

Federal Agency for Fisheries of the Russian Federation

**COMPARATIVE POPULATION-GENETIC ANALYSIS OF POLLOCK
THERAGRA CHALCOGRAMMA FROM THE BERING SEA AND WATER
OFF THE NORTH KURIL ISLANDS BASED ON POLYMORPHIC
MICROSATELLITE MARKERS**

(document submitted for IX Bering Sea Pollock Conference, Kushiro, Japan, 2004)

By

E.A.Shubina, M.N.Melnikova, A.I.Glubokov

2004

This contribution continues a series of studies of the genetic structure and dynamics of spawning concentrations of the Bering sea pollock.

The genome DNA cluster analysis carried out by us on the basis of determining the genetic distances using PCR – RAPD (amplification with random primers) has allowed us to suggest tentatively some features of differentiation between the Navarin and Olutor spawning concentrations. However, this analysis has indicated an insufficient level of polymorphism in the markers used. Population-genetic analysis was applied to the study of this problem during several recent years. This involved the use of microsatellite fragments of DNA which are the most polymorphic genetic markers of all the known ones. These parts of genome consist of tandemly (head-tail) repeated sequences; their main motive includes 2-6 nucleotide pairs. The length of the microsatellite loci reaches 200-300 nucleotide pairs. Microsatellites evolve at the expense of a shift during complementary pairing of the newly synthesized and matrix DNA chains at an early point of replication. The mutation rate among the microsatellite DNA is 10^{-2} – 10^{-4} events per locus per generation. The microsatellite markers in their nature are codominant, i.e. each individual possesses a characteristic set of two alleles (one in each chromosome). The intraspecies structure analysis of pollock is based on the comparative assessment of the allele frequency distribution in microsatellite loci from various samples. Four microsatellite loci were taken to examine the Olutor and Navarin spawning groups and Shirshov underwater Ridge aggregations. This study showed the separateness of the Olutor and Navarin groups, and that the Shirshov aggregation is equally distant from them. Our further research was aimed at proving or disproving those conclusions by way of independent sampling using as many loci as possible. The study was extended and covered the Sea of Okhotsk pollock to make the results comparable with those on the geographically remote group. The number of microsatellite markers used was increased up to 9 loci.

Material and methods.

The collection of tissue specimens (testes and blood) in alcohol included 120 samples from five conventional groups: Koryak, Olutor, Karagin, Okhotsk Sea (“Kamchatka”), and Navarin.

Extraction of DNA.

The lysis of tissue samples washed from alcohol was done with guanidineisocyanate during 10-12 hours at 60°C. Sorbing of lysate was made on hard phase carrier, and was deproteinized with an alcohol buffer solution. Following several washing procedures the clean DNA was eluated with buffer or water. A set of reagents DIAAtom™DNArep100 was used for eduction of DNA. Preparations were checked for polymery randomly using electrophoresis in an 0,6% agarose gel. The estimated size of preparations thus obtained is normally 40-50 thousand n.p. The outgoing DNA were up to 1 µg in 5 µl of solution (200µg/1µl).

PCR reactions.

The analysis was based on 9 microsatellite loci: Tch5, Tch10, Tch12, Tch13, Tch14, Tch15, Tch18, Tch19 and Tch22. The sequences of the main motive in the nine used microsattelite markers and their primers are given below (F=straight/R=reverse primers)

Tch5 (GATA)₁₄
 F: gcc tta ata tea cgc aca
 R: teg cat tga gcc tag ttt

Tch10 (GGCT)₆CTCT (GTCT)₂
 F: gtc tct atg tct gtc ttt cta ttt g
 R: acg aaa ccc aac cct gat t

Tch12	(GGTT) ₂₂	F: caa ttt gtc age etc tgt tac c R: agt aca get tga ttg ttt ctg gg
Tch13	(GT) ₉	F: ttt ccg atg agg tea tgg R: agt aca get tga ttg ttt ctg gg
Tch14	(GAAA) ₃₁	F: cat aca ttg gtc act ctt tct tac R: aaa ctg ata tac gcc caa ct
Tch15	(GA) ₃ (CA) ₂ GACA (GA) ₅ CAGATA(GA) ₈	F: aaa ctt cac ctg ace aac R: gca aca caa ctt aat cat ct
Tch18	(GT) ₁₅	F: gga gat ggt get aac tgg R: aac gca cat gca cat acg.
Tch19	(GTCT) ₁₅	F: tat get gat tgg tta ggc R: gat cat ttg ttt cag aga gc
Tch22	(GACA) ₆	F: atc ata tct ggc caa gtt c R: ctc tct ctg aat ccc tct g

