ΠΜΣ ΒΙΟΣΤΑΤΙΣΤΙΚΗΣ

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΙΑΤΡΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΒΟΡΡΙΑ ΑΙΚΑΤΕΡΙΝΗ

Εφαρμογή Μεθόδων Μοριακής Επιδημιολογίας στην Πρόληψη του HIV

ΑΘΗΝΑ, ΙΟΥΝΙΟΣ 2020

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

……

……

……

Περιεχόμενα

ΕΙΣΑΓΩΓΗ

Ο HIV είναι υπεύθυνος για μία από τις μεγαλύτερες ιικές πανδημίες στην ανθρώπινη ιστορία. Παρά τη συνολική ανταπόκριση σε παγκόσμιο επίπεδο για την πρόληψη και τη θεραπεία, συμπεριλαμβανομένης της προφύλαξης προ της έκθεσης από το στόμα, της έγκαιρης έναρξης της αντιρετροϊκής θεραπείας ως δευτερογενούς πρόληψης και των κοινωνικο-συμπεριφορικών παρεμβάσεων, ο ιός επιμένει. Οι λοιμώδεις λοιμώξεις από τον HIV παραμένουν υψηλές ειδικά στους άνδρες που έχουν σεξουαλική επαφή με άνδρες από χώρες μεσαίων και υψηλών πόρων, όπως στις ΗΠΑ όπου οι υψηλές επιπτώσεις των επιδημιών από τον ιό HIV στους άνδρες που έχουν σεξουαλική επαφή με άνδρες συνεχίζονται, με έντονη βαρύτητα μεταξύ των νέων και φυλετικών και εθνοτικών μειονοτήτων, και σε χώρες τόσο διαφορετικές όπως η Γαλλία, το Ηνωμένο Βασίλειο, η Ταϊλάνδη, η Κίνα, η Κένυα και η Ρωσία. Συνεπώς, απαιτείται επείγουσα δράση για τη δημόσια υγεία, χρησιμοποιώντας νέες παρεμβάσεις, προκειμένου να αποφευχθούν μελλοντικά συμβάντα μετάδοσης, κρίσιμα για την εξάλειψη του HIV. (German & Grabowski & Beyrer, 2017; Paraskevis et al, 2016) Ένα από τα καθοριστικά χαρακτηριστικά του HIV είναι η ικανότητά του να εξελίσσεται ταχέως και να παραμένει μέσα στα άτομα παρά την συνεχιζόμενη πίεση από τις κυτταρικές και χημικές ανοσοαποκρίσεις του ξενιστή. Αυτός ο αγώνας μεταξύ του ιού και του ανθρώπου έχει οδηγήσει σε μία από τις πιο γενετικά ποικίλες πανδημίες στην καταγεγραμμένη ιστορία. Η ποικιλομορφία του ιού HIV ήταν αναμφίβολα ζωτικής σημασίας για την ανθεκτικότητα και τη διάδοσή του σε όλο τον κόσμο. Ωστόσο, οι πρόσφατες εξελίξεις στις τεχνολογίες γενετικής ακολουθίας, στις υπολογιστικές μεθοδολογίες και στα στατιστικά στοιχεία παρέχουν στους ερευνητές νέα εργαλεία για να αξιοποιήσουν την ιική ποικιλομορφία για να καταπολεμήσουν την παγκόσμια πανδημία του ιού HIV. Καθοδηγούμενοι από την πληθυσμιακή γενετική και τις επιδημιολογικές αρχές, οι επιστήμονες χρησιμοποιούν ιογενή φυλογενετική για τη βελτίωση της κατανόησης της διαφορετικότητας του HIV μέσα σε άτομα και πληθυσμούς, δημιουργώντας μια άνευ προηγουμένου γνώση της ιογενούς δυναμικής για τη βελτίωση των στρατηγικών πρόληψης του HIV και τη θεραπεία των HIV-μολυσμένων ατόμων. Η μοριακή επιδημιολογική αξιολόγηση των δικτύων μετάδοσης του HIV μπορεί να διασαφηνίσει τα στοιχεία συμπεριφοράς της μετάδοσης που μπορούν να αποτελέσουν στόχους για την παρέμβαση.(Chan, 2015) Λόγω της ταχείας εξέλιξης του HIV, οι μολύνσεις με παρόμοιες γενετικές αλληλουχίες είναι πιθανό να σχετίζονται με πρόσφατα συμβάντα μετάδοσης. Συστάδες συσχετισμένων λοιμώξεων μπορεί να αντιπροσωπεύουν υποπληθυσμούς με υψηλά ποσοστά μετάδοσης του HIV. Συνεπώς είναι χρήσιμη η εφαρμογή ενός αυτοματοποιημένου συστήματος "κοντά σε πραγματικό χρόνο" χρησιμοποιώντας ανάλυση συστάδων συλλεγόμενων HIV ανθεκτικών γονότυπων ρουτίνας για την παρακολούθηση και το χαρακτηρισμό των σημείων (hotspots) μετάδοσης του ιού HIV. Ο χαρακτηρισμός των δικτύων μετάδοσης του HIV μπορεί να είναι σημαντικός για την κατανόηση της εξέλιξης των προτύπων και για τη γεωπεριβαλλοντική εξάπλωση της επιδημίας. Αρκετές μελέτες έχουν χρησιμοποιήσει τη φυλογενετική συσταδοποίηση για μία ανάλυση HIV pol συνόλων δεδομένων για να χαρακτηρίσουν αναδρομικά πιθανές συσχετίσεις υψηλών ρυθμών μετάδοσης του HIV, όπως ιστορικό θεραπείας, στάδιο μόλυνσης και διαδρομές μετάδοσης. (Poon et al., 2016) Οι πρόσφατες εξελίξεις στη μοριακή επιδημιολογία έχουν σημαντικά ενισχύσει την ικανότητά μας να χαρακτηρίζουμε τη δυναμική και τη δομή των δικτύων μετάδοσης του ιού HIV σε χώρο και χρόνο χρησιμοποιώντας pol αλληλουχίες HIV που παράγονται για ρουτίνα. (Chaillon et al, 2017) Προκειμένου ο σχεδιασμός της δημόσιας υγείας να αποδειχθεί αποτελεσματικός και επιτυχημένος, πρέπει να κατανοήσουμε τη δυναμική των περιφερειακών επιδημιών και να παρέμβουμε κατάλληλα. Τα εργαλεία μοριακής επιδημιολογίας του HIV, όπως εφαρμόζονται σε φυλογενετικές, φυλοδυναμικές

και φυλογεωγραφικές αναλύσεις, έχουν αποδειχθεί ισχυρά εργαλεία στον σχεδιασμό της δημόσιας υγείας σε πολλές μελέτες. (Paraskevis et al, 2016)

Η Φυλογενετική χρησιμοποιείται συχνά για μελέτες με βάση τον πληθυσμό μετάδοσης του του ιού HIV. Μελέτες των φυλογενιών του ιού HIV μπορούν να παρέχουν κρίσιμες πληροφορίες σχετικά με τις HIV επιδημίες του, όπως η μετάδοση του ανθεκτικού σε φάρμακα ιού, η ανάμειξη μεταξύ των δημογραφικών ομάδων και η ταχύτητα της εξάπλωσης του ιού στους πληθυσμούς, οι οποίες κατά τα άλλα είναι δύσκολο να αποκτηθούν μέσω του παραδοσιακού σχεδιασμού μελέτης. Η κατανόηση της δυναμικής μετάδοσης του HIV είναι σχετική τόσο στον έλεγχο όσο και στις στρατηγικές παρεμβάσεων της λοίμωξης HIV. Συνήθως, οι αλυσίδες μετάδοσης του HIV προσδιορίζονται με βάση την ομοιότητα αλληλουχίας που εκτιμάται είτε απευθείας από την ευθυγράμμιση της αλληλουχίας είτε με την εξαγωγή ενός φυλογενετικού δέντρου. Ωστόσο, τα πρόσφατα αποτελέσματα από εμπειρικές και θεωρητικές μελέτες των φυλογενιών του ιού HIV αμφισβητούν ορισμένες από τις βασικές παραδοχές και ερμηνείες από φυλογενετικές μελέτες. Τα πρόσφατα ευρήματα περιλαμβάνουν έλλειψη σημείων συμφόρησης στους άνδρες που έχουν σεξουαλική επαφή με άνδρες και είναι χρήστες ενδοφλέβιων ναρκωτικών, στοιχείων για την προτιμησιακή μετάδοση του ιού HIV σε ετεροφυλόφιλες επιδημίες και περιορισμένες ενδείξεις ότι οι τοπολογίες των δέντρων συσχετίζονται με τις υποκείμενες δομές δικτύου. Άλλες προκλήσεις περιλαμβάνουν την έλλειψη ενός τυποποιημένου ορισμού για ένα φυλογενετικό σύμπλεγμα μετάδοσης και μεροληπτική ή αραιή δειγματοληψία των δικτύων μετάδοσης του ιού HIV. Η Φυλογενετική είναι ένα σημαντικό εργαλείο για την έρευνα για τον ιό HIV και προσφέρει ευκαιρίες για κατανόηση των κρίσιμων πτυχών της επιδημίας του HIV, όμως όπως όλες οι επιδημιολογικές έρευνες, οι χρησιμοποιούμενες μέθοδοι και η ερμηνεία των αποτελεσμάτων των φυλογενετικών μελετών θα πρέπει να γίνονται προσεκτικά με προσεκτική εξέταση.

ΣΚΟΠΟΣ ΤΗΣ ΠΑΡΟΥΣΑΣ ΕΡΓΑΣΙΑΣ

Σκοπός της παρούσας εργασίας μου είναι η ανάλυση των φυλογενετικών μεθόδων ως εργαλείο για τη δημιουργία στρατηγικών παρέμβασης και πρόληψης του HIV.

ΛΕΞΕΙΣ-ΚΛΕΙΔΙΑ

HIV, φυλογενετική, δίκτυα, πρόληψη, δημόσια υγεία

ΚΕΦΑΛΑΙΟ 1

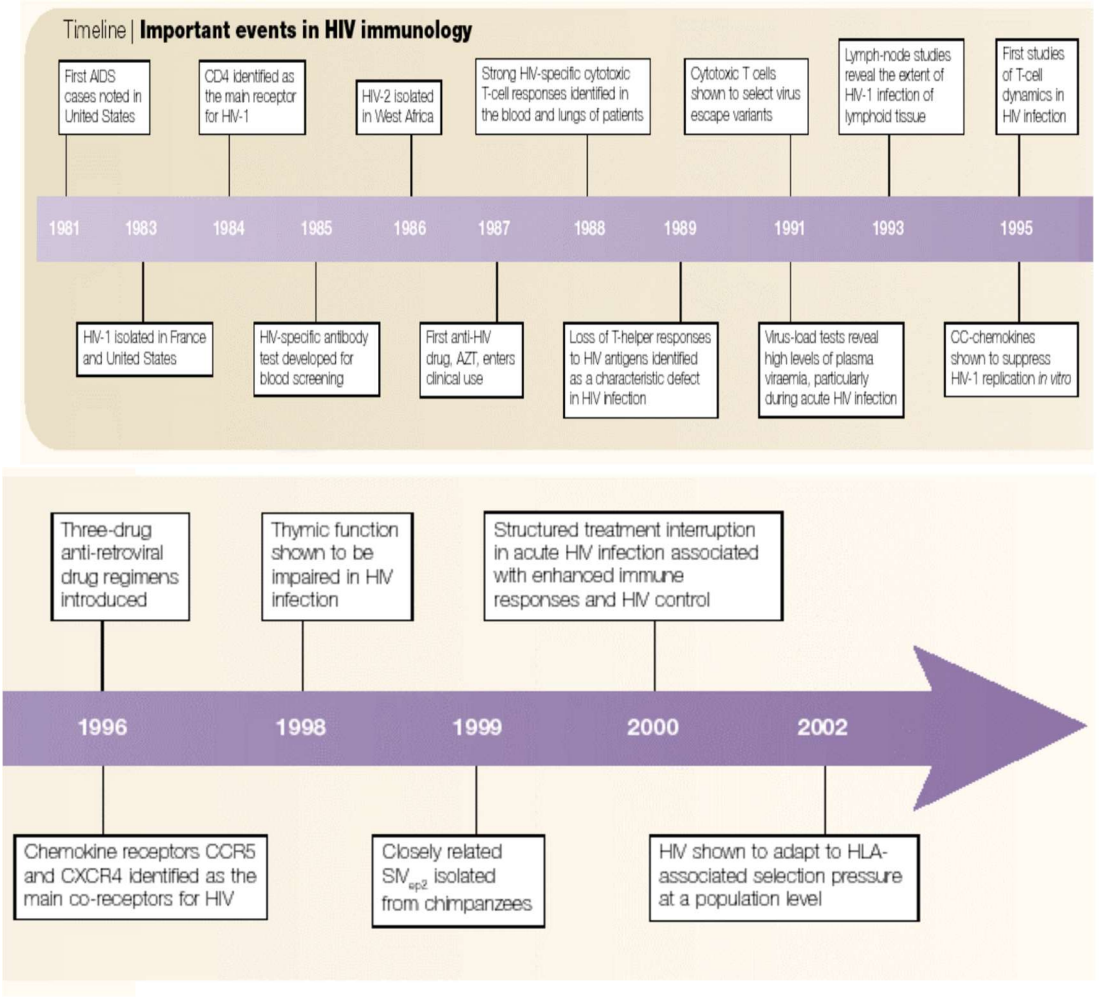
HIV

1.1 Η ανακάλυψη, η εξέλιξη και οι θεωρίες προέλευσης του HIV

Ο HIV εκτιμάται ότι γεννήθηκε στην Αφρική στις αρχές του 20ου αιώνα και πρόγονος του HIV θεωρείται ο Ιός της Ανοσοανεπάρκειας των Πιθηκοειδών (Simian Immunodeficiency Viruw, SIV) που ανήκει στο γένος των Lentiviruses. Πιο συγκεκριμένα, ο HIV-1 προέκυψε από τον SIV (cpz) που μολύνει χιμπατζήδες [Gao, F., et al.,1999], ενώ ο HIV-2 από τον SIV (smm) που μολύνει τον πίθηκο του είδους *Cercosebus alys alys* [Reeves, J.D. and R.W. Doms,2002]. Πιθανολογείται ότι ο SIV έχει περάσει από τον πίθηκο στον άνθρωπο σε πολλές διαφορετικές περιπτώσεις και οδήγησε σε πολλά αποκλίνοντα στελέχη [Tebit, D.M. and E.J. Arts,2011].

Τα πρώτα σημάδια της HIV επιδημίας ξεκίνησαν το 1981, όταν στις Ηνωμένες Πολιτείες Αμερικής έγινε επίσημη αναφορά ενός συνόλου νέων ομοφυλόφιλων ανδρών που παρουσίασαν συμπτώματα πνευμονίας από *Pneumocystis carinii* (PCP), που αποτελεί σπάνια και ευκαιριακή λοίμωξη. Σύντομα, εμφανίστηκαν αρκετές νέες περιπτώσεις ανθρώπων που ανέπτυξαν ασθένειες που προκαλούνται από ευκαιριακές λοιμώξεις, οι οποίες ήταν συχνά θανατηφόρες. Το 1982, το CDC ονόμασε την νέα αναδυόμενη νόσο AIDS (CDC, 1982) Παράλληλα με τους ομοφυλόφιλους, αναφέρθηκαν περιπτώσεις με σύνδρομο AIDS σε χρήστες ενδοφλέβιων ναρκωτικών, αιμορροφιλικούς και χρήστες που είχαν κάνει μετάγγιση αίματος και έτσι, οι ομάδες αυτές, αναγνωρίστηκαν ως υψηλού κινδύνου [Ryu, W.-S.,2017] .Λίγο αργότερα, το 1983, δύο ανεξάρτητες ερευνητικές ομάδες από τη Γαλλία και τις ΗΠΑ, με επικεφαλής τον L. Montagnier και τον R. Gallo αντίστοιχα, όπου δημοσίευσαν τα σχετικά ευρήματά τους στο ίδιο τεύχος του περιοδικού Science., κατάφεραν να απομονώσουν από το αίμα ασθενών ένα ρετροϊό, ο οποίος, στη συνέχεια, ονομάστηκε HIV και αναγνωρίστηκε ως ο αιτιολογικός παράγοντας του AIDS (Acheson N.H, 2011; Barre-Sinoussi et al., 1983, Gallo et al., 1983).Ο Gallo και οι συνεργάτες του υποστήριζαν ότι ο νέος αυτός ιός, που απομόνωσαν από ασθενή με AIDS, παρουσίαζε μορφολογική ομοιότητα με τους ιούς της ομάδας των ανθρώπινων Τ-λεμφοτρόπων ιών (Human T-Lymphotropic Virus, HTLVs) και τον ονόμασε HTLV-III. Ο Montagnier, ερχόμενος σε αντιπαράθεση με τον Gallo, τεκμηρίωσε ότι οι πυρηνικές πρωτεΐνες του ιού διέφεραν σημαντικά σε ανοσολογικό επίπεδο από τις αντίστοιχες των HTLVs και ονόμασε τον ιό LAV (lymphadenopathy-associated virus). Το 1986, απεδείχθη ότι οι δύο αυτοί ιοί ήταν οι ίδιοι και μετονομάστηκαν σε HIV. Τέλος, το 2008 ο Luc Montagnier τιμήθηκε με το βραβείο Nobel Ιατρικής για την ταυτοποίηση του HIV (Lever et al., 2008). Έκτοτε έχουν γίνει σημαντικές ανακαλύψεις που αφορούν στη διάγνωση και θεραπεία της συγκεκριμένης ιϊκής λοίμωξης. Στην Εικόνα 1 παρουσιάζονται διαχρονικά οι βασικότερες ανακαλύψεις στην ιστορία του HIV. Ο ιός HIV ανήκει ταξινομικά στο γένος Lentivirus της οικογένειας Retroviridae και έως σήμερα έχουν ταυτοποιηθεί δύο είδη του (HIV-1 και HIV-2), τα οποία εμφανίζουν ομοιότητα της τάξης του 40% στο γονιδίωμα τους (Clavel et al., 1986, Kanki et al., 1994).

Εικόνα 1: Ιστορική αναδρομή της ανακάλυψης του HIV. [Πηγή: Rowland- Jones, 2003]



Σήμερα, έχουν απομονωθεί και αναγνωριστεί πολλά στελέχη του HIV που ανήκουν σε διαφορετικές ομάδες και υπότυπους. Ο HIV, παρόλο που πιθανολογείται ότι είναι σχετικά πρόσφατος ιός, εμφανίζει μεγάλη ετερογένεια, η οποία οφείλεται στον υψηλό εξελικτικό του ρυθμό. Ο τελευταίος είναι αποτέλεσμα ορισμένων χαρακτηριστικών του κύκλου ζωής του και ιδιαίτερα της έλλειψης ενζυμικού επιδιορθωτικού μηχανισμού κατά το στάδιο της αντίστροφης μεταγραφής του γενετικού του υλικού, με αποτέλεσμα να προκύπτουν απόγονοι με συσσωρευμένες μεταλλάξεις στο γενετικό τους υλικό. Ο μεγάλος αριθμός μεταλλάξεων, σε συνδυασμό με τον μεγάλο αριθμό απογόνων που δημιουργούνται και τον μηχανισμό του γενετικού ανασυνδυασμού που λαμβάνει χώρα κατά τον αναπαραγωγικό κύκλο, αυξάνουν πολύ την πιθανότητα να δημιουργηθούν στελέχη με διαφορετικά χαρακτηριστικά από το προγονικό στέλεχος, με αποτέλεσμα ο ιός να εξελίσσεται με πολύ ταχύ ρυθμό [Tebit, D.M. and E.J. Arts, 2011].

Ο ιός της ανθρώπινης ανοσοανεπάρκειας (Human Immunodeficiency Virus, HIV) είναι ένας σύνθετος ρετροϊός που μολύνει και καταστρέφει τα ανθρώπινα CD4+ Τ λεμφοκύτταρα, τα οποία έχουν σημαντικό ρόλο στην ανοσοαπόκριση του οργανισμού. Ο HIV είναι υπεύθυνος για το Σύνδρομο της Επίκτητης Ανοσολογικής Ανεπάρκειας (Acquired Immune Deficiency Syndrome, AIDS), δηλαδή το καταληκτικό στάδιο της HIV λοίμωξης, κατά το οποίο το ανοσοποιητικό σύστημα αδυνατεί να αντιμετωπίσει ορισμένες μολύνσεις και καρκινογενέσεις και μπορεί να οδηγήσει σε θάνατο.

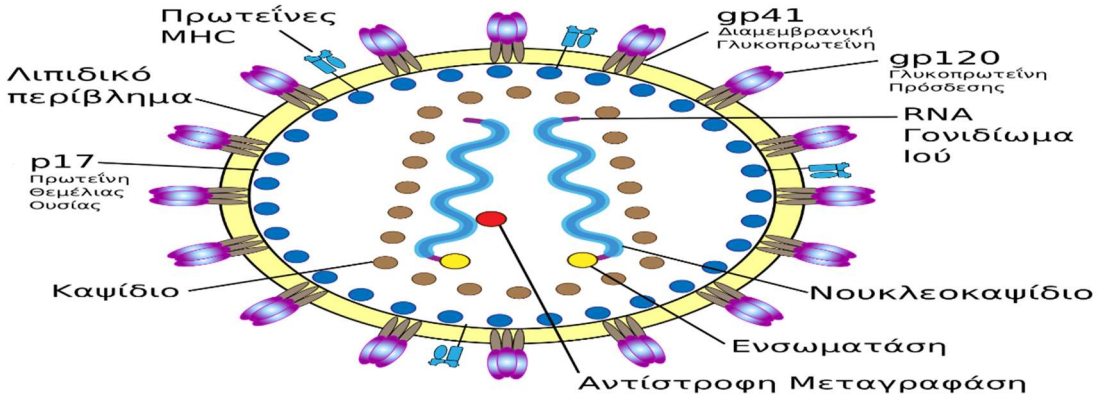
Ο HIV διακρίνεται σε δύο τύπους, τον HIV-1 και τον HIV-2, οι οποίο διαχωρίζονται σε επί μέρους ομάδες. Πρώτος ανακαλύφθηκε ο HIV-1, ο οποίος εμφανίζει μεγαλύτερο

επιπολασμό και μολυσματικότητα και είναι ο κύριος υπεύθυνος για την παγκόσμια HIV επιδημία, σε αντίθεση με τον HIV-2 που εντοπίζεται κατά κύριο λόγο σε περιοχές της Δυτικής Αφρικής [Ryu, W.-S, 2017; Tebit, D.M. and E.J. Arts, 2011].

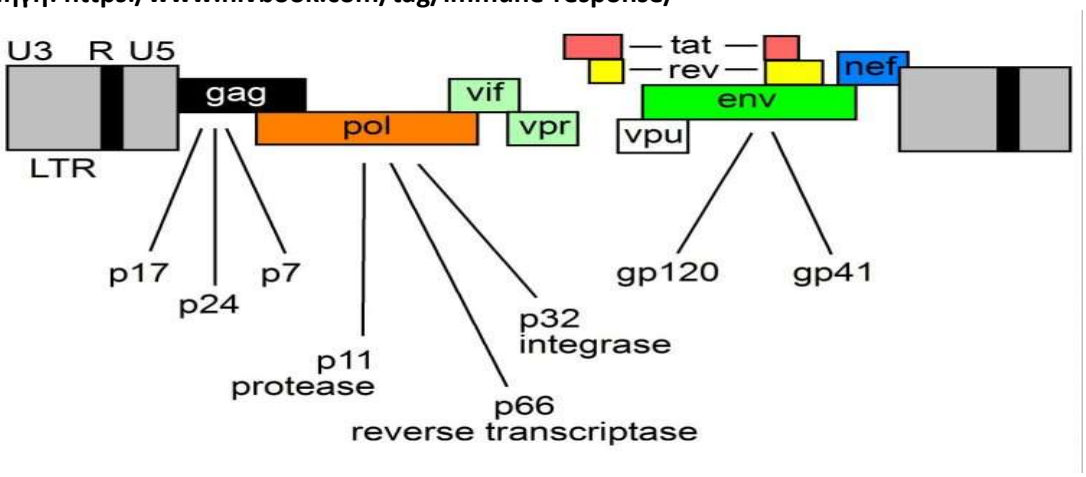
1.2. Γενετική οργάνωση του HIV

Ένα τυπικό στέλεχος του HIV έχει σφαιρικό σχήμα με διάμετρο 100-150 nm [Collier, L.H. and J.S. Oxford,2006]. Το γενετικό του υλικό αποτελείται από δύο μονόκλωνα και πανομοιότυπα αντίγραφα RNA θετικής πολικότητας, μήκους περίπου 9-10 kb [Collier, L.H. and J.S. Oxford,2006; Acheson, N.H., 2011; Ryu, W.-S, 2017]. Το γονιδίωμα του HIV-1 φέρει συνολικά εννέα γονίδια, εκ των οποίων τα δομικά γονίδια Gag και Env, και το γονίδιο Pol που βρίσκονται σε όλους τους ρετροϊούς, ενώ τα υπόλοιπα έξι (Tat, Rev, Vif, Vpr, Vpu και Nef) χαρακτηρίζουν αποκλειστικά τα στελέχη του HIV-1 και εντοπίζονται ανάμεσα στα γονίδια Pol και Env που παίζουν σημαντικό ρόλο στον κύκλο ζωής του ιού (Costin, 2007; Wang et al., 2000).Το γονίδιο Gag είναι υπεύθυνο για την παραγωγή δομικών πρωτεϊνών του κοψιδιού, ενώ το γονίδιο Env για την παραγωγή πρωτεϊνών που θα διαμορφώσουν τις γλυκοπρωτεΐνες του φακέλου. Το γονίδιο Pol κωδικοποιεί τα ένζυμα *πρωτεάση (protease), αντίστροφη μεταγραφάση (reverse transcription) και ιντεγκράση (integrase*).[Dimmock, N.J., A.J. Easton, and K.N. Leppard, 2016] Η δομή του γονιδιώματος του HIV-2 είναι σχεδόν πανομοιότυπη με αυτή του HIV-1, αλλά αντί του γονιδίου Vpu, φέρει το γονίδιο Vpx [Ryu, W.-S, 2017].

Εικόνα 2 Δομή Γονιδιώματος HIV-1 [Πηγή: Βικιπαιδεία: Ιός ανθρώπινης ανοσοανεπάρκειας]



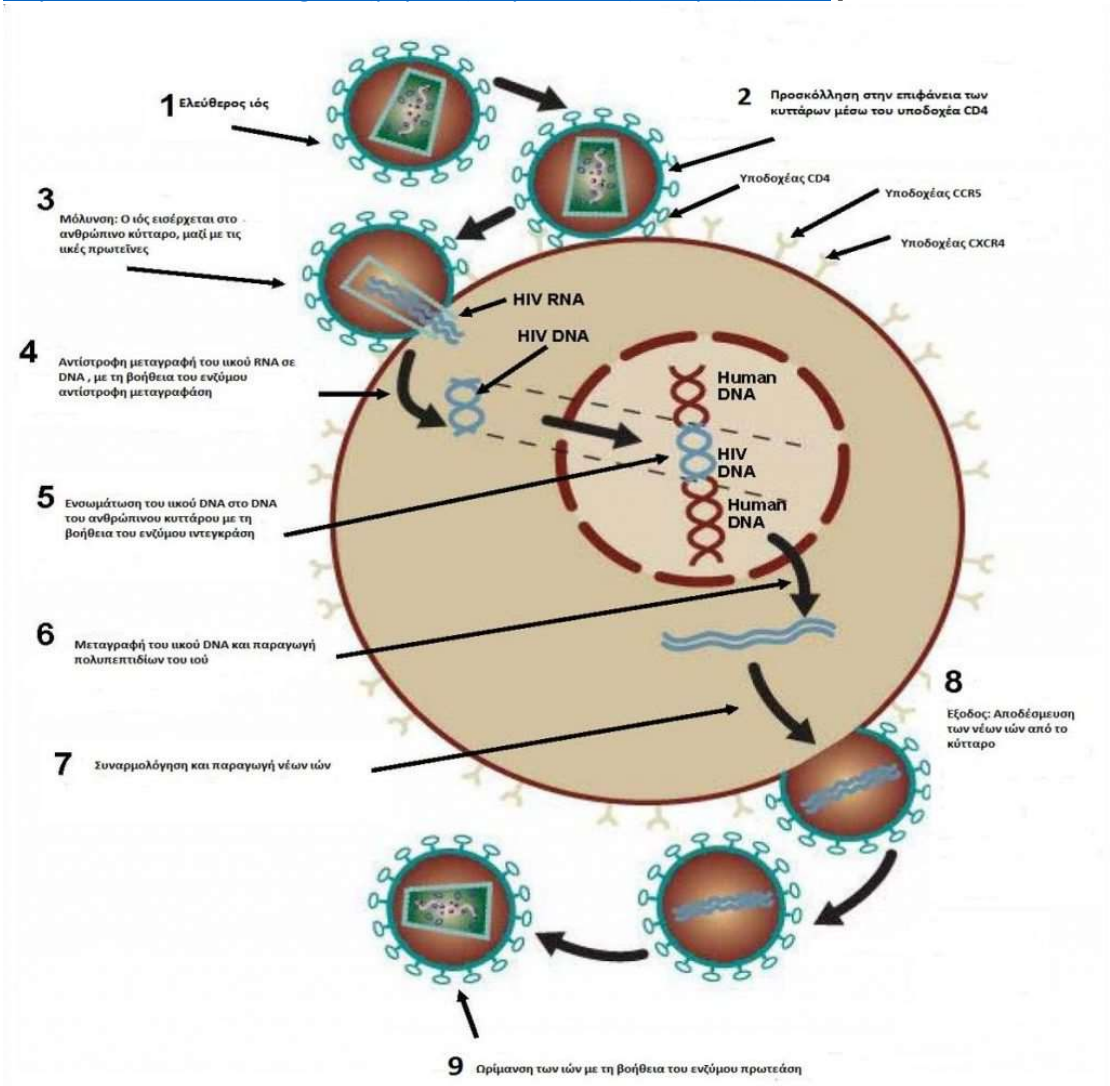
Εικόνα 3 Δομή Γονιδιώματος HIV-1 Το γονιδίωμα του HIV-1. Ο HIV-1 κωδικοποιεί τρία κύρια γονίδια, 5-gag-pol-env-3 που κωδικοποιούν δομικές, βοηθητικές και ρυθμιστικές πρωτεΐνες. **[Πηγή: https:/www.hivbook.com/tag/immune-response/**



1.3 Ο κύκλος ζωής του HIV

Ο κύκλος ζωής του ιού HIV συνοψίζεται στα ακόλουθα στάδια: Είσοδος του ιού στο κύτταρο-ξενιστή, σύνθεση και ενσωμάτωση του ιικού DNA στο γονιδίωμα του κυττάρου-ξενιστή, μεταγραφή των ιικών γονιδίων, σύνθεση των ιικών πρωτεϊνών και συγκρότηση και εκβλάστηση του ιού (Simon et al., 2006).

Εικόνα 4 Σχηματική απεικόνιση του κύκλου ζωής του ιού HIV [Πηγή: https://www.kentrozois.gr/ενημερωση/θεραπεια-τησ-λοιμωξησ-hiv/]



1.4 Γενετική ετερογένεια, η ταξινόμηση και οι υπότυποι του HIV

Δύο είδη του HIV έχουν καταγραφεί, ο HIV-1 και ο HIV-2, τα οποία, σύμφωνα με την ταξινόμηση της ICTV3, ανήκουν στο γένος των *Lentivirus* της οικογένειας των ρετροϊών (Retroviridae) [ICTV, 2017].

Ο HIV εμφανίζει μεγάλο βαθμό ετερογένειας, λόγω του υψηλού ρυθμού ανάπτυξης μεταλλαγών, του ανασυνδυασμού και της αντίστροφης μεταγραφάσης, η οποία είναι επιρρεπής σε λάθη και δεν έχει διορθωτική ικανότητα [Boeyer et al., 1992 ; Tebit, D.M. and E.J. Arts, 2011]. Επίσης, σημαντικό μέρος της γενετικής ετερογένειας του ιού οφείλεται στο *γενετικό ανασυνδυασμό*, δηλαδή έναν μηχανισμό ανταλλαγής γενετικού υλικού, ανάμεσα σε δύο διαφορετικά μόρια ικού γονιδιώματος που μπορεί να συμβεί κατά το στάδιο της

αντίστροφης μεταγραφής του ιικού RNA [Hu, W.S. and H.M. Temin,1990]. Η γενετική ετερογένεια του ιού περιπλέκει τη διάγνωση και τη θεραπεία της νόσου και δυσκολεύει την ανάπτυξη εμβολίου (Esparza and Bhamarapravati, 2000; Letvin, 2006; McCutchan, 2000; Tebit and Arts 2011; Van der Groen et al., 1998).

Λόγω της μεγάλης ετερογένειας, ο HIV διακρίνεται σε 2 τύπους, τον HIV-1 και HIV-2, καθένας από τους οποίους, διακρίνεται σε επί μέρους ομάδες. Πιο συγκεκριμένα, ο HIV-1 διακρίνεται σε 4 ομάδες, τις M (main), N (non-M), O (outlier) και P, ενώ ο HIV-2, σε 8 ομάδες, από τις οποίες οι πιο επικρατείς είναι οι A και B. Η ομάδα M ήταν η πρώτη που ανακαλύφθηκε και είναι υπεύθυνη για το μεγαλύτερο ποσοστό των μολύνσεων με HIV-1. Η ομάδα M του HIV-1 εμφανίζει τη μεγαλύτερη ετερογένεια και διακρίνεται επί πλέον σε 9 υπότυπους (A-D, F-H, J και K) (μερικοί από τους υπότυπους της ομάδας M υποδιαιρούνται σε υπό-υπότυπους (A1, A2, A3, A4, F1, F2)) και 49 ανασυνδυασμένες μορφές (Circulating Recombinant Forms, CRF) των υποτύπων, οι οποίοι διαφέρουν πολύ γενετικά.

Όταν δύο ιικά στελέχη, που προέρχονται από τον ίδιο ή από διαφορετικούς υπότυπους, αναμείξουν το γενετικό τους υλικό σε ένα κύτταρο ξενιστή, τότε δημιουργείται ένας νέος υβριδικός ιός. Τα συγκεκριμένα ιικά στελέχη ονομάζονται CRFs (κυρίαρχες ανασυνδυασμένες μορφές, Circulating Recombinant Forms) και ευθύνονται για το 20% των μολύνσεων σε ορισμένες περιοχές, όπως είναι η Νοτιοανατολική Ασία. Όταν ανασυνδυασμένα στελέχη συναντώνται σε λιγότερο από τρία άτομα, που δεν σχετίζονται επιδημιολογικά μεταξύ τους, τότε ονομάζονται μοναδικά ανασυνδυασμένα στελέχη (Unique Recombinant Forms, URFs).

Πρόσφατες μελέτες αναγνώρισαν μια νέα ομάδα στην Αφρική, την ομάδα P. Η ομάδα O τακτοποιήθηκε το 1990 και περιορίζεται κυρίως στο Καμερούν, την Γκαμπόν και τις γειτονικές χώρες. Η ομάδα N ανακαλύφθηκε το 1998 και μέχρι σήμερα έχουν αναφερθεί μόλις 13 περιστατικά μολύνσεων που περιορίζονται στην περιοχή του Καμερούν. Τέλος, η ομάδα P, που τακτοποιήθηκε το 2007, έχει βρεθεί σε δυο άτομα από το Καμερούν που ζουν στη Γαλλία (Charneau et al., 1994; Paraskevis et al., 2007; Rambaut et al., 2004; Sharp and Hahn, 2011; Takebe et al., 2008; Tebit and Arts, 2011 ; Knipe, D.M. and P.M. Howley, 2013; Tebit, D.M. and E.J. Arts, 2011; Plantier, J.C., et al.,2009).

Οι υπότυποι του ιού εμφανίζουν παρόμοια μολυσματικότητα, μοιράζονται κοινά βιολογικά και μορφολογικά χαρακτηριστικά και χρησιμοποιούνται συνήθως ως επιδημιολογικοί δείκτες για την αναγνώριση της διαδρομής που ακολούθησε ο ιός κατά τη διάδοσή του (Carr et al., 1998, McCutchan, 2000, Robertson et al., 2000, Sharp and Hahn, 2011, Simon et al., 2006, Takebe et al., 2008, Tebit and Arts, 2011). Ωστόσο, εμφανίζουν διαφορετικό ρυθμό εξέλιξης (Abecasis et al., 2009). Επιπλέον, ορισμένες μελέτες υποστηρίζουν ότι υπάρχει διαφορά στη φυσική ιστορία της νόσου ανάμεσα στους διαφορετικούς υπότυπους.

Η γεωγραφική διασπορά των υποτύπων και των CRFs αντικατοπτρίζει την πολυπλοκότητα της μοριακής επιδημιολογίας του HIV-1

• Ο υπότυπος A είναι κοινός στην Αφρική (όπως η δυτική Αφρική, η Κένυα και η Τανζανία) και χώρες Πρώην Σοβιετικής Ένωσης.

• Ο υπότυπος B είναι η κυρίαρχη μορφή της επιδημίας στην **Ευρώπη**, τη Βόρεια Αφρική, την **Αμερικανική ήπειρο**, την **Ιαπωνία** και την Αυστραλία.

• Ο υπότυπος C είναι η κυρίαρχη μορφή στην Αφρική (Νότια, Ανατολική, Υποσαχάρια) την Ινδία, το Νεπάλ και τα μέρη της Κίνας.

•Ο υπότυπος D βρίσκεται στην Ανατολική και Κεντρική Αφρική (Σουδάν, Ουγκάντα, Λιβύη).

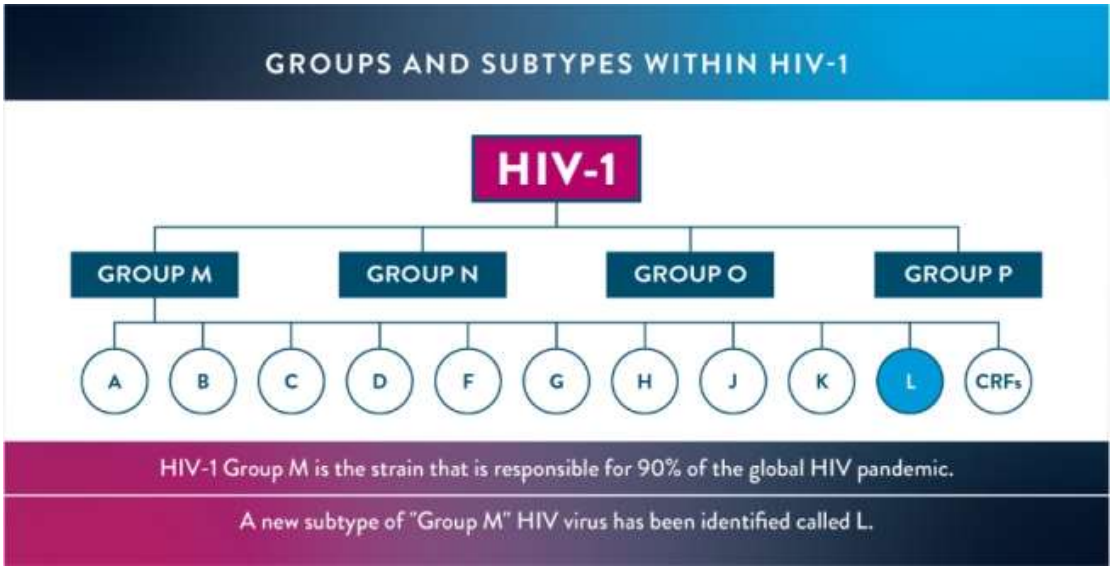
•Ο υπότυπος CRF01_AE βρίσκεται στη Νοτιοανατολική Ασία και θεωρείται η κυρίαρχη μορφή για ετεροφυλόφιλους.

•Ο υπότυπος F έχει σημαντική παρουσία στην κεντρική Αφρική, τη Νότια Αμερική και την Ανατολική Ευρώπη.

• Ο υπότυπος G βρίσκεται κυρίως στην Αφρική και την κεντρική Ευρώπη.

• Οι υπότυποι H, K και CRF04_cpx παρατηρούνται σε μικρές επιδημίες, κυρίως στην κεντρική Αφρική.
•Ο υπότυπος J βρίσκεται κυρίως σε μικρές επιδημίες στη Βόρεια, Κεντρική και Δυτική Αφρική, καθώς και την Καραϊβική.
•Στη Λατινική Αμερική, εκτός από τον υπότυπο B, συναντώνται και ανασυνδυασμένα BF στελέχη.
•Στην Ανατολική Ευρώπη και Κεντρική Ασία επικρατεί το ανασυνδυασμένο στέλεχος CRF03_AB, λόγω της εκτεταμένης χρήσης ενδοφλέβιων ουσιών, και σε μικρότερο ποσοστό οι υπότυποι A και B. Σε πολλές περιοχές κυριαρχούν ανασυνδυασμένα στελέχη, όπως το στέλεχος CRF01_AE στην Νοτιοανατολική Ασία και το CRF02_AG σε χώρες της Δυτικής Αφρικής (Robertson et al., 2000 ; Paraskevis and Hatzakis, 1999, Takebe et al., 2008).

Εικόνα 5 Σχήμα ταξινόμησης των HIV-1 με τις αντίστοιχες ομάδες, υπότυπους και ανασυνδυασμένες μορφές του ιού **[Πηγή: https://www.farmakeutikoskosmos.gr/article-f/anakalyf9hke-neos-ypotypos-toy-ioy-hiv/22780]**



1.5. Η επιδημιολογία του HIV

Η μελέτη των αλληλουχιών του ιού με μεθόδους μοριακής επιδημιολογίας έχει συμβάλλει τα μέγιστα στην κατανόηση της επιδημιολογίας της νόσου. Συγκεκριμένα, έχουν εκτιμηθεί η χρονολογική προέλευση της παγκόσμιας αλλά και επιμέρους επιδημιών (Korber et al., 2000), η γεωγραφική κατανομή των υποτύπων και ανασυνδυασμένων τύπων του ιού (Hamelaar et al., 2011), η προέλευση της επιδημίας σε διαφορετικές γεωγραφικές περιοχές και ο τρόπος διασποράς μεταξύ ή εντός πληθυσμών (Takebe et al., 2008).

Από την αρχή της επιδημίας HIV/AIDS μέχρι και σήμερα, εκτιμάται ότι 77,3 εκ. (59,9 εκ. – 100 εκ.) άνθρωποι έχουν μολυνθεί από τον ιό και 35,4 εκ. (25 εκ. – 49,9 εκ.) άνθρωποι έχουν πεθάνει από κάποια ασθένεια που σχετίζεται με το AIDS. Ο αριθμός των ανθρώπων παγκοσμίως, συμπεριλαμβανομένου όλων των ηλικιών, που ζούσαν το 2017 με HIV ήταν 36,9 εκ. (31,1 εκ.– 43,9 εκ.) και από αυτούς, το 1,8 εκ. ήταν παιδιά κάτω των 15 ετών. Ο συνολικός αριθμός θανάτων, για το 2017, εξαιτίας της HIV/AIDS λοίμωξης ήταν 0,94 εκ. (0,67 εκ. – 1,3 εκ.) και αντίστοιχα, η θνησιμότητα ήταν περίπου 12,7 θάνατοι ανά 100 χιλιάδες, ενώ η θνητότητα 25 θάνατοι ανά 1000 ασθενείς [UNAIDS, 2018].

