

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ  
ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΥΓΕΙΑΣ  
ΤΜΗΜΑ ΒΙΟΧΗΜΕΙΑΣ ΚΑΙ ΒΙΟΤΕΧΝΟΛΟΓΙΑΣ

ΕΞΕΛΙΚΤΙΚΗ ΑΝΑΛΥΣΗ  
ΚΟΡΩΝΟΪΩΝ ΜΕ  
ΒΙΟΠΛΗΡΟΦΟΡΙΚΕΣ  
ΜΕΘΟΔΟΥΣ

ΜΠΡΑΜΗΣ ΧΡΗΣΤΟΣ  
Νικόλαος

UNIVERSITY OF THESSALY  
FACULTY OF LIFE SCIENCES  
DEPARTMENT OF BIOCHEMISTRY AND BIOTECHNOLOGY

# EVOLUTIONARY ANALYSIS OF CORONAVIRUSES USING BIOINFORMATIC METHODS

BRAMIS CHRISTOS

Nikolaos

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

**Υπεύθυνος καθηγητής:**

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

**Τριμελής επιτροπή:**

- Αμούτζιας Γρηγόριος, Αναπληρωτής Καθηγητής Βιοπληροφορικής με έμφαση στην Μικροβιολογία, Τμήμα Βιοχημείας και Βιοτεχνολογίας, Π.Θ.
- Μόσιαλος Δημήτριος, Αναπληρωτής Καθηγητής Βιοτεχνολογίας Μικροβίων, Τμήμα Βιοχημείας και Βιοτεχνολογίας, Π.Θ.
- Κυριακοπούλου Ζαχαρούλα, Αναπληρώτρια Καθηγήτρια Μικροβιολογίας, Γενικό Τμήμα, Π.Θ.

## **ΕΥΧΑΡΙΣΤΙΕΣ**

Αρχικά, θα ήθελα να ευχαριστήσω ιδιαίτερα τον υπεύθυνο καθηγητή μου κ. Αμούτζια Γρηγόριο για όλη τη βοήθεια που παρείχε στην εκπόνηση αυτής της πτυχιακής εργασίας, για όλα τα αντικείμενα που με δίδαξε και όλες τις συμβουλές που μου έδωσε αυτόν τον ολόκληρο χρόνο που ήμουν μέλος του εργαστηρίου Βιοπληροφορικής. Επίσης θα ήθελα να ευχαριστήσω όλα τα μέλη του εργαστηρίου για τη βοήθειά τους και ιδιαίτερα τον κ. Νικολαΐδη Μάριο που βοήθησε τα μέγιστα κάθε φορά που χρειάστηκαν οι δεξιότητες και οι γνώσεις του. Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου και τους στενούς μου φίλους για την τεράστια υποστήριξή τους.

# Κατάλογος περιεχομένων

<b>ΠΕΡΙΛΗΨΗ</b> .....	6
<b>ABSTRACT</b> .....	7
<b>1 ΕΙΣΑΓΩΓΗ</b> .....	8
1.1 Ταξινόμηση.....	8
1.2 Γονιδιωματική οργάνωση.....	11
1.3 Δομή ιοσωμάτιου.....	13
1.4 Κύκλος ζωής κορωνοϊών.....	15
1.4.1 Προσκόλληση και είσοδος.....	15
1.4.2 Έκφραση του αντιγραφικού γονιδίου.....	16
1.4.3 Αντιγραφή και μεταγραφή.....	17
1.4.4 Συγκρότηση και απελευθέρωση.....	17
1.5 Προέλευση και εξέλιξη κορωνοϊών.....	18
<b>2 ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ</b> .....	20
2.1 Λογισμικό.....	20
2.1.1 Linux Ubuntu.....	20
2.1.2 Perl 5.26.1.....	20
2.1.3 SeaView 5.0.....	20
2.1.4 Usearch.....	21
2.1.5 MEGA - X.....	21
2.1.6 Gblocks.....	21
2.1.7 Visual TreeCmp.....	21
2.1.8 Dendroscope 3.....	21
2.2 Λήψη γονιδιωμάτων.....	22
<b>3 ΑΠΟΤΕΛΕΣΜΑΤΑ</b> .....	23
3.1 Απομόνωση γονιδίων από τις κωδικές αλληλουχίες.....	23
3.2 Φιλτράρισμα αλληλουχιών μέσω της πρωτεΐνης Spike.....	23
3.3 Πολλαπλή στοίχιση.....	23
3.4 Υπολογισμός γενετικών αποστάσεων.....	24
3.5 Σύγκριση φυλογενετικών δέντρων.....	29
3.5.1 Alphacoronavirus.....	29
3.5.2 Betacoronavirus.....	35
3.5.3 Gammacoronavirus.....	41
3.5.4 Deltacoronavirus.....	47
<b>4 ΣΥΖΗΤΗΣΗ</b> .....	50
<b>ΒΙΒΛΙΟΓΡΑΦΙΑ</b> .....	53

## ΠΕΡΙΛΗΨΗ

Η παγκόσμια κοινότητα από τον Δεκέμβρη του 2019 βρίσκεται αντιμέτωπη με μια συνεχιζόμενη πανδημία, με περίπου 900.000 θανάτους έως τώρα, που οφείλεται στο στέλεχος κορωνοϊού SARS – CoV – 2, όπως ονομάστηκε από τον Παγκόσμιο Οργανισμό Υγείας. Τόσο αυτό το στέλεχος όσο και τα στελέχη που ήταν υπεύθυνα για προηγούμενες επιδημίες που προκλήθηκαν από κορωνοϊούς (SARS, MERS) είναι ζωνοσογόνοι παθογόνοι ιοί που προήλθαν από άγρια ζώα. Οι κορωνοϊοί έχουν μεγάλα RNA γονιδιώματα που κωδικοποιούν για δομικές και μη δομικές πρωτεΐνες υπεύθυνες για την επιβίωση και αναπαραγωγή τους. Το RNA γονιδίωμα των ιών επιτρέπει την ύπαρξη πολλών ανασυνδυασμών και μεταλλάξεων γεγονός που ευνοεί την εξέλιξη τους. Σκοπός της εργασίας μας ήταν να πραγματοποιήσουμε βιοπληροφορική ανάλυση όλων των ομάδων των κορωνοϊών με υπολογιστικά εργαλεία ώστε να μελετήσουμε την εξελικτική τους πορεία. Από τις αναλύσεις που πραγματοποιήθηκαν παρατηρήσαμε σχετικά χαμηλά επίπεδα ομοιότητας των αμινοξικών αλληλουχιών του γονιδίου Spike σε όλα τα γένη κορωνοϊών αλλά και ενδεχόμενη ύπαρξη ανασυνδυασμών γύρω από το γονίδιο αυτό.

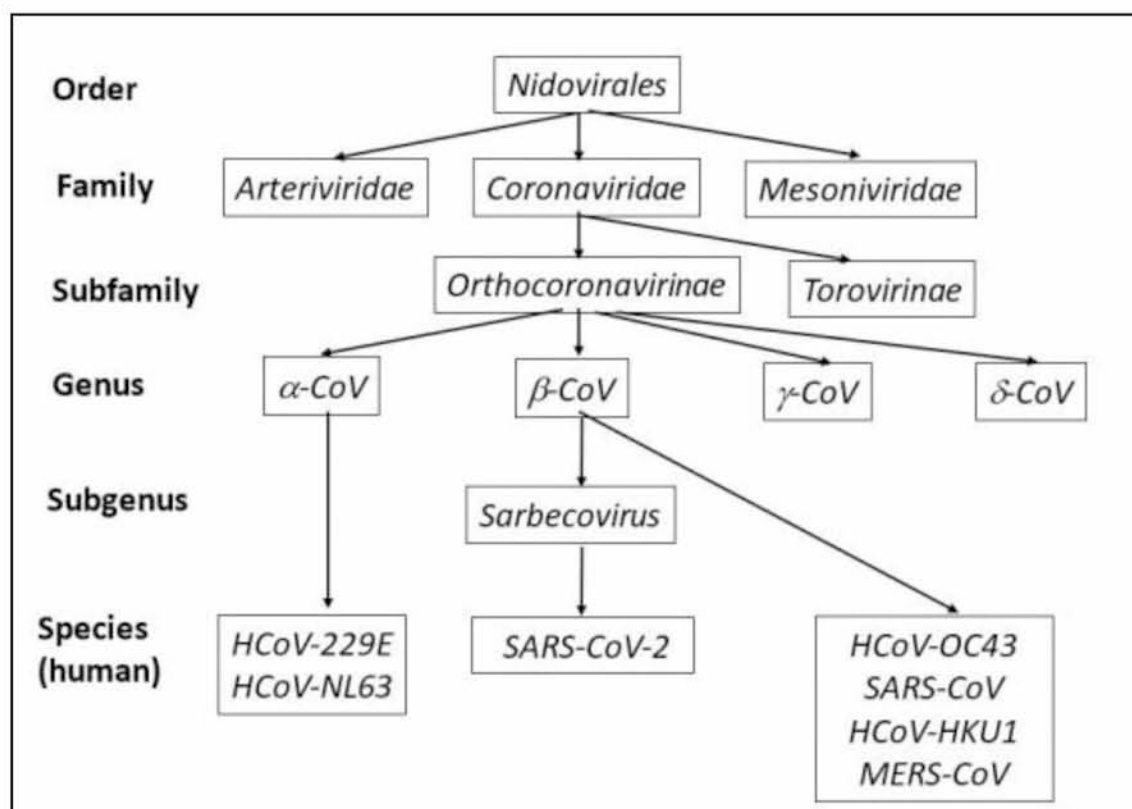
## **ABSTRACT**

Since December 2019, the global community faces an ongoing pandemic, which has cost the life of 900.000 people until now, caused by a strain of coronavirus, named SARS – CoV – 2 by the World Health Organization. This human coronavirus strain as well as the previous ones that were responsible for the earlier epidemics (SARS, MERS) are zoonotic pathogens that originated in wild animals. Coronaviruses have large RNA genomes that encode for structural and non structural proteins responsible for their survival and replication. RNA viruses allow plenty of recombination events and mutations in their genome, which favors quicker evolution of the viruses. The aim of this work was to perform a bioinformatics analysis of all the major coronavirus groups with the help of publicly available computational tools in order to understand their evolutionary trajectory. From the analysis that took place, we observed low similarity on the aminoacid sequences at the Spike gene of all the coronavirus genera, in addition to the possible existence of recombination events among different evolutionary groups, upstream and downstream of this gene.

# 1 ΕΙΣΑΓΩΓΗ

## 1.1 Ταξινόμηση

Οι κορωνοϊοί (CoVs) είναι η μεγαλύτερη ομάδα ιών που ανήκουν στην τάξη των *Nidovirales*, του βασιλείου των *Riboviria* και αποτελούν την μία από τις δύο υποοικογένειες, της οικογένειας *Coronaviridae*. Οι *Orthocoronavirinae* κατηγοριοποιούνται περαιτέρω σε τέσσερα γένη, τους *alpha*, *beta*, *gamma* και *delta* κορωνοϊούς (Εικόνα 1.1). Αρχικά οι ιοί, κατηγοριοποιούνταν με βάση τις ορολογικές αντιδράσεις των ιικών πρωτεϊνών καρπιών (*spikes*), αλλά πλέον η ταξινόμηση τους βασίζεται σε συγκριτικές αναλύσεις των αναπαραγωγικών πρωτεϊνών τους και πιο συγκεκριμένα τις πρωτεΐνες 3CLpro, NiRAN, RdRp, ZBD και HEL1 καθώς είναι συντηρημένες σε όλους τους ιούς της τάξης *Nidovirales*. Οι ιοί που έχουν 90% ομοιότητα στην αμινοξική αλληλουχία των συντηρημένων αυτών περιοχών, θεωρούνται ότι ανήκουν στο ίδιο είδος. (Saberī et al., 2018, *Coronaviridae Study Group of the International Committee on Taxonomy of Viruses*, 2020).



Εικόνα 1.1. Σχεδιάγραμμα ταξινόμησης κορωνοϊών (Malik, 2020).

Πριν την επιδημία του ιού SARS υπήρχαν μόλις 10 CoVs διαθέσιμοι με πλήρως αλληλουχημένο γονιδίωμα. Αρχικά κατηγοριοποιήθηκαν σε 3 ομάδες που ονομάστηκαν Group 1, Group 2, Group 3 αλλά το 2011 το Study Group of the International Committee for Taxonomy of Viruses μετονόμασε αυτές τις 3 ομάδες σε 3 γένη: *Alphacoronavirus*,



Betacoronavirus και Gammacoronavirus. Μετά την επιδημία του ιού SARS, ανακαλύφθηκε ένας απροσδόκητα μεγάλος αριθμός νέων CoVs που προϋπήρχαν, αλλά ήταν άγνωστοι μέχρι τότε. Αυτό οδήγησε στη δημιουργία νέων υπογενών καθώς και σε ένα καινούριο γένος, αυτό των Deltacoronaviruses όπως φαίνεται στον πίνακα 1.2 και στο φυλογενετικό δένδρο στην εικόνα 1.3.

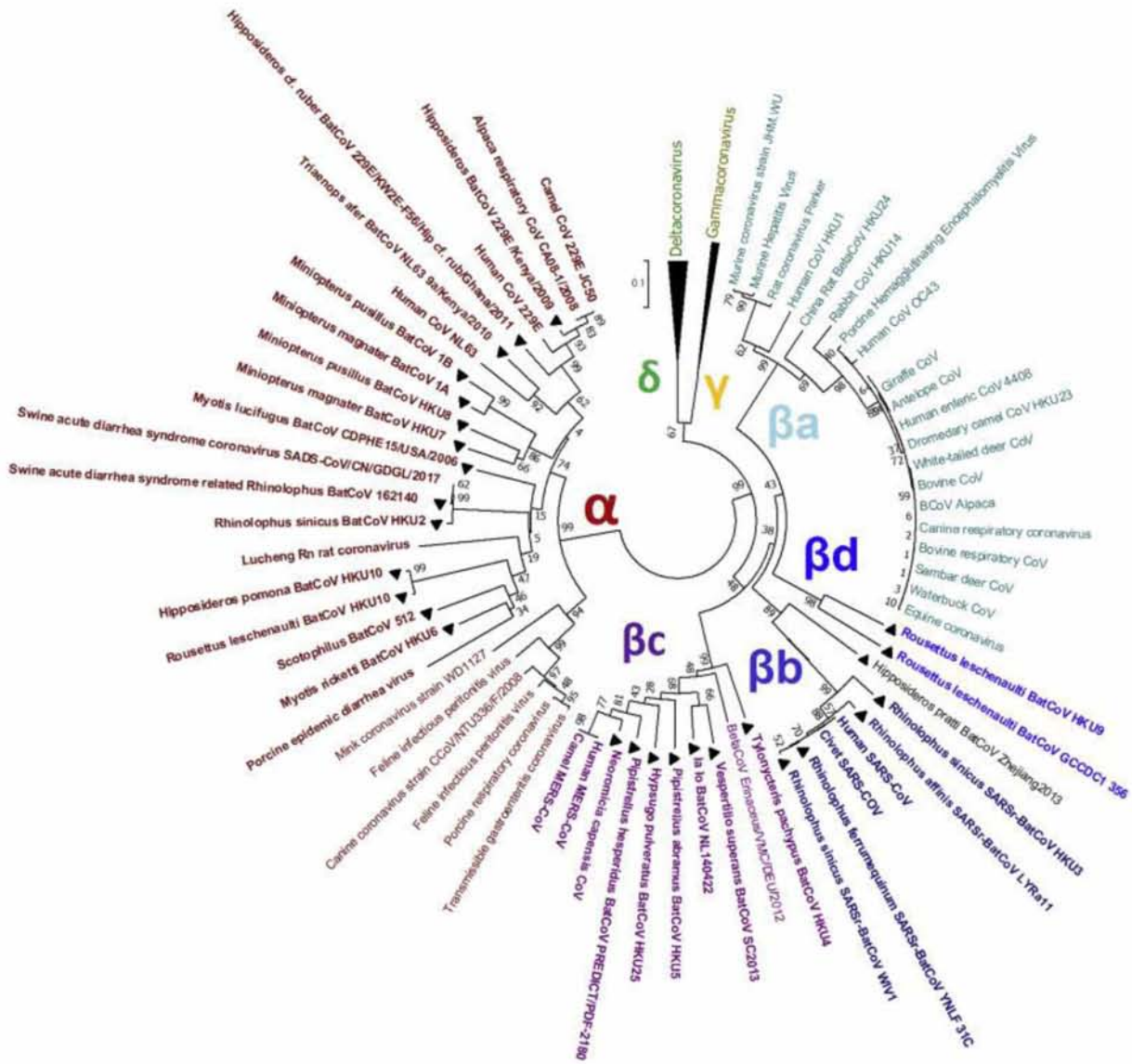
<b>Alphacoronavirus</b>	
Subgenus	Species
Colacovirus	Bat coronavirus CDPHE15
Decacovirus	Bat coronavirus HKU10 Rhinolophus ferrumequinum alphacoronavirus HuB-2013
Duvinacovirus	Human coronavirus 229E
Luchacovirus	Lucheng Rn rat coronavirus
Minacovirus	Mink coronavirus 1
Minunacovirus	Miniopterus bat coronavirus 1 Miniopterus bat coronavirus HKU8
Myotacovirus	Myotis ricketti alphacoronavirus Sax-2011
Nyctacovirus	Nyctalus velutinus alphacoronavirus SC-2013 Pipistrellus kuhlii coronavirus 3398
Pedacovirus	Porcine epidemic diarrhea virus Scotophilus bat coronavirus 512
Rhinacovirus	Rhinolophus bat coronavirus HKU2
Setracovirus	Human coronavirus NL63 NL63-related bat coronavirus strain BtKYNL63-9b
Soracovirus	Sorex araneus coronavirus T14
Sunacovirus	Suncus murinus coronavirus X74
Tegacovirus	Alphacoronavirus 1
<b>Betacoronavirus</b>	
Subgenus	Species
Embecovirus	Betacoronavirus 1 China Rattus coronavirus HKU24 Human coronavirus HKU1 Murine coronavirus Myodes coronavirus 2JL14
Hibecovirus	Bat Hp-betacoronavirus Zhejiang2013
Merbecovirus	Hedgehog coronavirus 1 Middle East respiratory syndrome-related coronavirus Pipistrellus bat coronavirus HKU5 Tylonycteris bat coronavirus HKU4
Nobecovirus	Eidolon bat coronavirus C704 Rousettus bat coronavirus GCCDC1 Rousettus bat coronavirus HKU9
Sarbecovirus	Severe acute respiratory syndrome-related coronavirus
<b>Gammacoronavirus</b>	
Subgenus	Species
Brangacovirus	Goose coronavirus CB17
Cegacovirus	Beluga whale coronavirus SW1
Igacovirus	Avian coronavirus Avian coronavirus 9203 (Turkey) Duck coronavirus 2714

## Deltacoronavirus

Subgenus	Species
Andecovirus	Wigeon coronavirus HKU20
Buldecovirus	Bulbul coronavirus HKU11
	Common moorhen coronavirus HKU21
	Coronavirus HKU15
	Munia coronavirus HKU13
	White-eye coronavirus HKU16
Herdecovirus	Night heron coronavirus HKU19

Πίνακας 1.2. Ταξινόμηση των subgenuses και οι πιο χαρακτηριστικοί CoVs αυτών (International Committee for Taxonomy of Viruses, 2020).

Όλοι οι ιοί της τάξης των Nidovirales είναι RNA ιοί θετικής πολικότητας με μη τεμαχισμένο γονιδίωμα το οποίο προστατεύεται από πυρηνικό φάκελο. Όλοι περιέχουν πολύ μεγάλα γονιδιώματα για RNA ιούς, με κάποιους από αυτούς να έχουν το μεγαλύτερο μήκος RNA γονιδιώματος που έχει παρατηρηθεί, φτάνοντας μέχρι και τις 33,5 κιλοβάσεις (kb). Μερικά άλλα κοινά χαρακτηριστικά των ιών της τάξης Nidovirales περιλαμβάνουν μια υψηλά συντηρημένη γενωμική οργάνωση, με ένα τεράστιο ανοικτό αναγνωστικό πλαίσιο (Open Reading Frame – ORF) που είναι υπεύθυνο για την αναπαραγωγή του ιού που το διαδέχονται δομικά και βοηθητικά ORFs. Επίσης, έχουμε έκφραση πολλών μη δομικών ORFs με ριβοσωμική ολίσθηση (ribosomal frameshifting) και έκφραση καθοδικών ORFs με σύνθεση 3' ένθετων υπογενωμικών mRNAs. Στην πραγματικότητα, η τάξη των Nidovirales πήρε το όνομα της από τα 3' ένθετα (nested) mRNAs καθώς nido στα λατινικά σημαίνει “φωλιά”. Οι μεγαλύτερες διαφορές μεταξύ των ιών αυτής της τάξης βρίσκονται στον αριθμό, τύπο και μέγεθος των δομικών πρωτεϊνών, οι οποίες διαφορές προκαλούν σημαντικές αλλαγές στη δομή και μορφολογία του νουκλεοκαψιδίου και των ιικών σωματίων (Fehr and Perlman, 2015).

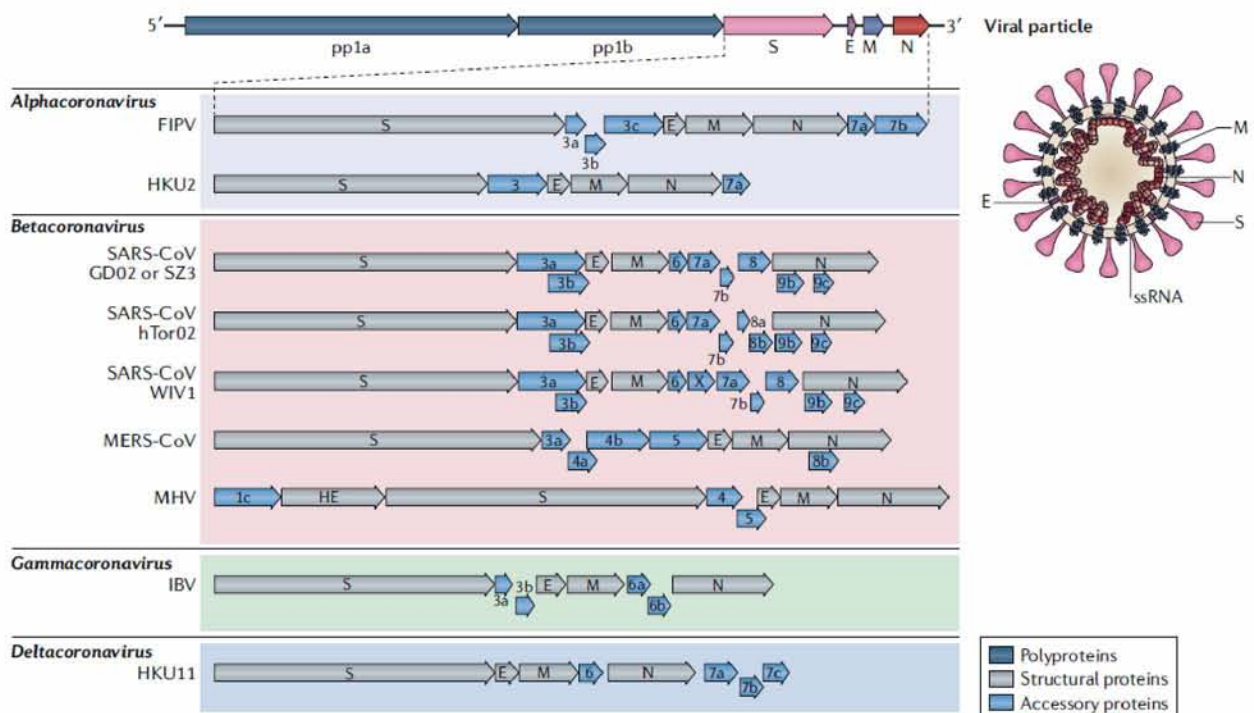


Εικόνα 1.3 Η ετερογένεια των κορωνοϊών όπως αποτυπώνεται στο φυλογενετικό δένδρο που έχει κατασκευασθεί μέσω της αμινοξικής αλληλουχίας της RdRp. Με μαύρα βέλη διακρίνονται οι bat CoVs. Οι κλάδοι των Gammacoronavirus και Deltacoronavirus έχουν συμπιεστεί (Luk et al., 2019).

