Molecular Phylogenetics and Evolution 55 (2010) 1195–1203

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Short Communication

Identifying coral reef fish larvae through DNA barcoding: A test case with the families Acanthuridae and Holocentridae

Nicolas Hubert [a,*], Erwan Delrieu-Trottin [b], Jean-Olivier Irisson [b], Christopher Meyer [c], Serge Planes [b,d]

[a] Laboratoire ECOMAR, Université de la Réunion, 15 Avenue René Cassin, BP 7151, 97715 Saint-Denis Messag Cedex 9, Réunion, France
[b] Équipe récifs coralliens, UMR 5244 CNRS-UPVD-EPHE, Centre de Biologie et Écologie Tropicale et Méditéranéenne, Université de Perpignan Via Domitia, 52 Avenue Paul Alduy, 66860 Perpignan Cedex, France
[c] Department of Invertebrate Zoology, National Museum of Natural History, Smithsonian Institution, P.O. Box 37012, MRC-163, Washington, DC 20013, USA
[d] Centre de Recherches Insulaires et Observatoire de l'Environnement (CRIOBE) UMS 2978 EPHE-CNRS, BP 1013 Papetoai, 98729 Moorea, Polynésie française, France

ARTICLE INFO

Article history:
Received 8 October 2009
Accepted 22 February 2010
Available online 25 February 2010

Keywords:
Coral reef
Barcoding
Larvae
Identification
Mitochondrial DNA

ABSTRACT

A reference collection of COI barcode (650 bp) for the Pacific Society Islands has been constituted for 22 species of Acanthuridae and 16 species of Holocentridae. Divergence between congeneric species was on average 20-fold to 87-fold higher than divergence between conspecific sequences and this set of DNA-identifiers was used to identify 40 larvae of both families. All larvae sequenced could be identified to species using DNA-barcodes. Pools of larvae constitute multi-specific assemblages and no additional species compared to adult reef communities were sampled in larval pools, suggesting that the larval assemblages originated from adult communities on neighboring reefs.

1. Introduction

DNA barcoding seeks to develop automated DNA-based identifications using molecular species tags based on short, standardized gene regions (Hebert et al., 2003; Hebert and Gregory, 2005). The primary goal of DNA barcoding is to create reference DNA-barcode libraries for known species used as DNA-identifiers (e.g. Kerr et al., 2007; Hubert et al., 2008). Mitochondrial DNA (mtDNA) has been widely used in evolutionary studies owing to its higher mutation rate and lower effective population size than nuclear DNA (Brown et al., 1979; Birky et al., 1989), and efforts have converged on a 650-bp portion of the mitochondrial cytochrome *c* oxidase I gene (COI) that can be readily recovered from a vast array of lineages with a limited set of primers. For a barcoding approach to succeed, within species DNA sequences need to be more similar to one another than those between species and recent studies confirmed that the majority of species examined are well delineated by a tight cluster of very similar sequences (Ward et al., 2005; Clare et al., 2006; Robins et al., 2007; Kerr et al., 2007; Hubert et al., 2008; Foottit et al., 2009; Sheffield et al., 2009). Nevertheless, some pitfalls have been identified due to the presence of pseudogenes, introgressive hybridization, and retention of ancestral polymorphism (Zhang and Hewitt, 1996; Funk and Omland, 2003; Meyer and Paulay, 2005). The occurrence of mixed genealogies among closely related species were estimated to reach 20% (Funk and Omland, 2003), although recent barcoding surveys suggest that it may not exceed 5–10% (Kerr et al., 2007; Hubert et al., 2008).

