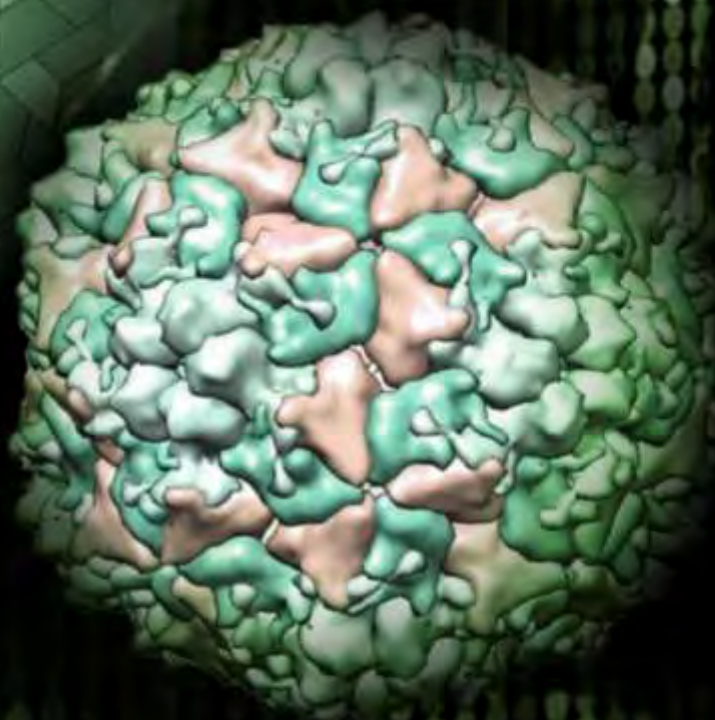




ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΤΜΗΜΑ ΒΙΟΧΗΜΕΙΑΣ & ΒΙΟΤΕΧΝΟΛΟΓΙΑΣ

ΓΟΝΙΔΙΩΜΑΤΙΚΗ ΚΑΙ ΕΞΕΛΙΚΤΙΚΗ ΑΝΑΛΥΣΗ ΑΝΘΡΩΠΙΝΩΝ ΕΝΤΕΡΟΪΩΝ ΜΕ ΜΕΘΟΔΟΥΣ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ



ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ ΤΣΙΜΠΙΔΗ ΜΙΧΑΗΛ

ΓΟΝΙΔΙΩΜΑΤΙΚΗ ΚΑΙ ΕΞΕΛΙΚΤΙΚΗ ΑΝΑΛΥΣΗ
ΑΝΘΡΩΠΙΝΩΝ ΕΝΤΕΡΟΪΩΝ ΜΕ ΜΕΘΟΔΟΥΣ
ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ

GENOMIC AND EVOLUTIONARY ANALYSIS
OF HUMAN ENTEROVIRUS WITH
BIOINFORMATIC METHODS

Τσιμπίδης Μιχαήλ

Τριμελής επιτροπή:

Αμούτζιας Γρηγόριος (Επιβλέπων)

Επίκουρος Καθηγητής Βιοπληροφορικής στη Γενωμική

Τμήμα Βιοχημείας και Βιοτεχνολογίας

Μαρκουλάτος Παναγιώτης

Καθηγητής Εφαρμοσμένης Μικροβιολογίας με έμφαση στη Βιοτεχνολογία

Τμήμα Βιοχημείας και Βιοτεχνολογίας

Κυριακοπούλου Ζαχαρούλα

Ειδικό Ερευνητικό και Διδακτικό Προσωπικό (ΕΕΔΙΠ)

Εκπαιδευτικό Εργαστήριο Βιολογίας και Ερευνητικό Εργαστήριο Ιολογίας

Τμήμα Βιοχημείας και Βιοτεχνολογίας

Αρχικά θα ήθελα να ευχαριστήσω τον κύριο Αμούτζια Γρηγόριο, Επίκουρο Καθηγητή Βιοπληροφορικής στη Γενωμική, του τμήματος Βιοχημείας και Βιοτεχνολογίας του Πανεπιστημίου Θεσσαλίας, για τη στήριξη και τη βοήθεια που μου παρείχε καθ' όλο το διάστημα εκπόνησης της διπλωματικής μου εργασίας.

Ακόμη, θα ήθελα να ευχαριστήσω τον κύριο Μαρκουλάτο Παναγιώτη, Καθηγητή Εφαρμοσμένης Μικροβιολογίας με έμφαση στη Βιοτεχνολογία, του τμήματος Βιοχημείας και βιοτεχνολογίας και μέλος της τριμελούς επιτροπής καθώς και την κυρία Κυριακοπούλου Ζαχαρούλα, μέλος του Ειδικού Ερευνητικού και Διδακτικού Προσωπικού (ΕΕΔΙΠ) στο Εκπαιδευτικό Εργαστήριο Βιολογίας και Ερευνητικό Εργαστήριο Ιολογίας και μέλος της τριμελούς επιτροπής για τις συμβουλές που μου έδωσαν κατά την εκπόνηση της διπλωματικής μου εργασίας.

Περιεχόμενα

Περίληψη	7
Summary	8
1 Εισαγωγή	9
1.1 Εντεροϊόι του ανθρώπου	9
1.2 Ταξινόμηση ανθρώπινων εντεροϊών	9
1.3 Δομή και γενετικό υλικό ανθρώπινων εντεροϊών	10
1.4 Ο κύκλος ζωής των εντεροϊών	11
1.5 Οι εντεροϊόι ανασυνδυάζονται συχνά	12
1.6 Ανασυνδυασμοί στα είδη των ανθρώπινων εντεροϊών	12
1.7 Φυλογενετικές αναλύσεις	14
1.8 Εργαλείο BLAST (Basic Local Alignment Search Tool)	14
1.9 Πολλαπλή στοίχιση (Multiple Sequence Alignment – MSA)	15
1.10 Κατασκευή Φυλογενετικών Δέντρων	15
1.11 Λογισμικό PhyML	16
1.12 Μοντέλα αντικατάστασης - Λογισμικό JModelTest	16
2 Υλικά και Μέθοδοι	18
2.1 Συλλογή νουκλεοτιδικών ακολουθιών ανθρώπινων εντεροϊών από τη Genbank	18
2.2 Συλλογή πρωτεϊνικών ακολουθιών ανθρώπινων εντεροϊών από την Genbank	18
2.3 Προετοιμασία 938 γονιδιωμάτων για την ανάλυση	18
2.4 Πολλαπλή στοίχιση ακολουθιών με το MUSCLE	19
2.5 Εύρεση κατάλληλου μοντέλου αντικατάστασης για την κατασκευή των φυλογενετικών δέντρων	19
2.6 Κατασκευή φυλογενετικών δέντρων	19
2.7 Απεικόνιση και επεξεργασία φυλογενετικών δέντρων	19
2.8 Αρχείο σχολιασμού δεδομένων	20
2.9 Αναζήτηση πιθανών περιπτώσεων ανασυνδυασμού	20
2.10 Διαθέσιμα αρχεία προς ανάλυση	21
2.11 Σύνοψη των βημάτων της ανάλυσης	22
3 Αποτελέσματα	23
3.1 JModelTest – Μοντέλα για την κατασκευή των φυλογενετικών δέντρων	23
3.2 ProtTest – Μοντέλα για την κατασκευή των φυλογενετικών δέντρων με βάση τις πρωτεϊνικές ακολουθίες	25
3.3 Phyml – Κατασκευή των φυλογενετικών δέντρων	27

3.4 Προετοιμασία των δεδομένων για τη χρήση τους στο Blastscript.....	29
3.5 Blastscript	29
3.5.1 Περιγραφή και τρόπος λειτουργίας.....	29
3.5.2 Έλεγχος πιθανών ανασυνδυασμών με το Blastscript.....	30
3.5.3 Χρησιμοποιώντας το Blastscript	31
3.6 Εφαρμογή του Blastscript στα δεδομένα.....	33
3.7 Μορφή αποτελεσμάτων του Blastscript	34
3.8 Κατασκευή αρχείων σχολιασμού.....	35
3.9 Απεικόνιση των φυλογενετικών δέντρων	36
3.10 Παρατήρηση μονοφυλετικότητας	48
3.11 Κατασκευή δικτύων δοτών – δεκτών για περιπτώσεις πιθανών γεγονότων ανασυνδυασμού βάσει του Blastscript	49
3.12 Γραφήματα τάσεων ανασυνδυασμών ανά περιοχή.....	49
4 Συζήτηση	54
4.1 Μονοφυλετικότητα.....	54
4.2 Φυλογενετικά δέντρα – Περιπτώσεις ανασυνδυασμών.....	54
4.3 Προβλήματα στην αξιολόγηση των φυλογενετικών δέντρων.....	56
Βιβλιογραφία.....	59

Περίληψη

Οι ανθρώπινοι εντεροϊοί μολύνουν μεγάλο αριθμό ατόμων κάθε χρόνο ανά τον κόσμο. Ανήκουν στην οικογένεια Picornaviridae και χωρίζονται σε 4 είδη (HEVA, HEVB, HEVC και HEVD). Το γενετικό τους υλικό είναι μονόκλωνο RNA το οποίο περιέχει 11 γονίδια (P2A, P2B, P2C, P3A, P3C, P3D, VP1, VP2, VP3, VP4, VPg) και 2 μη μεταγραφόμενες περιοχές (5'UTR, 3'UTR). Στους ανθρώπινους εντεροϊούς έχουν παρατηρηθεί φαινόμενα ανασυνδυασμού τα οποία επηρεάζουν άμεσα την εξελικτική πορεία τους. Στην παρούσα εργασία περιγράφεται η γονιδιωματική και εξελικτική ανάλυση των ανθρώπινων εντεροϊών για καθεμιά από τις παραπάνω περιοχές (εκτός της 3'UTR) με μεθόδους βιοπληροφορικής. Τα δεδομένα συλλέχθηκαν από την Genbank και περιείχαν νουκλεοτιδικές αλλά και πρωτεϊνικές ακολουθίες. Για καθεμιά από τις 12 περιοχές που αναλύθηκαν, δημιουργήθηκε το αντίστοιχο φυλογενετικό δέντρο ώστε να παρατηρηθεί η μονοφυλετικότητα μεταξύ των τεσσάρων ειδών. Επίσης δημιουργήθηκαν φυλογενετικά δέντρα (συνολικά 44) για καθεμιά από τις 12 περιοχές για το κάθε είδος ξεχωριστά. Ακόμη με τη χρήση ενός custom script που ονομάστηκε Blastscript, αναζητήθηκαν πιθανές περιπτώσεις ανασυνδυασμού μεταξύ διαφορετικών οροτύπων σε καθεμιά από τις περιοχές που μελετήθηκαν. Τέλος, κατασκευάστηκαν δίκτυα στα οποία παρουσιάζονται οι δότες και οι δέκτες σε πιθανά γεγονότα ανασυνδυασμού. Τέλος κατασκευάστηκαν γραφήματα που φανερώνουν τις τάσεις ανασυνδυασμού των ανθρώπινων εντεροϊών ανάμεσα στις περιοχές που μελετήθηκαν.

Summary

Human enteroviruses infect a remarkable number of people worldwide annually. They belong to the family of the Picornaviridae and are separated into 4 species (HEVA, HEVB, HEVC και HEVD). Their genetic material is a single strand RNA and contains 11 genes (P2A, P2B, P2C, P3A, P3C, P3D, VP1, VP2, VP3, VP4, VPg) and 2 non translated regions (5'UTR, 3'UTR). Among human enteroviruses there have been occurrences of recombinations, which affect directly their evolution . In this thesis, the genomic and evolutionary analysis of human enteroviruses is described for every single region (except from 3'UTR) using bioinformatic methods. All the data were acquired from Genbank and consisted of both nucleotide and protein sequences. For each one of the 12 regions that were analyzed, phylogenetic trees were constructed in order to study recombination among and within the four species. Furthermore, using a custom Perl script which was given the name Blastsript, a search for possible recombination events among the different serotypes for each one of the analyzed regions took place. Additionally, networks which illustrated donors and acceptors in possible recombination events were constructed. Finally, based on phylogenetic trees, graphs revealed which regions had the highest occurrences of recombination among the various serotypes of human enteroviruses.

1 Εισαγωγή

1.1 Εντεροϊοί του ανθρώπου

Οι εντεροϊοί του ανθρώπου μολύνουν εκατομμύρια άτομα παγκοσμίως κάθε χρόνο και παρουσιάζουν ένα ευρύ φάσμα συμπτωμάτων όπως άσηπτη μηνιγγίτιδα, μυοκαρδίτιδα και οξεία παράλυση. Σε πολλές περιπτώσεις όμως η μόλυνση μπορεί να είναι ασυμπτωματική [1]. Η κατανόηση της εξέλιξης των ιών αυτών μέσα από μεταλλάξεις σημείου και ανασυνδυασμούς θα μας επιτρέψει να σχεδιάσουμε αποτελεσματικά εμβόλια. Σκοπός αυτής της εργασίας είναι η μελέτη της εξέλιξης και ανασυνδυασμού των ιών αυτών χρησιμοποιώντας ακολουθίες και γονιδιώματα από ελεύθερα διαθέσιμες βάσεις δεδομένων και αναλύοντας τα δεδομένα αυτά με μεθόδους βιοπληροφορικής.

1.2 Ταξινόμηση ανθρώπινων εντεροϊών

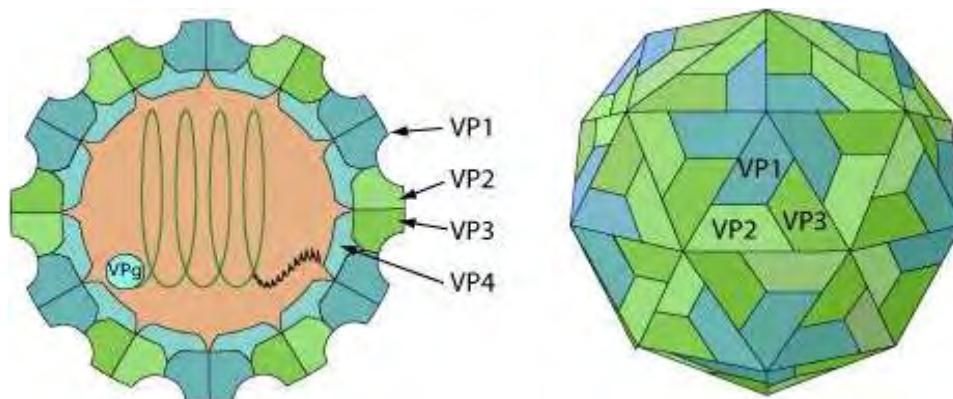
Οι ανθρώπινοι εντεροϊοί ανήκουν στην οικογένεια των *Picornaviridae* και χωρίζονται σε 4 είδη: *Human enterovirus A* (HEVA), *Human enterovirus B* (HEVB), *Human enterovirus C* (HEVC) και *Human enterovirus D* (HEVA) [2]. Στα 4 αυτά είδη κατατάσσονται οι διαφορετικοί ορότυποι των ανθρώπινων εντεροϊών οι οποίοι είναι οι *Coxsackievirus A* και *B*, *Echovirus*, *Poliovirus* καθώς και μερικοί ακόμα ορότυποι με την ονομασία *Enterovirus*. Η ταξινόμηση των ανθρώπινων εντεροϊών στα 4 είδη έγινε βάσει της γονιδιακής περιοχής VP1 και παρουσιάζεται παρακάτω στον πίνακα 1.

Human enterovirus A	
Enterovirus	71, 76, 89 - 92
Coxsackievirus	A2 – A8, A10, A12, A14, A16
Human enterovirus B	
Echovirus	E1 – E9, E11 – E21, E24 – E27, E29 – E33
Enterovirus	69, 73 – 75, 77 – 88, 93, 95, 97, 98, 100, 101, 107
Coxsackievirus	A9, B1 – B6
Human enterovirus C	
Poliovirus	1, 2, 3
Enterovirus	C95, 96, 99, 104, 105, 109, 116
Coxsackievirus	A1, A11, A13, A15, A17 – A22, A24
Human enterovirus D	
Enterovirus	68, 70, 94

Πίνακας 1: Ταξινόμηση ανθρώπινων εντεροϊών
από <http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=12059>

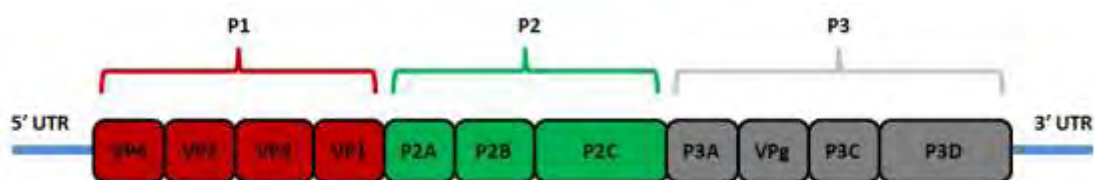
1.3 Δομή και γενετικό υλικό ανθρώπινων εντεροϊών

Οι ανθρώπινοι εντεροϊοί έχουν εικοσαεδρική συμμετρία, δεν έχουν φάκελο, το μέγεθός τους είναι περίπου 30 nm και το γενετικό τους υλικό είναι μονόκλωνο RNA μήκους περίπου 7000 – 8500 νουκλεοτιδίων. Η θεμελιώδης καψιδιακή δομή είναι ίδια σε όλα τα είδη ανθρώπινων εντεροϊών. Απαρτίζεται από 60 αντίγραφα καθεμιάς από τις 4 δομικές πρωτεΐνες VP1, VP2, VP3 και VP4 και προστατεύει το γενωμικό RNA [3].



Εικόνα 1: Δομή ενός εντεροϊού από http://viralzone.expasy.org/all_by_species/97.html

Το RNA των ανθρώπινων εντεροϊών περιλαμβάνει στο 5' άκρο μια αμετάφραστη περιοχή (5' UTR) καθώς και μια 3' πολυ(A) ουρά ποικίλου μήκους. Τα γονίδια οργανώνονται σε 3 περιοχές που ονομάζονται P1, P2 και P3 αντίστοιχα. Η περιοχή P1 περιέχει τα γονίδια των τεσσάρων δομικών πρωτεϊνών (VP1-4). Η P2 περιέχει ένα γονίδιο μιας πρωτεάσης (2A^{pro}) καθώς επίσης και δύο γονίδια (2B και 2C) που κωδικοποιούν πρωτεΐνες οι οποίες εμπλέκονται στην αντιγραφή του RNA και την καταστολή της γονιδιακής έκφρασης του κυττάρου-ξενιστή. Τέλος, στην περιοχή P3 περιλαμβάνονται τα γονίδια μιας πρωτεάσης (3C^{pro}) και μιας RNA-εξαρτώμενης RNA πολυμεράσης (3D^{pol}). Επίσης στην περιοχή P3 εντοπίζεται ένα γονίδιο που κωδικοποιεί την πρωτεΐνη VPg η οποία συνδέεται ομοιοπολικά στο 5' άκρο του γονιδιώματος του ιού [4]. Η σειρά των γονιδίων πάνω στο γονιδίωμα των ανθρώπινων εντεροϊών φαίνεται στην εικόνα 2.

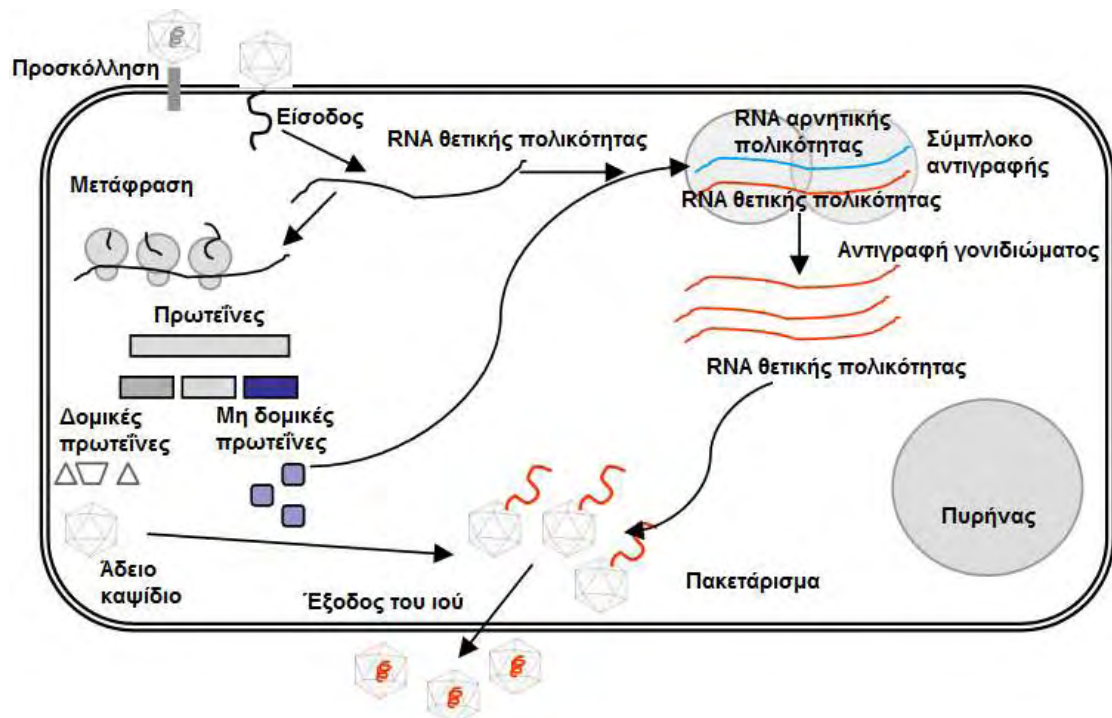


Εικόνα 2: Γονιδίωμα ανθρώπινων εντεροϊών

1.4 Ο κύκλος ζωής των εντεροϊών

Όπως προαναφέρθηκε, οι εντεροϊοί μολύνουν εκατομμύρια ανθρώπων ετησίως. Για την επίτευξη του στόχου τους ακολουθούν έναν πολύ συγκεκριμένο μηχανισμό ο οποίος έχει ως εξής:

Αρχικά αλληλεπιδρούν μέσω του καψιδίου τους με ειδικούς υποδοχείς στην κυτταρική επιφάνεια του ξενιστή. Οι υποδοχείς αυτοί διαφέρουν μεταξύ των διαφορετικών εντεροϊών. Η αλληλεπίδραση αυτή προκαλεί τη δημιουργία ανακατατάξεων στις καψιδιακές πρωτεΐνες που έχουν σαν αποτέλεσμα την απελευθέρωση της VP4 και του αμινοτελικού άκρου της VP1. Στη συνέχεια σχηματίζεται ένας μεμβρανικός πόρος στην κυτταρική επιφάνεια του κυττάρου του ξενιστή από μία εκ των δύο αυτών πρωτεϊνών ή ακόμα και από τις δύο μαζί, μέσα από τον οποίο είναι σε θέση πλέον να διέλθει το RNA του ιού στο κυτταρόπλασμα. Το RNA των εντεροϊών είναι θετικής πολικότητας επομένως μεταφράζεται κατευθείαν μετά την είσοδό του στον ξενιστή και μεταγράφεται από την 3D RNA εξαρτώμενη RNA πολυμεράση, μέσα σε κυστίδια, σε μια συμπληρωματική (αρνητικής πολικότητας) αλυσίδα που χρησιμεύει ως εκμαγείο για τη σύνθεση νέων ιικών RNA. Στο σχηματισμό των κυστιδίων παίζει ρόλο η πρωτεΐνη 3A. Οι πρωτεΐνες 2B, 2C και VPg παίζουν επίσης ρόλο στη δημιουργία του αρνητικής πολικότητας RNA. Τα παραγόμενα, θετικής πολικότητας, RNA εισέρχονται στα νεοσχηματισμένα καψίδια και τελικά εξέρχονται από το κύτταρο ξενιστή. Μια σχηματική αναπαράσταση της παραπάνω διαδικασίας περιγράφεται στην εικόνα 3.



