



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΤΜΗΜΑ ΒΙΟΧΗΜΕΙΑΣ ΚΑΙ ΒΙΟΤΕΧΝΟΛΟΓΙΑΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

ΝΙΚΟΛΑΙΔΗΣ ΜΑΡΙΟΣ

ΕΦΑΡΜΟΓΕΣ ΜΟΡΙΑΚΗΣ ΒΙΟΛΟΓΙΑΣ

ΜΟΡΙΑΚΗ ΓΕΝΕΤΙΚΗ

ΔΙΑΓΝΩΣΤΙΚΟΙ ΔΕΙΚΤΕΣ

ΛΑΡΙΣΑ 2021

**Φυλογενωμική και συγκριτική γονιδιωματική ανάλυση της
εξελικτικής ομάδας των βακίλλων *B. altitudinis* - *B. pumilus*
- *B. safensis* με βιοπληροφορικές μεθόδους**

**Phylogenomic and comparative genomic analysis of the
bacillus evolutionary group *B. altitudinis* - *B. pumilus* - *B.*
safensis with bioinformatic methods**

Η παρούσα διπλωματική εργασία εκπονήθηκε στο εργαστήριο Βιοπληροφορικής, του τμήματος Βιοχημείας και Βιοτεχνολογίας, Σχολής Επιστημών Υγείας του Πανεπιστημίου Θεσσαλίας (ΠΘ).

Επιβλέπων καθηγητής:

Αμούτζιας Γρηγόριος, Αναπληρωτής Καθηγητής Βιοπληροφορικής με έμφαση στην Μικροβιολογία, Τμήμα Βιοχημείας Και Βιοτεχνολογίας, Σχολή Επιστημών Υγείας, Πανεπιστήμιο Θεσσαλίας

Τριμελής Συμβουλευτική Επιτροπή:

Αμούτζιας Γρηγόριος, Αναπληρωτής Καθηγητής Βιοπληροφορικής με έμφαση στην Μικροβιολογία, Τμήμα Βιοχημείας Και Βιοτεχνολογίας, Σχολή Επιστημών Υγείας, Πανεπιστήμιο Θεσσαλίας

Μόσιαλος Δημήτριος, Αναπληρωτής Καθηγητής Βιοτεχνολογίας Μικροβίων, Τμήμα Βιοχημείας Και Βιοτεχνολογίας, Σχολή Επιστημών Υγείας, Πανεπιστήμιο Θεσσαλίας

Ηλιόπουλος Ιωάννης, Αναπληρωτής Καθηγητής Μοριακής Βιολογίας - Βιοπληροφορικής, Τμήμα Ιατρικής, Σχολή Επιστημών Υγείας, Πανεπιστήμιο Κρήτης

ΕΥΧΑΡΙΣΤΙΕΣ

Αρχικά, θα ήθελα να ευχαριστήσω ιδιαίτέρως τον υπεύθυνο καθηγητή μου τον κ. Αμούτζια Γρηγόριο για τη μοναδική βοήθεια και υποστήριξη του, για όλες τις γνώσεις και συμβουλές που μου προσέφερε και καθ' όλη τη διάρκεια της διπλωματικής εργασίας. Στη συνέχεια θα ήθελα να ευχαριστήσω τα υπόλοιπα μέλη της τριμελούς επιτροπής μου. Επίσης θα ήθελα να ευχαριστήσω το ίδρυμα Μποδοσάκη για την οικονομική στήριξη που μου παρείχε σε όλη την πορεία του μεταπτυχιακού προγράμματος σπουδών, με κωδικό υποτροφίας BDA-394. Τέλος, θα ήθελα να ευχαριστήσω τους κοντινούς μου ανθρώπους για την υποστήριξη τους σε αυτούς τους δύσκολους καιρούς.

ΠΕΡΙΛΗΨΗ

Τα βακτήρια του γένους *Bacillus* είναι πολύ σημαντικά στη βασική και την εφαρμοσμένη έρευνα. Η ομάδα *B. altitudinis* – *B. pumilus* – *B. safensis* αποτελείται από στελέχη που έχουν απομονωθεί κυρίως από αφιλόξενα περιβάλλοντα. Αξιοποιώντας 797 γονιδιώματα του γένους *Bacillus* πραγματοποιήσαμε μια φυλογενωμική και συγκριτική γονιδιωματική ανάλυση προκειμένου να αναγνωρίσουμε τις πρωτεΐνες πυρήνα της ομάδας και τα γονίδια που διαχωρίζουν τη συγκεκριμένη ομάδα από τις υπόλοιπες. Οι πρωτεΐνες πυρήνας της ομάδας (~2560) εμπλέκονται κυρίως σε βασικές κυτταρικές λειτουργίες αύξησης και επιβίωσης, ενώ αναγνωρίστηκαν 26 γονίδια - αποτυπώματα που χαρακτηρίζουν τον κλάδο.

Λέξεις – κλειδιά:

Bacillus, φυλογενωμική, πρωτεΐνες πυρήνας, μοριακό αποτύπωμα

ABSTRACT

The bacteria of the genus *Bacillus* are very important in basic and applied research. The *B. altitudinis* – *B. pumilus* – *B. safensis* group is comprised mostly of strains that are isolated from harsh environments. By utilizing 797 *Bacillus* genomes we performed a phylogenomic and comparative genomic analysis in order to identify the core proteins of the group and fingerprint proteins that separate this group from the rest. The core proteins (~2560) are mostly implicated in basic cellular growth and proliferation functions, while 26 protein fingerprints were recognized.

Keywords:

Bacillus, phylogenomics, core proteins, molecular fingerprint

Περιεχόμενα

1. ΕΙΣΑΓΩΓΗ.....	8
1.1. Παν-γονιδίωμα και γονιδίωμα πυρήνας.....	8
1.1.1. Παν-γονιδίωμα.....	8
1.1.2. Γονιδίωμα πυρήνας.....	9
1.1.3. Εύρεση γονιδιώματος πυρήνα.....	10
1.1.4. Γονιδίωμα πυρήνας και αριθμός γονιδιωμάτων.....	11
1.2. Διαχωρισμός των βακτηριακών ειδών.....	12
1.3. Το γένος <i>Bacillus</i>	14
1.3.1. <i>Bacillus subtilis sensu lato</i>	14
1.3.2. <i>Bacillus cereus sensu lato</i>	16
1.3.3. Εξελικτικές αναλύσεις του γένους.....	18
1.4. Ο κλάδος <i>B. altitudinis</i> - <i>B. pumilus</i> - <i>B. safensis</i>	19
1.4.1. <i>Bacillus altitudinis</i>	20
1.4.2. <i>Bacillus pumilus</i>	21
1.4.3. <i>Bacillus safensis</i>	22
1.4.4. Εξελικτικές αναλύσεις του κλάδου <i>B. altitudinis</i> - <i>B. pumilus</i> - <i>B. safensis</i>	23
2. ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ.....	26
2.1. Λογισμικό.....	26
2.1.1. Linux Ubuntu 20.....	26
2.1.2. Python 3.6.....	26
2.1.3. Seaview 5.....	26
2.1.4. PubMLST.....	26
2.1.5. Pyani.....	27
2.1.6. Treedyn.....	27
2.1.7. Λογισμικό αναγνώρισης πρωτεώματος πυρήνα.....	27
2.1.8. Λειτουργικός χαρακτηρισμός πρωτεϊνών.....	27
2.2. Λήψη πρωτεωμάτων.....	28
3. ΑΠΟΤΕΛΕΣΜΑΤΑ - ΣΥΖΗΤΗΣΗ.....	29
4. ΣΥΜΠΕΡΑΣΜΑΤΑ.....	35
5. ΒΙΒΛΙΟΓΡΑΦΙΑ.....	36

Περιεχόμενα (Πίνακες)

Table 1. Ο συνολικός αριθμός των πρωτεωμάτων που χρησιμοποιήθηκαν στην ανάλυση ανά φυλογενωμική ομάδα.....	28
--	----

Περιεχόμενα (Εικόνες)

Figure 1. Απεικόνιση των εννοιών πανγονιδίωμα και γονιδίωμα πυρήνας.....	9
Figure 2. Το γονιδίωμα πυρήνας του γένους <i>Pseudomonas</i> και η εξάρτηση του από τον αριθμό των πρωτεωμάτων υπό εξέταση.....	12
Figure 3. Διακριτική ικανότητα ανά μέθοδο διαχωρισμού βακτηριακών οργανισμών σε σχέση με τις ταξινομικές ομάδες.....	13
Figure 4. Φυλογένεση του κλάδου <i>Bacillus subtilis sensu lato</i> (σύμπλεγμα <i>Bacillus subtilis</i>).....	15
Figure 5. Φυλογένεση του κλάδου <i>Bacillus cereus sensu lato</i>	17
Figure 6. Φυλογενωμική ανάλυση του γένους <i>Bacillus</i>	19
Figure 7. Φυλογενωμική ανάλυση του κλάδου <i>B. altitudinis</i> – <i>B. pumilus</i> – <i>B. safensis</i> με τη	

μέθοδο Genome-BLAST Distance Phylogeny (GBDP)	25
Figure 8. Φυλογενωμική ανάλυση της ομάδας <i>B. altitudinis</i> - <i>B. pumilus</i> - <i>B. safensis</i>	30
Figure 9. Αριθμός των πρωτεϊνών πυρήνα για τον κλάδο <i>B. altitudinis</i> – <i>B. pumilus</i> – <i>B. safensis</i> και των υποομάδων που χαρακτηρίστηκαν.	31
Figure 10. Κύριες λειτουργικές κατηγορίες των πρωτεϊνών πυρήνα του κλάδου <i>B. altitudinis</i> – <i>B.</i> <i>pumilus</i> – <i>B. safensis</i>	32

1. ΕΙΣΑΓΩΓΗ

1.1. Παν-γονιδίωμα και γονιδίωμα πυρήνας

1.1.1. Παν-γονιδίωμα

Η ανάπτυξη που έχουν γνωρίσει οι τεχνολογίες αλληλούχησης νέας γενιάς έχει οδηγήσει στη κατάθεση πληθώρας γενωμικών δεδομένων στις δημόσιες βάσεις δεδομένων με φυσικό επακόλουθο τη θεμελίωση της γενωμικής (genomics). Οι γενωμικές αναλύσεις έχουν μετεξελιχθεί από αναλύσεις ενός ή λίγων γονιδιωμάτων σε έρευνες που χρησιμοποιούν εκατοντάδες ή και χιλιάδες γονιδιώματα. Κατά την γενωμική ανάλυση ενός φυλογενετικού κλάδου με συγκεκριμένους αντιπροσώπους το σύνολο των γονιδίων που περιέχουν απαρτίζουν το παν-γονιδίωμα του συγκεκριμένου συνόλου οργανισμών. Οι αναλύσεις του παν-γονιδιώματος παρέχουν πληροφορίες για την εκτίμηση της γονιδιωματικής ποικιλομορφίας του συνόλου υπό εξέταση και την πρόβλεψη του αριθμού των πρόσθετων γονιδιωμάτων που είναι απαραίτητα για τον πλήρη χαρακτηρισμό της (Vernikos et al., 2015).

Από δεδομένα παν-γονιδιωματικών αναλύσεων μπορεί κανείς να προβλέψει το αποτέλεσμα της εισαγωγής περισσότερων γονιδιωμάτων στο σύνολο στον αριθμό των γονιδίων. Η προσθήκη περισσότερων οργανισμών στο σύνολο των δεδομένων μπορεί να αποκαλύπτει νέα γονίδια που υπάρχουν στη γονιδιακή δεξαμενή της ταξινομικής ομάδας αυτής, όπου τότε γίνεται αναφορά σε ανοιχτό παν-γονιδίωμα (open pan-genome). Στη περίπτωση όπου δεν εμφανίζεται κάποια μεταβολή στο συνολικό αριθμό γονιδίων, τότε πρόκειται για ένα κλειστό παν-γονιδίωμα (Bosi et al., 2015).

Στη διεθνή βιβλιογραφία το παν-γονιδίωμα μπορεί να παρασταθεί ως διαγράμματα Venn, όπως φαίνεται και στην Figure 1 με κάθε υποσύνολο στο διάγραμμα να αντικατοπτρίζει το σύνολο των γονιδίων που περιέχει το εκάστοτε γονιδίωμα. Οι διάφοροι συνδυασμοί – αλληλοεπικαλύψεις μεταξύ ορισμένων υποσυνόλων αποτελούν τα πρόσθετα γονίδια (accessory genes), δηλαδή γονίδια που υπάρχουν μόνο σε ορισμένους οργανισμούς. Η τομή όλων των υποσυνόλων, δηλαδή τα γονίδια που είναι κοινά σε όλα τα γονιδιώματα αποτελεί το γονιδίωμα πυρήνα (core genome). Τέλος τα γονίδια που είναι μοναδικά για κάθε οργανισμό αναφέρονται ως μοναδικά γονίδια (unique genes) (Bosi et al., 2015).

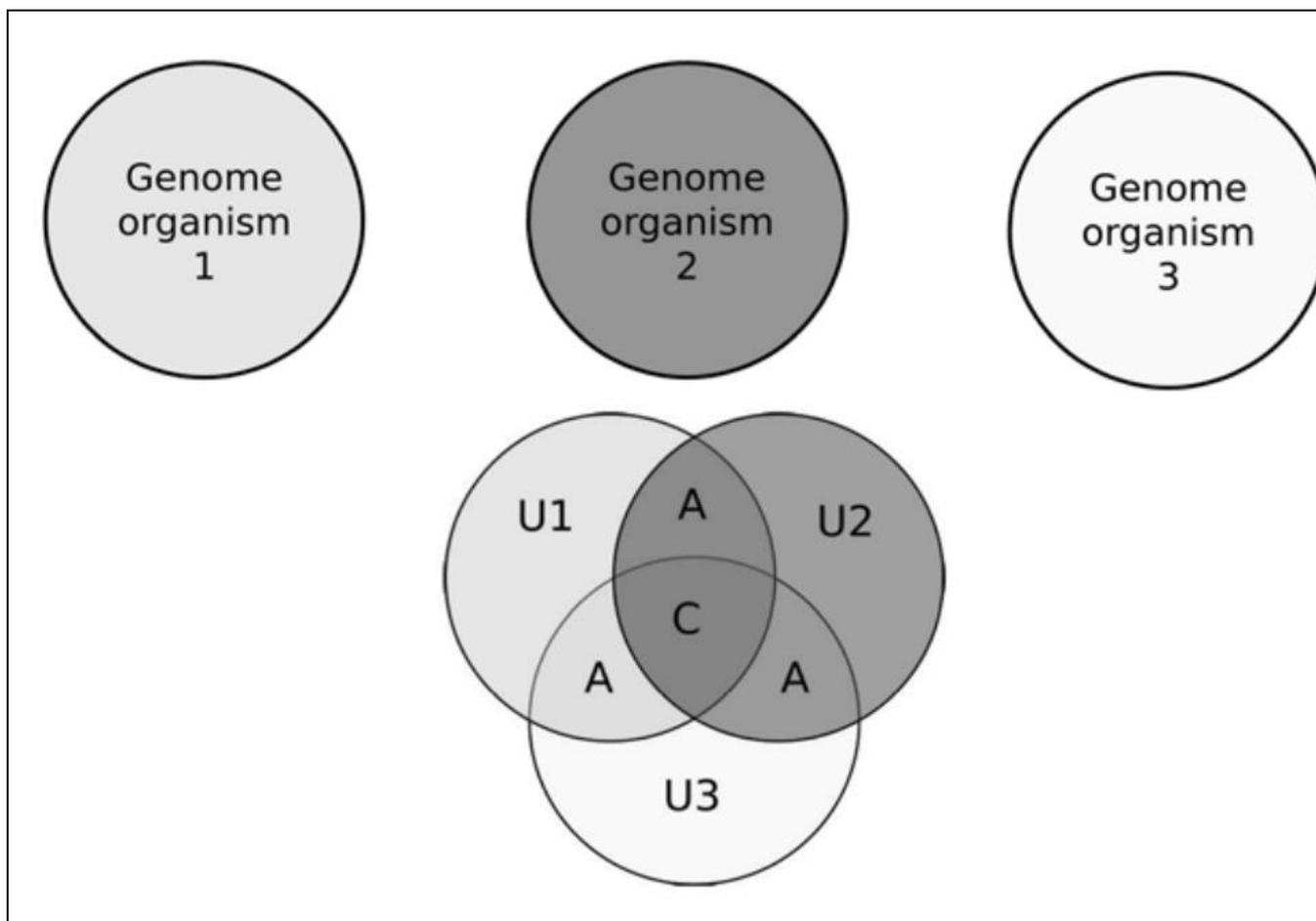


Figure 1. Απεικόνιση των εννοιών πανγονιδίωμα και γονιδίωμα πυρήνας. Το διάγραμμα Venn αναπαριστά ένα υποθετικό πανγονιδίωμα από 3 οργανισμούς (κύκλοι). Το γράμμα C αναπαριστά το γονιδίωμα πυρήνα (core genome), δηλαδή τα γονίδια που είναι κοινά σε όλους τους οργανισμούς. Το γράμμα A αναπαριστά το πρόσθετο γονιδίωμα (accessory genome) που είναι κοινό μεταξύ κάποιων οργανισμών αλλά όχι του συνόλου, ενώ το μοναδικό γονιδίωμα για κάθε οργανισμό απεικονίζεται με το γράμμα U (unique genome) (Bosi et al., 2015)

1.1.2. Γονιδίωμα πυρήνας

Το γονιδίωμα πυρήνας αποτελείται από τα γονίδια που είναι κοινά στους οργανισμούς που μελετώνται και έχει προταθεί ότι είναι απαραίτητα για βασικές κυτταρικές λειτουργίες (Bosi et al., 2015). Ένα παράδειγμα μελέτης που δείχνει ότι το γονιδίωμα πυρήνας συγκεκριμένων ταξινομικών ομάδων συγκαταλέγει γονίδια απαραίτητα για το ξεχωριστό τρόπο ζωής τους είναι η (Nikolaidis et al., 2020) στο γένος *Pseudomonas*.

