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.

When both start and stop codons were available and the unknown region was internal to the CDS, primers were designed based on the transcriptome sequence itself. This was the case for EeZPCc. Primers information for ORF extension are reported in Table 1A.

Due to the lack of the Japanese anchovy AcZPCb 5′ end, *Clupea pallasii* HgZPCb (GenBank accession number AB759549.1), a closely related species, was selected as template to design primers on. RACE cDNA from gonad samples was synthesized with the FirstChoice® RLM-RACE Kit (Life Technologies, AM1700M) following the manufacturer's instructions. Primers used for the 5′RACE-PCRs are available in Table 1B.

For each gene, 25 μl of 2× concentrated PCR Master Mix (Life Technologies, K0171), 2.5 μl of each forward and reverse specific primer designed on *E. encrasicolus*, 5 μl of liver (ZPBa and ZPCa) or ovary (ZPBb, ZPCb, ZPCc) cDNA and 15 μl of H_2O were used. The "touchdown" thermal protocol was used and consisted of the following conditions: (i) 1 cycle of initial denaturation at 94 °C for 3 min; (ii) 10 cycles of denaturation at 94 °C for 30 s, annealing at 62–52 °C (temperature decreased by 1 °C at each consecutive cycle) for 30 s and extension at 72 °C for 50 s; (iii) 25 cycles of denaturation at 94 °C for 30 s, annealing at 54 °C for 30 s and extension at 72 °C for 50 s. The last extension step was prolonged by ten minutes.

After a 2% agarose gel electrophoresis run stained with GelRed (Diatech Pharmacogenetics, 41003), specific PCR products were excised, purified with the MinElute Gel Extraction Kit (Qiagen, 28604) and cloned into the pGEM-T vector system (Promega, Madison, Wis., USA). DNAs were sequenced using a T7 sequenase version 2.0 DNA sequencing kit (Amersham, Piscataway, N.J., USA) and a dye-terminator kit (Perkin Elmer, Foster City, Calif., USA) on an ABI PRISM DNA sequencer (Perkin Elmer).

2.5. Protein prediction and identification

Several web-based bioinformatics tools were used for data mining. Protein prediction and characterization was performed by means of Expasy Translate (http://web.expasy.org/translate/), ORF Finder (http://www.ncbi.nlm.nih.gov/projects/gorf/), the BLAST suite (Madden, 2003 – http://blast.ncbi.nlm.nih.gov/Blast.cgi), SignalP (Petersen et al., 2011 – http://www.cbs.dtu.dk/services/SignalP/), SMART (Schultz et al., 1998; Letunic et al., 2015 – http://smart.embl-heidelberg.de/), PFAM (Finn et al., 2014 – http://pfam.xfam.org/), PDBsum (https://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?pdbcode=index.html).

2.6. Tissue-specificity RT-PCR

Tissue specificity PCR amplification was performed in cycles of 30 s at 95 °C, 30 s at 60 °C and 40 s at 72 °C using a reaction mix containing 5 μl of 2× concentrated PCR Master Mix (Life Technologies, K0171), 1 μl of 10 μM specific forward and reverse primer (Table 1C), 1 μl of template (either ovary or liver cDNA synthesized from stage 3 females) and 3 μl of RNAse free water. Amplification specificity was confirmed by PCR products sequencing and by retrieving gel bands of expected sizes.

Table 1A
Primer sequences used for ORF extension.

	Sequence (5′–3′)	
	Forward	Reverse
ZPBa	GCACCAGAGATGCTATAATCGT	CCACTTGCAATCCTCATTTCCAC
ZPBb	CAGTGTAGTGTAGTGTCGTGGT	GCAGGCAGTGACAGGGAAC
ZPCa	GGGCTGGACGTCACGACTC	TCCACCATAGAAACTGGACAGC
ZPCb_1	ATCTTAGCAATGGGGTTCAGAGC	GCCAGTGCCAAAGAAGTCCC
ZPCb_2	TGGTTTAGTGTTGTTTTGGTGCT	GATTCACCTCCACCTGCACA
ZPCc	CCCCACATGAAGCACCCAA	CTAGGTAAGGAGTTCACGTCTGG

Table 1B
Primers used for 5′RACE PCRs.

	Sequence (5′–3′)
5′ RACE outer primer	GCTGATGGCGATGAATGAACACTG
5′ RACE inner primer	CGCGGATCCGAACACTGCGTTTGCTGGCTTTGATG
ZPCb_f1	ATCTTAGCAATGGGGTTCAGAGC
ZPCb_r1	GCCAGTGCCAAAGAAGTCCC
ZPCb_f2	TGGTTTAGTGTTGTTTTGGTGCT
ZPCb_r2	GATTCACCTCCACCTGCACA

Table 1C
Primer sequences used for RT/qPCRs.