Η Υποσαχάρια Αφρική εξακολουθεί να πλήττεται περισσότερο και αντιπροσωπεύει το 69% των μολύνσεων παγκοσμίως. Ακολουθούν η Καραϊβική, η Ανατολική Ευρώπη και η Κεντρική Ασία.

Σύμφωνα με τις εκτιμήσεις του Παγκόσμιου Οργανισμού Υγείας (World Health Organization, WHO) για το 2017, ο παγκόσμιος επιπολασμός της επιδημίας για τις ηλικίες 15-49 ήταν 0,8% (0,6% - 0,9%). Ο υψηλότερος επιπολασμός της επιδημίας, στις αντίστοιχες ηλικίες, παρουσιάζεται στην Υποσαχάρια Αφρική και εκτιμάται στο 4,1% (3,4%-4,8%), ενώ για τις υπόλοιπες ηπείρους, ο επιπολασμός εκτιμάται πως δεν ξεπερνάει το 0,5% ανά ήπειρο [WHO, 2017].

Ο αριθμός των νέων λοιμώξεων για το έτος 2017, σε παγκόσμιο επίπεδο, εκτιμάται πως ήταν 1,8 εκ. (1,4 εκ. – 2,4 εκ.) και η αντίστοιχη επίπτωση 0,25 (0,19 – 0,33) νέες HIV λοιμώξεις ανά 1000 ανθρωποέτη. Στην Υποσαχάρια περιοχή της Αφρικής εκτιμάται ότι υπήρξαν 1,2 εκ. νέες λοιμώξεις για το ίδιο έτος, που αντιστοιχούν σε επίπτωση 1,22 (0,9 – 1,64) λοιμώξεις ανά 1000 ανθρωποέτη [WHO, 2017]

Ο αριθμός των νέων HIV λοιμώξεων παρουσιάζει πτωτική τάση τα τελευταία χρόνια και έχει σχεδόν υποδιπλασιαστεί, με 47% μείωση, σε σχέση με το 1996. Κάτι ανάλογο ισχύει και με τον αριθμό θανάτων που έχουν σχέση με το AIDS, αφού έχουν μειωθεί περισσότερο από 51% συγκριτικά με το 2004, γεγονός που οφείλεται σε μεγάλο βαθμό και στην αύξηση του αριθμού των οροθετικών που έχουν πρόσβαση σε αντιρετροϊκή θεραπεία [UNAIDS, 2018].

Στην Ελλάδα, μέχρι και το 2017, έχουν δηλωθεί στο εθνικό σύστημα επιδημιολογικής επιτήρησης, συνολικά, 16.669 HIV οροθετικά άτομα, από τους οποίους το 82,84% ήταν άντρες, το 2,82% γυναίκες, ενώ για το 0,24% δεν δηλώθηκε φύλο [ΚΕ.ΕΛ.Π.ΝΟ, 2017].

1.6. Το Σύνδρομο Επίκτητης Ανοσολογικής Ανεπάρκειας (AIDS)

Το Σύνδρομο της Επίκτητης Ανοσολογικής Ανεπάρκειας (Acquired Immune Deficiency Syndrome, AIDS) είναι η νόσος του ανθρώπινου ανοσοποιητικού συστήματος που προκαλείται από τον ιό της Ανθρώπινης Ανοσοανεπάρκειας (Human Immunodeficiency Virus, HIV). Η νόσος δρα στο ανοσοποιητικό σύστημα, καθιστώντας άτομα που πάσχουν από AIDS ευάλωτα σε ευκαιριακές λοιμώξεις και καρκινικούς όγκους, που το ανοσοποιητικό σύστημα υγιών ατόμων θα αντιμετώπιζε αποτελεσματικά με σχετική ευκολία [Sepkowitz, K.A, 2001].

Ένας φορέας του ιού HIV θεωρείται ότι έχει αναπτύξει τη νόσο AIDS όταν η συγκέντρωση των $CD4_+$ κυττάρων του στον ορό του αίματος είναι μικρότερη των 200 CD4+/μL ή όταν εμφανίσει συγκεκριμένες νόσους που σχετίζονται με προχωρημένο στάδιο της HIV λοίμωξης, όπως πνευμονία, σάρκωμα Kaposi (KS), σύνδρομο απίσχνασης (απώλεια βάρους), βλάβες μνήμης, καντιντίαση, κ.α. [Bennett, J.E. et al., 2015].

Η HIV/AIDS λοίμωξη διακρίνεται σε 4 κλινικά στάδια, σύμφωνα με τον Παγκόσμιο Οργανισμό Υγείας (World Health Organization, WHO), τα οποία διαχωρίζονται με βάση τον αριθμό $CD4_+$ κυττάρων στο αίμα, αλλά και με βάση την εμφάνιση ορισμένων συμπτωμάτων [WHO, 2007].

Πίνακας 1 *Τα κλινικά στάδια της HIV/AIDS* λοίμωξης με βάση των αριθμό των *CD4+ κυττάρων ανά mm3 αίματος και ο βαθμός των συμπτωμάτων ανά στάδιο* [WHO, 2007]

Κλινικό Στάδιο	Αριθμός CD4+/mm3 αίματος	Βαθμός Συμπτωμάτων
I	>500	Χωρίς Συμπτώματα
II	350-499	Ήπια Συμπτώματα
III	200-349	Προχωρημένα Συμπτώματα
IV ή AIDS	<200	Σοβαρά Συμπτώματα

Το AIDS είναι πλέον μια παγκόσμια επιδημία (Cohen et al., 2008). Το γεγονός ότι οι πρώτοι ασθενείς στην Αμερική και στην Ευρώπη ήταν άνδρες ομοφυλόφιλοι ή τοξικομανείς, οδήγησε στη διαμόρφωση της υπόθεσης ότι η νόσος σχετίζονταν με συγκεκριμένες πληθυσμιακές ομάδες (συνήθως περιθωριακές). Καθώς όμως φάνηκε ότι η νόσος μεταδίδεται και με την ετεροφυλοφιλική επαφή (Rambaut et al., 2004), αποδείχτηκε

ταυτόχρονα και το γεγονός ότι το AIDS δε σχετίζεται μόνο με τη σεξουαλική συμπεριφορά. Σχετικά με τα παραπάνω, είναι πλέον γνωστό ότι η εξάπλωση της επιδημίας σε παγκόσμια κλίμακα σχετίζεται με παράγοντες όπως οι ταυτόχρονες σεξουαλικές επαφές, η εναλλαγή συντρόφων σε συνδυασμό με την απουσία χρήσης προφυλακτικού, και η παρουσία άλλων σεξουαλικώς μεταδιδόμενων νοσημάτων. Πέραν από τους παραπάνω παράγοντες, η μετακίνηση πληθυσμών, ο μαζικός εμβολιασμός στην Αφρική καθώς και η χρήση ενδοφλέβιων ναρκωτικών έχουν συντελέσει σημαντικά στην εξάπλωση της λοίμωξης. Ειδικότερα, στην περίπτωση των μαζικών εμβολιασμών, η μετάδοση του ιού πραγματοποιήθηκε λόγω της χρήσης κοινών, μη αποστειρωμένων βελόνων στους εμβολιαζόμενους, και αντίστοιχα στην περίπτωση των χρηστών ενδοφλεβίων ναρκωτικών, η κοινή χρήση συρίγγων αποτελεί τον κύριο τρόπο μετάδοσης και διασποράς του ιού (Takebe et al., 2008; Taylor, 1995; Simon et al., 2006).

1.7. Αντιρετροϊκή θεραπεία της λοίμωξης με HIV και Μετάδοση ανθεκτικών HIV στελεχών σε φάρμακα

Ο ιός έχει την ικανότητα να ενσωματώνεται και να παραμένει στο ανθρώπινο γονιδίωμα, χωρίς απαραίτητα να προκαλεί την εκδήλωση κλινικών συμπτωμάτων. Ένας άνθρωπος που έχει μολυνθεί από HIV, καλείται «HIV οροθετικός» και κρίνεται απαραίτητο να λάβει έγκαιρα την κατάλληλη αντιρετροϊκή θεραπεία. Δεν υπάρχει, μέχρι σήμερα, θεραπεία για την εκρίζωση του HIV από τον οργανισμό του οροθετικού. Ωστόσο, οι υπάρχουσες αντιρετροϊκές θεραπείες είναι ικανές να επιβραδύνουν και να περιορίσουν την εξάπλωση του ιού στον οργανισμό, αυξάνοντας σημαντικά το προσδόκιμο και βελτιώνοντας την ποιότητα ζωής των οροθετικών, καθιστώντας την HIV λοίμωξη ως μια ανίατη μεν, αλλά χρόνια και μη θανατηφόρα νόσο. Επίσης, τα αντιρετροϊκά φάρμακα μειώνουν τη συγκέντρωση του ιού στα μολυσματικά υγρά του οροθετικού και έτσι, μειώνεται η πιθανότητα μετάδοσης σε άλλα άτομα [ΚΕ.ΕΛ.Π.ΝΟ, 2018].

Σημαντική πρόοδος έχει πραγματοποιηθεί στη θεραπεία της λοίμωξης με HIV, λόγω της ανάπτυξης αντιρετροϊκών φαρμάκων. Τα φάρμακα αυτά διακρίνονται σε τέσσερις κατηγορίες και στοχεύουν στη διαδικασία της σύντηξης του ιού με το κύτταρο-ξενιστή, στη διαδικασία της αντίστροφης μεταγραφής, στη διαδικασία της ενσωμάτωσης του ιικού γονιδιώματος στο γενωμικό DNA και στην ωρίμανση των ιικών πρωτεϊνών από την πρωτεάση του ιού. Παρακάτω συνοψίζονται τα σημαντικότερα από τα φάρμακα αυτά:

1) Φάρμακα που στοχεύουν στη διαδικασία της σύντηξης του ιού με το κύτταρο-ξενιστή
2) Αναστολείς ιντεγκράσης
3) Αναστολείς πρωτεάσης
4) Φάρμακα που στοχεύουν στη διαδικασία της αντίστροφης μεταγραφής

Ωστόσο ο HIV έχει την ικανότητα να αναπτύσσει αντοχή στα αντιρετροϊικά φάρμακα, κυρίως στους αναστολείς της πρωτεάσης και της αντίστροφης μεταγραφάσης. Η ανάπτυξη αντοχής οφείλεται στην γενετική ποικιλομορφία του ιού, ο οποίος έχει την ιδιότητα να δημιουργεί νέα γενετικά διαφορετικά στελέχη που μπορούν να παρακάμπτουν τη δράση των αντιρετροϊκών φαρμάκων (Kozal, 2009).

Φυλογενετικές μελέτες που διεξάγονται σε υψηλού εισοδήματος χώρες αναφέρουν σταθερά τη συσσώρευση των γονότυπων του ιού HIV που περιέχουν μεταλλάξεις που συνδέονται με την αντοχή σε φάρμακα. Αυτό υποδεικνύει ότι η μετάδοση του HIV με μετάλλαξη αντοχής σε φάρμακο είναι κυρίως μεταδιδόμενη μεταξύ των ατόμων που δεν έχουν λάβει αντιρετροϊκή θεραπεία. Ο HIV ανθεκτικός στο φάρμακο από άτομα που απέκτησαν μεταλλάξεις που σχετίζονται με την αντίσταση κατά τη διάρκεια αντιρετροϊκής θεραπείας με φάρμακα φαίνεται ότι μεταδίδεται ασυνήθιστα σε πρόσφατα μολυσμένα άτομα.

1.8. Τρόποι μετάδοσης του HIV Παράγοντες κινδύνου και ομάδες υψηλού κινδύνου

Ο HIV βρίσκεται σε υψηλές συγκεντρώσεις στα υγρά του σώματος, όπως το αίμα και τα σπερματικά ή κολπικά υγρά. Γι' αυτό το λόγο, η μετάδοση του HIV μπορεί να γίνει μέσω τριών κύριων οδών.

1. Έκθεση σε μολυσματικά σπερματικά ή κολπικά υγρά μέσω σεξουαλικής επαφής (κυρίως χωρίς τη χρήση προφυλακτικού). Η μετάδοση μπορεί να πραγματοποιηθεί, κατά τη διάρκεια της σεξουαλικής επαφής, είτε από μολυσμένο $CD4_+$ Τ-λεμφοκύτταρο ή από ελεύθερο ιό στα μολυσματικά σπερματικά/κολπικά υγρά [Dimmock, N.J., A.J. Easton, and K.N. Leppard, 2016]. Το ιικό φορτίο του οροθετικού παίζει σημαντικό ρόλο στον κίνδυνο μετάδοσης του ιού και έτσι, η μολυσματικότητα ενός οροθετικού ατόμου είναι σημαντικά υψηλότερη κατά το πρώτο χρονικό διάστημα της λοίμωξης, λόγω της υψηλής συγκέντρωσης ιικού φορτίου [Dosekun, O. and J. Fox, 2010].

2. Έκθεση σε μολυσμένο αίμα ή παράγωγά του, το οποίο μπορεί να πραγματοποιηθεί από μεταγγίσεις μολυσμένου αίματος, χρήση κοινής σύριγγας, χρήση μη αποστειρωμένων ιατρικών οργάνων, κλπ. [Modrow, S, 2013].

3. Κάθετη μετάδοση από μητέρα σε παιδί, κατά τη διάρκεια της κυοφορίας, του τοκετού ή του θηλασμού. Ο κίνδυνος μόλυνσης μέσω της συγκεκριμένης οδού μπορεί να περιοριστεί σημαντικά αν η οροθετική μητέρα λαμβάνει αντιρετροϊκή θεραπεία κατά τη διάρκεια της εγκυμοσύνης και του τοκετού, με αποτέλεσμα να μειωθεί το ιικό της φορτίο [Dimmock, N.J., A.J. Easton, and K.N. Leppard, 2016].

Ορισμένες ομάδες ανθρώπων εμφανίζουν αυξημένο κίνδυνο μόλυνσης από τον HIV, κυρίως, λόγω συγκεκριμένων παραγόντων κινδύνου που αυξάνουν την πιθανότητα έκθεσης στον ιό, όπως επικίνδυνη σεξουαλική συμπεριφορά ή κοινή χρήση βελόνας, σύριγγας ή άλλων ενδοφλέβιων εξαρτημάτων [CDC, 2018]. Κάθε ομάδα κινδύνου, ή συνώνυμα κατηγορία μετάδοσης, διακρίνεται από ορισμένα χαρακτηριστικά που πληροφορούν για τον τρόπο που πραγματοποιήθηκε η μετάδοση. Η γνώση των παραγόντων συμπεριφοράς της κάθε ομάδας έχει ιδιαίτερη επιδημιολογική σημασία, καθώς συμβάλει στην πληρέστερη κατανόηση των χαρακτηριστικών της επιδημίας και στη λήψη αποτελεσματικότερων μέτρων πρόληψης.

Ο κυριότερος τρόπος μετάδοσης του HIV, από την αρχή της επιδημίας μέχρι σήμερα, στην Ελλάδα είναι η απροφύλακτη σεξουαλική επαφή μεταξύ ανδρών με ποσοστό 48,4% και ακολουθούν η απροφύλακτη ετεροφυλοφιλική επαφή (21,5%), η χρήση ενέσιμων εξαρτησιογόνων ουσιών (11,5%), ενώ με μικρές συχνότητες η μόλυνση μέσω πολυμετάγγισης παράγωγων του αίματος (1,4%), μετάγγισης αίματος (0,6%) και μέσω κάθετης μετάδοσης (0,4%). Σημαντικό μειονέκτημα του εθνικού συστήματος επιδημιολογικής επιτήρησης αποτελεί ένα σχετικά υψηλό ποσοστό, που αντιστοιχεί σε HIV οροθετικούς για τους οποίους δεν είναι γνωστός ο τρόπος με τον οποίο μολύνθηκαν από τον ιό και έχουν δηλωθεί στο ΚΕ.ΕΛ.Π.ΝΟ. με ακαθόριστη κατηγορία μετάδοσης (16,2%)*)* [ΚΕ.ΕΛ.Π.ΝΟ, 2017].

Αν δεν ληφθούν υπόψιν οι οροθετικοί ακαθόριστης κατηγορίας μετάδοσης, το ποσοστά διαμορφώνεται σε 57,8% για τους άντρες με σεξουαλική επαφή με άλλους άντρες, 25,7% για τους ετεροφυλόφιλους και 13,6% για τους χρήστες ενδοφλέβιων ναρκωτικών [ΚΕ.ΕΛ.Π.ΝΟ, 2017].

Τα έτη 2011 και 2012, παρουσιάστηκε μια μεγάλη αύξηση στον αριθμό των νέων λοιμώξεων που αφορούσε κυρίως σε άτομα που έκαναν χρήση ενέσιμων εξαρτησιογόνων ουσιών [Paraskevis, D, et al, 2013]. Τα τελευταία χρόνια, οι περισσότερες νέες HIV διαγνώσεις ανά χρόνο αφορούσαν κυρίως σε άντρες που είχαν σεξουαλική επαφή με άντρες, με

εξαίρεση το 2012 που αφορούσαν σε χρήστες ενδοφλέβιων ναρκωτικών *)* [ΚΕ.ΕΛ.Π.ΝΟ, 2017].

Επομένως η οροθετικότητα στον HIV-1 σχετίζεται με διάφορους παράγοντες κινδύνου, οι οποίοι μπορεί να είναι περιβαλλοντικοί, παράγοντες συμπεριφοράς ή η προσωπική κατάσταση των ατόμων. Οι παράγοντες κινδύνου συνοψίζονται στους εξής:

Φύλο: Σύμφωνα με μελέτες οι άντρες είναι λιγότερο πιθανό να μολυνθούν με HIV-1 μέσω ετεροφυλόφιλης επαφής σε σχέση με τις γυναίκες. Ωστόσο, ορισμένα πολιτισμικά πρότυπα ενθαρρύνουν τους άντρες να έχουν πολλαπλούς συντρόφους αυξάνοντας τον κίνδυνο μόλυνσης. Επιπλέον, οι άντρες είναι περισσότερο πιθανό να προβούν σε χρήση ενέσιμων ουσιών και έτσι αυξάνεται ο κίνδυνος μετάδοσης από μολυσμένες βελόνες και σύριγγες. Οι γυναίκες αποτελούν σχεδόν το 50% των ατόμων που ζουν με HIV παγκοσμίως. Τα τελευταία 10 χρόνια ο παγκόσμιος αριθμός των οροθετικών γυναικών έχει μείνει σταθερός, αν και αυξομειώνεται σε διάφορες περιοχές. Κύριες αιτίες μόλυνσης των γυναικών είναι η χωρίς προστασία σεξουαλική επαφή με μολυσμένους συντρόφους και δευτερευόντως η χρήση ναρκωτικών (UNAIDS, 2012).

Χρήση ενδοφλέβιων ουσιών: Οι χρήστες ναρκωτικών είναι μια από τις πιο ευπαθείς ομάδες στη μόλυνση με τον ιό HIV. Η κοινή χρήση συρίγγων και βελόνων για την πρόσληψη ενδοφλέβιων ναρκωτικών αποτελεί τον κύριο τρόπο μετάδοσης της συγκεκριμένης ομάδας κινδύνου. Το γεγονός ότι η χρήση ναρκωτικών διώκεται ποινικά αποτρέπει τους χρήστες από την αναζήτηση βοήθειας σε νοσοκομεία, κλινικές και κέντρα υποστήριξης, ενώ η κατάσταση επιδεινώνεται μέσα στις φυλακές, όπου η διακίνηση ναρκωτικών και η κακοποίηση ανάμεσα σε κρατούμενους γίνεται χωρίς έλεγχο (UNAIDS, 2012).

Σεξουαλικές προτιμήσεις: Οι ομοφυλόφιλοι αποτελούν άλλη μια ομάδα κινδύνου. Κύρια αιτία μετάδοσης είναι η σεξουαλική επαφή χωρίς προφύλαξη. Τα τελευταία χρόνια, στις χώρες του Δυτικού Κόσμου, χάρη στην ενημέρωση πάνω στον ιό και τα μέτρα προστασίας, ο αριθμός των ομοφυλόφιλων οροθετικών έχει μειωθεί σημαντικά. Ωστόσο, στις χώρες του Τρίτου Κόσμου τα ποσοστά παραμένουν υψηλά (UNAIDS, 2012).

Ηλικία: Περίπου το 50% των νέων μολύνσεων παγκοσμίως αφορά νέους ηλικίας 18-24 ετών. Κύριες αιτίες μόλυνσης είναι η σεξουαλική επαφή χωρίς προστασία, η έλλειψη ενημέρωσης από την οικογένεια και το εκπαιδευτικό σύστημα και η χρήση ναρκωτικών. Σημαντικό ποσοστό των νέων μολύνσεων αποτελούν άτομα άνω των 50 ετών, τα οποία λόγω άγνοιας προβαίνουν σε μη ασφαλείς σεξουαλικές πρακτικές. Τέλος, τα παιδιά έως 10 ετών αποτελούν άλλη μια ομάδα υψηλού κινδύνου. Κύρια αιτία μόλυνσης των παιδιών είναι η παιδική εκμετάλλευση, ενώ σημαντικό ποσοστό καταλαμβάνουν τα παιδιά που γεννήθηκαν οροθετικά ή που απέκτησαν τον ιό κατά τον τοκετό από μητέρα – φορέα (UNAIDS, 2012).

Κοινωνικοοικονομική κατάσταση: Η κοινωνικοοικονομική κατάσταση των ατόμων καθορίζεται από το εισόδημα, το επάγγελμα και το μορφωτικό τους επίπεδο. Άτομα χαμηλού κοινωνικοοικονομικού επιπέδου βρίσκονται σε αυξημένο κίνδυνο μόλυνσης και μετάδοσης του HIV-1 κυρίως λόγω έλλειψης πληροφόρησης για τον ιό, ενώ έχουν αυξημένη πιθανότητα να προβούν σε χρήση ναρκωτικών και σε μη ασφαλείς ερωτικές πρακτικές. Επιπλέον, τα άτομα αυτά καταλήγουν γρηγορότερα σε σχέση με οροθετικούς μεγαλύτερης κοινωνικοοικονομικής κατάστασης λόγω έλλειψης πρόσβασης σε ιατρική φροντίδα (UNAIDS, 2012).

Τόπος κατοικίας: Η νοσηρότητα και η θνησιμότητα του HIV-1 διαφέρει ανάμεσα στις χώρες λόγω της περιορισμένης πρόσβασης σε ιατρική φροντίδα, της φτώχειας και των διακρίσεων (UNAIDS, 2012).

Μετανάστευση: Οι μετανάστες αποτελούν ομάδα υψηλού κινδύνου λόγω του ανθυγιεινού περιβάλλοντός τους, της νομικής τους κατάστασης η οποία προσδιορίζει τον τρόπο ζωής τους και της έλλειψης πρόσβασης σε ιατρική φροντίδα (Decosas et al., 1995; Decosas and Adrien, 1997; UNAIDS, 2012)

ΚΕΦΑΛΑΙΟ 2

Βασικές αρχές μοριακής εξέλιξης και φυλογενετικής ανάλυσης

2.1 Εισαγωγή στη μοριακή επιδημιολογία

Η Επιδημιολογία είναι ο επιστημονικός κλάδος, ο οποίος έχει ως στόχο να διερευνήσει την κατανομή και την αιτιολογία των παραγόντων που σχετίζονται με τα νοσήματα, τη συχνότητα και την κατανομή των νοσημάτων στους πληθυσμούς, την αξιολόγηση των υπηρεσιών υγείας και, επιπλέον, να εφαρμόσει τα ευρήματα ερευνών με σκοπό τον πρόληψη και αντιμετώπιση των νοσημάτων και γενικότερα τη βελτίωση της δημόσιας υγείας. Πολλές μέθοδοι μπορούν να χρησιμοποιηθούν για τη διεξαγωγή επιδημιολογικών ερευνών: περιγραφικές μελέτες και μελέτες παρατήρησης χρησιμοποιούνται για να διερευνηθεί η κατανομή και αναλυτικές μελέτες για την αποσαφήνιση των καθοριστικών παραγόντων της νόσου (World Health Organization, 2016).

Μια κατηγορία επιδημιολογικής έρευνας, είναι η μοριακή επιδημιολογία, η οποία ορίζεται ως η μελέτη γενετικών και περιβαλλοντικών παραγόντων για τη διερεύνηση της πιθανής συσχέτισης τους με νοσήματα σε ανθρώπινους πληθυσμούς, κάνοντας χρήση ανιχνεύσιμων σε μοριακό επίπεδο δεικτών που αφορούν γενετικούς και περιβαλλοντικούς παράγοντες. Η μοριακή επιδημιολογία θεωρείται ένας συνδυασμός της μοριακής βιολογίας και της επιδημιολογίας. Η μοριακή επιδημιολογία ουσιαστικά εφαρμόζει τις τεχνικές μεθόδους της μοριακής βιολογίας με χρήση βιολογικών δεικτών ή μετρήσεων σε επιδημιολογικές μελέτες. Οι μοριακές τεχνικές έχουν συμβάλει σημαντικά στην αύξηση της ακρίβειας των βιολογικών μετρήσεων προσφέροντας μεγαλύτερη ικανότητα να ανιχνεύονται πιθανές συσχετίσεις και συνεπώς καλύτερα και πιο αξιόπιστα στοιχεία για την ανακάλυψη και την κατανόηση της αιτιολογίας μιας νόσου. (Foxman, 2001).Επίσης η μοριακή επιδημιολογία βρίσκει εφαρμογές στη μελέτη επιδημιών από ταχέως εξελισσόμενα παθογόνα όπως οι ιοί. Αξίζει να σημειωθεί ότι τα αποτελέσματα μοριακών επιδημιολογικών ερευνών συνέβαλαν στην αναγνώριση της προέλευσης και των διαδρομών μετάδοσης πληθώρας λοιμωδών νοσημάτων, όπως επίσης και στην αξιολόγηση της αποτελεσματικότητας των προληπτικών μέτρων τα οποία έχουν διενεργηθεί (λ.χ. εμβολιασμοί).

2.2. Μοριακή εξέλιξη

Η μοριακή εξέλιξη αποτελεί πεδίο της εξελικτικής βιολογίας, δηλαδή της επιστήμης που μελετά την προέλευση και τη διαδικασία μεταβολής όλων των ζωντανών οργανισμών στο πέρασμα του χρόνου, από την πρωτοεμφάνιση της ζωής έως και σήμερα, αλλά και τους μηχανισμούς που δρουν για να πραγματοποιηθεί αυτή η μεταβολή. Η μοριακή εξέλιξη αξιοποιεί όλη τη διαθέσιμη πληροφορία που βρίσκεται στο γενετικό υλικό των οργανισμών. Αυτή η πληροφορία του γενετικού υλικού έχει την ιδιότητα να αποθηκεύεται με τη μορφή νουκλεοτιδίων (DNA ή RNA για ορισμένους ιούς), να κληρονομείται στους απογόνους και να ελέγχει την ανάπτυξη του αντίστοιχου οργανισμού που τη φέρει. Καθώς αυτή η γενετική πληροφορία κληρονομείται και “περνάει” από γενιά σε γενιά συσσωρεύει έναν αριθμό μεταλλάξεων, ικανών να οδηγήσουν σε μικρότερες ή μεγαλύτερες μεταβολές στη δομή του οργανισμού. Επομένως, οι νέες μεταλλάξεις που τελικά θα κληρονομηθούν από τους απογόνους θα αποτελέσουν την πηγή της γενετικής ποικιλότητας, πάνω στην οποία βασίζεται η μοριακή εξέλιξη. Η μοριακή εξέλιξη περιλαμβάνει δύο βασικά, αλλά αλληλεπικαλυπτόμενα πεδία:

i τη μελέτη των μεταλλάξεων που γίνονται στο γενετικό υλικό, του ρυθμού εμφάνισης των μεταλλάξεων και του μοντέλου που ακολουθούν αυτές οι μεταλλάξεις στην πορεία του χρόνου

ii της μοριακής φυλογένειας, στην οποία χρησιμοποιούνται μοριακά δεδομένα (πχ μόρια DNA) για να κατασκευαστούν φυλογενετικά δέντρα με κύριο στόχο την κατανόηση και αναπαράσταση της εξελικτικής ιστορίας των οργανισμών

Ο ρυθμός εμφάνισης και επικράτησης μεταλλάξεων ποικίλει ανάμεσα στα διάφορα είδη οργανισμών. Ο ιός της ανθρώπινης ανοσοανεπάρκειας (HIV) εμφανίζει πολύ υψηλό ρυθμό μεταλλάξεων, έτσι ώστε να συσσωρεύεται, εντός σχετικά μικρού χρονικού διαστήματος, ικανός αριθμός μεταλλάξεων για να μπορεί να τεκμηριωθεί επιδημιολογική σχέση ανάμεσα σε πιθανό δότη και δέκτη. [Παρασκευής Δ, et al., 2017]

2.3. Το γενετικό υλικό: η πηγή όλης της πληροφορίας

Το νουκλεϊκό οξύ είναι το βιομόριο που εμπεριέχει το σύνολο της γενετικής πληροφορίας σε όλους τους οργανισμούς και διακρίνεται σε δεοξυριβονουκλεϊκό οξύ (DNA) και ριβονουκλεϊκό οξύ (RNA). Είναι υπεύθυνο για τη διατήρηση, την αποθήκευση και τη μεταβίβαση της γενετικής πληροφορίας από γενιά σε γενιά. Στη μεγάλη πλειονότητα των οργανισμών, η γενετική πληροφορία οργανώνεται σε μορφή DNA, ενώ εξαίρεση αποτελούν ορισμένοι ιοί, όπως οι ρετροϊοί, στους οποίους η γενετική πληροφορία είναι σε μορφή RNA [Αλαχιώτης, Σ., 2005].

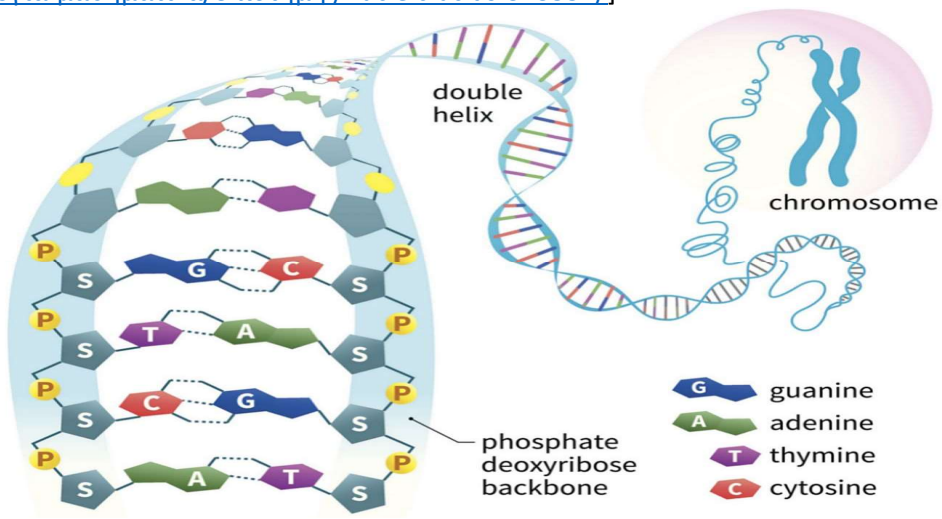
Το νουκλεϊκό οξύ είναι ένα πολυμερές βιομόριο, του οποίου τα μονομερή ονομάζονται νουκλεοτίδια. Κάθε νουκλεοτίδιο αποτελείται από μία πεντόζη ενωμένη με μία αζωτούχο βάση και μία φωσφορική ομάδα. Υπάρχουν δύο κατηγορίες πεντοζών, η δεοξυριβόζη που συναντάται στο DNA και η ριβόζη που βρίσκεται στο RNA. Επίσης, υπάρχουν δύο κατηγορίες αζωτούχων βάσεων, οι πουρίνες και οι πυριμιδίνες. Στις πουρίνες ανήκουν η αδενίνη (A) και η γουανίνη (G), ενώ στις πυριμιδίνες η κυτοσίνη (C), η θυμίνη (T) και η ουρακίλη (U). Κάθε νουκλεοτίδιο του DNA (δεοξυριβονουκλεοτίδιο) αποτελείται από μία δεοξυριβόζη, φωσφορικό οξύ και μία από τις ακόλουθες βάσεις: αδενίνη (A), γουανίνη (G), κυτοσίνη (C) ή θυμίνη (T). Οι παραπάνω βάσεις είναι συμπληρωματικές, δηλαδή η Αδενίνη συνδέεται πάντα με την Θυμίνη, ενώ η Γουανίνη θα συνδέεται πάντα με την Κυτοσίνη, για αυτό και οι δύο νουκλεοτιδικές αλυσίδες που αποτελούν το DNA λέμε πως είναι συμπληρωματικές. Έτσι, αρκεί η μια μόνο από τις δύο αλυσίδες ενός μορίου DNA για να μπορέσουμε να το διαβάσουμε, να καταγράψουμε δηλαδή το κείμενο που βασίζεται στο αλφάβητο των τεσσάρων γραμμάτων που αντιστοιχούν στις 4 αζωτούχες βάσεις. Έτσι, μια αλληλουχία βάσεων DNA θα είναι της μορφής:

…GCTCTCGAAATTTGCGCTCGCTATTTGCGCGCGCATATATA…

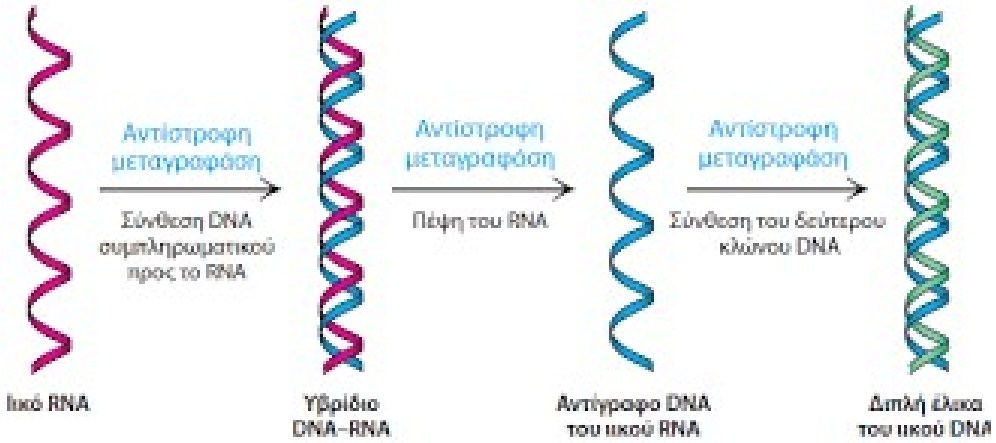
Η χημική σύσταση του RNA είναι παρόμοια με αυτή του DNA, αλλά διαφέρουν στα εξής: το RNA αποτελείται από ριβόζη αντί για δεοξυριβόζη και περιέχει ουρακίλη (U) αντί για θυμίνη (T). Τα νουκλεοτίδια ενώνονται με 3’-5’ φωσφοδιεστερικό δεσμό σχηματίζοντας μία πολυνουκλεοτιδική αλυσίδα και η αλληλουχία των νουκλεοτιδίων καθορίζει την γενετική πληροφορία που περιέχει το νουκλεϊκό οξύ.

Το RNA απαντάται στα κύτταρα κυρίως ως μονόκλωνο. Εξαίρεση αποτελεί το γονιδίωμα ορισμένων ιών, όπου ως γενετικό υλικό μπορεί να έχουν είτε DNA ή RNA, μονόκλωνο ή δίκλωνο [Μαρμάρας, Β. and Μ. Λαμπροπούλου-Μαρμάρα, 2005].

Εικόνα 6 Σχηματική απεικόνιση του DNA [Πηγή:https://www.greelane.com/el/επιστήμη-τεχνολογία-μαθηματικά/επιστήμη /nucleic-acids-373552/]



Εικόνα 7 Ικό RNA
[Πηγή:http://www.biology.uoc.gr/courses/BIOL154/documents/Lecture4.pdf]



2.3.1 Μεταλλάξεις του γενετικού υλικού

Οποιαδήποτε ποιοτική ή ποσοτική αλλαγή που συμβαίνει στο γενετικό υλικό καλείται μετάλλαξη. Οι μεταλλάξεις είναι τυχαία λάθη που συμβαίνουν στο γενετικό υλικό και μπορεί να οδηγήσουν σε παραγωγή διαφορετικής μορφής πρωτεΐνης. Αν η μετάλλαξη συμβεί σε γεννητικά κύτταρα, που μεταβιβάζουν τη γενετική πληροφορία σε θυγατρικά, τότε ο απόγονος μπορεί να διαφέρει φαινοτυπικά από τους γονείς, λόγω αυτής της μετάλλαξης. Οι μεταλλάξεις διακρίνονται σε σημειακές (point mutations), όταν η αλλαγή αφορά τη νουκλεοτιδική αλληλουχία ή τον αριθμό λίγων νουκλεοτιδίων και σε χρωμοσωματικές (chromosomal disorders), όταν επηρεάζουν τη δομή ή τον αριθμό των χρωμοσωμάτων. Οι σημειακές μεταλλάξεις μπορεί να αφορούν την αντικατάσταση, την προσθήκη ή την απαλοιφή ενός νουκλεοτιδίου. Αν η αλλαγή αυτή οδηγεί σε κωδικοποίηση του ίδιου αμινοξέος, τότε η μετάλλαξη καλείται συνώνυμη ή σιωπηλή (synonymous or silent), ενώ αν οδηγεί σε διαφορετικό αμινοξύ, μη συνώνυμη (non-synonymous). Μεταλλάξεις, λόγω προσθήκης ή απαλοιφής νουκλεοτιδίων, που οδηγούν σε αλλαγή της αλληλουχίας όλων των

επόμενων αμινοξέων της πεπτιδικής αλυσίδας ονομάζονται μεταλλάξεις τροποποίησης του αναγνωστικού πλαισίου (frameshift mutations).

Οι μεταλλάξεις είναι, συνήθως, βλαβερές για τον οργανισμό και μπορεί να οδηγήσουν σε θάνατο. Γι' αυτό όλα τα κύτταρα διαθέτουν ενζυμικούς επιδιορθωτικούς μηχανισμούς του γενετικού τους υλικού, οι οποίοι εντοπίζουν και επιδιορθώνουν ορισμένα «λάθη» του γενετικού υλικού. Έτσι, ο αριθμός μεταλλάξεων που θα συμβούν, στην πραγματικότητα, είναι πολύ μεγαλύτερος από αυτό αυτόν που, εν τέλει, θα επικρατήσουν στο κύτταρο.

Οι μεταλλάξεις αποτελούν την κύρια πηγή της γενετικής ποικιλότητας. Αν μια μετάλλαξη είναι ευνοϊκή για τον οργανισμό και βελτιώνει τις πιθανότητές του να επιβιώσει και να αφήσει απογόνους, τότε η μετάλλαξη αυτή θα επικρατήσει και θα συμβάλει στη βιολογική εξέλιξη του οργανισμού, ειδάλλως, θα οδηγήσει σε μικρότερη πιθανότητα επιβίωσης του οργανισμού και των απογόνων του και θα εξαλειφθεί [Αλαχιώτης, Σ., 2007].

Σε αυτή την παράγραφο θα ορίσουμε τι εννοούμε με τον όρο φυλογενετική ανάλυση και με ποιους βασικούς τρόπους μπορούμε να πραγματοποιήσουμε μία τέτοιου είδους ανάλυση. Η ομοιότητα των μοριακών μηχανισμών των οργανισμών που έχουν μελετηθεί υποδηλώνει ότι όλοι οι οργανισμοί στη Γη είχαν έναν κοινό πρόγονο. (Εμίρης, 2016) Αν το γενετικό υλικό των οργανισμών (γονιδίωμα) «εξελίσσεται» λόγω συνεχών μεταλλαγών που συμβαίνουν σε αυτό, τότε οι διαφορές στη νουκλεοτιδική αλληλουχία μεταξύ των γονιδιωμάτων δύο οργανισμών θα μπορούσε να υποδείξει το πόσο πρόσφατα οι δύο οργανισμοί μοιράζονταν τον κοινό τους πρόγονο (Brown, 2002). Ως εκ τούτου, κάθε ομάδα ειδών είναι συγγενική, και αυτή η συγγένεια ονομάζεται φυλογένεση. Η συγγένεια αυτή μπορεί να αναπαραστεί από ένα φυλογενετικό δέντρο. (Εμίρης, 2016) Έτσι, τα γονιδιώματα δύο οργανισμών που διαχωρίστηκαν εξελικτικά πολύ πρόσφατα, αναμένεται να διαφέρουν λιγότερο σε σχέση με τα γονιδιώματα δύο οργανισμών που διαχωρίστηκαν παλιότερα. Εφόσον μελετήσουμε αυτές τις σχέσεις για ένα μεγαλύτερο σύνολο οργανισμών, θα μπορέσουμε να πάρουμε πληροφορίες για τις εξελικτικές σχέσεις μεταξύ τους (Brown, 2002).

Γενικά λοιπόν, τα βασικά ερωτήματα της φυλογενετικής ανάλυσης είναι τα παραπάνω, δηλαδή, το πόσο κοντά εξελικτικά είναι δύο ή περισσότεροι οργανισμοί.

