάδου είναι ο μεγάλος αριθμός μικρών επιχειρήσεων που βρίσκονται σε όλη την έκταση της χώρας. Το αντικείμενο αποτελεί κατεξοχήν τμήμα του παραδοσιακού τομέα δραστηριότητας, με την εμφάνιση μεγαλύτερου βιομηχανικού χαρακτήρα τα τελευταία χρόνια. Παράμετρος που καθιστά ελκυστική την παραγωγή τυροκομικών προϊόντων, είναι οι κλιματολογικές συνθήκες που επικρατούν στην Ελλάδα και το γεωμορφολογικό ανάγλυφο των επιμέρους περιοχών. Κατά τη δεκαετία του 90', στον κλάδο της τυροκομίας εμφανίζονται μεγαλύτερες γαλακτοβιομηχανίες, με αποτέλεσμα την μείωση της παραγωγής σε μικρότερες τυροκομικές μονάδες. Το γεγονός αυτό έχει άμεσο αντίκτυπο στη διαφοροποίηση του μικροβιώματος από το οποίο αποτελούνται τα τυροκομικά προϊόντα και θα αναλυθεί στη συνέχεια του παρόντος πονήματος. Τα τυροκομικά προϊόντα παρουσιάζουν δυναμική προοπτική μεσοπρόθεσμα και μακροπρόθεσμα από τις αρχές του 21^{ου} αιώνα. Ακόμα και μετά τη δημοσιονομική κρίση στα τέλη της πρώτης δεκαετίας, η παραγωγή και κατανάλωση τυροκομικών προϊόντων επηρεάστηκε σε μεγάλο βαθμό αλλά παρουσίασε ανάκαμψη σε παγκόσμιο επίπεδο.

¹ <http://pmfood.dk/upl/9739/CHEESEARTICLE1.pdf>

Το 1992 με τον κανονισμό 2081/92 η Ε.Ε., θέσπισε για πρώτη φορά το καθεστώς για την προστασία των γεωγραφικών ενδείξεων και των ονομασιών προέλευσης των γεωργικών προϊόντων και τροφίμων. Το 2006, με σκοπό τη βελτίωση του συστήματος, οι κανονισμοί αντικαταστάθηκαν από τους 510/06 και 509/06, χωρίς να μεταβληθούν τα πεδία εφαρμογής τους.

Βασικός ορισμός που εφαρμόζεται στα γαλακτοκομικά και τυροκομικά προϊόντα είναι η «*Ονομασία Προέλευσης*». Βάσει του Υπουργείου Αγροτική Ανάπτυξης και Τροφίμων ορίζεται η ονομασία που ταυτοποιεί ένα προϊόν το οποίο:

1. Κατάγεται από συγκεκριμένο τόπο, περιοχή ή, σε εξαιρετικές περιπτώσεις, χώρα.
2. Η ποιότητα ή τα χαρακτηριστικά οφείλονται κυρίως ή αποκλειστικά στο ιδιαίτερο γεωγραφικό περιβάλλον που συμπεριλαμβάνει τους εγγενείς φυσικούς και ανθρώπινους παράγοντες.
3. Όλα τα στάδια της παραγωγής εκτελούνται εντός της οριοθετημένης γεωγραφικής περιοχής.

Ως «*γεωγραφική ένδειξη*» νοείται η ονομασία που ταυτοποιεί ένα προϊόν:

1. Το οποίο κατάγεται από συγκεκριμένο τόπο, περιοχή ή χώρα.
2. Του οποίου ένα συγκεκριμένο ποιοτικό χαρακτηριστικό, η φήμη ή άλλο χαρακτηριστικό μπορεί να αποδοθεί κυρίως στη γεωγραφική του προέλευση.
3. Του οποίου ένα τουλάχιστον από τα στάδια της παραγωγής εκτελείται εντός της οριοθετημένης γεωγραφικής περιοχής.

Για τα περισσότερα τυροκομικά προϊόντα που βρίσκονται στην αγορά μέχρι και σήμερα, δεν είναι γνωστή με βεβαιότητα η αρχική χρονική προέλευσή τους. Αυτή τη στιγμή στην Ελλάδα έχουν κατοχυρωθεί 23 Π.Ο.Π., τα οποία αποτυπώνονται στον ακόλουθο πίνακα.

Πίνακας 2: Π.Ο.Π. Τυριά παραγόμενα στην Ελλάδα²

A/A	ΠΡΟΙΟΝ ΟΝΟΜΑΣΙΑ	ΚΑΤΗΓΟΡΙΑ ΑΝΑΓΝΩΡΙΣΗΣ ΠΟΠ ή ΓΓΕ	ΔΗΜΟΣΙΕΥΜΕΝΕΣ ΠΡΟΔΙΑΓΡΑΦΕΣ
1	Ανεβατό	ΠΟΠ	313060/14.01.94 (ΦΕΚ 24/18.01.94) Επίσημη Εφημερίδα της Ε.Ε. L148/1996
2	Γαλοτύρι	ΠΟΠ	313031/11.01.94 (ΦΕΚ 8/11.01.94) Επίσημη Εφημερίδα της Ε.Ε. L148/1996
3	Γραβιέρα Αγράφων	ΠΟΠ	313045/14.01.94 (ΦΕΚ 16/14.01.94) 313045/14.01.94 (ΦΕΚ 101 Β'/16.02.94) Επίσημη Εφημερίδα της Ε.Ε. L148/1996
4	Γραβιέρα Κρήτης	ΠΟΠ	313047/14.01.94 (ΦΕΚ 16/14.01.94) Επίσημη Εφημερίδα της Ε.Ε. L148/1996
5	Γραβιέρα Νάξου	ΠΟΠ	Προδιαγραφές 313071/18.01.94(ΦΕΚ 23/18.01.94) 318849/21.08.08(ΦΕΚ 1725/28.08.08) Επίσημη Εφημερίδα της Ε.Ε. L148/1996
6	Καλαθάκι Λήμνου	ΠΟΠ	313044/14.01.94 (ΦΕΚ 16/14.01.94) Επίσημη Εφημερίδα της Ε.Ε. L148/1996
7	Κασέρι	ΠΟΠ	313027/11.01.94 (ΦΕΚ 8/11.01.94) 379116/19.07.2000 (ΦΕΚ 949 Β'/31.07.2000) Επίσημη Εφημερίδα της Ε.Ε. L148/1996 Επίσημη Εφημερίδα της Ε.Ε. L174/2000
8	Κατίκι Δομοκού	ΠΟΠ	313048/14.01.94 (ΦΕΚ 16/14.01.94) Επίσημη Εφημερίδα της Ε.Ε. L148/1996
9	Κεφαλογραβιέρα	ΠΟΠ	313032/11.01.94 (ΦΕΚ 8/11.01.94) Επίσημη Εφημερίδα της Ε.Ε. L148/1996
10	Κοπανιστή	ΠΟΠ	Προδιαγραφές Επίσημη Εφημερίδα της Ε.Ε. L148/1996 Επίσημη Εφημερίδα της Ε.Ε. C186/2012 Επίσημη Εφημερίδα της Ε.Ε. L129/2013
11	Λαδοτύρι Μυτιλήνης	ΠΟΠ	313058/17.01.94 (ΦΕΚ 25/18.01.94) Επίσημη Εφημερίδα της Ε.Ε. L148/1996
12	Μανούρι	ΠΟΠ	313028/11.01.94 (ΦΕΚ 8/11.01.94) Επίσημη Εφημερίδα της Ε.Ε. L148/1996
13	Μετσοβόνα	ΠΟΠ	313070/18.01.94 (ΦΕΚ 23/18.01.94) Επίσημη Εφημερίδα της Ε.Ε. L148/1996
14	Μπιάτζος	ΠΟΠ	313057/17.01.94 (ΦΕΚ 25/18.01.94) 313057/17.01.94 (ΦΕΚ 101 Β'/16.01.94) Επίσημη Εφημερίδα της Ε.Ε. L148/1996
15	Ξυνομυζήθρα Κρήτης	ΠΟΠ	313051/14.01.94 (ΦΕΚ 18/14.01.94) Επίσημη Εφημερίδα της Ε.Ε. L148/1996
16	Πηχτόγαλο Χανίων	ΠΟΠ	313062/17.01.94 (ΦΕΚ 24/18.01.94) Επίσημη Εφημερίδα της Ε.Ε. L148/1996
17	Σαν Μιχάλη	ΠΟΠ	313069/18.01.94 (ΦΕΚ 23/18.01.94) Επίσημη Εφημερίδα της Ε.Ε. L148/1996
18	Φέτα	ΠΟΠ	313025/11.01.94 (ΦΕΚ 8/11.01.1994) Επίσημη Εφημερίδα της Ε.Ε. L277/2002
19	Σφέλα	ΠΟΠ	313056/17.01.94 (ΦΕΚ 25/18.01.94) Επίσημη Εφημερίδα της Ε.Ε. L148/1996
20	Φορμαέλα Αράχωβας Παρνασσού	ΠΟΠ	313063/17.01.94 (ΦΕΚ 25/18.01.94) Επίσημη Εφημερίδα της Ε.Ε. L148/1996
21	Ξύγαλο Σητείας ή Ξίγαλο Σητείας	ΠΟΠ	Προδιαγραφές Επίσημη Εφημερίδα της Ε.Ε. C312/2010 Επίσημη Εφημερίδα της Ε.Ε. L200/2011
22	Κρασοτύρι Κω/ Τυρί της Πόσιας	ΠΓΕ	Προδιαγραφές Επίσημη Εφημερίδα της Ε.Ε. L332/23.12.2019 Επίσημη Εφημερίδα της Ε.Ε. C283/21.08.2019
23	ΑΡΣΕΝΙΚΟ ΝΑΞΟΥ ARSENICO NAXOU	ΠΟΠ	Προδιαγραφές Επίσημη Εφημερίδα ΕΕ L15/20.01.2020 Επίσημη Εφημερίδα ΕΕ C271/13.08.2019

² <http://www.minagric.gr/index.php/el/for-farmer-2/2012-02-02-07-52-07/ellinikaproionta/1270-tiria>

1.2. Στάδια Παραγωγής

1.2.1. Οργανικά και Ανόργανα Στοιχεία Γάλακτος - Τυριού

Τα τυροκομικά προϊόντα αποτελούνται από μικροβιακές κοινότητες, οι οποίες παρουσιάζουν αυξημένες διαφοροποιήσεις ανάλογα με το είδος των τυριών. Οι διαφοροποιήσεις αυτές ποικίλουν σε πολλές περιπτώσεις εντός του ίδιου δείγματος τυριού από τον πυρήνα έως της επιφάνεια του. Βασικός παράγοντας που επηρεάζει τη διακύμανση των διαφορετικών αποικιών μικροβιώματος είναι η διαδικασία παραγωγής του και τα στάδια ωρίμανσης. Η κατανόηση της πολύπλοκης σύνθεσης των μικροβιακών κοινοτήτων και η επίδραση τους στην ποιότητα, την ασφάλεια, τη διατροφική τους αξία και την ποικιλία που παρουσιάζουν είναι μείζονος σημασίας.

Αυτή η ετερογένεια των μικροοργανισμών που αναπτύσσεται στα διαφορετικά είδη τυριών διαδραματίζει ζωτικό ρόλο στην ανάπτυξη των οργανοληπτικών ιδιοτήτων του τυριού. Η μικροβιακή σύνθεση σε όλη τη ροή της διαδικασίας παραγωγής τυριού, διαφέρει από το αρχικό σημείο επεξεργασίας που είναι το γάλα, έως το τελικό προϊόν, που είναι το τυρί.

Το τυρί είναι προϊόν με υψηλή περιεκτικότητα σε λίπη. Εν συντομία, τα βασικά στοιχεία που εμπεριέχονται στο τυρί είναι τα ακόλουθα:

- Πρωτεΐνες με υψηλή βιολογική αξία και ομοιότητα με τις πρωτεΐνες που εμπεριέχονται στα ζωικά παράγωγα.
- Ανόργανα στοιχεία όπως ασβέστιο, φώσφορο, νάτριο, κάλιο και σε ελάχιστες ποσότητες σίδηρος. Οι ποσότητες των στοιχείων ποικίλουν ανάλογα με το είδος του τυριού.
- Βιταμίνες με κυρίαρχες τις λιποδιαλυτές, όπως Α, D, Ε και Κ, η ύπαρξη των οποίων οφείλεται στην υψηλή περιεκτικότητα σε λιπαρά οξέα.

Η πήξη του γάλακτος και η μετατροπή του σε τυρί με την αντίδραση της πυτιάς, τεχνική που με την πάροδο των χρόνων έχει συμπληρωθεί με πολλά στάδια επεξεργασίας, έχει ως αποτέλεσμα τη δημιουργία πολλών διαφορετικών ποικιλιών τυριού. Η παρασκευή των περισσότερων βασίζεται στην ενζυμική πήξη των καζεϊνών του γάλακτος [7]. Βασικές πρώτες ύλες για

την παραγωγή του τυριού είναι το γάλα, τα πηκτικά ένζυμα, οι οξυγαλακτικές καλλιέργειες, το αλάτι και οι πρόσθετες ύλες βάσει του επιθυμητού τελικού προϊόντος.

1.2.2. Παραγωγική Διαδικασία Τυριού

Το βασικό στοιχείο για την παραγωγή τυριών είναι το γάλα. Οι ποικιλίες γάλακτος διαφέρουν ανάλογα με το ζώο από το οποίο παράγονται, με πιο διαδεδομένο το αγελαδινό, και να ακολουθούν το πρόβειο, το κατσικίσιο, το βουβαλίσιο ή μείγμα διαφορετικών ειδών. Η παρασκευή γίνεται είτε από νωπό γάλα είτε από παστεριωμένο. Η διαδικασία της παστερίωσης γίνεται με σκοπό την καταστροφή παθογόνων μικροοργανισμών και την δημιουργία των κατάλληλων συνθηκών για τις καλλιέργειες εκκίνησης.

Η παραγωγική διαδικασία τυριών περιλαμβάνει τα ακόλουθα βασικά βήματα επεξεργασίας, τα οποία ανάλογα με το επιθυμητό τελικό προϊόν μπορούν να μεταβάλλονται, είτε παραλείποντας κάποιο βήμα, είτε με τροποποίηση και προσθήκη άλλων βημάτων:

1. Τυποποίηση γάλακτος με σκοπό τη ρύθμιση της σύστασης του και την ταυτοποίηση των κύριων συστατικών για την εξασφάλιση της τήρησης των επιτρεπόμενων τεχνολογικών προδιαγραφών για την επιτυχή παραγωγή τυριών [7].
2. Παστερίωση ή αλλιώς θερμική επεξεργασία για την καταστροφή ή αποφυγή ανάπτυξης παθογόνων μικροοργανισμών. Η παστερίωση γίνεται για 30 min στους 63°C ή για 15 sec στους 72°C. Στο συγκεκριμένο βήμα, συνδυαστικά με την θανάτωση των επιβλαβών μικροοργανισμών πραγματοποιείται και η καταστροφή βακτηρίων με σημαντική δράση στη γεύση των τυριών. Η πολυπλοκότητα του μικροβιώματος ελαχιστοποιείται και τα διαφορετικά είδη βακτηρίων καταστρέφονται. Τα τυριά που δέχονται ήπια θερμική επεξεργασία, πρέπει υποχρεωτικά να ωριμάσουν 60 ημέρες για να θεωρηθεί ασφαλής η κατανάλωση τους [8]. Στις περιπτώσεις μη αποστειρωμένου γάλακτος πραγματοποιείται ήπια θερμική επεξεργασία ώστε να καταστεί δυνατή η έναρξη της ζύμωσης.

3. Εμβολιασμός με εκκινητήρες και συμπληρωματικές καλλιέργειες μετά την ψύξη του γάλακτος με στόχο την ομαλή ανάπτυξη βακτηρίων και την εκκίνηση της ζύμωσης. Σε αυτό το σημείο, λαμβάνει χώρα η πτώση του pH και αναπτύσσεται η γεύση του τυριού.
4. Επόμενο βήμα είναι η πήξη του γάλακτος με τη δράση της πυτιάς, της οξύτητας ή και των δύο παραγόντων. Για το σχηματισμό του τυροπήγματος είναι απαραίτητη η υδρόλυση των πρωτεϊνών του γάλακτος, η οποία επιτυγχάνεται με την προσθήκη πυτιάς. Στην πυτιά εμπεριέχεται το πρωτεολυτικό ένζυμο χυμοσίνη ή αλλιώς ρενίνη που υδρολύει τις καζεΐνες του γάλακτος και σχηματίζεται το τυρόπηγμα [7]. Στη συνέχεια, για συγκεκριμένο χρονικό διάστημα παραμένει το τυρόπηγμα χωρίς κάποια διαταραχή, με στόχο τη σταθεροποίηση του.
5. Ακολουθεί διαίρεση του τυροπήγματος όταν το pH φτάσει στο επιθυμητό όριο ανάλογα με το είδος του τυριού [9]. Στη συνέχεια πραγματοποιείται τεμαχισμός του τυροπήγματος και αποβολή μέρους του τυρογάλαου με το καζεϊνικό δίκτυο να γίνεται συμπαγές σε μεγαλύτερο βαθμό [7].
6. Αποστράγγισή του ορού ή τυρογάλακτος από το πήγμα. Διαφαίνεται σταδιακά η μάζα που έχει δημιουργηθεί.
7. Ανάδευση τυροπήγματος με στόχο την αφαίρεση του τυρογάλακτος που δεν έχει αφαιρεθεί στο προηγούμενο βήμα [9].
8. Αναθέρμανση στις περιπτώσεις που επιδιώκεται έντονη στράγγιση του τυροπήγματος, ζητούμενο τις περισσότερες φορές στα σκληρά τυριά. Σταδιακή διακύμανση της θερμοκρασίας μεταξύ 32°C και 55°C [7].
9. Αλάτισμα με στόχο το τελικό προϊόν να περιέχει 1-4% αλάτι. Η τοποθέτηση γίνεται είτε επιφανειακά, είτε με προσθήκη υδατικού διαλύματος χλωριούχου νατρίου είτε με το συνδυασμό των δύο προαναφερθέντων μεθόδων.
10. Διαμόρφωση σχήματος του τυροπήγματος με στόχο τη μεταφορά και τη διατήρησή του.
11. Αποθήκευση και ωρίμανση για συγκεκριμένο χρονικό διάστημα, κατά το οποίο λαμβάνουν χώρα βιοχημικές και μικροβιακές διεργασίες που

αλλάζουν σταδιακά τη σύσταση και τη δομή του τυριού. Σε αυτό το σημείο διαμορφώνεται η γεύση και το άρωμα που επιδιώκεται.

Ουσιαστικά, το τυρί προέρχεται από μερική ή ολική πήξη του γάλακτος, με την προσθήκη και ενεργοποίηση της πυτιάς ή κατάλληλων πηκτικών μέσων όπως οξίνιση ή θερμική επεξεργασία. Έπειτα πραγματοποιείται μερική στράγγιση του ορού και διαχωρίζεται το στερεό τυρόπηγμα.

1.2.3. Είδη Τυριών

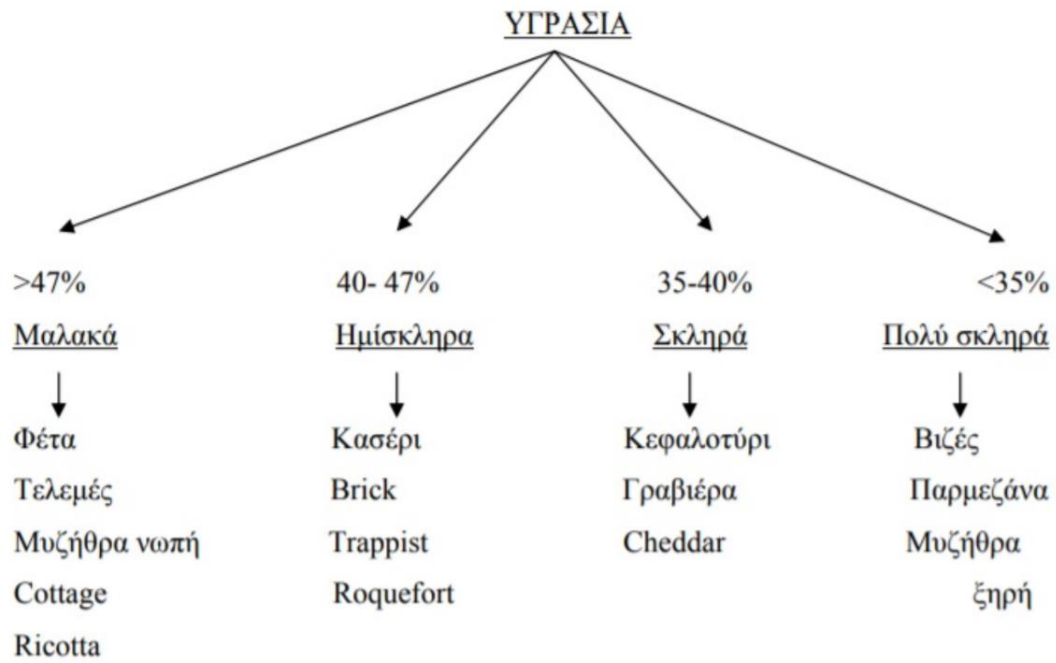
Βάσει του Κώδικα Τροφίμων, τα τυριά κατηγοριοποιούνται στις ακόλουθες διακριτές ομάδες:

1. Πολύ σκληρά
2. Σκληρά
3. Ημίσκληρα
4. Μαλακά

Τα κριτήρια που συμβάλλουν σε αυτή την κατηγοριοποίηση είναι η επί τοις % υγρασία στο άνευ λίπος τυρί, επί % λιποπεριεκτικότητα στην ξηρά ουσία του τυριού και τέλος ο τρόπος ωρίμανσής τους. Ειδικότερα, στα πολύ σκληρά τυριά η υγρασία δεν υπερβαίνει το 32%, στα σκληρά τυριά το 38%, στα ημίσκληρα τυριά το 46% και τέλος στα μαλακά τυριά το 58%.

Τέλος, διαφορετική κατηγορία καταλαμβάνουν τα τυριά που δεν υπόκεινται σε ωρίμανση και τα οποία αναγκαστικά πρέπει να παρασκευάζονται από παστεριωμένο γάλα.

Διάγραμμα 1: Κατάταξη τυριών με βάση την υγρασία [10]



1.3 Μικροοργανισμοί στα Τυριά

1.3.1. Μικροβίωμα Τυριού

Το μικροβίωμα του νωπού γάλακτος επηρεάζεται από το μικροβίωμα του μηρυκαστικού από το οποίο προέρχεται, και πιο συγκεκριμένα από το κανάλι της θηλής, την επιφάνεια του δέρματος, τις πρακτικές που χρησιμοποιούνται, τους χειριστές των οικόσιτων ζώων και τον εξοπλισμό που αξιοποιείται. Η προέλευση του γάλακτος φαίνεται να επηρεάζει εξίσου τα επίπεδα της ποικιλομορφίας, με το αγελαδινό γάλα να διαφοροποιείται σε μεγαλύτερο βαθμό από αυτό που παράγεται από αίγες και πρόβατα [11]. Ο τύπος του οικοσυστήματος βόσκησης που χρησιμοποιείται στις εκάστοτε περιπτώσεις, η χρονική διάρκεια και η δυναμική των μικροοργανισμών του δέρματος συντελούν στην διαμόρφωση της βιοποικιλότητας του μικροβιώματος στο νωπό γάλα. Έχουν μελετηθεί βακτήρια που επηρεάζουν τη γεύση, την οσμή, το χρώμα του νωπού γάλακτος όπως τα είδη *Brevibacterium linens*, *Staphylococcus equorum* και οξυγαλακτικά βακτήρια όπως *Lactococcus lactis*, *Lactococcus chungangensis*, *Lactobacillus casei/paracasei* τα οποία συμβάλλουν στο μεταβολισμό των πρωτεϊνών και των λιπαρών οξέων [12]. Η επεξεργασία μέσω της παστερίωσης του γάλακτος και το χαμηλό pH, συμβάλλουν επίσης στη διαφοροποίηση των μικροβιακών αποικιών, κάτι που συμβαίνει κυρίως στα τυριά που παράγονται από βιομηχανίες, τα οποία παρασκευάζονται με απουσία αρχικών βακτηρίων [13].

Επιπρόσθετοι παράγοντες όπως η περιεκτικότητα σε αλάτι, ο βαθμός ωρίμανσης, η προσθήκη συστατικών, όπως μπαχαρικά, επηρεάζουν εξίσου τη σύσταση του μικροβιώματος. Μελέτη ιρλανδικών τυριών βιομηχανικής παραγωγής, απέδειξε τη δημιουργία διαφοροποίησης μετά από την προσθήκη βοτάνων, καθώς σημειώθηκε μείωση της αναλογίας των γένων *Lactobacillus* και *Lactococcus*, ενώ η υψηλή περιεκτικότητα σε άλας κατέστειλε την ανάπτυξη βακτηρίων του γένους *Leuconostoc* και *Pseudomonas* [14] [15]. Υπάρχουν όμως και οι περιπτώσεις που το μικροβίωμα του νωπού γάλακτος δε διαφέρει σε μεγάλο βαθμό με το μικροβίωμα του τελικού προϊόντος, όπως στην περίπτωση του Βέλγικου Π.Ο.Π. τυριού Herve. Αρχικά η παρασκευή του γινόταν από το νωπό γάλα, χωρίς παστερίωση, αλλά λόγω της πιθανής

παρουσίας παθογόνου βακτηρίου *Listeria monocytogenes*, επιλέχθηκε η εισαγωγή του βήματος της παστερίωσης διότι βελτιώνει την ασφάλεια κατά των παθογόνων μικροοργανισμών. Το τελικό προϊόν μετά από ταυτοποίηση μέσω ανάλυσης 16S rDNA έδειξε παρόμοια σύσταση και στις δύο περιπτώσεις κατά 95% [16].

1.3.2. Κύριες καλλιέργειες εκκίνησης

Βακτηριακές καλλιέργειες, γνωστές και ως καλλιέργειες εκκίνησης, χρησιμοποιούνται στη βιομηχανία παραγωγής τυροκομικών προϊόντων σε ελεγχόμενες συνθήκες. Οι εκκινητές της διαδικασίας προστίθενται στο γάλα μετά από θερμική επεξεργασία ή παστερίωσή του. Τα κύρια στοιχεία των καλλιεργειών είναι τα οξυγαλακτικά βακτήρια που θα αναλυθούν στη συνέχεια. Τα βακτήρια ζυμώνουν τη λακτόζη του γάλακτος και τα προϊόντα της ζύμωσης διαμορφώνουν την οξύτητα (pH), τη συνεκτικότητα, την οσμή και την γεύση του τυριού.

Η σύσταση των καλλιεργειών εξαρτάται από το επιθυμητό προϊόν παραγωγής και κατηγοριοποιείται σύμφωνα με την προτιμώμενη θερμοκρασία ανάπτυξης των βακτηρίων, σε μεσόφιλα (20-30 °C) ή θερμόφιλα (40-45°C) και σύμφωνα με την παρουσία μονοκαλλιέργειας, δηλαδή με ένα μόνο στέλεχος ή μείγματα βακτηριακών ειδών που το καθένα αποδίδει διαφορετικά γνωρίσματα. Οι εκκινητές εξυπηρετούν στη μείωση της οξύτητας και τη διάσπαση των πρωτεϊνικών δεσμών [17]. Στις περισσότερες περιπτώσεις τα βακτήρια των καλλιεργειών εκκίνησης πεθαίνουν μετά την ωρίμανση αλλά τα προϊόντα τους συνεχίζουν να συμβάλλουν στην ωρίμανση του τυριού λόγω των ενζύμων που απελευθερώνουν κατά τη δράση τους. Τα πιο συνηθισμένα βακτήρια που χρησιμοποιούνται είναι: *L. lactis*, *Lactococcus cremoris*, *Streptococcus salivarius subsp. thermophilus*, *Lactobacillus delbruckii subsp. bulgaricus* και *Lactobacillus helveticus* [17].

Σε επέκταση της διαδικασίας, υπάρχουν περιπτώσεις που χρησιμοποιούνται δευτερεύουσες καλλιέργειες βακτηρίων, συνήθως μεσόφιλοι λακτοβάκιλλοι. Αυτό λαμβάνει χώρα για τον εμπλουτισμό της γεύσης, την επιτάχυνση της

ωρίμανσης σε περιπτώσεις που η χαμηλή υγρασία καθυστερεί τη διαδικασία, την προσθήκη συγκεκριμένων γευστικών χαρακτηριστικών σε τυριά με μειωμένη λιποπεριεκτικότητα και τέλος για την καταστολή της ανάπτυξης ανεπιθύμητων αποικιών και την αποφυγή της ανομοιομορφίας στην ποιότητα του τυριού [18].

1.3.3. Βακτήρια

1.3.3.1 Οξυγαλακτικά Βακτήρια

Τα οξυγαλακτικά βακτήρια συμβάλλουν περισσότερο από όλα τα είδη μικροοργανισμών στην παρασκευή του τυριού. Η έννοια των οξυγαλακτικών βακτηρίων (Lactic Acid Bacteria), ως διαφοροποιημένη ομάδα μικροοργανισμών, συναντάται για πρώτη φορά στις αρχές του 20^{ου} αιώνα. Βάσει ιστορικού αρχείου, γνωρίζουμε ότι η σχέση οξυγαλακτικών βακτηρίων με τα τρόφιμα μελετήθηκε για πρώτη φορά από τον Pasteur, με την πρώτη απομόνωση βακτηριακής καλλιέργειας του *Bacterium lactis* το 1873 από τον Lister [19]. Σε επόμενο χρονικό διάστημα, το 1890, ακολούθησε η αξιοποίηση των καλλιεργειών εκκίνησης για την παραγωγή τυριών και ξινογάλακτος. Ο αρχικός ορισμός των βακτηρίων βασίστηκε στην ικανότητά τους να δημιουργούν συνθήκες ζύμωσης και πήξης του γάλακτος, συμπεριλαμβανομένων και των κολοβακτηρίων (coliforms).

Ο ορισμός των οξυγαλακτικών βακτηρίων από τον Orla-Jensen (1919), που διατηρείται έως σήμερα, αναφέρει ότι αποτελούνται από μια ομάδα gram θετικών, μη σπορογόνων βακτηρίων, σε σχήμα ράβδου ή κόκκου που ζυμώνουν τους υδατάνθρακες και τις ανώτερες αλκοόλες παράγοντας γαλακτικό οξύ [20]. Τα οξυγαλακτικά βακτήρια ανήκουν στην τάξη *Lactobacillales* της κλάσης *Bacilli* του φύλλου *Firmicutes*. Έχουν χαμηλή περιεκτικότητα σε γουανίνη - κυτοσίνη (<53 mol %) στο DNA τους. Παρουσιάζουν υψηλή αντοχή σε χαμηλό pH κατά την πλειονότητα τους. Δεν διαθέτουν το ένζυμο καταλάση και είναι μικροαερόφιλα έως αναερόβια που μπορούν όμως να αναπτυχθούν με την παρουσία οξυγόνου.

Η ενέργεια που χρειάζεται για την ανάπτυξη των βακτηρίων, προέρχεται από το μεταβολισμό των σακχάρων. Αναπτύσσονται σε φυσικό περιβάλλον, είτε παρασιτούν σε ξενιστές αλληλεπιδρώντας με αυτούς ευεργετικά ή με αβλαβείς σχέσεις.

Τα οξυγαλακτικά βακτήρια είναι ετερογενής ομάδα που διαφοροποιείται μορφολογικά, στη φυσιολογία τους και στο μεταβολισμό τους. Τα γένη τους είναι τα ακόλουθα:

Lactobacillus, *Lactococcus*, *Leuconostoc*, *Oenococcus*, *Enterococcus*, *Tetragenococcus*, *Streptococcus*, *Pediococcus*, *Carnobacterium*, *Vagococcus*, *Weisella*, *Alloiococcus*, *Aerococcus* και *Symbiobacterium* [21].

Παρουσιάζονται και περιπτώσεις με τα γένη *Enterococcus* και *Streptococcus* που πολλά από αυτά ανήκουν στην κατηγορία των παθογόνων.

Τέλος, ορισμένα οξυγαλακτικά βακτήρια έχουν την ικανότητα να μεταβολίζουν το κιτρικό οξύ, που ανήκει στην κατηγορία των δευτερευόντων συστατικών του γάλακτος, και να το μετατρέπουν σε αρωματικές ενώσεις. Αυτός είναι και ο κύριος λόγος που αξιοποιούνται ως καλλιέργειες εκκίνησης.

Μια εναρκτήρια καλλιέργεια ορίζεται ως το μικροβίωμα που βρίσκεται σε μορφή παρασκευάσματος μεγάλου αριθμού κυττάρων, ενός ή πλήθους μικροοργανισμών. Το παρασκεύασμα προστίθεται ως πρώτη ύλη για την παραγωγή ζυμώμενης τροφής.

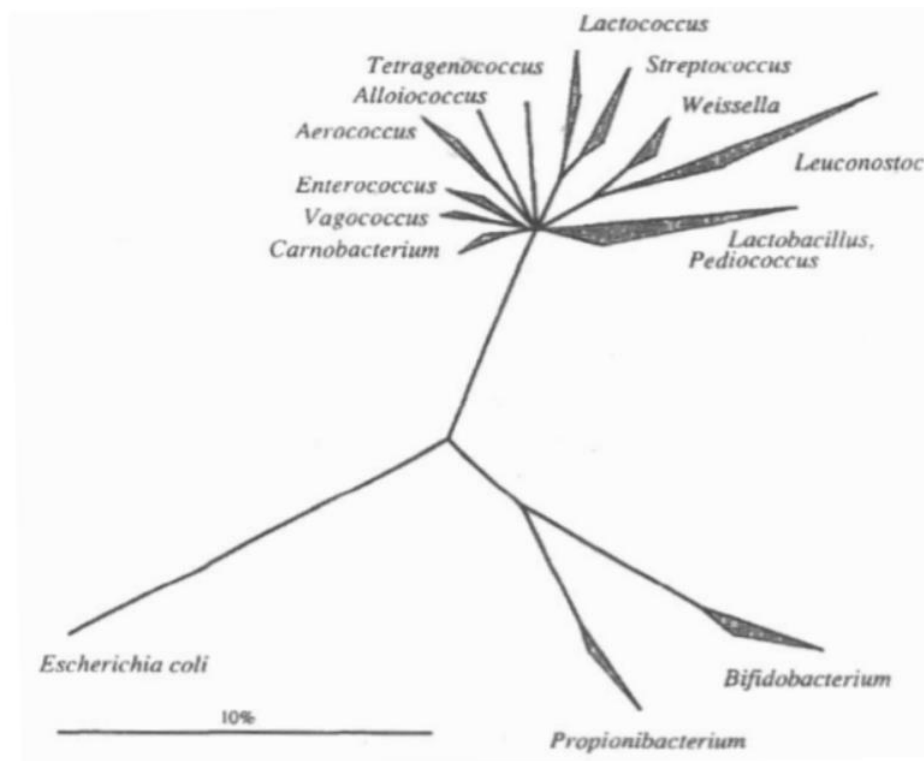
Χρησιμοποιείται σε όλα τα είδη τυριών με σκοπό τη ζύμωση της λακτόζης για την δημιουργία γαλακτικού οξέος, δηλαδή τη δημιουργία οξύτητας. Η οξύτητα ευνοεί τη δράση της πυτιάς και συμβάλει στην αποβολή του τυρογάλακτος από το τυρόπηγμα. Τα ένζυμα των οξυγαλακτικών βακτηρίων υδρολύουν το λίπος και τις καζεΐνες κατά τη διάρκεια της ωρίμανσης, συμβάλλοντας στη διαμόρφωση των τελικών ιδιοτήτων του τυριού [7]. Τα οξυγαλακτικά βακτήρια χρησιμοποιούν τη λακτόζη ως πηγή ενέργειας και σύμφωνα με τα κύρια προϊόντα της ζύμωσης και την παραγόμενη ποσότητα γαλακτικού οξέος ανά mol ζυμωμένης γλυκόζης, διακρίνονται σε ετεροζυμωτικά και ομοζυμωτικά. Τα ομοζυμωτικά είδη των οξυγαλακτικών βακτηρίων παράγουν αποκλειστικά γαλακτικό οξύ ως τελικό προϊόν, ενώ τα ετεροζυμωτικά είδη,

πέραν του γαλακτικού οξέος ως τελικό προϊόν, παράγουν διοξείδιο του άνθρακα και αιθανόλη [22].

Πίνακας 3: Είδη του γένους *Lactobacillus* με βάση τα προϊόντα ζύμωσης [23]

1η Ομάδα	2η Ομάδα	3η Ομάδα
Υποχρεωτικά ομοζυμωτικοί	Προαιρετικά ετεροζυμωτικοί	Υποχρεωτικά ετεροζυμωτικοί
<i>Lb. acidophilus</i>	<i>Lb. alimentarius</i>	<i>Lb. brevis</i>
<i>Lb. delbrueckii</i>	<i>Lb. casei</i>	<i>Lb. fermentum</i>
<i>Lb. helveticus</i>	<i>Lb. curvatus</i>	<i>Lb. kefir</i>
<i>Lb. kefirgranum</i>	<i>Lb. plantarum</i>	<i>Lb. parakefir</i>
<i>Lb. mali</i>	<i>Lb. sakei</i>	<i>Lb. reuteri</i>

Εικόνα 1: Γενεαλογικό δέντρο βασισμένο σε τεχνική που χρησιμοποιεί 16S rDNA και απεικονίζει τις φυλογενετικές σχέσεις μεταξύ των LAB. Η γραμμή υποδεικνύει 10% αναμενόμενη γραμμική απόκλιση [24]



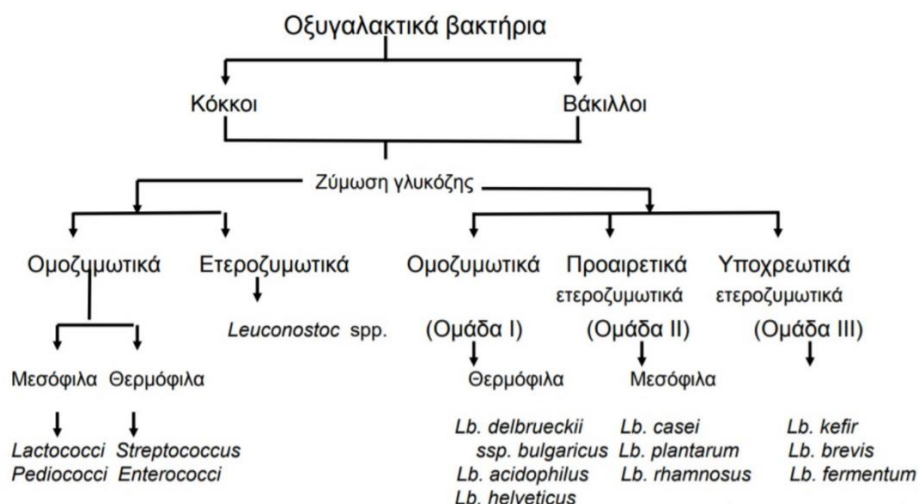
Απαντώνται σε πλούσια θρεπτικά υποστρώματα και αποτελούν μέρος της υγιούς μικροχλωρίδας του ανθρώπινου εντερικού σωλήνα. Από τα 12 γένη οξυγαλακτικών βακτηρίων, το γένος *Lactobacillus* είναι το πιο διαδεδομένο σε διάφορους τύπους τροφίμων.

Το είδος και ο συνδυασμός των οξυγαλακτικών βακτηρίων που θα προστεθούν ως καλλιέργεια εκκίνησης για την παραγωγή τυριών, εξαρτάται από τις συνθήκες της θερμοκρασίας στην επεξεργασία του τυροπήγματος, το ποσοστό άλατος και τις συνθήκες ωρίμανσης. Σε ειδικότερες περιπτώσεις υπάρχει η δυνατότητα να αξιοποιηθούν συγκεκριμένα στελέχη οξυγαλακτικών βακτηρίων, με ικανότητα παραγωγής συγκεκριμένων αρωματικών χαρακτηριστικών κατά τη διάρκεια της ωρίμανσης.

Συμπληρωματικά με την παραγωγή οξικού οξέος, σε πολλές περιπτώσεις παραγόμενα προϊόντα της ζύμωσης είναι οι αιθανόλες, οι αρωματικές ενώσεις, οι βακτηριοσίνες, οι εξωπολυσακχαρίτες και πλήθος ενζύμων που είναι ύψιστης σημασίας για τη βελτίωση της διάρκειας ζωής, τη μικροβιακή ασφάλεια, την υφή και το τελικό παραγόμενο αποτέλεσμα [25].

Στη διαδικασία δημιουργίας τυριού, τα πιο κοινά οξυγαλακτικά βακτήρια που επιλέγονται είναι *Lactobacillus*, *Streptococcus*, *Lactococcus*, *Leuconostoc* και *Enterococcus*.

Διάγραμμα 2: Διάκρισης Οξυγαλακτικών βακτηρίων βάσει της μορφής του κυττάρου, της ζύμωσης της γλυκόζης και της θερμοκρασίας ανάπτυξής τους [7]



1.3.3.2. Αντιμικροβιακές ιδιότητες των οξυγαλακτικών βακτηρίων

Πολλά είδη οξυγαλακτικών βακτηρίων δημιουργούν συνθήκες ανταγωνισμού έναντι των παθογόνων μικροοργανισμών που βρίσκονται στα τρόφιμα. Συγκεκριμένα, τα οξυγαλακτικά βακτήρια αποτρέπουν σε ορισμένο βαθμό την εγκατάσταση, τον πολλαπλασιασμό και την έκφραση παθογόνων μικροοργανισμών [26]. Μειώνουν το pH μέσω της παραγωγής πτητικών λιπαρών οξέων μικρής αλυσίδας, όπως το γαλακτικό και προπιονικό οξύ. Αυτό έχει ως αποτέλεσμα την καλύτερη εκμετάλλευση των θρεπτικών στοιχείων του περιβάλλοντα χώρου και την ποσότητα των θρεπτικών συστατικών ως τροφή για τους παθογόνους μικροοργανισμούς. Επίσης μειώνεται το οξειδοαναγωγικό δυναμικό του υποστρώματος, παράγεται υπεροξείδιο του υδρογόνου (H_2O_2), και τέλος παράγονται συγκεκριμένες ανασταλτικές ενώσεις όπως οι βακτηριοσίνες [26]. Ουσιαστικά, θα μπορούσαμε να αναφέρουμε ότι τα οξυγαλακτικά βακτήρια είναι υπεύθυνα για τη βιοδιατήρηση του προϊόντος, παράγοντας ένα ευρύ φάσμα αντιμικροβιακών μεταβολιτών. Τα οργανικά οξέα μεταβάλλουν το δυναμικό της κυτταρικής μεμβράνης διακόπτοντας την ενεργή μεταφορά, μειώνοντας το ενδοκυτταρικό pH και αναστέλλοντας λειτουργίες μεταβολισμού [27].

1.3.3.3. Βακτηριοσίνες

Ένα από τα μεταβολικά προϊόντα των οξυγαλακτικών βακτηρίων είναι οι βακτηριοσίνες. Οι βακτηριοσίνες είναι πρωτεϊνικές ουσίες που δύναται να αναστέλλουν την ανάπτυξη παθογόνων μικροοργανισμών, όπως τα γένη *Listeria*, *Clostridium*, *Staphylococcus*, *Bacillus sp.* και *Enterococcus sp.* [28].

Το 1988 δόθηκε από τον ερευνητή Klaenhammer ο ορισμός για τις βακτηριοσίνες, ως πρωτεΐνες ή ομάδες πρωτεϊνών, που συμβάλουν στην εξόντωση των βακτηρίων, και στοχεύουν κυρίως σε είδη μικροβιώματος που τις παράγουν. Χρησιμοποιούνται ευρέως στη βιομηχανία τροφίμων, καθώς καθιστούν δυνατή την αναστολή ή τη μείωση της ανάπτυξης αλλοιογόνων και παθογόνων μικροοργανισμών.

Ένα παράδειγμα βακτηριοσίνης είναι η νισίνη, η οποία παράγεται από το είδος *L. lactis*, που βρίσκεται στο μικροβίωμα των τυριών, και παρουσιάζει δράση εναντίων αρκετών θετικών κατά Gram βακτηρίων, όπως του γένους *Listeria*.

1.3.3.4. Παθογόνα Βακτήρια στα Τυριά

Σε αρκετά σημεία της επεξεργασίας του γάλακτος για την παρασκευή τυριών, δημιουργούνται συνθήκες που ευνοούν την ανάπτυξη παθογόνων μικροοργανισμών. Ειδικότερα, σε περιπτώσεις στις οποίες χρησιμοποιείται μη παστεριωμένο γάλα, αυξάνεται η πιθανότητα για ενδεχόμενες μολύνσεις.

Μια από τις ομάδες μικροοργανισμών που θεωρείται επιβλαβής είναι τα κολοβακτηρίδια. Συναντώνται σε ομάδες εντεροβακτηριοειδών, με βασική ιδιότητά τους τη ζύμωση της λακτόζης και την παραγωγή οξέος και αερίου. Είναι αρνητικά κατά Gram βακτήρια, αερόβια ή προαιρετικώς αερόβια και ασπορογόνα. Η παρουσία τους υποδηλώνει μόλυνση των τροφίμων από το εντερικό περιεχόμενο ζώου. Είναι δείκτης της κατάστασης ενός τροφίμου, καθώς η παρουσία του υποδεικνύει μόλυνση.

Επόμενη ομάδα παθογόνου βακτηρίου είναι η *E.coli*. Εμφανίζεται συχνά σε νωπά τρόφιμα ζωικής προέλευσης ή με ατελώς επεξεργασμένη μορφή. Η εντεροτοξινογόνος *E.coli* μπορεί να προκαλέσει σοβαρά συμπτώματα σε οξεία και αιφνίδια μορφή στον άνθρωπο. Δεν αναπτύσσεται με ευκολία κατά τη διάρκεια της επεξεργασίας στην τυροκομία, διότι το χαμηλό pH και το άλας λειτουργούν ανασταλτικά. Υπάρχουν όμως περιπτώσεις στις οποίες η γαλακτική καλλιέργεια δεν είναι αρκετά δραστική, με αποτέλεσμα την επιβίωση της *E.coli* κατά την τυροκόμηση [29].

Συνεχίζοντας, θα ακολουθήσει αναφορά στη Σαλμονέλα. Ανήκει στην Οικογένεια των εντεροβακτηριοειδών (*Enterobacteriaceae*), είναι αρνητικά κατά Gram βακτήρια, μικρών διαστάσεων 0.6x2 μm, με χαρακτηριστικό τις περίτριχες βλεφαρίδες, αερόβιο ή προαιρετικώς αναερόβιο, μη σπορογόνο που δε διαθέτει έλυτρο με εξαίρεση τα είδη *S.pullorum*, *S.gallinarum* που είναι ακίνητες και τα είδη *S.typhi*, *S.paratyphi* και *S.dublin* που παράγουν ειδικό περίβλημα που μοιάζει με έλυτρο [30]. Οι συνθήκες που επιτρέπουν την

ανάπτυξη της σαλμονέλας είναι 5.3°C – 5.4°C με βέλτιστη στους 35°C – 37 °C και pH 4.5 – 9.5 και βέλτιστο εύρος 6.5 – 7.5. Ομαδοποιείται στα πιο επικίνδυνα βακτήρια που μπορούν να προκαλέσουν λοιμώξεις, και για αυτόν το λόγο έχουν ορισθεί κανόνες για την αποφυγή της ανάπτυξης και του πολλαπλασιασμού τους. Τα τυριά δεν ανήκουν στα προϊόντα που είναι επικίνδυνα για σαλμονέλωση. Ωστόσο η πρόληψη είναι η βέλτιστη αποφυγή τέτοιων περιπτώσεων, καθώς χρησιμοποιείται νωπό γάλα για την παραγωγή μαλακών τυριών.

Επόμενο βακτήριο ανήκει στην κατηγορία των σταφυλόκοκκων, και αναφερόμαστε στο είδος *S.aurens*. Μεγάλος αριθμός κυττάρων του είδους οδηγεί στην παραγωγή εντεροτοξίνης. Στο στάδιο της ωρίμανσης παρουσιάζεται σημαντική μείωση των πληθυσμών, αλλά σε αρκετές περιπτώσεις η εντεροτοξίνη παραμένει στο τυρί.

Τέλος, θα αναφερθούμε στο βακτήριο που ανήκει στην Οικογένεια *Listeriaceae* και είναι το είδος *Lis. monocytogenes*. Αναπτύσσεται στα γαλακτοκομικά προϊόντα και συνεχίζει την αναπαραγωγή του σε τυροκομικά προϊόντα που έχουν παραχθεί από μη παστεριωμένο γάλα. Μπορεί να προκαλέσει ήπια γρίπη με γαστρεντερικά συμπτώματα έως και σηψαιμία, μηνιγγίτιδα ή εγκεφαλίτιδα σε ηλικιωμένους και μικρά παιδιά. Η υψηλή συχνότητα εμφάνισης πληθυσμών του είδους σε μαλακά τυριά και σε τυριά που ωριμάζουν παρουσία μυκητών είναι επιβεβαιωμένη από αρκετές μελέτες σε Η.Π.Α., Γαλλία, Ιταλία, Δανία, Κύπρο, Ισπανία, Ελβετία και Γερμανία [31]. Έχει διαπιστωθεί ότι το βακτήριο αναπτύχθηκε κατά τη βιομηχανική παρασκευή σε συγκεκριμένα είδη τυριών, όπως η Φέτα, και σχετίζεται με την αύξηση του pH κατά την ωρίμανση.

1.3.4. Μύκητες

Μύκητες που συμμετέχουν στο μικροβίωμα των τυριών είναι νηματώδεις οργανισμοί, εκτός από κάποιους κατώτερους μύκητες, που ονομάζονται υφές. Η αναπαραγωγή τους πραγματοποιείται με σπόρια και τα αναπαραγωγικά τους όργανα σχηματίζονται όταν στο περιβάλλον που ευδοκιμούν αναπτυχθεί το φυτικό μυκήλιο, σε βαθμό που να επιτρέπει την παραγωγή ποσότητας πρωτοπλάσματος για την εξασφάλιση της τροφής τους. Παρουσιάζουν ανθεκτικότητα σε ξηρό περιβάλλον, μικρό μέγεθος και μεγάλο ρυθμό αναπαραγωγής.

Στην παραγωγή συγκεκριμένων τύπων τυριού, αξιοποιούνται επιλεγμένα στελέχη, κυρίως του γένους *Penicillium*. Σε πολλές περιπτώσεις, η μη επιδιωκόμενη παρουσία μυκήτων, μπορεί να προκαλέσει ανεπιθύμητες μεταβολές όπως η υποβάθμιση της ποιότητας του τυριού. Η αρχική ένδειξη αλλοίωσης είναι η εμφάνιση αποικιών από μύκητες στην επιφάνεια των τυριών, που συνοδεύεται από ανωμαλίες στη μεταβολή του χρώματος και της οσμής.

Παραδείγματα μυκητών που βρίσκονται σε συγκεκριμένα είδη τυριών είναι τα ακόλουθα : *P. camembertii* και *P. caseicolum* για την παραγωγή Camembert και Brie, *P. roquefortii* για την ωρίμανση των τυριών Roquefort και Gorgonzola.

Στη παρούσα μελέτη, τα τυριά που αναλύθηκαν δεν ανήκουν στις κατηγορίες τυριών που χρησιμοποιούν μυκητιακές καλλιέργειες. Στην ανάλυση επιλέχθηκαν συγκεκριμένες προσεγγίσεις βιοπληροφορικής, για την εξακρίβωση της παρουσίας μυκητών στα δείγματα.

1.3.5. Ζύμες

Οι ζύμες είναι μονοκύτταροι οργανισμοί που ανήκουν στο Βασίλειο των μυκήτων. Η διαδικασία της ζύμωσης μετατρέπει τους υδατάνθρακες σε διοξείδιο του άνθρακα και αλκοόλη. Ανήκουν κυρίως στο Φύλο των Ασκομυκήτων ή Βασιδιομυκήτων, με τους περισσότερους να είναι αερόβιοι.

Οι ζύμες που απαντώνται στα γαλακτοκομικά προϊόντα ενώ είναι ευαίσθητες στις συνθήκες παστερίωσης, συναντώνται συχνά λόγω του μολυσμένου περιβάλλοντος των τυροκομείων. Μεγάλης βαρύτητας θεωρούνται ζύμες που συμβάλλουν στη διαδικασία ζύμωσης της λακτόζης, οι οποίες σε συγκεκριμένες περιπτώσεις μπορούν να αξιοποιηθούν ως καλλιέργειες για την παρασκευή των γαλακτοκομικών προϊόντων. Ορατά χαρακτηριστικά μόλυνσης από ζύμες θεωρούνται οι αλλοιώσεις στην οσμή, η παραγωγή αερίων και οι μεταβολές του χρώματος στο τυριού.

Παραδείγματα από ζύμες που αναπτύσσονται συμβιωτικά με λακτοβάκιλλους οξυγαλακτικών καλλιιεργειών και με το είδος *Streptococcus thermophilus* είναι η *Candida krusei*. Η συμβιωτική τους ανάπτυξη βασίζεται στην ικανότητά τους να οξειδώνουν το γαλακτικό οξύ και να μειώνουν την πίεση οξυγόνου. Το τελικό αποτέλεσμα είναι η δημιουργία κατάλληλων συνθηκών στο περιβάλλον που βρίσκονται για την ανάπτυξη μικροοργανισμών [32].

Εμφάνιση και ανάπτυξη ζυμών, παρουσιάζεται στην επιφάνεια των τυριών λόγω του χαμηλού pH, της σχετικά χαμηλής υγρασίας, της χαμηλής θερμοκρασίας ωρίμανσης και της υψηλής συγκέντρωσης άλατος.

