

Extent of mitochondrial DNA sequence variation in Atlantic cod from the Faroe Islands: a resolution of gene genealogy

H Sigurgíslason and E Árnason

Institute of Biology, University of Iceland, Grensásvegur 12, 108 Reykjavík, Iceland

Variation in a 250 base pair (bp) fragment of the mitochondrial cytochrome *b* (cyt *b*) has been used extensively for population studies in Atlantic cod *Gadus morhua*. To study the shape of the gene genealogy and the nature of the polymorphism, sequences of another region of the cyt *b* gene and the TP intergenic spacer were added, making a total of 566 bp from 74 cod from the Faroe Islands. A total of 44 segregating sites defined 41 haplotypes, many at frequencies greater than 5%. Haplotype diversity was 0.97 and nucleotide diversity 0.73% per base. A topology referred to as a constellation gene genealogy was observed with four major haplotypes at high frequencies, from each of which a

number of rare variants were derived. A young relative age of the haplotypes was gauged from the structure of the genealogy. The variation was mostly at synonymous sites within the coding region and thus likely to be neutral or under weak purifying selection. By comparative analysis this also applies to the TP spacer. Applying the locus to study population variation in the Faroe Islands by AMOVA revealed that the overall areas and localities within areas accounted for none of the variation, and all the variation was due to differences among individuals.

Heredity (2003) 91, 557–564, advance online publication, 15 October 2003; doi:10.1038/sj.hdy.6800361

Keywords: Atlantic cod; *Gadus morhua*; mtDNA genealogy; cyt *b*; TP spacer; Faroe Islands

Introduction

The Atlantic cod *Gadus morhua* supports an important commercial fishery in the North Atlantic. An understanding of the biology of cod is important for fisheries and their management, and also has an intrinsic value for biology in general. The question of stock structure or genetic population structure is one aspect of cod biology that has generated much interest for management and biologists alike. Important issues are at stake regarding management and conservation of the resource. Yet, in spite of much effort by population geneticists, a consensus has not been reached. Part of the difficulty is that different signals seem to emanate from the various genetic systems used. There is an apparent disagreement among workers about the utility of the various molecular markers for population studies (eg, Frydenberg *et al*, 1965; Jamieson and Jones, 1967; Mork *et al*, 1985; Carr *et al*, 1995; Árnason *et al*, 2000; Ruzzante *et al*, 2000; Jónsdóttir *et al*, 2002).

A number of studies have been made on population structure and gene flow in Atlantic cod using methods based on variation of mitochondrial DNA (mtDNA, Carr and Marshall, 1991a,b; Árnason *et al*, 1992; Pepin and Carr, 1993; Carr *et al*, 1995; Árnason and Pálsson, 1996; Árnason *et al*, 1998, 2000). In particular, studies using sequence variation of a certain fragment of the mitochondrial cytochrome *b* (cyt *b*) gene have revealed a rich polymorphism. On a global scale trans-Atlantic clines are

observed in frequencies of common haplotypes (Árnason, 1998, 2003) and the polymorphism also shows no substructure or net divergence among populations within conventionally divided cod stocks within the various countries. There is, however, some evidence for temporal components to the variation (Árnason and Pálsson, 1996; Árnason *et al*, 1998, 2000).

The cyt *b* sequence variation observed in cod is a polymorphism with four major haplotypes at frequencies greater than 5%. In addition, there are a number of haplotypes more frequent than 1% and a large number of rare types (Árnason, 2003). The genealogical relationships among the polymorphic haplotypes are shallow, meaning that the mutational distances among them are relatively short. Most of the observed variation is at synonymous sites and thus likely to be neutral or nearly neutral to natural selection. There also is considerable homoplasy at various sites due to repeat mutations implying a high mutation rate. Thus, there is a potential for divergence among populations by random drift of a number of haplotypes.

Gene genealogies and their geographical distributions provide an important inference tool in population genetics. The coalescent (eg, Kingman, 1982), a random process describing the genealogy, emphasises the importance of the topology of the genealogy and the age of alleles. This study is designed to resolve more clearly the genealogical relationships and homoplasies of cod mtDNA and to provide a clearer picture of the topology of the genealogy, when compared to our previous work. We ask whether the essential structure of the gene genealogies as a shallow relationship representing relatively young alleles (Árnason, 2003) is upheld or

changed by adding sequence variation of larger fragments of the mitochondrial genome.

To investigate these questions, we first study variation within the same 250 bp fragment of the *cyt b* gene previously studied. We then add another 242 bp fragment near the end of the *cyt b* gene as well as a 74 bp noncoding TP spacer region of the mitochondrial chromosome, located between the tRNA genes for proline and threonine (Johansen *et al.*, 1990; Johansen and Johansen, 1994; Johansen and Bakke, 1996).

