COMPARING THE DEHYDROGENASE'S 4L SUBUNIT POLYPEPTIDIC CHAIN FOR INDIVIDUALS OF *CYPRINUS CARPIO* L., 1785 AND *CARASSIUS GIBELIO* BLOCH. 1783 (*CYPRINIDAE*) SPECIES

Lucian D. GORGAN and Sorina Raluca GHERASIM
"Alexandru Ioan Cuza" University, Faculty of Biology, Carol I Bd., 20 A,700505, Iaşi
lucian.gorgan@uaic.ro

Abstract. The aim of present paper was to characterize and to compare the polypeptidic chains specific for the haplotypes of the gene which determines dehydrogenases' 4L sub-unit synthesis, for individuals of *Cyprinus carpio* L and *Carassius gibelio* Bloch species, with the origin in two different populations – Movileni and Iaşi, after the sequences translation, based on the mitochondrial genetic code for vertebrates. After the sequences alignment, based on the similarity percent, were established the characteristic haplotypes for the analyzed populations and for each haplotype, were studied different parameters used in polypeptide chains characterization.

Keywords: ND4L dehydrogenase, *Cyprinus carpio* L., 1785, *Carassius gibelio* Bloch., 1783.

Rezumat. Compararea catenelor polipeptidice ale subunităţii 4L a dehidrogenazei pentru indivizi ai speciilor *Cyprinus carpio* L. 1785 şi *Carassius gibelio* Bloch. 1783 (*Cyprinidae*). Scopul prezentei lucrări este de a caracteriza şi compara catenele polipeptidice caracteristice haplotipurilor genei care determină sinteza subunităţii 4L a dehidrogenazei, pentru indivizi aparţinând speciilor *Cyprinus carpio* L şi *Carassius gibelio* Bloch., proveniţi din populaţiile Movileni şi Iaşi. Analiza catenelor s-a efectuat după translaţia secvenţelor pe baza unui cod genetic mitocondrial pentru vertebrate. După alinierea secvenţelor, pe baza procentelor de similaritate, au fost stabilite haplotipurile caracteristice populaţiilor studiate, iar pentru fiecare haplotip, au fost studiaţi diferiţi parametri utilizaţi în caracterizarea lanţului polipeptidic.

Cuvinte cheie: ND4L dehidrogenază, *Cyprinus carpio* L., 1785, *Carassius gibelio* Bloch., 1783.

Introduction

NADH dehydrogenase has 7 sub-units on the mtDNA's (Fig. 1) structure. 4L sub-unit was chosen for amplification and sequencing, because it is considered to be a highly conservative region.

In the recent years, the use of mtDNA *ND4L* and *ND*4 genes for phylogenetic analysis has drawn attention (Arevalo *et al.*, 1994; Forstner *et al.*, 1995). Other studies based on sequences of ND4L for other fish species, were recorded for *Acipenseriformes*, endemic Chinese species and species from Europe and North America as well.

Material and Methods

The origins, tissues and preservation conditions of samples used in the study are presented in Table 1 and consist in dorsal muscle tissue samples from 40 individuals (10 individuals from both species and both populations – Movileni and Iaşi).

DNA was isolated according to the routine method (Zhang *et al.*, 1999) using phenol: chloroform: isoamyl alcohol (25 : 24 : 1) extraction and purification (Ausubel *et al.*, 1995).

The polymerase chain reaction (PCR) was performed in total reaction volume of 50 μL. For the ND4L gene amplification, have been used two complementary degenerated primers (L10420 5`-AAYAARAYCNTTGATTTCGRCTCA-3` for the direct

strand and H10720 5`-TCYGTGCCRTGYGTYCGNGC-3` for the reverse strand). The gene's amplifying reaction took place at 51°C as primer's aligning temperature, and a number of 40 replication cycles.

