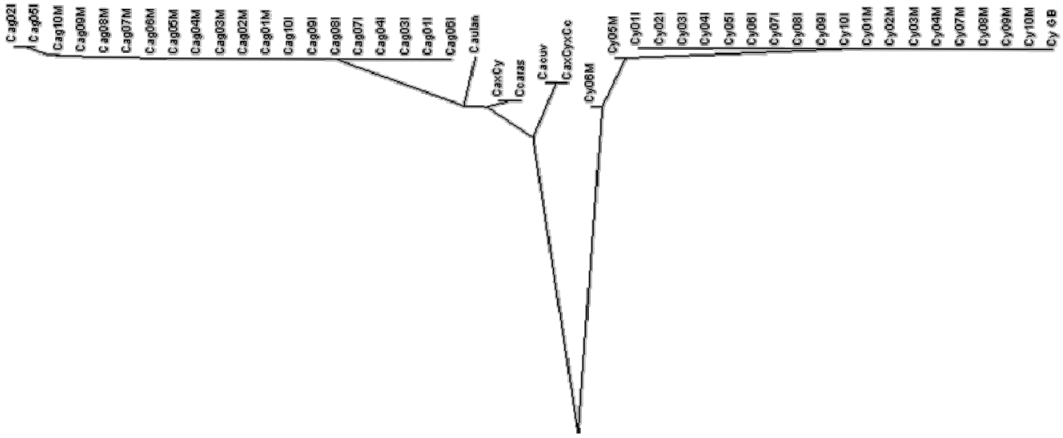
Dragoş Lucian Gorgan

INTRODUCERE ÎN STUDIUL FILOGENIEI ŞI FILOGEOGRAFIEI MOLECULARE



BIOFLUX
CLUJ-NAPOCA, 2008

Autor: Dragoş Lucian Gorgan

Referenți ştiinţifici:
Prof. univ. dr. Vlad Artenie
Cercetător dr. Ioan Valentin Petrescu-Mag

Director editură:
Dr. Ioan Valentin Petrescu-Mag

Consilier editorial:
Dr. Ruxandra Mălina Petrescu-Mag

Editura: Bioflux, Cluj-Napoca, 2008
ISBN 978-973-88929-9-6

CUPRINS

INTRODUCERE.....4
1. DEFINIȚIA ŞI CARACTERISTICILE POPULAȚIEI.....6
1.1. Metode utilizate în studiul genetic al populațiilor.....13
2. ASPECTE PRIVIND FILIAȚIA ŞI DIVERSITATEA MOLECULARĂ.....15
2.1. Importanța biologiei molecular în indicarea direcțiilor evolutive……...15
2.2. Rolul ADN mitocondrial (ADNmt)…….....16
2.3. Aspecte ale evoluției moleculare……...16
2.3.1.Evoluția secvențelor nucleotidice şi proteice……..16
2.3.2.Metode de deducere a direcțiilor de filiație, pornind de la date moleculare……17
2.3.3.Evoluția secvențelor de aminoacizi……..18
2.3.4.Mutațiile şi rata de substituție a aminoacizilor……22
2.4. Tipuri de modificări la nivelul macromoleculei de ADN……24
2.4.1.Diferențe nucleotidice între secvențe……..25
2.4.2.Modele de evoluție ale secvențelor nucleotidice……26
2.4.3.Estimarea numărului de substituții nucleotidice……30
2.4.3.1. Modelul Jukes-Cantor……..30
2.4.3.2. Modelul Kimura cu 2 parametri……..31
2.4.3.3. Modelul Tajima-Nei……..31
2.4.3.4. Modelul Tamura-Nei……32
2.4.3.5. Madelul Tamura……32
2.4.3.6. Distanțe Gamma……...34
2.4.4.Estimarea numerică a distanțelor evolutive……34
2.5. Arbori filogenetici……35
2.5.1.Arbori ai genelor şi arbori ai speciilor……...36
2.5.2.Metode de reconstruire a arborilor filogenetici……...38
2.5.2.1. Metode bazate pe distanțe……..38
2.5.2.2. Metoda evoluției minime (ME)……41
2.5.2.3. Metoda Neighbor-Joining (NJ)……41
2.5.2.4. Alegerea tipului de distanță evolutivă folosită în reconstrucțiile filogenetice……46
2.5.2.5. Metoda Maximum Parsimony (MP)……...47
2.5.2.6. Varianța şi covarianța bootstrap (replicărilor repetate)……..48
2.6. Situsuri informative şi homoplazia……49
2.7. Strategii de căutare a filogeniei MP……51
2.8. Metode ML (Maximum Likelihood)……52
3. METODOLOGIA DE OBȚINERE A SECVENȚELOR NUCLEOTIDICE……55
3.1. Determinare unor parametri şi indici fenotipici……...55
3.2. Protocolul de extracție al ADN total din țesut muscular……57
3.2.1.Prelucrarea statistică a datelor obținute din extracția ADN……..59
3.2.1.1. Estimarea valorilor medii ale populațiilor şi stabilirea intervalelor de variabilitate……...59

3.2.1.2. Compararea mai multor probe……60
3.3. Reacția de amplificare genică prin polimerizare în lanț (PCR)……61
3.3.1.Principii generale……61
3.3.2.Ciclurile de replicare……...63
3.3.2.1. Denaturarea……..63
3.3.2.2. Alinierea primerilor……...64
3.3.2.3. Extensia……64
3.3.2.4. Condițiile de reacție……65
3.3.3.Componentele PCR……65
3.3.4.Amplificarea genică la reprezentanți ai familiei *Cyprinidae*……..65
3.3.4.1. Amplificarea genei care determină sinteza citocromului B…66
3.3.4.2. Amplificarea regiunii de control mitocondrial……68
3.3.4.3. Amplificarea genei care codifică subunitatea 4L a dehidrogenazei (ND4L)……69
3.4. Testarea produşilor PCR prin electroforeză……70
3.5. Purificarea produşilor PCR……71
3.6. Cuantificarea produşilor PCR……72
3.6.1.Prepararea gelului de cuantificare……72
3.6.2.Pregătirea probelor pentru cuantificare……...72
3.6.3.Cuantificarea probelor……72
3.7. Reacția de secvențiere……73
3.7.1.Reacția de secvențiere a citocromului B……74
3.7.2.Reacția de secvențiere a regiunii de control mitocondrial……75
3.7.3.Reacția de secvențiere a ND4L……75
3.8. Precipitarea in etanol a probelor pentru secvențiere……75
3.9. Analiza secvențelor……76
4. Studiul relațiilor filogenetice……77
4.1. Evaluarea uniformității populațiilor……77
4.2. Testarea obținerii produşilor PCR prin electroforeză orizontală în gel de agaroză……82
4.3. Cuantificarea produşilor PCR……87
4.4. Secvențierea genelor utilizate în studiile de filogenie……95
4.5. Compararea secvențelor nucleotidice ale citocromului B provenite de la indivizi aparținând genului *Cyprinus*……96
4.5.1.Caracterizarea haplotipului CY01I……99
4.5.1.1. Frecvențele tipurilor de asocieri nucleotidice……..99
4.5.1.2. Translația şi caracterizarea catenei polipeptidice……...99
4.5.2.Caracterizarea haplotipului CY04I……102
4.5.2.1. Frecvențele tipurilor de asocieri nucleotidice……102
4.5.2.2. Translația şi caracterizarea catenei polipeptidice……102
4.5.3.Caracterizarea haplotipului CY0104N……...103
4.5.3.1. Frecvențele tipurilor de asocieri nucleotidice……103
4.5.3.2. Translația şi caracterizarea catenei polipeptidice……104
4.5.4.Caracterizarea haplotipului CY03M……...105
4.5.4.1. Frecvențele tipurilor de asocieri nucleotidice……105
4.5.4.2. Translația şi caracterizarea catenei polipeptidice……106
4.5.5.Caracterizarea haplotipului CY06M……...107
4.5.5.1. Frecvențele tipurilor de asocieri nucleotidice……107
4.5.5.2. Translația şi caracterizarea catenei polipeptidice……107

4.5.6.Rela□ii filogenetice în cadrul genului *Cyprinus* stabilite pe baza secvențelor citocromului B……..110
4.6. Compararea secvențelor nucleotidice ale regiunii de control mitocondrial provenite de la indivizi aparținând genului *Cyprinus*……113
4.6.1.Caracterizarea haplotipului general al celor două populații……...115
4.6.1.1. Frecvențele tipurilor de asocieri nucleotidice……115
4.6.2.Caracterizarea haplotipului CY01N……116
4.6.2.1. Frecvențele tipurilor de asocieri nucleotidice……116
4.6.2.2. Modelarea structurii secundare a ARNm……116
4.6.3.Caracterizarea haplotipului CY05M……...117
4.6.3.1. Frecvențele tipurilor de asocieri nucleotidice……117
4.6.4.Caracterizarea haplotipului CY06MD……117
4.6.4.1. Frecvențele tipurilor de asocieri nucleotidice……117
4.6.5.Relații filogenetice în cadrul genului *Cyprinus* bazate pe diferențe ale secvențelor D-LOOP……118
4.7. Analiza secvențelor nucleotidice ale ND4L dehidrogenazei provenite de la indivizi apar□inând genului Cyprinus……121
4.7.1.Caracterizarea haplotipului CY01IM……..124
4.7.1.1. Frecvențele tipurilor de asocieri nucleotidice……124
4.7.1.2. Translația şi caracterizarea catenei polipeptidice……124
4.7.2.Relații filogenetice în cadrul genului Cyprinus bazate pe diferențe ale secvențelor ND4L……127
4.8. Analiza secvențelor nucleotidice ale citocromului B provenite de la indivizi aparținând genului *Carassius*……..129
4.8.1.Caracterizarea haplotipului CAGIMY……133
4.8.1.1. Frecvențele tipurilor de asocieri nucleotidice……133
4.8.1.2. Translația şi caracterizarea catenei polipeptidice……134
4.8.2.Caracterizarea haplotipului CAG137IM……137
4.8.2.1. Frecvențele tipurilor de asocieri nucleotidice……137
4.8.2.2. Modelarea structurii secundare a ARNm……137
4.8.2.3. Translația şi caracterizarea catenei polipeptidice……137
4.8.3.Relații filogenetice în cadrul genului *Carassius* stabilite prin analiza secvențelor citocromului B……..138
4.9. Analiza secvențelor nucleotidice ale d-loop, provenite de la indivizi aparținând genului *Carassius*……..142
4.9.1.Caracterizarea haplotipului CAGIMD……145
4.9.1.1. Frecvențele tipurilor de asocieri nucleotidice……145
4.9.2.Caracterizarea haplotipului CAG25ID……146
4.9.2.1. Frecvențele tipurilor de asocieri nucleotidice……146
4.9.3.Relații filogenetice în cadrul genului *Carassius*, bazate pe diferențe ale secvențelor d-loop……146
4.10. Compararea secvențelor nucleotidice ale ND4L dehidrogenazei, provenite de la indivizi aparținând genului *Carassius*……..150
4.10.1. Caracterizarea haplotipului CAGIMN……153
4.10.1.1. Frecvențele tipurilor de asocieri nucleotidice……153
4.10.1.2. Translația şi caracterizarea catenei polipeptidice……153
4.10.2. Relații filogenetice în cadrul genului *Carassius* obținute prin analiza secvențelor genei ND4L……..155
4.11. Direcții evolutive în cadrul Familiei *Cyprinidae*……158
BIBLIOGRAFIE……..169

INTRODUCERE

Bazele geneticii populațiilor, puse în perioada 1920-1930, au permis descoperirea faptului că procesele evolutive afectează şi nivelul infraspecific populațional. Studiul genetic al populațiilor se realizează optim cu ajutorul matematicii, prin intermediul căreia putem calcula frecvența genelor în succesiunea de generații şi putem cuantifica factorii care contribuie la modificarea acestei frecvențe.

Încă Ch. Darvin a considerat evoluția ca proces statistic, pornind de la observația ca variațiile ereditare au caracter întâmplător, iar selecția naturală, realizată prin supraviețuirea şi reproducerea diferențiată a indivizilor mai apți, reprezintă factorul major al procesului.

F. Galton (1822-1911), a fost primul care a folosit metode statistico-matematice în studiul populațiilor. În lucrările sale „Typical laws of heredity in man" (1877) şi „Natural inheritance" (1889) redă considerații asupra fenomenului eredității, pornind de la cunoaşterea corelațiilor şi regresiilor existente la părinți şi descendenți.

K. Pearson (1857-1936) a continuat cercetările lui F. Galton şi a elaborat metodele de studiu a populațiilor mari, a dezvoltat metoda corelațiilor şi a elaborat testul $\chi 2$. Astfel a apărut biometria, disciplină care are ca obiect de studiu variabilitatea organismelor, interpretată cu ajutorul metodelor matematice.

În 1900, când au fost redescoperite legile lui Mnedel, s-a constatat, unanim, că acestea sunt în esență legi probabilistice, fenotipizarea diferitelor caractere la descendenți având un caracter aleatoriu dar determinist.

Matematicianul G.H. Hardy şi medicul W. Weinberg, au descoperit că, în cazul uni grup de indivizi izolați reproductiv, între care se realizează panmixia, frecvența genelor, respectiv a genotipurilor, se menține constantă atât timp cât nu intervin factori externi care pot modifica acest echilibru. Astfel a fost elaborată cunoscuta lege Hardy-Weinberg.

În deceniul trei al secolului trecut, iluştrii cercetători H. Nilson-Ehle, E.M. East şi R.A. Emerson au demonstrat că şi caracterele cantitative au o determinare poligenică şi se transmit mendelian. Astfel a luat naştere *Genetica cantitativă*, disciplină în cadrul ăcreia se studiază ereditatea caracterelor cantitative ale unei populații prin metode statistice.

R.A. Fisher prin lucrarea sa „Teoria genetică a selecției naturale" (1930), S. Wright prin lucrarea „Evoluția în populațiile mendeliene" (1931) şi H.B.S. Haldane prin lucrarea „Teoria matematică a selecției naturale şi artificiale" (1932) aduc o contribuție importantă la fundamentarea geneticii populațiilor.

Trebuie menționat că genetica populațiilor este strâns legată de genetica cantitativă, deoarece frecvența genelor, factorii care modifică această frecvență cât şi ereditatea caracterelor cantitive se pot studia numai la nivelul populațional.

Apariția geneticii populațiilor prezintă o mare importanță în ameliorarea plantelor şi animalelor. În prezent cu ajutorul geneticii populațiilor s-a putut da o explicație ştiințifică dinamicii speciilor, fenomen care se petrece nu la nivel individual ci la cel populațional. Tot cu ajutorul acestei discipline s-au evidențiat factorii care modifică frecvența genelor în populațiile panmictice, rorlul mecanismelor de izolare reproductivă în speciație, importanța mărimii populațiilor în procesul hazardului reprezentat de deriva genetică a populațiilor mici.

CAPITOLUL 1

DEFINIȚIA ŞI CARACTERISTICILE POPULAȚIEI

Populația reprezintă un nivel de organizare şi integrare a lumii vii care se caracterizează prin anumite trăsături specifice şi este subordonat acțiunii unor legități caracteristice. Pentru a specifica şi analiza aceste caracteristici trebuie de precizat că populația poate fi abordată atât ca unitate structurală şi funcțională în ierarhia organizării şi integrării sistemice a lumii vii, cât şi ca entitate strict taxonomică, în cadrul unităților sistematice infraspecifice.

Populațiile se caracterizează din punct de vedere al ratei naşterilor, ratei mortalității, ratei difuziei ş.a. Toți aceşti parametri reprezintă valori medii derivate din comportamentul unui anumit grup de indivizi. Alți parametri ca de exemplu supraviețuirea, reproducerea sunt caracteristici individuale. În mod cu totul convențional, în genetica populațiilor, se utilizează o simplificare extremă, luându-se în calcul caracteristicile individului la un moment dat, pentru delimitări arbitrare între populații. Dacă caracterele luate în considerare sunt determinate genetic, se pot defini cu o oarecare exactitate următorii termeni:

a) populația reprezintă o unitate interîncrucişabilă, din cadrul speciei. În cazul în care în populație sunt indivizi de două sexe şi fiecare individ are şanse egale de a se întâlni cu oricare alt individ de sex opus, populația este considerată *panmictică*;

b) migrațiile reprezintă trecerea indivizilor, de la o populație la alta. În loc de migrație, se mai foloseşte termenul de flux genic;

c) frecvența genelor într-o populație care reprezintă frecvența unui tip de alelă, la un locus dat, raportată la frecvența tuturor alelelor la respectivul locus.

În ceea ce priveşte *gena,* trebuie menționat faptul că aceasta a fost caracterizată prin investigațiile efectuate la trei niveluri: în pedigree sau în experimentele de hibridare, la cel al coziției chimice şi al implicațiilor în producerea fenotipului. *Genele majore*, care determină segregarea caracterelor, sunt diferite de genele asociate (funcțional) în poligene, care controlează caracterele cu variabilitate continuă.

Luând în considerație numărul de specii, diversitatea de forme şi grade de evoluție, se poate afirma că peştii formează „o lume aparte” (Taisescu, 1979), în lumea vertebratelor. Clasificarea lor complexă în 32 ordine, 80 subordine şi peste 400 de familii (Greenwood et al., 1996, citat de Taisescu, 1979) este o expresie elocventă a marii diversități care caracterizează clasa *Pisces*.

Chiar dacă nu se ia în considerare numărul mare de indivizi care aparțin unei specii, clasa peştilor numără mai multe specii decât toate celelalte clase de vertebrate la un loc. Astfel, numai speciile de teleosteeni sunt estimate la peste 20.000, ceea ce înseamnă aproximativ 60% din totalul vertebratelor actuale (la care se adaugă amfibienii, aproximativ 2.500 specii, reptilele 6.000 specii, păsări 8.600 specii şi mamifere 4.500 specii).

Pe lângă avantajul unei imense variabilități pe care o oferă numărul mare de specii, peştii sunt înzestrați cu cel mai bun material (cel puțin dintre vertebrate), pentru studii citogenetice. Deoarece prezintă avantajul încrucişării artificiale intra- şi interspecifice, la care se adaugă şi faptul că sunt foarte uşor de întreținut, peştii constituie pentru cercetările de genetică un material comparabil şi din unele puncte de vedere chiar superior *Drosophilei*.

Având în vedere că diferitele speciile de peşti prezintă anumite particularități care le diferențiază sub aspectul ritmului de creştere, rezistenței la boli, modului de valorificare a hranei etc. este necesară obținerea unor forme cu caracteristici bioproductive superioare acestora, care să îmbine caracterele lor pozitive.

Ciprinidele reprezintă una dintre cele mai răspândite familii de peşti, cu mai mult de 2.000 specii grupate în aproximativ 340 genuri (Bănărescu şi Coad, 1991). Peştii din această familie au atras atenția evoluționiştilor, datorită arealelor mari pe care le prezintă pe toate continentele şi prezenței lor în majoritatea ecosistemelor de apă dulce. Cu toate acestea, de când Cuvier (1817) a stabilit această familie, relațiile sistematice din cadrul său, au fost permanent discutate. Primele clasificări ale ciprinidelor se bazau în principal pe caractere externe (ex. prezența, tipul şi numărul mustăților), precum şi structura şi aranjarea dinților faringieni (Howes, 1991). Recent, caracterele osteologice au fost de asemenea folosite în determinarea relațiilor filogenetice în cadrul diferitelor grupe de ciprinide (ex. genul *Barbus*; Doadrio 1990). Dar, liniile monofiletice şi relațiile de înrudire dintre subfamiliile familiei *Cyprinidae* (ex. *Cyprininae*, *Leuciscinae* şi *Rasborinae*) trebuiesc stabilite cu exactitate (Howes, 1991).

Uşurința cu care pot fi manevrați gameții, fecundația externă şi posibilitatea de control a dezvoltării zigotului după fertilizare, fac ca tehnici de o deosebită importanță practică cum ar fi manipularea cromosomilor sau a seturilor de cromosomi, frecvent folosite în genetica plantelor, să se poată aplica la peştii de cultură în vederea creşterii productivității.

Ginogeneza reprezintă o formă rară de înmulțire sexuală. În cazul ginogenezei, fecundarea este necesară (prin fecundare înțelegând pătrunderea spermatozoidului în oosferă), dar cromosomii spermatozoidului nu se transmit descendenței.

De aceea, dezvoltarea descendențelor ginogenetice decurge numai pe baza eredității materne. În funcție de particularitățile citologice, descendențele ginogenetice fie că repetă în întregime forma maternă (ginogeneza ameiotică), fie că, datorită desbinării, se deosebesc de forma maternă prin anumite criterii (ginogeneza meiotică).

Ginogeneza este cunoscută în natură şi poate fi provocată experimental. Ginogeneza naturală (de regulă ameiotică) este asigurată de către nişte mecanisme biologice, care realizează inactivarea cromosomilor spermatozoidului care a pătruns în oosferă şi elimină reducerea cromosomilor femeli în cadrul procesului de maturare a oosferei. Datorită acestui fapt, icrele femelelor care se înmulțesc prin ginogeneză conțin un set complet de cromosomi materni şi unul redus la jumătate (aşa cum se întâmplă în cazul înmulțirii sexuale obişnuite).

În cazul ginogenezei induse, ambele particularități remarcate mai sus, adică inactivarea genetică a cromosomilor masculi şi eliminarea reducerii cromosomilor femeli, sunt asigurate prin intervenții experimentale (în cazuri rare diploidizarea cromosomilor femeli apare şi în mod spontan).

Prioritatea în elaborarea metodelor de obținere a ginogenezei induse la peşti aparține cercetătorilor sovietici D. D. Romaşev, K. A. Colovinscaia şi V. N. Beliaeva, care în anii 1950 – 1960 au demonstrat experimental posibilitatea ginogenezei

diploide induse la crap şi țipar şi au obținut primele date referitoare la proprietățile descendențelor ginogenetice. În prezent, ginogeneza meiotică indusă este realizată la multe specii de peşti, dar procentul cel mai mare îl constituie cea indusă la specii aparținând Familiei *Cyprinidae*.

Din punct de vedere teoretic, extinderea studiilor privind în principal, ginogeneza şi hibridarea, consangvinizarea şi heterozisul, a permis înțelegerea mai profundă a eredității organismelor, a modalităților de transmitere a caracterelor ereditare de-a lungul generațiilor. S-a putut da astfel explicația ştiințifică a importanței hibridării combinate cu sporirea numărului de cromosomi, în procesul natural de apariție a noilor specii.

De asemenea, hibridarea a dus la descoperirea celor mai eficiente modalități de modificare a eredității în sensul dorit de om. Astfel, în ultimii ani s-a acordat o atenție deosebită studiului mecanismului genetic al heterozisului şi s-au elaborat metode noi de obținere a organismelor ce manifestă heterozis pe bază de linii consangvinizate.

Ameliorarea peştilor, înțeleasă ca proces de modificare a potențialului productiv, a caracterelor ereditare, a genofondului populațiilor în direcția dorită de om, este o preocupare mai recentă. Selecția şi creşterea formelor de peşti ameliorate în scopul producerii de puiet şi a loturilor de reproducători cu calități superioare, reprezintă una dintre problemele majore ale cercetării şi producției piscicole.

Baza ereditară a individului, totalitatea genelor diseminate în nucleele tuturor celulelor din corpul său şi care îi controlează morfogeneza şi funcționarea organismului pe tot parcursul vieții este, deocamdată, imposibil de a fi dirijată după formarea zigotului. Singura soluție practică de a crea o bază ereditară dorită este aceea de a interveni în procesul formării zigotului, al reproducerii; o modificare în acest moment se referă însă la populație, nu la individ, se modifică generația următoare şi nu cea actuală. Cu fiecare generație nouă sunt posibile schimbări genetice şi este bine ca acestea să fie dirijate şi nu lăsate în voia hazardului.

În ultimele decenii cercetarea legată de creşterea animalelor a fost confruntată cu o multitudine de solicitări exprese impuse de necesitatea adaptării tehnologiilor de creştere şi ameliorare la nevoia de a produce mai mult, mai bun şi mai ieftin (Granciu et al., 1973). Ritmul dezvoltării şi industrializării creşterii animalelor reclamă introducerea în practică a unor metode moderne de investigație a potențialului productiv, reproductiv şi a stării de sănătate a efectivelor existente şi ameliorarea acestora într-un timp scurt.

Selecția practicată după schema clasică începe să fie depăşită atât sub raportul mijloacelor de investigație folosite, cât şi în raport cu factorul timp. Metodele clasice de apreciere a valorii de ameliorare a unui reproducător necesită timp îndelungat, de la naştere şi până când intră în producție primii săi descendenți. Practica ameliorării cere o metodă rapidă de selecție, aşadar cât mai economică, pentru a permite o preselecție a animalelor la o vârstă tânără, urmând ca ulterior să fie folosite şi alte criterii de selecție.

Dezvoltarea metodelor de investigație biochimică şi progresele realizate de genetică au determinat orientarea cercetării ştiințifice în direcția descifrării intimităților procesului metabolic, în special al sistemelor biochimice care îl compun (Cazacu, 1971). Acumularea datelor din acest domeniu ştiințific a condus totodată şi la naşterea uneia dintre cele mai noi discipline biologice – genetica biochimică – disciplină de graniță care înmănunchează tehnici de investigație şi concepte operaționale din domeniul biochimiei, geneticii, biofizicii, fiziologiei.

Progresele înregistrate au contribuit la conturarea unor noi concepții şi puncte de vedere în abordarea şi investigarea proceselor biologice. În acest context,

problematica majoră a biochimiei se circumscrie unor aspecte fundamentale cum sunt studiul biomoleculelor pe de o parte şi studiul biochimic al sistemelor biologice. Obiectivele esențiale care derivă din aceste aspecte se referă la organizarea biochimică a organismelor vii sub aspectul structurii, funcției şi proprietății biomoleculelor, la natura şi mecanismele reacțiilor biocatalitice asociate structurii celulare şi infracelulare, la procesele biochimice ale metabolismului, la reglarea şi controlul biochimic al activității celulare, la bazele moleculare ale proceselor biologice moleculare şi patologice (Dumitru, 1980; Tămaş et al., 1981).

În ultimele decenii, cercetările s-au orientat tot mai pregnant în direcția studiului amănunțit al metabolismului, în special al sistemelor biochimice care stau la baza lui, pornindu-se de la ideea că nivelul la care se desfăşoară procesele biochimice, influențează în mod direct capacitatea productivă, reproductivă, de adaptare la mediu sau starea de sănătate a animalelor (Cazacu, 1971). Desigur, este puțin probabil că se va putea izola o anumită genă care determină un proces fiziologic complex. Este însă posibil ca printr-o serie de metode biochimice, genetice sau imunoserologice să se identifice aşa-numitele „gene markeri" şi să se urmărească evoluția lor în cursul selecției şi de asemenea să se stabilească în ce măsură sunt sau nu corelate cu unele caractere cantitative şi calitative importante, din punct de vedere economic (Kirpicinicov, 1987; Paaver, 1983).

Caracteristicile biochimice ale animalelor de fermă sunt folosite pentru studierea şi aprofundarea unui număr de probleme teoretice ca şi pentru rezolvarea unor aspecte practice legate de creşterea, ameliorarea, adaptarea şi sănătatea animalelor. Privite din acest punct de vedere, sistemele biochimice reprezintă un exemplu elocvent cum un domeniu considerat ca o cercetare teoretică fundamentală exclusivă a putut să se implice în ameliorarea şi selecția peştilor şi să-şi dovedească aplicabilitatea.

Pentru genetica biochimică şi implicit practica ameliorării şi selecției peştilor sistemele biochimice prezintă importanță prin polimorfismul lor, adică prin existența mai multor forme distincte ale fiecărui sistem în parte (Bănărăscu, 1975, Maximilian, 1984). Polimorfismul se manifestă de la naştere şi, cu unele excepții, constând în virarea către monomorfism ca urmare a eliminării unor alele sub influența factorilor de mediu combinată cu cea a factorilor tehnologici, se menține constant tot timpul vieții. De aici rezultă şi posibilitatea folosirii lui ca un instrument de selecție timpurie a reproducătorilor de mare valoare. Avantajul de selecție conferit de un tip sau altul din cadrul unui sistem biochimic este suficient ca să atragă atenția asupra sa, în dependență cu formele sale de manifestare.

Ținând seama de faptul că peştii reprezintă clasa de vertebrate superioare aflată pe treapta cea mai de jos a evoluției acesteia, că sistemele lor de organe din punct de vedere anatomic şi fiziologic sunt încă primitive, abia anunțând evoluția din treaptă în treaptă şi înalta specializare de la mamifere, trebuie să acceptăm ideea că şi metabolismul lor şi implicit sistemele biochimice care compun acest metabolism, se caracterizează prin aceleaşi trăsături. Dacă adăugăm la acestea faptul că fecundația externă şi oviparitatea este regula, iar viviparitatea excepția, că dezvoltarea de la larvă la adult este, în general, caracteristică unor specii sau unui grup restrâns de specii, în timp ce altele nu trec prin stadiul de larvă (peştii vivipari), că unele specii de peşti reprezintă din punct de vedere evolutiv adevărate „puncte terminus" şi nu în ultimul rând numărul impresionant de specii, ne vom putea explica şi înțelege marea diversitate a situațiilor şi aparentele paradoxuri față de cunoştințele acumulate în diferitele domenii ale ihtiologiei.

Chen *et al.* (1984) au furnizat prima analiză cladistică clară a interrelațiilor din cadrul familiei *Cyprinidae*. Cavender şi Coburn (1992) au reanalizat datele furnizate

de Chen şi colaboratorii în 1984 şi au prezentat cea mai completă analiză filogenetică a familiei *Cyprinidae* din acea perioadă. În analiza lor asupra relațiilor filogenetice asupra ciprinidelor din America de Nord, Cavender şi Coburn (1992) au concluzionat existența a două subfamilii: *Cyprininae* care conține barbinele, cipriniele şi labeoninele şi Subfamilia *Leuscinae* care include ticine, gobionine, rasborine, leuciscine, cultrine xenocyprine, acheilognathine şi phoxinine. Totuşi, în ciuda acestor eforturi, interpretarea datelor morfologice s-a dovedit a fi dificilă. Verigile lipsă ale caracterelor morfologice valide care definesc diferite grupe sistematice au condus la stabilirea unor direcții sistematice în cadrul Familiei *Cyprinidae* (Nelson, 1994; Fink şi Fink, 1996).

Recent, s-au introdus alozime drept markeri pentru studiul proceselor de hibridare, structura populațiilor şi relațiilor filogenetice de nivel taxonomic mai mic în cadrul ciprinidelor europene (Coelho, 1992; Berrebi et al., 1995; Coelho et al., 1995; Karakousis et al., 1995; Carmona et al., 1997; Alves et al., 1997a). Cu toate acestea, aprecierea relațiilor filogenetice la nivele taxonomice superioare necesită utilizarea de markeri moleculari diferiți. Din acest punct de vedere, secvențierea ADN în scopul analizării principalelor linii filogenetice ale ciprinidelor este pe deplin necesară (Berrebi et al. 1996).

Distribuția mare a ciprinidelor duce la apariția unor întrebări de ordin biogeografic şi evoluționist referitoare la evoluția şi radiația viitoare a acestor specii. De exemplu, ciprinidele din interiorul Europei prezintă un interes particular în ce priveşte distribuția cu numeroase specii endemice în Peninsula Iberică şi sudul Greciei şi genuri cu destul de puține specii în Centrul Europei (Bănărescu 1973). Această distribuție caracteristică a fost explicată prin izolarea Peninsulei Iberice şi sudului Greciei de restul continentului, fapt ce a condus la limitarea numărului genurilor capabile să colonizeze ambele regiuni. Oricum, scenariul exact care a condus la actuala distribuție biogeografică nu este încă stabilit. Unele dintre cele mai vechi ciprinide fosile au fost descoperite în straturile din Oligocen din Europa Centrală (Obrhelova 1971), dar este general acceptat că ciprinidele din Europa au origine asiatică (Bănărescu 1989, 1992). Dispersia ciprinidelor asiatice în Europa a fost posibilă numai în timpul Oligocenului (Bănărescu 1989), a început odată cu formarea Munților Urali (Rögl şi Steininger 1983) desfăşurându-se în paralel cu formarea acestora. Este încă neclar dacă întreaga Europă a fost colonizată în timpul migrației ce traversa Uralii în formare sau daca unele ciprinide (în special genul *Barbus*) au ajuns în Peninsula Iberică prin nordul Africii (Doadrio 1990; Zardoya şi Doadrio; 1998).

Primele încercări de evidențiere şi numărare a cromosomilor la peşti, aparțin cercetătorilor Moore (1890), Van der Strict (1895) şi Sobota (1897), (citați de Taisescu, 1979) dar cercetările privind citogenetica şi filogenia peştilor, aparțin secolului nostru şi s-au dezvoltat preponderent, după 1960.

Ce ştim despre mecanismele eredității la peşti, se bazează în principal pe cercetări efectuate pe foarte puțini membri ai ordinului *Cyprinodontiformes* – *Lebistes reticulatus*, *Xiphophorus sp.* şi specii mexicane ale genului *Poeciliopsis*. Mai nou, cunoştințele în acest domeniu s-au lărgit odată cu efectuarea de cercetări pe crap şi caras (*Cyprinidae*), păstrăv (*Salmonidae*) şi alte specii reprezentative aparținând la grupe diferite.

În ce priveşte determinismul sexelor, aproape toate tipurile de reproducere sexuată cunoscute la animale pot fi întâlnite la peşti. Oricum, mecanismul determinismului sexual la majoritatea peştilor osoşi, nu este încă elucidat. Cromosomii sexului sunt atât de puțin diferențiați față de autosomi, încât foarte rar pot fi evidențiați. Se presupune că determinismul sexelor poate fi controlat, parțial dacă nu chiar

complet, de gene dispersate în lungul autosomilor. Determinismul sexelor poate fi diferit la specii diferite, precum şi în cazul încrucişărilor, între masculi puternici şi femele debilitate, rezultatul fiind un sex-ratio neechilibrat, în favoarea masculilor.

Datele existente în legătură cu determinismul sexelor în populații sălbatice de peşti şi populații de laborator, relevă mecanisme mai ample decât în cazul altor vertebrate. Pe când la mamifere femelele sunt homogametice (XX) şi masculii heterogametici (XY) şi la păsări tipul opus (femele, WY; masculi, YY), la peşti ambele mecanisme sunt prezente.

În ce priveşte numărul de cromosomi, ultimele cercetări arată că numărul diploid de cromosomi la peşti variază între 18 şi 104. Numărul cromosomilor determinat pe metafaze aparținând diviziunii mitotice, poate diferi radical pe parcursul diviziunilor repetate ale celulei-ou, deci în cadrul aceleiaşi specii şi aceluiaşi individ.

Observațiile asupra numărului de cromosomi la peşti sunt mult mai puține față de oricare alt grup de animale. Creşterea interesului asupra numărului de cromosomi şi morfologiei acestora promite importante contribuții la fundamentarea anumitor teorii privind evoluția şi filogenia peştilor.

Cunoaşterea unor elemente de genetică la crap prezintă o mare importanță pentru acțiunile de ameliorare, datorită faptului că la un reproducător se obține un număr foarte mare de descendenți (de ordinul zecilor sau chiar sutelor de mii), iar particularitățile genetice trebuiesc cunoscute în amănunt.

Lucrările de selecție sunt greu de abordat, deoarece fecundația, evoluția embrionului şi a larvei, precum şi desfăşurarea proceselor de creştere şi de dezvoltare au loc în apă, unde urmărirea acestora este mai greoaie. Pe lângă aceasta, mediul de viață acvatic are o influență pregnantă asupra creşterii şi dezvoltării descendenței, față de alte specii de animale (mai ales față de mamifere), toate acestea făcând ca rezultatele ameliorării să fie mai puțin spectaculoase.

Importanța cunoaşterii elementelor genetice la crap este accentuată şi de manifestarea unor fenomene caracteristice, cum ar fi letalitatea totală (Stan şi Păsărin 1996) întâlnită la indivizii homozigoți de crap oglindă, pentru a cărei înțelegere se dau următoarele explicații:

În primul rând, se apreciază că de învelişul solzos la crap sunt răspunzătoare două gene cu alelele lor şi anume:

- gena „S” şi alela sa „s” determină acoperirea totală a corpului cu solzi;
- gena „N” şi alela sa „n” determină neacoperirea cu solzi (sau tipul nud).

După unii autori, la crap se deosebesc 4 fenotipuri ale învelişului solzos şi anume:

- crap cu solzi pe toată suprafața corpului;
- crap cu solzi numai pe linia laterală;
- crap oglindă (cu puțini solzi răspândiți neuniform, mai ales sub formă de ramă);
- crap golaş (fără solzi sau puțini solzi şi răzleți);
- crap fără solzi;

Cu privire la transmiterea în descendență a tipului de înveliş solzos se apreciază că aceasta depinde de schemele de încrucişare aplicate, respectiv de combinarea genelor răspunzătoare de însuşirile respective.

Rezultatele încrucişării fenotipurilor de crap, în funcție de învelişul solzos, sunt conforme cu legile mendeliene. Aşa de exemplu, prin încrucişarea indivizilor homozigoți de crapi cu solzi (cu genotipul SSnn) cu cei oglindă (cu genotipul ssnn) se obțin produşi heterozigoți (cu genotipul Ssnn). Dacă aceşti indivizi sunt încrucişați în continuare cu indivizi de crap oglindă (ssnn), se obțin produşi cu genotipurile Ssnn (heterozigoți) şi ssnn (homozigoți oglindă) (Figura 2).

Trebuie menționat că indivizii homozigoți de crap oglindă cu genotipul NN, nu sunt viabili, deoarece sunt incompatibili cu procesele vitale (fenomenul de letalitate totală), aşa cum se poate observa în Tabelul 1.

Tabel 1. – Genotipurile caracteristice celor patru tipuri de crap

Specificare	Crap cu solzi	Crap oglindă	Crap cu solzi în linie	Crap golaş
În celule somatice	1. SSnn 2. Ssnn	ssnn	1. SSNn 2. SsNn	ssNn
În gameți	1. numai Ss 2. Ss sau sn	numai sn	1. SSNn 2. SsNn	sN sau sn

În cadrul familiei *Cyprinidae* s-au identificat două categorii de specii – prima cu complementul diploid variind în jurul cifrei 100 (*Cyprinus carpio* 2n=100, *Carassius auratus* 2n=104) şi cealaltă cuprinzând specii având 2n variind în jurul cifrei 50 (*Baarbus tetrazona* 2n=50, *Barbus fasciatus* 2n=52) (Minciu şi Băra, 1985).

În anul 1979, Taisescu a stabilit numărul diploid de cromosomi şi cariotipul la 10 specii de *Cyprinidae* (*Gobio gobio, Gobio uranoscopus, Gobio albipinnatus, Gobio kessleri, Pseudorasbora parva, Cyprinus carpio, Carassius carrasius, Carassius auratus, Scardinius erythrophthalmus şi Phoxinus phoxinus*), iar la specia *Gobio gobio* s-a întocmit idiograma.

Spre deosebire de alți autori care raportează la *Carassius carassius* numărul diploid identic cu cel al speciei *Carassius auratus* (2n=94, 100, 102, 104), cercetările prezente (Taisescu, 1979), stabilesc la caracudă 2n=50 (Figura 3) iar la carasul auriu 2n=98. Deoarece în literatura de specialitate foarte adesea este citată legătura filogenetică între cele două specii, s-a încercat stabilirea unei relații în sens evolutiv între complementele lor cromosomiale.

Studiul complementului cromosomial efectuat la patru linii şi doi hibrizi de crap (*Cyprinus carpio*), evidențiază în toate cazurile 2n=100 dintre care 24 sunt cromosomi metacentrici, 24 submetacentici şi subtelocentrici şi 52 sunt cromosomi acrocentrici. Neconcordanța rezultatelor din literatura de specialitate s-ar putea datora tehnicilor diferite folosite în evidențierea cromosomilor sau examinării unor linii sau populații diferite de crap care prezintă, probabil, un alt complement cromosomial.

La carasul auriu (*Carassius auratus*), a fost stabilit (Taisescu, 1979) numărul diploid de cromosomi 2n=98, iar în complementul cromosomial au fost identificați 48 de cromosomi cu două brațe şi 50 de cromosomi acrocentrici, totalizând un număr de brațe cromosomiale NF=146. Rezultatele par să mărească scara variabilității numerelor de cromosomi publicate pentru această specie (2n=94, 100, 102, 104).

La noi în țară, carasul este o specie ginogenetică (ceea ce presupune pentru perpetuare existența unor icre diploide, pentru a menține la urmaşi numărul de cromosomi caracteristic speciei), care prezintă o variabilitate fenotipică foarte pronunțată ca o consecință a participării diferitelor specii de *Cyprinidae* la „fecundație" şi populează condiții de mediu extrem de variate.

Faptul că se adaptează şi la condiții de mediu mai proaste demonstrează că specia se comportă ca un poliploid cu şanse mai mari de supraviețuire iar variația numărului de cromosomi în diferitele populații analizate s-ar putea datora unor rearanjamente cromosomiale secundare interacțiunii genotip – condiții de mediu.

Studiul complementului cromosomial la *Carassius carassius* stabileşte numărul diploid 2n=50 cromosomi, care se dispun în cariotip în 21 perechi de cromosomi cu 2 brațe (10 perechi metacentrici, 6 perechi submetacentrici, 5 perechi subterminali) şi 4

perechi de cromosomi acrocentrici realizând un număr fundamental de brațe cromosomiale, NF=92. În felul acesta, cercetările efectuate (Taisescu, 1979) sunt în discordanță cu cercetările lui S. Makino (1941) care raportează 2n=94 şi n=47 la fel ca şi la carasul auriu; cu cele ale lui H. Kobayasi (1977) care raportează 2n=94 fără să facă o descriere morfologică a complementului cromosomial şi cu cercetările lui U. Wolf şi colab., (1969), (citați de Taisescu, 1979), care raportează 2n=100 – 104, afirmație neînsoțită de material ilustrativ sau descrierea morfologică a complementului cromosomial.

Este posibil ca autorii citați anterior să fi analizat citogenetic o rasă geografică de caras auriu (specie cu o variabilitate individuală remarcabilă) care, datorită condițiilor din mediul pe care îl populează se aseamănă foarte mult, ca morfologie externă, cu caracuda.

Dacă admitem că cele două specii sunt atât de înrudite încât specia *Carassius carassius* constituie tipul sălbatic al speciei *Carassius auratus*, cea din urmă, fiind tetraploidă, numărul de cromosomi stabilit la *Carassius carassius* (2n=50) susține această evoluție prin poliploidizare însoțită de alte rearanjamente cromosomiale.

Analizând comparativ complementul cromosomial al mai multor specii de *Cyprinidae*, se pot trage două concluzii mai importante:

- în ceea ce priveşte numărul de cromosomi, se conturează două grupe de specii. Prima grupă cuprinde specii cu 2n= 48 – 50 (*Scardinius erythrophthalmus, Phoxinus phoxinus, Carassius carassius, Pseudorasbora parva* şi toate speciile genului *Gobio*) iar grupa a doua reuneşte specii cu 2n≅100 (*Cyprinus carpio, carassius auratus*) între cele două grupe ar putea fi o relație de tip diploid - tetraploid;
- subîmpărțirea familiei *Cyprinidae* în subfamilii este justificată citogenetic. Astfel, în subfamilia *Cyprininae* sunt cuprinse atât specii diploide cât şi specii tetraploide indicând modul şi direcția de apariție a unor specii. Gobioninaele au un complement cromosomial format din 50 – 52 de cromosomi, lipsit sau cu foarte puțini cromosomi acrocentrici. Leuciscinaele, mai puțin evoluate, au tot 2n=50 – 52 dar în complementul lor cromosomial se observă mai mulți cromosomi acrocentrici (caracter de primitivitate).

1.1. METODE UTILIZATE ÎN STUDIUL GENETIC AL POPULAȚIILOR

Populațiile pot fi studiate din punct de vedere genetic cu ajutorul a două metode: metoda descriptivă şi metoda genetica.

Metoda descriptivă se bazează pe caracterizarea fenotipică a indivizilor unei populații, considerându-se că fenotipul biotipurilor din populație reprezintă expresia genotipului indivizilor în strânsă relație cu mediul înconjurator. Studiind fenotipul organismelor se pot stabili unele raporturi de dependență existente între indivizii aceleiaşi colectivități. Deoarece într-o populație există indivizi asemănători fenotipic, dar desoebiți din punc de vedere genotipic, înseamnă că analiza populației prin metoda descriptivă nu oferă o imagine reală în ceea ca priveşte structura genetică a acesteia.

Metoda genetică de studiere a populațiilor a fst posibilă numai după redescoperirea legilor lui G. Mendel de segregare a caracterelor in descendențele hibride. Analiza hibridologică permite studierea formelor rezultate în urma segregării organismelor heterozigote. Cu ajutorul metodei genetice este posibilă evidențierea,

din diversitatea genotipurilor unei populații, a acelor genotipuri care corespund mai bine obiectivelor propuse de realizat în procesul de ameliorare.

În analiza genetică a populațiilor hibride se foloseşte metoda biometrică, completată cu calculul statistic al datelor, obținând, astfel, modelul teoretic al populației, pornind de la o anumită situație dată şi urmărind influența exercitată de diferiți factori care acționeaza asupra acesteia.

Astfel de studii genetice asupra populațiilor, bazate pe îmbinarea metodei biometrice cu calculul statistic, a efectuat W. Johansen (1903), care a definit atât *populația*, cât şi grupările componente de indivizi din cadrul populației, adică *liniile pure.*

Populațiile autogame, precum şi soiurile autogame, pot fi descompuse în *linii pure* prin alegerea descendenților pe parcursul mai multor generații şi prin consangvinizare.

De asemenea, în analiza genetică a populațiilor se studiază frecvența genelor şi a genotipurilor din populațiile naturale, după metoda elaborată de G. Hardz şi W. Weinberg (1908), care au demonstrat caracterul legic al echilibrul genetic al populațiilor.

ASPECTE PRIVIND FILIAȚIA ŞI DIVERSITATEA MOLECULARĂ

2.1. IMPORTANȚA BIOLOGIEI MOLECULARE ÎN INDICAREA DIRECȚIILOR EVOLUTIVE

Metodele comparative morfologice folosite în stabilirea liniilor filogenetice au fost foarte utile în studii asupra evoluției, atunci când nu existau verigi de legătură între fosilele descoperite. Oricum, aceste metode nu pot furniza o scală a evoluției corelată cu timpul. Progresul biologiei moleculare din ultimele decenii au furnizat noi metode de studiu asupra evoluției. Baza acestor metode o constituie gradul mare de stabilitate al secvențelor de ADN. Modificările evolutive în structura secvențelor de ADN sunt atât de lente, încât pot furniza informații detaliate asupra originii şi istoriei evoluției lor. Din momentul în care, secvențele ce intră în structura genelor au fost translate în secvențe de aminoacizi ale proteinelor cu ajutorul codului genetic, modificările evolutive ale aminoacizilor în cadrul proteinelor, de asemenea, pot furniza informații referitoare la procesele şi scala aproximativă a timpului evolutiv. De fapt, majoritatea rezultatelor obținute cu ajutorul studiilor la nivel molecular provin din analize ale secvențelor de aminoacizi ale anumitor proteine. Estimarea timpului evolutiv prin aceste metode se bazează pe descoperirea potrivit căreia, rata de substituție a aminoacizilor pe an pe site într-o anumită proteină se produce constant pentru toate organismele.

Este cunoscută existența a 20 de aminoacizi diferiți care intră în compoziția proteinelor. Numele şi modul de prescurtare sunt prezentate în Tabelul 2.

Tabel 2 Cei 20 de aminoacizi şi modul de prescurtare al numelor (Dayhoff, 1978)

Nr. crt.	Nume	Abrevieri		Nr. crt.	Nume	Abrevieri	
		trei litere	o literă			trei litere	o literă
1.	Alanină	Ala	A	11.	Leucină	Leu	L
2.	Arginină	Arg	R	12.	Lizină	Lzs	K
3.	Asparagină	Asn	N	13.	Metionină	Met	M
4.	Acid aspartic	Asp	D	14.	Fenilalanină	Phe	F
5.	Cisteină	Cys	C	15.	Prolină	Pro	P
6.	Glutamină	Gln	Q	16.	Serină	Ser	S
7.	Acid glutamic	Glu	E	17.	Treonină	Thr	T
8.	Glicină	Gly	G	18.	Triptofan	Trp	W
9.	Histidină	His	H	19.	Tirozină	Tyr	Z
10.	Izoleucină	Ile	I	20.	Valină	Val	V

2.2. ROLUL ADN MITOCONDRIAL (ADNmt)

Numărul mare de polimorfisme ale secvențelor de nucleotide din două zone cu hipervariabilitate ale regiunii de control mitocondrial necodificatoare pot permite discriminarea între indivizi sau probe biologice. Probabilitatea recuperării ADNmt din probe biologice foarte mici sau degradate este mai mare decât în cazul ADN nuclear deoarece moleculele de ADNmt sunt prezente în mii de copii per celulă, comparativ cu ADN nuclear care se găseşte în două copii per celulă. În plus, ADNmt este moştenit numai pe linie maternă, astfel încât, în cazul în care un individ nu este disponibil pentru o comparație directă cu o probă biologică, orice probă provenită de la genitorul matern, poate constitui o probă de referință.

Datorită recombinării meiotice şi moştenirii diploide a ADN nuclear, reconstituirea unui profil pe baza ADN nuclear chiar şi în cazul rudelor de gradul I este destul de greu de realizat. De asemenea, modelul moştenirii pe linie maternă a ADNmt poate fi de asemeni, considerat ca având unele deficiențe. Deoarece toți indivizii, pe linie maternă prezintă aceeaşi secvență de ADNmt, aceasta nu poate fi considerată un identificator unic. De fapt, indivizi neînrudiți aparent, pot avea o origine comună şi necunoscută, pe linie maternă.

Majoritatea animalelor au o reproducere sexuată, în timpul căreia, genele se transmit de la ambii părinți în urma recombinării prin crossing-over şi a segregării independente a cromosomilor din timpul meiozei I. Dar, genele din organite provenind din diferite linii de filiație niciodată nu se pot recombina şi aceasta tocmai datorită faptului că genomul organitelor (cum este şi cazul ADNmt) se transmite uniparental şi chiar daca s-ar transmite biparental, organitele provenite de la cei doi genitori nu ar putea fuziona astfel încât să se ajungă la o fuzionare a genomurilor. Dacă două celule parentale fuzionează, genotipurile recombinante nu se pot regăsi la nivel mitocondrial (Birky, 2001).

2.3. ASPECTE ALE EVOLUȚIEI MOLECULARE

2.3.1. EVOLUȚIA SECVENȚELOR NUCLEOTIDICE ŞI PROTEICE

De la Charles Darwin încoace reconstruirea istoriei evolutive a vieții şi redarea acesteia sub forma unui arbore filogenetic a fost un vis pentru mulți biologi, Haeckel (1866) fiind primul care a încercat sa realizeze acest lucru. Modul ideal pentru realizarea acestei sarcini imense este utilizarea registrului fosil, dar întrucât acesta este fragmentar mulți cercetători au încercat să folosească morfologia şi fiziologia comparată. Folosind aceste două metode s-a reuşit deducerea multor aspecte importante din istoria vieții. Însă schimbările morfologice şi fiziologice fiind extrem de complexe, iar aprecierea importanței evolutive a unui caracter fiind de cele mai multe ori subiective arborii filogenetici construiți au fost întotdeauna subiect pentru controverse.

În ultimul timp datorită realizărilor biologiei moleculare situația s-a schimbat drastic. Dacă în anul 1982 în GenBank erau doar 606 secvențe (680.338 pb) această bază de date fiind la începuturile sale, iar în 1997 se ajunsese după 15 ani abia la 1.765.847 de secvențe (1,160,300,687pb) astăzi baza de date numără 46 947 388 de secvențe cu un total de aproximativ 51 674 486 881 pb de la peste 165 000 de specii

(Benson, 2005). Această explozie se datorează în primul rând perfecționării metodelor de secvențiere şi a dus la apariția unui mare număr de studii de evoluție moleculară.

Planul structurii tuturor organismelor fiind înscris în ADN (sau ARN în cazul unor virusuri) apare posibilitatea de a studia relațiile evolutive dintre organisme comparându-le ADN-ul. Acest mod de a privi lucrurile are câteva avantaje față de cele clasice ce utilizează caractere morfologice şi fiziologice.

(1) Macromolecula de ADN este formată din patru tipuri de nucleotide: adenina (A), timina (T), citozina (C) şi guanina (G) în întreaga lume vie putând fi folosit astfel pentru a compara orice grupe de organisme, plante, bacterii, arhebacterii, animale, fungi. În metodele tradiționale aceasta este pur şi simplu imposibil.

(2) Întrucât modificările evolutive ale ADN urmează un model mai mult sau mai puțin regulat este posibilă utilizarea modelelor matematice pentru a formula modificările şi compara ADN-ul de la organisme foarte îndepărtate din punct de vedere filogenetic. Schimbările morfologice sau fiziologice urmează modele extrem de complexe chiar dacă analizăm perioade scurte de timp nefiind clar dacă diferitele aproximări necesare pentru realizarea unor filogenii sunt reale sau nu.

(3) Genomurile diferitelor organisme constau din lungi secvențe de nucleotide (de exemplu genomul uman are 9×1012 pb) conținând astfel cantități mult mai mari de caractere cu importanță filogenetică spre deosebire de alte tipuri de date.

Sistematica este una dintre cele mai controversate ramuri din biologie, definițiile speciilor, genurilor familiilor etc. fiind adesea subiectivă, dezacordul dintre specialişti privitoare la statutul unui anumit taxon fiind ceva obişnuit. Filogeneza este mai puțin controversată decât sistematica, ea dorind în primul rând să rezolve problema relațiilor dintre organisme, iar statutul sistematic al organismelor are o importanță secundară. Cele două domenii sunt însă strict corelate, clasificarea organismelor trebuind să reflecte istoria lor evolutivă (Ax, 1987). În acest sens filogenetica moleculară joacă un rol important în dezvoltarea bazelor ştiințifice ale sistematicii, însă nu trebuie privită ca atotputernică existând cazuri când unele probleme din sistematică nu pot fi rezolvate pe această cale.

Un exemplu interesant de utilizarea a datelor moleculare în stabilirea relațiilor filogenetice dintre diverşi taxoni îl constituie cetaceele a căror legături cu celelalte ordine de mamifere au fost mult discutate. Prin analiza evoluției genei γ fibrinogenului cât şi prin datele referitoare la inserția în genom a SINE (short interspersed repetitive elements) ce pot fi privite ca caractere sinapomorfe (Nei & Kumar, 2000) s-a ajuns la concluzia că printre mamiferele actuale, cele mai înrudite cu cetaceele sunt hipopotamii. Pe lângă faptul că la acealeaşi rezultate s-a ajuns prin două metode absolut diferite, ipoteza este susținută şi de datele paleontologice care indică înrudirea cetaceelor cu mezonichidele, un grup de ungulate carnivore din oligocen (Benton, 1997).

2.3.2. METODE DE DEDUCERE A DIRECȚIILOR DE FILIAȚIE PORNIND DE LA DATE MOLECULARE

Deducerea informațiilor filogenetice din datele moleculare necesită alegerea metodei potrivite din multele disponibile. Construirea unei ipoteze filogenetice este un procedeu de estimare, adică se face o estimare optimă a unei filogenii bazându-ne pe informația incompletă conținută în datele disponibile. În contextul sistematicii moleculare de obicei nu avem informații directe despre trecut – avem acces doar la organisme şi molecule actuale. Deoarece putem găsi un scenariu evolutiv prin care

orice filogenie poate fi explicată e nevoie de un criteriu pentru a selecta una sau câteva filogenii preferate dintr-un set de filogenii posibile. Metodele de deducere filogenetică încearcă să realizeze acest scop în unul din două moduri: (1) definind o secvență specifică de paşi (un algoritm) ce duce la obținerea unui arbore filogenetic ; sau (2) definind un criteriu pentru a compara filogeniile alternative între ele şi decizând care e mai bună.

Metodele pur algoritmice combină deducerea arborelui filogenetic şi definiția arborelui preferat într-o singură supoziție. Aceste metode includ toate formele de analiză a grupelor perechi (cum ar fi UPMGA) şi alte metode bazate pe distanțe cum ar fi NJ (neighbor joining). Ele sunt de obicei mai rapide deoarece merg direct spre arborele final fără a necesita evaluarea unui număr mare de soluții.

A doua grupă de metode presupune doi paşi logici. Primul este definirea unui criteriu de optimalitate (descris formal de printr-o funcție obiectivă) cu ajutorul căruia fiecărui arbore i se acordă un punctaj folosit pentru compararea arborilor unul cu celălalt. Al doilea pas este utilizarea unui algoritm specific pentru calcularea valorii funcției obiective şi găsirea arborelui cu cel mai bun punctaj. Astfel supozițiile evolutive făcute în primul pas sunt decuplate de folosirea calculatorului în pasul al doilea. Prețul acestei clarități logice este viteza lentă de lucru a calculatorului, iar pentru seturi date ce conțin mai mult de 20 de taxoni arborele optim nu poate fi întotdeauna găsit din cauza numărului mare de posibile soluții.

Folosirea algoritmilor este radical diferită între cele două metode. În cazul metodelor pur algoritmice, algoritmul este cel care defineşte criteriul de selecție al arborilor şi are deci o valoare crucială. În metodele bazate pe un criteriu algoritmul este doar unealtă pentru evaluarea valorii funcției obiective şi căutarea arborilor ce optimizează această valoare. Deoarece metodele bazate pe criterii acordă fiecărui arbore analizat un punctaj, filogeniile pot fi aranjate în ordinea preferințelor. Aceasta reprezintă un avantaj enorm în fața metodelor pur algoritmice. Dacă un set de date poate fi explicat egal de bine de câteva mii de arbori filogenetici, folosind metode bazate pe criterii nu vom fi duşi în eroare că un anumit arbore explică datele foarte bine. O metodă algoritmică dă un singur arbore, neputând şti câtă confidență putem avea în acesta. Unii autori (Hedges et al., 1992) au sugerat utilizarea metodelor statistice cum ar fi bootstrapping-ul pentru determinarea gradului de confidență pe care îl putem avea într-un arbore găsit prin metode pur algoritmice.

2.3.3. EVOLUȚIA SECVENȚELOR DE AMINOACIZI

Înainte de inventarea metodelor rapide de secvențiere în 1977 (Maxam & Gilbert, 1977; Sanger et al., 1977) majoritatea studiilor de evoluție moleculară se bazau pe secvențe proteice. Astăzi secvențierea ADN-ului este mult mai simplă şi rapidă decât secvențierea proteinelor şi majoritatea secvențelor proteice sunt obținute prin transcriere folosind codul genetic. Totuşi secvențele proteice şi-au păstrat utilitatea pentru studiile filogenetice : ele sunt mult mai conservate, decât secvențele de ADN şi pot oferi informații utile în cazul comparării grupelor îndepărtate din punct de vedere filogenetic. În plus, modelele matematice ale schimbării evolutive a aminoacizilor sunt mult mai simple decât cele pentru secvențe de acizi nucleici, motiv pentru care facem mai întâi o prezentare a evoluției secvențelor proteice şi doar apoi a celor nucleotidice.

Studiul schimbărilor evolutive al secvențelor polipeptidice începe prin compararea a două sau mai multe secvențe. Cel mai simplu indice este numărul de aminoacizi

diferiți (*nd*). Daca secvențele aliniate conțin inserții sau deleții acestea trebuiesc eliminate înainte de calcularea *nd*. Un alt indice simplu este distanța *p* ce reprezintă proporția de aminoacizi prin care două secvențe de aminoacizi diferă. Se calculează împărțind numărul de aminoacizi diferiți la numărul total de situsuri comparate.

$$p=\frac{n_d}{n}\ ; \qquad V(p)=\frac{p(1-p)}{n}$$

Acest indice nu face nici o corecție pentru substituții multiple în cadrul aceluiaşi situs sau pentru diferențele în ratele de evoluție între situsuri.

Următorul indice este distanța cu corecție Poisson (PC) deoarece în realitate în cadrul aceluiaşi situs pot avea loc substituții multiple ceea ce duce la o relație neliniară între *p* şi timpul *t*. Odată cu trecerea timpului discrepanța dintre *nd* şi numărul real de substituții creşte (Figura 1)

Dacă *r* este rata de substituție per an şi per situs (pentru simplitate se presupune că e acelaşi pentru toate situsurile) numărul mediu de substituții per situs într-o perioada de timp *t* este *rt*, iar probabilitatea a *k* substituții (*k* = 1, 2, 3, …) este dată de distribuția Poisson :

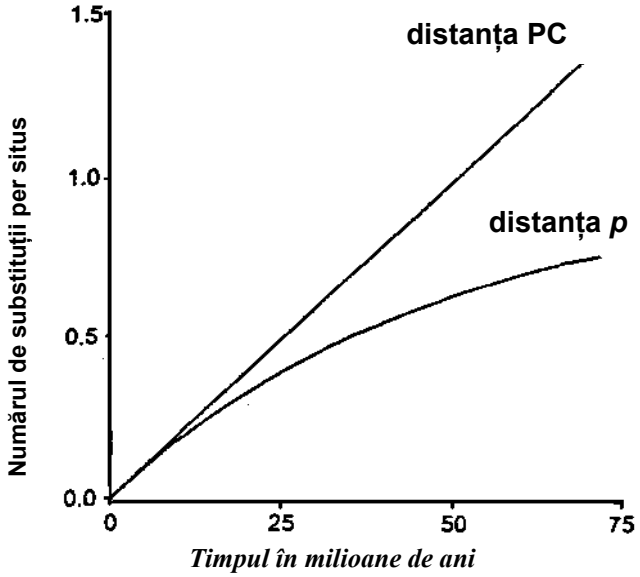


Figura 1. Relația dintre distanța PC şi distanța p odată cu trecerea timpului. Rata de substituție a aminoacizilor este de 10^{-8} per situs per an. *(După Nei & Kumar, 2000)*

$$P(k;t)=\frac{e^{-rt}(rt)^k}{k!}$$

Deci probabilitatea ca într-un anumit situs să nu aibă loc nici o substituție (*k* =0) este *P*(0;*t*)=*e-rt*. Pentru un număr de n aminoacizi, numărul de aminoacizi neschimbați este de *ne-rt.*

De obicei însă, secvența ancestrală nefiind cunoscută ecuația de mai sus nu este aplicabilă, trebuind să comparăm două secvențe actuale divergența cărora a început *t* timp în urmă. Probabilitatea (*q*) ca nici unul din situsurile omoloage ale celor două secvențe să nu se fi schimbat este:

$$q = (e\text{-}rt)2 = e\text{-}2rt$$

Ea poate fi estimată ca 1 - ^*p*, deoarece *q* = 1 – *p*, de unde 1- *p* = *e-2rt*. Daca *d* = 2*rt* este numărul total de substituții per situs pentru două secvențe atunci el poate fi calculat ca :

$$d = -\ln(1 - p)$$

O estimare a lui *d* ($\hat{d}$) poate fi obținută înlocuind în ecuația de mai sus *p* cu $\hat{p}$, iar varianța lui $\hat{d}$ este dată de

$$V(\hat{d}) = p / [(1 - p)\, n].$$

Pe de altă parte aceleaşi formule de mai sus pot servi pentru calcularea ratei de substituție a aminoacizilor

$$\hat{r} = \hat{d} / (2t),$$

dar pentru aceasta trebuie să cunoaştem momentul începerii divergenței dintre cele două secvențe. Daca cunoaştem rata de substituție a aminoacizilor putem folosi formula de mai sus pentru a estima timpul evolutiv

$$\hat{t} = \hat{d} / (2r).$$

În modul de calculare a distanței descris mai sus se presupune că rata de substituție aminoacizilor este identică pentru toate situsurile. Această aproximare de cele mai multe ori nu este valabilă, rata substituției fiind mai ridicată pentru situsurile mai puțin importante din punct de vedere funcțional, iar aminoacizii din centrul activ al unei proteine pot fi conservați pentru perioade de timp foarte mari (Kimura, 1983). Uzzel & Corbin (1971) au arătat că de fapt distribuția numărului de substituții per situs *k* are o varianță mai mare decât cea a distribuției *Poisson* şi urmăreşte de fapt o distribuție negativ binomială, iar distribuția lui *r* per situs urmăreşte o distribuție gamma.

f (*r*) = $\frac{b^a}{\Gamma(a)}$ *e -b r r a-1* unde *a*= , $\bar{r}/V(r)$ $b = \bar{r}/V(r)$, iar $\Gamma(a)$ este funcția gamma definită:

$$\Gamma(a) = \int_0^\infty e^{-t} t^{a-1} dt$$

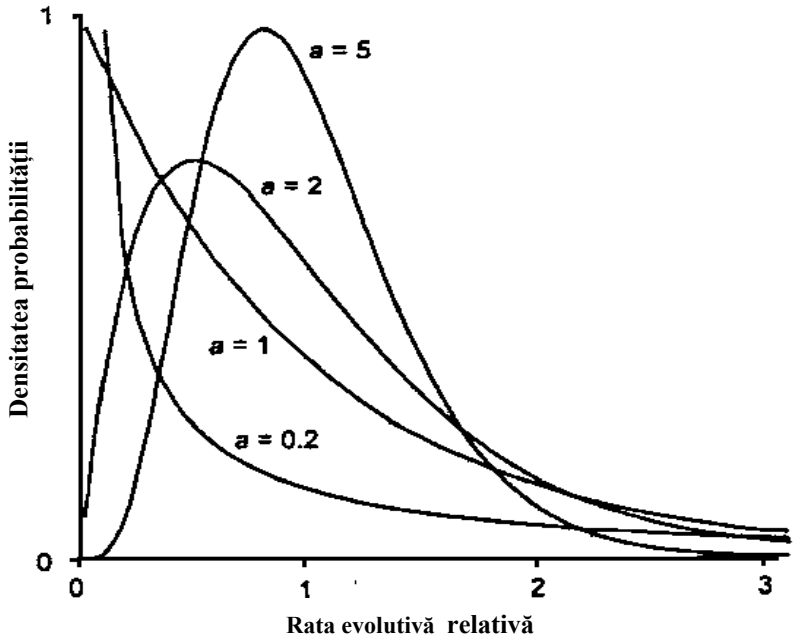


Figura 2 Distribuția gamma a ratei de substituție per situsuri pentru diferite valori ale parametrului gamma (*a*) (După Nei & Kumar, 2000)

Forma distribuției *f*(*r*) este dată de *a* care se mai numeşte şi parametru de formă sau parametru gamma. Distribuția gamma este foarte flexibilă şi ia diverse forme în dependență de *a*. Pentru $a = \infty$ *r* este acelaşi pentru toate situsurile, pentru *a*=1 *r* urmează o distribuție exponențială indicând că rata de substituție variază mult de la un situs la altul, iar pentru a < 1 *r* are o distribuție şi mai concavă indicând că un mare număr de situsuri au *r* = 0 (Figura 2).
Spre exemplu Uzzell & Corbin (1971) au estimat că *a* = 2 pentru citocromul c de la vertebrate, iar Zhang & Gu (1998)au estimat acest parametru între 0,2 şi 3,5 pentru un set de proteine nucleare şi mitocondriale. Când *r* variază conform unei distribuții gamma poate fi estimat numărul de substituții de aminoacizi per situs. Probabilitatea identității a doi aminoacizi din două situsuri omoloage peste un timp *t* este *q* = *e-2rt*, iar *q* mediu pentru toate situsurile este dat de

$$\overline{q} = \int_0^\infty q f(r) dr = \left(\frac{a}{a + 2\overline{r}t} \right)^a$$

(Nei & Kumar, 2000). daca notăm ca în cazul corecției Poisson *dG* $= 2\overline{r}t$ şi $\overline{q} = 1 - p$ avem următoarea formulă pentru distanța dintre două secvențe :

$$d_G = a\left[(1-p)^{-1/a} - 1\right] \qquad \text{cu varianța} \qquad V(\hat{d}_G) = p\left[(1-p)^{-(1+2/a)}\right]/n$$

iar estimarea ($\hat{d}_G$) a lui *dG* se obține înlocuind *p* prin $\hat{p}$. Efectul variației lui *r* asupra distanței *d* este semnificativ numai atunci când *p* > 0,2 iar *a* < 0,2, deci când *p* < 0,2 nu este necesară utilizarea distanței *dG*.

2.3.4. MUTAȚIILE ŞI RATA DE SUBSTITUȚIE A AMINOACIZILOR

Genele sau secvențe ale moleculelor de ADN care funcționează ca modele structurale pentru ARNm sunt numite gene structurale. De când secvențele de aminoacizi din structura unui lanț polipeptidic se pot determina cu ajutorul secvențelor nucleotidice din genele structurale, orice schimbare în secvența de aminoacizi este cauzată de mutațiile produse la nivelul ADN. De asemenea, o mutație care produce o modificare în structura ADN nu este necesar să se reflecte în modificarea secvenței aminoacizilor, fapt datorat degenerării codului genetic.

Înaintea inventării metodelor rapide de secvențiere a ADN în 1977 (Maxam şi Gilbert, 1977; Sanger et al., 1977), cea mai mare parte a studiilor de evoluție moleculară au avut la bază utilizarea secvențelor de aminoacizi. Unele principii importante ale evoluției moleculare, cum ar fi: evoluția prin duplicarea genelor (Ingram 1963; Ohno, 1970) şi ceasul evoluției moleculare (Zuckerkandl şi Pauling, 1962; Margoliash, 1963), au fost descoperite prin studiul secvențelor de aminoacizi. În prezent, secvențierea ADN este mult mai simplă decât secvențierea aminoacizilor care este de obicei dedusă din secvențierea nucleotidelor prin utilizarea codului genetic (Nei şi Kumar, 2000).

Nu este adevărat faptul că orice mutație poate fi încorporată într-o secvență, deoarece speciile sunt alcătuite din populații cu numeroşi indivizi, iar o mutație nou apărută într-un individ, poate dispare din populație prin şansă sau prin acțiunea selecției purificatoare. Numai în cazul în care mutația survine la nivelul întregii populații, atunci aceasta poate fi inclusă în genomul speciei. Acest proces poartă numele de fixare a mutației în populație. Odată fixată o mutație în cadrul unei populații, orice individ aparținând acelei populații va avea aceeaşi mutație. Când sunt comparate două secvențe de aminoacizi provenite de la două specii diferite, sunt urmăriți inițial, noii aminoacizi care au fost încorporați.

Când apare o mutație în cadrul populației, supraviețuirea alelelor depinde în principal de şansă, dacă este sau nu avantajoasă din punct de vedere selectiv şi de mărimea populației. De exemplu, considerând A1 gena sălbatică şi A2 alela. Într-un organism diploid, gena mutantă apare inițial în stare heterozigotă (A1A2). În urma încrucişării A1A2 x A1A1, dacă nu rezultă descendenți (generația F1) din diferite cauze biologice (de exemplu sterilitatea homozigotului A1A1), gena mutantă va dispare în generația următoare. Supraviețuirea genei alele, nu este sigură nici chiar dacă în urma încrucişării A1A2 x A1A1 se produc urmaşi, deoarece descendenții cu genotipul A1A2 apar cu o probabilitate de 0,5. În plus, există şi o şansă de 0,25 ca descendenții cu aceste genom să nu apară (Nei şi Kumar, 2000).

Când o mutație este neutră sau nu afectează fitness-ul individului purtător, frecvența relativă a mutantei poate creşte sau descreşte prin şansă în cadrul populației (Nei şi Kumar, 2000).

Până acum în toate modelele descrise s-a vorbit despre substituțiile aminoacizilor de parcă fiecare substituție este fixată în secvența considerată. De fapt secvențele de aminoacizi nu sunt independente, fiecare specie constând din populații, iar mutațiile au loc în individ ele putând dispare prin derivă genetică sau prin selecție

stabilizatoare. Doar când mutația se răspândeşte în întreaga populație putem spune că ea este fixată. Când comparăm două secvențe de aminoacizi de la două specii observăm doar schimbările ce au fost fixate în genofondul speciei. Când o mutație este neutră şi nu afectează fitnesul indivizilor frecvența relativă (*x*) a alelei mutante poate creşte sau descreşte la întâmplare în interiorul populației. La un organism cu generații discrete (cum ar fi insectele sau plantele anuale) şi număr *N* de indivizi adulți frecvența unei alele mutante este $x = 1/(2N)$, unde 2*N* este numărul de locusuri. Soarta alelei e determinată doar de şansă, *x* crescând şi descrescând până când alela e pierdută sau fixată. Probabilitatea fixării este de $u = 1/(2N)$ iar probabilitatea pierderii de 1-1/(2*N*). Daca rata de apariție a mutației este de *v* per locus şi per generație atunci numărul de mutații ce au loc în întreaga populație este de 2*Nv* per locus şi per generație. Cum probabilitatea fixării mutației este de 1/(2*N*) rata de substituție per locus şi per generație (*a*) este de :

$$a = 2Nv \times \frac{1}{2N} = v$$

Deci rata de substituție per locus şi per generație este egală cu rata mutației. Aceasta regulă simplă a fost descoperită în 1968 de către Kimura.

În cazul mutațiilor avantajoase selective situația este alta, fitnesul indivizilor fiind de 1 în cazul genotipului *AA*, 1+*s* în cazul genotipului *Aa* şi 1+2*s* în cazul genotipului *aa* (*s* este numit coeficient de selecție şi este pozitiv pentru mutațiile folositoare şi negativ pentru cele dăunătoare. Probabilitatea fixării mutațiilor avantajoase $u \approx 2s$ pentru populații mari unde $Ns > 1$, deci rata de fixare a mutațiilor avantajoase este

$$a = 2Nv \times 2s = 4Nsv$$

Să presupunem că *v* = 10-9, *N* = 105 iar *s* = 0,01. În acest caz *a* devine 4×10^{-6} care este de 4000 de ori mai mare ca în cazul mutației neutre (unde *a* = *v* =10-9). Acest exemplu este totuşi un pic exagerat deoarece odată ce o proteină are o funcție stabilită majoritatea mutațiilor vor fi dăunătoare sau neutre şi foarte puține vor amplifica activitatea proteinei. Mai multe studii au arătat că rata mutațiilor pozitive este mult mai mică ca a celor neutre.

Marea majoritate a mutațiilor sunt însă dăunătoare, ele fiind rapid eliminate din populație, ele necontribuind deci la substituțiile de aminoacizi nefiind fixate. Mutațiile puțin dăunătoare se comportă altfel însă. Ele pot fi totuşi fixate într-o populație cu o mică probabilitate în special în populații mici. Totuşi o acumulare continuă de mutații puțin dăunătoare i-ar putea deteriora structura şi funcția. Deci ca mutațiile uşor dăunătoare să fie fixate trebuie să se acumuleze şi mutații uşor benefice pentru ca proteina să-şi păstreze funcția. În acest caz rata medie de acumulare a mutațiilor poate deveni mai mult sau mai puțin (constantă pe termen lung bineînțeles).

În unele cazuri heterozigoții (*Aa*) pentru o nouă mutație pot avea un fitnes crescut față de oricare dintre homozigoți (*AA* sau *aa*) acest fenomen fiind cunoscut sub denumirea de selecțe supradominantă. În acest caz fitnesul indivizilor *AA*, *Aa* şi *aa* poate fi exprimat ca 1-*s1*, 1 şi 1-*s2* respectiv calculele probabilității fixării sunt prea complicate pentru a fi redate aici, însă când $s_1 \cong s_2$ probabilitatea este de câteva ori mai mare ca cea pentru mutațiile neutre.

