Journal of Coastal Life Medicine

journal homepage: www.jclmm.com

Original article doi: 10.12980/jclm.4.2016j5-196

Transferrin gene polymorphisms and population genetic studies of Atlantic cod (*Gadus morhua*)

Berhan Asmamaw[1,2*]

[1]*Aquatic Case Team, Animal Biodiversity Directorate, Ethiopian Biodiversity Institute, P.O. Box 30726, Addis Ababa, Ethiopia*

[2]*Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, P.O. Box 5003, 1432 Ås, Norway*

ARTICLE INFO

Article history:
Received 8 Oct 2015
Received in revised form 20 Oct 2015
Accepted 12 Nov 2015
Available online 21 Dec 2015

Keywords:
Transferrin
Polymorphisms
Atlantic cod
Gadus morhua

ABSTRACT

Objective: To detect single nucleotide polymorphisms (SNPs) in the cod transferrin gene by comparing the sequences from Norwegian (North East Atlantic Ocean) and Canadian (North West Atlantic Ocean) specimen, and to quantify the genetic variation and differentiation in East and West Atlantic cod populations.

Methods: cDNA sequences between individuals of Canadian (North West Atlantic Ocean) and Norwegian (North East Atlantic Ocean) origin were aligned. Allele frequencies of the SNPs were used to discriminate the different Atlantic cod populations in West/East Atlantic Ocean, and the Baltic Sea.

Results: The sequence alignment detected 19 SNPs, of which 18 of them resulted in amino acid changes in the transferrin protein. Nonsynonymous to synonymous site substitution ratio (dn/ds) was by far greater than 1 providing an evidence for the existence of positive selection. The West Atlantic cod populations showed high values of heterozygosity and the Baltic populations were found to be inbred.

Conclusions: This study identified and indicated transferrin gene polymorphisms that can be used for population differentiations.

1. Introduction

Transferrin is an iron transport protein that binds to iron atom, thus making it unavailable for catalysis of superoxide radical formation. It is a single monomeric glycoprotein with molecular weight of 80 kDa that transports iron involved in many metabolic processes, hence considered as the major iron binding protein in the plasma of vertebrate species.

A number of transferrin variants have been identified and characterized in many different fish species that include, but not limited to the following, goldfish *Carassius auratus* (*C. auratus*)[1], crucian carp *C. auratus*[2], scad *Trachurus trachurus* L.[3], coho salmon *Oncorhynchus kisutch*[4], haddock *Melanogrammus aeglefinus* L.[5], *Channa punctatus*[6], European common carp *Cyprinus carpio carpio* L.[7], Nile tilapia *Orechromis niloticus*[8], and carp[9].

Different parameters can be used to study the genetic variations that can exist in sub-populations to test whether the samples are related or not. In this study, the main population genetic parameters such as allele frequency, population pairwise Fst (fixation index or measure of population differentiation), observed heterozygosity, expected heterozygosity, genetic distance, and genetic relationship have been used to study the level of transferrin gene polymorphisms between the Atlantic cod sub-populations from 14 different localities in West and East Atlantic Ocean. The main objectives of the study were, to detect single nucleotide polymorphisms (SNPs) in the cod transferrin gene by comparing the sequences from Norwegian (North East) and Canadian (North West) specimen, and to quantify the genetic variation and differentiation in East and West Atlantic cod populations.

*Corresponding author: Berhan Asmamaw, Aquatic Case Team, Animal Biodiversity Directorate, Ethiopian Biodiversity Institute, P.O. Box 30726, Addis Ababa, Ethiopia.
Tel: +251912033025
E-mail: bichiruet@yahoo.co.uk

2. Materials and methods

2.1. Fish samples

A total of 375 adult Atlantic cod fish were collected from 14 different localities of North Atlantic during 2002–2008 (Table 1). Genomic DNA was extracted from fin clips, muscle tissue or gill arches.

2.2. Transferrin gene sequences

The transferrin gene sequences of Norwegian Atlantic cod (tf-NE, *i.e.* transferrin from Atlantic cod of the North East Atlantic Ocean) was identified from the database of the Norwegian coastal cod population[10], and aligned with a published transferrin cDNA sequences from Canadian cod population (tf-NW, *i.e.* transferrin from Atlantic cod of the North West Atlantic Ocean)[11]. The missing

N-terminal coding sequence of the latter variant named tf-NW was PCR amplified from a heterozygous Faroe cod.

Table 1
Analyzed Atlantic cod populations, sampling locations and sample size.

Locality	Latitude	Longitude	Sampling year	Sampling size	Population codes
Baltic Öland	56.04	16.41	2004	29	OG
Baltic Bornholm	55.50	16.00	2004	30	OO
Kattegat	56.90	12.15	2004	29	KAT
North Sea	55.57	05.85	2002	29	NO
Norwegian coast, Molde	62.80	06.44	2003	12	CCS
Norwegian coast, Malangen	69.71	17.33	2003	18	CCN
Faeroe Bank	61.10	-08.30	2008	50	FB
Faeroe Plateau	61.96	-06.02	2008	49	FP
North East Arctic	70.65	29.81	2003	10	NEAC
Greenland Nuuk	64.73	50.45	2003	25	A
Greenland Sisimiut	66.84	52.89	2003	25	B
Newfoundland, Labrador	52.06	53.39	2004	19	NF
Canada, Nova Scotia	45.74	58.41	2003	25	C
Georges Bank	42.15	67.01	2003	25	GB