The paper is part of our ongoing effort to apply population genetic methods of gene genealogies to understand the biology of cod and potential population structure in various areas. Up to now, mtDNA sequence variation had not been studied among cod from the Faroe Islands. We therefore chose to apply this effort to a sample from the Faroe Islands to also allow inference about population variation there using this mtDNA marker.

Materials and methods

Sampling

Samples of Atlantic cod were obtained in May 1997 from five localities of the Faroe Plateau surrounding the Faroe Islands. Three locations were North of and two were South-east and South-west of the islands (Table 1), thus comprising two overall areas, North and South, facilitating analysis of spatial pattern of the variation. Samples were obtained from local fishermen, who also provided precise sampling coordinates or, in the case of locality C, the centre of a fishing square (Table 1). This sampling gives the possibility of studying potential microspatial genetic differentiation within the Faroe Plateau cod stock (Jamieson and Jones, 1967; Jákupsstofu and Reinert, 1994).

Molecular analysis

Samples of tissue was taken from the anterior of the tongue and preserved in 95% ethanol. mtDNA was extracted by cutting preserved tissue and taking a sample the size of a pinhead and washing in high TE buffer (100 mM Tris-HCl, 40 mM EDTA pH 8.0). The samples were then digested overnight at 60°C in 200 µl of lysis buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 0.8% Tween-20) containing 2.5 µl of 20 µg/µl proteinase K. Following digestion, the proteinase K was inactivated by heating at 95°C for 5 min and the resulting sample was used directly for PCR amplification.

A pair of primers, L14413 and H14771 (the numbers refer to the 5' end of the primer and correspond to the base pair numbers in the Johansen and Bakke (1996), sequence of Atlantic cod mtDNA), were used to amplify

a 359 base pair (bp) fragment (PCR fragment 1) within the *cyt b* gene and primers L15076 and H15683 amplified a 608 bp fragment (PCR fragment 2) extending from within the *cyt b* gene through the tRNA^{Thr} and the TP spacer and into the tRNA^{Pro} gene.

PCR reactions contained 5 µl of DNA template, 0.32 µM (for fragment 1, see below) and 0.64 µM (for fragment 2) of each primer, 500 mM KCl, 15 mM MgCl₂, and 100 mM Tris-HCl, 166 µM of dNTPs and 1.5 U of *Taq* polymerase in a 52.5 µl volume. The reactions were performed in a Gene Amp PCR System 9700 (Applied Biosystems) with the following step cycle profiles; for PCR fragment 1: 5 min at 94°C followed by 35 cycles of denaturation (1 min at 94°C), annealing (30 s at 54°C) and extension (1 min at 72°C); for PCR fragment 2: 5 min at 94°C followed by 35 cycles of denaturation (45 s at 94°C), annealing (45 s at 48°C) and extension (1 min at 72°C).

Prior to sequencing, excess primers and nucleotides were enzymatically removed from PCR amplification products using a combination of exonuclease I and shrimp alkaline phosphatase (Amersham Pharmacia Biotech). Cycle sequencing was subsequently carried out using ThermoSequenase (Amersham Pharmacia Biotech) and Cy-5 fluorescent-labelled primers (corresponding to H14720 for fragment 1 and H15675 for fragment 2) using step cycle profile: 5 min at 94°C followed by 35 cycles of denaturation (30 s at 94°C), annealing (30 s at 58°C) and extension (30 s at 72°C). The resulting fragments were run on an automated ALFlex-press sequencer (Amersham Pharmacia Biotech).

Intralaboratory crosscontamination and mutations in early rounds of PCR present potential problems in a work of this kind. To address the question of fidelity of sequence, we re-extracted DNA from a separate cut of the tissue sample and repeated the PCR and sequencing from the new extraction (Arnason *et al.*, 2000). We applied this replication to rare variants and to a random sample of 14 individuals of common haplotypes. The replication gave identical results to previous results indicating fidelity of the sequences.

Fragment 1 yielded the sequence of a 250 bp region, forming approximately 22% of the *cyt b* gene, located between site numbers 14459 and 14708 (Johansen and Bakke, 1996). We refer to this sequence as the *cytbE* region. The second fragment yielded a sequence from site 15225 to 15466, another 21% of the *cyt b* gene, referred to as *cytbE*. The sequence of the last 18 base pairs of the *cyt b* gene (from 15467 to 15484) were frequently difficult in base calling of the sequencer, and we therefore excluded these base pairs from analysis, and limited the sequence to the 242 bp of coding sequence that consistently produced reliable sequence results. The sequence continued through the tRNA^{Thr} gene (sites 15485–15555)