Εικόνα 3: Κύκλος ζωής των εντεροϊών. Η εικόνα αυτή αποτελεί μια τροποποιημένη εκδοχή της <http://www.intechopen.com/books/type-1-diabetes/viruses-and-type-1-diabetes-focus-on-the-enteroviruses>

1.5 Οι εντεροϊοί ανασυνδυάζονται συχνά

Στους ανθρώπινους εντεροϊούς έχουν παρατηρηθεί συχνές περιπτώσεις ανασυνδυασμού και μεταλλάξεων που οφείλονται στην, επιρρεπή σε λάθη, πολυμεράση [5, 6]. Το γεγονός αυτό επηρεάζει άμεσα την εξελικτική πορεία των εντεροϊών δίνοντάς τους τη δυνατότητα να προσαρμόζονται εύκολα σε περιπτώσεις όπου το περιβάλλον μπορεί να μην είναι ευνοϊκό. Ανασυνδυασμοί έχουν παρατηρηθεί μεταξύ διαφορετικών οροτύπων (διαοροτυπικοί ανασυνδυασμοί), αλλά και μεταξύ διαφορετικών ειδών (διαειδικοί ανασυνδυασμοί). Οι διαοροτυπικοί και οι διαειδικοί ανασυνδυασμοί δε συμβαίνουν με την ίδια συχνότητα. Έχει παρατηρηθεί ότι οι ανασυνδυασμοί μεταξύ διαφορετικών οροτύπων είναι συχνότεροι από αυτούς μεταξύ διαφορετικών ειδών. Αξίζει να σημειωθεί πως συχνότεροι διαοροτυπικοί ανασυνδυασμοί έχουν παρατηρηθεί μεταξύ των στελεχών των ανθρώπινων εντεροϊών που ανήκουν στους Human enteroviruses B [7].

Οι θέσεις που συμβαίνουν οι ανασυνδυασμοί διαφέρουν ανάμεσα στα διαφορετικά είδη, μια παρατήρηση που φανερώνει πως μπορεί να υπάρχει διαφορά στη συμβατότητα των ανασυνδυαζόμενων περιοχών μεταξύ των στελεχών των διαφορετικών ειδών. Η περιοχή όπου έχει παρατηρηθεί συχνότερος ανασυνδυασμός για τους HEV-B είναι η P2 ενώ για τους μη Polio HEV-C είναι η P3 [8, 9, 10]. Η περιοχή P1 φαίνεται να μην παρουσιάζει συχνά το φαινόμενο του ανασυνδυασμού. Αρκετά επιρρεπής σε ανασυνδυασμούς φαίνεται να είναι και η περιοχή 5' UTR. Παρόλα αυτά η περιοχή 5' UTR είναι απαραίτητη για τον πολλαπλασιασμό των εντεροϊών καθώς αλληλεπιδρά με πρωτεΐνες όπως η 2B, VPg, 3A και 3D αλλά και με άλλες πρωτεΐνες του ξενιστή. Αυτό σημαίνει πως οι ανασυνδυασμοί που συμβαίνουν σε αυτήν την περιοχή θα πρέπει να διατηρούν τις αλληλεπιδράσεις αυτές ή να συμβαίνουν ταυτόχρονα συμπληρωματικές μεταλλάξεις στις μη δομικές πρωτεΐνες [11, 12]. Τη μελέτη για ανασυνδυασμούς πάνω στην περιοχή 5' UTR δυσχεραίνει ο σχετικά περιορισμένος αριθμός διαθέσιμων ακολουθιών της περιοχής αυτής, καθώς και το ότι είναι αρκετά συντηρημένη.

1.6 Ανασυνδυασμοί στα είδη των ανθρώπινων εντεροϊών

Human enterovirus A (HEV-A)

Η εξέλιξη ορισμένων οροτύπων (EV76, EV89, EV90 και EV91) καθορίστηκε από ανασυνδυασμούς. Τα τελευταία χρόνια έχουν παρατηρηθεί μεγάλες επιδημίες στη Νότια Ασία και τα στελέχη που απομονώθηκαν από αυτές ανήκαν στον ορότυπο EV71 [13]. Αυτός είναι ο κύριος λόγος για τον οποίο ο ορότυπος EV71 έχει μελετηθεί περισσότερο από κάθε άλλο του είδους του. Ανασυνδυασμοί μεταξύ διαφορετικών στελεχών καθώς και διαφορετικών οροτύπων έχουν συντελέσει σε μεγάλο βαθμό στην εξέλιξη των EV71 [14]. Άλλοι ορότυποι αυτού του είδους που φαίνεται να έχουν ανασυνδυαστεί είναι ο CAV16 όπου έχει παρατηρηθεί πως περιέχει κοινή την

περιοχή 5' UTR με τον CVA4, καθώς και παρόμοια περιοχή P2-P3 με στελέχη του EV71 [15]. Στην εικόνα 4 φαίνεται μια σχηματική αναπαράσταση ενός φαινομένου ανασυνδυασμού μεταξύ EV71 και CVA16. Άλλοι ορότυποι των οποίων τα στελέχη φαίνεται να ανασυνδυάζονται είναι οι CAV2, CAV4, CAV5 και CAV10 [16].

Human enterovirus B (HEV-B)

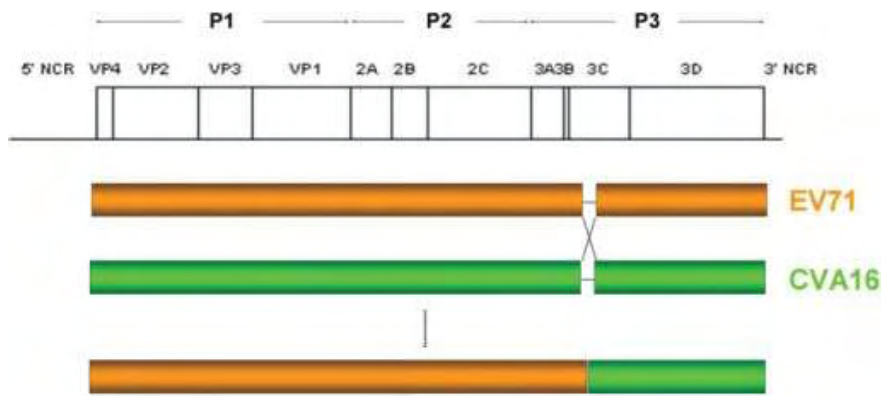
Όπως στους HEV-A έτσι και στους HEV-B έχουν παρατηρηθεί φαινόμενα ανασυνδυασμού μεταξύ των μελών αυτού του είδους. Τα στοιχεία αυτά έχουν προκύψει από μελέτες ολόκληρων γονιδιωμάτων στελεχών μέσα στους HEV-B [17]. Στην περίπτωση των HEV-B έχει παρατηρηθεί μεγαλύτερη συχνότητα στα φαινόμενα του ανασυνδυασμού από ότι στα υπόλοιπα είδη. Ορισμένοι ορότυποι στα στελέχη των οποίων έχει παρατηρηθεί ανασυνδυασμός είναι οι E7, E30, E11, E9 και CVB [18, 19].

Human enterovirus C (HEV-C)

Οι Polioviruses έχουν μελετηθεί σε μεγάλο βάθος και ανασυνδυάζονται συχνά μεταξύ τους. Όσον αφορά τους μη-Polioviruses έχουν αναφερθεί περιπτώσεις ανασυνδυασμού ανάμεσα σε Polioviruses και άλλα μέλη του είδους (μη-Polioviruses) [20, 21]. Οι γνώσεις για τους μη Polio εντεροϊούς της κατηγορίας αυτής είναι περιορισμένες. Ένας ορότυπος για τον οποίο έχει παρατηρηθεί ανασυνδυασμός είναι ο EV96. Ύστερα από φυλογενετική ανάλυση σε διάφορα στελέχη του EV96 για την περιοχή 3D, τρία στελέχη του οροτύπου αυτού ομαδοποιήθηκαν με στελέχη PV1 φανερώνοντας ότι ενδεχομένως οι PV1 δρουν ως πιθανοί δότες γενετικού υλικού ανάμεσα στους HEV-C.

Human enterovirus D (HEV-D)

Σε αντίθεση με τα υπόλοιπα είδη, φαινόμενα ανασυνδυασμού δεν έχουν παρατηρηθεί μέχρι στιγμής στους HEV-D. Σημαντικό ρόλο σε αυτό φαίνεται να παίζει το γεγονός ότι οι ορότυποι των εντεροϊών που αντιπροσωπεύουν αυτό το είδος είναι μόλις 3 (EV67, EV70 και EV94). Προς το παρόν δεν είναι ασφαλές να καταλήξουμε στο συμπέρασμα ότι δε συμβαίνουν ανασυνδυασμοί ανάμεσα στους HEV-D καθώς υπάρχει προφανής έλλειψη δεδομένων.



Εικόνα 4: Ανασυνδυασμός μεταξύ EV71 και CVA16

1.7 Φυλογενετικές αναλύσεις

Οι εξελικτικές σχέσεις μεταξύ γονιδίων/οργανισμών μπορούν να αποκαλυφθούν μέσα από τα φυλογενετικά δέντρα. Η κατασκευή τους και η ερμηνεία των αποτελεσμάτων αποτελούν την φυλογενετική ανάλυση. Τα βασικά βήματα μιας φυλογενετικής ανάλυσης είναι η απόκτηση/εύρεση ομόλογων ακολουθιών, η πολλαπλή στοίχιση των ακολουθιών και στη συνέχεια η κατασκευή των φυλογενετικών δέντρων

1.8 Εργαλείο BLAST (Basic Local Alignment Search Tool)

Η εύρεση ομόλογων ακολουθιών από μια βάση δεδομένων γίνεται με αλγόριθμους τοπικής στοίχισης κατά ζεύγη, ο πιο διαδεδομένος από τους οποίους είναι ο BLAST. Επιπλέον, μπορούν να εντοπιστούν οι ομοιότητες και οι διαφορές δύο ομόλογων ακολουθιών.

Τα αποτελέσματα που θα επιστρέψει το BLAST καθορίζονται από τις παραμέτρους με τις οποίες γίνεται η αναζήτηση. Η σημαντικότερη παράμετρος είναι το *e-value*. Πρόκειται για μια τιμή η οποία περιγράφει την πιθανότητα ενός αποτελέσματος να εμφανιστεί ως ομόλογο καθαρά από τύχη.

Το εργαλείο BLAST παρέχει τη δυνατότητα πραγματοποίησης διαφόρων μορφών στοίχισης. Στον πίνακα 2 φαίνονται τα προγράμματα που μπορούν να χρησιμοποιηθούν μέσα από το BLAST.

Πρόγραμμα	Βάση δεδομένων	Ακολουθία επερώτησης
BLASTN	Νουκλεοτιδική	Νουκλεοτιδική
BLASTP	Πρωτεϊνική	Πρωτεϊνική
BLASTX	Πρωτεϊνική	Νουκλεοτιδική μεταφρασμένη σε πρωτεϊνική
TBLASTN	Νουκλεοτιδική μεταφρασμένη σε πρωτεϊνική	Πρωτεϊνική
TBLASTX	Νουκλεοτιδική μεταφρασμένη σε πρωτεϊνική	Νουκλεοτιδική μεταφρασμένη σε πρωτεϊνική

Πίνακας 2: Προγράμματα του BLAST

1.9 Πολλαπλή στοίχιση (Multiple Sequence Alignment – MSA)

Κατά την πολλαπλή στοίχιση έχουμε την στοίχιση ορισμένων ακολουθιών μιας οικογένειας (από τρεις και πάνω) με αποτέλεσμα τον εντοπισμό συντηρημένων και μη περιοχών μεταξύ των ακολουθιών αυτών. Χρησιμοποιείται για τη δημιουργία profiles/motifs που χαρακτηρίζουν μια επικράτεια (domain), για την ανίχνευση συντηρημένων DNA-binding sites σε προαγωγείς γονιδίων, στην πρόβλεψη δευτεροταγούς και τριτοταγούς δομής πρωτεϊνών, στο σχεδιασμό εκφυλισμένων εκκινητών PCR και ειδικότερα στη φυλογενετική ανάλυση.

Η πολλαπλή στοίχιση μπορεί να πραγματοποιηθεί με διαφορετικές μεθόδους οι οποίες εφαρμόζονται από εξειδικευμένα προγράμματα. Οι μέθοδοι αυτοί μπορεί να είναι ο *Δυναμικός προγραμματισμός* ή κάποια *Ευρετική μέθοδος (Προοδευτική στοίχιση, Στοίχιση με διαδοχικές βελτιώσεις, Στοίχιση βασισμένη σε Blocks)*. Ορισμένα αρκετά διαδεδομένα προγράμματα που πραγματοποιούν πολλαπλή στοίχιση είναι το MUSCLE, το ClustalW και το T-coffee τα οποία βασίζονται στην προοδευτική στοίχιση.

Κανένα πρόγραμμα δεν παράγει τη βέλτιστη στοίχιση και για το λόγο αυτό θα πρέπει να πραγματοποιηθεί βελτίωση της στοίχισης χειροκίνητα. Δύο γνωστά προγράμματα που χρησιμοποιούνται για το σκοπό αυτό είναι το Seaview και το Bioedit.

1.10 Κατασκευή Φυλογενετικών Δέντρων

Βασίζεται σε δύο μεγάλες κατηγορίες αλγορίθμων, σε μεθόδους αποστάσεων (distance-based methods) και σε μεθόδους που βασίζονται σε χαρακτήρες (character-based methods).

Μέθοδοι βασιζόμενες σε χαρακτήρες: Οι μέθοδοι που περιλαμβάνονται στην κατηγορία αυτή είναι η *Μέγιστη Πιθανοφάνεια (maximum likelihood)* και η *Μέγιστη Φειδωλότητα (maximum parsimony)*. Η Μέγιστη Πιθανοφάνεια στηρίζεται στην αναζήτηση του δέντρου όπου τα δεδομένα που παρατηρούμε είναι τα πιο πιθανά, δεδομένου ενός δέντρου και ενός μοντέλου εξέλιξης. Στην περίπτωση της Μέγιστης

Φειδωλότητας η βασική αρχή (Ξυράφι του Okham) είναι ότι θα πρέπει να προτιμηθεί εκείνη η απεικόνιση της εξελικτικής διαδικασίας η οποία, με τις λιγότερες δυνατές αντικαταστάσεις χαρακτήρων (εξελικτικά βήματα), εξηγεί τα δεδομένα. Σήμερα, οι μέθοδοι Μέγιστης Πιθανοφάνειας θεωρούνται οι καλύτερες, αλλά με σχετικά μεγάλο υπολογιστικό κόστος.

Μέθοδοι βασιζόμενες σε αποστάσεις: Οι μέθοδοι αυτής της κατηγορίας είναι οι *UPGMA*, *Neighbor Joining*, *Fitch-Margoliash* και *Ελάχιστης εξέλιξης*. Στις μεθόδους αυτές αρχικά υπολογίζονται οι αποστάσεις ανάμεσα σε όλα τα πιθανά ζεύγη ακολουθιών και στη συνέχεια δημιουργείται ένας πίνακας αποστάσεων. Βάσει του πίνακα αυτού δημιουργούνται δέντρα που βασίζονται στην ομαδοποίηση (*UPGMA*, *Neighbour Joining*) ή στην βελτιστοποίηση (*Fitch-Margoliash*, *Ελάχιστη εξέλιξη*). Η πιο διαδεδομένη από τις παραπάνω μεθόδους είναι η *Neighbor Joining*. Η μέθοδος αυτή θεωρεί πως όλες οι ακολουθίες εξελίσσονται με διαφορετικό ρυθμό και παράγει ένα δέντρο το οποίο δεν έχει ρίζα. Το δέντρο αυτό αρχίζει να σχηματίζεται με την ομαδοποίηση των ακολουθιών οι οποίες έχουν τη μικρότερη απόσταση μεταξύ τους.

1.11 Λογισμικό PhyML

Το PhyML είναι μια εφαρμογή η οποία βασίζεται στην αρχή της *Μέγιστης Πιθανοφάνειας* και χρησιμοποιείται για την κατασκευή φυλογενετικών δέντρων. Το PhyML εφαρμόζει μεθόδους για την δημιουργία αξιοπιστία των κλάδων του δέντρου όπως είναι η *ανάλυση bootstrap* και το *Approximately-Likelihood Ratio Test* (aLRT).

1.12 Μοντέλα αντικατάστασης - Λογισμικό JModelTest

Για την κατασκευή ενός δέντρου είναι απαραίτητο να χρησιμοποιηθεί κάποιο μοντέλο αντικατάστασης για πού περιγράφει την εξελικτική διαδικασία. Υπάρχουν αρκετά μοντέλα διαθέσιμα όπως τα *Jukes-Cantor (JC)*, *Kimura 2 parameter (K2P)*, *Felsenstein (F81)*, *Hasegawa et al. (HKY85)*. Το πρόγραμμα *JModelTest* μπορεί και εντοπίζει το καταλληλότερο μοντέλο αντικατάστασης νουκλεοτιδίων για την κατασκευή ενός δέντρου από μία πολλαπλή στοίχιση.

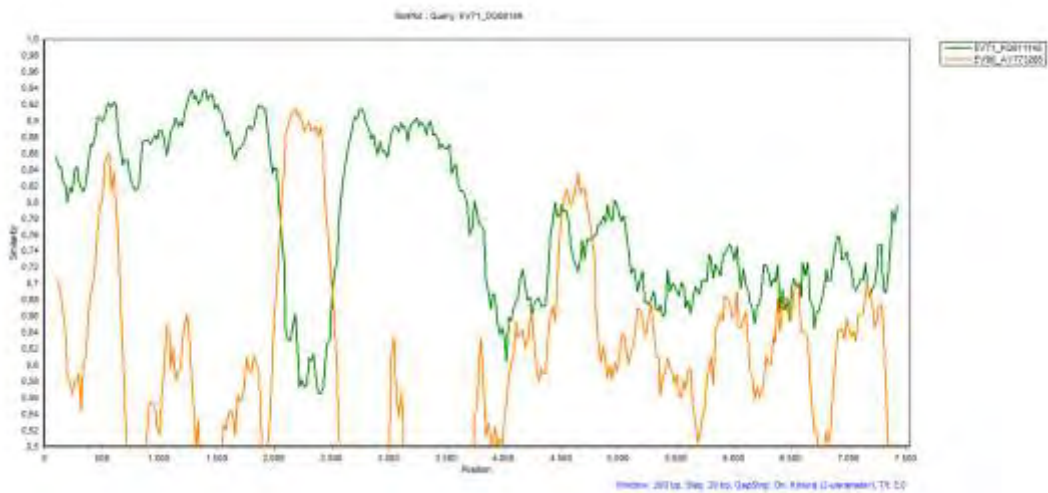
1.13 Προγράμματα εντοπισμού ανασυνδυασμού - SimPlot

Το πρόγραμμα *SimPlot* αποτελεί ένα υπολογιστικό εργαλείο για τη σύγκριση ακολουθιών σε μια πολλαπλή στοίχιση με τη μέθοδο συρόμενων παραθύρων ως προς μια ακολουθία επερώτησης, ώστε να αξιολογηθεί η μεταξύ τους ομοιότητα και να εντοπιστούν τυχόν ανασυνδυασμοί. Το αποτέλεσμα της διαδικασίας αυτής είναι ένα γράφημα όπου φαίνεται το ποσοστό ομοιότητας για όλες τις θέσεις της κάθε ακολουθίας που εξετάζεται ως προς την ακολουθία επερώτησης.

