Marine Genomics xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Marine Genomics

journal homepage: www.elsevier.com/locate/margen

Mitogenome sequence variation in migratory and stationary ecotypes of North-east Atlantic cod

Bård O. Karlsen [a,b,1,2], Åse Emblem [a,2], Tor E. Jørgensen [b], Kevin A. Klingan [b], Jarle T. Nordeide [b], Truls Moum [b], Steinar D. Johansen [a,b,*]

[a] Department of Medical Biology — RNA Lab, Faculty of Health Sciences, University of Tromsø, N-9037 Tromsø, Norway
[b] Marine Genomics Group, Faculty of Biosciences and Aquaculture, University of Nordland, N-8059 Bodø, Norway

ARTICLE INFO

Article history:
Received 4 October 2013
Received in revised form 11 December 2013
Accepted 6 January 2014
Available online xxxx

Keywords:
Atlantic cod
Marine genomics
mtDNA
Population mitogenomics
SOLiD sequencing

ABSTRACT

Sequencing of mitochondrial gene fragments from specimens representing a wide range of geographical locations has indicated limited population structuring in Atlantic cod (*Gadus morhua*). We recently performed whole genome analysis based on next-generation sequencing of two pooled ecotype samples representing offshore migratory and inshore stationary cod from the North-east Atlantic Ocean. Here we report molecular features and variability of the 16.7 kb mitogenome component that was collected from the datasets. These sequences represented more than 25 times coverage of each individual and more than 1100 times coverage of each ecotype sample. We estimated the mitogenome to have evolved 14 times more rapidly than the nuclear genome. Among the 365 single nucleotide polymorphism (SNP) sites identified, 121 were shared between ecotypes, and 151 and 93 were private within the migratory and stationary cod, respectively. We found 323 SNPs to be located in protein coding genes, of which 29 were non-synonymous. One synonymous site in *ND2* was likely to be under positive selection. F_{ST} measurements indicated weak differentiation in *ND1* and *ND2* between ecotypes. We conclude that the Atlantic cod mitogenome and the nuclear genome apparently evolved by distinct evolutionary constraints, and that the reproductive isolation observed from whole genome analysis was not visible in the mtDNA sequences.

1. Introduction

Knowledge on population structures is important in order to preserve fishery resources, local populations, and to assess the genetic pool of marine fish species. The Atlantic cod is an important species for fisheries and has in recent years become attractive for aquaculture (Johansen et al., 2009). Atlantic cod is widely distributed across the North Atlantic Ocean and complex patterns of genotypic and phenotypic variation have been revealed in recent years by means of single nucleotide polymorphisms (SNPs), microsatellites, mitochondrial genomes (mitogenomes; mtDNA), allozymes, restriction fragment length polymorphism loci, and phenotypic characters (e.g. Pogson et al., 1995; Árnason, 2004; Pogson and Mesa, 2004; Coulson et al., 2006; Carr and Marshall, 2008; Moen et al., 2008; Johansen et al., 2009; Hubert et al., 2010). Apparently, the population genetic structuring in Atlantic cod is both due to isolation by distance and vicariance, as well as adaptive variation along environmental clines and adaptive divergence among sympatric populations (e.g. Nielsen et al., 2009; Bradbury et al., 2010, 2013; Karlsen et al., 2013). The status of two phenotypically divergent ecotypes of cod that occur sympatrically in the North-east Atlantic, the offshore migratory North-east Arctic (NA) cod and inshore stationary Norwegian coastal (NC) cod, has been addressed by a number of studies and evidence concerning their connectedness is still incomplete (reviewed in Nordeide et al., 2011). Differential markers of NA and NC cod have been reported at blood type E, hemoglobin (*Hb* I) and pantophysin (*Pan* I) loci, while few significant differences have been found in microsatellite and SNP data, although temperature associations have been coupled with some SNPs supporting an adaptive diversity (Bradbury et al., 2010; Andersen, 2012).