Τα βακτήρια του συγκεκριμένου γένους απομονώνονται από πλειάδα ενδαιτημάτων και ένας από τους γνωστότερους αντιπροσώπους είναι το ευκαιριακό παθογόνο *Pseudomonas aeruginosa*. Το γονιδίωμα πυρήνας της *Pseudomonas aeruginosa* εμπεριέχει γονίδια σχετιζόμενα με τη δημιουργία βλέννας (mucoidy related genes) που είναι πολύ σημαντικά για την παθογονικότητα των στελεχών αυτών

σε ασθενείς με κυστική ίνωση. Επιπλέον, παρατηρήθηκε το οπερόνιο που είναι απαραίτητο για την παραγωγή του μεταλλοφόρου ψευδοπαλίνης. Η ψευδοπαλίνη χρησιμοποιείται από τα βακτήρια ως μέσο απορρόφησης σημαντικών μετάλλων από το ξενιστή και πιστεύεται ότι παίζει ρόλο στη παθογένεση. Μια ακόμα ενδιαφέρουσα περίπτωση αποτελούν τρία οπερόνια τοξίνης – αντιτοξίνης της οικογένειας Tsi τα οποία εικάζεται ότι παρέχουν ανταγωνιστικό πλεονέκτημα στην *P. aeruginosa* για την επικράτηση σε ένα ενδιαίτημα (Nikolaidis et al., 2020).

1.1.3. Εύρεση γονιδιώματος πυρήνα

Μια από τις στρατηγικές για τον εντοπισμό του γονιδιώματος πυρήνα είναι η χρήση των ομάδων ορθόλογων γονιδίων – πρωτεϊνών (orthologous groups) μεταξύ των διαφόρων οργανισμών που μελετώνται (Vernikos et al., 2015). Για την εύρεση των ομάδων ορθολόγων υπάρχουν διάφορες μεθοδολογίες και λογισμικά όπως το InParanoid, το OrthoMCL, το OMA, το OrthoFinder (Emms and Kelly, 2019) και η χρήση ανταποδοτικού BLAST (Hernández-Salmerón and Moreno-Hagelsieb, 2020; Nikolaidis et al., 2020).

Η μέθοδος του ανταποδοτικού BLAST είναι απλή στην εκτέλεση και έχει δειχθεί ότι έχει χαμηλό αριθμό ψευδώς θετικών αποτελεσμάτων. Μεταξύ δύο οργανισμών που αναλύονται συγκρίνεται το σύνολο των πρωτεϊνών που παράγει το κάθε γονιδίωμα (πρωτέωμα). Δύο πρωτεΐνες που παράγονται από τα διαφορετικά γονιδιώματα πρέπει να χαρακτηριστούν ως ανταποδοτικά τα καλύτερα χτυπήματα BLAST (Reciprocal Best Hits – RBH) για να είναι ορθόλογες. Αρχικά χρησιμοποιώντας το σύνολο των πρωτεϊνών του πρώτου οργανισμού ως ακολουθίες επερώτησης γίνεται αναζήτηση ομολόγων στο πρωτέωμα του δεύτερου οργανισμού. Στη συνέχεια η διαδικασία επαναλαμβάνεται με αντίστροφη φορά. Δύο πρωτεΐνες που εμφανίζουν η μία την άλλη ως το καλύτερο αποτέλεσμα ανεξάρτητα τη φορά αναζήτησης είναι ανταποδοτικά τα καλύτερα χτυπήματα BLAST. Το πρόγραμμα με το οποίο γίνεται η σύγκριση των πρωτεϊνών μέσω του αλγορίθμου BLAST είναι το blastp (Hernández-Salmerón and Moreno-Hagelsieb, 2020)

Σε πρόσφατη δημοσίευση της ερευνητικής ομάδας μας (Nikolaidis et al., 2020) ανέπτυξα μια σειρά προγραμμάτων (scripts) γραμμένα σε python 3 που χρησιμοποιούν τη μέθοδο ανταποδοτικού BLAST σε συνδυασμό με κάποια αυστηρά κριτήρια για την ανίχνευση γονιδιώματος πυρήνα. Η συγκεκριμένη σειρά διεργασιών βασίζεται στον ορισμό ενός πρωτεώματος αναφοράς, το οποίο χρησιμοποιείται ως βάση για τις συγκρίσεις των πρωτεωμάτων και τη δημιουργία των ομάδων ορθόλογων πρωτεϊνών (orthologous groups of proteins). Λόγω της ύπαρξης ενός σημείου αναφοράς η παραπάνω διαδικασία χαρακτηρίζεται ένας εναντίον πολλών (one to many) και είναι σημαντικά ταχύτερη από μεθόδους που πραγματοποιούν όλες τις δυνατές συγκρίσεις (many to many). Συνολικά πραγματοποιείται ανταποδοτικό BLAST για κάθε ζεύγος οργανισμών της ανάλυσης, όπου το ένα από τα δύο είναι το πρωτέωμα αναφοράς. Για κάθε ζεύγος συγκρίσεων υπολογίζεται ο μέσος όρος ομοιότητας όλων των πιθανών ορθολόγων και

αφαιρούνται εκείνα τα οποία αποκλίνουν κατά δύο τυπικές αποκλίσεις από το μέσο όρο αυτό, φιλτράροντας έτσι προβληματικά ορθόλογα. Στη συνέχεια δημιουργείται ένας πίνακας με τα καλύτερα ανταποδοτικά χτυπήματα για όλα τα πρωτεώματα επερώτησης με τις διάφορες γραμμές να αντιστοιχούν στις ορθόλογες πρωτεΐνες και οι στήλες στα διάφορα πρωτεώματα. Ο πίνακας αυτός φιλτράρεται και διατηρούνται μόνο οι πρωτεΐνες του πρωτεώματος αναφοράς που έχουν ορθόλογο στα υπόλοιπα πρωτεώματα, που αποτελούν και τις πρωτεΐνες πυρήνα του σετ που αναλύεται. Τέλος, πραγματοποιείται η φυλογενωμική ανάλυση όπου κάθε ομάδα ορθολόγων υποβάλλεται σε πολλαπλή στοίχιση με το MUSCLE (Edgar, 2004) και ενώνεται με τις υπόλοιπες δημιουργώντας μια υπερ-στοίχιση η οποία και φιλτράρεται με το Gblocks (Castresana, 2000). Η τελική στοίχιση χρησιμοποιείται για τον υπολογισμό ενός φυλογενωμικού δέντρου αποστάσεων με τη χρήση του αλγορίθμου BioNJ (Gouy et al., 2010). Όλες οι παράμετροι για τα παραπάνω βήματα και οι κατάλογοι των αρχείων εισόδου και εξόδου ορίζονται σε ειδικά αρχεία παραμέτρων και είναι μεταβλητά από το χρήστη.

Ένα λογισμικό που πραγματοποιεί όλες τις δυνατές συγκρίσεις των οργανισμών υπό ανάλυση είναι το Orthofinder. Το συγκεκριμένο λογισμικό χρησιμοποιεί για τις συγκρίσεις του τον αλγόριθμο BLAST είτε τον αλγόριθμο DIAMOND που είναι ταχύτερος (Emms and Kelly, 2019). Ωστόσο, η αύξηση της ταχύτητας μέσω του DIAMOND συνεπάγεται ένα κόστος ευαισθησίας. Προκειμένου να βρεθεί μια ισορροπία μεταξύ της αυξημένης ταχύτητας και των καλύτερων ανταποδοτικών χτυπημάτων (RBH) κρίνεται απαραίτητη η χρήση των πιο ευαίσθητων ρυθμίσεων στο DIAMOND (Hernández-Salmerón and Moreno-Hagelsieb, 2020).

1.1.4. Γονιδίωμα πυρήνας και αριθμός γονιδιωμάτων

Ο αριθμός των γονιδιωμάτων που εξετάζονται επηρεάζει το πλήθος των γονιδίων που αναγνωρίζονται ως γονιδίωμα πυρήνας. Πιο συγκεκριμένα όσο ο αριθμός των εξεταζόμενων οργανισμών αυξάνει, εμφανίζεται μια πτώση στον αριθμό των γονιδίων πυρήνα όπως περιγράφεται (Bosi et al., 2015; Nikolaidis et al., 2020). Για κάθε νέο οργανισμό που εισάγεται στην ανάλυση όλο και λιγότερες πρωτεΐνες παραμένουν παρούσες στο σύνολο των οργανισμών, με αποτέλεσμα τη μείωση των ομάδων ορθολόγων πρωτεϊνών που χρησιμοποιούνται για τη μέτρηση του γονιδιώματος πυρήνα.

Η γραφική παράσταση της μείωσης που προαναφέρθηκε μπορεί να χρησιμοποιηθεί για να εκτιμήσει κανείς τον πραγματικό αριθμό πρωτεϊνών πυρήνα για το σύνολο των οργανισμών καθώς και αν η ανάλυση περιλαμβάνει αρκετά στελέχη για σαφή συμπεράσματα. Σε μια πολύ πρόσφατη εργασία που αφορά το γονιδίωμα πυρήνα του γένους *Pseudomonas* (Nikolaidis et al., 2020) δείχθηκε ότι η μείωση αυτή έχει εκθετικό χαρακτήρα και ότι το πραγματικό σύνολο πρωτεϊνών πυρήνα για το συγκεκριμένο γένος είναι πολύ κοντά σε αυτό που υπολογίζεται με τη χρήση 491 οργανισμών όπως φαίνεται στην Figure 2.

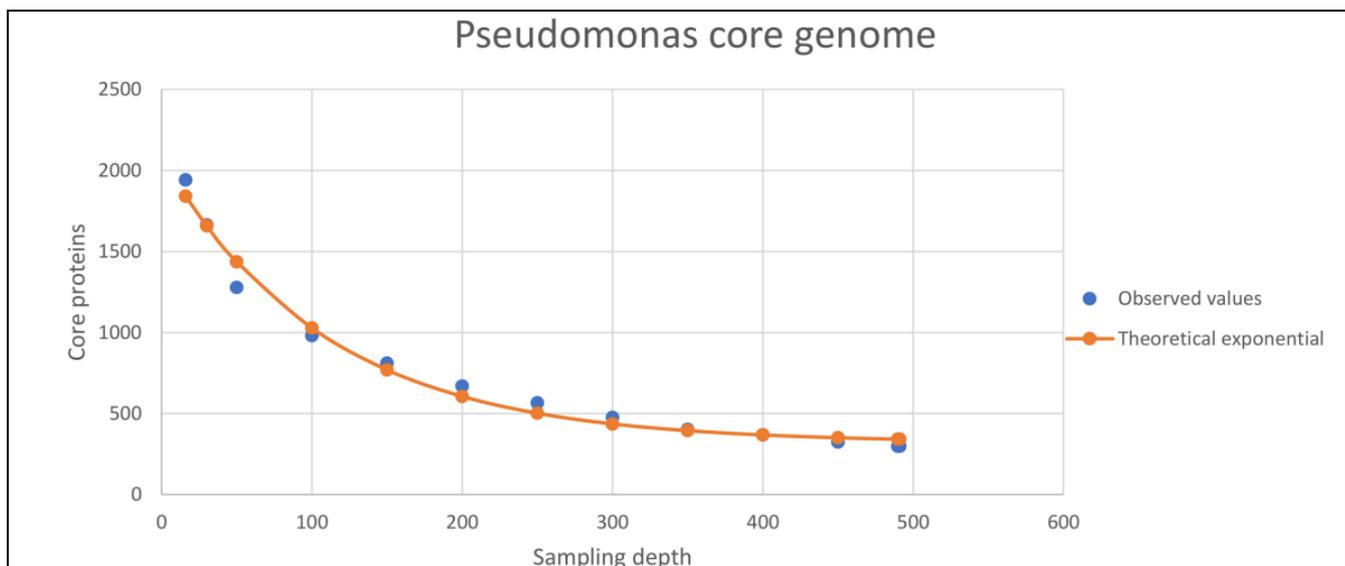


Figure 2. Το γονιδίωμα πυρήνας του γένους *Pseudomonas* και η εξάρτηση του από τον αριθμό των πρωτεωμάτων υπό εξέταση. Για τη δημιουργία του γραφήματος πραγματοποιήθηκαν τυχαίες δειγματοληψίες 16, 30, 50, 100, 150, έως 491 οργανισμών του γένους *Pseudomonas*. Κάθε τυχαία δειγματοληψία πραγματοποιήθηκε 20 φορές (Nikolaidis et al., 2020).

1.2. Διαχωρισμός των βακτηριακών ειδών

Το βασικό σύστημα για την ταξινόμηση των βακτηρίων και τον χαρακτηρισμό της βακτηριακής ποικιλότητας είναι μέσω της αλληλουχίας της μικρής ριβοσωμικής υπομονάδας 16S rRNA. Η χρήση αυτού του συστήματος προτάθηκε από τον Carl Woese και τους συνεργάτες του σε συνδυασμό με την εδραίωση των υπερ-βασιλείων των βακτηρίων, των αρχαίων και των ευκαρυωτών που μπορούν να διαχωριστούν μέσω του 16S rRNA (Woese et al., 1990). Οι παραλλαγές της αλληλουχίας αυτής μπορούν να εμφανίζουν αρκετά μεγάλη διακριτική ικανότητα και να χρησιμοποιούνται στο διαχωρισμό και ορισμό ακόμα και των βακτηριακών ειδών (Maiden et al., 2013). Ωστόσο παρά την επιτυχία της παραπάνω μεθόδου υπάρχουν περιπτώσεις όπου η ικανότητα διαχωρισμού της αποτυγχάνει για κάποια αδελφά βακτηριακά είδη, καθώς αυτά εμφανίζουν μεγάλη ομοιότητα στις ακολουθίες 16S rRNA (Maiden et al., 2013).

Η ανάγκη για ένα διαχωριστικό σύστημα μεγαλύτερης ευκρίνειας οδήγησε στην δημιουργία άλλων μεθόδων που χρησιμοποιούν περισσότερους του ενός γενετικούς τόπους (multi-locus sequence typing - MLST). Το MLST βασίζεται στη χρήση έξι έως οκτώ γενετικών τόπων και μπορεί να διαχωρίζει τα διάφορα βακτηριακά είδη στη πλειονότητα των περιπτώσεων (Maiden, 2006). Ωστόσο με την πάροδο των χρόνων και τη συσσώρευση γενωμικών δεδομένων έρευνες που θέλουν να αποσαφηνίσουν διαφορές μεταξύ στελεχών του ίδιου είδους χρειάζονται ακόμα μεγαλύτερη διακριτική ικανότητα. Η εξέλιξη του MLST που παρέχει τέτοιες δυνατότητες είναι το ριβοσωμικό MLST (ribosomal MLST - rMLST). Το rMLST εκμεταλλεύεται 53 γενετικούς τόπους των ριβοσωμικών πρωτεϊνών (rps) (Jolley et al., 2012), όπως φαίνεται και στην Figure 3.

Οι μέθοδοι διαχωρισμού που βασίζονται σε λίγα ή περιορισμένα γονίδια εξελίχθηκαν σε μεθόδους που αξιοποιούν δεδομένα από ολόκληρο το γονιδίωμα. Η χρήση πολλαπλών, κοινών μεταξύ των διαφόρων οργανισμών, γενετικών τόπων που βρίσκονται σε όλο το γονιδίωμα δίνει νέες δυνατότητες σε ερευνητικές ομάδες που θέλουν να πραγματοποιήσουν μελέτες σε πολύ κοντινά (εξελικτικά) στελέχη ή κλώνους (Figure 3), όπως για παράδειγμα σε συγκρίσεις δειγμάτων από έναν ασθενή (Jolley et al., 2017). Η συγκεκριμένη προσέγγιση αξιοποιεί το γονιδίωμα πυρήνα που αποτελείται από εκατοντάδες ή χιλιάδες ορθόλογα γονίδια/πρωτεΐνες (Kyriades et al., 2014) και θεωρείται ανθεκτική απέναντι σε γεγονότα που μπορούν να μπερδέψουν το εξελικτικό σήμα κάποιων μεμονομένων οικογενειών γονιδίων όπως η οριζόντια γονιδιακή μεταφορά που συμβαίνει εκτενέστατα στα βακτήρια (Gogarten and Townsend, 2005).