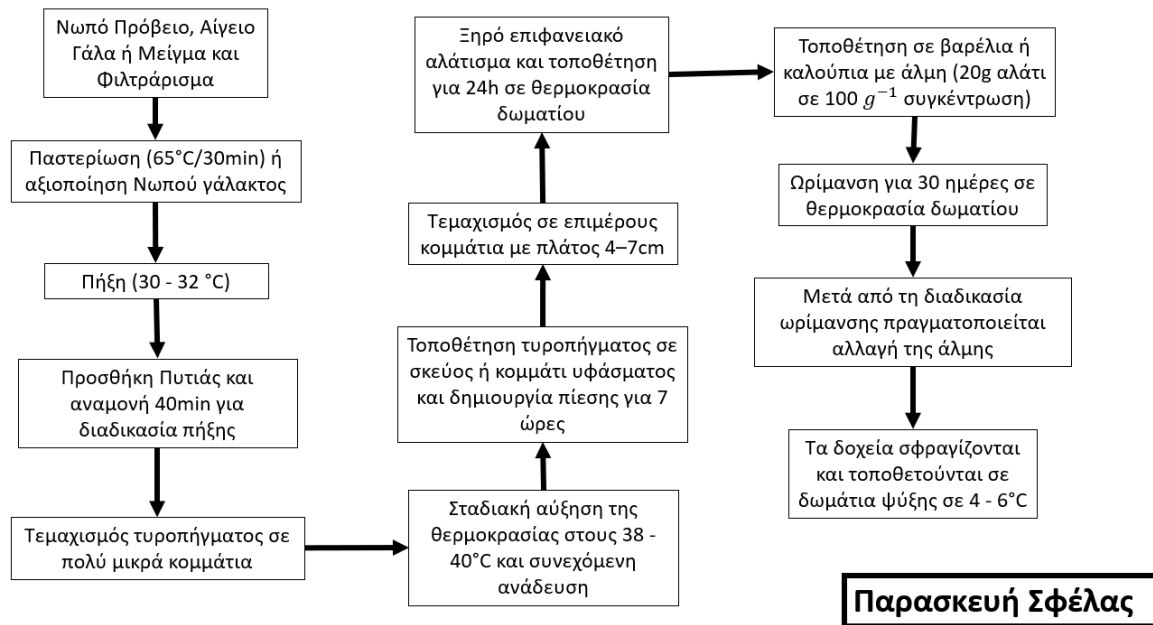
1.4 Τυριά που θα αναλυθούν στην Μελέτη

1.4.1. Σφέλα

Η Σφέλα είναι ημίσκληρο τυρί άλμης, με πολλές μικρές οπές στη μάζα του σώματός της, χρώματος λευκοκίτρινου, χωρίς επιδερμίδα, και σχήματος παραλληλεπίπεδων λωρίδων. Το γάλα που επιλέγεται για την παραγωγή της είναι πρόβειο, αίγαιο ή μείγμα των δύο ειδών. Η μέγιστη υγρασία που διαθέτει είναι 45% και η ελάχιστη λιποπεριεκτικότητα επί ξηρού βάρους 40%. Ο χρόνος ωρίμανσης της είναι τουλάχιστον τρεις μήνες και η παραγωγή της προέρχεται από τους Νομούς Μεσσηνίας και Λακωνίας. Η ονομασία που έχει δοθεί προέρχεται από τον τρόπο τεμαχισμού της τυρομάζας, καθώς σφέλα σημαίνει λωρίδα. (Υπουργείο Αγροτικής Ανάπτυξης & Τροφίμων, 2006). Αναγνωρίστηκε ως Π.Ο.Π. βάσει ΦΕΚ 25/18.01.94, με την υπ' αριθμό 3130568/17.01.94 Απόφαση Υφυπουργού Γεωργίας, στην οποία ορίζεται η περιοχή προέλευσης της, με τις φυλές αίγων και προβάτων να είναι αποκλειστικά εκτρεφόμενες με παραδοσιακό τρόπο από τη χλωρίδα της ίδιας περιοχής. Απαγορεύει ρητά τη προσθήκη σκόνης ή συμπυκνωμένου γάλακτος, και επιτρέπει τη προσθήκη παραδοσιακής πυτιάς ή ενζύμων σε ανάλογη ποσότητα.

Η διαδικασία παρασκευής του τυριού γίνεται είτε από παστεριωμένο είτε από νωπό γάλα και μετά το πέρας της επεξεργασίας υπάρχει δυνατότητα διατήρησης για δύο χρόνια με την προϋπόθεσή της κάλυψης του προϊόντος από άλμη για όλο το χρονικό διάστημα. Το βέλτιστο pH είναι 4.74 και η κύρια χημική σύσταση του σε συγκέντρωση ($g\ 100\ g^{-1}$) είναι η ακόλουθη: υγρασία 41.2, FDM 46.2 και SM 10.2. Ένα από τα βασικά χαρακτηριστικά του τυριού είναι η αυξημένη αλατότητα, η οποία μπορεί να μειωθεί μέσω του βήματος της παστερίωσης και τη χρήση συγκεκριμένων καλλιεργειών εκκίνησης [9].

Διάγραμμα 3: Ροή επεξεργασίας τυριού Σφέλα [33]



1.4.2. Χαλίτζι

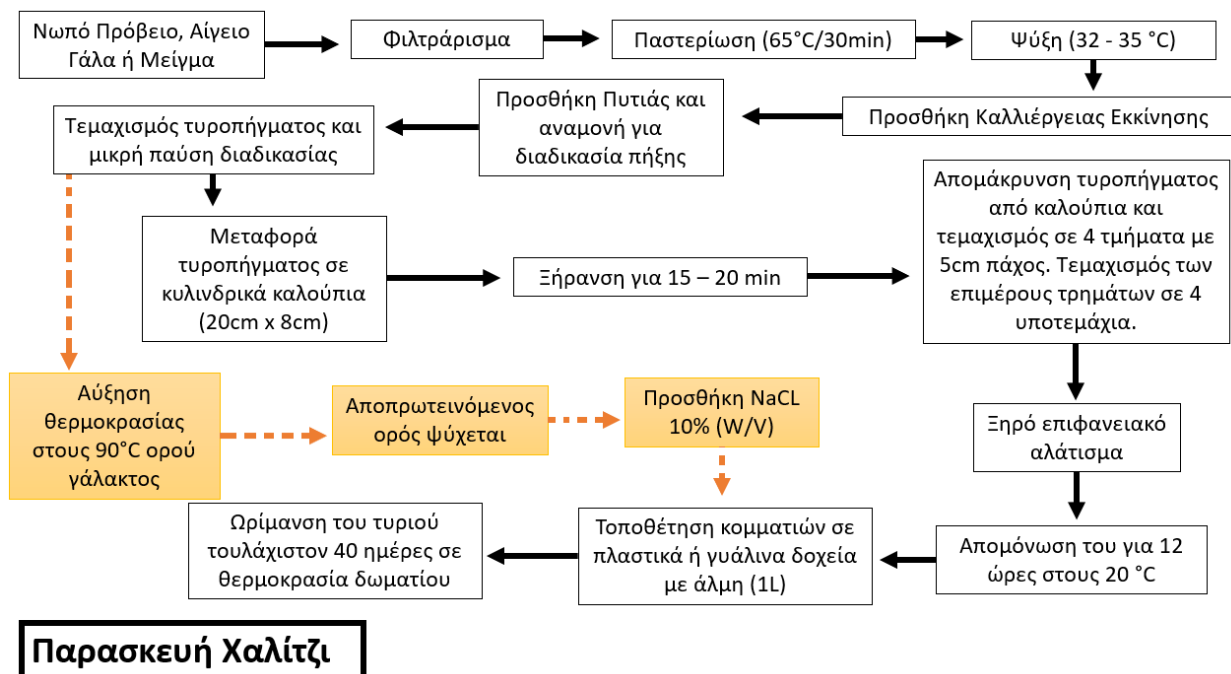
Τα Χαλίτζια ή Χαλίτσα Τηλλυρίας είναι προϊόν που παράγεται σε μικρά τοπικά τυροκομεία της ομώνυμης περιοχής, όπως και στην περιοχή της Τσακίστρα και σε συγκεκριμένες περιοχές της Κύπρου. Η ονομασία Χαλίτζι προέρχεται από τη λέξη χαλίκι, καθώς η οπτική μορφή του τυριού παραπέμπει σε κομμάτια που μοιάζουν με βότσαλα ή αλλιώς χαλίκια. Έχουν πραγματοποιηθεί ελάχιστες έρευνες για το συγκεκριμένο τυρί από το Γεωπονικό Πανεπιστήμιο Αθηνών (Γ.Π.Α.) και το Τεχνικό Πανεπιστήμιο Κύπρου (ΤΕ.ΠΑ.Κ.). Στόχος των ερευνών είναι η κατοχύρωση του τυριού ως προϊόν Προστατευόμενης Ονομασίας Προέλευσης.

Το Χαλίτζι είναι είδος ημίσκληρου τυριού, με λευκό χρώμα, διαμορφωμένες οπές χωρίς επιδερμίδα, γεύση υπόξινη, ελαφρώς αλμυρή και οσμή εξίσου υπόξινη. Παρασκευάζεται από νωπό γάλα, πρόβειο, αίγαιο ή μείγμα αυτών, το οποίο αξιοποιείται άμεσα μετά την άμελξη του.

Λίγα στοιχεία που γνωρίζουμε για την παρασκευή του τυριού είναι από το βιβλίο «Το γάλα και τα γαλακτοκομικά προϊόντα της Κύπρου με ιδιαίτερη αναφορά στο τυρί Χαλούμι» του Σωτήρη Οικονομίδη. Η διαδικασία παραγωγής

του διαθέτει ομοιότητες με τη διαδικασία παραγωγής του τυριού Χαλούμι. Για την πήξη του παστεριωμένου γάλακτος γίνεται θέρμανση στους 32 °C – 35 °C, στη συνέχεια προστίθεται πυτιά και παραμένει σε αυτή τη θερμοκρασία για 20 – 45 λεπτά. Σε αρκετές περιπτώσεις η θερμοκρασία πήξης, ανάλογα με τη περίοδο του χρόνου, ποικίλει διότι συσχετίζεται με την ποιότητα του γάλακτος. Το τυρόπηγμα τοποθετείται σε καλούπια, στα οποία στη συνέχεια μειώνεται σταδιακά το μέγεθος με στόχο την απομάκρυνση του γάλακτος. Τέλος, αφού απομακρυνθεί από τα καλούπια, κόβεται σε λεπτές λωρίδες 2 cm, τα κομμάτια αλατίζονται και αφήνεται σε θερμοκρασία δωματίου με στόχο τη μείωση της θερμοκρασίας του. Την επόμενη ημέρα, τα κομμάτια τοποθετούνται σε αλατισμένο τυρόγαλα, στο οποίο μπορούν να διατηρηθούν για ένα χρόνο. Ο χρόνος ωρίμανσης ποικίλει ανάλογα με τη θερμοκρασία. Σε χαμηλές θερμοκρασίες απαιτούνται μέχρι 40 ημέρες, ενώ σε υψηλότερες χρειάζεται λιγότερο διάστημα.

Διάγραμμα 4: Ροή επεξεργασίας τυριού Χαλίτζι [34]



1.5 Αλληλούχιση Επόμενης Γενιάς (Next Generation Sequence)

1.5.1. Θεωρητικό Πλαίσιο Αλληλούχισης Επόμενης Γενιάς

Η διερεύνηση της μικροβιακής ποικιλότητας σε αρχικό στάδιο, μπορεί να πραγματοποιηθεί μέσω μοριακών αναλύσεων. Η αρχική τεχνική αλληλούχιση Sanger δεν μπορούσε να ικανοποιήσει τις ανάγκες της έρευνας του μικροβιώματος σε άμεσο χρονικό διάστημα και με μειωμένο κόστος. Η Αλληλούχιση Επόμενη Γενιάς (NGS) είναι μια σχετικά πρόσφατη τεχνολογία με την οποία καθίσταται εφικτός ο προσδιορισμός της αλληλουχίας DNA με υψηλή απόδοση και σε μικρό χρονικό διάστημα. Επίσης, επιτρέπει την αλληλούχιση μεγαλύτερων τμημάτων DNA ή ακόμα και ολόκληρου του γονιδιώματος με σχετικά μικρό κόστος. Τα πρώτα εμπορικά μηχανήματα αλληλούχισης επόμενης γενιάς έγιναν διαθέσιμα το 2005, γεγονός που οδήγησε σε ταχεία αύξηση του ρυθμού παραγωγής γονιδιωματικών δεδομένων [35].

Παραδείγματα συσκευών αλληλούχισης επόμενης γενιάς είναι : 454 Pyrosequencing (Roche Diagnostics GMBH Ltd, Mannheim Germany), Illumina MiSeq και HiSeq (Illumina), Applied Biosystems – Life Technologies (SOLiD) και Ion Torrent – Life Technologies (Ion Proton, Personal Genome Machine). Κάθε μια από αυτές χρησιμοποιεί διαφορετική τεχνική και τρόπο ανίχνευσης των βάσεων της αλληλουχίας. Το κοινό χαρακτηριστικό και των τεσσάρων συσκευών είναι η παράλληλη αλληλούχιση πολλών μορίων DNA και στη συνέχεια η συναρμολόγηση των μικρών τμημάτων μέσω χρήσης εξελιγμένων αλγόριθμων βιοπληροφορικής. Με την αξιοποίηση των τεχνολογιών αλληλούχισης επόμενης γενιάς, μεγάλος αριθμός τμημάτων DNA μπορεί να αναλυθεί ταυτόχρονα στην ίδια αντίδραση, που ονομάζεται μαζική παράλληλη αλληλούχιση.

Στη μελέτη επιλέχθηκε η εταιρεία Illumina και χρησιμοποιήθηκαν τρεις διαφορετικές μεταγονιδιωματικές προσεγγίσεις, αλληλούχιση γονιδιωματικής περιοχής 16S rDNA και περιοχής ITS (Internal Transcribed

Spacer) και η τεχνική Shotgun (Τυχαίας Προσπέλασης) που επιτρέπει την μερική συναρμολόγηση μεγάλων αλληλουχιών.

Πίνακας 4: Πλεονεκτήματα και Μειονεκτήματα NGS

NGS	Πλεονεκτήματα	Μειονεκτήματα
Μειωση Κόστους Αλληλούχισης/Δείγμα	•	
Αυξηση Απόδοσης	•	
Αυξηση Ακρίβειας λόγω αύξησης επαναλήψεων	•	
Μείωση Χρόνου Αλληλούχισης	•	
Υπολογιστική Δύναμη		•
Μεγάλα Σύνολα Δεδομένων (δυσκολία διαχείρισης τους)		•
Εξειδικευμένες Γνώσεις		•
Εξειδικευμένα Εργαλεία Βιοπληροφορικής		•

1.5.2. Διαδικασία Αλληλούχισης Επόμενης Γενιάς

1.5.2.1. Βασικά βήματα αλληλούχισης

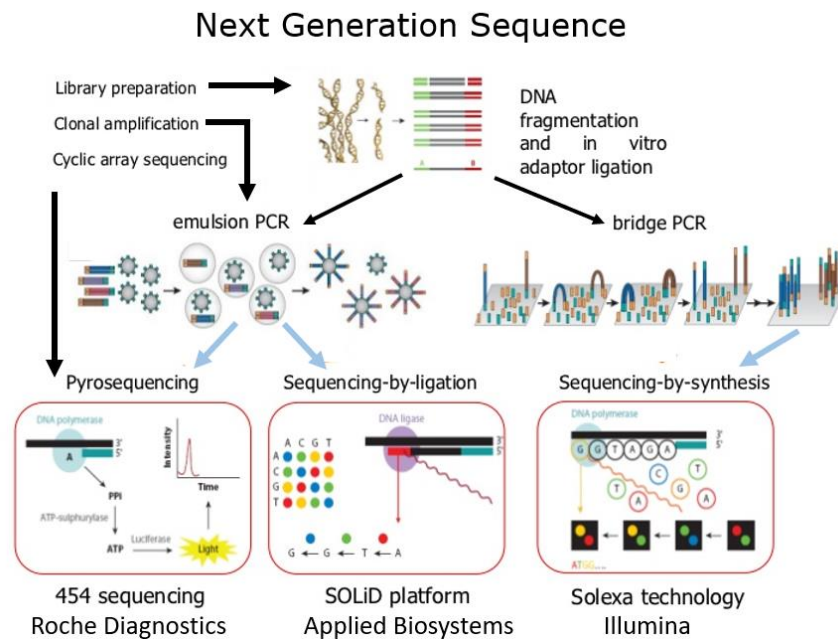
Η διαδικασία της μαζικής παράλληλης αλληλούχισης ξεκινά με την επιλογή του μορίου στόχου το οποίο μπορεί να είναι ολόκληρο το γονιδίωμα ενός οργανισμού, αλληλουχίες που κωδικοποιούν γονίδια – exome, επιλεγμένες περιοχές οι οποίες έχουν ενισχυθεί μέσω PCR, όπως η γονιδιωματική αλληλουχία των γονιδίων που η έκφραση τους εμπλέκεται σε συγκεκριμένη λειτουργία του οργανισμού, μόρια mRNA ενός οργανισμού ή ιστού, είτε το σύνολο των μορίων RNA όπως micro RNA, siRNA ενός οργανισμού ή ιστού [36].

Επόμενο βήμα είναι η απομόνωση του DNA ή RNA και ο τεμαχισμός του σε μικρά τμήματα. Πραγματοποιείται πρόσδεση γνωστών αλληλουχιών (adapters) στα άκρα των θραυσμάτων τα οποία επιτρέπουν στη συνέχεια την ενίσχυσή τους.

Στο σημείο αυτό πραγματοποιείται η δημιουργία βιβλιοθηκών που θα βοηθήσει στη συνέχεια την ταυτοποίηση των τμημάτων DNA ή RNA [37]. Για να δημιουργηθούν οι βιβλιοθήκες χρησιμοποιείται τις περισσότερες φορές, ένα στερεό υπόστρωμα, συνήθως μια γυάλινη επιφάνεια ή ένα σφαιρίδιο. Έπεται το βήμα της αλληλούχισης, μετά την πρόσδεση και την ενίσχυση των θραυσμάτων DNA ή RNA. Η διαδικασία την αλληλούχισης διαφέρει ανάλογα με την τεχνική που έχει επιλεγεί. Τα βασικά στάδια σε γενικότερο πλαίσιο είναι τα ακόλουθα : προσθήκη ενός διαφορετικού νουκλεοτιδίου κάθε φορά, ανίχνευση του νουκλεοτιδίου που προσδέθηκε στα θραύσματα με οπτική ή ηλεκτρονική μέθοδο, έκπλυση των νουκλεοτιδίων που δεν έχουν προσδεθεί και κυκλική επανάληψη των τριών βημάτων μέχρι τη στιγμή της αλληλούχισης όλου του τμήματος DNA ή RNA [38].

Τέλος, για την ανάλυση των ψηφιακών δεδομένων αλληλούχισης, χρησιμοποιούνται εργαλεία βιοπληροφορικής με εξειδικευμένους αλγόριθμους, αναλόγως το σκοπό και στόχο της ανάλυσης. Η τεχνική αλληλούχισης που πραγματοποιήθηκε για τα δείγματα της μελέτης είναι sequencing by synthesis που θα αναλυθεί εκτενέστερα στη συνέχεια.

Διάγραμμα 5: Διαφορετικές τεχνικές αλληλούχισης επόμενης γενιάς³



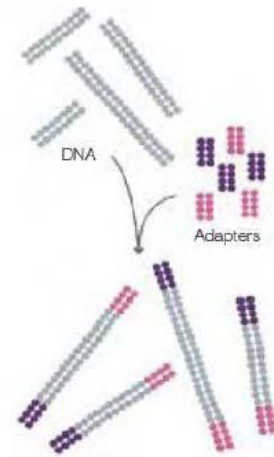
1.5.2.2. Sequencing by Synthesis (Illumina/Solexa)

Η αλληλούχιση μέσω σύνθεσης (by synthesis) βασίζεται στην ποσοτική ενίσχυση των θραυσμάτων DNA (DNA fragments) πριν τη διαδικασία αλληλούχισης και τη χρήση φθορίζουσας χρώσης για τη φωτοχημική ανίχνευση και καταγραφή των βάσεων που προσδένονται κατά το στάδιο επιμήκυνσης της αλληλουχίας.

Πιο συγκεκριμένα, τα θραύσματα DNA, για τα οποία επιδιώκεται η αλληλούχιση, προσδένονται ομοιοπολικά με εξειδικευμένες αλληλουχίες που ονομάζονται «προσδέτες» (adapters), και στα δύο άκρα τους ταυτόχρονα. Οι «προσδέτες» αποτελούνται από συγκεκριμένα και διαφορετικά barcodes για κάθε δείγμα. Τα barcodes είναι αλληλουχίες που προστίθενται στα DNA θραύσματα για να μπορέσει στη συνέχεια να γίνει η αντιστοίχιση του θραύσματος με το δείγμα στο οποίο ανήκει [39].

³ <https://genetics.thetech.org/3rd-generation-sequencing>

Εικόνα 2: Πρόσδεση adapters στα άκρα των θραυσμάτων DNA⁴

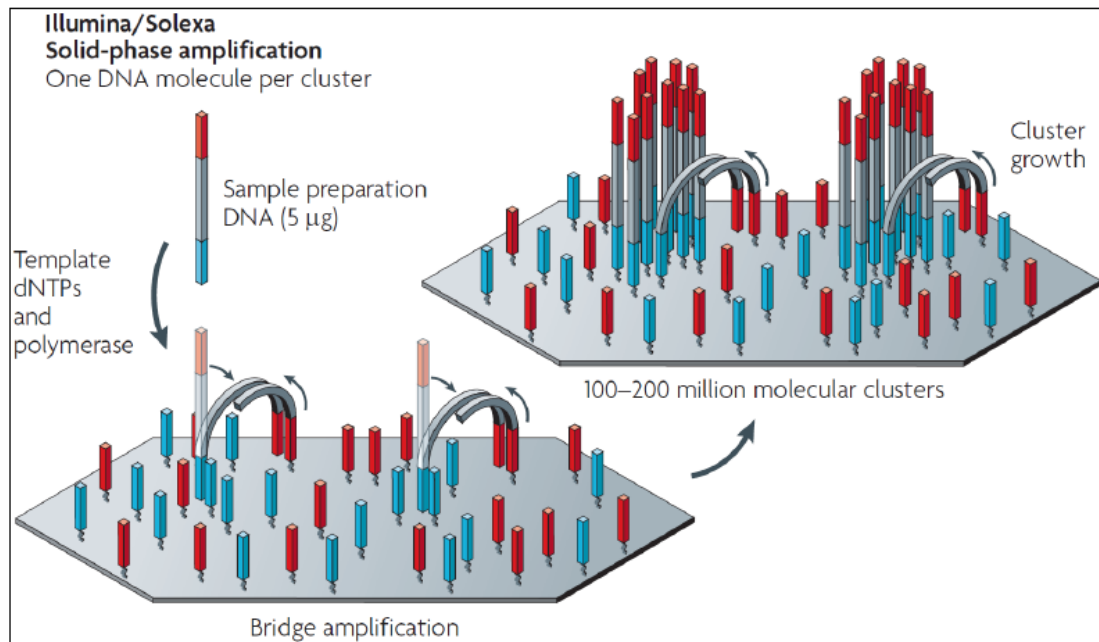


Εφόσον έχει επιτευχθεί η πρόσδεση, προκαλείται σπάσιμο των πλευρικών δεσμών της διπλής αλυσίδας, με αποτέλεσμα να εναπομένουν τα μονόκλινα άκρα του θραύσματος DNA με γνωστό μόνο ένα μικρό τμήμα στα άκρα του (sequencing binding site).

Στη συνέχεια, τα θραύσματα DNA ακινητοποιούνται σε ειδικό γυάλινο υπόστρωμα που φέρει μικρά κανάλια και ολιγονουκλεοτιδικές αλληλουχίες που είναι συμπληρωματικές με τμήματα της αλληλουχίας του προσδέτη των δύο άκρων του DNA θραύσματος. Σε αυτό το σημείο λαμβάνει χώρα τυχαίος υβριδισμός μεταξύ των ολιγονουκλεοτιδίων που βρίσκονται στην γυάλινη επιφάνεια με τους adapters του άκρου των μονόκλωνων θραυσμάτων DNA δημιουργώντας γέφυρες - bridge amplification [40] με την βοήθεια ισοθερμικής πολυμεράσης.

⁴ <https://emea.illumina.com/>

Εικόνα 3: Μέθοδος αλληλούχισης μέσω σύνθεσης Solexa/Illumina με ποσοτική ενίσχυση θραυσμάτων στερεάς φάσης (solid-phase amplification)⁵



Οι DNA πολυμεράσες αποτελούν μια ομάδα ενζύμων, η οποία έχει τη δυνατότητα να συναρμολογεί μόρια DNA, δηλαδή να καταλύει τον πολυμερισμό των δεσοξυριβονουκλεοτιδίων σε μια αλυσίδα DNA. Συμμετέχει στην αντιγραφή του DNA χρησιμοποιώντας μια αρχική αλυσίδα ως εκμαγείο για τη σύνθεση της συμπληρωματικής της.

Συνεχίζοντας, κατά την αντίδραση αυτή, ο πολυμερισμός συμβαίνει με ταυτόχρονη σύνδεση και των δύο άκρων των DNA θραυσμάτων, με τελικό αποτέλεσμα το σχηματισμό αλληλουχιών - κλώνων ανά θραύσμα. Έπειτα, η αλληλουχία αυτή μετουσιώνεται και απομακρύνεται το αρχικό θραύσμα DNA.

Η συμπληρωματική αλληλουχία που έχει εναπομείνει, προσδέεται μέσω σχηματισμού γέφυρας - bridge amplification με τον προσδέτη του άλλου άκρου της, που βρίσκεται στη γυάλινη επιφάνεια. Πραγματοποιείται εμποτισμός με DNA πολυμεράση, η οποία δημιουργεί κλώνο της αλληλουχίας που βρίσκεται προσδεμένη. Στη συνέχεια οι δίκλωνες αλυσίδες

5

https://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf

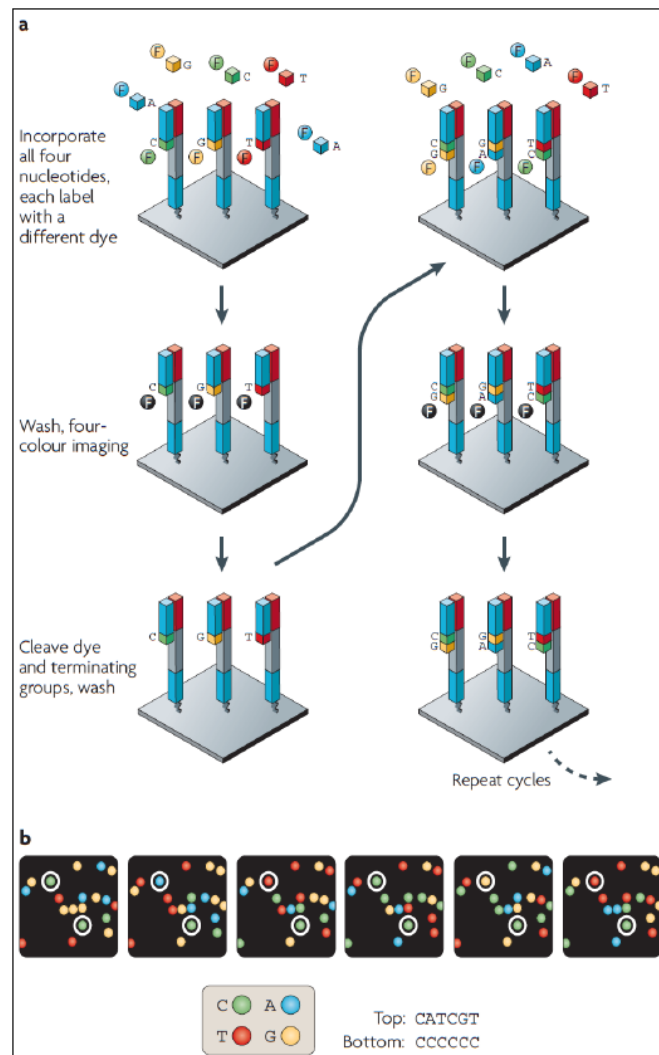
μετουσιώνονται με αποτέλεσμα τη δημιουργία δύο μονόκλωνων συμπληρωματικών αντιγράφων, προσδεμένων στη γυάλινη επιφάνεια.

Η διαδικασία αυτή επαναλαμβάνεται ταυτόχρονα πολλές φορές σε όλα τα DNA θραυσμάτων που επιδιώκεται αλληλούχιση. Μετά από κάθε υβριδοποίηση, η συμπληρωματικοί κλώνοι απομακρύνονται και στη γυάλινη επιφάνεια παραμένουν προσδεμένοι οι κλώνοι των αρχικών θραυσμάτων. Η κάθε βιβλιοθήκη θραυσμάτων αποτελείται πλέον από εκατοντάδες εκατομμύρια μοναδικά συμπλέγματα (clusters).

Σε αυτό το σημείο αρχίζει η αλληλούχιση με την βοήθεια σταδιακής εισαγωγής νουκλεοτιδίων, dNTPs (ddATP, ddGtp, ddCTP, ddTTP), που το καθένα έχει φθορίζουσα χρώση διαφορετικού χρώματος. Η εμφάνιση φωτεινού σήματος υποδεικνύει την πρόσδεση συμπληρωματικού νουκλεοτιδίου σε κάθε βήμα και η χρωματική διαφορά – χρωματικό ερέθισμα των τεσσάρων βάσεων καθιστά εφικτή την ορθή αποκρυπτογράφηση του νουκλεοτιδίου που προσδέεται κάθε φορά. Ο αριθμός των κύκλων εναπόθεσης και σύνδεσης νουκλεοτιδίων υποδεικνύει το μήκος των θραυσμάτων που γίνεται η αλληλούχιση. Τα προσλαμβάνουσα φωτεινά σήματα αποθηκεύονται για να γίνει στη συνέχεια η ανάλυση των αποτελεσμάτων.

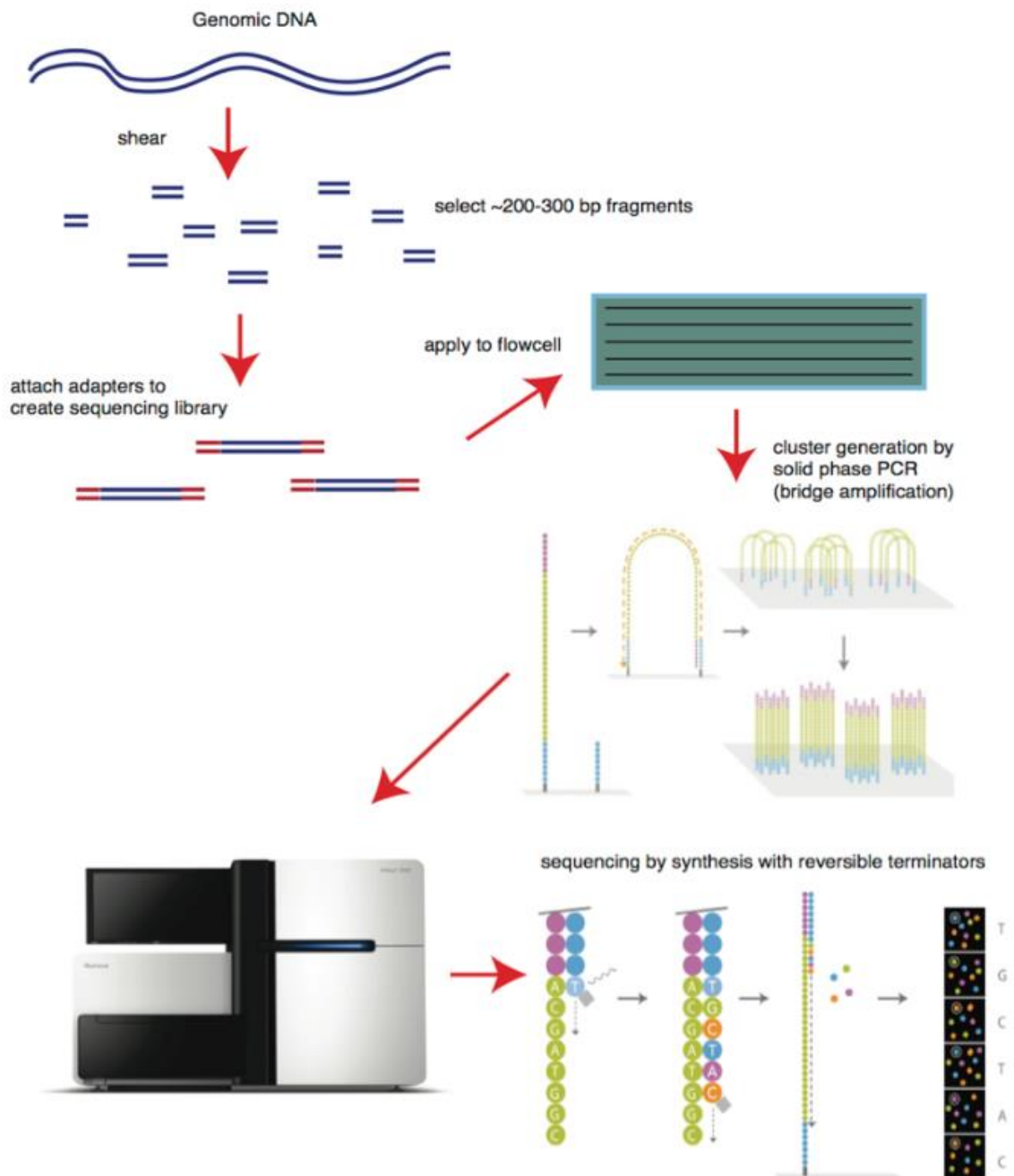
Μετά την ολοκλήρωση της πρώτης αναγνώρισης η ίδια διαδικασία διεξάγεται για την αναγνώριση της συμπληρωματικής αλληλουχίας των θραυσμάτων. Ο κύριος σκοπός της αυξημένης επανάληψης της διαδικασίας είναι η μείωση του ποσοστού σφάλματος. Η διαδικασία αλληλούχισης σε ξεχωριστά βήματα και στα δύο άκρα των μορίων ακολουθίας DNA (reads) καλείται paired-end sequence και παρέχει πληροφορίες που διευκολύνουν το alignment, και ιδιαίτερα για μικρά reads [41].

Εικόνα 4: Προσδιορισμός βάσεων στην αντίδραση επιμήκυνσης μέσω εμπλουτισμού dNTPs με φθορίζουσα χρώση⁶.



⁶ <https://emea.illumina.com/?langsel=/gr/>

Εικόνα 5: Ροή όλης της διαδικασίας αλληλούχισης μέσω σύνθεσης⁷



⁷ <https://emea.illumina.com/?langsel=/gr/>

1.6 Βιοπληροφορική ανάλυση Μεταγονιδιωματικών Δεδομένων

1.6.1 Μεταγονιδιωματική (Metagenomics)

Μεταγονιδιωματική ονομάζεται η ανάλυση του συνολικού γενετικού υλικού που έχει απομονωθεί από περιβαλλοντικά δείγματα. Ο όρος μεταγονιδιωματική προέρχεται από τον συνδυασμό δύο διαφορετικών προσεγγίσεων ανάλυσης, από τον όρο «μεταανάλυση» (meta-analysis) που αναφέρεται στη συγκέντρωση δεδομένων ίδιου ερευνητικού ενδιαφέροντος που έχουν αναλυθεί ξεχωριστά και την ανάλυση τους συγκεντρωτικά, και τον όρο «γενωμική» (genomics), η οποία αναφέρεται στην ανάλυση του γενετικού υλικού ενός οργανισμού [42].

Δημιουργείται η δυνατότητα απομόνωσης πολλών διαφορετικών γονιδιωμάτων εντός ενός δείγματος. Το 1985, μέσω της ερευνητικής ομάδας του Pace, έγινε η πρώτη αναφορά ως ιδέα, για την απευθείας γενετική ανάλυση περιβαλλοντικών δειγμάτων. Η πρώτη επιτυχημένη προσπάθεια μέσω κλωνοποίησης σε πλασμιδιακό φορέα, που δημιουργήθηκε από DNA φάγου, έγινε το 1991 από την ερευνητική ομάδα του Schmidt [43].

Μέσω της τεχνολογικής εξέλιξης και της ανάπτυξης των μεθόδων αλληλούχισης επόμενης γενιάς, η μεταγονιδιωματική αποτελεί ισχυρό εργαλείο διερεύνησης της δομής, της ποικιλότητας και της μεταβολικής δραστηριότητας σύνθετων μικροβιακών κοινοτήτων [44].

Μέσω αυτής της τεχνολογίας έγινε εφικτή η ποσοτική και ποιοτική ανίχνευση γονιδιωμάτων από βακτήρια, μύκητες, ζύμες, αρχαία και πρωτόζωα, η οποία δεν ήταν δυνατό να πραγματοποιηθεί κατά το παρελθόν. Κάποιοι από τους βασικούς τομείς που χρησιμοποιείται αυτή η τεχνολογία τα τελευταία χρόνια είναι στην ανάλυση του ανθρώπινου εντερικού μικροβιώματος (Human Microbiome Project), στη διαδικασία και το τελικό προϊόν της παραγωγής τροφίμων, σε περιβαλλοντικές μολύνσεις και σε μορφές παραγωγής ενέργεια.

Για την ανάλυση του μικροβιώματος ενός δείγματος υπάρχουν διαθέσιμες δύο βασικές μεθοδολογίες. Η αλληλούχιση του δείγματος των γονιδιωμάτων όλων των μικροοργανισμών (shotgun metagenomics – τυχαίας προσπέλασης)

και η αλληλούχιση του περιβαλλοντικού δείγματος μέσω στοχευμένων γενετικών τόπων που διαφοροποιούνται μεταξύ των μικροοργανισμών (DNA Barcoding) όπως η τεχνική 16S rDNA και ITS, που είναι συγκεκριμένα σημεία του γενετικού υλικού [45]. Για τον έλεγχο των βακτηριακών πληθυσμών χρησιμοποιείται η ταυτοποίηση της περιοχής γονιδίου 16S του ριβοσωμικού RNA [46], ενώ για την ταυτοποίηση των πληθυσμών μυκητών, χρησιμοποιείται η περιοχή του γονιδίου ITS (Internal Transcribed Spacer) [47].

Στην παρούσα μελέτη, χρησιμοποιήθηκαν και οι δύο τεχνικές μεταγονιδιωματικής ανάλυσης με τις επιμέρους κατηγορίες στοχευμένης αλληλούχισης 16S και ITS.

1.6.2 Τεχνική Shotgun – Τυχαίας Προσπέλασης

Η τεχνική μεταγονιδιωματικής ανάλυσης shotgun – τυχαίας προσπέλασης, είναι μια εναλλακτική προσέγγιση στην ανάλυση μικροβιακών κοινοτήτων που εμπεριέχονται σε ένα περιβαλλοντικό δείγμα. Η απομόνωση DNA γίνεται για το δείγμα χωρίς να υπάρχει συγκεκριμένος γνωστός γονιδιωματικός τόπος για ενίσχυση. Το συνολικό γενετικό υλικό του δείγματος, εφόσον έχει απομονωθεί, διαχωρίζεται σε μικρά θραύσματα ανεξαρτήτως αλληλουχίας. Αυτά τα μικρά τμήματα DNA (reads), αποτελούν το συνολικό γενετικό υλικό του δείγματος, χωρίς να υπάρχει προγενέστερη γνώση από ποιόν μικροοργανισμό έχουν απομονωθεί, της θέσης που κατελάμβαναν, αν υποδεικνύουν μικροοργανισμό ή είναι ‘θόρυβος’ που προκλήθηκε κατά τη διάρκεια της απομόνωσης κ.α..

Για να εξακριβωθεί η προέλευση των χιλιάδων ή εκατομμυρίων αλληλουχιών μικρού μήκους – reads (400 - 500 βάσεων) και να εξεταστούν τα δεδομένα της απομόνωσης πρέπει να ακολουθήσει χρήση εξειδικευμένων αλγορίθμων που θα αναλυθούν στο επόμενο κεφάλαιο της μελέτης. Τα μικρά αυτά θραύσματα DNA είναι τυχαία και επικαλύπτονται μερικώς ή πλήρως μεταξύ τους. Αυτή η επικάλυψη των reads μπορεί να βοηθήσει στη συναρμολόγηση μέσω στοιχίσεων μεγαλύτερων αλληλουχιών, τα λεγόμενα contigs. Στα σημεία που

δεν υπάρχει επαρκής πληροφορία και δεν μπορεί να συνεχίσει η διαδικασία της στοίχισης των reads, η συναρμολόγηση διακόπτεται και η συγκεκριμένη αλληλουχία του contig ολοκληρώνεται. Αυτό γίνεται για όλα τα reads των αρχικών δεδομένων, μέχρι όλα τα δεδομένα προς ανάλυση να πάρουν τη μορφή των contigs, και να προχωρήσει η ανάλυση στο επόμενο βήμα.

Η τεχνική μεταγονιδιωμιακής ανάλυσης shotgun, παρέχει τη δυνατότητα της έρευνας δύο διαφορετικών πτυχών μιας μικροβιακής κοινότητας, και την κάλυψη δύο κύριων ερωτημάτων. Από ποιους μικροοργανισμούς αποτελείται το δείγμα και ποια είναι η λειτουργία των μικροοργανισμών σε αυτό το δείγμα. Το δεύτερο ερώτημα δύναται να λάβει απαντήσεις καθώς το δείγμα αποτελείται από κωδικοποιημένες αλληλουχίες που παρέχουν πληροφορίες για την βιολογικές λειτουργίες των γονιδιωμάτων.

Οι νέες τεχνικές αλληλούχισης προσφέρουν πολλά οφέλη αλλά η ανάλυση που ακολουθηθεί για να απαντηθούν τα ερωτήματα που θέτονται, είναι μια περίπλοκη και χρονοβόρα διαδικασία. Σε κάποιες περιπτώσεις τα γονιδιώματα που βρίσκονται στο δείγμα, λόγω της ποικιλότητας που υπάρχει, δεν αντιπροσωπεύονται πλήρως μέσω των reads. Υπάρχει επίσης το ενδεχόμενο, δύο reads που είναι τμήματα από το ίδιο γονίδιο να μην αλληλοκαλύπτονται έτσι ώστε να μην επεκτείνεται η αλληλουχία του γονιδίου, και να γίνεται αδύνατη η αναγνώριση του μέσω στοίχισης αλληλουχιών [48] [49].

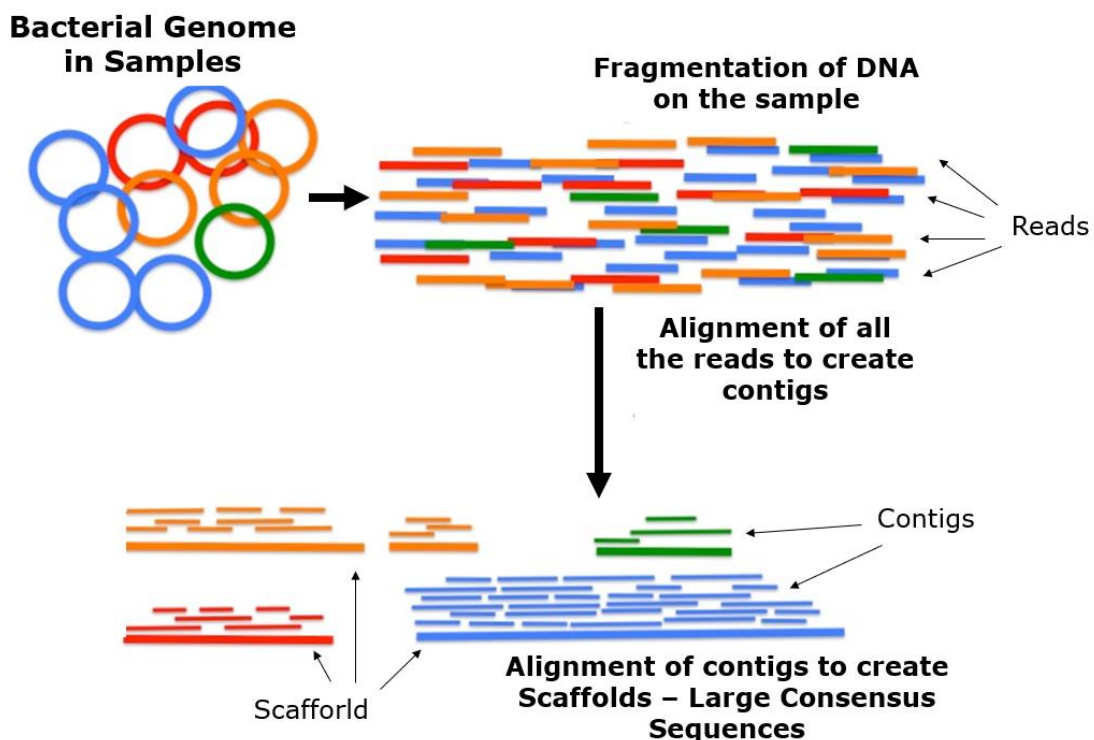
Επίσης κατά την ανάλυση των αρχικών δεδομένων, στις περιπτώσεις που υπάρχει αλληλοεπικάλυψη των reads, γεγονός που είναι το ζητούμενο για να επεκταθεί η αναγνώριση του είδους ή του λειτουργικού ρόλου της αλληλουχίας, δεν είναι απαραίτητο ότι τα reads προέρχονται από το γονιδίωμα του ίδιου μικροοργανισμού ή από διαφορετικό μικροοργανισμό με πολλές ομοιότητες. Το γεγονός αυτό μπορεί να προκαλέσει λάθη κατά την αναγνώριση [50, 51].

Συμπληρωματικά, υπάρχει το ενδεχόμενο στα δεδομένα του δείγματος που θέλουμε να αναλύσουμε, να βρίσκεται το γονιδίωμα ενός ξενιστή ή μιας μόλυνσης. Μέσω της χρήσης εξειδικευμένων εργαλείων βιοπληροφορικής, μας δίνεται η επιλογή της ταυτοποίησης και απομάκρυνσης των αλληλουχιών που

προκλήθηκαν από μολύνσεις. Στις περιπτώσεις που γνωρίζουμε τον ξενιστή η διαδικασία μπορεί αν γίνει μέσω φιλτραρίσματος των αρχικών δεδομένων, χωρίς όμως να είναι βέβαιο ότι τα αρχικά δεδομένα δεν θα παραποιηθούν. Δηλαδή, δημιουργείται το ενδεχόμενο να χαθούν πληροφορίες.

Τέλος, ένα ακόμα αρνητικό κατά την ανάλυση δεδομένων τεχνικής shotgun, είναι η ανάγκη μεγάλης υπολογιστικής δύναμης, καθώς ο όγκος των δεδομένων (Raw Data) είναι πολύ μεγάλος. Για αυτό το λόγο κατά την ανάλυση δημιουργούνται πολλά υπολογιστικά προβλήματα. Σε αυτό το πρόβλημα, υπάρχει η δυνατότητα λύσης μέσω της αξιοποίησης ανοιχτών λογισμικών και της χρήσης διαμορφωμένων Platforms που υπάρχουν διαθέσιμα.

Διάγραμμα 6: Ροή αλληλούχισης με τεχνική Shotgun⁸



⁸ <https://teachthemicrobiome.weebly.com/sequencing-the-microbiome.html>

1.6.3 Amplicon Based - 16S rDNA

Τα γονίδια που κωδικοποιούνται για τη δημιουργία ριβοσωμικών μορίων RNA είναι σταθερά ως προς τη λειτουργία τους και βρίσκονται σε όλους τους οργανισμούς. Η νουκλεοτιδική τους αλληλουχία αποτελείται από συντηρημένες περιοχές, που αξιοποιούνται ως προς την αναγνώριση και την ταξινόμηση των taxa, και από μεταβλητές περιοχές που μπορούν να χρησιμοποιηθούν για την ανίχνευση διαφορών σε συγγενικά είδη [52].

Πιο συγκεκριμένα, στους προκαρυωτικούς οργανισμούς βρίσκονται τρία μόρια rRNA, το 5S με μήκος 120 ζεύγη βάσεων, το 16S με 1542 ζεύγη βάσεων και το 23S με 2900 ζεύγη βάσεων. Το 5S, λόγω του μικρού μεγέθους, δεν αξιοποιείται στην ανάλυση ταξινόμησης των μικροοργανισμών, σε αντίθεση με το 16S και 23S, στα οποία η αλληλουχία έχει μεγαλύτερο μήκος και περιέχει φυλογενετικές πληροφορίες [52]. Επίσης, το γονίδιο 16S υπερτερεί έναντι του 23S σε θέμα χρόνου και κόστους, λόγω μικρότερου μήκους.

Το γονίδιο που κωδικοποιεί την υπομονάδα 16S του ριβοσωμικού RNA (16S rDNA) των βακτηρίων, χρησιμοποιείται ως εργαλείο για την ταξινόμηση των προκαρυωτικών οργανισμών [53]. Η συγκεκριμένη περιοχή του γενετικού υλικού, βρίσκεται σε όλους τους προκαρυωτικούς οργανισμούς και απαντάται ακόμη και στους ευκαρυωτικούς, ως μιτοχονδριακό ριβοσωμικό RNA, με μικρότερη ομολογία.

Ουσιαστικά, αναφερόμαστε σε έναν αρκετά καλά συντηρημένο γενετικό τόπο που περιέχει αλλαγές μικρής εμβέλειας, οι οποίες υποδεικνύουν την εξελικτική πορεία των βακτηρίων. Ο λόγος που χρησιμοποιείται ως δείκτης ταξινόμησης των βακτηρίων είναι διότι περιλαμβάνει εννέα διαφορετικές μεταβλητές περιοχές εντός της αλληλουχίας του, που είναι διακριτές μεταξύ των διαφορετικών βακτηριακών γονιδιωμάτων.

Πριν την τεχνολογική εξέλιξη της αλληλούχισης επόμενης γενιάς, η ταυτοποίηση των βακτηρίων ήταν εφικτή μέσω κλωνοποίησης του συγκεκριμένου γενετικού τόπου σε πλασμίδιο και ο προσδιορισμός της αλληλουχίας μέσω αλληλούχισης Sanger.

Η τεχνολογική εξέλιξη και η ανάπτυξη της τεχνικής αλληλούχισης επόμενης γενιάς, βοήθησε σε μεγάλο βαθμό στη μείωση του κόστους χρήσης αυτής της τεχνικής, καθώς δεν υπάρχει η ανάγκη κλωνοποίησης. Η μέθοδος αυτή συνήθως καλείται ως «προσδιορισμός αλληλουχίας αμπλικονίων 16S rDNA» που αναφέρεται στη στοχευμένη μεταγονιδιωματική ανίχνευση αλληλουχίας. Η διαδικασία γίνεται με τη βοήθεια ειδικών λογισμικών στοίχισης των rDNA αλληλουχιών με αλληλουχίες αναφοράς από βάσεις δεδομένων που περιλαμβάνουν αλληλουχίες εκατοντάδων χιλιάδων μικροοργανισμών.

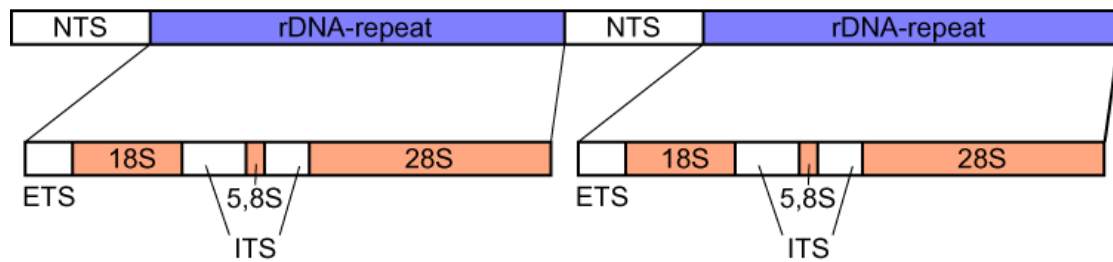
1.6.4 Amplicon Based - ITS (Internal Transcribed Spacer)

Η περιοχή της αλληλουχίας που ονομάζεται ITS – Internal Transcribed Spacer, βρίσκεται μεταξύ της περιοχής που κωδικοποιεί το γονίδιο της μικρής ριβοσωμικής υπομονάδας και το γονίδιο της μεγάλης ριβοσωμικής υπομονάδας. Στα βακτήρια και στα αρχαία, υπάρχει μόνο μία περιοχή ITS, μεταξύ των γονιδίων που κωδικοποιούν το 16S rDNA και το 23S rRNA, σε αντίθεση με τους ευκαρυωτικούς οργανισμούς που υπάρχουν δύο περιοχές ITS. Η περιοχή ITS1 βρίσκεται μεταξύ των γονιδίων 18S και του 5.8S και η περιοχή ITS2 βρίσκεται μεταξύ των γονιδίων 5.8S και 28S. Το σημείο του ITS1 αντιστοιχεί στην περιοχή ITS των βακτηρίων και των αρχαίων, ενώ το ITS2 προήλθε ως ένθετη διακοπή του προγονικού γονιδίου 23S rRNA.

Στους ευκαρυωτικούς οργανισμούς, γονίδια που κωδικοποιούνται για την δημιουργία ριβοσωμάτων, εμφανίζονται σε διαδοχικές επαναλήψεις κατά μήκος πολλών αντιγράφων, με διακοπτόμενα τμήματα μη μεταγραφικού DNA που ονομάζονται Intergenic Spacer – IGS ή αλλιώς non-transcribed spacer (NTS).

Στην παρακάτω εικόνα αποτυπώνεται η περιοχή που βρίσκεται η αλληλουχία ITS.

Εικόνα 6: Περιοχή ITS⁹



Η συγκεκριμένη περιοχή, χρησιμοποιείται ευρέως για τη διερεύνηση της συστηματικής ταξινόμησης των ειδών. Οι βασικοί λόγοι που έχει επιλεγεί αυτή η αλληλουχία είναι το σχετικό μικρό μήκος της, το γεγονός ότι είναι αρκετά συντηρημένη και η ανίχνευση είναι εύκολη ακόμη και σε μικρές ποσότητες DNA λόγω των πολλών αντιγράφων της. Επίσης έχει διαπιστωθεί υψηλός βαθμός παραλλαγών - variants στην αλληλουχία, ακόμη και μεταξύ ειδών που είναι συγγενικά σε μεγάλο βαθμό, διότι η εξελικτική πίεση που έχει δεχθεί είναι χαμηλή. Η περιοχή αυτή, χρησιμοποιείται ευρέως για την ταξινόμηση των μυκήτων.

