Molecular Phylogenetics and Evolution 100 (2016) 95–108

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Clupeiformes' Egg Envelope Proteins characterization: The case of *Engraulis encrasicolus* as a proxy for stock assessment through a novel molecular tool

Andrea Miccoli [a,b], Iole Leonori [b], Andone Estonba [c], Andrea De Felice [b], Chiara Carla Piccinetti [a], Oliana Carnevali [a,*]

[a] Laboratory of Developmental and Reproductive Biology, DiSVA, Università Politecnica delle Marche, Ancona, Italy
[b] CNR-National Research Council of Italy, ISMAR-Marine Sciences Institute, Ancona, Italy
[c] Department of Genetics, Physical Anthropology and Animal Physiology, University of the Basque Country, UPV/EHU, Leioa, Spain

ARTICLE INFO

Article history:
Received 23 February 2016
Revised 1 April 2016
Accepted 5 April 2016
Available online 7 April 2016

Keywords:
Engraulis encrasicolus
Zona radiata proteins
ZP domain
Clupeiformes
Stock management

ABSTRACT

Zona radiata proteins are essential for ensuring bactericidal resistance, oocyte nutrients uptake and functional buoyancy, sperm binding and guidance to the micropyle, and protection to the growing oocyte or embryo from the physical environment.

Such glycoproteins have been characterized in terms of molecular structure, protein composition and phylogenetics in several chordate models. Nevertheless, research on teleost has not been extensive. In Clupeiformes, one of the most biologically relevant and commercially important order which accounts for over 400 species and totally contributes to more than a quarter of the world fish catch, Egg Envelope Protein (EEP) information exist only for the *Clupea pallasii* and *Engraulis japonicus* species. The European anchovy, *Engraulis encrasicolus*, the target of a well-consolidated fishery in the Mediterranean Sea, has been ignored until now and the interest on the Otocephala superorder has been fragmentally limited to some Cypriniformes and Gonorynchiformes, as well.

The aim of the present study was to fill the ZP protein-wise gap of knowledge afflicting the understanding of the European anchovy's reproductive process and to expand the background on Clupeiformes. We cloned the five *Engraulis encrasicolus*' zp genes and deduced their products, determined their tissue distribution, quantified their mRNA expression throughout the reproductive cycle and provided an insight into their evolution through phylogenetic tools. Furthermore, we proposed a multivariate statistics-based method to objectively infer and/or confirm the classification of *Engraulis encrasicolus*' sexual maturity stages by analyzing data of zp mRNAs' relative abundance.

1. Introduction

Chordates, including mammals, amphibians, birds and fish, possess an egg surrounded by an extracellular coat that is mainly composed of several glycoproteins, known as Zona Pellucida (ZP) proteins. The eggshell serves many essential functions: it possesses bactericidal properties against both Gram-positive and -negative bacterial infections (Kudo, 2000). It plays important roles in biological processes such as oogenesis, nutrients uptake, functional buoyancy (Podolsky, 2002), specific sperm binding and guidance of the sperm to the micropyle (Dumont and Brummett, 1985); it also provides protection from the physical environment to the growing oocyte (Hedrick, 2008) as well as to the developing embryo until the hatching phase is reached (Babin et al., 2007), therefore directly and heavily influencing potential fecundity, hatching rate, larval survival and, ultimately, recruitment. As fertilization in fish relies on the entrance of a single sperm into the micropyle (Yue et al., 2014) and teleost sperm lacks acrosome (Hart, 1990), the need for sperm binding and acrosome reaction is diminished (Babin et al., 2007). For this reason, it appears that teleost egg envelope proteins have mainly a structural role (Litscher and Wassarman, 2007).

Over the years, Egg Envelope Proteins (EEPs) have been termed in a variety of ways: Vitelline Envelope proteins (Hyllner et al., 2001), Zona Radiata proteins (Oppen-Berntsen et al., 1992b) and

* Corresponding author.
E-mail addresses: a.miccoli@univpm.it (A. Miccoli), iole.leonori@an.ismar.cnr.it (I. Leonori), andone.estonba@ehu.es (A. Estonba), andrea.defelice@an.ismar.cnr.it (A. De Felice), c.piccinetti@univpm.it (C.C. Piccinetti), o.carnevali@univpm.it (O. Carnevali).

http://dx.doi.org/10.1016/j.ympev.2016.04.006
1055-7903/

Zona Pellucida proteins (Chang et al., 1996). This ambiguous nomenclature has generated considerable confusion, especially because it was used differently not only among chordates but also within teleost species (Babin et al., 2007). As to molecular phylogenetic analyses of vertebrate ZP proteins, Spargo and Hope (2003) proposed a four-subfamily classification of vertebrate *zp* genes: ZPA (not found in any teleost species), ZPB, ZPC and ZPX –identified in *Xenopus*, chicken and fish (Smith et al., 2005) but never in mammals. Subsequently, Goudet et al. (2008), in an attempt to establish an updated list of the ZP genes in mammals, chickens, frogs and some fish species, regarded Spargo and Hope's nomenclature system as logical and considered it as the starting point for further investigations. As to phylogenetic analyses, *ZP* genes were also classified into the ZPD subfamily.

