



ΤΜΗΜΑ
**Βιοχημείας &
Βιοτεχνολογίας**
ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

Τμήμα Βιοχημείας και Βιοτεχνολογίας
ΤΟΥ
Πανεπιστημίου Θεσσαλίας

“Βιοπληροφορική και φυλογενωμική ανάλυση του κοινού γονιδιώματος των μελών του γένους *Streptomyces*”

“Bionformatic and phylogenomic analysis of the core genome amongst members of the genus *Streptomyces*”

Νικολέττα Φράγγου
ΤΟΥ
Γεωργίου

2022

Υπεύθυνος καθηγητής:

Αμούτζιας Γρηγόριος, Αναπληρωτής Καθηγητής Βιοπληροφορικής με έμφαση στην Μικροβιολογία, Τμήμα Βιοχημείας και Βιοτεχνολογίας, Π.Θ.

Τριμελής επιτροπή:

1. Αμούτζιας Γρηγόριος, Αναπληρωτής Καθηγητής Βιοπληροφορικής με έμφαση στην Μικροβιολογία, Τμήμα Βιοχημείας και Βιοτεχνολογίας, Π.Θ.
2. Μόσιαλος Δημήτριος, Αναπληρωτής Καθηγητής Βιοτεχνολογίας Μικροβίων, Τμήμα Βιοχημείας και Βιοτεχνολογίας, Π.Θ.
3. Κοτταρίδη Χριστίνα, Επίκουρη Καθηγήτρια Μικροβιολογίας στο Τμήμα Βιολογίας, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

Ευχαριστίες:

Θα ήθελα να ευχαριστήσω θερμά κάποια άτομα, χωρίς την βοήθεια και αρωγή των οποίων το έργο αυτό δε θα ήταν δυνατό.

Πρώτον απ' όλους θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Αμούτζια Γρηγόριο για την σταθερή και αποτελεσματική καθοδήγησή του, τη συνέπεια, την προθυμία του να μου προσφέρει βοήθεια και πάνω απ' όλα την υπομονή του. Επίσης, είμαι ευγνώμων για το Μάριο Νικολαΐδη που μου προσέφερε συνεχή στήριξη, ως συνάδελφος και φίλος. Θα ήθελα να ευχαριστήσω, φυσικά, και τα υπόλοιπα μέλη της Τριμελούς Συμβουλευτικής Επιτροπής μου, κ. Μόσιαλο Δημήτριο και κα. Κοτταρίδη Χριστίνα.

Έπειτα, ευχαριστώ την οικογένειά μου, που ήταν δίπλα μου σε κάθε ακαδημαϊκό και μη βήμα και με βοήθησαν, ο καθένας με τη δυνατότητα και τον τρόπο του, να πλοηγηθώ στον κόσμο του Πανεπιστημίου.

Τέλος, δε θα μπορούσαν να λείπουν από αυτή τη λίστα οι φίλοι μου, που μου προσέφεραν αμέτρητες στιγμές χαράς αλλά και δύναμης σε δύσκολες στιγμές. Οι στιγμές που μου χάρισαν αποτέλεσαν εφιαλτήριο για να συνεχίσω να επιτελώ το ακαδημαϊκό μου έργο.

Περιεχόμενα

Περίληψη.....	3
---------------	---

Abstract	5
1. Εισαγωγή	6
1.1 Η τάξη <i>Streptomyces</i> και η ταξινόμησή της.....	6
1.2 Γονιδίωμα.....	10
1.3 Εφαρμογές στη Βιοτεχνολογία	12
1.4 Παλαιότερες μεθοδολογίες ταξινόμησης.....	12
1.5 Γονιδιωματικές αναλύσεις.....	13
2. Υλικά και μέθοδοι.....	17
2.1 Ubuntu	17
2.2 Windows Subsystem for Linux	17
2.3 Perl	17
2.4 Python.....	17
2.5 IQ-Tree2	18
2.6 fastANI	18
2.7 Interactive Tree of Life (iTOL)	19
2.8 NCBI Assembly.....	19
2.9 Markov Cluster Algorithm (MCL)	19
2.10 eggNOG-mapper.....	19
3. Αποτελέσματα - Συζήτηση	20
4. Βιβλιογραφία	29

Περίληψη

Η οικογένεια *Streptomycetaceae*, αποτελούμενη από οχτώ γένη, περιλαμβάνει θετικά κατά Gram βακτήρια με χαρακτηριστικά υψηλό ποσοστό GC. Αναφέρεται συχνά ότι συλλογικά, είναι ικανή να παράξει πάνω από τα δύο τρίτα ($\frac{2}{3}$) των ευρέως χρησιμοποιούμενων βιοενεργών ουσιών, με πιο γνωστή τη στρεπτομυκίνη. Χαρακτηρίζεται από ευμεγέθη γονιδιώματα, με πλασμίδια διαφόρων μεγεθών και μορφολογιών. Ο κύκλος ζωής των βακτηρίων αυτών αποτελείται από το στάδιο των σπορίων, που δίνει τη θέση του στο μυκήλιο υποστρώματος και μετά ακολουθεί το εναέριο μυκήλιο που σταδιακά διαφοροποιείται ξανά σε σπόρια. Βρίσκουν εφαρμογές στη Βιοτεχνολογία, ως άριστοι δέκτες ετερόλογης έκφρασης πρωτεϊνών, και ως εκ τούτου έχουν γίνει πολλές απόπειρες για τη μελέτη τους. Παλαιότερες μεθοδολογίες για τη διερεύνηση της ταξινόμησής τους αποτελούσαν η 16S rRNA και η Multi-locus Sequence Analysis, ενώ σήμερα οι φυλογενωμικές προσεγγίσεις φαίνονται πιο αποτελεσματικές. Μια τέτοια προσέγγιση αποτελεί και η παρούσα διπλωματική εργασία, που παράλληλα με άλλες διεργασίες έχει ως στόχο τη διαλεύκανση των εξελικτικών σχέσεων των μελών του γένους *Streptomyces* και σύγκριση με την ήδη υπάρχουσα βιβλιογραφία.

Abstract

The *Streptomycetaceae* family, consisting of eight (8) genera, includes Gram positive bacteria with high GC% content. It has the potential to produce two thirds ($\frac{2}{3}$) of all available bioactive compounds in use today. It is characterized by sizable genomes and plasmids of various sizes. Their life cycle consists of the spore phase, developing into the substrate mycelium (i.e., its vegetative state), then into the aerial mycelium which gradually differentiates into spores once again. Its members are commonly used as hosts for heterologous protein expression. Previous taxonomic analyses utilized 16S rRNA and Multi-locus Sequence Analysis (MLSA) methodologies to study them, while newly developed genomic approaches, promise to shed more light into their taxonomy. In this project, we use the phylogenomics approach, along with other methods, to better understand the evolutionary relationships in the *Streptomyces* genus.

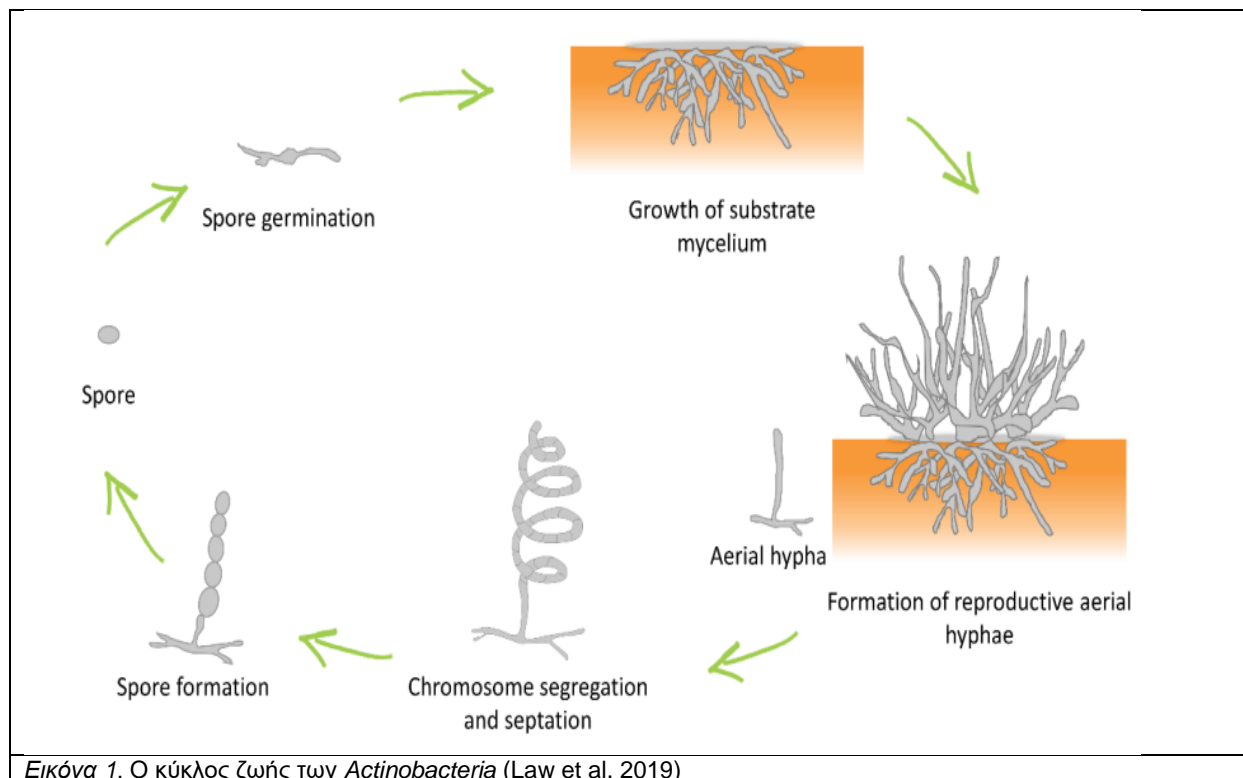
1. Εισαγωγή

1.1 Η τάξη *Streptomycetales* και η ταξινόμησή της

Η τάξη *Streptomycetales* ανήκει στο φύλο των Ακτινοβακτηρίων (*Actinobacteria*) και περιέχει την οικογένεια *Streptomycetaceae* (Schoch et al., 2020). Τα μέλη της οικογένειας αυτής, η οποία αναφέρθηκε για πρώτη φορά το 1943 από τους (Waksman & Henrici) είναι θετικά κατά Gram, αερόβια βακτήρια με υψηλό ποσοστό γουανινών και κυτοσινών στο γονιδίωμά τους (Paradkar et al. 2003).

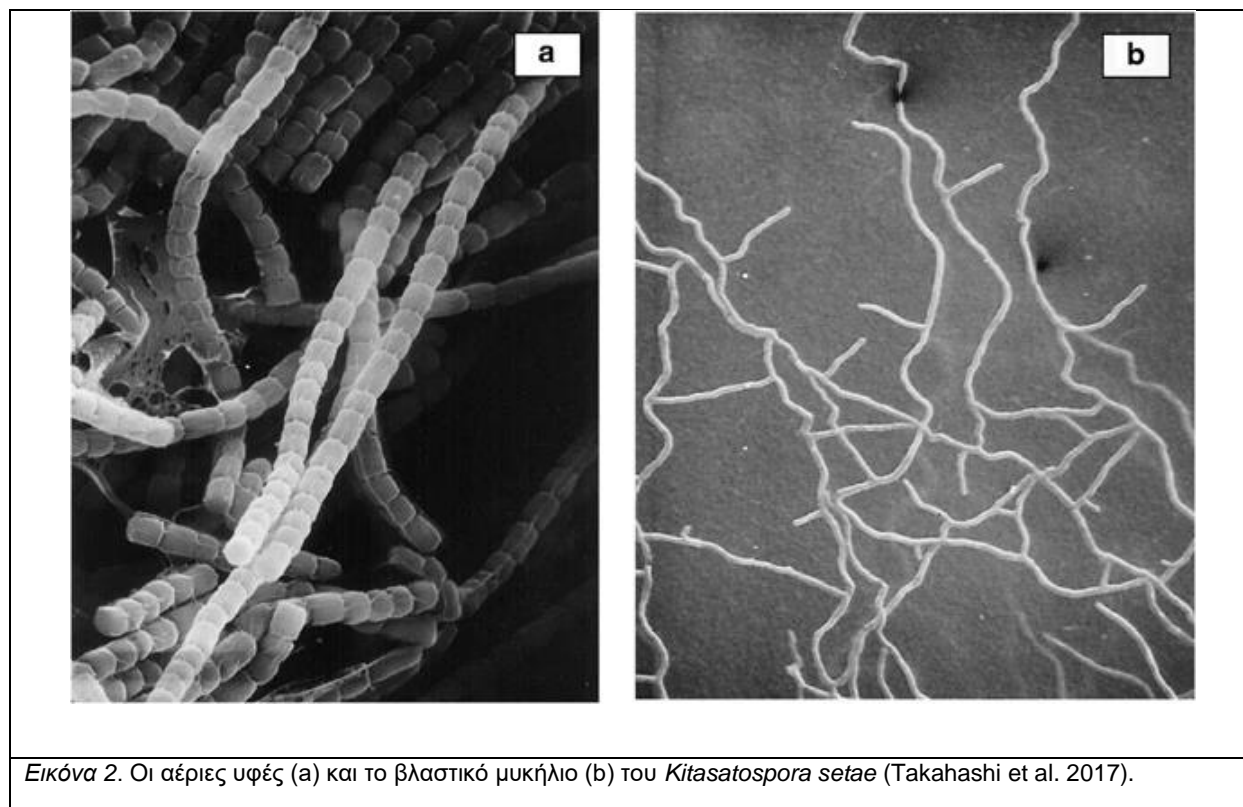
Πολλές ανακατατάξεις έχουν υπάρξει μέσα στην οικογένεια *Streptomycetaceae*. Αρχικά, σύμφωνα με ταξινόμηση βασιζόμενη σε μορφολογικά χαρακτηριστικά, σε αυτή την οικογένεια κατατάσσονταν τα τέσσερα γένη: *Microellobospora*, *Sporichthya*, *Streptomyces*, και *Streptoverticillium* (Bergey et al., 1984). Έπειτα προστέθηκε το γένος *Kitasatospora*, με το *K. setae* να είναι το χαρακτηριστικό είδος, που όμως αργότερα συμπίχθηκε με το γένος *Streptomyces*. Μια ακόμα ανατροπή στα ταξινομικά δεδομένα διαχώρισε εκ νέου τους *Kitasatospora*, λόγω της ύπαρξης γαλακτόζης στα κυτταρικά τους τοιχώματα, ένα χαρακτηριστικό των *Kitasatospora* που δε μοιράζονται με τους *Streptomyces*, καθώς κι επίσης στις μεγάλες ποσότητες μεσο-διαμινοπιμελικού οξέος (Zhang et al., 1997).

