MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2011) doi: 10.1111/j.1755-0998.2011.03059.x

High-throughput molecular identification of fish eggs using multiplex suspension bead arrays

LANI U. GLEASON and RONALD S. BURTON
Marine Biology Research Division, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093-0202, USA

Abstract

The location and abundance of fish eggs provide information concerning the timing and location of spawning activities and can provide fishery-independent estimates of spawning biomass. However, the full value of egg and larval surveys is severely restricted because many species' eggs and larvae are morphologically similar, making species-level identification difficult. Recent efforts have shown that nearly all species of fish may be identified by mitochondrial DNA (mtDNA) sequences (e.g. via 'DNA barcoding'). By taking advantage of a DNA barcode database, we have developed oligonucleotide probes for 23 marine fish species that produce pelagic eggs commonly found in California waters. Probes were coupled to fluorescent microspheres to create a suspension bead array. Biotin-labelled primers were used to amplify the mitochondrial cytochrome oxidase subunit I (COI) and 16S ribosomal rRNA genes from individual fish eggs. The amplicons were then hybridized to the bead array, and after the addition of a reporter fluorophore, samples were analysed by flow cytometry with Luminex 100 instrumentation. Probes specifically targeted eggs that are abundant and/or from morphologically indistinguishable species pairs. Results showed that the 33 different probes designed for this study accurately identified all samples when PCR was successful. Suspension bead arrays have a number of benefits over other methods of molecular identification; these arrays permit high multiplexing, simple addition of new probes, high throughput and lower cost than DNA sequencing. The increasing availability of DNA barcode data for numerous fish faunas worldwide suggests that bead arrays could be developed and widely used for fish egg, larval and tissue identifications.

Keywords: fish egg identification, molecular barcodes, suspension bead arrays

Received 17 March 2011; revision revised 13 June 2011; accepted 7 July 2011

Introduction

Changes in the ocean environment over the past several decades have led to documented alterations in the composition of marine communities (Perry *et al.* 2005; Occhipinti-Ambrogi 2007). Warming of sea surface temperatures has led to the most pronounced effects to date, as temperate regions witness an increase in abundance of warm water species while cold water species' abundance decreases (Barry *et al.* 1995). Although the pole-ward movement of warm water species is already underway (Hsieh *et al.* 2009; Fodrie *et al.* 2010), the ultimate impacts of these movements on marine ecosystems and world fisheries production remain poorly studied (Brander 2010; Jennings & Brander 2010). Understanding how individual component taxa respond to temperature is undoubtedly useful (Murawski 1993), but community characteristics, including changes in predator and prey abundances for example, will ultimately determine the net impacts of climate change on ecosystem composition and fishery productivity (Durant *et al.* 2007).

Methods for monitoring the dynamics of marine fish populations include approaches that focus on adult stocks (e.g. catch data and experimental trawls) as well as those that target early life stages via ichthyoplankton sampling. These methods prove to be rather complementary, because some species are poorly represented in one or the other form of survey. As a fisheries-independent approach to monitoring stock abundance, fish egg surveys play a significant role in the assessment and management of fish stocks in many regions of the world (e.g. Kawakami *et al.* 2010). In such applications, focus is typically on the abundance of relatively few species, and morphological identification of eggs often allows reasonably accurate enumeration (Parker 1980; Lasker 1985; Alheit 1993; Priede & Watson 1993). However, in focusing on these few commercially important species,

Correspondence: Lani Gleason, Fax: (858) 534 7313;
E-mail: lgleason@ucsd.edu

much of the potential value of egg surveys for broader ecological analyses is lost. Because egg surveys can reveal temporal and spatial patterns of spawning activity as well as changes in relative abundances of different species (Bellier *et al.* 2007), there would be great benefit to analysing all the species represented in the survey. Identification and enumeration of all species of eggs (and larvae) would not only enhance estimates of stock abundance but also permit comprehensive monitoring of changes in fish community composition in response to both natural and anthropogenic environmental changes (Hernandez *et al.* 2010a,b). Such an expanded analysis is, in fact, mandated by ongoing efforts to move from single-species to ecosystem-based management (Ecosystem Principles Advisory Panel 1999).

A major obstacle to broader use of egg surveys in monitoring the dynamics of fish communities is that many species of fish eggs and larvae are extremely difficult to morphologically distinguish. In the case of California's marine fish fauna, approximately 100 species produce pelagic eggs and most are spherical in shape and transparent; approximately 70% are between 0.7 and 1.6 mm in diameter (Ahlstrom & Moser 1980). Given natural variation in quantitative traits, identifications based only on size can lead to substantial error. For example, despite their great phylogenetic and ecological divergence, the eggs of *Scomber japonicus* (Pacific mackerel) and *Merluccius productus* (hake) cannot be reliably distinguished by morphology alone until the latter part of embryonic development. Thus, when fish egg surveys are used to determine spatial and temporal spawning patterns and biomass for commercially important species such as *S. japonicus* and *M. productus*, errors in identification can have important consequences for management (Fox *et al.* 2005; Perez *et al.* 2005).

Difficulties with morphological identifications of fish eggs and larvae have led to the increased use of molecular tools for identification (Hyde *et al.* 2005; Karaiskou *et al.* 2007; Carreon-Martinez *et al.* 2010; Lelièvre *et al.* 2010). In fact, this is precisely the type of problem that has led to the global effort known as 'DNA barcoding'. Because all species differ in their DNA sequences, it is often possible to identify organisms using short diagnostic sequences ('barcodes') once an appropriate database (assembling sequences of the same gene for all species of interest) is established (Hebert *et al.* 2003). Such databases are currently being developed for a number of fish faunas (http://www.fishbol.org), primarily focusing on the mitochondrial gene, cytochrome oxidase I (COI; Ward *et al.* 2009).

With a barcode database in hand, molecular identification of any developmental stage of a given species is straightforward, relying on standard methods of PCR amplification of target genes, subsequent DNA sequencing and comparison of the sequence to the database (Ward *et al.* 2005; Yancy *et al.* 2008; Lakra *et al.* 2011). These techniques are quite reliable, but are relatively slow and expensive. Kawakami *et al.* (2010) used sequencing to initially identify eggs but then relied on morphological characteristics of live eggs for identifications. Richardson *et al.* (2007) described a high-throughput system for molecular identification of eggs using robotic sample handling and DNA sequencing; however, the described system could only process approximately 800 samples per week. Furthermore, we estimate that sample costs in the described system would be in the range of $4 per sample. For many ecological applications, this technology would be too slow and costly.