⁹ https://www.wikiwand.com/en/Internal_transcribed_spacer

1.7 Σκοπός Διπλωματικής

Ο σκοπός της παρούσας διπλωματικής είναι η ανάλυση μεταγονιδιωματικών δεδομένων, για την εύρεση της πληθυσμιακής σύστασης στη μικροχλωρίδα τριών δειγμάτων Ελληνικού ΠΟΠ τυριού Σφέλα και τριών δειγμάτων παραδοσιακού Κυπριακού τυριού Χαλίτζι.

Τα δύο δείγματα τυριού ΠΟΠ Σφέλας παράχθηκαν σε μικρής κλίμακας βιομηχανικές παραγωγές στην περιοχή της Μεσσηνίας. Το δείγμα 1-Σφέλας και το δείγμα 2-Σφέλας παράχθηκε από διαφορετικούς παραγωγούς της περιοχής Σπερχογείας. Τα υπόλοιπα δείγματα τυριών, δείγμα 3-Σφέλας και τα τρία δείγματα Χαλίτζι, προέρχονται από οικιακές παραγωγές. Το γεγονός αυτό μας δίνει τη δυνατότητα συγκριτικής ανάλυσης της βιοποικιλότητας του μικροβιώματος των τυριών, βιομηχανικής και οικιακής παραγωγής.

Στη μελέτη χρησιμοποιήθηκαν δύο τεχνικές αλληλούχισης επόμενης γενιάς. Στοχευμένη αλληλούχιση των περιοχών 16S rDNA και ITS (amplicon based techniques), που έχουν ως στόχο την ταξινομική ανάλυση των μεταγονιδιωματικών δεδομένων, και η τεχνική αλληλούχισης τυχαίας προσπέλασης - Shotgun, που περιλαμβάνει αλληλούχιση του ολικού DNA των δειγμάτων.

Η δεύτερη τεχνική αλληλούχισης μας προσφέρει τη δυνατότητα, συμπληρωματικά με τη συστηματική ταξινόμηση των δειγμάτων, τη λειτουργική ανάλυση. Η ανάλυση αυτή παρέχει πληροφορίες για τις λειτουργικές ιδιότητες του συνολικού μικροβιώματος που βρίσκεται στο δείγμα.

Η μελέτη των μεταγονιδιωμάτων των δύο τυριών πραγματοποιήθηκε μέσω της χρήσης εργαλείων βιοπληροφορικής ανάλυσης. Χρησιμοποιήθηκαν τρεις διαφορετικές πλατφόρμες ανάλυσης μεταγονιδιωματικών δεδομένων που θα περιγραφούν εκτενέστερα στις υποενότητες που ακολουθούν. Η ροή εργασιών για τις τρεις πλατφόρμες ακολουθεί τη βασική δομή ανάλυσης με επιμέρους διαφοροποιήσεις.

Τα τέσσερα βασικά ερωτήματα που μπορούν να απαντηθούν μέσω της βιοπληροφορικής ανάλυσης των μεταγονιδιωματικών δεδομένων είναι η

διερεύνηση των μικροοργανισμών που βρίσκονται στα δείγματα, η σύνθεση του μικροβιώματος, οι λειτουργικές ιδιότητες των επιμέρους μικροοργανισμών και τέλος η συγκριτική ανάλυση των δειγμάτων.

2. ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ

2.1 Απομόνωση DNA

Για την απομόνωση του γενετικού υλικού των δειγμάτων χρησιμοποιήθηκε το τροποποιημένο πρωτόκολλο κατά Pitcher [54] ακολουθώντας τα εξής βήματα:

Προετοιμασία Buffer :

- PBS pH 7.4 1L (8g NaCl, 0.2 KCL, 1.44 Na₂HPO₂, 0.24 KH₂PO₂),
- TE pH 8.0. (10nmol Tris-HCl, 1nmol/L EDTA),
- GES reagent (5mol/L guanidium thiocyanate, 100mmol/L EDTA pH 8.0, 0.5% v/v sarkosyl).

Πρωτόκολλο:

1. 50gr δείγματος τυριού τοποθετήθηκε σε σκεύος θρυμματισμού και προστέθηκε 50ml αποσταγμένου H₂O.
2. Φυγοκέντρωση 2ml δείγματος για το διαχωρισμό του λίπους με το ίζημα
3. Προστέθηκε 1ml PBS buffer, pH 7.4, και ακολούθησε επώαση στους 65°C για 10min με στόχο την μείωση της δράσης των DNAases.
4. Πραγματοποιήθηκε φυγοκέντρωση 6.000rpm/3min, και επανάληψη του βήματος 3 και 4, 2 φορές.
5. Στο δείγμα προστέθηκαν 40μl λυσοζύμη (600ul 25mg/ml), 20μl RNase A (100ul 20mg/ml), 6μl mutanolysin (20ul 5U/ul) και διαλυμένη 5μl lyticase (20ul 10.000U/ml) σε μείγμα 2.5ml γλυκερόλης (500ul) και 2.5ml TE buffer (500ul pH 7.5).
6. Επώαστηκε για 3 ώρες στους 37°C με περιοδική ανάδευση.
7. Μετά την επώαση προστέθηκε 0.02ml πρωτεϊνάση K (25mg/ml) και επώαστηκε ξανά στους 55°C για 1 ώρα σε υδατόλουτρο με περιοδική ανάδευση, προκειμένου να επαναδιαλυθεί πλήρως το ίζημα.
8. Στη συνέχεια προστέθηκε 0.5ml διαλύματος GES και το δείγμα αναδεύτηκε με στόχο την πλήρη διάλυση του ιζήματος.
9. Ακολούθησε φυγοκέντρωση 12.500rpm/10min και μεταφέρθηκε 1ml του υπερκείμενου διαλύματος σε νέο σωλήνα.
10. Τοποθετήθηκε στον πάγο για 5-10min.

11. Προστέθηκε 0.25ml ammonium acetate (7.5mol/L), ακολούθησε ανάδευση και τοποθετήθηκε σε πάγο για 10min.
12. Προστέθηκαν 0.5ml χλωροφόρμιο και ακολούθησε ανάδευση.
13. Ακολούθησε φυγοκέντρηση 13.000rpm/10min και μετά μεταφέρθηκε 1ml του υπερκείμενου διαλύματος σε νέο σωλήνα.
14. Προστέθηκε φαινόλη : χλωροφόρμιο με αναλογία δείγματος 1:0.5:0.5 και ανακινήθηκε.
15. Φυγοκεντρήθηκε 12.500rpm/10min και μεταφέρθηκαν 0.9ml του υπερκείμενου διαλύματος σε νέο tube.
16. Επαναλήφθηκαν τα βήματα 11 και 12.
17. Μεταφέρθηκαν 0.9ml του υπερκείμενου και προστέθηκε ίδια ποσότητα χλωροφόρμιο.
18. Φυγοκεντρήθηκε 12.500rpm/10min και μεταφέρθηκαν 0.7ml του υπερκείμενου σε νέο tube.
19. Επαναλήφθηκαν τα βήματα 15 και 16.
20. Μεταφέρθηκαν 0.55-0.6ml σε νέο tube και προστέθηκαν 0.055ml sodium acetate pH 5.2 και 0.3ml κρύα ισοπροπανόλη.
21. Διατηρήθηκε όλη τη νύχτα στους -20°C.
22. Την επόμενη μέρα έγινε φυγοκέντρηση 12.500rpm/20min/4°C. Αφαιρέθηκε το υπερκείμενο και κρατήθηκε το pellet που βρισκόταν στο tube.
23. Ξεπλύθηκε το pellet 2 φορές σε 0.7ml κρύας αιθανόλης 70% και φυγοκεντρήθηκε σε 8.000rpm/5min/4°C.
24. Αφαιρέθηκε η αιθανόλη από το pellet και τοποθετήθηκε με ανοιχτό στόμιο σε 37°C για 5-10 min.
25. Διαλύθηκε το pellet με TE buffer pH 8.0
26. Τέλος επώασθηκε στους 37°C για 1 ώρα με περιοδικές ανακινήσεις των tubes.

2.2 Μεταγονιδιωματική Ανάλυση - Analysis Workflow

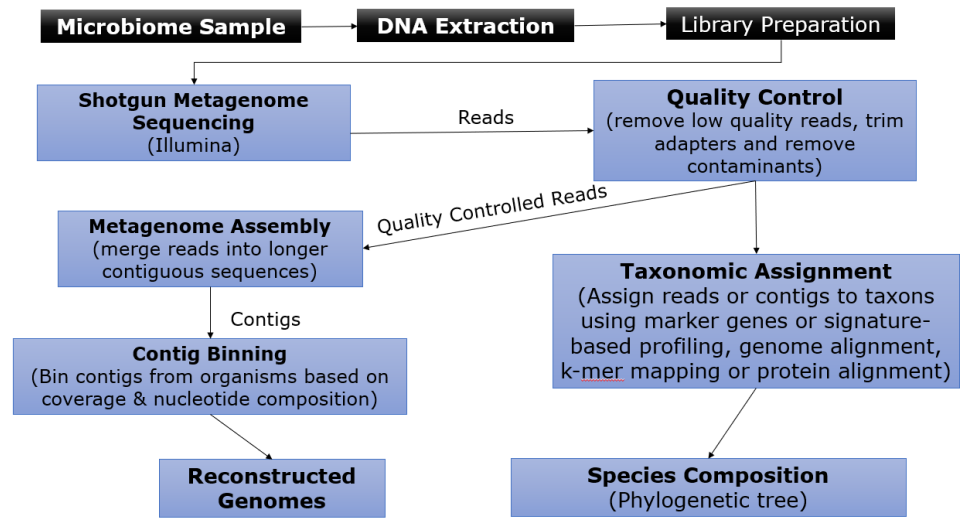
Η ανάλυση των δεδομένων μεταγονιδιωματικής είναι η διαδικασία κατά την οποία τα δεδομένα επιθεωρούνται, φιλτράρονται, μετασχηματίζονται και μοντελοποιούνται με στόχο την απάντηση των ερωτημάτων και την εξόρυξη χρήσιμων πληροφοριών για τη δημιουργία ολοκληρωμένων συμπερασμάτων. Ακολουθούνται αρκετά βήματα μέχρι το τελικό αποτέλεσμα, και διαφορετικές ροές εργασιών [55]. Η βασική ανάλυση ακολουθεί τον σχεδιασμό που αναγράφεται ακολούθως.

Εφόσον έχει προηγηθεί η απομόνωση του γενετικού υλικού, η αλληλούχιση των θραυσμάτων DNA και η δημιουργία εκατομμυρίων μικρών τμημάτων (reads), ακολουθεί η επεξεργασία των δεδομένων με μεθόδους βιοπληροφορικής. Πιο συγκεκριμένα, τα βασικά βήματα της μεταγονιδιωματικής ανάλυσης είναι τα ακόλουθα:

1. Επεξεργασία βιολογικού δείγματος
2. Αλληλούχιση
3. Ποιοτικός έλεγχος και φιλτράρισμα των δεδομένων αλληλούχισης
4. Συναρμολόγηση (Assembly)
5. Ομαδοποίηση (Binning)
6. Σχολιασμός (Annotation)

Διαφοροποιήσεις σχετικά με την ανάλυση των αρχικών δεδομένων Shotgun και με την ανάλυση των αρχικών δεδομένων 16S rDNA και ITS υπάρχουν και θα αναλυθούν στη συνέχεια.

Διάγραμμα 7: Βασική Ροή Εργασίας Μεταγονιδιωματικής Ανάλυσης

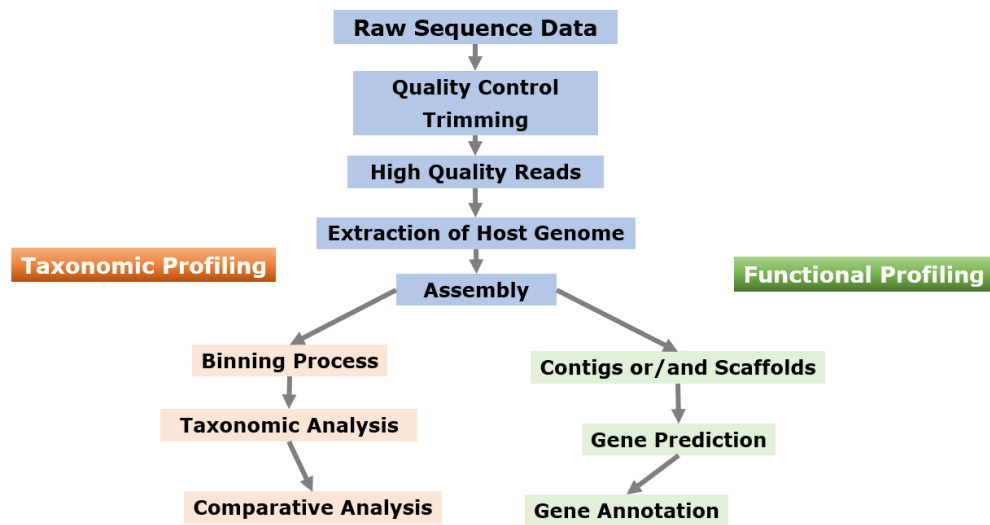


Το διάγραμμα που ακολουθεί παρουσιάζει τη ροή εργασιών στις δύο διαφορετικές τεχνικές που εφαρμόστηκαν. Δηλαδή αλληλούχιση όλου του γονιδιώματος των δειγμάτων - Shotgun, και αλληλούχιση εξειδικευμένων περιοχών του γονιδιώματος των μικροοργανισμών - 16S rDNA και ITS-Internal Transcribed Space [56].

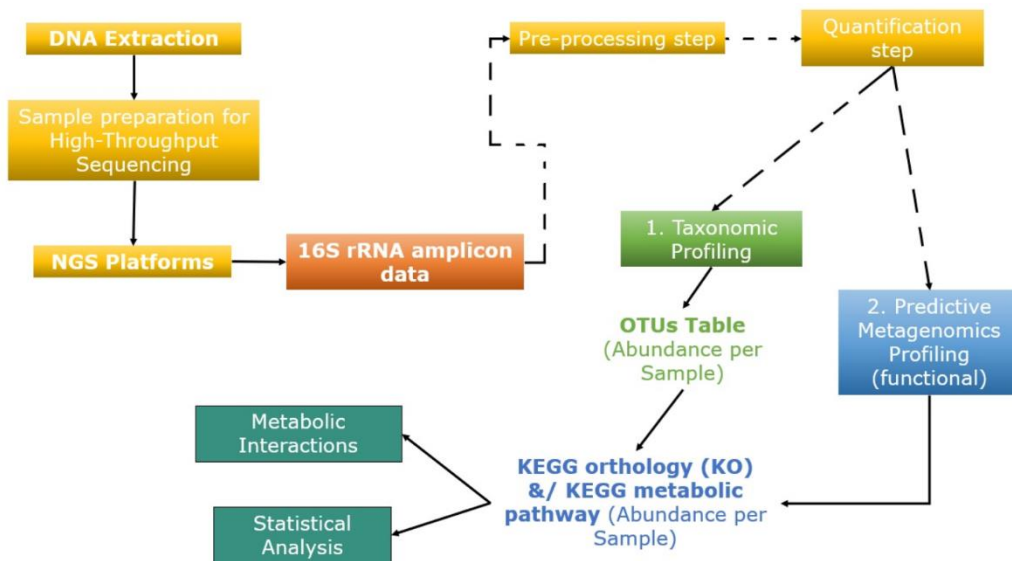
Στη πρώτη τεχνική αλληλούχισης, επιδιώκεται πέρα από την συστηματική ταξινόμηση των πληθυσμών που βρίσκονται στα δείγματα, η διεξαγωγή λειτουργικής ανάλυσης (Functional Analysis).

Στη δεύτερη τεχνική, στοχευμένη αλληλούχιση, επιδιώκεται η ταυτοποίηση των μικροοργανισμών που εμπεριέχονται στα δείγματα. Η συστηματική ταξινόμηση μπορεί να επιτευχθεί μέχρι τη βαθμίδα του γένους, καθώς η διαφοροποίηση στη βαθμίδα των ειδών ενδέχεται να περιέχει σφάλματα.

Διάγραμμα 8: Βιοπληροφορική Ροή Εργασιών, Taxonomic Profiling και Functional Analysis



Διάγραμμα 9: Ροή βιοπληροφορικής ανάλυσης Amplicon τεχνικής



Στη συνέχεια θα ακολουθήσει μια λεπτομερής αναφορά των εργασιών που πρέπει να επιτευχθούν και σε επόμενη ενότητα θα αναφερθούμε στα εργαλεία που χρησιμοποιήθηκαν ξεχωριστά για κάθε βήμα της ανάλυσης.

2.2.1 Αρχεία FastQ

Τα δείγματα DNA στάλθηκαν στην εταιρεία BGI, η οποία εδρεύει στο Hong Kong. Πραγματοποιήθηκε αλληλούχιση και στάλθηκαν τα αρχεία FastQ που περιείχαν όλες τις γενετικές πληροφορίες των 6 δειγμάτων.

Ένα από τα μεγαλύτερα προβλήματα της τεχνολογίας Επόμενης Γενιάς Αλληλούχισης, είναι η αποθήκευση και ο χειρισμός των δεδομένων. Τα αρχεία που εξάγονται από τις αναλύσεις NGS είναι σε συμπιεσμένη μορφή, και πολλές φορές διπλής εκτέλεσης (fq.gz). Αυτό γίνεται για να μειωθεί ο όγκος που καταλαμβάνουν και ο χρόνος που απαιτείται για τη μεταφορά των αρχείων [55].

Τα δεδομένα αλληλούχισης «paired-end», αποθηκεύονται σε δύο συμπιεσμένα αρχεία FastQ format. Τα αρχεία αυτά περιλαμβάνουν την ακολουθία και ένα σκορ βεβαιότητας για κάθε νουκλεοτίδιο που ανιχνεύθηκε κατά την αλληλούχιση. Ακολουθεί εικόνα, στην οποία αποτυπώνεται η μορφή των αρχείων FastQ.

Εικόνα 7: Μορφή Αρχείου FastQ

```
@ML-P2-14:9:000H003HG:1:11102:17290:1073 1:N:0:TCCTGAGC+GCGATCTA
TTTGGTAACAGCATGAATTATTCTAGCCACTAAAACCTATGAACATCTTGTGAAGGTTTCAGATAGAGCCTGAAGTACACAGAGAACAATTCTTAAAAAA
+
AAAAAEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE<AEEEEEEEE
```

Στην πρώτη γραμμή βρίσκεται το αναγνωριστικό της αλληλουχίας (όνομα αλληλουχίας - sequence identifier) το οποίο αρχίζει πάντα με το σύμβολο «@». Το όνομα που έχει δοθεί διαφέρει ανάλογα με το λογισμικό μετατροπής BCL (Binary Base Call) format σε FastQ.

Ακολουθεί στην επόμενη γραμμή η αλληλουχία που έχει διαβασθεί (A, T, C, G, N=not identified).

Στην τρίτη σειρά υπάρχει ένα διαχωριστικό συνήθως με το συμβολισμό (+), το οποίο είναι αναγνωριστικό για το σκορ ποιότητας (Quality Score Identifier).

Τέλος, στην τέταρτη σειρά βρίσκεται το Quality Score για κάθε βάση ξεχωριστά. Το Quality Score - Σκορ Ποιότητας, είναι η βαθμολογία που έχει δοθεί για την εγκυρότητα της βάσης που έχει διαβασθεί κατά την

αλληλούχιση. Είναι κωδικοποιημένη Phred +33 χρησιμοποιώντας χαρακτήρες ASCII (American Standard Code for Information Interchange) για την αντιπροσώπευση του σκορ ποιότητας μέσω αριθμού.

Η τιμή του Q (Quality Value) είναι μια ακέραια τιμή, αντίστοιχη του p (= P-Value), δηλαδή η πιθανότητα της αντίστοιχης βάσης να είναι εσφαλμένη κατά την αλληλούχιση. Χρησιμοποιούνται δύο διαφορετικές εξισώσεις για τον υπολογισμό της τιμής. Η πρώτη είναι σύμφωνα με τα δεδομένα αλληλούχισης κατά Sanger λεγόμενη αλλιώς Phred Quality Score,

$$Q_{Sanger} = -10 \log_{10} p$$

Η δεύτερη εξίσωση είναι βάσει της ροής εργασίας Solexa, η οποία χρησιμοποιεί την πιθανότητα $\frac{p}{1-p}$ αντί της πιθανότητας p:

$$Q_{solexa-prior\ to\ v.1.3} = -10 \log_{10} \frac{p}{1-p}$$

Οι δύο εξισώσεις είναι πανομοιότυπες και ασύμπτωτες για υψηλές τιμές σκορ, αλλά διαφέρουν στο επίπεδο των χαμηλών τιμών ($P > 0.05$ ή $Q < 13$).

Με Phred score μιας βάσης να είναι

$$Q_{phred} = -10 \log_{10} e ,$$

όπου e είναι η εκτιμώμενη πιθανότητα λάθους σε μια βάση.

Εν συντομία, το Q-score είναι ο αρνητικός λογάριθμος της πιθανότητας να έχει αναγνωριστεί λάθος η βάση στην οποία αντιστοιχεί, πολλαπλασιασμένος με το 10. Η αξιολόγηση της πιθανότητας λάθους ανάγνωσης, μίας ή περισσότερων βάσεων κατά τη διαδικασία της αλληλούχισης, είναι πολύ σημαντική καθώς αντιμετωπίζεται συχνά το ενδεχόμενο λάθους, εξαιτίας διάφορων τεχνικών παραγόντων κατά τη διαδικασία.

Εικόνα 8: Σύστημα κωδικοποίησης ASCII

ASCII Code Chart																
	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
1	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
2		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
6	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL

Το Phred είναι η πρώτη τιμή που χρησιμοποιήθηκε για την ακριβή και αξιόπιστη αποτύπωση της ποιότητας βαθμολόγησης των βάσεων κατά την αλληλούχιση. Η υψηλή ακρίβεια των αποτελεσμάτων συνδέεται λογαριθμικά με τις πιθανότητες λάθους. Η χρήση του δείκτη πιστότητας, είναι ο αυτόματος προσδιορισμός ακρίβειας και ποιότητας των αλληλουχιών. Βοηθάει επίσης στην εκτίμηση των διαφορών που δημιουργούνται μεταξύ των δύο αλυσίδων αλληλούχισης, δηλαδή οι διαφορές που προκύπτουν μεταξύ των δύο επικαλυπτόμενων αλληλουχιών προερχόμενες από τυχαία σφάλματα ή από διάφορα αντίγραφα μιας επαναλαμβανόμενης περιοχής στην αλληλουχία. Μεγάλη τιμή Quality score υποδεικνύει μικρή πιθανότητα λάθους, δηλαδή είναι ανάλογη με την ακρίβεια της αλληλούχισης [56].

Στη συνέχεια της ανάλυσης, τα δύο αρχεία FastQ για κάθε δείγμα, συνενώνονται έτσι ώστε να προκύψει ένα αρχείο FastQ για κάθε δείγμα. Αυτό συμβαίνει καθώς τα αρχικά δεδομένα είναι «paired-end». Η διαδικασία συνένωσης (stitching) γίνεται από ειδικούς αλγόριθμους, οι οποίοι στη σύγκριση των δύο αρχείων, επιλέγουν αυτόματα για κάθε θέση βάσεων, τη βάση που αναγνώστηκε με μεγαλύτερο βαθμό Phred-score, δηλαδή με μεγαλύτερη αξιοπιστία στην ανάγνωσή της.

Πίνακας 5: Phred quality scores

Phred quality scores are logarithmically linked to error probabilities

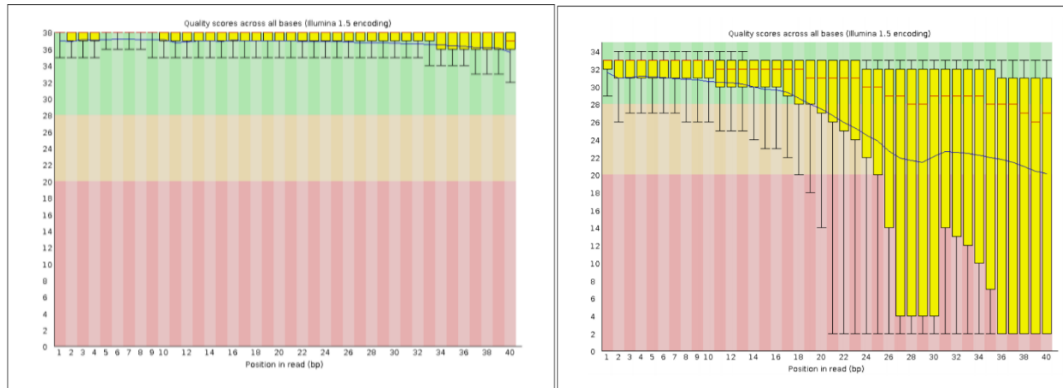
Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10000	99.99%
50	1 in 100000	99.999%

2.2.2. Quality Control - Trimming - Extraction of Host Genome

Ο ποιοτικός έλεγχος των reads και η αφαίρεση των περιοχών στις απολήξεις των αλληλουχιών είναι τα πρώτα βήματα επεξεργασίας των δεδομένων πριν την διαδικασία της ανάλυσης. Πραγματοποιείται αφαίρεση των «adapters», των τμημάτων των reads που ενδέχεται να έχουν προκληθεί από μολύνσεις και έχουν μικρές τιμές Q-score. Τα τελευταία χρόνια αναπτύχθηκαν πολλά εργαλεία που χρησιμοποιούν τη βαθμολογία του Q-score και παρέχουν πληροφορίες για τις πιθανότητες σφάλματος. Επίσης, για την διαδικασία του φιλτραρίσματος - trimming, έχουν σχεδιαστεί πολλοί αλγόριθμοι με σκοπό να αποκόπτουν βάσεις στα δύο άκρα των reads που διαθέτουν χαμηλό Q-score [56].

Εντοπίζονται σε αρκετές περιπτώσεις λάθη, ιδιαίτερα στα άκρα των reads. Το πρόβλημα δημιουργείται κατά την ενσωμάτωση φθοριζόντων dNTPs στα άκρα των θραυσμάτων. Το πιο διαδεδομένο εργαλείο ποιοτικού ελέγχου είναι το FastQC, το οποίο υλοποιείται με γλώσσα προγραμματισμού java. Επεξεργάζεται τα δεδομένα και δίνει εκτιμήσεις ως προς την ποιότητα των βάσεων της αλληλούχισης, το ποσοστό της περιεκτικότητας GC, τη ποσότητα πιθανών βάσεων που δεν ανιχνεύθηκαν και το ποσοστό των διπλασιασμένων βάσεων κατά το διάβασμα των αλληλουχιών. Στο παρακάτω γράφημα αποτυπώνονται δύο περιπτώσεις ποιοτικού ελέγχου.

Γράφημα 1: Χρήση εργαλείου FastQC. Στο πρώτο γράφημα αποτυπώνεται υψηλής ποιότητα αλληλούχισης και στο δεύτερο γράφημα χαμηλής ποιότητας αλληλούχιση¹⁰



Στις περιπτώσεις χαμηλής ποιότητας Q-score, πραγματοποιείται φιλτράρισμα με στόχο την βελτίωση της ποιότητας των δεδομένων πριν την ανάλυση. Τα περισσότερα εργαλεία trimming διαθέτουν επιλογές όπως αφαίρεση των duplicates – διπλασιασμένων αλληλουχιών που έχουν προκύψει. Επίσης, σε περιπτώσεις «paired-end» δεδομένων, το φιλτράρισμα γίνεται κατά ζεύγη και σε περίπτωση κακής ποιότητας της μίας αλληλουχίας, συνήθως αφαιρείται το ζεύγος.

Τέλος, σε κάποιες περιπτώσεις περιβαλλοντικών δειγμάτων, όπως στην περίπτωση της παρούσας μελέτης, τα δείγματα περιέχουν γενετικό υλικό από Host-Genome. Στη συγκεκριμένη περίπτωση, αναφερόμαστε σε γενετικό υλικό από τα είδη *Ovis aries* (πρόβατο) και *Capra aegagrus* (αίγα) που προέρχονται τα τυριά Σφέλα και Χαλίτζι. Για την απομάκρυνση του γενετικού υλικού των δύο «ξενιστών», πραγματοποιήθηκε αναζήτηση σε βάσεις δεδομένων το γενετικό υλικό των δύο μηρυκαστικών και στη συνέχεια διεξήχθη στοίχιση και απομάκρυνση των τμημάτων που υπήρχε alignment. Σε ανάλογες περιπτώσεις μπορεί να εφαρμοστεί παρόμοια διαδικασία.

¹⁰ https://rpubs.com/fmobegi/metagenomics_update

2.2.3 Assembly – Συναρμολόγηση των contigs

Συναρμολόγηση – Assembly ονομάζεται η διαδικασία κατά την οποία πραγματοποιείται δημιουργία μεγαλύτερου μήκους αλληλουχιών, τα λεγόμενα contigs, με βάση την αλληλοεπικάλυψη των reads που περιέχουν την αρχική πληροφορία. Σε γενικότερο πλαίσιο η διαδικασία assembly ορίζεται ως η προσπάθεια εύρεσης αλληλουχίας DNA μεγαλύτερου μήκους, από ένα συγκεκριμένο σύνολο δεδομένων, η οποία δημιουργείται βάσει στοίχισης των μικρών γραμμικών τμημάτων αλληλούχισης που βρίσκονται στα αρχικά δεδομένα της ανάλυσης [55].

Η διαδικασία απαιτεί μεγάλη υπολογιστική ισχύ, εξειδικευμένους αλγορίθμους και μεγάλο αποθηκευτικό χώρο για να μπορέσει να εκτελεστεί. Η πολυπλοκότητα της διαδικασίας έγκειται στη μικρή διαφοροποίηση των reads που έχουν προκύψει από πολλαπλά στελέχη, τα οποία διαφέρουν ελάχιστα σε μικρά τμήματα. Η συναρμολόγηση δημιουργεί μεγαλύτερου μήκους θραύσματα, που επιτρέπουν στη συνέχεια την ανίχνευση γονιδιακών τόπων [56]. Επίσης μπορεί να οδηγήσει σε πλήρη καταγραφή των γονιδιωμάτων, πάντα με την βοήθεια των κατατεθειμένων πληροφοριών στις βάσεις δεδομένων γενετικών πληροφοριών, ή σε ιδιαίτερες περιπτώσεις, μπορεί να οδηγήσει στην εύρεση νέων γονιδιωμάτων.

Μειονέκτημα αυτού του βήματος της ανάλυσης είναι η δημιουργία χιμαιρικών contigs. Οι χίμαιρες – chimera είναι τεχνητές αλληλουχίες που έχουν σχηματιστεί από δύο ή περισσότερες αλληλουχίες που συνδέονται λαθεμένα μεταξύ τους. Τείνουν να παράγουν στη συνέχεια της ανάλυσης, ψευδείς OTUs (Operational Taxonomic Units), τα οποία οδηγούν σε υποβαθμισμένες εκτιμήσεις της ποικιλομορφίας και αστοχίες στην συστηματική ταξινόμηση. Σε περίπτωση που δεν εντοπιστεί άμεσα η δημιουργία των χιμαιρικών contigs, δεν μπορεί να γίνει ακριβής αποτύπωση της βιολογικής ποικιλότητας του δείγματος.

Επίσης, η διαδικασία της συναρμολόγησης παραποιεί τις πληροφορίες σχετικά με την ποσότητα ενός είδους στο δείγμα, καθώς οι επικαλυπτόμενες αλληλουχίες από ένα είδος σε αφθονία, στοιχίζονται για τη συναρμολόγηση

και τη δημιουργία των contigs και κατά συνέπεια ενώνονται σε αυτό. Αυτό έχει ως αποτέλεσμα την ενδεχόμενη υποεκπροσώπηση των ειδών σε αφθονία.

Θα μπορούσαμε να αναφέρουμε ότι προκύπτουν δύο βασικά προβλήματα κατά το βήμα της συναρμολόγησης. Πρώτον, η λάθος διάταξη των reads που διαμορφώνεται κατά την τοποθέτησή τους σε μεγαλύτερου μήκους αλληλουχίες - contigs και δεύτερον η δημιουργία consensus αλληλουχίας από την αλληλοεπικάλυψη των reads με την παρουσία πολλών επαναλήψεων. Σε αυτές τις τοποθετήσεις επαναλήψεων των reads, δεν μπορεί να αξιολογηθεί η εγκυρότητα τους [56]. Τα περισσότερα εργαλεία που πραγματοποιούν συναρμολόγηση βασίζονται στο Quality-score και σε στατιστικά στοιχεία για την επιλογή των reads που θα συνενωθούν.

Χρησιμοποιούνται δύο βασικές στρατηγικές για τη διαδικασία της συναρμολόγησης:

1. Συναρμολόγηση με βάση γνωστή αλληλουχία αναφοράς - reference based assembly, όταν αυτή είναι γνωστή.
2. De novo συναρμολόγηση.

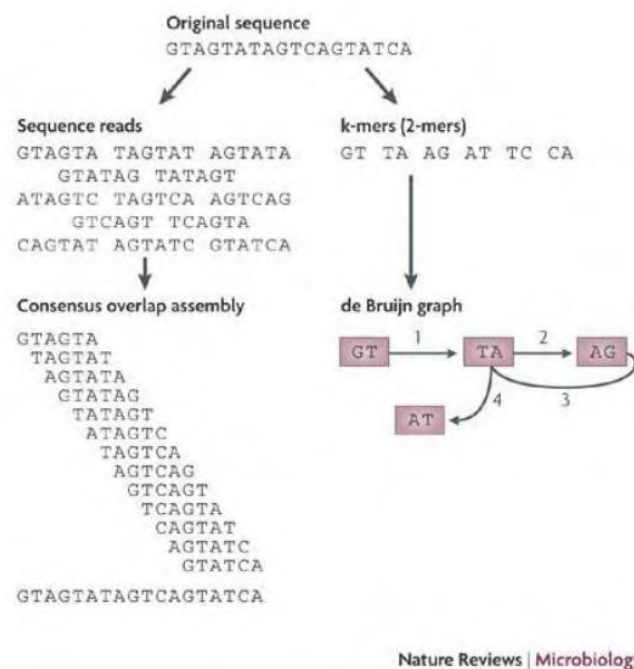
Στην πρώτη περίπτωση, co-assembly, γίνεται χαρτογράφηση με βάση αλληλουχία αναφοράς. Η διαδικασία απαιτεί μεγάλη υπολογιστική ισχύ και αποθηκευτικό χώρο. Εξελιγμένοι αλγόριθμοι βοήθησαν στη βελτίωση του χρόνου της διαδικασίας αλλά βιολογικοί παράμετροι, όπως οι διαφορές μεταξύ των γονιδιωμάτων λόγω προσθήκης, απαλοιφής ή πολυμορφισμού, δυσχεραίνουν τη δημιουργία αξιόπιστων αποτελεσμάτων.

Η de novo συναρμολόγηση απαιτεί μεγαλύτερη υπολογιστική ισχύ και αποθηκευτικό χώρο για να εκτελεσθεί. Μια βασική μέθοδος που βασίζεται η de novo συναρμολόγηση είναι η χρήση του γράφου de Bruijn. Είναι μια δομή δεδομένων για εργαλεία συναρμολόγησης που αναπαριστά συγκεντρωτικά τις αλληλοεπικαλύψεις των reads. Τα εργαλεία που βασίζονται σε αυτή τη δομή δεδομένων αναγνωρίζουν τις αλληλοεπικαλύψεις στα reads και τα συγχωνεύουν για να δημιουργήσουν μεγαλύτερου μήκους αλληλουχίες - contigs. Η μεθοδολογία του de Bruijn γράφου χρησιμοποιήθηκε πρώτη φορά

για αλληλούχιση μέσω υβριδισμού [57]. Η διαδικασία de novo assembling χωρίζεται σε δύο κατηγορίες:

1. Greedy extensions – Επεκτάσεις των reads βάσει στατιστικών παραμέτρων (prefix-tree, commonly base calls)
2. Graph based – Ακριβής προσέγγιση βασιζόμενη σε δημιουργία γράφου (over sampling due to short reads – calculate indexes, error removal step, de Bruijn Graph)

Εικόνα 9: Τα reads τμηματοποιούνται σε k-mers και υπολογίζεται η διαδρομή για τα k-mers, de Bruijn [58]

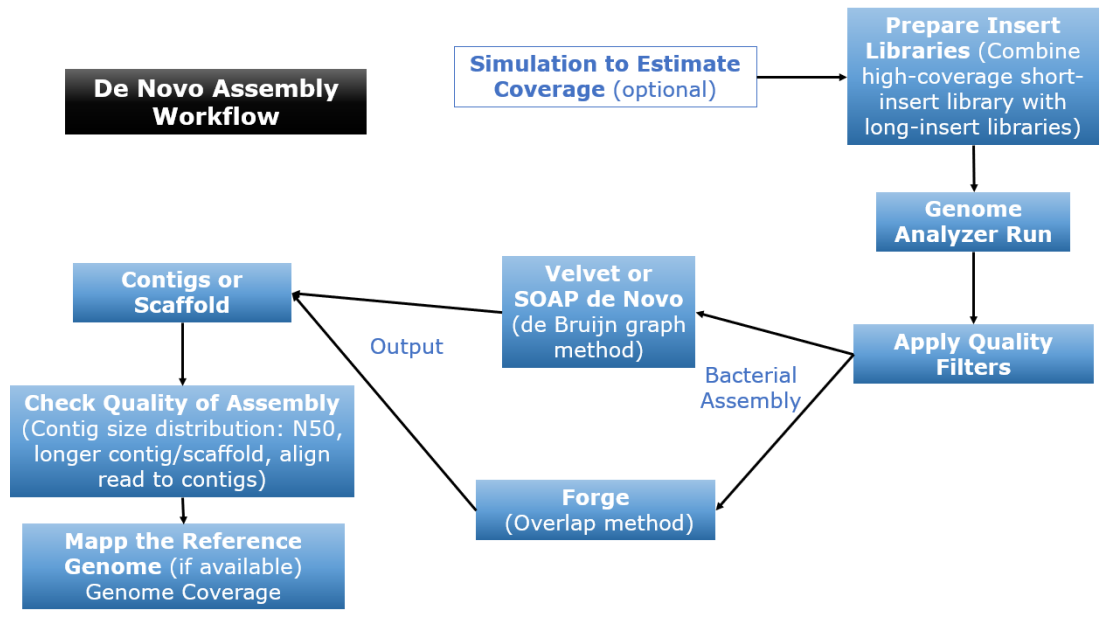


Nature Reviews | Microbiology

Η βασική δομή ενός γράφου de Bruijn είναι μέσω διαμόρφωσης k-mers που προέρχονται από τα reads, αλλά όχι τα reads αυτούσια. Επιπλέον, κάθε επανάληψη είναι εκπροσωπημένη μία μόνο φορά στον γράφο με ρητούς συνδυασμούς προς τα σημεία έναρξης και λήξης. Επίσης, κατά την εκτέλεση της διαδικασίας, ορίζεται βήμα που στοχεύει τη διόρθωση σφαλμάτων που χρησιμοποιεί τοπολογικά χαρακτηριστικά για τον εντοπισμό τους. Με τρεις επιμέρους υποκατηγορίες σφαλμάτων, αρχικά στα άκρα των reads, έπειτα στο ενδιάμεσο τμήμα των συνδεδεμένων reads και τέλος, τα σφάλματα των λάθος

συνδέσεων των reads, τα οποία γίνονται συνήθως από σφάλματα κλωνοποίησης. Ο σύνθετος γράφος de Bruijn χωρίζεται σε ξεχωριστούς υπογράφους με στόχο τη δημιουργία των scaffolds, τα οποία βασίζονται ξεχωριστά σε κάθε ένα γράφο που δημιουργεί. [59] [60].

Διάγραμμα 10: Ροή εργασιών για De Novo Assembly



2.2.4 Binning

Επόμενο βασικό βήμα της ανάλυσης είναι το Binning. Στόχος του βήματος είναι η κατηγοριοποίηση των δημιουργημένων contigs, σύμφωνα με στοιχεία της συστηματικής ταξινόμησης σε διαφορετικές υποομάδες. Η κατηγοριοποίηση γίνεται σύμφωνα με δύο διαφορετικούς τρόπους:

1. Ομαδοποίηση βάσει ταξινομικών συγκρίσεων με γνωστά δεδομένα/αλληλουχίες (π.χ. γένος, οικογένεια)
2. Ομαδοποίηση βάσει κοινών χαρακτηριστικών στις επιμέρους αλληλουχίες (π.χ. GC-content)

Η διαδικασία του Binning είναι ένα από τα πιο σημαντικά βήματα της ανάλυσης. Ανάλογα την μέθοδο που χρησιμοποιείται για τη διαδικασία του binning, παρέχονται πληροφορίες για την παρουσία νέων γονιδιωμάτων που θα ήταν δύσκολο έως αδύνατο να προσδιοριστούν με άλλο τρόπο [55]. Επίσης παρέχονται πληροφορίες για ποσοτικά δεδομένα των πληθυσμών σε ένα δείγμα και μπορούν να διεξαχθούν διαγράμματα διακύμανσης των υποπληθυσμών σε ταξινομική βαθμίδα γένους ή είδους. Επιπρόσθετα, παρέχεται μιας μορφής απλοποίηση της πολυπλοκότητας των δεδομένων, καθώς μετά τη διαδικασία του binning εξάγονται αρχεία με τις ομαδοποιήσεις που έχουν προκύψει, ανάλογα το εργαλείο που έχει επιλεγεί για το binning, και στη συνέχεια μπορεί να εκτελεσθεί ξανά η διαδικασία του assembly στις υποομάδες/bins που έχουν δημιουργηθεί.

Υπάρχουν τρεις βασικές μεθοδολογίες για τους αλγόριθμους των εργαλείων που εκτελούν binning:

1. Sequence composition (βάσει της σύνθεσης των contigs)
2. Sequence similarity (βάσει των ομοιοτήτων των contigs)
3. Fragment recruitment (βάσει της συναρμολόγησης των θραυσμάτων)

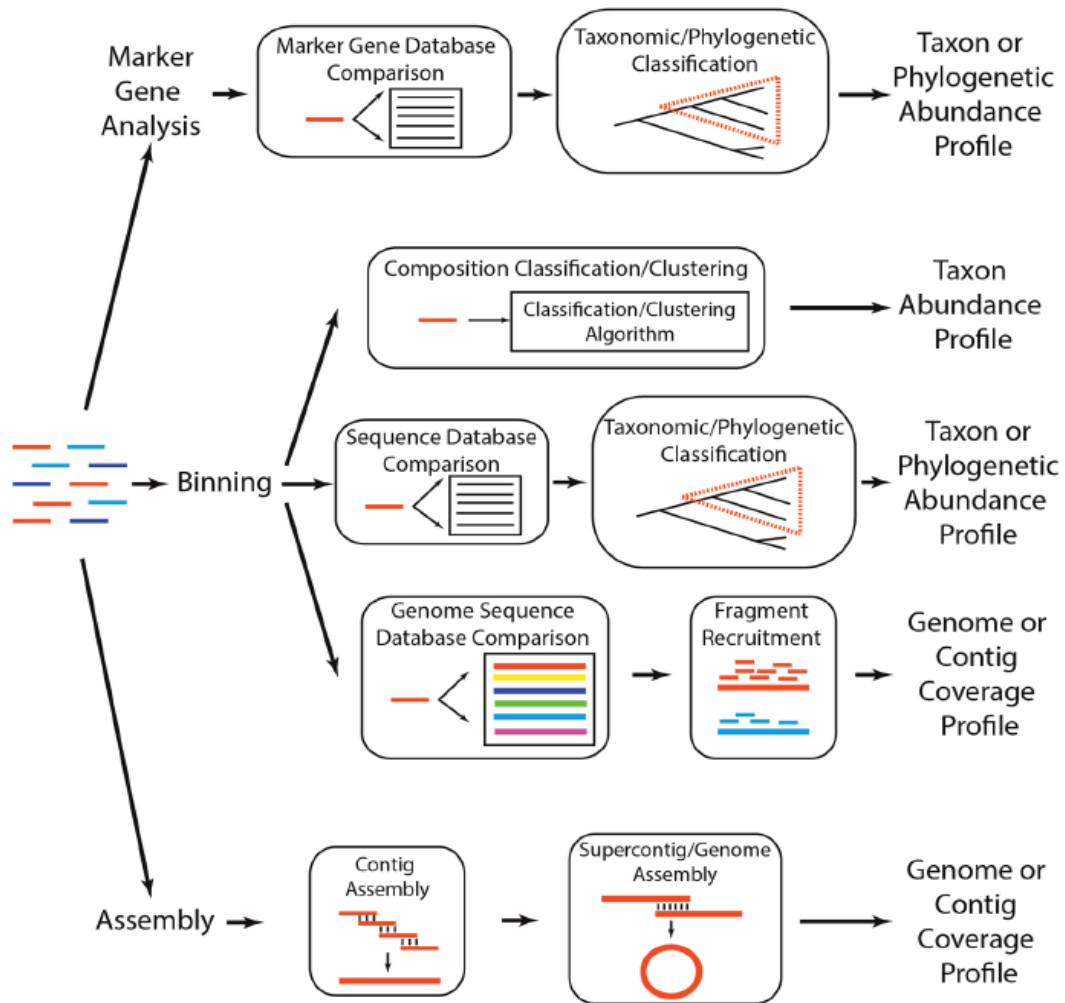
Οι περισσότεροι αλγόριθμοι binning αξιοποιούν δεδομένα των αλληλουχιών, όπως η συχνότητα τετραμερών, με σκοπό την ομαδοποίηση των contigs ή την ταξινόμηση σε ταξινομικές ομάδες. Αυτοί οι μέθοδοι δεν απαιτούν στοίχιση των αλληλουχιών σε γνωστές αλληλουχίες από βάσεις δεδομένων, και για

αυτό το λόγο η διαδικασία γίνεται σε μικρότερο χρονικό διάστημα από ότι στις άλλες δύο μεθόδους.

Επόμενης μορφής αλγόριθμοι εκπαιδεύουν τους λεγόμενους Classifiers με τα αρχικά δεδομένα που εισάγουμε και στη συνέχεια ομαδοποιούν σε ταξινομικές ομάδες τα contigs. Σε αυτή τη μεθοδολογία χρησιμοποιούνται συνήθων support vector machines. Η εκπαίδευση των αλγορίθμων γίνεται μέσω δεδομένων από φυλογενετικές ομάδες, με στόχο την αναγνώριση της συχνότητας εμφάνισης ολιγονουκλεοτιδίων ανά ταξινομική ομάδα και την εφαρμογή της ομαδοποίησης στα νέα δεδομένα που εισάγουμε.

Η τεχνική της ταξινομικής ομαδοποίησης των δεδομένων, χωρίς να πραγματοποιηθεί στοίχιση σε γνωστές αλληλουχίες γονιδιωμάτων από βάσεις δεδομένων, βοηθά στην εύρεση νέων γονιδιωμάτων. Σε αντίθετη περίπτωση, δηλαδή σε εργαλεία binning που πραγματοποιούν στοιχίσεις σε βάσεις δεδομένων, το χρονικό διάστημα ανάλυσης αυξάνεται σε μεγάλο βαθμό, καθιστώντας μη λειτουργική τη διαδικασία. Η επαλήθευση των αποτελεσμάτων της διαδικασίας binning είναι απαραίτητη για την εκτίμηση της εγκυρότητάς τους [56].

Διάγραμμα 11: Ροή εργασίας αναλύσεων [61]



2.2.5 Gene Detection – Gene Annotation

Οι λειτουργικές ιδιότητες επιτρέπουν την επιβίωση και την προσαρμογή των μικροοργανισμών στο περιβάλλον που βρίσκονται. Είναι κωδικοποιημένες πληροφορίες στο γενετικό υλικό και αρκετές φορές δημιουργούν λειτουργικά μοτίβα. Η ταυτοποίηση των γονιδίων στο σύνολο του γονιδιώματος είναι το βήμα που θα επιτρέψει την αποκρυπτογράφηση των ρυθμιστικών λειτουργιών και τον ορθό σχολιασμό τους.

Η ανίχνευση των γονιδίων ξεκινά μέσω της ανίχνευσης των ORFs, Open Reading Frames, που βρίσκονται στο γονιδίωμα και στη συνέχεια την διαπίστωση εάν τα ORFs μεταφράζονται σε λειτουργικά γονίδια, μέσω ανάλυσης Transcriptomics. Οι περισσότεροι αλγόριθμοι που χρησιμοποιούνται για αυτή την ανάλυση υλοποιούνται μέσω μεθοδολογία τεχνητής νοημοσύνης [56].

Απαραίτητη προϋπόθεση για την εξακρίβωση των πληροφοριών που διαθέτουν τα γονίδια και την εξόρυξη γνώσεων από αυτά, είναι ο κατάλληλος σχολιασμός της λειτουργίας τους. Υπάρχουν τρεις επιμέρους τεχνικές ανίχνευσης γονιδίων:

1. Συναρμολόγηση θραυσμάτων γονιδιακής αλληλουχίας.
2. Συναρμολόγηση βάσει ταξινόμησης πρωτεϊνικών οικογενειών.
3. De novo πρόβλεψη γονιδίων.

Στις περισσότερες περιπτώσεις χρησιμοποιείται η τεχνική διερεύνησης της ύπαρξης ή μη, ομολογίας σε γνωστά γονίδια, τα οποία είναι κατατεθειμένα σε βάσεις δεδομένων, περίπτωση τεχνικής 1 και 2.

Στην περίπτωση των μικροοργανισμών, τα δεδομένα μεταγονιδιωματικής που διαθέτουμε σε βάσεις δεδομένων είναι περιορισμένα. Για τη λύση αυτού του προβλήματος έχουν προκύψει νέες τεχνικές πρόβλεψης, όπως τα Hidden Markov Models και τεχνικές machine learning – μηχανικής μάθησης, περίπτωση τεχνικής 3.

Μέσω της χρήσης εξειδικευμένων εργαλείων επιδιώκεται η αποκρυπτογράφηση των ομοιοτήτων που βρίσκονται στο σύνολο των αλληλουχιών, όπως συγκεκριμένα μοτίβα. Στοιχεία τα οποία μπορούν να

συνεισφέρουν σε αυτό το σημείο είναι το μήκος των αλληλουχιών και η περιεκτικότητα σε GC. Στη συνέχεια, πραγματοποιείται σύγκριση των μοτίβων που βρέθηκαν στις αλληλουχίες της ανάλυσης και των μοτίβων που βρίσκονται σε γονίδια με γνωστή λειτουργία.

2.3 Πλατφόρμες και Εργαλεία

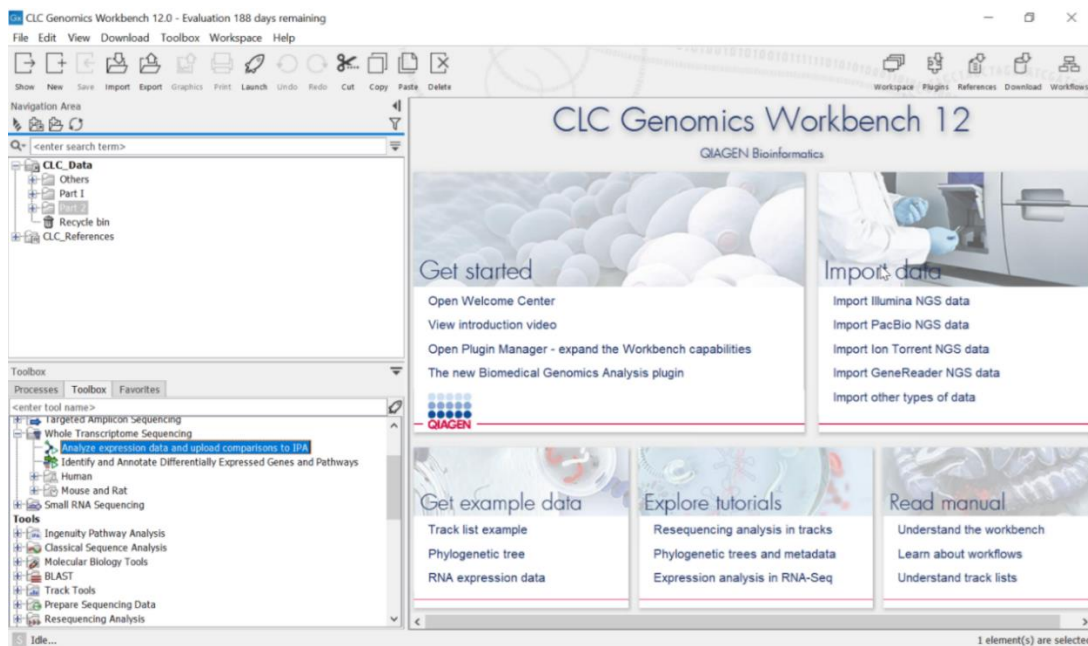
2.3.1 CLC Platform

Η πλατφόρμα CLC Workbench αναπτύχθηκε το 2005 από την εταιρεία Qiagen.¹¹ Η πρώτη έκδοση της πλατφόρμας περιείχε δύο ροές εργασιών, CLC Free Workbench και CLC Protein Workbench. Με την πάροδο του χρόνου και τη συνεχόμενη εξέλιξη των τεχνικών ανάλυσης δεδομένων αλληλούχισης, προέκυψαν διαφορετικές εκδόσεις με βελτιωμένες ροές εργασιών. Το 2008, λόγω της ανάγκης για ανάλυση δεδομένων που προκύπτουν από διαφορετικές πλατφόρμες αλληλούχισης (Illumina, Roche-454, SOLiD), αναπτύχθηκε η πλατφόρμα CLC Genomics Workbench. Τα επιμέρους εργαλεία έχουν υλοποιηθεί σε γλώσσα προγραμματισμού Java. Από το 2012 η εταιρεία αποφάσισε να παρέχει πρόσβαση σε εξελιγμένες version της πλατφόρμας μόνο επί πληρωμή. Η πλατφόρμα παρέχει ένα φιλικό περιβάλλον προς το χρήστη με υποχρεωτική επιλογή παραμέτρων σε όλα τα βήματα της ανάλυσης.

