



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΤΜΗΜΑ ΒΙΟΧΗΜΕΙΑΣ ΚΑΙ ΒΙΟΤΕΧΝΟΛΟΓΙΑΣ

ΜΠΑΧΟΥΜΗΣ ΓΙΩΡΓΟΣ



ΛΑΡΙΣΑ 2019

BAC-TRECs: Υπολογιστικό εργαλείο για την εύρεση ανασυνδυασμών σε γονιδιώματα βακτηρίων

BAC-TRECs: A Computational tool that detects recombination events among bacterial genomes

Όνομα: Μπαχούμης Γιώργος

Επιβλέπων Καθηγητής: Γρηγόριος Αμούτζιας

Εργαστήριο Βιοπληροφορικής του τμήματος Βιοχημείας και Βιοτεχνολογίας του Πανεπιστημίου Θεσσαλίας

ΤΡΙΜΕΛΗΣ ΣΥΜΒΟΥΛΕΥΤΙΚΗ ΕΠΙΤΡΟΠΗ

Αμούτζιας Γρηγόριος(Επιβλέπων)

Αναπληρωτής Καθηγητής Βιοπληροφορικής με έμφαση στη Μικροβιολογία
Τμήμα Βιοχημείας & Βιοτεχνολογίας

Μαρκουλάτος Παναγιώτης

Καθηγητής Εφαρμοσμένης Μικροβιολογίας με έμφαση στη Βιοτεχνολογία
Τμήμα Βιοχημείας & Βιοτεχνολογίας

Μόσιαλος Δημήτριος

Επίκουρος Καθηγητής Βιοτεχνολογίας Μικροβίων
Τμήμα Βιοχημείας & Βιοτεχνολογίας

Αρχικά θα ήθελα να ευχαριστήσω τον κύριο Αμούτζια Γρηγόριο, Αναπληρωτή Καθηγητή Βιοπληροφορικής με έμφαση στη Μικροβιολογία, του τμήματος Βιοχημείας και Βιοτεχνολογίας του Πανεπιστημίου Θεσσαλίας, για την πολύτιμη βοήθεια και τη στήριξη που μου παρείχε καθ'όλη τη διάρκεια της διπλωματικής μου εργασίας. Επιπλέον θα ήθελα να ευχαριστήσω τα μέλη της τριμελούς επιτροπής κύριο Μαρκουλάτο Παναγιώτη, Καθηγητή Εφαρμοσμένης Μικροβιολογίας με έμφαση στη Βιοτεχνολογία του τμήματος Βιοχημείας και Βιοτεχνολογίας καθώς και τον κύριο Μόσιαλο Δημήτριο, Επίκουρο Καθηγητή Βιοτεχνολογίας Μικροβίων του τμήματος Βιοχημείας και Βιοτεχνολογίας. Τέλος θα ήθελα να ευχαριστήσω το Μιχάλη Τσιμπίδη απόφοιτο του τμήματος Βιοχημείας και Βιοτεχνολογίας και τον Μάριο Νικολαΐδη φοιτητή του τμήματος Βιοχημείας και Βιοτεχνολογίας για την πολύ μεγάλη βοήθεια που μου παρείχαν κατά τη διάρκεια της διπλωματικής μου εργασίας.

Πίνακας περιεχομένων

Contents

Contents	5
Περίληψη	6
1 Εισαγωγή.....	7
1.1 Ο γενετικός ανασυνδυασμός	7
1.2 Ανασυνδυασμός στα Βακτήρια	8
1.3 Οριζόντια μεταφορά γενετικού υλικού (HGT).....	11
1.4 Παν-γονιδίωμα (Pan-genome).....	12
1.5 Μέθοδοι για τον εντοπισμό ανασυνδυασμών	14
1.6 Εργαλείο BLAST (Basic Local Alingment Search Tool).....	16
1.7 Mega Blast.....	17
2 Υλικά και Μέθοδοι	18
2.1 BAC-TRECs.....	18
2.2 Python 2.7 - QT-Designer 4.....	18
2.3 Blastn – makeblastdb	18
2.4 Biopython.....	19
2.5 Plotly	19
2.6 Αξιολόγηση του BAC-TRecs.....	19
3 Αποτελέσματα.....	20
3.1 Λήψη και εγκατάσταση του BAC-TRecs	20
3.2 Ο Αλγόριθμος του BAC-Trecs	20
3.3 Το γραφικό περιβάλλον του BAC-TRECs	22
3.4 Ανάλυση γραφήματος των αποτελεσμάτων	25
4 Συζήτηση.....	31
Βιβλιογραφία.....	32

Περίληψη

Το υπολογιστικό εργαλείο BAC-TRECs, που αναπτύχθηκε στην παρούσα πτυχιακή εργασία είναι ένα πρόγραμμα με γραφικό περιβάλλον για λειτουργικά συστήματα Linux το οποίο χρησιμοποιεί στοίχιση κατά ζεύγη με συρόμενα παράθυρα για τον εντοπισμό γεγονότων ανασυνδυασμού μεταξύ βακτηρίων. Τα γεγονότα αυτά οπτικοποιούνται και παρουσιάζονται σε ένα διαδραστικό γράφημα καθώς επίσης εμφανίζεται και ο σχολιασμός (annotation) των υπό μελέτη γονιδιωμάτων. Τα αποτελέσματα από χιμαιρικές ακολουθίες που κατασκευάστηκαν (mock data) έδειξαν ότι το πρόγραμμα είναι πολύ αποτελεσματικό. Το συγκεκριμένο εργαλείο βασίστηκε στο T-RECs, ένα δημοσιευμένο υπολογιστικό εργαλείο που εντοπίζει γεγονότα ανασυνδυασμού μεταξύ διαφορετικών εξελικτικών ομάδων ιών.

Abstract

The computational tool BAC-TRECs, which was developed in this thesis is a program with graphical environment for Linux operating systems which uses pairwise alignment with sliding windows for the detection of recombination events among bacteria. These events are being visualized and presented to an interactive graph, where the annotation of the studied genomes also appears. The results from the analysis of the mock data that we generated showed that this program is very effective. This tool is based on T-RECs, a published computational tool that detects recombination events among different evolutionary groups of viruses.

1 Εισαγωγή

1.1 Ο γενετικός ανασυνδυασμός

Ο ανασυνδυασμός στο γενετικό υλικό ορίζεται ως η ανταλλαγή γενετικού υλικού και χωρίζεται σε 2 κατηγορίες.

- **Γενικός ή ομόλογος ανασυνδυασμός**

Είναι ο ανασυνδυασμός μεταξύ εκτεταμένων ομόλογων περιοχών DNA που μπορεί να συμβεί σε οποιαδήποτε θέση σε όλο το μήκος τους. Στους ευκαρυώτες συμβαίνει κατά τη μείωση, και συγκεκριμένα στη φάση της σύναψης, τόσο στα αρσενικά άτομα (κατά τη διάρκεια της σπερματογένεσης) όσο και στα θηλυκά άτομα (κατά την ωογένεση).

- **Ειδικός ανασυνδυασμός ή ανασυνδυασμός ειδικής θέσης**

Ένας διαφορετικός μηχανισμός ακολουθείται σε περιπτώσεις ανασυνδυασμού ανάμεσα σε συγκεκριμένες αλληλουχίες-στόχους που δεν είναι υποχρεωτικά ομόλογες. Αυτή η διαδικασία είναι υπεύθυνη για την ενσωμάτωση του γονιδιώματος ορισμένων φάγων στο βακτηριακό χρωμόσωμα. Στον ανασυνδυασμό αυτό εμπλέκονται συγκεκριμένες αλληλουχίες DNA του φάγου και του βακτηρίου, οι οποίες είναι ομόλογες σε μια μικρή περιοχή. Τα ένζυμα που εμπλέκονται σε αυτή τη διαδικασία δρουν μόνο στο συγκεκριμένο ζεύγος αλληλουχιών – στόχων, στο πλαίσιο μιας διαμοριακής αντίδρασης. Παρόμοιες ενδομοριακές αντιδράσεις ευθύνονται για τον επανασχηματισμό δύο ανεξάρτητων κυκλικών χρωμοσωμάτων από ένα διμερές που μπορεί να έχει σχηματιστεί μέσω γενικού ανασυνδυασμού κατά τη βακτηριακή διαίρεση. Επίσης σε αυτή τη κατηγορία εντάσσονται και κάποιες διαδικασίες, μέσω των οποίων αναστρέφονται συγκεκριμένες περιοχές του βακτηριακού χρωμοσώματος. (Krebs, Lewin, Kilpatrick, & Goldstein, 2014).

Οι διαδικασίες αυτές είναι πολύ σημαντικές για την εξέλιξη των οργανισμών. Αν δεν ήταν εφικτή η ανταλλαγή γενετικού υλικού μεταξύ (ομόλογων) χρωμοσωμάτων, τότε δεν θα μπορούσαν να σχηματιστούν χρωμοσώματα τα οποία θα έφεραν νέους συνδυασμούς αλληλομόρφων. Έτσι κάθε φορά που θα συνέβαιναν μεταλλάξεις, δε θα ήταν δυνατόν να διαχωριστούν οι ευνοϊκές από τις μη ευνοϊκές αλλαγές. Επίσης, οι αρνητικές συνέπειες μιας μετάλλαξης δε θα συνδεόταν μόνο με ένα συγκεκριμένο γονίδιο που μεταλλάχθηκε, αλλά με ολόκληρο το χρωμόσωμα στο οποίο αυτό εδράζεται. Επομένως ένα χρωμόσωμα θα συσώρευε εν τέλει τόσες επιβλαβείς μεταλλάξεις, ώστε θα έπαυε πια να λειτουργεί. Ανακατανέμοντας τα αλληλόμορφα, ο ανασυνδυασμός επιτρέπει σε ευνοϊκές και μη ευνοϊκές μεταλλάξεις να διαχωριστούν και να δοκιμαστούν ως διακριτές μονάδες σε νέους

συνδυασμούς. Παρέχει ένα μέσο διατήρησης και εξάπλωσης των ευνοϊκών αλληλομόρφων, καθώς και έναν τρόπο εξάλειψης κάθε μη ευνοϊκού αλληλομόρφου, χωρίς την παράλληλη εξάλειψη των υπόλοιπων αλληλομόρφων με τα οποία αυτό συνδέεται πάνω στο χρωμόσωμα. Αυτή είναι η βάση της φυσικής επιλογής (Krebs, Lewin, Kilpatrick, & Goldstein, 2014)

1.2 Ανασυνδυασμός στα Βακτήρια

Το φαινόμενο του ανασυνδυασμού στα βακτήρια αναγνωρίζεται ως ένας πολύ σημαντικός εξελικτικός μηχανισμός κατά τον οποίο τμήματα του DNA στον οργανισμό δέκτη αντικαθίστανται από ξένα τμήματα του οργανισμού δότη. Ο ανασυνδυασμός στη μορφή παραλαβής και ενσωμάτωσης γονιδίων και γενετικών στοιχείων από άλλα στελέχη και είδη βακτηρίων παίζει αρκετά σημαντικό ρόλο ως πηγή ποικιλομορφίας στη προσαρμοστική εξέλιξη πολλών ειδών βακτηρίων. Λόγω της ικανότητάς τους να προσλαμβάνουν γονίδια και γενετικά στοιχεία από άλλους οργανισμούς, η προσαρμοστικότητα των βακτηρίων είναι πολύ μεγάλη. Ο ανασυνδυασμός κυρίτερα λαμβάνει χώρα μεταξύ στενών συγγενικά βακτηριακών στελεχών και η συχνότητα του μειώνεται όσο μειώνεται η ομοιότητα μεταξύ τους. Καθώς η δυνατότητα ανασυνδυασμού δεν εξαφανίζεται τελείως σε πολύ απομακρυσμένα είδη, παρατηρείται οριζόντια μεταφορά γενετικού υλικού για την οποία γίνεται λόγος παρακάτω και μέσω αυτής της διαδικασίας τα βακτήρια αποκτούν χρωμοσωμικά γονίδια και ομάδες γονιδίων καθώς επίσης πλασμίδια, τραπεζοζόνια και γονίδια προφάγων. Έτσι το γονιδίωμα ενός βακτηρίου μπορεί να είναι ένα μωσαϊκό γενετικού υλικού από διαφορετικούς πρόγονους κάνοντας τον καθορισμό του είδους του περίπλοκο. Γεγονότα ανασυνδυασμού έχουν την ικανότητα να αλλάζουν το φαινότυπο των βακτηρίων προσδίδοντας τους την ικανότητα να δημιουργούν αποικίες σε καινούργια ενδιαιτήματα, να μεταβολίζουν νέες πηγές ενέργειας, να αποκτούν ανθεκτικότητα σε τοξικούς παράγοντες όπως τα αντιβιοτικά και τέλος να αυξάνουν τη παθογένεια τους στον άνθρωπο και σε άλλους ξενιστές. (Levin & Cornejo, 2009, Marttinen et al., 2012)