2.3. SNP analysis

2.3.1. PCR amplification

For the PCR amplification of the genomic DNA of sampled fish, the following reagents were mixed gently: Nanopure H_2O 1.850 μL, PCR buffer with $MgCl_2$ (10×) 0.625 μL, $MgCl_2$ (25 mmol/L) 0.325 μL, diethyl-nitrophenyl thiophosphate mix (25 mmol/L) 0.100 μL, primer mix (500 nmol/L each) 1.000 μL, genomic DNA (5-10 ng/μL) 1.000 μL, Hostar Taq (5 IU/ μL) 0.100 μL. The PCR reaction was cycled in a standard thermo cycler at 94 ℃ for 15 min, (94 ℃ for 20 s, 56 ℃ for 30 s, 72 ℃ for 1 min) for 45 cycles, 72 ℃ for 3 min and finally 4 ℃.

2.3.2. SNP genotyping

Transferrin gene were genotyped for all individuals using the MassARRAY system from Sequenom (San Diego, USA). PCR-primers and extension-primers were designed using the software SpectroDESIGNER v3.0 (Sequenom) (Table 2). Genomic DNA was PCR amplified and SNP genotyping was performed according to the iPLEX protocol from Sequenom[12]. For allele separations the Sequenom MassARRAYTM analyzer (Autoflex mass spectrometer) was used. Genotypes were assigned in real time[13], by using the MassARRAY SpectroTYPER RT v3.4 software (Sequenom) based on the mass peaks present. All results were manually inspected using the MassARRAY Typer analyzer v3.3 software (Sequenom).

2.3.3. Data analysis

Synonymous and non-synonymous mutations were detected by DNAsp[14]. Standard measures of genetic diversity including the observed heterozygosity (Ho) and expected heterozygosity (He) were calculated for each population using the software Arlequine version 3.1[15]. With the same software, allele frequencies, pairwise Fst values and an exact test of population differentiation were calculated. The genetic structure (hierarchical partitioning of genetic diversity) of the sampled populations were analysed using an analysis of molecular variance (AMOVA) framework. Reduction in the average proportion of heterozygosity genotypes within populations (Fis) was calculated to address the Hardy-Weinberg equilibrium deviation in each population. The value for each population related to the loci set used was obtained through FSTAT[16]. Genetic divergence among populations was also estimated with GENDIST programe in Phylip[17], by calculating standard genetic distance, DA (*i.e.* Masatoshi Nei's standard genetic distance)[18], for all possible pairs of populations. The resulting distance matrix was used to build a neighbor-joining dendrogram, to visualize the genetic relationship among populations.

3. Results

3.1. SNPs

Nucleotide sequence of transferrin cDNA of Canadian (North West Atlantic Ocean) and Norwegian (North East Atlantic Ocean) populations were compared to identify polymorphic sites. The nucleotide sequence alignment revealed 18 nonsynonymous and 1 synonymous substitution between the two sequences (Table 3). The rates of nonsynonymous to synonymous mutations among DNA sequences (dn/ds ratio) of 5.8 between the north west and north east variants is greater than 1 indicating strong positive selection on the cod transferrin gene.

The cod transferrin polymorphisms were investigated in 14 cod populations covering the North Atlantic by genotyping 8 of the 19 identified SNPs. High heterozygosity was observed in the populations from Canada, George Bank, Newfoundland and Sisimut compared to the east Atlantic and Baltic populations. Genetic differentiation at all loci was observed among the sampled Atlantic cod populations (Table 4). The population pair-wise Fst test revealed significant differences ($P < 0.05$) between most of the sampled population (Tables 5-7). Cases of significant Fst values were largely found in pair-wise comparison of populations of Canada, George Bank, Newfoundland and Sisimut with the rest of the ten populations. The level of differentiation among the West Atlantic populations, Canada, George Bank and Newfoundland was very small or negative, indicating how identical these populations are. This is supported by the Neighbour-joining dendrogram constructed using Nei's genetic

Table 2
PCR primers used for this study (the sequences are shown in 5'–3' direction).

SNP_ID	2nd-PCR primer	1st-PCR primer
Cod_Tfgene_03_AG	ACGTTGGATGTCCCGCCCACGCTGTGGTTA	ACGTTGGATGGTGTGTGCATGCGTATTACC
Cod_Tfgene_10_CG	ACGTTGGATGTGCCATGTCTGTAGATGGAG	ACGTTGGATGGTATTGCTCGACCATGGCTG
Cod_Tfgene_05_AG	ACGTTGGATGGAACATCAAGTTCGCAGCAC	ACGTTGGATGTTACTTGTTCACTCTGCAGG
Cod_Tfgene_06_AC	ACGTTGGATGACCTTACCTCGACTCCATAG	ACGTTGGATGTTTTCTTCTCAGCGTGACCC
Cod_Tfgene_07_AT	ACGTTGGATGTGGAGTTCATTCAGCGTCTC	ACGTTGGATGTAATCCTCTGAAAGCAGCCC
Cod_Tfgene_18_CT	ACGTTGGATGCTGGCGACCATAATCTCTTG	ACGTTGGATGGTTTTACCCCTTTACACTAG
Cod_Tfgene_13_CG	ACGTTGGATGTTAAAGGGACAAGGTCCAGC	ACGTTGGATGTATCTCTGCCCTACCAACAC
Cod_Tfgene_01_AG	ACGTTGGATGTGTGATGCCCTGAAGCTTAG	ACGTTGGATGGGGTTTCTCACCTTGATTGC

Table 3

Comparison of the 19 substituted amino acids in the cod transferrin (TF) variants (NE and NW).