Χαρακτηριστικός είναι ο κύκλος ζωής τους (Εικόνα 1), που περιλαμβάνει πολλές μορφολογικές αλλαγές. Μετά το στάδιο των σπορίων, σχηματίζονται διακλαδισμένα νημάτια, το ονομαζόμενο βασικό μυκήλιο υπόστρωμα (substrate mycelium) που αποτελεί την βλαστική φάση. Με την πάροδο του χρόνου, οι μυκηλιακές υφές αναπτύσσονται κάθετα και εκτείνονται προς τα πάνω, δημιουργώντας τις επανομαζόμενες εναέριες υφές. Οι συγκεκριμένες δομές είναι πολυπύρηνες και υπόκεινται σε κυτταρικές διαιρέσεις ώστε να σχηματιστούν απλοειδή τμήματα που τελικά διαφοροποιούνται σε άλλα σπόρια (Law et al. 2019).



1.1.1 Γένος *Kitasatospora*

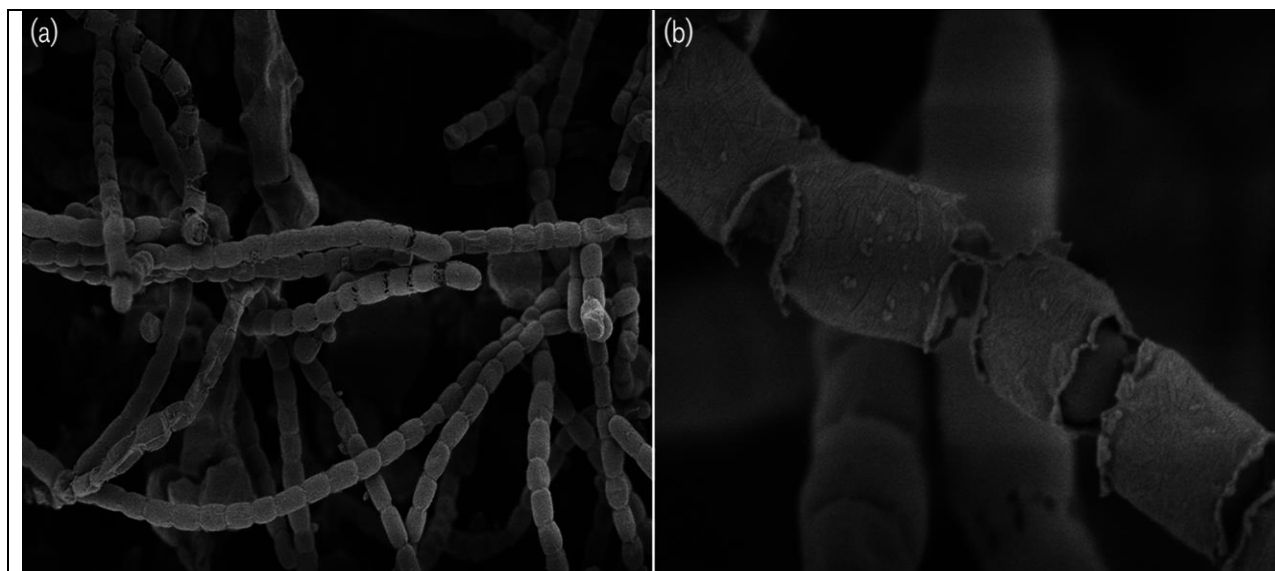
Σήμερα, η οικογένεια *Streptomycetaceae* εμπεριέχει τα γένη *Allostreptomyces*, *Embleya*, *Kitasatospora*, *Streptacidiphilus*, *Yinghuangia* και τους πλέον γνωστούς *Streptomyces* (Schoch et al., 2020). Το γένος *Kitasatospora* διαθέτει στο κυτταρικό του τοίχωμα LL-διαμινοπιμελικό οξύ όταν βρίσκεται υπό τη μορφή σπορίων, ενώ όταν είναι σε μυκηλιακή μορφή (Εικόνα 2b) διαθέτει κυρίως μεσο-διαμινοπιμελικό οξύ, τα οποία είναι σημαντικά εργαλεία για την χημειοταξινόμηση. Μέχρι τώρα 34 είδη έχουν ταυτοποιηθεί (Parte et al., 2020), με το χαρακτηριστικό είδος μεταξύ αυτών να είναι το *Kitasatospora setae*, στέλεχος KM-6054T. Από τους πιο γνωστούς δευτερογενείς μεταβολίτες που παράγει το συγκεκριμένο γένος, είναι η σεταμυκίνη, η οποία δρα κατά τριχομονάδων και μυκήτων, η φωσαλακίνη, ένα ζιζανιοκτόνο, οι propiochaxatins, οι οποίες έχει δειχθεί ότι δρουν ως αναστολείς εγκεφαλινασών. (Takahashi et al., 2017) Επίσης, έχει δειχθεί ότι ένα είδος του γένους *Kitasatospora* που απομονώθηκε από ίταμο, είναι ικανό να παράξει πακλιταξέλη, έναν αντικαρκινικό παράγοντα (Caruso et al., 2000).



1.1.2 Γένος *Streptacidiphilus*

Το γένος *Streptacidiphilus* αποτελεί ένα οξεόφιλο γένος που συνδέεται πιο στενά με το γένος *Kitasatospora* απ' ό τι με τα υπόλοιπα μέλη της οικογένειας (Li et al., 2021). Μέχρι σήμερα, έχουν ταυτοποιηθεί 16 μέλη του γένους *Streptacidiphilus* (Parte et al., 2020) και αναπτύσσονται σε ένα εύρος pH 3.5-6, με ιδανική τιμή το 5 (Kim et al., 2003). Το χαρακτηριστικό είδος του γένους αυτού είναι ο *Streptacidiphilus albus* (Parte et al., 2020).

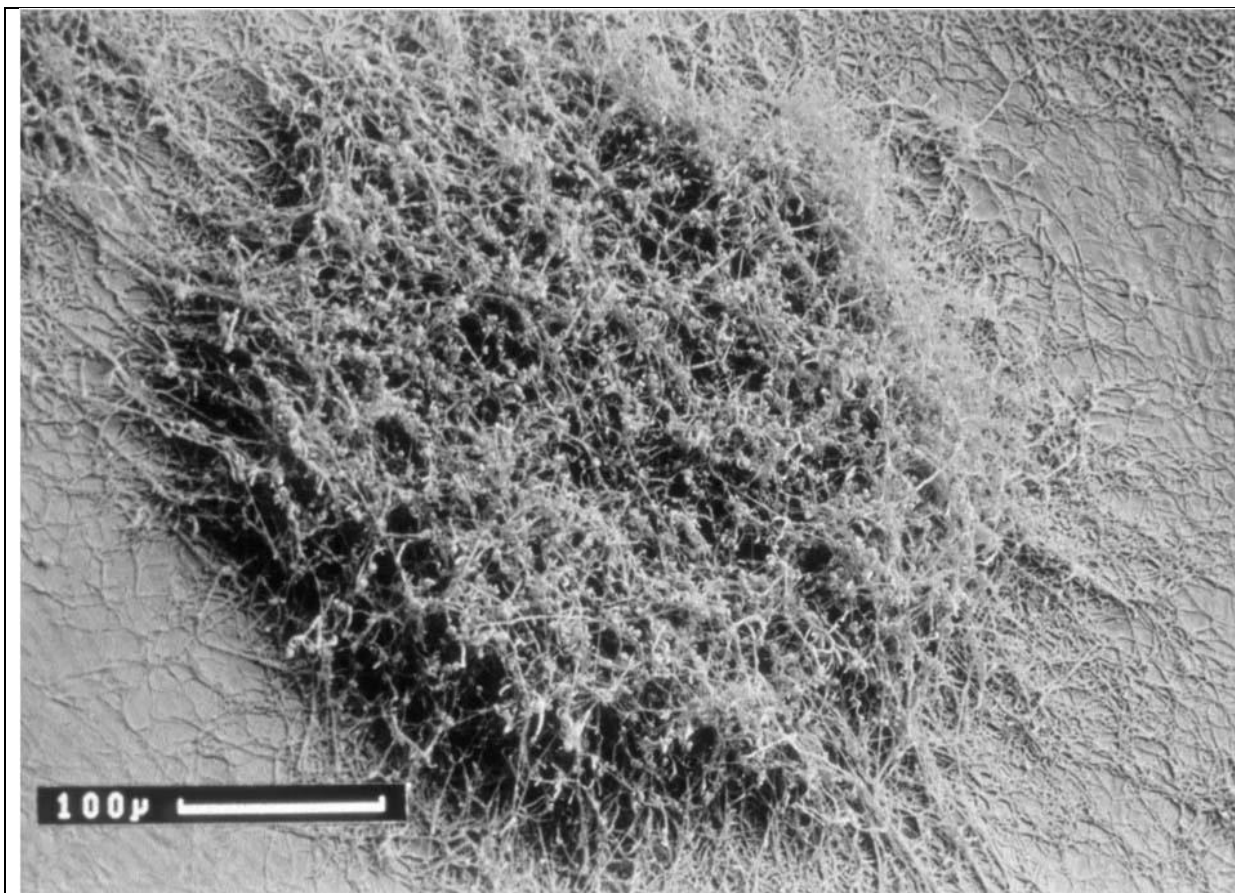
Μια λεπτομερή εικόνα της δημιουργίας των σπορίων, μέσω ηλεκτρονικού μικροσκοπίου φαίνεται στην Εικόνα 3.



Εικόνα 3. Α) Σπόρια στις αέριες υφές του *Streptacidiphilus monticola*. Β) Η επιφάνεια των σπορίων (Song et al. 2018)

1.1.3 Γένος *Streptomyces*

Τέλος, ένα ακόμη γένος μεγάλου ενδιαφέροντος είναι οι *Streptomyces*, των οποίων ο δευτερογενής μεταβολισμός είναι υπεύθυνος για την παραγωγή πάνω από των μισών βιοενεργών ουσιών που χρησιμοποιούνται ευρέως. Είναι επίσης ικανοί να αποδομήσουν αδιάλυτη νεκρή ύλη που περιέχει λιγνο-κυτταρίνη ή χιτίνη. Σημαντικά παραδείγματα ουσιών που παράγουν αποτελούν η αβερμεκτίνη, ένα παρασιτοκτόνο που παράγεται από τον *S. avermitilis*, η στρεπτομυκίνη του *S. griseus*, ένα αντιβιοτικό ευρέως φάσματος, αποτελεσματικό για την αντιμετώπιση της φυματίωσης και της πανώλης (Zhou et al. 2011) και η χλωραμφενικόλη που βρέθηκε ότι παράγεται από τον *S. venezuelae* (Kampfer et al 2014). Μερικοί από τους *Streptomyces* είναι αντιθέτως παθογόνοι, όπως ο *S. scabiei* που προσβάλλει πατάτες (Bignell et al. 2014), ενώ υπάρχουν κι άλλοι που προκαλούν λιγότερο οικονομικά επιζήμιες βλάβες σε φυτικά συστήματα. (Zhang et al. 2016). Μερικοί, όπως ο *S. somaliensis* είναι ικανοί να προκαλέσουν μυκήτωμα, μια τροπική ασθένεια, που αν αφαιρεθεί χωρίς θεραπεία μπορεί να προκαλέσει μόνιμη βλάβη σε ιστούς που έχει προσβάλλει, ακόμα και τα περιβάλλοντα οστά (Verma et al. 2019).



