Fin whale MDH-1 and MPI allozyme variation is not reflected in the corresponding DNA sequences

Morten Tange Olsen¹, Christophe Pampoulie², Anna K. Danielsdóttir³, Emmelie Lidh¹, Martine Bérube¹, Gisli A. Vikingsson² & Per J. Palsbøll¹

¹Evolutionary Genetics Group, Department of Genetics, Microbiology, and Toxicology, Stockholm University, Svante Arrhenius Väg 20C, S-106 91 Stockholm, Sweden
²Marine Research Institute, Skúlagata 4, IS-101 Reykjavik, Iceland
³Matís, Víniandsleið 12, IS-113 Reykjavik, Iceland
⁴Marine Evolution and Conservation, Centre for Ecological and Evolutionary Studies, University of Groningen, PO Box 11103, 9700 CC Groningen, The Netherlands

Keywords
Adaptation, marine mammals, metabolic enzymes, outlier loci, population structure, selection.

Correspondence
Morten Tange, Section for Evolutionary Genomics, Centre for Geogenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, DK-1350 Copenhagen K, Denmark. Tel: +45 42661525; Fax: +45 35323232; E-mail: mortentolsen@gmail.com

Per J. Palsbøll, Marine Evolution and Conservation, Centre for Ecological and Evolutionary Studies, University of Groningen, PO Box 11103, 9700 CC Groningen, The Netherlands. Tel: +31 (0)50 363 9882; Fax: +31 (0)50 363 9620; E-mail: p.j.palsboll@rug.nl

Funding Information
This work was in part supported by the International Whaling Commissions Scientific Committee (grant number 08-09 to PJP and MB) and by Stockholm University in the form of a doctoral fellowship to MTO.

Received: 4 February 2014; Accepted: 7 February 2014

Ecology and Evolution 2014; 4(10): 1787–1803
doi: 10.1002/ece3.1046

Introduction
Population genetic data have been utilized to infer intraspecific population genetic structure in ecology and conservation since the early 1960s when the advent of experimental methods enabled detection of individual genetic variation (Sick 1961). The appeal of genetic inference methods to assess population genetic structure is grounded in the correlation between the genetic similarity and gene flow among populations. Effects of such gene flow are typically genomewide; however, some loci may appear as outliers, displaying above or below average genetic divergence relative to the genomewide level. Above average population, genetic divergence may be due to divergent selection as a result of local adaptation. Consequently, substantial efforts have been directed toward such outlying loci in order to identify traits subject to local adaptation. Here, we report the results of an investigation into the molecular basis of the substantial degree of genetic divergence previously reported at allozyme loci among North Atlantic fin whale (Balaenoptera physalus) populations. We sequenced the exons encoding for the two most divergent allozyme loci (MDH-1 and MPI) and failed to detect any nonsynonymous substitutions. Following extensive error checking and analysis of additional bioinformatic and morphological data, we hypothesize that the observed allozyme polymorphisms may reflect phenotypic plasticity at the cellular level, perhaps as a response to nutritional stress. While such plasticity is intriguing in itself, and of fundamental evolutionary interest, our key finding is that the observed allozyme variation does not appear to be a result of genetic drift, migration, or selection on the MDH-1 and MPI exons themselves, stressing the importance of interpreting allozyme data with caution. As for North Atlantic fin whale population structure, our findings support the low levels of differentiation found in previous analyses of DNA nucleotide loci.
genetic divergence serves as the basis for delineating a species into conservation and management units (Moritz 1994; Waples and Gaggiotti 2006; Palsbøll et al. 2007). The effects of migration are typically genomewide; however, at occasion, some loci may appear as outliers displaying a substantially higher or lower degree of genetic divergence relative to the genomewide level of genetic divergence. Such signatures are usually inferred as loci subject to either divergent or balancing selection, respectively (Tajima 1989; McDonald and Kreitman 1991; Fu and Li 1993; Kreitman 2000). Divergent selection might be due to unique local adaptations (Protas et al. 2006, 2011; Storz et al. 2007, 2009; McCracken et al. 2009a; Scott et al. 2011; Nielsen et al. 2012), which in turn may warrant additional protective measures (Nielsen et al. 2009b; Allendorf et al. 2010; Ouborg et al. 2010; Hoffmann and Sgro 2011).

Allozymes are different variants of enzymes coded by the same locus (Hunter and Markert 1957; Ingram 1957; Markert and Moller 1959; Crick et al. 1961). Such expressed genetic variation is more likely to be subject to local selection and consequently detected as outliers in comparisons with selectively neutral DNA sequences, such as the mitochondrial control region or single tandem repeat (STR) loci (Ford 2002; Storz and Nachman 2003; Canino et al. 2005; Skarstein et al. 2007; Nielsen et al. 2009b). Allozyme analysis was the primary method to collect population genetic data (Hubby and Lewontin 1966; Lewontin and Hubby 1966) but was largely replaced when dideoxy-terminator nucleotide sequencing (Sanger et al. 1977), STR genotyping (Tautz 1989; Schlotterer et al. 1991) and other methods for detecting changes in the DNA sequence itself became more efficient. Recently, the increased focus on the genetics of adaptive variation in natural populations have renewed the interest in allozyme loci, as these may serve as a good starting point for detecting genomic regions under selection (Wheat et al. 2006; Hemmer-Hansen et al. 2007; Ellegren and Sheldon 2008; Nielsen et al. 2009a; Crease et al. 2011; Kirk and Freeland 2011; Schoville et al. 2012). Most studies of this kind make the implicit assumption that outlying allozyme loci are adaptive, and the different alleles arise due to nonsynonymous nucleotide substitutions in the DNA sequence coding the allozymes. However, cis-regulatory processes, such as alternative splicing of messenger RNA and/or post-translational modifications, may yield similar allozyme variation (King and Wilson 1975; Mann and Jensen 2003; Matlin et al. 2005; Marden 2008; Chen and Manley 2009; Choudhary et al. 2009; Keren et al. 2010; Kelemen et al. 2013). Assessing the relative contribution of protein-coding and cis-regulatory processes in shaping allozyme variation is not only fundamental in understanding locus evolution (Hoekstra and Coyne 2007; Carroll 2008; Barrett and Hoekstra 2011), but also central to the interpretation of allozyme variation in terms of estimating rates of gene flow and population divergence time (e.g., to delineate management units), as the observed variation may be transient and thus not represent the action of migration and/or local adaptation.