AA site	155	788	806	836	845	847	1000	1096	1097	1210	1211	1414	1415	1948	1949	1993	2031	2048	2054
TF-NE	R	D	G	R	N	L	T	Y	Y	G	G	T	T	Q	Q	V	D	T	E
TF-NW	K	V	E	T	S	F	P	S	S	Q	Q	H	H	L	L	I	E	I	S

Table 4

Allele frequencies of the transferrin SNPs analysed in different cod populations.

Locus		OO	C	CCN	CCS	FB	FP	GB	Kat	NF	NAEC	NO	A	B	OG
TfSNP3_AG	A	0.017	0.361	0.111	0.042	0.070	0.051	0.381	0.017	0.364	0.111	0.052	0.229	0.340	0.000
	G	0.983	0.639	0.889	0.958	0.930	0.949	0.619	0.983	0.636	0.889	0.948	0.771	0.660	1.000
TfSNP10_CG	C	0.017	0.360	0.111	0.042	0.070	0.042	0.340	0.017	0.395	0.150	0.052	0.152	0.340	0.000
	G	0.983	0.640	0.889	0.958	0.930	0.958	0.660	0.983	0.605	0.850	0.948	0.848	0.660	1.000
TfSNP5_AG	A	0.983	0.340	0.889	0.958	0.930	0.949	0.400	0.983	0.342	0.800	0.948	0.780	0.600	1.000
	G	0.017	0.660	0.111	0.042	0.070	0.051	0.600	0.017	0.658	0.200	0.052	0.220	0.400	0.000
TfSNP6_AC	A	0.983	0.333	0.889	0.958	0.930	0.949	0.380	0.983	0.316	0.800	0.948	0.750	0.625	1.000
	C	0.017	0.667	0.111	0.042	0.070	0.051	0.620	0.017	0.684	0.200	0.052	0.250	0.375	0.000
TfSNP7_AT	A	0.033	0.740	0.111	0.042	0.112	0.082	0.652	0.086	0.737	0.200	0.121	0.240	0.313	0.017
	T	0.967	0.260	0.889	0.958	0.888	0.918	0.348	0.914	0.263	0.800	0.879	0.760	0.688	0.983
TfSNP18_CT	C	0.017	0.560	0.111	0.042	0.070	0.051	0.400	0.017	0.500	0.200	0.052	0.205	0.400	0.000
	T	0.983	0.440	0.889	0.958	0.930	0.949	0.600	0.983	0.500	0.800	0.948	0.795	0.600	1.000
TfSNP13_CG	C	1.000	1.000	0.917	0.958	0.959	0.980	1.000	0.983	1.000	0.900	0.966	0.938	0.896	1.000
	G	0.000	0.000	0.083	0.042	0.041	0.020	0.000	0.017	0.000	0.100	0.034	0.063	0.104	0.000
TfSNP1_ AG	A	0.567	0.280	0.389	0.375	0.641	0.576	0.340	0.321	0.316	0.300	0.517	0.523	0.458	0.446
	G	0.433	0.720	0.611	0.625	0.359	0.424	0.660	0.679	0.684	0.700	0.483	0.477	0.542	0.554

Codes are explained in Table 1.

Table 5

Comparisons of pairs of populations sampled (population pairwise FSTs)

	OO	C	CCN	CCS	FB	FP	GB	Kat	NF	NEAC	NO	A	B	OG
OO	0.00000													
C	0.60431	0.00000												
CCN	0.06055	0.40547	0.00000											
CCS	-0.01417	0.48257	-0.00623	0.00000										
FB	0.02312	0.52130	-0.01032	-0.01355	0.00000									
FP	0.01248	0.57050	-0.00069	-0.02078	-0.00858	0.00000								
GB	0.52934	-0.00278	0.32272	0.40284	0.43968	0.49042	0.00000							
Kat	-0.00529	0.57486	0.03769	-0.01669	0.00732	0.00057	0.49777	0.00000						
NF	0.62773	-0.02140	0.40788	0.49087	0.53236	0.58713	-0.01105	0.59482	0.00000					
NEAC	0.18808	0.27990	-0.01519	0.05661	0.03725	0.07540	0.19891	0.13997	0.27493	0.00000				
NO	0.01429	0.50842	-0.00345	-0.01919	-0.01194	-0.01323	0.42722	-0.00497	0.52087	0.05288	0.00000			
A	0.19126	0.26384	0.02087	0.09319	0.07894	0.11653	0.18267	0.15953	0.26150	-0.03470	0.09020	0.00000		
B	0.33022	0.11821	0.13086	0.20968	0.23391	0.27886	0.06492	0.30174	0.11255	0.03497	0.23087	0.02790	0.00000	
OG	- 0.00538	0.63268	0.09962	0.01684	0.04439	0.03830	0.56057	0.01334	0.65997	0.25857	0.04069	0.23284	0.36365	0.00000

Distance method: Pairwise differences.