Εικόνα 4. Ο *Streptomyces lividans* σε ηλεκτρονικό μικροσκόπιο σάρωσης. Στα άκρα της εικόνας οι υφές του μυκηλίου υποστρώματος αναζητούν θρεπτικά συστατικά, ενώ στο κέντρο οι αέριες υφές αρχίζουν την παραγωγή σπορίων (Horwood et al., 2019).

1.2 Γονιδίωμα

1.2.1 Χρωμόσωμα

Το γονιδίωμα των *Streptomyces* αποτελείται από ένα ευμέγεθες γραμμικό χρωμόσωμα, με την περιεκτικότητα σε GC% να κυμαίνεται ανάμεσα σε 66% με 74% (Kampfer et al. 2014), ποσοστά ιδιαίτερα υψηλά που χαρακτηρίζουν αυτό το γένος. Το χρωμόσωμα του *S. coelicolor* A3(2), που είναι το στέλεχος αναφοράς, αρχικά θεωρούνταν κυκλικό. Μετά από αναλύσεις βιβλιοθηκών κοσμιδίων και συγκρίσεις με φυσικούς χάρτες διαπιστώθηκε ότι ήταν γραμμικό. Αυτή η γραμμικότητα του γονιδιώματος των *Streptomyces* θεωρείται ότι προέκυψε από γεγονός ανασυνδιασμού μονού χιάσματος (single-crossover) ανάμεσα σε ένα αρχικά κυκλικό χρωμόσωμα και ένα γραμμικό πλασμίδιο (Kampfer et al. 2014).

Αναλυτικότερα το στέλεχος αναφοράς *S. coelicolor* A3(2), διαθέτει περίπου 8.667 Mbp που κωδικοποιούν για 7.825 γονίδια (Bentley et al., 2002). Είναι σκόπιμο να παρατηρήσουμε σε αυτό το σημείο ότι ο *Saccharomyces cerevisiae* διαθέτει λιγότερα γονίδια, δηλαδή 6.275 (Liu et al., 2017), παρόλο που είναι ευκαρυώτης. Ο *S. avermitilis*, έχει ακόμα μεγαλύτερο γονιδίωμα, της τάξεως των 9.025 Mbp (Kampfer et al 2014).

Η κεντρική περιοχή αυτού του γονιδιώματος των *Streptomyces* είναι καλά συντηρημένη (Kampfer et al., 2014) όμως τα άκρα του έχουν τελικές ανάστροφες επαναλαμβανόμενες αλληλουχίες (Terminal Inverted Repeats, TIRs), το μέγεθος των οποίων κυμαίνεται από 20 ως 550kb και δεν είναι αρκετά συντηρημένες στους *Streptomyces*. Αυτές περιέχουν μη απαραίτητα γονίδια (Ikeda et al., 2003) και το μέγεθός τους διαφέρει έντονα από είδος σε είδος (Ventura et al. 2007). Μέσα, όμως, σε αυτές τις TIRs, τα πρώτα 200 ζεύγη βάσεων διαθέτουν πολλές παλίνδρομες αλληλουχίες και είναι καλά συντηρημένα. Τέλος, τα άκρα του χρωμοσώματος συνδέονται ομοιοπολικά με μια πρωτεΐνη, γνωστή ως terminal protein (TP), της οποίας το γονίδιο που την κωδικοποιεί βρίσκεται 100 kb πριν από το δεξί άκρο του χρωμοσώματος. Αυτή η πρωτεΐνη φαίνεται να εμπλέκεται στην προσκόλληση του DNA στη μεμβράνη κατά την κυτταρική διαίρεση και αντιγραφή. (Paradkar et al. 2003). Έχει επίσης παρατηρηθεί μεγάλη συντηρικότητα ανάμεσα στην κεντρική περιοχή του χρωμοσώματος του *S. coelicolor* A3(2) και μεγάλων τμημάτων των χρωμοσωμάτων των *Corynebacterium diphtheriae* και *Mycobacterium tuberculosis* (Bentley et al., 2002).

1.2.2 Πλασμίδια

Πέρα από το μεγάλο γραμμικό τους χρωμόσωμα, οι *Streptomyces* διαθέτουν και μεγάλα γραμμικά πλασμίδια, χωρίς να λείπουν βέβαια και τα κυκλικά. (Paradkar et al. 2003). Ένα πλέον γνωστό πλασμίδιο είναι το SCP1 του *S. coelicolor* A3(2), του οποίου το μέγεθος εκτιμάται στις 350 kb και φέρει τα βιοσυνθετικά γονίδια της μεθυλενομυκίνης. Το SCP1 φέρει κι αυτό TIRs περίπου 80 kb, τα οποία στα άκρα τους φέρουν κι αυτά μια terminal protein (Kinashi et al., 2011). Ένα άλλο πλασμίδιο, το pSLA2 του *S. rochei* έχει μέγεθος 17 kb και ομοιάζει το γενετικό υλικό του βακτηριοφάγου φ29. Ο *S. rochei* είναι γνωστός για την παραγωγή λανκασιδίνης (lanKacidin) και λανκαμυκίνης (lanKamycin), κι ενώ αρχικά θεωρούνταν ότι στη βιοσύνθεσή τους εμπλέκεται το εν λόγω πλασμίδιο, τελικά αποδείχθηκε ότι κάτι τέτοιο δεν ισχύει. Αντίθετα, στον *S. lasaliensis* ένα πλασμίδιο 520 kb, ονόματι pKSL επιβεβαιώθηκε ότι εμπλέκεται στην παραγωγή εχινομυκίνης (echinomycin), αφού όλο το βιοσυνθετικό μονοπάτι της, που βρισκόταν στο πλασμίδιο, εκφράστηκε στην *E. coli* και οδήγησε στην παραγωγή της (Watanabe et al., 2006). Διασκελισμοί έχουν περιγραφεί ανάμεσα στα γραμμικά χρωμοσώματα των *Streptomyces* και στα γραμμικά τους πλασμίδια, που μεταφέρουν τα μονοπάτια βιοσύνθεσης αντιβιοτικών από τη μια θέση στην άλλη. Ένα παράδειγμα είναι η

μεταφορά του μονοπατιού της οξυτετρακυκλίνης από το χρωμόσωμα του *S. rimosus* στο πλασμίδιο pPGZ3, η οποία κατέληξε σε ένα γιγάντιο μόριο 1 Mb (Kinashi et al., 2010).

1.3 Εφαρμογές στη Βιοτεχνολογία

Οι *Streptomyces* βρίσκουν εφαρμογή στη βιοτεχνολογία ως φορέας για την παραγωγή ετερόλογων πρωτεϊνών. Καταρχάς, όπως έχει ήδη αναφερθεί είναι θετικά κατά Gram βακτήρια. Αυτό σημαίνει ότι οι πρωτεΐνες που παράγουν, εκκρίνονται κατευθείαν στο θρεπτικό ανάπτυξης, καθώς, σε αντίθεση με τα Gram αρνητικά, δεν τους εμποδίζει το πολύπλοκο κυτταρικό τοίχωμα. Επιπλέον, κατ' αυτόν τον τρόπο δε σχηματίζονται (κυτταρικά) έγκλειστα και δεν υπάρχει κίνδυνος ανάμειξης των πρωτεϊνών με τα λιποσακχαρίδια της μεμβράνης, γεγονός που καθιστά εύκολη την περαιτέρω επεξεργασία τους (Hamed et al., 2018). Ιδιαίτερα, ο *S. lividans* χρησιμοποιείται για την παραγωγή ανασυνδυασμένων πρωτεϊνών εδώ και δύο δεκαετίες κι έχει μελετηθεί εκτενώς (Anné et al., 2014). Έχει αναδειχθεί (Busche et al., 2018, Ruckert et al., 2015) ως ιδανικός δέκτης λόγω της μειωμένης ενδογενούς δράσης των πρωτεασών του και του συστήματος restriction-modification. Ακόμα και πρωτεΐνες που δύσκολα μπορούν να παραχθούν στην *Escherichia coli*, η οποία ακόμα χρησιμοποιείται κατά κόρον στην ετερόλογη έκφραση πρωτεϊνών, παράχθηκαν με επιτυχία από τον *S. lividans* (Hamed et al., 2018).

1.4 Παλαιότερες μεθοδολογίες ταξινόμησης

Διάφορες τεχνολογίες και μεθοδολογίες έχουν χρησιμοποιηθεί για την εξελικτική μελέτη της οικογένειας *Streptomycetaceae*.

1.4.1 16s rRNA

Η χρήση του 16S rRNA, μέρος της μικρής υπομονάδας του προκαρυωτικού ριβοσώματος σε φυλογενετικές μελέτες βασίζεται στην ύπαρξη τμημάτων με υψηλή συντήρηση (Johnson et al., 2019), που απαντώνται σε όλα τα προκαρυωτικά ριβοσώματα. Για καιρό, υπήρξε το πιο κοινό ταξινομικό εργαλείο, ακριβώς λόγω της οικουμεικότητας και της σταθερότητάς του (Janda et al., 2007).

Η φυλογένεση με βάση αλληλουχίες 16S rRNA προσέφερε αρχικά το πλαίσιο για την συσχέτιση των διαφόρων ειδών της οικογένειας *Streptomycetaceae*, αλλά δεν παρείχε αρκετή ακρίβεια για την εγκυρότερη και καθοριστική ταυτοποίησή τους (Labeda

et al. 2017). Τέτοιες μελέτες οδήγησαν στο να συγκαταλλεγούν οι *Kitasatospora* μαζί με τους *Streptomyces* (Wellington et al. 1992), παρ' όλες τις διαφορές που παρουσίαζαν χημειοταξινομικά.

1.4.2 Multi Locus Sequence Analysis (MLSA)

Η βασική αρχή αυτής της μεθοδολογίας είναι η χρήση μερικών αλληλουχιών από γονίδια που κωδικοποιούν για πρωτεΐνες με καλά συντηρημένες λειτουργίες, δηλαδή “housekeeping genes” ώστε να δημιουργηθούν φυλογενετικά δέντρα, μέσα από τα οποία θα διαπιστωθούν σχέσεις ανάμεσα σε είδη (Glaeser et al. 2015). Από την ανάπτυξη της μεθοδολογίας αυτής κι έπειτα, έχει αποδειχθεί χρήσιμη και αποτελεσματική για την περαιτέρω διευκρίνηση των σχέσεων των ειδών της οικογένειας *Streptomycetaceae*, συμπληρώνοντας την μεθοδολογία 16S (Labeda et al. 2017). Χρησιμοποιήθηκε για να ρίξει φως στις σχέσεις μέσα σε ορισμένους κλάδους, όπως ο κλάδος των *Streptomyces albus* (Labeda et al. 2014) και *Streptomyces hirsutus* (Labeda et al. 2017). Το 2017, χρησιμοποιώντας MLSA, οι Labeda και συνεργάτες πρότειναν ότι ο αριθμός των ειδών του γένους *Streptomyces* θα μπορούσε, με διάφορες συμπύξεις, να μειωθεί δραματικά.

1.5 Γονιδιωματικές αναλύσεις

1.5.1 Ορολογία

Ο όρος pan-genome εδραιώθηκε το 2005 από τους Tettelin και συνεργάτες (Tettelin et al., 2005) και ορίζεται ως το σύνολο των γονιδίων μιας επιλεγμένης ομάδας οργανισμών. Η επιλογή της υπό μελέτη ομάδας οργανισμών έγκειται στον εκάστοτε ερευνητή και μπορεί να αφορά επίπεδο ορότυπου, είδους, φύλου ακόμα και βασιλείου (Vernikos et al., 2015).