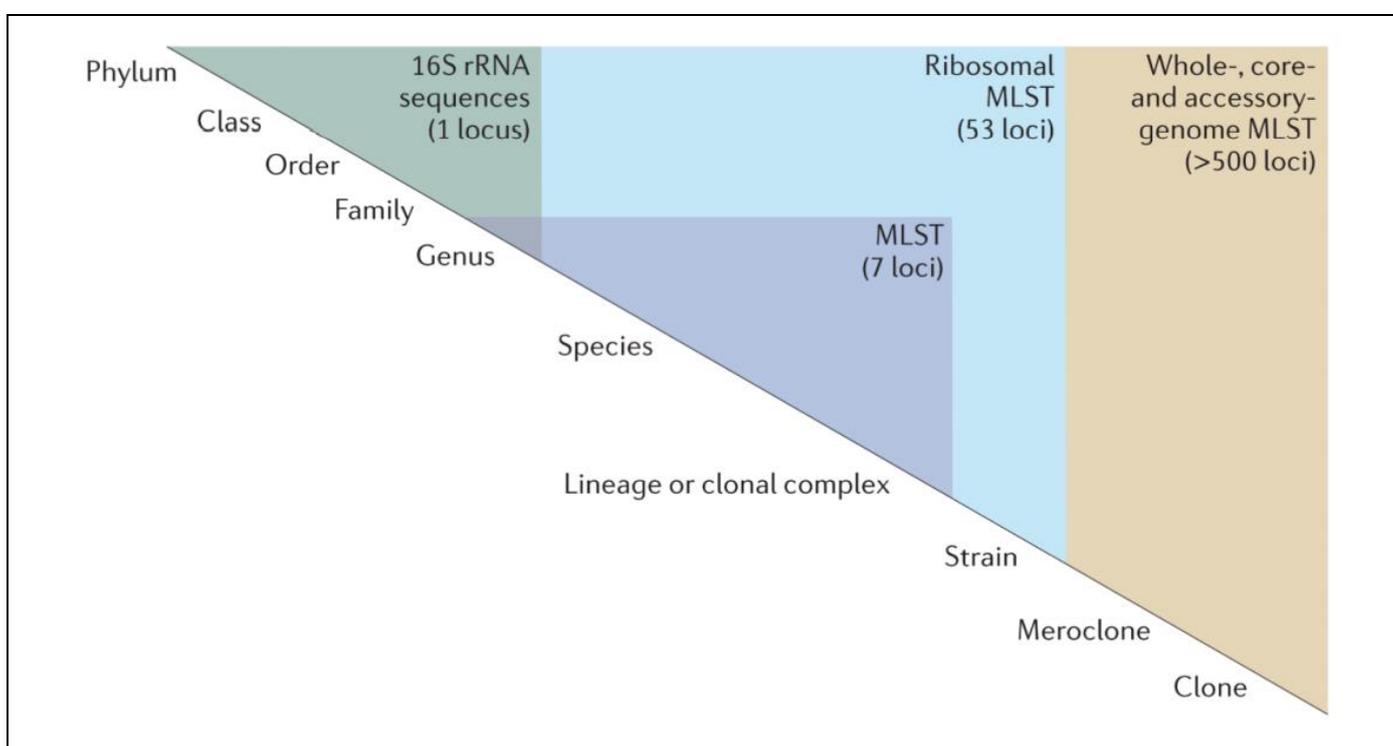


Figure 3. Διακριτική ικανότητα ανά μέθοδο διαχωρισμού βακτηριακών οργανισμών σε σχέση με τις ταξινομικές ομάδες. Η χρήση περισσότερων και διαφορετικών ορθόλογων αλληλουχιών προσφέρει μεγαλύτερη διακριτική ικανότητα (Maiden et al., 2013).

Η τελευταία προσέγγιση συγκρίσεων γονιδιωμάτων για το σωστό ταξινομικό χαρακτηρισμό τους αποτελεί η μέση νουκλεοτιδική ταύτιση (average nucleotide identity – ANI) και η υβριδοποίηση DNA, που προτάθηκε από τον Richter και το συνεργάτη του το 2009. Η βασική ιδέα στηρίζεται στις τεχνικές υβριδοποίησης DNA μεταξύ δύο οργανισμών (DNA-DNA hybridization; DDH), προκειμένου να αποσαφηνιστεί το αν ανήκουν στο ίδιο είδος ή όχι. Όσο πιο κοντά ταξινομικά είναι δύο οργανισμοί τόσο μεγαλύτερη θα είναι και η τιμή υβριδοποίησης καθώς εμφανίζουν περισσότερη ομοιότητα σε επίπεδο DNA. Η τιμή κατώφλι για την ένταξη δύο οργανισμών στο ίδιο είδος είναι τιμές DDH μεγαλύτερες ή ίσες

του 70%. Στην εποχή της γενωμικής όπου ολόκληρα γονιδιώματα είναι πλέον διαθέσιμα, η σύγκριση αυτή μπορεί να γίνει μέσω της τιμής μέσης νουκλεοτιδικής ταύτισης. Ο υπολογισμός της τιμής αυτής μπορεί να γίνει είτε μέσω του αλγορίθμου BLAST (ANiB) είτε μέσω του MUMmer (ANIm). Τιμές ANiB/ANIm μεγαλύτερες του 95% αντιστοιχούν σε τιμές DDH μεγαλύτερες του 70% (Pritchard et al., 2016; Richter and Rosselló-Móra, 2009) και υποδηλώνουν ότι τα δύο γονιδιώματα ανήκουν στο ίδιο είδος.

1.3. Το γένος *Bacillus*

Αρχικές αναφορές στο πιο κοινό είδος του γένους έγιναν από τον Ehrenberg το 1835 όπου ανέφερε τον οργανισμό "*Vibrio subtilis*", ενώ το γένος *Bacillus* εδραιώθηκε το 1872 από τους F. Cohn και Koch και έγινε η μετονομασία του "*Vibrio subtilis*" σε "*Bacillus subtilis*" (Harwood, 1989). Τα βακτήρια του συγκεκριμένου γένους είναι θετικά κατά Gram αερόβια ή προαιρετικά αναερόβια, έχουν ραβδωτό σχήμα και την ικανότητα να σχηματίζουν σπόρια (Fritze, 2004). Ο *Bacillus subtilis* είναι από τους πιο γνωστούς αντιπρόσωπος του γένους και αποτελεί οργανισμό μοντέλο για τη διαδικασία δημιουργίας σπορίων (σποριογένεση) (Barbe et al., 2009).

Τα συγκεκριμένα βακτήρια είναι προσαρμοσμένα και μπορούν να βρεθούν σε πληθώρα ενδιαιτημάτων όπως χώμα, νερό και διάφοροι ξενιστές μεταξύ άλλων και ο άνθρωπος (Hernández-González et al., 2018; Patel and Gupta, 2020). Πολλαπλές ερευνητικές ομάδες έχουν περιγράψει τον πιθανό ρόλο αντιπροσώπων του γένους ως παράγοντες ενίσχυσης της φυτικής ανάπτυξης (Radhakrishnan et al., 2017), παθογόνα (Ehling-Schulz et al., 2019), βιοαντιδραστήρες για την παραγωγή αντιμικροβιακών πεπτιδίων αλλά και προβιοτικά (Sumi et al., 2015). Τα αντιμικροβιακά πεππίδια βρίσκουν εφαρμογές ως συντηρητικά τροφίμων και ως εναλλακτική λύση στα υπάρχοντα αντιβιοτικά (Sumi et al., 2015).

Ταξινομικά το γένος *Bacillus* ανήκει στο φύλο *Firmicutes* και εμπεριέχει περισσότερα από 280 είδη (<https://lpsn.dsmz.de/genus/bacillus>) (Logan et al., 2009). Μέσω της χρήσης του 16S rRNA και άλλων γονιδίων δεικτών έχουν πραγματοποιηθεί πολλαπλές εξελικτικές αναλύσεις του γένους (Xu and Côté, 2003), ωστόσο αποτελεί μια περιορισμένη μέθοδο για τη θέσπιση των ορίων μεταξύ των ειδών μέσα σε ένα γένος (Tindall et al., 2010). Στη βιβλιογραφία είναι γνωστή και σταθερή η ύπαρξη δύο κύριων μεγάλων κλάδων εντός του γένους οι οποίοι αναφέρονται ως *Bacillus subtilis sensu lato* και *Bacillus cereus sensu lato* (Hernández-González et al., 2018; Patel and Gupta, 2020). Ο πρώτος εμπεριέχει τον *Bacillus subtilis* και τους εξελικτικά κοντινούς οργανισμούς ενώ ο δεύτερος έχει λάβει το όνομα του από το γνωστό παθογόνο *Bacillus cereus* και εμπεριέχει άλλους εξίσου σημαντικούς αντιπροσώπους.

1.3.1. *Bacillus subtilis sensu lato*

Ο κλάδος *B. subtilis sensu lato* είναι επίσης γνωστός και ως σύμπλεγμα *B. subtilis* (*B. subtilis*

species complex). Στη φυλογενωμική ανάλυση των (Fan et al., 2017) χρησιμοποίησαν το γονιδίωμα πυρήνα 20 τυπικών στελεχών του συγκεκριμένου κλάδου όπως φαίνεται στην Figure 4. Συνολικά υπολογίστηκαν τέσσερις κύριοι φυλογενετικοί κλάδοι που υποστηρίζονται από τιμές bootstrap 100%. Ο πρώτος (clade I) επωνομάζεται «*subtilis*» καθώς περιέχει τα υποείδη του *B. subtilis* (*subspecies subtilis*, *spizizenii*, *inaquosorum*), ενώ παράλληλα συγκαταλέγει τα είδη *B. valismortis*, *B. mojavenensis* και *B. atrophaeus*. Ο δεύτερος κλάδος (clade II) που φέρει την ονομασία «*amyloliquefaciens*» απαρτίζεται από τους *B. amyloliquefaciens*, *B. velezensis*, *B. siamensis* και *B. methylotrophicus*. Αξίζει να σημειωθεί ότι το στέλεχος του *B. amyloliquefaciens* subsp. *plantarum* έχει πιο άμεση σχέση με τα στελέχη *B. velezensis* και *B. methylotrophicus* παρά με τον *B. amyloliquefaciens* DSM7. Ο τρίτος φυλογενωμικός κλάδος (clade III) «*licheniformis*» αποτελείται από τους *B. licheniformis* και *B. sonorensis*, ενώ πλέον συγκαταλέγει και τον *B. paralicheniformis* που αποτελεί έναν πάρα πολύ κοντινό συγγενή του *B. licheniformis* (Du et al., 2019). Ο τέταρτος και τελευταίος κλάδος «*pumilus*» απαρτίζεται από τους *B. altitudinis*, *B. pumilus*, *B. safensis*, *B. xiamenensis*, *B. stratosphericus* και *B. aerophilus* αποτελεί και τον κλάδο ενδιαφέροντος της παρούσας διπλωματικής εργασίας, ωστόσο μέχρι την ημέρα της συγγραφής της δεν υπάρχουν πλήρη γονιδιώματα υψηλής ποιότητας *B. stratosphericus*.

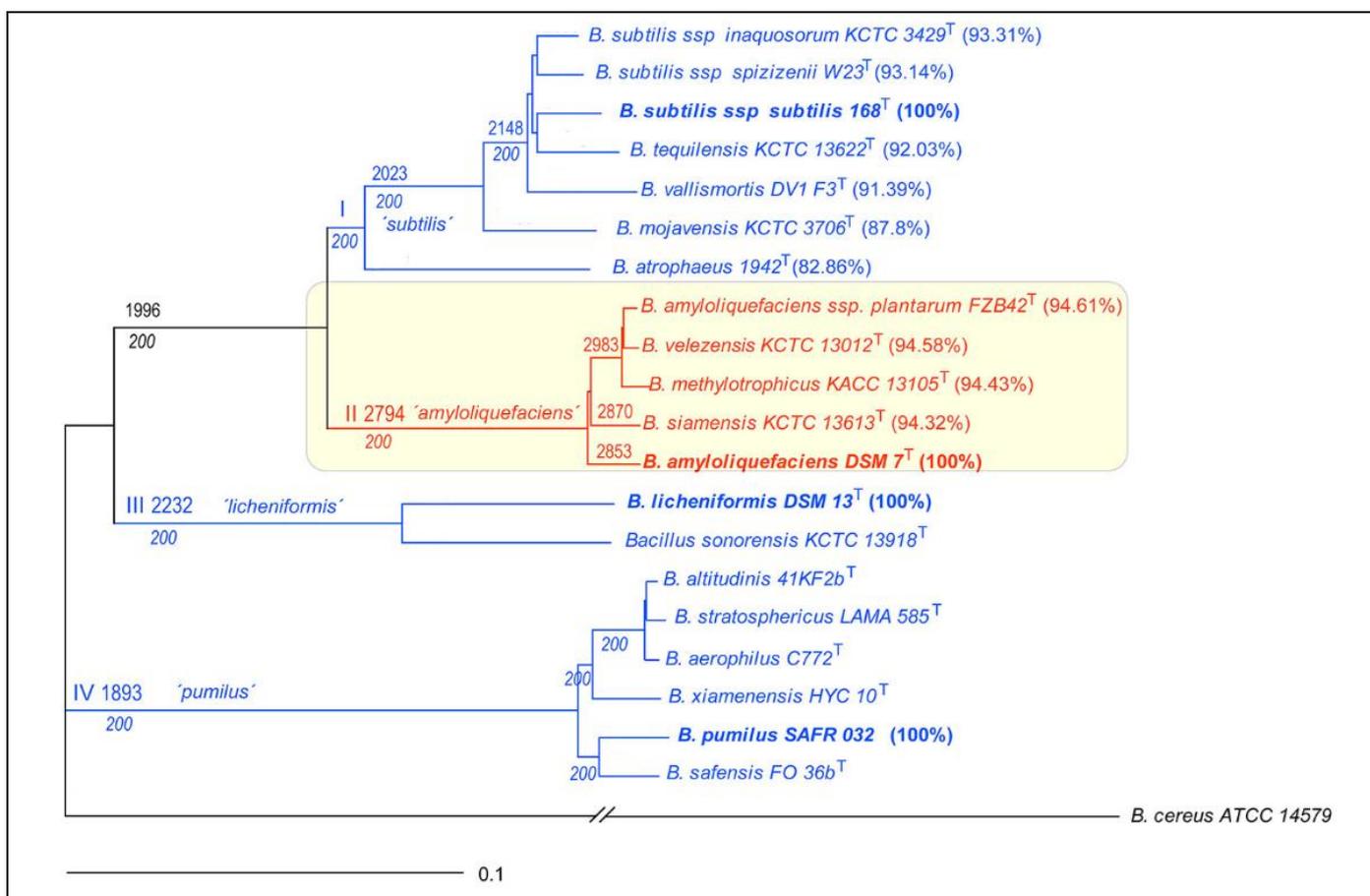


Figure 4. Φυλογένεση του κλάδου *Bacillus subtilis sensu lato* (σύμπλεγμα *Bacillus subtilis*). Για τον υπολογισμό του φυλογενετικού δέντρου χρησιμοποιήθηκε το γονιδίωμα πυρήνας των τυπικών στελεχών που φαίνονται στην εικόνα. Το γονιδίωμα του *B. cereus* ATCC 14579 χρησιμοποιήθηκε ως εξω-ομάδα.

Κάθε μεγάλος κλάδος απεικονίζεται αριθμείται με ένα ρωμαϊκό νούμερο I – IV στην αρχή του. Ο αριθμοί πάνω από τους κλάδους αναφέρονται στον αριθμό των κωδικών αλληλουχιών που απαρτίζουν τα γονιδιώματα πυρήνες για κάθε υποσύνολο γονιδιωμάτων. Οι αριθμοί κάτω από τους κλάδους αποτελούν τον αριθμό των επαναλήψεων bootstrap με μέγιστη τιμή το 200 (100%). Οι επί τοις εκατό τιμές δίπλα από τα διάφορα φύλλα κάθε μεγάλου κλάδου αποτελούν το ποσοστό ομοιότητας ANI με το στέλεχος αναφοράς κάθε υποομάδας. Τα στελέχη αναφοράς ανά υποομάδα εμφανίζονται με έντονα γράμματα (Fan et al., 2017).

1.3.2. *Bacillus cereus sensu lato*

Η ομάδα *Bacillus cereus sensu lato* είναι ευρέως γνωστή για τους αντιπροσώπους της που είναι σημαντικά παθογόνα των ζώων και του ανθρώπου. Συνολικά αποτελείται από τα οκτώ είδη τα οποία οργανώνονται σε 7 μεγάλες ομάδες και 3 φυλογενετικούς κλάδους όπως φαίνεται και στην Figure 5. Πιο συγκεκριμένα τα είδη *B. anthracis*, *B. cereus* και *B. thuringiensis* έχουν στελέχη που εμφανίζονται στους κλάδους I και II ενώ τα υπόλοιπα πέντε, *B. mycooides*, *B. pseudomycooides*, *B. weihenstephanensis*, *B. cytotoxicus*, και *B. toyonensis* αποτελούν τον κλάδο III (Bazinet, 2017).

Η συγκεκριμένη ομάδα έχει σχετικά σταθερό μέγεθος γονιδιώματος, με εξαίρεση το *B. cytotoxicus* που είναι ο πιο απομακρυσμένος οργανισμός. Επίσης ορισμένοι οργανισμοί φέρουν πλασμίδια που τους δίνουν την ικανότητα παραγωγής τοξινών με σημαντικές επιπτώσεις στην υγεία και την οικονομία. Πιο συγκεκριμένα ο *B. anthracis* φέρει πλασμίδιο που παράγει τις τοξίνες που προκαλούν την ασθένεια του άνθρακα, τα στελέχη *B. cereus* που προκαλούν έμετο κατά αντιστοιχία φέρουν ένα πλασμίδιο που τους επιτρέπει τη παραγωγή της τοξίνης cereulide, ενώ ο *B. thuringiensis* είναι γνωστός για την παραγωγή των τοξινών Cry και Cyt που δρουν ενάντια σε έντομα (Ehling-Schulz et al., 2019).