2.4. TIPURI DE MODIFICĂRI LA NIVELUL MACROMOLECULEI DE ADN

Există patru tipuri principale de modificări care au loc în macromolecula de ADN: înlocuirea nucleotidelor, deleția, adiția şi inversia acestora. Adiția, deleția şi inversia se pot produce pentru una sau mai multe nucleotide. Adiția şi deleția pot modifica secvențele cadrelor deschise de citire, porțiuni numite ORF (open reading frame), caz în care, sunt numite frameshift mutations (Figura 3). Înlocuirea nucleotidelor poate fi împărțită în două clase: tranziții şi transversii (Figura 4). Tranzițiile reprezintă înlocuirea unei baze azotate purinice (adenină sau guanină) cu altă bază purinică sau a unei baze azotate pirimidinice (timină sau citozină) cu o altă pirimidină. Substituția nucleotidelor purinice cu nucleotide pirimidinice sau invers, poartă numele de transversii. În majoritatea secvențelor de ADN, s-a constatat că substituția nucleotidelor prin tranziții, intervine mai frecvent decât substituția prin transversii (Fitch, 1967; Gojobori et al., 1982; Kocher et al., 1991).

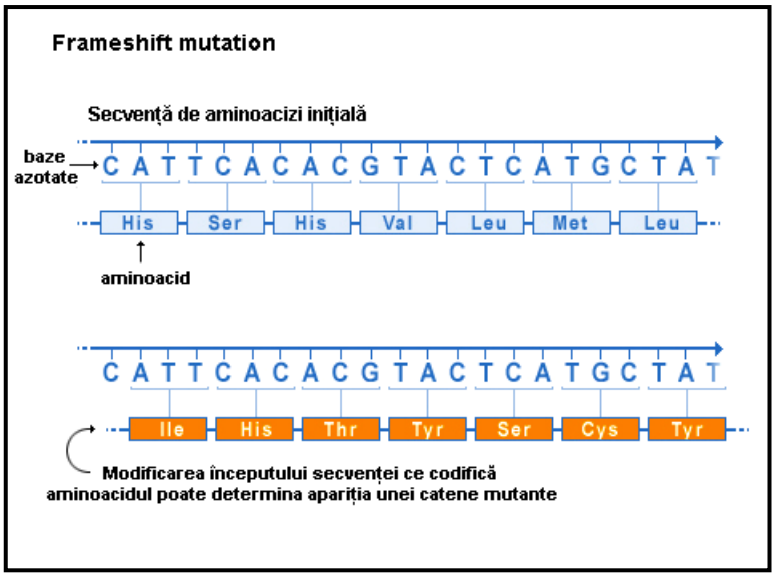


Figura 3 Modul de apariție al frameshift mutațiilor

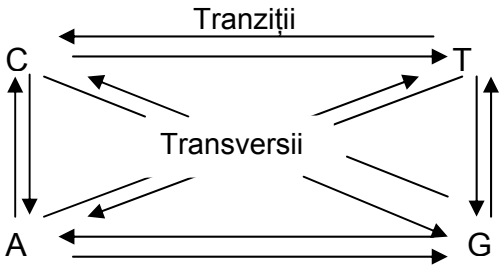


Figura 4 Cele două modalități de substituție a nucleotidelor
(după, Nei şi Kumar, 2000)

Modificările evolutive de la nivelul macromoleculelor de ADN sunt mai complicate decât cele din secvențele proteice, deoarece există mai multe tipuri de ADN cum ar fi regiunile codificatoare, regiuni necodificatoare, exoni, introni, regiuni terminale, regiuni

repetitive şi secvențe de inserție. Modificările produse de mutații la nivelul ADN; variază în funcție de zona în care apar. Chiar dacă luăm în considerare o singură regiune codificatoare, modelul substituției nucleotidelor în prima, a doua şi a treia poziție a unui codon, este diferit. Mai mult, unele regiuni se află sub acțiunea selecției naturale, mai mult decât altele, acest fapt contribuind la variații ale modelului de evoluție ale diferitelor regiuni ale ADN.

2.4.1 DIFERENȚE NUCLEOTIDICE ÎNTRE SECVENȚE

Când două secvențe de ADN provin din aceeaşi secvență ancestrală, evoluează divergent gradual, datorită substituției nucleotidice. Cea mai simplă măsură a divergenței dintre secvențe, este proporția siturilor nucleotidice care diferă între două nucleotide sau distanța *p* pentru secvențele nucleotidice.

În cazul substituției aminoacizilor, distanța p ne dă o estimare a numărului de nucleotide substituite per site, când secvențele au origine apropiată. Când p are valoare mare, ne dă o subestimare numerică, deoarece nu ține cont de mutațiile înapoi şi de cele paralele.

Pentru a estima numărul substituțiilor nucleotidice, este necesar să folosim un model matematic al substituțiilor nucleotidice. Din acest motiv, a fost constituită o matriță a ratei substituțiilor nucleotidice (Tabelul 3) (Nei şi Kumar, 2000).

În continuare, vor fi prezentate doar modelele utilizate în prelucrarea datelor din prezenta lucrare, modele alese pe baza numărului de nucleotide implicate precum şi a numărului de mutații apărute în cadrul secvenței unei gene.

Tabel 3 Modele ale substituției nucleotidice (Nei şi Kumar, 2000)

	A	T	C	G	A	T	C	G
Jukes – Cantor model					HKY model			
A	-	α	α	α	-	βgT	βgC	αgG
T	α	-	α	α	βgA	-	αgC	βgG
C	α	α	-	α	βgA	αgT	-	βgG
G	α	α	α	-	αgA	βgT	βgC	-
Kimura model					Tamura – Nei model			
A	-	β	β	α	-	βgT	βgC	α1gG
T	β	-	α	β	βgA	-	α2gC	βgG
C	β	α	-	β	βgA	α2gT	-	βgG
G	α	β	β	-	α1gA	βgT	βgC	-
Equal – input model					General reversible model			
A	-	αgT	αgC	αgG	-	agT	bgC	cgG
T	αgA	-	αgC	αgG	agA	-	dgC	egG
C	αgA	αgT	-	αgG	bgA	dgT	-	fgG
G	αgA	αgT	αgC	-	cgA	egT	fgC	-
Tamura model					Unrestricted model			
A	-	βθ2	βθ1	αθ1	-	a12	a13	a14
T	βθ2	-	αθ1	βθ1	a21	-	a23	a24
C	βθ2	αθ2	-	βθ1	a31	a32	-	a34
G	αθ2	βθ2	βθ1	-	a41	a42	a43	-

Notă: un element (*eij*) care intră în una din matricile anterioare de substituție nucleotidică va înlocui nucleotida din rândul *i* cu nucleotida din coloana *j*. gA, gT, gC şi gG sunt frecvențele nucleotidelor. θ1 = gG + gC; θ2 = gA + gT.

$$\hat{p} = n_d / n$$

unde n_d şi n sunt numărul de nucleotide diferite şi numărul total de situsuri, varianța lui p este aceiaşi ca pentru secvențele de aminoacizi. Pe lângă acest indice ades este necesar să cunoaştem frecvențele diferitelor perechi de nucleotide dintre două secvențe, X şi Y. Existând patru nucleotide, A, T, G şi C în fiecare secvență există 16 perechi diferite de nucleotide. Patru perechi sunt formate din nucleotide identice (AA, TT, CC şi GG), patru sunt perechi de tipul tranzițiilor (AG, GA, TC şi CT) şi opt sunt perechi de tipul transversiilor (restul perechilor). Frecvențele totale ale perechilor de nucleotide identice, perechilor de tipul tranzițiilor, şi perechilor de tipul transversiilor pot fi notate cu O, P şi Q respectiv, deci $p = P + Q$. Dacă substituțiile de nucleotide ar avea loc aleatoriu Q ar trebui să fie de două ori mai mare ca P (8 : 4) pentru un p mic. De fapt tranzițiile au loc mai frecvent ca transversiile şi deci P poate fi mai mare decât Q. Raportul R dintre tranziții şi transversii poate fi calculat după formula $\hat{R} = \hat{P} / \hat{Q}$ unde $\hat{P}$ şi $\hat{Q}$ sunt valorile observate ale lui P şi Q. R este de obicei de 0,5 – 2 pentru multe gene nucleare, iar pentru genele mitocondriale poate fi până la 15. $\hat{R}$ este supus unei mari erori când numărul de nucleotide studiate este mic şi varianța este dată de :

$$V(\hat{R}) = R^2 \left(\frac{1}{nP} + \frac{1}{nQ} \right)$$

fapt ce indică că dacă nP şi nQ nu sunt mari, varianța lui $\hat{R}$ poate fi foarte mare. În calculele lui $V(\hat{R})$, R din ecuația de mai sus este înlocuit cu $\hat{R}$.

Ca şi în cazul secvențelor de aminoacizi distanța p dă o estimare corectă a numărului de substituții per situs numai dacă secvențele sunt destul de apropiate., dacă p este mare apare o subestimare a numărului de substituții deoarece nu sunt luate în calcul substituțiile paralele şi cele ce au loc înapoi. În cazul secvențelor nucleotidice aceasta este mult mai grav deoarece există doar patru stări posibile în loc de 20.

2.4.2 MODELE DE EVOLUȚIE ALE SECVENȚELOR NUCLEOTIDICE

Pentru a putea estima numărul de substituții nucleotidice este necesară utilizarea unor modele matematice care să corecteze substituțiile multiple. Din această cauză au fost dezvoltate un număr mare de modele. Expresia matematică a unui model de substituție este un tabel al ratelor de substituție per situs şi per unitate de distanță evolutivă. Pentru secvențele de ADN aceste rate pot fi exprimate ca o matrice Ω de 4 $\times$ 4 în care fiecare element Ω_{ij} reprezintă rata schimbării de la baza i la baza j în timpul unei perioade de timp dt infinit de mică. Cea mai generală formă a acestei matrici este:

$$\Omega=\begin{pmatrix} -\mu(a\pi_C+b\pi_G+c\pi_T) & \mu a\pi_C & \mu b\pi_G & \mu c\pi_T \\ \mu g\pi_A & -\mu(g\pi_A+d\pi_G+e\pi_T) & \mu d\pi_G & \mu e\pi_T \\ \mu h\pi_A & \mu j\pi_C & -\mu(h\pi_A+j\pi_C+f\pi_T) & \mu f\pi_T \\ \mu i\pi_A & \mu k\pi_C & \mu l\pi_G & \mu(i\pi_A+k\pi_C+l\pi_G) \end{pmatrix}$$

rândurile şi coloanele corespunzând celor patru perechi de baze (A, C, G şi T). Factorul μ este dat de rata medie instantanee de substituție care este modificat de parametrii relativi *a*, *b*, *c*….*l*, care corespund câte unul pentru fiecare transformare posibilă dintr-o bază într-alta diferită. Produsul dintre μ şi parametrii relativi este *parametrul ratei de substituție*. Ceilalți parametri, π_A, π_C, π_G şi *πT* sunt *parametri de frecvență* şi corespund frecvențelor bazelor A, C, G şi T respectiv. Pentru a simplifica modelul se admite că frecvențele rămân constante în timp, aflându-se deci în echilibru, iar rata de schimbare spre oricare dintre baze este proporțională cu frecvența de echilibru şi este independentă de baza de la care se porneşte. În matrice elementele din diagonală sunt alese astfel încât în sumă toate elementele de pe un rând dau 0. O matrice analoagă poate fi construită şi pentru aminoacizi, în acest caz ea ar fi de 20×20 şi nu de 4×4.

Aproape toate modelele de substituție propuse până la ora actuală sunt cazuri particulare ale matricei de mai sus (motiv pentru care am şi prezentat-o). De obicei se admite că rata totală a schimbării de la o bază *i* la alta *j* pentru o perioadă mai lungă de timp este aceiaşi cu rata schimbării de la *j* la *i*. În termeni matematici aceasta corespunde cu afirmația că *g* = *a*, *h* = *b*, *i* = *c*, *j* = *d*, *k* = *e* şi *l* = *f*. Asemenea modele se numesc cu *timp reversibil* . Cel mai generalizat model cu timp reversibil (GTR) este reprezentat prin matricea simetrică :

$$\Omega=\begin{pmatrix} -\mu(a\pi_C+b\pi_G+c\pi_T) & \mu a\pi_C & \mu b\pi_G & \mu c\pi_T \\ \mu a\pi_A & -\mu(g\pi_A+d\pi_G+e\pi_T) & \mu d\pi_G & \mu e\pi_T \\ \mu b\pi_A & \mu d\pi_C & -\mu(h\pi_A+j\pi_C+f\pi_T) & \mu f\pi_T \\ \mu c\pi_A & \mu e\pi_C & \mu f\pi_G & \mu(i\pi_A+k\pi_C+l\pi_G) \end{pmatrix}$$

Majoritatea altor modele folosite fie pentru deducerea arborilor filogenetici pe baza maximei verosimilitudini (maximum likelihood) sau pentru estimarea distanțelor evolutive perechi pot fi obținute restrângând alți parametric ai modelului GTR. Spre exemplu dacă substituțiile sunt împărțite în transversii, tranziții între bazele purinice şi tranziții între bazele pirimidinice (adică a = c = d = f) obținem modelul lui Tamura şi Nei din 1993 (TrN). La fel putem obține modelul lui Kimura cu 3 tipuri de substituții (K3ST) cerând ca toate bazele să aibă aceiaşi frecvență ($\pi_A = \pi_C = \pi_G = \pi_T = 1/4$) şi împărțind substituțiile în tranziții (*b* = *e*), transversii de tipul A-T sau C-G (*c* = *d*) şi transversii de tipul A-C sau G-T (*a* = *f*). Alte restricții duc la modele mai simple şi mai des utilizate. Dacă admitem că frecvența de echilibru a bazelor este aceiaşi ($\pi_A = \pi_C = \pi_G = \pi_T = 1/4$), iar toate substituțiile au aceiaşi rată (a = b = c = d = e = f = 1) modelul se reduce la cel al lui Jukes şi Cantor din 1969 (JC).

$$\Omega = \begin{pmatrix} -\frac{3}{4}\mu & \frac{1}{4}\mu & \frac{1}{4}\mu & \frac{1}{4}\mu \\ \frac{1}{4}\mu & -\frac{3}{4}\mu & \frac{1}{4}\mu & \frac{1}{4}\mu \\ \frac{1}{4}\mu & \frac{1}{4}\mu & -\frac{3}{4}\mu & \frac{1}{4}\mu \\ \frac{1}{4}\mu & \frac{1}{4}\mu & \frac{1}{4}\mu & -\frac{3}{4}\mu \end{pmatrix}$$

iar dacă combinăm frecvența bazelor şi rata de substituție într-un singur parametru $\alpha = \mu / 4$ se obține o formă mai simplă. Acest parametru reprezintă probabilitatea schimbării unui nucleotid în oricare altul :

$$\Omega = \begin{pmatrix} -3\alpha & \alpha & \alpha & \alpha \\ \alpha & -3\alpha & \alpha & \alpha \\ \alpha & \alpha & -3\alpha & \alpha \\ \alpha & \alpha & \alpha & -3\alpha \end{pmatrix}$$

Un alt model utilizat frecvent este cel al lui Kimura cu 2 parametri propus în 1980 (K2P). El ia în calcul observația că tranzițiile şi transversiile au loc cu rate diferite, însă în continuare admite o frecvență egală a bazelor. Deci *a* = *c* = *d* = *f* =1 şi *b* = *e* = *k* obținându-se:

$$\Omega = \begin{pmatrix} -\frac{1}{4}\mu(k+2) & \frac{1}{4}\mu & \frac{1}{4}\mu k & \frac{1}{4}\mu \\ \frac{1}{4}\mu & -\frac{1}{4}\mu(k+2) & \frac{1}{4}\mu & \frac{1}{4}\mu k \\ \frac{1}{4}\mu k & \frac{1}{4}\mu & -\frac{1}{4}\mu(k+2) & \frac{1}{4}\mu \\ \frac{1}{4}\mu & \frac{1}{4}\mu k & \frac{1}{4}\mu & -\frac{1}{4}\mu(k+2) \end{pmatrix}$$

Dacă avem rata tranzițiilor de $\alpha = \mu k / 4$ şi cea a transversiilor de $\beta = \mu / 4$ matricea de mai sus poate fi rescrisă ca :

$$\Omega = \begin{pmatrix} -\alpha - 2\beta & \beta & \alpha & \beta \\ \beta & -\alpha - 2\beta & \beta & \alpha \\ \alpha & \beta & -\alpha - 2\beta & \beta \\ \beta & \alpha & \beta & -\alpha - 2\beta \end{pmatrix}$$

În matricea de mai sus $k = \alpha / \beta$ reprezintă un indice ce arată predominanța tranzițiilor asupra transversiilor. Dacă *k* = 1, adică $\alpha = \beta$ modelul K2P se reduce la cel JC.

Modelul K2P poate fi generalizat pentru a permite o frecvență de echilibru inegală (aşa numitul model HKY85, propus de Hasegawa et al. în 1985). Matricea ratelor de substituție a acestui model este dat de

$$\Omega = \begin{pmatrix} -\mu(k\pi_G + \pi_Y) & \mu\pi_C & \mu\pi_G & \mu\pi_T \\ \mu\pi_A & -\mu(k\pi_T + \pi_R) & \mu\pi_G & \mu\pi_T \\ \mu\pi_A & \mu\pi_C & -\mu(k\pi_A + \pi_Y) & \mu\pi_T \\ \mu\pi_A & \mu\pi_C & \mu\pi_G & \mu(k\pi_C + \pi_R) \end{pmatrix}$$

unde $\alpha = \mu, \beta = \mu k, \pi_R = \pi_A + \pi_G, \pi_Y = \pi_C + \pi_T$. Aceasta corespunde modelului GTR cu constrângerea ca *a* = *c* = *d* = *f* =1 şi *b* = *e* = *k*. Modelul JC poate de asemeni fi transformat pentru a permite frecvența inegală a bazelor azotate (modelul F81, propus de Felsenstein în 1981). Acest lucru se poate face punând k = 1 în matricea modelului HKY85 de mai sus :

$$\Omega = \begin{pmatrix} -\mu(\pi_G + \pi_Y) & \mu\pi_C & \mu\pi_G & \mu\pi_T \\ \mu\pi_A & -\mu(\pi_T + \pi_R) & \mu\pi_G & \mu\pi_T \\ \mu\pi_A & \mu\pi_C & -\mu(\pi_A + \pi_Y) & \mu\pi_T \\ \mu\pi_A & \mu\pi_C & \mu\pi_G & \mu(\pi_C + \pi_R) \end{pmatrix}$$

Felsenstein a mai propus ulterior (în 1984) o altă metodă ce ține cont de frecvența inegală a bazelor într-un model cu 2 parametri. Acest model împarte substituțiile în două componente : o rată *generală* de substituție ce produce toate tipurile de substituții, şi o rată *în grup* ce produce doar tranziții. Matricea acestui model poate fi obținută din cea a modelului GTR dacă *a* = *c* = *d* = *f* = 1, b = ($1 + K/\pi_R$), iar e = ($1 + K/\pi_Y$), unde *K* este parametrul ce arată raportul tranziții : transversii, $\pi_R = \pi_A + \pi_G, \pi_Y = \pi_C + \pi_T$, iar elementele din diagonală sunt calculate astfel încât să reprezinte suma negativă a celorlalte elemente de pe rândul respectiv.

$$\Omega = \begin{pmatrix} - & \mu\pi_C & \mu\pi_G(1 + K/\pi_R) & \mu\pi_T \\ \mu\pi_A & - & \mu\pi_G & \mu\pi_T(1 + K/\pi_Y) \\ \mu\pi_A(1 + K/\pi_R) & \mu\pi_C & - & \mu\pi_T \\ \mu\pi_A & \mu\pi_C(1 + K/\pi_Y) & \mu\pi_G & - \end{pmatrix}$$

Elementele din matricea de mai sus ce corespund tranzițiilor au fiecare câte două componente, deoarece tranzițiile au loc fie ratei generale fie ratei din grup. Când *K* = 0 acest model se rezumă la cel F81, când *K* creşte peste 0, tranzițiile au loc din ce în ce mai frecvent față de transversii. Acest model este implimentat în pachetul de programe PHYLIP 2.6 şi toate versiunile ulterioare (Felsenstein, 2004).

Aceste matrice ne dau doar rata instantanee a substituțiilor într-un moment de timp *dt*, însă în cazul construcției arborilor filogenetici după metoda maximei

verosimilitudini (ML) avem nevoie de probabilitățile schimbării oricărui caracter în oricare altul înt-un ram de lungime *t*. Matricea probabilității substituțiilor se calculează ca:

$$P(t) = e^{Qt}$$

Aceleaşi modele de mai sus pot fi folosite şi pentru a calcula distanța dintre două secvențe nucleotidice *X* şi *Y* care s-au diferențiat dintr-un strămoş comun *t* timp în urmă.

2.4.3. ESTIMAREA NUMĂRULUI DE SUBSTITUȚII NUCLEOTIDICE

2.4.3.1. MODELUL JUKES – CANTOR

Cel mai simplu model de substituție nucleotidică este cel propus de Jukes şi Cantor în 1969. Acest model presupune că substituția nucleotidică intervine în orice site cu frecvență egală, astfel încât, o nucleotidă este înlocuită cu una din celelalte trei nucleotide rămase cu o probabilitate *α* pe an (sau o altă unitate de timp). Mai mult, probabilitatea de înlocuire a unei nucleotide cu oricare din celelalte trei, este *r=3α*, unde r reprezintă probabilitatea de înlocuire pe site pe an. Dacă luăm în considerare două secvențe *X* şi *Y* care au evoluat divergent dintr-o secvență comună ancestrală, cu *t* ani în urmă. Dacă notăm cu *qt* nucleotidele comune dintre cele două secvențe şi cu *pt (pt=1 – qt)* proporția nucleotidelor diferite, procentul nucleotidelor comune *qt+1* în timpul *t+1* (măsurat în ani), poate fi obținut astfel – în primul rând, trebuie să precizăm că acest site care prezintă nucleotide comune pentru *X* şi *Y* în timpul *t*, va rămâne neschimbat în timp, adică în *t+1*, cu o probabilitate *(1-r)2* sau aproximativ *1-2r*, deoarece *r* reprezintă o cantitate foarte mică astfel încât *r2* poate fi neglijat. În plus, un site care prezintă nucleotide diferite în timpul *t*, va avea aceleaşi nucleotide în timpul *t+1* cu o probabilitate *2r/3*. Această probabilitate este obținută ținând cont de faptul că în momentul în care secvențele *X* şi *Z* au nucleotidele *i* respectiv *j*, în timpul *t*, ele devin identice dacă *i* în *X* se schimbă cu *j*, dar *j* din *Y* rămâne neschimbată sau dacă *j* din *Y* este înlocuită cu *i*, iar *i* din *X* rămâne neschimbată. Probabilitatea de apariție a primului eveniment este de *α(1-r)=r(1-r)/3*, deoarece *i* in *X* trebuie să fie înlocuită cu nucleotida *j* (cu o probabilitate mai mare decât cazul celorlalte două nucleotide), iar nucleotida *j* din *Y* trebuie să rămână neschimbată. Probabilitatea de apariție a celui de-al doilea eveniment este de asemenea *r(1-r)/3*. Mai mult, probabilitatea totală este *2r(1-r)/3*, care devine aproximativ *2r/3* dacă nu se ține cont de valoarea lui *r2*.

Dacă:

$q_{t+1} = (1-2r)q_t + \frac{2}{3}r(1-q_t)$, care poate fi scrisă ca:

$$q_{t+1} - q_t = \frac{2r}{3} - \frac{8r}{3}q_t$$

Dacă notăm diferența $q_{t+1} - q_t$ cu dq/dt, atunci vom abține:

$\frac{dq}{dt} = \frac{2r}{3} - \frac{8r}{3}q$; ținând cont de condiția inițială q=1 la t=0, atunci:

$$q = 1 - \frac{3}{4}\left(1 - e^{-8rt/3}\right).$$

În cazul prezentului model, numărul aşteptat de substituții nucleotidice pe site *(d)* pentru două secvențe este *2rt*. Din acest motiv, valoarea lui *d* este dată de:

$d = -(3/4)\ln[1-(4/3)p]$, unde: *p=1-q* este procentul nucleotidelor diferite dintre *X* şi *Y* (Jukes şi Cantor, 1969).

O estimare a valorii $(\hat{d})$ poate fi obținută prin înlocuirea lui *p* cu valoarea observată $(\hat{p})$; în acest caz, varianța este:

$V(\hat{d}) = \frac{9p(1-p)}{(3-4p)^2 n}$, (Kimura şi Ohta, 1972).

În cadrul acestui model, se presupune că rata substituției nucleotidelor este aceeaşi pentru fiecare pereche de nucleotide, astfel încât frecvențele aşteptate ale celor patru baze azotate A, T, C şi G sunt egale cu 0,25. Oricum, nefăcând nici o precizare referitoare la frecvența inițială, valoarea lui d rămâne aceeaşi; altfel spus, nu este necesară o constanță a valorii frecvențelor nucleotidice, pentru ca ecuația ce dă valoarea diferenței (d) să fie aplicabilă (Rzhetsky şi Nei, 1995).

2.4.3.2. MODELUL KIMURA CU 2 PARAMETRI

În cazul modelului lui Kimura cu 2 parametri se ea în considerare frecvența mai mare a tranzițiilor față de cea a translațiilor, rata de substituție *r* fiind astfel $\alpha + 2\beta$ şi nu 3α. Conform acestui model

$$P = (1/4)(1 - 2e^{-4(\alpha+\beta)t} + e^{-8\beta t})$$

$$Q = (1/2)(1 - e^{-8\beta t})$$

iar $d \equiv 2rt = 2\alpha t + 4\beta t = -(1/2)\ln(1-2P-Q) - (1/4)\ln(1-2Q)$

cu varianța $V(\hat{d}) = \frac{1}{n}\left[c_1^2 P + c_3^2 Q - (c_1 P + c_3 Q)^2\right]$

unde $c_1 = \frac{1}{1-2P-Q}, c_2 = \frac{1}{1-2Q}, c_3 = (c_1 + c_2)/2$

2.4.3.3. MODELUL TAJIMA - NEI

În modelul lui Tajima şi Nei frecvența în echilibru a nucleotidelor este considerată identică, iar $d = (1/2)\left[1 - \sum_{i=1}^{4} g_i^2 + p^2/c\right]$ unde $c = \sum_{i=1}^{3}\sum_{j=i+1}^{4} \frac{x_{ij}^2}{2g_i g_j}$ cu *xij* (*i* < *j*) fiind frecvența relativă a perechii de nucleotide *ij* din două secvențe de ADN comparate. În acest caz varianța lui $\hat{d}$ este dată de

$$V(d) = \frac{b^2 p(1-p)}{(b-p)^2 n}$$

Dacă rata de substituție a nucleotidelor este aceiaşi pentru toate perechile de nucleotide atunci *b* = ¾ în echilibru şi ecuația lui *d* se reduce la cea a metodei Jukes – Cantor, însă în practică *b* < ¾ şi metoda Tajima - Nei dă valori mai ridicate decât metoda Jukes – Cantor.

2.4.3.4. MODELUL TAMURA - NEI

Unul din modelele matematice ce dă concluziile de ordin filogenetic prin metoda maximum likelihood (ML) (metoda probabilității maxime), este cel al lui Hasegawa et al. (HKY) (1985). Acest model este un hibrid între modelul bazat pe 2 parametri al lui Kimura şi modelul inputului egal şi ia în considerare atât tranzițiile cât şi transversiile, precum şi valorile conținutului în GC. Oricum, formulele pentru valoarea lui $\left(\hat{d}\right)$ în cadrul acestui model sunt destul de complicate (Rzhetsky şi Nei, 1995), astfel încât nu ar trebui să nu-l luăm în considerare. În plus, acest model este inclus în modelul propus de Tamura şi Nei (1993), ca un caz special, permițând soluții analitice pentru valoarea distanței *(d)*. În acord cu acest model, formula pentru valoarea lui *d* este:

$$d = -\frac{2g_A g_G}{g_R}\ln\left[1 - \frac{g_R}{2g_A g_G}P_1 - \frac{1}{2g_R}Q\right] - \frac{2g_T g_C}{g_Y}\left[1 - \frac{g_Y}{2g_T g_C}P_2 - \frac{1}{2g_Y}Q\right] -$$
$$-2\left[g_R g_{Y-}\frac{g_A g_G g_R}{g_R} - \frac{g_T g_C g_Y}{g_Y}\right]\ln\left[1 - \frac{1}{2g_R g_Y}Q\right]$$

unde *P1* şi *P2* sunt proporțiile diferențelor de tranziții dintre A şi G, respectiv dintre T şi C, iar *Q* este proporția diferențelor datorate transversiilor (Tamura şi Nei, 1993). Atât formula pentru $\hat{d}$ cât şi pentru $V(\hat{d})$ sunt incluse în programele MEGA şi MEGA2 (Kumar et al., 2001).

2.4.3.5. MODELUL TAMURA

Un alt model, cel al lui Tamura din 1992 e deosebit de cel al lui Kimura prin aceea că ia în considerare faptul că frecvența celor patru nucleotide nu este egală. De exemplu la *Drosophila* sau la *Hymenoptera* ADN mitocondrial are un conținut mult mai ridicat în AT decât în GC. Deci acest model este un caz particular al celui al lui Kimura pentru cazurile în care conținutul de GC este foarte înalt sau foarte redus.

$$d = -h\ln(1 - P/h - Q) - (1/2)(1-h)\ln(1-2Q)$$

unde $h = 2\theta(1-\theta)$, iar θ este conținutul în GC. Această metodă de calculare a distanțelor dintre secvențe este inclusă în programele MEGA şi MEGA2 (Kumar et al., 2001).

Dacă considerăm un caz în care $n = \infty$ şi un set de date ce urmează modelul Tamura-Nei putem vedea cum se comportă diverse metode. În Figura 5 se observă bineînțeles că modelul Tamura-Nei se ajustează perfect datelor iar celelalte modele subestimează mai mult sau mai puțin numărul real de substituții. Diferențele sunt mari pentru d $\geq 0{,}6$, până la *d* = 0,5 modelul lui Tamura dând rezultate practic identice cu cel Tamura – Nei. Modelele Tamura, Kimura2P şi Jukes – Cantor dau aproape acelaşi rezultat cu modelul Tamura – Nei pentru *d* = 0,25, chiar şi distanța *p* devine foarte similară cu celelalte dacă $p \leq 0{,}1$. Deci în cazul analizării unor specii apropiate sau a unor gene foarte conservate cum ar fi citocromoxidaza I (COI) nu este necesară utilizarea metodelor prea complexe. În acest caz este mai indicată utilizarea unei metode simple pentru că are o varianță mai mică. De asemeni ca şi în cazul

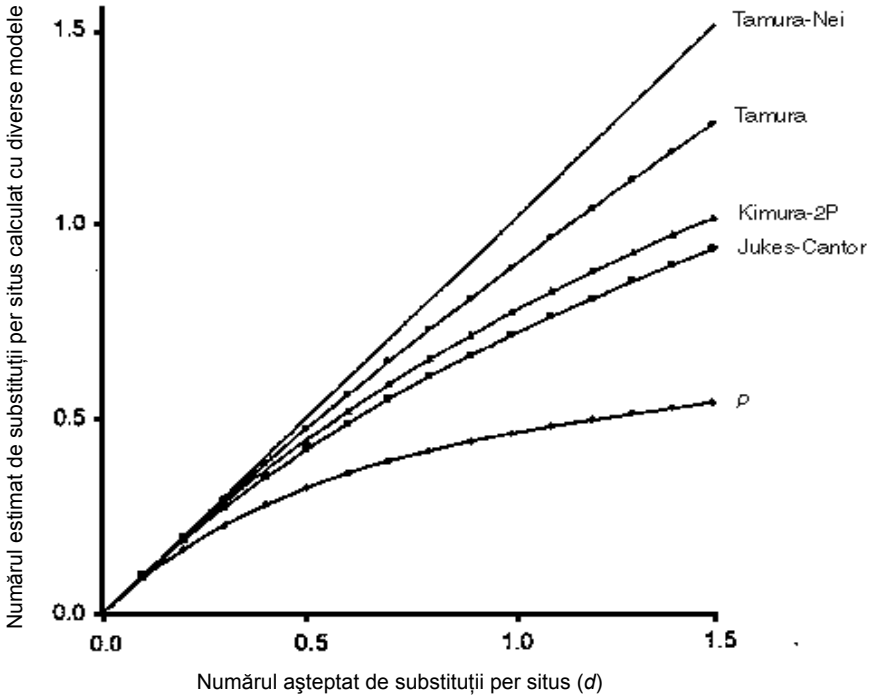


Figura 5 Numărul de substituții nucleotidice estimate folosind diferite metode (modelul actual al substituțiilor urmează modelul Tamura-Nei). Frecvențele nucleotidelor sunt g_A = 0,3; g_T =0,4; g_C = 0,2; g_G =0,1. Rata tranziții/transversii este $\alpha_1/\beta = 4$ şi $\alpha_2/\beta = 8$. (după Nei & Kumar, 2000)

secvențelor de aminoacizi şi în cazul celor de acizi nucleici poate fi utilizată distanța gamma pentru cazurile în care rata substituțiilor diferă substanțial în cadrul aceleiaşi secvențe. Mai uzuale sunt distanțele gamma pentru modelele Jukes-Kantor, Kimura 2P şi Tamura-Nei. Distanțele gamma sunt mai realiste dar au o varianță mai mare, din această cauză rezultatele obținute nu produc în mod obligatoriu rezultate mai bune în analizele filogenetice (doar dacă numărul de nucleotide folosite nu este foarte mare). Pentru estimarea lungimii ramificațiilor unui arbore filogenetic distanțele gamma dau de obicei rezultate mai bune.

2.4.3.6. DISTANȚE GAMMA

În cadrul distanțelor evolutive, rata de substituție a nucleotidelor, se presupune a fi aceeaşi pentru toate siturile nucleotidice. În realitate, această presupunere nu este pe deplin valabilă, deoarece rata substituției este diferită în situri diferite. În cazul genelor care codifică proteine, acest fapt este evident, deoarece prima, a doua şi a treia poziție a unui codon prezintă diferite rate de substituție. Existența zonelor pliate în zonele active ale proteinei, determină de asemenea rate diferite de substituție de-a lungul siturilor nucleotidice. Variația ratei de substituție poate fi observată şi în genele codificatoare din macromolecula de ARN, deoarece ARNs prezintă forțe de pliere şi adesea formează bucle şi porțiuni bicatenare la baza buclelor, care prezintă diferite rate de substituție. Analize statistice ale ratei de substituție *(r)* în diferite situsuri nucleotidice, au sugerat că aceasta variază urmând aproximativ distribuția gamma dată de ecuația:

$f(r) = \frac{b^a}{\Gamma(a)} e^{-br} r^{a-1}$, unde $a = \bar{r}^2 / V(r)$ şi $b = \bar{r}^2 / V(r)$, $\bar{r}$ şi $V(r)$ fiind media şi respectiv varianța ratei de substituție *r* (Johnson şi Kotz, 1970). $\Gamma(a)$ este funcția gamma, definită de formula:

$\Gamma(a) = \int_0^\infty e^{-t} t^{a-1} dt$. Forma distribuției $f(r)$ este dată de valoarea lui „a" care este adesea numită conformație gamma sau parametru gamma şi de valoarea lui *b* care este factor de mărime (Kocher şi Wilson, 1991; Tamura şi Nei, 1993; Wakeley, 1993, 1994).

Din acest motiv, unii autori dezvoltă distanțele gamma corespunzătoare, pentru substituția nucleotidelor. Distanțele gamma pot fi derivate ale aceleiaşi metode matematice, ca şi cele utilizate pentru calcularea distanțelor în cazul secvențelor de aminoacizi.

2.4.4. ESTIMAREA NUMERICĂ A DISTANȚELOR EVOLUTIVE

Formulele analitice prezentate mai sus sunt bazate pe modele matematice de substituție a nucleotidelor relativ simple, iar estimările sunt aceleaşi cu estimările prin metoda maximei verosimilitudini pentru acelaşi model. Dacă se doreşte însă efectuarea estimărilor pe baza unui model mai complicat cum ar fi Tamura-Nei gamma sau GTR este mai utilă estimarea numerică a distanțelor folosind metoda maximei verosimilitudini. Spre exemplu, formula analitică pentru estimarea distanței Tamura-Nei gamma necesită informații privind frecvența nucleotidelor şi a parametrului *a*. Frecvența nucleotidelor este estimată din două secvențe de ADN comparate, iar valoarea lui *a* este obținută utilizând secvențe adiționale.

Dacă utilizăm metode numerice este posibilă estimarea frecvenței nucleotidelor şi valoarea parametrului *a* din setul de date. Altfel spus putem estima atât toți parametrii cât şi distanțele maximizând verosimilitudinea pentru setul de date şi topologia

corespunzătoare. De asemeni este posibilă selecția celui mai apropriat model al substituțiilor folosind testul LRT (likelihood ratio test). Metoda numerică dă rezultate bune atâta timp cât numărul de nucleotide examinate *n* este foarte mare, iar topologia folosită este mai mult sau mai puțin corectă. Când *n* este mic distanțele estimate obținute astfel nu dau o estimare bună a topologiei arborelui.

2.5. ARBORI FILOGENETICI

Analizele filogenetice ale secvențelor de ADN sau proteine, au devenit un mijloc important pentru studiul istoriei evolutive a organismelor, de la bacterii la om. De când s-a constatat că rata de evoluție a ADN variază extensiv cu gena sau segmentul de ADN (Wilson et al., 1977; Dayhoff et al., 1978), se pot studia relațiile filogenetice existente între toate nivelele de clasificare ale organismelor (regnuri, filumuri, familii, genuri, specii şi populații intraspecifice), utilizând diferite gene sau secvențe de ADN. Analiza filogenetică este de asemeni importantă pentru elucidarea modelelor evolutive ale familiilor de gene (Atchley et al., 1994; Goodwin et al., 1996; Nei et al., 1997) ca şi pentru explicarea proceselor de evoluție adaptativă, la nivel molecular (Jermann et al., 1995; Chandrasekharan et al., 1996; Zhang et al., 1998).

Reconstrucția arborilor filogenetici utilizând metode statistice a fost inițiată independent în cadrul taxonomiei numerice, referitor la caracterele morfologice (Sokal şi Sneath, 1963) şi în genetica populațională pentru frecvența genelor (Cavalli – Sforza şi Edwards, 1964, 1967). Unele metode statistice dezvoltate în aceste scopuri sunt încă utilizate pentru analiza statistică a datelor de ordin molecular, dar , recent, numeroase metode noi au fost dezvoltate.

Relațiile filogenetice dintre organisme sau gene sunt de obicei prezentate sub formă de *arbori filogenetici* cu o rădăcină (Figura 6a) sau fără rădăcină (Figura 6b).

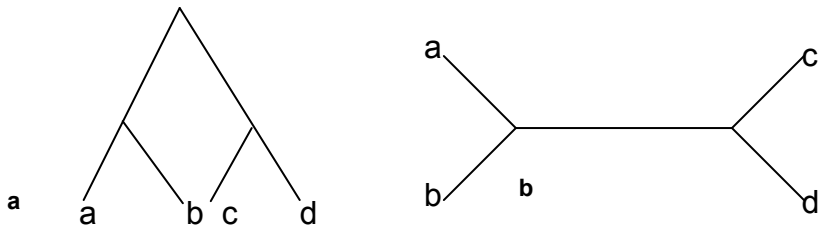


Figura 6 Cele două tipuri mari de arbori filogenetici: a-cu rădăcina, b-fără rădăcină

Modul în care se dispun ramurile într-un arbore filogenetic se numeşte *topologia arborelui*. Există un număr foarte mare de arbori (cu rădăcină sau fără) pentru un anumit număr de taxoni (orice tip de unitate taxonomică, fie ea familie, specie, populație, secvență de ADN sau aminoacizi).

Dacă numărul taxonilor (*m*) este 4, există 3 topologii posibile pentru arbori fără rădăcină şi 15 pentru arbori cu rădăcină. Numărul posibil de topologii creşte brusc odată cu creşterea lui *m*. Numărul de topologii posibile pentru un arbore cu rădăcină ce prezintă doar bifurcații (rezolvat complet) este

$$1 \cdot 3 \cdot 5 \cdots (2m-3) = \frac{[(2m-3)!]}{[2^{m-2}(m-2)!]}$$

pentru $m \geq 2$. Deci numărul posibil de topologii pentru *m* = 2, 3, 4, 5 şi 6 este de 1, 3, 15, 105 şi 945 respectiv. Când *m* = 10 el devine 34.459.425, doar una dintre aceste topologii fiind cea adevărată! Pentru a calcula numărul total de topologii pentru un arbore fără rădăcină înlocuim în formula de mai sus *m* prin *m* – 1.

Într-un arbore filogenetic fără rădăcină de *m* taxoni există 2*m* – 3 *ramificații*, *m* *ramificații externe* şi deci *m* – 3 *ramificații interne*. Numărul de *noduri interne* este de *m* – 2. Pentru un arbore cu rădăcină numărul de ramificații interne este de *m* – 2, numărul de noduri interne este de *m* – 1, iar numărul total de ramificații este de 2*m* – 2.

Teoretic o secvență de ADN se împarte dihotomic la fiecare speciație sau duplicație de gene, arborii filogenetici fiind de obicei bifurcați. Totuşi când se studiază o secvență relativ scurtă unele ramificații interne pot să nu prezinte nici o substituție putând apărea astfel arbori pluriramificați. Majoritatea metodelor folosite pentru deducerea arborilor filogenetici construiesc arbori bifurcați, dar un arbore obținut poate fi transformat într-unul pluriramificat eliminând orice ramificație cu lungime 0. Este posibil de asemeni ca chiar dacă adevărata filogenie este un fenomen dihotomic arborele reconstruit să apară pluriramificat datorită erorilor statistice. În realitate cele două cazuri sunt practic imposibil de deosebit.

2.5.1. ARBORI AI GENELOR ŞI ARBORI AI SPECIILOR

De cele mai multe ori suntem interesați să obținem un arbore filogenetic care să reprezinte istoria evolutivă a speciilor sau populațiilor. Acest tip de arbore este denumit arbore al speciilor sau al populațiilor. Într-un arbore al speciilor momentul divergenței dintre două specii corespunde cu apariția izolării reproductive. Totuşi când se construieşte un arbore filogenetic pornind de la o genă de la fiecare specie, arborele obținut poate să nu fie în concordanță cu filogenia speciilor. În cazul existenței alelelor polimorfe divergența genelor are loc anterior separării populațiilor (Figura 7)

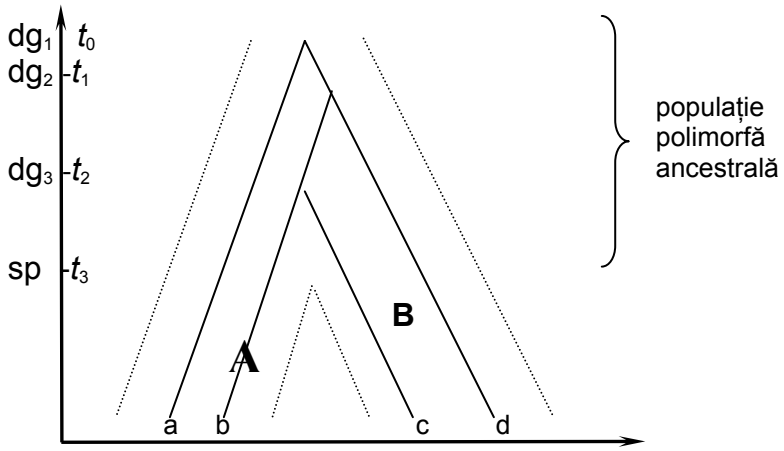


Figura 7 Diagramă ce arată că în cazul existenței polimorfismului, momentul separării genelor este anterior separării populațiilor

Când genele luate în studiu aparțin unei familii multigenice poate să apară altă problemă. Dacă spre exemplu două specii înrudite 1 şi 2 au două gene duplicate, a1 b1 şi a2 b2 respectiv, iar duplicarea a avut loc înaintea separării celor două specii atunci genele a1 şi a2 sunt *ortoloage*, adică una este descendentul direct al celeilalte, iar perechile a1 cu b1, a2 cu b2, a1 cu b2 şi a2 cu b1 sunt *paraloage*, adică nu sunt descinse direct una din cealaltă. Când construim un arbore filogenetic al speciilor trebuie să utilizăm doar gene ortoloage şi nu paraloage, deoarece doar primele reprezintă fenomenele de speciație. diferențierea între cele două tipuri de gene nu este întotdeauna simplă, mai ales în cazul familiilor multigenice, de aceia deducerea filogeniei speciilor din cea a genelor trebuie făcută mereu cu multă grijă.

Bineînțeles că nu întotdeauna se caută a obține un arbore al speciilor. În studiul familiilor multigenice se caută cunoaşterea istoriei evolutive a membrilor familiei şi procesul duplicării genelor.

O altă problemă în reconstruirea arborilor filogenetici este că de multe ori lungimea secvențelor utilizate este mică, apărând erori statistice. O filogenie teoretică construită dintr-un set de date cu lungime foarte mare şi un număr aşteptat de substituții se numeşte *arbore real*, iar cel din secvențele disponibile cu numărul real de substituții, *arbore realizat*, şi unul şi celălalt fiind diferiți de fapt de *arborele reconstituit* sau *dedus*. Majoritatea modelelor şi metodelor de reconstrucție a arborilor filogenetici au tendința de a produce filogeniile realizate, iar UPGMA pe cea reală. Din păcate această din urmă metodă este foarte sensibilă la erori şi dimensiunea setului de date. Desigur că cel mai util este arborele real, cel ce reprezintă evoluția istorică, însă în practică este mult mai simplu să obținem arborele realizat, deoarece setul de date se referă anume la acest tip de arbore. De asemeni topologia unei filogenii realizate este aceiaşi cu cea a celei reale, exceptând cazurile în care prima devine cu noduri pluriramificate datorită erorilor statistice (unele ramificații interne pot să nu prezinte

nici o substituție). Odată cu creşterea numărului de nucleotide utilizate arborele realizat se apropie de cel real.

Când se construieşte o filogenie a speciilor sau populațiilor arborele real trebuie să aibă lungimea ramificațiilor proporțională cu timpul de evoluție, iar două linii evolutive descendente din acelaşi nod intern trebuie să aibă aceiaşi lungime. În realitate însă rata de evoluție a genelor este supusă erorilor randomice cât şi anumitor tipuri de presiune selectivă, astfel încât în unele linii gena poate evolua mai rapid, iar în altele mai încet, arborele realizat nereuşind să producă tipul de arbore filogenetic descris mai sus.

2.5.2. METODE DE RECONSTRUIRE A ARBORILOR FILOGENETICI

Metodele utilizate la ora actuală pentru reconstruirea arborilor filogenetici pot fi împărțite în trei mari categorii:

1. metode bazate pe distanțe
2. metode MP (Maximum Parsimony)
3. metode ML (Maximum Likelyhood)

Reconstrucția unui arbore filogenetic este de obicei privită ca deducerea statistică a adevăratului arbore filogenetic care este necunoscut. Acest proces poate fi subâpărțit în două etape şi anume estimarea topologiei şi estimarea lungimii ramificațiilor. Când topologia este cunoscută estimarea lungimii ramificațiilor este relatv simplă, utilizându-se metode ca cea a pătratelor minime sau ML. Cum pentru un număr dat de secvențe există foarte multe topologii posibile au fost dezvoltate metode de optimizare a căutării cum ar fi cea a maximei verosimilitudini (ML) sau a evoluției minime (ME). Pentrun secvențe mici şi un număr de specii mare aceste metode tind însă să dea topologii greşite.

2.5.2.1. METODE BAZATE PE DISTANȚE

În cadrul metodelor bazate pe distanță sau pe matrixul de distanță, acestea sunt calculate pentru toate perechile de taxoni, iar arborii filogenetici sunt construiți ținând cont de relațiile stabilite în jurul valorilor acestor distanțe.

În acest caz, direcțiile filogenetice sunt construite considerând matrici ale distanțelor în care sunt calculate distanțele între secvențe luate două câte două. Există un număr foarte mare de metode prin care se pot estima direcțiile filogenetice pornind de la distanțe.

UPGMA (Unweighted Pair-Group Method using Arithmetic Averages)

Numele metodei provine de la „unweighted pair-group method using arithmetic averages" (UPGMA). Metoda este atribuită autorilor Sokal şi Michener (1958), dar s-a constatat ulterior că prezenta metodă este complet diferită de cea utilizată de aceştia. În schimb, algoritmul apare publicat de către Sneath şi Sokal (1973). Un arbore construit pe baza acestei metode este numit şi fenogramă, deoarece a fost inițial folosit la reprezentarea gradului de similaritate fenotipică pentru un grup de specii, în cadrul taxonomiei numerice. Oricum, poate fi folosit în filogenia moleculară atunci când rata de substituție a genelor este mai mult sau mai puțin constantă. În special în

cazul în care frecvența genelor este folosită în reconstrucția filogenetică, acest model produce arbori destul de buni, comparativ cu alte metode bazate pe distanță (Nei et al., 1983; Takezaki şi Nei, 1996). În acest caz, o măsură a distanței care prezintă un coeficient mai mic de variație, pare să ducă la construirea de arbori cu o mai mare precizie, comparativ cu alte metode de determinare a distanței, chiar dacă nu respectă exact numărul de substituții genice (Takezaki şi Nei, 1996). Prin UPGMA se pot construi arbori pentru evoluția în cadrul speciilor, cu toate că erori de topologie apar adesea când rata substituției genelor nu este constantă sau când numărul de gene sau nucleotide este mic.

În cadrul UPGMA, un mod de măsurare a distanțelor evolutive, este calcularea acestora pentru toate perechile de taxoni sau secvențe, valorile distanțelor fiind sub forma unei matrici (Tabelul 4)

Tabel 4 – Prima matrice a distanțelor (Nei şi Kumar, 2000)

Taxon	1	2	3	4
2	*d12*	-	-	-
3	*d13*	*d23*	-	-
4	*d14*	*d24*	*d34*	
5	*d15*	*d25*	*d35*	*d45*

Din această matrice, rezultă că *dij* reprezintă distanța dintre taxonii *i* şi *j*. Analiza grupată (de tip clustal) a taxonilor începe cu perechea de taxoni între care există cea mai mică distanță. Presupunând că *d12* este cea mai mică distanță dintre toate valorile matricei rezultate, taxonii 1 şi 2 vor fi grupați pornind dintr-un ram comun localizat la distanța *b=d12/2*. Din acestea, putem presupune că lungimea ramurilor celor doi taxoni care provin dintr-un ram comun, este aceeaşi. În acest caz, taxonii 1 şi 2 vor fi încadrați într-un singur taxon compus sau grup *[u=(1-2)]*, iar distanța dintre acest grup şi un alt taxon *k (k≠1,2)* este dată de: *duk=(d1k+d2k)/2*. În plus, obținem o nouă matrice (Tabel 5)

Tabel 5 A doua matrice a distanțelor (Nei şi Kumar, 2000)

Taxon	*u=(1-2)*	3	4
3	*du3*	-	-
4	*du4*	*d34*	-
5	*du5*	*d35*	*d45*

În continuare, presupunând că distanța *du3* este cea mai mică, taxonii *u* şi 3 vor fi încadrați într-un nou taxon compus sau cluster *[v=(1-2-3)]*, cu o lungime a ramului comun de *b=du3/2=(d13+d23)/(2x2)*. Această distanță între noul cluster (grup) creat *v* şi fiecare dintre taxonii rămaşi *(k)*, este dată de: *dkv=(dk1+dk2+dk3)/3*. Atunci matricea se va transforma în (Tabelul 6):

Tabel 6 A treia matrice a distanțelor

Taxon	*v=(1-2-3)*	4
4	*dv4*	-
5	*dv5*	*d45*

Presupunând că *dv4* este cea mai mică valoare din matricea anterioară, putem încadra *v=(1-2-3)* şi 4 ca având un ram comun dat de *b=dv4/2=(d14+d24+d34)/(3x2)*.

Este evident că ultimul taxon care se alătură celorlalți trei, este 5 şi ramul comun este dat de: *b=(d15+d25+d35+d45)/(4x2)*.

Este de asemenea posibil, ca în cea de-a doua matrice cea mai mică distanță poate fi *d45* (sau oricare alta în afară de *du3*). În acest caz, taxonii 4 şi 5, sunt uniți cu ramura comună dată de *b=d45/2*, creându-se astfel un nou taxon compus, *v=(4-5)*. Distanțele între *v* şi alt taxon (3 şi *u*) sunt date de *d3v=(d34+d35)/2* şi *duv=(d14+d15+d24+d25)/4*. Presupunând că *duv* este cel mai mic, taxonii u şi v sunt grupați iar taxonul 3 va fi ultimul care se va alătura grupului. De asemenea, dacă *d3v* este cel mai mic, taxonii 3 şi v se vor grupa primii.

După cum reiese din exemplul anterior, distanța dintre două grupuri (*A* şi *B*) este dată de următoarea formulă:

$d_{AB} = \sum_{ij} d_{ij} / (rs)$, unde *r* şi *s* sunt numerele de taxoni în grupele *A* şi *B*, iar d_{ij} este distanța dintre taxonul *i* din grupul *A* şi taxonul *j* din grupul *B*. Ramura comună dintre cele două grupe, este dată de *dAB/2*. Cu toate acestea, în scopul modelării pe calculator, ecuația de mai sus nu este utilizabilă, fiind folosiți algoritmi mai facili pentru a calcula *dAB* (Swofford et al., 1996).

METODA PĂTRATELOR MINIME (LS – Least Squares)

Când rata de substituție a nucleotidelor variază de la un taxon la altul metoda UPGMA dă adeseori filogenii cu topologie greşită. În acest caz poate fi utilizată metoda pătratelor minime.

În cazul metodei pătratelor minime obişnuite, se ia în considerare următoarea sumă reziduală a pătratelor

$$S_R = \sum_{i<j} (d_{ij} - e_{ij})^2$$

unde *dij şi eij* sunt distanțele observată şi respectiv patristică dintre taxonii *i* şi *j*. Distanța patristică este suma lungimii estimate a ramificațiilor ce unesc cei doi taxoni. *SR* este calculată pentru toate topologiile plauzibile şi topologia cu cea mai mică valoare este aleasă drept filogenie finală. O altă posibilitate de a calcula *SR* este aplicând formula :

$$S_R = \sum_{i<j} \left[(d_{ij} - e_{ij})^2 / d_{ij} \right]$$

În practică cele două formule dau rezultate identice sau aproape identice. O problemă ce se datorează criteriului de optimizare al acestei metode este obținerea frecventă a ramificațiilor cu lungime negativă (ceia ce din punct de vedere biologic este imposibil). Felsenstein (1995) a îmbunătățit metoda LS prin aplicarea constrângerii de a considera doar topologiile cu ramificații nenegative.

2.5.2.2. METODA EVOLUȚIEI MINIME (ME)

În cadrul acestei metode, suma lungimilor tuturor ramurilor estimate este:

$$S = \sum_{i}^{T} \hat{b}_i$$

fiind calculată pentru toate topologiile posibile, cea corectă, fiind topologia cu cea mai mică valoare a lui *S*. În acest caz, $\hat{b}_i$ este o estimare a lungimii ramului *i*, iar *T* este numărul total de ramuri *T=2m-3*.

Ideea metodei minimei evoluții (minimum evolution ME) a fost prima dată promovată de Edwards şi Cavalli-Sforza (1963), fără a da nici o explicație asupra algoritmului. Mai târziu, Kidd şi Sgaramella-Zonta (1971) au sugerat că lungimea totală a ramurilor *[L(S)]* poate fi calculată pe baza valorii absolute ($|\hat{b}_i|$) a tuturor lungimilor estimate ale ramurilor, fără nici o justificare teoretică. Dar, $L(\hat{S})$ nu prezintă proprietăți statistice care să permită un calcul rapid al valorii lui S, iar testele statistice dezvoltate de Rzhetsky şi Nei (1992, 1993)nu sunt aplicabile pentru *L(S)*. De menționat este faptul că în prezența erorilor estimate statistic ale ramurilor cu lungime mică, poate deveni negativă chiar şi pentru o topologie corectă a arborelui (Sitnikova et al., 1995).

ME presupune o examinare a tuturor topologiilor posibile şi găsirea celei cu cea mai mică valoare *S* până la epuizarea tuturor variantelor, iar arborele cu cea mai mică valoare *S*, este considerat ca fiind un arbore filogenetic ME. Baza teoretică a acestei strategii, este faptul că arborii filogenetici ME sunt în general identici sau asemănători cu arborii Neighbor – Joining (NJ), când valoarea lui *m* este relativ mică (Saitou şi Imanishi 1989; Rzhetsky şi Nei, 1992), astfel încât arborii de tip NJ pot fi folosiți ca punct de plecare când valoarea lui *m* este mare.

2.5.2.3. METODA NEIGHBOR JOINING (NJ)

Deşi metoda ME are proprietăți statistice utile, necesită o perioadă de calcul suficient de mare atunci când numărul taxonilor comparați este mare. Saitou şi Nei (1987) au dezvoltat o metodă eficientă de realizare a arborilor filogenetici, care se bazează pe principiul evoluției minime (ME). Această metodă nu examinează toate topologiile posibile, dar în fiecare stadiu de grupare a taxonilor, se foloseşte un principiu de evoluție minimă. Această metodă se numeşte Neighbour –Joining (NJ) şi este privită ca o versiune simplificată a metodei ME. Când se folosesc 4 din 5 taxoni, metodele NJ, ME dau rezultate identice (Saitou şi Nei, 1987). Există o oarecare asemănare între NJ şi metoda adițională de realizare a arborilor, care dă atât topologia cât şi lungimea ramurii simultan.

Unul dintre conceptele importante în această metodă, este reprezentat de „neighbors”, definiți ca doi taxoni conectați printr-un singur nod într-un arbore fără punct de origine. De exemplu, taxonii 1 şi 2 din arborele prezentat în Figura 37, sunt considerați „neighbors” (vecini), pentru că sunt conectați doar prin nodul A. Similar,

taxonii 5 şi 6 sunt consideraţi „neighbors”, dar toate celelalte perechi nu. Cu toate acestea dacă se combină taxonii 1 şi 2 şi se consideră ca fiind un singur taxon, acesta, (1-2) şi taxonul 3 sunt acum „neighbors”. Este posibilă definirea topologiei unui arbore prin alăturări succesive ale taxonilor vecini (nj) şi producerea unor noi perechi de taxoni vecini. De exemplu, topologia arborelui din Figura 8 poate fi descrisă prin următoarele perechi de taxoni vecini („neighbors”): (1, 2), (5, 6), (1 – 2, 3) şi (1 – 2 – 3,4). Astfel, prin găsirea acestor perechi de taxoni vecini, se poate obține topologia arborelui.

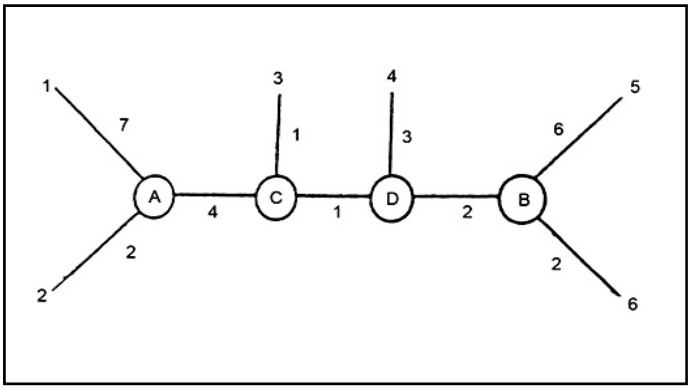


Figura 8 – Arborele filogenetic cu lungime ştiută a ramurilor, pentru şase secvențe (Nei şi Kumar, 2000)

Construirea unui arbore prin intermediul acestei metode, începe cu arborele în formă de stea care este produs, pe baza presupunerii că nu există o grupare a taxonilor (Figura 9). În practică, această presupunere este în general incorectă. Astfel, dacă se estimează lungimea ramului unui arbore în formă de stea şi se calculează suma tuturor ramurilor *(S0)*, această sumă ar trebui să fie mai mare decât suma *(SF)* pentru arborele de tip NJ final. Dacă se elimină taxonii vecini 1 şi 2 din cadrul arborelui prezentat în Figura 10, suma *(S12)* a tuturor lungimilor ramurilor, trebuie să fie mai mică decât *S0*, cu toate că este posibil să fie mai mare decât *SF*. În practică, deoarece nu se ştie exact care perechi de taxoni sunt vecini, se consideră toate perechile de taxoni, ca potențiale perechi de taxoni vecini şi se calculează suma lungimilor ramurilor *(Sij)* pentru taxonii *i* şi *j*, utilizând o topologie similară cu cea din Figura 39, se pot alege taxonii *i* şi *j* care au cea mai mică valoare pentru *Sij* Desigur, valorile distanțelor sunt subiectul erorilor stochastice, astfel încât taxonii vecini aleşi în acest mod, nu sunt întotdeauna adevărații taxoni vecini. O dată identificată o pereche de taxoni vecini, aceşti sunt încadrați într-un taxon compus, procedura repetându-se până la producerea arborelui final.

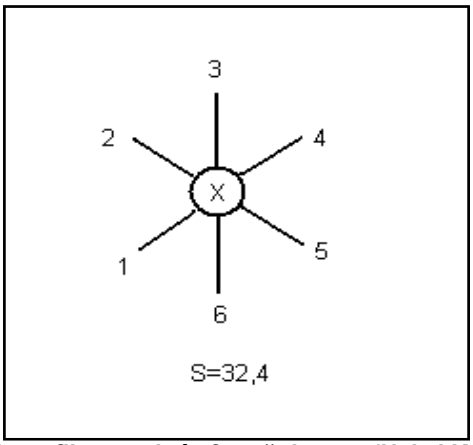


Figura 9 Arbore filogenetic în formă de stea (Nei şi Kumar, 2000)

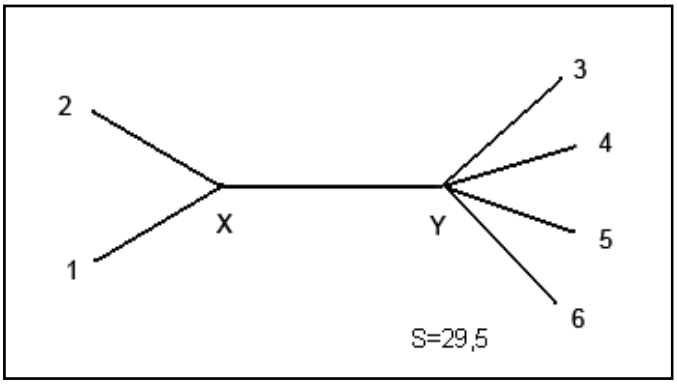


Figura 10 Separarea primilor taxoni vecini (1 şi 2), (Nei şi Kumar, 2000)

Matematic, valoarea sumei S0 pentru un arbore în formă de stea este:

$S_0 = \sum_{i=1}^{m} L_{iX} = \frac{1}{m-1} \sum_{i \prec j}^{m} d_{ij} = T/(m-1)$, unde L_{iX} este lungimea brațului estimat între nodurile *i* şi *X*, iar *T=Σi<jdij*. În acest caz, *i* reprezintă un anumit nod exterior, iar *X* nodul interior (Figura 9). În situația opusă, aşa cum este în cazul Figurii 10, din care reiese faptul că suma *S12* este dată de *L1X+L2X, LXY* şi $\sum_{i=3}^{m} L_{iZ}$. Dar,

$L_{1X} + L_{2X} = d_{12}$,

$$L_{XY} = \frac{1}{2(m-2)} \left[\sum_{i=3}^{m} (d_{1i} + d_{2i}) - (m-2)(L_{1X} + L_{2X}) - 2\sum_{i=3}^{m} d_{iY} \right]$$

$$\sum_{i=3}^{m} L_{iY} = \frac{1}{m-3} \sum_{3 \le i \prec j} dij$$

Din care, rezultă:

$$S_{12} = L_{1X} + L_{2X} + L_{XY} + \sum_{i=3}^{m} L_{iY} = \frac{1}{2(m-2)} \sum_{i=3}^{m} (d_{1i} + d_{2i}) + \frac{1}{m-2} \sum_{3 \le i \prec j} d_{ij}$$

Dacă notăm cu $R_1 = \sum_{i=3}^{m} d_{1i}$ şi $R_2 = \sum_{i=3}^{m} d_{2i}$, atunci S12 poate fi exprimată:

$$S_{12} = \frac{2T - R_1 - R_2}{2(m-2)} + \frac{d_{12}}{2}$$

Desigur, *Sij* poate fi calculată în acelaşi mod dacă înlocuim *1* şi *2* cu *i* şi *j* în ecuația de mai sus. Ultima ecuație necesită mult mai puțin timp de calcul decât cea anterioară. Mai mult, dacă *T* este acelaşi pentru toate perechile de tipul *i* şi *j*, *Sij* poate fi înlocuit cu:

$$Q_{ij} = (m-2)d_{ij} - R_i - R_j$$

în scopul de a calcula valoarea relativă a lui *Sij*.

O dată cu determinarea celei mai mici valori a lui *Sij*, poate fi creat un nou nod (A), care permite conectarea taxonilor *i* şi *j*. Lungimile brațelor (*bAi* şi *bAj*) din acest nod până la taxonii *i* şi *j* sunt date de:

$$\begin{matrix} b_{Ai} \frac{1}{2(m-2)} \left[(m-2)d_{ij} + R_i - R_j \right] \\ b_{Aj} \frac{1}{2(m-2)} \left[(m-2)d_{ij} + R_i - R_j \right] \end{matrix}$$, (Saitou şi Nei, 1987)

Aceste valori sunt determinate prin metoda pătratului minim, pentru actuala topologie a arborelui. Următorul pas, este calcularea distanței între noul nod A şi taxonii rămaşi (k; 3≤k≤m) (Figura 11).

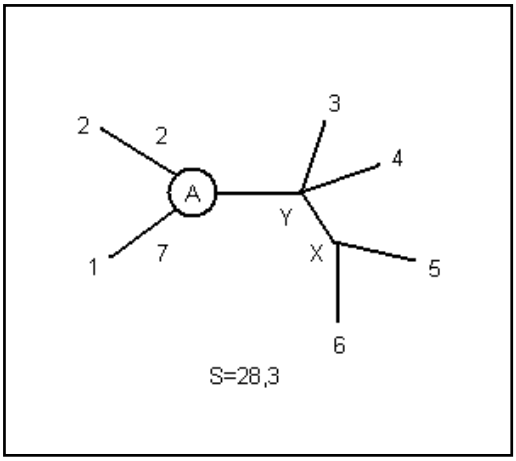


Figura 11 Separarea nodului A de restul taxonilor, (Nei şi Kumar, 2000)

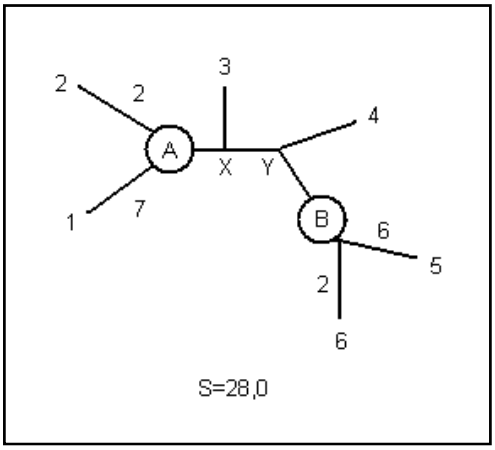


Figura 12 Separarea taxonilor 5 şi 6 prin formarea nodului B,
(Nei şi Kumar, 2000)

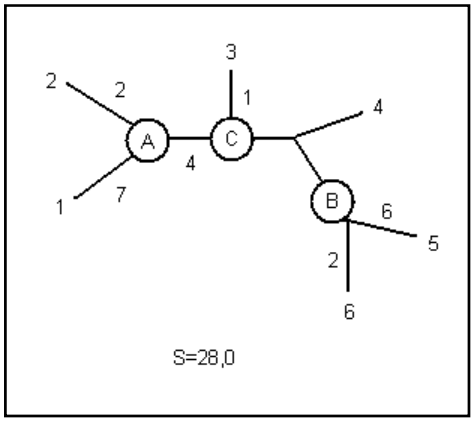


Figura 13 Separarea taxonului 3 prin apariția nodului C,
(Nei şi Kumar, 2000)

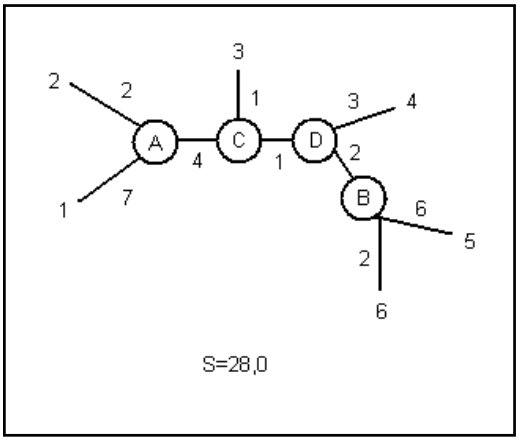


Figura 14 Separarea taxonului 4 prin apariția nodului D,
(Nei şi Kumar, 2000)

Această distanță este dată de:

$$d_{Ak} = \left(d_{ik} + d_{jk} - d_{ij}\right)/2$$

Dacă se calculează toate distanțele pe baza acestei valori, vom obține o nouă matrice de tipul *(m-1)(m-1)*. Pentru această nouă matrice, putem calcula o nouă sumă *Sij*, care va fi notată cu *Sij`*, deoarece nu include lungimea ramurilor externe pentru prima pereche de taxoni vecini identificați, astfel încât apare ca fiind mai mică decât suma totală *(Sij)* a lungimii brațelor la acest nivel al construcției arborelui filogenetic. Pentru găsirea unei noi perechi de taxoni vecini, se ia în considerare din nou perechea cu cea mai mică valoare *Sij*. Un nou nod B poate fi creat pentru această nouă pereche de taxoni şi o nouă valoare a matricei distanței *(m-2)(m-2)* este calculată. Această procedură se repetă, până când toți taxonii sunt grupați într-un singur arbore fără punct de origine (unrooted tree) de tip neighbour joining (Figurile 12 - 14).

2.5.2.4. ALEGEREA TIPULUI DE DISTANȚĂ EVOLUTIVĂ FOLOSITĂ ÎN RECONSTRUCȚIILE FILOGENETICE

Acuratețea unei filogenii deduse pe bază de secvențe depinde de doi factori : (1) relația liniară dintre distanța estimată şi numărul real de substituții şi (2) eroarea standard a coeficientului de variație a distanței utilizate. Până în momentul de față nu a fost pus la punct nici un test statistic general valabil pentru a putea alege modul de estimare a distanței evolutive. Totuşi în urma estimărilor pe calculator şi a datelor experimentale s-a ajuns la o serie de generalizări utile:

1. Când numărul de substituții nucleotidice estimat prin metoda JC este $\leq 0{,}05$ se recomandă utilizarea distanței *p* sau distanța JC chiar dacă există diferență între numărul de transversii şi tranziții sau dacă rata de substituție *r* variază de la situs la situs. În acest caz utilizarea modelului K2P sau a altor modele mai complexe dă acelaşi rezultat însă aceste metode au o varianță mai mare datorită numărului mai mare de parametri.

2. Când 0,05<*d*<1.0, iar numărul de nucleotide examinate este mare se recomandă utilizarea distanței JC. Dacă raportul *R* dintre tranziții şi transversii este mai mare ca 5 iar numărul de nucleotide examinate este mare este recomandată folosirea modelului K2P sau distanța gamma. Dacă însă numărul de secvențe folosite este mare, iar *n* este mic atunci distanța *p* dă cele mai bune rezultate (doar dacă rata de substituție nu variază foarte mult pentru diverse linii filogenetice). În ultimii ani s-a observat o utilizare frecventă a estimării gamma a modelului HKY, deoarece acesta ține cont de conținutul în perechi GC, raportul *R* dintre tranziții şi transversii şi variația ratei de substituție între situsuri. Datorită varianței mari modelul HKY dă însă de obicei rezultate mai proaste decât utilizarea distanțelor *p* sau JC. Doar în cazurile extreme când *n* > 10000, iar rata de substituție a nucleotidelor variază foarte mult de la o linie evolutivă la alta utilizarea modelelor complexe cum ar fi HKY este îndreptățită.

3. Când *d* > 1 pentru mai multe perechi de secvențe arborele filogenetic construit este de obicei greşit (din cauza varianței mari a distanțelor estimate şi a erorilor de aliniere). Acest tip de seturi de date trebuiesc ocolite pe cât este posibil eliminând porțiunile slab conservate folosindu-le pe cele mai puțin variabile (ca în cazul

imunoglobulinelor unde regiunile variabile nu sunt utilizate în construirea arborilor genelor).

4. În cazul utilizării secvențelor ce codifică proteine distincția între substituțiile sinonime (*ds*) şi cele nesinonime (*dn*) poate fi utilă deoarece rata substituțiilor sinonime este mai mare datorită faptului că acestea sunt neutre din punct de vedere selectiv. Astfel utilizând *ds* se poate evita diferența dintre rata de substituție între situsurile sinonime şi cele nesinonime. Acest lucru se poate face doar dacă *ds* < 0,5. Pentru specii mai îndepărtate filogenetic *dn* dă rezultate mai bune deoarece *ds* sau *ps* pot da estimări greşite datorită saturării rapide cu mutații a situsurilor.

5. Când mai multe tipuri de estimări a distanțelor evolutive dau rezultate similare este recomandată utilizarea celui mai simplu model de substituție deoarece cele mai complexe au o varianță mai mare. Când rata de substiție a nucleotidelor este aproapre identică pentru toate liniile evolutive şi aportul *R* dintre tranziții şi transversii nu este mare distanța *p* pare să dea topologii corecte mai frecvent decât alte metode, chiar dacă divergența dintre secvențe este mare.

2.5.2.5. METODA MAXIMUM PARSIMONY (MP)

Metodă inițial elaborată pentru caracterele morfologice, prezintă mai multe versiuni (Wiley, 1981; Felsenstein, 1982; Wiley et al., 1991; Maddison şi Maddison, 1992; Swofford şi Begle 1993). Eck şi Dayhoff în 1996 par a fi fost primii care au utilizat această metodă pentru realizarea arborilor, pornind de la secvențe de aminoacizi. Fitch (1971)şi Hartigan (1973) au elaborat un algoritm mai riguros al metodei MP pentru secvențe de nucleotide. În cadrul acestei metode se iau în considerare 4 sau mai multe secvențe aliniate de nucleotide sau aminoacizi *(m≥4)*, iar nucleotidele (aminoacizii) taxonilor ancestrali sunt luate în considerare separat pentru fiecare site şi o anumită topologie, presupunând că modificările produse de mutații apar la nivelul oricăreia dintre cele patru nucleotide (sau 20 de aminoacizi). Se calculează apoi, cel mai mic număr de substituții nucleotidice (sau de aminoacizi) care explică întregul proces evolutiv pentru o anumită topologie. Acest calcul este realizat pentru toate topologiile presupuse a fi corecte şi topologia care implică cel mai mic număr de substituții este aleasă ca reprezentând arborele corect. Cu cât este implicat mai puțin procesul evolutiv de creare a unei inferențe filogenetice, cu atât putem obține concluzii mai exacte.

