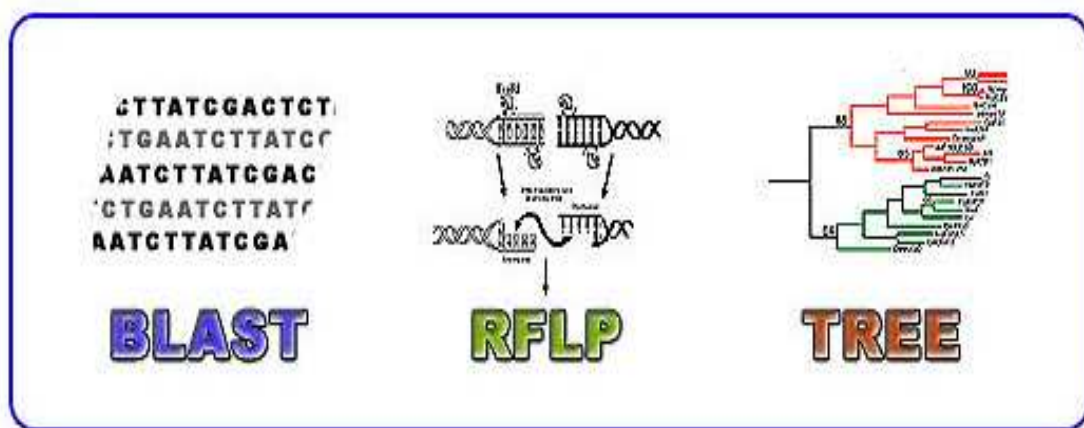


ΑΛΕΞΑΝΔΡΕΙΟ ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ
ΘΕΣΣΑΛΟΝΙΚΗΣ - ΠΑΡΑΡΤΗΜΑ Ν. ΜΟΥΔΑΝΙΩΝ
ΤΜΗΜΑ ΤΕΧΝΟΛΟΓΙΑΣ ΑΛΙΕΙΑΣ ΚΑΙ
ΥΔΑΤΟΚΑΛΛΙΕΡΓΕΙΩΝ

ΦΟΥΝΤΟΣ ΚΩΝΣΤΑΝΤΙΝΟΣ

ΕΦΑΡΜΟΓΕΣ ΣΤΑΤΙΣΤΙΚΩΝ -
ΦΥΛΟΓΕΝΕΤΙΚΩΝ ΠΡΟΓΡΑΜΜΑΤΩΝ ΣΤΗ
ΓΕΝΕΤΙΚΗ ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ
FISHTRACE



ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΝΕΑ ΜΟΥΔΑΝΙΑ (2007)

Στους φίλους μου και στην οικογένειά μου

ΜΕΛΗ ΤΗΣ ΕΞΕΤΑΣΤΙΚΗΣ ΕΠΙΤΡΟΠΗΣ

- 1. Ιμσιρίδου Αναστασία (Επίκουρος Καθηγήτρια)**
- 2. Μίνος Γεώργιος (Επίκουρος Καθηγητής)**
- 3. Κατσαρές Βασίλειος (Εργαστηριακός Συνεργάτης)**

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΡΟΛΟΓΟΣ.....	5
1. ΕΙΣΑΓΩΓΗ	6
1.1. ΓΕΝΙΚΑ	6
1.2. ΔΟΜΗ ΤΟΥ ΓΕΝΕΤΙΚΟΥ ΥΛΙΚΟΥ	7
1.3. BLAST (BASIC LOCAL ALIGNMENT SEARCH TOOL).....	9
1.3.1. Ο ΑΛΓΟΡΙΘΜΟΣ ΤΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ BLAST	9
1.4. RESTRICTION FRAGMENT LENGTH POLYMORPHISM	12
1.4.1. ΠΡΟΓΡΑΜΜΑ ΓΙΑ ΑΝΑΛΥΣΗ RFLP	15
1.5. PHYLIP (PHYLOGENY INFERENCE PACKAGE)	15
1.6. ΣΚΟΠΟΣ ΤΗΣ ΕΡΓΑΣΙΑΣ	16
2. ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ.....	17
2.1. ΑΝΑΖΗΤΗΣΗ ΣΤΟ ΔΙΑΔΙΚΤΥΟ	17
2.2. ΠΡΟΓΡΑΜΜΑ BLAST	17
2.3. ΠΡΟΓΡΑΜΜΑ ΓΙΑ ΑΝΑΛΥΣΗ RFLP	24
2.4. PHYLIP (PHYLOGENY INFERENCE PACKAGE)	29
3. ΑΠΟΤΕΛΕΣΜΑΤΑ.....	33
3.1. ΠΡΟΓΡΑΜΜΑ BLAST	33
3.2. ΠΡΟΓΡΑΜΜΑ ΓΙΑ ΑΝΑΛΥΣΗ RFLP	45
3.3. ΠΡΟΓΡΑΜΜΑ ΓΙΑ ΑΝΑΛΥΣΗ PHYLIP.....	47
4. ΣΥΖΗΤΗΣΗ	50
4.1. ΓΕΝΙΚΑ	50
4.2. ΕΦΑΡΜΟΓΕΣ ΤΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ BLAST	52
4.3. ΕΦΑΡΜΟΓΕΣ ΤΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ ΓΙΑ ΑΝΑΛΥΣΗ RFLP.....	54
4.4 ΕΦΑΡΜΟΓΕΣ ΤΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ PHYLIP	56
5. ΠΕΡΙΛΗΨΗ.....	58
6. SUMMARY	59
7. ΒΙΒΛΙΟΓΡΑΦΙΑ	60

ΠΡΟΛΟΓΟΣ

Η παρούσα εργασία πραγματοποιήθηκε στο ερευνητικό κέντρο της Ευρωπαϊκής Κοινότητας Joint Research Centre (Ispra – Ιταλία), κατά την περίοδο Ιούλιος 2005 – Μάιος 2006. Η εργασία έγινε στα πλαίσια του προγράμματος FISHREG που πραγματοποιήθηκε στο Ινστιτούτο Προστασίας και Ασφάλειας του Πολίτη (IPSC), και συγκεκριμένα στη μονάδα Agrifish.

Για το θέμα της εργασίας, τις απαραίτητες διορθώσεις καθώς και την ευκαιρία που μου δόθηκε να εργαστώ στην Ιταλία, θα ήθελα να ευχαριστήσω την καθηγήτρια μου κ. Αναστασία Ιμσιρίδου.

Για την πολύτιμη βοήθεια τους κατά την περίοδο εκπόνησης της εργασίας θα ήθελα να ευχαριστήσω τον κ. Philippe Carreau που με βοήθησε σε όλα τα βήματα, την κ. Νανά Κούρτη για την κατάλληλη επίβλεψη, καθώς και την κα. Παπαϊωάννου Στέλλα-Συλβάνα για τη συμπαράσταση και τις ιδέες που μου παρείχε.

1. ΕΙΣΑΓΩΓΗ

1. 1. ΓΕΝΙΚΑ

Στον τομέα της Μοριακής Βιολογίας, Βιοχημείας και Γενετικής έχουν γίνει πολλά βήματα καθώς και αρκετά τεχνολογικά επιτεύγματα, τα οποία οδήγησαν στην ανάπτυξη μιας πλειάδας γενετικών δεικτών που χρησιμοποιούνται για τη διάκριση, διαχείριση και διατήρηση των ιχθυοαποθεμάτων. Έχουν δημοσιευτεί πολλές επιστημονικές εργασίες, σχετικές με τις εφαρμογές των γενετικών δεικτών στους ιχθυοπληθυσμούς. Έτσι είναι διαθέσιμος ένας αυξανόμενος αριθμός από επιστημονικά δεδομένα τα οποία παρέχουν υποδείξεις για τη διαχείριση των ιχθυοαποθεμάτων.

Τα περισσότερα από αυτά τα γενετικά στοιχεία είναι καταχωρημένα στα διάφορα εργαστήρια ανά τον κόσμο, όπου γίνονται οι αντίστοιχες αναλύσεις. Η πρόσβαση αυτών των δεδομένων είναι δύσκολη και αυτό οφείλεται σε εξωτερικούς παράγοντες, καθώς επίσης και στον κίνδυνο απώλειας κάποιων στοιχείων με το πέρασμα του χρόνου. Έτσι δημιουργείται αναπόφευκτα η ανάγκη συγκέντρωσης και οργάνωσης όλης αυτής της πληροφορίας σε μία βάση δεδομένων, η οποία θα είναι προσβάσιμη σε ερευνητές, μοριακούς βιολόγους καθώς και σε όλη την επιστημονική κοινότητα.

Η βάση δεδομένων της **FishTrace** (Εικόνα 1) σχεδιάστηκε σε συνεργασία με πολλά ευρωπαϊκά εργαστήρια, έτσι ώστε όλες οι πληροφορίες από τις γενετικές αναλύσεις να μπορέσουν να μεταφερθούν σε μια βάση δεδομένων της οποίας η πρόσβαση να πραγματοποιείται μέσω του διαδικτύου. Έτσι, βασικός στόχος της ιστοσελίδας της FishTrace είναι να εμπλουτιστεί από δεδομένα που αφορούν σύγχρονες μεθοδολογίες γενετικής, καθώς επίσης και από εργαλεία που θα χρησιμοποιούν τα δεδομένα της βάσης για εφαρμογές αυτών στη διάκριση και διαχείριση των ιχθυοπληθυσμών.

Στόχος της παρούσας δουλειάς, είναι η εφαρμογή και χρήση τριών φυλογενετικών προγραμμάτων στη γενετική βάση δεδομένων FishTrace, με σκοπό τη διάκριση ιχθυοαποθεμάτων και ειδών. Τα στατιστικά αυτά πακέτα είναι τα ακόλουθα: BLAST, RFLPs, PHYLIP.



THE PROJECT

- AIMS
- THE CONSORTIUM
- PERSONNEL & EXPERTISE
- DATABASE STRUCTURE
- DATABASE LOADER
- PUBLICATIONS
- DISSEMINATION & PHOTOS

SAMPLING & TAXONOMY

- AIMS
- SAMPLING AREAS
- TARGETED SPECIES
- STANDARD PROTOCOLS

REFERENCE COLLECTIONS

- AIMS
- BIOLOGICAL COLLECTIONS
- ACCESS TO COLLECTIONS
- STANDARD PROTOCOLS

GENETIC CATALOGUE

- AIMS
- MOLECULAR ID TOOLS
- STANDARD PROTOCOLS

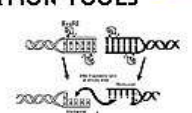
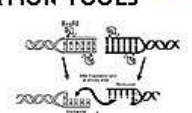
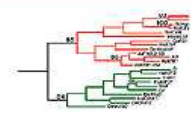
Contact us if you experience any problem

SEARCH

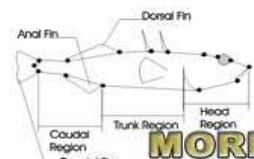
By scientific name

By common name (ex: *Bigueye tuna*)

MOLECULAR IDENTIFICATION TOOLS

BLAST  **RFLP**  **TREE** 

MORPHOLOGICAL TOOL



MORPHOLOGICAL TOOL

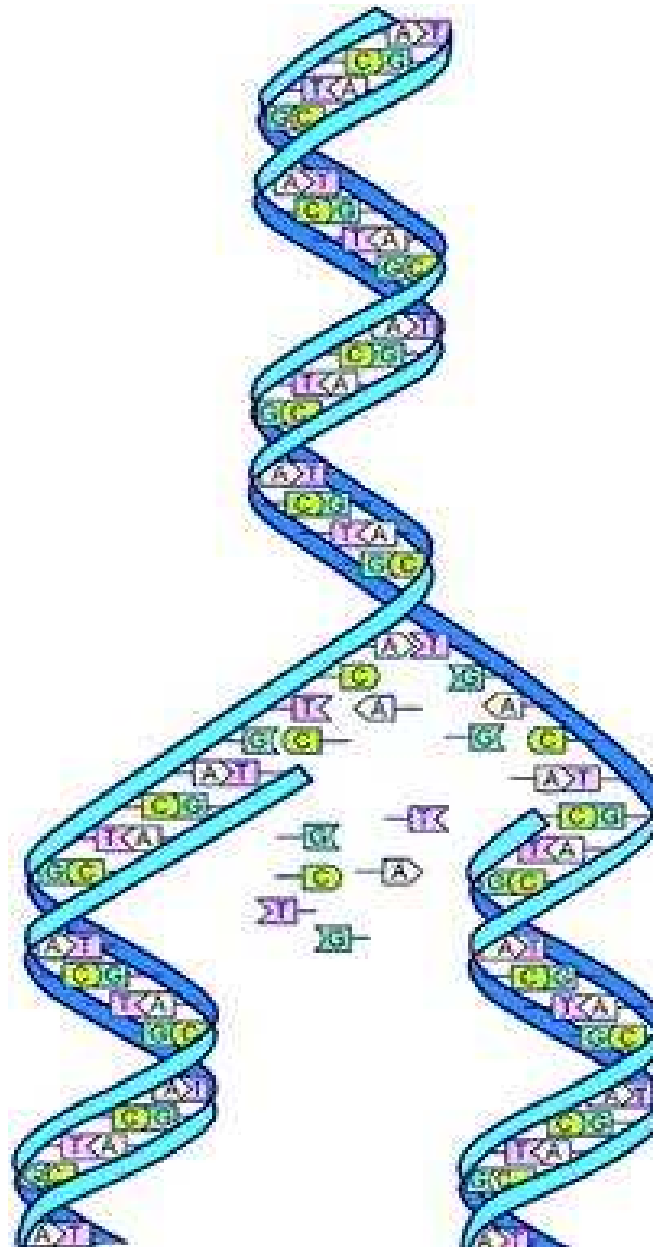
Εικόνα 1. Η κεντρική σελίδα της βάσης δεδομένων Fishtrace

1. 2. Η ΔΟΜΗ ΤΟΥ ΓΕΝΕΤΙΚΟΥ ΥΛΙΚΟΥ (DNA)

Το δεσοξυριβονουκλεϊκό οξύ (DNA) είναι ένα νουκλεϊνικό οξύ – συνήθως στη μορφή μιας διπλής έλικας – που περιέχει τις γενετικές πληροφορίες, οι οποίες καθορίζουν τη βιολογική εξέλιξη κάθε κυτταρικής μορφής ζωής. Το DNA είναι ένα μακρύ πολυμερές από νουκλεοτίδια και κωδικοποιεί την αλληλουχία των αμινοξέων σε πρωτεΐνες, χρησιμοποιώντας τον γενετικό κώδικα. Το DNA αναφέρεται συχνά ως η κληρονομική μονάδα, καθώς είναι υπεύθυνο για τη μεταφορά από γενιά σε γενιά μιας σειράς γνωρισμάτων (<http://en.wikipedia.org/wiki/DNA>).

Η δομή του DNA έχει τη μορφή μιας διπλής έλικας η οποία είναι φτιαγμένη από δύο αντιπαράλληλες αλυσίδες νουκλεοτιδίων, με τις βάσεις αδενίνη, θυμίνη, γουανίνη και κυτοσίνη ενωμένες με δεσμούς υδρογόνου (Εικόνα 2). Αυτές οι αζωτούχες βάσεις συμβολίζονται με τα αρχικά γράμματα των λατινικών ονομάτων. Έτσι, το γράμμα «Α» χρησιμοποιείται για την αδενίνη (Adenine), το γράμμα «Τ»

χρησιμοποιείται για τη θυμίνη (Thymine), το γράμμα «G» χρησιμοποιείται για τη γουανίνη (Guanine), και τέλος το γράμμα «C» χρησιμοποιείται για την κυτοσίνη (Cytosine). Οι βάσεις αυτές συνθέτουν μια γενετική ακολουθία. Για παράδειγμα μια γενετική ακολουθία μπορεί να είναι: «AACGTCGTGATC». Τα αποτελέσματα τα οποία προκύπτουν αφού συγκριθούν δυο γενετικές ακολουθίες είναι πολύ αξιόπιστα για τη ταυτοποίηση των ειδών (<http://en.wikipedia.org/wiki/DNA>).



Εικόνα 2. Η διπλή έλικα του DNA με τις βάσεις A, C, G, T

1. 3. BLAST (Basic Local Alignment Search Tool)

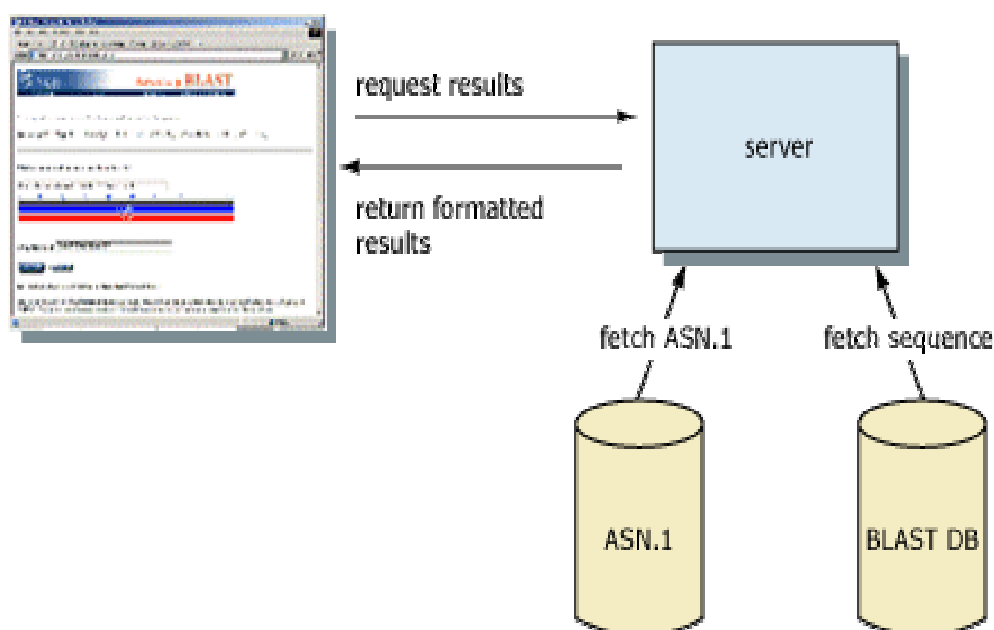
Στην Επιστήμη της Γενετικής, έχουν δημιουργηθεί διάφορα εργαλεία για την σύγκριση γενετικών αλληλουχιών. Το πιο γνωστό μεταξύ αυτών είναι το BLAST (Basic Local Alignment Search Tool). Αυτό το λογισμικό πρόγραμμα βρίσκει περιοχές ομοιότητας μεταξύ των γενετικών αλληλουχιών. Συγκρίνει νουκλεοτιδικές ή πρωτεϊνικές αλληλουχίες με αλληλουχίες που βρίσκονται ήδη σε μία βάση δεδομένων, και υπολογίζει τη στατιστική σημαντικότητα των αποτελεσμάτων. Το πρόγραμμα BLAST μπορεί να χρησιμοποιηθεί για την εκτίμηση της συγγένειας και της εξελικτικής σχέσης μεταξύ διαφορετικών αλληλουχιών. Το πρόγραμμα χρησιμοποιεί μια βάση δεδομένων η οποία περιέχει γενετικές ακολουθίες και συγκρίνει βάση προς βάση μια ακολουθία νουκλεοτιδίων που θα ορίσει ο χρήστης, με μία άλλη που προέρχεται από τη βάση δεδομένων (alignment) (<http://www.ncbi.nlm.nih.gov/BLAST/tutorial/Altschul-1.html>).