Η έκταση και η φύση του ανασυνδυασμού ποικίλει μεταξύ των ειδών των μικροοργανισμών αλλά και μεταξύ γενεαλογιών του ίδιου είδους. Είναι γνωστό ότι η έλλειψη γενετικής ανταλλαγής στα βακτήρια δεν αποτελεί συνηθισμένη κατάσταση πέρα από ορισμένα γενετικά μονομορφικά παθογόνα (Denamur & Matic, 2006). Τα βακτήρια επίσης εισάγουν γονίδια ή κομμάτια αυτών σε ένα ήδη υπάρχον ομόλογο γενετικό υλικό, μια διαδικασία που αναγνωρίστηκε από την παρατήρηση μωσαϊκών γονιδίων σε τόπους που κωδικοποιούν για αντιγόνα ή για ανθεκτικότητα σε αντιβιοτικά και συνήθως είναι αποτέλεσμα της δράσης της RecA (Feavers, Heath, Bygraves, & Maiden, 1992). Ο ομόλογος ανασυνδυασμός στα βακτήρια χωρίζεται σε αποτελεσματικός και μη αποτελεσματικός ανάλογα την επίδραση που θα έχει στο γονιδίωμα του οργανισμού δέκτη.

Σε περιπτώσεις ανασυνδυασμού που παρατηρείται ανταλλαγή ενός πανομοιότυπου κομματιού γενετικού υλικού δεν παρατηρείται το γεγονός παρ'όλο που στα συγγενικά στελέχη αυτός ο μη αποτελεσματικός ανασυνδυασμός αποτελεί μια πολύ σημαντική μέθοδο για την επιδιόρθωση του DNA και πιθανά την πιο σημαντική βιολογική λειτουργία του ομόλογου ανασυνδυασμού (Michod, Bernstein, & Nedelcu, 2008). Η πρόσφατη συσσώρευση των δεδομένων νουκλεοτιδικών αλληλουχιών από διαφορετικά είδη βακτηρίων διευκόλυε τις συγκριτικές μελέτες με αποτέλεσμα την καλύτερη κατανόηση όσον αφορά τη συχνότητα των ανασυνδυασμών, την επίδραση των ανασυνδυασμών στο γονιδίωμα καθώς και μοτίβα στην γονιδιακή ροή μεταξύ των βακτηριακών πλυθησμών (Didelot & Maiden, 2010).

- **Συχνότητα γεγονότων ανασυνδυασμού**

Υπάρχουν διάφορες προσεγγίσεις για τον υπολογισμό της συχνότητας των ανασυνδυασμών. Μια από αυτές είναι ο υπολογισμός της συχνότητας των ανασυνδυασμών σε σχέση με αυτήν των μεταλλάξεων που λαμβάνουν χώρα στο γονιδίωμα. Συγκριτικές μελέτες με τη μέθοδο MLST για την οποία γίνεται λόγος παρακάτω έδειξαν διαφορετικούς ρυθμούς ανασυνδυασμών στα υπό μελέτη βακτήρια. Για παράδειγμα, σε επτά διαφορετικά βακτήρια η ανάλυση έδειξε διαφορετικές συχνότητες ανασυνδυασμών με εύρος τιμών μεταξύ 0,05 για το *Bacillus cereus* και 7,21 για το *Streptococcus pyogenes* (Fraser, Hanage, & Spratt, 2005; Hanage, Fraser, & Spratt, 2006). Ένα θέμα που πρέπει να ληφθεί υπόψη είναι η ασυμφωνία που αφορά τα επίπεδα ανασυνδυασμού για κάποια είδη ακόμα και σε πρόσφατες μελέτες. Για παράδειγμα έχει βρεθεί ότι ο *B. cereus* ανασυνδυάζεται αρκετά σπάνια με ένα εύρος τιμών από 0,05 μέχρι 0,3 ενώ άλλη έρευνα έδειξε ότι ο ανασυνδυασμός συμβαίνει σχεδόν με διπλάσια συχνότητα (Pérez-Losada et al., 2006). Κάποιες από αυτές τις διαφορές μεταξύ των αποτελεσμάτων μπορεί να οφείλονται σε διαφορές μεταξύ των αναλυτικών εργαλείων καθώς και διαφορές στις στρατηγικές δειγματοληψίας. Κοντινά συγγενικά βακτήρια με διακριτά χαρακτηριστικά μπορούν να παρουσιάσουν μεγάλες διαφορές στη συχνότητα των γεγονότων ανασυνδυασμού. Αρκετά υψηλά εξειδικευμένα παθογόνα παρουσιάζουν λιγότερες περιπτώσεις ανασυνδυασμού απ'ότι τα όχι τόσο εξειδικευμένα συγγενικά με αυτά βακτήρια (Achtman, 2008), όπως στις περιπτώσεις των *Francisella tularensis* (Larsson et al., 2009) και *Bacillus anthracis* (Van Ert et al., 2007) όπου η συχνότητα των ανασυνδυασμών είχε μειωθεί λόγω της εξειδίκευσης των βακτηρίων. Παρατηρείται όμως και αύξηση της συχνότητας σε περιπτώσεις παθογένειας όπως στις παθολογικές γενεαλογίες της *E.coli* (Wirth et al., 2006). Επίσης αύξηση της παρατηρούμενης συχνότητας ανασυνδυασμού μπορεί να προκληθεί είτε από αύξηση στο καθεαυτό ρυθμό με τον οποίο συμβαίνει είτε με μείωση της αρνητικής επιλογής στα καινούργια αλληλόμορφα που δημιουργούνται (Didelot & Maiden, 2010). Το αποτέλεσμα της αλλαγής του τρόπου ζωής μπορεί να είναι είτε η αύξηση είτε η μείωση της παρατηρούμενης

συχνότητας ανασυνδυασμού. Επιπλέον η προσαρμοστική διαδικασία είναι σε θέση να αλλάξει της εξελικτικές πιέσεις που ασκούνται σε οποιαδήποτε γενιά, το οποίο θα οδηγήσει σε αύξηση ή μείωση της συχνότητας των ανασυνδυασμών εξαρτώμενες από τις συγκεκριμένες συνθήκες (Denamur & Matic, 2006) .

- **Επίδραση των ανασυνδυασμών στο γονιδίωμα**

Συνήθως δεν παρατηρούνται μεγάλου μεγέθους ανασυνδυασμοί στη φύση με κάποιες εξαιρέσεις που πιθανά προκλήθηκαν από συζευκτική μεταφορά και θετική επιλογή. Για παράδειγμα μεγάλα γεγονότα έχουν παρατηρηθεί σε ανθεκτικές γενεαλογίες στο *S. aureus* (Robinson & Enright, 2004) και στο *Streptococcus cerevisiae* (Brueggemann, Pai, Crook, & Beall, 2007). Η πλειονότητα των γεγονότων ανασυνδυασμού που παρατηρείται σήμερα κυμαίνεται μεταξύ λίγων εκατοντάδων ζευγών βάσεων έως και μερικών χιλιάδων όπως υπολογίστηκε χρησιμοποιώντας δεδομένα MLST από *N. meningitidis* (Jolley, Wilson, Kriz, McVean, & Maiden, 2005) και δεδομένα ολόκληρου γονιδιώματος από *S. enterica* (Didelot & Falush, 2007) . Τα γεγονότα ανασυνδυασμού στη φύση συνήθως έχουν πολύ μικρότερο μέγεθος από αυτά που παρατηρούνται *in vitro* όπως σημειώθηκε για παράδειγμα στα *C. jejuni* (Fearnhead, Smith, Barrigas, Fox, & French, 2005) και *E. coli* (Touchon et al., 2009) . Αυτή η διαφορά μεγέθους μπορεί να οφείλεται στο γεγονός ότι μεγάλα γεγονότα μπορεί να μειώνουν την αρμοστικότητα με αποτέλεσμα να εξαφανίζονται πριν προλάβουν να παρατηρηθούν στο γονιδίωμα (Jolley et al., 2005). Γονιδιωματικές περιοχές που κωδικοποιούν για πρωτεΐνες που παίζουν ρόλο στη παθογένεια βρίσκονται συχνά κάτω από θετική επιλογή (Petersen, Bollback, Dimmic, Hubisz, & Nielsen, 2007). Το γεγονός ότι ο ανασυνδυασμός παρατηρείται σε περιοχές θετικής επιλογής του βακτηριακού γονιδιώματος εξηγείται αν αναλογιστούμε ότι ο μόνος ανασυνδυασμός που μπορούμε να παρατηρήσουμε είναι αυτός που φέρει ευνοϊκές μεταλλάξεις και οδηγεί σε μια πολύ πιο μεγάλη αύξηση στην προσαρμοστικότητα απ'ότι θα συνέβαινε μόνο με μετάλλαξη (Vos, 2009).

- **Μοτίβα γονιδιακής ροής**

Αρκετοί παράγοντες επιτρέπουν στα γεγονότα ανασυνδυασμού να λαμβάνουν χώρα σε στενά συγγενικά βακτήρια πιο συχνά. Ένας παράγοντας είναι η φυσική εγγύτητα. Αυτή χρειάζεται για να μπορούν να πραγματοποιούνται διαδικασίες όπως ο μετασχηματισμός, η σύζευξη και η μεταγωγή για τις οποίες γίνεται λόγος παρακάτω. Η εγγύτητα αυτή συνήθως εμφανίζεται στα μέλη της ίδιας κοινότητας τα οποία πιθανά είναι και συγγενικά επειδή πολλά είδη βακτηρίων εμφανίζουν σημαντική οικολογική και γεωγραφική δομή. Συγκεκριμένα για το *C. jejuni* παρατηρήθηκε μια οικολογική δομή που εξαρτάται από το ξενιστή και έχει ως αποτέλεσμα την αύξηση της συχνότητας ανασυνδυασμών μεταξύ βακτηρίων διαφορετικών τύπων που μοιράζονται τον ίδιο ξενιστή

(Falush, 2009; McCarthy et al., 2007). Ένας άλλος παράγοντας που μειώνει τη συχνότητα γεγονότων ανασυνδυασμού μεταξύ μη συγγενικών βακτηρίων είναι ότι στα περισσότερα βακτηριακά είδη η πιθανότητα αποδοχής ενός γεγονότος ανασυνδυασμού μειώνεται εκθετικά με την αύξηση της γενετικής απόστασης του δότη και του δέκτη DNA (Majewski, 2001). Αναλύσεις γενετικής ποικιλότητας δείχνουν ότι γεγονότα ανασυνδυασμού συμβαίνουν πιο συχνά ανάμεσα σε στελέχη του ίδιου είδους σε σχέση με εκείνα που ανήκουν σε διαφορετικά είδη. Ακόμα και χωρίς κάποια σημαντική περιβαλλοντική αλλαγή, η εισαγωγή κάποιου μακρινού γενετικά υλικού θα μειώσει τη αρμοστικότητα του δέκτη με αποτέλεσμα την απομάκρυνση του μέσω αρνητικής επιλογής. Παρ' όλα αυτά έχει παρατηρηθεί υψηλή συχνότητα ανασυνδυασμών μεταξύ βακτηρίων διαφορετικών γενεαλογιών ακόμα και διαφορετικών ειδών. Για παράδειγμα στα *C.jejuni* και *C.coli* που ενώ απέχουν γενετικά περίπου 3,5% παρατηρείται υψηλή συχνότητα ανασυνδυασμών μεταξύ τους (Dingle, Colles, Falush, & Maiden, 2005; Wilson et al., 2009). Τέλος η εικόνα της γονιδιακής ροής στα περισσότερα είδη βακτηρίων αποτελείται από ανταλλαγές μεταξύ στενών συγγενικά βακτηρίων παρ' όλο που παρατηρούνται ανταλλαγές μεταξύ μελών διαφορετικών γενεών και ειδών που συχνά εξηγούνται από επιλεκτικές πιέσεις (Didelot & Maiden, 2010).

