Biodiversity genomics for species at risk: patterns of DNA sequence variation within and among complete mitochondrial genomes of three species of wolffish (*Anarhichas* spp.)

K.A. Johnstone, H.D. Marshall, and S.M. Carr

Abstract: The first marine fish species to be listed under the Canadian Species At Risk Act as Threatened with extinction are the spotted wolffish (*Anarhichas minor* Olafsen, 1774) and the broadhead wolffish (*Anarhichas denticulatus* Krøyer, 1844); a third species, the striped wolffish (*Anarhichas lupus* L., 1758), is listed as a species of special concern. As part of the recovery plan for wolffish, we determined the complete mitochondrial DNA (mtDNA) genome sequences of all three species to identify the most variable gene regions for population analysis. *Anarhichas* genomes comprise either 16 519 or 16 520 base pairs (bp), among which there are 449 single nucleotide polymorphisms (SNPs). The most variable protein-coding loci are ND4, CYTB, and ND2, with 4.40, 4.22, and 4.19 SNPs / 100 bp, respectively. Comparisons of rates of synonymous and nonsynonymous nucleotide substitutions indicate no evidence of selection. The control region, characterized in many species as hypervariable, was less variable than 9 of 13 protein-coding loci (2.45 SNPs / 100 bp). Phylogenetic analysis shows that *A. lupus* and *A. minor* are more closely related to each other than either is to *A. denticulatus*. Amplification and sequence analysis of a contiguous block of 6392 bp that spans the ND4, ND5, ND6, and CYTB loci is an efficient strategy for evaluating patterns of intraspecific mtDNA variability.

Résumé : Les premières espèces de poissons marins à être placées sur la liste des espèces « menacées » d'extinction selon la Loi canadienne sur les espèces en péril sont le loup tacheté (*Anarhichas minor* Olafsen, 1774) et le loup à tête large (*Anarhichas denticulatus* Krøyer, 1844); une troisième espèce, le loup atlantique (*Anarhichas lupus* L., 1758), est incluse comme espèce « préoccupante ». Comme contribution au plan de récupération des loups, nous avons déterminé les séquences du génome complet de l'ADN mitochondrial (ADNmt) chez les trois espèces afin d'identifier les régions les plus variables en vue d'une analyse de population. Les génomes d'*Anarhichas* comprennent ou 16 519 ou 16 520 paires de bases(pb), parmi lesquelles il y a 449 polymorphismes simple nucléotide (SNPs). Les locus codant pour les protéines les plus variables sont ND4, CYTB et ND2 qui possèdent respectivement 4,40, 4,22 et 4,19 SNPs / 100 pb. La comparaison des taux de substitutions synonymes et non synonymes de nucléotides ne révèle aucun signe de sélection. La région de contrôle, caractérisée d'hypervariable chez de nombreuses espèces, est moins variable que 9 des 13 locus codant pour les protéines (2,45 SNPs / 100 pb). L'analyse phylogénétique montre que les *A. lupus* et *A. minor* sont plus apparentés l'un à l'autre que chacun ne l'est avec l'*A. denticulatus*. L'amplification et l'analyse des séquences d'un bloc contigu de 6392 pb qui comprend les locus ND4, ND5, ND6 et CYTB constituent une stratégie efficace pour l'évaluation des patrons de variabilité interspécifique de l'ADNmt.

[Traduit par la Rédaction]

Introduction

The wolffish family Anarhichadidae is a small family of blenny-like marine fishes found on rocky and hard bottom areas of the northern oceans (O'Dea and Haedrich 2002). Wolffish are sedentary, slow-growing fish with stout bodies and large, blunt heads (Templeman 1986*a*, 1986*b*). Their dentition comprises large, conical anterior teeth and molariform lateral teeth, which gives them a wolf-like appearance and enables them to feed on crustaceans and invertebrates (Scott and Scott 1988). There are two genera, the monotypic Pacific wolfeel (*Anarrhichthys ocellatus* Ayres, 1855) and *Anarhichas* L., 1758, which comprises four species — the Bering wolffish (*Anarhichas orientalis* Pallas, 1814) and three species in the North Atlantic Ocean, the striped or Atlantic wolffish (*Anarhichas lupus* L., 1758), the spotted wolffish (*Anarhichas minor* Olafsen, 1774), and the broadhead or northern wolffish (*Anarhichas denticulatus* Krøyer, 1844). *Anarhichas lupus* are recognized by a series of dark transverse bars, whereas *A. minor* and *A. denticulatus* are covered in irregular blackish brown spots, which are somewhat larger in the former than in the latter. All three species occur on both sides of the Atlantic Ocean and into Arctic waters (Fig. 1).

Although not recently the targets of directed fisheries, wolffish populations and habitats have been heavily impacted by bottom-trawling associated with the fishery for

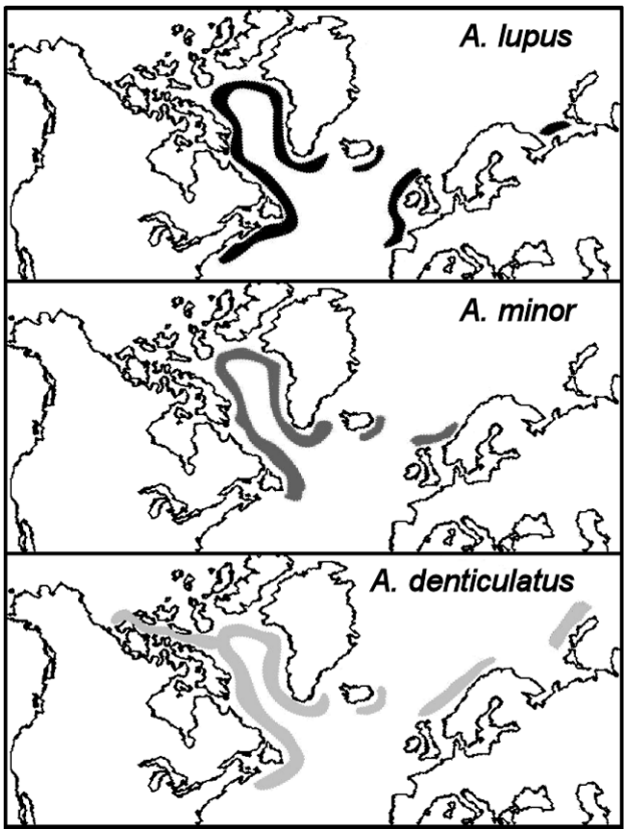
Received 15 March 2006. Accepted 21 November 2006. Published on the NRC Research Press Web site at http://cjz.nrc.ca on 22 February 2007.