1. 3. 1. Ο ΑΛΓΟΡΙΘΜΟΣ ΤΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ BLAST

Ο αλγόριθμος του προγράμματος BLAST είναι ένα λογισμικό πρόγραμμα το οποίο είναι ικανό να λειτουργεί με μερικές συντομεύσεις, για να πραγματοποιεί την αναζήτηση γρηγορότερα. Ο αλγόριθμος εκτελεί «τοπικές» συγκρίσεις ακολουθιών. Αν το BLAST ξεκινούσε προσπαθώντας να ευθυγραμμίσει δύο νουκλεοτιδικές αλληλουχίες σε ολόκληρο το μήκος τους (καθολική ευθυγράμμιση), πολύ λιγότερες ομοιότητες θα είχαν εντοπιστεί μεταξύ των συγκρινόμενων ακολουθιών (<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=handbook.chapter.610>).

Όταν μια εξέταση - έλεγχος υποβάλλεται μέσω μιας ιστοσελίδας του προγράμματος BLAST, η νουκλεοτιδική ακολουθία μαζί με οποιαδήποτε άλλη πληροφορία (όπως η βάση δεδομένων μέσα στην οποία θα γίνει η αναζήτηση, το μέγεθος της ακολουθίας, η αναμενόμενη τιμή) δίνονται στον αλγόριθμο του δικτυακού υπολογιστή. Η αρχή λειτουργίας του BLAST βασίζεται στη δημιουργία ενός πίνακα ελέγχου όλων των «λέξεων» (μικρές ακολουθίες) καθώς και των «γειτονικών λέξεων» (παρόμοιες ακολουθίες στην αλληλουχία αναζήτησης). Στη συνέχεια η νουκλεοτιδική βάση δεδομένων σαρώνεται, έτσι ώστε να μπορέσουν να βρεθούν όλα αυτά τα σημαντικά σημεία (λέξεις, γειτονικές λέξεις) (<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=handbook.chapter.610>).

Αφού ο αλγόριθμος έχει ελέγξει όλες τις πιθανές «λέξεις» από την εξεταζόμενη αλληλουχία και τις προεκτείνει μέγιστα, συγκεντρώνει τις καλύτερες ευθυγραμμίσεις νουκλεοτιδικών ακολουθιών (alignments) για κάθε ζεύγος ακολουθιών που συγκρίνει, και στη συνέχεια γράφει αυτή την πληροφορία σε μια «SeqAlign» δομή δεδομένων. Η δομή δεδομένων «SeqAlign» δεν περιέχει πληροφορία για την ακολουθία, αλλά αναφέρεται στις ακολουθίες οι οποίες βρίσκονται στη βάση δεδομένων αναζήτησης του προγράμματος BLAST (Εικόνα 3) (<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=handbook.chapter.610>).



Εικόνα 3. Η διάρθρωση των αποτελεσμάτων του προγράμματος BLAST.

Τη στιγμή που το πρόγραμμα βρει μια παρόμοια ακολουθία με αυτή του χρήστη κατά την αναζήτηση μέσα στη βάση δεδομένων, είναι καλό ο ερευνητής να έχει μια ιδέα εάν το αποτέλεσμα της ευθυγράμμισης (alignment) αντικατοπτρίζει μια πιθανή βιολογική σχέση, ή εάν η ομοιότητα που βρέθηκε οφείλεται στη τύχη. Το πρόγραμμα BLAST χρησιμοποιεί τη στατιστική θεωρία για να παράγει ένα αποτέλεσμα (bit score) και την αναμενόμενη τιμή (Expectation value), για κάθε ζεύγος ακολουθιών που συγκρίνεται. Η τιμή E είναι ο αριθμός των διαφορετικών συγκρίσεων ακολουθιών (alignments), για τις οποίες η ομοιότητα οφείλεται στη τύχη. Συνεπώς, όσο χαμηλότερη είναι η τιμή E, τόσο πιο σημαντικό είναι το αποτέλεσμα από την εξελικτική και βιολογική σκοπιά. Για παράδειγμα, αφού γίνει μια έρευνα με

το λογισμικό εργαλείο BLAST, τα αποτελέσματα δίνονται σε στήλες. Η τιμή E (Expectation value), είναι η δεύτερη στήλη και το πρώτο αποτέλεσμα που δίνεται είναι το πιο κοντινό αποτέλεσμα στην αναζήτησή μας (Εικόνα 4) (<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=handbook.chapter.610>).

```

Query= EST_Clone_DW1
          (587 letters)

Database: C:\fishtrace\blast\data\FT_specimen.nt
            648 sequences; 507,416 total letters

Sequences producing significant alignments:

Score      E
(bits)    Value

Aspitrigla-cuculus (CytB)          32    0.057
Aspitrigla-cuculus (CytB)          32    0.057
Aspitrigla-cuculus (CytB)          32    0.057
Aspitrigla-cuculus (CytB)          32    0.057
Chelidonichthys-lucernus (CytB)    26    3.5
Chelidonichthys-lucernus (CytB)    26    3.5
Chelidonichthys-lucernus (CytB)    26    3.5
Chelidonichthys-lucernus (CytB)    26    3.5
Chelidonichthys-lucernus (CytB)    26    3.5
Mullus-surmuletus (CytB)          26    3.5
Mullus-surmuletus (CytB)          26    3.5
Mullus-surmuletus (CytB)          26    3.5
Mullus-surmuletus (CytB)          26    3.5
Mullus-surmuletus (CytB)          26    3.5
Mullus-surmuletus (CytB)          26    3.5
Mullus-surmuletus (CytB)          26    3.5
Mullus-surmuletus (CytB)          26    3.5
Salmo-trutta (CytB)              26    3.5
Salmo-trutta (CytB)              26    3.5
Chelidonichthys-gurnardus (CytB)   26    3.5
Chelidonichthys-gurnardus (CytB)   26    3.5
Chelidonichthys-gurnardus (CytB)   26    3.5
Chelidonichthys-gurnardus (CytB)   26    3.5

>Aspitrigla-cuculus (CytB)>specimen:AspCuc-CB-02==Aspitrigla-cuculus (CytB)
      Length = 1141

Score = 32.2 bits (16), Expect = 0.057
Identities = 16/16 (100%)
Strand = Plus / Plus

```

Εικόνα 4. Η τιμή E στη σελίδα των αποτελεσμάτων του προγράμματος BLAST.

Η τιμή E λοιπόν είναι μία ένδειξη στατιστικής σημαντικότητας των αποτελεσμάτων ευθυγράμμισης των ακολουθιών (alignments), και αντικατοπτρίζει ως ένα βαθμό το μέγεθος της βάσης δεδομένων αναζήτησης καθώς και το σύστημα αποτελεσμάτων (scoring system) που έχει χρησιμοποιηθεί. Όσο πιο χαμηλή είναι η τιμή E, τόσο πιο σημαντική είναι η ομοιότητα μεταξύ των ακολουθιών. Ένα

αποτέλεσμα ευθυγράμμισης (sequence alignment) που έχει μια τιμή $E < 0.05$, δείχνει ότι αυτή η ομοιότητα μεταξύ των ακολουθιών έχει μία πιθανότητα 5% να οφείλεται στην τύχη. Εάν και ένας στατιστικολόγος μπορεί να θεωρήσει αυτό το αποτέλεσμα σημαντικό, εντούτοις μπορεί να μην αντικατοπτρίζει ένα σημαντικό αποτέλεσμα από βιολογική, εξελικτική και φυλογενετική σκοπιά.

Το αποτέλεσμα (bit score) δίνει μία ένδειξη για το πόσο καλή είναι η σύγκριση - ευθυγράμμιση των ακολουθιών κατά ζεύγη (alignment). Όσο πιο υψηλή είναι η τιμή, τόσο καλύτερο είναι το αποτέλεσμα της ευθυγράμμισης (Εικόνα 4). Σε γενικές γραμμές, αυτή η τιμή υπολογίζεται από έναν τύπο που λαμβάνει υπ' όψιν την ευθυγράμμιση παρόμοιων ή ταυτόσημων νουκλεοτιδικών κομματιών (μεταξύ συγκρινόμενων ακολουθιών), καθώς επίσης και τα κενά που εισάγονται για να ευθυγραμμιστούν οι ακολουθίες. Ένα στοιχείο κλειδί σε αυτό τον υπολογισμό είναι η «μήτρα αντικατάστασης» (substitution matrix), η οποία αναθέτει μία τιμή για την ευθυγράμμιση κάθε πιθανού ζεύγους συγκρινόμενων νουκλεοτιδικών ακολουθιών (<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=handbook.chapter.610>).

1. 4. RESTRICTION FRAGMENT LENGTH POLYMORPHISM

Στη Μοριακή Βιολογία, ο όρος Πολυμορφισμός Μήκους Περιοριστικών Θραυσμάτων (Restriction Fragment Length Polymorphism: RFLP) έχει δύο διαφορετικές σημασίες: α) χαρακτηρίζει μόρια DNA τα οποία διαφέρουν στη νουκλεοτιδική τους αλληλουχία και έτσι μπορούν να διακριθούν β) περιγράφει μία εργαστηριακή τεχνική με την οποία οι οργανισμοί μπορούν να διακριθούν, από την ανάλυση των περιοριστικών προτύπων του γενετικού τους υλικού. Η τεχνική αυτή είναι μία από τις κύριες γενετικές μεθοδολογίες ανίχνευσης πολυμορφισμών σε ιχθυοπληθυσμούς (Cronin et al. 1993, Apostolidis et al. 1996, Hansen & Loeschcke 1996, Imsiridou et al. 1998). Χρησιμοποιείται επίσης στη γενετική ανάλυση αποτυπωμάτων (fingerprinting).

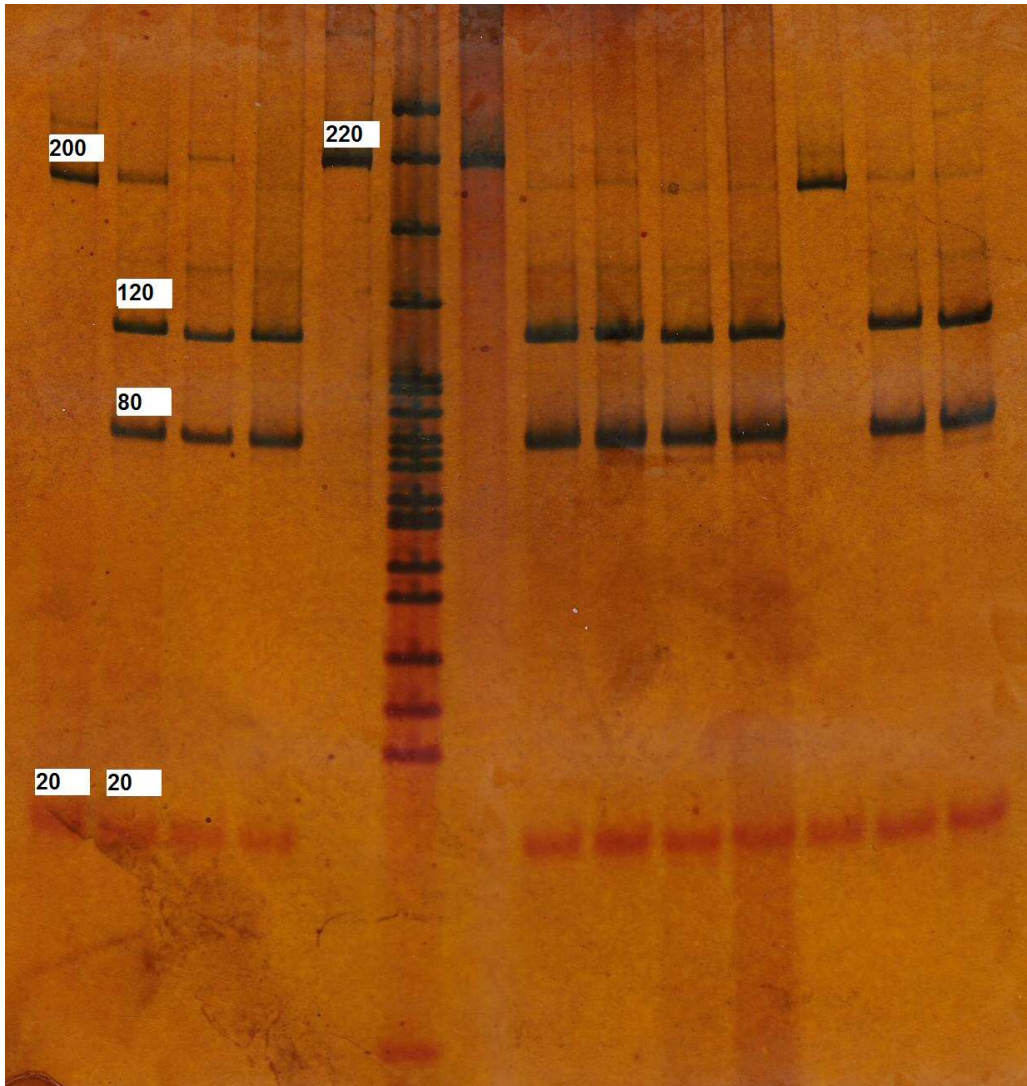
Αν δύο οργανισμοί διαφέρουν στην απόσταση μεταξύ των θέσεων κοπής του DNA τους από μία συγκεκριμένη περιοριστική ενδονουκλεάση, το μήκος και ο αριθμός των κομματιών DNA που θα προκύψουν μετά από πέψη με το ένζυμο θα διαφέρει μεταξύ των δύο. Τα διαφορετικά πρότυπα που θα προκύψουν μπορούν να χρησιμοποιηθούν για τη διάκριση και διαφοροποίηση των οργανισμών (Cespedes et al. 2000, Comesana & Abella, 2003)

Οι ενδονουκλεάσες περιορισμού είναι ένζυμα που κόβουν τα μόρια του DNA σε συγκεκριμένες νουκλεοτιδικές θέσεις, αναλόγως το ένζυμο που χρησιμοποιείται. Οι περιοχές αναγνώρισης του ενζύμου έχουν μέγεθος 4-6 ζεύγη βάσεων. Γενικά, όσο πιο μικρή είναι η αλληλουχία αναγνώρισης του ενζύμου, τόσο περισσότερα είναι τα κομμάτια DNA που προκύπτουν μετά από την πέψη. Αν δύο μόρια DNA διαφέρουν στη νουκλεοτιδική τους αλληλουχία, τότε κομμάτια διαφορετικού μεγέθους θα προκύψουν μετά από την πέψη. Τα τμήματα DNA μπορούν να διαχωριστούν με ηλεκτροφόρηση πηκτής αγαρόζης. Τα ένζυμα περιορισμού απομονώνονται από βακτήρια και αποτελούν μέρος της κυτταρικής άμυνας των βακτηρίων εναντίον των ιών, εφόσον καταστρέφουν το γενετικό υλικό του ιού μόλις αυτός προσβάλλει το βακτήριο. Οι ενδονουκλεάσες περιορισμού παίρνουν το όνομα τους από το πρώτο γράμμα του γένους του βακτηρίου που απομονώθηκαν, τα δύο πρώτα γράμματα του είδους του βακτηρίου και την σειρά ανακάλυψης του ενζύμου π.χ. EcoRI.

Συνήθως, γίνεται εξαγωγή DNA από ένα άτομο και ακολουθεί καθαρισμός του DNA. Το καθαρό DNA μπορεί να ενισχυθεί με την μέθοδο της Αλυσιδωτής Αντίδρασης Πολυμεράσης (Polymerase Chain Reaction – PCR). Στη συνέχεια το DNA κόβεται σε περιοριστικά κομμάτια από τις ενδονουκλεάσες περιορισμού, οι οποίες θα το κόψουν μόνο όπου υπάρχουν οι συγκεκριμένες αλληλουχίες αναγνώρισης των ενζύμων. Στη συνέχεια τα περιοριστικά θραύσματα διαχωρίζονται ανάλογα με το μήκος τους, σε μία ηλεκτροφόρηση πηκτής αγαρόζης ή πολυακρυλαμίδης (Εικόνα 5) (Ιμσιρίδου, 2003).

Η απόσταση μεταξύ των θέσεων κοπής ποικίλει μεταξύ των ατόμων, και έτσι το μήκος και ο αριθμός των κομματιών DNA θα διαφέρει (πολυμορφισμός). Η τεχνική αυτή μπορεί να χρησιμοποιηθεί για το χαρακτηρισμό διαφορετικών ατόμων μέσα σε ένα πληθυσμό ψαριών, για την αναγνώριση διαφορετικών πληθυσμών ψαριών μέσα σε ένα είδος ή για τη διάκριση διαφορετικών ειδών. Τα κύρια στάδια μιας ανάλυσης RFLP είναι τα παρακάτω:

- Επιλογή της ακολουθίας του DNA που θα αναλυθεί
- Εξαγωγή DNA
- Πέψη του DNA με ενδονουκλεάσες περιορισμού
- Χρώση της πηκτής
- Ανάγνωση των αποτελεσμάτων



Εικόνα 5. Περιοριστικό πρότυπο που προκύπτει μετά από μία ανάλυση RFLP

Έστω δύο ζωικοί οργανισμοί οι οποίοι μοιράζονται δύο κοινές θέσεις κοπής (έχουν και κοινό κομμάτι DNA στο περιοριστικό τους πρότυπο). Αν στο ένα από τα δύο άτομα δημιουργηθεί μια νέα θέση κοπής ανάμεσα στις δύο προϋπάρχουσες, τότε χάνεται το κοινό κομμάτι DNA και προκύπτουν δύο νέα μικρότερα (Ιμισιρίδου, 2003).

1. 4. 1. ΠΡΟΓΡΑΜΜΑ ΓΙΑ ΑΝΑΛΥΣΗ RFLP

Για τη βάση δεδομένων Fishtrace (Εικόνα 1), βρέθηκε και προσαρμόστηκε ένα εργαλείο για τη μέθοδο του Πολυμορφισμού Μήκους Περιοριστικών Θραυσμάτων (RFLP). Το πρόγραμμα αυτό χρησιμοποιεί μία λίστα περιοριστικών ενζύμων και τις αλληλουχίες του DNA από τα είδη που υπάρχουν στη Fishtrace, για να εφαρμόσει «εικονικά» την πέψη και τον διαχωρισμό θραυσμάτων DNA σε πηκτική αгарόζης. Τα αποτελέσματα της «πέψης» δίνονται στο τέλος σε γράφημα.

