DNA barcoding provides support for a cryptic species complex within the globally distributed and fishery important opah (*Lampris guttatus*)

JOHN R. HYDE,* KAREN E. UNDERKOFFLER† and MEAGAN A. SUNDBERG†

*National Marine Fisheries Service, Southwest Fisheries Science Center, 8901 La Jolla Shores Dr., La Jolla, CA 92037, USA, †Joint Institute for Marine and Atmospheric Research, University of Hawaii, Honolulu, HI 96818, USA

Abstract

The cornerstone of fisheries management relies on a solid taxonomic base and an understanding of how animals can be grouped into coherent management units. Surprisingly, little is known about the basic biology and ecology of opah (*Lampris guttatus*), a globally distributed species that is commercially exploited and regionally common in the North Pacific. Recent efforts to collect life history data on this species uncovered evidence of two North Pacific morphotypes. Sequencing of the mitochondrial cytochrome c oxidase I gene (655 bp) for these morphotypes and other specimens collected worldwide (n = 480) produced five strongly diverged and well-supported clades. Additional sequence data from the cytochrome b gene (1141 bp) as well as the nuclear recombination activating gene 1 (1323 bp) corroborated these results, suggesting these five clades probably represent separate species. Our conclusion that opah is a complex of five separate species has implications for management and indicates a need to gather additional data on these poorly understood fishes.

Keywords: cryptic species, *Lampris guttatus*, moonfish, opah

Received 14 March 2014; revision received 15 April 2014; accepted 17 April 2014

Introduction

Though not a direct fishery target, the opah or moonfish (*Lampris guttatus*) is often captured in pelagic longline fisheries targeting bigeye tuna (*Thunnus obesus*) and broadbill swordfish (*Xiphias gladius*). While often classified as a nontarget species, the quality of its flesh and relative abundance in central North Pacific fisheries has made it a valuable component of commercial fisheries in this region with 196,047 individuals landed during the period 1994–2011 (Hawn & Collette 2012). The only other extant recognized species in the genus is *Lampris immaculatus*.

Despite its commercial value and common occurrence in pelagic fisheries, little is known regarding the life history of opah. Fossil data show that morphologically similar lamprid ancestors (*Megalampris keyesi*, Gottfried et al. 2006) have existed since at least the Oligocene (approximately 26 Ma) in southern latitudes and have been present off the coast of California since the late Miocene (approximately 6–10 Ma, *Lampris zatima*, Jordan & Gilbert 1920; Jordan 1920). To infer the life history and habitat of this species, there have been a number of cursory studies including electronic tagging (Polovina et al. 2008), examination of unique eye-heater musculature (Runcie et al. 2009), diet studies (Polovina et al. 2008; Choy et al. 2013), early life history studies (Olney 1984), sexual dimorphism (Hawn et al. 2002), live colouration and maximum size (Hawn & Collette 2012), and preliminary age, growth and maturity studies (Francis et al. 2004). However, these studies were typically limited in sample size and geographic scope. In the course of collecting additional life history data on this species, observations were made by the authors and other researchers (Don (DH) Hawn, NMFS personal communication) that suggested the presence of two distinct morphotypes in the North Pacific. We subsequently initiated a study to investigate these differences and began collecting tissue samples for genetic analysis and compiling standard morphometric data.

Materials and methods

Sample collection

Fish were sampled opportunistically at the United Fishing Agency fish auction in Hawaii, by fishery observers.
aboard commercial fishing vessels, from museum collections, and by collaborating fishers. In most cases, a sample of pectoral fin tissue was preserved in 95% undenatured ethanol; in some cases, white muscle or skin samples were obtained instead. The overwhelming majority of samples were obtained from the North Pacific (see Table 1); however, additional samples were obtained from throughout the species’ known range to obtain a good representation of overall diversity (Fig. 1). Because a tissue specimen of the sister species Lampris immaculatus was unavailable, we used a sample of Regalecus glesne as an out-group and included L. immaculatus sequence data when available from GenBank.

