



Πανεπιστήμιο Πειραιώς – Τμήμα Πληροφορικής  
Πρόγραμμα Μεταπτυχιακών Σπουδών  
«Πληροφορική»

**Μεταπτυχιακή Διατριβή**

Τίτλος Διατριβής	<b>ΑΝΑΓΝΩΡΙΣΗ ΜΟΡΙΑΚΩΝ ΑΚΟΛΟΥΘΙΩΝ</b>
Όνοματεπώνυμο Φοιτητή	<b>ΑΘΑΝΑΣΙΟΣ ΠΑΥΛΙΔΗΣ</b>
Πατρώνυμο	<b>ΕΥΑΓΓΕΛΟΣ</b>
Αριθμός Μητρώου	<b>ΜΠΠΛ/07006</b>
Επιβλέπων	<b>ΑΓΓΕΛΟΣ ΠΙΚΡΑΚΗΣ, ΛΕΚΤΟΡΑΣ</b>

Ημερομηνία Παράδοσης **ΣΕΠΤΕΜΒΡΙΟΣ 2011**

---



**Τριμελής Εξεταστική Επιτροπή**

(υπογραφή)

ΑΓΓΕΛΟΣ ΠΙΚΡΑΚΗΣ  
ΛΕΚΤΟΡΑΣ

(υπογραφή)

ΜΙΧΑΗΛ ΨΑΡΑΚΗΣ  
ΕΠΙΚΟΥΡΟΣ ΚΑΘΗΓΗΤΗΣ

(υπογραφή)

ΧΑΡΑΛΑΜΠΟΣ  
ΚΩΝΣΤΑΝΤΟΠΟΥΛΟΣ  
ΛΕΚΤΟΡΑΣ

## ΠΕΡΙΛΗΨΗ

Για την κατανόηση των πληροφοριών που κωδικοποιούνται σε γενετικές ακολουθίες, βάσεις δεδομένων συχνά συγκρίνονται για τον εντοπισμό ομοιοτήτων, σε μια διαδικασία που καλείται ευθυγράμμιση ακολουθίας.

Η σύγκριση Γενετικών ακολουθιών είναι ένα από τα σημαντικότερα ερωτήματα στην υπολογιστική Βιολογία. Ο αλγόριθμος που χρησιμοποιείται περισσότερο για αυτό το σκοπό είναι ο BLAST, ο οποίος απαιτεί λιγότερο υπολογιστικό χρόνο, αλλά έχει μικρότερη ευαισθησία, π.χ. σημαντικές ομοιότητες μπορεί να μη ληφθούν υπόψη. Ο αλγόριθμος Smith-Waterman επιτυγχάνει υψηλή ευαισθησία, ωστόσο, απαιτεί περισσότερο υπολογιστικό χρόνο. Σε αυτή την εργασία γενετικό υλικό από Ιούς συγκρίνεται με τη χρήση των αλγορίθμων BLAST και Smith and Waterman

## ABSTRACT

In order to understand the information encoded by DNA sequences, databases containing large amount of DNA sequence information are frequently compared for matching patterns. This is called sequence alignment.

DNA sequence comparison is one of the most important and basic problems in computational biology. The most popular algorithm for this operation is BLAST which gives high speed but low sensitivity, i.e. significant matches may be missed by the searches. Another algorithm, the Smith-Waterman algorithm, is a more computationally expensive algorithm but achieves higher sensitivity. In this paper viral genetic material is compared by means of two algorithms, BLAST and Smith and Waterman algorithm.

# ΓΑΜΕΤΕΡΗΜΟ ΠΕΡΑΝ

## ΕΙΣΑΓΩΓΗ



Ελαιογραφία του μοναχού Gregor Mendel από τον Alois Zenker 1884, Μοναστήρι του St. Thomas, Brno, Τσεχική Δημοκρατία.

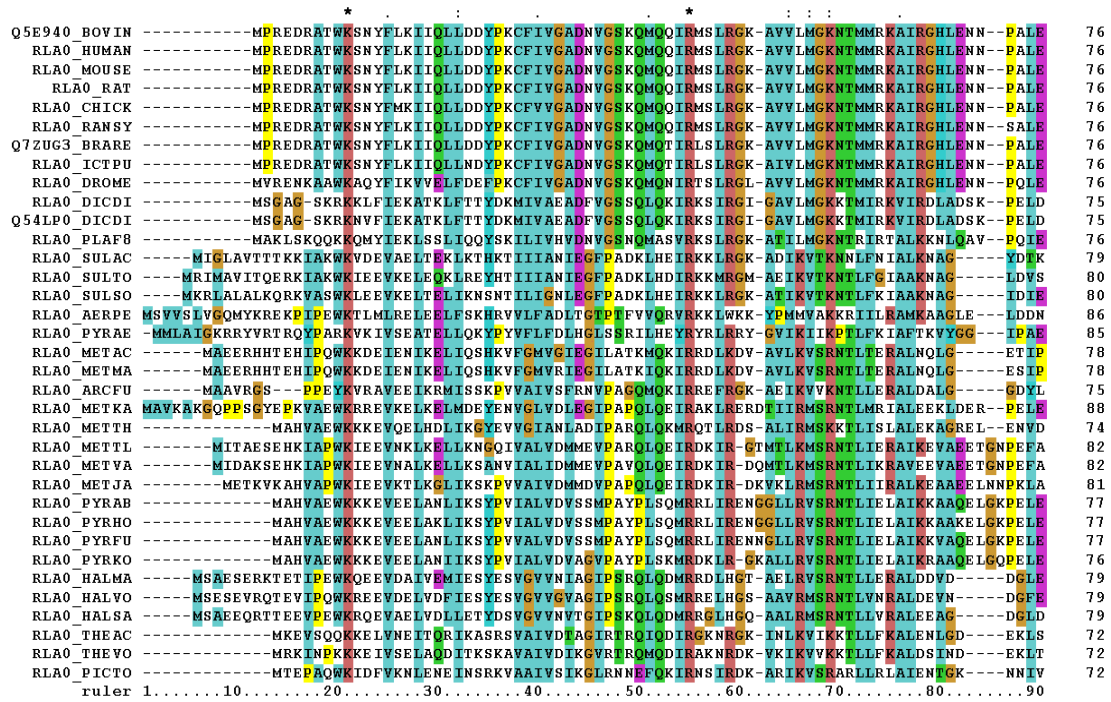
Η Γενετική κατέχει κεντρικό ρόλο σε ολόκληρο το αντικείμενο της βιολογίας. Για κάθε σοβαρό σπουδαστή της φυτικής, ζωικής ή μικροβιακής ζωής, η κατανόηση της γενετικής είναι απαραίτητη. Επιπλέον η γενετική, όσο λίγοι επιστημονικοί κλάδοι, έχει γίνει κεντρικός σταθμός σε πολλές πτυχές της ανθρώπινης διαβίωσης. Αρχικά πρέπει να καθορίσουμε τι είναι η γενετική. Μερικοί την καθορίζουν ως τη μελέτη της κληρονομικότητας, και κληρονομικά φαινόμενα έχουν ενδιαφέρον για τους ανθρώπους πριν από την αυγή του πολιτισμού. Πολύ πριν από τη ύπαρξη της βιολογίας και της γενετικής, ως τις επιστήμες που γνωρίζουμε σήμερα, οι αρχαίοι λαοί είχαν βελτιώσει καλλιέργειες φυτών και τα οικόσιτα ζώα, επιλέγοντας επιθυμητές ιδιότητες για αναπαραγωγή. Πρέπει, να ήταν επίσης προβληματισμένοι σχετικά με την κληρονομικότητα των χαρακτηριστικών στον άνθρωπο. Η γενετική ως ένα σύνολο αρχών και διαδικασιών ανάλυσης δεν άρχισε παρά μετά το 1860 όταν ένας Αυγουστιανός μοναχός με το όνομα Γκρέγκορ Μέντελ εκτέλεσε μια σειρά πειραμάτων και επισήμανε την ύπαρξη βιολογικών στοιχείων που ονομάζονται γονίδια. Οι γενετιστές είτε μελετούν σε μοριακό, κυτταρικό, οργανικό, οικογενειακό, πληθυσμιακό, ή εξελικτικό επίπεδο, τα γονίδια είναι πάντα κεντρικά στις σπουδές τους. Ένα γονίδιο είναι ένα τμήμα ενός μορίου που ονομάζεται δεσοξυριβονουκλεϊκό οξύ, η συντομογραφία του DNA. Το DNA, το κληρονομικό υλικό που περνά από τον έναν οργανισμό στον άλλο, υπαγορεύει τις εγγενείς ιδιότητες ενός είδους. Κάθε κύτταρο ενός οργανισμού έχει μια ή δύο σειρές του βασικού DNA και λέγεται γονιδίωμα. Το γονιδίωμα αποτελείται από ένα ή περισσότερα εξαιρετικά μεγάλα μόρια του DNA, που ονομάζονται χρωμοσώματα. Τα γονίδια είναι λειτουργικές περιοχές του DNA που δραστηριοποιούνται σε τμήματα κατά μήκος των χρωμοσωμάτων. Σε σύνθετους οργανισμούς τα χρωμοσώματα αριθμούν σε τάξεις των δεκάδων, αλλά τα γονίδια αριθμούν σε τάξεις των δεκάδων χιλιάδων.(20)

Ένας από τους τρέχοντες στόχους της πρωτεϊμικής (proteomics) είναι η χαρτογράφηση των δικτύων των πρωτεϊνικών αλληλεπιδράσεων ενός μεγάλου αριθμού μοντέλου οργανισμών. Οι πληροφορίες αλληλεπίδρασης μεταξύ πρωτεϊνών επιτρέπουν τη λειτουργία μιας πρωτεΐνης που θα καθορίζεται από τη θέση της σε ένα σύνθετο πλέγμα αλληλεπιδρώντων πρωτεϊνών. Η

πρόσβαση στις πληροφορίες αυτές θα βοηθήσει σε μεγάλο βαθμό τη βιολογική έρευνα και, ενδεχομένως, θα καταστεί ευκολότερη η ανακάλυψη φαρμακευτικών ουσιών. Αρχικά, η υπολογιστική πρόβλεψη της αλληλεπίδρασης μεταξύ πρωτεϊνών ήταν αυστηρά περιορισμένη στις πρωτεΐνες των οποίων οι τρισδιάστατες (3-D) δομές είχαν καθοριστεί.

Οι πρόσφατες εξελίξεις στο γονιδίωμα αλληλουχίας έχουν, ωστόσο, παρέχει ένα σημαντικό πλούτο γονιδιωματικών πληροφοριών. Είναι πλέον δυνατό να καθοριστεί το γονιδιωματικό πλαίσιο ενός δεδομένου γονιδίου σε ένα πλήρες γονιδίωμα. (23)

Η απευθείας ανάλυση των ακολουθιών που παράγονται καθημερινά είναι πρακτικά αδύνατη λόγω του μεγάλου αριθμού των αλληλουχιών που ανακαλύπτονται καθημερινά.



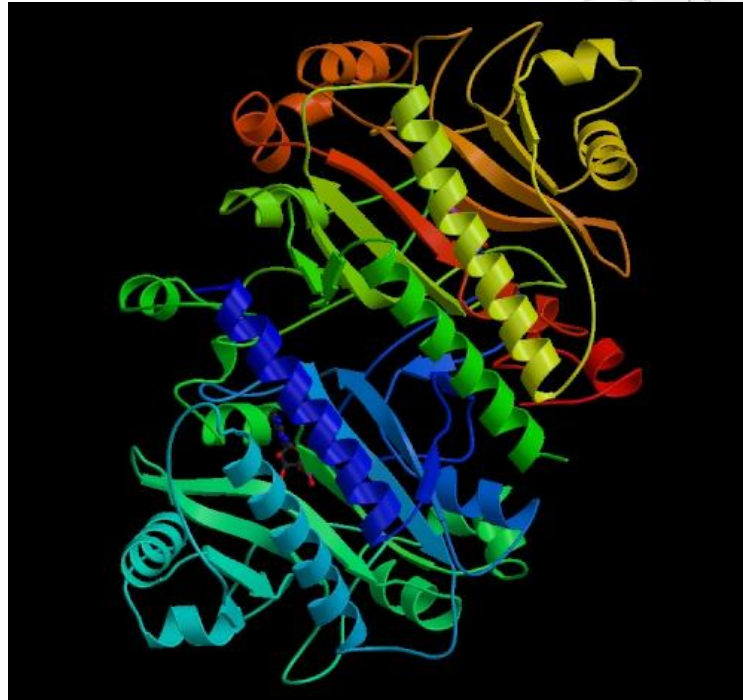
Σχήμα 2: Παράδειγμα συστοίχισης πρωτεϊνών (29)

Γενικά στοίχιση μπορεί να εφαρμοστεί είτε μεταξύ δύο είτε μεταξύ περισσότερων πρωτεϊνών. Ο αλγόριθμος Smith-Waterman εφαρμόζεται σε περιοχές της ακολουθίας, ενώ ο αλγόριθμος Needleman-Wunch συστοίχιζει ολόκληρες τις ακολουθίες.

Σε αυτή την εργασία γενετικό υλικό από ιούς επιλέχτηκε για σύγκριση χρησιμοποιώντας τον αλγόριθμο Smith-Waterman με τη χρήση του Blast και του Matlab. Το γενετικό υλικό επιλέχθηκε από τη βάση γενετικού υλικού του NCBI (National Center for Biotechnology Information). Επειδή η δομή είναι εκείνη που καθορίζει και τη λειτουργία των βιολογικών μορίων, σκοπός μας είναι η εξαγωγή συμπερασμάτων στην εξέλιξη του γενετικού υλικού και πως αυτή η εξέλιξη εμφανίζεται με γεωγραφικά, χρονικά, εξελικτικά και κοινωνικά κριτήρια. Επίσης, θα εξετασθούν πιθανές αδυναμίες των δυο αλγόριθμων στην προσπάθεια εξαγωγής συμπερασμάτων. Για τις βάσεις δεδομένων των νουκλεοτιδικών ακολουθιών οι παρακάτω ιστότοποι χρησιμοποιήθηκαν:

- α) GenBank – NCBI (<http://www.ncbi.nlm.nih.gov/Genbank/index.html>).
- β) EMBL DataBank (<http://www.ebi.ac.uk/embl/>).
- γ) DDBJ (<http://www.ddbj.nig.ac.jp/Welcomes.html>).

## ΤΟ ΓΕΝΕΤΙΚΟ ΥΛΙΚΟ ΚΑΙ ΟΙ ΠΡΩΤΕΙΝΕΣ



Σχήμα3: ΠΑΡΑΔΕΙΓΜΑ ΔΟΜΗΣ ΠΡΩΤΕΙΝΗΣ (Jinbo Xu, School of Computer Science, University of Waterloo)

Στο κύτταρο και *in vitro*, κάθε μόριο του RNA και των πρωτεϊνών διπλώνει σε μια τρισδιάστατη διάρθρωση, και αυτό είναι αυτό που καθορίζει τη δομή της βιοχημικής του λειτουργίας. Η κατανόηση της διαδικασίας αναδίπλωσης αυτών των βιομορίων και η πρόβλεψη της τρισδιάστατης δομής τους, από την πρωτογενή δομή τους είναι δύο από τα βασικά αναπάντητα ερωτήματα στη μοριακή βιολογία. (19)

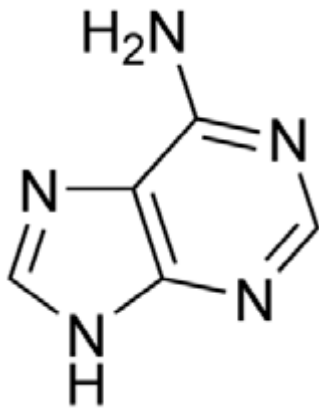
Τα νουκλεϊκά οξέα, DNA και RNA, είναι πολυνουκλεοτίδια, δηλαδή πολυμερή με βασική επαναλαμβανόμενη μονάδα το νουκλεοτίδιο. Κάθε νουκλεοτίδιο αποτελείται από τρία στοιχεία:

1. ένα κυκλικό σάκχαρο αποτελούμενο από πέντε άτομα άνθρακα (πεντόζη). Στην περίπτωση των ριβονουκλεοτιδίων (RNA) πρόκειται για μία ριβόζη, ενώ στην περίπτωση των δεσοξυριβονουκλεοτιδίων (DNA) το σάκχαρο είναι μία δεσοξυριβόζη. Η διαφορά ριβόζης και δεσοξυριβόζης εντοπίζεται στο άτομο  $C_2$ , όπου υπάρχει αντίστοιχα υδροξυλομάδα (-OH) ή υδρογόνο (H).
2. μία βάση, πουρίνη ή πυριμιδίνη
3. μία φωσφορική ομάδα που είναι συνδεδεμένη στον άνθρακα  $C_5$  του σακχάρου με φωσφοδιεστερικό δεσμό. (7)

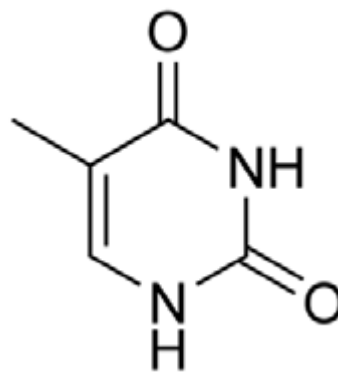
Στα μόρια αυτά που υπάρχουν σε όλα τα κύτταρα βρίσκονται κωδικοποιημένες όλες οι πληροφορίες που προσδιορίζουν τη φύση, ανάπτυξη και λειτουργία του κυττάρου, ιδιαίτερα την βιοσύνθεση των πρωτεϊνών που είναι ακρογωνιαία μόρια για τη φυσιολογία των κυττάρων. Τα νουκλεϊκά οξέα είναι πολυμερή, μεγάλου μοριακού βάρους αλυσίδες.

Οι νουκλεοβάσεις είναι : Αδενίνη, (Adenine, A), Γουανίνη (Guanine, G), Κυτοσίνη (Cytosine, C), Θυμίνη (Thymine, T) και Ουρακίλη (Uracil, U). Οι τέσσερις πρώτες είναι νουκλεοβάσεις του DNA και η ουρακίλη αντικαθιστά την θυμίνη στο RNA. Οι νουκλεοβάσεις παίζουν εξαιρετικά σπουδαία ρόλο για τη μεταφορά της γενετικής πληροφορίας. Η διάταξη που έχουν στους δύο κλώνους της διπλής έλικας των νουκλεϊνικών οξέων και τους δεσμούς-υδρογόνου (H...) που σχηματίζουν μεταξύ τους διατηρούν τους δύο κλώνους ενωμένους. Συγχρόνως μπορούν να διανοίγονται εύκολα (ενζυμικά) για να εκθέσουν τις νουκλεοβάσεις τους και να τις χρησιμοποιήσουν ως συμπληρωματικό καλούπι για το ζευγάρι των βάσεων: Αδενίνη...Θυμίνη (A...T) και Κυτοσίνη...Γουανίνη (C...G). Στο πρώτο ζευγάρι σχηματίζονται δύο δεσμοί-υδρογόνου, ενώ στο δεύτερο τρεις δεσμοί-υδρογόνου. Έτσι οι δύο κλώνοι του DNA είναι συμπληρωματικοί και περιελίσσονται σχηματίζοντας διπλό έλικα. Τα μόρια του DNA έχουν τεράστιο μοριακό βάρος που μπορεί να φθάσει και τα 150 δισεκατομμύρια, ενώ τα μόρια του RNA είναι πολύ μικρότερα, περίπου, μοριακό βάρος 35.000.

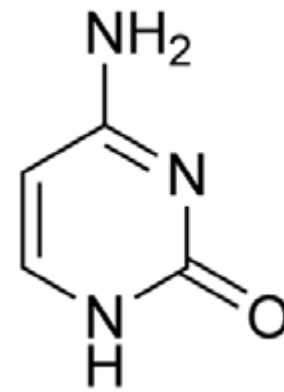
Τα μόρια του RNA και DNA υπόκεινται καθημερινά βλαβερές επιδράσεις από ενδογενή δραστικά χημικά μόρια και ακτινοβολίες, αλλά οι βλάβες επιδιορθώνονται ενζυμικά. Με τη γήρανση των οργανισμών και τη συσσώρευση των βλαβών (μεταλλάξεις) επέρχεται η γήρανση, αλλά και χρόνιες ασθένειες φθοράς όπως οι κακοήθειες νεοπλασίες και νευροεκφυλιστικές ασθένειες.



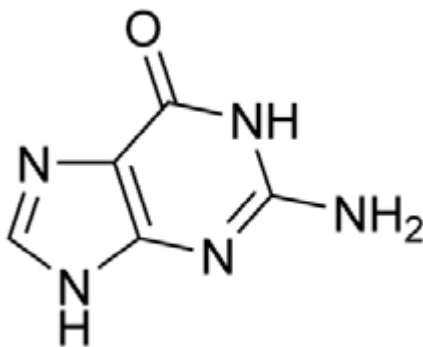
Αδενίνη



Θυμίνη

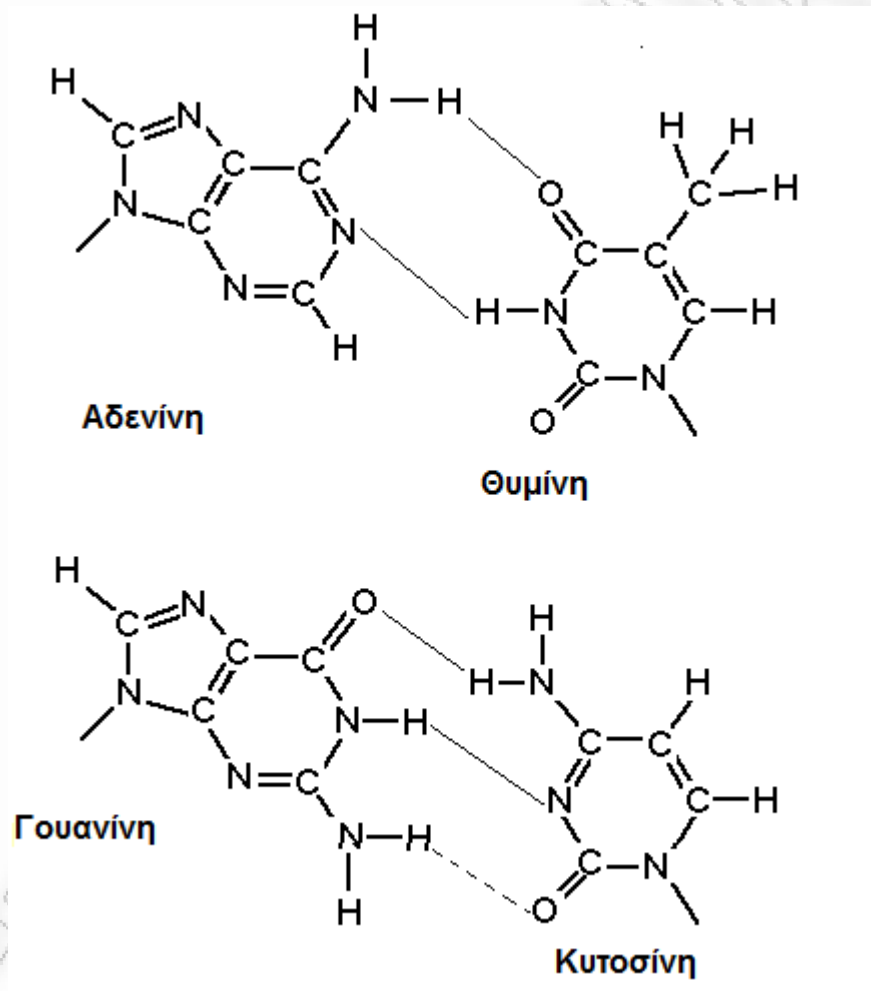


Κυτοσίνη

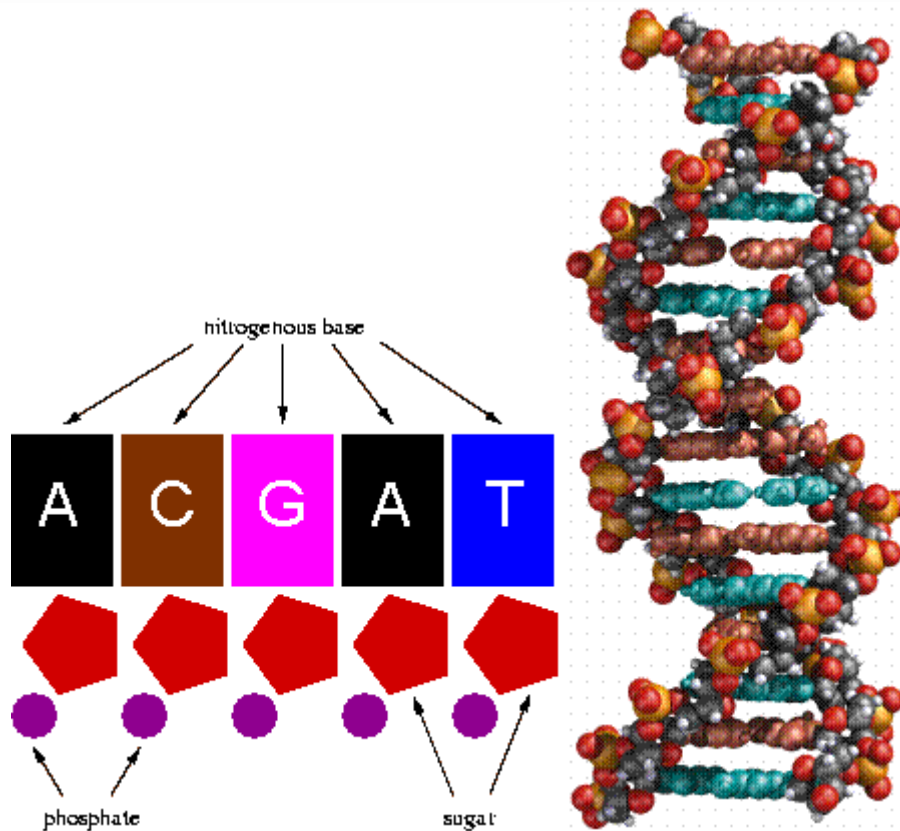


Γουανίνη





Σχήμα 4: Ο τρόπος ενώσεως Αδενίνης με Θυμίνη και Γουανίνης με Κυτοσίνη.

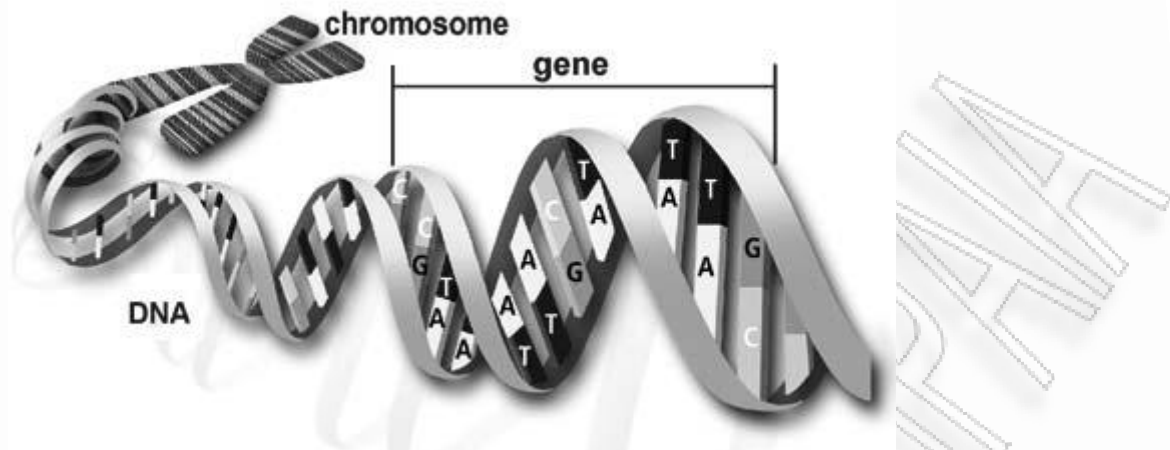


**Σχήμα 5: Το Μόριο του DNA (11)**

Το DNA έχει την πληροφορία, για το πώς να κατασκευαστεί ένας ζωντανός οργανισμός και για την λειτουργικότητά του. Αυτό επιτυγχάνεται με την εντολή στο κύτταρο κατασκευής χαρακτηριστικών πρωτεϊνών για τον οργανισμό αυτό. Για αυτόν τον λόγο οι διάφοροι επιστήμονες ενδιαφέρονται για τον ακριβή προσδιορισμό της ακολουθίας των νουκλεοτιδίων στο DNA.

Αρχικά, αυτό ήταν πολύ δύσκολο να επιτευχθεί και χρειάστηκε πάρα πολύς χρόνος, ακόμη και για μικρά τμήματα γενετικού υλικού. Με το πέρασμα του χρόνου, νέες τεχνικές εφευρέθηκαν και ο προσδιορισμός της αλληλουχίας του DNA μπορούσε να πραγματοποιηθεί πιο γρήγορα. Αναμφίβολα, ένα μεγάλο βήμα έγινε με την εφεύρεση της PCR (Polymerase Chain Reaction, Αλυσιδωτή Αντίδραση Πολυμεράσης), για την κατασκευή εκατομμυρίων αντιτύπων από μια έλικα του DNA. Ακόμα και έτσι όμως η συμπλήρωση της αλληλουχίας του γονιδιώματος ενός ολόκληρου οργανισμού ήταν ανέφικτη, με εξαίρεση εκείνης ενός ιού.

Το 1990, η κυβέρνηση των ΗΠΑ έθεσε σε εφαρμογή το Πρόγραμμα του Ανθρώπινου Γονιδιώματος (*Human Genome Project*). Το σχέδιο αφορούσε την χαρτογράφηση ολόκληρου του ανθρώπινου γονιδιώματος που αποτελείται από ένα σύνολο 3 δισεκατομμυρίων βάσεων με ακρίβεια ενός σφάλματος κάθε 10,000 βάσεις. Η επιστήμη της γονιδιωματικής είχε γεννηθεί, και με αυτήν ο κλάδος των Βιολογικών Επιστημών έχει αλλάξει για πάντα. Η Γονιδιωματική είναι η επιστήμη της κατανόησης των πληροφοριών που περιέχονται στο γονιδίωμα των οργανισμών σε παγκόσμιο επίπεδο. Νέοι κλάδοι της Βιολογίας γεννήθηκαν εξαιτίας της: πρωτεϊνωματική, συγκριτική γονιδιωματική, ιατρική και δομική γονιδιωματική, αλλά και φαρμακευτική. Σημαντικό χαρακτηριστικό της αποτελεί και η χρήση υπολογιστών για την επίτευξη των στόχων της, για αυτό και αναφέρεται συχνά ως «υπολογιστική βιολογία». (8)



**Σχήμα 6:** Η διπλή έλικα του DNA με ένα γονίδιο στην ακολουθία του οριοθετημένη. Τα γονίδια είναι συγκεκριμένες αλληλουχίες βάσεων που κωδικοποιούν τις οδηγίες για την δημιουργία πρωτεϊνών. (12)

**Πίνακας 1:** Ακολουθίες DNA από γονίδια β-σφαιρίνης για οκτώ διαφορετικά είδη.

A Human beta-globin 92 Βάσεις.

ATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGC  
AAGGTGAACGTGGATTAAGTTGGTGGTGAGGCCCTGGGCAG

B Goat alanine beta-globin 86 Βάσεις.

ATGCTGACTGCTGAGGAGAAGGCTGCCGTCACCGGCTTCTGGGGCAAGGTG  
AAAGTGGATGAAGTTGGTGCTGAGGCCCTGGGCAG

C Opossum beta-hemoglobin beta-M-gene 92 Βάσεις.

ATGGTGCACCTGACTTCTGAGGAGAAGAAGTGCATCACTACCATCTGGTCTA  
AGGTGCAGGTTGACCAGACTGGTGGTGAGGCCCTTGGCAG

D Gallus gallus beta-globin 92 Βάσεις.

ATGGTGCACCTGGACTGCTGAGGAGAAGCAGCTCATCACCGGCCTCTGGGGC  
AAGGTCAATGTGGCCGAATGTGGGGCCGAAGCCCTGGCCAG

E Lemur beta-globin 92 Βάσεις.

ATGACTTTGCTGAGTGCTGAGGAGAATGCTCATGTACCTCTCTGTGGGGCA  
AGGTGGATGTAGAGAAAGTTGGTGGCGAGGCCCTTGGGCAG

F Mouse beta-a-globin 94 Βάσεις.

ATGTTGCACCTGACTGATGCTGAGAAGTCTGCTGTCTCTTGCCTGTGGGCA  
AAGGTGAACCCCGATGAAGTTGGTGGTGAGGCCCTGGGCAGG

G Rabbit beta-globin 90 Βάσεις.

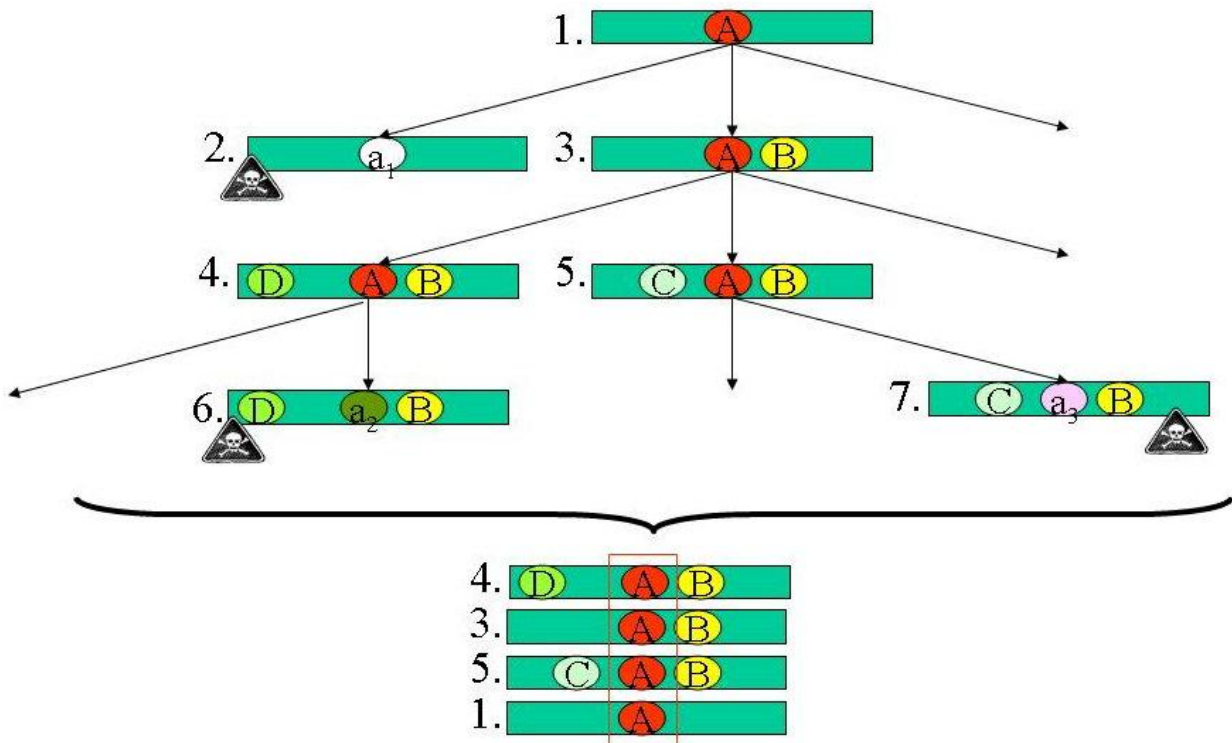
ATGGTGCATCTGTCCAGTGAGGAGAAGTCTGCCGTCCTGACCCTGTGGGGC  
AAGGTGAATGTGGAAGAAGTTGGTGGTGAGGCCCTGGGC

H Rat beta-globin 92 Βάσεις.

ATGGTGCACCTAACTGATGCTGAGAAGGCTACTGTTAGTGGCCTGTGGGGA  
AAGGTGAACCCCTGATAATGTTGGCGCTGAGGCCCTGGGCAG

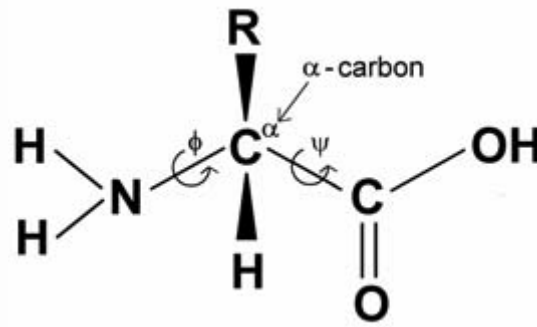
Σύμφωνα και με την θεωρία της εξέλιξης δεν θα πρέπει να ξεχνούμε πως όσα γονίδια είναι αρνητικά για την ζωή, εξαφανίζονται μαζί με τους φορείς τους. Για παράδειγμα, η εξέλιξη

ενός υποθετικού ενζύμου. Οι απόγονοι που δεν έχουν τα επιθυμητά χαρακτηριστικά δεν επιβιώνουν και συνεπώς δεν μπορούν να πολλαπλασιαστούν.

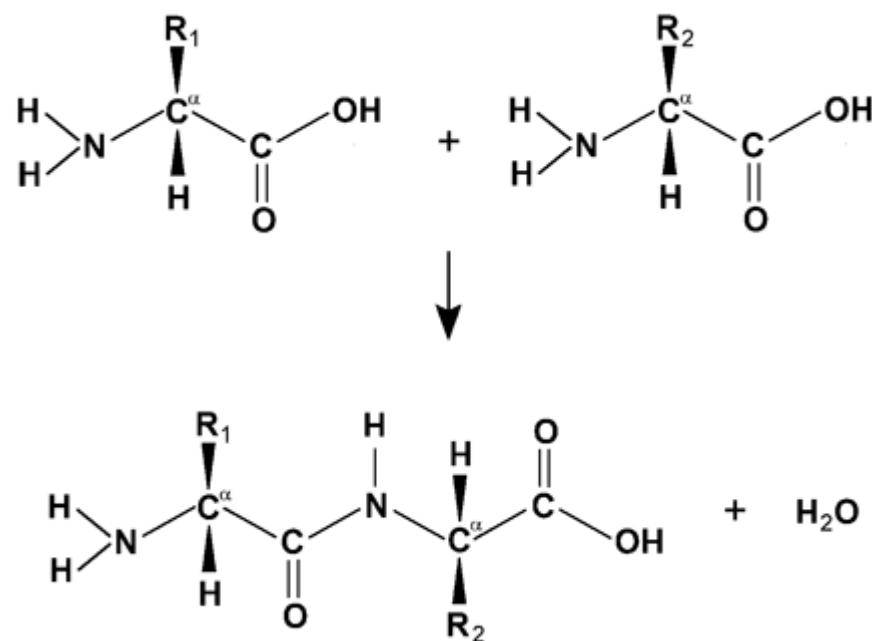


Σχήμα 7: Η εξέλιξη ενός υποθετικού ενζύμου. (22)

Οι πρωτεΐνες είναι αζωτούχες ενώσεις, μεγάλου μοριακού βάρους. Στο μόριο της πρωτεΐνης τα αμινοξέα ενώνονται μεταξύ τους με πεπτιδικό δεσμό. Η κάθε κυτταρική λειτουργία επιτελείται από μία ή περισσότερες πρωτεΐνες. Λειτουργούν ως μεταφορείς, υποστηρικτές του ανοσοποιητικού συστήματος, ένζυμα, ρυθμιστές γονιδίων, παρέχουν μηχανική στήριξη και γενικότερα προσφέρουν λειτουργικές δυνατότητες στον οργανισμό. 20 αμινοξέα χρησιμοποιούνται για τη σύνθεση πρωτεϊνών. Τα αμινοξέα είναι τα δομικά μόρια των πρωτεϊνών. Η χωροδιάταξή τους καθορίζει και τη μεγάλη λειτουργική τους εξειδίκευση. Έχουν όμοια δομή εκτός από την ομάδα R, η οποία και καθορίζει το είδος του αμινοξέος. Έτσι για παράδειγμα, αν η ομάδα R είναι ένα υδρογόνο το αμινοξύ θα είναι η Γλυκίνη (Gly).



Σχήμα 8: Δομή αμινοξέως.



Σχήμα 9: Η δημιουργία πεπτιδικού δεσμού μεταξύ 2 αμινοξέων για την δημιουργία πεπτιδικής αλυσίδας.

## ΒΙΟΠΛΗΡΟΦΟΡΙΚΗ

Μία από τις βασικές τεχνολογίες που συμμετέχουν στην επανάσταση της γονιδιακής ιατρικής είναι η βιοπληροφορική. Ένας από τους τομείς της βιοπληροφορικής είναι η ανάπτυξη ισχυρών αλγορίθμων που θα μας επιτρέψει να αντιμετωπίσουμε τον διαρκώς αυξανόμενο όγκο πληροφοριών που αφορούν μοριακές πληροφορίες από τα ζωντανά συστήματα. Σήμερα υπάρχει μεγάλη αισιοδοξία ότι οι κλινικές επιπτώσεις της μοριακής ιατρικής θα διαδραματίσουν κεντρικό ρόλο στην ανάπτυξη της βιοπληροφορικής.(1)

Στις βιολογικές συμβολοσειρές το πρόβλημα εύρεσης επαναλήψεων σε ένα σύνολο συμβολοσειρών εφαρμόζεται σε πολλά πεδία, όπως για παράδειγμα στην αναζήτηση σε βάσεις δεδομένων. Η εύρεση κανονικοτήτων σε ακολουθίες είναι ιδιαίτερως σημαντικές καθώς εκφράζουν και αναδεικνύουν αρκετές βιολογικές λειτουργίες. Για παράδειγμα, η εύρεση DNA ή πρωτεϊνικών υποσυμβολοσειρών που εμφανίζονται σε ένα ευρύ σύνολο ειδών βοηθά στην εύρεση προτύπων που περιγράφουν τη λειτουργία και τη δομή των ακολουθιών.(3)

Συνεπώς η Βιοπληροφορική γίνεται όλο και πιο σημαντική στον τομέα της έρευνας. Με την ικανότητα άμεσης σε πληροφορίες για την ακολουθία του DNA, οι βιολόγοι έχουν τα εργαλεία για την μελέτη της δομής και της λειτουργίας του Γενετικού υλικού, να μελετήσουν τις εξελικτικές τάσεις και να συσχετιστούν πληροφορίες για το DNA με τις ασθένειες. Για παράδειγμα, δύο γονίδια εντοπίστηκαν να συμμετέχουν στην προέλευση του καρκίνου του μαστού το 1994. Η εν λόγω έρευνα έγινε εφικτή μόνο με τη βοήθεια μιας μεθόδου που είναι γνωστή ως: υψηλών ταχυτήτων σύγκριση γενετικής ακολουθίας. Όλα τα κύτταρα ενός οργανισμού αποτελούνται από ένα είδος γενετικής πληροφορίας.

Μεταφέρονται από μια χημική ουσία γνωστή ως δεοξυριβονουκλεϊνικό οξύ (DNA) στον πυρήνα του κυττάρου. Το DNA είναι ένα πολύ μεγάλο μόριο και το νουκλεοτίδιο είναι η βασική μονάδα αυτού του τύπου του μορίου. Υπάρχουν 4 είδη νουκλεοτιδίων και κάθε ένα έχει διαφορετικές βάσεις, ήτοι: αδενίνη, κυτοσίνη, γουανίνη και θυμίνη. Τα συντετμημένα ονόματά τους είναι: "A", "C", "G" και "T", αντίστοιχα. Είναι δυνατό να συναχθεί ο αρχικός προσδιορισμός της αλληλουχίας του DNA που κωδικοποιεί ένα συγκεκριμένο αμινοξύ.