The amplification reaction was made in the volume of 25 µl containing about 100 ng DNA, 10 mM tris HCl (PH 8.3), 50 mM KCl, 2.0 mM MgCl₂, 200 mM of each dNTP, 0.5mM of direct and reverse primers and 2 Tag polymerase each. Conditions for polymerase reaction: (94° C – 3 min)x1, (94° C – 1min, 54° C – 0.5 min) x 5, (90° C –3 min, 54° C – 0.5 min, 72° C – 0.5 min) x 40.

All reactions were carried out of Tag – polymerase preparase containing monoclonal antibodies. The presence of antibodies averts premature start of polymerisation reactions. Antibodies denature only at temperatures higher than 70 °C which ensures the so-called “hot start”.

Electrophoresis of microsatellite PCR products was done in non-denaturing vertical 8% polyacrylamide gels (200x200x0.4 mm), TBE buffer under 400 V and 10 mA during 10-12 hours. One of the glass plates was covered with molecular glue (Bind-silan, Sigma), the other one – with silicone. Samples of up to 10 µl were applied in a solution of mixture of bromophenol and xylencianol dyes containing 50% of glycerine. Each reaction was subdivided in three gels for compilation, with due regard to the markers and cross reactions. The process was conducted for 6-8 hours depending on the position of dyes. They tried to maintain standard conditions for each locus. The marker was 20 bp Molecular Ruler placed in three holes of each gel. Upon the completion of each electrophoresis the siliconized glass was removed; the gel attached to the second glass was colored with ethidium bromide and read in UV light which records the results of electrophoresis as a computer file. Drawing editor Adobe Photo Shop was used to process files; in some cases the Gel-Quant program was applied to determine the exact size of microsatellite alleles. In typing monoallelic products the high intensity alleles were given the status of homozygotes; those of low intensity were considered heterozygotes by the zero allele.

The results of typing served as basis for matrices in the form of computer programs utilized for population genetic analysis.

Computer analysis.

The genetic parameters of the assumed populations studied were determined using TFPGA and GENEPOP 3.1 programs. TFPGA uses algorithms which do not make it possible to evaluate more than two alleles per locus (binomial distribution). That is why it was the homozygote and heterozygote

most frequently occurring in the given locus of the concrete population that are selected for analysis; all the other combinations of alleles are combined into a single pool. GENEPOP was developed for polynomial distribution based on Markov's chains, and is capable of multiple elimination of the previous data from the memory and establishing new ones with due regard to existing results. Such an approach allows us to evaluate all the existing genotype options. The application of both programs makes it possible to assess both the statistical status of populations (TFPGA), and the dynamics of processes therein. The primary analysis made through TFPGA estimated the frequencies of alleles, the observed and expected heterozygosity, the allele and genotype diversity, and the polymorphism of the loci studied. The statistical confidence of the pair interpopulation differences between the observed and expected heterozygosity by 9 loci was estimated for the Koryak and Olutor, Koryak and Navarin groups using Wilcoxon nonparametric criterion.

The attainment of Hardy-Weinberg equilibrium by individual loci of specific populations was verified by χ^2 criteria and precise Haldane test at 99 % significance threshold (TFPGA), and by Guo and Thompson method for polynomial distribution (GENEPOP). The balance of populations in Markov chain algorithms was checked using summary Haldane, Weir and Robertson, and Hill tests for each locus and for population; for all loci and populations in general Fisher's combined probability test was applied as based on tables of conjoint characteristics.

The genotype imbalance in cohesion was found by loci pairs within each population, and for all populations at the same time independence of genotypes of one locus from genotypes of the other was taken as zero hypothesis. Statistical confidence was assessed using χ^2 criterion.

Population divergence expressed in differences among the alleles and genotypes, as tested, was assessed for each locus and all populations, and for each population pair by RxG Fisher test (by tables of conjoint characteristics).