Recently, suspension bead arrays have been developed for various applications of species identification and quantification, ranging from human pathogens to marine microbial communities (Diaz & Fell 2004; Deak *et al.* 2010; Mayali *et al.* 2010). In brief, species-specific oligonucleotide capture probes are attached to fluorescently labelled beads (3- to 5-μm microspheres). Specialized flow cytometers (e.g. BD FACSArray™ Bioanalyzer, Luminex™ 100) use dual laser detection to classify the bead colour and quantify a reporter fluorophore attached to the target. Multiplexing is achieved by using multiple bead colours (each with an associated capture probe) in a single assay. Systems capable of multiplexing up to 500 probes are now commercially available (Luminex FLEXMAP 3D) and can process many samples efficiently and relatively cheaply.

We developed a 33-probe suspension bead array designed to identify 23 different species of California marine fish. We used biotin-labelled PCR primers to simultaneously amplify and label fragments of two mitochondrial genes, cytochrome oxidase subunit 1 (COI) and 16S rRNA. For a given fish species, 25-mer oligonucleotide probes were developed to either gene (because in some cases one gene amplifies better than the other) or both (reconfirming the specificity of the assay). Employing PCR amplification and the Luminex 100 instrument for suspension bead array analysis, we have developed a flexible, high-throughput method for molecular identification of fish eggs that can substantially reduce cost and increase throughput over DNA sequencing approaches. This identification method can be widely applied in any region where DNA barcodes of local fish fauna are available to facilitate the design of species-specific oligonucleotide capture probes.

Materials and methods

Sample collection

Fish egg samples were obtained from several sources. The Ichthyoplankton Laboratory of the Southwest Fisheries

Science Center (SWFSC), a division of the National Oceanic and Atmospheric Administration (NOAA) Fisheries Service, provided a variety of egg samples. For over 60 years, the California Cooperative Oceanic Fisheries Investigations (CalCOFI) programme (a partnership between NOAA Fisheries, Scripps Institution of Oceanography and the California Department of Fish and Game) has conducted quarterly research cruises collecting plankton samples on a prescribed grid along California's coast. Bongo net samples are taken at each of approximately 60 stations; one side of the net sample is preserved in formalin for morphological identification, while the other side is preserved in ethanol. Ethanol preservation was initiated only in the past 14 years of the sampling. We obtained two sets of ethanol-preserved samples, one collected during CalCOFI cruises in 1998 and 1999, and the second collected during a 2005 cruise that sampled the California Cowcod Conservation Area. This region encompasses Santa Barbara and San Nicolas Islands and the Tanner and Cortez Banks off southern California (http://www.dfg.ca.gov/marine/cowcod.asp).

Atractoscion nobilis (white sea bass) eggs were obtained from the Leon Raymond Hubbard, Jr. Marine Fish Hatchery in Carlsbad, California. Egg samples were obtained from three different spawning events in September and October of 2009 and stored in 95% ethanol.

DNA extraction

Eggs were isolated on depression slides and ethanol was blotted away with tissue. Eggs were then individually transferred to 0.2-mL tubes and soaked in 50 μL dH_2O for 20 min. Water was then removed with a pipette, and a clean pipette tip was used to smash the egg; the tip was left in the tube to ensure that no tissue was lost. Each sample was then digested in 20 μL of a lysis buffer/proteinase K solution [10 mM Tris–HCl (pH 8.3), 50 mM KCl and 0.05% Tween 20, with 0.2 μL of proteinase K (20 mg/mL; Qiagen)]. With the pipette tip that was left in the tube, samples were mixed by back-pipetting. Following brief centrifugation, samples were incubated in a thermal cycler at 65 °C for 60 min followed by 95 °C for 10 min and then stored at −20 °C; 3 μL of the sample was utilized for each PCR.

Gene amplification

Universal COI and 16S primers were used to amplify the target genes. The COI amplicon was 710 bp in length and was obtained using the forward primer COI VF1 (5′-TTCTCAACCAACCACAAAGACATTGG-3′) and the reverse primer COI VR1 [5′-TAGACTTCTGGGTGG CCAAAGAATCA-3′ (Ward *et al.* 2005; Ivanova *et al.* 2007)]. The VF1 primer was biotinylated at the 5′ end to allow fluorescent detection of the gene products during the suspension bead analysis. The mitochondrial 16S ribosomal rRNA gene was amplified using the forward primer 16Sar (5′-CGCCTGTTATCAAAAACAT-3′) and the reverse primer 16Sbr (5′-CCGGTCTGAACTCAGATCACGT-3′), yielding an amplicon of approximately 570 bp (Palumbi 1996). The 16Sbr primer was biotinylated. To preferentially amplify the biotinylated strand of DNA, asymmetric PCR was performed with a 3:1 ratio of primers. The reactions had a final volume of 25 μL and contained 3 μL of the DNA template, 0.2 mM dNTPs, 1.75 total mM $MgCl_2$, 0.8 μM COI VR1 primer, 0.27 μM COI VF1 primer (or 0.4 μM 16Sar primer and 0.13 μM 16Sbr primer), 25 μg BSA, 2.5 U *Taq* DNA polymerase, 1× *Taq* Master, 1× *Taq* Buffer and 150 ng of T4 gene 32 protein (NEB). [*Taq* polymerase, *Taq* Master and *Taq* Buffer are all components of the MasterTaq Kit (Cat. 2200210; 5 Prime, Inc, Gaithersburg, MD, USA).] The thermal cycler profile for both COI and 16S reactions was 95 °C for 2 min, 35 cycles of 95 °C for 30 s, 50 °C for 1 min, and 72 °C for 2 min, and a final extension of 72 °C for 7 min.