K.A. Johnstone, H.D. Marshall, and S.M. Carr.[1] Genetics, Evolution, and Molecular Systematics Laboratory, Department of Biology, Memorial University of Newfoundland, St. John's, NL A1B 3X9, Canada.

[1]Corresponding author (e-mail: scarr@mun.ca).

 doi:10.1139/Z06-191

Fig. 1. Distribution of three species of wolffish from the North Atlantic Ocean: from top to bottom, striped or Atlantic wolffish (*Anarhichas lupus*), spotted wolffish (*Anarhichas minor*), and broadhead or northern wolffish (*Anarhichas denticulatus*) (maps drawn from descriptions in Scott and Scott 1988).



Atlantic cod (*Gadus morhua* L., 1758). Although this decreased after the Atlantic cod moratorium in 1992 (O'Dea and Haedrich 2002), *A. minor* and *A. denticulatus* are still caught in directed fisheries for halibut and crab, and almost 1500 t were taken annually as bycatch between 2000 and 2002 (Department of Fisheries and Oceans 2004). Wolffish not caught may be injured by groundfish trawls, which dislocate the rocks and boulders where wolffish live, nest, and spawn. Dredging for scallops and clams also promotes resuspension of bottom sediments, and can destroy spawning habitats and damage fish gills (O'Dea and Haedrich 2002).

Indices of wolffish abundance declined by more than 90% on the Grand Banks and the Northeast Newfoundland and Labrador Shelf from the late 1970s to the early 1990s, about two wolffish generations (O'Dea and Haedrich 2002; Department of Fisheries and Oceans 2004). Consequently, *A. minor* and *A. denticulatus* became in 2001 the first marine species to be assessed by the Committee on the Status of Endangered Wildlife in Canada (COSEWIC) as Threatened with extinction, and were accordingly listed on schedule 1 of the Species At Risk Act (SARA) by the Minister in 2005. *Anarhichas lupus* were listed in 2000 as a species of Special Concern and *A. orientalis* were listed in 2002 as being Data Deficient (i.e., the available data are insufficient to assign a status).

Wolffish in the North Atlantic Ocean are currently the subjects of a Department of Fisheries and Oceans (now Fisheries and Oceans Canada) species recovery plan that is intended to identify and correct the factors responsible for the observed decline. This plan includes a genetic analysis of their population structure to determine whether any of the species comprises distinctive subcomponents consistent with a stock structure (cf. Pepin and Carr 1993) or designatable units under SARA. Preliminary genetic studies of mitochondrial DNA (mtDNA) sequence variation in more than 400 wolffish from all three species showed very little variation within the species in the control region (K.A. Johnstone, H.D. Marshall, and S.M. Carr, unpublished observation), a gene region considered "hypervariable" in many fish species (Faber and Stepien 1997). These data gave no indication of population structure within species, and pairwise genetic differentiation among species was small, which is indicative of recent speciation that might also explain the low intraspecific variation.

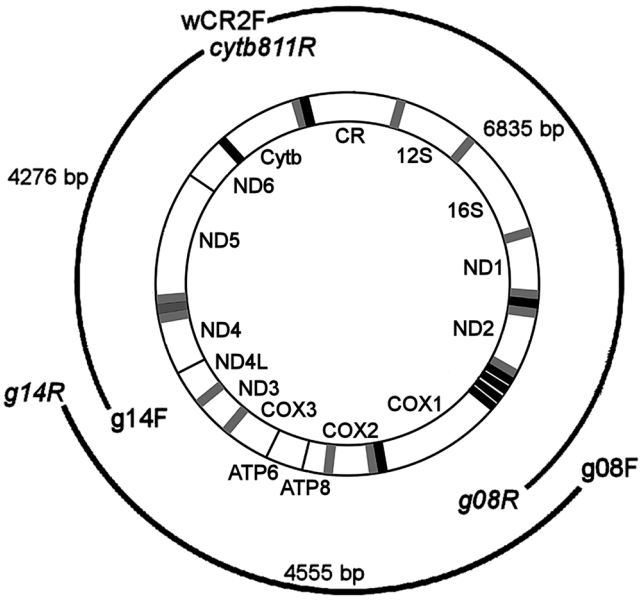
To clarify the evolutionary, population, and conservation genetic status of wolffish, we obtained the complete mtDNA genome sequences of all three species to identify loci most suitable for population genetic analysis. Interspecific divergence should predict intraspecific variability at selectively neutral loci, as both are functions of the neutral mutation rate (Kimura 1983). Along with comparative genomic data on codfish, including *G. morhua*, which has also been assessed by COSEWIC as Endangered, this study represents the first genomic study of Canadian species at risk under SARA.

Materials and methods

Frozen hearts were collected at sea by Department of Fisheries and Oceans' personnel from bycatch fish of all three species. Preliminary experiments indicated that DNA obtained from most of these samples consisted of fragments of low molecular weight. To obtain high molecular weight DNA for long-range polymerase chain reactions (PCR; see below), we obtained a live *A. lupus* under a Department of Fisheries and Oceans' incidental harm permit at the Bonne Bay Marine Station on the west coast of the island of Newfoundland (Northwest Atlantic Fisheries Organization (NAFO) division 4R). Fresh heart, muscle, and liver tissues were collected from the fish immediately after being euthanized. DNA from frozen specimens was obtained from individual *A. minor* and *A. denticulatus* from NAFO divisions 30 and 3L, respectively. mtDNA was extracted with the QIAamp DNA Mini Kit (QIAGEN, Valencia, California), and the tissue protocol provided by the manufacturer was followed.

