Mitochondrial DNA, October–December 2009; 20(5–6): 110–118

informa
healthcare

SHORT COMMUNICATION

Testing taxonomic boundaries and the limit of DNA barcoding in the Siberian sturgeon, *Acipenser baerii*

VADIM J. BIRSTEIN[1], ROB DESALLE[2], PHAEDRA DOUKAKIS[3], ROBERT HANNER[4,5], GEORGII I. RUBAN[6], & EUGENE WONG[5]

[1]*Sturgeon Conservation International, New York, USA,* [2]*Division of Invertebrate Zoology, American Museum of Natural History, New York, USA,* [3]*School of Marine and Atmospheric Sciences, Institute for Ocean Conservation Science, Stony Brook University, Stony Brook, NY, USA,* [4]*Canadian Barcode of Life Network, Biodiversity Institute of Ontario, Guelph, Ont., Canada,* [5]*Department of Integrative Biology, University of Guelph, Guelph, Ont., Canada, and* [6]*A. N. Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, Moscow, Russia*

(*Received 3 July 2008; revised 24 March 2009; accepted 7 July 2009*)

Abstract
DNA barcoding efforts involving animals have focused on the mitochondrial cytochrome *c* oxidase subunit I (*Cox1*) gene. Some authors suggest that this marker might under-diagnose young species. Herein, we examine *Cox1* and control region diversity in a sample of Siberian sturgeon (*Acipenser baerii*), a species with an extremely wide geographic distribution in the major rivers of Siberia and in Lake Baikal. Some authors currently recognize three subspecies within this species. These subspecies are reasonable candidates for species units detectable through DNA barcoding. The *Cox1* gene illustrated no variation within the species, while the control region displayed statistically significant differences among the subspecies using analysis of molecular variance (AMOVA). Given the uniformity of *Cox1* sequences recovered, *Cox1* is probably a good region for barcoding *A. baerii* at the species level. Although control region variation among subspecies was significant, diagnostic differences were not found for any of the subspecies.

Keywords: *DNA barcode,* Cox1*, control region, forensics, sturgeon, subspecies*

Introduction

The use of many different markers in species-level studies has resulted in the lack of a broadly comparable database needed to facilitate molecular diagnostics for biodiversity. DNA barcoding was proposed to meet this need and relies on patterns of sequence variation derived from a short standardized gene fragment for rapid, accurate, and cost-effective identification of species (Hebert et al. 2003). The reaction of the larger scientific community to DNA barcoding has been positive in general (e.g. Besansky et al. 2003; Schander and Willassen 2005; Dasmahapatra and Mallet 2006; Costa and Carvalho 2007; Neigel et al. 2007; Waugh 2007), but see below for papers critical of the approach. In particular, ichthyologists have increasingly begun to employ the technique for species identification and discovery (Ward et al. 2005, 2007, 2008a,b; Spies et al. 2006; Victor 2007, 2008; Hubert et al. 2008; Rock et al. 2008). These successes not withstanding, some researchers have raised questions about the utility of DNA sequence information in modern taxonomic and systematic biology (Lipscomb et al. 2003; Seberg et al. 2003; Moritz and Cicero 2004; Meyer and Paulay 2005; Prendini 2005; DeSalle 2006; Rubinoff 2007).

One of the more important questions for DNA barcoding concerns the utility of the marker designated as the universal barcode sequence in

Correspondence: P. Doukakis, School of Marine and Atmospheric Sciences, Institute for Ocean Conservation Science, Stony Brook University, Stony Brook, NY 11794, USA. Tel: + 1 631 632 9325. Fax: + 1 631 632 2648. E-mail: pdoukakis@gmail.com

ISSN 1940-1736 print/ISSN 1940-1744 online
DOI: 10.3109/19401730903168182

animals – the mitochondrial 5′ region of the cytochrome *c* oxidase subunit I gene (*Cox1*). Numerous studies have suggested that *Cox1* barcoding can delineate most closely related sister species (e.g. Hebert et al. 2003, 2004; DeSalle et al. 2005; Smith et al. 2005, Lefébure et al. 2006; Gómez et al. 2007). Yet few, if any, studies have explicitly compared the resolution of barcoding to more rapidly evolving markers in fishes to understand the taxonomic limits of the barcoding gene region.

In the present study, we conduct a test for the appropriateness of *Cox1* as a barcoding tool in the fish genus *Acipenser*. To conduct the test, we use the Siberian sturgeon (*Acipenser baerii*), a species with a detailed taxonomic record (see Materials and methods) and several historical subspecies designated within it (see below). By comparing markers for their ability to diagnose the previously established species designated from taxonomic work, and the ability to reject other below species-level hypotheses (subspecies), we can comment upon the adequacy of such markers to detect taxonomic boundaries. Such tests will help us understand the lower taxonomic limit of the utility of *Cox1* as a barcoding tool. In this study, we do not use DNA barcoding as a species "discovery" tool. Rather, we use it as a means to "identify" species (for the distinction between these two very different processes, see DeSalle 2006, 2007; Rubinoff 2007). The simple design of our study therefore is to ask several questions concerning diagnosis of species using molecular markers. First, does *Cox1* diagnose *A. baerii* clearly as distinct from other *Acipenser* species? Second, are diagnostics present in *Cox1* for previously designated subspecies? Third, can other more rapidly evolving markers such as the mitochondrial control region sequences be used to establish diagnostics for *A. baerii* or for the subspecies designated by previous taxonomic work?

In another context, Davis and Nixon (1992) highlighted two caveats concerning species diagnosis in a character-based context – over- and under-diagnosis. The first is the problem of over-diagnosis resulting from examining too few individuals in tests for species boundaries. In this case, two entities can be imprecisely diagnosed as distinct because of the lack of detection of existing variation due to limited sampling. The second caveat concerns the under-diagnosis of two entities and is the subject of this paper. This problem occurs when a marker reveals inadequate variation to precisely test hypotheses concerning species boundaries. Hickerson et al. (2006) suggested that this might be a problem for DNA barcoding. If DNA barcoding using *Cox1* is capable of rejecting a hypothesis of a species boundary and a more rapidly evolving marker (such as mitochondrial control region) does not, then it is possible that barcoding would under-diagnose species boundaries. In this study, we focus on the utility of *Cox1* in comparison with the more rapidly evolving mitochondrial control region. In this way, we test whether control region sequences are more sensitive to species boundaries than *Cox1* barcodes and hence, empirically examine how prone DNA barcodes using *Cox1* might be to under-diagnosis.