Table 6

Comparisons of FST *P* values of pairs of population samples.

	OO	C	CCN	CCS	FB	FP	GB	Kat	NF	NEAC	NO	A	B	OG
OO	-													
C	0.00000	-												
CCN	0.00901	0.00000	-											
CCS	0.87387	0.00000	0.51351	-										
FB	0.07207	0.00000	0.55856	0.51351	-									
FP	0.14414	0.00000	0.32432	0.80180	0.69369	-								
GB	0.00000	0.38739	0.00000	0.00000	0.00000	0.00000	-							
Kat	0.56757	0.00000	0.06306	0.77477	0.16216	0.35135	0.00000	-						
NF	0.00000	0.92793	0.00000	0.00000	0.00000	0.00000	0.54955	0.00000	-					
NEAC	0.02703	0.00000	0.53153	0.27928	0.09910	0.04505	0.00000	0.01802	0.00000	-				
NO	0.09910	0.00000	0.34234	0.64865	0.78378	0.82883	0.00000	0.54054	0.00000	0.13514	-			
A	0.00000	0.00000	0.10811	0.01802	0.00000	0.00000	0.00000	0.00000	0.00000	0.94595	0.00000	-		
B	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00901	0.00000	0.00000	0.11712	0.00000	0.09910	-	
OG	0.99099	0.00000	0.00901	0.30631	0.04505	0.03604	0.00000	0.11712	0.00000	0.01802	0.06306	0.00000	0.00000	-

Number of permutations: 110.

Table 7
Comparisons of matrix of significant Fst *P* values of pairs of population samples.

	OO	C	CCN	CCS	FB	FP	GB	Kat	NF	NEAC	NO	A	B	OG
OO	+	+	-	-	-	+	-	+	+	-	+	+	-	
C	+		+	+	+	+	-	+	-	+	+	+	+	+
CCN	+	+		-	-	-	+	-	+	-	-	-	+	+
CCS	-	+	-		-	-	+	-	+	-	-	+	+	-
FB	-	+	-	-		-	+	-	+	-	-	+	+	+
FP	-	+	-	-	-		+	-	+	+	-	+	+	+
GB	+	-	+	+	+	+		+	-	+	+	+	+	+
Kat	-	+	-	-	-	-	+		+	+	-	+	+	-
NF	+	-	+	+	+	+	-	+		+	+	+	+	+
NEAC	+	+	-	-	-	+	+	+	+		-	-	-	+
NO	-	+	-	-	-	-	+	-	+	-		+	+	-
A	+	+	-	+	+	+	+	+	+	-	+		-	+
B	+	+	+	+	+	+	+	+	+	-	+	-		+
OG	-	+	+	-	+	+	+	-	+	+	-	+	+	

Significance level = 0.0500; Number of permutations: 110.

distance, which also illustrated a clear differentiation between the samples analysed.

3.2. AMOVA analysis

Based on allele frequency information, the AMOVA test revealed that 74.37% of the total genetic variation was explained by variability within the individuals. The genetic variation among populations and individuals within populations were 29.81% and -4.18%, respectively (Table 8).

Table 8
AMOVA results showing the sources of variations and their value in percentage.

Source of variation	*df*	Sum of squares	Variance components	Percentages of variations
Among population	13	165.247	0.23023	29.81
Among individuals within populations	362	184.596	-0.03227	-4.18
Within individuals	376	216.000	0.57447	74.37
Total	751	565.843	0.77243	

These differences were then investigated on a locus by locus basis to determine which locus or loci were responsible for the differentiation. This showed that locus TfSNP5 and TfSNP6 were the major contributors, with the highest Fst value of 0.361 and 0.374, respectively. On the other hand, TfSNP13 and TfSNP1 contributed less (Fst 0.016 and 0.043, respectively) for the populations differentiation. The pattern of TfSNP1 was different from the others, probably because the primers used may have picked up the second transferrin gene (*TF2*) which was not part of this study (Table 9).

Table 9
Fixation index per locus of the populations analysed.

Locus	Fit (mean, SE)	Fst (mean, SE)	Fis (mean, SE)
Cod_Tfgene_03_AG	-0.138, 0.046	0.160, 0.030	-0.353, 0.077
Cod_Tfgene_10_CG	-0.147, 0.050	0.165, 0.027	-0.372, 0.077
Cod_Tfgene_05_AG	0.405, 0.115	0.361, 0.067	0.062, 0.086
Cod_Tfgene_06_AC	0.370, 0.113	0.374, 0.071	-0.012, 0.076
Cod_Tfgene_07_AT	0.384, 0.113	0.351, 0.091	0.047, 0.050
Cod_Tfgene_18_CT	0.218, 0.095	0.253, 0.053	-0.050, 0.061
Cod_Tfgene_13_CG	0.151, 0.113	0.016, 0.010	0.137, 0.111
Cod_Tfgene_01_AG	0.067, 0.047	0.043, 0.017	0.026, 0.049

Fis: The deficiency or excess of average heterozygotes in each population; Fst: The degree of gene differentiation among populations (in terms of allele frequencies); Fit: The deficiency or excess of average heterozygotes in a group of populations.

3.3. Genetic distance

The pair-wise genetic distance values indicated how far or close a population is genetically from the others and were used to generate a neighbour–joining dendrogram. The largest genetic distance was observed between Canada and Øland (0.268 228), while the smallest was between North Sea and Faroe Plateau (0.000 578) (Table 10).