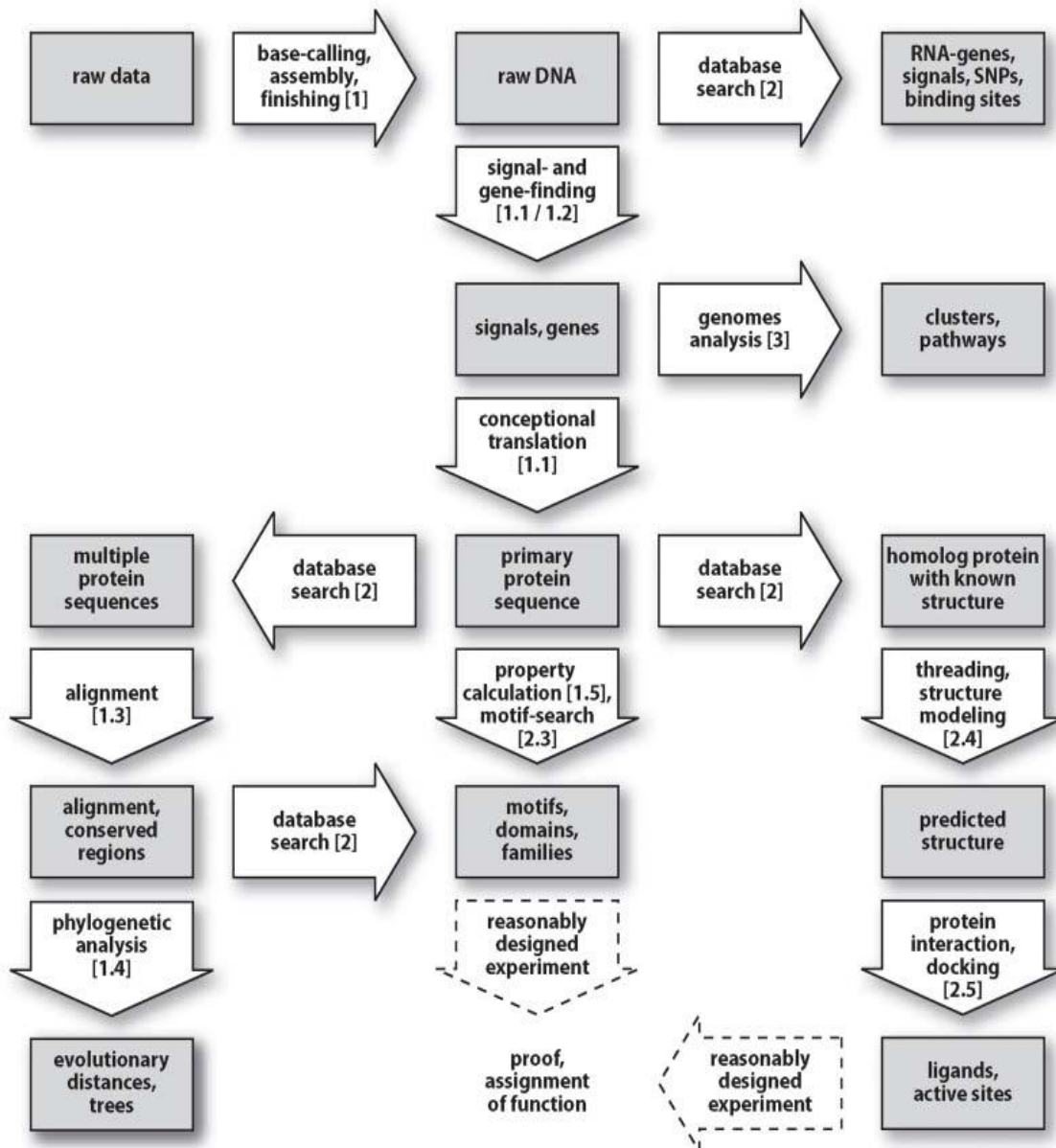
Με τη διαπίστωση της ομοιότητας μεταξύ ενός αριθμού "αμινοξέων που παράγουν " γενετικές αλληλουχίες του DNA και μια πραγματική ακολουθία DNA ενός ατόμου, μπορεί κανείς να εντοπίσει την πρωτεΐνη που κωδικοποιείται από την αλληλουχία του DNA του ανθρώπου. Επιπλέον, εάν επιτύχουμε να βρούμε την ομοιότητα μεταξύ των αλληλουχιών του DNA μεταξύ δύο διαφορετικών ειδών, μπορούμε να κατανοήσουμε την εξελικτική τάση μεταξύ τους. Μια άλλη σημαντική χρήση είναι η μελέτη της σχέσης μεταξύ της ασθένειας και της κληρονομικότητας. Αυτό επιτυγχάνεται με την ευθυγράμμιση των ειδικών αλληλουχιών του DNA ατόμων με νόσο του με εκείνα των απλών ανθρώπων. Αν βρεθούν συσχετισμοί που να μπορούν να χρησιμοποιηθεί για τον εντοπισμό εκείνων που προσβάλλονται από ορισμένες ασθένειες, νέα φάρμακα μπορεί να βρεθούν ή καλύτερες τεχνικές να εφευρευθούν για τη θεραπεία των νοσημάτων. Υπάρχουν πολλές άλλες εφαρμογές της βιοπληροφορικής και αυτός ο τομέας επεκτείνεται με έναν εξαιρετικά γρήγορο ρυθμό.

Ένα ανθρώπινο γονιδίωμα περιέχει περίπου 3 δισεκατομμύρια ζεύγη βάσεων DNA. Προκειμένου να ανακαλύψουμε ποιά αμινοξέα παράγονται από κάθε μέρος της ακολουθίας του DNA, είναι αναγκαίο να βρούμε την ομοιότητα μεταξύ δύο ακολουθιών. Αυτό γίνεται με την εύρεση της ελάχιστης απόστασης χορδών μεταξύ των δύο ακολουθιών και η διαδικασία είναι γνωστή ως ευθυγράμμιση ακολουθίας. Υπάρχουν πολλοί αλγόριθμοι για την ευθυγράμμιση ακολουθίας. Αυτοί που χρησιμοποιούνται πιο συχνά είναι οι: FASTA και BLAST. Οι BLAST και FASTA είναι γρήγοροι αλγόριθμοι που χρησιμοποιούνται για το 'κλάδεμα' της αναζήτησης και εμπλέκονται στην ευθυγράμμιση ακολουθίας χρησιμοποιώντας ευριστικές μεθόδους. Ο αλγόριθμος Smith-Waterman, αποτελεί τη βέλτιστη μέθοδο για την αναζήτηση ομολογίας και την ευθυγράμμιση ακολουθίας σε βάσεις δεδομένων γενετικού υλικού και κάνει όλες τις συγκρίσεις ζευγών μεταξύ των δύο χορδών. Επιτυγχάνει υψηλή ευαισθησία, ωστόσο, ο υπολογιστικός χρόνος που απαιτείται περιορίζει έντονα τη χρήση του.(6)

Η ολοένα και αυξανόμενη βελτίωση τεχνολογιών που χρησιμοποιούνται στις αναλύσεις θα συντελέσει στην ελαχιστοποίηση των όποιων δυσκολιών.(21)

**Σχήμα 10: Παράδειγμα ακολουθιών DNA για ένα τμήμα του L1 γονιδίου για επτά διαφορετικά γκρούπ Α9 Ιών των θηλωμάτων.**

```
HPV16 ATGTGGCTGCCTAGTGAGGCCACTGTCTACTTGCCTCCTGTCCAGTATCTAAGGTTG
HPV35h ATGTGGCGGTCTAACGAAGCCACTGTCTACCTGCCTCCAGTTCAGTGTCTAAGGTTG
HPV31 ATGTGGCGGCCTAGCGAGGCTACTGTCTACTTACCACCTGTCCAGTGTCTAAAGTTG
HPV52 ATGTGGCGGCCTAGTGAGGCCACTGTGTACCTGCCTCCTGTCTCTAAGGTTG
HPV33 ATGTGGCGGCCTAGTGAGGCCACAGTGTACCTGCCTCCTGTCTCTAAGGTTG
HPV58 ATGTGGCGGCCTAGTGAGGCCACTGTGTACCTGCCTCCTGTCTGTCTAAGGTTG
RhPV1 ATGTGGCGGCCTAGTGACTCCAAGGTCTACCTACCACCTGTCTGTCTAAGGTGG
```

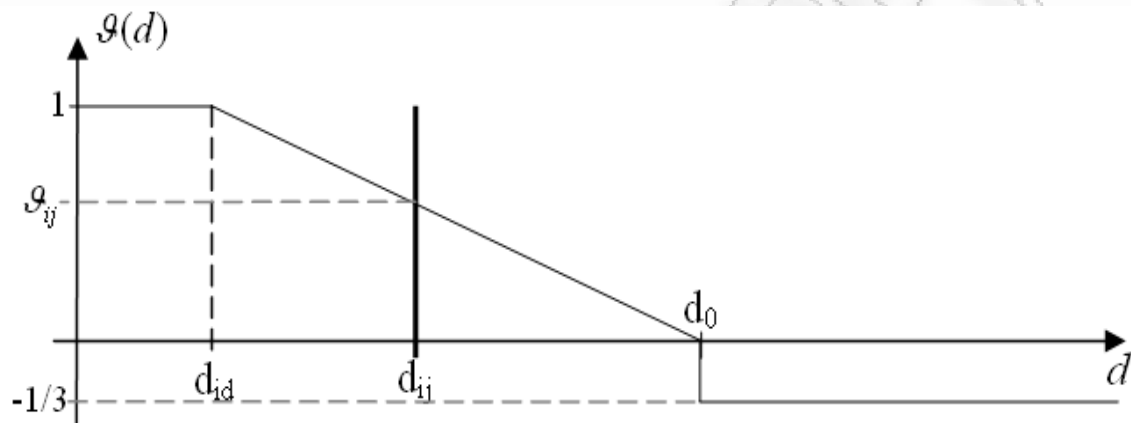


Σχήμα 11: Διαδοχή προγραμμάτων για την διερεύνηση της λειτουργίας DNA ή πρωτεϊνών. Τα βέλη δείχνουν τη δράση των εργαλείων της βιοπληροφορικής.(17)

Δεδομένης μιας ακολουθίας DNA ή μιας ακολουθίας πρωτεΐνης, το πρόβλημα είναι η εξαγωγή από μια βάση δεδομένων όλων των παρόμοιων ακολουθιών. Το βιολογικό κίνητρο πίσω από μια τέτοια λειτουργία είναι η αντιστοίχιση της λειτουργίας ενός άγνωστου γονιδίου ή μιας άγνωστης πρωτεΐνης: στο κύτταρο, μια πρωτεΐνη που υιοθετεί ένα συγκεκριμένο τρισδιάστατο (3D) σχήμα που σχετίζεται με την ακολουθία των αμινοξέων. Αυτό η 3D δομή είναι σημαντική γιατί καθορίζει την λειτουργία των πρωτεϊνών και την αλληλεπίδρασή τους με άλλα μόρια. Θεωρείται δεδομένο ότι οι δύο πρωτεΐνες με σχεδόν ταυτόσημο 3D σχήμα μπορεί να έχουν παρόμοιες λειτουργίες. Έτσι, η εύρεση παρόμοιων ακολουθιών αμινοξέων μπορεί να συσχετιστούν με παρόμοιες δομές, και στη συνέχεια σε παρόμοιες λειτουργίες. Η υπόθεση αυτή, δεν είναι πάντοτε εξακριβωμένη, αλλά, στην πράξη, δίνει καλές ενδείξεις για περαιτέρω έρευνες. Ως εκ τούτου, οι βάσεις δεδομένων της γονιδιωματικής ελέγχονται καθημερινά από χιλιάδες ερευνητές που εξερευνούν αυτή την γιγαντιαία πηγή πληροφοριών για την παρουσίαση νέων γνώσεων.(15)



Ψάχνοντας για παρόμοιες πρωτεΐνες, μέσω της σύγκρισης των δομών τους απαιτούνται αποτελεσματικές και πλήρως αυτοματοποιημένες μέθοδοι. Ως αποτέλεσμα η ανάπτυξη ενός αλγορίθμου και ενός συνόλου εργαλείων που ονομάζονται EAST (Energy Alignment Search Tool, εργαλείο αναζήτησης ταξινόμησης ενέργειας) έχει προταθεί. Αυτό, χρησιμεύει ως εργαλείο για την εξεύρεση ομοιοτήτων μεταξύ πρωτεϊνών από μια βάση δεδομένων με δομές πρωτεϊνών. Η αναζήτηση της ομοιότητας πραγματοποιείται με την σύγκριση και ευθυγράμμιση του ενεργειακού χαρακτηριστικού της πρωτεΐνης που ελήφθη κατά την υπολογιστική διαδικασία με βάση την μηχανική-μοριακή θεωρία.



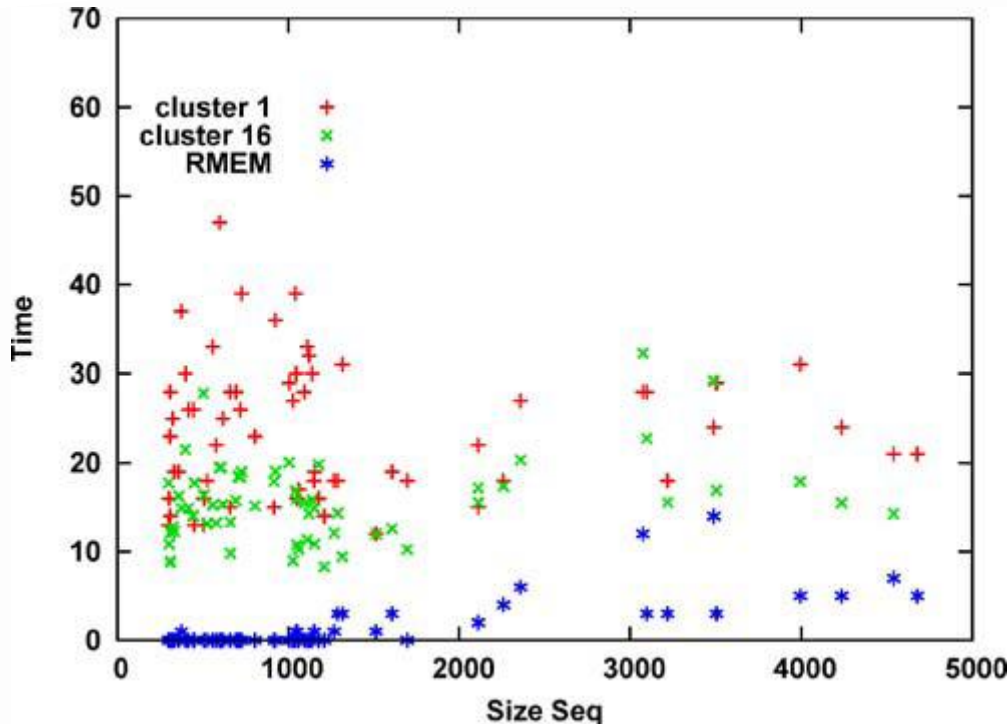
**Σχήμα 12: ο βαθμός ομοιότητας  $\theta(d)$  επιτρέπει την μέτρηση της ομοιότητας που βασίζεται στην απόσταση μεταξύ δυο ενεργειακών σημείων των συγκρινόμενων μορίων. (16)**

Η αναζήτηση της ομοιότητας σε βάση δεδομένων του DNA είναι μια σημαντική λειτουργία στην γονιδιωματική έρευνα. Είναι χρήσιμη για νέες ανακαλύψεις σχετικά με μια αλληλουχία DNA, συμπεριλαμβανομένης της αναζήτησης της θέσης των λειτουργικών εγκαταστάσεων και των επαναλαμβανόμενων. Είναι επίσης χρήσιμο για τη συγκριτική ανάλυση των διαφορετικών ακολουθιών DNA. Η κατά προσέγγιση αντιστοίχιση ακολουθιών είναι προτιμότερη από την ακριβή αντιστοίχιση σε βάσεις δεδομένων γονιδιωματικής λόγω των εξελικτικών μεταλλάξεων στις γονιδιακές ακολουθίες και την παρουσία θορύβου σε μια πραγματική βάση δεδομένων ακολουθίας. Πολλές προσεγγίσεις έχουν αναπτυχθεί για την κατά προσέγγιση αντιστοίχιση ακολουθίας.

Ο πιο θεμελιώδης είναι ο αλγόριθμος ευθυγράμμισης Smith-Waterman, ο οποίος χρησιμοποιεί μια προσέγγιση δυναμικού προγραμματισμού που επιδιώκει τη βέλτιστη ευθυγράμμιση μεταξύ ενός ερωτήματος και την ακολουθία στόχων σε  $O(mn)$  το χρόνο,  $m$  και  $n$  είναι το μήκος των δύο ακολουθιών. (9)

Η υπολογιστική ισχύς σε συνδυασμό με την χρήση αλγορίθμων διαδραματίζει σημαντικό ρόλο στην ανάλυση των αποτελεσμάτων.





Σχήμα 13: Ο χρόνος εκτέλεσης ανάλυσης από 60 (από 300 έως 5.000 νουκλεοτίδια) ερωτημάτων DNA με βάση το ανθρώπινο γονιδίωμα για την αρχιτεκτονική RMEM, με 1 επεξεργαστή, και σε ένα σύμπλεγμα 16 κόμβων. (15)

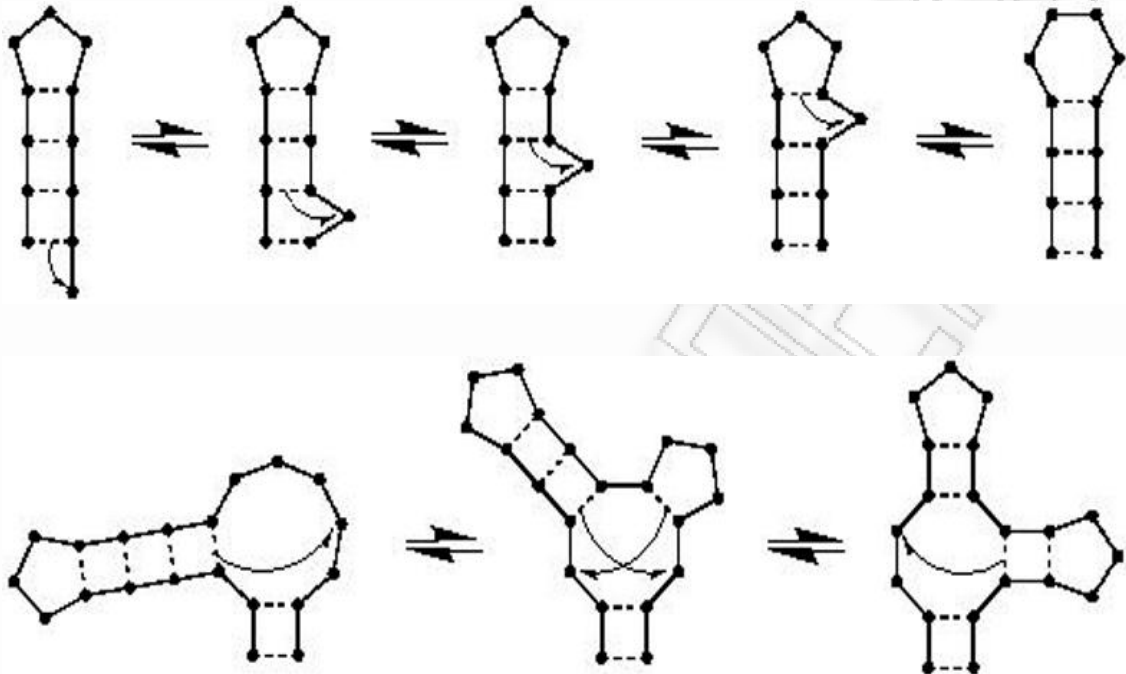
## Ανάλυση Μοριακών Ακολουθιών

Ένα σημαντικό τμήμα των μεθόδων επεξεργασίας κειμένου που χρησιμοποιούνται στον τομέα της μοριακής επεξεργασίας ακολουθίας είναι τυποποιημένοι αλγόριθμοι επεξεργασίας κειμένου. Αυτές οι μεθοδολογίες εξακολουθούν να ανήκουν, τουλάχιστον εν μέρει, στο πεδίο της βιοπληροφορικής και της μοριακής βιολογίας, καθώς και η αξιοποίηση των αποτελεσμάτων με χρήση εφαρμοσμένων αλγορίθμων απαιτεί την εφαρμογή βιολογικών γνώσεων. Η περιοχή επεξεργασίας μοριακών ακολουθιών έχει επίσης πολλές ειδικές ιδιότητες, οι οποίες οδηγούν στην ανάγκη εκπόνησης μεθοδολογιών αφιερωμένων σε ειδικές εφαρμογές. Μερικές σημαντικές ιδιότητες στην ανάλυση μοριακών ακολουθιών είναι οι εξής:

- (i) Το μέγεθος των υπό επεξεργασία δεδομένων.
- (ii) Ειδικά αλφάβητα, για παράδειγμα στον τομέα της γονιδιωματικής το αλφάβητο αποτελείται από τέσσερα γράμματα a, g, c, και t, το καθένα από τα οποία κωδικοποιεί τέσσερις διαφορετικές αζωτούχες βάσεις, αδενίνη, γουανίνη, κυτοσίνη και θυμίνη, που υπάρχουν στο DNA. Μια παρόμοια αλφάβητος καθορίζει το περιεχόμενο του RNA, εκτός από το ότι η θυμίνη (t) αντικαθίσταται από την ουρακίλη (u). Στην πρωτεϊνωματική η αλφάβητος έχει 20 γράμματα που αντιστοιχούν σε 20 αμινοξέα που χρησιμοποιούνται για την κατασκευή της πρωτογενούς δομής των πρωτεϊνών.
- (iii) Ειδικές μορφές επεξεργασίας προβλημάτων κειμένου, όπως προβλήματα που αφορούν για διαφορετικά είδη πολυμορφισμών και μεταλλάξεων σε DNA (π.χ., μονονουκλεοτιδικό ή συνδυασμός επαναλήψεων πολυμορφισμών), ή προβλήματα που σχετίζονται με τη χρήση των αλληλουχιών του DNA, ως δεικτών (πχ, ταξινόμηση των οργανισμών, φυλογενετική ανάλυσή τους), καθώς και προβλήματα που εκτελούν διάφορες συγκρίσεις ακολουθιών. (13)

Επίσης το πρόβλημα της υπολογιστικής πρόβλεψης της δευτεροταγούς δομής των μορίων RNA, αν και εισήχθη για πρώτη φορά πριν από τριάντα χρόνια, συνεχίζει να αποτελεί πεδίο εκτεταμένης έρευνας και ανάπτυξης, ιδίως λόγω της πρόσφατης ανακάλυψης μιας

μεγάλης ποικιλίας νέων τύπων μορίων RNA και της βιολογικής σημασίας τους. Επιπρόσθετο ενδιαφέρον για το πρόβλημα προέρχεται από τον τομέα της συνθετικής βιολογίας, όπου τροποποιημένα μόρια RNA έχουν σχεδιαστεί, και από την μελέτη του πλήρους γονιδιώματος του RNA ιών (το οποίο μπορεί να ανέλθει σε 11.000 ζεύγη βάσεων σε μήκος). (14)



Σχήμα 14: πτυσσόμενες διεργασίες στο RNA. (18)

### Αλγόριθμοι Ευθυγράμμισης Ακολουθιών – Sequence Alignment Algorithms

Ευθυγράμμιση (ή στοίχιση) ακολουθιών είναι η διαδικασία σύγκρισης τους για εύρεση ατομικών χαρακτήρων ή προτύπων χαρακτήρων με την ίδια σειρά στις δύο (ή περισσότερες) ακολουθίες. Πανομοιότυποι χαρακτήρες τοποθετούνται στην ίδια στήλη ενώ μη όμοιοι χαρακτήρες μπορούν να τοποθετηθούν είτε στην ίδια στήλη δηλώνοντας mismatch ή απέναντι από κενό. Υπάρχουν παραπάνω από μία δυνατές ευθυγραμμίσεις. Η βέλτιστη λύση πρέπει να ελαχιστοποιεί τις διαφορές ανάμεσα στις δυο ακολουθίες ή διαφορετικά να μεγιστοποιεί τη συνάρτηση ομοιότητας. Η διαδικασία αυτή στηρίζεται σε πίνακες που βαθμολογούν τις ομοιότητες (matches) και διαφορές (mismatches) μεταξύ διαδοχικών συμβόλων. Τέτοιου τύπου πίνακες είναι οι Dayhoff PAM, BLOSUM κτλ.

Για πολύ λίγες ακολουθίες γνωρίζουμε τη δομή και τη λειτουργία τους. Η ευθυγράμμιση μας παρέχει λειτουργικές, δομικές και εξελικτικές πληροφορίες. Μπορούμε να ευθυγραμμίσουμε δυο ακολουθίες και αν είναι αρκετά όμοιες πιθανόν να έχουν τον ίδιο πρόγονο, επίσης πιθανόν να έχουν την ίδια δομή και λειτουργία οι οργανισμοί. Αν για μία από τις ακολουθίες γνωρίζουμε δομή και λειτουργία και την έχουμε ευθυγραμμίσει με μια άγνωστη ακολουθία μπορούμε να εξαγάγουμε παρόμοια συμπεράσματα και για την άγνωστη. Με την ίδια λογική είναι δυνατός και ο εντοπισμός μεταλλάξεων στα γονίδια, πράγμα πολύ σημαντικό, ιδιαίτερα για την νόσο του καρκίνου.

Μπορούμε να έχουμε ολική ευθυγράμμιση (global alignment) ή τοπική ευθυγράμμιση (local alignment). Στην ολική ευθυγράμμιση, η ευθυγράμμιση γίνεται σε όλο το πεδίο της ακολουθίας και συμπεριλαμβάνει όσα περισσότερα matching ζευγάρια μπορεί. Περιοχές με υψηλή τοπική ομοιότητα αγνοούνται για να επιτευχθεί μεγαλύτερη ολική βαθμολογία. Χαρακτηριστικός αλγόριθμος ολικής ευθυγράμμισης είναι ο Needleman & Wunsch. Στην τοπική

ευθυγράμμιση αναζητούμε περιοχές τοπικής ομοιότητας. Η ευθυγράμμιση αυτή σταματά στις περιοχές που είναι ταυτόσημες ή που έχουν μεγάλη ομοιότητα και δίδεται μεγαλύτερη προτεραιότητα στην εύρεση τέτοιων περιοχών δηλαδή προτύπων (patterns) χαρακτήρων παρά για την εύρεση ατομικών ταυτίσεων χαρακτήρων. Αλγόριθμος τοπικής ευθυγράμμισης είναι ο Smith-Waterman.

### **Αλγόριθμοι Ευθυγράμμισης Πολλαπλών Ακολουθιών**

Η ιδέα της ευθυγράμμισης περισσότερων των 2 ακολουθιών οδήγησε στην ανάπτυξη αυτών των αλγορίθμων και αποτελεί φυσική γενίκευση της ευθυγράμμισης 2 ακολουθιών.

Οι αλγόριθμοι FASTA, BLAST και CLUSTALw είναι οι χαρακτηριστικότεροι αλγόριθμοι ευθυγράμμισης πολλαπλών ακολουθιών. Η ταχύτητά τους (περίπου 50 φορές ταχύτεροι από τους αλγόριθμους ολικής ευθυγράμμισης που προαναφέραμε) τους καθιστά και ένα σπουδαίο εργαλείο ταχείας αναζήτησης βιολογικών βάσεων δεδομένων.

Γενικότερα η ευθυγράμμιση Πολλαπλών ακολουθιών χρησιμοποιείται για

- 1) την αναγνώριση και αναπαράσταση πρωτεϊνικών οικογενειών και υπεροικογενειών,
- 2) στην αναπαράσταση των χαρακτηριστικών που μεταφέρονται στις ακολουθίες DNA ή στις πρωτεϊνικές ακολουθίες και
- 3) στην αναπαράσταση της εξελικτικής ιστορίας (φυλογενετικά δέντρα) από ακολουθίες DNA ή πρωτεϊνών.

### **Ο Αλγόριθμος FASTA**

Ο FASTA είναι ευριστικός αλγόριθμος τοπικής ευθυγράμμισης και μάλιστα ο πρώτος ευριστικός (heuristic) αλγόριθμος ευρείας χρήσεως αναζήτησης ομοιοτήτων σε βιολογικές βάσεις.

*Ευριστικός* είναι ο αλγόριθμος που βρίσκει λύση γρήγορα και εύκολα χωρίς όμως να εγγυάται ότι είναι και η βέλτιστη. Έτσι είναι περισσότερο προσεγγιστικός αλγόριθμος παρά ακριβής. Συνήθως επιτυγχάνεται η εύρεση λύσεως κοντά στην βέλτιστη.

Ο αλγόριθμος FASTA (fast-a) είναι μια συντομογραφία του FAST-ALL και ονομάστηκε έτσι επειδή χρησιμοποιείται για ταχεία (fast) σύγκριση είτε πρωτεϊνικών ακολουθιών είτε νουκλεοτιδικών.

Ψάχνει να βρει τις βέλτιστες τοπικές στοιχίσεις από την ανίχνευση της ακολουθίας για μικρές αντιστοιχίες που ονομάζονται "λέξεις" (words). Αρχικά, υπολογίζεται το σύνολο των τμημάτων ("init1") στα οποία υπάρχουν πολλαπλές "λέξεις" (words). Τα αποτελέσματα διάφορων τμημάτων μπορούν να αθροιστούν για να παράγουν ένα "initn". Μια βελτιστοποιημένη στοιχίση που περιλαμβάνει κενά (gaps) εμφανίζεται ως "opt". Η ευαισθησία και η ταχύτητα της αναζήτησης είναι αντιστρόφως σχετιζόμενες και ελεγχόμενες από τη μεταβλητή "k-tuple" που προσδιορίζει το μέγεθος μιας λέξης (word).

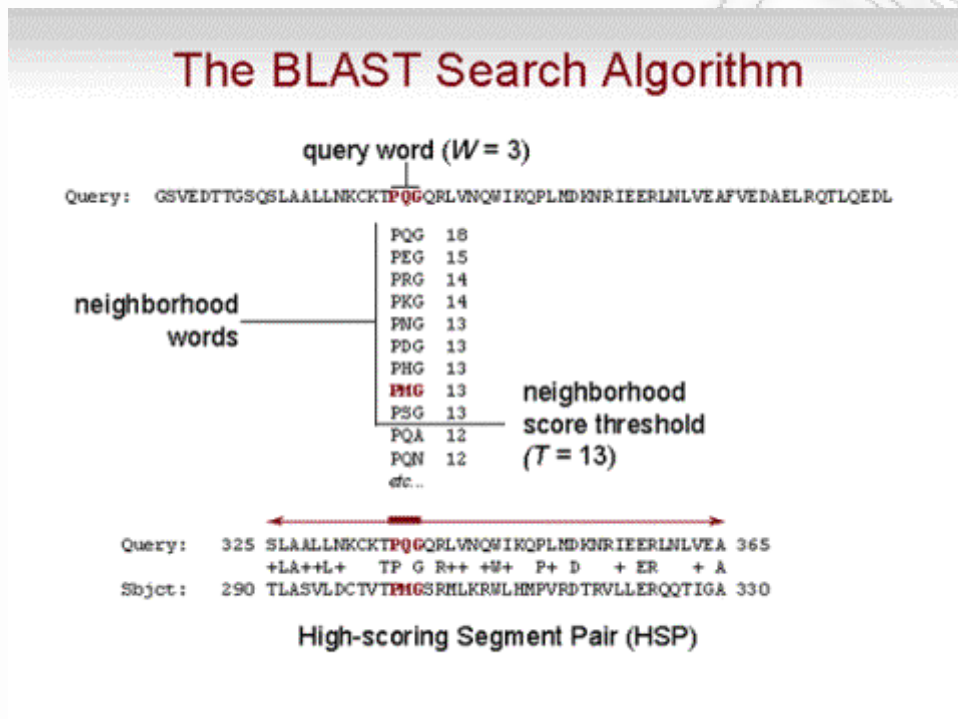
Σαν είσοδο δέχεται δεδομένα σε μορφή FASTA (FASTA format). Η ακολουθία σε μορφή FASTA ξεκινάει με μια γραμμή περιγραφής (single-line description), και ακολουθείται από γραμμές ακολουθίας με 60 χαρακτήρες για τα αμινοξέα σε κάθε γραμμή. Η γραμμή περιγραφής ξεχωρίζει από τις γραμμές της ακολουθίας επειδή στην αρχή της γραμμής περιγραφής υπάρχει το σύμβολο (">").

### **Ο Αλγόριθμος BLAST (Basic Local Alignment Search Tool)**

Ο αλγόριθμος BLAST χρησιμοποιείται για τη σύγκριση μιας ακολουθίας με μια βάση δεδομένων, έχει αρχικά αναπτυχθεί και διατηρείται από το NCBI (National Center for Biotechnology Information) (<http://www.ncbi.nlm.nih.gov/BLAST/>). Είναι ένας ευριστικός (heuristic) αλγόριθμος σύγκρισης ακολουθιών βελτιστοποιημένης ταχύτητας που χρησιμοποιείται για να ψάχνει σε βάσεις ακολουθιών την άριστη τοπική στοιχίση με μια αναζήτηση. Η αρχική αναζήτηση γίνεται για μια λέξη μήκους "W" (3 στο blastp) που δίνει αποτελέσματα για τουλάχιστον "T", όταν συγκρίνεται με την ζητούμενη ακολουθία, χρησιμοποιώντας έναν δεδομένο πίνακα υποκαταστάσεων (substitution matrix). Οι επιτυχείς λέξεις που έχουν score T ή μεγαλύτερο επεκτείνονται και προς τις δύο κατευθύνσεις σε μια

απόπειρα να παραχθούν στοιχείσεις που να υπερβαίνουν το προκαθορισμένο κατώφλι (threshold) "S". Οι περιοχές που ικανοποιούν αυτή τη συνθήκη ονομάζονται HSP (High-scoring Segment Pair). Η παράμετρος "T" καθορίζει την ταχύτητα και την ευαισθησία της αναζήτησης (βλέπε παρακάτω εικόνα). Ενώ το BLAST υπολογίζει τις παραμέτρους της EVD (Extreme Value Distribution), από τις οποίες θα υπολογίσει τη στατιστική σημαντικότητα από εξομοιώσεις που έχει πραγματοποιήσει από πριν, το FASTA τις υπολογίζει από όλες τις άλλες ακολουθίες της βάσης δεδομένων και για αυτό το λόγο είναι και πιο αργό.

Η περιοχή αναζήτησης ('Search' box) του BLAST δέχεται διαφορετικούς τύπους μορφοποιημένης εισόδου και αυτόματα καθορίζει τη μορφή. Οι τρεις μορφοποιήσεις (formats) με τις οποίες μπορεί να εισαχθεί μια ακολουθία και να γίνει η αναζήτηση είναι FASTA format, σκέτη ακολουθία (bare sequence) δηλαδή μόνο οι γραμμές της ακολουθίας χωρίς τη FASTA γραμμή ορισμού και με τη χρήση προσδιοριστών (identifiers).



Σχήμα 15: Ο αλγόριθμος BLAST.

Εισάγωντας μια ακολουθία σε μορφή FASTA δεν επιτρέπεται να υπάρχουν άδειες γραμμές ενδιάμεσα. Οι ακολουθίες παρουσιάζονται με τους πρότυπους αμινοξικούς και νουκλεοτιδικούς κώδικες (IUB/IUPAC), με μερικές εξαιρέσεις. Ο κώδικας γραμμάτων (letter codes) των νουκλεοτιδίων είναι:

- |                                   |                     |
|-----------------------------------|---------------------|
| A --> adenosine                   | M --> A C (amino)   |
| C --> cytidine                    | S --> G C (strong)  |
| G --> guanine                     | W --> A T (weak)    |
| T --> thymidine                   | B --> G T C         |
| U --> uridine                     | D --> G A T         |
| R --> G A (purine)                | H --> A C T         |
| Y --> T C (pyrimidine)            | V --> G C A         |
| K --> G T (keto)                  | N --> A G C T (any) |
| - --> gap of indeterminate length |                     |

Για τα προγράμματα που πραγματοποιούν αναζητήσεις αμινοξικών ακολουθιών ο κώδικας γραμμάτων των αμινοξέων είναι:

A	alanine	P	proline
B	aspartate or asparagine	Q	glutamine
C	cystine	R	arginine
D	aspartate	S	serine
E	glutamate	T	threonine
F	phenylalanine	U	selenocysteine
G	glycine	V	valine
H	histidine	W	tryptophan
I	isoleucine	Y	tyrosine
K	lysine	Z	glutamate or glutamine
L	leucine	X	any
M	methionine	*	translation stop
N	asparagine	-	gap of indeterminate length

*Mike Waterman**Temple Smith*

### **Ο Αλγόριθμος Smith-Waterman.**

Ο αλγόριθμος Smith-Waterman αποτελεί επέκταση του αλγόριθμου Levenshtein και βασίστηκε στον αλγόριθμο Needleman-Wunch. Χρησιμοποιεί την τεχνική του δυναμικού προγραμματισμού, η οποία παίρνει σειρές χαρακτήρων οποιουδήποτε μήκους και καθορίζει το αν υπάρχει μια βέλπστη κατάταξη μεταξύ τους. Βασισμένος σε αυτούς τους υπολογισμούς, ο αλγόριθμος αναθέτει κόστη για κάθε σύγκριση χαρακτήρων. Τα κόστη μπορεί να είναι θετικά για ομοιότητες και αντικαταστάσεις και αρνητικά για εισαγωγές και διαγραφές. (2)

Στον πίνακα Π που χρησιμοποιεί, τα κόστη καταγράφονται, αθροίζονται και το μεγαλύτερο αποτελεί το αποτέλεσμα. Ουσιαστικά, μετατρέπει μια συμβολοσειρά σε μια άλλη κάνοντας συγκεκριμένες ενέργειες σε κάθε μεμονωμένο ζευγάρι χαρακτήρων των συμβολοσειρών αυτών. Έτσι λοιπόν μπορούμε να εισάγουμε, να διαγράψουμε και να αντικαταστήσουμε κάποιον από τους χαρακτήρες που απαρτίζουν την πρώτη συμβολοσειρά. εξετάζει ακολουθίες αντί για μεμονωμένους χαρακτήρες, συγκρίνοντας τμήματα πολλαπλών μεγεθών. Κάθε θέση του πίνακα παίρνει την παρακάτω τιμή:

$$\Pi(i, k) = \max \begin{cases} 0 & // \text{ Αρχικά} \\ \Pi(i-1, k-1) + d(s_i, t_k) & // \text{ για αντικατάσταση ή αντιγραφή} \\ \Pi(i-1, k) - G & // \text{ για εισαγωγή} \\ \Pi(i, k-1) - G & // \text{ για διαγραφή} \end{cases}$$

Όπου  $d[s_i, t_k]$  είναι μια συνάρτηση για αλφαβητική αντιστοίχιση σε κόστη για μεταβολές και  $G$  η τιμή του κόστους για κάθε μεταβολή. Η επόμενη εικόνα παρουσιάζει τον τελικό πίνακα του αλγορίθμου για δύο τυχαίες συμβολοσειρές και για κόστος 0.5.

	A	B	C	N	J	R	Q	C	L	C	R	P	M
A	1	0	0	0	0	0	0	0	0	0	0	0	0
J	0	0.5	0.5	0.5	2	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
C	0	0.5	1.5	0	0	1.5	1.5	3	1.5	3	1.5	1.5	1.5
J	0	0.5	0	1	2.5	1.5	1	1	2.5	2.5	2.5	2.5	2.5
N	0	0.5	0	2.5	0.5	2	2	2	2.5	2	2.5	2	2
R	0	0.5	0	1	2	3.5	2	2	2.5	2	4	2	2
C	0	0.5	1.5	1	2	2	3	4.5	3	4.5	3	3.5	3.5
K	0	0.5	0	1	2	2	3	2.5	4	4	4	4	4
C	0	0.5	1.5	1	2	2	3	4	4	5	4	3.5	3.5
R	0	0.5	0	1	2	3.5	3	2.5	4	3.5	6	4.5	4.5
B	0	2	0	1	2	2	3	3	4	3.5	4.5	5.5	5.5
P	0	0.5	1.5	1.5	2	2	3	2.5	4	3.5	4.5	7	5

Σχήμα 16: Ο Αλγόριθμος Smith-Waterman για την Τοπική Ευθυγράμμιση Ακολουθίας (Local Sequence Alignment).

Για τη σύγκριση δύο ακολουθιών, πρέπει να βρούμε την καλύτερη ευθυγράμμιση μεταξύ τους, η οποία είναι να τοποθετήσουμε την μια ακολουθία κάνοντας σαφή την αντιστοίχια παρόμοιων χαρακτήρων ή συμβολοσειρών από τις αλληλουχίες (1). Σε μια ευθυγράμμιση, χώροι παρεμβάλλονται σε αυθαίρετες τοποθεσίες κατά μήκος των ακολουθιών, ώστε να καταλήξουν με το ίδιο μέγεθος.(4)

Για παράδειγμα σε μια ευθυγράμμιση μεταξύ δυο ακολουθιών  $s$  και  $t$ , το σκορ προκύπτει ως εξής: Για κάθε στήλη, προσθέτουμε +1, εάν οι δύο χαρακτήρες είναι ταυτόσημοι, -1 αν οι χαρακτήρες είναι διαφορετικοί και -2, εάν ένας από τους χαρακτήρες είναι κενός. Η βαθμολογία προκύπτει από το άθροισμα των βαθμολογιών για κάθε στήλη. Το μέγιστο σκορ είναι η ομοιότητα μεταξύ δύο ακολουθιών, που υποδηλώνεται από:  $sim(s, t)$ . Το σχήμα 16 δείχνει την ευθυγράμμιση των ακολουθιών  $s$  και  $t$ , με το σκορ για κάθε στήλη. Σε αυτή την περίπτωση, υπάρχουν εννέα στήλες με τα ίδια χαρακτηριστικά, μια στήλη με ξεχωριστό χαρακτήρα και μία στήλη με ένα κενό, δίνοντας μια συνολική βαθμολογία 6.



Ένα κεντρικό πρόβλημα στην υπολογιστική βιολογία είναι η κατάταξη των πρωτεϊνών σε λειτουργικές και δομικές κατηγορίες με βάση τις αλληλουχίες των αμινοξέων τους. Μέσω της εξέλιξης η δομή τους είναι πιο συντηρημένη από ότι η ακολουθία τους, έτσι ώστε η ανίχνευση ακόμη και μιας πολύ λεπτής σειράς ομοιοτήτων, ή απομακρυσμένης ομολογίας, είναι σημαντική για την πρόβλεψη και τη δομή και συνεπώς της λειτουργίας τους.

Μεταξύ αυτών των αλγορίθμων, ο αλγόριθμος των Smith-Waterman δυναμικού προγραμματισμού, είναι ένας από τις πιο ακριβείς μεθόδους, ενώ ευρετικοί αλγόριθμοι, όπως οι: BLAST και FASTA, μειώνουν την ακρίβεια για τη βελτίωση της αποτελεσματικότητας. Οι μέθοδοι στη συνέχεια έχουν επιτύχει υψηλότερα ποσοστά ακρίβειας μέσω της συλλογής στατιστικών πληροφοριών από μια σειρά παρόμοιων αλληλουχιών. Η PSI-BLAST χρησιμοποιώντας την BLAST δημιουργεί ένα πιθανοτικό προφίλ ενός ερωτήματος ακολουθίας και παράγοντας μια πιο ευαίσθητη ακολουθία σε αποτέλεσμα σύγκρισης. Παραγωγικά μοντέλα, όπως τα μοντέλα Markov (HMM-Hidden Markov Models) χρησιμοποιούν θετικά παραδείγματα μιας οικογένειας πρωτεϊνών. Τέλος, οι διακριτικοί αλγόριθμοι όπως ο Support Vector Machine (SVM) χρησιμοποιούν τόσο θετικά όσο και αρνητικά παραδείγματα και παράγουν εξαιρετικά αποτελέσματα με την χρησιμοποίηση κατάλληλου πυρήνα. Πολλοί SVM-μέθοδοι έχουν προταθεί, όπως SVM-ψαράς, SVM-k-φάσματος, Αναντιστοιχία-SVM, SVM-ζεύγη και άλλοι. Μια σύγκριση των SVM-με βάση τις μεθόδους έχει διενεργηθεί από τους Saigo et al.

Οι ομόλογες ακολουθίες αποτελούν μια σημαντική πηγή πληροφόρησης σχετικά με τις πρωτεΐνες. Πολλαπλές ευθυγραμμίσεις αλληλουχιών πρωτεϊνικών ακολουθιών περιέχουν σημαντικές πληροφορίες σχετικά με τις εξελικτικές διαδικασίες. Οι πληροφορίες αυτές μπορούν να ανιχνευθούν από την ανάλυση των αποτελεσμάτων της PSI-BLAST.(5)

Η σύγκριση ακολουθίας είναι ένα εκτεταμένο αντικείμενο μελέτης. Οι εφαρμογές είναι πολυάριθμες και περιλαμβάνουν σύγκριση αρχείων, διόρθωση ορθογραφίας, ανάκτηση πληροφορίας, και αναζήτηση ομοιοτήτων μεταξύ ακολουθιών βιολογικών δεδομένων.

Με δεδομένη μια συμβολοσειρά  $A = a_1a_2a_3\dots a_m$  και  $B = b_1b_2b_3\dots b_n$ , ζητείται η ευθυγράμμιση μεταξύ των δυο χορδών που εκθέτει την ομοιότητά τους. Τα αποτελέσματα ανατίθενται σε ευθυγραμμίσεις, σύμφωνα με την έννοια της ομοιότητας ή της διαφοράς που απαιτούνται από το πλαίσιο της εφαρμογής.

Ενώ για εφαρμογές όπως π.χ. η σύγκριση πρωτεϊνικών ακολουθιών οι μέθοδοι βαθμολόγησης μπορεί να περιλαμβάνουν αυθαίρετα αποτελέσματα για ζεύγη συμβόλων και για τα κενά μη ευθυγραμμισμένων συμβόλων, σε πολλές άλλες περιπτώσεις κόστους αρκεί. Δύο από αυτές, η μεγαλύτερη κοινή ακολουθία (LCS), καθώς και η επεξεργασία απόστασης, έχουν μελετηθεί εκτενώς στην επιστήμη των υπολογιστών. Κατά την επεξεργασία της απόστασης, κάθε ασυμφωνία ευθυγραμμισμένου ζεύγους και μη-ευθυγραμμισμένου σύμβολου καλείται διαφορά και δίνει αποτέλεσμα: 1. Όλα τα ζεύγη των ίσων ευθυγραμμισμένων χαρακτήρων δίνουν αποτέλεσμα: 0. Η επιδίωξη είναι ο εντοπισμός της ευθυγράμμισης που ελαχιστοποιεί το κόστος ή τον αριθμό των διαφορών και αυτό το ελάχιστο σκορ,  $ED(A, B)$ , καλείται η ελάχιστη απόσταση επεξεργασίας μεταξύ των  $A$  και  $B$ .(24)



Η μελέτη των εξελικτικών σχέσεων μεταξύ των οργανισμών είναι το αντικείμενο του ενδιαφέροντος για τους επιστήμονες για περισσότερο από έναν αιώνα. Οι πρώτες απόπειρες εύρεσης εξελικτικών συγγενειών στηρίχθηκαν αποκλειστικά σε διακριτά χαρακτηριστικά ειδών. Οι σύγχρονες μοριακές τεχνικές, ωστόσο, έχουν καταστήσει διαθέσιμη μια πληθώρα δεδομένων ακολουθιών του DNA, οι οποίες μπορούν να χρησιμοποιηθούν για τη μελέτη των σχέσεων αυτών. Σήμερα, είναι σύνηθες να εξετάζονται τα στοιχεία που περιέχονται και στα δύο είδη δεδομένων, ώστε να επιτυγχάνονται αυτοδύναμες εκτιμήσεις της εξελικτικής ιστορίας.

