Available online at www.sciencedirect.com
ScienceDirect
Gene 383 (2006) 12–23
GENE
SECTION
EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

The complete mitochondrial genome of the whiting, *Merlangius merlangus* and the haddock, *Melanogrammus aeglefinus*: A detailed genomic comparison among closely related species of the Gadidae family

Severine Roques [a], Clive J. Fox [b], Maria I. Villasana [a], Ciro Rico [a,*]

[a] *Estación Biológica de Doñana, CSIC, Pabellón del Perú, 41013 Sevilla, Spain*
[b] *The Centre for Environment, Fisheries and Aquaculture Science, Pakefield Road, Lowestoft, Suffolk, NR33 OHT, England, United Kingdom*

Received 2 June 2005; received in revised form 20 June 2006; accepted 29 June 2006
Available online 14 July 2006
Received by G. Pesole

Abstract

We determined the first complete mitochondrial DNA (mtDNA) sequences for the whiting (*Merlangius merlangus*, family Gadidae, order Gadiformes) and the haddock (*Melanogrammus aeglefinus*, family Gadidae, order Gadiformes). The entire mitogenomes were amplified and sequenced by primer walking using newly designed specific internal primers. Lengths were 16,569 and 16,585 bases for whiting and haddock respectively, lengths which lie within the range of previously reported gadiform sequences from Atlantic cod (*Gadus morhua*, 16,696 bases) and walleye pollock (*Theragra chalcogramma*, 16,570 bases). Gene arrangement in both species conformed to the order seen in most vertebrate mitochondrial genomes. We identified a long intergenic spacer located between the $tRNA^{Thr}$ and $tRNA^{Pro}$ genes (of 100 and 70 bp long for whiting and haddock, respectively), as previously described for other species of the order Gadiformes. Using nucleotide and amino acid divergence data of four complete gadoid mitogenomes (*M. merlangius*, *M. aeglefinus*, *G. morhua* and *T. chalcogramma*), we examined in detail the relative mtDNA mutation patterns across genes and among Gadidae species and tested for the performance of each protein-coding, transfer RNA and ribosomal RNA gene in depicting the expected phylogeny among the four species, as compared with the whole genome dataset. This comparison may be particularly useful in phylogenetic analyses of such a diverse fish family, as well as for the understanding of the patterns of nucleotide substitution of the mtDNA at low levels of divergence.

Keywords: Mitochondrial DNA; Complete sequence; Fish; Protein-coding genes

1. Introduction

From the numerous studies, it appears that the optimal resolution of a particular mitochondrial DNA (mtDNA) gene for population or phylogenetic inferences relies on an optimal combination of the mutation rate of the portion analysed and on the level of polymorphism studied. Because different regions of the mt genome evolve at different rates, certain regions of the mtDNA have been commonly targeted for particular types of study, e.g. phylogeny, phylogeography or population structure (Avise, 1986). Thus, in phylogenetic studies, one usually assumes that a single portion of mtDNA (generally 300–600 base pairs) is representative of both the complete molecule and the maternal evolutionary relationships of the organism studied. However, increasing evidence from gene-comparison studies has demonstrated that the phylogenetic performance may vary among different genes. A single gene may not be representative of the whole mitogenome and can provide a misleading inference of the true inter-specific evolutionary relationships (Cao

Abbreviations: A, adenosine; aa, amino acid (s); C, cytosine; CR, control region; CSB, conserved block sequence; DHU, dihydrouracil; G, guanosine; mt, mitochondrial; MRP, mean ranking position; NC, non-coding; ORF, open reading frames; PCG, protein-coding gene (s); rRNA, ribosomal RNA; T, thymidine; TAS, termination-associated sequence; tRNA, transfer RNA; U, uridine.

* Corresponding author. Tel.: +34 954232340; fax: +34 954621125.
E-mail address: ciro@ebd.csic.es (C. Rico).

0378-1119/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2006.06.018

et al., 1994; Meyer, 1994; Cummings et al., 1995; Russo et al., 1996; Zardoya and Meyer, 1996; Miya and Nishida, 2000). Moreover, the relationship between gene mutation rate and our ability to recover the true evolutionary relationships may not be straightforward. Obtaining complete genomic sequences from extensive taxonomic sampling may overcome these problems, thus improving estimates of tree topology and dates of divergence among clades (Pollock et al., 2000). In addition, these data may subsequently aid in choosing the appropriate genetic marker(s) for a given level of divergence (e.g. shallow or deep phylogenies or population genetics) among related species.

The large amount of recently available sequence data indicates that fish mtDNA is structurally very similar to that of other vertebrates with 22 tRNA genes, 13 protein-coding genes and a main non-coding region (control region) with common features such as TAS (termination-associated sequences) and CSB (conserved block sequences) (Inoue et al., 2000; Delarbre et al., 2001; Lee et al., 2001; Miya et al., 2001; Doiron et al., 2002; Miya et al., 2003; Kim et al., 2004; Manchado et al., 2004). Among approximately 140 complete sequences, however, nine patterns of gene rearrangement were found in euteleosts (Miya and Nishida, 1999; Inoue et al., 2001; Miya et al., 2001, 2003). Also, Miya and Nishida (1999) reported the first example of tRNA gene rearrangement in bony fishes between the ND6 and the control region (ND6–cytb–$tRNA^{Glu}$–$tRNA^{Pro}$–$tRNA^{Thr}$–control region), changing the "conserved" status of fish mt genome. Although complete genomes have been particularly helpful in resolving phylogenetic relationships, subsequent sequencing and analyses may still be necessary for resolving persistent controversies over higher-level relationships within teleosts which are the most diversified group of all vertebrates.