Table 1 Sampling localities and dates for Atlantic cod from the Faroe Islands

Locality	Area	Latitude	Longitude	Fishing region	Date
1. Station TFA7	North	62°20'	07°26'	DH 8	30/05/1997
2. Station TFB7	North	62°19'	07°09'	DH 8	30/05/1997
3. Station TFC7	North	62°25'	06°30'	DH 7	01/06/1997
4. Station TFD7	South	61°15'	05°30'	DF 6	01/06/1997
5. Station TFE7	South	61°45'	07°30'	DG 8	06/06/1997

and 3%, respectively, and nine haplotypes were singletons in the sample (Table 2). The high-frequency haplotypes were all internal to the tree (Figure 1). Haplotype C, also internal to the tree and a singleton in the Faroe Islands sample, is nevertheless widespread, found in the Baltic, Norway, Newfoundland, Greenland and Iceland, where it is at 4% frequency (Árnason *et al.*, 2000). Six haplotypes were new in our study, so far only being found in the Faroe Islands (haplotypes EK, EL, EX, EY, DL and DO) and all were singletons in the sample. Three other rare haplotypes were found. Haplotype GJ, a singleton in the Faroe Islands sample, has also been found in one individual from Greenland (Árnason *et al.*, 2000). Haplotype H, found in two individuals in the Faroe Islands, has also been found in various locations, one individual in Iceland, four in Newfoundland and two in the Baltic (Carr *et al.*, 1995; Árnason *et al.*, 1998, 2000). And finally three individuals of haplotype MI were found in the Faroe Islands, the same haplotype also having been found at low frequency both in Greenland and Iceland (Árnason *et al.*, 2000). Overall haplotype diversity was $\hat{h}=0.81$ and nucleotide diversity was $\hat{\pi}=0.0064$, both relatively high due to the large number of haplotypes at high and relatively even frequencies.

All variable sites in the cytb region were synonymous. The haplotypes differed from each other by one to five mutations and all differed by a single mutation from their nearest neighbour. The variation at the cytb region for cod at Faroe Islands thus looked much like that in other countries with the same four high-frequency haplotypes as internal nodes of a shallow genealogy, and with several rare haplotypes differing by single substitutions from each common haplotype-forming terminal nodes (Árnason, 2003).

Considering the 492bp cytb.cytbE coding region revealed four major high-frequency haplotypes all internal to the genealogy: A.E1, E.E1, D.E1 and G.E7 (Table 2 and Figure 2). A total of 32 segregating sites defined 34 haplotypes. All but two variable sites were synonymous. The two amino-acid replacements were

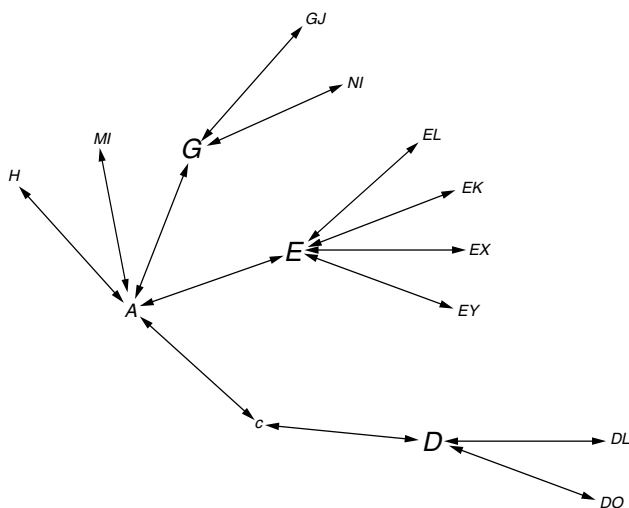


Figure 1 Network of relationships of haplotypes for a 250 bp cytb coding region found among 74 Atlantic cod from the Faroe Islands. Sizes of haplotype names is a rough indicator of relative frequencies.

functionally conservative changes from one nonpolar amino acid to another and both were found in singletons only. Haplotype G.E9 was a change from valine to leucine and haplotype G.E11 a change from valine to isoleucine. Overall haplotype diversity $\hat{h}=0.93$ increased compared to the cytb region alone due both to an increased number of haplotypes and due to a more even frequency distribution. However, nucleotide diversity $\hat{\pi}=0.0056$ decreased slightly. The branch between the high-frequency haplotypes A and G was longer relative to its length in the genealogy derived from the cytb region alone. There was also additional structure generated in the genealogy with the rarer haplotypes deriving both from common haplotypes and other rare haplotypes, thus adding new branches and increasing length of other branches (Figure 2).

A minimum spanning tree of the haplotypes was constructed from which 40 mutations were counted. In total, 22 were $g \leftrightarrow a$ purine transitions (21 synonymous), 15 were $c \leftrightarrow t$ pyrimidine transitions, two were $g \leftrightarrow c$ transversions and one was a $g \leftrightarrow t$ transversion. The transition/transversion ratio was 12.3:1 and the purine/pyrimidine transition mutation ratio was 1.5:1. The frequency of synonymous guanine mutations was high with eight of the nine synonymous guanine sites hit by mutation, one site being hit twice. A two by two comparison of the synonymous intraspecific purine to pyrimidine transition mutation ratio to the interspecific substitution ratio for cyt b between Atlantic cod and Greenland cod, *G. ogac* (Carr and Marshall, 1991a, b), and Walleye pollock, *Theragra chalcogramma* (Carr *et al.*, 1999) were significantly different (ratios 21:15 vs 3:11 and 4:13 and probabilities by Fisher's exact test $P=0.028$ and 0.022, respectively). The haplotypes differed from each other by one to nine mutations and all differed by one mutation from their nearest neighbour except haplotypes G.E11, which differed by three mutations from its nearest neighbour and NI.E10, EK.E16, EX.E2, D.E17 and DO.E8, each of which differed by two mutations from its nearest neighbour (Table 2).