Στη συνέχεια ακολουθεί ένα παράδειγμα για τη χρήση του SimPlot:



Εικόνα 5: Προσομίωση ανασυνδυασμού μεταξύ δύο γονιδιωμάτων (EV71 και EV90) στις θέσεις 2000 έως 2500



Εικόνα 7: Αποτελέσματα SimPlot της παραπάνω προσομίωσης ανασυνδυασμού

2 Υλικά και Μέθοδοι

2.1 Συλλογή νουκλεοτιδικών ακολουθιών ανθρώπινων εντεροϊών από τη Genbank

Κατά τη διαδικασία της ανάλυσης χρησιμοποιήθηκαν 39995 νουκλεοτιδικές ακολουθίες εντεροϊών οι οποίες συλλέχθηκαν από τη βάση δεδομένων Genbank του NCBI σε μορφή FASTA. Η λέξη-κλειδί που χρησιμοποιήθηκε για την αναζήτηση ήταν «enterovirus». Τα δεδομένα που συγκεντρώθηκαν αποτελούνταν από ακολουθίες-τμήματα του γονιδιώματος αλλά και από ακολουθίες ολόκληρων γονιδιωμάτων. Ακολούθως, έγινε επιμέρους επιλογή 938 γονιδιωμάτων ανθρώπινων εντεροϊών A, B, C και D από τις 39995 ακολουθίες. Οι 938 ακολουθίες-γονιδιώματα φαίνονται στον πίνακα 3.

938 ΑΚΟΛΟΥΘΙΕΣ ΑΝΘΡΩΠΙΝΩΝ ΕΝΤΕΡΟΪΩΝ	
Enterovirus A	375
Enterovirus B	200
Enterovirus C	353
Enterovirus D	10

Πίνακας 3: Προγράμματα του BLAST

Επίσης, δημιουργήθηκε μία λίστα με δύο στήλες, η μία εκ των οποίων περιείχε τους accession numbers των 39995 ακολουθιών και η δεύτερη τον ορότυπο που αντιστοιχούσε σε κάθε accession number. Το αρχείο αυτό ονομάστηκε **genbankexp**.

2.2 Συλλογή πρωτεϊνικών ακολουθιών ανθρώπινων εντεροϊών από την Genbank

Έπειτα από τη συγκέντρωση των 938 γονιδιωμάτων, συλλέχθηκαν και οι πρωτεϊνικές ακολουθίες των περιοχών P2A, P2B, P2C, P3A, P3D, VP1, VP2, VP3 και VPg για κάθε είδος εντεροϊών από την Genbank.

2.3 Προετοιμασία 938 γονιδιωμάτων για την ανάλυση

Αρχικά κάθε γονιδιωματική ακολουθία από τα 938 γονιδιώματα χωρίστηκε σε 12 επιμέρους τμήματα τα οποία αντιστοιχούσαν στις περιοχές 5'UTR, P2A, P2B, P2C, P3A, P3D, VP1, VP2, VP3 και VPg. Έτσι, δημιουργήθηκαν 12 αρχεία για κάθε ένα γονίδιο/περιοχή που περιείχαν τις 938 ακολουθίες και αποτέλεσαν το πρώτο σετ ακολουθιών. Ένα δεύτερο σετ ακολουθιών δημιουργήθηκε με το διαχωρισμό των ακολουθιών του κάθε είδους από τις ακολουθίες των υπολοίπων ειδών. Κάθε αρχείο του δεύτερου σετ συνεπώς περιείχε ακολουθίες μίας εκ των 12 περιοχών για ένα από τα τέσσερα είδη εντεροϊών. Έτσι, το σετ αυτό περιείχε συνολικά 48 αρχεία (12 περιοχές X 4 είδη εντεροϊών).

Η προετοιμασία των πρωτεϊνικών ακολουθιών έγινε με παρόμοιο τρόπο (εκτός της 5'UTR).

2.4 Πολλαπλή στοίχιση ακολουθιών με το MUSCLE

Στη συνέχεια τόσο στις ακολουθίες που εμπεριέχονταν σε κάθε αρχείο του πρώτου σετ νουκλεοτιδικών ακολουθιών όσο και σε αυτές στα αρχεία του δεύτερου σετ, εφαρμόστηκε πολλαπλή στοίχιση με τη βοήθεια του προγράμματος MUSCLE στο software Seaview. Μετά την εφαρμογή των πολλαπλών στοιχίσεων ακολούθησε χειρωνακτική επιδιόρθωση των αποτελεσμάτων με τη βοήθεια του προγράμματος Seaview. Οι ακολουθίες του πρώτου σετ μετατράπηκαν και σε μορφή fasta για μετέπειτα χρήση.

Επίσης, πολλαπλή στοίχιση έγινε και στα 44 αρχεία στα οποία εμπεριέχονταν οι πρωτεϊνικές ακολουθίες και πάλι με το πρόγραμμα MUSCLE.

2.5 Εύρεση κατάλληλου μοντέλου αντικατάστασης για την κατασκευή των φυλογενετικών δέντρων

Καθένα από τα αρχεία πολλαπλών στοιχίσεων ελέγχθηκε για την εύρεση του καταλληλότερου μοντέλου αντικατάστασης με το πρόγραμμα JModelTest (για πολλαπλές στοιχίσεις νουκλεοτιδικών ακολουθιών) και το ProtTest (για πολλαπλές στοιχίσεις αμινοξικών ακολουθιών).

2.6 Κατασκευή φυλογενετικών δέντρων

Έχοντας τις πολλαπλές στοιχίσεις και γνωρίζοντας τα καταλληλότερα μοντέλα εξέλιξης για κάθε μια από αυτές, κατασκευάστηκαν φυλογενετικά δέντρα με την μέθοδο maximum likelihood από το πρόγραμμα PhyML. Επιπλέον, κατασκευάστηκαν και φυλογενετικά δέντρα με τη μέθοδο Neighbor Joining και τα μοντέλα εξέλιξης Kimura (για πρωτεΐνες) και Jukes-Cantor (για νουκλεοτιδικές ακολουθίες).

2.7 Απεικόνιση και επεξεργασία φυλογενετικών δέντρων

Το πρόγραμμα Treedyn χρησιμοποιήθηκε για την απεικόνιση των φυλογενετικών δέντρων και την ενσωμάτωση άλλων πληροφοριών.

2.8 Αρχείο σχολιασμού δεδομένων

Για την κατανόηση του κάθε φυλογενετικού δέντρου ήταν απαραίτητη η κατασκευή ενός αρχείου για το σχολιασμό των ακολουθιών (annotation file). Η μορφή των περιεχομένων του κάθε annotation file παρουσιάζεται στον πίνακα 4.

Annotation file	
Accession number	Ο κωδικός πρόσβασης του στελέχους
Serotype	Ο ορότυπος του στελέχους
Organism	Το είδος στο οποίο ανήκει το στέλεχος
Country	Η χώρα προέλευσης του στελέχους
Year	Η χρονολογία απομόνωσης του στελέχους
938_to_938	Πιθανοί ανασυνδυασμοί βάσει του Blastscript
938_to_39995	

Πίνακας 4: Μορφή περιεχομένων ενός αρχείου σχολιασμού δεδομένων.

Στις τελευταίες δύο σειρές του πίνακα 4 εμφανίζονται δύο καταχωρήσεις που σχετίζονται με την ύπαρξη πιθανών γεγονότων ανασυνδυασμού. Ο ακριβής ορισμός του ονόματος αλλά και του περιεχομένου τους δίνεται στην ενότητα 3.5.

2.9 Αναζήτηση πιθανών περιπτώσεων ανασυνδυασμού

Ο εντοπισμός πιθανών γεγονότων ανασυνδυασμού μεταξύ διαφορετικών οροτύπων έγινε με τη βοήθεια ενός custom script το οποίο κατασκευάστηκε στη γλώσσα προγραμματισμού Perl και ονομάστηκε Blastscript, επειδή βασίστηκε στο Blast.

Το Blastscript αποτελεί ένα πρόγραμμα το οποίο μπορεί να χρησιμοποιεί τα προγράμματα του NCBI-Blast και να πραγματοποιεί το Blast τοπικά στον υπολογιστή ενώ ο χρήστης παρέχει ο ίδιος τη βάση δεδομένων καθώς και την ακολουθία (ή ακολουθίες) επερώτησης. Επίσης δίνεται η δυνατότητα φιλτραρίσματος των αποτελεσμάτων του Blast έτσι ώστε να μπορεί να εντοπίζει πιθανά ενδεχόμενα ανασυνδυασμού. Εκτενέστερη περιγραφή του προγράμματος καθώς και του τρόπου λειτουργίας του γίνεται στην ενότητα 3.3.

2.10 Διαθέσιμα αρχεία προς ανάλυση

Στη συνέχεια παρατίθενται τα αρχεία που χρησιμοποιήθηκαν στην ανάλυση, συγκεντρωτικά στον πίνακα 5.

Διαθέσιμα αρχεία προς ανάλυση			
EVA_P2A_aln_nucl.phy	EVB_P2A_aln_nucl.phy	EVC_P2A_aln_nucl.phy	EVD_P2A_aln_nucl.phy
EVA_P2B_aln_nucl.phy	EVB_P2B_aln_nucl.phy	EVC_P2B_aln_nucl.phy	EVD_P2B_aln_nucl.phy
EVA_P2C_aln_nucl.phy	EVB_P2C_aln_nucl.phy	EVC_P2C_aln_nucl.phy	EVD_P2C_aln_nucl.phy
EVA_P3A_aln_nucl.phy	EVB_P3A_aln_nucl.phy	EVC_P3A_aln_nucl.phy	EVD_P3A_aln_nucl.phy
EVA_P3C_aln_nucl.phy	EVB_P3C_aln_nucl.phy	EVC_P3C_aln_nucl.phy	EVD_P3C_aln_nucl.phy
EVA_P3D_aln_nucl.phy	EVB_P3D_aln_nucl.phy	EVC_P3D_aln_nucl.phy	EVD_P3D_aln_nucl.phy
EVA_VP1_aln_nucl.phy	EVB_VP1_aln_nucl.phy	EVC_VP1_aln_nucl.phy	EVD_VP1_aln_nucl.phy
EVA_VP2_aln_nucl.phy	EVB_VP2_aln_nucl.phy	EVC_VP2_aln_nucl.phy	EVD_VP2_aln_nucl.phy
EVA_VP3_aln_nucl.phy	EVB_VP3_aln_nucl.phy	EVC_VP3_aln_nucl.phy	EVD_VP3_aln_nucl.phy
EVA_VP4_aln_nucl.phy	EVB_VP4_aln_nucl.phy	EVC_VP4_aln_nucl.phy	EVD_VP4_aln_nucl.phy
EVA_VPg_aln_nucl.phy	EVB_VPg_aln_nucl.phy	EVC_VPg_aln_nucl.phy	EVD_VPg_aln_nucl.phy
938_P2A.phy	938_P3A.phy	938_VP1.phy	938_VP4.phy
938_P2A.fasta	938_P3A.fasta	938_VP1.fasta	938_VP4.fasta
938_P2B.phy	938_P3C.phy	938_VP2.phy	938_VPg.phy
938_P2B.fasta	938_P3C.fasta	938_VP2.fasta	938_VPg.fasta
938_P2C.phy	938_P3D.phy	938_VP3.phy	938_5UTR.phy
938_P2C.fasta	938_P3D.fasta	938_VP3.fasta	938_5UTR.fasta
938_ev_nucl.fasta		39995_ev_nucl.fasta	

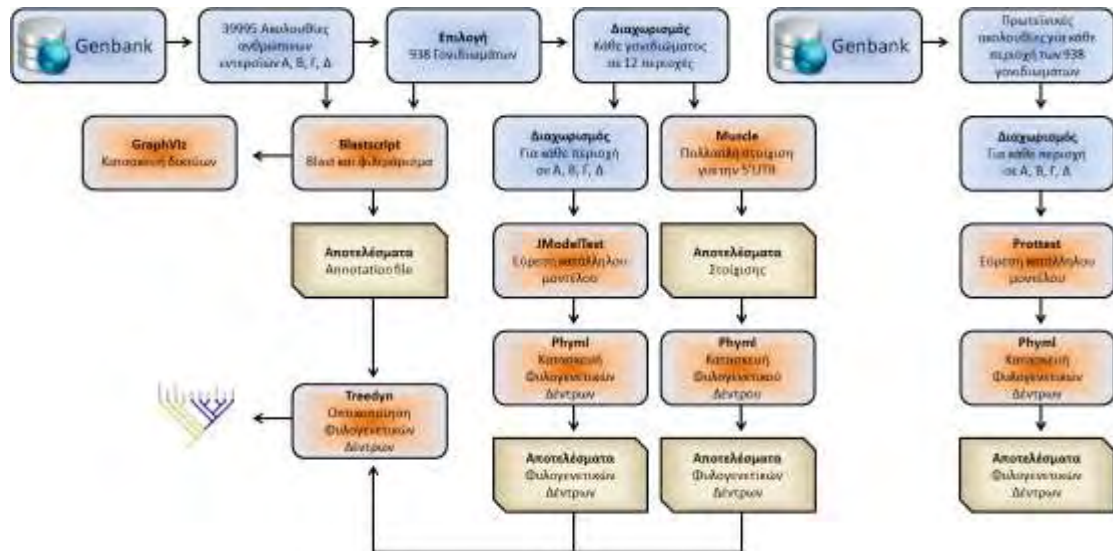
Πίνακας 5: Τα διαθέσιμα αρχεία προς ανάλυση

- Στοιχισμένες, σε κάθε περιοχή, ακολουθίες σε fasta μορφή για κάθε είδος
- Στοιχισμένες, σε κάθε περιοχή, ακολουθίες σε phyip και fasta μορφή (όλα τα είδη)
- 39995 ακολουθίες εντεροϊών και 938 ολόκληρα γονιδιώματα από αυτές σε ξεχωριστά αρχεία σε μορφή fasta

Σημειώνεται ότι στις αναλύσεις δεν χρησιμοποιήθηκαν τα αρχεία με τις περιοχές 5'UTR για καθένα από τα 4 είδη καθώς τα αποτελέσματα που λαμβάνονταν δεν ήταν εφικτό να αξιολογηθούν.

2.11 Σύνοψη των βημάτων της ανάλυσης

Στην εικόνα 8 παρουσιάζονται συνοπτικά τα βήματα που ακολουθήθηκαν κατά τη διάρκεια της ανάλυσης.



Εικόνα 8: Διαγραμματική σύνοψη των βημάτων της ανάλυσης

3 Αποτελέσματα

Στη συνέχεια γίνεται αναφορά στις εντολές και τις παραμέτρους που χρησιμοποιήθηκαν κατά την πορεία της φυλογενετικής ανάλυσης των ανθρώπινων εντεροϊών. Σημειώνεται πως η μελέτη έλαβε χώρα εξ' ολοκλήρου σε περιβάλλον Linux και συγκεκριμένα στην έκδοση Ubuntu 12.10.

3.1 JModelTest – Μοντέλα για την κατασκευή των φυλογενετικών δέντρων

Ύστερα από τη συλλογή και την πολλαπλή στοίχιση των δεδομένων ακολούθησε η αναζήτηση του καταλληλότερου μοντέλου για κάθε αρχείο που περιείχε τα δεδομένα της πολλαπλής στοίχισης. Η διαδικασία αυτή επιτεύχθηκε με τη χρήση του προγράμματος JModelTest. Η εντολή η οποία περιλαμβάνει όλες τις παραμέτρους που χρησιμοποιήθηκαν για την αναζήτηση είναι η ακόλουθη:

```
java -jar jModelTest.jar -d input_file -g 4 -i -f -s 7 -tr 10 -t BIONJ -AICc -p
```

όπου input_file κάθε ένα από τα 48 αρχεία που περιέχουν τις στοιχισμένες νουκλεοτιδικές ακολουθίες κάθε είδους εντεροϊών (A, B, C και D) για κάθε περιοχή.

Επεξήγηση παραμέτρων

-d → αρχείο εισόδου

-g → gamma

Αφορά τον τις κατηγορίες ρυθμών εξέλιξης.

-i → συμπερίληψη invariable sites

-f → χρήση συχνότητας

-s → αριθμός μοντέλων αντικατάστασης

Σημειώνεται ότι χρησιμοποιήθηκαν 7 μοντέλα αντικατάστασης διότι με χρήση περισσότερων προέκυπταν στα αποτελέσματα μοντέλα που δεν ήταν εφικτό να χρησιμοποιηθούν μετέπειτα στο πρόγραμμα rhyml.

-tr → threads

-t → χρησιμοποιούμενη μέθοδος κατασκευής δέντρου

-AICc → υπολογισμός διορθωμένου Akaike information criterion

-p → υπολογισμός parameter importances

Τα αποτελέσματα του υπολογισμού των καταλληλότερων μοντέλων για κατασκευή δέντρου από κάθε αρχείο πολλαπλών στοιχίσεων παρουσιάζεται στον πίνακα 6.

Αποτελέσματα μοντέλων	
EVA_P2A_aln_nucl.phy	GTR + G
EVA_P3A_aln_nucl.phy	GTR + G
EVA_P3D_aln_nucl.phy	GTR + G
EVB_P3A_aln_nucl.phy	GTR + G
EVB_P3C_aln_nucl.phy	GTR + G
EVD_P3A_aln_nucl.phy	GTR + G
EVD_VP1_aln_nucl.phy	GTR + G
EVA_P2B_aln_nucl.phy	GTR + I + G
EVA_VP4_aln_nucl.phy	GTR + I + G
EVA_VPg_aln_nucl.phy	GTR + I + G
EVB_P2B_aln_nucl.phy	GTR + I + G
EVB_P2C_aln_nucl.phy	GTR + I + G
EVB_P3D_aln_nucl.phy	GTR + I + G
EVB_VP1_aln_nucl.phy	GTR + I + G
EVB_VP2_aln_nucl.phy	GTR + I + G
EVB_VP3_aln_nucl.phy	GTR + I + G
EVB_VP4_aln_nucl.phy	GTR + I + G
EVB_VPg_aln_nucl.phy	GTR + I + G
EVC_P2A_aln_nucl.phy	GTR + I + G
EVC_P2B_aln_nucl.phy	GTR + I + G
EVC_P2C_aln_nucl.phy	GTR + I + G
EVC_P3A_aln_nucl.phy	GTR + I + G
EVC_P3C_aln_nucl.phy	GTR + I + G
EVC_P3D_aln_nucl.phy	GTR + I + G
EVC_VP1_aln_nucl.phy	GTR + I + G
EVC_VP4_aln_nucl.phy	GTR + I + G
EVC_VPg_aln_nucl.phy	GTR + I + G
EVD_P2A_aln_nucl.phy	GTR + I + G
EVD_P2C_aln_nucl.phy	GTR + I + G
EVD_P3C_aln_nucl.phy	GTR + I + G
EVD_P3D_aln_nucl.phy	GTR + I + G
EVD_VP3_aln_nucl.phy	GTR + I + G
EVD_P2B_aln_nucl.phy	HKY + G
EVD_VPg_aln_nucl.phy	HKY + G
EVA_VP1_aln_nucl.phy	HKY + I + G
EVA_VP2_aln_nucl.phy	K80 + G
EVA_VP3_aln_nucl.phy	K80 + G
EVC_VP3_aln_nucl.phy	K80 + G
EVB_P2A_aln_nucl.phy	K80 + I + G
EVA_P2C_aln_nucl.phy	TrN+G
EVD_VP4_aln_nucl.phy	TrN+G
EVA_P3C_aln_nucl.phy	TrN+I+G
EVC_VP2_aln_nucl.phy	TrN+I+G
EVD_VP2_aln_nucl.phy	TrN+I+G

Πίνακας 6: Αποτελέσματα κατάλληλων μοντέλων από το JModelTest για την κατασκευή των φυλογενετικών δέντρων βάσει νουκλεοτιδικών ακολουθιών

3.2 ProtTest – Μοντέλα για την κατασκευή των φυλογενετικών δέντρων με βάση τις πρωτεϊνικές ακολουθίες

Στη συνέχεια με τη βοήθεια του προγράμματος ProtTest έγινε ο αντίστοιχος έλεγχος εύρεσης κατάλληλων μοντέλων για την κατασκευή φυλογενετικών δέντρων με βάση τις πρωτεϊνικές ακολουθίες που κωδικοποιούνται από κάθε περιοχή των εντεροϊών. Η εντολή που χρησιμοποιήθηκε στο πρόγραμμα ProtTest ήταν η ακόλουθη:

```
java -jar prottest-3.2.jar -threads 14 -i inputfile -S 0 -JTT -LG -DCMut -MtREV -MtMam -MtArt -Dayhoff -WAG -RtREV -CpREV -Blosum62 -VT -HIVb -HIVw -o outputfile -all-distributions -all -AIC -t1
```

Όπου inputfile το αρχείο εισόδου των δεδομένων και outputfile το αρχείο εξόδου των αποτελεσμάτων. Ως αρχείο εισόδου χρησιμοποιήθηκε καθένα από τα 44 αρχεία με τις στοιχισμένες πρωτεϊνικές ακολουθίες.

Επεξήγηση παραμέτρων

-threads → χρησιμοποιούμενα threads

-i → αρχείο εισόδου

-S → Starting topology selection mode

-f → χρήση συχνότητας

-JTT έως και **-HIVw** → μοντέλα αντικατάστασης που θα χρησιμοποιηθούν

-o → αρχείο εξόδου των αποτελεσμάτων

-all-distributions → Συμπερίληψη των μοντέλων +I, +G και +IG

-all → Δημιουργία ενός πίνακα με όλα τα model selection frameworks

-AIC → Ταξινόμηση των μοντέλων βάσει του Akaike Information Criterion

-t1 → Δημιουργία ενός δέντρου σε Newick format βάσει του καλύτερου μοντέλου

Τα αποτελέσματα των μοντέλων που προέκυψαν από τον παραπάνω έλεγχο φαίνονται στον πίνακα 7.