1. 5. PHYLIP (PHYLOGENY INFERENCE PACKAGE)

Το λογισμικό πρόγραμμα PHYLIP (Felsenstein, 1989), είναι ένα φυλογενετικό πακέτο ανάλυσης. Οι μέθοδοι οι οποίες είναι διαθέσιμες στο πακέτο περιλαμβάνουν μεταξύ άλλων μεθόδους φειδωλότητας (parsimony methods), μεθόδους απόστασης (distance matrix methods), μεθόδους πιθανότητας (likelihood methods), μεθόδους εύρεσης ορίων εμπιστοσύνης (bootstrap methods) καθώς και μεθόδους κατασκευής φυλογενετικών δέντρων (consensus trees). Τα δεδομένα τα οποία μπορεί να χρησιμοποιήσει το πρόγραμμα περιλαμβάνουν νουκλεοτιδικές ακολουθίες, γονιδιακές συχνότητες, περιοριστικές θέσεις κοπής, μήτρες γενετικών αποστάσεων και 0/1 διακριτούς χαρακτήρες (Tuimala, 2004).

Το πρόγραμμα ελέγχεται μέσα από ένα μενού, το οποίο ζητά από τους χρήστες να ρυθμίσουν τις επιλογές που θέλουν και στη συνέχεια τους επιτρέπει να αρχίσουν τους υπολογισμούς. Τα δεδομένα διαβάζονται μέσα στο πρόγραμμα από ένα αρχείο τύπου «text», το οποίο ο χρήστης μπορεί να προετοιμάσει χρησιμοποιώντας τον «επεξεργαστή κειμένου» (word processor) ή τον «διορθωτή κειμένου» (text editor). Σημαντική λεπτομέρεια είναι αυτό το αρχείο «text» να μην είναι σε ειδική μορφή (format) αυτού του επεξεργαστή κειμένου, αλλά αντίθετα σε "flat ASCII" ή "Text Only" format. Μερικά λογισμικά προγράμματα ευθυγράμμισης ακολουθιών, όπως το ClustalX και το T-Coffee, μπορούν να γράψουν τα αρχεία δεδομένων σε format τύπου PHYLIP (Tuimala, 2004).

Στο πρόγραμμα PHYLIP υπάρχει μία λίστα από εφαρμογές, οι οποίες μπορούν να χρησιμοποιηθούν για ανάλυση νουκλεοτιδικών ακολουθιών. Οι εφαρμογές που χρειάστηκαν για τη βάση δεδομένων Fishtrace, είναι αυτές που

σχεδιάζουν φυλογενετικά δέντρα (Drawgram και drawtree). Αυτές οι εφαρμογές σχεδιάζουν ένα δέντρο από το εισαγόμενο αρχείο σε format τύπου PHYLIP, που στην περίπτωση μας είναι νουκλεοτιδικές ακολουθίες ψαριών. Οι εφαρμογές Drawgram και drawtree παράγουν ένα αρχείο σχεδίου (outtree). Συνεπώς το αρχείο «outtree» δεν μπορεί να χρησιμοποιηθεί ως εισαγόμενο δεδομένο (input) σε αυτό το πρόγραμμα (Tuimala, 2004).

1.6. ΣΚΟΠΟΣ ΤΗΣ ΕΡΓΑΣΙΑΣ

Ο σκοπός της παρούσας δουλειάς συνίσταται στα παρακάτω:

- Οργάνωση και σχεδιασμός ιστοσελίδων, για την εφαρμογή εργαλείων στη γενετική βάση δεδομένων FishTrace
- Έρευνα για την εύρεση **τριών** κατάλληλων στατιστικών προγραμμάτων, για εφαρμογές γενετικών δεδομένων
- Προγραμματισμός των προγραμμάτων αυτών στο λειτουργικό σύστημα των Windows
- Προσαρμογή των παραμέτρων **κάθε προγράμματος**, στις ανάγκες και στα επίπεδα της γενετικής βάσης δεδομένων FishTrace

2. ΥΛΙΚΑ ΚΑΙ ΜΕΘΟΔΟΙ

2.1. ΑΝΑΖΗΤΗΣΗ ΣΤΟ ΔΙΑΔΙΚΤΥΟ

Αρχικά, έγινε μια ανασκόπηση σε όλες τις σελίδες που υπάρχουν στο διαδίκτυο για το πρόγραμμα BLAST και για άλλες μεθοδολογίες και λογισμικά προγράμματα, που πραγματοποιούν σύγκριση και ευθυγράμμιση ακολουθιών. Ελέγχθηκαν πάνω από 5.000 ιστοσελίδες. Μετά από αυτή τη βιβλιογραφική έρευνα, ακριβείς πληροφορίες έδωσαν την λύση. Η μέθοδος του προγράμματος BLAST ήταν η πιο κατάλληλη, καθώς φαίνεται ότι είναι ο πιο δημοφιλής αλγόριθμος που χρησιμοποιείται από τους επιστήμονες.

2. 2. ΠΡΟΓΡΑΜΜΑ BLAST

Το πρόγραμμα αυτό βρίσκει περιοχές τοπικής ομοιότητας μεταξύ γενετικών ακολουθιών. Συγκρίνει νουκλεοτιδικές ή πρωτεϊνικές ακολουθίες με ακολουθίες που βρίσκονται σε μια βάση δεδομένων, και υπολογίζει τη στατιστική σημαντικότητα των συγκρίσεων. Για παράδειγμα, μία τοπική ευθυγράμμιση δύο ακολουθιών (alignment) χωρίς κενά (gaps) αποτελείται απλά από ένα ζεύγος τμημάτων με το ίδιο μήκος, τα οποία ανήκουν στις δύο συγκρινόμενες ακολουθίες. Μια τροποποίηση στον αλγόριθμο Smith - Waterman ή Sellers, θα βρει όλα τα ζεύγη τμημάτων των οποίων τα αποτελέσματα δεν μπορούν να βελτιωθούν με επέκταση. Αυτά καλούνται ζεύγη τμημάτων με καλά αποτελέσματα (High Scoring Segment Pairs – HSPs).

Για να αναλυθεί πόσο ένα καλό αποτέλεσμα μπορεί να οφείλεται στην τύχη, χρειάζεται ένα μοντέλο τυχαίων ακολουθιών. Ας θεωρήσουμε δύο τυχαίες νουκλεοτιδικές ακολουθίες m και n με αρκετά μεγάλο μήκος. Η στατιστική των ζευγών τμημάτων με καλά αποτελέσματα (HSPs), χαρακτηρίζεται από δύο παραμέτρους K και λ . Συνεπώς, η αναμενόμενη τιμή (Expectation value) των HSPs με αποτέλεσμα τουλάχιστον S , θα δίνεται από τον τύπο:

$$E = Kmn e^{-\lambda S}$$

Αυτή θα είναι η τιμή E για ένα αποτέλεσμα S . Διπλασιάζοντας το μήκος κάθε μιας ακολουθίας, θα έπρεπε να διπλασιαστεί και ο αριθμός των HSPs τα οποία

επιτυγχάνουν ένα δομένο αποτέλεσμα. Επίσης, για μπορέσει ένα ζεύγος τμημάτων (HSP) να πετύχει ένα αποτέλεσμα $2x$, πρέπει να πετύχει το αποτέλεσμα x δύο φορές στην ίδια σειρά. Έτσι, η τιμή E μειώνεται εκθετικά με το αποτέλεσμα S . Η παράμετρος K αντικατοπτρίζει το μέγεθος της βάσης δεδομένων αναζήτησης, ενώ η παράμετρος λ δηλώνει το σύστημα αποτελεσμάτων (scoring system) που έχει χρησιμοποιηθεί.

Το πρόγραμμα BLAST χρησιμοποιεί μία βάση δεδομένων η οποία έχει συντεθεί από γενετικές ακολουθίες. Έτσι το πρόγραμμα κάνει σύγκριση βάση προς βάση μία ακολουθία την οποία εισχωρεί ο χρήστης, με μία δεύτερη που προέρχεται από τη βάση δεδομένων.

Το πρόγραμμα BLAST (**B**asic **L**ocal **A**lignment **S**earch **T**ool) δημιουργήθηκε από το Εθνικό Κέντρο Βιοτεχνολογίας (National Center for Biotechnology Information - NCBI). Η βάση δεδομένων Fishtrace περιέχει ακολουθίες DNA από εμπορικά είδη ψαριών. Προκειμένου λοιπόν το πρόγραμμα BLAST να χρησιμοποιηθεί για τη βάση δεδομένων της Fishtrace, έπρεπε να προσαρμοστεί κατάλληλα.

Στην ιστοσελίδα του Εθνικού Κέντρου Βιοτεχνολογίας (<http://www.ncbi.nlm.nih.gov/BLAST/download.shtml>), βρέθηκαν τα εκτελέσιμα αρχεία για να ξεκινήσει η εγκατάσταση και προσαρμογή. Έτσι, σε πρώτη φάση χρησιμοποιήθηκε το εκτελέσιμο αρχείο «wwwblast», καθώς είναι αυτό που απευθύνεται στην δικτυακή δομή του BLAST. Παρ' όλο που το αρχείο ήταν διαθέσιμο μόνο για το λειτουργικό σύστημα Linux, αποφασίστηκε να προσαρμοστεί σε περιβάλλον Windows χρησιμοποιώντας στοιχεία Java. Η κύρια αποστολή ήταν η μεταφορά και η δοκιμή του προγράμματος BLAST, στο περιβάλλον του command line. Αυτό είχε ως αποτέλεσμα τη δημιουργία μιας απλής λειτουργικής φόρμας, προσαρμοσμένης στις ανάγκες της βάσης δεδομένων.

Στη συνέχεια, εξετάστηκαν οι τομείς και οι επιλογές του προγράμματος. Η φόρμα του εργαλείου του BLAST, έχει τρεις τομείς (Εικόνα 6).



NCBI BLAST BLAST Entrez ?

Choose program to use and database to search:

Program Database

Enter sequence below in FASTA format

The query sequence is filtered for low complexity regions by default.

Filter Low complexity Mask for lookup table only

Expect Matrix Perform ungapped alignment

Query Genetic Codes (blastx only)

Database Genetic Codes (tblast[nx] only)

Frame shift penalty for blastx

Other advanced options:

Graphical Overview Alignment view

Descriptions Alignments

Εικόνα 6. Το εργαλείο BLAST.

Το πρώτο κομμάτι της φόρμας (Εικόνα 7), παρέχει στους χρήστες τη δυνατότητα επιλογής λειτουργίας του.

NCBI BLAST BLAST Entrez ?

Choose program to use and database to search:

Program Database

Enter sequence below in FASTA format

Εικόνα 7. Το εργαλείο BLAST, τομέας πρώτος.

Ο χρήστης λοιπόν έχει να επιλέξει ανάμεσα στα παρακάτω προγράμματα BLAST:

1. blastn: συγκρίνει μία νουκλεοτιδική ακολουθία ελέγχου με νουκλεοτιδικές ακολουθίες που βρίσκονται σε μία βάση δεδομένων. Συνεπώς το πρόγραμμα μπορεί να χρησιμοποιηθεί για ταυτοποίηση νουκλεοτιδικών ακολουθιών, ή για να βρει ακολουθίες μέσα στη βάση δεδομένων παρόμοιες με την ακολουθία ελέγχου του χρήστη.
2. blastp: συγκρίνει μία ακολουθία αμινοξέων ελέγχου, με πρωτεϊνικές ακολουθίες που βρίσκονται σε μία βάση δεδομένων. Συνεπώς το πρόγραμμα μπορεί να χρησιμοποιηθεί για ταυτοποίηση πρωτεϊνικών ακολουθιών, ή για να βρει πρωτεϊνικές ακολουθίες μέσα στη βάση δεδομένων παρόμοιες με την ακολουθία ελέγχου του χρήστη.
3. blastx: συγκρίνει μία νουκλεοτιδική ακολουθία ελέγχου η οποία μεταφράζεται σε όλα τα αναγνωστικά πλαίσια, με πρωτεϊνικές ακολουθίες που βρίσκονται σε μία βάση δεδομένων. Συνεπώς το πρόγραμμα μπορεί να χρησιμοποιηθεί για να βρει παρόμοιες πρωτεΐνες με τη μεταφραζόμενη ακολουθία ελέγχου, μέσα σε μια πρωτεϊνική βάση δεδομένων.
4. tblastn: συγκρίνει μία πρωτεϊνική ακολουθία ελέγχου μέσα σε μία βάση νουκλεοτιδικών ακολουθιών, οι οποίες δυναμικά μπορούν να μεταφραστούν

σε όλα τα αναγνωστικά πλαίσια. Συνεπώς το πρόγραμμα μπορεί να χρησιμοποιηθεί για να βρει παρόμοιες πρωτεΐνες, σε μία βάση δεδομένων «μεταφραζόμενων νουκλεοτιδικών ακολουθιών».

Η βάση δεδομένων Fishtrace χρειάζεται μόνο το πρώτο πρόγραμμα, εφόσον αυτό έχει να κάνει με νουκλεοτιδικές ακολουθίες. Έτσι η μοναδική επιλογή στο πεδίο αυτό είναι το πρόγραμμα *blastn* (<http://www.ncbi.nlm.nih.gov/BLAST>).

Το δεύτερο πεδίο έχει να κάνει με την επιλογή της βάσης δεδομένων, μέσα στην οποία θα συγκριθεί η ακολουθία που θα εισαχθεί από τον χρήστη (query sequence). Το Εθνικό Κέντρο Βιοτεχνολογίας χρησιμοποιεί τις δικές του βάσεις δεδομένων οι οποίες αφορούν το ανθρώπινο γονιδίωμα, το γονιδίωμα ποντικών, αρουραίων, διαφόρων μικροβίων και εντόμων. Για το σκοπό της εργασίας χρησιμοποιήθηκε η βάση δεδομένων της Fishtrace (Εικόνα 7).

Το τρίτο πεδίο αφορά την εισαγωγή της καινούργιας και καθαρής ακολουθίας DNA από τον χρήστη, η οποία πρόκειται να συγκριθεί με τις αλληλουχίες της βάσης δεδομένων. Αυτή η ακολουθία δεν είναι απαραίτητο να είναι σε «FASTA format», όπως προτείνει το πρόγραμμα.

Αφού ο χρήστης τοποθετήσει την ακολουθία του DNA, μπορεί να προχωρήσει στην αναζήτηση πατώντας το κουμπί «search» (Εικόνα 7). Υπάρχει και η επιλογή «εκκαθάρισης», σε περίπτωση που ο χρήστης έχει κάνει λάθος στην εισαγόμενη ακολουθία.

Το δεύτερο κομμάτι της φόρμας του BLAST (Εικόνα 8), έχει επιλογές σχετικά με τη στατιστική ανάλυση της ακολουθίας από το πρόγραμμα.

Ως προεπιλογή το πρόγραμμα φιλτράρει την ακολουθία για περιοχές «χαμηλής πολυπλοκότητας». Αυτό σημαίνει ότι μπορεί να εξαλείψει στατιστικά σημαντικές αλλά βιολογικά «αδιάφορες» αναφορές από το αποτέλεσμα, αφήνοντας τις πιο ενδιαφέρουσες «βιολογικά» περιοχές της ελεγχόμενης ακολουθίας, συγκριτικά με τις ακολουθίες της βάσης δεδομένων. Το φιλτράρισμα εφαρμόζεται μόνο στην ελεγχόμενη ακολουθία και όχι στις ακολουθίες της βάσης δεδομένων. Η δεύτερη επιλογή όσον αφορά το φιλτράρισμα, έχει να κάνει με ακολουθίες προερχόμενες από το ανθρώπινο DNA, και γι' αυτό το λόγο η επιλογή αυτή δεν λειτουργεί.

The query sequence is **filtered** for low complexity regions by default.

Filter Low complexity Mask for lookup table only

Expect Matrix Perform ungapped alignment

Query Genetic Codes (blastx only)

Database Genetic Codes (tblast[nx] only)

Frame shift penalty for blastx

Other advanced options:

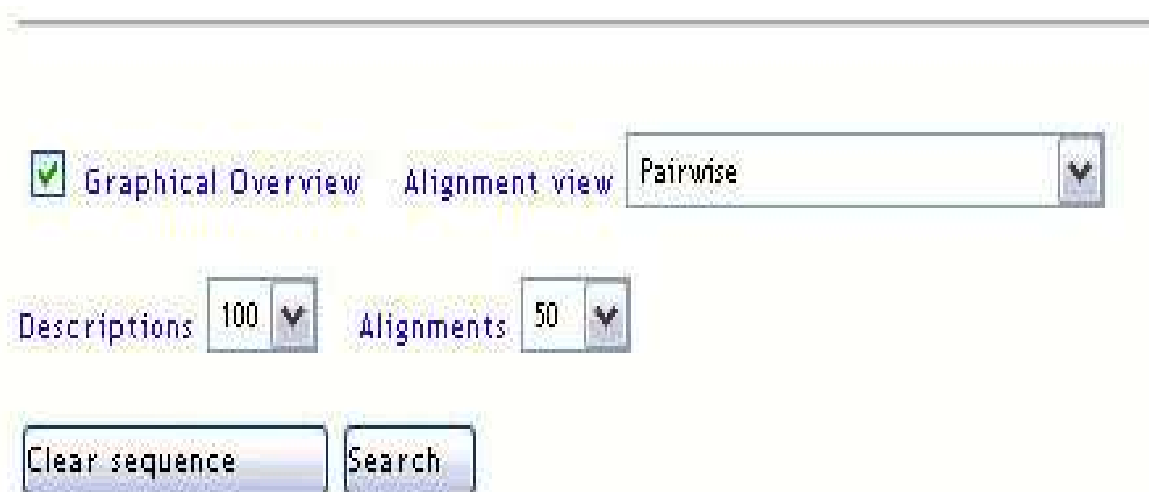
Εικόνα 8. Το εργαλείο BLAST, τομέας δεύτερος.

Η επόμενη επιλογή ονομάζεται «όριο στατιστικής σημαντικότητας» (the statistical significance threshold), και αναφέρεται στο ταίριασμα με τις ακολουθίες της βάσης δεδομένων. Η προκαθορισμένη τιμή είναι 10, δηλαδή 10 ταιριάσματα αναμένονται να βρεθούν κατά τύχη, σύμφωνα με το στοχαστικό μοντέλο των Karlin & Altschul (1990) (<http://www-bimas.cit.nih.gov/blastinfo/KAstat.html>).

Στη συνέχεια υπάρχει η επιλογή του μαθηματικού προτύπου. Ένα στοιχείο κλειδί σε αυτόν τον υπολογισμό είναι η «μήτρα αντικατάστασης» (substitution matrix), η οποία αναθέτει μία τιμή για την ευθυγράμμιση κάθε πιθανού ζεύγους συγκρινόμενων ακολουθιών. Τα αποτελέσματα (bit scores) ομαλοποιούνται, το οποίο σημαίνει ότι αποτελέσματα από διαφορετικές ευθυγραμμίσεις ακολουθιών (alignments) μπορούν να συγκριθούν, ακόμη και εάν έχουν χρησιμοποιηθεί διαφορετικές «μήτρες αντικατάστασης». Στο πρόγραμμα BLAST, προκαθορισμένη επιλογή είναι η «μήτρα αντικατάστασης» BLOSUM-62.