## 1.2 Γονιδιωματική οργάνωση

Οι κορωνοϊοί περιέχουν μονόκλωνο, μη τεμαχισμένο, θετικής πολικότητας RNA γονιδίωμα μήκους περίπου 30kb με 5' καλύπτρα και 3' poly-A ουρά. Το ORF αναπαραγωγής που κωδικοποιεί για τις μη δομικές πρωτεΐνες (non structural proteins - nsps) καταλαμβάνει τα 2/3 του συνόλου του γονιδιώματος, με τις δομικές και τις βοηθητικές πρωτεΐνες να καταλαμβάνουν τα υπόλοιπα 10kb περίπου (εικόνα 1.4). Το γονιδιωματικό RNA χρησιμοποιείται ως μήτρα για την απευθείας μετάφραση της πολυπρωτεΐνης 1a/1b (pp1a/pp1b) η οποία κωδικοποιεί για τις μη δομικές πρωτεΐνες οι οποίες σχηματίζουν το σύμπλοκο αντιγραφής-μεταγραφής (Replicase-Transcriptase Complex) μέσα σε κυστίδια με διπλή μεμβράνη (Double Membrane Vesicles). Επομένως, μια ομάδα ένθετων υπογονιδιωματικών RNAs συντίθεται από το σύμπλοκο αντιγραφής-μεταγραφής με ασυνεχή μεταγραφή. Αυτά τα υπογονιδιωματικά αγγελιοφόρα RNAs κατέχουν κοινές

αλληλουχίες στο 5' και 3' άκρο.



Εικόνα 1.4. Η οργάνωση του γονιδιώματος των πιο χαρακτηριστικών ιών κάθε γένους της οικογένειας των κορωνοϊών (Cui et al., 2019).

Το γονιδίωμα ενός τυπικού κορωνοϊού περιέχει τουλάχιστον 6 ORFs. Το πρώτο ORF (ORF 1a/b) κωδικοποιεί 16 μη δομικές πρωτεΐνες (πίνακας 1.5), με εξαίρεση τους Gammacoronaviruses στους οποίους δεν υπάρχει η 1η μη δομική πρωτεΐνη. Μεταξύ του ORF1a και ORF1b υπάρχει ολίσθηση του πλαισίου ανάγνωσης, μία θέση ανοδικά η οποία οδηγεί στην παραγωγή δύο πολυπεπτιδίων, την pp1a και pp1b. Αυτές οι πολυπεπτιδιακές πρωτεΐνες επεξεργάζονται κατάλληλα από ειδικές πρωτεάσες που κωδικοποιούνται από το ίδιο γονιδίωμα και δημιουργούν τις 16 μη δομικές πρωτεΐνες. Η pp1a περιέχει τις 1-11 nsps ενώ η pp1b τις 12-16. Μεταξύ αυτών των πρωτεασών είναι οι: chymotrypsin-like protease (3Clpro – nsp5), και η papain-like protease (nsp3) με ένα ή δύο domains. Τα επόμενα ORFs που βρίσκονται κοντά στο 3' άκρο του γονιδιώματος κωδικοποιούν για τουλάχιστον 4 απαραίτητες δομικές πρωτεΐνες: καρφιά (Spike), μεμβράνης (Membrane), φακέλου (Envelope) και νουκλεοκαψιδίου (Nucleocapsid). Εκτός από αυτές τις απαραίτητες πρωτεΐνες, διάφοροι κορωνοϊοί κωδικοποιούν ειδικές δομικές και βοηθητικές πρωτεΐνες, όπως τις HE, 3a/b και 4a/b οι οποίες μαζί με τις υπόλοιπες δομικές πρωτεΐνες μεταφράζονται από τα υπογενωμικά RNAs των κορωνοϊών (Chen et al., 2020).

<b>Μη δομική πρωτεΐνη</b>	<b>Λειτουργία</b>
nsp1	Αποδόμηση κυτταρικού mRNA, παρεμπόδιση μετάφρασης κυττάρου ξενιστή
nsp2	Άγνωστη λειτουργία
nsp3	Τεμαχισμός πολυπεπτιδίων, παρεμπόδιση ανοσολογικής απόκρισης του κυττάρου ξενιστή, επαγωγή έκφρασης κυτοκινών
nsp4	Σχηματισμός κυστιδίων διπλής μεμβράνης (DMV)
nsp5	3CL πρωτεάση, M πρωτεάση, τεμαχισμός πολυπεπτιδίων
nsp6	Διαμεμβρανική πρωτεΐνη σκαλωσιά
nsp7	Σχηματίζει εξαμερές σύμπλοκο με τα nsp8 και nsp12
nsp8	Σχηματίζει εξαμερές σύμπλοκο με τα nsp7 και nsp12
nsp9	Πρωτεΐνη δέσμευσης του RNA
nsp10	Πρωτεΐνη σκαλωσιά με τα nsp14 και nsp16
nsp11	Άγνωστη λειτουργία
nsp12	RNA-εξαρτώμενη RNA πολυμεράση
nsp13	RNA ελικάση, 5' τριφωσφατάση
nsp14	Εξωριβονουκλεάση, N7 ΜΤάση
nsp15	Ενδοριβονουκλεάση
nsp16	2'-Ο-Μεθυλοτρανσφεράση

Πίνακας 1.5. Οι 16 μη δομικές πρωτεΐνες των κορωνοϊών και οι λειτουργίες τους (Chen et al., 2020).

### 1.3 Δομή ιοσωμάτιου

Τα ιικά σωμάτια των κορωνοϊών είναι σφαιρικά με διάμετρο περίπου 125nm σύμφωνα με τελευταίες μελέτες με cryo-electron μικροσκοπία και τομογραφία. Το πιο εξέχων χαρακτηριστικό των συγκεκριμένων ιών είναι τα καρφιά σε σχήμα μπαστουιού (club-shaped spikes) που αναδύονται από την επιφάνεια του ιοσωμάτιου. Αυτά τα καρφιά είναι το χαρακτηριστικό στοιχείο του ιικού σωμάτιου και του δίνει την εικόνα ενός ηλιακού στέματος, μιας ηλιακής κορώνας, από την οποία η οικογένεια των κορωνοϊών πήρε το όνομά της. Μέσα στο φάκελο του ιοσωμάτιου βρίσκεται το νουκλεοκαψίδιο, το οποίο έχει ελικοειδή συμμετρία, ένα χαρακτηριστικό που είναι ασυνήθιστο μεταξύ RNA ιών θετικής πολικότητας αλλά συνηθισμένο σε αυτούς με αρνητική πολικότητα (Εικόνα 1.6, Lee et al., 2020).

Η S πρωτεΐνη (~150 kDa), χρησιμοποιεί μια αμινοτελική σηματοδοτούμενη περιοχή για να εισέλθει στο κύτταρο ξενιστή. Η χαρακτηριστική δομή των καρφίων προέρχεται από τριμερή των S πρωτεϊνών. Η τριμερής S γλυκοπρωτεΐνη είναι της τάξης 1 των fusion proteins και μεσολαβεί στην προσκόλληση στον υποδοχέα του ξενιστή. Στους περισσότερους κορωνοϊούς, η S πρωτεΐνη διασπάται από μια furin-like πρωτεάση του κυττάρου ξενιστή σε δύο ξεχωριστά πολυπεπτίδια, τα S1 και S2. Η S1 δημιουργεί το



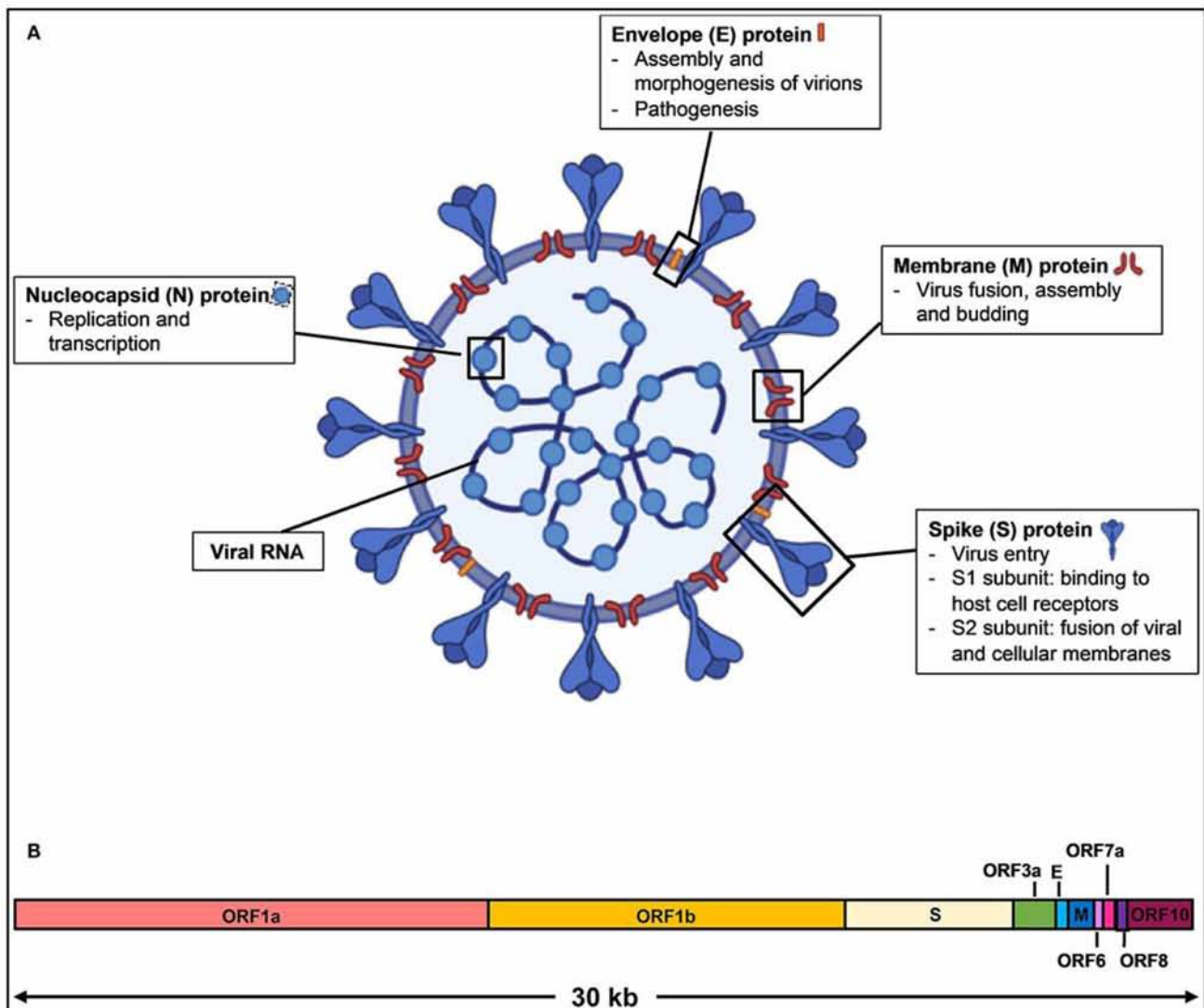
μεγάλο τμήμα της S πρωτεΐνης που προσδένεται στον υποδοχέα, ενώ η S2 σχηματίζει τη δομή κοτσανιού του καρφιού και είναι πιο συντηρημένη.

Η M πρωτεΐνη είναι αυτή με τη μεγαλύτερη αφθονία μεταξύ των δομικών πρωτεϊνών μέσα στο ιοσωμάτιο. Είναι μια μικρή πρωτεΐνη (~25-30 kDa) με 3 διαμεμβρανικές περιοχές και θεωρείται ότι είναι αυτή που δίνει το σχήμα στο ιοσωμάτιο. Έχει μια μικρή αμινοτελική γλυκοσυλιωμένη εξωδομή και ένα πολύ μεγαλύτερο καρβοξυτελικό ενδοτομέα που επεκτείνεται 6-8 nm μέσα στο ιικό σωμάτιο.

Η E πρωτεΐνη (~8-12 kDa) υπάρχει σε μικρές ποσότητες μέσα στο ιοσωμάτιο. Οι E πρωτεΐνες των κορωνοϊών είναι αρκετά αποκλίνουσες μεταξύ τους αλλά έχουν κοινή αρχιτεκτονική. Αν και δεν είναι πλήρως γνωστή ακόμα η μεμβρανική τοπολογία της πρωτεΐνης, τα περισσότερα δεδομένα όμως συνηγορούν ότι πρόκειται για διαμεμβρανική πρωτεΐνη με δραστηριότητα καναλιού ιόντων. Περιέχει μια αμινοτελική εξωπεριοχή και μια καρβοξυτελική ενδοπεριοχή. Εν αντιθέσει με τις υπόλοιπες δομικές πρωτεΐνες, οι χωρίς την E πρωτεΐνη μπορεί να μην είναι θανατηφόροι αν και εξαρτάται κυρίως από τον τύπο του ιού (DeDiego et al., 2007). Η E πρωτεΐνη χρησιμοποιείται για την κατασκευή και απελευθέρωση του ιού αλλά έχει και άλλες λειτουργίες. Για παράδειγμα, η E πρωτεΐνη του SARS-CoV δεν απαιτείται για την ιική αναπαραγωγή, αλλά είναι απαραίτητη για την παθογένεση.

Η N πρωτεΐνη αποτελεί τη μοναδική πρωτεΐνη που βρίσκεται παρούσα στο νουκλεοκαψίδιο. Αποτελείται από δύο περιοχές, μια αμινοτελική και μια καρβοξυτελική, οι οποίες είναι ικανές να δεσμεύουν το RNA *in vitro*, αλλά χρησιμοποιούν διαφορετικούς μηχανισμούς η καθεμία για να το κάνουν. Έχει προταθεί πως η βέλτιστη πρόσδεση του RNA απαιτεί τη συνεισφορά και των δύο περιοχών. Η N πρωτεΐνη είναι υψηλά φωσφορυλιωμένη, με την φωσφορυλίωση να προκαλεί μια δομική αλλαγή που αυξάνει τη συγγένεια δέσμευσης του ιικού RNA σε σχέση με το μη ιικό RNA.

Μια πέμπτη δομική πρωτεΐνη, η αιμοσυγκολλίνη-εστεράση (Hemagglutinin-Esterase) υπάρχει στους beta-coronaviruses. Αυτή η πρωτεΐνη δεσμεύει το σιαλικό οξύ στην επιφάνεια των γλυκοπρωτεϊνών και έχει δράση ακέτυλο-εστεράσης. Αυτές οι δραστηριότητες πιστεύεται ότι αυξάνουν την είσοδο στο κύτταρο μέσω της S πρωτεΐνης αλλά και την εξάπλωση του ιού μέσω του βλεννογόνου (Fehr and Perlman, 2015).



Εικόνα 1.6. Σχηματικό διάγραμμα του ιοσωματίου και της γενωμικής οργάνωσης του ιού SARS-CoV-2 (Lee et al., 2020).

## 1.4 Κύκλος ζωής κορωνοϊών

### 1.4.1 Προσκόλληση και είσοδος

Η αρχική προσκόλληση του ιού στο κύτταρο ξενιστή αρχίζει από αλληλεπίδραση μεταξύ της S πρωτεΐνης και του υποδοχέα. Οι περιοχές του receptor binding domain (RBD) μέσα στην S1 περιοχή ενός κορωνοϊού ποικίλει ανάλογα με τον ιό, με μερικούς να έχουν την RBD στο αμινοτελικό άκρο του S1, ενώ άλλοι στο καρβοξυτελικό ακρό του S1. Η αλληλεπίδραση μεταξύ S πρωτεΐνης και υποδοχέα είναι το πρωταρχικό σημείο που καθορίζει αν ένας κορωνοϊός θα μολύνει ένα ξενιστή. Πολλοί κορωνοϊοί χρησιμοποιούν πεπτιδάσες ως υποδοχείς, ενώ άλλοι χρησιμοποιούν το ένζυμο μετατροπής αγγειοτενσίνης 2 (Angiotensin-Converting Enzyme 2 - ACE2).

Μετά τη σύνδεση στον υποδοχέα, ο ιός πρέπει να αποκτήσει πρόσβαση στο κυτοσόλιο του κύτταρου ξενιστή. Αυτό επιτυγχάνεται με πρωτεολυτική διάσπαση της S

πρωτεΐνης από μια καθεψίνη ή άλλη πρωτεάση. Η διάσπαση αυτή, φανερώνει τα πεπτίδια που είναι υπεύθυνα για την ένωση με τον υποδοχέα τα οποία εισέρχονται μέσα στη μεμβράνη μαζί με άλλες βοηθητικές δομές σχηματίζοντας μια δεσμίδα. Αυτή η δεσμίδα επιτρέπει την ανάμιξη του ιού με τις κυτταρικές μεμβράνες, οδηγώντας σε σύντηξη και τελικώς απελευθέρωση του ιικού γονιδιώματος στο κυτταρόπλασμα (Fehr and Perlman, 2015).

### 1.4.2 Έκφραση του αντιγραφικού γονιδίου

Το επόμενο στάδιο στον κύκλο ζωής των κορωνοϊών είναι η έκφραση του αντιγραφικού ORF του ιικού γονιδιωματικού RNA. Το αντιγραφικό ORF κωδικοποιεί δύο μεγάλα πεπτίδια, pp1a και pp1b. Για να εκφραστούν και οι δύο πολυπρωτεΐνες, ο ιός χρησιμοποιεί μια αλληλουχία ολίσθησης (5'-UUUAAAC-3') και έναν RNA ψευδοκόμβο που προκαλούν αλλαγή στο πλαίσιο ανάγνωσης και μετάβαση από το *gp1a* στο *gp1b*. Στις περισσότερες περιπτώσεις, το ριβόσωμα ξετυλίγει τη δομή ψευδοκόμβου και συνεχίζει τη μετάφραση μέχρι να συναντήσει το κωδικόνιο λήξης του *gp1a*. Περιστασιακά, ο ψευδοκόμβος εμποδίζει το ριβόσωμα από το να συνεχίσει την επιμήκυνση, σταματώντας το στην αλληλουχία ολίσθησης, αλλάζοντας το πλαίσιο ανάγνωσης πηγαίνοντας πίσω ένα νουκλεοτίδιο, πριν μπορέσει το ριβόσωμα να λύσει τη δομή ψευδοκόμβου και να επεκτείνει τη μετάφραση στο *gp1b*. Είναι ακόμα άγνωστο γιατί ακριβώς αυτοί οι ιοί χρησιμοποιούν την αλλαγή του πλαισίου ανάγνωσης για να ελέγχουν την έκφραση της πρωτεΐνης, αλλά έχει προταθεί ότι συμβαίνει είτε για να ελέγξει την ακριβής αναλογία μεταξύ pp1a και pp1b πρωτεϊνών ή για να καθυστερήσει την παραγωγή των προϊόντων του *gp1b* εωσότου τα προϊόντα του *gp1a* να έχουν δημιουργήσει ένα κατάλληλο περιβάλλον για την αντιγραφή του RNA.

Οι πολυπρωτεΐνες pp1a και pp1b περιέχουν τις μη δομικές πρωτεΐνες 1-11 και 12-16 αντίστοιχα. Αυτές οι πολυπρωτεΐνες διασπώνται τελικώς στα ξεχωριστά μη δομικά πεπτίδια. Οι κορωνοϊοί κωδικοποιούν για δύο ή τρεις πρωτεάσες η οποίες διασπών τις αντιγραφικές πολυπρωτεΐνες. Αυτές είναι οι, papain-like πρωτεάσες (PLP<sub>pro</sub>), που κωδικοποιούνται από το ns<sub>p3</sub> και μια πρωτεάση 3CL<sub>pro</sub> που κωδικοποιείται από το ns<sub>p5</sub>. Οι PLP<sub>pro</sub>s διασπών τα σύνορα μεταξύ ns<sub>p1/2</sub>, ns<sub>p2/3</sub> και ns<sub>p3/4</sub>, ενώ η 3CL<sub>pro</sub> είναι υπεύθυνη για τις υπόλοιπες 11 κοπές.

Πολλές από τις μη δομικές πρωτεΐνες συνδέονται στο σύμπλοκο αντιγραφής-μεταγραφής για να δημιουργήσουν ένα περιβάλλον κατάλληλο για τη σύνθεση του RNA, και επομένως είναι υπεύθυνες και για τη μεταγραφή και αντιγραφή του RNA των υπογενωμικών RNAs. Στις μη δομικές πρωτεΐνες υπάρχουν και άλλα ένζυμα, μεταξύ αυτών απαραίτητα για την αντιγραφή του RNA, όπως η RNA-εξαρτώμενη RNA πολυμεράση (RdRp) η οποία κωδικοποιείται από το ns<sub>p12</sub>, η RNA ελικάση από το ns<sub>p13</sub> και η εξωριβονουκλεάση (ExoN) στο ns<sub>p14</sub>. Επίσης υπάρχουν και μη δομικές πρωτεΐνες με άγνωστη λειτουργία, όπως η ενδοριβονουκλεάση (NendoU) που κωδικοποιείται από το ns<sub>p15</sub>. Ενδιαφέρον στοιχείο είναι το γεγονός πως τόσο το ns<sub>p14</sub>-ExoN όσο και το ns<sub>p15</sub>-NendoU είναι μοναδικά στην τάξη των Nidovirales και αποτελούν γενετικούς δείκτες γι' αυτούς τους ιούς (Fehr and Perlman, 2015).