**DNA extraction and PCR**

A chelex-based boiling protocol (Walsh et al. 1991; Hyde et al. 2005) was employed to extract DNA from all samples. The ‘barcode’ region of the mitochondrial cytochrome c oxidase I gene (COI) was amplified by polymerase chain reaction (PCR) using primers Fish_COI-F 5’ TCW ACC AAC CAC AAA GAY ATY GGC AC and Fish_COI-R 5’ TAR ACT TCW GGG TGR CCR AAG AAT CA (modified from Ward et al. 2005). A subset of samples had their mitochondrial cytochrome b (cyt b) gene and a portion of their nuclear recombinein activating gene 1 (rag1) amplified using primers Opah_t-Glu-F 5’ TGCCCTGAAAAACCACCGTTG with Opah_tThr-R 5’ TGGCCTCAGCTTTACAAAGCT and Lampridae_Ragl-F 5’ CCTGCCCTCAAGAAYGGTGC with Lampridae_Ragl-R 5’CGAAAAACGCCGAAACAG TTTG, respectively. PCR was performed in 10 µL volumes containing 67 mM Tris-HCl pH 8.8, 16.6 mM (NH4)2SO4, 10 mM β-mercaptoethanol, 2 mM MgCl2, 800 µM dNTPs, 0.4 µM each primer, 0.5 mg/mL bovine serum albumin, 0.5 units Taq DNA polymerase (New England Biolabs) and 50 ng of DNA template. The thermal profile for all reactions consisted of 92°C for 2 min, followed by 35 cycles of 92°C (20 s), 55°C (60 s) and 72°C (60 s). No template, negative controls were run in all PCR sets to control for possible contamination of reagents. Products were electrophoresed through 2% agarose gel, stained with ethidium bromide and visualized on a UV transilluminator. PCR products were enzymatically cleaned using ExoSap-IT (Affymetrix). Cleaned products were sequenced in both directions using BIGDYE TERMINATOR v3.1 (Life Technologies) following manufacturer’s protocols and analysed using an ABI3730 Genetic Analyzer (Life Technologies). Raw sequences were edited and aligned using SEQUENCHER v4.9 (GeneCodes).

**Sequence analyses**

Edited sequences were compiled and distance-based analyses were performed using MEGA v4.1 (Tamura et al. 2007). Pairwise distance measurements between individual sequences were performed using the Kimura 2-parameter (K2P) model. The K2P model was chosen for these analyses so as to be directly comparable to the Ward (2009) study which specifically addressed levels of divergence at the COI gene. The distances obtained were used to assess intraspecific and interspecific divergence and compared to published data generated for other taxa. Standard measures of genetic diversity, including nucleotide and gene diversity as well as counts of transitions and transversions, were performed using ARLEQUIN v3.1.1. (Excoffier et al. 2005).

Initial analyses of COI sequences produced several distinct and deeply diverged lineages. To further investigate this finding, additional sequence data were generated for the rag1 and cyt b genes from a subset of samples from each lineage. Evolutionary model testing for each gene was performed using jModelTest (Posada 2008). Phylogenetic relationships were evaluated for each gene using a Bayesian framework as implemented in BEAST v1.6.1 (Drummond & Rambaut 2007). Trees were rooted with R. glesne because this lampriform species was available for analysis. Whenever possible, sequence

---

**Table 1** Regional sample sizes (n), number of unique haplotypes (H), haplotype diversity (h), nucleotide diversity (π), number of transitions (T), number of transversions (T), number of variable sites (V), uncorrected average pairwise differences between samples (P) and per cent nucleotide composition for cytochrome c oxidase I gene data