The observed heterozygosity of the studied populations varied from 0.049 88 in population Øland to 0.406 33 in population Sisimut. Six of the populations (Øland, North East Arctic Cod, Kattegat, Coastal Cod South, Canada and Bornholm) showed heterozygosity deficit, with Fis (inbreeding coefficient) values of 0.074, 0.310, 0.183, 0.070, 0.034, and 0.044, respectively. The remaining eight populations showed excess heterozygosity with positive Fis values. The expected heterozygosity across the loci varied between 0.053 77 and 0.073 16 (Table 11).

3.4. Genetic relationships among samples

Neighbor-Joining/unweighted pair group method with arithmetic means method dendrogram (unrooted tree) was constructed from the Nei's genetic distance in order to ascertain whether results obtained using population genetic analysis were reflected in the geographic distribution (Figure 1). It shows only the relationships between the different populations, and does not imply to direction to the evolutionary changes.

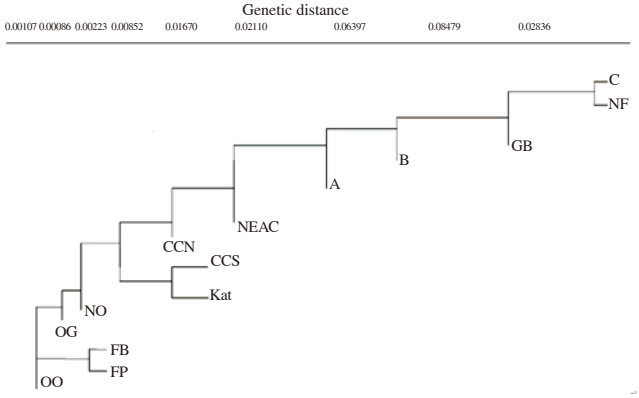

Figure 1. Neighbor-joining tree (unrooted) illustrating the relationship between the different cod populations analyzed for transferrin polymorphisms.

Table 10
Nei's genetic distance between different cod populations based on the analysed transferrin SNPs.

	OO	C	CCN	CCS	FB	FP	GB	Kat	NF	NEAC	NO	A	B	OG
OO	******	0.2650	0.0080	0.0040	0.0020	0.0006	0.1980	0.0070	0.2590	0.0240	0.0010	0.0230	0.0820	0.0020
C	0.2650	******	0.1940	0.2390	0.2310	0.2390	0.0060	0.2420	0.0009	0.1410	0.2280	0.1290	0.0610	0.2680
CCN	0.0080	0.1930	******	0.0030	0.0090	0.0060	0.1400	0.0040	0.1900	0.0050	0.0040	0.0100	0.0480	0.0070
CCS	0.0040	0.2390	0.0030	******	0.0090	0.0050	0.1790	0.0008	0.2360	0.0122	0.0030	0.0210	0.0720	0.0010
FB	0.0020	0.2310	0.0090	0.0090	******	0.0008	0.1690	0.0130	0.2250	0.02150	0.0020	0.0140	0.0640	0.0070
FP	0.0006	0.2390	0.0060	0.0050	0.0008	******	0.1770	0.0080	0.2340	0.0190	0.0006	0.0160	0.0690	0.0030
GB	0.1980	0.0060	0.1400	0.1790	0.1690	0.1770	******	0.1820	0.0040	0.1000	0.1680	0.0860	0.0370	0.2020
Kat	0.0069	0.2420	0.0040	0.0009	0.0130	0.0080	0.1820	******	0.2390	0.0130	0.0050	0.0260	0.0790	0.0030
NF	0.2580	0.0009	0.1900	0.2360	0.2250	0.2340	0.0040	0.2380	******	0.1390	0.2230	0.1250	0.0590	0.2640
NEAC	0.0240	0.1410	0.0050	0.0120	0.0210	0.0190	0.1000	0.0130	0.1390	******	0.0150	0.0090	0.0300	0.0200
NO	0.0010	0.2280	0.0040	0.0030	0.0020	0.0006	0.1680	0.0050	0.2230	0.0150	******	0.0150	0.0660	0.0030
A	0.0230	0.1290	0.0100	0.0210	0.0141	0.0160	0.0860	0.0260	0.1250	0.0090	0.0150	******	0.0190	0.0270
B	0.0820	0.0620	0.0480	0.0720	0.0640	0.0690	0.0370	0.0790	0.0590	0.0310	0.0660	0.0190	******	0.0870
OG	0.0020	0.2680	0.0070	0.0010	0.0070	0.0030	0.2020	0.0020	0.2640	0.0200	0.0030	0.0270	0.0870	******

Table 11
Standard diversity indices of different populations.

	OO	C	CCN	CCS	FB	FP	GB	Kat	NF	NEAC	NO	A	B	OG
He mean	0.07316	0.31636	0.18651	0.10725	0.14042	0.11577	0.33233	0.08114	0.32373	0.24565	0.12910	0.26932	0.35108	0.05377
Ho mean	0.0700	0.30556	0.19444	0.10000	0.15189	0.12825	0.37167	0.06650	0.36220	0.17222	0.13103	0.32277	0.40633	0.04988
Fis	0.044	0.034	-0.044	0.070	-0.083	-0.109	-0.122	0.183	-0.124	0.3100	-0 .015	-0.204	-0.161	0.074

He: Expected heterozygosity; Ho: Observed heterozygosity; Fis: Inbreeding coefficient; *P*-value = 0.05.

