

ΜΠΣ ΒΙΟΣΤΑΤΙΣΤΙΚΗ

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΙΑΤΡΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΓΚΟΒΑ ΜΑΡΙΑ

Μελέτη διασποράς της επιδημίας HIV-1 στην Ελλάδα: εκτίμηση
σημαντικών τοπικών δικτύων και χαρακτηρισμός τους

ΑΘΗΝΑ, 2018

Η παρούσα διπλωματική εργασία εκπονήθηκε στο πλαίσιο των σπουδών για την απόκτηση του Μεταπτυχιακού Διπλώματος Ειδίκευσης στη

ΒΙΟΣΤΑΤΙΣΤΙΚΗ

που απονέμει η Ιατρική Σχολή και το Τμήμα Μαθηματικών του Εθνικού & Καποδιστριακού Πανεπιστημίου Αθηνών

Εγκρίθηκε την..... από την εξεταστική επιτροπή:

ΟΝΟΜΑΤΕΠΩΝΥΜΟ

ΒΑΘΜΙΔΑ

ΥΠΟΓΡΑΦΗ

Δ. ΠΑΡΑΣΚΕΥΗΣ (Επιβλέπων)

**ΕΠΙΚΟΥΡΟΣ
ΚΑΘΗΓΗΤΗΣ**

.....

Β. ΣΥΨΑ

**ΕΠΙΚΟΥΡΗ
ΚΑΘΗΓΗΤΡΙΑ**

.....

Γ. ΜΑΓΙΟΡΚΙΝΗΣ

ΛΕΚΤΟΡΑΣ

.....

Στην οικογένειά μου

ΕΥΧΑΡΙΣΤΙΕΣ

Ολοκληρώνοντας την διπλωματική μου εργασία και κατά συνέπεια τις μεταπτυχιακές μου σπουδές στην «Βιοστατιστική» θα ήθελα να ευχαριστήσω θερμά όλους όσους με βοήθησαν, ο καθένας με τον δικό του τρόπο στην επίτευξη αυτού του στόχου.

Ευχαριστώ πρώτα απ' όλα τον Επίκουρο Καθηγητή και Επιβλέποντα μου κ. Παρασκευή Δημήτριο για την ανάθεση της διπλωματικής εργασίας, την καθοδήγηση και τις πολύτιμες συμβουλές που μου έδωσε καθώς και για την εμπιστοσύνη και το ενδιαφέρον του καθ' όλη την διάρκεια. Επίσης ευχαριστώ την Επίκουρη Καθηγήτρια κ. Σύψα Βάνα και τον Λέκτορα κ. Μαγιορκίνη Γκίκα, μέλη της εξεταστικής επιτροπής για τα εποικοδομητικά τους σχόλια

Ευχαριστώ ιδιαίτερα την υποψήφια Διδάκτωρ κ. Κωστάκη Ευαγγελία – Γεωργία για την πολύτιμη βοήθεια της, τον χρόνο της, την καλή της διάθεση και τις συμβουλές της όλο αυτό το χρονικό διάστημα.

Ευχαριστώ μέσα από την καρδιά μου τους γονείς μου για όλα όσα μου έχουν προσφέρει κατά την διάρκεια των σπουδών μου, ο καθένας από την δική του πλευρά και για την υποστήριξή τους σε κάθε μου βήμα.

Θέλω να εκφράσω ένα μεγάλο ευχαριστώ στον Πρωτ. Σταύρο Νικ. Γούλα για την ανιδιοτελή και πολύπλευρη βοήθεια και στήριξη του ώστε να ολοκληρώσω τις μεταπτυχιακές μου σπουδές .

Τέλος ευχαριστώ όλους τους φίλους μου και τους ανθρώπους που με αγαπάνε και πίστεψαν σε εμένα ενθαρρύνοντας με κάθε στιγμή.

Πίνακας περιεχομένων

Εισαγωγή – Σκοπός.....	7
Κεφάλαιο 1: Ο ιός HIV	8
1.1 Ιοί - Γενικά	8
1.1.1 Δομή	8
1.1.2 Πολλαπλασιασμός.....	8
1.2 Ιστορικά στοιχεία του HIV	8
1.3 Ταξινόμηση του HIV	10
1.4 Δομή του HIV	12
1.5 Γονιδίωμα του HIV-1	13
1.6 Ο κύκλος ζωής του HIV-1.....	16
1.6.1 Εισαγωγή.....	17
1.6.2 Αντίστροφη μεταγραφή	17
1.6.3 Ενσωμάτωση	17
1.6.4 Μεταγραφή	17
1.6.5 Μετάφραση.....	18
1.6.6 Συνέλευση	18
1.6.7 Ωρίμανση και απελευθέρωση.....	18
1.7 Παγκόσμια επιδημιολογικά δεδομένα	18
1.8 Ελληνικά επιδημιολογικά δεδομένα.....	21
Κεφάλαιο 2: Υλικό και μέθοδοι.....	24
2.1 Μοριακή Επιδημιολογία	24
2.2 Μοριακή Εξέλιξη	24
2.3 Φυλογενετική ανάλυση.....	24
2.4 Βάση Δεδομένων.....	25
2.4.1 Στοιχισή αλληλουχιών	26
2.4.2 Εξελικτικά μοντέλα - Υπολογισμός εξελικτικών αποστάσεων μεταξύ αλληλουχιών	27
2.4.3 Κατασκευή φυλογενετικών δέντρων	33
2.4.4 Αξιολόγηση των φυλογενετικών δέντρων	39
2.5 Μέθοδος εύρεσης δικτύων από τα τελικά δέντρα	40
2.6 Στατιστική ανάλυση.....	43

Κεφάλαιο 3: Αποτελέσματα	44
3.1 Περιγραφικά στοιχεία (A1&B – A1 – B).....	44
MSM: Men having Sex with Men (Άνδρες που κάνουν σεξ με άνδρες)	44
3.2 Υπότυπος A1	47
3.3 Υπότυπος B	51
3.4 Στατιστική Ανάλυση.....	55
Κεφάλαιο 4: Συμπεράσματα	57
Κεφάλαιο 5: Συζήτηση.....	58
Περίληψη.....	60
Abstract	61
Βιβλιογραφία	63

Εισαγωγή – Σκοπός

Ο ιός της ανθρώπινης ανοσοανεπάρκειας (Human immunodeficiency Virus, HIV) είναι ο αιτιολογικός παράγοντας του συνδρόμου επίκτητης ανοσοανεπάρκειας (Acquired Immune Deficiency Syndrome, AIDS). Σύμφωνα με στοιχεία του Παγκόσμιου Οργανισμού Υγείας (ΠΟΥ), εκτιμάτε ότι ο συγκεκριμένος ιός έχει προσβάλει περίπου 36,9 εκατομμύρια ανθρώπους παγκοσμίως μέχρι το τέλος του 2017. Χαρακτηριστικό του ιού είναι η εκτενής γενετική του ετερογένεια που αποτελεί συνέπεια του υψηλού ρυθμού μεταλλαγών, του υψηλού ρυθμού αναπαραγωγής και ανασυνδυασμού του ιού, με αποτέλεσμα να περιπλέκεται η διάγνωση και η θεραπεία της νόσου, ενώ παράλληλα δυσκολεύεται η διαδικασία ανάπτυξης εμβολίου.

Η μελέτη του ιού με μεθόδους μοριακής επιδημιολογίας έχει συμβάλλει σημαντικά στην κατανόηση της επιδημιολογίας του. Συγκεκριμένα, έχουν εκτιμηθεί:

- η χρονολογική προέλευση της παγκόσμιας αλλά και πολλών επιμέρους επιδημιών του
- η γεωγραφική κατανομή των υποτύπων και ανασυνδυασμένων τύπων του ιού
- η προέλευση της επιδημίας σε διαφορετικές γεωγραφικές περιοχές ανά τον κόσμο
- ο τρόπος διασποράς του ιού μεταξύ ή εντός διαφορετικών πληθυσμών

Η μεγαλύτερη γενετική ετερογένεια του ιού παρατηρείτε στην Υποσαχάρια Αφρική, ενώ στην Ευρώπη και στην Βόρειο Αμερική επικρατεί ο υπότυπος Β. Παράλληλα στην Ευρώπη παρατηρούνται και άλλοι υπότυποι (μη-Β υπότυποι).

Στην Ελλάδα σύμφωνα με στοιχεία της επιδημιολογικής επιτήρησης (Κέντρο Ελέγχου και Πρόληψης νοσημάτων, ΚΕΕΛΠΝΟ), ο συνολικός αριθμός ατόμων που έχουν προσβληθεί από τον ιό και έχουν δηλωθεί έως την 31^η Οκτωβρίου 2017 ανέρχεται σε 16.527. Προηγούμενες μελέτες έχουν δείξει ότι οι υπότυποι Α1 και Β είναι οι επικρατέστεροι υπότυποι του HIV-1 στην Ελλάδα και παρουσιάζουν το μεγαλύτερο ποσοστό διασποράς.

Σκοπός της παρούσας μελέτης είναι η διερεύνηση του τρόπου διασποράς των υποτύπων Α1 και Β στην Ελλάδα, η εύρεση όλων των τοπικών δικτύων μετάδοσης τους (επιμέρους επιδημίες) και η εκτίμηση των παραγόντων που σχετίζονται με την τοπική διασπορά του ιού, αναλύοντας με μεθόδους μοριακής επιδημιολογίας όλες τις διαθέσιμες αλληλουχίες των συγκεκριμένων υπότυπων με δειγματοληψία στην Ελλάδα το χρονικό διάστημα μεταξύ 1996 και Ιούνιος του 2015.

Κεφάλαιο 1: Ο ιός HIV

1.1 Ιοί - Γενικά

Οι ιοί είναι αυτόνομα μολυσματικά σωματίδια που διαφέρουν πολύ από τους υπόλοιπους μικροοργανισμούς. (1) Τα χαρακτηριστικά που τους διακρίνουν αφορούν την απλή δομική τους οργάνωση και το μηχανισμό πολλαπλασιασμού τους. Το μέγεθός τους κυμαίνεται από 20 έως 1000 nm σε μήκος. (2)

1.1.1 Δομή

Το ιικό σωματίδιο είναι ένα πλήρες, εντελώς αναπτυγμένο, λοιμογόνο σωματίδιο ιού που αποτελείται από γενετικό υλικό, το οποίο μπορεί να έχει ένα συγκεκριμένο τύπο **νουκλεϊκού οξέος** είτε DNA, είτε RNA. Το νουκλεϊκό οξύ ενός ιού μπορεί να είναι μονόκλωνο ή δίκλωνο, κυκλικό ή γραμμικό ή σε ξεχωριστά τμήματα. Το ποσοστό του νουκλεϊκού οξέος σε σχέση με τις ικές πρωτεΐνες κυμαίνεται από 1% έως 50%. Το νουκλεϊκό οξύ περιβάλλεται από ένα πρωτεϊνικό κάλυμμα, το **καψίδιο**, το οποίο αποτελείται από υπομονάδες, τα καψομερίδια, που μπορεί να είναι ένας ή περισσότεροι τύποι πρωτεϊνών. Ανάλογα με την αρχιτεκτονική του καψιδίου προσδιορίζεται και η γενική μορφολογία του ιού όπως ελικοειδείς ιοί, πολυεδρικοί ή σύνθετοι. Ορισμένες φορές το καψίδιο εσωκλείεται σε ένα λιπιδικό ή πρωτεϊνικό ή υδατανθρακικό **φάκελο** ο οποίος μπορεί να είναι καλυμμένος με ακίδες, δηλαδή σύμπλοκα πρωτεϊνών και υδατανθράκων.(2)

1.1.2 Πολλαπλασιασμός

Οι ιοί είναι υποχρεωτικώς ενδοκυτταρικά παράσιτα, δηλαδή απαιτούν ζωντανά κύτταρα προκειμένου να πολλαπλασιαστούν. Οι ιοί δεν διαθέτουν ένζυμα για την παραγωγή ενέργειας ή για την πρωτεϊνοσύνθεση. Έτσι για να πολλαπλασιαστεί ένας ιός πρέπει να εισβάλει σε ένα κύτταρο ξενιστή και να εκτρέψει τα μεταβολικά εργαλεία του κυττάρου προκειμένου να παράγει ιικά ένζυμα και άλλα ιικά συστατικά.(2)

1.2 Ιστορικά στοιχεία του HIV

Μέχρι το 1980 ο HIV ήταν άγνωστος και η μετάδοση του δεν συνοδευόταν με εμφανή συμπτώματα, με αποτέλεσμα μέχρι εκείνο το έτος να μη γνωρίζουμε πόσοι άνθρωποι είχαν μολυνθεί ή είχαν αναπτύξει AIDS. Μέχρι το 1980 μπορεί να είχαν ήδη μολυνθεί 100.000 - 300.000 άνθρωποι σε Ασία, Βόρεια και Νότια Αμερική, Αφρική, Ευρώπη και Αυστραλία. (3)

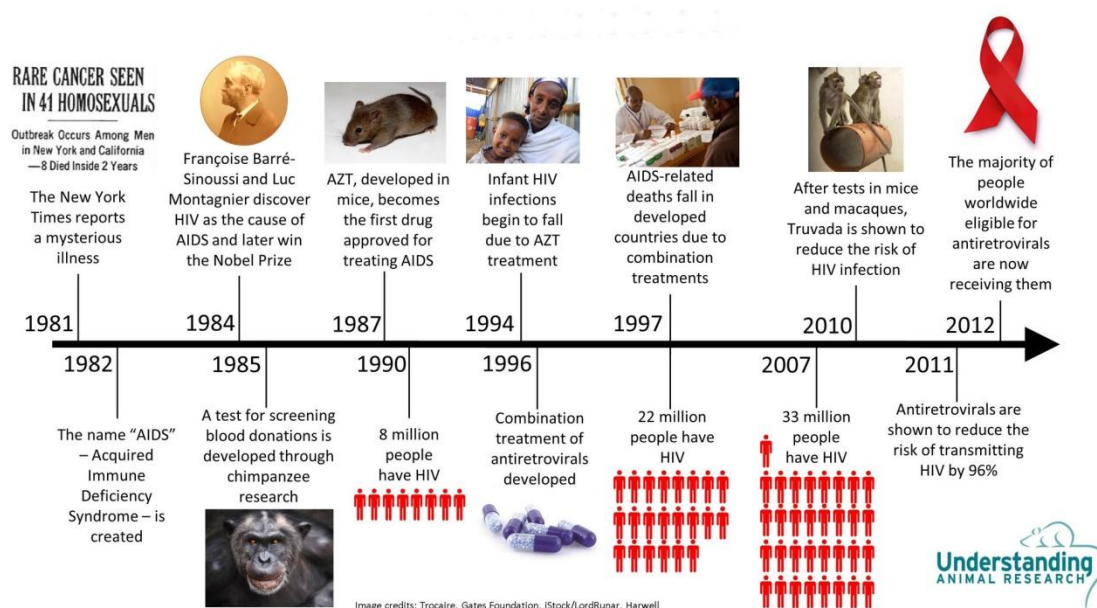
Το 1981 έγινε η πρώτη καταγραφή κλινικού περιστατικού AIDS (Εικόνα 1). (4) Τότε παρουσιάστηκαν σε νοσοκομεία του Λος Άντζελες των ΗΠΑ πέντε περιστατικά νέων ομοφυλόφιλων ανδρών με συμπτώματα της σπάνιας πνευμονικής λοίμωξης «Πνευμονία από Πνευμοκύστη» (Pneumocystis Carinii Pneumonia - PCP). (5) Την

ίδια περίοδο αναφέρθηκαν 73 κρούσματα ομοφυλόφιλων ανδρών στην Νέα Υόρκη και την Καλιφόρνια που είχαν αναπτύξει έναν σπάνιο καρκίνο δέρματος, το σάρκωμα Καπόσι (Sarkoma Karosi, KS). (6) Τον Δεκέμβριο του ίδιου έτους έγινε η πρώτη καταγραφή χρήστη ενδοφλέβιων ναρκωτικών με την PCP λοίμωξη. (7)

Αυτή η έντονη επίπτωση των νοσημάτων PCP και KS οδήγησε το Κέντρο Ελέγχου και Πρόληψης Νοσημάτων της Αμερικής (Communicable Disease Center, CDC) να δημιουργήσει μια ομάδα η οποία θα παρακολουθούσε και θα έθετε υπό έλεγχο τις επιδημίες. Τον Ιούνιο του 1982 υπέθεσαν ότι η αιτία της ανοσολογικής ανεπάρκειας ήταν η σεξουαλική επαφή και έδωσαν στην νόσο την ονομασία GRID (Gay-Related Immune Deficiency). (3) Τελικά όταν διαπιστώθηκε ότι αυτή η νόσος δεν περιοριζόταν μόνο στην κοινότητα των ομοφυλόφιλων εισήχθη ο όρος AIDS όπου το CDC ξεκίνησε να το χρησιμοποιεί τον Σεπτέμβριο του 1982. (8)

Το 1983 μια αμερικανική ομάδα με επικεφαλή τον Ρόμπερτ Γκάλλο (Robert Gallo) ανακάλυψε έναν καινούριο ιό, ο οποίος μπορεί να είχε μολύνει τους ασθενείς του AIDS. Ισχυρίστηκαν ότι ο ιός που απομόνωσαν από έναν ασθενή με AIDS, είχε παρόμοια μορφολογία με ανθρώπινους Τ-λεμφοτροπικούς ιούς (Human T-Lymphotropic Virus, HTLVs). Ο Γκάλλο έδωσε στον απομονωμένο ιό την ονομασία HTLV-III. (9) Ταυτόχρονα μια γαλλική ομάδα με επικεφαλή τον Λουκ Μοντανιέ (Luc Montagnier) απομόνωσε έναν ιό από έναν ασθενή με αυχενική λεμφαδενοπάθεια και σωματική εξασθένιση, κλασικά συμπτώματα του AIDS, δείχνοντας ότι οι πρωτεΐνες του πυρήνα του νέου ιού δεν ήταν ίδιες μορφολογικά με τις πρωτεΐνες p19 και p24 του HTLV-1, το οποίο ερχόταν σε αντίθεση με αυτό που είχε παρατηρήσει η ομάδα του Γκάλλο. Ο Μοντανιέ έδωσε στον ιό την ονομασία LAV (Lymphadenopathy – Associated Virus). (10) Το 1984 ο Τζέι Λεβί (Jay A. Levy) δημιούργησε μια Τ - κυτταρική σειρά από μολυσμένα άτομα στο Σαν Φρανσίσκο. Αυτές οι καλλιέργειες επιβεβαίωσαν τους ισχυρισμούς του Μοντανιέ χωρίς όμως άμεση σύγκριση με τον LAV. Ο ιός πήρε την ονομασία ARV-2 (AIDS –associated retrovirus). (11) Το 1986, έπειτα από πολλές μελέτες, οι απομονώσεις των ιών του Γκάλλο, του Μοντελιέ και του Λεβί αποδείχτηκε ότι οι ιοί ήταν σχεδόν ίδιοι δίνοντας την συλλογική ονομασία Ιός Ανθρώπινης Ανοσοανεπάρκειας (Human Immunodeficiency Virus Infection, HIV). (12)

Έκτοτε, γνωρίζοντας ότι ο αιτιολογικός παράγοντας του AIDS είναι ο HIV, έχουν γίνει σημαντικές ανακαλύψεις και μελέτες που αφορούν την πρόληψη, τη διάγνωση και την θεραπεία αυτής της λοίμωξης, η οποία αποτελεί μια από τις πιο μολυσματικές ασθένειες στον κόσμο, συνεχίζοντας την έρευνα μέχρι σήμερα. Ο Μοντελιέ και η Φρανσουά Μπαρέ – Σινουσί (Françoise Barré-Sinoussi) τιμήθηκαν με το βραβείο Νόμπελ Φυσιολογίας ή Ιατρικής το 2008 για την ανακάλυψη του HIV. (13)



Εικόνα 1: Ιστορική Αναδρομή του HIV. Πηγή: www.dailyrounds.org

1.3 Ταξινόμηση του HIV

Η ταξινόμηση των ιών γίνεται επίσημα από την Διεθνή Επιτροπή για την Ταξινόμηση των Ιών (International Committee for Taxonomy of Viruses, ICTV). Σύμφωνα με αυτήν, για το 2016, ο HIV τύπου 1 και 2 ανήκει στο γένος των Φακοϊών (Lentivirus), της υποοικογένειας Orthoretrovirinae, της οικογένειας των Ρετροϊών (Retroviridae), της τάξης Unassigned (Πίνακας 1). (14)

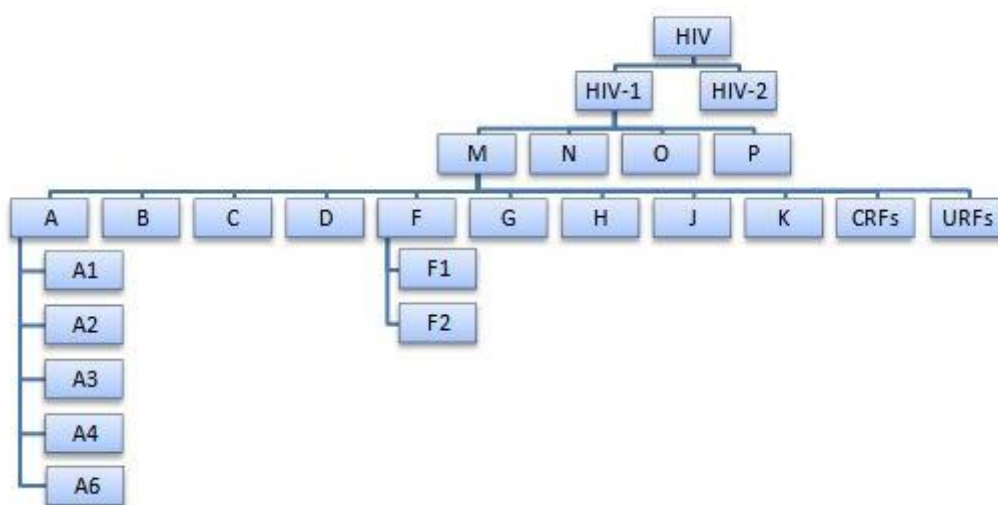
Πίνακας 1: Η ταξινόμηση του HIV για το 2016 σύμφωνα με την International Committee for Taxonomy of Viruses (ICTV)

Τάξη	Unassigned
Οικογένεια	Retroviridae
Υποοικογένεια	Orthoretrovirinae
Γένος	Lentivirus
Είδος	HIV
Τύπος	1 & 2

Ένα από τα χαρακτηριστικά του HIV είναι η εκτενής του ετερογένεια που αποτελεί συνέπεια του υψηλού ρυθμού μεταλλαγών, του υψηλού ρυθμού αναπαραγωγής και ανασυνδυασμού του ιού με αποτέλεσμα να περιπλέκεται η διάγνωση και η θεραπεία της νόσου, ενώ παράλληλα δυσκολεύεται η διαδικασία ανάπτυξης εμβολίου. (15)(16) Όπως είδαμε παραπάνω ο HIV χωρίζεται σε δύο τύπους, τον HIV-1 (HIV τύπου 1) και τον HIV-2 (HIV τύπου 2). Τα στελέχη του HIV-1 έχουν χωριστεί σε

τέσσερις ομάδες: την ομάδα M (Main), την ομάδα O (outlier), την ομάδα N (non-M, non-O) και την ομάδα P.

Η ομάδα M, που είναι υπεύθυνη για την πανδημία του AIDS, κατηγοριοποιείται σε 9 υπότυπους (A-D, F-H, J και K), σε υπο-υπότυπους (A1-A4, A6, F1-F2) καθώς και σε ανασυνδυασμένους τύπους. Όταν ανασυνδυασμένα γονιδιώματα παρουσιάζονται σε έναν μόνο ασθενή ονομάζονται «Unique Recombinant Forms» (URFs), ενώ όταν τα στελέχη του ιού με ταυτόσημο πρότυπο ανασυνδυασμού εντοπιστούν σε τουλάχιστον τρεις ασθενείς, οι οποίοι δε συνδέονται επιδημιολογικά, τακτοποιούνται ως ανασυνδυασμένοι τύποι του HIV, «Circulating Recombinant Forms» (CRFs) οι οποίοι μέχρι σήμερα είναι συνολικά 90. (Εικόνα 2)



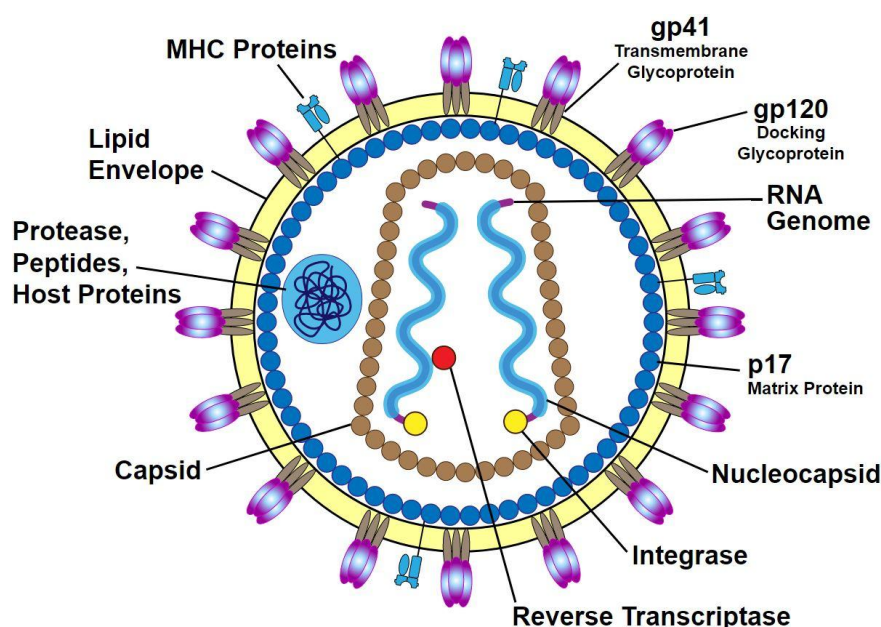
Εικόνα 2: Ταξινόμηση του HIV

Η ομάδα O περιέχει μεγάλη ποικιλία ετερογενών γενετικά στελεχών σε μικρό αριθμό ανθρώπων που ζουν στην Κεντρική Αφρική (Καμερούν, Γκαμπόν, Ισημερινή Γουινέα). Η ομάδα N είναι μια πολύ διακριτή μορφή του ιού που έχει εντοπιστεί σε λίγα άτομα στο Καμερούν. Η ομάδα P εντοπίστηκε πρόσφατα στην Αφρική.