Initial studies of *Anarhichas* mtDNA employed the "universal" primers of Kocher et al. (1989) for the CYTB and control region (CR) loci. Of the 18 primer pairs designed by Coulson et al. (2006) for amplification and sequencing of *Gadus* and related gadiforms, three pairs successfully amplified *Anarhichas* spp.: a fish primer wCR2F and a mammalian primer *Cytb811R* for a region spanning the 3′ end of the CYTB locus and the 5′ end of the CR locus, g14F and *g14R* for part of the ND4 locus, and g08F and *g08R* for part of the COX1 locus. By pairing *Cytb811R* with g14F, *g14R* with g08F, and *g08R* with wCR2F, and modifying

Fig. 2. Schematic diagram of the wolffish mitochondrial genome, with three regions >4 kbp amplified by long-range PCR. Forward primers are in roman type and reverse primers are in italic type. The 466 bp fragment of the genome between the 5′ ends of the *cytb811R* and wCR2F primers spans the 3′ end of the CYTB locus and the 5′ end of the control region (CR) locus, for which the sequence was already known.



standard PCR protocols to amplify fragments >5 kbp, we were able to amplify the entire genome (except for a 1 kbp portion that included the CR locus, the sequence of which was already known) of *A. lupus* as three overlapping fragments of >4 kbp each (Fig. 2). In the process of DNA sequencing, 42 *Anarhichas* specific primers were designed and were used along with 19 *Gadus* specific primers and 1 mammalian primer (Table 1).

For long-range PCR amplification of mtDNA from *A. lupus*, the TripleMaster system and High-Fidelity PCR protocol were used (Eppendorf, New York, New York). The reaction mixture was prepared in two parts. The first mixture comprised 8.4 μL of distilled deionized water (ddH_2O), 0.3 μL each of the 10 μmol/L forward and reverse primers, and 1 μL of the template DNA. The second mixture comprised 7.4 μL of ddH_2O, 2 μL of 10× High-Fidelity buffer with Mg^{2+}, 0.4 μL of 40 nmol/L dNTP, and 0.2 μL of TripleMaster polymerase mix. The two mixtures were combined and immediately preheated to 94 °C for 2 min to denature the double-stranded DNA template. The PCR cycle was 94 °C for 20 s, 54 °C for 15 s, and 68 °C for 4 min, which was repeated for 35 cycles. The reactions were thereafter cooled to 5 °C. Size of amplicons were estimated by electrophoresis of 4 μL of the PCR product through 1% or 2% agarose gel with 8 μL of ethidium bromide in 1× TBE buffer at 60 V, for up to 6 h according to the expected sizes. PCR products were purified with the QIAGEN QIAQuick PCR purification kit according to the manufacturer's directions.

For low molecular weight DNA from *A. minor* and *A. denticulatus*, standard PCR reactions for amplicons with expected sizes <1 kbp were done according to previously published methods (Coulson et al. 2006).

Table 1. Primers used to amplify and (or) sequence the mitochondrial genomes of *Anarhichas* species.

Primer	DNA sequence (5′→3′)	5′ base
g20R	GGCAGGACATTAAGGGCATTCTCAC	29
wCR1F	ACCTCCCACCCCTAACTCCCAAAGC	53
wCR1R	GGCATGCTGGGTTATCTCGCTTATG	161
wCR2F	ATTCCTGACATTTGGCTCCTACTTC	176
wCR2R	CTAGGGCCCATCTTAACATCTTCAG	739
wCR2F1	TGGTATCAGGCACATCTCTAGTGAG	870
g01F	CTGAAGATATTAGGATGGACCCTAG	1 358
g02F	CCAAAAACGTCAGGTCGAGGTGTAG	1 499
g02R	CTATTCATTTCACAGGCAACCAGCT	2 019
w02F	TGTTCCGCTGAAATTGGCCCTGAAG	2 176
g03F	ACCCCGAAACTGAGCGAGCTACTCC	2 618
g03R	TAAGCCCTCGTGATGCCATTCATAC	2 841
g04F	TTTACCAAAAACATCGCCTCTTG	3 276
g04R	TGAACCTCTGTAGAAAGGGCTTAGG	3 954
wO4F	TTCAGACCGGAGTAATCCAGGTCAG	4 020
g05R	ATGTTCGGGGTATGGGCCCAAGAGC	5 026
w05F	AGTATGCTCTTATCGGAGCCCTTCG	5 107
wO6F	GTGCTTCCACTACACCACTTCCTAG	5 225
w06R	GATGGCGACAGGAAGAGTAAGTTGC	5 293
g06R	AGCTTAATTAAAGTATTTGTTTTGC	5 931
w07R	CTGGTTTGAGCGCTTAGCTGTTAAC	6 296
g07F	AAACTAGACCAAGGGCCTTCAAAGC	6 407
w08R	CAGAGGTAAAGTAAGCGCGTGTGTC	6 950
w08F1	CCCTTCACCTAGCAGGAATTTCTTCAATCC	7 165
g08F	ATGGGTATAGTCTGAGCTATGATGG	7 498
g08R	TAACCCACAATTCTGCCTTGACAAG	7 908
w08F	ACAACGAATGTGGAGTGACTACACG	8 473
g09R	ACCCATATTAGCTTCTTAGTGAGG	8 585
w09F	GGCCATCAGTGGTACTGAAGCTATG	9 169
g10R	AGAGGGCGAATGAATAAACTAATTG	9 904
w10F1	ATGCGAAACCAACCAACCCATGCTC	10 274
w11R	TTGATCTCCTCAGGGTAGCGGGAGTAGTAG	10 459
w11F	CTACACTTGACCCATTTGAAGTGCC	10 812
w12R1	TGAGTGAGAGTTGTGGGTCACAGT	11 044
w13R	TACCTGCGTTTAGTCGTTCGGTTTG	11 258
g13F	CTTTCTCCGCTTGTGAAGCAAG	11 594
w14R	ACCCATATGGCTTACGGAGGAGTAGG	11 670
g14F	CTGTTGCAGGCTCAATAGTTCTTGC	11 893
g14R	TTCGAGGGAGCCTTGGGGTCTAACC	12 358
w14F	CCTCTATATATTTCTTATAACCCAGCGAGG	12 915
w14F2	GAACATCTTCTTATGGCCCTTCACC	13 025
w15R	CTAGCTGGCTTGATGTTGAGAAAGC	13 515
w15F	TCGCCATAGTCATTCTAGTGACAGC	13 535
w16R	GCTTGTTTGTTAGGGAGGCTAGTTC	13 842
w16F	CCACAGCTTGAATGACGAGCAAGAC	14 216
w17R	TAACGCGAGGATTAAGTCGAGGAAC	14 519
w17F	TTTACCACTCCACCACTTCTCCAAC	14 893
w18R	AGCAAAGGCCGAGTAGGGAACCAAAGTTTG	14 915
wCytbR	GTTGTCAACTGAGAAGCCTCCTCAG	15 593
w18F	CCGCTACAACAACCAACCCTAAAGC	15 631
cytb811R	CTGCCATTTTTTGGTTTACAACAC	16 060
g19F	GAGGAGGTTTCTCAGTAGATAATGC	16 174

Note: Primer codes beginning with "g" were designed for *Gadus* spp. (Coulson 2004). Forward primers are in Roman type and reverse primers are in italic type.