Here, we present the results of an assessment of two outlying allozyme loci detected among samples collected from North Atlantic fin whales, Balaenoptera physalus. The North Atlantic fin whale has been the target of multiple population genetic analyses of data collected from allozyme loci (Danielsdottir et al. 1991, 1992), as well as STR genotypes and mitochondrial control region sequences (Bérubé et al. 1998). Early work, based upon allozyme variation, revealed very high levels of genetic divergence among the summer feeding areas of Eastern Canada, around Iceland, Norway, and Atlantic Spain, indicative of low migration rates and substantial population structuring across the North Atlantic (Danielsdottir et al. 1991, 1992). In contrast, subsequent analyses of presumed selectively neutral genetic markers (the mitochondrial control region and STR loci) exhibited low levels of genetic differentiation across the North Atlantic (Bérubé et al. 1998). While intriguing in itself, this discrepancy have resulted in an unclear understanding of North Atlantic fin whale migration patterns and significantly hampered management efforts (IWC 2007, 2009).

The purpose of our study was to examine whether the variation observed in outlying allozyme loci was a result of mutations in the enzyme-encoding nucleotide sequences and thus possibly due to local adaptation. Specifically, we considered the following three possible scenarios that would result in the high levels of genetic divergence reported at the outlying allozyme loci: (1) nucleotide substitutions, possible due to divergent natural selection; (2) technical artifacts relating to differential treatment of samples during collection, storage, and processing; or (3) alternative splicing and/or post-translational modifications of the allozyme loci. To assess the possible effects of these different processes, we first re-analyzed the previously published allozyme dataset to identify the most extreme outlier allozyme loci relative to a novel dataset of 15 STR loci. Subsequently, we extracted DNA from a subset of the fin whales used in the previous allozyme study and sequenced the genes encoding the outlier allozyme loci to identify potential nucleotide substitutions that could account for the observed allozyme phenotypes (electromorphs). Surprisingly, we failed to detect any nonsynonymous substitutions in the exons encoding the outlier allozyme loci (MDH-1 and MPI), suggesting that factors other than genetic drift, migration, and selection may account for electrophoretic variation in allozyme loci.

While a great number of studies in nonmodel species have contrasted population genetic divergence estimated from selectively neutral STR and mitochondrial loci with those obtained from allozyme analyses (e.g., Lemaire et al.

1788 © 2014 The Authors. Ecology and Evolution published by John Wiley & Sons Ltd.
2000; De Innocentiis et al. 2001; Dufresne et al. 2002; Dhuyvetter et al. 2004; Vandewoestijne and Van Dyck 2010; Strand et al. 2012), few have proceeded to assess the variation at the DNA sequences encoding the divergent allozyme loci to assess the underlying molecular mechanisms (Eanes 1999; Pogson 2001; Brunelli et al. 2008; McCracken et al. 2009b; Schoville et al. 2012), and, to the best of our knowledge, none have found that the observed allozyme variation was not reflected in the corresponding DNA sequences.

Material and Methods

Samples

Fin whale liver and muscle tissue samples were collected by biologists during commercial whaling operations undertaken off western Iceland and Spain in the period 1985–1989 (Fig. 1; Table 1). Allozyme data were collected from liver samples of 327 individual fin whales, as detailed in the previous allozyme study (Danielsdottir et al. 1991). STR data were collected from genomic DNA extracted from muscle tissue samples from a total of 400 individuals. Included in these two datasets were 115 individuals from which both allozyme data and STR genotypes were available (i.e., both a liver and a muscle sample had been collected). In addition, we sequenced all exons in the DNA encoding the cytosolic malate dehydrogenase 1 (MDH-1) and mannose-6-phosphate isomerase (MPI) allozymes in a total of 34 animals from Iceland. Each of these individuals had known allozyme electromorphs and were selected to ensure an equal representation of each MDH-1 and MPI allozyme electromorph.

Experimental methods

Allozyme and STR genotyping

The experimental conditions used to generate the allozyme data are described in the study by Danielsdottir et al. (1991) (Table S1). Genomic DNA for STR genotyping was extracted using 15% Chelex 100 Resin (Bio-Rad Inc.) and Proteinase K as outlined by Walsh et al. (1991). The STR loci were amplified as detailed in Table S2 (Valsecchi and...
Amos 1996; Palsbøll et al. 1997; Bérubé et al. 2000). All polymerase chain reactions (PCR, Mullis and Faloona 1987) were performed in a total volume of 10 μL, including 2 μL genomic DNA, 0.6 U DyNAzyme™ DNA polymerase (Finnzymes, Thermo Scientific, Waltham, MA, USA), 10× DNA polymerase buffer with 10 mmol/L Tris-HCl, 1.5 mmol/L MgCl2 and 0.1% Triton X-100 (Finnzymes, Thermo Scientific, Waltham, MA, USA), 8 μmol/L dNTPs and between 0.7–4.0 μmol/L of each forward and reverse primer. PCR amplifications were performed using a Thermal Cycler 225 (MJ Research Inc., St. Bruno, Canada) with 4 min at 94 degrees celsius (°C) followed by 32–35 cycles of each 50 sec at 94°C, 50 sec at 54 to 64°C, and 90 sec at 72°C, and finally a single cycle of 7 min at 72°C. PCR amplification products were separated on an ABI3730 DNA Analyzer™, sized using a GeneScan™ – 500LIZ size standard (Applied Biosystems Inc., Waltham, MA, USA). STR alleles were scored manually using the GeneMapper™ Analysis Software version 4.0 (Applied Biosystems Inc.).