F_{is}, F_{it} and F_{st} values (f, F_{and} θ in Weir&Cockerham algorithms) were established for all populations by separate loci and alleles for all population pairs. The jackknife procedure by the sum of loci made it possible to obtain an assessment of variance F statistics, while the bootstrap procedure established confidence limits. In this case there were 1000 steps of bootstrap made for 95% level.

The interpopulational genetic distances by Nei were calculated proceeding from F_{st} values ($F_{st} = 2 D^2$), and are expressed both as a diagonal matrix, and graphically as an unrooted UPGMA tree (TFPGA). The genetic distance estimates by five microsatellite loci among the major spawning concentrations represented in this sample, and UPGMA clusters fitted on the basis of those distances disagreed with our previous conclusions regarding the existence a genetically isolated group around the Koryak shelf (Shirshov ridge sample). In order to clarify the situation we monitored changes of the genetic distances during the consecutive introduction into the analysis of additional loci, as well as previously used loci Tch18 and Tch19 upon their repeated amplification and typing. The genetic parameters were assessed without accounting for possible exchange of genetic information among populations.

Results and discussion.

The results of standard analysis of the main genetic parameters based on allele frequencies in 9 microsatellite loci within 5 geographic samples of pollock are given in table 1. As the table 1 shows, all the samples examined are marked by heterozygote deficiency of some degree. That is why homozygotes were substituted with heterozygotes by the zero allele in most cases, and for all loci, for monoallele pattern analysis. However, as it follows from the table, both loci mentioned remained most deficient in terms of the degree of heterozygosis, despite this method of treatment. No doubt, this is a molecular characteristic of specific parts of DNA. A detailed analysis also showed that together with the loci which are conventionally most deficient in heterozygotes against a theoretically estimated

value, some loci were seen to have many excessive heterozygotes (e.g., Tch13, Tch18). The data obtained show close agreement between the number of alleles observed in all the sets of samples considered. The ratio between the observed and expected heterozygosity provides the main indication of the instances of Hardy-Weinberg relationship in the samples examined. The evaluation of Hardy-Weinberg balance for polynomial distribution of gametes in the analysis of individual loci showed that of the nine loci studied the most doubtful as regards this criterion are Tch14 and Tch15. In Tch14 the imbalanced ones are Olutor, Kamchatka and Navarin sets of samples. As for Tch15, the Olutor and Kamchatka samples do not fit into the 95% interval. Fisher's combined probability method showed uncertainty in relation to loci Tch13 (with excess of heterozygotes in the samples examined) (Table 1), and Tch14. As for the other loci, Hardy-Weinberg law was followed.

In addition to agreeing with Hardy-Weinberg relationship, one compulsory condition for recognizing as neutral the genetic markers used is to prove their independence. The Fisher method testing (with conjoint characteristic tables) of the genotype imbalance, in other words, the imbalance of loci in coupling showed a statistically reliable pair independence of all the loci used ($P \geq 0.17$) on the 95% level of significance (the tables had been drawn up for all individuals of the summary sample by all loci). The entire further analysis was conducted on the basis of the data obtained on the genetic balance of the system, and it concerned the population differentiation directly. A summary analysis of the interpopulational differences by the allele diversity showed a lack of difference between samples in loci Tch12 ($P=0.23 > 0.05$). For the other loci the samples showed differences in the alleles presented; the result was corroborated by a combined Fisher test.

The testing made on individual loci among all the population pairs allowed us to identify loci Tch13 and Tch18 which showed radical differences virtually among all the samples represented. The sample divergence data based on allele variant distribution are given in table 2. The head of this table summarizes the data on the genotype differentiation of samples. As one could expect, the number of loci

which demonstrate the genotype difference among the samples is smaller than the number of allele variants. In both instances the same loci contribute to the diversification of concrete samples by the criteria examined. However, none of them brings to light the genotypic diversity in all the samples compared simultaneously.

It follows from the results obtained that all the samples examined are subdivided in some degree. The Navarin and Olutor sets of samples are not indetical in the distribution of alleles of two loci, similarly to the Olutor and Karagin ones. The Kamchatka set of samples is representated by differentiation in a large number of loci. The isolation of the Koryak concentration is unexpected. It is separated from the Olutor group by loci Tch5, Tch13, Tch15, Tch18 and Tch22. If compared with the Navarin group, Tch14 will add itself up to the above list. The least number of differences is with the Karagin sampling set: in two loci only. The genotype differences to follow are about the same except that Tch13 falls out in some cases.