Οι υπόλοιπες επιλογές αφορούν τα προγράμματα blastx, tblastn και tblastx, και όχι το πρόγραμμα blastn που χρησιμοποιεί η Fishtrace. Γι' αυτό το λόγο οι επιλογές αυτές είναι ανενεργές στο πρόγραμμα.

Το τρίτο κομμάτι της φόρμας του BLAST (Εικόνα 9), έχει επιλογές σχετικά με τα αποτελέσματα και την εμφάνισή τους.



The image shows a section of the BLAST search interface. It features a horizontal line at the top. Below it, there are several interactive elements: a checked checkbox for 'Graphical Overview', a radio button for 'Alignment view', and a dropdown menu set to 'Pairwise'. Below these are two more dropdown menus: 'Descriptions' set to '100' and 'Alignments' set to '50'. At the bottom of this section are two buttons: 'Clear sequence' and 'Search'.

Εικόνα 9. Το εργαλείο BLAST, τομέας τρίτος.

Η πρώτη επιλογή που είναι επιλεγμένη ως προεπιλογή, αφορά τη γραφική απεικόνιση των αποτελεσμάτων. Το πρόγραμμα δίνει μία αναφορά των ακολουθιών από τη βάση δεδομένων, οι οποίες έχουν ευθυγραμμιστεί με την ακολουθία ελέγχου. Το αποτέλεσμα κάθε ευθυγράμμισης δίνεται με ένα από πέντε διαφορετικά χρώματα, τα οποία χωρίζουν όλη τη γκάμα των αποτελεσμάτων του προγράμματος σε πέντε ομάδες.

Η δεύτερη επιλογή αφορά την όψη των αποτελεσμάτων ευθυγράμμισης (Alignment view). Συνήθως το πρόγραμμα παραθέτει τα αποτελέσματα κατά ζεύγη (pairwise) δηλαδή της σύγκρισης μεταξύ δύο ακολουθιών, της ελεγχόμενης και μιας ακολουθίας από τη βάση δεδομένων αναζήτησης.

Η τρίτη επιλογή αφορά τις περιγραφές. Η επιλογή αυτή περιορίζει τον αριθμό των περιγραφών των ακολουθιών που ταιριάζουν, μετά από τη σάρωση της βάσης δεδομένων αναζήτησης. Το προκαθορισμένο όριο είναι οι 100 περιγραφές.

Η τέταρτη και τελευταία επιλογή αφορά τις ευθυγραμμίσεις ακολουθιών (Alignments). Η επιλογή αυτή περιορίζει τα αποτελέσματα ευθυγράμμισης, μόνο σε εκείνες τις ακολουθίες για τις οποίες αναφέρονται ζεύγη τμημάτων με καλά αποτελέσματα (High Scoring Segment Pairs – HSPs). Το προκαθορισμένο όριο είναι 50 ακολουθίες. Εάν περισσότερες από 50 ακολουθίες από τη βάση δεδομένων

συμβαίνει να ικανοποιούν αυτό το όριο στατιστικής σημαντικότητας, τότε μόνο αυτές με τις υψηλότερες τιμές θα αναφερθούν.

2.3. ΠΡΟΓΡΑΜΜΑ ΓΙΑ ΑΝΑΛΥΣΗ RFLP

Η μέθοδος των RFLPs βασίζεται στη χρήση ειδικών ενζύμων περιορισμού που αποκαλούνται «ενδονουκλεάσες περιορισμού». Τα ένζυμα περιορισμού κόβουν το DNA σε μια συγκεκριμένη ακολουθία αναγνώρισης 4-6 ζευγών βάσεων. Συνεπώς οδηγούν σε ένα πρότυπο το οποίο αποτελείται από ορισμένο αριθμό κομματιών (θραυσμάτων) DNA. Οι αλλαγές στον αριθμό και στο μέγεθος των θραυσμάτων (πολυμορφισμοί), μπορούν να προέλθουν κυρίως από νουκλεοτιδικές αντικαταστάσεις μέσα στις θέσεις κοπής, από αναδιατάξεις ακολουθιών και από προσθήκες ή ελλείμματα. Έτσι οι νουκλεοτιδικές αντικαταστάσεις μέσα στις θέσεις κοπής οδηγούν στη δημιουργία ή την απώλεια θέσεων. Τα θραύσματα που προκύπτουν μετά από πέψη του DNA με ένζυμα περιορισμού, διαχωρίζονται σε πηκτές αγαρόζης ή πολυακρυλαμίδης.

Για τη βάση δεδομένων της Fishtrace έπρεπε να βρεθεί ένα εργαλείο που θα εξυπηρετούσε αυτή τη μέθοδο, για την εύρεση φυλογενετικής συγγένειας μεταξύ των ειδών. Αρχικά έγινε μια έρευνα στο διαδίκτυο για διάφορα εργαλεία που χρησιμοποιούν περιοριστικά ένζυμα και παρέχουν τη δυνατότητα στο χρήστη να τοποθετήσει μια ακολουθία ελέγχου, καθώς επίσης και ένα σύνολο περιοριστικών ενζύμων. Τελικά βρέθηκε ένα τέτοιο πρόγραμμα για ανάλυση των RFLPs, και προσαρμόστηκε στις ανάγκες της γενετικής βάσης δεδομένων της Fishtrace (<http://vm.cfsan.fda.gov/~frf/rflp.html>).

Για το λόγο αυτό δημιουργήθηκε ένα πεδίο με ένζυμα περιορισμού, ένα πεδίο με όλα τα είδη ψαριών που βρίσκονται στη βάση δεδομένων Fishtrace και ένα πεδίο όπου ο χρήστης έχει τη δυνατότητα να επιλέξει μεταξύ της ακολουθίας του γονιδίου για το κυτόχρωμα B (CytB), και της ακολουθίας του γονιδίου για τη ροδοψίνη (Rhodopsin). Στη βάση δεδομένων Fishtrace υπάρχουν αποθηκευμένες DNA ακολουθίες των δύο γονιδίων, για πολλά είδη ψαριών.

Η φόρμα του προγράμματος για την ανάλυση των RFLPs έχει τέσσερις τομείς (Εικόνα 10).

Select restriction enzymes:

<div style="border: 1px solid gray; padding: 2px;"> AarI AasI AatI AatII AacI AacII </div>	<div style="border: 1px solid gray; height: 60px; width: 100%;"></div>
<input type="button" value="Clear"/>	

Select species:

<div style="border: 1px solid gray; padding: 2px;"> Abudefduf luridus Agonus cataphractus Alosa alosa Alosa fallax Aluterus scriptus Ammodytes tobianus </div>	<div style="border: 1px solid gray; height: 60px; width: 100%;"></div>
<input type="button" value="Clear"/>	

Select sequence type and PCR primers:

CytB
 CytB3' Select primer
 CytB5' Select primer
 Rhodopsin

[VIEW](#)
[Enzyme list](#) [More info](#)

Use DNA sequence specified in the species list
 Use DNA sequence entered in the field below
 Use enzymes specified in the enzyme list
 Use defined oligonucleotide sequences below

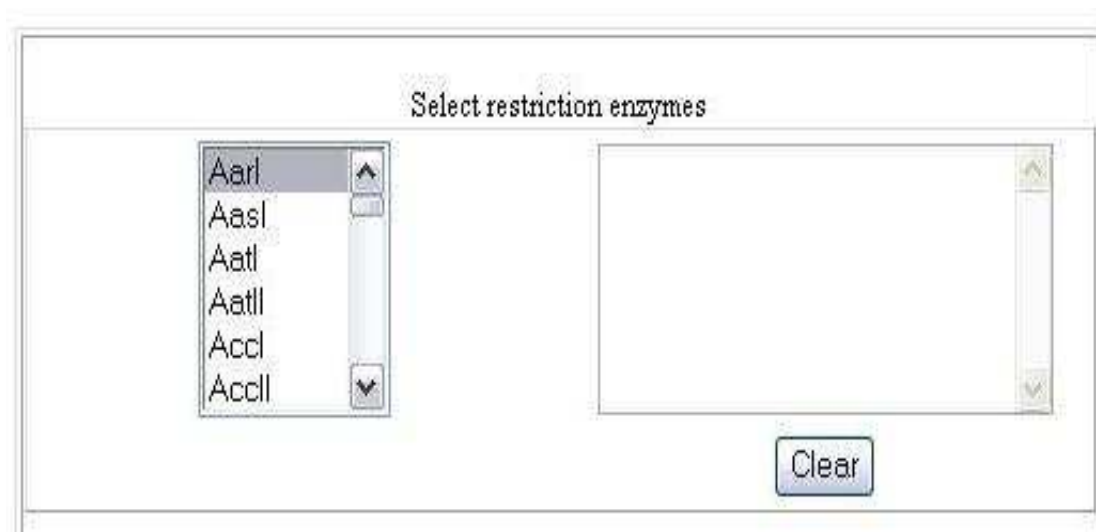
Enter DNA sequence ex: AAGAATTCGGTAA

Define oligonucleotide sequences and put symbol ^ at the cutting point (ex: aga^ggg)

Name(no space)	Oligonucleotide(^ to cut)

Εικόνα 10. Το εργαλείο για τη μέθοδο των RFLPs.

Το πρώτο κομμάτι της φόρμας (Εικόνα 11) δίνει τη δυνατότητα επιλογής ενός ενζύμου περιορισμού, μεταξύ 616 διαθέσιμων εμπορικών περιοριστικών ενζύμων.

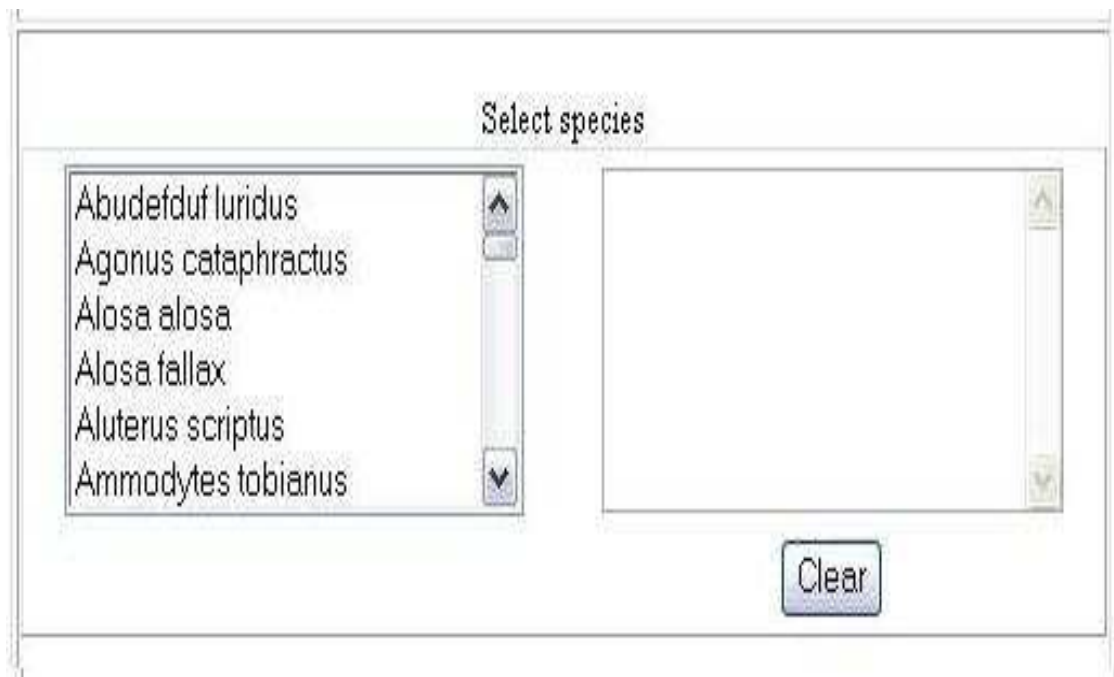


Εικόνα 11. Το εργαλείο για ανάλυση RFLPs, τομέας πρώτος.

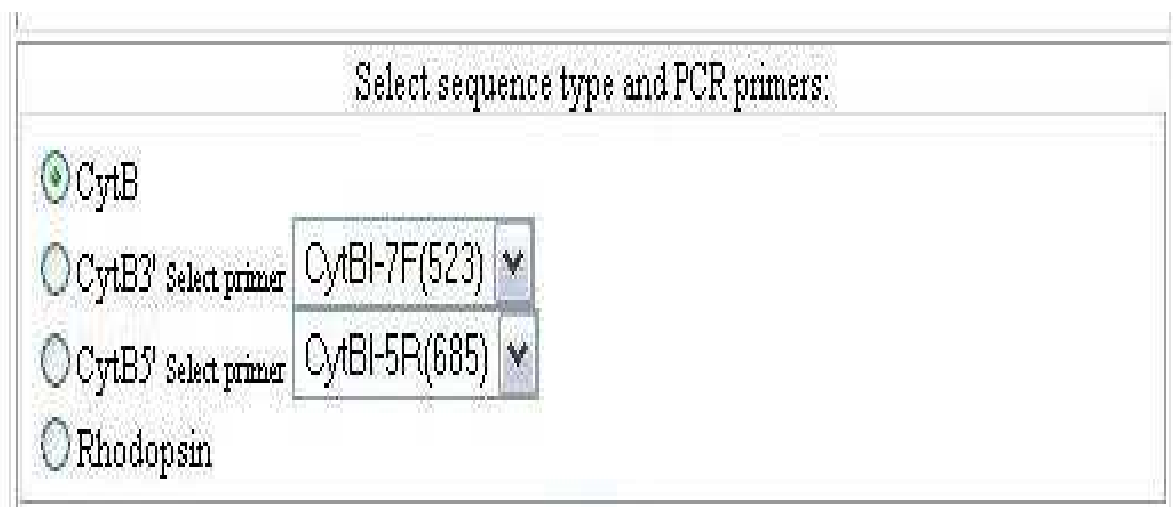
Σε αυτό το σημείο ο χρήστης έχει τη δυνατότητα να επιλέξει ένα ή περισσότερα ένζυμα περιορισμού από τη λίστα που βρίσκεται στο αριστερό μέρος, τα οποία εμφανίζονται στο δεξί μέρος. Υπάρχει και η επιλογή «εκκαθάρισης», σε περίπτωση που έγινε λάθος επιλογή.

Το δεύτερο κομμάτι της φόρμας (Εικόνα 12) παρέχει την επιλογή ενός είδους ψαριού μεταξύ ενός μεγάλου αριθμού ειδών, τα οποία προέρχονται από τη βάση δεδομένων της Fishtrace. Και σε αυτό το σημείο ο χρήστης έχει τη δυνατότητα να επιλέξει ένα ή περισσότερα είδη ιχθύων από τη λίστα που βρίσκεται στο αριστερό μέρος, τα οποία εμφανίζονται στο δεξί μέρος. Υπάρχει και η επιλογή «εκκαθάρισης», σε περίπτωση που έγινε λάθος επιλογή.

Το τρίτο κομμάτι της φόρμας (Εικόνα 13) παρέχει τη δυνατότητα επιλογής του τύπου της νουκλεοτιδικής ακολουθίας (κυτόχρωμα Β, ροδοψίνη), για το επιλεγμένο είδος ψαριού. Στο ίδιο πεδίο δίνεται και η δυνατότητα επιλογής των εκκινητών που χρησιμοποιήθηκαν στην αλυσιδωτή αντίδραση πολυμεράσης (PCR), για την ενίσχυση του συγκεκριμένου γονιδίου.



Εικόνα 12. Το εργαλείο για ανάλυση RFLPs, τομέας δεύτερος.



Εικόνα 13. Το εργαλείο για ανάλυση RFLPs, τομέας τρίτος.

Το τέταρτο και τελευταίο κομμάτι της φόρμας (Εικόνα 14), παρέχει ένα συνδυασμό επιλογών ανάλογα με την κάθε περίπτωση. Δηλαδή παρέχει τη δυνατότητα στο χρήστη να επιλέξει μια ακολουθία DNA η οποία προέρχεται από τη βάση δεδομένων της Fishtrace (κυτόχρωμα B, ροδοψίνη), ή να τοποθετήσει μια νέα ακολουθία DNA. Επίσης, παρέχει τη δυνατότητα στο χρήστη να επιλέξει εάν τα ένζυμα περιορισμού που θέλει να ελέγξει θα προέρχονται από τη λίστα που βρίσκεται

παραπάνω, ή να ελέγξει κάποια νέα ένζυμα που δε βρίσκονται στη λίστα. Στην τελευταία περίπτωση, θα πρέπει να εισάγει την ονομασία του ενζύμου καθώς και την

VIEW
[Enzyme list](#) [More info](#)

Use DNA sequence specified in the species list
 Use DNA sequence entered in the field below
 Use enzymes specified in the enzyme list
 Use defined oligonucleotide sequences below

Enter DNA sequence ex:AAGAATTCGGTTA

Define oligonucleotide sequences and put symbol ^ at the cutting point (ex:aga^ggg)

Name(no space)	Oligonucleotide(^ to cut)

Εικόνα 14. Το εργαλείο για ανάλυση RFLPs, τομέας τέταρτος.

ακολουθία αναγνώρισής του, μαζί με το σύμβολο « ^ » που θα ορίζει την ακριβή θέση κοπής του ενζύμου. Ο ερευνητής μπορεί να χρησιμοποιήσει αυτές τις δύο επιλογές συνδυαστικά π.χ. νουκλεοτιδική ακολουθία από τη Fishtrace ελεγχόμενη με ένζυμα περιορισμού που υπάρχουν στη λίστα του προγράμματος, νουκλεοτιδική ακολουθία από τη Fishtrace ελεγχόμενη με ένζυμα περιορισμού που δεν υπάρχουν στη λίστα του προγράμματος, νέα ακολουθία ελεγχόμενη με ένζυμα περιορισμού που υπάρχουν στη λίστα του προγράμματος, νέα ακολουθία ελεγχόμενη με ένζυμα περιορισμού που δεν υπάρχουν στη λίστα του προγράμματος.

2.4. PHYLIP (PHYLOGENY INFERENCE PACKAGE)

Αρχικά δημιουργήθηκε ένα αρχείο «text», με τις αλληλουχίες DNA σε format τύπου PHYLIP. Στη συνέχεια, γίνεται εκκίνηση του προγράμματος «Drawgram». Το αρχείο «text» πρέπει να είναι στον τρέχοντα κατάλογο. Αφού καθοριστεί αυτό, εμφανίζεται το μενού του προγράμματος:

DRAWTREE from PHYLIP version 3.61

```
Reading tree ...
Tree has been read.
Loading the font ...
Drawtree: can't find font file "fontfile"
Please enter a new file name> font1
Font loaded.
```

Rooted tree plotting program version 3.61

```
Here are the settings:
0 Screen type (IBM PC, ANSI): IBM PC
P Final plotting device: Postscript printer
V Previewing device: MS Windows display
H Tree grows: Horizontally
S Tree style: Phenogram
B Use branch lengths: Yes
```

L Angle of labels: 90.0
R Scale of branch length: Automatically rescaled
D Depth/Breadth of tree: 0.53
T Stem-length/tree-depth: 0.05
C Character ht / tip space: 0.3333
A Ancestral nodes: Weighted
F Font: Times-Roman
M Horizontal margins: 1.65 cm
M Vertical margins: 2.16 cm
Pages per tree: one page per tree
Y to accept these or type the letter for one to change

Εμφανίζονται τα κύρια πεδία που έχουν ήδη μια προκαθορισμένη παράμετρο. Σε αυτό το σημείο θα πρέπει να γίνει αλλαγή της συσκευής γραφικής αποτύπωσης σε MS-Windows Bitmap, πληκτρολογώντας το γράμμα «P» που αντιστοιχεί σε αυτή την επιλογή.