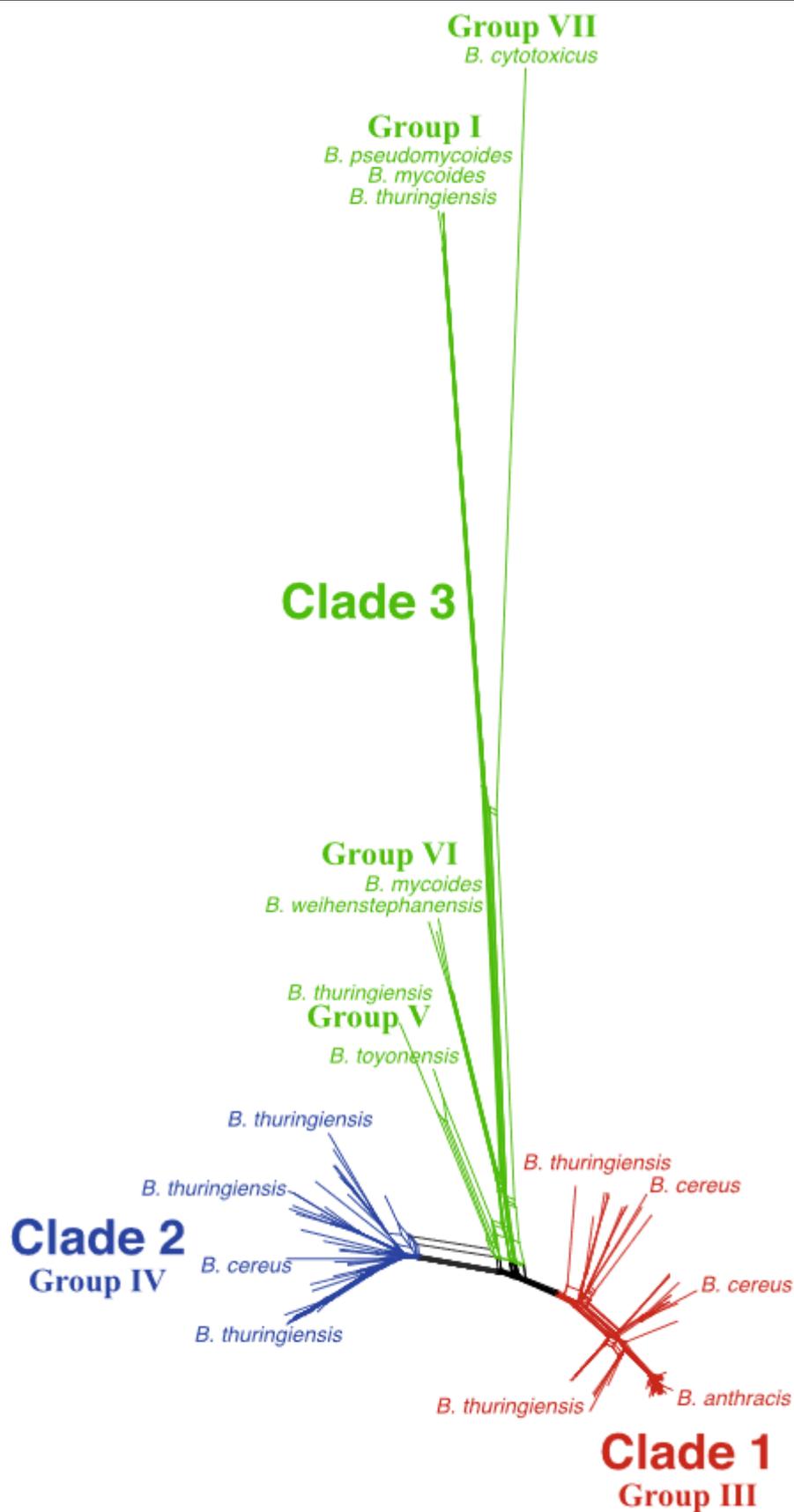


Figure 5. Φυλογένεση του κλάδου *Bacillus cereus sensu lato*. Για τον υπολογισμό του φυλογενετικού δικτύου χρησιμοποιήθηκαν δεδομένα παρουσίας/απουσίας γονιδίων από πλήρη γονιδιώματα 114 στελεχών όπως εξάγονται από το λογισμικό Roary. Στη συνέχεια, τα δεδομένα αυτά εισήχθησαν στο πρό-

γραμμα SplitsTree. Συνολικά αναγνωρίζονται οι 3 μεγάλοι κλάδοι και επτά ομάδες οργανισμών που είναι γνωστές από την υπάρχουσα βιβλιογραφία, με τη συγκεκριμένη μέθοδο που έρχεται σε συμφωνία με παλαιότερες φυλογενετικές αναλύσεις. Η ομάδα II φαίνεται να απουσιάζει καθώς δεν υπήρχε πλήρες γονιδίωμα που να την απαρτίζει (Bazinet, 2017).

1.3.3. Εξελικτικές αναλύσεις του γένους

Η σημασία των βακτηρίων του γένους *Bacillus* έχει οδηγήσει στην αλληλούχηση περίπου 5.000 γονιδιωμάτων (πηγή: NCBI Assembly; Σεπτέμβριος 2020), με περίπου 900 (16%) από αυτά να είναι υψηλής ποιότητας και συναρμολόγησης, χαρακτηρισμένα ως «πλήρες γονιδίωμα» ή «χρωμόσωμα». Οι φυλογενωμικές αναλύσεις χρησιμοποιούν τα γονίδια/πρωτεΐνες πυρήνα των οργανισμών υπό μελέτη ώστε να έχουν μια πιο εμπειριστατωμένη εικόνα της εξέλιξης των ειδών. Έτσι, πρόσφατες φυλογενωμικές μελέτες (Hernández-González et al., 2018; Patel and Gupta, 2020) έχουν χρησιμοποιήσει τον πλούτο αυτών των δεδομένων προκειμένου να αποσαφηνίσουν τις κύριες και δευτερεύουσες εξελικτικές ομάδες που απαρτίζουν το γένος *Bacillus* καθώς και τις σχέσεις μεταξύ τους με μεγαλύτερη ακρίβεια και πιστότητα.

Για παράδειγμα, στην ανάλυση τους, οι (Hernández-González et al., 2018) χρησιμοποίησαν πρωτεώματα από 83 οργανισμούς εκ των οποίων 4 ανήκουν στην ευρύτερη οικογένεια *Bacillaceae* ενώ τα υπόλοιπα είναι χαρακτηρισμένα ως *Bacillus*, όπως φαίνεται και στην εικόνα Figure 6. Φάνηκε ότι υπάρχουν 9 βασικοί κλάδοι στο γένος με τους *Bacillus subtilis sensu lato* (κίτρινος κλάδος) και *Bacillus cereus sensu lato* (καφέ κλάδος) να αποτελούν εκείνους με τους περισσότερους αντιπροσώπους. Ένας τρίτος κλάδος περιέχει τα βακτήρια που επιβιώνουν σε ακραία αλκαλικά περιβάλλοντα όπως ο *B. alcalophilus* (πράσινος κλάδος). Ένας τέταρτος, μονοφυλετικός, κλάδος αποτελείται από στελέχη *B. megaterium* (μωβ κλάδος) και είναι ο κοντινότερος στον *B. subtilis s.l.* Ο αμέσως επόμενος κλάδος (βαθύ μπλε) αποτελείται από μη χαρακτηρισμένα είδη του γένους και τον *B. horikoshii*. Ο έκτος κλάδος (πορτοκαλί) απαρτίζεται από τους *B. methanolicus*, *B. bataviensis*, *Bacillus sp. 1NLA3E* και *Bacillus sp. 2_A_57_CT2* και τα θαλάσσια στελέχη *B. oceanisediminidis*, *B. infantis* και *Bacillus sp. NRRL B-14911*. Βασικότερα του κλάδου αυτού έρχεται ο έβδομος κλάδος (θαλασσί) που αποτελείται εξ' ολοκλήρου από *Bacillus* υδάτινων οικοσυστημάτων. Οι τελευταίοι δύο κλάδοι αποτελούνται από τους *B. isronensis*, *Bacillus sp. B14905* (γκρι κλαδος) και *B. coagulans* (μαύρος κλάδος). Επίσης η φυλογενωμική ανάλυση δείχνει ότι οι οργανισμοί *O. iheyensis* και *Geobacillus sp. Y412MC52* μάλλον αποτελούν μέλη του γένους *Bacillus*.

Η πιο πρόσφατη φυλογενωμική ανάλυση των (Patel and Gupta, 2020) προτείνει την ύπαρξη μόνο των δύο κύριων κλάδων (*B. subtilis s.l.* και *B. cereus s.l.*) ενώ όλοι οι υπόλοιποι κλάδοι θα πρέπει να μεταφερθούν σε νέα γένη.

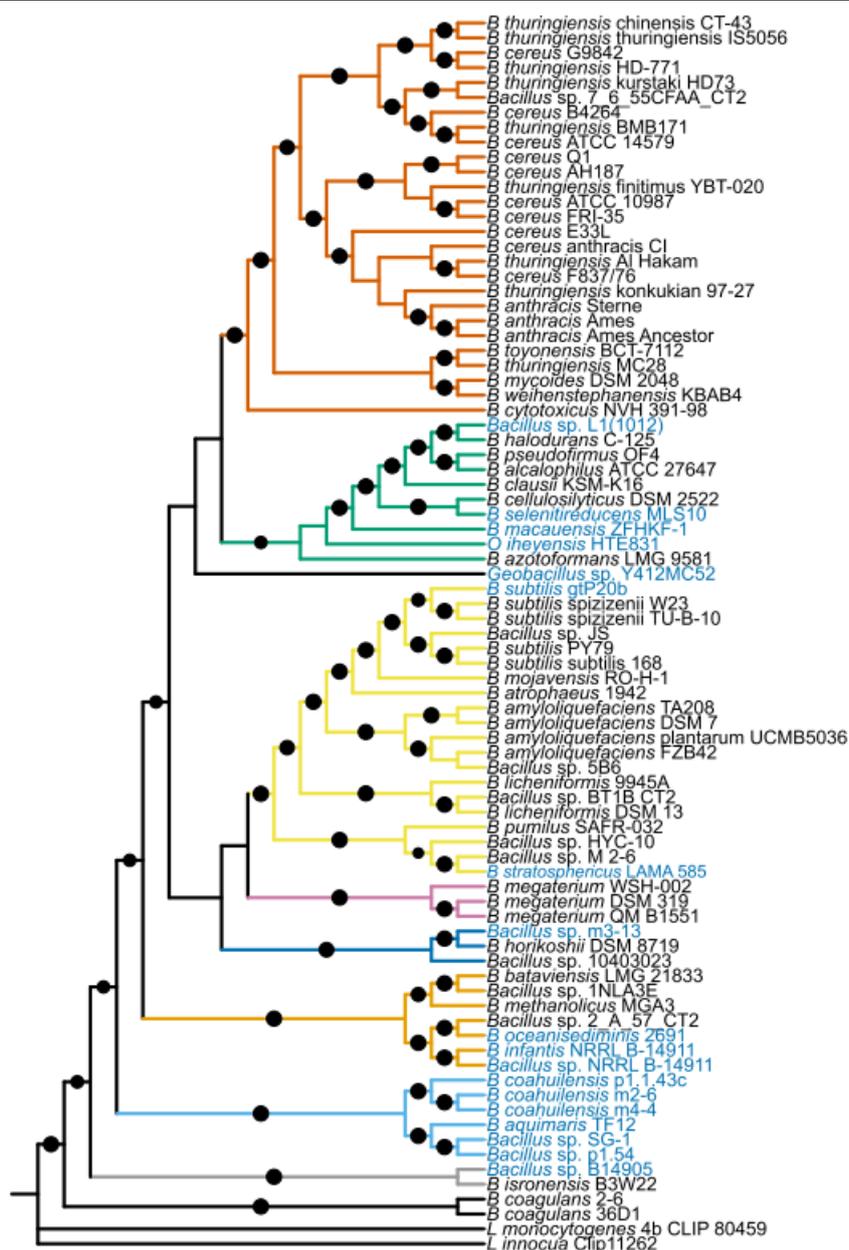


Figure 6. Φυλογενωμική ανάλυση του γένους *Bacillus*. Ο υπολογισμός του δέντρου μέγιστης πιθανοφάνειας έγινε με το πρόγραμμα PhyML και το μοντέλο LG+I+G+F αξιοποιώντας την υπερστοίχιση 196 πρωτεϊνών. Οι κλάδοι που έχουν υποστήριξη bootstrap πάνω από 80% εμφανίζονται με μαύρους κύκλους, ενώ με χρώμα απεικονίζονται οι 9 κλάδοι που χαρακτηρίζουν το γένος. Με μπλε γράμματα απεικονίζονται τα βακτήρια που απομονώθηκαν από υδάτινα οικοσυστήματα. Οι *L. monocytogenes* και *L. innocua* αποτελούν τις εξω-ομάδες του δέντρου. Η οπτικοποίηση του κλαδογράμματος έγινε στο iTOL (Hernández-González et al., 2018).

1.4. Ο κλάδος *B. altitudinis* - *B. pumilus* - *B. safensis*

Η ομάδα *B. pumilus* – *B. safensis* – *B. altitudinis* αποτελεί έναν κλάδο εντός του συμπλέγματος *B. subtilis* (Shivaji et al., 2006). Στην εργασία των (Fan et al., 2017) η ομάδα αναφέρεται ως ο τέταρτος κλάδος (clade IV) του συμπλέγματος *B. subtilis* όπως φαίνεται και στην Figure 4. Η πιο πρόσφατη

φυλογενωμική ανάλυση (Tirumalai et al., 2018) ανέδειξε την ύπαρξη των τριών μεγάλων υπο-ομάδων *pumilus*, *safensis* και *altitudinis*, οι οποίες εμφανίζουν πολύ υψηλή ομοιότητα σε επίπεδο 16S rRNA (Branquinho et al., 2014). Για το συγκεκριμένο λόγο έχουν γίνει αναφορές του εσφαλμένου χαρακτηρισμού στελεχών *B. safensis* ως *B. pumilus* (Branquinho et al., 2014; Lateef et al., 2015). Στην εργασία των (Branquinho et al., 2014) όπου αξιοποίησαν τα γονίδια 16S, *rpoB* και *gyrB* ανέφεραν ότι το ποσοστό των εσφαλμένων καταχωρήσεων στο σετ των οργανισμών που ανέλυσαν είναι πολύ υψηλό.

Συνολικά, τα μέλη της συγκεκριμένης ομάδας μπορούν να βρεθούν σε διάφορα εδάφη, σε ύδατα και σε υψηλά υψόμετρα και παράγουν σπόρια (Branquinho et al., 2014). Εμφανίζουν μεγάλο βιοτεχνολογικό ενδιαφέρον, ιδιαίτερα με εφαρμογές στη γεωργία καθώς εμφανίζουν θεραπευτική δράση σε μολυσμένα εδάφη μέσω της βιοαποδόμησης ρύπων (Kaur and Goyal, 2020) και αποτελούν παράγοντες ενίσχυσης της φυτικής ανάπτυξης. Η φυτοαυξητική δράση μπορεί να γίνεται είτε άμεσα μέσω παραγωγής αυξητικών παραγόντων και διαλυτοποίησης ανόργανων στοιχείων, είτε έμμεσα μέσω της άσκησης φυτοπροστασίας έναντι σε παθογόνα (Goswami and Deka, 2020; Lateef et al., 2015). Ένα επίσης ενδιαφέρον χαρακτηριστικό της ομάδας είναι η ιδιαίτερα υψηλή ανθεκτικότητα που εμφανίζουν σε συνθήκες υπερβολικής καταπόνησης όπως η υπεριώδης ακτινοβολία και άλλες κοινές πρακτικές αποστείρωσης (Satomi et al., 2006; Shivaji et al., 2006; Tirumalai et al., 2013).

1.4.1. *Bacillus altitudinis*

Το *Bacillus altitudinis* ανακαλύφθηκε σε σωλήνες που είχαν χρησιμοποιηθεί για τη συλλογή δειγμάτων αέρα σε ύψος 41 χιλιομέτρων και έτσι του δόθηκε το συγκεκριμένο όνομα που βασίζεται στη λέξη «altitude» δηλαδή υψόμετρο. Στελέχη του είδους έχουν την ικανότητα ανάπτυξης σε εργαστηριακές συνθήκες τόσο στους 8 όσο και τους 45 °C όσο και σε τιμές pH 5-8 (Shivaji et al., 2006), ενώ άλλα μπορούν να αντέχουν την ύπαρξη βαρέων μετάλλων στο ενδιαίτημα (Kumaravel et al., 2018). Επιπροσθέτως έχει γίνει αναφορά επιβίωσης σε συνθήκες υψηλής αλατότητας (μέχρι και 7.5%) και ξηρασίας (Kumaravel et al., 2018).