DNA sequencing

DNA sequencing was performed on exons encoding MDH-1 and MPI because they were the two most divergent allozyme loci (see Results). Genomic DNA was extracted using either standard phenol/chloroform extractions (Sambrook et al. 1989) or the DNeasy™ blood and tissue kit according to the manufacturer’s instructions (QIAGEN Inc., Venlo, The Netherlands). Sequencing primers were designed from the alignment of NCBI sequences, and primer performance was conducted using AmplifX v. 1.5.4 (Jullien 2008). When possible, primer pairs were placed in conserved regions in the introns flanking the targeted exons. In some cases, flanking intron sequences were insufficiently conserved in the alignment of NCBI sequences, necessitating the design of primers in the exon to sequence the flanking intron in a small panel of fin whale samples. The fin whale–specific intron sequences obtained in this manner were then subsequently employed as the basis for designing primers for sequencing the exons. PCR conditions consisted of 2 min at 94°C, followed by between 29–35 cycles at 94°C for 30 sec, at 54–60°C for 30 sec, and finally at 72°C for 45–74 sec followed by a single cycle at 72°C for 10 min (Table S4). PCR products were purified by shrimp alkaline exonuclease digestion (Werle et al. 1994) and sequenced using the forward or reverse primers used in the initial PCR, and the ABI BigDye™ Terminator Cycle Sequencing Kit v3.1 (Applied Biosystems Inc.) according to the manufacturer’s protocol. The order of sequencing fragments was resolved on an ABI 3130 Genetic Analyzer™ (Applied Biosystems Inc.), and chromatograms were aligned and manually edited in Geneious™ (v. 5.4, Drummond et al. 2011) using the corresponding human exon sequences as reference. As control, the 11 sequence loci containing single-nucleotide polymorphisms (SNPs) were re-amplified and resequenced in on average 21% (n = 7) of the individuals. In addition, to assess the authenticity of our DNA sequence data, we mapped them to the recently published minke whale (Balaenoptera acutirostrata) genome (Yim et al. 2014) using a BLAST search in the whole-genome shotgun database.

Data analysis

Genetic divergence at allozyme and STR loci

Input files for statistical analyses of the allozyme and microsatellite data were created using CONVERT ver 1.31
Nucleotide substitutions in the MDH-1 and MPI DNA sequences

SNPs in the MDH-1 and MPI sequences of the fin whale were identified as single-nucleotide differences either in the homozygote or in the heterozygote state. The frequencies of each SNP variant as well as the observed and expected heterozygosity were determined using SNPator (Morcillo-Suarez et al. 2008). Pairwise tests of linkage disequilibrium were performed using GENEPOP v. 4.0 (Rousset 2008) and significance assessed using the sequential Bonferroni correction (Holm 1979). We used ARLEQUIN (Excoffier and Lischer 2010) to estimate the sequence-level polymorphism, \(\theta\) (Watterson 1975), and average nucleotide diversity, \(\pi\) (Nei 1987), for the concatenated exon sequences only, as well as for exons and partial intron sequences combined.

Inferred amino acid variation in MDH-1 and MPI

The DNA sequences of the MDH-1- and MPI-coding regions were translated into the corresponding amino acid sequences to identify synonymous and nonsynonymous nucleotide substitutions. To examine homology and provide an additional indication of the authenticity of our inferred fin whale protein sequences, these were aligned and compared with the equivalent MDH-1 and MPI protein sequences from human, cow, pig, rat (Rattus norvegicus), and dog (Canis lupus familiaris) obtained from the NCBI GenBank database (www.ncbi.nlm.nih.gov/genbank) and UniProtKB database (www.uniprot.org) (Table S3). Translations and alignments were performed in Geneious v. 5.4 (Drummond et al. 2011) using a global alignment with free end-gaps, a Blosum62 cost matrix, a gap open penalty of 12 and gap extension penalty of 3.

Alternative factors causing electrophoretic variation in allozyme loci?

In order to assess potential alternative factors causing electrophoretic variation in the fin whale MDH-1 and MPI allozyme loci we performed additional assessments of experimental artifacts, alternative splicing and post-translational modifications (PTMs).

First, to explore potential experimental artifacts caused by sample storage the strength and statistical significance of the correlation between allozyme electromorph frequencies and sampling year was estimated by linear regression and an \(F\)-test as implemented in the Microsoft Excel Analysis ToolPak (Microsoft Inc.).

Second, in the absence of fin whale reference data and given the logistical and ethical difficulties associated with obtaining new high-quality fin whale tissue samples for laboratory testing, we extracted information about active sites, putative splice forms (isoforms), and PTMs in human, mouse, and rat from the UniProtKB, PhosphositePlus (Hornbeck et al. 2004), and PHOSIDA (Gnad et al. 2011) databases, as well as a novel atlas of tissue-specific phosphorylation in the mouse (Huttlin et al. 2010). Next, we used this information on known MDH-1 and MPI protein isoforms in human, mouse, and rat to infer putative protein isoforms in fin whales. The molecular weight, isoelectric point, net electric charge, and instability index of putative isoforms in the fin whale was estimated using the package ProtParam (Gasteiger et al. 2005). In addition, in silico prediction of putative PTM sites for acetylation, phosphorylation, and sumoylation in the fin whale MDH-1 and MPI proteins was performed using NetAcet (Kiemer et al. 2005), NetPhos (Blom et al. 1999), and SUMOsp (Ren et al. 2009), respectively. To reduce the frequency of false positives, we applied the most conservative cutoff values (i.e., “high”) for each of the estimations.