The noncoding 74 bp TP spacer region contained 12 variable nucleotide sites defining 14 haplotypes (Table 2). In addition, a 29bp tandem repeat was found in

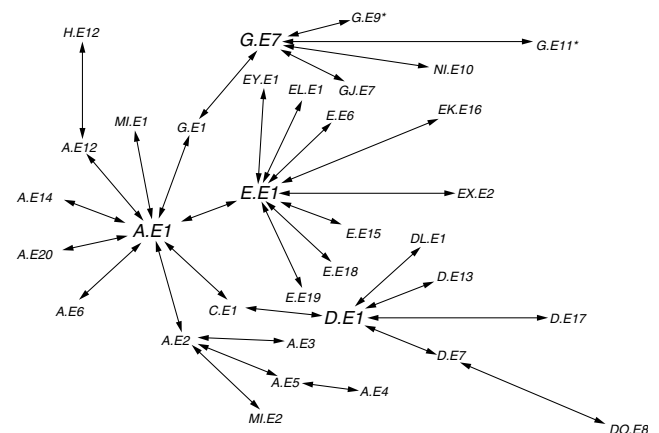


Figure 2 Network of relationships of 34 haplotypes at the 492 bp cytb and cytbE coding regions combined. Sizes of haplotype names is a rough indicator of relative frequencies. The starred haplotypes are amino-acid variants.

individual TFC7.06. Three haplotypes, *S1*, *S6* and *S3*, were at high frequencies of 51, 23 and 7%, respectively. They were all internal to the tree as was the fourth most frequent haplotype, *S2* at 4%. Haplotype *S11* was rare and nine singletons were found in the sample. Haplotype diversity $h = 0.68$ was somewhat lower than for the coding region but nucleotide diversity was more than three times that for the coding region or $\hat{\pi} = 0.0186$. A minimum spanning tree of the haplotypes was constructed from which 14 mutations were counted (Table 2). In all, 11 were $g \leftrightarrow a$ purine transitions, one was a $c \leftrightarrow t$ pyrimidine transition, one was a $g \leftrightarrow t$ transversion and one an $a \leftrightarrow t$ transversion. The transition/transversion ratio was 6:1 and the purine/pyrimidine mutation ratio was 11:1, even higher and significantly different from that for the coding region ($P = 0.04$, Fisher's exact test).

Adding information from the spacer region to make a 566 bp coding and noncoding region further resolved the gene genealogy (Figure 3). The genealogy of the full data contained 44 variable nucleotide sites defining 41 haplotypes (Table 2 and Figure 3), and from the minimum spanning tree a total of 54 mutations were counted. Given the tree, 10 mutations are homoplastic. The added information, however, allows a resolution of genealogy by parsimony because all homoplastic mutations occur in divergent branches of the tree. An example including the *A.E3* and *A.E4* haplotypes (Figure 2) that were not resolved by the coding sequences alone. Of the 54 mutations counted, 33 were $g \leftrightarrow a$ purine transitions, 16 were $c \leftrightarrow t$ pyrimidine transitions, two were $g \leftrightarrow t$ transversions, two were $g \leftrightarrow c$ transversions and one was

an $a \leftrightarrow t$ transversion. The transition/transversion ratio was 9.8:1 and the purine/pyrimidine mutation ratio was 2.1:1.

The gene genealogy had four major clades represented by haplotypes *A.E1.S1*, *E.E1.S1*, *D.E1.S6* and *G.E7.S1*, which all were relatively common. However, other haplotypes such as *A.E1.S6* also were relatively frequent. In fact, this particular haplotype, on the branch leading to the *D*-clade (Árnason, 2003), was slightly more frequent than *G.E7.S1* that defines the *G*-clade. Also some new clades or lineages were identified at a lower frequency, for example *A.E2.S3* and *A.E12.S1*. Adding the *cytbE* and spacer resulted in a general lengthening of internal branch lengths of the gene genealogy relative to the genealogy for the *cytb* region alone. There was one notable exception. A single mutation separated the two major haplotypes *A.E1.S1* and *E.E1.S1* (Figure 3), the same mutation that separates the *A* and *E* haplotypes and the *A*- and *E*-clade (Árnason, 2003). The high-frequency haplotypes were all internal to the tree. In addition, there were nine rare haplotypes and 28 singletons. The overall haplotype diversity was 0.97 and nucleotide diversity was 0.0073 per base. The mitochondrial DNA locus of Atlantic cod thus has a rich polymorphism with several closely related haplotypes at a relatively high frequency overlaid with a large number of rare and singleton haplotypes derived from various branches of the genealogy.