To sequence the PCR products, 5 μL of the PCR product was dried by evaporation or under vacuum. The sequencing cocktail comprised 1.0 or 2.0 μL of Big Dye (version 3.0 chemistry; Applied Biosystems Inc., Foster City, California), 1.6 μL of 2 μmol/L primer forward or reverse, and 2.2 μL ddH_2O. The sequencing cycle was an initial denaturation at 96 °C for 2 min, followed by 50 cycles of a 96 °C denaturation for 2 min, 50 °C annealing for 15 s, and 60 °C extension for 4 min. Excess reactants were removed by isopropanol precipitation. The DNA was resuspended in 5 μL of a 5:1 mixture of deionized formamide and

Table 2. Phenetic comparison of genome differentiation in *Anarhichas* species and *Gadus* codfish mtDNA genomes.

(A) Sequence differences among the genus *Lycodes* and *Anarhichas* species.				
	Lycodes	*A. denticulatus*	*A. minor*	*A. lupus*
Lycodes	0	1708	1706	1732
A. denticulatus	99.66	0	298	340
A. minor	98.79	17.44	0	283
A. lupus	100.85	20.22	16.64	0
(B) Number of SNP differences / kbp (*p* distances) among *Gadus* species.				
	G. (= *T.*) *chalcogrammus* EPac	*G. morhua* EAtl	*G. macrocephalus* EPac	*G. m. ogac* WAtl
G. (= *T.*) *chalcogrammus* WPac	*4.77*	38.46	40.18	40.25
G. morhua WAtl	38.26	*3.32*	39.58	39.25
G. macrocephalus WPac	39.85	38.72	*4.91*	*3.25*
G. m. ogac WAtl	39.72	38.46	*4.51*	[1.33]

Note: In A, the sequence differences were calculated as the absolute number of differences in the aligned complete sequences (upper half of the matrix), or the number of SNPs/kbp for the 15 082 bp aligned between zoarcoid and gadid sequences, less the CR locus, and missing or ambiguous sequences (lower half of the matrix). In B, the number of SNP differences/kbp (*p* distances) was calculated over the same 15 082 bp (after Coulson et al. 2006). In B, interspecific and intraspecific differences are compared across oceans (italic type); the *Gadus macrocephalus ogac* Richardson, 1836 comparison (brackets) is for two individuals from the western Atlantic Ocean (cf. Carr et al. 1999; Coulson et al. 2006). EPac, eastern Pacific Ocean; EAtl, eastern Atlantic Ocean; WAtl, western Atlantic Ocean.

25 mmol/L Na_2EDTA buffer, heated to 95 °C for 2 min, and cooled rapidly to 5 °C. The denatured samples were spotted onto a paper comb and electrophoresed in an ABI377 automated DNA sequencer. The ABI Data Collection (version 2.6) software was used to collect the trace data and the Sequence Analysis (version 3.2) software was used to extract the sequence chromatograms.

Sequencer version 4.1.2 (GeneCodes, Ann Arbor, Michigan) was used to assemble and edit the mtDNA fragments of the three *Anarhichas* species. These fragments were initially aligned against the *G. morhua* reference sequence (NCBI accession No. NC002081), in which the positions of the initial primer sets were known.

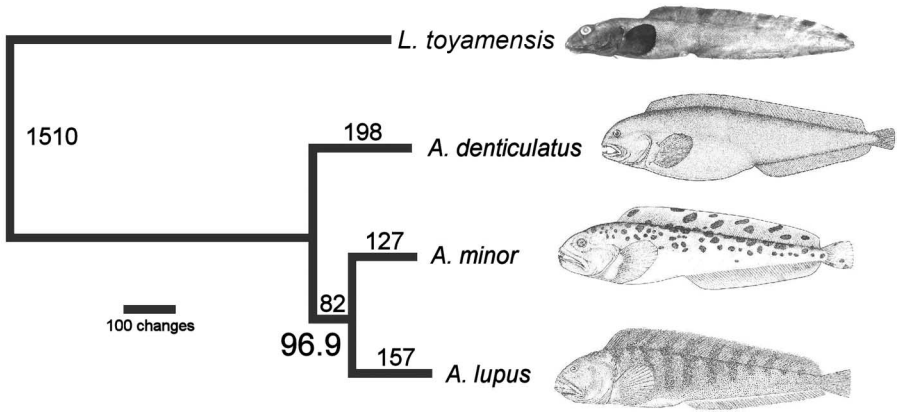
Phylogenetic and phenetic analyses were carried out using the phylogenetic analysis using parsimony (PAUP*) program version 4.0b10 (Swofford 2002). The exhaustive search algorithm was used to evaluate the properties of each of the three bifurcating trees possible with the complete genomes of the three *Anarhichas* species and another zoarcoid perciform species, the eelpout *Lycodes toyamensis* (Katayama, 1944) (NCBI accession No. NC004409), as the outgroup. A bootstrap analysis with 10 000 replicates was done to estimate the statistical significance for the nodes of the tree. Neighbor-joining phenetic analysis of uncorrected average pairwise distances (*p* distances) was also done. To compare the extent of intra- and inter-specific mtDNA genome differentiation with other fish species, the four zoarcoid genome sequences were aligned with those of seven gadid codfish (*Gadus* sensu lato; Carr et al. 1999) described in Coulson et al. (2006) (NCBI accession Nos. DQ356937–DQ356941 and DQ356946), with the addition of a second sequence from *Gadus* (= *Theragra*) *chalcogrammus* Pallas, 1814 (NCBI accession No. NC004449) from the western Pacific Ocean. CR locus sequences, which were not available for all gadines, were removed, and PAUP* was used to removed any missing and (or) ambiguous characters. For the consensus data set of 15 082 bp, *p* distances were calculated and are shown as numbers of changes per kilo base pairs in Table 2.