The mitogenome of Atlantic cod is small (about 16.7 kb) and circular, and holds a gene content identical to that of most vertebrates (Johansen et al., 1990, 2009; Johansen and Bakke, 1996). The use of mitochondrial gene sequences as biomarkers for intra- and interspecies studies has become popular. Presumably near-neutral evolution, maternal inheritance, and a small compact size have made mitogenomes an easy-to-use population tool. However, in the last years it has been argued that mitogenomes may be less suited for vertebrate population history assessments due to increasing evidence of recombination events, positive selection, erratic evolutionary rate, and heteroplasmy (e.g. Brown et al., 1979; Bazin et al., 2006; Mjelle et al., 2008; Galtier et al., 2009; He et al., 2010; Ameur et al., 2011). Árnason (2004) examined a

* Corresponding author at: Department of Medical Biology — RNA Lab, Faculty of Health Sciences, University of Tromsø, N-9037 Tromsø, Norway.
E-mail address: steinar.johansen@uit.no (S.D. Johansen).
[1] Current address: Department of Laboratory Medicine, Division of Molecular Diagnostics, Nordland Hospital Trust, N-8092 Bodø, Norway.
[2] Contributed equally.

1874-7787/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.margen.2014.01.001

250 bp segment of the mitochondrial *CytB* from almost 1300 Atlantic cod specimens representing the species' range in the North Atlantic Ocean. A shallow genealogy and lack of clear population structuring were detected, and he found that sequence variation was mainly due to synonymous substitutions and that limited selection was operating on these substitutions. Marshall et al. (2009) studied the complete mtDNA of 32 cod specimens and found that variation is consistent with mutation rate heterogeneity across the mitogenome and complies with the nearly neutral theory of molecular evolution. Elevated substitution rates were found in ND genes, but the rate heterogeneity among genes was mainly due to changes at synonymous sites (Marshall et al., 2009).

Sequencing of pooled samples of specimens offers an attractive approach to investigate mutation rate and heterogeneity of mitogenomes within and between populations, while the genealogical information content of individual haplotypes is sacrificed (Futschik and Schlotterer, 2010; Kofler et al., 2011). Next-generation sequencing (NGS) is well suited for such analyses and reveals sequence information of a number of mitochondrial genomes at the same time, keeping labor and running cost at a minimum. The pooled sample analyses of mitogenomes presented here is part of a population genomic project that apply NGS in assessing genetic diversity and divergence between the migratory NA and stationary NC cod ecotypes. Recently we reported the divergence in nuclear DNA of the NA and NC cod ecotypes, where mitochondrial genome sequences were not included in the genome assembly (Karlsen et al., 2013). Here we aim to address the utility and limitation of mitogenome sequences as a biomarker in population assessments. Each ecotype was represented by 44 specimens, which were pooled and deep sequenced by SOLiD ligation sequencing (Karlsen et al., 2013). Reads originating from the mitogenomes were mapped to a Sanger sequenced reference Atlantic cod mtDNA. Genetic variation and differentiation between ecotypes were then calculated from variation at SNP loci applying BayeScan and the PoPoolation toolbox.

2. Materials and methods

2.1. Sample collection, DNA isolation, and DNA sequencing

These procedures are essentially as described by Karlsen et al. (2013). In summary, muscle tissue was collected from 44 Atlantic cod specimens captured in the Barents Sea north of 72°N (migratory NA cod) and 44 specimens from Holandsfjord, 66° 71′N 13° 68′E (stationary NC cod). DNA was isolated applying Roche High Pure PCR Template Preparation Kit and Nexttec™ Genomic DNA Isolation Kit for Tissue and Cells according to manufacturers' specifications. Purified total DNA from each individual was quality verified, and equal amounts (100 ng) from each individual in one ecotype were titrated, pooled, and sheared to an average size of 125 bp with a Covaris S2 system. SOLiD Fragment libraries were prepared according to the standard protocol given by the manufacturer and sequencing performed on a SOLiD 3 Plus system (Applied Biosystems). The two libraries were sequenced on separate SOLiD runs. The NC run produced 298,850,963 usable reads and the NA run produced 328,107,403 usable reads (Table 1). Sanger sequencing of the complete mitogenome isolated from muscle tissue from the Norwegian coastal cod specimen NC3 (69° 44′N 19° 37′E) was prepared and performed according to the previously described protocol (Ursvik et al., 2007).