Οι ομάδες M και N έχουν προέλθει από τον αντίστοιχο ιό ανοσοανεπάρκειας του Πιθήκου (Simian Immunodeficiency Virus, SIV) που προσβάλλει τους χιμπατζήδες (Pan troglodytes troglodytes, Ptt) SIVcpzPtt ενώ οι άλλες δύο ομάδες O και P έχουν προέλθει από στελέχη SIV που μολύνουν γορίλλες (Gorilla gorilla) SIVgor. Αντίστοιχα, ο HIV-2 προέρχεται από στελέχη του SIV που μολύνουν τους πιθήκους του είδους Sooty Mangabeys (Cercopithecus atys) δηλαδή τον ιό SIVsm (Εικόνα 3).(15)(17)

κυτταρικές μεμβρανικές πρωτεΐνες προερχόμενες από το κύτταρο ξενιστή όπως ακτίνη και ουβικτίνη, που περιέχουν αντιγόνα μεγάλης ιστοσυμβατότητας.

Στο εσωτερικό του φακέλου βρίσκεται ένα στρώμα (matrix) που αποτελείται από 2000 αντίγραφα της πρωτεΐνης p17. Το στρώμα περικλείει το κωνικό καψίδιο το οποίο συγκροτείται από 2000 αντίγραφα της καψιδιακής πρωτεΐνης p24. Μέσα στο καψίδιο υπάρχει το νουκλεοκαψίδιο, το οποίο περιέχει το γενετικό υλικό, δηλαδή δύο αντίγραφα θετικής μονής έλικας RNA το οποίο σταθεροποιείται ως σύμπλοκο ριβονουκλεοπρωτεΐνης με 2000 αντίγραφα της πρωτεΐνης νουκλεοκαψιδίου p7 και είναι συνδεδεμένα με τα ένζυμα αντίστροφη μεταγραφάση (Reverse Transcriptase, RT), πρωτεάση (Protease, PR) και ιντεγκράση (Integrase, IN).(18)(19)

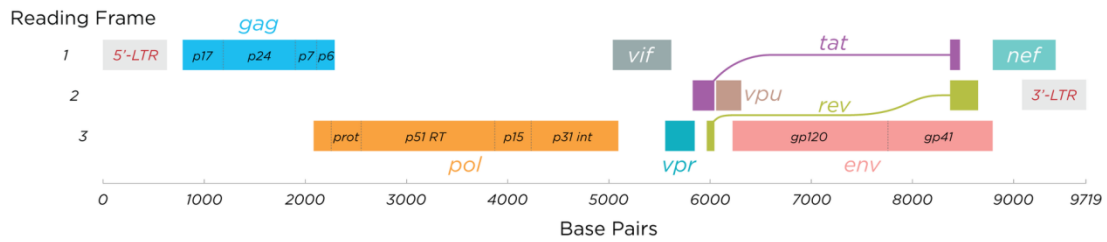


Εικόνα 4: Μορφολογία ιικού σωματίου του HIV.

Πηγή:<https://mappingignorance.org/2013/01/10/towards-a-vaccine-against-hiv/fig1-4/>

1.5 Γονιδίωμα του HIV-1

Το γονιδίωμα του ιού HIV διαφοροποιείται ανάλογα με τον τύπο του ιού. Εδώ θα αναλύσουμε το γονιδίωμα για τον HIV-1. Στην Εικόνα 5 βλέπουμε την μορφή του γονιδιώματος όταν είναι ελεύθερο σωματίδιο ιού. Όταν ο ιός ενσωματωθεί στο DNA γονιδίωμα του ξενιστή ως προϊός, τότε και οι πληροφορίες του κωδικοποιούνται σε μορφή DNA.(20)



Εικόνα 5: Το γονιδίωμα του HIV-1. Πηγή: Thomas Splettstoesser (www.scistyle.com)

Το γονιδίωμα του HIV έχει μήκος περίπου 9,7 Kb (kilo base pair) και αποτελείται από 10 γονίδια, εκ των οποίων 3 είναι κοινά σε όλους τους ρετροϊούς. Τα 3 κοινά και βασικά γονίδια είναι τα gag, env και pol, ενώ τα υπόλοιπα 7 βοηθητικά γονίδια είναι τα tat, rev, nef, vpr, vif, vpu και tev.

Το γονίδιο gag κατευθύνει την σύνθεση και την συναρμολόγηση μιας πρόδρομης πρωτεΐνης (p55), που υποβάλλεται σε επεξεργασία μέσω του γονιδίου πρωτεάσης, για τον σχηματισμό των δομικών πρωτεϊνών του στρώματος (MA, p17), του καψιδίου (CA, p24), του νουκλεοκαψιδίου (NC, p7) και της πρωτεΐνης p6 που υπάρχουν στα ώριμα μολυσματικά ισοσώματα. Το γονίδιο gag συνεργάζεται με τη μεμβράνη του πλάσματος, όπου εκεί γίνεται η συναρμολόγηση του ιού.

Το γονίδιο env κωδικοποιεί μια πρόδρομη γλυκοπρωτεΐνη (gp160), η οποία υποβάλλεται σε επεξεργασία με κυτταρικές πρωτεάσες για να δώσει την γλυκοπρωτεΐνη επιφάνειας gp120 και την διαμεμβρανική γλυκοπρωτεΐνη gp41. Η gp120 καθορίζει τον κυτταρικό τροπισμό με την προσκόλληση στους CD4 υποδοχείς και άλλους συνυποδοχείς και η gp41 εμπλέκεται επίσης στην είσοδο του ιού στα κύτταρα και συμμετέχει στην κυτταρική σύντηξη.

Το γονίδιο pol κωδικοποιεί τα ιικά ένζυμα πρωτεάση (PR, p10), αντίστροφη μεταγραφάση (RT, 51), ιντεγράση (IN, p31) και RNase H (p15). Η αντίστροφη μεταγραφάση αντιγράφει το γονιδίωμα του RNA ιού στο DNA. Αυτή η αντιγραφή είναι το πιο ιδιαίτερο σημείο στον κύκλο ζωής του HIV. Η πρωτεάση επεξεργάζεται τις πρωτεΐνες που παράγονται από το γονιδίωμα του HIV έτσι ώστε να μπορούν να γίνουν μέρος των νέων πλήρως λειτουργούντων σωματιδίων του ιού. Το ένζυμο RNase H διασπά το ρετροϊκό γονιδίωμα μετά από την μόλυνση ενός κυττάρου. Τέλος η ιντεγκράση ενσωματώνει το αντίγραφο DNA του γονιδιώματος HIV στο DNA του ξενιστή.

Τα βοηθητικά γονίδια tat και rev είναι απαραίτητα για την έναρξη της αναπαραγωγής του HIV, ενώ τα γονίδια nef, vif, vpr και vpu έχουν αντίκτυπο σχετικά με την αναπαραγωγή, την εκβλάστηση και παθογένεση του ιού (Πίνακας 2).

Πίνακας 2: Γονίδια (πρωτεΐνες και λειτουργία) του HIV-1

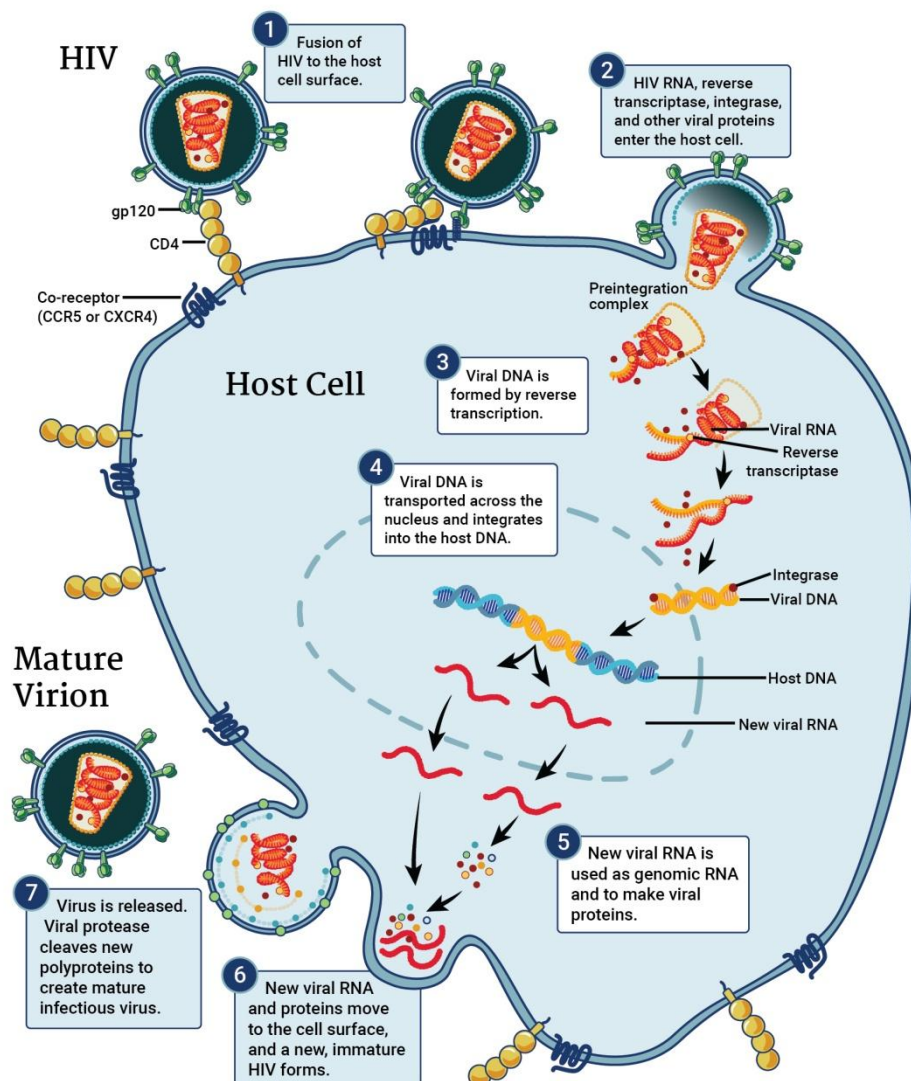
Γονίδιο	Μέγεθος*	Πρωτεΐνη	Λειτουργία
<i>gag</i>		Pr 55Gag	Πρόδρομος των εσωτερικών δομικών πρωτεϊνών
	P24	capsid protein (CA)	Σχηματισμό κωνικού καψιδίου
	P17	matrix protein (MA)	Σχηματισμός του εσωτερικού στρώματος
	P7	nucleoprotein (NC)	Σχηματισμός του σύμπλοκου νουκλεοπρωτεΐνης/RNA
	P6		Εμπλέκεται στην απελευθέρωση του ιικού σωματιδίου
<i>pol</i>		Pr160GagPol	Πρόδρομος των ιικών ενζύμων
	P10	protease (PR)	Πρωτεολυτική διάσπαση της πρόδρομης πρωτεΐνης Gag (p55) και Gag-Pol (Pr160GagPol): απελευθέρωση δομικών πρωτεϊνών και ιικών ενζύμων
	P51	reverse transcriptase (RT)	Μεταγραφή του HIV RNA σε προϊκό DNA
	P15 (66)	RNAse H	Υποβιβασμός του ιικού RNA στο ιικό RNA/DNA σύμπλοκο αντιγραφής
	P31	integrase (IN)	Ενσωμάτωση του προϊκού DNA στο γονίδιο του ξενιστή
<i>env</i>		PrGp160	Πρόδρομος των πρωτεϊνών του φακέλου gp120 και gp41, διάσπαση από την κυτταρική πρωτεάση
	gp120	surface glycoprotein (SU)	Προσκόλληση του ιού στον κυτταρικό στόχο
	gp41	transmembrane protein (TM)	«αγκύρωση» της gp120, σύντηξη ιικής και κυτταρικής μεμβράνης
<i>tat</i>	p14	transactivator protein	Ενεργοποιητής της μεταγραφής ιικών γονιδίων
<i>rev</i>	P19	RNA splicing regulator	Ρυθμίζει την εξαγωγή με συντεταγμένου και μερικής συναρμολογούμενου ιικού mRNA
<i>nef</i>	P27	negative regulating factor	Επίδραση στην αναπαραγωγή του HIV, ενίσχυση της μολυσματικότητας των ιικών σωματιδίων, μείωση της ρύθμισης των CD4 κυττάρων
<i>vif</i>	P23	viral infectivity protein	Κρίσιμη για την παραγωγή μολυσματικών ιών in vivo
<i>vpr</i>	P15	virus protein r	Συστατικό των σωματιδίων του ιού, αλληλοεπιδρά με την p6, διευκολύνει τη μολυσματικότητα του ιού, επιδρά στον κύκλο του κυττάρου
<i>vpu</i>	p16	virus protein uique	Αποτελεσματική απελευθέρωση σωματιδίων του ιού, έλεγχος αποικοδόμησης των CD4, ρυθμίζει την ενδοκυτταρική διακίνηση
<i>tev</i>	p26	tat/rev protein	Tat-Env-Rev συγχωνευτική πρωτεΐνη, ρυθμίζει την δραστηριότητα του tat και του rev στον πυρήνα

*Οι αριθμοί αντιστοιχούν στο μέγεθος της πρωτεΐνης (p) ή της γλυκοπρωτεΐνης (gp) σε 1000 Da

1.6 Ο κύκλος ζωής του HIV-1

Ο κύκλος ζωής του HIV-1 μπορεί να αποτυπωθεί συνοπτικά στα επόμενα βήματα (Εικόνα 6):

1. Εισαγωγή του ιού στο κύτταρο ξενιστή
2. Αντίστροφη μεταγραφή
3. Ενσωμάτωση
4. Μεταγραφή
5. Μετάφραση
6. Συνέλευση
7. Ωρίμανση και Εκβλάστηση (έξοδος)



Εικόνα 6: Ο κύκλος ζωής του HIV-1 Πηγή: www.niaid.nih.gov/diseases-conditions/hiv-replication-cycle

1.6.1 Εισαγωγή

Όπως αναφέραμε παραπάνω όλοι οι ιοί, κατά συνέπεια και ο HIV, δεν μπορεί να πολλαπλασιαστεί από μόνος του. Για να παρασιτήσει χρειάζεται να βρει ένα κύτταρο ξενιστή. Όταν οι γλυκοπρωτεΐνες το φακέλου gp120 και gp41 του ιού ακουμπήσουν πάνω στο μόριο CD4 που έχει στην επιφάνεια του μια ομάδα T-λεμφοκυττάρων (τα CD4 ή βοηθητικά λεμφοκύτταρα) του ξενιστή, τότε συνδέονται σταθερά και σχηματίζουν ένα σύμπλεγμα. Έτσι ο ιός αγκιστρώνεται πάνω στο κύτταρο στόχο, οι μεμβράνες τους συντήκονται και το περιεχόμενο του σωματιδίου του ιού μπορεί να εισέλθει στο κύτταρο.

Εκτός από το μόριο CD4, ο HIV πρέπει να συνδεθεί και με μια άλλη πρωτεΐνη την CCR5 ή την CXCR4 που είναι συν-υποδοχείς (co-receptors). Μερικοί άνθρωποι, λόγω μίας μεταλλάξεως, έχουν λιγότερους συν-υποδοχείς πάνω στα T-λεμφοκύτταρά τους με αποτέλεσμα να έχουν αργότερη εξέλιξη στην λοίμωξη, απ' ότι ο γενικός πληθυσμός.

1.6.2 Αντίστροφη μεταγραφή

Επειδή, όπως έχουμε ήδη αναφέρει, το γενετικό υλικό του HIV αποτελείται από δύο αντίγραφα θετικής μονής έλικας RNA (ριβονουκλεϊκό οξύ) δεν μπορεί να ενσωματωθεί από μόνο του με το γενετικό υλικό του κυττάρου που είναι DNA (δεοξυριβονουκλεϊκό οξύ). Εδώ παρεμβαίνει η ανάστροφη μεταγραφάση, ένα ένζυμο που μεταγράφει στην μονή αλυσίδα του ιικού RNA μία συμπληρωματική αλυσίδα DNA και σχηματίζει μια υβριδική διπλή αλυσίδα. Η ίδια ανάστροφη μεταγραφάση στη συνέχεια χρησιμοποιεί σαν μήτρα την έλικα του DNA και κατασκευάζει την συμπληρωματική της, έχοντας σαν αποτέλεσμα μια διπλή έλικα DNA με τις γενετικές πληροφορίες του ιού. Η διαδικασία της ανάστροφης μεταγραφής είναι ιδιαίτερα ευάλωτη σε λάθη και κατά τη διάρκειά της υπάρχει περίπτωση μετάλλαξης. Αυτές οι παραλλαγμένες μορφές ιού είναι πολλές φορές θνησιγενείς, άλλες εμφανίζουν ανθεκτικότητα στη θεραπεία και κάποιες δεν μπορούν να αναγνωριστούν από τα κύτταρα του ανοσοποιητικού συστήματος.

1.6.3 Ενσωμάτωση

Το διπλής έλικας DNA εισέρχεται στον πυρήνα του κυττάρου ξενιστή και ενσωματώνεται στο κυτταρικό DNA μέσω της δράσης ενός άλλου ενζύμου, της ιντεγκράσης (αλλιώς ενσωματάσης, integrase) και καλείται προϊός. Ορισμένες φορές ο προϊός παραμένει σε λανθάνουσα κατάσταση και πολλαπλασιάζεται μαζί με το DNA του κυττάρου ξενιστή.

1.6.4 Μεταγραφή

Το DNA του προϊού χρησιμεύει ως πρότυπο για τη δημιουργία του νέου ιικού RNA (messenger RNA ή mRNA, επειδή μεταφέρει μηνύματα από τον πυρήνα) μέσω μιας διαδικασίας γνωστή ως μεταγραφή. Το πρόσφατα σχηματισμένο ιικό RNA κινείται έξω από τον πυρήνα του μολυσμένου κυττάρου.

1.6.5 Μετάφραση

Το ιικό RNA φέρει τον κώδικα για την σύνθεση πρωτεϊνών και ενζύμων. Ο κώδικας μεταφράζεται σε αλυσίδες αμινοξέων, γνωστές ως πολυπεπτιδικές αλυσίδες, οι οποίες διπλώνουν για να σχηματίσουν τα συστατικά των πρωτεϊνών και των ενζύμων των νέων σωματιδίων του ιού.

1.6.6 Συνέλευση

Τα συστατικά που απαιτούνται για την κατασκευή των νέων σωματιδίων ιού (ικές πρωτεΐνες, ένζυμα και γενετικό υλικό), μετακινούνται στην εξωτερική μεμβράνη του κυττάρου όπου συσσωρεύονται και συναρμολογούνται με την μορφή μπουμπουκιού.

1.6.7 Ωρίμανση και απελευθέρωση

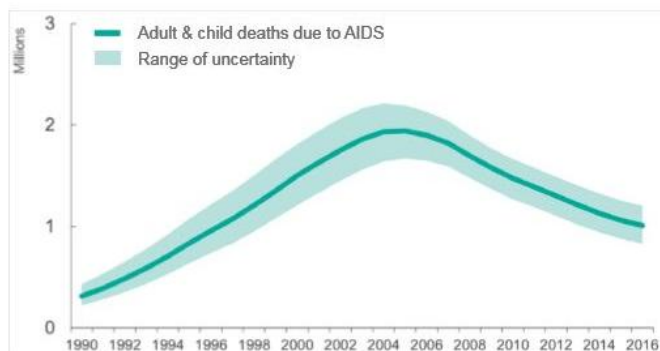
Ένα ένζυμο, η πρωτεάση (protease) τεμαχίζει σε μικρότερα κομμάτια τις μακριές πρωτεϊνικές αλυσίδες που προκύπτουν από τη μεταγραφή του ιικού DNA, και με αυτό τον τρόπο προκύπτουν οι τελικές μορφές πρωτεϊνών που χρειάζονται ώστε να δημιουργηθεί ένας ώριμος μολυσματικός ιός. Σε αυτό το σημείο αρχίζει η συναρμολόγηση των νέων ιών. Οι δομικές πρωτεΐνες gag και pol σχηματίζουν τα περιβλήματα, ενώ οι πρωτεΐνες gp120 και gp41 διατάσσονται στην εξωτερική επιφάνειά τους. Αμινοξέα που περιέχουν ψευδάργυρο και βρίσκονται στην εξωτερική επιφάνεια του ιού, το ιικό RNA και τα ένζυμα αντίστροφη μεταγραφάση, ιντεγκράση και πρωτεάση, συγκεντρώνονται στον πυρήνα και οι ιοί αρχίζουν να εκβλαστάνουν από την κυτταρική μεμβράνη του CD4 κυττάρου. Βγαίνοντας συμπαρασύρουν ένα κομμάτι από την κυτταρική μεμβράνη του ξενιστή, με το οποίο τυλίγονται πλήρως. Ένα μολυσμένο κύτταρο μπορεί να απελευθερώσει πολλά νέα σωματίδια του ιού HIV που μετακινούνται προς αναζήτηση νέων κυττάρων στόχων σε διάφορα μέρη του σώματος, όπου επαναλαμβάνεται ο κύκλος ζωής του ιού. Τα μολυσμένα κύτταρα τελικά καταστρέφονται.

1.7 Παγκόσμια επιδημιολογικά δεδομένα

Ο HIV εξακολουθεί να αποτελεί σημαντικό παγκόσμιο ζήτημα δημόσιας υγείας έχοντας οδηγήσει μέχρι σήμερα περισσότερους από 35 εκατομμύρια ανθρώπους στο θάνατο.(24) Η αντιρετροϊκή θεραπεία, την οποία μέχρι τον Ιούλιο του 2017 έχουν λάβει περίπου 20,9 εκατομμύρια άνθρωποι, έχει συμβάλει στη μείωση κατά 48% των θανάτων από AIDS, καθώς το 2015 υπήρχαν 1,9 εκατομμύρια [1,7 εκ. – 2,2 εκ.] θάνατοι ενώ το 2016 1,0 εκατομμύριο [830.000 - 1,2 εκ.]. Μέχρι τα τέλη του 2016 υπολογίζεται ότι 36,7 εκατομμύρια [30,8 εκ. – 42,9 εκ.] άνθρωποι σε όλο τον κόσμο ζούσαν με τον HIV. Το 2016 υπήρξαν 1,8 εκατομμύρια [1,6 εκ. – 2,1 εκ.] νέα περιστατικά μολύνσεων από τον ιό. (Πίνακας 3, Γραφήματα 1 και 2)

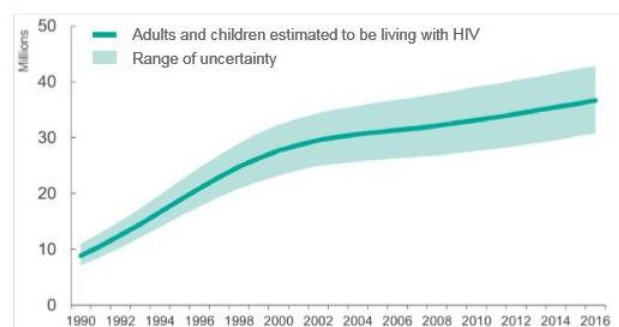
Πίνακας 3: Παγκόσμια στατιστικά για τον HIV και το AIDS για το 2016

Περιοχή	Ενήλικες και παιδιά που ζουν με HIV έως το 2016	Νέες HIV μολύνσεις Ενηλίκων και παιδιών το 2016	Θάνατοι Ενηλίκων και παιδιών λόγω AIDS το 2016
Ανατολική και Νότια Αφρική	19,4 εκ. [17,8 εκ. – 21,1 εκ.]	790.000 [710.000 – 870.000]	420.000 [350.000 – 510.000]
Δυτική και κεντρική Αφρική	6,1 εκ. [4,9 – 7,6 εκ.]	370.000 [270.000 – 490.000]	310.000 [220.000 – 400.000]
Μέση Ανατολή και Βόρεια Αφρική	230.000 [160.00 – 380.000]	18.000 [11.000 – 39.000]	11.000 [7.700 – 19.000]
Ασία και Ειρηνικός	5,1 εκ. [3,9 εκ. – 7,2 εκ.]	270.000 [190.000 – 370.000]	170.000 [130.000 – 220.000]
Λατινική Αμερική	1,8 εκ. [1,4 εκ. – 2,1 εκ.]	97.000 [79.000 – 120.000]	36.000 [28.000 – 45.000]
Καραϊβική	310.000 [280.000 – 150.000]	18.000 [15.000 – 22.000]	9.400 [7.300 – 12.000]
Ανατολική Ευρώπη και κεντρική Ασία	1,6 εκ. [1,4 εκ. – 1,7]	190.000 [160.000 – 220.000]	40.000 [32.000 – 49.000]
Δυτική και κεντρική Ευρώπη και Βόρεια Αμερική	2,1 εκ. [2,0 εκ. – 2,3]	73.000 [68.000 – 78.000]	18.000 [15.000 – 20.000]
Σύνολο	36,7 εκ. [30,8 εκ. - 42,9 εκ.]	1,8 εκ. [1,6 εκ. - 2,1 εκ.]	1.0 εκ. [830.000 – 1,2 εκ.]