Dacă nu există substituții în paralel sau în sens invers (nu există omologie) la fiecare site nucleotidic şi numărul de nucleotide examinate *(n)* este foarte mare, ar trebui ca utilizând MP să se obțină arborele real. Cu toate acestea, în practică, secvențele de nucleotide suferă adesea substituții în paralel sau în sens invers, şi *n* este relativ mic. În acest caz, MP tinde să confere o topologie incorectă. În plus, Felsenstein (1981) a arătat că atunci când rata substituției nucleotidelor variază proporțional cu direcția evolutivă, MP poate genera topologii incorecte chiar şi atunci când se examinează un număr foarte mare de nucleotide. În anumite condiții, aceasta se poate întâmpla chiar şi atunci când rata de substituție este constantă pentru toate direcțiile evolutive (Hendy şi Penny, 1989; Takezaki şi Nei, 1996). În acest caz, ramurile lungi (sau cele scurte) ale arborelui real, tind să se alăture, sau se atrag în cadrul arborelui final. Astfel, acest fenomen, este adesea numit „long branch attraction" (*engl.* atracția dintre ramurile lungi) (Hendy şi Penny, 1989) sau „Short branch attraction" (*engl*. atracția dintre brațele scurte) (Nei, 1996). În analiza parsimoniei este de asemenea dificilă abordarea inferenței filogenetice într-un cadru

statistic, deoarece nu există o cale de calcul a mediei şi varianței numărului minim de substituții obținute pe criteriul parsimoniei.

Cu toate acestea, MP prezintă avantaje comparativ cu alte metode de realizare a arborilor. În primul rând, nu implică folosirea unui număr mare de supoziții necesare în cadrul metodelor bazate pe distanță sau probabilitatea substituției nucleotidelor (sau aminoacizilor). În timp ce toate modelele matematice utilizate în prezent, reprezintă o aproximare a realității, MP poate furniza arbori adecvați comparativ cu alte metode atunci când gradul de divergentă al secvențelor este scăzut. De fapt, simulările au arătat că atunci când: 1. gradul de divergență al secvențelor este scăzut *(d≤0,1)*; 2. rata substituției nucleotidelor este mai mult sau mai puțin constantă; 3. numărul de nucleotide examinate este mare, metodele MP sunt adesea mai bune decât metodele bazate pe distanță, în obținerea unor topologii reale (Sourdis si Nei 1988;Nei 1991).În plus, analiza parsimoniei este foarte utilă pentru anumite tipuri de date moleculare, de tipul secvențelor inserate sau inserții/deleții.

2.5.2.6. VARIANȚA ŞI COVARIANȚA BOOTSTRAP (REPLICĂRILOR REPETATE)

Există unele metode de estimare a numărului de substituții ale aminoacizilor între două secvențe, în funcție de modelul matematic folosit. În practică, orice model este o aproximare a realității şi dă numai valori aproximative ale substituției aminoacizilor. Mai mult, formulele analitice pentru calcularea distanțelor varianțelor estimate sunt de asemenea aproximative.

Când trebuiesc estimate lungimile ramurilor unui arbore filogenetic pentru mai multe secvențe, aceasta se poate realiza prin metoda LS (least-square = metoda pătratului minim), dar trebuiesc cunoscute varianțele şi covarianțele distanțelor estimate pentru toate perechile de secvențe diferite.

În aceste cazuri, se poate folosi metoda bootstrap pentru calcularea varianțelor şi covarianțelor distanțelor măsurate. Această metodă nu necesită formularea unei ipoteze referitoare la distribuțiile distanțelor, exceptând faptul că fiecare site al unui aminoacid poate evolua independent (Efron, 1982a, 1982b; Efron şi Tibshirani, 1993).

Presupunând că avem trei secvențe proteice cu n aminoacizi, înrudite evolutiv:

x11 x12 x13 x14 x15.....x1n
x21 x22 x23 x24 x25.....x2n
x31 x32 x33 x34 x35.....x3n

Xij reprezintă aminoacidul din poziția *j* secvența *i*. Putem astfel, calcula valorile *dij* pentru secvențele *i* şi *j*. Mai mult, se poate obține chiar o estimare a distanței pentru fiecare pereche de secvențe.

În cadrul acestei metode, pentru calcularea varianțelor şi covarianțelor, o secvență aleatoare de trei aminoacizi cu *n* situsuri de aminoacizi este generată din setul original de date. Această probă (model) este produsă prin luarea de probe (separarea de modele) din cadrul situsurilor aminoacizior (pe coloane), cu înlocuire, utilizând numere pseudoaleatoare. Acest model aleator al secvențelor de aminoacizi poate include două sau mai multe reprezentări ale unui site, dar nici o reprezentare a altuia. De exemplu, un model aleator poate avea următoarele secvențe cu *n* situsuri de aminoacizi.

x11 x12 x12 x14 x15.....x1n-1
x21 x22 x22 x24 x25.....x2n-1
x31 x32 x32 x34 x35.....x3n-1

Acesta, poartă numele de *bootstrap resampled data set* şi odată obținut, o estimare a distanțelor este calculată cu ajutorul ecuațiilor:

$$\hat{p} = n_d / n$$
$$d = -\ln(1-p)$$
$$\hat{d}_R = \alpha\left[(1-p)^{-1/\alpha} - 1\right]$$

, pentru oricare trei perechi de secvențe.

Dacă aceasta se repetă de *B* ori, poate fi generată distanța *B* pentru un număr dat de perechi de secvențe. Se notează cu $\hat{d}_b$ şi $\hat{d}$ valorile pentru *b* număr de replicări. În acest caz, varianța devine:

$V_B(\hat{d}) = \frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{d}_b - \bar{d}\right)^2$, unde $\bar{d}$ = media valorilor $\hat{d}_b$ la sfârşitul tuturor replicărilor. Pentru calcularea $V_B(\hat{d})$, de obicei se folosesc 1000 de replicări (*B*=1000).

O presupunere făcută pentru bootstrap, este aceea că toate siturile evoluează independent. Dacă numărul siturilor examinate este destul de mare (n>100), majoritatea siturilor cu rată evolutivă diferită vor fi reprezentate în fiecare model de bootstrap.

Covarianța, este dată de:

$Cov_B(\hat{d}_{ij}, d_{kl}) = \frac{1}{1-B}\sum_{b=1}^{B}\left(\hat{d}_{ijb} - \bar{d}_{ij}\right)^2\left(\hat{d}_{klb} - \bar{d}_{kl}\right)^2$, unde *ij* şi *kl* se referă la perechile de secvențe *i* şi *j*, *k* şi *l*.

Un avantaj al acestei metode este acela că poate calcula varianța şi covarianța, chiar atunci când nu sunt disponibile formulele matematice şi oferă estimări mai bune decât alte formule analitice (Efron şi Tibshirani, 1993).

2.6. SITUSURI INFORMATIVE ŞI HOMOPLAZIA

Un situs nucleotidic este informativ numai dacă favorizează o anumită filogenie față de altele. În căutarea arborelui MP situsurile identice pentru toate speciile sunt eliminate, fiind folosite doar cele variabile. Nu toate situsurile variabile sunt utile pentru a găsi o filogenie MP. Spre exemplu situsurile la care există diferențe unice (Figura 15) nu sunt informative deoarece ele pot fi explicate prin acelaşi număr de substituții oricare ar fi topologia.

Luând în considerare următoarele patru secvențe ipotetice se poate ilustra mai exact diferența dintre situsurile informative şi neinformative

Secvența	Situsuri omoloage 1 2 3 4 5 6 7 8 9
1	T A G A G A T C G
2	T G C C G A T C A
3	T G A T A A C C G
4	T G A G A A C C A

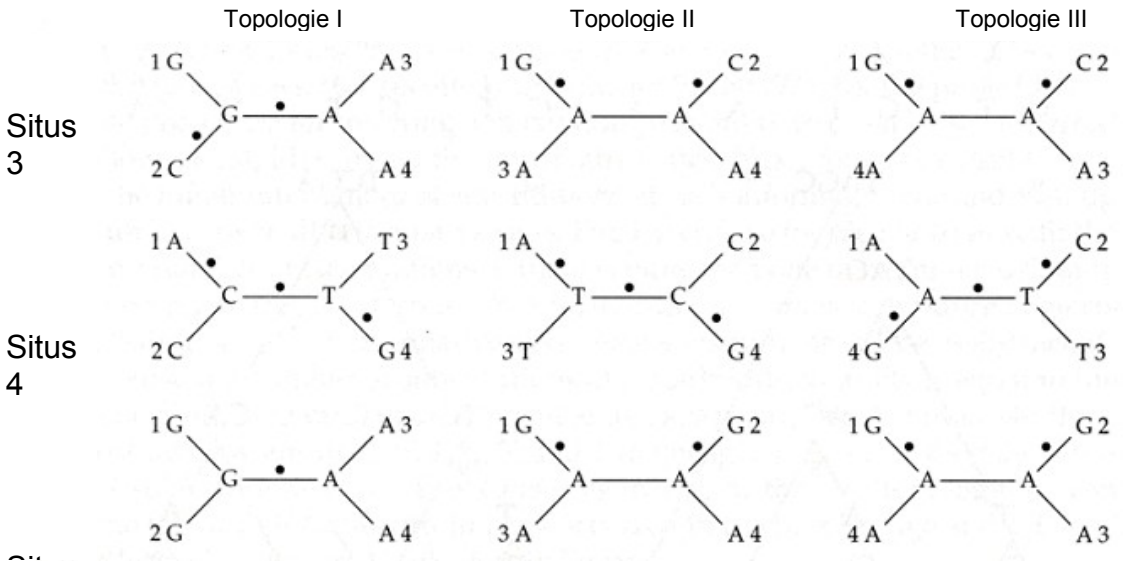


Figura 15 Tipuri de situsuri

Există trei arbori fără rădăcină posibili pentru 4 secvențe. Situsul 1 nu este informativ, toate secvențele având A în această poziție, deci nu este necesară nici o schimbare pentru oricare din cele trei topologii. Pentru situsul doi secvența 1 are A, celelalte 3 având G. Deci se poate presupune că în cadrul taxonului 1 a avut loc o mutație care a substituit G cu A. Nici acest situs nu este informativ deoarece oricare din cele 3 filogenii necesită o singură schimbare. După cum se arată în figura 12 de mai sus, pentru situsul 3 oricare dintre topologii necesită 2 mutații, deci nici acest situs nu este informativ. Se observă că dacă în topologia I în nodul intern ce duce la taxonii 1 şi 2 am avea C în loc de G, numărul de scimbări este tot de două. Pentru situsul 4 oricare ar fi topologia sunt necesare tot două schimbări, deci nici acesta nu

este informativ. Pentru situsul 5, filogenia I necesită o singură schimbare, pe când filogeniile II şi III necesită câte 2, deci abea acest situs este informativ.

Din acest exemplu şi din cele spuse mai sus se poate vedea că este informativ numai un situs ce are cel puțin două tipuri de nucleotide, fiecare fiind prezent de cel puțin două ori. Pentru aceste situsuri informative (situsurile 5, 7 şi 9) topologia I necesită 1, 1 şi 2 mutații respectiv, topologia II necesită 2, 2 şi 1 mutație, iar topologia III necesită 2, 2 şi 2 respectiv. Deci arborele I este ales ca fiind Arborele MP el necesitând cel mai mic număr de mutații.

Deoarece doar situsurile informative sunt utile în definirea filogeniilor MP pentru a obține o topologie viabilă sunt necesari un număr mare de asemenea situsuri. Dacă însă secvențele au un grad înalt de homoplazie (mutații paralele şi înapoi), topologia va fi incorectă chiar pentru un număr foarte mare de situsuri informative. Din acest motiv a fost propus un indice numit *indice de consistență* care măsoară gradul homoplaziei. Pentru un singur situs nucleotidic *i* acest indice este dat de *ci* = *mi* / *si* unde *m* este numărul minim posibil de substituții pentru respectivul situs pentru orice topologie şi este egal cu numărul de nucleotide observate minus unu, deoarece o stare este considerată cea ancestrală, iar *s* este numărul de substituții pentru respectiva topologie.

Limita de jos a indicelui de consistență nu este 0 şi *ci* variază pentru fiecare topologie în parte. Din aceste motive se folosesc alți doi indici şi anume *indicele de retenție* şi *indicele de consistență recalculat*. Indicele de retenție poate fi calculat ca *ri* = (*gi* - *si*)/(*gi* - *mi*) unde *gi* este numărul maxim posibil de substituții pentru situsul *i* pentru orice topologie şi este egal cu numărul de substituții necesare pentru o topologie în formă de stea admițând că cel mai frecvent nucleotid este ancestral şi este plasat în centru. Indicele de retenție devine 0 când *gi* = *si*, adică numărul maxim de substituții este identic cu numărul de substituții pentru respectiva topologie.

Indicele de consistență recalculat (*rci*) este dat de *ri* X *ci*, adică

$$rc_i = \frac{g_i - s_i}{g_i - m_i}\frac{m_i}{s_i}$$

Formulele de mai sus sunt date pentru un singur situs, practic însă este necesară calcularea acestor valori pentru toate situsurile informative, calculându-se în final indici totali (CI - consistenci index, RI - retention index şi RC - rescaled consistenci index) după formulele:

$$CI = \sum_i m_i \Big/ \sum_i s_i ; RI = \left(\sum_i g_i - \sum_i s_i\right) \Big/ \left(\sum_i g_i - \sum_i m_i\right); RC = \frac{CI}{RI}$$

2.7. STRATEGII DE CĂUTARE A FILOGENIEI MP

Când numărul de secvențe sau taxoni utilizați este mic, adică *m* < 10, se poate calcula lungimea tuturor topologiilor posibile, alegând topologia cu lungime minimă ca arborele MP. Acest tip de căutare se numeşte *exhaustivă*. Cum numărul de topologii posibile creşte foarte repede odată cu creşterea lui *m* examinarea tuturor topologiilor devine la un moment dat imposibilă. Dacă cunoaştem *a priori* că o parte din topologii sunt false, putem calcula doar lungimea celor rămase, acest tip de căutare fiind *căutarea arborilor specificați*.

Când *m* > 10 căutarea arborilor specificați nu este aplicabilă există alte două modalități numite *metoda ramificării şi unirii* (branch – and – bound) şi *căutarea euristică*. În cazul primei metode filogeniile care sunt evident mai lungi ca una examinată anterior sunt ignorate toate, iar arborele MP este determinat calculând lungimea unui grup de arbori care au potențialul de a avea lungimea minimă. În cea de a doua metodă este examinată doar o mică parte din toate topologiile posibile şi nu există nici o garanție că va fi găsită topologia cu o lungime minimă. Totuşi utilizând anumiți algoritmi, probabilitatea găsirii topologiei corecte creşte.

De multe ori nu există o singură topologie de lungime minimă ci mai multe. În acest caz se practică reprezentarea unei *filogenii-consens*, care este o sinteză a tuturor filogeniilor obținute. În *filogeniile-consens stric* toate modurile de ramificare ce intră în conflict sunt rezolvate prin crearea unui nod intern pluriramificat. În *filogeniile-consens după regula majoritară* acele moduri de ramificare care predomină sunt păstrate în arborele final. De obicei se utilizează regula majoritară de 50%, adică modurile de ramificare prezente în 50% sau mai mult dintre topologii sunt păstrate.

2.8. METODE ML (Maximum Likelihood)

Analizele filogenetice tind să deducă istoria evolutivă (sau un set de istorii probabile) care corespunde cel mai bine setului de date observat (în cazul de fată este vorba de secvențe nucleotidice sau de aminoacizi dar poate fi vorba şi de caractere morfologice, frecvențe ale genelor, situsuri de restricție etc.). Necunoscutele problemei sunt ordinea de ramificare şi lungimea ramurilor a filogeniei. Pentru a aplica metodele ML este nevoie de un model concret de substituție ce descrie transformarea unei secvențe în alta.

Metodele ML de reconstrucție filogenetică evaluează ipoteza despre istoria evolutivă în termeni probabilistici (Care este probabilitatea că o anumită istorie evolutivă - topologie şi un anumit model de substituție vor da naştere datelor observate). O istorie evolutivă ce are o probabilitate mai mare de a da naştere datelor observate are prioritate față de una cu o probabilitate mai mică. Această metodă a fost utilizată pentru prima dată în reconstrucțiile filogenetice de către Cavalli-Sforza şi Edwards în 1967 însă aceşti autori au considerat calculele necesare prea complicate pentru calculatoarele de la acea vreme şi au dezvoltat metode aproximative ca de exemplu ME (Huelsenbeck & Crandall, 1997). Felsenstein a utilizat pentru prima dată această metodă în 1981 pentru analiza filogenetică a secvențelor de nucleotide, după care metoda a început să fie din ce în ce mai utilizată. Printre avantajele acestei metode se numără varianța mică şi posibilitatea utilizării unui număr minim de parametri. Chiar pentru un număr mic de nucleotide acest tip de metode depăşesc de multe ori metodele bazate pe distanțe şi cele MP.

Principiul de bază al metodei implică calcularea verosimilitudinii unei filogenii. Deoarece majoritatea modelelor de substituție folosite sunt reversibile în timp, verosimilitudinea unei filogenii este independentă de localizarea rădăcinii. Presupunând că fiecare situs evoluează independent putem calcula verosimilitudinea fiecărui situs separat şi să combinăm aceste valori pentru obținerea unei valori finale. Pentru a calcula verosimilitudinea unui situs *j* trebuie să luăm în considerare toate scenariile posibile prin care starea finală a situsului ar fi putut evolua. Bineînțeles că unele scenarii sunt mult mai plauzibile decât altele, însă fiecare are o anumită şansă să se fi produs. Spre exemplu dacă considerăm o filogenie ca în fig. 14 a în rădăcină am fi putut avea teoretic oricare din cele 4 nucleotide (A, G, C, T), pentru oricare alt

nod intern este de asemeni posibilă existența oricărui nucleotid. Deci avem un total de 4 X 4 = 16 posibilități. Cum oricare dintre aceste 16 scenarii ar fi putut duce la setul final de date observate, probabilitățile tuturor evenimentelor trebuiesc calculate şi însumate pentru a obține probabilitatea unui situs *j* (această probabilitate depinde după cum am mai menționat de topologie şi de modelul de substituție nucleotidică) (fig. 14 c). Cum am presupus că fiecare situs evoluează independent de celelalte, probabilitatea unui scenariu evolutiv este egală cu produsul probabilităților tuturor situsurilor 1, 2, 3, … *j* … *n*. Pentru că probabilitatea unei singure observații este foarte mică în locul lor se utilizează logaritmii acestora, deci probabilitățile sunt acumulate ca sumă a logaritmilor verosimilitudinilor fiecărui situs luat în parte (fig. 14 e)

Din cauza volumului foarte mare de calcule ce trebuiesc efectuate dacă numărul de taxoni este mai mare ca 10 şi secvențele utilizate sunt mari au fost dezvoltate o serie de algoritmi de căutare (ca şi în cazul metodelor MP).

Cum modalitatea reală de substituție a nucleotidelor este foarte complicată, s-ar putea crede că un model matematic cu mulți parametri este mai bun decât un model cu mai puțini parametri. Totuşi chiar dacă un model cu mai mulți parametri corespunde datelor mai exact decât unul mai simplu, numărul mare de parametri îl face mai sensibil la erori. Deci este mai utilă utilizarea unui model mai simplu atâta timp cât acesta aproximează modul de substituție al nucleotidelor suficient de bine.

În cazul metodelor ML corespondența dintre model şi date poate fi examinată folosind testul *LR* (Likelihood Ratio Test) sau criteriul informației al lui Akaike (*AIC* – Akaike Information Criterion). Când avem două modele 1 şi 2 şi modelul 1 este caz particular al modelului 2, iar topologia corectă este cunoscută putem calcula testul LR după formula

$$LR = 2(\ln L_2 - \ln L_1)$$

unde ln*L*1 şi ln*L*2 sunt valorile ML pentru modelele 1 şi 2. în acest mod putem testa dacă modelul 2 este semnificativ mai bun decât modelul 1.

$$L = L_{(1)} \cdot L_{(2)} \cdot \ldots \cdot L_{(j)} \cdot \ldots \cdot L_{(n)} = \prod_{j=1}^{n} L_{(j)}$$

$$\ln L = \ln L_{(1)} + \ln L_{(2)} + \ldots + \ln L_{(j)} + \ldots + \ln L_{(n)} = \sum_{j=1}^{n} \ln L_{(j)}$$

Spre exemplu modelul lui Kimura este caz special al modelului HKY, primul având un parametru liber, celălalt patru. Deci diferența în gradul de corespondență dintre model şi date poate fi evaluată prin testul *LR* sau χ 2 cu 3 grade de libertate. Dacă dintre cele două două modele comparate unul nu este caz particular al celuilalt testul *LR* nu poate fi utilizat, un mod alternativ de a genera distribuția testului este simularea Monte Carlo (parametric bootstrapping) (Goldman, 1993) sau se aplică *AIC* definit ca

$$AIC = -2\ln L + 2p$$

unde ln*L* este logaritmul verosimilitudinii unui anumit model, iar *p* este numărul de parametri liberi ce trebuiesc estimați. Se consideră că predictibilitatea statistică a unui model este cu atât mai bună cu cât *AIC* are valori mai mici. Din ecuația de mai sus se vede că chiar dacă un model mai sofisticat are valori ale –ln*L* mici din cauza

numărului mare de parametri *p* valoarea indicelui *AIC* creşte. Totuşi în alte domenii acest criteriu dă indicații mult mai bune decât în reconstrucțiile filogenetice. De cele mai multe ori valoarea *AIC* pentru un model complex cum ar fi HKY este mai mică decât pentru un model simplu precum K2P ceia ce nu înseamnă însă obligatoriu că topologia obținută cu un model mai complex este întotdeauna mai bună.

Metode ML aplicabile în cazul secvențelor de aminoacizi şi a proteinelor

Când secvențele de ADN sunt destul de apropiate între ele din punct de vedere filogenetic metodele ML funcționează foarte bine, în special atunci când sunt utilizați prima şi a doua poziție a codonului. Dacă se folosesc secvențe (ce codifică proteine) îndepărtate filogenetic apar complicații deoarece rata substituțiilor sinonime este mult mai mare decât cea a substituțiilor nesinonime, frecvența celor 4 nucleotide în cea de a treia poziție a codonului poate varia mult arătând că este astfel violat modelul staționar de substituție. Din contra, schimbările evolutive ale secvențelor proteice sunt mult mai încete astfel încât se poate utiliza metoda ML pentru secvențele proteice utilizând spre exemplu matricea de tranziție a aminoacizilor a lui Dayhoff et al. din 1978.

METODOLOGIA DE OBȚINERE A SECVENȚELOR NUCLEOTIDICE

Un prim obiectiv al cercetărilor, îl constituie cel de stabilire a omogenității loturilor, urmat de stabilirea cantității totale de ADN pentru diferite specii şi subspecii ale genurilor *Carassius* şi *Cyprinus*, stabilirea unor intervale de variație a cantității de ADN pentru diferite țesuturi, provenind de la indivizi diferiți ai celor două genuri, amplificarea genelor luate în studiu, verificarea acestora în electroforeză, urmată de purificare şi reacția de secvențiere.

3.1. DETERMINAREA UNOR PARAMETRI ŞI INDICI FENOTIPICI LA INDIVIZII STUDIAȚI

Asupra indivizilor aparținând celor două specii de ciprinide utilizați pentru dozarea acizilor nucleici, s-au efectuat măsurători, urmărindu-se o serie de parametri fenotipici (*Lt* = lungime totală; *Ls* = lungime standard; *g* = greutate individuală; *C* = circumferință; *Hd* = înălțime maximă; *Ha* = înălțime minimă; *Lc* = lungime cap) şi de indici fenotipici (indice de profil = *Lt / Hd*; coeficient Fulton = *g* x 100 / *Ls3*; indice de circumferință (Kiselev) = *Ls / C* şi raportul *Lc* x 100 / *Lt*) (Gorgan et al., 2000)

Pentru determinarea parametrilor fenotipici se vor avea în vedere, următorii parametri (Voican et al., 1974, 1975):

- lungimea totală a peştelui *(Lt)* este distanța de la vârful botului până la linia care uneşte vârfurile lobilor înotătoarei caudale;
- lungimea standard *(Ls)* este distanța de la vârful botului până la extremitatea învelişului solzos, la baza înotătoarei caudale;
- lungimea capului *(Lc)* este distanța de la vârful botului până la marginea posterioară a osului opercular, fără membrana branhiostegală;
- înălțimea maximă a corpului *(Hd)* se măsoară în regiunea cea mai lată a corpului;
- înălțimea minimă a corpului *(Ha)* se măsoară acolo unde înălțimea corpului este cea mai redusă (la extremitatea posterioară a pedunculului caudal);
- circumferința maximă a corpului *(C)* se măsoară la nivelul grosimii maxime şi înălțimii maxime;
- greutatea individuală *(g)* s-a determinat prin cântărirea peştilor;

Indicii fenotipici menționați s-au determinat pe baza valorilor parametrilor fenotipici după cum urmează:

- *indicele de profil* = *Ls / Hd* unde:
Ls = lungimea standard, în cm.
Hd = înălțimea maximă a corpului, în cm;

- *coeficientul Fulton* (de îngrăşare):

$$\frac{g}{Ls^3} \times 100 \text{ , unde:}$$

g = greutatea corpului, în g;
Ls = lungimea standard, în cm.

Pentru a obține informații cu privire la gradul de uniformitate a populației, precum şi asupra nivelului de semnificație statistică a diferențelor dintre populațiile celor două specii, s-au efectuat măsurători biometrice la 50 exemplare/specie şi s-au calculat valorile unor parametri statistici (Snedecor, 1968): $\bar{x}$ = media aritmetică; *S*=deviația standard; *Es* = eroarea standard; *C.V.* % = coeficientul de variație.

1. Media aritmetică:

$$x = \frac{\sum xi}{n}$$

, în care: *Σxi* = suma valorilor individuale;
n = numărul indivizilor.

2. Estimarea deviației standard pornind de la probă către populație

$$S = \sum x^2 - \frac{\left(\sum x\right)^2}{n}$$

, unde:

Σx 2 = suma pătratelor valorilor individuale;
(Σx)2 = pătratul sumei valorilor individuale;
n = numărul indivizilor

3. Eroarea standard:

S2=varianța

$$S_{\bar{x}} = \sqrt{\frac{S^2}{n}} = \frac{S}{\sqrt{n}}$$

4. Coeficientul de variație:

$$C.V.\% = \frac{S \times 100}{\bar{x}}$$

S = deviația standard;

, în care:

$\bar{x}$ = media aritmetică.

Pentru a stabili uniformitatea loturilor care au fost utilizate în experimentele ulterioare de extracție a ADN total şi de secvențiere, au fost folosiți câte 50 de indivizi aparținând celor două specii, pentru fiecare populație analizată.

3.2. PROTOCOLUL DE EXTRACȚIE AL ADN TOTAL DIN ȚESUT MUSCULAR

Tehnica este folosită pentru extragerea ADN total din țesuturi proaspete, congelate sau păstrate în etanol (Ausubel şi colab., 1995). În vederea determinării cantităților totale de acizi nucleici şi a intervalelor de variabilitate ale acestui parametru pentru diferite țesuturi, au fost folosiți câte 20 de indivizi aparținând celor două specii.

Soluții:

1. Tampon pentru liza celulară:

Conține: Tris HCl 50mM (pH=8), SDS 1,0%, EDTA 25mM.

Pentru 50ml, într-o eprubetă se introduc 25ml H2O miliQ (apă distilată filtrată prin filtru de carbon, schimbător de ioni etc.) autoclavată, la care se adaugă 2,5ml soluție stoc 1M Tris HCl (pH=8), 5ml 10% SDS, 2,5ml 0,5M EDTA. Se completează volumul cu H2O miliQ până la 50ml şi se introduce într-un tub Corning nou cu volumul de 50ml, neautoclavat.

Soluții stoc:

1. Soluție Tris-HCl 1M (pH=8): - cantitățile sunt calculate pentru un volum de 100ml;

- se dizolvă 12,11g Tris în 80ml H2O miliQ;
- se corectează pH-ul la 8 cu HCl;
- se autoclavează.

2. Soluție Tris-EDTA (TE) (pH=8,0): -cantitățile sunt considerate pentru un volum de 100ml;

- într-o eprubetă de 100ml se introduce 1ml soluție stoc Tris-HCl 1M (pH=8);
- 200µl EDTA 0,5M;
- se completează cu H2O-miliQ până la 100ml şi se autoclavează.

3. Soluție NaOH 10N (100ml):

- se cântăresc 40g NaOH
- se adaugă 70ml H2O-miliQ autoclavată şi se agită până la dizolvare;
- se aduce volumul la 100ml;
- nu se autoclavează.

4. Soluție SDS 10% (dodecil sulfat de sodiu):

- se introduc 10g SDS în 80ml H2O-miliQ autoclavată şi se încălzeşte soluția la 80ºC pentru o mai bună dizolvare;
- se ajustează pH-ul la 7,2 (dacă este necesar), utilizând HCl;
- se aduce volumul la 100ml cu H2O-miliQ;
- nu se autoclavează.

5. Soluție EDTA 0,5M (etilendiaminotetraacetat·sare disodică – Ref. SIGMA ED2SS):

- se cântăresc 18,61g EDTA, peste care se adaugă 80ml H2O-miliQ;
- se agită puternic şi se adaugă 1ml soluție NaOH 10N;
- se adaugă NaOH până când soluția devine alcalină (pH=8,0), iar EDTA este dizolvat complet.
- se aduce volumul la 100ml cu H2O-miliQ şi se autoclavează;

Modul de lucru:

Se etichetează tuburi Eppendorf de 1,5ml autoclavate, în care se introduc câte 500µl soluție tampon (tampon pentru liză). Tuburile se țin închise, până în momentul introducerii țesutului.

Se cântăresc între 20 – 200mg de țesut muscular proaspăt, congelat sau conservat în alcool etilic 95% (în acest caz, țesutul se usucă înainte de cântărire, pe hârtie de filtru). Fragmentul de țesut se aşează pe o lamă sterilă şi se taie mărunt cu un bisturiu sau cu o lamă, sterilizate. Dacă fragmentul este mic (dacă provine de la un individ de 1 – 2cm), această etapă nu este necesară.

Se introduc fragmentele de țesut în tubul Eppendorf (de 1,5ml) corespunzător, care conține 500µl soluție tampon pentru liză. Această etapă se repetă pentru toți indivizii. În fiecare tub, se adaugă câte 10µl proteinază K (10µl reprezintă echivalentul a 200µg proteinază dintr-o soluție stoc 20mg/ml care este preparată şi congelată la -20ºC). Se agită fiecare probă pe un vortex şi se păstrează peste noapte în etuvă la 37ºC.

Notă: pentru realizarea acestei etape se recomandă să se lucreze cu un număr de probe echivalent cu numărul de cuve din centrifugă.

În ziua următoare, sub nişă, în fiecare tub de 1,5ml se adaugă 600µl de soluție fenol : cloroform : alcool izoamilic (25 : 24 : 1), până la semnul superior al tubului (Figura 12). Se agită tuburile timp de 30 – 60 secunde, apoi se centrifughează (utilizând o centrifugă de masă), la o turație 8000 rotații/minut, timp de 4 minute. În timpul centrifugării, se etichetează noi tuburi Eppendorf de 1,5ml. La sfârşitul centrifugării, tuburile prezintă o fază inferioară de culoare gălbuie care conține solventul organic, o fază superioară care conține ADN şi o fază intermediară unde sunt depozitate toate resturile pe care dorim să le eliminăm din soluția de ADN (Figura 13)

Notă: dacă această fază intermediară este foarte mare, se repetă spălarea cu fenol : cloroform : alcool izoamilic urmată de centrifugare. Dacă fazele sunt bine separate, se trece la etapa următoare.

Se separă faza superioară a fiecărui tub, în noile tuburi etichetate; - pentru aceasta se utilizează o pipetă Gilson P-100, cu ajutorul căreia se extrag 700µl de ADN (Figura 14). Dacă pipetarea se face repede, iar concentrația ADN este foarte mare în faza superioară a tubului, în momentul pipetării, se pot extrage resturi ale fazei intermediare. Se extrage faza superioară din fiecare tub menținându-le deschise sub nişă, iar la sfârşit, se adaugă câte 550µl cloroform (Figura 15). Se agită tuburile timp de 30 – 60 secunde, apoi se centrifughează la 8000 rotații/minut, timp de 3 minute, utilizând o centrifugă de masă (tip COBOS MC-10). În timpul centrifugării, se etichetează noi tuburi.

Se separă fazele superioare rezultate în urma centrifugării (Figura 16), în noile tuburi, în acelaşi mod ca în etapa precedentă (Figura 17), având grijă să nu se preia nimic din faza intermediară de separare.

Se adaugă 1ml etanol rece 95% (păstrat la -20ºC) în noile tuburi după separare până la semnul superior al fiecărui tub (Figura 18). Se agită tuburile şi se lasă la temperatura de -20ºC, timp de 30 – 60 minute, timp în care se va forma un precipitat albicios (ADN total) pe fundul tuburilor (Figura 19).

După precipitarea ADN, se centrifughează tuburile la 10.000 rotații/minut, timp de 5 minute, pentru sedimentare. Se elimină supernatantul şi se aşează tuburile deasupra unei hârtii de filtru, astfel încât să se scurgă ultimele picături de alcool. Urmează uscarea peletelor, mutând tuburile, deschise, într-o centrifugă cu vacuum, unde sunt menținute timp de 10 minute, cu temperatura în scădere. Dacă în acest

interval de timp sedimentele nu sunt uscate, se mai lasă tuburile încă 5 minute în centrifugă.

După uscarea peletelor, se adaugă în fiecare tub câte 200µl soluție TE (pH=8,0) şi se resuspendă manual sau pipetând de câteva ori. Pentru o mai bună resuspensie, se pot lăsa între 15 şi 30 minute la temperatura camerei, la 37ºC sau timp mai îndelungat la 4ºC. După resuspendarea peletelor de ADN, conținutul fiecărui tub este repartizat în trei tuburi de câte 1,5ml (etichetate anterior), punând câte 50µl din soluția resuspendată. Din cele trei tuburi ale unei probe, unul se păstrează la 4ºC, iar celelalte două se congelează (-20ºC - -80ºC).

3.2.1. PRELUCRAREA STATISTICĂ A DATELOR OBȚINUTE DIN EXTACȚIA ADN

3.2.1.1. ESTIMAREA VALORILOR MEDII ALE POPULAȚIILOR ŞI STABILIREA INTERVALELOR DE VARIABILITATE

Folosind valorile mediei $(\bar{x})$ şi deviației standard *(S)* (conform formulelor prezentate anterior), calculate pe baza probelor dintr-o populație, se pot estima valorile corespunzătoare pentru întreaga populație (sau specie). Estimarea mediei populației implică calculul erorii standard a mediei $(S_{\bar{x}})$, care ține cont de variabilitatea şi mărimea probei (Varvara et al., 2001).

Se caută o valoare critică pentru un anumit grad de confidență *α* şi un anumit număr de grade de libertate *t(α, t-1)*; atunci, probabilitatea estimării va fi dată de *1-α*, adică *α=0,05*. Valoare lui *t* reprezintă valoarea critică a distribuției *t* pentru un anumit nivel de confidență sau prag de semnificație *α* şi *n-1* grade de libertate. Pragul de semnificație reprezintă probabilitatea aleasă arbitrar, ca intervalul de confidență calculat să cuprindă media populației. Pentru majoritatea studiilor, probabilitatea de 95%, adică *α=0,05*, este considerată satisfăcătoare.

Eroarea standard a mediei este folosită pentru a calcula intervalul de confidență a mediei populației, interval care să cuprindă, cu o anumită probabilitate, media populației.

$\mu = \bar{x} \pm t(\alpha, n-1) \cdot S_{\bar{x}}$, unde μ reprezintă media populației, iar $\bar{x}$ valorile medii ale probelor.

Limitele intervalului: limita inferioară *(LI)* şi limita superioară *(LS)*, sunt date de formulele:

$$LI = \bar{x} - S_{\bar{x}} \cdot t(\alpha, n-1)$$

$$LS = \bar{x} + S_{\bar{x}} \cdot t(\alpha, n-1)$$

Astfel, intervalul LI-LS, include media populației cu o probabilitate de *1-α*.

3.2.1.2. COMPARAREA MAI MULTOR PROBE

ANOVA model I bifactorial cu număr egal de observații în celulă

Acest model se foloseşte în condiții asemănătoare cu modelul bifactorial cu o observație în celulă la care se adaugă interacțiunea dintre factori. Evidențierea interacțiunii poate fi sugerată de reprezentarea grafică a mediilor probelor urmărite.

Observația în acest caz este egală cu media generală + efectul nivelului *„i”* al factorului *A* + efectul nivelului *„j”* al factorului *B* + interacțiunea *AiBj* + eroarea întâmplătoare (Fowler et al., 2000).

		Ai	
		A1	A2
Bj	B1	- - - - $\bar{x}$	- - - - $\bar{x}$
	B2	- - - - $\bar{x}$	- - - - $\bar{x}$

În fiecare celulă trebuie să avem un număr egal de observații

$$S_t^2 = S_{ext}^2 + S_{\text{int}}^2$$
$$S_{ext}^2 = S_c^2 + S_r^2 + S_i^2 \quad ,$$
$$S_t^2 = S_c^2 + S_r^2 + S_i^2 + S_{\text{int}}^2$$

Unde: S_i^2 = varianța interacțiunii; S_{ext}^2 =varianța externă; S_t^2 =varianța totală

Se stabilesc ipotezele modelului:

H01: μA1=μA2=μAn	H11: μA1≠μA2≠μAn
H02: μB1=μB2=μBn	H12: μB1≠μB2≠μBn
H03: interacțiunea=0	H13: interacțiunea≠0

$S^2 = \overline{SS} = \dfrac{SS}{grade \quad libertate}$, SS =suma de pătrate; $\overline{SS}$ =suma de pătrate medie;

$SS_t = \sum x_t^2 - \dfrac{\left(\sum x_t\right)^2}{n_t}$, SS_i = suma de pătrate a interacțiunii

$$SS_{ext} = SS_c + SS_r + SS_i,$$

$$SS_c = \sum\left[\frac{\left(\sum x_c\right)^2}{n_t r}\right] - \frac{\left(\sum x_t\right)^2}{n_t}$$

$$SS_r = \sum\left[\frac{\left(\sum x_r\right)^2}{n_t c}\right] - \frac{\left(\sum x_t\right)^2}{n_t}$$

$$SS_i = \frac{\left(\sum x_{c1r1}\right)^2}{n_{c1r1}} + \frac{\left(\sum x_{c1r2}\right)^2}{n_{c1r2}} + \frac{\left(\sum x_{c2r1}\right)^2}{n_{c2r1}} + \frac{\left(\sum x_{c2r2}\right)^2}{n_{c2r2}} - \frac{\left(\sum x_t\right)^2}{n_t} - SS_c - SS_r$$

$$SS_{\text{int}} = SS_t - SS_c - SS_r - SS_i$$

glt =nt – 1, glt=grade de libertate totale, nt=număr total de valori;
glc = c – 1, glc=grade de libertate pe coloane, c=număr de valori pe coloane;
glr = r – 1, glr=grade de libertate pe rânduri, r=număr de valori pe rânduri;
gli = (c-1)(r-1), gli=grade de libertate ale interacțiunii;
glint = nt – cr, glint=grade de libertate interne

Tabel sintetic al valorilor calculate pentru testul ANOVA

Sursa de variabilitate	SS	Grade de libertate	$\overline{SS}$	F
c	SS_c	c - 1	SS_c / c - 1	$\overline{SS}_c / \overline{SS}_{\text{int}}$
r	SS_r	r - 1	SS_r / r - 1	$\overline{SS}_r / \overline{SS}_{\text{int}}$
i	SS_i	(c-1)(r-1)	SS_i /(c - 1)(r - 1)	$\overline{SS}_i / \overline{SS}_{\text{int}}$
int	SS_{int}	nt - cr	SS_{int} / nt - cr	
t	SS_t	nt - 1		

Dacă: Fc ≥ F (α, c – 1, g.l.int) ⇒ H01 (F), H11 (A)
Fr ≥ F (α, r – 1, g.l.int) ⇒ H02 (F), H12 (A)
Fi ≥ F (α, (c-1)(r-1), g.l.int) ⇒ H03 (F), H13 (A)

3.3. REACȚIA DE AMPLIFICARE GENICĂ PRIN POLIMERIZARE ÎN LANȚ (PCR)

3.3.1. PRINCIPII GENERALE

Apărută la începutul anilor ’80, reac□ia de polimerizare în lan□ (PCR) este atât de utilizată în zilele noastre în nenumărate domenii, încât s-a uitat meritul său revoluționar . Ştiința aşa numitei reacții de polimerizare în lanț este astăzi foarte bine cunoscută şi întâlnită în multe laboratoare. Tehnica PCR este considerată a fi un veritabil „fotocopiator genetic".

Reacția de polimerizare în lanț permite o rapidă amplificare a secvențelor ADN, chiar si atunci când materialul de inițiere este un ADN puternic degradat, acesta

fiind utilizat in studii de antropologie moleculara si paleogenetica, cum ar fi analiza ADN recuperat din oase fosile, din mumii sau din țesuturi formolizate. Se poate realiza amplificarea ADN-ului chiar si dintr-o singura celula. Practic, PCR-ul reprezintă o forma de clonare „in vitro” ce poate genera şi modifica fragmente de lungimi şi secvențe bine definite într-o simplă reacție automatizată.

Metaforic vorbind, metoda permite sa fie găsit „acul în carul cu fân”, deoarece prin ea, o singură genă dintre nenumăratele gene purtate de o moleculă de ADN poate fi reprodusă în câteva ore în miliarde de copii identice; ceea ce unei celule, în starea sa cea mai activă proliferativ (celula canceroasă) i-ar trebui o lună de zile, pentru a realiza acelaşi lucru.

Probele ADN şi reacția de polimerizare în lanț sunt descoperiri ce au fost puse imediat în practică, în medicina legală şi în identificarea genelor responsabile de boli ereditare.

Reacția de polimerizare în Lanț este o metodă foarte sensibilă de amplificare, prin care se realizează miliarde de copii identice ale unei anumite gene din macromolecula de ADN, în numai câteva ore, sistemul putând fi contaminat foarte uşor de elemente ale genelor ADN amplificate anterior sau de particule externe. Posibilitatea de contaminare în reacțiile de amplificare este un aspect cu implicații deosebite, atât în aplicații de cercetare cât şi în cele cu specific diagnostic.

În ciuda sensibilității deosebite, PCR este victima propriului său succes. Pot să apară reacții fals pozitive prin acizii nucleici contaminați datorită puterii de amplificare incredibile a reacției. Contaminarea noilor probe cu ADN amplificat anterior este o problemă importantă a laboratoarelor în care se execută PCR.

Reacția de polimerizare în lanț permite manipularea directă a secvențelor de acizi nucleici, putându-se realiza astfel secvențierea unui fragment ADN (determinarea ordinii de succesiune a bazelor azotate în molecula de ADN), determinarea variațiilor existente între diferite molecule ADN, modificarea secvenței moleculelor de acizi nucleici de interes. Este o tehnică ingenioasă bazată pe capacitatea ADN-polimerazei de a copia o catenă ADN prin elongarea catenelor complementare, inițiată de o pereche de primeri. Teoretic, fiecare ciclu dublează cantitatea de ADN țintă. Comportamentul şi manipularea ADN polimerazei a fost o concentrare a eforturilor pentru a înțelege felul în care materialul genetic este copiat, reparat şi moştenit.

Se poate porni cu o cantitate extrem de mică de ADN, de la o singură sau câteva molecule de ADN ce trebuie amplificat, denumit ADN țintă. Acest ADN țintă ce urmează a fi amplificat, are în mod curent, sub 45 kilobaze.

Reacția de polimerizare în lanț se desfăşoară „in vitro”, având nevoie de următorii reactivi: template-ul ADN ce urmează a fi amplificat, primerii necesari amplificării unui anumit sector din ADN țintă, deoxiribonucleotide sub formă de trifosfați (dATP, dGTP, dTTP şi dCTP) care pot fi dozați în soluție apoasă prin spectofotometrie UV la 260 nm, ADN polimeraza, capabilă de a rămâne funcțională la temperaturi înalte de pana la 100º C şi bufferul ($MgCl_2$ în concentrații de aproximativ 1,5 mM) care este de o importanță vitală, deoarece determină condițiile de hibridare a primerilor şi cele de funcționare a polimerazei (Taq polimeraza nu poate funcționa când salinitatea mediului este crescută). ADN ce urmează a fi amplificat se află în cantitate foarte mică şi este utilizat ca template pe un sector limitat (ADN țintă), delimitat de doi primeri, cu orientare opusă, unul pe catena directă 5` - 3` şi celălalt, pe catena 3` - 5`. Primerii sunt reprezentați de secvențe scurte oligonucleotidice sintetizate „in vitro“, de ADN monocatenar, de numai 20-30 nucleotide lungime. Secvența primerilor este proiectată înaintea sintezei artificiale,

astfel încât să fie realizată împerecherea complementară cu o secvență aflată înaintea sectorului din ADN țintă ce urmează a fi amplificat, primerii delimitând astfel acel sector. Cei doi primeri sunt proiectați să se asocieze unul aproape de celălalt dar sunt situați pe catene diferite şi sunt orientați în aşa fel încât ei copie secvența de ADN care se găseşte între ei. Cu cât primerii sunt mai lungi, cu atât temperatura de asociere poate fi mai mare ceea ce creează o specificitate sporită. Cu toate acestea, priimerii nepurificați si lungi vor conține cantități mai mari de produse nespecifice in amestecul final. Dacă este necesară utilizarea unor primeri mai lungi atunci se recomandă purificarea lor, în caz contrar aceasta nu este necesară.

Există şi aşa numiții primeri nespecifici sau degenerați care sunt de fapt un amestec de oligonucleotide ale unor secvențe de acizi nucleici diferite şi sunt utilizați pentru amplificarea simultană a mai multor loci în macromolecula de ADN țintă.

Majoritatea polimerazelor recunosc ADN mono catenar ca fiind un posibil model de amplificare şi se leagă temporar la acest lanț, pentru formarea unei duble catene. De asemenea, polimerazele se leagă şi de deoxinucleotidtrifosfații (dNTPs) valabili în mediu şi folosind energia înmagazinată în legăturile triple, catalizează reacția de ataşare a nucleotidelor la a doua catenă de ADN; apoi enzima se mută la sfârşitul porțiunii unde catena este dublă iar procesul este repetat de sute sau mii de ori pe secundă.

ADN polimeraza acționează unidirecțional; începe să acționeze de la capătul 3` al porțiunii dublucatenare, sinteza decurgând în direcția 5`→3`. Astfel, polimeraza poate începe să acționeze, oriunde găseşte o joncțiune de tip ADN monocatenar – bicatenar. Pentru direcționarea începutului sintezei unui anumit segment de ADN, este necesar un scurt segment oligonucleotidic (un mic număr de nucleotide unite într-un lanț scurt complementar unei secvențe de ADN) poziționat chiar deasupra porțiunii dorite pentru sinteză (Prescott şi Harley, 1999).

Acest fragment oligonucleotidic nu se poate alinia la ADN dublu catenar, în schimb, se leagă imediat la secvența complementară a ADN mono catenar. Prin încălzire, legăturile dintre cele două catene ale ADN se rup, formându-se porțiuni monocatenare. Prin urmare, denaturarea ADN prin căldură, urmată de o uşoară răcire, în prezența oligonucleotidelor complementare, determină alinierea acestora la macromolecula de ADN monocatenar, oriunde întâlnesc secvențe similare complementare. (Palumbi, 1996; Monod şi Francois 1996; Lisker şi Armendares, 2000; Mas Eva et al., 2001).

3.3.2. CICLURILE DE REPLICARE

Replicarea în PCR constă în trei faze mai importante: denaturare, aliniere şi extensie.

3.3.2.1. DENATURAREA

În această fază, căldura este folosită în scopul opririi tuturor reacțiilor enzimatice şi denaturării ADN de la dublu catenar, la mono catenar. De obicei, se foloseşte o temperatură de 94ºC sau 92ºC după metode mai noi. O temperatură prea joasă sau un timp de denaturare prea scurt, pot fi ineficiente, rămânând porțiuni de ADN bicatenar. Cu toate că *Taq* polimeraza este rezistentă la denaturarea termică, nu este totuşi imună la acțiunea căldurii, o încălzire excesivă, ducând la o scădere a activității

enzimatice. De exemplu, după 30 cicluri de replicare cu incubare la 94ºC pentru 60 de minute, activitatea enzimei, se reduce la jumătate.

3.3.2.2. ALINIEREA PRIMERILOR

În această fază, temperatura este mai joasă, astfel încât primerii oligonucleotidici se pot lega în zonele complementare din catena de ADN. Aceasta este faza critică a reacției de amplificare, deoarece dacă primerii se leagă corect doar în zona țintă (zona care urmează a fi amplificată), atunci există o probabilitate suficient de mare să se obțină produşi de calitate.

Alinierea este un proces randomic, care depinde de concentrația primerilor, disponibilitatea siturilor de legătură şi de prezența zonelor cu similaritate scăzută secvenței primerilor, care duc la alinieri nespecifice. Trebuie remarcat şi faptul că, dacă macromolecula de ADN care foloseşte drept matriță este fragmentată, atunci, mici fragmente din aceasta, pot acționa ca primeri, fiecare având o țintă precisă în cadrul genomului. Acesta este un motiv pentru care molecule de ADN cu greutate moleculară mare şi molecule mici, nu trebuiesc amestecate ca matrițe pentru viitoare amplificări (Palumbi, 1996).

O temperatură mare de aliniere a primerilor, sau o concentrație mare a acestora în mediul de reacție, pot determina o blocare a formării legăturilor şi implicit de aliniere a primelor cu matrița de ADN. O temperatură joasă, determină o aliniere nespecifică a primerilor în diferite porțiuni care prezintă o complementaritate scăzută, generându-se astfel numeroase artefacte. Dacă temperatura de aliniere este foarte scăzută, catenele de ADN se vor „realinia” formând structura inițială – ADN dublu catenar.

Timpul de aliniere a primerilor, constituie de asemenea o problemă foarte importantă pentru obținerea unui produs bun. Astfel, un timp de aliniere foarte lung, permite primerilor să găsească noi posibile situri de legătură şi să se alinieze nespecific. Cu toate acestea, un timp scurt de aliniere, poate determina o mare specificitate de aliniere în PCR. În general, timpul de aliniere este de 30 – 60 secunde, dar poate funcționa la fel de bine şi un timp mai scurt de 15 secunde la temperaturi mari de aliniere cu primeri specifici.

3.3.2.3. EXTENSIA

Faza de extensie sau elongare, determină activitatea enzimatică şi sinteza segmentului de ADN. *Taq* polimeraza funcționează de obicei bine la 72ºC, aceasta fiind şi temperatura utilizată de obicei pentru această etapă. Enzima se activează şi îşi începe activitatea la temperaturi scăzute, ceea ce este foarte important pentru succesul reacției de amplificare. Majoritatea primerilor au o temperatură de aliniere mai mică de 72ºC, astfel încât majoritatea vor fi foarte slab legați de secvența țintă, o dată cu creşterea temperaturii de extensie. O dată cu creşterea temperaturii de la temperatura de aliniere către 72ºC, începe şi reacția de polimerizare, dar destul de lent. Oricum, această reacție de polimerizare este suficientă pentru a adăuga câteva extra-baze, crescând astfel stabilitatea complexului primer – matriță. Astfel, o dată cu trecerea timpului (de obicei 30 de secunde), temperatura de extensie creşte, primerul fiind deja parte componentă a noii catene de ADN.

Timpul de extensie, este o altă variabilă importantă. În condiții optime, *Taq* poate lega mii de nucleotide pe minut. Prin urmare, produşii PCR cu o lungime mai mică de

500 perechi de baze (pb) nu necesită mult timp pentru o sinteză completă (aproximativ 30 de secunde fiind suficiente). În general, un timp de extensie de 30 de secunde este suficient pentru produşi sub 50pb, 60 de secunde sunt necesare pentru produşi cu lungimi cuprinse între 500 şi 1500pb şi 90 de secunde sunt necesare pentru produşi cu lungimi mai mari. Optimizarea timpului de elongare este necesară, deoarece un timp prea mare poate favoriza apariția de artefacte (alinieri nespecifice) (Palumbi, 1996).

3.3.2.4. CONDIȚIILE DE REACȚIE

Pentru sistematica moleculară, atenția a fost îndreptată către primeri şi cât de bine se aliniază la matrița de ADN. Dacă alinierea nu este eficientă, sau dacă primerii se aliniază nespecific, amplificarea poate să fie blocată sau se pot amplifica produşi alternativi. Astfel, complementaritatea perfectă primer – matriță, poate determina amplificări cu mare specificitate. Din păcate, temperaturile ridicate, adesea împiedică utilizarea primerilor universali (primeri realizați a fi eficienți pentru un anumit număr de taxoni) (Kocher et al., 1989).

3.3.3. COMPONENTELE PCR

Mediul de reacție este foarte important pentru specificitatea şi eficiența amplificării. Astfel, mediul de reacție trebuie să conțină un tampon, amestecul de nucleotide, polimerază, primeri, ADN şi Mg2+ (un cofactor necesar pentru activitatea enzimei) (Carbonari et al., 1993). Numai primerii şi nucleotidele se consumă în cursul reacției acestea fiind adăugate în exces, astfel încât sinteza să nu poată fi limitată de aceste componente.

Cofactorul, nu se consumă în timpul reacției şi nici nu este degradat de temperaturile extreme ale ciclurilor de amplificare, astfel încât concentrațiile inițiale şi finale sunt aceleaşi. Ionul, este un important cofactor în cataliza enzimatică a reacției de sinteză, astfel, o concentrație optimă poate accelera considerabil desfăşurarea reacției. Oricum, Mg2+ interacționează cu sarcinile negative ale grupării fosfat ale nucleotidelor, suficient de puternic încât Mg2+ să atragă grupările PO4- şi să rămână în cantitate destul de mică pentru a mai reacționa ca şi cofactor. Din acest motiv, concentrația de Mg2+ trebuie să fie mai mare decât concentrația dNTP.

Varierea concentrației de Mg2+ este una dintre cele mai comune metode de stabilire a condițiilor de reacție pentru PCR. De obicei se utilizează o cantitate de 1,5mM MgCl2.

3.3.4. AMPLIFICAREA GENICĂ LA REPREZENTANȚI AI FAMILIEI *CYPRINIDAE*

În cazul celor două specii analizate, a indivizilor ginogenetici şi a hibrizilor interspecifici, au fost amplificate 3 secvențe aparținând la trei gene diferite. Astfel, au fost amplificate regiunile cu hipervariabilitate aparținând citocromului b şi regiunii de control a ADNmt (D-loop), precum şi o regiune conservativă necodificatoare la speciile analizate, segmentul 4L al dehidrogenazei (ND4L).

3.3.4.1. AMPLIFICAREA GENEI CARE DETERMINĂ SINTEZA CITOCROMULUI B

În cazul speciilor analizate, a fost amplificat citocromul b diferențiat, in funcție de primerii utilizați şi condițiile de reacție specifice. La *Cyprinus carpio* L şi hibrizii intraspecifici analizați, datorită lungimii mari a acestei gene, de aproximativ 1140pb, a fost necesară amplificarea în două etape, pe secvențe diferite, utilizându-se două perechi de primeri (Figurile 16 şi 17).

H15149 5`-AAACTGCAGCCCTCAGAATGATATTTGTCCTCA-3`
L14724 5`-CGAAGCTTGATATGAAAAACCATCGTTG-3`
} Primerii pentru primul segment

H15910 5`-ATCTTCGGTTTACAAGACCGGT-3`
L15138 5`-ATGATGACCGCCTTCGTAGGCTA-3`
} Primerii pentru al doilea segment

Figura 16 Primerii utilizați pentru amplificarea citocromului b, la *Cyprinus carpio*

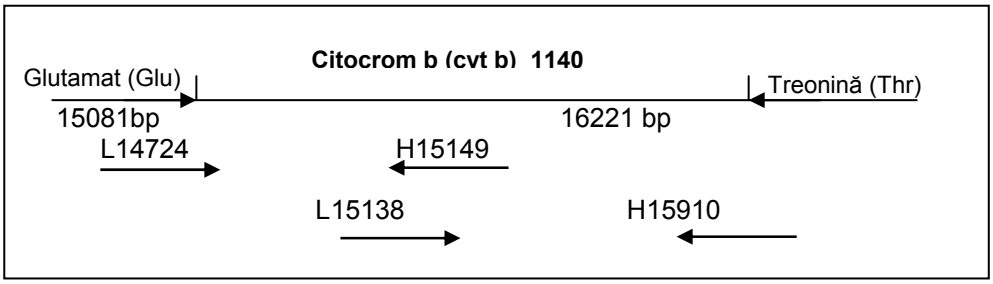


Figura 17 Poziția primerilor utilizați pentru amplificarea genei la *Cyprinus carpio* L şi hibrizii intraspecifici

Pentru prima parte a citocromului b, în cazul tuturor indivizilor aparținând genului *Cyprinus*, a fost folosit acelaşi mediu de reacție, precum şi aceleaşi condiții de amplificare. Referitor la mediul de reacție, prepararea acestuia se realizează într-un tub Eppendorf steril de 1,5ml, în care se adaugă toate componentele utilizate în amplificare, ținând cont de cantitățile necesare corelate cu numărul de probe plus 2 (pentru controlul negativ şi o rezervă de reactivi). Astfel, se multiplică cantitățile necesare, cu numărul de probe (Tabelul 6). Controlul negativ (C-) în toate cazurile, conține acelaşi mediu de reacție utilizat pentru toate celelalte probe, exceptând ADN folosit drept matriță pentru amplificare, înlocuit în acest caz cu H2O miliQ (apă bidistilată, filtrată prin filtru de carbon şi schimbător de ioni). Este utilizat pentru decelarea eventualelor contaminări ale mediului de reacție.

Primerii utilizați pentru amplificarea primului segment, sunt L14724 (Pääbo, 1990) pentru catena directă şi H15149 (Kocher et al., 1989)

Tabel 6 – Compoziția mediului de reacție şi cantitățile corespunzătoare de reactivi pentru amplificarea cytb prima parte la specii ale genului *Cyprinus*

Reactiv	Concentrația	Cantitatea pentru o probă (µl)	Nr. probe	Amestec final
ddH2O-miliQ	-	33,2	n	33,2n
PCR buffer	10X	4,5	n	4,5n
dNTPs	8Mm	5,0	n	5,0n
MgCl2	25mM	2,0	n	2,0n
Primer 1 (L14724)	2mM	1,5	n	1,5n
Primer 2 (H15149)	2mM	1,5	n	1,5n
Taq polimeraza	5000U/ml	0,3	n	0,3n

Compoziția soluțiilor din mediul de reacție:

1. 10X PCR buffer: 500mM KCl, 100mM Tris HCl (pH=8,4 la 20ºC), 15mM MgCl2, 1mg/ml gelatină.
2. 10X dNTP stoc: 2mM dATP, 2mM dCTP, 2mM dGTP, 2mM dTTP (neutralizat la pH=7 cu NaOH).
3. 10X Primeri: primerul liofilizat 2,5µM dizolvat în TE (10mM Tris, 0,1mM EDTA, pH=8).
4. *Taq* polimeraza: 5000 unități/ml.

Procesul de amplificare a primei părți a citocromului b a urmat acelaşi program pentru toate probele analizate aparținând genului *Cyprinus*, având o temperatură de aliniere a primerilor de 45ºC timp de 1 minut şi 30 cicluri de replicare.

Pentru al doilea segment al citocromului b în cazul speciilor şi hibrizilor aceluiaşi gen – *Cyprinus* au fost efectuate modificări ale temperaturii de aliniere a primerilor (pentru mărirea procentului de amplificări specifice) care a fost ridicată la 50ºC, timp de 1 minut, întregul proces având 30 cicluri de replicare. De asemenea, au fost schimbați primerii, în acest caz utilizându-se L15138 (Chang et al., 1994) pentru catena directă şi H15910 (Chang et al., 1994; Tatsuo et al., 2004) pentru catena complementară (Tabel 7).

Tabel 7 – Compoziția mediului de reacție şi cantitățile corespunzătoare de reactivi pentru amplificarea cytb a 2-a parte la specii ale genului *Cyprinus*

Reactiv	Concentrația	Cantitatea pentru o probă (µl)	Nr. probe	Amestec final
ddH2O-miliQ	-	33,2	n	33,2n
PCR buffer	10X	4,5	n	4,5n
dNTPs	8mM	5,0	n	5,0n
MgCl2	25mM	2,0	n	2,0n
Primer 1 (L15138)	2mM	1,5	n	1,5n
Primer 2 (H15910)	2mM	1,5	n	1,5n
Taq polimeraza	5000U/ml	0,3	n	0,3n

În cazul speciilor genului *Carassius*, amplificarea citocromului b s-a desfăşurat într-o singură etapă, datorită imperfecțiunii de amplificare a primerilor intermediari. Astfel, pentru specificitatea amplificării şi pentru a mări cantitatea de produs obținut, au fost modificate cantitatea de *Taq* polimerază din amestecul de reacție de la 0,3µl la 0,5µl (Tabelul 8), iar numărul de cicluri a fost modificat la 40 etape de amplificare.

Primerii utilizați pentru amplificarea acestui fragment, sunt Glu-F şi Thr-R (Zardoya şi Doadrio, 1998), având ca regiuni complementare, secvențe nucleotidice din afara citocromului b şi anume, genele care codifică sinteza acidului glutamic (Glu) şi a treoninei (Thr). (Figura 18).

Tabel 8 – Compoziția mediului de reacție şi cantitățile corespunzătoare de reactivi pentru amplificarea cytb la specii ale genului *Carassius*

Reactiv	Concentrația	Cantitatea pentru o probă (µl)	Nr. probe	Amestec final
ddH2O-miliQ	-	33	n	33n
PCR buffer	10X	4,5	n	4,5n
dNTPs	8mM	5,0	n	5,0n
MgCl2	25mM	2,0	n	2,0n
Primer 1 (Glu-F)	2mM	1,5	n	1,5n
Primer 2 (Thr-R)	2mM	1,5	n	1,5n
Taq polimeraza	5000U/ml	0,5	n	0,5n

Glu-F 5`-GAAGAACCACCGTTGTTATTCAA-3`
Thr-R 5`-ACCTCCRATCTYCGGAGGACA-3`

Figura 18 Structura primerilor utilizați în amplificarea cytb la *Carassius gibelio* Bloch.

3.3.4.2. AMPLIFICAREA REGIUNII DE CONTROL MITOCONDRIAL

Replicarea ADNmt este inițiată pe o singură catenă, în interiorul regiunii de control, (regiune necodificatoare). Pe măsură ce această catenă îşi sintetizează catena complementară, cealaltă catenă a duplexului inițial, este liberă, formând o buclă monocatenară numită D-loop (*engl*. displacement lopp). Când replicarea a avansat pentru aproximativ două treimi din suprafața moleculei, este inițiată replicarea şi la nivelul buclei, până când se va forma un nou genom circular bicatenar (Clayton, 1982, Abilock şi Stephenson, 2001).

În cazul speciei *Cyprinus carpio* L. şi a varietății crap oglindă, a fost folosit un mediu de reacție standard (Tabel 9) şi primeri universali specifici regiunii de control mitocondrial control A pentru catena directă şi control E pentru catena reversă (Figura 19). Reacția de amplificare, a decurs după un program a cărui temperatură de aliniere a primerilor a fost de 62ºC, la un număr de 40 cicluri de amplificare. Temperatura ridicată utilizată pentru alinierea primerilor, s-a datorat dificultăților de amplificare, datorate alinierilor imperfecte concretizate în benzi multiple; prin mărirea temperaturii, se măreşte specificitatea alinierilor.

Tabel 9 – Compoziția mediului de reacție şi cantitățile corespunzătoare de reactivi pentru amplificarea d-loop

Reactiv	Concentrația	Cantitatea pentru o probă (µl)	Nr. probe	Amestec final
ddH2O-miliQ	-	33,2	n	33,2n
PCR buffer	10X	4,5	n	4,5n
dNTPs	8mM	5,0	n	5,0n
MgCl2	25mM	2,0	n	2,0n
CTRL A	2mM	1,5	n	1,5n
CTRL E	2mM	1,5	n	1,5n
Taq polimeraza	5000U/ml	0,3	n	0,3n

CTRL A 5`-TTCCACCTCTAACTCCCAAAGCTAG-3`
CTRL E 5`-CCTGAAGTAGGAACCAGATG-3`

Figura 19 Structura primerilor utilizați în amplificarea regiunii de control mitocondrial la speciile analizate de ciprinide

Pentru specii aparținând genului *Carassius*, a fost utilizat acelaşi mediu de reacție şi aceiaşi primeri ca în cazul speciilor genului *Cyprinus*. Referitor la parametrii programului de amplificare, acesta s-a desfăşurat la o temperatură de aliniere a primerilor de 60ºC şi 30 cicluri de replicare.

3.3.4.3. AMPLIFICAREA GENEI CARE CODIFICĂ SUBUNITATEA 4L A DEHIDROGENAZEI (ND4L)

NADH dehidrogenaza prezintă 7 subunități pe toată suprafața ADNmt. A fost amplificată şi secvențiată subunitatea 4L, deoarece este considerată a fi o regiune puternic conservativă, astfel încât eventualele substituții nucleotidice de la acest nivel, ne vor putea confirma (în cazul apariției acestor substituții) diferențe decelate în cadrul celorlalte gene secvențiate, hipervariabile – citocromul b şi d-loop.

La speciile genului *Cyprinus,* pentru amplificarea genei, s-a folosit un amestec de reacție modificat prin dublarea cantității de *Taq* polimerază de la 0,3µl, la 0,6µl (Tabelul 9). Pentru amplificarea segmentului dorit, s-au folosit doi primeri degenerați L10420 pentru catena directă şi H10720 pentru reversă (Figura 20)

Reacția de amplificare a genei s-a desfăşurat după un program care avea temperatura de aliniere a primerilor de 51ºC şi un număr de 40 cicluri de replicare.

În cazul speciilor genului *Carassius*, pentru amplificarea genei codificatoare a ND4L, a fost folosit acelaşi mediu de reacție (inclusiv aceiaşi primeri) şi aceleaşi condiții de amplificare ale programului PCR ca în cazul speciilor genului *Cyprinus*.

Tabel 10 – Compoziția mediului de reacție şi cantitățile corespunzătoare de reactivi pentru amplificarea ND4L

Reactiv	Concentrația	Cantitatea pentru o probă (µl)	Nr. probe	Amestec final
ddH2O-miliQ	-	31,9	n	31,9n
PCR buffer	10X	4,5	n	4,5n
dNTPs	8mM	5,0	n	5,0n
MgCl2	25mM	2,0	n	2,0n
CTRL A	2mM	1,5	n	1,5n
CTRL E	2mM	1,5	n	1,5n
Taq polimeraza	5000U/ml	0,6	n	0,6n

L10420 5`-AAYAARAYCNTTGATTTCGRCTCA-3`
H10720 5`-TCYGTGCCRTGYGTYCGNGC-3`

Figura 20 Structura primerilor utilizați în amplificarea ND4L la ciprinide

3.4. TESTAREA PRODUŞILOR PCR PRIN ELECTROFOREZĂ

La sfârşitul fiecărei reacții de amplificare, este necesară o verificare a produşilor obținuți pentru a observa dacă a avut loc amplificarea, dacă au avut loc amplificări multiple sau pentru decelarea eventualelor contaminări.

Soluții necesare:

- agaroză SEAKEM;
- tampon TBE: - soluție stoc 10X (108g Tris base, 55g acid boric, 40ml 0,5M EDTA (pH=8,0), se completează cu apă bidistilată până la 1l) şi 1X soluție de lucru (89mM Trios base, 89mM acid boric, 2mM EDTA);
- loading buffer: 20% glicerol, 1mg/ml bromfenol albastru, 1mg/ml xilencianol;
- soluție bromură de etidiu: pentru 1000X soluție stoc, cu o concentrație 0,5mg/ml sunt necesare: 50mg bromură de etidiu, 100ml H2O bidistilată. Se diluează soluția stoc în proporție 1:1000.
- marker molecular de 100pb

Concentrația gelului de agaroză depinde de mărimea fragmentelor de ADN care urmează a fi evidențiate prin electroforeză (Tabelul11) (Ausubel et al., 1992)

Tabel 11 Concentrația gelului de agaroză corelată cu mărimea fragmentului de ADN

Concentația gelului de agaroză (%)	Lungimea fragmentului de ADN (kb)
0,5	1 – 30
0,7	0,8 – 12
1,0	0,5 – 10
1,2	0,4 – 7
1,5	0,2 – 3

Având valori aproximative ale lungimilor genelor amplificate (valori provenite din banca de gene – GenBank), pentru verificarea produşilor PCR obținuți, au fost folosite geluri de agaroză având concentrații de 1,5% şi volume de 30ml. Prepararea unui

astfel de gel, necesită o cantitate de 0,45g agaroză (SEAKEM) la 30ml TBE 1X. Astfel, se adaugă soluția de TBE 1X peste agaroză, se amestecă până la omogenizare şi se introduce în cuptorul cu microunde pentru aproximativ 20 de secunde. După expirarea timpului, se scoate, se agită şi se reintroduce în cuptor pentru alte 20 de secunde. Flaconul ce conține gelul, se răceşte în jet de apă; după răcire, se introduc în flacon 1,5µl bromură de etidiu şi se agită până la dispariția culorii roşii a bromurii.

Se toarnă gelul în suportul pentru gel, se introduce un pieptene şi se lasă să se întărească. (Notă: cuva de electroforeză trebuie să fie aşezată pe un loc perfect plan înaintea turnării gelului). După întărire, se scoate pieptenele prin culisare şi se aşează gelul în cuva de electroforeză care conține tampon TBE sau TAE 1X.

Pregătirea probelor pentru migrarea în gel, necesită un fag, în care se introduc câte 2,5µl loading buffer pentru fiecare probă, plus aceeaşi cantitate pentru markerul molecular de 100pb şi controlul negativ. În continuare, peste buffer se adaugă câte 10µl din fiecare probă (produs PCR). Se omogenizează conținutul fiecărei celule a fagului, în momentul introducerii probei. După realizarea amestecurilor, se încarcă probele în gelul de agaroză, păstrând ordinea, întotdeauna pe ultima poziție disponibilă de încărcare, fiind situat controlul negativ (C-).

Electroforeza produşilor PCR a fost efectuată în cuve tip FOTODYNE, într-un câmp electric de 90V şi 50 – 60mA, timp de 40 – 50 de minute, în funcție de mărimea fragmentelor.

3.5. PURIFICAREA PRODUŞILOR PCR

Purificarea tuturor produşilor PCR obținuți s-a realizat utilizând un kit de reacție QIAGEN (QIAquick PCR purification kit) şi un protocol furnizat de aceeaşi firmă. Acestea sunt utilizate pentru purificarea directă a produşilor PCR mono- sau bicatenari cu lungimi cuprinse între 10pb – 10kb de primerii neataşați, nucleotide libere, polimeraze şi săruri. Kit-ul conține coloane de purificare, tampon PB, tampon PE, tampon EB (producătorul nu furnizează compoziția fiecărui tampon) şi tuburi colectoare.

Ca primă etapă a purificării, se adaugă 5 volume tampon PB la 1 volum produs PCR şi se amestecă. Nu este necesară îndepărtarea uleiului mineral adăugat în tuburi pentru a împiedica evaporarea în timpul amplificării). Se ataşează coloanele la tuburile colectoare (tuburi cu un volum de 2ml). Produşii PCR amestecați cu alcool sunt introduşi în coloane şi se centrifughează la 13000 rotații/minut (rpm), timp de 30 – 60 secunde, utilizând o centrifugă de masă. Se descarcă tuburile colectoare, plasând din nou coloanele în aceleaşi tuburi.

Pentru etapa de spălare, se adaugă 0,75ml tampon PE în coloane şi se centrifughează la 13000 rpm timp de 30 – 60s. Se golesc tuburile colectoare, şi se centrifughează din nou coloanele timp de 1 minut la 14000 rpm, pentru a elimina complet alcoolul din probe. Notă: etanolul rezidual nu va fi eliminat complet dacă nu este golit tubul colector înaintea acestei ultime centrifugări.

Se aşează coloanele în tuburi Eppendorf sterile de 1,5ml. Pentru a dizolva ADN din coloane, se adaugă 50µl tampon EB (10mM Tris-Cl, pH=8,5) sau H2O miliQ în centrul membranei şi se centrifughează coloanele, timp de 1 minut la 13000 rpm. Pentru a creşte concentrația de ADN, se adaugă numai 30µl tampon EB în centrul membranei, se lasă în repaus 1 minut şi se centrifughează.

3.6. CUANTIFICAREA PRODUŞILOR PCR

3.6.1. PREPARAREA GELULUI DE CUANTIFICARE

Pentru cuantificarea ADN din probele purificate în coloane QIAGEN, se va prepara un gel de agaroză 3%, cu un volum de 30ml. Astfel, pentru un volum de 30ml sunt necesari 3ml TBE 10X care se diluează cu 27ml H2O bidistilată.

În prepararea gelului sunt utilizate două tipuri de agaroză, cu densități diferite; Seakem : Nusieve în raport de 2 : 1., adică, 0,6g Seakem : 0,3g Nusieve. Peste acest amestec (0,9g agaroză) se adaugă soluția TBE 1X, se agită până la omogenizare şi se introduce în cuptorul cu microunde, unde se lasă aproximativ 20 de secunde; se scoate flaconul, se agită pentru o mai bună omogenizare şi se introduce din nou în cuptor pentru alte 20 de secunde. Se răceşte flaconul în jet de apă, se introduc 1,5µl bromură de etidiu şi se agită flaconul până la dispariția culorii roşii a bromurii de etidiu.

Se toarnă gelul într-o cuvă de gel, se ataşează un pieptene şi se lasă să se răcească. După răcire, se scoate pieptenele şi se introduce gelul în cuva de electroforeză.

3.6.2. PREGĂTIREA PROBELOR PENTRU CUANTIFICARE

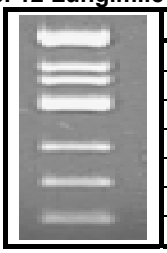
Amestecul pentru electroforeză, ca şi în cazul testării produşilor PCR, se realizează într-un fag. În fiecare celulă a fagului (corespunzător numărului de probe plus una pentru marker) se introduc câte 2µl loading buffer. Peste bufferul de încărcare a probelor, se adaugă câte 5µl din fiecare mostră de ADN purificat şi 8µl marker ØX174-*Hinf I*. Se omogenizează conținutul fiecărei celule a fagului, în momentul introducerii probei. După realizarea amestecurilor, se încarcă probele (amestecul corespunzător fiecărei probe) în gelul de electroforeză.

Migrarea probelor în gel, s-a realizat în cuve de electroforeză tip FOTODYNE, într-un câmp electric de 90V, 50 – 60mA, timp de 40 – 70 minute, în funcție de mărimea fragmentelor.

3.6.3 CUANTIFICAREA PROBELOR

Markerul molecular este reprezentat de fagul ØX174, cu o lungime de 3500pb. *Hinf I* este o enzimă care taie macromolecula de ADN a fagului, dând fragmente cu lungimi (măsurate în pb) şi greutăți cunoscute (Tabel 12). Concentrația fagului în marker este de 0,0375µg/µl fag; dacă se utilizează 8µl marker, atunci vom avea 0,3µg/µl fag.

Tabel 12 Lungimile şi greutățile fragmentelor de ADN ale fagului ØX174



ØX174-*Hinf I* (pb)	Greutatea fragmentului (ng)
726/713	80
533	29,7
500	28
427/417/413	70
311	17
249	13,9
200	11

Cuantificarea se face prin compararea intensității benzilor date de probele purificate cu intensitatea benzilor date de marker. Astfel, pentru fiecare bandă de ADN, se va alege câte o bandă a marker-ului (pe criteriul similarității intensităților), iar datele vor fi notate într-un Tabel sintetic.