Which plotter or printer will the tree be drawn on?
(Many other brands or models are compatible with these)

Type: **to choose one compatible with**

L	Postscript printer file format
M	PICT format (for drawing programs)
J	HP Laserjet PCL file format
W	MS-Windows Bitmap
F	FIG 2.0 drawing program format
A	Idraw drawing program format
Z	VRML Virtual Reality Markup

Language file

P	PCX file format (for drawing programs)
K	TeKtronix 4010 graphics terminal

- X X Bitmap format
- V POVRAY 3D rendering program file
- R Rayshade 3D rendering program file
- H Hewlett-Packard pen plotter (HPGL file format)
- D DEC ReGIS graphics (VT240 terminal)
- E Epson MX-80 dot-matrix printer
- C Prowriter/Imagewriter dot-matrix printer
- T Toshiba 24-pin dot-matrix printer
- O Okidata dot-matrix printer
- B Houston Instruments plotter
- U other: one you have inserted code for

Το θεμιτό αποτέλεσμα πρέπει να βγαίνει σε MS-Windows Bitmap. Το γράμμα - επιλογή που αντιστοιχεί σε MS-Windows Bitmap είναι το «W». Μετά την πληκτρολόγηση του γράμματος, το πρόγραμμα ζητάει και τις διαστάσεις του «δέντρου». Έτσι αρχικά χρησιμοποιούνται οι εξής διαστάσεις: 640 X 400. Στη συνέχεια, οι ρυθμίσεις γίνονται αποδεκτές χρησιμοποιώντας το κουμπί «Enter» του πληκτρολογίου.

Choose one:

W

Please select the MS-Windows bitmap file resolution

X resolution?

640

Y resolution?

400

Στην τελική οθόνη του προγράμματος φαίνεται ότι το πρόγραμμα είναι έτοιμο να δημιουργήσει το δέντρο, και ότι για να γίνει αυτό θα πρέπει να γίνει επιβεβαίωση πληκτρολογώντας το γράμμα «y».

Rooted tree plotting program version 3.65

Here are the settings:

- O Screen type (IBM PC, ANSI): IBM PC
- P Final plotting device: MS-Windows Bitmap (640 by 400 resolution)
- V Previewing device: MS Windows display
- H Tree grows: Horizontally
- S Tree style: Phenogram
- B Use branch lengths: Yes
- L Angle of labels: 90.0
- R Scale of branch length: Automatically rescaled
- D Depth/Breadth of tree: 0.53
- T Stem-length/tree-depth: 0.05
- C Character ht / tip space: 0.3333
- A Ancestral nodes: Weighted
- M Horizontal margins: 51.20 cm
- M Vertical margins: 32.00 cm
- # Pages per tree: one page per tree

Y to accept these or type the letter for one to change

Το πρόγραμμα «Drawgram» ανοίγει ένα καινούργιο παράθυρο, όπου μπορούμε να δούμε μια επισκόπηση του δέντρου.

3. ΑΠΟΤΕΛΕΣΜΑΤΑ

3.1. ΠΡΟΓΡΑΜΜΑ BLAST

Αφού ακολουθηθούν όλα τα βήματα πάνω στη φόρμα του BLAST (Εικόνα 6), ο χρήστης πιέζει την επιλογή «Search» για να δει τα αποτελέσματα. Η ιστοσελίδα παραμένει ως έχει, και τα αποτελέσματα εμφανίζονται σε καινούργιο «παράθυρο» που ενεργοποιείται αυτόματα από το πρόγραμμα.

Η σελίδα των αποτελεσμάτων παραθέτει αρκετές πληροφορίες, σχετικά με την έρευνα που έγινε μέσω του προγράμματος BLAST. Επίσης τα αποτελέσματα χωρίζονται σε τρεις διαφορετικούς τομείς. Για την παρουσίαση των αποτελεσμάτων, έγινε μία έρευνα στο πρόγραμμα BLAST και χρησιμοποιήθηκε η νουκλεοτιδική ακολουθία του είδους *Mugil cephalus* για το γονίδιο του κυτοχρώματος B (cytB), η οποία εξάχθηκε από μία νουκλεοτιδική βάση δεδομένων. Η ακολουθία αυτή συγκρίθηκε με τις νουκλεοτιδικές ακολουθίες της βάσης δεδομένων Fishtrace και τα αποτελέσματα παραθέτονται παρακάτω.

Στον πρώτο τομέα των αποτελεσμάτων (Εικόνα 15) εμφανίζεται η έκδοση του προγράμματος καθώς και η ημερομηνία. Ακολουθεί η αναφορά στους επιστήμονες που εργάστηκαν για το πρόγραμμα BLAST, καθώς και τα πλήρη στοιχεία της εργασίας (έτος δημοσίευσης, τίτλος εργασίας, περιοδικό). Στη συνέχεια αναφέρεται το μήκος της ελεγχόμενης ακολουθίας σε αριθμό γραμμάτων, δηλαδή στη συγκεκριμένη αναζήτηση η ακολουθία είχε 1141 γράμματα που σημαίνει 1141 αζωτούχες βάσης (A,T,G,C). Στην επόμενη σειρά, δηλώνεται η βάση δεδομένων μέσα στην οποία θα γίνει η αναζήτηση (στην προκειμένη περίπτωση είναι η βάση δεδομένων της Fishtrace) μαζί με το συνολικό αριθμό των νουκλεοτιδικών ακολουθιών που αυτή διαθέτει, μεταφρασμένο και σε γράμματα.

Στη συνέχεια αναφέρονται οι ακολουθίες που βρέθηκαν να ευθυγραμμίζονται σημαντικά με την ελεγχόμενη ακολουθία, συγκριτικά με τις ακολουθίες της βάσης δεδομένων Fishtrace πάντα. Εδώ παρατηρούνται τρεις στήλες. Στην πρώτη στήλη αναγράφεται η επιστημονική ονομασία του είδους καθώς και ο τύπος της νουκλεοτιδικής ακολουθίας. Στη δεύτερη στήλη αναγράφονται τα αποτελέσματα (score) της σύγκρισης, δηλαδή πόσα γράμματα της ελεγχόμενης ακολουθίας μετά από την ευθυγράμμιση, βρέθηκαν κοινά με την ακολουθία της βάσης δεδομένων που συγκρίθηκε. Στην τρίτη στήλη αναγράφεται η αναμενόμενη τιμή E, που όπως

αναφέρθηκε παραπάνω όσο πιο μικρή είναι τόσο πιο πολύ ταιριάζουν τα συγκρινόμενα δείγματα.



FishTrace Blast results

BLASTN 2.2.12 [Aug-07-2005]

Reference:

Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", *Nucleic Acids Res.* 25:3389-3402.

Query=

(1141 letters)

Database: C:\fishtrace\blast\data\FT_specimen.nt
1677 sequences; 1,340,621 total letters

Sequences producing significant alignments:	Score (bits)	E Value
Mugil-cephalus (CytB)	2262	0.0
Mugil-cephalus (CytB)	2246	0.0
Liza-ramado (CytB)	456	e-129
Liza-ramado (CytB)	456	e-129
Liza-aurata (CytB)	414	e-116
Liza-aurata (CytB)	410	e-115
Chelon-labrosus (CytB)	406	e-114
Liza-aurata (CytB)	402	e-113
Liza-aurata (CytB)	402	e-113
Chelon-labrosus (CytB)	398	e-111
Liza-aurata (CytB)	394	e-110
Epinephelus-costae (CytB)	379	e-105
Epinephelus-costae (CytB)	379	e-105
Epinephelus-caninus (CytB)	379	e-105
Epinephelus-costae (CytB)	371	e-103
Agonus-cataphractus (CytB)	341	2e-094
Agonus-cataphractus (CytB)	341	2e-094

<u>Thunnus-thynnus (CytB)</u>	<u>335</u>	<u>1e-092</u>
<u>Thunnus-thynnus (CytB)</u>	<u>335</u>	<u>1e-092</u>
<u>Thunnus-alalunga (CytB)</u>	<u>335</u>	<u>1e-092</u>
<u>Thunnus-alalunga (CytB)</u>	<u>335</u>	<u>1e-092</u>
<u>Thunnus-alalunga (CytB)</u>	<u>335</u>	<u>1e-092</u>
<u>Thunnus-alalunga (CytB)</u>	<u>335</u>	<u>1e-092</u>
<u>Thunnus-alalunga (CytB)</u>	<u>335</u>	<u>1e-092</u>
<u>Thunnus-alalunga (CytB)</u>	<u>335</u>	<u>1e-092</u>
<u>Thunnus-alalunga (CytB)</u>	<u>335</u>	<u>1e-092</u>
<u>Thunnus-alalunga (CytB)</u>	<u>335</u>	<u>1e-092</u>
<u>Thunnus-alalunga (CytB)</u>	<u>335</u>	<u>1e-092</u>
<u>Thunnus-alalunga (CytB)</u>	<u>335</u>	<u>1e-092</u>
<u>Thunnus-alalunga (CytB)</u>	<u>335</u>	<u>1e-092</u>
<u>Thunnus-alalunga (CytB)</u>	<u>335</u>	<u>1e-092</u>
<u>Dactylopterus-volitans (CytB)</u>	<u>317</u>	<u>3e-087</u>
<u>Tylosurus-acus acus (CytB)</u>	<u>311</u>	<u>2e-085</u>
<u>Tylosurus-acus acus (CytB)</u>	<u>311</u>	<u>2e-085</u>
<u>Tylosurus-acus acus (CytB)</u>	<u>311</u>	<u>2e-085</u>
<u>Polyprion-americanus (CytB)</u>	<u>311</u>	<u>2e-085</u>
<u>Polyprion-americanus (CytB)</u>	<u>311</u>	<u>2e-085</u>
<u>Chromis-limbata (CytB)</u>	<u>305</u>	<u>1e-083</u>
<u>Taractichthys-longipinnis (CytB)</u>	<u>305</u>	<u>1e-083</u>
<u>Taractichthys-longipinnis (CytB)</u>	<u>305</u>	<u>1e-083</u>
<u>Diplodus-sargus (CytB)</u>	<u>305</u>	<u>1e-083</u>
<u>Beryx-splendens (CytB)</u>	<u>303</u>	<u>5e-083</u>
<u>Polyprion-americanus (CytB)</u>	<u>303</u>	<u>5e-083</u>
<u>Polyprion-americanus (CytB)</u>	<u>303</u>	<u>5e-083</u>
<u>Boops-boops (CytB)</u>	<u>303</u>	<u>5e-083</u>
<u>Boops-boops (CytB)</u>	<u>303</u>	<u>5e-083</u>
<u>Thunnus-albacares (CytB)</u>	<u>303</u>	<u>5e-083</u>
<u>Chromis-limbata (CytB)</u>	<u>297</u>	<u>3e-081</u>
<u>Chromis-limbata (CytB)</u>	<u>297</u>	<u>3e-081</u>
<u>Chromis-limbata (CytB)</u>	<u>297</u>	<u>3e-081</u>
<u>Diplodus-sargus (CytB)</u>	<u>297</u>	<u>3e-081</u>
<u>Diplodus-sargus (CytB)</u>	<u>297</u>	<u>3e-081</u>
<u>Diplodus-sargus (CytB)</u>	<u>297</u>	<u>3e-081</u>
<u>Diplodus-sargus (CytB)</u>	<u>297</u>	<u>3e-081</u>
<u>Diplodus-sargus (CytB)</u>	<u>297</u>	<u>3e-081</u>
<u>Diplodus-sargus (CytB)</u>	<u>297</u>	<u>3e-081</u>
<u>Diplodus-sargus (CytB)</u>	<u>297</u>	<u>3e-081</u>
<u>Diplodus-sargus (CytB)</u>	<u>297</u>	<u>3e-081</u>
<u>Myoxocephalus-scorpilus (CytB)</u>	<u>297</u>	<u>3e-081</u>
<u>Boops-boops (CytB)</u>	<u>295</u>	<u>1e-080</u>
<u>Boops-boops (CytB)</u>	<u>295</u>	<u>1e-080</u>
<u>Boops-boops (CytB)</u>	<u>295</u>	<u>1e-080</u>
<u>Boops-boops (CytB)</u>	<u>295</u>	<u>1e-080</u>
<u>Boops-boops (CytB)</u>	<u>295</u>	<u>1e-080</u>

<u>Thunnus-thynnus (CytB)</u>	<u>287</u>	<u>3e-078</u>
<u>Thunnus-thynnus (CytB)</u>	<u>287</u>	<u>3e-078</u>
<u>Thunnus-thynnus (CytB)</u>	<u>287</u>	<u>3e-078</u>
<u>Thunnus-thynnus (CytB)</u>	<u>287</u>	<u>3e-078</u>
<u>Thunnus-thynnus (CytB)</u>	<u>287</u>	<u>3e-078</u>
<u>Thunnus-albacares (CytB)</u>	<u>287</u>	<u>3e-078</u>
<u>Thunnus-thynnus (CytB)</u>	<u>280</u>	<u>8e-076</u>
<u>Pagellus-acarne (CytB)</u>	<u>278</u>	<u>3e-075</u>
<u>Euthynnus-alletteratus (CytB)</u>	<u>278</u>	<u>3e-075</u>
<u>Euthynnus-alletteratus (CytB)</u>	<u>278</u>	<u>3e-075</u>
<u>Pomadasys-perotaei (CytB)</u>	<u>276</u>	<u>1e-074</u>
<u>Taurulus-bubalis (CytB)</u>	<u>276</u>	<u>1e-074</u>
<u>Seriola-rivoliana (CytB)</u>	<u>274</u>	<u>5e-074</u>
<u>Seriola-rivoliana (CytB)</u>	<u>274</u>	<u>5e-074</u>
<u>Anguilla-anguilla (CytB)</u>	<u>274</u>	<u>5e-074</u>
<u>Oblada-melanura (CytB)</u>	<u>272</u>	<u>2e-073</u>
<u>Oblada-melanura (CytB)</u>	<u>272</u>	<u>2e-073</u>
<u>Oblada-melanura (CytB)</u>	<u>272</u>	<u>2e-073</u>
<u>Oblada-melanura (CytB)</u>	<u>272</u>	<u>2e-073</u>
<u>Oblada-melanura (CytB)</u>	<u>272</u>	<u>2e-073</u>
<u>Oblada-melanura (CytB)</u>	<u>272</u>	<u>2e-073</u>

Εικόνα 15. Αποτελέσματα του προγράμματος BLAST, τομέας πρώτος.

Όπως φαίνεται και από την Εικόνα 15, τα αποτελέσματα της αναζήτησης εμφανίζονται σε αύξουσα ταξινόμηση.

Θα πρέπει εδώ να αναφερθεί ότι τα αποτελέσματα της πρώτης στήλης αποτελούν links για τα δεδομένα του δείγματος (Εικόνα 16). Αυτό σημαίνει ότι εάν ο χρήστης κάνει κλικ με τον δείκτη του ποντικιού επάνω στο όνομα του κάθε δείγματος, τότε θα μπορέσει να πάρει μία σειρά από πληροφορίες οι οποίες αφορούν μορφομετρικά χαρακτηριστικά του δείγματος, τόπο δειγματοληψίας, τρόπο αποθήκευσης κλπ. (Εικόνα 16).

Specimen data**FishTrace code:** MugCep-EM-04**Specimen**

Species code: 785
Identification date: 10/10/2004
Identified by: Leontarakis,P
Voucher fixation: formol 10pc
Voucher preservation: alcohol_70deg
Voucher collection institute: MNHN
Voucher collection No.: 2006-1169 (Link to MNHN)
Total length(mm): 497
Fork length(mm): 445
Standard length(mm): 408
Total weight(g): 1094
Ontogenic stage: Adult
Head length(mm): 93
Body depth(mm): 78
Sex: Unknown
Dorsal fin-rays: IV
2nd Dorsal fin-rays: 9
Anal fin-rays: III 8
Pectoral fin-rays: 18
Scales on lateral line: 43
Otolith collection: MNHN
Otolith collection No.: Right: ICOT-001612, Left: ICOT-001613
Id biogeography: NAGREF_56
Remarks: Measurements taken in fresh specimen.

Collecting data

Locality: Northern Aegean
Depth: 1
Fishing method: other
Latitude(dec deg): 40°50.24' N
Longitude(dec deg): 25°56.24' E
Coordinate source: GPS
Date of capture or purchase : 10/10/2004
Collector or purchaser: Leontarakis,P

Collecting data

Locality: Northern Aegean
Depth: 1
Fishing method: other
Latitude(dec deg): 40°50.24' N
Longitude(dec deg): 25°56.24' E
Coordinate source: GPS
Date of capture or purchase : 10/10/2004
Collector or purchaser: Leontarakis,P

Tissue sample

id tissue: MugCep-EM-04a
Tissue collection: MNHN
Storage medium: ethanol 100pc
 •
id tissue: MugCep-EM-04b
Tissue collection: MNHN
Storage medium: ethanol 100pc

Pictures

Type: voucher
Name: MugCep-EM-04.jpg
 •
Type: otolith
Name: MugCep-EM-04-oto.jpg
Comments Right otolith

GO UP **Εικόνα 16.** Δεδομένα του δείγματος, που βρέθηκε με τα αποτελέσματα του BLAST.

Στη συνέχεια ακολουθεί ο δεύτερος τομέας των αποτελεσμάτων. Σε αυτό τον τομέα εμφανίζονται αναλυτικά όλα τα στάδια της συγκριτικής έρευνας, βάση προς βάση, καθώς και η επιτυχία της εκάστοτε σύγκρισης. Δηλαδή αναγράφονται, η επιστημονική ονομασία του είδους καθώς και η νουκλεοτιδική ακολουθία από τη βάση δεδομένων Fishtrace, η οποία βρέθηκε να έχει το μεγαλύτερο ποσοστό ευθυγράμμισης με την ελεγχόμενη ακολουθία. Στη συνέχεια αναγράφεται ο κωδικός του συγκεκριμένου δείγματος και επαναλαμβάνεται ο τύπος της νουκλεοτιδικής ακολουθίας. Τέλος το πρόγραμμα παραθέτει το μήκος της συγκρινόμενης ακολουθίας, τα αποτελέσματα της σύγκρισης, την αναμενόμενη τιμή E και το ποσοστό ομοιότητας των συγκρινόμενων ακολουθιών.