Materials and methods

Current taxonomy of the study species

The Siberian sturgeon *A. baerii* Brandt, 1869, is a potamodromous species that occupies a wide geographic area, inhabiting all of the major river systems of Siberia, from the Ob River in the west to the Kolyma River in the east, and Lake Baikal (Ruban 1999, 2005; Figure 1). This species is closely related to the Russian sturgeon *Acipenser gueldenstaedtii*, which inhabits basins of the Caspian and Black seas, and probably originated recently after the last glaciation (Birstein and DeSalle 1998; Birstein et al. 2000, 2005; Birstein and Ruban 2004). Molecular diagnostics clearly delineate *A. baerii* from its close relatives (Birstein et al. 2005).

While studying mostly museum specimens, four subspecies have been historically described within *A. baerii* based on geographic location, and morphometric and meristic differences: (1) the Ob River sturgeon (*Acipenser baerii baerii*) in the Ob River basin, (2) the Yakutian sturgeon (*Acipenser baerii chatys*) in the Lena River basin, (3) the long-nose Siberian

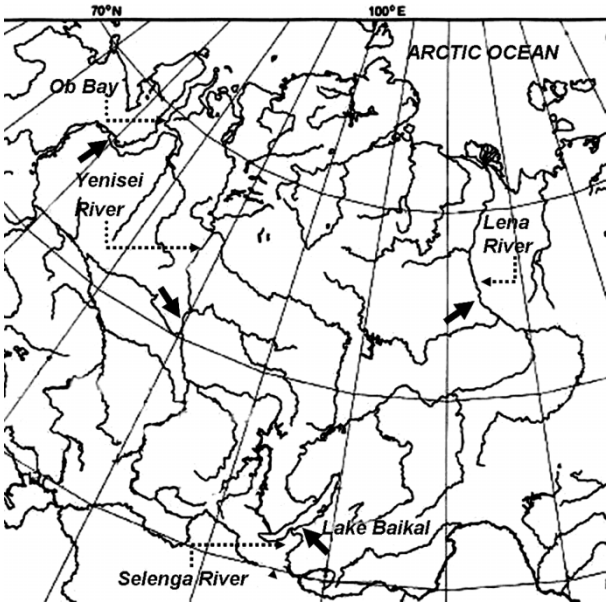

Figure 1. Map of the distribution of the Siberian sturgeon, *A. baerii* and sampling locations for wild samples identified by black arrows. *A. b. baicalensis* sampled from the Selenga River (approximately 20 km above the mouth), *A. b. baerii* from the Ob River, and *A. b. stenorrhynchus* from the Yenisei River. The Lena River is shown for reference only as samples for this population of *A. b. stenorrhynchus* were taken from a hatchery population.

sturgeon (*Acipenser baerii stenorrhynchus*, at first considered a separate species *Acipenser stenorrhynchus*) in the Yenisei River and East Siberian rivers, and (4) the Baikal sturgeon (*A. baerii baicalensis*) in Lake Baikal (reviewed in Sokolov and Vasil'ev 1989; Ruban 1997, 2005). Subsequent taxonomic revisions led to the recognition of only three subspecies, *A. b. baerii*, *A. b. stenorrhynchus* (it includes the former *A. baerii chatys*) and *A. b. baicalensis* (Ruban 2005). In the present study, we consider these three subspecies as potential species and test the hypothesis that they are such, by searching for *Cox1* sequence diagnostics.

However, the recent extensive detailed statistical analysis of morphological data obtained for a large number of specimens studied in the field and representing various river populations demonstrated that there are no differences among subspecies of *A. baerii* (Ruban 1997, 2005). Surprisingly, Eschmeyer's "Catalog of Fishes" still lists subspecies of *A. baerii* (Eschmeyer 2008). Therefore, a *Cox1*-based DNA barcode test could confirm whether *A. baerii* should be considered a species or whether it consists of considerably diversified forms described as subspecies.

The Siberian sturgeon is threatened throughout its range and understanding the taxonomic structure of the species is critical to conservation (Sokolov 2001). Rapid disappearance of the Lake Baikal population became clear about 20 years ago, with natural reproduction practically ceasing in 1995 (Afanasiev and Afanasieva 1996; Ruban 1999; Afanasiev 2006). Hatchery supplementation has been attempted as a conservation strategy with little success. We examine the Siberian sturgeon using samples from wild and hatchery populations and both *Cox1* and the control region to understand the limits of DNA barcoding in this species. Hatchery samples are used to represent subspecies for which wild collection is no longer possible.

Samples

All of the samples listed here were newly examined for *Cox1*. For the control region, the only new samples considered included fins fixed in alcohol from 11 immature *A. b. baicalensis* individuals caught in the Selenga River (approximately 20 km above the mouth) in the late 1990s (Baikal 1–10, 12 wild; Figure 1). Samples from other populations were examined in a previous study (Doukakis et al. 1999) as follows: six *A. b. baicalensis* individuals from the Konakovo hatchery (Baikal 1–3, 5, 7, 10 K; GenBank accession numbers AF168496–AF168500 and AF168502), 11 *A. b. stenorrhynchus* individuals from the Konakovo hatchery (Lena 1–10, 12; accession numbers AF168484–AF168493 and GQ262745), one *A. b. stenorrhynchus* egg sample from the Konakovo hatchery (Lena fc, accession number AF168494), 11 *A. b. baerii* individuals from the Ob River (Ob 1, 3, 4, 6, 7, 9–14; accession numbers AF168469–AF168479) and four *A. b. stenorrhynchus* individuals from the Yenisei River (Yen 2–5; accession numbers AF168480–AF168483, Figure 1). The *A. gueldenstaedtii* sequence with GenBank accession number AF238721 was used as the outgroup in the phylogenetic analysis.