Finally, as the allozyme variation observed at MDH-1 and MPI may be a response to metabolic processes (Slein 1950; Gracy and Noltmann 1968; Proudfoot et al. 1994), we employed linear regression, ANOVA, and Student’s \(t\)-tests implemented in the Microsoft Excel Analysis ToolPak (Microsoft Inc.) to assess the strength and statistical significance of potential correlations in allozyme electromorph frequencies with fin whale body condition. Morphological data for the fin whales included in the allozyme study was obtained from Vikingsson (1990). Estimates of half girth-width, blubber thickness, and total body length, were converted into measures of blubber thickness/body length and half girth-width/body length, respectively, as a measure of the relative body condition of each individual fin whale.
Results

Genetic divergence at allozyme and STR loci

Ten of the 40 enzyme loci screened for allozymes by Daníelsdóttir et al. (1991) and 15 of the STRs yielded consistent and polymorphic genotypes in the majority of fin whale samples (Tables S1 and 3). As expected, the estimates of genetic diversities were higher for STR loci (\(H_O = 0.77, 95\% CI = 0.73–0.81; H_E = 0.80, 95\% CI = 0.76–0.83\)) than allozyme loci (\(H_O = 0.259, 95\% CI = 0.160–0.358; H_E = 0.332, 95\% CI = 0.212–0.452\)). Several allozyme (\(n = 6\)) and STR (\(n = 5\)) loci exhibited a statistically significant degree of heterozygote deficiency and a single STR loci exhibited heterozygote excess. None of the microsatellite and allozyme loci exhibited significant linkage disequilibrium after sequential Bonferroni correction.

Among the allozyme loci, the degree of genetic divergence between sample areas was moderate to high with \(FST = 0.028–0.197\) and significantly different from zero all estimations (Table 2). In contrast, the degree of genetic divergence between sampling areas at the STR loci was low. Estimates of \(FST\) ranged from zero to 0.0008 and did not differ significantly from zero in any of the tests (Table 3). The overall degree of genetic divergence was significantly higher (two-tailed sign test, \(P < 0.002\)) at the allozyme loci (\(FST = 0.103, 95\% CI = 0.049–0.165\)) than at the STR loci (\(FST < 0.001, 95\% CI = 0.000–0.001\)).

In the outlier test, the width of the 99% CIs varied slightly depending upon which mutation model was assumed (infinite alleles vs. stepwise mutation) and the choice of simulation model (“Neutral mean \(FST\)” vs. “Neutral + Forced mean \(FST\)”) as well as whether the allozyme and STR data were analyzed together or separately. However, three allozyme loci MDH-1, MPI, and AK-1 were consistently identified as outlier loci with above average \(FST\) values, suggesting the possibility of divergent selection (Fig. 2). Eleven of the 15 STR loci had lower than average \(FST\)’s in estimations including allozyme and STR data; however, this pattern was not observed when only STR data were analyzed.

Table 2. Estimates of genetic differentiation at 10 allozyme loci among the five sampling groups of North Atlantic fin whales.

<table>
<thead>
<tr>
<th>Loci</th>
<th>ICB5</th>
<th>ICB6</th>
<th>ICB7</th>
<th>ICB8</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESP85</td>
<td>0.1206</td>
<td>0.1149</td>
<td>0.1984</td>
<td>0.1972</td>
</tr>
<tr>
<td>ICB5</td>
<td>0.015–0.231</td>
<td>0.0470</td>
<td>0.1101</td>
<td>0.1625</td>
</tr>
<tr>
<td>ICB6</td>
<td>0.029–0.203</td>
<td>0.003–0.082</td>
<td>0.0280</td>
<td>0.0644</td>
</tr>
<tr>
<td>ICB7</td>
<td>0.050–0.345</td>
<td>0.031–0.178</td>
<td>0.009–0.049</td>
<td>0.0634</td>
</tr>
<tr>
<td>ICB8</td>
<td>0.046–0.373</td>
<td>0.052–0.267</td>
<td>0.025–0.109</td>
<td>0.016–0.113</td>
</tr>
</tbody>
</table>

Pairwise \(FST\) estimates above diagonal (Weir and Cockerham 1984); 95% bootstrap confidence interval below diagonal.

Table 3. Estimates of genetic differentiation at 15 microsatellite loci among the five sampling groups of North Atlantic fin whales.

<table>
<thead>
<tr>
<th>Loci</th>
<th>ICB5</th>
<th>ICB7</th>
<th>ICB8</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESP85</td>
<td>0.0009</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>ICB3</td>
<td>0.0–0.004</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>ICB5</td>
<td>0.0–0.005</td>
<td>0.0–0.001</td>
<td>0.0000</td>
</tr>
<tr>
<td>ICB7</td>
<td>0.0–0.013</td>
<td>0.0–0.003</td>
<td>0.0–0.004</td>
</tr>
<tr>
<td>ICB9</td>
<td>0.0–0.002</td>
<td>0.0–0.001</td>
<td>0.0–0.001</td>
</tr>
</tbody>
</table>

Pairwise \(FST\) estimates above diagonal (Weir and Cockerham 1984); 95% bootstrap confidence interval below diagonal.

Nucleotide substitutions in the MDH-1 and MPI DNA sequences

We designed 18 primer pairs to sequence the MDH-1 and MPI exons and partial introns (Table 4 and Table S4). Each of the inferred fin whale MDH-1 and MPI exons mapped to a single location on the minke whale genome, strongly suggesting that we sequenced the correct genes (Fig. 3). The total sequence coverage was more than 11,000 base pairs (bp), but bidirectional coverage in at least 95% of the animals were only obtained for 3300 bps of the MPI gene and 3908 bp of the MDH-1 gene. In these regions, a total of 18 SNPs were identified. Nine SNPs were detected in the MPI gene; five SNPs were located in the introns and four in the exons. Nine SNPs were also detected in the MDH-1 gene, all of which were located in the introns. Two SNPs located in the introns of MDH-1 and MPI, respectively, could not be consistently genotyped and were therefore omitted from further analysis. As genotyping control, on average, 21% of the individuals were resequenced per sequence locus, revealing a single mismatch in a SNP and thus genotyping error-rate below 1%. There was no correlation between genotyping success and allozyme phenotype (data not shown).