### 1.4.3 Αντιγραφή και μεταγραφή

Μετά τη μετάφραση και συγκρότηση του ιικού αντιγραφικού συμπλόκου ακολουθεί



η σύνθεση του ιικού RNA. Η σύνθεση αυτή, παράγει τόσο τα γενωμικά όσο και τα υπογενωμικά RNAs. Τα υπογενωμικά RNAs δρουν ως mRNAs για τα δομικά και βοηθητικά γονίδια τα οποία βρίσκονται καθοδικά μετά τις αντιγραφικές πολυπρωτεΐνες. Όλα τα υπογενωμικά RNAs θετικής πολικότητας έχουν κοινό 3' τελικό άκρο με το πλήρους μήκους ιικό γονιδίωμα, σχηματίζοντας έτσι μια ομάδα ένθετων RNAs, ένα χαρακτηριστικό που είναι σήμα κατατεθέν των ιών της τάξης Nidovirales.

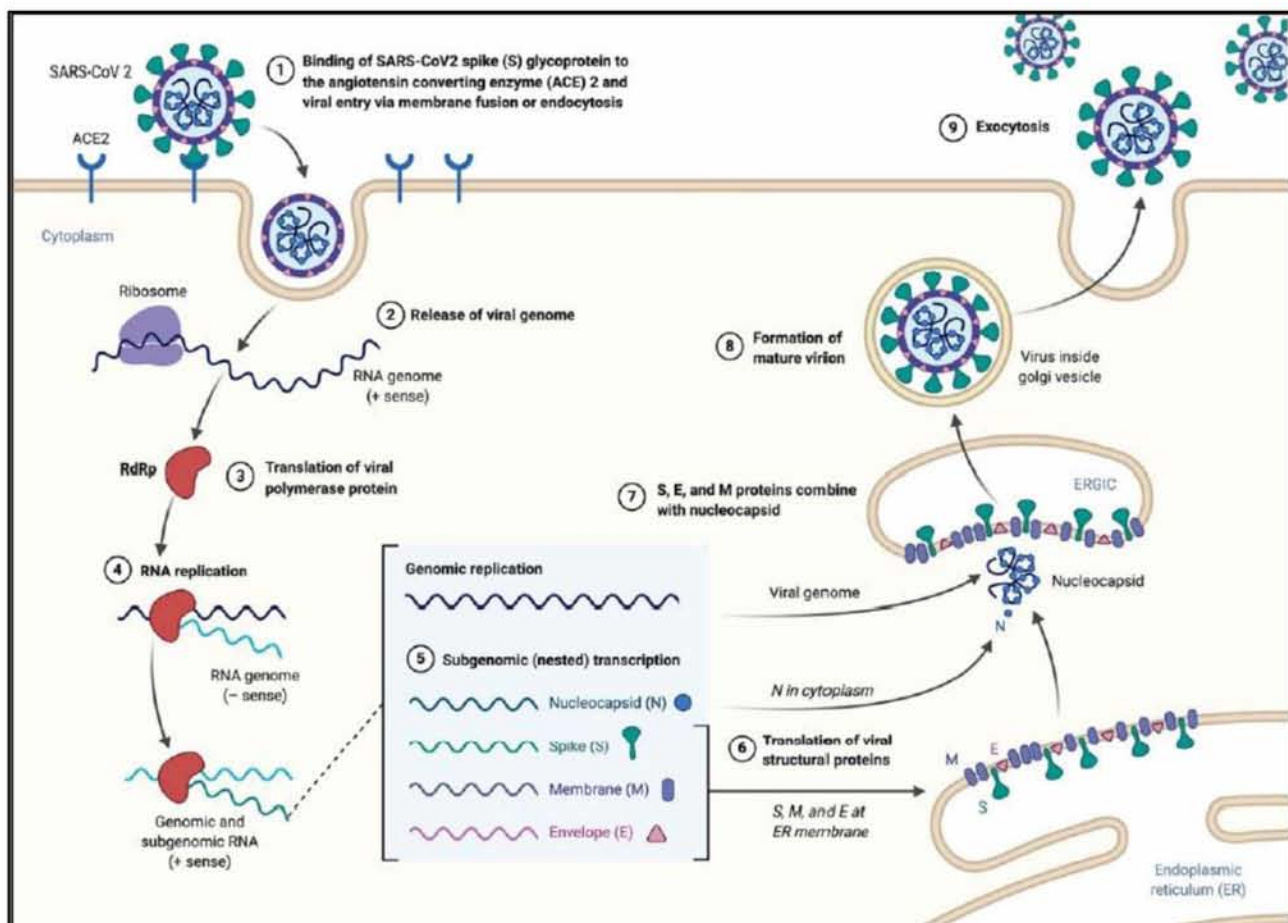
Πληθώρα cis αλληλουχιών είναι απαραίτητες για την αντιγραφή του ιικού RNA. Μέσα στην 5' UTR του γονιδιώματος υπάρχουν επτά δομές stem-loop που μπορεί να φτάνουν έως και το γονίδιο 1a. Η 3' UTR περιέχει μια διογκωμένη stem-loop, έναν ψευδοκόμβο και μια υπερμεταβλητή περιοχή. Όλες αυτές οι διαφορετικές δομές θεωρείται πως ρυθμίζουν τα διάφορα στάδια της σύνθεσης του RNA, αν και ο ακριβής μηχανισμός αλλά και ποια στάδια ρυθμίζονται είναι ακόμα άγνωστα.

Οι κορωνοϊοί είναι επίσης γνωστοί για την ικανότητα τους να ανασυνδυάζονται τόσο με ομόλογο όσο και με μη ομόλογο τρόπο. Καθοριστικό ρόλο σε αυτήν την ιδιότητα είναι η ικανότητα της RNA-εξαρτώμενης-RNA-πολυμεράσης να αλλάζει κλώνο. Ο ανασυνδυασμός όπως είναι γνωστό, παίζει σημαντικό ρόλο στην εξέλιξη των ιών (Fehr and Perlman, 2015).

#### 1.4.4 Συγκρότηση και απελευθέρωση

Μετά την αντιγραφή και τη σύνθεση των υπογενωμικών RNAs, οι ιικές δομικές πρωτεΐνες μεταφράζονται και εισάγονται στο ενδοπλασματικό δίκτυο. Εκεί, τα ιικά γενώματα κασιδιώνονται από το νουκλεοκαψίδιο σχηματίζοντας τα ώριμα ιοσωμάτια.

Η μεμβρανική πρωτεΐνη καθοδηγεί τις περισσότερες αλληλεπιδράσεις μεταξύ των πρωτεϊνών που απαιτούνται για το σχηματισμό των κορωνοϊών, ωστόσο δεν είναι αρκετή από μόνη της. Η μεμβρανική πρωτεΐνη όταν εκφράζεται μαζί με την πρωτεΐνη φάκελο, σχηματίζονται τα virus-like particles (VLPs), ενώ επίσης η πρωτεΐνη του νουκλεοκαψιδίου αυξάνει και αυτή το σχηματισμό των VLPs. Αν και η πρωτεΐνη Spike εισάγεται στο ιοσωμάτιο σε αυτό το στάδιο, δεν είναι απαραίτητη για τον σχηματισμό του. Η μεμβρανική πρωτεΐνη συνδέεται στο νουκλεοκαψίδιο, δίνοντας το σήμα για την ολοκλήρωση του σχηματισμού του ιικού σωμάτιου. Μετά την ολοκλήρωση του σχηματισμού, τα ιοσωμάτια μεταφέρονται στην επιφάνεια των κυττάρων μέσα σε κυστίδια και απελευθερώνονται με εξωκύτωση (Fehr and Perlman, 2015).



Εικόνα 1.5. Ο κύκλος ζωής του SARS-CoV-2 στο κύτταρο ξενιστή (Alanagreh et al., 2020).

## 1.5 Προέλευση και εξέλιξη κορωνοϊών

Οι ζωϊκοί κορωνοϊοί είναι γνωστοί στην ανθρωπότητα περίπου από το 1930 καθώς είχαν απομονωθεί στελέχη τους από μολυσμένα ζώα, μεταξύ αυτών γαλοπούλες, ποντίκια, αγελάδες, γουρούνια, γάτες και σκύλοι. Αντίθετα, το πρώτο στέλεχος κορωνοϊού που ανιχνεύθηκε σε άνθρωπο με κοινό κρύωμα ήταν αυτό του ιού HCoV-229E περίπου 30 χρόνια αργότερα (Ye et al., 2020).

Ένα ζώο λειτουργεί ως ξενιστής ενός νέου κορωνοϊού εάν φιλοξενεί κάποιον άλλον στενό συγγενή του νέου ιού, με τον οποίο έχουν υψηλή ομολογία σε νουκλεοτιδικό επίπεδο. Ο προγονικός αυτός ιός είναι συνήθως καλά προσαρμοσμένος και μη παθογόνος για το ζώο ξενιστή. Σε αυτήν την περίπτωση, τα ζώα αποτελούν τους φυσικούς ξενιστές των ιών αυτών εν αντιθέσει με την περίπτωση που ένας ιός εισέρχεται σε έναν ενδιάμεσο ξενιστή ακριβώς πριν μολύνει τον άνθρωπο. Τότε, ο ιός δεν είναι καλά προσαρμοσμένος στον ενδιάμεσο ξενιστή ζώο, με αποτέλεσμα συχνά να είναι και παθογόνος. Αυτός ο ενδιάμεσος ξενιστής αποτελεί την πηγή της ανθρώπινης μόλυνσης και επιτρέπει στον ιό να πολλαπλασιαστεί και στη συνέχεια να μεταδοθεί με μεγαλύτερο ιικό φορτίο (Ye et al., 2020).

Ο ιός MERS - CoV αποτελεί εξαιρετικό παράδειγμα της μετάδοσης μέσω ενδιάμεσου ξενιστή από τις νυχτερίδες στις καμήλες και από τις καμήλες στον άνθρωπο. Έχει αποδειχθεί πως η αρχική προέλευση του MERS - CoV είναι οι νυχτερίδες (Annan et al., 2013). Οι νυχτερίδες παρέχουν μια μεγάλη δεξαμενή για μετάδοση ιών μεταξύ διαφόρων ειδών καθώς τα χαρακτηριστικά τους όπως η μακροζωία, η ικανότητα τους να πετάνε αλλά και οι στενές σχέσεις που αναπτύσσουν δημιουργώντας αποικίες, ευνοούν τη διάδοση ιών. Στη συνέχεια, ο MERS - CoV προσαρμόστηκε στις καμήλες και

μετατράπηκε από ενδιάμεσο ξενιστή σε μια φυσική δεξαμενή του ιού. Ο συγκεκριμένος ιός προκαλεί ήπια ασθένεια και έχει σχετικά χαμηλό ρυθμό μεταλλάξεων σε αυτά τα ζώα.

## 2 ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ

### 2.1 Λογισμικό

Για την πραγμάτωση των αναλύσεων της διπλωματικής εργασίας ήταν απαραίτητη η χρήση εξειδικευμένου λογισμικού και γλωσσών προγραμματισμού τα οποία και παρατίθενται.

#### 2.1.1 Linux Ubuntu

Τα Linux Ubuntu αποτελούν ένα λειτουργικό σύστημα, ανοιχτού λογισμικού για υπολογιστές που διατίθενται δωρεάν για τους χρήστες. Το βασικό περιβάλλον τερματικού του Ubuntu χρησιμοποιεί το BASH και ενδείκνυται για τη διαχείριση μεγάλου όγκου δεδομένων (Canonical, n.d.).

#### 2.1.2 Perl 5.26.1

Η PERL (Practical Extraction and Report Language) είναι μια γλώσσα προγραμματισμού που εξειδικεύεται στο χειρισμό αρχείων κειμένου και την εξαγωγή ή εισαγωγή πληροφορίας σε αυτά. Ενδείκνυται για την ανάλυση βιοπληροφορικών δεδομένων ("About Perl - [www.perl.org](http://www.perl.org)," n.d.).

#### 2.1.3 SeaView 5.0

Το λογισμικό Seaview μπορεί να χρησιμοποιηθεί σε όλα τα λειτουργικά συστήματα (Linux, Windows, MacOS) και έχει σχεδιαστεί για να πραγματοποιεί πολλαπλές στοιχίσεις ακολουθιών DNA ή πρωτεϊνών και να δημιουργεί φυλογενετικά δένδρα. Το συγκεκριμένο λογισμικό υποστηρίζει τόσο γραφικό περιβάλλον (Graphical User Interface – GUI) όσο και περιβάλλον τερματικού (Command Line Interface – CLI) (Gouy et al., 2010). Για τις πολλαπλές στοιχίσεις δίνεται η δυνατότητα χρήσης τόσο του αλγορίθμου Clustal Omega όσο και του MUSCLE, ο οποίος και χρησιμοποιήθηκε στη παρούσα εργασία. Ο υπολογισμός των φυλογενετικών δένδρων μπορεί να γίνει είτε με μεθόδους αποστάσεων και πιο συγκεκριμένα με Neighbor Joining είτε με μεθόδους χαρακτήρων, όπως η μέγιστη φειδωλότητα (Maximum Parsimony) ή η μέγιστη πιθανοφάνεια (Maximum Likelihood – Αλγόριθμος PhyML).

## 2.1.4 Usearch

Στη σημερινή εποχή όπου η βιολογική πληροφορία συσσωρεύεται συνεχώς με ταχείς ρυθμούς είναι απαραίτητη η χρήση αλγορίθμων που μπορούν ταχύτατα να πραγματοποιούν αναζητήσεις και ομαδοποιήσεις μέσα στον τεράστιο αυτό όγκο πληροφορίας. Το Usearch μέσω του αλγορίθμου Uclust έχει την ικανότητα σε πολύ σύντομο χρόνο να ομαδοποιεί πρωτεϊνικές ή νουκλεοτιδικές αλληλουχίες με βάση την ομοιότητά τους και να κρατάει έναν αντιπρόσωπο από κάθε ομάδα (Edgar, 2010).

## 2.1.5 MEGA - X

Το λογισμικό MEGAX (Molecular Evolutionary Genetics Analysis) μπορεί να χρησιμοποιηθεί σε όλα τα λειτουργικά συστήματα (Windows, Linux, MacOS) και παρέχει χρήσιμα εργαλεία για την εξερεύνηση και ανάλυση πρωτεϊνικών και DNA αλληλουχιών από εξελικτική σκοπιά. Μεταξύ των δυνατοτήτων που παρέχονται από το λογισμικό είναι η πολλαπλή στοίχιση αλληλουχιών μέσω των αλγορίθμων ClustalW και MUSCLE, ο υπολογισμός των εξελικτικών αποστάσεων και η δημιουργία φυλογενετικών δένδρων (Kumar, Stecher, Li, Knyaz, and Tamura 2018).

## 2.1.6 Gblocks

Το Gblocks είναι ένα υπολογιστικό εργαλείο που επιτρέπει την απομόνωση και επιλογή των συντηρημένων περιοχών πολλαπλών στοιχισμένων αλληλουχιών DNA ή πρωτεϊνών με σκοπό τη χρήση τους για φυλογενετική ανάλυση (Jose Castresana 2002).

## 2.1.7 Visual TreeCmp

Το Visual TreeCmp είναι ένα λογισμικό που σχεδιάστηκε για να υπολογίσει τις διαφορές μεταξύ φυλογενετικών δέντρων. Επιτρέπει τη σύγκριση σχετικά μεγάλων δέντρων μέσω της εφαρμογής μετρικών αλγορίθμων όπως το Robinson – Foulds που χρησιμοποιήθηκε στην παρούσα εργασία (Bogdanowicz 2020).

## 2.1.8 Dendroscope 3

Το Dendroscope 3 είναι ένα λογισμικό για επεξεργασία ένριζων φυλογενετικών δένδρων και δικτύων. Παρέχει πληθώρα μεθόδων για τη σχεδίαση και σύγκριση φυλογενετικών δέντρων, όπως ο αλγόριθμος tanglegram που χρησιμοποιήθηκε στην παρούσα εργασία (Huson and Scornavacca 2012).

## 2.2 Λήψη γονιδιωμάτων

Για την εργασία χρησιμοποιήθηκαν γονιδιώματα κορωνοϊών που αποκτήθηκαν από την GenBank (Benson et al., 2013), χρησιμοποιώντας ως λέξεις κλειδιά για την αναζήτηση τους όρους “Alphacoronavirus complete genome”, “Betacoronavirus complete genome”, “Gammacoronavirus complete genome” και “Deltacoronavirus complete genome”. Οι αλληλουχίες που αποκτήθηκαν φιλτραρίστηκαν σε επίπεδο νουκλεοτιδικού μήκους, με τη χρήση perl script, θέτοντας ως κατώτατο όριο τις 20.000 βάσεις. Ως εκ τούτου, ανακτήθηκαν συνολικά 2.912 αλληλουχίες ( α-CoV - 1026, β-CoV – 1334, γ-CoV – 415 και δ-CoV – 137). Επίσης, από την GenBank ανακτήθηκαν και οι αντίστοιχες κωδικές αλληλουχίες και πληροφορίες όπου ήταν διαθέσιμες για κάθε αλληλουχία.

## 3 ΑΠΟΤΕΛΕΣΜΑΤΑ

### 3.1 Απομόνωση γονιδίων από τις κωδικές αλληλουχίες

Από τις κωδικές αλληλουχίες που λήφθηκαν, πραγματοποιήθηκε απομόνωση των γονιδίων που κωδικοποιούν για polyprotein ORF1ab, ORF1a, ORF1b, spike, envelope, membrane και nucleocapsid για κάθε γένος κορωνοϊών με τη χρήση perl script για περαιτέρω μελέτη καθώς αυτές οι πρωτεΐνες έχουν κοινή ύπαρξη και στα 4 γένη κορωνοϊών.

### 3.2 Φιλτράρισμα αλληλουχιών μέσω της πρωτεΐνης Spike

Όπως αναφέρθηκε και στην εισαγωγή, η πρωτεΐνη Spike είναι από τις πιο σημαντικές στο γονιδίωμα των κορωνοϊών καθώς είναι υπεύθυνες για την είσοδο του ιού στο κύτταρο-ξενιστή. Επίσης, η αμινοξική ποικιλία που εμφανίζουν είναι η αιτία για την ύπαρξη πληθώρας διαφορετικών ενδιάμεσων οργανισμών ξενιστών μεταξύ αυτών οι μωσχογαλές, κυνόμορφα, καμήλες, βοειδή, τρωκτικά κ.α. (Li, 2016). Γι' αυτό, οι πρωτεϊνικές αλληλουχίες των Spikes δώθηκαν ως αρχείο εισόδου στο Usearch και πραγματοποιήθηκε clustering με ποσοστό ομοιότητας 90% παίρνοντας ως αποτέλεσμα 193 αλληλουχίες. Ως εκ τούτου, με βάση αυτές τις 193 αλληλουχίες, απομονώθηκαν και οι αντίστοιχες κωδικές αλληλουχίες για τις πρωτεΐνες polyprotein, envelope, membrane και nucleocapsid.

<b>Spike</b>		
	FULL	90,00%
<b>Alphacoronavirus</b>	1028	82
<b>Betacoronavirus</b>	1270	51
<b>Gammacoronavirus</b>	418	48
<b>Deltacoronavirus</b>	140	13

Πίνακας 3.1 Πλήθος γονιδιωμάτων πριν και μετά το clustering με βάση την πρωτεΐνη Spike.

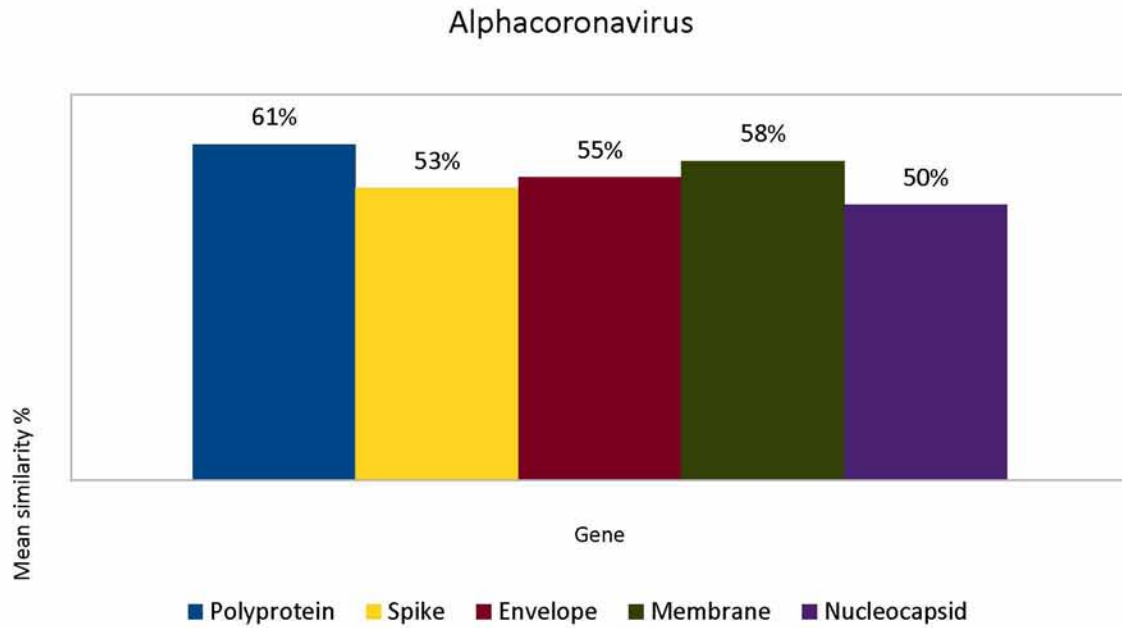
### 3.3 Πολλαπλή στοίχιση

Τα 5 γονίδια για κάθε γένος κορωνοϊών υποβλήθηκαν σε πολλαπλή στοίχιση στο λογισμικό SeaView μέσω του αλγορίθμου MUSCLE. Εν συνεχεία, αυτές οι στοιχήσεις εισήχθησαν στο online tool GBLOCKS με στόχο την απομόνωση των καλά συντηρημένων περιοχών του κάθε γονιδίου ώστε οι περαιτέρω φυλογενετικές αναλύσεις να εξάγουν πιο αντιπροσωπευτικά αποτελέσματα.