Για την ανάλυση των αρχικών δεδομένων χρησιμοποιήθηκε η ροή εργασίας CLC Microbial Genomic Module της CLC Genomic Workbench 12.

¹¹ <https://digitalinsights.qiagen.com/>

Εικόνα 10: Interface της πλατφόρμας CLC Genomic Workbench



Μέσω της CLC Microbial Genomic Module παρέχεται η δυνατότητα προσδιορισμού των μικροβιωμάτων που αποτελείται το αρχικό σετ δεδομένων. Υπάρχει η δυνατότητα ανάλυσης και εξαγωγής στατιστικών αποτελεσμάτων για την αφθονία των μικροβιωμάτων σε επίπεδο είδους. Δίνεται η επιλογή φιλτραρίσματος των αρχικών δεδομένων από τυχόν μολύνσεις που προέρχονται από DNA ξενιστών.

Στην περίπτωση μας πραγματοποιήθηκε απομόνωση των θραυσμάτων DNA των ειδών *Ovis aries* (πρόβατο) και *Capra aegagrus* (αίγα) από τα αρχικά δεδομένα, καθώς τα δείγματα προέρχονται από τα δύο είδη μηρυκαστικών. Η διαδικασία έλαβε χώρα μετά από αναζήτηση σε βάσεις δεδομένων, GenBank και RefSeq. Τα δεδομένα που βρίσκονται κατατεθειμένα στη RefSeq προέρχονται από αντίγραφα καταχωρημένα στην GenBank σε assembled μορφή. Σε ορισμένες περιπτώσεις τα assembled δεδομένα της RefSeq δεν είναι πανομοιότυπα με τα δεδομένα της GenBank καθώς υπάρχουν προσθέσεις non-nuclear organelle units ή αποκλεισμός μικρού μήκους αλληλουχιών. Και στις δύο περιπτώσεις βρέθηκαν δεδομένα των χρωμοσωμικών αλληλουχιών των δύο ειδών με GenBank Assembly Accession: GCA_000978405.1 για το είδος *Capra aegagrus* και GenBank Assembly Accession: GCA_002742125.1 του είδους *Ovis aries*.

CLC Microbial Genomic Module

Σε συνέχεια της ροής εργασιών στην πλατφόρμα CLC Microbial Genomic Module, εφόσον έχει γίνει η απομάκρυνση του γενετικού υλικού των δύο μηρυκαστικών από τα αρχικά δεδομένα, πραγματοποιήθηκε Assembly των reads μέσω αλγορίθμου που χρησιμοποιεί τεχνική de novo Assembly για την δημιουργία contigs. Η πρώτη παράμετρος που πρέπει να ορίσει ο χρήστης είναι το ελάχιστο μήκος που απαιτείται για την συναρμολόγηση ενός contig από τα reads. Βάσει των αρχικών δεδομένων και της ανάλυσης που επιδιώκεται, υπάρχει η επιλογή παραμέτρων για εκτέλεση του αλγορίθμου, μια φορά με χρήση $k\text{-mers}=21$, ή τρεις φορές με διαφορετικά ορίσματα $k\text{-mers}=(21,41,61)$ για κάθε έναν από τους τρεις κύκλους εκτέλεσης.

Μετά από τη διαδικασία του de novo Assembly, παρέχεται η δυνατότητα Taxonomic Analysis. Ο αλγόριθμος πραγματοποιεί στοίχιση των reads σε ολοκληρωμένα γονιδιώματα που βρίσκονται κατατεθειμένα σε βάσεις δεδομένων που ορίζει ο χρήστης. Μετά την εκτέλεση της στοίχισης υλοποιείται ποσοτική και ποιοτική ανάλυση των δεδομένων αφθονίας για κάθε taxon ξεχωριστά και στο τέλος δημιουργείται συγκεντρωτικό πίνακα στατιστικής αφθονίας των ειδών για το δείγμα.

Ο στόχος του ποιοτικού ελέγχου είναι να καθορίσει εάν τα συγκεκριμένα taxa που υπολογίστηκαν στο δείγμα δεν έχουν προκύψει από στατιστικό λάθος. Ο τρόπος με τον οποίο επιτυγχάνεται είναι μέσω του confidence-score. Σε περιπτώσεις που η μεταβλητή είναι <0.995 , οι αλληλουχίες που στοιχήθηκαν στο συγκεκριμένο taxon απορρίπτονται και εκχωρούνται στον πλησιέστερο συγγενικά πρόγονο.

Ο ποσοτικός έλεγχος γίνεται για την εκτίμηση της αφθονίας των taxon που βρίσκονται στο δείγμα και βασίζεται στον αριθμό των reads που έχουν εκχωρηθεί σε κάθε taxon ξεχωριστά. Ο χρήστης ορίζει τις παραμέτρους για την υλοποίηση του αλγορίθμου και τη δημιουργία του πίνακα αφθονίας του δείγματος.

Στην ροή εργασίας CLC Genomic Workbench πραγματοποιήθηκε μια σειρά διεργασιών για τη διερεύνηση των μικροοργανισμών που αποτελούνται τα δείγματα.

Αρχικά πραγματοποιείται φιλτράρισμα των δεδομένων για δημιουργία συγκρίσιμου μήκους στα reads. Η διαδικασία αυτή γίνεται μέσω παραμετροποιήσεων του χρήστη, με όρισμα σταθεράς για το κατώτατο όριο του ελαχίστου αριθμού των reads που υποδεικνύει εάν ένα δείγμα έχει επαρκή κάλυψη. Σε όλα τα βήματα της ανάλυσης, υπάρχει η δυνατότητα εξαγωγής αρχείου με τα αποτελέσματα.

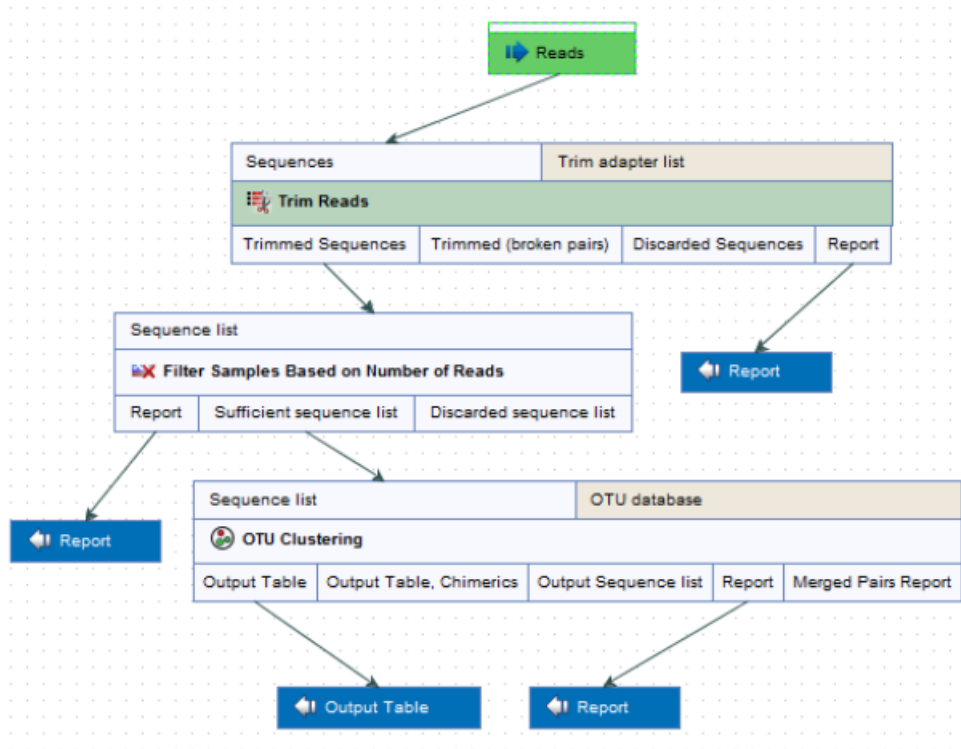
Στη συνέχεια πραγματοποιήθηκε ανάλυση OTUs clustering μέσω στοίχισης των reads με γνωστές αλληλουχίες που είναι καταχωρημένες σε βάσεις δεδομένων. Η βάση δεδομένων που χρησιμοποιήθηκε για την ανάλυση 16S στη μελέτη είναι η Greengenes και για την ανάλυση ITS δεδομένων η UNITE. Η Greengenes περιέχει ολοκληρωμένες αλληλουχίες του 16S rDNA γονιδίου με επιμελημένους σχολιασμούς. Τη δεδομένη στιγμή περιέχει 1.012.863 αλληλουχίες του γονιδίου 16S rDNA από 92.684 οργανισμούς, με μέσο μήκος των αλληλουχιών 1.402 νουκλεοτίδια. Η βάση δεδομένων UNITE περιέχει 2.480.043 αλληλουχίες της περιοχής ITS από ευκαρυωτικούς οργανισμούς αλλά εστιάζει κυρίως στους μύκητες.

Η στοίχιση που πραγματοποιείται δημιουργεί πίνακα βαθμολογήσεων - alignment score για κάθε OTUs. Το αρχείο εξόδου της ανάλυσης περιέχει λίστες αλληλουχιών των OTUs, πίνακα με δεδομένα αφθονίας και γραφήματα αποτύπωσης των αποτελεσμάτων. Οι βασικοί παράμετροι που ορίστηκαν είναι η βάση δεδομένων που θα χρησιμοποιηθεί για την ανάλυση, το ποσοστό ομοιότητας των αλληλουχιών για κάθε OTUs, τιμή σταθεράς για την δημιουργία νέου OTUs, το ποσοστό ταξινομικής ομοιότητας των μικροοργανισμών όπως και το μέγεθος των k-mers. Τέλος, υπάρχει η δυνατότητα ανάλυσης της alpha και beta diversity των δειγμάτων.

Οι βάσεις δεδομένων που υποστηρίζονται από την πλατφόρμα CLC Genomic Workbench είναι η Greengenes για 16S rDNA δεδομένα αλληλούχισης προκαρυωτικών οργανισμών, Silva για 16S/18S rRNA για προκαρυωτικούς

και ευκαρυωτικούς οργανισμούς και τέλος η βάση δεδομένων UNITE για δεδομένα ITS αλληλούχισης για ταξινόμηση μυκητών.

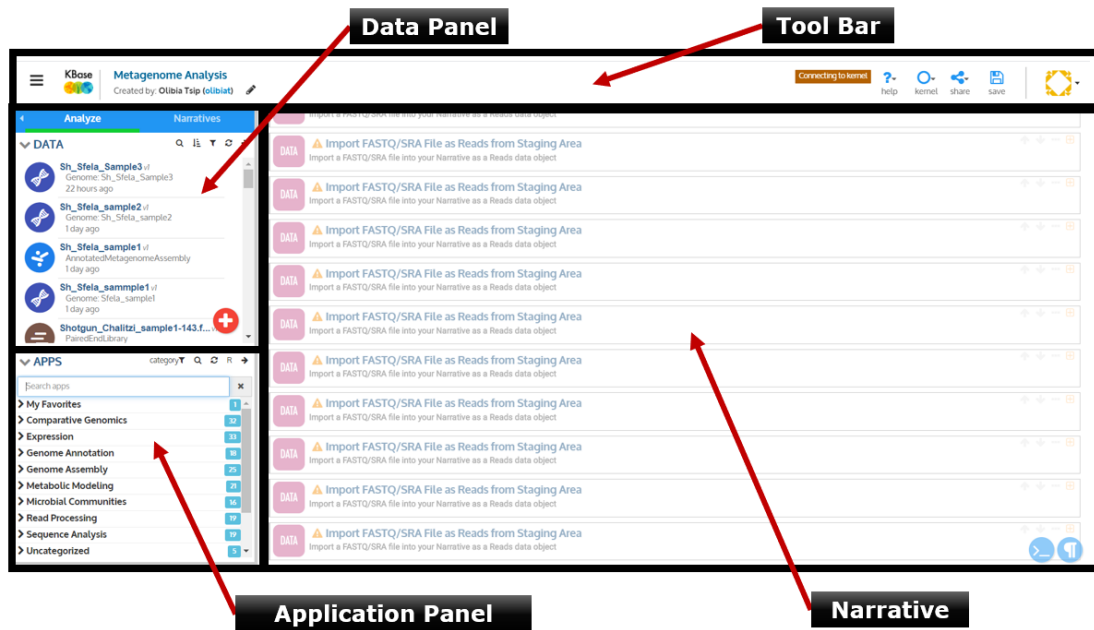
Εικόνα 11: Ροή εργασιών CLC Genomic Workbench για Amplicon ανάλυση



2.3.2. KBase Platform

To Department of Energy Systems Biology Knowledgebase (KBase), Washington DC, υλοποίησε ανοιχτό λογισμικό σχεδιασμένο να ενσωματώνει δεδομένα και επιμέρους εργαλεία σε μια ενοποιημένη μορφή [62]. Το διαμορφωμένο περιβάλλον είναι φιλικό προς το χρήστη και διαθέτει σε κάθε εργαλείο τη δυνατότητα παραμετροποίησης πριν την εκτέλεση του αλγορίθμου. Υπάρχει η δυνατότητα εκτέλεσης αναλύσεων μεγάλης κλίμακας, με συνδυασμό των εργαλείων και μοντελοποίηση της ροής εργασιών. Η διάθεση του προς το κοινό πραγματοποιήθηκε το 2018 και μέχρι αυτή τη στιγμή είναι ελεύθερο προς χρήση. Για την αξιοποίηση της πλατφόρμας, ο χρήστης πρέπει να δημιουργήσει δικό του περιβάλλον και να επιλέξει τα εργαλεία που θα χρησιμοποιήσει στην ανάλυσή του.

Εικόνα 12: Interface της πλατφόρμας KBase

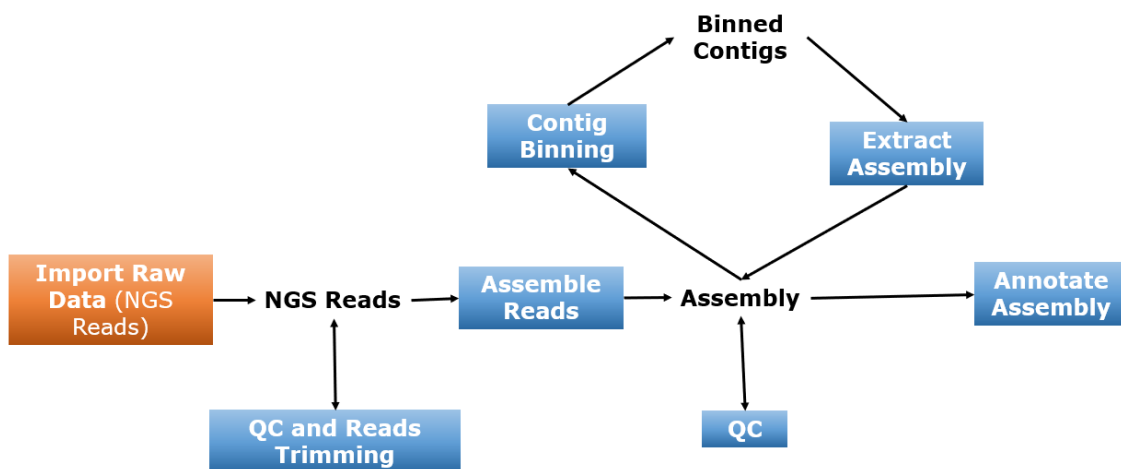


Το πρώτο βήμα για την ανάλυση στην πλατφόρμα KBase είναι η δημιουργία του Narrative στο οποίο θα γίνει το upload των FastQ files των αρχικών δεδομένων. Η χρονική διάρκεια για την ολοκλήρωση του upload των αρχείων ενός δείγματος μεγέθους $\approx 5-7\text{GB}$ είναι ≈ 26 ώρες και διατηρούνται έως και 3 μήνες στην πλατφόρμα. Σε επόμενο βήμα, επιλέγονται τα επιμέρους εργαλεία

που θέλουμε να χρησιμοποιήσουμε. Ο κώδικας των εργαλείων μπορεί να εξαχθεί από την πλατφόρμα του GitHub και είναι διαθέσιμος για τοπική εγκατάσταση.

Στην παρακάτω εικόνα αποτυπώνεται η ροή εργασίας. Στη συνέχεια θα αναφερθούμε στα επιμέρους εργαλεία που χρησιμοποιήθηκαν για την ανάλυση.

Εικόνα 13: Ροή εργασίας της Ανάλυσης



2.3.2.1 Εργαλεία Assembly

IDBA-UD - v1.1.3

Το εργαλείο IDBA-UD v1.1.3 εκτελεί *de novo assembly* με εξελιγμένο αλγόριθμο του αρχικού IDBA, που βασίζεται σε μέθοδο πολλαπλών *de Bruijn* γράφων, για μικρού μήκους reads με υψηλό βάθος ανομοιογένειας στον βαθμό πιστότητας της αλληλούχισης [63] [64]. Η αρχική έκδοση του αλγορίθμου εκτελούσε υπολογισμούς για την επιλογή της βέλτιστης τιμής *k-mer* και τη δημιουργία των *de Bruijn* γράφων. Αποθηκεύονται διαδοχικά οι υπολογισμοί των διαφορετικών τιμών *k-mers* και στη συνέχεια εκτελούσε σύγκριση των αποτελεσμάτων για την επιλογή της βέλτιστης τιμής.

Στην περίπτωση που η τιμή του *k-mer* είναι μεγάλη δημιουργούνται προβλήματα κενών - *gaps* στους επιμέρους γράφους, ενώ στις περιπτώσεις που η τιμή του *k-mer* είναι μικρή δημιουργούνται προβλήματα πολλαπλών διακλαδώσεων στο γράφο.

Η νέα έκδοση του αλγορίθμου λειτουργεί με την ίδια μεθοδολογία, δηλαδή σταδιακή αύξηση της τιμής του *k-mer* για επιλογή της βέλτιστης τιμής μετά από τους απαραίτητους υπολογισμούς. Επιπρόσθετο χαρακτηριστικό της νέας έκδοση είναι η απομάκρυνση των *contigs* με μικρό μήκος και χαμηλό βαθμό πιστότητα σε κάθε επανάληψη. Η βελτίωση έχει στόχο τη μείωση των *errors* στο τελικό αποτέλεσμα. Με αυτή την τροποποίηση δίνεται η δυνατότητα στον αλγόριθμο να δημιουργεί *de Bruijn* γράφους μεγάλων τιμών *k-mers* με λιγότερες διακλαδώσεις για τη δημιουργία μεγαλύτερου μήκους *contigs*.

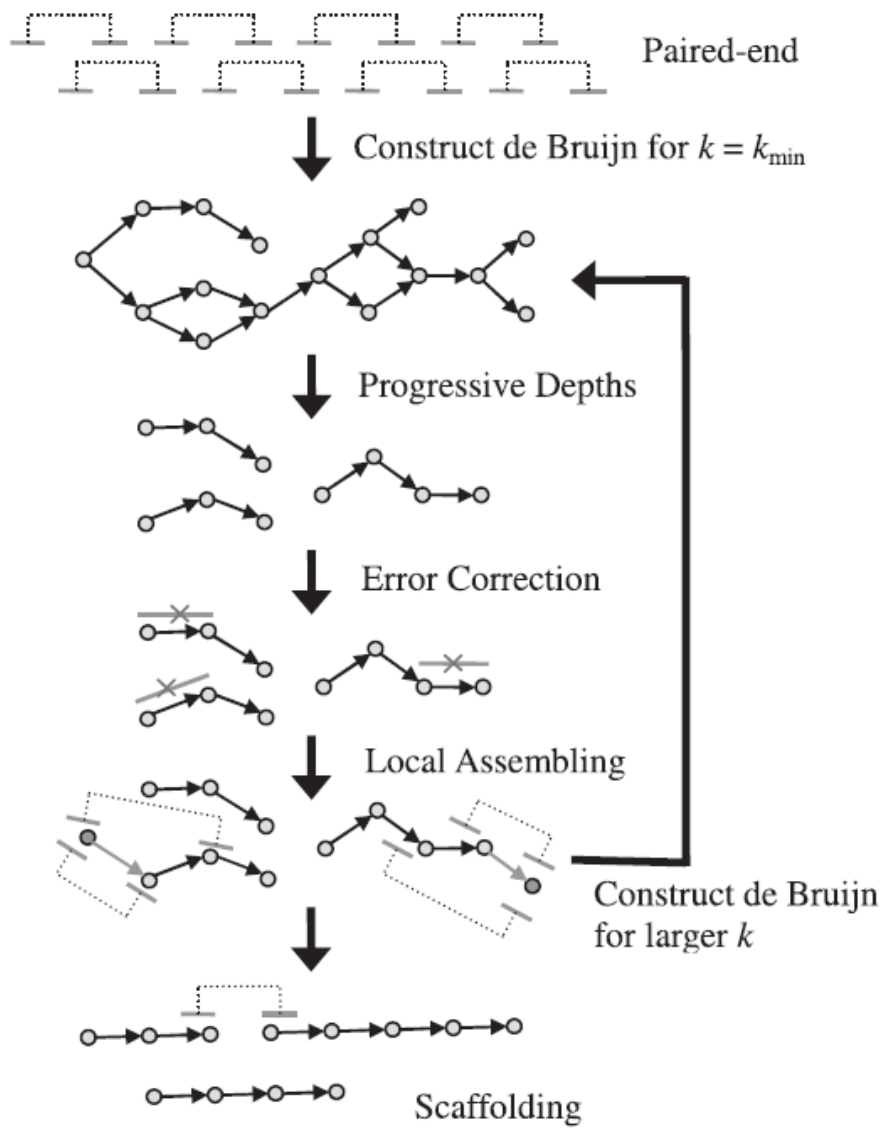
Με τον όρο «βαθμό κάλυψης» εννοούμε τον αριθμό που έχει διαβαστεί κατά την αλληλούχιση μια συγκεκριμένη βάση ανά φορά διαβάσματος. Ο υψηλός βαθμός κάλυψης αυξάνει τον βαθμό εμπιστοσύνης στο τελικό αποτέλεσμα.

Το IDBA-UD είναι εργαλείο ανοιχτού λογισμικού¹² ¹³, σε γλώσσα προγραμματισμού Python και εκτελείται σε περιβάλλον UNIX.

¹² https://github.com/kbaseapps/kb_IDBA/

¹³ <https://github.com/loneknightpy/idba>

Εικόνα 14: Ροή εκτέλεσης υπολογισμών για δημιουργία του de Bruijn γράφου στο IDBA [63]



2.3.2.2 Εργαλείο Binning

MaxBin2 - v2.2.4

Το εργαλείο MaxBin εκτελεί διαδικασία binning των contigs και δίνει αποτελέσματα με υψηλή απόδοση για τη διαδικασία της διερεύνησης των μικροβιακών πληθυσμών που βρίσκονται στα μεταγονιδιωματικά δεδομένα [65]. Η έκδοση 2.2.4 που χρησιμοποιεί η πλατφόρμα KBase είναι μια εξελιγμένη μορφή με διευρυμένο αλγόριθμο ο οποίος αναλύει περισσότερα βακτηριακά γονιδιώματα και συγκρίνει τη σύνθεση μίας μικροβιακής κοινότητας με άλλα περιβαλλοντικά δείγματα.

Πιο συγκεκριμένα, ο αλγόριθμος χρησιμοποιεί το βαθμό κάλυψης (depth of coverage), τη σύνθεση των νουκλεοτιδίων στην αλληλουχία και γονιδιακούς δείκτες για να υλοποιήσει φιλτράρισμα και διαχωρισμό των contigs σε διαφορετικά bins. Ομαδοποιεί τα contigs που έχουν δημιουργηθεί από τη εκτέλεση του assembly σε ομάδες-bins, καθένα από τα οποία αντιστοιχεί σε ξεχωριστό πληθυσμό. Χρησιμοποιεί τη σύνθεση των νουκλεοτιδίων που βρίσκονται στα contigs και υπολογίζεται η αφθονία από τα αρχικά δεδομένα με μετρήσεις του βάθους κάλυψης μέσω στοίχισης των reads στα contigs.

Ο αλγόριθμος Expectation-Maximization (EM) που χρησιμοποιεί, βασίζεται στην επαναληπτική μέθοδο για την εκτίμηση της μέγιστης πιθανοφάνειας - maximum likelihood (ML) ή την εκτίμηση της μέγιστης μεταγενέστερης πιθανοφάνειας - maximum posteriori (MAP). Η μέγιστη μεταγενέστερης πιθανοφάνεια είναι μπαγιεσιανή στατιστική μέθοδος εκτίμησης μιας άγνωστης μεταβλητής και βασίζεται σε εμπειρικά δεδομένα.

Και στις δύο περιπτώσεις γίνεται εκτίμηση των παραμέτρων από ένα στατιστικό μοντέλο δεδομένων. Για παράδειγμα, στις περιπτώσεις που οι παράμετροι ενός συστήματος είναι πολλοί και δεν έχουν ορισθεί με ακρίβεια, δημιουργείται ένα στατιστικό μοντέλο που διαθέτει γνωστές παραμέτρους. Ο αλγόριθμος στη συνέχεια ορίζει τις άγνωστες παραμέτρους σύμφωνα με την κατανομή των γνωστών για να μπορέσει να εξάγει αποτελέσματα βασισμένα σε γνωστά δεδομένα. Η μέθοδος εκτίμησης μέγιστης πιθανοφάνειας, λαμβάνει τη μέση τιμή και τη διακύμανση των δεδομένων του συστήματος ως

παράμετρο. Στη συνέχεια ορίζονται ειδικές τιμές για τις παραμέτρους στο στατιστικό μοντέλο, με στόχο να μεγιστοποιηθεί η βελτίωση της συνάρτησης πιθανότητας. Η διαφορά των δύο στατιστικών μεθόδων έγκειται στο γεγονός ότι η MAP χρησιμοποιεί αυξημένο στόχο βελτιστοποίησης. Ενσωματώνει μια προγενέστερη κατανομή, ποσοτικοποιώντας τις επιπρόσθετες πληροφορίες από προγενέστερα δεδομένα, και συσχετίζοντάς αυτά με τα δεδομένα του προβλήματος. Η εκτίμηση MAP μπορεί να θεωρηθεί ως μια κανονικοποίηση της εκτίμησης ML.

Τα αρχεία εισόδου για την υλοποίηση του αλγορίθμου είναι το αρχείο FastQ των αρχικών δεδομένων (reads) και το αρχείο που εξήχθη από την εκτέλεση του assembly. Στη συνέχεια της εκτέλεσης του αλγορίθμου δημιουργούνται τα διαφορετικά bins. Τα αρχεία εξόδου του εργαλείου περιλαμβάνουν τα διαφορετικά bins και στατιστικά δεδομένα, όπως το ποσοστό πληρότητας, περιεχόμενο GC των αλληλουχιών και το μέγεθος των γονιδιωμάτων. Ο κώδικας του αλγορίθμου είναι υλοποιημένος σε Perl και C++, ανοιχτός προς τοπική εγκατάσταση και επεξεργασία¹⁴.

¹⁴ <http://sourceforge.net/projects/maxbin/>

2.3.2.3 Εργαλείο για Βελτιστοποίηση και Φιλτράρισμα των Bins

2.3.2.3.1 CheckM - v1.0.18

Ένα από τα εργαλεία που χρησιμοποιήθηκε για το φιλτράρισμα των διαμορφωμένων bins είναι το CheckM v1.0.18 [66]. Το CheckM εκτελεί ροή εργασιών για την εκτίμηση της ποιότητας των γονιδιωμάτων που ομαδοποιήθηκαν μέσω συγκρίσεων με βάσεις δεδομένων. Δημιουργεί ένα νέο «αντικείμενο» που ονομάζει Binned Contigs, στο οποίο τοποθετεί τα bins με υψηλή ποιότητα σύμφωνα με τις παραμετροποιήσεις που εκτελέστηκαν. Η ποιότητα είναι βάση της τιμής κατώφλι που έχει ορίσει ο χρήστης για τις παραμέτρους πληρότητας και μόλυνσεων. Μετά την εκτέλεση του αλγορίθμου, διατηρούνται μόνο εκείνα τα bins που έχουν περάσει το κατώφλι των παραμέτρων που έχουν ορισθεί.

Στην περίπτωση των μεταγονιδιωματικών δεδομένων, το εργαλείο χρησιμοποιείται πριν την εκτέλεση του Annotation – Σχολιασμό των bins, καθώς μπορεί να δώσει πληροφορίες σχετικά με την αξιοπιστία των bins και θα απορρίψει εκείνα που δημιουργήθηκαν από στατιστικά λάθη.

Τα αρχεία εξόδου του εργαλείου είναι γραφικές παραστάσεις της πληρότητας και των μόλυνσεων των bins, συνδυαστικά με πίνακα που περιέχουν πληροφορίες, όπως ταξινομικά δεδομένα. Το εργαλείο χρησιμοποιεί σύνολα συντηρημένων γονιδιακών δεικτών για κάθε bin με στόχο την εξαγωγή αποτελεσμάτων.

Η βάση δεδομένων που έχει χρησιμοποιηθεί για τους γονιδιακούς δείκτες είναι η IMG/M – Integrated Microbial Genomes & Microbiomes, και συγκεκριμένα 39 Classes και 20 Phyla. Αυτοί οι γονιδιακοί δείκτες, επιλέχθηκαν με τυχαίο τρόπο και θραυσματοποιήθηκαν σε μη επικαλυπτόμενα 5-50kbp τμήματα. Το εργαλείο χρησιμοποιεί γλώσσα προγραμματισμού Python και είναι ελεύθερου λογισμικού¹⁵.

¹⁵ <https://github.com/Ecogenomics/CheckM>

2.3.2.3.3 KB - Compare Assembled Contig Distributions - v1.1.2

Το εργαλείο KB v1.1.2 υλοποιεί έλεγχο των συναρμολογημένων-assembled αρχείων, βάσει του μήκους και του αριθμού της κατανομής των επιμέρους contigs [62]. Η είσοδος του εργαλείου είναι τα αρχεία εξόδου των εργαλείων που εκτελούν Assembly και μπορεί να πραγματοποιήσει σύγκριση των assembled αρχείων από τα αποτελέσματα διαφορετικών εργαλείων Assembly. Τα αρχεία εξόδου διαθέτουν διαγράμματα, όπως ιστογράμματα, με υπολογισμούς που εκτελέστηκαν για την κατανομή και τη διακύμανση των contigs με μεγαλύτερο μήκος. Μέσω του αποτελέσματος της υλοποίησης, ο χρήστης μπορεί να επιλέξει τα αρχεία Assembled με τα οποία θα συνεχίσει την ανάλυση των μεταγονιδιωματικών δεδομένων. Είναι υλοποιημένο σε γλώσσα Python και υπάρχει δυνατότητα εκτέλεσης του τοπικά μέσω του λειτουργικού Docker¹⁶.

2.3.2.4 Εργαλεία για Taxonomic Analysis

Ένα από τα μεγαλύτερα ερωτήματα της ανάλυσης μεταγονιδιωματικών δειγμάτων είναι η σύνθεση της μικροβιακής κοινότητας μέσω της συστηματικής ταξινόμησης. Εργαλεία που έχουν αναπτυχθεί για εκτέλεση τέτοιων υπολογισμών αντιμετωπίζουν προβλήματα κυρίως λόγω της έλλειψης υπολογιστικής ισχύς και του αυξημένου αποθηκευτικού χώρου που απαιτείται για τις αναλύσεις. Στη συνέχεια παρουσιάζονται τα εργαλεία που εκτελέστηκαν για ταξινομικές αναλύσεις στην παρούσα μελέτη.

2.3.2.4.1. KAIJU - v1.7.2

Το KAIJU είναι εργαλείο που πραγματοποιεί ταξινομική ανάλυση σε δεδομένα μεγάλης κλίμακας. Χρησιμοποιεί τα αρχικά δεδομένα της αλληλούχισης (Raw Data) και εκτελεί ταξινομική διερεύνηση για κάθε read μέσω στοίχισης σε βάσεις δεδομένων. Αξιοποιεί πληροφορίες που είναι κατατεθειμένες στη βάση

¹⁶ https://github.com/kbaseapps/kb_assembly_compare

RefSeq του NCBI. Σε μικροβιακά δείγματα χρησιμοποιεί το υποσύνολο του NCBI BLAST της Non-Redundant πρωτεϊνικής βάσης δεδομένων [67].

Η διαδικασία που ακολουθεί ο αλγόριθμος του εργαλείου είναι να μεταφράζει τα reads σε αμινοξικά κατάλοιπα βάση επιλεγμένου γενετικού κώδικα από τον χρήστη ή να εκτελεί στοίχιση των νουκλεοτιδικών αλληλουχιών.

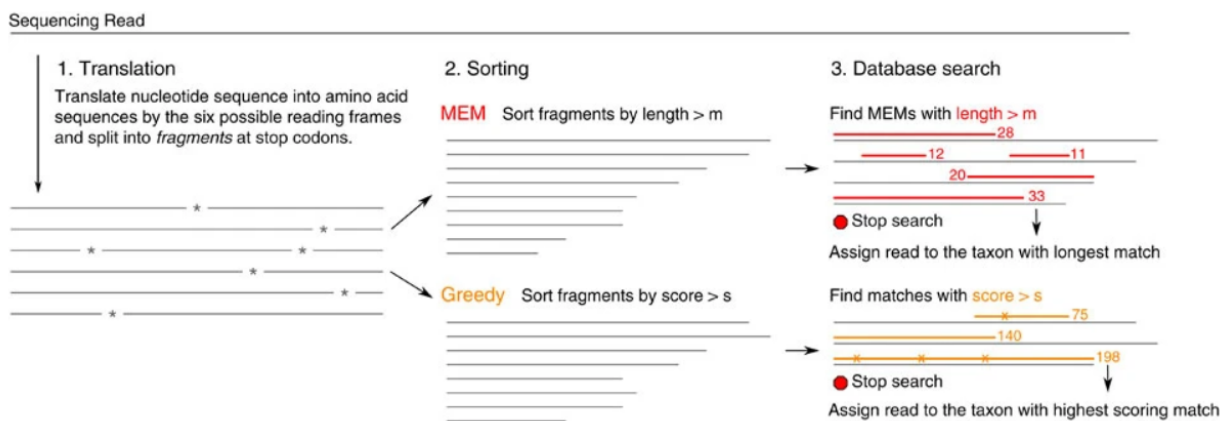
Στη συνέχεια πραγματοποιεί αναζήτηση στη βάση δεδομένων αξιοποιώντας μια τροποποιημένη αναζήτηση μέσω του μετασχηματισμού Burrows – Wheeler (BWT) που ονομάζεται αλλιώς block-sorting-compression. Ο μετασχηματισμός αυτός αναδιατάσσει μια συμβολοσειρά χαρακτήρων σε σειρά παρόμοιων χαρακτήρων. Αυτού του είδους η μεθοδολογία χρησιμοποιείται από εργαλεία συμπίεσης αρχείων. Το σημαντικότερο χαρακτηριστικό της είναι ότι είναι αναστρέψιμη μέσω αποθήκευσης της θέσης του πρώτου χαρακτήρα στη σειρά, χωρίς να χαθούν πληροφορίες. Θεωρείται μέθοδος βελτίωσης της αποτελεσματικότητας των αλγορίθμων συμπίεσης κειμένων και χρησιμοποιείται στη συγκεκριμένη περίπτωση γιατί τα δεδομένα καταλαμβάνουν μεγάλο αποθηκευτικό χώρο. Χωρίς τη χρήση αυτής της τεχνικής για την εκτέλεση του αλγορίθμου, δεν θα ήταν δυνατή η στοίχιση του κάθε read σε βάση δεδομένων. Η αναζήτηση στις βάσεις δεδομένων μπορεί να εκτελεστεί για εκατομμύρια reads ανά λεπτό έχοντας το ελάχιστο 10 GB RAM για μια βάση δεδομένων που περιλαμβάνει 4.821 πλήρη γονιδιώματα μικροοργανισμών.

Το KAIJU εκτελεί ταξινόμηση των contigs με υψηλή ευαισθησία και ακρίβεια σε σύγκριση με άλλα εργαλεία ταξινομικής ανάλυσης, ιδιαίτερα σε γένη που υποεκπροσωπούνται στις βάσεις δεδομένων. Ωστόσο, όπως αναφέρεται στις επιμέρους υλοποιήσεις του αλγορίθμου, η υπολογιστική διαδικασία με αρχικά δεδομένα τα contigs, δημιουργούν πρόσθετες υπολογιστικές προκλήσεις. Η ταχύτερη μέθοδος γίνεται μέσω της ανάλυσης των reads από τα αρχικά μεταγονιδιωματικά δεδομένα.

Το εργαλείο αξιοποιεί τέσσερεις βάσεις δεδομένων: RefSeq Complete Genomes, proGenomes, NCBI BLAST nr και NCBI BLAST nr+euk. Στην περίπτωση της μελέτης επιλέχθηκε η RefSeq Genomes (no Euks). Η διαδικασία στοίχισης που ακολουθεί είναι βασισμένη στη δημιουργία δομών ευρετηρίου,

hashbased index structures, για το σύνολο των αλληλουχιών αναφοράς που είναι κατατεθειμένες στις βάσεις δεδομένων πριν ξεκινήσει η στοίχιση με τα δεδομένα των δειγμάτων. Το ευρετήριο που δημιουργείται από τις βάσεις δεδομένων (k-mers words), χρησιμοποιείται στη συνέχεια για την αναζήτηση στοίχισης με δεδομένα τα reads. Το εργαλείο είναι υλοποιημένο σε γλώσσα προγραμματισμού C/C++ σε περιβάλλον LINUX, ελεύθερο για τοπική εγκατάσταση¹⁷.

Εικόνα 15: Ροή εργασιών του αλγορίθμου που υλοποιείται από το εργαλείο KAIJU [67]



2.3.2.4.2. GOTTCOA2 - v2.1.6

Το εργαλείο GOTTCOA (Genomic Origins Through Taxonomic CHallenge) εκτελεί ταξινομική ανάλυση σε μεταγονιδιωμιακά δεδομένα χωρίς να βασίζεται τις επιμέρους υπολογιστικές υλοποιήσεις σε γονιδιακούς δείκτες, αλλά σε signature-based μεθόδους, παρέχοντας μικρότερες τιμές FDR – False Discovery Rate [68].

Πιο συγκεκριμένα, ο αλγόριθμος παρακάμπτει τα ψευδώς θετικά αποτελέσματα καθώς εκτελεί αναζήτηση των γονιδιακών αλληλουχιών που ανήκουν σε γονιδιακές περιοχές που προκαλούν μεγάλα ποσοστά FDR και στη συνέχεια τις εξαιρεί. Εξάγει τις πληροφορίες των γονιδιακών περιοχών μέσω

¹⁷ <https://github.com/bioinformatics-centre/kaiju>

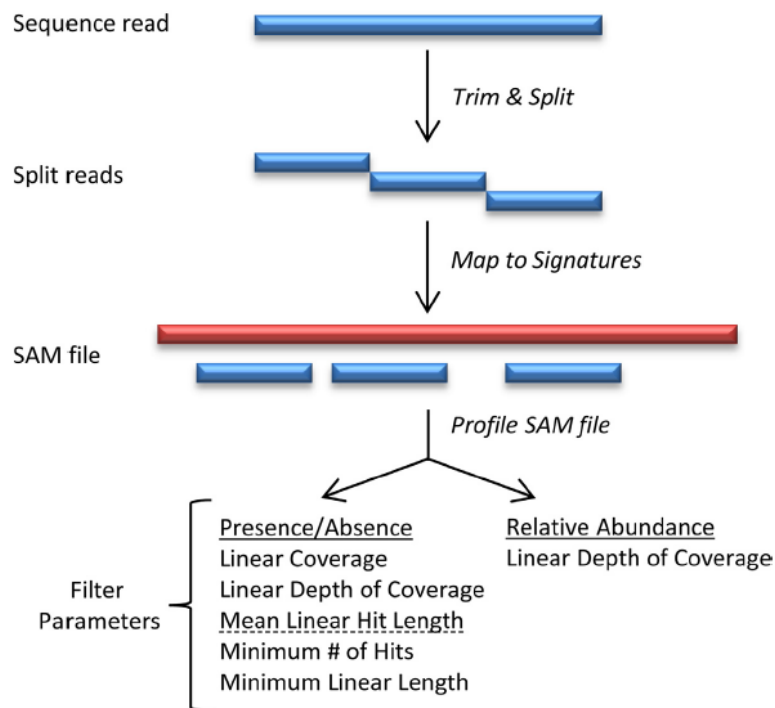
ανάλυσης κατανομή και depth of coverage μέσω δημιουργίας ενός κλάσματος για κάθε γονιδίωμα αναφοράς. Με αυτόν τον τρόπο προσδιορίζει την πληθυσμιακή σύνθεση της κοινότητας και την σχετική αφθονία των πληθυσμών που την αποτελούν.

Χρησιμοποιεί εμπειρικούς αλγορίθμους, οι οποίοι υποστηρίζονται από προσεγγίσεις μηχανικής μάθησης, με στόχο να θέσει τα όρια των ανιχνεύσεων που εκτελεί. Το αποτέλεσμα της εκτέλεσης είναι μια αναβαθμισμένη δημιουργία ενός Community Profiling. Το αρχικό βήμα του αλγορίθμου είναι το φιλτράρισμα των reads και η θραυσματοποίηση τους σε μήκος ενιαίας μορφής. Τα reads με μεγαλύτερο μήκος από 30bp διαιρούνται (trimmed subreads) με στόχο τη μείωση του ποσοστού λάθους κατά την υλοποίηση. Η σύγκριση των τιμών υπολογίζεται μέσω trail-error με διάφορες τιμές μήκους των θραυσμάτων και χρησιμοποιείται η πιο αποδοτική. Τα trimmed subreads που έχουν δημιουργηθεί στοιχίζονται με βάση δεδομένων που έχει δημιουργήσει το εργαλείο με:

```
BWA: mem -k 24 -T 0 -B 100 -O 100 -E 100 -t 12
```

-k= το ελάχιστο μήκος αλληλουχίας, -T το ελάχιστο σκορ στοίχισης, -B το penalty του mismatch, -O το gap σχετικά με το penalty, -E το gap σχετικά με την επέκταση της στοίχισης και τέλος το -t ο αριθμός των threads. Στη συνέχεια ακολουθεί εικόνα με τη ροή εργασίας του αλγορίθμου.

Εικόνα 16: Ροή αλγορίθμου με τα τρία βήματα που εκτελεί [68]



Η γλώσσα προγραμματισμού που είναι υλοποιημένο είναι Python, με κώδικα ανοιχτού λογισμικού¹⁸ ¹⁹ και λειτουργικό σύστημα σε Linux ή MacOS με ελάχιστη μνήμη RAM 8GB. Επίσης απαιτείται C/C++ για την εγκατάσταση των επεκτάσεων.

2.3.2.4.3. GTDB-Tk v1.1.0

Το GTDK-Tk είναι λογισμικό που περιέχει εργαλεία για την ταξινομική ανάλυση σε πληθυσμούς βακτηρίων και αρχαίων [69]. Η βάση δεδομένων που έχει δημιουργήσει λέγεται Genome Database Taxonomy GTDB [70] στην οποία γίνεται ενημέρωση κάθε φορά που γίνεται ενημέρωση στην NCBI Assembly database. Έχει σχεδιαστεί για να αναλύει εκατοντάδες χιλιάδες συναρμολογημένα γονιδιώματα μεταγονιδιωματικής (Metagenome Assembled Genomes – MAGs) από περιβαλλοντικά δείγματα.

¹⁸ <https://github.com/LANL-Bioinformatics/GOTTCHA>

¹⁹ <https://github.com/seninp/GOTTCHA2>

Η ροή που υλοποιεί ο αλγόριθμος βασίζεται σε συνδυασμό της δημιουργίας ενός φυλογενετικού δένδρου μέσω Hidden Markov Model ER, τον υπολογισμό της σχετικής εξελικτικής απόκλισης στους κόμβους των φυλογενετικών δένδρων (Relative Evolutionary Divergence - RED) και μιας μεταβλητής που υπολογίζεται από την απόσταση του αρχικού ερωτήματος και του αντιπροσωπευτικού γονιδιώματος που υπολογίζει ANI (Average Nucleotide Identity) [71]. Όταν υλοποιηθούν οι υπολογισμοί μέσω αυτών των τριών βημάτων τότε η αλληλουχία ερωτήματος ταξινομείται σε μια γνωστή ταξινομική ομάδα μικροοργανισμών ή χαρακτηρίζεται ως novel είδος μέσα σε ένα συγκεκριμένο γένος. Στην εικόνα 18 αποτυπώνεται η ροή του αλγορίθμου.

Κάποιοι από τους επιμέρους αλγορίθμους που χρησιμοποιεί το λογισμικό είναι ο αλγόριθμος Prodigal - PROkaryotic Dynamic programming Gene-finding Algorithm για πρόβλεψη γονιδίων [72], ο Pplacer που υλοποιεί γραμμική συνάρτηση μέγιστης πιθανοφάνειας και Μπαγιεσιανή στατιστική για τοποθέτηση των δεδομένων σε σταθερά φυλογενετικά δένδρα αναφοράς [73], τον αλγόριθμο FastTree2 που δημιουργεί δένδρα μέσω του μέγιστου βαθμού πιθανοφάνειας για μεγάλου μήκους στοιχίσεις [74] και τέλος ο αλγόριθμος Mash που είναι η επέκταση του αλγορίθμου MinHash, ο οποίος εισάγει στους υπολογισμούς μια μεταβλητή απόστασης για τον υπολογισμό του ποσοστού μεταλλάξεων δύο ακολουθιών συνδυαστικά με τη τιμή του P-value. Στόχος είναι η αποτελεσματική υλοποίηση ομαδοποίησης και αναζήτηση μαζικών συνόλων από βάσεις δεδομένων [75].

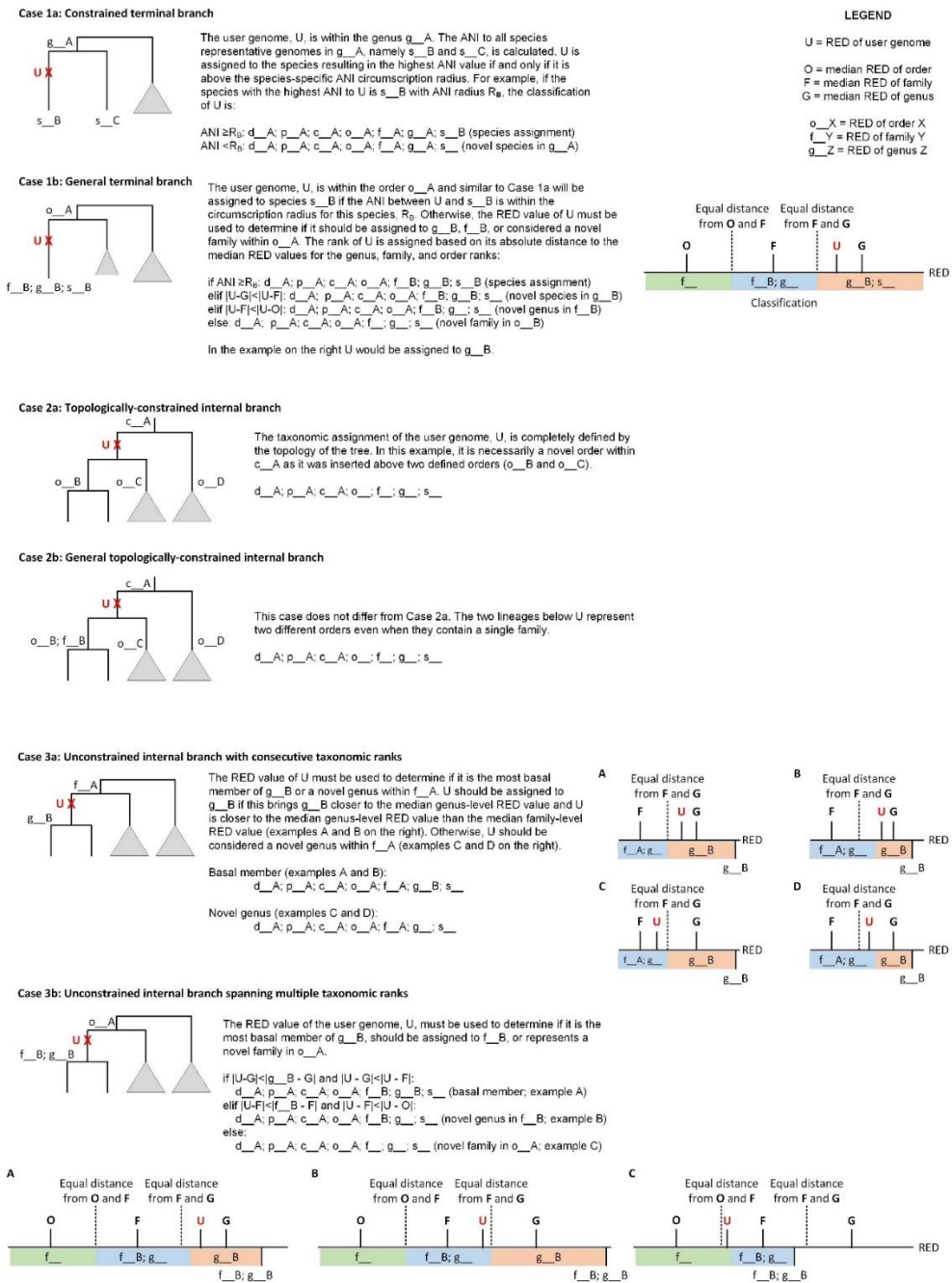
Το λογισμικό είναι ελεύθερης πρόσβασης, σε γλώσσα προγραμματισμού Python και απαιτεί 100GB μνήμη RAM και 27GB συμπληρωματική μνήμη ως αποθηκευτικό χώρο²⁰. Λόγω των αυξημένων απαιτήσεων υπάρχει η δυνατότητα χρήσης του μέσω της πλατφόρμας KBase²¹ και μέσω της σελίδας του λογισμικού²².

²⁰ <https://github.com/ECogenomics/GTDBTk>

²¹ https://github.com/kbaseapps/kb_gtdbtk

²² <https://gtdb.ecogenomic.org/>

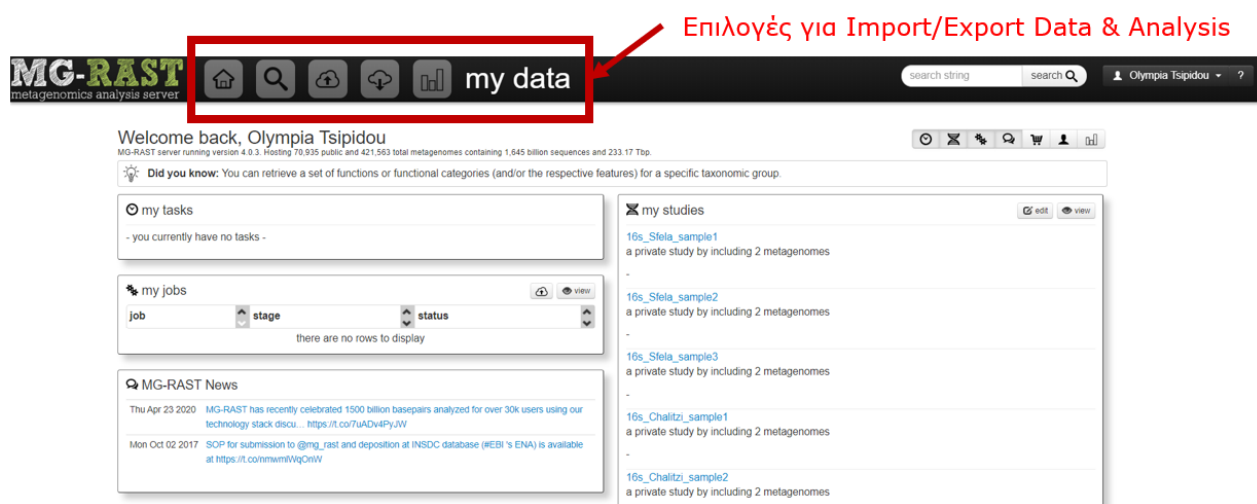
Εικόνα 17: Ροή αλγορίθμου για εκτέλεση ταξινομικής ανάλυσης πληθυσμών [69]



2.3.3. MG-Rast Platform

To MG-Rast (MetaGenomic Rapid Annotation using Subsystems Technology) δημιουργήθηκε και ξεκίνησε να λειτουργεί το 2008 ως ανοιχτή πλατφόρμα ανάλυσης μεταγονιδιωματικών δεδομένων [76], παρέχοντας ταυτόχρονα τη δυνατότητα της απόθεσης των αναλύσεων στο ευρύ κοινό. Η ανάλυση που υλοποιεί η πλατφόρμα είναι σε αυτοματοποιημένη μορφή και δεν παρέχεται στο χρήστη η δυνατότητα αλλαγής του κώδικα που εκτελεί. Παρέχεται όμως η δυνατότητα χρήσης του λογισμικού μέσω cmd [77]. Η εντολές για την υλοποίηση της ροής εργασίας βρίσκονται στη σελίδα του MG-Rast²³ και δίνεται η επιλογή της παραμετροποίησης από το χρήστη.