Γράφημα 1: Θάνατοι ενηλίκων και παιδιών λόγω του AIDS από το 1990 έως το 2016

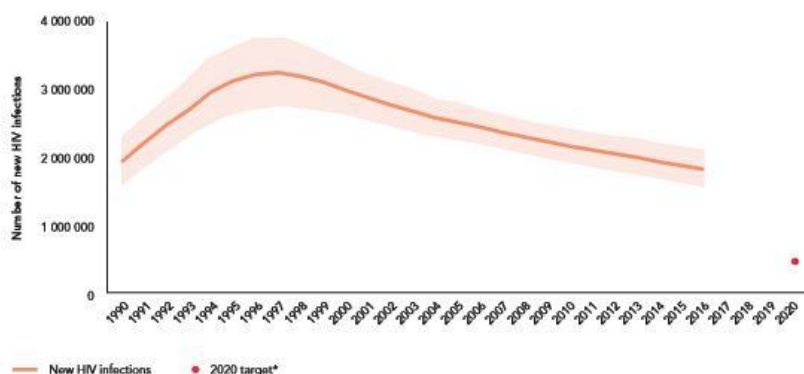
Πηγή: <http://www.unaids.org/>



Γράφημα 2: Ενήλικες και παιδιά που εκτιμάται ότι ζουν με τον HIV από το 1990 έως το 2016

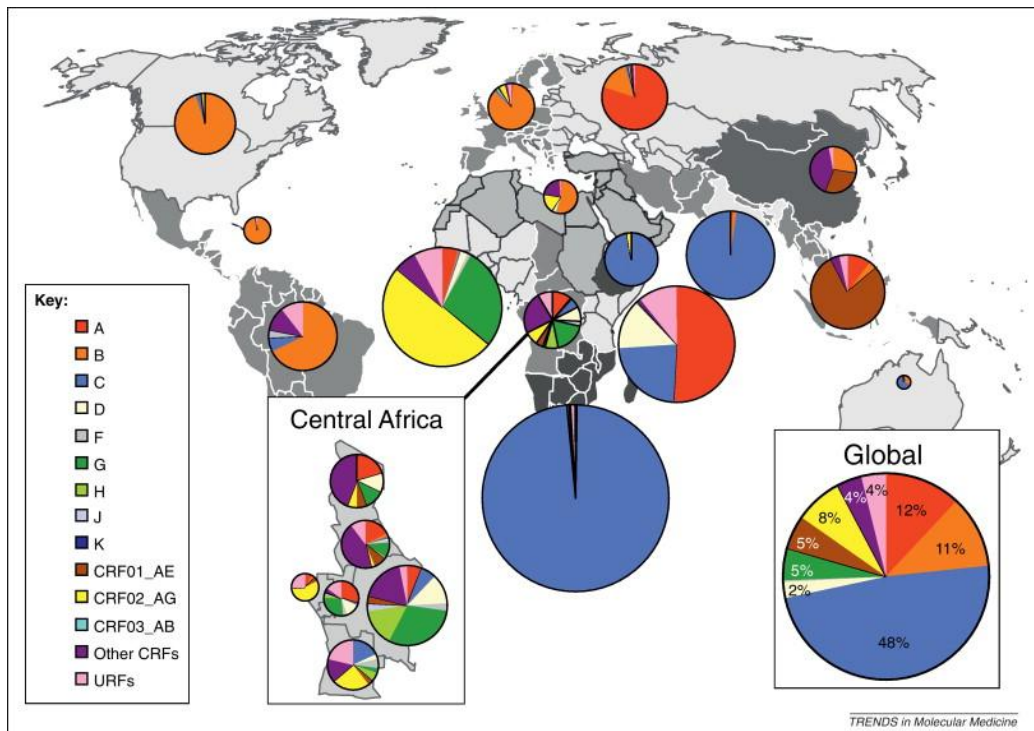
Κάθε μέρα για το 2016 καταγράφονταν περίπου 5.000 νέες μολύνσεις σε όλο τον κόσμο, εκ των οποίων τα 400 ήταν παιδιά. Το υψηλότερο ποσοστό παρατηρήθηκε στην Υποσαχάρια Αφρική και ήταν περίπου 64%.

Παγκόσμιες προσπάθειες πρόληψης και θεραπείας του HIV έχουν μειώσει τα επίπεδα νέων μολύνσεων καθώς από το 2010 ο ετήσιος αριθμός νέων μολύνσεων από τον ιό μειώθηκε κατά 16%. Ωστόσο ο ρυθμός μείωσης των νέων λοιμώξεων είναι πολύ αργός για να επιτευχθεί ο στόχος ταχείας παρακολούθησης που συμφωνήθηκε από την Γενική Συνέλευση των Ηνωμένων Εθνών το 2016, δηλαδή λιγότερες από 500.000 νέες μολύνσεις ετησίως έως το 2020 (Γράφημα 3). (25)



Γράφημα 3: Νέες HIV μολύνσεις, όλων των ηλικιών από το 1990 έως το 2016 και ο στόχος για το 2020 Πηγή: <http://www.unaids.org/>

Η μεγαλύτερη γενετική ετερογένεια του HIV-1 παγκόσμια εντοπίζεται στην κεντρική Υποσαχάρια Αφρική. Εκεί οι υπότυποι A και C είναι οι επικρατέστεροι, όμως έχουν αναφερθεί όλοι οι υπότυποι. Στην Νότια και Ανατολική Αφρική όπως και στην Ινδία κυριαρχεί ο υπότυπος C. Ο υπότυπος B επικρατεί στη Βόρεια και Νότια Αμερική, την Δυτική και Κεντρική Ευρώπη, την Αυστραλία, σε κάποιες χώρες της Ασίας, τη Μέση Ανατολή και τη Βόρεια Αφρική. Στην Δυτική Αφρική κυριαρχεί ο ανασυνδιασμένος τύπος CRF02_AG, ενώ ο CRF01_AE έχει έντονη παρουσία στην Ταϊλάνδη και στις γειτονικές χώρες (Εικόνα 7). (16)(26)(27)

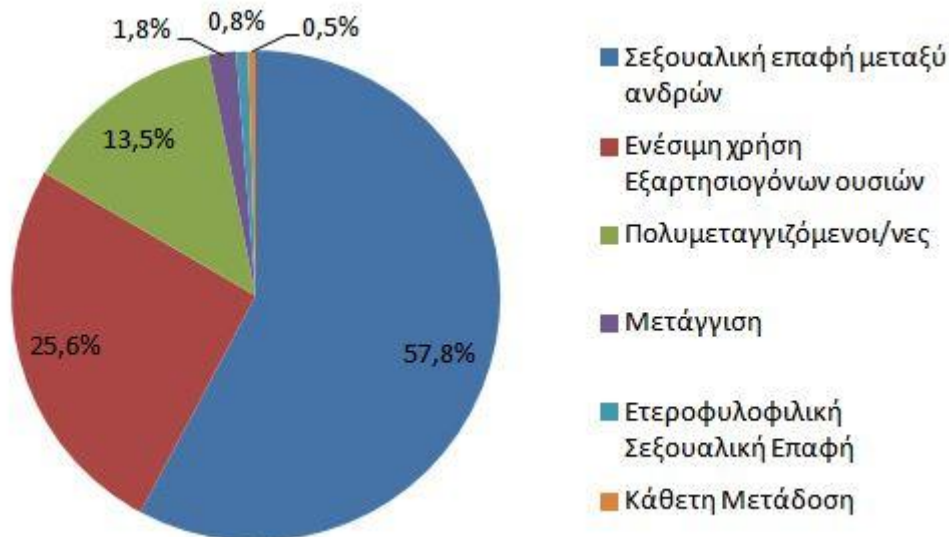


Εικόνα 7: Γεωγραφική διασπορά των υποτύπων και ανασυνδυασμένων τύπων του HIV-1.
Πηγή: Joris Hemelaar 2012

1.8 Ελληνικά επιδημιολογικά δεδομένα

Ο συνολικός αριθμός των HIV οροθετικών ατόμων που έχουν δηλωθεί στην Ελλάδα μέχρι την 31^η Οκτωβρίου 2017 ανέρχεται σε 16.527, εκ των οποίων 82.75% ήταν άνδρες και 17.25% γυναίκες. Από το σύνολο των ατόμων αυτών, 4.083 έχουν εμφανίσει AIDS και περίπου 9.500 λαμβάνουν αντιρετροϊκή θεραπεία. Ο αριθμός των θανάτων ανέρχεται στους 2.746.(28)

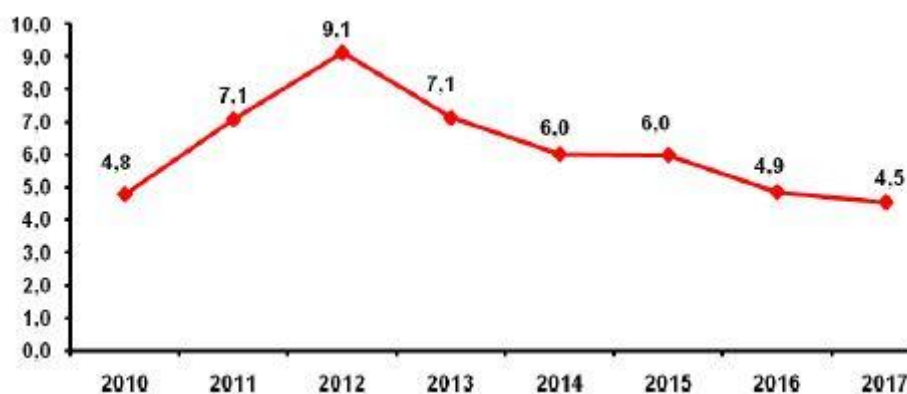
Συνολικά η απροφύλακτη σεξουαλική επαφή μεταξύ ανδρών αποτελεί τον κυριότερο τρόπο μετάδοσης του HIV στην χώρα μας.(28) Έως το τέλος του 2016 το 57.2% των περιστατικών (7.564) που είχαν δηλώσει ότι είναι άνδρες, ανέφεραν ότι μολύνθηκαν μέσω της σεξουαλικής επαφής με άλλους άνδρες. Το συνολικό ποσοστό των δηλωθέντων οροθετικών που μολύνθηκαν μέσω ετεροφυλοφιλικής επαφής ήταν 21% (3.355 περιστατικά) ενώ το ποσοστό λόγω ενέσιμης χρήσης εξαρτησιογόνων ουσιών ανέρχεται στο 11.1% (1.774 περιστατικά). Ακαθόριστος παραμένει ο τρόπος μετάδοσης του ιού για το 18% (2.866 περιστατικών) των οροθετικών ατόμων. Στο Γράφημα 4 παρουσιάζονται τα ποσοστά των HIV διαγνώσεων ανά κατηγορία μετάδοσης για το σύνολο των ατόμων με γνωστή κατηγορία μετάδοσης στην Ελλάδα μέχρι τις 31/12/2016. (29)



Γράφημα 4: Συνολικά ποσοστά δηλωθέντων HIV οροθετικών ατόμων ανά κατηγορία μετάδοσης στην Ελλάδα έως το τέλος του 2016 (δεν περιλαμβάνονται άτομα στα οποία δεν έχει προσδιοριστεί ο τρόπος μετάδοσης). Πηγή: <http://www.keelpno.gr/>

Τα τελευταία 8 έτη η ηλικιακή ομάδα με το υψηλότερο ποσοστό κατά τη διάγνωση είναι αυτή των 30 – 39 ετών. Αύξηση παρατηρείται στις ηλικιακές ομάδες 40 – 49 και 50 – 64 ετών, ενώ υπάρχει πτωτική τάση στην ηλικιακή ομάδα των 19 – 29 ετών.

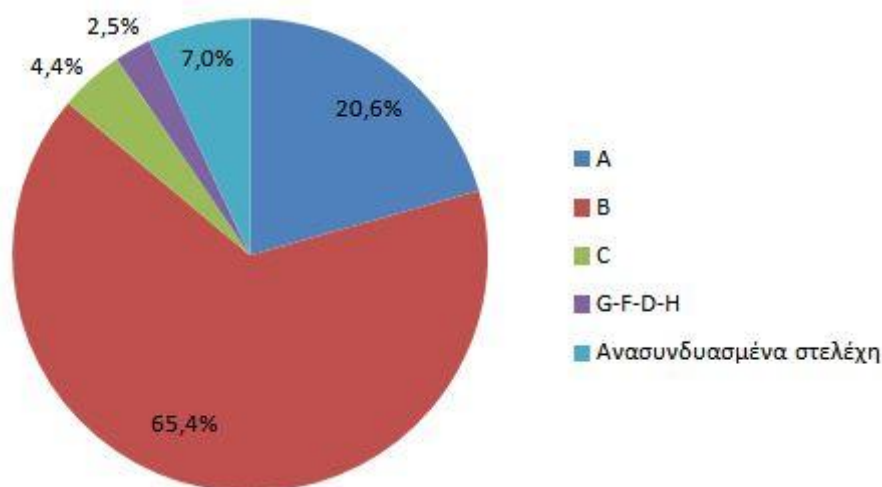
Σημαντική αύξηση του HIV σημειώθηκε την περίοδο 2011 – 2013 στους χρήστες ενδοφλέβιων εξαρτησιογόνων ουσιών, όπως παρουσιάζεται στο Γράφημα 5. Στη συνέχεια παρατηρήθηκε πτωτική τάση του ποσοστού των νέων περιστατικών σε αυτή την ομάδα.



Γράφημα 5: Νέες διαγνώσεις HIV λοιμώξεις ανά 100.000 πληθυσμού (περίοδος Ιανουαρίου-Οκτωβρίου 2010-2017) Πηγή: <http://www.keelpno.gr/>

Σύμφωνα με μελέτη σε 1.158 οροθετικά άτομα στην Ελλάδα από το 1998 έως το 2005, βρέθηκε ότι ο επικρατέστερος υπότυπος με επιπολασμό 65,4% είναι ο Β. Ο αμέσως επόμενος είναι ο υπότυπος Α με ποσοστό 20,6%, ενώ ακολουθούν οι

υπότυποι C (4.4%), G (0.9%), F (0.8%), D (0.7%), H (0.1%) καθώς και ανασυνδυασμένοι τύποι περίπου στο 7% (Γράφημα 6). Στον υπότυπο A παρατηρήθηκε σημαντική αύξηση κατά την διάρκεια του χρόνου καθώς το 1984 ήταν στο 4% ενώ το 2004 στο 42%. Χαρακτηριστικό του υπότυπου A είναι ότι υπερισχύει στον τοπικό πληθυσμό (στην φυλογενετική ανάλυση δημιουργείται ένα μεγάλο μονοφυλετικό δίκτυο), γεγονός που δείχνει ότι δεν σχετίζεται με την μετανάστευση.(30)(31) Αξιοσημείωτο είναι ότι στην Ελλάδα, όπως επίσης στην Κύπρο και τη Βουλγαρία, έχει βρεθεί το ανασυνδυασμένο στέλεχος CRF04_crx σε ποσοστό 2-10%. (32)(33)(34)



Γράφημα 6: Διασπορά των υποτύπων του HIV-1 στην Ελλάδα

Κεφάλαιο 2: Υλικό και μέθοδοι

2.1 Μοριακή Επιδημιολογία

Η μοριακή επιδημιολογία είναι κλάδος της επιδημιολογίας που μελετά τη συμβολή γενετικών και περιβαλλοντικών παραγόντων, που ανιχνεύονται σε μοριακό επίπεδο, στην αιτιολογία, την κατανομή της συχνότητας και την πρόληψη των νοσημάτων σε ανθρώπινους πληθυσμούς.(35) Μέθοδοι της μοριακής επιδημιολογίας είναι η φυλογενετική ανάλυση, η φυλοδυναμική ανάλυση και η φυλογεωγραφία.

2.2 Μοριακή Εξέλιξη

Σκοπός της εξελικτικής βιολογίας είναι η μελέτη της ιστορίας των ζωντανών οργανισμών και η διαλεύκανση των μηχανισμών που ευθύνονται για αυτό, βασιζόμενοι σε απομεινάρια του παρελθόντος και σε πληροφορίες του παρόντος που βρίσκονται στο γενετικό υλικό του κάθε οργανισμού. Συγκεκριμένα, η μοριακή εξέλιξη έχει δύο βασικά πεδία, τη μελέτη των αλλαγών που συμβαίνουν στο γενετικό υλικό σε συνδυασμό με τη μελέτη του ρυθμού και του μοντέλου που ακολουθούν αυτές οι αλλαγές στην πορεία του χρόνου, και της μοριακής φυλογένειας, με αντικείμενο τη μελέτη της εξελικτικής ιστορίας των οργανισμών βασισμένη σε μοριακά δεδομένα, χρησιμοποιώντας αναπαράσταση με φυλογενετικά δέντρα. Επομένως, το αντικείμενο της μοριακής εξέλιξης είναι η μελέτη της εξελικτικής πληροφορίας που περιέχεται στο γενετικό υλικό των οργανισμών και πώς αυτή η πληροφορία μπορεί να επεξεργαστεί. Χρησιμοποιώντας μαθηματικά μοντέλα μοριακής εξέλιξης μπορούμε να εξάγουμε συμπεράσματα για την εξελικτική ιστορία παθογόνων, όπως ιών, και στην περίπτωση μας για τον HIV.

2.3 Φυλογενετική ανάλυση

Η φυλογενετική ανάλυση έχει σκοπό την εκτίμηση και την αναπαράσταση των εξελικτικών σχέσεων διαφορετικών οργανισμών ή διαφορετικών ατόμων του ίδιου οργανισμού, χρησιμοποιώντας ως δεδομένα το γενετικό τους υλικό ή τις αντίστοιχες πρωτεϊνικές αλληλουχίες του γονιδιώματος, υπολογίζοντας το βαθμό ετερογένειας των ομόλογων περιοχών τους.

Η φυλογενετική ανάλυση από πλευράς διαδικασίας μπορεί να αναλυθεί σε τέσσερα διακριτά βήματα τα οποία αναλύονται στη συνέχεια:

1. Στοίχιση αλληλουχιών βιολογικών μορίων και υπολογισμός βαθμών ανομοιότητας
2. Εξελικτικά μοντέλα - Υπολογισμός εξελικτικών αποστάσεων μεταξύ των αλληλουχιών
3. Κατασκευή φυλογενετικών δέντρων
4. Αξιολόγηση των φυλογενετικών δέντρων

2.4 Βάση Δεδομένων

Όσον αφορά τις αλληλουχίες με δειγματοληψία στην Ελλάδα, η βάση δεδομένων αποτελούνταν από 1.230 αλληλουχίες υπότυπου A1 και 2.156 υπότυπου B. Χρησιμοποιήθηκαν όλες οι διαθέσιμες αλληλουχίες υπότυπου A1 και B, οι οποίες συλλέχτηκαν μεταξύ 1996 και Ιουνίου του 2015 στη Νότιο Ελλάδα. Οι αλληλουχίες είχαν συλλεγεί στα πλαίσια του ελέγχου ανοχής και η έγκριση για την ανάλυσή τους δόθηκε από την επιτροπή βιοηθικής του Εθνικού και Καποδιστριακού Πανεπιστημίου Αθηνών.

Για την εύρεση των επιμέρους επιδημιών χρειάστηκαν αλληλουχίες αναφοράς για τους δύο υπό μελέτη υπότυπους, δηλαδή αλληλουχίες με παγκόσμια δειγματοληψία, με γνωστό και επιβεβαιωμένο υπότυπο. Η αναζήτησή τους έγινε στην παγκόσμια βάση δεδομένων HIV Sequence Database (<https://www.hiv.lanl.gov/components/sequence/HIV/search/search.html>) όπου εντοπίστηκαν όλες οι διαθέσιμες αλληλουχίες μέχρι τον Αύγουστο του 2016 για τους υπότυπους A1 και B. Αρχικά βρέθηκαν 6.418 αλληλουχίες αναφοράς για τον A1 και 58.993 για τον B αντίστοιχα. Από αυτές διαγράφηκαν όσες δεν είχαν πληροφορία για την χώρα προέλευσης καταλήγοντας σε 6.373 αλληλουχίες αναφοράς για τον A1 και 49.112 για τον B.

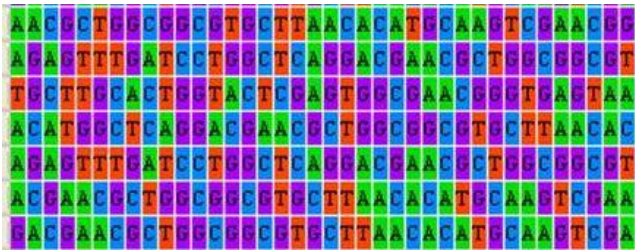
Στη συνέχεια για να αρχίσουν τα βήματα της φυλογενετικής ανάλυσης έπρεπε να ενωθούν οι αλληλουχίες αναφοράς με τις αλληλουχίες από Ελλάδα για κάθε υπότυπο ξεχωριστά, έχοντας ένα σύνολο δεδομένων για τον A1 ($6.373 + 1.230 = 7.603$ αλληλουχίες) και ένα σύνολο δεδομένων για τον B ($49.112 + 2.156 = 51.268$ αλληλουχίες). Λόγω του ότι τα διαθέσιμα προγράμματα αδυνατούν να αναλύσουν τόσο μεγάλο όγκο δεδομένων, ο αριθμός των αλληλουχιών αναφοράς έπρεπε να περιοριστεί. Έτσι, επιλέχτηκε με τυχαίο τρόπο, ένα υποσύνολο του συνόλου των διαθέσιμων αλληλουχιών αναφοράς για τον κάθε υπότυπο. Η δειγματοληψία έγινε στην ιστοσελίδα www.random.org με γεννήτρια τυχαίων αριθμών χωρίς επανάθεση και ακολουθούσε ομοιόμορφη κατανομή. Για να διασφαλιστεί ότι δεν επιλέχθηκαν τυχαία πολλές αλληλουχίες από την ίδια περιοχή, η δειγματοληψία έγινε 5 φορές για τον κάθε υπότυπο. Κατά συνέπεια δημιουργήθηκαν 5 sets με 1500 αλληλουχίες αναφοράς το κάθε ένα για τον υπότυπο A1 και 5 sets με 2000 αλληλουχίες αναφοράς το κάθε ένα για τον υπότυπο B. Το κάθε set αλληλουχιών αναφοράς ενώθηκε με όλες τις αλληλουχίες από Ελλάδα για τον κάθε υπότυπο έχοντας τελικά:

- 5 sets για τον υπότυπο A1 με 2.730 αλληλουχίες το κάθε ένα (1.500 αναφοράς και 1.230 από Ελλάδα)
- 5 sets για τον υπότυπο B με 4.156 αλληλουχίες το κάθε ένα (2000 αναφοράς και 2.156 από Ελλάδα)

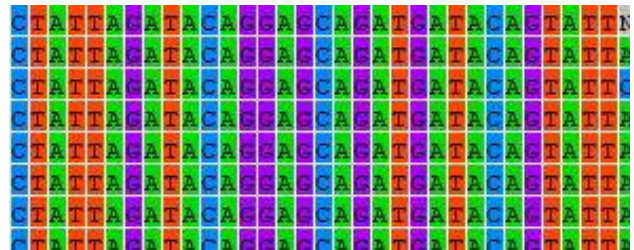
Πραγματοποιήθηκαν 5 ξεχωριστές φυλογενετικές αναλύσεις για τον κάθε υπότυπο. Τα βήματα της φυλογενετικής ανάλυσης τα είδαμε συνοπτικά (στοίχιση αλληλουχιών, επιλογή εξελικτικού μοντέλου, κατασκευή φυλογενετικών δέντρων και αξιολόγηση δέντρων) και τώρα θα τα δούμε αναλυτικά.