Αποτελέσματα μοντέλων	
EVA_P2A_aln_prot.phy	JTT+G
EVA_P3A_aln_prot.phy	JTT+G
EVB_P3C_aln_prot.phy	JTT+G
EVA_P2B_aln_prot.phy	JTT+G
EVA_VP4_aln_prot.phy	JTT+G
EVC_VP4_aln_prot.phy	JTT+G
EVC_VPg_aln_prot.phy	JTT+G
EVB_P2C_aln_prot.phy	JTT+G
EVB_P2A_aln_prot.phy	JTT+G
EVA_P2C_aln_prot.phy	JTT+G
EVA_P3C_aln_prot.phy	JTT+G
EVB_VP1_aln_prot.phy	LG+I+G
EVB_VP2_aln_prot.phy	LG+I+G
EVB_VP3_aln_prot.phy	LG+I+G
EVC_P2B_aln_prot.phy	LG+I+G
EVC_VP1_aln_prot.phy	LG+I+G
EVC_VP2_aln_prot.phy	LG+I+G
EVD_VP2_aln_prot.phy	LG+I+G
EVA_VP1_aln_prot.phy	LG+I+G
EVC_P2C_aln_prot.phy	LG+G
EVC_P3A_aln_prot.phy	LG+G
EVC_P3C_aln_prot.phy	LG+G
EVC_P2A_aln_prot.phy	LG+G
EVB_VP4_aln_prot.phy	LG+G
EVB_P2B_aln_prot.phy	LG+G
EVD_P3A_aln_prot.phy	LG+G
EVD_VP1_aln_prot.phy	LG+G
EVD_P2A_aln_prot.phy	LG+G
EVD_P2C_aln_prot.phy	LG+G
EVD_P3C_aln_prot.phy	LG+G
EVD_P3D_aln_prot.phy	LG+G
EVD_VP3_aln_prot.phy	LG+G
EVD_P2B_aln_prot.phy	LG+G
EVD_VP4_aln_prot.phy	LG+G
EVA_VP2_aln_prot.phy	LG+G
EVA_VP3_aln_prot.phy	LG+G
EVC_VP3_aln_prot.phy	LG+G
EVC_P3D_aln_prot.phy	HIVb+I+G
EVA_P3D_aln_prot.phy	HIVb+I+G
EVB_P3D_aln_prot.phy	HIVb+I+G
EVB_P3A_aln_prot.phy	HIVb+G
EVA_VPg_aln_prot.phy	HIVb+G
EVD_VPg_aln_prot.phy	RtREV
EVB_VPg_aln_prot.phy	RtREV+I

Πίνακας 7: Αποτελέσματα κατάλληλων μοντέλων από το ProtTest για την κατασκευή των φυλογενετικών δέντρων βάσει πρωτεϊνικών ακολουθιών

3.3 PhymI – Κατασκευή των φυλογενετικών δέντρων

Αφού συλλέχθηκαν τα κατάλληλα μοντέλα για την κατασκευή των φυλογενετικών δέντρων έλαβε χώρα ο υπολογισμός των δέντρων αυτών με τη βοήθεια του προγράμματος PhymI. Για εξοικονόμηση χρόνου, όσα αρχεία στοιχισμένων ακολουθιών είχαν κοινό αποτέλεσμα όσον αφορά το μοντέλο κατασκευής φυλογενετικού δέντρου που απέδωσε το JModelTest, συγκεντρώθηκαν με τη βοήθεια ενός custom script σε ένα ενιαίο αρχείο (dataset). Βασική προϋπόθεση για να αναγνωριστεί κάθε dataset από το PhymI ήταν η εξής: τα περιεχόμενα κάθε αρχείου, που περιλάμβανε τα δεδομένα της στοίχισης σε μορφή phylip, έπρεπε να απέχουν μεταξύ τους διάστημα μιας κενής γραμμής.

Αφού κατασκευάστηκαν τα datasets (8 στον αριθμό) ακολούθησε η κατασκευή των φυλογενετικών δέντρων βάσει των νουκλεοτιδικών ακολουθιών και του μοντέλου με χρήση του προγράμματος PhymI. Η εντολή που χρησιμοποιήθηκε για το PhymI ήταν η παρακάτω:

```
phymI -i dataset -d nt -b -4 -m model -a e -s NNI -f e -t e -v e -c 4 -o tlr --no_memory_check --quiet -n number
```

Επεξήγηση παραμέτρων

-d → Τύπος δεδομένων

Χρήση του τύπου «nt» (nucleotide) διότι οι ακολουθίες στα datasets είναι νουκλεοτιδικές.

-b → bootstrap

Η τιμή 4 αφορά το SH-like branch supports alone.

-m → Μοντέλο που θα χρησιμοποιηθεί για την κατασκευή του δέντρου

Όπου **model** στην παραπάνω εντολή το αποτέλεσμα του καταλληλότερου μοντέλου που αφορά το τρέχον dataset και προέκυψε από τα αποτελέσματα του JModelTest.

-a → gamma

Η τιμή e αφορά το Maximum likelihood

-s → Χρησιμοποιούμενη μέθοδος

NNI (Nearest Neighbor Interchange)

-f → Συχνότητα νουκλεοτιδίων

-t → tv_ratio

-v → invariable sites

-c → Substitution categories

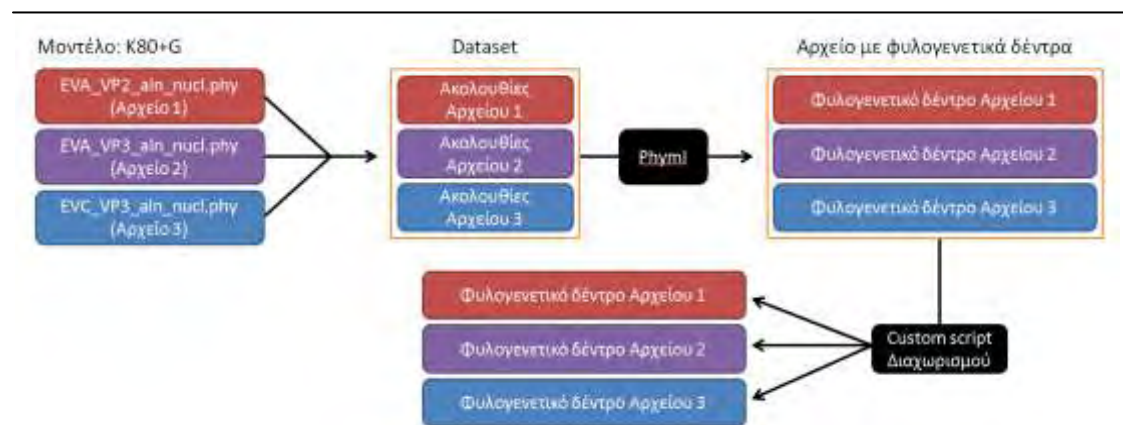
-o → Παράμετροι tlr

Tree topology (t), branch length (l), substitution rate parameters (r)

-n → Αριθμός διαφορετικών αρχείων που περιέχονται στο dataset

Το Rhyml παρήγαγε δύο αρχεία αποτελεσμάτων. Το ένα περιείχε τα φυλογενετικά δέντρα για κάθε αρχείο που είχε προστεθεί στο dataset και το άλλο περιείχε τις στατιστικές λεπτομέρειες. Από τα αρχεία αυτά, εκείνο που χρησιμοποιήθηκε στη συνέχεια της πορείας της φυλογενετικής ανάλυσης ήταν αυτό που περιείχε μόνο τα φυλογενετικά δέντρα.

Καθένα από τα αποτελέσματα του Rhyml περιείχε διαφορετικό αριθμό φυλογενετικών δέντρων ο οποίος ήταν ανάλογος του αριθμού των διαφορετικών αρχείων στοιχισμένων ακολουθιών που περιείχε κάθε dataset. Τα δέντρα αυτά έπρεπε να διαχωριστούν και να μπουν σε ξεχωριστά αρχεία ώστε να είναι εφικτή η επεξεργασία τους και συνεπώς η συνέχεια της ανάλυσης. Ο σκοπός αυτός επιτεύχθηκε με ένα custom script το οποίο προγραμματίστηκε σε γλώσσα προγραμματισμού perl. Στην εικόνα 9 παρουσιάζεται σχηματικά η πορεία της ανάλυσης από τη δημιουργία ενός dataset (συγκεκριμένα του dataset που περιείχε τα αρχεία με τις ακολουθίες που βάσει του jModelTest παρουσίαζαν ως καταλληλότερο μοντέλο το K80+G) μέχρι τον διαχωρισμό των δέντρων που παράχθηκαν από το Rhyml.



Εικόνα 9: Διαγραμματική σύνοψη της δημιουργίας των φυλογενετικών δέντρων από τα αρχεία στα οποία εμπεριέχονταν οι πολλαπλές στοιχίσεις.

Η παραπάνω διαδικασία εφαρμόστηκε και για τα αρχεία με τις πρωτεϊνικές ακολουθίες. Η εντολή για το Rhyml ήταν ίδια, με μοναδική διαφορά στην παράμετρο «-d» όπου αντί για «nt» έγινε χρήση του «aa» (amino acid) εφ' όσον επρόκειτο για πρωτεϊνικές ακολουθίες.

3.4 Προετοιμασία των δεδομένων για τη χρήση τους στο Blastscript

Οι στοιχισμένες νουκλεοτιδικές ακολουθίες που βρίσκονταν στα 48 αρχεία πολλαπλών στοιχίσεων ήταν αυτά στα οποία θα γινόταν ο έλεγχος για ύπαρξη πιθανών περιπτώσεων ανασυνδυασμού. Οι ακολουθίες αυτές βρίσκονταν σε μορφή rhylip, μια μορφή που δεν αναγνωρίζεται από το Blastscript. Έτσι έπρεπε να γίνει η μετετροπή τους από rhylip σε fasta μορφή. Η διαδικασία αυτή έγινε με τη βοήθεια μιας επιλογής που βρίσκεται στο πρόγραμμα seaview και μετατρέπει τις ακολουθίες σε διαφορετικές μορφές (format).

3.5 Blastscript

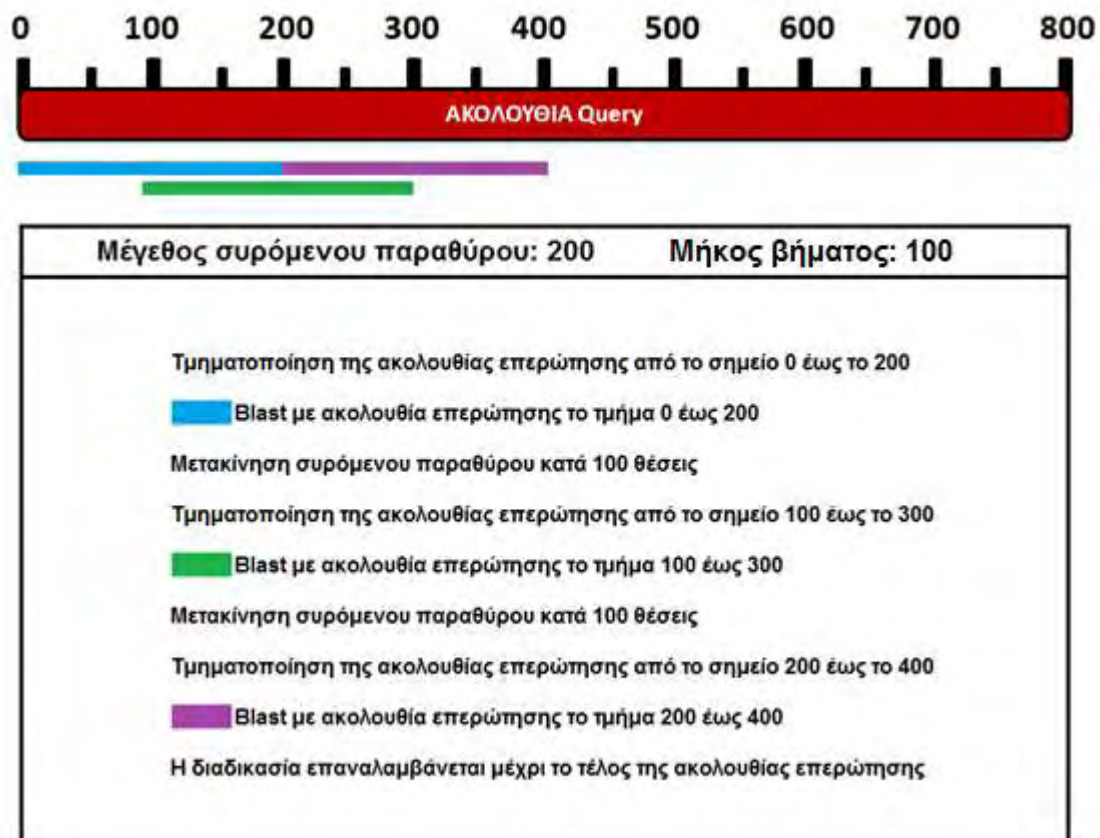
3.5.1 Περιγραφή και τρόπος λειτουργίας

Όπως έχει ήδη αναφερθεί, το Blastscript αποτελεί ένα πρόγραμμα το οποίο μπορεί να πραγματοποιεί Blast τοπικά στον υπολογιστή ενώ ο χρήστης παρέχει ο ίδιος τη βάση δεδομένων καθώς και την ακολουθία (ή ακολουθίες) επερώτησης. Τόσο οι ακολουθίες της βάσης δεδομένων όσο και η ακολουθία/ες επερώτησης θα πρέπει να βρίσκονται σε δύο αρχεία (database _file και query _file αντίστοιχα) σε μορφή FASTA και η μία κάτω από την άλλη με την ακόλουθη δομή:

```
>accession_number1  
Αλληλουχία1  
  
>accession_number2  
Αλληλουχία2
```

Το πρόγραμμα δέχεται τροποποίηση σε όλες τις παραμέτρους που χρησιμοποιούνται από το Blast.

Βασικό χαρακτηριστικό του Blastscript είναι ο έλεγχος για ανασυνδυασμούς μεταξύ μιας ακολουθίας επερώτησης και μιας ακολουθίας από αυτές της βάσης δεδομένων. Ο έλεγχος αυτός πραγματοποιείται με τμηματικό Blast της ακολουθίας επερώτησης: η ακολουθία αυτή «κόβεται» από το Blastscript σε μικρότερα τμήματα, των οποίων το μέγεθος καθορίζεται από τον χρήστη και στη συνέχεια με κάθε ένα από αυτά τα τμήματα γίνεται Blast επάνω στις ακολουθίες της βάσης δεδομένων. Επίσης το πρόγραμμα είναι ικανό να χωρίζει την ακολουθία επερώτησης σε μικρότερα τμήματα συγκεκριμένου μεγέθους βάσει ενός συρόμενου παραθύρου. Η βασική αυτή λειτουργία παρουσιάζεται σχηματικά στην εικόνα 10.



Εικόνα 10: Συνοπτική παρουσίασης τρόπου λειτουργίας του Blastscript. Στο συγκεκριμένο παράδειγμα το μέγεθος των επιμέρους τμημάτων που προκύπτουν από τον «τεμαχισμό» της ακολουθίας επερώτησης (μέγεθος συρόμενου παραθύρου) είναι 200 ενώ το μήκος βήματος είναι 100 θέσεις.

Θα πρέπει να σημειωθεί ότι σε πολλές περιπτώσεις ίσως υπάρξει πρόβλημα με το τελευταίο τμήμα που θα πρέπει να αποσπαστεί από την ακολουθία επερώτησης ώστε να γίνει το Blast με αυτό. Το πρόβλημα είναι ότι το μέγεθος του τμήματος αυτού ενδεχομένως να είναι αρκετά μικρότερο από το μέγεθος των τμημάτων που έχει οριστεί από τον χρήστη. Στην περίπτωση αυτή ο χρήστης μπορεί να επιλέξει ένα κατώφλι επιτρεπτών μεγεθών, ώστε να συμπεριληφθούν ακολουθίες με μέγεθος μεγαλύτερο του κατωφλίου ή να απορριφθούν αυτές με μικρότερο μέγεθος.

3.5.2 Έλεγχος πιθανών ανασυνδυασμών με το Blastscript

Αφού πραγματοποιηθούν τα Blast για όλα τα τμήματα κάθε μιας από τις ακολουθίες που βρίσκονται στο αρχείο με τις ακολουθίες επερώτησης ξεκινά μια σειρά από φιλτραρίσματα των αποτελεσμάτων του Blast ώστε να φανερωθούν τυχόν γεγονότα ανασυνδυασμού. Αρχικά από όλα τα αποτελέσματα αφαιρείται η πρώτη γραμμή καθώς αυτή αντιπροσωπεύει την ομολογία μεταξύ της ίδιας της ακολουθίας επερώτησης η οποία συμπεριλαμβάνεται στη βάση δεδομένων. Η ουσιαστική πορεία του φιλτραρίσματος είναι η ακόλουθη:

Ως αποτέλεσμα πιθανού ανασυνδυασμού θεωρείται το αποτέλεσμα του Blast που ικανοποιεί τις παρακάτω συνθήκες:

- Το ελεγχόμενο αποτέλεσμα θα πρέπει να έχει το μεγαλύτερο bitscore
- Ο ορότυπος του ελεγχόμενου αποτελέσματος θα πρέπει να είναι διαφορετικός από αυτόν της ακολουθίας επερώτησης
- Η διαφορά του ποσοστού ταύτισης μεταξύ του ελεγχόμενου αποτελέσματος και του αμέσως καλύτερου αποτελέσματος που έχει ίδιο ορότυπο με αυτόν της ακολουθίας επερώτησης να ξεπερνάει ένα ποσοστό που ορίζεται από τον χρήστη

Με τον τρόπο αυτό εξασφαλίζονται οι καλύτερες δυνατές προϋποθέσεις για την ύπαρξη τυχόν περιπτώσεων ανασυνδυασμού.

Τέλος, για τη σωστή λειτουργία του Blastscript είναι αναγκαία η ύπαρξη του αρχείου `genbankchr` το οποίο όπως αναφέρθηκε και στην ενότητα 2.2 περιέχει μία στήλη με τους accession numbers και μία με τους οροτύπους που αντιστοιχούν σε καθέναν από αυτούς.

3.5.3 Χρησιμοποιώντας το Blastscript

Όπως ήδη αναφέρθηκε, το Blastscript κατασκευάστηκε στη γλώσσα προγραμματισμού perl. Η χρήση και η δοκιμή του προγράμματος έγινε σε περιβάλλον Linux και συγκεκριμένα στα Ubuntu 12.04. Ύστερα από την εκτέλεση του προγράμματος θα πρέπει να οριστούν μερικές παράμετροι:

- Αρχείο ακολουθίας επερώτησης
- Αρχείο βάσης δεδομένων
- Το e-value με το οποίο θα γίνει το Blast

Οι υπόλοιπες παράμετροι βρίσκονται σε μια προκαθορισμένη (default) τιμή και είναι εφικτό να μεταβληθούν χειροκίνητα μέσω του προγράμματος. Οι παράμετροι αυτοί είναι:

- Word size
- Gap open
- Gap extend
- Penalty
- Reward

Στην εικόνα 10 παρουσιάζεται το Blastscript σε λειτουργία. Αρχικά το πρόγραμμα προτρέπει τον χρήστη να τοποθετήσει τα αρχεία που θα χρησιμοποιηθούν με τέτοιο τρόπο ώστε το αρχείο με τις ακολουθίες επερώτησης, το αρχείο με τις ακολουθίες της βάσης δεδομένων καθώς και το `genbankchr` να συνυπάρχουν στον ίδιο

κατάλογο. Χωρίς αυτή την προϋπόθεση, το πρόγραμμα ενδέχεται να μη λειτουργήσει σωστά.

```

x _ □
---IMPORTANT---
Ensure you have the query, database and genbank file in the current directory!
Press n for new blast or f to format results: n
Query file name: query_file
Database file name: database_file
e-value: 1e-10
Current parameters:
Database file: database_file
Query file: query_file
value: 1e-10 word size: 16 gapopen: 5
gapextend: 2 penalty: 3 reward: 2
Would you like to change any parameter? (y/n)█

```

Εικόνα 10: Εκτέλεση του προγράμματος *Blastscript* σε περιβάλλον *Linux, Ubuntu 12.10*

Στη συνέχεια, το πρόγραμμα ρωτάει τον χρήστη αν θέλει να ξεκινήσει ένα νέο Blast ή αν απλώς θέλει να φιλτράρει τα αποτελέσματα από ένα Blast που ο χρήστης έχει ήδη κάνει. Πληκτρολογώντας «n» επιλέγεται το μονοπάτι του νέου Blast και το πρόγραμμα ζητάει τα ονόματα των αρχείων με τις ακολουθίες επερώτησης και τις ακολουθίες της βάσης δεδομένων. Έπειτα ζητά την τιμή του e-value που θα χρησιμοποιηθεί για το Blast που θα ακολουθήσει. Τέλος εμφανίζει τις τρέχουσες παραμέτρους με τις οποίες θα τρέξει το Blast και ρωτά τον χρήστη αν θέλει να αλλάξει κάποια από τις τιμές αυτές. Σε περίπτωση που ο χρήστης πληκτρολογήσει «y» τότε εμφανίζεται μια λίστα με τις παραμέτρους όπου ο χρήστης επιλέγει εκείνη που θέλει να αλλάξει και ορίζει την τιμή της. Αν ο χρήστης πληκτρολογήσει «n» τότε το *Blastscript* ξεκινά τη διαδικασία «τεμαχισμού» των ακολουθιών επερώτησης και προχωρά στη διαδικασία του Blast.