<table>
<thead>
<tr>
<th>Location</th>
<th>n</th>
<th>H</th>
<th>h ± SD</th>
<th>π ± SD</th>
<th>T</th>
<th>T</th>
<th>V</th>
<th>P ± SD</th>
<th>% C</th>
<th>% T</th>
<th>% A</th>
<th>% G</th>
</tr>
</thead>
<tbody>
<tr>
<td>North Pacific</td>
<td>408</td>
<td>66</td>
<td>0.894 ± 0.009</td>
<td>0.036 ± 0.018</td>
<td>115</td>
<td>19</td>
<td>125</td>
<td>23.91 ± 10.53</td>
<td>31.5</td>
<td>27.0</td>
<td>22.2</td>
<td>19.3</td>
</tr>
<tr>
<td>Indian</td>
<td>12</td>
<td>7</td>
<td>0.773 ± 0.128</td>
<td>0.039 ± 0.021</td>
<td>44</td>
<td>7</td>
<td>49</td>
<td>25.52 ± 12.06</td>
<td>31.4</td>
<td>27.1</td>
<td>22.3</td>
<td>19.3</td>
</tr>
<tr>
<td>South Pacific</td>
<td>21</td>
<td>8</td>
<td>0.719 ± 0.099</td>
<td>0.002 ± 0.001</td>
<td>6</td>
<td>1</td>
<td>7</td>
<td>1.01 ± 0.71</td>
<td>31.8</td>
<td>26.7</td>
<td>22.4</td>
<td>19.1</td>
</tr>
<tr>
<td>South Atlantic</td>
<td>25</td>
<td>15</td>
<td>0.907 ± 0.045</td>
<td>0.019 ± 0.010</td>
<td>48</td>
<td>8</td>
<td>52</td>
<td>12.33 ± 5.76</td>
<td>31.7</td>
<td>26.8</td>
<td>22.4</td>
<td>19.2</td>
</tr>
<tr>
<td>North Atlantic</td>
<td>4</td>
<td>4</td>
<td>1.000 ± 0.177</td>
<td>0.063 ± 0.042</td>
<td>60</td>
<td>11</td>
<td>71</td>
<td>41.33 ± 22.94</td>
<td>31.4</td>
<td>27.2</td>
<td>22.6</td>
<td>18.9</td>
</tr>
<tr>
<td>S. Atlantic/Indian</td>
<td>9</td>
<td>6</td>
<td>0.889 ± 0.091</td>
<td>0.003 ± 0.002</td>
<td>6</td>
<td>1</td>
<td>7</td>
<td>2.22 ± 1.35</td>
<td>31.8</td>
<td>26.7</td>
<td>22.4</td>
<td>19.1</td>
</tr>
</tbody>
</table>

*These samples lacked locality data other than they came from either of these two regions.

Published 2014. This article is a U.S. Government work and is in the public domain in the USA.
data from the sister taxon \textit{L. immaculatus} were included in the analyses.

**Results**

Analysis of 655 base pairs (bp) of sequence data for the COI gene from 480 \textit{Lampris guttatus} specimens produced 82 unique haplotypes (GenBank accession# JF931865–JF931947). Overall genetic diversity was exceptionally high with 140 polymorphic sites and overall nucleotide diversity of $\pi = 0.038$ (Table 2). These high levels of diversity were largely retained when samples were broken into oceanographic regions (Table 1). Per cent divergence (K2P) between individual \textit{L. guttatus} sequences ranged from 0% to 13.2%, divergence between \textit{L. guttatus} and a single \textit{L. immaculatus} sequence ranged from 12.0% to 14.4%. Levels of divergence were not normally distributed and instead presented as several discrete modes (Fig. 2A). Even though only a small subset of these samples ($n = 19$) were analysed at the other genes, the high diversity was largely retained (Table 2) with per cent divergence for cyt b (13 unique haplotypes, GenBank accession# JF931948–JF931967), and rag1 (14 unique genotypes, GenBank accession# JF931968–JF931987), sequences ranging from 0% to 21.4% (Fig. 2B) and 0% to 0.69% (Fig. 2C), respectively. This variation in

**Table 2** Molecular diversity measures for analysed genes with overall sample sizes ($n$), number of unique haplotypes (H), haplotype diversity ($h$), nucleotide diversity ($\pi$), number of transitions ($T_s$), number of transversions ($T_v$), number of variable sites ($V$), uncorrected average pairwise differences between samples ($P$), per cent nucleotide composition and total base pairs analysed (bp)