	Sequence (5′–3′)	
	Forward	Reverse
EeZPCa	CGCCGTATTCTGCCAAAAGG	CCCTGTGACCTGTGCATCTT
EeZPCb	TGTGGCAGCGAACTTGAGAT	CCAGAATGTCCTCCGCAGTT
EeZPCc	GCTGCACAAATGTGGAAGCA	AGGTGGGCTTCAGATCGTTG
EeZPBa	TGTGAGGTTTCTGTGTGCCA	ACGGTAGTCCCTTGCCTTTG
EeZPBb	TGCAGTCAGAGATGATGGCC	GGTCCCGGATCATCTTGGTC
β-actin	CGTGACATCAAGGAGAAGCTGTGC	CAGACTCATCGTACTCCTGCTTGC
EF-α	GAGACAGCAAGAACGACCCA	AGAACTTGCAGGCGATGTGA

2.7. Quantitative polymerase chain reaction (Real Time q-PCR)

q-PCRs were performed in an iQ5 iCycler thermal cycler (Bio-Rad, 179-8891) with the SYBR green method according to Miccoli et al. (2015). The optimal annealing temperature was found to be 60 °C for each of the EEPs and housekeeping genes. Primer specificity was obtained after optimization with different times and temperatures. A single peak was observed in the melt curve at the end of the amplification cycle, indicating the lack of primer-dimer formation. Primer sequences are available in Table 1C.

2.8. Phylogenetic analyses

The evolutionary history of 45 ZP protein sequences of 20 worldwide distributed fish species belonging to the Euteleostei, Elopomorpha and Otocephala clades was inferred by using the Maximum Likelihood method based on the Whelan and Goldman model (2001). Phylogenetic analyses were performed with the MEGA suite, version 6 (Tamura et al., 2013), according to Hall (2013). Protein alignment was made with the MUSCLE algorithm (Edgar, 2004). Before constructing a Maximum Likelihood phylogenetic tree, aligned protein sequences were screened for the best substitution model to be employed, which resulted in the Whelan And Goldman one (Whelan and Goldman, 2001). The initial tree for the heuristic search was obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 2.9549)). Inferred phylogeny was tested with the Bootstrap method and the number of bootstrap replications was set at 1000.

2.9. Statistical analysis

Real Time q-PCRs for assessing relative mRNA abundance levels throughout the sexual reproductive cycle were analyzed with an ordinary one-way analyses of variance (ANOVA). Values were compared regardless of sexual maturity stages. Pairwise comparisons were tested with the *post hoc* Tukey test and confidence interval was set at 95% ($p < 0.05$).

The same data were also used as input in a non-parametric permutational multivariate analyses of variance (PERMANOVA or NPMANOVA – Anderson, 2006) with the PERMANOVA+ add-on package of the PRIMER software, version 6, according to the software tutorial (Anderson et al., 2008). Multivariate patterns were checked from data in a normalized Euclidean distance matrix, and pairwise comparisons between groups were performed. The number of permutations was always set at 999 and the *p*-value was calculated through the Monte Carlo method. Eventually, on the basis of the same distance measure chosen before, the matrix was subjected to the CAP constrained ordination (Anderson and Willis, 2003) with the sexual maturity stage as response variable.

3. Results

3.1. ZPB and ZPC gene annotation

The ORFs of five *Engraulis encrasicolus* egg envelope proteins were obtained by a combination of NGS and RT-PCRs/RACE-PCR. Regarding the former, the results from the BLASTn and tBLASTx searches against the trimmed reads, contigs and isotigs allowed us to identify five putative *zp* genes displaying a high level of similarity with the Japanese anchovy's egg envelope proteins. Details of the BLASTn searches are reported in Table 2. Three candidate genes (*zpcc*, *zpba* and *zpbb*) were represented by more than one isotig and therefore assembled manually based on obvious regions of identity highlighted by sequence analyses through ClustalW (Sievers et al., 2011 –http://www.ebi.ac.uk/Tools/msa/) and Blast (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

About the latter, specific pairs of primers were designed to cover the unknown regions. Eventually, Ee*zpba*, Ee*zpbb*, Ee*zpca*, Ee*zpcb*, Ee*zpcc* nucleotide sequences comprised 1677, 1521, 1276, 1122 and 1427 bp, encoded for 480, 493, 401, 373 and 445 AAs, respectively. Importantly, we did not succeed in obtaining the full-length open reading frame of Ee*zpcb*. The sequences were deposited into the NCBI's GenBank database and are identified by the following accession numbers: KU518807, KU518808, KU518809, KU518810 and KU518811, respectively.