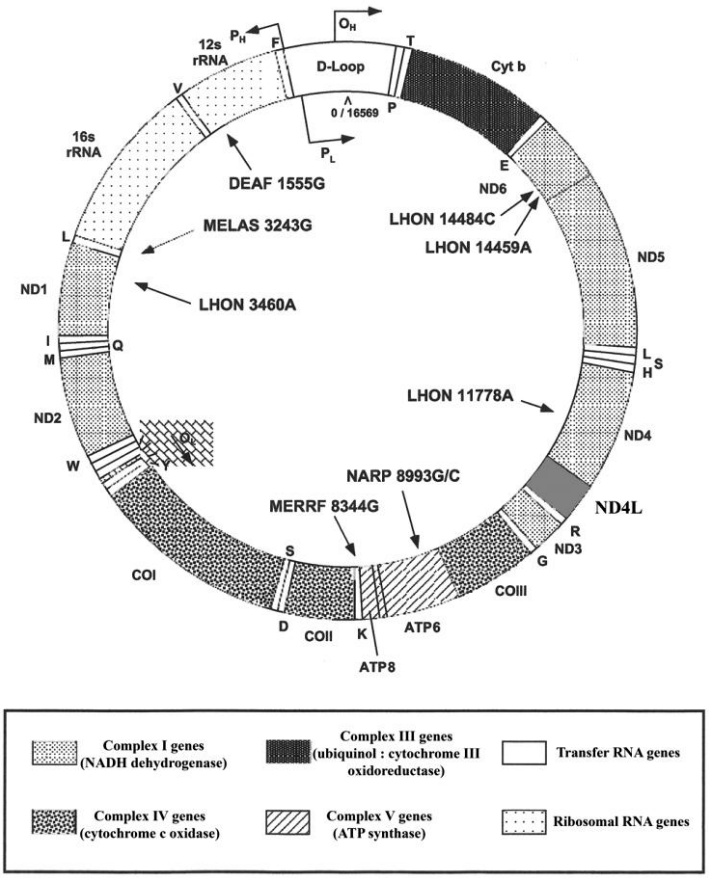

Figure 1. ND4L's position inside of mitochondrial DNA (Mitomap web page).

Table 1. Origin, tissue and preservation condition of samples.

Species	Origin or sampling site	No. of individuals	Tissue	Preservation condition
Cyprinus carpio L., 1785	Movileni	10	muscle	95% ethanol
Cyprinus carpio L., 1785	Iasi	10	muscle	95% ethanol
Carassius gibelio Bloch., 1783	Movileni	10	muscle	95% ethanol
Carassius gibelio Bloch., 1783	Iasi	10	muscle	95% ethanol

Towards sequencing, the codifying gene for NADH 4L sub-unit's synthesis information, for both analyzed species, we used the same primers: L10420 for the direct chain and H10720 for the reverse chain in 10µM concentration. The sequencing reaction program for species of *Cyprinus* genera had a primer's aligning temperature of 47ºC in 30 replication cycles and for *Carassius* individuals the aligning temperature was 50°C with 30 replication cycles.

Sequence's primary process as fluorograms and correction were made using CEQ2000 program by Beckman Coulter and the obtained data was exported as text, towards alignment.

Binding the direct and reverse chains for each individual, was made using ESEE 32 (The Eyeball Sequence Editor) program, through alignment with primer's sequences and cutting out the two chains, followed by the association in an unique chain.

Aligning all the gene's sequences, resulted from different individuals, (individuals from different populations, obtained by ginogenesys or hybrids) was carried out through Clustal W (Thompson *et al.*, 1994) and Clustal V (Higgins & Sharp, 1989; Higgins, 1994; Wheeler, 2001) methods, using the MegAlign module, from DNASTAR 5 (full trial version) program.

Based on the observed differences, between the analyzed sequences, the characteristic haplotype for the two populations was established. Based on this haplotype's sequence, translation was later established, using a mitochondrial genetic code for vertebrates. Based on the obtained polypeptidic chain after translation, the total number of amino acids, their type, frequency and percents from total weight were quantified.

Results and Discussions

Comparing each species sequences, similarity and divergence tables were obtained (Table 2 for *Cyprinus* genera and Table 3 for *Carassius* genera).

For *Cyprinus* genera individuals, the similarity percentage was of 100% for Cy01I and Cy01M sequences (representative for the two populations) and 98.9% for the comparison between CyGB sequence (taken from the Gene bank) and the other sequences, which represent the general haplotypes, characteristic for the two populations.

Table 2. Similarity and divergence percentage for *Cyprinus'* genera representative's established haplotypes.
(Cy01I = *Cyprinus'* 01 sequence from Iasi population; Cy01M = *Cyprinus'* 01 sequence from Movileni population; Cy GB = *Cyprinus'* sequence from GenBank)

	Similarity percentage				
Divergence percentage		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	

By aligning the obtained sequences for *Carassius* genera individuals we established the existence of 51 differences, of which 35 are transitions and 15 transversions, only between sequences taken from the Gene Bank. We have also established that there are no differences between the two analysed populations' sequences; therefore, taking into account the small number of mutations existent between the other sequences, we concluded that the codifying gene for dehydrogenases 4L subunit synthesis information is a conservative one, and it expresses very low mutagen potential. Therefore, we can conclude that a new halotype was discovered (CgIM) for ND4L, characteristic for *Carassius gibelio* species, sampled from Movileni and Iaşi populations.