An additional confirmation of the isolation of the Koryak sample came from the test of statistical reliability of paired multiloci differences between the expected and observed heterozygosis Koryak and Navarin, or Koryak and Olutor sample sets. The results of analysis have confirmed reliably the difference between the Koryak group, and the Olutor or Navarin groups.

As a rule, it is the statistical value of F_{st} that serves as a quantitative expression of the measure of population differentiation. For the sea fish populations marked by periodic mixing of undividuals this value is within 0-0,05. Table 3 shows F_{st} values (F_{is} is the variance of allele frequencies of individuals within a separate population; F_{it} = variance in a summarized undivided population; F_{st} = interpopulation variance). A more accurate assessment, in our view, came when TEPGA program was applied.

Statistical verification of the values obtained made by resampling (jaknife and bootstrap procedures) showed that the differentiation estimates for the samples examined was correct at 95% confidence interval.

A precise test for overall sample subdivision for each of the nine loci, and for all loci at the same time, showed lack of differentiation only for locus Tch19. A multiloci test found that the value of χ^2 - 109.77, with the number of degree of freedom 18, and a 100% probability of differentiation.

Table 1. Main estimates parameters of the pollock samples examined by 9 microsatellite loci. He – unbiased expected heterozygosity; Ho – heterozygosity observed.

Description of loci			Group 1 (Koryak)			Group 2 (Olutor)			Group 3 (Kamchatka)			Group 4 (Karagin)			Group 5 (Navarin)								
	Number of allele	Size {in n.p.}	N	Number of allele	He	Ho	N	Number of allele	He	Ho	N	Number of allele	He	Ho	N	Number of allele	He	Ho	N	Number of allele	He	Ho	
			20				20				20				26				34				
Tch5	26	170-300		16	0,9269	0,9000		15	0,9231	0,9000		17	0,9063	0,9000		16	0,9282	0,9200		19	0,9236	0,8529	
Tch10	25	130-324		14	0,8667	0,8000		14	0,8974	0,9000		13	0,8987	0,6500		16	0,9299	0,8462		17	0,9100	0,8182	
Tch12	10	125-158		6	0,7833	0,4500		7	0,7715	0,5500		9	0,8769	0,6500		6	0,7572	0,6154		8	0,7524	0,6765	
Tch13	11	78-98		7	0,8410	0,8500		7	0,8051	0,9000		9	0,8679	1,0000		9	0,8160	0,9615		7	0,7113	0,9063	
Tch14	26	124-240		17	0,9774	0,7000		18	0,9462	0,6500		16	0,9295	0,5000		13	0,9184	0,5600		16	0,9196	0,5166	
Tch15	11	76-116		8	0,8115	0,6000		8	0,6551	0,5500		9	0,7423	0,6000		9	0,7534	0,6538		10	0,6864	0,5667	
Tch18	15	66-102		12	0,8987	0,9500		10	0,8397	0,9500		10	0,8564	0,9000		9	0,8653	1,0000		14	0,9148	0,9412	
Tch19	21	98-180		15	0,9282	0,6000		15	0,9192	0,8500		17	0,9459	0,7368		17	0,9299	0,7692		17	0,9425	0,7188	
Tch22	11	80-104		6	0,7179	0,6000		8	0,8372	0,6500		8	0,7885	0,8500		7	0,8122	0,6923		9	0,8371	0,7941	
Average	17,3			11,2	0,8580	0,7167		11,3	0,8427	0,7667		12	0,8706	0,7541		11,3	0,8567	0,7798		13,0	0,8442	0,7545	

Table 3. Estimation of F_{ST} values by allele frequency distribution of 9 microsatellite loci in pollock.