Αυτές οι εξελικτικές ιστορίες συνήθως εκπροσωπούνται από ένα φυλογενετικό δέντρο, το οποίο περιγράφεται μαθηματικά ως ένα άκυκλο συνδεδεμένο γράφημα  $(V, E)$ , όπου  $V$  είναι το σύνολο των κορυφών και  $E$  είναι το σύνολο των ακμών. Οι κορυφές συνδέονται μόνο με μια πλευρά και ονομάζονται κόμβοι τερματικών σταθμών, ενώ οι κορυφές συνδέονται με περισσότερο από μια πλευρές και ονομάζονται εσωτερικοί κόμβοι.

Η εκτίμηση των φυλογενετικών σχέσεων ανάμεσα σε μια συλλογή οργανισμών δεδομένου γενετικών δεδομένων για τους οργανισμούς αυτούς, μπορεί να χωριστούν σε δύο διαφορετικές κατηγορίες προβλημάτων. Η πρώτη κατηγορία καθορίζει το συγκεκριμένο κριτήριο με το οποίο θα συγκρίνουμε την καταλληλότητα μιας συγκεκριμένης φυλογενετικής υπόθεσης για τα παρατηρούμενα δεδομένα. Η δεύτερη είναι η αναζήτηση στο χώρο των πιθανών φυλογενέσεων για το συγκεκριμένο δέντρο ή δέντρα που παρέχουν την καλύτερη προσαρμογή στα δεδομένα.

Όπως προαναφέρθηκε, ένα φυλογενετικό δέντρο μπορεί να θεωρηθεί ως μια γραφική παράσταση για την οποία οι τερματικοί κόμβοι αντιπροσωπεύουν οργανισμούς για τους οποίους δεδομένα έχουν παρατηρηθεί, που ονομάζονται ταξινομικές μονάδες ή *taxa*, ενώ οι εσωτερικοί κόμβοι αντιπροσωπεύουν υποθετικούς προγονικούς οργανισμούς. Οι ακμές που συνδέουν τους κόμβους αναφέρονται γενικά ως *παρακλάδια* και χαρακτηρίζουν τις σχέσεις μεταξύ προγόνων-απογόνων. Έρριζα φυλογενετικά δέντρα μπορεί και να μην ικανοποιούν την υπόθεση του μοριακού ρολογιού. Η υπόθεση του μοριακού ρολογιού είναι ότι ο ρυθμός της εξέλιξης είναι περίπου σταθερός στο χρόνο. (10)

Σύγκριση πρωτεϊνών χρησιμοποιώντας MATLAB (30)

Σχήμα 17: Παραδείγμα αντιστοίχισης χρησιμοποιώντας τον αλγόριθμο NW. Στο παρακάτω παράδειγμα γίνεται αντιστοίχιση χρησιμοποιώντας τις πρωτεΐνες που προκύπτουν από την έφραση του γενετικού υλικού.

```

Identities = 445/529 (84%), Positives = 501/529 (95%)
1  MTSRLLWFSLLLAAGFAGRATALWPWPQNFTSDQRYVLYPNHFQFDVSSAACPGCSVLDEA
  :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1  MAGCRLLVSLLLAAALACLATALWPWPQYIQTVHRRYTLYPNNHFQFRHVSSAACAGCVVLEA

65  FQRYRDLDFGSGWPRPYLTGKRHTLERNVLVSVVTFGCNQLP TLESVENYTLTINDDQCCLL
  :||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
65  FRYRNLDFGSGWPRPFSFNKCOOTLGENILVUSVVTAECPNPLESVENYTLTINDDQCLLA

129  SETVVGALRGLETFSQLVWKSAGETFFINKTEIEDPFRPHRGULLDTSRHYLPLESSILDLDV
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
129  SETVVGALRGLETFSQLVWKSAGETFFINKTKIKDFPFRPHRGVLLDTSRHYLPLESSILDLDV

193  MAYNKLNVFHHLLVDDSPFPYESFTPELMRKGSYNPVTHIYTAQDVKEVIEYAKLRGIRVLAE
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
193  MAYNKFNVHHLLVDDSPFPYESFTPELTRKGSFNPVTHIYTAQDVKEVIEYAKLRGIEVLAE

257  FDTPGHTLSUGPGIPGLLTPCYSGSEPSGTFGPVNPSSLNNTYEFMSTFFLEVSVVFPDFYLHLG
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
257  FDTPGHTLSUGPCAPGLLTPCYSGSLSGTFGPVNPSSLNNTYDFMSTLFLFISSVFPDFYLHLG

321  GDEVDFTCWKSNEIQDFHRKKGFGZDFKQLESFYIQTLLDIVSYYGKGYVWQEVFQVFNKVKIQ
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
321  GDEVDFTCWKSNPNIQAFWKKGF-TDFKQLESFYIQTLLDIVSDYDKCVVWQEVFQVFNKVKVUP

385  PDTTIQWREDIPVNYNKELELVTRAGFRALLSAPWYLNRISTGPDWEDFYVVEPLAFEGTPEQ
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
384  PDTTIQWREEMRPVYMLENQDITRAGFRALLSAPWYLNRYKTPGPDWEDMYKVEPLAFHGTFEQ

449  KALVIGGEACHWGEYVDNTNLVPRLWPRAGAVAERLWSNKLTSDLTFAYERLSSHFRCELLRRGV
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
448  KALVIGGEACHWGEYVDS-TNLVPRLWPRAGAVAERLUSNLTNIDFAFERLSHFRCELVRGCI

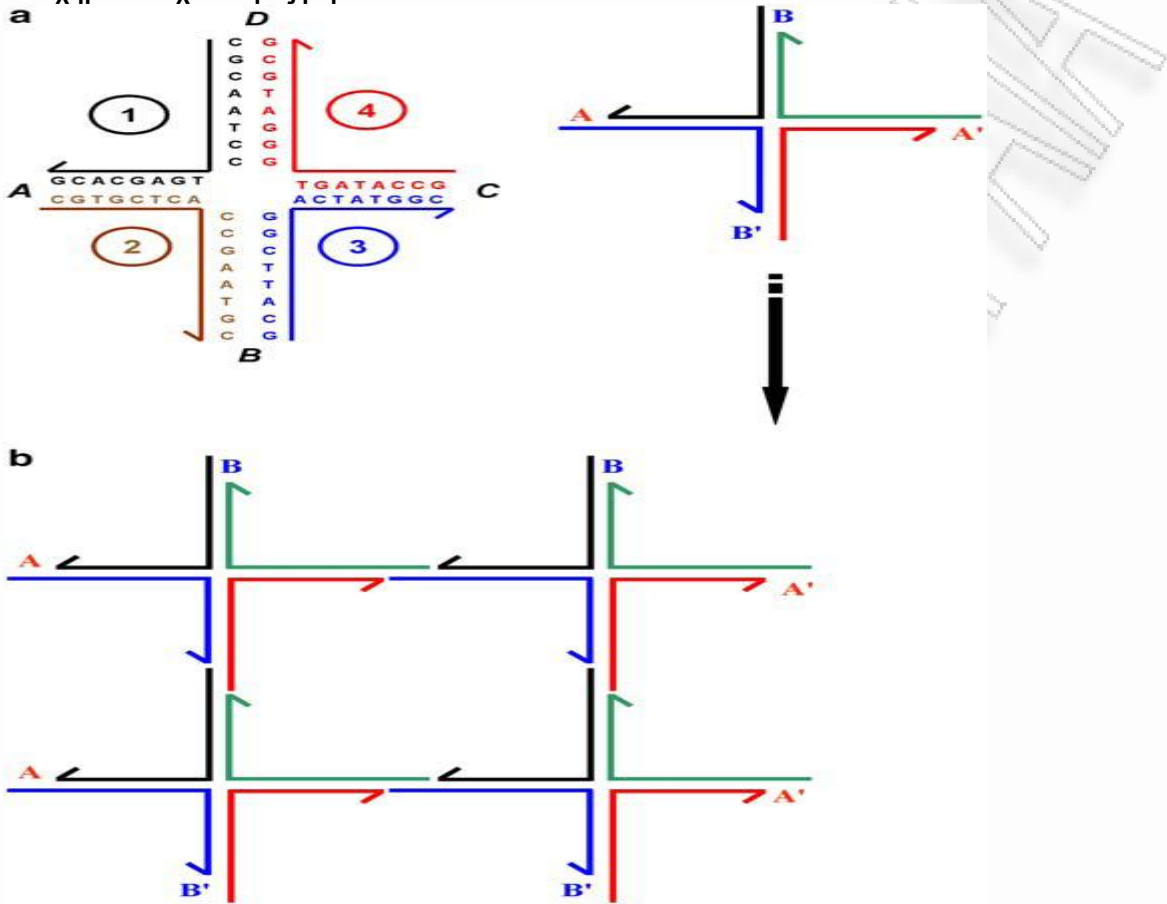
513  QAOPLNVGFCEQEFEQI
  ||||| ||| ||||| |||
512  QAOPIVGCCEQEFEQI
    
```

Σχήμα 18: Παραδείγμα αντιστοίχισης χρησιμοποιώντας τον αλγόριθμο Smith and Waterman. Στο παρακάτω παράδειγμα γίνεται αντιστοίχιση χρησιμοποιώντας τις πρωτεΐνες που προκύπτουν από την έκφραση του γενετικού υλικού.

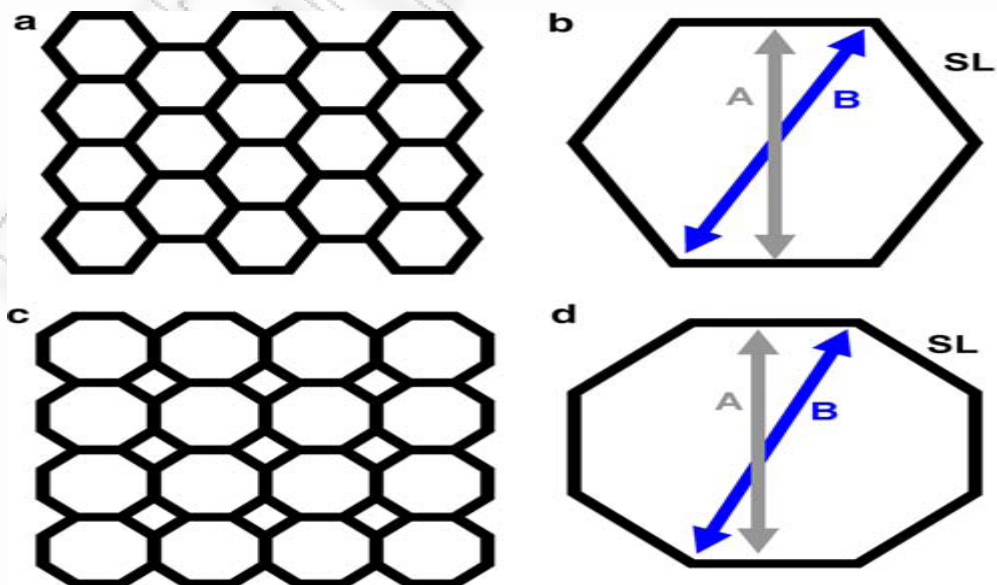
```
Identities = 454/547 (83%), Positives = 514/547 (94%)
1  RGDQR-AMTSSRLWFSLLLAAAFAGRATALWPPWQNFQTSQORYLVLPNNFQYQDVSAAQCPG
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1  RGAGRWAMAGCRLWVSLLLAALACLATALWPPWQYIQTYHRRYTLYPNNFQYRVSSAAQAG
64  C5VLDEAFQRYRDLDFSGSWPRFLYTGKRHLTKHNLVWVSVTGCMQLPFLSEVENYTLIN
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
65  CVLDEAFRYYRNLLFSGSWPRPFYFVSNKQQTLGKNIWVSVVTAECNEFFNLESEVENYTLIN
128  DPQCLLSETVWGALRGLFTSQLVWKSAGETFFINKTIEZDFPRFRHGLLSDTSHRYLPSS
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
129  DDQCLLASETVWGALRGLFTSQLVWKSAGETFFINKTIEZDFPRFRHGLLSDTSHRYLPSS
192  ILDTLDVMAYNKLWVFWHLVDDPSFPYESFTFPELMKKSYPNVTTHIYTAQDVKEVIEYARLR
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
193  ILDTLDVMAYNKFVWFHHLVDDPSFPYESFTFPELMKKSYPNVTTHIYTAQDVKEVIEYARLR
256  GIRVLAEFDTPGHTLSWGPGLGILLTPCYSGSEFSGTFGVNPSLNHTYEFHSTFLEVSSVFP
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
257  GIRVLAEFDTPGHTLSWGPGLGILLTPCYSGSEHSGTFGVNPSLNHTYEFHSTFLEVSSVFP
320  DFLHLGGEVDFTCWKSNEIQDFHKKGFQZDFKQLEZFYIQTLLDIVSYGKGYVWVQEVF
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
321  DFLHLGGEVDFTCWKSNNIQAFFMKKGF-TDFKQLESFYIQTLLDIVSDYDKGVVWVQEVF
384  DNKVKIQPDTIIQVUREDIPVNYHKELELVTHAGFRALLSAPVYLNRESYGPDWKDFYVVEFLA
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
384  DNKVKVVRDPTIIQVUREENPVEYHLEKQDITPACFRALLSAPVYLNRESYGPDWKDYKVEPLA
448  FECTPEQKALVICGEACHUGEYVDNLTNLPRLWPRACAVAERLUSNLTSDLTFAVERLSHFRC
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
448  FHCPTPEQKALVICGEACHUGEYVDSTNLPRLWPRACAVAERLUSNLTNIDFAFRKRLSHFRC
512  ELLRRGVQAOPLNVCFCQEFEQTS*APGTEECAGC
   ||| | | | | | | | | | | | | | | | | | | | | | | | |
512  ELVRRGIAQAPIVSGCCQEFEQTS*ATSAEHPGCG
```

**Μελλοντική έρευνα για σχεδιασμό μορίων.**

Σχήμα 19: Σχεδιασμός μορίων DNA.



Σχήμα 20: Εξαγωνικά και οκταγωνικά δίκτυα κατασκευασμένα με την βοήθεια της βιοτεχνολογίας.





Νέοι μέθοδοι για τον σχεδιασμό και την δημιουργία βιολογικών νανοφίλτρων έχει προταθεί. Οι υπολογισμοί δείχνουν ότι θα είναι εφικτό να χρησιμοποιηθούν για τον διαχωρισμό διαφόρων μορίων, ακόμη και ιών και βακτηρίων. Θα μπορούσαν να συμβάλουν όχι μόνο στην αποτροπή εξάπλωσης παθογόνων μικροοργανισμών, αλλά και στην προστασία του περιβάλλοντος, καθώς αυτά τα βιολογικά νανοφίλτρα θα μπορούσαν να χρησιμοποιηθούν και για τον διαχωρισμό νερού-πετρελαίου. (25)

Μία ομάδα ερευνητών από το Πανεπιστήμιο του Hong Kong στην Κίνα έχουν επινοήσει έναν τρόπο για την αποθήκευση δεδομένων σε ζωντανά βακτήρια. Ο μηχανισμός κωδικοποίησης είναι αρκετά απλός, καθώς μετατρέπουν τα ηλεκτρονικά δεδομένα (π.χ. κείμενο ASCII) σε σύστημα τετραψηφίων, το οποίο στη συνέχεια τους επιτρέπει να χαρτογραφήσουν τα δεδομένα στα 4 βασικά στοιχεία που αποτελούν το DNA: A, G, T και C. Μόλις τα δεδομένα μετατραπούν σε μια πλήρη αλληλουχία DNA, τότε συμπιέζονται (με τη χρήση του αλγόριθμου DEFLATE), πριν «καταγραφούν» σε ένα ζωντανό κύτταρο.

Η ομάδα ισχυρίζεται ότι μπορεί να κάνει κρυπτογράφηση και αποκρυπτογράφηση (Bioencryption), ωστόσο όμως από την παρουσίασή τους βλέπουμε ότι μπορούν μόνο να κωδικοποιήσουν και να αποκωδικοποιήσουν τα δεδομένα από δυαδικά σε DNA. Επιπλέον ισχυρίζονται ότι μπορούν να αποθηκεύσουν 900TB (terabytes!) σε ένα μόνο γραμμάριο κυττάρων του βακτηριδίου E.coli, το οποίο είναι αρκετά εντυπωσιακό. (26), (27).

## BLAST

Πριν ξεκινήσουμε με το BLAST πρέπει να έχουμε επιλέξει τις ακολουθίες που θέλουμε να συγκρίνουμε και οι οποίες θα μας απασχολήσουν.

Για παράδειγμα:

The screenshot shows the NCBI Nucleotide database interface. The search bar contains the accession number 'L06042.1'. The page title is 'Simian immunodeficiency virus complete genome'. The main content area displays the following information:

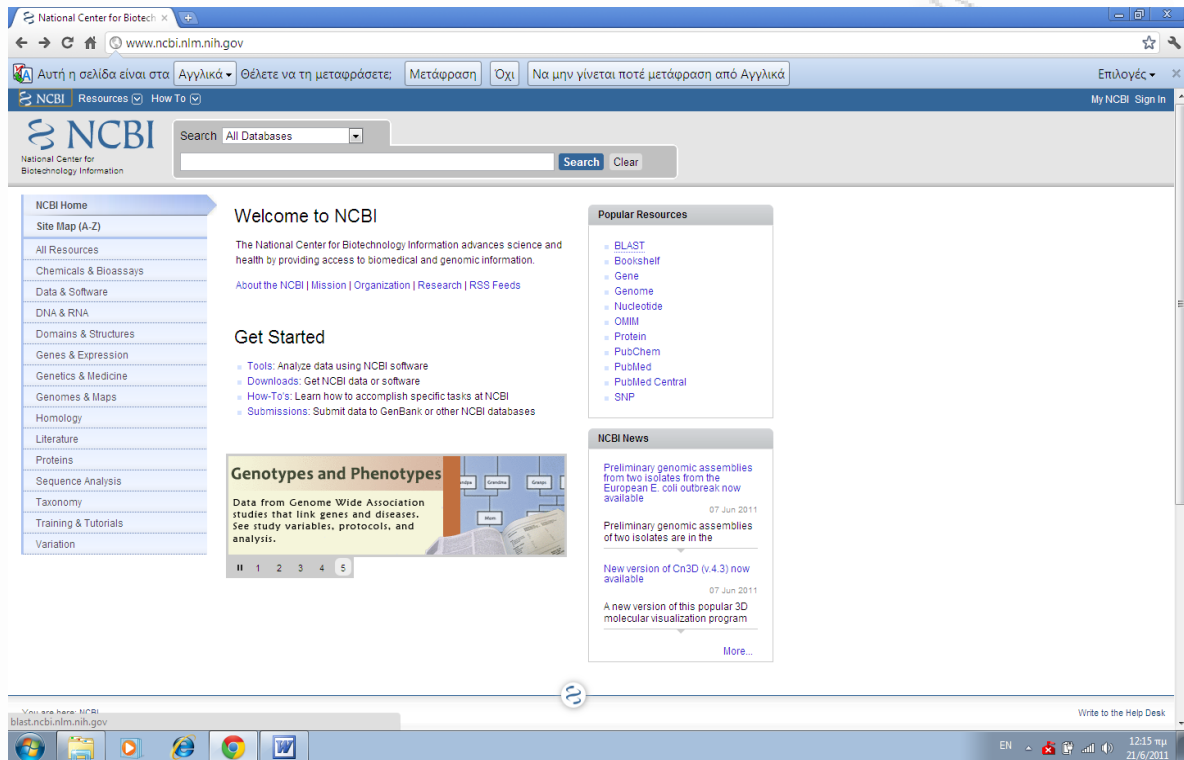
- LOCUS:** SIVCOMGNM 9597 bp RNA linear VRL 17-JUL-2000
- DEFINITION:** Simian immunodeficiency virus complete genome.
- ACCESSION:** L06042
- VERSION:** L06042.1 GI:294960
- KEYWORDS:** -
- SOURCE:** Simian immunodeficiency virus
- ORGANISM:** [Simian immunodeficiency virus](#)
- REFERENCE:** Hirsch, V.M., Dapolito, G.A., Goldstein, S., McClure, H., Emau, P., Fultz, P.N., Isehakia, M., Lenroot, R., Myers, G. and Johnson, P.R. A distinct African lentivirus from Sykes' monkeys *J. Virol.* 67 (3), 1517-1528 (1993)
- FEATURES:**
  - source** Location/Qualifiers
    - 1..9597
      - /organism="Simian immunodeficiency virus"
      - /proviral
      - /mol\_type="genomic RNA"
      - /isolate="syk173/1.2"
      - /db\_xref="taxon:11723"
      - /cell\_line="CEMx174"
    - 1..602
      - [LTR](#)
      - [gene](#)
        - 795..2459
          - /gene="gag"
          - /note="gag protein"
          - /odon\_start=1
          - /product="gag protein"
          - /protein\_id="AB01716.1"

The right sidebar contains several utility sections: 'Change region shown', 'Customize view', 'Analyze this sequence' (with options for Run BLAST, Pick Primers, Find in this Sequence), 'Retrovirus Resource', 'LinkOut to external resources' (LANL HIV Databases), 'All links from this record' (Full text in PMC, Protein, PubMed, RefSeq Genome for Species), and 'Recent activity' (Simian immunodeficiency virus complete genome).



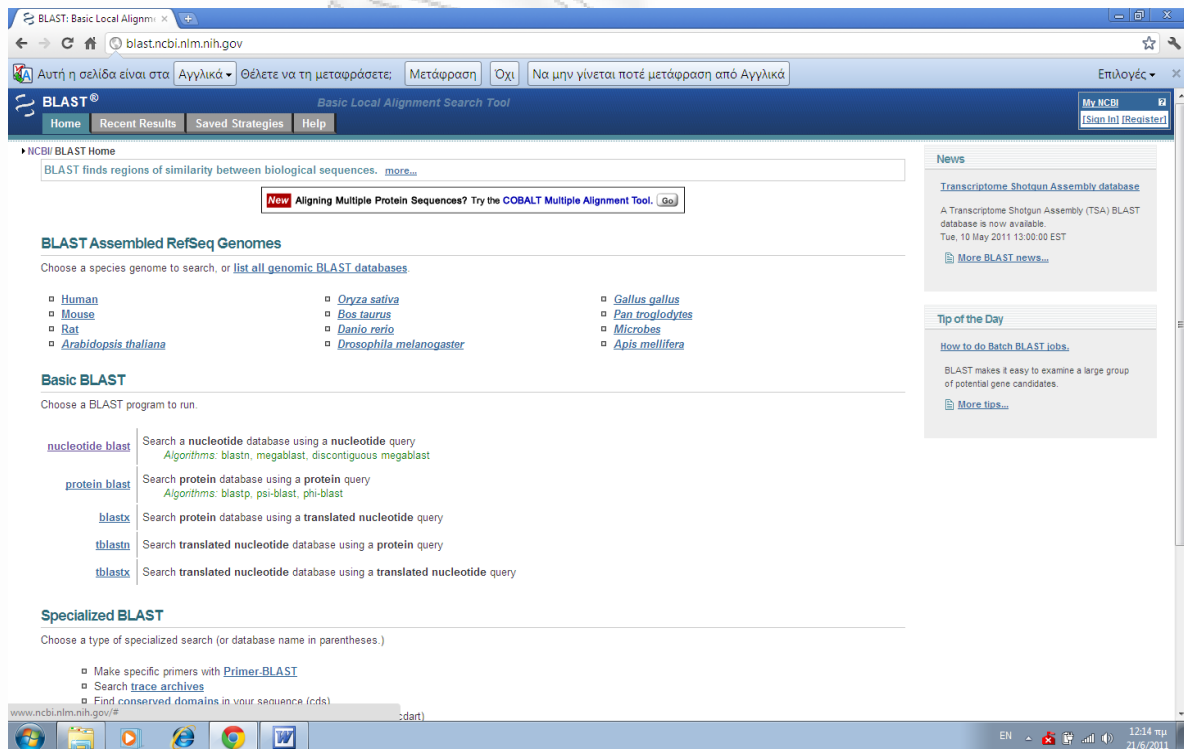
Για πρόσβαση στο blast θα πρέπει να εισέλθουμε στην σελίδα του ncbi που είναι η εξής: <http://www.ncbi.nlm.nih.gov/>

Από την ίδια σελίδα έχουμε εντοπίσει και τις ακολουθίες που μας ενδιαφέρουν.

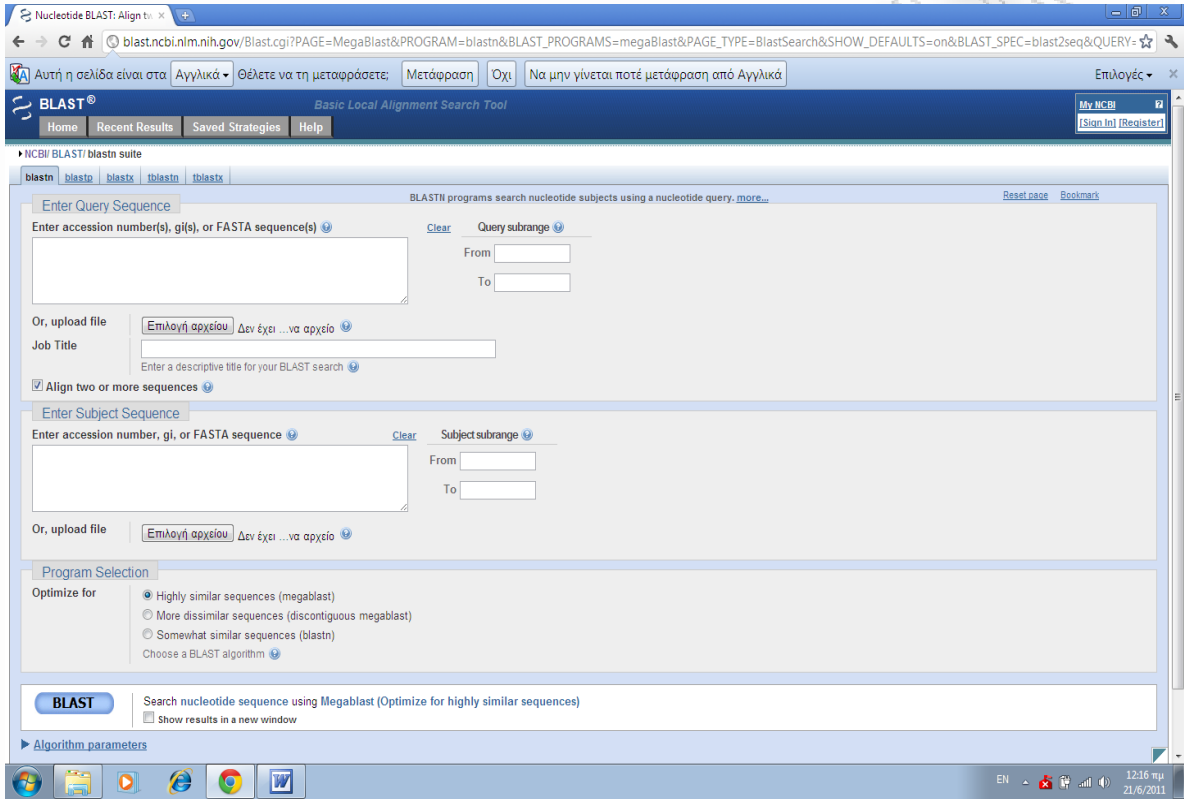


Στην συνέχεια θα επιλέξουμε από την δεξιά στήλη την επιλογή BLAST.

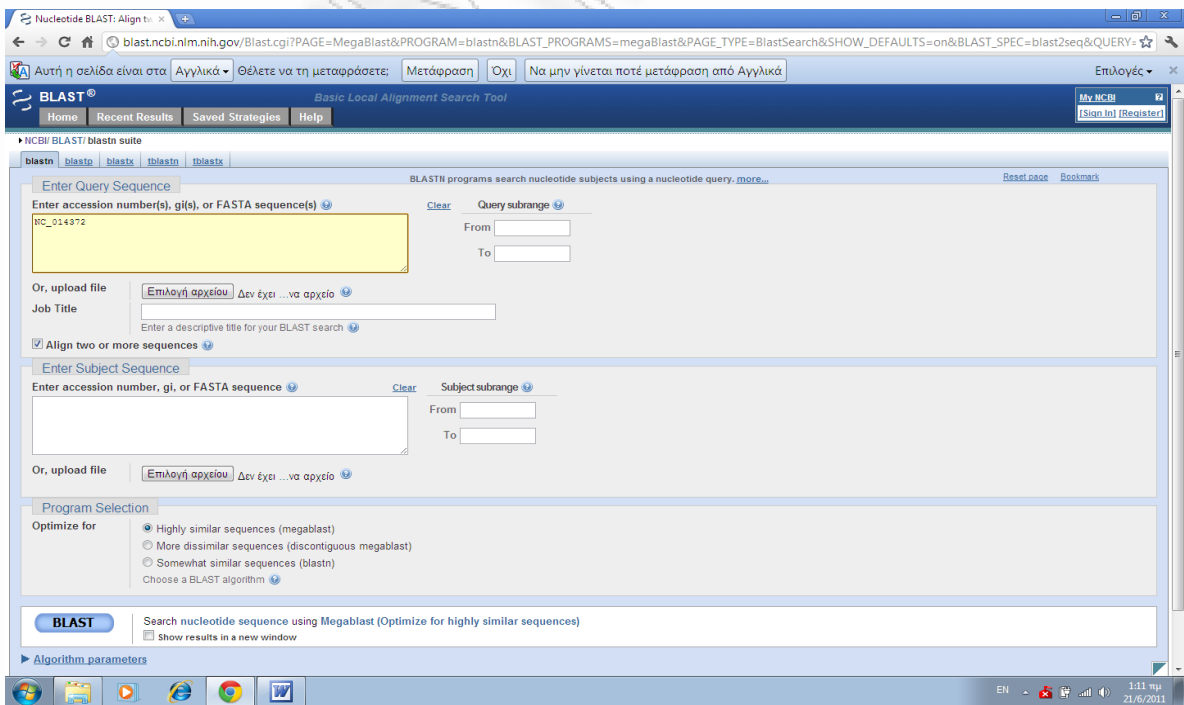
Από τις παρακάτω επιλογές θα επιλέξουμε την επιλογή στα δεξιά που αναγράφει nucleotide blast



Επιλέγοντας την επιλογή: Align two or more sequences μπορούμε δίνοντας τους κωδικούς για κάθε συμβολοσειρά που θέλουμε να συγκρίνουμε να γίνει η σύγκριση.

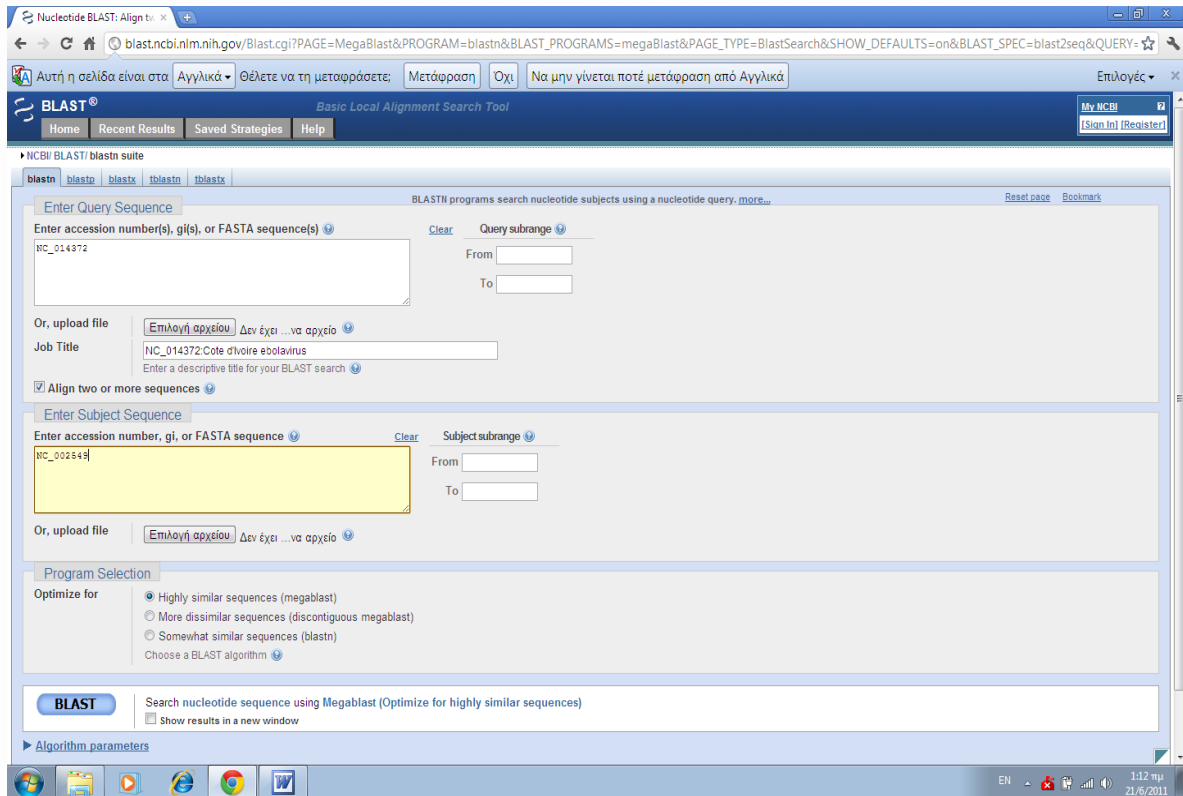


Πρώτα εισάγουμε την πρώτη ακολουθία που θέλουμε να συγκρίνουμε:

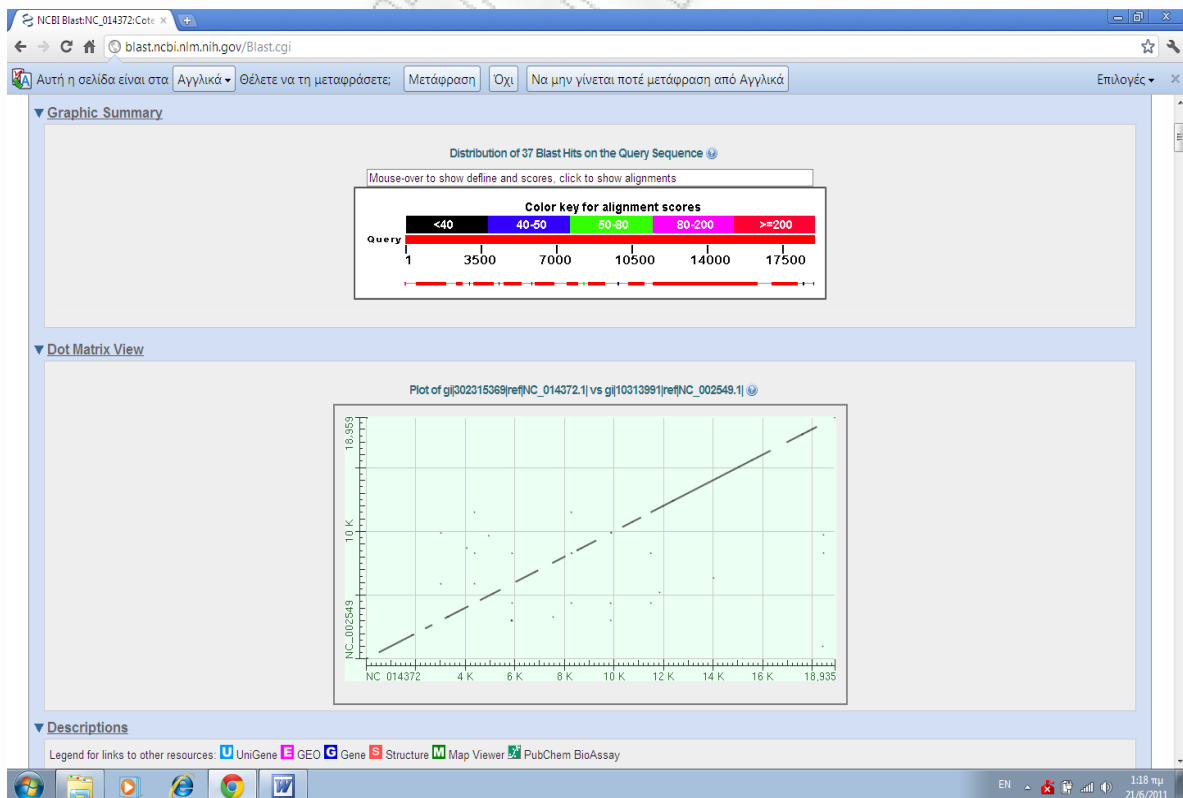




Και στην συνέχεια τη δεύτερη:

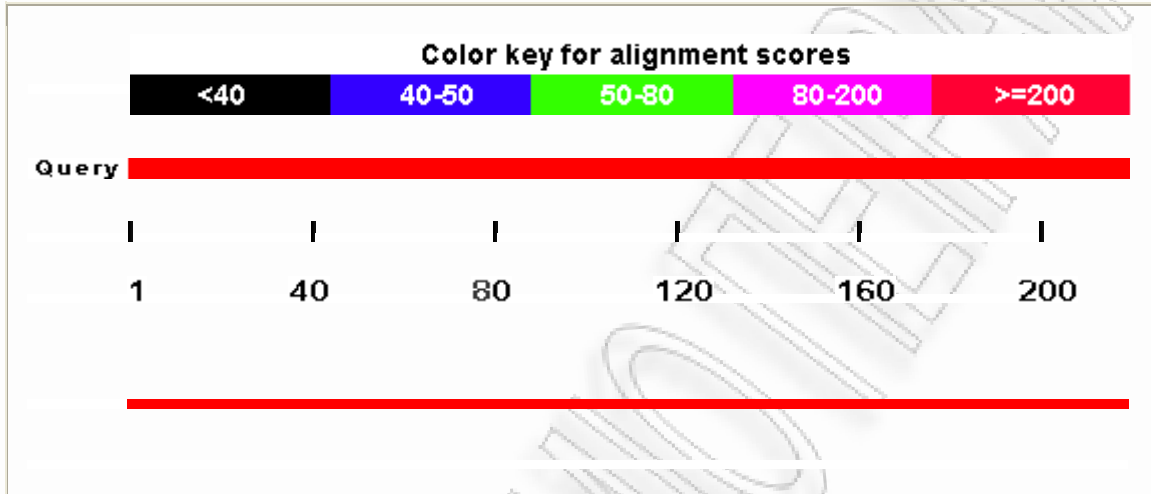


Μπορούμε να επιλέξουμε blastn για παράδειγμα ανάλογα με το πόσο όμοιες είναι οι συμβολοσειρές και έχουμε:



Ένα παράδειγμα χρησιμοποίησης του BLAST φαίνεται παρακάτω:

Το αποτέλεσμα για κάθε ευθυγράμμιση υποδεικνύεται από ένα από τα πέντε διαφορετικά χρώματα, τα οποία χωρίζουν το φάσμα των βαθμολογιών σε πέντε ομάδες. Στο παρακάτω παράδειγμα έχει γίνει σύγκριση μεταξύ δύο στελεχών του ιού της γρίπης που έχει εμφανιστεί σε δυο διαφορετικές περιοχές. Η σύγκριση αφορά στέλεχος του ίδιου ιού που έχει εμφανιστεί στο Αμβούργο της Γερμανίας και στην Βαρκελόνη της Ισπανίας. Η ομοιότητα είναι εμφανής από την κόκκινη γραμμή που υποδηλώνει την μέγιστη δυνατή ομοιότητα.