Based on the sequences comparison, the similarity and divergence percentage table was obtained (Table 3), and we observed that the similarity percentage for the established haplotype is of 97.7% for the comparison with the *Carassius auratus x Cyprinus carpio* (CaxCy) hybrid and of 95.1% for the comparison with *Carassius auratus langsdörfi* (Clan) species' sequence.

Table 3. Similarity and divergence percentage for *Carassius* genera individuals established haplotypes.
(CgIM = *Carassius gibelio* general sequence from Iasi and Movileni populations; Ccuv = *Carassius cuvieri* sequence from GenBank; Clan = *Carassius auratus langsdörfi* from GenBank; CaxCy = *Carassius auratus x Cyprinus carpio* hybrid; Ca x Cy x Ccuv = *Carassius auratus* x *Cyprinus carpio* x *Carassius cuvieri* hybrid; Ccaras = *Carassius carassius*)

	Similarity percentage							
Divergence percentage		CgIM	Ccuv	Clan	CaxCy	CaxCy xCcuv	Ccaras	
	CgIM		93.5	95.1	97.7	97.0	97.3	CgIM
	Ccuv	6.8		93.5	93.9	94.7	94.3	Ccuv
	Clan	5.2	6.9		97.3	95.8	97.0	Clan
	CaxCy	2.3	6.5	2.7		98.5	99.6	CaxCy
	CaxCyxCcuv	3.1	5.6	4.4	1.5		98.9	CaxCyx Ccuv
	Ccaras	2.7	6.1	3.1	0.4	1.2		Ccaras
		CgIM	Ccuv	Clan	CaxCy	CaxCy xCcuv	Ccaras	

Polypeptidic chain translation (Fig. 2) for both genera haplotypes was established based on a mitochondrial genetic code for vertebrates and using EditSeq module.

MTPVHFSFSS AFILGLMGLA FHRTHLLSAL LCLEGMMLSL FIALALWALQ
FESTGFSTAP MLLLAFSACE ASTGLALLVA ARTHGT

Figure 2. ND4L polypeptide chain, characteristic for *Cyprinus carpio* species' Cy01IM halotype.

Following the *Cyprinus*'general haplotypes sequences translation, we observed that the entire polypeptides chain has 87 amino acids. Table 4 indicates that leucyne has the highest frequency (21.84%) and the lowest, glycine and tryptophan (1.15%). Likewise, we observed that the aspartic acid, lysine, asparagine and tyrosine, are missing. Also, from the total of 87 amino acids, 3 are acidics, 2 basics, 20 polar and 45 hydrophobic. More, the polypeptide chain has a charge of -0.49 (at pH=7) and an isoelectric point of 6.73 (Table 5 and Fig. 3).

Table 4. Polypeptide chain composition.

Amino acids		Number	Percentage from total weight (%)	Frequency (%)
Symbol	Name			
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

Table 5. Polypeptide chain characterisation.

Molecular weight (Daltons)	Number of amino acids	Amino acids				Isoelectric point	Charge at pH=7	Concentration for absorbtion = 1 and λ=280nm
		Basic	Acidic	Hydrophobic	Polar			
9247.07	87	2	3	45	20	6.736	-0.479	1.56

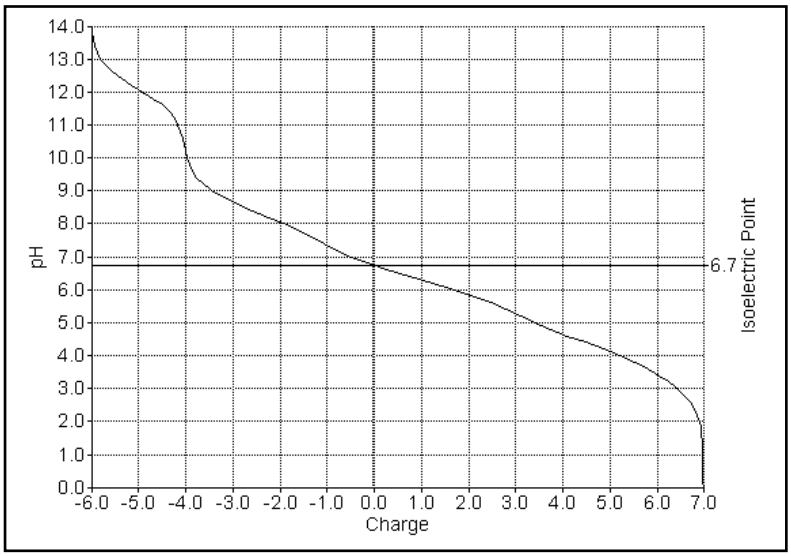

Figure 3. ND4L polypeptidic chain titration curve for *Cyprinus carpio* species’ Cy01IM haplotype.