Coral reefs are among the most diverse ecosystems and the Indo-Pacific region alone hosts 10,490 fish species, nearly 32 percent of Earth's ichthyofauna (Froese and Pauly, 2000). In ecosystems with no obvious physical barriers, assessing the determinants of connectivity is a priority for conservation practices (Mora et al., 2006; Claudet et al., 2008). In marine systems, connectivity is widely assessed through analysis of gene flow (e.g. Doherty et al., 1995; Jones et al., 1999; Almany et al., 2007). However, community level processes such as competitive exclusion, assortative settlement and habitat selection may strongly influence species distribution and thereby, communities connectivity (Loreau and Mouquet, 1999; Mouquet and Loreau, 2002; Webb et al., 2002; Leibold et al., 2004). Given their high diversity and dramatic phenotypic changes during development, coral reef fish species identification is no easy task and only feasible up to the genera at best for early ontogenetic stages based on diagnostic morphological characters (e.g. Leis and Carson-Ewart, 2004). Species interactions, however, are likely to vary largely depending on ontogenetic stages through which interactions occur (e.g. Webb, 2000).

Here, we explore the efficacy of the barcoding approach in the identification of coral reef fish larvae to the species level in order to address the following questions. First, larvae are aggregated in

* Corresponding author. Fax: +33 02 62 93 87 65.
E-mail address: myloplus@excite.com (N. Hubert).

1055-7903/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2010.02.023

patches and schools (Doherty, 1987) and are often collected in pools of several phylogenetically-related individuals, e.g. several individuals from one family or one genus. Do these pools host multi-specific assemblages or correspond to single-species schools? Second, without additional knowledge, larvae collected are assumed to come from the neighboring reefs. With a more precise species-level identification, one can ask whether larvae samples contain species that are not present in adult communities? In such context, we first assessed the genetic variability at COI for two of the most abundant coral reef fish families, namely Acanthuridae and Holocentridae, and further explored its use in a species-level tagging for the identification of early ontogenetic stages up to the species level.

2. Materials and methods

2.1. Sampling reef fishes larvae in the pelagic realm

Fish larvae were sampled during an oceanographic campaign aboard the N.O. Alis in May 2006 all around the atoll of Tetiaroa in the Society archipelago (17°S, 149°55W). Owing to its relative isolation by surface currents, the atoll of Tetiaroa is likely to be an autonomous system where self-recruitment sustains most of the local populations. In order to describe the spatial distribution of larvae, thirteen stations laid in a radiating pattern around the atoll were sampled twice. Samples were obtained by trawling with a Multiple Opening-Closing Net and Environmental Sampling System (MOCNESS) with an 800 μm mesh and 4 m^2 opening. Trawling was conducted between the surface and 100 m where the majority of coral reef fish larvae are found (Boehlert et al., 1992). Sampling was stratified vertically: net M0 sampled from 0 down to 100 m and nets M1 to M4 sampled back up to the surface, in 25 m layers. M0 samples were preserved in 90% ethanol to allow the use of DNA-based identification. M1–M4 samples were preserved in a 4% buffered formaldehyde-sea water solution, which preserves pigmentation, for morphological identification and further oceanographic analysis.

In the M0 net, larvae were identified to the lowest taxonomic level possible using a stereo-microscope to look for morphological criteria described in previous systematic surveys (e.g. Moser, 1996; Leis and Carson-Ewart, 2004; Miller and Tsukamoto, 2004) such as general shape, fin rays and spines count, head spine location, and pigmentation. Equivocal identifications were confirmed by using an online photography database developed for this purpose (http://cbetm.univ-perp.fr/larvae). Over all dataset, morphological identification discriminated 82 families, with a number of larvae ranking from 4 to 231 individuals per station. The families Acanthuridae and Holocentridae were the most abundant in the formaldehyde-preserved samples, particularly in stations 9 and 10 during the first round of sampling since 87 out of the 158 Acanthuridae and 78 out of the 120 Holocentridae were sampled there (Irisson, 2008). Likewise, Acanthuridae and Holocentridae were the most abundant families in ethanol-preserved samples and stations 9 and 10 of the first sampling round also contained high concentrations of larvae, with 15 Acanthuridae and 20 Holocentridae at stations 9 and 11 Acanthuridae at station 10. Finally, a total of 46 larvae (sum of Holocentridae and Acanthuridae of stations 9 and 10) were used for DNA-based identifications.