Εικόνα 18: Το Interface της πλατφόρμας MG-Rast

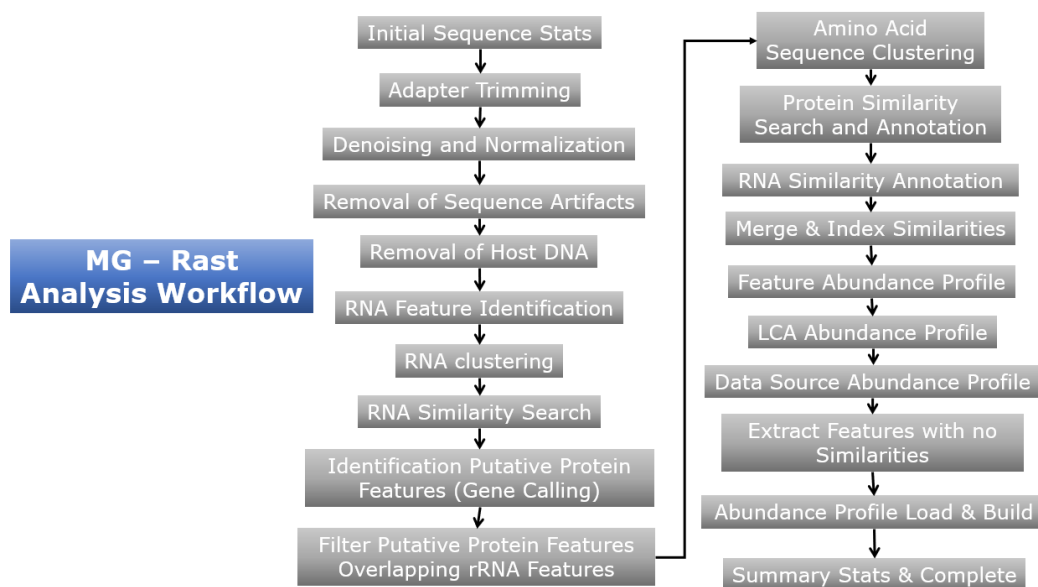


Τα αρχεία εισόδου είναι FastQ format και μετά το επιτυχές ανέβασμα στο server πραγματοποιείται αξιολόγηση της ποιότητας των ακολουθιών. Η υλοποίηση στατιστικού ελέγχου διευκολύνει τον έγκαιρο προσδιορισμό της ποιότητας των δεδομένων, βοηθάει στη μείωση των υπολογιστικών πόρων και την αποφυγή εσφαλμένων βιολογικών συμπερασμάτων. Επίσης, στην περίπτωση των paired-end αρχικών δεδομένων, ο χρήστης μπορεί να πραγματοποιήσει συνένωση των αρχείων πριν την εκτέλεση της ανάλυσης.

²³ <http://api.metagenomics.anl.gov/api.html>

Η μέση χρονική διάρκεια του pre-processing των δεδομένων της ανάλυσης της μελέτης διήρκησε 30 λεπτά/δείγμα για τα Amplicon based (16S-ITS) δεδομένα και 22-27 ώρες/δείγμα για τα Shotgun δεδομένα.

Εικόνα 19: Επιμέρους ροή εργασίας MG-Rast Ανάλυσης



Οι αναλύσεις που παρέχονται στο αρχικό στάδιο περιλαμβάνουν ποιοτικό έλεγχο των αλληλουχιών, σχολιασμό, συγκριτική ανάλυση σε δεδομένα Shotgun και Amplicon based (16S-ITS). Η ροή εργασίας περιλαμβάνει πέντε βασικά στάδια:

1. Έλεγχο και φιλτράρισμα δεδομένων. Περιοχές με χαμηλή ποιότητα και μικρό μήκος αποκόπτονται χρησιμοποιώντας το εργαλείο SolexaQA [78]. Στη συνέχεια της ροής, μέσω του εργαλείου DRISSE (Duplicate Read Inferred Sequencing Error Estimation) [79] γίνεται εκτίμηση των σφαλμάτων αλληλούχισης που δημιουργούν επαναλήψεις βάσεων. Στο τέλος του πρώτου βήματος, μέσω του εργαλείου Bowtie προσφέρεται η οπτικοποίηση των reads και η αφαίρεση αλληλουχιών που προέρχονται από ξενιστή. Το εργαλείο ως ξενιστή περιλαμβάνει δεδομένα αλληλουχίας του ανθρώπινου γονιδιώματος, της μύγας, της αγελάδας και του ποντικιού. Στην περίπτωση των δικών μας δειγμάτων, δεν δύναται η επιλογή της

απομάκρυνσης των reads που προέρχονται από τους ξενιστές, δηλαδή το γονιδίωμα των μηρυκαστικών.

2. Επόμενο βήμα της ροής είναι ο προσδιορισμός γονιδιακών αλληλουχιών στα δεδομένα. Χρησιμοποιεί το εργαλείο FragGeneScan [80] που εκτελεί αλγόριθμο μηχανικής μάθησης. Στην περίπτωση των Amplicon based δεδομένων, οι αλληλουχίες προσδιορίζονται μέσω του εργαλείου BLAT έναντι της βάσεις δεδομένων SILVA [81].
3. Ακολουθεί ο σχολιασμός - annotation προκειμένου να εντοπιστούν τυχόν λειτουργικές περιοχές. Δημιουργεί ομάδες - clusters πρωτεϊνών με 90% ταύτιση χρησιμοποιώντας την εφαρμογή UCLUST²⁴ από την πλατφόρμα του QIIME²⁵. Η μεγαλύτερου μήκους αλληλουχία για κάθε ομάδα επιλέγεται για τη διαδικασία της στοίχισης. Η στοίχιση υλοποιείται μέσω του εργαλείου BLAT²⁶ [82] και η βάση δεδομένων που χρησιμοποιείται είναι η M5nr [83]. Η M5nr περιέχει μοναδικές καταχωρήσεις πρωτεϊνικών ακολουθιών που έχουν προκύψει από τις βάσεις δεδομένων GenBank²⁷, SEED [84], IMG²⁸, UniProt²⁹, KEGG³⁰ και eggNOGs³¹ για shotgun δεδομένα. Στην περίπτωση των Amplicon based δεδομένων η στοίχιση υλοποιείται πάλι μέσω του BLAT έναντι της M5rna βάσης, με δεδομένα από τις βάσεις SILVA, Greengenes³² και RDP [85].
4. Προτελευταίο βήμα είναι η δημιουργία προφίλ. Τα δεδομένα που αναλύθηκαν και παρουσιάζουν στατιστικά σημαντικά αποτελέσματα εξάγονται με τη μορφή πίνακα αφθονίας. Τα προφίλ που δημιουργούνται είναι ανάλογα με την ανάλυση που επιλέχθηκε από το χρήστη προς υλοποίηση, δηλαδή ταξινομική ανάλυση ή λειτουργική ανάλυση.
5. Στο τελευταίο βήμα πραγματοποιείται η οπτικοποίηση των αποτελεσμάτων της ανάλυσης μέσω γραφημάτων. Τα γραφήματα περιλαμβάνουν: Sequence Breakdown (QC), Predicted Features (unknown

²⁴ https://drive5.com/usearch/manual/uclust_algo.html

²⁵ <http://qiime.org/>

²⁶ <https://genome.ucsc.edu/cgi-bin/hgBlat>

²⁷ <https://www.ncbi.nlm.nih.gov/genbank/>

²⁸ <https://img.jgi.doe.gov/>

²⁹ <https://www.uniprot.org/>

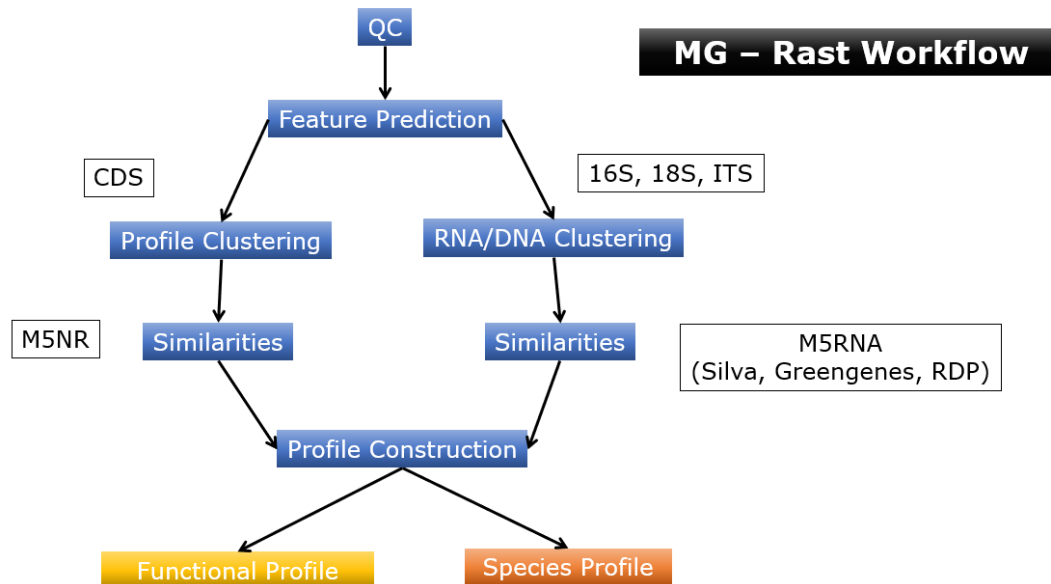
³⁰ <https://www.genome.jp/kegg/>

³¹ <http://eggnog5.embl.de/#/app/home>

³² <https://greengenes.secondgenome.com/>

or annotated proteins), Statistical Analysis, K-mer Profile, Nucleotide Histogram, Taxonomic Hit Distribution, Rank Abundance Plot, Alpha Diversity.

Εικόνα 20: Γενική ροή εργασιών κατά την εκτέλεση του MG-Rast

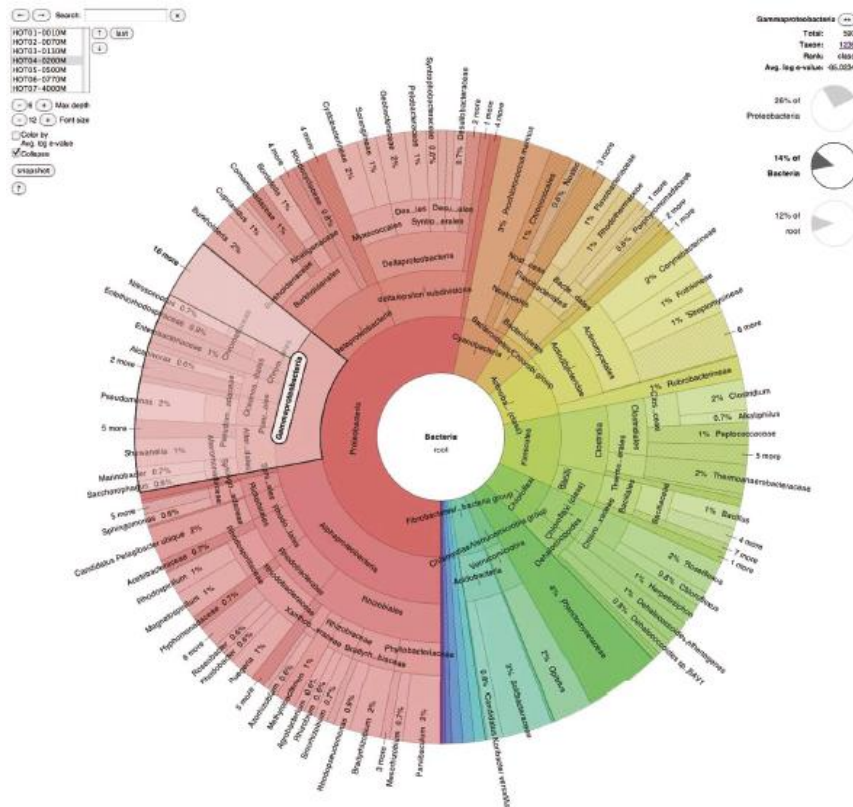


Τέλος υπάρχει η δυνατότητα δημιουργίας γραφημάτων Heatmap, Stacked bar, Barcharts κ.α. με συγκριτική ανάλυση των δειγμάτων.

Ένα από τα βασικά εργαλεία οπτικοποίησης των αποτελεσμάτων που παρέχεται από την πλατφόρμα του MG-Rast είναι το εργαλείο KRONA [86]. Το εργαλείο αυτό επιτρέπει την οπτικοποίηση των ταξινομικών αποτελεσμάτων μέσω πολυεπίπεδου γραφήματος πίτας και τη δυνατότητα zoom-in και zoom-out στα δεδομένα. Το KRONA μπορεί να χρησιμοποιηθεί και ανεξάρτητα από την πλατφόρμα του MG-Rast μέσω τοπικής εγκατάστασης στον υπολογιστή³³ σε περιβάλλον UNIX-like.

³³ <https://github.com/marbl/Krona/wiki/KronaTools>

Εικόνα 21: Οπτικοποίηση ταξινομικών αποτελεσμάτων που υλοποιεί το KRONA [87]



Οι βάσεις δεδομένων που χρησιμοποιεί το MG-Rast για την λειτουργική ανάλυση είναι η RefSeq [88], η Subsystem (SEED) [89], η KO (Kegg Orthology)³⁴, η COG [90] και η NOG [91].

Στην περίπτωση των Amplicon-based αναλύσεων, οι βάσεις δεδομένων που χρησιμοποιεί η πλατφόρμα είναι η RDP³⁵, η SILVA SSU για 16S/18S και SILVA LSU για 23S/28S³⁶ και τέλος η GreenGenes³⁷.

³⁴ <https://www.genome.jp/kegg/ko.html>
³⁵ <https://rdp.cme.msu.edu/index.jsp>
³⁶ <https://www.arb-silva.de/>
³⁷ <https://greengenes.secondgenome.com/>

2.3.4 Άλλα εργαλεία που χρησιμοποιήθηκαν

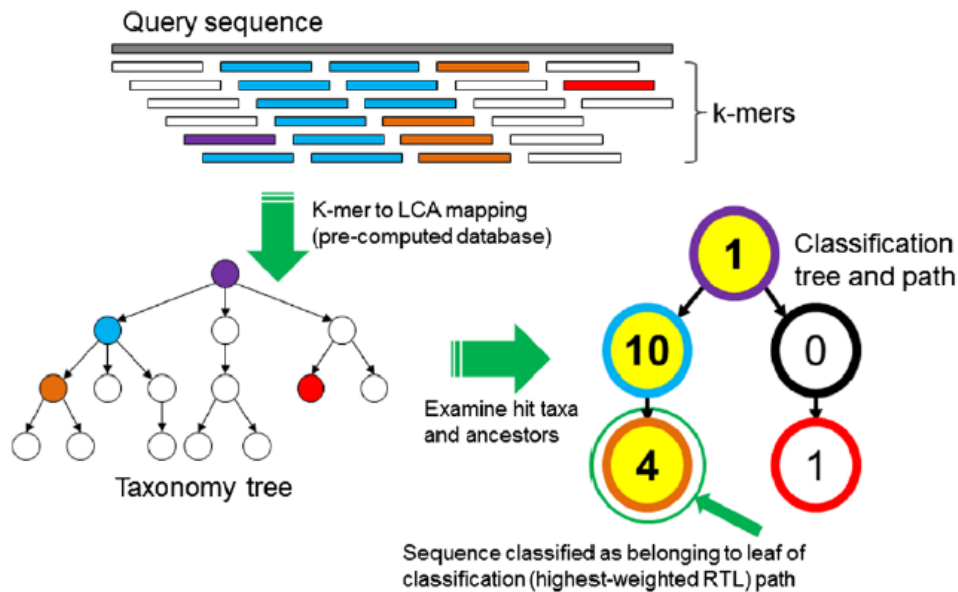
2.3.4.1. BusyBee

Το εργαλείο BusyBee³⁸ υλοποιεί ανάλυση μεταγονιδιωματικών δεδομένων με στόχο τη δημιουργία ομαδοποίησης των επιμέρους δεδομένων. Είναι εργαλείο που εκτελεί binning σε assembled αρχεία, με unsupervised τεχνική. Δηλαδή δεν χρησιμοποιεί αλληλουχίες αναφοράς για τη δημιουργία των bins που ομαδοποιούνται τα assembled δεδομένα. Ο αλγόριθμος εκτελεί στατιστικό υπολογισμό με βάση τα contigs που έχουν εισαχθεί. Δίνεται η επιλογή δύο ειδών αναλύσεων, Taxonomic και Functional analysis. Η ταξινομική ανάλυση γίνεται με τη βοήθεια του αλγορίθμου Kraken [92].

Ο αλγόριθμος Kraken εκτελεί εισαγωγή ταξινομικής ετικέτας σε μικρές αλληλουχίες DNA. Η υλοποίηση δημιουργεί μια βάση δεδομένων που προέρχεται από το NCBI και αποτελείται από k-mer συνοδευόμενο από τον κατώτατο κοινό πρόγονο (Lower Common Ancestor) όλων των οργανισμών που περιέχει το k-mer που έχει ορισθεί. Η δημιουργία αυτής της βάσης επιτρέπει την γρήγορη αναζήτηση σε ένα ταξινομικό δένδρο. Κάθε κόμβος του δένδρου είναι μια taxa και έχει μια τιμή «βάρους» (weight) ίση με τον αριθμό των k-mers που έχουν στοιχηθεί σε αυτή.

³⁸ <https://ccb-microbe.cs.uni-saarland.de/busybee>

Εικόνα 22: Αλγόριθμος Kraken [92]



Μετά το ανέβασμα των αρχείων στη σελίδα, ακολουθεί η παραμετροποίηση από τον χρήστη. Βασικοί παράμετροι είναι το ελάχιστο μήκος των contigs για να δημιουργηθεί ένα νέο bin, η τιμή του k-mer για την ανάλυση, μια τιμή για την πιθανότητα - probability η οποία κρίνει την ευαισθησία και την ακρίβεια του συστήματος που εκτελείται.

Εφόσον ορισθούν οι παράμετροι και υποβληθεί αίτηση στο server προς εκτέλεση, η αυτοματοποιημένη ροή εργασίας ξεκινάει. Σε κάθε βήμα, παρέχεται ειδοποίηση στο χρήστη για την σωστή εκτέλεσή του.

Η σελίδα αποτελεσμάτων περιλαμβάνει οπτικοποιήσεις των αποτελεσμάτων όπως δισδιάστατα γραφήματα. Το πρώτο γράφημα υποδεικνύει τα bins που δημιουργήθηκαν με διαφορετικούς χρωματισμούς για τα επιμέρους taxa, ακολουθούμενο από στατιστικά αποτελέσματα. Αυτά περιλαμβάνουν τον αριθμό των contigs που αποτελούνται τα bins, την ποιότητα των bins μέσω του βαθμού πληρότητας, επιμολύνσεων και ετερογένειας για το κάθε bin μεμονομένα. Εφόσον έχει επιλεγεί από το χρήστη η ταξινομική ανάλυση με σχολιασμούς, ακολουθούν τα αποτελέσματα μέσω γραφημάτων για το κάθε bin ξεχωριστά και τα taxa που περιλαμβάνει. Τέλος, παρέχεται η δυνατότητα εξαγωγής των αποτελεσμάτων για περαιτέρω ανάλυση.

Το εργαλείο χρησιμοποιείται μόνο μέσω απομακρυσμένου διακομιστή, χωρίς την δυνατότητα αλλαγών στον κώδικα που εκτελεί. Τα αρχεία εξόδου αποτελούνται από γραφήματα οπτικοποίησης των αποτελεσμάτων της ανάλυσης.

3. ΑΠΟΤΕΛΕΣΜΑΤΑ

3.1 Αποτελέσματα Amplicon based Αναλύσεων

3.1.1 Αποτελέσματα Δεδομένων Ανάλυσης 16S rDNA

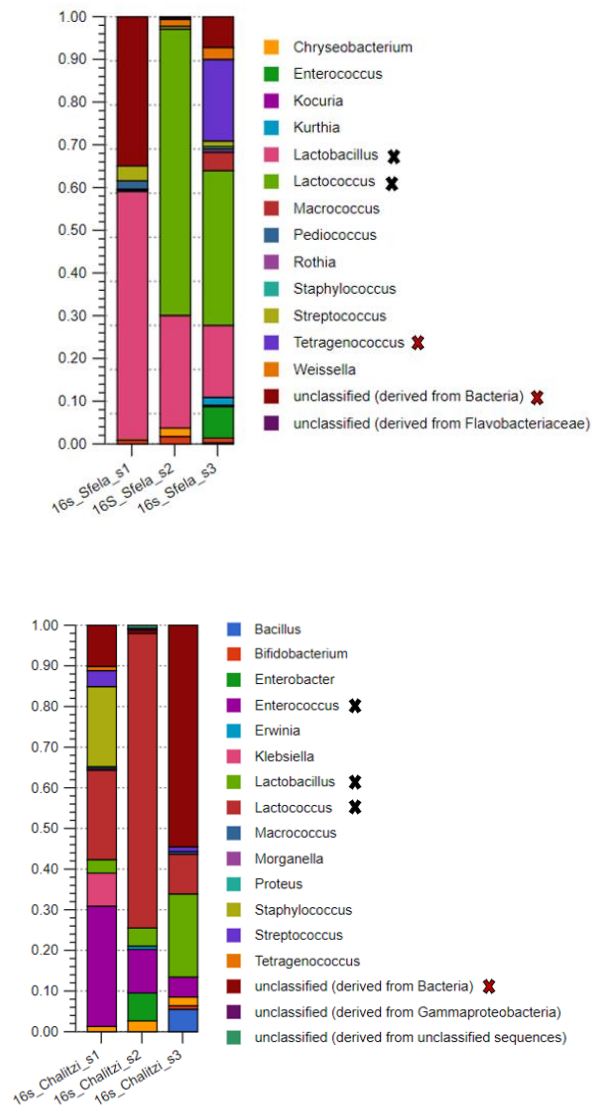
Ακολουθούν τα αποτελέσματα της ανάλυσης των δεδομένων για τις Amplicon based τεχνικές και συγκεκριμένα των περιοχών 16S rDNA. Η αλληλουχία του γονιδίου 16S rDNA χρησιμοποιείται στη μεταγονιδιωματική για την αναγνώριση των βακτηρίων σε ταξινομικό επίπεδο γένους. Οι πλατφόρμες που χρησιμοποιήθηκαν είναι η MG-Rast και η CLC.

Στο δείγμα 1 υπολογίσθηκε το γένος *Lactococcus* με ποσοστό άνω του 70%. Στο δείγμα 2 υπολογίσθηκε το γένος *Lactococcus* και στη συνέχεια το γένος *Lactobacillus*, τα οποία ανήκουν στην ομάδα LAB βακτηρίων, σε ποσοστό 67% και 26% αντίστοιχα. Στο δείγμα 3 υπολογίσθηκε διαφοροποίηση συγκριτικά με τα δείγματα 1-2 Σφέλας και αυξημένη ποικιλομορφία. Με σειρά αφθονίας βρέθηκαν τα γένη *Lactococcus* (36%), *Tetragenococcus* (20%), *Lactobacillus* (16%), *Enterococcus* (6%) και *Macrooccus* (5%).

Συνεχίζοντας για τα δείγματα Χαλίτζι, στο δείγμα 1 υπολογίσθηκαν τα γένη *Enterococcus* (30%), *Lactococcus* (22%) και *Staphylococcus* (20%). Στο δείγμα 2 υπολογίσθηκαν τα γένη *Lactococcus* (75%) και *Enterococcus* (10%). Στο δείγμα 3, στο 55% του δείγματος δεν μπόρεσε να διαπιστωθεί το γένος των βακτηρίων, με τα γένη *Lactobacillus* και *Lactococcus* να έπονται με ποσοστά 23% και 10% αντίστοιχα.

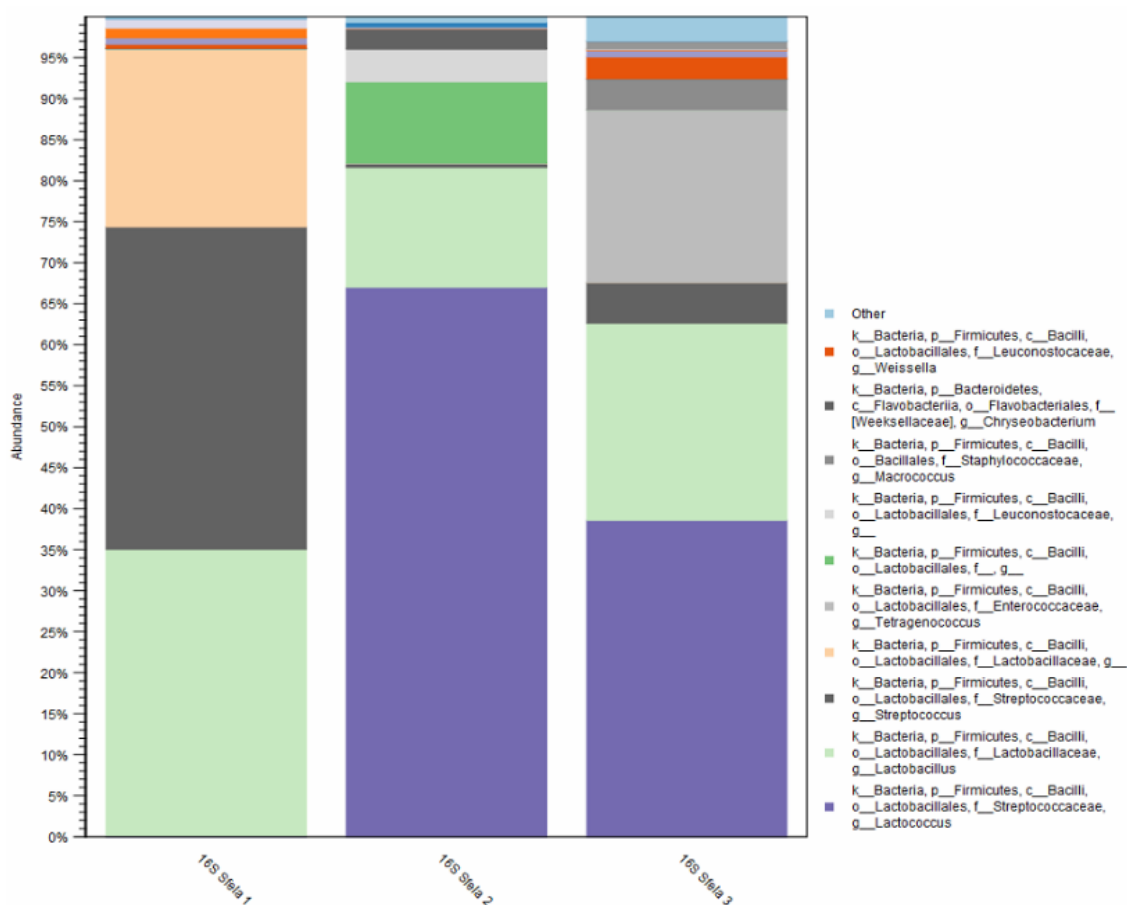
Τα αποτελέσματα παρουσιάζονται στην εικόνα 23. Μαύρος συμβολισμός βρίσκεται στα γένη που υπολογίσθηκαν και στα τρία δείγματα και κόκκινος συμβολισμός στα γένη που βρίσκονται με αυξημένο ποσοστό αλλά όχι σε όλα τα δείγματα. Για τα δείγματα Σφέλας υπολογίσθηκαν τα γένη *Lactobacillus* και *Lactococcus* στα τρία δείγματα. Και για τα τρία δείγματα Χαλίτζι με σειρά αφθονίας τα γένη *Lactococcus*, *Enterococcus* και *Lactobacillus*.

Εικόνα 23: Δείγματα Σφέλα (α) και Χαλίτζι (β). Με μαύρο X συμβολίσθηκαν γένη που υπολογίσθηκαν ότι βρίσκονται και στα 3 δείγματα διαφορετικών λευκών τυριών άλμης



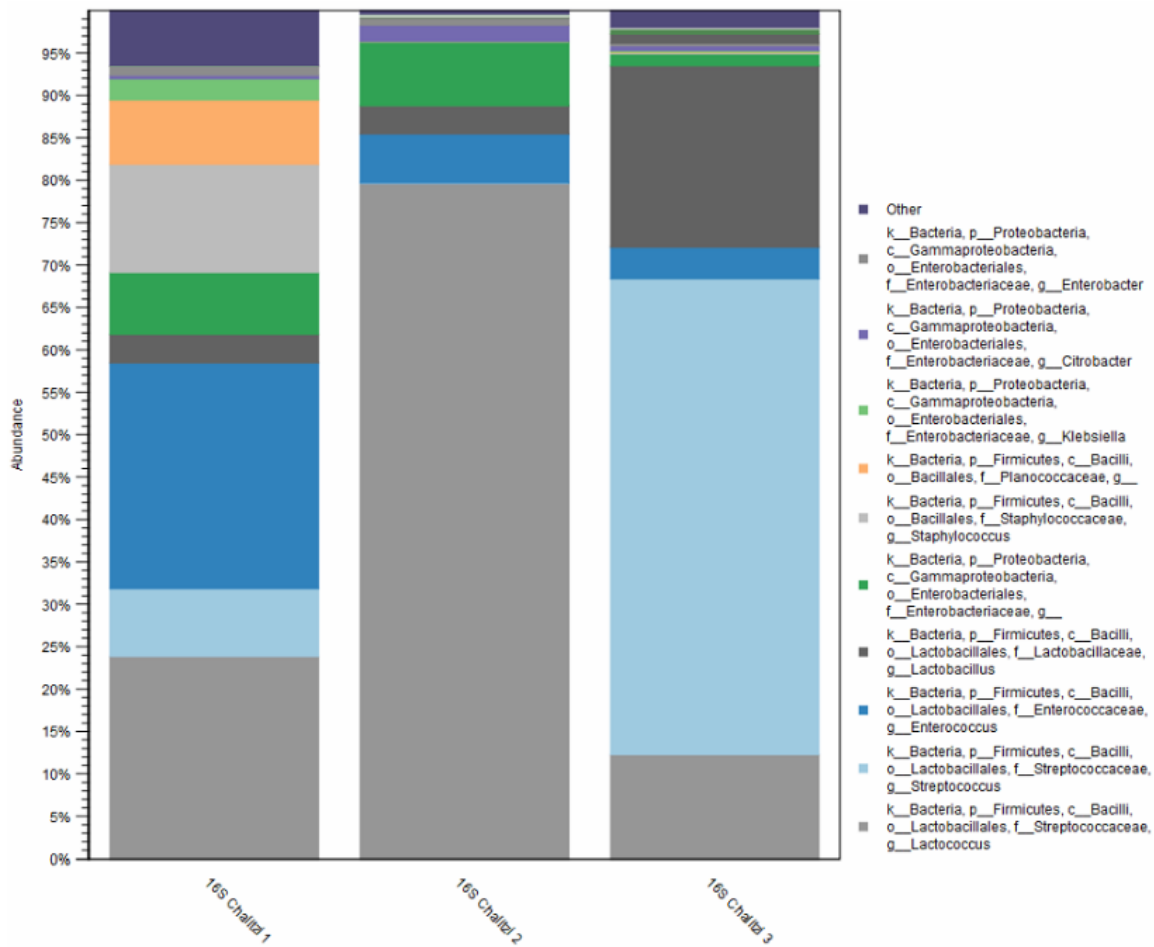
Στα αποτελέσματα από την πλατφόρμα CLC, εικόνα 24, παρατηρείται διαφοροποίηση στο δείγμα 1 Σφέλας, με 35% του δείγματος να ανήκει στο γένος *Lactobacillus* και 40% στο γένος *Streptococcus*. Σε αντίθεση με τα δείγματα 2 και 3, το 65% και 37% αντίστοιχα, υπολογίσθηκε το γένος *Lactococcus*. Το γένος *Lactobacillus* υπολογίσθηκε στα δείγματα 2 με ποσοστό 16% και στο δείγμα 3 με ποσοστό 25%. Τέλος, το δείγμα 3 Σφέλας είναι σε συμφωνία με τα αποτελέσματα του MG-Rast καθώς υπολογίσθηκε το γένος *Tetragenococcus* με ποσοστό 20%.

Εικόνα 24: Δείγματα Σφέλα 16S rDNA



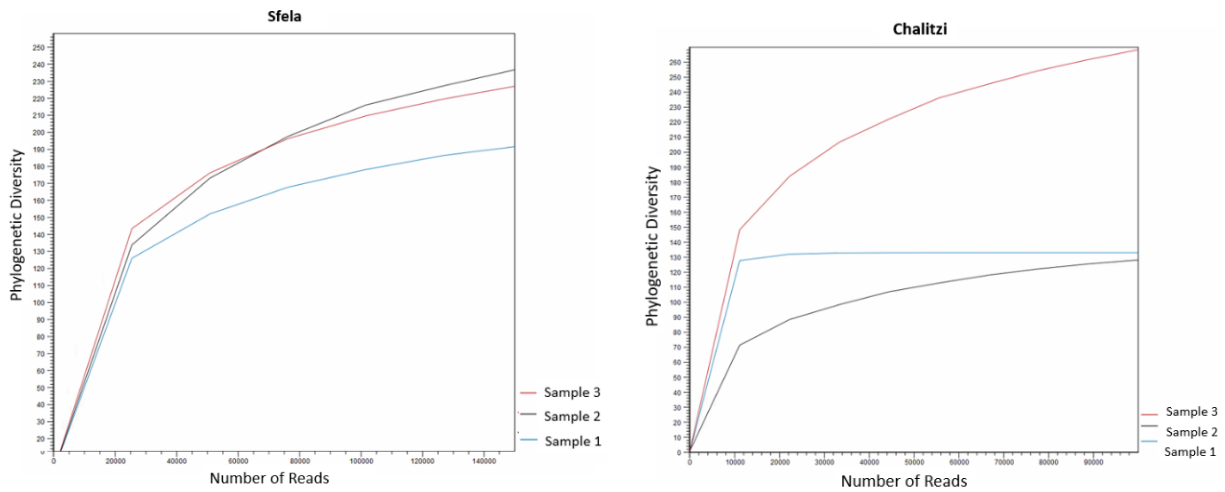
Για την ανάλυση των δεδομένων 16S rDNA μέσω της πλατφόρμας CLC στα τρία δείγματα Χαλίτζι, ακολουθεί η εικόνα 25. Τα Stacked Bar που δημιουργήθηκαν για τα τρία δείγματα έχουν κοινά αποτελέσματα, με διαφοροποιήσεις στα ποσοστά αφθονίας. Το γένος *Lactococcus* υπολογίσθηκε με ποσοστό 24%, 80% και 12% αντίστοιχα για τα τρία δείγματα. Ακολουθεί το γένος *Streptococcus* για τα δείγματα 1 και 3, με 8% και 58% αντίστοιχα. Τρίτο σε σειρά βρίσκεται το γένος *Enterococcus* για τα τρία δείγματα, με 26%, 5% και 4% αντίστοιχα. Και τέλος το γένος *Lactobacillus* με 4%, 3% και 20% ποσοστό αφθονίας.

Εικόνα 25: Δείγματα Χαλίτζι 16S rDNA



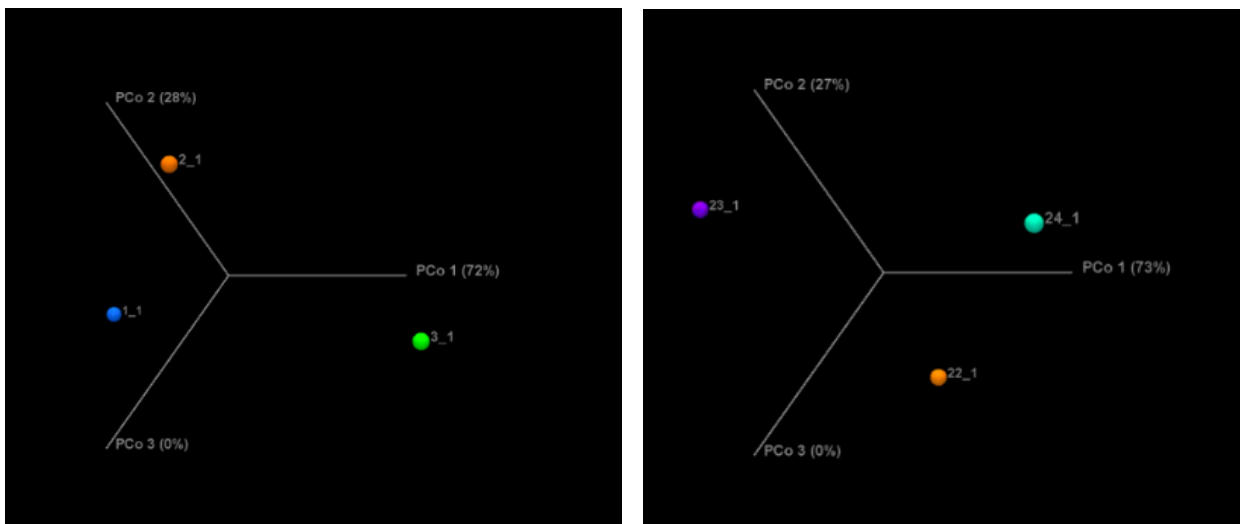
Ακολουθούν γραφικές παραστάσεις της καμπύλης Rarefaction των τυριών Σφέλα και Χαλίτζι. Τα γραφήματα οπτικοποιούν τον υπολογισμό της ποικιλότητας των ειδών, στη συγκεκριμένη περίπτωση των γένων, με βάση την κατασκευή καμπύλης σε συνάρτηση με τον αριθμό των Reads. Μέσω των καμπυλών αποδίδεται η Alpha-Diversity των δειγμάτων δηλαδή η ποικιλότητα εντός των δειγμάτων.

Εικόνα 26: Alpha-diversity δειγμάτων μέσω της καμπύλης rarefaction
α) Σφέλα β) Χαλίτζι



Τέλος, υπολογίσθηκε η Beta-Diversity των δειγμάτων, η οποία αποτυπώνει την ποικιλομορφία μεταξύ των δειγμάτων. Η γραφική παράσταση πολλαπλών διαστάσεων υλοποιείται μέσω ανάλυσης κύριων συντεταγμένων (PCoA, Principal Coordinates Analysis = MDS, Multidimensional scaling) στις σταθμισμένες αποστάσεις για τα τρία δείγματα Σφέλας και Χαλίτζι. Κάθε δείγμα είναι ένα σημείο και η απόσταση των σημείων αντιπροσωπεύει την τιμή ποικιλομορφίας μεταξύ των δειγμάτων.

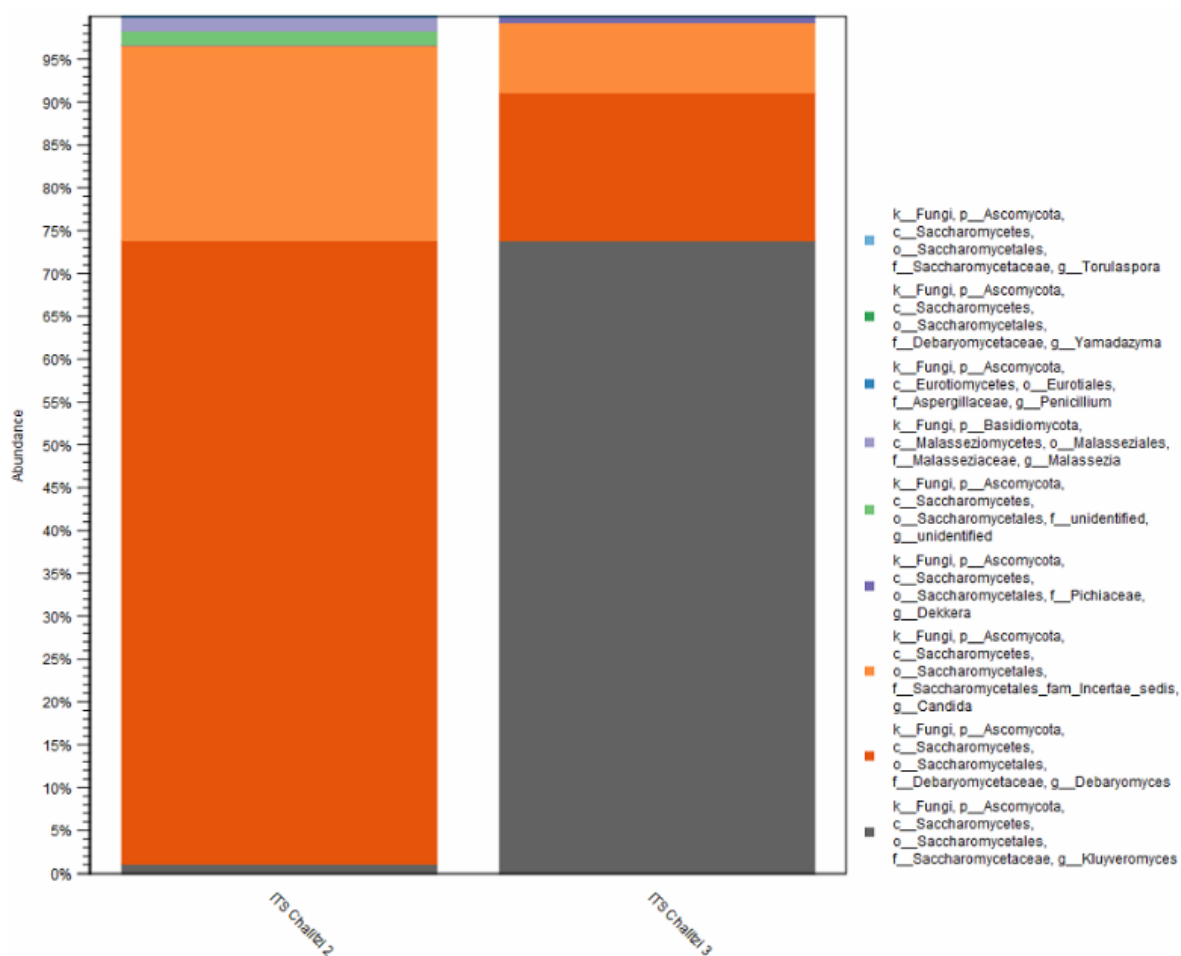
Εικόνα 28: Υπολογισμός Beta-diversity μέσω ανάλυσης PCoA α) Σφέλα
β) Χαλίτζι



3.1.2. Αποτελέσματα Δεδομένων Ανάλυσης ITS

Στη συνέχεια, πραγματοποιήθηκε ανάλυση των δεδομένων ITS (Internal Transcribed Spacer) για την εύρεση των μυκήτων και ζυμών στα δείγματα 2 και 3 Χαλίτζι. Στην εικόνα 27 αποτυπώνονται τα αποτελέσματα, τα οποία διαφέρουν σε μεγάλο βαθμό. Στο δείγμα 3 υπολογίστηκε το γένος *Kluyveromyces* με ποσοστό 74%, σε αντίθεση με το δείγμα 2, στο οποίο υπολογίστηκε με ποσοστό μόλις 1%. Στην περίπτωση του δείγματος 2 υπολογίστηκε το γένος *Debaryomyces* με 74% και στο δείγμα 3 με 16%. Επίσης, παρατηρείται και στις δύο περιπτώσεις μειωμένη ποικιλομορφία. Και τα δύο γένη που υπολογίστηκαν με αυξημένα ποσοστά ανήκουν στην οικογένεια των *Saccharomycetaceae*.

Εικόνα 27: Stacked Bars για τα γένη μυκητών που υπολογίστηκαν στα δείγματα 2 και 3 Χαλίτζι



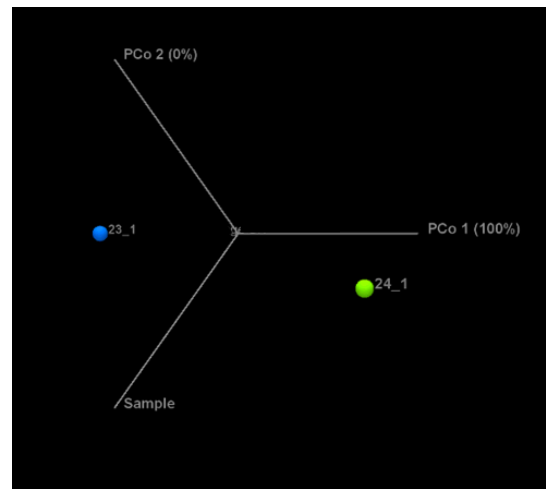
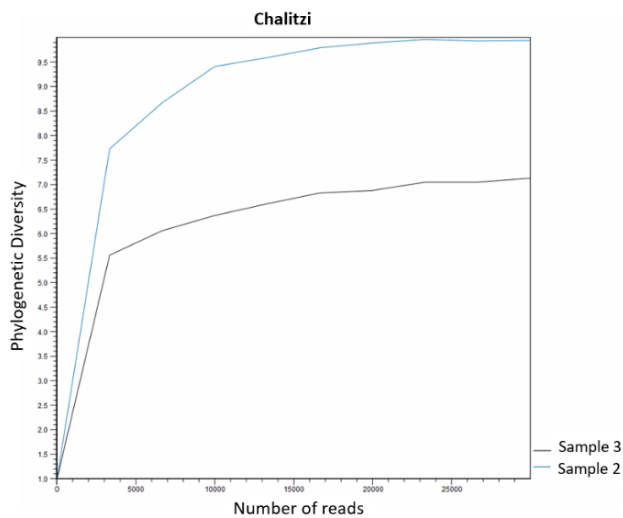
Οι ζύμες διαδραματίζουν σημαντικό ρόλο στην ποιότητα και τις οργανοληπτικές ιδιότητες των τυριών. Υπάρχουν περιπτώσεις αλλοιώσεων στα τυροκομικά προϊόντα όταν οι πληθυσμοί ζυμών αυξηθούν. Αυτό μπορεί να συμβεί λόγω ευνοϊκών συνθηκών όπως η αύξηση της υγρασία και η αλλαγή του pH.

Ζύμες του γένους *Kluyveromyces* βρίσκονται σε λευκά τυριά άλμης, όπως το είδος *Kluyveromyces lactis*, προκαλώντας τη δημιουργία μικρών οπών. Η παρουσία τους εμφανίζεται αρκετά γρήγορα, συνήθως μετά την πήξη εντός 20 έως 45 ωρών. Επίσης χρησιμοποιείται κατά την παραγωγή τυριών διαφορετικής κατηγορίας, ως δευτερογενής καλλιέργεια με στόχο την βράχυνση της περιόδου ωρίμανσης. Η χρήση του είδους γίνεται για οικονομικής σημασίας λόγους, όπως τη μείωση στα έξοδα ωρίμανσης και αποθήκευσης των τυριών. Το συγκεκριμένο είδος, όπως και άλλα είδη του ίδιου γένους, παράγει β-γαλακτοζιδαση, η οποία συνεπάγεται την υδρόλυση της λακτόζης και μετατροπής της σε αιθανόλη. Συμπληρωματικά, το είδος *Kluyveromyces marxianus* χρησιμοποιείται για την παραγωγή αιθανόλης από το τυρόγαλο [93]. Παρουσιάζει ιδιαίτερο βιοτεχνολογικό ενδιαφέρον για την παραγωγή αιθανόλης, καθώς δύναται να αναπτύσσεται σε υψηλές θερμοκρασίες χωρίς να προηγείται ψύξη.

Στην περίπτωση του γένους *Debaryomyces*, το είδος *Debaryomyces hansenii*, με συνώνυμο *Candida famata*, έχει μεγάλη ανθεκτικότητα σε συνθήκες αυξημένης αλατότητας (NaCl). Χρησιμοποιείται για την παραγωγή σκληρών τυριών και συμμετέχει στην επιφανειακή ωρίμανση. Το είδος χρησιμοποιείται για την παραγωγή λιπασών. Οι λιπάσες ανήκουν στην κατηγορία των υδρολυτικών ενζύμων και είναι υπεύθυνες για την υδρόλυση των λιπών και των ελαίων. Αναπτύσσεται σε περιβάλλοντα με υψηλή περιεκτικότητα σακχάρων. Αξιοποιείται σε καλλιέργειες εκκίνησης συνδυαστικά με γαλακτικά βακτήρια. Η λιπολυτική και πρωτεολυτική του δράση παράγει ουσίες που προσδίδουν ιδιαίτερα χαρακτηριστικά στα σκληρά κίτρινα τυριά. Σε αρκετές περιπτώσεις όμως το είδος δρα ως αλλοιογόνος παράγοντας όταν αυξηθεί η ποσότητα του πληθυσμού.

Τέλος, ακολουθεί γραφική παράσταση της καμπύλης Rarefaction για την ανάλυση της Alpha diversity των δύο δειγμάτων. Διαπιστώνεται αυξημένη ποικιλότητα για το δείγμα 2 Χαλίτζι, σε αντίθεση με το δείγμα 3. Ακολουθεί οπτικοποίηση της Beta diversity των δειγμάτων μέσω γραφικής παράστασης πολλαπλών διαστάσεων (PCoA) για τις αποστάσεις των δειγμάτων.

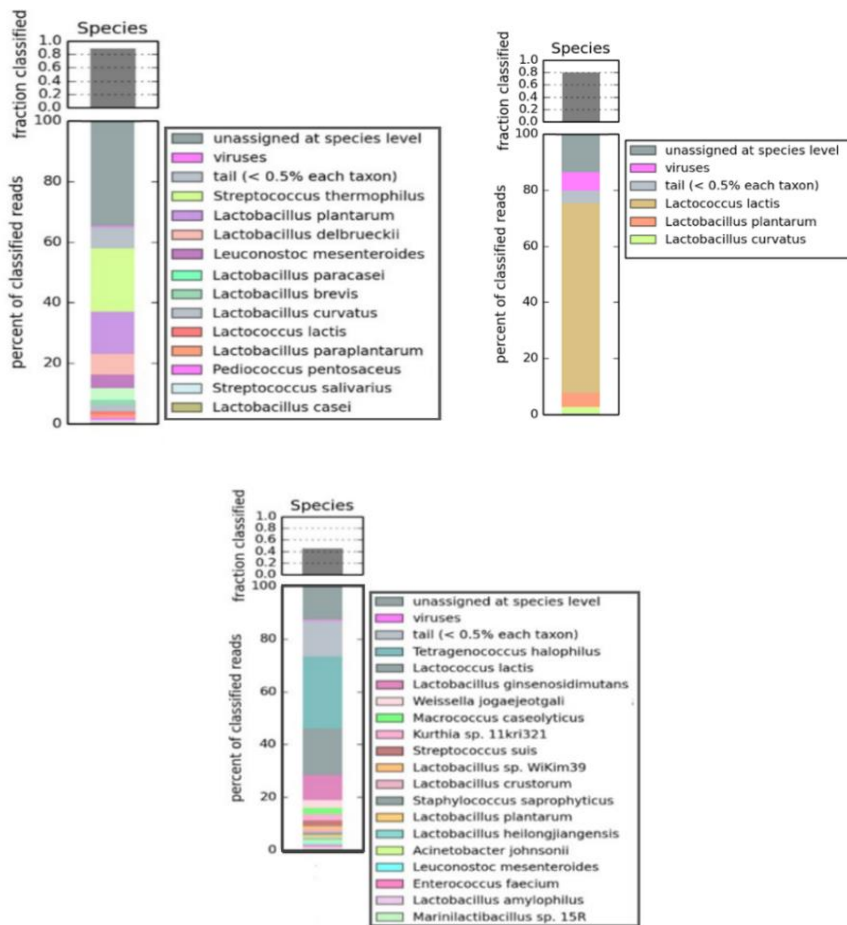
Εικόνα 28: α) Απόδοση Alpha-diversity μέσω καμπύλης rarefaction, β) Απόδοση Beta-diversity μέσω γραφικής παράστασης PCoA



3.2. Αποτελέσματα Ανάλυσης Δεδομένων Shotgun

Ακολουθεί η ανάλυση των δεδομένων της τεχνικής shotgun μέσω της χρήσης του εργαλείου KAIJU. Τα αρχεία εισόδου είναι τα FastQ files των reads της αλληλούχισης. Αρχεία εξόδου περιέχουν Stacked Bars στα οποία αποτυπώνονται τα είδη που υπολογίσθηκαν στα δείγματα.

Εικόνα 29: Δείγματα Σφέλας 1 - 2 - 3

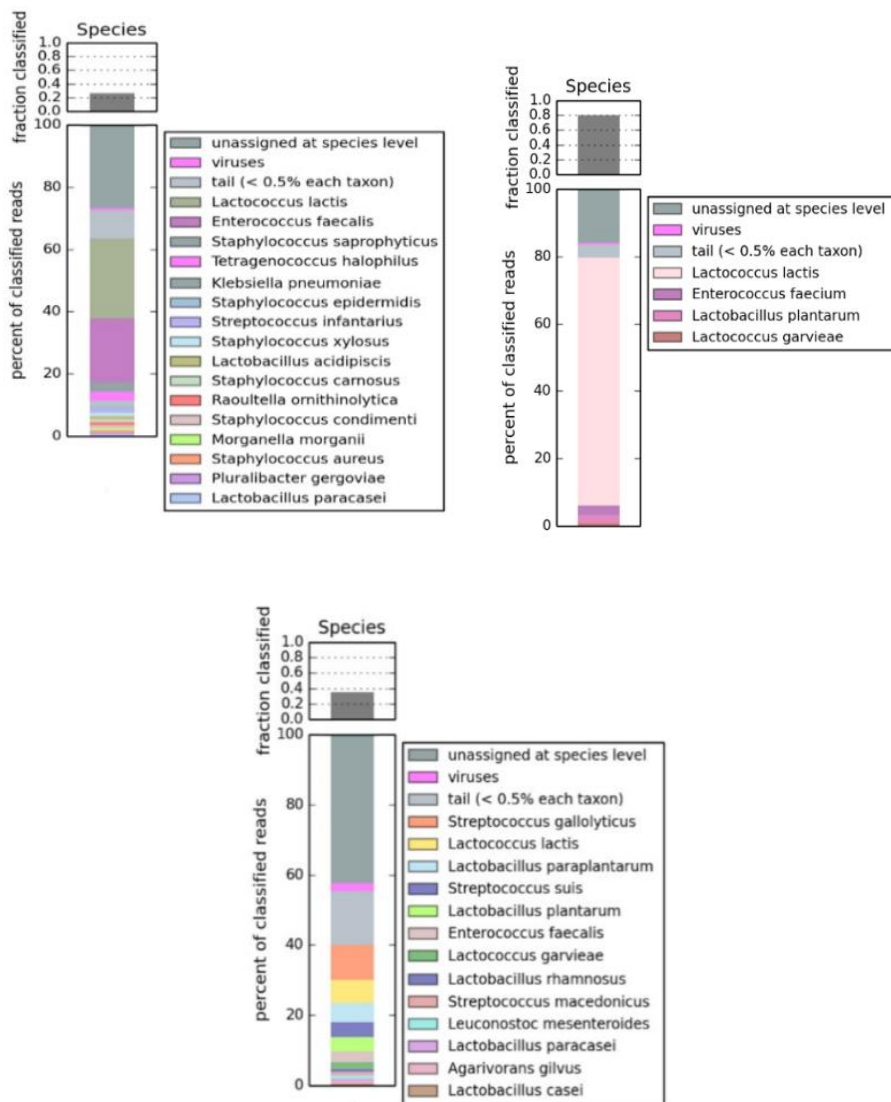


Όπως αποτυπώνεται στην εικόνα 29 με σειρά δειγμάτων, τα δύο πρώτα δείγματα Σφέλας (δείγμα 1 και 2), τα οποία προέρχονται από μικρής κλίμακας βιομηχανική παραγωγή, περιέχουν βακτήρια του γένους *Lactococcus* και *Lactobacillus*. Στο δείγμα 1 υπολογίσθηκε το γένος *Streptococcus* με ποσοστό 20%.