MTPVHFSFSS AFILGLMGLA FHRTHLLSAL LCLEGMMLSL FIALALWALQ
FESTGFSTAP MLLLAFSACE ASTGLALLVA TARTHGT

Figure 4. ND4L polypeptide chain characteristic for *Carassius gibelio* species' CagIMD haplotypes.

After translation (Fig. 4), the entire polypeptide chain for *Carassius* genera general haplotype, consists of 87 amino acids (Table 6), from which leucyne has the highest frequency (24.84%), and glutamic acid and tryptophane have the lowest frequency (1.15%).

Table 6. Polypeptide chain composition.

Amino acids		Number	Percentage from total weight (%)	Frequency (%)
Symbol	Name			
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

From the total of 87 amino acids, 3 are acidic, 2 are basic, 20 are polar and 45 are hydrophobic. The polypeptidic chain has an electric charge of -0.48 (at pH=7), and an isoelectric point of 6.74 (Table 7).

Tabel 7. Polypeptide chain caracterisation.

Molecular weight (Daltons)	Number of amino acids	Amino acids				Isoelectric point	Charge at pH=7	Concentration for absorbtion = 1 and λ=280nm
		Basic	Acidic	Hydrophobic	Polar			
9247.07	87	2	3	45	20	6.736	-0.479	1.56

Subsequently, a protein structure analysis for both genera general haplotypes was carried out (Figs 5; 6), and the alpha regions, beta regions and turn regions using two comparative methods Garnier-Robson (Garnier *et al.* 1978) and Chou-Fasman (Chou & Fasman, 1978), hydrophilicity plot (Kyte & Doolittle, 1982), flexible regions (Karplus & Schultz, 1985), the antigenic index (Jameson & Wolf, 1988) and surface probability plot (Emini *et al.* 1985) have been determined.

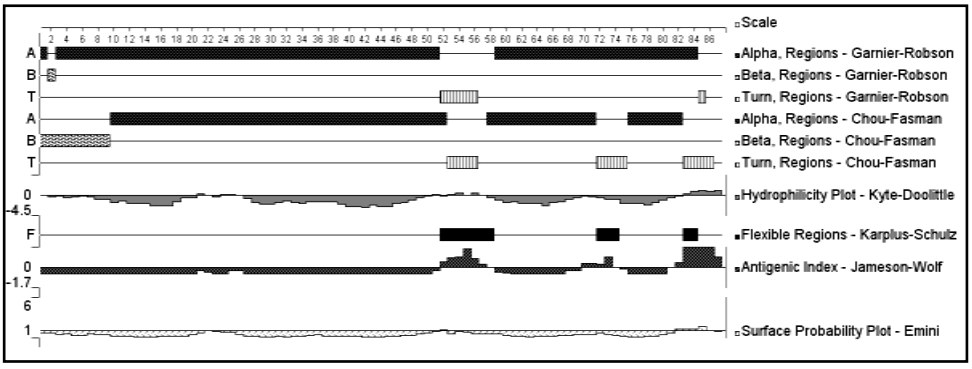

Figure 5. Polypeptidic chain structure analysis, for *Cyprinus* genera general haplotype.

Based on the polypeptidic chain structure, we concluded that there are no major differences between Garnier-Robson and Chou-Fasman methods, concerning the alpha helicoidal regions. Thus, for both methods, we noticed the existence of 3 alpha helicoidal regions and a partial surface superposition comparative between both methods. Concerning the beta folded regions, for both methods, the existence of one single beta region has been recorded, at the starting point of the polypeptidic chain, with the distinction given by those regions' surface differences, detected through both methods.

With reference to the inflexion regions, for the first method (Garnier - Robson), two such zones were recorded, whereas for the Chou – Fasman method 3 inflexion regions were presented, with the difference that in this case larger surface regions are presented.

Based on the hydrophilicity graphic, made after a Kyte – Doolitle model (Kyte & Doolittle, 1982), 6 hydrophobic and 6 hydrophilic regions were recorded.

From the antigenic index, computed after a Jameson–Wolf model (Jameson & Wolf, 1988), 3 regions with antigenic potential were recorded.