After establishing that the COI and 16S primers could individually amplify the target genes, a primer cocktail was then used to co-amplify the COI and 16S genes in the same reaction (again using asymmetric PCR). These multiplex reactions had a final volume of 25 μL and contained 3 μL of DNA template, 0.3 mM dNTPs, 2.25 mM total $MgCl_2$, 0.8 μM COI VR1 primer, 0.27 μM COI VF1 primer, 0.4 μM 16Sar primer, and 0.13 μM 16Sbr primer, 25 μg BSA, 3.75 U *Taq* DNA polymerase, 1× *Taq* Master, 1× *Taq* Buffer and 150 ng of T4 gene 32 protein (NEB); slightly different PCR conditions were determined to be optimal: 95 °C denaturation for 2 min, 40 cycles of a 95 °C denaturation for 30 s, 48 °C annealing for 30 s, and 69 °C extension for 30 s, and a final 69 °C elongation for 9 min.

All PCR products were run on 2% agarose gels and visualized with ethidium bromide to verify successful amplification.

Probe design

We used the mitochondrial cytochrome oxidase subunit I (COI) and 16S ribosomal rRNA (16S) sequences of California fishes publicly available in GenBank to design species-specific probes. There is thorough coverage of California marine fishes in GenBank due in part to an ongoing DNA-barcoding project (marine fishes of California) carried out at the Scripps Institution of Oceanography Marine Vertebrate Collection (resulting sequences are all deposited in GenBank). Species with difficult-to-identify eggs were priority targets. Oligonucleotide probes were designed using Primer BLAST available

through the NCBI website. Twenty-five-base pair potential probes identified by Primer BLAST were then tested using nucleotide BLAST in NCBI to identify any nontarget matches in the GenBank database. All fish species that could potentially complement the probe besides the target species of interest were then examined to determine whether (i) they are found in southern California coastal waters and (ii) they have pelagic eggs. If the species of the nonspecific match inhabit southern California waters and produce pelagic eggs, additional potential probes were tested until a unique probe was found that returned no such nontarget matches. These unique oligonucleotide probes were specifically designed to anneal close to the 5′ end of the target DNA sequence to enhance hybridization. COI probes were designed for all fish species used in this study, and 16S probes were also designed for many species to further corroborate egg identifications.

In some cases, only a single sequence was available for probe development, while ten or more sequences were available for others (Table 1). In all cases where multiple GenBank sequences were available, no evidence for intraspecific variation at the probe site was observed. Although within-species variation could impact the identification process as single-base variation can effect probe hybridization (Diaz & Fell 2004), such variation would typically lead to a failure of the identification (false negative) process rather than an erroneous identification (false positive). Conversely, false positives could occur if other unknown species, including species that do not have sequences represented in GenBank, match our probes. The fish fauna in the region is quite well known and barcoding is nearing completion, so the most likely scenario for a false positive would be from a cryptic invasion of non-native species. Indeed, our probes do match some non-native species that have sequences listed in GenBank (see footnotes to Table 1).

Oligonucleotide probes were synthesized with a 5′ C6 amino linker (Bioneer, Inc) and coupled to differently coloured Luminex xMAP fluorescent beads (See Table 1) following the protocol developed by Mayali *et al.* (2010). We routinely coupled aliquots of ~1.2 million beads (0.1 mL) of each stock type at a time (enough for approximately 3300 egg identifications). Following coupling, beads were washed in 1 mL of 0.1% SDS (sodium dodecyl sulphate), resuspended in 100 μL of 10 mM Tris–EDTA and then stored in the dark at 4 °C.

Hybridization to the array and egg identification

For each hybridization, 0.03 μL of each bead solution (about 300–500 beads total) and 1× tetramethyl ammonium chloride (TMAC) hybridization buffer (5 M TMAC, n-lauryl sarcosine 20%, 1 M Tris–HCl and 500 mM EDTA) were combined to create a bead master mix. Hybridizations were performed in 96-well skirted plates; 10 μL of bead/TMAC mixture (including all 33 probes coupled to their respective beads) was aliquoted into each well followed by 4.5 μL of PCR product (either amplified 16S, COI or both). The species-specific coupled oligonucleotide probes and labelled DNA were incubated at 95 °C for ten minutes to denature the amplified labelled DNA and then hybridized to the probes at 56 °C for two hours. During each hybridization, a negative control was run with each set of samples: 4.5 μL of water was added to the reaction well instead of 4.5 μL of PCR product. After hybridization, 35 μL of 1× TMAC was added to each well of the plate. The plate was then centrifuged at 2164 *g*, 22 °C, for three minutes. After removing the supernatant, 15 μL of a streptavidin-conjugated phycoerythrin (SPE) solution (Invitrogen; 250× dilution in 1× TMAC) was added to each well, and the samples were incubated in the Luminex 100 (Luminex, Austin, TX, USA) for ten minutes at 56 °C. SPE provides the reporter fluorophore, binding to the biotin label on the appropriate PCR primer. Excess SPE was washed from the samples by the addition of 35 μL of 1× TMAC followed by centrifugation and resuspension of the beads in 50 μL of 1× TMAC. The plate was then returned to the Luminex 100 for analysis. The median fluorescence for each of the probes was determined using LUMINEX software (version 1.7), with a minimum or 40 beads of each colour analysed.

Results

DNA extraction and amplification

In the course of these experiments, many different extraction techniques were tested to determine which method would yield the best PCR amplification results. The method presented resulted in greater than 90% success in PCR amplifications of ethanol-preserved eggs from field collections made in 2005. It is important to note that even early embryos with few cells have many copies of mtDNA, making amplification of our two target genes (both encoded in the mtDNA) more reliable than nuclear-encoded genes. In our bead array analysis, the PCR amplification step not only increases the number of target DNA molecules, but it also labels the end of each target DNA strand with biotin (from the labelled primer).