Παρακάτω δίνονται τα αποτελέσματα της πρώτης ευθυγράμμισης (alignment) των δύο ακολουθιών, δηλαδή αυτής που βρέθηκε να ταιριάζει πιο πολύ στην ακολουθία που αναζητήθηκε μέσω του BLAST:

```
>Mugil-cephalus\(CytB\)> specimen:MugCep-CB-04==Mugil cephalus(CytB)
Length = 1141
```

```
Score = 2262 bits (1141), Expect = 0.0
Identities = 1141/1141 (100%)
Strand = Plus / Plus
```

```
Query: 1
atgaccagcatccggaaaagccaccccctgttaaagattgcgaaactccgcgctagttgac 60
```

```
|||||
Sbjct: 1
atgaccagcatccggaaaagccaccccctgttaaagattgcgaaactccgcgctagttgac 60
```

```
Query: 61
ctcccagccccagctaacaatttctgcatgatgaaactttggctcactccttggactttgc 120
```

```
|||||
Sbjct: 61
ctcccagccccagctaacaatttctgcatgatgaaactttggctcactccttggactttgc 120
```

```
Query: 121
ttaatttcacaaaattgtcacaggcctattccttgccatacactatacccctgaaacctca 180
```

```
|||||
Sbjct: 121
ttaatttcacaaaattgtcacaggcctattccttgccatacactatacccctgaaacctca 180
```

```
Query: 181
tccgccttttccctccgtagccacatctgccgagacgtaaaactacggctgactaatccgc 240
```

```
|||||
```

Sbjct: 181
tccgccttttctccgtagccacatctgccgagacgtaaactacggctgactaatccgc 240

Query: 241
aacatacacgctaacggggcatctttttctttatctgcatttacattcacatcggacga 300

|||||
Sbjct: 241
aacatacacgctaacggggcatctttttctttatctgcatttacattcacatcggacga 300

Query: 301
ggcctttattacggctcgtacctatacaaaagaaacttgaaacattggagtcgtcctcctc 360

|||||
Sbjct: 301
ggcctttattacggctcgtacctatacaaaagaaacttgaaacattggagtcgtcctcctc 360

Query: 361
ctcctagtaataataaccgcatttggttgatacgtcctcccctgaggacaaatccttc 420

|||||
Sbjct: 361
ctcctagtaataataaccgcatttggttgatacgtcctcccctgaggacaaatccttc 420

Query: 421
tgaggtgccaccgtaatcaccaacctcctctctgctgtcccctacattggagaatcccta 480

|||||
Sbjct: 421
tgaggtgccaccgtaatcaccaacctcctctctgctgtcccctacattggagaatcccta 480

Query: 481
gttcaatggatctgagggggattctcagtagacaacgccacactcacccgattcttcgcc 540

|||||
Sbjct: 481
gttcaatggatctgagggggattctcagtagacaacgccacactcacccgattcttcgcc 540

Query: 541
ttccacttccttctcccatttgtaatcctcgcctcacattaattcatcttatcttcctc 600

|||||
Sbjct: 541
ttccacttccttctcccatttgtaatcctcgcctcacattaattcatcttatcttcctc 600

Query: 601
cacgagaccggctcaaacaacccccctcggattaccctccaactcagacaaaattcccttc 660

|||||
Sbjct: 601
cacgagaccggctcaaacaacccccctcggattaccctccaactcagacaaaattcccttc 660

Query: 661
caccctattacacttataaggacattccttggtttctagtagtacttctccttacactcatc 720

|||||

Sbjct: 661
caccctattacacttataaggacattccttggtttctagtagtacttctccttacactcatc 720

Query: 721
tcccttgactatgtgctcccaatcttttaggagaccagacaacttcaccctgccaac 780

|||||

Sbjct: 721
tcccttgactatgtgctcccaatcttttaggagaccagacaacttcaccctgccaac 780

Query: 781
cctatggtcactcctgcacacattaaccagaatgatactttctggttgctatgccatt 840

|||||

Sbjct: 781
cctatggtcactcctgcacacattaaccagaatgatactttctggttgctatgccatt 840

Query: 841
cttcgatcaatcccaaaacttgaggagtaattgcacttctcgcttccatcctcgtc 900

|||||

Sbjct: 841
cttcgatcaatcccaaaacttgaggagtaattgcacttctcgcttccatcctcgtc 900

Query: 901
ctactagttgtaccaatcctccacacatccaaacaccgaagcctaactccgccccgtc 960

|||||

Sbjct: 901
ctactagttgtaccaatcctccacacatccaaacaccgaagcctaactccgccccgtc 960

Query: 961
acacaattcctggttctgactcctcatcgcggatgtgcttgtccttacctgaatcggagga 1020

|||||

Sbjct: 961
acacaattcctggttctgactcctcatcgcggatgtgcttgtccttacctgaatcggagga 1020

Query: 1021
ataccagtagaagaccctacattatcacggggcaaatcgcttcagcactgtacttctcc 1080

|||||

Sbjct: 1021
ataccagtagaagaccctacattatcacggggcaaatcgcttcagcactgtacttctcc 1080

Query: 1081
ctattcctaatacctaataccaatggcagcatgagtcgagaataaaaatgttaacctgatgt 1140

|||||

Query: 352
gtcctcctcctcctagtaataataaacccgatttggtagatagctcctcccctgaggacaa 411
|||
Sbjct: 352
atcctactccttttagtaataataaacccgattcgtaggatagctcctcccctgaggccaa 411

Query: 412
atataccttctgagggtgccaccgtaatacacaacctcctctctgctgtcccctacattgga 471
||
Sbjct: 412
atataccttctgagggtgctactgtaatacacaacctcctttccgctgtcccctacatcggg 471

Query: 472
gaatccctagttcaatggatctgagggggattctcagtagacaacgccacactcaccgga 531
|||
Sbjct: 472
gattcccttgtccaatgaatttgaggaggcttttcagttgacaatgctactctcacacga 531

Query: 532 ttctttgccttccacttctctccc 557
|||
Sbjct: 532 ttctttgccttccacttctctccc 557

Score = 131 bits (66), Expect = 4e-031
Identities = 186/226 (82%)
Strand = Plus / Plus

Query: 731
tatttgctcccaatcttttaggagaccagacaacttcaccctgccaaccctatgggtca 790
|
Sbjct: 731
tatttgcaaccaacttctaggggaccagacaacttcaccctgcaaaccactagttta 790

Query: 791
ctcctgcacacattaaccagaatgatactttctgtttgcgatgccattcttcgatcaa 850
|
Sbjct: 791
cccccccatatcaagcccgaatgatacttctctttgcatagccattctccgctcca 850

Query: 851
tcccaaacaaacttgaggagtaattgcacttctcgcttccatcctcgtcctactagttg 910
|||
Sbjct: 851
tcctaacaactaggaggcgttcttgcacttctatgctcgtccttaaatagttg 910

Query: 911 taccaatcctccacacatccaaacaccgaagcctaacattccgccc 956
|||
Sbjct: 911 tccccatcctccacacatctaaacatcgaagcctaacattccgccc 956

Score = 63.9 bits (32), Expect = 8e-011
Identities = 41/44 (93%)
Strand = Plus / Plus

Query: 1001 tccttacctgaatcggaggaataaccagtagaagaccctacatt 1044
 ||||| ||||||||||||||||| || |||||||||||||||||||||
Sbjct: 1001 tcctaacctgaatcggaggcatgccagtagaagaccctacatt 1044

Score = 36.2 bits (18), Expect = 0.019
Identities = 33/38 (86%)
Strand = Plus / Plus

Query: 592 atcttcctccacgagaccggctcaaacaaccctcgcg 629
 ||||||| || || || || ||||||||||||||||| |||||
Sbjct: 592 atcttccttcatgaaacaggctcaaacaaccctcgcg 629

Στον τρίτο και τελευταίο τομέα των αποτελεσμάτων, αναγράφεται ξανά η βάση δεδομένων και πληροφορίες σχετικά με το περιεχόμενο καθώς και την ημερομηνία ανανέωσής της. Στη συνέχεια δίνονται στατιστικές πληροφορίες, σχετικές με τον τρόπο λειτουργίας του προγράμματος.

Database: C:\fishtrace\blast\data\FT_specimen.nt
Posted date: Aug 3, 2006 11:18 AM
Number of letters in database: 1,340,621
Number of sequences in database: 1677

Lambda	K	H
1.37	0.711	1.31

Gapped

Lambda	K	H
1.37	0.711	1.31

Matrix: blastn matrix:1 -3
Gap Penalties: Existence: 5, Extension: 2
Number of Hits to DB: 6458
Number of Sequences: 1677
Number of extensions: 6458
Number of successful extensions: 5494
Number of sequences better than 10.0: 835
Number of HSP's better than 10.0 without gapping: 835
Number of HSP's successfully gapped in prelim test: 0
Number of HSP's that attempted gapping in prelim test: 0
Number of HSP's gapped (non-prelim): 4585
length of query: 1141
length of database: 1,340,621
effective HSP length: 15
effective length of query: 1126

effective length of database: 1,315,466
effective search space: 1481214716
effective search space used: 1481214716
T: 0
A: 0
X1: 11 (21.8 bits)
X2: 15 (29.7 bits)
S1: 12 (24.3 bits)
S2: 14 (28.2 bits)

3.2. ΠΡΟΓΡΑΜΜΑ ΓΙΑ ΑΝΑΛΥΣΗ RFLP

Ξεκινώντας το λογισμικό πρόγραμμα για την ανάλυση RFPL, γίνεται επιλογή των περιοριστικών ενζύμων, επιλογή των ειδών και επιλογή μιας νουκλεοτιδικής ακολουθίας η οποία βρίσκεται στη βάση δεδομένων Fishtrace, ή γίνεται τοποθέτηση μιας νέας ακολουθίας DNA. Στη συνέχεια, ο χρήστης πιέζει την επιλογή της φόρμας «submit» με αποτέλεσμα να ανοίξει μια καινούργια σελίδα με τα αποτελέσματα.

Ας υποθέσουμε ότι ο χρήστης έχει εισάγει μία νέα ακολουθία DNA. Στη σελίδα των αποτελεσμάτων αναγράφεται η ανεπεξέργαστη αλληλουχία που εισήχθηκε, και ένα διάγραμμα στο οποίο ο άξονας Y είναι η λίστα των περιοριστικών ενζύμων που χρησιμοποιήθηκαν και ο άξονας X είναι η προσομοίωση της μετακίνησης των περιοριστικών θραυσμάτων σε πηκτή αгарόζης 2%. Οι αριθμοί στον άξονα X, δηλώνουν το μέγεθος των περιοριστικών θραυσμάτων σε ζεύγη βάσεων. Επίσης, πάνω στο γράφημα φαίνονται τα σημεία κοπής του ενζύμου και έτσι μπορεί να προβλεφθεί το μέγεθος των περιοριστικών θραυσμάτων.

Τα περιοριστικά ένζυμα αναγνωρίζουν μία νουκλεοτιδική ακολουθία πάνω στην αλληλουχία DNA, και κόβουν την ακολουθία σε αυτά τα σημεία. Οι αριθμοί, στον άξονα X του γραφήματος, είναι βαθμίδες που βοηθούν στον υπολογισμό του μεγέθους των κομματιών. Έτσι υπάρχουν οι βαθμίδες 100, 200, 300, 500 και 1000. Σε αυτό το παράδειγμα φαίνεται ότι μόνο ένα ένζυμο ήταν ικανό να κόψει την αλληλουχία DNA, και αυτό ήταν το ένζυμο με την ονομασία AluI. Αυτό το ένζυμο έκοψε την αλληλουχία σε τέσσερα σημεία, τα οποία δηλώνονται στο γράφημα με μικρές κάθετες μαύρες γραμμές. Τα νούμερα 437, 148, 555 και 29 είναι τα μεγέθη των κομματιών σε ζεύγη βάσεων. Έτσι το νούμερο 29 βρίσκεται πριν την βαθμίδα 100, το νούμερο 148 βρίσκεται πριν την βαθμίδα 200, το νούμερο 437 βρίσκεται πριν την βαθμίδα 500 και το νούμερο 555 βρίσκεται πριν την βαθμίδα 1000. Το άθροισμα αυτών των κομματιών δίνουν ολόκληρο το μέγεθος της αλληλουχίας (Εικόνα 17).

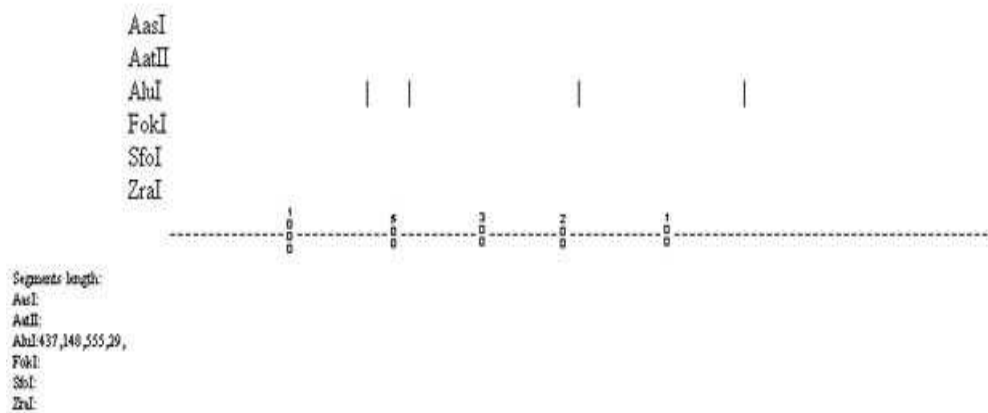


RFLP gel simulation tool v1
Joint Research Center-Agrifish 2006
Contact philippe.carreau@jrc.it

Sequence type Entered sequence ATGGCAAGCCTCCGAAAACTCACCCGCTACTAAAAATCGCTAACGACGCACTAGTTGACCTTCCTACCCCTCTAATAT
CTCTGCATGATGAAACTTTGGCTCACTACTTGGCCTTTGGCCTTATTTCTCAGATCCTTACAGGACTATTCTCGCAATAC
ACTACACCCCTGATGTGGAATCAGCCTTCGCCTCAGTAGCCACATTTCCCGAGATGTCAACTTCGGTTGACTTATCCGG
AACCTCCACGCAAAACGGGGCTCTTTCTTTTATCTGTATCTACTCCACATCGGCGGAGGACTTTACTACGGCTCTTA
CCTATACAAAGAAAACATGAAACATCGGAGTAGTACTCCTACTCCTAGTTATGATGACCGCCTTCGTTGGCTACGTCCTCC
CCTGAGGACAAAATGTCTTTCTGAGGAGCTACCGTCATTACTAACCTCCTATCCCGAGTCCCATATGTCCGAACTACTCTC
GTTGAATGAAATCTGAGGAGGCTTTTTCAGTAGACAATGCCACCCCTCACCCGATTCTTCCGATTCCACTTCCTATTCCCAT
CGTCATCCGAGCTATGACAATTCTTCACCTTCTTTTCTTCACGAAAACAGGTTCAAATAATCCAATCGGATTAAACTCAA
ATGCAGATAAAATCTCATTCCACCCATACTTCTCTTACAAAGACCTCCTTGGTTTGGTGATCCTGCTAGTAGCACTCGCC
TCTTAGCACTATTCTCCCTAACCTCCTAGGAGATCCAGACAACCTTCAACCCTGCCAACCCAAATGGTTACTCCACCTCA
CATTAAACCTGAATGATATTTCTATTTGCATACGCAATCTTCGGTCCATTCCAAACAACTAGGAGGAGTACTAGCCC
TCCTAGCCTCCATCCTTGTACTTATAGTAGTTCCCTTCCTGCACACTTCAAAAACAGGGAACCTTAACATTCCGACCCAGTT
TCCCAATTCCTATTCTGAACCCCTTATTGCAGACGTAGCCATTCTTACCTGAATCGGGGTATGCCAGCAGAACAGCCCTT
CATTATTATCGGCCAAGTAGCCTCCGCTCCTACTTCTCCCTATTCTTGTTTTCTTCCCACTTGCAGGCTGAGCTGAGA ACAAATCCTTGGATGATCCT

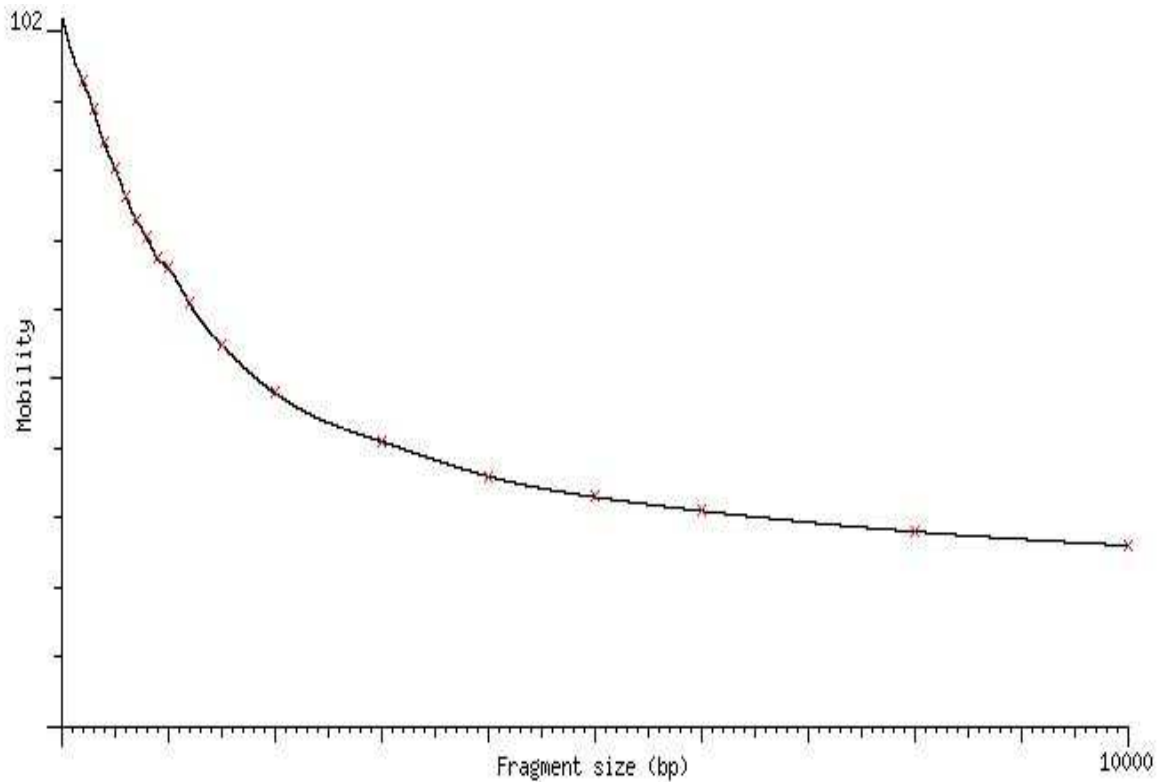
Y axis: List of enzymes
X axis: Simulation of fragment migration in 2% agarose gel (Numbers are fragment size indication)

Species Not specified



Εικόνα 17. Σελίδα αποτελεσμάτων του προγράμματος για ανάλυση RFLP.