Variability of individual *Anarhichas* genes within genomes was calculated three ways: first as the average of the three pairwise sequence divergences among species (*p* distance) and second as the density of interspecific single nucleotide polymorphisms (SNPs), calculated as the ratio of the number of SNPs observed among all three species to the total gene length, expressed as the number of SNPs / 100 bp. The two indices are closely related: the former will underestimate the actual extent of nucleotide divergence by not correcting for multiple-hit nucleotide substitutions at the same site, whereas the latter will underestimate variability by not counting the homoplasic changes implied by the inferred phylogeny. The latter provides a more intuitive measure of observed variability. The third measure of variability is the difference between the rates of synonymous (d_S) and nonsynonymous (d_N) nucleotide substitutions in each gene, according to the algorithm of Nei and Gojobori (1986) as implemented in MEGA version 3.1 (Kumar et al. 2004). This measure detects unusual patterns of amino acid substitution, for example, as a result of selection (cf. Coulson et al. 2006). The expected Poisson distribution of SNPs over tRNA loci was calculated by the recursive algorithm described by Zar (1999: 181). Nonparametric rank-order correlations of SNP density among protein-coding loci between the three *Anarhichas* species and the three *Gadus* codfish species (*G. morhua*, *G.* (= *T.*) *chalcogrammus*, and *Gadus macrocephalus* Tilesius, 1810) (Coulson et al. 2006) (DQ356938, DQ356939, and DQ356946), and comparisons of variability indices among *Anarhichas* loci made independently of absolute magnitude, were calculated by Kendall's coefficient of rank correlation (τ) as implemented in BIOMstat version 3.30m (Applied Biostatistics, Inc. 2002; Sokal and Rohlf 1995: 593).

Results

The long-range PCR procedure amplified more than 97% of the complete mtDNA genome of *A. lupus* in three over-

Fig. 3. Phylogenetic relationships among the mitochondrial genomes of *Anarhichas* species and *Lycodes toyamensis*. The minimum length tree requires 2074 changes, with branch lengths as shown. The tree indicates that *A. lupus* and *A. minor* are each other's closest relatives; this arrangement is supported by 96.9% of bootstrap replications.



lapping fragments of 6835, 4555, and 4276 bp (Fig. 2). The sequence of the remaining 466 bp segment of the genome (including the 3′ end of the CYTB locus and the 5′ end of the CR locus) had been determined previously. The genomes of the other two wolffish species were amplified with various combination of the *Anarhichas* and *Gadus* specific primers in Table 1. The three genomes were sequenced with the same set of primers.

The mtDNA genome sequences comprised 16 519 bp in *A. lupus* and 16 520 bp in the other two species. The sequences have been submitted to GenBank and assigned the accession numbers EF427916, EF427917, and EF427918 for *A. lupus*, *A. minor*, and *A. denticulatus*, respectively. The gene order is identical to that of *L. toyamensis*, except that *L. toyamensis* has an extra 169 bp segment 3′ to the tRNA–Val locus that is not present in *Anarhichas* species. This segment has been deleted from the alignment analyzed here.

A total of 449 SNP sites were observed among the individuals representing the three species. *Anarhichas lupus* and *A. minor* differed by 248 pairwise differences (226 transitions and 22 transversions), *A. minor* and *A. denticulatus* by 274 differences (248 transitions and 26 transversions), and *A. lupus* and *A. denticulatus* by 286 differences (254 transitions and 32 transversions) (Table 2A). Interspecific *p* distances among *Anarhichas* species range from 16.64 to 20.22 substitutions / kbp (subs/kbp), which is approximately one-half that among *Gadus* species (from 38.46 to 40.25 subs/kbp) (Table 2B) and about four to five times as large as the differences among transoceanic individuals within *Gadus* species (from 3.25 to 4.77 subs/kbp).

With *L. toyamensis* included as an outgroup, there are 1944 variable sites, of which 146 are phylogenetically informative (sensu Nei 1987). The three possible bifurcating trees for four terminal taxa had lengths of 2074, 2094, and 2100 changes. The shortest tree indicates that *A. lupus* and *A. minor* are more closely related to each other than they are to *A. denticulatus*; this tree is supported by 96.9% of 10 000 bootstrap replicates (Fig. 3). Random resampling of 1, 2, 4, or 8 kbp of the complete genome supported the same tree in 57.2%, 66.2%, 77.4%, and 89.4% of bootstraps, respectively. This tree is also supported by neighbour-joining distance analysis (*p* distances) in 96.3% of bootstraps, as well as by weighted parsimony and various distance models with >95% bootstrap support in all cases (results not shown).

The most variable protein-coding loci as measured by SNP density were ND4, CYTB, and ND2, with 4.40, 4.22, and 4.19 SNPs / 100 bp, respectively. The least variable loci were ATP8 (also the shortest region), COX2, and ND3, with 1.19, 1.57, and 1.99 SNPs / 100 bp, respectively (Table 3). The CR locus was less variable than 9 of 13 protein-coding regions (2.45 SNPs / 100 bp) (Table 4). Most tRNA loci were either invariant or had only 1 SNP site, and the 20 tRNA loci collectively showed only 18 SNP sites over 1410 bp (1.28 SNPs / 100 bp), which was lower than all except 1 protein-coding locus. Although 3 tRNA loci had three or more SNP sites and were thus more variable than any protein-coding locus (tRNA–Val, tRNA–Arg, and tRNA–Trp: >4 SNPs / 100 bp), the observed distribution did not depart significantly from the expected random Poisson distribution of 18 SNPs over 20 tRNA loci ($\chi^2_{0.05,[3]} = 4.21$, $0.10 < p < 0.25$). The 16S and 12S rRNA loci were the least variable gene regions (1.18 and 0.35 SNPs / 100 bp, respectively). The relative ranks of locus variability as measured by SNP density and mean *p* distance are highly correlated (Kendall's $\tau = 0.9731$, $p << 0.01$), as are the rankings of SNP density and the rate of synonymous substitution (d_S) (Kendall's $\tau = 0.7871$, $p << 0.01$) and *p* distance vs. d_S (Kendall's $\tau = 0.8387$, $p << 0.01$). The correlations of SNP density and *p* distance with $d_N - d_S$ were slightly lower (Kendall's $\tau = 0.7179$ and 0.7949, respectively), but still $p << 0.01$ in both cases.