Probe specificity and egg sample identification

Because we do not have eggs of known identity available from a diversity of species, much of our specificity testing relies on DNA extracted from adult voucher specimens from the Scripps Institution of Oceanography Marine Vertebrate Collection. Double blind testing was achieved

Table 1 Fish egg capture probes

Species	GenBank accession	Probe sequence	Bead
Atractoscion nobilis	COI: EU547246.1 (2)	CAAGAGGAAAGAAGGGGGCAGGAGT	63
	16S: AY958646.2 (1)	ACAGACCACGTCAAAGCTCCCTGAT	65
Citharichthys sordidus	COI: EU520655 (>10)	CATCCTGTACCTGCCCCAGCTTCTA	5
	16S: AY952496.2 (1)	CGACCATGGGGAAAACAAATCCCCC*	13
Engraulis mordax	COI: AM911171.1 (>10)	GTTCCAGCTCCTGCTTCAACACCAG	82
Girella nigricans	COI: GU440329.1 (1)	TTGAAGAGACACCGGCGAGGTGAAG	19
Glyptocephalus zachirus	COI: FJ164636.1 (>10)	GGCCAGATTACCAGCGAGGGGAAG	80
Hypsopsetta guttulata	COI: EU522916.1 (2)	CAACCCCAGAAGAGGCCAGAAGGAG	56
	16S: AY958637.2 (2)	ATTTCTCCTCCTCCCACAAGCCAGA	50
Icichthys lockingtoni	COI: EU403053.1 (3)	TATTGGGAAATGGCTGCCGGCTTCA	84
Leuroglossus stilbius	COI: EU400164.1 (2)	CAAGGTTGCTTGCTAGAGGAGGGTA†	25
Lyopsetta exilis	COI: EU522917.1 (>10)	TACAGTTCACCCAGTTCCTGCTCCG	39
Medialuna californiensis	COI: GU440401.1 (1)	GCATGGGCTAAGTTACCAGCGAGAG	90
Merluccius productus	COI: EU489713.1 (>10)	GAATTGAGGAAACGCCTGCTAAGTG‡	33
	16S: AY947850.2 (>10)	ACCTCCATGTGGACGGGGATACTTT‡	86
Microstomus pacificus	COI: EU522918.1 (>10)	GTACACGGTTCACCCAGTACCTGCC	42
Ophidion scrippsae	COI: GU440437.1 (2)	TGGGCCGTGACGATCACGTTATAGA	1
Oxyjulis californica	COI: GU440440.1 (2)	GGGCAAGATTCCCAGATAGAGGGGG	29
Paralichthys californicus	COI: EU520657.1 (2)	CGGTGCCGGCTCCAGCTTCGACACC	35
	16S: AY952499.2 (2)	GGAGTACACCCCTACGTTCCTCTCC	22
Pleuronichthys coenosus	COI: GU936488.1 (3)	AAGATCGTAAGGTCTACGGACGCCC	92
Sardinops sagax	COI: FJ165130.1 (>10)	GAAGCGAAAGCAGGAGGAGAACAGC	88
	16S: GQ412303.1 (>10)	AGTGGATGGGGGACACCCTAAAACC	15
Scomber japonicus	COI: EF607526.1 (7)	TGATACTGGGATGTGGCTGCAGGTT	95
	16S: EU099472.1 (>10)	TACTCCTACAGTCAAGAGCCGCCAC	97
Scopelogadus mizolepis bispinosus	COI: EU489712.1 (1)	CCAGATTGCCTGCAAGAGGGGGATA	72
	16S: AY947847.1 (1)	TAACAAAATGGGGCCCCGCTCAATG	70
Semicossyphus pulcher	COI: EU489705.1 (4)	GCTAACAGGAGGAGGAAGGATGGGG	45
	16S: AY920487.2 (2)	TCTGACCAAATGGATCCGGCAAAGC§	46
Sphyraena argentea	COI: EU752212.1 (3)	TCGTGGAAAGGCCATGTCGGGGGCG	67
	16S: EU099477.1 (1)	ACCCTTCTCTCCCAAAACCAAGGGT	60
Symphurus atricaudus	COI: EU403075.1 (2)	CCGAAGCCTCCAATTAGCACGGGTA	36
Vinciguerria lucetia	COI: HQ010067.1 (1)	CAACACCCGATGATGCCAAGAGGAG	54

'COI' and '16S' designations before each GenBank accession number indicate which mitochondrial gene is being referenced. Numbers in parentheses after each accession number indicate the number of different sequences available in GenBank that were used to design the probe. Twenty-five-base pair-length oligonucleotide sequences are written in 5′ to 3′ orientation. Bead pairings listed refer to different fluorescent bead colours designated by the xMAP system of Luminex. Footnotes denote potential matches to non-native species for the probe as identified by NCBI BLAST search, with the nontarget's geographic home range listed in parentheses. (It is worth noting that only species with sequences available in GenBank were identified as false positives. Species with sequences not yet available in the NCBI database could also cause false-positive identifications, but this needs to be tested empirically.)

*Matches *Etropus crossotus* (Central and South America), *Citharichthys stigmaeus*, and *Citharichthys xanthostigma*. While *C. sordidus*, *C. stigmaeus*, and *C. xanthostigma* are found in the same geographic regions, these species occupy different inshore and offshore habitats. Furthermore, these species can be distinguished using the *C. sordidus* COI probe. This 16S probe is merely for additional identification confidence.

†Matches *Leuroglossus schmidti* (Canada).

‡Matches *Merluccius hubbsi* (Argentina) and *Merluccius gayi* (Chile).

§Matches *Semicossyphus darwini* (Argentina).

by randomly distributing DNA samples of target fish species and deionized water (negative controls) into a 96-well plate. The plate was then analysed, including the PCR and bead array analysis steps, by a naïve investigator; results proved that identification of amplified COI and 16S gene products with the bead array is accurate and consistent. All 33 probes utilized in this study showed strong specificity to the target sequence and low cross-reactivity with other nontarget DNA sequences.

This was further tested as follows: COI and 16rRNA genes were amplified from DNA samples from 50 additional nontarget California species (Table S1, Supporting information) and analysed individually with the bead

array; each of the 33 probes gave a clear signal only to its single target species. No false positives were observed.

The criterion for identification was a signal-to-noise ratio (defined as the fluorescent signal of the target probe divided by the fluorescent signal of the negative control) of greater than five. Eggs were identified with both uniplex PCR using only the COI or 16S probe or using both COI and 16S probes following multiplex PCR to improve identification success when one gene fails to amplify.

As discussed earlier, there is the possibility that intraspecific variation could interfere with species-specific probe hybridization and lead to false-negative identifications. Although levels of sequence variation within species in the region used for probe development are rather low, in our iterative process of probe development (please see Discussion), we have come across one instance in which a *Leuroglossus stilbius* target sequence had one mismatch with the intended probe (mismatch at nucleotide position number 6 of 25). While the signal for this sample was reduced 50% compared to *L. stilbius* samples whose sequences perfectly matched the probe, the signal-to-noise ratio was still >5 and the sample was readily identified.