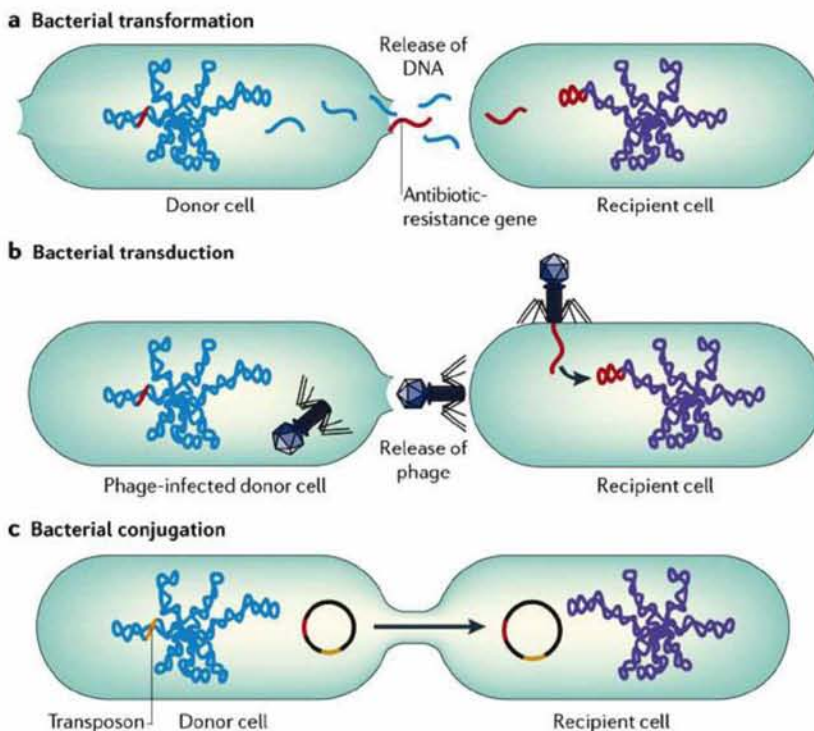
1.3 Οριζόντια μεταφορά γενετικού υλικού (HGT)

Η πρώτη περιγραφή αυτού του φαινομένου αποτέλεσε μεγάλη πρόοδο για τη μοριακή βιολογία. Αποδεικνύοντας το 1928 ότι μη παθογόνα βακτήρια πνευμονιόκοκκου μπορούσαν να μετατραπούν σε παθογόνα απλά με την επαφή με παθογόνα βακτήρια έδειξε ότι υπάρχει μια αρχή ικανή να επηρεάζει τη κληρονομικότητα (Griffith, 1928). Αυτή η αρχή αργότερα αναγνωρίστηκε ως το DNA. Αυτή η ανακάλυψη συνέβη χάρη στην ικανότητα του πνευμονιόκοκκου να προσλαμβάνει DNA οριζόντια.

Η οριζόντια μεταφορά γενετικού υλικού (Horizontal Gene Transfer - HGT) ορίζεται ως η κίνηση της γενετικής πληροφορίας ανάμεσα στους οργανισμούς (Burmeister, 2015). Το κυτταρόπλασμα του βακτηρίου στο οποίο μέσα εντοπίζεται το γονιδίωμα του είναι απομονωμένο από εξωτερικά μέσα μέσω ενός η περισσότερων μεμβρανών ανάλογα το είδος του βακτηρίου. Το DNA δεν μπορεί να διασχίσει παθητικά αυτά τα εμπόδια. Έτσι η HGT λαμβάνει χώρα με τρεις γνωστούς μηχανισμούς στη φύση όπως φαίνονται στην Εικόνα 1 : Οριζόντια μεταφορά γενετικού υλικού

- **Μετασχηματισμός-Transformation:** Είναι ένας ενεργός μηχανισμός κατά τον οποίο ελεύθερο DNA που υπάρχει στο θρεπτικό μέσο, κυρίως από νεκρούς οργανισμούς, εισέρχεται στο κυτταρόπλασμα. Αυτό μπορεί να συμβαίνει κυρίως για θρεπτικούς σκοπούς αλλά κάποια βακτήρια είναι πολύ επιλεκτικά στο είδος του DNA που επιτρέπουν στο να εισέλθει στο κύτταρο δηλώνοντας έτσι την ύπαρξη γεγονότων ανασυνδυασμού μεταξύ στενών συγγενών (Daubin & Szöllösi, 2016)

- **Σύζευξη-Conjugation:** Τα βακτήρια μεταφέρουν απευθείας γονίδια σε άλλα κύτταρα. Τα γονίδια που είναι υπεύθυνα για τη σύζευξη υπάρχουν μέσα σε πλασμίδια ή φάγους που ονομάζονται συζευκτικά, χρησιμοποιώντας τη σύζευξη για να καταφέρουν την μεταφορά του γενετικού υλικού (Daubin & Szöllösi, 2016)
- **Μεταγωγή-Transduction :** Βακτηριοφάγοι μεταφέρουν γονίδια από ένα κύτταρο σε ένα άλλο. Όταν ένας φάγος μεταφέρει τα γονίδια μέσα σε ένα βακτηριακό κύτταρο εκείνα ενσωματώνονται στο γονιδίωμα με τη διαδικασία του ανασυνδυασμού (Griffiths et al., 2000).



Εικόνα 1 : Οριζόντια μεταφορά γενετικού υλικού (Πηγή: Furuya and Lowy Nature, 2006.)

1.4 Παν-γονιδίωμα (Pan-genome)

Το Παν-γονιδίωμα είναι μια έννοια στο χώρο της γονιδιωματικής και της βιοπληροφορικής που περιλαμβάνει τον συνολικό αριθμό γονιδίων που υπάρχουν σε όλα τα γνωστά στελέχη ενός είδους ή ακόμα και σε όλα τα στελέχη ενός γένους. Βασίζεται στη σύγκριση γονιδιωμάτων διαφορετικών στελεχών του ίδιου είδους ή του ίδιου γένους. Με την Αλληλούχιση Νέας Γενιάς (Next Generation Sequencing) το κόστος και ο χρόνος μιας

αλληλούχισης ενός γονιδιώματος μειώθηκε δραματικά. Αυτό είχε ως αποτέλεσμα την αύξηση των αλληλουχιών ολόκληρων γονιδιωμάτων στις δημόσιες βάσεις δεδομένων όπως του NCBI. Σε αυτό το πλαίσιο η συγκεκριμένη αύξηση οδήγησε στην ανάγκη για το σχεδιασμό εργαλείων συγκριτικής γονιδιωματικής για την ανάλυση τόσων πολλών δεδομένων. Αυτό είχε ως αποτέλεσμα την ανάπτυξη του αντικειμένου της Παν-γονιδιωματικής (Guimarães et al., 2015).

Η μελέτη δεδομένων με την Παν-γονιδιωματική προσέγγιση παρέχει πληροφορίες για την εξέλιξη των βακτηρίων και την προσαρμογή τους καθώς και σε εφαρμοσμένες τεχνικές όπως στην παρασκευή εμβολίων και φαρμάκων. Το Παν-γονιδίωμα αποτελείται από 3 μέρη όπως φαίνονται και στην εικόνα 2.

1. Core Genome

Το core-genome είναι το υποσύνολο των γονιδίων τα οποία υπάρχουν σε όλα τα γονιδιώματα της υπό μελέτη ομάδας και καθορίζεται συγκρίνοντας τα γονιδιώματα αυτά. Είναι γνωστό πλέον ότι παραπάνω από 250 οικογένειες γονιδίων έχουν χαρακτηριστεί ως κομμάτι του core-genome και βάσει αυτών των οικογενειών αποδεικνύεται η συντηρημένη φύση των γονιδίων. Τα γονίδια που υπάρχουν στο core – genome εμπλέκονται σε βασικούς μηχανισμούς της βιολογίας των οργανισμών (Guimarães et al., 2015).

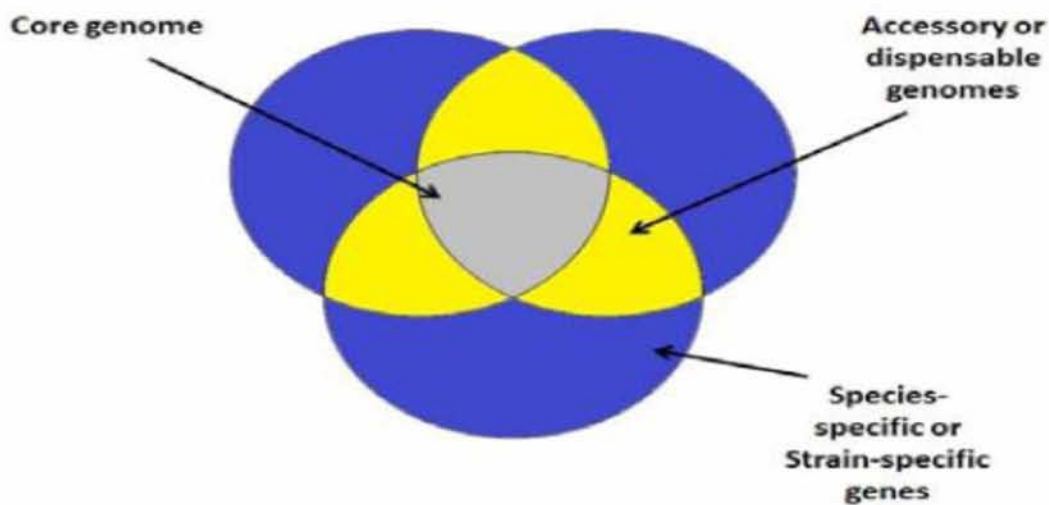
2. Accessory Genome

Το accessory genome είναι το υποσύνολο των γονιδίων που υπάρχει σε κάποιους οργανισμούς της υπό μελέτη ομάδας, αλλά δεν υπάρχει σε όλους τους οργανισμούς. Περίπου 8000 οικογένειες γονιδίων αποτελούν το accessory genome . Το υποσύνολο αυτό περιλαμβάνει γονίδια απαραίτητα για την επιβίωση των οργανισμών σε συγκεκριμένες περιπτώσεις/συνθήκες και συνήθως σχετίζονται με τη τοξικότητα ή την ανθεκτικότητα σε αντιβιοτικά. Το accessory-genome προέκυψε πιθανόν από την οριζόντια μεταφορά των γονιδίων και την παραφυλετική εξέλιξη τους λόγω διπλασιασμού και μετάλλαξης (Guimarães et al., 2015).

3. Species-Specific Genome ή Strain-Specific Genome

Στο species- specific genome εντοπίζονται γονίδια τα οποία υπάρχουν σε ένα μόνο είδος από αυτά που μελετούνται. Στο strain-specific genome εντοπίζονται γονίδια που υπάρχουν σε ένα μόνο από τα υπό μελέτη στελέχη ενός είδους. Συνήθως, αυτά τα γονίδια προέρχονται από την οριζόντια μεταφορά μεταξύ διαφορετικών ειδών. Σύμφωνα με την παν-γονιδιωματική μελέτη του γένους *Streptococcus*, τα ειδο-ειδικά γονίδια (species-specific genes) οργανώνονται σε 139.000 οικογένειες γονιδίων. Η ύπαρξη αυτών των γονιδίων μπορεί να επιβεβαιώσει το προσαρμοστικό πλεονέκτημα σε σχέση με εκείνα που τη στερούνται. Επιπλέον μελέτες έδειξαν ότι αυτά τα γονίδια σχετίζονται με την τοξικότητα ή την παθογένεια σε παθογόνους μικροοργανισμούς (Jordan, Makarova, Sprouge, Wolf, &

Koonin, 2001). Γενικά αυτή η ομάδα γονιδίων βρίσκεται κάτω από χαλαρή πίεση μεταλλάξεων με μεταλλάξεις να συμβαίνουν συνέχεια στις ακολουθίες των γονιδίων σε αντίθεση με τα γονίδια του core-genome τα οποία βρίσκονται σε συνεχόμενη επιλογή για να διατηρήσουν τις συντηρημένες αλληλουχίες τους (Daubin & Ochman, 2004). Κάποιες μεταλλάξεις αυξάνουν τη βακτηριακή προσαρμογή σε συγκεκριμένα περιβάλλοντα και συνθήκες, έτσι τα γονίδια αυτά διατηρούνται στο γονιδίωμα. Διαφορετικά οι μεταλλάξεις μπορούν να οδηγήσουν στη δημιουργία ψευδογονιδίων (γονίδια χωρίς λειτουργία) που κατά τη διάρκεια της εξελικτικής διαδικασίας εξαιρούνται από το γονιδίωμα.