<table>
<thead>
<tr>
<th>Gene</th>
<th>$n$</th>
<th>H</th>
<th>$h \pm SD$</th>
<th>$\pi \pm SD$</th>
<th>$T_s$</th>
<th>$T_v$</th>
<th>$V$</th>
<th>$P \pm SD$</th>
<th>% C</th>
<th>% T</th>
<th>% A</th>
<th>% G</th>
<th>bp</th>
</tr>
</thead>
<tbody>
<tr>
<td>COI</td>
<td>480</td>
<td>82</td>
<td>0.896 ± 0.009</td>
<td>0.038 ± 0.018</td>
<td>131</td>
<td>22</td>
<td>140</td>
<td>24.65 ± 10.84</td>
<td>31.6</td>
<td>27.0</td>
<td>22.2</td>
<td>19.3</td>
<td>655</td>
</tr>
<tr>
<td>cyt b</td>
<td>19</td>
<td>13</td>
<td>0.947 ± 0.033</td>
<td>0.090 ± 0.045</td>
<td>245</td>
<td>80</td>
<td>286</td>
<td>102.39 ± 46.04</td>
<td>33.3</td>
<td>27.8</td>
<td>23.0</td>
<td>16.0</td>
<td>1141</td>
</tr>
<tr>
<td>rag1</td>
<td>19</td>
<td>14</td>
<td>0.964 ± 0.028</td>
<td>0.005 ± 0.003</td>
<td>6</td>
<td>7</td>
<td>13</td>
<td>7.12 ± 3.50</td>
<td>23.9</td>
<td>23.1</td>
<td>25.5</td>
<td>27.5</td>
<td>1323</td>
</tr>
</tbody>
</table>

Published 2014. This article is a U.S. Government work and is in the public domain in the USA.
Evolutionary rate among these genes is in line with other studies that have shown COI evolves at approximately 60–70% the rate of the cyt b gene but much faster than the rag genes (e.g. Hyde & Vetter 2007). As is typical for genes that code for functional proteins, the majority of nucleotide substitutions were at 3rd codon positions.

Published 2014. This article is a U.S. Government work and is in the public domain in the USA.
with no resultant change in amino acid sequence. The majority of substitutions that resulted in amino acid changes were confined to interlineage comparisons.

Evolutionary model testing indicated the general time-reversible (GTR) model considering a proportion of invariant sites (I) and gamma distribution (G) (Rodriguez et al. 1990) had the highest likelihood for each of the three data sets (COI, cyt b, rag1) and was used for construction of Bayesian-derived trees using BEAST. The tree with maximum posterior probability for COI data produced five deeply diverged and strongly supported clades within *L. guttatus* (Fig. 3). Individuals

![Bayesian-derived tree with maximum posterior probability for cytochrome c oxidase I gene haplotypes found in this study. Values above nodes represent posterior support values. Lineage names for each of the clades present in this figure are referred to in Figs 4 and 5. Sequence data from *Lampris immaculatus* (GenBank DQ108066) and *Regalecus glesne* (SIO 97–226) are used as out-groups.](image)

Table 3 Average per cent divergence (Kimura 2–P) between cytochrome c oxidase I gene (COI) haplotype clades below diagonal, net divergence above diagonal, average within clade haplotype divergence in bold. Lineages containing only a single specimen are denoted NA for within lineage distance.

<table>
<thead>
<tr>
<th>Lineage</th>
<th>Lineage 1</th>
<th>Lineage 2</th>
<th>Lineage 3</th>
<th>Lineage 4</th>
<th>Lineage 5</th>
<th><em>Lampris immaculatus</em></th>
<th>within lineage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lineage 1</td>
<td>*</td>
<td>11.55</td>
<td>12.35</td>
<td>10.73</td>
<td>11.17</td>
<td>13.87</td>
<td>0.31</td>
</tr>
<tr>
<td>Lineage 2</td>
<td>11.70</td>
<td>*</td>
<td>5.57</td>
<td>6.12</td>
<td>5.71</td>
<td>11.85</td>
<td>NA</td>
</tr>
<tr>
<td>Lineage 3</td>
<td>12.71</td>
<td>5.78</td>
<td>*</td>
<td>6.95</td>
<td>6.65</td>
<td>13.89</td>
<td>0.41</td>
</tr>
<tr>
<td>Lineage 4</td>
<td>11.14</td>
<td>6.37</td>
<td>7.41</td>
<td>*</td>
<td>2.68</td>
<td>12.73</td>
<td>0.50</td>
</tr>
<tr>
<td>Lineage 5</td>
<td>11.55</td>
<td>5.93</td>
<td>7.08</td>
<td>3.15</td>
<td>*</td>
<td>12.17</td>
<td>0.45</td>
</tr>
<tr>
<td>Lampris immaculatus</td>
<td>14.02</td>
<td>11.85</td>
<td>14.10</td>
<td>12.98</td>
<td>12.39</td>
<td>*</td>
<td>NA</td>
</tr>
</tbody>
</table>