No statistically significant deviations from the expected Hardy–Weinberg genotype frequencies were observed for any SNP. In the MPI gene, four of the 28 pairwise tests of linkage disequilibrium among SNPs were statistically significant at the 5% level after sequential Bonferroni correction, and two of the 36 pairwise linkage disequilibrium tests were significant at the 5%
level in MDH-1 gene. All nonsignificant tests were between SNPs with minor allele frequencies (i.e., <5%), suggesting that the lack of significant LD within genes could be due to small sample sizes (i.e., few gene copies of the rare SNP allele), rather than recombination. We did not detect linkage between SNPs located on different genes. The average nucleotide diversity \( \pi \) and level of polymorphism \( \theta \) was low for the exons and partial introns of MDH-1 (\( \pi = 0.0005; \theta = 0.0008 \)), for the MDH-1 exons alone (\( \pi = 0.0006; \theta = 0.0013 \)) and for the exons and partial introns of the MDH-1 gene (\( \pi = 0.0002; \theta = 0.0007 \)) and was zero for the MDH-1 exons alone (which did not contain SNPs).

**Inferred amino acid variation in MDH-1 and MPI**

The amino acid sequences inferred from the fin whale MDH-1- and MPI-coding DNA sequences were similar to the annotated amino acid sequences from other mammals. The pairwise identity scores averaged 96% and 88% for inferred MDH-1 and MPI amino acid sequences, respectively (Figs. S1 and S2). Interestingly, as no SNPs were detected in the exons of the MDH-1 gene and all SNPs located in the exons of the fin whale MPI gene were synonymous substitutions, our DNA sequence data did not indicate variation in the fin whale MDH-1 and MPI proteins. That is, the nucleotide sequences obtained from the exons coding the MDH-1 and MPI allozymes appeared incompatible with the previously reported allozyme variation and differentiation among North Atlantic fin whales.

**Table 4.** Characteristics of the 18 single-nucleotide polymorphisms (SNPs) detected in the MDH-1 and MPI genes of the fin whale. A SNP in intron 6 of the MPI gene (MPI6-367) and a SNP in intron 8 of MDH-1 (MDH9-022) could not be genotyped consistently and were omitted from further analyses.

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP ID</th>
<th>PCR Locus</th>
<th>Locus position</th>
<th>Hs Region</th>
<th>Hs Position</th>
<th>Alleles</th>
<th>N</th>
<th>MAF</th>
<th>( H_0 )</th>
<th>( H_E )</th>
<th>( P )</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDH-1</td>
<td>MDH1-184</td>
<td>Mdh1-1</td>
<td>150</td>
<td>Intron 1</td>
<td>184</td>
<td>A/G</td>
<td>34</td>
<td>0.029</td>
<td>0.059</td>
<td>0.057</td>
<td>0.965</td>
</tr>
<tr>
<td></td>
<td>MDH2-393</td>
<td>Mdh1-2</td>
<td>193</td>
<td>Intron 2</td>
<td>5873</td>
<td>C/T</td>
<td>34</td>
<td>0.147</td>
<td>0.235</td>
<td>0.251</td>
<td>0.834</td>
</tr>
<tr>
<td></td>
<td>MDH5-418</td>
<td>Mdh1-5</td>
<td>258</td>
<td>Intron 5</td>
<td>10564</td>
<td>A/G</td>
<td>34</td>
<td>0.074</td>
<td>0.147</td>
<td>0.136</td>
<td>0.854</td>
</tr>
<tr>
<td></td>
<td>MDH6-257</td>
<td>Mdh1-6</td>
<td>257</td>
<td>Intron 6</td>
<td>15505</td>
<td>A/G</td>
<td>34</td>
<td>0.191</td>
<td>0.324</td>
<td>0.309</td>
<td>0.857</td>
</tr>
<tr>
<td></td>
<td>MDH6-333</td>
<td>Mdh1-6</td>
<td>333</td>
<td>Intron 6</td>
<td>15581</td>
<td>A/C</td>
<td>34</td>
<td>0.029</td>
<td>0.059</td>
<td>0.057</td>
<td>0.965</td>
</tr>
<tr>
<td></td>
<td>MDH6-562</td>
<td>Mdh1-6</td>
<td>562</td>
<td>Intron 6</td>
<td>15817</td>
<td>A/T</td>
<td>34</td>
<td>0.029</td>
<td>0.059</td>
<td>0.057</td>
<td>0.965</td>
</tr>
<tr>
<td></td>
<td>MDH6-563</td>
<td>Mdh1-6</td>
<td>563</td>
<td>Intron 6</td>
<td>15818</td>
<td>A/T</td>
<td>34</td>
<td>0.015</td>
<td>0.029</td>
<td>0.029</td>
<td>0.988</td>
</tr>
<tr>
<td></td>
<td>MDH8-345</td>
<td>Mdh1-8</td>
<td>345</td>
<td>Intron 8</td>
<td>17312</td>
<td>C/T</td>
<td>33</td>
<td>0.015</td>
<td>0.030</td>
<td>0.030</td>
<td>0.988</td>
</tr>
<tr>
<td></td>
<td>MDH9-022</td>
<td>Mdh1-9</td>
<td>22</td>
<td>Intron 8</td>
<td>17788</td>
<td>C/T</td>
<td>34</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPI</td>
<td>MPI1-161</td>
<td>Mpi-1</td>
<td>161</td>
<td>Intron 1</td>
<td>76</td>
<td>C/T</td>
<td>34</td>
<td>0.353</td>
<td>0.294</td>
<td>0.457</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>MPI3-280</td>
<td>Mpi-3</td>
<td>280</td>
<td>Exon 3</td>
<td>1469</td>
<td>C/T</td>
<td>34</td>
<td>0.044</td>
<td>0.088</td>
<td>0.084</td>
<td>0.935</td>
</tr>
<tr>
<td></td>
<td>MPI3-296</td>
<td>Mpi-3</td>
<td>296</td>
<td>Exon 3</td>
<td>1485</td>
<td>C/T</td>
<td>34</td>
<td>0.044</td>
<td>0.088</td>
<td>0.084</td>
<td>0.935</td>
</tr>
<tr>
<td></td>
<td>MPI5-397</td>
<td>Mpi-5</td>
<td>397</td>
<td>Exon 5</td>
<td>3246</td>
<td>C/G</td>
<td>34</td>
<td>0.441</td>
<td>0.353</td>
<td>0.493</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>MPI6-326</td>
<td>Mpi-6</td>
<td>420</td>
<td>Exon 6</td>
<td>6234</td>
<td>C/T</td>
<td>33</td>
<td>0.046</td>
<td>0.030</td>
<td>0.087</td>
<td>0.249</td>
</tr>
<tr>
<td></td>
<td>MPI6-367</td>
<td>Mpi-6</td>
<td>367</td>
<td>Intron 6</td>
<td>6275</td>
<td>G/T</td>
<td>33</td>
<td>NA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MPI6-420</td>
<td>Mpi-6</td>
<td>420</td>
<td>Intron 6</td>
<td>6328</td>
<td>A/G</td>
<td>34</td>
<td>0.059</td>
<td>0.118</td>
<td>0.111</td>
<td>0.898</td>
</tr>
<tr>
<td></td>
<td>MPI78-456</td>
<td>Mpi-78</td>
<td>456</td>
<td>Intron 7</td>
<td>7347</td>
<td>C/T</td>
<td>34</td>
<td>0.059</td>
<td>0.118</td>
<td>0.111</td>
<td>0.898</td>
</tr>
<tr>
<td></td>
<td>MPI78-486</td>
<td>Mpi-78</td>
<td>486</td>
<td>Intron 7</td>
<td>7386</td>
<td>A/G</td>
<td>34</td>
<td>0.015</td>
<td>0.029</td>
<td>0.029</td>
<td>0.988</td>
</tr>
</tbody>
</table>