Population variation

An analysis of molecular covariance (Table 3) showed that covariances for both overall areas (North vs South) and sampling localities within areas had a negative sign. Localities within areas and areas within the country accounted for none of the variation and all variation was due to genetic differences among individuals within localities.

Discussion

Neutrality and selection

The variation at coding regions observed in this study was primarily at synonymous sites. The rare amino-acid variants are conservative changes that are not expected to affect function of the protein (Taylor, 1986) radically. Similarly, in the more extensive data set on the *cytb* region from around the Atlantic most of the rare amino-acid variants involved aliphatic isoleucine, leucine and valine residues, as observed here (Árnason, 2003). On this basis, one may expect that purifying natural selection already has removed deleterious amino-acid mutations leaving variation that is largely neutral to fitness at the coding regions. The presence of rare

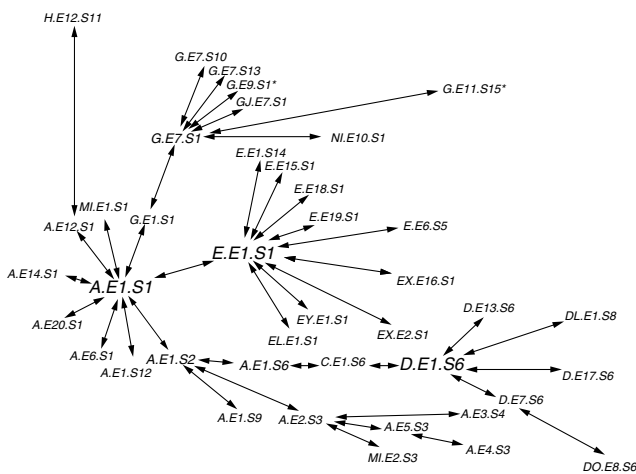


Figure 3 Network of relationships of 41 haplotypes at the 566 bp *cytb* and *cytbE* coding regions and the noncoding spacer region combined. Sizes of haplotype names is a rough indicator of relative frequencies. The starred haplotypes are amino-acid variants.

Table 3 Analysis of molecular variance of variation of a 566 bp fragment of the combined mitochondrial *cytb*, *cytbE* and spacer regions of Atlantic cod from the Faroe Islands

Source of variation	df	Sum of squares	(Co)variance	% total explained	Φ statistic	ASL
Among areas	1	1.2	$V_a = -0.026$	-1.22	$\Phi_{CT} = -0.012$	0.80
Localities in areas	3	6.5	$V_b = -0.001$	-0.05	$\Phi_{SC} = -0.0005$	0.44
Among individuals	69	151.2	$V_c = 2.191$	101.27	$\Phi_{ST} = -0.013$	0.62
Total	73	159.0	2.164			

conservative amino-acid variants intraspecifically relative to their absence in interspecific comparisons is frequently taken as evidence of weak purifying natural selection (Rand, 2001). Although there is little information about the potential function of the TP intergenic spacer, the variation observed in this region also is likely to be largely neutral or nearly neutral to selection. Of the 549 possible single-base mutations of amino-acid codons, 134 (one-fourth) are mutations to synonymous codons (King and Jukes, 1969). If variation of the spacer was strictly neutral, it could be argued to represent the baseline neutral mutation rate per base pair. Under that assumption, the ratio of nucleotide diversities between the coding *cytb.cytbE* and the noncoding spacer should similarly be one-fourth. The observed ratio, however, was $0.0056/0.0186 = 0.30$ in the right direction but slightly higher. This implies that there may be some purifying selective constraints on the spacer. The spacer also differed from the coding region in having a higher frequency of transversion mutation and a considerably higher frequency of purine over pyrimidine mutation, although *cyt b* also had a higher frequency of purine than pyrimidine mutation. The neutral theory (King and Jukes, 1969) regards polymorphism as a transient phase of molecular evolution and this provides a rationale for a test of neutrality. The test of *cytb.cytbE* intraspecific purine/pyrimidine mutation ratios *vs* the interspecific substitution ratios (McDonald and Kreitman, 1991) was significant, implying that the purine mutation bias is sorted out by purifying natural selection. This is all synonymous variation and therefore the selection must arise from the action of the cellular machinery, that is, from phenotypic expression of DNA and RNA in replication, transcription and translation. Such selection is expected to be weak (Akashi, 1995).

Shape of gene genealogies and relative allele ages

This paper is part of our ongoing work using mtDNA sequence variation for inference about population quantities and demographic parameters and history of cod by sampling sequence variation from the population (Árnason and Pálsson, 1996; Árnason *et al*, 1998, 2000; Árnason, 2003). Such sequence data are samples from a true but unknown genealogy that represents the relationships of the alleles in the population. The behaviour of such a genealogy can be studied by a stochastic process called a coalescent (Kingman, 1982), in particular time-inhomogeneous coalescent processes (Möhle, 2002) for an organism like cod showing multifurcating genealogies. Although we shall not attempt estimation of relevant coalescent parameters in this paper, we use the ideas of the coalescent to discuss qualitatively the topology of the genealogy and relative allele ages. These are important issues for inference from the polymorphism about population phenomena (Árnason, 2003).