Concentrația ADN (ng/µl) este dată de raportul dintre greutatea (ng) benzii respective şi cantitatea (µl) de ADN purificat încărcată în gel (5µl). cantitatea de ADN (µl) necesară secvențierii se aproximează astfel încât să obținem 60 – 90fmoli ADN dublu catenar, conform Tabelului 13:

De exemplu, pentru a obține 50fmol ADN cu o lungime de 500pb, sunt necesare 16ng ADN. Ştiind câte ng sunt necesare pentru a obține un anumit număr de fmoli, putem aproxima cantitatea de ADN dublu catenar pe care o vom utiliza.

Pentru a obține cantitatea de fmoli, se înmulțeşte cantitatea de ADN aproximată, cu concentrația ADN (ng/µl) corespunzătoare. În continuare se face echivalența cu numărul de fmoli corespunzători la 33ng pentru un segment de 0,5Kpb, date furnizate de Tabelul 12 (Beckman Coulter Protocol).

Tabel 13 Estimarea concentrației ADN

Lungimea (perechi de kilobaze)	ng pentru 25fmoli	ng pentru 50fmoli	ng pentru 100fmoli
0,2	3,3	6,5	13
0,3	4,9	9,8	20
0,4	6,5	13	26
0,5	8,1	16	33
1,0	16	33	65
2,0	33	65	130
3,0	50	100	195
4,0	65	130	260
5,0	80	165	325
6,0	100	195	390
8,0	130	260	520
10,0	165	325	650
12,0	195	390	780

În continuare, va fi calculată cantitatea de H_2O miliQ necesară pentru fiecare probă în reacția de secvențiere. Astfel, ştiind că pentru fiecare probă se vor folosi 12,5µl amestec de reacție şi 7µl de ADN, ştiind cantitatea reală de ADN care va fi folosită, se va scade această cantitate din 7, obținându-se cantitatea de apă necesară în reacție pentru fiecare probă.

3.7. REACȚIA DE SECVENȚIERE

Probele sunt pregătite pentru reacția de secvențiere, în tuburi de 0,2ml.

Notă: toți reactivii kit-ului de secvențiere trebuiesc decongelați pe gheață, iar probele de ADN pentru secvențiere, trebuiesc de asemenea păstrate permanent pe mediu cu gheață.

Ca şi în cazul reacției de amplificare genică în PCR, în cazul reacției de secvențiere se foloseşte un kit pe baza căruia se prepară mediul de reacție (Tabel 14).

Tabel 14 Compoziția mediului pentru reacția de secvențiere

Reactivi	Cantități (µl)
10X buffer	2
dNTP mix	1
ddATP	2
ddGTP	2
ddCTP	2
ddTTP	2
Primer (10µM)	0,5
ADN polimerază	1

Cantitatea totală a amestecului de reacție este de 12,5µl, iar volumul final al reacției de secvențiere este de 20µl; restul de 7,5µl este constituit din ADN sau ADN+H2O. Odată calculate cantitățile de ADN necesare pentru fiecare probă, astfel încât să obținem între 60 – 90fmoli, iar dacă această cantitate este mai mică de 7,5µl, se adaugă H2OmiliQ sterilizată, până când volumul final de ADN plus apă, ajunge la 7,5µl.

După adăugarea în toate tuburile a cantităților de apă necesare, se prepară amestecul de reacție, multiplicând fiecare din cantitățile menționate în Tabelul 14 cu numărul de probe plus 0,5. După prepararea amestecului de reacție, se repartizează câte 12,5µl în fiecare tub de 0,2ml (tuburile trebuiesc menținute pe gheață). În continuare, se adaugă cantitatea calculată de ADN şi se omogenizează cu amestecul de reacție.

3.7.1. REACȚIA DE SECVENȚIERE A CITOCROMULUI B

În cazul tuturor indivizilor aparținând celor două genuri *Carassius* şi *Cyprinus*, au fost folosiți aceiaşi primeri L14724 şi H15149 pentru prima parte a genei, ca şi în cazul reacției de amplificare, cu deosebirea că au fost utilizate concentrații mai mari (10µM) pentru reacția de secvențiere. Programul utilizat, a avut o temperatură de aliniere a primerilor de 45,0ºC şi un număr de 30 cicluri de replicare.

Reacția de secvențiere pentru al doilea fragment al cyt b, ca şi în cazul amplificării, a decurs diferențiat pentru speciile celor două genuri. În cazul speciilor genului *Cyprinus*, s-au folosit primerii L15138 şi H15910 în concentrație 10µM, urmând un program de secvențiere cu temperatura de aliniere a primerilor de 47ºC.

În cazul speciilor genului *Carassius*, reacția de secvențiere a ultimei părți a genei ce codifică sinteza citocromului b, s-a desfăşurat utilizând primerii L15138 pentru catena directă şi Cyt-R pentru catena complementară (Figura 21), în concentrații de 10µM. Reacția s-a desfăşurat după un program a cărui temperatură de aliniere a primerilor a fost de 45ºC şi 30 cicluri de replicare.

L15138	5`-ATGATGACCGCCTTCGTAGGCTA-3`
Cyt-R	5`-TCTTTATATGAGAARTANGGGTG-3`

Figura 21 Structura primerilor utilizați în reacția de secvențiere a părții a doua a cyt b la specii ale genului *Carassius*

3.7.2. REACȚIA DE SECVENȚIERE A REGIUNII DE CONTROL MITOCONDRIAL (D-LOOP)

Pentru reacția de secvențiere a d-loop, în cazul speciilor ambelor genuri au fost folosiți cei doi primeri universali CTRL-A pentru catena directă şi CTRL-E pentru catena antisens, în concentrații de 10μM. De asemeni, programul utilizat pentru reacția de secvențiere, a avut o temperatură de aliniere a primerilor de 50ºC şi 30 cicluri de replicare

3.7.3. REACȚIA DE SECVENȚIERE A ND4L

Pentru secvențierea genei care codifică informația sintezei nicotinamid dehidrogenazei subunitatea 4L, pentru toate speciile analizate, indivizi ginogenetici şi hibrizi, au fost folosiți aceiaşi primeri L10420 pentru catena sens şi H10720 pentru catena antisens, în concentrație 10μM. Programele reacției de secvențiere au fost diferite; pentru speciile genului *Cyprinus*, a fost folosită o temperatură de aliniere a primerilor de 47ºC în 30 cicluri de replicare.

În cazul speciilor genului *Carassius*, programul de secvențiere a avut temperatura de aliniere a primerilor de 50ºC şi 30 cicluri de replicare.

3.8. PRECIPITAREA ÎN ETANOL A PROBELOR PENTRU SECVENȚIERE

Sunt folosite tuburi Eppendorf de 0,5ml care se etichetează cu inițiala numelui genei şi a catenei.

Se prepară soluția STOP (1,5M NaOAc şi 0,05M EDTA). Aceste soluții se prepară în tuburi Eppendorf de 1,5ml în volum de 50μl (sau 100μl pentru precipitarea a 24 de probe). Astfel, se amestecă 25μl de apă miliQ sterilă cu 25μl 3M NaOAc soluție stoc şi 45μl de apă miliQ sterilă cu 5μl de 0,5M EDTA stoc. Se adaugă în fiecare tub câte 5μl soluție STOP care conține 2μl de 1,5M NaOAc, 2μl 0,05M EDTA şi 1μl de glicogen (20mg/ml) care se găseşte în kit-ul de secvențiere.

Pentru a adăuga această soluție STOP în toate tuburile se prepară un amestec într-un tub Eppendorf de 1,5ml, astfel (cantitățile sunt date în μl/probă): 1,5M NaOAc 2μl, 0,05M EDTA 2μl şi glicogen 1μl.

Se înmulțesc aceste cantități cu numărul de probe plus 1 şi se adaugă câte 5μl din acest amestec în fiecare tub etichetat anterior. Se centrifughează câteva secunde tuburile de 0,2ml care conțin reacția de secvențiere şi se transferă tot conținutul (20μl) în tuburile de 0,5ml, unde se află soluția stop şi se omogenizează cu pipeta. Se deschid toate tuburile şi se adaugă câte 65μl etanol 95% rece (-20ºC). Se închid tuburile şi se agită bine.

Se centrifughează imediat la 4ºC şi 14500rpm, timp de 15 minute.

Notă: cea mai importantă etapă este cea de amestecare a ADN cu soluția stop şi alcoolul, urmată de centrifugarea imediată după amestecare.

După centrifugare, se extrage cu pipeta tot lichidul supernatant. Nu trebuie atins sedimentul, pentru a nu-l desprinde de pereții tubului şi astfel se poate folosi acelaşi vârf de pipetă (dacă se atinge sedimentul, vârful pipetei trebuie schimbat). Se adaugă câte 190μl etanol 70% în fiecare tub (alcoolul se introduce în tub, prelingându-l pe pereți pentru a nu fragmenta peleta de ADN).

Se centrifughează la 4ºC şi 14500rpm, timp de 5 minute (alcoolul trebuie menținut la -20ºC). După centrifugare, se elimină cu pipeta tot alcoolul supernatant, urmărind permanent unde se află sedimentul, pentru a nu-l elimina). Se adaugă 190µl etanol 70%, rece, şi se repetă etapa anterioară. După eliminarea alcoolului, se centrifughează (utilizând o centrifugă normală de masă) la 8000rpm, 1 minut. Pentru a sedimenta toate picăturile de alcool, care se află pe pereți. Cu un vârf de pipetă foarte fin (pentru pipete de 10µl) se elimină tot alcoolul din tuburi. Se lasă tuburile deschise 10 – 15 minute pentru a usca sedimentele într-o nişă unde nu pătrunde praf. Se scoate formamida din congelator pentru decongelare. După evaporarea completă a alcolului din tuburi, se adaugă în fiecare 40µl formamidă şi se ametecă pe vortex, timp de 5 secunde. Se centrifughează câteva secunde pentru a colecta toate picăturile de pe pereții tuburilor în partea inferioară a acestora. Se lasă la temperatura camerei, 10 minute. Se amestecă din nou pe vortex 5 secunde şi se centrifughează. Se lasă în repaus la temperatura camerei, timp de 10 minute.

Dacă probele vor fi introduse direct în secvențiator, se transferă câte 40µl pe placa pentru probe. În această etapă, se pipetează de câteva ori pentru a resuspenda ADN în formamidă. Dacă probele nu sunt secvențiate imediat, se păstrează tuburile la -20ºC. Probele nu pot fi păstrate congelate mai mult de o lună.

Secvențierea propriu-zisă s-a realizat utilizând un secvențiator BECKMAN COULTER CEQ8000 cu opt capilare. Pentru toate genele s-a folosit un timp de secvențiere de 120 minute, exceptând partea a doua a citocromului b, care, datorită lungimii (aproximativ 800pb), a necesitat un timp de 160 minute.

3.9. ANALIZA SECVENȚELOR

Pelucrarea primară a secvențelor sub formă de fluorograme şi corectarea, au fost realizate utilizând programul CEQ2000 furnizat de BECKMAN COULTER, iar datele au fost exportate sub formă de text pentru aliniere.

Unirea secvențelor directă şi reversă pentru fiecare individ, a fost efectuată utilizând programul ESEE 32 (The Eyeball Sequence Editor), prin aliniere cu secvențele primerilor şi tăierea celor două catene, urmată de unirea într-o catenă unică.

Alinierea tuturor secvențelor unei gene, provenite de la indivizi diferiți (indivizi aparținând la populații diferite, indivizi obținuți prin ginogeneză, precum şi hibrizi), a fost efectuată prin metodele ClustalW (Thompson et al., 1994) şi Clustal V (Higgins şi Sharp, 1989; Higgins, 1994; Wheeler, 2000) utilizând modulul MegAlign din cadrul programului LaseGene 7.

Compararea secvențelor, ca şi trasarea arborilor filogenetici, a fost realizată prin utilizarea programului MEGA 4 (Nei şi Kumar, 2000).

STUDIUL RELAȚIILOR FILOGENETICE

4.1. EVALUAREA UNIFORMITĂȚII POPULAȚIILOR

Pe baza parametrilor fenotipici înregistrați conform Figurii 11 şi a indicilor fenotipici calculați, utilizând metodele prezentate în capitolul 4.1., au fost obținute valorile pentru loturile analizate, diferențiat pentru cele două genuri *Cyprinus* şi *Carassius*. Astfel, în cazul speciilor genului *Cyprinus*, din datele prezentate în Tabelul 15 şi Figura 22, se poate observa că valorile medii ale erorilor standard sunt destul de scăzute, iar limitele intervalelor de variabilitate (calculate cu o probabilitate de α=0,05 şi $n-1$ grade de libertate, adică, la o valoare a lui t de 2,021) ale parametrilor fenotipici investigați sunt foarte restrânse, ceea ce ne sugerează o omogenitate a loturilor investigate. În plus, în cazul indicilor analizați se constată de asemeni, valori foarte mici ale erorii standard (între 0,009 pentru indicele de circumferință şi 0,027 pentru indicele de profil) şi ale coeficientului de variație (0,38 pentru indicele de profil), fapt care ne confirmă încă o dată uniformitatea fenotipică a indivizilor analizați, cu toate că provin din regiuni geografice diferite (Iaşi şi Larga-Jijia – Movileni) şi au origini diferite (indivizi ginogenetici hibrizi, hibrizi şi indivizi normali).

Tabel 15 Parametrii fenotipici calculaţi, la indivizii genului *Cyprinus*

Populația		Nr.crt.	Lt (cm)	Ls (cm)	Lc (cm)	Ha (cm)	Hd (cm)	G (g)	C (cm)	Ip	Ic K	CF
Larga-Jijia – Movileni	Indivizi cu solzi	1	27,6	24,6	7,3	3,9	9,7	442	20,5	2,55	1,20	2,96
		2	27,0	23,8	7,1	3,7	11,7	413	18,2	2,03	1,31	3,07
		3	27,7	24,9	7,4	3,9	11,5	400	18,8	2,17	1,32	2,60
		4	28,2	24,9	7,5	4,1	10,4	487	19,6	2,40	1,27	3,17
		5	27,4	23,7	7,2	3,9	11,1	408	19,4	2,14	1,23	3,05
	Indivizi fără solzi	6	25,8	22,5	6,8	3,4	10,9	430	19,5	2,06	1,15	3,80
		7	28,3	25,3	7,6	4,1	10,9	452	19,1	2,32	1,32	2,79
		8	27,9	24,8	7,4	4,0	10,9	487	20,1	2,27	1,23	3,20
		9	28,6	25,7	7,8	4,2	10,3	463	19,5	2,49	1,31	2,74
		10	27,8	24,6	7,1	4,1	9,5	451	20,6	2,59	1,19	3,03
		11	25,3	21,7	6,3	3,2	10,9	358	19,1	1,99	1,14	3,51
		12	29,1	25,9	7,6	4,2	11,5	506	20,6	2,25	1,26	2,90
		13	26,9	22,8	6,7	3,4	11,9	351	18,4	1,92	1,24	2,96
		14	29,3	25,6	7,2	4,2	9,8	345	20,5	2,62	1,25	2,06
		15	26,0	21,4	6,0	3,2	11,2	502	18,3	1,91	1,17	5,15
	Indivizi ginogenetici	16	28,1	24,7	7,2	3,9	10,0	369	19,0	2,46	1,30	2,44
		17	26,2	21,9	6,2	3,4	9,6	372	20,3	2,29	1,08	3,53
		18	29,2	25,7	7,6	4,1	10,9	500	20,6	2,35	1,24	2,96
		19	27,9	22,1	6,4	3,9	10,3	485	19,3	2,14	1,14	4,51
		20	25,6	24,7	7,2	3,2	11,2	475	19,8	2,20	1,25	3,15
	Crap oglindă	21	25,4	22,9	6,7	3,4	9,8	402	18,5	2,34	1,24	3,36
		22	29,2	25,5	7,1	4,2	10,5	427	20,4	2,44	1,25	2,57
		23	26,9	22,6	6,6	3,6	10,2	436	18,6	2,21	1,22	3,78
		24	26,2	22,3	6,2	3,4	11,1	341	18,5	2,02	1,20	3,06
		25	26,2	22,4	6,3	3,4	10,6	358	18,8	2,11	1,19	3,18
Iaşi		26	28,1	24,8	7,2	3,8	11,0	343	18,8	2,25	1,32	2,25
		27	28,1	24,1	6,1	4,0	11,3	403	19,5	2,12	1,23	2,89
		28	29,1	25,9	7,4	4,3	11,3	437	20,5	2,30	1,26	2,52
		29	27,5	23,7	6,6	3,6	11,3	389	19,3	2,10	1,23	2,93
		30	26,9	22,8	6,2	3,4	10,2	454	18,9	2,25	1,21	3,83
		31	28,0	24,3	7,6	4,1	9,8	507	20,1	2,49	1,21	3,53
		32	27,2	24,0	7,6	3,8	10,6	360	20,0	2,26	1,20	2,60
		33	25,3	21,8	6,1	3,2	9,3	347	20,5	2,34	1,07	3,35
		34	26,9	22,7	6,4	3,5	9,8	475	19,5	2,32	1,16	4,06
		35	28,5	24,1	7,2	4,0	11,0	406	18,3	2,19	1,31	2,92
		36	25,7	22,3	6,1	3,5	10,4	450	18,5	2,15	1,21	4,06
		37	27,4	24,6	7,3	4,0	11,3	464	19,7	2,18	1,25	3,11
		38	27,2	24,0	7,1	4,0	9,4	398	18,2	2,55	1,32	2,88
		39	28,4	23,5	6,8	4,1	9,3	410	19,9	2,51	1,18	3,16
		40	26,4	23,0	6,6	3,7	11,5	410	18,7	2,01	1,23	3,37
		41	25,4	21,9	6,1	3,5	9,6	368	19,6	2,27	1,11	3,52

Populația	Nr.crt.	Lt (cm)	Ls (cm)	Lc (cm)	Ha (cm)	Hd (cm)	G (g)	C (cm)	Ip	Ic K	CF
	42	26,7	24,0	7,4	3,9	11,4	501	18,8	2,11	1,28	3,64
	43	29,3	25,8	7,7	4,2	10,5	341	20,3	2,47	1,27	1,99
	44	28,3	24,9	7,5	3,9	9,7	451	20,2	2,57	1,23	2,92
	45	29,9	25,6	7,3	4,6	11,1	463	20,3	2,31	1,26	2,76
	46	28,5	24,4	7,0	3,8	11,4	398	20,1	2,14	1,22	2,73
	47	26,0	22,6	6,8	3,4	11,6	404	19,3	1,95	1,17	3,48
	48	26,7	22,3	6,7	3,6	10,9	469	19,0	2,06	1,17	4,21
	49	25,4	22,2	6,7	3,3	11,3	497	18,7	1,96	1,19	4,52
	50	28,1	24,7	7,3	3,9	9,7	344	20,0	2,53	1,23	2,28
Media		27,3	23,8	6,95	3,78	10,6	423	19,4	2,25	1,22	3,18
Deviația standard		1,25	1,34	0,53	0,35	0,73	52,7	0,76	0,19	0,06	0,64
Coeficient de variație		2,50	2,67	1,05	0,69	1,46	105,	1,52	0,38	0,12	1,27
Eroarea standard		0,17	0,18	0,07	0,04	0,10	7,46	0,1	0,02	0,01	0,09
Limita superioară a intervalului de confidență		27,7	24,1	7,30	4,14	10,9	423,	19,8	2,61	1,58	3,53
Limita inferioară a intervalului de confidență		27	23,4	6,59	3,42	10,2	422	19	1,89	0,86	2,82

Ip=indice profil; Ic K=indice de circumferință Kieslev; CF= coeficient Fulton; Lt=lungimea totală, Ls=lungime standard, Lc=lungimea capului, Ha=înălțimea la nivelul înotătoarei anale, Hd=înălțimea la nivelul înotătoarei dorsale, G=greutate, C=circumferință

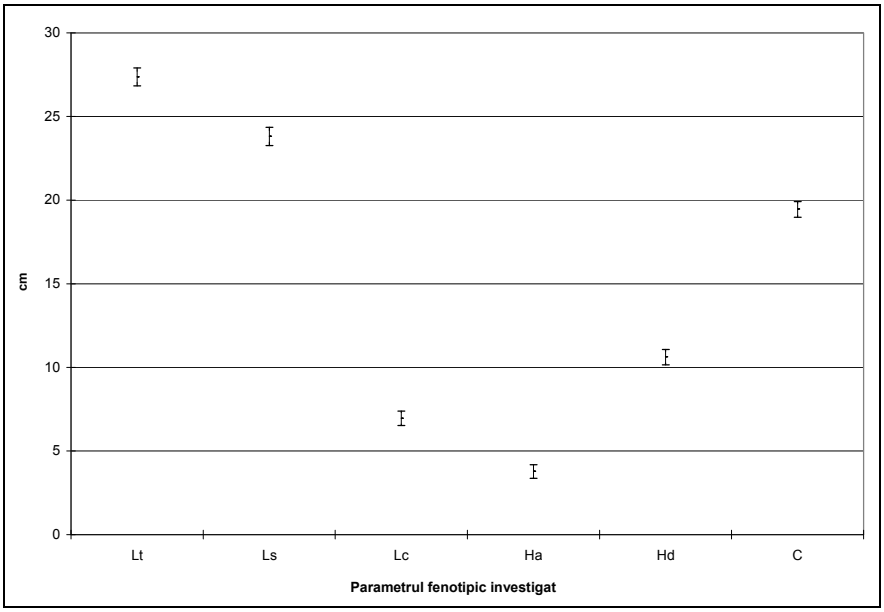


Figura 22 Intervalele de variabilitate ale parametrilor fenotipici investigați la speciile genului *Cyprinus*

Analizând datele observate în cazul speciilor genului *Carassius* (Tabelul 16), putem constata că valorile erorii standard sunt foarte mici, ceea ce semnifică o grupare a tuturor datelor foarte aproape de valoare medie pe probă şi parametru investigat. Pe baza erorii standard a mediei şi a valorii lui *t* dată de *α=0,05* (adică, o probabilitate de 95%) şi *n-1* grade de libertate (unde *n* reprezintă numărul de valori din fiecare probă), s-au calculat limitele intervalelor de variabilitate pentru toți parametrii şi indicii fenotipici investigați. Din reprezentarea grafică a intervalelor de variabilitate (Figura 23), se observă limitele reduse ale tuturor parametrilor, ceea ce semnifică o uniformitate a indivizilor analizați. De asemenea, analizând indicele de profil, se constată că acesta variază între limite foarte înguste la cei 50 de indivizi (2,19 – 3,19cm). Chiar şi pentru coeficientul Fulton, valorile sunt destul de restrânse, fiind cuprinse între 4,77 – 9,19g/cm. Toate aceste valori, ne confirmă încă o dată că loturile analizate sunt uniforme din punct de vedere morfologic.

Tabel 16 Parametrii fenotipici calculați, la indivizii genului *Carassius*

Populația	Nr. crt.	Lt (cm)	Ls (cm)	Lc (cm)	Ha (cm)	Hd (cm)	G (g)	C (cm)	Ip	IK	CF
Larga-Jijia - Movileni	1	18,2	15,8	4,0	2,0	5,9	257	15,3	2,68	1,03	6,52
	2	19,5	15,6	3,7	2,4	6,3	274	15,2	2,48	1,03	7,22
	3	19,2	15,9	4,1	2,2	6,1	252	14,7	2,61	1,08	6,27
	4	21,7	17,8	4,4	2,6	6,9	371	17,6	2,58	1,01	6,58
	5	18,5	14,7	4,1	1,9	5,2	248	15,2	2,83	0,97	7,81
	6	19,2	15,6	4,1	2,3	5,5	262	14,6	2,84	1,07	6,90
	7	19,9	16,6	4,2	2,3	6,2	268	15,8	2,68	1,05	5,86
	8	17,3	14,4	3,6	1,4	5,0	243	14,4	2,88	1,00	8,14
	9	18,3	14,6	3,8	2,0	5,9	260	15,6	2,47	0,94	8,35
	10	21,0	17,2	4,4	2,4	6,7	342	16,8	2,57	1,02	6,72
	11	20,7	17,2	4,3	2,3	6,5	309	16,3	2,65	1,06	6,07
	12	18,6	15,5	4,3	2,1	5,9	249	15,7	2,63	0,99	6,69
	13	20,6	16,7	4,3	2,3	6,6	284	16,3	2,53	1,02	6,10
	14	20,0	16,4	4,8	2,0	5,8	289	15,6	2,83	1,05	6,55
	15	20,2	16,7	4,4	2,1	6,3	277	16,1	2,65	1,04	5,95
	16	20,1	16,6	4,1	2,4	6,6	280	16,6	2,52	1,00	6,12
	17	18,6	15,3	3,9	2,0	5,7	261	14,8	2,68	1,03	7,29
	18	21,3	17,3	4,9	2,4	6,7	340	17,0	2,58	1,02	6,57
	19	18,4	15,4	4,0	1,8	5,4	264	14,0	2,85	1,10	7,23
	20	18,4	15,2	3,6	2,1	5,9	263	14,3	2,58	1,06	7,49
	21	18,1	15,0	3,9	2,0	6,1	266	14,8	2,46	1,01	7,88
	22	18,7	15,3	3,9	2,0	5,7	266	14,8	2,68	1,03	7,43
	23	17,7	14,4	3,6	1,9	5,4	243	13,2	2,67	1,09	8,14
	24	18,0	14,8	4,1	1,9	5,5	237	14,2	2,69	1,04	7,31
	25	19,2	15,7	4,1	2,1	6,0	250	16,1	2,62	0,98	6,46
Iaşi	26	19,4	16,2	4,1	2,1	6,2	255	15,3	2,61	1,06	6,00

Populația	Nr. crt.	Lt (cm)	Ls (cm)	Lc (cm)	Ha (cm)	Hd (cm)	G (g)	C (cm)	Ip	IK	CF
	27	18,8	15,4	4,1	2,0	5,9	264	14,8	2,61	1,04	7,23
	28	17,1	13,9	3,7	2,0	5,6	231	14,3	2,48	0,97	8,60
	29	18,9	15,9	3,6	2,0	5,8	250	15,2	2,74	1,05	6,22
	30	17,3	14,2	3,9	2,1	5,6	243	14,3	2,54	0,99	8,49
	31	18,0	14,8	3,9	1,9	5,9	247	14,6	2,51	1,01	7,62
	32	17,4	15,4	4,3	1,4	5,4	269	15,1	2,84	1,02	7,43
	33	20,0	15,1	4,1	2,6	5,6	283	14,3	2,70	1,06	8,24
	34	17,9	14,9	4,5	2,0	5,5	243	15,7	2,70	0,95	7,38
	35	19,5	14,5	3,7	2,3	5,5	253	16,6	2,62	0,87	8,31
	36	18,1	16,9	4,1	2,4	6,6	257	16,6	2,56	1,02	5,37
	37	17,3	16,3	4,2	1,4	6,7	244	16,8	2,43	0,97	5,68
	38	20,7	16,2	3,8	2,0	5,8	273	14,0	2,82	1,16	6,36
	39	18,6	16,8	3,8	1,4	6,4	253	15,2	2,61	1,11	5,31
	40	19,6	17,0	4,4	2,6	5,3	268	16,5	3,19	1,03	5,40
	41	19,6	16,6	4,2	1,4	5,8	331	14,4	2,86	1,15	7,29
	42	18,7	16,4	3,9	2,6	6,1	289	14,7	2,68	1,12	6,52
	43	18,6	17,1	3,7	2,4	6,6	240	14,1	2,58	1,21	4,77
	44	17,6	16,7	4,8	1,9	5,9	237	14,1	2,82	1,19	5,09
	45	19,4	16,0	4,6	1,4	5,5	302	14,5	2,90	1,11	7,37
	46	18,3	16,3	4,6	1,5	6,2	243	14,6	2,63	1,12	5,57
	47	21,1	14,6	4,5	1,6	6,7	287	14,0	2,19	1,04	9,19
	48	20,5	16,8	4,5	2,4	5,8	244	14,8	2,90	1,13	5,14
	49	19,6	17,3	4,4	1,5	5,8	278	15,0	3,00	1,15	5,39
	50	17,7	14,3	3,8	1,9	6,0	240	15,6	2,37	0,92	8,21
Media		19,02	15,83	4,12	2,04	5,96	267,59	15,20	2,66	1,04	6,84
Deviația standard		1,17	0,99	0,34	0,34	0,46	29,47	0,97	0,17	0,07	1,07
Coeficient de variație		2,34	1,99	0,67	0,67	0,91	58,93	1,93	0,34	0,13	2,13
Eroarea standard		0,16	0,14	0,04	0,04	0,06	4,16	0,13	0,02	0,01	0,15
Limita superioară a intervalului de confidență		19,38	16,18	4,47	2,39	6,32	267,94	15,56	3,02	1,40	7,19
Limita inferioară a intervalului de confidență		18,66	15,46	3,75	1,68	5,60	267,22	14,84	2,30	0,68	6,47

Ip=indice profil; Ic K=indice de circumferință Kieslev; CF= coeficient Fulton; Lt=lungimea totală, Ls=lungime standard, Lc=lungimea capului, Ha=înălțimea la nivelul înotătoarei anale, Hd=înălțimea la nivelul înotătoarei dorsale, G=greutate, C=circumferință

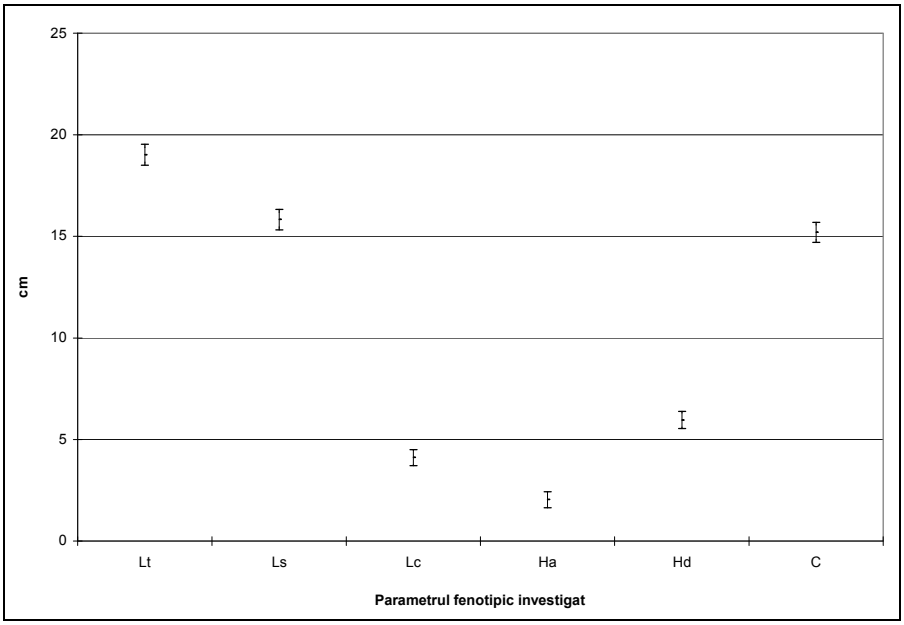


Figura 23 Intervalele de variabilitate ale parametrilor fenotipici analizați, la speciile genului *Carassius*

Dacă comparăm valorile medii ale celor doi indici şi al coeficientului Fulton, pentru ambele genuri investigate (Tabel 17), putem constata că valorile sunt aproximativ egale, ceea ce denotă că există similaritate fenotipică chiar şi între indivizii celor două genuri. Astfel, se constată că există o diferență foarte mică (0,41) în ceea ce priveşte indicele de profil pentru indivizii celor două genuri, de unde rezultă că morfologic, populațiile studiate sunt asemănătoare.

Tabel 17 Compararea indicilor calculați pentru cele două genuri

Nr. crt.	Indici fenotipici	*Cyprinus*	*Carassius*
1	Indice de profil	2,25	2,66
2	Indice de circumferință	1,22	1,04
3	Coeficient Fulton	3,18	6,84

4.2. TESTAREA OBȚINERII PRODUŞILOR PCR PRIN ELECTROFOREZĂ ORIZONTALĂ ÎN GEL DE AGAROZĂ

Amplificarea genică, s-a realizat pentru 10 indivizi din fiecare populație (Larga-Jijia – Movileni şi Iaşi), aparținând celor două genuri, inclusiv hibrizi şi indivizi ginogenetici. Trebuie menționat faptul că toate experimentele de amplificare genică prin PCR şi secvențierea, au fost efectuate în cadrul Laboratorului de Genetică şi Filogenie

Moleculară a Facultății de Ştiințe, Universitatea din Vigo, Spania şi completate cu rezultate obținute în cadrul Laboratorului de Genetică Moleculară şi Arheogenetică, Universitatea „Alexandru Ioan Cuza”

Pentru toate cele trei gene, reacțiile de amplificare s-au desfăşurat conform protocoalelor prezentate în capitolul anterior, pe baza unei matrițe de ADNmt provenit din țesutul muscular, deoarece s-a constatat că purificarea este mult mai bună în ciuda faptului că este prezent în cantitate mai mică la nivelul acestui țesut, fapt care, pe de altă parte, determină şi o mai bună dizolvare a acestuia în buffer TE.

În cazul speciilor genului *Cyprinus*, s-a constatat că pentru prima parte a citocromului b, s-au obținut în urma amplificării, fragmente de aproximativ 480 perechi de baze (pb) (Figura 24).

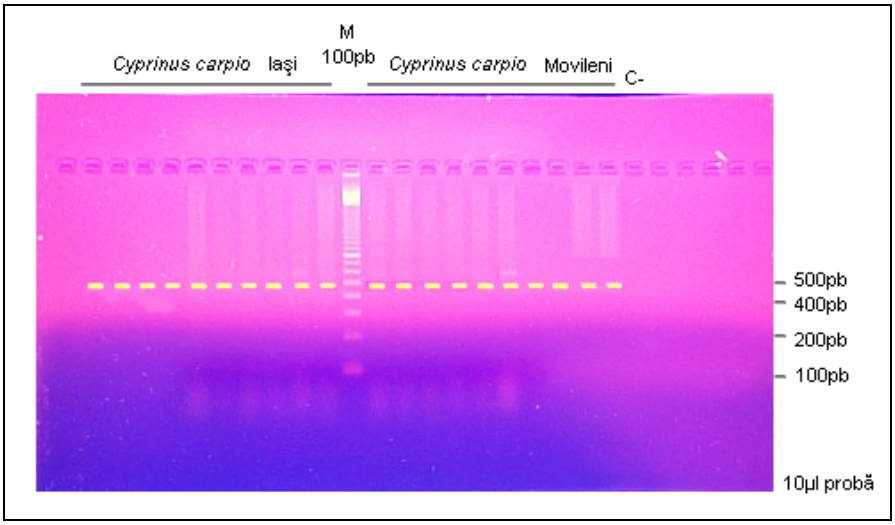


Figura 24 Electroforeză în gel de agaroză 1,5%, pentru prima parte a Cytb la indivizi ai genului *Cyprinus* (C-=control negativ, M=marker molecular de 100pb)

După cum s-a precizat în capitolul anterior, s-a folosit un marker molecular de 100pb pentru determinarea lungimilor fragmentelor amplificate. De asemenea, rolul controlului negativ (C-), este de determinare a eventualelor contaminări.

Pentru a doua parte a citocromului b (cyt b) în cazul speciilor aceluiaşi gen, s-au obținut fragmente cu lungimi de aproximativ 800pb (Figura 25).

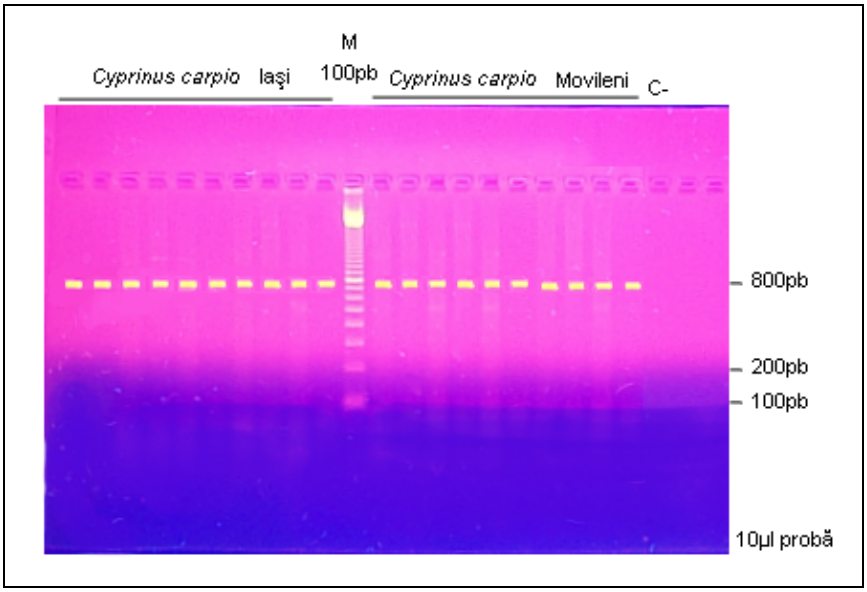


Figura 25 Electroforeză în gel de agaroză 1,5%, pentru a doua parte a Cytb la indivizi ai genului *Cyprinus* (C-=control negativ, M=marker molecular de 100pb)

În cazul speciilor genului *Carassius gibelio* Bloch, după cum a fost precizat şi în capitolul anterior, amplificarea citocromului b a avut loc într-o singură etapă, obținându-se un segment de aproximativ 1200pb (Figura 26).

Potrivit datelor furnizate de banca de gene (GenBank din cadrul NCBI – National Center for Biotechnology Information), lungimea totală a citocromului b la ambele genuri studiate, este de aproximativ 1140pb. Rezultatele obținute de noi în urma clonării acestei gene, după cum reiese din figurile anterioare, nu corespund acestor date, deoarece segmentele amplificate includ secvențele primerilor şi fragmente din secvențele genelor limitrofe, deoarece primerii utilizați, în majoritatea cazurilor se aliniază în aceste zone.

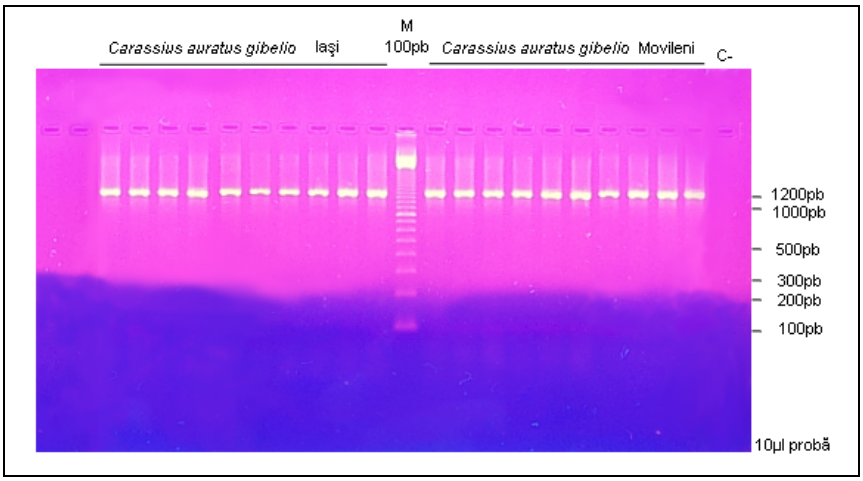


Figura 26 Electroforeză în gel de agaroză 1,5%, pentru Cytb la *Carassius gibelio* Bloch (C-=control negativ, M=marker molecular de 100pb)

În cazul regiunii de control mitocondrial (d-loop), pentru indivizii genului *Cyprinus,* s-au obținut în urma clonării, fragmente cu lungimi de aproximativ 420pb (Figura 27).

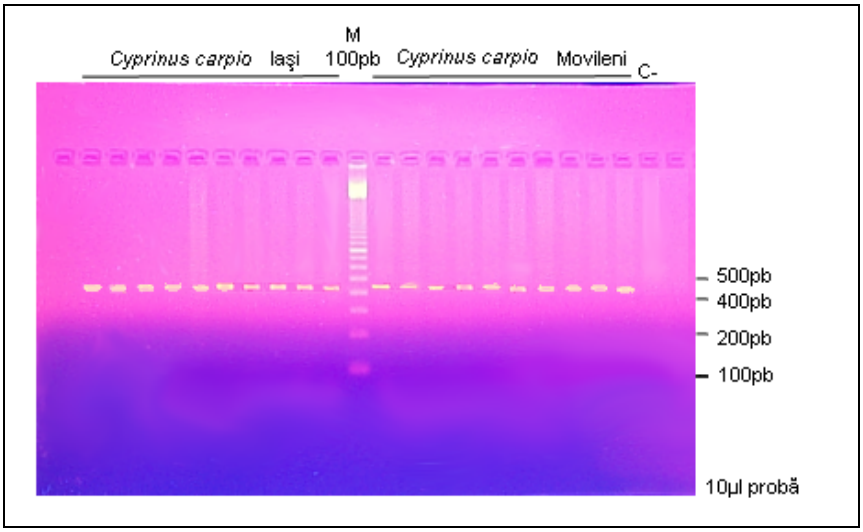


Figura 27 Electroforeză în gel de agaroză 1,5%, pentru d-loop la indivizi ai genului *Cyprinus* (C-=control negativ, M=marker molecular de 100pb)

La *Carassius gibelio*, pentru toți indivizii analizați, în urma amplificării d-loop, s-au obținut fragmente cu lungimi de aproximativ 300pb (Figura 28), datorită poziției primerilor utilizați, care determină amplificarea doar a primului segment hipervariabil al acestei gene.

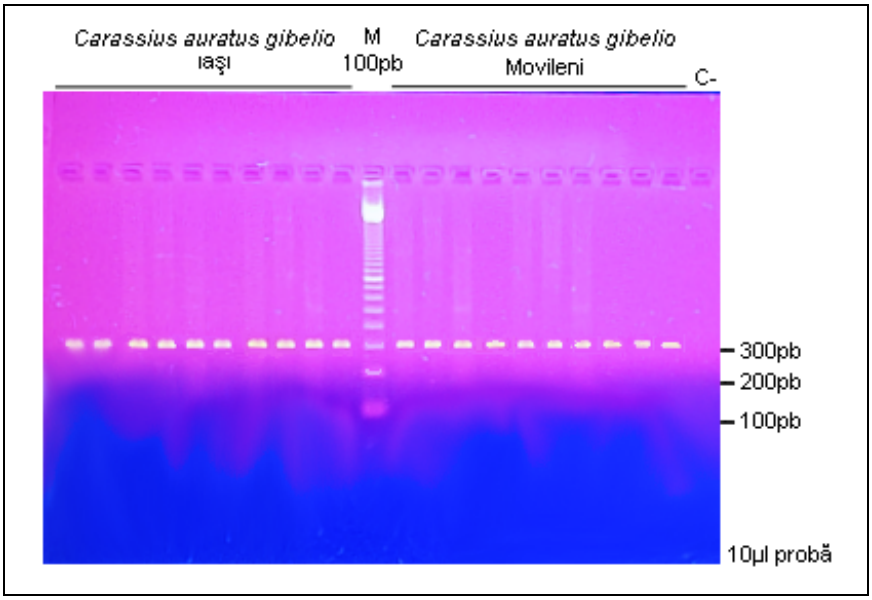


Figura 28 Electroforeză în gel de agaroză 1,5%, pentru d-loop la *Carassius gibelio* Bloch (C-=control negativ, M=marker molecular de 100pb)

Amplificarea subunității 4L a dehidrogenazei, a determinat obținerea unui segment de aproximativ 280pb pentru indivizii ambelor genuri *Cyprinus* (Figura 29) şi *Carassius* (Figura 30).

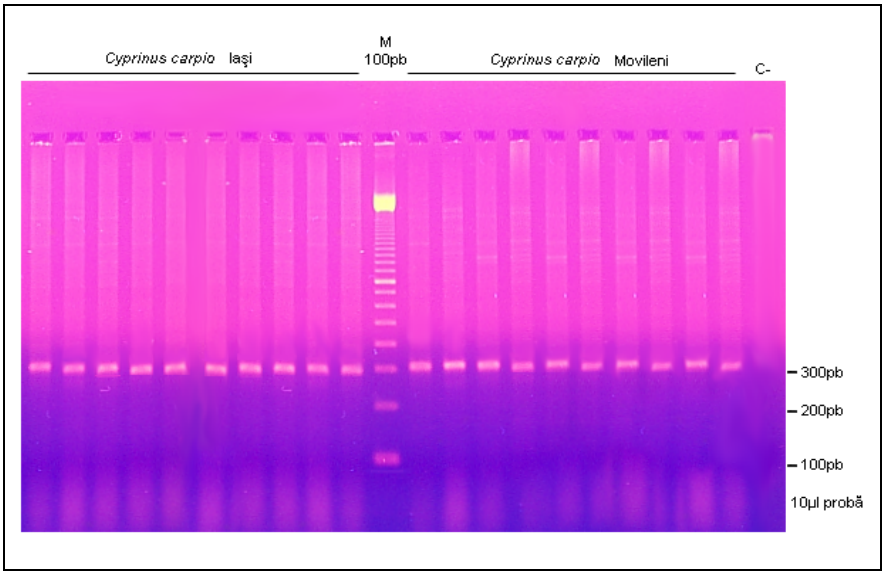


Figura 29 Electroforeză în gel de agaroză 1,5%, pentru ND4L la indivizi ai genului *Cyprinus* (C-=control negativ, M=marker molecular de 100pb)

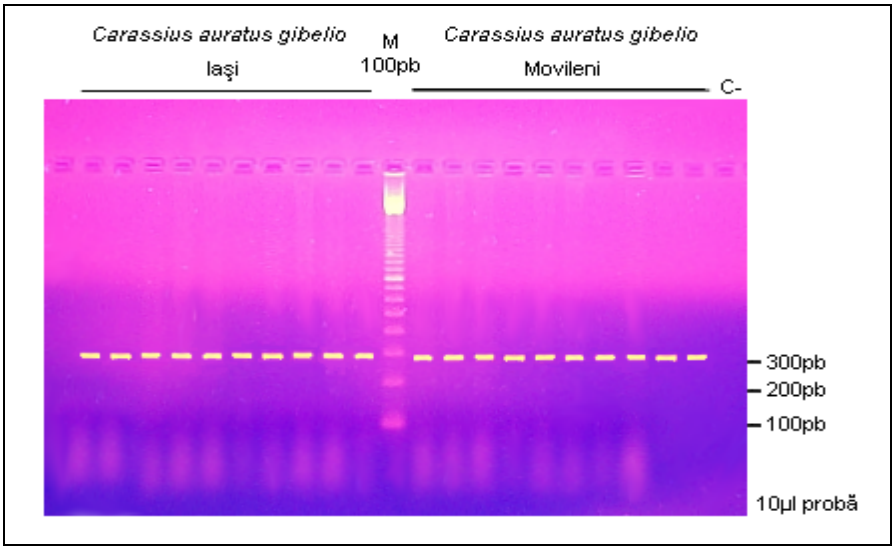


Figura 30 Electroforeză în gel de agaroză 1,5%, pentru ND4L la *Carassius gibelio* Bloch (C-=control negativ, M=marker molecular de 100pb)

4.3. CUANTIFICAREA PRODUŞILOR PCR

Cuantificarea produşilor PCR, reprezintă estimarea cantităților de produs purificat şi determinarea cantităților necesare reacției de secvențiere. După cum a fost precizat anterior, cuantificarea se realizează utilizând markerul molecular ØX174-*Hinf I* prin migrarea produşilor PCR purificați, într-un gel de agaroză 3%.

Pentru citocromul b prima parte, la speciile genului *Cyprinus*, pe baza electroforegramei (Figura 31), au fost estimate cantitățile de ADN existente în fiecare probă (Tabelul 18).

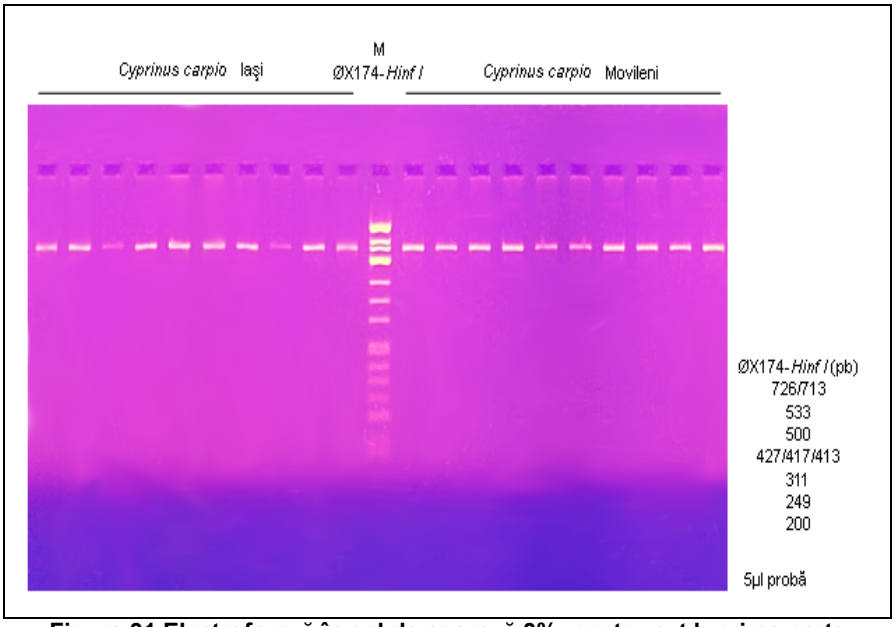


Figura 31 Electroforeză în gel de agaroză 3%, pentru cyt b prima parte la indivizi ai genului *Cyprinus*

Din Tabelul 18, putem constata că valorile concentrațiilor de ADN amplificat în PCR şi purificat, corespunzătoare primei părți a citocromului b, sunt suficient de ridicate, astfel încât să se obțină concentrații suficiente (între 60 – 90 fmoli) pentru reacția de secvențiere.

Tabel 18 Estimarea cantităților de ADN necesare secvențierii primei părți a cyt b, pentru indivizi ai genului *Cyprinus*

Nr. probă	Banda	Concentrația ADN (ng/ul)	Cantitatea de ADN (µl)	fmol
Cy01I	533	5,9	4	71,5
Cy02I	500	5,6	4,5	76,3
Cy03I	1/2x200	1,1	7,6	25
Cy04I	500	5,6	4	67,8
Cy05I	0,8x500	4,48	5	67,5
Cy06I	533	5,9	4	71,5
Cy07I	500	5,6	4	67,8
Cy08I	533	5,6	4	71,5
Cy09I	533	5,9	4	71,5
Cy10I	0,8x500	4,48	5	67,8
Cy01M	533	5,6	4	71,5
Cy02M	533	5,9	4	71,5
Cy03M	533	5,6	4	71,5
Cy04M	533	5,9	4	71,5
Cy05M	0,8x500	4,48	5	67,8
Cy06M	0,8x500	4,48	5	67,8
Cy07M	500	5,6	4	67,8
Cy08M	500	5,6	4	67,8
Cy09M	500	5,6	4	67,8
Cy10M	500	5,6	4	67,8

Pentru partea a doua a citocromului b la specii ale genului *Cyprinus*, a fost realizată cuantificarea pe baza Figurii 32, iar datele au fost sintetizate în Tabelul 19.

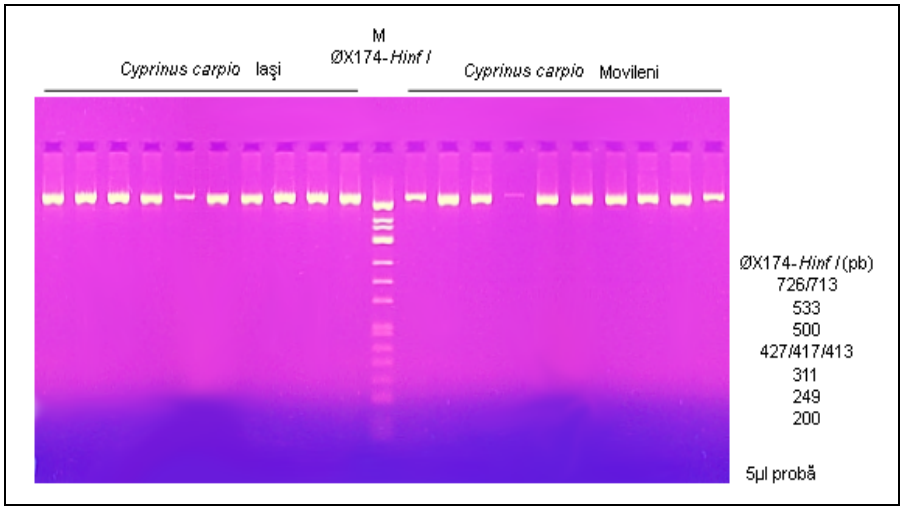


Figura 32 Electroforeză în gel de agaroză 3%, pentru cyt b partea a doua, la indivizi ai genului *Cyprinus*

Tabel 19 Estimarea cantităților de ADN necesare secvențierii cyt b a doua parte, pentru indivizi ai genului *Cyprinus*

Nr. probă	Banda	Concentrația ADN (ng/ul)	Cantitatea de ADN (µl)	fmol
Cy01I	533	5,94	6	67,5
Cy02I	1,8x726/713	28,8	1,5	81,8
Cy03I	1,8x726/713	28,8	1,5	81,8
Cy04I	1,5x726/713	24	1,5	68,18
Cy05I	1,8x726/713	28,8	1,5	81,8
Cy06I	1,8x726/713	28,8	1,5	81,8
Cy07I	2x726/713	32,0	1	60,6
Cy08I	1,8x726/713	28,8	1,5	81,8
Cy09I	1,8x726/713	28,8	1,5	81,8
Cy10I	1,3x726/713	28,8	1,5	81,8
Cy01M	427/417/413	14	3	79,54
Cy02M	1,8x726/713	28,8	1,5	81,8
Cy03M	1,8x726/713	28,8	1,5	81,8
Cy04M	0,5x200	1,1	7,5	15,6
Cy05M	1,8x726/713	28,8	1,5	81,8
Cy06M	1,8x726/713	28,8	1,5	81,8
Cy07M	1,8x726/713	28,8	1,5	81,8
Cy08M	1,8x726/713	28,8	1,5	81,8
Cy09M	1,8x726/713	28,8	1,5	81,8
Cy10M	427/417/413	14	3	79,54

La indivizii genului *Carassius* din ambele populații, s-a realizat cuantificarea întregului citocrom b (Figura 33 şi Tabelul 20).

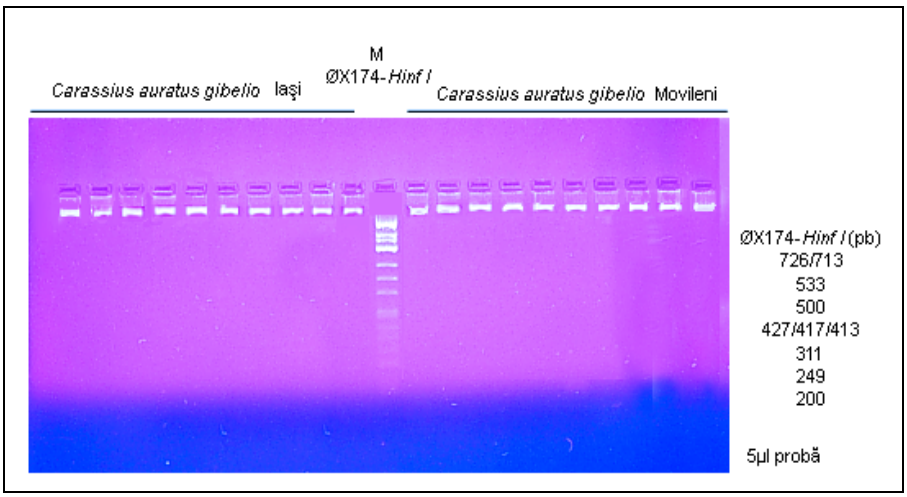


Figura 33 Electroforeză în gel de agaroză 3%, pentru cyt b, la indivizi ai genului *Carassius*

Se observă că, doar în cazul unui singur individ (Cy04M) nu s-a obținut cantitatea de ADN dorită pentru reacția de secvențiere, cu toate că este posibil ca aceasta să decurgă normal.

Tabel 20 Estimarea cantităților de ADN necesare secvențierii cyt b pentru indivizi ai genului *Carassius*

Nr. probă	Banda	Concentrația ADN (ng/ul)	Cantitatea de ADN (µl)	fmol
Ca01I	427/417/413	14	2	70,7
Ca02I	0,7x427/417/413	9,8	3	74,2
Ca03I	0,7x427/417/413	9,8	3	74,2
Ca04I	427/417/413	14	2	70,7
Ca05I	427/417/413	14	2	70,7
Ca06I	726/713	16	1,55	60,6
Ca07I	726/713	16	1,55	60,6
Ca08I	427/417/413	14	2	70,7
Ca09I	427/417/413	14	2	70,7
Ca10I	427/417/413	14	2	70,7
Ca01M	427/417/413	14	2	70,7
Ca02M	726/713	16	1,55	60,6
Ca03M	427/417/413	14	2	70,7
Ca04M	427/417/413	14	2	70,7
Ca05M	427/417/413	14	2	70,7
Ca06M	427/417/413	14	2	70,7
Ca07M	427/417/413	14	2	70,7
Ca08M	427/417/413	14	2	70,7
Ca09M	427/417/413	14	2	70,7
Ca10M	0,7x427/417/413	9,8	3	74,2

În acest caz, au fost obținute valori ale concentrației de ADN, cuprinse între 60,6fmoli şi 74,2fmoli ADN, ceea ce asigură un optim al cantității de produs PCR pentru reacția de secvențiere.

Pentru d-loop la speciile genului *Cyprinus,* s-au calculat cantitățile de ADN necesar reacției de secvențiere, pe baza Figurii 34.

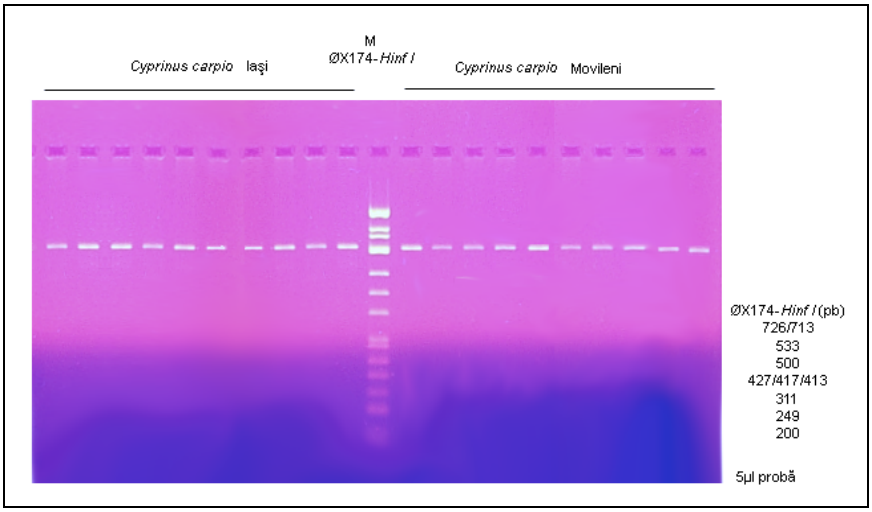


Figura 34 Electroforeză în gel de agaroză 3%, pentru d-loop, la indivizi ai genului *Cyprinus*

Comparativ, la indivizii genului *Carassius*, s-au obținut cantități mai mari de produs PCR purificat şi implicit cantități mai mari de ADN ce pot fi utilizate pentru reacția de secvențiere (Figura 35 şi Tabelul 21).

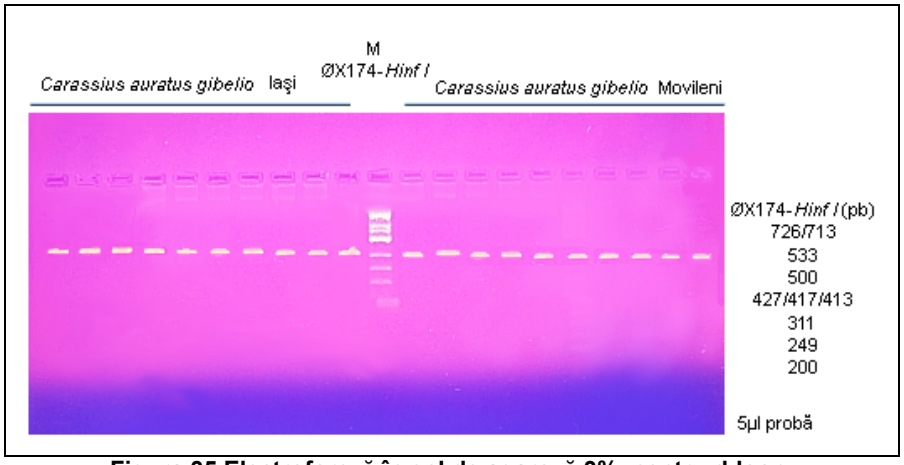


Figura 35 Electroforeză în gel de agaroză 3%, pentru d-loop, la indivizi ai genului *Carassius*

Tabel 21 Estimarea cantităților de ADN necesare secvențierii d-loop pentru indivizi ai genului *Cyprinus*

Nr. probă	Banda	Concentrația ADN (ng/ul)	Cantitatea de ADN (µl)	fmol
Cy01I	200	2,2	7,5	50
Cy02I	311	3,4	7,5	77,2
Cy03I	249	2,7	7,5	61,3
Cy04I	500	5,6	4,5	76,3
Cy05I	200	2,2	7,5	50
Cy06I	200	2,2	7,5	50
Cy07I	200	2,2	7,5	50
Cy08I	249	2,7	7,5	61,3
Cy09I	311	3,4	7,5	77,2
Cy10I	200	2,2	7,5	50
Cy01M	200	2,2	7,5	50
Cy02M	200	2,2	7,5	50
Cy03M	200	2,2	7,5	50
Cy04M	249	2,7	7,5	61,3
Cy05M	311	3,4	7,5	77,2
Cy06M	200	2,2	7,5	50
Cy07M	200	2,2	7,5	50
Cy08M	249	2,7	7,5	61,3
Cy09M	200	2,2	7,5	50
Cy10M	200	2,2	7,5	50

Din Tabelul 21, se observă că s-au obținut cantități de ADN cu valori cuprinse între 50fmoli pentru majoritatea indivizilor şi 77,2fmoli pentru trei dintre indivizii analizați.

Tabel 22 Estimarea cantităților de ADN necesare secvențierii d-loop pentru indivizi ai genului *Carassius*

Nr. probă	Banda	Concentrația ADN (ng/ul)	Cantitatea de ADN (µl)	fmol
Ca01I	533	5,94	4	72
Ca02I	0,7x427/417/413	9,8	2,5	74,24
Ca03I	0,8x427/417/413	11,2	2	67,87
Ca04I	0,8x427/417/413	11,2	2	67,87
Ca05I	533	5,94	4	72
Ca06I	533	5,94	4	72
Ca07I	0,8x427/417/413	11,2	2	67,87
Ca08I	0,5x427/417/413	7	3,5	74,24
Ca09I	500	5,6	4,5	76,36
Ca10I	0,7x427/417/413	9,8	2,5	74,24
Ca01M	533	5,94	4	72
Ca02M	500	5,6	4,5	76,36
Ca03M	0,8x427/417/413	11,2	2	67,87
Ca04M	0,8x427/417/413	11,2	2	67,87
Ca05M	533	5,94	4	72
Ca06M	500	5,6	4,5	76,36
Ca07M	500	5,6	4,5	76,36
Ca08M	0,8x427/417/413	11,2	2	67,87
Ca09M	0,8x427/417/413	11,2	2	67,87
Ca10M	0,8x427/417/413	11,2	2	67,87

Se poate observa din Tabelul 22 că s-au obținut cantitățile optime de ADN necesare reacției de secvențiere, pentru toți indivizii analizați, aparținând celor două populații Iaşi şi Larga-Jijia – Movileni.

Pentru subunitatea 4L a dedidrogenazei (ND4L) în cazul speciilor şi varietăților analizate pentru genul *Cyprinus*, s-au calculat cantitățile necesare de ADN pe baza Figurii 36.

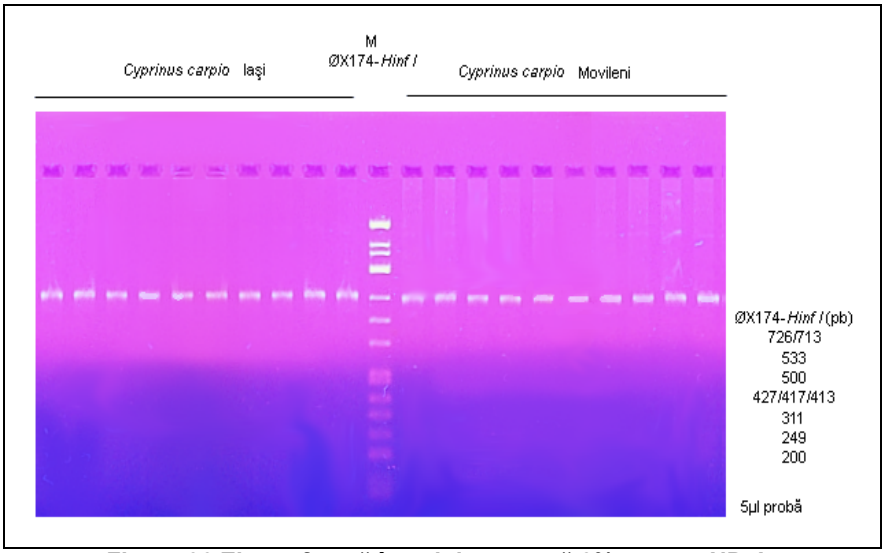


Figura 36 Electroforeză în gel de agaroză 3%, pentru ND4L, la indivizi ai genului *Cyprinus*

La indivizii genului *Carassius* din ambele populații analizate, s-au calculat cantitățile de ADN (produs PCR purificat) necesare reacției de secvențiere, pe baza Figurii 37.

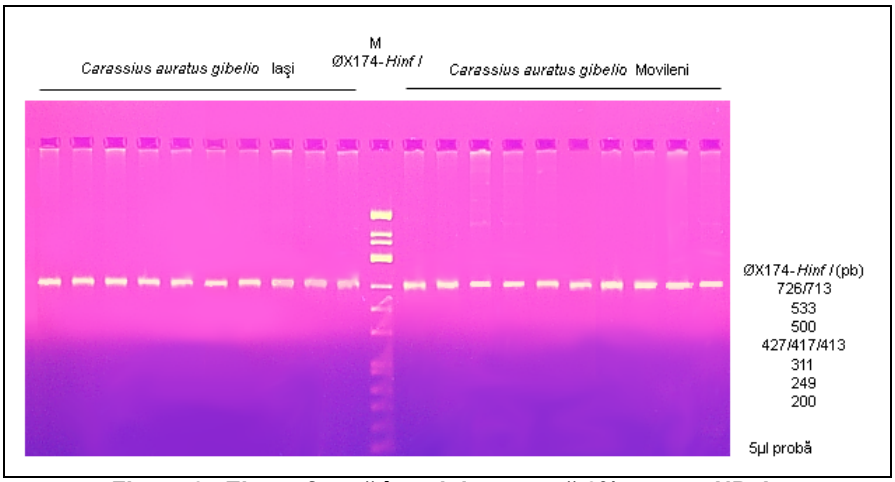


Figura 37 Electroforeză în gel de agaroză 3%, pentru ND4L, la indivizi ai genului *Carassius*

Tabel 23 Estimarea cantităților de ADN necesare secvențierii ND4L pentru indivizi ai genului *Cyprinus*

Nr. probă	Banda	Concentrația ADN (ng/ul)	Cantitatea de ADN (µl)	fmol
Cy01I	249	2,78	4,5	62,5
Cy02I	249	2,78	4,5	62,5
Cy03I	311	3,4	3,5	59,5
Cy04I	311	3,4	3,5	59,5
Cy05I	0,4x427/417/413	5,6	2,5	70
Cy06I	0,5x427/417/413	7	2	70
Cy07I	0,5x427/417/413	7	2	70
Cy08I	0,5x427/417/413	7	2	70
Cy09I	311	3,4	3,5	59,5
Cy10I	311	3,4	3,5	59,5
Cy01M	0,5x427/417/413	7	2	70
Cy02M	0,5x427/417/413	7	2	70
Cy03M	0,5x427/417/413	7	2	70
Cy04M	0,5x427/417/413	7	2	70
Cy05M	0,5x427/417/413	7	2	70
Cy06M	311	3,4	3,5	59,5
Cy07M	0,5x427/417/413	7	2	70
Cy08M	0,5x427/417/413	7	2	70
Cy09M	0,5x427/417/413	7	2	70
Cy10M	0,5x427/417/413	7	2	70

Din Tabelul 23 rezultă că s-au obținut cantități de ADN necesare secvențierii, cuprinse între 59,5 şi 70fmoli.

Se poate constata din Tabelul 24, că şi în acest caz, pentru indivizii genului *Carassius*, cantitățile de ADN obținute pentru reacția de secvențiere a ND4L sunt mai mari decât în cazul varietăților genului *Cyprinus*.

Tabel 24 Estimarea cantităților de ADN necesare secvențierii d-loop pentru indivizi ai genului *Carassius*

Nr. probă	Banda	Concentrația ADN (ng/ul)	Cantitatea de ADN (µl)	fmol
Ca01I	0,7x427/417/413	9,8	1,5	73,5
Ca02I	0,8x427/417/413	11,2	1,5	84
Ca03I	0,8x427/417/413	11,2	1,5	84
Ca04I	0,7x427/417/413	9,8	1,5	73,5
Ca05I	0,7x427/417/413	9,8	1,5	73,5
Ca06I	0,8x427/417/413	11,2	1,5	84
Ca07I	0,7x427/417/413	9,8	1,5	73,5
Ca08I	0,7x249	1,94	7,5	72,75
Ca09I	0,7x249	1,94	7,5	72,75
Ca10I	0,5x249	1,39	7,5	52,12
Ca01M	0,7x427/417/413	9,8	1,5	73,5
Ca02M	0,7x427/417/413	9,8	1,5	73,5
Ca03M	0,7x427/417/413	9,8	1,5	73,5
Ca04M	0,7x427/417/413	9,8	1,5	73,5
Ca05M	0,7x427/417/413	9,8	1,5	73,5
Ca06M	0,7x427/417/413	9,8	1,5	73,5
Ca07M	0,8x427/417/413	11,2	1,5	84
Ca08M	0,8x427/417/413	11,2	1,5	84
Ca09M	0,8x427/417/413	11,2	1,5	84
Ca10M	0,7x427/417/413	9,8	1,5	73,5

4.4. SECVENȚIEREA GENELOR UTILIZATE ÎN STUDIILE DE FILOGENIE

Secvențierea propriu-zisă a celor 3 gene luate în studiu pentru toți indivizii aparținând la două genuri, fiecare cu câte două populații diferite, s-a realizat cu ajutorul unui secvențiator Beckman Coulter cu 8 capilare. Pentru fiecare din cele trei gene, au fost secvențiate ambele catene (directă şi reversă), secvențele fiind obținute sub formă de fluorograme.

Toate fluorogramele au fost corectate utilizând programul CEQ 2000 DNA Analysis System, furnizat de Beckman Coulter. Această corectare este necesară pentru a verifica dacă citirea electronică a fost corectă şi a completa sau înlocui eventualele erori date de suprapunerea curbelor fluorogramei. Ulterior, secvențele corectate au fost exportate pentru aliniere în programul ESEE 32 (The Eyeball Sequence Editor) care a permis alinierea catenelor complementare pe baza secvențelor primerilor, tăierea şi unirea ambelor catene într-o unică secvență care are lungimea reală.

4.5. COMPARAREA SECVENȚELOR NUCLEOTIDICE ALE CITOCROMULUI B, PROVENITE DE LA INDIVIZI APARȚINÂND GENULUI *CYPRINUS*

În cazul speciilor genului *Cyprinus*, pentru citocromul b, au fost aliniate cele 20 de secvențe (19 indivizi aparținând celor două populații şi 1 individ de crap oglindă Cy06M), utilizând metoda Clustal V (Higgins şi Sharp, 1989).

Din alinierea secvențelor pentru 20 de indivizi ai genului *Cyprinus*, se constată că există un număr de 15 diferențe între cele douăzeci de secvențe.

Cele mai multe diferențe, apar la individul de crap oglindă (Cy06M). Astfel, pentru acesta au fost înregistrate un număr de 12 diferențe, dintre care, 11 tranziții şi 1 transversie. Tranzițiile au fost observate în pozițiile: 177 (substituția citozinei cu timină); 198 şi 864 (substituția timinei cu citozină); 352, 568, 624, 693,897 şi 990 (substituția adeninei de guanină); 591 şi 732 (substituția guaninei de către adenină). Transversia s-a produs în poziția 870, unde a avut loc înlocuirea timinei de către adenină. În cazul acestei secvențe, se poate constata că rata de apariție a tranzițiilor este foarte mare comparativ cu cea de apariție a transversiilor.

Alte diferențe, au fost constatate pentru individul Cy03M (individ normal din populația Movileni) în cazul căruia s-a produs o transversie în poziția 396, unde a fost substituită timina de către guanină; la indivizii Cy01M şi Cy04M s-a înregistrat o tranziție în poziția 507 (înlocuirea adeninei de către guanină) şi în final, pentru individul Cy04I (din populația Iaşi), a avut loc o transversie în poziția 432, unde a fost substituită adenina cu timină.

Referitor la restul indivizilor, se constată o rată destul de scăzută a diferențelor existente între cele două populații.

Din cele prezentate anterior, putem concluziona că au fost găsite 5 noi haplotipuri pentru citocromul b: Cy06M pentru crapul oglindă din populația Mărgineni, Cy0104M şi Cy03M pentru crapul normal cu solzi, din aceeaşi populație şi Cy04I pentru crapul cu solzi din populația Iaşi, fiind stabilit haplotipul general pentru cele două populații (dat de celelalte secvențe care nu prezintă diferențe).

Pe baza comparației între secvențe, s-a realizat tabelul similarității şi divergenței (Tabel 25), din care se constată că procentele de similaritate sunt cuprinse între 98,9% (pentru secvențele Cy06M şi restul care nu prezintă foarte multe diferențe) şi 99,9% pentru secvențele care diferă între ele doar printr-o singură nucleotidă.

Tabel 25 Procentele de similaritate şi divergenţă pentru secvenţele haplotipurilor stabilite

	Procent de similaritate						
Procent de divergenţă		Cy01I	Cy04I	Cy0104M	Cy03M	Cy06M	
	Cy01I		99,9	99,9	99,9	98,9	Cy01I
	Cy04I	0,1		99,8	99,8	98,9	Cy04I
	Cy03M	0,1	0,2		99,8	98,9	Cy03M
	Cy0104M	0,1	0,2	0,2		98,9	Cy0104M
	Cy06M	1,1	1,2	1,2	1,2		Cy06M
		Cy01I	Cy04I	Cy0104M	Cy03M	Cy06M	

În plus, pentru a stabili o mai clară diferenţiere între toate noile haplotipuri, au fost analizate numărul şi procentele în care se găsesc bazele azotate în fiecare dintre acestea (Tabel 26).