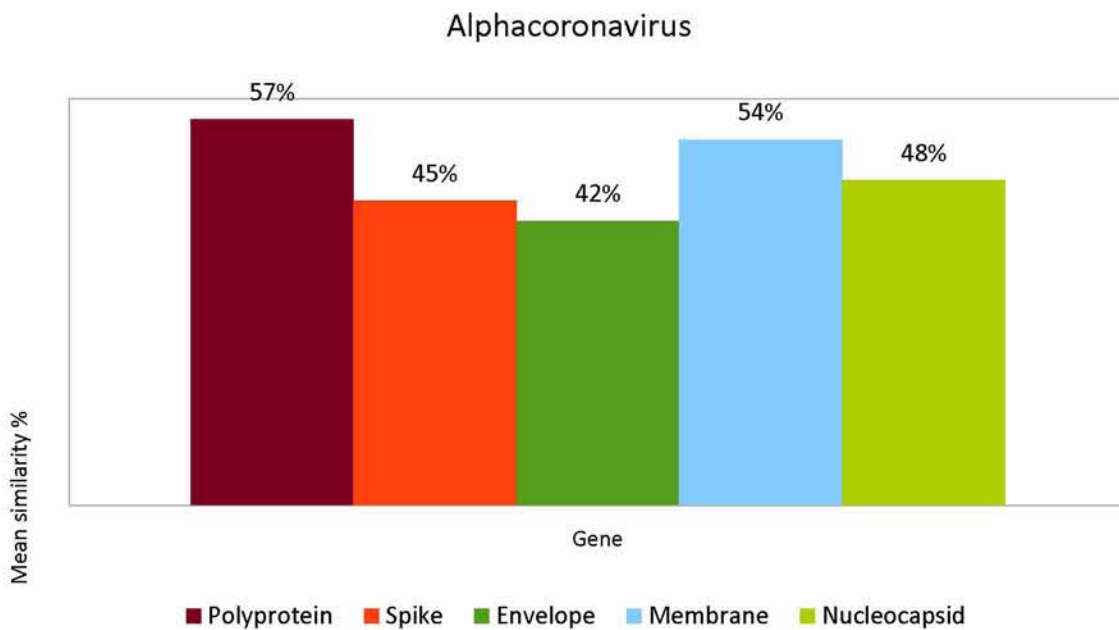
### 3.4 Υπολογισμός γενετικών αποστάσεων

Στη συνέχεια, με τη χρήση του λογισμικού MEGA - X, υπολογίσθηκαν οι μέσες τιμές των αποστάσεων τόσο σε αμινοξικό όσο και σε νουκλεοτιδικό επίπεδο. Τόσο στο νουκλεοτιδικό όσο και στο αμινοξικό επίπεδο, υπολογίστηκε ο μέσος όρος των διαφορών για κάθε περιοχή των γονιδίων (p-differences), ενώ χρησιμοποιήθηκε και η παράμετρος pairwise deletion, δηλαδή η διαγραφή των κενών για κάθε ζευγάρι αλληλουχιών. Οι τιμές (x) που ελήφθησαν αφαιρέθηκαν από τη μονάδα (1 - x) ώστε να πάρουμε το τελικό μέσο ποσοστό ομοιότητας των αλληλουχιών το οποίο και παρουσιάζεται στα γραφήματα που ακολουθούνε.

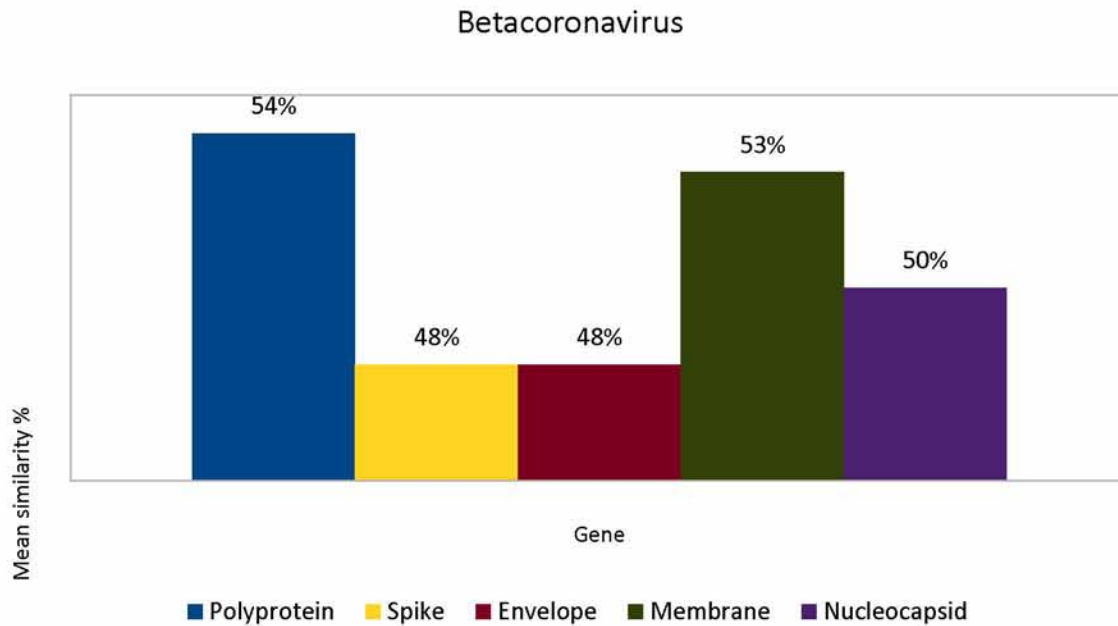




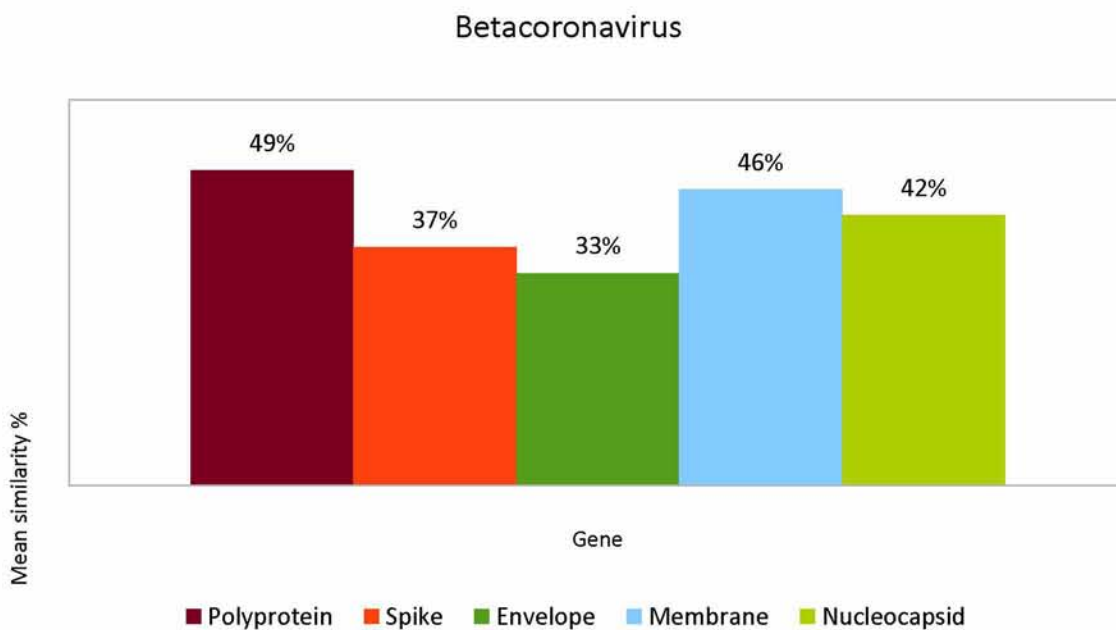
Εικόνα 3.2 Μέσος όρος ομοιότητας των νουκλεοτιδικών αλληλουχιών των γονιδίων που κωδικοποιούν για τις πρωτεΐνες: polyprotein, spike, envelope, membrane και Nucleocapsid, των ιών του γένους Alphacoronavirus.



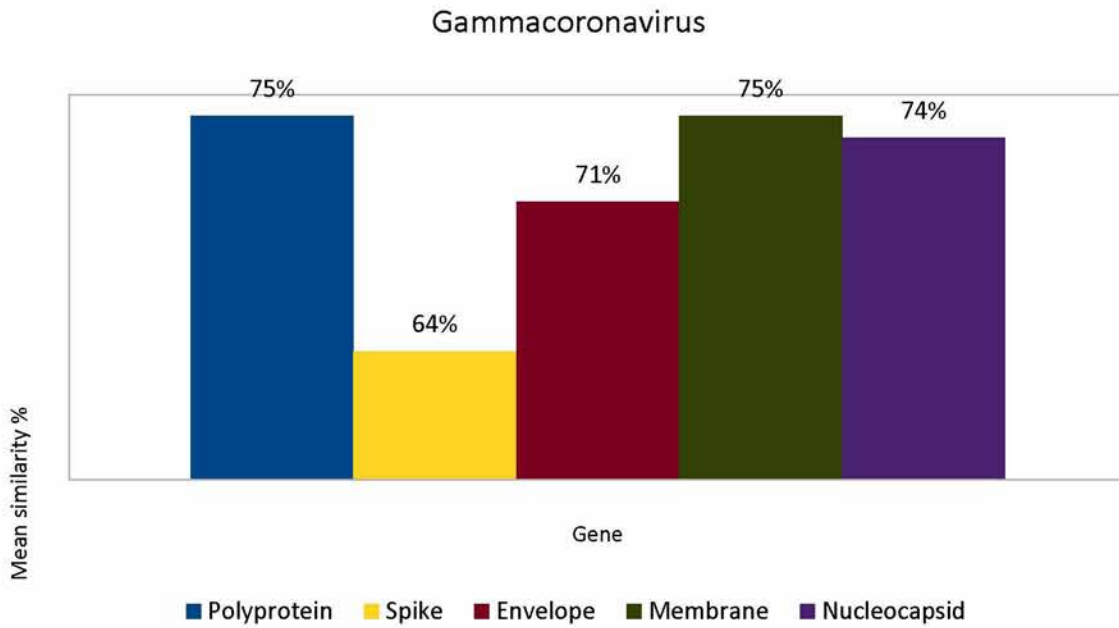
Εικόνα 3.3 Μέσος όρος ομοιότητας των αμινοξικών αλληλουχιών των πρωτεϊνών: polyprotein, spike, envelope, membrane, nucleocapsid των ιών του γένους Alphacoronavirus.



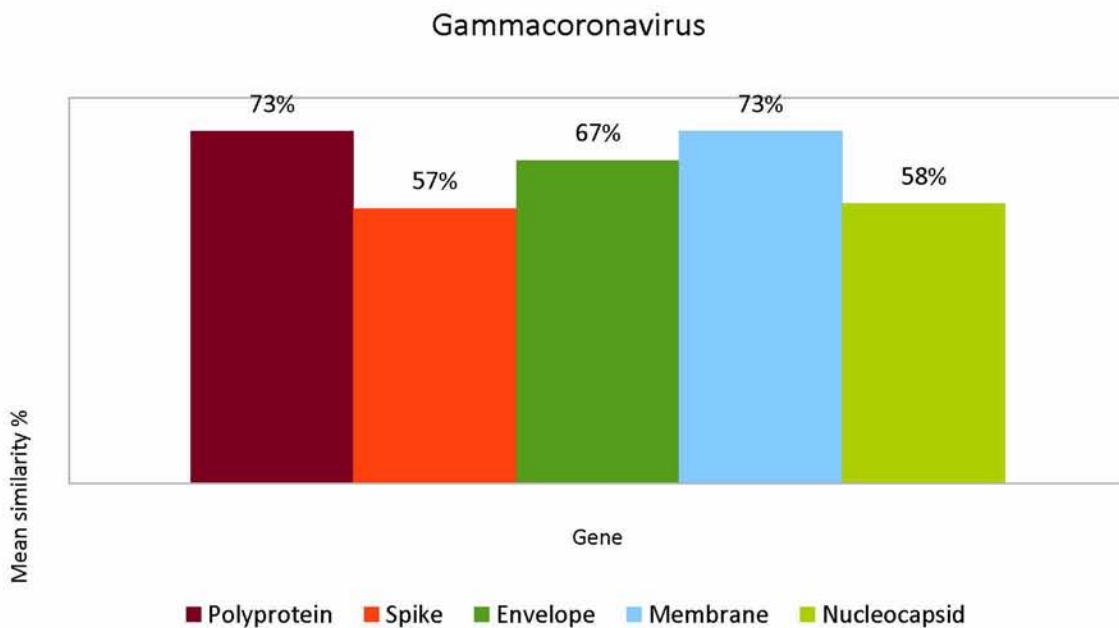
Εικόνα 3.4 Μέσος όρος ομοιότητας των νουκλεοτιδικών αλληλουχιών των γονιδίων που κωδικοποιούν για τις πρωτεΐνες: polyprotein, spike, envelope, membrane και Nucleocapsid, των ιών του γένους Betacoronavirus.



Εικόνα 3.5 Μέσος όρος ομοιότητας των αμινοξικών αλληλουχιών των πρωτεϊνών: polyprotein, spike, envelope, membrane, nucleocapsid των ιών του γένους Betacoronavirus.

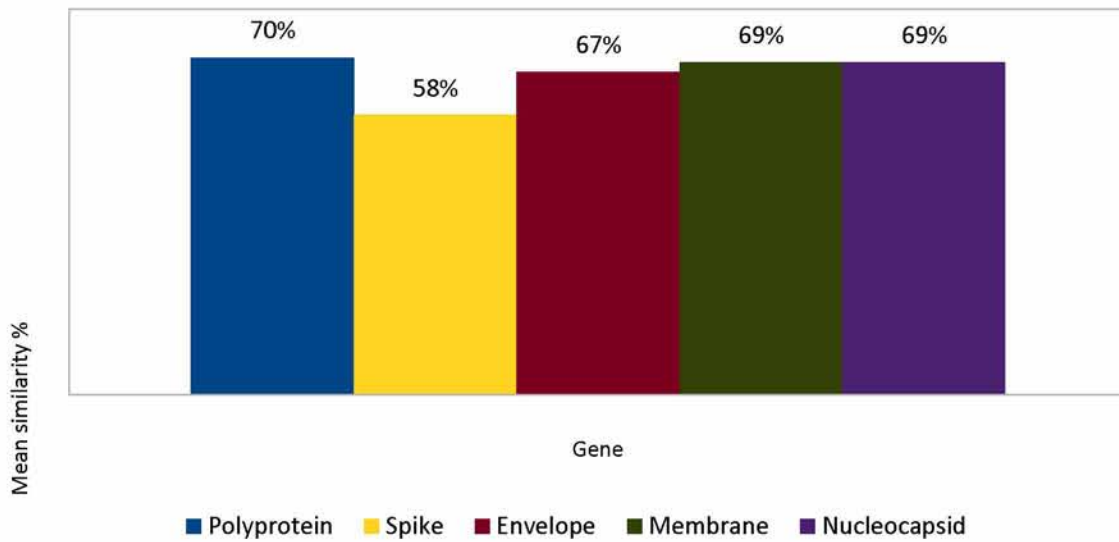


Εικόνα 3.6 Μέσος όρος ομοιότητας των νουκλεοτιδικών αλληλουχιών των γονιδίων που κωδικοποιούν για τις πρωτεΐνες: polyprotein, spike, envelope, membrane και Nucleocapsid, των ιών του γένους Gammacoronavirus.



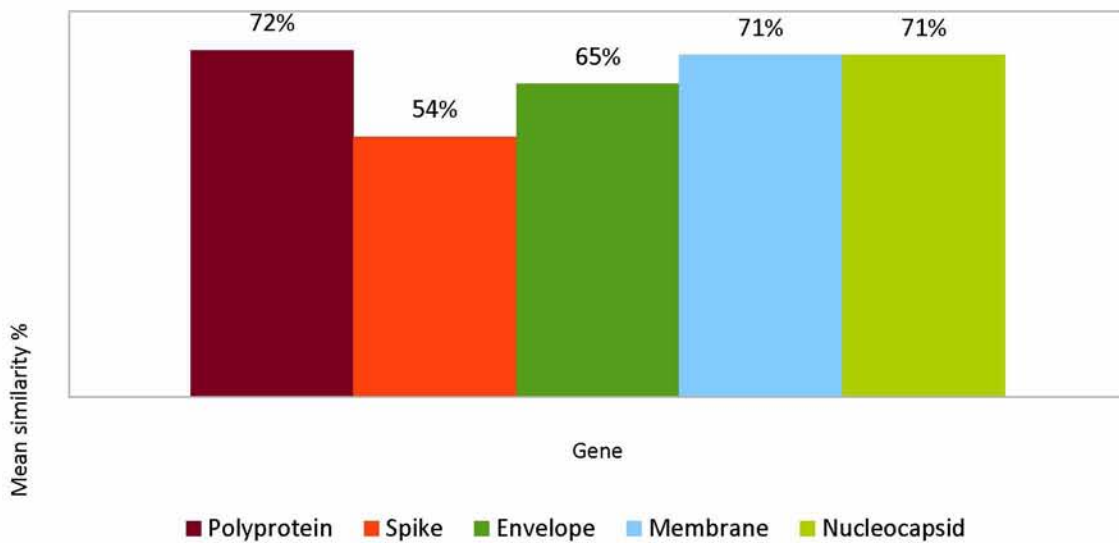
Εικόνα 3.7 Μέσος όρος ομοιότητας των αμινοξικών αλληλουχιών των πρωτεϊνών: polyprotein, spike, envelope, membrane, nucleocapsid των ιών του γένους Gammacoronavirus.

### Deltacoronavirus



Εικόνα 3.8 Μέσος όρος ομοιότητας των νουκλεοτιδικών αλληλουχιών των γονιδίων που κωδικοποιούν για τις πρωτεΐνες: polyprotein, spike, envelope, membrane και Nucleocapsid, των ιών του γένους Deltacoronavirus.

### Deltacoronavirus



Εικόνα 3.9 Μέσος όρος ομοιότητας των αμινοξικών αλληλουχιών των πρωτεϊνών: polyprotein, spike, envelope, membrane, nucleocapsid των ιών του γένους Deltacoronavirus.

### 3.5 Σύγκριση φυλογενετικών δέντρων

Το επόμενο βήμα της παρούσας εργασίας, ήταν η δημιουργία των φυλογενετικών δέντρων στο λογισμικό SeaView. Τα δένδρα κατασκευάστηκαν με μέθοδο αποστάσεων και πιο συγκεκριμένα με Bio-Neighbor Joining και το μοντέλο αποστάσεων του Poisson. Στη συνέχεια τα δένδρα για κάθε γονίδιο που κωδικοποιεί για τις πέντε σημαντικές πρωτεΐνες που εξετάζουμε, ORF1ab, ORF1a, ORF1b, spike, envelope, membrane και nucleocapsid υποβλήθηκαν σε μεταξύ τους σύγκριση στο διαδικτυακό εργαλείο Visual TreeCmp. Το σύστημα μέτρησης αποστάσεων μεταξύ των δέντρων που χρησιμοποιήθηκε είναι το Robinson – Foulds (RF), που κατά κοινή παραδοχή είναι το πιο ευρέως διαδεδομένο σύστημα μέτρησης για συγκρίσεις φυλογενετικών δέντρων (Robinson & Foulds, 1981). Όσο μικρότερος είναι ο αριθμός RF τόσο μεγαλύτερη ομοιότητα έχουν τα δύο φυλογενετικά δένδρα και εξαρτάται και από το πλήθος των αλληλουχιών που συγκρίνονται.

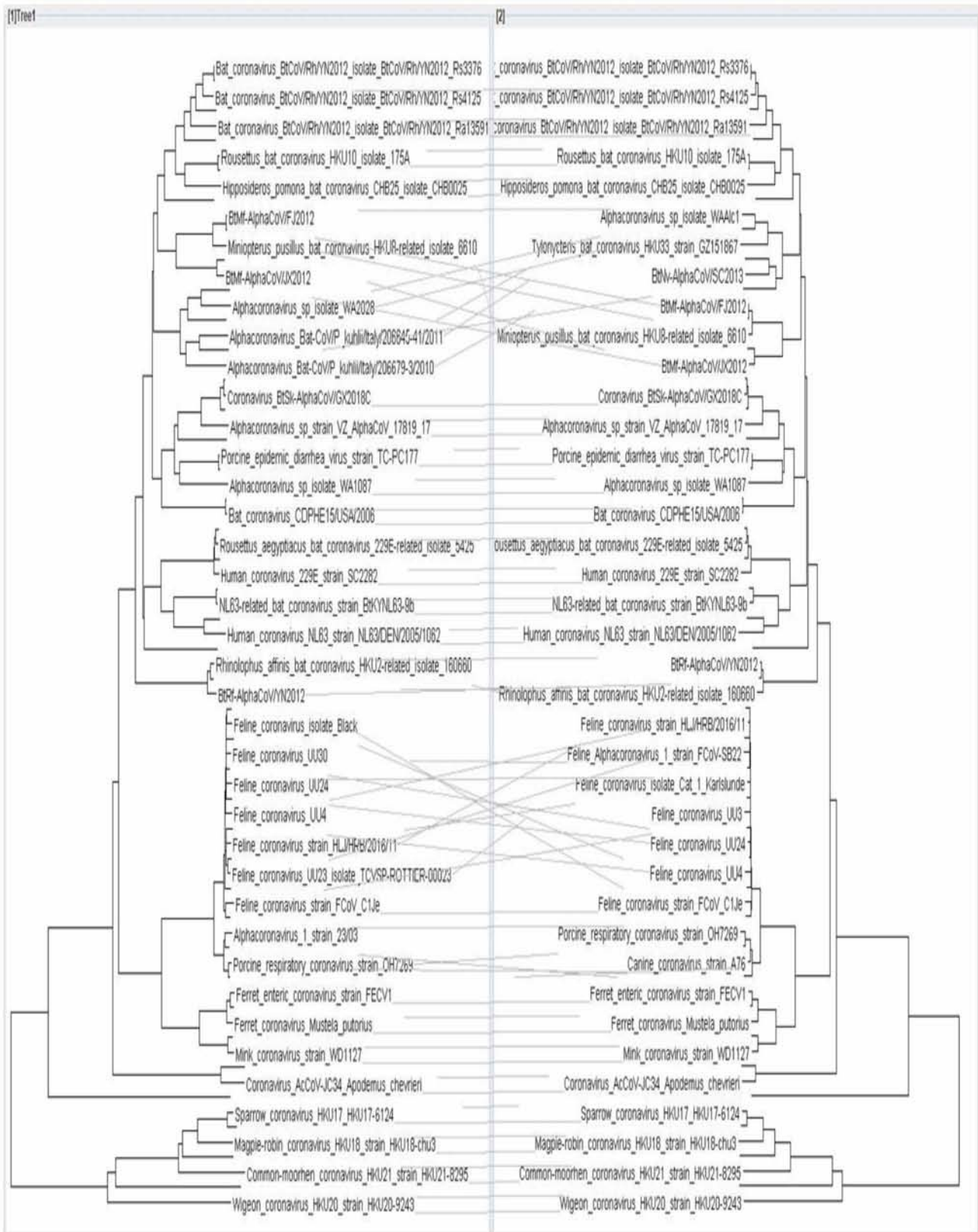
Παρακάτω, παρουσιάζονται τα συγκριτικά φυλογενετικά δένδρα των χαρακτηριστικών ιών για κάθε γένος όπως παρατέθηκαν στον πίνακα 1.2 αλλά και οι τιμές Robinson – Foulds των διαδοχικών γονιδίων που κωδικοποιούν για τις πέντε πρωτεΐνες που εμπλέκονται τα μέγιστα στον κύκλο ζωής των κορωνοϊών.

#### 3.5.1 Alphacoronavirus

Η πρώτη φυλογενετική σύγκριση δέντρων (tanglegram) με τη βοήθεια του λογισμικού προγράμματος Dendroscope 3 έγινε για τα δένδρα του ORF1a και ORF1b και ακολούθως υπολογίστηκε η τιμή Robinson – Foulds (RF) μέσω του διαδικτυακού εργαλείου Visual TreeCmp. Στη συνέχεια η διαδικασία αυτή ακολουθήθηκε για κάθε διαδοχικό σημαντικό γονίδιο του γονιδιώματος των κορωνοϊών. Στον πίνακα που ακολουθεί (Πίνακας 3.10) παραθέτονται οι τιμές RF.