4. Discussion

The 19 observed SNPs between the two sequences indicated the existence of variation. This is in line with what has been reported about the existence of sequence variation[1], whereby 113 substitutions and 77 amino acid replacements were detected between coding sequences of the goldfish transferrin alleles A1 and B13. In studying the effects of natural selection on patterns of DNA sequence[19], variation within and among four populations of chinook salmon was described. In the same manner a large number of variations in the amino acid transferrin sequences of *C. auratus* was observed[20]. Using the software DNAsp, the nucleotide sequence was translated into the peptide sequence, whereby 691 and 690 amino acids were derived in Norwegian and Canadian variants, respectively. None of the substitution was found at the iron binding sites, so it was impossible to tell about the iron binding capacity of the SNPs. Nonsynonymous-to-synonymous site substitution ratios (dn/ds ratios) between transferrin genes that is used for testing the existence of positive selection[21], should be greater than 1 so as to provide evidence for positive selection at a locus. Evidence for positive selection at transferrin has been limited to salmonids with a dn/ds ratios estimated for non-salmonid lineages were generally less than 1[1,22]. However, this study revealed a dn/ds ratio of 5.8 rejecting the conclusion that positive selection at transferrin is limited to salmonids. Positive selection at transferrin locus is also a phenomenon of Atlantic cod. This result is supported by a report for salmonids in which approximately 13% of the transferrin codons were subject to positive selection with an average Ka/Ks ratios (an alternate designation of dn/ds ratio) of 7[22]. Moreover, positive selection of transferrin in *C. auratus*[1], and directional positive selection of transferrin gene in Antarctic Notothenioids[23], have been reported.

The allele frequencies of the SNPs that were detected in the cod transferrin gene are different in the different sampled populations reflecting the genetic diversity. On the basis of Canadian transferrin variant, the allele frequencies of TfSNP1, TfSNP3, TfSNP5, TfSNP6, TfSNP7, TfSNP10, TfSNP13, and TfSNP18 were observed more frequently in the West Atlantic populations (Canada, George Bank, Newfoundland, Sisimut) than the Baltic populations (Bornholm, Øland, Kattegat), which can be explained by the high values of observed heterozygosity in the West Atlantic populations. In the same way, increased haplotype frequency of mitochondrial cytochrome b DNA was reported for Atlantic cod population from Newfoundland[24]. On the other hand, next to North East Arctic cod, Baltic and East Atlantic populations have high rate of inbreeding (Fis) as compared to the west populations indicating why less SNP allele variants were observed there. This is in consistence with a result in which the Baltic populations were less genetically diverse than populations of the same species from the North Sea/Atlantic Ocean[25].

The hierarchical AMOVA analysis also indicated the presence of significant levels of genetic variations in the Atlantic cod samples studied. In particular, although most of the total variance in the SNP allele frequency distribution occurred between individuals (74.37%), almost all of the remainder (29.81%) was attributable to differences among populations. Overall, there was a large amount of genetic heterogeneity among the samples of Atlantic cod from west and east of Atlantic Ocean. Accordingly, the value of Fst for all of these samples combined was 0.298 06 clearly showing the variations.

In this study, departure from Hardy-Weinberg equilibrium that might be caused by non-random mating, mutation, gene flow, drift and small sample size was observed. The highest value of observed heterozygosity (0.406 33) was obtained in the West Atlantic population of Sisimut and the lowest (0.049 88)

in the Baltic population of Øland. Out of the 14 populations sampled, six (Øland, North East Arctic Cod, Kattegat, Coastal South Norway, Canada and Bornholm) showed a deficit in heterozygosity with Fis values of 0.074, 0.310, 0.183, 0.070, 0.034, and 0.044, respectively. This indicated that the mating was probably not random; therefore, the relationship or the similarities were increased in the population, allowing genetic drift to play a primary role. The other possible justification for deficit in heterozygosity in these populations is the Wahlund effect, which is the mixing of undetected genetically divergent stocks within the samples[26]. It is also likely that the presence of null alleles and/or mis-scoring of heterozygotes were responsible for the deficit of heterozygosity. For the populations North East Arctic and coastal south, lower observed heterozygosity values could be correlated to small sample size (10 and 12, respectively) analyzed. The remaining eight populations showed excess heterozygosity with positive Fis values, probably because of the existence of balancing selection which favours heterozygotes in these populations.

Genetic differentiation can be influenced by different evolutionary forces, such as mutation, genetic drift, migration and their interactions. Estimates of genetic differentiation for transferrin gene polymorphisms between all 14 populations using population pairwise differences Fst gave significant differences ($P < 0.05$) for many of the sampled populations. The populations of Canada, George Bank, Newfoundland and Sisimut showed high levels of genetic differentiation. Fst values were high for all pair-wise comparisons with the West Atlantic populations (*e.g.* 0.659 97 for Newfoundland and Øland, and 0.632 68 for Canada and Øland) with significant *P*-values ($P < 0.05$). The same trend as this result was reported[27], whereby microsatellites revealed that the Canadian Atlantic cod population was significantly divergent from the European populations. A clear differentiation between west and east Atlantic population units after analyzing the nuclear DNA restricted fragment length polymorphisms of Atlantic cod collected from six locations across the North Atlantic was also reported[28]. Significant genetic differentiation was not detected between the populations of Bornholm and Øland, Bornholm and Kattegat, Bornholm and Coastal South, and Coastal North and North East Arctic, probably because of the extensive ongoing gene flow among these populations. The absence of significant differentiation between the populations Coastal North and North East Arctic was also reported[29], in which the spatial analysis of molecular variance revealed that the outer coastal samples (CCN) cannot be discriminated from North East Arctic by means of microsatellite markers, supporting the similarity of the two groups at the *Pan* I locus.