Όπως και οι κοντινοί συγγενείς του (*B. pumilus* – *B. safensis*), το *B. altitudinis* αποτελεί έναν ακόμα οργανισμό στον εκτενή κατάλογο των ριζικών βακτηρίων με ενίσχυση της φυτικής ανάπτυξης. Στελέχη όπως το *B. altitudinis* FD48 (Kumaravel et al., 2018) και το *B. altitudinis* MS16 (Goswami and Deka, 2020) εμφανίζουν αξιοσημείωτες ιδιότητες που επηρεάζουν την ανάπτυξη των φυτών τόσο άμεσα όσο και έμμεσα. Πιο συγκεκριμένα το FD48 διαλυτοποιεί τον ανόργανο φώσφορο, απαραίτητο για τους φυτικούς οργανισμούς, ενώ παράγει την απαμινάση της ένωσης 1-aminocyclopropane-1-carboxylate (ACC) που έχει άμεση επίδραση στην δημιουργία των φυματίων από τα ψυχανθή. Η ενίσχυση της δημιουργίας φυματίων δίνει σημαντική αύξηση στη βιομάζα του φυτού μέχρι και 40% (Kour et al., 2020)

Το στέλεχος *B. altitudinis* MS16 ασκεί την ευεργετική του δράση μέσω της παραγωγής χρήσιμων μορίων και την καταπολέμηση παθογόνων. Ένα τέτοιο μόριο είναι η αμμωνία που μάλιστα παράγεται σε

μεγάλες ποσότητες. Η παραγωγή της από βακτήρια θεωρείται ένα σημαντικό χαρακτηριστικό για την ανάπτυξη των φυτών καθώς τους παρέχει το άζωτο που χρειάζονται για την σύνθεση αζωτούχων ενώσεων (πρωτεΐνες, νουκλεϊκά οξέα) και επομένως την αύξηση της βιομάζας και του μεγέθους τους. Επίσης παράγει ινδολοξικό οξύ (IAA) μια ισχυρή φυτο-ορμόνη που διαδραματίζει σημαντικό ρόλο σε διάφορες αναπτυξιακές διαδικασίες. *In vitro* και *in planta* πειράματα σε σπόρους μουστάρδας έδειξαν ότι το συγκεκριμένο στέλεχος εμφανίζει ανταγωνιστική δράση έναντι αρκετών παθογόνων με ύψιστη δράση έναντι στο *S. sclerotiorum* (Goswami and Deka, 2020).

Τέλος, στελέχη του *B. altitudinis* μπορούν να χρησιμοποιηθούν για την βιο-αποδόμηση ρύπων με χαρακτηριστικό παράδειγμα το ζιζανιοκτόνο Butachlor (*N*-(butoxymethyl)-2-chloro-*N*-(2,6-diethylphenyl)acetamide). Το συγκεκριμένο χημικό χρησιμοποιείται ευρέως στην Ασία για την καταπολέμηση χορταριών που εμφανίζονται κάθε χρόνο και έχει χρόνο ημιζωής 1.5 έως 29 μέρες. Το συγκεκριμένο ζιζανιοκτόνο θεωρείται ένας νευροτοξικός παράγοντας με πιθανή καρκινογόνο δράση, ενώ έχει κατασταλτική δράση στην ανάπτυξη των σκουληκιών. Το στέλεχος *B. altitudinis* A16 απομονώθηκε από εδάφη με οργανικούς ρύπους (ανθρακόπισσα) και έχει την ικανότητα να το αποικοδομεί. Η αποικοδόμηση του ζιζανιοκτόνου μετρήθηκε μέσω της παραγωγής πέντε ενδιάμεσων μεταβολητών που μπορούν να οδηγήσουν σε πλήρη αποικοδόμηση προς διοξείδιο του άνθρακα και νερό (Kaur and Goyal, 2020).

1.4.2. *Bacillus pumilus*

Το *B. pumilus* αποτελεί ένα βακτήριο με πολλαπλές εφαρμογές στη γεωργία, στη φαρμακευτική και στη βιοτεχνολογία. Αποτελεί ένα βακτήριο με ικανότητα ενίσχυσης της φυτικής ανάπτυξης (Pérez-García et al., 2011), ενώ από το 2005 είναι γνωστή η χρήση *B. pumilus* σε σκευάσματα προβιοτικών που προορίζονται για χρήση από ανθρώπους, αν και προβιοτικά με άλλα βακτήρια *Bacillus* χρησιμοποιούνται και στην κτηνοτροφία. Η ικανότητα των συγκεκριμένων βακτηρίων να δημιουργούν πολύ ανθεκτικά σπόρια τους επιτρέπει να περάσουν το αφιλόξενο περιβάλλον του στομάχου και να βρεθούν στον εντερικό αυλό όπου και δρουν. Τα συγκεκριμένα προβιοτικά χρησιμοποιούνται κυρίως για θεραπευτικούς σκοπούς, μέσω της ενίσχυσης της ιθαγενούς εντερικής χλωρίδας λόγω της παραγωγής ενζύμων και βιταμινών. Παράλληλα δρουν ενάντια σε παθογόνα και ρυθμίζουν τη φλεγμονώδη αντίδραση της γαστρεντερικής οδού, καταπραΰνοντας συμπτώματα γαστρεντερικών διαταραχών (Elshaghabe et al., 2017; Hong et al., 2005).

Ωστόσο, στελέχη *B. pumilus* μπορεί να προκαλέσουν ασθένειες (Branquinho et al., 2014). Πιο συγκεκριμένα έχουν συσχετιστεί με τροφικές δηλητηριάσεις, ειδικά σε ρύζι το οποίο έχει μαγειρευτεί και ξαναζεσταθεί (From et al., 2007). Οι τοξίνες που αναφέρεται να προκαλούν τις δηλητηριάσεις αυτές είναι οι πουμιλασιδίνες (pumilacidins) (From et al., 2007). Επίσης έχουν συσχετιστεί με δερματικές αλλοιώσεις που παρομοιάζουν την μόλυνση από άνθρακα (*Bacillus anthracis*) (Branquinho et al., 2014).

Στελέχη του βακτηρίου *Bacillus pumilus* έχουν απομονωθεί σε εργαστήρια και εγκαταστάσεις της NASA που συγκαταλέγονται στα πιο αυστηρά ελεγχόμενα και αποστειρωμένα περιβάλλοντα του κόσμου. Η ικανότητα ορισμένων ενδοσπορίων να αντιστέκονται σε πολλούς χημικούς παράγοντες και διαδικασίες αποστείρωσης αποτελεί μια μεγάλη πρόκληση στην καταπολέμηση των επιμολύνσεων. Για το λόγο αυτό θεωρείται ένας οργανισμός που μπορεί να δημιουργήσει προβλήματα σε διαστημικές αποστολές με πιθανή επιμόλυνση εξωγήινων εδαφών (Tirumalai et al., 2013). Ένα από τα παραπάνω στελέχη, που ονομάζεται SAFR-032, αναφέρεται πως είναι ικανό να αντέχει σε ακραία αλκαλικά περιβάλλοντα με pH 11, υψηλές θερμοκρασίες μέχρι και 65 °C, υπεριώδη ακτινοβολία, το κενό του διαστήματος και προσομοιωμένες συνθήκες της επιφάνειας του Άρη. (Tirumalai et al., 2013).

1.4.3. *Bacillus safensis*

Το *Bacillus safensis* αποικίζει τόσο χερσαία όσο και υδάτινα περιβάλλοντα (Branquinho et al., 2014), ενώ το πρώτο στέλεχος απομονώθηκε σε μια μονάδα συναρμολόγησης διαστημόπλοιων (space-assembly facility – SAF) από όπου πήρε και το όνομα του. Το συγκεκριμένο περιβάλλον είναι ιδιαίτερα φιλόξενο προς τους μικροοργανισμούς καθώς έχει σχεδόν μηδενικές παροχές θρεπτικών υλικών και μεγάλη ξηρότητα ενώ η απολύμανση του γίνεται με υπεροξειδίο του υδρογόνου (Satomi et al., 2006). Αξιοσημείωτη επίσης είναι η ικανότητα που έχει να επιβιώνει σε συνθήκες υψηλής συγκέντρωσης αλάτων (1-25%), σε pH 4-9, ενώ άλλα στελέχη φέρουν ανθεκτικότητα σε αρσενικό και βαρέα μέταλλα. Συνολικά πιστεύεται ότι η ανθεκτικότητα σε μεγάλο εύρος περιβαλλόντων τα οποία είναι φιλόξενα για την επιβίωση πολλών μικροοργανισμών οφείλεται σε ειδικά γενετικά χαρακτηριστικά τους και την ικανότητα δημιουργίας ανθεκτικών σπορίων (Lateef et al., 2015).

Τα συγκεκριμένα βακτήρια είναι και σημαντικοί παραγωγοί ενζύμων και δευτερογενών μεταβολιτών που χρησιμοποιούνται στη βιομηχανία. Πιο συγκεκριμένα μπορούν να παράγουν ένζυμα όπως αμυλάση, κυτταρινάση, πρωτεάσες και λιπάσες που χρησιμοποιούνται σε πολλαπλές διεργασίες αλλά και πιο εξειδικευμένα όπως η κερατινάση που χρησιμοποιείται στην αποδόμηση φτερών, τριχών και την απομάκρυνση λεκέδων (Lateef et al., 2015).

Επίσης, το *B. safensis* χρησιμοποιείται ως παράγοντας ενίσχυσης της φυτικής ανάπτυξης, αυξάνοντας την αντοχή των φυτών σε αβιοτικές καταπονήσεις όπως την ξηρασία σε κάκτους (Kanamura, 2013) και σε πολλαπλές ποικιλίες σιταριού (Lateef et al., 2015). Ωστόσο η ενισχυτική ικανότητά του βασίζεται και στην απομάκρυνση φυτικών παθογόνων και κυρίως μυκήτων που προκαλούν μεγάλες ζημιές στις καλλιέργειες. Σε αντίθεση με τα εξωγενώς εφαρμοζόμενα φάρμακα που είναι ιδιαίτερα μεγάλη παρεμβολή στο οικοσύστημα και δεν είναι τόσο αποτελεσματικά, η εφαρμογή ειδικών στελεχών *B. safensis* προστατεύει τα φυτά έναντι σε παθογόνα μέσω της παραγωγής αντιβακτηριδιακών, αντιμυκητιακών και αντιικών παραγόντων (Lateef et al., 2015).

Τέλος έχουν γίνει αναφορές για την εκμετάλλευση του *B. safensis* σε συνδυασμό με άλλους

οργανισμούς για την απορρύπανση μολυσμένων εδαφών από ρύπους όπως βαρέα μέταλλα, τοξικά χημικά (υδροκυάνιο) και αντιβιοτικά. Η γρήγορη και επιτυχής απομάκρυνση τέτοιων ρύπων είναι ύψιστης σημασίας καθώς έχουν τοξικές και μεταλλαξιγόνες επιπτώσεις στους οργανισμούς, ενώ τα αντιβιοτικά προάγουν την ανάπτυξη ανθεκτικών στελεχών (Lateef et al., 2015).

1.4.4. Εξελικτικές αναλύσεις του κλάδου *B. altitudinis* - *B. pumilus* - *B. safensis*

Δύο αναλύσεις της εξελικτικής γραμμής με πολλαπλά στελέχη (Liu et al., 2013; Tirumalai et al., 2018) έχουν αναδείξει την ύπαρξη των τριών μεγάλων ομάδων που προαναφέρθηκαν ως ξεχωριστές. Επίσης αναφέρουν τους *B. pumilus* και *B. safensis* ως πιο γενετικά όμοιους και τον *B. altitudinis* να είναι πιο βασικός κλάδος στην ομάδα.

Οι (Liu et al., 2013) χρησιμοποίησαν τα 7 housekeeping γονίδια *gyrB-rpoB-rycA-pyrE-mutL-aroE-trpB* από 76 θαλάσσια στελέχη και 3 τυπικά για τον υπολογισμό φυλογενετικών δέντρων από τα επιμέρους γονίδια καθώς και την υπερστοίχιση τους. Όλα αυτά τα δέντρα είχαν παρόμοια τοπολογία με μικρές μεταβολές. Η φυλογενετική ανάλυση του γονιδίου *gyrB* θαλάσσιων και άλλων στελεχών από άλλες πηγές έδειξε ότι τα βακτήρια που απομονώνονται από υδάτινα περιβάλλοντα σχηματίζουν επιμέρους συστάδες στους κλάδους και τα περισσότερα ανήκουν στην ομάδα *B. altitudinis*. Σε αντίθεση ο κλάδος *B. safensis* αποτελείται κυρίως από βακτήρια χερσαίας προέλευσης με ποσοστό χερσαίων προς θαλάσσιων στελεχών 3/1.

Η πιο πρόσφατη ανάλυση (Tirumalai et al., 2018) συμπεριέλαβε πολλαπλά στελέχη *B. pumilus*, *B. safensis* και τέσσερα *B. altitudinis*, συνδυαστικά με ένα στέλεχος *Geobacillus kaustophilus*, ένα στέλεχος *B. subtilis* και ένα στέλεχος *E. coli* ως εξω-ομάδες (Figure 7). Για τη φυλογένεση χρησιμοποίησαν τη μεθοδολογία Genome – BLAST Distance Phylogeny (GBDP). Συνοπτικά οι γονιδιωματικές ακολουθίες από πλήρη ή μερικώς αλληλουχημένα γονιδιώματα συγκρίνονται ανά ζεύγη μέσω του αλγορίθμου BLAST και υπολογίζεται μια συνολική απόσταση μεταξύ του κάθε ζεύγους όπως περιγράφεται από τους (Meier-Kolthoff et al., 2013). Στη συνέχεια οι αποστάσεις αυτές χρησιμοποιούνται για τον υπολογισμό ενός φυλογενετικού δέντρου. Η παραπάνω ανάλυση βρίσκεται σε συμφωνία με αναλύσεις μικρότερου βεληνεκού (λίγων γονιδίων). Επιπλέον, στελέχη *B. pumilus/B. safensis* απαρτίζουν μια φυλογενετική ομάδα με 3 υπο-συστάδες, ενώ η ομάδα των *B. altitudinis* είναι πιο βασική και αποτελείται από διάφορα στελέχη.

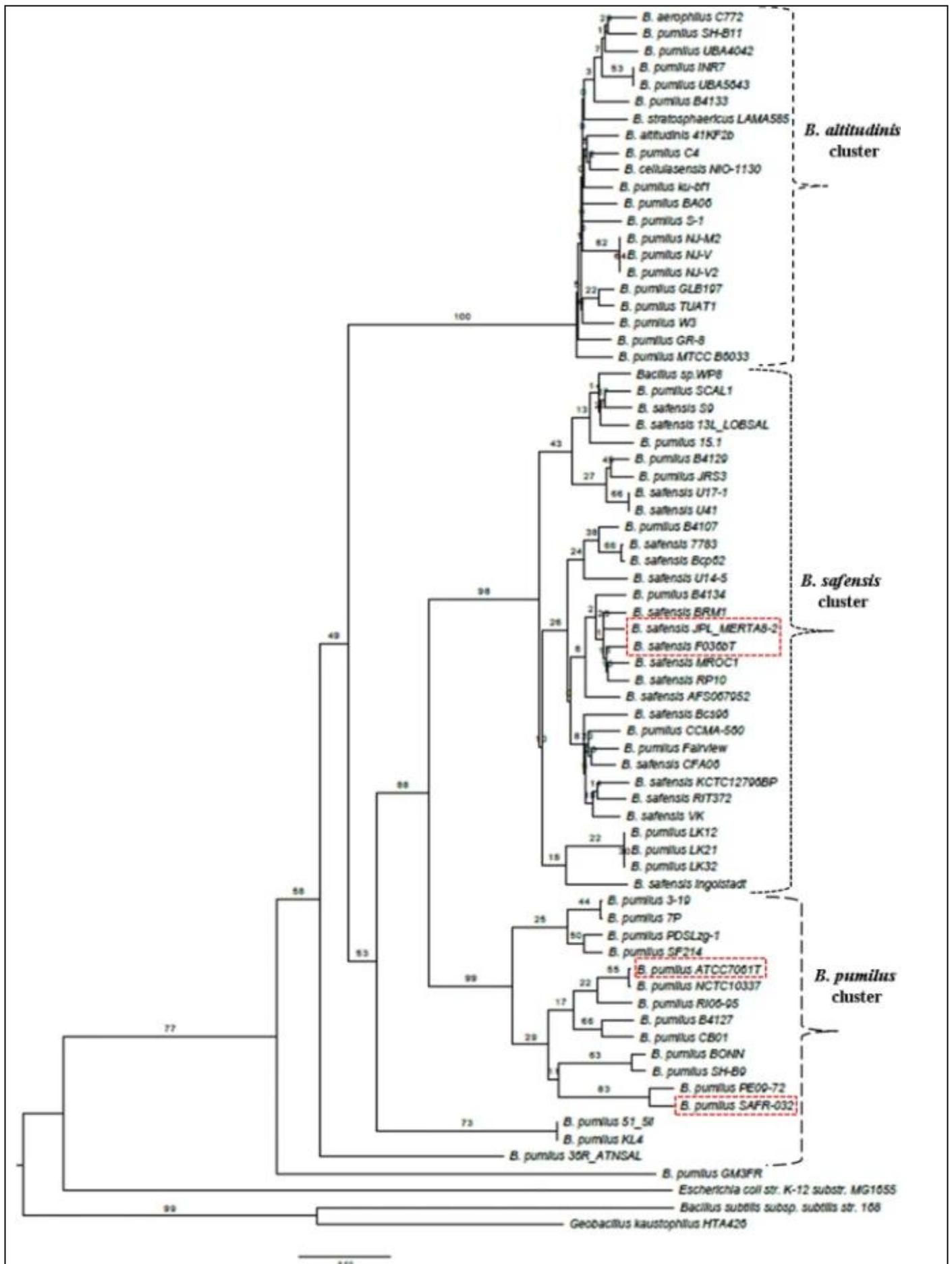


Figure 7. Φυλογενωμική ανάλυση του κλάδου *B. altitudinis* – *B. rumilus* – *B. safensis* με τη μέθοδο Genome-BLAST Distance Phylogeny (GBDP). Τα αποτελέσματα της ανάλυσης βρίσκονται σε συμφωνία με αναλύσεις μεμονωμένων ή λίγων γονιδίων. Σε κόκκινα τετράγωνα πλαισιώνονται 3 στελέχη που εμφανίζουν υψηλή ανθεκτικότητα έναντι στην υπεριώδη ακτινοβολία. Το στέλεχος *B. rumilus* ATCC7051T είναι το τυπικό (Tirumalai et al., 2018).