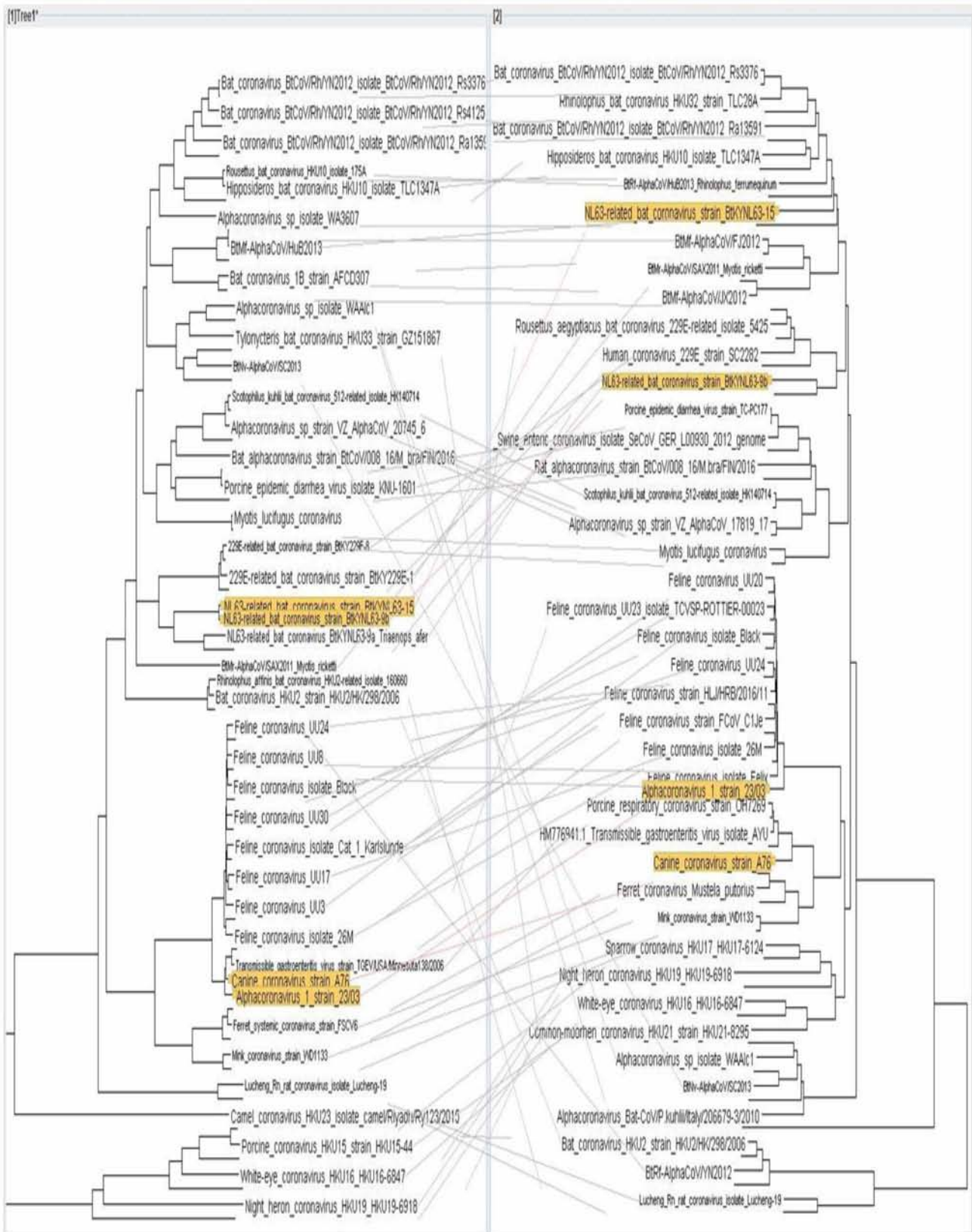
	ORF1ab	ORF1a	ORF1b	Spike	Envelope	Membrane
ORF1ab						
ORF1a	7					
ORF1b	13	15				
Spike	45	47	44			
Envelope	38	39	37	50		
Membrane	29	30	31	54	39	
Nucleocapsid	27	28	26	50	39	34

Πίνακας 3.10 Τιμές Robinson – Foulds των συγκριτικών φυλογενετικών δέντρων των ιών Alphacoronavirus.

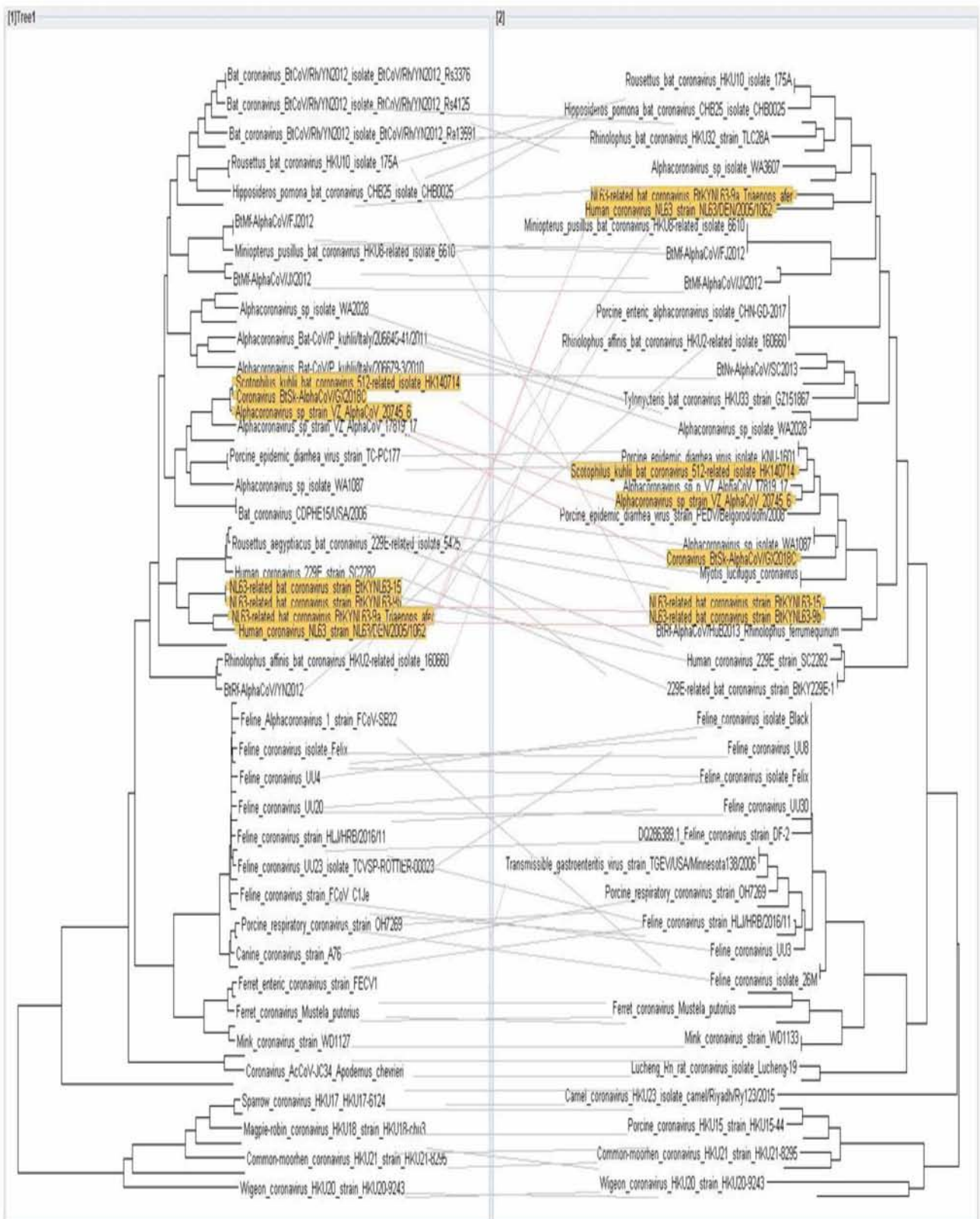


Εικόνα 3.11 Tanglegram των πρωτεϊνών ORF1a-ORF1b των ιών Alphacoronavirus.



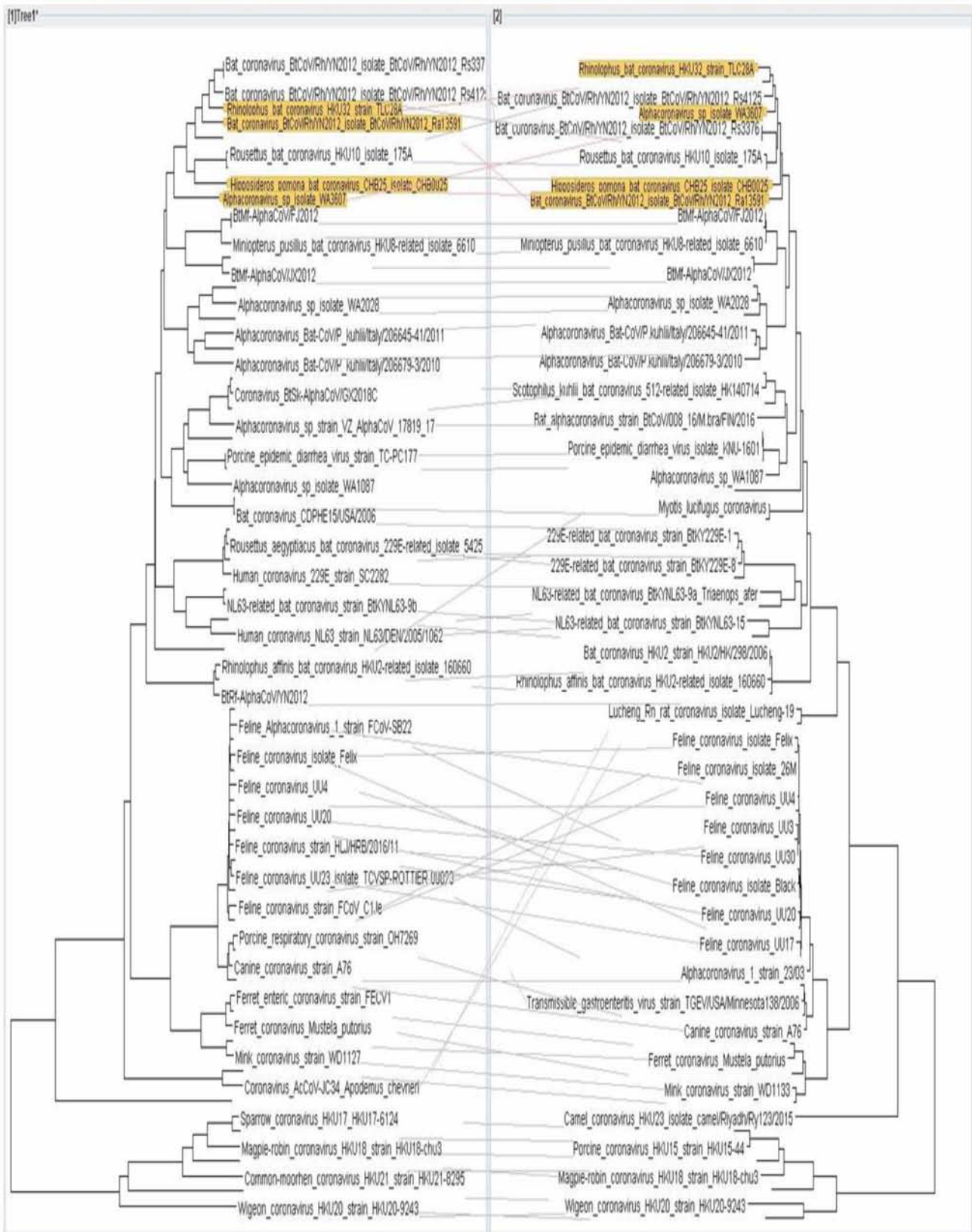


Εικόνα 3.12 Tanglegram των πρωτεϊνών ORF1ab - Spike των ιών Alphacoronavirus. Με κίτρινο χρώμα διακρίνονται στελέχη που εμφανίζουν εξελικτική διαφορά.

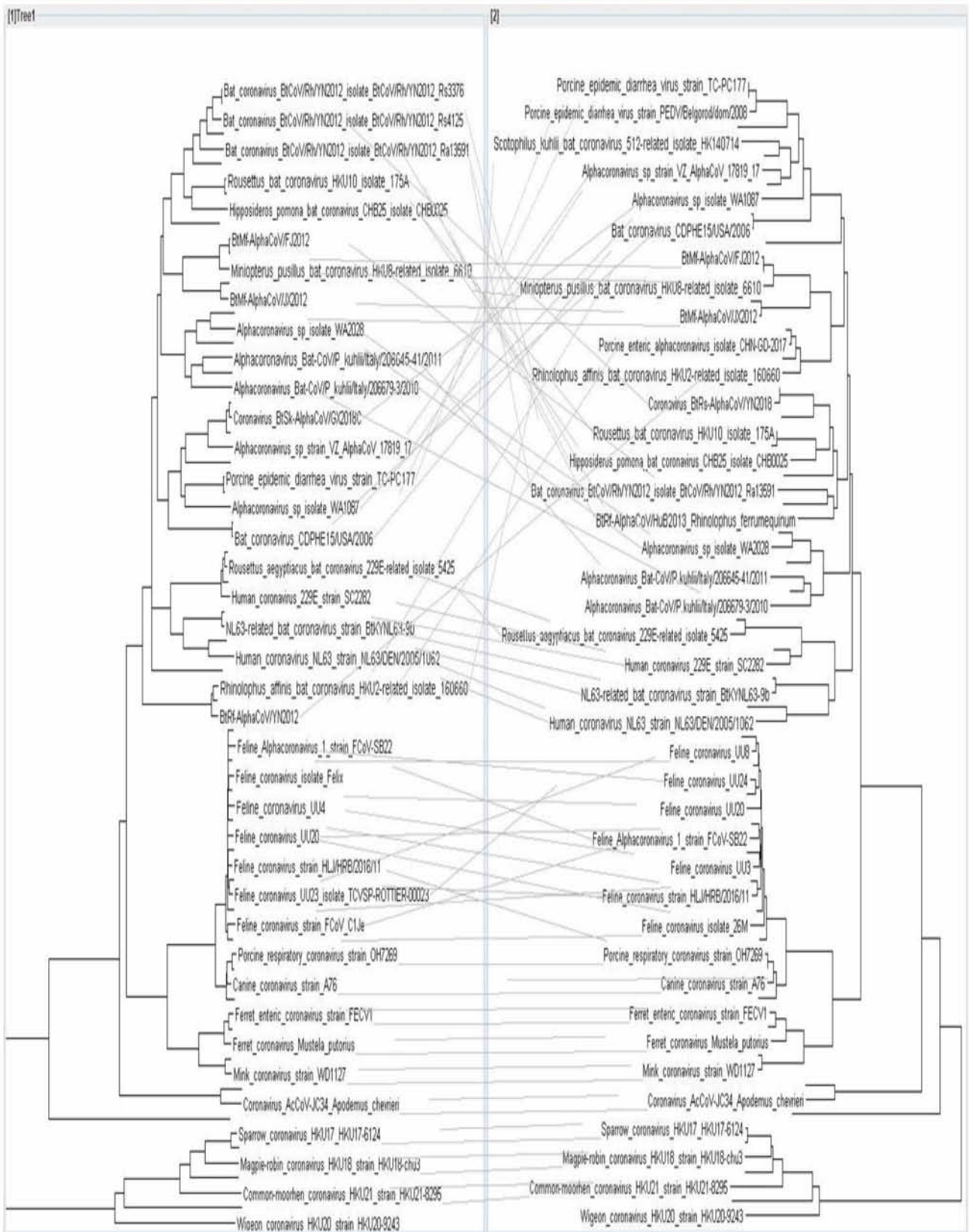


Εικόνα 3.13 Tanglegram των πρωτεϊνών ORF1ab - Envelope των ιών Alphacoronavirus. Με κίτρινο χρώμα διακρίνονται στελέχη που εμφανίζουν εξελικτική διαφορά.





Εικόνα 3.14 Tanglegram των πρωτεϊνών ORF1ab - Membrane των ιών Alpacoronavirus. Με κίτρινο χρώμα διακρίνονται στελέχη που εμφανίζουν εξελικτική διαφορά.



Εικόνα 3.15 Tanglegram των πρωτεϊνών ORF1ab - Nucleocapsid των ιών Alphacoronavirus.

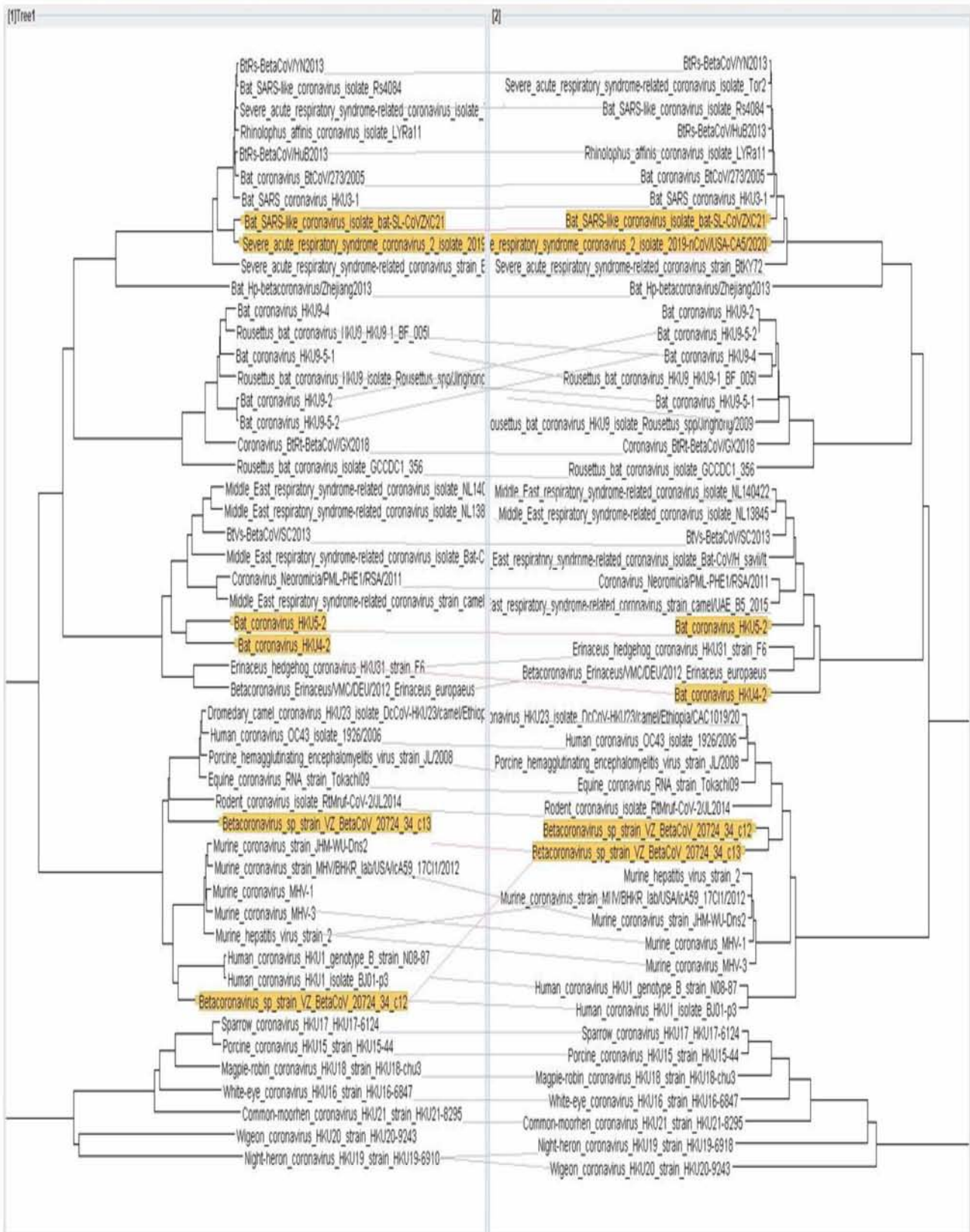
### 3.5.2 Betacoronavirus

Η ίδια διαδικασία που ακολουθήθηκε για τους ιούς του γένους Alphacoronavirus, ακολουθήθηκε και για τους ιούς Betacoronavirus αλλά και για τους Gammacoronavirus και Deltacoronavirus. Ακολουθεί ο πίνακας με τις τιμές Robinson – Foulds (πίνακας 3.18) και τα συγκριτικά φυλογενετικά δένδρα.

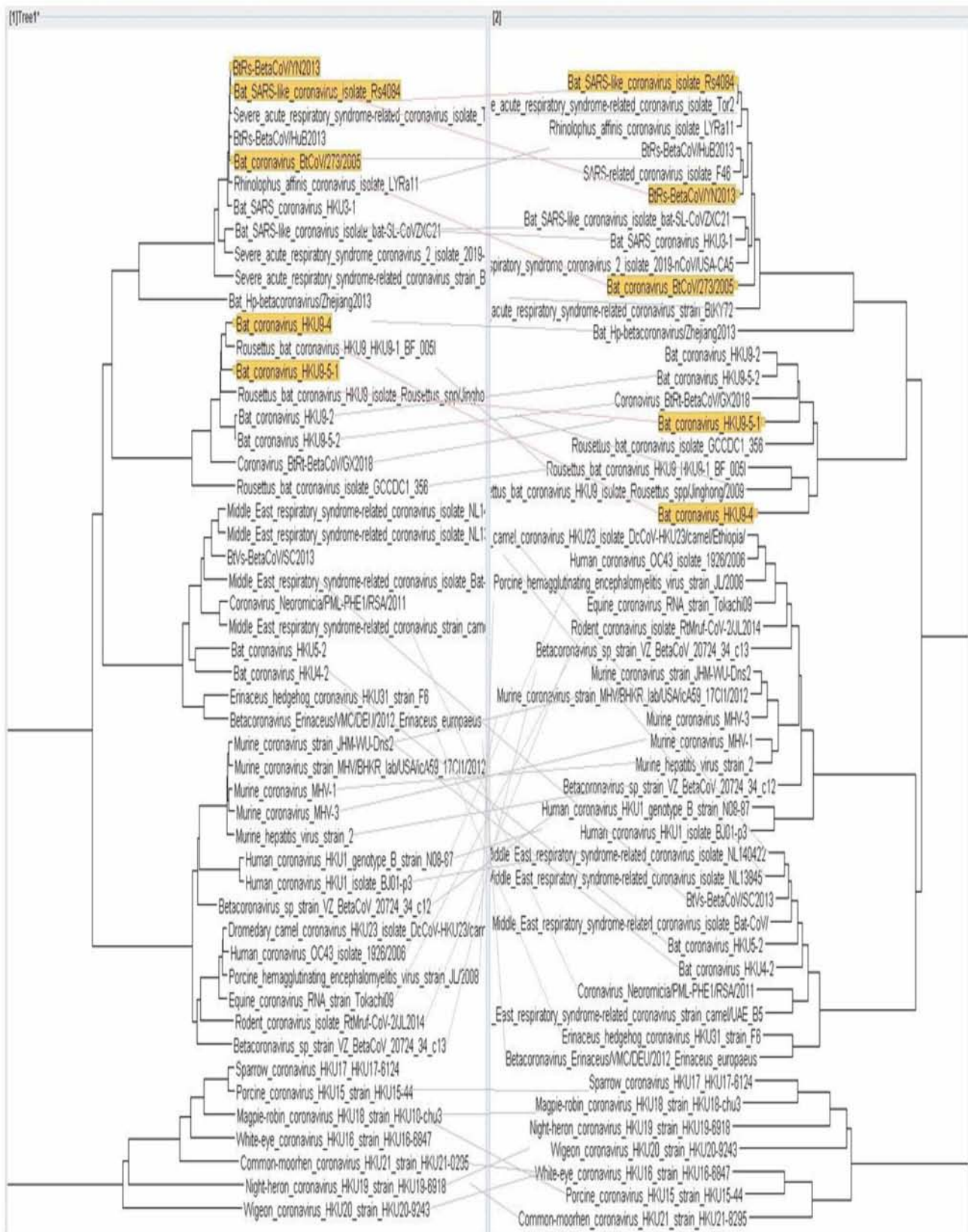
	ORF1ab	ORF1a	ORF1b	Spike	Envelope	Membrane
ORF1ab						
ORF1a	3					
ORF1b	14	15				
Spike	23	23	24			
Envelope	20	21	22	26		
Membrane	23	21	26	30	22	
Nucleocapsid	22	21	24	25	19	25

Πίνακας 3.16 Τιμές Robinson – Foulds των συγκριτικών φυλογενετικών δένδρων των ιών Betacoronavirus.