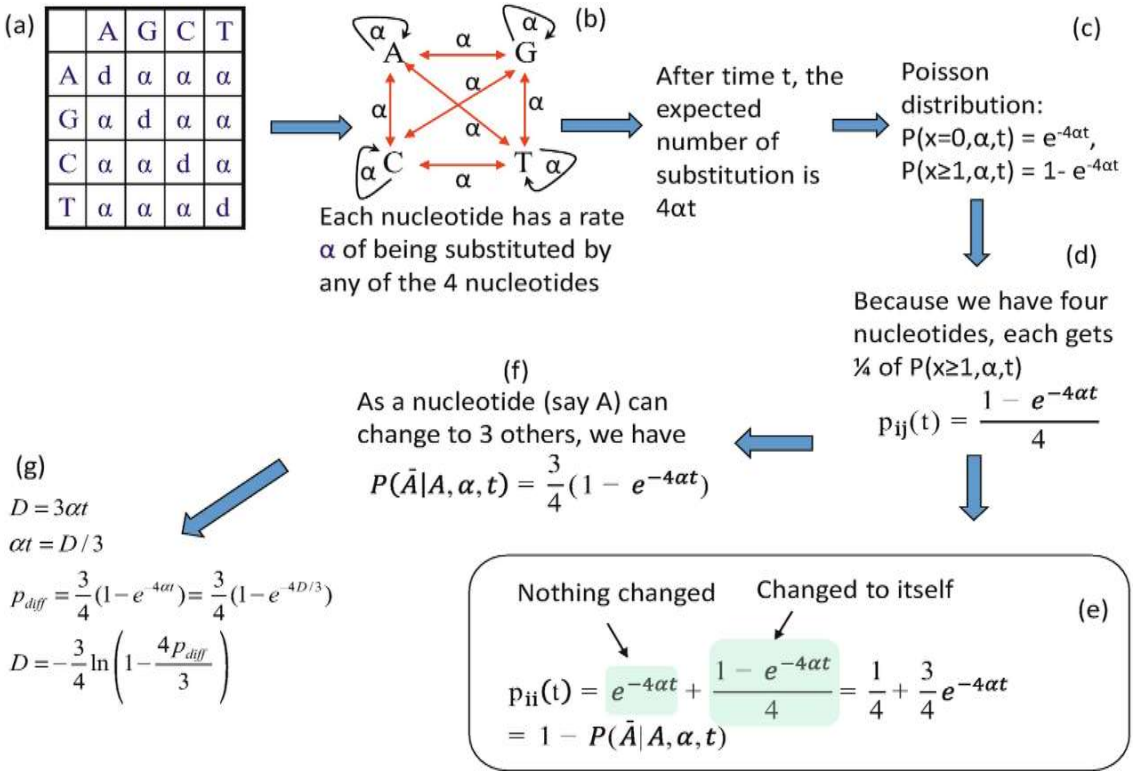
Για να μπορέσει κάποιος να πραγματοποιήσει φυλογενετική ανάλυση χρειάζεται ένα δείγμα από αλληλουχίες νουκλεοτιδίων ή πρωτεϊνών. Αυτές μπορεί να είναι ένα συγκεκριμένο γονίδιο που απομονώθηκε από ένα σύνολο οργανισμών του ίδιου είδους, μπορεί να είναι το συνολικό γονιδίωμα ενός πλήθους οργανισμών του ίδιου είδους, ή μέρος του γονιδιώματος, από διαφορετικά είδη οργανισμών, ενώ μπορούμε να μελετήσουμε τις ίδιες σχέσεις και μέσω ανάλυσης των πρωτεϊνικών αλληλουχιών. Το σημαντικό, μιας και κάνουμε σύγκριση, είναι να συγκρίνουμε το αντίστοιχο μέρος του γονιδιώματος των οργανισμών και όχι τυχαία τμήματα. Τέλος, θα πρέπει οι υπό μελέτη αλληλουχίες να είναι στοιχισμένες, ώστε να μπορέσει να γίνει σωστά η φυλογενετική ανάλυση (Brown, 2002).

Στο DNA ενός οργανισμού υπάρχουν περιοχές που περιέχουν την γενετική πληροφορία σχετικά με τη μορφή, τις ιδιότητες και όλα τα χαρακτηριστικά του οργανισμού. Οι περιοχές αυτές λέγονται γονίδια, ενώ υπάρχουν και κάποιες περιοχές πριν και μετά τα γονίδια που συμμετέχουν στις διαδικασίες με τις οποίες αξιοποιείται η πληροφορία των γονιδίων.

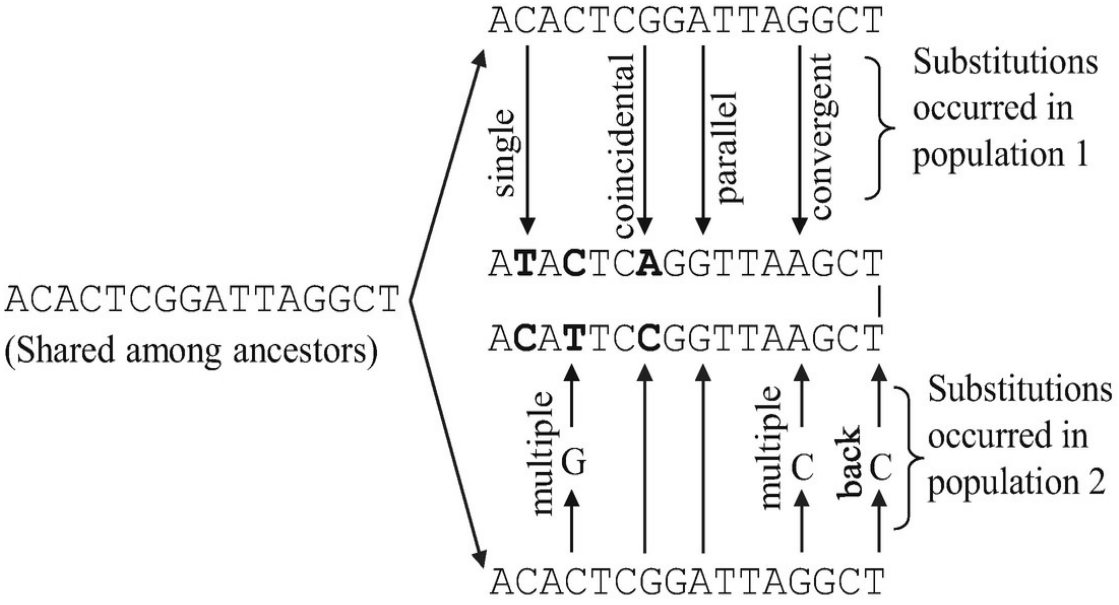
Είναι γεγονός πως με την πάροδο του χρόνου συμβαίνουν διάφορες αλλαγές στην αλληλουχία του DNA, οι μεταλλαγές. Συγκεκριμένα, μια αζωτούχος βάση μπορεί να μετατραπεί σε μια άλλη (πχ A σε T, ή σε G κλπ.), μια βάση μπορεί να χαθεί από την αλληλουχία (έλλειψη) πχ: ATCGCT AT…GCT, ή μια βάση μπορεί να προστεθεί στην αλληλουχία ενώ δεν υπήρχε (προσθήκη) πχ ATCGCT ATCGCTG. Μια μετάλλαξη η οποία οδηγεί σε αντικατάσταση της μορφής: A G, ή T C, ονομάζεται Μετάβαση (transition), ενώ μια μετάλλαξη που οδηγεί σε αντικατάσταση της μορφής: A T, A C, T G, ή G C, ονομάζεται Μεταστροφή-Μετάπτωση (transversion). Όπως και να έχει, αφού γνωρίζουμε ότι

συμβαίνουν τέτοιες αλλαγές, αν γνωρίζουμε με ποιον τρόπο, με ποια διαδοχικότητα και πόσο γρήγορα συμβαίνουν, τότε μπορούμε να συμπεράνουμε αρκετά πράγματα για τη γενετική απόσταση δύο οργανισμών.

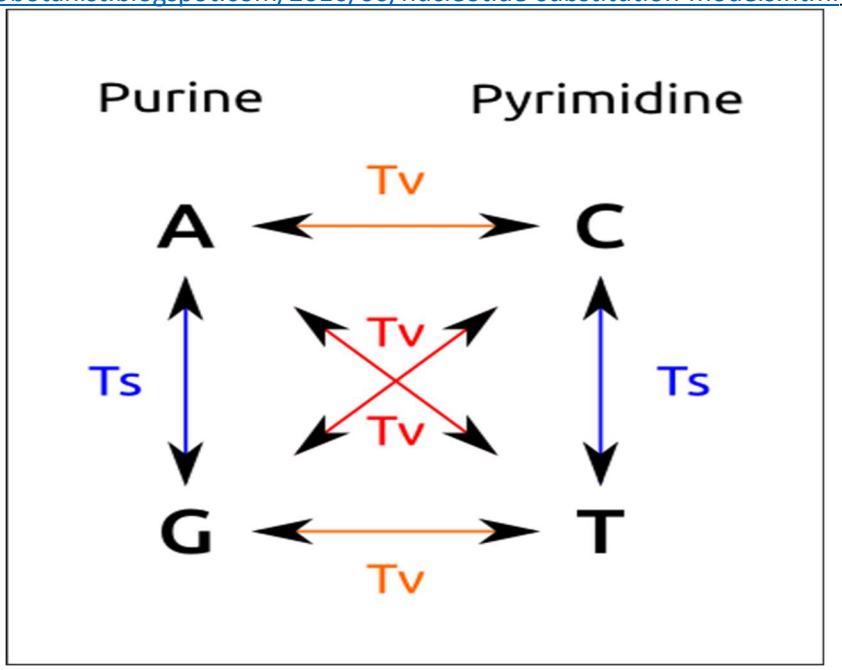
- Εικόνα 8 Ρυθμός υποκατάστασης **[Πηγή:** Xia X, 2018]



Εικόνα 9 Νουκλεοτιδικές Υποκαταστάσεις **[Πηγή:** Xia X, 2018]



Εικόνα 10 Μετάβαση και Μετάπτωση [Πηγή: http://phylobotanist.blogspot.com/2016/06/nucleotide-substitution-models.html]



2.4 Τεχνολογίες Ιικής Αλληλούχισης

Αφού απομονωθεί ο HIV από το αίμα ή άλλα σωματικά υγρά και στη συνέχεια ενισχύεται μέσω μίας μόνο σειράς αλυσιδωτών αντιδράσεων πολυμεράσης, μπορεί να ληφθεί μια ιική αλληλουχία από το προκύπτον ιικό αντίγραφο. Οι μεθοδολογίες προσδιορισμού αλληλουχίας του HIV περιλαμβάνουν άμεση Sanger αλληλούχιση (Sanger F, Nicklen S, Coulson AR, 1977), ενίσχυση ή κλωνοποίηση απλού γονιδιώματος (SGA/κλωνοποίηση) (Salazar-Gonzalez JF et al., 2008) και επόμενης γενεάς ακολουθία (NGS) (Metzker ML, 2010). Η υπεροχή των ιστορικών αλληλουχιών που σήμερα είναι διαθέσιμες σε μεγάλες βάσεις δεδομένων αλληλουχιών HIV (δηλ. Βάση δεδομένων HIV του Los Alamos) δημιουργήθηκαν χρησιμοποιώντας την αλληλούχιση Sanger μίας μόνο περιοχής γονιδίου. Τα αποτελέσματα αλληλουχίας Sanger είναι μία απλή συναινετική αλληλουχία ή αλληλουχία χύμα (“bulk”) από το ιικό αντίπυλο PCR (post-PCR viral amplicon), όπου κάθε βάση στην αλληλουχία είναι η πιο συχνή βάση σε αυτή τη θέση σε όλες τις μοναδικές αλληλουχίες εντός του αμπλικονίου (“amplicon”) (Sanger F, Nicklen S, Coulson AR, 1977). Αντίθετα, η SGA/κλωνοποίηση και η NGS δίνουν πολλαπλές διακριτές αλληλουχίες ιού, που επιτρέπουν σε κάποιον να ποσοτικοποιήσει και να εξετάσει την ιική ποικιλότητα εντός των δειγμάτων. Το επίπεδο ιικής ποικιλότητας που μπορεί να μετρηθεί σχετίζεται άμεσα με το γενετικό μήκος και το βάθος της αλληλουχίας που λαμβάνεται (Redd AD et al., 2012). Στην περίπτωση της SGA/κλωνοποίησης αυτό γίνεται συνήθως σε επίπεδο μερικών έως και πολλαπλών δεκάδων μακρύτερων αλληλουχιών από ένα δείγμα. Οι στρατηγικές της NGS Αλληλούχισης είναι σημαντικά πιο ισχυρές σε σύγκριση με την SGA και μπορούν να παράγουν δεκάδες χιλιάδες μικρότερες αλληλουχίες από ένα μόνο δείγμα. Εντούτοις, ορισμένες τεχνολογίες NGS μπορεί να είναι σφάλματα στο εσωτερικό των ομοιοπολυμερών τμημάτων DNA (Margulies M et al., 2005) και μία από τις μεγαλύτερες προκλήσεις στη χρήση της NGS ήταν να διακρίνει συστηματικό και τυχαίο τεχνικό σφάλμα από την πραγματική ιική ποικιλομορφία (Zagordi O

et al., 2010a; Beerenwinkel N, Zagordi O, 2011; Zagordi O et al., 2010b). Λόγω αυτών των εγγενών περιορισμών, καθώς και των διαφορών στο κόστος και την ευκολία της ανάλυσης των δεδομένων, οι στρατηγικές αυτές έχουν διαφορετικές εφαρμογές.

Πίνακας 2 Τεχνολογίες Ιικής Αλληλούχισης

Τεχνολογία	Αρ. Ακολουθιών	Πλεονεκτήματα	Περιορισμοί	Χρησιμότητα
Sanger 'bulk'	Μία	•Χαμηλή εργασία •Ταχεία ανάκαμψη • Εύκολη ανάλυση • Μεγάλο ιστορικό αρχείο • Χαμηλότερο συνολικό κόστος ανά δείγμα	• Δεν είναι δυνατή η ακριβή ανίχνευση πολλαπλών μολύνσεων • Μεσαίου μήκους Ακολουθία • Περιορισμένη ικανότητα λήψης intra-host ιικής ποικιλομορφίας	•Κλινικός Έλεγχος αντίστασης • Μεγάλες επιδημιολογικές μελέτες που χρησιμοποιούν φυλογενετική • Σάρωση υποτύπων • Αρχική ανάλυση σύνδεσης μεταξύ ατόμων με γνωστή ή υποψία σύνδεσης
•Μονή γονιδιακή αλληλούχιση (Single genome sequencing) (SGS) / Κλωνοποίηση (Cloning)	2-50 +	• Τεχνικά απλή • Ακριβότερες κλήσεις βάσης • Μεγάλη διάρκεια αλληλουχίας • Μέτριο κόστος σε χαμηλού αριθμού αντιγράφων ή σε αριθμό δείγματος	• Ιογενης ποικιλομορφία / Η ικανότητα να ανιχνεύει πολλαπλές λοιμώξεις εξαρτάται από τον αριθμό των κλώνων που αναλύθηκαν • Ένταση εργασίας και υψηλότερο κόστος σε υψηλού αριθμού αντιγράφων ή αριθμού δείγματος	• Μελέτες ιογενής ποικιλομορφίας των μεγάλων γενετικών περιοχών • Πλήρης αλληλούχιση του ιικού γονιδιώματος • Μελέτες ανασυνδυασμού
Επόμενη γενεά αλληλουχιών (NGS) Next generation sequencing	50000000-25000000000 εκατομμύρια + §	• Ακριβέστερα συλλαμβάνει intra-host ιική ποικιλομορφία, συμπεριλαμβανομένων των πολλαπλών λοιμώξεων • Υψηλής απόδοσης • Το χαμηλότερο κόστος ανά αλληλούχιση νουκλεοτιδίου βάσης	• Πολύ ευαίσθητο σε μόλυνση • Απαιτεί καθαρισμό εξειδικευμένων δεδομένων και ανάλυση • ακριβό εκ των προτέρων εξοπλισμό και το κόστος αντιδραστηρίου • Περιορισμένη πρόσβαση παγκοσμίως	• υπερμόλυνση HIV και διπλής μόλυνσης μελέτες •HIV μελέτες υπερμόλυνσης και μελέτες διπλής μόλυνσης •Επιδημιολογικές μελέτες των πληθυσμών σε υψηλό κίνδυνο για πολλαπλές λοιμώξεις (IDUs, πληθυσμούς υψηλής επίπτωσης) • Ανίχνευση παραλλαγών ήσσονος σημασίας

Το γονίδιο pol του γονιδιώματος HIV-1, ο στόχος ("target") της δοκιμής αντοχής στα φάρμακα HIV, χρησιμοποιείται πιο συχνά στην κατασκευή μοριακών συστάδων. Προτείνεται σε όλους τους ασθενείς να υποβληθούν σε δοκιμή αντοχής στα φάρμακα HIV πριν από την ART στις ανεπτυγμένες χώρες. Όταν οι ασθενείς παρουσιάζουν ιολογική ανεπάρκεια κατά τη διάρκεια της θεραπείας με ART σε αναπτυσσόμενες χώρες, τεράστιες ποσότητες

σχετιζόμενων δεδομένων μπορούν να ληφθούν χωρίς επιπλέον έξοδα. Ωστόσο, το γονίδιο pol θεωρείται λιγότερο ενημερωτικό και έχει σχετικά χαμηλό ποσοστό υποκατάστασης στο γονιδίωμα HIV [Hué S et al, 2004]. Ολόκληρη η αλληλουχία γονιδιώματος ή η env αλληλουχία γονιδίου του HIV-1 θεωρείται ότι αντικατοπτρίζει καλύτερα την πραγματική σχέση μετάδοσης [Trask SA et al., 2002; Yebra G et al., 2016]. Σε μια αλυσίδα μετάδοσης του HIV-1 που αποτελείται από εννέα ασθενείς, το εξελικτικό ιστορικό που συνήχθη από το φυλογενετικό δέντρο με τις αλληλουχίες γονιδίου pol δεν ήταν πλήρως συμβατό με το γνωστό ιστορικό μετάδοσης και οι ανθεκτικοί στα πολλαπλά φάρμακα ιοί συσταδοποιήθηκαν λανθασμένα. Αντιθέτως, το env φυλογενετικό δέντρο ήταν πλήρως συμβατό με το γνωστό ιστορικό μετάδοσης [Yerly S et al, 2001]. Ωστόσο, η χρήση ολόκληρης της αλληλουχίας γονιδιώματος ή της env αλληλουχίας γονιδίου δεν ισχύει για την πρακτική δημόσιας υγείας λόγω των αυστηρών τεχνικών απαιτήσεων, του υψηλού κόστους και των πολυμορφισμών μεγάλου μήκους.

Μια πτυχή της ιολογιας του HIV που είναι ιδιαίτερα σημαντική για να ληφθεί υπόψη στις μελέτες αλληλούχισης των συγκεντρωμένων επιδημιών του HIV, όπου ο κύκλος των συντρόφων (“partner turnover”) είναι πιο συχνός και ως εκ τούτου ο αριθμός των συντρόφων κατά την διάρκεια της ζωής είναι υψηλότερος, είναι η ιική συν-μόλυνση (“coinfection”) ή η υπερμόλυνση (“superinfection”) με πολλαπλούς γενετικά διακριτούς ιούς HIV (Redd and Tobian, 2013). Όταν τα άτομα έχουν μολυνθεί με πολλαπλούς ιούς είτε κατά την αρχική μόλυνση είτε αργότερα, μια bulk αλληλουχία δεν θα αντιπροσωπεύει επαρκώς όλους τους ιούς που ένα άτομο έχει αποκτήσει ή ενδεχομένως μεταδώσει. Σε αυτές τις περιπτώσεις, οι μέθοδοι SGA και NGS μπορούν να παρέχουν μια λεπτομερέστερη απεικόνιση της ποικιλομορφίας του HIV εντός του ξενιστή. Οι φυλογένειες που ανασυντάσσονται από bulk αλληλουχίες μπορεί να υποβάλλονται σε μεγαλύτερη μεροληψία με αυξανόμενο επιπολασμό της συν-μόλυνσης από τον ιό HIV σε έναν πληθυσμό, αν και αυτό παραμένει μια μελετημένη περιοχή στη φυλογενετική.

2.5 Στοίχιση

Όπως είπαμε και παραπάνω, στη φυλογενετική ανάλυση, προσπαθούμε να εκτιμήσουμε τις εξελικτικές σχέσεις μεταξύ των διαφόρων οργανισμών. Προκειμένου να συμβεί αυτό, λαμβάνουμε υπόψη το γενετικό υλικό του οργανισμού, ή, τις πρωτεϊνικές αλληλουχίες που παράγονται απ’ αυτό.

Για να μπορέσει να γίνει όμως μία τέτοια ανάλυση, αρχικά θα πρέπει να επιλέξουμε το αντίστοιχο τμήμα του γενετικού υλικού των οργανισμών που θα μελετήσουμε, δηλαδή να επιλέξουμε ομόλογες περιοχές για να συγκρίνουμε. Αυτό θα μπορούσε να σημαίνει να πάρουμε το αντίστοιχο γονίδιο από οργανισμούς του ίδιου είδους, αλλά και αυτό δεν είναι αρκετό. Στην πραγματικότητα, θα πρέπει να πραγματοποιήσουμε μια διαδικασία στις αλληλουχίες που έχουμε στο δείγμα μας, η οποία καλείται «στοίχιση» και κατά την οποία οι ομόλογες περιοχές δύο ή περισσότερων αλληλουχιών «στοιχίζονται» μαζί (η μία κάτω από την άλλη).

Γενικά, για την πραγματοποίηση οποιασδήποτε φυλογενετικής ανάλυσης, απαιτείται πρώτα η στοίχιση των αλληλουχιών. Η στοίχιση χωρίζεται στις εξής κατηγορίες 1)στοίχιση ζευγών αλληλουχιών (pairwise alignment) (σπάνια χρήση) και 2) στοίχιση πολλαπλών αλληλουχιών.

Η διαδικασία αυτή γίνεται σήμερα με τη βοήθεια Η/Υ, ενώ έχουν αναπτυχθεί στην πορεία του χρόνου διάφορα μοντέλα για αυτό τον σκοπό, τα οποία βασίζονται σε ένα ποσοτικό βαθμό (score) βέλτιστης στοίχισης. Ονομαστικά αναφέρουμε ορισμένα, όπως το κριτήριο μέγιστης φειδωλότητας, το διάγραμμα κουκίδων, και την εισαγωγή μονάδων κόστους, για κάθε κενό που απαιτείται να προσθέσουμε σε μία αλληλουχία ώστε να μπορέσει να στοιχηθεί με τις άλλες. Οι αλγόριθμοι στοίχισης πολλαπλών αλληλουχιών που χρησιμοποιούνται έχουν ως βάση κυρίως το δυναμικό προγραμματισμό (dynamic

programming), γενετικούς αλγορίθμους (genetic algorithms), κρυμμένα Μαρκοβιανά Μοντέλα (Hidden Markov Models HMMs) ή προοδευτικούς αλγορίθμους (progressive algorithms). Ο αλγόριθμος προοδευτικής στοίχισης είναι η πιο ευρέως διαδεδομένη μέθοδος στοίχισης.

Γενικά είναι δυνατόν να στοιχίσουμε οποιεσδήποτε αλληλουχίες DNA αφού εισάγουμε κενά (gaps) και αντικαταστάσεις (substitutions) σε διαφορετικά σημεία της συστοιχίας (alignment). Μετρώντας το σύνολο των παραπάνω (κενών και αντικαταστάσεων) είναι δυνατόν να υπολογίσουμε ένα μέτρο κόστους μιας συγκεκριμένης στοίχισης.

Η βασική λειτουργία των αλγορίθμων στοίχισης είναι να βρουν την συστοιχία με το μικρότερο συνολικό κόστος. Αυτό μεταφράζεται στην εύρεση της καλύτερης δυνατής συστοιχίας με το μικρότερο συνολικό κόστος.

2.6 Εξελικτικά μοντέλα

Το πρώτο βήμα στην ανάλυση των ευθυγραμμισμένων αλληλουχιών είναι η εκτίμηση της γενετικής ή εξελικτικής απόστασης μεταξύ των αλληλουχιών. Είναι ένα μέτρο του πόσο διαφορετικές είναι οι αλληλουχίες και εκφράζει τον αριθμό των εξελικτικών αλλαγών που έχουν συμβεί από τη στιγμή της απόκλισης τους.

Προκειμένου να πραγματοποιήσουμε φυλογενετική ανάλυση, θα πρέπει να καθορίσουμε με κάποιο μαθηματικό μοντέλο, τον τρόπο με τον οποίο υποθέτουμε πως πραγματοποιούνται οι διάφορες μεταλλάξεις στην αλυσίδα του DNA. Έτσι, με τη βοήθεια ενός τέτοιου μοντέλου και την υπόθεση πως όταν συγκρίνουμε δύο αλυσίδες τότε είτε η μία έχει προκύψει από την άλλη μετά από μια σειρά μεταλλάξεων, ή οι δυο υπό μελέτη αλληλουχίες μοιράζονται κάποιο κοινό πρόγονο, μπορούμε να εκτιμήσουμε την εξελικτική απόσταση που χωρίζει τις δύο αυτές αλληλουχίες, ή κατ' επέκταση τις ανά δύο αποστάσεις ενός συνόλου αλληλουχιών.

Η πιο κοινή και απλούστερη μέτρηση της εξελικτικής απόστασης είναι η εκτίμηση του αριθμού των θέσεων που διαφέρουν μεταξύ 2 αλληλουχιών δια του συνολικού μήκους της αλληλουχίας (p-distances). Ωστόσο αυτή η μέτρηση υστερεί σε πολλά σημεία, π.χ. εάν ο ρυθμός αλλαγών (υποκατάστασης) είναι υψηλός, μπορεί να έχουμε υποεκτίμηση της πραγματικής γενετικής απόστασης (διαδοχικές, παράλληλες, ανάστροφες, πολλαπλές, τυχαίες, συγκλίνουσες υποκαταστάσεις) που έχουν συμβεί. Δηλαδή, οι απλές p-αποστάσεις δεν χρησιμοποιούν ένα μοντέλο υποκατάστασης που περιγράφει την εξελικτική διαδικασία και ως εκ τούτου δεν λαμβάνουν υπόψη πολλαπλές αλλαγές ή μεταλλάξεις στο ίδιο σημείο. Κατά συνέπεια, συνήθως υποτιμούν την πραγματική γενετική απόσταση μεταξύ δύο αλληλουχιών.

Δεδομένου ότι η απόσταση p μπορεί να υποεκτιμήσει την πραγματική ποσότητα της εξελικτικής αλλαγής, έχει γίνει μια μεγάλη προσπάθεια ανεύρεσης μοντέλων που μετατρέπουν την παρατηρούμενη απόσταση σε πραγματική εξελικτική απόσταση. Για την διόρθωση αυτών των κρυμμένων αλλαγών πραγματοποιείται χρήση ενός στατιστικού μοντέλου για το πώς συμβαίνουν οι σημειακές μεταλλαγές, ώστε να είναι εφικτή η εκτίμηση της πραγματικής γενετικής απόστασης από την παρατηρούμενη απόσταση.

Τα μοντέλα αυτά ονομάζονται μοντέλα εξέλιξης ή μέθοδοι διόρθωσης αποστάσεων ή μοντέλα νουκλεοτιδικής υποκατάστασης

• Τα μοντέλα μας βοηθούν να παρεμβληθούμε ανάμεσα στις παρατηρήσεις μας ώστε να κάνουμε προβλέψεις

• Η προσθήκη παραμέτρων σε ένα μοντέλο σε γενικές γραμμές αυξάνει τη προσαρμογή του μοντέλου στα παρατηρούμενα δεδομένα

• Η υποπαραμετροποίηση των μοντέλων οδηγεί σε κακή προσαρμογή

• Η υπερπαραμετροποίηση των μοντέλων οδηγεί σε υψηλή διασπορά

• Τα κριτήρια για την επιλογή των μοντέλων είναι πολλά και περιλαμβάνουν ελέγχους του λόγου πιθανοτήτων (likelihood ratio tests), AIC, BIC, Bayes Factors, κλπ.

➢ Όλα αυτά παρέχουν έναν τρόπο να επιλεγεί ένα μοντέλο που δεν είναι ούτε ύπο- ούτε ύπερ-παραμετροποιημένο

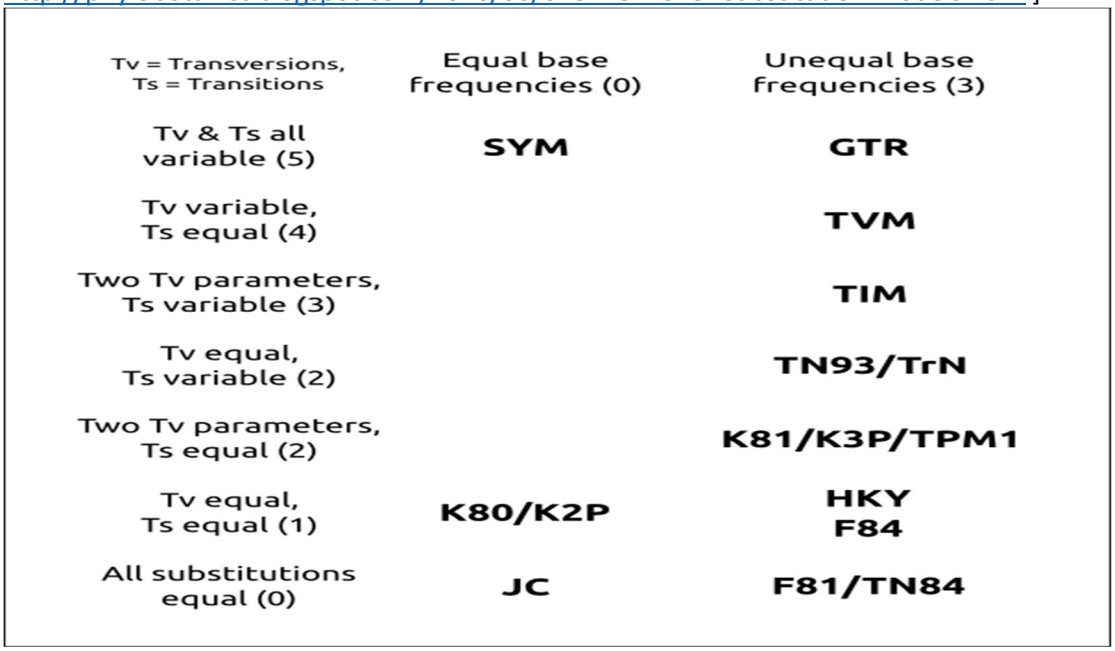
Ο αριθμός των υποκαταστάσεων = ρυθμός × χρόνο
Εάν ένα κλαδί είναι σχετικά μακρύ τότε δύο φαινόμενα μπορεί να ισχύουν:
• Ο ρυθμός υποκατάστασης είναι υψηλός
• Η γενεαλογική γραμμή υπάρχει για ένα μεγάλο χρονικό διάστημα

➢ Κάθε μήκος κλαδιού μπορεί να είναι και μια παράμετρος σε ένα μοντέλο

Τα μοντέλα νουκλεοτιδικής αντικατάστασης όπως τα HKY, F81, TN, K2, JC, GTR και άλλα, με ή χωρίς Γ-κατανομής ρυθμό μεταβολών μεταξύ των περιοχών (Γ-distributed rate variation across sites) και συνυπολογίζοντας ή όχι τις αναλογίες των αμετάβλητων περιοχών (the proportion of invariant sites - I), έχουν ένα εξέχοντα ρόλο - και μάλλον σημαντικότερο ακόμα και από τις ίδιες τις μεθοδολογίες ανάλυσης - για την κατά το δυνατόν ορθότερη φυλογενετική συμπερασματολογία. Τα πιο διαδεδομένα εξελικτικά μοντέλα είναι:

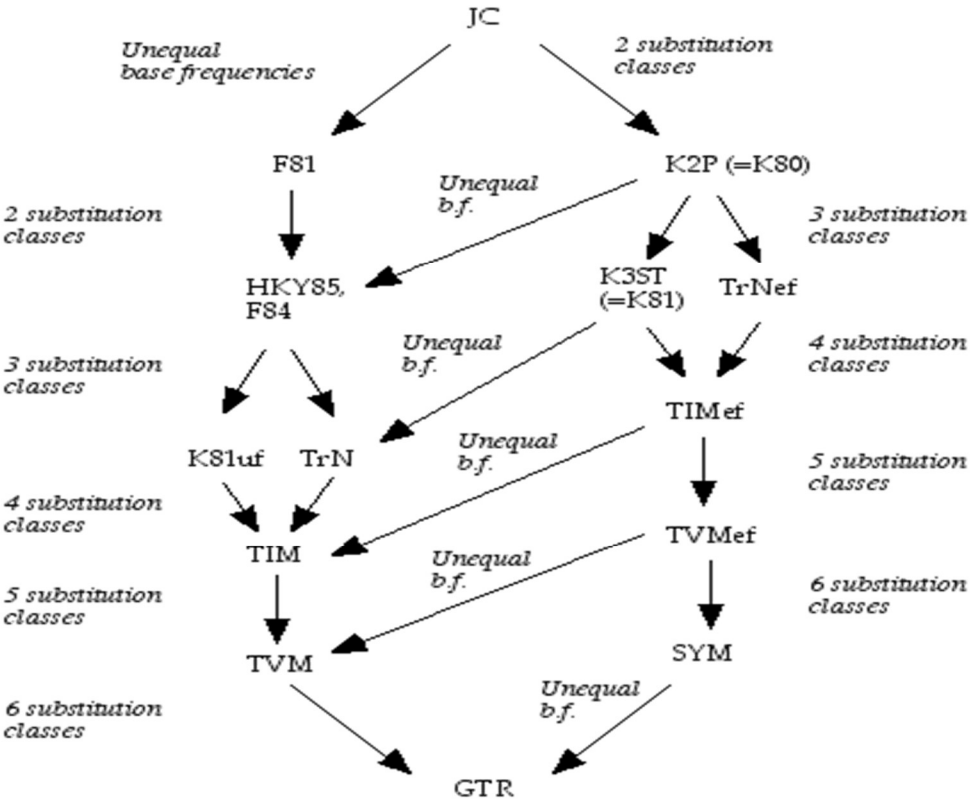
- Το μοντέλο **JC69** (Jukes & Cantor, 1969)
- Το μοντέλο K2P ή **K80** (Kimura, 1980)
- Το μοντέλο **F81** (Felsenstein, 1981)
- Το μοντέλο **HKY85** (Hasegawa, Kishino, & Yano, 1985)
- Το μοντέλο **TN93** (Tamura and Nei,1992)

Εικόνα 11 Μοντέλα Νουκλεοτιδικής Υποκατάστασης (Εξελικτικά Μοντέλα) [Πηγή: http://phylobotanist.blogspot.com/2016/06/overview-over-substitution-models.html]



Εικόνα 12 Μοντέλα Νουκλεοτιδικής Υποκατάστασης (Εξελικτικά Μοντέλα)

[Πηγή: http://carrot.mcb.uconn.edu/~olgazh/bioinf2010/class24.html]



1. Το μοντέλο **JC69** (Jukes & Cantor, 1969)

Αποτελεί το πιο απλό εξελικτικό μοντέλο για το DNA. Το μοντέλο αυτό υποθέτει πως το πλήθος των τεσσάρων βάσεων σε ένα μόριο DNA είναι σε ισορροπία, έχουν την ίδια συχνότητα ενώ οι μεταβάσεις και μεταστροφές δηλαδή οι αντικαταστάσεις κάθε τύπου θεωρούνται ισοπίθανες.

Θεωρήσεις:

- οι βάσεις σε ίση συχνότητα (όλες 0,25 ή 1/4)
- ο ρυθμός υποκατάστασης είναι ο ίδιος (μ) για όλες τις πιθανές υποκαταστάσεις
- 1 παράμετρος μ

➢ Οι αποστάσεις με βάση αυτό το μοντέλο υπολογίζονται με τον μαθηματικό τύπο

$$d = -\left({}^{3}/_{4}\right) \times \ln\left(1 - \left({}^{4}/_{3}\right) \times p\right)$$

2. Το μοντέλο K2P ή **K80** (Kimura, 1980)

Το μοντέλο αυτό επιτρέπει την ύπαρξη διαφορετικού ρυθμού εμφάνισης μεταβάσεων απ' ότι μεταστροφών, αν και παραμένει συμμετρικό. Θεωρούνται ίσες οι συχνότητες εμφάνισης των αζωτούχων βάσεων.

Θεωρήσεις:

- οι βάσεις σε ίση συχνότητα (όλες 0,25 ή 1/4)
- δύο ρυθμοί υποκατάστασης: (a = c = d = f , kappa = b = e)
- ανεξάρτητοι ρυθμοί για μεταπτώσεις και μεταστροφές

Υπάρχουν

- 4 πιθανές μεταπτώσεις (Ti): A→G, C→T, G→A, T→C!

- 8 πιθανές μεταστροφές (Tv): A→C, A→T, C→A, C→G, G→C, G→T, T→A, T→G

➢ Η πιθανότητα για μία μετάπτωση (π.χ. A G)

$$Pr(AG) = Pr(startwithA) \times Pr(changetoG) = {}^{1}/_{4} \times (\kappa\mu\, dt)$$

¼ η συχνότητα της A, κμ ο ρυθμός μεταπτώσεων, dt ο χρόνος

➢ Αν μ είναι ο ρυθμός των μεταστροφών και κμ ο ρυθμός των μεταπτώσεων , τότε ο ρυθμός της αναλογίας μεταπτώσεων/μεταστροφών είναι: $\kappa\mu/\mu = \kappa$

$$\left({}^{T_i}/_{T_v}\right) ratio = \frac{Pr(\kappa\acute{a}\theta\varepsilon\ \mu\varepsilon\tau\acute{a}\pi\tau\omega\sigma\eta)}{Pr(\kappa\acute{a}\theta\varepsilon\ \mu\varepsilon\tau\alpha\sigma\tau\rho o\varphi\acute{\eta})} = \frac{4 \times ({}^{1}/_{4}) \times \kappa\mu\, dt}{8 \times ({}^{1}/_{4}) \times \mu\, dt} = \frac{\kappa}{2}$$

➢ Στο μοντέλο K80 η αναλογία Ti/Tv είναι η μισή σε σχέση με το ρυθμό της αναλογίας μεταπτώσεων/μεταστροφών επειδή ο αριθμός των μεταστροφών είναι διπλάσιος των μεταπτώσεων.

➢ Ίδιο με JC69 εάν κ = 1

3. Το μοντέλο **F81** (Felsenstein, 1981)

Στο μοντέλο αυτό πέραν του ότι επιτρέπεται ο ρυθμός μεταβάσεων και μεταστροφών να διαφέρουν μεταξύ τους, επιτρέπεται οι δύο αυτοί ρυθμοί να μην είναι σταθεροί αλλά να μπορούν να διαφέρουν μεταξύ διαφορετικών ειδών οργανισμών.

οι συχνότητες των αζωτούχων βάσεων διαφέρουν, ενώ οι αντικαταστάσεις κάθε τύπου θεωρούνται ισοπίθανες

Θεωρήσεις:

- οι βάσεις σε διαφορετικές συχνότητες (π_A, π_C, π_G, μ)
- ίδιος ρυθμός υποκατάστασης: (a = c = d = f = b = e)
- 4 παράμετροι (π_A, π_C, π_G, μ)

➢ Ίδιο με το μοντέλο JC69 εάν οι συχνότητες των βάσεων είναι ίσες (¼)

4. Το μοντέλο **HKY85** (Hasegawa, Kishino, & Yano, 1985)

Αποτελεί είναι μοντέλο το οποίο συνδυάζει χαρακτηριστικά των K80 και F81. Συγκεκριμένα, το μοντέλο υποθέτει πως ο ρυθμός μεταβάσεων και μεταστροφών διαφέρει ανά νουκλεοτίδιο. Ο ρυθμός αφορά την πιθανότητα με την οποία ένα νουκλεοτίδιο μπορεί να μετατραπεί σε κάποιο άλλο συγκεκριμένο νουκλεοτίδιο. Παράδειγμα, η τιμή του ρυθμού για την Θυμίνη (T), αφορά την πιθανότητα με την οποία ένα νουκλεοτίδιο μπορεί να μεταλλαχθεί σε T. Επομένως οι συχνότητες των αζωτούχων βάσεων διαφέρουν, ενώ δίδεται ένας ρυθμός μεταπτώσεων (transitions) και ένας ρυθμός μεταστροφών (transversions) για τις αντικαταστάσεις

Θεωρήσεις:

- οι βάσεις σε διαφορετικές συχνότητες (π_A, π_C, π_G, μ)
- δύο ρυθμοί υποκατάστασης: (a = c = d = f , kappa = b = e)
- 5 παράμετροι (π_A, π_C, π_G, μ, k)

➢ Ίδιο με το μοντέλο F81 εάν κ=1 και ίδιο με JC69 εάν κ=1 και οι συχνότητες των βάσεων είναι ίσες (¼)

5. Το μοντέλο **TN93** (Tamura and Nei,1992)

Σε συνέχεια των προηγούμενων μοντέλων, οι συχνότητες των αζωτούχων βάσεων διαφέρουν και εδώ πέραν των διαφορετικών ρυθμών μεταβάσεων και μεταστροφών, δίνεται η δυνατότητα ο ρυθμός μεταβάσεων να είναι διαφορετικός για το ζεύγος (A,G) από τον ρυθμό για το ζεύγος (T,C) και το αντίστοιχο συμβαίνει και με τον ρυθμό μεταστροφών (transversions) για τα ζεύγη (A,T) και (G,C).

6. Το μοντέλο **GTR** (Generalised time-reversible) (Tavaré, 1986)

Το τελευταίο είναι αυτό με τις πιο πολλές παραμέτρους. Επικεντρώνεται στις αντικαταστάσεις, δέχεται πως αυτές συμβαίνουν με διαφορετικούς ρυθμούς. Συνεπώς, ο ρυθμός για κάθε δυνατή μεταστροφή ή μετάβαση (από και προς οποιαδήποτε βάση) μπορεί να είναι διαφορετικός. Τέλος, το ποσοστό κάθε βάσης μπορεί να διαφέρει μεταξύ των αλληλουχιών. Τα χαρακτηριστικά αυτά μετατρέπουν το μοντέλο GTR σε ένα αρκετά ευέλικτο και ρεαλιστικό μοντέλο για την όσο το δυνατό καλύτερη περιγραφή της όλης διαδικασίας με την οποία συμβαίνουν οι αντικαταστάσεις των βάσεων στις αλληλουχίες DNA ενός δείγματος. Επομένως οι συχνότητες των αζωτούχων βάσεων διαφέρουν, ενώ υπάρχει συμμετρική μήτρα αντικαταστάσεων (διαφορετικός ρυθμός μεταστροφών και μεταπτώσεων για κάθε μία από τις τέσσερεις πιθανές περιπτώσεις).

Θεωρήσεις:

- οι βάσεις σε διαφορετικές συχνότητες (π_A, π_C, π_G, μ)
- διαφορετικοί ρυθμοί υποκατάστασης: (a, b, c, d, e)
- 9 παράμετροι (π_A, π_C, π_G, μ, a, b, c, d, e)

- Ίδιο με το JC69 εάν a = b = c = d = e = f = 1 και όλες οι νουκλεοτιδικές συχνότητες είναι ίσες (1/4).

- Ετερογένεια στα δεδομένα μας :
  - Ρυθμός ετερογένειας (r)

  • Διαφοροποίηση (ποικιλία) στο ρυθμό υποκατάστασης μεταξύ των θέσεων ενός ευθυγραμμισμένου συνόλου αλληλουχιών DNA
  - Αιτίες

  • η συστηματική μεροληψία στο ρυθμό μεταλλαγής
  • διαφορές στη λειτουργικότητα τμημάτων των αλληλουχιών

- Έτσι λοιπόν, επιπρόσθετα του παραμέτρων του ρυθμού υποκατάστασης και των συχνοτήτων των βάσεων που είδαμε έως τώρα, τα περισσότερα σύγχρονα προγράμματα κατασκευής δέντρων επιτρέπουν στον χρήστη να προσθέσει περισσότερο ρεαλισμό στο μοντέλο που επιλέγει, συμπεριλαμβάνοντας επιπλέον παραμέτρους, όπως η ετερογένεια του ρυθμού υποκατάστασης κατά μήκος των ευθυγραμμισμένων αλληλουχιών.
- Μια συνεχής γάμμα κατανομή χρησιμοποιείται για να μοντελοποιηθεί αυτή η ετερογένεια στη φυλογένεση (Yang, 1996). Μια απλή παράμετρος (alpha, α ή shape parameter) ελέγχει τη μορφή της γάμμα κατανομής.
  - Όταν το α < 1 υπάρχει σημαντική ποικιλομορφία στο ρυθμό.
  - Όσο μεγαλύτερο είναι το α τόσο μικρότερη είναι η ετερογένεια.
- Μερικές φορές χρησιμοποιείται και μια άλλη παράμετρος (proportion of invariant sites, I) για να μοντελοποιήσει την ετερογένεια.