DNA extraction, polymerase chain reaction (PCR) and sequencing

DNA was extracted using phenol chloroform techniques after overnight Proteinase K incubation for all samples (DeSalle et al. 1993). The entire control region was amplified using primers in Doukakis et al. (1999). Polymerase chain reaction (PCR) conditions in a Perkin-Elmer 480 thermocycler (PE Biosystems, Foster City, CA, USA) were: 94°C during 1 min, 46°C during 1 min, and 72°C during 110 s for 33 cycles in a 25 μl reaction. All PCR products were purified with BIO 101 Gene Clean system (BIO 101, Inc., La Jolla, CA, USA) before sequencing. Sequencing protocols followed manufacturer specifications (PE Biosystems) and used primers listed in Birstein et al. (2000, 2005). The 643 bp fragment of the control region examined corresponds to positions 287–922 of the *Acipenser transmontanus* GenBank sequence X54348, outside the heteroplasmic region (Buroker et al. 1990; GenBank accession numbers GQ262734–GQ262744).

The conditions for *Cox1* PCR amplification in a Mastercycler gradient thermal cycler (Eppendorf, New York, NY, USA) were: 94°C during 2 min, 35 cycles of 94°C during 30 s, 52°C during 40 s and 72°C during 1 min, and finally 72°C during 10 min. A 652 bp segment of the 5′-end of the mitochondrial *Cox1* gene was amplified using the Folmer primers (Folmer et al. 1994). Each PCR reaction mixture consisted of 6.25 μl of 10% trehalose, 3 μl ultrapure ddH_2O, 1.25 μl of 10 × PCR buffer for Platinum *Taq* (Invitrogen, Carlsbad, CA, USA), 0.625 μl of 50 mM $MgCl_2$, 0.125 μl of each primer (10 μM), 0.0625 μl of 10 mM dNTP mix, 0.06 μl Platinum *Taq* DNA polymerase, and 0.5–2.0 μl template DNA. PCR products were visualized on a 2% w/v agarose E-gel 96-well plate (Invitrogen). Bi-directional sequencing reactions were carried out on an ABI 3730 DNA Analyzer using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Inc., Foster City, CA, USA). Each cycle sequencing reaction mixture consisted of 5.0 μl of 10% trehalose, 0.917 μl ultrapure ddH_2O, 1.917 μl of 5 × buffer (400 mM Tris–HCl, pH 9.0 and 10 mM $MgCl_2$), 1 μl primer (10 μM), 0.167 μl BigDye, and 1.5 μl PCR product. Bi-directional contig assembly was carried out in SeqScape 2.1.1 (Applied Biosystems). Finished *Cox1* barcode sequence assemblies were recorded on BOLD (http://www.barcodinglife.org; Ratnasingham and Hebert 2007). *Cox1* was successfully sequenced for

all samples except Baikal 9 and 12, Lean fc and 12, and Ob 3 (GenBank accession numbers GQ328783–GQ328816).

Genetic diversity and phylogenetic analyses

Sequences were aligned using CLUSTAL (Larkin et al. 2007). Arlequin 3.01 was used to calculate genetic diversity measures and to examine subspecies differences using analysis of molecular variance (AMOVA) with 10,000 permutations and a significance level of 0.05 (Excoffier et al. 2005). Phylogenetic analysis was performed using three optimality criteria. Maximum parsimony (MP) was employed using heuristic searches with 100 random addition replicates and tree bisection–reconnection branch-swapping under equal character weighting and *A. gueldenstaedtii* as an outgroup in PAUP 4.0b10 (Phylogenetic analysis using parsimony, Swofford 2003). Node support was estimated using 1000 bootstrap pseudoreplicates with full heuristic searches and 10 random taxon additions. We estimated the best-fit substitution model for our data in the web-based implementation of ModelTest (Posada and Crandall 1998) using the Akaike information criterion in FindModel (http://hcv.lanl.gov/content/hcv-db/findmodel/findmodel.html). Maximum likelihood inference was carried out in PAUP with a heuristic search and 100 random sequence additions. The same model was used for Bayesian inference in MrBayes 3.1.2 (Ronquist and Huelsenbeck 2003). We ran two independent analyses for 3 million generations each sampling every 1000 steps from the posterior density and used a 10% burn-in. The post-burn-in trees were summarized in the form of a 50% majority-rule consensus tree.

Sequences were examined for the presence of diagnostic nucleotide sites unique to a population or a subspecies. Since sequence variation was minimal and easily interpreted by visual inspection, haplotypes were assigned by eye for both the *Cox1* gene and the control region. We used MacClade (Maddison and Maddison 2002; http://macclade.org) to visualize and organize the haplotypes. Haplotype diversity (H) and nucleotide diversity (π) were calculated for each population in Arlequin. Distance-based analyses for comparison with the diagnostic characters approach were accomplished with neighbor joining (NJ) using Kimura two-parameter distances (Kimura 1980).