Published 2014. This article is a U.S. Government work and is in the public domain in the USA.
within each clade were subdivided and further analysed by these unique COI lineages. Interclade divergence (K2P) ranged from 2.7% to 12.4% while values for intraclad divergence ranged from 0% to 0.9% with average intraclad haplotype divergence ranging from 0.3% to 0.5% (Table 3).

Sequence data from rag1 and cyt b for individuals from each COI lineage produced highly similar trees to that derived from the COI data (Figs 4 and 5). In all cases, there was a 100% concordance between COI clades and those generated for the other two genes; high posterior support resulted in all cases. The placement of Lineage 2 relative to the other lineages was the only structural difference noted between gene trees. Such discrepancies between gene trees are common and can be caused by multiple factors including incomplete lineage sorting, selective or constraining forces acting upon individual genes and hybridization (e.g. Maddison 1997).

Discussion

Ward (2009) found that 92% of conspecific fishes showed <1% haplotype divergence at the barcode region of the COI gene, while 98% showed divergence <2%. Based on these findings, he suggested that intraspecific divergence >2% probably characterized cryptic species complexes. Pairwise distances between COI haplotypes for *L. guttatus* ranged from 0% to 13.2% with the presence of five distinct clades with interclade divergences of 2.7% to 12.4%. Values for intraclad divergence ranged from 0%
to 0.9%. Following Ward (2009), the data suggest that these individual clades represent cryptic species as all interclade divergences are >2% and intraclade divergences are <1%.

To further evaluate the COI findings, we sequenced a subset of samples representing members of the different COI clades, for an additional mitochondrial gene (cyt b) and a nuclear gene (rag1). Trees for all three genes produced the same five distinct clades (Figs 3–5) with strong posterior support. Concordance of the two mitochondrial genes and one nuclear gene strongly support that the five clades found in the COI analysis are on separate evolutionary trajectories; given current divergence and comparison to other fishes (Ward 2009), all five warrant separate species status based upon the genetic data and forthcoming evidence of distinguishing morphological characters (KEU & MAS unpublished data).

Genetic evaluation of *L. guttatus* was initiated due to the presence of two morphotypes observed in the North Pacific. These morphotypes corresponded 100% with two of the COI clades found in this study (Lineages 3 and 5) and are approximately 6.7% diverged at this gene. The morphotypes were originally categorized by relative eye size, but subsequent detailed morphologic study has uncovered several other diagnostic characters that differentiate these fishes (KEU & MAS unpublished data). Although fewer specimens from the other lineages have been available for study, initial data support the conclusion that significant morphologic distinctions exist among them as well. The agreement between morphology and genetics, especially because several of these lineages occur in sympatry (see Fig. 1), adds further support to the assertion that these are separate species. The genetic and morphological data generated in this study are in the process of being combined into formal descriptions for several of the species.

The finding of a cryptic species complex in a widely distributed marine fish that is well-known or

![Bayesian-derived tree with maximum posterior probability for rag1 sequences generated in this study. Values above nodes represent posterior support values. Sequences are labelled with cytochrome c oxidase I gene lineage from Fig. 3. A single sequence from Regalecus glesne (SIO 97-226) is used as an out-group.](image)

Published 2014. This article is a U.S. Government work and is in the public domain in the USA.
an important fishery component is not unique (e.g. Brendtro et al. 2008; Hyde et al. 2008; Yoshita et al. 2009) and may be more common in poorly studied pelagic mid-water species (e.g. Craig et al. 2004; Roberts 2012). Review of literature and sequences deposited in BOLD and GenBank suggest that several lampridiform genera contain cryptic lineages and are probably in need of taxonomic revision. Following detailed genetic and morphologic comparisons, the crestfish genus Lophotus was found to contain at least two species in the Pacific with uncertainty as to the status of fish in other ocean basins (Craig et al. 2004). A thorough review by Roberts (2012) offers multiple lines of evidence for existence of at least two species within Regalecus. Comparisons of conspecific sequences for Zu cristasus and Metavelifer multiradiatus available from the GenBank and BOLD databases show evidence for strongly diverged lineages.