2.4.1 Στοιχίση αλληλουχιών

Οι αλληλουχίες DNA ή πρωτεϊνών αποτελούνται από σειρές χαρακτήρων. Μια αλληλουχία DNA περιέχει 4 διαφορετικούς χαρακτήρες και πιο συγκεκριμένα τις βάσεις A, C, T και G (π.χ. TACTCAATCCGTATATAAA), ενώ μια αλληλουχία πρωτεϊνών περιέχει 20 διαφορετικούς χαρακτήρες, τα αμινοξέα (π.χ. PWLRNEFKYFQMRTLVGSI). Προκειμένου να συγκρίνουμε τις αλληλουχίες θα πρέπει να βρούμε τις ομόλογες περιοχές και να τις στοιχίσουμε. Με τον όρο στοιχίση εννοούμε την τοποθέτηση των διαφορετικών αλληλουχιών σε γραμμές, κάθετα η μία από την άλλη έτσι ώστε όμοιοι χαρακτήρες να τοποθετούνται ο ένας κάτω από τον άλλο στην ίδια κατακόρυφη θέση (Εικόνες 8(α) και 8(β)). Όπου οι χαρακτήρες σε διαφορετικές αλληλουχίες διαφέρουν, είναι συνήθως επειδή μία αλληλουχία έχει διαφορετική βάση στην ίδια θέση σε σύγκριση με μία άλλη αλληλουχία είτε λόγω μετάλλαξης, είτε εισαγωγής, είτε διαγραφής. Σε αυτές τις περιπτώσεις μπορεί να εισαχθούν κενά σε ορισμένες από τις αλληλουχίες προκειμένου να στοιχηθούν.(36)



Εικόνα 8 (α): Άποψη μη στοιχισμένων αλληλουχιών



Εικόνα 8 (β): Άποψη στοιχισμένων ακολουθιών

Για τη στοιχίση χρησιμοποιούνται αλγόριθμοι που έχουν ως βάση κυρίως:

- Δυναμικό προγραμματισμό (dynamic programming)
- Γενετικούς αλγορίθμους (genetic algorithms)
- Κρυμμένα Μαρκοβιανά μοντέλα (Hidden Markov Models, HMMs)
- Προοδευτικούς αλγορίθμους (progressive algorithms)

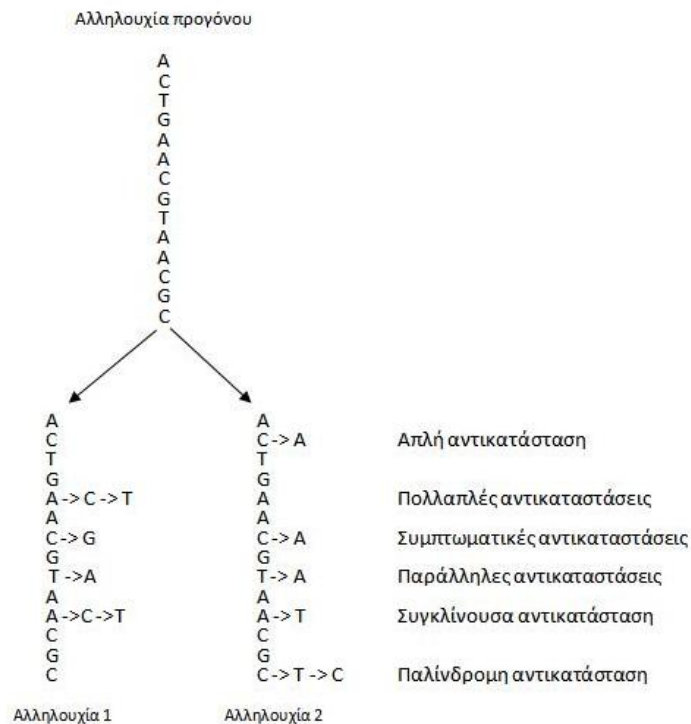
Στην παρούσα εργασία το κάθε set αλληλουχιών στοιχήθηκε *de novo*, χωρίς δηλαδή κάποιο πρότυπο αλληλουχιών, στο πρόγραμμα MEGA 5.2 χρησιμοποιώντας το MUSCLE. Το MUSCLE είναι ένα πρόγραμμα που χρησιμοποιεί προοδευτικούς αλγορίθμους για να δημιουργήσει πολλαπλές στοιχήσεις αλληλουχιών αμινοξέων ή νουκλεοτιδίων. Συγκρινόμενο με άλλα προγράμματα όπως το T-Coffee, το MAFFT και το ClustalW πέτυχε την υψηλότερη κατάταξη στην ακρίβεια σε όλες τις δοκιμές. (37)(38)

Από το γονιδίωμα του ιού το κομμάτι που μελετήθηκε ήταν το γονίδιο *pol*, καθώς σε αυτό το τμήμα του γονιδιώματος γίνεται ο έλεγχος αντοχής του ιού, και συγκεκριμένα η περιοχή της πρωτεάσης και της αντίστροφης μεταγραφάσης. Έπειτα

από τη στοίχιση DNA και τις κατάλληλες διορθώσεις στη στοίχιση (απαλοιφή των κενών θέσεων –ελλείψεων) με μη αυτοματοποιημένη διαδικασία σύμφωνα με το κωδικοποιημένο πλαίσιο ανάγνωσης, οριστικοποιήθηκαν 5 στοιχισμένα sets για τον υπότυπο A1 με 897 νουκλεοτίδια το κάθε ένα και 5 στοιχισμένα sets για τον υπότυπο B με 891 νουκλεοτίδια αντίστοιχα.

2.4.2 Εξελικτικά μοντέλα - Υπολογισμός εξελικτικών αποστάσεων μεταξύ αλληλουχιών

Στη φυλογενετική ανάλυση αφού προηγηθεί η στοίχιση αλληλουχιών DNA ή πρωτεϊνών, επόμενο βασικό βήμα είναι η εκτίμηση των γενετικών αποστάσεων (genetic distances) των υπό μελέτη αλληλουχιών. Ο πιο απλός τρόπος να εκτιμήσουμε την γενετική απόσταση είναι να βρούμε το ποσοστό ανομοιότητας, δηλαδή τον αριθμό των θέσεων (sites) που δύο αλληλουχίες διαφέρουν μεταξύ τους. Αυτή η μέθοδος όμως υποεκτιμά τον πραγματικό αριθμό των αλλαγών (αντικαταστάσεων, μεταλλαγών) που έχουν συμβεί με την πάροδο του χρόνου στην πορεία εξέλιξης των αλληλουχιών από τον κοινό τους πρόγονο, καθώς σε κάποιες θέσεις μπορεί να έχουν συμβεί περισσότερες από μία αντικαταστάσεις (πολλαπλές αντικαταστάσεις) ή να έχουν γίνει αντικαταστάσεις οι οποίες επαναφέρουν ένα νουκλεοτίδιο στην αρχική του κατάσταση (παλίνδρομες αντικαταστάσεις) (Εικόνα 9). Επίσης και οι συγκλίνουσες αντικαταστάσεις οδηγούν σε αύξηση της ανομοιότητας.



Εικόνα 9: Διαφορετικά είδη νουκλεοτιδικών αντικαταστάσεων μεταξύ δύο αλληλουχιών DNA

Επομένως η παρατηρούμενη ετερογένεια είναι μικρότερη από πραγματική. Προκειμένου να μπορεί να γίνει διόρθωση στην εκτίμηση των γενετικών αποστάσεων αναπτύχθηκαν διάφορα μοντέλα νουκλεοτιδικής αντικατάστασης, τα λεγόμενα εξελικτικά μοντέλα (evolutionary models) τα οποία ανήκουν στη γενική κατηγορία των: ομοιογενή στο χρόνο (time-homogeneous) και διαρκώς στάσιμα στο χρόνο μοντέλα (time – continuous stationary Markov models) και χρησιμοποιούν τις ακόλουθες παραδοχές:

1. Σε οποιαδήποτε χρονική στιγμή, ο ρυθμός αντικατάστασης από ένα νουκλεοτίδιο i σε j είναι ανεξάρτητος από το νουκλεοτίδιο που υπήρχε σε αυτή τη θέση πριν το i (ιδιότητα Markov).
2. Οι ρυθμοί αντικατάστασης δεν μεταβάλλονται με το χρόνο (ομοιογένεια)
3. Οι σχετικές συχνότητες των νουκλεοτιδίων A, C, G και T είναι σε ισορροπία (στασιμότητα – stationarity).

Οι παραπάνω παραδοχές δεν ανταποκρίνονται πάντα στην πραγματικότητα, αλλά χρησιμοποιούνται στα μοντέλα θεωρώντας ότι η διαδικασία είναι τυχαία.

Η διαδικασία νουκλεοτιδικής αντικατάστασης των αλληλουχιών DNA μπορεί να γενικευτεί σε μια Μαρκοβιανή διαδικασία και μπορεί να περιγραφεί από έναν πίνακα Q, ο οποίος καθορίζει όλες τις παραμέτρους του εξελικτικού μοντέλου. Παρακάτω φαίνεται η πιο γενική μορφή του πίνακα Q.

$$Q = \begin{pmatrix} \text{A} & \text{C} & \text{G} & \text{T} \\ \begin{matrix} -(a\pi_C + b\pi_G + c\pi_T) \\ g\pi_A \\ h\pi_A \\ j\pi_A \end{matrix} & \begin{matrix} a\pi_C \\ -(g\pi_A + d\pi_G + e\pi_T) \\ i\pi_C \\ k\pi_C \end{matrix} & \begin{matrix} b\pi_G \\ d\pi_G \\ -(h\pi_A + i\pi_C + f\pi_T) \\ l\pi_G \end{matrix} & \begin{matrix} c\pi_T \\ e\pi_T \\ f\pi_T \\ -(j\pi_A + k\pi_C + l\pi_G) \end{matrix} \end{pmatrix}$$

Οι γραμμές ακολουθούν την σειρά A, C, G και T έτσι ώστε για παράδειγμα ο δεύτερος όρος στην πρώτη γραμμή « $a\pi_C$ » είναι ο στιγμιαίος ρυθμός αλλαγής από A σε C. Το « π_C » συμβολίζει την συχνότητα εμφάνισης της βάσης C και το « a » τον ρυθμό αντικατάστασης από A σε C. Επομένως κάθε μη διαγώνια καταχώρηση στον πίνακα αναπαριστά τη ροή του νουκλεοτιδίου από το i στο j , ενώ τα διαγώνια στοιχεία επιλέγονται ώστε το άθροισμα της σειράς να είναι ίσο με το μηδέν, καθώς δείχνουν την συνολική ροή που εγκαταλείπει το νουκλεοτίδιο i .

Όταν στον πίνακα Q οι παράμετροι του εξελικτικού μοντέλου καθοριστούν τότε είναι δυνατός ο υπολογισμός της πιθανότητας να παρατηρηθεί νουκλεοτιδική αντικατάσταση σε κάθε θέση σε χρόνο t (θεωρώντας ότι η πιθανότητα αυτή προκύπτει από μια κατανομή Poisson) με τον ακόλουθο τύπο:

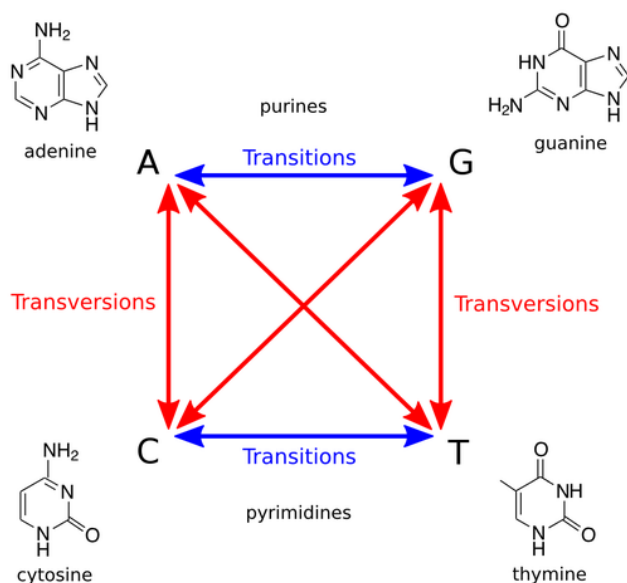
$$P(t) = \exp(Qt)$$

Όταν οι πιθανότητες $P(t)$ είναι γνωστές, η παραπάνω εξίσωση μπορεί να χρησιμοποιηθεί για να υπολογιστούν οι αναμενόμενες αποστάσεις (expected distances) σύμφωνα με το εξελικτικό μοντέλο στον πίνακα Q . Η εκτίμηση των αποστάσεων προκύπτει από την λύση της ακόλουθης εξίσωσης:

$$\frac{d}{dt}P(t) = P(t)Q = QP(t)$$

Μέσα σε αυτό το γενικό πλαίσιο μπορούν να αναπτυχθούν διάφορα εξελικτικά μοντέλα. Αυτά που χρησιμοποιούνται συνήθως στην πράξη είναι αναστρέψιμα στο χρόνο (time reversible), δηλαδή ο ρυθμός αντικατάστασης προς ένα νουκλεοτίδιο ισούται με τον ρυθμό αντικατάστασης του νουκλεοτιδίου αυτού σε κάποιο άλλο (π.χ. ο ρυθμός αντικατάστασης από το i στο j είναι ίδιος με τον αντίστοιχο από το j στο i).

Να σημειωθεί πως οι νουκλεοτιδικές αλλαγές κατατάσσονται σε δύο ομάδες. Οι αντικαταστάσεις στις οποίες μια πουρίνη αλλάζει σε μια πυριμιδίνη ή το αντίστροφο ονομάζονται μεταστροφές (transversions, Tv), ενώ όλες οι υπόλοιπες ονομάζονται μεταβάσεις (transitions, Ts) (Εικόνα 10).



Εικόνα 10: Μεταστροφές και μεταβάσεις στα μοντέλα νουκλεοτιδικής αντικατάστασης

Μερικά από τα μοντέλα αυτά είναι τα ακόλουθα:

- **Jukes and Cantor (JC69):** Το μοντέλο JC69 είναι το πιο απλό εξελικτικό μοντέλο έχοντας μόνο μία παράμετρο που χαρακτηρίζεται ως ρυθμός αντικατάστασης και ισούται με τον ρυθμό όλων των πιθανών αντικαταστάσεων που συμβαίνουν στη μονάδα του χρόνου. Ο ρυθμός αυτός

είναι κοινός σε οποιαδήποτε αντικατάσταση και γι' αυτό το μοντέλο χαρακτηρίζεται ως μονοπαραμετρικό. (39)

$$Q = \begin{pmatrix} * & \frac{\mu}{4} & \frac{\mu}{4} & \frac{\mu}{4} \\ \frac{\mu}{4} & * & \frac{\mu}{4} & \frac{\mu}{4} \\ \frac{\mu}{4} & \frac{\mu}{4} & * & \kappa \\ \frac{\mu}{4} & \frac{\mu}{4} & \frac{\mu}{4} & * \end{pmatrix}$$

- **Felsenstein 1981 (F81):** Το μοντέλο F81 υποθέτει ότι οι συχνότητες των αζωτούχων βάσεων διαφέρουν, ενώ οι αντικαταστάσεις κάθε τύπου θεωρούνται ισοπίθανες. (40)

$$Q = \begin{pmatrix} * & \pi_C & \pi_A & \pi_G \\ \pi_T & * & \pi_A & \pi_G \\ \pi_T & \pi_C & * & \pi_G \\ \pi_T & \pi_C & \pi_A & * \end{pmatrix}$$

- **Kimura (K80):** το μοντέλο K80 θεωρεί ίσες συχνότητες εμφάνισης των αζωτούχων βάσεων, ενώ δίδεται ένας ρυθμός μεταπτώσεων και ένας ρυθμός μεταστροφών για τις αντικαταστάσεις. (41)

$$Q = \begin{pmatrix} * & \kappa & 1 & 1 \\ \kappa & * & 1 & 1 \\ 1 & 1 & * & \kappa \\ 1 & 1 & \kappa & * \end{pmatrix}$$

- **Hasegawa–Kishino–Yano (HKY85):** στο μοντέλο HKY85 οι συχνότητες αζωτούχων βάσεων διαφέρουν, ενώ δίδεται ένας ρυθμός μεταπτώσεων και ένας ρυθμός μεταστροφών για τις αντικαταστάσεις. (42)

$$Q = \begin{pmatrix} * & \kappa\pi_C & \pi_A & \pi_G \\ \kappa\pi_T & * & \pi_A & \pi_G \\ \pi_T & \pi_C & * & \kappa\pi_G \\ \pi_T & \pi_C & \kappa\pi_A & * \end{pmatrix}$$

- **Tamura-Nei (TN93):** στο μοντέλο NT οι συχνότητες των αζωτούχων βάσεων διαφέρουν, ενώ δίδεται ένας ρυθμός μεταστροφών αλλά διαφορετικοί ρυθμοί μεταπτώσεων για τις αντικαταστάσεις. (43)

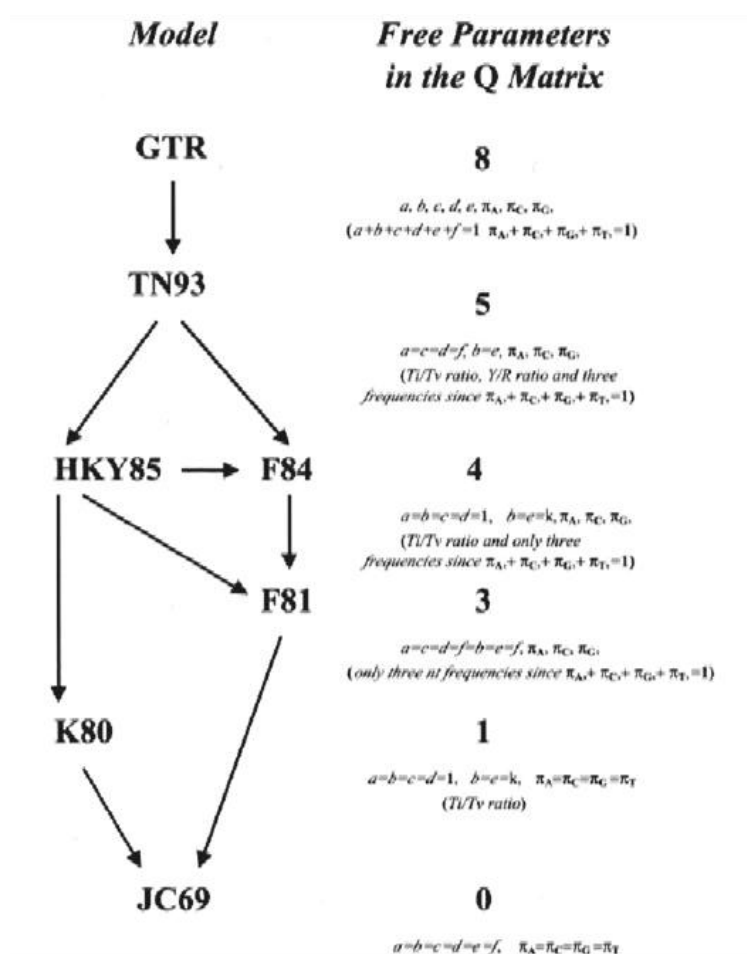
$$Q = \begin{pmatrix} * & \kappa_1\pi_C & \pi_A & \pi_G \\ \kappa_1\pi_T & * & \pi_A & \pi_G \\ \pi_T & \pi_C & * & \kappa_2\pi_G \\ \pi_T & \pi_C & \kappa_2\pi_A & * \end{pmatrix}$$

- **General Time Reversible (GTR):** στο μοντέλο GTR (γενικευμένο μοντέλο αναστροφής του χρόνου), το οποίο είναι το πιο σύνθετο. Οι συχνότητες των

αζωτούχων βάσεων διαφέρουν, ενώ υπάρχει διαφορετικός ρυθμός μεταστροφών και μεταπτώσεων για κάθε μία από τις τέσσερις πιθανές περιπτώσεις. (44)

$$Q = \begin{pmatrix} -(\alpha\pi_C + \beta\pi_A + \gamma\pi_G) & \alpha\pi_C & \beta\pi_A & \gamma\pi_G \\ \alpha\pi_T & -(\alpha\pi_T + \delta\pi_A + \epsilon\pi_G) & \delta\pi_A & \epsilon\pi_G \\ \beta\pi_T & \delta\pi_C & -(\beta\pi_T + \delta\pi_C + \eta\pi_G) & \eta\pi_G \\ \gamma\pi_T & \epsilon\pi_C & \eta\pi_A & -(\gamma\pi_T + \epsilon\pi_C + \eta\pi_A) \end{pmatrix}$$

Στο παρακάτω σχήμα (Εικόνα 11) παρουσιάζονται τα εξελικτικά μοντέλα με ιεραρχία πολυπλοκότητας.

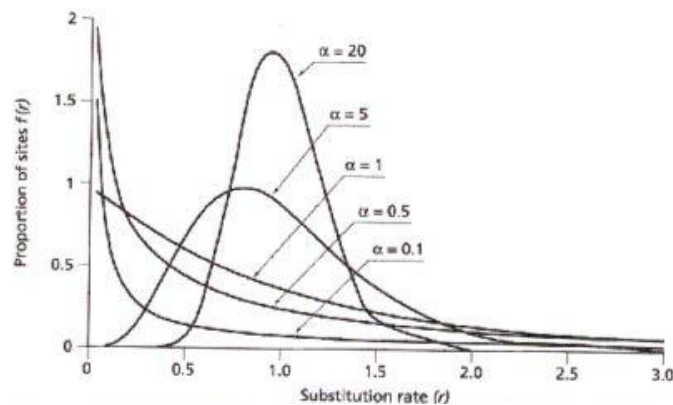


Εικόνα 11: Μοντέλα νουκλεοτιδικής αντικατάστασης με ιεραρχία πολυπλοκότητας (στην παρένθεση φαίνονται οι παράμετροι που εκτιμώνται κάθε φορά. Με Ti/Tv συμβολίζεται ο λόγος των μεταβάσεων (transitions) προς των μεταπτώσεων (transversions))

Στα μοντέλα που είδαμε παραπάνω ο ρυθμός αντικατάστασης κάθε νουκλεοτιδίου είναι σταθερός σε όλες τις θέσεις της αλληλουχίας του DNA. Είναι όμως γνωστό ότι τις περισσότερες φορές αυτό δεν ισχύει, δηλαδή ο ρυθμός αντικατάστασης ποικίλει σημαντικά σε διαφορετικές θέσεις της νουκλεοτιδικής αλληλουχίας. Αυτό σημαίνει

ότι οι μεταλλαγές δεν παρατηρούνται με την ίδια συχνότητα σε όλο το μήκος της αλληλουχίας καθώς υπάρχουν διαφορετικές εξελικτικές πιέσεις σε διαφορετικές περιοχές του γονιδιώματος, όπως επίσης σχετίζεται και με την λειτουργικότητα της κάθε περιοχής. Έτσι για παράδειγμα σε περιοχές του γενετικού υλικού που κωδικοποιούν για πρωτεΐνες, μεταλλαγές στην 1^η και 2^η θέση κάθε κωδικονίου, που οδηγούν σε αλλαγές αμινοξέως, εξελίσσονται με πιο αργούς ρυθμούς σε σχέση με την 3^η θέση που κατά κανόνα δεν οδηγεί σε αντικατάσταση αμινοξέως. Στην φυλογενετική ανάλυση είναι πολύ σημαντικό να γνωρίζουμε αυτή την πληροφορία, καθώς θα έχουμε πιο αξιόπιστα αποτελέσματα από την στιγμή που θα μπορούμε να επιλέξουμε για ανάλυση το σωστό τμήμα του γενετικού υλικού.

Ο διαφορετικός ρυθμός νουκλεοτιδικής αντικατάστασης στις αλληλουχίες DNA μπορεί να προσεγγιστεί με μια συνάρτηση Γ-κατανομής, αντιπροσωπεύοντας το ποσοστό των θέσεων που έχουν κοινό ρυθμό νουκλεοτιδικής αντικατάστασης με παράμετρο α . Όταν η παράμετρος $\alpha < 1$ τότε ο ρυθμός αντικατάστασης διαφέρει σημαντικά μεταξύ των διαφορετικών θέσεων και η συνάρτηση παίρνει σχήμα μορφής L. Όσο μεγαλώνει η τιμή της παραμέτρου α ($\alpha > 1$) τόσο μικρότερη είναι η ετερογένεια, δηλαδή οι περισσότερες θέσεις παρουσιάζουν παρόμοιους ρυθμούς αντικατάστασης με μικρές αποκλίσεις και η συνάρτηση έχει κωδωνοειδή μορφή (Γράφημα 7). Σε τέτοιες περιπτώσεις η διόρθωση στη μοντελοποίηση μπορεί να προσεγγίσει καλύτερα την πραγματική εξελικτική διαδικασία βελτιώνοντας και τις εκτιμήσεις της γενετικής απόστασης. (45)



Γράφημα 7: Συνάρτηση γάμα κατανομής για διαφορετικές τιμές της παραμέτρου α
Πηγή: <http://www.ccg.unam.mx>

Στις 10 φυλογενετικές αναλύσεις αυτής της μελέτης χρησιμοποιήθηκε το εξελικτικό μοντέλο GTR γιατί υπερισχύει έναντι των υπολοίπων για τους ακόλουθους σημαντικούς λόγους (46):

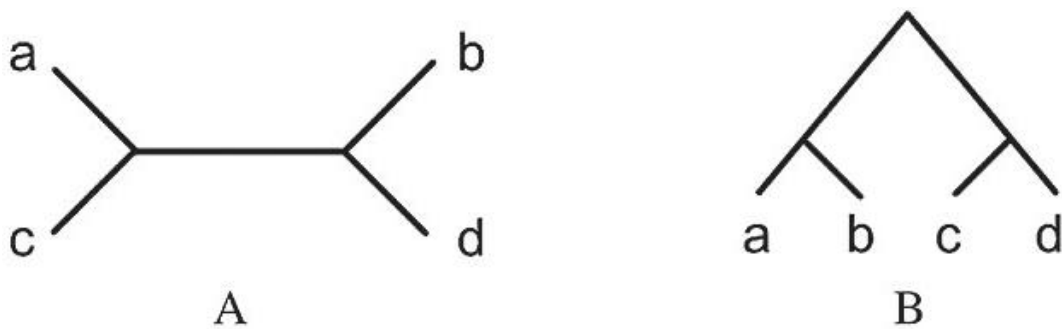
- Υποθέτει ότι μια βάση μπορεί να αλλάξει και μετά να αλλάξει στην αρχική ($A \rightarrow T$ και μετά $T \rightarrow A$)
- Υποθέτει ξεχωριστές παραμέτρους ρυθμού νουκλεοτιδικής αντικατάστασης
- Υποθέτει διαφορετική συχνότητα παραμέτρων για κάθε νουκλεοτιδικό τύπο
- Είναι το πιο ανεξάρτητο μοντέλο αντικαταστάσεων
- Τα υπόλοιπα μοντέλα είναι ειδικές περιπτώσεις του GTR μοντέλου.