Tabel 26 Procentul bazelor azotate, în cadrul haplotipurilor stabilite

Baze azotate		Specii analizate				
		Cy01I	Cy04I	Cy0104M	Cy03M	Cy06M
A	Nr.	350	349	349	350	347
	%	30,70	30,60	30,60	30,70	30,44
G	Nr.	155	155	156	156	159
	%	13,60	13,60	13,68	13,68	13,95
T	Nr.	295	297	295	294	293
	%	25,88	26,0	25,88	25,79	25,70
C	Nr.	340	340	340	340	341
	%	29,82	20,80	29,82	29,82	29,91
A+T	Nr.	645	645	644	644	640
	%	56,58	56,58	56,49	56,49	56,14
C+G	Nr.	495	495	496	496	500
	%	43,42	43,42	43,51	43,51	43,86

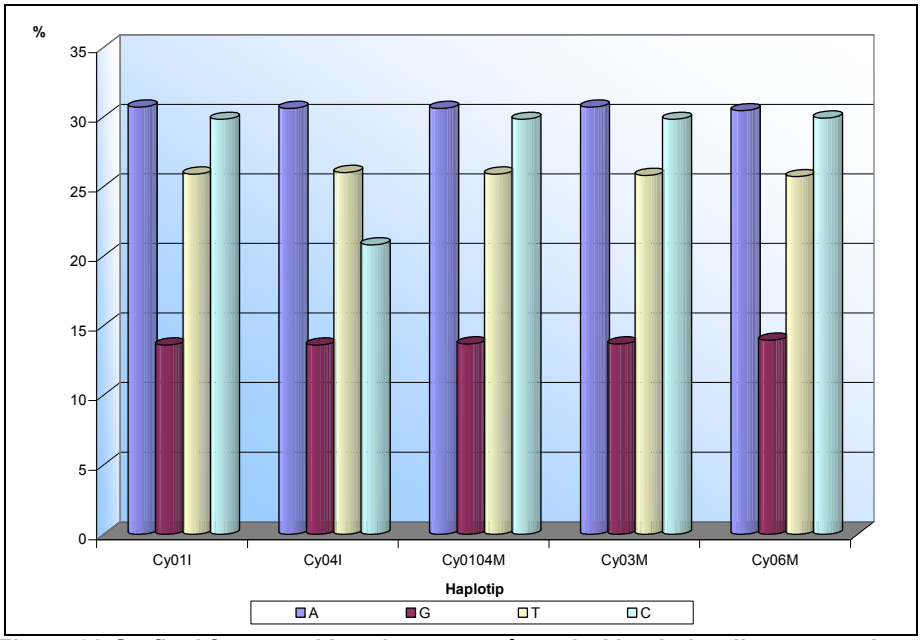


Figura 38 Graficul frecvenței bazelor azotate în cadrul haplotipurilor caracterizate

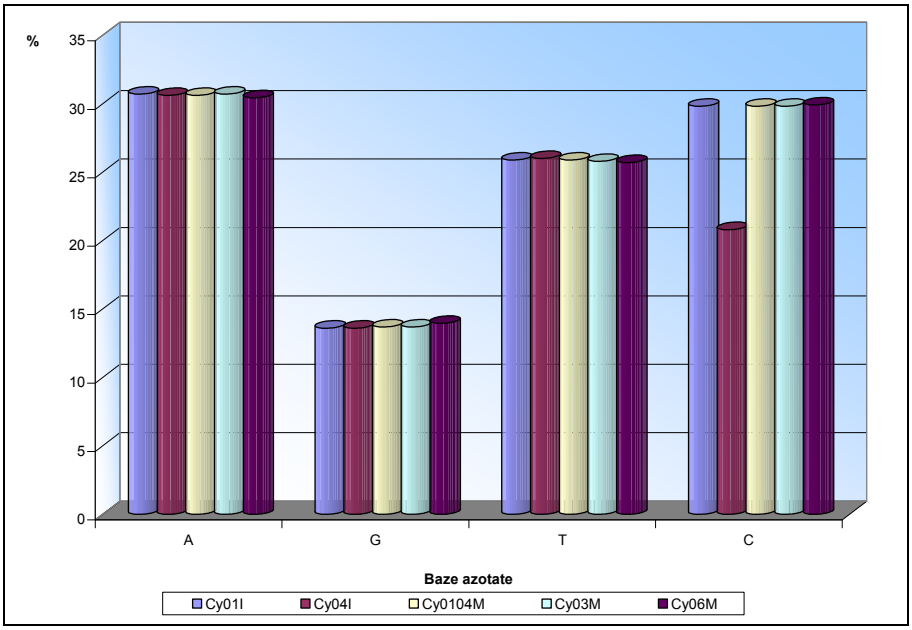


Figura 39 Graficul frecvenței bazelor azotate, comparativ între haplotipurile analizate

Din Tabelul 26 şi Figurile 38 şi 39, se poate observa că diferențe majore se constată pentru secvența Cy06M, care prezintă valori diferite pentru toate bazele azotate, comparativ cu celelalte haplotipuri.

Referitor la cantitatea de adenină (A), aceasta este egală pentru Cy01I şi Cy03M (30,70%) şi pentru haplotipurile Cy04I – Cy0104M (30,60%). Cantitatea de guanină este aproximativ egală pentru toate haplotipurile, fiind cuprinsă între 13,60% (Cy01I şi Cy04I) şi 13,68% pentru Cy0104M şi Cy03M. De asemeni, se constată diferenţe şi pentru celelalte baze azotate: timina se găseşte în proporţie de 25,88% pentru Cy01I, Cy0104M, având valori diferite pentru celelalte haplotipuri, iar citozina, în proporţii aproximativ egale pentru toate haplotipurile.

Procentul de baze complementare este de 56,58% în cazul A+T pentru primele două haplotipuri şi 56,49% pentru Cy0104M şi Cy03M, pentru Cy06M fiind de 56,14%. Procentul de G+C este de 43,42% pentru primele două haplotipuri (Cy01I, Cy04I), de 43,51% pentru (Cy0104M, Cy03M) şi 43,86% pentru Cy06M.

4.5.1. CARACTERIZAREA HAPLOTIPULUI CY01I

4.5.1.1. FRECVENŢELE TIPURILOR DE ASOCIERI NUCLEOTIDICE

În această categorie au fost încadrate toate secvenţele între care nu există diferenţe (Cy01I, Cy02I, Cy03I, Cy05I, Cy06I, Cy07I, Cy08I, Cy09I, Cy10I, Cy02M, Cy05M, Cy07M, Cy08M, Cy09M, Cy10M), aparţinând celor două populaţii luate în studiu.

Referitor la procentele dinucleotidelor, constatăm, că valoarea cea mai mare o are perechea AC (10,3%), iar valoarea cea mai mică perechea GG, în proporţie de 2,3%. Referitor la trinucleotide, cea mai mare frecvenţă o înregistrează tripletul CAT (3,9%), iar valoarea cea mai mică, tripletele GGG şi GCG, ambele având o frecvenţă de 0,2%.

Tetramerii cu frecvenţa cea mai mare, sunt ACTA (1,6%), iar cei cu frecvenţă redusă: TCTC, CTCA, ATAC, ACCC, toţi în procent de 1,0%. În cazul pentamerilor, cea mai mare frecvenţă o are pentamerul TCCTA (0,9%), iar cea mai mică frecvenţă, o au TTCCT, TACCA, CTCCT, CTCAT şi CTAGT (0,4%). În ceea ce priveşte hexamerii, aceştia au frecvenţe cuprinse între 0,3% şi 0,4%.

Tetramerii, pentamerii şi hexamerii cu frecvenţă joasă, sunt cei care se găsesc într-un singur exemplar în întreaga genă, având o frecvenţă de 0,1%.

Cu ajutorul modulului EditSeq al DNA Star, a fost calculată temperatura de topire a ADN, după un model Davis-Botstein-Roth; în acest caz, prezintă o valoare de 82,26ºC.

4.5.1.2. TRANSLAŢIA ŞI CARACTERIZAREA CATENEI POLIPEPTIDICE

Translaţia s-a realizat pe baza unui cod genetic mitocondrial pentru vertebrate, utilizând modulul EditSeq din cadrul programului DNA Star.

Catena polipeptidică obţinută este prezentată în Figura 40.

```
MASLRKTHPL IKIANDALVD LPTPSNISAW WNFGSLLGLC LITQILTGLF
LAMHYTSDIS TAFSSVTHIC RDVNYGWLIR NVHANGASFF FICIYMHIAR
GLYYGSYLYK ETWNIGVILL LLVMMTAFVG YVLPWGQMSF WGATVITNLL
SAVPYMGDML VQWIWGGFSV DNATLTRFFA FHFLLPFVIT AATIIHLLFL
HETGSNNPIG LNSDADKVSF HPYFSYKDLL GFVIMLLALT LLALFSPNLL
GDPENFTPAN PLVTPPHIKP EWYFLFAYAI LRSIPNKLGG VLALLFSILV
LMVVPLLHTS KQRGLTFRPI TQFLFWTLVA DMIILTWIGG MPVEHLIMWH
          ASSYSPHSHY A.RC.MAMTA MLQNKALKWA
```

Figura 40 Catena polipeptidică a citocromului b caracteristică haplotipului Cy01I al speciei *Cyprinus carpio*

După realizarea translației, se constată că întreg lanțul polipeptidic, este alcătuit din 378 aminoacizi (Tabelul 27), din care se observă că frecvența cea mai mare, o are leucina (15,08%), iar cea mai scăzută o are cisteina (1,06%). Se constată de asemenea, existența a doi codoni stop (notați cu „.”) către sfârşitul lanțului polipeptidic.

Se constată că din totalul de 378 aminoacizi, 16 sunt acizi, 19 sunt bazici, 92 polari şi 177 hidrofobi (Tabelul 28). De asemenea, se constată că lanțul polipeptidic are o sarcină electrică de 5,28 (la pH=7) şi un punct izoelectric de 8,42 (Figura 41).

Tabel 27 Compoziția lanțului polipeptidic

Aminoacizi		Număr	Procent din greutatea totală (%)	Frecvența (%)
Simbol	Nume			
A	Ala	30	5,01	7,94
C	Cys	4	0,97	1,06
D	Asp	11	2,97	2,91
E	Glu	5	1,52	1,32
F	Phe	26	8,98	6,88
G	Gly	23	3,08	6,08
H	His	15	4,83	3,97
I	Ile	29	7,70	7,67
K	Lys	10	3,01	2,65
L	Leu	57	15,14	15,08
M	Met	16	4,93	4,23
N	Asn	17	4,55	4,50
P	Pro	20	4,56	5,29
Q	Gln	6	1,80	1,59
R	Arg	9	3,30	2,38
S	Ser	25	5,11	6,61
T	Thr	25	5,93	6,61
V	Val	22	5,12	5,82
W	Trp	13	5,68	3,44
Y	Tyr	15	5,75	3,97
B	Asx	0	0,00	0,00
Z	Glx	0	0,00	0,00
.	Ter	2	0,00	0,53

Tabel 28 Caracterizarea lanțului polipeptidic

Greutate molecular ă (Daltoni)	Număr de aminoaciz i	Aminoacizi				Punct izoelectri c	Sarcina electric ă la pH=7	Concentrați a la o extincție =1 și λ=280nm
		Puterni c bazici	Puterni c acizi	Hidrofob i	Polar i			
42592,56	378	19	16	177	92	8,423	5,279	0,45

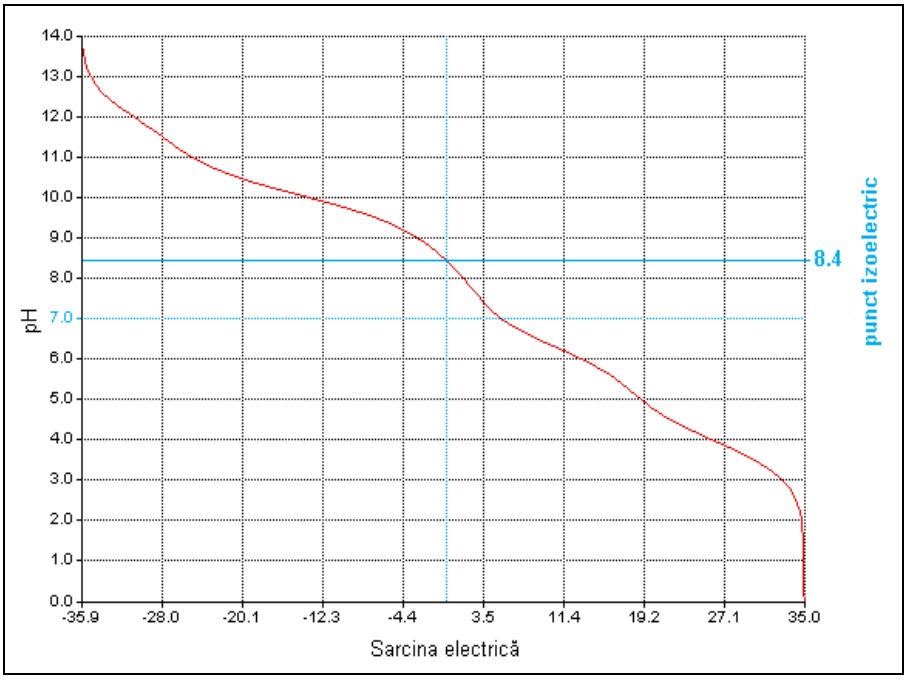


Figura 41 Curba de titrare pentru lanțul polipeptidic al citocromului b, haplotip Cy01I al speciei *Cyprinus carpio*

Ulterior, a fost realizată o analiză a structurii proteinei, fiind determinate zonele alfa helicoidale, beta pliate şi de inflexiune, prin două metode Garnier-Robson (Garnier et al., 1978) şi Chou-Fasman (Chou şi Fasman, 1978); reprezentarea grafică a zonelor hidrofobe (Kyte şi Doolittle, 1982), zonele de flexie (Karplus şi Schultz, 1985), indexul antigenic (Jameson şi Wolf, 1988) şi graficul probabil al suprafeței (Emini et al., 1985).

Din analiza structurii lanțului polipeptidic, se poate observa că referitor la regiunile alfa helicoidale, există diferențe între cele două metode Garnier-Robson şi Chou-Fasman. Astfel, în cazul primei metode, se observă că există 18 regiuni alfa helicoidale, în timp ce, potrivit celei de-a doua metode, există doar 5, existând doar 3 puncte de suprapunere ale suprafețelor. De asemenea, pentru regiunile beta pliate, se observă că există o similaritate crescută între cele două metode comparativ cu regiunile alfa. Pentru determinarea efectuată prin prima metodă, se constată existența a 30 de regiuni beta pliate, în timp ce, pentru cea de-a doua metodă (Chou-Fasman), sunt considerate ca existând 12 astfel de regiuni. Trebuie remarcat faptul că similaritatea în acest caz este dată de suprafețele ample considerate ca fiind beta pliate de cea de-a doua metodă. Referitor la zonele de inflexiune, de asemenea,

comparativ pentru cele două metode, se constată că pentru ambele metode (Garnier –Robson şi Chou-Fasman) sunt date 22 astfel de regiuni, cu diferența că în cazul celei de-a doua, sunt prezentate ca regiuni cu suprafețe mai mari.

Pe baza graficului hidrofobicității realizat după un model Kyte-Doolitle (Kyte şi Doolittle, 1982), se constată că există 15 zone hidrofobe şi 13 zone hidrofile.

Din graficul indexului antigenic calculat pe baza modelului Jameson –Wolf (Jameson şi Wolf, 1988), se constată existența a 23 de regiuni cu potențial antigenic.

4.5.2. CARACTERIZAREA HAPLOTIPULUI CY04I

4.5.2.1. FRECVENȚELE TIPURILOR DE ASOCIERI NUCLEOTIDICE

În acest haplotip intră o singură secvență care diferă de haplotipul general prin existența unei transversii (substituția adeninei de către timină) în poziția 432.

Referitor la procentele dinucleotidelor constatăm, că valoarea cea mai mare o are perechea AC (10,3%), iar valoarea cea mai mică perechea GG, în proporție de 2,3%. În cazul trinucleotidelor, cea mai mare frecvență o înregistrează tripletul CAT (3,9%), iar valoarea cea mai mică, tripletele GGG şi GCG, ambele având o frecvență de 0,2%.

Tetramerii cu frecvența cea mai mare, sunt ACTA (1,6%), iar cei cu frecvența redusă: TCTC, CTCA, ATAC, ACCC, toți în procent de 1,0%. În cazul pentamerilor (Tabelul 38), cea mai mare frecvență o are pentamerul TCCTA (0,9%), iar cea mai mică frecvență, o au TTCCT, TACCA, CTCCT, CTCAT şi CTAGT (0,4%). În ceea ce priveşte hexamerii, aceştia au frecvențe cuprinse între 0,3% şi 0,4%.

Tetramerii, pentamerii şi hexamerii cu frecvență joasă, sunt cei care se găsesc într-un singur exemplar în întreaga genă, având o frecvență de 0,1%.

Cu ajutorul modulului EditSeq al DNA Star, a fost calculată temperatura de topire a secvenței de ADN, după un model Davis-Botstein-Roth; în acest caz, prezintă o valoare de 82,26ºC.

4.5.2.2. TRANSLAȚIA ŞI CARACTERIZAREA CATENEI POLIPEPTIDICE

Translația s-a realizat pe baza codului genetic mitocondrial pentru vertebrate, utilizând modulul EditSeq din cadrul programului DNA Star (Figura 42).

```
MASLRKTHPL IKIANDALVD LPTPSNISAW WNFGSLLGLC LITQILTGLF
LAMHYTSDIS TAFSSVTHIC RDVNYGWLIR NVHANGASFF FICIYMHIAR
GLYYGSYLYK ETWNIGVILL LLVMMTAFVG YVLPWGQMSF WGATVITNLL
SAVPYMGDML VQWIWGGFSV DNATLTRFFA FHFLLPFVIT AATIIHLLFL
HETGSNNPIG LNSDADKVSF HPYFSYKDLL GFVIMLLALT LLALFSPNLL
GDPENFTPAN PLVTPPHIKP EWYFLFAYAI LRSIPNKLGG VLALLFSILV
LMVVPLLHTS KQRGLTFRPI TQFLFWTLVA DMIILTWIGG MPVEHLIMWH
           ASSYSPHSHY A.RC.MAMTA MLQNKALKWA
```

Figura 42 Catena polipeptidică a citocromului b caracteristică haplotipului Cy04I al speciei *Cyprinus carpio*

După realizarea translației, se constată că întreg lanțul polipeptidic, este alcătuit din 378 aminoacizi (Tabelul 29), din care, frecvența cea mai mare, o are leucina (15,08%), iar cea mai scăzută o are cisteina (1,06%). Se constată de asemenea,

existența a doi codoni stop (notați cu „.”) către sfârşitul lanțului polipeptidic ceea ce corespunde cu structura lanțului polipeptidic al haplotipului anterior, de unde putem concluziona că in cazul de față, în poziția 432, s-a produs o mutație nonsens, deci, o mutație care nu determină o modificare a aminoacidului.

Tabel 29 Compoziția lanțului polipeptidic pentru haplotipul Cy04I

Aminoacizi		Număr	Procent din greutatea totală (%)	Frecvența (%)
Simbol	Nume			
A	Ala	30	5,01	7,94
C	Cys	4	0,97	1,06
D	Asp	11	2,97	2,91
E	Glu	5	1,52	1,32
F	Phe	26	8,98	6,88
G	Gly	23	3,08	6,08
H	His	15	4,83	3,97
I	Ile	29	7,70	7,67
K	Lys	10	3,01	2,65
L	Leu	57	15,14	15,08
M	Met	16	4,93	4,23
N	Asn	17	4,55	4,50
P	Pro	20	4,56	5,29
Q	Gln	6	1,80	1,59
R	Arg	9	3,30	2,38
S	Ser	25	5,11	6,61
T	Thr	25	5,93	6,61
V	Val	22	5,12	5,82
W	Trp	13	5,68	3,44
Y	Tyr	15	5,75	3,97
B	Asx	0	0,00	0,00
Z	Glx	0	0,00	0,00
.	Ter	2	0,00	0,53

4.5.3. CARACTERIZAREA HAPLOTIPULUI CY0104M

4.5.3.1. FRECVENȚELE TIPURILOR DE ASOCIERI NUCLEOTIDICE

În acest haplotip, se încadrează 2 secvențe (Cy01M şi Cy04M), care diferă de haplotipul general prin existența unor tranziții (substituția adeninei de către guanină) în poziția 507.

Referitor la procentele dinucleotidelor constatăm, că valoarea cea mai mare o are perechea AC (10,3%), iar valoarea cea mai mică perechea GG, în proporție de 2,4%. În cazul trinucleotidelor, cea mai mare frecvență o înregistrează tripletul CAT (3,9%), iar valoarea cea mai mică, tripletele GGG şi GCG, ambele având o frecvență de 0,2%.

Tetramerul cu frecvența cea mai mare, este ACTA (1,6%), iar cel cu frecvența redusă: TTCC în procent de 0,9%. În cazul pentamerilor, cea mai mare frecvență o are pentamerul TCCTA (0,9%), iar cea mai mică frecvență, o au TTCCT, TACCA, CTCCT, CTCAT şi CTAGT (0,4%). În ceea ce priveşte hexamerii, aceştia au frecvențe cuprinse între 0,3% şi 0,4%.

Tetramerii, pentamerii şi hexamerii cu frecvență joasă, sunt cei care se găsesc într-un singur exemplar în întreaga genă, având o frecvență de 0,1%.

Cu ajutorul modulului EditSeq al DNA Star, a fost calculată temperatura de topire a ADN, după un model Davis-Botstein-Roth; în acest caz, prezintă o valoare de 82,30ºC.

4.5.3.2. TRANSLAȚIA ŞI CARACTERIZAREA CATENEI POLIPEPTIDICE

Translația s-a realizat pe baza codului genetic mitocondrial pentru vertebrate, utilizând modulul EditSeq din cadrul programului DNA Star (Figura 43).

```
MASLRKTHPL IKIANDALVD LPTPSNISAW WNFGSLLGLC LITQILTGLF
LAMHYTSDIS TAFSSVTHIC RDVNYGWLIR NVHANGASFF FICIYMHIAR
GLYYGSYLYK ETWNIGVILL LLVMMTAFVG YVLPWGQMSF WGATVITNLL
SAVPYMGDML VQWIWGGFSV DNATLTRFFA FHFLLPFVIT AATIIHLLFL
HETGSNNPIG LNSDADKVSF HPYFSYKDLL GFVIMLLALT LLALFSPNLL
GDPENFTPAN PLVTPPHIKP EWYFLFAYAI LRSIPNKLGG VLALLFSILV
LMVVPLLHTS KQRGLTFRPI TQFLFWTLVA DMIILTWIGG MPVEHLIMWH
           ASSYSPHSHY A.RC.MAMTA MLQNKALKWA
```

Figura 43 Catena polipeptidică a citocromului b caracteristică haplotipului Cy0104M al speciei *Cyprinus carpio*

După realizarea translației, se constată că întreg lanțul polipeptidic, este alcătuit din 378 aminoacizi (Tabelul 30), din care, frecvența cea mai mare, o are leucina (15,08%), iar cea mai scăzută o are cisteina (1,06%); aceste date arată că există o similaritate cu structura lanțului polipeptidic al haplotipului, de unde putem concluziona că şi in cazul de față, în poziția 507, s-a produs o mutație nonsens, deci, o mutație care nu determină o modificare a aminoacidului.

Tabel 30 Compoziția lanțului polipeptidic pentru haplotipul Cy0104M

Aminoacizi		Număr	Procent din greutatea totală (%)	Frecvența (%)
Simbol	Nume			
A	Ala	30	5,01	7,94
C	Cys	4	0,97	1,06
D	Asp	11	2,97	2,91
E	Glu	5	1,52	1,32
F	Phe	26	8,98	6,88
G	Gly	23	3,08	6,08
H	His	15	4,83	3,97
I	Ile	29	7,70	7,67
K	Lys	10	3,01	2,65
L	Leu	57	15,14	15,08
M	Met	16	4,93	4,23
N	Asn	17	4,55	4,50
P	Pro	20	4,56	5,29
Q	Gln	6	1,80	1,59
R	Arg	9	3,30	2,38
S	Ser	25	5,11	6,61
T	Thr	25	5,93	6,61
V	Val	22	5,12	5,82
W	Trp	13	5,68	3,44
Y	Tyr	15	5,75	3,97
B	Asx	0	0,00	0,00
Z	Glx	0	0,00	0,00
.	Ter	2	0,00	0,53

4.5.4. CARACTERIZAREA HAPLOTIPULUI CY03M

4.5.4.1. FRECVENȚELE TIPURILOR DE ASOCIERI NUCLEOTIDICE

În acest haplotip se încadrează o singură secvență (Cy03M) care diferă de haplotipul general prin existența unei transversii (substituția timinei de către guanină) în poziția 396.

Cu ajutorul modulului EditSeq al DNA Star, a fost calculată temperatura de topire a ADN, după un model Davis-Botstein-Roth; în acest caz, prezintă o valoare de 82,30ºC.

Referitor la procentele dinucleotidelor constatăm, că valoarea cea mai mare o are perechea AC (10,3%), iar valoarea cea mai mică perechea GG, în proporție de 2,3%. În cazul trinucleotidelor, cea mai mare frecvență o înregistrează tripletul CAT (3,9%), iar valoarea cea mai mică, tripletele GGG şi GCG, ambele având o frecvență de 0,2%.

Tetramerul cu frecvența cea mai mare, este ACTA (1,6%), iar cei cu frecvența redusă sunt: TCTC, CTCA, ATAC, ACCC, toți având o frecvență de 1.0%. În cazul pentamerilor, cea mai mare frecvență o are pentamerul TCCTA (0,9%), iar cea mai mică frecvență, o au TTCCT, TACCA, CTCCT, CTCAT şi CTAGT (0,4%). În ceea ce priveşte hexamerii, aceştia au frecvențe cuprinse între 0,3% şi 0,4%.

Tetramerii, pentamerii şi hexamerii cu frecvență joasă, sunt cei care se găsesc într-un singur exemplar în întreaga genă, având o frecvență de 0,1%.

4.5.4.2. TRANSLAȚIA ŞI CARACTERIZAREA CATENEI POLIPEPTIDICE

Translația s-a realizat pe baza codului genetic mitocondrial pentru vertebrate, utilizând modulul EditSeq din cadrul programului DNA Star (Figura 44).

```
MASLRKTHPL IKIANDALVD LPTPSNISAW WNFGSLLGLC LITQILTGLF
LAMHYTSDIS TAFSSVTHIC RDVNYGWLIR NVHANGASFF FICIYMHIAR
GLYYGSYLYK ETWNIGVILL LLVMMTAFVG YVLPWGQMSF WGATVITNLL
SAVPYMGDML VQWIWGGFSV DNATLTRFFA FHFLLPFVIT AATIIHLLFL
HETGSNNPIG LNSDADKVSF HPYFSYKDLL GFVIMLLALT LLALFSPNLL
GDPENFTPAN PLVTPPHIKP EWYFLFAYAI LRSIPNKLGG VLALLFSILV
LMVVPLLHTS KQRGLTFRPI TQFLFWTLVA DMIILTWIGG MPVEHLIMWH
          ASSYSPHSHY A.RC.MAMTA MLQNKALKWA
```

Figura 44 Catena polipeptidică a citocromului b caracteristică haplotipului Cy03M al speciei *Cyprinus carpio*

După realizarea translației, se constată că întreg lanțul polipeptidic, este alcătuit din 378 aminoacizi (Tabelul 31), dintre care, frecvența cea mai mare, o are leucina (15,08%), iar cea mai scăzută o are cisteina (1,06%); aceste date arată că există o similaritate cu structura lanțului polipeptidic al haplotipului, de unde putem concluziona că şi in cazul de față, în poziția 396, s-a produs o mutație nonsens, deci, o mutație care nu determină o modificare a aminoacidului.

Tabel 31 Compoziția lanțului polipeptidic pentru haplotipul Cy03M

Aminoacizi		Număr	Procent din greutatea totală (%)	Frecvența (%)
Simbol	Nume			
A	Ala	30	5,01	7,94
C	Cys	4	0,97	1,06
D	Asp	11	2,97	2,91
E	Glu	5	1,52	1,32
F	Phe	26	8,98	6,88
G	Gly	23	3,08	6,08
H	His	15	4,83	3,97
I	Ile	29	7,70	7,67
K	Lys	10	3,01	2,65
L	Leu	57	15,14	15,08
M	Met	16	4,93	4,23
N	Asn	17	4,55	4,50
P	Pro	20	4,56	5,29
Q	Gln	6	1,80	1,59
R	Arg	9	3,30	2,38
S	Ser	25	5,11	6,61
T	Thr	25	5,93	6,61
V	Val	22	5,12	5,82
W	Trp	13	5,68	3,44
Y	Tyr	15	5,75	3,97
B	Asx	0	0,00	0,00
.	Ter	2	0,00	0,53

4.5.5. CARACTERIZAREA HAPLOTIPULUI CY06M

4.5.5.1. FRECVENȚELE TIPURILOR DE ASOCIERI NUCLEOTIDICE

În acest haplotip se încadrează o singură secvență (Cy06M) caracteristică crapului oglindă care diferă de haplotipul general prin mai multe substituții de tipul:

- tranziție în poziția 177 (substituția citozinei de către timină);
- tranziție în poziția 198 (substituția timinei de către citozină);
- tranziție în poziția 352 (substituția adeninei de către guanină);
- tranziție în poziția 568 (substituția adeninei de către guanină);
- tranziție în poziția 591 (substituția guaninei de către adenină);
- tranziție în poziția 624 (substituția adeninei de către guanină);
- tranziție în poziția 693 (substituția adeninei de către guanină);
- tranziție în poziția 732 (substituția guaninei cu adenină);
- tranziție în poziția 864 (substituția timinei cu citozină);
- transversie în poziția 870 (substituția timinei cu adenină);
- tranziție în poziția 897 (substituția adeninei cu guanină);
- tranziție în poziția 990 (substituția adeninei de către guanină).

Referitor la procentele dinucleotidelor constatăm, că valoarea cea mai mare o are perechea AC (10,3%), iar valoarea cea mai mică perechea GG, în proporție de 2,5%. Referitor la trinucleotide, cea mai mare frecvență o înregistrează tripletul CAT (3,9%), iar valoarea cea mai mică, tripletul GTG, ambele având o frecvență de 0,2%.

Tetramerul cu frecvența cea mai mare, este ACTA (1,7%), iar cei cu frecvența redusă sunt: CTAT, CACC, ATAC, ACCC, toți având o frecvență de 1.0%. În cazul pentamerilor, cea mai mare frecvență o are pentamerul TCCTA (0,9%), iar cea mai mică frecvență, o au TTCCT, TACCA şi CTCCT (0,4%). În ceea ce priveşte hexamerii, aceştia au frecvențe cuprinse între 0,3% şi 0,4%.

Tetramerii, pentamerii şi hexamerii cu frecvență joasă, sunt cei care se găsesc într-un singur exemplar în întreaga genă, având o frecvență de 0,1%.

Cu ajutorul modulului EditSeq al DNA Star, a fost calculată temperatura de topire a ADN, după un model Davis-Botstein-Roth; în acest caz, prezintă o valoare de 82,26ºC.

4.5.5.2. TRANSLAȚIA ŞI CARACTERIZAREA CATENEI POLIPEPTIDICE

Translația s-a realizat pe baza unui cod genetic mitocondrial pentru vertebrate, utilizând modulul EditSeq din cadrul programului DNA Star.

Catena polipeptidică obținută este prezentată în Figura 45.

```
MASLRKTHPL IKIANDALVD LPTPSNISAW WNFGSLLGLC LITQILTGLF
LAMHYTSDIS TAFSSVTHIC RDVNYGWLIR NVHANGASFF FICIYMHIAR
GLYYGSYLYK ETWNIGVVLL LLVMMTAFVG YVLPWGQMSF WGATVITNLL
SAVPYMGDML VQWIWGGFSV DNATLTRFFA FHFLLPFVIA AATIIHLLFL
HETGSNNPIG LNSDADKVSF HPYFSYKDLL GFVIMLLALT LLALFSPNLL
GDPENFTPAN PLVTPPHIKP EWYFLFAYAI LRSIPNKLGG VLALLFSILV
LMVVPLLHTS KQRGLTFRPI TQFLFWTLVA DMIILTWIGG MPVEHLIMWH
          ASSYSPHSHY A.RC.MAMTA MLQNKALKWA
```

Figura 45 Catena polipeptidică a citocromului b caracteristică haplotipului Cy06M al speciei *Cyprinus carpio* var. crap oglindă

Tabel 32 Compoziția lanțului polipeptidic

Aminoacizi		Număr	Procent din greutatea totală (%)	Frecvența (%)
Simbol	Nume			
A	Ala	31	5,18	8,20
C	Cys	4	0,97	1,06
D	Asp	11	2,98	2,91
E	Glu	5	1,52	1,32
F	Phe	26	8,99	6,88
G	Gly	23	3,08	6,08
H	His	15	4,83	3,97
I	Ile	28	7,45	7,41
K	Lys	10	3,01	2,65
L	Leu	57	15,16	15,08
M	Met	16	4,93	4,23
N	Asn	17	4,56	4,50
P	Pro	20	4,57	5,29
Q	Gln	6	1,81	1,59
R	Arg	9	3,30	2,38
S	Ser	25	5,12	6,61
T	Thr	24	5,70	6,35
V	Val	23	5,36	6,08
W	Trp	13	5,69	3,44
Y	Tyr	15	5,57	3,97
B	Asx	0	0,00	0,00
.	Ter	2	0,00	0,53

După realizarea translației, se constată că întreg lanțul polipeptidic este alcătuit din 378 aminoacizi (Tabelul 32), din care se observă că frecvența cea mai mare, o are leucina (15,16%), iar cea mai scăzută o are cisteina (1,06%). Se constată de asemenea, existența a doi codoni stop (notați cu „.”) către sfârşitul lanțului polipeptidic.

Se poate observa că din totalul de 378 aminoacizi, 16 sunt acizi, 19 sunt bazici, 91 polari şi 178 hidrofobi (Tabelul 33). De asemenea, se constată că lanțul polipeptidic are o sarcină electrică de 5,28 (la pH=7) şi un punct izoelectric de 8,42 (Figura 46).

Tabel 33 Caracterizarea lanțului polipeptidic

Greutate moleculară (Daltoni)	Număr de aminoacizi	Aminoacizi				Punct izoelectric	Sarcina electrică la pH=7	Concentrația la o extincție =1 şi λ=280nm
		Puternic bazici	Puternic acizi	Hidrofobi	Polari			
42548,51	378	19	16	178	91	8,423	5,279	0,45

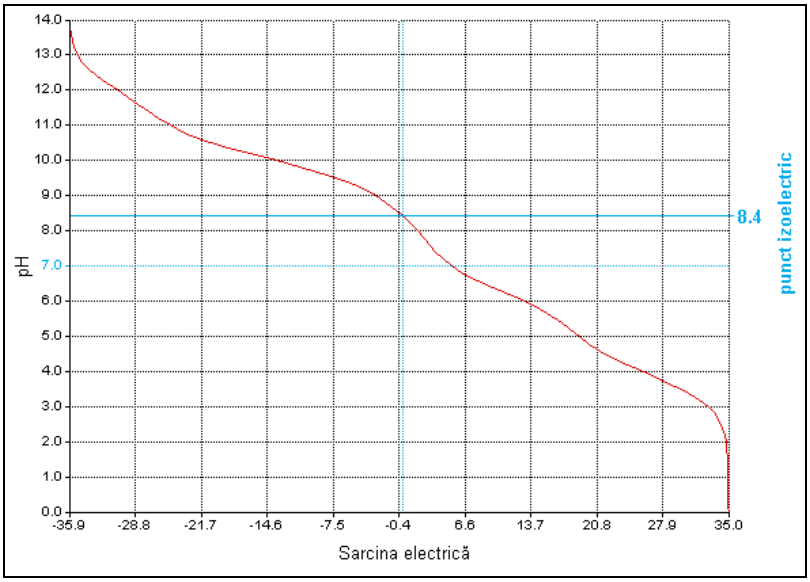


Figura 46 Curba de titrare pentru lanțul polipeptidic al citocromului b, haplotip Cy06M al speciei *Cyprinus carpio* var. crap oglindă

Ulterior, a fost realizată o analiză a structurii proteinei; astfel, au fost determinate zonele alfa helicoidale, beta pliate şi de inflexiune, prin două metode Garnier-Robson (Garnier et al., 1978) şi Chou-Fasman (Chou şi Fasman, 1978); reprezentarea grafică a zonelor hidrofobe (Kyte şi Doolittle, 1982), zonele de flexie (Karplus şi Schultz, 1985), indexul antigenic (Jameson şi Wolf, 1988) şi graficul probabil al suprafeței (Emini et al., 1985).

Din analiza structurii lanțului polipeptidic, se poate observa că referitor la regiunile alfa helicoidale şi în acest caz există diferențe între cele două metode Garnier-Robson şi Chou-Fasman. Astfel, în cazul primei metode, se observă că există 17 regiuni alfa helicoidale, în timp ce, potrivit celei de-a doua metode, există doar 6, existând 5 puncte de suprapunere ale suprafețelor comparativ între cele două metode. De asemenea, pentru regiunile beta pliate, se observă că, pentru determinarea efectuată cu ajutorul primei metode, se constată existența a 30 de regiuni beta pliate, în timp ce, pentru cea de-a doua metodă (Chou-Fasman), sunt considerate ca existând doar 11 astfel de regiuni. Trebuie remarcat faptul că similaritatea în acest caz este dată de suprafețele ample considerate ca fiind beta pliate de cea de-a doua metodă. Referitor la zonele de inflexiune, de asemenea, comparativ pentru cele două metode, (Garnier-Robson şi Chou-Fasman) sunt date 22 astfel de regiuni, cu diferența că în cazul celei de-a doua, sunt prezentate ca regiuni cu suprafețe mai mari.

Pe baza graficului hidrofobicității realizat după un model Kyte-Doolitle (Kyte şi Doolittle, 1982), se constată că există 15 zone hidrofobe şi 12 zone hidrofile.

Din graficul indexului antigenic calculat pe baza modelului Jameson-Wolf (Jameson şi Wolf, 1988), se constată existența a 26 de regiuni cu potențial antigenic.

4.5.6. RELAȚII FILOGENETICE ÎN CADRUL GENULUI *CYPRINUS*, STABILITE PE BAZA SECVENȚELOR CITOCROMULUI B

Metodele folosite în filogenie şi modalitățile de realizare a arborilor filogenetici, utilizate în acest caz, sunt cele care includ tranzițiile şi transversiile, deoarece haplotipurile analizate pentru cele două populații, prezintă ambele tipuri de mutații.

Astfel, prin metoda UPGMA bazată pe modelul Tamura – Nei, care include ambele tipuri de mutații posibile (tranziții şi transversii), au fost obținuți arbori filogenetici (Figurile 47, 48 şi 49), din care se observă că Cy06 derivă direct dintr-un nod comun ancestral, în care îşi are originea şi brațul comun al celorlalți indivizi.

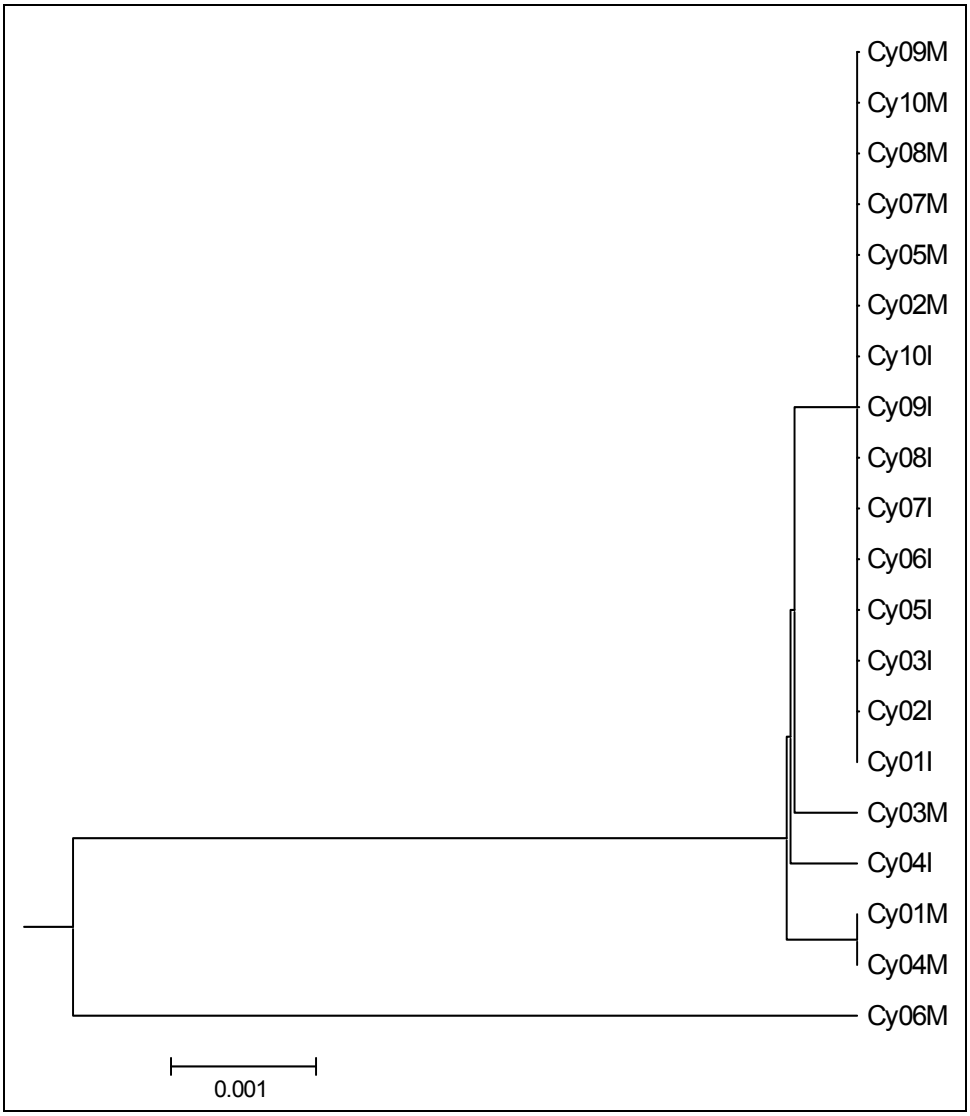


Figura 47 Arbore filogenetic rectangular al genului *Cyprinus*, realizat prin metoda UPGMA pe baza secvențelor citocromului B

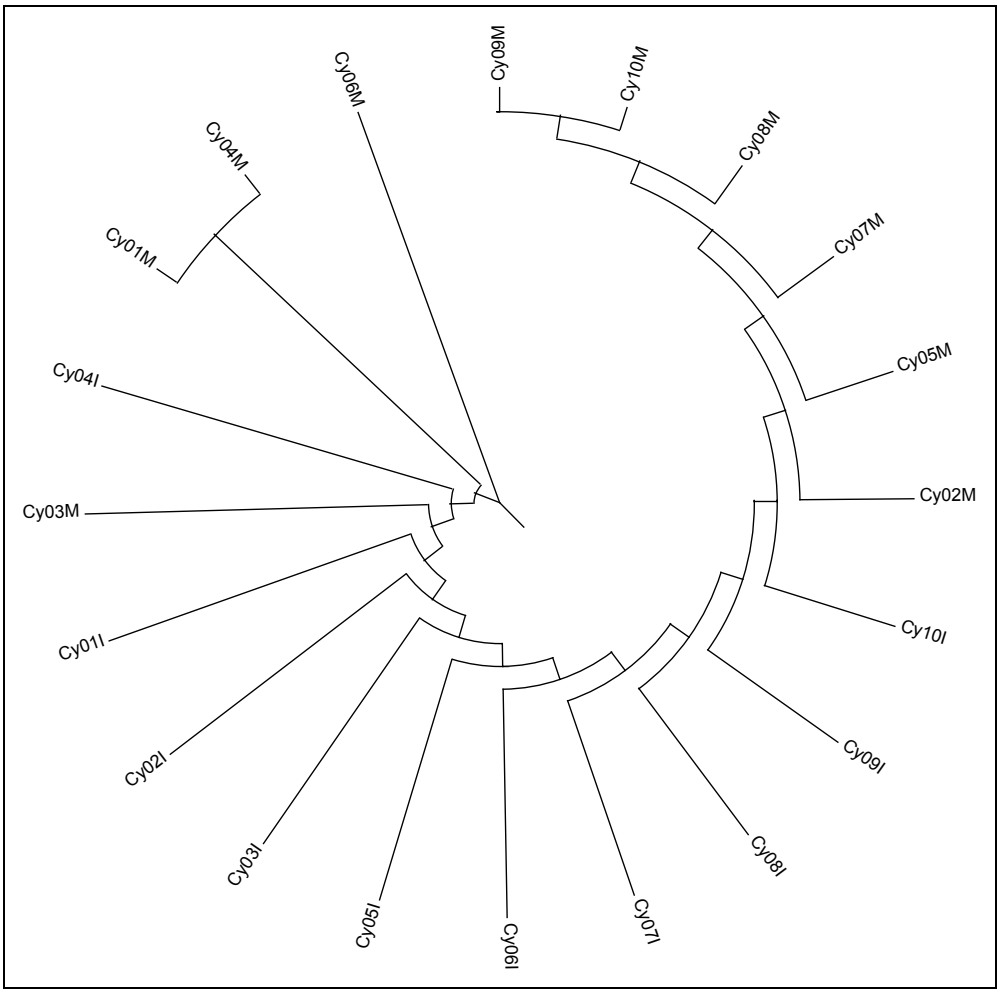


Figura 48 Arbore filogenetic circular al genului *Cyprinus*, realizat prin metoda UPGMA pe baza secvențelor citocromului B

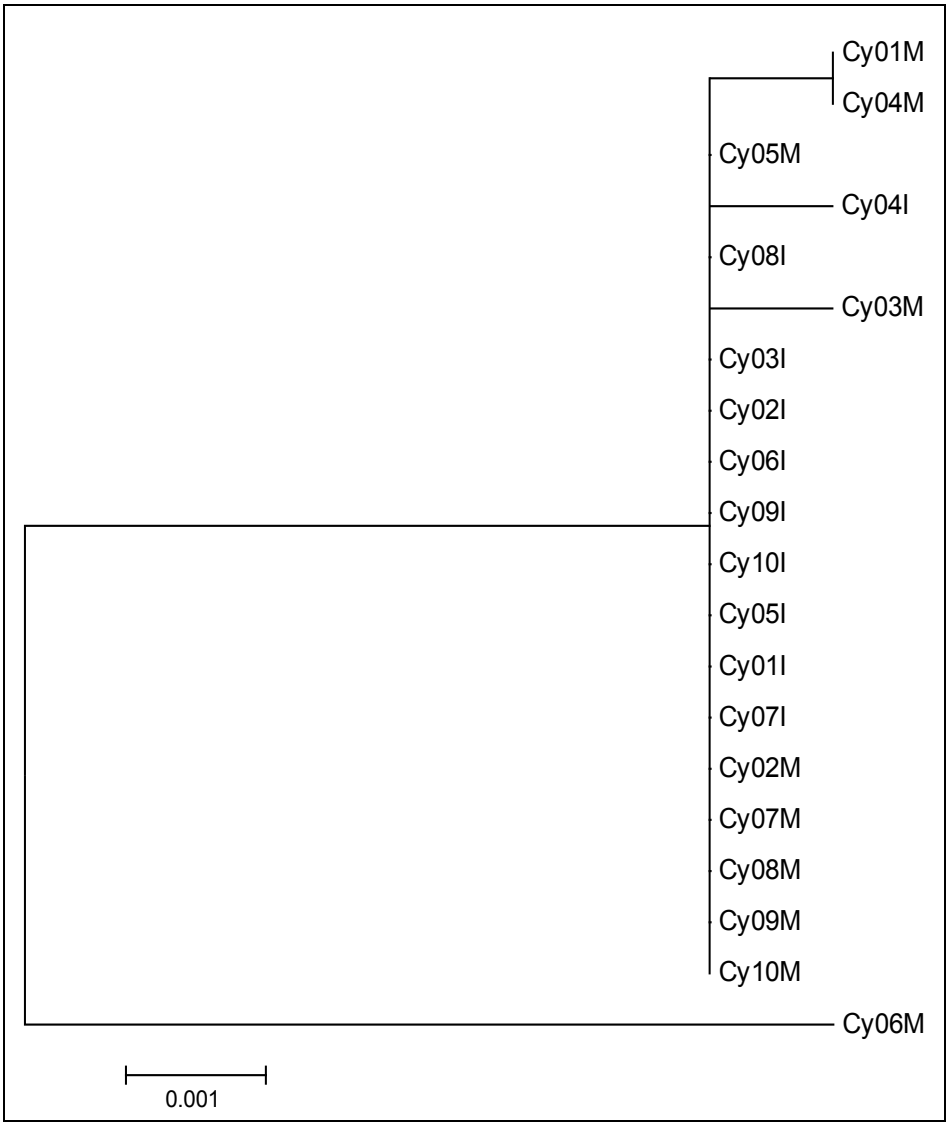


Figura 49 Arbore filogenetic rectangular al genului *Cyprinus*, realizat prin metoda evoluției minime (ME) după un model neighbor joining, pe baza secvențelor citocromului B

Se observă că există o similaritate între cele două tipuri de arbori realizați prin metode diferite.

Nu a fost folosită metoda maximum parsimony în acest caz, deoarece diversitatea taxonilor este scăzută, ceea ce poate duce la o creştere a probabilității de apariție a erorilor.

4.6. COMPARAREA SECVENȚELOR NUCLEOTIDICE ALE REGIUNII DE CONTROL MITOCONDRIAL, PROVENITE DE LA INDIVIZI APARȚINÂND GENULUI *CYPRINUS*

Pentru regiunea de control mitocondrial, în cazul speciilor genului *Cyprinus*, a fost secvențiată prima parte hipervariabilă a d-loop (fapt precizat şi în capitolele anterioare). Ulterior, au fost aliniate cele 20 de secvențe (19 indivizi aparținând celor două populații, 1 individ de crap oglindă Cyo06M şi o secvență provenită de la un individ de crap, preluată din banca de gene pentru comparație – GenBank Ref. nr, AF 508931), utilizând metoda Clustal V (Higgins şi Sharp, 1989).

Din alinierea celor 21 de secvențe, se constată existența a 11 diferențe, majoritatea fiind întâlnite în cadrul secvenței Cy06M, provenită de la un individ de crap oglindă. Astfel, pentru acest individ, comparativ cu celelalte secvențe, au fost constatate 8 diferențe, dintre care 5 tranziții, o transversie şi o deleție.

Alte diferențe, au fost constatate pentru indivizii Cy01M (o tranziție în poziția 388), Cy05M (o tranziție în poziția 285) şi pentru secvența provenită din banca de gene, o tranziție în poziția 388.

Pentru restul indivizilor, nu au fost înregistrate mutații la acest nivel al ADNmt. În plus, putem constata că pentru populația Iaşi, nu au fost constatate diferențe, deci, rata de apariție a mutațiilor în această populație este mai scăzută.

Din cele prezentate anterior, putem estima că au fost descoperite 4 noi haplotipuri pentru prima parte a regiunii de control mitocondrial: Cy01M, Cy05M şi Cy06M pentru populația Movileni (Cy06M pentru populația de crap oglindă Movileni) şi haplotipul general, caracteristic celor două populații, dat de secvențele similare.

Pe baza comparației între secvențe, s-a realizat tabelul similarității şi divergenței (Tabel 34), din care se constată că procentele de similaritate sunt cuprinse între 100% (pentru secvențele Cy01M şi secvența preluată din banca de gene) şi 68,7% pentru secvența Cy06M şi celelalte secvențe care constituie haplotipul general caracteristic celor două populații.

Tabel 34 Procentele de similaritate şi divergență pentru secvențele haplotipurilor stabilite

	Procent de similaritate						
Procent de divergență		Cy01I	Cy01M	Cy05M	Cy06M	Cy GB	
	Cy01I		99,7	99,7	68,7	99,7	Cy01I
	Cy01M	0,3		99,5	69,0	100	Cy01M
	Cy05M	0,3	0,5		69,0	99,5	Cy05M
	Cy06M	40,6	40,2	40,2		69,0	Cy06M
	Cy GB	0,3	0,0	0,5	40,2		Cy GB
		Cy01I	Cy01M	Cy05M	Cy06M	Cy GB	

În plus, pentru a stabili o mai clară diferențiere între toate noile haplotipuri, au fost analizate numărul şi procentele în care se găsesc bazele azotate în fiecare dintre acestea (Tabel 35).

Tabel 35 Procentul bazelor azotate, în cadrul haplotipurilor stabilite

Baze azotate		Specii analizate				
		Cy01I	Cy01M	Cy05M	Cy06M	Cy GB
A	Nr.	145	145	146	143	145
	%	36,90	36,90	37,15	36,39	36,90
G	Nr.	54	54	53	55	54
	%	13,74	13,74	13,49	13,99	13,74
T	Nr.	129	130	129	131	130
	%	32,82	33,08	32,82	33,33	33,08
C	Nr.	65	64	65	64	64
	%	16,54	16,28	16,54	16,28	16,28
A+T	Nr.	274	275	275	274	275
	%	69,72	69,97	69,97	69,72	69,97
C+G	Nr.	119	118	118	119	118
	%	30,28	30,03	30,03	30,28	30,03

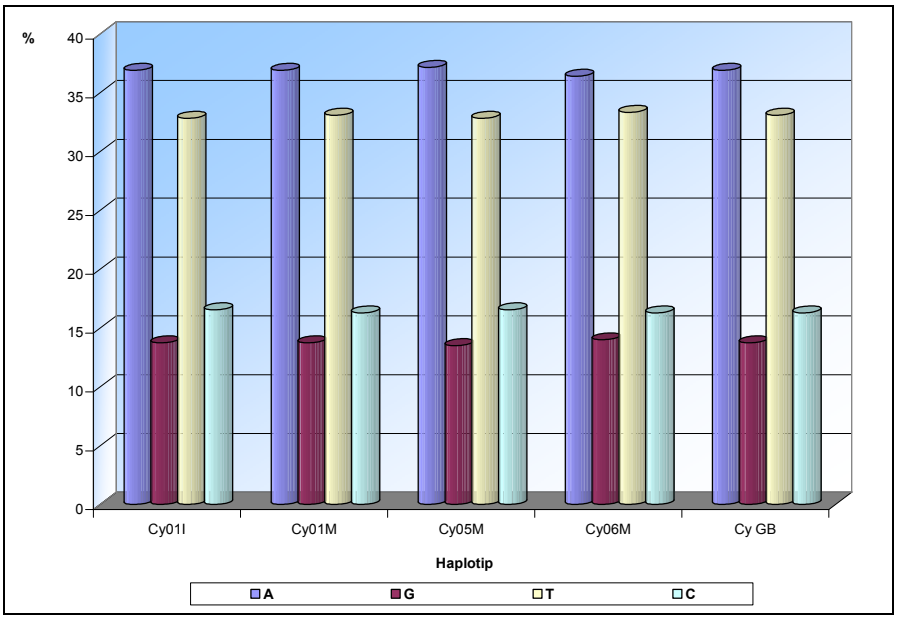


Figura 50 Graficul frecvenței bazelor azotate în cadrul haplotipurilor caracterizate

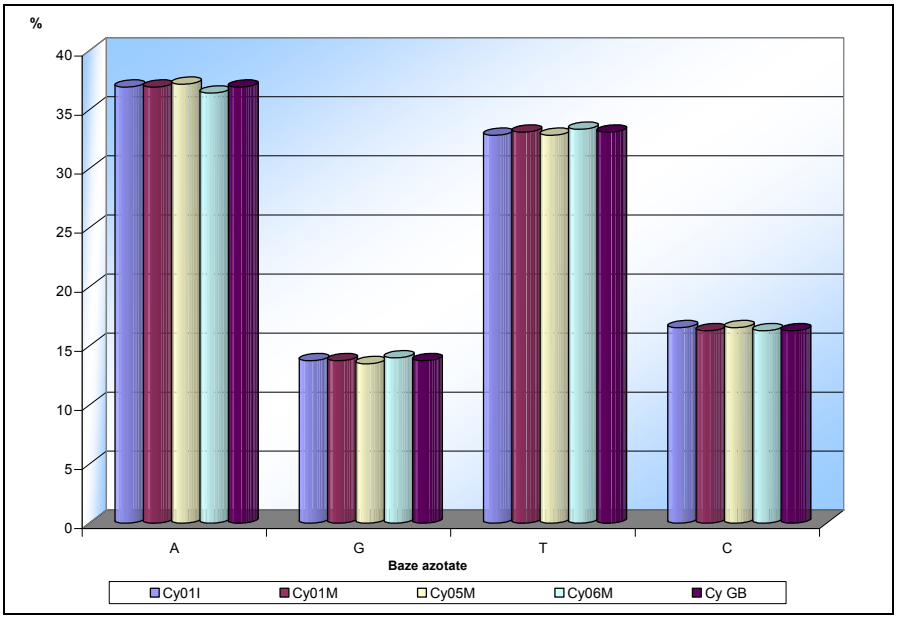


Figura 51 Graficul frecvenței bazelor azotate, comparativ între haplotipurile analizate

Din Tabelul 35 şi Figurile 50 şi 51, se poate observa că diferențe majore se constată pentru secvența Cy06M, care prezintă valori diferite pentru trei dintre bazele azotate (A, G, T), comparativ cu celelalte haplotipuri.

Referitor la cantitatea de adenină (A), aceasta este egală pentru Cy01I, Cy01M şi Cy GB (36,90%), pentru haplotipul Cy05M se găseşte în proporție de 37,15%, iar pentru Cy06M, în proporție de 36,39%. Cantitatea de guanină este cuprinsă între 13,49% (Cy05M) şi 13,99 pentru Cy06M. De asemeni, se constată diferențe şi pentru celelalte baze azotate: cantitatea de timină variază între 32,82% pentru Cy01I, Cy05M şi 33,33% pentru Cy06M, iar citozina, între 18,28% pentru Cy01I (ca reprezentant al haplotipului general caracteristic celor două populații), Cy06M şi Cy GB – 18,54% pentru haplotipurile Cy01I şi Cy05M.

Procentul de baze complementare este de 69,72% în cazul A+T pentru Cy01I şi Cy06M; 69,97% pentru celelalte trei haplotipuri. Procentul de G+C este de 30,28% (Cy01I, Cy06M) şi de 30,03% pentru Cy01M, Cy05M şi Cy GB.

4.6.1. CARACTERIZAREA HAPLOTIPULUI GENERAL AL CELOR DOUĂ POPULAȚII

4.6.1.1. FRECVENȚELE TIPURILOR DE ASOCIERI NUCLEOTIDICE

În acest haplotip, au fost încadrate toate secvențele între care procentele de similaritate sunt de 100% (Cy01I, Cy02I, Cy03I, Cy04I, Cy05I, Cy06I, Cy07I, Cy08I, Cy09I, Cy10I, Cy02M, Cy03M, Cy04M, Cy07M, Cy08M, Cy09M, Cy10M), aparținând celor două populații luate în studiu (Movileni şi Iaşi).

Referitor la procentele dinucleotidelor constatăm, că valoarea cea mai mare o are perechea AT (14,5%), iar valoarea cea mai mică perechea GC, în proporție de 1,8%, fiind absente dinucleotidele de tipul CG.

În ceea ce priveşte trinucleotidele, cea mai mare frecvență o înregistrează tripletul TAT (6,6%), iar valoarea cea mai mică, tripletele GTC şi TGC, ambele având o frecvență de 0,3%.

Tetramerul cu frecvența cea mai mare, este TATT (3,6%), iar cel cu frecvența cea mai mică TTTT, în procent de 1,0%. În cazul pentamerilor, cea mai mare frecvență o are pentamerul TATTA (2,8%), iar cea mai mică frecvență, o au TTAAT, TATTT, TATAT, TAGTA, TAATC şi TAATA (0,8%). În ceea ce priveşte hexamerii, aceştia au frecvențe cuprinse între 1,5% (ATATTA) şi 0,5%.

Tetramerii, pentamerii şi hexamerii cu frecvență joasă, sunt cei care se găsesc într-un singur exemplar în întreaga genă, având o frecvență de 0,3%.

Cu ajutorul modulului EditSeq al DNA Star, a fost calculată temperatura de topire a ADN, după un model Davis-Botstein-Roth; în acest caz, prezintă o valoare de 76,04ºC.

4.6.2. CARACTERIZAREA HAPLOTIPULUI CY01M

4.6.2.1. FRECVENȚELE TIPURILOR DE ASOCIERI NUCLEOTIDICE

În acest haplotip, se încadrează o singură secvență Cy01M din populația Larga-Jijia – Mărgineni, caracteristică unui individ de tip sălbatic aparținând speciei *Cyprinus carpio*. După cum s-a observat în tabelul de similaritate, această secvență este identică cu secvența preluată din banca de gene, ceea ce ne permite să concluzionăm că cele două populații studiate prezintă haplotipuri distincte de cele secvențiate până acum, aflându-se în proces de speciație. În plus, se poate constata că mutația apărută în populațiile studiate, datorită frecvenței cu care apare, este deja fixată în genofondul acestora.

Din analiza procentelor dinucleotidelor, constatăm că valoarea cea mai mare o are perechea AT (14,8%), iar valoarea cea mai mică perechea GC, în proporție de 1,8%; şi în acest caz sunt absente dinucleotidele CG. Referitor la trinucleotide, cea mai mare frecvență o înregistrează tripletul TAT (6,9%), iar valoarea cea mai mică, tripletele GTC şi TGC, ambele având o frecvență de 0,3%.

Tetramerul cu frecvența cea mai mare, este TATT (3,8%), iar cel cu frecvența redusă TTTT (1,0%). În cazul pentamerilor, cea mai mare frecvență o are pentamerul TATTA (2,8%), iar cea mai mică frecvență, o au TTAAT, TATAT, TAGTA, TAATC ŞI TAATA, toți cu o frecvență de 0,8%. În ceea ce priveşte hexamerii, aceştia au frecvențe cuprinse între 0,5% şi 1,5%.

Tetramerii, pentamerii şi hexamerii cu frecvență joasă, sunt cei care se găsesc într-un singur exemplar în întreaga genă, având o frecvență de 0,3%.

Cu ajutorul modulului EditSeq al DNA Star, a fost calculată temperatura de topire a ADN, după un model Davis-Botstein-Roth; în acest caz, prezintă o valoare de 75,94ºC.

4.6.2.2. MODELAREA STRUCTURII SECUNDARE A ARNm

Modelarea structurii ARN, s-a realizat pe baza unei energii minime de -49,0. Deoarece acest haplotip diferă de cel general doar printr-o singură bază azotată – înlocuirea citozinei cu timină în poziția 388 în urma unei tranziții, nu apar modificări majore în aspectul general al macromoleculei de ARNm. Singura diferență existentă

în acest caz, între ARNm din haplotipul general pentru cele două populații şi cel al prezentei secvențe de ADN constă în existența uracilului în locul citozinei.

De asemenea, se constată ca şi în cazul precedent, că structura secundară a ARNm prezintă 31 de porțiuni monocatenare şi 14 regiuni bicatenare.

4.6.3. CARACTERIZAREA HAPLOTIPULUI CY05M

4.6.3.1. FRECVENȚELE TIPURILOR DE ASOCIERI NUCLEOTIDICE

În acest haplotip, se încadrează o singură secvență Cy05M din populația Larga-Jijia – Mărgineni, caracteristică unui individ de tip sălbatic aparținând speciei *Cyprinus carpio*. Diferența dintre acest haplotip şi cel general al celor două populații, constă în substituția guaninei de către adenină, printr-un proces de tranziție, în poziția 285.

Referitor la procentele dinucleotidelor constatăm că valoarea cea mai mare o are perechea AT (14,5%), iar valoarea cea mai mică perechea GC, în proporție de 1,8%; şi în acest caz fiind absente dinucleotidele de tip CG.

Cea mai mare frecvență o înregistrează tripletul TAT (6,6%), iar valoarea cea mai mică, tripletele GTC şi TGC, ambele având o frecvență de 0,3%.

Tetramerul cu frecvența cea mai mare, este TATT (3,6%), iar cel cu frecvența redusă TTTT (1,0%). În cazul pentamerilor, cea mai mare frecvență o are pentamerul TATTA (2,8%), iar cea mai mică frecvență, o au TTAAT, TATTT, TATAT, TAGTA, TAATC ŞI TAATA, toți cu o frecvență de 0,8%. În ceea ce priveşte hexamerii, aceştia au frecvențe cuprinse între 0,5% şi 1,5%.

Tetramerii, pentamerii şi hexamerii cu frecvență joasă, sunt cei care se găsesc într-un singur exemplar în întreaga genă, având o frecvență de 0,3%.

Cu ajutorul modulului EditSeq al DNA Star, a fost calculată temperatura de topire a ADN, după un model Davis-Botstein-Roth; în acest caz, prezintă o valoare de 75,94ºC.

4.6.4. CARACTERIZAREA HAPLOTIPULUI CY06MD

4.6.4.1. FRECVENȚELE TIPURILOR DE ASOCIERI NUCLEOTIDICE

Haplotipul cuprinde o singură secvență Cy06M care diferă de haplotipul general prin apariția următoarelor modificări ale catenei nucleotidice:

- transversie în poziția 87 (substituția adeninei cu citozină);
- o deleție în poziția 217;
- tranziție în poziția 257 (substituția citozinei cu timina);
- tranziție în poziția 262 (înlocuirea citozinei cu timină);
- tranziție în poziția 263 (înlocuirea timinei cu citozină); practic, în aceste două ultime poziții, putem presupune că de fapt s-a produs o inversie care a cuprins cele două nucleotide;
- tranziție în poziția 310 (substituția citozinei cu timină);
- tranziție în poziția 340 (substituția guaninei cu adenină);
- tranziție în poziția 353 (substituția adeninei cu guanină).

Din analiza procentelor dinucleotidelor constatăm că valoarea cea mai mare o are perechea AT (14,3%), iar valoarea cea mai mică perechea GC, în proporție de 1,8%; şi în acest caz fiind absente dinucleotidele de tip CG.

Referitor la trinucleotide, cea mai mare frecvență o înregistrează tripletul TAT (6,6%), iar valoarea cea mai mică, tripletele GTC şi GAT, ambele având o frecvență de 0,3%.

Tetramerul cu frecvența cea mai mare, este TATT (3,6%), iar cei cu frecvență redusă TTTT, TGAA, TCAA (1,0%). În cazul pentamerilor, cea mai mare frecvență o are pentamerul TATTA (2,8%), iar cea mai mică frecvență, o au TTAAT, TATTT, TATAT, TAGTA, TAATA, TAAAG, GTACA, GCATA şi CATAT, toți cu o frecvență de 0,8%. În ceea ce priveşte hexamerii, aceştia au frecvențe cuprinse între 0,5% şi 1,5%.

Tetramerii, pentamerii şi hexamerii cu frecvență joasă, sunt cei care se găsesc într-un singur exemplar în întreaga genă, având o frecvență de 0,3%.

Cu ajutorul modulului EditSeq al DNA Star, a fost calculată temperatura de topire a ADN, după un model Davis-Botstein-Roth; în acest caz, prezintă o valoare de 76,04ºC.

4.6.5. RELAȚII FILOGENETICE ÎN CADRUL GENULUI *CYPRINUS*, BAZATE PE DIFERENȚE ALE SECVENȚELOR D-LOOP

Ca şi în cazul citocromului b, au fost utilizate metode care se bazează pe distanțe şi rata de substituție, dar care implică şi mutații de genul tranzițiilor şi transversiilor. Astfel şi în acest caz au fost utilizate UPGMA (Figura 52) pe baza unui model Tamura-Nei, metoda evoluției minime (ME) după un algoritm Neighbour-Joining (Figura 53) şi Maximum-Parsimony (Figura 54).

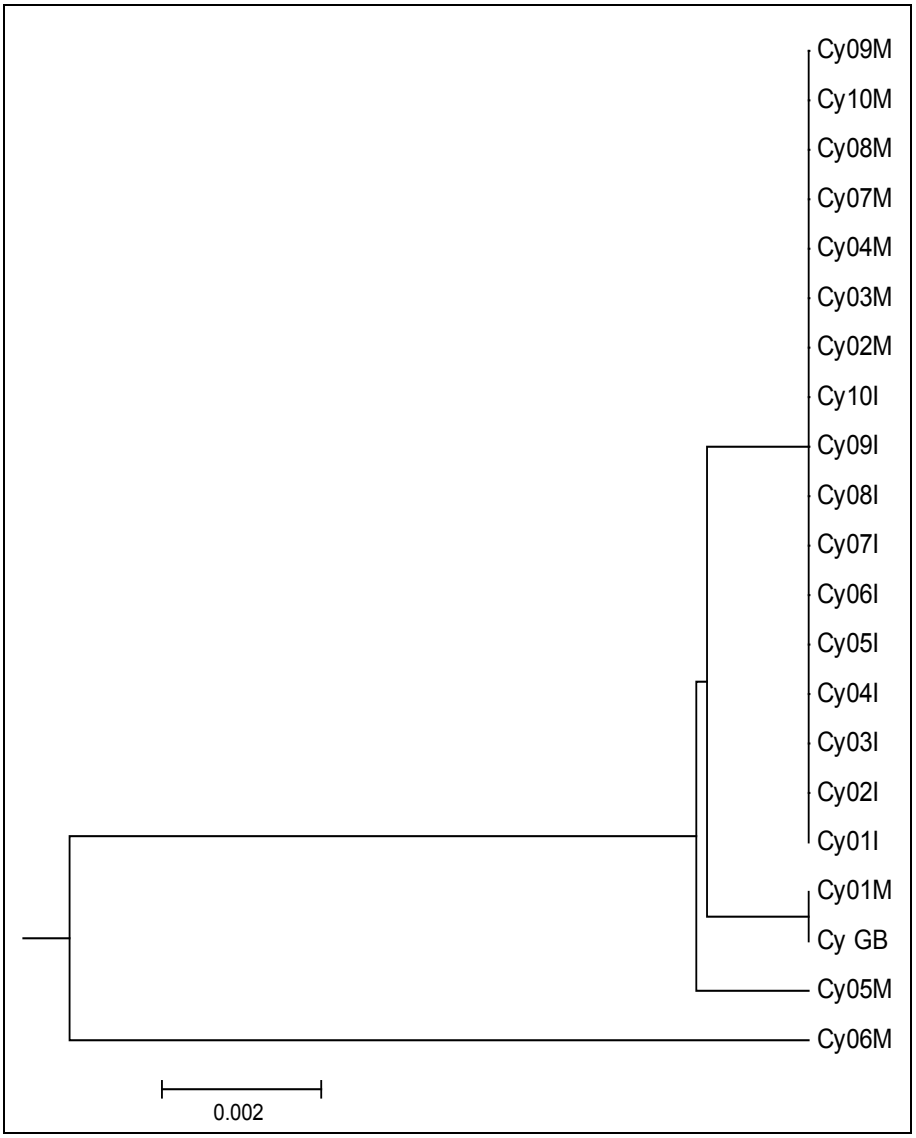


Figura 52 Arbore filogenetic rectangular al genului *Cyprinus*, realizat prin metoda UPGMA pe baza secvențelor d-loop

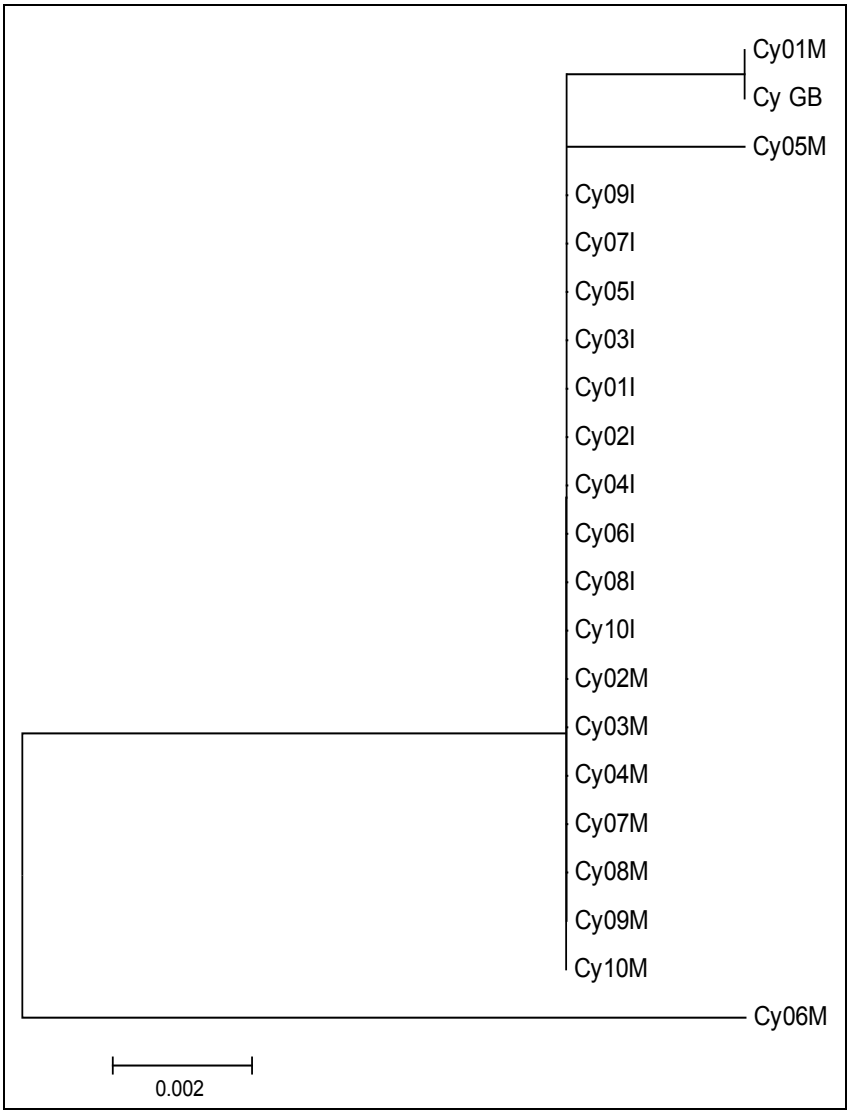


Figura 53 Arbore filogenetic al genului *Cyprinus*, realizat prin metoda ME pe baza secvențelor d-loop

Din cele două figuri, se observă că şi în acest caz, Cy06M provine dintr-un nod comun ancestral, din care provin şi ramurile comune pentru celelalte secvențe. De asemenea, se observă că secvențele Cy01M şi Cy GB, provin dintr-un ram comun, din care provine şi ramul pentru Cy05M.

Ulterior, a fost construit un arbore filogenetic bazat pe metoda Maximum Parsimony (MP) (Figura 54). Din acesta, se poate observa că secvența Cy06M a urmat o linie evolutivă monofiletică dintr-o specie ancestrală. De asemenea, se constată că secvențele Cy01I şi Cy02I, precum şi Cy01M, CyGB, au evoluat din noduri comune.