3.6 Εφαρμογή του Blastscript στα δεδομένα

Έπειτα από την κατασκευή των φυλογενετικών δέντρων και πριν από την επεξεργασία τους έπρεπε να γίνει ένας έλεγχος για τυχόν γεγονότα ανασυνδυασμού μεταξύ διαφορετικών οροτύπων του ίδιου είδους εντεροϊών. Ο έλεγχος αυτός επιτεύχθηκε με τη βοήθεια του Blastscript. Όπως περιγράφηκε σε προηγούμενη ενότητα, το Blastscript βασίζεται στα προγράμματα του NCBI-Blast και στη συγκεκριμένη μελέτη στο πρόγραμμα blastn. Έτσι, η εντολή του blastn που ενσωματώνει το Blastscript είναι η εξής:

```
blastn -query queryseq -db database -evalue 1e-10 -word_size 16 -gapopen 5 -gapextend 2 -penalty -3 -reward 2 -outfmt "6 qseqid sseqid evalue qstart qend sstart send bitscore qlen slen pident ppos" > results1
```

Επεξήγηση βασικών παραμέτρων

-query → Ακολουθία επερώτησης

Όπου queryseq η ακολουθία επερώτησης που εισάγει το Blastscript στο blastn.

-db → Η βάση δεδομένων με τις νουκλεοτιδικές αλληλουχίες

Όπου database η κατάλληλα διαμορφωμένη βάση δεδομένων όπου έχει τροποποιηθεί το αρχείο βάσης δεδομένων που εισάγει ο χρήστης, έτσι ώστε αυτό να μπορεί να αναγνωριστεί από το blastn.

-evalue → Η τιμή e-value

Για το σκοπό της ανάλυσης έγιναν δύο σειρές από blast με το Blastscript:

Σειρά blast 1: Στην πρώτη σειρά blast, ως βάση δεδομένων χρησιμοποιήθηκαν οι 938 ακολουθίες γονιδιωμάτων ανθρώπινων εντεροϊών και ως ακολουθίες επερώτησης οι 938 ακολουθίες για καθεμιά από τις 11 περιοχές (εκτός της 5'UTR) καθενός από τα 4 είδη εντεροϊών.

Σειρά blast 2: Στην δεύτερη σειρά blast, ως βάση δεδομένων χρησιμοποιήθηκαν οι 39995 νουκλεοτιδικές ακολουθίες ανθρώπινων εντεροϊών και ως ακολουθίες επερώτησης οι 938 ακολουθίες για καθεμιά από τις 11 περιοχές (εκτός της 5'UTR) καθενός από τα 4 είδη εντεροϊών.

Στον πίνακα 7 παρουσιάζονται τα blast που έλαβαν χώρα στις δύο σειρές από blast.

Blast 938 (Σειρά 1)				Blast 39980 (Σειρά 2)			
Ακολουθίες επερώτησης				Ακολουθίες επερώτησης			
EVA_P2A	EVB_P2A	EVC_P2A	EVD_P2A	EVA_P2A	EVB_P2A	EVC_P2A	EVD_P2A
EVA_P2B	EVB_P2B	EVC_P2B	EVD_P2B	EVA_P2B	EVB_P2B	EVC_P2B	EVD_P2B
EVA_P2C	EVB_P2C	EVC_P2C	EVD_P2C	EVA_P2C	EVB_P2C	EVC_P2C	EVD_P2C
EVA_P3A	EVB_P3A	EVC_P3A	EVD_P3A	EVA_P3A	EVB_P3A	EVC_P3A	EVD_P3A
EVA_P3C	EVB_P3C	EVC_P3C	EVD_P3C	EVA_P3C	EVB_P3C	EVC_P3C	EVD_P3C
EVA_P3D	EVB_P3D	EVC_P3D	EVD_P3D	EVA_P3D	EVB_P3D	EVC_P3D	EVD_P3D
EVA_VP1	EVB_VP1	EVC_VP1	EVD_VP1	EVA_VP1	EVB_VP1	EVC_VP1	EVD_VP1
EVA_VP2	EVB_VP2	EVC_VP2	EVD_VP2	EVA_VP2	EVB_VP2	EVC_VP2	EVD_VP2
EVA_VP3	EVB_VP3	EVC_VP3	EVD_VP3	EVA_VP3	EVB_VP3	EVC_VP3	EVD_VP3
EVA_VP4	EVB_VP4	EVC_VP4	EVD_VP4	EVA_VP4	EVB_VP4	EVC_VP4	EVD_VP4
EVA_VPg	EVB_VPg	EVC_VPg	EVD_VPg	EVA_VPg	EVB_VPg	EVC_VPg	EVD_VPg
Database				Database			
938 Γονιδιώματα Εντεροϊών				39995 Ακολουθίες Εντεροϊών			

Πίνακας 7: Τα blast που έλαβαν χώρα σε καθεμιά από τις 2 σειρές blast με χρήση του Blastscript

3.7 Μορφή αποτελεσμάτων του Blastscript

Τα αποτελέσματα του Blastscript βρίσκονται σε ένα αρχείο κειμένου με όνομα «final_results». Παρακάτω περιγράφεται η μορφή των αποτελεσμάτων η οποία απαρτίζεται από 20 στήλες κατά μήκος μιας γραμμής του αρχείου αποτελεσμάτων:

Στήλη 1, Q-ORGANISM: Ορότυπος ακολουθίας επερώτησης

Στήλη 2, Q-ACCESSION: Accession number του συγκεκριμένου στελέχους

Στήλη 3, S-ORGANISM: Ορότυπος ακολουθίας βάσης δεδομένων

Στήλη 4, S-ACCESSION: Accession number του S-ORGANISM

Στήλη 5, E-VALUE: Τιμή e-value

Στήλη 6, S-START: Θέση έναρξης στοίχισης στην ακολουθία του S-ORGANISM

Στήλη 7, S-END: Θέση που ολοκληρώνεται η στοίχιση στον S-ORGANISM

Στήλη 8, S-LENGTH: Μήκος ακολουθίας του S-ORGANISM

Στήλη 9, BITSCORE: Τιμή Bitscore για το συγκεκριμένο αποτέλεσμα

Στήλη 10, PIDENT: Ποσοστό ταύτισης μεταξύ των S-ORGANISM και Q-ORGANISM

Στήλη 11, PPOS: Percentage of positive-scoring matches

Οι υπόλοιπες στήλες είναι ίδιες με τις παραπάνω (ξεκινώντας από τη στήλη 3) και αφορούν το blasthit με τον αμέσως επόμενο εντεροϊό που έχει ίδιο ορότυπο με αυτόν της ακολουθίας επερώτησης.

Επίσης πρέπει να σημειωθεί ότι στη στήλη 2 εκτός από το Accession number της ακολουθίας επερώτησης, εμφανίζεται και η τοποθεσία της επάνω στην ακολουθία ολόκληρης της περιοχής από την οποία έχει προέλθει η ακολουθία επερώτησης. Υπενθυμίζεται ότι το Blastscript δε χρησιμοποιεί κατευθείαν μια ακολουθία που

βρίσκεται στο αρχείο με τις ακολουθίες επερώτησης που έχει εισάγει ο χρήστης. Αντίθετα, κόβει την καθεμιά από αυτές σε μικρότερα κομμάτια τα οποία στη συνέχεια χρησιμοποιεί για ακολουθίες επερώτησης. Οι συντεταγμένες των κομματιών αυτών εμφανίζονται στη στήλη 2 με τη μορφή:

X---START-->END

Όπου X, ο Accession number του στελέχους, START η θέση από την οποία ξεκινά το κομμάτι επάνω στην αρχική ακολουθία και END η θέση στην οποία τελειώνει το κομμάτι αυτό επάνω στην αρχική ακολουθία.

3.8 Κατασκευή αρχείων σχολιασμού

Η κατασκευή του αρχείου σχολιασμού με σκοπό την ευκολότερη επεξεργασία των φυλογενετικών δέντρων έγινε με τη βοήθεια δύο custom script τα οποία προγραμματίστηκαν στη γλώσσα προγραμματισμού Perl. Απαραίτητη ήταν η δημιουργία ενός αρχείου με όλες τις απαραίτητες πληροφορίες για κάθε ένα από τα 938 στελέχη εντεροϊών που ήταν διαθέσιμα για την ανάλυση. Για το σκοπό αυτό από τη βάση δεδομένων της Genbank, συλλέχθηκαν οι πληροφορίες για κάθε έναν από τους 938 εντεροϊούς σε ένα αρχείο.

Στη συνέχεια, με τη βοήθεια των δύο custom script, δημιουργήθηκαν τα annotation files για κάθε φυλογενετικό δέντρο. Η εργασία που κλήθηκαν να διεκπεραιώσουν τα δύο προαναφερθέντα custom scripts ήταν η συνένωση των διαθέσιμων πληροφοριών για κάθε ένα από τα 938 στελέχη εντεροϊών με τα αποτελέσματα του Blastscript για το κάθε στέλεχος.

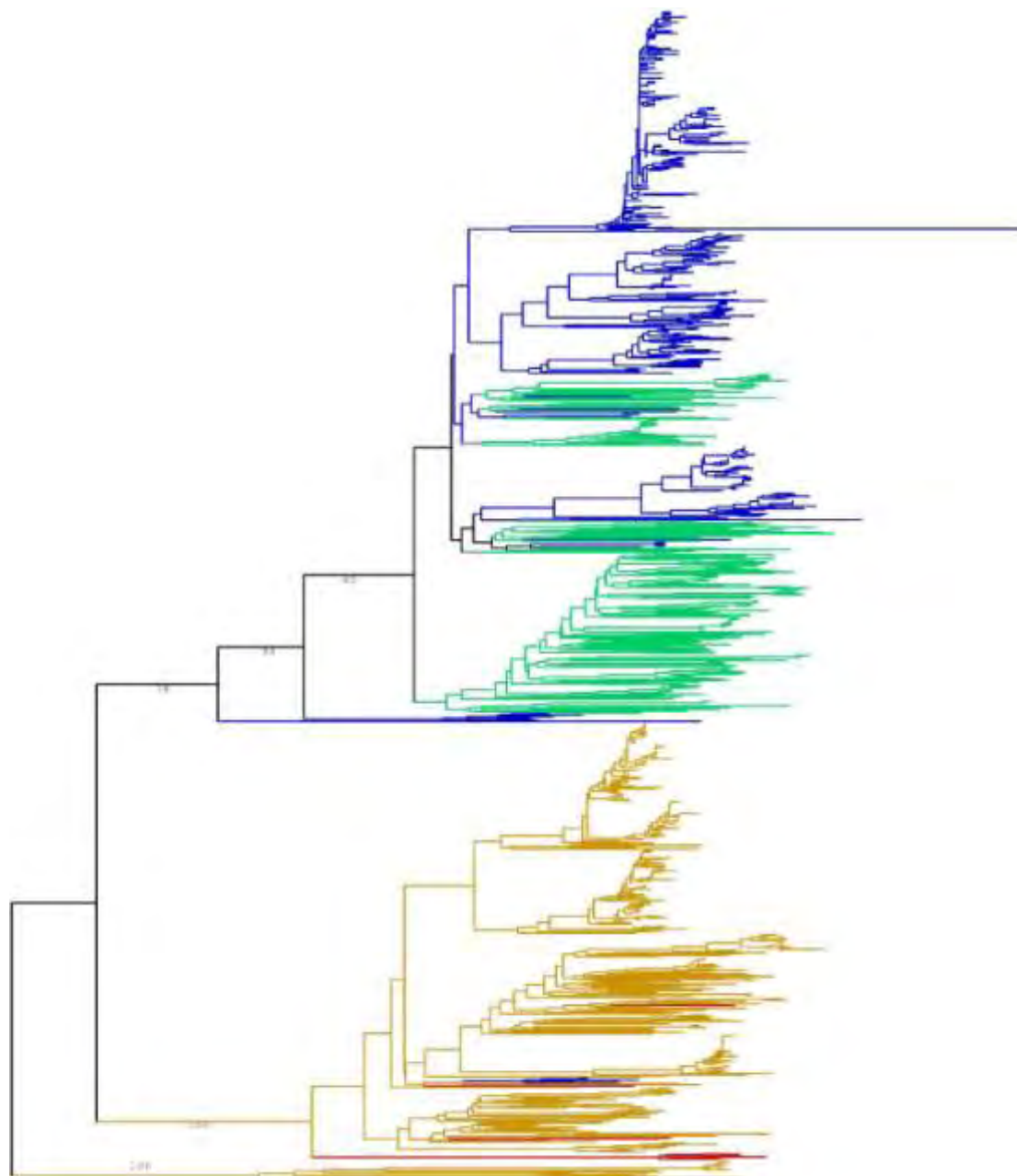
Έτσι, κατασκευάστηκαν συνολικά 11 αρχεία σχολιασμού (ένα για κάθε περιοχή), τα οποία είχαν την επέκταση (format) «.tlf». Η τελική μορφή ενός αρχείου σχολιασμού όπως παρουσιάστηκε στον πίνακα 4, περιλαμβάνει δύο σειρές των οποίων τα ονόματα είναι «938_to_938» και «938_to_39995» αντίστοιχα. Πρόκειται για τις δύο τελευταίες στήλες καθενός από τα αρχεία σχολιασμού που δημιουργήθηκαν και καθεμιά από αυτές περιλαμβάνει τους οροτύπους των στελεχών με τους οποίους, σύμφωνα με το Blastscript, ενδέχεται να έχει ανασυνδυαστεί κάποιο στέλεχος από τα υπό μελέτη. Η διαφορά τους είναι στη βάση δεδομένων που χρησιμοποιήθηκε κατά το blast. Αντίστοιχα λοιπόν η στήλη 938_to_938 αντιπροσωπεύει τα blast με βάση δεδομένων τα 938 γονιδιώματα ενώ η 938_to_39995 εκείνα με βάση δεδομένων τις 39995 ακολουθίες.

Σημειώνεται πως αν για κάποιο από τα υπό μελέτη στελέχη δεν υπάρχει αποτέλεσμα πιθανού ανασυνδυασμού από το Blastscript, τότε η αντίστοιχη στήλη θα περιέχει ένα «x» στη γραμμή που αφορά το συγκεκριμένο στέλεχος. Ο ίδιος χαρακτήρας περιέχεται επίσης και σε άλλες στήλες για τις οποίες δεν ήταν διαθέσιμα τα συγκεκριμένα στοιχεία.

3.9 Απεικόνιση των φυλογενετικών δέντρων

Στη συνέχεια τόσο τα 11 φυλογενετικά δέντρα πρωτεϊνών (ένα για κάθε περιοχή) και το φυλογενετικό δέντρο νουκλεοτιδικών ακολουθιών της περιοχής 5'UTR, όσο και τα 44 (ένα για κάθε είδος –A, B, C και D- για κάθε περιοχή εκτός της 5'UTR) που προέκυψαν από το Phyml, περάστηκαν στο πρόγραμμα TreeDyn ώστε να καταστεί εύκολη η επεξεργασία τους. Στα 44 δέντρα προστέθηκε σε κάθε περιοχή το αντίστοιχο αρχείο σχολιασμού. Στις επόμενες 11 εικόνες απεικονίζονται τα 11 φυλογενετικά δέντρα όπου με διαφορετικό χρώμα φαίνεται το καθένα από τα τέσσερα είδη. Η VPg έμεινε εκτός διότι ήταν πολύ μικρή. Αντιστοιχία χρωμάτων:

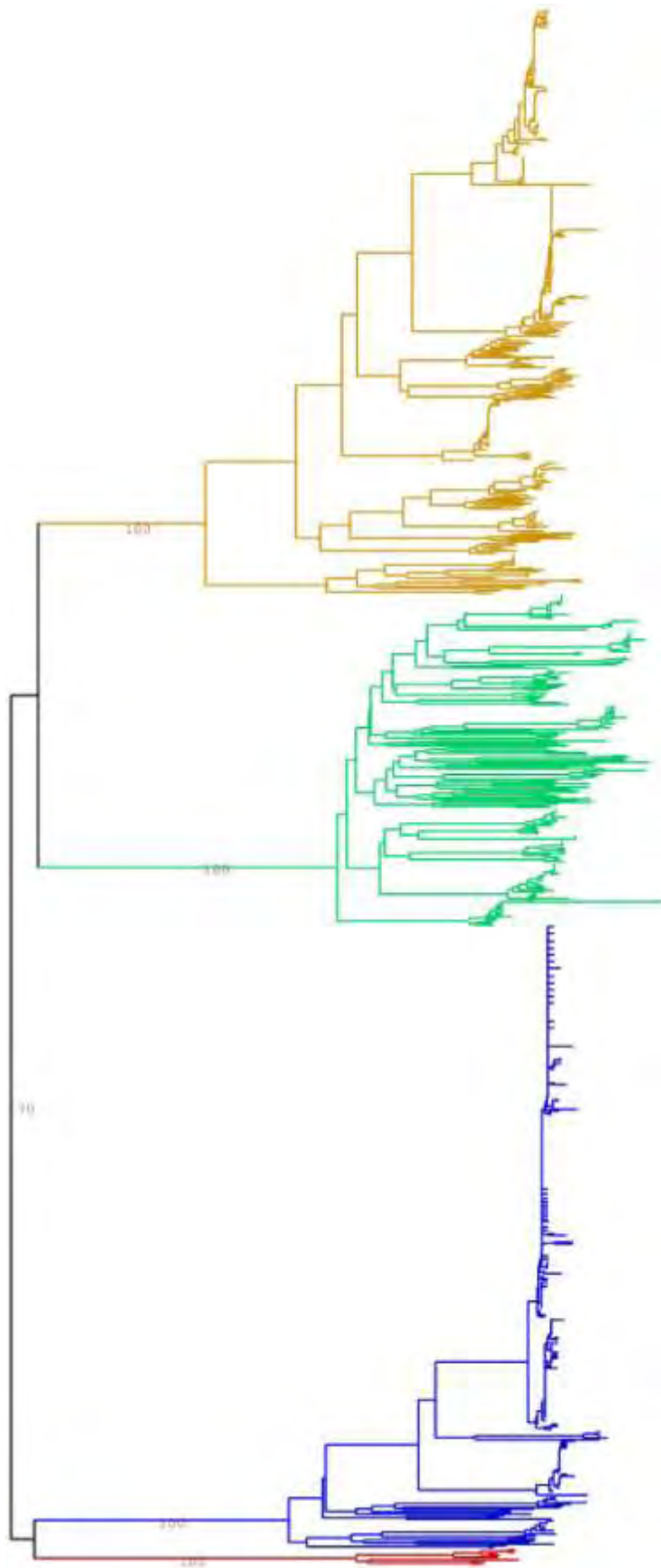
Μπλε - EVA, Πράσινο - EVB, Κίτρινο - EVC, Κόκκινο - EVD.



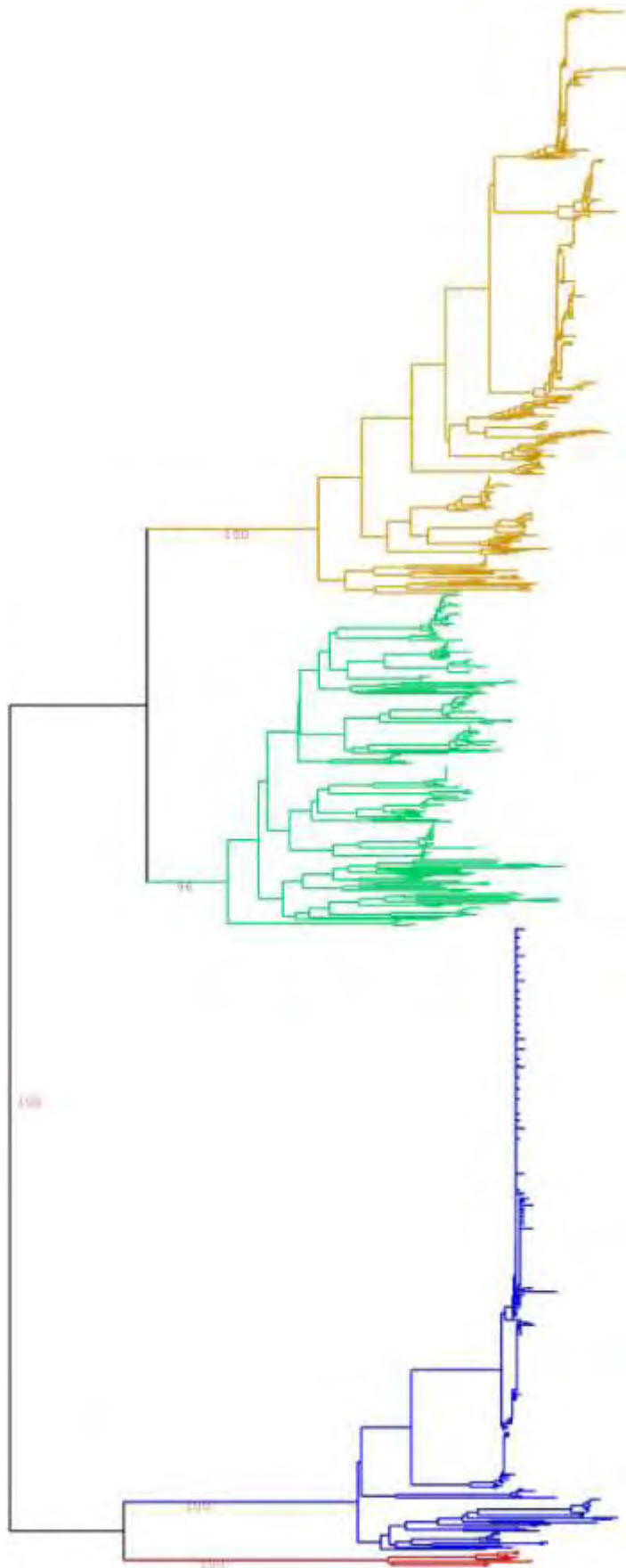
Εικόνα 11: Φυλογενετικό δέντρο της περιοχής 5'UTR. Μέθοδος NJ, Μοντέλο JC, bootstrap 500