Εικόνα 2 Παν- Γονιδίωμα Adapted: (Πηγή: Muzzi et al., 2007.)

1.5 Μέθοδοι για τον εντοπισμό ανασυνδυασμών

Τα γεγονότα ανασυνδυασμού μαζί με τις σημειακές μεταλλάξεις αποτελούν μια από τις κινητήριες δυνάμεις της εξέλιξης των βακτηρίων. Με την αλληλούχιση νέας γενιάς (NGS) έχει αυξηθεί ο όγκος γονιδιωμάτων που έχουν αλληλουχηθεί με αποτέλεσμα την ανάγκη σχεδιασμού νέων υπολογιστικών εργαλείων για την ανάλυση των ανασυνδυασμών. Τα εργαλεία αυτά βασίζονται στις εξής αναλυτικές προσεγγίσεις :

1. Σύγκριση ακολουθιών κατά ζεύγη.
2. Φυλογενετικές μέθοδοι
3. Πληθυσμιακή γενετική

4. Μοτίβα περιοχών

Συγκεκριμένα για τα βακτήρια δημιουργήθηκε μια μέθοδος για την ανάλυση των γονιδιωμάτων τους με τη χρήση γονιδίων του core – genome (housekeeping genes) ως γονίδια αναφοράς στα υπο μελέτη γονιδιώματα. Η μέθοδος αυτή ονομάζεται MLST (Multi Locus Sequence Typing). Σε περιπτώσεις ανασυνδυασμού παρατηρείται ασυμφωνία φυλογενετικών δέντρων. Η μέθοδος αυτή έχει αρκετούς περιορισμούς όσον αφορά τον εντοπισμό των ανασυνδυασμών λόγω του μικρού αριθμού των γονιδίων αναφοράς που μπορεί να μην είναι αντιπροσωπευτικά για ολόκληρο το βακτηριακό γονιδίωμα καθώς επίσης και ότι εντοπίζει μόνο πολύ πρόσφατα γεγονότα (Maiden, 2006).

Ο μεγάλος όγκος πληροφορίας που υπάρχει πλέον στις βάσεις δεδομένων έχει οδηγήσει στην ανάγκη ανάπτυξης υπολογιστικών εργαλείων για την ανάλυση αυτών των δεδομένων. Παρακάτω παρουσιάζονται ορισμένα δημοφιλή λογισμικά που εντοπίζουν ανασυνδυασμούς σε βακτηριακά γονιδιώματα.

- **BratNextGen**

Το λογισμικό αυτό χρησιμοποιεί μια μπεϋζιανή προσέγγιση ομαδοποίησης για τον εντοπισμό γενετικά διαφορετικών ομάδων που εκπροσωπούν διαφορετικές εξελικτικές γενεαολογιές. Το υπο-μελέτη γονιδίωμα διαχωρίζεται σε τμήματα των 5kb και πραγματοποιείται ανάλυση ομαδοποίησης ξεχωριστά για κάθε τμήμα. Τα αποτελέσματα παρουσιάζονται σε μορφή γενεαλογικού δέντρου όπου φαίνονται τα τμήματα που έχουν ομαδοποιηθεί μαζί και παρουσιάζουν κοινά γεγονότα ανασυνδυασμού (Marttinen et al., 2012)

- **SimBac**

Το συγκεκριμένο λογισμικό προσομοιώνει την εξέλιξη των βακτηρίων που εξετάζονται προς τα πίσω χρονικά εφαρμόζοντας ένα συμφυή μοντέλο με μετατροπή γονιδίων (gene conversion (Wiuf & Hein, 2000)) για τη προσομοίωση της εξέλιξης βακτηριακών γονιδιωμάτων. Τέλος παρουσιάζει ένα γράφημα με τους πιθανούς ανασυνδυασμούς που έχουν λάβει χώρα κατά την εξέλιξη (Brown & Wilson, n.d.).

- **Gubbins (Genealogies Unbiased By recomBinations In Nucleotide Sequences)**

Το λογισμικό αυτό εφαρμόζει στατιστικές μεθόδους για τον εντοπισμό περιοχών με πολυμορφισμούς σε βακτηριακά γονιδιώματα που υποδηλώνουν οριζόντια μεταφορά μεταξύ τους. Στη συνέχεια κατασκευάζει φυλογενετικά δέντρα με τη μέθοδο της μέγιστης πιθανοφάνειας (maximum likelihood) χρησιμοποιώντας

το υπολογιστικό εργαλείο RaxML-Light (Stamatakis et al., 2012). Έτσι εντοπίζονται τα πιθανά γεγονότα ανασυνδυασμού (Croucher et al., 2015).

Τα εργαλεία αυτά παρόλες τις δυνατότητες που προσφέρουν δεν είναι ικανά να αναλύσουν εκατοντάδες η και χιλιάδες ολόκληρα γονιδιώματα βακτηρίων με ένα απλό βήμα και σε ένα λογικό χρονικό διάστημα. Τέλος, αυτά τα λογισμικά δεν προσφέρουν ένα φιλικό γραφικό περιβάλλον προς το χρήστη. Η αλληλούχιση νέας γενιάς καθιστά αναγκαία πλέον τη δημιουργία υπολογιστικών εργαλείων που να καλύπτουν τις παραπάνω ιδιότητες καθώς ο όγκος της πληροφορίας αυξάνεται καθημερινά μαζί με την ανάγκη για την ανάλυσή της.

1.6 Εργαλείο BLAST (Basic Local Alingment Search Tool)

Το εργαλείο που αναπτύξαμε για γρήγορη ανάλυση μεγάλου αριθμού βακτηριακών γονιδιωμάτων βασίστηκε στο εξαιρετικά δημοφιλές βιοπληροφορικό εργαλείο BLAST.

Για την εύρεση ομολόγων ακολουθιών από μια βάση δεδομένων χρησιμοποιούνται αλγόριθμοι τοπικής στοίχισης κατά ζεύγη και ο πιο διαδεδομένος είναι ο αλγόριθμος BLAST. Εντοπίζει τις ομόλογες ακολουθίες αρκετά αποτελεσματικά και πολύ γρήγορα σε σχέση με άλλους αλγόριθμους. Καθώς επίσης μπορεί να εντοπίσει ομοιότητες και διαφορές μεταξύ των ομόλογων ακολουθιών.

Για την πραγματοποίηση του BLAST πρέπει να οριστούν ορισμένες παράμετροι. Η πιο σημαντική παράμετρος είναι το e-value το οποίο είναι μια τιμή που ορίζει τη πιθανότητα μιας ακολουθίας στα αποτελέσματα να εμφανιστεί ως ομόλογη καθαρά από τύχη.

Το BLAST παρέχει προγράμματα τα οποία ειδικεύονται για στοιχίσεις ακολουθιών διαφορετικής φύσης όπως νουκλεοτιδικές, πρωτεϊνικές καθώς επίσης και νουκλεοτιδικές μεταφρασμένες σε πρωτεϊνικές. Τα προγράμματα φαίνονται παρακάτω στον Πίνακα 1

Πρόγραμμα	Βάση δεδομένων	Ακολουθία επερώτησης
BLASTN	Νουκλεοτιδική	Νουκλεοτιδική
BLASTP	Πρωτεϊνική	Πρωτεϊνική
BLASTX	Πρωτεϊνική	Νουκλεοτιδική μεταφρασμένη στα 6 πιθανά ORFs

TBLASTN	Νουκλεοτιδική μεταφρασμένη πιθανά ORFs	στα 6	Πρωτεϊνική
TBLASTX	Νουκλεοτιδική μεταφρασμένη πιθανά ORFs	στα 6	Νουκλεοτιδική μεταφρασμένη πιθανά ORFs

Πίνακας 1 Προγράμματα του BLAST

1.7 Mega Blast

Το megablast ανήκει στο πακέτο του blastall και χρησιμοποιεί έναν αλγόριθμο ο οποίος δίνει τη δυνατότητα πραγματοποίησης του Blast χρησιμοποιώντας πολλές ακολουθίες επερώτησης (multiple queries) και εξειδικεύεται σε ανάλυση μεγάλων γονιδιωμάτων όπως στη συγκεκριμένη περίπτωση των βακτηριακών γονιδιωμάτων. Ο αλγόριθμος αυτός εξετάζει τη βάση δεδομένων για παρόμοιες αλληλουχίες πολύ γρήγορα και αποτελεσματικά (McGinnis & Madden, 2004).

2 Υλικά και Μέθοδοι

2.1 BAC-TRECS

Το εργαλείο που κατασκευάστηκε και παρουσιάζεται στη συγκεκριμένη εργασία ονομάζεται Bac-TRECS (Bacterial-Tool for Recombinations) και βασίζεται στην ευρετική μέθοδο τοπικής στοίχισης κατά ζεύγη του Blast με τη χρήση «συρόμενων παραθύρων». Επειδή το συγκεκριμένο εργαλείο σχεδιάστηκε για τον εντοπισμό ανασυνδυασμών σε βακτήρια χρησιμοποιείται το Megablast το οποίο αναλύει γονιδιώματα μεγάλου μεγέθους με μεγαλύτερη ταχύτητα. Το Bac-TRECS στη συνέχεια φιλτράρει τα αποτελέσματα του Megablast με παραμέτρους που ρυθμίζει ο χρήστης όπως το μέγεθος του HSP (High-scoring Segment Pair). Στη συνέχεια τα αποτελέσματα αυτά παρουσιάζονται σε γράφημα διαδραστικό σε μορφή html στο οποίο ο χρήστης μπορεί να μεγενθύνει την περιοχή που τον ενδιαφέρει, να απομονώσει συγκεκριμένες ακολουθίες καθώς επίσης αναγράφονται επάνω οι σχολιασμοί (annotation) των υπό μελέτη ακολουθιών. Ακόμα δίνεται η επιλογή αποθήκευσης του γραφήματος σε εικόνα της μορφής png. Τέλος το πρόγραμμα παρέχει ένα φιλικό περιβάλλον προς τον χρήστη και είναι διαθέσιμο για συστήματα Linux.

2.2 Python 2.7 - QT-Designer 4

Για την ανάπτυξη του προγράμματος χρησιμοποιήθηκε η γλώσσα προγραμματισμού Python 2.7. Η κατασκευή του γραφικού περιβάλλοντος έγινε με το πρόγραμμα QT-Designer 4. Αυτό καθιστά απαραίτητη την εγκατάσταση της βιβλιοθήκης PyQt 4.

2.3 Blastn – makeblastdb

Για τη λειτουργία του Bac-TRECS είναι απαραίτητο το Blastn το οποίο περιέχει το Megablast με το οποίο γίνεται η λήψη της ακολουθίας επερώτησης και η σύγκρισή της με τις ακολουθίες που βρίσκονται στη βάση δεδομένων που ορίζει ο χρήστης. Με το Blastn εγκαθίσταται αυτόματα και το πρόγραμμα makeblastdb το οποίο μετατρέπει τη προκαθορισμένη βάση δεδομένων σε αξιοποιήσιμη μορφή από το Blastn.

Τα 2 αυτά προγράμματα περιλαμβάνονται στη βιβλιοθήκη προγραμμάτων του BLAST+ τα οποία μπορούν να ληφθούν μέσω του παρακάτω συνδέσμου: https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastDocs&DOC_TYPE=Download

2.4 Biopython

Για τη μορφοποίηση της ακολουθίας επερώτησης και το φιλτράρισμα των αποτελεσμάτων από το Megablast χρησιμοποιήθηκε η βιβλιοθήκη Biopython. Η Biopython έχει δημιουργηθεί από μια διεθνή ομάδα προγραμματιστών που χρησιμοποιούν τη γλώσσα προγραμματισμού Python για την ανάπτυξη υπολογιστικών εργαλείων που χρησιμοποιούνται στο τομέα της βιοπληροφορικής, τα οποία διατίθενται δωρεάν.