Although mining from online databases has been shown to yield false levels of diversity due to species misidentifications and sequence errors (Harris 2003), these hypotheses can be tested if voucher specimens archived in permanent museum collections can be examined. Much of the taxonomic confusion within this group of fishes is likely due to the lack of type specimens available and relative rarity of capture, which has led to a preponderance of synonyms for several species including opah. In the case of opah, although at least six synonyms currently exist for L. guttatus (L. gunneri, L. imperialis, L. lauta, L. luna, L. regius, L. stroemii), a complete lack of type specimens makes reappraisals of synonymy difficult. Application of genetic barcoding techniques to many of these wide-ranging species will probably yield additional examples of cryptic diversity, better clarify the taxonomy of this group of fishes and add to our understanding of speciation in the mesopelagic realm.

Implications for management

Even though no lampridiform species is at present directly targeted by commercial fisheries, the opah has become an important component in some fisheries, particularly those targeting bigeye tuna and swordfish in the North Pacific. The value of US commercial landings of opah in 1999 to almost $2.5 million in 2009. Results from this study provide clear evidence that L. guttatus is not a single, globally distributed species. The overwhelming majority (99.9%) of US landings of opah occur in Hawaii where it seems at least two species (Lineages 3 and 5) interact with fisheries, though one dominates the catch in this region (Lineage 3, approximately 90%, KEA & MAS unpublished data). However, initial monitoring of catch in California waters shows that the species composition is strongly skewed towards the other species (Lineage 5, ~95%, JRH unpublished data). Clearly, further work needs to be done to delineate the distribution, habitat and life history of these cryptic lineages and how they interact with world fisheries to improve future management plans for this growing fishery. Further, previous studies of life history, diet and habitat will need to be re-evaluated in the light of these findings.

Acknowledgements

Tissue samples were graciously provided by B. Takenaka, N. Arberilla (UFA, Hawaii); B. Luduc (Euronor, France); J. Palsøn (Marine Research Institute, Iceland); S. Kullander, B. Deling, E. Ahlander (Swedish Museum of Natural History, Sweden); M. Bruni (Oceanographic Museum of Monaco, France); A. Hay, S. Reader (Australian Museum, Australia); J. Friel (Cornell University, USA); D. Nelson, A. Belasen (University of Michigan Museum of Zoology, USA); A. Connell, R. Bills (South Africa Institute for Aquatic Biodiversity, South Africa); S. Lamberth (Marine and Coastal Management, South Africa); A. Graham (CSIRO, Australia); S. Kostelnik and participating American Samoa longline observers (NMFS, USA); D. Roje (University of Washington, USA); A. de Vera Hernandez (Museo de Ciencias Naturales de Tenerife, Spain); K. Busscher, J. Kelly, M. Marsik, and all participating Hawaii Longline observers and debriefers (NMFS, USA); G. Shinahara (National Museum of Nature and Science, Japan); J. Clayton, J. Finan, D. Johnson, and J. Williams (National Museum of Natural History, USA); H. J. Walker (Scripps Institution of Oceanography, USA). We are grateful to E. DeMartini, A. Lang, and B. Mundy for providing comments on earlier drafts of this manuscript and G. Kitazaki (Diamond Head Seafood, Hawaii) for pointing out morphological variations in opah specimens to KEA and MAS during UFA visits.

References


Yoshita Y, Yamanoue Y, Sagara K et al. (2009) Phylogenetic relationship of two Mola sunfishes (Tetraodontiformes: Molidae) occurring around the coast of Japan, with notes on their geographical distribution and morphological characteristics. Ichthyological Research, 56, 232–244.

J.R.H., K.E.U., and M.A.S. designed the study. K.E.U. and M.A.S. collected specimens and morphometric data. J.R.H. obtained, edited and analysed genetic data. All authors contributed to the final version of this manuscript.

Data Accessibility