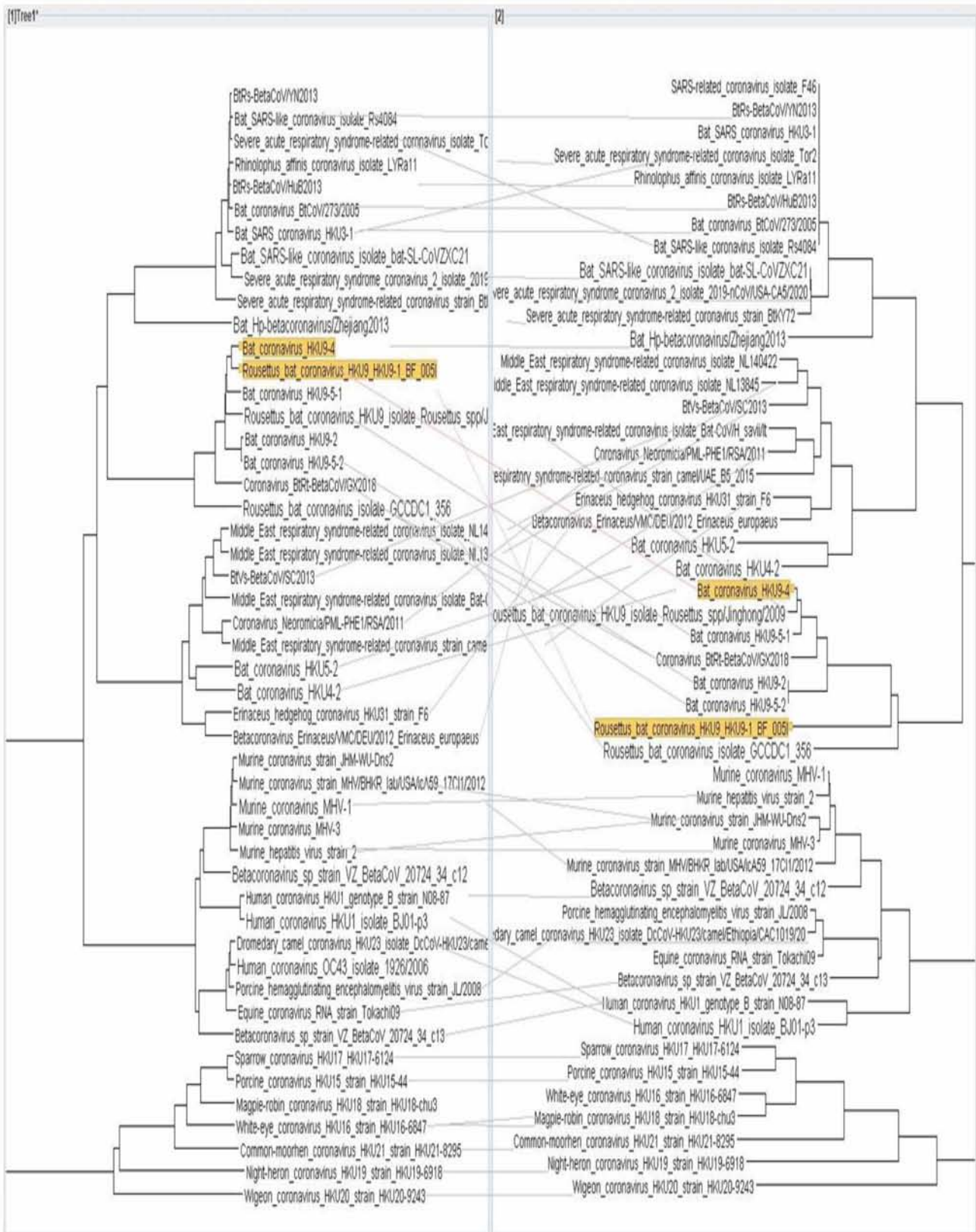


Εικόνα 3.17 Tanglegram των πρωτεϊνών ORF1a-ORF1b των ιών Betacoronavirus. Με κίτρινο χρώμα διακρίνονται στελέχη που εμφανίζουν εξελικτική διαφορά.

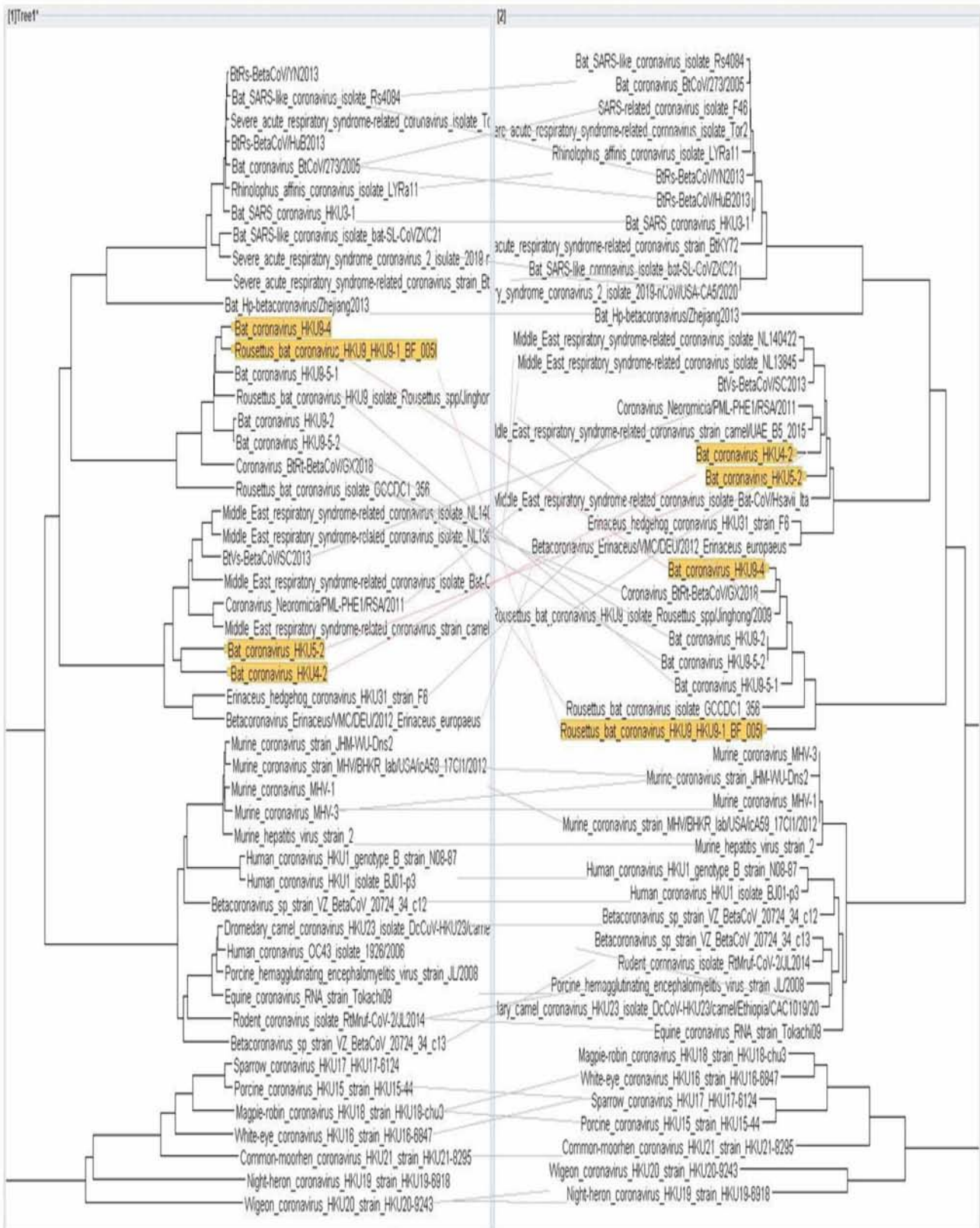


Εικόνα 3.18 Tanglegram των πρωτεϊνών ORF1ab - Spike των ιών Betacoronavirus. Με κίτρινο χρώμα διακρίνονται στελέχη που εμφανίζουν εξελικτική διαφορά.



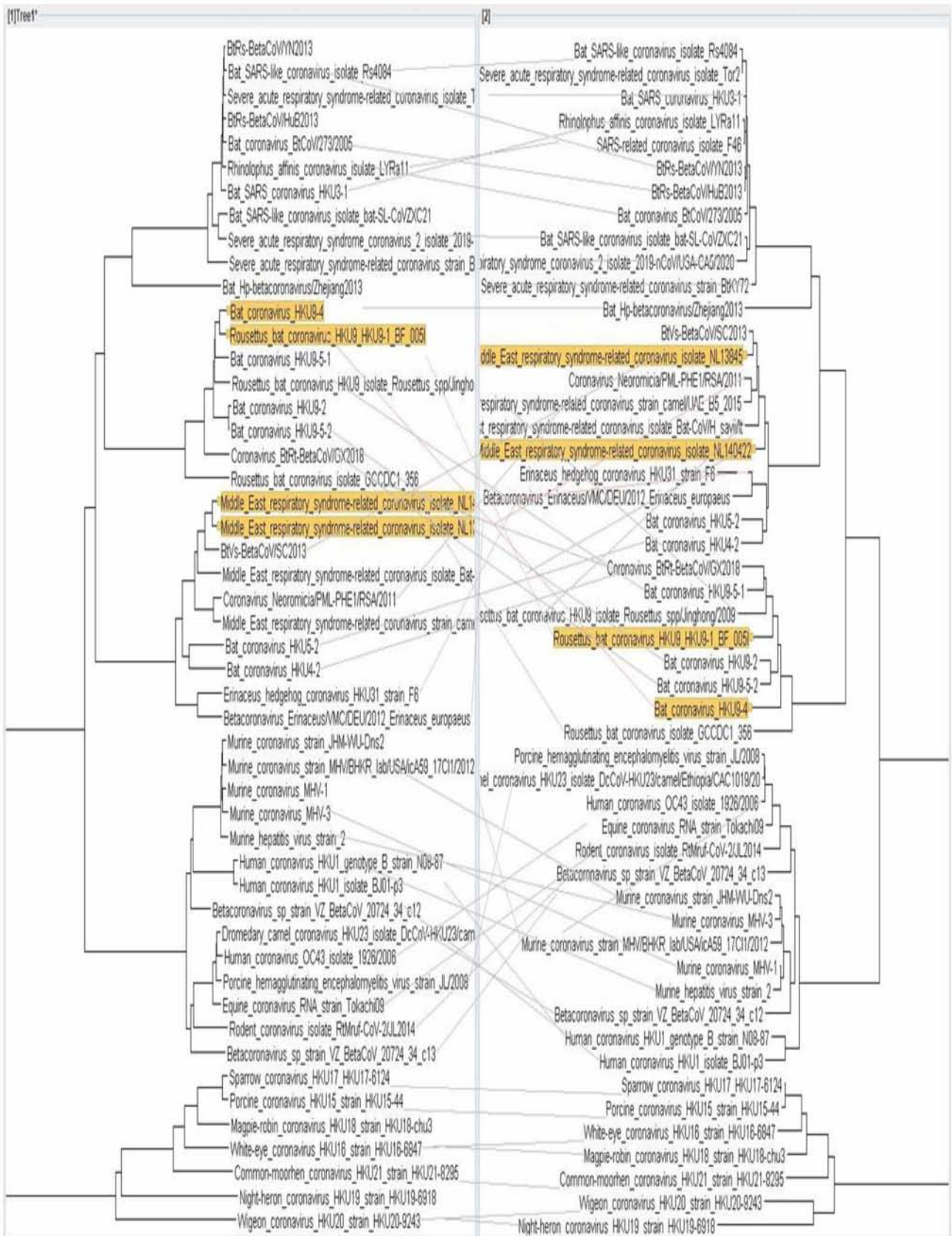


Εικόνα 3.19 Tanglegram των πρωτεϊνών ORF1ab - Envelope των ιών Betacoronavirus. Με κίτρινο χρώμα διακρίνονται στελέχη που εμφανίζουν εξελικτική διαφορά.



Εικόνα 3.20 Tanglegram των πρωτεϊνών ORF1ab - Membrane των ιών Betacoronavirus. Με κίτρινο χρώμα διακρίνονται στελέχη που εμφανίζουν εξελικτική διαφορά.





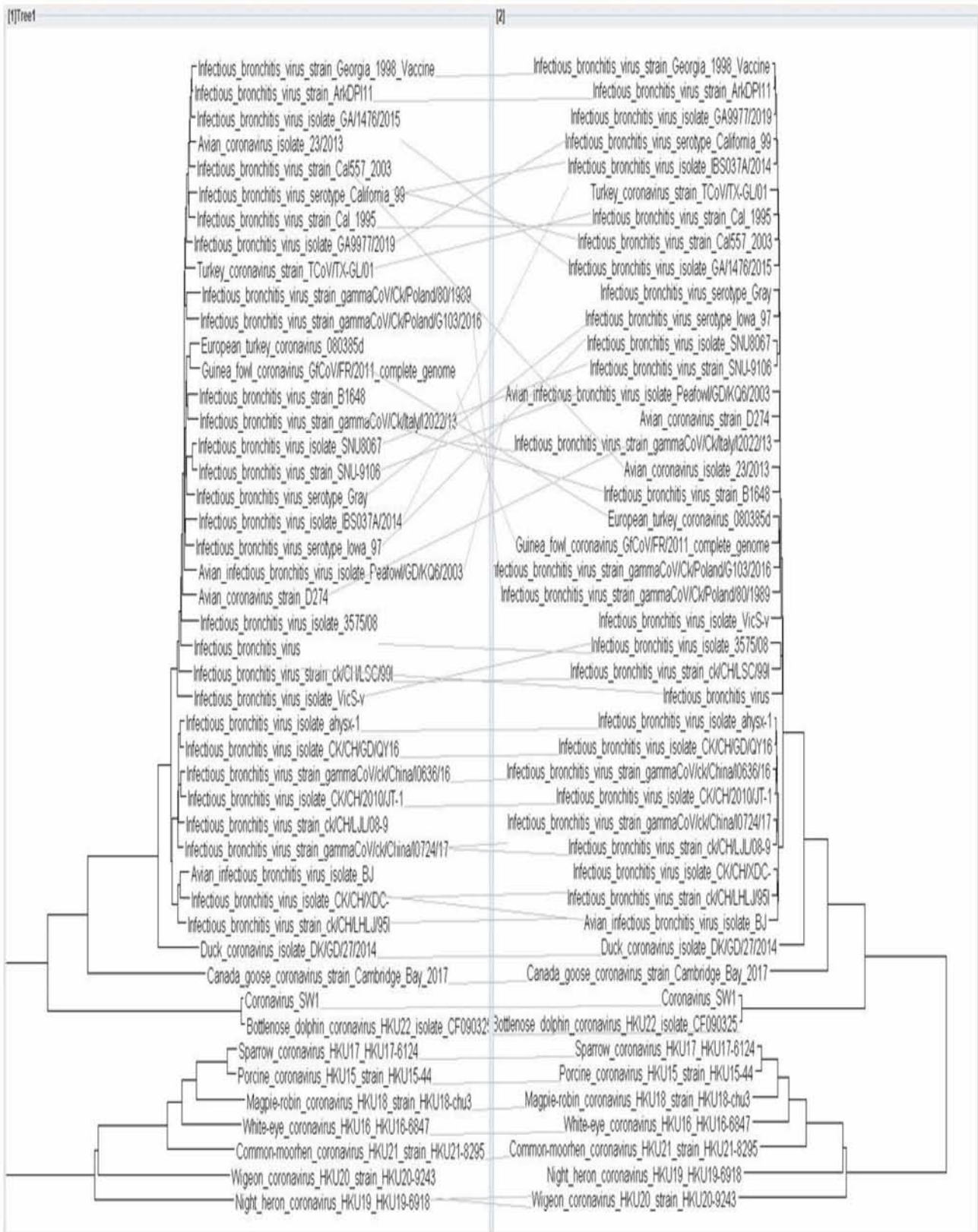
Εικόνα 3.21 Tanglegram των πρωτεϊνών ORF1ab - Nucleocapsid των ιών Betacoronavirus. Με κίτρινο χρώμα διακρίνονται στελέχη που εμφανίζουν εξελικτική διαφορά.



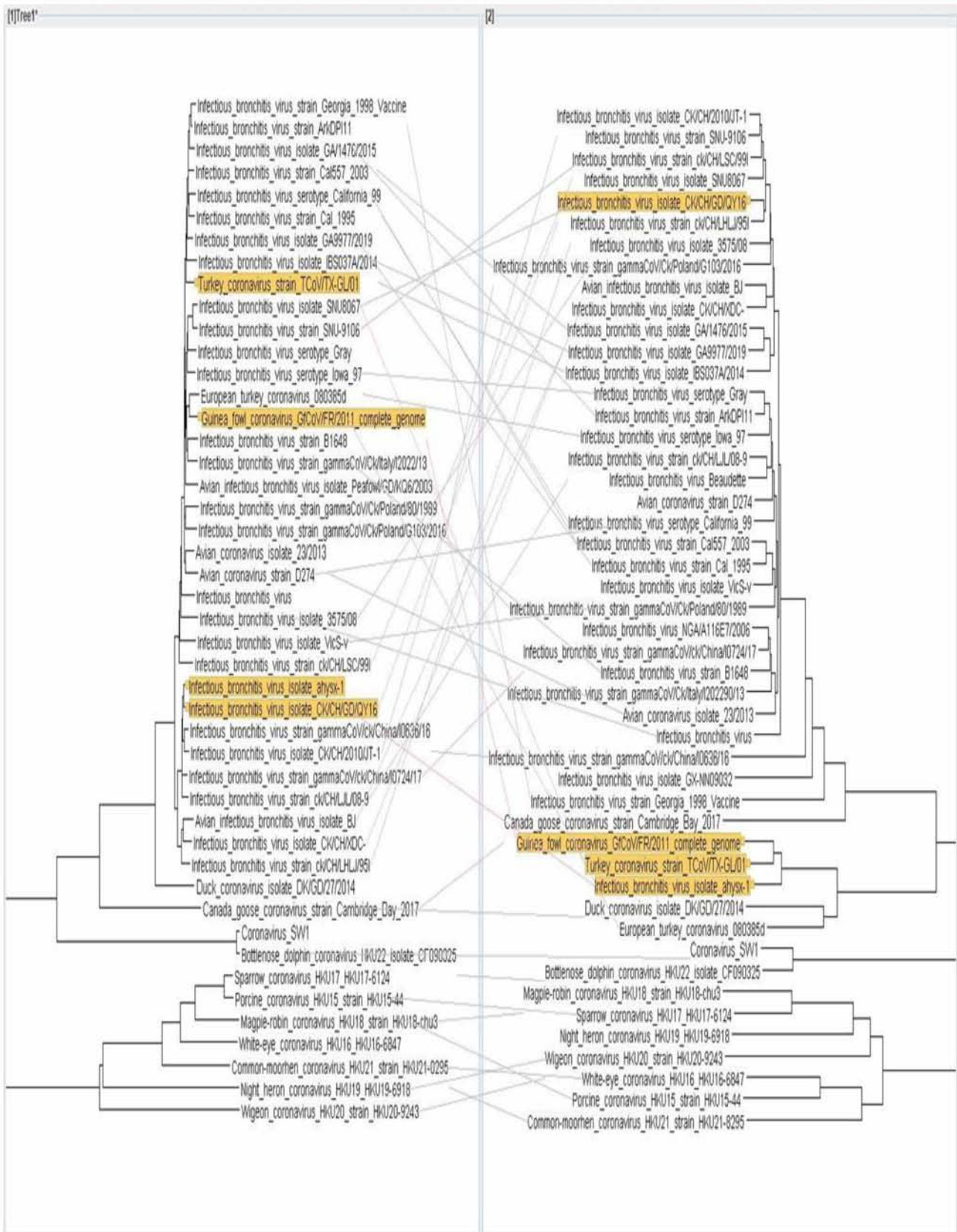
### 3.5.3 Gammacoronavirus

	ORF1ab	ORF1a	ORF1b	Spike	Envelope	Membrane
ORF1ab						
ORF1a	15					
ORF1b	21	23				
Spike	37	37	38			
Envelope	34	35	34	39		
Membrane	33	33	33	35	36	
Nucleocapsid	33	32	34	40	39	34

Πίνακας 3.22 Τιμές Robinson – Foulds των συγκριτικών φυλογενετικών δένδρων των ιών Gammacoronavirus.