Σε αντίθεση με τα δύο πρώτα δείγματα, το δείγμα 3 (Σφέλα Τουλουμοτύρι) το οποίο προέρχεται από οικιακή παραγωγή, παρουσιάζει διαφοροποίηση στην ποικιλομορφία της μικροχλωρίδας του. Μεγάλο ποσοστό καταλαμβάνει το

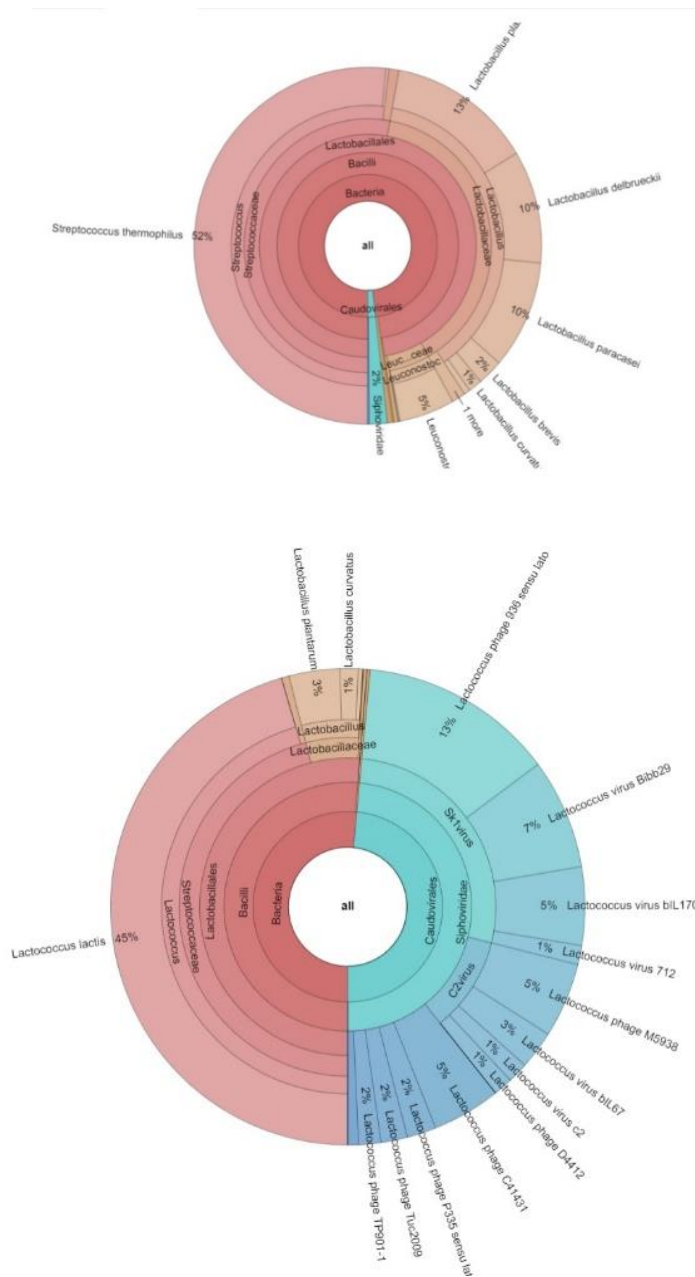
γένος *Tetragenococcus*, το οποίο ανήκει στη LAB ομάδα βακτηρίων και υπολογίσθηκε μόνο στο δείγμα 3. Το γένος ανήκει στην ομάδα βακτηρίων που παράγει βακτηριοσίνες, όπως αναφέρθηκε στην εισαγωγή, είναι ομάδα πρωτεϊνών η οποία δύναται να αναστέλλει την ανάπτυξη παθογόνων μικροοργανισμών. Επόμενο γένος που υπολογίσθηκε στο δείγμα 3 είναι *Lactococcus*, *Lactobacillus*, *Weissella*, *Macrococcus* που συναντάται σε νωπό γάλα, *Kurthia* που συναντάται συχνότερα σε μικροβίωμα του κόκκινου κρέατος, *Streptococcus* και *Staphylococcus* που χρησιμοποιούνται ως καλλιέργειες εκκίνησης στην παραγωγή τυριών.

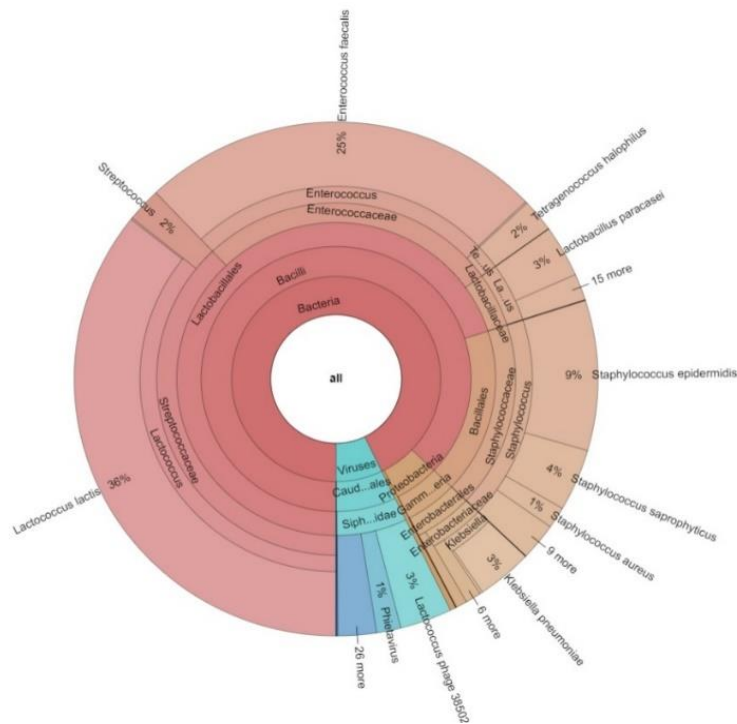
Εικόνα 30: Δείγματα Χαλίτζι 1 - 2 - 3



Συνεχίζοντας με το εργαλείο KAIJU, τα είδη βακτηρίων που κυριαρχούν στα τρία δείγματα Χαλίτζι ανήκουν στα γένη *Lactococcus*, *Streptococcus* και *Enterococcus*. Βακτήρια που ανήκουν στα τρία γένη περιέχονται στα τυροκομικά προϊόντα σε αυξημένα ποσοστά. Στο δείγμα 1 υπολογίσθηκε παρουσία παθογόνων βακτηρίων του γένους *Klebsiella*, *Raoultella* και *Morganella*, σε αντίθεση με τα δείγματα 2 και 3. Στο δείγμα 2 Χαλίτζι παρατηρείται μικρή ποικιλομορφία σε σύγκριση με τα αποτελέσματα των δειγμάτων 1 και 3, με το είδος *Lactococcus lactis* σε ποσοστό 70%.

Εικόνα 31: Δείγματα Σφέλας 1 - 2 - 3





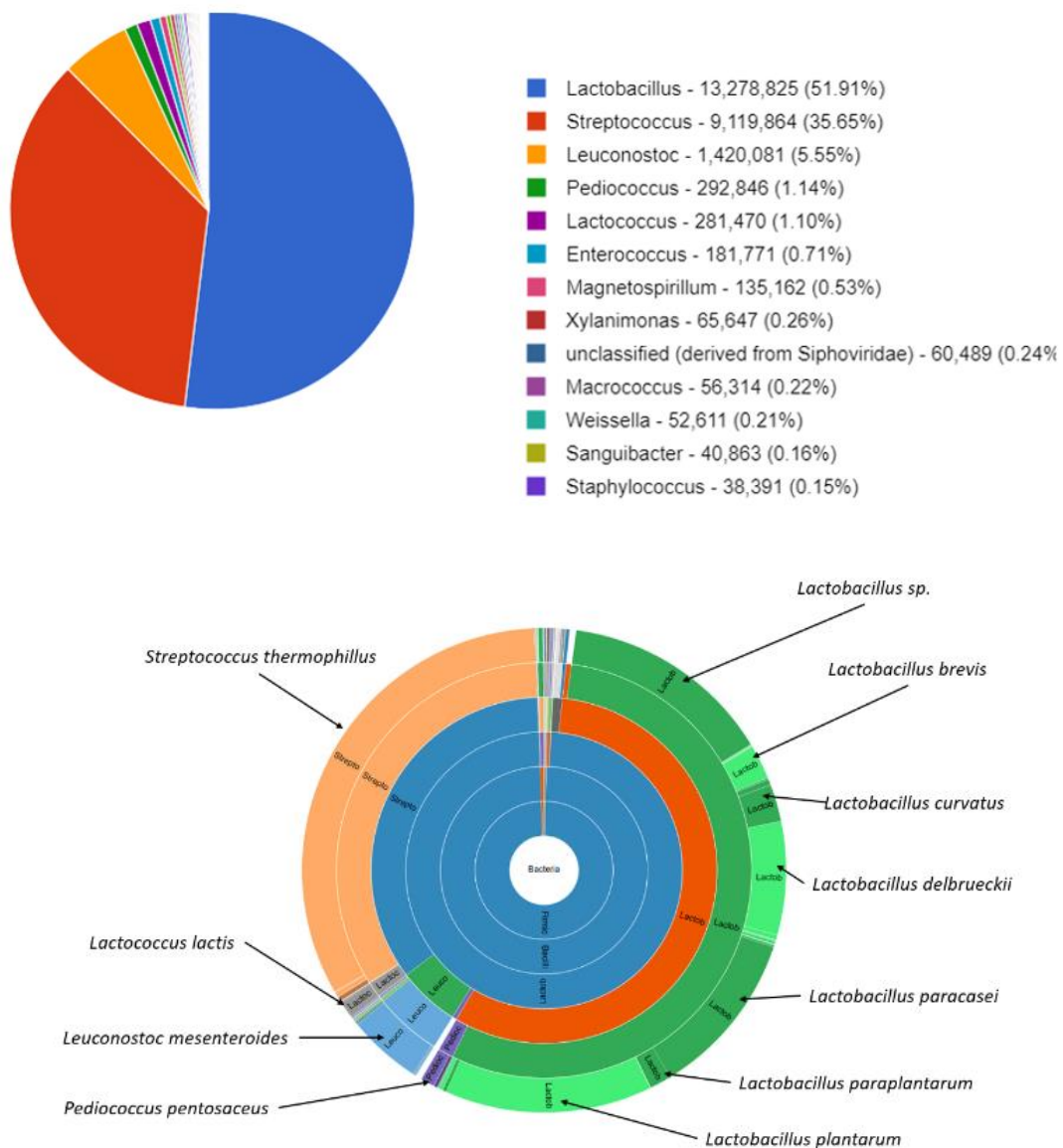
Το εργαλείο GOTTCHA (Genomic Origin Through Taxonomic CHallenge), αξιοποιώντας το διαδραστικό περιβάλλον οπτικοποίησης KRONA, επιτρέπει την εξερεύνηση ταξινομικών δεδομένων μέσω Pie Charts. Παρουσιάζεται συμφωνία με τα αποτελέσματα του εργαλείου KAIJU.

Στο δείγμα 1 Σφέλας, υπολογίσθηκε το είδος *Str. thermophilus* με 52% ποσοστό και είδη του γένους *Lactobacillus* με 36%. Στο δείγμα 2 υπολογίσθηκε το είδος *L. lactis* με ποσοστό 45% και ιοί της Οικογένειας *Siphonviridae*. Στο δείγμα 3 παρουσιάζεται μεγαλύτερη ποικιλομορφία συγκριτικά με τα δείγματα 1 και 2, τα είδη που υπολογίσθηκαν ανήκουν στα γένη *Enterococcus* και *Lactococcus*.

παρατηρείται η ύπαρξη παθογόνων βακτηρίων, κυρίως του γένους *Staphylococcus* και *Klebsiella*. Στο δείγμα 2 με ποσοστό 82% βρίσκεται το είδος *L. lactis*, σε πλήρη συμφωνία με τα αποτελέσματα του εργαλείου KAIJU. Στο δείγμα 3 υπολογίσθηκαν τα γένη *Lactococcus*, *Lactobacillus* και *Streptococcus*.

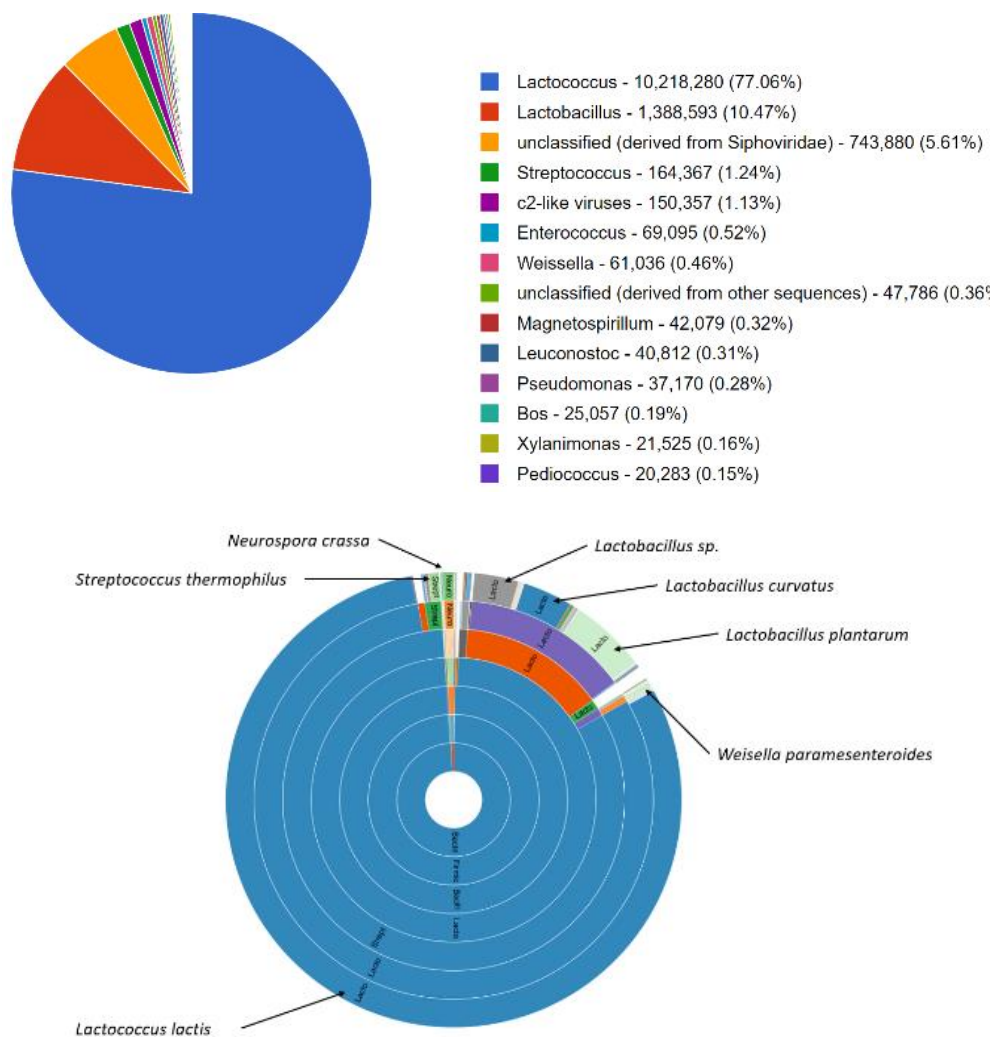
Συνεχίζοντας θα γίνει απόδοση των αποτελεσμάτων από τις πλατφόρμες MG-Rast και CLC για τα δεδομένα ανάλυσης Shotgun.

Εικόνα 33: Δείγμα 1 Σφέλας, α) MG-Rast Platform β) CLC Microbial Genomic Module



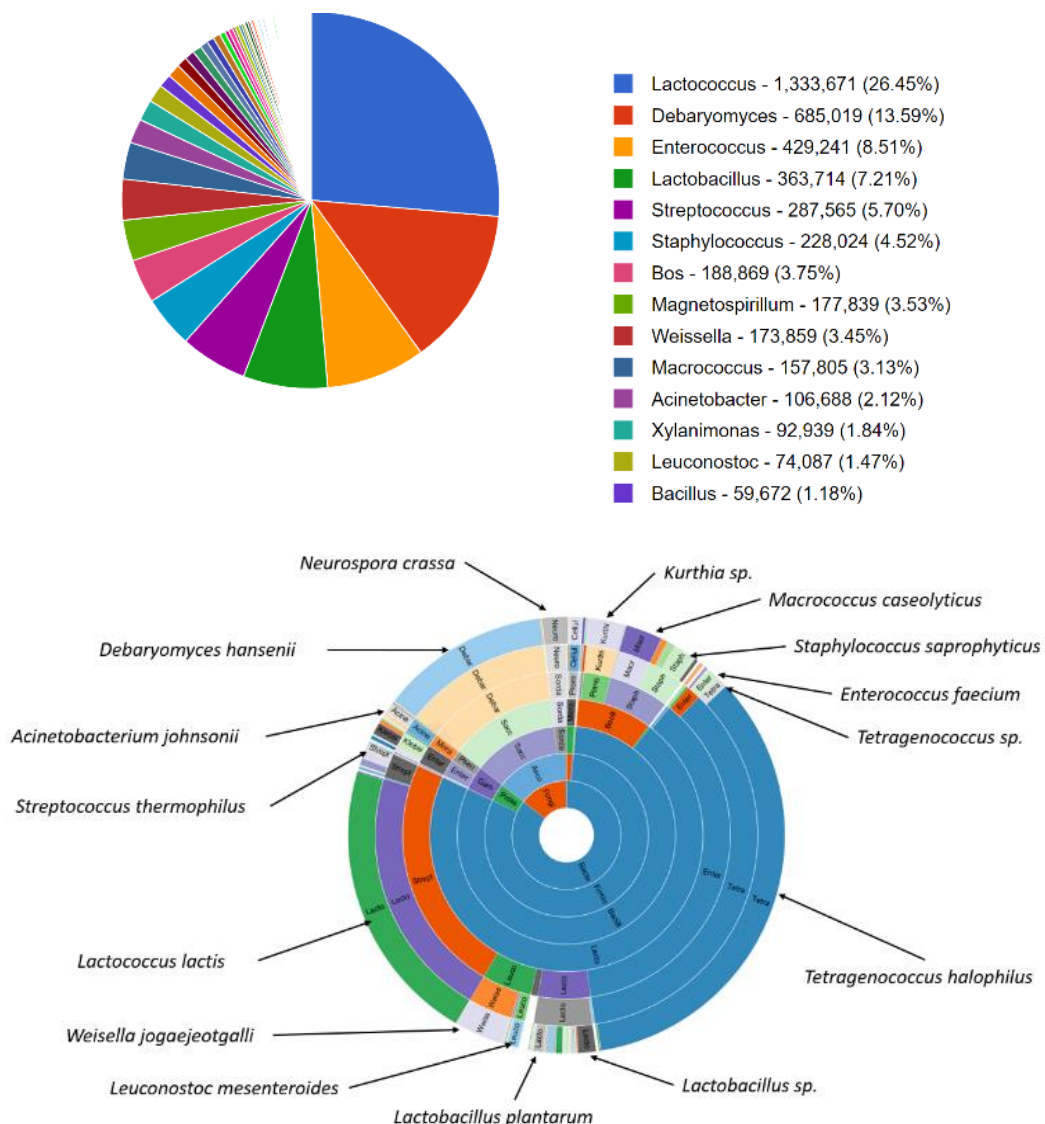
Όπως αποτυπώνεται στην εικόνα 33, στο δείγμα 1 Σφέλας υπολογίσθηκαν τα γένη *Streptococcus* και *Lactobacillus* με ποσοστό μεγαλύτερο των υπολοίπων. Τα αποτελέσματα μέσω της πλατφόρμας MG-Rast δείχνουν παρουσία των ειδών *Lb. plantarum*, *Lb. casei*, *Lb. delbrueckii* και *Lb. brevis* σε αφθονία. Τα αποτελέσματα της πλατφόρμας CLC συμπληρώνουν τα είδη *Lb. paracasei* και *Lb. curvatus*. Αυξημένο ποσοστό αφθονίας και στις δύο περιπτώσεις καταλαμβάνει το είδος *Str. thermophilus*. Τα συγκεκριμένα είδη έχουν αυξημένη παρουσία σε λευκά τυριά άλμης. Έπονται τα είδη *Leuconostoc mesenteroides* και σε μικρότερο ποσοστό *L. lactis*. Σημειώνεται ότι τα αποτελέσματα της ανάλυσης των δεδομένων shotgun στις δύο πλατφόρμες παρουσιάζουν ομοιότητα.

Εικόνα 34: Δείγμα 2 Σφέλας, α) MG-Rast Platform β) CLC Microbial Genomic Module



Τα αποτελέσματα για το δείγμα 2 Σφέλας εμφανίζουν ομοιότητες στις πλατφόρμες MG-Rast και CLC. Το είδος *L. lactis* υπολογίσθηκε 75%-85% μέσω της πλατφόρμας MG-Rast και 85%-90% μέσω της πλατφόρμας CLC. Επόμενα είδη που υπολογίσθηκαν ανήκουν στο γένος *Lactobacillus*. Προβληματισμό δημιουργεί η διαφοροποίηση των αποτελεσμάτων στα δείγματα 1 και 2.

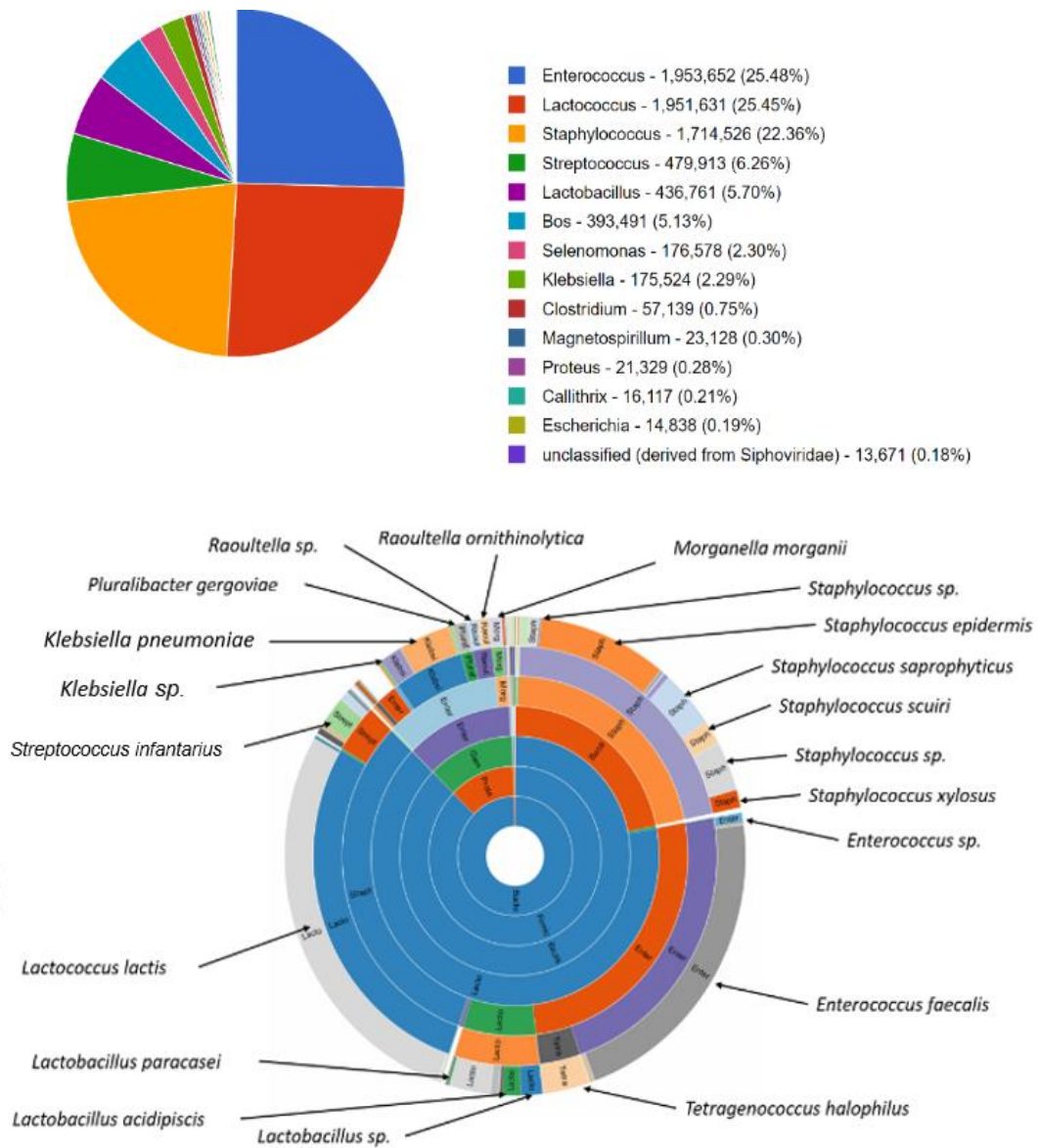
Εικόνα 35: Δείγμα 3 Σφέλας, α) MG-Rast Platform β) CLC Microbial Genomic Module



Τα αποτελέσματα του δείγματος 3 παρουσιάζουν μεγαλύτερη ποικιλότητα σε σύγκριση με τα δείγματα 1 και 2. Μέσω της πλατφόρμας MG-Rast υπολογίσθηκαν είδη του γένους *Lactococcus* με μεγάλο ποσοστό αφθονίας, σε αντίθεση με τα αποτελέσματα της πλατφόρμας CLC που υπολογίσθηκε το

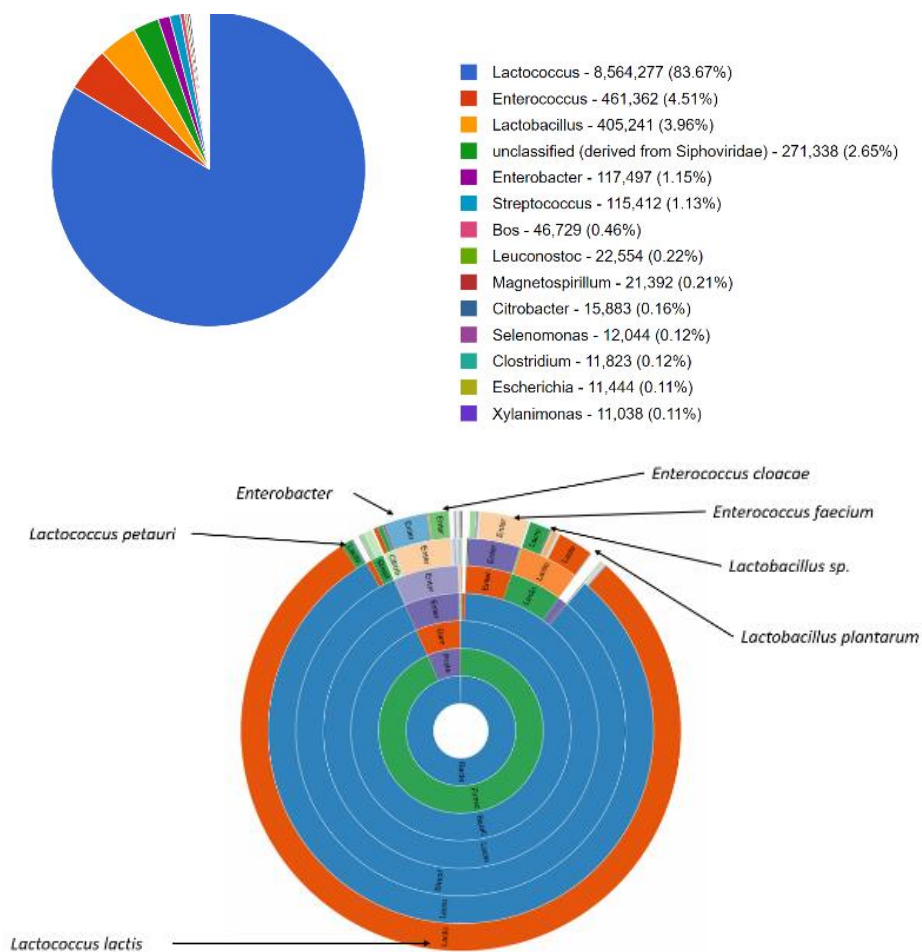
είδος *Tetragenococcus halophilus*, το οποίο ανήκει στην ομάδα LAB βακτηρίων. Επίσης, στα αποτελέσματα του MG-Rast παρατηρείται η παρουσία ζύμης του γένους *Debaryomyces* η οποία ανήκει στην οικογένεια *Saccharomycetaceae*. Στη δεύτερη θέση των αποτελεσμάτων της πλατφόρμας CLC βρίσκεται το είδος *L. lactis*. Ακολουθεί το είδος ζύμης *D. hansenii* που έρχεται σε συμφωνία με το αποτέλεσμα της πλατφόρμας MG-Rast. Και οι δύο πλατφόρμες εντοπίζουν είδη του γένους *Lactobacillus*, *Enterococcus*, *Streptococcus* και *Staphylococcus* σε μικρότερα ποσοστά.

Εικόνα 36: Δείγμα 1 Χαλίτζι, α) MG-Rast Platform β) CLC Microbial Genomic Module



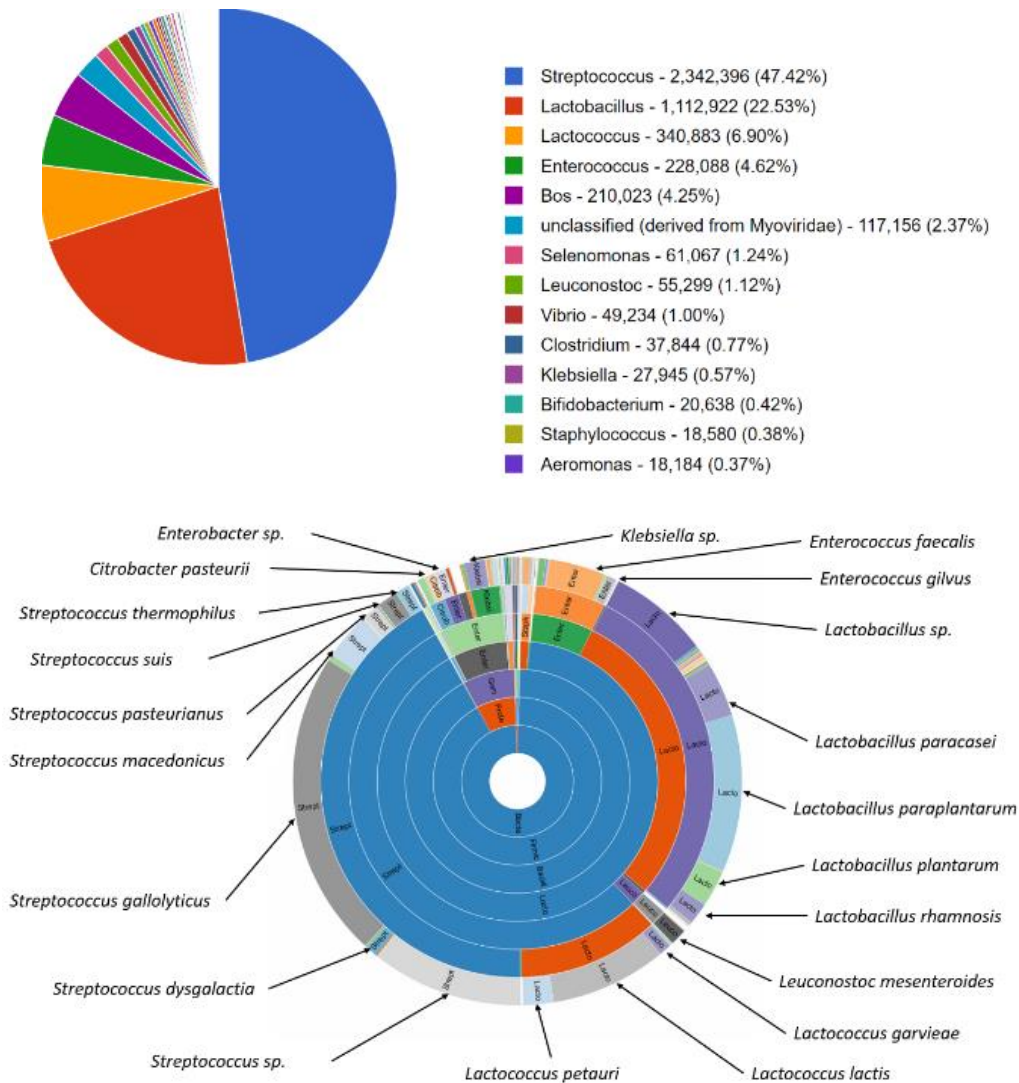
Τα αποτελέσματα της ανάλυσης του δείγματος 1 Χαλίτζι, βρίσκονται σε συμφωνία στα δύο εργαλεία. Τα είδη που υπολογίσθηκαν σε αφθονία ανήκουν στα γένη *Enterococcus*, *Lactococcus*, *Staphylococcus* και *Streptococcus*. Πιο συγκεκριμένα, υπολογίσθηκε το είδος *L. lactis* και ακολουθεί το είδος *Enterococcus faecalis*. Παρατηρείται επίσης ποσοστό παθογόνων βακτηρίων του γένους *Klebsiella* και *Morganella* με ποσοστό $\approx 45\%$. Τα επιμέρους είδη του γένους είναι, *K. pneumoniae*, *Raoultella ornithinolytica*, *M. morganii* και ακολουθούν είδη του γένους *Staphylococcus*. Αυξημένο ποσοστό του γένους *Staphylococcus*, βάσει βιβλιογραφικών αναφορών, βρίσκεται σε τυροκομικά προϊόντα σε αρχικό στάδιο ωρίμανσης του τυριού, με σταδιακή πτώση του ποσοστού.

Εικόνα 37: Δείγμα 2 Χαλίτζι, α) MG-Rast Platform β) CLC Microbial Genomic Module



Σε αντίθεση με το δείγμα 1 Χαλίτζι, το δείγμα 2 παρουσιάζει απλή σύνθεση. Το 85% καταλαμβάνει το είδος *L. lactis* και έπονται τα γένη *Lactobacillus* και *Enterococcus*. Τα αποτελέσματα είναι κοινά στα δύο εργαλεία.

Εικόνα 38: Δείγμα 3 Χαλίτζι, α) MG-Rast Platform β) CLC Microbial Genomic Module

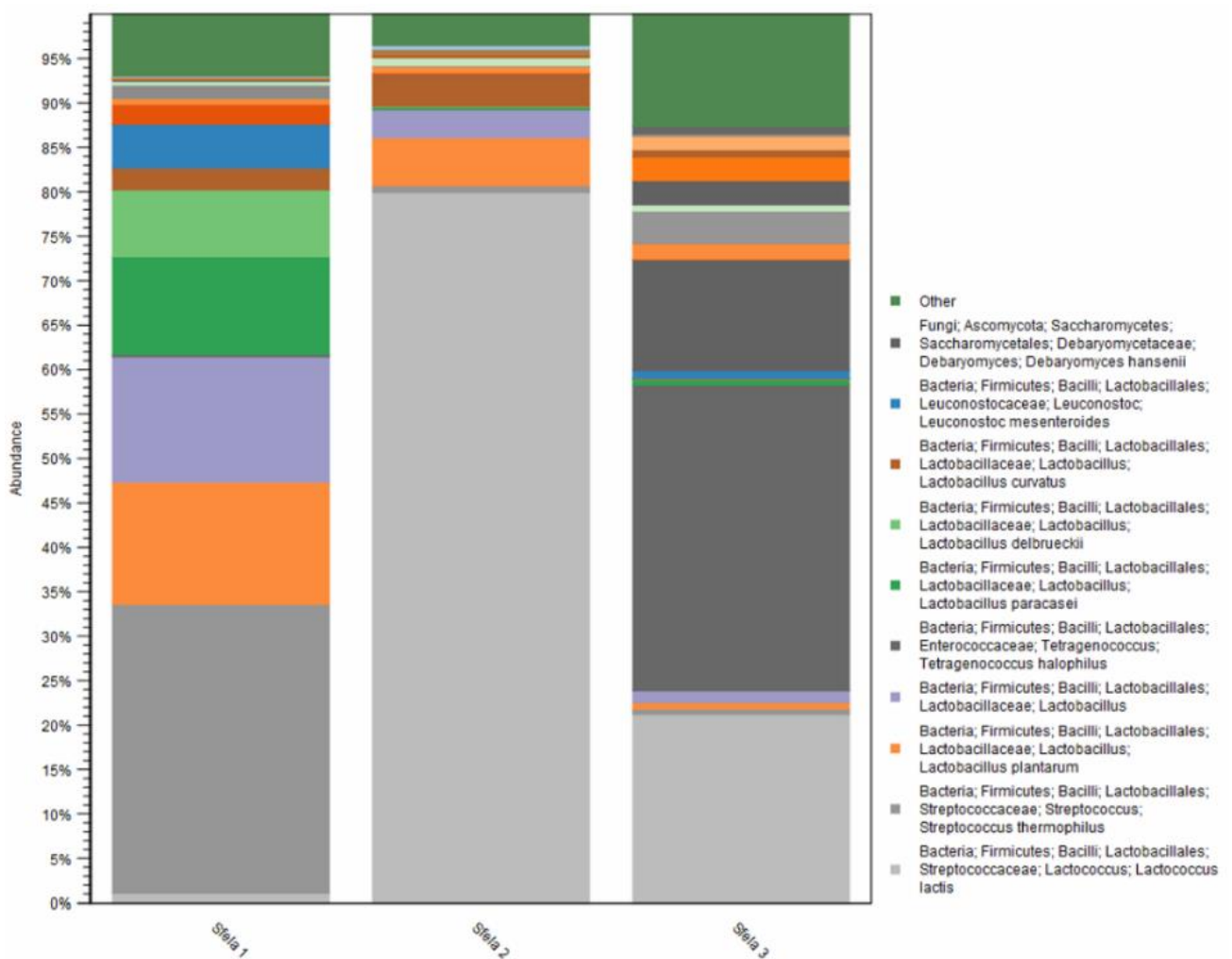


Τέλος, τα αποτελέσματα του δείγματος 3 Χαλίτζι είναι κοινά στα δύο υπολογιστικά εργαλεία. Με ποσοστό 50% υπολογίστηκαν είδη του γένους *Streptococcus*, πιο συγκεκριμένα *S. gallolyticus* 20% και *S. Macedonicus* 5%. Έπεται το γένος *Lactobacillus* με ποσοστό 20% και κυρίαρχα είδη *Lb. paraplantarum* και *Lb. paracasei*. Επόμενα γένη που υπολογίστηκαν είναι του γένους *Lactococcus*, με το είδος *L. lactis* και *Enterococcus* με το είδος *E.*

faecalis. Μικρό ποσοστό περιέχεται από παθογόνα βακτηρίων του γένους *Klebsiella* όπως στην περίπτωση του δείγματος 1 Χαλίτζι.

Στη συνέχεια ακολουθεί Stacked Bar, εικόνα 39, για τα τρία δείγματα Σφέλας και εικόνα 41, για τα τρία δείγματα Χαλίτζι. Αποτυπώνονται τα είδη που υπολογίστηκαν από την πλατφόρμα CLC και οι διαφοροποιήσεις των δειγμάτων.

Εικόνα 39: Δείγματα Σφέλας

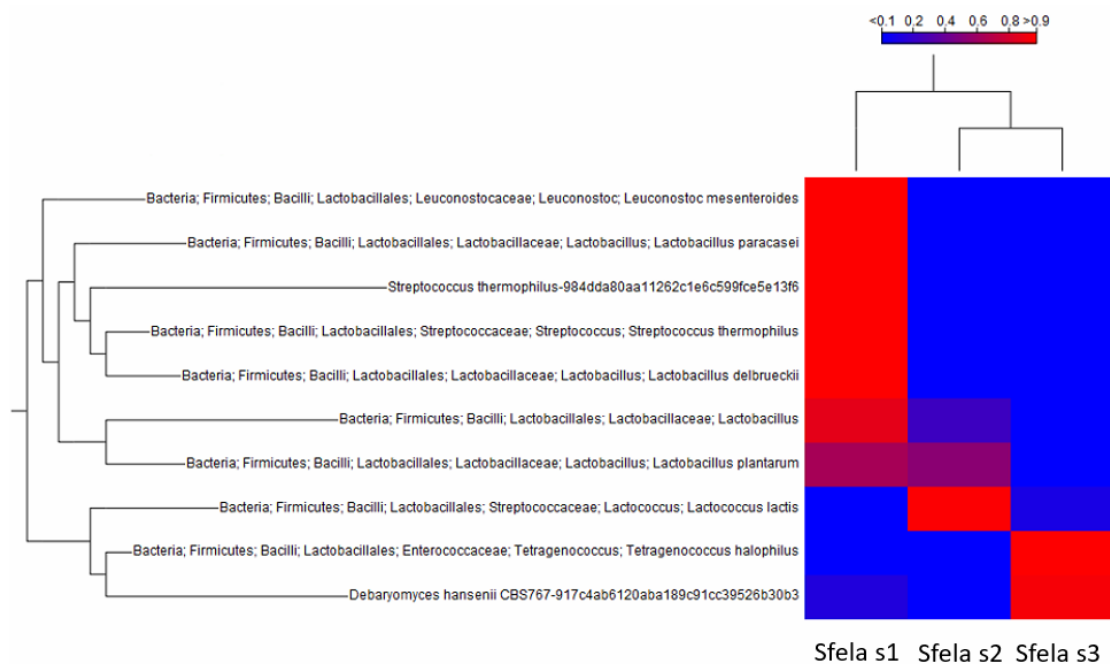


Το είδος *L. lactis* υπολογίστηκε στο δείγμα 2 και 3 Σφέλας, 80% και 20% αντίστοιχα, όπως και στο δείγμα 1 με 1%. Στο δείγμα 1 υπολογίστηκε το είδος *Str. thermophilus* με ποσοστό 35% και στα δείγματα 2 και 3 με 1-2% αντίστοιχα. Τα επόμενα είδη ανήκουν στα γένη *Lactobacillus* και *Tetragenococcus*. Με το γένος *Lactobacillus* να περιέχει τα είδη *Lb. plantarum*,

Lb. paracasei, *Lb. delbrueckii*, *Lb. curvatus* και το γένος *Tetragenococcus* το είδος *T. halophilus*.

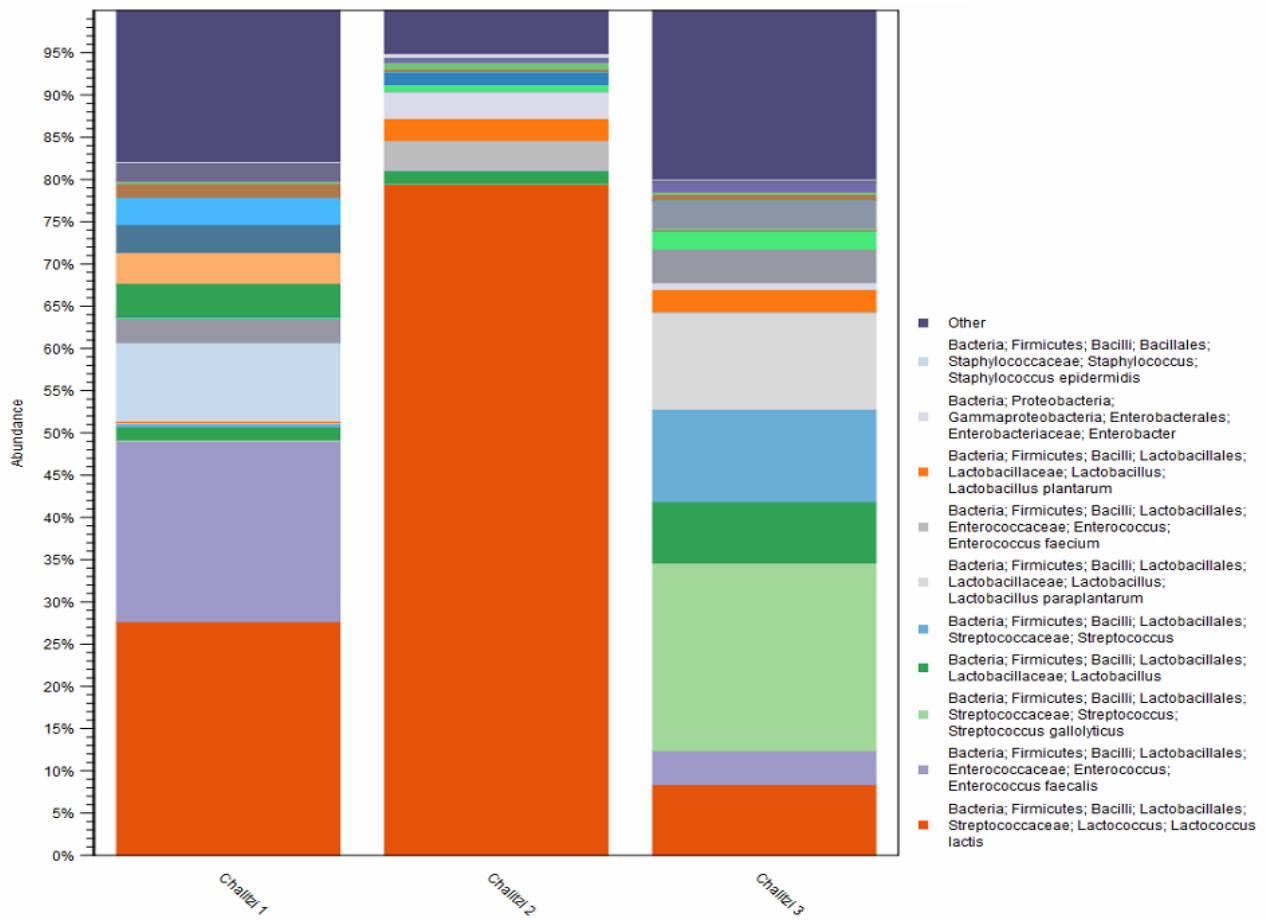
Παρατηρούμε στη συνέχεια μέσω του Heatmap, εικόνα 40, τη διαφοροποίηση που δημιουργείται μεταξύ των τριών δειγμάτων Σφέλας. Το δείγμα 1 παρουσιάζει διαφορετική σύσταση συγκριτικά με τα δείγματα 2 και 3 Σφέλας, στα οποία τα γένη που υπολογίσθηκαν είναι κοινά. Στην πρώτη περίπτωση υπολογίσθηκαν τα βακτήρια *Leuc. mesenteroides*, *Lb. paracasei*, *Str. thermophilus*, *Lb. delbrueckii* ενώ η παρουσία τους στα δείγματα 2 και 3 είναι αμελητέα. Τα είδη *L. lactis* και *Tetragenococcus halophilus* υπολογίσθηκαν με μεγαλύτερα ποσοστά στα δείγματα 2 και 3. Τέλος, παρατηρούμε την παρουσία του μύκητα *D. hansenii* στο δείγμα 3 και σε μικρότερο ποσοστό στο δείγμα 1 Σφέλας.

Εικόνα 40: Heatmap για τα τρία δείγματα Σφέλας



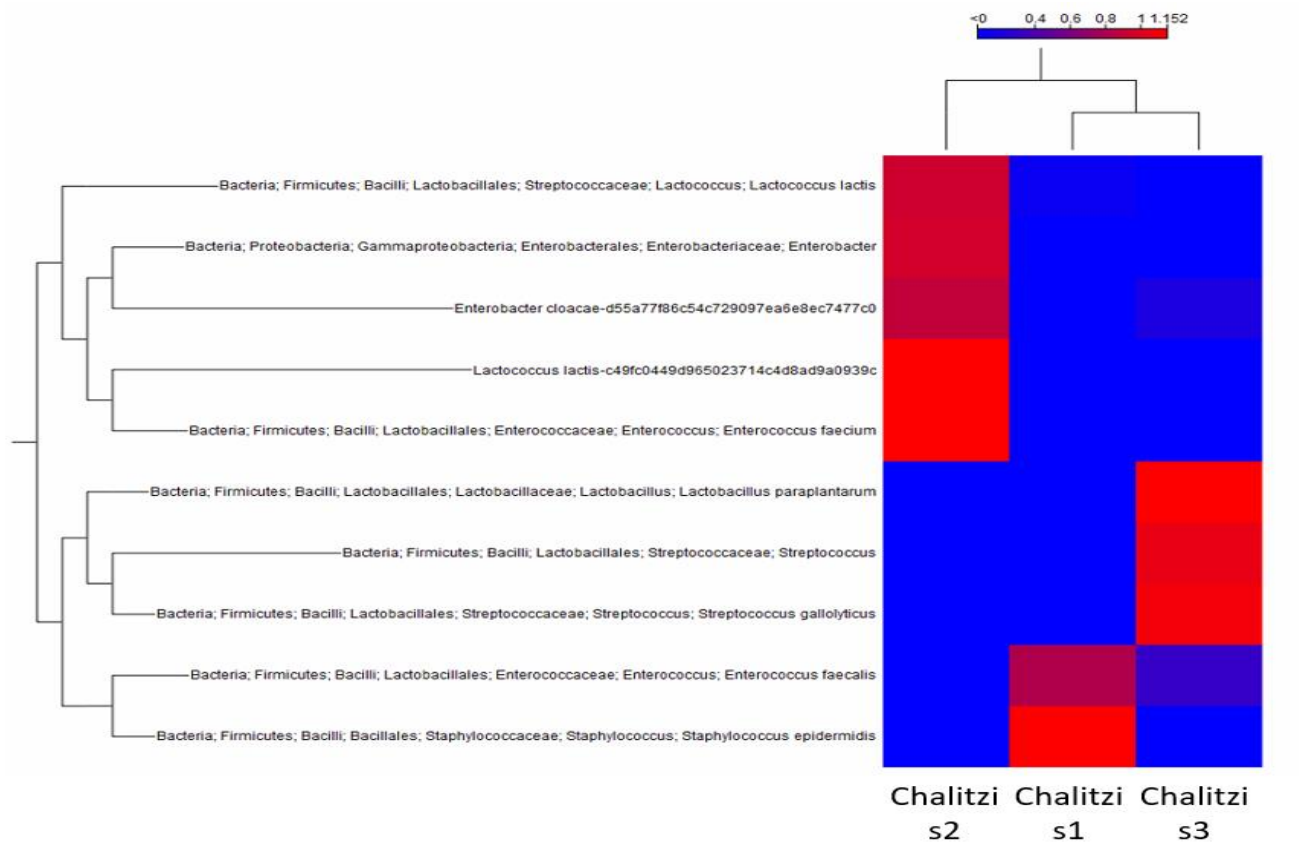
Συνεχίζοντας, παρατηρούμε στην εικόνα 41 συγκριτικά Stacked Bars για τα τρία δείγματα Χαλίτζι. Όπως αναφέραμε, το δείγμα 2 παρουσιάζει απλή μικροβιακή σύνθεση, με το είδος *L. lactis* σε ποσοστό 70%. Στα υπόλοιπα δύο δείγματα (1 και 3 Χαλίτζι) η σύνθεση του μικροβιώματος εμφανίζει αυξημένη ποικιλομορφία. Τα είδη που βρίσκονται σε αφθονία για το δείγμα 1 είναι *E. faecalis* και για το δείγμα 3 *Streptococcus gallolyticus*.

Εικόνα 41: Δείγματα Χαλίτζι



Μέσω του Heatmap, εικόνα 42, φαίνονται οι διαφοροποιήσεις που υπολογίσθηκαν για τα τρία δείγματα Χαλίτζι. Στο δείγμα 1 τα είδη *L. lactis* και *E. faecium*, στο δείγμα 2 το είδος *L. lactis* και στο δείγμα 3 τα είδη *Str. gallolyticus* και *Lb. paraplantarum*.

Εικόνα 42: Heatmap για τα τρία δείγματα Χαλίτζι

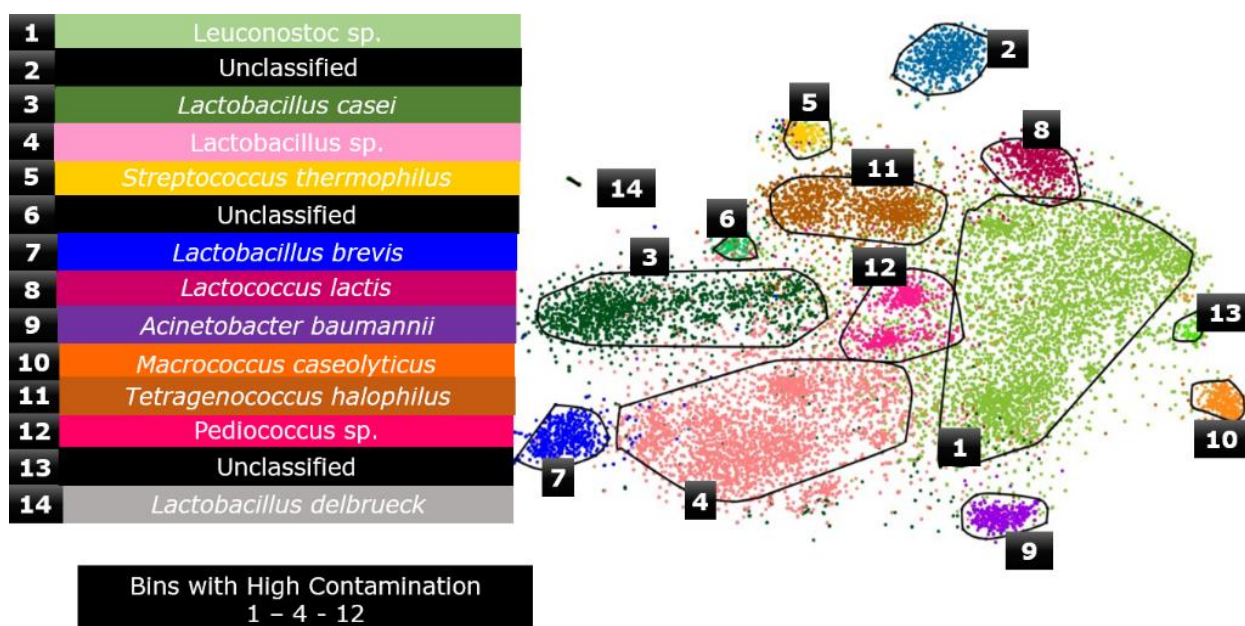


Ολοκληρώνοντας τις αναλύσεις των δεδομένων της τεχνικής shotgun, υλοποιήθηκε η διαδικασία Assembly μέσω της πλατφόρμας CLC και του εργαλείου IDBA. Τα αρχεία που δημιουργήθηκαν περιείχαν τα reads των αρχικών δεδομένων συναρμολογημένα σε contigs. Στη συνέχεια χρησιμοποιήθηκε το εργαλείο BusyBee για τη διαδικασία του binning και αναλύθηκαν τα έξι διαφορετικά δείγματα τυριών. Στις ακόλουθες εικόνες παρουσιάζονται τα αποτελέσματα του binning, δηλαδή η ομαδοποίηση των contigs και τα κυρίαρχα είδη βακτηρίων και ζυμών για κάθε bin.