Results

Diversity and distance analysis

The genetic distance between *A. gueldenstaedtii* and *A. baerii* was 6.3–7.9%, while the intraspecific distance did not exceed 3%. In previous DNA-based studies of *A. baerii* and its close relatives *A. gueldenstaedtii* and *Acipenser nacarii* (Birstein et al. 2005), we have shown that clear diagnostics using DNA markers exist for these species. Consequently, the first step in the present study was to determine the extent of variability for *Cox1* sequences within *A. baerii*. This analysis revealed that all *A. baerii* individuals, regardless of subspecies designation, were identical at the *Cox1* sequence level. For the control region, haplotype diversity (H%) and nucleotide diversity (π%) is highest for the Ob River (Ob River $H = 94.55 \pm 6.59$, $\pi = 1.0812 \pm 0.6217$; Yenisei River: $H = 83.33 \pm 22.24$, $\pi = 0.4710 \pm 0.3682$; Konakovo: $H = 60.0 \pm 21.52$, $\pi = 0.209 \pm 0.173$; Lena River: $H = 45.45 \pm 17.01$, $\pi = 0.23 \pm 0.1683$; and Selenga River: $H = 18.18 \pm 0.1436$, $\pi = 0.0285 \pm 0.0448$). Subspecies-level variation was also highest for *A. b. baerii* ($H = 94.55 \pm 6.59$, $\pi = 1.0812 \pm 0.6217$) followed by *A. b. stenorrhynchus* ($H = 69.17 \pm 12.38$, $\pi = 0.3534 \pm 0.2296$) and *A. b. baicalensis* ($H = 59.56 \pm 9.91$, $\pi = 0.7076 \pm 0.4099$). The AMOVA returned a significant differentiation among all subspecies (Table I). The NJ tree recovered for the control region sequences did not display clustering that corresponded directly to the geographic or predesignated subspecies designations.

Character-based tree analysis

According to the Akaike information criterion, the best-fit model for the control region alignment was HKY (Hasegawa et al. 1985) with rate heterogeneity modeled by the Γ distribution (Yang 1994) and accounting for a proportion of invariant sites. The MP analysis recovered one tree with 71 steps based on 16 parsimony-informative characters and a consistency

Table I. AMOVA comparisons for subspecies of *A. baerii*.

	A. b. baicalensis	*A. b. stenorrhynchus*	*A. b. baerii*
A. b. baicalensis	–	0.00069	0.01465
A. b. stenorrhynchus	0.29694	–	0.00228
A. b. baerii	0.16715	0.11939	–

Notes: F_{ST} below the diagonal, p-value above the diagonal. *A. b. baicalensis* refers to samples from Lake Baikal (Selenga River and Konakovo hatchery), *A. b. stenorrhynchus* to samples from the Lena and Yenisei Rivers, and *A. b. baerii* to samples from the Ob River.

index of 0.9437, a retention index of 0.9592, and a rescaled consistency index of 0.9051.

Figure 2 shows the phylogenetic tree recovered for the control region under Bayesian inference with bootstrap support for the clades illustrated using NJ, MP, maximum likelihood, and Bayesian inference. The Bayesian and MP tree supported the highest degree of structure. All analyses illustrate that wild individuals from the Selenga River form a monophyletic group, with three of the four Yenisei River individuals comprising a sister group to this cluster. As illustrated in Table II, very little sequence diversity was detected in the wild Selenga River samples. Konakovo Hatchery individuals formed a single clade with one individual from the Ob River. Phylogenetic structure corresponding to subspecies or populations was not observed.

Character diagnostics

Of the 38 variable sites recovered for the control region, 36 were found in more than one individual (Table II). Selenga River individuals are distinct from other populations, captive and wild, and can be diagnosed at one position in the control region (position 320, Table II). This population is distinct from the Konakovo hatchery population at five additional positions: 91, 501, 538, 601, and 605.

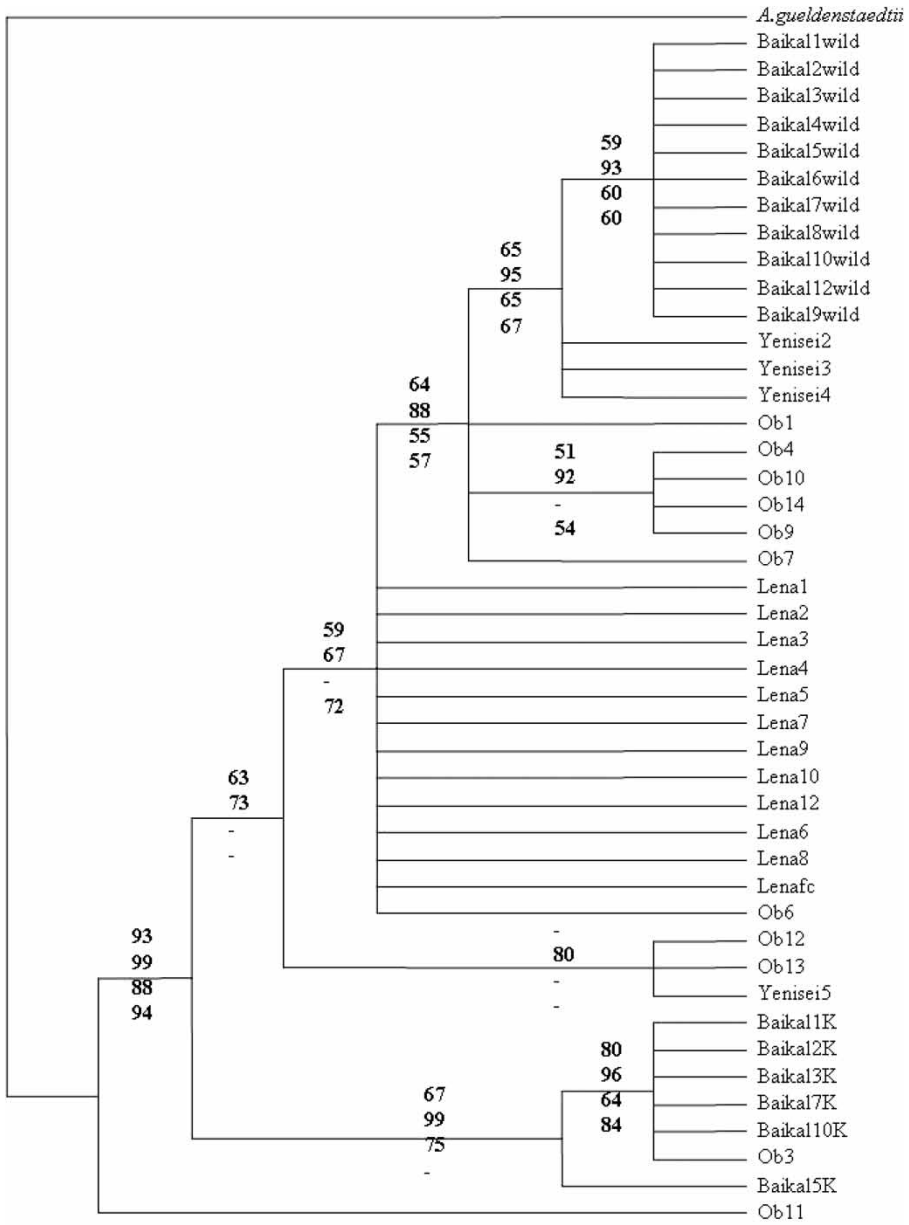

Figure 2. Bayesian phylogenetic tree for control region sequence data. Node support >50% is shown in the following order: MP (top), Bayesian, maximum likelihood, and NJ (bottom). –, <50% support.