Έτσι το πιο πολύπλοκο μοντέλο είναι GTR+I+G, με 10 ελεύθερες παραμέτρους. Γενικώς, το μοντέλο GTR, είναι το πιο γενικό, ουδέτερο, ανεξάρτητο και χρονικά αναστρέψιμο κατά το δυνατό μοντέλο. Όσον αφορά τις παραμέτρους G και I, όταν χρησιμοποιείται η πρώτη σημαίνει ότι οι μεταβολές μεταξύ των ποικίλων γενετικών περιοχών κατανέμονται με Γ-κατανομή, ενώ όταν εφαρμόζεται η δεύτερη παράμετρος ουσιαστικά δίδεται μία

ποσόστωση για τις περιοχές που θα θεωρηθούν αμετάβλητες. Οι Leitner et al. (1997) υποστηρίζουν ότι το μοντέλο GTR+Γ παρέχει την καλλίτερη περιγραφή της εξέλιξης του HIV-1, ενώ τα απλούστερα μοντέλα (π.χ. τα TN, HKY, F81, K2, JC) είναι μη ρεαλιστικά και ανακριβή. Γενικά, φαίνεται ότι η εκάστοτε μέθοδος χωρίς την Γ-κατανομή αποδίδει χειρότερα. Σε αυτό το σημείο φαίνεται, επίσης, ότι συνυπολογίζοντας και τις αναλογίες των αμετάβλητων περιοχών (GTR+Γ+I), το μοντέλο γίνεται ακόμη πιο ρεαλιστικό και αποδοτικό, προσεγγίζοντας ακόμα περισσότερο το φυσικό εξελικτικό μοντέλο.

- Φυσικά δεν είναι μόνο αυτά που τα μοντέλα των νουκλεοτιδικών υποκαταστάσεων. Ο αριθμός τους ανέρχεται σε 203.

Συνοψίζοντας, όλα τα μοντέλα αποτελούν μια ειδική περίπτωση του GTR. Αυτό σημαίνει ότι αν θέσουμε συγκεκριμένους περιορισμούς στις τιμές των παραμέτρων του GTR θα οδηγηθούμε σε μια ειδική περίπτωση του γενικευμένου μοντέλου (π.χ. το JC69 είναι μια περίπτωση του K2P ενώ το K2P είναι μια περίπτωση του TN93 το οποίο είναι περίπτωση του GTR).

Στη συνέχεια, με τη γνώση αυτή και ομαδοποιώντας τις αλληλουχίες σύμφωνα με την απόσταση που εκτιμήθηκε (οι αλληλουχίες που απέχουν εξελικτικά λιγότερο θα ομαδοποιούνται μαζί), μπορούμε να κατασκευάσουμε ένα γράφημα, το φυλογενετικό δέντρο, όπου το μήκος των κλαδιών του αντιστοιχεί στις εκτιμώμενες εξελικτικές αποστάσεις μεταξύ των αλληλουχιών.

2.7 Μεθοδοι Αποστασεων

Όπως αναφέρθηκε παραπάνω, στη φυλογενετική ανάλυση καλούμαστε να εκτιμήσουμε και στη συνέχεια να αναπαραστήσουμε γραφικά, με τη βοήθεια των φυλογενετικών δέντρων, τις εξελικτικές σχέσεις μεταξύ ενός συνόλου αλληλουχιών DNA ή άλλου τύπου γενετικού υλικού (πρωτεΐνες). Οι σχέσεις αυτές επί της ουσίας μας πληροφορούν για το πόσο μοιάζουν ή διαφέρουν εξελικτικά οι υπό μελέτη αλληλουχίες μεταξύ τους. Οι δύο βασικές κατηγορίες μεθοδολογικών προσεγγίσεων για την πραγματοποίηση μιας τέτοιας ανάλυσης είναι οι μέθοδοι αποστάσεων (distances methods) και οι μέθοδοι χαρακτήρων (character state) (Παρασκευής, Μαγιορκίνης, & Χατζάκης, 2015).

Η γενική ιδέα των μεθόδων αποστάσεων είναι αρχικά να υπολογίζονται όλες οι ανά δύο γενετικές αποστάσεις και στη συνέχεια να εκτιμάται ένα δέντρο το οποίο αναπαριστά τις αποστάσεις όσο το δυνατόν καλύτερα. Το μέγεθος των υπολογισθέντων αποστάσεων αντιστοιχούν στο μήκος των κλάδων που συνδέουν τις αντίστοιχες ανά δύο αλληλουχίες (Felsenstein, 2004). Οι γενετικές αποστάσεις που αντιστοιχούν στον παρατηρούμενο αριθμό μεταλλαγών αποτελούν εκτιμήσεις (υποεκτιμήσεις) της πραγματικής εξελικτικής απόστασης μεταξύ των αλληλουχιών και τις συναντάμε ως p-distance, ή αποκαλούμενη απόσταση p, ή Hamming απόσταση, που ισούται με τον παρατηρούμενο αριθμό διαφορών νουκλεοτιδίων μεταξύ δύο αλληλουχιών, ενώ για των υπολογισμό των εξελικτικών αποστάσεων (πραγματικού αριθμού μεταλλαγών) βασιζόμαστε σε κάποιο εξελικτικό μοντέλο-νουκλεοτιδικό μοντέλο υποκατάστασης, και αυτές τις αποστάσεις τις συναντάμε ως η αναμενόμενη γενετική απόσταση ή η λεγόμενη απόσταση d (Vandamme, 2009). Αν οι γενετικές αποστάσεις υπολογίζονται ως το άθροισμα του μήκους των κλαδιών ανάμεσα σε δύο άκρες σε ένα δέντρο, τότε είναι γνωστές ως πατερικές αποστάσεις .

Αν και θα περίμενε κανείς μια τέτοια μέθοδος να αφήνει ανεκμετάλλευτο ένα τόσο μεγάλο μέρος της πληροφορίας, (αφού τη συνοψίζει σε έναν απλό πίνακα αποστάσεων) ώστε να μη μπορεί να παραχθεί ένα δέντρο που να εκτιμά σωστά την πραγματική φυλογένεια των δεδομένων, έχει φανεί πως τελικά η απώλεια πληροφορίας είναι ιδιαίτερα μικρή (Felsenstein, 2004). Παρακάτω περιγράφονται ορισμένες από τις βασικότερες μεθόδους αποστάσεων.

2.7.1 Unweighted Pair-Group Method with Arithmetic mean UPGMA Αλγόριθμος Συνάθροισης Τοπική Μέθοδος Κατασκευής Φυλογενετικού Δέντρου

Η μέθοδος αυτή αποτελεί την απλούστερη των μεθόδων αποστάσεων. Αρχικά αναπτύχθηκε για κατασκευή φαινογραμμάτων (Cluster analysis) δηλαδή είναι μια μέθοδος ταξινομικών μονάδων-κλάσεων (clustering). Η UPGMA χρησιμοποιεί έναν επαναληπτικό αλγόριθμο, κατά τον οποίο αρχικά η κάθε αλληλουχία θεωρείται μια ταξινομική μονάδα (OTU). Αφού υπολογισθούν όλες οι ανά δύο αποστάσεις των υπό μελέτη αλληλουχιών, οι δύο κοντινότερες (εξελικτικά) αλληλουχίες τοποθετούνται σε μια κοινή ομάδα, η οποία πλέον θεωρείται ως μια νέα ταξινομική μονάδα και η διαδικασία επαναλαμβάνεται από την αρχή, μέχρι να απομείνουν δύο μόνο ταξινομικές μονάδες. Αξίζει να σημειωθεί πως η γενετική απόσταση μεταξύ δύο OTUs διαιρείται διά 2 και αυτό ισούται με το μήκος των κλάδων που τις συνδέουν (Van de Peer, 2009).

- Βασική προϋπόθεση: ο εξελικτικός ρυθμός είναι σταθερός σε όλες τις γενεαλογικές γραμμές ή ισοδύναμα οι εξελικτικές αποστάσεις είναι ευθέως ανάλογες με τους χρόνους απόκλισης
- Παράγονται υποχρεωτικά χωρίς κλίμακα (unscaled) EPPIZA δένδρα.

2.7.2 Neighbor-Joining NJ Αλγόριθμος Συνάθροισης Τοπική Μέθοδος Κατασκευής Φυλογενετικού Δέντρου

αναζητούν να κατασκευάσουν την τοπολογία η οποία αντιπροσωπεύει καλύτερα την μήτρα των αποστάσεων μεταξύ των ζευγών των ταξινομικών μονάδων. Έτσι, σύντομα, το εύρος των λύσεων γίνεται χαώδες με διαρκώς αυξανόμενο αριθμό αλληλουχιών Brocchieri (2001) Μια κλασική ευριστική τεχνική που χρησιμοποιείται για να υπεκεράσει αυτό το εμπόδιο είναι η μέθοδος αποστάσεων σύνδεσης γειτόνων (NJ) (Saitou and Nei, 1987) και για τον λόγο αυτό αποτελεί και τον κυριότερο εκπρόσωπο αυτής της ομάδας μεθόδων. Brocchieri (2001)

Η μέθοδος του πλησιέστερου γείτονα, όπως συχνά συναντάται σε ελληνικά βιβλία και σημειώσεις, αποτελεί μια άλλη μέθοδο αποστάσεων. Το μοντέλο αυτό λέγεται «προσθετικό» (additive), καθώς στο δέντρο που κατασκευάζεται βάσει αυτού, η γενετική απόσταση μεταξύ δύο OTUs αποτελεί το άθροισμα του μήκους των κλαδιών που τις συνδέουν. (Van de Peer, 2009). Η NJ έχει ως στόχο, να εκτιμήσει το δέντρο ελάχιστης εξέλιξης (minimum evolution), δηλαδή το δέντρο με το μικρότερο συνολικό μήκος (το μικρότερο άθροισμα των κλάδων). Αν και η NJ μοιάζει με την UPGMA, καθώς σχετίζεται και πάλι με OTUs, καθώς και πάλι χρησιμοποιεί τη βασική ιδέα του να τοποθετεί σε κάθε βήμα μαζί σε ένα OTU τις κοντινότερες αλληλουχίες (ή OTUs), χωρίς όμως να υποθέτει πως οι αλληλουχίες έχουν σταθερούς ρυθμούς εξέλιξης (clock-like behavior) (Van de Peer, 2009).

- Το δένδρο που παράγεται είναι άρριζο και συνήθως απαιτεί μια εξωομάδα για να βρεθεί η ρίζα.
- Η αρχή της μεθόδου στηρίζεται στην εύρεση των «γειτόνων» διαδοχικά ώστε να μειώνεται το συνολικό μήκος του δέντρου

Παράδειγμα: Έστω ο πίνακας αποστάσεων 5 OTUs (A–F)
Για κάθε OTU υπολογίζουμε τα μεγέθη:

- r_i: το άθροισμα των αποστάσεων της OTU i από όλες τις άλλες και
- $r_i/(n-2)$ όπου n ο αριθμός των OTUs

Εν συνεχεία υπολογίζουμε τις τροποποιημένες αποστάσεις (Dij) ως εξής:

- $D_{ij} = d_{ij} - {}^{r_i}/_{(n-2)} - {}^{r_j}/_{(n-2)}$

Η απόσταση των δύο OTUs από τον κόμβο υπολογίζεται ως εξής:

- $d_i - node = {}^{d_{ij}}/_2 + {}^{\frac{r_i - r_j}{n-2}}/_2$
- $d_j - node = {}^{d_{ij}}/_2 + {}^{\frac{r_j - r_i}{n-2}}/_2$

Καταρτίζεται νέος πίνακας αποστάσεων στον οποίο τα OTUs A και B εμφανίζονται ως ένα σύνθετο OTU, node1 (κόμβος-1) και ακολουθείται η ίδια διαδικασία. Οι νέες αποστάσεις των OTUs από τον κόμβο 1 υπολογίζονται από τη σχέση:

- $D_k - node_{ij} = {}^{d_{ik} + d_{jk} - d_{ij}}/_2$

Καταρτίζεται νέος πίνακας αποστάσεων στον οποίο τα OTUs C και node 1 εμφανίζονται ως ένα σύνθετο OTU (node 2) και ακολουθείται η ίδια διαδικασία.

- Η μικρότερη (πιο αρνητική) απόσταση υποδεικνύει τα δύο OTUs που ομαδοποιούνται πρώτα, μέσω ενός εσωτερικού κόμβου 1 (node1)

❖ Κάποιες γενικοποιήσεις της μεθόδου NJ έχουν προταθεί. Αυτές, κάνουν αναζητήσεις σε πολλαπλά μονοπάτια χαμηλού επιπέδου λαθών και προοδευτικά βάζουν σε συστάδες τις αλληλουχίες (Kumar, 1996; Pearson et al., 1999 ; Brocchieri 2001)

2.7.3 Minimum Evolution Κριτήριο Βελτιστοποίησης Καθολική Μέθοδος Κατασκευής Φυλογενετικού Δέντρου

❖ Βασίζεται στη θεώρηση ότι το δέντρο με το μικρότερο άθροισμα των μηκών των κλάδων επιλέγεται ως το αληθινό δέντρο
❖ Για όλα τα πιθανά εναλλακτικά δέντρα, γίνεται εκτίμηση των μηκών των κλάδων από τις εκτιμώμενες ανά ζεύγη αποστάσεις μεταξύ των taxa.
❖ Υπολογίζεται το άθροισμα (S) όλων των μηκών των κλάδων
❖ Το ελάχιστο κριτήριο επιλέγει το δέντρο με τη μικρότερη τιμή S
❖ Θα μπορούσε να χαρακτηριστεί ανάλογη της μέγιστης φειδωλότητας

2.7.4 Fitch – Margoliash algorithm (Least Squares) Κριτήριο Βελτιστοποίησης Καθολική Μέθοδος Κατασκευής Φυλογενετικού Δέντρου

❖ Παρόμοια της UPGMA, χωρίς τη θεώρηση ότι ο εξελικτικός ρυθμός είναι σταθερός σε όλες τις γενεαλογικές γραμμές
❖ Παράγει άρριζα δέντρα
❖ Η μέθοδος (LS) παίρνει τις ανά ζεύγη αποστάσεις ως δεδομένο και εκτιμά τα μήκη των κλάδων σε ένα δέντρο ταιριάζοντας αυτές τις αποστάσεις, ελαχιστοποιώντας το άθροισμα των τετραγώνων μεταξύ των δοσμένων και των αναμενόμενων αποστάσεων. Η αναμενόμενη απόσταση υπολογίζεται ως το άθροισμα των μηκών των κλάδων κατά μήκος ενός μονοπατιού που συνδέει 2 είδη. Το ελάχιστο άθροισμα των τετραγώνων των διαφορών στη συνέχεια μετρά την αρμοστικότητα του δέντρου στα δεδομένα (αποστάσεις) και χρησιμοποιείται ως η τιμή (score) για το δέντρο.

2.7.5 Συνοψίζοντας Γενική προσέγγιση

Μεθοδοι Αποστασεων (Unweighted Pair-Group Method with Arithmetic mean UPGMA, Neighbor-Joining NJ, Minimum Evolution ME)

1. Εκτιμούν τη φυλογένεση με τον υπολογισμό πινάκων αποστάσεων ανά ζεύγη OTUs
2. Οι φυλογενετικές σχέσεις εκτιμώνται βάσει των ανά ζεύγος αποστάσεων με διάφορους αλγόριθμους
3. Θεωρούν ότι η φυλογενετική απόκλιση είναι αντίστοιχη με την αλληλουχική διαφοροποίηση
4. Είναι διαθέσιμοι πολλοί διαφορετικοί τύποι αποστάσεων

Πλεονεκτήματα

- Είναι υπολογιστικά γρήγορη
- Για πολύ μεγάλα σετ δεδομένων είναι πιθανά η πιο ελεγχόμενη προσέγγιση
- Ορισμένα δεδομένα μπορούν να αναλυθούν μόνο με μεθόδους αποστάσεων
- Η προσέγγιση είναι αλγοριθμική
- Βασίζονται σε εξελικτικά μοντέλα που περιέχουν διορθώσεις για πολλαπλές αντικαταστάσεις
- Με ισχυρά δεδομένα και σωστά εκτιμημένες αποστάσεις, τα παραγόμενα δέντρα είναι ακριβή με τοπολογίες που προσεγγίζουν τις αντίστοιχες πιο πολύπλοκων διεργασιών

Μειονεκτήματα ΑΡΝΗΤΙΚΑ ΜΕΙΟΝΕΚΤΗΜΑΤΑ ΠΕΡΙΟΡΙΣΜΟΙ

- Χάνεται πληροφορία, αφού οι αλληλουχίες μετατρέπονται σε αποστάσεις
- Υπολογισμός 1 μόνο τοπολογίας
- Εξαρτώνται από το εξελικτικό μοντέλο
- Ελαχιστοποίηση της πληροφορίας των αλληλουχιών σε ένα και μόνο αριθμό, τη γενετική απόσταση
- αναζητούν να κατασκευάσουν την τοπολογία η οποία αντιπροσωπεύει καλύτερα την μήτρα των αποστάσεων μεταξύ των ζευγών των ταξινομικών μονάδων. Έτσι, σύντομα, το εύρος των λύσεων γίνεται χαώδες με διαρκώς αυξανόμενο αριθμό αλληλουχιών (Brocchieri, 2001)
- απαιτούν ένα εξαιρετικά ακριβές μέτρο εξελικτικών αποστάσεων μεταξύ των αλληλουχιών, πράγμα που συχνά δεν είναι εφικτό (Brocchieri, 2001)
- δεν παρέχει εγγυήσεις ότι θα βρει την γενικότερα καλύτερη λύση (Pearson et al., 1999 ; Brocchieri 2001)
- δεν εκμεταλλεύονται πλήρως τις πληροφορίες που εμπεριέχονται στις αλληλουχίες DNA (Brocchieri, 2001)

❖ Εάν οι αλληλουχίες είναι λάθος ευθυγραμμισμένες, η επιλογή μεθόδου κατασκευής δέντρου δεν έχει σημασία!

2.8. Αλγόριθμοι εύρεσης ιδεατών δέντρων

2.8.1 Ακριβείς αλγόριθμοι (Exact algorithms)

I. <u>Exhaustive (<11 taxa)</u>

Αποτίμηση όλων των δέντρων και εύρεση του πιο «καλού»

1) Ο αλγόριθμος ξεκινάει φτιάχνοντας ένα δέντρο με όλα τα taxa, το οποίο δεν είναι απαραίτητα και το βέλτιστο και στη συνέχεια συναρμολογεί ένα δέντρο προσθέτοντας ένα taxon κάθε φορά.

2) Αρχίζει από ένα δέντρο με 3 taxa.
3) Το 4ο taxon προστίθεται με την προσθήκη ενός νέου κλάδου στο μέσο κάθε προϋπάρχοντος κλάδου.
4) Εκτιμά το παραγόμενο δέντρο, βάσει κάποιου κριτηρίου (π.χ. μήκος).

Πιθανά δέντρα

- για έρριζα δένδρα (n≥2):
$$N_R = {(2n-3)!}\Big/{2^{n-2}(n-2)!}$$

- για άρριζα δένδρα (n≥3):
$$N_U = {(2n-5)!}\Big/{2^{n-3}(n-3)!}$$

- Το PAUP* (ένα από τα πιο γρήγορα μηχανήματα σε ένα σύστημα ΗΥ) μπορεί να εκτιμήσει 1.089.798 δέντρα / sec υπό το κριτήριο της φειδωλότητας με 10 αλληλουχίες μήκους 1.296 βάσεων. Αν ο αριθμός των OTUs=n=20 τότε ο συνολικός αριθμός των άρριζων δέντρων θα ήταν 221.643.095.476.699.771.875 και των έρριζων θα ήταν 8.200.794.532.637.891.559.375. Για τον συγκεκριμένο αριθμό των έρριζων δέντρων αυτή η ανάλυση θα τέλειωνε σήμερα αν κάποιος την είχε βάλει να τρέχει στο Τριαδικό, λίγο μετά την μεγάλη εξαφάνιση του Περμίου (238 εκ. χρόνια πριν)

II. Branch and Bound (BandB) (11<taxa<20)

- Εγγυάται την εύρεση του καλύτερου δέντρου, χωρίς να απαιτείται η αποτίμηση κάθε δέντρου
- Ο αλγόριθμος ξεκινάει φτιάχνοντας ένα δέντρο με όλα τα taxa, το οποίο δεν είναι απαραίτητα και το βέλτιστο και στη συνέχεια συναρμολογεί ένα δέντρο προσθέτοντας ένα taxon κάθε φορά. Εάν ένα δέντρο που προκύπτει με τη προσθήκη του νέου τάξου έχει μήκος που ξεπερνάει το τρέχων κατώτερο όριο (bound) του βέλτιστου δέντρου, τότε αυτό το μονοπάτι τερματίζεται και ο αλγόριθμος γυρίζει πίσω και πάει στο αμέσως επόμενο διαθέσιμο μονοπάτι. Όταν το ψάξιμο σε ένα μονοπάτι φτάσει στο τέλος του (έχουν προστεθεί όλα τα τάξα, το τελικό δέντρο είναι είτε το βέλτιστο και διατηρείται είτε ένα τοπικό βέλτιστο και απορρίπτεται. Όταν όλα τα μονοπάτια, ξεκινώντας από το αρχικό 3 τάξων δέντρο, ολοκληρωθούν ο αλγόριθμος ολοκληρώνεται και τότε σχεδόν όλα τα πιο φειδωλά δέντρα θα έχουν ανακτηθεί. Η μέθοδος είναι υπολογιστικά εφικτή για αναλύσεις μέχρι 20 taxa που έχουν ~8.2*1021 δέντρα

2.8.2 Ευρετικοί αλγόριθμοι (Heuristic algorithms)

- Όταν ο αριθμός των πιθανών δέντρων είναι μεγάλος, τότε η εκτίμηση κάθε δέντρου, χρησιμοποιώντας ακριβείς μεθόδους είναι πρακτικά αδύνατη.

Η ευρετική μέθοδος (heuristic search) είναι ουσιαστικά ένας αλγόριθμος αναρρίχησης λόφου (hill climbing), όπου επιλέγεται ένα αρχικό δέντρο και στη συνέχεια γίνονται αναδιευθετήσεις, επιζητώντας τη βελτίωση του δέντρου, βάσει του δεδομένου κριτηρίου επιλογής. Οι ευρετικοί αλγόριθμοι ομαδοποιούνται σε δύο κύριες κατηγορίες:

1. Η πρώτη κατηγορία περιλαμβάνει **αλγόριθμους ιεραρχικής ομαδοποίησης (hierarchical clustering algorithms)**, οι οποίοι διακρίνονται σε 2 υποομάδες:
   - **Agglomerative methods (Συσωρευτικές μέθοδοι)**, οι οποίες προχωρούν με τη διαδοχική συγχώνευση των n ειδών (αλληλουχιών) σε ομάδες, και
   - **Divisive methods (Διχαστικές μέθοδοι)**, οι οποίες διαχωρίζουν τα n είδη (αλληλουχίες) διαδοχικά σε μικρότερες ομάδες. Αν και κάθε βήμα περιλαμβάνει μια συγχώνευση ή μια διάσχιση, οι αλγόριθμοι πρέπει να επιλέξουν μία από τις

πολυάριθμες εναλλακτικές περιπτώσεις και η επιλογή γίνεται με βάση το κριτήριο που έχει καθοριστεί

2. Η δεύτερη κατηγορία περιλαμβάνει **<u>αλγόριθμους που κάνουν αναδιευθετήσεις στα επιλεγόμενα δέντρα (tree-rearrangement ή branchswapping algorithms)</u>**
   - Προτείνουν νέα δέντρα μέσω τοπικών «διαταραχών» στο τρέχων δέντρο, όπου το προκαθορισμένο κριτήριο (φειδωλότητα ή πιθανότητα) χρησιμοποιείται για να αποφασιστεί εάν θα κινηθεί προς το νέο δέντρο ή όχι. Η διαδικασία επαναλαμβάνεται μέχρι να μην μπορούν να γίνουν βελτιώσεις στο παραγόμενο δέντρο.

2.8.2.1.1η κατηγορία Ευρετικών αλγορίθμων ιεραρχικής ομαδοποίησης (hierarchical clustering algorithms)

i. **Stepwise addition (Agglomerative method)**
   a. Αρχίζει με ένα δέντρο 3 αλληλουχιών.
   b. Προσθέτει ένα taxon.
   c. Εκτιμά όλα τα δέντρα
   d. Επιλέγει το δέντρο με το καλύτερο score και προσθέτει νέο taxon,
      - είτε τυχαία από τα εναπομείναντα τάξα (αλληλουχίες) (Random)
      - είτε με τη σειρά που είναι στον πίνακα δεδομένων (alignment) που δίνουμε στο πρόγραμμα (Asis)
      - είτε επιλέγοντας το taxon που αυξάνει την τιμή του δέντρου (score) στο μέγιστο (furthest)
      - είτε επιλέγοντας το taxon που αυξάνει την τιμή του δέντρου (score) στο ελάχιστο (closest)

❖ **Μειονέκτημα**

Εάν το καλύτερο δέντρο σε ένα επίπεδο είναι το Α, αλλά τελικά το καλύτερο δέντρο με όλα τα taxa προέρχεται από το Β του ίδιου επιπέδου, τότε το καλύτερο δέντρο δεν θα βρεθεί. Η τεχνική stepwise θα σκαρφαλώσει στη κορυφή ενός λόφου, αλλά ο λόφος αυτός δεν είναι ο ψηλότερος.

ii. **Star Decomposition (Divisive method)**
   a. Ο αλγόριθμος ξεκινάει με όλα τα taxa να συνδέονται σε δέντρο με μορφή άστρου (star topology, όλα τα taxa συνδέονται σε ένα εσωτερικό κόμβο).
   b. Στη συνέχεια εκτιμώνται όλα τα δέντρα που δημιουργούνται με σύνδεση δύο ακραίων taxa (terminal nodes) σε μία ομάδα.
   c. Το δέντρο με τη καλύτερη τιμή (best score) διατηρείται για το επόμενο στάδιο.
   d. Σε κάθε βήμα, όταν δημιουργούμε μία νέα ομάδα, ο αριθμός των κλαδιών μειώνεται κατά ένα. Και αυτό συνεχίζεται μέχρι να έχουμε ένα διχοτομούμενο δέντρο.

<u>Σχόλια για την 1η κατηγορία</u>

❖ Αμφότεροι οι αλγόριθμοί stepwise-addition και star-decomposition παράγουν επιλυμένα δέντρα για το σύνολο των υπό εξέταση ειδών, τάξων, αλληλουχιών (n). Εάν σταματήσουμε στο τέλος κάθε αλγόριθμου, έχουμε μια αλγοριθμική μέθοδο συνάθροισης για την κατασκευή δέντρων βάσει του κριτηρίου που έχουμε επιλέξει.

❖ Ωστόσο στα περισσότερα φυλογενετικά προγράμματα, τα δέντρα που παράγονται από αυτούς τους αλγόριθμους χρησιμοποιούνται ως δέντρα εκκίνησης (starting trees) και υποβάλλονται σε τοπικές αναδιευθετήσεις.

2.8.2.2.2η κατηγορία Ευρετικών αλγορίθμων που κάνουν αναδιευθετήσεις στα επιλεγόμενα δέντρα (tree-rearrangement ή branchswapping algorithms).

- Στοχεύει στη βελτίωση της αρχικής εκτίμησης πραγματοποιώντας προκαθορισμένες διευθετήσεις στο δέντρο.
- Στην ουσία είναι τρόποι να «σπρώξεις» το δέντρο να ξεκολλήσει από το τοπικό βέλτιστο και να οδηγηθεί στο συνολικό βέλτιστο.
- Η μέθοδος αυτή περιλαμβάνει κόψιμο του δέντρου σε ένα ή περισσότερα σημεία (subtrees) και συναρμολόγηση του με τέτοιο τρόπο ώστε να διαφέρει από το αρχικό δέντρο.

Υπάρχουν **3 είδη μετακίνησης των υποδέντρων (subtrees)**

i. **NNI** (nearest-neighbor interchange)
ii. **SPR** (subtree pruning and regrafting)
iii. **TBR** (tree bisection and recombination)

i. **NNI (nearest-neighbor interchange)**

- Η απλούστερη μέθοδος, γνωστή ως NNI, αλλάζει τη συνδεσιμότητα των 4 υποδέντρων του κύριου δέντρου.
- Κάθε εσωτερικός κλάδος ενός άρριζου δέντρου έχει 4 υποδέντρα που συνδέονται σε αυτόν (ένα υποδέντρο μπορεί να αποτελείται από 1 και μόνο κόμβο).
- Η NNI αλλάζει τη θέση αυτών, παράγοντας νέα δέντρα.
- Υπάρχουν μόνο 2 αλλαγές που οδηγούν σε νέα δέντρα.
- Η διαδικασία συνεχίζει για κάθε εσωτερικό κλάδο έως ότου να μην γίνονται βελτιώσεις του αρχικού δέντρου βάσει του αρχικού κριτηρίου.
- Ένα δέντρο με N>2 φύλλα (κόμβους) έχει N-3 εσωτερικούς κλάδους και έτσι η NNI, που ελέγχει 2 δέντρα για κάθε εσωτερικό κλάδο, θα εξετάσει 2(N-3) νέα δέντρα.

ii. **SPR (Sub-tree Pruning and Regrafting) («κλαδεύω και μπολιάζω»)**

1. Η SSR επιλέγει το υποδέντρο του αρχικού δέντρου που θα κλαδέψει (pruning).
2. Αφαιρεί το υποδέντρο και το μπολιάζει σε άλλο σημείο του δέντρου, δημιουργώντας ένα νέο δέντρο.
3. Η διαδικασία συνεχίζεται για κάθε πιθανό υποδέντρο και για κάθε πιθανό κλαδί.

iii. **TBR (Tree bisection and reconnection) (Διχοτόμηση και επανασύνδεση)**

1. Η μέθοδος Tree-Bisection-Reconnection (TBR) κόβει το δέντρο σε 2 κομμάτια (υποδέντρα) και στη συνέχεια επανασυνδέει τα 2 υποδέντρα σε όλους τους πιθανούς κλάδους.
2. Εάν βρεθεί ένα δέντρο που είναι «καλύτερο» από το αρχικό, τότε αυτό διατηρείται και αρχίζει ένας νέος γύρος TBR.

- Όπως και στις προηγούμενες περιπτώσεις δεν εγγυάται ότι θα βρει το βέλτιστο δέντρο, ωστόσο είναι πιο ισχυρή από τις SPR και NNI.
- Τόσο η SPR όσο και η TBR αφαιρούν ένα υποδέντρο, όμως η SPR διατηρεί το αρχικό υποδέντρο ριζωμένο.

Σχόλιο 2ης κατηγορίας

- Ο χώρος με τα δέντρα μπορεί να είναι γεμάτος από τοπικά ελάχιστα και νησιά φειδωλών δέντρων!

2.9 Μεθοδοι Χαρακτηρων

Στις μεθόδους χαρακτήρων, τα ποιοτικά δεδομένων των αλληλουχιών DNA πολλών taxa αναλύονται ευθέως, ώστε να δώσουν φυλογενετικές εκτιμήσεις, βασισμένες και πάλι σε υποθέσεις και εξελικτικά μοντέλα. Αυτές οι μέθοδοι λοιπόν επικεντρώνονται στις διαφορές χαρακτήρων (A,T,G,C) ανά θέση μεταξύ των υπό μελέτη αλληλουχιών και όχι απλά στη συνολική γενετική απόσταση. Μια τέτοια μέθοδος είναι αυτή της μέγιστης φειδωλότητας (Maximum parsimony), ενώ ιδιαίτερα διαδεδομένες είναι και οι μέθοδοι μέγιστης πιθανοφάνειας. Η τελευταία κατηγορία θα αναλυθεί χωριστά, καθώς συνδυάζει χαρακτηριστικά μεθόδων αποστάσεων αλλά και μεθόδων χαρακτήρων.

2.9.1 Maximum Parsimony MP

Η βασική ιδέα των μεθόδων φειδωλότητας είναι η δημιουργία ενός φυλογενετικού δέντρου (ή ενός συνόλου δέντρων) το οποίο να ελαχιστοποιεί το πλήθος των εξελικτικών αλλαγών/μεταλλάξεων. Με άλλα λόγια, θεωρούν πως το πιο αποδεκτό και πιο κοντά στην πραγματικότητα δέντρο(-α), είναι εκείνο με το μικρότερο δυνατό συνολικό μήκος κλαδιών (Avice, 2006).

Η μέθοδος Μέγιστης Φειδωλότητας (MP) είναι εκείνη που απαιτεί τον ελάχιστο δυνατό αριθμό εξελικτικών αλλαγών, ώστε να ερμηνεύσει τις παρατηρούμενες διαφορές χαρακτήρων-θέσης μεταξύ των αλληλουχιών που μελετώνται. Με άλλα λόγια, με τη μέθοδο αυτή εκτιμάται ένα φυλογενετικό δέντρο το οποίο να ελαχιστοποιεί τα απαιτούμενα εξελικτικά βήματα για την περιγραφή του δείγματος. Πάντως, και παρά τη διαδεδομένη χρήση της συγκεκριμένης μεθόδου υπάρχει η κριτική που λέει πως η φύση γενικά δεν λειτουργεί με τους κανόνες της «μέγιστης φειδωλότητας», δηλαδή δεν επιλέγει πάντα τον πιο άμεσο και με τα λιγότερα βήματα δρόμο για να μεταβεί από μια κατάσταση σε μια άλλα (εξέλιξη), συνεπώς μέθοδοι φειδωλότητας παράγουν τοπολογίες, οι οποίες δεν περιγράφουν απαραίτητα την πραγματική εξελικτική ιστορία των δεδομένων (Avice, 2006).

Σχετικά με την εφαρμογή της μεθόδου, όταν το σύνολο των αλληλουχιών που πρέπει να αναλυθούν είναι μικρό, τότε είναι εφικτό να δημιουργηθούν όλες οι πιθανές τοπολογίες και να επιλεχθεί εκείνη που απαιτεί το ελάχιστο συνολικό πλήθος εξελικτικών βημάτων. Όταν το πλήθος όμως των αλληλουχιών μεγαλώνει, μια τέτοια διαδικασία απαιτεί πολύ χρόνο και επεξεργαστική ισχύ και συνεπώς έχουν προταθεί διάφορες μέθοδοι ευρετικής αναζήτησης, ώστε να μπορεί να ξεπεραστεί αυτό το πρόβλημα (Παρασκευής et al., 2015).

<u>Πληροφοριακή θέση (στήλη)</u>

Θέση που ευνοεί κάποιο δέντρο έναντι των υπολοίπων. Όταν υπάρχουν 2 τουλάχιστον καταστάσεις χαρακτήρων κάθε μια από τις οποίες αντιπροσωπεύεται σε τουλάχιστον 2 από τα εξεταζόμενα taxa.

1ο βήμα: Εντοπισμός Πληροφοριακών θέσεων
2ο βήμα: Υπολογισμός των απαιτούμενων εξελικτικών αλλαγών για κάθε δένδρο
3ο βήμα: Άθροισμα του αριθμού των αλλαγών
4ο βήμα: Επιλογή του πιο φειδωλού δέντρου

Ασυνέπεια στην φειδωλότητα

- Ο Felsenstein (1978) δημιούργησε ένα απλό μοντέλο φυλογένεσης που περιλάμβανε 4 τάξα και ένα μίγμα από μακριά και κοντά κλαδιά και προς έκπληξη όλων, η μέγιστη φειδωλότητα παράγει λάθος δέντρο. Τα μακριά κλαδιά έλκονται, όμως η ομοιότητα είναι λόγω ομοπλασίας.

- ❖ Όσο περισσότερα δεδομένα διαθέτουμε τόσο αυξάνει η πιθανότητα η φειδωλότητα να δώσει λάθος δέντρα – συνεπώς η φειδωλότητα είναι στατιστικά ασυνεπής.
- ❖ Υποστηρικτές της φειδωλότητας θεωρούν ότι το μοντέλο του Felsenstein είναι μη ρεαλιστικό.

Σήμερα αναγνωρίζεται ως Long-Branch Attraction (LBA) (ή η ζώνη Felsenstein) και αποτελεί ένα από τα πιο σοβαρά προβλήματα στην φυλογένεση.

Ασυνέπεια (LBA) όχι μόνο στη φειδωλότητα

- ❖ Δεν περιορίζεται μόνο στη μέγιστη φειδωλότητα.
- ❖ Αποτελεί μεγάλο πρόβλημα και στις μεθόδους αποστάσεων, ακόμα και στα δέντρα της μέγιστης πιθανότητας (ML).
- ❖ Όταν το μοντέλο νουκλεοτιδικής υποκατάστασης έχει κακή εφαρμογή (δεν είναι το πλέον κατάλληλο), μπορεί να οδηγήσει σε LBA.
- ❖ Ωστόσο τα ML δέντρα είναι πιο ανθεκτικά σε παραβιάσεις (μη καλή προσαρμογή) του μοντέλου.

Πλεονεκτήματα Maximum Parsimony MP

- ➢ Απλή μέθοδος – εύκολα κατανοήσιμη σε λειτουργία.
- ➢ Δεν φαίνεται να εξαρτάται από κάποιο συγκεκριμένο μοντέλο εξέλιξης.
- ➢ Θα μπορούσε να δώσει πολύ αξιόπιστα δέντρα εάν τα δεδομένα ήταν πολύ καλά δομημένα και η ομοπλασία ήταν σπάνια ή εμφανίζονταν τυχαία σε όλο το δέντρο.

Μειονεκτήματα Maximum Parsimony MP

- ➢ Μπορεί να δώσει λάθος αποτέλεσμα εάν η ομοπλασία είναι συχνή μέσα στα δεδομένα μας ή συγκεντρώνεται σε ένα συγκεκριμένο σημείο του δέντρου:
  - o thermophilic convergence
  - o base composition biases
  - o long branch attraction
- ➢ Υποεκτιμά τα μήκη των κλάδων
- ➢ Όταν η παραδοχή της φειδωλότητας δεν ισχύει, μπορεί να οδηγήσει σε λανθασμένες εκτιμήσεις εξελικτικών ρυθμών και διακλαδώσεων του δέντρου (Brocchieri, 2001)
- ➢ Απαιτεί πολύ αυστηρές παραδοχές αμεταβλητότητας των ρυθμών αντικατάστασης μεταξύ των γενετικών περιοχών αλλά και παρόμοιους ρυθμούς αντικατάστασης μεταξύ των γενεών για να αποδώσει ορθές εκτιμήσεις. (Brocchieri, 2001)
- ➢ Δραματική μείωση της απόδοσης της όταν οι ρυθμοί μετάλλαξης διαφέρουν μεταξύ των νουκλεοτιδίων και μεταξύ των προς ανάλυση γενετικών τόπων (Yang, 1996 ; Brocchieri, 2001)
- ➢ Αποτυγχάνει όταν οι εξελικτικοί ρυθμοί είναι υψηλά μεταβλητοί μεταξύ των γενεών ή εάν τα εσωτερικά κλαδιά του φυλογενετικού δένδρου είναι κοντά (Hendy , Penny , 1989 ; DeBry, 1992 ; Brocchieri, 2001)
- ➢ Εχει επίσης υποστεί κριτική και γιατί δεν παρέχει συνεπείς εκτιμήσεις και αυτό γιατί ο σχετικός αλγόριθμος της μεθόδου δεν συγκλίνει στην σωστή τοπολογία όταν αυξάνεται ο αριθμός των στοιχιζόμενων θέσεων (Felsenstein, 1978 ; Brocchieri, 2001)

2.9.2 Maximum Likelihood ML

Η μέθοδος ML (Felsenstein, 1981 & 1988) βασίζεται σε ειδικά πιθανολογικά μοντέλα εξέλιξης και ψάχνει το φυλογενετικό δένδρο με την μέγιστη πιθανοφάνεια υπό εκάστο

μοντέλο. Brocchieri (2001). Επιλέγει δηλαδή το/τα δέντρο(-α) εκείνο, από το οποίο είναι πιο πιθανό να έχουν προκύψει τα δεδομένα (Page & Holmes, 1998; Schmidt & von Haeseler, 2009). Η πιθανοφάνεια ορίζεται ως η δεσμευμένη πιθανότητα να παρατηρηθούν τα συγκεκριμένα δεδομένα του δείγματος, δεδομένου συγκεκριμένου εξελικτικού μοντέλου και συγκεκριμένου φυλογενετικού δέντρου. Έτσι, επί της ουσίας, σε αυτή τη μέθοδο γίνεται προσπάθεια να μεγιστοποιηθεί η ποσότητα:

$$L(\text{εξελικτικό μοντέλο}, \text{δέντρο}) = P(Data/\text{εξελικτικό μοντέλο}, \text{δέντρο})$$

Με τον όρο «δέντρο» αναφερόμαστε τόσο στην τοπολογία, δηλαδή στο πως ομαδοποιούνται οι αλληλουχίες του δείγματος, όσο και στα μήκη των κλαδιών του δέντρου, που όπως έχουμε πει συνήθως αντιστοιχούν σε εκτιμήσεις των εξελικτικών αποστάσεων. Έτσι, η μέθοδος μέγιστης πιθανοφάνειας καλείται να απαντήσει σε δύο διαφορετικά ερωτήματα: Για κάθε συγκεκριμένη τοπολογία, ποιο σύνολο μηκών κλαδιών κάνουν τα δεδομένα πιο πιθανά και ποιο δέντρο από τα πιθανά δέντρα τελικά (τοπολογία + μήκη κλαδιών), είναι αυτό με τη μεγαλύτερη πιθανοφάνεια (Page & Holmes, 1998).

Αμέσως γίνεται αντιληπτό πως, αν και σαν μέθοδος θα μπορούσε να θεωρηθεί μέθοδος χαρακτήρων, καθώς είναι μια διακριτή μέθοδος που αξιολογεί όλες (ή όσο το δυνατόν περισσότερες) τις πιθανές τοπολογίες, βασιζόμενη στις παρατηρούμενες διαφορές χαρακτήρων-θέσης μεταξύ των αλληλουχιών, ταυτόχρονα επιχειρεί να εκτιμήσει τη βέλτιστη τοπολογία, εκτιμώντας και λαμβάνοντας υπόψη τις εξελικτικές αποστάσεις μεταξύ των αλληλουχιών. Έτσι, συγκεντρώνει στοιχεία τόσο των μεθόδων αποστάσεων, όσο και των μεθόδων χαρακτήρων (Page & Holmes, 1998).