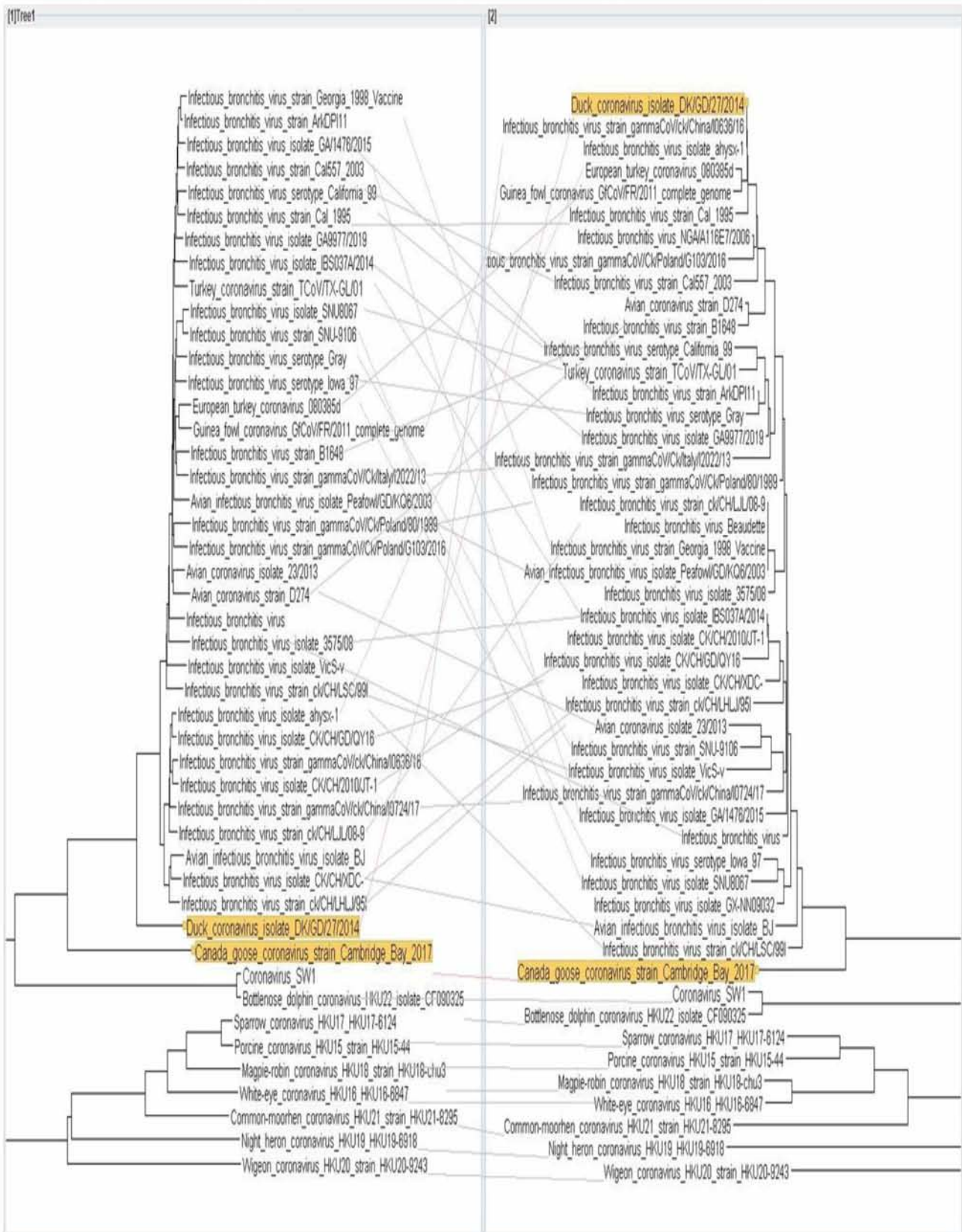


Εικόνα 3.23 Tanglegram των πρωτεϊνών ORF1a-ORF1b των ιών Gammacoronavirus.

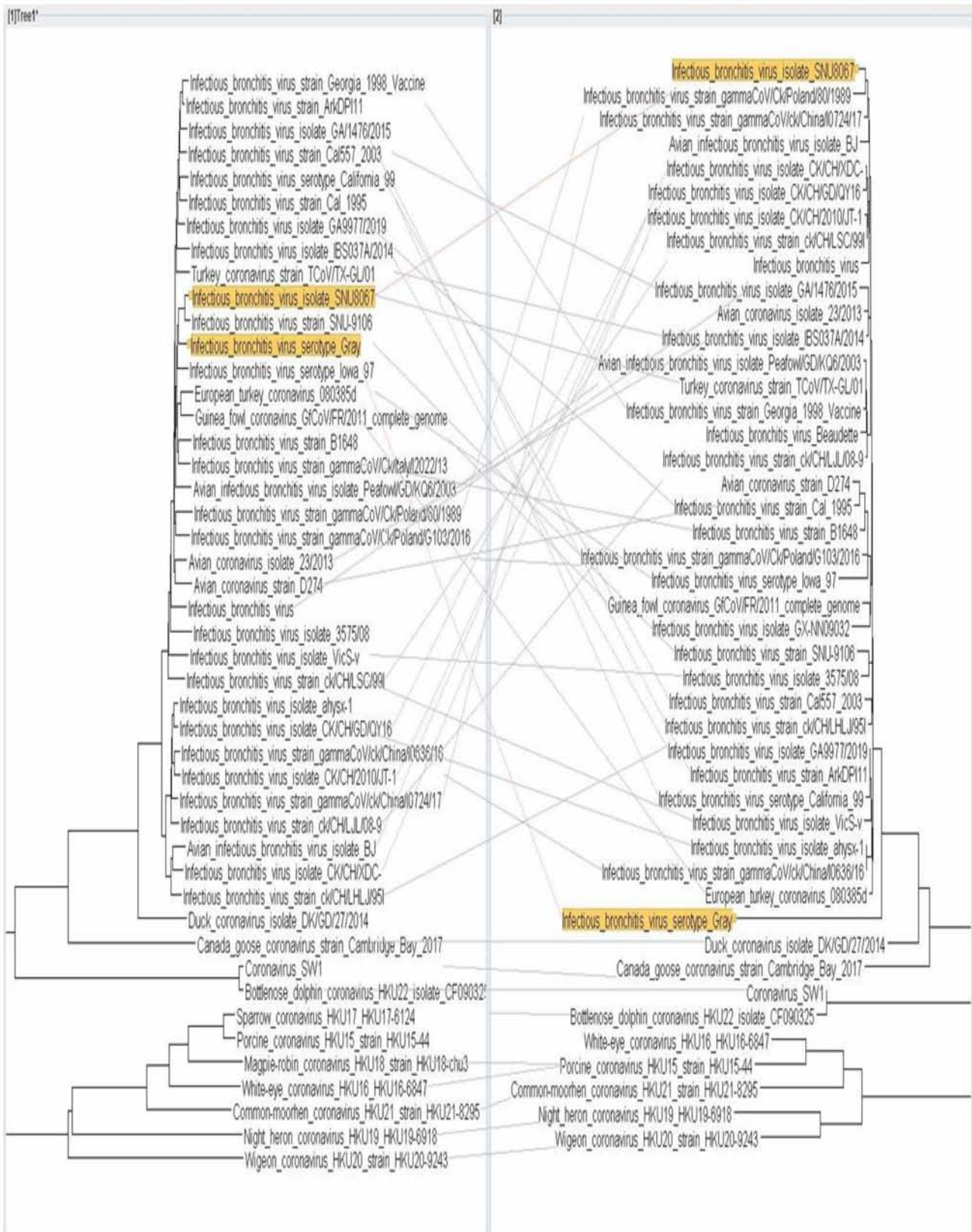


Εικόνα 3.24 Tanglegram των πρωτεϊνών ORF1ab - Spike των ιών Gammacoronavirus. Με κίτρινο χρώμα διακρίνονται στελέχη που εμφανίζουν εξελικτική διαφορά.



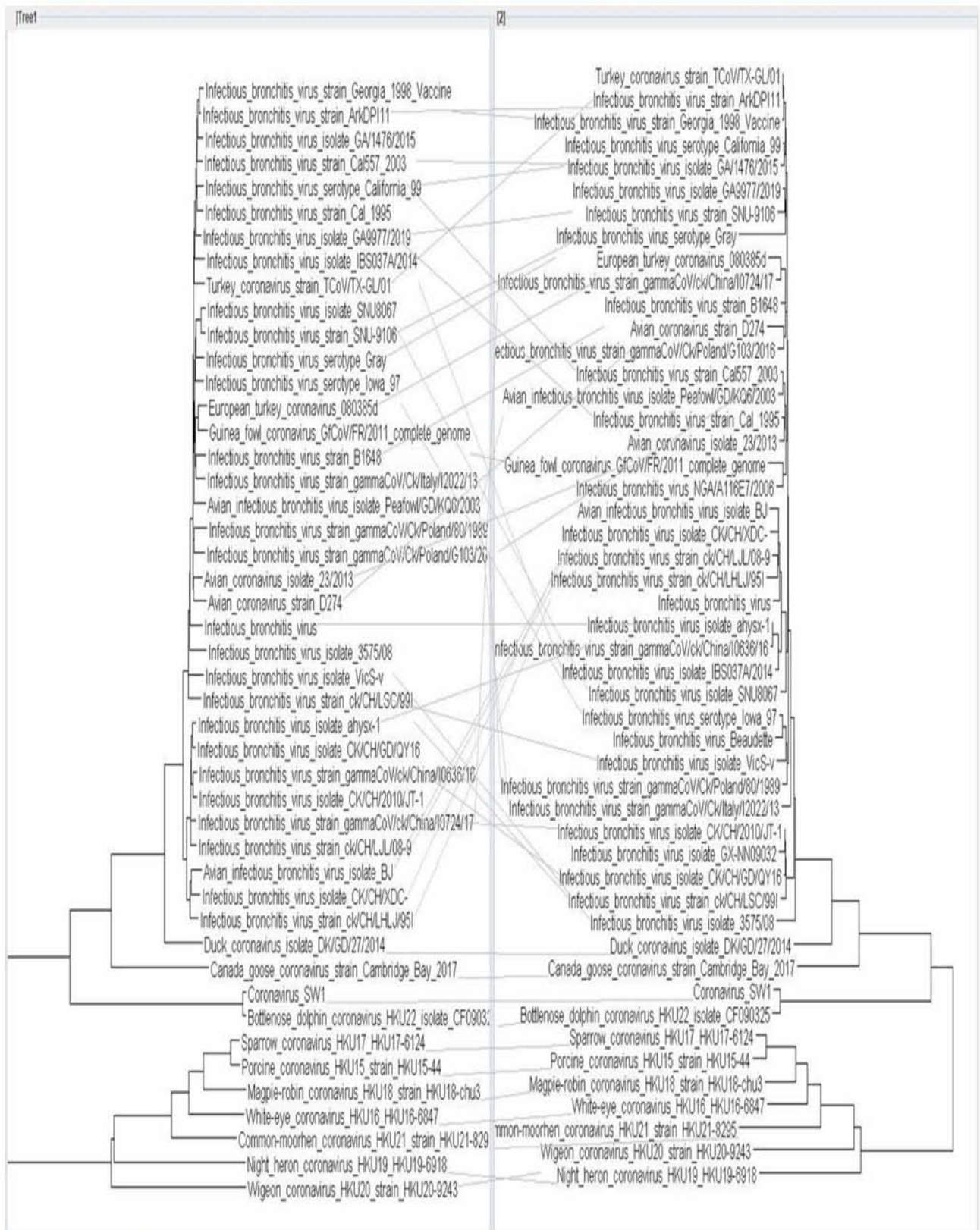


Εικόνα 3.25 Tanglegram των πρωτεϊνών ORF1ab – Envelope των ιών Gammacoronavirus. Με κίτρινο χρώμα διακρίνονται στελέχη που εμφανίζουν εξελικτική διαφορά.



Εικόνα 3.26 Tanglegram των πρωτεϊνών ORF1ab – Membrane των ιών Gammacoronavirus. Με κίτρινο χρώμα διακρίνονται στελέχη που εμφανίζουν εξελικτική διαφορά.



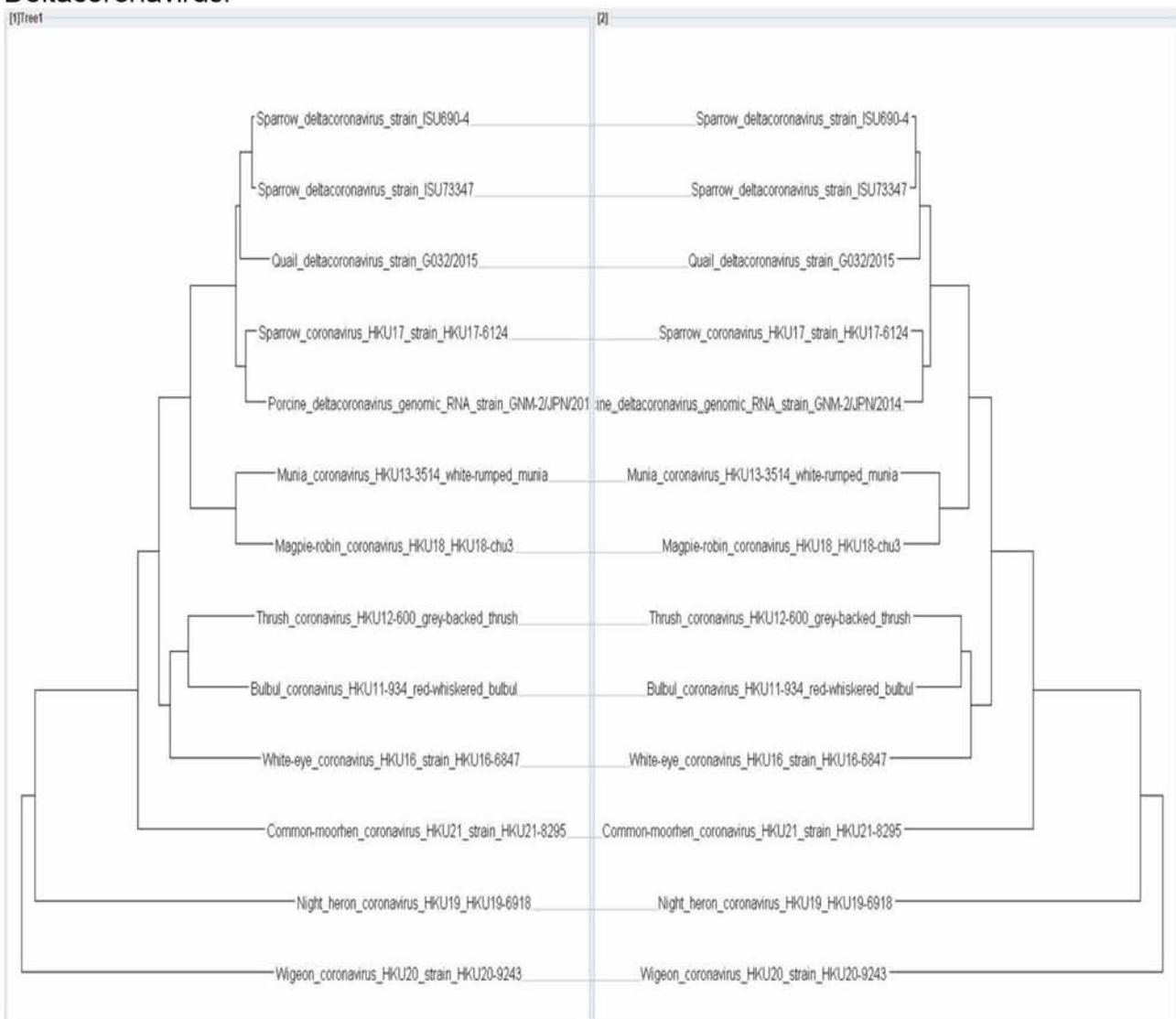


Εικόνα 3.27 Tanglegram των πρωτεϊνών ORF1ab – Nucleocapsid των ιών Gammacoronavirus.

### 3.5.4 Deltacoronavirus

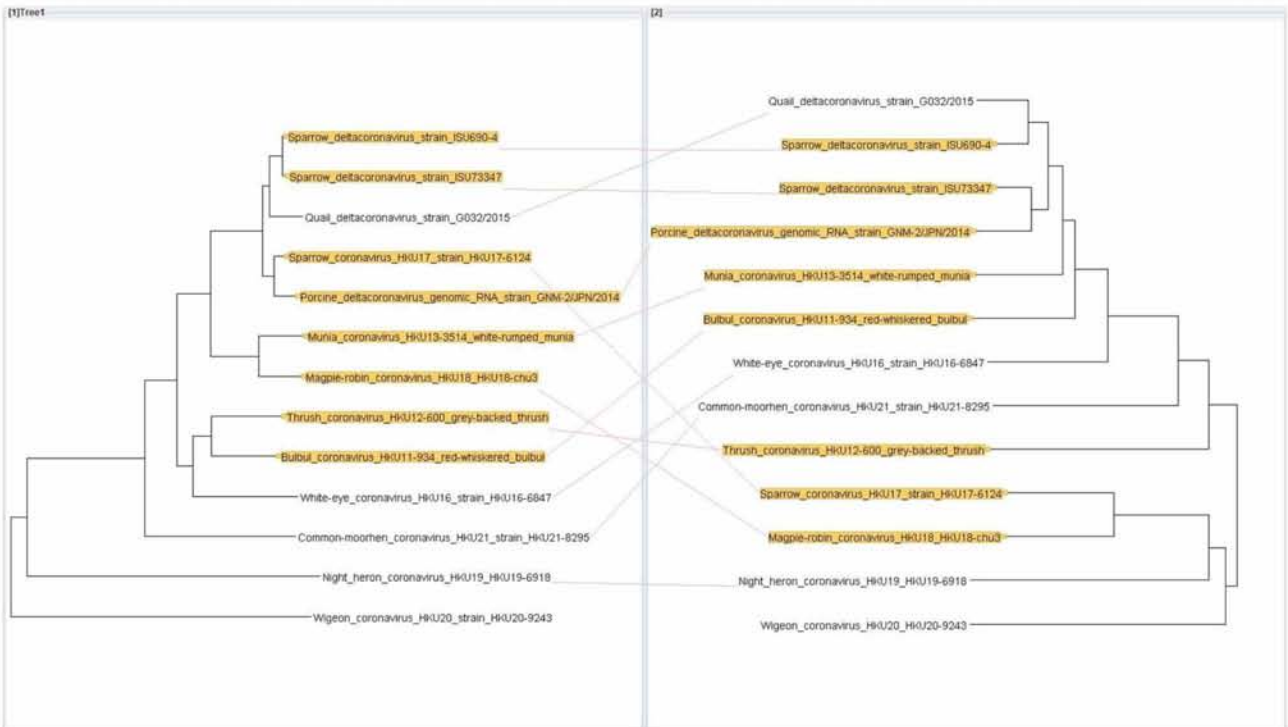
	ORF1ab	ORF1a	ORF1b	Spike	Envelope	Membrane
ORF1ab						
ORF1a	0					
ORF1b	0	0				
Spike	9	9	9			
Envelope	3	3	3	9		
Membrane	4	4	4	10	5	
Nucleocapsid	2	2	2	10	4	2

Πίνακας 3.28 Τιμές Robinson – Foulds των συγκριτικών φυλογενετικών δένδρων των ιών Deltacoronavirus.

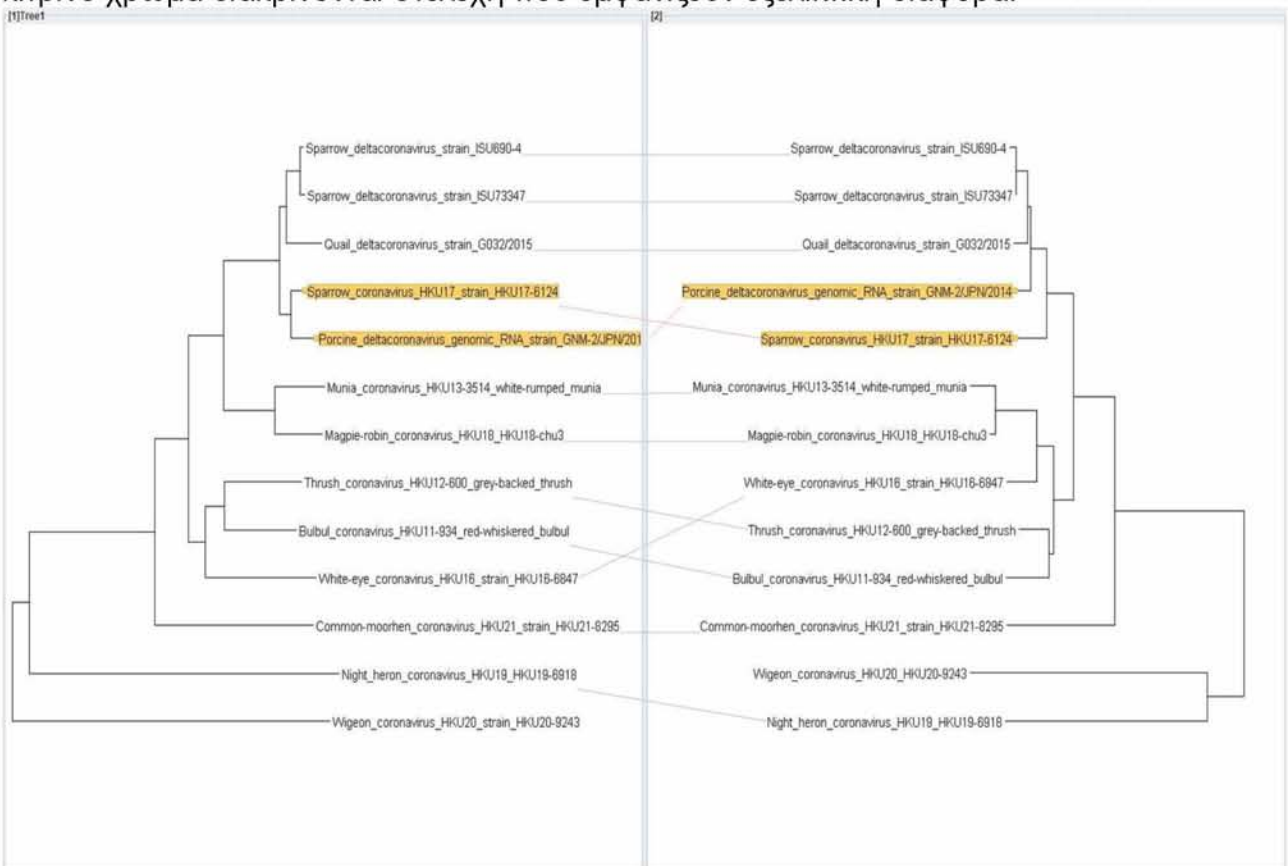


Εικόνα 3.29 Tanglegram των πρωτεϊνών ORF1a-ORF1b των ιών Deltacoronavirus.