Table II. Polymorphic nucleotide sites in the control region, haplotype designations (*H*) and absolute haplotype frequency (*N*) for Siberian sturgeon.

	1111111222334444455555555566666 1479902456780572600467011112334790022 5921273704379050379489136787486160151 9	*H*	*N*
Baikal1-10wild	GTTCAGTGGCT-AACTGCTTAAG-GAGAGA-TTTGGAT	A	10
Baikal-12wild	. . . . . G	B	1
Baikal1K-3K,7K,10K	A . .T . . C . . . .- . . TC . . . . . A- . . . . .G- . . . AT . .	C	5
Baikal5K	. . .T . . . . . - . .TC . . . . . A- . . . . . G- . . . AT . .	D	1
Lena1-5,7,9,10,12	. . . . . .- . . TC . . . . . A- . . . . . -	E	9
Lena6	. . . . . .- . . TC . . . . . AG . . . . . -	F	1
Lena8	. . . . . .- . . TC . . . . . AGTG- . . . -	G	1
Lenafc	. . . . C . . . . . - . . TC . . . . . A- . . . . . -	H	1
Ob1	. . . . . -. . . CA . . . . . A- . . . . . - C	I	1
Ob3	A . .T . . C . . . - . . TC . . . . . A- . . . . . .G- . . . AT . .	C	1
Ob4,10,14	. . . . . A . .- . . . C . . . . . A- . . . . . -	J	3
Ob6	. . . . . .- . . TC . . . . . A- . . . . A. -	K	1
Ob7	. . . . . .- . . . C . . . . . A- . . . . . -	L	1
Ob9	. . . . . A. .- . . . C . . . . . A-. . . . . - . -	M	1
Ob11	. CCT . C . AA.CT . . TC.T . .GGA- . . . G.G-C . . . . G.	N	1
Ob12	A . .T . . . . . -. GTC . . C. . A- . . . . . G.	O	1
Ob13	A . .T . . . . . T. - . . TC . . . . . A- . . . . . -.	P	1
Yenisei2	. . . . . .- . . . C . . . . . -. . . . . - . . C	Q	1
Yenisei3,4	. . . . . .- . . . C . . . . . -. . . . . -.	R	2
Yenisei5	A . .T . . . . . -. . TC . . C . . .A- . . . . . -.	S	1
A. gueldenstaedtii	. . .T.A . . AA . . .T . . TC.TCG.A- . . . . . G-C . . AAG .		

Discussion

DNA barcoding and taxonomic boundaries within the Siberian sturgeon

The control region trees offer little resolution, with the paraphyletic assemblage recovered comprised of individuals from all four subspecies and geographic regions (Figure 2). As such, neither the tree nor the character-based approach can be used to reject the hypothesis that all four subspecies are the same single specific unit. The analysis of molecular variance does, however, support population-level differentiation amongst subspecies, although the relatively limited and unequal sample sizes involved should temper over-interpretation of these results. Ultimately, subspecies within *A. baerii* illustrate no definitive taxonomic distinction, as suggested by previous works (Ruban 1998, 1999, 2005; Doukakis et al. 1999; Birstein et al. 2000).

The structure recovered in the tree supports previous findings as well as biogeographic history. The clustering of individuals from the Selenga River (Lake Baikal) and Yenisei River populations supports the findings of previous morphological studies (Ruban 1999, 2005). The Yenisei River was historically connected with Lake Baikal through the Angara River, the only river that flows out from the lake. In the 1950–1980s, a series of hydroelectric dams were built on the Angara River, effectively eliminating connectivity between these populations. The one Yenisei sturgeon with a haplotype similar to the captive group of Baikal sturgeon (from Konakovo) shows that the genetic variability within this population might be even higher than what we observe in this study.

DNA barcoding and Acipenser taxonomy

Our *Cox1* data for *A. baerii* and the data for five American *Acipenser* species (Hubert et al. 2008) point to the possibility of using the *Cox1* barcode as a tool in the future reevaluation of taxonomy of the Eurasian *Acipenser* species. In addition, the *Cox1* barcodes can be used to address questions about recognizing subspecies designations in sturgeons. In our 1997 review, we suggested that 12 Eurasian *Acipenser* species should be recognized, but with some reservation about the Persian sturgeon *A. persicus* (Birstein and Bemis 1997). Described as a species on the basis of a study of a museum specimen, later it was considered a subspecies of *A. gueldenstaedtii*, *A. gueldenstaedtii persicus*. Subsequently, it was elevated to the species level (reviewed in Ruban et al. 2008). However, our cytochrome *b* data suggest that *A. persicus* should be included in *A. gueldenstaedtii*

(Birstein et al. 2000, 2005). A more detailed combined morphological and cytochrome *b* analysis of *A. persicus* in comparison with *A. gueldenstaedtii* also did not support the species validity of *A. persicus* (Ruban et al. 2008). Therefore, currently 11 Eurasian *Acipenser* species should be recognized: *A. baerii*, *Acipenser dabryanus*, *A. gueldenstaedtii*, *Acipenser mikadoi*, *Acipenser naccarii*, *Acipenser nudiventris*, *Acipenser ruthenus*, *Acipenser schrenkii*, *Acipenser sinensis*, *Acipenser stellatus,* and *Acipenser sturio.* In the future, *Cox1* barcoding should be applied for testing these taxonomic units.