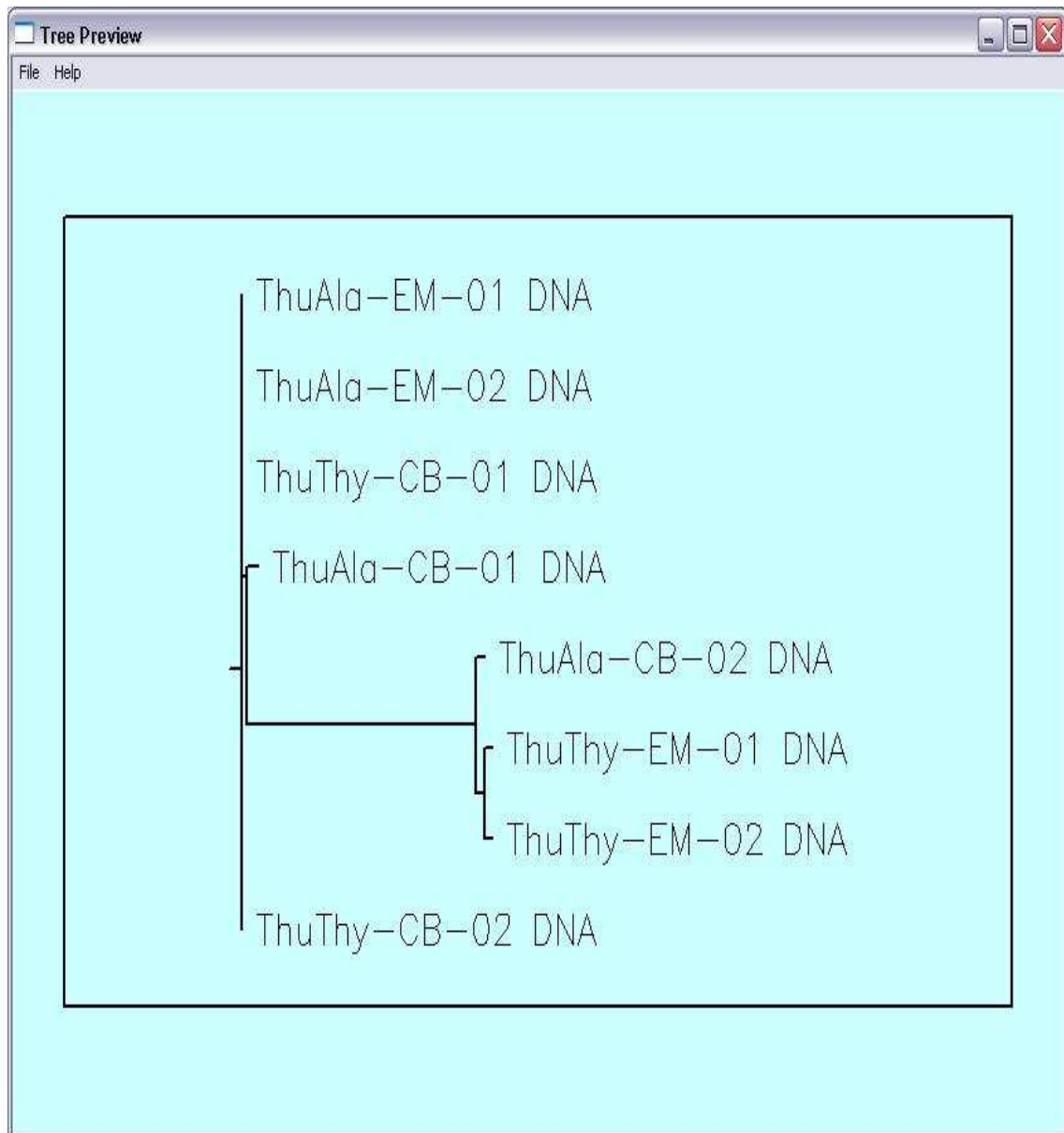
Τα αποτελέσματα της παραπάνω διαδικασίας μπορούν να δοθούν και σαν γραφική παράσταση συνάρτησης (Εικόνα 18). Στον άξονα Y φαίνεται η κινητικότητα των περιοριστικών θραυσμάτων και στον άξονα X φαίνεται το μέγεθος των περιοριστικών θραυσμάτων σε ζεύγη βάσεων.



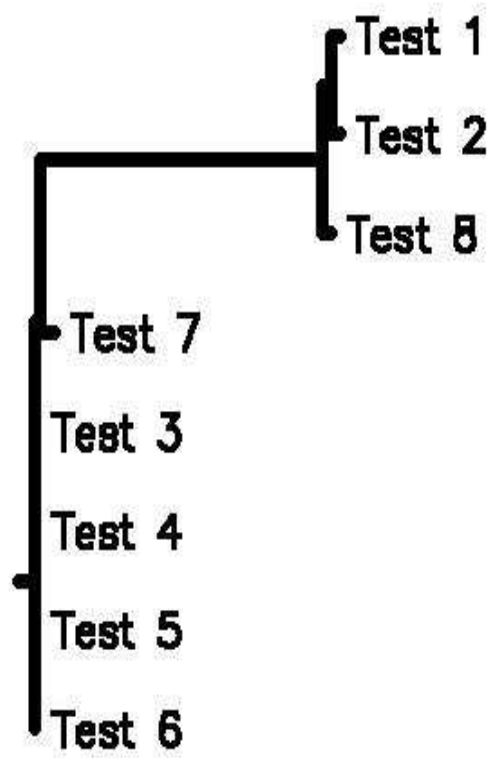
Εικόνα 18. Γράφημα συνάρτησης αποτελεσμάτων του προγράμματος για ανάλυση RFLP.

3.3. PHYLIP (PHYLOGENY INFERENCE PACKAGE)

Μετά την εισαγωγή των δεδομένων και την προσαρμογή των ρυθμίσεων στο λογισμικό πρόγραμμα PHYLIP, το επόμενο βήμα είναι η εξαγωγή των αποτελεσμάτων. Η εφαρμογή «Drawgram» ανοίγει ένα καινούργιο παράθυρο, για να γίνει μια επισκόπηση του δέντρου πριν την επιβεβαιώσουμε (Εικόνα 19). Αν τα αποτελέσματα είναι τα επιθυμητά, από το μενού «File» (από το παράθυρο που μόλις άνοιξε) γίνεται η επιλογή «plot». Ένα νέο αρχείο με την ονομασία «plotfile» θα πρέπει να εμφανιστεί στον τρέχοντα κατάλογο. Στη συνέχεια υπάρχει η δυνατότητα μετονομασίας του αρχείου αυτού ως «plotfile.bmp». Το αρχείο αυτό είναι ικανό να ανοίξει πολύ εύκολα, με ένα πρόγραμμα προβολής εικόνας των windows (Εικόνα 20).



Εικόνα 19. Επισκόπηση φυλογενετικού δέντρου μέσω του προγράμματος «Drawgram».



Εικόνα 20. Το αρχείο «plotfile» σε μορφή «.bmp»

4. ΣΥΖΗΤΗΣΗ

4.1. ΓΕΝΙΚΑ

Η εργασία πραγματοποιήθηκε στο Joint Research Centre (Ιταλία) στα πλαίσια του προγράμματος FISHREG, κατά την περίοδο Ιούλιος 2005 - Μάιος 2006. Ένας από τους βασικούς στόχους του προγράμματος, είναι η δημιουργία μιας ιστοσελίδας που θα περιέχει γενικές πληροφορίες σχετικές με διάφορα εμπορικά είδη ιχθύων (www.fishtrace.com). Κομμάτι της ιστοσελίδας στο οποίο αναφέρεται η εργασία, αφορά τα στατιστικά εργαλεία.

Σκοπός της παρούσας εργασίας ήταν η εύρεση των κατάλληλων εργαλείων, η προσαρμογή τους στις παραμέτρους της βάσης δεδομένων και η ένταξη τους στην ιστοσελίδα FishTrace.

Για την επιλογή και την προσαρμογή των προγραμμάτων απαιτήθηκε ο συνδυασμός δύο βασικών επιστημών, της Πληροφορικής και της Βιολογίας. Η επιστήμη της Πληροφορικής είναι απαραίτητη για τη διαχείριση του όγκου των σημερινών δεδομένων που παράγονται στο χώρο της Μοριακής Βιολογίας, τα οποία πρέπει να αναλυθούν και να επεξεργαστούν. Ο επιστημονικός χώρος της ένωσης αυτών των πεδίων ονομάζεται διεθνώς Βιοπληροφορική.

Για την ένταξη αυτών των προγραμμάτων στην ιστοσελίδα της FishTrace, έπρεπε πρώτα να γίνει μια εκτίμηση των αναγκών της ιστοσελίδας και των δυνατοτήτων των προγραμμάτων. Έτσι τα απαραίτητα βήματα για την επιλογή και την ένταξη τους στην ιστοσελίδα είναι τα εξής: α) καθορισμός των αναγκών της ιστοσελίδας, σε εργαλεία επεξεργασίας γενετικών δεδομένων β) εύρεση των απαιτούμενων προγραμμάτων γ) προσαρμογή αυτών στις παραμέτρους της βάσης δεδομένων σύμφωνα με τα χαρακτηριστικά και το είδος των δεδομένων δ) δοκιμή της λειτουργίας και της επιτυχίας με τη χρήση των δεδομένων της FishTrace και ε) ένταξη τους στην ιστοσελίδα της FishTrace. Είναι απαραίτητος ο σωστός σχεδιασμός, για την υποστήριξη και εξασφάλιση της σωστής λειτουργίας μιας ιστοσελίδας τέτοιου τύπου.

Η Fishgen (<http://fishgen.jrc.it>) ήταν μια πρώτη απόπειρα δημιουργίας μιας ιστοσελίδας που συλλέγει όλες τις γενετικές πληροφορίες για μεγάλο αριθμό εμπορικών ειδών (Imsiridou et al. 2003). Κύριος στόχος της δημιουργίας αυτής της

βάσης δεδομένων ήταν να συλλεχθούν όλες οι γενετικές πληροφορίες σχετικές με το διαχωρισμό και την αναγνώριση πληθυσμών, σε ένα διαδουκτιακό τόπο. Αυτό θα βοηθήσει τους επιστήμονες να πραγματοποιήσουν σύγκριση και ταυτοποίηση των δεδομένων τους με τα ήδη υπάρχοντα στην ιστοσελίδα. Με τη χρήση των δεδομένων της σελίδας θα μπορεί να γίνεται αναγνώριση της προέλευσης ενός πληθυσμού ψαριών ή ενός αποθέματος, από ένα μόνο μικρό δείγμα. Δηλαδή οι επιστήμονες θα είναι σε θέση με μια DNA ανάλυση ενός δείγματος, να γνωρίζουν την ακριβή τοποθεσία σύλληψης του δείγματος.

Η σελίδα της FishTrace δημιουργήθηκε από τον Philippe Carreau. Η τελική ιστοσελίδα που αφορά το κομμάτι των εργαλείων επεξεργασίας γενετικών πληροφοριών, λειτουργεί κανονικά. Η σελίδα φτιάχτηκε για να είναι πολύ εύκολη στη χρήση για όλους τους χρήστες.

Η διατήρηση των ιχθυοπληθυσμών επιβάλει την ύπαρξη της ιστοσελίδας. Τη τελευταία δεκαετία η υπεραλίευση σε συνδυασμό με τις παράνομες αλιευτικές δραστηριότητες σε όλο τον κόσμο, έχει οδηγήσει στη δραματική μείωση των πληθυσμών των κυριότερων εμπορικών ιχθύων. Οι αρμόδιες αρχές παγκοσμίως σε συνεργασία με επιστήμονες και οικολογικές οργανώσεις εργάζονται προς την κατεύθυνση προστασίας των ιχθυοαποθεμάτων. Παρόλα τα δραστικά μέτρα που λαμβάνονται κατά καιρούς, η παράνομη αλιεία ανθίζει. Η αποτελεσματική διαχείριση της αλιείας και των ιχθυοαποθεμάτων, διευκολύνεται και από τις γενετικές τεχνικές διάκρισης των πληθυσμών.

Πολλοί δείκτες στο παρελθόν έχουν χρησιμοποιηθεί στην αναγνώριση ειδών και διαχείριση των πληθυσμών. Χαρακτηριστικά παραδείγματα είναι οι δείκτες σωματικής και γοναδικής ανάπτυξης καθώς και η χρήση μορφομετρικών παραμέτρων και αναλύσεων πτερυγίων. Οι δείκτες αυτοί έχουν το βασικό μειονέκτημα ότι είναι άμεσα συνδεδεμένοι με το περιβάλλον διαβίωσης, και είναι περισσότερο φαινοτυπικοί και όχι γενοτυπικοί δείκτες.

Τα γενετικά δεδομένα τα οποία προκύπτουν από την εφαρμογή των γενετικών δεικτών σε ιχθυοαποθέματα, μπορούν να συγκριθούν με αντίστοιχα από παρόμοιες έρευνες και να βοηθήσουν στη ταυτοποίηση της προέλευσης ενός πληθυσμού ψαριών. Πιο άμεση μέθοδος ανίχνευσης πολυμορφισμού είναι η ανάλυση πρωτοδιάταξης, γιατί παρέχει απευθείας ανάγνωση της ακολουθίας του DNA. Πέρα από την ύπαρξη των δεδομένων για την εκάστοτε χρήση, δίνεται η δυνατότητα σε

επιστήμονες που έχουν πραγματοποιήσει κάποια σχετική έρευνα να καταχωρήσουν τα δεδομένα τους στην ιστοσελίδα.

4.2. ΕΦΑΡΜΟΓΕΣ ΤΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ BLAST

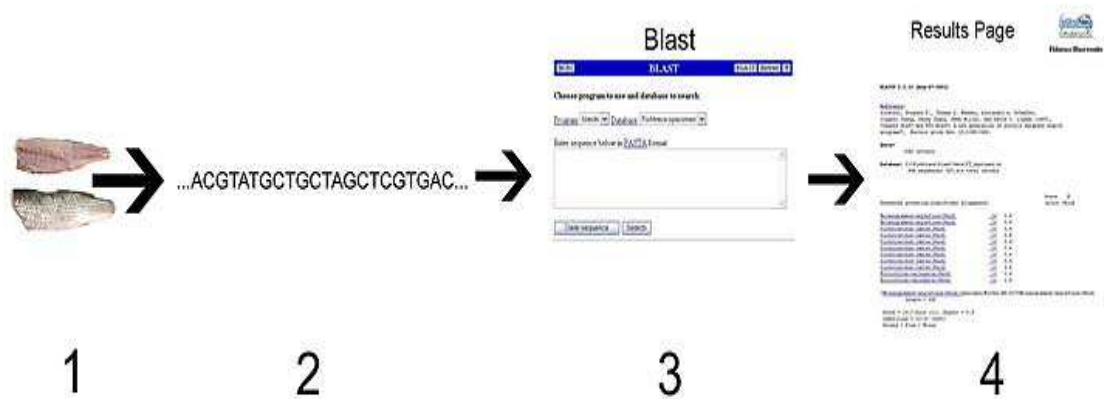
Το πρόγραμμα BLAST βρίσκει εφαρμογή στον εντοπισμό της απάτης στο εμπόριο ψαριών και στην παράνομη αλιεία. Για την πρόληψη της απάτης στο εμπόριο ψαριών (νωπά ή επεξεργασμένα), γίνεται δειγματοληψία πριν την πώληση και ακολουθεί εξαγωγή DNA. Στη συνέχεια οι επιστήμονες μπορούν να προχωρήσουν στην ταυτοποίηση του είδους του ψαριού που πωλείται στην αγορά σε οποιαδήποτε μορφή. Τα αποτελέσματα της σύγκρισης που θα γίνει με το πρόγραμμα BLAST, θα αποκαλύψουν την ύπαρξη απάτης ή μη στο εμπόριο αλιευμάτων.

Μια από τις εφαρμογές του προγράμματος είναι η εξακρίβωση της παράνομης αλιείας. Οι άγριοι πληθυσμοί ψαριών που ζουν στην θάλασσα έχουν κάποιες γενετικές διαφορές οι οποίες περιγράφονται με τον όρο «απλότυπος» (haplotype). Ο απλότυπος είναι μια ομάδα πολυμορφικών νουκλεοτιδίων, τα οποία συνδέονται με ένα συγκεκριμένο χρωμόσωμα. Η ταυτοποίηση ορισμένων νουκλεοτιδίων σε έναν απλότυπο, αναμφίβολα αναγνωρίζει και τις υπόλοιπες πολυμορφικές θέσεις στην ίδια περιοχή. Ο όρος «απλοτυπική ποικιλότητα» περιγράφει τη μοναδικότητα κάποιων απλοτύπων σε έναν πληθυσμό ψαριών. Έτσι, οι απλότυποι μπορούν να χρησιμοποιηθούν για τη σύγκριση και ταυτοποίηση διαφορετικών πληθυσμών ψαριών (<http://en.wikipedia.org/wiki/Haplotype>).

Οι ομάδες απλοτύπων (haplogroups) μπορούν να χρησιμοποιηθούν για να ορίσουν γενετικούς πληθυσμούς ψαριών, οι οποίοι είναι συνήθως γεωγραφικά προσανατολισμένοι. Στο μέλλον λοιπόν το πρόγραμμα BLAST μπορεί να χρησιμοποιηθεί για τον εντοπισμό της παράνομης, μη καταγεγραμμένης και μη ρυθμισμένης αλιείας (IUU - Illegal Unreported and Unregulated), καθώς επίσης και για τον έλεγχο της υπεραλίευσης σημαντικών θαλάσσιων ιχθυοαποθεμάτων τα οποία απειλούνται.

Παράδειγμα : Προσομοίωση μιας εφαρμογής για τον εντοπισμό απάτης με την χρήση του προγράμματος BLAST (Εικόνα 21):

1. Ο επιθεωρητής λαμβάνει δείγμα από το φρέσκο ψάρι ή από το φιλέτο ψαριού.
2. Ο ιστός που λήφθηκε, μεταφέρεται στο εργαστήριο και ακολουθεί εξαγωγή DNA. Στη συνέχεια, γίνεται η επεξεργασία του με τις μεθόδους της PCR και της ανάλυσης πρωτοδιάταξης, για την αποκάλυψη της νουκλεοτιδικής ακολουθίας.
3. Αφού διαβαστεί η νουκλεοτιδική ακολουθία, ο επιθεωρητής χρησιμοποιεί το εργαλείο BLAST εισάγοντας την και συγκρίνοντας την με ακολουθίες της βάσης δεδομένων της Fishtrace.
4. Στο τέλος λαμβάνει τα αποτελέσματα μέσω της ιστοσελίδας τα οποία αποκαλύπτουν την πραγματική ταυτότητα του ψαριού από όπου λήφθηκε το δείγμα προς εξέταση, δηλαδή αν υπήρξε ή όχι απάτη.