Στις μελέτες του pan-genome χρησιμοποιείται συχνά η ίδια ορολογία. Αρχικά, ως “core genome” ή γονιδίωμα πυρήνα εννοούμε τα γονίδια που είναι κοινά σε όλους τους οργανισμούς της ομάδας. Αυτά συχνά είναι υπεύθυνα για τη βασική βιολογία τους και τα κύρια φαινοτυπικά χαρακτηριστικά τους. Από την άλλη, το “accessory” ή “dispensible” γονιδίωμα είναι κοινό σε υπο-ομάδες των οργανισμών ενδιαφέροντος και συμβάλλει στην ποικιλότητα, ανθεκτικότητα και προσαρμογή σε ιδιαίτερα περιβαλλοντικά ενδιαίτηματα. (Vernikos et al., 2015). Επίσης, υπάρχουν τα γονίδια που είναι μοναδικά

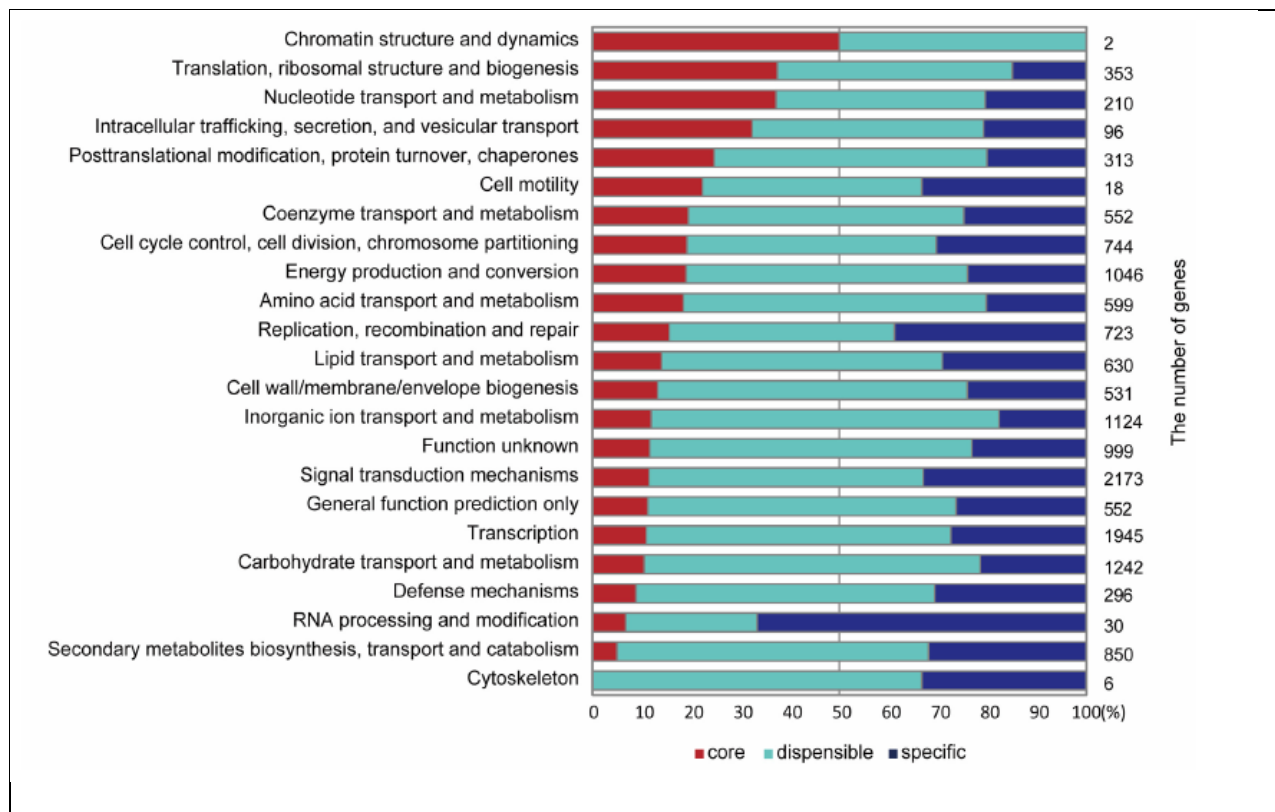
σε συγκεκριμένη υπο-ομάδα μελέτης, τα επονομαζόμενα γονίδια υπογραφές (fingerprints).

Πιστεύεται ότι η εν λόγω μεθοδολογία είναι η πλέον κατάλληλη για να διαλευκάνει λιγότερο εμφανείς διαφορές στο γονιδιακό περιεχόμενο, και ως εκ τούτου να διευκολύνει την πιο ορθή και αποτελεσματική (ανα)ταξινόμηση πολλών βακτηριακών ειδών (Caruto et al., 2019). Ως εκ τούτου, μπορεί να εφαρμοστεί και μέσα στην οικογένεια *Streptomycetaceae*.

1.5.2 Προηγούμενες γενωμικές μελέτες

Το 2012, οι (Zhou et al., 2012) διεξήγαν μια τέτοια μελέτη, χρησιμοποιώντας πέντε γονιδιώματα, αυτά των *S. avermitilis* MA-4680, *S. bingchenggensis* BCW-1, *S. coelicolor* A3(2), *S. griseus* IFO 13350 και *S. scabiei* 87-22. Από τα πειράματά τους προέκυψε ότι ανάμεσα στο ran-genome αυτών, το οποίο αποτελούταν από 17.362 ορθόλογες οικογένειες, υπήρχαν 3.096 core γονίδια. Τα υπόλοιπα συνίσταντο σε 5.066 γονίδια που υπήρχαν σε δύο ή περισσότερα είδη (accessory), αλλά και 9.200 γονίδια τα οποία ήταν μοναδικά στον εκάστοτε οργανισμό (fingerprint).

Μια επόμενη μελέτη ήρθε το 2015 (Kim et al., 2015), χρησιμοποιώντας αυτή τη φορά 17 γονιδιώματα. Το ran-genome στη συγκεκριμένη ανάλυση βρέθηκε ότι αποτελούταν από 1.129.413 συνολικά γονίδια, που σχημάτιζαν 34.592 ορθόλογα clusters. Από αυτά τα clusters, τα 11.743 εμφανίζονταν σε δύο ή περισσότερα είδη (accessory), τα συνολικά clusters που ήταν μοναδικά για κάθε οργανισμό ήταν 20.831, ενώ τα κοινά σε όλους (core) ήταν 2.018. Αναλύοντας τα δεδομένα περαιτέρω με βάση τις λειτουργικές κατηγορίες COG, βρέθηκε ότι στο core συχνότερα εμφανίζονταν οι κατηγορίες της μετάφρασης και του μεταβολισμού νουκλεοτιδίων (Εικόνα 5). Αντίθετα, τα γονίδια που εμπλέκονται στον δευτερογενή μεταβολισμό βρίσκονται ως επί το πλείστον στο accessory γονιδίωμα. Αυτό εξηγεί την εξαιρετική ποικιλότητα και μοναδικότητα του κάθε είδους στο γένος *Streptomyces*.

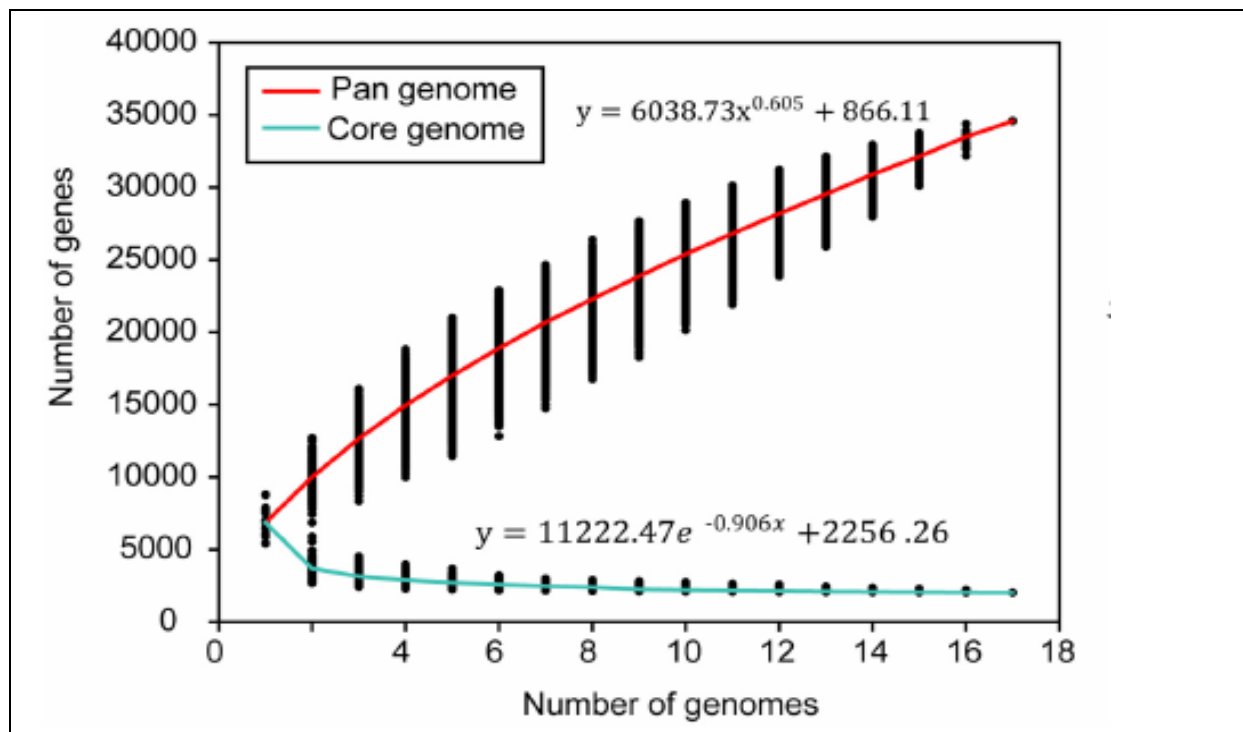


Εικόνα 5. Οι λειτουργικές κατηγορίες των core, dispensible και fingerprint (δηλ. μοναδικών) σύμφωνα με την βάση δεδομένων COG (Kim et al., 2015).

Μια επόμενη μελέτη (Jackson et al., 2018) χρησιμοποιώντας 13 δείγματα από θαλάσσιους σπόγγους εντόπισε 1699 core γονίδια, αποτελώντας μόλις το 12% ολόκληρου του pan-genome. Τα δεδομένα αυτών των ερευνητών έδειξαν πως υπάρχει μεγάλη διαφοροποίηση και ποικιλότητα εντός του γένος και ότι οι *Streptomyces* αποτελούν εύφορο έδαφος για την ανακάλυψη νέων δευτερογενών μεταβολιτών αφού από τα μεταβολικά μονοπάτια που διαπιστώθηκαν, μόνο περίπου το 18% είχε ομολογία με μονοπάτια γνωστής λειτουργίας.

Μια σημαντικά μεγαλύτερη μελέτη του 2019 (Xu et al., 2019) με 87 θαλάσσια είδη, ανέδειξε 996 core genes. Για άλλη μια φορά, αναδείχθηκε η τρομερή ικανότητα των *Streptomyces* να παράγουν πληθώρα δευτερογενών μεταβολιτών αλλά παράλληλα και η ανάγκη για περαιτέρω μελέτη τους.

Γενικά, όπως φαίνεται και στην Εικόνα 6, κατά την προσθήκη οργανισμών στο dataset, το pan-genome σε ορισμένες περιπτώσεις τείνει συνεχώς να αυξάνεται. Στη συγκεκριμένη περίπτωση θεωρείται ανοιχτό (open pan-genome), ενώ το core genome συνεχώς μειώνεται μέχρι να φτάσει σε πλατώ (Kim et al. 2015, Jackson et al. 2018).



Εικόνα 6. Μεταβολή των pan-genome και core genome με την εισαγωγή περισσότερων οργανισμών (Kim et al. 2015). Αξίζει να αναφερθεί ότι το συγκεκριμένο γράφημα αναφέρεται σε μικρό αριθμό γονιδιωμάτων, καθώς οι γενωμικές αναλύσεις πλέον γίνονται με εκατοντάδες γονιδιώματα.

Σκοπός της παρούσας εργασίας ήταν να πραγματοποιηθεί ανάλυση γονιδιώματος πυρήνα στα μέλη του γένους *Streptomyces*, ώστε να διαπιστωθεί αν υπάρχουν γονίδια υπογραφές που χαρακτηρίζουν συγκεκριμένα είδη.