2.2. Adult sampling, reference library and barcoding

Fish adults used as DNA-identifiers for larvae in the reference library were sampled apart in the context of the Moorea biocode project and FISH-BOL campaign (Ward et al., 2009) in 2006 all around the atoll of Moorea in the Society Archipelago. The reference library of adult sequences was built from ethanol-preserved fin clips on specimens identified by experts using morphological criteria (Randall, 2005). For each species, reference specimens were deposited as vouchers in publicly available collections, namely the 'Muséum National d'Histoire Naturelle' (MNHN) in Paris and the Moorea Biocode collection (MBIO) in Berkeley. In order to fit with the BARCODE criteria in GenBank (Hubert et al., 2008), sequences were bi-directionally sequenced for at least 500 bp and electropherogram trace files were made accessible in the NCBI Trace Archive as well as forward and reverse PCR amplification primers. In addition, a link with the Barcode of Life Data System (BOLD) has been created to provide access to detailed voucher data including collection record and photographs. These data are also publicly available in the Moorea Biocode databases (http://biocode.berkeley.edu) (Table 1).

Genomic DNA was extracted using the Gentra System Puregene DNA Purification Kit according to manufacturer specifications. A 650-bp segment was amplified from the 5′ region of the mitochondrial COI gene using the primers FF2d-5′TTCTCCACCAACCAC AARGAYATYGG3′ and FR1d-5′CACCTCAGGGTGTCCGAARAAYCARA A3′ (Ivanova et al., 2007). PCR amplifications were performed in 27 μl including 10.7 μl of ultrapure water, 2.5 μl of 10× PCR buffer, 3 μl of $MgCl_2$ (25 mM), 2.5 μl of each primers (10 mM), 3 μl of each dNTP (2 mM), 0.3 μl of Taq DNA polymerase (5U/μl), and 4 μl of template DNA. The PCR conditions consisted of 94 °C (5 min), 10 cycles of 94 °C (1 min), 60–50 °C decreasing 1 °C per cycle (1 min), 72 °C (1 min 30 s) followed by 25 cycles of 94 °C (1 min), 50 °C (1 min), 72 °C (1 min 30 s), with a final extension at 72 °C (5 min).

All sequences have been deposited in GenBank (Accession Numbers XX–XX). Accession Numbers for the barcodes, specimen and collection data, sequences, trace files, and primers details are available within the project "in progress" in BOLD (http://www.barcodinglife.org). Sequence divergence was calculated using the JC69 model (Jukes and Cantor, 1969) and ultrametric trees were computed using the UPGMA algorithm as implemented in PAUP* 4.0b10 (Swofford, 2002) to provide a graphic representation of species divergence. Finally, several metrics were computed from the pairwise distance matrice using the package APE for R (Paradis et al., 2004), namely the mean, minimum and maximum of the distance within species ($D_{\text{within species}}$), the distance to the nearest neighbor (D_{NN}) and the distance between species ($D_{\text{between species}}$).

3. Results and discussion

All adult specimens were successfully amplified using the primers FF2d and FR1d. Four genera and 22 species were discriminated among the 53 adults of Acanthuridae. Another four genera and 16 species were characterized from 53 adults of Holocentridae (Table 1). Those 106 sequences constituted the reference library of DNA-identifiers used to assign larvae to known species. Among the 46 larvae collected in this study, only three Acanthuridae and three Holocentridae failed to amplify using FF2d and FR1d. The 40 remaining larvae provided COI consensus established through bi-directional sequencing. A total of 146 COI barcodes of 650-bp were thus obtained for 38 species and eight genera. No insertions, deletions or codon stops were found, supporting that all amplified sequences constitute functional mitochondrial COI. All the amplified sequences were larger than 600-bp, the maximum size limit observed for non-functional nuclear copies of mtDNA genes (Zhang and Hewitt, 1996).

The distribution of pairwise differences among COI barcodes of adults revealed little to no overlap in the distribution of divergence within and between species (Fig. 1). The majority of COI sequences were identical within species, while reaching up to four differences