Επίσης παρατίθεται πίνακας με τις τιμές Completeness, Contamination και Strain heterogeneity. Μέσω των τιμών του πίνακα μπορούμε να συμπεράνουμε την παρουσία μολύνσεων σε ένα bin. Με τον όρο μόλυνση στη συγκεκριμένη περίπτωση εννοούμε την τοποθέτηση contigs από διαφορετικό μικροοργανισμό σε κοινό bin. Η μόλυνση μπορεί να προκληθεί από λάθη κατά το clustering των contigs (μεγάλη τιμή Contamination και χαμηλή τιμή Strain Heterogeneity) ή λόγω της ομοιότητας των αλληλουχιών, που μπορεί να προέρχονται από κοινό πρόγονο, και τοποθετήθηκαν σωστά στο ίδιο bin (μεγάλος βαθμός Contamination και μεγάλος βαθμός Strain Heterogeneity). Στη δεύτερη περίπτωση μπορούμε να εμπιστευτούμε το αποτέλεσμα του Binning για το συγκεκριμένο bin λαμβάνοντας υπόψιν την τιμή του Completeness.

Το εργαλείο BusyBee υλοποιεί ταξινομική ανάλυση έως την ταξινομική βαθμίδα των species. Η χρωματική διαφοροποίηση των bins συμβάλει στην καλύτερη ευκρίνεια των ορίων της ομαδοποίησης. Συμπληρωματικά, αναγράφεται το κυρίαρχο είδος που κατηγοριοποιήθηκε ανά bin. Τέλος, με τον όρο Undefined νοείται η αδυναμία του εργαλείου να αποδώσει συγκεκριμένο είδος ή γένος.

Εικόνα 43: Σφέλα Δείγμα 1

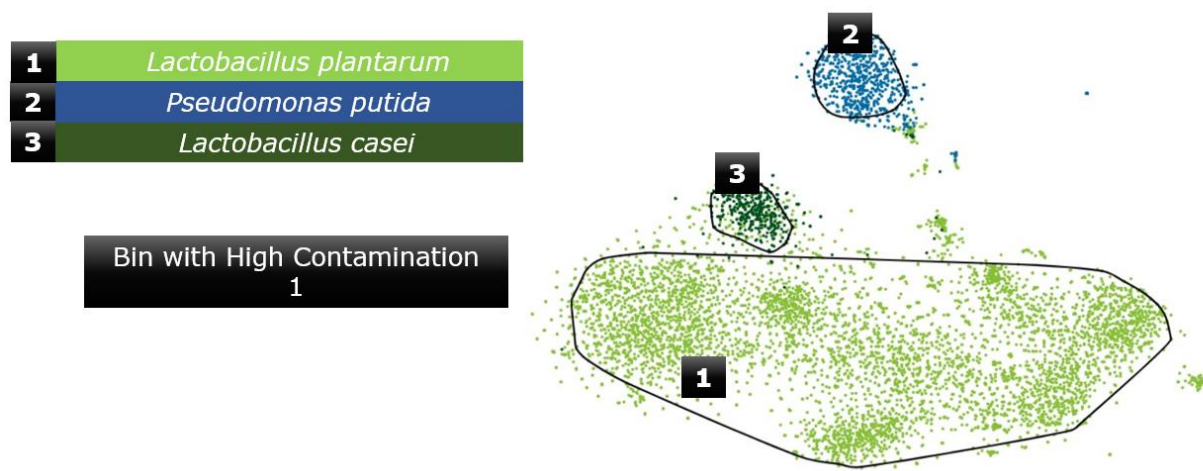


Πίνακας 6: Επιμέρους στοιχεία του bin

Bin Id	Completeness	Contamination	Strain heterogeneity
bin.1	95.5	382.88	9.3
bin.2	66.67	12.61	100
bin.3	90.09	11.71	20
bin.4	95.5	227.93	14.61
bin.5	64.86	2.7	0
bin.6	0.9	0	0
bin.7	95.5	43.24	1.89
bin.8	72.97	3.6	50
bin.9	83.78	4.5	60
bin.10	92.79	7.21	44.44
bin.11	90.99	50.45	22.37
bin.12	86.49	83.78	7.21
bin.13	0	0	0
bin.14	94.59	0.9	0

Παρατηρούμε στα αποτελέσματα για το δείγμα 1 Σφέλας, τα bins 3-5-14 να παρουσιάζουν αυξημένη τιμή στη μεταβλητή Completeness και μειωμένες τιμές για τις μεταβλητές Contamination και Strain heterogeneity. Συγκεκριμένα, τα κυρίαρχα είδη των bins είναι *Lb. casei* (bin 3), *Str. thermophilus* (bin 5) και *L. delbrueck* (bin 14). Είδη τα οποία κυριαρχούν στα λευκά τυριά άλμης, όπως στην περίπτωση των δειγμάτων Σφέλας και Χαλίτζι.

Εικόνα 44: Σφέλας δείγμα 2

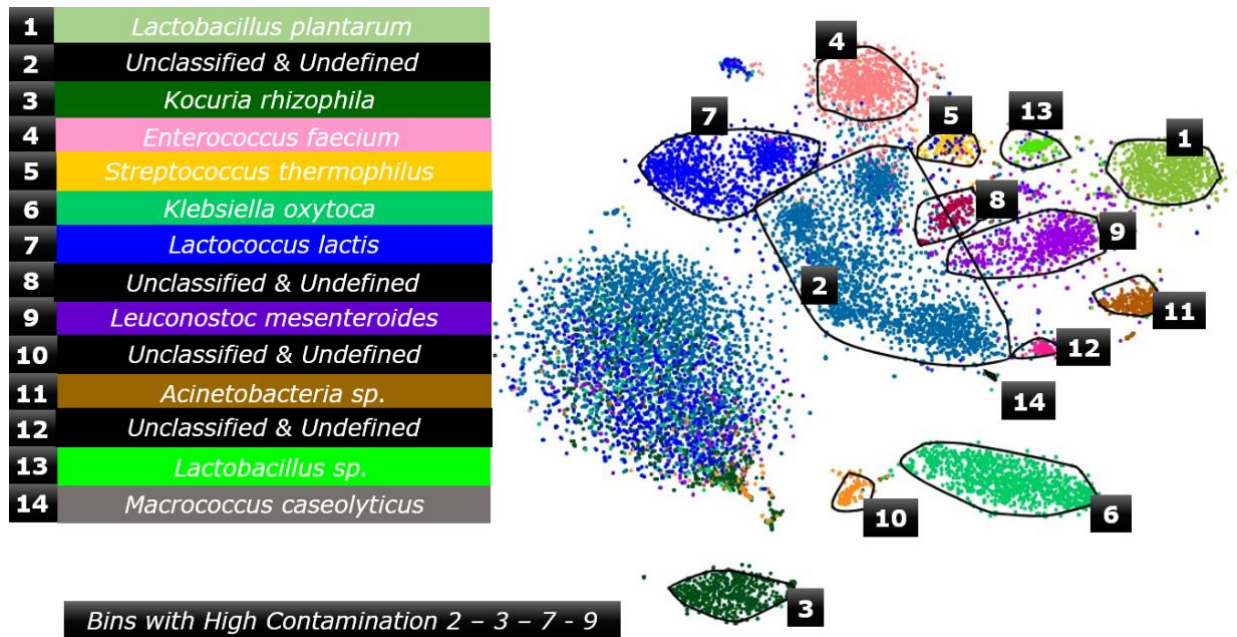


Πίνακας 7: Επιμέρους στοιχεία του bin

Bin Id	Completeness	Contamination	Strain heterogeneity
bin.1	95.5	609.01	5.34
bin.2	95.5	14.41	0
bin.3	18.02	2.7	0

Στο δείγμα 2 Σφέλας, παρατηρούμε τη δημιουργία τριών bins, εκ των οποίων στο bin 1 υπολογίσθηκε μεγάλο ποσοστό μόλυνσης και στο bin 3 μικρό ποσοστό πληρότητας. Το bin 1 παρουσιάζει μόλυνση λόγω τοποθέτησης διαφορετικών μικροοργανισμών στο ίδιο bin, με αυξημένα ποσοστά του γένους *Lactococcus* και *Lactobacillus*. Το bin 2 παρουσιάζει μικρό ποσοστό μόλυνσης, με κυρίαρχο είδος *Pseudomonas putida* 20%. Το bin 3 παρουσιάζει σε ποσοστό 100% το είδος *Lb. casei*.

Εικόνα 45: Τρίτο δείγμα Σφέλας



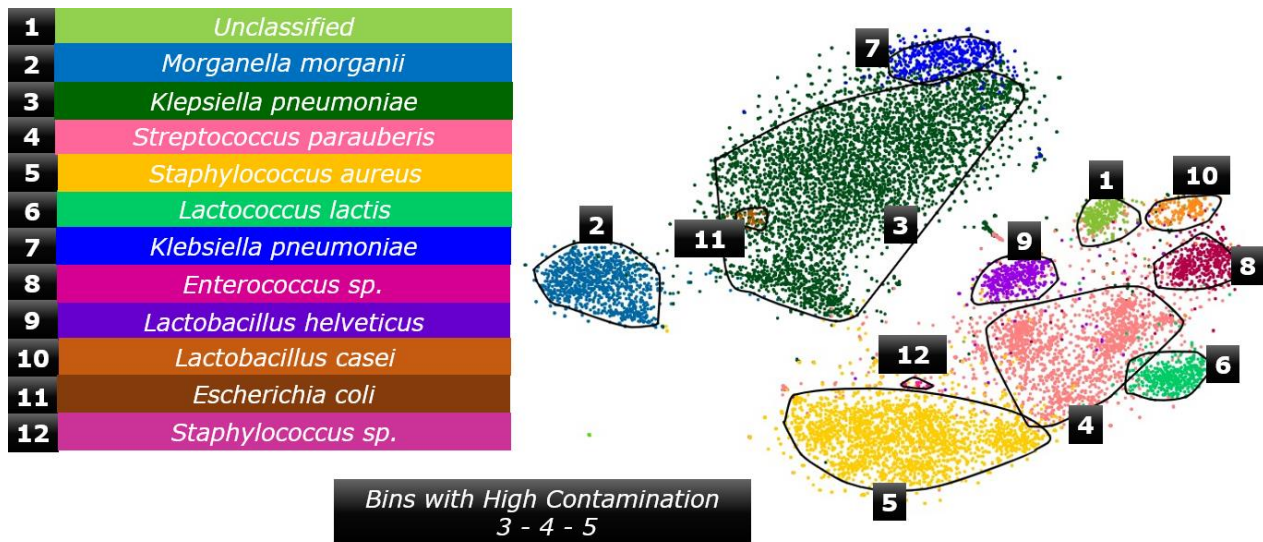
Πίνακας 8: Επιμέρους στοιχεία του bin

Bin Id	Completeness	Contamination	Strain heterogeneity
bin.1	55.86	4.5	60
bin.2	97.3	318.02	13.91
bin.3	94.59	88.29	5.66
bin.4	51.35	9.91	69.23
bin.5	91.89	17.12	52.38
bin.6	83.78	12.61	73.33
bin.7	93.69	82.88	11.97
bin.8	54.95	19.82	2.94
bin.9	94.59	173.87	3.49
bin.10	93.69	1.8	0
bin.11	95.5	1.8	100
bin.12	95.5	16.22	0
bin.13	68.47	0.9	0
bin.14	95.5	0.9	0

Το δείγμα 3 Σφέλας πραγματοποιήθηκε ομαδοποίηση σε περισσότερα bins. Τρία από αυτά παρουσιάζουν αυξημένο ποσοστό πληρότητας και απουσία μολύνσεων. Το bin 14 με κυρίαρχο είδος το *Macrocooccus caseolyticus* και ποσοστό 100%. Η παρουσία του είδους εμφανίζεται περισσότερο στα σκληρά κίτρινα τυριά. Τα επόμενα δύο bins 10-12, ενώ παρουσιάζουν αυξημένο ποσοστό πληρότητας, δεν ήταν εφικτός ο προσδιορισμός των ειδών που

περιέχονται. Ενδιαφέρον παρουσιάζεται σε bins με αυξημένες τιμές Contamination και Strain heterogeneity. Το bin 1 περιέχει το είδος *Lactobacillus plantarum* σε ποσοστό 95%, το bin 4 το είδος *E. faecium* σε ποσοστό 85% με αυξημένη την τιμή του Strain heterogeneity και το bin 6 το είδος *Klebsiella oxytoca* σε ποσοστό 70%. Το τελευταίο είδος ανήκει στην Οικογένεια των *Enterobacteriaceae* και προμηνύει την παρουσία μόλυνσης στο δείγμα καθώς ανήκει στην κατηγορία των παθογόνων. Τα δύο προηγούμενα είδη, *Lb. plantarum* και *E. faecium* ανήκουν στα είδη βακτηρίων που βρίσκονται σε αφθονία στα λευκά τυριά άλμης.

Εικόνα 46: Δείγμα 1 Χαλίτζι

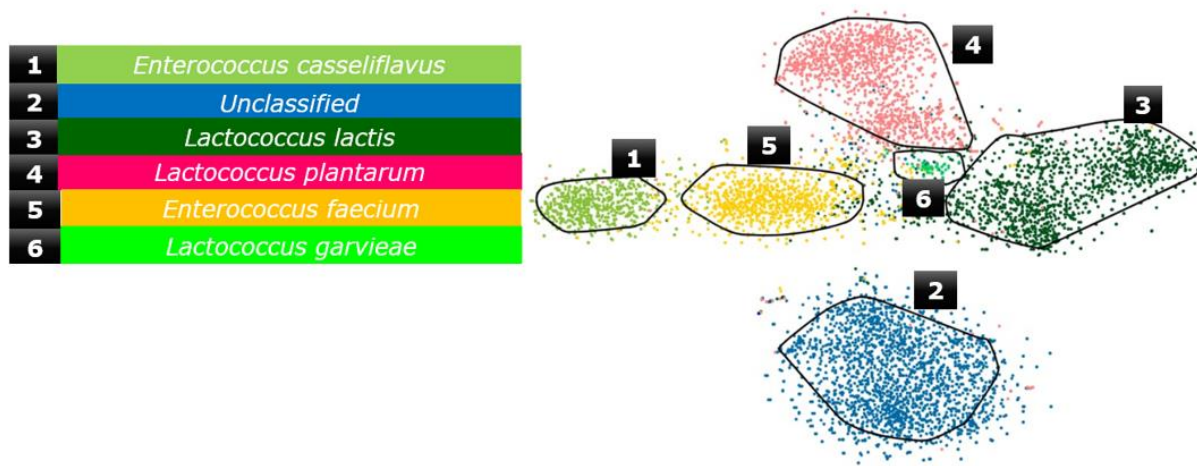


Πίνακας 9: Επιμέρους στοιχεία του bin

Bin Id	Completeness	Contamination	Strain heterogeneity
bin.1	93.69	5.41	0
bin.2	41.44	4.5	40
bin.3	91.89	115.32	78.87
bin.4	96.4	326.13	23.43
bin.5	95.5	460.36	35.4
bin.6	54.05	3.6	75
bin.7	8.11	0	0
bin.8	96.4	47.75	20
bin.9	63.96	2.7	33.33
bin.10	95.5	2.7	0
bin.11	0.9	0	0
bin.12	1.8	0	0

Ενδιαφέρον παρουσιάζεται στα αποτελέσματα της ανάλυσης των δειγμάτων Χαλίτζι. Στο δείγμα 1 πραγματοποιήθηκε ταξινόμηση σε 12 διαφορετικά bins. Στο bin 1 ο αλγόριθμος δεν μπόρεσε να πραγματοποιήσει αναγνώριση των αλληλουχιών. Υπολογίσθηκε όμως η παρουσία του γένους *Lactobacillus* σε ποσοστό 21%. Στο bin 10 υπολογίσθηκε το είδος *Lb. casei* σε ποσοστό 50%, με μικρές τιμές Contamination και Strain heterogeneity. Παρατηρείται η παρουσία παθογόνων βακτηρίων του γένους *Klebsiella* και *Staphylococcus* με μικρά ποσοστά. Η παρουσία παθογόνων προέρχεται από τα μειωμένα μέτρα υγιεινής κατά την παραγωγή οικιακών τυριών.

Εικόνα 47: Δείγμα 2 Χαλίτζι



Πίνακας 10: Επιμέρους στοιχεία του bin

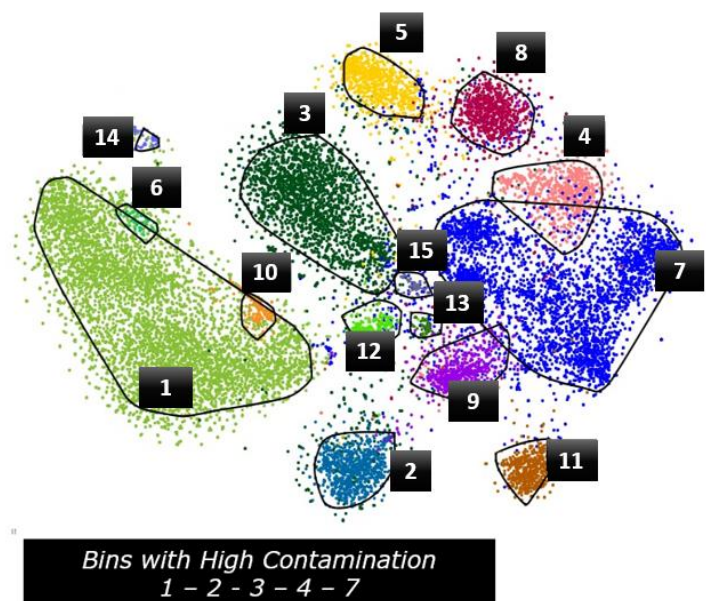
Bin Id	Completeness	Contamination	Strain heterogeneity
bin.1	19.82	1.8	100
bin.2	93.69	72.07	86.9
bin.3	72.07	20.72	58.62
bin.4	94.59	46.85	11.76
bin.5	90.99	34.23	62.5
bin.6	87.39	1.8	50

Στην περίπτωση των αποτελεσμάτων της ανάλυσης του δείγματος 2 Χαλίτζι, τα bins 2-3-5 έχουν μεγάλη τιμή Contamination και το bin 1 μεγάλη τιμή Strain heterogeneity. Αυτό υποδεικνύει την τοποθέτηση αλληλουχιών με κοινή προέλευση στο ίδιο bin. Η παρουσία του βακτηρίου *Enterococcus casseliflavus* στο bin 1 βρίσκεται σε ποσοστό 93%. Τα βακτήρια που ανήκουν στην

οικογένεια των *Enterococcaceae*, εμφανίζονται στα τυροκομικά προϊόντα με διαφοροποιήσεις ως προς τα είδη και τα στάδια που ανιχνεύονται. Από προγενέστερες μελέτες έχει διατυπωθεί η παρουσία τους στο γάλα και σε στάδια κατά την ωρίμανση του τυριού, χωρίς να υπάρχουν μελέτες σχετικά με την παρουσία τους στο μεσοδιάστημα. Κάποια είδη της οικογένειας, εάν βρεθούν στον ανθρώπινο οργανισμό σε μεγαλύτερες ποσότητες, μπορεί να προκαλέσουν παθογένειας. Στη συγκεκριμένη περίπτωση, καθώς γνωρίζουμε ότι τα δείγματα Χαλίτζι προέρχονται από οικιακή παραγωγή, η παρουσία βακτηρίων της οικογένειας *Enterococcaceae* προμηνύει την έλλειψη παστερίωσης ή τις μειωμένες συνθήκες υγιεινής κατά την παρασκευή του τυριού. Στο bin 4, το βακτήριο *Lactococcus plantarum* που είναι από τα είδη βακτηρίων στα λευκά τυριά άλμης, υπολογίσθηκε σε ποσοστό μεγαλύτερο του 65%.

Εικόνα 48: Δείγμα 3 Χαλίτζι

1	<i>Klebsiella oxytoca</i>
2	<i>Lactobacillus plantarum</i>
3	<i>Lactobacillus casei</i>
4	<i>Streptococcus macedonicus</i>
5	Unclassified
6	<i>Enterobacter cloacae</i>
7	<i>Streptococcus sp. & Enterococcus sp.</i>
8	<i>Enterococcus faecalis</i>
9	<i>Leuconostoc mesenteroides</i>
10	Unclassified
11	<i>Macroccoccus caseolyticus</i>
12	Undefined & Unclassified
13	<i>Lactobacillus plantarum</i>
14	Unclassified
15	<i>Lactobacillus sp.</i>



Πίνακας 11: Επιμέρους στοιχεία του bin

Bin ID	Completeness	Contamination	Strain heterogeneity
bin.1	61.26	76.58	82.07
bin.2	93.69	36.04	70.45
bin.3	95.5	94.59	74.42
bin.4	79.28	126.13	29.14
bin.5	27.03	2.7	33.33
bin.6	0	0	0
bin.7	95.5	269.37	16.45
bin.8	67.57	14.41	50
bin.9	87.39	8.11	22.22
bin.10	0	0	0
bin.11	76.58	9.91	45.45
bin.12	95.5	2.7	66.67
bin.13	0.9	0	0
bin.14	92.79	2.7	0
bin.15	0	0	0

Στην περίπτωση των αποτελεσμάτων του δείγματος 3 Χαλίτζι, ο αλγόριθμος ταξινόμησης δημιούργησε 15 bins. Στο bin 9 παρατηρούνται χαμηλές τιμές Contamination και Strain heterogeneity, με το είδος *Leuc. mesenteroides* σε ποσοστό 65%. Το είδος αυτό το συναντάμε συχνά σε καλλιέργειες εκκίνησης. Συνεχίζοντας, τα αποτελέσματα δείχνουν την παρουσία του γένους *Bifidobacterium*. Τα βακτήρια αυτού του γένους είναι αναερόβια και κάποια από αυτά χρησιμοποιούνται ως προβιοτικά. Η αρχική εκτίμηση του όρου προβιοτικά αναφέρεται σε βακτήρια με ανοχή σε ιδιαίτερα περιβάλλοντα όπως το γαστρεντερικό σύστημα του ανθρώπου, την ικανότητα αποικισμού σε ξενιστή με χαρακτηριστικά την αντιμικροβιακή δραστηριότητα, τις ανοσορυθμιστικές ιδιότητες και την πρωτεολυτική δράση [94, 95]. Το γένος *Bifidobacterium* βρίσκεται συνήθως σε κίτρινα ημίσκληρα τυριά.

Στα αποτελέσματα του δείγματος 3 Χαλίτζι, παρατηρήθηκαν βακτήρια που βρίσκονται σε λευκά τυριά άλμης, όπως *Lb. plantarum* 90% (bin2), *Lb. casei* 50% (bin3), *Str. gallolyticus* 50% (bin4) το οποίο βρίσκεται σε μεγάλες συγκεντρώσεις στο γιαούρτι, *Leuc. mesenteroides* 60% (bin 9) και τέλος *Lb. plantarum* 40% (bin 13).

Τέλος, για τα αποτελέσματα της ανάλυσης του δείγματος 3, μπορούμε να συμπεράνουμε την παρουσία παθογόνων βακτηρίων. Τα γένη *Klebsiella* και *Staphylococcus*, όπως αναφέραμε και για το δείγμα 1, βρίσκονται με αυξημένα

ποσοστά. Το γεγονός αυτό έχει προκληθεί από τα μειωμένα μέτρα υγιεινής κατά την τυροκόμηση.

Η διαδικασία assembly υλοποιήθηκε μέσω του εργαλείου IDBA και της πλατφόρμας CLC. Τα αρχεία εξόδου της υλοποίησης assembly χρησιμοποιήθηκαν ως αρχεία εισόδου στο εργαλείο BusyBee. Τα αποτελέσματα και στις δύο περιπτώσεις ήταν κοινά.

3.3. Αποτελέσματα Λειτουργικής Ανάλυσης Δεδομένων

Shotgun

Σε επόμενο και τελευταίο στάδιο της μελέτης, επιχειρήθηκε μια αρχική διερεύνηση στις λειτουργικές ιδιότητες των δειγμάτων, μέσω της αξιοποίησης των αρχικών δεδομένων shotgun. Στην περίπτωση των μεταγονιδιωματικών δεδομένων, η λειτουργική ανάλυση προσπαθεί να διερευνήσει τις λειτουργικές ιδιότητες που ενδείκνυται να υλοποιούνται στο σύνολο του μικροβιακού δείγματος. Η διερεύνηση έχει ως δεδομένο τις μεταγονιδιωματικές αλληλουχίες και υλοποιείται με τη βοήθεια βάσεων δεδομένων απόθεσης καλά σχολιασμένων πληροφοριών.

Η εκτέλεση της λειτουργικής ανάλυσης πραγματοποιήθηκε με τη βοήθεια της πλατφόρμας MG-Rast μέσω της διασύνδεσής της με τη βάση δεδομένων KO (Kegg Orthology). Η KO περιέχει πληροφορίες για γνωστές μοριακές λειτουργίες που εκπροσωπούνται με τον όρο Functional Orthologs. Ένα λειτουργικό ορθόλογο έχει ορισθεί στα πλαίσια των μοριακών δικτύων της KEGG. Πιο συγκεκριμένα, ένα δίκτυο (pathway) που δημιουργείται στην KEGG περιέχει κόμβους δεδομένων, σε αυτούς τους κόμβους αποδίδεται αναγνωριστικός αριθμό/γράμμα. Αυτοί οι κόμβοι συσχετίζονται με πειραματικά χαρακτηρισμένα γονίδια ή/και πρωτεΐνες από συγκεκριμένους οργανισμούς με στόχο την εύρεση ορθόλογων σε άλλους οργανισμούς.

Η πλατφόρμα MG-Rast προκειμένου να εντοπίσει τις πιθανές λειτουργίες των γονιδίων, χρησιμοποιεί τα reads των αρχικών δεδομένων και δημιουργεί ομάδες πρωτεϊνών με επίπεδο ταύτισης 90%, χρησιμοποιώντας την εφαρμογή UCLUST³⁹ του QIIME⁴⁰. Οι αλληλουχίες με το μεγαλύτερο μήκος για κάθε ομάδα επιλέγονται για τις αναλύσεις ομοιότητας που εκτελούνται στη συνέχεια. Η ανάλυση ομοιότητας βασίζεται στην στοίχιση μέσω του αλγορίθμου BLAT [96] και της βάσης δεδομένων πρωτεϊνών M5nr [83].

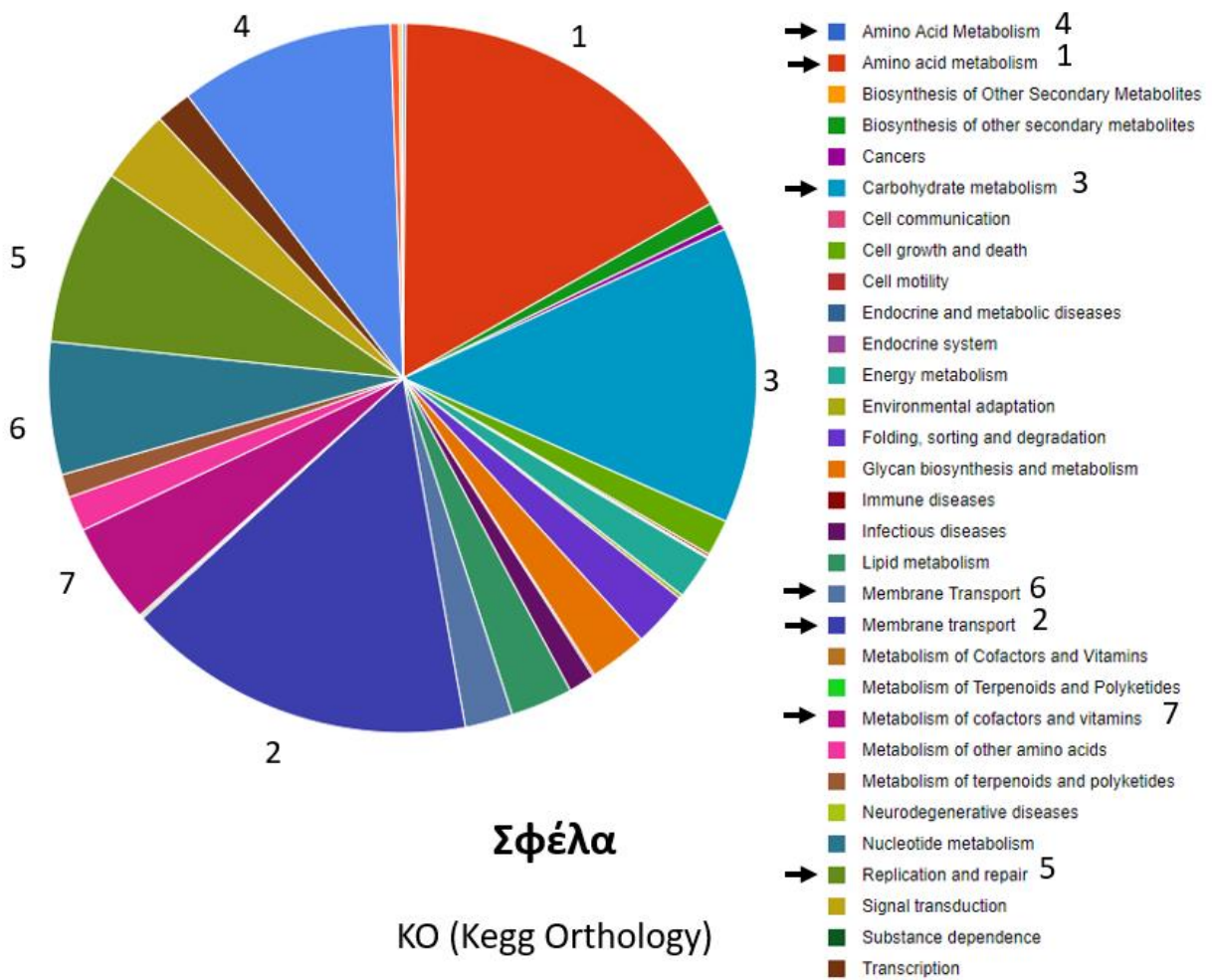
³⁹ <https://omictools.com/uclust-tool>

⁴⁰ <http://qiime.org/index.html>

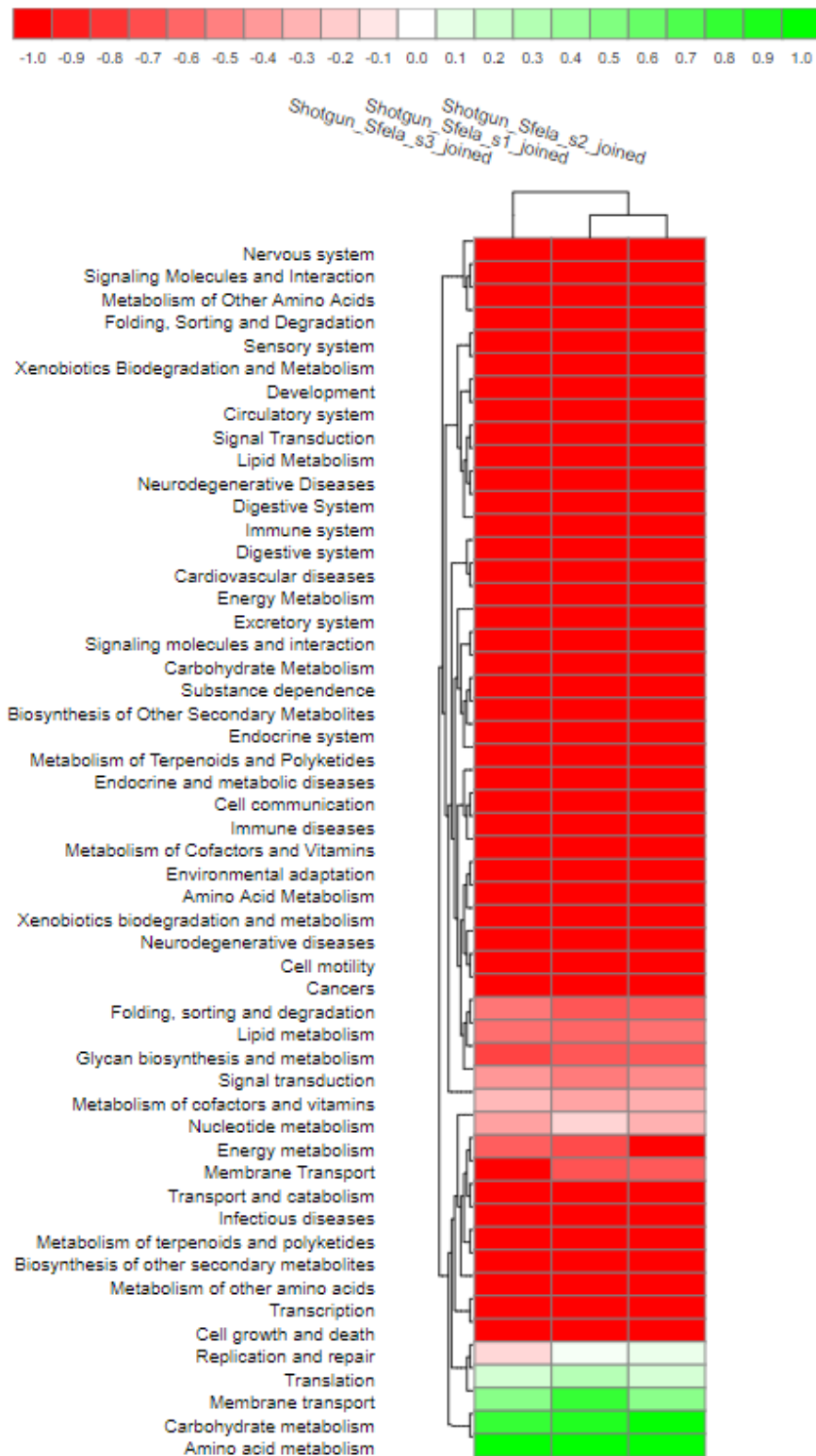
Τα δεδομένα που εισήχθησαν στην ανάλυση ήταν τα τρία δείγματα Σφέλας και τα τρία δείγματα Χαλίτζι. Τα αποτελέσματα που εξήχθησαν από την ανάλυση είναι αρκετά κοινά για τις δύο περιπτώσεις λευκών τυριών άλμης.

Στις εικόνες 49 έως 52 αποτυπώνονται οι βασικές λειτουργικές ιδιότητες στο σύνολο του μικροβιώματος των δύο διαφορετικών τυριών. Και στις δύο περιπτώσεις μεγαλύτερο ποσοστό καταλαμβάνει η λειτουργία του μεταβολισμού των αμινοξικών καταλοίπων και έπεται με ελάχιστα μικρότερο ποσοστό η λειτουργία της μεμβρανικής μεταφοράς. Και στις δύο περιπτώσεις, το αποτέλεσμα ήταν αναμενόμενο καθώς το σύνθετο δυναμικό σύμπλεγμα του μικροβιώματος εντός των τυροκομικών προϊόντων, έχει ως απαραίτητη προϋπόθεση τον μεταβολισμό των αμινοξέων και τη μεταφορά μορίων εντός και εκτός κυττάρου. Οι μικροοργανισμοί λαμβάνουν τα απαραίτητα θρεπτικά στοιχεία μέσω των μεμβρανών τους και τα μετατρέπουν με ενζυμικές αντιδράσεις σε προϊόντα απαραίτητα για την ανάπτυξη και τον πολλαπλασιασμό τους. Φυσικό επακόλουθο νοείται η παρουσία γονιδίων που κωδικοποιούν λειτουργίες αποδόμησης των αμινοξικών καταλοίπων για την παραγωγή της απαραίτητης ενέργειας. Συμπληρωματικά, η επιβίωση των μικροοργανισμών δεν θα ήταν δυνατή εάν δεν επιτρεπόταν η μεμβρανική μεταφορά. Επόμενη λειτουργική ιδιότητα είναι ο μεταβολισμός των υδατανθράκων, εξίσου απαραίτητος για την παραγωγή ενέργειας. Ακολουθεί ο μεταβολισμός των νουκλεϊκών οξέων, η αναπαραγωγή - επιδιόρθωση και η μεταγωγή σήματος. Η ομοιότητα των αποτελεσμάτων για τα δύο διαφορετικά λευκά τυριά άλμης ήταν αναμενόμενη, καθώς οι λειτουργίες που επιτελεί το μικροβίωμα των τυροκομικών προϊόντων βασίζονται στις ίδιες ανάγκες κάλυψης ενέργειας για επιβίωση.

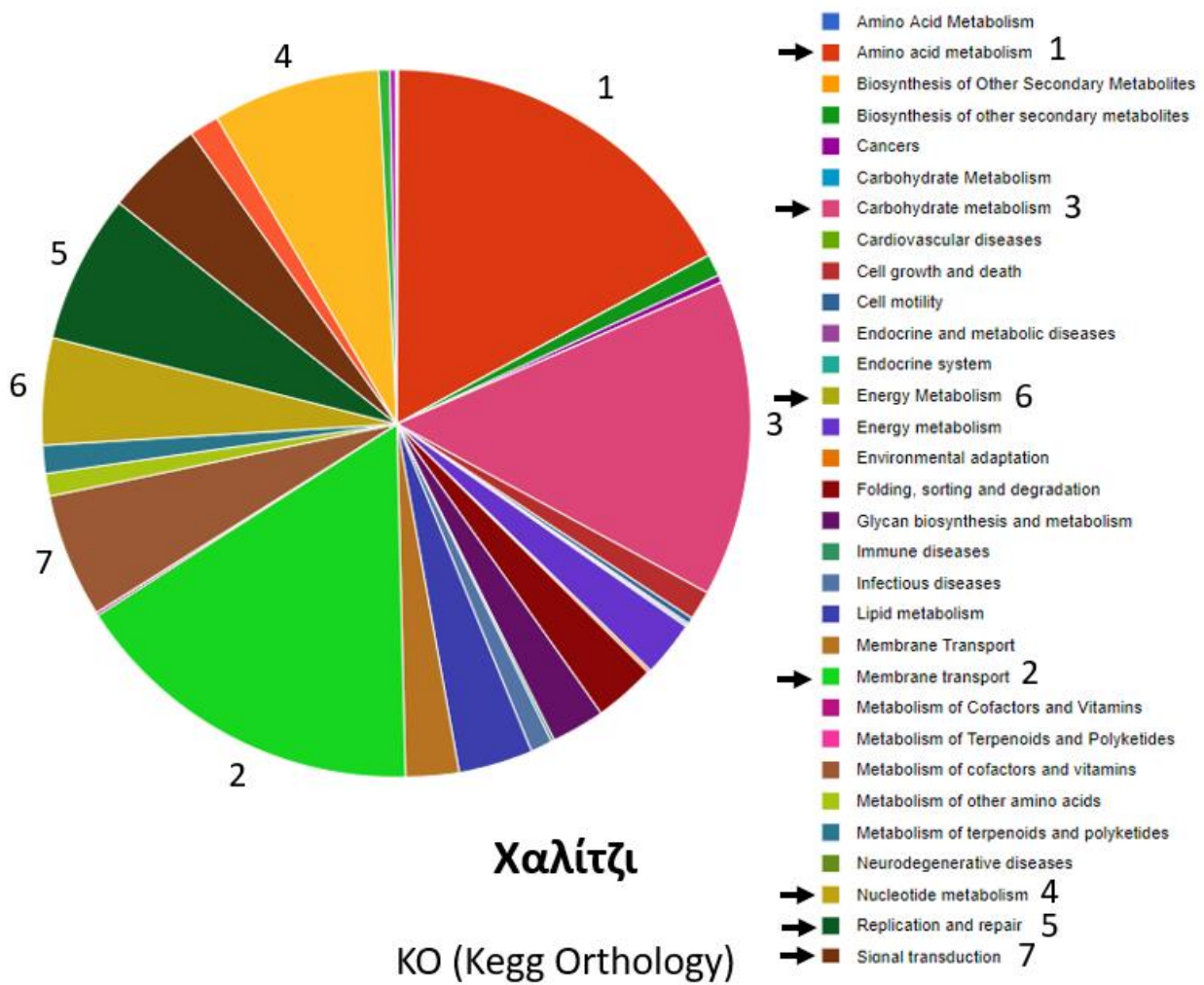
Εικόνα 49: Chart με τα αποτελέσματα λειτουργικής ανάλυσης για τα 3 δείγματα Σφέλας



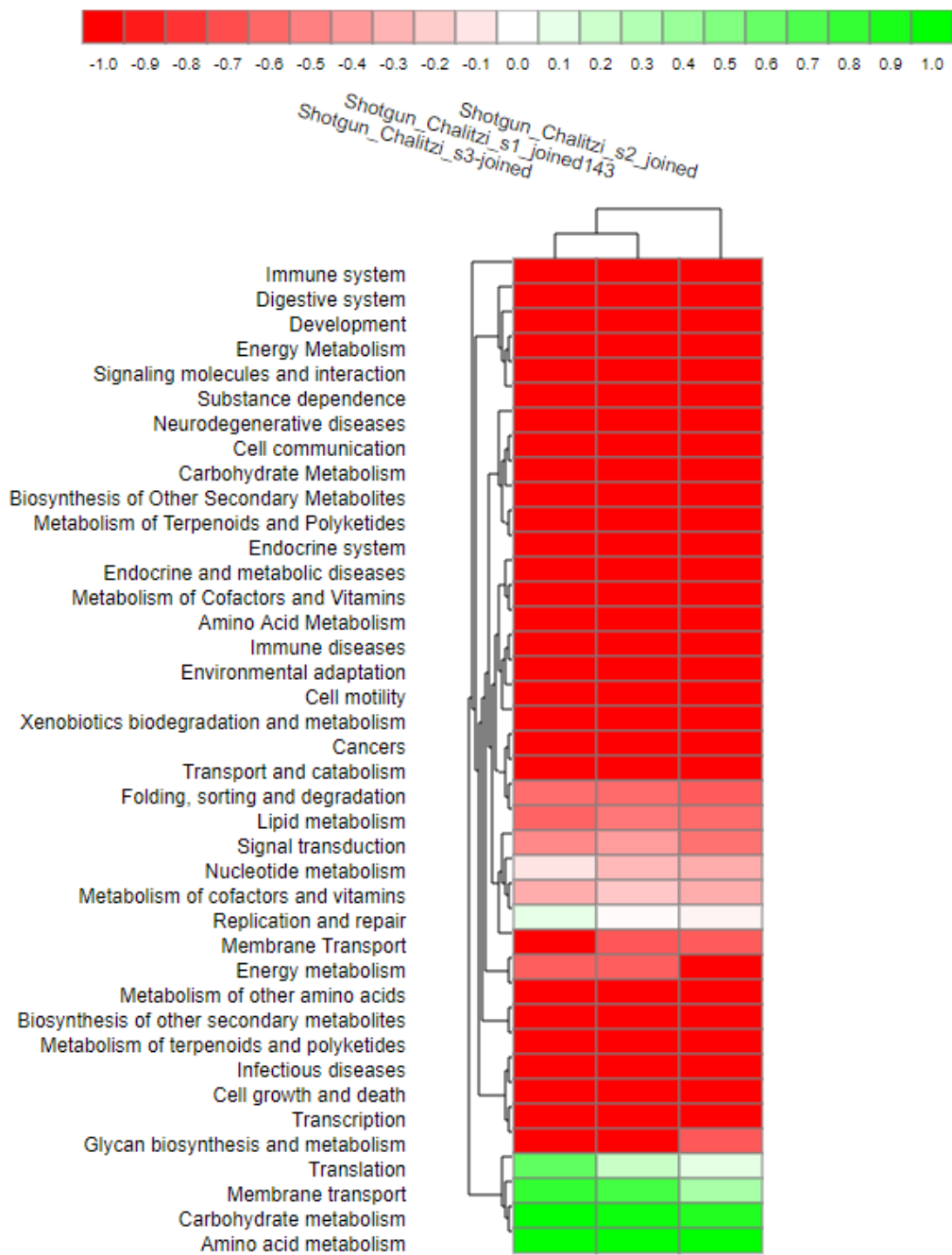
Εικόνα 50: Heatmap με τα αποτελέσματα λειτουργικής ανάλυσης για τα 3 δείγματα Σφέλας



Εικόνα 51: Chart με τα αποτελέσματα λειτουργικής ανάλυσης για τα 3 δείγματα Χαλίτζι



Εικόνα 52: Heatmap με τα αποτελέσματα της λειτουργικής ανάλυσης για τα 3 δείγματα Χαλίτζι



4. ΣΥΜΠΕΡΑΣΜΑΤΑ

4.1 Μεταγονιδιωματική Ανάλυση - Εργαλεία

Η βιοπληροφορική ανάλυση των μεταγονιδιωματικών δεδομένων συνεισφέρει στη συσσώρευση πληροφοριών και γνώσεων που με παλαιότερες μεθόδους δεν ήταν εφικτή. Η εξόρυξη δεδομένων από περιβαλλοντικά δείγματα έχει επεκτείνει σε μεγάλο βαθμό την αντίληψη του δυναμικού των μικροβιωμάτων, καθιστώντας τη μεταγονιδιωματική έναν από τους ανερχόμενους επιστημονικούς κλάδους. Τα αποτελέσματα που εξάγονται επιτρέπουν την ταξινόμηση των ειδών καθώς και την εξαγωγή δεδομένων αφθονίας των πληθυσμών. Η μέθοδος shotgun επιτρέπει επιπρόσθετα τη διερεύνηση των λειτουργικών ιδιοτήτων στο σύνολο των δειγμάτων. Συμπληρωματικά, οι μέθοδοι Amplicon, δηλαδή επιλεγμένες περιοχές αλληλούχισης, 16S rDNA περιοχή για βακτήρια και ITS περιοχή για μύκητες - ζύμες, συνεισφέρουν στη συστηματική ταξινόμηση του μικροβιώματος έως την ταξινομική βαθμίδα του γένους.

Η βιοπληροφορική ανάλυση των δεδομένων αποτελεί μια σύνθετη και πολύπλοκη διαδικασία. Απαραίτητη σε όλες τις περιπτώσεις αναλύσεων είναι η αυξημένη υπολογιστική δύναμη. Κάποια από τα επιμέρους εργαλεία που αξιοποιήθηκαν για την ανάλυση, έχουν ως προαπαιτούμενο η μνήμη τυχαίας προσπέλασης να ξεπερνά τα 50GB. Γεγονός που υποδεικνύει την αυξημένη υπολογιστική ισχύ που είναι αναγκαία για την ανάλυση. Το πρόβλημα αυτό μπορεί να ξεπεραστεί αξιοποιώντας πλατφόρμες και εργαλεία ανοιχτού λογισμικού, όπως στην περίπτωση της μελέτης. Επίσης, το χρονικό διάστημα που απαιτείται για την υλοποίηση των υπολογιστικών διεργασιών είναι αρκετά μεγάλο, ακόμα και σε περιπτώσεις χρήσης συγκεκριμένων servers διαμορφωμένων για ανάλυση δεδομένων metagenomics. Στη συγκεκριμένη μελέτη η ανάλυση των δεδομένων shotgun στην πλατφόρμα MG-Rast και στα εργαλεία IDBA και GOTTECHA της πλατφόρμας KBase ξεπέρασε τις 48 ώρες για κάθε δείγμα.

Η επιμέρους ροή εργασιών διαφοροποιείται ανάλογα με το βιολογικό ερώτημα που θέτουμε. Στην παρούσα μελέτη τα βιολογικά ερωτήματα ήταν η

εξακρίβωση της ποικιλότητας του μικροβιώματος των δύο διαφορετικών λευκών τυριών άλμης, η ταυτοποίηση των βακτηρίων, μυκητών και ζυμών που εμπεριέχονται σε επίπεδο είδους, η ανάλυση αφθονίας και η διερεύνηση των κεντρικών λειτουργικών ιδιοτήτων που κωδικοποιούνται μέσω του γενετικού υλικού των οικοσυστημάτων.

4.2 Μικροβίωμα Λευκών Τυριών Άλμης

Τα λευκά τυριά άλμης είναι η πιο δημοφιλής κατηγορία τυριών που παρασκευάζεται στη Βορειοανατολική Μεσόγειο και στα Βαλκάνια [19]. Για αυτό το είδος τυριών χρησιμοποιείται αρχικά στερεή μορφή άλατος, ενώ ωριμάζουν και αποθηκεύονται σε άλμη. Βασική διαφορά των δειγμάτων που μελετήθηκαν, μεταξύ άλλων τυροκομικών προϊόντων, είναι η παραγωγή τους από αίγαιο και πρόβειο νωπό ή παστεριωμένο γάλα. Η προέλευση του γάλατος και η διαδικασία παραγωγής τυροκομικών προϊόντων συμβάλει σημαντικά στη σύνθεση του μικροβιώματος των τυριών. Επίσης, η παρασκευή τους βασίζεται και επηρεάζεται από πολλούς παράγοντες όπως η ποιότητα του γάλακτος και η θερμική επεξεργασία προς αποφυγή ενδεχόμενων μολύνσεων κατά τη διάρκεια παραγωγής και ωρίμανσής [34]. Η άλμη που χρησιμοποιείται στα λευκά τυριά αυτής της κατηγορίας, αποσκοπεί στη σωστή αποθήκευση διατηρώντας την κατάλληλη ποσότητα άλατος στο τελικό προϊόν. Οι συνθήκες του περιβάλλοντος που δημιουργούνται κατά την τοποθέτηση σε άλμη βοηθούν στην ανάπτυξη διαφορετικών ειδών βακτηρίων. Η υψηλή συγκέντρωση άλατος, το χαμηλό pH και η χαμηλή θερμοκρασία συμβάλουν στην επιβίωση συγκεκριμένων βακτηρίων που επηρεάζουν την ανάπτυξη δευτερογενών μικροοργανισμών στο τελικό προϊόν. Οι μικροοργανισμοί που αναπτύσσονται είναι είτε επιθυμητοί, όπως *Lactobacillus*, είτε μη επιθυμητοί αλλά όχι βλαβεροί όπως κάποια είδη *Lactobacillus*, *Micrococcus*, *Staphylococcus*, είτε παθογόνοι όπως τα είδη *L. monocytogenes*, *E. coli* και *Staph. aureus*.

Βασική παράμετρος για τη μικροβιακή ποικιλότητα του τυριού είναι η καλλιέργεια εκκίνησης [97]. Στην περίπτωση των λευκών τυριών άλμης χρησιμοποιούνται καλλιέργειες εκκίνησης με βακτήρια μεσόφιλα, θεرمόφιλα

ή συνδυασμός των δύο ειδών. Οι καλλιέργειες με μεσόφιλα βακτήρια περιλαμβάνουν είδη του γένους *Lactococcus* όπως το είδος *L. lactis* ενώ στις θερμόφιλες καλλιέργειες περιλαμβάνονται τα γένη *Streptococcus* και *Lactobacillus*, όπως τα είδη *S. thermophilus* και *Lb. delbrueckii subsp. bulgaricus* [98]. Άμεση συνάρτηση με τις καλλιέργειες εκκίνησης έχει η διακύμανση του pH στο εσωτερικό του τυριού κατά τη διάρκεια της ωρίμανσης [99]. Σε πολλές περιπτώσεις, χρησιμοποιούνται δευτερογενείς συμπληρωματικές καλλιέργειες για την παρασκευή των τυριών, με κυριαρχία του γένους *Lactobacillus* όπως το είδος *Lb. plantarum* [100]. Τέλος, λόγω της μείωση του pH κατά την παρασκευή των τυριών, παρουσιάζονται σε αρκετές περιπτώσεις Non Starter Lactic Acid Bacteria ομάδες βακτηρίων τα οποία προκαλούν αλλαγές στο μικροπεριβάλλον. Βακτήρια που επιβίωσαν κατά την παστερίωση, όπως του γένους *Lactobacillus* και *Enterococcus*, ενδέχεται να προκαλέσουν την ανάπτυξη NSLAB βακτηρίων [100, 101]. Τα είδη που περιλαμβάνονται συχνότερα είναι *Lb. casei*, *Lb. paracasei*, *Lb. rhamnosus*, *Lb. plantarum*, *Lb. paraplantarum*, *Lb. pentosus*, *Leuconostoc sp.*, *Weisella sp.*, *E. faecium*, *E. faecalis* και *Enterococcus durans* [102]. Τα είδη αυτά είναι απαραίτητα για να προσδώσουν τα επιθυμητά φυσικοχημικά χαρακτηριστικά στο τυρί, όπως και την προσδοκώμενη γεύση [103].