In this study, *F*-statistics indicated clear population differentiations. The Fis values are positive in the TfSNP5, TfSNP7, TfSNP13 and TfSNP1 loci, and negative in the TfSNP3, TfSNP10, TfSNP6 and TfSNP18 loci (Table 9). The positive Fis value indicated that the homozygote genotype frequency was high in the TfSNP5, TfSNP7, TfSNP13 and TfSNP1 loci, whereas the negative findings indicated that the frequency of the heterozygote genotype was high in the TfSNP3, TfSNP10, TfSNP6, and TfSNP18 loci. The Fit values were positive in all loci except TfSNP3 and TfSNP10. This is advantageous as increases in the frequency of homozygote genotypes in a population allow possible selection factors on these loci to be detected. The estimated Fst values of all loci indicated the differences in genetic structure that exist in the sampled populations. The two loci (TfSNP5 and TfSNP6) showed higher Fst values than the other loci, suggesting that the differentiation among populations is most evident in these two loci.

The pairwise genetic distance indicates the genetic differentiations between the different populations. Populations that are closer to each other have less genetic differentiation and those further apart correspond to more differentiated populations. The largest genetic distance (0.268 228) was observed between Canada, one of the West Atlantic populations and Øland, the Baltic population clearly indicating the high level of differentiation, which could be explained by the different geographical locations of the two samples. This is in agreement with the reported result[30], in which the cod populations surveyed from more distant locations around Newfoundland and Labrador showed genetic distinction. The smallest genetic distance was between North Sea and Faroe Plateau (0.000 578) which is tightly grouped demonstrating that there is high genetic similarity between these populations. The possible explanation for this is that the amount of gene flow between these sub-populations is sufficient to prevent or retard genetic differentiation of the assemblage in this area.

The neighbor joining method was used to assess the relationship among the different Atlantic cod sub-populations. It showed the genetic similarity of geographically adjacent populations. That is, the first top branch in the dendrogram (Canada, Newfoundland, George Bank) was separated from the other part of the dendrogram (Baltic populations, Bornholm and Øland) and this obviously corresponds to the geographical pattern. Generally, the analyses of Neighbour-Joining dendrogram and *F* statistics clearly showed the apparent genetic differentiation between the East and West Atlantic cod populations. This result is in agreement with the findings in which the mitochondrial DNA sequence variation in terms of allele frequency of samples of Atlantic cod from the Northeast (Norway and Barents Sea) and Northwest Atlantic sea was high[31].

Atlantic cods occur as several biologically distinct stocks which may have different production capacity that require different breeding management systems. The identification of these genetically different sub-populations contributes a lot for the application of proper stock management systems. Therefore, the study of the genetic diversity of these economically important species is vital. The result from the present study, based on the identified transferrin gene polymorphisms, has revealed the existence of strong population differentiation in East and West Atlantic, and Baltic oceans. The West Atlantic cod populations (*e.g.* Canada, Newfoundland) showed remarkably high heterozygosity than the East Atlantic populations. The result also showed how inbred the Baltic and East Atlantic cod populations are as compared to the West Atlantic populations.

Conflict of interest statement

I declare that I have no conflict of interest.

Acknowledgments

This study was supported by the Norwegian and Swedish Research Councils.

References

[1] Yang L, Zhou L, Gui JF. Molecular basis of transferrin polymorphism in goldfish (*Carassius auratus*). *Genetica* 2004; **121**: 303-13.

[2] Long H, Yu QX. Studies on resistance characteristic and cDNA sequence conservation of transferrin from crucian carp, *Carassius auratus*. *Mol Cell Biochem* 2007; **303**: 201-9.

[3] Borges MF, Turner RJ, Casey J. Plasma transferrin polymorphism in scad (*Trachurus trachurus* L.) populations from the north-east Atlantic. *ICES J Mar Sci* 1993; **50**: 299-301.

[4] Van Doornik DM, Milner GB, Winans GA. An improved method of data collection for transferrin polymorphism in coho salmon (*Oncorhynchus kisutch*). *Biochem Genet* 1995; **33**: 257-60.

[5] Jamieson A, Birley AJ. The distribution of transferrin alleles in haddock stocks. *ICES J Mar Sci* 1989; **45**: 248-62.

[6] Nabi N, Arif SH, Hasnain A. Genetic structure of natural populations of air-breathing Murrel *Channa punctatus* Bloch in the Rohilkhand Plains of India. *Asian Fish Sci* 2003; **16**: 77-84.

[7] Jurecka P, Wiegertjes GF, Rakus KŁ, Pilarczyk A, Irnazarow I. Genetic resistance of carp (*Cyprinus carpio* L.) to *Trypanoplasma borreli*: Influence of transferrin polymorphisms. *Vet Immunol Immunopathol* 2009; **127**: 19-25.