Εικόνα 21. Προσομοίωση μιας εφαρμογής του προγράμματος BLAST, για τον εντοπισμό απάτης

Έτσι, γίνεται σύγκριση των εργαστηριακών αποτελεσμάτων με τα δεδομένα της ιστοσελίδας Fishtrace. Στην περίπτωση που το δείγμα του ερευνητή έχει παρόμοια νουκλεοτιδική ακολουθία με ένα είδος ψαριού που βρίσκεται στη γενετική βάση δεδομένων Fishtrace, τότε μάλλον το δείγμα του ανήκει στο παραπάνω είδος. Πρέπει βέβαια να σημειωθεί το γεγονός ότι μόνο ο συνδυασμένος έλεγχος με αποτελέσματα από διάφορες τεχνικές, θα επιτρέψει στον ερευνητή να είναι αρκετά σίγουρος για την ταυτότητα του δείγματος. Συνεπώς, πρέπει να γίνει ανάλυση του δείγματος με διαφορετικές τεχνικές DNA και για το λόγο αυτό η σελίδα περιέχει τρία διαφορετικά εργαλεία για την ταυτοποίηση των ειδών. Το γεγονός αυτό μπορεί να βοηθήσει τις αρμόδιες αρχές στον έλεγχο της παράνομης αλιείας και στη διατήρηση πολλών ιχθυοπληθυσμών.

Στόχος των αρμόδιων αρχών, είναι να ελέγχουν τον τόπο αλιείας των ψαριών που βρίσκονται στη αγορά καθώς και την ακριβή ταυτότητα των ειδών. Οι πληροφορίες από τη βάση δεδομένων σε συνδυασμό με τα στατιστικά εργαλεία μπορούν να βοηθήσουν τους επιστήμονες και τις αρμόδιες αρχές στη ταυτοποίηση των ειδών και τον διαχωρισμό των πληθυσμών, ώστε να προβούν στην πάταξη της παράνομης αλιείας και στη διατήρηση των ιχθυοποθεμάτων.

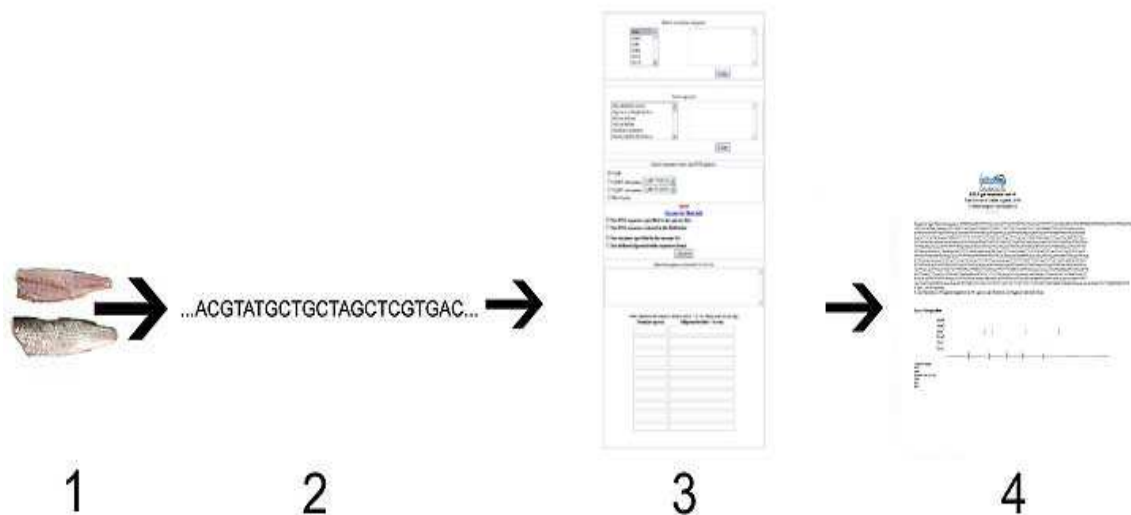
4.3. ΕΦΑΡΜΟΓΕΣ ΤΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ ΓΙΑ ΑΝΑΛΥΣΗ RFLP

Όπως το πρόγραμμα BLAST έτσι και το πρόγραμμα για ανάλυση RFLP, βρίσκει εφαρμογή στον εντοπισμό της απάτης στο εμπόριο ψαριών και στην παράνομη αλιεία. Έτσι, χρησιμοποιείται για την ταυτοποίηση του είδους του ψαριού που πρόκειται να ελεγχθεί.

Παράδειγμα : Προσομοίωση μιας εφαρμογής για τον εντοπισμό απάτης με την χρήση του προγράμματος για ανάλυση RFLP (Εικόνα 22):

1. Ο επιθεωρητής λαμβάνει δείγμα από το φρέσκο ψάρι ή από το φιλέτο ψαριού.
2. Ο ιστός που λήφθηκε, μεταφέρεται στο εργαστήριο και ακολουθεί εξαγωγή DNA. Στη συνέχεια γίνεται η επεξεργασία του με τις μεθόδους της PCR και της ανάλυσης πρωτοδιάταξης, για την αποκάλυψη της νουκλεοτιδικής ακολουθίας.

3. Αφού αποκαλυφθεί η γενετική ακολουθία, ο επιθεωρητής χρησιμοποιεί το εργαλείο RFLP εισάγοντας την και κόβοντας την με τα περιοριστικά ένζυμα που περιέχονται στη βάση δεδομένων, είτε εισάγοντας ένα καινούργιο ένζυμο.
4. Στο τέλος λαμβάνει τα αποτελέσματα μέσω της ιστοσελίδας, στα οποία μπορεί να δει τις θέσεις κοπής της νουκλεοτιδικής ακολουθίας και το μέγεθος των περιοριστικών θραυσμάτων.
5. Η ίδια νουκλεοτιδική ακολουθία του προς "υποψία" είδους ψαριού από τη βάση δεδομένων Fishtrace, εισάγεται στο πρόγραμμα και "κόβεται" με τα ίδια ένζυμα περιορισμού.
6. Γίνεται σύγκριση των αποτελεσμάτων.



Εικόνα 22. Προσομοίωση μιας εφαρμογής του προγράμματος για ανάλυση RFLP, για τον εντοπισμό απάτης

Συνεπώς, γίνεται και πάλι σύγκριση των εργαστηριακών αποτελεσμάτων με τα δεδομένα της ιστοσελίδας Fishtrace. Στην περίπτωση που οι θέσεις κοπής καθώς και το μέγεθος των περιοριστικών θραυσμάτων της προς εξέταση ακολουθίας, ταιριάζουν με αυτά του προς «υποψία» είδους ψαριού (το οποίο βρίσκεται στη

γενετική βάση δεδομένων Fishtrace), τότε μάλλον το προς εξέταση δείγμα ανήκει στο παραπάνω είδος.

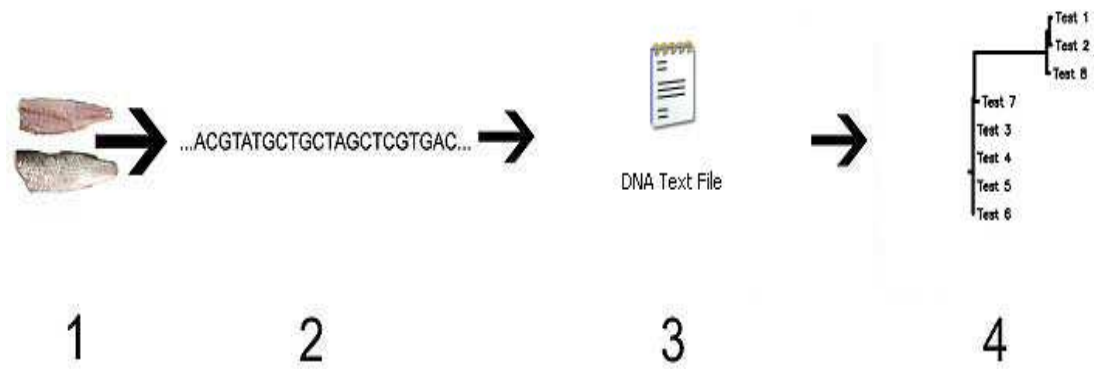
4.4. ΕΦΑΡΜΟΓΕΣ ΤΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ PHYLIP

Όπως τα προγράμματα BLAST και RFLP, έτσι και το πρόγραμμα PHYLIP βρίσκει εφαρμογή στον εντοπισμό της απάτης στο εμπόριο ψαριών και στην παράνομη αλιεία. Έτσι χρησιμοποιείται για την ταυτοποίηση του είδους του ψαριού που πρόκειται να ελεγχθεί.

Παράδειγμα: Προσομοίωση μιας εφαρμογής για τον εντοπισμό απάτης με τη χρήση του προγράμματος PHYLIP (Εικόνα 23):

1. Ο επιθεωρητής λαμβάνει δείγμα από το φρέσκο ψάρι ή από το φιλέτο ψαριού
2. Ο ιστός που λήφθηκε, μεταφέρεται στο εργαστήριο και ακολουθεί εξαγωγή DNA. Στη συνέχεια γίνεται η επεξεργασία του με τις μεθόδους της PCR και της ανάλυσης πρωτοδιάταξης, για την αποκάλυψη της νουκλεοτιδικής ακολουθίας.
3. Αφού αποκαλυφθεί η νουκλεοτιδική ακολουθία, ο επιθεωρητής χρησιμοποιεί το εργαλείο PHYLIP εισάγοντας την ακολουθία αυτή καθώς και τις νουκλεοτιδικές ακολουθίες ψαριών που έχουν παρόμοιο φαινότυπο με το προς εξέταση δείγμα. Τις ακολουθίες αυτές ο ερευνητής τις παίρνει από τη βάση δεδομένων Fishtrace.
4. Στο τέλος λαμβάνει τα αποτελέσματα μέσω μιας εικόνας «bitmap» στην οποία μπορεί να δει τη σχέση της νουκλεοτιδικής ακολουθίας του αναζητούμενου ψαριού με τις νουκλεοτιδικές ακολουθίες των υπολοίπων ψαριών, ως ένα φυλογενετικό δέντρο.

Έτσι, γίνεται σύγκριση των εργαστηριακών αποτελεσμάτων με τα δεδομένα της ιστοσελίδας Fishtrace. Στην περίπτωση που η νουκλεοτιδική ακολουθία από το δείγμα του ερευνητή ομαδοποιείται μαζί με κάποιες από τις νουκλεοτιδικές ακολουθίες ψαριών με παρόμοιο φαινότυπο (οι οποίες βρίσκονται στη γενετική βάση δεδομένων Fishtrace), τότε μάλλον το προς εξέταση δείγμα ανήκει σε αυτό το είδος ψαριού.



Εικόνα 23. Προσομοίωση μιας εφαρμογής του προγράμματος PHYLIP για τον εντοπισμό απάτης

5. ΠΕΡΙΛΗΨΗ

Η βάση δεδομένων της FishTrace σχεδιάστηκε σε συνεργασία με πολλά ευρωπαϊκά εργαστήρια, έτσι ώστε όλες οι πληροφορίες από τις γενετικές αναλύσεις να μπορέσουν να μεταφερθούν σε μια βάση δεδομένων της οποίας η πρόσβαση να πραγματοποιείται μέσω του διαδικτύου. Βασικός στόχος της ιστοσελίδας της FishTrace είναι να εμπλουτιστεί από δεδομένα που αφορούν σύγχρονες μεθοδολογίες γενετικής, καθώς επίσης και από εργαλεία που θα χρησιμοποιούν τα δεδομένα της βάσης για εφαρμογές αυτών στη διάκριση και διαχείριση των ιχθυοπληθυσμών. Τα στατιστικά αυτά εργαλεία είναι τα ακόλουθα: BLAST, RFLPs, PHYLIP. Τα προγράμματα αυτά μπορούν να χρησιμοποιηθούν για τον εντοπισμό της παράνομης, μη καταγεγραμμένης και μη ρυθμισμένης αλιείας (IUU - Illegal Unreported and Unregulated), καθώς επίσης και για τον έλεγχο της υπεραλίευσης σημαντικών θαλάσσιων ιχθυοαποθεμάτων τα οποία απειλούνται.

Το BLAST (**B**asic **L**ocal **A**lignment **S**earch **T**ool) είναι ένα λογισμικό πρόγραμμα που βρίσκει περιοχές ομοιότητας μεταξύ των γενετικών αλληλουχιών. Συγκρίνει νουκλεοτιδικές ή πρωτεϊνικές αλληλουχίες με αλληλουχίες που βρίσκονται ήδη σε μία βάση δεδομένων, και υπολογίζει τη στατιστική σημαντικότητα των αποτελεσμάτων. Το πρόγραμμα BLAST μπορεί να χρησιμοποιηθεί για την εκτίμηση της συγγένειας και της εξελικτικής σχέσης μεταξύ διαφορετικών αλληλουχιών.

Το πρόγραμμα για ανάλυση RFLP, είναι εργαλείο για τη μέθοδο του Πολυμορφισμού Μήκους Περιοριστικών Θραυσμάτων. Το πρόγραμμα αυτό χρησιμοποιεί μία λίστα περιοριστικών ενζύμων και τις αλληλουχίες του DNA από τα είδη που υπάρχουν στη Fishtrace, για να εφαρμόσει «εικονικά» την πέψη και τον διαχωρισμό θραυσμάτων DNA σε πηκτή αγαρόζης. Τα αποτελέσματα της «πέψης» δίνονται στο τέλος σε γράφημα.

Το λογισμικό πρόγραμμα PHYLIP, είναι ένα φυλογενετικό πακέτο ανάλυσης. Οι εφαρμογές που χρειάστηκαν για τη βάση δεδομένων Fishtrace, είναι αυτές που σχεδιάζουν φυλογενετικά δέντρα (Drawgram και drawtree). Αυτές οι εφαρμογές σχεδιάζουν ένα δέντρο από το εισαγόμενο αρχείο σε format τύπου PHYLIP.

6. SUMMARY

FishTrace database has been created with the collaboration of many European partners. The plan was to transfer all the genetic information in a database which can be accessed from a website. Basic task of the Fishtrace website is to fill the database with the latest genetic methodologies, as well as with tools that can use this data for discrimination and management of fish populations. BLAST, RFLPs and PHYLIP are the statistic tools. These programs can be used for the detection of Illegal Unreported and Unregulated fishing (IUU), as well as for the control of fishing of important sea stocks which are threatened.

BLAST (**B**asic **L**ocal **A**lignment **S**earch **T**ool) is a heuristic program which finds domains or shorter stretches of sequence similarity. It compares nucleotide or protein sequences with the sequences that are already included in a database, and calculates the statistical significance of the results. BLAST can be used for the relation between several sequences estimation.

RFLP is a tool which uses the Restriction Fragment Length Polymorphism method. This program uses a list of restriction enzymes and the DNA sequences from the Fishtrace database, to apply virtually the procedure and the migration of the fragments in agarose gel. The results are given on a graph in a new page.

PHYLIP is a comprehensive phylogenetic analysis package. There is a list of the programs that can be used for the molecular sequence data analysis. The programs that are used in our work are the tree drawing programs (Drawgram και drawtree). These programs draw a tree from the specifications (the input file with the sequence in Phylip format).

7. ΒΙΒΛΙΟΓΡΑΦΙΑ

7.1. ΕΛΛΗΝΙΚΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

- ✚ Ιμσιρίδου, Α. (2003). Γενετική Μηχανική και Βιοτεχνολογία. ΑΤΕΙ Θεσ/κης - Παράρτημα Ν. Μουδανιών.

7.2. ΞΕΝΟΓΛΩΣΣΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

- ✚ Apostolidis, A. P., Karakousis, Y. and Triantaphyllidis, C. (1996). Genetic differentiation and phylogenetic relationships among Greek *Salmo trutta* L. (brown trout) populations as revealed by RFLP analysis of PCR amplified mitochondrial DNA segments. *Heredity* **77**, 608-618
- ✚ Cespedes, A., Garcia, T., Carrera, E., Gonzalez, I., Fernandez, A., Asensio, L., Hernandez, P. and Martin, R. (2000). Genetic differentiation between sole (*Solea solea*) and Greenland halibut (*Reinhardtius hippoglossoides*) by PCR-RFLP analysis of a 12S rRNA gene fragment. *J. Sci. Food Agric.* **80**, 29-32.
- ✚ Comesana, A. S. and Abella, S. (2003). Molecular identification of five commercial flatfish species by PCR-RFLP analysis of a 12S rRNA gene fragment. *J. Sci. Food Agric.* **83**, 752-759.
- ✚ Cronin, M. A., Spearman, W. J., Wilmot, R. L., Patton, J. C. and Bickham, J. W. (1993). Mitochondrial DNA variation in chinook (*Oncorhynchus tshawytscha*) and chum salmon (*O. keta*) detected by restriction enzyme analysis of polymerase chain reaction (PCR) products. *Can. J Fish. Aquat. Sci.* **50**, 708-715.
- ✚ Felsenstein, J. (1989) *PHYLIP, Phylogenetic Inference Package, Version 3.2*. University of Washington, Seattle, WA.
- ✚ Hansen, M. M. and Loeschcke, V. (1996). Genetic differentiation among Danish brown trout populations, as detected by RFLP analysis of PCR amplified mitochondrial DNA segments. *J. Fish Biol.* **48**, 422-436.
- ✚ Imsiridou, A., Apostolidis, A., Durand, J. D., Briolay, J., Bouvet, Y. and Triantaphyllidis, C. (1998). Genetic differentiation and phylogenetic relationships among Greek chub *Leuciscus cephalus* L. (Pisces, Cyprinidae)

populations as revealed by RFLP analysis of mitochondrial DNA. *Biochem. Syst. Ecol.* **26**, 415-429.

- ✚ Imsiridou, A., Hardy, H., Maudling, N., Amoutzias, G. and Zaldivar, J. M. (2003). Web Database of Molecular Genetic Data from Fish Stocks. *J. Heredity.* **94**:265-267.
- ✚ Tuimala, J. (2004). *A Primer to Phylogenetic Analysis using Phylip Package*. 2nd edition. Center for Scientific Computing, Espoo, Finland.

7.3. ΙΣΤΟΣΕΛΙΔΕΣ ΣΤΟ ΔΙΑΔΙΚΤΥΟ

- ✚ Wikipedia, the free encyclopedia, article for the “DNA”
<http://en.wikipedia.org/wiki/DNA>
- ✚ The NCBI: The Statistics of Sequence Similarity Scores
<http://www.ncbi.nlm.nih.gov/BLAST/tutorial/Altschul-1.html>
- ✚ Wikipedia, the free encyclopedia, article for the “haplotype”
<http://en.wikipedia.org/wiki/Haplotype>
- ✚ The NCBI handbook, article for “How BLAST Works: The Basics”, article for “BLAST Scores and Statistics” and article for “BLAST output”
<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=handbook.chapter.610>
- ✚ The NCBI BLAST main page, sector for program choice
<http://www.ncbi.nlm.nih.gov/BLAST>
- ✚ Restriction Fragment Length Polymorphism
<http://en.wikipedia.org/wiki/RFLP>
- ✚ Restriction Fragment Length Polymorphism (RFLP) Tool
<http://vm.cfsan.fda.gov/~frf/rflp.html>
- ✚ Fishtrace BLAST, RFLP and PHYLIP on line tools <http://www.fishtrace.org/>
- ✚ PHYLIP Frequently Asked Questions
<http://evolution.genetics.washington.edu/phylip/faq.html>
- ✚ KARLIN-ALTSCHUL STATISTICS
<http://www-bimas.cit.nih.gov/blastinfo/KAstat.html>