Loci \ Estimates	GENEPOP			TFPGA		
	F_{IT}	F_{ST}	F_{IS}	F_{IT}	F_{ST}	F_{IS}
Tch5	-0,028636	0,008217	-0,037159	0,0463	0,0077	0,0388
Tch10*	0,051261	0,018609	0,033271	0,1168	0,0089	0,1088
Tch12	0,0198324	-0,002046	-0,195877	0,2386	0,0042	0,2354
Tch13	-0,157660	0,076098	-0,253012	-0,0812	0,0685	-0,1607
Tch14*	0,196319	-0,0022956	0,198687	0,3846	0,0003	0,3845
Tch15	0,170112	0,004107	0,166689	0,1927	0,0090	0,1854
Tch18	-0,106001	0,028927	-0,138947	-0,0481	0,0310	-0,0817
Tch19*	-0,043047	0,004002	-0,047238	0,2097	-0,0088	0,2166
Tch22	-0,057439	0,035971	-0,96896	0,1467	0,0519	0,1000
Average for all loci	-0,0181	0,0203	-0,0392	0,1334	0,0189	0,1168
Resampling						
Jackknife for all loci				0,1334	0,1888	0,1172
Standard deviation				0,0502	0,0086	0,0565
Bootstrap 95% int.						
Upper limit				0,2332	0,0370	0,2263
Lower limit				0,0369	0,0041	0,0014

Note: * - loci showing highest differentiation when different programs are used

The table 4 summarized paired interpopulational genetic distances for 9 loci in F_{ST} values. Unbiased distances in Nei units are given in table 5. The Koryak sample is isolated accordingly too.

Table 4. Fst genetic distances for group pairs by individual loci, and totally by 9 loci. Numbers show samples:

1 – Koryak group; 2 – Olutor group; 3 – Okhotsk Sea group; 4 – Karagin group, 5 – Navarin group. Highest values of Fst were identified as basis for isolation of Koryak sample.

	Tch5				Tch10				Tch12				Tch13				Tch14			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
2	0,0203				-0,009				-0,0303				0,0461				-0,0164			
3	0,0128	0,0188			0,0384	0,0389			-0,0083	-0,0021			0,0910	0,0965			0,0072	-0,0024		
4	0,0020	0,0104	0,0031		0,0122	0,0033	-0,0018		-0,0282	-0,0129	0,0183		0,1109	0,0310	0,0358		0,0057	0,0112	-0,0081	
5	0,0231	-0,003	0,0060	0,0007	0,0181	0,0151	0,0520	0,0213	-0,0059	-0,0061	0,0217	-0,0046	0,1675	0,0329	0,1302	0,0172	0,0005	0,0018	-0,0088	-0,0177
	Tch15				Tch18				Tch19				Tch22				All loci			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
2	0,0395				0,0524				0,0015				0,0817				0,0206			
3	0,0012	-0,0037			0,0229	0,0293			0,0158	-0,0029			0,1173	0,0094			0,0337	0,0211		
4	-0,0030	0,0004	-0,0092		0,0516	0,0396	0,0223		0,0204	0,0069	-0,0047		0,1251	-0,0036	0,0021		0,0346	0,0100	0,0065	
5	0,0280	-0,0044	0,0055	-0,0037	0,0223	0,0288	0,0246	0,0126	0,0121	0,0088	-0,0031	-0,0062	0,0673	0,0118	0,0184	0,0068	0,0373	0,0097	0,0276	0,0029

Table 5. Paired genetic Nei distances (original ones above the diagonal; unbiased ones under diagonal) among various groups of pollock, by 9 microsatellite loci.

Group	Koryak	Olutor	Kamchatka	Karagin	Navarin
Koryak	***	0,3013	0,4412	0,3564	0,3603
Olutor	0,1681	***	0,2833	0,1729	0,1725
Kamchatka	0,2927	0,1422	***	0,1957	0,3015
Karagin	0,2309	0,0549	0,0624	***	0,1314
Navarin	0,2501	0,0698	0,1835	0,0365	***

Based on the bias distances, the largest distance separates the Koryak and Okhotsk Sea groups. Genetic distances between the Bering Sea samples are relatively identical between Navarin and Olutor, Olutor and Karagin, Navarin and Karagin ones. The unbiased distances copy the features of differentiation of the Bering Sea groups by other characteristics.

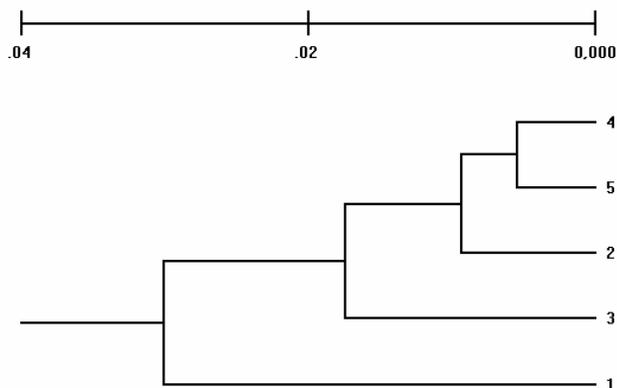


Figure 1. Cluster UPGMA of pollock concentrations based on ubias distances of Nei by 9 microsatellite loci.