2.4.3 Κατασκευή φυλογενετικών δέντρων

Η ανασύσταση των εξελικτικών σχέσεων των αλληλουχιών αναπαρίσταται μέσω ειδικών διαγραμμάτων που ονομάζονται φυλογενετικά δέντρα (phylogenetic trees), τα οποία αποτελούν το βασικό τρόπο αναπαράστασης της εξελικτικής ιστορίας όπως προκύπτει από το εξελικτικό μοντέλο που χρησιμοποιήθηκε. Τα φυλογενετικά δέντρα απεικονίζουν τις σχέσεις γειννίας (ομοιότητας) μεταξύ βιολογικών αντικειμένων. Οι σχέσεις ομοιότητας αντανακλούν την εξελικτική συγγένεια μεταξύ των αντικειμένων αυτών και με αυτόν τον τρόπο τα δέντρα αντιστοιχούν στην γενεαλογία των αλληλουχιών – οργανισμών.

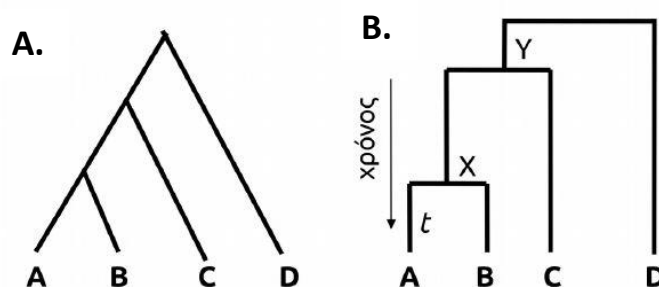
Ένα δέντρο αποτελείται από κόμβους (nodes) και κλάδους (branches). Οι εξωτερικοί κόμβοι (terminal nodes), δηλαδή τα «φύλλα» του δέντρου, αντιστοιχούν σε αλληλουχίες και χαρακτηρίζονται ως λειτουργικές ταξινομικές μονάδες (operational taxonomic units, OTU). Ο αριθμός των OTU αντιστοιχεί στο μέγεθος του δέντρου και στην τελική του μορφή το δέντρο θα έχει τόσες απολήξεις – φύλλα, όσα και τα OTU. Οι εσωτερικοί κόμβοι (internal nodes) αναπαριστούν υποθετικούς προγόνους των υπό μελέτη αλληλουχιών. Το μήκος των κλάδων (branch lengths) αναλογεί στον αριθμό των εξελικτικών αλλαγών που έχουν συμβεί στο χρονικό διάστημα που έχουν εξελιχθεί οι υπό μελέτη αλληλουχίες. Ο υπολογισμός των εξελικτικών αλλαγών μπορεί να γίνει με τις περισσότερες φυλογενετικές μεθόδους. Επειδή ο αριθμός των εξελικτικών αλλαγών είναι ανάλογος του χρόνου, η διεύθυνση που απεικονίζει το μήκος των κλάδων, απεικονίζει επίσης χρόνο.

Ένα φυλογενετικό δέντρο μπορεί να έχει ρίζα (rooted) (Εικόνα 12A), η οποία συμβολίζει τον κοινό πρόγονο όλων των υπόλοιπων οργανισμών, ή όχι (unrooted) (Εικόνα 12B). Ο κοινός πρόγονος πρέπει να είναι γνωστός εκ των προτέρων προκειμένου να γίνει η αναπαράσταση του δέντρου, αφού αυτή η πληροφορία δεν λαμβάνεται από την φυλογενετική ανάλυση. Η απόσταση από την ρίζα αντιπροσωπεύει την εξελικτική απόσταση των αλληλουχιών. Στα δέντρα χωρίς ρίζα δεν υπάρχει πληροφορία για τον κοινό πρόγονο, καθώς ούτε και για τον χρόνο που έχει μεσολαβήσει για την εξέλιξη των αλληλουχιών.



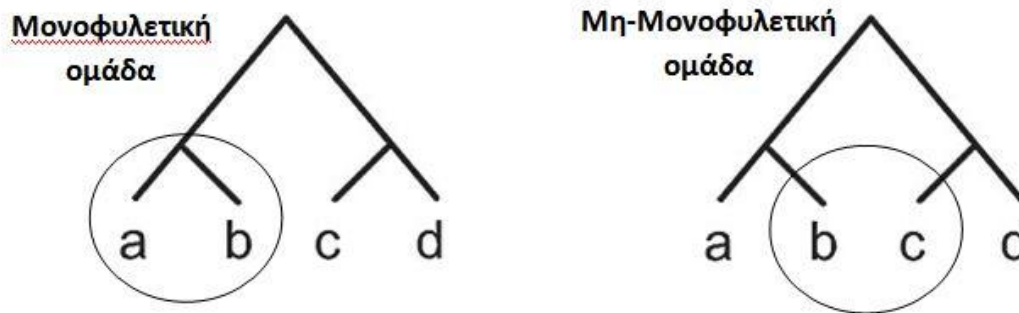
Εικόνα 12: Φυλογενετικό δέντρο (A) χωρίς ρίζα και (B) με ρίζα

Ο τρόπος με τον οποίο οι OTU διατάσσονται στο δέντρο είναι αυτό που ονομάζουμε τοπολογία. Η τοπολογία είναι ένα από τα βασικά χαρακτηριστικά ενός φυλογενετικού δέντρου καθώς μας δείχνει την σειρά με την οποία έχουν αποκλίνει οι διάφορες ταξινομικές μονάδες μεταξύ τους. Για παράδειγμα η τοπολογία του δέντρου της εικόνας 13(A) μας λέει ότι η αλληλουχία A βρίσκεται πιο κοντά εξελικτικά στη B, έπειτα στη C και τέλος στη D. Διατρέχοντας το δέντρο από πάνω προς τα κάτω η σειρά των διακλαδώσεων του δέντρου δείχνει τη σειρά με την οποία οι αλληλουχίες προβλέπονται ότι απέκλιναν από τον κοινό τους πρόγονο. Στο δέντρο της εικόνας 13(B) βλέπουμε την ίδια πληροφορία με το δέντρο της εικόνας 13(A) όσον αφορά την τοπολογία, μας δίνει όμως επιπλέον μια ποσοτική πληροφορία όσον αφορά την προβλεπόμενη απόσταση σε εξελικτικό χρόνο μεταξύ των αλληλουχιών και των κοινών τους προγόνων. Για παράδειγμα το μήκος σε κάθετη κατεύθυνση μεταξύ των εσωτερικών κόμβων x και y και των φύλλων αντιστοιχεί σε ένα μέτρο εξελικτικής απόστασης μεταξύ τους.



Εικόνα 13: (A) Κλαδόγραμμα χωρίς μήκος κλάδων, (B) Φυλόγραμμα όπου τα μήκη των κλάδων αντιστοιχούν σε χρόνους απόκλισης

Μια ομάδα αλληλουχιών μπορεί να ανήκει σε μια μονοφυλετική ομάδα (monophyletic group ή clade) ή όχι (Εικόνα 14). Στις μονοφυλετικές ομάδες ανήκουν οι αλληλουχίες από οργανισμούς με κοινή εξελικτική ιστορία, δηλαδή ανήκουν όλοι οι απόγονοι από έναν κοινό πρόγονο.



Εικόνα 14: Διάκριση μεταξύ μονοφυλετικής και μη-μονοφυλετικής ομάδας

Για την κατασκευή των φυλογενετικών δέντρων υπάρχουν διάφορες μέθοδοι οι οποίες χωρίζονται σε δύο ομάδες:

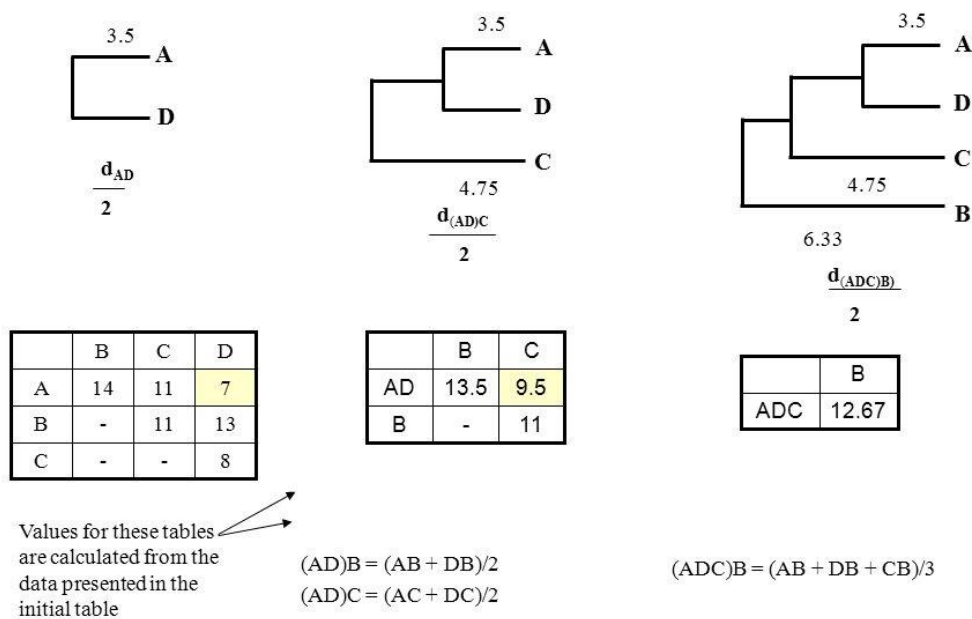
- Τις μεθόδους αποστάσεων (distance methods)
 - UPGMA (Unweighted pair-group method with arithmetic mean)
 - Neighbor - Joining
- Τις μεθόδους χαρακτήρων (character methods)
 - Μέθοδος μέγιστης φειδωλότητας (maximum parsimony)
 - Μέθοδος μέγιστης πιθανοφάνειας (maximum likelihood)
 - Μέθοδο Μπεϋζιανής συμπερασματολογίας (Bayesian inference)

Οι **μέθοδοι αποστάσεων** υπολογίζουν έναν πίνακα που περιέχει όλες τις ανά ζεύγος αποστάσεις μεταξύ των αλληλουχιών που θα συγκριθούν. Με βάση αυτόν τον πίνακα δημιουργείται ένα φυλογενετικό δέντρο του οποίου τόσο η τοπολογία, όσο και τα μήκη των κλάδων να συμφωνούν όσο το δυνατό περισσότερο με τις παρατηρούμενες αποστάσεις. (47)

Οι μέθοδοι αποστάσεων έχουν το πλεονέκτημα ότι είναι γρήγορες και βασίζονται σε εξελικτικά μοντέλα που περιέχουν διορθώσεις. Τα μειονεκτήματά τους είναι ότι υπολογίζουν μόνο μία τοπολογία, ότι εξαρτώνται από το εξελικτικό μοντέλο, είναι ευαίσθητες με τα κενά που τοποθετούνται κατά την στοίχιση των αλληλουχιών και ότι ελαχιστοποιούν την πληροφορία των αλληλουχιών σε έναν και μόνο αριθμό, την γενετική απόσταση.

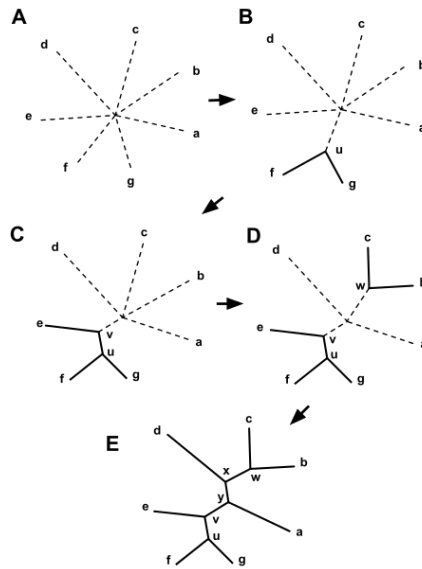
- **Unweighted pair-group method with arithmetic mean (UPGMA):** Η μέθοδος UPGMA αποτελεί την απλούστερη μέθοδο κατασκευής

φυλογενετικών δέντρων. Χρησιμοποιεί έναν επαναληπτικό αλγόριθμο ομαδοποίησης κατά τον οποίο κάθε αλληλουχία είναι μια OTU. Αρχικά υπολογίζονται όλες οι ανά δύο αποστάσεις των υπό μελέτη αλληλουχιών και αφού εντοπιστούν μέσα στον πίνακα οι δύο κοντινότερες εξελικτικά ακολουθίες, τότε ενώνονται μαζί σε μία κοινή ομάδα η οποία θεωρείται ως μία νέα OTU. Αυτή η διαδικασία επαναλαμβάνεται από την αρχή μέχρι να μείνουν ουσιαστικά δύο OTU. Έτσι σταδιακά δημιουργείται το φυλογενετικό δέντρο από τα φύλλα προς την ρίζα (Εικόνα 15). Η UPGMA βασίζεται στην παραδοχή ότι όλες οι ταξινομικές μονάδες παρουσιάζουν κοινό ρυθμό εξέλιξης, πράγμα που σπάνια ισχύει στην πράξη. (48)



Εικόνα 15: Σταδιακή κατασκευή φυλογενετικού δέντρου με τη μέθοδο UPGMA

- Neighbor-joining (NJ):** Η NJ μέθοδος παρουσιάζει ομοιότητες με την μέθοδο UPGMA χωρίς όμως να θεωρεί ότι ο ρυθμός εξέλιξης των υπό μελέτη OTUs είναι σταθερός, δημιουργώντας δέντρα χωρίς ρίζα. Στόχο έχει να εκτιμήσει το δέντρο ελάχιστης εξέλιξης, δηλαδή το δέντρο με το μικρότερο συνολικό μήκος. Η μέθοδος ξεκινά με μια αστεροειδή τοπολογία Εικόνα 16(A). Υπολογίζεται ο πίνακας αποστάσεων και επιλέγονται ως ζευγάρι «γείτονες» οι δύο OTUs με την ελάχιστη απόσταση. Αυτό το ζευγάρι συνδέεται με έναν νέο κόμβο Εικόνα 16(B). Στη συνέχεια συνδέεται ο κόμβος με το γειτονικό ζευγάρι OTUs με την μικρότερη εξελικτική απόσταση. Αυτό συνεχίζεται μέχρι να ολοκληρωθεί το δέντρο. (49)



Εικόνα 16: Σταδιακή κατασκευή φυλογενετικού δέντρου με τη μέθοδο Neighbor-Joining

Οι **μέθοδοι χαρακτήρων** βασίζονται και αυτές σε υποθέσεις και εξελικτικά μοντέλα, σε αντίθεση όμως με τις μεθόδους αποστάσεων, δεν μετασχηματίζουν τις αλληλουχίες. Επικεντρώνονται στις διαφορές χαρακτήρων ανά θέση μεταξύ των υπό μελέτη αλληλουχιών και όχι απλά στη γενετική τους απόσταση. (47)

- **Μέθοδος μέγιστης φειδωλότητας (maximum parsimony):** Η μέθοδος μέγιστης φειδωλότητας επικεντρώνεται στην ταυτοποίηση της τοπολογίας που απαιτεί τον μικρότερο αριθμό εξελικτικών αλλαγών προκειμένου να καταλήξει κανείς στις παρατηρούμενες διαφορές των υπό μελέτη ταξινομικών μονάδων. Για την κατασκευή του φυλογενετικού δέντρου πρώτα ταξινομούνται όλες οι θέσεις των υπό εξέταση αλληλουχιών οι οποίες μπορεί να διαφέρουν ή όχι και προσδιορίζονται ποιες είναι πληροφοριακές ή μη. Μια μεταβλητή θέση αποτελεί θέση πληροφορίας (informative site) αν ευνοεί μια τοπολογία σε σχέση με τις υπόλοιπες. Στη συνέχεια για κάθε δυνατή τοπολογία επιλέγεται ο ελάχιστος αριθμός αντικαταστάσεων που απαιτείται και τέλος επιλέγεται η τοπολογία με τον μικρότερο αριθμό αντικαταστάσεων. Αυτή η διαδικασία επαναλαμβάνεται για όλες τις θέσεις πληροφορίας. Με αυτή την μέθοδο υπολογίζεται μόνο η τοπολογία και όχι το μέτρο των εξελικτικών αποστάσεων (μήκος κλάδων). Πλεονεκτήματα της μεθόδου είναι η δυνατότητα μελέτης όλων των διαφορετικών τοπολογιών και η μη ελαχιστοποίηση της πληροφορίας που περιέχουν οι αλληλουχίες σε έναν αριθμό, όπως η γενετική απόσταση. Ως μειονεκτήματα μπορούν να αναφερθούν η εξαιρετικά αργή εκτέλεσή της, όταν ο αριθμός των αλληλουχιών είναι μεγάλος. Επίσης ότι ο αλγόριθμος δεν περιλαμβάνει διορθώσεις για τις πολλαπλές αντικαταστάσεις καθώς και ότι μειονεκτεί όταν ο ρυθμός εξέλιξης διαφέρει μεταξύ των λειτουργικών ταξινομικών μονάδων. Προκειμένου να ξεπεραστεί το πρόβλημα του μεγάλου

υπολογιστικού χρόνου έχουν προταθεί μέθοδοι ευρετικής αναζήτησης. (36)
(50)

- **Μέθοδος μέγιστης πιθανοφάνειας (maximum likelihood, ML):** Η μέθοδος μέγιστης πιθανοφάνειας εστιάζει στον υπολογισμό του φυλογενετικού δέντρου, του μήκους των κλάδων του δέντρου και των παραμέτρων του εξελικτικού μοντέλου, που μεγιστοποιούν την πιθανότητα να παρατηρηθούν τα υπό μελέτη δεδομένα. Χρειάζεται μεγάλη υπολογιστική δύναμη καθώς εκτιμάται κάθε πιθανό δέντρο. Σκοπός είναι να επιλεγθεί το δέντρο εκείνο από το οποίο είναι πιο πιθανό να έχουν προκύψει τα δεδομένα, το δέντρο δηλαδή με την μεγαλύτερη πιθανοφάνεια. (36) Η πιθανοφάνεια ορίζεται ως η δεσμευμένη πιθανότητα να παρατηρηθούν τα συγκεκριμένα δεδομένα του δείγματος, δεδομένου συγκεκριμένου εξελικτικού μοντέλου και συγκεκριμένου φυλογενετικού δέντρου, δηλαδή:

$$L(\text{εξελικτικό μοντέλο, δέντρο}) = \\ = \text{Prob}(\text{συστοιχία αλληλουχιών} \mid \text{εξελικτικό μοντέλο, δέντρο})$$

Σημαντικό είναι να προσδιοριστεί ότι η συνάρτηση πιθανοφάνειας εκφράζει την πιθανότητα να παρατηρήσουμε τα δεδομένα του δείγματος όπως προαναφέρθηκε και όχι την πιθανότητα των εκτιμώμενων παραμέτρων. Επίσης η πιθανοφάνεια δεν μας παρέχει πληροφορία για το αν είναι σωστές οι όχι οι παραδοχές του χρησιμοποιούμενου μοντέλου.

Τα πλεονεκτήματα της μεθόδου είναι ότι υπάρχει μεγάλη διαθεσιμότητα εξελιγμένων εξελικτικών μοντέλων έναντι της μέγιστης φειδωλότητας. Είναι κατάλληλη για αλληλουχίες DNA καθώς μπορούμε να μοντελοποιήσουμε την στοχαστική διαδικασία. Χρησιμοποιεί όλη την πληροφορία της αλληλουχίας και αξιολογεί διαφορετικές τοπολογίες δέντρων. Στα μειονεκτήματα συγκαταλέγεται η υπολογιστική δυσκολία που την καθιστά αργή μέθοδο (γι' αυτό και εδώ χρησιμοποιούνται μέθοδοι ευρετικής αναζήτησης) και ότι έχει φτωχές στατιστικές ιδιότητες αν το μοντέλο είναι κακώς προσδιορισμένο. (36)

- **Μέθοδο Μπεϋζιανής συμπερασματολογίας (Bayesian inference):** Η Μπεϋζιανή συμπερασματολογία βασίζεται στην ύστερη πιθανότητα μιας τοπολογίας (την πιθανότητα του δέντρου δοθέντος των δεδομένων), η οποία είναι διαφορετική από την μέθοδο μέγιστης πιθανοφάνειας, η οποία αναζητά την καλύτερη πιθανοφάνεια του δέντρου (την πιθανότητα των δεδομένων δοθέντος του δέντρου). Η Μπεϋζιανή προσέγγιση βασίζεται στο θεώρημα του Bayes και για την φυλογενετική συμπερασματολογία παίρνει την ακόλουθη μορφή:

$$P(Tree | Data) = \frac{P(Data | Tree)P(Tree)}{P(Data)}$$

Συνδυάζει την εκ των προτέρων πιθανότητα μια φυλογένειας $P(Tree)$, με την συνάρτηση πιθανοφάνειας $P(Data | Tree)$, για να παράγει μια ύστερη πιθανότητα στα δέντρα $P(Tree | Data)$. Η ύστερη πιθανότητα μπορεί να μεταφραστεί ως η πιθανότητα το δέντρο να είναι σωστό.

Για την προσέγγιση της ύστερης πιθανότητας χρησιμοποιείται ο αλγόριθμος Markov Chains Monte Carlo (MCMC). (51) Συνήθως όλα τα δέντρα τα θεωρούμε εκ των προτέρων ισοπίθανα. Έτσι κατασκευάζεται αρχικά ένα τυχαίο δέντρο και ένα νέο προτείνεται που γίνεται δεκτό ή απορρίπτεται σύμφωνα με τον τρόπο δειγματοληψίας που περιγράφηκε για πρώτη φορά από τους Metropolis και Hastings(52)(53)(54). Στο πρόσφατα αποδεκτό δέντρο γίνεται βελτιστοποίηση των παραμέτρων. Η διαδικασία συνεχίζεται για όλα τα δέντρα. Τελικά επιλέγεται το δέντρο με την μεγαλύτερη ύστερη πιθανότητα ως ο καλύτερος εκτιμητής της εξελικτικής ιστορίας.

Η κατασκευή των φυλογενετικών δέντρων στην παρούσα ανάλυση έγινε με το πρόγραμμα RAxML (Randomized Axelerated Maximum Likelihood), ιδανικό για μεγάλες βάσεις δεδομένων. Μεγάλο πλεονέκτημά του είναι ο γρήγορος αλγόριθμος του βασισμένος στην μέγιστη πιθανοφάνεια, επιστρέφοντας δέντρα με μεγάλες τιμές πιθανοφάνειας χρησιμοποιώντας το εξελικτικό μοντέλο GTR. (55) Ταυτόχρονα χρησιμοποιήθηκε σαν παράμετρος η Γάμμα κατανομή για την προσέγγιση του ρυθμού νουκλεοτιδικής αντικατάστασης μεταξύ των θέσεων του DNA όπως περιγράφηκε και στην προηγούμενη παράγραφο *Εξελικτικά μοντέλα*.

2.4.4 Αξιολόγηση των φυλογενετικών δέντρων

Το τελευταίο στάδιο της φυλογενετικής ανάλυσης έχει να κάνει με τον έλεγχο αξιοπιστίας των φυλογενετικών δέντρων. Έλεγχοι μπορούν να γίνουν και στα υπόλοιπα στάδια της ανάλυσης, όμως η κατασκευή των δέντρων περιέχει την μεγαλύτερη πιθανότητα σφάλματος εξαιτίας της επαγωγικής διαδικασίας και της πολυπλοκότητας των υπολογισμών που περιλαμβάνει. Ανάλογα με το είδος της φυλογενετικής ανάλυσης που έχει προηγηθεί υπάρχουν διάφοροι έλεγχοι. Ο πιο συνήθης στατιστικός έλεγχος είναι ο μη παραμετρικός bootstrapping (bootstrap)

Η μέθοδος Bootstrap είναι μια πολύ διαδεδομένη μέθοδος ελέγχου αξιοπιστίας δεδομένων, βασισμένη σε μια τεχνική επαναλαμβανόμενης δειγματοληψίας. Στο πλαίσιο της φυλογενετικής ανάλυσης πραγματοποιείται με τα ακόλουθα βήματα (56):

- **Βήμα 1:** Ανακατατάσσονται στήλες νουκλεοτιδίων από την αρχική αλληλουχία με τυχαίο τρόπο και προστίθενται σε μια νέα bootstrap στοίχιση με αντικατάσταση. Αυτό συνεχίζεται μέχρι η στοίχιση στην αλληλουχία bootstrap να έχει το ίδιο μήκος με την αρχική.
- **Βήμα 2:** Κατασκευάζονται δέντρα από την αρχική στοίχιση και την ανακαταταγμένη.
- **Βήμα 3:** Στο δέντρο που προκύπτει από κάθε ανακατάταξη καταμετρώνται οι κόμβοι που βρίσκονται στην ίδια θέση με το αρχικό δέντρο. Αν ένας κόμβος ενώνει τα ίδια δύο φύλλα ή τον ίδιο συνδυασμό κόμβων τότε η τιμή bootstrap του αυξάνεται κατά 1.
- **Βήμα 4:** Η διαδικασία των βημάτων 1 έως 3 επαναλαμβάνεται για ένα μεγάλο αριθμό.