Σκοπός της συγκεκριμένης εργασίας είναι η αξιοποίηση των διαθέσιμων γονιδιωμάτων του γένους *Bacillus* για την πραγματοποίηση φυλογενωμικών και συγκριτικών αναλύσεων ώστε να αποσαφηνιστούν οι πρωτεΐνες πυρήνας του κλάδου *B. altitudinis* - *B. rumilus* - *B. safensis* και τα μοριακά αποτυπώματα που τον διαχωρίζουν από υπόλοιπα είδη του γένους και πιθανώς να εμπλέκονται στην εμφάνιση των ιδιαίτερων ιδιοτήτων ανθεκτικότητας των διάφορων στελεχών του.

2. ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ

2.1. Λογισμικό

2.1.1. Linux Ubuntu 20

Τα Linux Ubuntu αποτελούν ένα ασφαλές λειτουργικό σύστημα ανοιχτού λογισμικού για υπολογιστές που διατίθεται δωρεάν για τους χρήστες με γραφικό περιβάλλον GNOME. Το βασικό περιβάλλον τερματικού των Ubuntu χρησιμοποιεί το BASH (Bourne Again Shell) και ενδείκνυται για τη διαχείριση μεγάλου όγκου δεδομένων και την εκτενή αυτοματοποίηση διεργασιών (“Enterprise Open Source and Linux,” n.d.).

2.1.2. Python 3.6

Η Python είναι μια ευρέως διαδομένη αντικειμενοστρεφής (object – oriented) γλώσσα προγραμματισμού παρόμοια με τη Ruby (scripting languages). Διαθέτει πλήθος βιβλιοθηκών που την καθιστούν μια γλώσσα προγραμματισμού με τεράστιο εύρος εφαρμογών (“Welcome to Python.org,” n.d.). Η συγκεκριμένη γλώσσα φέρει βιβλιοθήκες για επιστημονική έρευνα και δη βιολογική – βιοπληροφορική έρευνα όπως για παράδειγμα τη biopython, τη pandas και τη scipy.

2.1.3. Seaview 5

Το λογισμικό Seaview είναι γραμμένο σε C++, προσφέρεται για όλα τα λειτουργικά συστήματα (Linux, Windows, MacOS) και έχει σχεδιαστεί για να πραγματοποιεί πολλαπλές στοιχίσεις ακολουθιών DNA ή πρωτεϊνών και να δημιουργεί φυλογενετικά δέντρα. Το συγκεκριμένο λογισμικό πρωτοεκδόθηκε ως γραφικό περιβάλλον (Graphical User Interface – GUI) αλλά υποστηρίζει και διεργασίες σε περιβάλλον τερματικού (Command Line Interface – CLI). Για τις πολλαπλές στοιχίσεις δίνεται η δυνατότητα χρήσης των αλγορίθμων Clustal Omega και MUSCLE που υπάρχουν ενσωματωμένα ή και εξωτερικών μεθόδων. Ο αλγόριθμος MUSCLE χρησιμοποιήθηκε στη παρούσα εργασία. Ο υπολογισμός των φυλογενετικών δένδρων μπορεί να γίνει είτε με μεθόδους αποστάσεων και πιο συγκεκριμένα με τον αλγόριθμο BioNJ είτε με μεθόδους χαρακτήρων, όπως η μέγιστη φειδωλότητα (Maximum Parsimony) ή η μέγιστη πιθανοφάνεια (Maximum Likelihood – Αλγόριθμος PhyML) (Gouy et al., 2010).

2.1.4. PubMLST

Για το διαχωρισμό των μικροοργανισμών σε συνδυασμό με τη βιβλιοθήκη ryan1 χρησιμοποιήθηκε η μέθοδος PubMLST (multi-locus sequence typing). Η συγκεκριμένη μέθοδος βασίζεται σε χαρακτηρισμούς μικροοργανισμών από τμήματα πολλαπλών γονιδίων (Maiden et al., 2013) και έχει σημαντικό πλεονέκτημα διακριτικής ικανότητας έναντι τις κλασσικής ταξινόμησης με 16S rRNA όπως

περιγράφηκε νωρίτερα και φαίνεται στην Figure 3. Η επερώτηση των γονιδιωμάτων προς αναγνώριση έγινε μέσω της διεπαφής (API) του PubMLST rest api στη βάση δεδομένων BIGSdb (Jolley et al., 2017).

2.1.5. Pyani

Το pyani είναι ένα πρόγραμμα και πακέτο της python3 που βοηθά στον υπολογισμό της μέσης ομοιότητας σε νουκλεοτιδικό επίπεδο (average nucleotide identity – ANI) προκειμένου να είναι εφικτή η ακριβής σύγκριση διαφόρων γονιδιωμάτων. Για τη στοίχιση των γονιδιωμάτων και τον υπολογισμό της ομοιότητας χρησιμοποιεί το BLAST (ANIb) ή το MUMmer (ANIm) με την προσέγγιση «όλα εναντίων όλων» όπου υπολογίζεται η ομοιότητα για κάθε δυνατό ζευγάρι γονιδιωμάτων. Τα τελικά αποτελέσματα των συγκρίσεων γράφονται σε αρχεία tsv. Όπου είναι δυνατή γίνεται χρήση πολλαπλών πυρήνων επεξεργασίας (Pritchard et al., 2016).

2.1.6. Treedyn

Το λογισμικό Treedyn αποτελεί μια εφαρμογή οπτικοποίησης, επεξεργασίας και εισαγωγής πληροφοριών μεμονομένων ή πολλαπλών φυλογενετικών ή φυλογενωμικών δέντρων. Η εφαρμογή είναι γραμμένη στη γλώσσα προγραμματισμού Tcl/Tk και είναι ελεύθερα διαθέσιμη για λειτουργικά συστήματα Mac OS, Windows και Linux. Από τις βασικότερες λειτουργίες του Treedyn είναι η εισαγωγή λειτουργικών χαρακτηρισμών και γραφικών στους κλάδους κάθε δέντρου και ο χρωματισμός ανάλογα με τα κριτήρια που επιλέγει ο χρήστης. Επίσης διαθέτει λειτουργίες σύγκρισης κοινών κλάδων μεταξύ δέντρων. Συνολικά αποτελεί ένα πλήρες εργαλείο με απλό χειρισμό και δυνατότητα αποθήκευσης των τελικών αποτελεσμάτων σε διάφορους τύπους αρχείων γραφικών όπως postscript (.ps) αλλά και αρχεία εικόνων (.png, .jpg) (Chevenet et al., 2006).

2.1.7. Λογισμικό αναγνώρισης πρωτεώματος πυρήνα

Για την αναγνώριση του πρωτεώματος πυρήνα των οργανισμών υπό εξέταση χρησιμοποιήθηκε μια σειρά διεργασιών (pipeline) που είχε αναπτυχθεί και εφαρμοσθεί σε βακτήρια του γένους *Pseudomonas* (Nikolaidis et al., 2020). Η μέθοδος βασίζεται στην αναγνώριση ορθολόγων μέσω ανταποδοτικού BLAST και περαιτέρω φιλτράρισμα των αποτελεσμάτων με εξεζητημένα κριτήρια, όπως περιγράφεται στην εισαγωγή. Ανάλογα με την ταξινομική ομάδα που μελετήθηκε χρησιμοποιήθηκαν συγκεκριμένα στελέχη ως σημεία αναφοράς.

2.1.8. Λειτουργικός χαρακτηρισμός πρωτεϊνών

Οι πρωτεΐνες πυρήνας για κάθε ταξινομική ομάδα που μελετήθηκε με βάση τα στελέχη αναφοράς ελέγχθηκαν ως προς την λειτουργία τους με το EGGNOG database v4.5 μέσω του ενσωματωμένου εργαλείου eggNOG mapper (Huerta-Cepas et al., 2016).

2.2. Λήψη πρωτεωμάτων

Για την παρούσα εργασία χρησιμοποιήθηκαν 797 πρωτεώματα βακτηρίων του γένους *Bacillus* (NCBI taxonomy ID: 1386) και πιο συγκεκριμένα όπως φαίνεται στον πίνακα Table 1. Τα πρωτεώματα λήφθηκαν από τη βάση δεδομένων NCBI Assembly το Σεπτέμβριο του 2020 με κριτήρια η συναρμολόγηση να ήταν σε επίπεδο πλήρους γονιδιώματος ή χρωμοσώματος. Σκοπός ήταν η χρήση ύψιστης ποιότητας δεδομένων, καθώς η ενσωμάτωση κακής ποιότητας συναρμολογήσεων μπορεί να επιφέρει δραματικά αποτελέσματα στον τελικό αριθμό των γονιδίων πυρήνα.

Table 1. Ο συνολικός αριθμός των πρωτεωμάτων που χρησιμοποιήθηκαν στην ανάλυση ανά φυλογενωμική ομάδα

Phylogenomic group	Number of organisms
<i>B. subtilis</i>	159
<i>B. atrophaeus</i>	6
<i>B. amyloliquefaciens</i>	10
<i>B. licheniformis</i>	32
<i>B. paralicheniformis</i>	10
<i>B. velezensis</i>	161
<i>B. altitudinis</i>	19
<i>B. pumilus</i>	7
<i>B. safensis</i>	8
<i>B. subtilis lineage (rest)</i>	28
<i>B. cereus - B. thuringiensis</i>	102
<i>B. anthracis</i>	97
<i>B. mycoides</i>	9
<i>B. pseudomycoides</i>	5
<i>B. cytotoxicus</i>	11
<i>B. cereus lineage (rest)</i>	46
<i>B. megaterium</i>	18
<i>B. coagulans</i>	10
<i>Bacillus (other)</i>	59
Total	797

3. ΑΠΟΤΕΛΕΣΜΑΤΑ - ΣΥΖΗΤΗΣΗ

Η φυλογενωμική ανάλυση της ομάδας *B. altitudinis* - *B. pumilus* - *B. safensis* πραγματοποιήθηκε με 289 πρωτεΐνες που αποτελούν πυρήνα του συμπλέγματος *Bacillus subtilis sensu lato* όπως φαίνεται στην Figure 8. Για την ορθή οριοθέτηση των ειδών και υποομάδων χρησιμοποιήθηκε το όριο της τιμής ANIm 95% και ο υπολογισμός της έγινε με 3 στελέχη αναφοράς από κάθε είδος. Σύμφωνα με τα παραπάνω διακρίνονται 3 κύριες υποομάδες μια για κάθε ένα από τα είδη *B. altitudinis* (19 οργανισμοί), *B. pumilus* (7 οργανισμοί) και *B. safensis* (8 οργανισμοί). Η πρώτη υποομάδα έχει κυρίως αντιπροσώπους *B. altitudinis* μαζί με ορισμένους *B. pumilus* και έναν *B. aerophilus*, ενώ οι άλλες είναι μονοφυλετικές. Επίσης φαίνεται ότι οι *B. pumilus* αποτελούν μια ομάδα με εξαίρεση τον *B. pumilus* με κωδικό GCF_003431975 που έρχεται πιο βασικός και η τιμή ANIm του είναι 93%, οπότε θεωρείται ένα ξεχωριστό είδος. Οι αμέσως κοντινότεροι συγγενείς είναι οι *B. safensis* που σχηματίζουν έναν μονοφυλετικό κλάδο. Οι *B. pumilus* – *B. safensis* σχηματίζουν έναν ξεχωριστό κλάδο από τους *B. altitudinis* που περιέχει συνολικά 3 είδη. Τα παραπάνω αποτελέσματα έρχονται σε συμφωνία και με τις αναλύσεις των (Liu et al., 2013; Tirumalai et al., 2018). Τέλος διακρίνεται ένας κλάδος με 2 γονιδιώματα *B. xiamenensis*, τα οποία δεν έχουν χρησιμοποιηθεί στις προηγούμενες αναλύσεις. Αν και οι *B. xiamenensis* έχουν διαφοροποιηθεί από τους κοντινούς τους *B. altitudinis*, δεν διακρίνονται ως ξεχωριστή υποομάδα.

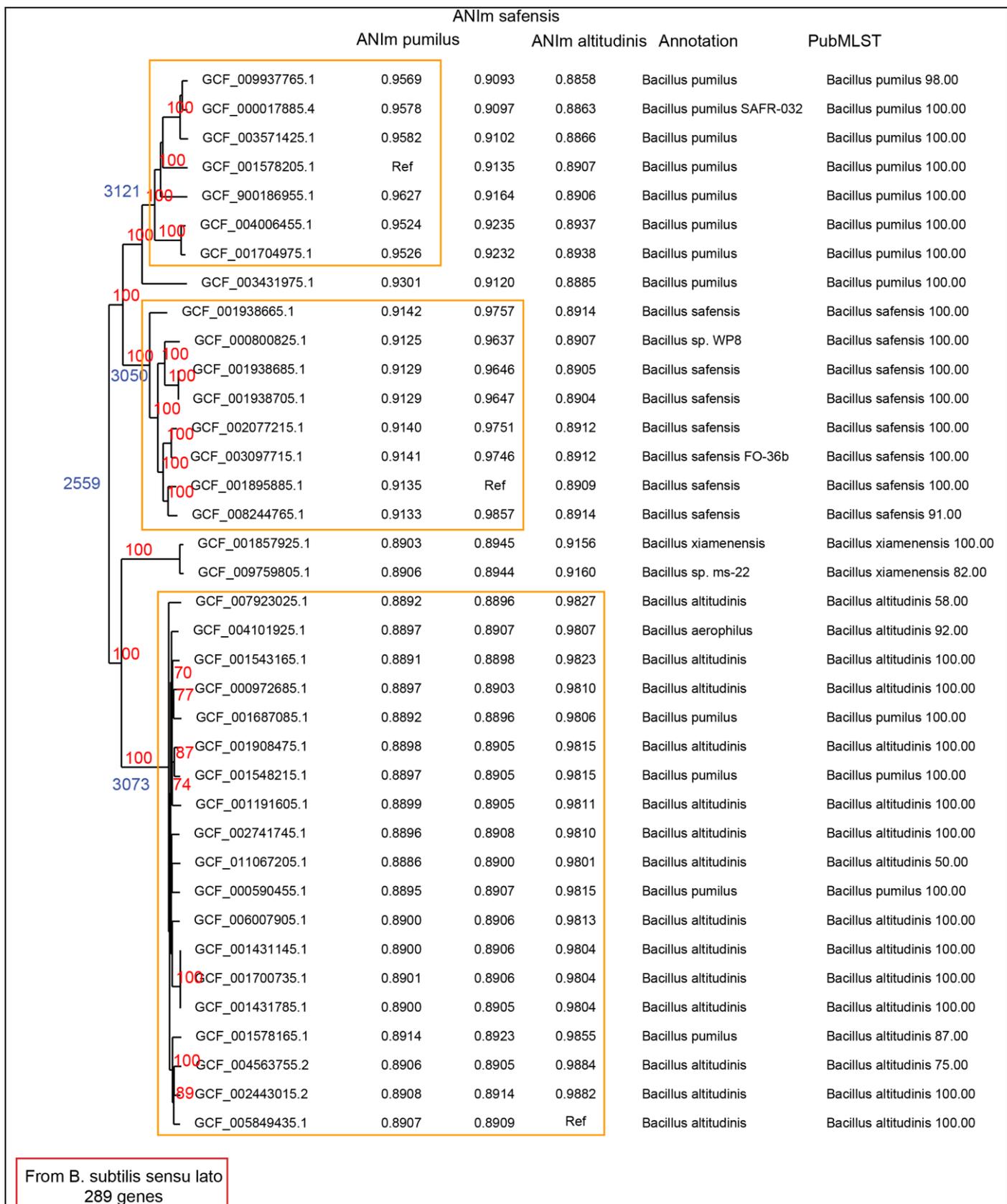


Figure 8. Φυλογενωμική ανάλυση της ομάδας *B. altitudinis* - *B. pumilus* - *B. safensis*. Για τον υπολογισμό του δέντρου χρησιμοποιήθηκε η υπερστοίχιση από 289 πρωτεΐνες που αποτελούν τις πρωτεΐνες πυρήνα του κλάδου *B. subtilis sensu lato*, ο αλγόριθμος BioNJ και 500 επαναλήψεις bootstrap. Για τον υπολογισμό των ANIm χρησιμοποιήθηκαν τα στελέχη με τιμή Ref ως σημεία

αναφοράς. Με μπλε γράμματα αναγράφεται ο αριθμός των πρωτεϊνών πυρήνα (σκληρός πυρήνας) για κάθε κλάδο, ενώ με κόκκινο η επί τοις εκατό υποστήριξη bootstrap. Αναγνωρίζονται οι 3 βασικές υποομάδες σε συμφωνία με την υπάρχουσα βιβλιογραφία με τους κλάδους *B. pumilus* και *B. safensis* να είναι οι πιο κοντινοί. Η τιμή ANIm 95% χρησιμοποιήθηκε για το διαχωρισμό των ειδών (πορτοκαλί πλαίσια).