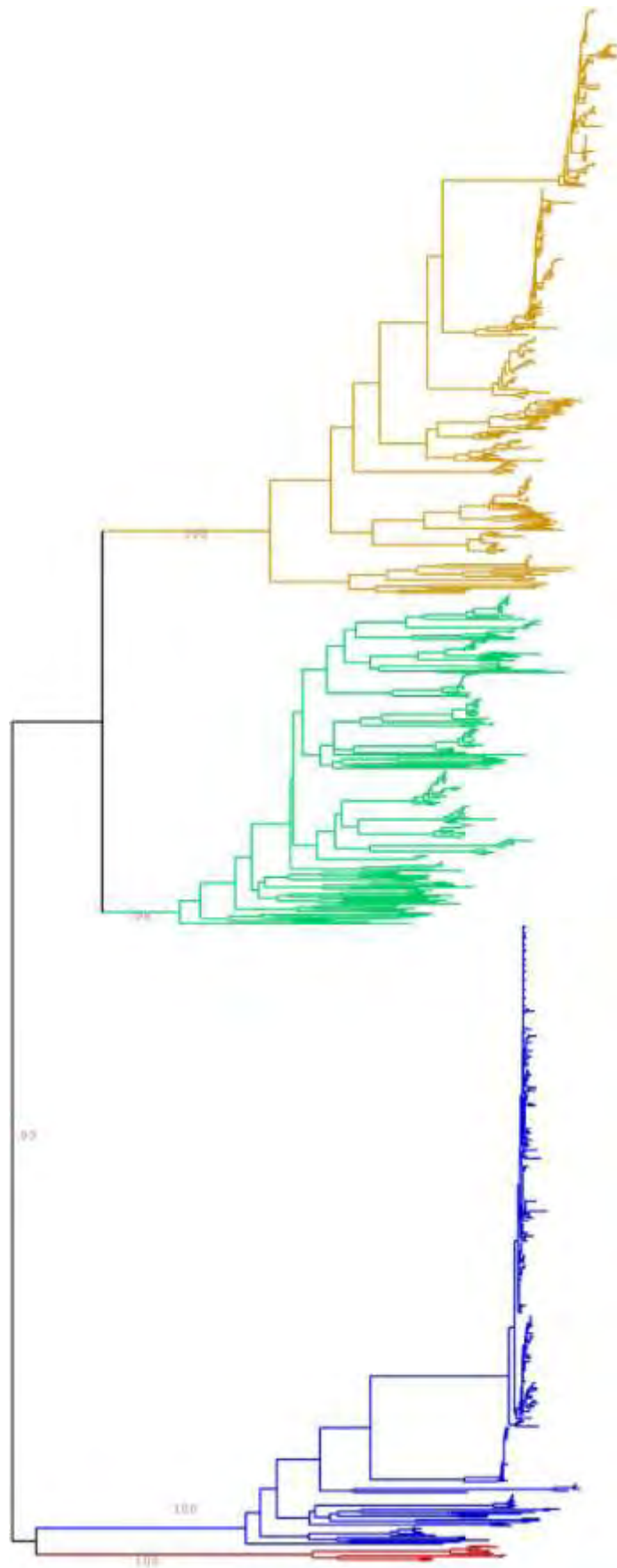
Εικόνα 21: Φυλογενετικό δέντρο της περιοχής VP4. Μέθοδος NJ, Μοντέλο K2P, bootstrap 500



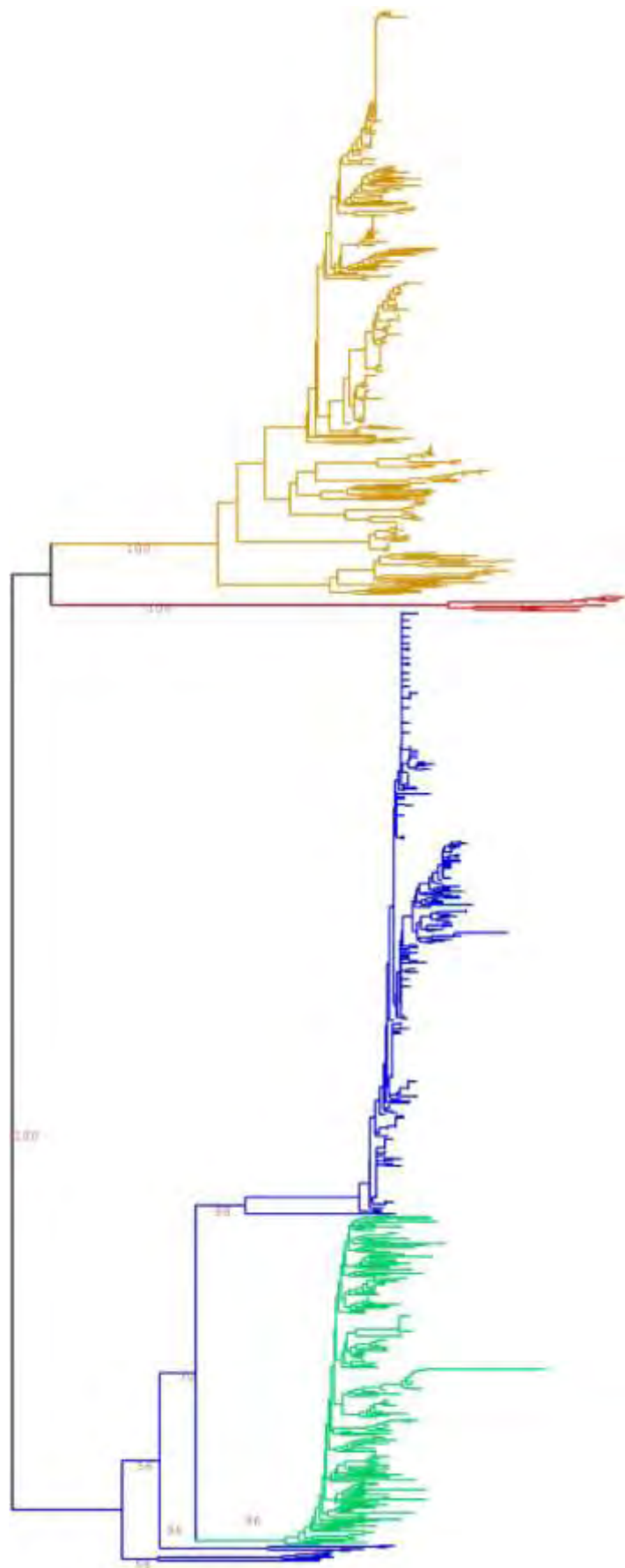
Εικόνα 19: Φυλογενετικό δέντρο της περιοχής VP2. Μέθοδος NJ, Μοντέλο K2P, bootstrap 500



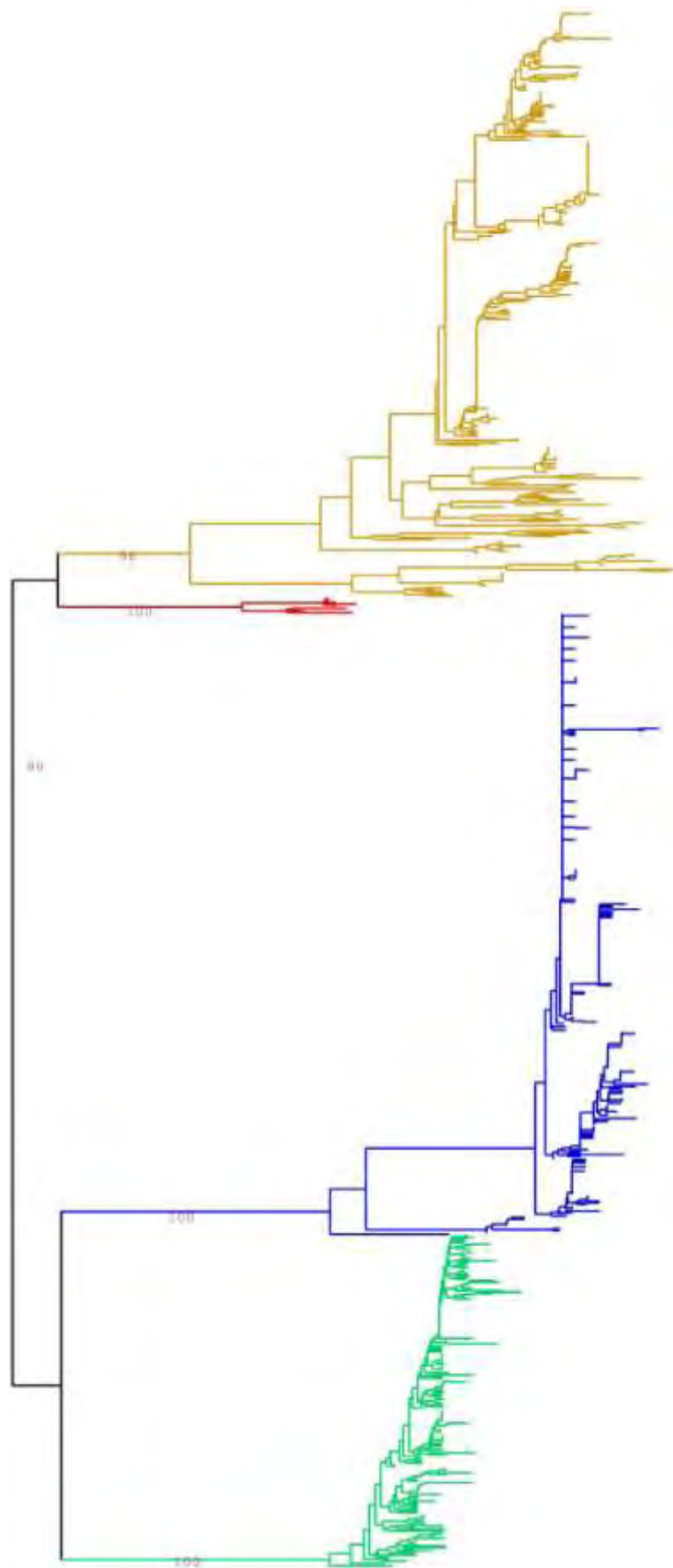
Εικόνα 20: Φυλογενετικό δέντρο της περιοχής VP3. Μέθοδος NJ, Μοντέλο K2P, bootstrap 500



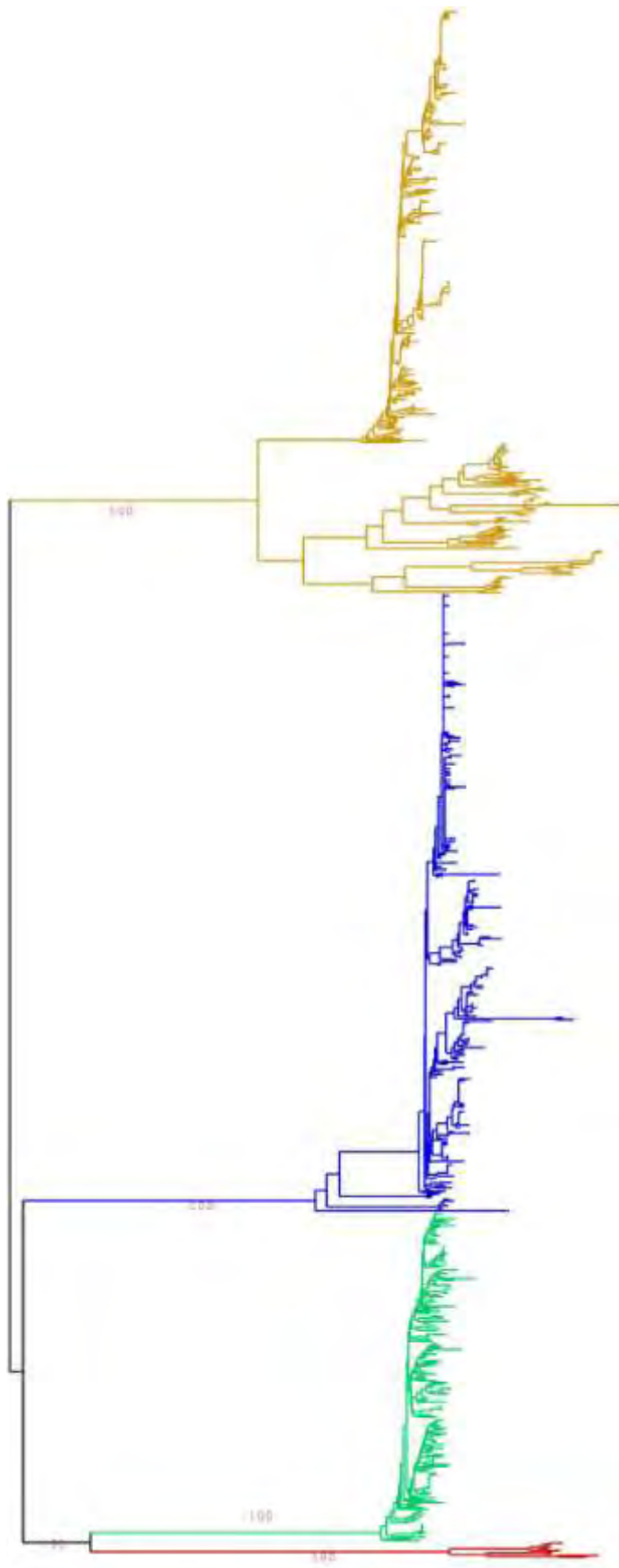
Εικόνα 18: Φυλογενετικό δέντρο της περιοχής VP1. Μέθοδος NJ, Μοντέλο K2P, bootstrap 500



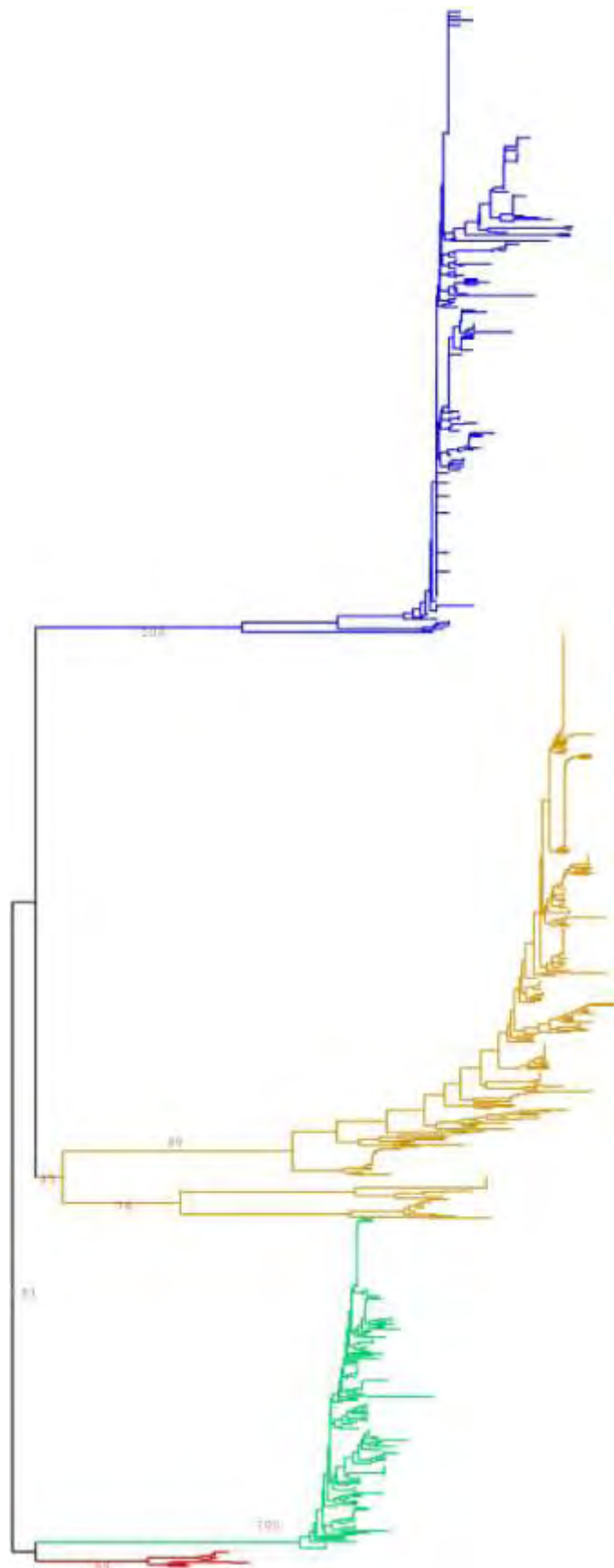
Εικόνα 12: Φυλογενετικό δέντρο της περιοχής P2A. Μέθοδος NJ, Μοντέλο K2P, bootstrap 500



Εικόνα 13: Φυλογενετικό δέντρο της περιοχής P2B. Μέθοδος NJ, Μοντέλο K2P, bootstrap 500



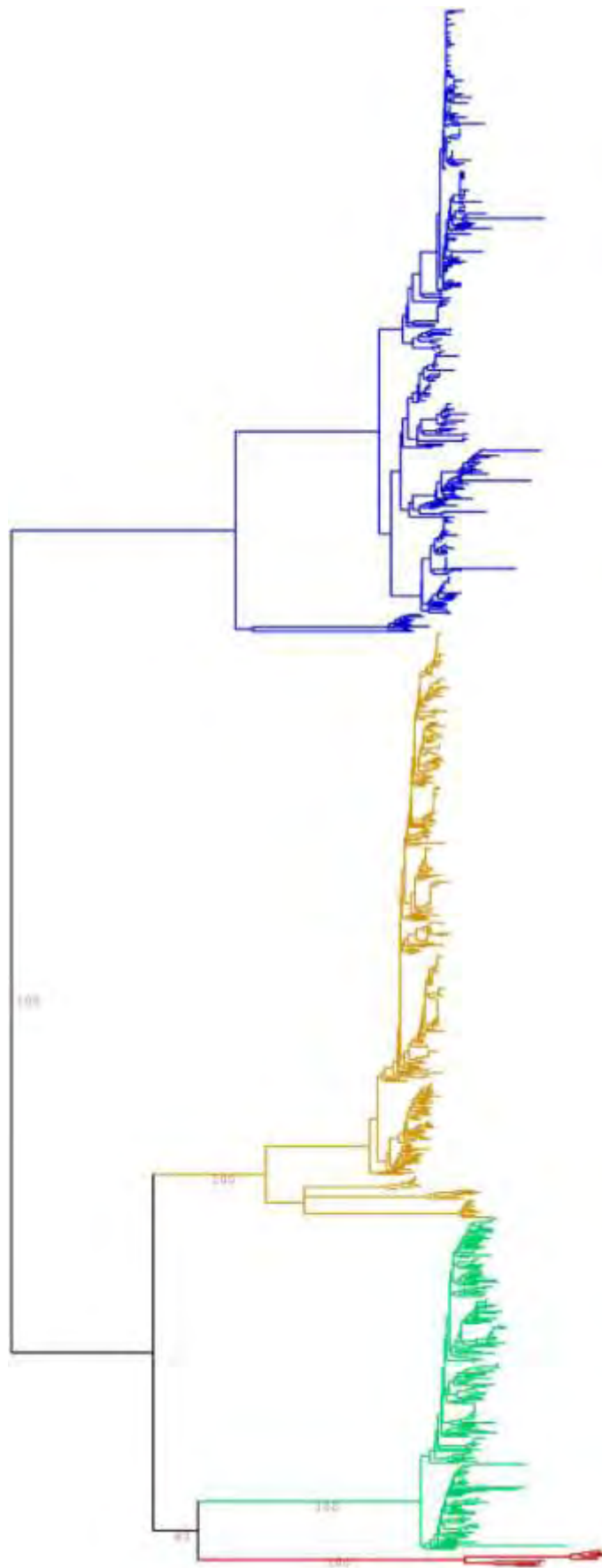
Εικόνα 14: Φυλογενετικό δέντρο της περιοχής P2C. Μέθοδος NJ, Μοντέλο K2P, bootstrap 500



Εικόνα 15: Φυλογενετικό δέντρο της περιοχής P3A. Μέθοδος NJ, Μοντέλο K2P, bootstrap 500

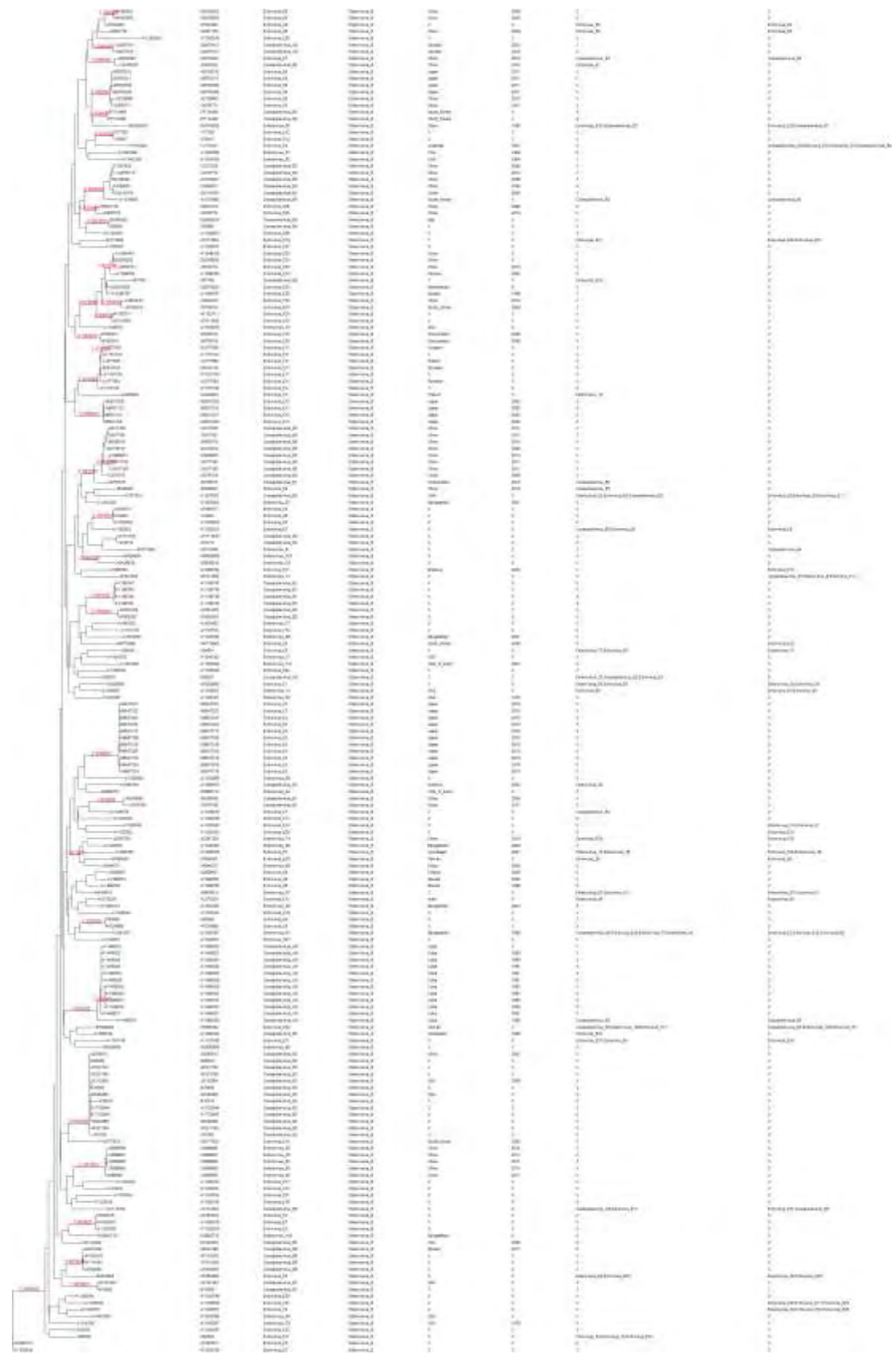


Εικόνα 16: Φυλογενετικό δέντρο της περιοχής P3C. Μέθοδος NJ, Μοντέλο K2P, bootstrap 500



Εικόνα 17: Φυλογενετικό δέντρο της περιοχής P3D. Μέθοδος NJ, Μοντέλο K2P, bootstrap 500

Στη συνέχεια παρουσιάζεται ένα από τα 44 φυλογενετικά δέντρα (συγκεκριμένα το δέντρο με τους EVB για την περιοχή P2A) ενώ παρακάτω αναλύεται το αρχείο σχολιασμού όπως αυτό φαίνεται πάνω στο συγκεκριμένο φυλογενετικό δέντρο.