PCR Locus, the PCR locus referred to in Table S4; Hs Region, regional location in the human gene (MDH-1 gene ID: 154200; MPI gene ID: 4351); Hs Position, position in the human gene; N, number of samples genotyped; MAF, minor allele frequency; \( H_0 \), observed heterozygosity; \( H_E \), expected heterozygosity; \( P \), probability of the SNP being in Hardy–Weinberg equilibrium; NA, not analyzed because of genotyping uncertainties.
Alternative factors causing electrophoretic variation in allozyme loci?

The assessment of potential effects of sample storage upon allozyme variation revealed that the frequencies of the two allozyme electromorphs detected at the MDH-1 locus were strongly and highly significantly correlated with sampling year among the Icelandic fin whale samples. The frequency of the most negatively charged MDH-1 allozyme electromorph decreased at a rate at approximately 17% per year (Fig. 4A). In contrast, the change in MPI allozyme electromorph frequencies did not correlate with time (Fig. 4B).

In our in silico inference of likely alternative splicing we found three known isoforms of the human MDH-1 enzyme, here denoted I-III (Table S5). Isoform III appears to be specific to humans and was disregarded in the subsequent analyses. To examine whether the two fin whale electromorphs reported in the allozyme study may correspond to the remaining MDH-1 isoforms I and II observed in humans, we estimated their electrical charge of the putative MDH-1 isoforms I and II using the inferred fin whale amino acid sequence. A putative fin whale isoform I was inferred which corresponded to the human isoform II and which would carry a net negative charge, whereas a putative isoform II would carry a net positive charge in fin whales, suggesting that the two isoforms would migrate in different directions on a polyacrylamide gel. As this contrasts with the electrophoretic pattern reported for the slow and fast MDH-1 electromorphs observed in the allozyme study (Fig. S3), we assume that those are different from the human isoforms II and III. The human MPI protein exists in four known isoforms denoted I-IV. Again assuming that the above isoforms occur in fin whales, we found that all isoforms carries a net negative charge and hence could be the slow and fast electromorphs observed in the allozyme study (Fig. S3). Thus, we did not find support for alternative splicing in fin whale MDH-1, but it could occur in MPI.

We identified 18 MDH-1 amino acid residues that were known targets of PTM in human, mouse, and/or rat, and an additional four inferred amino acid residues were predicted in silico as PTM sites using NetAcet, NetPhos, and SUMOsp (Table S6). Two PTM sites are known in
human, mouse, and/or rat MPI, and the NetPhos and SUMOsp approaches inferred additional 10 nucleotide sites based on the inferred fin whale MPI amino acid sequence. In MDH-1, one of the 22 known or predicted PTM sites appeared variable across different mammal species, whereas five of the 12 sites in MPI were variable (Figs. S1, S2 and Table S6). Of these, residue 332 of MDH-1 was alanine and a putative acetylation site in fin whale, bottlenose dolphin and cow, but a serine and potential phosphorylation site in human, rat, dog, and pig. Also, residue 389 of MPI was serine and predicted phosphorylation site in fin whale, humpback whale and dolphin, but proline in human, rat, dog, pig, and cow. These two sites may be subject to post-translational modifications in fin whales.

Finally, in assessing the potential correlation between allozyme electromorphs and fin whale body condition, we found that the index of relative body condition in individual fin whales was significantly higher in animals that carried the fast MDH-1 electromorph in homozygote or heterozygote state compared with animals that did not carry this electromorph (Fig. 5A and B), suggesting that this electromorph may be associated with a metabolic process in fin whales. No such patterns were detected for the MPI allozyme electromorphs.