The combined *cytb.cytbE.spacer* of the current paper, relative to the *cytb* of our previous work considerably increases the observed amount of variation. For the 250 bp *cytb*, a total of 59 distinct haplotypes were observed in a sample of 1278 individuals (Arnason, 2003), whereas here for a combined fragment of 566 bp, 34 haplotypes were found among 74 individuals. Thus, merely doubling the length of DNA sequence per individual considerably increases the amount of varia-

tion observed as is also reflected in the haplotype diversity. The inclusion of the noncoding TP spacer, having fewer selective constraints than the coding regions, also increased the per base nucleotide diversity.

An important aspect of the topology of the *cytb* gene tree (Árnason, 2003) was a shallow genealogy with four major haplotypes overlaid with rare haplotypes. The essential shallow structure was retained. However, important details were added. The branch lengths of some lineages increased somewhat. Almost all haplotypes representing internal nodes are present as real individuals in the sample and they are at a higher frequency than haplotypes representing terminal nodes. An example of such branch lengthening are haplotypes of the G-clade (ie haplotypes characterised by a $g \leftrightarrow t$ transversion at site 14691, Figure 3), where a mutation at 15375 in *cytbE* has been added, pushing the *G.E7.S1* and its derivative haplotypes one additional mutational step away from the most frequent *A.E1.S1* haplotype relative to the relationships observed at *cytb* only. Both changes probably occurred prior to the G rising appreciably in frequency or else they would be evident as separate branches on the genealogy. The shape of the genealogy contrasts with networks of mtDNA variation in the human gene pool, where haplotypes representing internal nodes frequently are either missing from the sample and are inferred or they are rare variants in the sample (see for example, Bandelt *et al*, 1995; Watson *et al*, 1997; Macaulay *et al*, 1999). Furthermore, haplotypes forming terminal nodes frequently are at a higher frequency in human mtDNA studies and human mtDNA haplotypes frequently differ by several mutations from their nearest neighbour. This contrast and the relatively shallow gene genealogy taken together imply a recent turnover and regeneration of the cod mtDNA polymorphism.

Another fact worth noting about the gene genealogy is that the short distance of a single mutational step between *A* and *E* (Árnason, 2003) was retained even with the added variation. Thus, the distance between the *A.E1.S1* and the *E.E1.S1* haplotypes also was one mutation. The effect of adding the extra sequence information was to add a number of rare (singleton) variants around *E.E1.S1*, a feature also seen for the other major haplotypes. This implies that the age of $t \leftrightarrow c$ transition at 14508 that characterises the *E* and *E.E1.S1* haplotypes, as well as the whole *E*-clade of haplotypes (Árnason, 2003) is very recent, at least relative to other high-frequency mutations in cod mtDNA. The *E* haplotype, however, is the second most common haplotype over all in the North Atlantic, second only to the *A* haplotype. The *A* haplotype is most common in Newfoundland and the *E* in the Baltic. The short genetic distance implies that the latter haplotype has risen in frequency very rapidly, so rapidly that it has not created divergent branches or lines. The high frequency of *E* in the Baltic in conjunction with its wide geographic distribution and its genealogical standing closest to *A* provides important insights into the forces responsible. To judge from its frequency and geographical distribution (this haplotype is likely to have originated in the Baltic (its likely centre of origin) and that it is currently spreading westwards and northwards by gene flow. Although we have found indications for weak selection purifying purine/pyrimidine bias, it is highly unlikely that there is positive selection operating directly on the

synonymous site defining the *E*-clade of haplotypes. Its rapid rise in frequency is unlikely to be due to positive selection of this kind. Since mtDNA is a single linkage group, it is possible that the *t*↔*c* transition at 14508 defining *E* is in full linkage disequilibrium with adaptive variation at another site on the chromosome, which could explain its rise in frequency. Alternatively, there might be sweepstakes reproduction (Hedgercock, 1994) generating a much higher than Poisson's variance in offspring number. This could drastically reduce effective population size leading to bottlenecks and increased drift of the mtDNA variation. Under this explanation, nuclear genes could be the targets of selection. However, mtDNA variation, such as the *E* haplotype, being neutral or nearly neutral, could be carried to a high frequency by numerous offspring of highly fit females winning reproductive sweepstakes (Árnason, 1998; Árnason *et al*, 1998; Árnason, 2003). Under this scenario, an mtDNA haplotype could be neutral in the physiology of the mitochondrion and not contribute to individual fitness, but its relatively rapid rise in frequency would be due to phenomena at the level of individuals.