[8] Rengmark AH, Lingaas F. Genomic structure of the Nile tilapia (*Orechromis niloticus*) transferrin gene and a haplotype associated with saltwater tolerance. *Aquaculture* 2007; **272**: 146-55.

[9] Csizmadia C, Jeney Z, Szerencsk I, Gorda S. Transferrin polymorphism of some races in a live gene bank of common carp. *Aquaculture* 1995; **129**: 193-8.

[10] The cod genome initiative. About the cod genome project. Noway: the Research Council of Norway; 2011. [Online] Available from: www.codgenome.no [Accessed on 13th January, 2015]

[11] Denovan-Wright E, Ramsey NB, McCormick CJ, Lazier CB, Write JM. Nucleotide sequence of transferring cDNA and tissue-specific expression of the transferring gene in atlantic cod (*Gadus morhua*). *Comp Biochem Physiol B* 1996; **113**: 269-73.

[12] Sequenom. The science of interpreting the genome. San Diego, CA: Sequenom. [Online] Available from: www.sequenom.com [Accessed on 12th January, 2015]

[13] Tang K, Fu DJ, Julien D, Braun A, Cantor CR, Köster H. Chip-based genotyping by mass spectrometry. *Proc Natl Acad Sci U S A* 1999; **96**: 10016-10020.

[14] Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 2009; **25**: 1451-2.

[15] Excoffier L, Laval G, Schneider S. Arlequin ver 3.1. An integrated software for population genetics data analysis. Switzerland: Computational and Molecular Population Genetics Lab; 2006. [Online] Available from: http://cmpg.unibe.ch/software/arlequin3 [Accessed on 23rd Febuary, 2015]

[16] Goudet J. Fstat version 2.9.3.2. Lausanne: Goudet J; 2002. [Online] Available from: http://www2.unil.ch/popgen/softwares/fstat.htm [Accessed on 23rd Febuary, 2015]

[17] Felsenstein J. PHYLIP Phylogeny Inference Package, Version 3.69. Washington DC: University of Washington; 2008. [Online] Available from: http://evolution.genetics.washington.edu/phylip/getme.html [Accessed on 12th Febuary, 2015]

[18] Nei M. Genetic distance between populations. *Am Nat* 1972; **106**: 283-92.

[19] Ford MJ. Effects of natural selection on patterns of DNA sequence variation at the transferrin, somatolactin, and p53 genes within and among chinook salmon (*Oncorhynchus tshawytscha*) populations. *Mol Ecol* 2000; **9**: 843-55.

[20] Yang L, Gui JF. Positive selection on multiple antique allelic lineages of transferrin in the polyploidy *Carassius auratus*. *Mol Biol Evol* 2004; **21**: 1264-77.

[21] Hughes AL, Nei M. Patterns of nucleotide substitution at major histocompatability complex class I loci reveals overdominant selection. *Nature* 1988; **335**: 167-70.

[22] Ford MJ. Molecular evolution of transferrin: evidence for positive selection in salmonids. *Mol Biol Evol* 2001; **18**: 639-47.

[23] Trinchella F, Parisi E, Scudiero R. Evolutionary analysis of the transferrin gene in Antarctic Notothenioidei: a history of adaptive evolution and functional divergence. *Mar Genomics* 2009; **1**: 95-101.

[24] Arnason E. Mitochondrial cytochrome B DNA variation in the high-fecundity atlantic cod: trans-atlantic clines and shallow gene genealogy. *Genetics* 2004; **166**: 1871-85.

[25] Johannesson K, Andre C. Life on the margine: genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic sea. *Mol Ecol* 2006; **15**: 2013-29.

[26] Halliburton R. *Introduction to population genetics*. Upper Saddle River: Pearson Education Inc.; 2004, p. 650.

[27] Hutchinson W, Carvalho GR, Rogers SI. Marked genetic structuring in localised spawning populations of cod (*Gadus morhua*) in the North Sea and adjoining waters, as revealed by microsatellites. *Mar Ecol Prog Ser* 2001; **223**: 251-60.

[28] Jónsdóttir ÓDB, Imsland AK, Atladóttir ÓÝ, Daníelsdóttir AK. Nuclear DNA RFLP variation of Atlantic cod in the North Atlantic Ocean. *Fisheries Research* 2003; **63**: 429-36.

[29] Westgaard JI, Fevolden SE. Atlantic cod (*Gadus morhua* L.) in inner and outer coastal zones of northern Norway display divergent genetic signature at non-neutral loci. *Fish Res* 2007; **85**: 306-15.

[30] Beacham TD, Brattey J, Miller KM, Le KD, Withler RE. Multiple stock structure of Atlantic cod (*Gadus morhua*) off Newfoundland and Labrador determined from genetic variation. *ICES J Mar Sci* 2002; **59**: 650-65.

[31] Carr S, Crutcher D. Population genetic structure in Atlantic cod (*Gadus morhua*) from the North Atlantic and Barents Sea: contrasting or concordant patterns in mtDNA sequence and microsatellite data? In: von Herbing IH, Kornfield I, Tupper M, Wilson J, editors. The implications of localized fishery stocks. New York: Northeast Regional Agricultural Engineering Service; 1998, p. 91-103.