2.5 Plotly

Το plotly είναι μια βιβλιοθήκη της γλώσσας προγραμματισμού Python για τη δημιουργία γραφημάτων. Χρησιμοποιήθηκε η συγκεκριμένη βιβλιοθήκη διότι δίνει τη δυνατότητα κατασκευής ενός διαδραστικού γραφήματος κατά το οποίο ο χρήστης έχει τη δυνατότητα να μεγενθύνει τις περιοχές που επιθυμεί, να δει ακριβώς τη θέση τους σε σχέση με την ακολουθία επερώτησης, να απομονώσει συγκεκριμένα αποτελέσματα και τέλος να αποθηκεύσει το γράφημα ως εικόνα με τη μορφή που εκείνος επιθυμεί.

2.6 Αξιολόγηση του BAC-TRecs

Τα δεδομένα που χρησιμοποιήθηκαν για την αξιολόγηση του προγράμματος είναι τα εξής. Έγινε λήψη 13 βακτηριακών γονιδιωμάτων εκ των οποίων τα 5 ήταν του γένους *Escherichia*, τα 4 του γένους *Salmonella* και τα 4 του γένους *Pseudomonas* από τη Genbank. Αυτά αποτελούσαν τη βάση δεδομένων που καθορίσαμε. Ως ακολουθία επερώτησης χρησιμοποιήθηκε ένα στέλεχος του είδους *Escherichia coli*. Κάνοντας την ανάλυση παρατηρήσαμε αρκετά γεγονότα πιθανού ανασυνδυασμού μεταξύ των διαφορετικών στελεχών της *E.coli*. Στη συνέχεια κατασκευάσαμε μια χιμαιρική ακολουθία (mock data) παίρνοντας ένα μεγάλο τμήμα της ακολουθίας επερώτησης και τοποθετώντας το σε ένα στέλεχος του γένους *Salmonella*. Ξανακάνοντας την ανάλυση το τμήμα αυτό που περιμέναμε εμφανίστηκε ως αποτέλεσμα στο γράφημα

3 Αποτελέσματα

3.1 Λήψη και εγκατάσταση του BAC-TRecs

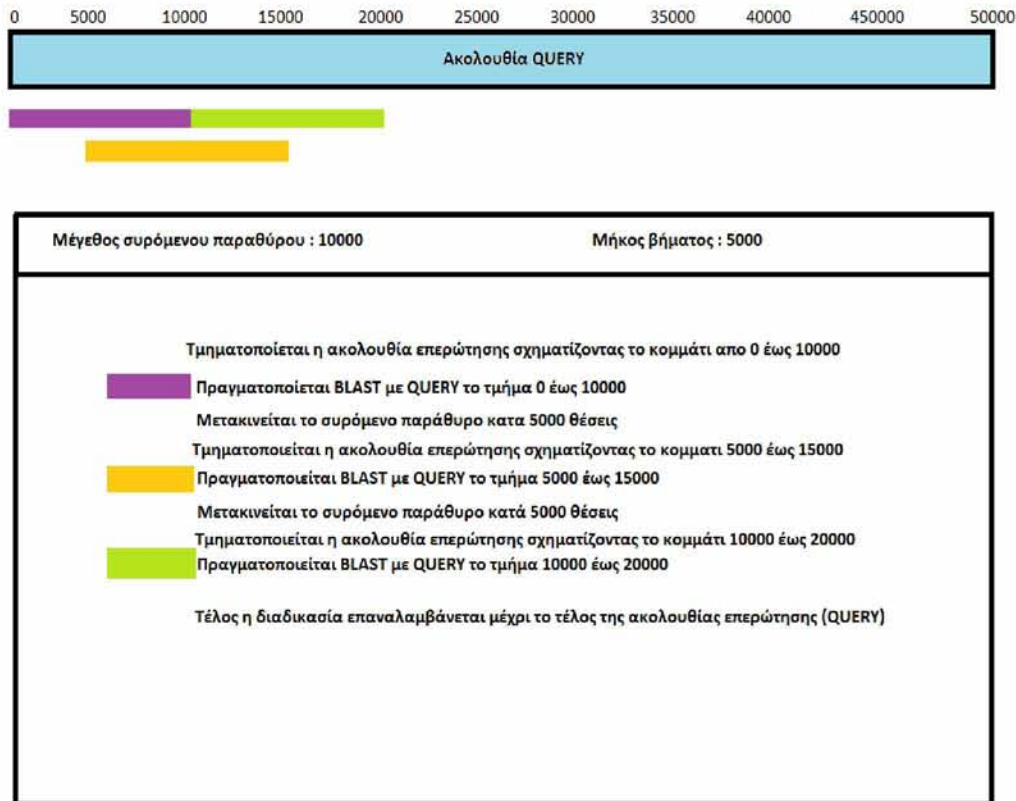
1. Απαραίτητο για τη λειτουργία του προγράμματος είναι η εγκατάσταση του plotly η οποία μπορεί να γίνει γράφοντας στο τερματικό του Linux το εξής : `sudo pip install plotly`. Περισσότερες πληροφορίες μπορείτε να βρείτε εδώ <https://plot.ly/python/getting-started/>
2. Σε όλα τα λειτουργικά συστήματα Linux υπάρχει το πακέτο της γλώσσας προγραμματισμού Python. Σε αυτό το πακέτο υπάρχει ένα πρόγραμμα διαχείρισης πακέτων το pip το οποίο είναι απαραίτητο για την εγκατάσταση του plotly. Σε περίπτωση που αυτό το πρόγραμμα δεν υπάρχει μπορείτε να το βρείτε εδώ <https://pip.pypa.io/en/latest/installing/>
3. Για τη λειτουργία του Bac-TRecs χρειάζεται εγκατάσταση του λογισμικού Blast+ ώστε να μπορεί να πραγματοποιείται η ανάλυση του Blast τοπικά στον υπολογιστή του χρήστη. Το λογισμικό αυτό μπορείτε να το βρείτε σε αυτό το σύνδεσμο : <ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/>
4. Τέλος απαιτείται η εγκατάσταση της βιβλιοθήκης PyQt 4. Για τη πραγματοποίηση της εγκατάστασης πληκτρολογούμε στο τερματικό το εξής : `sudo apt-get install python-qt4`

3.2 Ο Αλγόριθμος του BAC-Trecs

Είναι βασικό να αναφερθούμε στο μηχανισμό με τον οποίο το πρόγραμμα αναζητά και εντοπίζει πιθανά γεγονότα ανασυνδυασμού πριν την ανάλυση των επιλογών του. Αρχικά το Bac-TRecs μπορεί να πραγματοποιεί Blast τοπικά στον υπολογιστή εφόσον ο χρήστης του παρέχει την ακολουθία επερώτησης, τη βάση δεδομένων και ένα αρχείο που θα ορίζονται οι ομάδες των οργανισμών που θα αναλυθούν. Οι μορφές των αρχείων θα περιγραφούν αναλυτικότερα παρακάτω. Η βασική λειτουργία του BAC-TRecs είναι ο οπτικός έλεγχος για ανασυνδυασμούς μεταξύ της ακολουθίας επερώτησης και μιας ακολουθίας από τη βάση δεδομένων που του παρέχεται. Αυτό γίνεται με τον εξής τρόπο:

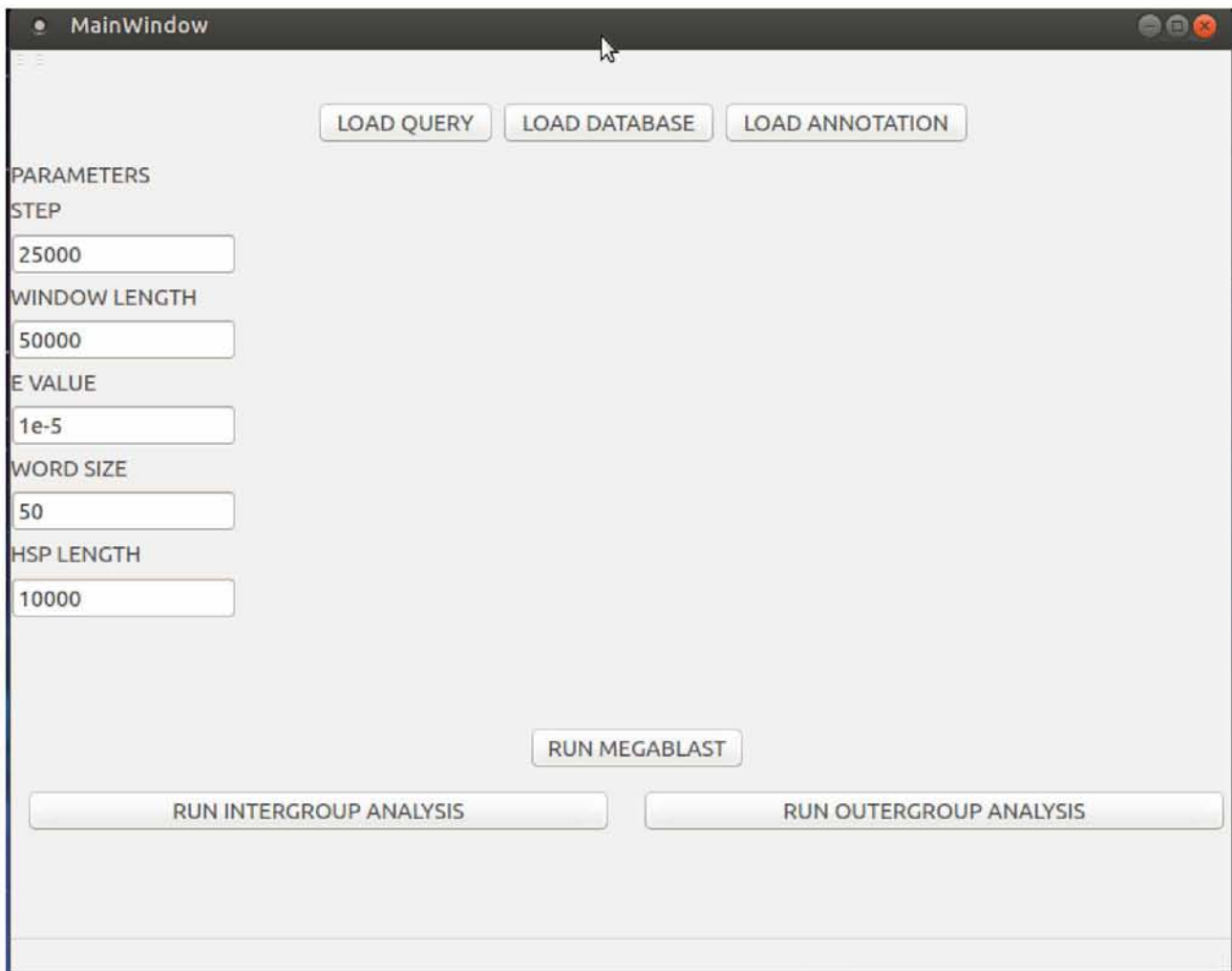
Η ακολουθία επερώτησης “κόβεται” από το πρόγραμμα σε μικρότερα τμήματα των οποίων το μέγεθος ορίζεται από το χρήστη και στη συνέχεια το κάθε τμήμα λειτουργεί ως ακολουθία επερώτησης για το Blast που πραγματοποιείται με τις ακολουθίες στη βάση

δεδομένων. Τα τμήματα χωρίζονται με τη χρήση ενός συρόμενου παραθύρου που επίσης ορίζεται από το χρήστη. Η βασική αυτή λειτουργία παρουσιάζεται σχηματικά στην



Εικόνα 3 Μηχανισμός λειτουργίας του BAC-T-RECS

3.3 Το γραφικό περιβάλλον του BAC-TREcs



Εικόνα 4: Γραφικό περιβάλλον του BAC-TREcs

Η φόρμα του BAC-TREcs όπως φαίνεται και στην Εικόνα 4: Γραφικό περιβάλλον του BAC-TREcs αποτελείται από τα εξής :

- **Load Query**

Η διεργασία που εκτελείται με αυτό το κουμπί είναι η φόρτωση (upload) της ακολουθίας επερώτησης σε μορφή Fasta.