Population variation

The mtDNA variation shows many criteria for a locus well suited for population studies (Smouse and Chevilion, 1998). It has a number of neutral or nearly neutral haplotypes at high frequencies and a host of rare haplotypes that provide a rich genealogical structure of variation that is useful for phylogeographic analysis (Avice, 1989). The high-frequency haplotypes provide genetic variance that could register population structuring. The observed number of mutations on the genealogy is high and if there was isolation of population changes in their frequency by random genetic drift could lead to further divergence of the populations. The mtDNA thus represents a polymorphic locus that ought to be able to register historical population differentiation (Jamieson and Jones, 1967; Møller, 1968) in cod if, indeed, it exists. The only caveat is that mtDNA is a single locus in this respect.

The results obtained from applying this locus to detect possible microspatial pattern of variation of cod in the Faroe Islands is simple. There is no spatial component to the variation. The variation is entirely among individuals within localities and neither localities nor overall areas account for any of the variation. All the sampling in this study was within areas considered as the Faroe Plateau (Jákupsstofu and Reinert, 1994). These results are similar to our results for *cytb* variation within other countries (Árnason and Pálsson, 1996; Árnason *et al*, 1998, 2000) that show no evidence for historical population structure within a country. They contrast with results obtained from microsatellite loci that show some microspatial structuring in the western Atlantic (Ruzzante *et al*, 1998; Beacham *et al*, 2000). Microsatellite data are not available for Faroese cod. However, only a weak structure (average $F_{ST}=0.0023$) thought to be due to passive transport of eggs or larvae was detected along the Norwegian coast (Knutson *et al*, 2003) on a geographical scale comparable to that studied here. Variation at the mtDNA locus, however, does register differentiation on an oceanic scale. The major mtDNA haplotypes show trans-Atlantic frequency clines and differentiation

among countries that is primarily due to the clinal frequency variation, as discussed elsewhere (Árnason, 1998, 2003).

Acknowledgements

We thank the many fishermen from the Faroe Islands, who provided us with samples and Eyðfinn Magnussen for help while collecting. Many students and staff at the Population Genetics laboratory at the Institute of Biology, University of Iceland provided technical and moral support for which we thank them. The work was supported by grants from the University of Iceland Research Fund and the Icelandic Science Foundation to Einar Árnason.

References

- Akashi H (1995). Inferring weak selection from patterns of polymorphism and divergence at 'silent' sites in *Drosophila* DNA. *Genetics* **139**: 1067–1076.
- Árnason E (1998). Mitochondrial cytochrome *b* DNA sequence variation of Atlantic cod, *Gadus morhua*. In: Hunt von Herbing I, Kornfield I, Tupper M, Wilson J (eds) *The Implication of Localized Fishery Stocks*. NRAES: Ithaca, NY. pp 129–137.
- Árnason E (2003). Mitochondrial cytochrome *b* DNA variation in the high fecundity Atlantic cod: trans-Atlantic clines and shallow gene-genealogy. *Genetics*. (in press).
- Árnason E, Pálsson S (1996). Mitochondrial cytochrome *b* DNA sequence variation of Atlantic cod, *Gadus morhua*, from Norway. *Mol Ecol* **5**: 715–724.
- Árnason E, Pálsson S, Arason A (1992). Gene flow and lack of population differentiation in Atlantic Cod, *Gadus morhua* L. from Iceland, and comparison of cod from Norway and Newfoundland. *J Fish Biol* **40**: 751–770.
- Árnason E, Pálsson S, Petersen PH (1998). Mitochondrial cytochrome *b* DNA sequence variation of Atlantic cod, *Gadus morhua*, from the Baltic and the White Seas. *Hereditas* **129**: 37–43.
- Árnason E, Petersen PH, Kristinsson K, Sigurgíslason H, Pálsson S (2000). Mitochondrial cytochrome *b* DNA sequence variation of Atlantic cod from Iceland and Greenland. *J Fish Biol* **56**: 409–430.
- Avice JC (1989). Gene trees and organismal histories: A phylogenetic approach to population biology. *Evolution* **43**: 1192–1208.
- Bakke I, Shields GF, Johansen S (1999). Sequence characterization of a unique intergenic spacer in mtDNA of Gadiformes. *Mar Biotechnol* **1**: 411–415.
- Bandelt H, Forster P, Sykes BC, Richards MB (1995). Mitochondrial portraits of human populations using median networks. *Genetics* **141**: 743–753.
- Beacham TD, Bratley J, Miller KM, Le KD, Schulze AD, Withler RE (2000). Population structure of Atlantic cod (*Gadus morhua*) in the Newfoundland and Labrador area determined from genetic variation. *Fisheries and Oceans Science. Canadian Stock Assessment Secretariat. Research Document 2000/099*. ISSN 1480-4883. Ottawa, 2000 Canada.
- Carr SM, Kivlichan DGS, Pepin P, Crutcher DC (1999). Molecular phylogeny of Gadid fishes: implications for the biogeographic origins of Pacific species. *Can J Zool* **77**: 19–26.
- Carr SM, Marshall HD (1991a). Detection of intraspecific DNA sequence variation in the mitochondrial cytochrome *b* gene of Atlantic cod (*Gadus morhua*) by the polymerase chain reaction. *Can J Fish Aquat Sci* **48**: 48–52.
- Carr SM, Marshall HD (1991b). A direct approach to the measurement of genetic variation in fish populations: applications of the polymerase chain reaction to studies of Atlantic cod (*Gadus morhua*). *J Fish Biol* **39**: 101–107.