Στο τέλος των επαναλήψεων αποδίδεται σε κάθε κόμβο η τιμή bootstrap που έχει υπολογιστεί ως ποσοστό με βάσει τον αριθμό N των επαναλήψεων. Αν για παράδειγμα σε 100 ανακατασκευασμένα δέντρα, τα 70 αναπαράγουν την παρουσία ενός συγκεκριμένου κλάδου, η τιμή bootstrap θα είναι 70 η οποία συνήθως λαμβάνεται ως το ελάχιστο για την εκκίνηση μιας τοπολογίας. Όσο μεγαλύτερη είναι η τιμή τόσο ο κόμβος είναι στατιστικά σημαντικός και κατά συνέπεια η συναγόμενη τοπολογία στη συγκεκριμένη θέση είναι βιολογικά σημαντική. Μεγάλες τιμές bootstrap για όλους τους κόμβους του δέντρου δηλώνει γενική αξιοπιστία.

Το πρόγραμμα RAxML που χρησιμοποιήθηκε κάνει bootstrap ανάλυση κάνοντας bootscanning analysis. (55) Όμως στην συνέχεια της μελέτης για την εύρεση όλων των τοπικών επιδημιών δεν χρησιμοποιήθηκε σαν κριτήριο η bootstrap τιμή αξιοπιστίας καθώς σε μία τόσο μεγάλη βάση δεδομένων εκ των πραγμάτων η ίδια η μέθοδος δεν μπορεί να δώσει υψηλά bootstrap values, όχι γιατί δεν υπάρχουν, αλλά γιατί όσο αυξάνεται ο αριθμός των αλληλουχιών τόσο μειώνεται η αξιοπιστία των εκτιμήσεων. (βιβλιογραφία) Παράλληλα σε ένα πλήρες γονιδίωμα δεν θα υπήρχε πρόβλημα με τα bootstrap values ανεξάρτητα από το πόσες αλληλουχίες υπάρχουν στη βάση δεδομένων, καθώς μελετάτε όλο το κομμάτι του ιού. Σε αυτή την περίπτωση όπως προαναφέρθηκε μελετάτε μόνο το γονίδιο *pol* και συγκεκριμένα το τμήμα της πρωτεΐσης και της αντίστροφης μεταγραφάσης.

2.5 Μέθοδος εύρεσης δικτύων από τα τελικά δέντρα

Ως μονοφυλετική ομάδα (τοπικό δίκτυο μετάδοσης) ορίστηκε κάθε φυλογενετική ομάδα αποτελούμενη από τουλάχιστον 2 αλληλουχίες με δειγματοληψία >70% από Ελλάδα. Σε κάθε set βρέθηκαν όλα τα τοπικά δίκτυα και κρατήθηκαν στο τέλος αυτά που επιβεβαιώθηκαν και από τις πέντε αναλύσεις για κάθε υπότυπο.

Η τοπική διασπορά του ιού δηλαδή το ποσοστό μονοφυλετικότητας εκτιμήθηκε υπολογίζοντας το πηλίκο του αριθμού των αλληλουχιών σε μονοφυλετικές ομάδες

και του αριθμού του συνόλου των διαθέσιμων αλληλουχιών από την Ελλάδα, ανά υπότυπο.

$$\% \text{ μονοφυλετικότητας υπότυπου} = \frac{\text{αριθμός αλληλουχιών σε μονοφυλετικές ομάδες}}{\text{σύνολο διαθέσιμων ελληνικών αλληλουχιών}}$$

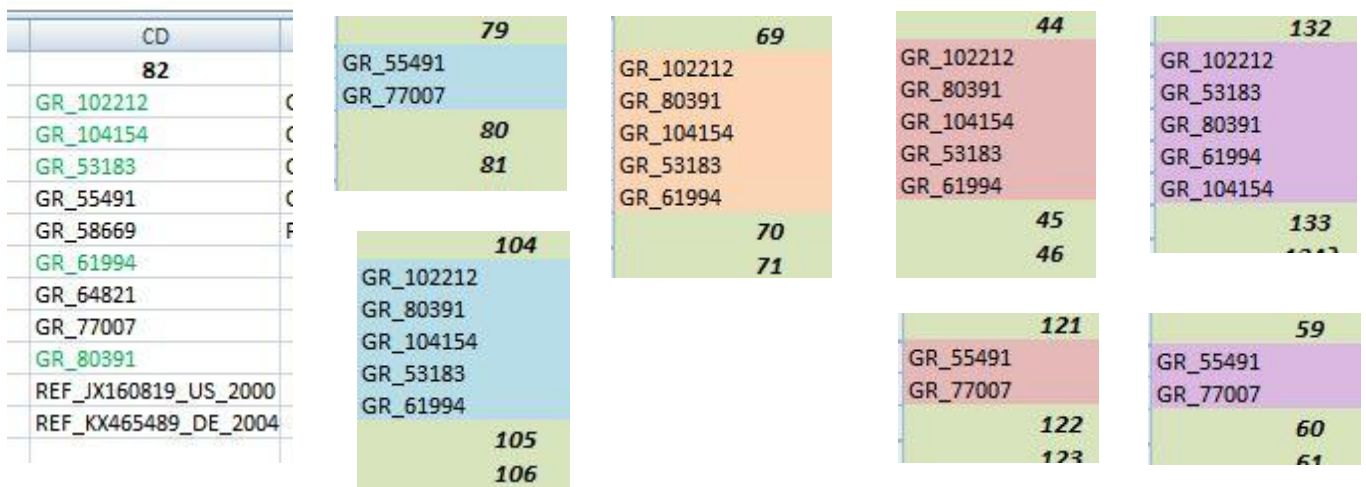
Η αναπαράσταση των δέντρων πραγματοποιήθηκε κυρίως με το πρόγραμμα Dendroscope 3.5.7. Το Dendroscope είναι ένα φιλικό προς τον χρήστη λογισμικό που επιτρέπει την επεξεργασία δέντρων με πολλές αλληλουχίες, καθώς επίσης μας επιτρέπει να απομονώσουμε από ένα δέντρο ένα δίκτυο και να αντιγράψουμε μόνο τις ονομασίες των αλληλουχιών του συγκεκριμένου δικτύου (<http://dendroscope.org>), κάτι που το πρόγραμμα FigTree δεν το επέτρεπε. Παρ' όλα αυτά το FigTree χρησιμοποιήθηκε και αυτό σε κάποια στάδια της ανάλυσης (Fig Tree v1.3.1). (57)

Οι ονομασίες των αλληλουχιών κάθε δικτύου από κάθε σετ αντιγράφονταν σε ένα αρχείο Microsoft Office Excel 2007. (58) Κάθε υπότυπος είχε πέντε φύλλα εργασίας, ένα για την κάθε ανάλυση, ενώ κάθε στήλη στα φύλλα εργασίας είχε τα ονόματα των αλληλουχιών που αποτελούσαν ένα τοπικό δίκτυο (Εικόνα 17).

Εικόνα 17: Άποψη από το πρόγραμμα Microsoft Office Excel για τον υπότυπο Β. Υπάρχουν 5 φύλλα εργασίας, ένα για κάθε ανάλυση, ενώ η κάθε στήλη περιλαμβάνει τα ονόματα των αλληλουχιών που βρέθηκαν στο ίδιο τοπικό δίκτυο.

Στη συνέχεια έπρεπε να βρεθούν ποια δίκτυα επιβεβαιώνονταν και από τις 5 αναλύσεις. Αυτό το σημείο ήταν το πιο δύσκολο και χρονοβόρο, ενώ απαιτούσε ιδιαίτερη προσοχή. Αυτό το βήμα θα γίνει πιο κατανοητό μέσα από το ακόλουθο παράδειγμα. Στην εικόνα 17 Παρατηρούμε ότι το 109^ο δίκτυο της δεύτερης ανάλυσης του υπότυπου Β αποτελείται από 3 αλληλουχίες (έχει επισημανθεί με κόκκινο ορθογώνιο σχήμα). Για να οριστεί ως επιβεβαιωμένο δίκτυο και από τις 5 αναλύσεις έπρεπε να βρεθεί ότι και στις υπόλοιπες 4 αναλύσεις αυτές οι

αλληλουχίες ομαδοποιούνταν μαζί. Ιδιαίτερη δυσκολία παρουσίασαν τα μεγαλύτερα τοπικά δίκτυα τα οποία σε πολλές περιπτώσεις δεν μπορούσαν να επιβεβαιωθούν και από τις 5 αναλύσεις, με αποτέλεσμα να επιμερίζονται σε μικρότερα επιμέρους δίκτυα στα οποία έπρεπε να γίνει εκ νέου η διαδικασία του ελέγχου και της επιβεβαίωσης από όλες τις αναλύσεις. Για παράδειγμα βλέπουμε στην εικόνα 18 το 82^ο δίκτυο από την πρώτη ανάλυση του υπότυπου Β που αποτελείται από 9 αλληλουχίες από Ελλάδα και 2 αλληλουχίες με παγκόσμια δειγματοληψία. Στην ίδια εικόνα παρατηρούμε ότι στις υπόλοιπες τέσσερις αναλύσεις οι αλληλουχίες του δικτύου 82 της πρώτης ανάλυσης εντοπίζονται σε ένα ή δυο επιμέρους δίκτυα. Εν τέλη μόνο 5 αλληλουχίες από Ελλάδα (πράσινη επισήμανση) εντοπίστηκαν στο ίδιο τοπικό δίκτυο σε όλες τις αναλύσεις.



Εικόνα 18: Τοπικά δίκτυα αλληλουχιών υπότυπου Β. Η κάθε ανάλυση έχει επισημανθεί με διαφορετικό χρώμα (1^η: λευκό, 2^η: γαλάζιο, 3^η: πορτοκαλί, 4^η: κόκκινο και 5^η: μωβ).

Με κατάλληλη επεξεργασία των δεδομένων, χρήση συναρτήσεων και φίλτρων βρέθηκαν τα δίκτυα που επιβεβαιώθηκαν και από τις πέντε αναλύσεις για τον κάθε υπότυπο.

Η φυλογεωγραφία (φυλογεωγραφική ανάλυση) περιλαμβάνει τις αρχές και τις διαδικασίες που καθορίζουν τα γεωγραφικά πρότυπα των γενεαλογικών γραμμών, ειδικά εκείνων εντός ενός είδους, αλλά ακόμα και μεταξύ στενά συγγενικών εξελικτικών ειδών. (59) Η πιο απλή της μορφή αφορά τη χωρική κατανομή των απλοτύπων, των οποίων οι φυλογενετικές σχέσεις είναι γνωστές ή μπορούν να εκτιμηθούν. Η φυλογεωγραφική ανάλυση στην παρούσα μελέτη πραγματοποιήθηκε με το πρόγραμμα Mesquite v3.01, λογισμικό που χρησιμοποιείται στην εξελικτική βιολογία αναλύοντας συγκριτικά στοιχεία για τους οργανισμούς. Το Mesquite εκτιμά για κάθε κόμβο του φυλογενετικού δέντρου τον προγονικό χαρακτήρα. (60)

2.6 Στατιστική ανάλυση

Η εκτίμηση των παραγόντων και τα χαρακτηριστικά του πληθυσμού που σχετίζονται με την τοπική διασπορά του HIV για τους υπότυπους A1 και B πραγματοποιήθηκε με τη μέθοδο της λογιστικής παλινδρόμησης στο πρόγραμμα STATA 12. (61) Πραγματοποιήθηκαν δύο αναλύσεις, μία για τον υπότυπο A1 και μια για τον B αντίστοιχα. Και στα δύο μοντέλα η εξαρτημένη μεταβλητή ήταν δίτιμη με τιμές 0: το άτομο δεν βρίσκεται σε τοπικό δίκτυο και 1: το άτομο βρίσκεται σε τοπικό δίκτυο. Οι ανεξάρτητες μεταβλητές ήταν:

- το φύλο (Άνδρας, Γυναίκα)
- η εθνικότητα (Ελληνική, μη-Ελληνική, Άγνωστη)
- η ομάδα κινδύνου (MSW: Men having Sex with Women (Ετεροφυλόφιλοι), PWID: People Who Inject Drugs (Χρήστες ενδοφλέβιων ναρκωτικών), MSM: Men having Sex with Men (Άνδρες που κάνουν σεξ με άνδρες), Άλλη, Άγνωστη)
- η περίοδος δειγματοληψίας ([1996-2001), [2001-2006), [2006-2011), [2011-2016))
- η ηλικία

Κεφάλαιο 3: Αποτελέσματα

3.1 Περιγραφικά στοιχεία (A1&B – A1 – B)

Συνολικά και για τους δύο υπότυπους (A1 και B) υπήρχαν 3.386 αλληλουχίες. Από αυτές το 87,27% ήταν άνδρες, το 62,02% ελληνικής εθνικότητας, το 67,67% MSM (Men having Sex with Men (Άνδρες που κάνουν σεξ με άνδρες)), το 41,11% είχε διαθέσιμο δείγμα με περίοδο δειγματοληψίας από το 2011 έως τον Ιούνιο του 2015 και τέλος το 63,73% είχε μολυνθεί με στελέχη υπότυπου B (Πίνακας 4). Αναφορικά με την ηλικία, το εύρος ήταν από 1 έως 87 έτη, ενώ η μέση ηλικία των οροθετικών της μελέτης ήταν 38 έτη (τυπική απόκλιση 11,3 έτη).

Πίνακας 4: Συχνότητες ποιοτικών χαρακτηριστικών συνολικά και για τους δύο υπότυπους (A1 και B), N = 3.386 αλληλουχίες

Μεταβλητή		Συχνότητα	Ποσοστό (%)
Φύλο	Άνδρας	2955	87,27
	Γυναίκα	424	12,52
	Άγνωστο	7	0,21
Εθνικότητα	Ελληνική	2100	62,02
	Μη-ελληνική	250	7,38
	Άγνωστη	1036	30,60
Ομάδα κινδύνου	MSW	580	17,13
	PWID	315	9,30
	MSM	2122	62,67
	Άλλη	90	2,66
	Άγνωστη	279	8,24
Περίοδος δειγματοληψίας	[1996 – 2001)	386	11,40
	[2001 – 2006)	788	23,27
	[2006 – 2011)	820	24,22
	[2011 – 2016)	1392	41,11
Υπότυπος	A1	1230	36,33
	B	2156	63,67

MSW: Men having Sex with Women (Ετεροφυλόφιλοι)
PWID: People Who Inject Drugs (Χρήστες ενδοφλέβιων ναρκωτικών)

MSM: Men having Sex with Men (Άνδρες που κάνουν σεξ με άνδρες)

Για τον υπότυπο A1 αναλύθηκαν 1230 αλληλουχίες από Ελλάδα. Από αυτές το 84,63% είχε απομονωθεί από άνδρες, το 55,20% από οροθετικούς ελληνικής εθνικότητας, το 57,24% από MSM και τέλος το 49,51% είχε συλλεγεί το διάστημα μεταξύ 2011 και Ιούνιο του 2015 (Πίνακας 5). Το εύρος της ηλικίας των οροθετικών που είχαν μολυνθεί με στελέχη υπότυπου A1 ήταν από 1 έως 79 έτη, ενώ η μέση ηλικία ήταν 38,5 έτη και η τυπική απόκλιση 12,3 έτη.

Πίνακας 5: Συχνότητες ποιοτικών χαρακτηριστικών και για τον υπότυπο A1, N = 1230 αλληλουχίες

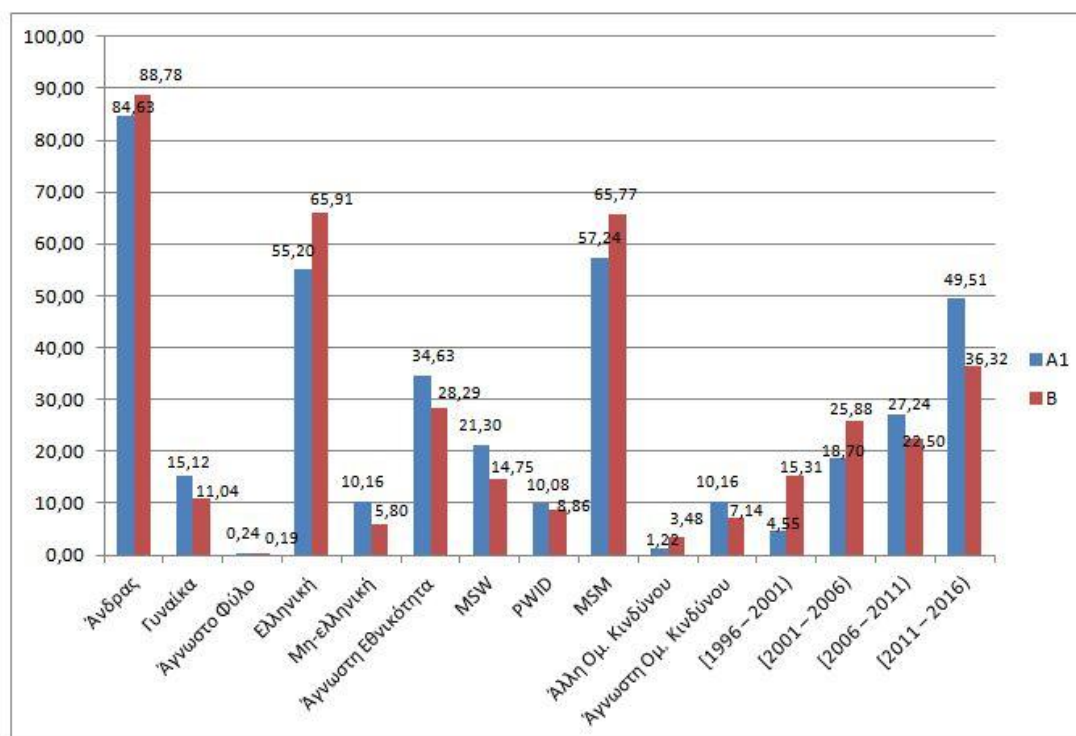
Μεταβλητή		Συχνότητα	Ποσοστό (%)
Φύλο	Άνδρας	1041	84,63
	Γυναίκα	186	15,12
	Άγνωστο	3	0,24
Εθνικότητα	Ελληνική	679	55,20
	Μη-ελληνική	125	10,16
	Άγνωστη	426	34,63
Ομάδα κινδύνου	MSW	262	21,30
	PWID	124	10,08
	MSM	704	57,24
	Άλλη	15	1,22
	Άγνωστη	125	10,16
Περίοδος δειγματοληψίας	[1996 – 2001)	56	4,55
	[2001 – 2006)	230	18,70
	[2006 – 2011)	335	27,24
	[2011 – 2016)	609	49,51
MSW: Men having Sex with Women (Ετεροφυλόφιλοι)			
PWID: People Who Inject Drugs (Χρήστες ενδοφλέβιων ναρκωτικών)			
MSM: Men having Sex with Men (Άνδρες που κάνουν σεξ με άνδρες)			

Για υπότυπο Β αναλύσαμε 2.156 αλληλουχίες από Ελλάδα. Το 88,78% είχε απομονωθεί από άνδρες, το 65,91% από οροθετικούς ελληνικής εθνικότητας, το 65,77% από MSM και το 36,32% είχε συλλεγεί τη περίοδο από το 2011 έως τον Ιούνιο του 2015 (Πίνακας 6). Τέλος, η ηλικία των οροθετικών αυτών ήταν από 1 έως 87 έτη, η μέση ηλικία 37,7 έτη (τυπική απόκλιση 10,5 έτη).

Πίνακας 6: Συχνότητες ποιοτικών χαρακτηριστικών και για τον υπότυπο Β, N = 2.156 αλληλουχίες

Μεταβλητή		Συχνότητα	Ποσοστό (%)
Φύλο	Άνδρας	1914	88,78
	Γυναίκα	238	11,04
	Άγνωστο	4	0,19
Εθνικότητα	Ελληνική	1421	65,91
	Μη-ελληνική	125	5,80
	Άγνωστη	610	28,29
Ομάδα κινδύνου	MSW	318	14,75
	PWID	191	8,86
	MSM	1418	65,77
	Άλλη	75	3,48
	Άγνωστη	154	7,14
Περίοδος δειγματοληψίας	[1996 – 2001)	330	15,31
	[2001 – 2006)	558	25,88
	[2006 – 2011)	485	22,50
	[2011 – 2016)	783	36,32
MSW: Men having Sex with Women (Ετεροφυλόφιλοι)			
PWID: People Who Inject Drugs (Χρήστες ενδοφλέβιων ναρκωτικών)			
MSM: Men having Sex with Men (Άνδρες που κάνουν σεξ με άνδρες)			

Στο Γράφημα 8 παρατηρούμε ότι και οι δύο υπότυποι έχουν το μεγαλύτερο ποσοστό τους στις ίδιες κατηγορίες σε όλες τις υπο μελέτη μεταβλητές.



Γράφημα 8: Ραβδογράμμα περιγραφικών χαρακτηριστικών των υποτύπων A1 και B

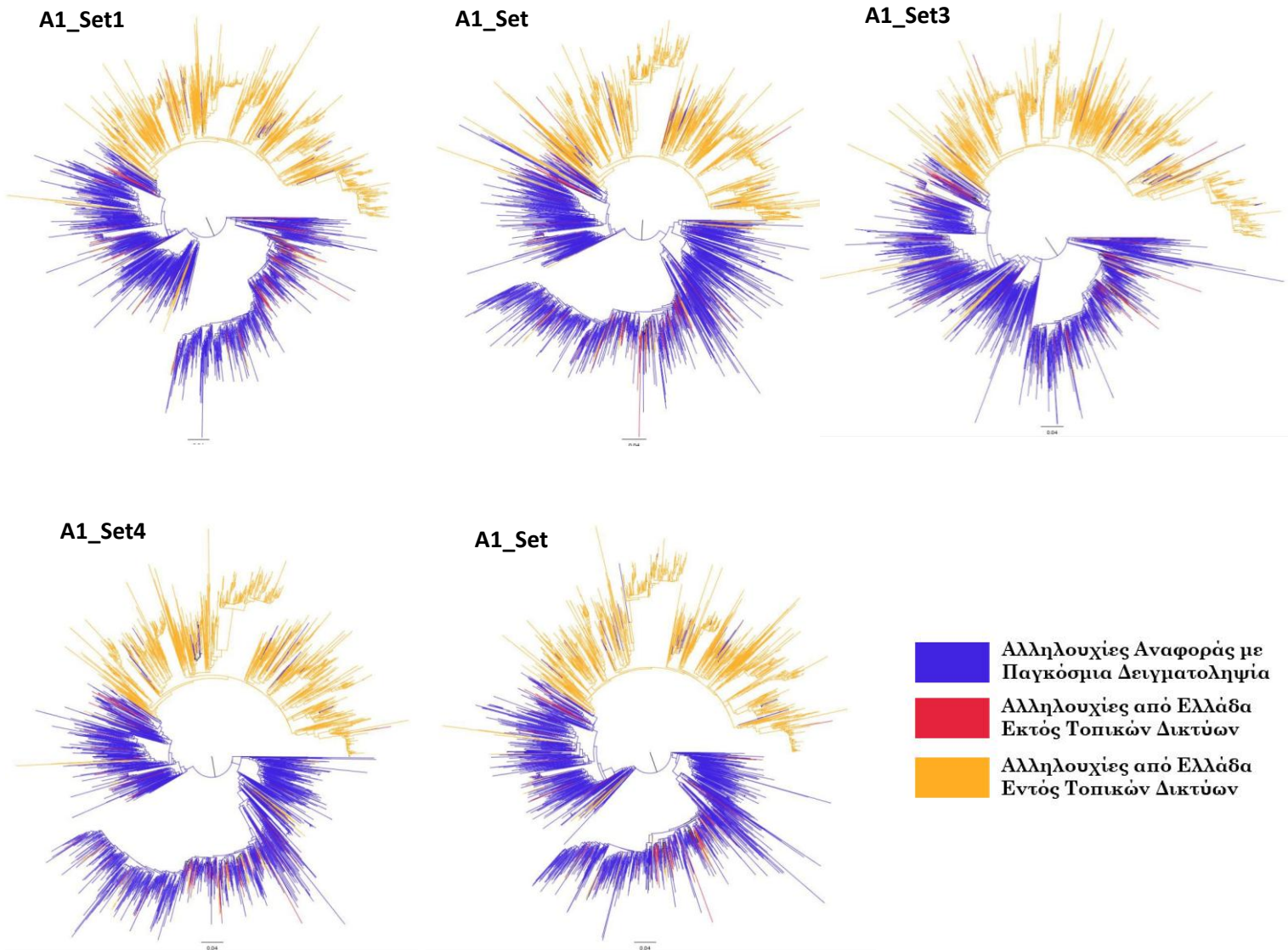
3.2 Υπότυπος A1

Για τον υπότυπο A1 πραγματοποιήθηκαν 5 φυλογενετικές αναλύσεις όπου το κάθε σύνολο δεδομένων αποτελούνταν από όλες τις διαθέσιμες αλληλουχίες από Ελλάδα για τον υπότυπο A1 (1.230) και από 1.500 τυχαία επιλεγμένες αλληλουχίες αναφοράς. Το μήκος των νουκλεοτιδικών αλληλουχιών ήταν 897 βάσεις. Η κάθε ανάλυση έδωσε 360 bootstrap φυλογενετικά δέντρα καθώς και το δέντρο μέγιστη πιθανοφάνειας. Στον πίνακα 7 που ακολουθεί αναγράφονται για κάθε μια από τις πέντε αναλύσεις που πραγματοποιήθηκαν ο αριθμός των τοπικών δικτύων μετάδοσης του ιού που εντοπίστηκαν, καθώς και ο μέγιστος και ελάχιστος αριθμός ελληνικών αλληλουχιών που βρέθηκαν εντός των δικτύων.