- **Load Database**

Εδώ φορτώνεται το αρχείο που αποτελεί τη βάση δεδομένων που επιθυμεί να εξετάσει ο χρήστης σε μορφή Fasta. Περιέχει όλες τις ακολουθίες των οργανισμών που θα αναλυθούν και είναι της μορφής:

>Accession Number

Γονιδίωμα

>Accession Number

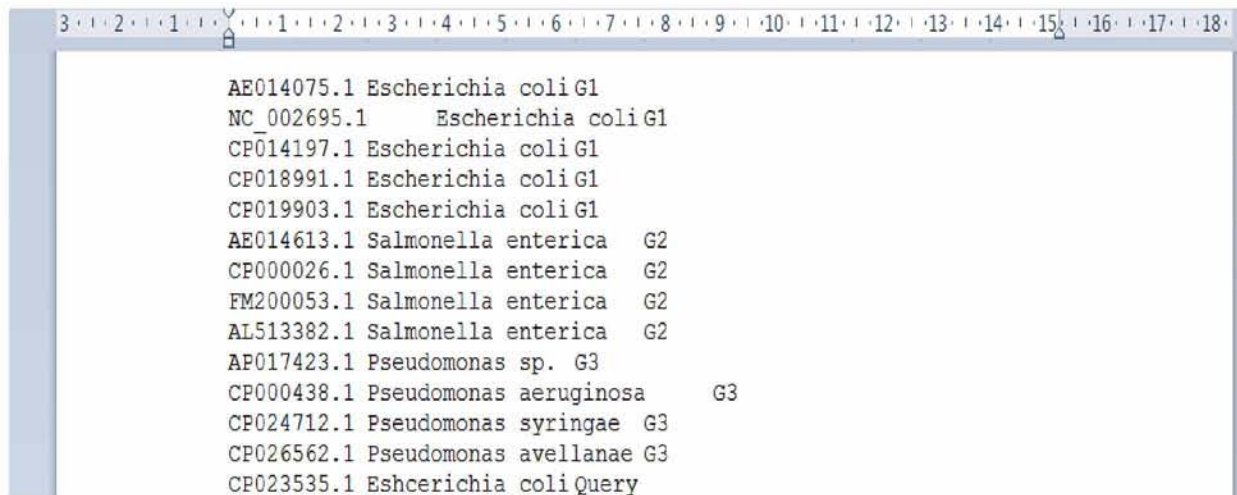
Γονιδίωμα

- **Load Annotation**

Εδώ ο χρήστης φορτώνει (upload) το αρχείο με τις ομάδες των οργανισμών που έχει δημιουργήσει και έχει τη μορφή :

Accession_number Είδος Ομάδα

Ένα παράδειγμα ενός τέτοιου αρχείου φαίνεται στην Εικόνα 5 Αρχείο Annotation



```
3 2 1 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
AE014075.1 Escherichia coli G1
NC_002695.1 Escherichia coli G1
CP014197.1 Escherichia coli G1
CP018991.1 Escherichia coli G1
CP019903.1 Escherichia coli G1
AE014613.1 Salmonella enterica G2
CP000026.1 Salmonella enterica G2
FM200053.1 Salmonella enterica G2
AL513382.1 Salmonella enterica G2
AP017423.1 Pseudomonas sp. G3
CP000438.1 Pseudomonas aeruginosa G3
CP024712.1 Pseudomonas syringae G3
CP026562.1 Pseudomonas avellanae G3
CP023535.1 Escherichia coli Query
```

Εικόνα 5 Αρχείο Annotation

Είναι βασικό κάθε στήλη να χωρίζεται με \TAB από την επόμενη για να λειτουργήσει το πρόγραμμα.

Επίσης στο περιβάλλον παρατηρείται ένα πλαίσιο με παραμέτρους (Parameters) οι οποίες έχουν κάποιες ενδεικτικές τιμές με βάση την ανάλυση που έγινε. Οι παράμετροι είναι οι εξής.

- **Window_Length**

Είναι το μέγεθος του παραθύρου και κατά συνέπεια το μέγεθος των τμημάτων που θα τμηματοποιηθεί η ακολουθία επερώτησης.

- **Step**

Αποτελεί το βήμα με το οποίο ορίζεται το συρόμενο παράθυρο στη μορφοποίηση της ακολουθίας επερώτησης.

- **E-Value**

το e-value είναι μια τιμή που ορίζει τη πιθανότητα μιας ακολουθίας στα αποτελέσματα να εμφανιστεί ομόλογη καθαρά από τύχη.

- **Word-Size**

Αποτελεί το μέγεθος της αρχικής ακολουθίας που θα πρέπει να ταιριάζει απόλυτα ανάμεσα στην ακολουθία επερώτησης και στη συγκρινόμενη ακολουθία της βάσης δεδομένων κατά τη διαδικασία του Blast.

- **Hsp length**

Είναι το μέγεθος των περιοχών που παρουσιάζουν τη μεγαλύτερη ομοιότητα στα αποτελέσματα του Blast.

Τέλος παρουσιάζονται τρεις ακόμα λειτουργίες:

1. Run Megablast

Εφόσον έχει γίνει εισαγωγή στο πρόγραμμα των αρχείων που προορίζονται για ανάλυση με τη συγκεκριμένη λειτουργία θα πραγματοποιηθεί το Megablast για τις ακολουθίες αυτές και τα αποτελέσματα του θα χρησιμοποιηθούν για το φιλτράρισμα και την παρουσίαση τους στο γράφημα στο τέλος της ανάλυσης.

2. Run Intergroup_analysis

Με τη συγκεκριμένη λειτουργία το πρόγραμμα φιλτράρει τα αποτελέσματα του megablast και τα παρουσιάζει σε γράφημα στο οποίο φαίνεται η ακολουθία επερώτησης και τα τμήματα που έχουν προκύψει από το megablast των ακολουθιών στη βάση δεδομένων στην ακριβή θέση τους σε σχέση με την ακολουθία επερώτησης. Ακόμα εμφανίζεται ο σχολιασμός των ακολουθιών (annotation). Ο χρωματισμός των ακολουθιών στη συγκεκριμένη ανάλυση έχει γίνει με βάση το στέλεχος του οργανισμού. Έτσι παρατηρούνται πιθανά συμβάντα ανασυνδυασμού μεταξύ των διαφορετικών στελεχών τόσο της ίδιας ομάδας όσο και μεταξύ διαφορετικών ομάδων.

3. Run Outergroup_analysis

Εδώ η λειτουργία είναι ακριβώς η ίδια με παραπάνω με τη διαφορά ότι στο γράφημα που προκύπτει ο χρωματισμός γίνεται με βάση την ομάδα που έχει ορίσει ο χρήστης.

Όποιος οργανισμός είναι στην ίδια ομάδα χρωματίζεται το ίδιο και έτσι ξεχωρίζουν τα τμήματα που έχουμε ως αποτελέσματα από άλλες ομάδες.

3.4 Ανάλυση γραφήματος των αποτελεσμάτων

Οι λειτουργίες `Intergroup_analysis` και `Outergroup_analysis` παρουσιάζουν τα αποτελέσματα της ανάλυσης με διαφορετικό χρωματισμό ανάλογα με την προσέγγιση που επιθυμεί ο χρήστης. Παρακάτω παρουσιάζονται τα γραφήματα που προκύπτουν από τις συγκεκριμένες λειτουργίες καθώς και οι επιλογές που έχει ο χρήστης στον χειρισμό του γραφήματος για καλύτερη οπτικοποίηση των αποτελεσμάτων.

Γράφημα `Intergroup_analysis`:

Στο συγκεκριμένο γράφημα παρουσιάζεται η ακολουθία επερώτησης και τα hits που έχουν προκύψει από τα αποτελέσματα του BLAST τα οποία έχουν φιλτραριστεί με την παραμέτρο `HSP_length` που όπως αναφέρθηκε είναι οι περιοχές με τη μεγαλύτερη ομοιότητα και το μέγεθος τους ορίζεται από το χρήστη. Επίσης παρατηρούνται με διαφορετικό χρωματισμό όλα τα στελέχη που εμφανίζονται στα αποτελέσματα στην ακριβή θέση που εντοπίζονται πάνω στην ακολουθία επερώτησης (QUERY). Ακόμα εμφανίζεται και ο σχολιασμός (annotation) δεξιά του γραφήματος με το accession number του κάθε οργανισμού καθώς επίσης το είδος και το στέλεχος. Έτσι διακρίνονται όλα τα πιθανά γεγονότα εντός και εκτός των ομάδων που έχουν οριστεί. Το γράφημα φαίνεται στην Εικόνα 6.



Εικόνα 6 Γράφημα *Intergroup_analysis*

Γράφημα *Outergroup_analysis*:

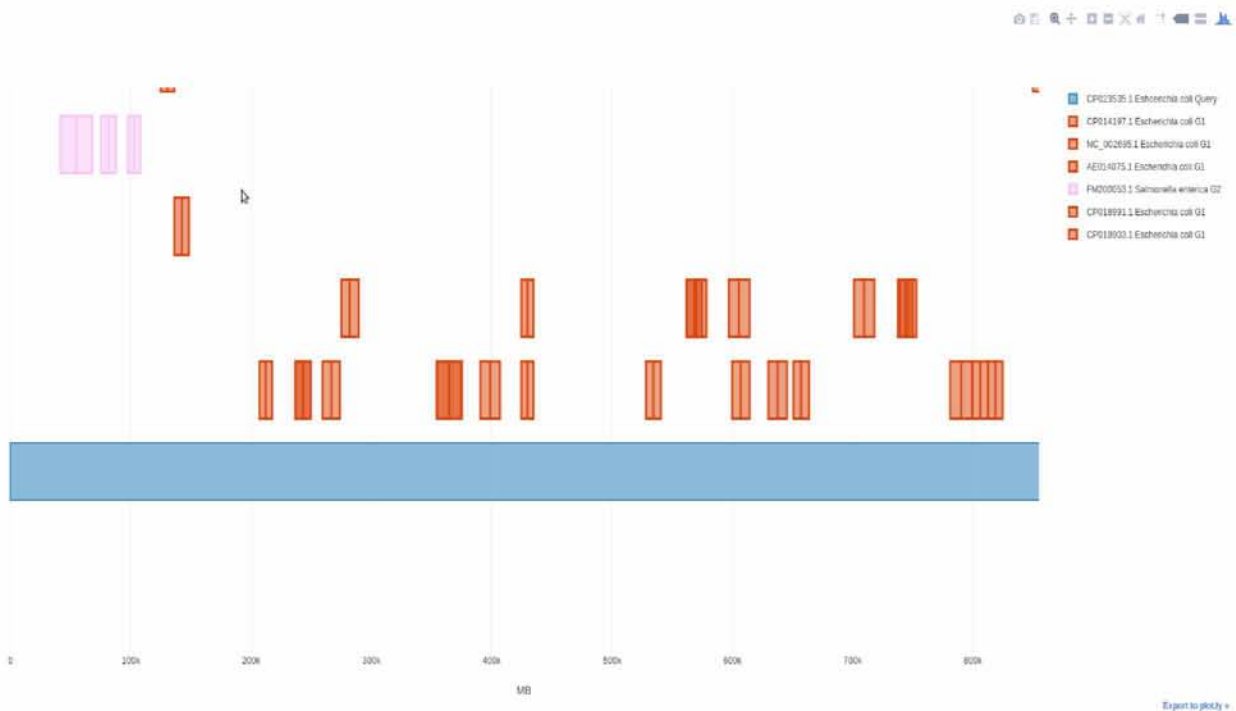
Στο συγκεκριμένο γράφημα εμφανίζονται τα ίδια στοιχεία με παραπάνω με τη διαφορά ότι χρωματίζονται όλα τα στελέχη μιας ομάδας με το ίδιο χρώμα και με διαφορετικό εκείνα από άλλες ομάδες. Έτσι διακρίνεται οποιοδήποτε hit του BLAST έχει δημιουργηθεί από διαφορετική ομάδα της ακολουθίας επερώτησης. Το γράφημα φαίνεται στην Εικόνα 7 Γράφημα *Outergroup_analysis*



Εικόνα 7 Γράφημα *Outergroup_analysis*

Λειτουργία μεγένθυσης περιοχής ενδιαφέροντος :