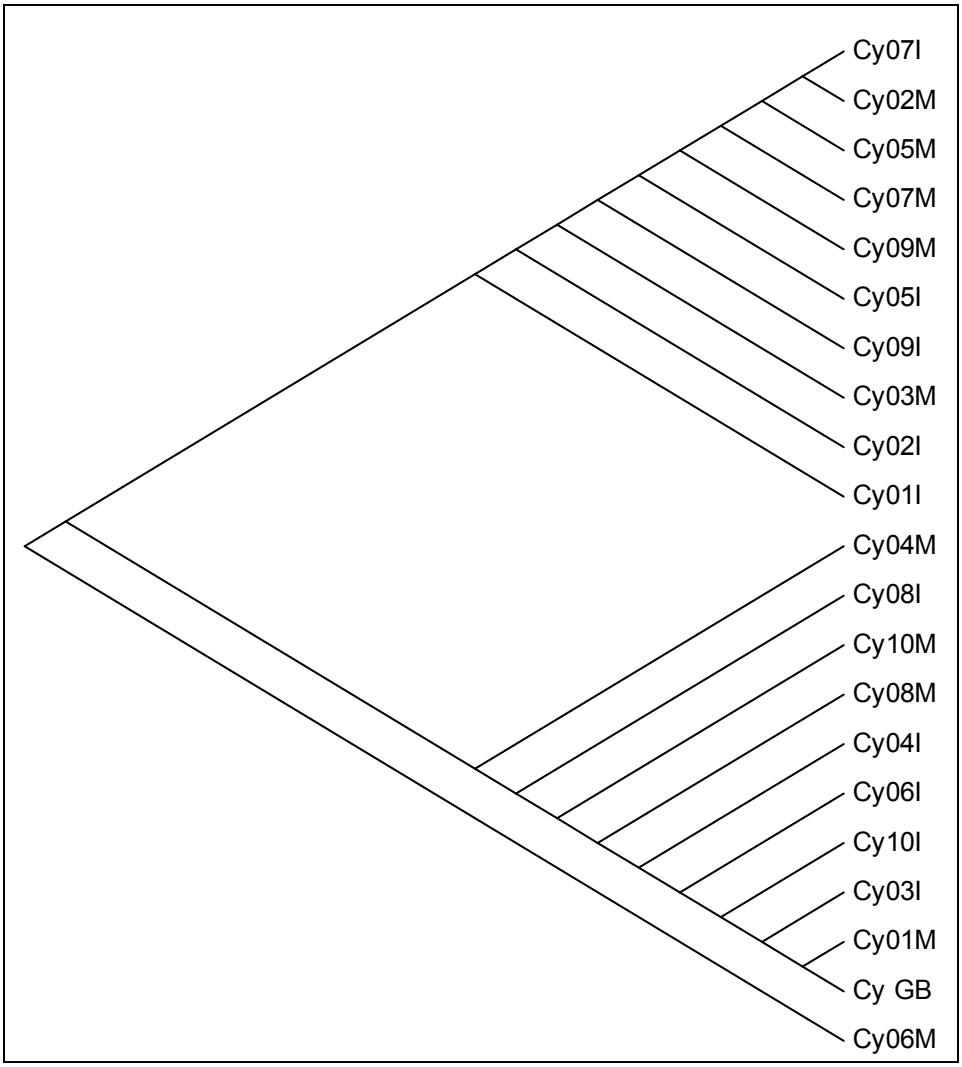


Figura 54 Arbore filogenetic al genului *Cyprinus*, realizat prin metoda MP pe baza secvențelor d-loop

4.7. ANALIZA SECVENȚELOR NUCLEOTIDICE ALE ND4L DEHIDROGENAZEI, PROVENITE DE LA INDIVIZI APARȚINÂND GENULUI *CYPRINUS*

Pentru subunitatea 4L a dehidrogenazei (ND4L), în cazul speciilor genului *Cyprinus*, au fost aliniate cele 20 de secvențe (19 indivizi aparținând celor două populații, 1 individ de crap oglindă Cyo06M şi o secvență provenită de la un individ de crap, preluată din banca de gene pentru comparație – GenBank Ref. nr. AF 508931), utilizând metoda Clustal V (Higgins şi Sharp, 1989).

Din alinierea celor 21 de secvențe, se constată că nu există diferențe între indivizii celor două populații (Larga-Jijia – Movileni şi Iaşi) la acest nivel al ADNmt. În plus,

putem concluziona că ND4L este o genă conservativă, cu variabilitate scăzută. Totuşi, comparând secvența Cy GB cu haplotipul general al celor două populații, se constată existența a 3 diferențe. Astfel pentru secvențele celor două populații, în poziția 165 a avut loc o tranziție (adenina a fost înlocuită cu guanină), o tranziție în poziția 180 (substituția timinei cu citozină) şi o transversie în poziția 258 (substituția adeninei cu citozină).

Din cele precizate anterior, deducem că, a fost găsit un haplotip nou, caracteristic celor două populații (Cy01IM).

Pe baza comparației între secvențe, s-a realizat tabelul similarității şi divergenței (Tabel 36), din care se constată că procentele de similaritate sunt de 100% pentru secvențele Cy01I şi Cy01M (ca reprezentanți ai celor două populații) şi 98,9% pentru comparația între secvența Cy GB şi celelalte secvențe care constituie haplotipul general caracteristic celor două populații.

Tabel 36 Procentele de similaritate şi divergență pentru secvențele haplotipurilor stabilite

	Procent de similaritate				
Procent de divergență		Cy01I	Cy01M	Cy GB	
	Cy01I		100	98,9	Cy01I
	Cy01M	0,0		98,9	Cy01M
	Cy GB	1,2	1,2		Cy GB
		Cy01I	Cy01M	Cy GB	

În plus, pentru a stabili o mai clară diferențiere între toate noile haplotipuri, au fost analizate numărul şi procentele în care se găsesc bazele azotate în fiecare dintre acestea (Tabel 37).

Tabel 37 Procentul bazelor azotate, în cadrul haplotipurilor stabilite

Baze azotate		Specii analizate		
		Cy01I	Cy01M	Cy GB
A	Nr.	63	63	65
	%	24,14	24,14	24,90
G	Nr.	44	44	43
	%	16,86	16,86	16,48
T	Nr.	74	74	75
	%	28,35	28,35	28,74
C	Nr.	80	80	78
	%	30,65	30,65	29,89
A+T	Nr.	137	137	140
	%	52,49	52,49	53,64
C+G	Nr.	124	124	121
	%	47,51	47,51	46,36

Din Tabelul 37 şi Figurile 55 şi 56, după cum a fost precizat şi anterior, se observă că diferenţe majore se constată pentru secvenţa Cy GB, care prezintă valori diferite pentru toate bazele azotate, comparativ cu celelalte haplotipuri.

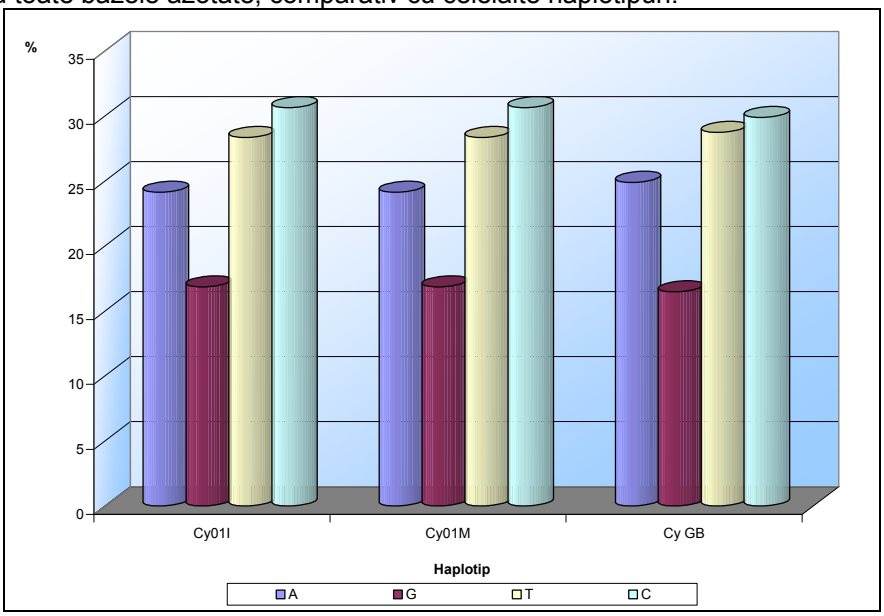


Figura 55 Graficul frecvenţei bazelor azotate în cadrul haplotipurilor caracterizate

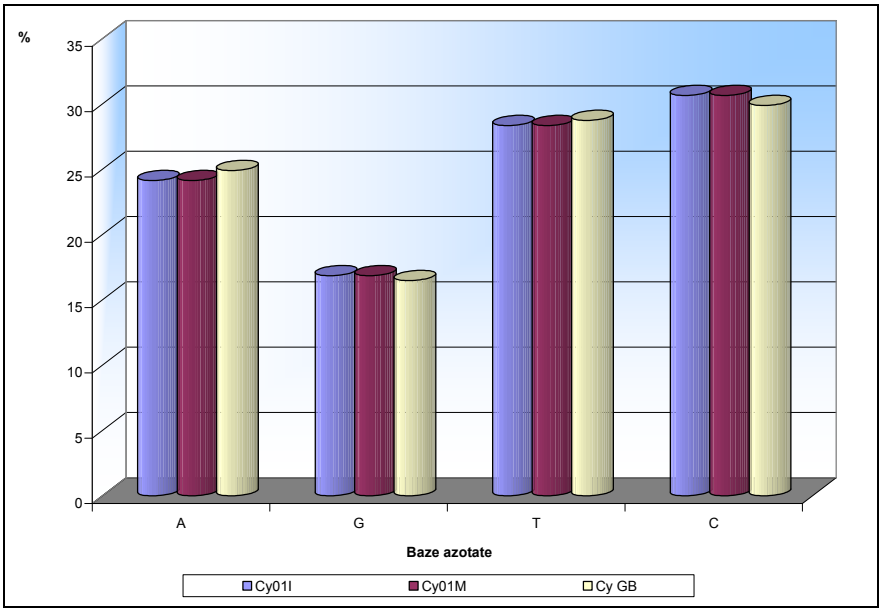


Figura 56 Graficul frecvenţei bazelor azotate, comparativ între haplotipurile analizate

Referitor la cantitatea de adenină (A), aceasta este egală pentru Cy01I şi Cy01M, fiind în proporţie de 24,14%, iar pentru Cy GB, în proporţie de 24,90%. Cantitatea de guanină este cuprinsă între 16,86% (Cy01I, Cy01M) şi 16,48% pentru Cy GB. De asemeni, se constată diferenţe şi pentru celelalte baze azotate: cantitatea de timină

are valori de 28,35% pentru haplotipul general şi 28,74% pentru Cy GB, iar citozina, 30,65% pentru Cy01I (ca reprezentant al haplotipului general caracteristic celor două populații) şi 29,89% pentru haplotipul Cy GB.

Procentul de baze complementare este de 52,49% în cazul A+T pentru Cy01I şi 53,64% pentru Cy GB. Procentul de G+C este de 47,51% (Cy01I, Cy01M) şi de 46,36% pentru Cy GB.

4.7.1. CARACTERIZAREA HAPLOTIPULUI CY01IM

4.7.1.1. FRECVENȚELE TIPURILOR DE ASOCIERI NUCLEOTIDICE

În acest haplotip, au fost încadrate toate secvențele, deoarece procentele de similaritate sunt de 100%.

Din analiza procentelor dinucleotidelor, s-a constatat că valoarea cea mai mare o are perechea CT (10.3%), iar valoarea cea mai mică perechea GT, în proporție de 2,3%.

Referitor la trinucleotide, cea mai mare frecvență o înregistrează tripletul CTA (4,6%), iar valoarea cea mai mică, tripletele ACG, AGA, AGT, ATC, CGT, GAG, GCG, GTA, GTG şi TCG, având o frecvență de 0,4%.

Tetramerii cu frecvența cea mai mare, sunt TACT, CATT (2,3%), iar cei cu frecvența cea mai mică TTTT, TTTA, TTCA, TGCT, GCTA, GCAT, CTTT, CTCA, CTAT în procent de 1,2%. În cazul pentamerilor, cea mai mare frecvență o are pentamerul CTACT (1,5%). În ceea ce priveşte hexamerii, aceştia au frecvențe cuprinse între 1,2% (AGCATT) şi 0,4%.

Tetramerii, pentamerii şi hexamerii cu frecvență joasă, sunt cei care se găsesc într-un singur exemplar în întreaga genă, având o frecvență de 0,4%.

Cu ajutorul modulului EditSeq al DNA Star, a fost calculată temperatura de topire a ADN, după un model Davis-Botstein-Roth; în acest caz, prezintă o valoare de 82,46ºC.

4.7.1.2. TRANSLAȚIA ŞI CARACTERIZAREA CATENEI POLIPEPTIDICE

Translația s-a realizat pe baza codului genetic mitocondrial pentru vertebrate, utilizând modulul EditSeq din cadrul programului DNA Star.

Catena polipeptidică obținută este prezentată în Figura 57.

```
MTPVHFSFSS AFILGLMGLA FHRTHLLSAL LCLEGMMLSL
FIALALWALQ FESTGFSTAP MLLLAFSACE ASTGLALLVA ARTHGT
```

Figura 57 Catena polipeptidică a ND4L, caracteristică haplotipului Cy01IM al speciei *Cyprinus carpio*

După realizarea translației, se constată că întreg lanțul polipeptidic, este alcătuit din 87 aminoacizi. Din Tabelul 38, se observă că frecvența cea mai mare, o are leucina (21,84%), iar cea mai scăzută o au glicina şi triptofanul (1,15%). Se constată de asemenea, că lipsesc acidul aspartic, lizina, asparagina şi tirozina.

Se observă că, din totalul de 87 aminoacizi, 3 au caracter acid, 2 sunt bazici, 20 polari şi 45 hidrofobi (Tabelul 39). De asemenea, se constată că lanțul polipeptidic are

o sarcină electrică de -0,49 (la pH=7) şi un punct izoelectric de 6,73 (Tabelul 39 şi Figura 58).

Tabel 38 Compoziția lanțului polipeptidic

Aminoacizi		Număr	Procent din greutatea totală (%)	Frecvența (%)
Simbol	Nume			
A	Ala	13	9,99	14,94
C	Cys	2	2,23	2,30
D	Asp	0	0,00	0,00
E	Glu	3	4,19	3,45
F	Phe	8	12,73	9,20
G	Gly	6	3,70	6,90
H	His	4	5,93	4,60
I	Ile	2	2,45	2,30
K	Lys	0	0,00	0,00
L	Leu	19	23,25	21,84
M	Met	5	7,09	5,75
N	Asn	0	0,00	0,00
P	Pro	2	2,10	2,30
Q	Gln	1	1,39	1,15
R	Arg	2	3,38	2,30
S	Ser	9	8,48	10,34
T	Thr	8	8,75	9,20
V	Val	2	2,14	2,30
W	Trp	1	2,01	1,15
Y	Tyr	0	0,00	0,00
B	Asx	0	0,00	0,00
Z	Glx	0	0,00	0,00
.	Ter	0	0,00	0,00

Tabel 39 Caracterizarea lanțului polipeptidic

Greutate moleculară (Daltoni)	Număr de aminoacizi	Aminoacizi				Punct izoelectric	Sarcina electrică la pH=7	Concentrația la o extincție =1 şi λ=280nm
		Puternic bazici	Puternic acizi	Hidrofobi	Polari			
9247,07	87	2	3	45	20	6,736	-0,479	1,56

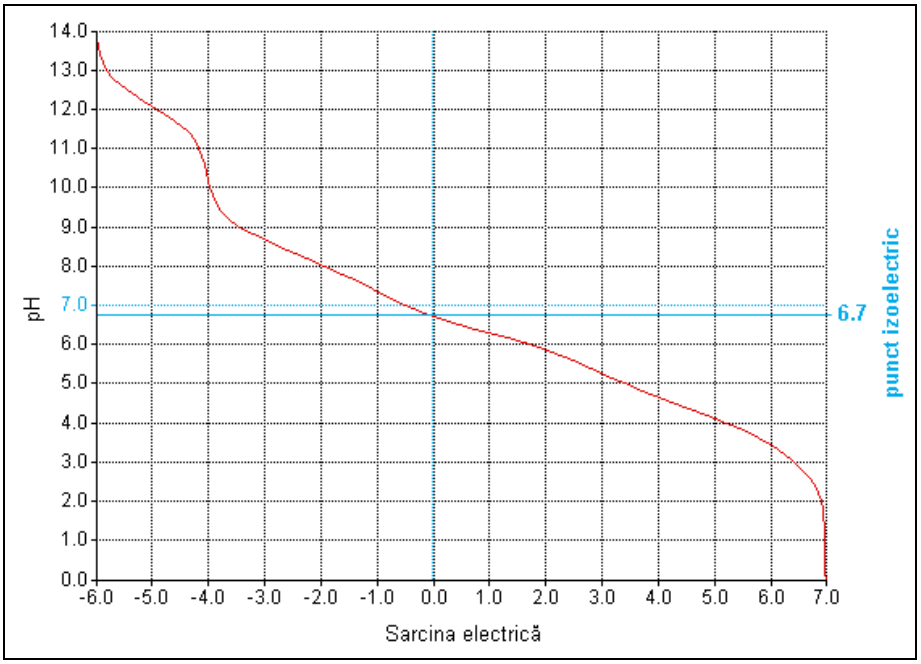


Figura 58 Curba de titrare pentru lanțul polipeptidic al ND4L, haplotip Cy01IM, al speciei *Cyprinus carpio*

Ulterior, a fost realizată o analiză a structurii proteinei astfel, au fost determinate zonele alfa helicoidale, beta pliate şi de inflexiune, prin două metode Garnier-Robson (Garnier et al., 1978) şi Chou-Fasman (Chou şi Fasman, 1978); reprezentarea grafică a zonelor hidrofobe (Kyte şi Doolittle, 1982), zonele de flexie (Karplus şi Schultz, 1985), indexul antigenic (Jameson şi Wolf, 1988) şi graficul probabil al suprafeței (Emini et al., 1985).

Din analiza structurii lanțului polipeptidic, se poate observa că, referitor la regiunile alfa helicoidale, nu există diferențe majore între cele două metode Garnier-Robson şi Chou-Fasman. Astfel, în cazul ambelor metode, se observă că există 3 regiuni alfa helicoidale, existând o suprapunere parțială a suprafețelor comparativ între cele două metode. De asemenea, pentru regiunile beta pliate, se observă că, pentru ambele metode, se constată existența unei singure regiuni beta pliate situate la începutul catenei polipeptidice, cu deosebirea dată de diferențele de mărime dintre suprafețele acestor regiuni determinate prin cele două metode.

Referitor la zonele de inflexiune, pentru prima metodă (Garnier–Robson) sunt date 2 astfel de zone, în timp ce prin metoda Chou-Fasman sunt prezentate ca fiind 3 astfel de regiuni, cu diferența că în cazul celei de-a doua, sunt prezentate ca regiuni cu suprafețe mai mari.

Pe baza graficului hidrofobicității realizat după un model Kyte-Doolitle (Kyte şi Doolittle, 1982), se constată că există 6 zone hidrofobe şi 6 zone hidrofile.

Din graficul indexului antigenic calculat pe baza modelului Jameson–Wolf (Jameson şi Wolf, 1988), se constată existența a 3 regiuni cu potențial antigenic.

4.7.2. RELAȚII FILOGENETICE ÎN CADRUL GENULUI *CYPRINUS*, BAZATE PE DIFERENȚE ALE SECVENȚELOR ND4L

În acest caz, arborii filogenetici au fost realizați, utilizând aceleaşi metode UPGMA pe baza modelului Tamura-Nei (Figura 59) şi Minimum Evolution pe baza unui algoritm Neighbour-Joining (Figura 60 şi 61). Maximum Parsimony, nu poate fi folosită în acest caz, deoarece nu există situri informative pentru această metodă (situri conservative care să intervină cu o frecvență de minim 2).

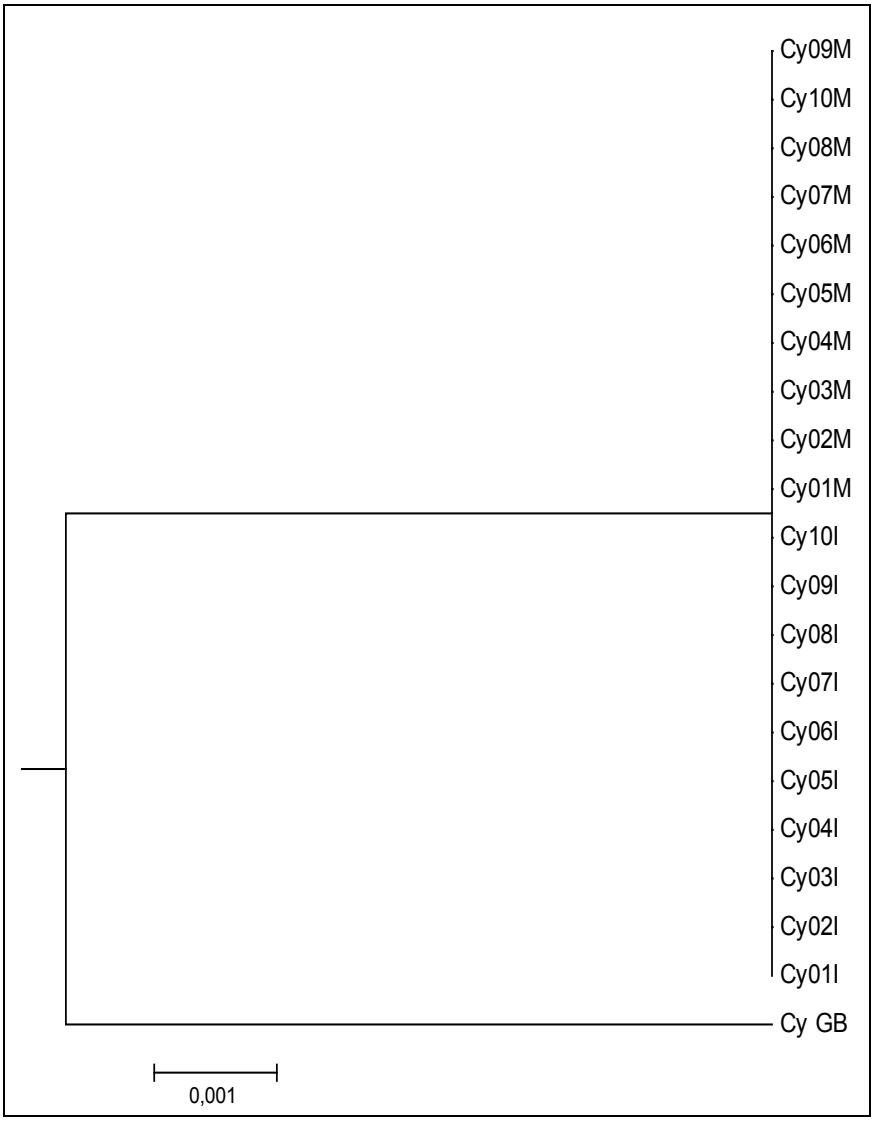


Figura 59 Arbore filogenetic al genului *Cyprinus*, realizat prin metoda UPGMA pe baza secvențelor ND4L

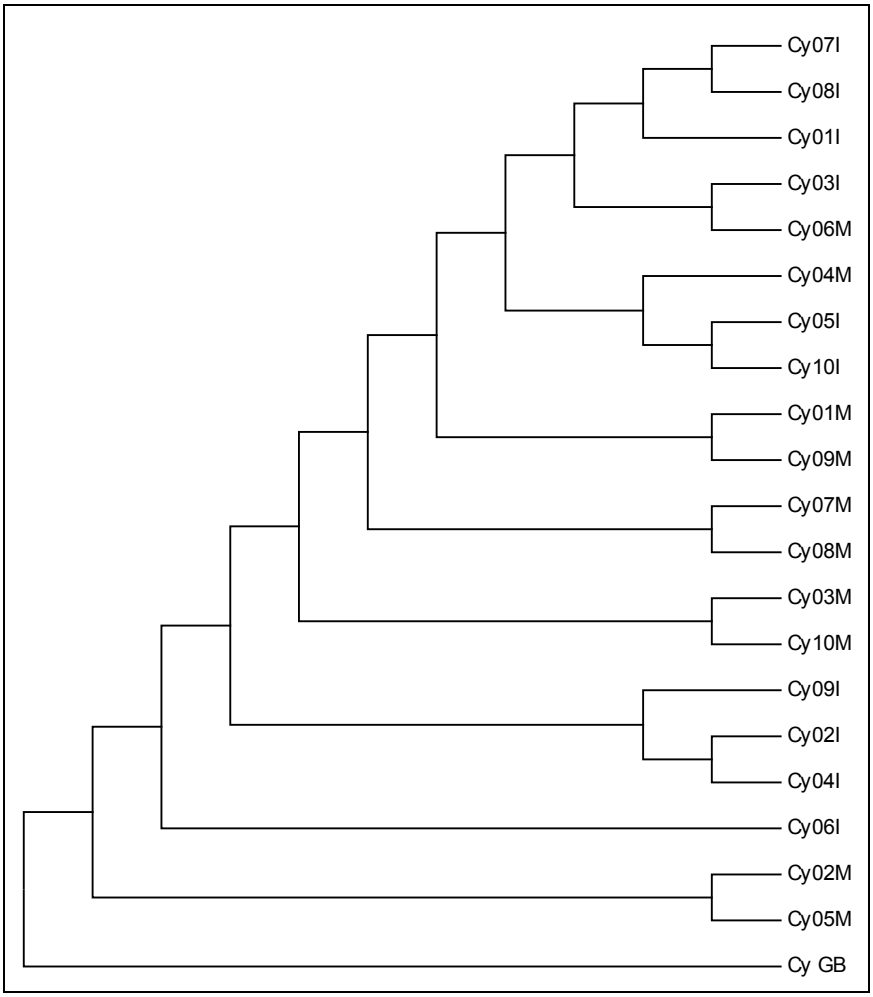


Figura 60 Arbore filogenetic rectangular al genului *Cyprinus*, realizat prin metoda ME pe baza secvențelor ND4L

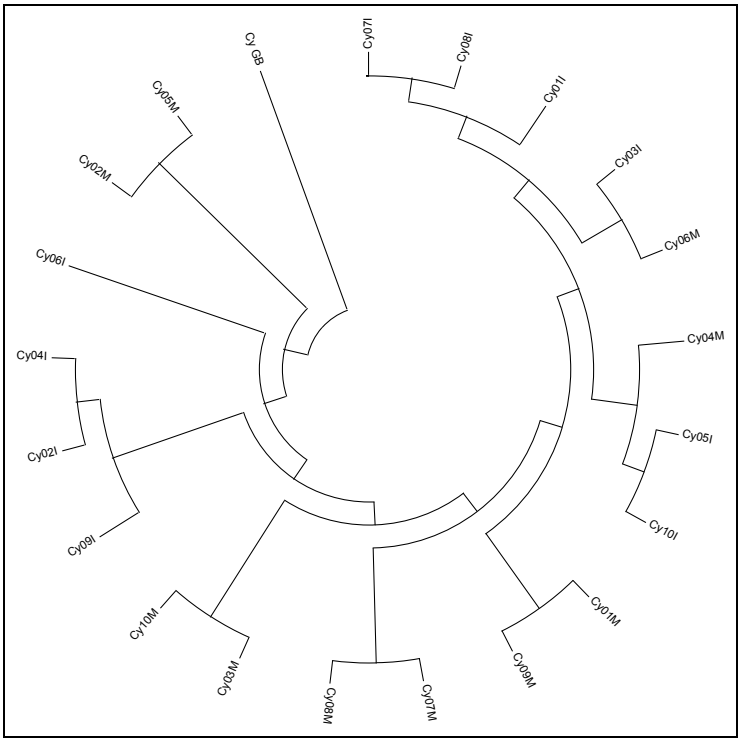


Figura 61 Arbore filogenetic circular al genului *Cyprinus*, realizat prin metoda ME pe baza secvențelor ND4L

Din Figurile 60 şi 61, pe baza diferențelor dintre secvențe şi a sumei lungimilor brațelor, se constată că individul a cărui secvență a fost preluată din banca de gene, provine dintr-un nod comun ancestral, cu ramurile din care au evoluat speciile celor două populații. De asemeni, pentru cele două populații analizate, se observă că indivizii acestora se grupează diferențiat. Se constată din cele două figuri, că există diferențe în cadrul arborelui, date de pozițiile ramurilor interioare, dar care nu prezintă o importanță la fel de mare ca topologia (aspectul general al arborelui) care rămâne neschimbată şi care ne furnizează informațiile necesare pentru direcțiile filogenetice ale haplotipului analizat.

4.8. ANALIZA SECVENȚELOR NUCLEOTIDICE ALE CITOCROMULUI B, PROVENITE DE LA INDIVIZI APARȚINÂND GENULUI *CARASSIUS*

Au fost aliniate 20 de secvențe ale citocromului b prima parte, provenite de la indivizi ai celor două populații luate în studiu (Larga-Jijia – Movileni şi Iaşi), precum şi secvențe preluate din banca de gene GenBank, provenite de la specii (*Carassius carassius* Ref. nr. AY714387 şi *Carassius cuvieri* Ref. nr. AB045144), subspecii (*Carassius auratus langsdörfi* Ref. nr. AB006953) şi hibrizi (*Carassius auratus x Cyprinus carpio* Ref. nr. AY694420 şi *Carassius auratus x Cyprinus carpio x Carassius cuvieri* Ref. nr. AY771781) ai genului *Carassius*.

Alinierea s-a realizat, utilizând metoda Clustal V (Higgins şi Sharp, 1989), (Anexa 10).

Din alinierea secvențelor pentru 23 de indivizi ai genului *Carassius* şi doi hibrizi, se constată că există un număr de 149 diferențe. Dintre acestea, comparând secvențele care provin de la indivizi ai celor două populații analizate, se constată existența a trei diferențe (sau a unei diferențe, dar care implică trei secvențe) şi anume, existența unei tranziții în poziția 258 (substituirea guaninei cu adenină), pentru secvențele Cag01I, Cag07I şi Cag03M. Se poate observa că această mutație este prezentă în ambele populații, dar cu o rată destul de scăzută.

Cele mai multe diferențe (146), apar pentru secvențele preluate din banca de gene. Astfel, se constată existența a 137 de tranziții şi 9 transversii. Secvențele care prezintă cele mai multe diferențe față de haplotipul general al celor două populații, sunt secvențele speciei *Carassius cuvieri* şi a hibridului *Carassius auratus x Cyprinus carpio x Carassius cuvieri*; acestea diferă prin 38 de tranziții şi 6 transversii. Majoritatea tranzițiilor produse au avut loc între citozină şi timină (25), în timp ce pentru adenină – guanină au fost înregistrate doar 15. Referitor la transversii, 2 s-au produs între guanină – citozină, 1 între adenină – timină şi una între adenină – citozină.

În cazul individului *Carassius auratus langsdörfi*, acesta diferă față de haplotipul general al celor două populații, prin 38 de mutații, dintre care 37 de tranziții şi 1 transversie.

Specia *Carassius carassius* şi hibridul *Carassius auratus x Cyprinus carpio*, prezintă câte 11 diferențe fată de haplotipul general, din care 10 tranziții şi o transversie (Gorgan, 2004a).

Din cele prezentate anterior, putem concluziona că au fost găsite 2 noi haplotipuri pentru citocromul b: un haplotip general (care va fi notat CagIMY) caracteristic secvențelor care nu prezintă diferențe în cadrul celor două populații şi un haplotip specific secvențelor care prezintă acea mutație în poziția 258 Cag137IM, de asemeni, caracteristic populațiilor studiate.

Pe baza comparației între secvențe, s-a realizat tabelul similarității şi divergenței (Tabel 40), din care se constată că procentele de similaritate pentru cele două haplotipuri sunt de 99,9%, deoarece diferă printr-o singură nucleotidă. De asemeni, procente foarte ridicate de similaritate (100%) se constată între secvențele *Carassius carassius* şi *Carassius auratus x Cyprinus carpio*, precum şi între secvențele *Carassius cuvieri* şi *Carassius auratus x Cyprinus carpio x Carassius cuvieri*. În primul caz, din cauza similarității dintre secvența speciei *Carassius carassius* şi a hibridului *Carassius auratus x Cyprinus carpio*, care s-a observat că este de 100%, putem deduce că, de fapt, la încrucişarea de obținere a hibridului a participat un individ aparținând speciei *Carassius carassius* şi nu *Carassius auratus* cum s-a precizat. Această afirmație, se bazează pe faptul că ADNmt se transmite doar pe linie maternă, iar în acest caz, dacă femela ar fi aparținut speciei *Cyprinus carpio*, cu siguranță, procentul de similaritate cu celelalte secvențe ale genului *Carassius*, ar fi fost mai scăzut. Deci, femela a aparținut unei specii a genului *Carassius*, iar masculul, speciei *Cyprinus carpio*. Din compararea secvențelor, constatăm că secvența hibridului corespunde în procent de 100% cu a unui individ aparținând speciei *Carassius carassius*, deci, femela care a participat la încrucişare, aparținea acestei specii. Această eroare rezidă probabil din faptul că femela care a fost introdusă în hibridare a fost determinată doar pe baza caracterelor morfologice sau cele două specii sunt de fapt subspecii ale aceleiaşi specii.

În al doilea caz, în situația similarității dintre secvența speciei *Carassius cuvieri* şi cea a hibridului *Carassius auratus x Cyprinus carpio x Carassius cuvieri*, ținând cont de faptul că ADNmt se transmite pe linie maternă, cea mai probabilă situație în acest

caz, este cea în care, hibridul obținut în urma încrucişării *Carassius auratus x Cyprinus carpio* este încrucişat cu o femelă aparținând speciei *Carassius cuvieri*, caz în care, la nivel mitocondrial, va fi transmis doar ADN-ul provenit de la genitorul matern.

Ținând cont de faptul că la *Cyprinidae* în general şi la reprezentanții genului *Carassius* în special, este întâlnit fenomenul de ginogeneză şi în aceste cazuri, este posibil să fie doar falşi hibrizi, adică indivizi ginogenetici. În acest caz, specia *Cyprinus carpio*, nu a contribuit cu informația sa genetică la formarea descendenților, ci, materialul seminal, a avut rol, doar de inițiere a diviziunii celulare.

Aceste două exemple, demonstrează încă o dată importanța studiilor de genetică şi filogenie moleculară în determinarea hibrizilor, a genitorilor acestora precum şi a sexului genitorilor.

Tabel 40 Procentele de similaritate şi divergență pentru secvențele haplotipurilor stabilite

	Procent de similaritate								
Procent de divergență		CagIMY	Cag137IM	Cacuv	Caulan	CaxCy	CaxCyx Ccuv	Ccaras	
	CagIMY		99,9	93,6	94,3	98,2	93,6	98,2	CagIMY
	Cag137IM	0,1		93,7	94,5	98,4	93,7	98,4	Cag137IM
	Cacuv	6,8	6,7		93,7	93,9	100,0	93,9	Cacuv
	Caulan	6,0	5,9	6,7		94,6	93,7	94,6	Caulan
	CaxCy	1,8	1,6	6,5	5,7		93,9	100,0	CaxCy
	CaxCyx Ccuv	6,8	6,7	0,0	6,7	6,5		93,9	CaxCyx Ccuv
	Ccaras	1,8	1,6	6,5	5,7	0,0	6,5		Ccaras
		CagIMY	Cag137IM	Cacuv	Caulan	CaxCy	CaxCyx Ccuv	Ccaras	

În plus, pentru a stabili o mai clară diferențiere între toate noile haplotipuri, au fost analizate numărul şi procentele în care se găsesc bazele azotate în fiecare dintre acestea (Tabel 41).

Tabel 41 Procentul bazelor azotate, în cadrul haplotipurilor stabilite

Baze azotate		Specii analizate						
		CagIMY	Cag137IM	Ccaras	Ccuv	Caulan	CaxCy	CaxCyxCcuv
A	Nr,	198	199	202	194	198	202	194
	%	28,95	29,09	29,53	28,36	28,95	29,53	28,36
G	Nr,	100	99	97	101	101	97	101
	%	14,62	14,47	14,18	14,77	14,77	14,18	14,77
T	Nr,	203	203	200	205	198	200	205
	%	29,68	29,68	29,24	29,97	28,95	29,24	29,97
C	Nr,	183	183	185	184	187	185	184
	%	26,75	26,75	27,05	26,90	27,34	27,05	26,90
A+T	Nr,	401	402	402	399	396	402	399
	%	58,63	58,77	58,77	58,33	57,89	58,77	58,33
C+G	Nr,	283	282	282	285	288	282	285
	%	41,37	41,23	41,23	41,67	42,11	41,23	41,67

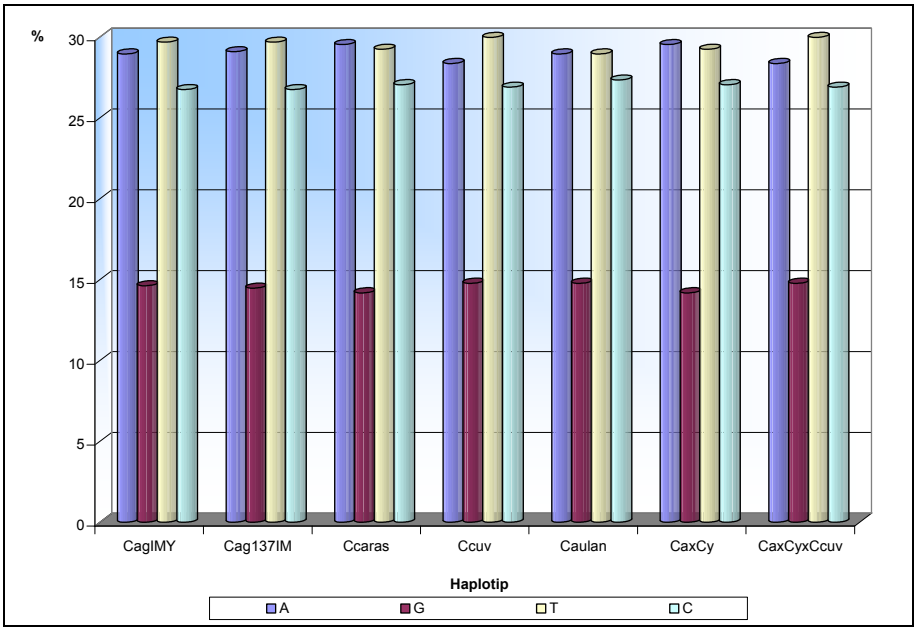


Figura 62 Graficul frecvenței bazelor azotate în cadrul haplotipurilor caracterizate

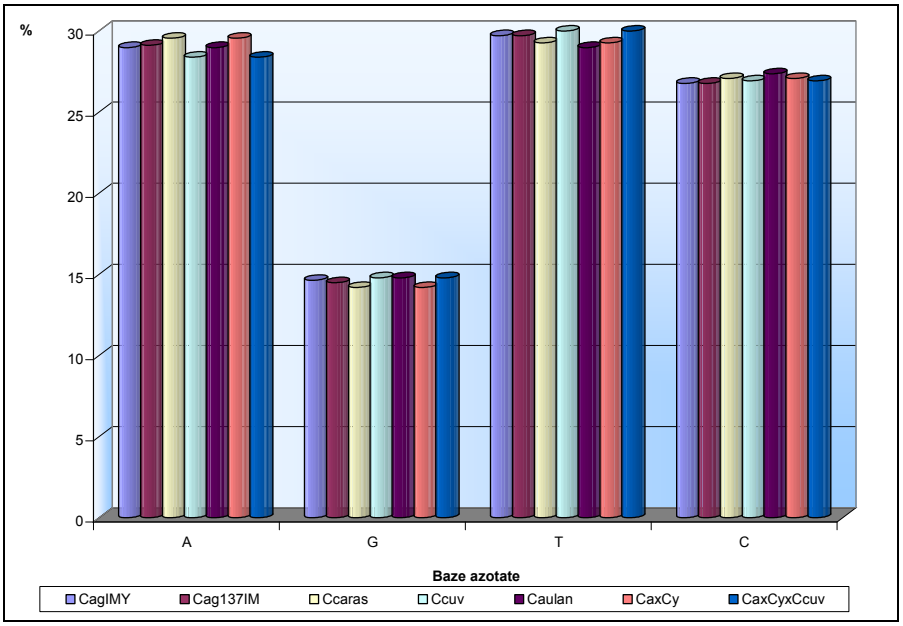


Figura 63 Graficul frecvenței bazelor azotate, comparativ între haplotipurile analizate

Din Tabelul 41 şi Figurile 62 – 63, constatăm că pentru haplotipurile descoperite în cele două populații analizate, concentrația de adenină este de 28,95% pentru CagIMY şi 29,09% pentru Cag137IM. Cantitatea de guanină, variază între 14,62% în cazul primului haplotip şi 14,47% pentru cel de-al doilea, timina are valori de 29,68% pentru ambele haplotipuri, ca şi citozina, cu o valoare de 26,75%.

Procentul bazelor complementare este de 58,63% pentru A+T în cazul primului haplotip şi de 58,77% pentru cel de-al doilea, iar procentele pentru C+G sunt de 41,37% pentru CagIMY şi 41,23% pentru Cag137IM.

Dacă urmărim valorile pentru întreg tabelul, constatăm că există o variație a cantității de adenină, cuprinsă între 28,36%pentru haplotipurile Ccuv, CyxCyxCcuv şi 29,53% pentru haplotipurile Ccaras şi CaxCy. Cantitatea de guanină variază asemănător, având valori cuprinse între 14,18% pentru CaxCy şi 14,77% pentru Ccuv, Caulan şi CaxCyxCcuv. Referitor la cantitatea de timină, se constată că variază între 28,95% pentru Caulan şi 29,97% pentru Ccuv şi hibridul CaxCyxCcuv. Cantitatea de citozină, variază între 26,75% pentru primele două haplotipuri şi 27,34% pentru Caulan.

Referitor la procentele în care se găsesc bazele complementare, pentru A+T se înregistrează valori cuprinse între 58,77% pentru haplotipurile Cag137IM, Ccaras şi CaxCy şi 57,89% pentru Caulan, iar pentru C+G, valori cuprinse între 41,23% (Cag137IM, Ccaras şi CaxCy) şi 42,11% pentru secvența Caulan.

4.8.1. CARACTERIZAREA HAPLOTIPULUI CAGIMY

4.8.1.1. FRECVENȚELE TIPURILOR DE ASOCIERI NUCLEOTIDICE

În acest haplotip au fost încadrate toate secvențele între care nu există diferențe (Cag02I, Cag03I, Cag04I, Cag05I, Cag06I, Cag08I, Cag09I, Cag10I, Cag01M,

Cag02M, Cag04M, Cag05M, Cag06M, Cag07M, Cag08M, Cag09M, Cag10M), aparţinând celor două populații luate în studiu, Larga-Jijia – Movileni şi Iaşi.

Pentru a caracteriza acest haplotip, au fost calculate frecvențele dinucleotidelor (Tabelul 84), trinucleotidelor (Tabelul 85), tetramerilor, pentamerilor şi hexamerilor (Tabelul 86), precum şi tetramerii, pentamerii şi hexamerii cu frecvență joasă (Tabelul 87).

Din analiza procentelor dinucleotidelor constatăm că valoarea cea mai mare o are perechea AT (9,8%), iar valoarea cea mai mică perechea GG, în proporție de 3,2%. Referitor la trinucleotide, cea mai mare frecvență o înregistrează tripletul CAT (3,8%), iar valoarea cea mai mică, tripletele GCG, GGT şi GTG, având o frecvență de 0,1%.

Tetramerul cu frecvența cea mai mare, este CATT (1,9%), iar cei cu frecvență redusă: TTAT, TAAT, CTTC, CTCA ŞI CCAT toți în procent de 0,9%. În cazul pentamerilor (Tabelul 86), cea mai mare frecvență o are pentamerul ACATT (0,9%), iar cea mai mică frecvență, o au TTTCT, TTCTA, TTCAT şi TTCAC (0,4%). În ceea ce priveşte hexamerii, aceştia au frecvențe cuprinse între 0,3% şi 0,4%.

Tetramerii, pentamerii şi hexamerii cu frecvență joasă, sunt cei care se găsesc într-un singur exemplar în întreaga genă, având o frecvență de 0,1%.

Cu ajutorul modulului EditSeq al DNA Star, a fost calculată temperatura de topire a ADN, după un model Davis-Botstein-Roth; în acest caz, prezintă o valoare de 81,13ºC.

4.8.1.2. TRANSLAȚIA ŞI CARACTERIZAREA CATENEI POLIPEPTIDICE

Translația s-a realizat pe baza unui cod genetic mitocondrial pentru vertebrate, utilizând modulul EditSeq din cadrul programului DNA Star.

Catena polipeptidică obținută este prezentată în Figura 64.

```
MASLRKTHPL IKIANDALVD LPTPSNISAW WNFGSLLGLC LITQILTGLF LAMHYTSDIS
TAFSSVTHIC RDVNYGWLIR NIHANGASFF FICIYMHIAR GLYYGSYLYK ETWNIGVVLL
LLVMMTAFVG YVLPWGQMSF WGATVITNLL SAVPYMGDML VQWIWGGFSV DNATLTRFFA
                       FHFLLPFIIA AATVIHLLFL
```

Figura 64 Catena polipeptidică a citocromului b caracteristică haplotipului CagIMY al speciei *Carassius gibelio* Bloch.

După realizarea translației, se constată că întreg lanțul polipeptidic, este alcătuit din 228 aminoacizi (Tabelul 42), din care se observă că frecvența cea mai mare, o are leucina (15,08%), iar cea mai scăzută o are cisteina (1,06%).

Se constată că din totalul de 228 aminoacizi, 11 sunt acizi, 10 sunt bazici, 61 polari şi 105 hidrofobi (Tabelul 43). De asemenea, se constată că lanțul polipeptidic are o sarcină electrică de 0,31 (la pH=7) şi un punct izoelectric de 7,09 (Figura 65).

Tabel 42 Compoziția lanțului polipeptidic

Aminoacizi		Număr	Procent din greutatea totală (%)	Frecvența (%)
Simbol	Nume			
A	Ala	18	4,98	7,89
C	Cys	3	1,21	1,32
D	Asp	9	4,04	3,95
E	Glu	2	1,01	0,88
F	Phe	17	9,75	7,46
G	Gly	16	3,56	7,02
H	His	9	4,81	3,95
I	Ile	20	8,82	8,77
K	Lys	5	2,50	2,19
L	Leu	29	12,79	12,72
M	Met	8	4,09	3,51
N	Asn	12	5,33	5,26
P	Pro	8	3,03	3,51
Q	Gln	3	1,50	1,32
R	Arg	5	3,04	2,19
S	Ser	17	5,77	7,46
T	Thr	13	5,91	6,58
V	Val	15	5,02	5,70
W	Trp	8	5,80	3,51
Y	Tyr	11	6,99	4,82
B	Asx	0	0,00	0,00
Z	Glx	0	0,00	0,00
.	Ter	0	0,00	0,00

Tabel 43 Caracterizarea lanțului polipeptidic

Greutate moleculară (Daltoni)	Număr de aminoacizi	Aminoacizi				Punct izoelectric	Sarcina electrică la pH=7	Concentrația la o extincție =1 şi λ=280nm
		Puternic bazici	Puternic acizi	Hidrofobi	Polari			
25666,07	228	10	11	105	61	7,098	0,311	0,43

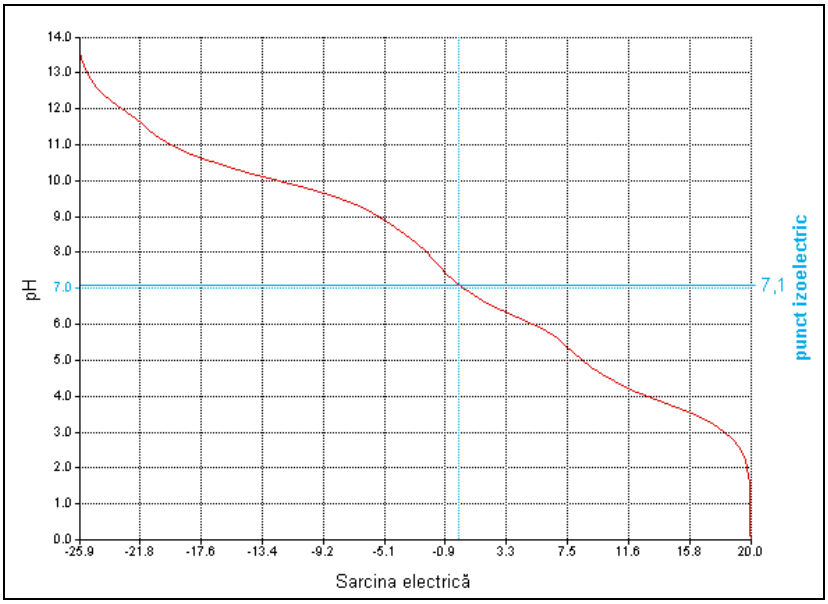


Figura 65 Curba de titrare pentru lanțul polipeptidic al citocromului b, haplotip CagIMY al speciei *Carassius gibelio*

Ulterior, a fost realizată o analiză a structurii proteinei (Figura 98); astfel, au fost determinate zonele alfa helicoidale, beta pliate şi de inflexiune, prin două metode Garnier-Robson (Garnier et al., 1978) şi Chou-Fasman (Chou şi Fasman, 1978); reprezentarea grafică a zonelor hidrofobe (Kyte şi Doolittle, 1982), zonele de flexie (Karplus şi Schultz, 1985), indexul antigenic (Jameson şi Wolf, 1988) şi graficul probabil al suprafeței (Emini et al., 1985).

Din analiza structurii lanțului polipeptidic, se poate observa că referitor la regiunile alfa helicoidale, există diferențe între cele două metode Garnier-Robson şi Chou-Fasman. Astfel, în cazul primei metode, se observă că există 9 regiuni alfa helicoidale, în timp ce, potrivit celei de-a doua metode, există doar 4, cu 3 puncte de suprapunere ale suprafețelor. De asemenea, pentru regiunile beta pliate, se observă că există o similaritate relativă între cele două metode, comparativ cu regiunile alfa, în ceea ce priveşte suprafața acestor zone. Pentru determinarea efectuată prin prima metodă, se constată existența a 18 regiuni beta pliate, în timp ce, pentru cea de-a doua metodă (Chou-Fasman), sunt considerate ca existând 7 astfel de regiuni. Referitor la zonele de inflexiune, de asemenea, comparativ pentru cele două metode, se constată că pentru prima metodă (Garnier–Robson) sunt prezentate 14 regiuni, iar prin a doua metodă (Chou-Fasman) sunt date 12 astfel de regiuni, cu precizarea că în cazul celei de-a doua, sunt prezentate ca regiuni cu suprafețe mai mari.

Pe baza graficului hidrofobicității, realizat după un model Kyte-Doolitle (Kyte şi Doolittle, 1982), se constată că există 9 zone hidrofobe şi 7 zone hidrofile.

Din graficul indexului antigenic calculat pe baza modelului Jameson –Wolf (Jameson şi Wolf, 1988), se constată existența a 9 regiuni cu potențial antigenic.

4.8.2. CARACTERIZAREA HAPLOTIPULUI CAG137IM

4.8.2.1. FRECVENȚELE TIPURILOR DE ASOCIERI NUCLEOTIDICE

Acest haplotip este caracteristic secvențelor Cag01I, Cag07I şi Cag03S, aparținând celor două populații luate în studiu, Larga-Jijia – Movileni şi Iaşi. Acestea diferă de haplotipul general prin intervenția unei tranziții în poziția 258 (substituția guaninei cu adenină).

Din analiza procentelor dinucleotidelor, constatăm că valoarea cea mai mare o are perechea AT (9,8%), iar valoarea cea mai mică perechea GT, în proporție de 2,6%. Referitor la trinucleotide, cea mai mare frecvență o înregistrează tripletul CAT (3,8%), iar valoarea cea mai mică, tripletele GCG, GGT şi GTG, având o frecvență de 0,1%.

Tetramerul cu frecvența cea mai mare, este CATT (1,9%), iar cei cu frecvență redusă: TTAT, TAAT, CTTC, CTCA ŞI CCAT, toți în procent de 0,9%. În cazul pentamerilor, cea mai mare frecvență o are pentamerul ACATT (0,9%), iar cea mai mică frecvență, o au TTTCT, TTCAT, TTCTA şi TTCAC (0,4%). În ceea ce priveşte hexamerii, aceştia au frecvențe cuprinse între 0,3% şi 0,4%.

Tetramerii, pentamerii şi hexamerii cu frecvență joasă, sunt cei care se găsesc într-un singur exemplar în întreaga genă, având o frecvență de 0,1%.

Cu ajutorul modulului EditSeq al DNA Star, a fost calculată temperatura de topire a ADN, după un model Davis-Botstein-Roth; în acest caz, prezintă o valoare de 81,07ºC.

4.8.2.2. MODELAREA STRUCTURII SECUNDARE A ARNm

Deoarece există o singură diferență, constând în înlocuirea unei nucleotide, structura generală a ARNm este similară structurii ARNm al haplotipului general pentru această genă. Singura diferență în acest caz, față de structura anterioară, este dată de existența adeninei în locul guaninei. Ca şi în cazul precedent, ARNm prezintă 47 regiuni monocatenare, alternând cu 24 regiuni bicatenare.

4.8.2.3. TRANSLAȚIA ŞI CARACTERIZAREA CATENEI POLIPEPTIDICE

Translația s-a realizat pe baza unui cod genetic mitocondrial pentru vertebrate, utilizând modulul EditSeq din cadrul programului DNA Star.

Catena polipeptidică obținută este prezentată în Figura 66.

```
MASLRKTHPL IKIANDALVD LPTPSNISAW WNFGSLLGLC LITQILTGLF
LAMHYTSDIS TAFSSVTHIC RDVNYGWLIR NIHANGASFF FICIYMHIAR
GLYYGSYLYK ETWNIGVVLL LLVMMTAFVG YVLPWGQMSF WGATVITNLL
SAVPYMGDML VQWIWGGFSV DNATLTRFFA FHFLLPFIIA AATVIHLLFL
            HETGSNNPIG LNSDADKISF HPYFSYKD
```

Figura 66 Catena polipeptidică a citocromului b caracteristică haplotipului Cag137IM al speciei *Carassius gibelio Bloch.*

După realizarea translației, se constată că întreg lanțul polipeptidic, este alcătuit din 228 aminoacizi (Tabelul 44), din care se observă că frecvența cea mai mare, o are

leucina (15,08%), iar cea mai scăzută o are cisteina (1,06%). Se constată de asemenea, existența a doi codoni stop (notați cu „.”) către sfârşitul lanțului polipeptidic.

Se constată că din totalul de 228 aminoacizi, 11 sunt acizi, 10 sunt bazici, 61 polari şi 105 hidrofobi. De asemenea, se constată că lanțul polipeptidic are o sarcină electrică de 0,31 (la pH=7) şi un punct izoelectric de 7,09.

Tabel 44 Compoziția lanțului polipeptidic

Aminoacizi		Număr	Procent din greutatea totală (%)	Frecvența (%)
Simbol	Nume			
A	Ala	18	4,98	7,89
C	Cys	3	1,21	1,32
D	Asp	9	4,04	3,95
E	Glu	2	1,01	0,88
F	Phe	17	9,75	7,46
G	Gly	16	3,56	7,02
H	His	9	4,81	3,95
I	Ile	20	8,82	8,77
K	Lys	5	2,50	2,19
L	Leu	29	12,79	12,72
M	Met	8	4,09	3,51
N	Asn	12	5,33	5,26
P	Pro	8	3,03	3,51
Q	Gln	3	1,50	1,32
R	Arg	5	3,04	2,19
S	Ser	17	5,77	7,46
T	Thr	13	5,91	6,58
V	Val	15	5,02	5,70
W	Trp	8	5,80	3,51
Y	Tyr	11	6,99	4,82
B	Asx	0	0,00	0,00
Z	Glx	0	0,00	0,00
.	Ter	0	0,00	0,00

4.8.3. RELAȚII FILOGENETICE ÎN CADRUL GENULUI *CARASSIUS*, STABILITE PRIN ANALIZA SECVENȚELOR CITOCROMULUI B

Metodele folosite în filogenie şi modalitățile de realizare a arborilor filogenetici, utilizate în acest caz, sunt cele care includ tranzițiile şi transversiile, deoarece haplotipurile analizate pentru cele două populații, prezintă ambele tipuri de mutații.

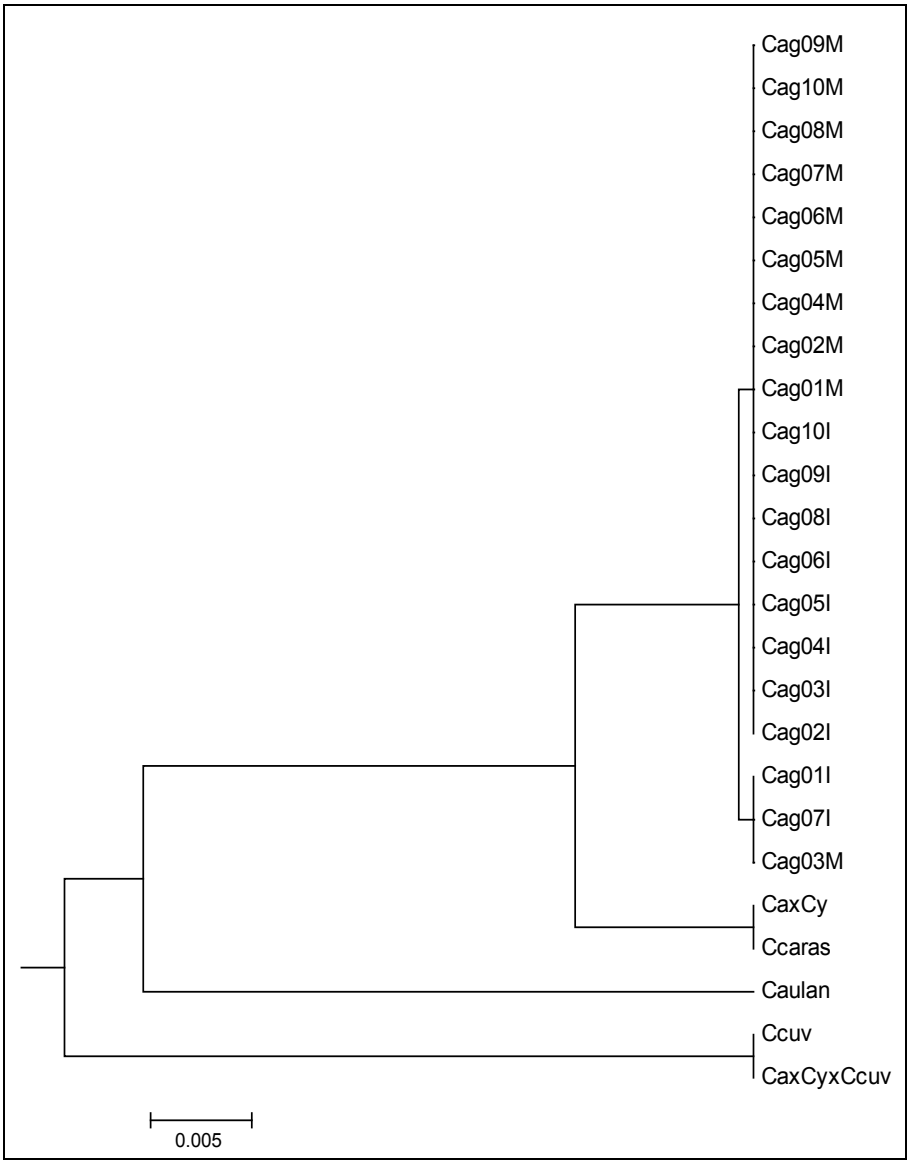


Figura 67 Arbore filogenetic rectangular al genului *Carassius*, realizat prin metoda UPGMA pe baza secvențelor citocromului B

Astfel, prin metoda UPGMA bazată pe modelul Tamura – Nei, care include ambele tipuri de mutații posibile (tranziții şi transversii), a fost obținut un arbore filogenetic (Figura 67), din care se observă că Cy06 derivă direct dintr-un nod comun ancestral, din care derivă şi brațul comun al celorlalte secvențe.

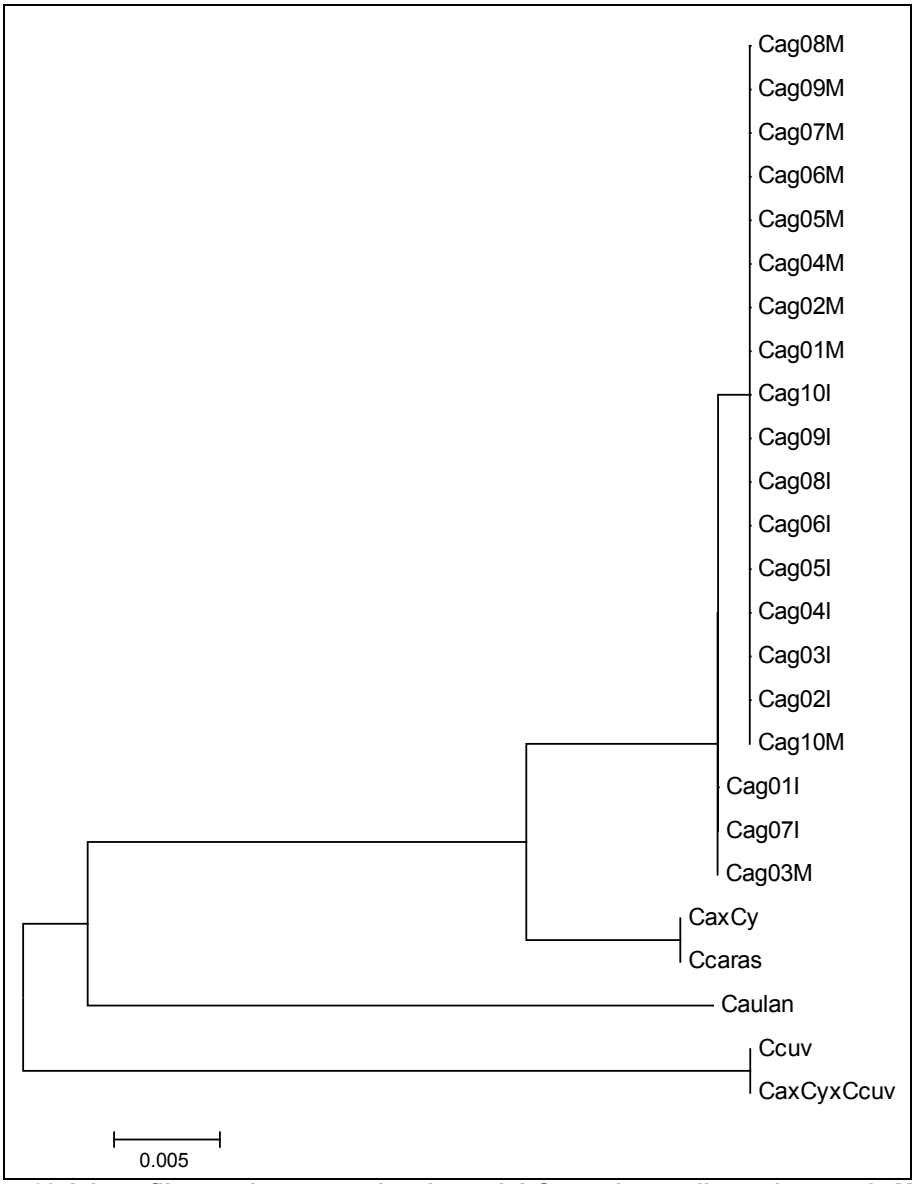


Figura 68 Arbore filogenetic rectangular al genului *Carassius*, realizat prin metoda ME pe baza secvențelor citocromului B

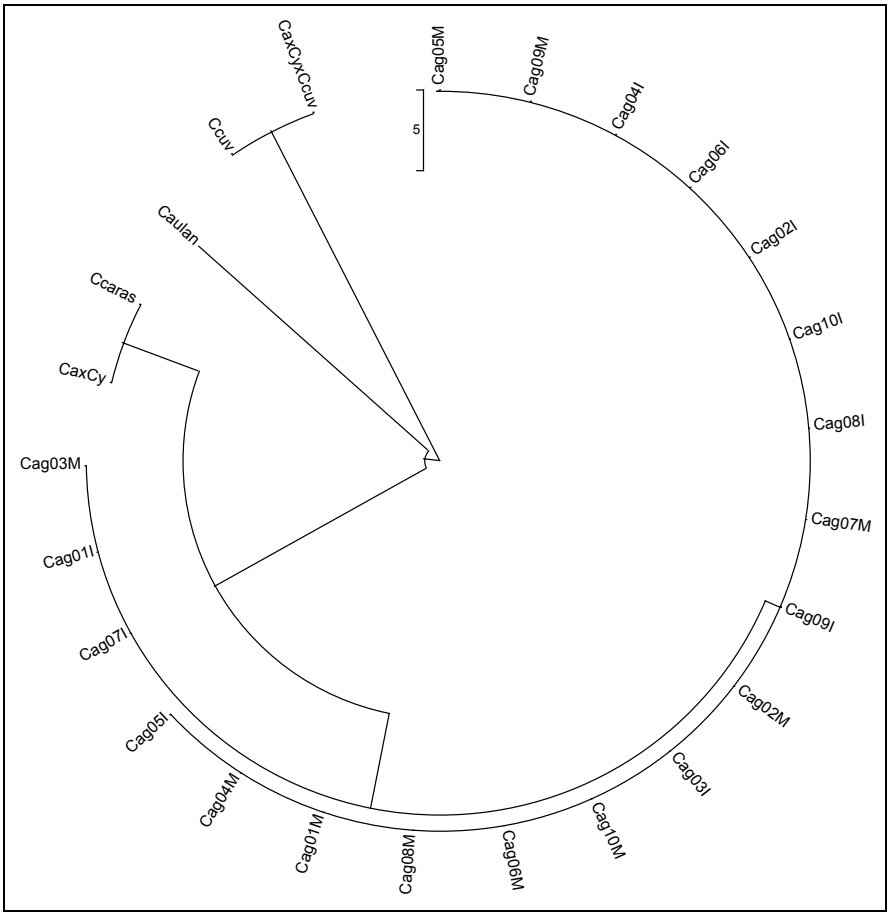


Figura 69 Arbore filogenetic circular al genului *Carassius*, realizat prin metoda MP pe baza secvențelor citocromului B

Din Figura 68 se constată grupele evolutive ale speciilor analizate, în care se încadrează secvențele pe baza similarităților existente, aşa cum au fost încadrate şi în cazul analizei procentelor de similaritate.

Arborele filogenetic din Figura 68, realizat prin metoda evoluției minime (ME), utilizând un algoritm Neighbour-Joining, ne furnizează informații referitoare la distanțele existente între grupele stabilite anterior. Astfel, constatăm că pe ultima treaptă a evoluției se situează haplotipul general stabilit pentru cele două populații, urmat de haplotipul caracteristic celorlalte trei secvențe. Acesta din urmă se situează pe acelaşi palier evolutiv cu specia *Carassius auratus langsdörfi*, specie care însă, a urmat alt traseu evolutiv, până în acest punct. De asemeni, se poate observa că specia cea mai apropiată de speciile comune ancestrale, din cadrul acestui grup evolutiv, este *Carassius carassius* (caracuda), care se găsea şi în apele țării noastre ca specie dominantă, dar care astăzi, a fost înlocuită de specia *Carassius gibelio*, persistând însă doar în populații izolate, pe areale foarte restrânse, limitate la doar câteva ecosisteme acvatice. Faptul că a fost înlocuită , dovedeşte o adaptare mai bună a speciei *Carassius gibelio* la condițiile actuale, fapt ce o situează pe o treaptă

evolutivă superioară speciei *Carassius carassius*, fapt confirmat şi de prezenta analiză.

Figura 69 prezintă un arbore circular realizat prin metoda Maximum Parsimony (MP) care ne confirmă rezultatele discutate anterior, cu precizarea că specia *Carassius cuvieri* a evoluat monofiletic dintr-un nod filogenetic comun ancestral, ceea ce ne determină să concluzionăm că este o specie ancestrală care stă la originea celorlalte specii analizate. Bineînțeles, dacă s-ar mări numărul de taxoni analizați, este posibil să se constate o altă specie ancestrală, iar *Carassius cuvieri* să fie doar un rezultat al evoluției, dar, în acest caz, cele trei metode diferite utilizate (UPGMA; Minimum Evolution şi Maximum Parsimony) la care se adaugă şi modelul utilizat (Tamura-Nei), ne confirmă că este specie ancestrală pentru celelalte specii analizate.

4.9. ANALIZA SECVENȚELOR NUCLEOTIDICE ALE D-LOOP, PROVENITE DE LA INDIVIZI APARȚINÂND GENULUI *CARASSIUS*

Pentru regiunea de control mitocondrial, în cazul speciilor genului *Carassius*, a fost secvențiată prima parte hipervariabilă a acesteia. Au fost aliniate 20 de secvențe provenite de la indivizi ai celor două populații luate în studiu (Larga-Jijia – Movileni şi Iaşi), precum şi secvențe preluate din banca de gene GenBank, provenite de la specii (*Carassius carassius* Ref. nr. AY714387 şi *Carassius cuvieri* Ref. nr. AB045144), subspecii (*Carassius auratus langsdörfi* Ref. nr. AB006953) şi hibrizi (*Carassius auratus x Cyprinus carpio* Ref. nr. AY694420 şi *Carassius auratus x Cyprinus carpio x Carassius cuvieri* Ref. nr. AY771781) ai genului *Carassius*.

Alinierea s-a realizat, utilizând metoda Clustal V (Higgins şi Sharp, 1989) şi a fost verificată prin metoda Clustal W (Thompson et al., 1994), obținându-se în ambele cazuri acelaşi rezultat.

Din alinierea secvențelor, se observă că există un număr de 82 diferențe între cele 25 de secvențe, dintre care 64 tranziții, 12 transversii, 3 inversii şi 3 deleții.

Între secvențele celor două populații analizate, se constată existența a două diferențe (sau o diferență care implică două secvențe, comparativ cu celelalte secvențe similare ale aceloraşi populații) ce constau în existența unei tranziții pentru fiecare dintre cele două secvențe în poziția 133, unde a avut loc substituția timinei de către citozină.

Referitor la celelalte secvențe preluate din banca de gene, se constată că cele mai multe diferențe (21) față de haplotipul general al celor două populații analizate, îl prezintă secvența speciei *Carassius cuvieri*, care are acelaşi număr de diferențe cu secvența hibridului *Carassius auratus x Cyprinus carpio x Carassius cuvieri*. Pentru secvențele speciei *Carasius carassius* şi ale hibridului *Carassius auratus* x *Cyprinus carpio*, de asemeni, se înregistrează un număr egal de mutații (11), dintre care 10 tranziții şi o deleție.

Pe baza comparației între secvențe, s-a realizat tabelul similarității şi divergenței (Tabel 45), din care se constată că procentele de similaritate pentru cele două haplotipuri ale populațiilor analizate sunt de 99,6%. De asemeni, procente foarte ridicate de similaritate (100%) se constată între secvențele *Carassius carassius* şi *Carassius auratus x Cyprinus carpio*, precum şi între secvențele *Carassius cuvieri* – *Carassius auratus x Cyprinus carpio x Carassius cuvieri*.

Tabel 45 Procentele de similaritate şi divergenţă pentru secvenţele haplotipurilor stabilite

	Procent de similaritate								
Procent de divergenţă		CagIMD	Cag25ID	Cacuv	Caulan	CaxCy	CaxCyx Ccuv	Ccaras	
	CagIMD		99,6	91,1	74,3	73,1	91,1	73,1	CagIMD
	Cag25ID	0,4		90,7	73,9	72,7	90,7	72,7	Cag137IM
	Cacuv	9,7	10,2		70,2	69,0	100,0	69,0	Cacuv
	Caulan	31,5	32,1	38,0		94,7	70,2	94,7	Caulan
	CaxCy	33,4	34,0	40,0	5,6		69,0	100,0	CaxCy
	CaxCyx Ccuv	9,7	10,2	0,0	38,0	40,0		69,0	CaxCyx Ccuv
	Ccaras	33,4	34,0	40,0	5,6	0,0	40,0		Ccaras
		CagIMD	Cag25ID	Cacuv	Caulan	CaxCy	CaxCyx Ccuv	Ccaras	

Pe baza comparaţiei între secvenţe, precum şi a procentelor de similaritate, constatăm că au fost găsite 2 noi haplotipuri pentru specia *Carassius gibelio* Bloch şi cele două populaţii analizate – un haplotip general CagIMD şi unul caracteristic doar populaţiei Iaşi Cag25ID.

În plus, pentru a stabili o mai clară diferenţiere între toate noile haplotipuri, au fost analizate numărul şi procentele în care se găsesc bazele azotate în fiecare dintre acestea (Tabel 46).

Tabel 46 Procentul bazelor azotate, în cadrul haplotipurilor stabilite

Baze azotate		Specii analizate						
		CagIMD	Cag25ID	Ccaras	Ccuv	Caulan	CaxCy	CaxCyxCcuv
A	Nr,	91	91	91	94	91	91	94
	%	36,99	36,99	37,45	38,21	37,45	37,45	38,21
G	Nr,	38	38	35	38	33	35	38
	%	15,45	15,45	14,40	15,45	13,58	14,40	15,45
T	Nr,	77	76	74	73	74	74	73
	%	31,30	30,89	30,45	29,67	30,45	30,45	29,67
C	Nr,	40	41	43	41	45	43	41
	%	16,26	16,67	17,70	16,67	18,52	17,70	16,67
A+T	Nr,	168	167	165	167	165	165	167
	%	68,29	67,89	67,90	67,89	67,90	67,90	67,89
C+G	Nr,	78	79	78	79	78	78	79
	%	31,71	32,11	32,10	32,11	32,10	32,10	32,11

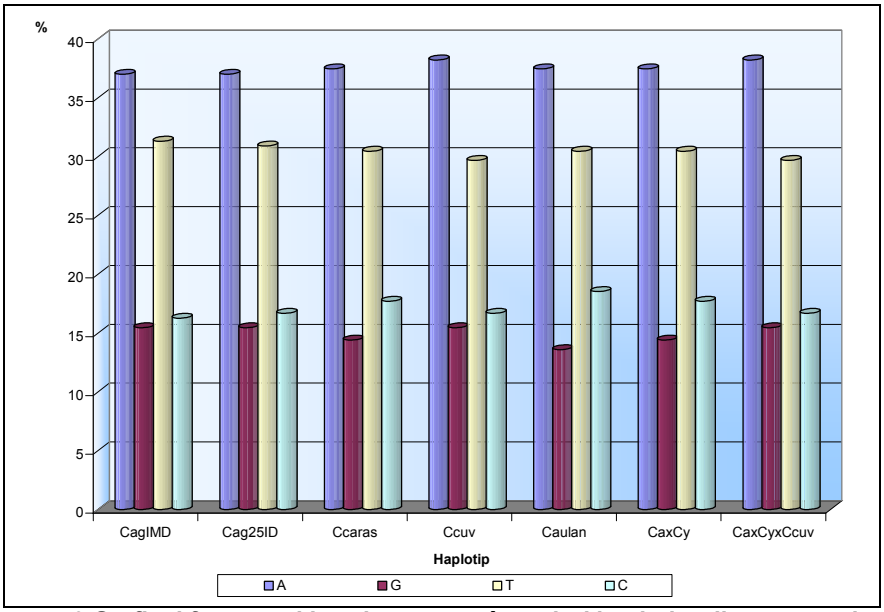


Figura 70 Graficul frecvenței bazelor azotate în cadrul haplotipurilor caracterizate

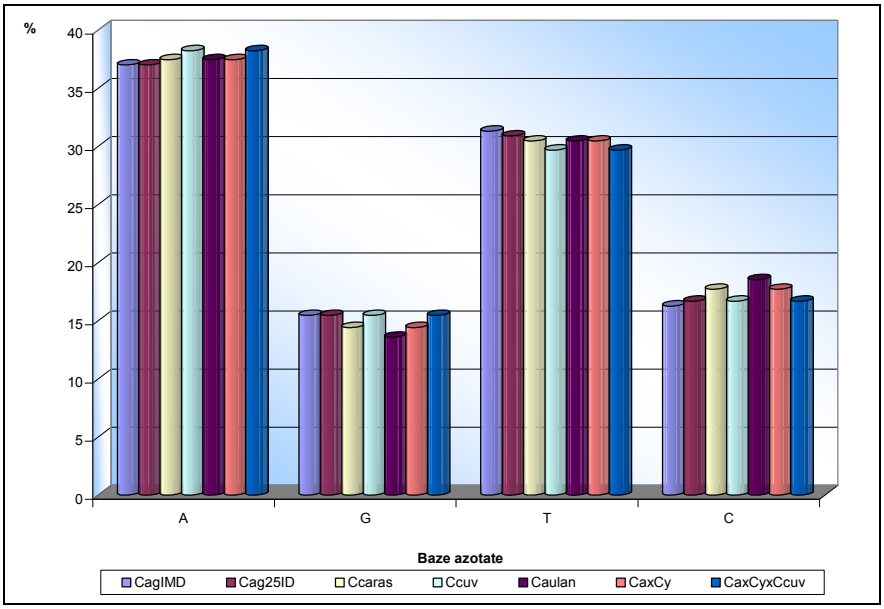


Figura 71 Graficul frecvenței bazelor azotate, comparativ între haplotipurile analizate

Din Tabelul 46 şi Figurile 70 – 71, constatăm că pentru haplotipurile descoperite în cele două populații analizate, concentrația de adenină este de 36,99% pentru ambele haplotipuri analizate, ca şi cantitatea de guanină, care este de 15,45%. Timina are valori de 31,30% pentru haplotipul CagIMD şi 30,89% pentru haplotipul Cag25ID, iar citozina, valori de 18,26% pentru primul haplotip şi 16,67% pentru cel de-al doilea.