**Σύγκριση στελέχους Ζαίρ με στέλεχος Ακτής Ελεφαντοστού**

Στο παρακάτω παράδειγμα θα γίνει σύγκριση με το Blast μεταξύ του Ιού Έμπολα Ζαίρ και Ακτής Ελεφαντοστού: Cote d'Ivoire ebolavirus, RNA; linear; Length: 18,935 nt Replicon Type: viral genome, Created: 2010/08/05, [NC\\_014372](#) με Zaire ebolavirus, complete genome, RNA; linear; Length: 18,959 nt Replicon Type: viral genome, Created: 1999/02/10, [NC\\_002549](#)

Score = 1110 bits (1230), Expect = 0.0  
 Identities = 1104/1426 (77%), Gaps = 4/1426 (0%)  
 Strand=Plus/Plus

```

Query  527  GATTACCACAAGATCTTGACAGCAGGTCTGTCCGTTCAACAGGGATTGTTTCGGCAAAGA  586
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Sbjct  521  GATTACCATAAGATTTTAACAGCAGGATTGTCAGTCCAACAAGGCATTGTGAGACAACGG  580

Query  587  GTCATCCAGTGTATCAAGTAAACAATCTTGAAGAAATTTGCCAACTTATCATACAGGCC  646
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Sbjct  581  GTCATTCAAGTCCACCAGGTTACAAACCTAGAAGAAATATGCCAATTGATCATTCAAGCC  640

Query  647  TTTGAAGCAGGTGTTGATTTTCAAGAGAGTGC GGACAGTTTCCTTCTCATGCTTTGTCTT  706
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Sbjct  641  TTTGAAGCTGGTGTGATTTTCAAGAGAGTGCAGACAGTTTCTTGCTGATGCTATGTTTA  700

Query  707  CATCATGCGTACCAGGGAGATTACAACTTTTCTTGAAAGTGGCGCAGTCAAGTATTTG  766
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Sbjct  701  CATCATGCTTATCAGGGTACTACAAGCAATCTTGAAAGCAATGCAGTCAAGTACCTT  760
    
```

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΑΘΑΝΑΣΙΟΣ ΠΑΥΛΙΔΗΣ

```

Query 767 GAAGGGCACGGGTTCCGTTTTGAAGTCAAGAAGCGTGATGGAGTGAAGCGCCTTGAGGAA 826
          || || || || || || || || || || || || || || || || || || || || ||
Sbjct 761 GAGGGTCATGGCTTTCGCTTTGAGGTCAAGAAAAGGAAGGAGTCAAGCGACTCGAAGAA 820

Query 827 TTGCTGCCAGCAGTATCTAGTGGAAAAACATTAAGAGAACAACCTTGCTGCCATGCCGGAA 886
          || || || || || || || || || || || || || || || || || || || || ||
Sbjct 821 TTGCTTCCTGCTGCATCCAGTGGCAAGAGCATCAGGAGAACACTGGCTGCAATGCCTGAA 880

Query 887 GAGGAGACAACTGAAGCTAATGCCGGTCAGTTTCTCTCCTTTGCAAGTCTATTCTTCCG 946
          || || || || || || || || || || || || || || || || || || || || ||
Sbjct 881 GAGGAGACAACAGAAGCAAATGCCGGACAGTTCTCTCTTTGCTAGCTTATTCTTCTCT 940

Query 947 AAATTGGTAGTAGGAGAAAAGGCTTGCCTTGAGAAGGTTCAAAGGCAAATTC AAGTACAT
1006
          || | || || || || || || || || || || || || || || || || || || || ||
Sbjct 941 AAGCTAGTTGTCGGAGAAAAGCCTGTCTAGAAAAGGTCAGCGCAAATTC AAGTTCAT
1000

Query 1007 GCAGAGCAAGGACTGATACAATATCCAACAGCTTGGCAATCAGTAGGACACATGATGGT
1066
          | || || || || || || || || || || || || || || || || || || || || ||
Sbjct 1001 TCTGAGCAGGGATTGATCCAATACCCACAGCCTGGCAGTCAGTTGGACACATGATGGT
1060

Query 1067 ATTTTCCGTTTGATGCGAACAAATTTTCTGATCAAATTTCTCCTAATACACCAAGGGAT
1126
          || || || | || || || || || || || || || || || || || || || || || ||
Sbjct 1061 ATTTTCAGACTGATGAGAACAAATTTTCTAATTAAGTTCCTCCTTATACATCAAGGGAT
1120

Query 1127 CACATGGTTGCCGGGCATGATGCCAACGATGCTGTGATTTCAAATTCAGTGGCTCAAGCT
1186
          || || || || || || || || || || || || || || || || || || || || || ||
Sbjct 1121 CATATGGTAGCAGGACACGATGCTAACGATGCTGTCATCGCAAACCTCTGTAGCTCAAGCA
1180

Query 1187 CGTTTTTCAGGCTTATTGATTGTCAAACAGTACTTGATCATATCCTACAAAAGACAGAA
1246
          || || || || || || || || || || || || || || || || || || || || || ||
Sbjct 1181 CGTTTTTCAGGATTATTGATCGTTAAACAGTGCTAGATCACATCCTTCAGAAAACAGAG
1240

Query 1247 CGAGGAGTTCGTCTCCATCCTCTTGCAAGGACCGCCAAGGTAAAAAATGAGGTGAACTCC
1306
          | || || || || || || || || || || || || || || || || || || || || ||
Sbjct 1241 CACGGAGTTCGTCTTCATCCTTTGGCAAGAACTGCTAAGGTCAAGAACGAAGTAAATTC
1300

Query 1307 TTTAAGGCTGCACTCAGCTCCCTGGCCAAGCATGGAGAGTATGCTCCTTTCGCCCGACTT
1366
          || || || || || || || || || || || || || || || || || || || || || ||
Sbjct 1301 TTTAAGGCTGCCCTTAGCTCGCTAGCACAACATGGAGAGTATGCTCCTTTTGCTCGCTTG
1360
    
```

```

Query 1367 TTGAACSTTTCTGGAGTAAATAATCTTGAGCATGGTCTTTTCCCTCAACTATCGGCAATT
1426
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Sbjct 1361 CTGAATCTTTCTGGAGTCAACAATCTCGAGCACGGACTGTTTCTCAGCTTCTGCAATT
1420

Query 1427 GCACTCGGAGTCGCCACAGCACACGGGAGTACCCTCGCAGGAGTAAATGTTGGAGAACAG
1486
    || || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Sbjct 1421 GCCCTAGGTGTCGCAACGGCACACGGCAGTACCCTGGCAGGAGTAAATGTGGGGGAACAG
1480

Query 1487 TATCAACAACACTCAGAGAGGCTGCCACTGAGGCTGAGAAGCAACTCCAACAATATGCAGAG
1546
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Sbjct 1481 TATCAGCAACTACGAGAAGCAGCCACTGAGGCAGAAAAACAATTGCAGAAATACGCTGAA
1540

Query 1547 TCTCGGAACTTGACCATCTTGACTTGATGATCAGGAAAAGAAAATCTTATGAACTTC
1606
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Sbjct 1541 TCTCGCGAGCTTGACCATCTAGGTCTCGATGATCAAGAGAAGAAGATCTGAAAGACTTC
1600

Query 1607 CATCAGAAAAAGAACGAAATCAGCTTCCAGCAAACAAACGCTATGGTAACTCTAAGAAAA
1666
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Sbjct 1601 CATCAGAAGAAAAATGAAATCAGCTTCCAGCAGACAACAGCCATGGTCACACTACGGAAG
1660

Query 1667 GAGCGCTGGCCAAGCTGACAGAAGCTATCACTGCTGCGTCACTGCCCAAACAAGTGA
1726
    || || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Sbjct 1661 GAAAGGCTAGCCAAGCTCACTGAGGCAATCACCTCCACATCCCTTCTCAAGACAGGAAAA
1720

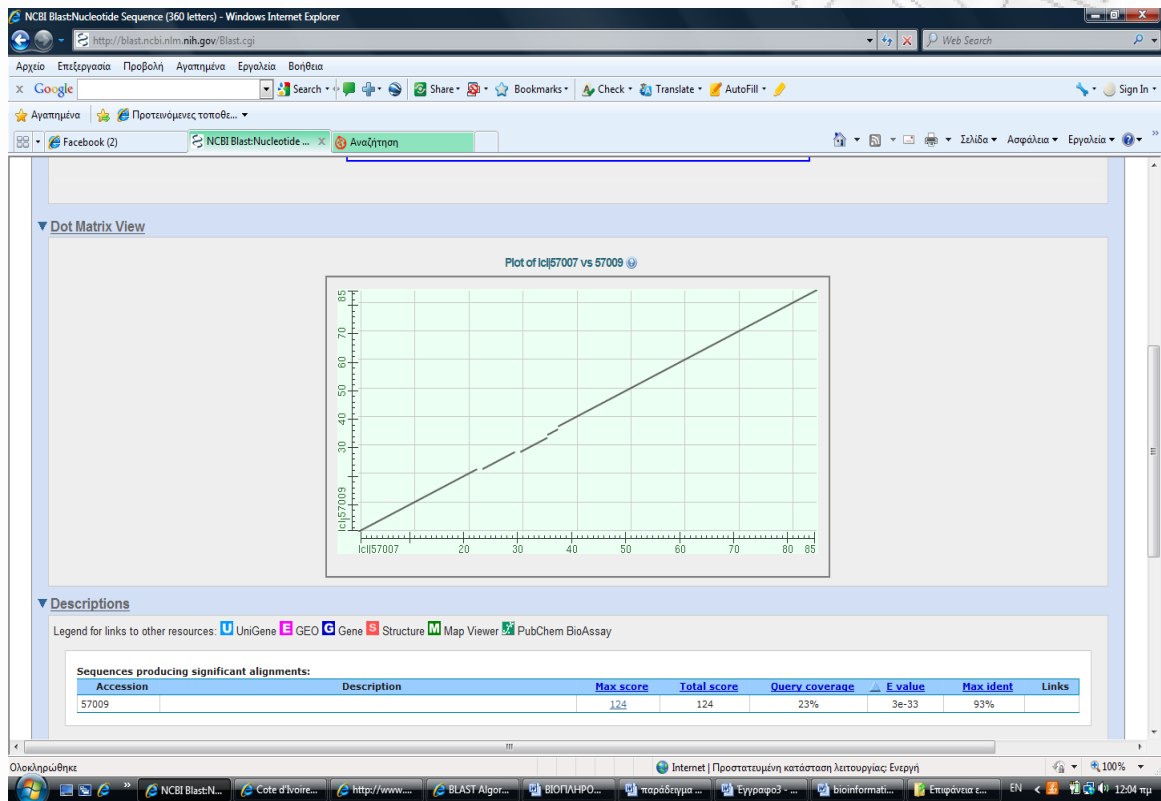
Query 1727 CATTACGATGATGATGACGACATTCCTTTCCAGGACCCATCAATGATGACGACAATCCT
1786
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Sbjct 1721 CAGTATGATGATGACAACGATATCCCCTTCTGCGCCATCAATGATAACGAAAACCTCA
1780

Query 1787 GGCCATCAAGATGATGATCCGACTGACTCACAGGATACGACCATTCCCGATGTGGTGGTT
1846
    | || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Sbjct 1781 GAACAGCAAGACGATGATCCAACAGATTCTCAGGACACTACCATCCCTGATATCATGTGT
1840

Query 1847 GATCCCGATGATGGAAGCTACGGCGAATA-CCAGAGT-TACTCGGAAAACGGCATGAATG
1904
    || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Sbjct 1841 GACCCGATGATGGCAGATACAACAATATGGGAGACTATCCTAGTGAGACGGC--GAATG
1898
    
```

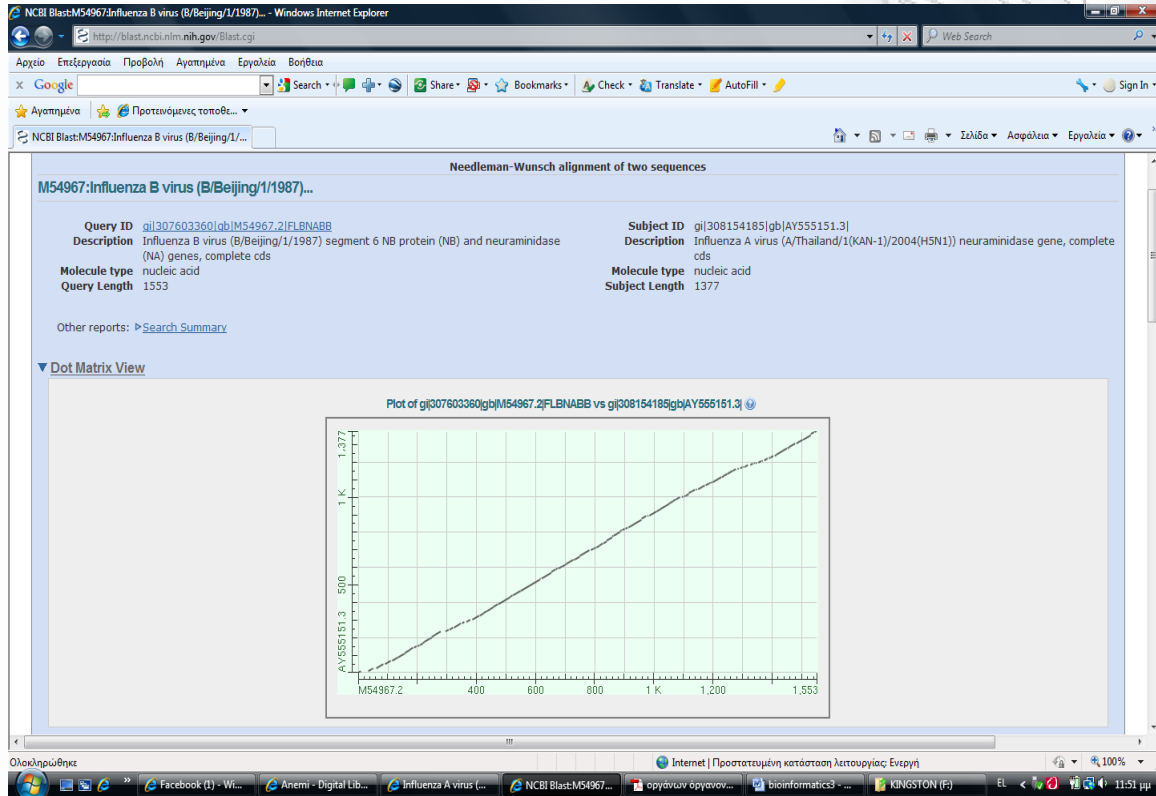
```

Query  1905  CACCAGATGACTTGGTCCTATTCGATCTAGACGAGGACGACGAGGA 1950
      | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  1899  CCCCTGAAGACCTTGTTCTTTTGGACCTTGAAGATGGTGACGAGGA 1944
  
```



**Σύγκριση στέλεχους Πεκίνου με στέλεχος Ταϊλάνδης.**

Στο συγκεκριμένο παράδειγμα θα γίνει σύγκριση των στελεχών του Πεκίνου και της Ταϊλάνδης του Ιού γρίπης τύπου Β. (Influenza B virus (B/Beijing/1/1987) segment 6 NB protein (NB) and neuraminidase (NA) genes, **complete** cds GenBank: M54967.2 with Influenza A virus (A/Thailand/1(KAN-1)/2004(H5N1)) neuraminidase gene, **complete** cds GenBank: AY555151.3).



Identities = 830/1636 (51%), Gaps = 342/1636 (20%)

Strand=Plus/Plus

```

Query 1      AGCAGAAGCNNNNNNNNNNNNNNNNNNNNNNNNAACAGGCCAAAAATGAACAATGCTAC 60
            ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct 1      AGCAAAAGC-----AGGAGTTCAAAA-TGAA----- 25

Query 61     CTTCAACTATACAAACGTTAACCCATTCTCACATCAGGGGAGTGTTATTATCACTAT 120
            | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct 26     -TCCAAATAAGAAGATAATAACCAT-----CGGATCAATCTGTATGGTA-----ACTGG 73

Query 121    ATGTGTCAGCTTCACTGTCATACTTACTGTATTCGGAT----ATATTGCTAAAAATTTCA 176
            | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct 74     AATGGTTAGCTTAA-TGTTACAAATTGGGAAGCTG-ATCTCAATATGGGTCAGTCATTC 131

Query 177    --TCAAAAAATAACTGCACCAACAATGACATTGGATTGCGCGAAC--GCATCA--AATGTT 230
            | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct 132    ATTACACAGG---GAATCAACA---CAAAGC--TGAACCAATCAGCAATACTAATTTT 182

Query 231    CAGGCTGTGAACCGCTCTG----CAACAAAAG AGATGACATTTCTTCTCCAGAACCGG 285
            | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct 183    CTTACTGAGAAA-GCTGTGGCTTCAGTAAAATTAGCGGGCAATTCATCTCTTTG----- 235

Query 286    AGTGGACATACCCTCGTTTATCTTGCCAGGGCTCAACCTTTCAGAAAGCACTCCTA-ATT 344
            | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct 236    -----CCC-CATTAAT---GGATGGGCTGTA---TACAGTAAGGACAACAGTATA 278
    
```

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΑΘΑΝΑΣΙΟΣ ΠΑΥΛΙΔΗΣ

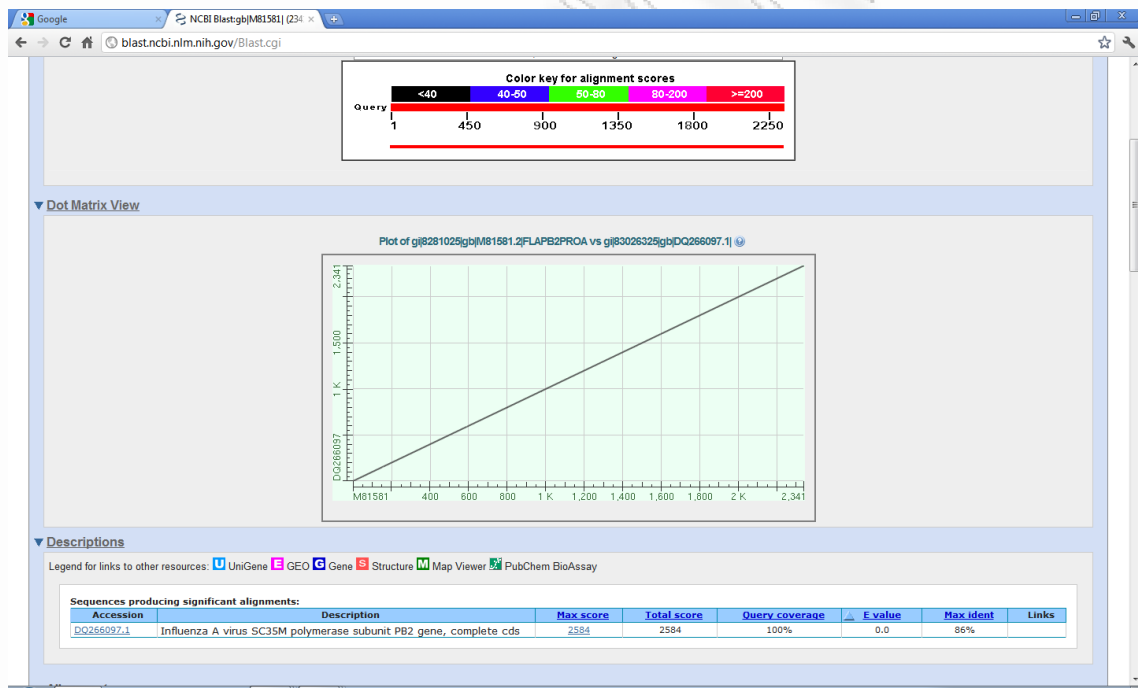
Query	345	AGCCCTCATAGATTTCGGAGAAGCCAGAGGAAACTCAGCTCCCTTAATAATAAGGGAACCT	404
Sbjct	279	AGG----ATCGGTTC-----CAAGGGGATGTG-----TTTGTATAAGAGAGCCA	320
Query	405	TTTATTGCTTGTGGACAAA--GGAATGCAAACACTTTGCTCTA-ACCCATTATGCA--G	459
Sbjct	321	TTCAT--CTCATGCTCCCACTTGGAAATGCAGA-ACTTCTTTTGGACTCAGGGAGCCTTG	377
Query	460	CTCAACCGGGAGGGTACTACAATGGAACAAGAGAGGACAGAAACAAGCTGAGG-CATC--	516
Sbjct	378	CTGAATGACAAGC--ACTCCAATGGGACTGTCAAAGACAGAAGCCCTCACAGAACATTAA	435
Query	517	TGATTTTCAGTCAAATTAGGCAAAATCCCAACTGTAGAAAACCTCCATTTTCCACATGGCAG	576
Sbjct	436	TGAGTT--GCCTGTGGGTGAGGCTCCCTCCC-ATATAACTCAAGGTTTGAGTCTGTTG	492
Query	577	CTTGAGTGGGTCTGCATGCCATGATGGTAGAGAATGG--ACATATATCGGAGTTGATG	633
Sbjct	493	CTTGGTCAGCAAGTGTCCATGATGGCACCAGTTGGTTGACA--ATTGGAATTTCTG	549
Query	634	GCCCTGACAGTAATGCATTG----ATCAAAATAAAATATGGAGAAGCATATACTGACAC	688
Sbjct	550	GCCCAGACAATGGGGCTGTGGCTGTATTGAAATACAATG-GCATAA---TA-ACAGACAC	604
Query	689	ATACCATTCCATGCA--AACACATCCTAAGAACACAAGAAAGTGCCTGCAATTGCATC	746
Sbjct	605	-TATCAAGAGT-TGGAGGAATAACATACTGAGAACTCAAGA--GT--CTGAATGTGCATG	658
Query	747	GGG----GGAGATTGTTATCTTATGATAACTGATGGCTCAGCTTCAGGAATTAGTAAATG	802
Sbjct	659	TGTAATGGCTCTTGCTTTACTGTAATGACTGACGGACCA----AGTAATG-GTCAG-G	711
Query	803	CAGATTTCTTAAGAT--TCGAGAGGGTCAATAATAAAAG----AAATATTTCCAACAG	855
Sbjct	712	CA--TCACATAAGATCTTCAAAATGGA--AAAAGGAAAAGTGGTTAAATCAGTCGAATTG	767
Query	856	GAAG-----AGTAGAGCATACTGAAGAATGCACATG-----CGGATTTGCCAGCAATAAA	905
Sbjct	768	GATGCTCCCAATT-ATCACTATGAGGAATGCTCCTGTTATCCTGAT--GCCGGCGA---A	821
Query	906	ACCATAGAATGTGCCTGTAGAGATAACAGTTACACAGCAAAAAGACCCTTTGT-----C	959
Sbjct	822	ATCACA---TGTGTGTGCAGGGATAATTGGCATGGCTCAAATCGGCCATGGGTATCTTTC	878
Query	960	AAATTTAAATGTGGAGACTGATACAGCTGAGATAAGATTGATGTGCACAGAGACTTATTTG	1019
Sbjct	879	AATCAAAATTTGGAG-----TA----TCAAATAGGATATATATGCAGTG-GAGTTTTTGG	928
Query	1020	-GACACCCCCAGACCAGATGATGGAAGCA--TAACAG-GGCCTTGCGAA-TCTAATGGGG	1074
Sbjct	929	AGACAATCCACGCCCAATGATGGAA-CAGGTAGTTGTGGTCCGGTGTCTCTAACGGGG	987
Query	1075	ACAAAGGGCGTGGAGGCATCAAGGGAGGATTTGTCCATCAAAGA-----ATGGCATCCAA	1129
Sbjct	988	-CATATGGGGTAAA-----AGGG----TTT-TCATTTAAATACGGCAATGGTGTCTG-	1033
Query	1130	GATTGGAAGATGGTACTCCCGAACGATGTCTAAAACCTGAGAGAATGGGGATGGAACCTGTA	1189
Sbjct	1034	GATCGGGAGA----AC-----AAAAAGCACTAATCCAGGAGC---GGCTTTGAAATG-A	1080
Query	1190	TGTCAGGTATGATGGAGACCCATGGACTGACAGTGACGCCCTTGCTCAT-AGTGGAGTAA	1248
Sbjct	1081	TTT--GG---GATCCAAATGGGTGGACTGAAACGGACAG--TAGCTTTTCAGTGAAACAA	1133
Query	1249	TGGTCTCAATGAAAGAACCTGGTTGGTATTCTTTGGCTTCGAAATAAAAGATAAGAAAT	1308
Sbjct	1134	-GATATCGTAGCAATAAC-TGATTGGTCA-----GGAT-----ATAGCGG----GA---	1173
Query	1309	GTGATGTCCCCTGTATTGGGATAGAGATGGTACACGATGGTGGAATAAAGACTTGGCACT	1368
Sbjct	1174	GTTTTGTCCA-----GCATCCAGA----AC----TG-----ACAGGACTAG--ATT	1209
Query	1369	CAGCAGCAACAGCCATTTACTGTTTGGATGGGCTCAGGACAATTGCT--GTGGGACACTGT	1426

```

Sbjct 1210  -----GCATAAGACCTT-----GTTTC-TGGGTGAG-----TTGATCAGAGGGCGGCCCA 1254
          ||| || | | | ||| ||| | | | ||| | | ||| |
Query 1427  CACAGGTGTTTGTATATGGCTCTGTAATGGAGGA--ATGGTTGAATCTGTTCTAAACCCTTT 1484
          || | || | || | || | || | || | || | || | || | || | || | || | || |
Sbjct 1255  AAGAGAGCACAAATTTGGACTAGTG--GGAGCAGCATATCTTTTTGTTGTTGATAA----TA 1308
          || | | | | | || | || | || | || | || | || | || | || | || | || |
Query 1485  GTTCCTATTTTTGTTTGAACAATGTTCCTTACTGGACTTAATTGT--TTC-----TGAAAA 1537
          || | | | | || | || | || | || | || | || | || | || | || | || | || |
Sbjct 1309  GTGACACTGTGGTTGGTC--TTGGCCAGACGGTGCTGAATTGCCATTACCATTGACAA 1366
          || | | | | || | || | || | || | || | || | || | || | || | || | || |
Query 1538  ATGCTCTGTTACTAC 1553
          | | |||
Sbjct 1367  GTAGT--TGTTCA 1377
  
```

**Στέλεχος Ιού γρίπης τύπου A Leningrad και Ιού γρίπης τύπου A USA.**

Στο παρακάτω παράδειγμα θα γίνει σύγκριση μεταξύ στελεχών του Ιού γρίπης που είχε πρωτο- εμφανιστεί στο Leningrad το 1957, και Ιού γρίπης που είχε εμφανιστεί στις ΗΠΑ το 2005. (**Influenza virus type A/Leningrad/134/17/57 (H2N2) PB2 gene, complete cds GenBank: M81581.2** με **Influenza A virus SC35M polymerase subunit PB2 gene, complete cds 2005 USA GenBank: DQ266097.1**)



Influenza virus type A/Leningrad/134/17/57 (H2N2) PB2 gene, complete cds

Length=2341

Score = 2584 bits (1399), Expect = 0.0

Identities = 2039/2353 (87%), Gaps = 24/2353 (1%)

Strand=Plus/Plus

```

Query 1   AGCGAAAGCAGGTCAAATATATTC AATATGGAGAGAATAAAAGA ACTA-AGAGACCTAAT 59
          ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | |||
Sbjct 1   AGCAAAAGCAGGTCAAATATATTC AATATGGAAAGAATAAAAGA ACTACGGA-ATCTGAT 59
  
```



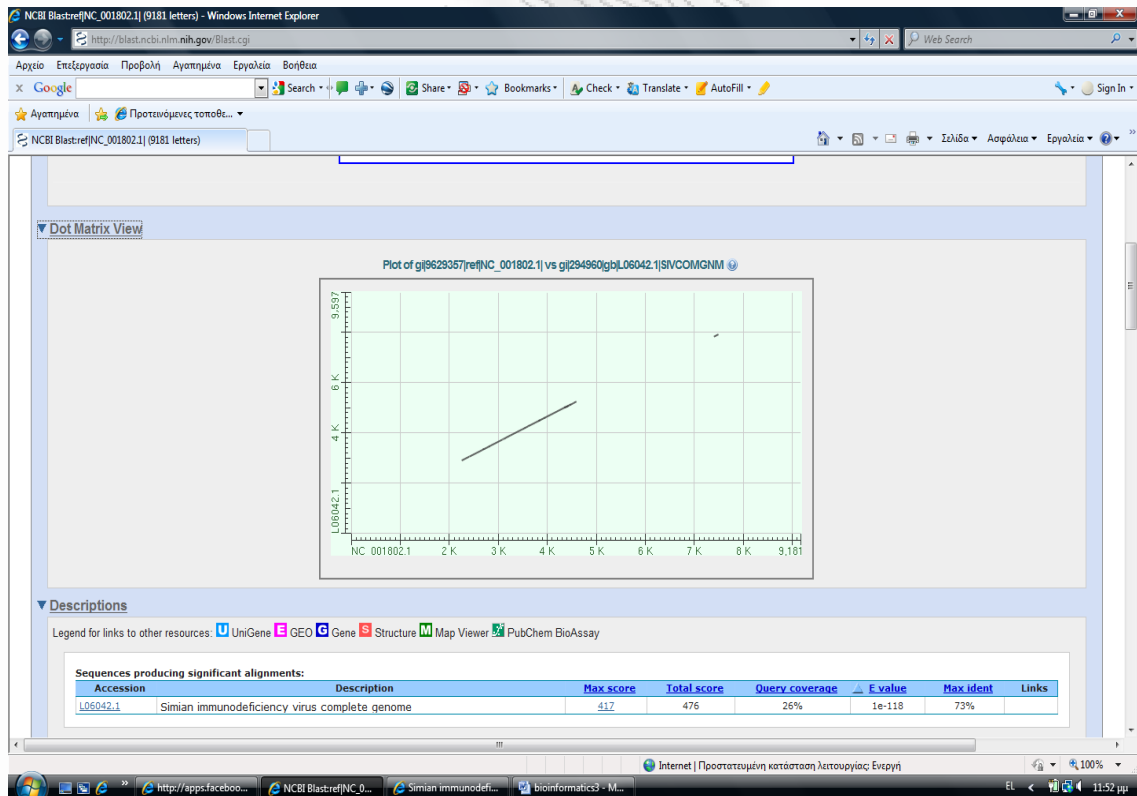
Query	60	GTCACAATCCCGCACCCGCGAGATACTCACTAAAACCACTGTGGACCATATGGCCATAAT	119
Sbjct	60	GTCGCAGTCTCGCACTCGCGAGATACTAACAAAAACCACAGTGGACCATATGGCCATAAT	119
Query	120	CAAAAAGTACACATCAGGAAGACAGGAGAAGAACCCCG-CGCTCAGGATGAAGTGGATGA	178
Sbjct	120	TAAGAAGTACACATCAGGGAGACAGGAAAAGAA-CCCGTCACCTAGGATGAAATGGATGA	178
Query	179	TGGCAATGAAATATCCAATTACGGCAGATAAGAGGATAATGGAAATGATTCCTGAGAGGA	238
Sbjct	179	TGGCAATGAAATATCCGATTACAGCTGACAAGAGGATAACAGAAATGATTCCTGAGAGAA	238
Query	239	ATGAACAAGGACAAAACCTCTGGAGCAAAAACAAATGACGCCGCTCAGACCGAGTGATGG	298
Sbjct	239	ATGAGCAAGGGCAAACCTCTATGGAGTAAAATGAGTGTGCCGGATCGGATCGAGTGATGG	298
Query	299	TATCACCTCTGGCTGTGACATGGTGGAAATAGGAATGGGCCAACAAACAAGTACAGTTCACT	358
Sbjct	299	TATCACCTCTGGCTGTGACATGGTGGAAATAGAAATGGACCAATGACAAGTACGGTTCATT	358
Query	359	ATCCAAAGGTATATAAACTTATTTTCGAAAAAGTCGAAAGGATGAAACATGGGACCTTTG	418
Sbjct	359	ATCCAAAAATCTACAAAACCTTATTTTGAGAAAAGTCGAAAGGTAAAACATGGAACCTTTG	418
Query	419	GCCCTGTTCACTTTAGGAATCAAGTTAAGATAAGACGGAGAGTTGACATAAACCCCTGGAC	478
Sbjct	419	GCCCTGTCCATTTAGAAAACCAAGTCAAAATACGCCGAAGAGTTGACATAAATCCTGGTC	478
Query	479	ACGCAGACCTTAGTGCCAAAGAGGCACAGGATGTAATCATGGAAGTTGTTTTCCCAAACG	538
Sbjct	479	ATGCAGACCTCAGTGCCAAGGAGGCACAGGATGTAATCATGGAAGTTGTTTTCCCTAACG	538
Query	539	AAGTGGGGGCCAGAACTACTAACATCGGAATCACAACACTGACAATAACAAAAGAG-AAAAAG	597
Sbjct	539	AAGTGGGGGCCAGGATACTAACCTCGGAATCGCAATTAACAATAACCAAGAGAAAAAAG	598
Query	598	GAGAACTCCAGGATTGTAATAATTGCTCCCTTGTATGGTAGCATACTAGTAAAGAGAA	657
Sbjct	599	AAG-AACTCCAGGATTGCAAAATTTCTCCCTTGTATGGTTGCGTACATGTTAGAGAGAGAA	657
Query	658	-TTGGTCCGCAAAAACAAGTTCCCTCCAGTGGCTGGTGGGACAAGCAGTGTGTATATTGA	716
Sbjct	658	CTT-GTCCGAAAAACGAGATTTCTCCAGTGGCTGGTGGGAACAAGCAGTGTGTACATTGA	716
Query	717	GGTGTTCGACTTTGACCCAAGGAACATGCTGGGAGCAGATGTACACTCCAGGAGGGGAAGT	776
Sbjct	717	AGTGTTCGACTTGAATCAAGGAACATGCTGGGAAACAGATGTACACTCCAGGTGGAGAAGT	776
Query	777	GAGAAATGATGATATTGACCAAAGCT-TAATTATTGCTGCCAGAAACATAGTGAGAAGGG	835
Sbjct	777	GAGGAATGATGATGTTGATCAAAG-TCTAATTATTGCAGCCAGGAACATAGTGAGAAGAG	835

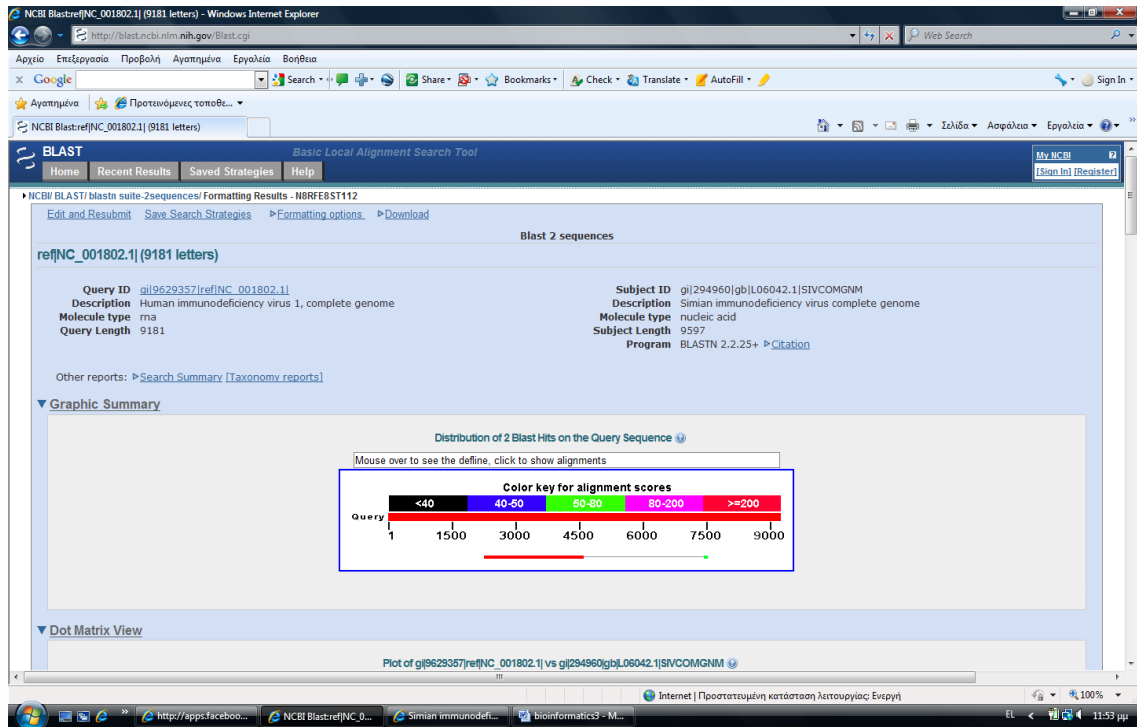
Query	836	CAACAGTATCAGCAGATCCACTAGCATCTCTATTGGAGATGTGCCACAGCACACAGATTG	895
Sbjct	836	CAGCAGTATCAGCAGATCCACTAGCATCTTATGGAGATGTGCCACAGCACACAGATTG	895
Query	896	GTGGAATAAGAATGGTAGACATCCTGAGGCAGAATCCAACAGAGGAACAAGCCGTGGATA	955
Sbjct	896	GCGGGACAAGGATGGTGACATCTTAGGCAGAACCCAACGGGAAGAGCAAGCTGTGGATA	955
Query	956	TATGCAAGGCAGCAATGGG-CTTGAGGATTAGCTCATCTTTCAGCTTCGGTGGATTCACT	1014
Sbjct	956	TATGCAAGGCTGCAATGGGAC-TGAGAATCAGCTCATCTTTCAGTTTTGGCGGGTTCACA	1014
Query	1015	TTTAAAAGAATAAGCGGATCGTCAGTCAAGAGAGAAGAAGAAGTCTTACCGGCAACCTT	1074
Sbjct	1015	TTTAAGAGAACAAGCGGATCATCAGTCAAGAGAGAGGAAGAAGTCTTACGGGAATCTT	1074
Query	1075	CAAACACTGAAAATACGAGTGCATGAGGGGTATGAAGAGTTCACAATGGTTGGGAGAAGA	1134
Sbjct	1075	CAAACATTGAAAATAAGGGTGCATGAGGGATACGAGGAGTTCACAATGGTTGGGAAAAGG	1134
Query	1135	GCAACAGCTATCCTCAGAAAAGCAACCAGGAGATTGATTCAGCTAATAGTAAGTGGGAGA	1194
Sbjct	1135	GCAACAGCTATACTCAGAAAAGCAACCAGGAGATTGATTCAGCTGATAGTGAGTGGGAAGA	1194
Query	1195	GACGAACAGTCAATTGCTGAAGCAATAATGTGGCCATGGTATTTTCAAGAGGATTGC	1254
Sbjct	1195	GACGAACAGTCGATAGCCGAAGCAATAATGTGGCCATGGTATTTTCAAGAAGATTGT	1254
Query	1255	ATGATCAAAGCAGTTCGAGGTGACCTGAACTTT-GTCAATAGAGCAAACCAGCGGTTGAA	1313
Sbjct	1255	ATGATAAAAGCAGTTAGAGGTGATCTGAA-TTTCGTTAATAGGGCAAATCAGCGATTGAA	1313
Query	1314	CCCCATGCATCAACTCTTGAGGCATTTCCAAAAGGATGCAAAAGTGCTTTTCCAAAACCTG	1373
Sbjct	1314	TCCCATGCATCAACTTTAAGACATTTTCCAGAAGGATGCGAAAGTGCTTTTCCAAAATTG	1373
Query	1374	GGGAATTGAGCCCATCGACAATGTGATGGGAATGATCGGGATATTACCCGACATGACCCC	1433
Sbjct	1374	GGGAATTGAACATATCGACAATGTGATGGGAATGATTGGGGTATTACCAGACATGACTCC	1433
Query	1434	AAGTACTGAGATGTCGCTGAG-GGGAATAAGAGTCAGCAAGATGGGAGTAGATGAATACT	1492
Sbjct	1434	AAGCACAGAGATGTCAATGAGAGGG-TTAAGAGTCAGCAAAATGGGCGTAGATGAATACT	1492
Query	1493	CCAGTACAGAGAGGGTGGTGGTGAGCATCGACCGTTTTTAAAGAGTTCGGGATCAACGAG	1552
Sbjct	1493	CCAGCGCGGAGAGAGTGTGGTGGTGAGCATTCACCGTTTTTGAGAGTTCGAGACCAACGAG	1552
Query	1553	GGAAATGTACTACTATCACCCGAGGAAGTCAGCGAGACACAAGGG-ACGGAGAAGCTAACA	1611
Sbjct	1553	GAAATGTACTATTATCTCCTGAGGAGGTGAGTGAACAC-AGGGAACAGAGAAACTGACA	1611
Query	1612	ATAACTTATTCATCATCAATGATGTGGGAGATCAATGGTCTGAGTCAGTGCTGATCAAT	1671



### Σύγκριση με Blast μεταξύ Ιού HIV και SIV

Στο παρακάτω παράδειγμα θα γίνει σύγκριση μεταξύ του SIV1 (L06042.1) με τον HIV1 (NC\_001802.1)





Score = 417 bits (462), Expect = 1e-118  
 Identities = 1560/2397 (66%), Gaps = 114/2397 (4%)  
 Strand=Plus/Plus

```

Query 2255 AATCCATACAATACTCCAGTATTTGCCATAAAGAAAAAGACAGTACTAAATGGAGAAA 2314
          ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 2906 AATCCTTATAATACCCAGTATTTGTAATTAGGAAAAAGACAAAAGAAATGGAGGATG 2965

Query 2315 TTAGTAGATTTAGAGAACTTAATAAGAGAACTCAAGACTTCTGGGAAGTTCATTAGGA 2374
          | | ||||| ||| || | ||| || | ||||| || | || ||||| |||
Sbjct 2966 CTCATTGATTTAGAAAATTAATGAGCCACACAAGATTTTTTCGAGGTACAATTGGGA 3025

Query 2375 ATACCACATCCCGCAGGGTTaaaaaagaaaaaTCAGTAACAGTACTGGATGTGGGTGAT 2434
          ||||| || || | ||| ||||| || ||||| ||||| || ||| | |||
Sbjct 3026 ATACCCACCCAGGAGGACTAAAACGGCAAAAAC---TAACAATTATAGATTTAAAAGAT 3082

Query 2435 GCATATTTTTCAGTTCCCTTAGATGAAGACTTCAGGAAGTACTGCATTTACCATACCT 2494
          ||||| ||||| ||| || | ||| ||||| || || ||||| ||||| |||
Sbjct 3083 GCATATTACTCAGTCCCCCTGGACAAAGAGTTTAGGCCATACACGGCATTACAGTGCCA 3142

Query 2495 AGTATAAACAATGAGACACCAGGGATTAGATATCAGTACAATGTGCTCCACAGGGATGG 2554
          || ||||| ||||| | ||| || ||||| || | ||| | ||||| |||||
Sbjct 3143 AGCATAAATAATGCCTCCCGGGAGAAAGATATCAATTTACAGTGTTACCACAGGGATGG 3202

Query 2555 AAAGGATCACCAGCAATATCCAAAGTAGCATGACAAAATCTTAGAGCCTTTAGAAAA 2614
          ||||| || | || ||||| || | ||| | || ||||| |||||
Sbjct 3203 AAAGGATCTCCCACTATCTCCAAAGCACAATAAATCAGATACTACAGCCATTTAGAAAG 3262
    
```

```

Query 2615 CAAAATCCAGACATAGTTATCTATCAATACATGGATGATTTGTATGTAGGATCTGAC--- 2671
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct 3263 AAGTATTCAGATCTCACTCTAATACAGTACATGGATGACTTATGATTGGCACTGATAGA 3322

Query 2672 TTAGAAATAGGGCAGCA-----TAGAACAA-AAATAGAGGAGCTGAGACAACATCTGTTG 2725
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct 3323 TCAGAGAAGGCACATCAGGAGATAGTACAACAAATAGTG-----ACAGCA-CTATTA 3373

Query 2726 AGGTGGGGACTTACCACACCAGACAAAAAACATCAGAAAGAACCTCCATTCCTTTGGATG 2785
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct 3374 AAAGTGGGATTTAAGGTCCCAAAGAAAAATGCGAGGACCAATATCCAATGCAGTGGTTA 3433

Query 2786 GGTATGAACCTCCTGATAAATGGACAGTA-CAGCCTATAGTGTGCC-----AGA 2838
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct 3434 GGTACACTCTTCATCCAGACAAGTGG-CAGTTGCAGAAAATAGAACTCCCTAATATAGA 3492

Query 2839 AAAAGACAGCTGGACTGTCAATGACATACAGAAGTTAGTGGGAAATGAATTGGGCAAG 2898
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct 3493 TGACGAAATC---ACAGTGAATCAGTTACAGAAATGATAGGAGTACTTAATGGGCAAG 3549

Query 2899 TCAGATTTACCCAGGGATTAAGTAAGGCAATTATGTAAGCTCCTTAGAGGAACCAAAGC 2958
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct 3550 CCAGATTTATTCAGGAATTAAGACAAAAGAGCTCTGTAAAGTGTATCAGAGGAACTAAACA 3609

Query 2959 ACTAACAGAAGTAATACCACTAACAGAAAGAGCAGAGCTAGAAGTGGCAGAAAAACAGAGA 3018
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct 3610 ATTAACAGAAGTACTCACACTTACAGAAAGCAGCTGAGGCAGAATTAGAAGAGAATAGACA 3669

Query 3019 GATTCTAAAAGAACCAGTACATGGAGTGT---ATTATGACCCATCAAAGACTTAATAGC 3075
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct 3670 AATCTTAAAGGAA--GAACAGGCAGGCTCTTATTATGACCCTAAAAAACCTTTAGAAGC 3726

Query 3076 AGAAAT-ACAGAAGCAGGGCAAGGCCAATGGACATATCAAATTTATCAAGAGCCATTTA 3134
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct 3727 ACACATCACTAAACTAGGGTCCCAGC-AATGGGGATACATGATTAAGCAAGAGC-----A 3780

Query 3135 AAAAT-----CTGAAAACAGGAAAATATGCAAGAATGAGGGGTGCCCACACTAATGA 3186
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct 3781 AAAAGGCCACCCTAATAACAGGAAAGACAGCAAAAACCTTTGCAGCACATAGTAATGA 3840

Query 3187 -TGTAAAACAATTAACAGAGGCAGTGCAAAAATAACCACAGAA--AGCATAGTAATATG 3243
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct 3841 CTATCAATCTCTTG-CTCAATTGTTAAATAAAATAGGGATCCAATCACTATGGTATT--G 3897

Query 3244 GGGAAAGACTCSTAAATTTAAACTGCCCATACAAAAGGAAACATGGGAAACATGGTGGAC 3303
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct 3898 GGGAAAGGTACCAACATTCATCTCCAGTAAAAAGAGAGGAATGGGAAAAATGGTGGAC 3957

Query 3304 AGAGTATTGGCAAGCCACCTGGATTTCCTGAGTGGGAGTTGTTAATACCCCTCCCTTAGT 3363
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct 3958 AGACTATTGGCAAGCCACATGGGTACCAGAAGTCAAATTCATAAGTACCCCGCCATTGGT 4017

Query 3364 GAAATTTATGGTACCAGTTAGAGAAAGAACCATAGTAGGAGCAGAAACCTTCTATGTAGA 3423

```





ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΑΘΑΝΑΣΙΟΣ ΠΑΥΛΙΔΗΣ

Sbjct 4783 ACCAAACTTTGTCAGTGACAAGTTCAAAGCAGCATGTTGGTGGTGTGGCATAGAACACAC 4842

Query 4189 ATTTGGAATTCCTTACAATCCCCAAAGTCAAGGAGTAGTAGAATCTATGAATAAAGAATT 4248  
 || ||||| ||||| || ||||| || ||| ||||| || | |||| | ||

Sbjct 4843 TACAGGTATCCATACACCCACAAAGCCAGGAATAGTAGAGTCCAAAAATAGGTATTT 4902

Query 4249 AAAGAAAATTATAGGACAGGTAAGAGATCAGGCTGAACATCTTAAGACAGCAGTACAAAT 4308  
 ||| || ||| ||| | ||||| | | ||||| | | ||||| ||||

Sbjct 4903 AAAAGAAGCTATCTCACAAATTAGAGACGACGTCACACATTTGCAAACAGCAGTGGCAAT 4962

Query 4309 GGCAGTATTCATCCACAATTTTAAAAGAAAAGggggggattgggggg-TA-CAGTGCAGGG 4366  
 |||| | ||||| | ||||| ||||| || || || ||||| || || |||||

Sbjct 4963 GGCAACCTTCATCTTAAATTTTAAAAGAAAGGGAGGAATAGGGGGGATATCA--CCAGGA 5020

Query 4367 GAAAGAATAGTAGACATAATA-GCAACAGACATacaaacta-aagaattacaaa-acaa 4423  
 ||||| ||| | ||||| || | | ||| ||||| || | ||||| ||||

Sbjct 5021 GAAAG----GTACA-TTAATATGCTATATACA---GAACTACAATTACAACAAAATACAA 5072

Query 4424 -atta-caaaaattcaaaattTTCGGGTTTATTACAGGGACAGCAGAA-ATCCACTTTGG 4480  
 || | ||||| || | ||||| ||||| ||| | ||| ||| | | | |||

Sbjct 5073 CATCACCAAAATTTTCGAATTTTAGGGTTTACTAC--CGCCAGGGGAAGAACGA--GTGG 5128

Query 4481 AAAGGACCAGCAAAGCTCCTCTGGAAAGGTGAAGGGGCAGTAGTAATACA---AGATAAT 4537  
 || ||||| || || ||||| ||||| ||||| || || ||| |

Sbjct 5129 AAGGGACCAGCTCGTCTACTTTGGAAAGGTGAAGGTGCAGTAGTGGTGCAAACAGAGGAG 5188

Query 4538 AGTGACATAAAAGTAGTGCCAAGAAGAAAAGCAAAGATCATTAGGGATTATGGAAAA 4594  
 | |||| | || | ||||| ||||| ||||| || ||||| || ||||| ||

Sbjct 5189 GGAGACATTTTTGCAGTACCAAGAAGAAAGGCAAAAATAATTACAGACCATGGAGAA 5245

Score = 59.0 bits (64), Expect = 2e-10  
 Identities = 68/92 (74%), Gaps = 0/92 (0%)  
 Strand=Plus/Plus

Query 7400 TTATTGTCGGTATAGTGCAGCAGCAGAACAATTTGCTGAGGGCTATTGAGGCGCAACAG 7459  
 || |||| |||| | ||||| ||||| || ||||| ||| ||||| || ||