Discussion

Our combined analysis of 10 allozyme and 15 STR loci in North Atlantic fin whales identified three outlier allozyme loci MDH-1, MPI, and AK-1, all of which exhibited well above average levels of genetic divergence among sampling years and localities. However, when sequencing the exons of the two most divergent allozyme loci, MDH-1 and MPI, we only identified four synonymous nucleotide substitutions and no nonsynonymous substitutions. In itself, the low level of genetic polymorphisms in these nuclear loci is consistent with
other studies of baleen whales (Palumbi and Baker 1994; Gaines et al. 2005; Jackson et al. 2009), but the observed absence of nonsynonymous substitutions is in contrast with the generally applied notion that allozyme variation is governed by nucleotide substitutions in the underlying coding DNA sequence (Kreitman 1983; Griffith and Powell 1997; Fields and Somero 1998; Hasson et al. 1998; Pogson 2001; Protas et al. 2006; Wheat et al. 2006, 2010; Linnen et al. 2009; McCracken et al. 2009a, b; Storz et al. 2009; Crease et al. 2011; Scott et al. 2011; Schoville et al. 2012). Still, our findings may not be an uncommon phenomenon in natural populations as the majority of previous reports were based upon readily observable selective agents and differences in phenotype (Griffith and Powell 1997; Fields and Somero 1998; Protas et al. 2006; Wheat et al. 2006, 2010; McCracken et al. 2009a, b; Storz et al. 2009; Crease et al. 2011; Scott et al. 2011) and thus potentially biased toward organisms and genes with clear links between phenotype and genotype. In contrast, natural populations and species typically do not exhibit clear phenotypic differences, may be subject to weaker and cryptic selective agents and/or are difficult to study because of their elusive nature, suggesting that observations like ours may be underrepresented, or unreported. The question is what governs the observed allozyme variation in fin whales if not nonsynonymous substitutions?

**Experimental artifacts?**

It is well known that experimental artifacts may result in the detection of false polymorphisms in analyses of allozyme loci (May 1998). In our analyses of the allozyme data, we made two observations, which could suggest that the polymorphisms reported for the MDH-1 and MPI allozyme loci result from such experimental artifacts. First; the MDH-1 locus exhibited a gradual change in allozyme electromorph frequencies across sampling years, a pattern similar to those associated with experimental bias. For example, electromorph frequencies did not change with time in samples collected from sei whales (Balaenoptera borealis), which were processed simultaneously with the fin whale samples (Danielsdottir et al. 1991). Third, the nonvariable MDH-1 and MPI enzyme loci reported in other studies and species could result from the use of starch gel electrophoresis, which has a lower resolution compared with the polyacrylamide gels used to generate the fin whale allozyme data reported here (Danielsdottir et al. 1991).

Such “hidden” polymorphism owing to the use of different electrophoretic conditions is common among allozyme studies (Bernstein et al. 1973; Cochrane 1976; Coyne 1976, 1982). Finally, artifacts resulting from us sequencing the incorrect DNA regions seems unlikely as our fin whale MHD-1 and MPI DNA sequences each mapped to a single region of the recently published minke whale genome. Also, we observed a large degree of similarity between our inferred exon and protein sequences and the publically available MDH-1 and MPI exon and protein sequences obtained from other mammals. Hence, there is little to suggest that experimental artifacts account for the discrepancy between MDH-1 and MPI enzyme- and DNA-level variation, although the possibility cannot be completely ruled out.

**Alternative splicing and post-translational modifications?**

A plausible explanation for the observed discrepancy between enzyme- and DNA-level variation involves alternative splicing and post-translational modifications (King and Wilson 1975; Matlin et al. 2005; Marden 2008; Chen and Manley 2009; Keren et al. 2010; Kelemen et al. 2013). In our assessment of the fin whale MDH-1 and MPI enzymes, we assumed homology to the corresponding proteins in human, mouse, and rat and found no indication of alternative splicing in fin whale MDH-1, a finding that agrees with preliminary fin whale transcriptome data (Per Palsbøll, unpublished). In contrast, we cannot rule out alternative splicing as a cause for the observed MPI enzyme polymorphisms.

PTMs may result in several differently charged or folded states of the protein through enzyme-catalyzed modifications of the side chains or backbones of the folded protein (Walsh et al. 2005). Apparent polymor-
Function of MDH-1 and MPI in a biological context?

Assuming that the MDH-1 and MPI genes indeed are post-translational modified and alternatively spliced, respectively, in fin whales, are there any characteristics of their cellular function and the biology of fin whales that may provide a clue as to why? North Atlantic fin whales are believed to undertake seasonal movements between feeding and breeding areas (Rorvik and Jonsgard 1981; Donovan 1991; Vikingsson et al. 2009). Fin whales are filter feeders, preying primarily on zooplankton (e.g., euphasiids) to build up adequate fat storages for periods with limited nutritional intake (Hinga 1979; Lockyer 1986, 2007; Vikingsson 1990, 1997). Measurements of blubber thickness and girth-width in the period 1975–1988 document annual variations in fin whale body condition and female fecundity (Lockyer 1986, 2007; Vikingsson 1990), correlating with similar variations in zooplankton biomass (Beare et al. 2000; Lockyer 2007). The combination of high energetic requirements, a relatively short feeding season with unpredictable fluctuations in prey availability, and prolonged periods of reliance on stored lipids (Lockyer 1986, 2007; Potvin et al. 2012) could necessitate a degree of flexibility in the function of key metabolic enzymes that cannot be allowed for by amino acid substitutions, but could be obtained by cis-regulatory processes, such as alternative splicing and PTMs.