2. Υλικά και μέθοδοι

2.1 Ubuntu

Τα Ubuntu αποτελούν ένα λειτουργικό σύστημα που ως πλατφόρμα των Linux, διατίθεται δωρεάν. Από την εμφάνισή του το 2004, έχουν συμβάλλει με την ευελιξία και την ευχρηστία τους αποτελεσματικά στη διαχείριση μεγάλου όγκου δεδομένων (Sobell, 2015)

2.2 Windows Subsystem for Linux

Το σύστημα αυτό επιτρέπει τη χρήση του τερματικού των Linux απευθείας από το λειτουργικό σύστημα Windows, χωρίς να είναι απαραίτητη κάποια εικονική μηχανή (virtual machine) ή η εγκατάσταση ενός ολόκληρου λειτουργικού συστήματος των Linux παράλληλα με τα Windows (Microsoft, 2021).

2.3 Perl

Η γλώσσα προγραμματισμού Perl (Practical Extraction and Report Language), που αναπτύχθηκε το 1988 από τον Larry Wall, είναι μια εύχρηστη γλώσσα που προσφέρει αρκετές δυνατότητες και υποστηρίζεται σε πολλές πλατφόρμες. Αρχικά, ο Wall δημιούργησε την Perl με σκοπό να λύσει ένα δυσανάλογο μικρό πρόβλημα. Έκτοτε, η ανάπτυξη της εν λόγω γλώσσας ήταν ραγδαία. Έγινε δημοφιλής στους τομείς του CGI (Computer-generated Imagery) και της Βιοπληροφορικής (Schwartz, 2008).

2.4 Python

Η Python είναι μια φιλική για αρχάριους, εύχρηστη και ευέλικτη γλώσσα. Τα εκτεταμένα και λεπτομερή εγχειρίδια καθώς και η πληθώρα βιβλιοθηκών την καθιστούν ιδανική για την επιστημονική έρευνα (Van Rossum και Drake, 2009).

2.4.1 Εργαλείο εύρεσης γονιδιώματος πυρήνα (Core genome analysis tool)

Το συγκεκριμένο εργαλείο (Nikolaidis et al., 2020) αποτελεί ένα pipeline για την εύρεση του γονιδιώματος πυρήνα ενός συνόλου οργανισμών μέσω του Basic Local Alignment Search Tool (BLAST). Με τη χρήση ενός οργανισμού αναφοράς (reference genome), εκτελεί ανταποδοτικό BLAST με το εργαλείο blast+ έναντι όλων των υπολοίπων και υπολογίζει τα ορθόλογα γονίδια για κάθε ζεύγος σύγκρισης. Ως παράμετρο E-value χρησιμοποιήθηκε τιμή μικρότερη από 1×10^{-5} . Στη συνέχεια, για κάθε ζεύγος σύγκρισης (οργανισμός αναφοράς – οργανισμός επερώτησης), υπολογίζει τη μέση τιμή και την τυπική απόκλιση της αμινοξικής ταυτότητας για όλα τα ορθόλογα και

απορρίπτει τα ορθόλογα που έχουν τιμή ταύτισης μικρότερη του μέσου όρου μετά από αφαίρεση δύο τυπικών αποκλίσεων. Με την ολοκλήρωση των αναζητήσεων και της επιλογής ορθολόγων, κατασκευάζονται δύο πίνακες. Ο πρώτος περιέχει όλες τις πρωτεΐνες του γονιδιώματος που έχουμε επιλέξει ως γονιδίωμα αναφοράς (μια ανά σειρά) αλλά και τα ορθολογά τους στα υπόλοιπα πρωτεώματα, με κάθε πρωτέωμα να εμφανίζεται ως μια στήλη. Ο δεύτερος πίνακας περιλαμβάνει τις πρωτεΐνες και τα ορθόλογα που απαντώνται σε όλα τα επιλεγμένα πρωτεώματα, δηλαδή εμφανίζει πλέον το core proteome. Στη συνέχεια, με τη βοήθεια του λογισμικού MUSCLE (Edgar, 2004), στοιχίζονται οι συστάδες πρωτεϊνών του core proteome, ενώνονται σε μια μεγαλύτερη αλληλουχία (super-alignment) και φιλτράρονται με το λογισμικό Gblocks (Castresana et al., 2000). Οι τελικές αυτές αλληλουχίες είναι αυτές που αξιοποιούνται για τη δημιουργία φυλογενωμικού δέντρου.

2.4.2 Plotly

Το plotly είναι μια βιβλιοθήκη της γλώσσας Python, η οποία χρησιμοποιείται για τη σχεδίαση γραφημάτων στους τομείς της Στατιστικής, των Οικονομικών αλλά και άλλων επιστημών. Παρέχει πλήθος απλών αλλά και πολύπλοκων γραφημάτων τα οποία είναι πλήρως διαδραστικά και διαμορφώσιμα από τον χρήστη (Plotly Technologies Inc., 2015).

2.5 IQ-Tree2

Το IQ-Tree2 είναι ένα εύχρηστο λογισμικό για την δημιουργία φυλογενετικών δέντρων, που λαμβάνει υπόψη τις εξελίξεις στον όγκο δεδομένων γονιδιωματικής. Λειτουργεί με βάση την εκτίμηση μέγιστης πιθανοφάνειας, ενώ επιτρέπει την επιλογή ανάμεσα σε διάφορα μοντέλα αντικατάστασης και ετερογένειας ρυθμού (Minh et al., 2020). Στη συγκεκριμένη εργασία χρησιμοποιήθηκε για τον υπολογισμό του φυλογενωμικού δέντρου από το εργαλείο εύρεσης γονιδιώματος πυρήνα. Ως παράμετροι (που τέθηκαν ως βέλτιστοι από το ίδιο το λογισμικό) χρησιμοποιήθηκε το μοντέλο αντικατάστασης LG+I+G4 και η αξιολόγηση των κλάδων έγινε με το approximate Likelihood Ratio Test (aLRT).

2.6 fastANI

Το fastANI είναι μια γρήγορη μέθοδος υπολογισμού της μέσης τιμής νουκλεοτιδικής ταύτισης (Average Nucleotide Identity; ANI) μεταξύ προκαρυωτικών γονιδιωμάτων (Jain et al., 2018). Με βάση την υπάρχουσα βιβλιογραφία έχει καθιερωθεί ότι το ποσοστό ομοιότητας ANI άνω του 95% σημαίνει ότι δύο οργανισμοί ανήκουν στο ίδιο είδος (Figueras et al., 2014).

2.7 Interactive Tree of Life (iTOL)

Το iTOL είναι ένα διαδικτυακό εργαλείο για την οπτικοποίηση και χαρακτηρισμό φυλογενετικών/φυλογενωμικών δέντρων. Παρέχει διάφορους τύπους προβολής των δέντρων, πολλαπλά εργαλεία για την διαχείριση των δεδομένων που προβάλλονται πάνω σε αυτά και στηρίζει αρκετές μορφές αρχείων εισόδου (Newick, Nexus, PhyloXML) (Letunic & Bork, 2021).

2.8 NCBI Assembly

Το NCBI Assembly αποτελεί δημόσια βάση δεδομένων για συναρμολογημένα γονιδιώματα (assemblies). Στη παρούσα εργασία χρησιμοποιώντας ως επερώτηση το taxonomy ID της τάξης *Streptomyetales* (txid85011) και ως επίπεδο συναρμολόγησης το πλήρες γονιδίωμα ή το χρωμόσωμα παρελήφθησαν δεδομένα τριακοσίων ογδόντα επτά (387) καταχωρίσεων.

2.9 Markov Cluster Algorithm (MCL)

Ο αλγόριθμος Markov Cluster (MCL) στοχεύει στην ομαδοποίηση με βάση τη θεωρία γράφων. Ουσιαστικά χρησιμοποιεί δύο διεργασίες για να διασχίσει έναν σταθμισμένο γράφο, το “inflation” και το “expansion”. Κατά την πρώτη, οι σχέσεις ανάμεσα σε γειτονικούς κόμβους ενισχύονται, ενώ οι σχέσεις με μακρινούς κόμβους αποδυναμώνονται, και ακολουθεί κανονικοποίηση. Από την άλλη, το “expansion” είναι υπεύθυνο για την επισήμανση των σχέσεων ανάμεσα στις ομάδες κόμβων, δηλαδή clusters, που δημιουργήθηκαν προηγουμένως (Enright et al., 2002).

2.10 eggNOG-mapper

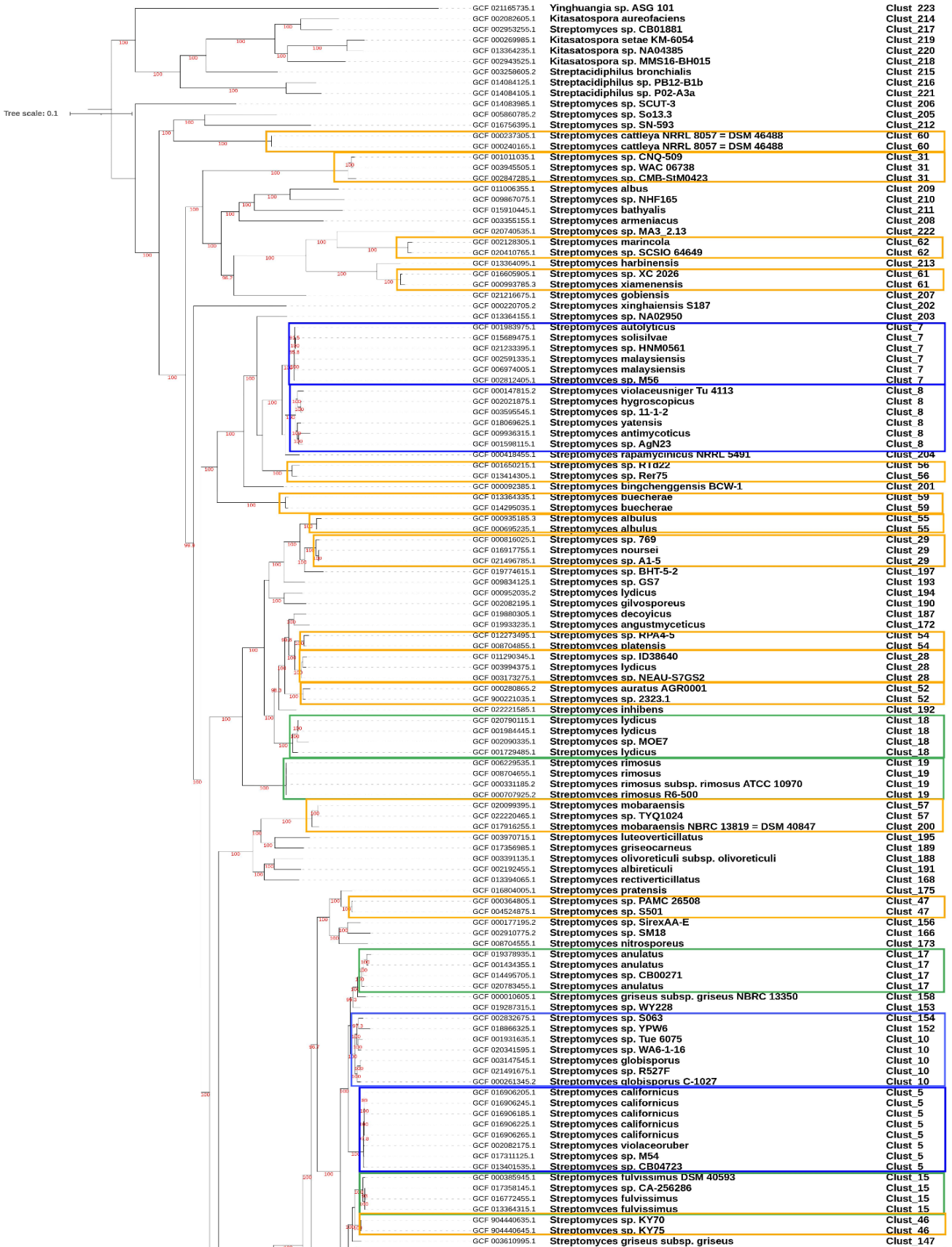
Το συγκεκριμένο εργαλείο αξιοποιήθηκε στην κατάταξη των core πρωτεϊνών σε λειτουργικές κατηγορίες. Το eggNOG-mapper χρησιμοποιεί συσχετίσεις ανάμεσα σε ορθόλογες ομάδες (Orthologous Groups, OGs) της βάσης δεδομένων eggNOG για το χαρακτηρισμό της λειτουργίας της εκάστοτε αλληλουχίας επερώτησης (Cantalapiedra et al., 2021)(Huerta-Cepas et al., 2019).