1 – Koryak group; 2 – Olutor group; 3 – Kamchatka group; 4 – Karagin group; 5 – Navarin group.

As follows from the dendrogram, no monophyletic cluster pointing to a common origin of some or other tree branches could be found (Fig. 1). Bootstrap analysis of the cluster showed that the knot uniting the Karagin and Navarin samples have index of 60% and is supported by five loci. The Olutor sample branching knot has the same index but is not supported by any loci. This means that those three samples are not resolvable within the analysis made which corresponds approximately to equal distances (Table 5). The identification of Kamchatka sample has a 78% bootstrap coefficient and three loci supporting this knot. Finally, identification of the Koryak sample has a 100% index and support from all the nine loci. However, the summary index of bootstrap is only 0,8%, i.e. given one thousand resamplings, the tree represented had been formed only 8 times.

Previously, this sample was analysed by 5 loci (all of them have been included in this analysis), and a monophyletic cluster was obtained which logically combined the geographically close samples: Olutor and Karagin on the one hand, Navarin and Koryak on the other. The genetic distancing of the Okhotsk Sea group agreed with its remoteness (Fig. 2).

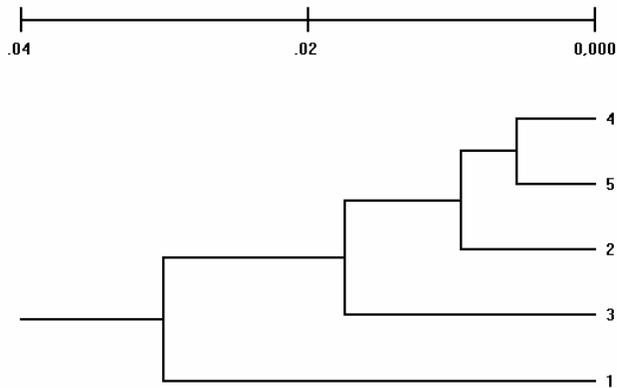


Figure 2. UPGMA tree without roots constructed on Nei genetic distances by 5 loci: Tch12, Tch14, Tch15, Tch18, Tch19.

1 – Koryak group; 2 – Olutor group; 3 – Kamchatka group; 4 – Karagin group; 5 – Navarin group.

In order to eliminate this contradiction the formation of clusters was monitored (UPGMA); first, at the partial substitution of homozygotes with heterozygotes by 0-allele (mostly, locus Tch12); second, at the gradual increase in the number of loci introduced in the analysis; finally, when exempting from the analysis of the loci which are the most informative as regards differentiation. It was found that the introduction of additional loci Tch5 and Tch10 into the analysis did not alter the cluster design. The loss of monophyly and correlation of branching knots with the geographical closeness had occurred when locus Tch13 typing data were added to the database. Meanwhile, the Bering Sea sample branches remained

in group whereas population from the Okhotsk Sea turned out to be the most remote. The Koryak sample showed the longest genetic distance out of all the Bering Sea samples.

The removal of locus Tch22 entails the establishment of a monophyletic UPGMA cluster fully identical with the one in Figure 2. Hence, it was shown additionally that the way of grouping of the Bering Sea pollock where the Koryak sample happens to be the most detached genetically, while the Olutor, Navarin and Karagin concentrations are nearly equally placed interpopulationally, results totally from loci Tch13 and Tch22. We should only point out that of all the UPGMA clusters presented the one obtained in our work (Fig. 1) only is bootstrap-supported by the entire structure. All the other clusters named here were supported in part only in some branching knots.

Therefore, we probably ought to consider the Koryak concentration of the Bering Sea pollock as detached from the Olutor, Karagin and Navarin ones. At the same time, each of these groups is undoubtedly a genetically distinct stock. These results should probably be explained by the special characteristics of the shelf and slope currents which might create conditions for the isolation of the individual groups of pollock.