Εικόνα 22: Αναλυτικά η μορφή του δέντρου και του αρχείου σχολιασμού. **(1)** Περιοχή σχεδίασης του φυλογενετικού δέντρου, **(2)** Accession number του στελέχους, **(3)** Ορότυπος του στελέχους, **(4)** Είδος στο οποίο ανήκει το στέλεχος, **(5)** Χώρα προέλευσης του στελέχους, **(6)** Έτος απομόνωσης του στελέχους, **(7)** Αποτελέσματα του Blastscript με βάση δεδομένων τα 938 γονιδιώματα, **(8)** Αποτελέσματα του Blastscript με βάση δεδομένων τις 39995 ακολουθίες εντεροϊών.

3.10 Παρατήρηση μονοφυλετικότητας

Ο σκοπός της κατασκευής των 11 φυλογενετικών δέντρων από τις 938 πρωτεϊνικές ακολουθίες καθώς και του φυλογενετικού δέντρου της 5' UTR, ήταν η παρατήρηση της μονοφυλετικότητας μεταξύ των ειδών. Οι παρατηρήσεις που έγιναν ήταν οι ακόλουθες:

Οι εντεροϊοί στην περιοχή 5' UTR χωρίζονται σε δύο μεγάλες ομάδες στις οποίες δεν υπάρχει ξεκάθαρα μονοφυλετικότητα. Στην πρώτη μεγάλη ομάδα περιλαμβάνονται οι Enterovirus A σε ανάμιξη με τους Enterovirus B. Στη δεύτερη ομάδα παρατηρείται παρόμοια κατάσταση μεταξύ των C & D, ενώ φαίνεται να περιλαμβάνονται σε αυτή και κάποια στελέχη από τους Enterovirus A (Εικόνα 11 – Φυλογενετικό δέντρο της περιοχής 5' UTR).

Στην περιοχή της P2A παρατηρείται μονοφυλετικότητα μόνο στους Enterovirus A-B, C και D. Αντίθετα οι Enterovirus A εμφανίζονται παραφυλετικοί, με τους B να εξέρχονται μέσα από τους A (Εικόνα 12 – Φυλογενετικό δέντρο της περιοχής P2A).

Στις περιοχές P2B, P2C, P3A, P3C, P3D, VP1, VP2 και VP3 διακρίνεται ξεκάθαρα μονοφυλετικότητα ανάμεσα στα τέσσερα είδη των εντεροϊών (εικόνες 13 έως 20)

Τέλος, στην περιοχή VP4 παρατηρείται παρόμοια κατάσταση με εκείνη της περιοχής P2A, αλλά αυτή τη φορά οι Enterovirus B φαίνεται να εξέρχονται μέσα από τους Enterovirus C (Εικόνα 21 – Φυλογενετικό δέντρο περιοχής VP4).

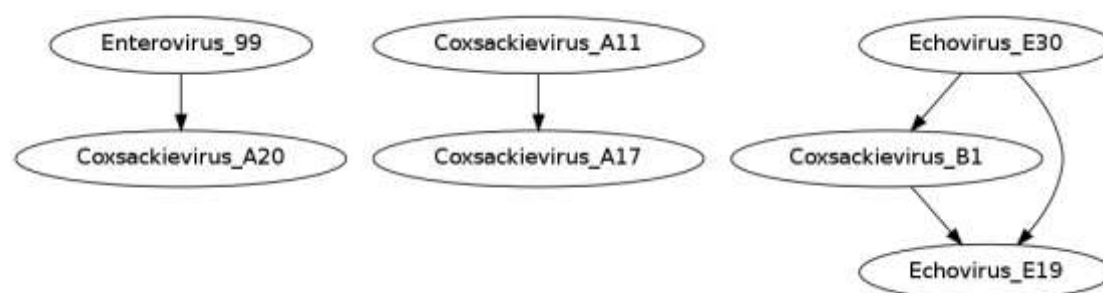
Τα αποτελέσματα συμφώνησαν με την υπάρχουσα βιβλιογραφία καθώς υπάρχουν σποραδικές αναφορές. Αυτή είναι όμως η πρώτη φορά που τέτοιου είδους ανάλυση

γίνεται με τέτοια πληθώρα δεδομένων και ουσιαστικά επιβεβαιώνει ξεκάθαρα τις σποραδικές αναφορές που βασίζονταν σε λίγα δεδομένα.

3.11 Κατασκευή δικτύων δοτών – δεκτών για περιπτώσεις πιθανών γεγονότων ανασυνδυασμού βάσει του Blastscript

Στη συνέχεια της ανάλυσης κατασκευάστηκαν για κάθε περιοχή δίκτυα στα οποία παρουσιάζονται οι δότες και οι δέκτες σε πιθανά γεγονότα ανασυνδυασμού βάσει των αποτελεσμάτων του Blastscript. Πρόκειται για εικόνες που αποδίδουν τους πιθανούς ανασυνδυασμούς μεταξύ των διαφορετικών οροτύπων που ανήκουν στο ίδιο είδος και παριστάνονται με βέλη. Με τη βοήθεια των δικτύων είναι δυνατό να παρατηρηθεί πόσοι διαφορετικοί δέκτες έχουν ανασυνδυαστεί με έναν πιθανό δότη.

Τέτοιου είδους δίκτυα κατασκευάστηκαν για κάθε μια από τις 11 περιοχές με τη χρήση του GraphViz που προσδίδει τη δυνατότητα δημιουργίας δικτύων μέσω της γλώσσας προγραμματισμού Perl. Στην εικόνα 23 παρουσιάζεται το δίκτυο για την περιοχή VP4.



Εικόνα 23: Δίκτυο δοτών – δεκτών πιθανών γεγονότων ανασυνδυασμού για την περιοχή VP4. Τα βελάκια ξεκινούν από τον πιθανό δότη και καταλήγουν στον πιθανό δέκτη

3.12 Γραφήματα τάσεων ανασυνδυασμών ανά περιοχή

Στο τελευταίο στάδιο της ανάλυσης δημιουργήθηκαν γραφήματα στο πρόγραμμα excel στα οποία αναπαραστάθηκαν οι τάσεις με τις οποίες συνέβαιναν τα γεγονότα ανασυνδυασμού σε κάθε περιοχή για κάθε είδος εντεροϊών.

Για λόγους αξιοπιστίας των αποτελεσμάτων, από κάθε είδος μελετήθηκαν μόνο οι ορότυποι των οποίων οι διαθέσιμες ακολουθίες ήταν πάνω από 10. Στους EVD οι συνολικές διαθέσιμες ακολουθίες ήταν εξ' αρχής 10 και συνεπώς δεν ήταν δυνατόν

να χρησιμοποιηθεί κάποιος από τους οροτύπους αυτού του είδους για τη συγκεκριμένη ανάλυση.

Στον πίνακα 8 παρουσιάζονται τα δεδομένα που χρησιμοποιήθηκαν, ύστερα από κανονικοποίηση, για τη δημιουργία κάθε γραφήματος.

	P2A	P2B	P2C	P3A	P3C	P3D	VP1	VP2	VP3	VP4	Vpg	Seqs	M.O.
EVA													
Enterovirus_71	2	3	2	5	6	6	1	1	1	1	3	302	2,8182
Coxsackievirus_A16	2	2	2	2	2	2	1	1	1	1	2	34	1,6364
EVB													
Echovirus_E6	4	5	8	10	9	8	1	1	1	3	9	15	5,3636
Coxsackievirus_A9	3	3	6	6	5	4	1	1	1	3	5	15	3,4545
Coxsackievirus_B3	5	6	8	8	8	8	1	1	1	4	8	22	5,2727
Coxsackievirus_B5	4	4	5	6	6	7	1	1	1	4	7	11	4,1818
Echovirus_E30	3	8	10	10	10	10	1	1	1	2	10	14	6
Echovirus_E11	5	5	5	4	7	7	1	1	1	3	8	11	4,2727
Echovirus_E3	3	3	3	3	3	3	1	1	1	1	3	13	2,2727
EVC													
Poliovirus_2	7	6	8	10	7	7	1	1	1	1	12	91	5,5455
Poliovirus_3	4	4	7	9	9	9	1	1	1	1	8	37	4,9091
Poliovirus_1	6	5	8	9	9	11	1	1	1	1	14	107	6
Coxsackievirus_A13	3	1	2	4	6	8	1	1	1	2	7	16	3,2727
Coxsackievirus_A24	1	1	1	1	2	4	1	1	1	4	5	22	2
Coxsackievirus_A21	1	2	1	3	3	4	1	1	1	1	3	14	1,9091

Πίνακας 8: Τα δεδομένα που χρησιμοποιήθηκαν για την κατασκευή των γραφημάτων με τις τάσεις ανασυνδυασμού σε κάθε περιοχή. Οι στήλες «P2A» έως και «VPg» στον πίνακα 8 δείχνουν τον αριθμό των μονοφυλετικών ομάδων στις οποίες χωρίζεται ο κάθε ορότυπος του πίνακα σύμφωνα με το φυλογενετικό δέντρο κάθε περιοχής. Η στήλη «seqs» υποδηλώνει τον αριθμό των διαθέσιμων ακολουθιών κάθε οροτύπου ανάμεσα στις 938 γονιδιωματικές ακολουθίες ενώ η στήλη «M.O.» αφορά τον μέσο όρο των ομάδων που φαίνεται να χωρίζεται ο κάθε ορότυπος του πίνακα σε κάθε περιοχή.

Οι στήλες «P2A» έως και «VPg» στον πίνακα 8 δείχνουν τον αριθμό των μονοφυλετικών ομάδων στις οποίες χωρίζεται ο κάθε ορότυπος του πίνακα σύμφωνα με το φυλογενετικό δέντρο κάθε περιοχής. Η στήλη «seqs» υποδηλώνει τον αριθμό των διαθέσιμων ακολουθιών κάθε οροτύπου ανάμεσα στις 938 γονιδιωματικές ακολουθίες ενώ η στήλη «M.O.» αφορά τον μέσο όρο των ομάδων που φαίνεται να χωρίζεται ο κάθε ορότυπος του πίνακα σε κάθε περιοχή.

Αφού συγκεντρώθηκαν τα παραπάνω δεδομένα, έγινε κανονικοποίηση σε αυτά ώστε να καταστεί δυνατή η επεξεργασία τους. Για κάθε τιμή, η κανονικοποίηση έγινε διαιρώντας τον αριθμό των ομάδων ενός οροτύπου σε μια περιοχή με τον

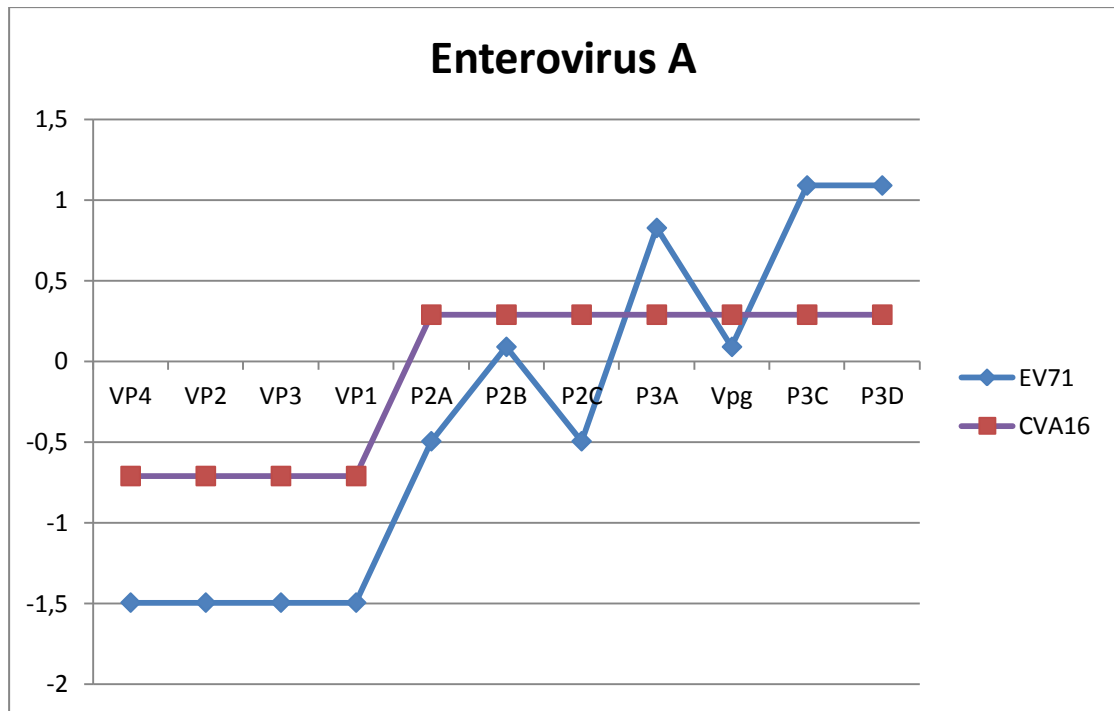
μέσο όρο των ομάδων για τον συγκεκριμένο ορότυπο και το αποτέλεσμα της διαίρεσης αυτής μετατράπηκε σε λογάριθμο με βάση το 2.

Στη συνέχεια παρατίθενται τα κανονικοποιημένα δεδομένα στον πίνακα 9.

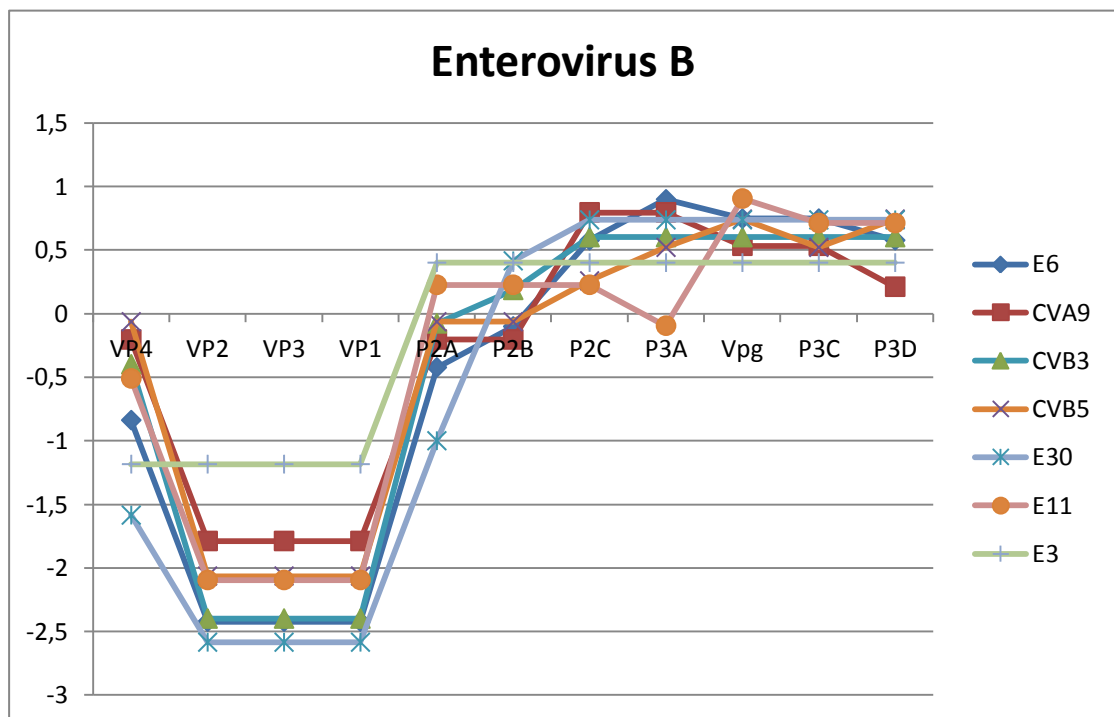
	VP4	VP2	VP3	VP1	P2A	P2B	P2C	P3A	Vpg	P3C	P3D
EVA											
Enterovirus_71	-1,49	-1,49	-1,49	-1,49	-0,49	0,09	-0,49	0,83	0,09	1,09	1,09
Coxsackievirus_A16	-0,71	-0,71	-0,71	-0,71	0,29	0,29	0,29	0,29	0,29	0,29	0,29
EVB											
Echovirus_E6	-0,84	-2,42	-2,42	-2,42	-0,42	-0,10	0,58	0,90	0,75	0,75	0,58
Coxsackievirus_A9	-0,20	-1,79	-1,79	-1,79	-0,20	-0,20	0,80	0,80	0,53	0,53	0,21
Coxsackievirus_B3	-0,40	-2,40	-2,40	-2,40	-0,08	0,19	0,60	0,60	0,60	0,60	0,60
Coxsackievirus_B5	-0,06	-2,06	-2,06	-2,06	-0,06	-0,06	0,26	0,52	0,74	0,52	0,74
Echovirus_E30	-1,58	-2,58	-2,58	-2,58	-1,00	0,42	0,74	0,74	0,74	0,74	0,74
Echovirus_E11	-0,51	-2,10	-2,10	-2,10	0,23	0,23	0,23	-0,10	0,90	0,71	0,71
Echovirus_E3	-1,18	-1,18	-1,18	-1,18	0,40	0,40	0,40	0,40	0,40	0,40	0,40
EVC											
Poliovirus_2	-2,47	-2,47	-2,47	-2,47	0,34	0,11	0,53	0,85	1,11	0,34	0,34
Poliovirus_3	-2,30	-2,30	-2,30	-2,30	-0,30	-0,30	0,51	0,87	0,70	0,87	0,87
Poliovirus_1	-2,58	-2,58	-2,58	-2,58	0,00	-0,26	0,42	0,58	1,22	0,58	0,87
Coxsackievirus_A13	-0,71	-1,71	-1,71	-1,71	-0,13	-1,71	-0,71	0,29	1,10	0,87	1,29
Coxsackievirus_A24	1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	-1,00	1,32	0,00	1,00
Coxsackievirus_A21	-0,93	-0,93	-0,93	-0,93	-0,93	0,07	-0,93	0,65	0,65	0,65	1,07

Πίνακας 9: Τα δεδομένα που χρησιμοποιήθηκαν για την κατασκευή των γραφημάτων με τις τάσεις ανασυνδυασμού σε κάθε περιοχή ύστερα από κανονικοποίηση. Για κάθε τιμή, η κανονικοποίηση έγινε διαιρώντας τον αριθμό των ομάδων ενός οροτύπου σε μια περιοχή με τον μέσο όρο των ομάδων για τον συγκεκριμένο ορότυπο και το αποτέλεσμα της διαίρεσης αυτής μετατράπηκε σε λογάριθμο με βάση το 2.