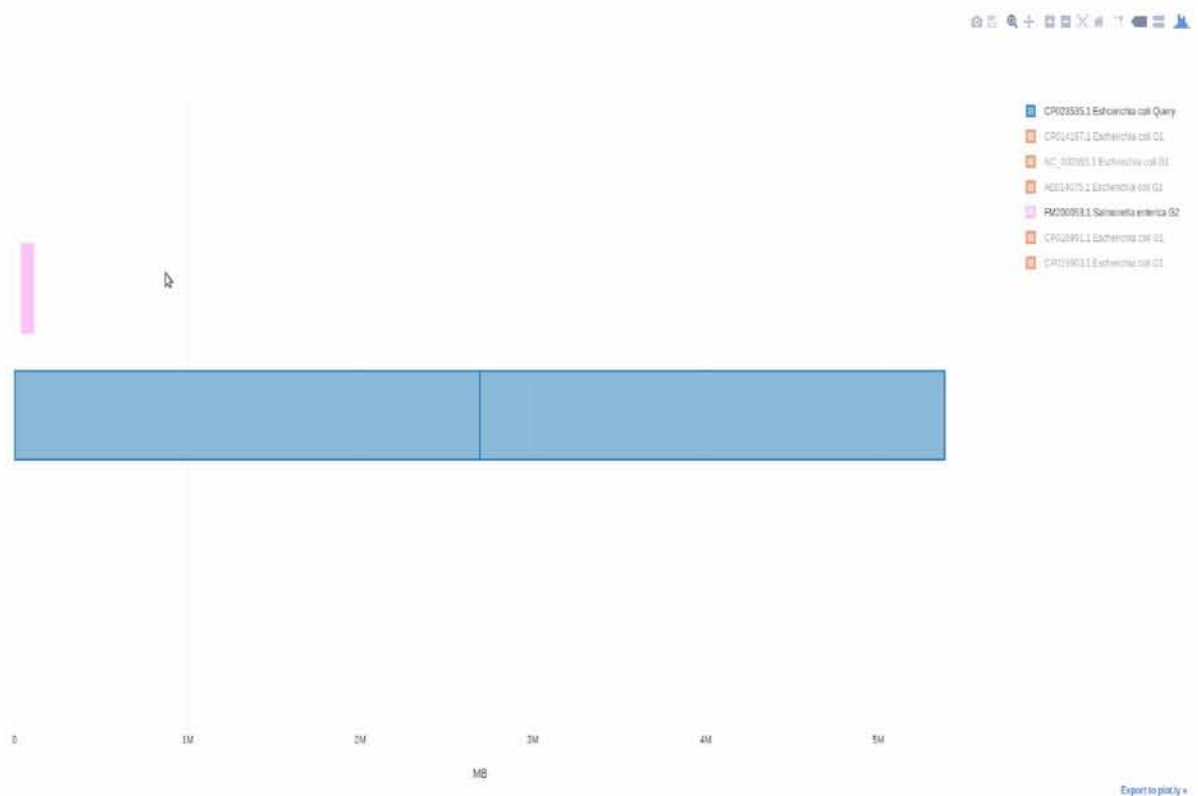
Η συγκεκριμένη λειτουργία δίνει τη δυνατότητα στο χρήστη να μεγενθύνει το γράφημα στη περιοχή που τον ενδιαφέρει. Αυτό του επιτρέπει να δει λεπτομερώς το γεγονός καθώς και τα όρια της περιοχής που παρουσιάζεται ανασυνδυασμένη. Έτσι κρατώντας πατημένο το αριστερό κλικ απο το ποντίκι επιλέγει τη περιοχή και εκείνη μεγενθύνεται όπως παρουσιάζεται στην Εικόνα 8 Μεγένθυση περιοχής ενδιαφέροντος



Εικόνα 8 Μεγένθυση περιοχής ενδιαφέροντος

Λειτουργία απομόνωσης αλληλουχιών :

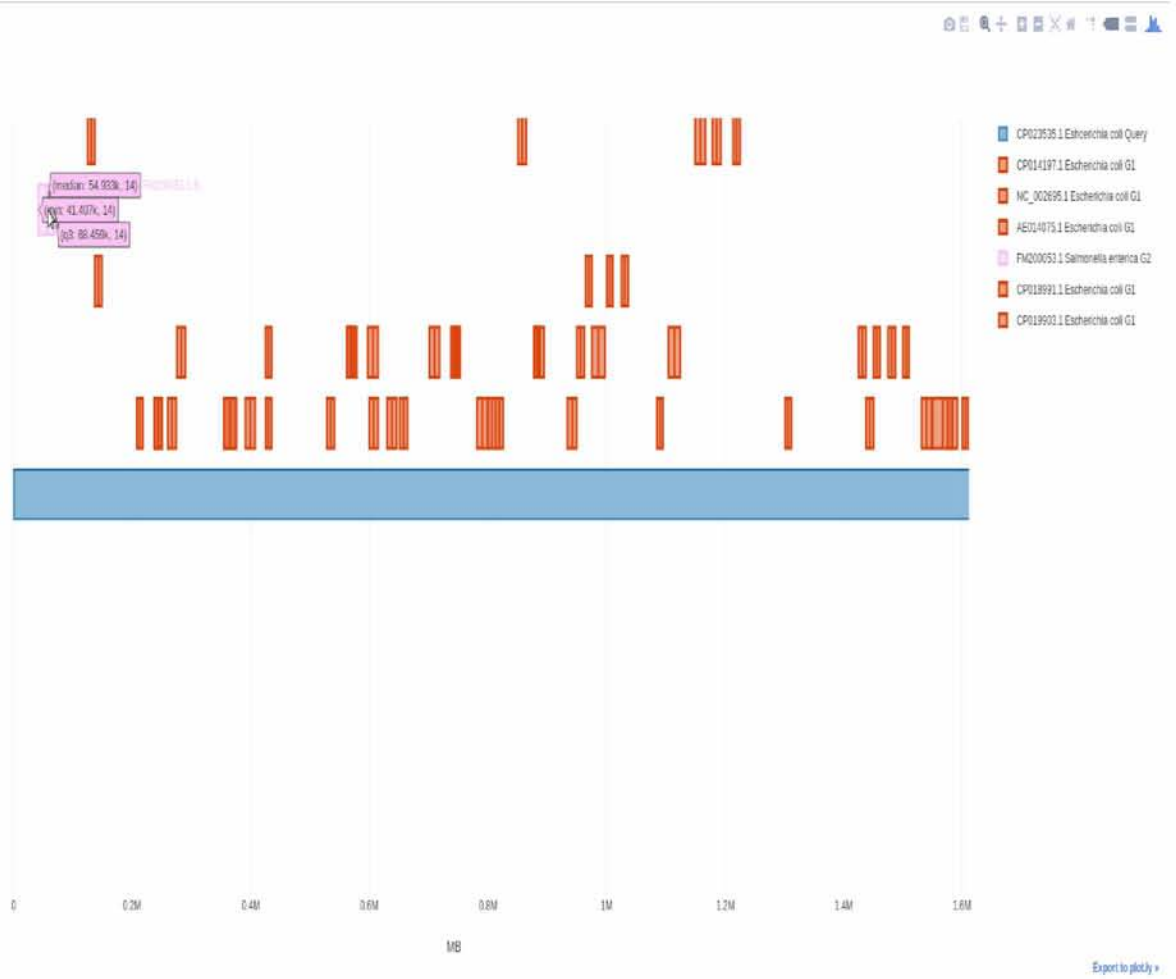
Η λειτουργία αυτή επιτρέπει στο χρήστη να απομονώσει την αλληλουχία ή τις αλληλουχίες που τον ενδιαφέρουν από όλα τα αποτελέσματα του γραφήματος. Αυτό του δίνει τη δυνατότητα να παρατηρήσει πιο εύκολα τις περιοχές που εμφανίζονται ως πιθανά γεγονότα ανασυνδυασμού καθώς απλοποιείται το γράφημα αφαιρώντας αποτελέσματα. Η επιλογή γίνεται αφαιρώντας αποτελέσματα κάνοντας κλικ στα τετράγωνα που εμφανίζονται δίπλα από τα στοιχεία κάθε οργανισμού. Η λειτουργία αυτή φαίνεται στην Εικόνα 9 Απομόνωση επιθυμητού αποτελέσματος



Εικόνα 9 Απομόνωση επιθυμητού αποτελέσματος

Λειτουργία εμφάνισης ορίων και σχολιασμού :

Τέλος ο χρήστης έχοντας το ποντίκι σε ένα αποτέλεσμα μπορεί να δει τα όρια του συγκεκριμένου hit καθώς και το σχολιασμό του αποτελέσματος. Έτσι εμφανίζεται ακριβώς η θέση στην οποία παρατηρείται ανταλλαγή γενετικού υλικού μεταξύ των οργανισμών της βάσης δεδομένων με την ακολουθία επερώτησης. Η λειτουργία αυτή παρουσιάζεται στην Εικόνα 10 Εμφάνιση ορίων και σχολιασμού



Εικόνα 10 Εμφάνιση ορίων και σχολιασμού

4 Συζήτηση

Τα τελευταία χρόνια η ανάγκη για ανάλυση των γονιδιωμάτων που βρίσκονται στις βάσεις δεδομένων έχει αυξηθεί με αποτέλεσμα την ανάπτυξη εργαλείων βιοπληροφορικής που έχουν την ικανότητα να πραγματοποιήσουν μια τέτοια ανάλυση. Ο εντοπισμός γεγονότων ανασυνδυασμού στους οργανισμούς και συγκεκριμένα στα βακτήρια είναι πολύ σημαντικός διότι παρέχει πληροφορίες για την εξέλιξη των ειδών αλλά και τις φαινοτυπικές διαφορές που παρουσιάζουν στη φύση. Τα βακτήρια έχουν την ικανότητα μεταφοράς γενετικού υλικού και ανταλλαγής του με τη διαδικασία της οριζόντιας μεταφοράς γενετικού υλικού. Η διαδικασία αυτή γίνεται με 3 τρόπους. Το μετασχηματισμό, τη σύζευξη και τη μεταγωγή. Με τη χρήση της προσέγγισης του παν-γονιδιώματος η ανάλυση των γεγονότων αυτών στα βακτήρια δίνει τη δυνατότητα γνώσης της εξελικτικής διαδικασίας των βακτηρίων εις βάθος. Κατά τη προσέγγιση αυτή τα γονίδια ομαδοποιούνται σε 3 κατηγορίες. Το core-genome, το accessory-genome και το species-specific-genome. Στο core-genome τα γονίδια βρίσκονται σε όλα τα γονιδιώματα των υπο-μελέτη βακτηρίων. Στο accessory-genome παρατηρούνται σε αρκετούς οργανισμούς αλλά όχι σε όλους. Τέλος στο species-specific-genome παρατηρούνται γονίδια που υπάρχουν μόνο στο ίδιο είδος. Η γνώση αυτή παρέχει στην επιστημονική κοινότητα καλύτερη κατανόηση της παθογένειας των βακτηρίων καθώς και της ανοσίας που παρουσιάζουν στα διάφορα αντιβιοτικά. Εργαλεία που έχουν αναπτυχθεί για τον εντοπισμό ανασυνδυασμών στηρίζονται σε διάφορες μεθόδους όπως τις μεθόδους αποστάσεων, τις φυλογενετικές μεθόδους κ.α. Στις μεθόδους αποστάσεων παρατηρείται τοπική στοίχιση κατά ζεύγη με τη χρήση συρόμενου παράθυρου, μία μέθοδος που δίνει την ικανότητα ανάλυσης γονιδιωμάτων γρήγορα και αξιόπιστα. Έχουν αναπτυχθεί διάφορα υπολογιστικά εργαλεία για τον εντοπισμό ανασυνδυασμών. Γνωστά για εντοπισμό ανασυνδυασμών σε βακτηριακά γονιδιώματα είναι το BratNextGen, SimBac και το Gubbins. Παρ' όλες τις δυνατότητες που προσφέρουν τα εργαλεία αυτά δεν μπορούν να διαχειριστούν τον τόσο μεγάλο όγκο πληροφορίας που υπάρχει στις βάσεις δεδομένων με ένα απλό βήμα και ακόμα δεν προσφέρουν ένα γραφικό περιβάλλον χειρισμού τους φιλικό προς το χρήστη. Το Bac-T-Recs είναι ένα εργαλείο που δημιουργήθηκε με τη γλώσσα προγραμματισμού Python 2.7 και η λειτουργία του βασίζεται στο BLAST, στο φιλτράρισμα των αποτελεσμάτων του Blast με τη χρήση της βιβλιοθήκης Biopython καθώς και στη βιβλιοθήκη Plotly για την οπτικοποίηση των αποτελεσμάτων. Βασίζεται στην στοίχιση ακολουθιών κατά ζεύγη του BLASTN και συγκεκριμένα του MEGABLAST το οποίο χρησιμοποιείται για μεγαλύτερα σε μέγεθος γονιδιώματα όπως αυτά των βακτηρίων με τη χρήση συρόμενων παραθύρων στην ακολουθία επερώτησης έναντι μιας προκαθορισμένης βάσης δεδομένων που δίνεται από το χρήστη. Το BAC-T-Recs συνιστά ένα εργαλείο εξειδικευμένο για βακτηριακά