Sbjct 7821 TTGTTGGCTGGGATAGTGCAGCAGCAGCAAAAGTTGCTGGAGGCAGTTGAGGCACAGCAA 7880

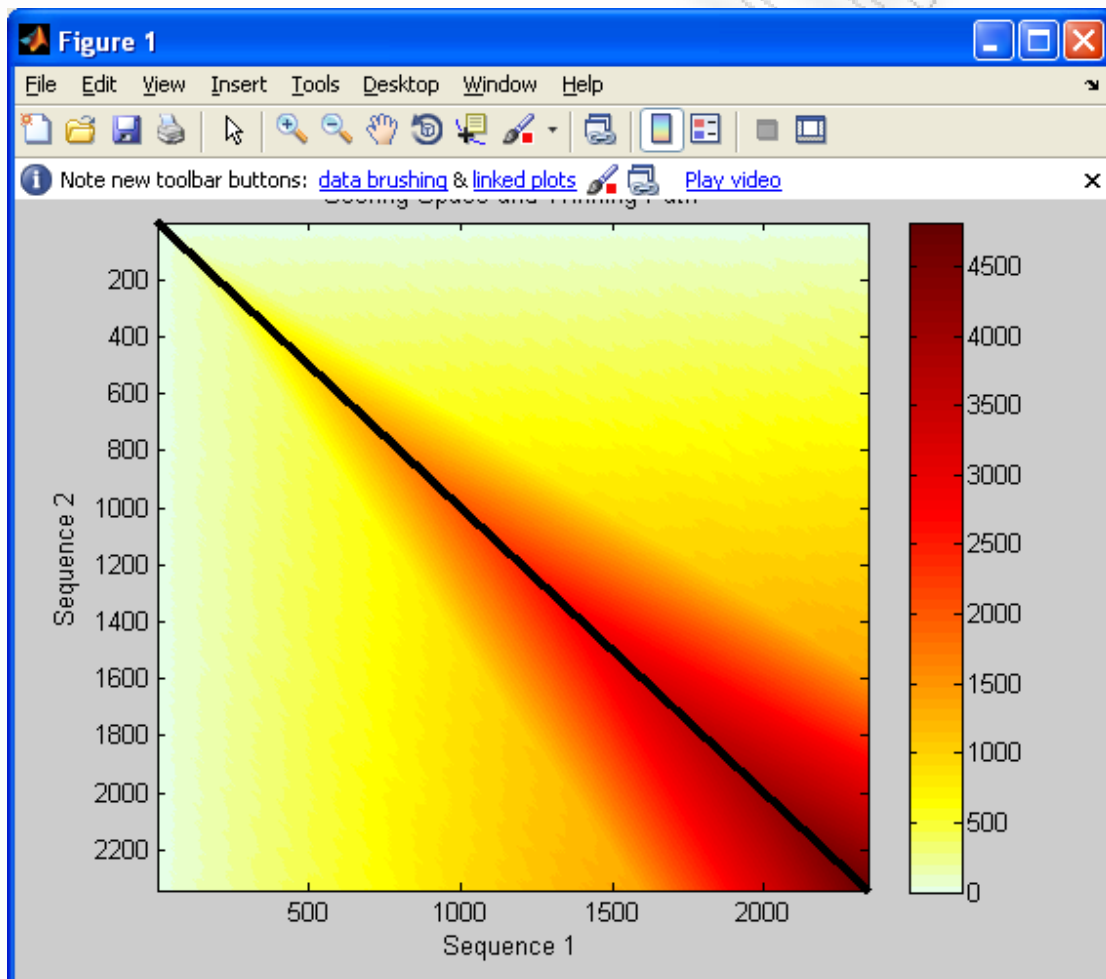
Query 7460 CATCTGTGCAACTCACAGTCTGGGGCATCAA 7491  
 || || || | || ||||| ||||

Sbjct 7881 CACCTCTAGGCTTAACTGTCTGGGGTGTCAA 7912

### Αποτελέσματα από matlab:

Σύγκριση των ακολουθιών M81581.2 και DQ266097.1 με τον αλγόριθμο Smith - Waterman και παρουσίαση αποτελέσματος. Με την παρακάτω εντολή παράγεται το γράφημα του score της σύγκρισης

```
[Score1, Alignment1, Start1] = swalign(A,B, 'Showscore', true);
```



Με την παρακάτω εντολή παράγεται το αποτέλεσμα του alignment των ακολουθιών `showalignment(Alignment1);`

Identities = 2027/2341 (87%), Positives = 2199/2341 (94%)



0833 GGGCAACAGTATCAGCAGATCCACTAGCATCTCTATTGGAGATGTGCCACAGCACACAGATTGG  
 0897 CGGGACAAGGATGGTGGACATTTCTTAGGCAGAACCCAACGGAAGAGCAAGCTGTGGATATATGC  
 ||:| |||:|||||:|||| | | ||||| | ||||:|:|:| ||| | ||||| |||  
 0897 TGGAATAAGAAATGGTAGACATCCTGAGGCAGAAATCCAACAGAGGAACAAGCCGTGGATATATGC  
  
 0961 AAGGCTGCAATGGGACTGAGAAATCAGCTCATCCTTCAGTTTTGGCGGTTTCACATTTAAGAGAA  
 |||||:||||||| | |||:| | ||||| | ||| | | | | |:|||||:|||||:||||  
 0961 AAGGCAGCAATGGGCTTGAGGATTAGCTCATCTTTCAGCTTCGGTGGATTCACITTTAAAAGAA  
 1025 CAAGCGGATCATCAGTCAAGAGAGAGGAAGAAGTGCTTACGGGCAATCTTCAAACATTTGAAAAT  
 |||||:|||||||:|||||||:||||||| | ||| | ||||| | |||||  
 1025 TAAGCGGATCGTCAGTCAAGAGAGAAAGAAGAAGTGCTTACCGGCAACCTTCAAACACTGAAAAT  
  
 1089 AAGGGTGCATGAGGGATACGAGGAGTTCACAATGGTTGGGAAAGGGCAACAGCTATACTCAGA  
 | |:|||||||:| | |:||||||| |||||:|:|:| ||||| | |||||  
 1089 ACGAGTGCATGAGGGGTATGAAGAGTTCACAATGGTTGGGAGAAGAGCAACAGCTATCCTCAGA  
 1153 AAAGCAACCAGGAGATTGATTCAGCTGATAGTGAGTGGAGAGACGAACAGTCGATAGCCGAAG  
 |:|||||||:|||||||:|||||||:|||||||:|:|:| | | | |  
 1153 AAGGCAACCAGGAGATTGATTCAGCTAATAGTAAGTGGGAGAGACGAACAGTCAATTGCTGAAG  
 1217 CAATAATTGTGCCATGGTATTTTCACAAGAAGATTGTATGATAAAAGCAGTTAGAGGTGATCT  
 |||||:|||||||:||||||| | ||| | ||||| | ||||| |  
 1217 CAATAATTGTGCCATGGTATTTTCACAAGAGGATTGCATGATCAAAGCAGTTAGAGGTGACCT  
 1281 GAATTTTCGTTAATAGGGCAAATCAGCGATTGAAATCCCATGCATCAACTTTAAGACATTTTCAG  
 || | | | | |:||||| | |:||||| | ||||| | |:|:|:| | | | |  
 1281 GAACTTTGTCGAATAGAGCAAACCAGCGTTGAACCCCATGCATCAACTCTTGAGGCATTTCCAA  
 1345 AAGGATGCGAAAGTGCTTTTCAAATTTGGGGAATTGAACATATCGACAATGTGATGGGAATGA  
 |||||:||||||| | ||| | |||||:| | ||||| | ||||| |  
 1345 AAGGATGCAAAAGTGCTTTTCAAACCTGGGGAATTGAGCCCATCGACAATGTGATGGGAATGA  
 1409 TTGGGGTATTACCAGACATGACTCCAAGCACAGAGATGTCAATGAGAGGGTTAAGAGTCAGCAA  
 | |:||||| | ||||| | ||| |:|||||: | |:|:|:| ||||| | |||||  
 1409 TCGGGATATTACCCGACATGACCCCAAGTACTGAGATGTCTGAGGGGAATAAGAGTCAGCAA  
 1473 AATGGGCGTAGATGAATACTCCAGCGCGGAGAGAGTAGTGGTGAGCATTGACCGTTTTTTGAGA  
 |:|||| | |||||:|:|:|:|:| ||||| | ||||| |:| |  
 1473 GATGGGAGTAGATGAATACTCCAGTACAGAGAGGGTGGTGGTGAGCATCGACCGTTTTTTAAGA  
 1537 GTTCGAGACCAACGAGGAATGTACTATATCTCCTGAGGAGGTCAGTGAACACAGGGAACAG  
 |||||:| | |||||:||||||| | |:| | |||||:| | |:|:|:|:|:| |  
 1537 GTTCGGGATCAACGAGGCAATGTACTACTATCACCCGAGGAAGTCAGCGAGACACAAGGGACGG  
 1601 AGAAACTCACAATAACTTACTCATCGTCAATGATGTGGGAGATTAATGGCCCTGAGTCAGTGTT  
 |||:|:|:| ||||| | |:| | ||||| | ||| | ||||| | |  
 1601 AGAAGCTAACAAATAACTTATTCATCATCAATGATGTGGGAGATCAATGGTCTGAGTCAGTGCT  
 1665 GGTCAATACCTATCAGTGGATCATCAGAAACTGGGAAACTGTTAAAATTCAGTGGTCTCAGAAT  
 |:||||| | | |:|||||||:||||||| | |||||:|:|:|:|:| | |  
 1665 GATCAATACTTACCAATGGATCATCAGGAACTGGGAAACTGTGAAAATTCAGTGGTACAGGAT  
 1729 CCTACAATGCTATACAATAAAATGGAATTTGAGCCATTTAGTCTTTAGTTTCTAAGGCCATTA  
 | | ||||| | | |:||||||| | |||||:| | |:| | ||||| | |  
 1729 CCCACAATGTTATATAACAAGATGGAATTCGAGCCATTTCAATCTCTGGTCCCTAAGGCAGCCA



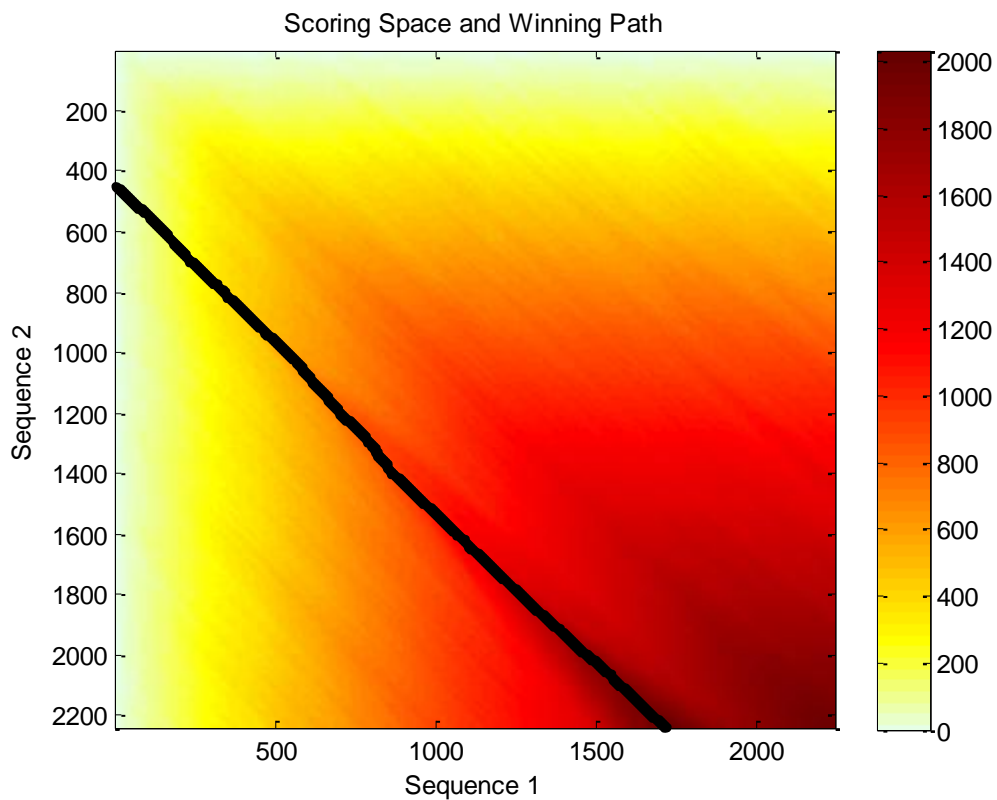
Με την παρακάτω εντολή παράγεται ο πίνακας (scoring matrix) που χρησιμοποιείται από τον αλγόριθμο

Matrix1 = blosum(90);

5	-2	-2	-3	-1	-1	-1	0	-2	-2	-2	-1	-2	-3	-1	1	0	-4	-3	-1	-2	-1	-1	-6
-2	6	-1	-3	-5	1	-1	-3	0	-4	-3	2	-2	-4	-3	-1	-2	-4	-3	-3	-2	0	-2	-6
-2	-1	7	1	-4	0	-1	-1	0	-4	-4	0	-3	-4	-3	0	0	-5	-3	-4	4	-1	-2	-6
-3	-3	1	7	-5	-1	1	-2	-2	-5	-5	-1	-4	-5	-3	-1	-2	-6	-4	-5	4	0	-2	-6
-1	-5	-4	-5	9	-4	-6	-4	-5	-2	-2	-4	-2	-3	-4	-2	-2	-4	-4	-2	-4	-5	-3	-6
-1	1	0	-1	-4	7	2	-3	1	-4	-3	1	0	-4	-2	-1	-1	-3	-3	-3	-1	4	-1	-6
-1	-1	-1	1	-6	2	6	-3	-1	-4	-4	0	-3	-5	-2	-1	-1	-5	-4	-3	0	4	-2	-6
0	-3	-1	-2	-4	-3	-3	6	-3	-5	-5	-2	-4	-5	-3	-1	-3	-4	-5	-5	-2	-3	-2	-6
-2	0	0	-2	-5	1	-1	-3	8	-4	-4	-1	-3	-2	-3	-2	-2	-3	1	-4	-1	0	-2	-6
-2	-4	-4	-5	-2	-4	-4	-5	-4	5	1	-4	1	-1	-4	-3	-1	-4	-2	3	-5	-4	-2	-6
-2	-3	-4	-5	-2	-3	-4	-5	-4	1	5	-3	2	0	-4	-3	-2	-3	-2	0	-5	-4	-2	-6
-1	2	0	-1	-4	1	0	-2	-1	-4	-3	6	-2	-4	-2	-1	-1	-5	-3	-3	-1	1	-1	-6
-2	-2	-3	-4	-2	0	-3	-4	-3	1	2	-2	7	-1	-3	-2	-1	-2	-2	0	-4	-2	-1	-6
-3	-4	-4	-5	-3	-4	-5	-5	-2	-1	0	-4	-1	7	-4	-3	-3	0	3	-2	-4	-4	-2	-6
-1	-3	-3	-3	-4	-2	-2	-3	-3	-4	-4	-2	-3	-4	8	-2	-2	-5	-4	-3	-3	-2	-2	-6
1	-1	0	-1	-2	-1	-1	-1	-2	-3	-3	-1	-2	-3	-2	5	1	-4	-3	-2	0	-1	-1	-6
0	-2	0	-2	-2	-1	-1	-3	-2	-1	-2	-1	-1	-3	-2	1	6	-4	-2	-1	-1	-1	-1	-6
-4	-4	-5	-6	-4	-3	-5	-4	-3	-4	-3	-5	-2	0	-5	-4	-4	11	2	-3	-6	-4	-3	-6
-3	-3	-3	-4	-4	-3	-4	-5	1	-2	-2	-3	-2	3	-4	-3	-2	2	8	-3	-4	-3	-2	-6
-1	-3	-4	-5	-2	-3	-3	-5	-4	3	0	-3	0	-2	-3	-2	-1	-3	-3	5	-4	-3	-2	-6
-2	-2	4	4	-4	-1	0	-2	-1	-5	-5	-1	-4	-4	-3	0	-1	-6	-4	-4	4	0	-2	-6
-1	0	-1	0	-5	4	4	-3	0	-4	-4	1	-2	-4	-2	-1	-1	-4	-3	-3	0	4	-1	-6
-1	-2	-2	-2	-3	-1	-2	-2	-2	-2	-2	-1	-1	-2	-2	-1	-1	-3	-2	-2	-2	-1	-2	-6
-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	1

Σύγκριση των ακολουθιών NC\_001802.1 και L06042.1 με τον αλγόριθμο Smith - Waterman και παρουσίαση αποτελέσματος

Με την παρακάτω εντολή παράγεται το γράφημα του score της σύγκρισης  
`[Score2, Alignment2, Start2] = swalign(C,D, 'Showscore', true);`



Με την παρακάτω εντολή παράγεται το αποτέλεσμα του alignment των ακολουθιών:  
`showalignment(Alignment2);`

Identities = 1074/1850 (58%), Positives = 1390/1850 (75%)

```

0001 GGTCCTCTGGTTAGACCAGATCTGAGCCT-GGAGCTCTCTG--GCTAACTAGGGAACCCACT
    |: || | |||:|:| | | |: | | || ||||| | |||:|:| | : |:|
0001 GAGGTC-CGGGTTGGCC-G-CCCGGTCCTCGGCAGCTCTCCGTTGCTTGCTT-TGCTTACGCT

0062 GCTTAAGCCCAATAAAGCTTGCCCTGAGTGCTTCAAGTAGTGTGTGC-C-CGTCTGTTGTG-T
    |: | ||| | ||||| || | |:::| | |||||: || ||| | |:||| | || |
0061 G-ATCAGCCCAATAAAG--TG--TGGTAAAC-GCAAGTAACGTCTGCTCTCATCTGTTCTGCT
0123 GACTCTGGTAACTAGAGATCCCTCAGA-C-C-CTTTTAGT--CA--GTG-TGGAAATCTCTA-
    
```



0119 GGGTCA CCTCCCGAAGGTGCTCAGGGGAGCGGCC TT CAGTGGCGCCGAA CAGGGACTTGACTAG  
 0178 G-CAGTGGCGCC-CGAA CAGG-GA--CCTG--AAAGCGAAAGGGAAACC-AGA-GGAGC-TCTC  
 0183 GCCAGGGAC-CCGAGAGGAGGTGACGCCAGTCAAAGGGAGTCCGGATCCAAGAGGGATCGGCTG  
 0232 T-CGACGCAGGACTCGGCT--TGCTGAAGCGCGCACGGC-AAGAGGCGAGGGCGGC GACTGGT  
 0246 TCCGTC-CTAGGC-CGACCGGTGCTGGTG-GCTCGTGGTG TACACACGAGACGAGGTAAAGTAT  
 0292 GAGTACGCCAAAAATTTGACTAGCGGA-GGCTAGA-AGGAGAGAGATGGGTGC-G-AGAG-CG  
 0307 -CCAGCGCC-AGACTGCCAC-AGCAGTCCGGGAAATAGGGGA-CCATGGGAGCAGCGGGTCCG  
 0351 TCAGTATTAAGCGGG-GGAGAA TTAGATCGATGGAAAAAATTCGGTTAAGGCCAGGGGAAAG  
 0367 GCAATATTAA-CAGGTAGAGAGCTAGACAGATATGACAAAATTCGGCTCCGACCGAAAGGGAAG  
 0414 AAAAATATAAATTA AACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTC-GCAGTT-AA  
 0430 AAGCGCTACCTAGTTGACATTTAGTTGGGCAAAGAAGGAGTTGGACAGATTCGGCCTTTCCG  
 0476 TCC TGGCCTGT TAGAAACATCAGAAGGCTGTAGACAAATACTGGGACAG--CTACAACCATCC-  
 0494 ACC-AA-CTCATGGAGTCTAAAGAAGGCTGTGAAA AAATTCCT--CACAGTCCTACTTCCATTAG  
 0537 CTT CAGACAGGATCAGAAGAACTTAGAT--CATTAT---ATAA-TACAGTAGCA---ACC---C  
 0554 AAGCTAA-TGGGTCAGAAAATTTAAAAAGCCTTTTTGGCATAATTTCCGGTCGTATGGGCCGTTCC  
 0589 -T-CTATTGTGTGCATCAAAGG-A-T-AG--AG-ATAAAAGACA-CCAAGGAAGCTTTAGACAA  
 0617 ATGCCAAAAAGGAAGTAGAAGACACTGAGCAAGCAAACAGAAAGTAAAGAAGCATGTAATTG  
 0644 GATAGAGGA---A-GAGCAA-AAC-AAAAG--TAAGAAAAAGCACAG-C-AAGCA--GC-AGC  
 0681 GAAAGATGACCCACCAGCGACATCTGGTGGACAAAGTGAAAAATAGCAGTCAAACATGGCTAGT  
 0695 -TGACA-C-AG-GACACAGCAATCAGGTCAGCCAAAAT TACCC TATAGTGCAG-AACATCCAGG  
 0745 GAGACATCTAGTGGC-CAAAGGTAGTACAG-CAGGAAAA-ACAAAAG-GCAGCAACA-CCGCC  
 0754 GGCAAATGG-TAC-AT-CA-GGC-CATATCACCT-AGAA-CTTTAAATGCAT---GG-GTAAAA  
 0804 ACCTAGAGGA AACTATCCACTGCTCAGAAATCCTCAAATCAGTGGATTCATACAGGAGTCCCA  
 0807 G-TA-G----T--AGA-A---GAGAAGGCTTTCAGC-CC-AGAAG--TGATAC--CC--A-TG-  
 0868 GTTAGGACTCTAAAAACATGGGTGGAGGCTGTGAAC TCCAAAAAATTTGATGCATCCATAGTGC  
 0849 ---T-TTCA-GCATTATCAGAAGGAGCCACCCCAAGATTTAAACACCATGCTAAACACAGT  
 0932 CATTAATTT CAGGCTCTAACAGAAGGATTTATTCCTTATGATCTCAATGGCATGCTTAATGCTGT  
 0908 GGGGGACATCAAGCAGCCATGCAAATGTTAAAAAGACCATCAATGAGGAAGCTGCAGAAATGG  
 0996 AGGAGATCATCAGGGAGCCATGCAGGTAATTAAGACATCATCAATGAGGAAGGGCGGCATGG  
 0972 GATAGAGTGCATCCAGTGCATGCAG-GGCCTATTGCACCAGGCCAGA-TGAGAGAACCAAGGGG

```

    | | | :|:| | | | :|:| | | | | :| | | | :| | | | | | :| | : |
1060 GATCTTGAACATCCCCAAC-CGCAACAACCTGCTCCA-CAAGCAGGATTGAGGGATCCCTCCGC
1034 AAGTGACATAGCAGGAACTACTAGTACCCTT-CAGGAACAAATAGGATGGATGAC-AAATA-AT
    : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1122 CTCAGACATAGCAGGAACAACAAGCTCAATTGCT-GAGCAAATAGAATGGATTACTAGGCAGAA
1095 CCACCTA-TCCAGTAGGAGAAAATTATATAAAGATGGATAATCCTGGGATTAATAAAAATAGTA
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1185 CAACCCAGTACAAGTAGGAGAAAATATATAGAAGATGGATCATTCTAGGGCTGCAGAAGTGTGTG
1158 AGAATGTATAGCCCTACCAAGCATTC-TGGACATAAGACAAGGACCAAAGGAACCCCTTAGAGAC
    : : | | | | | | : : | : : | | : | | | | | | | | | | | | | | | | | | | | | | | | |
1249 CAGGTCTACAACCCTGTTT-CTATCTTAGACATAAAACAGGGCCCAAAAGAAATTTCAAAGAC
1221 TATGTAGACCGGTTCTA-TAAAACTCTAAGAGCCGAGCAAGC-TTCACAGGAGGTAAAAAATTG
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1312 TATGTAGACAGATTCTACCATTTGCT-TAAGAGCAGAGCAGGCGGACCATCA-GTCAAAGGGTG
1283 GATGACAGAAACCTTGTTGGTCCAAAATGCGAACCAGATTGTAAGACT-ATTTAAAAGCATT
    | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1374 GCTAACTCAAACACTGTTGATTCAAAATGCCAACCAGAAATG-CAGACAGATCCTCAAGGCAAT
1346 GGGACCAGCGGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTAGGAGGACCCGGCCAT
    | | : | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1437 -GGTCAAAC--CTACCTGGAGGAAATGCTGCAAGCTTGTCAAGGAGTCGGAGGACCCTTACAC
1410 AAGGCAAGAGTTTGGCTGAAGCAATGAGCCAAGTAAACA-ATTCAGCTACCATAATGATGCAG
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1498 AAGGCAAAACTCATGGCAGAGGCCATG-GTAATGGCCAGCAGTCAG-TA-AAT-ATGGTACA-
1473 AGAGGCAATTTAGGAACCAAAGAAAGATTCTTAAAGTGTTC AATTGT-GGCAAAG AAGGGCAC
    | | | | | | : | : | | : | | : | | : | | | | | | | | | | | | | | | | | | | | | | | | | |
1557 AG-GGCCATCAAAG----GGAGATCAATGATTAAATGCTACAATTGTGGGCAA-TTGGGCAC
1536 ACAGCCAGAA-ATTGC--AGGGCCCTAGGAA-AAAGGGCTGTTGGAAATGTGGAAAGG AAGGA
    | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1615 A-TGCAAAAAGACTGCAAAAGCCACTTAAAGCTAAATGCT-TT--AATTGTGGAAAAACAGGA
1596 CACCAATGAAAGATTGTA--CTG--AGAGACAGGCTAATTTTTTTAGGGAAGATCTGGCCTTCC
    | | | | | | : | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1675 CACCTAGCAAGGCCGTAGGCAACCCAGAGAAATCAAGGCCCTCCAGTTGCACAGGCTAATT
1656 TACAAGGGAAGGCCAGGGAAATTTCTTCAGAGCAGACCAGAGCCAACAGCCCCACCAG
    | : | | | | | | : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1739 TTTTAGGGAAGGATGGGG--GTCTCCAG-GAGGCCCCAGCCA ACTTCCAGTGAG
  
```

Με την παρακάτω εντολή παράγεται ο πίνακας (scoring matrix) που χρησιμοποιείται από τον Αλγόριθμο:

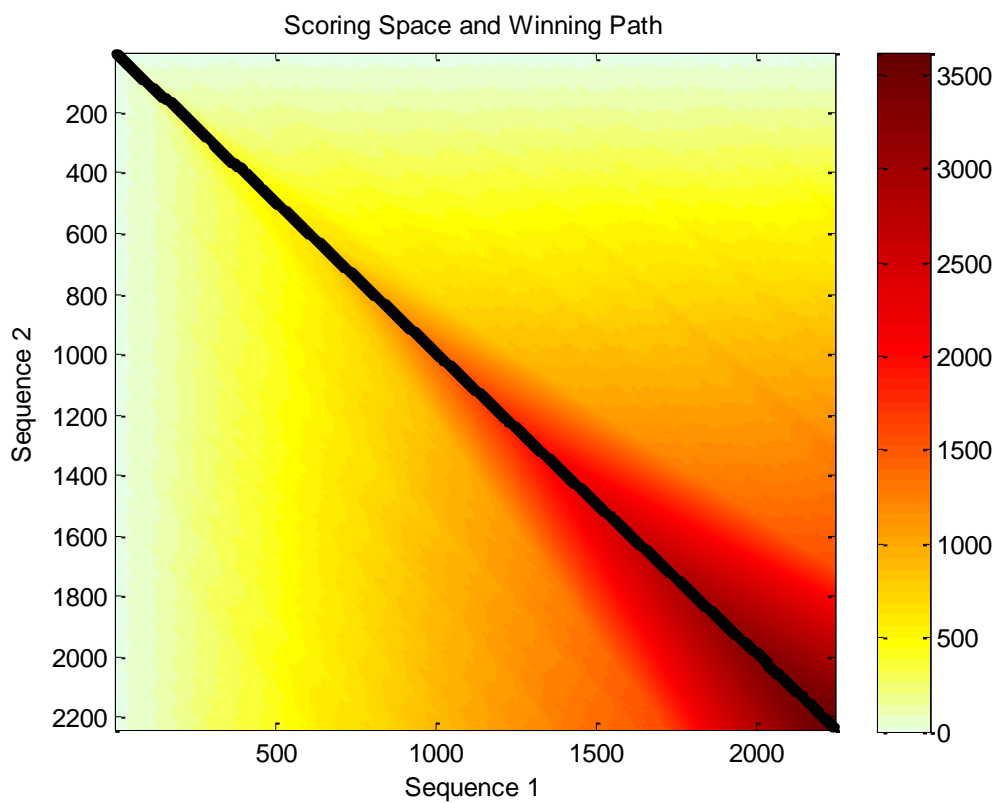
Matrix2 = blosum(90);

5	-2	-2	-3	-1	-1	-1	0	-2	-2	-2	-1	-2	-3	-1	1	0	-4	-3	-1	-2	-1	-1	-6
-2	6	-1	-3	-5	1	-1	-3	0	-4	-3	2	-2	-4	-3	-1	-2	-4	-3	-3	-2	0	-2	-6
-2	-1	7	1	-4	0	-1	-1	0	-4	-4	0	-3	-4	-3	0	0	-5	-3	-4	4	-1	-2	-6
-3	-3	1	7	-5	-1	1	-2	-2	-5	-5	-1	-4	-5	-3	-1	-2	-6	-4	-5	4	0	-2	-6
-1	-5	-4	-5	9	-4	-6	-4	-5	-2	-2	-4	-2	-3	-4	-2	-2	-4	-4	-2	-4	-5	-3	-6
-1	1	0	-1	-4	7	2	-3	1	-4	-3	1	0	-4	-2	-1	-1	-3	-3	-3	-1	4	-1	-6
-1	-1	-1	1	-6	2	6	-3	-1	-4	-4	0	-3	-5	-2	-1	-1	-5	-4	-3	0	4	-2	-6
0	-3	-1	-2	-4	-3	-3	6	-3	-5	-5	-2	-4	-5	-3	-1	-3	-4	-5	-5	-2	-3	-2	-6
-2	0	0	-2	-5	1	-1	-3	8	-4	-4	-1	-3	-2	-3	-2	-2	-3	1	-4	-1	0	-2	-6
-2	-4	-4	-5	-2	-4	-4	-5	-4	5	1	-4	1	-1	-4	-3	-1	-4	-2	3	-5	-4	-2	-6
-2	-3	-4	-5	-2	-3	-4	-5	-4	1	5	-3	2	0	-4	-3	-2	-3	-2	0	-5	-4	-2	-6
-1	2	0	-1	-4	1	0	-2	-1	-4	-3	6	-2	-4	-2	-1	-1	-5	-3	-3	-1	1	-1	-6
-2	-2	-3	-4	-2	0	-3	-4	-3	1	2	-2	7	-1	-3	-2	-1	-2	-2	0	-4	-2	-1	-6
-3	-4	-4	-5	-3	-4	-5	-5	-2	-1	0	-4	-1	7	-4	-3	-3	0	3	-2	-4	-4	-2	-6
-1	-3	-3	-3	-4	-2	-2	-3	-3	-4	-4	-2	-3	-4	8	-2	-2	-5	-4	-3	-3	-2	-2	-6
1	-1	0	-1	-2	-1	-1	-1	-2	-3	-3	-1	-2	-3	-2	5	1	-4	-3	-2	0	-1	-1	-6
0	-2	0	-2	-2	-1	-1	-3	-2	-1	-2	-1	-1	-3	-2	1	6	-4	-2	-1	-1	-1	-1	-6
-4	-4	-5	-6	-4	-3	-5	-4	-3	-4	-3	-5	-2	0	-5	-4	-4	11	2	-3	-6	-4	-3	-6
-3	-3	-3	-4	-4	-3	-4	-5	1	-2	-2	-3	-2	3	-4	-3	-2	2	8	-3	-4	-3	-2	-6
-1	-3	-4	-5	-2	-3	-3	-5	-4	3	0	-3	0	-2	-3	-2	-1	-3	-3	5	-4	-3	-2	-6
-2	-2	4	4	-4	-1	0	-2	-1	-5	-5	-1	-4	-4	-3	0	-1	-6	-4	-4	4	0	-2	-6
-1	0	-1	0	-5	4	4	-3	0	-4	-4	1	-2	-4	-2	-1	-1	-4	-3	-3	0	4	-1	-6
-1	-2	-2	-2	-3	-1	-2	-2	-2	-2	-2	-1	-1	-2	-2	-1	-1	-3	-2	-2	-2	-1	-2	-6
-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	1

Σύγκριση των ακολουθιών NC\_002549.1 και NC\_014372.1 με τον αλγόριθμο Smith - Waterman και παρουσίαση αποτελέσματος

Με την παρακάτω εντολή παράγεται το γράφημα του score της σύγκρισης.

```
[Score3, Alignment3, Start3] = swalign(E,F, 'Showscore', true);
```







1698 CTGCTGCGTCACTGCCAAAACAAGTGGACATTACGATGATGATGACGACATTCCTTTCCAGG  
 | | : : || | | | | : | | | : | | | | | | | | : | | | | | | | | : | |  
 1692 CCTCCACATCCCTTCTCAAGACAGGAAAACAGTATGATGATGACAACGATATCCCTTTCCCTGG  
 1762 ACCCATCAATGATGACGACAATCCTGGCCATCAAGATGATGATCCGACTGACTCACAGGATACG  
 : | | | | | | | | : | | | | | | | | : | | | | | | | | : | | | | | | | | : | |  
 1756 GCCCATCAATGATAACGAAAACCTCAGAACAGCAAGACGATGATCCAACAGATCTCAGGACACT  
 1826 ACCATTCCTGATGTGGTGGTTGATCCC-GATGATGGAAGCTACGGCGAATAC-CAGAGT-TACT  
 | | | | | | | | : | : | | | | | | | | | | | | | | | | | | | | | | | | | |  
 1820 ACCATCCCTGATATCATTTGTTGA-CCCGATGATGGCAGATACAACAATTATGGAGACTATCCT  
  
 1887 CGGAAAACGGCATGAATGCACCAGATGACTTGGTCCTATTCGATCTAGACGAGGACGACGAGGA  
 | : : | | | | | | | | | | | | : : | | | | | | | | | | : | | | | | | | | : | | | | | | | |  
 1883 AGTGAGACGGC--GAATGCCCTGAAGACCTTGTCTTTTGACCTTGAAGATGGTGCAGGAGGA  
 1951 CACTAAGCCAGTGCCTAATAGATCGACCAAGGGTGGACAACAGAAGAACAGTCAAA-AGGGCCA  
 : | | | | | | : | | : : | | : | | | | : : : | | | | | | | | | | : | | | : | | : |  
 1945 TGATCA-CC-G-ACCGTCAAGTTC-ATCAGAGAACACAACA-AA-CACAGTCTTACAGGA ACT  
 2014 G-CATATAGAGGGC-AG--AC-AGACAC-AATCCAGGCCAAT-T-CAAAATGTCCCAGGC-CCT  
 | | | | : | | : : | | | | | | : | | : | | | | | | | | | | : | | | | | | : | | | | | |  
 2003 GACAGTAACAAAACAAGTAACTGGAATCGAAACCCGACTAATATGCCAAA-GA-AAGACTCCA  
  
 2069 CACAGAACAATCCACCACGCCAGTGC-GC-CACTCACGGACAATGACAGAAGAAATGAACCCCTC  
 | | | | : | | | | | | | | | | | | : | | : | | | | | | : | | : | | | | | | : | |  
 2064 CACAAAACAAT-GACAAAT-CCTGCACAGCGGGCTCA-AGAATACGCCAG-GGATAACATCCAGG  
 2131 CGGCTCAACCAGCCC-TCGCATGCTGACACCAATTAACGAAGAGGCAGAC-CCACTGG--ACGA  
 : | | | | | | | | | | : | | : | | | | | | | | | | | | | | : | | | | | | | | : | |  
 2124 ATACACCAACACCCCATCGAGCTCTAACTCCCATCAGCGAAGAAACCGGCTCCAATGGTCACAA  
 2191 TGCCGACGACGAGACGTCTAGCCTTCCGCCCTTGGAGTCAGATGATGAA  
 | | | | | | | | : | | : | | | | | | | | | | | | | | | | | | | | | | | | | |  
 2188 TGAAGATGAC-A-TTGA-TAGCATCCCTCCTTTGGAATCAGACGAAGAA



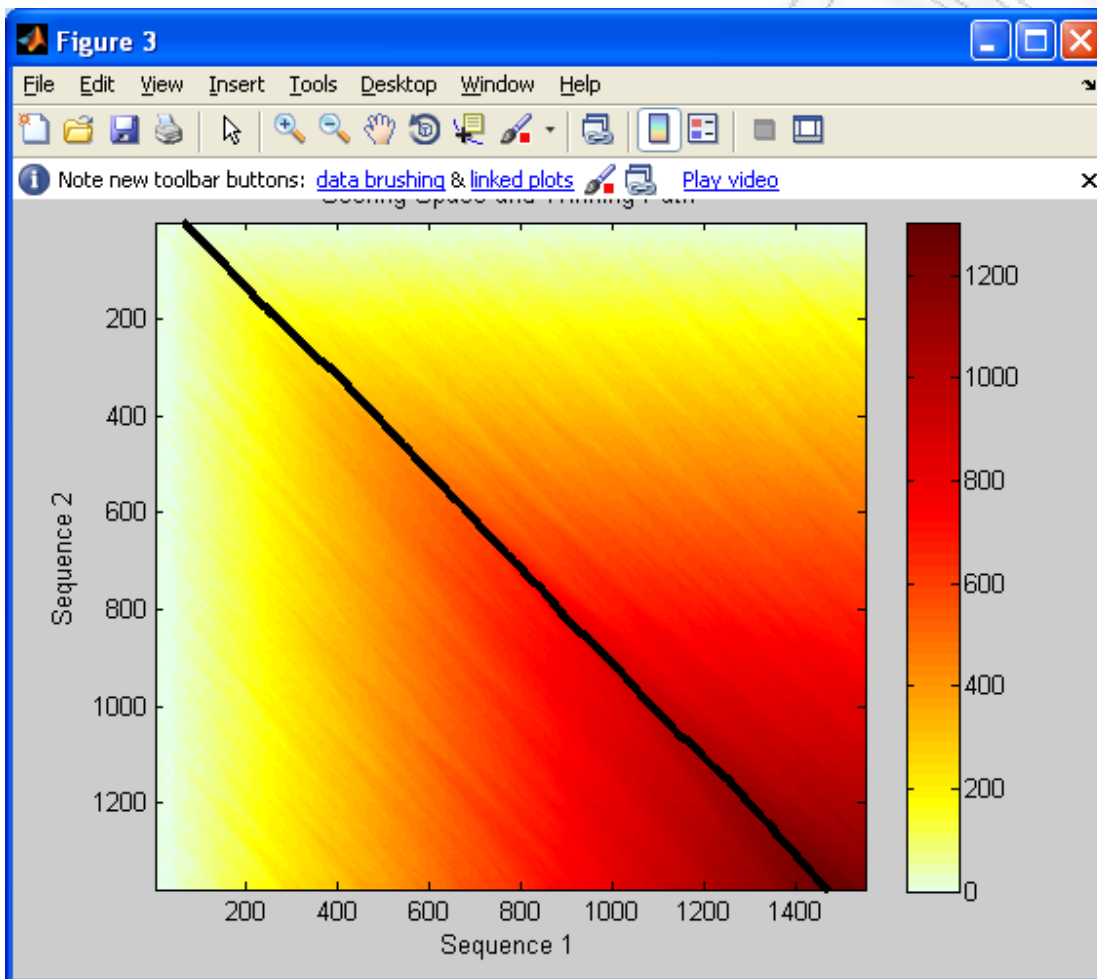
Με την παρακάτω εντολή παράγεται ο πίνακας (scoring matrix) που χρησιμοποιείται από τον Αλγόριθμο:

```
Matrix3 = blosum(90);
```

5	-2	-2	-3	-1	-1	-1	0	-2	-2	-2	-1	-2	-3	-1	1	0	-4	-3	-1	-2	-1	-1	-6
-2	6	-1	-3	-5	1	-1	-3	0	-4	-3	2	-2	-4	-3	-1	-2	-4	-3	-3	-2	0	-2	-6
-2	-1	7	1	-4	0	-1	-1	0	-4	-4	0	-3	-4	-3	0	0	-5	-3	-4	4	-1	-2	-6
-3	-3	1	7	-5	-1	1	-2	-2	-5	-5	-1	-4	-5	-3	-1	-2	-6	-4	-5	4	0	-2	-6
-1	-5	-4	-5	9	-4	-6	-4	-5	-2	-2	-4	-2	-3	-4	-2	-2	-4	-4	-2	-4	-5	-3	-6
-1	1	0	-1	-4	7	2	-3	1	-4	-3	1	0	-4	-2	-1	-1	-3	-3	-3	-1	4	-1	-6
-1	-1	-1	1	-6	2	6	-3	-1	-4	-4	0	-3	-5	-2	-1	-1	-5	-4	-3	0	4	-2	-6
0	-3	-1	-2	-4	-3	-3	6	-3	-5	-5	-2	-4	-5	-3	-1	-3	-4	-5	-5	-2	-3	-2	-6
-2	0	0	-2	-5	1	-1	-3	8	-4	-4	-1	-3	-2	-3	-2	-2	-3	1	-4	-1	0	-2	-6
-2	-4	-4	-5	-2	-4	-4	-5	-4	5	1	-4	1	-1	-4	-3	-1	-4	-2	3	-5	-4	-2	-6
-2	-3	-4	-5	-2	-3	-4	-5	-4	1	5	-3	2	0	-4	-3	-2	-3	-2	0	-5	-4	-2	-6
-1	2	0	-1	-4	1	0	-2	-1	-4	-3	6	-2	-4	-2	-1	-1	-5	-3	-3	-1	1	-1	-6
-2	-2	-3	-4	-2	0	-3	-4	-3	1	2	-2	7	-1	-3	-2	-1	-2	-2	0	-4	-2	-1	-6
-3	-4	-4	-5	-3	-4	-5	-5	-2	-1	0	-4	-1	7	-4	-3	-3	0	3	-2	-4	-4	-2	-6
-1	-3	-3	-3	-4	-2	-2	-3	-3	-4	-4	-2	-3	-4	8	-2	-2	-5	-4	-3	-3	-2	-2	-6
1	-1	0	-1	-2	-1	-1	-1	-2	-3	-3	-1	-2	-3	-2	5	1	-4	-3	-2	0	-1	-1	-6
0	-2	0	-2	-2	-1	-1	-3	-2	-1	-2	-1	-1	-3	-2	1	6	-4	-2	-1	-1	-1	-1	-6
-4	-4	-5	-6	-4	-3	-5	-4	-3	-4	-3	-5	-2	0	-5	-4	-4	11	2	-3	-6	-4	-3	-6
-3	-3	-3	-4	-4	-3	-4	-5	1	-2	-2	-3	-2	3	-4	-3	-2	2	8	-3	-4	-3	-2	-6
-1	-3	-4	-5	-2	-3	-3	-5	-4	3	0	-3	0	-2	-3	-2	-1	-3	-3	5	-4	-3	-2	-6
-2	-2	4	4	-4	-1	0	-2	-1	-5	-5	-1	-4	-4	-3	0	-1	-6	-4	-4	4	0	-2	-6
-1	0	-1	0	-5	4	4	-3	0	-4	-4	1	-2	-4	-2	-1	-1	-4	-3	-3	0	4	-1	-6
-1	-2	-2	-2	-3	-1	-2	-2	-2	-2	-2	-1	-1	-2	-2	-1	-1	-3	-2	-2	-2	-1	-2	-6
-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	1

Σύγκριση των ακολουθιών M54967.2 και AY555151.3 με τον αλγόριθμο Smith - Waterman και παρουσίαση του αποτελέσματος.