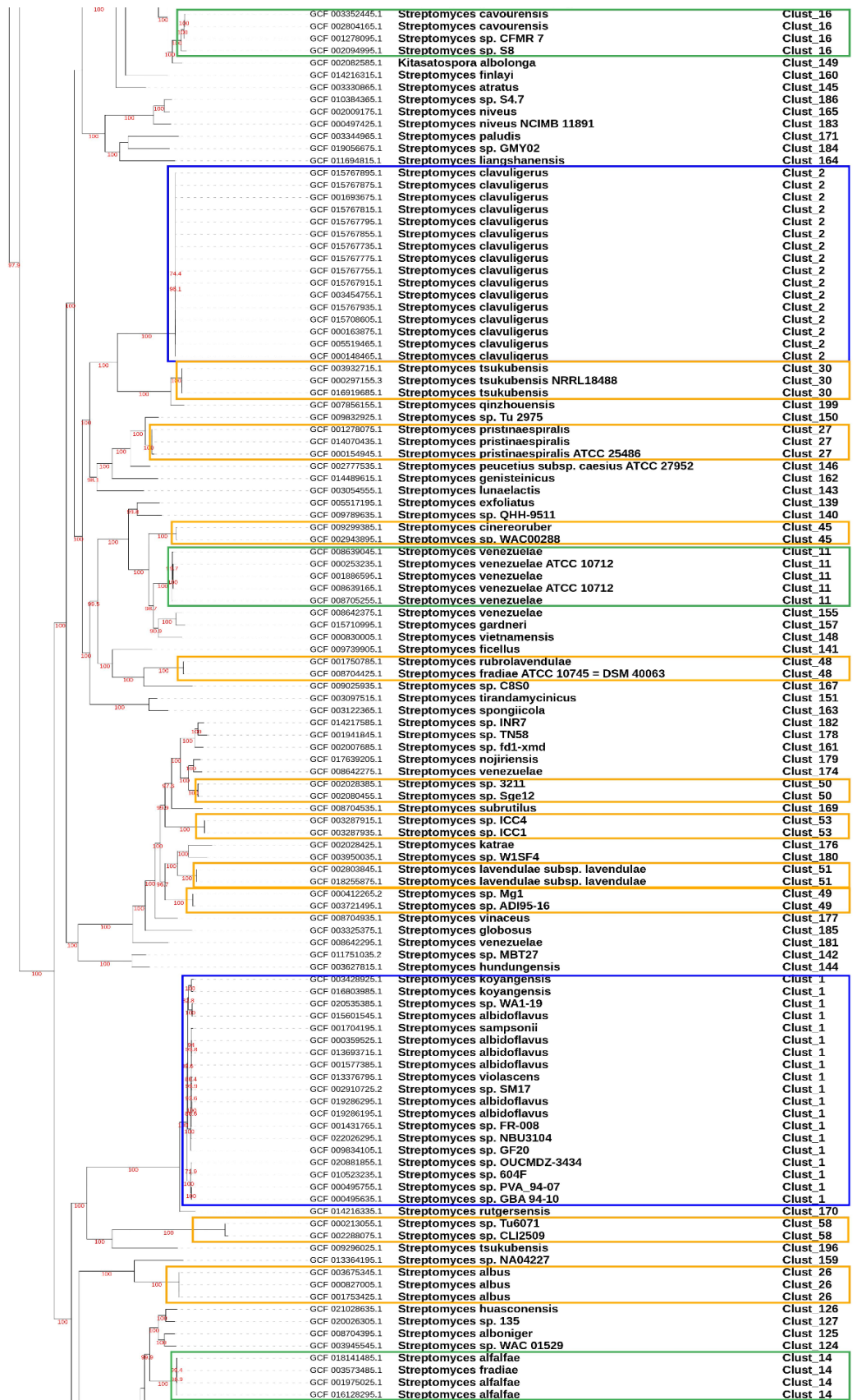
3. Αποτελέσματα - Συζήτηση

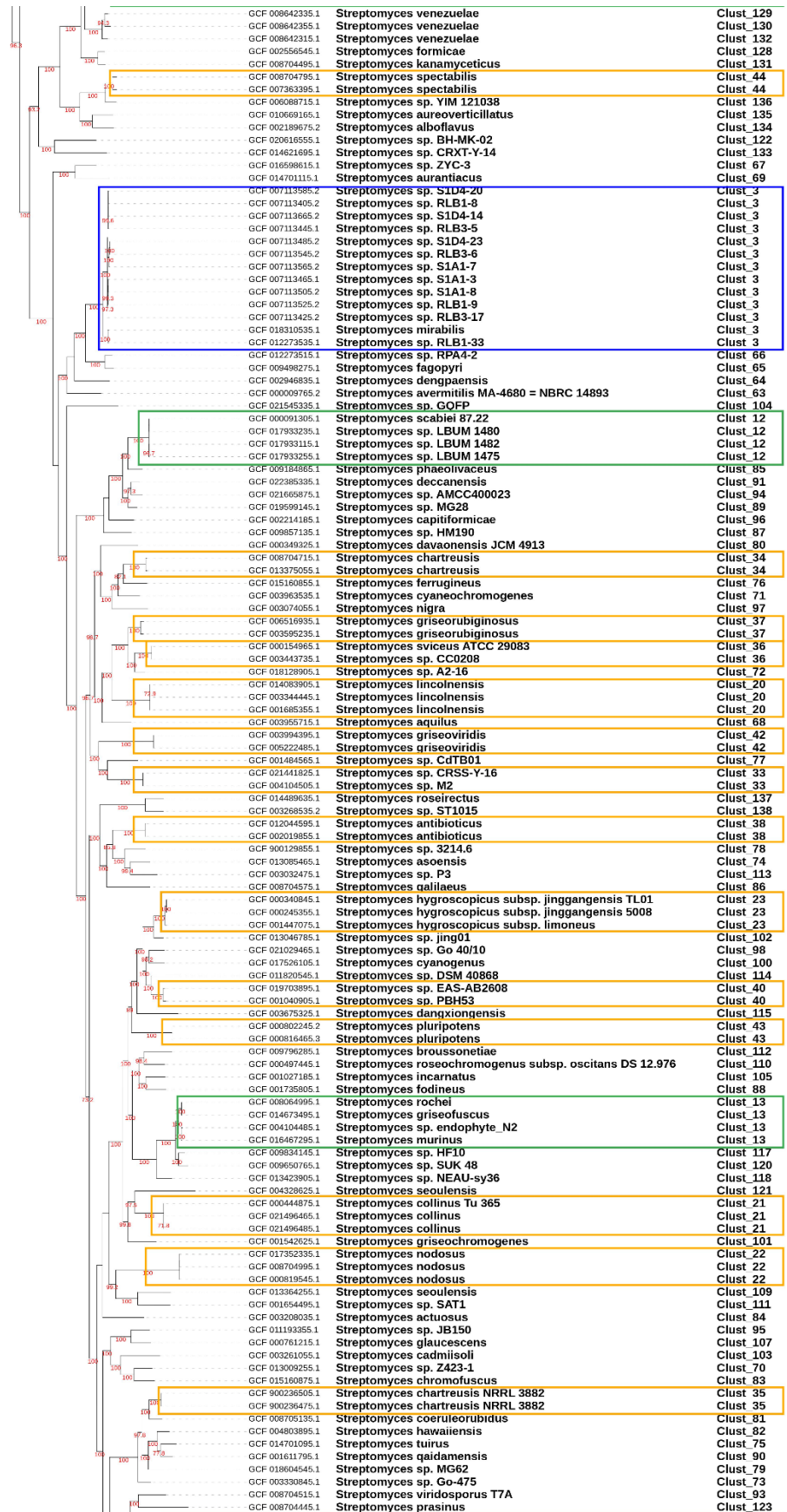
Τα 387 πρωτεύματα που ελήφθησαν από το NCBI Assembly δόθηκαν ως δεδομένα εισόδου στο εργαλείο εύρεσης του γονιδιώματος πυρήνα που προαναφέρθηκε, ώστε να προκύψει το παρακάτω φυλογενωμικό δέντρο (Εικόνα 7). Τα μέλη του γένους *Kitasatospora* (*K. aureofaciens*, *K. setae* KM-6054, *Kitasatospora* sp. NA04385, *Kitasatospora* sp. MMS16-BH015) και *Streptacidiphilus* (*S. bronchialis*, *S. sp.* PB12-B1b, *Streptacidiphilus* sp. P02-A3a), καθώς και ο μοναδικός οργανισμός του γένους *Yinghuangia* (*Yinghuangia* sp. ASG 101), διαχωρισμένοι αποτελούν εξω-ομάδα από τους *Streptomyces* με μοναδική εξαίρεση τον *Streptomyces* sp. CB01881. Ο κοντινότερος συγγενής αυτού του οργανισμού με κάποιο άλλο συναρμολογημένο γονιδίωμα είναι με το *Kitasatospora misakiensis*, που ήταν αρχικά γνωστό ως *Streptomyces misakiensis* (NCBI Assembly, 2022).

Αξίζει να σημειωθεί ότι υπάρχει ένας οργανισμός *Kitasatospora albolonga* εντός του κλάδου των *Streptomyces*, περί τη μέση του δέντρου, που πιθανώς είναι mis-annotation. Το εν λόγω είδος ήταν γνωστό ως *Streptomyces albolongus* έως το 2017, όταν μετονομάστηκε (Labeda et al., 2017). Και οι δύο εξαιρέσεις έχουν υψηλή τιμή aLRT support. Για να διασαφηνίσουμε τον ορθό διαχωρισμό σε είδη των καταχωρήσεων που χρησιμοποιήθηκαν στο παρόν πείραμα, εφαρμόστηκε το fastANI. Το output του fastANI χρησιμοποιήθηκε ως input για τον αλγόριθμο Markov Cluster ώστε τα γονιδιώματα με τιμές ANI μεγαλύτερες του 95% να δημιουργήσουν μια συστάδα (cluster) που αντιπροσωπεύει ένα είδος. Από αυτή τη διαδικασία προέκυψαν πολλά clusters και συγκεκριμένα διακόσια είκοσι τρία (223) για όλη την τάξη *Streptomycetales* και διακόσια δεκατέσσερα (214) για το γένος *Streptomyces*, καταδεικνύοντας την εξαιρετική ετερογένεια του γένους.

Έπειτα, επιλέχθηκαν τα clusters με περισσότερα από 5 μέλη (Εικόνα 7) για περαιτέρω ανάλυση, που είναι δέκα (10) σε αριθμό. Επειδή το κάθε cluster είχε διαφορετικό αριθμό μελών, κανονικοποιήθηκαν όλα τα clusters ώστε να έχουν το μέγιστο πέντε μέλη. Τα πέντε αυτά μέλη επιλέχθηκαν με βάση το δέντρο ως εκπρόσωποι του κάθε cluster που δεν γειτνιάζουν.









Εικόνα 7. Το φυλογενωμικό δέντρο της οικογένειας *Streptomycetales*. Ο υπολογισμός του δέντρου έγινε με το IQ-Tree2 χρησιμοποιώντας 259 πρωτεΐνες πυρήνα (56013 sites). Σε μπλε πλαίσια παρατηρούνται τα είδη (MCL clusters) με > 5 μέλη, με πράσινο αυτά με 4 μέλη, ενώ με πορτοκαλί τα είδη με 2-3 μέλη.

Με βάση τα κανονικοποιημένα είδη υπολογίστηκε το core proteome, δηλαδή πρωτεΐνες που είναι κοινές σε όλους τους πέντε επιλεγμένους οργανισμούς του εκάστοτε είδους. Επίσης υπολογίστηκε ο αριθμός των πρωτεϊνών που είναι παρούσες σε όλα τα μέλη του είδους και ταυτόχρονα μοναδικές για το συγκεκριμένο είδος, δηλαδή οι fingerprint πρωτεΐνες (Πίνακας 1).

Παρατηρούμε το Cluster 2 να κυριαρχεί με μεγάλη διαφορά από τα υπόλοιπα στον αριθμό των μοναδικών πρωτεϊνών (fingerprints). Η συγκεκριμένη ομάδα αποτελείται αποκλειστικά από στελέχη του είδους *Streptomyces clavuligerus*, που ξεχωρίζει λόγω της αδυναμίας του να χρησιμοποιεί τη γλυκόζη ως πηγή άνθρακα, αφού δε διαθέτει το γονίδιο περμεάσης της γλυκόζης *glP1* (Liras και Martin, 2021). Παρόλα αυτά, ο *S. clavuligerus* έχει αναδειχθεί για την ικανότητα του να παράγει πληθώρα δευτερογενών μεταβολιτών, πολύ περισσότερων από το μέσο *Streptomyces* (Hwang et al., 2021), κάτι που ενδεχομένως εξηγεί τις υπεράριθμες fingerprint πρωτεΐνες.

Πίνακας 1. Τα επιλεγμένα clusters, το στέλεχος αναφοράς (reference strain), οι αριθμοί γονιδιωμάτων πριν και μετά την κανονικοποίηση, οι core πρωτεΐνες μετά την κανονικοποίηση και ο αριθμός των κανονικοποιημένων fingerprint.