By addressing the question of subspecies within *A. baerii*, it becomes clear that the subspecies level should not be recognized within this *Acipenser* species. The only other example of two subspecies within the genus *Acipenser* the American Atlantic sturgeon *Acipenser oxyrhynchus*, *A. oxyrhynchus oxyrhynchus* and *A. oxyrhynchus desotoi* needs special attention. The validity of these two subspecies is questionable since morphologically they are similar, and the only significant difference is the length of the spleen. Evidently, a detailed combined morphological and molecular comparison is necessary for establishing the validity of the subspecies.

DNA barcoding and hatchery individuals

In our previous study (Doukakis et al. 1999), we assumed that individuals kept at the Konakovo Hatchery represented the historic wild population of the Baikal sturgeon. Our new results demonstrate a more complex structure within this species than we assumed. The wild-collected samples from the Selenga River show a single diagnostic relative to all other populations surveyed for the control region (position 320C → T). These two populations are also significantly differentiated based on AMOVA ($F_{ST} = 0.96081$, $p = 0.0002$). The uniqueness of the Selenga River population could be due to a number of factors. First, since only a single female was used to start the Selenga Hatchery population in 1998, a unique haplotype could have been selected and propagated in captive breeding, assuming that the fish sampled here represent the animals that originated in the hatchery. Although unlikely, the wild sample could also be from wild reproduction of a population lacking genetic diversity. Environmental selection could also be acting upon juveniles released from the hatcheries or the wild population as the Selenga River is highly polluted by PCBs from the Selenga Cardboard Mill located upstream of the sturgeon spawning site (Tarasova et al. 1997). Still another hypothesis is that the low genetic diversity in wild individuals is due to the sampling strategy used during field collection. Whether the distinction observed is representative of population distinction or sampling cannot be confirmed without further field collection, which is nearly impossible at present.

The population maintained at the Konakovo hatchery was initiated using females captured in Lake Baikal and thus the distinction from the wild-caught individuals is puzzling. As this captive population exhibits more similarity to one individual from the Ob River than to the individuals collected in the Selenga River (Figure 2), it may harbor genetic diversity close to that in the original population. Such diversity could have been lost during the bottleneck that occurred when the Selenga Hatchery was established or in the wild population. If not an artifact of sampling, the absence of the haplotypes present in the Konakovo hatchery within the wild-caught individuals could indicate that release programs from this hatchery may have been ineffective.

Conservation implications and DNA barcoding of Siberian sturgeons

Our findings present a challenge to future efforts to stock Lake Baikal using available hatchery animals as it is unclear which population represents the historic genetic make-up and therefore should be used for restocking. Furthermore, if selection on released individuals is occurring, restocking with inadequate forms may have little benefit.

Our data show that the *Cox1* gene fragment displays no phylogeographic structure and has no utility in diagnosing putative subspecies of *A. baerii*. While DNA barcoding was unable to provide resolution at this level, the unique barcode haplotype recovered from all specimens analysed suggests that barcoding may be beneficial for species identification. Barcoding has been validated for use in forensic genetic species identification (Dawnay et al. 2007) and yields promising results for shark conservation (Ward et al. 2008b). Given the strong commercial interest in sturgeon roe and the importance of effective species monitoring for conservation enforcement, DNA barcoding may provide an effective way to determine the species of origin of commercial sturgeon products and an additional tool to those already developed (DeSalle and Birstein 1996; Birstein et al. 1998). The utility of DNA barcoding in identifying larval fish (Pegg et al. 2006), fish fillets (Wong and Hanner 2008), and smoked fish products (Smith et al. 2008) has already been demonstrated.

In addition, in North America, the five species of *Acipenser* (*Acipenser brevirostrum*, *Acipenser fulvescens*, *Acipenser medirostris*, *Acipenser oxyrinchus,* and *A. transmontanus*) each possesses a unique *Cox1* barcode haplogroup that is readily identified by DNA barcoding (Hubert et al. 2008). Our data suggest this is also likely for *A. baerii*, although a comprehensive database is critical for confident barcode-based identification (Ekrem et al. 2007).

Until more specimens and more data can be collected from the control region, an effective probabilistic method of assigning unknowns to the population level remains tenuous. Alternatively, the lack of *Cox1* variation among *A. baerii* populations suggests that a molecular character-based approach toward species-level identifications through DNA barcoding could be possible. Either way, the development of a sound reference collection for all sturgeon species will need to be made in order to substantiate results (Ruedas et al. 2000; Por 2007; McKelvey et al. 2008), which is particularly important for molecular diagnostic applications of a regulatory nature (e.g. Yancy et al. 2007).

Acknowledgements

The authors are thankful to Russian colleagues S. G. Afanasiev (Vostsibrybtsentr, Irkutsk, Russia) and S. V. Kirilchik (Limnological Institute, Siberian Division of the Russian Academy, Irkutsk, Russia) for providing the tissue samples. The present work was partially supported through funding to the Canadian Barcode of Life Network from Genome Canada (through the Ontario Genomics Institute), NSERC, and other sponsors listed at www.BOLNET.ca.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

Afanasiev SG. 2006. Ecological basis for breeding of the Baikal Sturgeon (*Acipenser baeri baikalensis* A. Nikolski), Dissertation. Irkutsk, Russia.

Afanasiev GA, Afanasieva VG. 1996. The condition and reproduction of the Baikal sturgeon population. Rybovodstvo i Rybolovstvo 2:6–7, (in Russian).

Besansky NJ, Severson DW, Ferdig MT. 2003. DNA barcoding of parasites and invertebrate disease vectors: What you don't know can hurt you. Trends Parasitol 12:545–546.

Birstein VJ, Bemis WE. 1997. How many species are there within the genus *Acipenser?* In: Birstein VJ, Waldman JR, Bemis WE, editors. Sturgeon biodiversity and conservation. Dordrecht: Kluwer Academic Publishers. p 157–163.