ΠΛΕΟΝΕΚΤΗΜΑΤΑ

- Απέδωσε ελαφρώς καλύτερα σε δένδρα με ίσους ρυθμούς ανά κλάδο, και παρατηρητέα καλύτερα όταν είχαμε άνισους ρυθμούς ανά κλάδο σε σχέση με την NJ (Hasegawa and Yano, 1984)
- ML ήταν καλλίτερος εκτιμητής από τις μεθόδους MP, ME, NJ, Fitch-Margoliash και μεθόδους συμβατότητας, στην πλειοψηφία των περιπτώσεων (Kuhner and Felsenstein, 1994)
- ML υπερτερούσε της NJ όταν συγκρίθηκαν από τον (Huelsenbeck, 1995),
- οι καλλίτερες αναδομήσεις φυλογενετικών δένδρων, όσον αφορά τον HIV-1 και άλλα συγγενικά συστήματα, έγιναν με χρήση της ML (Leitner et al., 1996 και 1997; Felsenstein, 1981).

ΜΕΙΟΝΕΚΤΗΜΑΤΑ

- υπολογιστικά κοστοβόρα (Brocchieri, 2001)
- όταν το πλήθος των αλληλουχιών αυξάνει, ακόμη και για σχετικά μικρά μεγέθη δείγματος, ο υπολογισμός όλων των τοπολογιών είναι πρακτικά αδύνατος, έτσι και πάλι χρησιμοποιούνται μέθοδοι ευρετικής αναζήτησης, προκειμένου εκτιμώντας ένα σημαντικό πλήθος τοπολογιών, να φτάσουν σε όσο το δυνατό καλύτερο αποτέλεσμα (Παρασκευής et al., 2015).
- τα μοντέλα που χρησιμοποιούνται για την εφαρμογή της ανωτέρω μεθόδου, συχνά είναι λιγότερο σύνθετα από όσο απαιτείται για να ερμηνεύσουν τους μεταβλητούς παράγοντες και τις ποικίλες εμπλοκές που χαρακτηρίζουν την φυσική εξελικτική διαδικασία (Zhang, 1999 ; Brocchieri, 2001)

❖ μπορούν να χρησιμοποιηθούν σε σχετικά μικρού όγκου δεδομένα (Huelsenbeck, 2001 ; Brocchieri, 2001)

2.10 Υποστήριξη Κόμβου Έλεγχοι Στατιστικης Σημαντικοτητας

Ένας επιστήμονας που θέλει να ελέγξει την αξιοπιστία των αποτελεσμάτων του επαναλαμβάνει το πείραμα του με άλλα δεδομένα. Για τους φυλογενετιστές η δημιουργία ενός φυλογενετικού δέντρου είναι ένα στατιστικό πρόβλημα και ο καθένας μπορεί να επιθυμήσει την εκτίμηση της αξιοπιστίας του.

Μετά τη δημιουργία ενός δέντρου μπορεί να αναδυθούν δύο ερωτήματα

1) Πόσο αξιόπιστο είναι το δέντρο; και

2) Είναι το δέντρο αυτό σημαντικά καλύτερο από κάποιο άλλο;

Η αξιοπιστία μετριέται ως η πιθανότητα τα μέλη ενός κλάδου να είναι πάντα μέλη αυτού του κλάδου.

Επομένως η έρευνα μας δεν ολοκληρώνεται ούτε με την δημιουργία του φυλογενετικού δέντρου. Ένας έλεγχος στατιστικής σημαντικότητας πρέπει να εφαρμοστεί στα όποια αποτελέσματα. Υπάρχουν πολλών τύπων έλεγχοι, όπως η υποστήριξη κατά Bremer (Bremer, 1994), η στάθμιση κατά Goloboff (Goloboff, 1993), ο έλεγχος των λόγων μέγιστων πιθανοφανειών κατά Kishino-Hasegawa (Kishino and Hasegawa, 1989), η ανάλυση T-PTP (Faith, 1991; Faith and Trueman, 1996), η jackknife ανάλυση αλλά, ο πιο συχνά χρησιμοποιούμενος είναι η μέθοδος bootstrap (Efron, 1982; Efron et al., 1996) και τον τελευταίο καιρό, η κατά πολύ ταχύτερη Μπεϋζιανή μέθοδος με χρήση των εκ των υστέρων πιθανοτήτων (posterior probabilities - PP - εκ των υστέρων πιθανότητες). Η μεταξύ τους επιλογή υπήρξε διαχρονικά θέμα μεγάλων συζητήσεων.

❖ Ωστόσο, και οι μέθοδοι αξιοπιστίας των δέντρων δεν είναι ελεύθεροι λαθών.

2.10.1 Bootstrap Method

Χρήση μιας μεθόδου δειγματοληψίας που ονομάζεται bootstrapping η οποία δημιουργεί ψεύτικα σύνολα δεδομένων μέσω των οποίων γίνεται εκτίμηση της αξιοπιστίας των δέντρων.

Παραδοσιακά, η στατιστική υποστήριξη κόμβων για τις σχέσεις σε ένα φυλογενετικό δέντρο έχει αξιολογηθεί με μια στατιστική τεχνική που ονομάζεται bootstrapping. Κατά τη διάρκεια του φυλογενετικού bootstrapping, η θέση των θέσεων (site positions) στην αρχική ευθυγράμμιση επαναδειγματίζεται τυχαία με αντικατάσταση για να παραχθεί μια σειρά ψευδο-επαναλήψεων ευθυγραμμίσεων. Στην συνέχεια εφαρμόζεται η προσέγγιση κατασκευής σε κάθε μία από αυτές τις ευθυγραμμίσεις. Συστάδες σχετιζόμενων ταξινομικών μονάδων taxa που υπάρχουν σε ένα μικρό ποσοστό των bootstrap δέντρων υποστηρίζονται ασθενώς και αντιστρόφως.

Ωστόσο, η ακριβής ερμηνεία των bootstrap τιμών είναι δύσκολη. Οι υψηλότερες τιμές είναι βεβαίως καλύτερες, αλλά ποιο είναι ένα λογικό όριο cut-off; Έχει προταθεί ότι οι bootstrap τιμές άνω του 70% δείχνουν ισχυρή υποστήριξη για μια ομάδα, με βάση το συμπέρασμα ότι οι υποστηρίξεις bootstrap είναι συντηρητικές μετρήσεις.

Όμως, τι είναι πρακτικά η μέθοδος bootstrap; Στην ουσία είναι μία κληρωτίδα. Αυτό είναι απόλυτα κυριολεκτικό μιας και η μέθοδος bootstrap ανήκει στην μεγάλη οικογένεια των δειγματοληπτικών μεθόδων. Η μέθοδος διέπεται από δύο κύριες παραδοχές, της ανεξαρτησίας και της ισοκατανομής των παρατηρήσεων. Γενικά, κατά την εφαρμογή της μεθόδου γίνονται συνεχείς επαναδειγματοληψίες από μία δεξαμενή χαρακτηριστικών και έτσι δομούνται νέα πακέτα δεδομένων αντίστοιχα με το αρχικό μας πακέτο δεδομένων. Εν συνεχεία, ελέγχεται ο βαθμός αντιστοιχίας που παρατηρείται στα νέα πακέτα δεδομένων

που προέκυψαν από τις επαναδειγματοληψίες και συμπεραίνουμε το ποσοστό αξιοπιστίας των αρχικών μας δεδομένων. Πιο ειδικά, όσον αφορά τα φυλογεντικά δεδομένα, η μέθοδος bootstrap εφαρμόζεται ως εξής: από τις αρχικές αλληλουχίες των εμπλεκομένων ατόμων ή ειδών, γίνονται τυχαιοποιημένα αντικαταστάσεις νουκλεοτιδίων ώστε με βάση τις αρχικές αλληλουχίες να προκύπτουν νέες αντίστοιχου μεγέθους και αριθμού, και πάντα τηρούμενων των δύο βασικών παραδοχών που αναφέρουμε ανωτέρω. Εν συνεχεία, δομούνται φυλογενετικά δένδρα με αυτές. Ακολούθως, στο αρχικό φυλογενετικό δένδρο που έχουμε δομήσει με τα αρχικά δεδομένα μας, δίδεται ανά κλάδο ένα ποσοστό. Το ποσοστό αυτό αντιστοιχεί στον αριθμό των δένδρων bootstrap τα οποία είχαν και αυτά αναδομήσει τον κλάδο αυτόν (Felsenstein, 1985). Έτσι, εάν π.χ. ζητήσουμε 100 δειγματοληψίες τύπου bootstrap, αφού γίνουν οι τυχαιοποιημένες αντικαταστάσεις στα δεδομένα μας και δομηθούν τα νέα δένδρα, στην περίπτωση που δούμε ένα ποσοστό 97% σε έναν εκ των κλάδων του δένδρου που υποθέτουμε ότι ισχύει, τότε σημαίνει ότι 97 από τα 100 δένδρα επαναδειγματοληψίας είχαν αυτόν τον κλάδο.

Σχόλια για το Bootstrapping
• Αυτού του είδους το bootstrapping είναι μη παραμετρικό
• Θεωρεί ότι οι θέσεις που έχει συλλέξει είναι αντιπροσωπευτικές
• Τιμές 80-85 θεωρούνται γενικά αποδεκτές ως οι ελάχιστες τιμές
αξιόπιστης στατιστικής υποστήριξης
• Το Bootstrapping είναι μεγάλη και χρονοβόρα δουλειά
• το φαινόμενο του Long Branch Attraction (LBA)

ΜΕΙΟΝΕΚΤΗΜΑΤΑ

- η διαδικασία εκτίμησης της αξιοπιστίας των αποτελεσμάτων με χρήση της μεθόδου bootstrap είναι γεμάτη με εμπόδια λόγω της έλλειψης ανεξαρτησίας και της ομοιογένειας των μοριακών δεδομένων (Brocchieri, 2001).
- παρέχει έναν συντηρητικό έλεγχο της σημαντικότητας της εκτιμόμενης τοπολογίας - υποεκτιμάται η πιθανότητα όταν η πιθανότητα είναι υψηλή και υπερεκτιμάται η πιθανότητα όταν η πιθανότητα είναι χαμηλή (Zharkikh and Li, 1992; Hillis and Bull, 1993).
- δεν εφαρμόζει ορθά στις μοριακές αλληλουχίες μιας και η συσχέτιση μεταξύ των θέσεων στις αλληλουχίες παραβαίνει τις παραδοχές της ανεξαρτησίας και ισοκατανομής των παρατηρήσεων οι οποίες απαιτούνται (Brocchieri, 2001).
- Άνισοι ρυθμοί εξέλιξης σε διαφορετικές γενεαλογίες εισαγάγουν επιπρόσθετους συγχυτικούς παράγοντες (Brocchieri, 2001).

2.10.2 Approximate likelihood-ratio test (aLRT) & Zero-branch length test

Δύο άλλοι τύποι στατιστικών τεστ, οι οποίοι είναι ουσιαστικά ταχύτεροι από την προσέγγιση bootstrap, είναι the approximate likelihood-ratio test και the zero-branch length test . Στην ουσία, αυτά ελέγχουν εάν κάθε κλάδος σε ένα δέντρο είναι σημαντικά μεγαλύτερος από το μηδέν ή όχι (δηλαδή εάν υπάρχει κλάδος) και οι πιθανότητες αποκοπής cut-off probabilities άνω του 0,9 έχουν προταθεί να είναι συντηρητικές και να αντιστοιχούν σχετικά καλά στις τιμές bootstrap περισσότερο από 70%.

Τέλος, μελέτες έδειξαν ότι η σχέση στα αποτελέσματα μεταξύ των PP και του bootstrap με ML είναι υψηλά μεταβλητή, αλλά και ότι υπάρχουν πολύ ισχυρές συσχετίσεις όταν η εκτίμηση με Μπεϋζιανές μεθόδους γίνει σε χαρακτήρες ήδη επεξεργασμένους μέσω bootstrap. Σε μία προσπάθεια να «δεθούν» αυτά τα στοιχεία, οι Douady et al. (2003) εφάρμοσαν μη παραμετρική δειγματοληψία bootstrap σε Μπεϋζιανή μεθοδολογία. Έτσι, τα

όποια προβλήματα προέκυπταν στην τοπολογία με την Μπεϋζιανή μεθοδολογία ελαττώθηκαν με την εφαρμογή επί αυτών του bootstrap συζεύγοντας έτσι τις δύο μεθόδους και κάνοντας τις να αποδώσουν καλύτερα.

2.11. Μπευζιανή Μέθοδος

Οι Μπεϋζιανές μέθοδοι, με χρήση διαφόρων αλγορίθμων όπως ο Markov Chains Monte Carlo (MC2) και ακόμα καλύτερα ο Metropolis Coupled Markov Chains Monte Carlo (MC3), ήρθαν ως το μεγάλο αντίπαλο δέος.

Αντί να στηρίζονται σε ένα «βέλτιστο δέντρο» ή σε ένα σύνολο του bootstrap ψευδο-επαναλήψεων, οι Bayesian φυλογενετικές προσεγγίσεις χρησιμοποιούν τη δειγματοληψία Markov της αλυσίδας Monte Carlo για να συναγάγουν μια πλήρη εκ των υστέρων κατανομή πιθανότητας εύλογων δέντρων, η οποία θα πρέπει να περιέχει όλες τις διαφορετικές τοπολογίες δέντρων που υποστηρίζονται καλά από τα δεδομένα. Αυτό το σύνολο δέντρων μπορεί να χρησιμοποιηθεί για την παραγωγή ενός δέντρου συναίνεσης a consensus tree [ονομάζεται δέντρο μέγιστης αξιοπιστίας κλάδου (MCC) a maximum clade credibility] όπου κάθε κλάδος και σύμπλεγμα έχει σχετική associated πιθανότητα. Σε ένα δέντρο MCC, αυτή η πιθανότητα είναι η αναλογία των δέντρων στην κατανομή της εκ των υστέρων πιθανότητας στην οποία υπάρχει η συστάδα ενδιαφέροντος.

Η Μπευζιανή μέθοδος είναι μια μέθοδος για την εκτίμηση της πιθανότητας να συμβεί ένα γεγονός από την εξέταση τόσο της εκ των προτέρων πιθανότητας να συμβεί αυτό το γεγονός όσο και των δεδομένων που έχουμε στη διάθεσή μας. Η Μπευζιανή συμπερασματολογία δίνει την εκ των υστέρων πιθανότητα ως συνέπεια δυο πρωτύτερων γεγονότων i) της εκ των προτέρων πιθανότητας και ii) της λειτουργικής πιθανότητας, η οποία προκύπτει από ένα στατιστικό μοντέλο για τα παρατηρούμενα δεδομένα. Η Μπευζιανή συμπερασματολογία υπολογίζει την εκ των υστέρων πιθανότητα βάσει του θεωρήματος του Bayes

$$P(\upsilon\pi\acute{o}\theta\varepsilon\sigma\eta/\delta\varepsilon\delta o\mu\acute{\varepsilon}\nu\alpha) = \frac{Pr(\delta\varepsilon\delta o\mu\acute{\varepsilon}\nu\alpha/\upsilon\pi\acute{o}\theta\varepsilon\sigma\eta)Pr(\upsilon\pi\acute{o}\theta\varepsilon\sigma\eta)}{\sum_{\upsilon\pi o\theta\acute{\varepsilon}\sigma\varepsilon\iota\varsigma} Pr(\delta\varepsilon\delta o\mu\acute{\varepsilon}\nu\alpha/\upsilon\pi\acute{o}\theta\varepsilon\sigma\eta)Pr(\upsilon\pi\acute{o}\theta\varepsilon\sigma\eta)}$$

Δηλαδή,

$$P(\theta/D) = \frac{Pr(D/\theta)Pr(\theta)}{\sum_{\theta} Pr(D/\theta)Pr(\theta)}$$

- Το D αναφέρεται στα παρατηρούμενα (δηλ. δεδομένα όπως αλληλουχίες)
- Το θ αναφέρεται σε ένα ή περισσότερα μη παρατηρήσιμα στοιχεία (δηλ. οι παράμετροι ενός μοντέλου):
  i. δέντρο (π.χ. η τοπολογία ενός δέντρου, μήκος κλάδων)
  ii. το μοντέλο νουκλεοτιδικής υποκατάστασης (π.χ. JC, F84, GTR, κ.α.)
  iii. οι παράμετροι ενός μοντέλου υποκατάστασης (π.χ. συχνότητα βάσεων, αναλογία μεταπτώσεων / μεταστροφών κ.α.)
  iv. υπόθεση (π.χ. μια ειδική περίπτωση μοντέλου)
  v. μια μη προφανής μεταβλητή (π.χ. προγονική κατάσταση)

Το βασικό σημείο που η μπευζιανή αποκλίνει από την likelihood είναι επειδή η εκ των προτέρων πληροφορία μπορεί να ενσωματωθεί στην Μπευζιανή ανάλυση μέσω του θεωρήματος του Bayes, ενώ η ίδια πληροφορία δεν μπορεί να χρησιμοποιηθεί υπό το πρίσμα της πιθανοφάνεια. Η εκ των προτέρων πιθανότητα έχει σημαντική επίδραση στην εκ των υστέρων πιθανότητα όταν τα δεδομένα είναι λίγα. Ανάλογα με την οπτική του καθενός, η

ενσωμάτωση εκ των προτέρων στοιχείων σχετικά με μια παράμετρο είναι είτε πλεονέκτημα είτε μειονέκτημα της μπευζιανής.

Η Μπευζιανή συμπερασματολογία στη φυλογένεση βασίζεται στην εκ των υστέρων πιθανότητα ενός φυλογενετικού δέντρου, τ_i. Η εκ των υστέρων πιθανότητα του i_{th} φυλογενετικού δέντρου, τ_i , βάσει ενός συγκεκριμένου συνόλου ευθυγραμμισμένων DNA αλληλουχιών (X) δίνεται από την εξίσωση:

$$f(\tau_i/X) = \frac{f(X/\tau_i)f(\tau_i)}{\sum_{j=1}^{B(s)} f(X/\tau_j)f(\tau_j)}$$

όπου,

- $f(\tau_i/X)$ είναι η εκ των υστέρων πιθανότητα της i_{th} φυλογένεσης και μπορεί να ερμηνευθεί ως η πιθανότητα το δέντρο τ_i να είναι το σωστό δέντρο δεδομένου των δοσμένων DNA αλληλουχιών.
- Η πιθανότητα εμφάνισης των αλληλουχιών δεδομένου του i_{th}δέντρου είναι $f(X/\tau_i)$
- Η εκ των προτέρων πιθανότητα του i_{th} δέντρου είναι $f(\tau_i)$. Είναι ουσιαστικά η πιθανότητα εμφάνισης μίας τοπολογίας προ της επιλογής των δεδομένων. Τυπικώς, όλα τα δένδρα θεωρούνται ισοπίθανα.

❖ Το άθροισμα στον παρονομαστή προέρχεται από όλα τα δέντρα B(s) που είναι πιθανά για s είδη. Αυτός ο αριθμός είναι

- <u>για έρριζα δένδρα (n≥2):</u>
  $$\mathrm{B(s)} = {(2n-3)!}\Big/{2^{n-2}(n-2)!}$$
- <u>για άρριζα δένδρα (n≥3):</u>
  $$\mathrm{B(s)} = {(2n-5)!}\Big/{2^{n-3}(n-3)!}$$

❖ Τυπικά ένα μη πληροφοριακό εκ των προτέρων στοιχείο (prior) χρησιμοποιείται για τα δέντρα, όπως το

- $f(\tau_i) = 1/B(s)$

Συνοπτικά λοιπόν, οι εκ των υστέρων πιθανότητες (posterior probabilities - PP) ενός δένδρου μπορούν απλά να ερμηνευτούν ως η πιθανότητα του δένδρου να είναι σωστό. Έτσι, μετά από εκτιμήσεις για έναν μεγάλο αριθμό δένδρων, αυτό με την μεγαλύτερη PP μπορεί να επιλεγεί ως το πιο πιθανό για την εκτίμηση της τοπολογίας μας.(Huelsenbeck et al., 2001).

Η εκτίμηση της κατανομής των εκ των υστέρων πιθανοτήτων όλων των δέντρων που εμπεριέχονται σε μια φυλογενετική ανάλυση περιλαμβάνει αμφότερα το άθροισμα όλων αυτών δέντρων και για κάθε δέντρο την ολοκλήρωση όλων πιθανών συνδυασμών των μηκών των κλάδων και των τιμών των παραμέτρων του μοντέλου (Huelsenbeck et al., 2001).

Δυστυχώς τέτοια προβλήματα δεν μπορούν να λυθούν αναλυτικά, οπότε στρεφόμαστε σε στοχαστικές προσομοιώσεις, δειγματοληπτώντας κατά προσέγγιση από την από την κατανομή των εκ των υστέρων πιθανοτήτων των δέντρων.

Τα πιο χρήσιμα εργαλεία που διαθέτουμε για τέτοιου είδους προσεγγίσεις βασίζονται στην θεωρία του **Markov chain Monte Carlo (MCMC)**. Στην ουσία χωρίς τις μεθόδους MCMC θα ήταν αδύνατη η εφαρμογή των αρχών της Bayesian σε προβλήματα φυλογένεσης.

Η Markov chain Monte Carlo (MCMC) είναι, με απλά λόγια, μια προσομοίωση ενός τυχαίου περιπάτου στο χώρο των τιμών των παραμέτρων με σκοπό τη συλλογή από την κατανομή των εκ των υστέρων πιθανοτήτων που μας ενδιαφέρει. Πιο συγκεκριμένα, η MCMC επιτρέπει σε ένα φυλογενετιστή να συλλέξει φυλογενετικά δέντρα σύμφωνα με τις εκ των υστέρων πιθανότητές τους (Huelsenbeck and Ronquist, 2001a).

Η Markov αλυσίδα είναι μια προγραμματισμένη αλληλουχία (ή αλυσίδα) τυχαία επιλεγμένων δειγμάτων από μια καθορισμένη περιοχή κατά τη διάρκεια του περιπάτου στο χώρο των παραμέτρων.

Μπορούμε να συνοψίσουμε όλες τις απαραίτητες φυλογενετικές παραμέτρους σε μία μεταβλητή ψ = {t, n, θ, α}, όπου ψ είναι

i) το δέντρο με συγκεκριμένη τοπολογία και μήκη κλάδων και

ii) οι παράμετροι του μοντέλου υποκατάστασης στις οποίες περιλαμβάνεται και το Γ (G) .

Στο πλαίσιο της MCMC είναι εύκολο να φανταστούμε το ψ ως ένα μοναδικό σημείο στο χώρο των παραμέτρων. Μια «διαταραχή» στην τιμή κάποιας από τις παραμέτρους που συνιστούν το ψ θα αλλάξει το ψ σε ψ΄ και έτσι θα καθοριστεί ένα νέο σημείο στο χώρο των παραμέτρων. Η αλυσίδα Markov δουλεύει συλλέγοντας διαφορετικά ψ καθώς κινείται τυχαία μέσα στο χώρο των παραμέτρων.

Εάν η αλυσίδα κατασκευαστεί συνετά και κινηθεί αρκετά μέσα στο χώρο των παραμέτρων, τότε τα δείγματα που θα συλλέξει η αλυσίδα θεωρείται ότι προέρχονται από την περιοχή κατανομής των εκ των υστέρων πιθανοτήτων που μας ενδιαφέρει.

Η μεθοδολογία Markov chain Monte Carlo (MCMC) είναι παρόμοια με τον αλγόριθμο εύρεσης του δέντρου.

- Από ένα αρχικό δέντρο προτείνεται ένα νέο δέντρο. Η κίνηση για την επιλογή του νέου δέντρου είναι τυχαία.
- Ο αλγόριθμος MCMC καθορίζει μεταξύ των άλλων και τους κανόνες για να γίνει αποδεκτό ή όχι το νέο δέντρο

Στην MCMC, όπως αυτή εφαρμόζεται σε ένα από πιο χρησιμοποιημένα προγράμματα για Bayesian στη φυλογένεση (MrBayes), οι τιμές των παραμέτρων «διαταράσσονται» με 2 τρόπους. Ο πρώτος προκαλεί μεταβολές στις παραμέτρους του μοντέλου υποκατάστασης, ενώ ο δεύτερος στην τοπολογία και τα μήκη των κλάδων. Αμφότεροι κάνουν χρήση του αλγόριθμου Metropolis-Hastings, ο οποίος απλά καθορίζει την τιμή της πιθανότητας ώστε να γίνει αποδεκτή η νέα κατάσταση.

Το αν θα γίνει αποδεκτή η νέα θέση στο χώρο των παραμέτρων εξαρτάται ουσιαστικά από τη πιθανότητα της νέας κατάστασης

$$R = min\left[1, \frac{f(X/\Theta')f(\Theta')f(\Theta/\Theta')}{f(X/\Theta)f(\Theta)f(\Theta'/\Theta)}\right]$$
$$= min[1, likelihood\ ratio \times prior\ ratio \times proposal\ ratio]$$

Πρακτικά $R = {}^{καινούριο\ υψόμετρο}\Big/_{τρέχων\ υψόμετρο}$

- εάν R< 1, τότε τραβάμε τυχαία μια τιμή από μία ομοιόμορφή κατανομή με τιμές από 0 έως 1 {U{1,0}] και
- εάν R > της τιμής της μεταβλητής από την ομοιόμορφη κατανομή, τότε το βήμα γίνεται αποδεκτό.
- Όσο το R τείνει στο 0, τόσο μικρότερη είναι πιθανότητα να ληφθεί τυχαία μια τιμή της μεταβλητής που να είναι μικρότερη του R, ώστε να γίνει αποδεκτό το βήμα.
- Γι' αυτό και μεγάλα άλματα προς τα κάτω δεν γίνονται συνήθως αποδεκτά.

- ❖ Μικρά προς τα κάτω βήματα συνήθως γίνονται αποδεκτά γιατί R->1.
- ❖ Μεγάλα προς τα κάτω βήματα σχεδόν ποτέ δεν γίνονται αποδεκτά γιατί R->0.
- ❖ Προς τα πάνω βήματα γίνονται πάντα αποδεκτά, γιατί R>1.

Μια σωστά κατασκευασμένη αλυσίδα Markov θα δειγματοληπτήσει από την σταθερή (επιθυμητή) περιοχή που μας ενδιαφέρει, αλλά αυτό μπορεί να πάρει πολλές γενεές (βήματα) για να συμβεί. Ο λόγος που συμβαίνει αυτό είναι γιατί μια αλυσίδα συνήθως (ιδανικά) ξεκινάει από ένα τυχαίο σημείο του χώρου των παραμέτρων που μπορεί να έχει σημαντική απόσταση από την κορυφή των εκ των υστέρων πιθανοτήτων που μας ενδιαφέρει. Τα δείγματα που λαμβάνονται από την αλυσίδα Markov πριν φτάσει στην περιοχή ενδιαφέροντος δεν προέρχονται από την περιοχή κατανομής που μας ενδιαφέρει (έχουν ουσιαστικά μηδενική πιθανότητα) και έτσι απορρίπτονται από το σύνολο των δειγμάτων που λαμβάνει το ρομποτάκι μας. Τα δείγματα αυτά που απορρίπτονται αναφέρονται ως το «καμένο» τμήμα της αλυσίδας (“burnin” of the chain).

ΣΥΝΟΨΗ ΒΗΜΑΤΩΝ MCMC

- ❖ Ξεκινάμε με ένα τυχαίο δέντρο και τυχαίες τιμές για τα μήκη των κλάδων και των παραμέτρων του μοντέλου
- ❖ Κάθε γενεά (μετακίνηση του ρομπότ) περιλαμβάνει ένα από τα επόμενα (η επιλογή είναι τυχαία):
  - Προτείνεται ένα νέο δέντρο και είτε γίνεται δεκτό είτε απορρίπτεται
  - Προτείνεται μία νέα τιμή για μια παράμετρο (και είτε γίνεται δεκτή είτε απορρίπτεται)
- ❖ Κάθε k γενεές, σώζεται η τοπολογία του δέντρου, τα μήκη των κλάδων και όλες οι παράμετροι
- ❖ Μετά από n γενεές, ανακεφαλαιώνονται (συνοψίζονται) τα δείγματα που σώθηκαν χρησιμοποιώντας ιστογράμματα, μέσες τιμές διαστήματα εμπιστοσύνης κ.α.

Metropolis Coupled Markov Chain Monte Carlo MCMCMC (MC)^3

- ❖ Τα μεγάλα βήματα βοηθούν να πηδάμε από τον ένα λόφο (νησί) στον άλλο μέσα στην περιοχή που μας ενδιαφέρει (εκ των υστέρων πιθανότητες)
- ❖ Τα μικρά βήματα βοηθούν στο να γίνει καλύτερη δειγματοληψία (mixing)

Πως θα γίνει ο παραπάνω συμβιβασμός;

- ➢ Απάντηση : MCMCMC (MC)^3

Η MCMCMC περιλαμβάνει το τρέξιμο μερικών (n) αλυσίδων ταυτόχρονα

• Η μία εξ αυτών, καλείται κρύα αλυσίδα (cold chain) και είναι αυτή που μετράει (λαμβάνει τιμές από το χώρο των παραμέτρων). Οι υπόλοιπες (n-1) ονομάζονται ζεστές αλυσίδες (heated chains)

Κάθε αλυσίδα υπολογίζει την εκ των υστέρων πιθανότητα της θέσης που πηγαίνει (π.χ. τοπολογία του δέντρου) και στη συνέχεια υποβάλει την εκ των υστέρων πιθανότητα σε μια δύναμη β. Το β είναι ο βαθμός της θέρμανσης (η απλά η θερμοκρασία) και παίρνει τιμές από $0 < \beta < 1$.

Η κρύα αλυσίδα παίρνει τιμή β ίση με 1, καθιστώντας την ανεπηρέαστη από τη θέρμανση. Οι ζεστές αλυσίδες παίρνουν τιμές από 0 έως 1. Συνεπώς η εκ των υστέρων πιθανότητα δίνεται από την εξίσωση

- ➢ $Pr(t/X)$ για την κρύα αλυσίδα
- ➢ $Pr(t/X)^{\beta}$ για τις ζεστές αλυσίδες.

Η διαδικασία της θέρμανσης «λιώνει» το ανάγλυφο των εκ των υστέρων πιθανοτήτων μόνο για τις ζεστές αλυσίδες, καθιστώντας τις κοιλάδες αναμεσά στους λόφους πιο ήπιες και τις κορυφές λιγότερο ψηλές, αν και όλες οι αλυσίδες εξερευνούν τον ίδιο χώρο παραμέτρων.

- Όταν β = 1 η αλυσίδα είναι ίδια με την κρύα
- Όταν β = 0 το τοπίο δεν έχει πλέον ανάγλυφο (επίπεδο) και η πιθανότητα σε κάθε σημείο είναι ίση με 1.

❖ Οι ζεστές αλυσίδες εξερευνούν την ίδια περιοχή αλλά η θέρμανση (λιώσιμο) του ανάγλυφου έχει μειώσει τις απότομες κορυφές και τις κοιλάδες. Μια ζεστή αλυσίδα μπορεί πιο εύκολα να διασχίσει κοιλάδες επειδή τα προς τα κάτω βήματα είναι μικρότερα σε μέγεθος. Παρά την αυξημένη κινητικότητα των ζεστών αλυσίδων, η μοναδική τους λειτουργία είναι να παρέχουν στην κρύα αλυσίδα νέες καταστάσεις.

❖ Μόνο η κρύα αλυσίδα καταγράφει τιμές από την περιοχή που κάνει δειγματοληψία.

❖ Οι ζεστές αλυσίδες λειτουργούν ως πρόσκοποι (ανιχνευτές), διερευνώντας την επιφάνεια των εκ των υστέρων πιθανοτήτων για απομονωμένες περιοχές (κορυφές) με υψηλή πιθανότητα. Ως ανιχνευτές, οι αλυσίδες θα πρέπει περιοδικά να επικοινωνούν μεταξύ τους, προσθέτοντας ένα ακόμα επίπεδο πολυπλοκότητας στην ανάλυση. Πολλοί επιστήμονες αρέσκονται στο να παρομοιάζουν τις διαφορετικές αλυσίδες ως ανεξάρτητα ρομπότ που εξερευνούν το χώρο των δέντρων, καθένα εκ των οποίων έχει ένα ασύρματο επικοινωνίας για να ενημερώνουν το ένα το άλλο για το υψόμετρο στο οποίο βρίσκονται και να ανταλλάσσουν θέση όταν αυτό είναι αποδεκτό. KRHTHS

Μέγιστη Πιθανοφάνεια vs. Μπευζιανή συμπερασματολογία

Η Μέγιστη Πιθανοφάνεια ψάχνει το δέντρο που μεγιστοποιεί τη πιθανότητα να παρατηρηθούν τα δεδομένα [P (Δεδομένα |Δέντρο)]

Η Μπευζιανή συμπερασματολογία ψάχνει το δέντρο που μεγιστοποιεί τη πιθανότητα να παρατηρηθεί το δέντρο δεδομένου των δεδομένων [P (Δέντρο | Δεδομένα)] KRHTHS

Σχόλια για MCMC

❖ Η MCMC ένας αλγόριθμος που προσπαθεί να προσεγγίσει την εκ των υστέρων κατανομή. Όσο πιο πολλά δείγματα ληφθούν τόσο το καλύτερο! Αλλά πόσα είναι αρκετά;
  - Η αλήθεια είναι ότι ΠΟΤΕ δεν μπορούμε να πούμε με σιγουριά ότι έχουμε συλλέξει ένα ικανοποιητικό αριθμό δειγμάτων (Lewis, 2002). Με άλλα λόγια δεν είναι εφικτό να καθορίσουμε, θεωρητικά, το κατάλληλο μήκος του τρεξίματος.

❖ Τα δείγματα μιας αλυσίδας Markov είναι αυτοσυσχετιζόμενα Η αυτοσυσχέτιση αυξάνει όσο το δ είναι μικρό. Με άλλα λόγια, ο απόλυτος αριθμός των δειγμάτων που λαμβάνεται είναι κατά πολύ μεγαλύτερος από το δραστικό μέγεθος των δειγμάτων. Υπάρχουν 2 στρατηγικές εδώ:
  - 1) αύξηση του αριθμού των δειγμάτων ή
  - 2) «λέπτυνση» της αλυσίδας Markov (δηλαδή να δειγματοληπτεί κάθε 100 ή περισσότερα βήματα).

❖ Πόσο σίγουροι είμαστε ότι η επιθυμητή περιοχή έχει εντοπιστεί και από εκεί δειγματοληπτούμε; Τα δείγματα που λαμβάνει η αλυσίδα Markov καθοδόν προς την επιθυμητή περιοχή. Αυτά τα δείγματα έχουν μηδενική πιθανότητα και δεν τα θέλουμε στο δείγμα αφού πιθανά θα το προκαλέσουν μια στρέβλωση. Πρέπει να τα αποβάλλουμε, αλλά πως καθορίζεται το κατώφλι απομάκρυνσης (αποβολής) κάποιων δειγμάτων;

- Αυτό μπορεί να γίνει μέσω διαγραμμάτων. Καθώς οι τιμές αρχίζουν να κάνουν ένα «plateau» μπορούμε θεωρητικά να υποθέσουμε ότι η επιθυμητή περιοχή έχει προσεγγιστεί.

- Τα «priors» που καθορίζονται εξ αρχής είναι το πιο αμφιλεγόμενο κομμάτι της ανάλυσης
- Κακή επιλογή μοντέλου οδηγεί σε λανθασμένα αποτελέσματα
- Διάφορα διαγνωστικά MCMC (αν και ποιοτικά) υπάρχουν και θα πρέπει να λαμβάνονται υπόψη για την αποφυγή λανθασμένων συμπερασμάτων, ενώ παράλληλα η επιλογή του μοντέλου είναι εξαιρετικής σημασίας σε οποιαδήποτε φυλογενετική άσκηση.

Μπευζιανή Συμπερασματολογία ΠΛΕΟΝΕΚΤΗΜΑΤΑ

- Αυτόματα παρέχει τις εκ των υστέρων πιθανότητες (PP-Posterior Probabilities) από τις οποίες συντελεστές εμπιστοσύνης, για το εκάστοτε δένδρο, μπορούν να υπολογιστούν εάν κάποιος το επιθυμεί.
- Υπάρχει ένα σαφές πλεονέκτημα για τις Μπεϋζιανές μεθόδους, γιατί το να υπολογίζει πέραν των τοπικών μεγίστων φτάνει μέχρι την «καρδιά» της μεθοδολογίας του αλγορίθμου MC^2 και ακόμα περισσότερο, όταν εφαρμόζεται ο αλγόριθμος MC^3 οι μέθοδοι είναι λιγότερο επίφοβο να «παγιδευτούν» σε τοπικά μέγιστα (Huelsenbeck et al., 2001).
- Η Μπευζιανή συμπερασματολογία λαμβάνει υπόψιν το πρόβλημα των φυλογενετικών και κάνει την ανάλυση μεγάλου όγκου δεδομένων πιο αποτελεσματική:
- αντί να ψάχνει για το ιδανικό δένδρο, δειγματοληπτεί δένδρα ανάλογα με τις PP τους.
- Όταν δημιουργηθεί ένα τέτοιο δείγμα, τα χαρακτηριστικά τα οποία είναι κοινά μεταξύ των δένδρων μπορούν να διακριθούν.
    - Για παράδειγμα, το δείγμα μπορεί να χρησιμοποιηθεί για να σχεδιαστεί ένα ομόφωνο δένδρο, με τις PP του εκάστοτε κλάδου να παρουσιάζονται στο δένδρο.
    - Αυτό είναι περίπου ισοδύναμο με το να εφαρμοστεί μια ανάλυση με την χρήση ML και να ακολουθήσει δειγματοληψία bootstrap, αλλά κατά πολύ ταχύτερο.
- Επίσης, οι Μπεϋζιανές αναλύσεις είναι ευρέως σύμφωνες με τις αναλύσεις φειδωλότητας. Πάρα ταύτα, η υποστήριξη σε μεγαλύτερες αποκλίσεις είναι γενικά υψηλότερη στις Μπεϋζιανές μεθόδους (Huelsenbeck et al., 2001 ; Brocchieri, 2001)
- Οι Bayesian οπίσθιες πιθανότητες έχουν προταθεί ότι είναι ένας γενικά λιγότερο προκατειλημμένος προγνωστικός παράγοντας της φυλογενετικής ακρίβειας από το bootstrapping.

Μπευζιανή Συμπερασματολογία ΜΕΙΟΝΕΚΤΗΜΑΤΑ

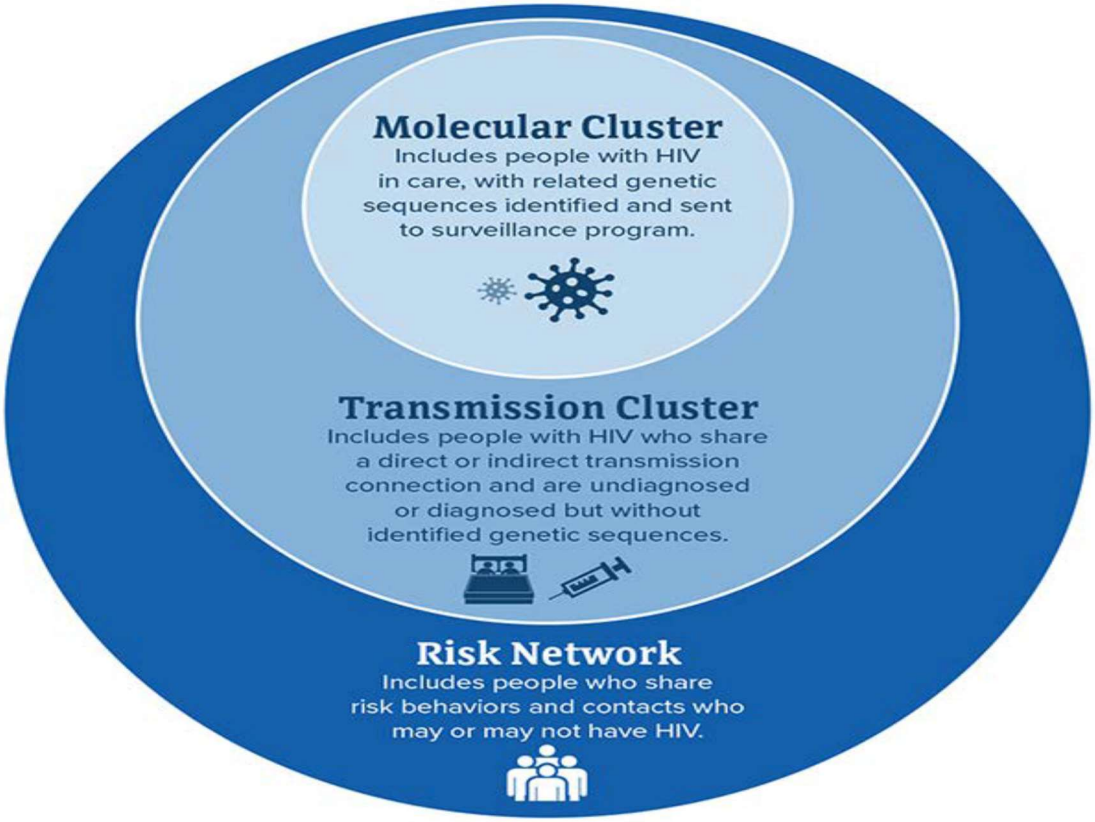
οι PP είναι σημαντικά υψηλότερες από τις αντίστοιχες μη παραμετρικές συχνότητες του bootstrap για τους πραγματικούς κλάδους, αλλά δείχθηκε και ότι λανθασμένα συμπεράσματα μπορούν να προκύψουν με μεγαλύτερη συχνότητα. Αυτά τα λάθη εντείνονται όταν χρησιμοποιούνται μοντέλα νουκλεοτιδικής αντικατάστασης υποπαραμετροποιημένα (όπως το μοντέλο JC). Όταν τα δεδομένα αναλύονται υπό καταλληλότερο μοντέλο (GTR+Γ), το μη παραμετρικό bootstrap είναι πολύ συντηρητικό. Οι PP είναι επίσης συντηρητικές αλλά λιγότερο από ότι η μέθοδος bootstrap, προσεγγίζοντας περισσότερο την πραγματική φύση της σχέσης (Erixon et al., 2003).