Όσον αφορά τις πρωτεΐνες πυρήνα, πραγματοποιήθηκαν αναλύσεις τόσο σε ολόκληρο τον κλάδο όσο και στις διάφορες υποομάδες, με τα στελέχη αναφοράς να είναι εκείνα που αναγράφονται στην Figure 8. Το στέλεχος αναφοράς για τον υπολογισμό των πρωτεϊνών πυρήνα ολόκληρου του κλάδου είναι το *B. altitudinis* με κωδικό GCF_005849435. Οι πρωτεΐνες που υπάρχουν σε όλους τους οργανισμούς της ομάδας αναφέρονται ως «σκληρός πυρήνας – hard core», ενώ οι πρωτεΐνες που υπολογίζονται να υπάρχουν σε όλους τους οργανισμούς εκτός ενός αναφέρονται ως «χαλαρός πυρήνας – soft core». Για ολόκληρο τον κλάδο αναγνωρίστηκαν 2559 πρωτεΐνες ως «σκληρός» πυρήνας και 2893 ως «χαλαρός» πυρήνας όπως φαίνεται και στην Figure 9. Κατά αντιστοιχία για την ομάδα *B. altitudinis* αναγνωρίστηκαν 3073 και 3270, για την ομάδα *B. pumilus* 3121 και 3253, ενώ για τον *B. safensis* 3050 και 3265. Όπως είναι αναμενόμενο η εφαρμογή χαλαρότερων κριτηρίων οδηγεί στην αναγνώριση περισσότερων πρωτεϊνών.

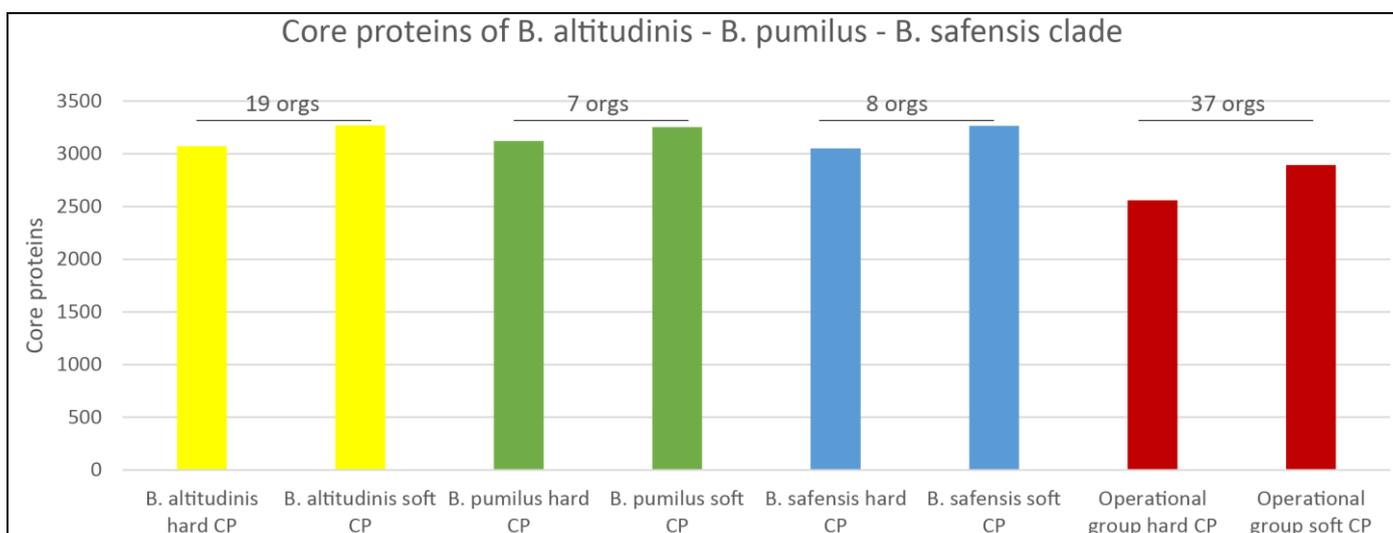


Figure 9. Αριθμός των πρωτεϊνών πυρήνα για τον κλάδο *B. altitudinis* – *B. pumilus* – *B. safensis* και των υποομάδων που χαρακτηρίστηκαν. Ο συνολικός κλάδος αναφέρεται ως operational group. Ως “hard CP” αναφέρονται οι πρωτεΐνες πυρήνας (CP; Core Proteins) που υπάρχουν σε όλα τα στελέχη της ομάδας που μελετάται, ενώ ως “soft CP” οι πρωτεΐνες που υπάρχουν σε όλους εκτός από έναν οργανισμό. Πάνω από τις μπάρες αναφέρονται το πλήθος των οργανισμών που απαρτίζουν τη συγκεκριμένη ταξινομική ομάδα.

Εν συνέχεια, οι πρωτεΐνες πυρήνας του κλάδου *B. altitudinis* – *B. pumilus* – *B. safensis* χαρακτηρίστηκαν ως προς τη λειτουργία τους (Figure 10). Ο λειτουργικός χαρακτηρισμός έγινε μέσω της

βάσης δεδομένων eggNOG v4.5 (Huerta-Cepas et al., 2016). Αναγνωρίστηκαν οι έξι κύριες κατηγορίες που έχουν όλες, πλην μιας, ποσοστό περισσότερο από 5%. Στις “hard core proteins” συγκαταλέγονται οι 2559 πρωτεΐνες του κλάδου ενώ στις “soft core” συγκαταλέγονται οι 2893 πρωτεΐνες που προαναφέρθηκαν. Οι πρωτεΐνες πυρήνα του κλάδου στη πλειοψηφία τους δεν έχουν γνωστό λειτουργικό χαρακτηρισμό, ενώ οι υπόλοιπες εμπλέκονται στις βασικές κυτταρικές λειτουργίες της μεταγραφής και της μετάφρασης, στο μεταβολισμό υδατανθράκων και αμινοξέων και την παραγωγή ενέργειας. Ωστόσο οι αμέσως επόμενες κατηγορίες σε πλήθος πρωτεϊνών (4.6 – 4.87 % του συνόλου) αφορούν τη μεταφορά και μεταβολισμό ανόργανων ιόντων και τη δημιουργία κυτταρικού τοιχώματος.

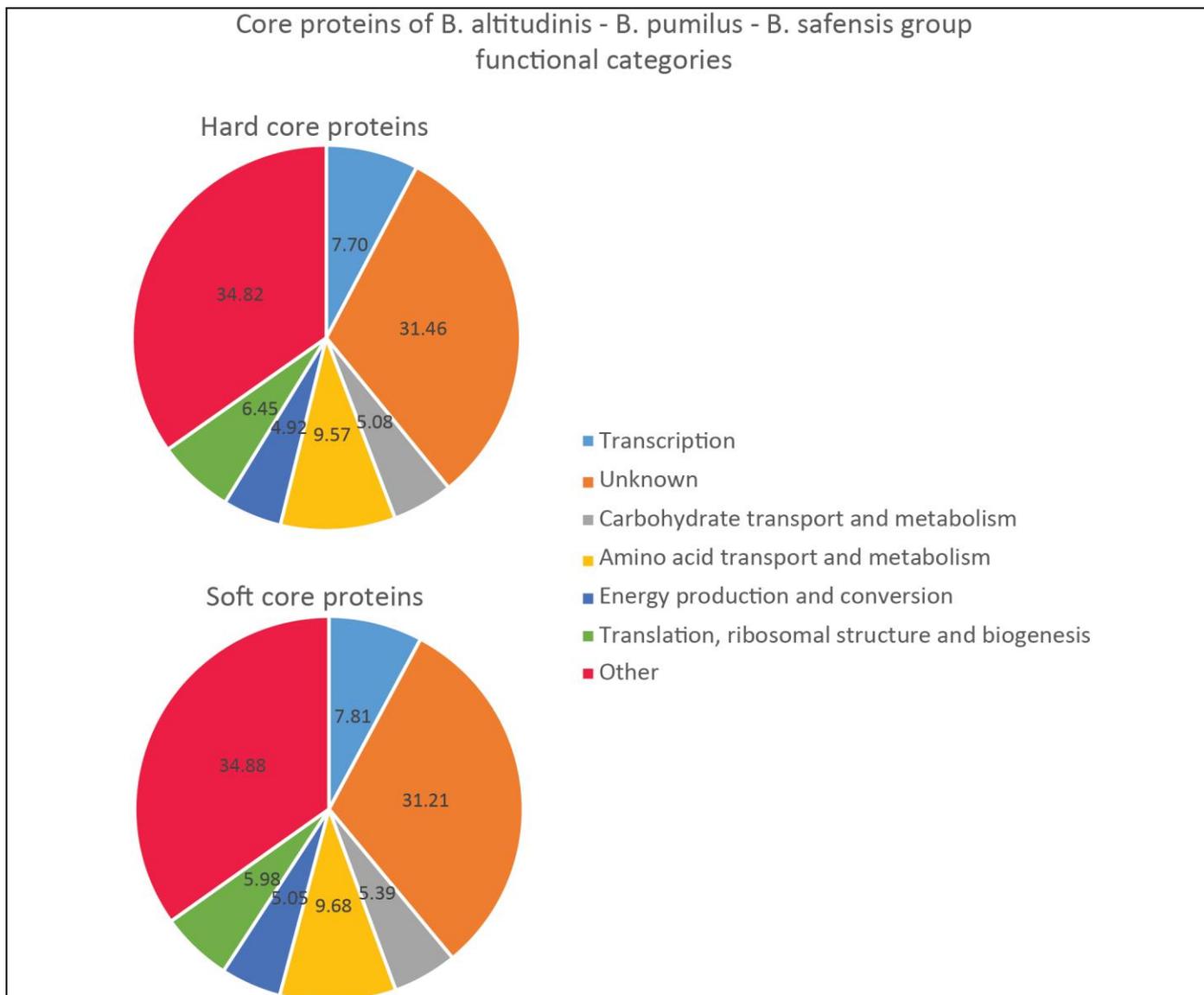


Figure 10. Κύριες λειτουργικές κατηγορίες των πρωτεϊνών πυρήνα του κλάδου *B. altitudinis* – *B. pumilus* – *B. safensis*. Ως “hard core proteins” αναφέρονται οι πρωτεΐνες που υπάρχουν σε όλους τους οργανισμούς του κλάδου ενώ ως «soft core» οι πρωτεΐνες που υπάρχουν σε όλους εκτός από έναν οργανισμό (2893). Για το λειτουργικό χαρακτηρισμό χρησιμοποιήθηκε η βάση δεδομένων eggNOG v4.5 (Huerta-Cepas et al., 2016). Οι πρωτεΐνες πυρήνα του κλάδου στη πλειοψηφία τους δεν έχουν γνωστό λειτουργικό χαρακτηρισμό, ενώ οι υπόλοιπες αφορούν βασικές κυτταρικές διεργασίες όπως ο

Ως μοριακό αποτύπωμα (fingerprint) ορίζουμε μια πρωτεΐνη που υπάρχει σε όλα τα στελέχη εντός της ομάδας ενδιαφέροντος και σε κανένα από τα υπόλοιπα του γένους που συμπεριλαμβάνονται στην ανάλυση. Για τις τρεις υποομάδες δεν βρέθηκαν πρωτεΐνες αποτυπώματα, αναδεικνύοντας την γενετική ομοιότητα που υπάρχει μεταξύ τους. Ωστόσο για ολόκληρο τον κλάδο αναγνωρίστηκαν 26 πρωτεΐνες αποτυπώματα. Η συντριπτική πλειοψηφία αυτών αναφέρονται ως υποθετικές. Συνολικά οι υποθετικές πρωτεΐνες είναι 20 από τις 26, δηλαδή το 77%. Οι υπόλοιπες 6 είναι χαρακτηρισμένες ως ένας *helix-turn-helix* μεταγραφικός παράγοντας, μια *YozQ family protein*, μια *YjcZ family sporulation protein*, μια πρωτεΐνη της οικογένειας ROK (Repressor, ORF, Kinase), μια λιγάση βιοτίνης και μια πρωτεΐνη ρυθμιστής απόκρισης (response regulator).

Οι πρωτεΐνες *YozQ* και *YjcZ* αποτελούν πρωτεΐνες απαραίτητες για τη διαδικασία της σποριογένεσης στους *Bacillus*. Η *YozQ* στον *B. subtilis 168* αποτελεί μια πρωτεΐνη 97 αμινοξέων, ενώ η πρωτεΐνη της συγκεκριμένης ομάδας έχει μήκος μόλις 45 αμινοξέα. Η *YjcZ* αποτελεί ένα από τα πολλά παράλογα (6) που υπάρχουν σε αντιπροσώπους της φυλογενετικής ομάδα και μάλιστα το μικρότερο με μήκος 29 αμινοξέων. Πιθανόν το μικρό τους μήκος να αποτελεί μια ειδική προσαρμογή, αλλά δεν υπάρχουν επαρκή στοιχεία που να στηρίζουν τη συγκεκριμένη εικασία.

Η πρωτεΐνη λιγάση της βιοτίνης αποτελεί ομόλογο του γονιδίου *birA* της *E.coli*. Η βιοτίνη αποτελεί ενζυμικό συμπ παράγοντα πολλαπλών ενζύμων και το συγκεκριμένο γονίδιο είναι υπεύθυνο για την τροφοδότηση της βιοτίνης στο μεταβολισμό αλλά και τη καταστολή του οπερονίου *bio* μέσω του οποίου αυτή παράγεται. Όταν τα επίπεδα βιοτίνης είναι υψηλά η πρωτεΐνη *birA* συνδέεται στο DNA και αποτρέπει τη μεταγραφή του οπερονίου (Beckett, 2009).

Η πρωτεΐνη της οικογένειας ROK που αναγνωρίστηκε αποτελεί ένα από τα 5 παράλογα που εντοπίζονται στο στέλεχος αναφοράς και αποτελεί πρωτεΐνη αίσθησης της συγκέντρωσης υδατανθράκων στο φύλο *Thermotogae* (Kazanov et al., 2013). Υπάρχουν ομόλογα ROK στον *B. subtilis* τα οποία εμπλέκονται στη μεταγραφική ρύθμιση πολλαπλών γονιδίων. Μεταξύ άλλων εμπλέκονται στη διαδικασία της σποριογένεσης καθώς μεταλλάγματα εμφάνιζαν αδυναμία δημιουργίας σπορίων (Albano et al., 2005).

Συνολικά οι πρωτεΐνες αποτυπώματα, που έχουν χαρακτηρισμό, εμπλέκονται σε πολύ γενικευμένες βιολογικές διεργασίες και κατά βάση αποτελούν παράγοντες ρύθμισης της γονιδιακής έκφρασης. Καθώς δεν αναγνωρίστηκαν γονίδια που να σχετίζονται άμεσα με τους ιδιαίτερους φαινοτύπους της ομάδας, μάλλον ίσως τα ιδιαίτερα χαρακτηριστικά αυτής της ομάδας να είναι αποτέλεσμα γονιδιακής ρύθμισης και όχι ύπαρξης εξειδικευμένων γονιδίων. Η αναγνώριση ρυθμιστικών παραγόντων ως γονίδια αποτυπώματα ευνοεί την υπόθεση αυτή. Αξίζει να σημειωθεί ότι η πιο πρόσφατη συγκριτική ανάλυση

που έγινε στα στελέχη *B. pumilus* και *B. safensis* που απομονώθηκαν από τους χώρους της NASA (Tirumalai et al., 2018) απέτυχε να αναγνωρίσει γονίδια που προσδίδουν άμεσα την ανθεκτικότητα των στελεχών αυτών.

4. ΣΥΜΠΕΡΑΣΜΑΤΑ

Συνολικά, το γένος *Bacillus* αποτελεί μια πολύ ετερογενή ταξινομική ομάδα και τα είδη που την απαρτίζουν τίθενται ακόμα υπό συζήτηση με εξαίρεση δύο κύριους κλάδους. Η ομάδα *B. altitudinis* – *B. pumilus* – *B. safensis* αποτελεί μέρος ενός εξ αυτών. Η φυλογενωμική και συγκριτική ανάλυση μας σε υψηλής ποιότητας γονιδιώματα αναγνώρισε αρκετές υποθετικές και λίγες ρυθμιστικές πρωτεΐνες που χαρακτηρίζουν το συγκεκριμένο κλάδο. Υπάρχει πιθανότητα οι ρυθμιστικές αυτές πρωτεΐνες να σχετίζονται έμμεσα με την ικανότητα επιβίωσης σε ακραία περιβάλλοντα που εμφανίζουν τα περισσότερα από τα στελέχη της συγκεκριμένης ομάδας. Μια εμπειριστατωμένη ανάλυση σαν τη συγκεκριμένη είναι απαραίτητη για την εξακρίβωση του αριθμού των διαφόρων υποομάδων, των εξελικτικών τους σχέσεων και των γενετικών στοιχείων που τις κάνουν να διαφέρουν από άλλους συγγενικούς οργανισμούς.