This study focused on developing probes for species that have morphologically identical eggs. For instance, the eggs of *Scomber japonicus* (Pacific mackerel) and *Merluccius productus* (hake) cannot be reliably differentiated until late in embryonic development, *Sphyraena argentea* (Pacific barracuda) and *Atractoscion nobilis* (white seabass) have similarly indistinguishable eggs, and eggs of the flatfishes *Paralichthys californicus* (California halibut), *Citharichthys sordidus* (Pacific sanddab) and *Hypsopsetta guttulata* (diamond turbot) are often confused. For these species, both COI and 16S probes were designed when possible to improve identification confidence. Identifications were performed using both probes simultaneously (Fig. 1a, b).

For species that are abundant in CalCOFI samples but are morphologically distinct, such as *Engraulis mordax* and *Sardinops sagax*, only one probe was needed for identification (Fig. 1c, d).

For several species, including *Sardinops sagax*, we found that the 16S gene amplified much more consistently than

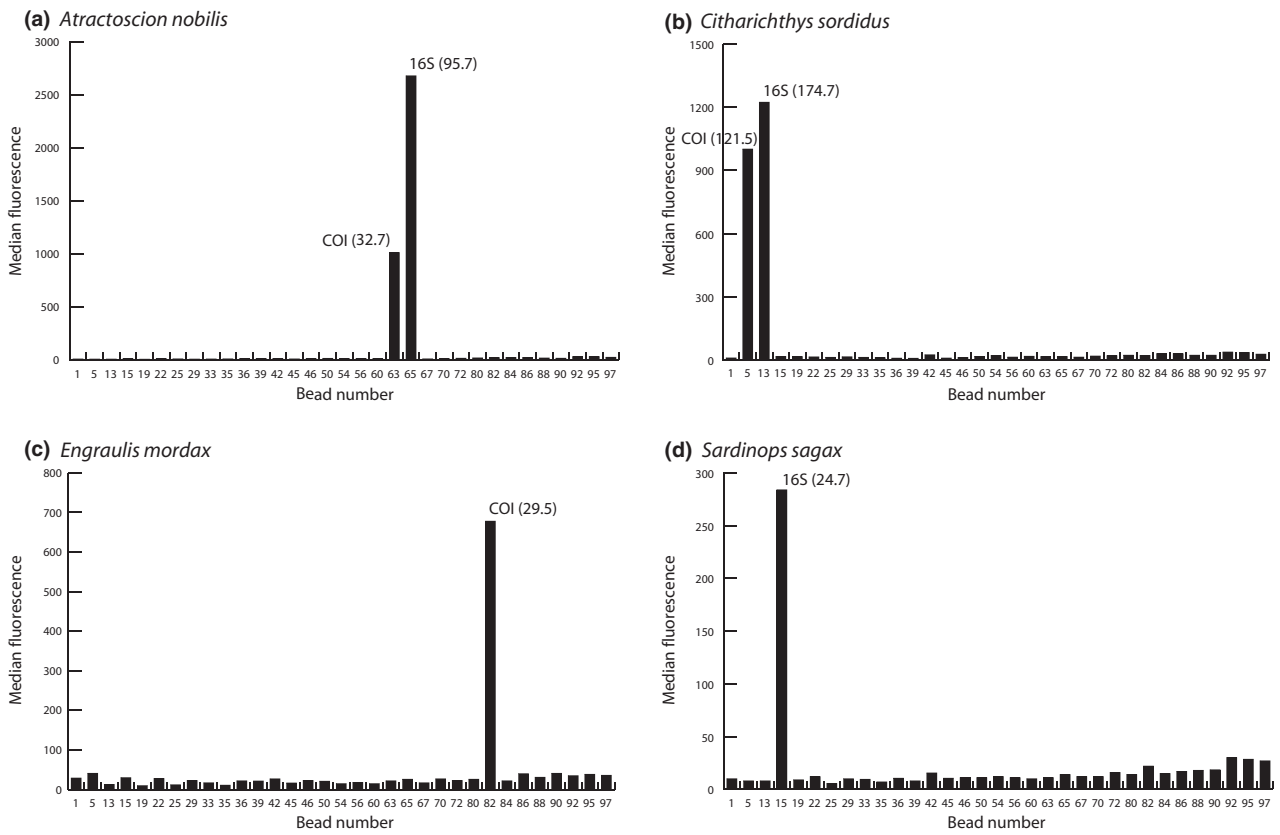

Fig. 1 Graphs depicting probe median fluorescence values for fish eggs of the following species: (a) *Atractoscion nobilis*, (b) *Citharichthys sordidus*, (c) *Engraulis mordax*, and (d) *Sardinops sagax*. Each bar depicts the median fluorescence value for one of the 33 probes. To make identification more robust, both COI and 16S probes were used for species such as *A. nobilis* and *C. sordidus* with difficult-to-identify eggs. For species that have common and morphologically distinguishable eggs such as *E. mordax* and *S. sagax*, only one probe was used for identification. Signal-to-noise ratios are shown in parentheses next to the probe type.

COI. Therefore, 16S probes were used to identify these species (Fig. 1d).

Discussion

Suspension bead arrays present a promising system for performing high-throughput molecular identification of fish eggs. The Luminex 100 system is composed of a flow cytometry-based instrument and xMAP-carboxylated beads to which custom-made oligonucleotide probes can be coupled. The xMAP microspheres are available in 100 different colours, allowing the utilization of up to 100 different probes in a single reaction. Existing data indicate that there are approximately 100 species of California marine fish that produce pelagic eggs (Ahlstrom & Moser 1980), so, in principle, samples can be rapidly identified by simultaneously testing them against all one hundred species-specific probes. Where fish fauna are more diverse, for instance in tropical regions such as the Great Barrier Reef and Moorea, multiple sets of probes could be used to test each egg, or systems capable of analysing up to 500 probes such as the Luminex FLEXMAP 3D could be employed.