ΚΕΦΑΛΑΙΟ 3

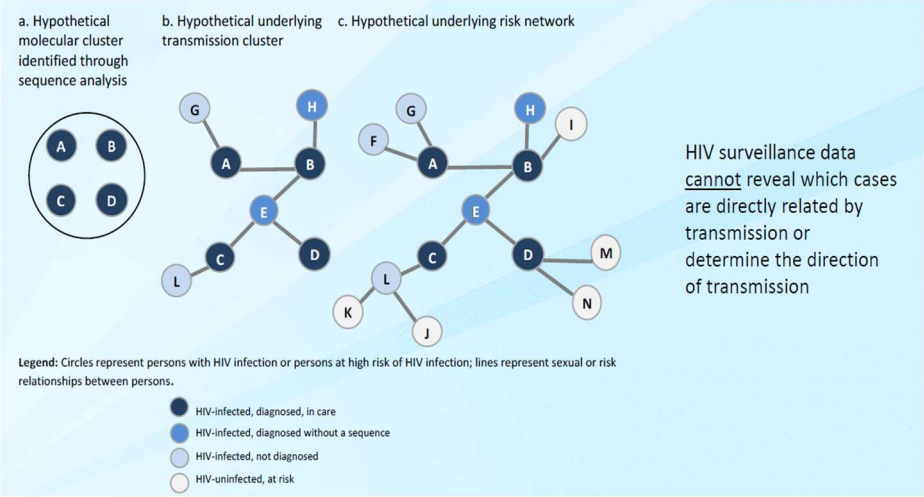
Εργαλεία ταυτοποίησης συστάδων μετάδοσης HIV και φυλογενετική βελτιστοποίηση στρατηγικών πρόληψης του HIV

Ένα κρίσιμο βήμα προς την κατεύθυνση να φέρει το έθνος πιο κοντά στο στόχο της είναι ο εντοπισμός και η ανταπόκριση σε ομάδες του HIV-μολυσμένα άτομα που έχουν επιδημιολογική σύνδεση που σχετίζονται με τη μετάδοση του HIV (δηλαδή, HIV σύμπλεγμα μετάδοσης των ατόμων με διαγνωσμένη ή αδιάγνωστη λοίμωξη HIV). Τα στοιχεία δείχνουν ότι η επιτήρηση του HIV μπορεί να εντοπίσει ομάδες μετάδοσης που διαφορετικά δεν θα αναγνωρίζονταν. Οι πληροφορίες σχετικά με αυτές τις συστάδες μετάδοσης και τα σχετικά δίκτυα κινδύνου μπορούν να μας βοηθήσουν να εστιάσουμε αποδεδειγμένα εργαλεία πρόληψης του HIV εκεί όπου χρειάζονται περισσότερο. Με τον τρόπο αυτό, η εκτεταμένη χρήση της επιτήρησης του HIV έχει τη δυνατότητα να βελτιώσει σημαντικά τις προσπάθειες. (CDC, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention Division of HIV/AIDS Prevention, 2018)

Εικόνα 13. Μοριακό σύμπλεγμα και το υποκείμενο σύμπλεγμα μετάδοσης και το δίκτυο κινδύνου. (CDC, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention Division of HIV/AIDS Prevention, 2018)



Εικόνα 14 Υποθετικό μοριακό σύμπλεγμα (a) και αντίστοιχο υποκείμενο σύμπλεγμα μετάδοσης (b) και δίκτυο κινδύνου (c). (CDC, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention Division of HIV/AIDS Prevention, 2018)



Τι είναι ένα σύμπλεγμα μετάδοσης;

Ένα σύμπλεγμα μετάδοσης αντιπροσωπεύει ένα υποσύνολο ενός **υποκείμενου δικτύου κινδύνου**. Ένα δίκτυο κινδύνου περιλαμβάνει την ομάδα των ατόμων μεταξύ των οποίων η μετάδοση του HIV έχει συμβεί και θα μπορούσε να είναι σε εξέλιξη. Το δίκτυο αυτό περιλαμβάνει άτομα που δεν έχουν προσβληθεί από τον ιό HIV, αλλά ενδέχεται να διατρέχουν κίνδυνο μόλυνσης, καθώς και τα άτομα που έχουν προσβληθεί από τον ιό HIV στο σύμπλεγμα μετάδοσης. Οι συστάδες μετάδοσης παρουσιάζουν ευκαιρίες πρόληψης στο μεγαλύτερο υποκείμενο δίκτυο κινδύνου.

❖ Οι συστάδες μετάδοσης μπορούν να προσδιοριστούν μέσω πολλαπλών μηχανισμών:

- Δεδομένα παρακολούθησης κρουσμάτων HIV
- Υπηρεσίες-εταίροι για τον ιό HIV και έρευνες επικοινωνίας.
- Έξυπνο προσωπικό του υπουργείου Υγείας, πάροχοι φροντίδας ή μέλη της κοινότητας.
- Μοριακά δεδομένα επιτήρησης του HIV.

Η ανάλυση των μοριακών δεδομένων επιτήρησης του HIV μπορεί να εντοπίσει συστάδες περιπτώσεων με στενά συνδεδεμένα στελέχη του HIV (π.χ. μοριακές συστάδες). Η μέθοδος αυτή μπορεί να είναι ιδιαίτερα χρήσιμη για τον εντοπισμό συμπλεγμάτων μετάδοσης που δεν εντοπίζονται μέσω άλλων μηχανισμών.(CDC, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention Division of HIV/AIDS Prevention, 2018)

Τι είναι ένα μοριακό σύμπλεγμα, και πώς σχετίζεται με ένα σύμπλεγμα μετάδοσης;

Η ταυτοποίηση μοριακών συστάδων παρέχει ένα εργαλείο για τον εντοπισμό συστάδων μετάδοσης. Ένα **μοριακό σύμπλεγμα** είναι μια ομάδα ατόμων με διαγνωσμένη λοίμωξη HIV που έχουν γενετικά παρόμοια στελέχη του HIV. Επειδή ο ιός HIV εξελίσσεται

συνεχώς, τα άτομα των οποίων τα ιικά στελέχη είναι γενετικά παρόμοια μπορεί να συνδέονται στενά με τη μετάδοση.

Ένα **μοριακό σύμπλεγμα** περιέχει μόνο εκείνους τους ανθρώπους για τους οποίους τα μοριακά δεδομένα είναι διαθέσιμα και μπορούν να αναλυθούν, και περιέχει ένα υποσύνολο αυτού που είναι πιθανό ένα μεγαλύτερο υποκείμενο σύμπλεγμα μετάδοσης**.**

Οι μοριακές συστάδες εντοπίζονται μέσω της ανάλυσης των δεδομένων μοριακής αλληλουχίας του HIV που παράγονται μέσω των δοκιμών αντοχής στα φάρμακα HIV. Ο έλεγχος αντοχής στα φάρμακα διεξάγεται για τον εντοπισμό μεταλλάξεων που σχετίζονται με την αντίσταση σε αντιρετροϊκά φάρμακα HIV και βοηθά τον πάροχο περίθαλψης HIV να επιλέξει ένα κατάλληλο θεραπευτικό σχήμα. Αυτός ο έλεγχος συνιστάται για όλα τα άτομα με λοίμωξη HIV, με τη σύσταση ότι οι δοκιμές πρέπει να διεξάγονται κατά την είσοδο στη φροντίδα του HIV.

- Ως αποτέλεσμα, οι μοριακές συστάδες περιλαμβάνουν άτομα με διαγνωσμένη λοίμωξη HIV που έχουν εισέλθει στην περίθαλψη και είχαν γενετικές δοκιμές αντοχής, και είχαν αλληλουχίες για ανάλυση.
- Αυτό αντιπροσωπεύει ένα υποσύνολο του υποκείμενου συμπλέγματος μετάδοσης, το οποίο μπορεί επίσης να περιλαμβάνει:
  - Άτομα με διαγνωσμένη λοίμωξη HIV που δεν έχουν διαθέσιμη ακολουθία για ανάλυση, είτε επειδή:
    - Δεν μπήκαν στην περίθαλψη
    - Μπήκαν στην περίθαλψη, αλλά δεν είχαν πραγματοποιήσει έλεγχο γενετικής αντίστασης
    - Μπήκαν στην περίθαλψη και είχαν μια γενετική δοκιμή αντίστασης, αλλά η ακολουθία δεν διαβιβάστηκε στο τμήμα υγείας για ανάλυση, ή ήταν κακής ποιότητας και δεν μπορούσε να αναλυθεί
  - Άτομα με αδιάγνωστη λοίμωξη
- Εκτός από τα πρόσωπα του συμπλέγματος μετάδοσης, το υποκείμενο δίκτυο κινδύνου θα περιλαμβάνει:
  - HIV-αρνητικά άτομα που διατρέχουν κίνδυνο για την απόκτηση του HIV

Τα μοριακά δεδομένα δεν μπορούν να αποκαλύψουν ποιες περιπτώσεις σχετίζονται άμεσα με τη μετάδοση ή να καθορίζουν την κατεύθυνση μετάδοσης. Ο περιορισμός αυτός οφείλεται στο γεγονός ότι δύο άτομα με γενετικά παρόμοια στελέχη του HIV δεν συνδέονται απαραίτητα άμεσα με τη μετάδοση: η σχέση θα μπορούσε να είναι έμμεση, και θα μπορούσαν να υπάρχουν μη αναγνωρισμένα άτομα που εμπλέκονται στις σχέσεις μετάδοσης.

Μόλις εντοπιστεί ένα μοριακό σύμπλεγμα, τα αντίστοιχα δίκτυα μετάδοσης και δίκτυα κινδύνου μπορούν να εντοπιστούν μόνο μέσω διερεύνησης. (CDC, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention Division of HIV/AIDS Prevention, 2018)

Τρεις βασικοί τύποι ορισμού συμπλεγμάτων.

1) Καθαροί ορισμοί φυλογενετικής συστάδας μετάδοσης που βασίζονται αποκλειστικά σε υποστήριξη φυλογενετικών κόμβων
2) Καθαροί ορισμοί που βασίζονται σε αποστάσεις που βασίζονται αποκλειστικά σε γενετικές αποστάσεις ανά ζεύγη pairwise genetic distances
3) Συνδυασμένοι ορισμοί συμπλέγματος μετάδοσης που βασίζονται τόσο στη στήριξη των φυλογενετικών κόμβων όσο και στις γενετικές αποστάσεις ανά ζεύγη

Καμία τυπική προσέγγιση δεν είναι προς το παρόν διαθέσιμη για τον καθορισμό μοριακών δικτύων. Δύο γενικές κατηγορίες προσεγγίσεων έχουν χρησιμοποιηθεί συνήθως ανεξάρτητες ή συνδυασμένες για τον προσδιορισμό μοριακών συστάδων HIV. Η πρώτη είναι μια προσέγγιση βασισμένη στη φυλογένεια, στην οποία αλληλουχίες που μοιράζονται έναν

κοινό πρόγονο ορίζονται ως ένα σύμπλεγμα. Και οι προσεγγίσεις μέγιστης πιθανοφάνειας και Bayesian κατασκευής δέντρου χρησιμοποιούν μοντέλα πιθανότητας για την αξιολόγηση της σχετικής αξιοπιστίας των διαφορετικών φυλογενετικών τοπολογιών, ενώ η προσέγγιση της γειτονικής σύνδεσης (neighbour-joining) χρησιμοποιεί έναν ντετερμινιστικό αλγόριθμο οικοδόμησης-δέντρου που παράγει μόνο μία φυλογενετική τοπολογία. Διάφορες φυλογενετικές μέθοδοι μπορούν να χρησιμοποιηθούν για να συναχθεί ένα φυλογενετικό δέντρο, όπως ένα δέντρο γειτονικής σύνδεσης (NJ), δέντρο μέγιστης πιθανοφάνειας (ML) ή δέντρο μέγιστης αξιοπιστίας κλάδου clade (MCC), υποστηριζόμενο από την τιμή bootstrap, LRtest, zero-branch length test ή εκ των υστέρων πιθανότητα. Το δέντρο NJ βασίζεται σε ένα μοντέλο απόστασης και μπορεί να κατασκευαστεί γρηγορότερα από τα ML δέντρα και MCC. Ως εκ τούτου, έχει συνήθως χρησιμοποιηθεί για την κατασκευή φυλογενετικών δέντρων σε προηγούμενες μελέτες. Τα ML και δέντρα MCC χρησιμοποιούν και τα δύο μοντέλα νουκλεοτιδικών υποκατάστασεων (εξελικτικά μοντέλα) για να αξιολογήσουν τη σχετική πιθανότητα (relative likelihood) διαφορετικών φυλογενετικών τοπολογιών, που προκαλούν υψηλό υπολογιστικό φορτίο. Το δέντρο MCC θα μπορούσε επίσης να επιτρέψει στον τύπο του μοριακού ρολογιού και στο δημογραφικό μοντέλο να εκτιμήσει το χρόνο στον πιο πρόσφατο κοινό πρόγονο (tMRCA), τον εξελικτικό ρυθμό και το πραγματικό (effective) μέγεθος του παρελθόντος πληθυσμού (ο αριθμός των ατόμων σε έναν πληθυσμό που συνεισφέρουν απογόνους στον επόμενη γενιά) με την πάροδο του χρόνου, που μπορεί να αντανακλά την αυξανόμενη ή φθίνουσα δημογραφική ιστορία της ιικής επιδημία [Hui S et al.,2004 ; Yerly S et al.,2001 ; Saitou N et al.,1987] Οι υποκαταστάσεις νουκλεοτιδίων, το μοριακό ρολόι και τα δυναμικά μοντέλα πληθυσμού θα πρέπει να δοκιμαστούν για να προσδιοριστεί ποιο θα ταιριάζει καλύτερα στο σύνολο δεδομένων ακολουθίας στόχου πριν από την ανακατασκευή του εξελικτικού ιστορικού [Baele G et al.,2012 ; Baele G et al.,2013]. Τα λογισμικά πακέτα (BEAST 1 και BEAST 2) χρησιμοποιούνται ευρέως για φυλοδυναμικά και φυλογεωγραφικά συμπεράσματα [Bouckaert R et al.,2019 ; Suchard MA et al.,2018]. Αρκετές πρόσφατες μελέτες χρησιμοποίησαν επίσης ιογενείς αλληλουχίες με χωροχρονικά χαρακτηριστικά για να συμπεράνουν την προέλευση και την εξάπλωση του συμπλέγματος μετάδοσης ή του δικτύου μέσω φυλοδυναμικών και φυλογεωγραφικών προσεγγίσεων [Dennis AM et al.,2019 ; Wilkinson E et al.,2019] Η βασική ιδέα του μοριακού δικτύου είναι να ταξινομηθούν οι ιικές αλληλουχίες σύμφωνα με γενετικές ομοιότητες. Ωστόσο, με την προσέγγιση βασισμένη στη φυλογένεια, μια πολύ αποκλίνουσα αλληλουχία απογόνων δεν μπορεί να αποκλειστεί από τις άλλες με έναν κοινό πρόγονο [Poon AF, 2016], πράγμα που θα μπορούσε να σημαίνει ότι η συλλογή δειγμάτων πολύ μετά τη μετάδοση δεν συνάγει πρόσφατο ενεργό δίκτυο μετάδοσης.

Η άλλη προσέγγιση είναι οι ορισμοί συμπλέγματος που βασίζονται σε GD Genetic Distances (Γενετικές αποστάσεις). Οι Pairwise genetic distances (ζευγαρώδεις γενετικές αποστάσεις) μέσα σε ένα συγκρότημα μετάδοσης περισσοτέρων από δύο αλληλουχιών μπορούν να συνοψιστούν με διάφορους τρόπους, για παράδειγμα χρησιμοποιώντας τη μέση, διάμεση ή την μέγιστη ζευγαρωτή απόσταση. Μια άλλη προσέγγιση είναι να συσχετίσει μια ακολουθία με ένα συγκεκριμένο σύμπλεγμα, εάν η απόσταση από αυτή την ακολουθία σε οποιαδήποτε άλλη ακολουθία σε αυτή τη συστάδα είναι χαμηλότερη από μια τιμή κατωφλίου - ανεξάρτητα από τις αποστάσεις σε άλλες ακολουθίες του συμπλέγματος. Υπάρχουν πλεονεκτήματα και μειονεκτήματα στις διαφορετικές προσεγγίσεις, για παράδειγμα η μέγιστη γενετική απόσταση έχει προταθεί ότι είναι λιγότερο ευαίσθητη στο μέγεθος συμπλέγματος από ορισμούς συμπλέγματος που βασίζονται σε μέσες γενετικές αποστάσεις στις οποίες μία ή λίγες «μη συνδεδεμένες» αλληλουχίες μπορούν να συμπεριλαμβάνονται σε μεγάλες συστάδες επειδή έχουν ελάχιστη επίδραση στη μέση απόσταση. Επιπλέον, οι προσεγγίσεις μέγιστης γενετικής απόστασης είναι γρήγορες στον υπολογισμό και έχει προταθεί να συσχετιστούν με το χρόνο του MRCA των συστάδων σε φυλογένειες μοριακού ρολογιού. Η Pairwise GD υπολογίζεται συνήθως χρησιμοποιώντας το

TN93 μοντέλο υποκατάστασης. Άτομα με απόσταση κατά ζεύγη κάτω από το προκαθορισμένο όριο GD αντιστοιχίζονται στα ίδιες συστάδες [Aldous JL et al.,2012]. Συνιστώνται διάφορα όρια GD βάσει του στόχου της ανάλυσης. Ένα γενετικό όριο 0,5%, με περίπου πέντε διαφορετικά νουκλεοτίδια για ακολουθίες μήκους 1000 νουκλεοτιδίων, προτείνεται για τον εντοπισμό περιπτώσεων που σχετίζονται με πρόσφατη και ταχεία εξάπλωση. Αυτό το όριο αντιστοιχεί σε περίπου 2-3 χρόνια ανεξάρτητης ιογενούς εξέλιξης [National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention 2018]. Εάν ο στόχος είναι να προσδιοριστούν όλες οι πιθανές περιπτώσεις που πιθανώς σχετίζονται με μια δεδομένη περίπτωση, ένα μεγαλύτερο όριο GD 1,5% που αντιστοιχεί σε ένα μέγιστο 7-8 ετών διαχωριστικών στελεχών ιογενούς εξέλιξης προτείνεται από τις οδηγίες CDC των ΗΠΑ [National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention 2018]. Το HIVTRACE είναι ένα οπτικό λογισμικό με βάση την απόσταση που χρησιμοποιείται για την κατασκευή μοριακών δικτύων (http: //hivtrace.datamonkey.Org / hivtrace) και έχει εφαρμοστεί στις ΗΠΑ και σε αρκετές ασιατικές χώρες [Chin BS et al.,2016 ; Wang X et al.,2015] .

Οι μέθοδοι με βάση τη GD και με βάση τη φυλογένεια δεν είναι ούτε καλές ούτε κακές. Ωστόσο, είναι απαραίτητη η επιλογή της κατάλληλης μεθόδου με βάση τα χαρακτηριστικά ακολουθίας και τους ερευνητικούς στόχους. Το λογισμικό HIV-TRACE τείνει να ανιχνεύει μεγαλύτερα και λιγότερα σμήνη από το Cluster Picker, το οποίο ανιχνεύει περισσότερα σμήνη που περιέχουν μόνο δύο αλληλουχίες [Rose R et al.,2017]. Όταν ο στόχος είναι η ανίχνευση μεγαλύτερων δικτύων σε μια βαθιά περιοχή δειγματοληψίας, το HIV-TRACE μπορεί να αποδώσει ευνοϊκότερα και αναμένεται να εντοπίσει περισσότερες αλυσίδες μετάδοσης [Rose R et al.,2017]. Η προσέγγιση που βασίζεται στη GD μπορεί να χρησιμοποιηθεί για πιο γρήγορο υπολογισμό, αλλά δεν μπορεί να διακρίνει διαφορετικούς ρυθμούς εξέλιξης και μπορεί να υποτιμήσει τον χρόνο απόκλισης του ιού. Επιπλέον, η GD συνδέεται στενά με την πιθανή εξελικτική απόσταση, γεγονός που καθιστά την προσέγγιση δημοφιλή για την ανοικοδόμηση μοριακών δικτύων σε πραγματικό χρόνο και την παρακολούθηση δυναμικών τάσεων [Little SJ et al.,2014]. Όσον αφορά την προσέγγιση φυλογένειας για την κατασκευή μοριακού δικτύου, η απόκτηση υψηλής τιμής υποστήριξης κόμβου και σταθερών (steady) τοπολογιών είναι δύσκολη κατά την επεξεργασία μεγάλου αριθμού αλληλουχιών. Συγκεκριμένα λογισμικά, όπως το PhyloPart [Prosperi MC et al.,2011] και το Cluster Picker [Ragonnet-Cronin M et al.,2013], μπορούν να συνδυάσουν φυλογενετικά δέντρα εκκίνησης bootstraps και GDs για τον εντοπισμό συστάδων μετάδοσης. Όποια και αν είναι η εγγενής μεροληψία (inherent bias) της γενετικής μεθόδου ομαδοποίησης, η ταχεία διαδοχή νέων μολυσμένων ατόμων σε ένα προκαθορισμένο σύμπλεγμα υποδηλώνει τοπικό ξέσπασμα μόλυνσης από HIV [Hassan AS et al.,2017].

Μοριακή δίκτυα οδηγός στοχευμένη παρέμβαση

Πρόσφατα, αρκετές μελέτες έχουν χρησιμοποιήσει φυλογενετικές αναλύσεις για να αξιολογήσουν πώς οι στρατηγικές πρόληψης του HIV μπορούν να βελτιστοποιηθούν. Τα δίκτυα μετάδοσης του ιού HIV διευκρίνισαν την εξάπλωση του HIV μεταξύ πληθυσμού και προσφέρουν τη δυνατότητα παρέμβασης, οι οποίες είναι παραδοσιακά προσδιοριζόμενες από την επιτήρηση του ιού HIV, υπηρεσίες συντρόφων και έρευνες επαφών. Η ανάλυση μοριακού δικτύου είναι συμπληρωματική των υφιστάμενων partnerships που βασίζονται στα κοινωνικά HIV δίκτυα ή τα δίκτυα μετάδοσης και συμβάλλει στην προώθηση ειδοποιήσεων συντρόφων [Smith DM et al.,2009], γεφυρώνοντας έτσι το προηγουμένως μη αναγνωρισμένο στοιχείο δικτύου ειδοποίησης συντρόφων. Αυτή η ανάλυση παρέχει πιο αξιόπιστα στοιχεία από τη ονομασία συντρόφου για τον εντοπισμό πιθανών συνδέσμων μετάδοσης. [Pasquale DK et al.,2018 ; Avila D et al.,2014 ; Wertheim JO et al.,2017 ; Kostaki EG et al.,2018]

Μία μελέτη στην Ελβετία συνδύασε αναλύσεις φυλογενετικής και λανθάνουσας τάξης [Lazarsfeld PF, Henry NW.,1968 ; McLachlan G, Peel D,2000].για την κατανόηση της

μετάδοσης του HIV-1 Η ανάλυση λανθάνουσας τάξης πραγματοποιήθηκε για τον εντοπισμό ομάδων ασθενών με κοινωνικόδημογραφικά και συμπεριφοριστικά χαρακτηριστικά.

Προδιαγραφή του μοντέλου λανθάνουσας τάξης

Για ένα σύνολο q κατηγορικών δηλωτικών manifest μεταβλητών Y_l με κατηγορίες c_l (l= 1, ..., q), ένα μοντέλο λανθάνουσας τάξης με g λανθάνουσες τάξεις μπορεί να προσδιοριστεί ως εξής:

$$P(Y = y) = \sum_{c=1}^{g} \pi_c \prod_{l=1}^{q} \varphi_{cly_l}$$

Όπου,

- Το Y είναι διάνυσμα που περιέχει τα στοιχεία Y_l
- Η π_c αντιπροσωπεύει τον επιπολασμό prevalence της τάξης c
- Η φ_{cly_l} αντιπροσωπεύει την πιθανότητα ότι το Y παίρνει την κατηγορία y_l δεδομένης της κατηγορίας c (c=1, …, g, l= 1, …, q, y_l= 1, …, c_l).

Έτσι, εντός των τάξεων, οι μεταβλητές θεωρούνται ανεξάρτητες (υπόθεση της τοπικής ανεξαρτησίας) και να κατανέμεται σύμφωνα με τις πολυωνυμικές πιθανότητες φ_{cly_l}.

Συμπερίληψη των ακόλουθων 10 κατηγορηματικών μεταβλητών:

1. Φύλο (αρσενικό, θηλυκό)

2. Ηλικία (<25, 25–34, 35–44, 45–54, ≥55 ετών)

3. Περιοχή προέλευσης (Ελβετία και Βορειοδυτική Ευρώπη, Νότια Ευρώπη, Υποσαχάρια Αφρική, Λατινική Αμερική, Ασία και Ανατολική Ευρώπη)

4. Επίπεδο εκπαίδευσης (υποχρεωτική εκπαίδευση, επαγγελματική κατάρτιση, τριτοβάθμια εκπαίδευση)

5. Επάγγελμα (αυτοαπασχολούμενος, μαθητευόμενος ή ασκούμενος, ανώτατα διοικητικά στελέχη, μεσαία ή επίπεδο διαχείρισης, εργαζόμενος, νοικοκυρά/ νοικοκυρά)

6. Κύρια πηγή εισοδήματος (μισθωτή εργασία, στήριξη από οικογένεια ή σύντροφο, παροχές κοινωνικής πρόνοιας)

7. Σεξουαλική προτίμηση (ετεροφυλόφιλοι, αμφιφυλόφιλοι, ομοφυλόφιλοι)

8. Σεξουαλικές επαφές (κανένας σύντροφος, απροστάτευτο σεξ με σταθερό σύντροφο, προστατευμένο σεξ με σταθερό σύντροφο, απροστάτευτο σεξ με περιστασιακό σύντροφο, προστατευμένο σεξ με περιστασιακό σύντροφο)

9. Ιστορικό χρήσης ναρκωτικών ένεση (ποτέ, ποτέ, τρέχουσα)

10. Κατανάλωση οινοπνεύματος (σοβαρή, μέτρια ή ελαφριά σύμφωνα με την Παγκόσμιος Οργανισμός Υγείας (3) [World Health Organization,2004])

Η μεταβλητή 8 είναι ένας συνδυασμός απαντήσεων σε δύο ερωτήσεις, μία για τη χρήση προφυλακτικού (πάντα, μερικές φορές, ποτέ, δεν απαντά) και ένα για το είδος των εταίρων (σταθερή, περιστασιακή). Σεξουαλικές επαφές ορίστηκαν ως απροστάτευτες εάν η χρήση προφυλακτικού αναφέρθηκε ως μερικές φορές ή ποτέ. Για όλες τις ερωτήσεις, οι κατηγορίες απόκρισης "καμία απάντηση" ή "άλλη" επανακωδικοποιήθηκαν σε ελλείπουσες τιμές.

Τοποθέτηση του μοντέλου λανθάνοντος τάξης

Τα μοντέλα τοποθετήθηκαν με τη χρήση του λογισμικού Mplus έκδοση 6.1 [Muthén BO, Muthén LK.,1998] με τον αριθμό των κατηγοριών g να κυμαίνεται μεταξύ 1 και 10. Η προσαρμογή του μοντέλου βασιζόταν στη μέγιστη πιθανοφάνεια χρησιμοποιώντας τον αλγόριθμο μεγιστοποίησης προσδοκιών (EM)[Dempster AP, Laird NM, Rubin DB,1977]. Δεν αποκλείστηκαν ασθενείς με ελλείπουσες τιμές σε οποιαδήποτε από τις μεταβλητές. Ο αλγόριθμος EM επιτρέπει σε κάποιον να

ενσωματώσει τα διαθέσιμα δεδομένα από ασθενείς με την παραδοχή ότι οι ελλείπουσες τιμές λείπουν τυχαία [Little RJA, Rubin DB,2002].

Χρήση του Bayesian Κριτήριου Πληροφοριών BIC για την επιλογή του αριθμού των τάξεων, εξισορροπώντας έτσι την φειδωλότητα και την προσαρμογή του μοντέλου [McLachlan G, Peel D,2000]. Αναφορά επίσης του κριτήριου πληροφόρησης Akaike AIC και της εντροπίας [Ramaswamy V et al.,1993] για κάθε μοντέλο.

Για κάθε ασθενή, υπολογισμός των εκ των υστέρων πιθανοτήτων να ανήκουν στις διαφορετικές λανθάνουσες τάξεις ενός προσαρμοσμένου μοντέλου. Διάθεση ατόμων στις ομάδες για τις οποίες είχαν την υψηλότερη εκ των υστέρων πιθανότητα ιδιότητας μέλους [McLachlan G, Peel D,2000].

<u>Συμφωνία Concordance των ομάδων κοινωνικο-συμπεριφοράς και των φυλογενετικών συστάδων</u>

Συγκέντρωση διασταυρούμενων (cross-tabulated) ομάδων κοινωνικο-συμπεριφορικής ομάδας (στήλες) έναντι των φυλογενετικών SHCS συστάδων (σειρές) και ανάλυση του πίνακα με δύο τρόπους. Καταρχάς, υπολογισμός συντελεστών συσχέτισης Pearson ανά μεταξύ στηλών, δηλαδή τη συσχέτιση μεταξύ κάθε ζεύγος ζευγαριού ομάδων κοινωνικο-συμπεριφοράς σχετικά με τη συχνότητα με την οποία εμφανίζονται στις διάφορες ομάδες. Μια ουσιαστική συσχέτιση μεταξύ δύο ομάδων υποδεικνύει ότι οι ασθενείς από αυτές τις ομάδες εμφανίζονται στα ίδια φυλογενετικά συμπλέγματα, υποδηλώνοντας μεταξύ των ομάδων μετάδοση. Οι συντελεστές έως 0,4 θεωρήθηκαν ότι αντικατοπτρίζουν αδύναμους συσχετισμούς, συντελεστές μεταξύ 0,4 και 0,7 μέτριων συσχετίσεων και συντελεστές άνω των 0,7 ισχυρών συσχετίσεων.

Δεύτερον, χρήση της Multiple Correspondence Ανάλυσης για την απόκτηση μιας δισδιάστατης αναπαράστασης στηλών και σειρών και των αλληλεξαρτήσεων τους [Greenacre M.,1984]. Στα γραφήματα Multiple Correspondence Ανάλυσης, οι ομάδες κοινωνικο-συμπεριφοράς βρίσκονται κοντά όταν τα μέλη τους εμφανίζονται στις ίδιες φυλογενετικές συστάδες, ενώ οι φυλογενετικοί συστάδες τοποθετούνται το ένα κοντά στο άλλο εάν έχουν παρόμοια σύνθεση ομάδων κοινωνικο-συμπεριφοράς [Greenacre M.,1984]. Στην πλήρους διαστάσεων αναπαράσταση της Multiple Correspondence Ανάλυσης, από την οποία προβλήθηκαν μόνο οι δύο κύριες διαστάσεις, οι φυλογενετικές συστάδες τοποθετούνται στο κέντρο βάρους των ομάδων κοινωνικο-συμπεριφοράς από τις οποίες αποτελούνται, με τις ομάδες κοινωνικο- συμπεριφοράς να έχουν βάρη ίση με την αναλογία τους στο φυλογενετικό σύμπλεγμα. Πραγματοποιήθηκαν αναλύσεις για όλους τους υπότυπους, τους υπότυπους Β και τους μη υπότυπους Β.

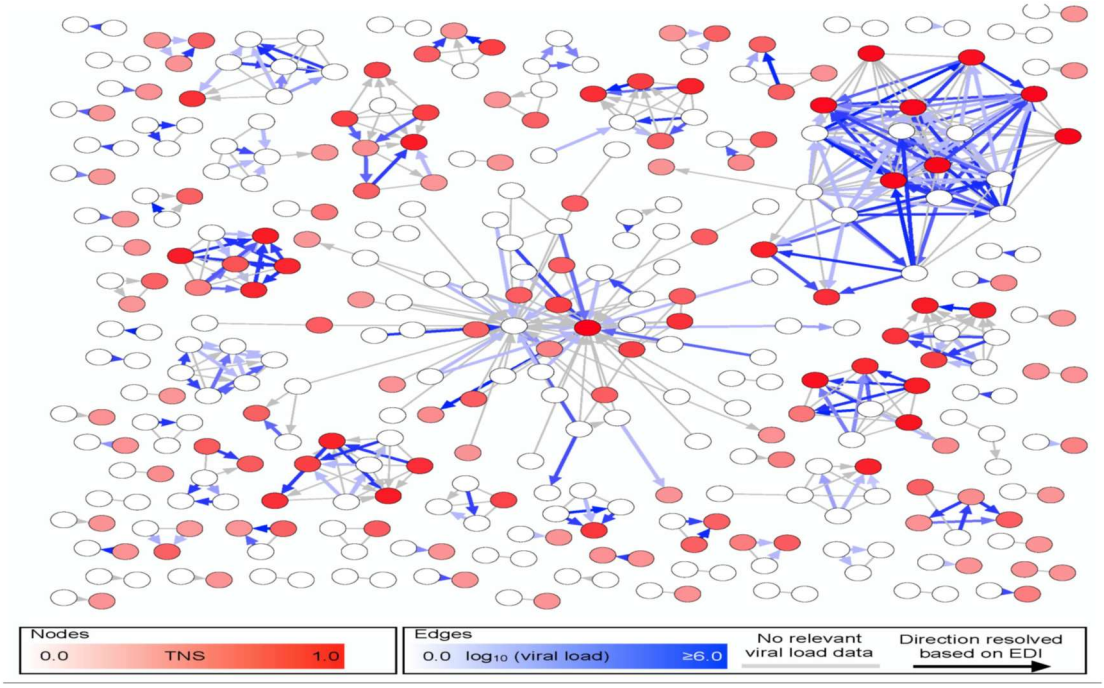
Το 2009, οι Smith et al. εισήγαγαν ένα επιστημονικό μοντέλο σχεδιασμένο για τη μελέτη της μοριακής παρακολούθησης της μετάδοσης του HIV χρησιμοποιώντας πληροφορίες δημόσιας υγείας [Smith DM et al.,2009]. Βρήκαν ότι η μοριακή επιδημιολογία σε συνδυασμό με ιχνηλάτηση επαφών συντρόφων μπορεί να χρησιμοποιηθεί για την ταυτοποίηση ατόμων εντός ενός πληθυσμού που ανήκουν σε πολύ σχετικές ομάδες μετάδοσης του ιού HIV. Αυτές οι μέθοδοι θα μπορούσαν να χρησιμοποιηθούν για την εφαρμογή επιλεκτικά στοχευμένων προληπτικών παρεμβάσεων. Ορισμένες παράμετροι έχουν χρησιμοποιηθεί για τον ποσοτικό προσδιορισμό του κινδύνου των ατόμων μεταξύ των μοριακών δικτύων και για την καθοδήγηση της στοχευμένης παρέμβασης. Ένας δείκτης ονομάζεται σύνδεσμος ή βαθμός. Όσο περισσότερους συνδέσμους έχουν τα άτομα στο δίκτυο, τόσο περισσότερους πιθανούς συντρόφους μετάδοσης θα μπορούσαν να έχουν και τόσο μεγαλύτερος είναι ο κίνδυνος επικοινωνίας. Οι Oster et al. χρησιμοποίησαν τον σύνδεσμο ως δείκτη σε μια μελέτη για τις ομάδες με υψηλούς κινδύνους και μεταξύ ομάδων με διαφορετικές φυλετικές / εθνοτικές καταστάσεις στο Εθνικό Σύστημα Εποπτείας του HIV στις ΗΠΑ. Οι μολυσμένες με HIV ετεροφυλόφιλες γυναίκες συνδέονταν κυρίως με MSM. Οι

παρεμβάσεις που μπόρεσαν να μειώσουν τις μεταδόσεις HIV μεταξύ ατόμων στον MSM πληθυσμό έδειξαν μεγάλες δυνατότητες μείωσης της απόκτησης HIV, καθώς και μεταξύ άλλων ομάδων με υψηλούς κινδύνους [Oster AM et al.,2015]. Οι Leigh Brown et al. χρησιμοποίησαν βαθμούς για να κατηγοριοποιήσουν τη μόλυνση από τον ιό HIV σε μια φυλοδυναμική ανάλυση σχετικά με τους MSM στο Ηνωμένο Βασίλειο και έδειξαν την προτιμησιακή συσχέτιση των MSM του Ηνωμένου Βασιλείου και ζήτησαν παρέμβαση με στόχο άτομα υψηλού βαθμού [Leigh Brown AJ et al.,2011]. Μια "what-if" προσέγγιση, η οποία ανασυντάσσει ένα παρελθόν δίκτυο μετάδοσης και στη συνέχεια αναλύει πόσες μολύνσεις θα μπορούσαν να αποφευχθούν εάν μια συγκεκριμένη στρατηγική πρόληψης του HIV είχε εφαρμοστεί σε ένα δεδομένο χρονικό σημείο στο παρελθόν, πραγματοποιήθηκε από μια ερευνητική ομάδα του Σαν Ντιέγκο ανέπτυξε μια παράμετρο που ονομάζεται σκορ δικτύου μετάδοσης (TNS) για να εκτιμήσει τον κίνδυνο μετάδοσης του HIV από ένα νεοδιαγνωσθέν άτομο στον σύντροφό του. Σε αυτήν την αναδρομική ανάλυση προσομοίωσης, διαπίστωσαν ότι σε σύγκριση με τα συγκεντρωμένα άτομα από ένα τυχαία επιλεγμένο υποσύνολο, η ART που στοχεύει άτομα με το υψηλότερο TNS έδειξε ένα σημαντικά μειωμένο HIV επίπεδο μετάδοσης σε δίκτυο [Little SJ et al.,2014]. Ορισμός του TNS ως τη συνάρτηση του συνολικού βαθμού (d) του κόμβου κατά την έναρξη, που βασίζεται στο δίκτυο που είναι γνωστό κατά τη στιγμή της εγγραφής του υποκειμένου (N). Συγκεκριμένα,

$$TNS(d, N) = Prob(\beta\alpha\theta\mu\acute{o}\varsigma\ \epsilon\nu\acute{o}\varsigma\ \kappa\acute{o}\mu\beta ou\ \sigma\tau o\ N \leq\ d)$$

με την πιθανότητα να υπολογίζεται χρησιμοποιώντας την καλύτερη προσαρμογή παραμετρικής πυκνότητας για το δίκτυο N. Έτσι, ένα TNS 0,92 σημαίνει ότι ο βαθμός ενός συγκεκριμένου κόμβου βρίσκεται στο 8% όλων των ατόμων που συγκεντρώνονται στο υπάρχον δίκτυο.

Εικόνα 15 *Το συμπεραινόμενο δίκτυο μετάδοσης (εξαιρουμένων των μη συνδεδεμένων ατόμων) στο SDPIC. Εμφανίζονται μόνο άτομα μέσα σε σύμπλεγμα (κόμβοι) εντός του δικτύου (52,3%). Παρά την πιθανή παρουσία κόμβων χωρίς δείγμα (δηλαδή, κόμβων που να λείπουν), ένα μερικό δίκτυο μετάδοσης HIV-1 είναι χρωματικά κωδικοποιημένο. Η ένταση του χρωματισμού των κόμβων καθορίζεται από τη βαθμολογία TNS τους, ενώ αυτή για τις κατευθυνόμενες άκρες αντιστοιχεί στο ιικό φορτίο του υποτιθέμενου αρχικού συντρόφου στο χρονικό σημείο που βρίσκεται πιο κοντά στο συμβάν μετάδοσης. Η απουσία μπλε σκίασης υποδηλώνει ότι κανένα VL δεν ήταν διαθέσιμο για το δειγματισμένο άτομο σε οποιοδήποτε χρονικό σημείο ή ότι η κατεύθυνση του άκρου δεν ήταν δυνατό να εξακριβωθεί χρησιμοποιώντας EDI. Η απουσία κόκκινης σκίασης υποδηλώνει ένα TNS = 0 (δηλ. κόμβους που ήταν μη συνδεδεμένοι κατά τη στιγμή της εγγραφής). Οι κόμβοι συνδέονται με ένα άκρο (δηλαδή, μια γραμμή που υποδεικνύει πιθανή μετάδοση) εάν η ελάχιστη απόσταση μεταξύ των αντίστοιχων pol αλληλουχιών (δηλαδή, πιθανά ζεύγη μετάδοσης) είναι μικρότερη από 1,5%. Μια κατεύθυνση εκχωρείται σε ένα άκρο εάν το EDI για το δευτερεύων σύντροφο (δηλαδή, ο υποτιθέμενος «παραλήπτης») είναι τουλάχιστον 30 ημέρες μετά την ημερομηνία δειγματοληψίας του υποτιθέμενου εταίρου διαβίβασης (δηλαδή, υποτιθέμενη «πηγή»). Η κατεύθυνση της μετάδοσης επιλύθηκε σε 332 από τα 540 άτομα (61,5%).* [**Πηγή**: (Little SJ et al., 2014) doi: 10.1371 / journal.pone.0098443.g001]



Το αποτέλεσμα της στοχευμένης ART που βασίζεται στο TNS υποστηρίχθηκε περαιτέρω από μια άλλη μελέτη σχετικά με κινέζους MSM. Υψηλά επίπεδα με βάση το TNS ART προσομοιώθηκαν και συγκρίθηκαν με τον αριθμό των κυττάρων T CD4, μια στρατηγική που βασίζεται σε ιικό φορτίο και μια στρατηγική θεραπείας όλων των MSM πρωταρχικά (primary) μολυσμένων με HIV σε μια κοορτή στο Πεκίνο. Τα αποτελέσματα έδειξαν ότι η άμεση θεραπεία μετά τη διάγνωση θα μπορούσε να είχε αποτρέψει 31 από τις 134 (23%) νέες μολύνσεις το 2010. Η θεραπεία με μικρότερο αριθμό κυττάρων CD4 + ή υψηλότερο ιικό φορτίο εμπόδισε το 10-18% των νέων μολύνσεων. Η αποτελεσματικότητα πρόληψης του με υψηλή TNS ART κυμαινόταν μεταξύ 30% και 42%, η οποία ήταν σημαντικά υψηλότερη από τις άλλες τρεις στρατηγικές που αξιολογήθηκαν. Αυτή η μελέτη υπονοούσε ότι οι στρατηγικές που βασίζονται στο TNS μπορεί να είναι ένας αποτελεσματικός τρόπος παροχής προληπτικών παρεμβάσεων [Wang X et al.,2015].

<u>Επιπτώσεις της στοχοθετημένης πρόληψης</u> [Wang X et al.,2015].

1) Καθορισμός του εάν θα ήταν εφικτή η χρήση δεδομένων δικτύου ή κλινικών δεδομένων για την αποτελεσματική στοχοθετημένη παρέμβαση πρόληψης.
    1) Εξαίρεση αλληλουχιών ατόμων με άγνωστη ημερομηνία συλλογής (n = 16).
    2) Διαχωρισμός της κοόρτης σε διαγνωσμένους πριν και μετά το 2010.
    3) Ανάλυση των κλινικών χαρακτηριστικών και των χαρακτηριστικών του δικτύου μεταξύ των συμμετεχόντων στην κοορτη μεταξύ του 2007 και του 2009 για την πρόβλεψη του αριθμού των λοιμώξεων που θα είχαν προληφθεί το 2010, εάν η παρέμβαση ήταν στοχοθετημένη με βάση αυτά τα χαρακτηριστικά.
2) Βασικοί κλινικοί παράγοντες που χρησιμοποιήθηκαν για την προσομοίωση της στοχοθετημένης παρέμβασης πρόληψης ήταν:

μετρήσεις CD4

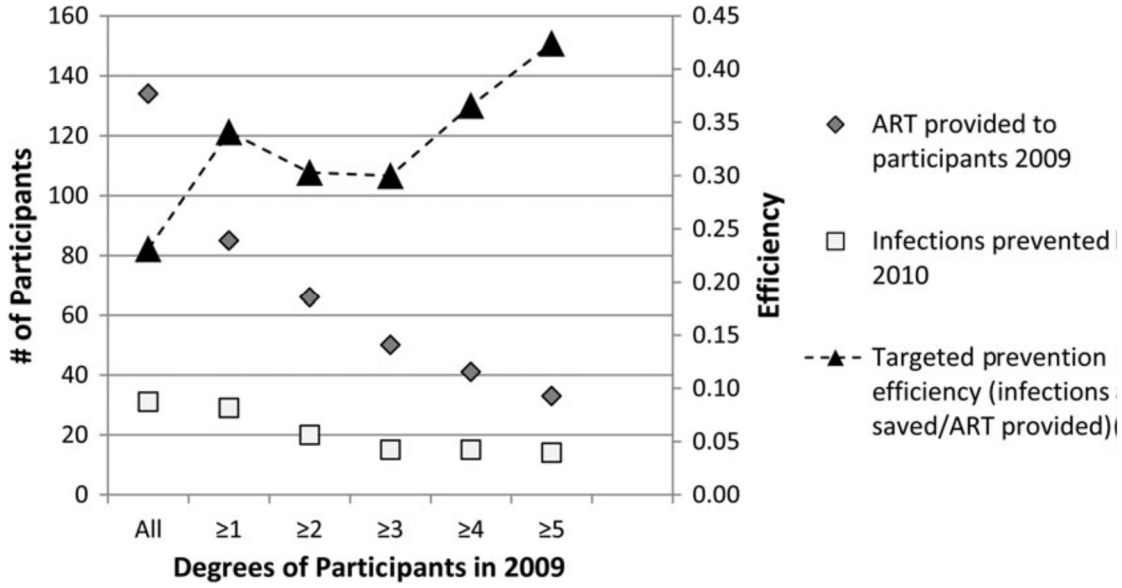
- <200 κύτταρα / mL,
- <350 κύτταρα / mL
- <500 κύτταρα / mL

και ιικά φορτία HIV

- 100.000 αντίγραφα / mL
- 50.000 αντίγραφα / mL.