Similarly, there is a significant rank-order correlation of SNP densities between *Anarhichas* species and *Gadus* species (Kendall's $\tau = 0.4615$, $p = 0.0140$), but the association is far from exact. ND4 ranked first in *Anarhichas* and second in *Gadus*; however, the second- and third-ranked loci in *Anarhichas* (CYTB and ND2) ranked seventh and sixth in *Gadus*. The most highly ranked locus in *Gadus* (ND6) ranked fourth in *Anarhichas*, and the ATP6 and ND3 loci also had substantially lower ranks in *Anarhichas* (ninth and eleventh vs. fourth and fifth, respectively) (Fig. 4).

Discussion

The determination of the complete mtDNA genome se-

Table 3. Length (bp), SNP occurrence, SNP density (no. of SNPs / 100 bp), mean pairwise difference (p), rates of synonymous (d_S) and nonsynonymous (d_N) nucleotide substitutions, and the difference of the latter two measures for protein-coding loci of the mitochondrial genomes among *Anarhichas* species.

Gene	Length	SNP occurrence	SNP density	p	d_S	d_N	$d_N - d_S$
ND1	975	36	3.69	2.57	0.0858	0.0000	–0.0858
ND2	1048	44	4.19	2.54	0.0857	0.0054	–0.0803
COX1	1554	50	3.15	2.11	0.0728	0.0018	–0.0710
COX2	699	11	1.57	1.05	0.0424	0.0000	–0.0424
ATP8	168	2	1.19	0.80	0.0374	0.0000	–0.0374
ATP6	684	17	2.49	1.66	0.0568	0.0014	–0.0554
COX3	786	19	2.42	1.61	0.0538	0.0036	–0.0503
ND3	352	7	1.99	1.23	0.0416	0.0030	–0.0385
ND4L	297	8	2.69	1.80	0.0671	0.0000	–0.0671
ND4	1386	61	4.40	3.01	0.0837	0.0079	–0.0758
ND5	1839	59	3.21	1.90	0.0609	0.0060	–0.0550
ND6	523	21	4.02	2.55	0.0857	0.0035	–0.0822
CYTB	1162	49	4.22	2.64	0.0905	0.0047	–0.0851

Note: The regions are listed in their linear order in the genome.

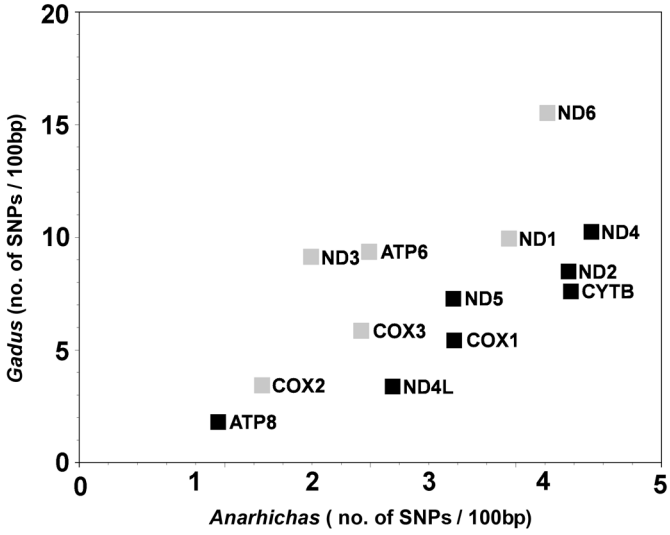
Table 4. Length (bp), SNP occurrence, and SNP density (no. of SNPs / 100 bp) for the Control Region (CR),rRNA-coding, and tRNA-coding loci of the mitochondrial genomes of *Anarhichas* species.

Gene	Length	SNP occurrence	SNP density
CR	981	24	2.45
12S rRNA	948	3	0.32
16S rRNA	1697	20	1.18
tRNA–Phe	68	0	0
tRNA–Val	72	4	5.56
tRNA–Leu	74	0	0
tRNA–Ile	70	1	1.43
tRNA–Gln	68	0	0
tRNA–Met	69	0	0
tRNA–Trp	70	3	4.29
tRNA–Ala	72	0	0
tRNA–Asn	74	0	0
tRNA–Cys	66	1	1.52
tRNA–Tyr	67	1	1.49
tRNA–Ser	71	0	0
tRNA–Asp	69	1	1.45
tRNA–Lys	74	1	1.35
tRNA–Gly	69	1	1.45
tRNA–Arg	69	3	4.35
tRNA–His	64	0	0
tRNA–Ser	72	1	1.39
tRNA–Leu	73	0	0
tRNA–Glu	79	1	1.27
All tRNAs	1410	18	1.28

quences of *Anarhichas* species provides benchmark data for the investigation of their genetic biodiversity. The analysis presented here can better inform a recovery plan for these species by identifying the gene regions of highest variability, which would best be able to identify any distinctive fishery stocks and (or) designatable units under SARA.

The molecular analysis shows that *A. lupus* and *A. minor* are more closely related to each other than either is to *A. denticulatus*. The genomic difference between *A. lupus* and *A. minor* (~20 subs/kbp) is about half that found between pairs of *Gadus* species (~40 subs/kbp), including another COSEWIC-assessed species, *G. morhua*. Because of the close relationships of the three species, the entire genome is required to resolve their branching order at >95% confidence. In contrast, Coulson et al. (2006) found that a random sample of about one-half of the genome sequence was sufficient to resolve almost all interspecific branches among gadines at >95% confidence. Given the estimated time of divergence of Atlantic *G. morhua* and Pacific *G. macrocephalus* at 3.5 million years before present, at the time of the last opening of the Bering strait (Carr et al. 1999; Coulson et al. 2006), extrapolation suggests that the

Fig. 4. Total number of SNPs observed among three species of *Anarhichas* and three species of *Gadus* (*G. morhua*, *G. macrocephalus*, and *G.* (= *T.*) *chalcogramma*) (Coulson et al. 2006), expressed as number of SNPs / 100 bp of gene length. Solid boxes indicate loci at which the relative rank of SNP density is higher in *Anarhichas* than in *Gadus* and shaded boxes indicate the reverse; ATP8 ranked last in both genera.



three *Anarhichas* species separated from one another ca. 2 million years ago.