Με την παρακάτω εντολή παράγεται το γράφημα του score της σύγκρισης:  
`[Score4, Alignment4, Start4] = swalign(G,H, 'Showscore', true);`



Με την παρακάτω εντολή παράγεται το αποτέλεσμα του alignment των ακολουθιών:

```

showalignment (Alignment4);
Identities = 745/1453 (51%), Positives = 1039/1453 (72%)
0001 AACTATA-CAAAACGTTAACCCATTTCACATCAGGGGGAGTGTATTATC-ACTATATGTGT
    |:|:|:| |:|:| |:| | | | | |:|:|:| |:|:|:| |:| | | |:| |:| |:| |:|
0001 AGCAAAAGCAGGAGTTCAAAATGAATC-CAAAT-AAGAAGATAATAACCATCGGATCAATCTGT
0063 CAGCTTCACTGTCATACTTA-CTGTATTCCGGATATATTGCTAAAATTTCATCAA-AAATAACT
    :| | | | | |:| |:| | | |:| | | |:| | | |:| |:| |:| |:| |:| |:|
0063 ATG-GTAACTGGAATGTTAGCT-TAATGTTACAAATGGGAAC TTGATC-TCAATATGGGTCA
0125 GCACCAACAATG-ACAT TGGATTGC-GCG-AACGCATCAAATGTTCAAGGCTGTGAACCGCTCTG
    | :| :|:| | | |:|:|:|:| | :|:| | | | | |:| | | |:| | | |:| | |
0124 G-TCATTC AATTCACACAGGGAA TCAACA CAAGC-TGAACC AATCA-GCAAT--ACTAATTTT
0186 C-AACAAAAGAGATGACATTTCTTCTCCAGAACCGGAGTGGACATACCCTCGT TTA TCTTGCC
    | :| | :|:| |:| | | | | | | | | | | |:|:|:|:| | | | |:| | | | |
0183 CTTAC-TGAGA-AAG-C-TGT-GGCTT CAGTAA AATTAGCGGGCAATTCATCTCTTTGC-C-CC
0249 AGGGCTCAACCTTT CAGAAAGCACTCCTAATTAGCCCTCATAGATTCGGAGAAGCCAGAGGAAA
    | :| | :| |:|:|:|:| | | | | | | | | | |:|:|:|:|:| | |:| |:|:|:|:|
0240 ATTAATGGA-TGGGCTGTATACAGTAAGGA-CAACAGTATAAGGATC-G-GTT-CCAAGGGGGA
0313 CTCAGCTCCCTTAATAATAAGGGAACCTTTTATTGCTTGTGGACCA-AAGGAATGCAAACACTT
    | | | | |:|:|:|:|:|:|:|:| | | | | | | | | | |:| | | |:|:|:|:|:| | | |
0299 -T--G-T--GTTTGTATAAGAGAGCCATT CATCTCATGC-TCCCAC TTGGAATGCAGA-ACTT
0376 T-GCTCTAACCCAT TATGCAGCTCAA CCGGGAGGGTACTACAATGGAACAAGAGAGGACAGAAA
    | | | |:| | | |:| | | | | | | | | | | | | | |:|:|:|:|:|:| |:|:|:|:|:|
0355 TCTTTT T GACTCAGG GAGCCTTGCTGAATGCAAGCACTCCAATGGGACTGTC AAAGACAGAAG
0439 CAAGCTGAG-GCATCTGAT-TTCAGTCAAATTAGGCAAAATCCCAACTGTAGAAA ACTCCATTT
    | | | | |:| | | |:| | | |:| | | |:| | | |:| | | |:| | | |:| | | |:| | |
0419 CCCTCACAGAACAT-TAATGAGTGTCCGTGGGTGAGGCTCCCTCCC-CATATAACTCAAGGT
0501 TCCACATGGCAGCTTGGAGTGGGTCTGCATGCCATGATGGTAGAGAATGGACATATATCGGAGT
    | | | | |:| | | |:| | | |:| | | |:| | | |:| | | |:| | | |:| | | |:| | |
0481 TTGAGTCTGTGCTTGGTCAGCAA GTGCTTGCCATGATGGCACCAGTTGGTTGACAATTGGAAT
0565 TGATGGCCCTGACAGTAATGCATTGATCAAAATAAAATATGGAGAAGCAT-AT-ACTGACACAT
    | | | | |:| | | |:| | | |:| | | |:| | | |:| | | |:| | | |:| | | |:| | |
0545 TTCTGGCC CAGACAATGGGGCTGTGG-CTGTATTGAA-ATACAATGGCATAATAACAGACAC-T
0627 ACCATTCCTATGC-AAACAACATCCTAAGAACAAGAAAGTGCCCTGC AATTGCATCGGGGGAG
    | | | | |:| | | |:| | | |:| | | |:| | | |:| | | |:| | | |:| | | |:| | |
0606 ATCAAAGAGTTGGAGGAATAACATACTGAGAACTCAAGAGTCTGAATGTGCATGTGTAAATGGCT
0690 ATTGTTATCTTATGATAACTGATGGCTCAGCTTCAGGAATTAGTAAATGCA-GATTTCTTAAGA
    | | | |:| | | |:| | | |:| | | |:| | | |:| | | |:| | | |:| | | |:| | |
0670 CTTGCTT TACTGTAATGACTGACGGACCAAGTAATGG--TCAGGCATCACATAAGATCTTCAA
0753 TTTCAGAGGGTTCGAATAATAAAA GAAATATTTCACACAGGAAG-AGTAGGCAT-AC--TGAAG
    :| | | |:|:|:| | | | |:|:|:|:| | | | | | | | | | | |:| | | |:| | | |:| |

```

0732 ATGGAAAAAGG-GAA-AGTGGTT-AAATCAGTCGAATGGATGCTCCC AATTTATCCTATGAGG  
 0813 AATGCACATGCGGA--T-TTGCCAGCAATAAAACCATAGAATGTGCCTGTAGAGATAACAGTTA  
 |||||:| || :: | :|||:|:|:| | || | :||| | | | :|:| | | :|  
 0793 AATGCTCCTGTTATCCTGATGCCGCGCAAATCA-CAT-GTGTGTGCAGGATAAATTGGCA-TG  
  
 0874 CACAGCAAAAAGACCC TTTGTC A AATTA AATGTGGAGACTGATACAGCTGAGATAAGATTGATG  
 :|: ||||: |:|| | || : :|| ||| ::||: |||:| | | | :|||:|:|:| | :|  
 0853 -GCT-CAAAATCGGCCATGGGTATCTTTCAAT-CAAAATTTGGAGTATC--AAATAGGATAATATA  
 0938 TGCACAGAGACTTATTTG-GACACCCCCAGACCAGATGATGGAAAGCA--TAAACAG-GGCCTTGC  
 |||| :| || ||:| | |||| || | || :||| |||| | | ||| :| | | | | | |  
 0912 TGCAGTG-GAGTTTTGGAGACAATCCACGCCCAATGATGGAA-CAGGTAGTTGTGGTCCGGT  
 0998 GA-ATCTAATGGGGACAAAAGGGCGT-GGAGGCATCAAGGGAGGATTTGTCCATCAAAGATGGC  
 |: |||| | || | ||:| || || :||| :| : : :|||:| | | | :||:| | |  
 0974 GTCCCTCTAACGGGG-CATATGGGGTAAAAGGTTTTTCATTFAAATACGGCAAT-GGTGTCTGG-  
 1060 ATC-CAAGATTTGGAAGATGGTACTCCCGAACGATGCTCTAAACTGAGAGAATGGGGATGGAAC  
 ||| :|||: :|||: : || | || | : || | | :||| :|| | |||: : |||:|  
 1035 ATCGGGAGAACAAAAAGCACTAATTCC-AG-GA-G-C---GGCTTGA-AATGATTTGGGATCC  
 1123 GTATGTCAGGTATGATGGAGACCCATGGACTGACAGTGACGCCCTTGCTCATAGTGGAGTAATG  
 :||| :||: |||: : || | :|| | :||| | : |||| | : ||| | || | | || | |  
 1091 AAATGGGTGGA CTGAAA CCGA-CAGTAGCTTTTCAGTGA-A-ACAAGAT-ATCGT--AGCAAT-  
 1187 GTCTCAATGAA-AGAACC TGGTTGGTATTCT-TTTGGCTT-C-GAAATAA-AAGATAAGAAATG  
 :|| | :||: ||:| |:| ||:| | | :||:| | || | :| | :|| :|| :||  
 1148 AACTGATTTGGTCAGGATATAGCGGGAGTTTTGTCCAGCATCCAGA ACTGACAGGACTAG-ATTG  
 1246 TGATGTCCCCTGTATTTGGATAGAGATGTTACACGATGGTGG---AAA-AAAG-AC--TTGGCA  
 :||: || |||:| ||:|:|:|:| | | | | | | || | :| | | | | | |  
 1211 CATAAAGACCTTGTCTTCTGGSTTGAGTTGAT-CA-GAGGGCGGCCCAAAGAGAGCACAAATTTGGA  
  
 1303 CTCAGCAGCAACAGCCATTTACTGTTTGATG-G-GCTCAG-GACAAT-TGCTGTGGGACACTGT  
 || || :| | :|| | ||:| || ||||: | | : : || |||| | || | | ||:|: |  
 1273 CT-AGTGGGAGCAG-CATAT-CTTTTGTGGTGTAAATAGTGACACTGTG-GGTTGGTCTTGGC  
 1363 CACA-GGTGTTGATATGGCTCT-GTAATGGAGGAATGGTTGAATC  
 || | |||| | ||:| | :| : || || :||:| ||| :||  
 1333 CAGACGGTGCTGAATTGCCATTCACCATTGACAAGTAGTTG-TTC

Με την παρακάτω εντολή παράγεται ο πίνακας (scoring matrix) που χρησιμοποιείται από τον Αλγόριθμο:

```
Matrix4 = blosum(90);
```

5	-2	-2	-3	-1	-1	-1	0	-2	-2	-2	-1	-2	-3	-1	1	0	-4	-3	-1	-2	-1	-1	-6
-2	6	-1	-3	-5	1	-1	-3	0	-4	-3	2	-2	-4	-3	-1	-2	-4	-3	-3	-2	0	-2	-6
-2	-1	7	1	-4	0	-1	-1	0	-4	-4	0	-3	-4	-3	0	0	-5	-3	-4	4	-1	-2	-6
-3	-3	1	7	-5	-1	1	-2	-2	-5	-5	-1	-4	-5	-3	-1	-2	-6	-4	-5	4	0	-2	-6
-1	-5	-4	-5	9	-4	-6	-4	-5	-2	-2	-4	-2	-3	-4	-2	-2	-4	-4	-2	-4	-5	-3	-6
-1	1	0	-1	-4	7	2	-3	1	-4	-3	1	0	-4	-2	-1	-1	-3	-3	-3	-1	4	-1	-6
-1	-1	-1	1	-6	2	6	-3	-1	-4	-4	0	-3	-5	-2	-1	-1	-5	-4	-3	0	4	-2	-6
0	-3	-1	-2	-4	-3	-3	6	-3	-5	-5	-2	-4	-5	-3	-1	-3	-4	-5	-5	-2	-3	-2	-6
-2	0	0	-2	-5	1	-1	-3	8	-4	-4	-1	-3	-2	-3	-2	-2	-3	1	-4	-1	0	-2	-6
-2	-4	-4	-5	-2	-4	-4	-5	-4	5	1	-4	1	-1	-4	-3	-1	-4	-2	3	-5	-4	-2	-6
-2	-3	-4	-5	-2	-3	-4	-5	-4	1	5	-3	2	0	-4	-3	-2	-3	-2	0	-5	-4	-2	-6
-1	2	0	-1	-4	1	0	-2	-1	-4	-3	6	-2	-4	-2	-1	-1	-5	-3	-3	-1	1	-1	-6
-2	-2	-3	-4	-2	0	-3	-4	-3	1	2	-2	7	-1	-3	-2	-1	-2	-2	0	-4	-2	-1	-6
-3	-4	-4	-5	-3	-4	-5	-5	-2	-1	0	-4	-1	7	-4	-3	-3	0	3	-2	-4	-4	-2	-6
-1	-3	-3	-3	-4	-2	-2	-3	-3	-4	-4	-2	-3	-4	8	-2	-2	-5	-4	-3	-3	-2	-2	-6
1	-1	0	-1	-2	-1	-1	-1	-2	-3	-3	-1	-2	-3	-2	5	1	-4	-3	-2	0	-1	-1	-6
0	-2	0	-2	-2	-1	-1	-3	-2	-1	-2	-1	-1	-3	-2	1	6	-4	-2	-1	-1	-1	-1	-6
-4	-4	-5	-6	-4	-3	-5	-4	-3	-4	-3	-5	-2	0	-5	-4	-4	11	2	-3	-6	-4	-3	-6
-3	-3	-3	-4	-4	-3	-4	-5	1	-2	-2	-3	-2	3	-4	-3	-2	2	8	-3	-4	-3	-2	-6
-1	-3	-4	-5	-2	-3	-3	-5	-4	3	0	-3	0	-2	-3	-2	-1	-3	-3	5	-4	-3	-2	-6
-2	-2	4	4	-4	-1	0	-2	-1	-5	-5	-1	-4	-4	-3	0	-1	-6	-4	-4	4	0	-2	-6
-1	0	-1	0	-5	4	4	-3	0	-4	-4	1	-2	-4	-2	-1	-1	-4	-3	-3	0	4	-1	-6
-1	-2	-2	-2	-3	-1	-2	-2	-2	-2	-2	-1	-1	-2	-2	-1	-1	-3	-2	-2	-2	-1	-2	-6
-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	-6	1

## ΣΥΜΠΕΡΑΣΜΑΤΑ:

Οι παρακάτω συγκρίσεις έχουν γίνει στο επίπεδο του γενετικού υλικού (νουκλεοτιδίων) και όχι στο επίπεδο των πρωτεϊνών που προκύπτουν από το γενετικό υλικό. Ο λόγος είναι ο εξής: Μικρές αλλαγές στο γενετικό υλικό ενδέχεται να οδηγήσουν σε λανθασμένα συμπεράσματα, καθώς εντελώς διαφορετικές πρωτεΐνες θα δημιουργηθούν. Για τον παραπάνω λόγο κρίθηκε σημαντικό η μελέτη να γίνει σε πρωτογενές υλικό. Ως πρωτογενές υλικό ορίζω το γενετικό υλικό. Επίσης κατά την διάρκεια της μελέτης δεν ήταν εφικτό για κάποιες ακολουθίες να τρέξουν λόγω μεγάλου μεγέθους αριθμού συμβολοσειρών στο matlab. Σε αυτές τις περιπτώσεις επιλέχθηκε η σύγκριση να γίνει σε μέρος των συμβολοσειρών. Μια συμβολοσειρά είναι μια ακολουθία χαρακτήρων, οι οποίοι ανήκουν σε ένα αλφάβητο. Οι τέσσερις χαρακτήρες (A,C,G,T), είναι τμήματα γενετικών ακολουθιών.

Κατά την ευθυγράμμιση των ακολουθιών οι ομοιότητες καταγράφονται με χρώμα κόκκινο και εμφανίζονται με κάθετη κόκκινη γραμμή ( | ). Τα αποτελέσματα που δεν είναι απολύτως όμοια, αλλά θα μπορούσαν να ληφθούν ως θετικά με καταγράφονται με ροζ διαλυτικά ( : ).

Σε πολύ μεγάλες συμβολοσειρές εμφανιζόταν σφάλμα που δημιουργείται από ανεπάρκεια μνήμης του υπολογιστή. Οι προσπάθειες για εκτέλεση του κώδικα έγιναν τόσο σε διπύρηνους, όσο και σε τετραπύρηνους υπολογιστές με τουλάχιστον 4GB μνήμης RAM, με παρόμοια αποτελέσματα. Γίνεται εμφανές πως μερικοί αλγόριθμοι απαιτούν περισσότερη μνήμη. (28) Ίσως η χρήση ισχυρότερων υπολογιστών να δώσει λύση στο πρόβλημα. Σε αντίθεση, η χρησιμοποίηση του BLAST δεν εμφάνισε κάποιο πρόβλημα, ανεξάρτητα του μεγέθους των συμβολοσειρών.

Η επιλογή των ιών έγινε με κριτήρια γεωγραφικά, χρονικά, εξελικτικά (Εξέλιξη είναι η διαδικασία μέσω της φυσικής επιλογής κατά την οποία οι οργανισμοί αποκτούν χαρακτηριστικά τα οποία ενίοτε τους παρέχουν πλεονέκτημα στην επιβίωσή τους, με την προϋπόθεση ότι το περιβάλλον τους παραμένει σχετικά αμετάβλητο). Τα παρακάτω παραδείγματα είναι ένας δείκτης της μοναδικής ικανότητας των ιών να μεταδίδονται ξεπερνώντας τα διάφορα γεωγραφικά εμπόδια.

Εμπόδια που σε άλλες εποχές θα ήταν πρακτικά αδύνατο να ξεπεραστούν. Σήμερα, όμως με την εκπληκτική βελτίωση των συγκοινωνιών οποιοσδήποτε μπορεί να μεταφερθεί σε λίγες ώρες στην άλλη άκρη του πλανήτη και μαζί του τα μικρόβια στα οποία είναι φορέας. Γίνεται επίσης εμφανές πως ο χρονικός ορίζοντας δεν αποτελεί εμπόδιο στην μετάδοση ιών. Κάποιοι μπορούν κάλλιστα να επανεμφανιστούν ακόμα και μετά από αρκετές δεκαετίες και μάλιστα σε εντελώς απομακρυσμένες περιοχές καταφέρνοντας να ανασυνδιάσουν και ικανότητες τις οποίες δεν είχαν αρχικά. Ίσως λοιπόν ο μεγαλύτερος κίνδυνος επιβίωσης της ανθρωπότητας, αν καταφέρει να ξεπεράσει τις τάσεις αυτοκαταστροφής που διαθέτει μέσω κάποιου πυρηνικού ολοκαυτώματος ή ενός νέου παγκόσμιου πολέμου είναι σίγουρα οι ιοί, οι οποίοι άμεσα μέσω κάποιας πανδημίας, είτε έμμεσα (π.χ. AIDS), θα απειλούν πάντα την

ύπαρξή μας. Με σωστή μελέτη, και κυρίως πρόληψη ίσως να μπορούμε να κερδίσουμε την μάχη. Για παράδειγμα ο ιός της ευλογιάς εξαφανίστηκε για πάντα και υπάρχει μόνο σε εργαστήρια αναφοράς. Αν καταφέρουμε να ξεπεράσουμε τις τάσεις αυτοκαταστροφής μας, ίσως καταφέρουμε να γίνουμε πράγματι ένας υπερπολιτισμός.

### **Σύγκριση του στελέχους M81581.2 με το στέλεχος DQ266097.1**

Για την σύγκριση των ακολουθιών του ιού της γρίπης του στελέχους του Leningrad/1957 και του Ιού του τελέχους των ΗΠΑ (USA/2005) τα αποτελέσματα από το BLAST δείχνουν ομοιότητα 87%. Τα αντίστοιχα αποτελέσματα από το MATLAB με χρησιμοποίηση του Smith and Waterman δείχνουν ομοιότητα επίσης 87% με τα θετικά να ανέχονται στο 94%. Θετικές εκλαμβάνονται οι αντιστοιχίες που μπορεί να μην είναι εντελώς όμοιες, αλλά δεν επηρεάζουν σημαντικά την έκβαση του αποτελέσματος.

Το παραπάνω ζεύγος σύγκρισης αποδεικνύει ότι στελέχη ιών με πολύ μεγάλες ομοιότητες και συνεπώς παρόμοια συμπεριφορά μπορούν να εμφανιστούν σε εντελώς διαφορετικές περιοχές ακόμη και μετά από 50 έτη. Αυτό σημαίνει ότι στελέχη ιών που είχαν προκαλέσει στο παρελθόν πανδημίες μπορούν να ξαναεμφανιστούν αρκετά χρόνια αργότερα διατηρώντας όμως την ικανότητα να προκαλέσουν αυξημένη θνησιμότητα –αν όχι και μεγαλύτερη.

Χαρακτηριστικό είναι το γεγονός της Ισπανικής γρίπης που προκάλεσε την απώλεια περισσότερων από 25,000,000 ψυχών στον κόσμο στις αρχές του προηγούμενου αιώνα.

### **Σύγκριση του στελέχους NC001802.1 με το στέλεχος L06042.1**

Στο επόμενο παράδειγμα γίνεται σύγκριση μεταξύ του ιού HIV1 – του ιού που είναι υπεύθυνος για το AIDS ( Σύνδρομο Επίκτητης Ανοσολογικής Ανεπάρκειας) και του αντίστοιχου ιού SIV1 – του ιού που προσβάλλει τα πιθηκοειδή. Πρέπει να τονιστεί πως ο ιός δεν μπορεί να μεταδοθεί από τον πίθηκο στον άνθρωπο με κανένα μέσο μετάδοσης –ακόμη και με το αίμα. Με δεδομένο ότι ο ιός SIV1 προϋπήρχε του αντίστοιχου HIV1, πρέπει να οδήγησε με κάποιο τρόπο στην δημιουργία του HIV1. αν και δεν έχουν δοθεί αποδείξεις από την επιστημονική κοινότητα, υπάρχει η θεωρία ότι με κάποιο τρόπο ο ιός HIV προέκυψε από τον SIV1. Για ένα μεγάλο χρονικό διάστημα τοπικές φυλές της Αφρικής είχαν κατηγορηθεί ότι πραγματοποίησαν τελετές στις οποίες αναμείγνυαν αίμα από πίθηκο και άνθρωπο οδήγησαν στην δημιουργία του HIV1. Ο ιός όμως δεν μπορεί να μεταδοθεί παρά μόνο με σωματικά υγρά και όχι δια της πεπτικής οδού. Προσθέτωντας στα παραπάνω ‘περίεργα’ ερωτήματα στα μέσα τις δεκαετίας του 1980 συκλόνησαν το Ευρωκοινοβούλιο. Σύμφωνα με αυτά ο ιός μεταλλάχτηκε με την συνδρομή επιστημόνων σε χώρα της Νότιας Αφρικής με σκοπό να χρησιμοποιηθεί ως βιολογικό όπλο. Τα σημερινά πάντως δεδομένα για την εξάπλωση του ιού στην Αφρική είναι απογοητευτικά, καθώς σε αρκετές χώρες τις κεντρικής Αφρικής το ποσοστό μόλυνσης του πληθυσμού αγγίζει σχεδόν το 50%.

Χρησιμοποιώντας το BLAST, η ομοιότητα μεταξύ HIV1 και SIV1 αγγίζει το 66%. Με την χρησιμοποίηση του MATLAB είναι 58% με τα θετικά όμως να ανέρχονται στο 75%.

### **Σύγκριση του στελέχους NC002549.1 με το στέλεχος NC014373.1**

Επίσης γίνεται σύγκριση μεταξύ διαφορετικών στελεχών του ιού ebola σε δυο διαφορετικές χώρες της Αφρικής με διαφορά κάθε επιδημίας 11 ετών. Η σύγκριση με το BLAST δείχνει ομοιότητα 77%, ενώ με το MATLAB η ομοιότητα είναι 70%. Τα παραπάνω αποτελέσματα αποδεικνύουν την ανθεκτικότητα των ιών και την ικανότητά τους να εμφανιστούν σε διαφορετικές γεωγραφικές τοποθεσίες ακόμη και μετά από αρκετά χρόνια. Επίσης εμπορικοί, ή και μεταναστευτικοί λόγοι συντελούν στην μετάδοσή τους. Το Ζαίρ από την Ακτή Ελεφαντοστού είναι σχετικά κοντά. Επίσης για το συγκεκριμένο παράδειγμα έγινε και σύγκριση με το BLAST, αλλά χρησιμοποιώντας τον αλγόριθμο Needleman-Wunsch με ομοιότητα 67%. (παράρτημα).

### **Σύγκριση του στελέχους M54967.2 με το στέλεχος AY555151.3**



Στο επόμενο παράδειγμα γίνεται σύγκριση μεταξύ του ιού της γρίπης τύπου Β, που είχε εμφανιστεί στο Πεκίνο το 1987 και του ιού της γρίπης τύπου Α, που είχε εμφανιστεί στην Ταϊλάνδη το 2004. Συνήθως οι συγκρίσεις πρέπει να γίνονται με σχετικά συγγενείς ιούς. Στο συγκεκριμένο παράδειγμα όμως ο σκοπός είναι να γίνει εμφανής η ομοιότητα μεταξύ δυο διαφορετικών στελεχών ιών που απασχόλησαν την επιστημονική κοινότητα. Πιθανές αλληλεπιδράσεις μεταξύ ακόμα και μεταξύ εντελώς διαφορετικών ιών όχι μόνο δεν πρέπει να αποκλείονται, αλλά αποτελούν και τον εφιάλη της παγκόσμιας επιστημονικής κοινότητας.

Τα παραπάνω παραδείγματα επιλέχτηκαν ακριβώς επειδή σύμφωνα πάντα και με τη διεθνή βιβλιογραφία πιθανή μετάλλαξη του ιού της γρίπης με τον ιό ebola, θα μπορούσε θεωρητικά πάντα να οδηγήσει σε ένα καινούριο ιό, οποίος αν είχε το ποσοστό θνησιμότητας του ebola, που συνήθως είναι πάνω από 50% και την ευκολία μετάδοσης των περισσοτέρων ιών της γρίπης, τότε σοβαροί κίνδυνοι θα εμφανίζονταν για τη δημόσια υγεία.

Σήμερα ευτυχώς η μετάδοση του ιού ebola δεν είναι καθόλου εύκολη και μπορεί να περιοριστεί αποτελεσματικά από τις αρχές, ενώ ο ιός της γρίπης συνήθως δεν έχει μεγάλη θνησιμότητα σε υγιή άτομα.

## **BIBΛΙΟΓΡΑΦΙΑ:**

### **Identification of Common Molecular subsequences,**

T.F. Smith and M.S. Waterman,  
*J. Mol. Biol.* (1981) 147, pp: 195-197.

**Βιοχημεία Στην Ιατρική**, ΒΙΟΜΟΡΙΑ τόμος1, Αμαλία Γ.Διονυσίου – Αστερίου, Κωνσταντίνος Τρούγκος. Ιατρικές Εκδόσεις Π.Χ. Πασχαλίδης. pp: 157-175. 2006

**An Introduction To Genetic Analysis**, Anthony J.F. Griffiths, Jeffrey H. Miller, David T. Suzuki, Richard C. Lewontin, William M. Gelbart. W.H. Freeman and Co. New York pp: 313-422. 1996

**Biochemistry 2<sup>nd</sup> edition**. Donald Voet and Judith G. Voet. John Wiley & Sons, Inc. pp: 55-67. 1995

**Genetics, a molecular approach 2<sup>nd</sup> edition**, T.A. Brown. Chapman and Hall, pp: 12-56, 1993.

**Molecular and Cellular Biology**, Stephen L. Wolfe, Wadsworth Publishing Company, 1993.

**Human Molecular Genetics 2<sup>nd</sup> edition**, Tom Strachan and Andrew P. Raed, Bios Scientific Publishers Ltd, 1999.

**Molecular Biology of the Cell, 3<sup>rd</sup> edition**, Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD. Garland Publishing Inc, New York, pp:98, 223.

**Biochemistry 3<sup>rd</sup> edition**, Struyer L. WH Freeman, New York, 1988.

1) **BIOINFORMATICS**, Chapter 26,

*Overview and Research Opportunities*, J.David Schaffer, Nevenka Dimitrova, Michael Q.Zhang, G. Spekowius and T. Wendler (Eds.), *Advances in Healthcare Technology*, 421-438. 2006 Springer.

2) Λεξικογραφική Ομοιότητα, Κεφάλαιο 3, Συγκριτική Αξιολόγηση Μεθόδων Λεξικογραφικής και Σημασιολογικής Ομοιότητας, Κωνσταντίνος Μ. Κολομβάτσος.

3) [http://nemertes.lis.upatras.gr/dspace/bitstream/123456789/1656/1/phd\\_theodori.pdf](http://nemertes.lis.upatras.gr/dspace/bitstream/123456789/1656/1/phd_theodori.pdf)

4) **Parallel Smith-Waterman Algorithm for Local DNA Comparison in a cluster of Workstations**

Azzedine Boukerche, Alba Cristina Magalhaes Alves de Melo, Mauricio Ayala – Rincon, and Thomas M. Santana.

WEA 2005, LNCS 3503, PP. 464-475, 2005. (Ed.): S.E. Nikolettseas.

5) **Protein Remote Homology Detection Based on Binary Profiles**

Qiwen Dong, Lei Lin, and Xiaolong Wang

LNBI 4414, pp. 212-223, 2007.

(Eds): S. Hochreiter and R. Wagner.

6) **A Smith-Waterman Systolic Cell,**

C.W.Yu, K.H.Kwong, K.H.Lee and P.H.W.Leong

New Algorithms, Architectures and Applications for Reconfigurable Computing, 291-300.

2005 Springer. P.Lysaght and W.Rosenstiel (eds.).

7) **ΒΙΟΧΗΜΕΙΑ ΣΤΗΝ ΙΑΤΡΙΚΗ,**

Αμαλία Γ.Διονυσίου-Αστερίου, Κωνσταντίνος Β.Στρούγκος, pp 157-169, Ιατρικές Εκδόσεις Πασχαλίδης 2006.

8) **Genomics**

R.W. Shonkwiler and J.Herod, *Mathematical Biology: An Introduction with Maple and Matlab*, Undergraduate Texts in Mathematics, Springer Science + Business Media, LLC 2009.

9) **Indexing DNA Sequences Using q-Grams**

Xia Cao, Shuai Cheng Li, and Anthony K.H.Tung. Department of Computer Science, National University of Singapore. L.Zhou, B.C.Ooi, and X.Meng (Eds): DASFAA 2005, LNCS 3453, pp 4-16,2005

10) **Inference of Phylogenetic Trees**

L.S. Kubatko

Departments of Statistics and Evolution, Ecology, and Organismal Biology, The Ohio State University, Columbus, OH 43210, USA.

11) <http://www.acm.org/crossroads/xrds9-3/alignment.html>

12) **Data Mining for Bioinformatics**

A.W.-C.Liew, Hong Yan, and Mengsu Yang

13) **A Survey of Text Processing Tools for the Automatic Analysis of Molecular Sequences**

Andrzej Polański<sup>1,2</sup>, Rafał Pokrzywa<sup>1</sup>, and Marek Kimmel<sup>3,1</sup>

1 Silesian University of Technology, Gliwice, Poland

andrzej.polanski@polsl.pl, pokrzywek@poczta.onet.pl,

2 Polish Japanese Institute of Information Technology, Bytom, Poland

3 Rice University, Houston, USA

LNCS 5070, PP. 359-378, 2009. (Eds): M. Marciniak and A. Mykowiecka

**14) A Simple, Practical and Complete  $O(n^3 \log n)$ -Time**

**Algorithm for RNA Folding**

**Using the *Four-Russians* Speedup**

Yelena Frid and Dan Gusfield

Department of Computer Science, U.C. Davis

LNBI 5724, pp 97-107, 2009, WABI 2009, (Eds): S.L. Salzberg and T. Warnow

**15) A Reconfigurable Index FLASH Memory tailored to Seed-Based Genomic Sequence Comparison Algorithms**

D. LAVENIER AND G. GEORGES

IRISA/CNRS, Campus de Beaulieu, Rennes, 35042, France

X. LIU

Key Laboratory of Computer System and Architecture, Institute of Computing Technology, CAS, Beijing, China

**16) Agent-Supported Protein Structure Similarity Searching**

Dariusz Mrozek, Bożena Małysiak, and Wojciech Augustyn

Silesian University of Technology, Department of Computer Science, Akademicka 16,

44-100 Gliwice, Poland

**17) Bioinformatic Tools for Gene and Protein Sequence Analysis**

**Bernd H. A. Rehm and Frank Reinecke**

**18) Flamm, C., Fontana, W., Hofacker, I., Schuster, P.: Kinetic folding of RNA at elementary step resolution. RNA 6 (2000) 325–338**

**19) Graph Transformation in Molecular Biology**

Francesc Rossell<sup>1</sup> and Gabriel Valiente<sup>2</sup>

1 Department of Mathematics and Computer Science,

Research Institute of Health Science (IUNICS), University of the Balearic Islands, Palma de Mallorca, Spain

2 Department of Software, Technical University of Catalonia, Barcelona, Spain

**20) An introduction to genetic analysis.** Griffiths et al. 6<sup>th</sup> edition, 1996, w.h.freeman, New york.

**21) Combining a High-Throughput Bioinformatics Grid and Bioinformatics Web Services**

Chunyan Wang<sup>1</sup>, Paul M.K. Gordon<sup>1</sup>, Andrei L. Turinsky<sup>1</sup>,

Jason Burgess<sup>2</sup>, Terry Dalton<sup>2</sup>, and Christoph W. Sensen<sup>1</sup>  
1 Sun Center of Excellence for Visual Genomics, University of Calgary,  
HS 1150, 3330 Hospital Dr. NW, Calgary, Alberta, T2N 4N1, Canada  
2 National Research Council Institute for Marine Biosciences,  
1411 Oxford Street, Halifax, Nova Scotia, B3H 3Z1, Canada

**22) Manifestation and Exploitation of  
Invariants in Bioinformatics**

Limsoon Wong  
School of Computing, National University of Singapore  
3 Science Drive 2, Singapore 117543

**23) Computational Prediction of Protein–Protein Interactions**  
**Anton J. Enright, Lucy Skrabanek, and Gary D. Bader**

**24) INCREMENTAL STRING COMPARISON**

GAD M. LANDAU<sup>y</sup>, EUGENE W. MYERS<sup>z</sup>, AND JEANETTE P. SCHMIDT<sup>x</sup>  
SIAM J. COMPUT. c 1998 Society for Industrial and Applied Mathematics  
Vol. 27, No. 2, pp. 557{582, April 1998 012

**25) Computer-aided design of nano-filter construction using DNA  
self-assembly**

Reza Mohammadzadegan Z Hassan Mohabatkar  
Nanoscale Res Lett (2007) 2:24–27  
DOI 10.1007/s11671-006-9024-6

26) <http://www.secnews.gr/archives/6366>

27) [http://2010.igem.org/files/presentation/Hong\\_Kong-CUHK.pdf](http://2010.igem.org/files/presentation/Hong_Kong-CUHK.pdf)

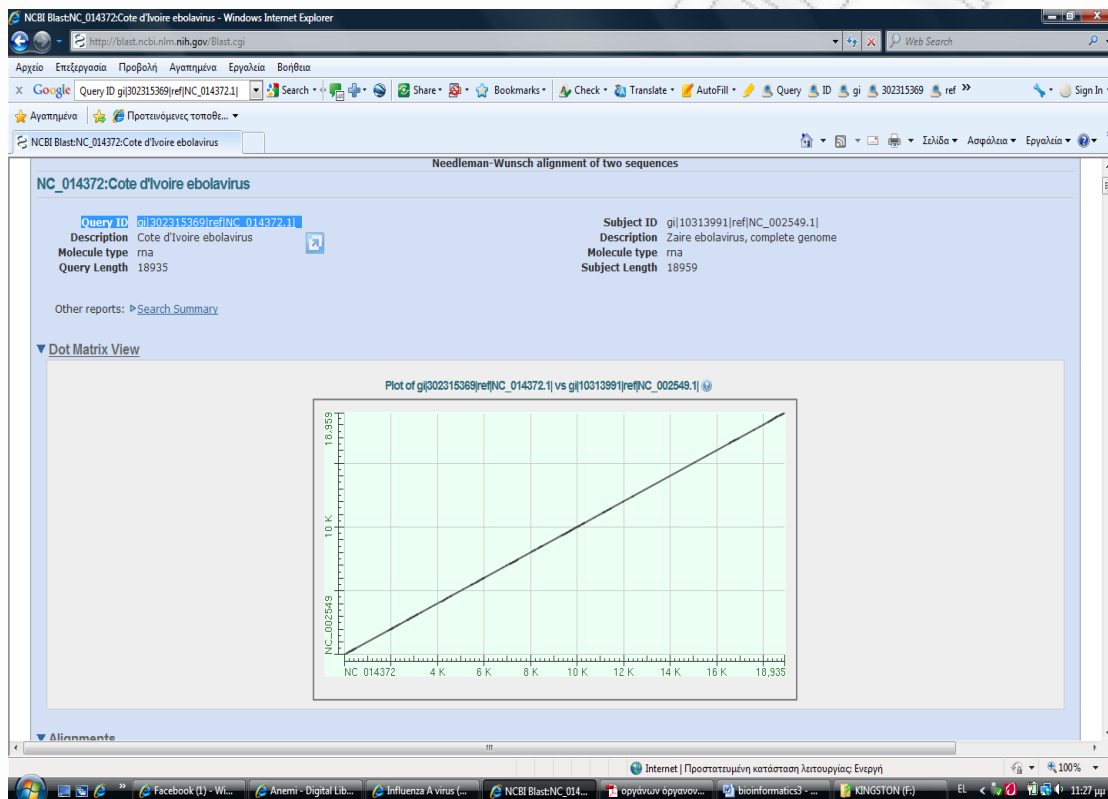
28) **Ανάλυση Και Σχεδίαση Αλγορίθμων 2<sup>η</sup> έκδοση**, Anany Levitin, Εκδόσεις Τζιόλα,  
Κεφάλαιο 1, 2008.

29) <http://dwave.wordpress.com/ideasforapplicationsandexperiments/applicationscomputationalbiology/generatingmultiplesequencealignments/>

30) From: <http://www.bioinformatics.kmutt.ac.th/download/seminar/bif03/SarawutPPT2.pdf>

## Παράρτημα

Σύγκριση με blast χρησιμοποιώντας τον αλγόριθμο Needleman – Wunsch μεταξύ του στελέχους του Ζαίρ του Ιού Έμπολα και του αντίστοιχου στελέχους της Ακτής Ελεφαντοστού.



NW Score = 5502

Identities = 13095/19564 (67%), Gaps = 1234/19564 (6%)

Strand=Plus/Plus

```

Query 1      CGGACACACAAAAGAAAGAAGGTTTTT--GATCTTATTGTGCGAATAACTATGAG  58
             ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct 1      CGGACACACAAAAGAAAGAAGAATTTTAGGATCTTT--TGTGTCGAATAACTATGAG  58

Query 59     GAAGATTAATAATTTTCCCTCATTTGACACTTACATTAAGATTAAGATTCTCATTGAT-C  117
             ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct 59     GAAGATTAATAATTTTCCCTCATTTGAAATTTATATCGGAATTTAAATTGAAATTGTTAC  118

Query 118    TGTTA-CTTAC-----TCTGAGGATAATAATTGGTGTTCAGAAGTACCCCATTC  166
    
```

Sbjct	119	TGTAATCACACCTGGTTTGTTCAGAGCCACATCACAAAGATAGAGAACAACCTAGGTCT		178
Query	167	CCAGTGGGGGCAAAGACAGTCCAAAAGACTCAACTTG----TCCTAT-TCAACTAATCTG		221
Sbjct	179	CCGAAGGGAGCAAGGGCA-TCAGTGTG-CTCAG-TTGAAAATCCCTTGTCAACACCTAGG		235
Query	222	TTTTGTC---TCAGTAGTTCACATATTTGATCATACCCAGGAGTTGGACCTAATCCAAAG		278
Sbjct	236	TCTTATCACATCACAAGTTC-CACCT----CAGACTCTGCAGGGTGATCCAA---CAAC-		286
Query	279	CTTAGAGTGGGACCTAGTGTATCCTCGGGCTGTAATATAATCAGCCAT--TTAACA--C		334
Sbjct	287	CTTA-ATAGAAACAT----TAT-----TGTTAAAGGA-CAGCATTAGTTCACAGTC		331
Query	335	ATAACAAGCCCTACTG----TTTTC-TTG-TTTTGC-CGTG--CATTTAGA--AT--AAG		381
Sbjct	332	A-AACAAGCAAGATTGAGAATTAACCTTGGTTTGAACCTGAACACTAGGGGATTGAAG		390
Query	382	AGACAACCTTAAACCTCC-GATTCCGC-AACACA--GGGAATAATCTCACCAGACCCGGCA		437
Sbjct	391	ATTC AAC--AACCTAAAGCTTGGGGTAAAACATTGGAAATAGTTAAA--AGACAA---A		443
Query	438	GTGTCTTCAGGCTT-CATAGCCCAAG-ATGGAGAGTCGGGCCACAAAGCATGGATGAC		495
Sbjct	444	TTG-CT-CGGAATCACAAAATCCGAGTATGGATTCTCGTCTCAGAAAATCTGGATGGC		501
Query	496	GCACAC-CGCATCAGGTTTCGAAACAGATTACCATAAGATTTTAACAGCAGGATTGTCAG		554
Sbjct	502	GCCGAGTCTCA-CTGAATCTGACATGGATTACCACAAGATCTTGACAGCAGGTCTGTCCG		560
Query	555	TCCAACAAGGCATTGTGAGACAACGGGTCAATCAAGTCCACCAGGTTACAAACCTAGAAG		614
Sbjct	561	TTCAACAGGGGATTGTTCCGCAAAGAGTCATCCAGTGTATCAAGTAAACAATCTTGAAG		620
Query	615	AAATATGCCAATTGATCATTCAAGCCTTTGAAGCTGGTGTGATTTTCAAGAGAGTGCAG		674
Sbjct	621	AAATTTGCCAACTTATCATACAGGCCTTTGAAGCAGGTGTGATTTTCAAGAGAGTGCAG		680
Query	675	ACAGTTTCTTGTGATGCTATGTTTACATCATGCTTATCAGGGTGACTACAAGCAATCTCT		734
Sbjct	681	ACAGTTTCTTCTCATGCTTTGTCTTTCATCATGCGTACCAGGGAGATTACAACTTTTCT		740
Query	735	TGGAAGCAATGCAGTCAAGTACCTTGAGGGTCATGGCTTTTCGCTTTGAGGTCAGGAAA		794
Sbjct	741	TGGAAGTGGCGCAGTCAAGTATTTGGAAGGGCACGGGTTCCGTTTTGAAGTCAAGAAGC		800
Query	795	AGGAAGGAGTCAAGCGACTCGAAGAATTGCTTCCTGCTGCATCCAGTGGAAGAGCATCA		854
Sbjct	801	GTGATGGAGTGAAGCGCCTTGAGGAATTGCTGCCAGCAGTATCTAGTGGA AAAACATTA		860
Query	855	GGAGAACACTGGCTGCAATGCCTGAAGAGGAGACAACAGAAGCAAATGCCGGACAGTTCC		914
Sbjct	861	AGAGAACACTTGCTGCCATGCCGGAAGAGGAGACAACCTGAAGCTAATGCCGGTCAGTTTC		920

Query	915	TCTCTTTTGCTAGCTTATTTCTTCCSTAAGCTAGTTGTCGGAGAAAAAGCCTGTCTAGAAA	974
Sbjct	921	TCTCCTTTGCAAGTCTATTCCTTCCGAAATTTGGTAGTAGGAGAAAAGGCTTGCCTTGAGA	980
Query	975	AGGTGCAGCGGCAAAATCAAGTTCATTCTGAGCAGGATTGATCCAATACCCACAGCCT	1034
Sbjct	981	AGGTCAAAGGCAAAATCAAGTACATGCAGAGCAAGGACTGATACAATATCCAACAGCTT	1040
Query	1035	GGCAGTCAGTTGGACACATGATGGTCATTTTCAGACTGATGAGAACAAATTTCTAATTA	1094
Sbjct	1041	GGCAATCAGTAGGACACATGATGGTGATTTTCCGTTTGTGCGAACAAATTTCTGATCA	1100
Query	1095	AGTTCCTCCTTATACATCAAGGGATGCATATGGTAGCAGGACAGATGCTAACGATGCTG	1154
Sbjct	1101	AATTTCTCCTAATACACCAAGGGATGCACATGGTTGCCGGGCATGATGCCAACGATGCTG	1160
Query	1155	TCATCGCAAATCTGTAGCTCAAGCAGCTTTTTCAGGATTATTGATCGTTAAAACAGTGC	1214
Sbjct	1161	TGATTTCAAATTCAGTGGCTCAAGCTCGTTTTTTCAGGCTTATTGATTGTCAAAAACAGTAC	1220
Query	1215	TAGATCACATCCTTTCAGAAAACAGAGCACGGAGTGCCTTTTCATCCTTTGGCAAGAACTG	1274
Sbjct	1221	TTGATCATATCCTACAAAAGACAGAACGAGGAGTTCGCTCCATCCTCTTGCAAGGACCG	1280
Query	1275	CSTAAGTCAAGAACGAAGTAAATTCCTTTAAGGCTGCCCTTAGCTCGCTAGCACAACATG	1334
Sbjct	1281	CCAAGTAAAAAATGAGGTGAATCCTTTAAGGCTGCACTCAGCTCCCTGGCCAAGCATG	1340
Query	1335	GAGAGTATGCTCCTTTGCTCGCTTGGTGAATCTTCTGGAGTCAACAATCTCGAGCAG	1394
Sbjct	1341	GAGAGTATGCTCCTTTGCGCCGACTTTTGAACCTTCTGGAGTAAATAATCTTGAGCATG	1400
Query	1395	GACTGTTTCCTCAGCTTTCTGCAATTCGCCCTAGGTGTGCGAACGGCACACGGCAGTACCC	1454
Sbjct	1401	GTCTTTTCCTCAACTATCGGCAATTCGCACTCGGAGTCGCCACAGCACACGGGAGTACCC	1460
Query	1455	TGGCAGGAGTAAATGTGGGGGAACAGTATCAGCAACTACGAGAAGCAGCCACTGAGGCAG	1514
Sbjct	1461	TGCGAGGAGTAAATGTGGGAGAACAGTATCAACAACACAGAGAGGCTGCCACTGAGGCTG	1520
Query	1515	AAAAACAATTCAGAAATACGCTGAATCTCGCGAGCTTGACCATCTAGGTCTCGATGATC	1574
Sbjct	1521	AGAAGCAACTCCAACAATATGCAGAGTCTCGCGAAGCTTGACCATCTTGACTTGATGATC	1580
Query	1575	AAGAGAAGAAGATCTTGAAAGACTTCCATCAGAAGAAAAATGAAATCAGCTTCCAGCAGA	1634
Sbjct	1581	AGGAAAAGAAAAATTTATGAACTTCCATCAGAAAAAGAACGAAATCAGCTTCCAGCAAA	1640
Query	1635	CAACAGCCATGGTCACACTACGGAAGGAAAGGCTAGCCAAGCTCACTGAGGCAATCACCT	1694
Sbjct	1641	CAAACGCTATGGTAACTCTAAGAAAAGAGCGCTGGCCAAGCTGACAGAAGCTATCACTG	1700
Query	1695	CCACATCCCTTCTCAAGACAGGAAAACAGTATGATGATGACAACGATATCCCTTTCTCG	1754

Sbjct	1701	CTGCGTCACTGCCCAAACAAGTGGACATTACGATGATGATGACGACATTCCTTTCCAG	1760
Query	1755	GGCCCATCAATGATAACGAAACTCAGAACAGCAAGACGATGATCCAACAGATTCTCAGG	1814
Sbjct	1761	GACCCATCAATGATGACGACAATCCTGGCCATCAAGATGATGATCCGACTGACTCACAGG	1820
Query	1815	ACACTACCATCCCTGATATCATTGTTGACCCGGATGATGGCAGATACAACAATTATGGAG	1874
Sbjct	1821	ATACGACCATTCCCGATGTGGTGGTTGATCCCGATGATGGAAGCTACGGCGAATACC-AG	1879
Query	1875	ACTATCCTAGTGAGACGGC--GAATGCCCTGAAGACCTTGTCTTTTTGACCTTGAAGA	1932
Sbjct	1880	AGT-TACTCGAAAACGGCATGAATGCACCAGATGACTTGGTCTTATTCGATCTAGACGA	1938
Query	1933	TGGTGACGAGGATGATCACCGACCGTCAAGTTCATCAGAGAACAACAACAACACAGTCT	1992
Sbjct	1939	GGACGACGAGGA----CACTAAGC--CA-GTGCCTAATAGATCGAC--CAAGGGTGGACA	1989
Query	1993	TACAGAACTGACAGTAACAAAA---CAAGTAACTGGAATCGAAACCCGACT-AAT--AT	2046
Sbjct	1990	-ACAGAAGA--ACAGT--CAAAGGGCCAGCATATAGAGGGCAGACA-GACACAATCCAG	2043
Query	2047	GCCAA---AGAAAGACTC-----CACACAAAACAATGACAATCCTGCACAGCGGGC---T	2095
Sbjct	2044	GCCAAATCAAATGTCCAGGCCCTCACAGAACAAT--CCACCAGGC-CAGTGCGCCACT	2100
Query	2096	CAAG---AAT-ACGCCAGGGAT-AACATCCAGGATACACCAACACCCCATCGAGCTCTAA	2150
Sbjct	2101	CACGGACAATGACAGAAGAAATGAACCTCCGGCTCAACCAGCCC---TCGCATGCTGA	2156
Query	2151	CTCCCATCAGCGAAGAAACCGGCTCCAATGGTCACAATGAAGATGAC---ATTGATAGCA	2207
Sbjct	2157	CACCAATTAACGAAGAGGCAGAC-CCACTGG--ACGATGCCGACGACGAGACGTCTAGCC	2213
Query	2208	TCCCTCCTTTGGAATCAGACGAAGAAAA-CAACACTGAGAC--AACCATACCACCACAA	2264
Sbjct	2214	TTCCGCCCTTGGAGTCAGATGATGAAGAGCAGGACAGGGACGGAAC--TTCCAACCGCAC	2271
Query	2265	AAAATACCCTG---CTCCACCAGCACCTGTTTATCGGAG-TAATTCAGAAAAGGAGCCC	2320
Sbjct	2272	AC----CCACTGTGCCCCACCGGCTCCCGTATA-CAGAGATCACTCTGAAAAGAAAGAA	2326
Query	2321	CTCCCGCAAGAAAAATCCCAGAAGCA--ACCAAAC-CAAGTGAGTGGTAGTGAGAATACC	2377
Sbjct	2327	CTCCCGCAAGACGAGCAACAAGATCAGGACCACACTCAAGAGGCCAGGAACCAGGACAGT	2386
Query	2378	GACAATAAACCTCACTCAGAGCAATCAGTGGGAAGAAATGTATCGACACATCCTCCAAACA	2437
Sbjct	2387	GACAACA--CC-CAGTCAGAACACTCTTTTGAGGAGATGTATCGCCACATTCTAAGATCA	2443
Query	2438	CAAGGACCATTGATGCCATCCTATACTATTACATGATGACGGAGGAGCCGATTGTCTTT	2497
Sbjct	2444	CAGGGGCCATTGATGCTGTTTTGTATTATCATATGATGAAGGATGAGCCTGTAGTTTTC	2503



ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΑΘΑΝΑΣΙΟΣ ΠΑΥΛΙΔΗΣ

Query 2498 AGCACTAGTGTGGGAAAGAATACGTATACCCTGATCTCTTGAAGGGGAGCATCCACCG 2557  
 || || ||||| ||||| ||| || || || || ||||| ||| |||||

Sbjct 2504 AGTACCAGTGTGGCAAAGAGTACACGTATCCAGACTCCCTTGAAGAGGAATATCCACCA 2563

Query 2558 TGGCTCAGTAAAAAGAGGCCTTGAATGAGGACAATAGGTTTATCACAATGGATGATCAA 2617  
 ||||| ||||| ||||| ||||| || ||||| ||| ||| ||||| ||||

Sbjct 2564 TGGCTCACTAAAAAGAGGCTATGAATGAAGAGAATAGATTTGTTACATTGGATGGTCAA 2623

Query 2618 CAATTCTACTGGCCTGTAATGAATCACAGGAACAAATTCATGGCTATCCTTCAGCACCAC 2677  
 ||||| || ||||| || ||||| ||||| ||||| ||||| ||||| |||||

Sbjct 2624 CAATTTTATTGGCCGGTGATGAATCACAGAATAAATTCATGGCAATCCTGCAACATCAT 2683

Query 2678 AAGTAATTTCTTCATAATGACAGATCATTGTAAGGTTATTACCACCA--TCCCTGCAACA 2735  
 ||| || | ||| ||||| ||| ||| ||| ||| ||| ||| ||| |||

Sbjct 2684 CAGTGAATG----AGCATG---GAACAATGGGATGATTC AACCGACAAATAGCT--AACA 2734

Query 2736 --AAGCA-TGAA--AACCACACTCAACAACGCCCTACCACAGGATACCTTGGAGACCAT- 2789  
 ||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Sbjct 2735 TTAAGTAGTCAAGGAACGAAAA-CAGGAAGAATTTTGTAT-GTCTAAGGTGTGAATTATT 2792

Query 2790 ---ACACCAA----GATCAGCAGCTGTGCAACCACCCCATGCGAATCCACCACCA--CA 2840  
 ||| || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Sbjct 2793 ATCACAATAAAGTGATTCTTATTTTGAATTTAAAGCTA-GCTTATT-ATTACTAGCCG 2850

Query 2841 ACCACCAA---CAAT--AATCCCAAGACCAAACCGCACACATCCAGA-TCA-ACCCA- 2891  
 |||| |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Sbjct 2851 TTTTTCAAAGTCAATTTGAGTCTTAATGCAATAGGCGTTAAGCCACAGTTATAGCCAT 2910

Query 2892 AACCCTCAAACACCACCCACTCCGCGATCCCAGACCAAACCTCCGCCCCAGACAAGCACC 2951  
 || ||| ||| ||| ||||| ||||| ||||| ||||| |||||

Sbjct 2911 AATTGTAACCTCAATATCTAAGTAGCGATTAT--CTAAATTA-----AATTACA 2958

Query 2952 CCACCCATCCCAGAAACCGCACGGCCGAGAATCGATCCCAGCATTCAAA--ATGCGTT- 3008  
 | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Sbjct 2959 TTATGCTTTT-ATAA--CTTACCTACTAGCCTG----CCCAACATTTACACGAT-CGTTT 3010

Query 3009 ----ATTAAGAAAAACATATGATGAAGATTA AACCTTCATCAACATTGCA-CA---GA 3060  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Sbjct 3011 TATAATTAAGAAAAACTAATGATGAAGATTA AACCTTCATCATCCTTACGTCAATTGA 3070

Query 3061 CTTTGAT--CCTTAGGAGTTTATTCTAG-CTATCTACAA--AACG--GGTCCAAAACGGA 3113  
 || | ||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| |||

Sbjct 3071 ATTCTTAGCACTCGAAGCTTATTGTCTTCAATGTAAAAGAAAAGCTGGTCTAACAAG-- 3128

Query 3114 ATGATTTCCACTAG-----GGCTGCAGCAATCAATGATCCTTCATTACCAATCAGAAA 3166  
 |||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Sbjct 3129 ATGA---CACTAGAACAAAGGGCAGGGCCAT-ACTGCGGC--CACGAC---TCA-AAA 3178

Query 3167 CCAGTGTACACGTGGCCCTGAACTATCAGGATGGATCTCCGAACAATTAATGACAGGCCAA 3226  
 | | | | | ||||| ||| ||| ||||| ||| ||| ||||| ||| |

Sbjct 3179 CGACAGAATGCCAGGCCCTGAGCTTTCGGGCTGGATCTCTGAGCAGCTAATGACCGGAAG 3238

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΑΘΑΝΑΣΙΟΣ ΠΑΥΛΙΔΗΣ

Query 3227 AATTCGGTACATGAAATCTTCAACGACACTGAGCCCCACATAAGCTCAGGGTCCGACTG 3286  
 ||||| ||| || ||||| || | |||| | | | | | | | | | |  
 Sbjct 3239 AATTCCTGTAAGCGACATCTTCTGTGATATTGAGAACAATCCAGGATTATGCTACG---- 3294

Query 3287 CCTTCCCAGACCCAA-AAACACGGCCC--CCCG--GACTCGCAACACCCAGACACAGACC 3341  
 | ||||| | ||| |||| ||| |||| ||| ||||| ||||| |||||  
 Sbjct 3295 -CATCCCAAATGCAACAAACGAAGCCAAACCCGAAGACGCGCAACAGTCAAACCCAAACG 3353

Query 3342 GATCCGGTTTGAATCACAATTTTGAAGACGTTACACAAGCACTAACATCATTAAACCAAT 3401  
 || | | ||||| || | ||||| || | | |||| | | | | | | | | | |  
 Sbjct 3354 GACCCAATTTGAATCATAGTTTGGAGGAGTAGTACAAACATTGGCTTCATTGGCTACT 3413

Query 3402 GTCATACAAAAACAGGCTCTTAACTTAGAGTCTCTCGAACACGCATCATAGATCTAGAG 3461  
 || | ||| |||| | | | ||| || | ||||| ||||| | ||| |||  
 Sbjct 3414 GTTGTGCAACAACAACCATCGCATCAGAATCATTAGAACAACGCATTACGAGTCTTGAG 3473

Query 3462 AATGGCTTAAAGCCAATGTATGACATGGCTAAAGTCATTTCTGCATTGAATAGATCTTGT 3521  
 ||||| ||||| | ||||| ||||| ||| || | | ||||| || | |||||  
 Sbjct 3474 AATGGTCTAAAGCCAGTTTATGATATGGCAAAAACAATCTCCTCATGAACAGGGTTTGT 3533

Query 3522 GCTGAGATGGTAGCAAAATATGATCTCTGGTGATGACAACGGCCGCAACCGCCACC 3581  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Sbjct 3534 GCTGAGATGGTTGCAAAATATGATCTCTGGTGATGACAACGGTCGGGCAACAGCAACC 3593

Query 3582 GCCGCTGCAACTGAGGCTTATTGGGAGGAACATGGACAACCACCACCTGGACCATCACT 3641  
 || | | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Sbjct 3594 GCTGCGCAACTGAGGCTTATTGGGCGGAACATGGTCAACCACCACCTGGACCATCACT 3653

Query 3642 TATGAAGAGAGTGCGATTAGAGGCAAGATTAACAAGCAAGAG--GATAAAGTACCTAAGG 3699  
 ||||| ||||| | || | |||| | | | |||| | | | | | | | |  
 Sbjct 3654 TATGAAGAAAGTGCGATTGCGGGTAAGATTGA--ATCTAGAGATGAGACCGTCCCTCAA 3711

Query 3700 AAGTTCAGAAGCTTTTCGTAATCTGGACAGTACCAGCTCACTAACAGAAGAGAACTTTG 3759  
 ||| || || | | ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Sbjct 3712 GTGTTAGGGAGGCATCAACAATCTAAACAGTACCCTTCACTAACGAGGAAAATTTG 3771

Query 3760 GCAAGCCAGATATATCTGCAAGGACCTACGAGACATCATGTATGACCACCTACCAGGCT 3819  
 | || | | | | | | ||||| | | |||| | ||||| ||||| ||||| |||||  
 Sbjct 3772 GGAAACCTGACATTCGGCAAGGATTTGAGAAACATTATGTATGATCACTTGCCTGGTT 3831

Query 3820 TCGGTACGGCTTTTCCACCAACTGGTCCAGGTAATTTGCAAGCTAGGAAAAGA---CAATT 3876  
 | || | | ||||| ||||| | || | | |||| | | | ||||| |||||  
 Sbjct 3832 TTGGAAGTGCCTTCCACCAATTAGTACAAGTATTGTAATTTGGGAAAAGATAGCAACT 3891

Query 3877 CTGCATTGGACATTATTCATGCTGAGTTCAGCCAGCCTTGCTGAAGGTGATTCTCCCC 3936  
 | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 Sbjct 3892 C---ATTGGACATCATTCATGCTGAGTTCAGGCCAGCCTGGCTGAAGGAGACTCTCCTC 3948

Query 3937 AATGTGCCCTGATCCAAATAACAAAACGGATCCCCATCTTCCAGGATGCCACTCCGCCA 3996  
 ||||| || | |||| | |||| | | | ||||| ||||| ||||| |||||  
 Sbjct 3949 AATGTGCCCTAATTCAAATTACAAAAGAGTTCCAATCTTCCAAGATGCTGCTCCACCTG 4008

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΑΘΑΝΑΣΙΟΣ ΠΑΥΛΙΔΗΣ

Query 3997 CAATTCACATCCGCTCTCGTGGTGACATCCACGTGCCTGCCAAAAAGTCTCCGTCCAG 4056  
 || ||| ||||| ||||| ||| || | ||||| ||||| | |||||

Sbjct 4009 TCATCCACATCCGCTCTCGAGGTGACATTCCCCGAGCTTGCCAGAAAAGCTTGCGTCCAG 4068

Query 4057 TTCCTCCATCACCAAAAAATAGACAGAGGTGGGTTCGATTTTCCAATTGCAGGACGGGA 4116  
 | | | |||| | | | | | | | ||||| || | ||| | | | | | | |

Sbjct 4069 TCCCACCATCGCCCAAGATTGATCGAGGTGGGTATGTGTTTTTCAGCTTCAAGATGGTA 4128

Query 4117 AGACACTTGGGCTCAAGATATAGGGTCCCCAGTCAAAGACACGTGCGGTCCCATCTCC 4176  
 | ||||| | | | | | | | | | ||||| | | | | | | |

Sbjct 4129 AAACACTTGGACTCAAATTTGAG-----CCAATC-----TCCCTCCCTC 4169

Query 4177 CTCACCTTCAGACATCAACGCATGGCAGTCCCAAACACCGGTGAGGGAGGCGCCCGGGA 4236  
 | | | | | | | | | | | | | | | | | | | | | | |

Sbjct 4170 CGAA-----AGAGCGAATA-ATAGCAGAGGCTTCAACTGCTGAATA-----GGTA 4218

Query 4237 CAACACATGA-TGATAGGCT---GATCTTCGG-----GATAAGAGACATGAA--AAAC-- 4283  
 | |||| | |||| | | | | | | | | |||| | | | | | | |

Sbjct 4219 CGTTACATTAATGATACACTTGTGAGTATCAGCCCTGGATAATATAAGTCAATTAACGA 4278

Query 4284 -CAA---AAAGCCGTTTACATCCAGATCCAAGATCAAGAGTGGCTTGGAAAT--AAGGG 4337  
 ||| ||| ||| | ||| | | | | | | | | | | | | | |

Sbjct 4279 CCAAGATAAAATTGTTCAATCTCGTAGCAGCTTAAAA-----TATAAATGTAATAGG 4332

Query 4338 CACTTGTCTTTGTC-----TCAAAG-----ACTTACAAAAAC-AAGGGTGAT 4380  
 || ||| || | | |||| | | | | |||| | | |||| |

Sbjct 4333 AGCTATATCTCTGACAGTATTATAATCAATTGTTATTAAGTAACCCAAACAAAAGTGAT 4392

Query 4381 GAAGATTAAGAAAAAGCCTCCTTCAGTTGCAAG-GAGCTAATTCCTTAAACTTCATCTAG 4439  
 ||||| ||||| ||| | | | | | | | | | | | | | | | | | |

Sbjct 4393 GAAGATTAAGAAAAA-CCTACCTCGGCTGAGAGAGTGTTTTTTCATTAACCTTCATCTTG 4451

Query 4440 ---AC--TAAGGATAAATCGATTCCAATCACGATGAGGAGAATCATCTACCCACGGCAC 4494  
 || | | | | | | | | | | | | | | | | | | | | | | |

Sbjct 4452 TAAACGTTGAGCA-AAATTGTTAAAAAT----ATGAGGCGGGTTATATGCCTACTGCTC 4506

Query 4495 CACCTGAATACATGGAGGCTGTTACCCAATGAGAACA--TGAA---TT-CTGGTGCGAG 4548  
 | ||||| | | ||||| | | | | | | | | | | | | | | | | |

Sbjct 4507 CTCCTGAATATATGGAGGCCATATACCCTGTCAGGTCAAATCAACAATTGCTAGAGGTG 4566

Query 4549 ACAACACTGCCAGTGGCCCTAATTACACAACAAGTGGTGTGATGACAAATG--ATACTCC 4606  
 |||| | | | | | | | | | | | | | | | | | | | | | |

Sbjct 4567 GCAACAGCAATACAGGCT-TCCTGACACC-----GGAGTCA-GTCAATGGGGACTCC 4618

Query 4607 CTCTAATTCACTCCGACCAGTTGCAGATGATAAATTTGATCATCCGAGCCACACGCCATA 4666  
 || ||| |||| | | | |||| |||| | | | | | | | | | | | |

Sbjct 4619 ATCGAATCCACTCAGGCCAATTGCCGATGACACCATCGACCATGCCAGCCACACACCAGG 4678

Query 4667 CAGTGTGCCTCTGCATTTATATTGGAAGCTATGGTGAATGTAATATCTGGCCCGAAAGT 4726  
 ||||| | | |||| | | | ||||| ||||| ||||| ||||| |||||

Sbjct 4679 CAGTGTGCATCAGCATTCATCCTTGAAGCTATGGTGAATGTCATATCGGGCCCCAAAGT 4738

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΑΘΑΝΑΣΙΟΣ ΠΑΥΛΙΔΗΣ

Query	4727	GCTGATGAAGCAAATCCCAATCTGGCTTCCTCTGGGTGTCTCTGACCAGAAGACATATAG	4786
Sbjct	4739	GCTAATGAAGCAAATCCCAATTTGGCTTCCTCTAGGTGTCGCTGATCAAAAGACCTACAG	4798
Query	4787	CTTTGATTCAACCACTGCTGCCATTATGCTAGCATCATATACCATCACTCATTTTGGCAA	4846
Sbjct	4799	CTTTGACTCAACTACGGCCGCATCATGCTTCTTCATACACTATCACCCATTTTCGGCAA	4858
Query	4847	AACCTCAAATCCCCTTGTGAGAATCAACCGACTTGGTCCTGGCATACTGATCACCCACT	4906
Sbjct	4859	GGCAACCAATCCACTTGTGAGAGTCAATCGGCTGGGTCTGGAATCCCGGATCATCCCT	4918
Query	4907	ACGACTCCTAAGAATAGGAAATCAAGCCTTCCTACAAGAGTTTGTGCTACCTCCTGTACA	4966
Sbjct	4919	CAGGCTCCTGCGAATTGAAACCAGGCTTTCCTCCAGGAGTTCGTTCTTCCGCCAGTCCA	4978
Query	4967	ACTGCCACAATACTTCACTTTTGATCTGACAGCGCTGAAGCTGATCACCCAGCCACTCCC	5026
Sbjct	4979	ACTACCCAGTATTTACCTTTGATTTGACAGCACTCAAACGATCACCCAACCACTGCC	5038
Query	5027	AGCGGCAACCTGGACAGATGAAACTCCAGCTGTGTCAACTGGCACGCTCCGCCAGGGAT	5086
Sbjct	5039	TGCTGCAACATGGACCGATGACACTCCAACAGGATCAAATGGAGCGTTGCGTCCAGGAAT	5098
Query	5087	CTCATTCATCCCAAATTAAGGCCTATCCTGCTACC-AGGAAGAGCTGGAAAGAAGGGCT	5145
Sbjct	5099	TTCATTTATCCAAAACCTTCGCCCATCTTTTACCCAACAAAAG-TGGGAAGAAGGGGA	5157
Query	5146	CCAACCTCCGATCTAACATCTCCTGACAAAATCCAGGCTATAATGAATTTCTACAAGACC	5205
Sbjct	5158	ACAGTGCCGATCTAACATCTCCGGAGAAAATCCAAGCAATAATGACTTCACTCCAGGACT	5217
Query	5206	TCAAAATTGTACCAATCGATCCAACCAAGAATATCATGGGTATTGAAGTGCCAGAATCC	5265
Sbjct	5218	TTAAGATCGTTCCAATTGATCCAACCAAAAATATCATGGGAATCGAAGTGCCAGAACTC	5277
Query	5266	TGGTTCACAGGCTGACTGGGAAGAAGACAACCTACCAAGAATGGTCAACCAATCATTCCAA	5325
Sbjct	5278	TGGTCCACAAGCTGACCGGTAAGAAGGTGACTTCTAAAAATGGACAACCAATCATCCCTG	5337
Query	5326	TTCTGTACCAAAAGTACATTGGTCTTGATCCTCTATCTCAAGGTGATCTCACAAATGGTGA	5385
Sbjct	5338	TTCTTTTGCCAAAAGTACATTGGGTGGACCCGGTGGCTCCAGGAGACCTCACCATGGTAA	5397
Query	5386	TCACTCAGGACTGTGATTCTGCCACTCCCCGGCCAGTCTTCCCCAGTCAATGAAAAAT	5445
Sbjct	5398	TCACACAGGATTGTGACACGTGCATTTCTCCTGCAAGTCTTCCAGCTGTGATTGAGAAGT	5457
Query	5446	GA---CCATGA--GACTCAACATCACACTGCCAGAGCACCTCACCGCAAGTCTATACAAC	5500
Sbjct	5458	AATTGCAATAATTGACTCAG-ATC-CAGTTTATAGAATCTT--CTCAGGGATAGTGAT-	5512

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΑΘΑΝΑΣΙΟΣ ΠΑΥΛΙΔΗΣ

Query 5501 AATCAACCCCGCATCTACAACCTGCAAAAACCAGCCCATCTGAT---ACTCCTGGCATC 5557  
 ||| |||| | | | | |||| | | | | |  
 Sbjct 5513 ----AAC-----ATCTA-----TTTAGTAATCCGTCCATTAGAGGAGACACTTTTAATT 5557

Query 5558 GGGGGCAAGAC-AAGGCAGCCAAGCAGCAGCCCCGAGCCGAGCCCAAACCCAT-TACAC 5615  
 | | | | | | | | | | | | | | | | | | | | | | |  
 Sbjct 5558 GATCAATATACTAAAGGTGCTTTACACCATTGTCTTTTTTTCTCTCTAA---ATGTAGAA 5614

Query 5616 CCGAGCCCAACACCCATCCAGCAACCCACAACCGTCAAACGCACAGATGGACAAGCAAAG 5675  
 | | | | | | | | | | | | | | | | | | | | | | |  
 Sbjct 5615 CTTAACAAAAGACTCAT-----AATATACTTGTTTTAAAGGATTGATTGATGA---AAG 5666

Query 5676 AACATCAAGCCAGGAGCA--ACACAGACCCCAAGTCTAAGC-----TGAT-CAACCC- 5724  
 | | | | | | | | | | | | | | | | | | | | | | |  
 Sbjct 5667 ATCATAA--CTAATAACATTACAATAATCTACTATAATCAATACGGTGATTCAAAATGT 5724

Query 5725 ----CTCCGCAATCCCACCAACGCCA-GCAAAAATCCCCAA----CTCGATACCAACC 5775  
 | | | | | | | | | | | | | | | | | | | | | | |  
 Sbjct 5725 TAATCTTTCTCATTGC-ACATACTTTTTGCCCTTATCTCAAATTGCCTGCATGCTTACA 5783

Query 5776 CCA-AGCAAATC-AG-----CTCAAACCGTCTATCTCTCCCGCTTCACTCCACACCCC- 5827  
 | | | | | | | | | | | | | | | | | | | | | | |  
 Sbjct 5784 TCTGAGGATAGCCAGTGTGACTTGGATTGGAAATGTGGA---GAAAAAATCGGGACCCAT 5840

Query 5828 ----AGATTCAAGCAACGATCAACGCACCTTCTTATGCCACAGCTTATA-TTAAGAAAAAG 5882  
 | | | | | | | | | | | | | | | | | | | | | | |  
 Sbjct 5841 TTCTAGGTTGTTACACA--ATCCAAGTACAGACATTGCC---CTTCTAATTAAGAAAAAA 5894

Query 5883 AACTTGATGAAGATTAAGGCCAACAGTG-GTGCTATCTTCACTCTCTTT--GATT-TGAGT 5938  
 | | | | | | | | | | | | | | | | | | | | | | |  
 Sbjct 5895 TCGGCGATGAAGATTAAGCCGAC-AGTGAGCGTAATCTTCACTCTCTTTAGATTATTTGT 5953

Query 5939 CTTA-AGTGAA----TACACAGGTTCTAATACTGTTCTTCTGTCCAACGGTATAATTTCAG 5993  
 | | | | | | | | | | | | | | | | | | | | | | |  
 Sbjct 5954 TTTCCAGAGTAGGGTTCGTCAGGTCCTTTT-CAAT-CGTGTAACCAAAA-TAAACTCCA- 6009

Query 5994 CCAGGCCTAAGACA--GTAGCTAATCACAGTCATCATTGGGAGCGTCAGGGATTCTGCAAT 6051  
 | | | | | | | | | | | | | | | | | | | | | | |  
 Sbjct 6010 CTAGA---AGGATATTGTGGGCAACA-ACACA--ATGGGCGTTACAGGAATATTGCAGT 6063

Query 6052 TGCCCCGTGAGCGCTTCAAGAAAACATCTTTCTTTGTTGGGTAATAATCCTATTCCATA 6111  
 | | | | | | | | | | | | | | | | | | | | | | |  
 Sbjct 6064 TACCTCGTGATCGATTCAAGAGGACATCTTTCTTTGGGTAATTATCCTTTTCCAAA 6123

Query 6112 AAGTCTTTTCAATCCCGTTGGGGTTGTACACAACAATACCCTACAAGTGAGTGATATTG 6171  
 | | | | | | | | | | | | | | | | | | | | | | |  
 Sbjct 6124 GAACATTTTCCATCCCCTTGGAGTCATCCACAATAGCACATTACAGGTTAGTGATGTGC 6183

Query 6172 ACAAGTTTGTGTGCCGAGACAAACTCTTCAACTAGCCAATTGAAGTCAGTCGGGTGTA 6231  
 | | | | | | | | | | | | | | | | | | | | | | |  
 Sbjct 6184 ACAAACTAGTTTGTGTCGTGACAAACTGTCATCCACAATCAATTGAGATCAGTTGGACTGA 6243

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΑΘΑΝΑΣΙΟΣ ΠΑΥΛΙΔΗΣ

```

Query  6232  ACTTGGAGGGCAATGGAGTAGCAACTGATGTACCAACGGCAACCAAAGATGGGGTTTTC  6291
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  6244  ATCTCGAAGGGAATGGAGTGGCAACTGACGTGCCATCTGCAACTAAAAGATGGGGCTTCA  6303

Query  6292  GAGCTGGTGTTCACCAAAGGTGGTAAATTGCGAAGCTGGAGAATGGGCTGAGAACTGTT  6351
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  6304  GGTCCGGTGTCCACCAAAGGTGGTCAATTATGAAGCTGGTGAATGGGCTGAAAACCTGCT  6363

Query  6352  ATAACCTGGCTATAAAGAAAGTTGATGGTAGTGAGTGCCTACCAGAAGCCCCTGAGGGAG  6411
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  6364  ACAATCTTGAATCAAAAACCTGACGGGAGTGAGTGTCTACCAGCAGCGCCAGACGGGA  6423

Query  6412  TGAGGGATTTTCCCCTGTCGCCTATGTACACAAAGTCTCAGGAACCTGGACCATGCCAG  6471
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  6424  TTCGGGGCTTCCCCGGTGGCCGATGTGCACAAAGTATCAGGAACGGGACCGTGTGCCG  6483

Query  6472  GAGGACTCGCCTTTCACAAAGAAGGAGCCTTCTCCTGTATGACCGACTCGCATCAACAA  6531
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  6484  GAGACTTTCCTTCCATAAAGAGGGTGCCTTCTCCTGTATGATCGACTTGCCTCCACAG  6543

Query  6532  TCATTTATCGGGGTACAACCTTGGCCGAAGGAGTTATTGCATTTCTGATCTTGCCTAAGG  6591
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  6544  TTATCTACCGAGGAACGACTTTCGCTGAAGGTGTCGTGCATTTCTGATACTGCCCAAG  6603

Query  6592  CGCGAAAGGATTTTTTCCAGTCTCCTCCATTGCATGAGCCTGCCAACATGACCACGGATC  6651
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  6604  CTAAGAAGGACTTCTTCAGCTCACACCCTTGAGAGAGCCGGTCAATGCAACGGAGGACC  6663

Query  6652  CCTCCAGTTACTATCACACGACAACAATAAACTACGTGGTTGATAATTTTGGAAACCAACA  6711
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  6664  CGTCTAGTGGCTACTATTCTACCACAATTAGATATCAGGCTACCGGTTTTTGGAAACCAATG  6723

Query  6712  CCACAGAGTTTCTGTTCAGTGCATATTTGACGTATGTGCAGCTCGAGGCAAGATTCA  6771
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  6724  AGACAGAGTACTTGTTCGAGGTTGACAATTGACCTACGTCCAACCTGAATCAAGATTCA  6783

Query  6772  CACCACAATTCCTTGTCTCCTAAATGAAACCATCTACTC---TGATAACCGCAGAAGTA  6828
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  6784  CACCACAGTTTCTGTCTCAGCTGAATGAGACAATATATACAAGTGGGAAA---AGGAGCA  6840

Query  6829  ACACAAACAGGAAAACATACTGGAAAATAAATCCCAGTGTGATACCAGCATGGGTGAGT  6888
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  6841  ATACCACGGGAAAACATAATTTGGAAGGTCACCCCGAAATTGATACAACAATCGGGGAGT  6900

Query  6889  GGGCTTCTGGGAAAATAAAAAAACCCTCACAAAAACCCTTCAAGTGAAGAGTTGTCTTT  6948
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  6901  GGGCCTTCTGGGAAACTAAAAAAACCTCACTAGAAAAATTCGCAGTGAAGAGTTGTCTTT  6960

Query  6949  CGTACCTGTACCAGA---AACCAGAAC--CAG--GTCCTTGACACGACAGCGACGGTCT  7001
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  6961  CACAGTTGTATCAAACGGAGCCAAAAACATCAGTGGTCAGAGTC--CGGC--GCGAACTTCT  7018
    
```

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΑΘΑΝΑΣΙΟΣ ΠΑΥΛΙΔΗΣ

Query 7002 CTCCTCCCATCTCCGCCACAACCACGAGCCGAAGACCACAAAGAATTGGTTTCAGAGG 7061  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 7019 TCCGACCCAGGGACCAACACACAACAACACT-----GAAGACCACAAAATCATGGCTTCAGAAA 7073

Query 7062 ATTCCTCCTCCAGTGGTTTCAGATGCAAAAACATCAAGGGAAAGGACACAATGCCAACCACAG 7121  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 7074 ATTCTCTGCAATGGTTCAAGTGCACAG--TCAAGG-AAGGGAAGC--TGC-----AG 7121

Query 7122 TGACGGGTGTACCAACAA---CCACACCTCTCCA---TTTCCAATCAATGCTCGCAAC 7174  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 7122 TGTGCGATCTAACAACCCTTGGCCACA---TCTCCACGAGTCCCAATCC---CTCACAAAC 7176

Query 7175 ACTGATCATAACCAATCATTTATCGGCTGGAGGGGCCCAAGAAGACCACAGC----- 7228  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 7177 -----CAAACCAGGTC-----CGGACAACAG---CACCCATAATACCCCGTGTATAA 7221

Query 7229 ACCACACAGC-CTGCCAAG-ACCACCAGCCAACCAACACAGCACAGAATCGACGAC-A 7285  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 7222 ACTTGACATCTCTG--AGGCAACTCAAGTTGAACAAC-ATCACCAGCAGAACAGACAACGA 7278

Query 7286 CTA AACCCCAACATCAGAGCCCTCCAGTAGAGGCACGGGACCATCCAGCCCCACGGTCCCC 7345  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 7279 C--AGCACAGCCTCCGA-CACTCCCTCTGC--CACG--ACCG--CAGCCGGAC---CCCC 7326

Query 7346 AACACCACAGAAAGCCACGCGGAACCTTG-GCAA-GACA-----ACCCCA--ACCACAC- 7394  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 7327 AAAAGCAGAGAACACCAACACGAGCAAGAGCACTGACTTCCTGGACCCCGCCACCACAAC 7386

Query 7395 ---TCCGAGAACAGCAGCACTGCCGCCAGTGCCATTTC-CAAGAGCCGTGCACC-----CC 7444  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 7387 AAGTCCCCAAA-ACCACAGCGAGACCCGTGGCAACAACAACACTCAT-CACCAAGATACC 7444

Query 7445 GACGAACTCAGTG-----GAC-CT-GGCTTCCTGACGAACACAATACGGGGGGTT 7492  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 7445 GGAGAAGAGAGTGCCAGCAGCGGGAAGCTAGGCTTAATTACCAATACTATTGCTGGAGTC 7504

Query 7493 ACAAACTCTCCTGACAGGATCCAGAAGAAAGCGAAGGGATGTCACCTCCCAATACACAACCC 7552  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 7505 GCAGGACTGATCACAGGCGGGAGAAGA ACTCGAAGAGAAGCAATTGTCAATGCTCAACCC 7564

Query 7553 AAATGCAACCCAAACCTGCACATAATTGGACAGCCTTGGATGAGGGTGCTGCCATAGGTTTA 7612  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 7565 AAATGCAACCCATAATTACATTA TACTGGACTACTCAGGATGAAGGTGCTGCAATCGGACTG 7624

Query 7613 GCCTGGATACCATACTTCGGGCCAGCAGCTGAGGGAATTTACACTGAAGGCATAATGGAG 7672  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 7625 GCCTGGATACCATATTTTCGGGCCAGCAGCCGAGGGAATTTACATAGAGGGGCTAATGCAC 7684

Query 7673 AATCAAAATGGATTGATCTGTGGATTGAGGCAGCTGGCCAACGAAACGACACAAGCTCTT 7732  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 7685 AATCAAGATGGTTTAACTGTGGGTTGAGACAGCTGGCCAACGAGACGACTCAAGCTCTT 7744

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

Query 7733 CAAATGTTCTTAAGGGCAACTACTGAGTTGCGTACATCTCTCTATACTAAATCGGAAAGCA 7792  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 7745 CAACTGTTCCCTGAGAGCCACAACTGAGCTACGCACCTTTTCAATCCTCAACCGTAAGGCA 7804

Query 7793 ATAGACTTCTTGCTCCAAAGATGGGGAGGAACATGTCACATTCTAGGGCCTGATTGTTGC 7852  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 7805 ATTGATTCTTGCTGCGAGCGATGGGGCGGCACATGCCACATTCTGGGACCGGACTGCTGT 7864

Query 7853 ATTGAACCCCAAGATTGGACCAAAAAATATCACTGATAAAATTGATCAAATAATCCATGAC 7912  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 7865 ATCGAACCCATGATTGGACCAAGAACAATAACAGACAAAATTGATCAGATTATTCATGAT 7924

Query 7913 TTTGTCGATAATAATCTTCCAAATCAGAATGATGGCAGCAACTGGTGGACTGGATGGAAA 7972  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 7925 TTTGTTGATAAAACCCTTCCGGACCAGGGGACAATGACAATTGGTGGACAGGATGGAGA 7984

Query 7973 CAATGGGTTCCCTGCTGGAATAGGAATCACAGGAGTAATCATTTGCTATTATTGCTTTGCTG 8032  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 7985 CAATGGATACCGGCAGGTATTGGAGTTACAGGCGTTATAATTGCAGTTATCGCTTTATTC 8044

Query 8033 TGCATTTGCAAATTCATGCTTTGAACATAAT-----ATAGCATCAT--ACTTCTAATAT 8084  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 8045 TGTATATGCAAATTTGT-CTTTTAGTTTTTCTTCAGATTGCTTCATGGAAAAGCTCAG-- 8101

Query 8085 TCCCCAAT--ATGAATTTTGTTTTCGATTTTATTTAATGATATATCCTCTGTATACCT 8142  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 8102 -CCTCAAATCAATGAAACCAGGATTTA-ATTATATGGATTACTTGAATCTAAGATTACTT 8159

Query 8143 CACTAATG-TACTCGAGCATAATTTCCCTGATAGACTTGATTGTATTGTATG--ATTAAG 8199  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 8160 GACAAATGATAAT-----ATAATA-CACTGG-AGCTTTAAACATAGCCAATGTGATTCTA 8212

Query 8200 GACCTCACAAAATTCCTGGGGATTGAAAAGAACTGGATAAC-TCAATAAAATTTTATGCTA 8258  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 8213 -ACTCCTTTAAACTCACAGTTAATCATAAACAAGGTTTGACATCAATCTAGTTATCTCTT 8271

Query 8259 GGACCACAAATACACTTGATGAAGATTAAGAAAAGATAATCTTATGATTATCATTGATC 8318  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 8272 TGAGAATGA-TAAACTTGATGAAGATTAAGAAAAGGTAATCTTTCGATTATCTTTAATC 8330

Query 8319 TTCATCTATACCTTAAATACTCTATTCAAGGAGAGTA-----TGACAAAACCAAGTA 8370  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 8331 TTCATC-----CTTGATT-CTACAATCATG-ACAGTTGTCTTTAGTGACAAGGGAAAGAA 8383

Query 8371 GTATTGG-ATAAACTTGTCTGCATTCAAATCTGAAGACA--TACGGCTTAT-----C 8420  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 8384 GCCTTTTATTAAGTTGTAAT--AATCAGATCTGCGAACCGGTAGAGTTTAGTTGCAACC 8441

Query 8421 TATT-CAC-TATTGTATTAG--AAAATCTAGGGAATATCATTGAAACTAATTAGTGACT 8476  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 8442 TAACACACATAAAGCATTGGTCAAAA---AGTCAATAGAAATTTAAAC-----AGTGAGT 8493



ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΑΘΑΝΑΣΙΟΣ ΠΑΥΛΙΔΗΣ

Query 8477 AAAACACACAACACTCAAGTCGGCCAGAATGGAAGTTGTTTCATGAAAGAGGTCGCTCCAGGA 8536  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 8494 GGAG---ACAAC TTTTA-----AATGGAAGCTTCATATGAGAGAGGACGC-CCACGA 8541

Query 8537 TCTCCC-GACAAAACACAAGGGATGGACCTAGTCATTTAGTACGGGCGAGATCATCCTCT 8595  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 8542 GCTGCCAGACAGCATTC AAGGGATGGACACGACCACCATGTTTCGAGCAGCATCATCATCC 8601

Query 8596 CGAGCTAGTTATCGAAGTGAATACCATACACCAAGGAGTGCCTCGCAGATCCGTGTCCCC 8655  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 8602 AGAGAGAATTATCGAGGTGAGTACCGTCAATCAAGGAGCGCCTCACAAGTGC GCGTTCCCT 8661

Query 8656 ACTGTCTTTCATCGGAAAAAGACAGATTTA---TTGACAGTTCACCAGCACCTAAAGAT 8712  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 8662 ACTGTATTTCAT---AAGAAGAGAGTGAACCATTAACAGTTCCTCCAGCACCTAAAGAC 8718

Query 8713 GTATGCCCGACTTTAAAGAAAGGTTTCTATGTGACAGCAATTTCTGTAAAAAGGATCAC 8772  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 8719 ATATGTCCGACCTTGA AAAAAGGATTTTGTGTGACAGTAGTTTTTGCA AAAAAGATCAC 8778

Query 8773 CAACTTGAAAGCTTACAGATAGAGAGTTACTCTTGCTGATTGCACGCAAGACATGTGGA 8832  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 8779 CAGTTGGAGAGT TTAAGTATAGGGAATTACTCTACTAATCGCCCGTAAGACTTGTGGA 8838

Query 8833 TCCACGGAACAACAAC TAAGCATAGTTGCTCCAAAAGATTCACGTCTGGCTAATCCTATT 8892  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 8839 TCAGTAGAACAACAAT TAAATATAACTGCACCCAAGGACTCGCGCTTAGCAAAATCCAAG 8898

Query 8893 GCTGAGGATTTCCAACAAAAGAGTGGGCTAAGGTAACACTGTCGATGCTTATAGAGACA 8952  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 8899 GCTGATGATTTCCAGCAAGAGGAAGTCCAAA AATTACCTTGTGACACTGATCAAGAGC 8958

Query 8953 GCAGAGTATTTGGTCCAAAACAGGACATTAAGAACATCGATGATTCAAGATTAAGAGCTTTA 9012  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 8959 GCAGAACACTGGGCGAGACAAGACATCAGAACCATAGAGGATTCAAAAATTAAGAGCATTG 9018

Query 9013 TTGACCCTTTGTGCTGTTATGACGCGCAAATTTTCAAAATCTCAACTTAGCTTGTGCTATGT 9072  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 9019 TTGACTCTATGTGCTGATGACGAGGAAATTC TCAAAATCCCAGCTGAGTCTTTTATGT 9078

Query 9073 GAAAGCCTTACGGCGAGAAGGACTTGGTCAAGACCAATCAGAGTCAGTTCTGGAGGTA 9132  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 9079 GAGACACACCTAAGGCGGAGGGGCTTGGGCAAGATCAGGCAGAACCCGTTCTCGAAGTA 9138

Query 9133 TATCAACGCTTACACAGCGATAAAGGTGGGAATTTTCGAGGCAGCATAATGGCAGCAGTGG 9192  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 9139 TATCAACGATTACACAGTGATAAAGGAGGCAGTTTGAAGCTGCACATATGGCAACAATGG 9198

Query 9193 GATCGGCAATCATTTGATAATGTTTCATAACAGCATTTTTAAATATTGCATTACAATTACCA 9252  
| | | | | | | | | | | | | | | | | | | | | | | | | |  
Sbjct 9199 GACCGACAATCCCTAATTATGTTTATCACTGCATTTCTGAATATTGCTCTCCAGTTACCG 9258

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΑΘΑΝΑΣΙΟΣ ΠΑΥΛΙΔΗΣ

Query 9253 TGTGAGAGTTCATCTGTGTGTTATTTTCAGGTTTGAGAATGCTGATACCCAGTCGGAAGCC 9312  
 ||||| ||||| ||||| ||| ||| ||||| || ||||| || ||| ||| ||| |||  
 Sbjct 9259 TGTGAAAGTTCGTGCTGCTGTTGTTTCAGGGTTAAGAACATTGGTTCCTCAATCAGAT--- 9315

Query 9313 ACTGAGGTTGT---AACCCCTCCGAAACCTGCACATGGTCAGAAGGAGGAAGTTCCCAT 9369  
 | ||||| | ||||| | | | | ||||| || | ||| || ||| |  
 Sbjct 9316 AATGAGGAAGCTTCAACCAACCCGGGACATGCTCATGGTCTGATG-AGG--GTACCCCT 9372

Query 9370 TGAAGCCCCAAATCACAAGGCGAGCTAAAAAATCCCTTTTGAACATGC-ATAACATCACA 9428  
 | ||| ||||| ||||| | || | ||| | | | |||||  
 Sbjct 9373 T-----AAT---AAGGCTGACTAAAACA---CTATATAACCTTCTACTTGATCACA 9417

Query 9429 -TACAATTTCAAAGGCATTGGAATAAATGGTGAATTCAGGAAGATTAGTGTTTGCCCTCA 9487  
 ||| | | | | | | | ||||| | ||||| | | | | | |  
 Sbjct 9418 ATAC---TCCGTATACCTATCATCATAT----ATTTAATCAAGACGA---TATCCTTTAA 9467

Query 9488 AAATCAGATCCGAGCAATAATCA-TCTACTCTACAGCCAGTTAAT-----TT 9533  
 || | | | | | ||||| ||| | | | | ||||| ||  
 Sbjct 9468 AACTTAT-TCAGTACTATAATCACTCTCGTTT-CA---AATTAATAAGATGTGCATGATT 9522

Query 9534 ---CTAATATAAAGGTTAAAAAA-ATGCTGCAGGCCAGCTATTGTTCCAC--AGGTCCCA 9587  
 ||||| || | | | ||| | | | | | | | | | | | | |  
 Sbjct 9523 GCCCTAATATA----TGAAGAGGTATGATACAACCCTAACAGTGATCAAAGAAAATCATA 9578

Query 9588 AT-TCTTCTTGTTA--AATTGTAGGAGCTA-GCACAACTGATGCAATTAATGATACTAG 9643  
 || | | | | | ||| | | | | ||| | | | ||||| || |  
 Sbjct 9579 ATCTCGTATCGCTCGTAATATAACCTGCCAAGCATACTCTTGACACA-AAGTGATTCTTG 9637

Query 9644 TATATACAATGCCACCAAC--TTAATTCTA-AG-ATTTTGAT--ATC-----TCGGA 9690  
 || | | ||| || | || | ||||| || | | | | ||| || |  
 Sbjct 9638 TACACA-AAT-----AATGTTTACTCTACAGGAGGTAGCAACGATCCATCCCATCAA 9690

Query 9691 AATTCAAAATTAATGCTACGT--TATTGAT-TC----AAT--TAAGAAAA--GACAAT 9739  
 || | | ||| ||| || | | | ||||| || ||| ||||| ||||| |||  
 Sbjct 9691 AAATAAGTATTTTCATG--ACTTACTAATGATCTCTTAAAATATTAAGAAAACTGAC--- 9745

Query 9740 GGACCATCAAAATTAGTTCAATACCTGAACCTAATGCAC----TTATAGAAACAGGAGAAC 9795  
 ||| ||| ||||| || | | | | || | | | | | | | |  
 Sbjct 9746 GGAACAT--AAATCTTT--ATGCTTCAAGCTGTGGAGGAGGTGTTTGGTATTGGCTATT 9801

Query 9796 CAGCCAGACAGCAGACAAATAACAA---TGAACCACAATATGTACTGCTATAATGAAGT 9852  
 | | | | ||| ||||| || | | ||||| | | | | | | | |  
 Sbjct 9802 --GTTATATTACAATCAA-TAACAAGCTTGTA--AAAATATTGTTCTTGTTCAGAGGT 9856

Query 9853 ---TCGTTAATTCAAA-----AACAAATGATGAAGATTAATGCAGATGTCTAA--AGGAT 9902  
 | || | ||| ||| ||||| ||||| ||||| ||||| ||||| |||||  
 Sbjct 9857 AGATTGTGACCGGAAATGCTAAACTAATGATGAAGATTAATGCGGAGGTCTGATAAGAAT 9916

Query 9903 AAACACTCCATGCATCAGTGTATAATTGGGCTCTGTAGAAAATCTTCATCTCCTCCAAC 9962  
 |||| | || ||||| ||| || | | || | ||||| ||||| |  
 Sbjct 9917 AAACCTT--ATTATTCAGA-TTA-----GGCCCAAGAGGCATTCTTCATCTCCTTTTAG 9968