Taxonomy	Reference strain	Number of genomes	Number of genomes in Normalized dataset	Core proteins (normalized dataset)	Number of fingerprints (normalized dataset)
Order	GCF_008931305.1	387	344	264	-
Cluster_1	GCF_016803985.1	19	5	4363	0
Cluster_2	GCF_015767775.1	16	5	5216	158

Cluster_3	GCF_018310535.1	13	5	5532	0
Cluster_4	GCF_003665435.1	9	5	5393	45
Cluster_5	GCF_016906185.1	8	5	6077	36
Cluster_6	GCF_008931305.1	6	5	5180	4
Cluster_7	GCF_006974005.1	6	5	6110	23
Cluster_8	GCF_009936315.1	6	5	5093	5
Cluster_9	GCF_021383925.1	5	5	5673	6
Cluster_10	GCF_003147545.1	5	5	4781	0
Cluster_11	GCF_000253235.1	5	5	6085	0

Ακολούθως, χρησιμοποιήθηκε το eggNOG-mapper για την κατάταξη των core (βλ. Πίνακας 2), αλλά και των fingerprint πρωτεϊνών (βλ. Πίνακας 3) σε λειτουργικές κατηγορίες. Το όνομα της κάθε κατηγορίας αναγράφεται στον Πίνακας 4.

Όπως ήδη έχει αναφερθεί, στο Cluster2 (*S. clavuligerus*) παρατηρείται μεγάλος αριθμός fingerprint πρωτεϊνών. Με τη βοήθεια του eggNOG-mapper, φαίνεται ότι η πλειονότητα αυτών των πρωτεϊνών είναι ακόμα άγνωστης λειτουργίας (κατηγορία S, Πίνακας 4). Διασταυρώνοντας με τους χαρακτηρισμούς του NCBI, προκύπτει ότι η συντριπτική πλειοψηφία από τις συγκεκριμένες πρωτεΐνες άγνωστης λειτουργίας είναι χαρακτηρισμένες είτε ως υποθετικές (115/130; ~85%), είτε ως πρωτεΐνες με επικράτεια άγνωστης λειτουργίας DUF domain-containing proteins (10/130; ~8%). Ωστόσο, αξίζει να σημειωθεί ότι βρέθηκαν δύο αναστολείς της β-λακταμάσης (beta-lactamase inhibitor protein; BLIP) ως γονίδια υπογραφές. Ο *S. clavuligerus* είναι γνωστός για την παραγωγή του κλαβουλανικού οξέος (clavulanic acid), ενός αναστολέα της β-λακταμάσης (Liras και Martin, 2021). Επίσης, έχει βρεθεί ότι μια DUF397 είναι μέρος μιας πρωτεΐνης που επίσης περιέχει στοιχείο απόκρισης σε ξеноβιοτικές ουσίες (xenobiotic response element, XRE) (Hwang et al., 2021).

Πίνακας 2. Οι λειτουργικές κατηγορίες (COG categories) των πρωτεϊνών πυρήνα (core proteins) των υπό-μελέτη ειδών.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	S	T	U	V
Order	0	0	20	7	20	14	6	13	16	43	28	5	5	0	13	6	2	55	8	2	1
Cluster_1	2	4	240	62	434	106	266	140	202	201	462	147	176	16	130	158	129	1137	206	40	105
Cluster_2	3	3	256	70	489	121	256	149	258	201	525	201	226	20	155	169	167	1556	231	35	125
Cluster_3	5	2	310	60	570	121	398	139	290	218	643	154	202	20	152	182	166	1484	261	34	121
Cluster_4	5	4	276	67	471	104	375	149	245	207	573	178	225	25	153	165	154	1590	255	40	132
Cluster_5	1	3	306	58	570	111	314	156	297	209	612	162	226	14	153	221	177	2069	258	40	120
Cluster_6	3	3	294	59	485	117	400	142	224	223	629	148	193	24	139	171	142	1355	276	37	116
Cluster_7	2	3	386	51	610	118	486	159	349	208	755	159	236	20	158	206	231	1527	296	35	115
Cluster_8	1	2	320	52	515	101	380	130	274	194	653	136	186	20	141	166	173	1275	248	29	97
Cluster_9	4	3	315	69	540	122	451	151	236	236	643	173	228	21	163	194	151	1509	292	36	136

Cluster_10	2	2	240	52	446	99	303	125	217	187	501	140	192	10	126	176	119	1510	213	33	88
Cluster_11	3	3	343	80	619	126	355	165	307	225	741	174	255	19	167	187	181	1666	271	41	157

Πίνακας 3. Οι λειτουργικές κατηγορίες (COG categories) των μοναδικών (fingerprint) πρωτεϊνών των υπό-μελέτη ειδών.

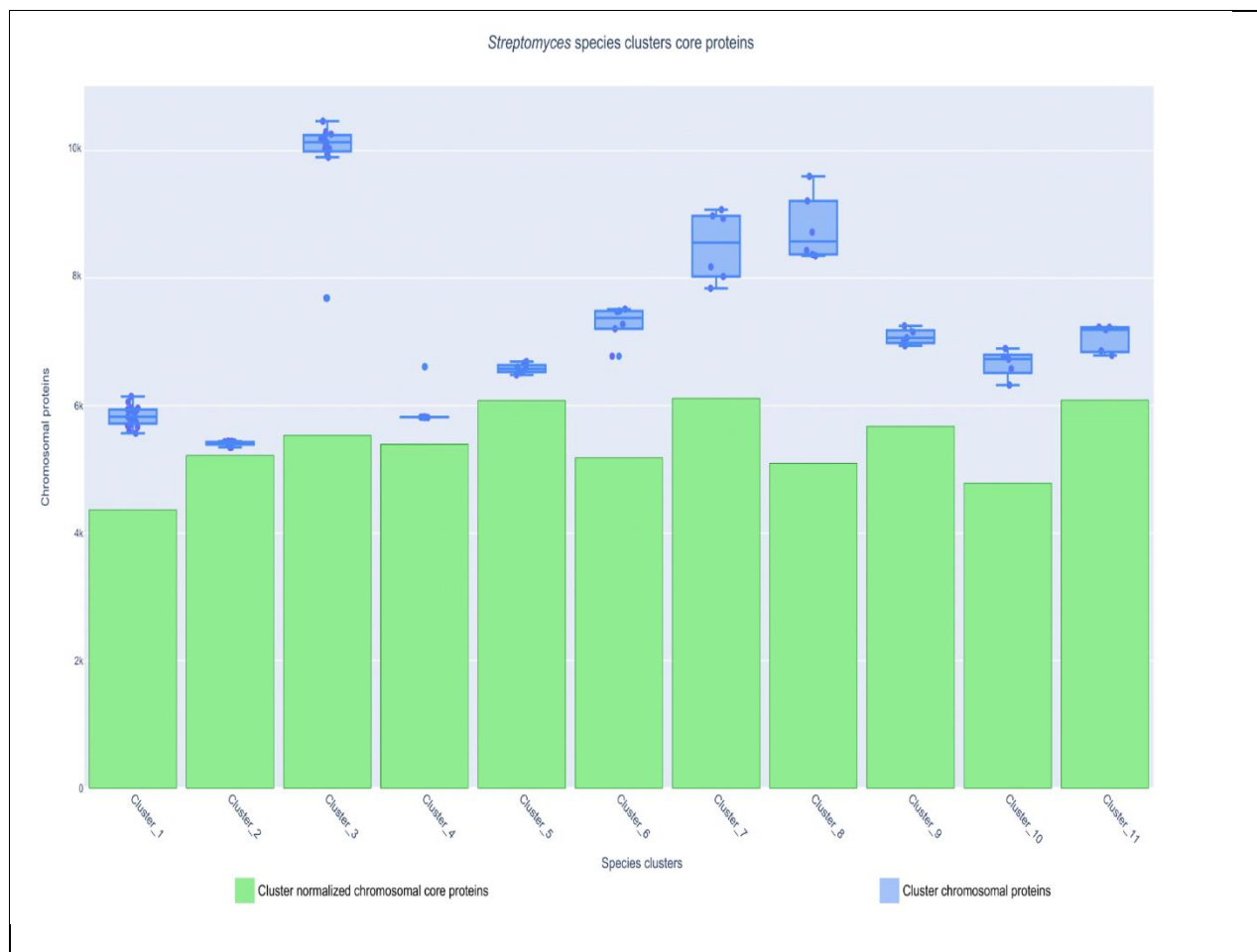
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	S	T	U	V
Cluster_1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cluster_2	0	0	1	0	5	0	1	0	0	1	8	0	1	0	0	4	4	130	2	0	1
Cluster_3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cluster_4	0	0	0	0	0	0	1	0	3	0	0	0	0	0	0	0	0	39	0	1	1
Cluster_5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	32	0	0	1	
Cluster_6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0
Cluster_7	0	0	1	0	4	0	0	0	1	0	1	0	0	0	1	2	12	0	0	1	
Cluster_8	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	2	0	0	0	
Cluster_9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	1	0	0
Cluster_10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cluster_11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Πίνακας 4. Περιγραφή των λειτουργικών κατηγοριών

Category symbol	Annotation
A	RNA processing and modification
B	Chromatin structure and dynamics
C	Energy production and conversion
D	Cell cycle control, cell division, chromosome partitioning
E	Amino acid transport and metabolism
F	Nucleotide transport and metabolism
G	Carbohydrate transport and metabolism
H	Coenzyme transport and metabolism
I	Lipid transport and metabolism
J	Translation, ribosomal structure and biogenesis
K	Transcription
L	Replication, recombination and repair
M	Cell wall/membrane/envelope biogenesis
N	Cell motility
O	Post-translational modification, protein turnover, and chaperones
P	Inorganic ion transport and metabolism
Q	Secondary metabolites biosynthesis, transport, and catabolism
S	Function unknown
T	Signal transduction mechanisms

U	Intracellular trafficking, secretion, and vesicular transport
V	Defense mechanisms

Συγκρίνοντας τις κανονικοποιημένες πρωτεΐνες πυρήνα κάθε είδους με τον συνολικό αριθμό των πρωτεϊνών όλων των οργανισμών του συγκεκριμένου είδους, λαμβάνεται μια εικόνα για το accessory γονιδίωμα (Εικόνα 8). Γενικά, φαίνεται ότι ο αριθμός των πρωτεϊνών πυρήνα για κάθε είδος είναι σχετικά σταθερός, ενώ τα επιμέρους είδη εμφανίζουν μεγάλες διακυμάνσεις στον αριθμό των accessory πρωτεϊνών.



Εικόνα 8. Bar-chart των core πρωτεϊνών του κάθε cluster (πράσινο) σε συνδιασμό με boxplot των συνολικών πρωτεϊνών του κάθε cluster (μπλε).

Συμπεράσματα

Συμπερασματικά, η φυλογενωμική ανάλυση έδειξε ότι το γένος *Streptomyces* αποτελείται από ένα μεγάλο αριθμό ειδών. Γενικά, ο αριθμός των πρωτεϊνών πυρήνα για κάθε είδος είναι σχετικά σταθερός, ενώ τα επιμέρους είδη εμφανίζουν μεγάλες διακυμάνσεις στον αριθμό των accessory πρωτεϊνών. Κάποια από αυτά τα είδη έχουν εξελίξει ένα μεγάλο αριθμό πρωτεϊνών fingerprints που χαρακτηρίζουν το εκάστοτε είδος. Η πλειονότητα των πρωτεϊνών αυτών είναι αγνώστου λειτουργίας, γεγονός που καταδεικνύει ότι υπάρχει μεγάλη ανάγκη για βιοχημικό/λειτουργικό χαρακτηρισμό των πρωτεϊνών αυτών, ώστε να κατανοήσουμε καλύτερα πως προσαρμόστηκαν στο ενδιαίτημά τους.