Procentul bazelor complementare este de 68,29% pentru A+T în cazul primului haplotip şi de 67,89% pentru cel de-al doilea, iar procentele pentru C+G sunt de 31,71% pentru CagIMD şi 32,11% pentru Cag25ID.

Din analiza valorilor, constatăm că există o variație a cantității de adenină, cuprinsă între 36,99%pentru haplotipurile celor două populații şi 38,21% pentru haplotipurile Ccuv şi CaxCyxCcuv. Cantitatea de guanină variază asemănător, având valori cuprinse între 13,58% pentru Caulan şi 15,45% pentru cele două haplotipuri analizate, Ccuv, şi CaxCyxCcuv. Referitor la cantitatea de timină, se constată că variază între 29,67% pentru Ccuv şi hibridul CaxCyxCcuv şi 31,30% pentru haplotipul CagIMD. Cantitatea de citozină, variază între 16,26% pentru primul haplotip şi 18,52% pentru Caulan.

Referitor la procentele în care se găsesc bazele complementare, pentru A+T se înregistrează valori cuprinse între 67,89% pentru haplotipurile Cag25ID, Ccuv şi CaxCyxCcuv şi 68,29% pentru CagIMD, iar pentru C+G, valori cuprinse între 31,71% (CagIMD) şi 32,11% pentru secvențele Cag25ID, Ccuv şi CaxCyxCcuv.

4.9.1. CARACTERIZAREA HAPLOTIPULUI CAGIMD

4.9.1.1. FRECVENȚELE TIPURILOR DE ASOCIERI NUCLEOTIDICE

În acest haplotip au fost încadrate toate secvențele între care nu există diferențe (Cag01I, Cag03I, Cag04I, Cag06I, Cag07I, Cag08I, Cag09I, Cag10I, Cag01M,

Cag02M, Cag03M, Cag04M, Cag05M, Cag06M, Cag07M, Cag08M, Cag09M, Cag10M), aparţinând celor două populaţii luate în studiu, Larga-Jijia – Movileni şi Iaşi.

Din analiza procentelor dinucleotidelor, constatăm că valoarea cea mai mare o au perechile AA şi AT (13,1%), iar valoarea cea mai mică perechea CG, în proporţie de 0,4%. Referitor la trinucleotide, cea mai mare frecvenţă o înregistrează tripletele AAT şi TAA (4,9%), iar valoarea cea mai mică, tripletele AGC, CCG, CGG, CTG, GAT şi GTC, având o frecvenţă de 0,4%. În plus, se constată lipsa trimerilor CGA, CGC, CGT, GCC, GCG, GCT, GGC şi TCG.

Tetramerul cu frecvenţa cea mai mare, este TATT (2,5%), iar cei cu frecvenţă redusă se găsesc în procent de 1,2%. În cazul pentamerilor (Tabelul 99), cea mai mare frecvenţă o au pentamerii TTATT, TATTA, AACTA şi AAATT (1,2%), iar toţi ceilalţi au frecvenţe foarte mici (0,8%). În ceea ce priveşte hexamerii, aceştia au frecvenţe cuprinse între 0,8% şi 0,4%.

Tetramerii, pentamerii şi hexamerii cu frecvenţă joasă (Tabelul 100), sunt cei care se găsesc într-un singur exemplar în întreaga genă, având o frecvenţă de 0,4%.

Cu ajutorul modulului EditSeq al DNA Star, a fost calculată temperatura de topire a ADN, după un model Davis-Botstein-Roth; în acest caz, prezintă o valoare de 75,87ºC.

4.9.2. CARACTERIZAREA HAPLOTIPULUI CAG25ID

4.9.2.1. FRECVENŢELE TIPURILOR DE ASOCIERI NUCLEOTIDICE

În acest haplotip, se încadrează două secvenţe Cag02I şi Cag05I din populaţia Iaşi. În plus, se poate constata că mutaţia apărută în populaţiile studiate, datorită frecvenţei cu care apare, este în curs de a fi fixată în genofondul acestora.

Din analiza procentelor dinucleotidelor, constatăm că valoarea cea mai mare o are perechea AA (13,1%), iar valoarea cea mai mică perechea CG, în proporţie de 0,4%. Referitor la trinucleotide, cea mai mare frecvenţă o înregistrează tripletul AAT (4,9%), iar valoarea cea mai mică, tripletele AGC, CCG, CGG, CTG, GAT şi GTC, având o frecvenţă de 0,4%. În plus, se constată lipsa trimerilor ACG, CGA, CGC, CGT, GAG, GCC, GCG, GCT, GGC şi TCG.

Tetramerii cu frecvenţa cea mai mare, sunt TATT, ATTA, ATAT, ACAT şi AATT, (2,1%), iar cei cu frecvenţă redusă se găsesc în procent de 1,2%. În cazul pentamerilor (Tabelul 103), cea mai mare frecvenţă o au pentamerii TATTA, ACTCA, AACTA şi AAATT (1,2%), iar toţi ceilalţi au frecvenţe foarte mici (0,8%). În ceea ce priveşte hexamerii, aceştia au frecvenţe cuprinse între 0,8% şi 0,4%.

Tetramerii, pentamerii şi hexamerii cu frecvenţă joasă, sunt cei care se găsesc într-un singur exemplar în întreaga genă, având o frecvenţă de 0,4%.

Cu ajutorul modulului EditSeq al DNA Star, a fost calculată temperatura de topire a ADN, după un model Davis-Botstein-Roth; în acest caz, prezintă o valoare de 76,03ºC.

4.9.3. RELAŢII FILOGENETICE ÎN CADRUL GENULUI *CARASSIUS*, BAZATE PE DIFERENŢE ALE SECVENŢELOR D-LOOP

Ca şi în cazul citocromului b, au fost utilizate metode care se bazează pe distanţe şi rata de substituţie, dar care implică şi mutaţii de genul tranziţiilor şi transversiilor.

Astfel şi în acest caz, au fost utilizate UPGMA (Figura 72) pe baza unui model Tamura-Nei, metoda evoluției minime (ME) după un algoritm Neighbour-Joining (Figura 73) şi Maximum Parsimony (Figura 74).

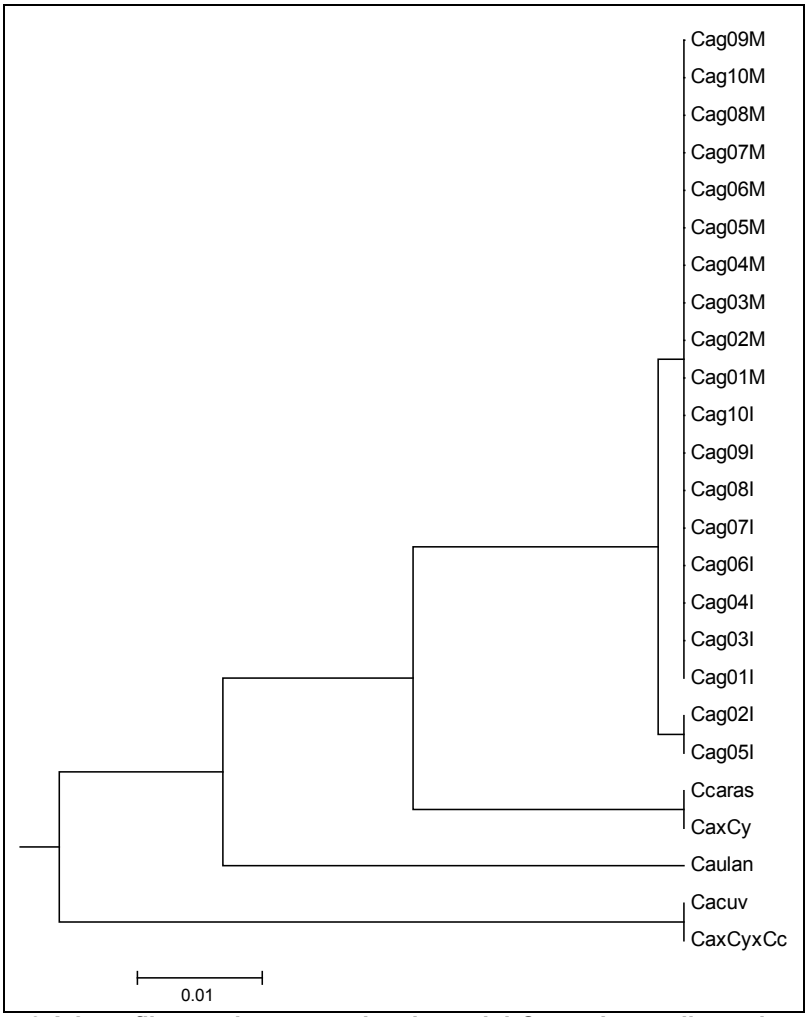


Figura 72 Arbore filogenetic rectangular al genului *Carassius*, realizat prin metoda UPGMA pe baza secvențelor d-loop

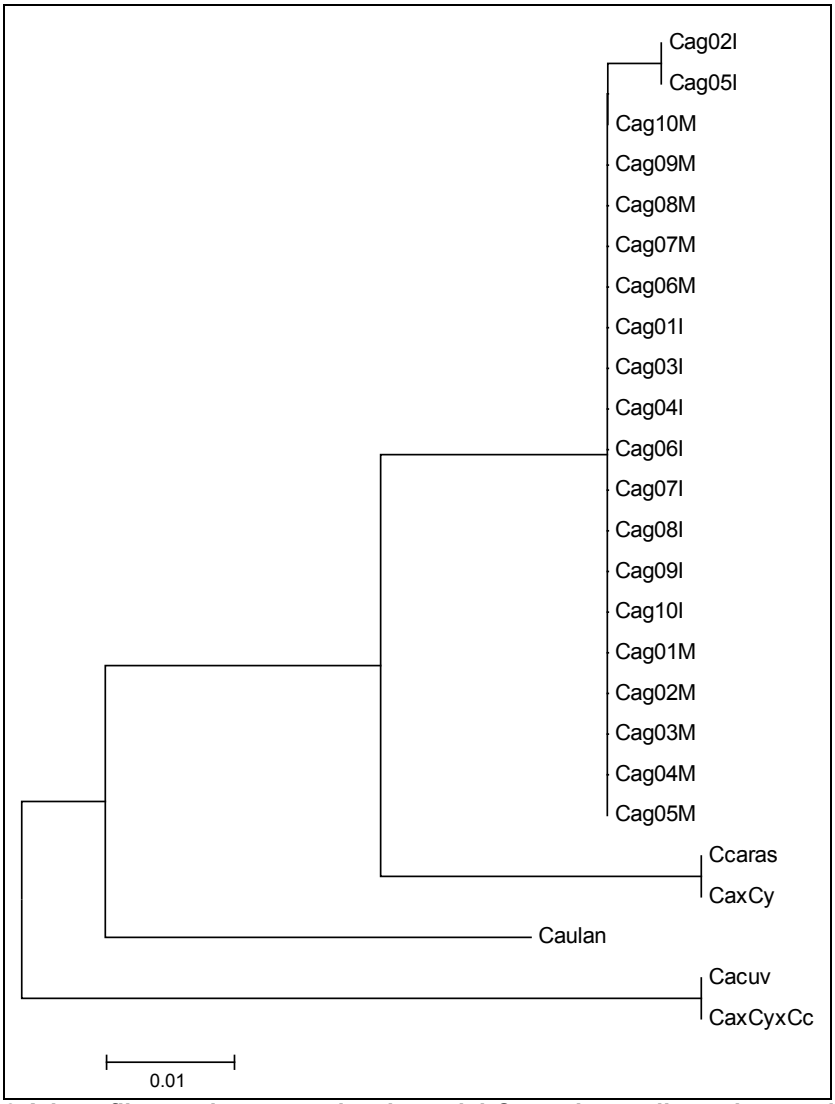


Figura 73 Arbore filogenetic rectangular al genului *Carassius*, realizat prin metoda ME pe baza secvențelor d-loop

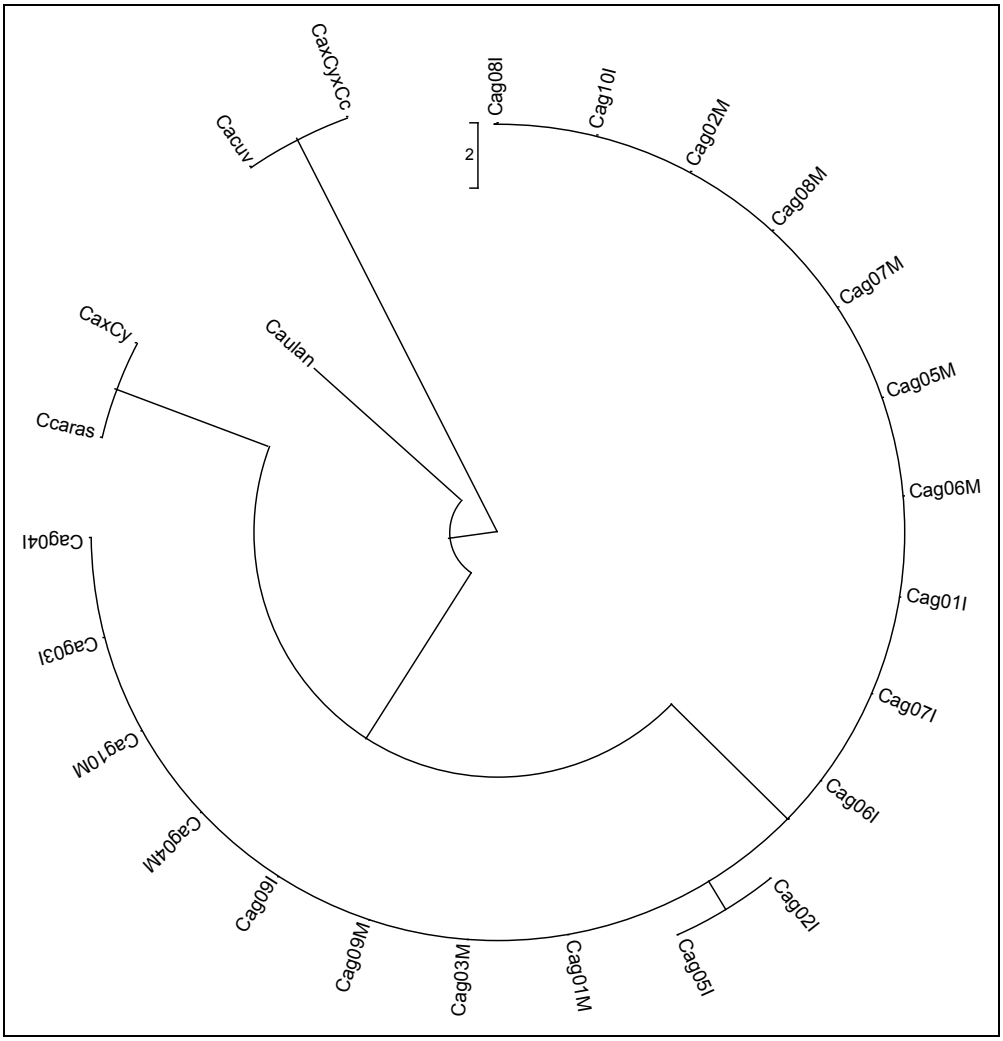


Figura 74 Arbore filogenetic circular al genului *Carassius*, realizat prin metoda MP pe baza secvențelor d-loop

Ca şi în cazul citocromului b, topologia arborilor obținuți prin toate cele trei metode este identică, evidențiind gruparea taxonilor în categorii de similaritate precum şi treapta evolutivă a fiecărui grup. Astfel, se observă că specia *Carassius auratus langsdörfi* provine din aceeaşi specie de origine ca şi *Carassius cuvieri*, precum şi grupul din care au derivat speciile *Carassius carasssius* şi *Carassius gibelio*. De asemeni, se observă că hibrizii sunt situați în aceleaşi grupe cu speciile de origine.

4.10. COMPARAREA SECVENȚELOR NUCLEOTIDICE ALE GENEI CARE DETERMINĂ SINTEZA ND4L DEHIDROGENAZEI, PROVENITE DE LA INDIVIZI APARȚINÂND GENULUI *CARASSIUS*

Pentru subunitatea 4L a dehidrogenazei, în cazul speciilor genului *Carassius*, au fost aliniate 20 de secvențe provenite de la indivizi ai celor două populații luate în studiu (Larga-Jijia – Movileni şi Iaşi), precum şi secvențe preluate din banca de gene GenBank, provenite de la specii (*Carassius carassius* Ref. nr. AY714387 şi *Carassius cuvieri* Ref. nr. AB045144), subspecii (*Carassius auratus langsdörfi* Ref. nr. AB006953) şi hibrizi (*Carassius auratus x Cyprinus carpio* Ref. nr. AY694420 şi *Carassius auratus x Cyprinus carpio x Carassius cuvieri* Ref. nr. AY771781) ai genului *Carassius* (Gorgan, 2004b; Gorgan şi Cîmpeanu, 2004f).

Alinierea s-a realizat, utilizând metoda Clustal V (Higgins şi Sharp, 1989) şi a fost verificată prin metoda Clustal W (Thompson et al., 1994), obținându-se în ambele cazuri acelaşi rezultat.

Din alinierea secvențelor, se constată existența a 51 diferențe, dintre care 36 de tranziții şi 15 transversii, doar între secvențele preluate din banca de gene. De asemeni, se constată că între secvențe provenite din cele două populații analizate, nu există diferențe, de unde putem concluziona ținând cont şi de numărul mic de mutații existent între celelalte secvențe, că gena care codifică informația sintezei subunității 4L a dehidrogenazei, este o genă conservativă, prezentând un potențial mutagen foarte scăzut. Din cele discutate anterior, putem concluziona că a fost găsit un nou haplotip (CagIMN) pentru gena ND4L, caracteristic speciei *Carassius auratus gubelio*, din populațiile Iaşi şi Movileni.

Pe baza comparației între secvențe, s-a realizat Tabelul procentelor de similaritate şi divergență (Tabelul 47)., din care observăm că procentele de similaritate pentru haplotipul stabilit sunt de 97,7% din comparația cu hibridul *Carassius auratus x Cyprinus carpio* şi 95,1% pentru comparația cu secvența speciei *Carassius auratus langsdörfi*.

Tabel 47 Procentele de similaritate şi divergenţă pentru secvenţele haplotipurilor stabilite

	Procent de similaritate							
Procent de divergenţă		CagIMN	Cacuv	Caulan	CaxCy	CaxCyxCcuv	Ccaras	
	CagIMN		93,5	95,1	97,7	97,0	97,3	CagIMN
	Cacuv	6,8		93,5	93,9	94,7	94,3	Cacuv
	Caulan	5,2	6,9		97,3	95,8	97,0	Caulan
	CaxCy	2,3	6,5	2,7		98,5	99,6	CaxCy
	CaxCyxCcuv	3,1	5,6	4,4	1,5		98,9	CaxCyxCcuv
	Ccaras	2,7	6,1	3,1	0,4	1,2		Ccaras
		CagIMN	Cacuv	Caulan	CaxCy	CaxCyxCcuv	Ccaras	

În plus, pentru a stabili o mai clară diferenţiere între toate noile haplotipuri, au fost analizate numărul şi procentele în care se găsesc bazele azotate în fiecare dintre acestea (Tabel 48).

Tabel 48 Procentul bazelor azotate, în cadrul haplotipurilor stabilite

Baze azotate		Specii analizate					
		CagIMN	Ccaras	Ccuv	Caulan	CaxCy	CaxCyxCcuv
A	Nr,	64	59	57	59	60	59
	%	24,52	22,61	21,84	22,61	22,99	22,61
G	Nr,	43	45	47	45	44	45
	%	16,48	17,24	18,01	17,24	16,86	17,24
T	Nr,	78	80	79	78	80	81
	%	29,89	30,65	30,27	29,89	30,65	31,03
C	Nr,	76	77	78	79	77	16
	%	29,12	29,50	29,89	30,27	29,50	29,12
A+T	Nr,	142	139	136	137	140	140
	%	54,41	53,26	52,11	52,49	53,64	53,64
C+G	Nr,	119	122	125	124	121	121
	%	45,59	46,74	47,09	47,51	46,36	46,36

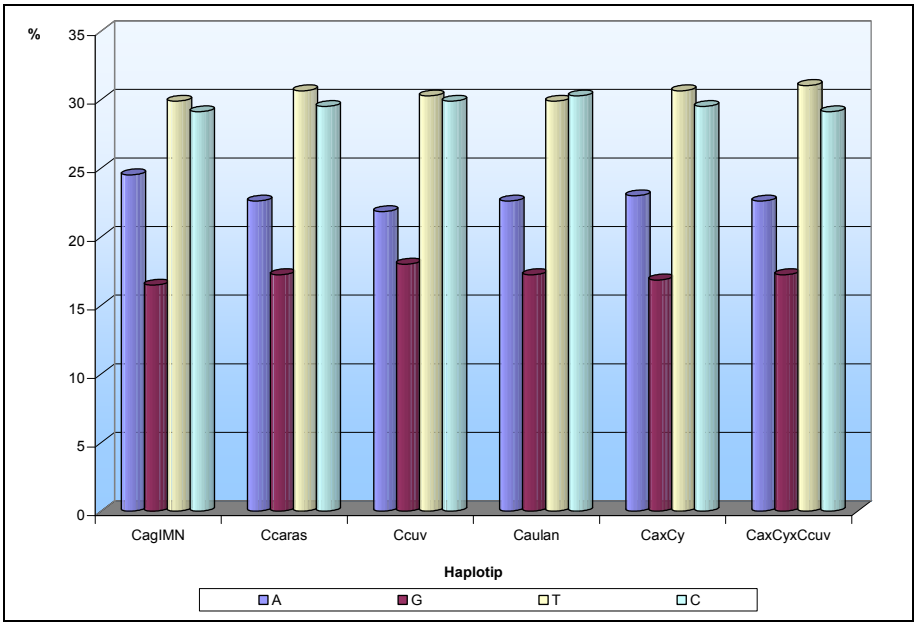


Figura 78 Graficul frecvenței bazelor azotate în cadrul haplotipurilor caracterizate

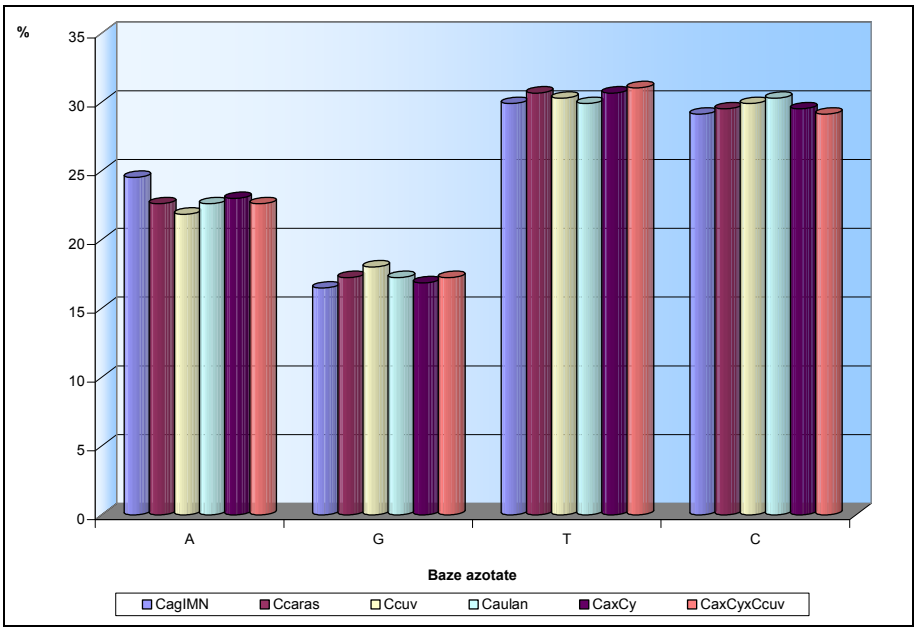


Figura 79 Graficul frecvenței bazelor azotate, comparativ între haplotipurile analizate

Din Tabelul 48 şi Figurile 78 - 79, constatăm că, referitor la concentrația de adenină se înregistrează valori cuprinse între 24,52% pentru haplotipul populației Iaşi şi 21,84% pentru *Carassius cuvieri*. Cantitatea de guanină are valori cuprinse între 18,01% pentru Ccuv şi 16,48% pentru haplotipul CagIMN. De asemeni, cantitatea de

timină, are valori cuprinse între 31,03% pentru hibridul *Carassius auratus x Cyprinus carpio x Carassius cuvieri* şi 29,89% pentru haplotipurile CagIMN şi Caulan. Pentru citozină, se înregistrează valori cuprinse între 30,27% pentru Caulan şi 29,12% pentru secvențele CaugIMN şi CyxCyxCcuv.

Referitor la procentele în care se găsesc bazele complementare, pentru A+T se înregistrează valori cuprinse între 54,41% pentru haplotipul CagIMN şi 52,11% pentru Ccuv, iar pentru C+G, valori cuprinse între 45,59% (CagIMN) şi 47,51% pentru secvența Caulan.

4.10.1. CARACTERIZAREA HAPLOTIPULUI CAGIMN

4.10.1.1. FRECVENȚELE TIPURILOR DE ASOCIERI NUCLEOTIDICE

În acest haplotip au fost încadrate toate secvențele, aparținând celor două populații luate în studiu, Larga-Jijia – Movileni şi Iaşi.

Din analiza, procentelor dinucleotidelor constatăm că valoarea cea mai mare o are perechea TT (10,7%), iar valoarea cea mai mică perechea GT, în proporție de 1,1%. Referitor la trinucleotide, cea mai mare frecvență o înregistrează tripletul TTT (4,2%), iar valoarea cea mai mică, tripletele AGT, CGC, CGG, CGT, GAG, GAT, GCG, GTA, TCG şi TGT, având o frecvență de 0,4%. Se constată că există şi trinucleotide lipsă: AAA, ACG, GGG, GGT, GTC şi GTG.

Tetramerul cu frecvența cea mai mare, este CCTA (2,7%), iar cei cu frecvență redusă: TTTT, TTGC, TTCA, TTAT, TACT şi TACA, toți în procent de 1,2%. În cazul pentamerilor, cea mai mare frecvență o are pentamerul CCCTA (1,5%), iar restul, au valori de 1,2% - 0,8%. În ceea ce priveşte hexamerii, aceştia au frecvențe cuprinse între 1,2% şi 0,4%.

Tetramerii, pentamerii şi hexamerii cu frecvență joasă, sunt cei care se găsesc într-un singur exemplar în întreaga genă, având o frecvență de 0,4%.

Cu ajutorul modulului EditSeq al DNA Star, a fost calculată temperatura de topire a ADN, după un model Davis-Botstein-Roth; în acest caz, prezintă o valoare de 81,68ºC.

4.10.1.2. TRANSLAȚIA ŞI CARACTERIZAREA CATENEI POLIPEPTIDICE

Translația s-a realizat pe baza unui cod genetic mitocondrial pentru vertebrate, utilizând modulul EditSeq din cadrul programului DNA Star.

Catena polipeptidică obținută este prezentată în Figura 80.

```
MTPVHFSFSS AFILGLMGLA FHRTHLLSAL LCLEGMMLSL FIALALWALQ
FESTGFSTAP MLLLAFSACE ASTGLALLVA TARTHGT
```

Figura 80 Catena polipeptidică a ND4L caracteristică haplotipului CagIMN al speciei *Carassius gibelio*

După realizarea translației, se constată că întreg lanțul polipeptidic, este alcătuit din 87 aminoacizi (Tabelul 49), din care se observă că frecvența cea mai mare, o are leucina (24,84%), iar cea mai scăzută o au acidul glutamic şi triptofanul (1,15%).

Tabel 49 Compoziția lanțului polipeptidic

Aminoacizi		Număr	Procent din greutatea totală (%)	Frecvența (%)
Simbol	Nume			
A	Ala	13	9,99	14,94
C	Cys	2	2,23	2,30
D	Asp	0	0,00	0,00
E	Glu	3	4,19	3,45
F	Phe	8	12,73	9,20
G	Gly	6	3,70	6,90
H	His	4	5,93	4,60
I	Ile	2	2,45	2,30
K	Lys	0	0,00	0,00
L	Leu	19	23,25	21,84
M	Met	5	7,09	5,75
N	Asn	0	0,00	0,00
P	Pro	2	2,10	2,30
Q	Gln	1	1,39	1,15
R	Arg	2	3,38	2,30
S	Ser	9	8,48	10,34
T	Thr	8	8,75	9,20
V	Val	2	2,14	2,30
W	Trp	1	2,01	1,15
Y	Tyr	0	0,00	0,00
B	Asx	0	0,00	0,00
Z	Glx	0	0,00	0,00
.	Ter	0	0,00	0,00

Se constată că din totalul de 87 aminoacizi, 3 sunt acizi, 2 sunt bazici, 20 polari şi 45 hidrofobi (Tabelul 50). De asemenea, se constată că lanțul polipeptidic are o sarcină electrică de -0,48 (la pH=7) şi un punct izoelectric de 6,74 (Figura 81).

Tabel 50 Caracterizarea lanțului polipeptidic

Greutate molecular ă (Daltoni)	Număr de aminoaciz i	Aminoacizi				Punct izoelectri c	Sarcina electric ă la pH=7	Concentrați a la o extincție =1 și λ=280nm
		Puterni c bazici	Puterni c acizi	Hidrofob i	Polar i			
9247,07	87	2	3	45	20	6,736	-0,479	1,56

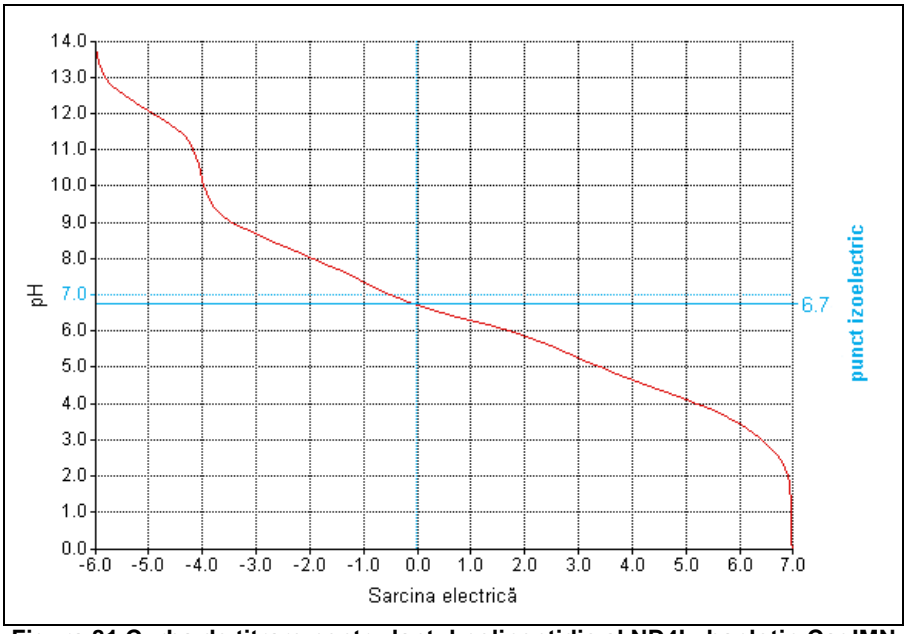


Figura 81 Curba de titrare pentru lanțul polipeptidic al ND4L, haplotip CagIMN al speciei *Carassius gibelio*

Ulterior, a fost realizată o analiză a structurii proteinei, fiind determinate zonele alfa helicoidale, beta pliate şi de inflexiune, prin două metode Garnier-Robson (Garnier et al., 1978) şi Chou-Fasman (Chou şi Fasman, 1978); reprezentarea grafică a zonelor hidrofobe (Kyte şi Doolittle, 1982), zonele de flexie (Karplus şi Schultz, 1985), indexul antigenic (Jameson şi Wolf, 1988) şi graficul probabil al suprafeței (Emini et al., 1985).

Din analiza structurii lanțului polipeptidic, se poate observa că referitor la regiunile alfa helicoidale, pentru ambele metode, sunt date ca fiind 2 astfel de zone. De asemenea, pentru regiunile beta pliate, se constată existența unei singure regiuni, în cazul ambelor metode. Referitor la zonele de inflexiune, comparativ pentru cele două metode, se constată că pentru prima metodă (Garnier–Robson), sunt prezentate 2 regiuni, iar prin a doua metodă (Chou-Fasman), sunt date 3 astfel de regiuni.

Pe baza graficului hidrofobicității, realizat după un model Kyte-Doolitle (Kyte şi Doolittle, 1982), se constată că există 6 zone hidrofobe şi 4 zone hidrofile.

Din graficul indexului antigenic, calculat pe baza modelului Jameson –Wolf (Jameson şi Wolf, 1988), se constată existența a 9 regiuni cu potențial antigenic.

4.10.2. RELAȚII FILOGENETICE ÎN CADRUL GENULUI *CARASSIUS*, OBȚINUTE PRIN ANALIZA SECVENȚELOR GENEI ND4L

În acest caz, arborii filogenetici au fost realizați utilizând aceleaşi metode UPGMA pe baza modelului Tamura-Nei (Figura 82) şi Minimum Evolution, pe baza unui algoritm Neighbour-Joining (Figura 83) şi Maximum Parsimony (Figura 84).

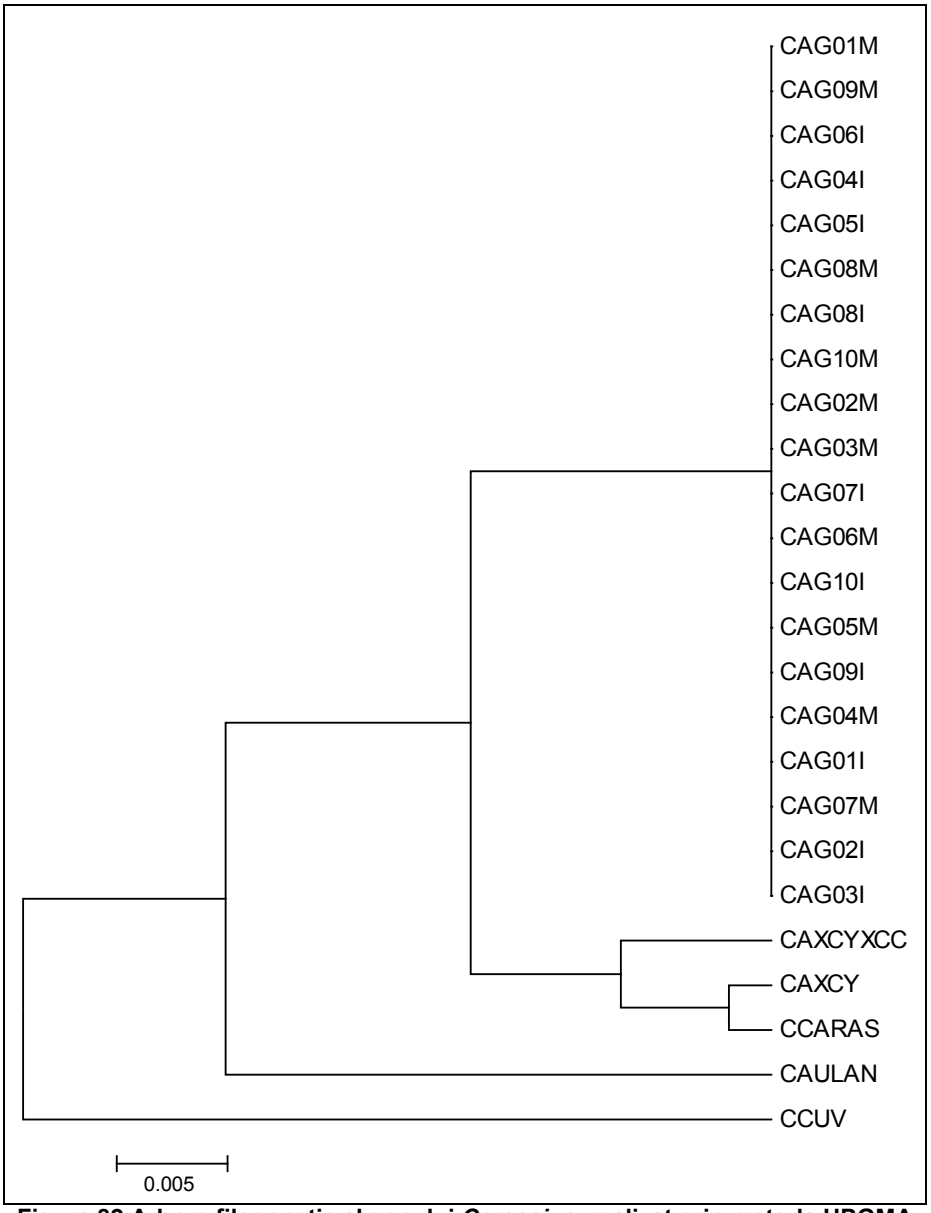


Figura 82 Arbore filogenetic al genului *Carassius*, realizat prin metoda UPGMA pe baza secvențelor ND4L

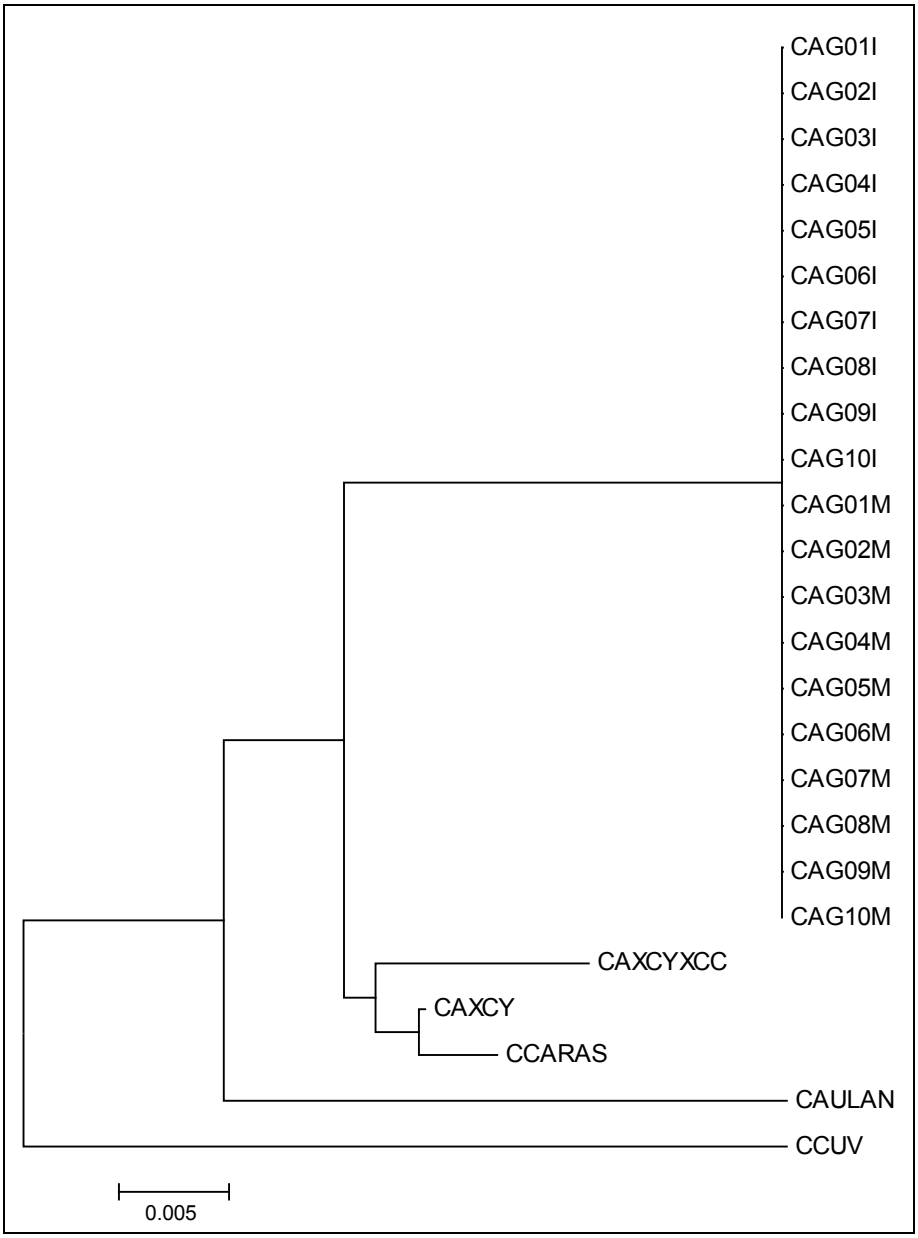


Figura 83 Arbore filogenetic al genului *Carassius*, realizat prin metoda ME pe baza secvențelor ND4L

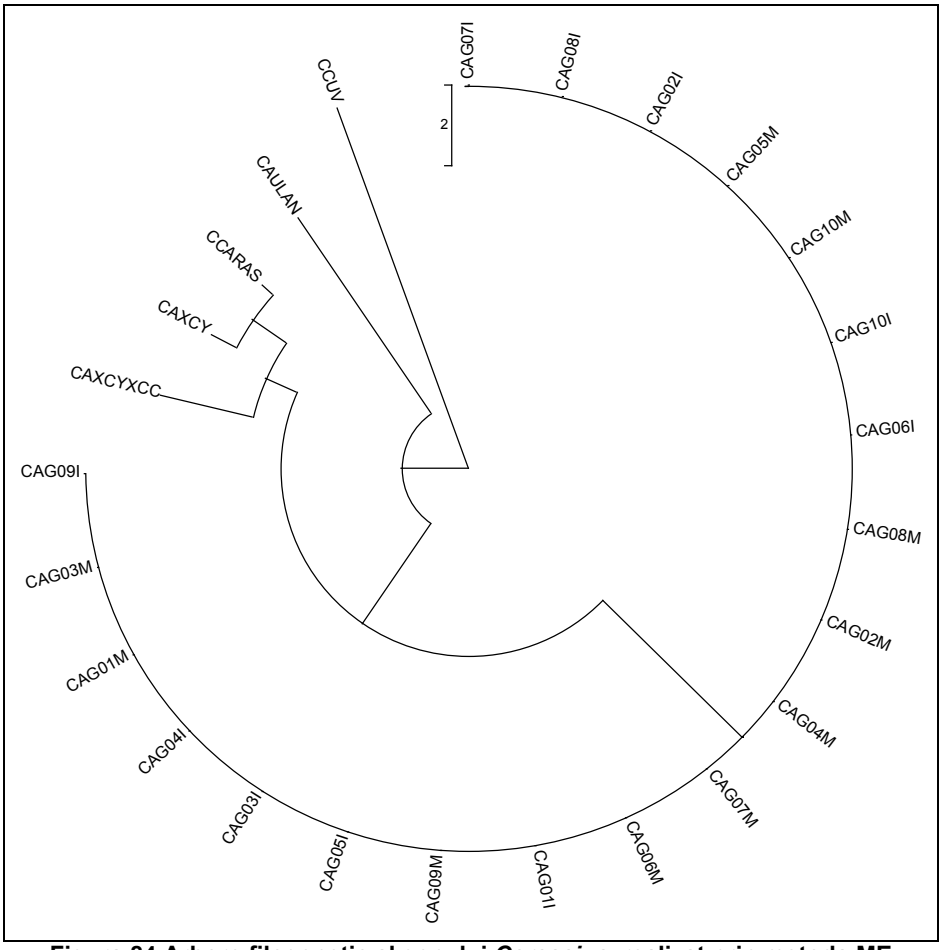


Figura 84 Arbore filogenetic al genului *Carassius*, realizat prin metoda ME pe baza secvențelor ND4L

Din Figurile 83 şi 84, pe baza diferențelor dintre secvențe şi a sumei lungimilor brațelor, se constată că individul a cărui secvență a fost preluată din banca de gene, provine dintr-un nod comun ancestral, cu ramurile din care au evoluat speciile celor două populații. De asemeni, pentru cele două populații analizate, se observă că indivizii acestora se grupează diferențiat. Se constată din cele două figuri, că există diferențe în cadrul arborelui, date de pozițiile ramurilor interioare, dar care nu prezintă o importanță la fel de mare ca topologia (aspectul general al arborelui), care rămâne neschimbată şi care ne furnizează informațile necesare pentru direcțiile filogenetice ale haplotipului analizat.

4.11. DIRECȚII EVOLUTIVE ÎN CADRUL FAMILIEI *CYPRINIDAE*

Acest subcapitol se doreşte a fi o sinteză a datelor de ordin molecular prezentate anterior şi mai exact o cumulare a direcțiilor evolutive ale speciilor aparținând la cele două genuri. În acest sens, au fost analizate liniile de evoluție caracteristice fiecărui

gen, cu decelarea speciilor de origine, diferențiat, pentru fiecare dintre cele trei gene secvențiate, deci pentru trei regiuni diferite ale ADNmt. Astfel, pentru citocromul b în cazul celor două genuri, *Carassius* şi *Cyprinus*, au fost trasați arborii filogenetici, utilizând cele trei metode menționate şi în capitolele anterioare UPGMA (Figura 85), Minimum Evolution (Figura 86) şi Maximum Parsimony (Figura 87).

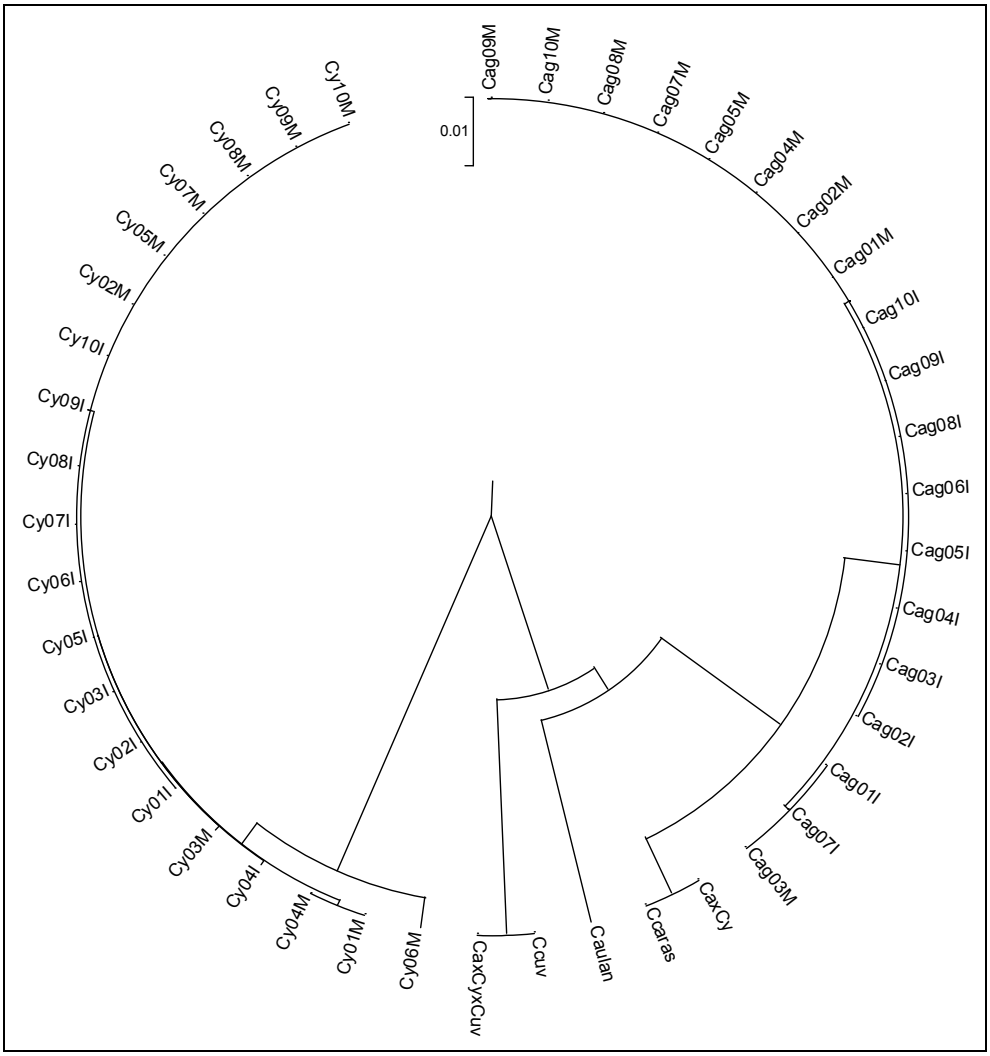


Figura 85 Arbore filogenetic circular realizat pe baza secvențelor citocromului b, pentru cele două genuri analizate, prin metoda UPGMA

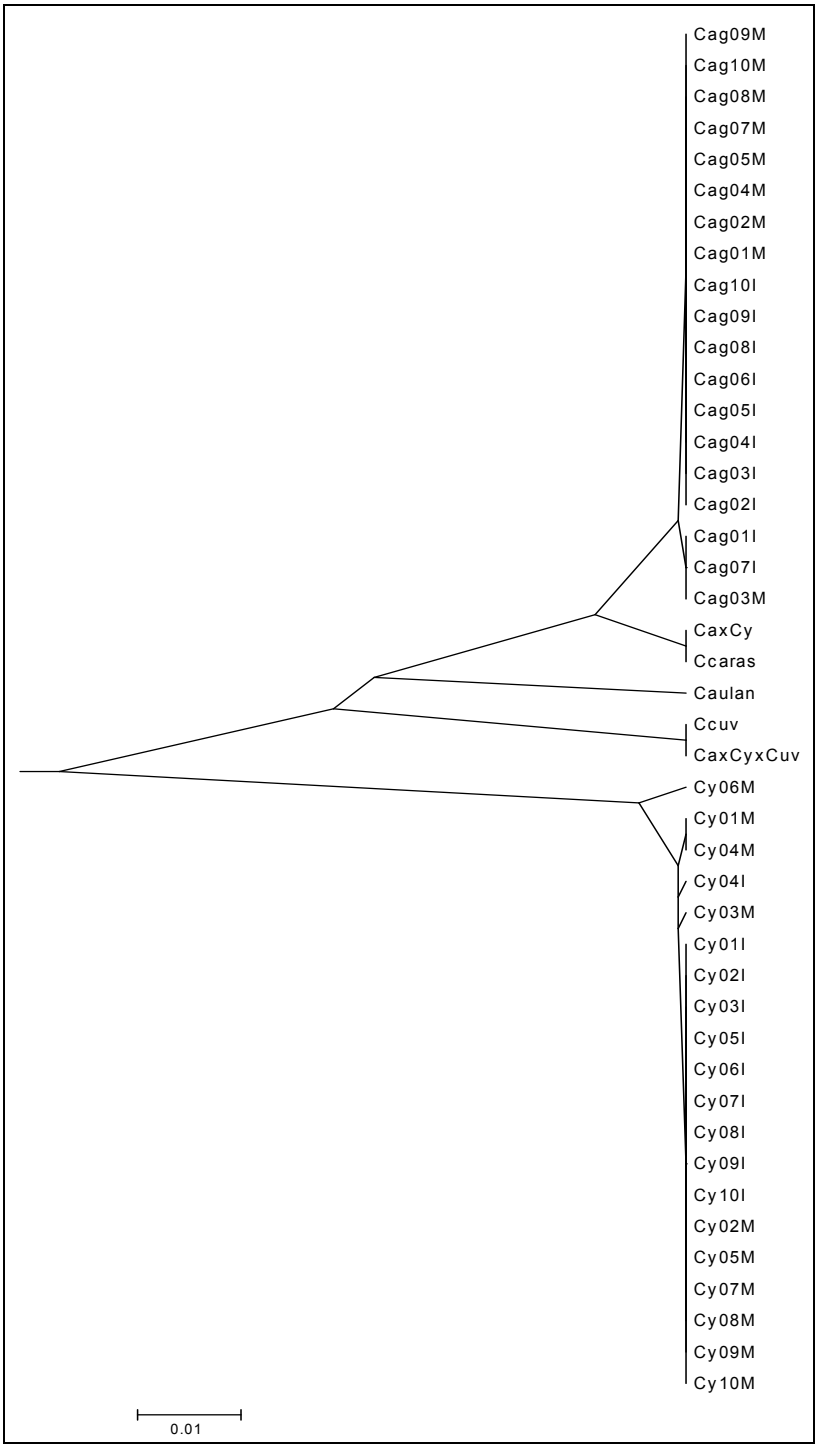


Figura 86 Arbore filogenetic realizat pe baza secvențelor citocromului b, pentru cele două genuri analizate, prin metoda ME

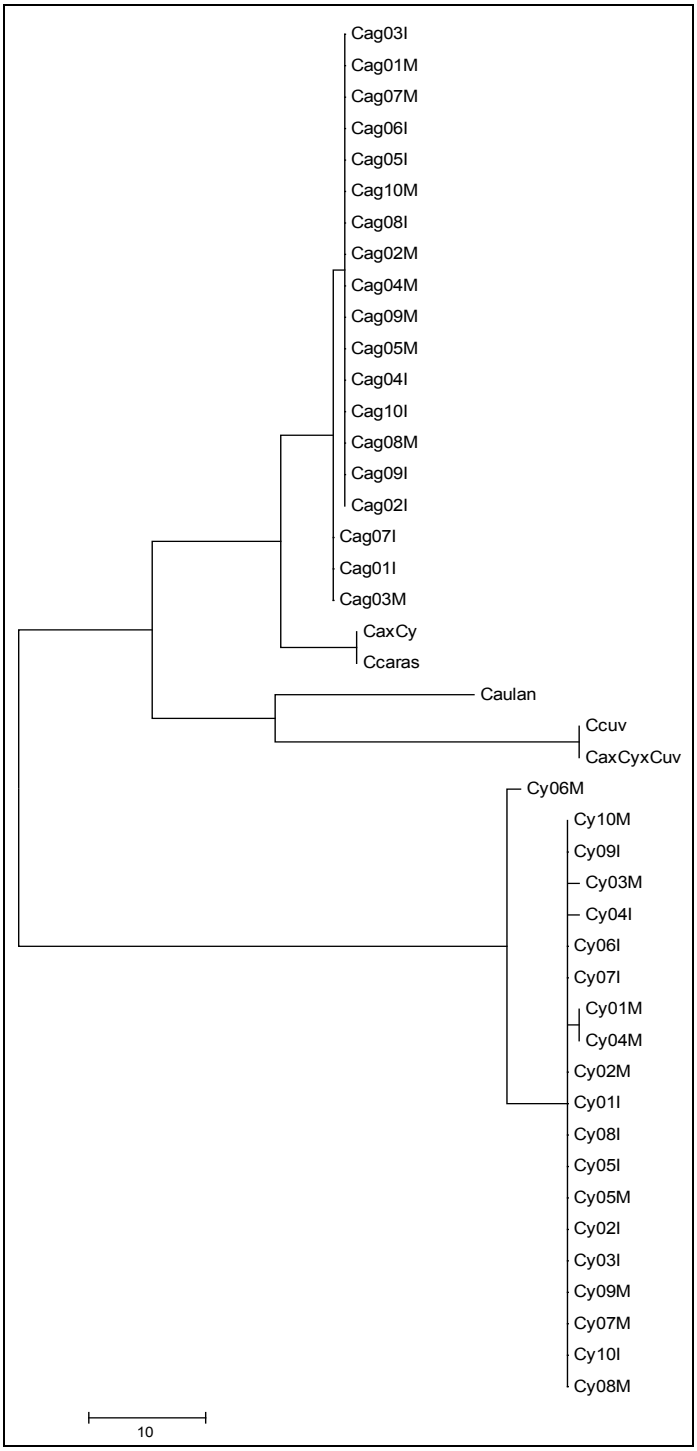


Figura 87 Arbore filogenetic realizat pe baza secvențelor citocromului b, pentru cele două genuri analizate, prin metoda MP

Referitor la direcțiile evolutive, trasate pe baza secvențelor citocromului b, se costată în toate cele trei cazuri, că specia *Carassius gibelio* a evoluat în paralel, dintr-un nod filogenetic comun, cu specia *Carassius carassius*. De asemeni, se constată că ambele au provenit dintr-o specie ancestrală comună, din care s-au diferențiat şi speciile *Carassius auratus langsdörfi* şi *Carassius cuvieri*; în plus această specie este foarte apropiată de speciile genului *Cyprinus*.

În cazul speciilor genului *Cyprinus*, se constată că direcțiile evolutive sunt mai restrânse, probabil şi datorită numărului mic de specii utilizat pentru analiză. Cu toate acestea, constatăm că individul de crap oglindă (Cy06M) este cel mai apropiat filogenetic de speciile genului *Cyprinus*, ceea ce ne demonstrează faptul că este obținut prin hibridare dintr-o specie ancestrală comună, din care provin şi restul indivizilor. În plus, se constată inițierea unei noi direcții evolutive în cazul haplotipului Cy0104M (dat de cele două secvențe Cy01M şi Cy04M), care poate constitui la rândul său un nod filogenetic pentru viitoare linii evolutive.

Pentru analiza filogenetică a regiunii de control mitocondrial au fost folosite secvențe de nucleotide comune celor două genuri, având lungimi de 246pb.

Din arborele reprezentat în Figura 88, realizat pe baza UPGMA după un model Tamura-Nei, constatăm şi în acest caz că specia *Carassius gibelio* provine ditr-un nod evolutiv comun cu specia *Carassius carassius*. În plus, din Figurile 88 şi 89, observăm că pentru regiunea de control mitocondrial, o specie ancestrală, care a evoluat monofiletic, poate fi considerată specia *Carassius cuvieri*. Figura 90, în schimb, ne sugerează că speciile *Carassius cuvieri* şi *Carassius auratus langsdörfi*, provin dintr-un nod filogenetic comun. Referitor la speciile genului *Cyprinus*, ca şi în cazul arborilor realizați pe baza secvențelor citocromului b, se constată că individul de crap oglindă se diferențiază evolutiv de restul indivizilor aparținând tipului „sălbatic”, fiind urmat pe scara evolutivă de individul Cy05M, aparținând aceleiaşi populații – Movileni.

Din arborii realizați pe baza secvențelor ND4L, se observă că pentru această regiune, specia *Carassius cuvieri* a evoluat monofiletic dintr-un nod filogenetic comun, din care au evoluat şi celelalte specii (Figurile 91 - 93). De asemeni, se consideră în cazul metodei MP, că speciile *Carassius auratus langsdörfi* şi *Carassius carassius*, derivă dintr-un hibrid interspecific (*Carassius aurats x Cyprinus carpio)*. Pentru această regiune, speciile genului *Cyprinus* se află pe acelaşi palier evolutiv, diferențiindu-se doar individul a cărui secvență a fost preluată din banca de gene.

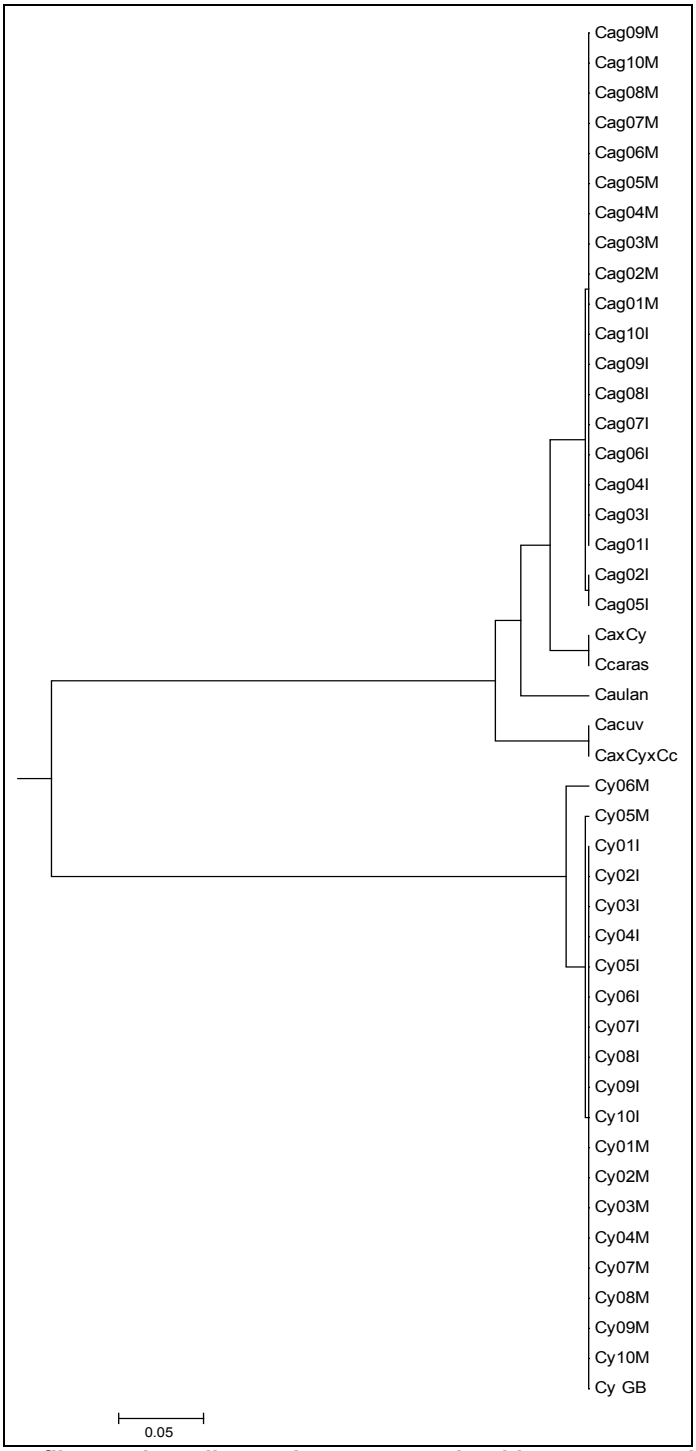


Figura 88 Arbore filogenetic realizat pe baza secvențelor d-loop, pentru cele două genuri analizate, prin metoda UPGMA

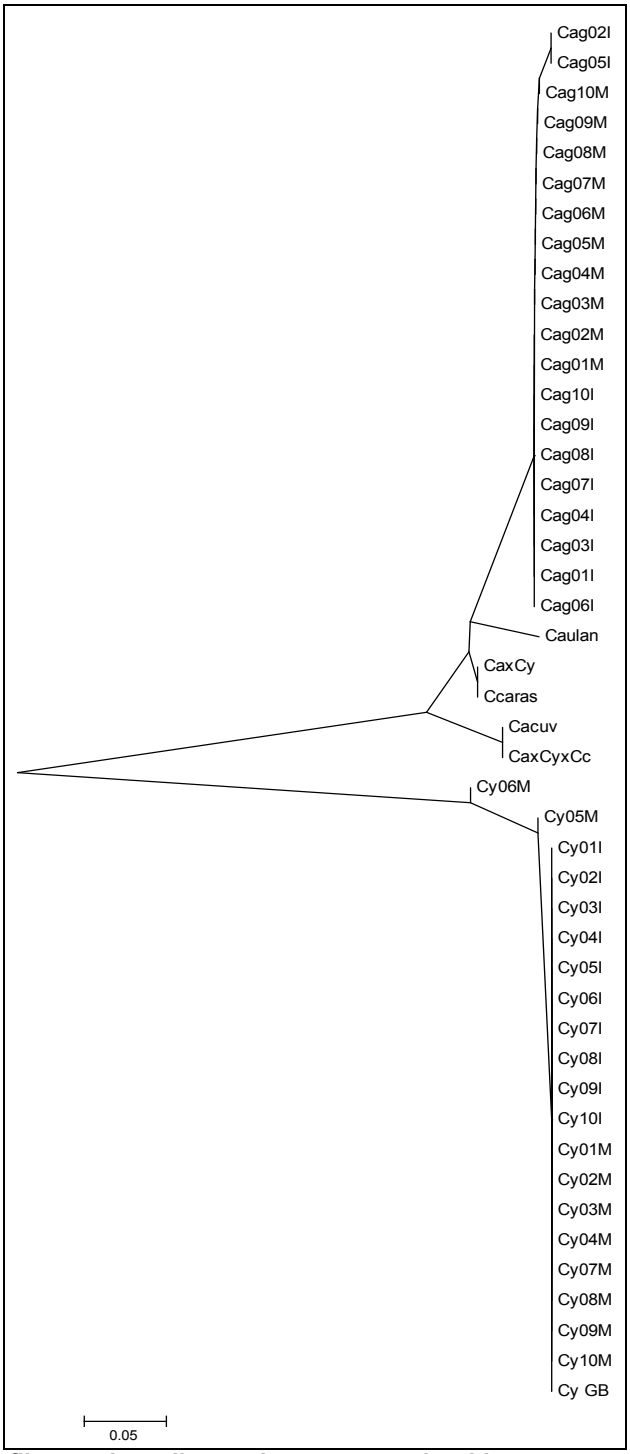


Figura 89 Arbore filogenetic realizat pe baza secvențelor d-loop, pentru cele două genuri analizate, prin metoda ME

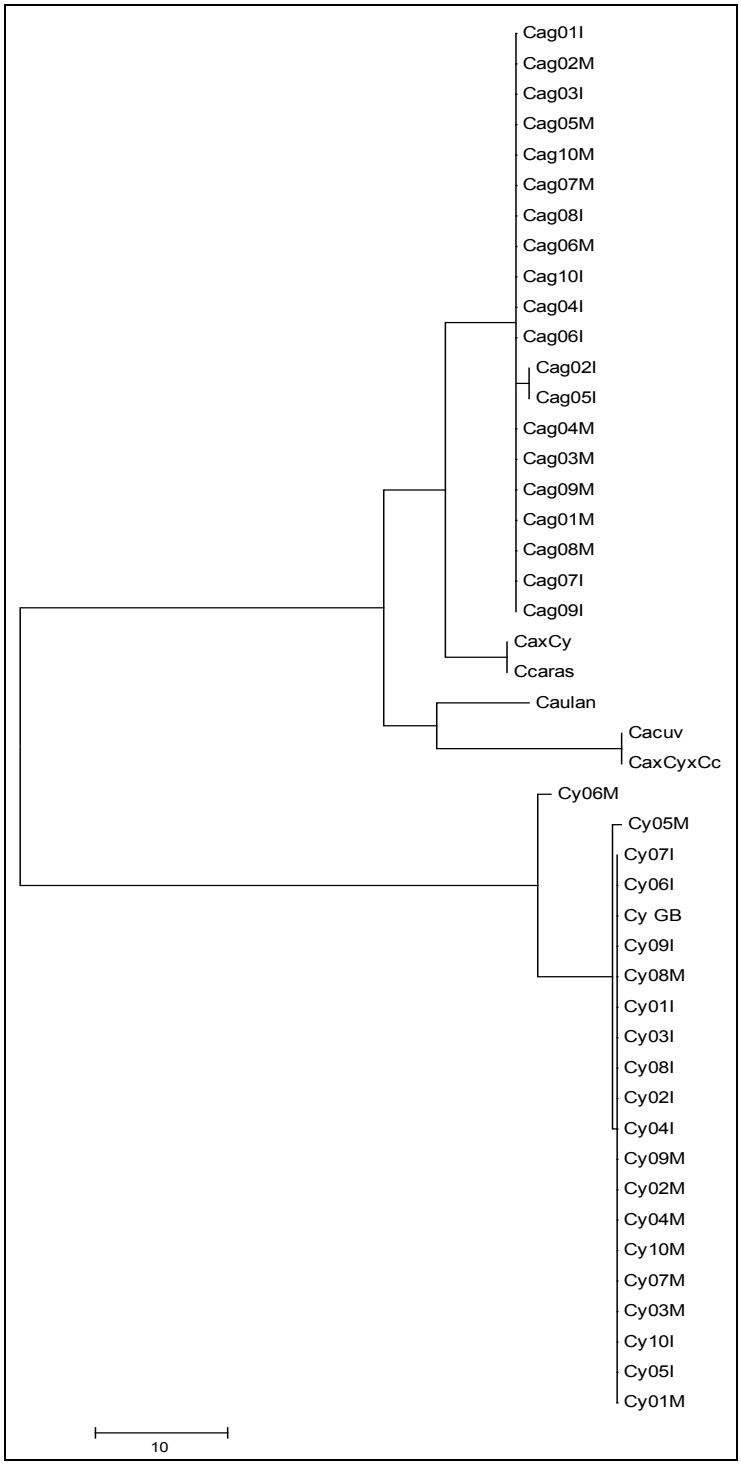


Figura 90 Arbore filogenetic realizat pe baza secvențelor d-loop, pentru cele două genuri analizate, prin metoda MP

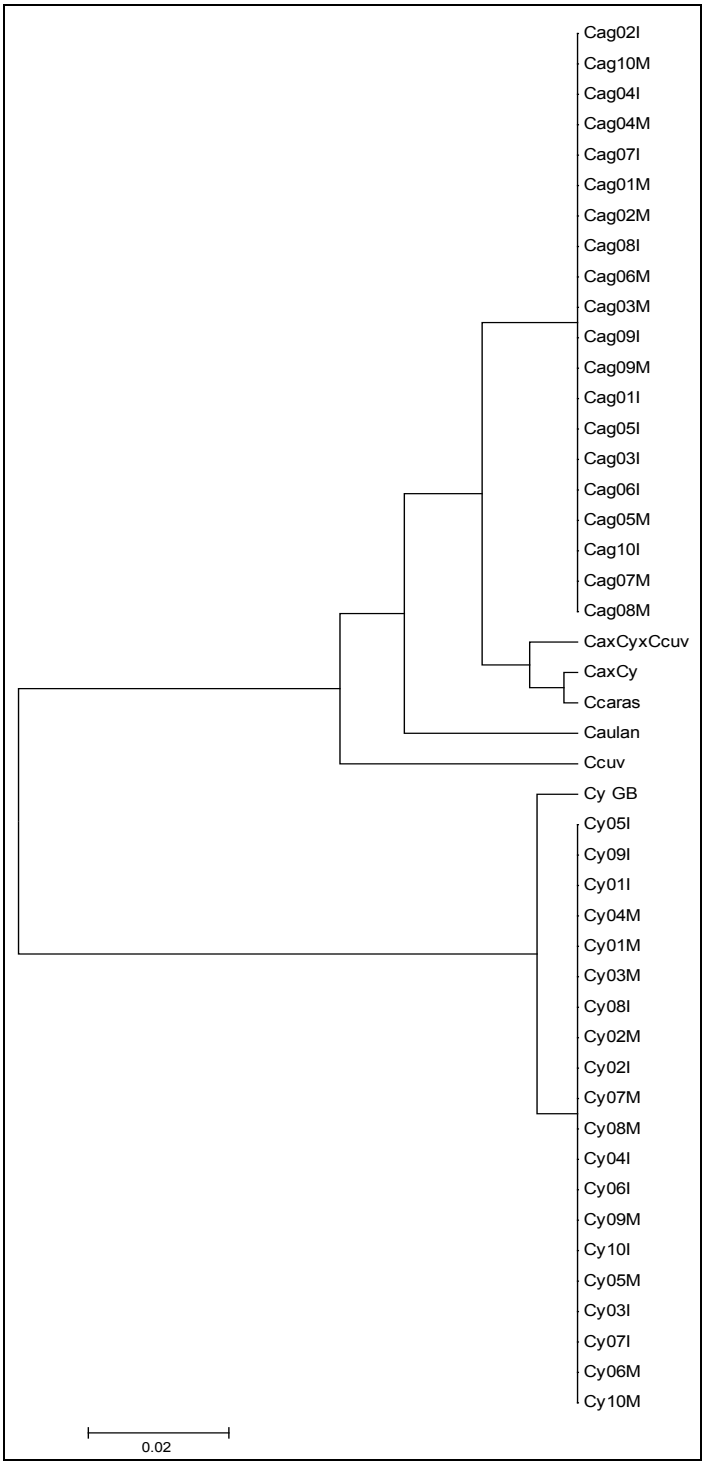


Figura 91 Arbore filogenetic realizat pe baza secvențelor ND4L, pentru cele două genuri analizate, prin metoda UPGMA

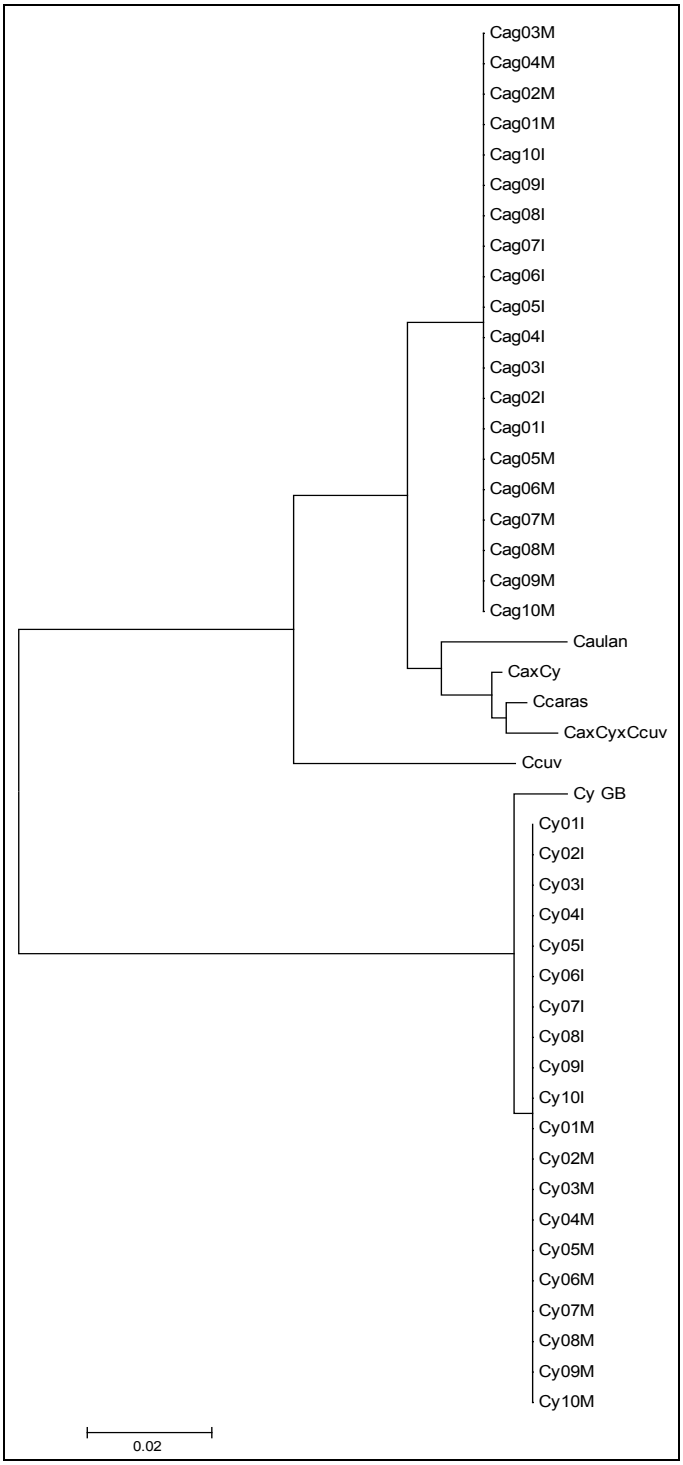


Figura 92 Arbore filogenetic realizat pe baza secvențelor ND4L, pentru cele două genuri analizate, prin metoda ME

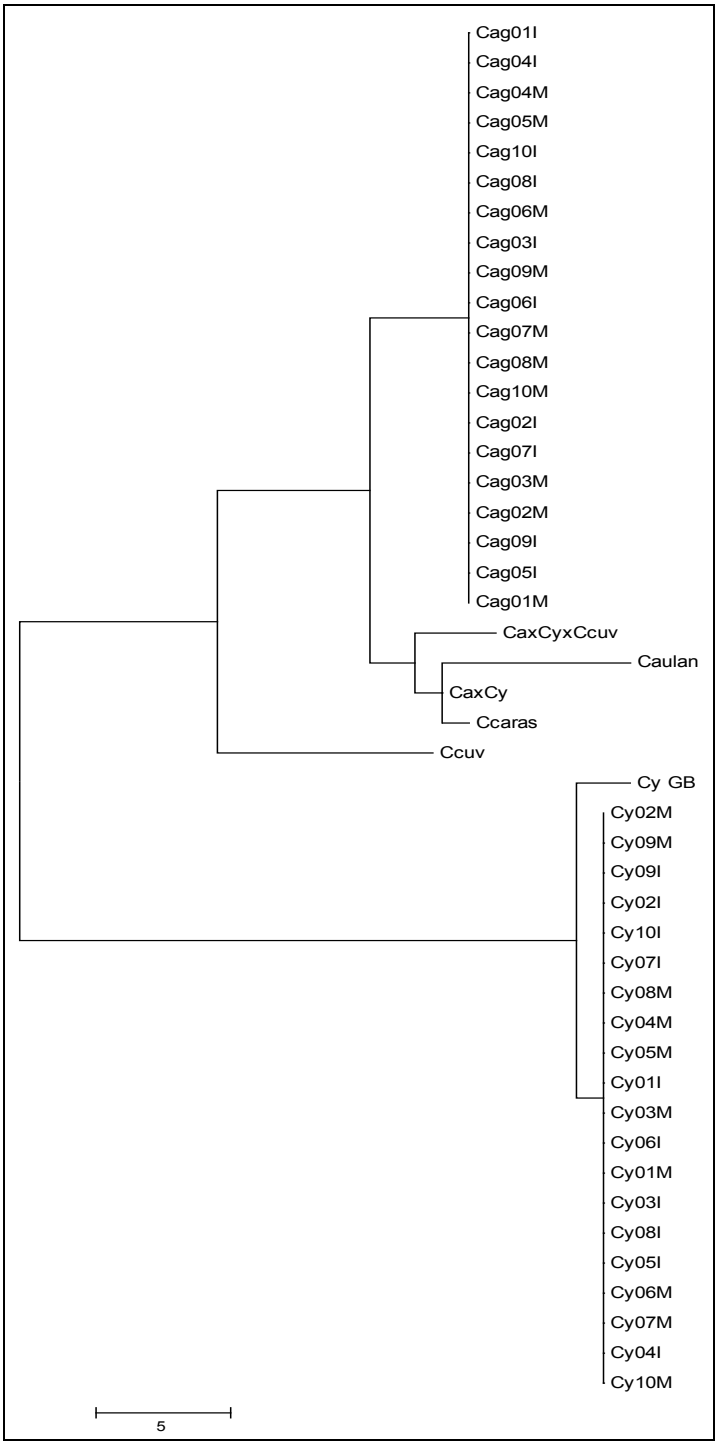


Figura 93 Arbore filogenetic realizat pe baza secvențelor ND4L, pentru cele două genuri analizate, prin metoda MP

BIBLIOGRAFIE

1. ABILOCK MARIA, STEPHENSON F. H., 2001 - M*itochondrial DNA as a Molecular Clock*, In *Mitochondrial DNA Sequencing Student Guide*. Ed. Applied Biosystems, p. 1 – 25.
2. ACKERFORS, H., 1989 – *A regional survey on the acvaculture sector in eastern and northwestern Europe* – din Aquacult. dev. and coord. progr. ADCP. REP. p. 38 – 89.
3. ALVES M. J., COELHO M. M., COLLARES-PEREIRA M. J., 1997a – *The Rutilus alburnoides complex (Cyprinidae): evidence for a hybrid origin.* J. Zool. Syst. Evol. Res., Nr. 35, p. 1 – 10.
4. ALVES M. J., COELHO M. M., COLLARES-PEREIRA M. J., DOWLING T. E., 1977b – *Maternal ancestry of the Rutilus alburnoides complex (Teleostei, Cyprinidae) as determined by analysis oh cytochrome b sequences*. Evolution Nr. 51, 1584 – 1592.
5. ANDERSON E., STEBBINS G. L., 1954 – *Hybridization as an evolutionary stimuls*, Evolution, Nr. 8, p. 378 – 388.
6. APETROAEI, MARIA, PRICOPE, F., 1998 – *Cercetări privind obținerea unor hibrizi de salmonide din genitori cu perioade de reproducere diferite* – Lucrările Simpozionului Internațional AQUAROM ′98 Galați, p. 322 – 323.
7. ARNOLD S. J., PFRENDER M. E., JONES A. G., 2001 – *The adaptative landscape as a conceptual bridge between micro- and macroevolution*, în *Genetica* nr. 112 – 113, p. 9 – 32.
8. ATCHLEY W. R., FITCH W. M., BRONNER-FRASER M., 1994 – *Molecular evolution of the MyoD family of transcription factors*, Proc. Natl. Acad Sci., USA, Nr. 91, p. 11522 – 11526. Atteson K., 1999 - *The performance of the neighbor-joining methods of phylogeneticreconstruction*, în *Algorithmica* nr. 25, p. 251 – 278.
9. AUSUBEL, F. M., BRENT, R., KINGSTON R. E., MOORE D. D., SEIDMAN J. G., SMITH J. A., STRUHL K., 1992 – *Short protocols in molecular biology*, 3-rd Edition, cap. 2 – *Preparation and analysis of DNA*, Ed. by John Wiley & Sons, Inc., p. 2-3 – 2-5.
10. AUSUBEL F. M., BRENT R., KINGSTON R. E., MOORE D. D., SEIDMAN J. G., SMITH J. A., STRUHL K., 1995.– *Current protocols in molecular biology*, vol. 1, cap. 2 – *Preparation and analysis of DNA. Phenol extraction and ethanol precipitation of DNA*, Ed by John Wiley & Sons, Inc., p. 2.1.1. – 2.1.3.
11. BADER D., MORET B., YAN M., 2001 – *A linear-time algorithm for computing in-version distance between signed permutations with an experimental study*, în J. Comput. Biol., nr. 8 (5), p. 483 – 491.
12. BAFNA V., PEVZNER P., 1995 - *Sorting permutations by transpositions*, în Proc. 6th Ann. ACM/SIAM Symp. Discrete Algs. (SODA'95). SIAM Press, Philadelphia, p. 614 – 623.