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΑΘΑΝΑΣΙΟΣ ΠΑΥΛΙΔΗΣ

```

Query 9963 CTACCTCAAAGAAGGATTTTACCG-----CGATT---GGGA-GT-TATAACGACAATAGG 10012
|      ||| | ||| | | | | | | | | | | | | | | | | |
Sbjct 9969 C-----AAAGTACTATTTTCAGGGTAGTCCAATTAGTGGCACGTCTTTTAGCTGTATAT- 10021

Query 10013 GACAACCACCTTTGACACTAGCCAAGCTTGTCGTGGGCACACAGCATTTTATCTTGCAAC 10072
|| | | | | | | | | | | | | | | | | | | | | | |
Sbjct 10022 --CAGTCGCCCTTGAGATACGCCACAAAAGT-GT---CTCTAAGC---TAAATTGGTCT 10071

Query 10073 GTCGACATTCCCAT-CAATCTG--AGGAGTAACAGCTATCAAAAC-AACGCATATGTAGA 10128
|| ||| |||| | | | | | | | | | | | | | | | | |
Sbjct 10072 GTACACAT-CCCATACATTGTATTAGGGCAATAA-TATCTAATTGAACTTAGCCGTTA 10129

Query 10129 CATTGTCGGTAATAGTACTGC-CTAAGACAACATTTATAAT--AACAGTTGGAATTCAT 10185
| | | | | | | | | | | | | | | | | | | | | | |
Sbjct 10130 AAATTTAGTGCATAAATCTGGGCTAACACCACCAGGTCACCTCCATTGGCTGAAAAGAAG 10189

Query 10186 TTT---TTCACCCAAGCTATCTCAAGT----TAACAGTTGAAACA--GGAC--TCGACC 10234
|| | | | | | | | | | | | | | | | | | | | | |
Sbjct 10190 CTTACSTACAACGAACATCACTTTGAGCGCCCTCACAATTAATAAATAGGAACGTCGTT 10249

Query 10235 CAGGACAACCTCCGGATACGTAACATAAGAAAAGAACAACCCCTTGACCCAGAGTGA----A 10290
|| ||| | | | | | | | | | | | | | | | | | | |
Sbjct 10250 CA--ACAA-TCGAG---CGCAAGGTTTCAAG-GTTGAAC---TGA---GAGTGTCTAGA 10295

Query 10291 CAAGCTCATACTATCAAGGCT--AATCCTCGGGCTG-CCTG-GAGTCCACAATGGCCAA 10346
|| | | | | | | | | | | | | | | | | | | | | |
Sbjct 10296 CAA---CAAAATATTGATACTCCAGACACCAAGCAAGACCTGAGAAAAACCATGGCTAA 10352

Query 10347 GGCTACTGGGAGGTACAACCTTATCTCCCCAAAGAAAGATCTTGAAAAAGGGCTGGTTCT 10406
|||| | | | | | | | | | | | | | | | | | | | | |
Sbjct 10353 AGCTACGGGACGATACAATCTAATATCGCCCAAAAGGACCTGGAGAAAGGGTTGTCTT 10412

Query 10407 GAATGACCTTTGCACTCTCTCAGTGGCCAGACGGTCCAGGGATGGAAGGTTACCTGGGC 10466
| |||| | | | | | | | | | | | | | | | | | | | |
Sbjct 10413 AAGCGACCTCTGTAACCTTCTAGTTAGCCAAACTATTCAGGGTGGAAAGGTTATTGGGC 10472

Query 10467 TGGGATTGAATTTGATGTTACACAGAAAGGGATGGCCTTATTGCACAGGCTCAAGACCAG 10526
||| |||| |||| | | | | | | | | | | | | | | | |
Sbjct 10473 TGGTATTGAGTTGATGTGACTCACAAAGGAATGGCCCTATTGCATAGACTGAAAATAA 10532

Query 10527 TGATTTTGCTCCAGCCTGGTCAATGACCAGGAACCTTATTTCCACATCTCTTTCAAACCC 10586
||| |||| | | | | | | | | | | | | | | | | | |
Sbjct 10533 TGACTTTGCCCTGCATGGTCAATGACAAGGAATCTCTTTCCTCATTTATTCAAATCC 10592

Query 10587 GAACTSTACAATTGAGTCGCCACTTTGGGCACTGCGGGTCATACTAGCAGCAGGTATTCA 10646
||| | | | | | | | | | | | | | | | | | | | | |
Sbjct 10593 GAATTCACAATTGAATCACCCTGTGGGCATTGAGAGTCATCCTTGCAGCAGGGATACA 10652

Query 10647 AGATCAGSTAATTGATCAATCGTTGATCGAACCTTGGCAGGAGCGCTAGGCTTAATTGC 10706
|| |||| |||| | | | | | | | | | | | | | | | |
Sbjct 10653 GGACCAGCTGATTGACCAGTCTTTGATTGAACCCTTAGCAGGAGCCCTGGTCTGATCTC 10712
    
```

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΑΘΑΝΑΣΙΟΣ ΠΑΥΛΙΔΗΣ

Query	10707	TGATTGGCTTCTTACTACTGGAACAAACCACCTTTCAA-ATGCGCACACAACAGGCTAAGG	10765
Sbjct	10713	TGATTGGCTGCTAACCAACCAACTAACC-TTCAACATGCGAACACAACGTGTCAAGG	10771
Query	10766	AGCAACTAAGTCTAAAAATGTTGTCCTGGTGCGATCAAACATCCTAAAGTTCATCAACC	10825
Sbjct	10772	AACAATTGAGCCTAAAAATGCTGTCGTTGATTGATCGATCAATATTCTCAAGTTATTAACA	10831
Query	10826	AACTAGATGCACTACATGTTGTGAATTACAATGGACTTCTCAGTAGCATTGAAATTGGCA	10885
Sbjct	10832	AATTGGATGCTCTACATGTCGTAACACAACGGATTGTTGAGCAGTATTGAAATTGGAA	10891
Query	10886	CCAAAAGCCATACAATTATAATTACCCGGACAAATATGGGTTTTTTGGTAGAGTTGCAAG	10945
Sbjct	10892	CTCAAAATCATACAATCATCATAACTCGAACAACATGGGTTTTCTGGTGGAGCTCCAAG	10951
Query	10946	AGCCTGACAAATCAGCCATGAACACCAGAAAACCAGGACCAGTCAAATTCTCCCTCCTCC	11005
Sbjct	10952	AACCCGACAAATCGGCAATGAACCGCATGAAGCCTGGGCCGGCGAAATTTCCCTCCTTC	11011
Query	11006	ATGAATCAACCTTGAAGACACTTGCTAAAAAACCTGCGACCCAGATGCAAGCACTAATCT	11065
Sbjct	11012	ATGAGTCCACACTGAAAGCATTTACACAAGGATCCTCGACACGAATGCAAAGTTTGATTC	11071
Query	11066	TAGAATTCAATAGTCTCTCGCTATTTAACTCAACT-----CATCAAAATG--CTAACT	11117
Sbjct	11072	TTGAATTAAATAGCTCTCTTGTCTATCTAACTAAGGTAGAATACTTCATATTGAGCTAACT	11131
Query	11118	TGTGATCCTTAAGCTGCACCTTAGACTTTTGATAAGAATACTAACTATTGATGATTGTCT	11177
Sbjct	11132	CAT-AT---ATGCTG-ACTCAATAGTTATCTTGACATCTCTG-CT-TTCATAATCAGAT	11183
Query	11178	TTGACATGAGGATAAGAAC--ACTGCCCAT-----TAGATAGATGGGGTTCACCATTAAT	11230
Sbjct	11184	AT---ATAAGCATAATAAATAAATACTCATATTTCTTGATA-ATTTGTTAACCACAGAT	11239
Query	11231	ACACAAATACC-ATCATGTTAACAGCAGTTAGATCCCTCAAGTATATCAAGT-TCATT	11288
Sbjct	11240	AAATCCTCACTGTAAGCCAGCTTCCA--AGTTGACACCCTTACAAAACCAGGACTCAGA	11297
Query	11289	CTACCSTTTGCATTGTCACTCTAATTAATCACCTGATACAATT-----ATGTTAAT	11340
Sbjct	11298	ATCCCTCAAACAA-GAGATCCAAGACAA-CATC--ATAGAATTGCTTTATTATATGAAT	11353
Query	11341	TAGCTAGATTCTCTCATTTTTAGACTTGTTTGTAGAATAATTGATCATCCACTTGATTA	11400
Sbjct	11354	AAGC---ATTTATCACCAGAAATCCTATATACTA-AATGGTTAATTGTA-ACT-GAACC	11407
Query	11401	CACATCCAAC TAGGGTCTAG--TTCATAGATTGCTAATAATCTTTAGTTCAATACT---A	11455
Sbjct	11408	CGCAGGTCACATGTGT-TAGGTTTACAGATT-CTATATATTACTAACTCTATACTCGTA	11465

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

Query	11456	ATGACAAAGAGAT---TAGATTAGCTATA--GCTTGAGGAAGATTAAGAAAAAGTGTCTG	11510
Sbjct	11466	ATTAACATTAGATAAGTAGATTAAGAAAAAGCCTGAGGAAGATTAAGAAAAACTGCTTA	11525
Query	11511	TGGGGTCTTTCGGTGTAGAAGGGCACACAGCCATAATCTTCCTCTTTATACAACATGGC	11570
Sbjct	11526	TTGGGTCTTTCGGTGTTTAGATGAAGCAGTTGAAATCTTCCTCTTGATATTAAATGGC	11585
Query	11571	TACACAACATACGCAATATCCAGACGCAAGTTATCATCACCTATAGTTTATAGATCAGTG	11630
Sbjct	11586	TACACAACATACCCAATACCCAGACGCTAGTTATCATCACCAATTGTATTGGACCAATG	11645
Query	11631	TGATCTGTCACTCGTGTCTTGTTGGATGTATTCCGCATACTCCTTAAATCCCAACTAAA	11690
Sbjct	11646	TGACCTAGTCACTAGAGCTTGCGGGTTATATTCATCATACTCCCTTAAATCCGCAACTAG	11705
Query	11691	GAACTGTAGACTACCGAAACATATATACCGACTAAAATATGACACCACTGTTACAGAGTT	11750
Sbjct	11706	CAACTGTAAACTCCCGAAACATATCTACCGTTTGAAATACGATGTAAGTGTACCAAGTT	11765
Query	11751	TTTGAGTGATGTGCCGGTAGCAACATTGCCAGCGGATTTTTAGTACCTACATTTCTTAG	11810
Sbjct	11766	CTTGAGTGATGTACCAGTGGCGACATTGCCCATAGATTTTCATAGTCCAGTCTTCTCAA	11825
Query	11811	GACTCTATCAGGAAATGGTTCTGTGCCAATTGATCCAAAATGCAGTCAATTTTAGAAGA	11870
Sbjct	11826	GGCACTGTCAAGCAATGGATTCTGTCTCTTGGCCGGTGCCAACAGTTCCTTAGATGA	11885
Query	11871	AATGTCAATTACTCTACAAGATATTCGCTTCCTAAACTATTACCTCAATCGAGCCGG	11930
Sbjct	11886	AATCATTAAGTACACAATGCAAGATGCTCTCTCTTGAATATATATCTCAAAAATGTGGG	11945
Query	11931	AGTGCATAACGATCATGTGGATAGGGATTTGGACAAAAAATTCGCAATCT--AATTTGC	11988
Sbjct	11946	TGCTCAAGAAGACTGTGTGATGAACACTTCAAGAGAAAAT-CTTA-TCTTCAATTCAG	12003
Query	11989	GACAATGAGGTTTACATCAAAATGTTTCACTGGTATGATCTTGCAATCTAGCACGTAGA	12048
Sbjct	12004	GGCAATGAATTTTACATCAAAATGTTTTCTGGTATGATCTGGCTATTTTAACTCGAAGG	12063
Query	12049	GGCGACTAAATAGAGGGAATAATCGCTCAACATGGTTTGCAAGTGATAATTTGGTAGAT	12108
Sbjct	12064	GGTAGATTAATCGAGGAACTCTAGATCAACATGGTTTGTTCATGATGATTTAATAGAC	12123
Query	12109	ATCCTAGGTTATGGAGATTATATTTTTGGAATAACCATTATCACTACTACCAGTGGAT	12168
Sbjct	12124	ATCTTAGGCTATGGGACTATGTTTTTGGAAAGATCCCAATTTCAATGTTACCAGTGAAC	12183
Query	12169	ACACAAGGCCCTCCCATGCGCAAGGACTGGTATCATGAATCGGTTTTCAAGGAGGCT	12228
Sbjct	12184	ACACAAGGAATCCCCATGCTGCTATGGACTGGTATCAGGCATCAGTATTCAAAGAAGCG	12243



Table with 4 columns: Label (Sbjct/Query), ID, DNA Sequence, and ID. Contains 30 pairs of sequence entries for IDs 13019 through 13799.

Query	13844	AGCAAGAATTAAGTTCTGAAGATAATGCCGCTAGAGTAGCTGCTAGCTTAGCAAAAGTCA	13903
Sbjct	13859	AGCAGGAACAGAGCGCCGAAGACAATGCAGCGAGGGTGGCCGCCAGCCTAGCAAAAGTTA	13918
Query	13904	CAAGTGCCTGCGGCATCTTTTAAACCTGATGAACTTTTGTCTACTCAGGTTTCATTT	13963
Sbjct	13919	CAAGTGCCTGTGGAATCTTTTAAACCTGATGAACTTTTGTACATTCAGGTTTATCT	13978
Query	13964	ATTTTGGCAAAAACAATATTTGAATGGAGTACAATTACCTCAATCACTGAAAACCTGCTA	14023
Sbjct	13979	ATTTTGGAAAAACAATATTTGAATGGGTCCAATTCCTCAGTCCCTTAAAACGGCTA	14038
Query	14024	CTAGAATTGCACCCTTGTCAGATGCTATCTTTGATGATCTTCAAGGGACACTAGCTAGCA	14083
Sbjct	14039	CAAGAATGGCACCATTGTCTGATGCAATTTTGGATGATCTTCAAGGGACCCTGGCTAGTA	14098
Query	14084	TAGGCACGGCTTTTGAAGATCTATCTCCGAACTAGGCACGTAGTCCCTGTAGAGTAG	14143
Sbjct	14099	TAGGCACTGCTTTGAGCGATCCATCTCTGAGACACGACATATCTTTTCCTTGCAAGATAA	14158
Query	14144	CAGCTGCATTCCATACCTTTTTTCGGTAAAGATCTTACAATATCATCATCTGGCTTCA	14203
Sbjct	14159	CCGCAGCTTTCCATACGTTTTTTTCGGTGAAGATCTTGCAATATCATCATCTCGGGTCA	14218
Query	14204	ACAAGGAACAGACCTGGGTCAATGTCAATTAAGCAAGCCATTAGATTTTGGAACTATAA	14263
Sbjct	14219	ATAAAGTTTTTACCTTGGACAGTTAACTCCGCAACCTCTGGATTTTCGGAACAATAT	14278
Query	14264	CTTTGGCCTTGGCAGTACCACAAGTCTTGGGTGGCTTATCATTCTAAATCCAGAAAAAT	14323
Sbjct	14279	CATTGGCACTAGCGGTACCAGGTGCTTGGAGGGTTATCCTTCTTGAATCCTGAGAAAT	14338
Query	14324	GTTTTTATAGAAATCTGGGTGATCCTGTACTTCAGGGCTGTTTCAGCTCAAGACATATC	14383
Sbjct	14339	GTTTCTACCGGAATCTAGGAGATCCAGTTACCTCAGGCTTATTCAGTTAAAACCTTATC	14398
Query	14384	TTCAAATGATCCACATGGATGATTTGTTTTTACCTTTGATCGCAAAGAACCAGGGAAC	14443
Sbjct	14399	TCCGAATGATGAGATGGATGATTTATTCTTACCTTTAATTCGAAGAACCCTGGGAAC	14458
Query	14444	GTAGCGCAATTGACTTTGTGTAAACCTTAGTGGGTAAACGTACCGGGGTCACAGGATT	14503
Sbjct	14459	GCACTGCCATTGACTTTGTGCTAACTCTAGCGGATTAATGTCCCTGGGTCGCAAGACT	14518
Query	14504	TGACATCCTTCTACGTCAGATAGTGCGCCGAACAATTACTTAAGTGCTAAAAATAAAT	14563
Sbjct	14519	TAACTTCATTTCTGCGCCAGATTGTACGCAGGACCATCACCTAAGTGCGAAGCAAC	14578
Query	14564	TAATAAACACTTTGTTCCATTTCTCTGCTGATTTAGAAGATGAAATGGTTTGCAAATGGT	14623
Sbjct	14579	TTATTAAACCTTATTTTCATGCGTCAGCTGACTTCGAAGACGAAATGGTTTGTAAATGGC	14638
Query	14624	TGCTTTCTTCTACACCAGTCATGAGTAGGTTTGCCGCCGATATATTTTCTCGCACTCCCA	14683



Sbjct	14639 TATTATCATCAACTCCTGTTATGAGTCGTTTTGCGGCCGATATCTTTTCACGCACGCCGA	14698
Query	14684 GTGGGAAACGTTTACAGATCTTAGGTACCTGAAGGGACTAGAACATTGTTAGCCTCTA	14743
Sbjct	14699 GCGGGAAGCGATTGCAAATTTCTAGGATACCTGGAAGGAACACGCACATTTATAGCCTCTA	14758
Query	14744 AAATTATAAATCATAAATACTGAGACACCTATCCTAGATCGATTGAGGAAAATTACGCTGC	14803
Sbjct	14759 AGATCATCAACAATAATACAGAGACACCGTTTTGGACAGACTGAGGAAAATAACATTGC	14818
Query	14804 AAAGGTGGAGCCTGTGGTTTTAGTTATCTCGACCACTGTGATCAAGTTCTGGCTGATGCC	14863
Sbjct	14819 AAAGGTGGAGCCTATGGTTTTAGTTATCTTGATCATTGTGATAATATCCTGGCGGAGGCTT	14878
Query	14864 TAACTCAGATAACCTGCCTGTGGACTTAGCACAGATTTCTCGCGAGTACACTGGGCAC	14923
Sbjct	14879 TAACCCAAATAACTTGCACAGTTGATTTAGCACAGATTTCTGAGGGAATATTCATGGGCTC	14938
Query	14924 ACATACTAGAGGGAAGGCAGCTCATTTGGAGCAACACTTCTTGTATACTAGAACAACTAA	14983
Sbjct	14939 ATATTTTAGAGGGAAGACCTCTATTGGAGCCACACTCCCATGTATGATTGAGCAATTCA	14998
Query	14984 ATGTCATCTGGCTCAAACCATATGAGCATTGCCCTAAATGTGCAAAGTCA--GCAAACCC	15041
Sbjct	14999 AAGTGTTTTGGCTGAAACCTACGAACAATGCCGCGAGTTCAAATGCAAAGCAA--CC	15056
Query	15042 TAAAGGGGAACCTTTTGTTCCTATTGCAATTAAAAACATGTAGTAAGTGCTTGGCCTGA	15101
Sbjct	15057 AGGTGGGAAACCATTCGTGTACAGTGGCAGTCAAGAAACATATTGTTAGTGCATGGCCGAA	15116
Query	15102 TCAATCAGACTTAGTTGGACAATTGGAGATGGCATCCCTTATATCGGATCTCGAACAGA	15161
Sbjct	15117 CGCATCCCGAATAAGCTGGACTATCGGGGATGGAATCCCATACATTGGATCAAGGACAGA	15176
Query	15162 GGATAAGATTGGGCAGCCAGCCATCAAACCAAATGCCCTTACGAGCCTTACGTGAAGC	15221
Sbjct	15177 AGATAAGATAGGACAACCTGCTATTAAACCAAATGTCTTCCGAGCCTTAAGAGAGGC	15236
Query	15222 AATTGAGTTGACATCAAGATTGACTTGGGTTACTCAAGGTGGAGCAACAGCGACTTACT	15281
Sbjct	15237 CATTGAATTGGCGTCCCGTTTAAACATGGGTAACCTCAAGGCAGTTCGAACAGTACTTGGCT	15296
Query	15282 AGTTAAACCTTCATAGAAGCAGAGTAAATTTAAGCGTACAGGAAATCTCCAAATGAC	15341
Sbjct	15297 AATAAAACCATTTTGGAGCAGAGTAAATTTAAGTGTTCAGAAATACTTCAAATGAC	15356
Query	15342 ACCTTCTCATTACTCCGGCAACATTTGTCATCGATATAATGATCAATATAGTCCACACTC	15401
Sbjct	15357 CCCTTCACTTACTCAGGAAATATTGTTACAGGTACAACGATCAATACAGTCCCTCATTC	15416
Query	15402 ATTTATGGCAAATAGGATGAGTAATTCTGCTACTAGGTAGTTGTTTCGACAAACACTCT	15461
Sbjct	15417 TTTCATGGCCAATCGTATGAGTAATTCAGCAACGCGATTGATTGTTTCTACAAACACTTT	15476

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΑΘΑΝΑΣΙΟΣ ΠΑΥΛΙΔΗΣ

Query 15462 TGGAGAATTTTCAGGAGGAGGTCAGTCAGCAAGAGATAGTAATATTATCTTCCAGAATGT 15521  
 || || ||||| || ||||| ||| | || ||||| ||||| |||||  
 Sbjct 15477 AGGTGAGTTTTTCAGGAGGTGGCCAGTCTGCACGCGACAGCAATATTATTTCCAGAATGT 15536

Query 15522 CATTAATTTTGCTGTTGCACTTTTTGATCTACGATTTAGGAAC---GTGGCTACTTCTTC 15578  
 || |||| ||| ||||| || |||| | ||||| ||| | ||||| |  
 Sbjct 15537 TATAAATTATGCAGTTGCACTGTTTCGATATTAATTTAGAAACACTGAGGCTACAGAT-- 15594

Query 15579 TATACAACATCATCGGGCTCATCTTCATTTGTCAAAGTGTTCACGCGAGAGGTTCCAGC 15638  
 || |||| || ||||| ||||| ||||| | | ||||| || || || |||||  
 Sbjct 15595 -ATCCAATATAATCGTGCTCACCTTCATCTAACTAAGTGTTCACCCGGGAAGTACCAGC 15653

Query 15639 CCAATATTTAGTTTATACATCAACATTGCCATTGGACCTTACACGGTATCGGGATAATGA 15698  
 || ||||| || ||||| ||||| || || | ||| | || || || || || ||  
 Sbjct 15654 TCAGTATTTAACATACACATCTACATTGGATTTAGATTTAACAAAGATACCGAGAAAACGA 15713

Query 15699 GTTGATTTACGATGACAATCCATTAAGAGGTGGTTTAAATGCAATCTTCTTTTGATAA 15758  
 ||||| || ||||| ||| ||| || | ||||| || || || |||||  
 Sbjct 15714 ATTGATTTATGACAGTAATCCTCTAAAAGGAGGACTCAATTGCAATATCTCATTCGATAA 15773

Query 15759 TCCGCTTTTCAAGGGCCAGAGACTTAAACATAAATTGAAGAAGACTTGATTAGACTACCTTA 15818  
 ||| |||| | || | || ||||| || ||||| || | ||| ||||| ||| |  
 Sbjct 15774 TCCATTTTCCAAGGTAAACGGCTGAACATTATAGAAGATGATCTTATTCGACTGCCTCA 15833

Query 15819 CTTATCAGGATGGGAGCTAGCTAAAACCTGTTATCCAATCTATAATTTCTGACAGCAACAA 15878  
 ||||| ||||| ||||| || || | || ||||| || ||||| || ||||| ||  
 Sbjct 15834 CTTATCTGGATGGGAGCTAGCCAAGACCATCATGCAATCAATTATTTAGATAGCAACAA 15893

Query 15879 TTCATCAACGGATCCAATCAGTAGTGGGAAACACGATCATTCACCACTCACTTCTTGAC 15938  
 ||||| || || ||||| || ||||| ||||| ||||| || || ||||| ||  
 Sbjct 15894 TTCATCTACAGACCCAATTAGCAGTGGAGAAACAAGATCATTCACTACCCATTTCTTAAC 15953

Query 15939 ATATCCTAAGATTGGACTACTATATAGTTTTGGTGCCTCATCAGTTATTATCTAGGCAA 15998  
 ||||| ||||| ||||| || || ||||| || || | ||||| ||||| |||||  
 Sbjct 15954 TTATCCCAAGATAGGACTTCTGTACAGTTTTGGGCCTTTGTAAGTTATTATCTTGCAA 16013

Query 15999 CACCATTATTAGAACCAAAAAATTGACTCTTAACAACCTCATATATTACCTAGCTACTCA 16058  
 || ||| || | || || ||||| || ||| ||||| || ||||| || ||||| ||  
 Sbjct 16014 TACAATCTTCGGACTAAGAAATTAACACTTGACAATTTTTATATTACTTAACTACTCA 16073

Query 16059 AATACATAATTACCTCATCGCTCGTTGAGAATCCTTAAACCTACTTTGAAACACGCTAG 16118  
 ||| ||||| ||||| ||||| ||||| ||||| || || || ||||| || ||  
 Sbjct 16074 AATTCATAATCTACCACATCGCTCATTCGCAATACTTAAGCCAACATTCAAACATGCAAG 16133

Query 16119 TGTATCTCGAGATTAATAAGTATTGACTCTCACTTCTCAATTTATATTGGAGGAAGTGC 16178  
 ||||| || | ||||| ||||| ||||| || || || ||||| || || || |||||  
 Sbjct 16134 CGTTATGTCACGGTTAATGAGTATTGATCCTCATTTTTCTATTTACATAGGCGGTGCTGC 16193

Query 16179 TGGTATCGAGGACTTCCGATGCGGCAAGATTGTTTCTTAGAACTGCCATTACTGTCTT 16238  
 ||||| ||||| || ||||| || || ||| | ||||| ||||| | ||  
 Sbjct 16194 AGGTGACAGAGGACTCTCAGATGCGGCCAGGTATTTTTGAGAACGTCCATTTCTCTTT 16253

Query 16239 CCTTCAATTCGTTAGAAAAGTGATAGTTGAACGCAAGACAGCTATTCCACTGTGGGTCTAT 16298

Sbjct	16254 TCTTACATTTGTAAAAGAATGGATAATTAATCGCGGAACAATTGTCCTTTTATGGATAGT	16313
Query	16299 CTACCCTCTAGAAGGTCAAAGTCCTAGTCCGATCAACAGTTTTCTACACCACGTCATCGC	16358
Sbjct	16314 ATATCCGCTAGAGGGTCAAAACCCAACACCTGTGAATAATTTCTCTATCAGATCGTAGA	16373
Query	16359 ATTGTTGCAACATGAGTCCTCCACGATCATGTTTGTGCTGCAGAAGCCCACAGTCGAGT	16418
Sbjct	16374 ACTGCTGGTGCATGATTCATCAAGACAACAGGCTTTTA----AAACTACCATAAGTGATC	16429
Query	16419 GGAGACATTT---GATAATTTAGTTTTATATGTGTAAAAGCACAGCAAGTAACTTCTTTC	16474
Sbjct	16430 ATGTACATCCTCACGACAATCTGTTTACACATGTAAGAGTACAGCCAGCAATTTCTTCC	16489
Query	16475 ATGCTTCATTAGCATACTGGAGAAGTCGATCTAAAAATCAAGACAAAAGAGAGATGACAA	16534
Sbjct	16490 ATGCATCATTTGGCGTACTGGAGGAGCAGACACAGAAACAGCAACCGAAAATACTTGGCAA	16549
Query	16535 -AGA-TATT-ATCTT--TGACGCAAACGGAAAAGAAAATTCAT-----TCGGC	16578
Sbjct	16550 GAGACTCTTCAACTGGATCAAGCACAAACAACAGTGATGGTCATATTGAGAGAAGTCAAG	16609
Query	16579 TATACAGCAC-----ATCCA---GAAAGCACTG--CTGTTCTTGGTTCCCTCCAGACCAG	16628
Sbjct	16610 AACAAACCACCAGAGATCCACATGATGGCACTGAACGGAATCTAGT--CCTACAAATGAG	16667
Query	16629 CCTTGCTCCACCTCCATCTGCTGACGAGGCTACATATGATAGGAAAAACAAAGTTTTGAA	16688
Sbjct	16668 CCATGAAATAAAAAGAAC---GACAATTCACAA--GAAAACACGCACCAGGGTCCGTC	16721
Query	16689 AGCTTCCAGACCTGGC--AAGT-ATTC--CCAGAA-TACAACCAAAGCCCCACCCAACCAA	16743
Sbjct	16722 G--TTCAGTCCTTTCTAAGTGACTCTGCTTGTGGTACAGC-AAATCCAAA---ACTAA	16774
Query	16744 ACCAGTTGTCG--CGATGTATCTCCCAATATCA-----CAGGCACAGATGGGTGCCCTTC	16796
Sbjct	16775 A---TT-TCGATCGATCGAGA-CACAATGTGAAATTCAGGATCATAA-----CTCGGC	16823
Query	16797 TGCCAAATGAGGGTTCTAACAGCAATAACAATAATTTAGTCTCGCACAGAATTGTACTGCC	16856
Sbjct	16824 ATCCAA-GAGGG---AAGGTCATCAA--ATAAT-----CTCACACCGTCTAGTCCTACC	16871
Query	16857 GTTTTTTACATTGTCTCATAATTATAA-CGAAAGACCCTCTATCAGAAAGTCTGAGGGGA	16915
Sbjct	16872 TTTCTTTACATTATCTCA-AGGGACACGCCAATTAACGTCATCCAATGAGTCACAAACC-	16929
Query	16916 CAACA-GAGATTGTAAGG--CTTACTCGGCAGCTGAGGGCAATACCAGACACCACAATAT	16972
Sbjct	16930 CAAGACGAGATATCAAAGTACTTAC--GGCAATTGAGATCCGTCATTGATACCACAGTTT	16987
Query	16973 ATTGCCGCTTACGGGAATAGTTTCTTCAATGCACTATAAGCTCGATGAAGTCCTTTGGG	17032
Sbjct	16988 ATTGTAGATTTACCGGTATAGTCTCGTCCATGCATTACAAACTTGATGAGGTCCTTTGGG	17047

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΑΘΑΝΑΣΙΟΣ ΠΑΥΛΙΔΗΣ

Query	17033	AATTTGATAATTTTAAAGTCTGCTATAACACTTGCCGAAGGTGAAGGTCGGGTGCATTAC	17092
Sbjct	17048	AAATAGAGAGTTTCAAGTCGGCTGTGACGCTAGCAGAGGGAGAAGGTGCTGGTGCCTTAC	17107
Query	17093	TCTTATTACAAAAATATAAAGTAGAAACCTTGTTTTTTAATACACTAGCCACAGAACACA	17152
Sbjct	17108	TATTGATTTCAGAAATACCAAGTTAAGACCTTATTTTTCAACACGCTAGCTACTGAGTCCA	17167
Query	17153	GCATTGAAGCAGAAATATTTCTGGAATAACTACACCAAGAATGCTTCTCCCTATTATGT	17212
Sbjct	17168	GTATAGAGTCAGAAATAGTATCAGGAATGACTACTCCTAGGATGCTTCTACCTGTTATGT	17227
Query	17213	CTAGGTCCATGGTGGACAAATAAAAGTCACTTTAAACAATTCTGCAAGCCAGATTACCG	17272
Sbjct	17228	CAAAATTCATAATGACCAAATTGAGATTATCTTAACTCAGCAAGCCAAATAACAG	17287
Query	17273	ATATTACTAATCCAAGTTGGTTGGCAGACCAAAAATCTAGGATCCCTAAGCAAGTAGAGA	17332
Sbjct	17288	ACATAACAAATCCTACTTGGTTTAAAGACCAAGAGCAAGGCTACCTAAGCAAGTCGAGG	17347
Query	17333	TTATAACCATGGATGCTGAAACAACAGAAAACATTAATCGGTCAAAATGTACGAAGCAG	17392
Sbjct	17348	TTATAACCATGGATGCAGAGACAACAGAGAATATAAACAGATCGAAATGTACGAAGCTG	17407
Query	17393	TCCAACAGCTGATTGTCTCACATATTGATCCGAATGCACTCAAAGTTGTGGTCTTAAAG	17452
Sbjct	17408	TATATAAATGATCTTACACCATATTGATCCTAGCGTATTGAAAGCAGTGGTCCTTAAAG	17467
Query	17453	TTTTCTAAGTGACATTTGATGGAATCCTATGGCTGAATGATAACCTTACCCCTTTGTTG	17512
Sbjct	17468	TCTTTCTAAGTGATACTGAGGGTATGTTATGGCTAAATGATAATTTAGCCCGTTTTTTG	17527
Query	17513	GGCTGGGTACTTGAACAAGCCGATCACCTCTAGCCCCAAAATCTAGTGAGTGGTACCTAT	17572
Sbjct	17528	CCACTGGTTATTTAATTAAGCCAATAACGTCAGTGCTAGATCTAGTGAGTGGTATCTTT	17587
Query	17573	GTCTCTCAAACCTTCTT-TCAACTTCAAG-ACGATTACCTCATCAGAGTCATACTACTTG	17630
Sbjct	17588	GTCTGACGAAC-TTCTTATCAACTACACGTAAGATG-CCACACAAAACCATCTCAGTTG	17645
Query	17631	CATGCATGTTATTCAAACAGCACTCCAGCTACAAATTCAGAGGAGCTCATATTGGCTTAG	17690
Sbjct	17646	TAAACAGGTAATACTTACGGCATTGCAACTGC AAATTC AACGAAGCCATACTGGCTAAG	17705
Query	17691	CCACSTGTCCAGTATGCCAATCATAATTTGCATTTAGATTATATTAATCTCGGTTTCCC	17750
Sbjct	17706	TCATTTAACTCAGTATGCTGACTGTGAGTTACATTTAAGTTATATCCGCTTGGTTTTCC	17765
Query	17751	TTCATTGGAGAGGGTTTTATACCATAGATACAATTTAGTCGATTCTCAGAAAGGCCCTTT	17810
Sbjct	17766	ATCATTAGAGAAAAGTACTATACCACAGGTATAACCTCGTCGATTCAAAAAGAGGTCCTACT	17825
Query	17811	GACTTCCATTGTCCAACATCTAGCGCACCTGCAGACC-GAGATTAGGGAGTTGGTTAATG	17869

Sbjct	17826	AGTCTCTATCACTCAGCACTTAGCACATCTT-AGAGCAGAGATTTCGAGAATTAACATAATG	17884
Query	17870	ACTATAATCAACAAAGACAAAGTCGAACCCAAACATATCATTTTCATTAACAATAAAAAG	17929
Sbjct	17885	ATTATAATCAACAGCGACAAAGTCGGACTCAAACATATCACTTTATTTCGTACTGCAAAAAG	17944
Query	17930	GTCGTATTACAAAATTGGTAAATGATTACCTTAAGTCTTTCTAATAATACAAGCCTTAA	17989
Sbjct	17945	GACGAATCACAAAAC TAGTCAATGATTATTTAAAATCTTTCTTATTGTGCAAGCATTAA	18004
Query	17990	AGCACAATTGCACATGGCAAGAGGAACT-AAGAGCTCTCCAGATCTAATTAGTGTCTGC	18048
Sbjct	18005	AACATAATGGGACATGGCAAGCTGAGTTAAGAAAT-TACCAGAGTTGATTAGTGTGTGC	18063
Query	18049	ACTCGATTCTATCATACTCGAAACTGTTTCATGTGAAAACCGGTTCCTAGTACAGACTTTA	18108
Sbjct	18064	AATAGGTTCTACCATATTAGAGATTGCAATTGTGAAGAACGTTCTTAGTTCAAACCTTA	18123
Query	18109	TACTTATCACGCATGCAGGATTCGGAATCAAACATAATAGATAGATTGACCGGCTTCTT	18168
Sbjct	18124	TATTTACATAGAATGCAGGATCTGAAGTTAAGCTTATCGAAAGGCTGACAGGGCTTCTG	18183
Query	18169	AGTCTATGTCCAAATGGTTTTTTTCGGTAAGGACTC--TTGACGTACAACT--CCACAT	18224
Sbjct	18184	AGTTTATTTCGGATGGTCTCTACAGTTTG-ATTGAATTACCGTGCATAGTATCCTGAT	18242
Query	18225	AGTTATACAA--TGGT-ACCAGGACACTATATGTAATTGACCCTAAGAAAGAG---TAA	18278
Sbjct	18243	ACTTGCAAAGGTTGGTTATTAACATACAGATTATAAA---AAACTCATAAATTGCTCTCA	18299
Query	18279	TTCGACACACAGAGTT--CTCAAGTGAAACCCCTCATCTCAGATTATCTGTGGTTGCAA	18335
Sbjct	18300	TACATCATATTGATCTAATCTCAAT--AAACAAC-ATTTAA-ATAACGAAAGGAGTCCC	18355
Query	18336	T-TCTAATATCCGATTGTTACCC-----CGTGAGT-ATAA-CTCCAGATTAAATAAGA	18386
Sbjct	18356	TATATTATATACTATATTAGCCTCTCTCCCTGCGTGATAATCAAAAAATTACAATGCA	18415
Query	18387	AAAT----ACSTTTTG-TCCTGCAA---ATTATCTTAAATTCAGTACA-TACGCTCC	18436
Sbjct	18416	GCATGTGTGACATATTACTGCCGCAATGAATTTAACGCAACATAATAAACTCTGCACTCT	18475
Query	18437	AAATCGTATAAAAATTAAGAAAAAGT-----TAATCTGCTTGCTTTAATTATAACTTT	18490
Sbjct	18476	TTATAAT-TAAGCT-TTAACGAAAGGCTGGGCTCATATTGTTATTGATATAATAATGTT	18533
Query	18491	-----AATATTC-GACA-----AATAGT-TAACGGTCTCATCACTCAAAAATTCATTA	18538
Sbjct	18534	GATCAATATCCTGTGATGGAATAGTGTGTTGGT-TGATAACACAAC---TTC-TTAA	18588
Query	18539	CAAAAGAAGTACTCT----GAGTA---TATTCACA--TATCATA--TGTGATTAA-CATA	18586
Sbjct	18589	AACAA-AATTGATCTTTAAGATTAAGTTTTTTATAAATTATCATTACTTTAATTTGTCGTT	18647

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΑΘΑΝΑΣΙΟΣ ΠΑΥΛΙΔΗΣ

Query 18587 TAAGCAACGCATGATGCGCCTTCCTCTACTTATTGTGTTG--TCACGCAGTCGTTGTAC 18644  
 | | |||| |||| |||| |||| |||| |||| | | | | ||||  
 Sbjct 18648 TAAAAACGG-TGATA-GCCTTAATCTT-----TGTGAAAAATAAGAGATTAGGTGTAA 18699

Query 18645 TACCTCGAAAATTCCAAACAATAAATCGTGTCTATCCCGCATTTAGTGTCTTTAATTAA 18704  
 || | || ||| | |||| || | | | | || | |  
 Sbjct 18700 TAACSTTAACATTTT-----TGTCTAGTAAGC---TACTAT--TTCATACAG 18741

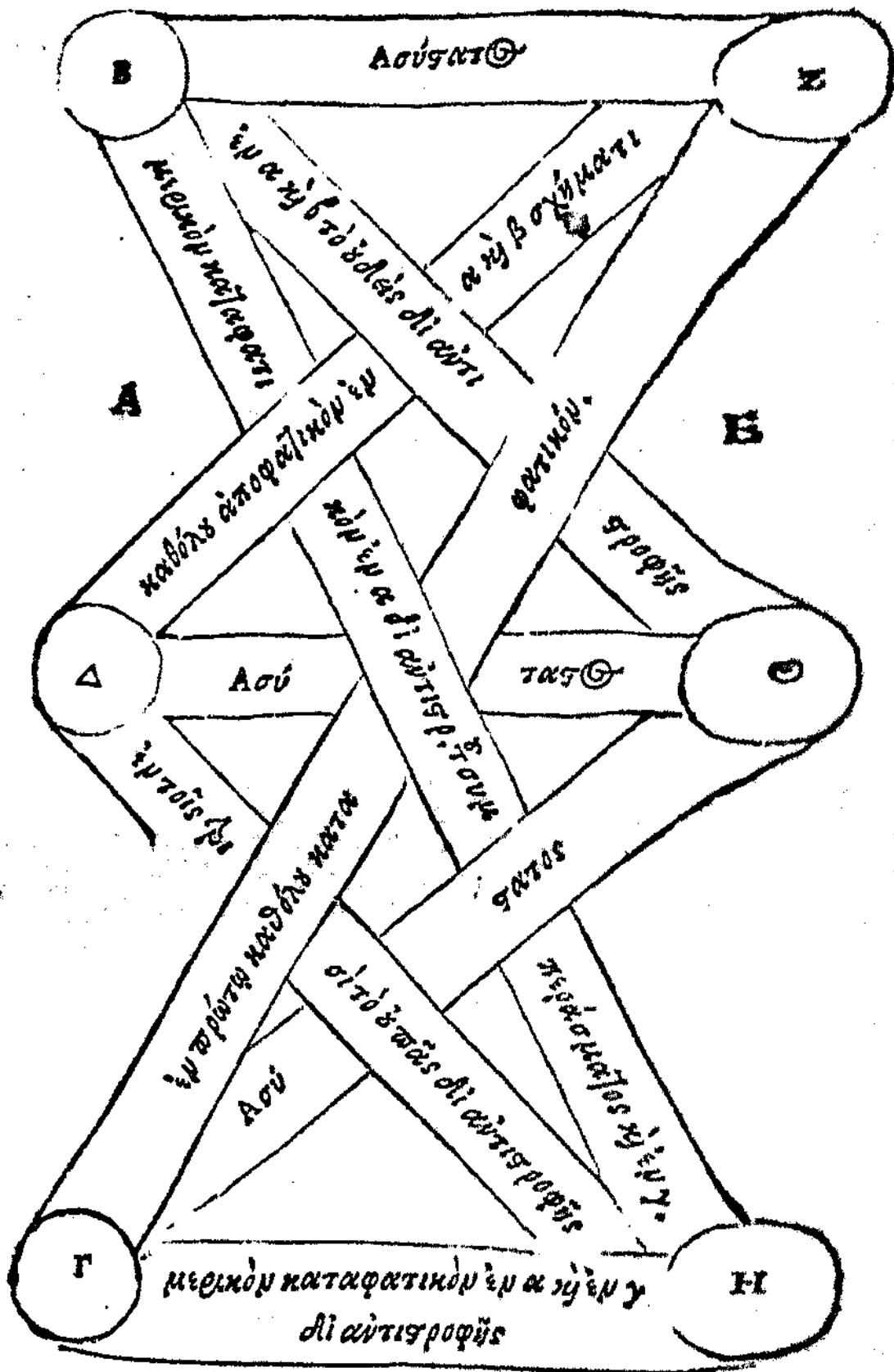
Query 18705 GATC-TCAAATCCAAAAAAGCTGGGTTTATGTTGATGTAAA-TCAATAATACCGAAATTGC 18762  
 || | |||| || || | | | | |||| || || || || || ||  
 Sbjct 18742 AATGATAAAATTAAGAAAAAGGCAGGAC-----TGAAAAATCAGAAATACC----TT-C 18791

Query 18763 TTGATATTAAAAATAAGCTTAAAGGAT-----TTTTCTT-TAAACGGTGATGTAGGTAT 18816  
 || || |||| || | || | || | | | | | | | | ||  
 Sbjct 18792 TT-----TACAATATAGCAGACTAGATAATAATCTTCGTGTTAATGATAAT-TAAGACAT 18845

Query 18817 ATAGGAAAGCTCGATCAGATGTCCCTTACTCAGAAAA-----GAAAAACGGAAGCCCT 18871  
 | | |||| |||| || | | | |||| || | |||| || ||  
 Sbjct 18846 TGACCAC-GCTC-ATCA-GAAGGCTCGC--CAGAATAACGTTGCAAAAAGGATTCC-- 18897

Query 18872 ATTGGCCATTTAATCGTACACAAAAATATCTTTACCAAATTGTTTTCTTTTTTGTGTG 18931  
 ||| | | || |||| |||| | | || | || || |||| |||| ||||  
 Sbjct 18898 --TGGAAAAATGGTCGCACACAAAAATTAATAAATAATCTATTCTCTTTTTTGTGTG 18955

Query 18932 TCCA 18935  
 ||||  
 Sbjct 18956 TCCA 18959



Τὸ τοῦ διαγράμματος περὶ δύπορίας προσησίων ὄσφι καὶ  
 νοῦν ἢ α, καὶ ε, πρώτασις ἔχουσα καταγορέμενον, τὸ α  
 διπλαδὴ ἀγαθὸν ὑποκάμενον τὸ ε, διπλαδὴ ὑδρονή, τὸ  
 ζ β, τὰ ἰπόμενα τῷ α, τὸ ἦ γ, οἷς ἔπιτῆ τὸ α, τὸ ἦ δ, τὰ  
 ἀλλότρια αὐτῆ, καὶ πάλιν τὸ ζ, τὰ ἰπόμενα τῷ ε, τὸ ἦ  
 κ, οἷς ἔπιτῆ τὸ ε, τὸ ἦ θ, τὰ ἀλλότρια αὐτῆ, γίνονται δὲ  
 συζυγίαι θ', ὅτως, β ζ, γ κ, δ θ, γ ζ, β θ, δ ζ, γ θ, δ κ,  
 β κ, ὧν αἱ μὲν ἐξ σιωπῆσιν καὶ ποιῶσι συλλογισμῶν,  
 αἱ δὲ ῥῆσις, ἦτε β ζ, γ θ, δ θ, ἀσύστατοι.

Ἐπόμενα τῷ Α, ὠφέλιμον, αἰρετὸν, δικαῖον, οἰκῆον, ἐ-  
 φικτὸν, λυσιτελίε, συμφέρον. Τὰ ἀλλότρια τῷ Α, ἀτε-  
 λῆς, φουκῆον, βλαβερὸν, κακὸν, ζημιώδεις, ἀλλότριον, ἀσύμ-  
 φερον. Οἷς ἔπιτῆ τὸ Α, σὺδαιμονία, σὺξία, κατὰ φύ-  
 σιν, δ' ἰνικον, τέλειον, κατ' αἰρετικῶ βίῳ.

Τὰ ἰπόμενα τῷ Β, κίνησις, ἐνέργεια κατὰ φύσιν, ἀνεμ-  
 κόδιτος ζωῆ, τὸ κατὰ φύσιν ὄρεκτὸν, τὸ ἀνεμδῆχλητον, τὸ  
 αἰρετὸν. Ἀλλότρια τῷ Β, νόσος, πόνος, ἀπορία, φόβος,  
 παρὰ φύσιν κίνησις. Οἷς ἔπιτῆ τὸ Β, ὑγία, δύπορία,  
 σὺδοξία, σὺτικνία, κατ' αἰρετικῶ βίῳ, ἀπονός ζωῆ.





[Αριστοτέλους Όργανον. : Aristotelis Stagiritae Peripateticorum principis Organum: hoc est libri omnes ad logicam pertinentes, Graece, & Latine / Iul. Pacio a Beriga interprete. Cum triplici indice]. Francofurti, MDXCII [=1592], ΦΣΑ 3979  
<http://anemi.lib.uoc.gr/metadata/2/3/4/metadata-106-0000098.tkl>