We have developed a DNA extraction and PCR protocol that resulted in an amplification success rate of approximately 90% using field-collected eggs that had been stored in ethanol for 5–10 years. Although the duplex PCR (COI and 16S) typically produces two amplicons, we did observe that some species produced stronger results for one of the two target genes and we accordingly target that gene for probe development. The use of mitochondrial DNA target genes assures that even early-stage eggs will have many copies to the target; even prior to fertilization, fish eggs typically contain over 10^6 copies of mtDNA (Gangfeng & Shaoy 1992).

Although all of the probes ultimately utilized for this study were specific, several probes that we initially tested showed cross reactivity with nontarget species and thus had to be redesigned. These nonspecific probes were subjected to the same rigorous protocols in the probe design process as all other probes, but for unknown reasons, when these probes were utilized in the actual suspension bead arrays, they showed reactivity with species other than the target. However, in all instances, after carefully redesigning the probes, species-specific probes were obtained. This empirical probe validation is most readily accomplished when performed in conjunction with DNA barcoding projects where DNA extracted from voucher adult specimens is available for most members of the regional fauna. As mentioned previously, DNA barcoding sequences of fish species worldwide are becoming increasingly available, and thus probe validation could be readily performed for species from any number of geographic locales, such as other regions of North America, South America and Europe.

Aside from the Luminex 100 instrument itself, the methods described here require only standard laboratory equipment, including thermal cyclers (for PCR and incubations) and a centrifuge with a rotor for 96-well plates (for washes). The Luminex analysis takes approximately 45 min to analyse one 96-well plate, so if DNA extraction and PCR amplification have already been performed, six to eight plates, or 560–750 eggs (plus negative and positive controls) could be identified in one working day. If two or three thermal cyclers are available, placing individual eggs into plates for processing is likely the rate-limiting step.

When taking into account reagents utilized in DNA extraction, PCR amplification, hybridization and Luminex analysis, the estimated cost for identifying one egg using a 100-probe array is approximately $1.00, with the biggest post-PCR expense being the microsphere beads (~$0.60 per egg). It is important to note that this cost estimate is scalable based on the number of probes employed; if a 50-probe array satisfies the intended purpose, the cost of beads would be cut in half. Although it was not attempted here, further savings could be achieved by using smaller PCR reactions (and reducing reagent costs) and possibly by using fewer beads of each colour per identification; at our current usage of ~400 beads of each colour per well, single vials of beads last for over 30 000 individual identifications.

Several approaches are available for molecular identification of fish eggs. Fluorescent in situ hybridization (FISH) methods, recently coupled with large bore flow cytometry, offers a no-PCR approach but lacks potential for high multiplexing, making it potentially attractive in situations where the goal is enumeration of only a few target species in unsorted plankton samples (e.g. Henzler *et al.* 2010). Where discrimination of more target species is required, one valuable approach uses multiplex PCR with species-specific primers designed to produce a different amplicon size for each target species. Hyde *et al.* (2005) successfully used multiplex PCR on board a ship to identify billfish larvae in near real time. This is a simple, fast and low-cost approach; however, it is best suited to identifications of 5–20 species at a time and cannot reach the level of multiplexing available with bead arrays.

When discrimination among a large number of species is required, DNA sequencing of each individual egg provides a straightforward, if somewhat costly, approach (Shao *et al.* 2002; Richardson *et al.* 2007; Kawakami *et al.* 2010). Richardson *et al.* (2007) present a 'high-throughput' solution for sequencing using a liquid-handling robot to isolate the DNA from their samples and perform the sequencing reactions. Although this approach

reduces the amount of manual pipetting required, reagent costs per sample (not discussed in the paper) are high (we estimate that extraction, PCR, cleanup steps and sequencing reactions would likely total at least $4 per egg). Our estimated costs per egg sample (~$1.00 for the 100-probe array) are significantly lower, and our estimated throughput of ~600+ identifications per day compares favourably to Richardson *et al.*'s estimate of 800+ identifications per week. The major limitation of the bead array approach when compared to sequencing is that an array can only identify an egg if a probe to that species is included on the array; no such restrictions apply to DNA sequencing identification. However, where potential target species are well defined (as in the case with ichthyoplankton of the California Current, the island of Moorea, additional regions of North America and areas of Australia, Europe and Antarctica), the bead array approach appears to hold a significant cost and throughput advantage.

We are proceeding with further development of the probe set for coastal California by iteratively processing egg samples from research cruises (samples provided by Southwest Fisheries Science Center Ichthyoplankton Lab). When an individual egg sample is not identified by the existing 33-probe array, the PCR product is sequenced and identified by GenBank BLAST search (approximately 600 California marine species are now in the database, R. S. Burton and P. A. Hastings, Scripps Institution of Oceanography, unpublished data). The sequence information obtained is then used to design additional species-specific probes.

As a first test on field collected eggs, the current set of 33 probes was used to analyse samples of eggs obtained from the 2005 SWFSC Cowcod Conservation Area ichthyoplankton survey. Of the 151 eggs tested, 14 (<10%) failed to produce PCR products and could not be identified. Using the bead array, we identified 116 of the 137 eggs that yielded PCR products: *Merluccius productus* (45 eggs), *Leuroglossus stilbius* (38 eggs), *Citharichthys sordidus* (18 eggs), *Atractoscion nobilis* (9 eggs), *Engraulis mordax* (2 eggs), *Sardinops sagax* (1 egg), *Lyopsetta exilis* (2 egg) and *Paralichthys californicus* (1 egg). The remaining 21 eggs were identified by sequencing and GenBank BLAST search as described earlier. These 21 eggs were found to represent four species not yet included on the array (and eight squid eggs that had been incorrectly sorted from the plankton tow): *Nansenia candida* (8 eggs), *Trachipterus trachipterus* (1 egg), *Embassichthys bathybius* (1 egg) and *Icostus aenigmaticus* (3 eggs). Probes to these four species can now be added to the array to increase the proportion of eggs identified in future analyses.