- Carr SM, Snellen AJ, Howse KA, Wroblewski JS (1995). Mitochondrial DNA sequence variation and genetic stock structure of Atlantic cod (*Gadus morhua*) from bay and offshore locations on the Newfoundland continental shelf. *Mol Ecol* **4**: 79–88.
- Frydenberg O, Møller D, Nævdal G, Sick K (1965). Haemoglobin polymorphism in Norwegian cod populations. *Hereditas* **53**: 257–271.
- Hedgecock D (1994). Does variance in reproductive success limit effective population size of marine organisms?. In: Beaumont AR (ed) *Genetics and Evolution of Aquatic Organisms*. Chapman & Hall: London. pp 122–134.
- Jákupsstofu SH, Reinert J (1994). Fluctuations in the Faroe Plateau cod stock. *ICES Mar Sci Symp* **198**: 194–211.
- Jamieson A, Jones BW (1967). Two races of cod at Faroe. *Heredity* **22**: 610–612.
- Johansen S, Bakke I (1996). The complete mitochondrial DNA sequence of Atlantic cod (*Gadus morhua*): relevance to taxonomic studies among codfishes. *Mol Mar Biol Biotechnol* **5**: 203–214.
- Johansen S, Guddal PH, Johansen T (1990). Organization of the mitochondrial genome of Atlantic Cod, *Gadus morhua*. *Nucleic Acids Res* **18**: 411–419.
- Johansen S, Johansen T (1994). Sequence analysis of 12 structural genes and a novel non-coding region from mitochondrial DNA of Atlantic cod, *Gadus morhua*. *Biochem Biophys Acta* **1218**: 213–217.
- Jónsdóttir ÓDB, Imsland AK, Daniélsdóttir AK, Marteinsdóttir G (2002). Genetic heterogeneity and growth properties of different genotypes of Atlantic cod (*Gadus morhua* L.) at two spawning sites off south Iceland. *Fish Res* **55**: 37–47.
- King JL, Jukes TH (1969). Non-Darwinian evolution. *Science* **164**: 788–798.
- Kingman JFC (1982). On the genealogy of large populations. *J Appl Probab* **19A**: 27–43.
- Knutsen H, Jorde PE, André C, Stenseth NC (2003). Fine-scaled geographical population structuring in a highly mobile marine species: the Atlantic cod. *Mol Ecol* **12**: 385–394.
- Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V *et al* (1999). The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. *Am J Hum Genet* **64**: 232–249.
- McDonald JH, Kreitman M (1991). Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**: 652–654.
- Möhle M (2002). The coalescent in population models with time-inhomogeneous environment. *Stochastic Proc Appl* **97**: 199–227.
- Møller D (1968). Genetic diversity in spawning cod along the Norwegian coast. *Hereditas* **60**: 1–32.
- Mork J, Ryman N, Ståhl G, Utter FM, Sundnes G (1985). Genetic variation in Atlantic Cod (*Gadus morhua*) throughout its range. *Can J Fish Aquat Sci* **42**: 1580–1587.
- Pepin P, Carr SM (1993). Morphological, meristic, and genetic analysis of stock structure in juvenile Atlantic cod (*Gadus morhua*) from the Newfoundland shelf. *Can J Fish Aquat Sci* **50**: 1924–1933.
- Rand DM (2001). The units of selection on mitochondrial DNA. *Ann Rev Ecol Syst* **32**: 415–448.
- Ruzzante DE, Taggart CT, Cook D (1998). A nuclear basis for shelf- and bank-scale population structure in NW Atlantic cod (*Gadus morhua*) stocks: Labrador to Georges Bank. *Mol Ecol* **7**: 1663–1680.
- Ruzzante DE, Wroblewski JS, Taggart CS, Smedbol RK, Cook D, Goddard SV (2000). Bay-scale population structure in coastal Atlantic cod in Labrador and Newfoundland, Canada. *J Fish Biol* **56**: 431–447.
- Smouse PE, Chevillon C (1998). Analytical aspects of population-specific DNA fingerprinting for individuals. *J Hered* **89**: 143–150.
- Taylor WR (1986). The classification of amino acid conservation. *Theor Popul Biol* **119**: 205–218.
- Watson E, Forster P, Richards M, Bandelt H (1997). Mitochondrial footprints of human expansion in Africa. *Am J Hum Genet* **61**: 691–704.