3) Εξέταση επίσης της αποτελεσματικότητας χρήσης συνδεσιμότητας δικτύου (network connectivity) για την στοχευμένη παρέμβαση πρόληψης με την εκτίμηση του αριθμού των λοιμώξεων που θα είχαν προληφθεί το 2010, αν οι συμμετέχοντες στην κοορτή είχαν λάβει την παρέμβαση με βάση τον αριθμό των συνδέσεων (π.χ., βαθμοί) που είχαν το 2009.
4) Στη συνέχεια, θεώρηση ενός εύρους αποτελεσματικότητας της παρέμβασης πρόληψης μεταξύ 30% και 100% για όλα τα άτομα με τον στοχοθετημένο παράγοντα σε κάθε συστάδα μέχρι το 2009.
5) Καθορισμός του αριθμού των συστάδων που καλύφθηκαν πλήρως από την παρέμβαση πρόληψης πριν από 2010 και αν ένα διαγνωσμένο άτομο το 2010 ανήκε σε ένα τέτοιο σύμπλεγμα, θεώρηση τότε ότι η μόλυνση θα είχε αποφευχθεί.
6) Στη συνέχεια υπολογισμός πόσων νέων συσταδοποιημένων μολύνσεων το 2010 θα είχαν αποτραπεί με βάση την αποτελεσματικότητα της παρέμβασης. Για κάθε κλινικό ή παράγοντα δικτύου που χρησιμοποιήθηκε για τις στοχευμένες παρεμβάσεις, υπολογισμός της αποτελεσματικότητας της παρέμβασης πρόληψης ως ο αριθμός των μολύνσεων που παρεμποδίστηκαν ανά αριθμό ατόμων που έλαβαν την παρέμβαση.

Εικόνα 16 *Στοχευμένη Αντιρετροϊκή θεραπεία (ART) με βάση τη συνδεσιμότητα δικτύου. Εδώ, δείχνουμε ότι όταν η ART στοχεύει σε άτομα (αριστερός άξονας y) με υψηλότερο αριθμό συνδέσεων το 2009 (άξονας x) αυξάνεται η στοχευμένη απόδοση πρόληψης (δεξί άξονας y) της ART. Για παράδειγμα, αν και στα 134 άτομα παρέχονταν ART το 2009 (◊), τότε 31 νέες μολύνσεις θα είχαν προβλεφθεί ότι θα προληφθούν το 2010 (□), με στοχευμένη αποτελεσματικότητα πρόληψης (▲) 23%. Εάν η ART είχε παρασχεθεί μόνο σε αυτά τα 33 άτομα με* ≥ 5 *συνδέσεις το 2009, τότε 14 μολύνσεις θα είχαν προβλεφθεί ότι θα προληφθούν το 2010 με στοχευμένη αποτελεσματικότητα πρόληψης 42%.* [**Πηγή:** Wang X et al., 2015]



Η "what-if" προσέγγιση χρησιμοποιήθηκε επίσης και σε μία μελέτη στην Γαλλία όπου και αυτή περιλάμβανε μόνο προσφάτως μολυσμένους ασθενείς και ανέφερε ότι η άμεση θεραπεία μετά τη διάγνωση θα είχε αποτρέψει 60 από 143 λοιμώξεις (42%) [Chaillon A et al.,2017].

Αυτές οι μελέτες υποδεικνύουν ότι η άμεση θεραπεία μετά τη διάγνωση δεν θα είναι αρκετή για να περιορίσει έντονα την επιδημία του ιού HIV. Οι μελέτες δεν εξέτασαν την επίδραση της πιο συχνού ελέγχου για τον ιό HIV για τον εντοπισμό ασθενών νωρίτερα κατά τη διάρκεια της λοίμωξης.

Εκτός από τις στοχευμένες παρεμβάσεις που βασίζονται σε ατομικό κίνδυνο μετάδοσης HIV, ο ρυθμός μετάδοσης έχει χρησιμοποιηθεί για τον εντοπισμό πρόσφατων και ταχέως αναπτυσσόμενων συστάδων και την ιεράρχηση των απαντήσεων (responses) στη δημόσια υγεία. Μια μελέτη από το CDC των ΗΠΑ εντόπισε 60 συστάδες από 27 δικαιοδοσίες επιτήρησης στο Εθνικό Σύστημα Παρακολούθησης του HIV από το 2013 έως το 2017. Ο ρυθμός μετάδοσης 11 συμπλεγμάτων ήταν έντεκα φορές υψηλότερος από αυτόν των εθνικών εκτιμήσεων για 4/100 άτομα-έτη και θα έπρεπε να δοθεί προτεραιότητα στη παρέμβαση δημόσιας υγείας [France AM et al.,2018]. Εκτός από τις παραμέτρους του μοριακού συμπλέγματος, το υψηλό ιικό φορτίο μπορεί επίσης να λειτουργήσει ως δείκτης στοχευμένων παρεμβάσεων. Σύμφωνα με την τελευταία έρευνα σχετικά με το Εθνικό Σύστημα Παρακολούθησης HIV του Ηνωμένου Βασιλείου, τα συχνά μεταδιδόμενα στελέχη βρέθηκαν σε μεγάλες μοριακές συστάδες και είχαν σημαντικά υψηλότερα ιογενή φορτία και αυξημένη συνδεσιμότητα στο δίκτυο και ότι αυτές οι ομάδες πρέπει να έχουν την υψηλότερη προτεραιότητα για δημόσιας υγείας παρεμβάσεις για διακοπή της μετάδοσης Wertheim JO et al.,2019].

Οι έρευνες μοριακών συστάδων θα μπορούσαν επίσης να βοηθήσουν στην αποκάλυψη διαγνωσμένων ή μη διαγνωσμένων περιπτώσεων μολυσμένων με HIV χωρίς συνδέσεις με ιατρική περίθαλψη και μη μολυσμένα περιστατικά με πολύ υψηλό κίνδυνο μόλυνσης. Ένα εξαιρετικό παράδειγμα προσέγγισης άμεσης παρέμβασης εφαρμόστηκε στον Καναδά. Η προσέγγιση άμεσης παρέμβασης είναι μια άμεση παρέμβαση στην οποία επινοείται μια στρατηγική πρόληψης του HIV σχετικά με τις φυλογενετικές αναλύσεις που πραγματοποιήθηκαν σε πραγματικό χρόνο. Σε αυτή τη μελέτη, ένα αυτοματοποιημένο φυλογενετικό σύστημα εντόπισε ένα πρόσφατο HIV ξέσπασμα του και υποστήριξε μια ενισχυμένη παρακολούθηση της δημόσιας υγείας για να εξασφαλίσει τη σύνδεση με την έναρξη της φροντίδας και της θεραπείας στον προσβεβλημένο υποπληθυσμό. Μείωση της προς τα εμπρός της μετάδοσης ανθεκτικότητας στα φάρμακα παρατηρήθηκε κατά τη διάρκεια της παρακολούθησης [Poon AFY et al.,2016]. Η μελέτη περιελάμβανε όλα τα άτομα που συμμετείχαν στο πρόγραμμα θεραπείας της Βρετανικής Κολομβίας (BC Drug Treatment Program (DTP)). Η μελέτη διεξήγαγε φυλογενετική ανάλυση γονότυπων HIV ρουτίνας που συλλέχθηκαν στην κλινική περίθαλψη. Τα αποτελέσματα της φυλογενετικής ανάλυσης συνοψίζονται σε μηνιαίες και τριμηνιαίες αναφορές σχετικά με την ανάπτυξη και τα χαρακτηριστικά των ομάδων στις οποίες νέες περιπτώσεις εντοπίστηκαν κατά την περίοδο αναφοράς. Τον Ιούνιο του 2014, διαπίστωσαν ότι ένα συγκεκριμένο σύμπλεγμα είχε αυξηθεί κατά 11 νέους ασθενείς σε 3 μήνες σε σχέση με τις αντίστοιχες περιόδους δειγματοληψίας. Μια πρόσθετη νέα περίπτωση που ήταν μέρος αυτού του συμπλέγματος εντοπίστηκε τον Ιούλιο. Είναι σημαντικό ότι εννέα από αυτούς τους 12 (75% των ομάδων) ασθενείς έφεραν τη μετάλλαξη K103N που προσδίδει αντίσταση σε νουκλεοτιδικους αναστολείς της μη ανάστροφης μεταγραφάσης. Οι αρχές δημόσιας υγείας ξεκίνησαν μια ενισχυμένη παρακολούθηση των 9 ατόμων που μετέφεραν τη μετάλλαξη K103N. Η παρακολούθηση εξασφάλισε ότι η θεραπεία ξεκίνησε αμέσως και κάθε ειδοποίηση συνεργαζόμενου συντρόφου ολοκληρώθηκε. Μετά την έναρξη της ενισχυμένης παρακολούθησης, δεν εντοπίστηκαν νέοι ασθενείς. Εντούτοις, 12 νέοι ασθενείς σταδιακά ήρθαν στο φως το 2015. Η αντοχή στα φάρμακα μειώθηκε σημαντικά από 75% των ομάδων ασθενών σε 25% των ομάδων ασθενών (τρεις από τους 12 νεοδιαγνωσθέντες ασθενείς) πριν και μετά την απόκριση στη δημόσια υγεία, αντίστοιχα. Συνεπώς, μια αυτοματοποιημένη φυλογενετική ανάλυση μπορεί να ανιχνεύσει πρόσφατες εστίες HIV και ως εκ τούτου μπορεί να υποστηρίξει την επακόλουθη απόκριση της δημόσιας υγείας.

Στην συνέχεια, μία άλλη προσέγγιση αυτή της Μοντελοποίησης χρησιμοποιήθηκε σε μια μελέτη που περιελάμβανε αλληλουχίες HIV από MSM στις Κάτω Χώρες. [Ratmann O. et al.,2016] Η Προσέγγιση Μοντελοποίησης βασίζεται στη μοντελοποίηση του αναμενόμενου αντίκτυπου των στρατηγικών πρόληψης του HIV χρησιμοποιώντας ανακατασκευασμένα δίκτυα μετάδοσης. Σε αυτή τη μελέτη, χρησιμοποιήθηκε η ιογενή φυλογενετική σχέση μεταξύ μερικών (partial) αλληλουχιών πολυμεράσης υποτύπου Β του HIV-1 για την ανακατασκευή παρελθοντικών πιθανών συμβάντων μετάδοσης στην Ολλανδία και χρησιμοποιήθηκαν κλινικά αρχεία για να τον προσδιορισμό της σταδιοποίησης (staging) των πιθανών συμβάντων μετάδοσης εντός της λοίμωξης και της συνεχούς περίθαλψης. Αυτό έδωσε τη δυνατότητα υπολογισμού της αναλογίας των μεταδόσεων μεταξύ των ανασυσταθέντων συμβάντων μετάδοσης που αντιστοιχούν σε 14 στάδια της λοίμωξης και του συνεχούς φροντίδας. Οι μεταδόσεις μπορούν να αποδοθούν στα στάδια πριν από τη διάγνωση, επειδή οι αλληλουχίες του ιού HIV, που συλλέγονται πάντοτε μετά τη διάγνωση, αποκλίνουν αρκετά γρήγορα ώστε να υποδηλώνουν προηγούμενα συμβάντα μετάδοσης (Lam TT, Hon CC, Tang JW, 2010) . Παρομοίως, οι μεταδόσεις θα μπορούσαν επίσης να αποδοθούν σε άνδρες χωρίς επαφή σε περίθαλψη για τουλάχιστον 18 μήνες. Τέλος, χρησιμοποιώντας αυτές τις εκτιμήσεις, υπολογίστηκε ο δυνητικός αντίκτυπος των διαθέσιμων προγραμμάτων πρόληψης, αλλά επί του παρόντος που δεν εφαρμόστηκαν στον ολλανδικό πληθυσμό MSM, αν αυτά είχαν χρησιμοποιηθεί τα τελευταία 3 χρόνια. Συγκεκριμένα, αξιολογήθηκε αν οι αναθεωρημένες κατευθυντήριες γραμμές του ΠΟΥ για την άμεση ART και PREP θα μπορούσαν να έχουν αλλάξει ουσιαστικά την πορεία της ολλανδικής επιδημίας του HIV μεταξύ MSM. Η κατανόηση σε ποιων παρεμβάσεων θα πρέπει να δοθεί προτεραιότητα στην ολλανδική επιδημία MSM είναι μια σημαντική μελέτη περίπτωσης.

Η μελέτη μοντελοποίησης εντοπίζει πηγές μετάδοσης για τους MSM που έχουν μολυνθεί για περισσότερο από 1 χρόνο (recipient MSM). Οι πηγές μετάδοσης, συμπεριλαμβανομένων όλων των ανδρών ασθενών που ζουν με τον ιό HIV στις Κάτω Χώρες, ήταν άνδρες των οποίων η λοίμωξη αλληλοεπικαλύφθηκε με το παράθυρο λοίμωξης ενός λήπτη MSM, άνδρες για τους οποίους ήταν διαθέσιμος ένας γονότυπος HIV και άνδρες των οποίων η ιική αλληλουχία συσταδοποιήθηκε φυλογενετικά με έναν αποδέκτη MSM . Η φυλογενετική πιθανότητα ότι ένας συγκεκριμένος άνδρας είχε μεταδώσει τον ιό HIV σε έναν αποδέκτη MSM στη συνέχεια βασίστηκε στη γενετική απόσταση των αλληλουχιών του. Κατά τη στιγμή της μετάδοσης, αξιολογήθηκε για κάθε πιθανό HIV πομπό εάν ήταν οξεία ή χρόνια μολυσμένος, αν είχε διαγνωστεί με HIV και εάν έλαβε αντιρετροϊκή θεραπεία.

Σε σενάρια αντίστοιχης προσομοίωσης, η μελέτη υπολόγισε ότι το 19% των παρελθουσών λοιμώξεων θα μπορούσε να αποφευχθεί με μια θεραπεία ως προσέγγιση πρόληψης (υποθέτοντας ότι τα άτομα είχαν ελεγχθεί ετησίως).

Επιπλέον, η μελέτη ανέφερε ετήσιους ελέγχους, ακολουθούμενους από άμεση θεραπεία από εκείνους με θετικό αποτέλεσμα και αν το 50% των MSM που έδειξαν αρνητικό θα είχαν χρησιμοποιήσει PrEP, τότε το 66% των νέων λοιμώξεων θα μπορούσε να αποφευχθεί.

Μία αναδρομική μελέτη που διεξήχθη στο Σαν Αντόνιο του Τέξας από το 2013 έως το 2015 εντόπισε ένα σύμπλεγμα 27 ατόμων, το οποίο επεκτάθηκε ραγδαία σε μεταγενέστερη παρακολούθηση. Η περαιτέρω διερεύνηση των υπηρεσιών συντρόφων (partner services) και των αρχείων συνεντεύξεων εντόπισαν 87 άτομα που είχαν μολυνθεί από τον ιό HIV που ήταν σεξουαλικοί σύντροφοι, σύντροφοι βελόνας (needle-sharing partners) ή επαφές κοινωνικού δικτύου επιβεβαιωμένων περιπτώσεων. Επομένως, αυτά τα 87 άτομα ήταν πολύ πιθανό να ανήκουν στο ίδιο σύμπλεγμα μετάδοσης με τα προαναφερθέντα 27 άτομα. Ωστόσο, αυτά τα άτομα δεν κατάφεραν να λάβουν κατάλληλη ιατρική περίθαλψη και τα δεδομένα προσδιορισμού αλληλουχίας HIV δεν ήταν διαθέσιμα για ανάλυση μοριακού δικτύου. Ως αποτέλεσμα, παρέμειναν σε υψηλό κίνδυνο μετάδοσης HIV [Oster AM, France AM, Mermin J,2018]. Ένα άλλο παράδειγμα είναι ένα ξέσπασμα HIV

μεταξύ των χρηστών ενέσιμων ναρκωτικών στην Ινδιάνα. Μέσω μοριακών δικτύων και αυτοαναφερόμενες επαφές με σεξ υψηλού κινδύνου, κοινής χρήσης βελονών, καθώς και με το σεξ και τις βελόνες, περισσότερα από 200 άτομα με HIV εντοπίστηκαν με στενούς κοινωνικούς δεσμούς με συστάδες μολυσμένων από τον HIV μελών, μεταξύ των οποίων συνέβη το ξέσπασμα [Campbell EM et al.,2017 ; Peters PJ et al.,2016]. Αυτή η μελέτη απέδειξε ότι η διερεύνηση ενεργά αναπτυσσόμενων μοριακών συστάδων παρέχει ευκαιρίες για την ιεράρχηση των ατόμων που σχετίζονται με αυτές τις συστάδες για σύνδεση με τη φροντίδα και παραπομπή PrEP [Monterosso A et al.,2017]. Μια μελέτη από το H.B έδειξε επίσης ότι οι προσεγγίσεις που βασίζονται στο δίκτυο μπορούν να καθοδηγήσουν στοχευμένες προσπάθειες πρόληψης για άτομα που ήταν επί του παρόντος αρνητικά στον HIV αλλά με πολύ υψηλούς κινδύνους μόλυνσης με οικονομικά αποδοτικό τρόπο. Οι προσομοιωμένες παρεμβάσεις έδειξαν ότι η εστίαση του PrEP σε νέους MSM μπορεί να αποτρέψει τέσσερις φορές περισσότερες μολύνσεις σε διάστημα 5 ετών από την τυχαία κατανομή [Volz EM et al.,2018]. Αν και το PrEP είναι ευρέως αποδεκτό στις ανεπτυγμένες χώρες, η χρήση του στην Κίνα είναι αμφιλεγόμενη. Οι 1,2 εκατομμύρια εκτιμώμενοι MSM στην Κίνα με την υψηλότερη συχνότητα εμφάνισης HIV [Shang H et al.,2012] είναι ο δυνητικός στοχευμένος πληθυσμός για το PrEP. Η ανάλυση μοριακού δικτύου βοηθά στην παροχή PrEP σε άτομα που βρίσκονται στο δίκτυο προτεραιότητας και μπορεί ουσιαστικά να βελτιώσει την επίδραση του PrEP σε αυτήν την περίπτωση.

Τα μοριακά δίκτυα αξιολογούν την αποτελεσματικότητα των παρεμβάσεων

Τα μοριακά δίκτυα μπορούν επίσης να χρησιμοποιηθούν για την αξιολόγηση των επιπτώσεων της παρέμβασης. Έχουν αναπτυχθεί διάφορες μέθοδοι για την αξιολόγηση του κατά πόσον η υπό έρευνα στρατηγική παρέμβασης μπορεί να διακόψει τη μετάδοση σε επίπεδο πληθυσμού. Μια πρόσφατη αναδρομική μελέτη για το δίκτυο μετάδοσης HIV στη Νέα Υόρκη έδειξε ότι η προηγούμενη δυναμική ανάπτυξης των συστάδων μπορεί να προβλέψει τη μελλοντική ανάπτυξη των συστάδων. Επομένως, τα σχήματα ιεράρχησης σε επίπεδο συμπλέγματος, λαμβάνοντας υπόψη τη σχέση μεταξύ της προηγούμενης ανάπτυξης και μεγέθους συμπλέγματος, μπορεί να συμβάλλουν στη βελτίωση των τελικών αποτελεσμάτων στη δημόσια υγεία [Wertheim JO et al.,2018]. Μια μελέτη από το Σαν Ντιέγκο αξιολόγησε τις επιδράσεις του ελέγχου του HIV σε ένα πρόγραμμα πρώιμης δοκιμής βασισμένο στο νουκλεϊκό οξύ χρησιμοποιώντας παρακολούθηση μοριακών συστάδων. Οι συγγραφείς διαπίστωσαν ότι με το πρόωρο πρόγραμμα ελέγχων, περίπου 100 λιγότερες μολύνσεις από τον HIV σημειώθηκαν σε σύγκριση με τον αναμενόμενο αριθμό στην κεντρική περιοχή του Σαν Ντιέγκο το 2012. Η γενετική ανάλυση επίσης έδειξε ότι οι αλυσίδες μετάδοσης του HIV είναι πιθανότερο να σταματήσουν σε περιοχές με πρώιμο έλεγχο [Mehta SR et al.,2016]. Ο αριθμός αναπαραγωγής (reproduction number) (R), μια παράμετρος που αντικατοπτρίζει πόσο αποτελεσματικά μεταδίδονται οι μολυσματικοί παράγοντες, χρησιμοποιείται συνήθως για τη μοντελοποίηση της δυναμικής της μόλυνσης. R> 1 αντιπροσωπεύει ότι οι μολυσματικοί παράγοντες μπορούν να συνεχίσουν να εξαπλώνονται. Χρησιμοποιούνται δύο κύριοι εκτιμητές: ο βασικός αναπαραγωγικός αριθμός (R_0) και ο πραγματικός (effective) αριθμός αναπαραγωγής (Re). Οι R_0 και Re είναι οι μέσοι αριθμοί δευτερογενών λοιμώξεων που προκαλούνται από ένα τυπικό μολυσμένο άτομο σε έναν εντελώς ευαίσθητο πληθυσμό και μόνο σε ένα μέρος του πληθυσμού που είναι ευαίσθητο, αντίστοιχα [Heesterbeek H et al.,2015]. Για χαμηλά διαδεδομένες επιδημίες όπως ο HIV, το Re ισούται με R_0 [Wertheim JO et al.,2014]. Το 2012, μια ομάδα από την Ελβετία ανέπτυξε μια νέα φυλογενετική μπευζιανή μέθοδο βασισμένη σε ένα μοντέλο γέννησης-θανάτου για να εκτιμήσει το R_0 απευθείας χρησιμοποιώντας τα δεδομένα της ιογενούς αλληλουχίας, στα οποία τα ποσοστά μετάδοσης και θανάτου εκτιμήθηκαν ανεξάρτητα για να βελτιώσουν ουσιαστικά την ακρίβεια σε σύγκριση με άλλα εκτιμήσεις συνένωσης (coalescent) [Stadler T et al.,2012]. Το 2017, η ίδια ομάδα εκτίμησε το R_0 των επιδημιών HIV μεταξύ ενός

ετεροφυλόφιλου πληθυσμού στην Ελβετία χρησιμοποιώντας την ανάλυση φυλογενετικών συστάδων με βάση τον πληθυσμό και διαπίστωσε ότι το R_0 του πληθυσμού ήταν πολύ κάτω από το όριο επιδημίας [Turk T et al.,2017]. Αυτή η μέθοδος μπορεί να είναι σε θέση να εκτιμήσει τις επιπτώσεις των τρεχόντων εφαρμοζόμενων προληπτικών μέτρων. Μια άλλη πρόσφατη μελέτη ενσωμάτωσε τη φυλοδυναμική σε μια μοριακών συστάδων ανάλυση για την παρακολούθηση της δυναμικής του συμπλέγματος με πυκνές δειγματοληπτικές αλληλουχίες από τη Βόρεια Καρολίνα. Το εκτιμώμενο Re των ενεργών συστάδων ήταν σημαντικά υψηλότερο από αυτό των ιστορικών ομάδων. Ο καθορισμός ενεργά αναπτυσσόμενων συστάδων είναι ζωτικής σημασίας για τη βελτιστοποίηση των προσεγγίσεων για την ανταπόκριση στη δημόσια υγεία, και μια αποτελεσματική παρέμβαση αναμένεται να μειώσει το Re [Dennis AM et al.,2019 ; France AM, Oster AM ,2019]

ΚΕΦΑΛΑΙΟ 4

Συμπεράσματα και μελλοντικές κατευθύνσεις ερευνάς

Πλεονεκτήματα

Η γενικευμένη γονοτύπιση ρουτίνας HIV σε αναπτυγμένες περιοχές, όπου η συγκριτικά χαμηλή, αλλά εξαιρετικά ετερογενής επικράτηση του ιού HIV θέτει σημαντικές προκλήσεις για την αποτελεσματική από πλευράς κόστους ανάπτυξη των πόρων πρόληψης του HIV, αποτελεί ισχυρό κίνητρο για παρακολούθηση κοντά σε πραγματικό χρόνο μέσω της δευτερογενούς φυλογενετικής ανάλυσης δεδομένων τα οποία συλλέγονται ως πρότυπο φροντίδας. Οι Poon et al, 2016 παρουσίασαν μια περίπτωση όπου το σύστημά παρακολούθησης έδωσε προτεραιότητα σε μια συγκεκριμένη εστία μεταδιδόμενης αντίστασης σε φάρμακα (την συστάδα 55) για την αντιμετώπιση της δημόσιας υγείας. Η ενισχυμένη παρακολούθηση της δημόσιας υγείας στην εστία του συμπλέγματος 55 ακολουθήθηκε από την καταστολή των ιικών φορτίων στα περισσότερα άτομα που είχαν προσβληθεί και αρκετούς μήνες χωρίς νέες περιπτώσεις. Είναι δύσκολο να αποδειχθεί μια αιτιώδης επίδραση στην παρέμβαση αυτή, δεδομένου ότι δεν υπήρχε ομάδα ελέγχου, η οποία να υφίσταται παρόμοια έκρηξη όπου παρακρατήθηκε η δράση δημόσιας υγείας. Ωστόσο, η πλειοψηφία (75%) των 12 νέων περιπτώσεων HIV που εμφανίστηκαν στο σύμπλεγμα 55 από τον Ιανουάριο του 2015 ήταν μη ανθεκτικά στελέχη που αντιστοιχούσαν σε τμήματα της φυλογένειας του HIV που δεν είχε στοχεύσει η ενισχυμένη παρακολούθηση της δημόσιας υγείας. Αυτό το αποτέλεσμα υποδηλώνει ότι η παρακολούθηση εκπλήρωσε εν μέρει τον πρωταρχικό της στόχο να αποτρέψει την περαιτέρω μετάδοση της ανθεκτικότητας σε φάρμακο HIV, αν και δεν ήταν απολύτως επιτυχής στην πρόληψη περαιτέρω μεταδόσεων στο σύμπλεγμα. Τέτοιες ενέργειες θα καταστούν τελικά σημαντικές για την υποστήριξη των στοχοθετημένων προσπαθειών πρόληψης του HIV και τη διατήρηση των θεραπευτικών επιλογών για τον πληθυσμό και μπορεί να μεταφραστούν σε άλλους τομείς μολυσματικών ασθενειών.

Περιορισμοί-Μειονεκτήματα

Ένας από τους γενικούς περιορισμούς στην παρακολούθηση των σημείων εστίασης μετάδοσης του ιού HIV από τη γονοτυπία ρουτίνας είναι ότι περιορίζεται σε πληθυσμούς με μέλη που παρουσιάζονται για εξέταση HIV και έχουν δοκιμασία ιικού φορτίου πλάσματος. Ο μη διαγνωσμένος πληθυσμός μπορεί να συμβάλει δυσανάλογα στην περαιτέρω μετάδοση του HIV. Ως εκ τούτου, ο αντίκτυπος της παρακολούθησης σε πραγματικό χρόνο μπορεί να εξαρτάται από τη βελτιστοποίηση όλων των σταδίων της φροντίδας του HIV. Επιπλέον, ο χρόνος μεταξύ της λοίμωξης από HIV και της διάγνωσης ποικίλει αναπόφευκτα μεταξύ των ατόμων.

Η χρήση συλλεχθέντων από το κοινό κλινικών δεδομένων σχετικά με το HIV για την ενημέρωση των παρεμβάσεων στον τομέα της δημόσιας υγείας εγείρει σημαντικές ηθικές παραμέτρους. Η κατευθυντήρια αρχή του πλαισίου δεοντολογίας της δημόσιας υγείας είναι να χρησιμοποιηθούν οι λιγότερο παρεμβατικές αλλά αποτελεσματικότερες παρεμβάσεις. Καθώς η χρήση της φυλογενετικής είναι μια αναδυόμενη περιοχή στη δημόσια υγεία, δεν υπάρχουν κατευθυντήριες γραμμές βέλτιστης πρακτικής για τη χρήση αυτών των πληροφοριών. Ο πρωταρχικός στόχος της διαχείρισης της δημόσιας υγείας είναι να αποτραπεί η περαιτέρω μετάδοση του ιού HIV με την προσέγγιση των πληθυσμών που βρίσκονται σε κίνδυνο - να μην αποδίδεται η μετάδοση σε συγκεκριμένα άτομα. Η ταυτοποίηση των ομάδων για παροχή συμβουλών, δοκιμών και θεραπείας σε ταχέως

αναπτυσσόμενα φυλογενετικά συμπλέγματα είναι συνεπής με τους στόχους της πρόληψης του HIV και αποθαρρύνει την απόδοση σφάλματος σε οποιοδήποτε άτομο

Μια άλλη πρόκληση για την εφαρμογή της φυλογενετικής παρακολούθησης του HIV είναι η εξεύρεση της ισορροπίας ανάμεσα στην προστασία του δικαιώματος του ατόμου στην ιδιωτική ζωή και του δικαιώματος άρνησης της ιατρικής περίθαλψης και στην ευθύνη της δημόσιας υγείας για την πρόληψη της περαιτέρω μετάδοσης του HIV.

Συμπεράσματα

Το μέλλον της φυλογενετικής επιστήμης φαίνεται πολύ ελπιδοφόρο. Η εντυπωσιακή πρόοδος της γενωμικής σημαίνει ότι θα γίνουν προσιτοί τεράστιοι όγκοι δεδομένων αλληλουχιών, και είναι πιθανό ότι νέοι τύποι πληροφορίας αλληλουχιών θα χρησιμοποιηθούν για φυλογένεση. Παρά τους αναπόφευκτους περιορισμούς, αποδείξαμε ότι η φυλογενετική παρακολούθηση μπορεί να συμπληρώσει τις τυποποιημένες επιδημιολογικές μεθόδους και να ενημερώσει τις δράσεις δημόσιας υγείας σε χρονική κλίμακα που επαρκεί για να μεταβάλει ενδεχομένως την πορεία μιας τοπικής εστίας.

Η φυλογενετική είναι ένας αυξανόμενος και συναρπαστικός τομέας στις επιστήμες πρόληψης του HIV. Όπως όλες οι επιστήμες, η φυλογενετική έχει περιορισμούς και όταν χρησιμοποιείται για επιδημιολογικούς σκοπούς, υπόκειται σε πολλές από τις ίδιες μεροληπτικές μετρήσεις και δειγματοληψίες των παραδοσιακών σχεδίων επιδημιολογικής μελέτης. Από τη σκοπιά της δημόσιας υγείας, η φυλογενετική του HIV είναι ισχυρότερη όταν συνδυάζεται με λεπτομερή κλινικά και επιδημιολογικά δεδομένα, οπότε οι HIV φυλογένειες του μπορούν να αποκαλύψουν κρίσιμες πληροφορίες σχετικές με τον έλεγχο της νόσου, συμπεριλαμβανομένης της μετάδοσης του ανθεκτικού σε φάρμακα ιού, των συσχετισμών μεταξύ των κοινωνικοδημογραφικών χαρακτηριστικών και της διάδοσης των ιών στους πληθυσμούς και τις χρονικές κλίμακες κατά τις οποίες εμφανίζονται οι επιδημίες του ιού HIV. Επιπρόσθετες θεωρητικές μελέτες που αφορούν τις φυλογένειες του ιού HIV σε δομή δικτύου και διαδικασίες μετάδοσης είναι απαραίτητες.

Όταν αξιολογήσαμε μελέτες που έχουν αναλύσει σύνολα δεδομένων αλληλουχιών που καλύπτουν σχετικά μεγάλο ποσοστό του μολυσμένου πληθυσμού σε εθνικές ή περιφερειακές κλίμακες, κατέστη σαφές ότι δεν υπάρχει κοινή στρατηγική για τον ορισμό συστάδων μετάδοσης. Ο αυξανόμενος αριθμός διαθέσιμων αλληλουχιών του HIV θα καταστήσει όλο και πιο δύσκολη την εξαγωγή φυλογενετικών δέντρων για τον προσδιορισμό συστάδων μετάδοσης. Μία μελλοντική πρόκληση θα είναι επομένως η εκτίμηση του επιπέδου κάλυψης αλληλουχίας (δηλ. του κλάσματος fraction του συνολικού αριθμού μολυσμένων ατόμων σε έναν πληθυσμό) στην οποία οι τρέχουσες μέθοδοι καθορισμού της υποστήριξης των κλάδων καθίστανται μη πρακτικές ή και μη ενημερωτικές.

Οι πρόσφατες εξελίξεις στις στρατηγικές αλληλουχίας δεν έχουν ως αποτέλεσμα μόνο έναν αυξημένο αριθμό ακολουθιών, αλλά και μια ευρεία ποικιλία στην ποιότητα και την ακρίβεια των ιικών ακολουθιών που υποβάλλονται σε δημόσιες βάσεις δεδομένων. Η Επόμενης γενιάς ακολουθία (NGS) είναι ανώτερη από την αλληλούχιση του Sanger στην ανίχνευση παραλλαγών χαμηλού επιπέδου, αλλά μερικές μεθοδολογίες NGS υποφέρουν από σχετικά υψηλότερα ποσοστά σφάλματος και μία από τις μεγαλύτερες προκλήσεις ήταν η διάκριση των τεχνικών και αναλυτικών σφαλμάτων από την πραγματική ιική ποικιλομορφία . Οι Eshleman et al. ανέλυσε τις αλληλουχίες HIV-1 από οκτώ ζεύγη index partner με μη συνδεδεμένες αλληλουχίες HIV-1 (όπως προσδιορίστηκε προηγουμένως με ανάλυση bulk Sanger αλληλουχιών) και ανέφερε ότι ένα από τα οκτώ ζεύγη στην πραγματικότητα συνδέθηκε όταν οι ιικοί πληθυσμοί αναλύθηκαν εκ νέου από το NGS. Αυτό δείχνει ότι αν και η αντιστοιχία μεταξύ των αλληλουχιών Sanger και NGS γενικά φαίνεται υψηλή, μπορεί να υπάρχουν περιπτώσεις στις οποίες οι αλληλουχίες bulk Sanger δεν θα αντιπροσωπεύουν επαρκώς ολόκληρο τον πληθυσμό του ιού μέσα σε ένα άτομο. Στην ανασκόπηση της βιβλιογραφίας μας, καμία από τις μελέτες δεν χρησιμοποίησε ακολουθίες

NGS για να μελετήσει τη δυναμική της μετάδοσης του HIV-1 σε μια γεωγραφική περιοχή ή χώρα. Ωστόσο, με τον αυξανόμενο αριθμό αλληλουχιών NGS που δημιουργούνται τα τελευταία χρόνια, θα πρέπει να μελετηθεί τόσο η επίδραση της ανάλυσης των Sanger έναντι αλληλουχιών NGS σε μεγαλύτερη κλίμακα βασισμένη στον πληθυσμό όσο και τα αποτελέσματα συνδυασμού και των δύο τύπων αλληλουχιών στην ίδια ανάλυση των συστάδων μετάδοσης.

Η εξελικτική δυναμική του HIV-1 και οι γενετικές δυνάμεις του πληθυσμού διαφέρουν ουσιαστικά μεταξύ των επιπέδων εντός του ξενιστή intra-host και των επιπέδων μεταξύ των ξενιστών inter-host και με τη διαδρομή μετάδοσης. Συνεπώς ένα άλλο θέμα που χρειάζεται περαιτέρω διερεύνηση είναι το πώς η συμπερίληψη αρκετών αλληλουχιών ανά ασθενή (είτε διαχρονικώς συλλεγμένες είτε πολλαπλές κλωνικές αλληλουχίες από ένα χρονικό σημείο) επηρεάζει την αναγνώριση των συστάδων μετάδοσης σε μεγάλα σύνολα δεδομένων. Παρομοίως, χρειάζονται περαιτέρω μελέτες σχετικά με τις επιδράσεις της μίξης αλληλουχιών mixing sequences από άτομα που έχουν μολυνθεί μέσω διαφορετικών οδών μετάδοσης. Ένα ξέσπασμα outbreak μεταξύ των IDUs μπορεί για παράδειγμα να φαίνεται πολύ διαφορετικό σε σύγκριση με ένα συγκρότημα μετάδοσης με αλληλουχίες κυρίως από MSM ή ετεροφυλόφιλων. Έχει προταθεί ότι τέτοιες διαφορές μπορεί να συνδέονται με ταχείες μεταδόσεις HIV-1 και έλλειψη σημείων συμφόρησης μεταδόσεων στα σημεία έκρηξης της IDU.

Οι ασθενείς με HIV-1 είναι συνδεδεμένοι με ιστορικό μετάδοσης και οι HIV-1 πληθυσμοί συσσωρεύουν γενετική απόσταση με την πάροδο του χρόνου. Ως εκ τούτου, οι γενετικές αποστάσεις σε ένα συγκρότημα μετάδοσης θα εξαρτηθούν από το πόσο καιρό πριν δημιουργήθηκε. Ο πλέον κατάλληλος ορισμός μίας συστάδας μετάδοσης του HIV-1 θα εξαρτηθεί από την υπόθεση που εξετάζεται και τη σύνθεση του υπό μελέτη συνόλου δεδομένων αλληλουχίας HIV-1. Συνεπώς, καμία μέθοδος ή όριο αποκοπή cut-off δεν θα ταιριάζει σε όλους τους ερευνητικούς σκοπούς. Ωστόσο, μια προσέγγιση που συνδυάζει ένα κατώφλι γενετικής απόστασης με μια υποστήριξη φυλογενετικού κλάδου φαίνεται να ταιριάζει με τις περισσότερες υποθέσεις και σύνολα δεδομένων. Επιπλέον, το χαλαρά καθορισμένο γενετικό όριο (π.χ. μεγαλύτερο από τα συνήθως χρησιμοποιούμενα κατώφλια 1,5 ή 4,5%) επιτρέπει την συμπερίληψη ομάδων που καλύπτουν μεγαλύτερες χρονικές περιόδους. Αυτό φαίνεται κατάλληλο για σύνολα δεδομένων με κάλυψη υψηλών ακολουθιών high-sequence coverage των πληθυσμών που ακολουθούνται σε μακροχρόνιες περιόδους αν ο κύριος στόχος είναι να κατανοήσουμε τη μακροχρόνια δυναμική μετάδοσης. Εντούτοις, ένα υψηλότερο όριο θα αυξήσει την πιθανότητα να συμπεριληφθούν συστάδες μετάδοσης με ελλείπουσες συνδέσεις (δηλαδή ακολουθίες χωρίς δειγματοληψία) και ένα αυστηρότερο γενετικό όριο ή ανάλυση μοριακού ρολογιού μπορεί να είναι καταλληλότερο όταν ο στόχος είναι να προσδιοριστούν πρόσφατα και επιδημιολογικά ενεργά σύνολα μετάδοσης (οι πρόσφατα σχηματισμένες ομάδες που έχουν μεγαλύτερη πιθανότητα να είναι ακόμα ενεργοί).

Παρότι όμως τόσες πολλές μελέτες έχουν γίνει σχετικά με την σύγκριση διαφόρων στατιστικών μεθόδων, έχει δειχθεί ότι η ακρίβεια του αναδομούμενου φυλογενετικού δένδρου εξαρτάται περισσότερο από τον όγκο της γενετικής πληροφορίας που θα αναλυθεί παρά από την φυλογενετική μέθοδο (Leitner et al., 1996; Kaye et al., 2009). Ως εκ τούτου, η έρευνα γύρω από τις περιοχές του προς ανάλυση γενετικού υλικού απαιτεί την κύρια προσοχή μας.

Όμως, ως προς το λιθαράκι το οποίο βάζουν με την σειρά τους οι φυλογενετικές μέθοδοι έχει γίνει πια σαφές ότι το επίπεδο ακρίβειας αυτών δικαιολογεί την χρήση τους στην έρευνα γύρω από τον HIV-1 και διαφωνεί με την θεωρία της συγκλίνουσας εξέλιξης και της επιλεκτικής μετάδοσης συγκεκριμένων στελεχών του ιού - θεωρίες οι οποίες είχαν αρχικώς παρουσιασθεί ως βασικοί ανασταλτικοί παράγοντες ως προς την επάρκεια των

παρουσών επιστημονικών τεχνικών να είναι ακριβείς υπό την παρουσία αυτών, θεωρίες που όμως τελικώς αποδείχθηκαν ανεπαρκείς (Leitner et al., 1996).

Το 2009 οι Kaye et al., με στόχο να ενισχύσουν την φυλογενετική συμπερασματολογία προτείνουν η ερμηνεία των ιολογικών συνδέσεων να βασίζεται όχι μόνο στην φυλογενετική, αλλά και στην ύπαρξη μικρών γενετικών αποστάσεων, οι οποίες θα πρέπει να εκτιμώνται με περισσότερη ευαισθησία κατά την κατακλίδα των αποφάσεών μας.

θα πρέπει να γίνει μία γενικότερη εκτίμηση για τα εύρη των γενετικών αποστάσεων. Δηλαδή, να γίνει μία προτυποποίηση του εύρους μεταξύ ασυσχέτιστων στελεχών, μερικώς συσχετιζόμενων και άμεσα συσχετιζόμενων. Ως προς αυτό, βέβαια, την μεγαλύτερη σημασία έχει ο υπολογισμός του πόσο αποκλίνουν λόγω εξέλιξης τα στελέχη μεταξύ τους ανά κάποιο χρονικό διάστημα που έχει παρέλθει (π.χ. ανά έτος) αλλά και ανά γενετικό τόπο. Για την εκτίμηση όλων των ανωτέρω θα πρέπει να ληφθεί υπόψιν εκτεταμένος αριθμός συνιστωσών οι οποίες μπορεί να επηρεάσουν, η καθεμία με την δική της βαρύτητα, αυτήν την γενετική απόκλιση μεταξύ των στελεχών του HIV από άτομο σε άτομο (π.χ. ηλικία, ανοσολογικές πιέσεις, φαρμακευτική αγωγή κ.τ.λ.). Μέχρι την μελέτη και θέσπιση αυτών των ορίων, η συνεκτίμηση των γενετικών αποστάσεων θα πρέπει να γίνεται με ιδιαίτερη προσοχή.