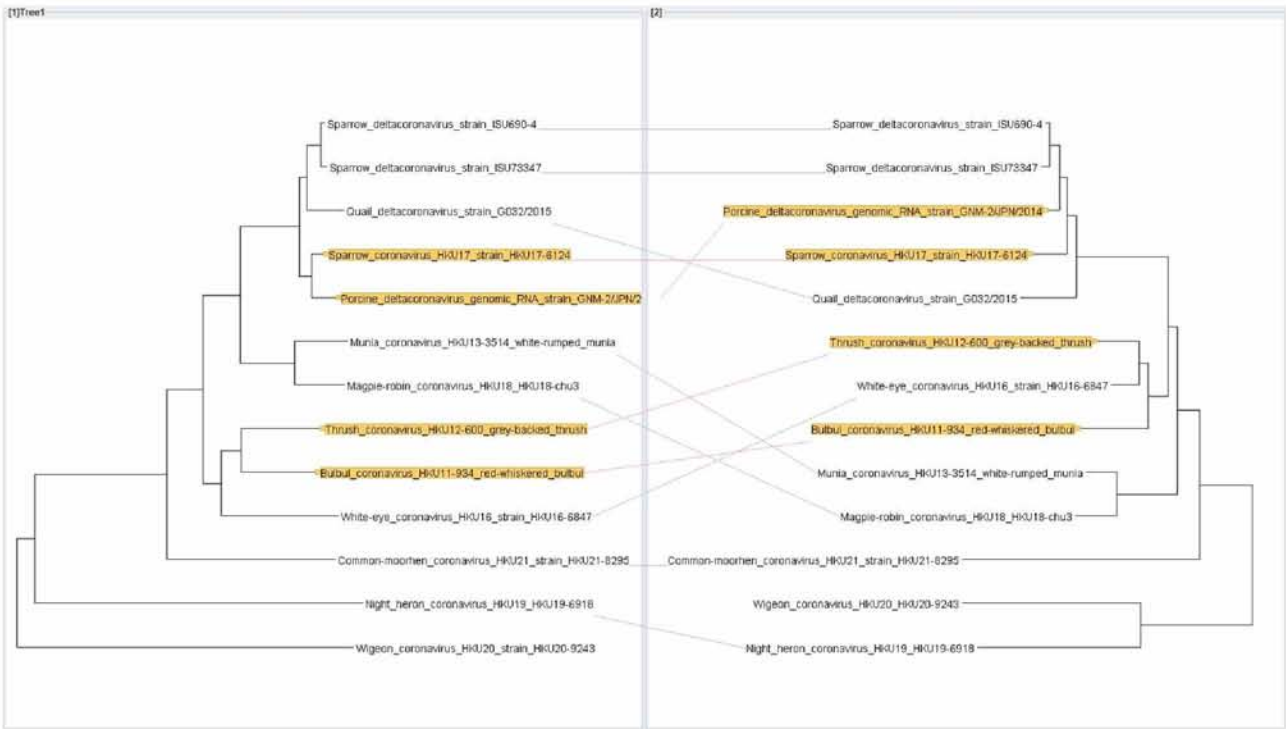




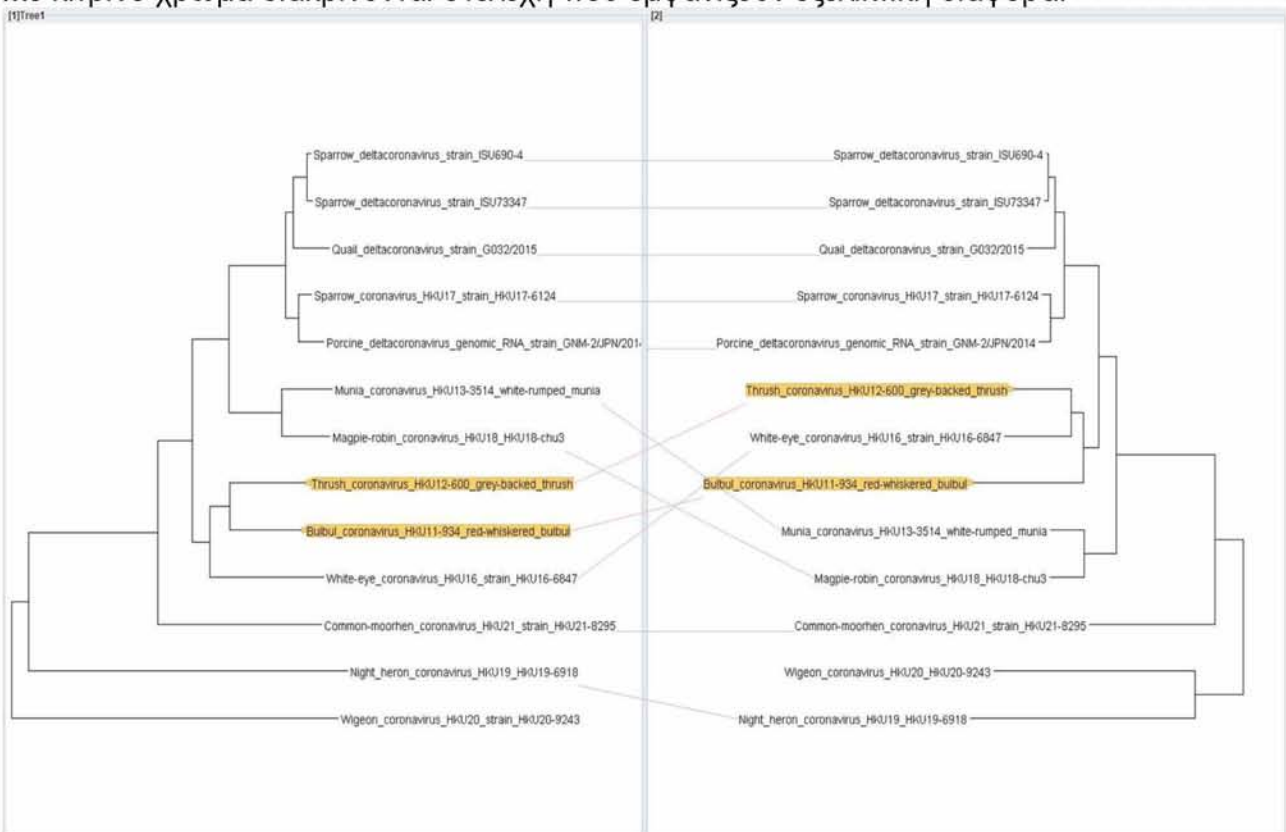
Εικόνα 3.30 Tanglegram των πρωτεϊνών ORF1ab – Spike των ιών Deltacoronavirus. Με κίτρινο χρώμα διακρίνονται στελέχη που εμφανίζουν εξελικτική διαφορά.



Εικόνα 3.31 Tanglegram των πρωτεϊνών ORF1ab – Envelope των ιών Deltacoronavirus. Με κίτρινο χρώμα διακρίνονται στελέχη που εμφανίζουν εξελικτική διαφορά.



Εικόνα 3.32 Tanglegram των πρωτεϊνών ORF1ab – Membrane των ιών Deltacoronavirus. Με κίτρινο χρώμα διακρίνονται στελέχη που εμφανίζουν εξελικτική διαφορά.



Εικόνα 3.33 Tanglegram των πρωτεϊνών ORF1ab – Nucleocapsid των ιών Deltacoronavirus. Με κίτρινο χρώμα διακρίνονται στελέχη που εμφανίζουν εξελικτική διαφορά.

## 4 ΣΥΖΗΤΗΣΗ

Από τις αρχές του 21ου αιώνα έχουμε παρατηρήσει μια ραγδαία αύξηση στην ανακάλυψη νέων κορωνοϊών αλλά και στα γονιδιώματα αυτών που έγιναν διαθέσιμα στους ερευνητές. Σημαντικό ρόλο σε αυτό το γεγονός έπαιξαν οι τρεις επιδημίες (SARS-2002, MERS 2012, SARS-CoV-2 2019), με την τρίτη και πιο πρόσφατη να συνεχίζεται μέχρι και σήμερα αλλά και τα υπόλοιπα 4 στελέχη κορωνοϊού που προσβάλλουν τον άνθρωπο (229-E, OC-43, NL63, HKU1) καθώς οδήγησαν στην περαιτέρω εξερεύνηση των κορωνοϊών ανακαλύπτοντας έτσι την τεράστια ποικιλομορφία τους. Η παρούσα μελέτη εκμεταλλεύτηκε όλα τα διαθέσιμα γονιδιώματα και κωδικές αλληλουχίες καθώς και πληθώρα βιοπληροφορικών εργαλείων έχοντας ως στόχο να αποκτήσουμε μια καλύτερη εικόνα για την εξέλιξη των κορωνοϊών.

Όπως αναφέρθηκε και προηγουμένως, επιλέχθηκε να γίνει το φιλτράρισμα των στελεχών των ιών βάσει του γονιδίου Spike καθώς η αμινοξική ποικιλία του ανάμεσα στα διάφορα είδη είναι η αιτία ύπαρξης πολλών ενδιάμεσων οργανισμών ξενιστών για τους κορωνοϊούς. Το γεγονός της μικρής ομοιότητας τόσο της νουκλεοτιδικής όσο και της αμινοξικής αλληλουχίας του γονιδίου της πρωτεΐνης Spike μεταξύ των διαφόρων στελεχών επιβεβαιώνεται και από τα διαγράμματα (Εικόνα 3.2 – Εικόνα 3.9). Το επόμενο αξιοσημείωτο χαρακτηριστικό που παρατηρούμε στα διαγράμματα είναι το σχετικά μικρό ποσοστό ομοιότητας των πέντε σημαντικών γονιδίων των ιών του γένους Alphacoronavirus (48% έως 61%) και Betacoronavirus (33% έως 54%) διακρίνοντας έτσι μια μεγάλη ετερογένεια των ιών που ανήκουν σε αυτά τα γένη. Αντίθετα, στους ιούς του γένους Gammacoronavirus αλλά και Deltacoronavirus τα επίπεδα ομοιότητας των γονιδίων μεταξύ των διαφόρων ιών που ανήκουν σε αυτά είναι ιδιαίτερα υψηλά από 57% έως 75% για τους Gamma και 54% έως 72% για τους Delta, με εξαίρεση φυσικά το γονίδιο Spike του οποίο το ποσοστό είναι χαμηλότερο σε σχέση με τα γειτονικά του γονίδια. Στο σημείο αυτό θα έπρεπε να μπει μια υποσημείωση για τους ιούς του γένους Gammacoronavirus καθώς το ποσοστό ομοιότητάς τους είναι μεροληπτικά υψηλό διότι από τα 48 στελέχη που παρέμειναν προς ανάλυση μετά το φιλτράρισμα, περισσότερα από 30 ήταν διαφορετικά στελέχη του ιού Avian coronavirus (Infectious Bronchitis).

Προχωρώντας στην εξελικτική σύγκριση των φυλογενετικών δένδρων, και πιο συγκεκριμένα για τους ιούς του γένους Alphacoronavirus δεν παρατηρούμε κάποια σημαντική διαφοροποίηση μεταξύ ORF1a – ORF1b, ούτε και στο πιο απομακρυσμένο γονίδιο, δηλαδή μεταξύ ORF1ab και Nucleocapsid. Αντίθετα στο Spike (Εικόνα 3.12), μεταξύ των στελεχών *NL63-related bat CoV 63-15* και *63-95*, *Canine CoV\_A76* και *AlphaCoV\_2303*, στο γονίδιο Envelope (Εικόνα 3.13) μεταξύ των *Human CoV\_NL63* και *NL63-related\_bat\_CoV*, *Scotophilus\_Kuhli* και *CoV\_BtSK\_63-9a* και *Alphacoronavirus\_sp* παρατηρήθηκε φυλογενετική ασυμφωνία (phylogenetic incongruence) που ενδέχεται να ωφείλεται σε ανασυνδυασμός στο γονιδίωμα των ιών. Επίσης, όσον αφορά το γονίδιο Membrane (Εικόνα 3.14) ασυμφωνία παρατηρήθηκε μεταξύ *Rhinolophus\_HKU32* και *Bat\_CoV\_Ra13591*, *Hipposideros\_pomona* και *Alphacoronavirus sp*.

Στο γένος των ιών Betacoronavirus και ειδικότερα μεταξύ των στελεχών *bat-CoV-HKU9-4* και *HKU9-BF\_0051* παρατηρείται φυλογενετική ασυμφωνία για όλα τα σημαντικά γονίδια εκτός από το ORF1a – ORF1b. Διαφορές στο ORF1a – ORF1b (Εικόνα 3.17) βλέπουμε μεταξύ των ιών *Bat\_Sars\_like\_ZCC21* και *SARS\_CoV\_2019*, *Bat\_CoV\_sp\_strain\_C13* και *C12*, *Bat\_CoV\_HKU5* και *HKU4*. Το τελευταίο ζευγάρι ιών, εμφανίζουν ασυμφωνία και όσον αφορά το γονίδιο της Membrane (Εικόνα 3.20). Για το γονίδιο Spike, εκτός από το HKU9, ασυμφωνία παρατηρήθηκε και για τα στελέχη *BtRS-2013* και *Bat\_CoV-2005* (Εικόνα 3.18). Τέλος, στο γονίδιο του Nucleocapsid, άξια αναφοράς

είναι η διαφοροποίηση μεταξύ των στελεχών του ιού **MERS NL13845 και NL140422** (Εικόνα 3.21).

Σε σχέση με τα υπόλοιπα γένη, στους ιούς Gammacoronavirus παρατηρήθηκαν οι λιγότερες διαφορές αναλογικά, καθώς όπως αναφέρθηκε και προηγουμένως οι περισσότερες αλληλουχίες ήταν διαφορετικά στελέχη του ιού Infectious Bronchitis. Ωστόσο, τόσο στο γονίδιο Spike (Εικόνα 3.24) τα στελέχη **Infectious Bronchitis ahysx-1** και **CK/CHGDQY1C** όσο και στο γονίδιο Membrane (Εικόνα 3.26) τα στελέχη **SNU8067** και **Serotype Gray** ήταν τόσο απομακρυσμένα που δεν ήταν δυνατό να αγνοηθεί αυτή η διαφορά. Επίσης για το Spike, φυλογενετική ασυμφωνία παρατηρήθηκε για τους ιούς **Turkey-CoV** και **Guinea\_found\_CoV** αλλά και στο γονίδιο Envelope (Εικόνα 3.25) για τους ιούς **Duck\_CoV** και **Canada\_goose\_CoV**.

Περνώντας στο πιο πρόσφατα ανακαλυμμένο γένος κορωνοϊών, η μεγαλύτερη φυλογενετική ασυμφωνία παρατηρήθηκε στο ζευγάρι ιών **Sparrow\_CoV\_HKU17** και **Porcine\_deltacoronavirus** στα γονίδια Spike, Envelope και Membrane και στο ζευγάρι ιών **Thrush\_CoV\_HKU12** και **Bulbul\_CoV\_HKU11** στα γονίδια Spike, Membrane και Nucleocapsid (Εικόνα 3.30 - 3.32). Τέλος, στο γονίδιο Spike παρουσιάζουν ασυμφωνία άλλα 2 ζευγάρια ιών, **Sparrow\_delta\_ISU630-4** και **ISU73347**, **Munia\_CoV\_HKU13** και **Magpie\_CoV\_HKU18** (Εικόνα 3.29).

Αναλύοντας τις τιμές Robinson – Foulds και συγκρίνοντας τα διάφορα φυλογενετικά δένδρα μεταξύ τους παρατηρούμε για τους Alphacoronavirus πως οι τιμές μεταξύ ORF1ab και Nucleocapsid είναι συγκριτικά μικρότερες με τα υπόλοιπα γονίδια (RF=27) και άρα υπάρχουν λιγότερες διαφορές στους κλάδους του φυλογενετικού δένδρου (Εικόνα 3.10). Αντίθετα, οι υψηλότερες τιμές παρατηρούνται στις συγκρίσεις του Spike με τα υπόλοιπα γονίδια (RF>45) γεγονός που υποδεικνύει την ενδεχόμενη ύπαρξη ανασυνδυασμών στο ενδιάμεσο τμήμα του γονιδιώματος μεταξύ του ORF1ab και του Nucleocapsid. Ανάλογη ερμηνεία των αποτελεσμάτων με τους Alphacoronavirus έχουμε και για τους Deltacoronavirus, καθώς έχουμε μικρή τιμή Robinson – Foulds μεταξύ ORF1ab και Nucleocapsid (RF=2), και υψηλές τιμές του Spike με τα υπόλοιπα γονίδια (RF>9) (Εικόνα 3.28). Όσον αφορά τους Betacoronavirus, οι τιμές Robinson – Foulds είναι σχετικά σταθερές μεταξύ ORF1ab και Envelope, Membrane, Nucleocapsid (RF=20-23), αλλά και σε αυτήν την περίπτωση οι συγκριτικές τιμές του γονιδίου Spike παρουσιάζονται αυξημένες (RF>25) (Εικόνα 3.16). Σε παρόμοιο μήκος κύματος με τους ιούς Betacoronavirus, βρίσκονται και οι τιμές Robinson – Foulds των ιών Gammacoronavirus καθώς παρατηρείται μια σταθερά τιμών μεταξύ 33-34 για τα ORF1ab με τα υπόλοιπα γονίδια εκτός από το Spike που οι τιμές του είναι μεγαλύτερες από 35 έπειτα από σύγκριση με τα υπόλοιπα γονίδια (Εικόνα 3.22).

Από τα δεδομένα που μας έχουν γίνει διαθέσιμα έως τώρα, γίνεται σαφές ότι το γονιδίωμα των κορωνοϊών εμφανίζει υψηλή παραλλακτικότητα στο γονιδιακό περιεχόμενο τους. Ωστόσο υπάρχουν ερωτήματα που θα πρέπει να διερευνηθούν στο μέλλον, όπως ποια είναι η σημασία των προσαρμοστικών αλλαγών στις non-structural proteins του ORF1ab αλλά και ο ρόλος των βοηθητικών πρωτεϊνών στη μετάδοση των ιών μεταξύ των διαφόρων ειδών. Επίσης, ένας ακόμη τομέας που χρήζει εξερεύνησης είναι η δυνατότητα δημιουργίας εργαλείων, χρησιμοποιώντας τις εξελικτικές πληροφορίες που είναι γνωστές, για την πρόβλεψη ιών που μπορούν να οδηγήσουν σε μια νέα πανδημία.

## BIBΛΙΟΓΡΑΦΙΑ

- Alanagreh, L., Alzoughool, F., Atoum, M., 2020. The Human Coronavirus Disease COVID-19: Its Origin, Characteristics, and Insights into Potential Drugs and Its Mechanisms. *Pathogens* 9, 331. <https://doi.org/10.3390/pathogens9050331>
- Annan, A., Baldwin, H.J., Corman, V.M., Kloze, S.M., Owusu, M., Nkrumah, E.E., Badu, E.K., Anti, P., Agbenyega, O., Meyer, B., Oppong, S., Sarkodie, Y.A., Kalko, E.K.V., Lina, P.H.C., Godlevska, E.V., Reusken, C., Seebens, A., Gloza-Rausch, F., Vallo, P., Tschapka, M., Drosten, C., Drexler, J.F., 2013. Human Betacoronavirus 2c EMC/2012–related Viruses in Bats, Ghana and Europe. *Emerg Infect Dis* 19, 456–459. <https://doi.org/10.3201/eid1903.121503>
- Chen, Y., Liu, Q., Guo, D., 2020. Emerging coronaviruses: Genome structure, replication, and pathogenesis. *J Med Virol* 92, 418–423. <https://doi.org/10.1002/jmv.25681>
- Cui, J., Li, F., Shi, Z.-L., 2019. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol* 17, 181–192. <https://doi.org/10.1038/s41579-018-0118-9>
- DeDiego, M.L., Álvarez, E., Almazán, F., Rejas, M.T., Lamirande, E., Roberts, A., Shieh, W.-J., Zaki, S.R., Subbarao, K., Enjuanes, L., 2007. A Severe Acute Respiratory Syndrome Coronavirus That Lacks the E Gene Is Attenuated In Vitro and In Vivo. *J Virol* 81, 1701–1713. <https://doi.org/10.1128/JVI.01467-06>
- Fehr, A.R., Perlman, S., 2015. Coronaviruses: An Overview of Their Replication and Pathogenesis. *Coronaviruses* 1282, 1–23. [https://doi.org/10.1007/978-1-4939-2438-7\\_1](https://doi.org/10.1007/978-1-4939-2438-7_1)
- Lee, C.Y.-P., Lin, R.T.P., Renia, L., Ng, L.F.P., 2020. Serological Approaches for COVID-19: Epidemiologic Perspective on Surveillance and Control. *Front. Immunol.* 11. <https://doi.org/10.3389/fimmu.2020.00879>
- Li, F., 2016. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol* 3, 237–261. <https://doi.org/10.1146/annurev-virology-110615-042301>
- Luk, H.K.H., Li, X., Fung, J., Lau, S.K.P., Woo, P.C.Y., 2019. Molecular epidemiology, evolution and phylogeny of SARS coronavirus. *Infect. Genet. Evol.* 71, 21–30. <https://doi.org/10.1016/j.meegid.2019.03.001>
- Malik, Y.A., 2020. Properties of Coronavirus and SARS-CoV-2. *Malays J Pathol* 42, 3–11.
- Saberi, A., Gulyaeva, A.A., Brubacher, J.L., Newmark, P.A., Gorbalenya, A.E., 2018. A planarian nidovirus expands the limits of RNA genome size. *PLoS Pathog* 14, e1007314. <https://doi.org/10.1371/journal.ppat.1007314>
- Ye, Z.-W., Yuan, S., Yuen, K.-S., Fung, S.-Y., Chan, C.-P., Jin, D.-Y., 2020. Zoonotic origins of human coronaviruses. *Int. J. Biol. Sci.* 16, 1686–1697. <https://doi.org/10.7150/ijbs.45472>
- Canonical, n.d. The leading operating system for PCs, IoT devices, Servers and the cloud [WWW Document]. Ubuntu. URL [ubuntu.com](http://ubuntu.com) (accessed 9/9/2020)
- About Perl – [www.perl.org](http://www.perl.org) [WWW Document], n.d. URL [www.perl.org/about.html](http://www.perl.org/about.html) (accessed 9/9/2020)
- Gouy, M., Guindon, S. & Gascuel, O. (2010) SeaView version 4 : a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution* 27(2):221-224.
- R.C. Edgar,(2010), Search and clustering orders of magnitude faster than BLAST, *Bioinformatics* 26(19) 2460-2461
- Kumar S, Stecher G, Li M, Knyaz C, and Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution* 35:1547-1549.
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* 17, 540-552.
- T. Goluch, D. Bogdanowicz, K. Giaro, Visual TreeCmp: Comprehensive Comparison of Phylogenetic Trees on the Web, *Methods in Ecology and Evolution*, (2020).



- D.H. Huson and C. ScordDendroscope 3: An interactive tool for rooted phylogenetic trees and networksnavacca, , Systematic Biology (2012).
- Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Sayers, E.W., 2013. GenBank. Nucleic Acids Res 41, D36–D42. <https://doi.org/10.1093/nar/gks1195>
- Zuckerkandl E. and Pauling L. (1965). Evolutionary divergence and convergence in proteins. Edited in *Evolving Genes and Proteins* by V. Bryson and H.J. Vogel, pp. 97-166. Academic Press, New York.
- Robinson, D. F., & Foulds, L. R. (1981). Comparison of phylogenetic trees. *Mathematical Biosciences*, 53(1-2), 131–147. doi:10.1016/0025-5564(81)90043-2
- Böcker, S., Canzar, S., & Klau, G. W. (2013). The generalized Robinson-Foulds metric. In A. Darling & J. Stoye (Eds.), *Algorithms in Bioinformatics. WABI 2013. Lecture Notes in Computer Science, vol 8126* (pp. 156–169). Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-40453-5\_13
- <https://talk.ictvonline.org/taxonomy/>