Αυτή η μεταπτυχιακή εργασία βασίστηκε στην έρευνα που ξεκίνησα (Σεπτέμβριος 2019) και δημοσίευσα (Ιούλιος 2020) ως μεταπτυχιακός φοιτητής στο γονιδίωμα πυρήνα του γένους *Pseudomonas*. Ωστόσο, επειδή η δημοσίευση πραγματοποιήθηκε πριν την επίσημη έναρξη της μεταπτυχιακής μου εργασίας, προχώρησα στην επιπλέον ανάλυση των συγκεκριμένων βακίλλων.

The screenshot shows the MDPI journal website interface. At the top, there is a navigation bar with links for Journals, Information, Author Services, Initiatives, and About, along with Sign In / Sign Up and Submit buttons. Below this is a search bar with fields for Title / Keyword, Author / Affiliation, Diversity, and All Article Types, and a Search button. The main content area displays the article title, authors (Marios Nikolaidis, Dimitris Mossialos, Stephen G. Oliver, and Grigorios D. Amoutzias), and the journal information (Diversity 2020, 12(8), 289). The article is marked as Open Access, Feature Paper, and Article. The page also includes an Article Menu on the left with options like Abstract, Supplementary Material, and Open Access and Permissions. On the right, there are social media sharing icons.

5. ΒΙΒΛΙΟΓΡΑΦΙΑ

- Albano, M., Smits, W.K., Ho, L.T.Y., Kraigher, B., Mandic-Mulec, I., Kuipers, O.P., Dubnau, D., 2005. The Rok Protein of *Bacillus subtilis* Represses Genes for Cell Surface and Extracellular Functions. *JB* 187, 2010–2019. <https://doi.org/10.1128/JB.187.6.2010-2019.2005>
- Barbe, V., Cruveiller, S., Kunst, F., Lenoble, P., Meurice, G., Sekowska, A., Vallenet, D., Wang, T., Moszer, I., Médigue, C., Danchin, A., 2009. From a consortium sequence to a unified sequence: the *Bacillus subtilis* 168 reference genome a decade later. *Microbiology (Reading, Engl.)* 155, 1758–1775. <https://doi.org/10.1099/mic.0.027839-0>
- Bazinet, A.L., 2017. Pan-genome and phylogeny of *Bacillus cereus sensu lato*. *BMC Evol Biol* 17, 176. <https://doi.org/10.1186/s12862-017-1020-1>
- Beckett, D., 2009. Biotin Sensing at the Molecular Level. *The Journal of Nutrition* 139, 167–170. <https://doi.org/10.3945/jn.108.095760>
- Bosi, E., Fani, R., Fondi, M., 2015. Defining Orthologs and Pangenome Size Metrics, in: Mengoni, A., Galardini, M., Fondi, M. (Eds.), *Bacterial Pangenomics, Methods in Molecular Biology*. Springer New York, New York, NY, pp. 191–202. https://doi.org/10.1007/978-1-4939-1720-4_13
- Branquinho, R., Meirinhos-Soares, L., Carriço, J.A., Pintado, M., Peixe, L.V., 2014. Phylogenetic and clonality analysis of *Bacillus pumilus* isolates uncovered a highly heterogeneous population of different closely related species and clones. *FEMS Microbiol Ecol* 90, 689–698. <https://doi.org/10.1111/1574-6941.12426>
- Castresana, J., 2000. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Molecular Biology and Evolution* 17, 540–552. <https://doi.org/10.1093/oxfordjournals.molbev.a026334>
- Chevenet, F., Brun, C., Bañuls, A.-L., Jacq, B., Christen, R., 2006. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* 7, 439. <https://doi.org/10.1186/1471-2105-7-439>
- Du, Y., Ma, J., Yin, Z., Liu, K., Yao, G., Xu, W., Fan, L., Du, B., Ding, Y., Wang, C., 2019. Comparative genomic analysis of *Bacillus paralicheniformis* MDJK30 with its closely related species reveals an evolutionary relationship between *B. paralicheniformis* and *B. licheniformis*. *BMC Genomics* 20, 283. <https://doi.org/10.1186/s12864-019-5646-9>
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32, 1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Ehling-Schulz, M., Lereclus, D., Koehler, T.M., 2019. The *Bacillus cereus* Group: *Bacillus* Species with Pathogenic Potential. *Microbiology Spectrum* 7. <https://doi.org/10.1128/microbiolspec.GPP3-0032-2018>
- Elshaghabee, F.M.F., Rokana, N., Gulhane, R.D., Sharma, C., Panwar, H., 2017. *Bacillus* As Potential Probiotics: Status, Concerns, and Future Perspectives. *Front. Microbiol.* 8, 1490. <https://doi.org/10.3389/fmicb.2017.01490>
- Emms, D.M., Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 20, 238. <https://doi.org/10.1186/s13059-019-1832-y>
- Enterprise Open Source and Linux [WWW Document], n.d. . Ubuntu. URL <https://ubuntu.com/> (accessed 2.8.21).
- Fan, B., Blom, J., Klenk, H.-P., Borriss, R., 2017. *Bacillus amyloliquefaciens*, *Bacillus velezensis*, and *Bacillus siamensis* Form an “Operational Group *B. amyloliquefaciens*” within the *B. subtilis* Species Complex. *Front. Microbiol.* 8. <https://doi.org/10.3389/fmicb.2017.00022>
- Fritze, D., 2004. Taxonomy of the Genus *Bacillus* and Related Genera: The Aerobic Endospore-Forming Bacteria. *Phytopathology* 94, 1245–1248. <https://doi.org/10.1094/PHYTO.2004.94.11.1245>
- From, C., Hormazabal, V., Granum, P.E., 2007. Food poisoning associated with pumilacidin-producing *Bacillus pumilus* in rice. *International Journal of Food Microbiology* 115, 319–324. <https://doi.org/10.1016/j.ijfoodmicro.2006.11.005>
- Gogarten, J.P., Townsend, J.P., 2005. Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.* 3, 679–687. <https://doi.org/10.1038/nrmicro1204>
- Goswami, M., Deka, S., 2020. Isolation of a novel rhizobacteria having multiple plant growth promoting traits and antifungal activity against certain phytopathogens. *Microbiological Research* 240, 126516. <https://doi.org/10.1016/j.micres.2020.126516>

- Gouy, M., Guindon, S., Gascuel, O., 2010. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Molecular Biology and Evolution* 27, 221–224. <https://doi.org/10.1093/molbev/msp259>
- Harwood, C.R., 1989. Introduction to the Biotechnology of *Bacillus*, in: Harwood, C.R. (Ed.), *Bacillus*. Springer US, Boston, MA, pp. 1–4. https://doi.org/10.1007/978-1-4899-3502-1_1
- Hernández-González, I.L., Moreno-Hagelsieb, G., Olmedo-Álvarez, G., 2018. Environmentally-driven gene content convergence and the *Bacillus* phylogeny. *BMC Evol Biol* 18, 148. <https://doi.org/10.1186/s12862-018-1261-7>
- Hernández-Salmerón, J.E., Moreno-Hagelsieb, G., 2020. Progress in quickly finding orthologs as reciprocal best hits: comparing blast, last, diamond and MMseqs2. *BMC Genomics* 21, 741. <https://doi.org/10.1186/s12864-020-07132-6>
- Hong, H.A., Duc, L.H., Cutting, S.M., 2005. The use of bacterial spore formers as probiotics. *FEMS Microbiol Rev* 29, 813–835. <https://doi.org/10.1016/j.femsre.2004.12.001>
- Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., Rattei, T., Mende, D.R., Sunagawa, S., Kuhn, M., Jensen, L.J., von Mering, C., Bork, P., 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* 44, D286-293. <https://doi.org/10.1093/nar/gkv1248>
- Jolley, K.A., Bliss, C.M., Bennett, J.S., Bratcher, H.B., Brehony, C., Colles, F.M., Wimalarathna, H., Harrison, O.B., Sheppard, S.K., Cody, A.J., Maiden, M.C.J., 2012. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology (Reading, Engl.)* 158, 1005–1015. <https://doi.org/10.1099/mic.0.055459-0>
- Jolley, K.A., Bray, J.E., Maiden, M.C.J., 2017. A RESTful application programming interface for the PubMLST molecular typing and genome databases. *Database* 2017. <https://doi.org/10.1093/database/bax060>
- Kaur, R., Goyal, D., 2020. Biodegradation of Butachlor by *Bacillus altitudinis* and Identification of Metabolites. *Curr Microbiol* 77, 2602–2612. <https://doi.org/10.1007/s00284-020-02031-1>
- Kavamura, V.N., 2013. Screening of Brazilian cacti rhizobacteria for plant growth promotion under drought. *Microbiological Research* 9.
- Kazanov, M.D., Li, X., Gelfand, M.S., Osterman, A.L., Rodionov, D.A., 2013. Functional diversification of ROK-family transcriptional regulators of sugar catabolism in the Thermotogae phylum. *Nucleic Acids Research* 41, 790–803. <https://doi.org/10.1093/nar/gks1184>
- Kour, D., Rana, K.L., Yadav, A.N., Yadav, N., Kumar, M., Kumar, V., Vyas, P., Dhaliwal, H.S., Saxena, A.K., 2020. Microbial biofertilizers: Bioresources and eco-friendly technologies for agricultural and environmental sustainability. *Biocatalysis and Agricultural Biotechnology* 23, 101487. <https://doi.org/10.1016/j.bcab.2019.101487>
- Kumaravel, S., Thankappan, S., Raghupathi, S., Uthandi, S., 2018. Draft Genome Sequence of Plant Growth-Promoting and Drought-Tolerant *Bacillus altitudinis* FD48, Isolated from Rice Phylloplane. *Genome Announc* 6, e00019-18, /ga/6/9/e00019-18.atom. <https://doi.org/10.1128/genomeA.00019-18>
- Kyrpides, N.C., Hugenholtz, P., Eisen, J.A., Woyke, T., Göker, M., Parker, C.T., Amann, R., Beck, B.J., Chain, P.S.G., Chun, J., Colwell, R.R., Danchin, A., Dawyndt, P., Dedeurwaerdere, T., DeLong, E.F., Detter, J.C., De Vos, P., Donohue, T.J., Dong, X.-Z., Ehrlich, D.S., Fraser, C., Gibbs, R., Gilbert, J., Gilna, P., Glöckner, F.O., Jansson, J.K., Keasling, J.D., Knight, R., Labeda, D., Lapidus, A., Lee, J.-S., Li, W.-J., Ma, J., Markowitz, V., Moore, E.R.B., Morrison, M., Meyer, F., Nelson, K.E., Ohkuma, M., Ouzounis, C.A., Pace, N., Parkhill, J., Qin, N., Rossello-Mora, R., Sikorski, J., Smith, D., Sogin, M., Stevens, R., Stingl, U., Suzuki, K.-I., Taylor, D., Tiedje, J.M., Tindall, B., Wagner, M., Weinstock, G., Weissenbach, J., White, O., Wang, J., Zhang, L., Zhou, Y.-G., Field, D., Whitman, W.B., Garrity, G.M., Klenk, H.-P., 2014. Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains. *PLoS Biol.* 12, e1001920. <https://doi.org/10.1371/journal.pbio.1001920>
- Lateef, A., Adelere, I.A., Gueguim-Kana, E.B., 2015. The biology and potential biotechnological applications of *Bacillus safensis*. *Biologia* 70. <https://doi.org/10.1515/biolog-2015-0062>
- Liu, Y., Lai, Q., Dong, C., Sun, F., Wang, L., Li, G., Shao, Z., 2013. Phylogenetic Diversity of the *Bacillus pumilus* Group and the Marine Ecotype Revealed by Multilocus Sequence Analysis. *PLOS ONE* 8, 11.
- Logan, N.A., Berge, O., Bishop, A.H., Busse, H.-J., De Vos, P., Fritze, D., Heyndrickx, M., Kämpfer, P., Rabinovitch, L., Salkinoja-Salonen, M.S., Seldin, L., Ventosa, A., 2009. Proposed minimal stand-

- ards for describing new taxa of aerobic, endospore-forming bacteria. *Int. J. Syst. Evol. Microbiol.* 59, 2114–2121. <https://doi.org/10.1099/ijs.0.013649-0>
- Maiden, M.C.J., van Rensburg, M.J.J., Bray, J.E., Earle, S.G., Ford, S.A., Jolley, K.A., McCarthy, N.D., 2013. MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat Rev Microbiol* 11, 728–736. <https://doi.org/10.1038/nrmicro3093>
- Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.-P., Göker, M., 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14, 60. <https://doi.org/10.1186/1471-2105-14-60>
- Nikolaidis, M., Mossialos, D., Oliver, S.G., Amoutzias, G.D., 2020. Comparative Analysis of the Core Proteomes among the *Pseudomonas* Major Evolutionary Groups Reveals Species-Specific Adaptations for *Pseudomonas aeruginosa* and *Pseudomonas chlororaphis*. *Diversity* 12, 289. <https://doi.org/10.3390/d12080289>
- Patel, S., Gupta, R.S., 2020. A phylogenomic and comparative genomic framework for resolving the polyphyly of the genus *Bacillus*: Proposal for six new genera of *Bacillus* species, *Peribacillus* gen. nov., *Cytobacillus* gen. nov., *Mesobacillus* gen. nov., *Neobacillus* gen. nov., *Metabacillus* gen. nov. and *Alkalihalobacillus* gen. nov. *International Journal of Systematic and Evolutionary Microbiology* 70, 406–438. <https://doi.org/10.1099/ijsem.0.003775>
- Pérez-García, A., Romero, D., de Vicente, A., 2011. Plant protection and growth stimulation by microorganisms: biotechnological applications of Bacilli in agriculture. *Current Opinion in Biotechnology* 22, 187–193. <https://doi.org/10.1016/j.copbio.2010.12.003>
- Pritchard, L., Glover, R.H., Humphris, S., Elphinstone, J.G., Toth, I.K., 2016. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal. Methods* 8, 12–24. <https://doi.org/10.1039/C5AY02550H>
- Radhakrishnan, R., Hashem, A., Abd Allah, E.F., 2017. *Bacillus*: A Biological Tool for Crop Improvement through Bio-Molecular Changes in Adverse Environments. *Front Physiol* 8, 667. <https://doi.org/10.3389/fphys.2017.00667>
- Richter, M., Rosselló-Móra, R., 2009. Shifting the genomic gold standard for the prokaryotic species definition. *PNAS* 106, 19126–19131. <https://doi.org/10.1073/pnas.0906412106>
- Satomi, M., La Duc, M.T., Venkateswaran, K., 2006. *Bacillus safensis* sp. nov., isolated from spacecraft and assembly-facility surfaces. *International Journal of Systematic and Evolutionary Microbiology* 56, 1735–1740. <https://doi.org/10.1099/ijs.0.64189-0>
- Shivaji, S., Chaturvedi, P., Suresh, K., Reddy, G.S.N., Dutt, C.B.S., Wainwright, M., Narlikar, J.V., Bhargava, P.M., 2006. *Bacillus aerius* sp. nov., *Bacillus aerophilus* sp. nov., *Bacillus stratosphericus* sp. nov. and *Bacillus altitudinis* sp. nov., isolated from cryogenic tubes used for collecting air samples from high altitudes. *International Journal of Systematic and Evolutionary Microbiology* 56, 1465–1473. <https://doi.org/10.1099/ijs.0.64029-0>
- Sumi, C.D., Yang, B.W., Yeo, I.-C., Hahm, Y.T., 2015. Antimicrobial peptides of the genus *Bacillus*: a new era for antibiotics. *Can. J. Microbiol.* 61, 93–103. <https://doi.org/10.1139/cjm-2014-0613>
- Tindall, B.J., Rosselló-Móra, R., Busse, H.-J., Ludwig, W., Kämpfer, P., 2010. Notes on the characterization of prokaryote strains for taxonomic purposes. *Int. J. Syst. Evol. Microbiol.* 60, 249–266. <https://doi.org/10.1099/ijs.0.016949-0>
- Tirumalai, M.R., Rastogi, R., Zamani, N., O'Bryant Williams, E., Allen, S., Diouf, F., Kwende, S., Weinstock, G.M., Venkateswaran, K.J., Fox, G.E., 2013. Candidate Genes That May Be Responsible for the Unusual Resistances Exhibited by *Bacillus pumilus* SAFR-032 Spores. *PLoS ONE* 8, e66012. <https://doi.org/10.1371/journal.pone.0066012>
- Tirumalai, M.R., Stepanov, V.G., Wünsche, A., Montazari, S., Gonzalez, R.O., Venkateswaran, K., Fox, G.E., 2018. *Bacillus safensis* FO-36b and *Bacillus pumilus* SAFR-032: a whole genome comparison of two spacecraft assembly facility isolates. *BMC Microbiol* 18, 57. <https://doi.org/10.1186/s12866-018-1191-y>
- Vernikos, G., Medini, D., Riley, D.R., Tettelin, H., 2015. Ten years of pan-genome analyses. *Curr. Opin. Microbiol.* 23, 148–154. <https://doi.org/10.1016/j.mib.2014.11.016>
- Welcome to Python.org [WWW Document], n.d. . Python.org. URL <https://www.python.org/> (accessed 2.9.21).
- Woese, C.R., Kandler, O., Wheelis, M.L., 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences* 87, 4576–4579. <https://doi.org/10.1073/pnas.87.12.4576>

Xu, D., Côté, J.-C., 2003. Phylogenetic relationships between *Bacillus* species and related genera inferred from comparison of 3' end 16S rDNA and 5' end 16S-23S ITS nucleotide sequences. *Int. J. Syst. Evol. Microbiol.* 53, 695–704. <https://doi.org/10.1099/ijs.0.02346-0>