Indeed, several studies have documented the role of PTMs in regulating the activity of metabolic enzymes such as MDH-1 in response to cellular demands (Choudhary et al. 2009; Wang et al. 2010; Zhao et al. 2010). In a recent study, Kim et al. (2012) found that increased acetylation of MDH-1 during adipogenesis dramatically enhanced its enzymatic activity. They proposed that this activity supports acetyl coenzyme A (acetyl-CoA) and NADPH in lipid synthesis by accelerating the citrate shuttle and that MDH-1 performs a key function as a cross-talk mechanism between lipid synthesis and intracellular energy levels (Kim et al. 2012).

The observed differences in relative body condition index of fin whale individuals with and without the fast MDH-1 electromorph does point to a link between MDH-1 and fin whale body condition. More specifically, given: (1) reported decreases in zooplankton biomass (Beare et al. 2000) and in fin whale lipid content in the period 1985–1988 (Vikingsson 1990); (2) our findings that the frequency of the MDH-1 fast electromorph decreased during that same period (Fig. 4A); and (3) that the absence of this electromorph was associated with significantly reduced body condition (Fig. 5A–B), we tentatively propose that the fast MDH-1 electromorph result from acetylation of MDH-1 and that its observed decreasing frequency is associated reduced lipid synthesis as a result of limited prey availability.

Implications for the study of natural populations

Regardless of the causative agent, the observed discrepancy between enzyme- and DNA-level variation has important implications for the study of selection and adaptation in natural populations, and for the general use of allozyme markers in population genetic studies. The detection of allozyme, microsatellite, or SNP loci deviating from neutral expectations has often been inferred as evidence for selection in the marker itself or in closely linked genes, and consequently local adaptation (e.g., Eanes 1999; Lemaire et al. 2000; Pogson and Fevolden 2003; Hemmer-Hansen et al. 2007; Larsson et al. 2007; Skarstein et al. 2007; Nielsen et al. 2009b; White et al. 2010; Andre et al. 2011; Kirk and Freeland 2011; Richter-Boix et al. 2011; Chaoui et al. 2012). Our findings stress that deviations from neutral patterns in outlier loci does not imply that such loci are under selection, and if the deviation from neutrality indeed is of biological significance, the underlying mechanisms may be governed by a complex, but more flexible, interplay of protein-coding and cis-regulatory processes. This may in particular concern allozyme loci, which were the markers of choice for several decades (e.g., Bonnell and Selander 1974; Ferguson and Mason 1981; Simonsen et al. 1982a; O’Brien et al. 1983) and still find their use in population genetic studies (Clarke and Whyte 2003; Toda et al. 2003; Curole et al. 2004; Vuorinen and Eskelinen 2005; Matsui et al. 2006; Larson et al. 2007; Addison et al. 2008; Silva and Skibinski 2009; Andre et al. 2011; Chaturvedi et al. 2011; Crease et al. 2011; Pinho et al. 2011; Sa-Pinto et al. 2012; Strand et al. 2012). The data presented here suggest that reports of outlier loci should be interpreted with great caution. Inferred levels of genetic divergence and polymorphisms may not be the product of random genetic drift,
migration or even selection. For this reason, it is advisable to explore the molecular background before making any conclusive inference about population structuring, migration rates, demographic history, and local adaptation from the spatial and temporal distribution of allozyme variation.

Finally, with regard to the North Atlantic fin whale, our findings imply that the population structure inferred by previous allozyme studies (Danielsdóttir et al. 1991, 1992; Danielsdóttir 1994) should be disregarded in future assessments of the population’s management status. Rather, such assessments will require much larger sample sets and number of nuclear genetic markers, as well as the adoption of novel analytical approaches (Økland et al. 2010; Palsbøll et al. 2010) to facilitate the discrimination between recent divergence and high gene flow, both of which are consistent with the low levels of population divergence reported for mtDNA and microsatellite loci (Bérubé et al. 1998).

Acknowledgments

The authors wish to thank staff at the Icelandic Marine Research Institute for sampling, Fred Allendorf, Fred Utter, Bernie May, Anssi Saura and Vibeke Simonsen for fruitful discussions. Also, we wish to thank Steve Palumbi, Allen Moore and Ward Watt, as well as two anonymous reviewers, for their valuable comments on earlier versions of the manuscript. This study was in part supported by the International Whaling Commissions Scientific Committee (grant number 08-09 to PJP and MB) and by Stockholm University in the form of a doctoral fellowship to MTO.

Conflict of Interest

None declared.

References


Gracy, R. W., and E. A. Noltmann. 1968. Studies on phosphomannose isomerases 3. A mechanism for catalysis and...


Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Characteristics of the previously published 10 polymorphic enzyme loci in North Atlantic fin whales that were statistically re-analyzed in the present study.

Table S2. Characterization of the 15 microsatellite loci in North Atlantic fin whale populations included in the analyses.

Table S3. Gene and Protein IDs for selected reference species.

Table S4. Characteristics of the primer pairs designed to amplify and sequence the exons and partial introns of the MDH-1 and MPI nuclear genes of the North Atlantic fin whale.

Table S5. Characteristics of putative alternative splice isoforms of fin whale MDH-1 and MPI. In the three MDH-1 isoforms I-III, isoform I is the canonical isoform and isoform III appears to be specific to humans as exon 1 is not conserved in other mammals (data not shown).

Table S6. Protein residues known or predicted to be targets of post-translational modifications (PTMs).

Table S7. MDH-1 and MPI enzyme loci polymorphisms reported for other fin whale populations and cetacean species.

Figure S1. The fin whale MDH-1 protein aligned with protein sequences from other mammals.

Figure S2. The fin whale MPI protein aligned with protein sequences from other mammals.

Figure S3. Pictures of the original gel electrophoresis of the two enzyme loci, MDH-1 (a) and MPI (b) loci in North Atlantic fin whales. Modified from Daniëlsdóttir (1994).