Σε περιπτώσεις όπου τα αποτελέσματα δεν είναι απολύτως ξεκάθαρα συνιστάται να συγκρίνονται τα αποτελέσματα μεταξύ ποικίλων φυλογενετικών μεθόδων και σε περίπτωση διαφωνίας μεταξύ τους να εκλαμβάνονται ως ανεπαρκή τα στοιχεία για να υποστηρίξουν μία ιική σύνδεση.

Συνοψίζοντας στην χρήση της στατιστικής συμπερασματολογίας, η μέθοδος της μέγιστης πιθανοφάνειας ακολουθούμενη από ανάλυση bootstrap αποτελεί μία από τις πιο συχνά χρησιμοποιημένες μεθόδους. Βεβαίως, μια γενικότερη εκτίμηση είναι ότι οι Μπεϋζιανές μέθοδοι ακολουθούμενες από εκτίμηση των PP, αν όχι ακόμα, σύντομα θα φτάσουν ή και θα ξεπεράσουν σε αποτελεσματικότητα και ακρίβεια την μέθοδο ML. Περαιτέρω μελέτες όμως απαιτούνται ακόμα

ΠΕΡΙΛΗΨΗ

Ο ιός της ανθρώπινης ανοσοανεπάρκειας (HIV) είναι ο αιτιολογικός παράγοντας του συνδρόμου επίκτητης ανοσοανεπάρκειας (AIDS). Ο ιός χαρακτηρίζεται από εκτενή γενετική ετερογένεια ως συνέπεια του υψηλού ρυθμού αλλαγών που ενσωματώνονται στο γενετικό του υλικό. Ο ρυθμός αυτός εκτιμάται ότι είναι 1 εκατομμύριο φορές πιο γρήγορος από τον αντίστοιχο στο ανθρώπινο γενετικό υλικό, αφήνοντας έτσι γενετικά αποτυπώματα σε σχετικά σύντομα χρονικά διαστήματα. Ταξινομείται παγκοσμίως σε τέσσερις ομάδες, από τις οποίες η Μ είναι υπεύθυνη για την πανδημία του AIDS και ταξινομείται φυλογενετικά σε εννέα υπότυπους, υπό-υπότυπους και σε ανασυνδυασμένους τύπους.

Η επιδημία HIV/AIDS παραμένει ακόμα και σήμερα ένα σημαντικό πρόβλημα της δημόσιας υγείας, τόσο σε παγκόσμιο όσο και εθνικό επίπεδο. Ο κίνδυνος μόλυνσης από τον HIV διαφέρει ανάμεσα σε άτομα που ανήκουν σε διαφορετικές κατηγορίες μετάδοσης και επομένως, η γνώση του τρόπου με τον οποίο μεταδόθηκε ο ιός έχει ιδιαίτερη σημασία, καθώς συμβάλει, αφενός, στη βαθύτερη κατανόηση των επιδημιολογικών χαρακτηριστικών της λοίμωξης και αφετέρου, στη λήψη αποτελεσματικότερων μέτρων πρόληψης για τον περιορισμό νέων λοιμώξεων

Λόγω αυτών των χαρακτηριστικών του ιού, μπορεί με μεθόδους φυλογενετικής ανάλυσης, να διερευνηθεί η πιθανή επιδημιολογική σχέση ανθρώπων με HIV λοίμωξη. Καθοδηγούμενοι από την πληθυσμιακή γενετική και τις επιδημιολογικές αρχές, οι επιστήμονες χρησιμοποιούν ιογενή φυλογενετική για τη βελτίωση της κατανόησης της διαφορετικότητας του HIV μέσα σε άτομα και πληθυσμούς, δημιουργώντας μια άνευ προηγουμένου γνώση της ιογενούς δυναμικής για τη βελτίωση των στρατηγικών πρόληψης του HIV και τη θεραπεία των HIV-μολυσμένων ατόμων Η μοριακή επιδημιολογική αξιολόγηση των δικτύων μετάδοσης του HIV μπορεί να διασαφηνίσει τα στοιχεία συμπεριφοράς της μετάδοσης που μπορούν να αποτελέσουν στόχους για την παρέμβαση. . Τα εργαλεία μοριακής επιδημιολογίας του HIV, όπως εφαρμόζονται σε φυλογενετικές, φυλοδυναμικές και φυλογεωγραφικές αναλύσεις, έχουν αποδειχθεί ισχυρά εργαλεία στον σχεδιασμό της δημόσιας υγείας σε πολλές μελέτες.

Σκοπός της παρούσας εργασίας μου είναι η ανάλυση των φυλογενετικών μεθόδων ως εργαλείο για τη δημιουργία στρατηγικών παρέμβασης και πρόληψης του HIV.

ABSTRACT

The Human Immunodeficiency Virus (HIV) is the etiologic factor of the Acquired Immunodeficiency Syndrome (AIDS). The virus is characterized from high genetic diversity as a consequence of the high rate of genetic changes that are integrated in his genetic material. This rate is estimated to be 1 million times faster than that of humans, letting this way, genetic fingerprints in relatively short time. HIV is globally classified into four groups, of which group M is responsible for AIDS pandemic and is phylogenetically classified into nine subtypes, sub-subtypes and recombinant forms.

HIV/AIDS epidemic still remains a hot issue in the field of both global and local public health. The risk of transmission differs among persons of different transmission groups and thus, the knowledge of transmission route of each patient leads to deeper understanding of viral epidemiology and better preventing measures

Because of these characteristics of the virus, we can, by the use of phylogenetic analysis methods, investigate the potential epidemiological relation of humans with HIV-1 infection.

Due to these characteristics of the virus, it is possible to investigate the possible epidemiological relationship between people with HIV infection by methods of phylogenetic analysis. Guided by population genetics and epidemiological principles, scientists are using viral phylogenetics to improve understanding of HIV diversity within individuals and populations, creating an unprecedented knowledge of viral dynamics to improve HIV prevention strategies and treat HIV-infected individuals. Molecular epidemiological evaluation of HIV transmission networks can clarify the behavioural elements of transmission that can be targeted for intervention. HIV molecular epidemiology tools, as applied to phylogenetic, phylodynamic and phylogeographical analyses, have been shown to be powerful tools in public health planning in many studies.

The purpose of my current work is to analyze phylogenetic methods as a tool for creating HIV intervention and prevention strategies.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1) Abecasis AB, Vandamme AM, Lemey P. 2009. Quantifying differences in the tempoof human immunodeficiency virus type 1 subtype evolution. J Virol 83: 12917-12924
2) Acheson, N.H., Fundamentals of molecular virology. 2nd ed. 2011, Hoboken, NJ: John Wiley & Sons. xxv, 500 p.
3) Aldous JL, Pond SK, Poon A, Jain S, Qin H, Kahn JS, Kitahata M, Rodriguez B, Dennis AM, Boswell SL, Haubrich R, Smith DM. Characterizing HIV transmission networks across the United States. Clin Infect Dis 2012; 55(8): 1135–1143
4) Avice, J. (2006). Evolutionary Pathways in Nature. A Phylogenetic Approach. New York: Cambridge University Press.
5) Avila D, Keiser O, Egger M, Kouyos R, Bφni J, Yerly S, Klimkait T, Vernazza PL, Aubert V, Rauch A, Bonhoeffer S, Gónthard HF, Stadler T, Spycher BD; Swiss HIV Cohort Study. Social meets molecular: combining phylogenetic and latent class analyses to understand HIV-1 transmission in Switzerland. Am J Epidemiol 2014; 179(12): 1514–1525
6) Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. Improving the accuracy of demographic andmolecular clock model comparison while accommodating phylogenetic uncertainty. Mol Biol Evol 2012; 29(9): 2157 2167
7) Baele G, Li WL, Drummond AJ, Suchard MA, Lemey P. Accurate model selection of relaxed molecular clocks in bayesian phylogenetics. Mol Biol Evol 2013; 30(2): 239–243
8) Barre-Sinoussi F, Chermann J, Rey F, Nugeyre M, Chamaret S, Gruest J, Dauguet C, Axler-Blin C, Vezinet-Brun F, Rouzioux C, Rosenbaum W, Montagnier L. 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science. 220: 868–871
9) Beerenwinkel N, Zagordi O. Ultra-deep sequencing for the analysis of viral populations. Curr Opin Virol. 2011; 1:413–8. [PubMed: 22440844]
10) Bennett, J.E., R. Dolin, and M.J. Blaser, Mandell, Douglas, and Bennett's principles and practice of infectious diseases. Eighth edition. ed. 2015, Philadelphia, PA: Elsevier/Saunders. 2 volumes.
11) Bouckaert R, Vaughan TG, Barido-Sottani J, Duchкne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kóhnert D, De Maio N, Matschiner M, Mendes FK, Móller NF, Ogilvie HA, du Plessis L, Popinga A, Rambaut A, Rasmussen D, Siveroni I, Suchard MA, Wu CH, Xie D, Zhang C, Stadler T, Drummond AJ. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol 2019; 15(4): e1006650
12) Bremer K. Branch support and tree stability, Cladistics, 10, 295-304 (1994).
13) Brocchieri L. Phylogenetic Inferences from Molecular Sequences: Review and Critique, Theoretical Population Biology, 59, 27-40 (2001).
14) Boeyer JC, Babenek K, Kunkel TA. 1992. Unequal human immunodeficiency virus type 1 reverse transcriptase error rates with RNA and DNA templates. Proc Natl Acad Sci USA 89: 6919-6923
15) Brown, T. A. (2002) Genomes (2nd ed.). Oxford.
16) Campbell EM, Jia H, Shankar A, Hanson D, Luo W, Masciotra S, Owen SM, Oster AM, Galang RR, Spiller MW, Blosser SJ, Chapman E, Roseberry JC, Gentry J, Pontones P, Duwve J, Peyrani P, Kagan RM, Whitcomb JM, Peters PJ, Heneine W, Brooks JT, Switzer WM. Detailed transmission network analysis of a large opiate-driven outbreak of HIV infection in the United States. J Infect Dis 2017; 216(9): 1053–1062
17) Carr JK, Salminen MO, Albert J, Sanders-Buell E, Cotte D, Birx DL, McCutchan FE. 1998. Full genome sequences of human immunodeficiency virus type 1 subtypes G and A/G intersubtype recombinants. Virology 247: 22-31
18) Chaillon A, Essat A, Frange P, et al. Spatiotemporal dynamics of HIV-1 transmission in France (1999–2014) and impact of targeted prevention strategies. Retrovirology 2017; 14:15

19) Centers for Disease Control and Prevention. HIV by Group. [cited 2018; Available from: https://www.cdc.gov
20) CDC, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention Division of HIV/AIDS Prevention, 2018
21) Centers for Disease Control (CDC). 1982. Update on acquired immune deficiency syndrome (AIDS)—United States. MMWR Morb Mortal Wkly Rep. 31: 507–508; 513–514
22) Clavel F, Guyader M, Guetard D, Salle M, Montagnier L, Alizon M. 1986. Molecular cloning and polymorphism of the human immune deficiency virus type 2. Nature 324: 691-695
23) Chaillon A, Essat A, Frange P, Smith DM, Delaugerre C, Barin F, Ghosn J, Pialoux G, Robineau O, Rouzioux C, Goujard C, Meyer L, Chaix ML; on behalf the ANRS PRIMO Cohort Study. Spatiotemporal dynamics of HIV-1 transmission in France (1999-2014) and impact of targeted prevention strategies. Retrovirology. 2017 Feb 21;14(1):15. doi: 10.1186/s12977-017-0339-4. PubMed PMID: 28222757; PubMed Central PMCID: PMC5322782
24) Chan PA, Hogan JW, Huang A, DeLong A, Salemi M, Mayer KH, Kantor R. Phylogenetic Investigation of a Statewide HIV-1 Epidemic Reveals Ongoing and Active Transmission Networks Among Men Who Have Sex With Men. J Acquir Immune Defic Syndr. 2015 Dec 1;70(4):428-35. doi: 10.1097/QAI.0000000000000786. PubMed PMID: 26258569; PubMed Central PMCID: PMC4624575
25) Charneau P, Borman AM, Quillent C, Guetard D, Chamaret S, Cohen J, Remy G, Montagnier L, Clavel E. 1994. Isolation and envelope sequence of a highly divergent HIV-1 isolate: Definition of a new HIV-1 group. Virology 205: 247-253
26) Chin BS, Chaillon A, Mehta SR, Wertheim JO, Kim G, Shin HS, Smith DM. Molecular epidemiology identifies HIV transmission networks associated with younger age and heterosexual exposure among Korean individuals. J Med Virol 2016; 88(10): 1832–1835
27) Cohen MS, Hellmann N, Levy JA, DeCock K, Lange J. 2008. The spread, treatment and prevention of HIV-1: evolution of a global pandemic. J Clin Invest 118: 1244-1254
28) Collier, L.H. and J.S. Oxford, Human virology : a text for students of medicine, dentistry, and microbiology. 3rd ed. 2006, Oxford ; New York: Oxford University Press. xviii, 303 p.
29) Costin JM. 2007. Cytopathic mechanisms of HIV-1. Virol J 4: 100
30) DeBry R. W. The consistency of several phylogeny-inference methods under varying evolutionary rates, Mol. Biol. Evol., 9, 537-51 (1992).
31) Decosas J, Kane F, Anarfi JK, Sodji KD, Wagner HU. 1995. Migration and AIDS. Lancet 346: 826-828
32) Decosas J, Adrien A. 1997. Migration and HIV. AIDS 11 Supple A: S77-S84
33) Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J Roy Statist Soc Ser B 1977;39(1):1–38.
34) Dennis AM, Huı S, Billock R, Levintow S, Sebastian J, Miller WC, Eron JJ. Human immunodeficiency virus type 1 phylodynamics to detect and characterize active transmission clusters in North Carolina. J Infect Dis 2019 Apr 27. [Epub ahead of print] doi: 10.1093/infdis/jiz176
35) Dimmock, N.J., A.J. Easton, and K.N. Leppard, Introduction to modern virology. 2016, Wiley Blackwell: Chichester, West Sussex ; Hoboken, NJ,. p. xviii, 519 pages.
36) Dosekun, O. and J. Fox, An overview of the relative risks of different sexual behaviours on HIV transmission. Curr Opin HIV AIDS, 2010. 5(4): p. 291-7.
37) Douady J. C., Delsuc F., Boucher Y., Doolittle F. W. and Douzery J. P. E. Comparison of Bayesian and Maximum Likelihood Bootstrap Measures of Phylogenetic Reliability, Mol. Biol. Evol., 20, 248-54 (2003).
38) Efron B. The Jacknife, the Bootstrap and Other Resampling Plans, CBMS-NSF Regional Conference Series in Applied Mathematics, Philadelphia: Society Industrial Applied Mathematics (1982).

39) Efron B., Halloran E. and Holmes S. Bootstrap confidence levels for phylogenetic trees, Proc. Natl. Acad. Sci. USA, 93, 7085-90 (1996).
40) Erixon P., Svennblad B., Britton T. and Oxelman B. Reliability of Bayesian Posterior Probabilities and Bootstrap Frequencies in Phylogenetics, Syst. Biol., 52, 665-73 (2003).
41) Esparza J and Bhamarapravati N. 2000. Accelerating the development and future availability of HIV-1 vaccines: Why, when, where and how? Lancet 355: 2061-2066
42) Faith P. D. Cladistic permutation tests for monophyly and nonmonophyly, Syst. Zool., 40, 366-75 (1991).
43) Faith P. D. and Trueman H. W. J. When the topology-dependent permutation test (T-PTP) for monophyly returns significant support for monophyly, should that be equated with (a) rejecting a null hypothesis of nonmonophyly, (b) rejecting a null hypothesis of "nostructure", (c) failing to falsify a hypothesis of monophyly or (d) none of the above?, Syst. Biol., 45, 580-6 (1996).
44) Felsenstein J. Cases in which parsimony or compatibility methods will be positively misleading, Sys. Zool., 27, 401-10 (1978).
45) Felsenstein J. Evolutionary trees from DNA sequences: a maximum likelihood approach, J. Mol. Evol., 17, 368-76 (1981).
46) Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap, Evolution, 39, 783-91 (1985).
47) Felsenstein, J. (2004). Inferring Phylogenies. Sunderland, MA: Sinauer Associates.
48) Felsenstein J. Phylogenies from molecular sequences: inference and reliabily, Annu. Rev. Genet., 22, 521-65 (1988).
49) Foxman, B. and L. Riley, Molecular epidemiology: focus on infection. Am J Epidemiol, 2001. 153(12): p. 1135-41.
50) France AM, Oster AM. The promise and complexities of detecting and monitoring HIV transmission clusters. J Infect Dis 2019 Apr 27. [Epub ahead of print] doi: 10.1093/infdis/jiz177
51) France AM, Panneer N, Ocfemia CB, Saduvala N, Campbell E, Switzer WM, Wertheim J, Oster AM. Rapidly growing HIV transmission clusters in the Unites States, 2013–2016. 2018 Conference on Retroviruses and Opportunistic Infections. March 4–7, 2018
52) Gallo RC, Sarin PS, Gelmann EP, Robert-Guroff M, Richardson E, Kalyanaraman VS, Mann D, Sidhu GD, Stahl RE, Zolla-Pazner S, Leibowitch J, Popovic M. 1983. Isolation of human T-cell leukemia virus in acquired immune deficiency syndrome (AIDS). Science 220: 865–867
53) Gao, F., et al., Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature, 1999. 397(6718): p. 436-41.
54) German, D., Grabowski, M., & Beyrer, C. (2017). Enhanced use of phylogenetic data to inform public health approaches to HIV among men who have sex with men. Sexual Health, 14(1), 89-96. https://doi.org/10.1071/SH16056
55) Goloboff P. A. Estimating character weights during tree search, Cladistics, 9, 83-91 (1993).
56) Greenacre M. Theory and Application of Correspondence Analysis,. London: Academic Press; 1984.
57) Hamelaar J, Gouws E, Ghys PD, Osmanov S; WHO-UNAIDS Network for HIV Isolation and Characterisation. 2011. Global trends in molecular epidemiology of HIV-1 during 2000-2007. AIDS 25: 679-689
58) Hasegawa, M., Kishino, H., & Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol, 22(2), 160-174.
59) Hasegawa M. and Yano T. Maximum likelihood method of phylogenetic inference from DNA sequence data, Bull. Biometr. Sot. Jpn., 5, l-7 (1984).
60) Hassan AS, Pybus OG, Sanders EJ, Albert J, Esbjörnsson J. Defining HIV-1 transmission clusters based on sequence data. AIDS 2017; 31(9): 1211–1222

61) Heesterbeek H, Anderson RM, Andreasen V, Bansal S, De Angelis D, Dye C, Eames KT, Edmunds WJ, Frost SD, Funk S, Hollingsworth TD, House T, Isham V, Klepac P, Lessler J, Lloyd-Smith JO, Metcalf CJ, Mollison D, Pellis L, Pulliam JR, Roberts MG, Viboud C; Isaac Newton Institute IDD Collaboration. Modeling infectious disease dynamics in the complex landscape of global health. Science 2015; 347(6227): aaa4339
62) Hendy M. D. and Penny D. A framework for the quantitative study of evolutionary trees, Sys. Zool., 38, 297-309 (1989).
63) Hillis D. M. and Bull J. J. An empirical test of bootstrapping as a method for assessing the confidence in phylogenetic analysis, Syst. Biol., 42, 182-92 (1993).
64) Hu, W.S. and H.M. Temin, Retroviral recombination and reverse transcription. Science, 1990. 250(4985): p. 1227-33.
65) Huelsenbeck J. P. The Robustness of Two Phylogenetic Methods: Four-Taxon Simulations Reveal a Slight Superiority of Maximum Likelihood over Neighbor Joining, Mol. Biol. Evol., 12, 843-9 (1995).
66) Huelsenbeck J. P. and Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees, Bioinformatics, 17, 754-5 (2001a).
67) Huelsenbeck J. P., Ronquist F., Nielsen R. and Bollback P. J. Bayesian inference of phylogeny and its impact on evolutionary biology, Science, 294, 2310-4 (2001).
68) Huι S, Clewley JP, Cane PA, Pillay D. HIV-1 pol gene variation is sufficient for reconstruction of transmissions in the era of antiretroviral therapy. AIDS 2004; 18(5): 719–728
69) International Commitee on Taxonomy of Virus (ICTV). Virus Taxonomy: 2017 Release. 2017 [cited 2018 April]; Available from: https://talk.ictvonline.org/taxonomy/
70) Jukes, T. H., & Cantor, C. R. (1969). Evolution of protein molecules. In H. M. Munro (Ed.), Mammalian protein metabolism. New York: Academic Press.
71) Kanki PJ, Travers KU, Mboup S, Hsieh CC, Marklink RG, Gueye-Ndiaye A, Siby T, Thior I, Hernandez-Avila M, Sankale JL, Ndoye I, Essex ME. 1994. Slower heterosexual spread of HIV-2 than HIV-1. Lancelot 343: 943-946
72) Kaye M., Chibo D. and Birch C. Comparison of Bayesian and Maximum-Likelihood Phylogenetic Approaches in Two Legal Cases Involving Accusations of Transmission of HIV, Aids Research And Human Retroviruses, 25, 741-8 (2009).
73) Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol, 16(2), 111-120.
74) Kishino H. and Hasegawa M. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data and the branching order in Hominoidea, J. Mol. Evol., 29, 170-9 (1989).
75) Knipe, D.M. and P.M. Howley, Fields virology. 2013, Wolters Kluwer/Lippincott Williams & Wilkins Health: Philadelphia, PA. p. 2 volumes.
76) Korber B, Muldoon M, Theiler J, Gao F, Gupta R, Lapedes A, Hahn BH, Wolinsky S, Bhattacharya T. 2000. Timing the ancestor of HIV-1 pandemic strains. Science 288: 1789-1796
77) Kostaki EG, Nikolopoulos GK, Pavlitina E, Williams L, Magiorkinis G, Schneider J, Skaathun B, Morgan E, Psichogiou M, Daikos GL, Sypsa V, Smyrnov P, Korobchuk A, Malliori M, Hatzakis A, Friedman SR, Paraskevis D. Molecular analysis of human immunodeficiency virus type 1 (HIV-1)-infected individuals in a network-based intervention (Transmission Reduction Intervention Project): phylogenetics identify HIV-1-infected individuals with social links. J Infect Dis 2018; 218(5): 707–715
78) Kozal MJ. 2009. Drug-resistant human immunodeficiency virus. Clin Microbiol Infect 15 Suppl 1: 69-73
79) Kuhner K. M. and Felsenstein J. A Simulation Comparison of Phylogeny Algorithms under Equal and Unequal Evolutionary Rates, Mol. Biol. Evol., 11, 459-68 (1994).

80) Kumar S. A stepwise algorithm for finding minimum evolution trees, Mol. Biol. Evol., 13, 584-93 (1996).
81) Lam TT, Hon CC, Tang JW. Use of phylogenetics in the molecular epidemiology and evolutionary studies of viral infections. Crit Rev Clin Lab Sci. 2010; 47:5–49. [PubMed: 20367503]
82) Lazarsfeld PF, Henry NW. Latent Structure Analysis. Boston, MA: Houghton Mifflin; 1968.
83) Leigh Brown AJ, Lycett SJ, Weinert L, Hughes GJ, Fearnhill E, Dunn DT; UK HIV Drug Resistance Collaboration. Transmission network parameters estimated from HIV sequences for a nation wide epidemic. J Infect Dis 2011; 204(9): 1463–1469
84) Leitner T., Escanilla D., Franzen C., Uhlen M. and Albert J. Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis, Proc. Natl. Acad. Sci. USA, 93, 10864-9 (1996).
85) Leitner T., Kumar S. and Albert J. Tempo and Mode of Nucleotide Substitutions in gag and env Gene Fragments in Human Immunodeficiency Virus Type 1 Populations with a Known Transmission History, J. Virol., 71, 4761-70 (1997).
86) Letvin NL. 2006. Progress and obstacles in the development of an AIDS vaccine. Nat Rev Immunol 6: 930-939
87) Lever AM, Berkhout B. 2008. 2008 Nobel prize in medicine for discoverers of HIV. Retrovirology 5:91
88) Little SJ, Kosakovsky Pond SL, Anderson CM, Young JA, Wertheim JO, Mehta SR, May S, Smith DM. Using HIV networks to inform real time prevention interventions. PLoS One 2014; 9(6): e98443
89) Little RJA, Rubin DB. Statistical Analysis With Missing Data. New York, NY: John Wiley & Sons; 2002.
90) Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005; 437:376–80. [PubMed: 16056220]
91) McCutchan FE. 2000. Understanding the genetic diversity of HIV-1. AIDS 14 (Suppl 3): S31-S44
92) McLachlan G, Peel D. Finite Mixture Models. New York, NY: John Wiley & Sons; 2000.
93) Mehta SR, Murrell B, Anderson CM, Kosakovsky Pond SL, Wertheim JO, Young JA, Freitas L, Richman DD, Mathews WC, Scheffler K, Little SJ, Smith DM. Using HIV sequence and epidemiologic data to assess the effect of self-referral testing for acute HIV infection on incident diagnoses in San Diego, California. Clin Infect Dis 2016; 63(1): 101–107
94) Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010; 11:31–46. [PubMed: 19997069]
95) Modrow, S., Molecular virology. 2013, New York: Springer. pages cm.
96) Monterosso A, Minnerly S, Goings S, Morris A, France AM, Dasgupta S, Oster AM, Fanning M. Identifying and investigating a rapidly growing HIV transmission cluster in Texas. Conference on Retroviruses and Opportunistic Infections. March 8, 2017. Seattle,Washington
97) Muthén BO, Muthén LK. Mplus [computer software]. Los Angeles, CA: Muthén & Muthén; 1998.
98) National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention; Division of HIV/AIDS. Prevention Detecting and Responding to HIV Transmission Clusters: A Guide for Health Department. 2018. https://www.cdc.gov/hiv/pdf/funding/ announcements/ps18-1802/CDC-HIV-PS18-1802-AttachmentE Detecting-Investigating-and-Responding-to-HIV-Transmission- Clusters.pdf
99) Oster AM, France AM, Mermin J. Molecular epidemiology and the transformation of HIV prevention. JAMA 2018; 319(16): 1657–1658

100) Oster AM, Wertheim JO, Hernandez AL, Ocfemia MC, Saduvala N, Hall HI. Using molecular HIV surveillance data to understand transmission between subpopulations in the United States. J Acquir Immune Defic Syndr 2015; 70(4): 444–451
101) Page, R. D. M., & Holmes, E. C. (1998). Molecular Evolution: A Phylogenetic Approach. UK: Wiley-Blackwell.
102) Paraskevis, D., et al., Economic recession and emergence of an HIV-1 outbreak among drug injectors in Athens metropolitan area: a longitudinal study. PLoS One, 2013. 8(11): p. e78941.
103) Paraskevis D, Hatzakis A. 1999. Molecular epidemiology of HIV-1 infection. AIDS Rev 1: 238-249
104) Paraskevis D, Magiorkinis E, Magiorkinis G, Sypsa V, Paparizos V, Lazanas M, Gargalianos P, Antoniadou A, Panos G, Chrysos G, Sambatakou H, Karafoulidou A, Skoutelis A, Kordossis T, Koratzanis G, Theodoridou M, Daikos GL, Nikolopoulos G, Pybus OG, Hatzakis A. 2007. Increasing prevalence of HIV-1 subtype A in Greece: estimating epidemic history and origin. J Infect Dis 15: 1167-1176
105) Paraskevis D, Nikolopoulos GK, Magiorkinis G, Hodges-Mameletzis I, Hatzakis A. The application of HIV molecular epidemiology to public health. Infect Genet Evol. 2016;46:159-168
106) Pasquale DK, Doherty IA, Sampson LA, Huı S, Leone PA, Sebastian J, Ledford SL, Eron JJ, Miller WC, Dennis AM. Leveraging phylogenetics to understand HIV transmission and partner notification networks. J Acquir Immune Defic Syndr 2018; 78(4): 367–375
107) Pearson W. R., Robins G. and Zhang T. Generalized neighbor-joining: more reliable phylogenetic tree reconstruction, J. Mol. Evol., 16, 806-16 (1999).
108) Peters PJ, Pontones P, Hoover KW, Patel MR, Galang RR, Shields J, Blosser SJ, Spiller MW, Combs B, Switzer WM, Conrad C, Gentry J, Khudyakov Y, Waterhouse D, Owen SM, Chapman E, Roseberry JC, McCants V, Weidle PJ, Broz D, Samandari T, Mermin J, Walthall J, Brooks JT, Duwve JM; Indiana HIV Outbreak Investigation Team.HIV infection linked to injection use of oxymorphone in Indiana, 2014-–2015. N Engl J Med 2016; 375(3): 229–239
109) Plantier, J.C., et al., A new human immunodeficiency virus derived from gorillas. Nat Med, 2009. 15(8): p. 871-2.
110) Poon AFY, Gustafson R, Daly P, Zerr L, Demlow SE, Wong J, Woods CK, Hogg RS, Krajden M, Moore D, Kendall P, Montaner JSG, Harrigan PR. Near real-time monitoring of HIV transmission hotspots from routine HIV genotyping: an implementation case study. Lancet HIV 2016; 3(5): e231–e238
111) Prosperi MC, Ciccozzi M, Fanti I, Saladini F, Pecorari M, Borghi V, Di Giambenedetto S, Bruzzone B, Capetti A, Vivarelli A, Rusconi S, Re MC, Gismondo MR, Sighinolfi L, Gray RR, Salemi M, Zazzi M, De Luca A; ARCA collaborative group. A novel methodology for large-scale phylogeny partition. Nat Commun 2011; 2(1): 321
112) Ragonnet-Cronin M, Hodcroft E, Hué S, Fearnhill E, Delpech V, Brown AJ, Lycett S; UK HIV Drug Resistance Database. Automated analysis of phylogenetic clusters. BMC Bioinformatics 2013; 14(1): 317
113) Ramaswamy V, Desarbo WS, Reibstein DJ, et al. An empirical pooling approach for estimating marketing mix elasticities with PIMS data. Market Sci 1993;12(1):103–124
114) Rambaut A, Posada D, Crandall KA, Holmes EC. 2004. The causes and consequences of HIV evolution. Nat Rev Genet 5: 52-61
115) Ratmann O, van Sighem A, Bezemer D, et al. Sources of HIV infection among men having sex with men and implications for prevention. Sci Transl Med 2016; 8:320ra322
116) Redd AD, Mullis CE, Serwadda D, Kong X, Martens C, et al. The rates of HIV superinfection and primary HIV incidence in a general population in rakai, uganda. J Infect Dis. 2012; 206:267–274. [PubMed: 22675216]

117) Redd AD, Quinn TC, Tobian AA. Frequency and implications of HIV superinfection. Lancet Infect Dis. 2013; 13:622–628. [PubMed: 23726798]
118) Reeves, J.D. and R.W. Doms, Human immunodeficiency virus type 2. J Gen Virol, 2002. 83(Pt 6): p. 1253-65.
119) Robertson DL, Anderson JP, Bradac JA, Carr JK, Foley B, Funkhouser RK, Gao F, Hahn BH, Kalish ML, Kuiken C, Learn GH, Leitner T, McCutchan F, Osmanov S, Peeters M, Pieniazek D, Salminen M, Sharp PM, Wolinsky S, Korber B. 2000. HIV-1 nomenclature proposal. Science 288: 55-56
120) Rose R, Lamers SL, Dollar JJ, Grabowski MK, Hodcroft EB, Ragonnet-Cronin M, Wertheim JO, Redd AD, German D, Laeyendecker O. Identifying transmission clusters with cluster picker and HIV-TRACE. AIDS Res Hum Retroviruses 2017; 33(3): 211–218
121) Rowland-Jones SL. 2003. Timeline: AIDS pathogenesis: what have two decades of HIV research taught us? Nat Rev Immunol 3: 343-348
122) Ryu, W.-S., Molecular virology of human pathogenic viruses. 2017, Amsterdam: Academic Press. xv, 423 pages.
123) Saitou N. and Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol., 4, 406-425 (1987).
124) Salazar-Gonzalez JF, Bailes E, Pham KT, Salazar MG, Guffey MB, et al. Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. J Virol. 2008; 82:3952–3970. [PubMed: 18256145]
125) Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977; 74:5463–5467. [PubMed: 271968]
126) Schmidt, H. A., & von Haeseler, A. (2009). Phylogenetic inference using maximum likelihood methods:Theory. In P. Lemey, M. Salemi & A. M. Vandamme (Eds.), The Phylogenetic Handbook. A Practical Approach to Phylogenetic Analysis and Hypothesis Testing. New York: Cambridge University Press.
127) Sepkowitz, K.A., AIDS--the first 20 years. N Engl J Med, 2001. 344(23): p. 1764-72.
128) Shang H, Xu J, Han X, Spero Li J, Arledge KC, Zhang L. HIV prevention: bring safe sex to China. Nature 2012; 485(7400): 576– 577
129) Sharp PM, Hahn BH. 2011. Origins of HIV and the AIDS pandemic. Cold Spring Harp Perspect Med 1: a006841
130) Simon V, Ho DD, Karim QA. 2006. HIV/AIDS epidemiology, pathogenesis, prevention, and treatment. Lancet 368: 489-504
131) Smith DM, May SJ, Tweeten S, Drumright L, Pacold ME, Kosakovsky Pond SL, Pesano RL, Lie YS, Richman DD, Frost SD, Woelk CH, Little SJ. A public health model for the molecular surveillance of HIV transmission in San Diego, California. AIDS 2009; 23(2): 225–232
132) Stadler T, Kouyos R, von Wyl V, Yerly S, Bφni J, Bórgisser P, Klimkait T, Joos B, Rieder P, Xie D, Gónthard HF, Drummond AJ, Bonhoeffer S; Swiss HIV Cohort Study. Estimating the basic reproductive number from viral sequence data. Mol Biol Evol 2012; 29(1): 347–357
133) Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol 2018; 4(1): vey016
134) Takebe Y, Uenishi R, Li X. 2008. Global molecular epidemiology of HIV: Understanding the genesis of AIDS pandemic. Advances in Pharmacology 56: 1-25
135) Tavare, S. (1986). Some probabilistic and statistical problems in the analysis of DNA sequences. In R. M.Miura (Ed.), Some mathematical questions in biology—DNA sequence analysis (pp. 57-86). Providence (RI): American Mathematical Society.

136) Taylor JO. 1995. Expectation, appraisal outcome, and coping for persons with AIDS. The intervention of a patient-based health information system. Doctoral dissertation. University of Wisconsin, Madison, WI
137) Tebit, D.M. and E.J. Arts, Tracking a century of global expansion and evolution of HIV to drive understanding and to combat disease. Lancet Infect Dis, 2011. 11(1): p. 45-56.
138) Trask SA, Derdeyn CA, Fideli U, Chen Y, Meleth S, Kasolo F, Musonda R, Hunter E, Gao F, Allen S, Hahn BH. Molecular epidemiology of human immunodeficiency virus type 1 transmission in a heterosexual cohort of discordant couples in Zambia. J Virol 2002; 76(1): 397–405
139) Turk T, Bachmann N, Kadelka C, Bóni J, Yerly S, Aubert V, Klimkait T, Battegay M, Bernasconi E, Calmy A, Cavassini M, Furrer H, Hoffmann M, Gónthard HF, Kouyos RD, Aubert V, Battegay M, Bernasconi E, Bфni J, Braun DL, Bucher HC, Calmy A, Cavassini M, Ciuffi A, Dollenmaier G, Egger M, Elzi L, Fehr J, Fellay J, Furrer H, Fux CA, Gónthard HF, Haerry D, Hasse B, Hirsch HH, Hoffmann M, Hфsli I, Kahlert C, Kaiser L, Keiser O, Klimkait T, Kouyos RD, Kovari H, Ledergerber B, Martinetti G, Martinez de Tejada B, Marzolini C, Metzner KJ, Móller N, Nicca D, Pantaleo G, Paioni P, Rauch A, Rudin C, Scherrer AU, Schmid P, Speck R, Stфckle M, Tarr P, Trkola A, Vernazza P, Wandeler G, Weber R, Yerly S. Assessing the danger of self-sustained HIV epidemics in heterosexuals by population based phylogenetic cluster analysis. eLife 2017; 6: e28721
140) UNAIDS. Global HIV & AIDS statistics - 2018 fact sheet. 2018 [cited 2018; Available from: http://www.unaids.org/en/resources/fact-sheet.
141) UNAIDS. 2012. Global report: UNAIDS report on the global AIDS epidemic 2012
142) Van de Peer, Y. (2009). Phylogenetic inference based on distance methods: Theory. In P. Lemey, M. Salemi & A. M. Vandamme (Eds.), The Phylogenetic Handbook. A Practical Approach to Phylogenetic Analysis and Hypothesis Testing. New York: Cambridge University Press.
143) Van der Groen G, Nyambi PN, Beirnaert E, Davis D, Fransen K, Heyndrickx L, Ondoa P, Van der Auwera G, Janssens W. 1998. Genetic variation of HIV type 1: Relevance of interclade variation to vaccine development. AIDS Res Hum Retroviruses. 14 (Suppl. 3): S211-S221
144) Vandamme, A. M. (2009). Basic concepts of molecular evolution. In P. Lemey, M. Salemi & A. M.Vandamme (Eds.), The Phylogenetic Handbook. A Practical Approach to Phylogenetic Analysis and Hypothesis Testing. New York: Cambridge University Press.
145) Volz EM, Le Vu S, Ratmann O, Tostevin A, Dunn D, Orkin C, O'Shea S, Delpech V, Brown A, Gill N, Fraser C; UK HIV Drug Resistance Database. Molecular epidemiology of HIV-1 subtype B reveals heterogeneous transmission risk: implications for intervention and control. J Infect Dis 2018; 217(10): 1522–1529
146) Wang WK, Chen MY, Chuang CY, Jeang KT, Huang LM. 2000. Molecular biology of human immunodeficiency virus type 1. J Microbiol Immunol Infect 33: 131-140
147) Wang X, Wu Y, Mao L, Xia W, Zhang W, Dai L, Mehta SR, Wertheim JO, Dong X, Zhang T,Wu H, Smith DM. Targeting HIV prevention based on molecular epidemiology among deeply sampled subnetworks of men who have sex with men. Clin Infect Dis 2015; 61(9): 1462–1468
148) Wertheim JO, Kosakovsky Pond SL, Forgione LA, Mehta SR, Murrell B, Shah S, Smith DM, Scheffler K, Torian LV. Social and genetic networks of HIV-1 transmission in New York City. PLoS Pathog 2017; 13(1): e1006000
149) Wertheim JO, Leigh Brown AJ, Hepler NL, Mehta SR, Richman DD, Smith DM, Kosakovsky Pond SL. The global transmission network of HIV-1. J Infect Dis 2014; 209(2): 304–313

150) Wertheim JO, Murrell B, Mehta SR, Forgione LA, Kosakovsky Pond SL, Smith DM, Torian LV. Growth of HIV-1 molecular transmission clusters in New York City. J Infect Dis 2018; 218(12): 1943–1953
151) Wilkinson E, Junqueira DM, Lessells R, Engelbrecht S, van Zyl G, de Oliveira T, Salemi M. The effect of interventions on the transmission and spread of HIV in South Africa: a phylodynamic analysis. Sci Rep 2019; 9(1): 2640
152) World Health Organization (WHO). Number of people (all ages) living with HIV, Estimates by WHO region. 2017 [cited 2018; Available from: http://apps.who.int/gho/data/view.main.22100WHO?lang=en.
153) World Health Organization (WHO), WHO case definitions of HIV for surveillance and revised clinical staging and immunological classification of HIV-related disease in adults and children. 2007.
154) Xia X. (2018) Nucleotide Substitution Models and Evolutionary Distances. In: Bioinformatics and the Cell. Springer, Cham
155) Yang Z. Phylogenetic analysis using parsimony and likelihood methods, J. Mol. Evol., 42, 294-307 (1996).
156) Yebra G, Hodcroft EB, Ragonnet-Cronin ML, Pillay D, Brown AJ; PANGEA_HIV Consortium; ICONIC Project. Using nearly full genome HIV sequence data improves phylogeny reconstruction in a simulated epidemic. Sci Rep 2016; 6(1): 39489
157) Yerly S, Vora S, Rizzardi P, Chave JP, Vernazza PL, Flepp M, Telenti A, Battegay M, Veuthey AL, Bru JP, Rickenbach M, Hirschel B, Perrin L; Swiss HIV Cohort Study. Acute HIV infection: impact on the spread of HIV and transmission of drug resistance. AIDS 2001; 15(17): 2287–2292
158) Zagordi O, Geyrhofer L, Roth V, Beerenwinkel N. Deep sequencing of a genetically heterogeneous sample: Local haplotype reconstruction and read error correction. J Comput Biol.2010b; 17:417–428. [PubMed: 20377454]
159) Zagordi O, Klein R, Daumer M, Beerenwinkel N. Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies. Nucleic Acids Res. 2010a; 38:7400–7409. [PubMed: 20671025]
160) Zang J. Performance of likelihood ration tests of evolutionary hypotheses under inadequate substitution models, Mol. Biol. Evol., 16, 868-75 (1999).
161) Zharkikh A. and Li W.-H. Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences: I. four taxa with a molecular clock, Mol. Biol. Evol., 9, 1119-47 (1992).
162) Αλαχιώτης, Σ., Εισαγωγή στη Γενετική. 2005, Αθήνα: Ελληνικά γράμματα.
163) Αλαχιώτης, Σ., Εισαγωγή στην Εξέλιξη. 2007, Αθήνα: Εκδοτικός Οίκος ΛΙΒΑΝΗ.
164) Κέντρο Ελέγχου & Πρόληψης Νοσημάτων (ΚΕ.ΕΛ.Π.ΝΟ.). [cited 2018; Available from: http://www.keelpno.gr.
165) Εμίρης Ζ. Ιωάννης, Ανάλυση βιολογικών αλληλουχιών: Πιθανοκρατικά μοντέλα πρωτεινών και νουκλεικών οξέων
166) Κέντρο Ελέγχου & Πρόληψης Νοσημάτων (ΚΕ.ΕΛ.Π.ΝΟ.), Επιδημιολογική Επιτήρηση της HIV/AIDS λοίμωξης στην Ελλάδα, Δηλωθέντα Στοιχεία έως 31.12.2017, Υπουργείο Υγείας, Editor. 2017.
167) Μαρμάρας, Β. and Μ. Λαμπροπούλου-Μαρμάρα, Βιολογία Κυττάρου Μοριακή Προσέγγιση. 5η ed. 2005, Πάτρα: Εκδόσεις Typorama. 63-65.
168) Παρασκευής, Δ., Μαγιορκίνης, Γ., & Χατζάκης, Α. (2015). Βασικές αρχές μοριακής εξέλιξης και φυλογενετικής ανάλυσης. Αθήνα: Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών, Ιατρική σχολή, Εργαστήριο υγιεινής και επιδημιολογίας.