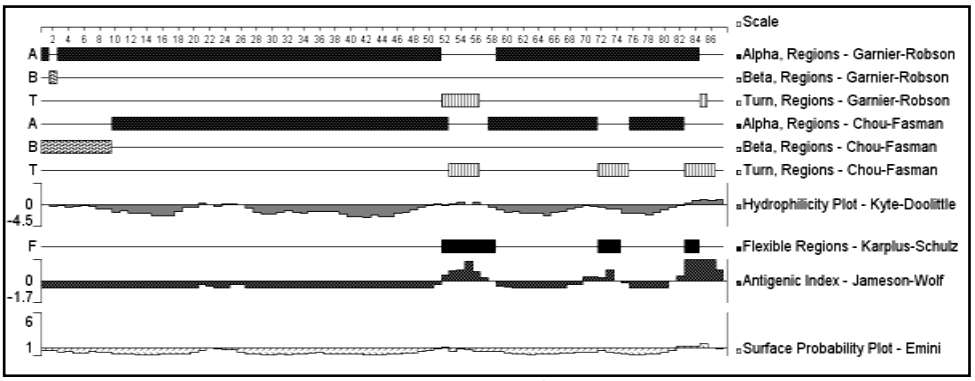

Figure 6. Polypeptidic chain structure analysis, for *Carassius* genera general haplotype.

Based on the polypeptidic chain structure, concerning the alpha helicoidal regions, for both methods, two such zones were presented. Also, for the beta folded zones, one single region was recorded for both methods. Regarding the inflexion regions, comparative for the two methods, we recorded 2 such regions for the first one (Garnier–Robson), and 3 for the second one (Chou – Fasman).

Based on the hydrophilicity graphic, made after a Kyte – Doolitle model (Kyte & Doolittle, 1982), 6 hydrophobic and 4 hydrophilic zones were recorded.

From the antigenic index, computed after a Jameson – Wolf model (Jameson & Wolf, 1988), 9 regions with antigenic potential were established.

Conclusions

From sequences alignment, we can conclude that ND4L is a high conservative region, reason for similarity percents between species of the same family.

Dehidrogenase's 4L subunit polypeptidic chain for *Cyprinus* genera individuals consists of 87 amino acids, from which leucine has the highest frequency (21.84%), and glycine and tryptophan the lowest (1.15%).

For individuals of *Carassius* genera, dehydrogenases' 4L subunit has the same chemical composition and structure like the *Cyprinus* genera individuals, fact which confirm once again the invariability of this region in evolutionary time.

References

Arevalo, E., Davis, S. K., Site, J. W., 1994. Mitochondrial DNA sequence divergence and phylogenetic relationships among eight chromosomes races of the *Sceloporus grammicus* complex (Phrynosomatidae) in central Mexico. *Syst. Biol*., **43**: 387.

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, **1**, cap. 2 – *Preparation and analysis of DNA. Phenol extraction and ethanol precipitation of DNA*, Ed. by John Wiley & Sons, Inc.: 211 – 213.

Chou, P. Y., Fasman, G. D., 1978. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol*., **47**: 45-148.

Emini, E. A., Hughes, J., Perlow, D., Boger, J., 1985. Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide. In *J. Virology*,. **55**: 836 – 839.

Forstner, M. R. J., Davise, S. K., Arevalo, E., 1995. Support for the hypothesis of Anguimorph ancestry for the suborder serpents from phylogenetic analysis of mitochondrial sequences. *Mol. Phylogenet. Evol.*, **4**: 93.

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*., **120**: 97-120.

Jameson, B. A., Wolf, H., 1988. The antigenic index: a novel algorithm for predicting antigenic determinants. *CABIOS*, **4**: 181 – 186.

Karplus, P. A., Schultz, G. E., 1985. Prediction of chain flexibility in proteins. In *Naturwissenschaften*, **72**: 212 – 213.

Kyte, J., Doolittle, R. F., 1982. A simple method for displaying the hydropathic character of a protein. *Jour. Mol. Biol.*, **157**: 105 – 132.

Mitomap, *A human mitochondrial genome database*, http://www.mitomap.org/mitomapgenome.pdf.

Thompson, J. D., 1994. *Nucleic Acids Research*, **22**: 4673 - 4680.

Zhang, S. M., Deng, H., Wang, D. Q., 1999. Systematic relationships among seven species of Acipenseriformes inferred from random amplified polymorphic DNA, Progress in *Natural Sciences* (in Chinese), **9 (9)**: 818.