The order Gadiformes comprises more than 450 species in 21 genera, arranged in 11 families: Bregmacerotidae, Euclichthyidae, Gadidae, Lotidae, Macrouridae, Macruronidae, Melanonidae, Merlucciidae, Moridae, Muraenolepidae and Muraenolepididae (Cohen et al., 1990). The family Gadidae is a large family with 15 known species in 12 genera and includes a large number of commercially important species, such as saithe, pollock, cod, haddock and whiting. Complete mtDNA sequences have been reported for six species of Gadiformes [*Lota lota*, family Lotidae, *Caelorinchus kishineryei*, family Macrouridae, *Pysiculus japonicus*, family Moridae *Bregmaceros nectabanus*, family Bregmacerotidae, *Gadus morhua* (Johansen and Bakke, 1996) and *Theragra chalcogramma*, family Gadidae (Yanagimoto et al., 2004; Miya et al., 2003)]. Further, special features of Gadiformes mt genomes have also been reported. For example, Miya et al. (2001) reported the presence of a unique mitogenome *B. nectabanus*, characterized by many rearrangements, such as the unusual position of the cytb and some tRNAs ($tRNA^{Ala}$, $tRNA^{Asn}$, $tRNA^{Cys}$, $tRNA^{Ser}$, $tRNA^{Glu}$, $tRNA^{Thr}$) genes. Also, Bakke et al. (1999) determined the unique presence of intergenic spacers between the $tRNA^{Thr}$ and $tRNA^{Pro}$ in the mt DNA of *G. morhua*, *T. chalcogramma*, *Boreogadus saida*, *Melanogrammus aeglefinus*, *Micromesistius poutassou*, *Brosme brosme*, *Enchelyopus cimbrius* and *Gaidropsarus argentatus*.

Among the gadiform taxa, genetic variation of the cytochrome *b* gene has proved very useful for inter-specific evolutionary relationships, but has given incoherent results within some genera of the family Gadidae for which the monophyly is still debated (Carr et al., 1999; Moller et al., 2002). Lack of genetic resolution may be the result of the low performance of the gene analysed to describe several speciation events within a relatively short period of time (Carr et al., 1999). However, the recent alignment of ten gadiform mt ribosomal and cytochrome *b* genes has clarified some of the unresolved phylogenetic relationships, corroborating that the Gadidae is the most derived group, with the most recent genus being *Gadus*, *Boreogadus*, *Merlangius*, *Melanogrammus* and *Pollachius* (Bakke and Johansen, 2005). Although at the intra-specific level mtDNA genetic variation has been extensively studied in Atlantic cod (*G. morhua*) (Carr and Marshall, 1991; Carr et al., 1995; Arnason and Palsson, 1996; Arnason et al., 1998, 2000; Sigurgislason and Arnason, 2003; Arnason, 2004) very little is known on most other commercial gadoid species.

The whiting (*Merlangius merlangus*) and the haddock (*M. aeglefinus*) belong to the family Gadidae and are among the most abundant and commercially important gadoid fish species of the North Atlantic. There are marked differences in geographic distribution with haddock widely distributed across both sides of the Atlantic. In contrast, whiting occur only on the eastern shelves but penetrate further south reaching the Mediterranean Sea (Wheeler, 1978). Based on mitochondrial cytochrome *b* and cytochrome oxidase I genes, previous phylogenetic studies concluded that both species are closely related and sister taxa to the more recently diverged group of cod-like and pollock species (Carr et al., 1999; Moller et al., 2002; Bakke and Johansen, 2005). At the intra-specific level, little resolution was found in the haddock mt DNA variation (12 restriction endonucleases), either because the fragment was too short to contain sufficient variability or because the gene was not necessarily suitable for the purpose of this study (Zwanenburg et al., 1992).

Our study describes for the first time, the complete mtDNA sequences of whiting and haddock. Gene arrangement and basics statistics are given providing information on the extent of polymorphism and divergence between these closely related species. The mtDNA genomes of these species will be an important source of sequence information for improving primer selection and design and for the identification of diagnostic regions suitable for species-specific identification, particularly of early stage eggs from plankton surveys (Taylor et al., 2002; Fox et al., 2005). Using nucleotide and amino acid divergence data of the four complete gadoid mitogenomes so far sequenced (whiting, haddock, this study, and Atlantic cod and walleye pollock, from GenBank), we examined the relative mtDNA mutation patterns across genes and among gadoid species. In addition, we tested for the performance of each protein-coding gene (PCG), the combined transfer RNA genes and the two ribosomal RNA subunits in depicting the expected phylogeny among the four species, as compared to the whole genome and the combined PCG dataset. This should be of particular use for phylogenetic studies in such a diverse species family, as well as