Βιβλιογραφία

- Anné J, Vrancken K, Van Mellaert L, Van Impe J & Bernaerts K (2014) Protein secretion biotechnology in Gram-positive bacteria with special emphasis on *Streptomyces lividans*. *Biochim Biophys Acta* 1843: 1750-1761.
- Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., von Haeseler A., Lanfear R. (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.*, 37:1530-1534. <https://doi.org/10.1093/molbev/msaa015>
- Bentley, S.D., Chater, K.F., Cerdeño-Tárraga, A.-M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.-H., Kieser, T., Larke, L., Murphy, L., Oliver, K., O’Neil, S., Rabinowitsch, E., Rajandream, M.-A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J., Hopwood, D.A., 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417, 141–147. <https://doi.org/10.1038/417141a>
- Bergey, D. H., Krieg, N. R., & Holt, J. G. (1984). *Bergey's manual of systematic bacteriology*. Baltimore, MD: Williams & Wilkins.
- Brangsch, H., Höller, M., Krauß, T., Waqas, M., Schroeckh, V., Brakhage, A.A., Bunk, B., Spröer, C., Overmann, J., Kothe, E., (2022). Extremophile Metal Resistance: Plasmid-Encoded Functions in *Streptomyces mirabilis*. *Appl Environ Microbiol* 88, e0008522. <https://doi.org/10.1128/aem.00085-22>
- Busche T, Tsolis KC, Koepff J, et al. (2018) Multi-Omics and Targeted Approaches to Determine the Role of Cellular Proteases in *Streptomyces* Protein Secretion. *Front Microbiol* 9: 1174.
- Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., Huerta-Cepas, J., (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution* 38, 5825–5829. <https://doi.org/10.1093/molbev/msab293>
- Caputo, A., Fournier, P.-E., Raoult, D., (2019). Genome and pan-genome analysis to classify emerging bacteria. *Biol Direct* 14, 5. <https://doi.org/10.1186/s13062-019-0234-0>
- Caruso, M., Colombo, A., Crespi-Perellino, N., Fedeli, L., Malyszko, J., Pavesi, A., Quaroni, S., Saracchi, M., Ventrella, G., 2000. Studies on a Strain of *Kitasatospora* sp. Paclitaxel Producer. *Annals of Microbiology* 50, 89–102.
- 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>

- Enright, A.J., Van Dongen, S., Ouzounis, C.A., (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30, 1575–1584. <https://doi.org/10.1093/nar/30.7.1575>
- Figueras, M.J., Beaz-Hidalgo, R., Hossain, M.J., Liles, M.R., (2014). Taxonomic Affiliation of New Genomes Should Be Verified Using Average Nucleotide Identity and Multilocus Phylogenetic Analysis. *Genome Announc* 2, e00927-14. <https://doi.org/10.1128/genomeA.00927-14>
- Glaeser, S. P., & Kämpfer, P. (2015). Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Systematic and Applied Microbiology*, 38(4), 237–245. <https://doi.org/10.1016/j.syapm.2015.03.007>
- Hamed, M. B., Anné, J., Karamanou, S., & Economou, A. (2018). Streptomyces protein secretion and its application in biotechnology. *FEMS Microbiology Letters*, 365(22). <https://doi.org/10.1093/femsle/fny250>
- Hopwood, D. A. (2019). Highlights of Streptomyces genetics. *Heredity*, 123(1), 23–32. <https://doi.org/10.1038/s41437-019-0196-0>
- Hwang, S., Lee, N., Choe, D., Lee, Y., Kim, W., Jeong, Y., Cho, S., Palsson, B.O., Cho, B.-K., (2021) Elucidating the Regulatory Elements for Transcription Termination and Posttranscriptional Processing in the Streptomyces clavuligerus Genome. *mSystems* 6, e01013-20. <https://doi.org/10.1128/mSystems.01013-20>
- Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M., Ōmura, S., 2003. Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis. *Nat Biotechnol* 21, 526–531. <https://doi.org/10.1038/nbt820>
- Jackson, S.A., Crossman, L., Almeida, E.L., Margassery, L.M., Kennedy, J., Dobson, A.D.W., (2018). Diverse and Abundant Secondary Metabolism Biosynthetic Gene Clusters in the Genomes of Marine Sponge Derived Streptomyces spp. Isolates. *Mar Drugs* 16, E67. <https://doi.org/10.3390/md16020067>
- Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T., Aluru, S., (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 9, 5114. <https://doi.org/10.1038/s41467-018-07641-9>
- Janda, J.M., Abbott, S.L., (2007). 16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls. *J Clin Microbiol* 45, 2761–2764. <https://doi.org/10.1128/JCM.01228-07>
- Johnson, J.S., Spakowicz, D.J., Hong, B.-Y., Petersen, L.M., Demkowicz, P., Chen, L., Leopold, S.R., Hanson, B.M., Agresta, H.O., Gerstein, M., Sodergren, E., Weinstock, G.M., (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun* 10, 5029. <https://doi.org/10.1038/s41467-019-13036-1>
- Kämpfer, P., Glaeser, S. P., Parkes, L., van Keulen, G., & Dyson, P. (2014). The Family Streptomycetaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F.

- Thompson (Eds.), *The Prokaryotes: Actinobacteria* (pp. 889–1010). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30138-4_184
- Kim, J.-N., Kim, Y., Jeong, Y., Roe, J.-H., Kim, B.-G., Cho, B.-K., (2015). Comparative Genomics Reveals the Core and Accessory Genomes of *Streptomyces* Species. *J Microbiol Biotechnol* 25, 1599–1605. <https://doi.org/10.4014/jmb.1504.04008>
 - Kim, S. B., Lonsdale, J., Seong, C.-N., & Goodfellow, M. (2003). *Streptacidiphilus* gen. Nov., acidophilic actinomycetes with wall chemotype I and emendation of the family Streptomycetaceae (Waksman and Henrici (1943)AL) emend. Rainey et al. 1997. *Antonie Van Leeuwenhoek*, 83(2), 107–116. <https://doi.org/10.1023/a:1023397724023>
 - Kinashi, H. (2011). Giant linear plasmids in *Streptomyces*: A treasure trove of antibiotic biosynthetic clusters. *The Journal of Antibiotics*, 64(1), 19–25. <https://doi.org/10.1038/ja.2010.146>
 - Labeda, D. P., Dunlap, C. A., Rong, X., Huang, Y., Doroghazi, J. R., Ju, K.-S., & Metcalf, W. W. (2017). Phylogenetic relationships in the family Streptomycetaceae using multi-locus sequence analysis. *Antonie Van Leeuwenhoek*, 110(4), 563–583. <https://doi.org/10.1007/s10482-016-0824-0>
 - Labeda, D.P., Doroghazi, J.R., Ju, K.-S., Metcalf, W.W., (2014). Taxonomic evaluation of *Streptomyces albus* and related species using multilocus sequence analysis and proposals to emend the description of *Streptomyces albus* and describe *Streptomyces pathocidini* sp. nov. *Int J Syst Evol Microbiol* 64, 894–900. <https://doi.org/10.1099/ijs.0.058107-0>
 - Law, J., Pusparajah, P., Ab Mutalib, N.-S., Wong, S., Goh, B. H., & Lee, L. H. (2019). A Review on Mangrove Actinobacterial Diversity: The Roles of *Streptomyces* and Novel Species Discovery. *Progress In Microbes & Molecular Biology*, 2. <https://doi.org/10.36877/pmmb.a0000024>
 - Letunic, I., Bork, P., (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research* 49, W293–W296. <https://doi.org/10.1093/nar/gkab301>
 - Li, Y., Wang, M., Sun, Z.-Z., Xie, B.-B., (2021). Comparative Genomic Insights Into the Taxonomic Classification, Diversity, and Secondary Metabolic Potentials of *Kitasatospora*, a Genus Closely Related to *Streptomyces*. *Frontiers in Microbiology* 12.
 - Liras, P., Martín, J.F., (2021). *Streptomyces clavuligerus*: The Omics Era. *J Ind Microbiol Biotechnol* 48, kuab072. <https://doi.org/10.1093/jimb/kuab072>
 - Liu, W., Li, L., Ye, H., Chen, H., Shen, W., Zhong, Y., Tian, T., He, H., (2017). From *Saccharomyces cerevisiae* to human: The important gene co-expression modules. *Biomed Rep* 7, 153–158. <https://doi.org/10.3892/br.2017.941>
 - Madeira, F., Pearce, M., Tivey, A.R.N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., Lopez, R., (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res* gkac240. <https://doi.org/10.1093/nar/gkac240>

- Madhaiyan, M., Saravanan, V.S. and See-Too, W.-S. (2020) 'Genomic and phylogenomic insights into the family Streptomycetaceae lead to proposal of Charcoactinosporaceae fam. nov. and 8 novel genera with emended descriptions of *Streptomyces calvus*', bioRxiv, p. 2020.07.08.193797. Available at: <https://doi.org/10.1101/2020.07.08.193797>.
- Nikolaidis, M., Mossialos, D., Oliver, S., Amoutzias, G., (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>
- Paradkar, A., Trefzer, A., Chakraborty, R., & Stassi, D. (2003). *Streptomyces* genetics: A genomic perspective. *Critical Reviews in Biotechnology*, 23(1), 1–27. <https://doi.org/10.1080/713609296>
- Parte, A.C., Sardà Carbasse, J., Meier-Kolthoff, J.P., Reimer, L.C. and Göker, M. (2020). List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *International Journal of Systematic and Evolutionary Microbiology*, 70, 5607-5612; DOI: 10.1099/ijsem.0.004332
- Plotly Technologies Inc., (2015), Collaborative data science, Plotly Technologies Inc., <https://plot.ly>
- Ruckert C, Albersmeier A, Busche T, et al. (2015) Complete genome sequence of *Streptomyces lividans* TK24. *J Biotechnol* 199: 21-22.
- Schwartz R. L. (2008), *Learning Perl*, 5th Edition, California, U.S.A.: O' Reilly Media Inc.
- Sobell, M.G., (2015). *A practical guide to Ubuntu Linux*, Pearson Education.
- Song, W., Duan, L., Jin, L., Zhao, J., Jiang, S., Sun, T., Guo, X.W., Xiang, W., Wang, X., (2018). *Streptacidiphilus monticola* sp. nov., a novel actinomycete isolated from soil. *Int J Syst Evol Microbiol* 68, 1757–1761. <https://doi.org/10.1099/ijsem.0.002751>
- Takahashi, Y. (2017). Genus *Kitasatospora*, taxonomic features and diversity of secondary metabolites. *The Journal of Antibiotics*, 70(5), 506–513. <https://doi.org/10.1038/ja.2017.8>
- Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L., Durkin, A.S., DeBoy, R.T., Davidsen, T.M., Mora, M., Scarselli, M., Margarit y Ros, I., Peterson, J.D., Hauser, C.R., Sundaram, J.P., Nelson, W.C., Madupu, R., Brinkac, L.M., Dodson, R.J., Rosovitz, M.J., Sullivan, S.A., Daugherty, S.C., Haft, D.H., Selengut, J., Gwinn, M.L., Zhou, L., Zafar, N., Khouri, H., Radune, D., Dimitrov, G., Watkins, K., O'Connor, K.J.B., Smith, S., Utterback, T.R., White, O., Rubens, C.E., Grandi, G., Madoff, L.C., Kasper, D.L., Telford, J.L., Wessels, M.R., Rappuoli, R., Fraser, C.M., (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial “pan-genome.” *Proc Natl Acad Sci U S A* 102, 13950–13955. <https://doi.org/10.1073/pnas.0506758102>
- Van Rossum, G. & Drake, F.L., (2009). *Python 3 Reference Manual*, Scotts Valley, CA: CreateSpace.

- Ventura M, Canchaya C, Tauch A, Chandra G, Fitzgerald GF, Chater KF, van Sinderen F (2007) Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiol Mol Biol Rev* 71:495–548
- Verma, P., & Jha, A. (2019). Mycetoma: Reviewing a neglected disease. *Clinical and Experimental Dermatology*, 44(2), 123–129. <https://doi.org/10.1111/ced.13642>
- 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>
- Waksman SA, Henrici AT (1943) The nomenclature and classification of the actinomycetes. *J Bacteriol* 46:337–341.
- Watanabe, K., Hotta, K., Praseuth, A. P., Koketsu, K., Migita, A., Boddy, C. N., Wang, C. C. C., Oguri, H., & Oikawa, H. (2006). Total biosynthesis of antitumor nonribosomal peptides in *Escherichia coli*. *Nature Chemical Biology*, 2(8), 423–428. <https://doi.org/10.1038/nchembio803>
- Xu, L., Ye, K.-X., Dai, W.-H., Sun, C., Xu, L.-H., Han, B.-N., (2019). Comparative Genomic Insights into Secondary Metabolism Biosynthetic Gene Cluster Distributions of Marine *Streptomyces*. *Mar Drugs* 17, E498. <https://doi.org/10.3390/md17090498>
- Zhang, Z., Wang, Y. and Ruan, J. (1997) 'A proposal to revive the genus *Kitasatospora* (Omura, Takahashi, Iwai, and Tanaka 1982)', *International Journal of Systematic Bacteriology*, 47(4), pp. 1048–1054. Available at: <https://doi.org/10.1099/00207713-47-4-1048>.
- Zhou, Z., Gu, J., Li, Y.-Q., Wang, Y., (2012). Genome plasticity and systems evolution in *Streptomyces*. *BMC bioinformatics* 13 Suppl 10, S8. <https://doi.org/10.1186/1471-2105-13-S10-S8>