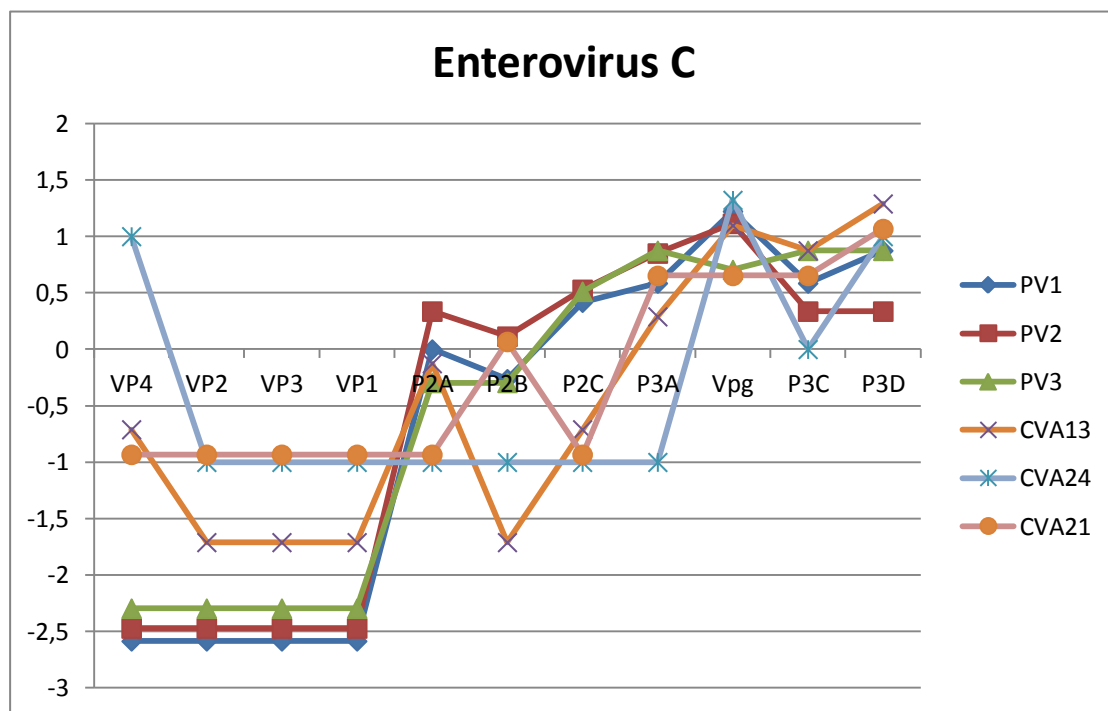
Έπειτα, βάσει του πίνακα 9 δημιουργήθηκαν τρία γραφήματα (ένα για κάθε είδος εντεροϊών) στα οποία αναπαρίστανται γραφικά οι τάσεις ανασυνδυασμού για κάθε είδος ανάμεσα σε όλες τις περιοχές. Τα γραφήματα αυτά παρουσιάζονται στις εικόνες 24, 25 και 26.



Εικόνα 24: Γράφημα τάσεων ανασυνδυασμού ανά περιοχή στους Enterovirus A



Εικόνα 25: Γράφημα τάσεων ανασυνδυασμού ανά περιοχή στους Enterovirus B



Εικόνα 26: Γράφημα τάσεων ανασυνδυασμού ανά περιοχή στους Enterovirus C

Συμπεράσματα

Η παρατήρηση των παραπάνω γραφημάτων οδηγεί στο συμπέρασμα ότι η τάση ανασυνδυασμού στις περιοχές VP2, VP3 και VP1 είναι μειωμένες. Αντίθετα, στις περιοχές P2A, P2B, P2C, P3A, Vpg, P3C και P3D η τάση ανασυνδυασμού είναι πιο αυξημένη ενώ μια μέτρια κατάσταση με σχετικά αυξημένη τάση ανασυνδυασμού παρατηρείται στην περιοχή VP4.

4 Συζήτηση

4.1 Μονοφυλετικότητα

Η ανάλυση έδειξε ότι στην περιοχή 5'UTR εμφανίζονται δύο μεγάλες ομάδες εντεροϊών στις οποίες δεν υπάρχει ξεκάθαρη μονοφυλετικότητα. Στην πρώτη ομάδα περιλαμβάνονται οι Enterovirus A διεσπαρμένοι μαζί με τους Enterovirus B, ενώ στη δεύτερη παρατηρείται παρόμοια κατάσταση μεταξύ των Enterovirus C & D. Επίσης στη δεύτερη ομάδα εμφανίζονται και κάποια στελέχη από τους Enterovirus A. Στην περιοχή P2A παρατηρείται μονοφυλετικότητα μόνο στους Enterovirus A-B, C και D. Αντίθετα οι Enterovirus A εμφανίζονται παραφυλετικοί, με τους B να εξέρχονται μέσα από τους A. Παρόμοια περίπτωση αποτελεί το φυλογενετικό δέντρο της περιοχής VP4 όπου οι Enterovirus B φαίνεται να εξέρχονται μέσα από τους Enterovirus C. Στις περιοχές P2B, P2C, P3A, P3C, P3D, VP1, VP2 και VP3 διακρίνεται ξεκάθαρη μονοφυλετικότητα ανάμεσα στα τέσσερα είδη των εντεροϊών.

4.2 Φυλογενετικά δέντρα – Περιπτώσεις ανασυνδυασμών

Η κατασκευή των 44 φυλογενετικών δέντρων καθώς και η προσθήκη των αρχείων σχολιασμού σε αυτά οδήγησε στην αξιολόγηση των δέντρων όσον αφορά περιπτώσεις πιθανών ανασυνδυασμών μεταξύ διαφορετικών οροτύπων. Για κάθε ένα από τα 44 δέντρα συγκεντρώθηκαν περιπτώσεις που αξιολογήθηκαν ως πιθανά γεγονότα ανασυνδυασμού σε αρχεία excel.

Κατά την αξιολόγηση των δέντρων, η αναζήτηση πιθανών γεγονότων ανασυνδυασμού έγινε με βάση 3 κριτήρια:

Ως περίπτωση πιθανού ανασυνδυασμού χαρακτηριζόταν εκείνη όπου:

- Το υπό μελέτη στέλεχος βρισκόταν στο φυλογενετικό δέντρο στον ίδιο κλάδο με πολλά άλλα στελέχη τα οποία είχαν τον ίδιο ορότυπο μεταξύ τους αλλά διαφορετικό ορότυπο από το υπό μελέτη στέλεχος. Για να ληφθεί μια τέτοια περίπτωση υπ' όψιν θα έπρεπε και η τιμή aLRT του κλάδου που περιλάμβανε όλα αυτά τα στελέχη να είναι μεγαλύτερη από 0,95.
- Η στήλη 938_to_938 του αρχείου σχολιασμού έδινε κάποιο αποτέλεσμα ανασυνδυασμού για τον υπό μελέτη ορότυπο.
- Η στήλη 938_to_39995 του αρχείου σχολιασμού έδινε κάποιο αποτέλεσμα ανασυνδυασμού για τον υπό μελέτη ορότυπο.

Τα παραπάνω κριτήρια μεμονωμένα ήταν εφικτό να χαρακτηρίσουν την ύπαρξη πιθανότητας ανασυνδυασμού του υπό μελέτη στελέχους. Σε περιπτώσεις όμως όπου ίσχυαν και τα τρία κριτήρια ταυτόχρονα και ειδικά όταν συμφωνούσαν στα αποτελέσματά τους οι στήλες 938_to_938 και 938_to_39995 του αρχείου σχολιασμού, τότε η υπόθεση για περίπτωση ανασυνδυασμού ήταν σχεδόν βέβαιη.

Όλες οι παρατηρήσεις που βρίσκονται στα αρχεία excel περιλαμβάνουν τις παρακάτω στήλες:

A/N (Accession Number): Κωδικός του πιθανώς ανασυνδυασμένου στελέχους

Μέσα Σε: Σε αυτή τη στήλη αναγράφεται ο ορότυπος δότης με τον οποίο φαίνεται να έχει ανασυνδυαστεί το υπό μελέτη στέλεχος

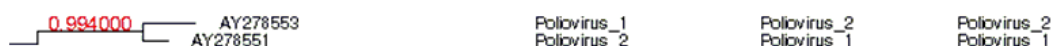
Δέντρο: Η στήλη αυτή συμπληρώνεται με «ΝΑΙ» ή «ΟΧΙ» και αφορά το αν υποστηρίζεται η υπόθεση ανασυνδυασμού από το φυλογενετικό δέντρο

938-938: Η στήλη αυτή συμπληρώνεται με «ΝΑΙ» ή «ΟΧΙ» και αφορά το αν υποστηρίζεται η υπόθεση ανασυνδυασμού από τη στήλη 938_to_938 του αρχείου σχολιασμού

938-39995: Η στήλη αυτή συμπληρώνεται με «ΝΑΙ» ή «ΟΧΙ» και αφορά το αν υποστηρίζεται η υπόθεση ανασυνδυασμού από τη στήλη 938_to_39995 του αρχείου σχολιασμού

Σχόλια: Σε αυτή τη στήλη αναγράφεται ο ορότυπος του υπό μελέτη στελέχους

Στην εικόνα 27 φαίνεται μια καθαρή περίπτωση ανασυνδυασμού. Η εικόνα αυτή αποτελεί κομμάτι του φυλογενετικού δέντρου των Enterovirus C για την περιοχή P3A και έχει τροποποιηθεί κατάλληλα ώστε από το αρχείο σχολιασμού να φαίνονται μόνο οι στήλες με τον ορότυπο των στελεχών που βρίσκονται στο δέντρο και οι δύο στήλες με τα αποτελέσματα του Blastsript (938_to_938 και 938_to_39995).



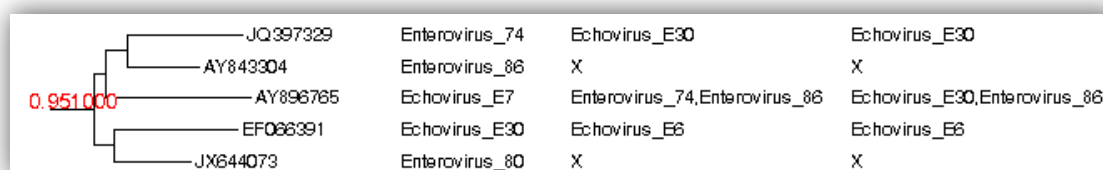
Εικόνα 27: Περίπτωση ανασυνδυασμού όπου ισχύουν και τα 3 κριτήρια στο φυλογενετικό δέντρο για τους Enterovirus C στην περιοχή P3A. Αριστερά, η περιοχή σχεδίασης του δέντρου. Δεξιά, οι 3 στήλες από το annotation file με τη σειρά που παρουσιάζονται: Ορότυπος στελέχους του δέντρου, αποτέλεσμα Blastsript 938_to_938, αποτέλεσμα Blastsript 938_to_39995.

4.3 Προβλήματα στην αξιολόγηση των φυλογενετικών δέντρων

Σε αρκετές περιπτώσεις κατά την αξιολόγηση των φυλογενετικών δέντρων παρουσιάστηκαν ορισμένα προβλήματα. Στις επόμενες εικόνες παρατίθενται τα προβλήματα αυτά.

Πρόβλημα 1^ο – Ο δότης δεν είναι ξεκάθαρος

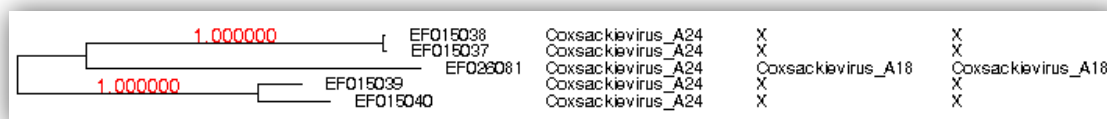
Στην περίπτωση που παρουσιάζεται στην εικόνα 28 και αφορά το φυλογενετικό δέντρο της περιοχής P2A για τους Enterovirus B παρατηρείται μια ομαδοποίηση ορισμένων στελεχών με διαφορετικούς οροτύπους όπου αν και σε 3 από αυτά τα στελέχη οι δύο στήλες του Blastscript συμφωνούν στα αποτελέσματά τους, η ακριβής ομαδοποίηση δεν είναι ξεκάθαρη. Κατά συνέπεια είναι πολύ δύσκολο να αξιολογηθούν αυτές οι περιπτώσεις όσον αφορά τον δότη.



Εικόνα 28: Φυλογενετικό δέντρο της περιοχής P2A για τους Enterovirus B. Πρόβλημα προσδιορισμού του δότη.

Πρόβλημα 2^ο – Η θέση του ανασυνδυασμού

Στην εικόνα 29 παρουσιάζεται μια περίπτωση προβλήματος που αφορά τη θέση όπου έχει γίνει ο ανασυνδυασμός και εντοπίζεται στο φυλογενετικό δέντρο της θέσης P3D για τους Enterovirus C. Ενώ οι στήλες του Blastscript στο αρχείο σχολιασμού υποστηρίζουν πως το στέλεχος EF026801 (Coxsackievirus_A24) έχει πιθανώς ανασυνδυαστεί με ένα στέλεχος του οποίου ο ορότυπος είναι Coxsackie_A18, το φυλογενετικό δέντρο το δείχνει ομαδοποιημένο με Coxsackie_A24. Υπενθυμίζεται ότι τα φυλογενετικά δέντρα κατασκευάστηκαν χρησιμοποιώντας ολόκληρες τις νουκλεοτιδικές ακολουθίες κάθε περιοχής, ενώ το Blastscript χρησιμοποιεί τμήματα αυτών των περιοχών και συνεπώς τα αποτελέσματά του αφορούν τα τμήματα αυτά. Έτσι, αν η θέση όπου έγινε ο ανασυνδυασμός βρίσκεται ανάμεσα σε 2 περιοχές τότε το Blastscript θα εντοπίσει τον ανασυνδυασμό ενώ το φυλογενετικό δέντρο όχι. Σχηματική αναπαράσταση του προβλήματος αυτού παρουσιάζεται στην εικόνα 30.



Εικόνα 29: Φυλογενετικό δέντρο της περιοχής P3D για τους Enterovirus C. Πρόβλημα με τη θέση του ανασυνδυασμού.

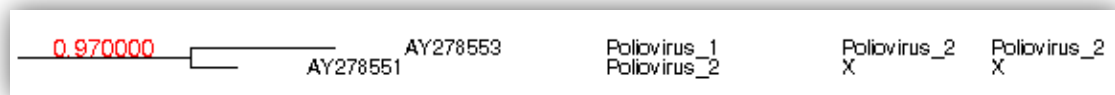


Εικόνα 30: Σχηματική απόδοση του προβλήματος θέσης του ανασυνδυασμού

Πρόβλημα 3^ο – Πρόβλημα με το ποσοστό ταύτισης

Όπως ήδη έχει αναφερθεί, μία από τις προϋποθέσεις για να συμπεριληφθεί ένα αποτέλεσμα του blast στα τελικά αποτελέσματα του Blastscript είναι το ποσοστό ταύτισης. Στην ανάλυση που περιγράφεται στην εργασία αυτή η διαφορά στο ποσοστό ταύτισης που χρησιμοποιήθηκε ήταν 5%. Αυτό σημαίνει ότι έγιναν δεκτά μόνο τα αποτελέσματα του blast όπου η διαφορά του ποσοστού ταύτισης, μεταξύ του ελεγχόμενου αποτελέσματος και του αμέσως καλύτερου αποτελέσματος που έχει ίδιο ορότυπο με αυτόν της ακολουθίας επερώτησης, ήταν μεγαλύτερη από 5%.

Στην εικόνα 31 εμφανίζεται το συγκεκριμένο πρόβλημα στο φυλογενετικό δέντρο της περιοχής P3A για τους Enterovirus C. Στην εικόνα, το φυλογενετικό δέντρο υποστηρίζει μια ομαδοποίηση μεταξύ ενός στελέχους Poliovirus 1 με ένα στέλεχος Poliovirus 2. Όσον αφορά το στέλεχος του Poliovirus 1, τα αποτελέσματα του Blastscript τον θέλουν να έχει ανασυνδυαστεί με το στέλεχος του Poliovirus 2 που φαίνεται στην εικόνα. Αντίθετα, το στέλεχος με ορότυπο Poliovirus 2 δεν εμφανίζει τον Poliovirus 1 στα αποτελέσματα του Blastscript όπως θα ήταν αναμενόμενο. Ύστερα από έρευνα στα αποτελέσματα του blast αποκαλύφθηκε ότι πράγματι το συγκεκριμένο στέλεχος Poliovirus 1 εντοπιζόταν ως πρώτο στα αποτελέσματα του blast με ποσοστό ταύτισης με την ακολουθία επερώτησης 89,5% αλλά το αμέσως καλύτερο blasthit με ορότυπο Poliovirus 2 είχε ποσοστό ταύτισης με την ακολουθία επερώτησης 87%. Η σύγκριση των δύο ποσοστών δείχνει ότι διαφέρουν κατά 2,5%, τιμή που δεν συμπεριλαμβάνεται στα αποδεκτά αποτελέσματα (αφού έχει οριστεί ως κατώφλι η τιμή 5%).



Εικόνα 31: Φυλογενετικό δέντρο της περιοχής P3A για τους Enterovirus C. Πρόβλημα με το ποσοστό ταύτισης

Βιβλιογραφία

- [1] Pallansch MA, Ross RP: Enteroviruses: polioviruses, coxsackieviruses, echoviruses, and newer enteroviruses. In Fields Virology Knipe DM, Howley PM, Griffin DE, et al., (eds). 4th ed. Lippincott Williams & Wilkins 2001; 723-775.
- [2] Brown B, Oberste MS, Maher K, Pallansch MA. Complete genomic sequencing shows that polioviruses and members of human enterovirus species C are closely related in the noncapsid coding region. *J Virol* 2003; 77: 8973-8984.
- [3] Smyth, M.S., & Martin, J.H. (2002). Picornavirus uncoating. *Mol Pathol.* 55(4), 214-9.
- [4] Oberste MS¹, Maher K, Pallansch MA. Evidence for frequent recombination within species human enterovirus B based on complete genomic sequences of all thirty-seven serotypes. *J Virol.* 2004 Jan;78(2):855-67.
- [5] Domingo E, Martin V, Perales C, Escarmis C. Coxsackieviruses and quasispecies theory: evolution of enteroviruses. *Curr Top Microbiol Immunol* 2008; 323: 3-32.
- [6] Savolainen-Kopra C, Blomqvist S. Mechanisms of genetic variation in polioviruses. *Rev Med Virol* 2010; 20: 358-371. DOI: 10.1002/rmv.663
- [7] Simmonds P, Welch J. Frequency and dynamics of recombination within different species of human enteroviruses. *J Virol* 2006; 80: 483-493. DOI: 10.1128/JVI.80.1.483-493.2006
- [8] Lukashev AN, Lashkevich VA, Ivanova OE, Koroleva GA, Hinkkanen AE, Ilonen J. Recombination in circulating Human enterovirus B: independent evolution of structural and non-structural genome regions. *J Gen Virol* 2005; 86: 3281-3290. DOI: 10.1099/vir.0.81264-0
- [9] Mirand A, Henquell C, Archimbaud C, Peigue-Lafeuille H, Bailly JL. Emergence of recent echovirus 30 lineages is marked by serial genetic recombination events. *J Gen Virol* 2007; 88: 166-176. DOI: 10.1099/vir.0.82146-0
- [10] Bessaud M, Joffret ML, Holmblat B, Razafindratsimandresy R, Delpeyroux F. Genetic relationship between cocirculating Human enteroviruses species C. *PLoS One* 2011; 6: e24823. DOI: 10.1371/journal.pone.0024823

- [11] Martinez-Salas E, Pacheco A, Serrano P, Fernandez N. New insights into internal ribosome entry site elements relevant for viral gene expression. *J Gen Virol* 2008; 89: 611-626. DOI: 10.1099/vir.0.83426-0
- [12] Florez de Sessions P, Dobrikova E, Gromeier M. Genetic adaptation to untranslated region-mediated enterovirus growth deficits by mutations in the nonstructural proteins 3AB and 3CD. *J Virol* 2007; 81: 8396-8405. DOI: 10.1128/jvi.00321-07
- [13] Solomon T, Lewthwaite P, Perera D, Cardoso MJ, McMinn P, Ooi MH. Virology, epidemiology, pathogenesis, and control of enterovirus 71. *Lancet Infect Dis* 2010; 10: 778-790. DOI: 10.1016/s1473-3099(10)70194-8
- [14] Huang SW, Kiang D, Smith DJ, Wang JR. Evolution of re-emergent virus and its impact on enterovirus 71 epidemics. *Exp Biol Med (Maywood)* 2011; 236: 899-908. DOI: 10.1258/ebm.2010.010233
- [15] Zhao K, Han X, Wang G, Hu W, Zhang W, Yu XF. Circulating coxsackievirus A16 identified as recombinant type A human enterovirus, China. *Emerg Infect Dis* 2011; 17: 1537-1540. DOI: 10.3201/eid1708.101719
- [16] Hu YF, Yang F, Du J, et al. Complete genome analysis of coxsackievirus A2, A4, A5, and A10 strains isolated from hand, foot, and mouth disease patients in China revealing frequent recombination of human enterovirus A. *J Clin Microbiol* 2011; 49: 2426-2434. DOI: 10.1128/JCM.00007-11
- [17] Oberste MS, Penaranda S, Pallansch MA. RNA Recombination Plays a Major Role in Genomic Change during Circulation of Coxsackie B Viruses. *Journal of Virology* 2004; 78: 2948-2955. DOI: 10.1128/jvi.78.6.2948-2955.2004
- [18] Oprisan G, Combiescu M, Guillot S, et al. Natural genetic recombination between co-circulating heterotypic enteroviruses. *J Gen Virol* 2002; 83: 2193-2200.
- [19] Bolanaki E, Kottaridi C, Markoulatos P, Kyriakopoulou Z, Margaritis L, Katsorchis T. Partial 3D gene sequences of Coxsackie viruses reveal interspecies exchanges. *Virus Genes* 2007; 35: 129-140. DOI: 10.1007/s11262-007-0083-2

- [20] Kew O, Morris-Glasgow V, Landaverde M, et al. Outbreak of poliomyelitis in Hispaniola associated with circulating type 1 vaccine-derived poliovirus. *Science* 2002; 296: 356-359. DOI: 10.1126/science.1068284
- [21] Paximadi E, Karakasiliotis I, Papaventsis D, Papageorgiou G, Markoulatos P. Recombinant Sabin environmental isolates in Greece and Cyprus. *J Appl Microbiol* 2008; 104: 1153-1162. DOI: 10.1111/j.1365-2672.2007.03649.x