Πίνακας 7: Αριθμός τοπικών δικτύων μετάδοσης των 5 φυλογενετικών αναλύσεων για τον υπότυπο A1 καθώς και ο μέγιστος και ελάχιστος αριθμός αλληλουχιών που εντοπίστηκαν στα δίκτυα

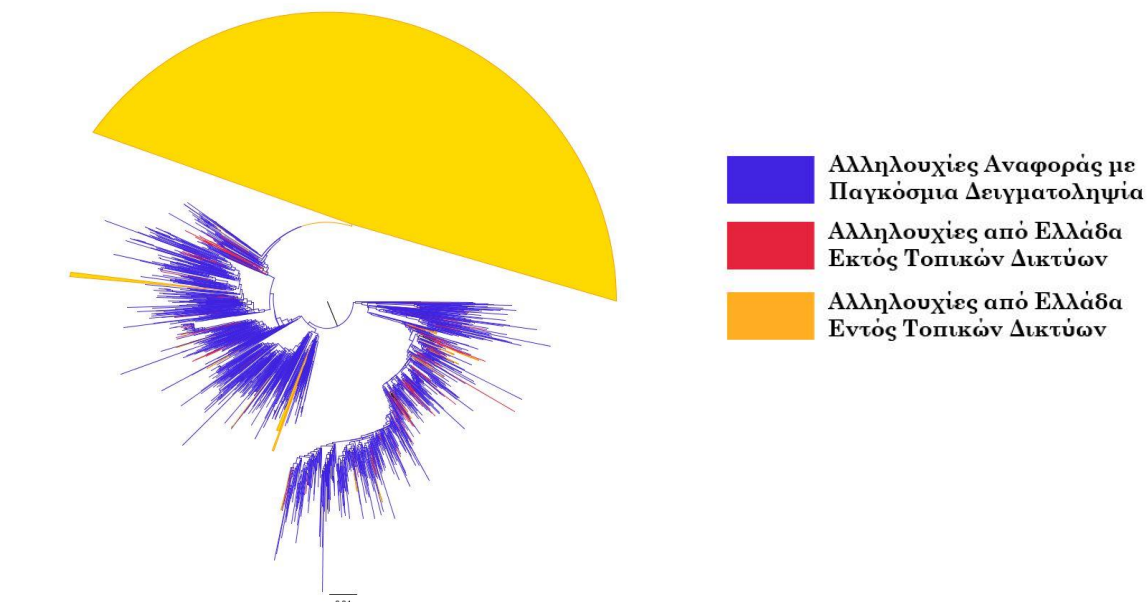
Set	Αριθμός τοπικών δικτύων	Ελάχιστος αριθμός αλληλουχιών	Μέγιστος αριθμός αλληλουχιών
1	29	2	1077
2	25	2	1078
3	27	2	1078
4	25	2	1060
5	26	2	1063

Οι πέντε αναλύσεις μας έδωσαν τα ακόλουθα φυλογενετικά δέντρα (Εικόνα 19):



Εικόνα 19: Τα 5 φυλογενετικά δέντρα που προέκυψαν από τις 5 αναλύσεις (5 Sets) για τον υπότυπο A1

Τα τοπικά δίκτυα μετάδοσης του ιού που επιβεβαιώθηκαν και από τις πέντε αναλύσεις ήταν 23 (Εικόνα 20). Συνολικά βρέθηκαν σε τοπικά δίκτυα 1.125 ελληνικές αλληλουχίες υπότυπου A1 από τις 1.230 που αναλύθηκαν. Τα δίκτυα είχαν εύρος από 2 έως 1.059 αλληλουχίες (Πίνακας 8)



Εικόνα 20: Επιβεβαιωμένα τοπικά δίκτυα και από τις 5 αναλύσεις για τον υπότυπο A1

Πίνακας 8: Αριθμός αλληλουχιών εντός των επιβεβαιωμένων δικτύων, όπως αυτά προέκυψαν από τις 5 αναλύσεις για τον υπότυπο A1

Αριθμός αλληλουχιών	Αριθμός δικτύων	Ποσοστό %
2	15	65,22
3	3	13,04
5	2	8,70
8	1	4,35
9	1	4,35
1059	1	4,35
Σύνολο	23	100,00

Στον Πίνακα 9 που ακολουθεί παρουσιάζονται τα ποσοστά μονοφυλετικότητας της κάθε ανάλυσης για τον υπότυπο A1 αλλά και το ποσοστό μονοφυλετικότητας από τα τοπικά δίκτυα που επιβεβαιώθηκαν και από τις 5 αναλύσεις το οποίο ήταν 91,46%.

Όπως περιγράφηκε και στην ενότητα *Υλικό και Μέθοδος* τα ποσοστά μονοφυλετικότητας βγήκαν με τις ακόλουθες πράξεις:

π.χ.

$$\begin{aligned}
 & \% \text{ μονοφυλετικότητας υπότυπου A1} = \\
 & = \frac{\text{αριθμός αλληλουχιών σε μονοφυλετικές ομάδες οι οποίες επιβεβαιώθηκαν από όλες τις αναλύσεις}}{\text{σύνολο διαθέσιμων ελληνικών αλληλουχιών}} = \\
 & = \frac{1125}{1230} = 91,46\%
 \end{aligned}$$

$$\% \text{ μονοφυλετικότητας υπότυπου A1 set 1} = \frac{\text{αριθμός αλληλουχιών σε μονοφυλετικές ομάδες της πρώτης ανάλυσης}}{\text{σύνολο διαθέσιμων ελληνικών αλληλουχιών}} = \frac{1166}{1230} = 94,80\%$$

Πίνακας 9: Ποσοστά μονοφυλετικότητας για τον υπότυπο A1 για την κάθε ανάλυση ξεχωριστά, και για το σύνολο των επιβεβαιωμένων δικτύων από τις 5 αναλύσεις

# όλων των ελληνικών αλληλουχιών υπότυπου A1	1230	Ποσοστό % μονοφυλετικότητας σε κάθε ανάλυση	Ποσοστό % μονοφυλετικότητας επιβεβαιωμένο και από τα 5 sets
# αλληλουχιών σε επιβεβαιωμένα δίκτυα και από τα 5 sets	1125		91,46
Set 1 # Αλληλουχιών που βρέθηκαν σε τοπικό δίκτυο	1166	94,80	
Set 2 # Αλληλουχιών που βρέθηκαν σε τοπικό δίκτυο	1158	94,15	
Set 3 # Αλληλουχιών που βρέθηκαν σε τοπικό δίκτυο	1166	94,80	
Set 4 # Αλληλουχιών που βρέθηκαν σε τοπικό δίκτυο	1155	93,90	
Set 5 # Αλληλουχιών που βρέθηκαν σε τοπικό δίκτυο	1165	94,72	
#: αριθμός, πλήθος			

Πίνακας 10: Συχνότητες περιγραφικών χαρακτηριστικών για τον υπότυπο A1 με κριτήριο αν είναι οι αλληλουχίες σε τοπικό δίκτυο ή όχι

Μεταβλητή		Αλληλουχίες σε τοπικό δίκτυο N=1.125		Αλληλουχίες εκτός τοπικού δικτύου N=105	
		Συχνότητα	Ποσοστό (%)	Συχνότητα	Ποσοστό (%)
Φύλο	Άνδρας	984	87,47	57	54,29
	Γυναίκα	138	12,27	48	45,71
	Άγνωστο	3	0,27	-	-
Εθνικότητα	Ελληνική	653	58,04	26	24,76
	Μη-ελληνική	80	7,11	45	42,86
	Άγνωστη	392	34,84	34	32,38
Ομάδα κινδύνου	MSW	208	18,49	54	51,43
	PWID	106	9,42	18	17,14
	MSM	692	61,51	12	11,43
	Άλλη	14	1,24	1	0,95
	Άγνωστη	105	9,33	20	19,05
Περίοδος δειγματοληψίας	[1996 – 2001)	52	4,62	4	3,81
	[2001 – 2006)	212	18,84	18	17,14
	[2006 – 2011)	296	26,31	39	37,14
	[2011 – 2016)	565	50,22	44	41,90

MSW: Men having Sex with Women (Ετεροφυλόφιλοι)
PWID: People Who Inject Drugs (Χρήστες ενδοφλέβιων ναρκωτικών)
MSM: Men having Sex with Men (Άνδρες που κάνουν σεξ με άνδρες)

Στον πίνακα 10 παρατηρούμε ότι οι αλληλουχίες των ατόμων που εντοπίστηκαν σε τοπικά δίκτυα μετάδοσης του ιού είχαν απομονωθεί στην πλειονότητα τους από MSM ελληνικής εθνικότητας, ενώ οι αλληλουχίες οι οποίες εντοπίστηκαν εκτός δικτύων από MSW μη – ελληνικής εθνικότητας.

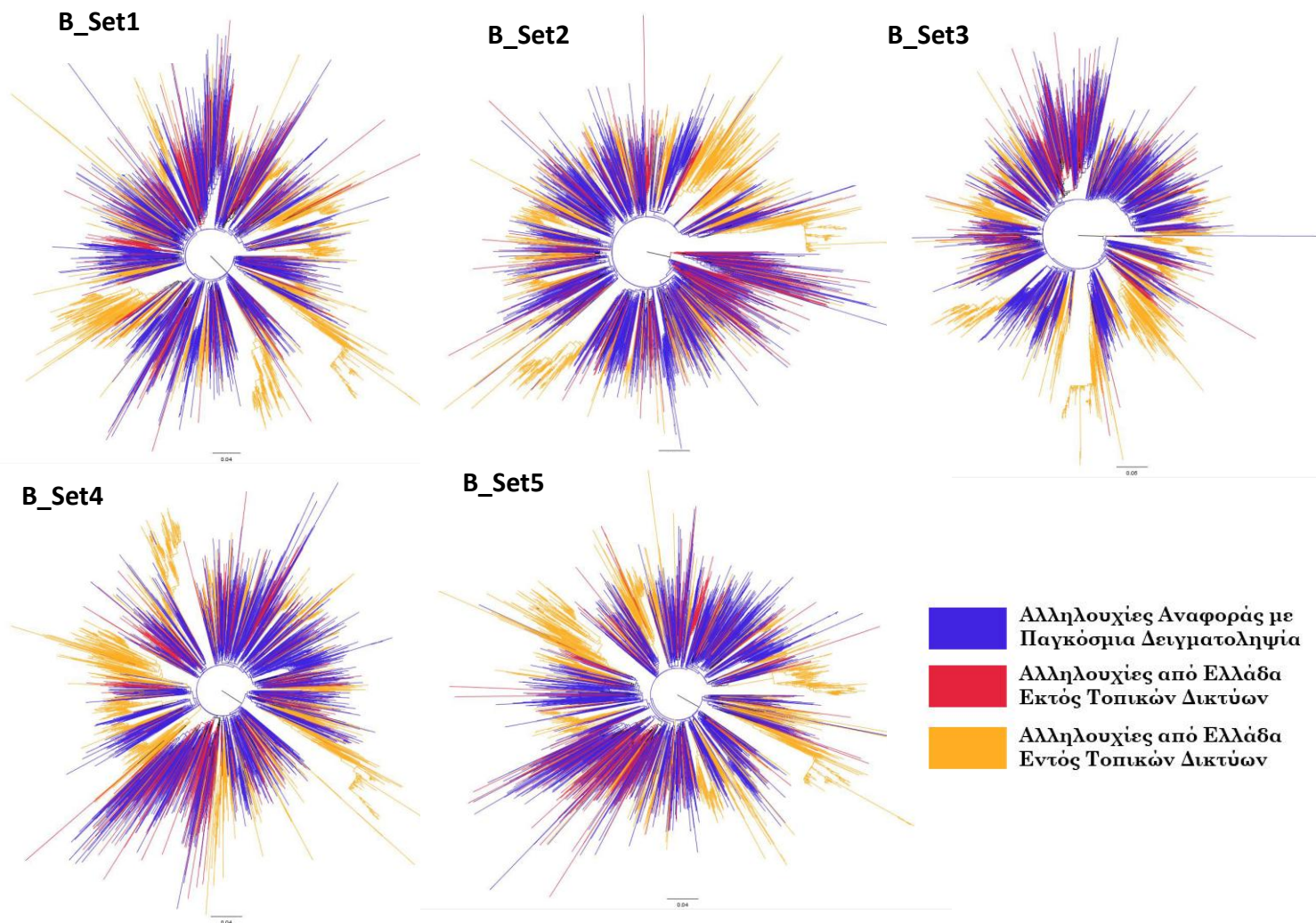
3.3 Υπότυπος Β

Όπως και στον υπότυπο Α1, στον υπότυπο Β πραγματοποιήθηκαν 5 φυλογενετικές αναλύσεις όπου το κάθε σύνολο δεδομένων αποτελούνταν από όλες τις διαθέσιμες ελληνικές αλληλουχίες υπότυπου Β (2.156) και από 2.000 τυχαία επιλεγμένες αλληλουχίες αναφοράς. Τα μήκος των αλληλουχιών ήταν 891 νουκλεοτιδικές βάσεις. Η κάθε ανάλυση έδωσε 300 φυλογενετικά δέντρα όπου μελετήθηκε τελικά το δέντρο μέγιστης πιθανοφάνειας. Στον Πίνακα 11 που ακολουθεί αναφέρονται για κάθε ανάλυση ο αριθμός των τοπικών δικτύων μετάδοσης που εντοπίστηκαν και ποιος ήταν ο μέγιστος και ο ελάχιστος αριθμός ελληνικών αλληλουχιών που εντοπίστηκαν σε δίκτυα.

Πίνακας 11: Αριθμός τοπικών δικτύων των 5 φυλογενετικών αναλύσεων για τον υπότυπο Β καθώς και ο μέγιστος και ελάχιστος αριθμός αλληλουχιών που εντοπίστηκαν σε δίκτυα για κάθε ανάλυση.

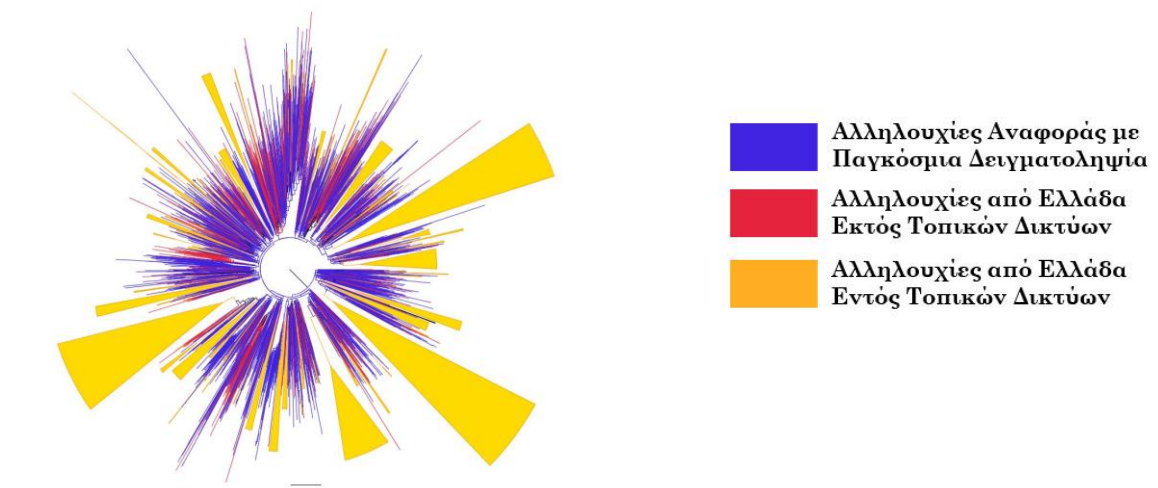
Set	Αριθμός τοπικών δικτύων	Ελάχιστος αριθμός αλληλουχιών	Μέγιστος αριθμός αλληλουχιών
1	150	2	311
2	152	2	225
3	153	2	238
4	161	2	253
5	134	2	303

Οι πέντε αναλύσεις μας έδωσαν τα ακόλουθα φυλογενετικά δέντρα (Εικόνα 21):



Εικόνα 21: Τα 5 φυλογενετικά δέντρα που προέκυψαν από τις 5 αναλύσεις (5 Sets) για τον υπότυπο B

Τα τοπικά δίκτυα μετάδοσης του ιού που επιβεβαιώθηκαν και από τις πέντε αναλύσεις ήταν 133 (Εικόνα 22). Συνολικά βρέθηκαν σε τοπικά δίκτυα 1.570 από τις 2.156 ελληνικές αλληλουχίες υπότυπου B που αναλύθηκαν. Τα δίκτυα είχαν εύρος από 2 έως 199 αλληλουχίες (Πίνακας 12).



Πίνακας 12: Αριθμός αλληλουχιών εντός των επιβεβαιωμένων δικτύων, όπως αυτά προέκυψαν από τις 5 αναλύσεις για τον υπότυπο Β

Αριθμός αλληλουχιών	Αριθμός δικτύων	Ποσοστό %
2	51	38,35
3	24	18,05
4	13	9,77
5	12	9,02
6	5	3,76
7	1	0,75
8	2	1,50
9	1	0,75
10	2	1,50
11	2	1,50
13	1	0,75
17	3	2,26
21	1	0,75
24	1	0,75
25	2	1,50
28	1	0,75
30	2	1,50
35	1	0,75
37	1	0,75
46	1	0,75
60	1	0,75
92	1	0,75
131	1	0,75
160	1	0,75
173	1	0,75
199	1	0,75
Σύνολο	133	100,00

Στον Πίνακα 13 που ακολουθεί παρουσιάζονται τα ποσοστά μονοφυλετικότητας της κάθε ανάλυσης για τον υπότυπο Β αλλά και το ποσοστό μονοφυλετικότητας από τα τοπικά δίκτυα που επιβεβαιώθηκαν και από τις 5 αναλύσεις το οποίο ήταν 72,82%.

Όπως περιγράφηκε και στην ενότητα *Υλικό και Μέθοδος* τα ποσοστά μονοφυλετικότητας βγήκαν με τις ακόλουθες πράξεις:

π.χ.

$$\begin{aligned} & \% \text{ μονοφυλετικότητας υπότυπου B} = \\ & = \frac{\text{αριθμός αλληλουχιών σε μονοφυλετικές ομάδες οι οποίες επιβεβαιώθηκαν από όλες τις αναλύσεις}}{\text{σύνολο διαθέσιμων ελληνικών αλληλουχιών}} = \\ & = \frac{1570}{2156} = 72,82\% \end{aligned}$$

$$\begin{aligned} & \% \text{ μονοφυλετικότητας υπότυπου B set 1} = \\ & = \frac{\text{αριθμός αλληλουχιών σε μονοφυλετικές ομάδες της πρώτης ανάλυσης}}{\text{σύνολο διαθέσιμων ελληνικών αλληλουχιών}} = \\ & = \frac{1908}{2156} = 88,50\% \end{aligned}$$

Πίνακας 13: Ποσοστά μονοφυλετικότητας για τον υπότυπο B για την κάθε ανάλυση ξεχωριστά αλλά και το αντίστοιχο ποσοστό από τα επιβεβαιωμένα δίκτυα

Όλες οι ελληνικές αλληλουχίες υπότυπου B	2156	Ποσοστό % αλληλουχιών σε δίκτυο σε κάθε set	Ποσοστό % αλληλουχιών σε δίκτυο επιβεβαιωμένο και από τα 5 sets
# αλληλουχιών σε δίκτυο επιβεβαιωμένο από τα 5 sets	1570		72,82
Set 1 # Αλληλουχιών που βρέθηκαν σε τοπικό δίκτυο	1908	88,50	
Set 2 # Αλληλουχιών που βρέθηκαν σε τοπικό δίκτυο	1858	86,18	
Set 3 # Αλληλουχιών που βρέθηκαν σε τοπικό δίκτυο	1866	86,55	
Set 4 # Αλληλουχιών που βρέθηκαν σε τοπικό δίκτυο	1883	87,34	
Set 5 # Αλληλουχιών που βρέθηκαν σε τοπικό δίκτυο	1914	88,78	
#: αριθμός, πλήθος			

Πίνακας 14: Συχνότητες περιγραφικών χαρακτηριστικών για τον υπότυπο Β με κριτήριο αν οι αλληλουχίες εντοπίστηκαν σε τοπικό δίκτυο ή όχι

Μεταβλητή		Αλληλουχίες σε τοπικό δίκτυο N=1.570		Αλληλουχίες εκτός τοπικού δικτύου N=586	
		Συχνότητα	Ποσοστό (%)	Συχνότητα	Ποσοστό (%)
Φύλο	Άνδρας	1410	89,81	504	86,01
	Γυναίκα	157	10,00	81	13,82
	Άγνωστο	3	0,19	1	0,17
Εθνικότητα	Ελληνική	998	63,57	423	72,18
	Μη-ελληνική	85	5,41	40	6,83
	Άγνωστη	487	31,02	123	20,99
Ομάδα κινδύνου	MSW	206	13,12	112	19,11
	PWID	157	10,00	34	5,80
	MSM	1063	67,71	355	60,58
	Άλλη	23	1,46	52	8,87
	Άγνωστη	121	7,71	33	5,63
Περίοδος δειγματοληψίας	[1996 – 2001)	154	9,81	176	30,03
	[2001 – 2006)	357	22,74	201	34,30
	[2006 – 2011)	384	24,46	101	17,24
	[2011 – 2016)	675	42,99	108	18,43

MSW: Men having Sex with Women (Ετεροφυλόφιλοι)
PWID: People Who Inject Drugs (Χρήστες ενδοφλέβιων ναρκωτικών)
MSM: Men having Sex with Men (Άνδρες που κάνουν σεξ με άνδρες)

Στον πίνακα 14 παρατηρούμε ότι τόσο οι αλληλουχίες που εντοπίστηκαν σε τοπικά δίκτυα μετάδοσης όσο και οι αλληλουχίες που εντοπίστηκαν εκτός δικτύων, είχαν απομονωθεί κυρίως από MSM ελληνικής εθνικότητας. Το μεγαλύτερο ποσοστό των αλληλουχιών που εντοπίστηκε σε δίκτυα είχε πρόσφατη δειγματοληψία.

3.4 Στατιστική Ανάλυση

Η ανάλυση λογιστικής παλινδρόμησης έδειξε ότι η τοπική διασπορά του υπότυπου A1 σχετίζεται με άνδρες (OR=1,8; $p=0,034$), που κάνουν σεξ με άνδρες (MSM) (OR=7,2 ; $p<0.001$) και είναι ελληνικής εθνικότητας (OR=7,2 ; $p<0.001$)

Για την τοπική διασπορά του υπότυπου Β η ανάλυση λογιστικής παλινδρόμησης (Πίνακας 15) έδειξε ότι σχετίζεται με οροθετικούς ελληνικής εθνικότητας (OR=1,70 ; $p<0.019$). Επίσης αναφορικά με την περίοδο δειγματοληψίας, παρατηρήθηκε μια αυξητική τάση (trend) στο Odds Ratio όσο η περίοδος δειγματοληψίας ήταν πιο πρόσφατη, ξεκινώντας με OR=2,09 ($p<0.001$) για το διάστημα 2001-2006 και φτάνοντας στο OR= 7,21 ($p<0.001$) για το 2011-2015.

Πίνακας 15: Λογιστική παλινδρόμηση για τους υπότυπους Α1 και Β. (Η εξαρτημένη μεταβλητή είναι δίτιμη με τιμές 0: το άτομο δεν βρίσκεται σε τοπικό δίκτυο, 1: το άτομο βρίσκεται σε τοπικό δίκτυο)

Μεταβλητή	Υπότυπος Α1			Υπότυπος Β		
	OR	95% Δ.Ε.	p-value	OR	95% Δ.Ε.	p-value
Φύλο (*Γυναίκες)						
Άνδρες	1.80	1.04 3.04	0.034	1.16	0.77 1.74	0.473
Άγνωστο	---	---	---	0.67	0.60 7.39	0.744
Εθνικότητα (*Ελληνική)						
Μη – Ελληνική	0.14	0.08 0.25	<0.001	0.59	0.38 0.92	0.019
Άγνωστη	0.35	0.19 0.67	0.002	0.78	0.58 1.04	0.089
Ομάδα Κινδύνου (*MSW)						
PWID	0.78	0.38 1.62	0.504	1.14	0.69 1.90	0.611
MSM	7.20	3.36 15.30	<0.001	1.12	0.78 1.62	0.525
Άλλη	2.63	0.23 23.63	0.388	0.20	0.11 0.39	<0.001
Άγνωστη	1.11	0.57 2.15	0.763	1.06	0.62 1.80	0.835
Περίοδος δειγματοληψίας (*[1996 – 2001])						
[2001 – 2006)	1.61	0.47 5.47	0.445	2.09	1.56 2.80	<0.001
[2006 – 2011)	1.39	0.42 4.54	0.588	4.09	2.94 5.69	<0.001
[2011 – 2015)	2.25	0.69 7.37	0.181	7.21	5.10 10.19	<0.001
Ηλικία	0.99	0.97 1.01	0.586	0.95	0.94 0.96	<0.001

Κεφάλαιο 4: Συμπεράσματα

Η επιδημία του υπότυπου A1 παρουσιάζει υψηλότερα επίπεδα τοπικής διασποράς σε σχέση με του B. Και στους δύο υπότυπους βρέθηκε υψηλό ποσοστό μονοφυλετικότητας (A1: 91,46% - B: 72,82%), όμως στον υπότυπο A1 το ποσοστό αυτό αφορά πολύ μικρότερο αριθμό δικτύων συγκριτικά με τον B.