2.2. Data analysis

The NC3 Atlantic cod mitogenome (16,696 bp) was used as a reference for mapping of the NA and NC pooled mitochondrial genome reads. Reference mapping was performed in the SOLiD Bioscope Software 1.3 with standard settings for fragment libraries. The CLC Genomics Workbench 4.7.2 was used to create SNP statistics for mapping of pooled samples with the following stringency, and for each position there had to be at least 400 reads mapping to the reference nucleotide. For the setting of quality window length 11 (default), max gaps 2 (default), minimum quality of central base 25 (default: 20), minimum average quality of surround bases 25 (default: 15), and minimum variant frequency 2%. Gff3 files were created with GMAP version 2012-01-11 (Wu and Watanabe, 2005) for the mitochondrial genes mapping to NC3. SNPs were then extracted from Bam files dictated by the gff3 files.

Table 1
Summary statistics for SOLiD sequencing and mitogenome mapping.

	NA ecotype[a]	NC ecotype[b]
Total raw reads for mapping[c]	328,107,403	298,850,963
NC3 reference length (bp)[d]	16,696	16,696
Total mapped reads	497,307	387,233
Average mapped read length (nt)	46.81	47.52
Average coverage	1,394	1,102

[a] Migratory NA cod (n = 44).
[b] Stationary NC cod (n = 44).
[c] From Karlsen et al. (2013).
[d] Sanger sequenced NC3 mitogenome (HG514359) applied for mapping in CLC Genomic Workbench.

2.3. F_{ST} statistics and BayeScan

F_{ST} measures the genetic differentiation between populations, with values ranging from 0 to 1. If the F_{ST} is high, the allele frequencies are highly differentiated while low F_{ST} signifies little differentiation. Large populations with migration generally show low F_{ST} values while small populations with little migration tend to show higher F_{ST} values (Holsinger and Weir, 2009). We created an R script (v 2.13.1, R Development Core Team, 2011), provided in Karlsen et al. (2013), to calculate F_{ST} value for each of the detected SNPs. To identify outlier loci putatively under selection, the Bayesian approach implemented in the BayeScan Version 2.01 (Foll and Gaggiotti, 2008) was applied (default settings). This method is based on the finite-island model of Beaumont and Nichols (1996), and infers significance from the posterior probability of a locus.

2.4. Tajima's D and π analysis

Tajima's D is a statistical test that aims to distinguish between neutral evolution of a locus (genetic drift) versus a non-random process like natural selection. It compares the average number of nucleotide differences between pairs of sequences to the total number of segregating sites. A neutrally evolving locus in a population in equilibrium, where the allele frequencies remain constant over time, will have a D value of approximately zero. A locus with increased genetic variation will generate a positive Tajima's D value and indicates balancing selection, while a locus with decreased genetic variation gives a negative D value and indicates purifying selection. Tajima's D is also influenced by demographic processes such that a population reduction or bottleneck can give the same locus a transient positive value, while a population expansion (i.e. after a bottleneck) can create a low or negative D value (Tajima, 1989; Oleksyk et al., 2010). The Tajima's π (nucleotide diversity) is the mean number of pairwise differences. Tajima's D and π analyses were performed applying the PoPoolation toolbox (Kofler et al., 2011), with the pileup files created with SAMtools (Li et al., 2009) as instructed in the PoPoolation manual. Window size was set to 30 nt and step size to 15 nt. We used Gnuplot Version 4.4 patchlevel 3 to plot the genome-wide scan of Tajima's D and π. The Gnuplot scripts for statistic plotting were taken from the webpage: http://www.phyast.pitt.edu/~zov1/gnuplot/html/statistics.html.

Please cite this article as: Karlsen, B.O., et al., Mitogenome sequence variation in migratory and stationary ecotypes of North-east Atlantic cod, Mar. Genomics (2014), http://dx.doi.org/10.1016/j.margen.2014.01.001