γονιδιώματα το οποίο εντοπίζει γεγονότα ανασυνδυασμού και τα οπτικοποιεί σε ένα διαδραστικό γράφημα στο οποίο φαίνεται η θέση του γεγονότος σε σύγκριση με την ακολουθία επερώτησης που μπορεί να αναλυθεί περαιτέρω και πιο εξειδικευμένα με άλλα προγράμματα. Ακόμα το Bac-TRecs προσφέρει ένα φιλικό γραφικό περιβάλλον στο χρήστη. Η επαλήθευση της σωστής λειτουργίας του BAC-T-Recs έγινε με τη χρήση ακολουθιών κατασκευασμένων από πρίν (mock data) στις οποίες τοποθετήθηκαν τμήματα από ένα είδος σε ένα άλλο και εντοπίστηκαν από το πρόγραμμα. Το BAC-T-Recs έχει την ικανότητα να αναλύσει μεγάλα μεγέθους γονιδιώματα όπως είναι αυτά των βακτηρίων καθώς και να οπτικοποιήσει τα αποτελέσματα με τέτοιο τρόπο που να είναι κατανοητά από το χρήστη. Τέλος η δημιουργία τέτοιων εργαλείων πλέον είναι πιο αναγκαία από ποτέ με τον όγκο των δεδομένων που παράγεται από την αλληλούχιση νέας γενιάς (Next Generation Sequencing)

Βιβλιογραφία

- Achtman, M. (2008). Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annual Review of Microbiology*, 62, 53–70. <https://doi.org/10.1146/annurev.micro.62.081307.162832>
- Brown, T., & Wilson, D. D. (n.d.). *SimBac: A Stochastic Simulator of Bacterial Evolution with Homologous Recombination*. 11.
- Burmeister, A. R. (2015). Horizontal Gene Transfer. *Evolution, Medicine, and Public Health*, 2015(1), 193–194. <https://doi.org/10.1093/emph/eov018>
- Croucher, N. J., Page, A. J., Connor, T. R., Delaney, A. J., Keane, J. A., Bentley, S. D., ... Harris, S. R. (2015). Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Research*, 43(3), e15. <https://doi.org/10.1093/nar/gku1196>
- Daubin, V., & Ochman, H. (2004). Bacterial genomes as new gene homes: The genealogy of ORFans in *E. coli*. *Genome Research*, 14(6), 1036–1042. <https://doi.org/10.1101/gr.2231904>

- Daubin, V., & Szöllösi, G. J. (2016). Horizontal Gene Transfer and the History of Life. *Cold Spring Harbor Perspectives in Biology*, 8(4), a018036. <https://doi.org/10.1101/cshperspect.a018036>
- Denamur, E., & Matic, I. (2006). Evolution of mutation rates in bacteria. *Molecular Microbiology*, 60(4), 820–827. <https://doi.org/10.1111/j.1365-2958.2006.05150.x>
- Didelot, X., & Maiden, M. C. J. (2010). Impact of recombination on bacterial evolution. *Trends in Microbiology*, 18(7), 315–322. <https://doi.org/10.1016/j.tim.2010.04.002>
- Feavers, I. M., Heath, A. B., Bygraves, J. A., & Maiden, M. C. (1992). Role of horizontal genetic exchange in the antigenic variation of the class 1 outer membrane protein of *Neisseria meningitidis*. *Molecular Microbiology*, 6(4), 489–495. <https://doi.org/10.1111/j.1365-2958.1992.tb01493.x>
- Fraser, C., Hanage, W. P., & Spratt, B. G. (2005). Neutral microepidemic evolution of bacterial pathogens. *Proceedings of the National Academy of Sciences of the United States of America*, 102(6), 1968–1973. <https://doi.org/10.1073/pnas.0406993102>
- Griffith, Fred. (1928). The Significance of Pneumococcal Types. *The Journal of Hygiene*, 27(2), 113–159.
- Griffiths, A. J., Miller, J. H., Suzuki, D. T., Lewontin, R. C., Gelbart, W. M., Griffiths, A. J., ... Gelbart, W. M. (2000). *An Introduction to Genetic Analysis* (7th ed.). W. H. Freeman.
- Guimarães, L. C., Florczak-Wyspianska, J., de Jesus, L. B., Viana, M. V. C., Silva, A., Ramos, R. T. J., ... Soares, S. de C. (2015). Inside the Pan-genome—Methods and Software Overview. *Current Genomics*, 16(4), 245–252. <https://doi.org/10.2174/1389202916666150423002311>
- Hanage, W. P., Fraser, C., & Spratt, B. G. (2006). The impact of homologous recombination on the generation of diversity in bacteria. *Journal of Theoretical Biology*, 239(2), 210–219. <https://doi.org/10.1016/j.jtbi.2005.08.035>
- Jordan, I. K., Makarova, K. S., Spouge, J. L., Wolf, Y. I., & Koonin, E. V. (2001). Lineage-specific gene expansions in bacterial and archaeal genomes. *Genome Research*, 11(4), 555–565. <https://doi.org/10.1101/gr.gr-1660r>

- Krebs, J. E., Lewin, B., Kilpatrick, S. T., & Goldstein, E. S. (2014). *Lewin's genes XI* (11th ed). Burlington, Mass: Jones & Bartlett Learning.
- Larsson, P., Elfsmark, D., Svensson, K., Wikström, P., Forsman, M., Brettin, T., ... Johansson, A. (2009). Molecular evolutionary consequences of niche restriction in *Francisella tularensis*, a facultative intracellular pathogen. *PLoS Pathogens*, *5*(6), e1000472.
<https://doi.org/10.1371/journal.ppat.1000472>
- Levin, B. R., & Cornejo, O. E. (2009). The population and evolutionary dynamics of homologous gene recombination in bacterial populations. *PLoS Genetics*, *5*(8), e1000601.
<https://doi.org/10.1371/journal.pgen.1000601>
- Maiden, M. C. J. (2006). Multilocus sequence typing of bacteria. *Annual Review of Microbiology*, *60*, 561–588. <https://doi.org/10.1146/annurev.micro.59.030804.121325>
- Majewski, J. (2001). Sexual isolation in bacteria. *FEMS Microbiology Letters*, *199*(2), 161–169.
<https://doi.org/10.1111/j.1574-6968.2001.tb10668.x>
- Marttinen, P., Hanage, W. P., Croucher, N. J., Connor, T. R., Harris, S. R., Bentley, S. D., & Corander, J. (2012). Detection of recombination events in bacterial genomes from large population samples. *Nucleic Acids Research*, *40*(1), e6. <https://doi.org/10.1093/nar/gkr928>
- McGinnis, S., & Madden, T. L. (2004). BLAST: At the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research*, *32*(Web Server issue), W20-25.
<https://doi.org/10.1093/nar/gkh435>
- Michod, R. E., Bernstein, H., & Nedelcu, A. M. (2008). Adaptive value of sex in microbial pathogens. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *8*(3), 267–285.
<https://doi.org/10.1016/j.meegid.2008.01.002>
- Petersen, L., Bollback, J. P., Dimmic, M., Hubisz, M., & Nielsen, R. (2007). Genes under positive selection in *Escherichia coli*. *Genome Research*, *17*(9), 1336–1343.
<https://doi.org/10.1101/gr.6254707>

- Stamatakis, A., Aberer, A. J., Goll, C., Smith, S. A., Berger, S. A., & Izquierdo-Carrasco, F. (2012). RAxML-Light: A tool for computing terabyte phylogenies. *Bioinformatics*, *28*(15), 2064–2066. <https://doi.org/10.1093/bioinformatics/bts309>
- Van Ert, M. N., Easterday, W. R., Huynh, L. Y., Okinaka, R. T., Hugh-Jones, M. E., Ravel, J., ... Keim, P. (2007). Global genetic population structure of *Bacillus anthracis*. *PLoS One*, *2*(5), e461. <https://doi.org/10.1371/journal.pone.0000461>
- Vos, M. (2009). Why do bacteria engage in homologous recombination? *Trends in Microbiology*, *17*(6), 226–232. <https://doi.org/10.1016/j.tim.2009.03.001>
- Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L. H., ... Achtman, M. (2006). Sex and virulence in *Escherichia coli*: An evolutionary perspective. *Molecular Microbiology*, *60*(5), 1136–1151. <https://doi.org/10.1111/j.1365-2958.2006.05172.x>
- Wiuf, C., & Hein, J. (2000). The coalescent with gene conversion. *Genetics*, *155*(1), 451–462.
- Brueggemann, A. B., Pai, R., Crook, D. W., & Beall, B. (2007). Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. *PLoS Pathogens*, *3*(11), e168. <https://doi.org/10.1371/journal.ppat.0030168>
- Didelot, X., & Falush, D. (2007). Inference of bacterial microevolution using multilocus sequence data. *Genetics*, *175*(3), 1251–1266. <https://doi.org/10.1534/genetics.106.063305>
- Dingle, K. E., Colles, F. M., Falush, D., & Maiden, M. C. J. (2005). Sequence typing and comparison of population biology of *Campylobacter coli* and *Campylobacter jejuni*. *Journal of Clinical Microbiology*, *43*(1), 340–347. <https://doi.org/10.1128/JCM.43.1.340-347.2005>
- Falush, D. (2009). Toward the use of genomics to study microevolutionary change in bacteria. *PLoS Genetics*, *5*(10), e1000627. <https://doi.org/10.1371/journal.pgen.1000627>
- Fearnhead, P., Smith, N. G. C., Barrigas, M., Fox, A., & French, N. (2005). Analysis of recombination in *Campylobacter jejuni* from MLST population data. *Journal of Molecular Evolution*, *61*(3), 333–340. <https://doi.org/10.1007/s00239-004-0316-0>

- Jolley, K. A., Wilson, D. J., Kriz, P., McVean, G., & Maiden, M. C. J. (2005). The influence of mutation, recombination, population history, and selection on patterns of genetic diversity in *Neisseria meningitidis*. *Molecular Biology and Evolution*, 22(3), 562–569.
<https://doi.org/10.1093/molbev/msi041>
- McCarthy, N. D., Colles, F. M., Dingle, K. E., Bagnall, M. C., Manning, G., Maiden, M. C. J., & Falush, D. (2007). Host-associated genetic import in *Campylobacter jejuni*. *Emerging Infectious Diseases*, 13(2), 267–272. <https://doi.org/10.3201/eid1302.060620>
- Pérez-Losada, M., Browne, E. B., Madsen, A., Wirth, T., Viscidi, R. P., & Crandall, K. A. (2006). Population genetics of microbial pathogens estimated from multilocus sequence typing (MLST) data. *Infection, Genetics and Evolution: Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, 6(2), 97–112.
<https://doi.org/10.1016/j.meegid.2005.02.003>
- Robinson, D. A., & Enright, M. C. (2004). Evolution of *Staphylococcus aureus* by large chromosomal replacements. *Journal of Bacteriology*, 186(4), 1060–1064.
<https://doi.org/10.1128/jb.186.4.1060-1064.2004>
- Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., ... Denamur, E. (2009). Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genetics*, 5(1), e1000344.
<https://doi.org/10.1371/journal.pgen.1000344>
- Wilson, D. J., Gabriel, E., Leatherbarrow, A. J. H., Cheesbrough, J., Gee, S., Bolton, E., ... Fearnhead, P. (2009). Rapid evolution and the importance of recombination to the gastroenteric pathogen *Campylobacter jejuni*. *Molecular Biology and Evolution*, 26(2), 385–397. <https://doi.org/10.1093/molbev/msn264>