Μεταξύ όλων των LAB βακτηρίων, αυτά που συμμετέχουν σε μεγαλύτερο βαθμό στο τελικό προϊόν είναι τα γένη *Lactococcus (L.lactis)*, *Enterococcus sp. (E.faecalis, E.faecium)*, *Lactobacillus sp. (Lb. plantarum, Lb. fermentum, Lb. paraplantarum, Lb. casei, Lb. brevis)*, *Leuconostoc sp. (Leuc. mesenteroides, Leuc. pseudomesenteroides)* και *Streptococcus sp. (Str. thermophilus)* [104].

Συμπληρωματικά, την χρονική περίοδο πριν την ωρίμανση των τυριών, αναπτύσσονται είδη που έχουν ανθεκτικότητα σε περιβάλλοντα με αυξημένη ποσότητα άλατος. Τέτοια είδη είναι οι εντερόκοκκοι, όπως τα είδη *E. faecium* και *E. durans* [105]. Το είδος *E. faecium* έχει βρεθεί σε μεγάλους πληθυσμούς στο τυρί Μπατζος και Χαλούμι που ανήκουν στην κατηγορία λευκών τυριών άλμης [97]. Τέλος, υψηλοί πληθυσμοί ειδών που ανήκουν στην κατηγορία των coliforms υποδεικνύουν την κακή υγιεινή κατά την παρασκευή των τυροκομικών προϊόντων. Όμως κατά τη διάρκεια της ωρίμανσης των τυριών,

λόγω της μείωση του pH και σε αρκετές περιπτώσεις την έλλειψη σακχάρων, προκαλεί μείωση της ανάπτυξης των coliforms [106, 107].

4.3 Μικροβίωμα Σφέλα - Χαλίτζι

Στην παρούσα μελέτη χρησιμοποιείται για πρώτη φορά η τεχνική Shotgun, που συνεπάγεται την ανάλυση μεταγονιδιωματικών δεδομένων με τη βοήθεια βιοπληροφορικών μεθόδων, για τα δείγματα Σφέλα και Χαλίτζι. Έχουν πραγματοποιηθεί προγενέστερες μελέτες μόνο για το τυρί Χαλίτζι στην κλίμακα των Amplicon based αναλύσεων.

Στο μικροβίωμα των τυριών Σφέλα και Χαλίτζι απαντώνται σε αφθονία ομάδες οξυγαλακτικών βακτηρίων, όπως ήταν αναμενόμενο. Τα γένη που εμφανίστηκαν με μεγαλύτερα ποσοστά μετά την ανάλυση των δειγμάτων, ήταν *Lactococcus*, *Lactobacillus*, *Streptococcus*, *Leuconostoc* και σε μικρότερες ποσότητες τα γένη *Staphylococcus* και *Klebsiella*.

Πιο συγκεκριμένα, στα τρία δείγματα Σφέλας, παρουσιάζεται διαφοροποίηση μεταξύ των δειγμάτων 1 – 2 με το δείγμα 3. Η διαφοροποίηση αυτή ενδέχεται να προκλήθηκε από το διαφορετικό τρόπο παρασκευής των τυριών, καθώς στις δύο πρώτες περιπτώσεις τα δείγματα ανήκουν σε μικρής εμβέλειας βιομηχανικές παραγωγές, ενώ το δείγμα 3 προέρχεται από οικιακή παραγωγή. Στο δείγμα 1 συναντάμε τα είδη *Str. thermophilus*, *Lb. plantarum*, *Lb. paracasei*, *Lb. delbrueckii* και *Lb. brevis*. Στο δείγμα 2 συναντάμε μικρότερη ποικιλία βακτηρίων και κυριαρχία του είδους *L. lactis*, με το είδος *Lb. plantarum* να έπεται σε μικρότερο ποσοστό. Στο δείγμα 3 εμπεριέχονται τα είδη βακτηρίων *T. halophilus* σε αρκετά μεγάλο ποσοστό και ακολουθεί το είδος *L. lactis* με μικρότερο ποσοστό. Συνεχίζοντας, την τρίτη θέση του δείγματος 3, καταλαμβάνει το είδος ζύμης *D. hansenii* και σε μικρότερη ποσότητα τα γένη βακτηρίων *Lactobacillus*, *Weissella*, *Leuconostoc*, *Kurthia*, *Streptococcus* και παθογόνα είδη του γένους *Staphylococcus*.

Η μικροβιακή σύνθεση των δειγμάτων Σφέλας είναι σε συμφωνία με προγενέστερες έρευνες στα λευκά τυριά άλμης. Η κυριαρχία των *Lactococcus* και *Lactobacillus* ήταν αναμενόμενη. Το είδος *L. lactis* χρησιμοποιείται ευρέως

σε καλλιέργειες εκκίνησης και αρκετές φορές κατά τη διαδικασία ωρίμανσης των λευκών τυριών άλμης [108]. Έχει διαπιστωθεί από παλαιότερες μελέτες ότι ο *L. lactis* επηρεάζει σε πολύ μεγάλο βαθμό την υφή και το άρωμα του τυριού [109].

Είδη του γένους *Lactobacillus* ανήκουν επίσης στην ομάδα LAB βακτηρίων, και διαδραματίζουν σπουδαίο ρόλο στη διεργασία ζύμωσης των τυριών [110]. Οι λακτοβάκιλλοι βρίσκονται στο νωπό γάλα και αυξάνουν τους πληθυσμούς τους στα τυροκομικά προϊόντα κατά την περίοδο της ωρίμανση [111]. Τα πιο διαδεδομένα προς χρήση σε καλλιέργειες εκκίνησης είδη του γένους *Lactobacillus* είναι *Lb. paracasei* και *Lb. rhamnosus*. Η αύξηση της παραγωγής γαλακτικού οξέος συμβάλει στη γρήγορη μείωση του pH που είναι παράγοντας αναστολής της μικροβιακής ανάπτυξης ανεπιθύμητων ειδών δηλαδή βοηθάει με έμμεσο τρόπο στην αποφυγή μολύνσεων. Τέλος, έχει διατυπωθεί ότι οι λακτοβάκιλλοι έχουν αντιβακτηριακή δράση στα είδη *Y. enterocolitica*, *E. coli* και *B. cereus* [112].

Είδη του γένους *Leuconostoc* δεν παρουσιάζονται σε ίδιες ποσότητες εντός των τυροκομικών προϊόντων όπως οι λακτοβάκιλλοι, αλλά εφόσον βρεθούν εντός των προϊόντων συνεχίζουν την ανάπτυξη τους σε χαμηλές τιμές pH. Μία λειτουργία του είδους *Leuc. mesenteroides* είναι η δημιουργία αερίων που συμβάλλουν στη διαμόρφωση μικρών οπών. Επόμενη βασική λειτουργία του είδους είναι η μετατροπή της αιθανόλης σε αιθυλεστέρες κατά την ετεροζυμωτική δράση, για την αποφυγή ανάπτυξης παθογόνων βακτηρίων [113].

Το γένος *Tetragenococcus* ευδοκιμεί σε ακραία περιβάλλοντα ως προς τη συνθήκη του άλατος. Στο μόνο δείγμα που βρέθηκε σε υψηλούς πληθυσμούς ήταν το δείγμα 3 Σφέλας. Το βακτήριο αυτό χρησιμοποιείται ως καλλιέργεια εκκίνησης με στόχο τη βελτίωση της γεύσης σε σάλτσες ζύμωσης, όχι όμως στην περίπτωση των τυριών [114]. Επίσης, το είδος έχει βρεθεί σε δύο τυριών μεξικανικής προέλευσης παραγόμενα από αγελαδινό νωπό γάλα [115], εκ των οποίων το ένα παρομοιάζεται με το ΠΟΠ τυρί Φέτα. Το ποσοστό που υπολογίσθηκε ότι βρισκόταν στο δείγμα 3 Σφέλας, και στις τρεις πλατφόρμες και τα επιμέρους εργαλεία που χρησιμοποιήθηκαν στη μελέτη, ήταν αυξημένο.

Η εμφάνισή του μόνο σε ένα από τα τρία δείγματα Σφέλας, δεν καθιστά ικανή τη διατύπωση ολοκληρωμένου συμπεράσματος. Απαραίτητη είναι η περαιτέρω ανάλυση του δείγματος, ή δειγμάτων παραγόμενα από τον ίδιο οικιακό παραγωγό, με εργαστηριακές μεθόδους.

Συνεχίζοντας, το γένος *Weisella*, που υπολογίστηκε στα δείγματα 2 και 3 Σφέλας, παλαιότερα αναγνωριζόταν εσφαλμένα ως *Lactobacillus* και ομαδοποιούταν με γένη όπως *Leuconostoc* και *Pediococcus*. Η χρήση νέων μεθόδων αλληλούχισης διευκόλυνε τον προσδιορισμό του ως ξεχωριστό γένος. Έχει εντοπισθεί σε ένα ευρύ φάσμα ενδαιτημάτων, με ένα από αυτά να είναι το νωπό γάλα, χωρίς να έχει αναλυθεί εκτενέστερα [116]. Το είδος του γένους που έχει βρεθεί στη μικροχλωρίδα τυροκομικών προϊόντων είναι *W. paramesenteroides*. Συνεισφέρει στην ετερογαλακτική ζύμωση παράγοντας με τη βοήθεια των γαλακτικών αφυδρογονάσης και των ενζύμων πυροσταφυλικής δεκαρβοξυλάσης, γαλακτικό οξύ, αιθανόλη και διοξείδιο του άνθρακα.

Το γένος *Kurthia*, το οποίο βρέθηκε σε μικρή ποσότητα μόνο στο δείγμα 3 Σφέλας, έχει χαρακτηριστεί ως παθογόνο, χωρίς να έχει αναλυθεί εκτενέστερα. Η παρουσία του έχει αναφερθεί σε περιβαλλοντικά δείγματα στο γάλα, στο έδαφος και στο κρέας [117]. Το γένος, σε παλαιότερες μελέτες, είχε ομαδοποιηθεί με τα γένη *Listeria* και *Lactobacillus* λόγω του μικρού GC περιεχομένου στην αλληλουχία του [118].

Το γένος *Streptococcus* και πιο συγκεκριμένα το είδος *S. thermophilus*, εντοπίστηκε και στα τρία δείγματα Σφέλας με μεγαλύτερο ποσοστό στο δείγμα 1. Χρησιμοποιείται σε συνδυασμό με άλλα είδη βακτηρίων ως καλλιέργεια εκκίνησης σε τυριά και στο γιαούρτι [17]. Είναι ένα από τα ευρέως γνωστά εμπορικά σκευάσματα για τις καλλιέργειες εκκίνησης. Επίσης, ο *S. thermophilus* είναι το μόνο είδος του γένους που έχει χαρακτηριστεί ως ασφαλές για καλλιέργειες εκκίνησης από την Υπηρεσία Φαρμάκων και Τροφών (FDA) στις Η.Π.Α. και από την Ευρωπαϊκή Αρχή για την Ασφάλεια των Τροφίμων (EFSA) το 2007 [119]. Το γονιδίωμα του έχει μελετηθεί σε μεγάλο βαθμό και έχουν αποδειχθεί εξελικτικές αλλαγές που έχει υποστεί, οι οποίες οφείλονται στην προσαρμογή του στο γαλακτοκομικό περιβάλλον που είναι

ιδιαίτερα πλούσιο σε θρεπτικά συστατικά [120]. Κάποια από τα χαρακτηριστικά του γένους είναι η γρήγορη ανάπτυξη και η συμβολή του στη μείωση του pH που μπορεί να αλλάξει δραστικά την μικροβιακή σύνθεση του τυριού [102].

Η παρουσία ειδών του γένους *Staphylococcus*, και πιο συγκεκριμένα το είδος *Staphylococcus aureus*, συντελεί κοινή αιτία δηλητηριάσεων από τρόφιμα. Στις περισσότερες περιπτώσεις αναπτύσσεται σε τυριά που παρήχθησαν από νωπό γάλα ή τυριά που είχαν ελλιπή θερμική επεξεργασία [121]. Αναπτύσσεται σε μεγάλο εύρος pH 4.5-9.3 [122]. Υψηλές συγκεντρώσεις του είδους μπορούν να οδηγήσουν στην παραγωγή εντεροτοξίνης, η οποία δεν επηρεάζεται από τις συνθήκες αποθήκευσης του τυριού [123]. Στην ανάλυση βρέθηκε σε μικρή ποσότητα στο δείγμα 3 Σφέλας, το είδος *S. saprophyticus*. Το είδος αυτό του γένους *Staphylococcus*, έχει μελετηθεί ως προς την παρουσία του σε νωπό γάλα, τελικό τυροκομικό προϊόν Γαλλικής προέλευσης και κλινικά δείγματα [124, 125].

Τέλος, στα δείγματα Σφέλας, δεν παρουσιάζονται μολύνσεις από παθογόνους μικροοργανισμούς σε μεγάλη ποσότητα, πέρα από την παρουσία ειδών *Kurthia* και *Staphylococcus*.

Συνεχίζοντας, για τα τρία δείγματα Χαλίτζι υπάρχει δυνατότητα σύγκρισης των αποτελεσμάτων με προγενέστερες μελέτες στις οποίες υλοποιήθηκαν αναλύσεις των περιοχών 16S rDNA και ITS, για τη διερεύνηση της σύνθεσης του μικροβιώματος [126]. Αρχικά πρέπει να αναφέρουμε τις διαφοροποιήσεις που παρατηρήθηκαν μεταξύ των τριών δειγμάτων της μελέτης, και συγκεκριμένα των δειγμάτων 1 και 3 με το δείγμα 2. Η μικροβιακή σύνθεση του δείγματος 2 Χαλίτζι παρουσιάζει μειωμένη ποικιλομορφία. Κυρίαρχο είδος είναι ο *L. lactis* με ποσοστό μεγαλύτερο του 85%. Κοινό αποτέλεσμα έδωσαν και οι τρεις πλατφόρμες και τα επιμέρους εργαλεία που χρησιμοποιήθηκαν για την ανάλυση. Ακολουθεί το είδος *Lb. plantarum* και *E. faecium* όπως ήταν αναμενόμενο βάσει της μικροβιακής σύνθεσης των λευκών τυριών άλμης. Το είδος *E. faecium* έχει παρουσιαστεί επίσης σε υψηλούς σχετικά πληθυσμούς στο κυπριακό ΠΟΠ τυρί Χαλούμι [34, 127]. Αυτό έρχεται σε αντίθεση με τα αποτελέσματα της έρευνας των *Parademas et al. 2019*, καθώς στη μελέτη

αναφέρεται η παρουσία των γενών *Leuconostoc* και *Pseudomonas*, τα οποία δεν εμφανίζονται στα αποτελέσματα των δειγμάτων Χαλίτζι αυτής της μελέτης. Το γένος *Lactobacillus* που αναφέρεται στη μελέτη των *Parademas et al. 2019*, εντοπίστηκε στα δείγματα 1 - 2 σε πολύ χαμηλούς πληθυσμούς, ενώ βρίσκεται σε μεγαλύτερο πληθυσμό στο δείγμα 3 ($\approx 35\%$). Παρουσιάζονται επίσης, σε πολύ μικρά ποσοστά παθογόνα βακτήρια του γένους *Enterobacter*. Στο δείγμα 1 Χαλίτζι εντοπίστηκαν σε μεγάλη συγκέντρωση το είδος *L. lactis* και να έπεται το είδος *E. faecalis* για τα οποία υπάρχει αναφορά σε προηγούμενες μελέτες [1]. Άλλα είδη που εντοπίστηκαν στο δείγμα 1 με σειρά αφθονίας είναι του γένους *Staphylococcus* (*S. epidermis*, *S. saprophyticus*, *S. infantarius*, *S. scuri*, *S. xylosus*), *Lactobacillus* (*Lb. paracasei*, *Lb. acidipiscis*), *Klebsiella* (*K. pneumoniae*), *Tetragenococcus* (*T. halophilus*), *Raoultella sp.* και *Morganella morganii*.

Τα τελευταία γένη *Klebsiella* και *Morganella* ανήκουν στην οικογένεια των *Enterobacteriaceae*, το γένος *Raoultella* στην οικογένεια *Morganellaceae* και το γένος *Staphylococcus* στην οικογένεια *Staphylococcaceae*, τα οποία φανερώνουν προβλήματα υγιεινής κατά την τυροκόμηση. Τα ποσοστά των παθογόνων βακτηρίων που υπολογίστηκαν είναι μεγαλύτερα από τα ποσοστά που αναφέρονται σε προηγούμενες μελέτες. Το ίδιο ισχύει και για το δείγμα 3 Χαλίτζι.

Το δείγμα 3 εμφανίζει αυξημένη ποικιλομορφία στη σύνθεσή του, όπως και στην περίπτωση του δείγματος 2, με κάποιες διαφοροποιήσεις. Τα κυρίαρχα γένη που υπολογίστηκαν είναι *Streptococcus*, *Lactobacillus* και *Lactococcus*. Τα επιμέρους είδη που κυριαρχούν είναι *S. gallolyticus*, *S. macedonicus*, *Lb. paraplantarum*, *Lb. paracasei* και *L. lactis*. Τα είδη αυτά έρχονται σε συμφωνία με τη μελέτη των *Paradema et al. 2019*. Τέλος, παρουσία παθογόνων βρίσκεται στο δείγμα 3, και συγκεκριμένα της οικογένειας *Enterobacteriaceae*, όπως και στην περίπτωση του δείγματος 1. Συμπεραίνουμε λοιπόν, προβλήματα με τις συνθήκες υγιεινής στην παρασκευή του τυριού Χαλίτζι. Η παρουσία coliforms στα δείγματα φανερώνουν την ελλιπή θερμική επεξεργασία του γάλακτος ή τη μη ορθή οξίνιση του τυροπήγματος [1]. Τέλος, πρέπει να γίνει αναφορά στο είδος *Str. gallolyticus* (*S. gallolyticus* subsp.

macedonicus). Η παρουσία του σε γαλακτοκομικά προϊόντα που έχουν υποστεί ζύμωση έχει διερευνηθεί ιδιαίτερα, καθώς μελέτες έχουν αποδείξει παρουσία παθογένειας σε στελέχη του είδους. Ο βαθμός συσχέτισής του με ασθένειες, δεν έχει εξακριβωθεί για την περίπτωση της παρουσίας του σε τυροκομικά προϊόντα, αλλά η στενή φυλογενετική του σχέση με μέλη του γένους που ανήκουν στην κατηγορία των παθογόνων είναι αδιαμφισβήτητη [128].

Στην ανάλυση των ITS περιοχών για τη διαπίστωση της σύνθεσης των μυκητών και των ζυμών, τα αποτελέσματα φανερώνουν διαφοροποίηση μεταξύ των δύο δειγμάτων και απουσία αυξημένης ποικιλομορφίας. Αναφερόμαστε στο δείγμα 2 και 3 του τυριού Χαλίτζι. Το μεγαλύτερο μέρος των μυκητών που αναπτύσσονται κατά την παρασκευή των τυριών, καταστρέφεται κατά την παστερίωση. Η ανθεκτικότητα σε υψηλές συγκεντρώσεις άλατος και η επιβίωσής τους σε ένα ευρύ φάσμα pH, μπορεί να βοηθήσει στην ανάπτυξη τους κατά τη διάρκεια της ωρίμανσης. Η πηγή ενέργειας που χρησιμοποιούν προέρχεται από τη διάσπαση του γαλακτικού οξέος, γεγονός που αυξάνει το pH και ευνοεί την ανάπτυξη ανεπιθύμητων βακτηρίων [9].

Οι μύκητες και οι ζύμες μπορεί να προκαλέσουν αλλοιώσεις όπως η ανεπιθύμητη οσμή και γεύση, λόγω της αύξησης στη διάσπαση πρωτεϊνών και λιπών. Δηλαδή αύξηση της παραγωγής αλκοολών, αλδευδών και εστέρων. Επίσης, δρουν ανασταλτικά ως προς την ανάπτυξη των οξυγαλακτικών βακτηρίων εκκίνησης. Η αυξημένη κατανάλωση σακχάρων, οδηγεί στην παραγωγή διοξειδίου του άνθρακα. Το αέριο αυτό, δημιουργεί οπές στη μάζα του τυριού με αλλοίωση της εξωτερικής και εσωτερικής σύστασής του. Τέλος τα προϊόντα που παράγουν, μπορούν να οδηγήσουν στην υποβάθμιση της ποιότητας των γαλακτοκομικών προϊόντων. Παραδείγματος χάρη, ορισμένοι μύκητες παράγουν βιογενείς αμίνες, οι οποίες με τη σειρά τους προκαλούν αλλαγή στη γεύση των τυριών [129].

Το κυρίαρχο γένος ζύμης που εμφανίστηκε στο δείγμα 2 είναι το *Debaryomyces* με ποσοστό άνω του 74%, ενώ στην περίπτωση του δείγματος 3, το γένος που κατέχει εξίσου άνω του 74% ποσοστό είναι το *Kluyveromyces*. Τα ευρήματα της ανάλυσης των ITS περιοχών, συμφωνούν με τα αποτελέσματα της μελέτης

των Parademas et. al 2019, με διαφοροποίηση να αποτυπώνεται στα ποσοστά της παρουσίας των ζυμών εντός των δειγμάτων. Η αυξημένη ανάπτυξη ζυμών μπορεί να προκαλέσει αλλαγή στις οργανοληπτικές ιδιότητες των τυριών, με χαρακτηριστικό παράδειγμα την διαφοροποίηση στην οσμή. Επίσης, στην περίπτωση των λευκών τυριών άλμης, οι ζύμες του γένους *Kluyveromyces* μπορούν να προκαλέσουν εσωτερική πίεση στο σκεύος αποθήκευσης των τυριών δημιουργώντας προβλήματα. Αυτό συμβαίνει λόγω της ζύμωσης της λακτόζης και της παραγωγής αερίων [34] [130].

Οι δύο ζύμες που εντοπίστηκαν στα δείγματα Χαλίτζι, συμβάλλουν συχνά στη βελτίωση της γεύσης του τυριού. Όμως η αυξημένη ανάπτυξη της ζύμης *Debaryomyces hansenii* μπορεί να προκαλέσει την ανάπτυξη μη επιθυμητών κηλίδων στην επιφάνεια του τυριού [131].

Ανάπτυξη των δύο ζυμών έχει αναφερθεί και σε άλλα λευκά τυριά άλμης κατά το παρελθόν [99]. Επίσης το είδος *Kluyveromyces lactis* έχει αναφερθεί ότι χρησιμοποιείται για την μετατροπή της λακτόζης σε γαλακτικό οξύ, και αξιοποιείται στη παραγωγή τυροκομικών προϊόντων. Επόμενο γένος ζύμης που υπολογίστηκε στα δείγματα Χαλίτζι είναι η *Candida* σε ποσοστό 22% για το δείγμα 2, και 9% για το δείγμα 3. Το γένος καλείται ως ευκαιριακά παθογόνο, αλλά έχει αναφερθεί η παρουσία του σε λευκά τυριά άλμης με μικρές ποσότητες, χωρίς να προκαλεί δυσμενείς συνθήκες στις οργανοληπτικές ιδιότητες των τυριών [97].

4.4 Επέκταση Ανάλυσης – Μελλοντική Έρευνα

Συμπερασματικά, η βιοπληροφορική ανάλυση περιβαλλοντικών δειγμάτων έχει συνεισφέρει σε αξιοσημείωτο βαθμό στην καλύτερη κατανόηση των μικροβιακών κοινοτήτων. Τα εργαλεία που έχουν αναπτυχθεί, βοηθούν στην αποσαφήνιση της σύστασης των οικοσυστημάτων και των λειτουργικών ιδιοτήτων τους. Απαραίτητη είναι η εξέλιξη και ο εμπλουτισμός των εργαλείων για την πιο γρήγορη ανάλυση των δεδομένων, καθώς οι αυξημένες ανάγκες υπολογιστικής δύναμης, δυσχεραίνουν σε πολλές περιπτώσεις τη ροή εργασίας. Στη μελέτη αναλύθηκαν για πρώτη φορά δείγματα των τυριών

Σφέλα και Χαλίτζι με τεχνικές Amplicon και shotgun NGS. Τα περισσότερα αποτελέσματα συμφωνούν με προγενέστερες βιβλιογραφικές καταχωρήσεις για το μικροβίωμα των λευκών τυριών άλμης. Επιμέρους διαφοροποιήσεις αναλύθηκαν εκτενέστερα στις προηγούμενες παραγράφους, με τις κυριότερες να συνοψίζονται ως εξής : Διαφοροποίηση μεταξύ δειγμάτων 1 – 2 Σφέλας από το δείγμα 3, ενδεχομένως λόγω των διαφορετικών συνθηκών τυροκόμησης. Αυξημένη ποικιλομορφία στη σύνθεση του μικροβιώματος στο δείγμα 1 και 3, με την παρουσία μη επιθυμητών παθογόνων βακτηρίων στο δείγμα 3. Συμφωνία των αποτελεσμάτων με προγενέστερες μελέτες για το τυρί Χαλίτζι. Αυξημένη παρουσία παθογόνων βακτηρίων σε όλα τα δείγματα Χαλίτζι, που προέρχεται από ελλιπή θέρμανση του νωπού γάλακτος και προβλήματα υγιεινής κατά την παρασκευή. Διαφοροποιήσεις στα επιμέρους είδη βακτηρίων που αναπτύσσονται στη Σφέλα και στο Χαλίτζι από άλλα λευκά τυριά άλμης, όπως Φέτα, Μπάτζος και Χαλούμι. Τα αποτελέσματα της ανάλυσης των λειτουργικών ιδιοτήτων του μικροβιακού συνόλου των δειγμάτων, απέδωσαν τις βασικές λειτουργίες του μικροβιώματος, όπως το μεταβολισμό των αμινοξικών καταλοίπων, των υδατανθράκων και των νουκλεϊκών οξέων. Αυτές οι λειτουργίες έχουν ζωτική σημασία για την επιβίωση των μικροοργανισμών και του συστήματος αλληλεξάρτησης των μικροβίων που αναπτύσσονται. Μελλοντική έρευνα θα μπορούσε να πραγματοποιηθεί για τη διερεύνηση των γονιδίων που περιέχονται στα βακτήρια με το μεγαλύτερο ποσοστό αφθονίας. Επίσης, μέσω εργαστηριακών πειραμάτων θα μπορούσε να διερευνηθεί η αλληλεπίδραση και αλληλεξάρτηση των επιμέρους βακτηρίων που υπολογίσθηκαν με βιοπληροφορικές μεθόδους. Η επέκταση στην έρευνα του μικροβιώματος θα μπορέσουν να αποδώσουν μια ολική αλλά και ταυτόχρονα λεπτομερειακή εικόνα της σύνθεσης τους. Σε εξέλιξη της γενικότερη έρευνας του μικροβιώματος των τυροκομικών προϊόντων, η δημιουργία εξειδικευμένων βάσεων δεδομένων που θα εμπεριέχονται πληροφορίες της μικροβιακής σύστασης των τελικών προϊόντων, σε διαφορετικές ποικιλίες τυριών, θα μπορέσει να συμβάλει στην εξέλιξη της βιομηχανικής παραγωγής και τη δημιουργία νέων ποικιλιών. Όπως επίσης την ανάπτυξη του τομέα της μεταβολικής μηχανικής και την αποσαφήνιση των γονιδιακών χαρακτηριστικών στα μικροβιώματα.

REFERENCES

1. Τσιγκριμάνη Μ: **Μελέτη της μικροβιακής χλωρίδας δειγμάτων του Κυπριακού τυριού Χαλίτζι**. Αθήνα: Γεωπονικό Πανεπιστήμιο Αθηνών; 2019.
2. Fox PF, Guinee TP, Cogan TM, McSweeney PLH: **Fundamentals of Cheese Science**, Second Edition edn. Springer New York 2017: Springer, Boston, MA; 2000.
3. Salque M, Bogucki PI, Pyzel J, Sobkowiak-Tabaka I, Grygiel R, Szmyt M, Evershed RP: **Earliest evidence for cheese making in the sixth millennium BC in northern Europe**. *Nature* 2013, **493**(7433):522-525.
4. Tamime AY, Robinson RK: **Feta and Related Cheeses**, 1st Edition edn. Sawston, Cambridge: Woodhead Publishing, Elsevier; 1991.
5. Γ. Μ: **Λεξικό της νέας ελληνικής γλώσσας**. Αθήνα; 1998.
6. World Health Organization F: **STANDARDS FOR CREAM CHEESE**. In: *International Food Standards*. Edited by Nations FaAOotU: FAO/WHO/UNU Expert Consultation; 1973.
7. Καμιναρίδης Σ, Μοάτσου Γ: **Γαλακτοκομία**, 1st Edition edn. Αθήνα: Έμβρυο Publications; 2009.
8. D'Amico DJ, Druart MJ, Donnelly CW: **60-day aging requirement does not ensure safety of surface-mold-ripened soft cheeses manufactured from raw or pasteurized milk when *Listeria monocytogenes* is introduced as a postprocessing contaminant**. *Journal of food protection* 2008, **71**(8):1563-1571.
9. Ανυφαντάκης Ε: **Τυροκομία: Χημεία, Φυσικοχημεία, Μικροβιολογία**. Αθήνα: Σταμούλη Α.Ε.; 2004.
10. Ζερφυρίδης Γ: **Technology of Milk Products**: Γιαχούδη Γιαπούλη; 2001.
11. Quigley L, McCarthy R, O'Sullivan O, Beresford TP, Fitzgerald GF, Ross RP, Stanton C, Cotter PD: **The microbial content of raw and**

- pasteurized cow milk as determined by molecular approaches.** *Journal of dairy science* 2013, **96**(8):4928-4937.
12. Fretin M, Martin B, Rifa E, Isabelle VM, Pomies D, Ferlay A, Montel MC, Delbes C: **Bacterial community assembly from cow teat skin to ripened cheeses is influenced by grazing systems.** *Scientific reports* 2018, **8**(1):200.
 13. De Filippis F, Vannini L, La Storia A, Laghi L, Piombino P, Stellato G, Serrazanetti DI, Gozzi G, Turrone S, Ferrocino I *et al*: **The same microbiota and a potentially discriminant metabolome in the saliva of omnivore, ovo-lacto-vegetarian and Vegan individuals.** *PloS one* 2014, **9**(11):e112373.
 14. Quigley L, O'Sullivan O, Beresford TP, Ross RP, Fitzgerald GF, Cotter PD: **High-throughput sequencing for detection of subpopulations of bacteria not previously associated with artisanal cheeses.** *Applied and environmental microbiology* 2012, **78**(16):5717-5723.
 15. O'Sullivan O, Cotter PD: **Chapter 12 - Microbiota of Raw Milk and Raw Milk Cheeses.** In: *Cheese (Fourth Edition)*. Edited by McSweeney PLH, Fox PF, Cotter PD, Everett DW. San Diego: Academic Press; 2017: 301-316.
 16. Delcenserie V, Taminiau B, Delhalle L, Nezer C, Doyen P, Crevecoeur S, Roussey D, Korsak N, Daube G: **Microbiota characterization of a Belgian protected designation of origin cheese, Herve cheese, using metagenomic analysis.** *Journal of dairy science* 2014, **97**(10):6046-6056.
 17. Beresford T, Williams A: **The Microbiology of Cheese Ripening.** In: *Cheese: Chemistry, Physics and Microbiology*. Edited by Fox PF, McSweeney PLH, Cogan TM, Guinee TP, vol. 1: Academic Press; 2004: 287-317.
 18. Fox PF, Cogan TM: **Factors that Affect the Quality of Cheese.** In: *Cheese: Chemistry, Physics and Microbiology*. Edited by Fox PF,

- McSweeney PLH, Cogan TM, Guinee TP, vol. 1: Academic Press; 2004: 583-608.
19. Moatsou G, Govaris A: **White brined cheeses: A diachronic exploitation of small ruminants milk in Greece.** *Small Ruminant Research* 2011, **101**(1):113-121.
 20. Orla-Jensen S, Orla-Jensen AD, Kjaer A: **On the ensiling of lucerne by means of lactic acid fermentation.** *Antonie van Leeuwenhoek* 1947, **12**(1-4):97-114.
 21. Gutiérrez-Cortés C, Suarez H, Buitrago G, Nero LA, Todorov SD: **Enhanced Bacteriocin Production by *Pediococcus pentosaceus* 147 in Co-culture With *Lactobacillus plantarum* LE27 on Cheese Whey Broth.** *Front Microbiol* 2018, **9**:2952-2952.
 22. König H, Fröhlich J: **Lactic Acid Bacteria.** In: *Biology of Microorganisms on Grapes, in Must and in Wine.* Edited by König H, Unden G, Fröhlich J. Cham: Springer International Publishing; 2017: 3-41.
 23. Βασιλειάδης ΘΒ: **Μελέτη της Ανθεκτικότητας Προβιοτικών Βακτηρίων και της Χρήσης τους στα Τρόφιμα μέσω Εδώδιμων Μεμβρανών.** Θεσσαλονίκη: Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης; 2017.
 24. Amann RI, Ludwig W, Schleifer KH: **Phylogenetic identification and in situ detection of individual microbial cells without cultivation.** *Microbiological reviews* 1995, **59**(1):143-169.
 25. Gänzle MG: **Lactic metabolism revisited: metabolism of lactic acid bacteria in food fermentations and food spoilage.** *Current Opinion in Food Science* 2015, **2**:106-117.
 26. Naidu AS, Bidlack WR, Clemens RA: **Probiotic Spectra of Lactic Acid Bacteria (LAB).** *Critical Reviews in Food Science and Nutrition* 1999, **39**(1):13-126.
 27. Rattanachaikunsopon P, Phumkhachorn P: **Assessment of factors influencing antimicrobial activity of carvacrol and cymene against**

- Vibrio cholerae in food.** *Journal of Bioscience and Bioengineering* 2010, **110**(5):614-619.
28. Soomro AH, Masud T. , Anwaar K: **Role of Lactic Acid Bacteria (LAB) in Food Preservation and Human Health - A Review.** *Pakistan Journal of Nutrition* 2002.
 29. Frank JF, Marth EH: **Inhibition of Enteropathogenic Escherichia coli by Homofermentative Lactic Acid Bacteria in Skimmilk: I. Comparison of Strains of Escherichia coli.** *Journal of food protection* 1977, **40**(11):749-753.
 30. Edwards PR, Ewing WH: **Identification of Enterobacteriaceae;** 1972.
 31. Ryser ET, Maisnier-Patin S, Gratadoux JJ, Richard J: **Isolation and identification of cheese-smear bacteria inhibitory to Listeria spp.** *International Journal of Food Microbiology* 1994, **21**(3):237-246.
 32. Τζανετάκη Λ-: **Μικροβιολογία Γάλακτος.** Αθήνα: Αιβάζης; 1990.
 33. Tamime A: **Production of Kefir, Koumiss and other Related Products.** In: *Fermented Milks.* Edited by Tamime AY; 2006: 174-216.
 34. Bintsis T, Papademas P: **Microbiological quality of white-brined cheeses: a review.** *International Journal of Dairy Technology* 2002, **55**(3):113-120.
 35. Pevsner J: **Bioinformatics and Functional Genomics;** Wiley Publishing; 2009.
 36. Head SR, Ordoukhanian P, Salomon DR: **Next Generation Sequencing.** Springer Science+Business Media, LLC: Humana Press; 2018.
 37. Feng K, Costa J, Edwards JS: **Next-generation sequencing library construction on a surface.** *BMC Genomics* 2018, **19**(1):416.
 38. Chen C-Y: **DNA polymerases drive DNA sequencing-by-synthesis technologies: both past and present.** *Front Microbiol* 2014, **5**(305).
 39. Bragg L. TGW: **Metagenomics Using Next-Generation Sequencing;** Humana Press, Totowa, NJ; 2014.

40. Mardis ER: **A decade's perspective on DNA sequencing technology.** *Nature* 2011, **470**(7333):198-203.
41. Voelkerding KV, Dames SA, Durtschi JD: **Next-generation sequencing: from basic research to diagnostics.** *Clinical chemistry* 2009, **55**(4):641-658.
42. Schloss PD, Handelsman J: **Biotechnological prospects from metagenomics.** *Current Opinion in Biotechnology* 2003, **14**(3):303-310.
43. Handelsman J: **Metagenomics: Application of Genomics to Uncultured Microorganisms.** *Microbiology and Molecular Biology Reviews* 2004, **68**(4):669-685.
44. Simon C, Daniel R: **Metagenomic Analyses: Past and Future Trends.** *Applied and environmental microbiology* 2011, **77**(4):1153-1161.
45. Jovel J, Patterson J, Wang W, Hotte N, O'Keefe S, Mitchel T, Perry T, Kao D, Mason AL, Madsen KL *et al*: **Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics.** *Front Microbiol* 2016, **7**(459).
46. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO: **Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies.** *Nucleic Acids Research* 2012, **41**(1):e1-e1.
47. Bokulich NA, Mills DA: **Improved internal transcribed spacer (ITS) primer selection enables quantitative, ultra-high-throughput fungal community profiling.** *Applied and environmental microbiology* 2013:AEM.03870-03812.
48. Schloss PD, Handelsman J: **A statistical toolbox for metagenomics: assessing functional diversity in microbial communities.** *BMC Bioinformatics* 2008, **9**(1):34.
49. Sharpton TJ: **An introduction to the analysis of shotgun metagenomic data.** *Frontiers in Plant Science* 2014, **5**(209).
50. Mende DR, Waller AS, Sunagawa S, Järvelin AI, Chan MM, Arumugam M, Raes J, Bork P: **Assessment of Metagenomic Assembly Using**

- Simulated Next Generation Sequencing Data.** *PloS one* 2012, 7(2):e31386.
51. Mavromatis K, Ivanova N, Barry K, Shapiro H, Goltsman E, McHardy AC, Rigoutsos I, Salamov A, Korzeniewski F, Land M *et al*: **Use of simulated data sets to evaluate the fidelity of metagenomic processing methods.** *Nature Methods* 2007, 4(6):495-500.
 52. Rudi K, Zimonja M, Trosvik P, Næs T: **Use of multivariate statistics for 16S rRNA gene analysis of microbial communities.** *International Journal of Food Microbiology* 2007, 120(1):95-99.
 53. Woese CR, Fox GE: **Phylogenetic structure of the prokaryotic domain: The primary kingdoms.** *Proceedings of the National Academy of Sciences* 1977, 74(11):5088-5090.
 54. Pitcher DG, Saunders NA, Owen RJ: **Rapid extraction of bacterial genomic DNA with guanidium thiocyanate.** *Letters in Applied Microbiology* 1989, 8(4):151-156.
 55. Low L, Tammi M: **Bioinformatics : a practical handbook of next generation sequencing and its applications.** New Jersey: World Scientific; 2017.
 56. Brown SM: **Next-generation DNA sequencing informatics;** 2015.
 57. Idury RM, Waterman MS: **A new algorithm for DNA sequence assembly.** *Journal of computational biology : a journal of computational molecular cell biology* 1995, 2(2):291-306.
 58. MacLean D, Jones JDG, Studholme DJ: **Application of 'next-generation' sequencing technologies to microbial genetics.** *Nature Reviews Microbiology* 2009, 7(4):96-97.
 59. Ratan A, Olson TL, Loughran TP, Miller W: **Identification of indels in next-generation sequencing data.** *BMC Bioinformatics* 2015, 16(1):42.
 60. Miller JR, Koren S, Sutton G: **Assembly algorithms for next-generation sequencing data.** *Genomics* 2010, 95(6):315-327.

61. Sharpton TJ: **An introduction to the analysis of shotgun metagenomic data.** *Front Plant Sci* 2014, **5**:209.
62. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S *et al*: **KBase: The United States Department of Energy Systems Biology Knowledgebase.** *Nature Biotechnology* 2018, **36**(7):566-569.
63. Peng Y, Leung HCM, Yiu SM, Chin FYL: **IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth.** *Bioinformatics (Oxford, England)* 2012, **28**(11):1420-1428.
64. Peng Y, Leung HCM, Yiu SM, Chin FYL: **IDBA - A Practical Iterative de Bruijn Graph De Novo Assembler.** In: *2010; Berlin, Heidelberg.* Springer Berlin Heidelberg: 426-440.
65. Wu YW, Simmons BA, Singer SW: **MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets.** *Bioinformatics (Oxford, England)* 2016, **32**(4):605-607.
66. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW: **CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes.** *Genome Res* 2015, **25**(7):1043-1055.
67. Menzel P, Ng KL, Krogh A: **Fast and sensitive taxonomic classification for metagenomics with Kaiju.** *Nature communications* 2016, **7**:11257.
68. Freitas TA, Li PE, Scholz MB, Chain PS: **Accurate read-based metagenome characterization using a hierarchical suite of unique signatures.** *Nucleic Acids Res* 2015, **43**(10):e69.
69. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH: **GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database.** *Bioinformatics (Oxford, England)* 2019, **36**(6):1925-1927.
70. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarszewski A, Chaumeil P-A, Hugenholtz P: **A standardized bacterial taxonomy based on**

- genome phylogeny substantially revises the tree of life.** *Nature Biotechnology* 2018, **36**(10):996-1004.
71. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S: **High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries.** *Nature communications* 2018, **9**(1):5114.
 72. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ: **Prodigal: prokaryotic gene recognition and translation initiation site identification.** *BMC Bioinformatics* 2010, **11**:119.
 73. Matsen FA, Kodner RB, Armbrust EV: **pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree.** *BMC Bioinformatics* 2010, **11**:538.
 74. Price MN, Dehal PS, Arkin AP: **FastTree 2--approximately maximum-likelihood trees for large alignments.** *PloS one* 2010, **5**(3):e9490-e9490.
 75. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM: **Mash: fast genome and metagenome distance estimation using MinHash.** *Genome Biology* 2016, **17**(1):132.
 76. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A *et al*: **The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes.** *BMC Bioinformatics* 2008, **9**(1):386.
 77. Keegan KP, Glass EM, Meyer F: **MG-RAST, a Metagenomics Service for Analysis of Microbial Community Structure and Function.** *Methods Mol Biol* 2016, **1399**:207-233.
 78. Cox MP, Peterson DA, Biggs PJ: **SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data.** *BMC Bioinformatics* 2010, **11**(1):485.
 79. Keegan KP, Trimble WL, Wilkening J, Wilke A, Harrison T, D'Souza M, Meyer F: **A Platform-Independent Method for Detecting Errors in**

- Metagenomic Sequencing Data: DRISSE.** *PLoS Computational Biology* 2012, **8**(6):e1002541.
80. Rho M, Tang H, Ye Y: **FragGeneScan: predicting genes in short and error-prone reads.** *Nucleic Acids Res* 2010, **38**(20):e191.
81. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO: **The SILVA ribosomal RNA gene database project: improved data processing and web-based tools.** *Nucleic acids research* 2013, **41**(Database issue):D590-D596.
82. Kent WJ: **BLAT--the BLAST-like alignment tool.** *Genome Res* 2002, **12**(4):656-664.
83. Wilke A, Harrison T, Wilkening J, Field D, Glass EM, Kyrpides N, Mavrommatis K, Meyer F: **The M5nr: a novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools.** *BMC Bioinformatics* 2012, **13**(1):141.
84. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M *et al*: **The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST).** *Nucleic acids research* 2014, **42**(Database issue):D206-D214.
85. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM *et al*: **The Ribosomal Database Project: improved alignments and new tools for rRNA analysis.** *Nucleic Acids Res* 2009, **37**(Database issue):D141-145.
86. Ondov BD, Bergman NH, Phillippy AM: **Interactive metagenomic visualization in a Web browser.** *BMC Bioinformatics* 2011, **12**(1):385.
87. Ondov BD, Bergman NH, Phillippy AM: **Krona: Interactive Metagenomic Visualization in a Web Browser.** In: *Encyclopedia of Metagenomics*. Edited by Nelson KE. New York, NY: Springer New York; 2013: 1-8.
88. Pruitt KD, Tatusova T, Maglott DR: **NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes,**

- transcripts and proteins.** *Nucleic acids research* 2007, **35**(Database issue):D61-D65.
89. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang H-Y, Cohoon M, de Crécy-Lagard V, Diaz N, Disz T, Edwards R *et al*: **The Subsystems Approach to Genome Annotation and its Use in the Project to Annotate 1000 Genomes.** *Nucleic Acids Research* 2005, **33**(17):5691-5702.
90. Tatusov RL, Galperin MY, Natale DA, Koonin EV: **The COG database: a tool for genome-scale analysis of protein functions and evolution.** *Nucleic acids research* 2000, **28**(1):33-36.
91. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen Lars J *et al*: **eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses.** *Nucleic Acids Research* 2018, **47**(D1):D309-D314.
92. Wood DE, Salzberg SL: **Kraken: ultrafast metagenomic sequence classification using exact alignments.** *Genome Biology* 2014, **15**(3):R46.
93. Fonseca GG, Heinzle E, Wittmann C, Gombert AK: **The yeast *Kluyveromyces marxianus* and its biotechnological potential.** *Applied microbiology and biotechnology* 2008, **79**(3):339-354.
94. Zoumpopoulou G, Tzouvanou A, Mavrogonatou E, Alexandraki V, Georgalaki M, Anastasiou R, Papadelli M, Manolopoulou E, Kazou M, Kletsas D *et al*: **Probiotic Features of Lactic Acid Bacteria Isolated from a Diverse Pool of Traditional Greek Dairy Products Regarding Specific Strain-Host Interactions.** *Probiotics and Antimicrobial Proteins* 2018, **10**(2):313-322.
95. Papadimitriou K, Zoumpopoulou G, Foligné B, Alexandraki V, Kazou M, Pot B, Tsakalidou E: **Discovering probiotic microorganisms: in vitro, in vivo, genetic and omics approaches.** *Front Microbiol* 2015, **6**:58.

96. Kent WJ: **BLAT--the BLAST-like alignment tool**. *Genome Res* 2002, **12**(4):656-664.
97. A. Y. Tamine RKR: **Feta and Related Cheeses**, vol. 2. Cambridge, England: Woodhead Publishing Limited; 1991.
98. Kehagias C, Koulouris S, Samona A, Malliou S, Koumoutsos G: **Effect of various starters on the quality of cheese in brine**. *Food Microbiology* 1995, **12**:413-420.
99. Abd El-Salam MH, Alichanidis E: **Cheese varieties ripened in brine**. In: *Cheese: Chemistry, Physics and Microbiology*. Edited by Fox PF, McSweeney PLH, Cogan TM, Guinee TP, vol. 2: Academic Press; 2004: 227-249.
100. Hynes E, Ogier J-C, Delacroix-Buchet A: **Proteolysis during ripening of miniature washed-curd cheeses manufactured with different strains of starter bacteria and a Lactobacillus plantarum adjunct culture**. *International Dairy Journal* 2001, **11**(8):587-597.
101. Nieto-Arribas P, Seseña S, Poveda JM, Palop L, Cabezas L: **Genotypic and technological characterization of Lactococcus lactis isolates involved in processing of artisanal Manchego cheese**. *Journal of applied microbiology* 2009, **107**(5):1505-1517.
102. C. A. Oluk OBK: **Traditional Cheeses from Selected Regions in Asia, Europe, and South America**: Bentham Science Publishers; 2020.
103. Uymaz B, Akçelik N, Yüksel Z: **Physicochemical and Microbiological Characterization of Protected Designation of Origin Ezine Cheese: Assessment of Non-starter Lactic Acid Bacterial Diversity with Antimicrobial Activity**. *Food Science of Animal Resources* 2019, **39**(5):804-819.
104. Milani E, Shahidi F, Mortazavi SA, Saedi M: **Isolation and Identification of Lactic Acid Bacteria in Kurdish Cheese during Ripening using 16S rRNA Gene Sequence Analysis**. *Journal of Food Processing and Preservation* 2017, **41**(4):e13009.

105. Sarantinopoulos P, Kalantzopoulos G, Tsakalidou E: **Effect of *Enterococcus faecium* on microbiological, physicochemical and sensory characteristics of Greek Feta cheese.** *International Journal of Food Microbiology* 2002, **76**(1):93-105.
106. Tzanetakis N, Vafopoulou-Mastrojiannaki A, Litopoulou-Tzanetaki E: **The quality of white-brined cheese from goat's milk made with different starters.** *Food Microbiology* 1995, **12**:55-63.
107. Manolopoulou E, Sarantinopoulos P, Zoidou E, Aktypis A, Moschopoulou E, Kandarakis IG, Anifantakis EM: **Evolution of microbial populations during traditional Feta cheese manufacture and ripening.** *International Journal of Food Microbiology* 2003, **82**(2):153-161.
108. Flórez AB, Ammor MS, Mayo B: **Identification of tet(M) in two *Lactococcus lactis* strains isolated from a Spanish traditional starter-free cheese made of raw milk and conjugative transfer of tetracycline resistance to lactococci and enterococci.** *International Journal of Food Microbiology* 2008, **121**(2):189-194.
109. Pedersen TB, Vogensen FK, Ardö Y: **Effect of heterofermentative lactic acid bacteria of DL-starters in initial ripening of semi-hard cheese.** *International Dairy Journal* 2016, **57**:72-79.
110. Zago M, Fornasari ME, Carminati D, Burns P, Suárez V, Vinderola G, Reinheimer J, Giraffa G: **Characterization and probiotic potential of *Lactobacillus plantarum* strains isolated from cheeses.** *Food Microbiology* 2011, **28**(5):1033-1040.
111. Bulut C, Gunes H, Okuklu B, Harsa S, Kilic S, Sevgi Coban H, Fazil Yenidunya A: **Homofermentative lactic acid bacteria of a traditional cheese, Comlek peyniri from Cappadocia region.** *Journal of Dairy Research* 2005, **72**(1):19-24.
112. Durlu-Ozkaya F, Xanthopoulos V, Tunail N, Litopoulou-Tzanetaki E: **Technologically important properties of lactic acid bacteria isolates from Beyaz cheese made from raw ewes' milk.** *Journal of applied microbiology* 2001, **91**(5):861-870.

113. Frantzen CA, Kot W, Pedersen TB, Ardö YM, Broadbent JR, Neve H, Hansen LH, Dal Bello F, Østlie HM, Kleppen HP *et al*: **Genomic Characterization of Dairy Associated Leuconostoc Species and Diversity of Leuconostocs in Undefined Mixed Mesophilic Starter Cultures.** *Front Microbiol* 2017, **8**(132).
114. Udomsil N, Rodtong S, Choi YJ, Hua Y, Yongsawatdigul J: **Use of Tetragenococcus halophilus as a starter culture for flavor improvement in fish sauce fermentation.** *Journal of agricultural and food chemistry* 2011, **59**(15):8401-8408.
115. Morales F, Morales JI, Hernández CH, Hernández-Sánchez H: **Isolation and Partial Characterization of Halotolerant Lactic Acid Bacteria from Two Mexican Cheeses.** *Applied Biochemistry and Biotechnology* 2011, **164**(6):889-905.
116. Kamboj K, Vasquez A, Balada-Llasat J-M: **Identification and significance of Weissella species infections.** *Front Microbiol* 2015, **6**:1204-1204.
117. Finglas LTA: **Encyclopedia of Food Sciences and Nutrition**, 2nd edn. Baltimore, Maryland 21205-2179, USA: Elsevier Science Ltd; 2003.
118. Cohen J, Powderly WG, Opal SM: **Infectious Diseases:** Elsevier Health Sciences UK; 2016.
119. Alexandraki V, Kazou M, Blom J, Pot B, Papadimitriou K, Tsakalidou E: **Comparative Genomics of Streptococcus thermophilus Support Important Traits Concerning the Evolution, Biology and Technological Properties of the Species.** *Front Microbiol* 2019, **10**:2916.
120. Goh YJ, Goin C, O'Flaherty S, Altermann E, Hutkins R: **Specialized adaptation of a lactic acid bacterium to the milk environment: the comparative genomics of Streptococcus thermophilus LMD-9.** *Microbial Cell Factories* 2011, **10**(1):S22.
121. Jöhler S, Macori G, Bellio A, Acutis PL, Gallina S, Decastelli L: **Short communication: Characterization of Staphylococcus aureus isolated**

- along the raw milk cheese production process in artisan dairies in Italy. *Journal of dairy science* 2018, **101**(4):2915-2920.**
122. Minor TE, Marth EH: **FATE OF STAPHYLOCOCCUS AUREUS IN CULTURED BUTTERMILK, SOUR CREAM, AND YOGURT DURING STORAGE**. *Journal of Milk and Food Technology* 1972, **35**(5):302-306.
123. Tamime A: **Brined Cheeses**. Dairy Science and Technology Consultant Ayr, UK: Blackwell Publishing; 2006.
124. Coton E, Desmonts MH, Leroy S, Coton M, Jamet E, Christieans S, Donnio PY, Lebert I, Talon R: **Biodiversity of coagulase-negative Staphylococci in French cheeses, dry fermented sausages, processing environments and clinical samples.** *Int J Food Microbiol* 2010, **137**(2-3):221-229.
125. Widerström M, Wiström J, Sjöstedt A, Monsen T: **Coagulase-negative staphylococci: update on the molecular epidemiology and clinical presentation, with a focus on Staphylococcus epidermidis and Staphylococcus saprophyticus.** *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology* 2012, **31**(1):7-20.
126. Papademas P, Aspri M, Mariou M, Dowd SE, Kazou M, Tsakalidou E: **Conventional and omics approaches shed light on Halitzia cheese, a long-forgotten white-brined cheese from Cyprus.** *International Dairy Journal* 2019, **98**:72-83.
127. Kirilov N, Dimov S, Dalgalarrrondo M, Ignatova T, Kambarev S, Stoyanovski S, Danova S, Iliev I, Haertlé T, Chobert J-M *et al*: **Characterization of enterococci isolated from homemade Bulgarian cheeses and katuk.** *European Food Research and Technology* 2011, **233**(6):1029-1040.
128. Papadimitriou K: **Novel insight into the pathogenicity of Streptococcus gallolyticus subsp. gallolyticus belonging to the Streptococcus bovis/Streptococcus equinus complex.** *Virulence* 2018, **9**(1):662-665.

129. Fröhlich-Wyder M-T, Arias-Roth E, Jakob E: **Cheese yeasts**. *Yeast* 2019, **36**(3):129-141.
130. Bintsis T, Robinson RK: **A study of the effects of adjunct cultures on the aroma compounds of Feta-type cheese**. *Food Chemistry* 2004, **88**(3):435-441.
131. Donnelly CW: **Cheese and Microbes**. Hoboken, New Jersey: John Wiley & Sons, 2014; 2014.