A closer relationship of *A. lupus* and *A. minor* is supported by an examination of several morphological and ecological characters. For example, both are typical benthophages with teeth adapted for feeding on bottom-dwelling invertebrates (echinoderms, molluscs, and crustaceans), in contrast to *A. denticulatus*, which has teeth adapted for feeding on mobile invertebrates (ctenophores and jellyfish) (Albikovskaya 1982). All three species are pan-Atlantic (Fig. 1) but appear to have rather different niche requirements in this region. *Anarhichas denticulatus* have the widest distribution, found in the eastern Atlantic Ocean from Iceland to Novaya Zemlya and in the western Atlantic Ocean from Mould Bay to Sable Island Bank. They also have the greatest range of depths, from 151 to 600 m, and are known to feed close to the surface in open water — behaviours that are uncommon in the other two species. *Anarhichas lupus* are found in the eastern Atlantic Ocean from Iceland to the west coast of France, and from Greenland to the Gulf of Maine, and are thus the most southerly distributed species. They are the most abundant species in the area of Newfoundland and Labrador, where they are found from 101 to 350 m (Albikovskaya 1982). *Anarhichas minor* have the most restricted distribution, from Iceland to Bergen and from Greenland to the Grand Banks, where they are usually restricted to depths >450 m (Scott and Scott 1988). Given the phylogenetic relationships shown here, we can hypothesize that *A. lupus* and *A. minor* represent alternative, derived specializations at more restrictive, benthic strata, in contrast to a more widespread, generalist ancestral type represented by *A. denticulatus*.

The 449 SNPs observed in the mtDNA of the three individual representatives of the *Anarhichas* species occur at an average density of 26.8 SNPs/kbp, which varies from 3.2 to 44.0 SNPs/kbp among the major gene regions. SNP density among protein-coding loci in these individuals is exactly one-half of that among three individuals representative of three *Gadus* species (3.0 vs. 6.0 SNPs/100 bp), which is in agreement with the twofold difference in intrageneric genetic distances (18.8 vs. 39.58 subs/kbp) (Table 2). Such a correspondence is expected on the assumption of selective neutrality (Kimura 1983), as is the observation that all measures of locus variability are highly correlated. Rates of nonsynonymous substitution are very low at every locus, and the difference ($d_N - d_S$) is always negative, so that Fisher's exact test of the distribution of synonymous and nonsynonymous substitutions (Zhang et al. 1997) would not approach significance for any comparison. There are thus no indications of unusual patterns of selection at work on *Anarhichas* genomes. The significant rank-order correlation of SNP density among protein-coding loci between *Anarhichas* and *Gadus* indicates that patterns are broadly similar; however, there are many exceptions. The three most variable loci in *Anarhichas* ranked second, seventh, and sixth, respectively, in *Gadus*; this includes the CYTB locus (second vs. seventh) for which a large number of SNP variants have been identified in *G. morhua* (Carr and Marshall 1991; Pepin and Carr 1993; Arnason 2004). Whether such intergeneric differences are a function simply of variation in the neutral mutation rate among loci, or instead indicate differing patterns of natural selection, can be evaluated by further comparison of intraspecific patterns of nucleotide substitution in *Anarhichas* and *Gadus* with the interspecific patterns shown here (Kreitman 2000).

Although many studies of fish and other populations have used the CR locus on the expectation that it is routinely hypervariable (Brown et al. 1979; Kocher and Carleton 1997; Ingman et al. 2000), in wolffish this region is one of the least variable (less than 9 of 13 protein-coding loci) and is not a good candidate for intraspecific comparisons. The available population data confirm this (K.A. Johnstone, H.D. Marshall, and S.M. Carr, unpublished observation). As shown by this study and others (Faber and Stepien 1997), hypervariability of the CR locus is often not the rule in nonmammalian taxa. Other studies have advocated the use of loci in the NADH gene complex for single- and multi-locus population analyses (Russo et al. 1996). Recently, use of the 5′ end of the COX1 locus as a "barcode" across species and taxa, including fish, has been advocated, with the expectation that the locus will be routinely variable among species and less so within (Ward et al. 2004). Again, the present data do not show a uniform pattern among taxa: COX1 is more variable interspecifically in *Anarhichas* species than in *Gadus* species (ranked sixth vs. tenth). The present data should emphasize that patterns of locus variability are not readily generalized across even relatively closely related taxa (zoarcoid perciform fishes) and that preliminary investigation of taxon-specific patterns is routinely warranted before carrying out a full-scale population genetic study.

The results of this study suggest several strategies for population genetic investigation of *Anarhichas* species. A single-locus analysis would best target the ND4 locus with primers g13F and *g14R*, which would amplify this gene within a fragment of 1476 bp. Where amplification of fragments of this size from samples with predominantly low molecular weight DNA is difficult, amplification of the ND4 region in two PCR reactions, with primer pairs of g13F–*w14R* (896 bp) and g14F–*g14R* (781 bp), should be feasible. A second strategy would be a multi-locus analysis of the contiguous ND4, ND5, CYTB, and COX1 genes, which includes two of the most variable loci. Use of long-range PCR amplification with the primers w13F and *cytb811R* would amplify a 6329 bp fragment that includes these four loci, and provide more than one-third of the genome for analysis. Alternatively, whole-genome approaches are also available that would examine variation in the complete 16.5 kbp genomes (Carr and Marshall 2005; Carr et al. 2007).

Note added in proof

The original description of the walleye pollock is *Gadus chalcogrammus* Pallas, 1814 in Pallas' *Zoographia rosso-asiatica*. Volume 1 of this work was published in 1811, which is often cited as the year of publication of this combination (e.g., Cohen et al. 1990), but the description of the pollock occurs in volume 3, which was published in 1814, according to *International Commission of Zoological Nomenclature* opinion 212. The correction should also be applied to the discussion in Carr et al. (1999) and Coulson et al. (2006).

Acknowledgments

We thank the staff of the Northwest Atlantic Fisheries Science Centre, Fisheries and Oceans Canada, St. John's, especially Mr. Dave Kulka, Dr. Mark Simpson, and Mr. Chris Holloway for their support on the project and for obtaining the bycatch samples; Drs. R. Hooper and M. Burton for aid in obtaining DNA from a fresh striped wolffish; and two anonymous reviewers for helpful comments on an earlier draft of the manuscript. We thank Mark Coulson for sharing his data in advance of publication, and Sarah Flynn, Angela Pope, Stephanie Royston, and Corinne Wilkerson for technical and other assistance. The data were collected as part of the B.Sc. (Hons.) thesis research of the senior author, who thanks her friends and her parents, Terry and Kathy Johnstone. Funding was provided by the Grants and Contributions program of Fisheries and Oceans Canada.