Birstein VJ, DeSalle R. 1998. Molecular phylogeny of Acipenserinae. Mol Phylogenet Evol 9:141–155.

Birstein VJ, Ruban G. 2004. A comment on the Siberian, *Acipenser baerii*, and Russian, *Acipenser gueldenstaedtii*, sturgeons. Environ Biol Fishes 70:91–92.

Birstein VJ, Doukakis P, Sorkin B, DeSalle R. 1998. Population aggregation analysis of caviar producing species of sturgeon and implications for diagnosis of black caviar. Conserv Biol 12: 766–775.

Birstein VJ, Doukakis P, DeSalle R. 2000. Polyphyly of mtDNA lineages in the Russian sturgeon, *Acipenser gueldenstaedtii*: Forensic and evolutionary implications. Conserv Genet 1: 81–88.

Birstein VJ, Ruban G, Ludwig A, Doukakis P, DeSalle R. 2005. The enigmatic Caspian Sea Russian sturgeon: How many cryptic forms does it contain? Syst Biodivers 3:203–218.

Buroker NE, Brown JR, Gilbert TA, O'Hara PJ, Beckenbach AT, Thomas WK, Smith MJ. 1990. Length heteroplasmy of sturgeon mitochondrial DNA: An illegitimate elongation model. Genetics 124:157–163.

Costa FO, Carvahlo GR. 2007. The barcode of life initiative: Synopsis and prospective societal impacts of DNA barcoding of fish. Genomics Soc Policy 3:29–40.

Dasmahapatra KK, Mallet J. 2006. DNA barcodes: Recent successes and future prospects. Heredity 97:254–255.

Davis JI, Nixon KC. 1992. Populations, genetic variation, and the delimitation of phylogenetic species. Syst Biol 41:421–435.

Dawnay N, Ogden R, McEwing R, Carvalho GR, Thorpe RS. 2007. Validation of the barcoding gene CO1 for use in forensic genetic species identification. Forensic Sci Int 173:1–6.

DeSalle R. 2006. Species discovery versus species identification in DNA barcoding efforts: Response to Rubinoff. Conserv Biol 20: 545–1447.

DeSalle R. 2007. Phenetic and DNA taxonomy; a comment on Waugh. Bioessays 29:1289–1290.

DeSalle R, Birtsein VJ. 1996. PCR identification of black caviar. Nature 381:197–198.

DeSalle R, Williams AK, George M. 1993. Isolation and characterization of animal mitochondrial DNA. In: Zimmer EA, White TJ, Cann RL, Wilson AC, editors. Molecular evolution: Producing biochemical data. Vol. 224. San Diego, CA: Academic Press. p 176–204.

DeSalle R, Egan MG, Siddall M. 2005. The unholy trinity: Taxonomy, species delimitation and DNA barcoding. 2005. Philos Trans R Soc Lond Ser B 360:1905–1916.

Doukakis P, Birstein VJ, Ruban GI, DeSalle R. 1999. Molecular genetic analysis among subspecies of two Eurasian sturgeon species, *Acipenser baerii* and *A. stellatus*. Mol Ecol 8:S117–S127.

Ekrem T, Willassen E, Stur E. 2007. A comprehensive DNA sequence library is essential for identification with DNA barcodes. Mol Phylogenet Evol 43:530–542.

Eschmeyer WN, editor. 2008. Catalog of fishes, [online version]. Available at: http://research.calacademy.org/research/ichthyology/catalog/fishcatmain.asp.

Excoffier L, Laval G, Schneider S. 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. 1994. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol 3:294–299.

Gómez A, Wright PJ, Lunt DH, Cancino JM, Carvalho GR, Hughes RN. 2007. Mating trials validate the use of DNA barcoding to reveal cryptic speciation of a marine bryozoan taxon. Proc R Soc Lond Ser B 274:199–207.

Hasegawa M, Kishino H, Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22:160–174.

Hebert PDN, Ratnasingham S, deWaard JR. 2003. Barcoding animal life: Cytochrome *c* oxidase subunit 1 divergences among closely related species. Philos Trans R Soc Ser B 270:S96–S99.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM. 2004. Identification of birds through DNA barcodes. PLoS Biol 2: e312.

Hickerson MJ, Meyer CP, Moritz C. 2006. DNA barcoding will often fail to discover new animal species over broad parameter space. Syst Biol 55:729–739.

Hubert N, Hanner R, Holm E, Mandrak NE, Taylor E, Burridge M. et al. 2008. Identifying Canadian freshwater fishes through DNA barcodes. PLoS One 3:e2490.

Kimura M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–2948.

Lefébure T, Douady CJ, Gouy M, Gibiert J. 2006. Relationship between morphological taxonomy and molecular divergence within Crustacea: Proposal of a molecular threshold to help species delimitation. Mol Phylogenet Evol 40:435–447.

Lipscomb D, Platnick N, Wheeler Q. 2003. The intellectual content of taxonomy: A comment on DNA taxonomy. Trends Ecol Evol 18:64–66.

Maddison WP, Maddison DR. 2002. MacClade: Analysis of phylogeny and character evolution. Version 3.0. Sunderland, MA: Sinauer Associates.

McKelvey KS, Aubry KB, Schwartz MK. 2008. Using anecdotal occurrence data for rare or elusive species: The illusion of reality and a call for evidentiary standards. BioScience 58:549–555.

Meyer CP, Paulay G. 2005. DNA barcoding: Error rates based on comprehensive sampling. PLoS Biol 3:e422.

Moritz C, Cicero C. 2004. DNA barcoding: Promises and pitfalls. PLoS Biol 2:e354.

Neigel J, Domingo A, Stake J. 2007. DNA barcoding as a tool for coral reef conservation. Coral Reefs 26:487–499.

Pegg GG, Sinclair B, Briskey L, Aspden WJ. 2006. MtDNA barcode identification of fish larvae in the south Great Barrier Reef, Australia. Sci Mar 70:7–12.