Οι οροθετικοί με A1 για τους οποίους δεν παρατηρήθηκε τοπική διασπορά είναι κυρίως ετεροφυλόφιλοι μη-Ελληνικής εθνικότητας, γεγονός που υποδεικνύει ότι αυτή η ομάδα πληθυσμού που έχει μολυνθεί με στελέχη A1 με γεωγραφική προέλευση εκτός Ελλάδας, δεν σχετίζεται με περαιτέρω διασπορά και συνεπώς δημιουργία τοπικών επιδημιών.

Επίσης παρατηρήσαμε μία έντονη αντίθεση μεταξύ των δύο υπότυπων καθώς στον A1 εντός δικτύων είναι άτομα ελληνικής Εθνικότητας και MSM ενώ εκτός δικτύων είναι μη-Έλληνες και MSW, σε αντίθεση με τον B που ανεξάρτητα από το αν είναι εντός ή εκτός δικτύων είναι Έλληνες και MSM.

Ο μεγαλύτερος αριθμός δικτύων για τον υπότυπο B σημαίνει πολλαπλές πηγές εισαγωγής στην Ελλάδα. Αυτό οφείλεται πιθανόν στο ότι ο υπότυπος B είναι πολύ συχνός στο Δυτικό κόσμο και του ότι οι ανθρώπινες δραστηριότητες που σχετίζονται με μετάδοση HIV (σεξουαλικές επαφές χωρίς προφύλαξη, μεταγγίσεις, ανθρώπινες μετακινήσεις) είναι συχνές στο Δυτικό κόσμο. Αντίθετα αυτό δεν ισχύει για τον A1 που ο υψηλότερος επιπολασμός του εντοπίζεται στην Αφρική και πρόσφατα στην Ανατολική Ευρώπη,

Για τον B παρατηρείται επίσης μια διαχρονική τάση στην περίοδο δειγματοληψίας, δηλαδή με την πάροδο του χρόνου τα άτομα τα οποία μολύνονται με στελέχη υπότυπου B έχουν μεγαλύτερη πιθανότητα να έχουν μολυνθεί εντός δικτύων.

Αυτή η διαχρονική τάση μπορεί να οφείλεται, εν μέρει, και στο γεγονός ότι για τις αλληλουχίες με προγενέστερη δειγματοληψία η πιθανότητα να βρεθεί ο πιθανή πηγή τους είναι μικρότερη (απουσιάζουν από το δείγμα).

Κεφάλαιο 5: Συζήτηση

Η παρούσα μελέτη πραγματοποιήθηκε στο μεγαλύτερο δυνατό δείγμα μειώνοντας σημαντικά το σφάλμα δειγματοληψίας (selection bias). Συγκεκριμένα αναλύθηκαν όλες οι διαθέσιμες ελληνικές αλληλουχίες των υπότυπων A1 και B με δειγματοληψία από το 1996 έως τον Ιούνιο του 2015, με σκοπό την εύρεση όλων των τοπικών επιδημιών. Η συγκεκριμένη μελέτη συμπληρώνει ένα σημαντικό κενό στη βιβλιογραφία δεδομένου ότι, από όσο γνωρίζουμε, παρόμοια μελέτη σε τόσο μεγάλο δείγμα δεδομένων από Ελλάδα δεν έχει πραγματοποιηθεί μέχρι σήμερα. Επίσης η ανάλυση σε μεγάλο δείγμα βελτιώνει την πιθανότητα να εκτιμηθεί αξιόπιστα το ποσοστό των αλληλουχιών σε δίκτυα και συνεπώς τα αντίστοιχα ποσοστά που αφορούν τοπική διασπορά της επιδημίας.

Η μεθοδολογία που ακολουθήθηκε στην επιλογή των αλληλουχιών αναφοράς για τον εντοπισμό των δικτύων μετάδοσης του ιού είναι καινοτόμα και συμβάλλει σημαντικά στην αύξηση της αξιοπιστίας των ευρημάτων μας. Μέχρι σήμερα δεν υπάρχει μια μοναδική βέλτιστη μέθοδος (gold standard) για τον τρόπο επιλογής των αλληλουχιών αναφοράς. Αναλύοντας τα δεδομένα μας χρησιμοποιώντας 5 διαφορετικούς συνδυασμούς τυχαία επιλεγμένων αλληλουχιών αναφοράς διασφάλισαμε ότι η σύνθεση των δικτύων που εντοπίσαμε δεν έχει επηρεαστεί από τις αλληλουχίες αναφοράς που χρησιμοποιήσαμε. Επίσης η επιλογή των αλληλουχιών που βρέθηκαν σε δίκτυα και στις 5 διαφορετικές δειγματοληψίες βελτιώνει σημαντικά την αξιοπιστία της ανάλυσης μας.

Η ταυτοποίηση των τοπικών δικτύων και το ποσοστού μεταδόσεων σε τοπικό επίπεδο, έχει μεγάλη σημασία για τη διερεύνηση της επιδημίας HIV και της πληρέστερης κατανόησης των χαρακτηριστικών της. Τα χαρακτηριστικά αυτά έχουν ιδιαίτερη σημασία για την αποτελεσματική αντιμετώπιση και την πρόληψη της επιδημίας. Συγκεκριμένα στην Ελλάδα, όπως και σε άλλες περιοχές της Ευρώπης, παρατηρείται μεγάλο ποσοστό καθυστερημένης διάγνωσης HIV (επίπεδα CD4-λεμφοκυττάρων < 350 c/μl) ή σε προχωρημένα στάδια της νόσου της νόσου (επίπεδα CD4-λεμφοκυττάρων < 200 c/μl). Αυτό το δεδομένο έχει σημαντικές συνέπειες τόσο στη δυνατότητα αποτελεσματικής θεραπείας όσο και στη δυνατότητα πρόληψης νέων μεταδόσεων. Από προηγούμενες μελέτες είναι γνωστό ότι κατά το διάστημα που κάποιος ζει με την HIV-λοίμωξη και δεν έχει διαγνωσθεί, μπορεί να επιδεικνύει υψηλή συμπεριφορά κινδύνου, σε αντίθεση με κάποιον που γνωρίζει ότι έχει μολυνθεί με HIV. Επίσης κατά το διάστημα αυτό και όσο δεν λαμβάνει αντιρετροϊκή θεραπεία μπορεί να συμβούν νέες μεταδόσεις σε άλλα άτομα, λόγω του ότι το HIV-RNA παραμένει ανιχνεύσιμο στα βιολογικά υγρά. Η καθυστερημένη διάγνωση παραμένει ένα από τα καίρια προβλήματα για την αντιμετώπιση και πρόληψη της HIV-λοίμωξης παγκοσμίως. Η γνώση των επιμέρους

χαρακτηριστικών της επιδημίας, όπως το ποσοστό τοπικών μεταδόσεων ανά ομάδα πληθυσμού, αποτελεί μια σημαντική παράμετρο για το αν η καθυστερημένη διάγνωση μπορεί να αντιμετωπιστεί και σε τι ποσοστό για κάθε ομάδα κινδύνου.

Επίσης η λεπτομερειακή γνώση των χαρακτηριστικών της επιδημίας, όπως η γνώση των δικτύων, μπορεί να παρέχει πληροφορίες για τα χαρακτηριστικά του πληθυσμού που σχετίζονται με δίκτυα. Αν αυτή η πληροφορία συνδυαστεί και με εκτίμηση της ημερομηνίας μετάδοσης τότε καθίσταται εξαιρετικά σημαντική για την πρόληψη HIV σε άτομα που ανήκουν σε ενεργά δίκτυα διασποράς.

Περίληψη

Εισαγωγή: Προηγούμενες μελέτες έχουν δείξει ότι οι A1 και B είναι οι επικρατέστεροι HIV-1 υπότυποι στην Ελλάδα, και παρουσιάζουν το μεγαλύτερο ποσοστό τοπικής διασποράς σε σχέση με άλλους υπότυπους και ανασυνδυασμένους τύπους του ιού.

Σκοπός: Η διερεύνηση του τρόπου διασποράς των υπότυπων A1 και B στην Ελλάδα, η εύρεση όλων των τοπικών δικτύων μετάδοσης τους (επιμέρους επιδημίες), και η εκτίμηση των παραγόντων που σχετίζονται με την τοπική διασπορά του ιού, χρησιμοποιώντας μεθόδους μοριακής επιδημιολογίας.

Υλικό: Μελετήθηκαν 1.230 αλληλουχίες υπότυπου A1 και 2.156 υπότυπου B με δειγματοληψία το διάστημα 1996-06/2015 στη Νότια Ελλάδα.

Μέθοδοι: Τα φυλογενετικά δέντρα εκτιμήθηκαν με τη μέθοδο μέγιστης πιθανοφάνειας και ανάλυση bootstrap. Η ανάλυση επαναλήφθηκε 5 φορές για κάθε υπότυπο χρησιμοποιώντας κάθε φορά ένα τυχαία επιλεγμένο δείγμα αλληλουχιών αναφοράς με παγκόσμια δειγματοληψία μεγέθους 1.500 αλληλουχιών για τον A1 και 2.000 για τον B, αντίστοιχα. Η τοπική διασπορά του ιού (ποσοστό μονοφυλετικότητας) εκτιμήθηκε υπολογίζοντας το πηλίκο του αριθμού των αλληλουχιών σε φυλογενετικές ομάδες με δειγματοληψία >70% από Ελλάδα και του συνόλου των διαθέσιμων αλληλουχιών από Ελλάδα, ανά υπότυπο. Η στατιστική ανάλυση πραγματοποιήθηκε με τη μέθοδο της λογιστικής παλινδρόμησης.

Αποτελέσματα: Για τον υπότυπο A1 βρέθηκαν 23 τοπικά δίκτυα μετάδοσης με εύρος από 2 έως 1.059 αλληλουχίες, ενώ για τον B βρέθηκαν 133 τοπικά δίκτυα με εύρος από 2 έως 157 αλληλουχίες. Τα ποσοστά μονοφυλετικότητας των A1 και B ήταν 91,46% (N=1.125) και 72,82% (N=1.570), αντίστοιχα. Το 86,1% (N=1.059) των A1 αλληλουχιών σχημάτισε μια μονοφυλετική ομάδα (τοπικό δίκτυο μετάδοσης). Οι HIV(+) με A1 των οποίων οι αλληλουχίες εντοπίστηκαν εκτός δικτύων μετάδοσης (N=105) ήταν, κυρίως, ετεροφυλόφιλοι (N=54; 51,4%), μη-Ελληνικής εθνικότητας (N=45; 42,9%). Η ανάλυση λογιστικής παλινδρόμησης έδειξε ότι η τοπική διασπορά του υπότυπου A1 σχετίζεται με άνδρες (OR=1,8; p=0.034), MSM (OR=7,2; p<0.001), Ελληνικής εθνικότητας (OR=7,2; p<0.001), ενώ του B με δειγματοληψία το διάστημα 2011-2015 (OR=7,21; p<0.001).

Συμπεράσματα: Η επιδημία του υπότυπου A1 παρουσιάζει υψηλότερα επίπεδα τοπικής διασποράς σε σχέση με του B. Οι HIV(+) με A1 για τους οποίους δεν παρατηρήθηκε τοπική διασπορά είναι κυρίως ετεροφυλόφιλοι μη-Ελληνικής εθνικότητας, γεγονός που υποδεικνύει ότι αυτή η ομάδα πληθυσμού που έχει μολυνθεί με στελέχη A1 με γεωγραφική προέλευση εκτός Ελλάδας, δεν σχετίζεται με περαιτέρω διασπορά και συνεπώς δημιουργία τοπικών επιδημιών.

Abstract

Introduction: Previous studies have shown that HIV-1 subtypes A1 and B are the predominant clades in Greece.

Aim: To investigate the patterns of subtypes A1 and B dispersal in Greece, to identify the proportion of infections occurred locally, and to estimate the factors that are associated with local transmissions, using molecular epidemiology methods.

Material: We studied 1,230 A1 and 2,156 B sequences isolated from HIV-1 diagnosed patients during 1996-06/2015 in Greece. Sequences were available in the protease (PR) and partial reverse transcriptase (RT) regions.

Methods: Maximum-likelihood (ML) phylogeny reconstruction with bootstrap evaluation was conducted in RAxML8, using GTR+G as nucleotide substitution model. Phylogenetic analysis was performed separately on the 1,230 A1 and 2,156 B sequences from Greece along with a random set of globally sampled sequences (A1: N=1,500; B: N=2,000), used as references. Phylogenetic analysis was performed for each subtype in five replicates, using a different reference dataset each time. Local transmission networks (LTNs) were defined as phylogenetic clusters including sequences from Greece at proportions >70% (geographic criterion), found in all five replicates (phylogenetic confidence). Local dispersal was estimated by dividing the number of sequences found within LTNs with the total number of sequences sampled in Greece for each subtype. The statistical analysis was carried out using multivariate logistic regression models. Presence in LTNs was the binary outcome variable, while age, period of sampling, risk group, nationality and gender were chosen as possible explanatory variables.

Results: Phylogenetic analysis revealed that 91.5% (1,125 out of 1,230) of the A1 sequences sampled in Greece formed 23 LTNs. The size of the LTNs ranged between 2 and 1,059 sequences. The 86.1% of the A1 sequences fell within a single LTN consisted of 1,059 sequences. The rest of the A1 sequences (N=105, 8.5%) clustered at different points in the ML tree. The majority of the unclustered sequences had been isolated mainly from heterosexuals with non-Greek nationality. Phylogenetic analysis also revealed that 72.8% (1,570 out of 2,156) of the B sequences sampled in Greece formed 133 LTNs, with a range of 2 to 157 sequences. Multivariate logistic regression analysis showed that gender (male; OR=1.8; p=0.034), risk group (MSM; OR=7.2; p<0.001) and nationality (Greek; OR=7.2; p<0.001) were associated with regional clustering of subtype A1, while period of sampling (2011-2015; OR=7.21; p<0.001) with local transmissions of subtype B.

Conclusions: Our analysis suggests considerable differences in the levels of regional clustering of HIV-1 subtype A1 and B in Greece. For subtype A1 we found a lower number of introductions than for subtype B. This probably due to that subtype B is more frequently found in developed countries, suggesting that cross-border transmissions are more likely than non-B clades. It was also shown that the majority of A1 unclustered sequences had been isolated mainly from heterosexuals with non-Greek nationality. This finding suggests that this population group was infected with A1 strains with non-Greek origin and it is not related to further dispersal. Moreover, we found that local dispersal of subtype B transmissions is mostly associated with recent diagnosis.

Βιβλιογραφία

1. Kayser FH, Bienz KA, Eckert J, R.M. Zinkernagel. Medical Microbiology. Thieme. 2005.
2. Tortora GJ, Funke BR, Case CL. Microbiology: An introduction. 11th ed. 2013. 975 p.
3. www.avert.org.
4. Sepkowitz Kent A. AIDS — The First 20 Years. N Engl J Med. 2001;344(23).
5. Michael S. Gottlieb. Pneumocystis Pneumonia - Los Angeles. MMWR Wkly [Internet]. 1981;30(21):1–3. Available from: http://www.cdc.gov/mmwr/preview/mmwrhtml/june_5.htm
6. Friedman-Kien et al. Disseminated Kaposi ' s Sarcoma in Homosexual Men. 1982;96(6):693–700.
7. Masur H, Michelis A, Greene JB, Onorato I, Stouwe RA V, Holzman RS, et al. An outbreak of community- acquired Pneumocystis carinii pneumonia. New Engl. 1981;305(24):1431–7.
8. Centers for Disease Control (CDC) (1982). «Update on acquired immune deficiency syndrome (AIDS)—United States». MMWR Morb Mortal Wkly Rep. 31 (37): 507–508; 513–514. PMID 6815471.
9. Gallo C. Robert et al. Isolation of Human T-Cell Leukemia Virus in Acquired Immune Deficiency SyndroColombme (AIDS). Science (80-). 1983;220:865–7.
10. Barre-Sinoussi F, Chermann J, Rey F, Nugeyre M, Chamaret S, Gruest J, et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science (80-) [Internet]. 1983;220(4599):868–71. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/6189183>
11. Levy A. Jay et al. Isolation of Lymphocytopathic Retroviruses from San Francisco Patients with AIDS. Science (80-). 1984;225.
12. Basavapathruni A, Anderson KS. Reverse transcription of the HIV-1 pandemic. FASEB J [Internet]. 2007;21(14):3795–808. Available from: <http://www.fasebj.org/cgi/doi/10.1096/fj.07-8697rev>
13. The New York Times [Internet]. Available from: <http://www.nytimes.com/2008/10/07/health/07nobel.html?hp>
14. International Committee on Taxonomy of Viruses (ICTV) [Internet]. Available from: <https://talk.ictvonline.org/taxonomy/>
15. Tebit DM, Arts EJ. Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. Lancet Infect Dis [Internet]. 2011;11(1):45–56. Available from: [http://dx.doi.org/10.1016/S1473-3099\(10\)70186-9](http://dx.doi.org/10.1016/S1473-3099(10)70186-9)
16. Paraskevis D, Hatzakis A. Molecular Epidemiology of HIV-1 Infection. 1999;238–49.
17. HIV Database [Internet]. Available from: <https://www.hiv.lanl.gov/content/sequence/HelpDocs/subtypes-more.html>

18. Seitz R. Human Immunodeficiency Virus (HIV). *Transfus Med Hemotherapy*. 2016;43(3):203–22.
19. Turner BG, Summers MF. Structural biology of HIV. *J Mol Biol*. 1999;285(1):1–32.
20. i-base [Internet]. Available from: <http://i-base.info/qa/factsheets/hiv-genome-explained>
21. King SR. HIV: Virology and mechanisms of disease. *Ann Emerg Med*. 1994;24(3):443–9.
22. Mushahwar IK. Human Immunodeficiency Viruses: Molecular Virology, Pathogenesis, Diagnosis and Treatment. *Perspect Med Virol*. 2006;13(6):75–87.
23. International Committee on Taxonomy of Viruses (ICTV) [Internet]. Available from: <https://talk.ictvonline.org/>
24. World Health Organization:WHO [Internet]. Available from: <http://www.who.int>
25. UNAIDS [Internet]. Available from: www.unaids.org
26. Takebe Y, Uenishi R, Li X. Global Molecular Epidemiology of HIV: Understanding the Genesis of AIDS Pandemic. *Adv Pharmacol*. 2008;56(7):1–25.
27. Hemelaar J. Implications of HIV diversity for the HIV-1 pandemic. *J Infect* [Internet]. 2013;66(5):391–400. Available from: <http://dx.doi.org/10.1016/j.jmolmed.2011.12.001>
28. ΚΕΕΛΠΝΟ. HIV λοίμωξη: νεότερα επιδημιολογικά δεδομένα, Οκτώβριος 2017 [Internet]. 2017. Available from: <http://www.keelpno.gr/>
29. Επιδημιολογική Επιτήρηση της HIV/AIDS λοίμωξης στην Ελλάδα Δηλωθέντα Στοιχεία έως 31.12.2016. Επιδημιολογική Επιτήρηση της HIV/AIDS λοίμωξης στην Ελλάδα [Internet]. 2016;(31). Available from: <http://www.keelpno.gr/>
30. Paraskevis D, Magiorkinis E, Magiorkinis G, Sypsa V, Papanizos V, Lazanas M, et al. Increasing prevalence of HIV-1 subtype A in Greece: estimating epidemic history and origin. *J Infect Dis*. 2007;196(8):1167–76.
31. Beloukas A, Psarris A, Giannelou P, Kostaki E, Hatzakis A, Paraskevis D. Molecular epidemiology of HIV-1 infection in Europe: An overview. *Infect Genet Evol* [Internet]. 2016;46:180–9. Available from: <http://dx.doi.org/10.1016/j.meegid.2016.06.033>
32. Kostrikis LG, Bagdades E, Cao Y, Zhang L, Dimitriou D, Ho DD. Genetic analysis of human immunodeficiency virus type 1 strains from patients in Cyprus: identification of a new subtype designated subtype I. *J Virol* [Internet]. 1995;69(10):6122–30. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=189509&tool=pmcentrez&rendertype=abstract>
33. Paraskevis D, Magiorkinis E, Magiorkinis G, Anastassopoulou C, Lazanas M, Chrysos G, et al. Molecular characterization of a complex, recombinant human immunodeficiency virus type 1 (HIV-1) isolate (A/G/J/K/?): evidence to support the existence of a novel HIV-1 subtype. *J Gen Virol*. 2001;82(Pt 10):2509–14.
34. Paraskevis D, Magiorkinis E, Magiorkinis G, Kiosses VG, Lemey P, Vandamme AM, et

- al. Phylogenetic reconstruction of a known HIV-1 CRF04_cpx transmission network using maximum likelihood and Bayesian methods. *J Mol Evol.* 2004;59(5):709–17.
35. ΚΕΕΛΠΝΟ [Internet]. Available from: <http://www.keelpno.gr/>
 36. Lam TTY, Hon CC, Tang JW. Use of phylogenetics in the molecular epidemiology and evolutionary studies of viral infections. Vol. 47, *Critical Reviews in Clinical Laboratory Sciences.* 2010. 5-49 p.
 37. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 2011;28(10):2731–9.
 38. Edgar RC. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics.* 2004;5:1–19.
 39. JUKES, T.H., CANTOR C. Evolution of protein molecules. In: *Mammalian Protein Metabolism.* New York Acad Press. 1969;21–132.
 40. Felsenstein J. Evolutionary Trees from DNA Sequences : A Maximum Likelihood Approach. *J Mol Evol.* 1981;17:368–76.
 41. Kimura M. A Simple Method for Estimating Evolutionary Rates of Base Substitutions Through Comparative Studies of Nucleotide Sequences. *J Mol Evol.* 1980;16(1330):111–20.
 42. Hasegawa M et al. Dating of the Human-Ape Splitting by a Molecular Clock of Mitochondrial DNA. *J Mol Evol.* 1985;22:160–74.
 43. Tamura K, Nei M. Estimation of the Number of Nucleotide Substitutions in the Control Region of Mitochondrial DNA in Humans and Chimpanzees. *Mol Biol Evol.* 1993;10(3):512–25.
 44. Tavaré Simon. Some probabilistic and statistical problems in the analysis of DNA sequence. *Lect Math Life Sci (American Math Soc.* 1986;17:57–86.
 45. Yang Z. of Molecular Evolution Maximum Likelihood Phylogenetic Estimation from DNA Sequences with Variable Rates over Sites : Approximate Methods. *J Mol Evol.* 1994;39:306–14.
 46. Sullivan J, Joyce P. Model Selection in Phylogenetics. *Annu Rev Ecol Evol Syst.* 2005;36:445–66.
 47. Project HG, Durbin R, Division I, Centre S, Eddy S, Investigators P, et al. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids.* 1998.
 48. Sokal, R. R., & Michener CD. A Statistical Method for Evaluating Systematic Relationships. *Univ Kansas Sci Bull.* 1958;38:1409–38.
 49. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 1987;406–25.
 50. Eck R V, Dayhoff MO. Evolution of the Structure of Ferredoxin Based on Living Relics of Primitive Amino Acid Sequences. *Science.* 1966;152:363–6.
 51. Gilks WR, Richardson S, Spiegelhalter D. *Markov Chain Monte Carlo in Practice*

- Chapman & Hall/CRC. 1996.
52. Metropolis N, Ulam S. The Monte Carlo Method. *J Am Stat Assoc* [Internet]. 1949;44(247):335–41. Available from: <http://www.jstor.org/stable/2280232>%5Cr
 53. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of State Calculations by Fast Computing Machines. *J Chem Phys* [Internet]. 1953;21(6):1087–92. Available from: <http://aip.scitation.org/doi/10.1063/1.1699114>
 54. Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*. 1970;57(1):97.
 55. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30(9):1312–3.
 56. Felsenstein J. Confidence Limits on Phylogenies : An Approach Using the Bootstrap Author. *Evolution (N Y)*. 1985;39(4):783–91.
 57. FigTree [Internet]. Available from: <http://tree.bio.ed.ac.uk/software/figtree/>
 58. Microsoft Office Excel 2007 [Internet]. Available from: <https://products.office.com/en-us/microsoft-excel-2007>
 59. Avise JC. *Phylogeography - The History and Formation of Species*. Harvard University Press; 2000.
 60. Mesquite [Internet]. Available from: [www.mesquiteproject.wikispaces.com](http://www.mesquiteproject.org)
 61. STATA [Internet]. Available from: <https://www.stata.com/>