References

Albikovskaya, L.K. 1982. Distribution and Abundance of Atlantic wolffish, Spotted wolffish and northern wolffish in the Newfoundland Area. Northwest Atl. Fish. Organ. Sci. Counc. Stud. **3**: 29–32.

Applied Biostatistics, Inc. 2002. BIOMstat. Version 3.30m [computer program]. Applied Biostatistics, Inc., Setauket, N.Y.

Arnason, E. 2004. Mitochondrial cytochrome *b* DNA variation in the high-fecundity Atlantic cod: trans-Atlantic clines and shallow gene genealogy. Genetics, **166**: 1871–1885. doi:10.1534/genetics.166.4.1871. PMID:15126405.

Brown, W.M., George, M., Jr., and Wilson, A.C. 1979. Rapid evolution of animal mitochondrial DNA. Proc. Natl. Acad. Sci. U.S.A. **76**: 1967–1971. doi:10.1073/pnas.76.4.1967. PMID:109836.

Carr, S.M., and Marshall, H.D. 1991. 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, S.M., and Marshall, H.D. 2005. The ArkChip — a multispecies mitochondrial genome DNA "re-sequencing" strategy for biodiversity and conservation genomics. Comp. Biochem. Physiol. A, **141**: S88–S89.

Carr, S.M., Kivlichan, D.S., Pepin, P., and Crutcher, D.C. 1999. Molecular systematics of gadid fishes: implications for the biogeographic origins of Pacific species. Can. J. Zool. **77**: 19–26. doi:10.1139/cjz-77-1-19.

Carr, S.M., Marshall, H.D., Duggan, A.T., Flynn, S.M.C., Johnstone, K.A., Pope, A.M., and Wilkerson, C.D. 2007. Phylogeographic genomics of mitochondrial DNA: patterns of intraspecific evolution and a multi-species, microarray-based DNA sequencing strategy for biodiversity studies. Comp. Biochem. Physiol. D Genomics and Proteomics. In press.

Cohen, D.M., Inada, T., Iwamoto, T., and Scialabba, N. 1990. FAO species catalogue. Vol. 10. Gadiform fishes of the World. Food and Agriculture Organization of the United Nations, Rome.

Coulson, M.W. 2004. Mitochondrial genomic phylogeny of gadid fish: implications for biogeographic origins and taxonomy. M.Sc. thesis, Memorial University of Newfoundland, St. John's.

Coulson, M.W., Marshall, H.D., Pepin, P., and Carr, S.M. 2006. Mitochondrial genomics of gadid fish: implications for biogeographic origins and taxonomy. Genome, **49**: 1115–1130. doi:10.1139/G06-083. PMID:17110992.

Department of Fisheries and Oceans. 2004. Allowable harm assessment for spotted and northern wolffish. Department of Fisheries and Oceans, Can. Sci. Advis. Sec. Stock Status Rep. No. 2004/031.

Faber, J.E., and Stepien, C. 1997. The utility of mitochondrial DNA control region sequence for analyzing phylogenetic relationships among populations, species, and genera of the Percidae. *In* Molecular systematics of fishes. *Edited by* T.D. Kocher and C.A. Stepien. Academic Press, New York. pp. 129–143.

Ingman, M., Kaessmann, H., Paabo, S., and Gyllensten, U. 2000. Mitochondrial genome variation and the origin of modern humans. Nature (London), **408**: 708–713. PMID:11130070.

Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge, UK.

Kocher, T.D., and Carleton, K.L. 1997. Base substitutions in fish mitochondrial DNA: patterns and rates. *In* Molecular systematics of fishes. *Edited by* T.D. Kocher and C.A. Stepien. Academic Press, New York. pp. 13–24.

Kocher, T.D., Thomas, W.K., Meyer, A., Edwards, S.V., Pääbo, S., Villablanca, F.X., and Wilson, A.C. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. U.S.A. **86**: 6196–6200. doi:10.1073/pnas.86.16.6196. PMID:2762322.

Kreitman, M. 2000. Methods to detect selection in populations with applications to the human. Annu. Rev. Genomics Hum. Genet. **1**: 539–559. doi:10.1146/annurev.genom.1.1.539. PMID:11701640.

Kumar, S., Tamura, K., and Nei, M. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. **5**: 150–163.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Nei, M., and Gojobori, T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3**: 418–426. PMID:3444411.

O'Dea, N.B., and Haedrich, R.L. 2002. A review of the status of the Atlantic wolffish, *Anarhichas lupus*, in Canada. Can. Field-Nat. **116**: 423–432.

Pepin, P., and Carr, S.M. 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.

Russo, C.A.M., Takezaki, N., and Nei, M. 1996. Efficiencies of different genes and different tree-building methods in recovering a known vertebrate phylogeny. Mol. Biol. Evol. **13**: 522–536.

Scott, W.B., and Scott, M.G. 1988. Atlantic fishes of Canada. University of Toronto Press, Toronto.

Sokal, R.R., and Rohlf, J.M. 1995. Biometry. 3rd ed. W.H. Freeman and Co., New York.

Swofford, D.L. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4.0b10. Sinauer Associates Inc., Sunderland, Mass.

Templeman, W. 1986*a*. Contribution to the biology of the Spotted wolffish (*Anarhichas minor*) in the Northwest Atlantic. J. Northw. Atl. Fish. Sci. **7**: 47–55.

Templeman, W. 1986*b*. Some biological aspects of the Atlantic wolffish (*Anarhichas lupus*) in the Northwest Atlantic. J. Northw. Atl. Fish. Sci. **7**: 57–65.

Ward, R.D., Zemlak, T.S., Innes, B.H., Last, P.R., and Hebert, P.D.N. 2004. DNA barcoding Australia's fish species. Philos. Trans. R. Soc. Lond. B Biol. Sci. **360**: 1847–1857.

Zar, J.H. 1999. Biostatistical Analysis. 4th ed. Prentice-Hall Inc., Upper Saddle River, N.J.

Zhang, J., Kumar, S., and Nei, M. 1997. Small-sample tests of episodic adaptive evolution: a case study of primate lysozymes. Mol. Biol. Evol. **14**: 1335–1338. PMID:9402743.