In the large majority of vertebrates, the genes codifying for the egg envelope proteins are only transcribed in developing oocytes. The exception to this rule is observed in some Teleosts – particularly, in the Euteleostei clade (e.g. Salmoniformes) – and in birds (Kanamori et al., 2003). For instance, chicken ZPB1 is expressed in the liver while its ZPC and ZPD are expressed in ovarian granulosa cells (Smith et al., 2005). ZP proteins in rainbow trout and a in large number of other non-cyprinoid fish – e.g. cod – (Oppen-Berntsen et al., 1992b) are synthesized in the liver upon estrogen induction, and the consequently secreted glycoproteins are then transported to the ovary via the bloodstream, where they are incorporated into the vitelline envelope of growing oocytes (Berg et al., 2004). The reason behind such double site of expression is still uncertain. Possibly, it was the result of a gene duplication event that took place in Euteleosts and led to genome complexity, after which one of the two resulting ovarian *zp* gene copies was substituted with another expressed in the liver, even though it is unclear what the advantages of such a substitution might be (Conner and Hughes, 2003).

A great deal of information about the molecular structure (Yamagami et al., 1992), characterization (Darie et al., 2004) and phylogenetic relationships (Goudet et al., 2008; Spargo and Hope, 2003) are therefore available on fish egg envelope components. This statement does not apply to the globally distributed Clupeiformes order, nor to the European anchovy *Engraulis encrasicolus*, an important coastal pelagic species included in the order. In fact, many aspects of the European anchovy's biology, which sometime exemplify those of other Clupeiformes species, have been thoroughly investigated. For instances, vertical distribution (e.g. Sabatés et al., 2008), reproductive strategies (e.g. Schismenou et al., 2012) and the effect of the environment on them (e.g. Basilone et al., 2006), feeding ecology (e.g. Borme et al., 2009), population dynamics (e.g. Bakun, 2010) and genetics (e.g. Montes et al., 2013). Nonetheless, Clupeiformes' reproductive physiology is largely uncharacterized and, specifically, no reproductive physiology-related information whatsoever exist on the European anchovy. Furthermore, at present days, the zonagenesis process in the Otocephala superorder, to which the *Engraulis sp.* belongs, has been limited to few group of fishes, that are Cypriniformes (e.g. *Danio rerio*, *Cyprinus carpio*, *Carassius carassius* and *Carassius auratus*), Clupeiformes (*Engraulis japonicus* and *Clupea pallasii*) and Gonorynchiformes (*Chanos chanos*).

With the aim of filling the ZP protein-wise gap of knowledge afflicting the understanding of the European anchovy's reproductive process as well as expanding the current limited background on Clupeiformes and the Otocephala superorder, we identified the *Engraulis encrasicolus zp* genes, determined the female tissue distribution, quantified their mRNA expression by Real Time-qPCR along the sexual maturity stages as well as provided an insight into their evolution by comparing ZP proteins belonging to Euteleostei, Elopomorpha and Otocephala species through phylogenetic analyses. Furthermore, we propose a solid, reliable and broadly applicable method to infer and/or confirm the classification of bony fishes' sexual maturity stages by analyzing Real Time quantification data of *zp* mRNAs' relative abundance with statistical tools such as the multivariate analyses of variance (PERMANOVA) and the Canonical Analyses of Principal coordinates (CAP).

2. Materials and methods

2.1. Samples collection

Engraulis encrasicolus female specimens at different sexual maturity stages were collected during the 2014 MEDIAS GSA17 and GSA18 (Leonori et al., 2011, 2012; MEDIAS, 2012) research cruises carried out in the months of July, August and September in the Adriatic Sea by the acoustic research group of the CNR-ISMAR of Ancona. The sexual maturity classification was macroscopically assigned based on the guidelines for bony fish reported in the Instruction Manual of the MEDITS working group (2012). Abbreviated stages are explained as follows. Ind: "Indeterminate"; F1: "Immature"; F2c: "Maturing"; F3: "Mature/Spawner"; F4a: "Spent"; F4b: "Resting".

Five individuals per sexual stage were collected, hence a total of 30 specimens were analyzed. Liver and gonad tissues were immersed in RNA*later* and stored at −20 °C until processing.

2.2. RNA extraction and cDNA synthesis

Total RNA was extracted using RNAzol® RT reagent (SIGMA–ALDRICH®, R4533) according to the manufacturer's instructions and eluted in RNAse-free water. Final RNA concentrations were determined using the Nanophotometer TM P-Class (Implem GmbH, Munich, Germany). Its integrity was verified by gel red staining of 28S and 18S ribosomal RNA bands on a 1% agarose gel. RNA was kept at −80 °C until cDNA was synthesized with the Tetro Reverse Transcriptase cDNA synthesis kit (Bioline, BIO-65050) and a total amount of 3 μg of RNA used as input template. Nucleic acids were then kept at −20 °C until use.

2.3. Homology searches

Recently, raw reads and the *de novo* assembled transcriptome of a pool of individuals of *Engraulis encrasicolus* were made available (Montes et al., 2013). Such reads derive from four different tissues (muscle, brain, liver and gonad; Bioproject accession number PRJNA193183) sampled from specimens belonging to three genetically divergent populations (Bay of Biscay, Mediterranean Sea and Atlantic Ocean). Both partial and full-length nucleotide sequences of EEPs belonging to closely related fish species present in NCBI GenBank were used to query the database. Genes were named and annotated according to the nomenclature of Spargo and Hope (2003) and by means of sequence similarity identified either by BLASTn or tBLASTx searches setting an E-value of 10^{-3} (Table 2).

2.4. ORF extension

EEPs' open reading frames were obtained by conventional/RACE PCRs, cloning and sequencing performed on either liver or gonad tissues. Because of the high genomic homology between the two species, forward primers for extending the European anchovy sequences missing the start codon were designed based on the Japanese anchovy ones (AcZPCa and AcZPBb, whose GenBank accession numbers are AB759551.1 and AB759550.1, respectively), while reverse primers were designed on the partial transcripts retrieved from the transcriptome.