Por FD. 2007. A "taxonomic affidavit": Why it is needed? Integr Zool 2:57–59.

Posada D, Crandall KA. 1998. Modeltest: Testing the model of DNA substitution. Bioinformatics 14:817–818.

Prendini L. 2005. Comment on "Identifying spiders through DNA barcoding". Can J Zool 83:498–504.

Ratnasingham S, Hebert PDN. 2007. The barcode of life data systems (www.barcodinglife.org). Mol Ecol Notes 7:355–364.

Rock J, Costa FO, Walker DI, North AW, Hutchinson WF, Carvalho GR. 2008. DNA barcodes of fish of the Antarctic Scotia Sea indicate priority groups for taxonomic and systematics focus. Antarct Sci 20:253–262.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.

Ruban GI. 1997. Species structure, contemporary distribution and status of the Siberian sturgeon, *Acipenser baerii*. Environ Biol Fishes 48:221–230.

Ruban GI. 1998. On the species structure of the Siberian sturgeon *Acipenser baerii* Brandt (Acipenseridae). Voprosy Ikhtiologii 38: 307–327, (in Russian) (English translation: J Ichthyol 38, 345-365).

Ruban GI. 1999. Siberian sturgeon *Acipenser baerii* (species structure and ecology). Moscow: (in Russian) GEOS. p 235.

Ruban GI. 2005. The Siberian sturgeon *Acipenser baerii* Brandt: Species structure and ecology. Special Publication No. 1. Norderstedt, Germany: Books on Demand, GmbK World Sturgeon Conservation Society.

Ruban GI, Kholodova MV, Kalmykov VA, Sorokin PA. 2008. Morphological and molecular genetic study of the Persian sturgeon *Acipenser persicus* Borodin (Acipenseridae) taxonomic status. J Ichthyol 48:891–903.

Rubinoff D. 2006. Utility of mitochondrial DNA barcodes in species conservation. Conserv Biol 20:1026–1033.

Ruedas LA, Salazar-Bravo J, Dragoo JW, Yates TL. 2000. The Importance of being earnest: What, if anything, constitutes a "specimen examined?". Mol Phylogenet Evol 17:129–132.

Schander C, Willassen E. 2005. What can biological barcoding do for marine biology? Mar Biol Res 1:79–83.

Seberg O, Humphries CJ, Knapp S, Stevenson DW, Petersen G, Scharff N, Andersen NM. 2003. Shortcuts in systematics? A commentary on DNA-based taxonomy. Trends Ecol Evol 18: 63–65.

Smith MA, Fisher B, Hebert PDN. 2005. DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: The ants of Madagascar. Philos Trans R Soc Lond Ser B 360:1825–1834.

Smith PJ, McVeagh SM, Steinke D. 2008. DNA barcoding for the identification of smoked fish products. J Fish Biol 72:464–471.

Sokolov LI. 2001. The Siberian sturgeon—Its Lake Baikal subspecies, the Baikal sturgeon, *Acipenser baerii baicalensis* A. Nikolski, 1896. In: Danilov-Danielyan VI, editor. The Russian Federation red data book. Moscow: (in Russian) Astrel. p 259–260.

Sokolov LI, Vasil'ev VP. 1989. *Acipenser baeri* Brandt, 1869. General introduction to fishes Acipenseriformes. In: Holcik J, editor. The freshwater fishes of Europe. Vol. 1. Part-II Weisbaden: Aula-Verlag. p 263–284.

Spies IB, Gaichas S, Stevenson DE, Orr JW, Canino MF. 2006. DNA-based identification of Alaska skates (Amblyraja, Bathyraja and Raja: Rajidae) using cytochrome *c* oxidase subunit I (coI) variation. J Fish Biol 69:283–292.

Swofford DL. 2003. PAUP*: Phylogenetic analysis using parsimony (and other methods). Version 4.0 beta. Sunderland, MA: Sinauer Associates.

Tarasova EN, Mamontov AA, Pastukhov MV. 1997. Dioxin and furan content in the body of Baikal seals. Dokl Biol Sci 354: 297–302.

Victor BC. 2007. *Coryphopterus kuna*, a new goby (Perciformes: Gobiidae: Gobinae) from the western Caribbean, with the identification of the late larval stage and an estimate of the pelagic larval duration. Zootaxa 1526:51–61.

Victor BC. 2008. Redescription of *Coryphopterus tortugae* (Jordan) and a new allied species *Coryphopterus bol* (Perciformes: Gobiidae: Gobiinae) from the tropical western Atlantic Ocean. J Ocean Sci Found 1:1–19.

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN. 2005. DNA barcoding Australia's fish species. Philos Trans R Soc Lond Ser B 360:1847–1957.

Ward RD, Holmes BH, Zemlak TS, Smith PJ. 2007. DNA barcoding discriminates spurdogs of the genus *Squalus*. In: Last PR, White WT, Pogonoski JJ, editors. Descriptions of new dogfishes of the genus *Squalus* (Squaloidea: Squalidae). CSIRO Marine and Atmospheric Research Paper 014. Hobart, Australia, p 117–130.

Ward RD, Holmes BH, Yearsley GK. 2008a. DNA barcoding reveals a likely second species of Asian seabass (barramundi) (*Lates calcarifer*). J Fish Biol 72:458–463.

Ward RD, Holmes BH, White WT, Last PR. 2008b. DNA barcoding Australian chondrichthyans: Results and potential uses in conservation. Mar Freshwater Res 59:57–71.

Waugh J. 2007. DNA barcoding in animal species: Progress, potential and pitfalls. Bio Essays 29:188–197.

Wong E, Hanner RH. 2008. DNA barcoding detects market substitution in North American seafood. Food Res Int 41: 828–837.

Yancy HF, Zemlak TS, Mason JA, Washington JD, Tenge BJ, Nguyen NLT, et al. 2007. The potential use of DNA barcodes in regulatory science: Applications of the regulatory fish encyclopedia. J Food Prot 70:210–217.

Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. J Mol Evol 39:306–314.