13. BARNDORFF-NIELSEN O., 1983 – *On a formula for the distribution of the maximum likelihood estimator*, în *Biometrika* Nr. 70, p. 343-365.
14. BĂNĂRESCU P., 1964 – *Fauna Republicii Populare Române – Pisces-Osteichthyes (Peşti ganoizi şi osoşi)*, vol. XIII, , Editura Academiei Republicii Populare Române, Bucureşti, p. 297-503
15. BĂNĂRESCU P., 1973 – *Some reconsiderations on the zoogeography of the Euro-mediterranean freshwater fish fauna*, Rev. Roum. Biol. Zool. Nr. 18, p. 257-264.
16. BĂNĂRESCU P., 1975 – *Orientări moderne în taxonomia şi sistematica animală*. *Probleme actuale de biologie I*, , Editura Didactică şi Pedagogică, Bucureşti, p. 5 – 20.
17. BĂNĂRESCU P., 1989 – *Zoogeography and history of the freshwater fish fauna of Europe*. In *The freshwater fishes of Europe*, vol. 1, Ed. J. Holeik, Wiesbaden: AULA, p. 88-107.
18. BĂNĂRESCU P., 1992 – *Distribution and dispersal of freshwater animals in North America and Eurasia. Zoogeography of fresh-waters: distribution and dispersal of freshwater animals*. În *North America and Eurasia*, vol. 2,. Wiesbaden: AULA, p. 519-1091.
19. BĂNĂRESCU P., Coad B. W., 1991 – *Cyprinids of Eurasia* În *Cyprinid fishes Systematics, biology and exploitation* Ed. I. J.Winfield & J. S. Nelson,. London: Chapman &Hall, p. 127-155.
20. BĂRA I. I., 1996 – *Vademecum în genetică*, Editura Corson, Iaşi, p.136 – 138.
21. BEERLI P, FELSENSTEIN J., 1999 – *Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach*, în *Genetics*, Nr. 152, p. 763 – 773.
22. BENDER M., GE D., HE S., HU H., PINTER R., SKIENA S., SWIDAN F., 2004 - *Improved bounds on sorting with length-weighted reversals*, în: Proc. 15th Ann. ACM/SIAM Symp. Discrete Algs. (SODA'04). SIAM Press, Philadelphia, p. 912 – 921.
23. BENSCH S., HASSELQUIST D., 1999 – *Phylogeographic population structure of great reed warblers: an analysis of mtDNA control region sequences*. Biol. Jour. Linn. Soc. Nr. 66, p. 171 – 185.
24. BERNATCHEZ L., 2001 – *The evolutionary history of brown trout (Salmo trutta L.) inferred from phylogeographic, nested clade, and mismatch analyses of mitochondrial DNA variation*, în *Evolution*, nr. 55(2), p. 351 – 379.
25. BERREBI P., 1995 – *Speciation of the genus Barbus in the north mediterranean basin: recent advances from biochemical genetics*. Biol. Conserv., Nr. 72, p. 237 – 249.
26. BERREBI P., KOTTELAT M., SKELTON P., RAB P., 1996 – *Systematics of barbus: state of the art and heuristic comments* Folia Zool., Nr. 45, p. 5 – 12.
27. BERREBI P., KRAIEM M. M., DOADRIO I., EL-GHARBI S., CATTANEO-BERREBI G., 1995 – *Ecological and genetic differentiation of Barbus callensis populations (Teleostei, Cyprinidae) in Tunisia*. J. Fish. Biol., Nr. 47, p. 850 – 864.
28. BIDWELL C. A., CHRISMAN C. L., LIBEZ G. S., 1985 – *Aquaculture*, 51, (1), p. 25 – 33.
29. BILLARD R., 1980 – *L`étang et l`agriculture des eaux* I.N.R.A., Publ. Paris, pag. 15 – 28.
30. BIRKY W. C., 2001 – *The inheritance of genes in mitochondria and chloroplasts: Laws, Mechanisms and models*. În Annu. Rev. Genet., Nr. 35, p. 125 – 148.

31. BLOOMER P., IMPSON N. D., 2000 – *Mitochondrial DNA differentiation in the critically endangered berg river redfin*. Journal of Heredity, Nr. 91, p. 122 – 127.
32. BOESIGER E., 1970 – *L`hybridation facteur de l`évolution*, Science, Progrès, Découverte, 3426, p. 30 – 35.
33. BOGOESCU C., DABIJA A., SANIELEVICI E., 1980 – *Atlas zoologic*, Editura Didactică şi Pedagogică, Bucureşti, p. 112 – 137.
34. BONGERS A. B. J., VELD E. P. C., ABO-HASHEMA K., BREMMER I. M., EDING E. H., KOMEN J., RICHTER C. J. J., 1994 – *Aquaculture*, 122, 2/3, p. 119 – 133.
35. BOORE J., COLLINS T., STANTON D., DAEHLER L., BROWN W., 1995 – *Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements*., *Nature* nr. 376, p. 163 – 165.
36. BOORE J., BROWN W., 1998 - *Big trees from little genomes: Mitochondrial gene orderas a phylogenetic tool*, în *Current Opinion in Genet and Development*, nr. 8 (6), p. 668 – 674.
37. BOTEZ SANDA, 1967 – *Ce sunt hibrizii?*, Editura Ştiinţifică, Bucureşti p.32 – 66.
38. BOURQUE G., PEVZNER P., 2002 - *Genome-scale evolution: reconstructing gene orders in the ancestral species*, în *Genome Research* nr. 12, p. 26 – 36.
39. BRAWLEY SUSAN , 1999 – *Submission and retrieval of an aligned set of nucleic acid sequences* J. Phycol. Nr. 35, p. 433 – 437.
40. BRUNO W. J., 2000 – *Weighted Neighbor Joining: A Fast Approximation to Maximum-Likelihood Phylogeny Reconstruction*, în Mol. Biol. Evol., nr. 17(1), p. 189-197.
41. BRUNO W. J., SOCCI N. D., HALPERN A. L., 2000 – *Weighted neighbor joining: A likelihood-based approach to distance-based phylogeny reconstruction*, în Mol. Biol. Evol., nr. 17 (1), p. 189 – 197.
42. BRYANT D., 2000 – *The complexity of calculating exemplar distances*. În: Sanko D., Nadeau J. (Eds.), *Comparative Genomics: Empirical and Analytical Approaches to Gene Order Dynamics, Map Alignment, and the Evolution of Gene Families. Kluwer Academic* Pubs. Dordrecht, Netherlands, p. 207 – 212.
43. BRYANT D., WADDELL P., 1998 – *Rapid evaluation of least-squares and minimum-evolution criteria on phylogenetic trees*, în Mol. Biol. Evol., nr. 15 (10), p. 1346 – 1359.
44. BUŞNIŢĂ TH., 1971 – *Atlasul peştilor din apele R.S.R. România*, Edit. Ceres, Bucureşti, p. 7 – 50.
45. CARAIMAN GH., 1982 – *Creşterea intensivă a puilor de crap (Cyprinus carpio) în condiţiile eutrofizării dirijate a heleşteielor în anul 1981*, Bulet. de cercetări piscicole, Anul IV (XXXV), Nucet – Dîmboviţa, Nr. 1 – 2, p. 35 – 36.
46. CARBONARI M, SBARIGIA D, CIBATI M, FIORILLI M., 1993 – *Optimization of PCR performance* Department of Clinical Immunology, University of Rome La Sapienza, Italy. Trends Genet. Nr. 9(2), p. 42 – 43.
47. CARMONA J. A., SANJUR O., DOADRIO I., MACHORDOM A., VRIJENHOEK R., 1997 – *Hybridojenetic reproduction and maternal ancestry of polyploid european fish: the Tropidophoxinellus alburnoides complex.* În *Genetics* Nr. 146, p. 983 – 993.
48. CASTRESANA J., 2000 – *Selection of conserved blocks form multiple Alignments for their use in phylogenetic analysis*. Mol. Biol. Evol. Nr. 17 (4), p. 540 – 552.
49. CAVALI-SFORZA L. L., EDWARDS A. W. F., 1964 – *Analysis of human evolution*. În *Proc. 11th Intl. Cong. Genet.*, Pergamon, New York, p. 923 – 933.

50. CAVALI-SFORZA L. L., EDWARDS A. W. F., 1967 – *Phylogenetic analysis*: Models and estimation procedures. Am. J. Hum. Genet., Nr. 19, p. 233 – 257.
51. CAVENDER T., COBURN M., 1992 – *Phylogenetic relationships of North American Cyprinidae*, În *Systematics, historical ecology, andNorth American freshwater fishes* (Editat de R. Mayden), p. 293 – 327, Stanford University Press.
52. CHANDRASEKHARAN U. M., SANKER S., GLYNIAS M. J., KARNIK S. S., HUSAIN A., 1996 – *Angiotensisn II-forming activity in a reconstructed ancestral chymase*. *Science* Nr. 271, p. 502 – 505.
53. CHANG Y. S., HUANG F. L., LO T. B., 1994 – *The complete nucleotide sequence and gene organization of carp (Cyprinus carpio) mitochondrial genome*, J. Mol. Evol., Nr. 38, p. 138 – 155.
54. CHARLESWORTH D., 1981 – *The cost of meiosis with alternation of sexual and asexual gnerations*, J. Theor., Biol., 87, p. 517 – 528.
55. CHAUDHURI P., DAS S., 2002 – *Swords: a statistical tool for analysing large DNA sequences*. J. Biosci, vol. 27, nr. 1, p. 1 – 6.
56. CHEN X. L., YUE P. Q., LIN R. D., 1984 – *Major groups within the family Cyprinidae and their phylogenetic relationships*, Acta Zootaxon. Sin.; Nr. 9, p. 424 – 440.
57. CHEN S. J., DILL K. A., 2000 – *RNA folding energy landscapes*, PNAS, Nr. 2, p. 646 – 651.
58. CHERFAS N. B., 1975 – *Investigation of radiation-induced diploid gynogenesys in the carp Cyprinus carpio L.. Experiment son obtaining the diploid gynogenetic progeny in mass quantities*. Genetika, Nr. 11 (7), p. 78 – 86.
59. CHERFAS N. B., GOMELSKY B. I., EMELYANOVA O. V., RECOUBRATSKY A. V., 1994 – *Induced diploid gynogenesis and polyploidy in crucian carp Carassius auratus gibelio (Block) x common carp Cyprinus carpio L. Hybrids*, Aquaculture Research Issue, vol. 25 nr. 9; p.943 – 955.
60. CHERFAS N. B., HULATA G., GOMELSKY B. I., BEN-DOM N., PERETZ Y., 1995 – *Chromosome set manipulation in the common carp (Cyprinus carpio)*, Aquaculture, 129 (1-4), p. 271 – 281.
61. CHERFAS N. B., HULATA G., KOZINSKY O., 1993 – *Gynogenesis in common carp IV Growth, phenotypic variation and gonad differentiation in normal and methyltestosterone – treated homozygous clones and F1 hybrids. 2. Timing of heat shock during the first cleavage*. Aquaculture, 111, (1 – 4); 271 – 281.
62. CHOU P. Y., FASMAN G. D., 1978 – *Prediction of the secondary structure of proteins from their amino acid sequence*. Adv. Enzymol., nr. 47, p. 45-148.
63. CHOURROUT, D., CHVASSNE, B., GYYOMARD, R., 1986 – *L'amelioration genetique des poissons*, din *La recherche* nr. 180, vol. 17, 1028 – 1038.
64. CLAYTON D. A., 1982 – *Replication of animal mitochondrial DNA*. În *Cell* Nr. 28, p. 693 – 705.
65. COCOLIN L., D`AGARO E., MANZANO M., LANARI D., COMI G., 2000 – *Rapid PCR-RFLP method for the identification of marine fish fillets*. În *Journal of Food Science*, vol 65, nr. 8, p. 1315 – 1317.
66. COELHO M. M., 1992 – *Genetic differentiation in the Iberian Cyprinids Chondrostoma polylepis Steinchdaner, 1885 and Ch. willkommii Steinchdaner 1866*, Arch. Hydrobiol. Nr. 125, p. 487 – 498.
67. COELHO M. M., BRITO R. M., PACHECO T. R., FIGUEIREDO D., PIRES A. M., 1995 – *Genetic variation anddivergence of L. pyrenaicus and L. caroliterlii (Pisces, Cyprinidae)*. J. Fish Biol., Nr. 47, p. 243 – 258.

68. COSNER M., JANSEN R., MORET B., RAUBESON L., WANG L., WARNOW T., WYMAN S., 2000 – *A new fast heuristic for computing the breakpoint phylogeny and experimental phylogenetic analyses of real and synthetic data*. În: Proc. 8th International Conf. on Intelligent Systems for Mol. Biol. p. 104 – 115.
69. CRISTEA M., 1991 – *Genetica ecologică şi evoluția*, Ed. Ceres, Bucureşti, p. 24 – 36.
70. DANZMANN R. G., GHARBI K., 2001 – *Gene mapping in fishes: a means to an end*, în *Genetics*, nr. 111, p. 3 – 23.
71. DAY W., 1983 – *Computationally di_cult parsimony problems in phylogenetic systematics*, în *Journal of Theoretical Biology* nr. 103, p. 429 – 438.
72. DAYHOFF M. O., SCHWARTZ R. M., ORCUTT B. C., 1978 *A model of evolutionary change in proteins*, In *Atlas of protein sequence and structure* (M. O. Dayhoff, ed.), National Biomedical Research Foundation , Silver Spring , MD., p. 345 – 352.
73. DESPER R., GASCUEL O., 2004 – *Theoretical Foundation of the Balanced Minimum Evolution Method of Phylogenetic Inference and Its Relationship to Weighted Least-Squares Tree Fitting* în Mol. Biol. Evol., vol 21, nr. 3, p. 587 - 598.
74. DILL W., 1988 – *The inland fisheries of Israel*, din *The Israeli gomn. of Aquac.*, Bamidgeh, nr. 403, p. 75 – 104.
75. DOADRIO I., 1990 – *Phylogenetic relationships and classification of west palearctic species of the genus Barbus (Osteichthyes, Cyprinidae)*. Aquat. Living Resour. Nr. 3, p. 265-282.
76. DOBZHANSKY TH., 1970 – *Genetics of the evolutionary process*, Columbia Univ. Press, New York, p. 391 – 393.
77. DOBZHANSKY TH., AYALA F. J., STEBBINS L. G., VALENTINE J. W., 1977 – *Evolution*, W. H. Ferman and Comp., San Francisco, p. 72 – 75.
78. DOBZHANSKY TH., BOESIGER E., 1968 – *Essais sur l`évolution*, Masson et CieEditeurs, Paris, p. 61.
79. ECK R. V., DAYHOFF M. O., 1966 – *Atlas of protein sequence and structure* National Biomedical Research Foundation, (abstract).
80. EDWARDS A. W. F., CAVALLI-SFORZA L. L., 1963 – *The reconstruction of evolution*. În *Heredity* Nr. 18, p. 553.
81. EFRON B., 1982a – *Maximum likelihood and decision theory*, în *Annals of Statistics*, Nr.10, p. 340-356.
82. EFRON B., 1982b – *The jackknife, the bootstrap and other resampling plans*, Society for Industrial and Applied Mathemathics, Philadelphia, PA., p. 38.
83. EFRON B., TIBSHIRANI R. J., 1993 – *An introduction to the bootstrap.* Ed. Chapman & Hall, New York, p. 193 – 274.
84. EL-MABROUK N., 2000 – *Genome rearrangement by reversals and insertions-deletions of contiguous segments*, în Proc. 11th Ann. Symp. Combin. Pattern Matching Vol. 1848 of Lecture Notes in Computer Science. Springer Verlag, p. 222 – 234.
85. EMINI E. A., HUGHES J., PERLOW D., BOGER J., 1985 – *Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide*. În J. Virology Nr. 55, p. 836 – 839.
86. FEIDER Z., GROSSU V. AL., GYURKO ŞT., POP V., 1964 – *Zoologia vertebratelor*,; Editura Didactică şi Pedagogică, Bucureşti, p.182-253.
87. FELSENSTEIN J., 1981 – *Evolutionary trees from DNA sequences: A maximum likeli-hood approach*, în J. Mol. Evol. nr. 17, p. 368 – 376.

88. FELSENSTEIN J., 1981 – *Numerical methods for inferring evolutionary trees*. Quart. Rev. Biol., Nr. 57, p. 379 – 404.
89. FINK S. V., FINK W. L., 1996 – *Interrelationships of Ostariophysan fishes*, în *Interrelationships of fishes*, Editat de M. L. J. Stiassny, L. R. Parenti şi G. D. Johnson, , San Diego, CA: Academic Press, p 209 – 250.
90. FITCH W. M., 1967 – *Evidence suggesting a non-random character to nucleotide replacements in naturally occuring mutations*. J. Mol. Biol., Nr. 26, p. 499 – 507.
91. FITCH W. M., 1971 – *Toward defining the course of evolution: Minimum change for a specific tree topology*. Syst. Zool. Nr. 20, p. 406 – 416.
92. FITCH W. M., 1977 – *On the problem of discovering the most parsimonious tree*. American Naturalist nr. 111, p. 223 – 257.
93. FOWLER J., COHEN L., JARVIS P., 2000 – *Practical statistics for field biology*, Second Edition, Ed. John Wiley & Sons Ltd., England p. 186-207.
94. FUJIKAWA N., NAKAYAMA I., ONOZATO H., 1993 – Ultraviolet-induced androgenesis in Nile tilapia, Oreochromis niloticus (L.), and hybrid Nile × blue tilapia, O. aureus (Steindachner) Bull. Nat. Res. Inst. of Aquaculture, nr. 22, p. 11 – 21.
95. GALLUT CYRIL, 1998 – *Codage de l`ordre des genes du genome mitochondrial animal en vue d]une analyse phylogenetique*. În DEA Biodiversité, p. 1 – 35.
96. GARNIER J., OSGUTHORPE D. J., ROBSON B., 1978 – *Analysis of the accuracy and implications of simple method for predicting the secondary structure of globular proteins*. Jour. Mol. Biol., nr. 120, p. 97-120.
97. GASCUEL O., 1997 – *BIONJ: An improved version of the NJ algorithm based on a simple model of sequence data*, în Mol. Biol. Evol. Nr. 14 (7), p. 685 – 695.
98. GASCUEL O., 2000 – *On the Optimization Principle in Phylogenetic Analysis and the Minimum-Evolution Criterion* în Mol. Biol. Evol., vol.17, nr.3, p. 401 - 405.
99. GERBER ANNE, LOGGINS R., KIMAR S., DOWLING T. E., 2001 – *Does nonneutral evolution shepe observed patterns of DNA variation in Animal mitochondrial genomes?*. Annu. Rev. Genet., Nr. 35, p. 539 – 566.
100. GERVAI J., MARIAN T., KRASZNAI Z., NAGY A., CSANYI V., 1980 – *Occurence of aneuploidy in radiation gynogenesis of carp*, J. Fish. Biol., 16 (4), 435 – 439.
101. GOLOVINSKAIA K. A., 1968 – *Genetics and selection of fish and artificial gynogenesis of the carp, Cyprinus carpio*. În T. V. R. Pillay, Proc. World Symposium on warm water Pond Fish Culture, Nr. 44, vol. 4, p. 215 – 222.
102. GOJOBORI T., LI W. H, GRAUR D., 1982 – *Patterns of nucleotide substitution in pseudogenes and functional genes*. J. Mol. Evol., Nr. 18, p. 360 – 369.
103. GOMELSKY B. I., CHERFAS N. B., HULATA G., PERETZ Y., BEN N., 1993 – *Chromosome set manipulation in common carp Cyprinus carpio*, Aquaculture Symposium of the carp, Budapest, Hungary, 6 – 9 sept, abstract.
104. GOODWIN R. L., BAUMANN H., BERGER F. G., 1996 – *Patterns of divergence during evolution of α1-proteinase inhibitors in mammals.* Mol Biol Evol. Nr. 13, p 346 – 358.
105. GORGAN D. L., APETROAEI MARIA, BĂRA I. I., 2000 - *Variabilitatea fenotipică la specii ale genului Salmo*, în „*Genetică şi Evoluționism*”, vol. 4, p.99 – 110, Editura Corson, Iaşi.
106. GORGAN D. L., BĂRA I. I., 2002 – *The content in total nucleic acid at Cyprinus carpio L. and Carassius carassius L. species*, An. Şt. ale Univ. „Al. I. Cuza”, secțiunea II, a. Genetică şi Biologie Moleculară, Tom III, p. 62 – 65, Iaşi.

107. GORGAN D. L., BĂRA I. I., 2003 – *The nucleic acid content at Cyprinus carpio L. and Carassius auratus gibelio Bloch.*, An. Şt. ale Univ. „Al. I. Cuza”, Genetică şi Biologie Moleculară, Tom IV, p. 119 – 123, Iaşi.
108. GORGAN D. L., 2004a – *Stabilirea secvenței nucleotidice a citocromului b la Carassius auratus gibelio* Bloch., *Acta Ihtyologica* Nr.1, sub tipar.
109. GORGAN D. L., 2004b – *Stabilirea secvenței nucleotidice a ND4L la Carassius auratus gibelio* Bloch. *şi Cyprinus carpio* L., în *Studii şi cercetări*, Biologie, Serie nouă, vol. IX, Universitatea din Bacău, p. 103 – 106.
110. GORGAN D. L., CÎMPEANU S. C., OLTEANU ZENOVIA, 2004c – *Determinarea cantității totale de ADN la Carassius auratus gibelio Bloch*, sub tipar.
111. GORGAN D. L., CÎMPEANU S. C., GHERASIM SORINA RALUCA, 2004d – *Comparative researches about total DNA quantity from three different organs from Carassius auratus gibelio Bloch and Cyprinus carpio L.*, sub tipar.
112. GORGAN D. L., SANJUAN L. A., COMENSAÑA S. A., BĂRA I. I., 2004e – *The D-loop sequence determination from Carassius auratus gibelio Bloch. species*, An. Şt. ale Univ. „Al. I. Cuza”, Genetică şi Biologie Moleculară, Tom V, p. 103 – 106, Iaşi.
113. GORGAN D. L., CÎMPEANU MIRELA MIHAELA, 2004f – *ND4L nucleotide sequence compare for different species of Carassius genera*, sub tipar.
114. GRANT V., 1977 – *Organismic evolution*, W. H. Ferman and Comp., San Francisco, p. 418.
115. GUINDON S., GASCUEL O., 2003 – *PHYML: a simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood*. În Syst. Biol. Nr. 52 (5), p. 696 – 704.
116. GUSTAFSSON J. P., STEBBINS G. L., AYALA F. J., 1986 – *Genetics, Development and evolution*, Plenum Press, New York (abstract).
117. HALL T. A., 1999 – *BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/ NT*. În Nucleic Acids Symp. Ser. Nr. 41, p.95–98.
118. HARTIGAN J. A., 1973 – *Minimum evolution fits to a given tree*. Biometrics Nr. 29, p. 53 – 65.
119. HARTMAN T., 2003 – *A simpler 1.5-approximation algorithm for sorting by transpositions*. În: Proc. 14th Ann. Symp. Combin. Pattern Matching, Vol. 2676 of Lecture Notes in Computer Science. Springer Verlag, p. 156 – 169.
120. HARTMAN T., SHARAN R., 2004 – *A 1.5-approximation algorithm for sorting by transpositions and transreversals*. În: Proc. 4th Internationall Workshop Algs. in Bioinformatics. Vol. 3240 of Lecture Notes in Computer Science. Springer Verlag, p. 50 – 61.
121. HASEGAWA M., KISHINO H., YANO T., 1985 – *dating of the human-ape splitting by a molecular clock of mitochondrial DNA*. J. Mol. Evol., nr. 22, p. 160 – 174.
122. HEARD S., 1996 – *Patterns in phylogenetic tree balance with variable and evolvingspeciation rates*. Evol. Nr. 50, p. 2141 – 2148.
123. HENDY M. D., PENNY D., 1989 – *A framework for the quantitative study of evolutionary trees*, Sysyt. Zool. Nr. 38, p. 297 – 309.
124. HICKERSON M. J., ROSS J. R. P., 2001 – *postglacial population history and genetic structure of the northern clingfish (Gobbiesox maeandricus), revealed from mtDNA analysis*. În Marine Biology nr. 138, p. 407 – 419.
125. HIGGINS D. G., SHARP P.M., (1989). *Fast and sensitive multiple sequence alignments on a microcomputer.* CABIOS, Vol. 5, Nr. 2:, p. 151-153.

126. HIGGINS D. G., 1994 – *Clustal V: Multiple alignment of DNA and protein sequences*. În *Methods Mol. Biol.*, Nr. 25, p. 307 – 318.
127. HILLIS D. M., LARSON A., DAVIS S. K., ZIMMER ELIZABETH, 1996 - *Nucleic Acids III: Sequencing*, în *Molecular systematics*, 2nd ed., Sinauer Assoc., Sunderland MA, Chapter 9, p. 318 – 370.
128. HOLLEBECQ M. G., CHOURROUT D., WOHLFARTH G., BILLARD R., 1986 – *Diploid gynogenesis induced by heat shock after fertilization with the UV-irradiated sperm of common carp*, Aquaculture, 54, p. 69 – 76.
129. HORVATH L., ORBAN L., 1995 – *Genome and gene manipulation in common carp*, Aquaculture, 129, (1-4), p. 157 – 183.
130. HOULE D., MEZEY J., GALPERN P., 2002 – *Interpretation of the results of common principal components analysis*, în *Evolution*, nr. 56(3), p. 433 – 440.
131. HOWES G. J. 1991 - Systematics and biogeography: an overview In Cyprinid fishes: Systematics, biology and exploitation, Ed. I. J.Winfield & J. S. Nelson,. 1991. London: Chapman &Hall, p. 1-33.
132. HULATA G., 2001 - *Genetic manipulations in aquaculture: a review of stock improvement by classical and modern technologies*, în *Genetica*, nr. 111, p. 155 – 173.
133. HUSON, D., SMITH, K., WARNOW, T., 1999 – *Correcting large distances for phylogenetic reconstruction*. În Proc. 3rd International Workshop Alg. Engineering. Vol. 1668 of Lecture Notes in Computer Science. Springer Verlag, p. 273 – 286.
134. HUSSAIN M. G., CHATTERGI A., Mc ANDREW B. J., JOHNSTONE R., 1991 – *Triploidy induction in Nile tilapia,* Oreochromis niloticus L. *using pressure, heat and cold shocks* în *Theory of Applicable Genetics*, 81, p. 6 – 12.
135. HUXLEY J., 1974 – *Evolution – the modern synthesis*; Third Edition, Baker, Allen et Unwin Ltd, p. 565 – 566.
136. IGUCHI K., YAMAMOTO G., MATSUBARA N., NISHIDA M., 2003 – *Morphological and genetic analysis of fish of a Carassius complex (Cyprinidae) in Lake Kasumigaua with reference to the taxonomic status of two all-female triploid morphs*. Biol. J. Linn. Soc., London, Nr. 79, p. 351 – 357.
137. IJIRI K., EGAMI N., 1980 – *Hertwig effect caused by UV-irradiation of sperm of Oryazis latipes (Teleost) and its photoreactivation*, Mut. Res., 69, p. 241 – 248.
138. IMSIRIDOU A., ZALDIVAR J. M., 1999 – *Methodology and formats for genetic identification of fish species*. Joint Research Centre, p. 1 – 36.
139. INGRAM V. M., 1963 – *The hemoglobins in genetics and evolution*. Columbia University Press, New York (abstract).
140. JAMESON B. A., WOLF H., 1988 – *The antigenic index: a novel algorithm for predicting antigenic determinants*. CABIOS vol. 4, p. 181 – 186.
141. KARAKOUSIS Z., MACHORDOM A., DOADRIO I., ECONOMIDIS P.S., 1995 – *Phylogenetic relationships of Barbus peloponensis Valenciennes, 1842 (Osteichthyes: Cyprinidae from Greece and other speciei of barbus as revealed by allozyme electrophoresis*. Biochem, System. Ecol., Nr. 23, p. 39 – 45.
142. KARPLUS P. A., SCHULTZ G. E., 1985 – *Prediction of chain flexibility in proteins*. În *Naturwissenschaften*, nr. 72, p. 212 – 213.
143. KAVUMPURATH S., PANDIAN T. J., 1994 – *Induction of heterozygous and homozygous diploid gynogenesis in Betta splendens (Regan) using hydrostatic pressure*, Aquaculture and Fisheries Management, Aquaculture Research Issue, Nr. 25, (1), p. 133 – 143.

144. KIDD K. K., SGARAMELLA-ZONTA L. A., 1971 – *Phylogenetic analysis: Concepts and methods*. Am. J. Hum. Genet. Nr. 23, p. 235 – 252.
145. KIMURA M., OHTA T., 1972 – *On the stochastic model for estimation of mutational distancebetween homologous proteins*, J. Mol. Evol. Nr. 2, p. 87 – 90.
146. KIRPICHNIKOV V. S., 1973 – *On Karyotype Evolution in Cyclostomata and Pisces*, Ichthyologia, Nr. 5/1, p. 55 – 77.
147. KIRPICINIKOV V. S., 1987 – *Genetische Grunglagen der Fischzuchtung*, Veb. Deut. Land. Berlin (abstract).
148. KOBAYASI H., 1971 – *A cytological study on gynogeneis of the triploid ginbuna (Carassius auratus langdorfi)*, Zool. Mag., 80, 9, p. 316 – 322.
149. KOBAYASI H., NAKANO K., NAKAMURA M., 1977 – *On the Hybrids, 4n Ginbuna (Carassius auratus langsdorfii) x Kinbuna (Carassius auratus), and their chromosomes*. Bulletin of the Japanese Society of Scientific Fisheries, Nr. 43/1, p. 31 – 37.
150. KOBAYASI H., OCHI H., TAKEUCHI N., 1973 – *Chromosome studies in the genus Carassius: Comparison of C. auratus grandoculis, C. auratus buergeri and Carassius auratus langsdorfii*. Japanese Journal of Ichthyology, Nr. 20 (1), p. 83 – 88.
151. KOCHER T. D., THOMAS W. K., MEYER A., EDWARDS S. V., PAABO S., 1989, *Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers*,.. Proc. Nati., Acad. Sci. USA, Nr. 86, p. 6196 – 6200.
152. KOCHER T. D., WILSON A. C., 1991 – *Sequence evolution of mitocondrial DNA in humans and chimpanzees: Control region and protein coding region*. In *Evolution of life*,. Springer-Verlag, New York. p. 391 – 413.
153. KOMEN J., DUYNHOVER J., RICHTER C. J. J., HUSMAN E. A., 1989 – *Gynogenesis in common carp (Cyprinus carpio L.). I. Effects of genetic manipulation of sexual products and incubation of eggs*, Aquaculture, 69, p. 227 – 239.
154. KREISER B. R., MITTON J. B., WOODLING J. D., 2001 – *Phylogeography of the plains killifish Fundulus Zebrinus*, în *Evolution*, nr. 55(2), p. 339 – 350.
155. KUMAR S., 1996 – *A stepwise algorithm for finding minimum evolution trees*, în Mol. Biol. Evol., nr. 13 (4), p. 584 – 593.
156. KUMAR S., TAMURA K., JAKOBSEN I.B., NEI M., 2001 – *MEGA2: molecular evolutionary genetics analysis software.* În *Bioinformatics*, Nr. 17, p. 1244– 1245.
157. KUMAR S., GADAGKAR S. R., 2000 – *Efficiency of the neighbor-joining method in reconstructing deep and shallow evolutionary relationships in large phylogenies*. J. Mol. Evol. Nr. 51, p 544 – 553.
158. KYTE J., DOOLITTLE R. F., 1982 – *A simple method for displaying the hydropathic character of a protein*. Jour. Mol. Biol., nr. 157, p. 105 – 132.
159. LARGET B., SIMON D., KADANE J., 2002 – *Bayesian phylogenetic inference from animal mitochondrial genome arrangements*. În J. Royal Stat. Soc. B Nr. 64 (4), p. 681 – 694.
160. LARHAMMAR D., RISINGER C., 1994 – *Molecular genetic aspects of tetraploidy in the common carp Cyprinus carpio*. Mol. Phylogenet. Evol. Nr. 3, p. 59 – 68.
161. LEFEBVRE J. F., EL-MABROUK N., TILLIER E., SANKO D., 2003 - *Detection and validation of single gene inversions*. În: Proc. 11th International Conf. on

Intelligent Systems for Mol. Biol., vol. 19 of Bioinformatics. Oxford U. Press, p. i190 – I196.
162. LEVIN D. A., 1979 – *Hybridization – an evolutionary perspective*, Dowden, Hutchinson & Rose, Inc. Stroudsburg, Pennsylvania (abstract).
163. LEWIN B., 1987 – *Genes III (Third Edition)*. John Wiley & Sons, New York, Chichester, Brisbane, Toronto, Singapore, 226 – 237.
164. LEWIS J., WOLPERT L., 1979 – *Diploidy, evolution and sex*, J. Theor. Biol., 78, p.425 – 438.
165. LEWONTIN R. C., 1978 – *Adaptation*, American Scientist, 239, p. 157 – 169.
166. LINDER C., MORET B., NAKHLEH L., WARNOW T., 2004 – *Network (reticulated) evolution: Biology, models, and algorithms*. Prezentare disponibilă la: compbio.unm.edu/papers.html.
167. LINHART O., KVASNICKA P., SLECHTOVA V., POKORNY J., 1986 – *Induced gynogenesis by retention of the second polar body in tha common carp, Cyprinus carpio L., and heterozigosity of gynogenetic progeny in trensferrin and Ldh-B1 loci*, Aquaculture, 54, p. 63 – 67.
168. LIOR D., BLUM S., FELDMAN W. M., LAVI U., HILLEL JOSSI, 2003 – *Recent duplication of the common carp (Cyprinus carpio L.) genome as revealed by analyses of microsatellite loci*. Mol. Biol. Evol. Nr. 20 (9), p. 1425 – 1424.
169. LISKER R., ARMENDARES S., 2000 – *Introducción a la Genética Humana*, primera edición Editorial El Manual Moderno México D.F., p. 1 – 20.
170. LIU T., MORET B., BADER D., 2003 – *An exact, linear-time algorithm for computing genomic distances under inversions and deletions*. Tech. Rep., Univ. of New Mexico, p. 31.
171. LUDWIG A., MAY B., DEBUS L., JENNECKENS I., 2000 – *Heteroplasmy in the mtDNA control region of sturgeon*, în *Genetics*, nr. 156 p. 1933 – 1947.
172. LUKOWICZ, M., 1980 – *Intensitätistufen der modernen Aquakultur*, din *Intens. Fisch*, Moglichkeitem und Probleme der modernen Aqukultur, p. 56.
173. LUKOWICZ, M., 1985 – *Fischzucht in China*, din *Fisch. m. Teichwirt,* 3, (abstract).
174. LUSTUN V., VOICAN V., RĂDULESCU I., 1978 – *Dicționar piscicol*, Editura Ceres, Bucureşti., p. 63.
175. MADDISON W., 1990 – *A method for testing the correlated evolution of two binary characters: are gains or losses concentrated on certain branches of a phylogenetic tree?*. Evol. Nr. 44, 539 – 557.
176. MADDISON W., 1997 – *Gene trees in species trees*. Syst. Biol. Nr. 46, p. 523 – 536.
177. MADDISON W., MADDISON D. R., 1992 – *MacClade: Analysis of phylogeny and character evolution*. Sinauer Associates, Sunderland, (abstract).
178. MAKEEVA A. P., VERIGIN B. V., 1974 – *Gibrid karpa sz belüm tolsztolobikom*. Genetika, T. 10, Nr. 2, p. 73.
179. MARGOLIASH E., 1963 – *Primary structure and evolution of cytochrome c*. Proc. Natl., Acad. Sci. USA, Nr. 50, p. 672 – 679.
180. MÁRIÁN T., KRASZNAI Z., 1978 – *Comparative karyological investigations on chinese carps*, International seminary on *Increasing the productivity of fishes by selection and hybridisation*, Szarvas, p. 79 – 97.
181. MARRON M., SWENSON K., MORET B., 2003 – *Genomic distances under deletions and insertions. In: Proc. 9th Int'l Conf. Computing and Combinatorics*. vol. 2697 Lecture Notes in Computer Science, Springer Verlag, p. 537 – 547.

182. MARTIN A., 2001 – *The phylogenetic placement of Chondrichtyes: inferences from analysis of multiple genes and implications for comparative studies*, în *Genetica*, nr. 111, p. 349 – 357.
183. MAS EVA, POZA J.; CIRIZA JESÚS 2001 – *Fundamento de la reacción en cadena de la Polimerasa*. Revista Acuatic No. 15, Universidad de Zaragoza, España, p. 1 – 20.
184. MASUDA, H., K. AMAOKA, C. ARAGA, T. UYENO AND T. YOSHINO, 1984 *The fishes of the Japanese Archipelago* Vol. 1.. Tokai University Press, Tokyo, Japan. p. 57.
185. MATEI D., 1990 – *600 de ani de atestare a pisciculturii în Moldova (sec. XIV – XX) şi 35 de ani de activitate ştiinţifică în sprijinul producţiei piscicole 1954 – 1989)*, Piscicultura Moldovei, Lucrările S.C.P. Piscicolă, Iaşi, vol.1, p. 15 – 17.
186. MATEI, D., 1985 – *Tehnologii de valorificare a potenţialului trofic şi piscicol al lacurilor de acumulare*, din Consf. pt. probl. pisc. şi mec. ap. din sect. pisc., p. 23 – 30.
187. MATHER K., JINKS J. L., 1971 – *Biometrical Genetics*, Chapman and Hall Ltd, New Fetter Lane, London (abstract).
188. MATHEWS CHRISTOPHER 1999 – *Bioquímica* segunda edición MacGraw Hill España p. 94, 101-102, 114.
189. MATSUO T., OGAWA Y., KUMAMARU A., OCHI K., ADACHI Y., 2000 – *Complete nucleotide sequence of the Cytochrome b gene of channel catfish Ictalurus punctatus and comparison of sequence homology among channel catfish and other fishes*, Mol. Biol.p 207 – 210.
190. MAXAM A. M., GILBERT W., 1977 – *A new method for sequencing DNA.* Proc. Natl. Acad. Sci. USA, Nr. 74, p. 560 – 564.
191. MAXIMILIAN C., IOAN DOINA MARIA, 1984 – *Dicţionar de genetică*, Editura Ştiinţifică şi Enciclopedică, Bucureşti, p. 27.
192. MELTON T., 2002 – *Mitochondrial DNA*, Mitotyping Technologies, p. 1 – 4.
193. MILLER-TAILLON PATRICIA, PIERNOT ELLEN, KWOK P. Y., 1999 – *Efficient approach to unique single-nucleotide polymorphism discovery*. Genom Research Nr. 9, p. 499 – 505.
194. MINCIU DOINA-RODICA, BĂRA I. I., 1985 – *Studiul complementului cromosomial la Pseudorasbora parva (Schlegel) (Pisces, Cyprinidae)*, Sem. Ştiinţ. *Progrese în ameliorarea şi tehnologia creşterii animalelor în sistem intensiv*, Timişoara (abstract).
195. MINCIU, DOINA – RODICA, 1991 – *Posibilităţi de corelare a unor proteine polimorfe cu anumiţi indici morfoproductivi la peşti*, Ecologia et Aquacultura Limnica, Piatra Neamţ, Nr. 2(11), p. 195 – 200.
196. MIROL PATRCIA, GARCIA PERAL P., DULOUT N. F., 2004 – *Mitochondrial variability in the D-loop of four equine breeds shown by PCR-SSCP analysis*, Ed. Sociedade Brasileira de Genénetica, p. 1 – 7.
197. MOAV R., HULATA G., WOHLFARTH G., 1975 – *Genetic differences between the Chinese and European races of the common carp. I: Analysis of genotype-enviroment interactions for growth rate*. În *Heredity* Nr. 34, p. 323–340.
198. MONOD J., FRANCOIS J., 1996, *Biología Molecular* décimo segunda edición Editorial Ciencia y Desarrollo México D.F. p. 9 – 10.
199. MORET B., BADER D., WARNOW T., 2002 – *High-performance algorithm engineering for computational phylogenetics*. În J. Supercomputing, nr. 22, p. 99 - 111.

200. MORET B., ROSHAN U., WARNOW T., 2002 – *Sequence length requirements for phylogenetic methods*. În Proc. 2nd International Workshop Algs. in Bioinformatics., vol. 2452 Lecture Notes in Computer Science, Springer Verlag, p. 343 – 356.
201. MORET B., TANG J., WANG L. S., WARNOW T., 2002 – *Steps toward accurate reconstructions of phylogenies from gene-order data*. J. Comput. Syst. Sci. Nr. 65 (3), p. 508 – 525.
202. MORET B., TANG J., WARNOW T., 2005 – *Reconstructing phylogenies from gene content and gene-order data*. În: Gascuel, O. (Ed.), *Mathematics of Evolution and Phylogeny*. Oxford University Press, p. 321 – 352.
203. MÜLLER S., 1989 – *Geschichte der Fiszucht in Europa*, Teil VII. Einige wichtige Etappen in der Entwicklung der Karpfenteichwirtschaft, Z. Binnenfisch D.D.R., Berlin, Nr. 36, p. 10 – 14.
204. MURAKAMI M., MATSUBA C., FUJITANI H., 2001 – *The maternal origins of the triploid ginbuna (Carassius auratus langsdörfi); phylogenetic relationships within the Carassius auratus taxa by partial mitochondrial D-loop sequnecing*. Genes Genet. Syst. Nr. 76 (1), p. 25 – 32 (abstarct).
205. NAGY A., RAJKI K., HORVATH L., CSANYI V., 1978 – *Investigation on carp (Cyprinus carpio L.) gynogenesis*. J. Fish. Biol., Nr. 13, p. 215 – 224 (abstract).
206. NAGY A., CSANYI V., BAKOS J., BERCSENYI, M., 1984 – *Utilization of gynogenesis and sex-reversal in commercial carp breeding: growth of the first gynogenetic hybrids.* Aquacult. Hung., Nr. 4, p. 7 – 16.
207. NAKHLEH L., ROSHAN U., ST. JOHN K., SUN J., WARNOW T., 2001 – *Designing fast converging phylogenetic methods*. În: Proc. 9th International Conf. on Intelligent Systems for Mol. Biol., vol. 17 of Bioinformatics. Oxford U. Press, p. S190 – S198.
208. NAKHLEH L., MORET B., ROSHAN U., JOHN K. S., WARNOW T., 2002 – *The accuracy of fast phylogenetic methods for large datasets*. In: Proc. 7th Pacific Symp. On Biocomputing. World Scientific Pub., p. 211 – 222.
209. NEE S., 2001 – *Inferring speciation rates from phylogenies*, în *Evolution*, nr. 55(4), p. 661 – 668.
210. NEI M., 1986 – *Stochastic errors in DNA evolution and molecular phylogeny*. În *Evolutionary perspectives and the new genetics*. p. 133 - 147 Ed. Alan R. Liss, New York.
211. NEI M., 1991 – *Relative efficiencies of different tree making methods for molecular data*. În *Recent advances in phylogenetic studies of Dna sequences*. Oxford University Press, p. 133 – 147.
212. NEI M., 1996 – *Phylogenetic analysis in molecular evoutionary genetics*. Annu. Rev. Genet., Nr. 30, p. 371 – 403.
213. NEI M., TAJIMA F., 1983 – *Maximum likelihood estimation of the number of nucleotide sbstitutions from restriction sites data*. În *Genetics*, Nr. 105, p. 207 – 217.
214. NEI M., TAJIMA F., TATENO Y., 1983 – *Accuracy of estimated phylogenetic trees from molecular data II. Genes frequency data*. J. Mol. Evol. Nr. 19, p. 153 – 170.
215. NEI M., MILLER J.C, 1990 – *A simple method for estimating average number of nucleotide substitutions within and between populations from restriction data*. În *Genetics*, nr. 125, p. 873-879.

216. NEI M., GU X., SITNIKOVA T., 1997 – *Evolution by the birth-and-death process in multigene families of the vertebrate immune system.* Proc. Natl. Acad Sci. USA, Nr. 94, p. 7799 – 7806.
217. NEI M., KUMAR S., TAKAHASHI K., 1998 – *The optimization principle in phylogenetic analysis tends to give incorrect topologies when the number of nucleotides or amino acids used is small*, în PNAS, nr. 95 (21) p. 12390 - 12397.
218. NEI M., KUMAR S., 2000 – *Molecular evolution and phylogenetics*, Oxford University Press.
219. NELSON G. J., 1994 – *Fishes of the world*, Editat de Whilez, New York.
220. NICOLAU AURELIA, BREZEANU GH., CALOIANU I., BUŞNIȚA A., 1973 – *Reproducerea artificială şi dezvoltarea la peşti*, Editura Academiei R.S.R., Bucureşti (abstract).
221. NICOLESCU CARMEN, 2002 – *Types of gynogenetic carps obtained using genetic manipulated gametes*, An. Şt. ale Univ. „Al. I. Cuza", secțiunea II Genetică şi Biologie Moleculară, Tom III, p.68 – 77, Iaşi.
222. NICOLESCU CARMEN, 1994 – *Das Studium der Wirksamkeit einiger Methoden in der genetischen Inaktivierung der Karpfengameten*, Analele Universității „Valachia" Târgovişte, anul II, fascicola I, p. 101 – 111.
223. NICOLESCU CARMEN, 1996 – *Studiul genetic al unor forme de peşti obținute prin manipulări genetice (ginogeneză, androgeneză, poliploidizare, hibridare interspecifică)* – teză de doctorat (rezumat).
224. NICOLESCU CARMEN, 1997 – *Studiul eficienței unor tehnici în diploidizarea meterialului genetic al gameților la crap*, „Al XXII-lea Congres al Academiei Româno – Americane de Ştiințe şi Arte, 26 – 27 iunie 1997", vol. I, secțiunea V – Ştiințe medicale şi biologie, p. 146 – 163.
225. NIXON C K., CARPENTER J. M., 2000 – *„On the other phylogenetic systematics*, în *Cladistics* Nr. 16, p. 298 – 318.
226. NOGUSA S., 1960 – *A comparative study of the chromosomes in fishes with particular considerations on taxonomy and evolution*, Mem. Univ. Agr., nr. 3., Hyogo p. 1 – 62.
227. NORMAN A. J, PORTER A. H., 2001 – *Toward a new synthesis: population genetics and evolutionary developmental biology*, în *Genetica*, nr. 112 – 113, p. 45 – 58.
228. NUȚĂ GH., BUŞNEAG C., 1977 – *Investigații biochimice*, Editura Didactică şi Pedagogică, Bucureşti. p. 188 – 189.
229. OBRHELOVA, N. P., 1971 Vergleinchende osteologie der gattung *Leuciscus* (*Pisces*) aus tertiaren schichten der nordlichen und westlichen CSSR, Paleontolog. Abhandl. Nr. 4, p. 549-660.
230. OHNO S., 1970 – *Evolution by gene duplication*. Springer – Verlag, Berlin (abstract).
231. OJIMA Y., HAYASHI M., UENO K., 1972 – *Cytogenetic studies in lower vertebrates – Karyotype and DNA studies in 15 Species of Japanese Cyprinidae*, Genetics, 476, p. 431 – 440, Japan.
232. OLSEN G., MATSUDA H., HAGSTROM R., OVERBEEK R., 1994 – *Fast DNAmt: A tool for construction of phylogenetic trees of DNA sequences using maximum likelihood*. Computations in Applied Biosciences, nr. 10 (1), p. 41 – 48.
233. OPRESCU S., 1980 – *Introducere în citogenetica animală*, Ed. Ştiințifică, Bucureşti.

234. ORNDUFF R., 1979 – *Reproductive biology in relation to sistematics*; Taxon, 18, p. 121 – 133.
235. OWCZARZY R., VALLONE M. P., GALLO J. F., PANER T. M., LANE M. J., BENIGHT A. S., 1998 – *Predicting sequence-dependent melting stability of the short duplex DNA oligomers*. Ed. John Wiley & Sons, p. 217 – 239.
236. PÄÄBO S., 1990 – *amplifying ancient DNA* In *PCR protocols: A guide to methods and applications*, Eds. M. A. Tunes, D.H. Gelfand, J. J. Sninsky şi T. J. White, Academic Press San Diego, p. 159 – 166.
237. PAAVER T., 1983 – *Biochimiceskaia genetika karpa Cyprinus carpio L.*, Akad. nauk Estonsk. S. S. R. Inst. Zool. Bot., p. 85 – 92.
238. PAGE R., 1998 – *Introduction to tree building*, DEEB IBLS, University of Glasgow, p. 1 – 5.
239. PAGE R., 1998 – G*eneTree: comparing gene and species phylogenies using reconciled trees*. Bioinf. Nr. 14 (9), p. 819 – 820.
240. PAGE R., CHARLESTON M., 1997 – *From gene to organismal phylogeny: Reconciled trees and the gene tree/species tree problem.* Mol. Phyl. Evol. Nr. 7, p. 231 – 240.
241. PAGE R., CHARLESTON M., 1997 – *Reconciled trees and incongruent gene and species trees*. În: Mirkin, B., McMorris, F. R., Roberts, F. S., Rzehtsky, A. (Eds.), Mathematical Hierarchies in Biology, vol. 37. American Math. Soc., p. 57 – 70.
242. PALUMBI S., 1996 – *Nucleic Acids II: The polymerase chain reaction*, în *Molecular systematics*, 2nd ed., Sinauer Assoc., Sunderland MA, Chapter 7, p. 205 – 317.
243. PALUMBI S. R., CIPRIANO F., HARE M. P., 2001 – *Predicting nuclear gene coalescence from mitochondrial data: the three-times rule*, în *Evolution*, nr. 55(5), p. 859 – 868.
244. PAMILO P., NEI M., 1998 – *Relationship between gene trees and species trees*. Mol. Biol. Evol. Nr. 5, p. 568 – 583.
245. PETRE A., NEGRUȚIU E., 1975 – *Genetică animală*, Editura Didactică şi Pedagogică, Bucureşti, p. 237 – 260.
246. PIONTKIVSKA H., ROONEY A. P., NEI M., 2002 – *Purifying selection and birth-and-death evolution in the histone H4 gene family*. În *Mol. Biol. Evol.* Nr. 19, p.689-697.
247. POJOGA I., 1977 – *Piscicultura modernă în apele interioare*, Editura Ceres, Bucureşti, 34 – 58.
248. POJOGA I., 1977 – *Piscicultura*, Editura Ceres, Bucureşti, p. 15 – 33.
249. POJOGA I., NEGRIU R., 1988 – *Piscicultura practică*, Editura Ceres, Bucureşti, p. 67.
250. POSADA D., CRANDALL KEITH, 1998 – *Modeltest: testing the model of DNA substitution*, în *Bioinformatics applications note*, vol. 14, Nr. 9, p. 817 – 818.
251. PRESCOTT L. M., Harley Johon P. 1999 – *Microbiología* cuarta edición Editorial MacGraw Hill Madrid España p. 320-344.
252. QUILLET E., 1994 – *Survival growth and reproductive traits in mitotic gynogenetic rainbow trout females*, Aquaculture, 123, p. 223 – 237.
253. QUINTEIRO J., VIDAL R., MENDEY M. R., 2000 – *Phylogeny and biogeographic history of hake (genus Merluccius), inferred from mitochondrial DNA control region sequences.* În Marine Biology, nr. 136, p. 163 – 174.
254. RAICU P, 1980 – *Genetica*, , Editura Didactică şi Pedagogică, Bucureşti, p. 533 – 553.

255. RAICU P., TAISESCU ELENA, CRISTIAN ALEXANDRINA, 1972 – *Diploid chromosome complement of the carp*, în *Cytologia*, p. 37.
256. RAM J. L., RAM M. L., BALDOUN F. F., 1996 – *Identification of canned tuna and bonito by sequence and restriction site analysis of polymerase chain rection products of mitochondrial DNA*. J. Agric. Food. Chem., Nr. 44, p. 2460 – 2467.
257. RANNALA B., YANG Z., 1996 – *Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference*. J. Mol. Evol. Nr. 43, p. 304 – 311.
258. RJABOV I. N., 1975 – *Oszobennoszti embrional, no-licsinocsnogo razvitija gibridov karpa (Cyprinus carpio L.) i koreszkoj vosztrobrjuski Hemiculter eignmanni*, Jordan et Metz., Vopr. Ihtiol., T. 15 vüp. 3, (abstract).
259. ROKAS A., HOLLAND P. W. H., 2000 – *Rare genomic changes as a tool for phylogenetics*. Trends in Ecol. and Evol. Nr. 15, p. 454 – 459.
260. RÖGL, F., STEININGER, F. F., 1983 Vom Zarfall der Tethys zu Mediterran und Paratethys. Die neogene palaösgegraphie und palinspastik des zirkum-mediterranen Raumes. 85A, Ann. Naturhist. Mus.Wien. p. 135-163.
261. RYMAN N., JORDE P. E., 2001 – *Statistical power when testing for genetic differentiation*. Mol. Ecol. Nr. 10, p. 2361 – 2373.
262. RZHETSKY A., NEI M., 1992 – *A simple method for estimating and testing minimum –evolution trees*. Mol. Biol. Evol. Nr. 9, p. 945 – 967.
263. RZHETSKY A., NEI M., 1992 – *Statistical properties of the ordinary least squares, generalized least-squares, and minimum–evolution methods of phylogenetic inference*. J. Mol. Evol. Nr. 35, p. 367 – 375.
264. RZHETSKY A., NEI M., 1993 – *Theoretical foundation of the minimum-evolution method of phylogenetic inference*. Mol. Biol. Evol. Nr. 10, p. 1073 – 1095.
265. RZHETSKY A., NEI M., 1995 – *Tests of applicability of several substitution models for DNA sequence data.* Mol. Biol. Evol., Nr. 12, p. 131 – 151.
266. SACCONE C., ATTIMONELLI M., SBISA E., 1987 – *Structural elements highly preserved during the evolution of the D-loop-containing region in vertebrate mitochondrial DNA*. J. Mol. Evol. Nr. 26, p. 205 – 211.
267. SAITOU N., IMANISHI M., 1989 – *Relative efficiencies of the Fitch-Margoliash, maximum-parsimony, maximum-likelihood, minimum-evolution, and neighbor-joining methods of phylogenetic reconstructions in obtaining the correct tree.* Mol. Biol. Evol. Nr. 6, p. 514-525.
268. SAITOU N., NEI M., 1987 – *The neighbor-joining method: A new method for reconstructing phylogenetic trees.* Mol. Biol. Evol. Nr. 4, p. 406 – 425.
269. SAMONTE I. E., PAGULAYAN R. C., MAYAER W. E., 2000 – *Molecular phylogeny of Philippine freshwater sardines based on mitochondrial DNA analysis*. Journal of Heredity, Nr. 91, p. 247 – 253.
270. SANGER F., AIR G. M., BARELL B. G., BROWN N. L., COULSON A. R., FIDDES C. A., 1977 – *Nucleotide sequence of bacteriophage ØX174 DNA*. In *Nature*, Nr. 265, p. 687 – 695.
271. SANKO D., NADEAU J., 1996 – *Conserved synteny as a measure of genomic distance*. Disc. Appl. Math. Nr. 71, p. 247 – 257.
272. SCHÄPERKLAUS W., 1961 – *Lehrbuch der Teichwirtschaft*, Aufl., Paul Perey, Berlin, Hamburg, (abstract).
273. SCHULTZ R. J. 1967 – *Gynogenesis and triploidy in the viviparous fish Poeciliopsis*, Science, 157, p. 1564 – 1567.
274. SHELTON W. L., 1986 – *Control of sex in Cyprinids for aquaculture*, Aquaculture al Cyprinids, I.N.R.A., Paris.

275. SHEN S.C., 1993 *Fishes of Taiwan* Ed. Department of Zoology, National Taiwan University, Taipei. p. 960.
276. SITNIKOVA T., RZHETSKY A., NEI M., 1995 – *Interior branch and bootstrap tests of phylogenetic trees.* Mol. Biol. Evol. Nr. 12, p. 319 – 333.
277. SNEATH P. H. A., SOKAL R. R., 1973 – *Numerical taxonomy*. Ed. Freeman, San Francisco CA. (abstract).
278. SNEDECOR, G. W., 1968 – *Metode statistice aplicate în cercetările de agricultură şi biologie*, Editura Didactică şi Pedagogică, Bucureşti, p. 37.
279. SOKAL R. R., MICHENER C. D., 1958 – *A statistical method for evaluating systematic relationships*. Univ. Kansas Sci. Bull. Nr. 28, p. 1409 – 1438.
280. SOKAL R. R., SNEATH P. H. A., 1963 – *Principles of numerical taxonomy*. Freeman, San Francisco CA. (abstract).
281. SOURDIS J., NEI M., 1988 – *Relative efficiencies of the maximum parsimony and distance-matrix methods in obtaining the correct phylogenetic tree*, în Mol. Biol. Evol., nr. 5, p. 298 – 311.
282. SPIRIN A., 1958 – *Spektrofotometriceskoe opredelenie summarnovo kolicestva nucleinovih kislot. Biohimia*, p.656.
283. STAN TR., 1985 – *Piscicultura*, lucr. practice, lito. I. A. Iaşi, p. 15 – 16.
284. STAN, T., PĂSĂRIN, B., 1996 – *Acvacultură*, Universitatea Agronomică şi de Medicină Veterinară „Ion Ionescu De la Brad”, Iaşi, p 31.
285. STEBBINS G. L., 1969 – *The basis of progressive evolution*, Univ. North Carolina, Press, Chapel Hill (abstract).
286. STUART G. W., MOFFETT KAREN, LEADER J. J., 2001 – *A comprehensive vertebrate phylogeny using vector representations of protein sequences from whole genomes*. Mol. Biol. Evol. p. 1 – 31.
287. SUCIU MARIA, POPESCU ALEXANDRINA, 1981 – *Lucrări practice de zoologie*, ediția a II-a, Editura Didactică şi Pedagogică, Bucureşti.
288. SUSKO E., INAGAKI Y., ROGER A. J., 2004 – *On Inconsistency of the Neighbor-Joining, Least Squares, and Minimum Evolution Estimation When Substitution Processes Are Incorrectly Modeled*, în Mol. Biol. Evol., vol 21 nr. 9, p. 1629 – 1642.
289. SWOFFORD D. L., BEGLE D. P., 1993 – *PAUP: Phylogentic analysis using parsimony, ver. 3.1. users manual*, Illinois Natural History Survey, Champaign, p. 74 – 102.
290. TAISESCU ELENA, 1979 – *Studiul complementului cromosomial la unele specii de peşti din fauna țării noastre*, Rez., Teză Dr., Fac. Biol., Univ. Bucureşti (rezumat).
291. TAKAHASHI K., NEI M., 2000 – *Efficiencies of fast algorithms of phylogenetic inference under the citeria of Maximum-Parsimony, Minimum Evolution, and Maximum Likelihood when a large number of sequences are used*. Mol. Biol. Evol. Nr. 17(8), p. 1251 – 1258.
292. TAKEZAKI N., NEI M., 1994 – *Inconsistency of the maximum parsimony method when the rate of nucleotide substitution is constant*. J. Mol. Evol. Nr. 39, p. 210 – 218.
293. TAKEZAKI N., 1998 – *The trees generated by distance methods of phylogenetic reconstruction*. Mol. Biol. Evol. Nr. 15, p. 727 – 737.
294. TAKEZAKI N., NEI M., 1996 – *Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA*. În *Genetics* Nr., 144, p. 389 – 399.

295. TAKEZAKI N., GOJOBORI T, 1999 – *Correct and incorrect vertebrate phylogenies obtained by the entire mitochondrial DNA sequences*. Mol. Biol. Evol. Nr. 16, p. 590 – 601.
296. TAMURA K., 1992 – *Estimation of the number of nucleotide substitution when there are strong transition-trensversion and G+C content biases*. Mol. Biol. Evol. Nr. 9, p. 678 – 687.
297. TAMURA K., NEI M., 1993 – *Estimaton of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees*, Mol. Biol. Evol., Nr. 10, p 512 – 526.
298. TANG J., MORET B., 2003 – *Phylogenetic reconstruction from gene rearrangement data with unequal gene contents*. În: Proc. 8th International. Workshop on Algs. and Data Structures. Vol. 2748 of Lecture Notes in Computer Science, Springer Verlag, p. 37 – 46.
299. TANG J., MORET B., CUI L., DEPAMPHILIS C., 2004 – *Phylogenetic reconstruction from arbitrary gene-order data*. În: Proc. 4th IEEE Symp. on Bioinformatics and Bioengineering, IEEE Press, Piscataway, NJ, p. 592 – 599.
300. TATSUO O., Chaudhry M. S., Sohail B., Masatoshi Y., Nor A. H., Hideki E., Hisashi Y., Ryuichi M., 2004 - *Phylogenetic position of the small Kashmir flying squirrel, Hylopetes fimbriatus (≡ Eoglaucomys fimbriatus), in the subfamily Pteromyinae*, Can. J. Zool./Rev. Can. Zool. 82(8) p. 1336-1342.
301. TEROL J., MASCARELL ROSARIO, FERNANDEZ-PEDROSA VICTORIA, PEREY A. M., 2002 – *Statistical validation of the identification of tuna species: bootstrap analysis of mitochondrial DNA sequences*. J. Agric. Food Chem., nr. 50, p. 963 – 969.
302. THOMPSON J.D., 1994 – *Nucleic Acids Research*, Vol 22, p. 4673 - 4680.
303. VAITHILIMGAM G., RAMANA T., 1988 – *Technics for chromosome analisis*, The University of Texas Health Science Center at Dallas, Texas (abstract).
304. VAN VELLER M. G. P., KORNET D. J., ZANDEE M., 2000 – *Methods in vicariance biogeography: assessment of implementation of assumption 0, 1 and 2*, în *Cladistics*, Nr. 16, p. 319 – 345.
305. VARVARA M., ZAMFIRESCU Ş., NEACŞU P., 2001 – *Lucrări practice de ecologie*, Editura Univ. „Al. I. Cuza" Iaşi, p. 2 – 25.
306. VASCO A. D., CRANDALL A. K., FU Y. X., 2001 – *Molecular populations genetics: coalescent methods based on summary statistics*, p. 173 – 216.
307. VASZJILJEV V. P., MAKEEVA A. P., RJABOV I. N., 1975 – *O Triploidii Gibridov Karpa sz Drugimi Predsztaviteljami Szemejsztva Cyprinidae*. Genetika, T. XI, No. 8 (abstract).
308. VOICAN V., LUSTUN L., RĂDULESCU I., 1975 – *Practica selecției şi reproducerii la peşti*, Editura Ceres, Bucureşti.
309. VOLCKAERT F. A. M., GALBUSERA P. H. A., HELLEMANS B. A. S., Van den HAUTE, VANSTAEN D., OLLEVIER F., 1994 – *Gynogenesis in the African catfish (Clanas gariepinus). I. Induction of meiogynognesis with thermal and pressure shocks*, Aquaculture, 128, p. 221 – 235.
310. YANG Y.P., WOMACK J. E., 1998 – *Parallel radiation hybrid mapping: a powerful tool for high-resolution genomic comparison*. Genome research, Nr. 8, p. 731 – 736.
311. YANG Z., 1996 –*Maximum-likelihood model for combined analyses of multiple sequence data*, J. Mol. Evol. Nr. 42, p. 587 – 596.
312. YANG Z., 1996 – *Phylogenetic analysis using parsimony and likelihood methods*. J. Mol. Evol. Nr. 42, p. 294 – 307.

313. WAKELEY J., 1993 – *Substitution rate variation among sites in hypervariable region I of human mitochondrial DNA*, J. Mol. Evol. Nr. 37, p. 613 – 623.
314. WAKELEY J., 1994 – *Substitution rate variation among sites and the estimation of transition bias*, Mol. Biol. Evol., Nr. 11, p. 436 – 442.
315. WANG H. Y., LEE S. C., 2002 - *Secondary structure of mitochondrial 12S rRNA among fish and its phylogenetic applications*. Mol. Biol. Evol., Nr. 19, p. 138 – 148.
316. WANG L. S., WARNOW T., 2001 – *Estimating true evolutionary distances between genomes*. În: Proc. 33rd Ann. ACM Symp. Theory of Comput. ACM Press, New York, p. 637 – 646.
317. WANG L. S., JANSEN R., MORET B., RAUBESON L., WARNOW T., 2002 – *Fast phylogenetic methods for genome rearrangement evolution: An empirical study*. În: Proc. 7th Pacific Symp. on Biocomputing. World Scientific Pub., p. 524 – 535.
318. WANG L. S., WARNOW T., 2005 – *Distance-based genome rearrangement phylogeny*. În: Gascuel O. (Ed.), *Mathematics of Evolution and Phylogeny*. Oxford University Press, p. 353 – 383.
319. WELCOMME R.L., 1988 – *International introductions of inland aquatic species*, Rome, FAO. (abstract).
320. WERNER S., 1979 – *Industrienmabige fischproduction*, veb Deutscher Lanswirtschaf sverlang, Berlin, D. D. R. (abstract).
321. WHEELER W., 2001 – *Homology and the optimization of DNA sequence data*, în *Cladistics*, Nr. 17, p. 5 – 11.
322. WILLIAMS G. C., 1966 – *Adaptation and natural selection. A critique of some current. Evolutionary thought*, Princeton Univ. Press. (abstract).
323. WILEY E. O., 1981 – *Phylogenetics: The theory and practice of phylogenetic systematics*. Ed. by Wiley, New York (abstract).
324. WILEY E. O., BROOKS D. R., SIEGEL-CAUSEY D., FUNK V. A., 1991 – *The Compleat cladist: A primer of phylogenetic procedures*. Museum of Natural History, University of Kansas, Lawrence, (abstract).
325. WILLS C., 1981 – *Genetic variability*, Clarendon Press, Oxford (abstract).
326. WILSON A. C., CARLSON S. S., WHITE T. J., 1977 – *Biochemical Evolution*. Annu. Rev. Biochem., Nr. 46, p. 573 – 639.
327. WOLF C., RENTSCH J., HUBNER P., 1999 – *PCR-RFLP analysis of mitochondrial DNA: a reliable method for species identification*. J. Agric. Food Chem. Nr. 47, p. 1350 – 1355.
328. WOO-JAI L., CONROY JANET, HOWELL H. W., KOCHER T. D., 1995 – *Structure and evolution of teleost mitochondrial control regions*. J. Mol. Evol. (1995) Nr. 41, p. 54 – 66.
329. WU C., YE Y., CHEN R., 1986 – *Genome manipulation in carp (Cyprinus carpio L.)*, Aquaculture, 54, p. 57 – 61.
330. ZARDOYA R., MEYER A., 1996 – *Phylogenetic performance of mitochondrial protein-coding genes in resolving ralationships among vertebrate*. Mol. Biol. Evol. Nr. 13)7), p. 933 – 942.
331. ZARDOYA R., DOADRIO, I., 1998 – *Phylogenetic relationships of Greek Cyprinidae: molecular evidence for at least two independent origins of the Greek cyprinid fauna*, sub tipar.(abstract).
332. ZARDOYA R., DOADRIO, I., 1998 – *Phylogenetic relationships of Iberian cyprinids: systematic and biogeographical implications*, Proc. R. Soc. Lond. B Nr. 265, p. 1365-1372.

333. ZHANG H., OKAMOTO N., IKEDA Y., 1995 - *Two c-myc genes from a tetraploid fish, the common carp (Cyprinus carpio)*. În *Gene* Nr. 153, p. 231–236.
334. ZHANG J., ROSENBERG H. F., NEI M., 1998 – *Positive Darwinian selection after gene duplication in primate ribonuclease genes.* Proc. Natl. Acad. Sci. USA, Nr. 95, p. 3708 – 3713.
335. ZUCKERKANDL E., PAULING L., 1962 – *Molecular disease, evolution and genetic heterogenity*. In *Horizons in biochem8istry*,. Acad. Soc. New York. p. 189 – 252.