Finally, in this study, we have focused on identification of pelagic fish eggs in the California Current. However, this identification method can be utilized in any region where some pre-existing knowledge regarding fish species assemblages and their DNA sequences are present. Many recently published papers provide such DNA sequences for fish fauna from areas as diverse as North America, Australia, Cuba, the Amazon, Africa, Italy, Israel and India (Ward *et al.* 2005; Swartz *et al.* 2008; Rasmussen *et al.* 2009; Shirak *et al.* 2009; Ardura *et al.* 2010; Filonzi *et al.* 2010; Lara *et al.* 2010; Lakra *et al.* 2011). More broadly, worldwide comprehensive efforts to barcode all fish species (Ward *et al.* 2009) have resulted in the COI DNA barcode sequences of over 8000 species, which are available in databases such as NCBI's GenBank and the barcode of life data system (BOLD, Ratnasingham & Hebert 2007). Thus, in any number of geographic regions, species-specific probes for local fish fauna could easily be designed from the available barcoding sequences. Furthermore, once these oligonucleotide capture probes have been designed, bead arrays can also be used to identify any other type of fish tissue in addition to eggs, including fish larvae, fillets at fish markets or gut contents of piscivores. Consequently, the development of a cost-effective molecular identification method has important implications for both ecological research and fisheries management worldwide. Assessing the impact of marine reserves on ichthyoplankton (Watson *et al.* 2002) or the effects of climate change on regional fish faunas (e.g. Vilchis *et al.* 2009) could greatly benefit from the application of bead arrays as a species identification tool.

Acknowledgements

The Ichthyoplankton Lab of the SWFSC and the Raymond Hubbard, Jr. Marine Fish Hatchery Fish Hatchery kindly provided fish eggs for testing. Asako Yamamoto, Cindy Hon and Aryana Sepassi contributed to data collection. William Watson, Carolina Bonin and Kathleen Fisch provided valuable comments on the manuscript. This work was supported under NOAA Grant #NA10OAR4170060, California Sea Grant Project #R/Fish 207, through NOAA's National Sea Grant College Program, U.S. Department of Commerce. The statements, findings, conclusions and recommendations are those of the authors and do not necessarily reflect the views of the aforementioned organizations.

References

Ahlstrom EH, Moser HG (1980) Characters useful in identification of pelagic marine fish eggs. *California Cooperative Oceanic Fisheries Investigations Report*, **21**, 121–131.

Alheit J (1993) Use of the daily egg production method for estimating biomass of clupeoid fishes: a review and evaluation. *Bulletin of Marine Science*, **53**, 750–767.

Ardura A, Linde AR, Moreira JC, Garcia-Vazquez E (2010) DNA barcoding for conservation and management of Amazonian commercial fish. *Biological Conservation*, **143**, 1438–1443.

Barry JP, Baxter CH, Sagarin RD, Gilman SE (1995) Climate-related, long-term faunal changes in a California rocky intertidal community. *Science*, **267**, 672–675.

Bellier E, Planque B, Petitgas P (2007) Historical fluctuations in spawning location of anchovy (*Engraulis encrasicolus*) and sardine (*Sardina pilchardus*) in the Bay of Biscay during 1967–73 and 2000–2004. *Fisheries Oceanography*, **16**, 1–15.

Brander K (2010) Impacts of climate change on fisheries. *Journal of Marine Systems*, **79**, 389–402.

Carreon-Martinez LB, Holt SA, Nunez BS, Faulk CK, Holt GJ (2010) The use of polymerase chain reaction for the identification of sciaenid eggs. *Marine Biology*, **157**, 1889–1895.

Deak E, Etienne KA, Lockhart SR, Gade L, Chiller T, Balajee SA (2010) Utility of a Luminex-based assay for multiplexed, rapid species identification of *Candida* isolates from an ongoing candidemia surveillance. *Canadian Journal of Microbiology*, **56**, 348–351.

Diaz MR, Fell JW (2004) High-throughput detection of pathogenic yeasts of the genus *Trichosporon*. *Journal of Clinical Microbiology*, **42**, 3696–3706.

Durant JM, Dag OH, Otterson G, Stenseth NC (2007) Climate and the match or mismatch between predator requirements and resource availability. *Climate Research*, **33**, 271–283.

Ecosystem Principles Advisory Panel (1999) *Ecosystem-Based Fishery Management*. National Oceanic and Atmospheric Administration National Marine Fisheries Service, Silver Spring, Maryland.

Filonzi L, Chlesa S, Vaghi M, Marzano FN (2010) Molecular barcoding reveals mislabeling of commercial fish products in Italy. *Food Research International*, **5**, 1383–1388.

Fodrie FJ, Heck KL, Powers SP, Graham WM, Robinson KL (2010) Climate-related, decadal-scale assemblage changes of seagrass-associated fishes in the northern Gulf of Mexico. *Global Change Biology*, **16**, 48–59.

Fox CJ, Taylor MI, Pereya R, Villasana MI, Rico C (2005) TaqMan DNA technology confirms likely overestimation of cod (*Gadus morhua* L.) egg abundance in the Irish Sea: implications for the assessment of the cod stock and mapping of spawning areas using egg-based methods. *Molecular Ecology*, **14**, 879–884.

Gangfeng W, Shaoy Y (1992) Mitochondrial DNA content and mitochondrial transcriptional activities in the early development of loach and goldfish. *International Journal of Developmental Biology*, **36**, 477–482.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B*, **270**, 313–321.

Henzler CM, Hoagland EA, Gaines SD (2010) FISH-CS-a rapid method for counting and sorting species of marine zooplankton. *Marine Ecology Progress Series*, **410**, 1–11.

Hernandez FJ, Powers SP, Graham WM (2010a) Seasonal variability in ichthyoplankton abundance and assemblage composition in the northern Gulf of Mexico off Alabama. *Fishery Bulletin*, **108**, 193–207.

Hernandez FJ, Powers SP, Graham WM (2010b) Detailed examination of ichthyoplankton seasonality from a high-resolution time series in the northern Gulf of Mexico during 2004–2006. *Transactions of the American Fisheries Society*, **139**, 1511–1525.

Hsieh CH, Kim HJ, Watson W, DiLorenzo E, Sugihara G (2009) Climate – driven changes in abundance and distribution of larvae of oceanic fishes in the southern California region. *Global Change Biology*, **15**, 2137–2152.

Hyde JR, Lynn E, Humphreys R, Musyl M, West AP, Vetter R (2005) Shipboard identification of fish eggs and larvae by multiplex PCR, and description of fertilized eggs of blue marlin, shortbill spearfish, and wahoo. *Marine Ecology Progress Series*, **286**, 269–277.

Ivanova NV, Zemlak TS, Hanner RH, Hebert PD (2007) Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes*, **7**, 544–548.

Jennings S, Brander K (2010) Predicting the effects of climate change on marine communities and the consequences for fisheries. *Journal of Marine Systems*, **79**, 418–426.

Karaiskou N, Triantafyllidis A, Alvarez P, Lopes P, Garcia-Vazquez E, Triantaphyllidis C (2007) Horse mackerel egg identification using DNA methodology. *Marine Ecology*, **28**, 429–434.

Kawakami T, Aoyama J, Tsukamoto K (2010) Morphology of pelagic fish eggs identified using mitochondrial DNA and their distribution in waters west of the Mariana Islands. *Environmental Biology of Fishes*, **87**, 221–235.

Lakra WS, Verma MS, Goswami M *et al*. (2011) DNA barcoding Indian marine fishes. *Molecular Ecology Resources*, **11**, 60–71.

Lara A, de Leon JLP, Rodriguez R *et al*. (2010) DNA barcoding of Cuban freshwater fishes: evidence for cryptic species and taxonomic conflicts. *Molecular Ecology Resources*, **10**, 421–430.

Lasker R (1985) An egg production method for estimating spawning biomass of pelagic fish: application to the northern anchovy, *Engraulis mordax*. *National Oceanic and Atmospheric Administration Technical Report, National Marine Fisheries Service*, **36**, 99.

Lelièvre S, Verrez-Bagnis V, Jérôme M, Vaz S (2010) PCR-RFLP analyses of formalin-fixed fish eggs for the mapping of spawning areas in the Eastern Channel and southern North Sea. *Journal of Plankton Research*, **32**, 1527–1539.

Mayali X, Palenik B, Burton R (2010) Dynamics of marine bacterial and phytoplankton populations using multiplex liquid bead array technology. *Environmental Microbiology*, **12**, 975–989.

Murawski SA (1993) Climate-change and marine fish distributions—forecasting from historical analogy. *Transactions of the American Fisheries Society*, **122**, 647–658.

Occhipinti-Ambrogi A (2007) Global change and marine communities: alien species and climate change. *Marine Pollution Bulletin*, **55**, 342–352.

Palumbi SR (1996) Nucleic acids II: the polymerase chain reaction. In: *Molecular Systematics* (eds. Hillis DM, Moritz C, Mable BK). Sinauer & Associates Inc., Sunderland, Massachusetts.

Parker K (1980) A direct method for estimating northern anchovy, *Engraulis mordax*, spawning biomass. *Fishery Bulletin*, **78**, 541–544.

Perez J, Álvarez P, Martinez J, Garcia-Vazquez E (2005) Genetic identification of hake and megrim eggs in formaldehyde-fixed plankton samples. *ICES Journal of Marine Science*, **62**, 908–914.

Perry AL, Low PJ, Ellis JE, Reynolds JD (2005) Climate change and distribution shifts in marine fishes. *Science*, **308**, 1912–1915.

Priede IF, Watson JJ (1993) An evaluation of the Daily Egg Production Method for estimating biomass of Atlantic mackerel (*Scomber scombrus)*. *Bulletin of Marine Science*, **53**, 891–911.

Rasmussen R, Morrissey MT, Hebert PDN (2009) DNA barcoding of commercially important salmon and trout species (*Oncorhynchus* and *Salmon)* from North America. *Journal of Agricultural and Food Chemistry*, **57**, 8379–8385.

Ratnasingham S, Hebert PDN (2007) BOLD: the barcode of life data system http://www.barcodinglife.org. *Molecular Ecology Notes*, **7**, 355–364.

Richardson DE, Vanwye JD, Exum AM, Cowen RK, Crawford DL (2007) High-throughput species identification: from DNA isolation to bioinformatics. *Molecular Ecology Notes*, **7**, 199–207.

Shao KT, Chen KC, Wu JH (2002) Identification of marine fish eggs in Taiwan using light microscopy, scanning electric microscopy and mtDNA sequencing. *Marine and Freshwater Resources*, **53**, 355–365.

Shirak A, Cohen-Zinder M, Barroso RM, Seroussi E, Ron M, Hulata G (2009) DNA barcoding of Israeli indigenous and introduced cichlids. *Israeli Journal of Aquaculture-Bamidgeh*, **61**, 83–88.

Swartz ER, Mwale M, Hanner R (2008) A role for barcoding in the study of African fish diversity and conservation. *South African Journal of Science*, **104**, 293–298.

Vilchis LI, Ballance LT, Watson W (2009) Temporal variability of neustonic ichthyoplankton assemblages of the eastern Pacific warm pool: can community structure be linked to climate variability? *Deep Sea Research I*, **56**, 125–140.

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, **360**, 1847–1857.

Ward RD, Hanner R, Hebert PDN (2009) The campaign to DNA barcode all fishes, FISH-BOL. *Journal of Fisheries Biology*, **74**, 329–356.

Watson W, Charter RL, Moser HG *et al*. (2002) Distributions of planktonic fish eggs and larvae off two state ecological reserves in the Santa Barbara Channel vicinity and two nearby islands in the Channel Islands National Marine Sanctuary. *California Cooperative Oceanic Fisheries Investigations Report*, **43**, 141–154.

Yancy HF, Zemlak TS, Mason JA *et al.* (2008) Potential use of DNA barcodes in regulatory science: applications of the Regulatory Fish Encyclopedia. *Journal of Food Protection*, **71**, 210–217.

Data Accessibility

16S and COI DNA sequences: GenBank accessions EU547246.1, AY958646.2, HQ010049.1, EU520655, AY952496.2, AM911171.1, GU440329.1, FJ164636.1, EU522916.1, AY958637.2, EU403053.1, EU400164.1, EU522917.1, GU440401.1, EU489713.1, AY947850.2, EU522918.1, GQ891091.1, GU440437.1, GU440440.1, EU520657.1, AY952499.2, GU936488.1, FJ165130.1, GQ412303.1, EF607526.1, EU099472.1, EU489712.1, AY947847.1, EU489705.1, AY920487.2, EU752212.1, EU0994770.1, EU403075.1, GU440558.1, HQ010067.1

All tested probe sequences: DRYAD entry doi:10.5061/dryad.8p425.

Supporting Information

Additional supporting information may be found in the online version of this article.

Table S1 California fish species used for probe specificity testing.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.