Amino acid deduced sequence analyses revealed the presence of common elements in each EEP. They all had a N-terminal signal peptide (with the exception of Ee*zpcb*) and shared a conserved characteristic domain known as ZP domain. Its length was not constant in every AA sequence; rather it consisted in 280, 282, 257, 254 and 255, residues for Ee*zpba*, Ee*zpbb*, Ee*zpca*, Ee*zpcb* and Ee*zpcc*. The number of the ZP domain's conserved Cysteines varied from 8 (Ee*zpca*, Ee*zpcb* and Ee*zpcc*) to 12 (Ee*zpba* and Ee*zpbb*).

A structural element that was proposed to have a role in the polymerization of ZP proteins (Jovine et al., 2004), the Internal Hydrophobic Patch, was found in the ZP domains following Cys residues 1–4 in all five forms.

A 40- and 42-AAs long Cysteine-rich so called Trefoil Domain was exclusively present in Ee*zpba* and Ee*zpbb*. Protein sequence alignment among *Engraulis encrasicolus* EEPs and other available Clupeiformes species (so far, *Engraulis japonicus* – Ac – and *Clupea pallasii* – Hg) are reported in Fig. 1A and B. Residues are very much conserved, as indicated by the asterisks and the colon punctuation marks.

A graphical overview of *Engraulis encrasicolus*' egg envelope proteins showing conserved domain of such glycoprotein family is reported in Fig. 2.

At last, the percentages of nucleotide identity among both Clupeiformes transcripts and ZP domains are present in Tables 3A and 3B. In both cases, the highest degree of similarity is found within the *Engraulis* genus for all EEP isoforms; yet, *zpba* and *zpca* (and the corresponding proteins) are much conserved also between *E. encrasicolus* and *C. pallasii*, with the lowest identity percentage equaling 55.29% (EeZPCa – HgZPCa).

3.2. zp gene expression – tissue specificity

The expression of *zp* genes was analyzed qualitatively by means of RT-PCR using cDNA synthesized from the ovary and liver as templates (Fig. 3).

After 35 cycles of amplification, three *zp* transcripts (Ee*zpba*, Ee*zpcb* and Ee*zpcc*) were shown to be exclusively transcribed in one tissue: Ee*zpba* was amplified solely in the liver, whereas Ee*zpcb* and Ee*zpcc* just in the ovary.

Ee*zpbb* showed a predominant expression in the ovary, yet a band of the same molecular size was very faintly visible also in the liver. At last, Ee*zpca* had a peculiar expression pattern, in that its transcription appeared in both tissues, even though a stronger band was found in the liver. Two housekeeping genes were used, *ef-α* and *β-actin*.

Taken together, these results demonstrate that the European anchovy, a member of the Clupeiformes order, similarly to its Japanese sibling *Engraulis japonicus,* possesses two subfamilies of *zp* genes that predominantly express in either the liver or the ovary, with the exception of Ee*zpca*.

3.3. zp gene expression – Real Time q-PCR

The relative abundance of the various EEPs was investigated semi-quantitatively among different sexual maturity stages in the tissue where predominant expression was previously found by tissue specificity PCR (Fig. 4). Specifically, Ee*zpba* and Ee*zpca* (Fig. 4A and C) were monitored in the liver, while Ee*zpbb*, Ee*zpcb* and Ee*zpcc* (Fig. 4B, D and E) in the ovary.

Two distinct general expression trends were found. Ee*zpbb*, Ee*zpca* and Ee*zpcc* were present in more or less stable levels throughout the active reproductive cycle. The few statistically significant stages highlight the transcripts' abundance evenness.

Instead, a decreasing pattern along the sexual maturity stages was found for Ee*zpba* and Ee*zpcb*, and maximum relative abundance interested the F2C ("Maturing") and the F1 ("Immature") ones, respectively.

Table 2
BLASTn searches against trimmed reads, contigs and isotigs databases. E-value cutoff set between 0 and 10^{-3}.

Query	GenBank accession number	Subject	% identity	Alignment length (nt)	Query start	Query end	E-value
AcZPBa	Dr. Yasumasu, Sophia University (Tokyo), personal communication	isotig_v2_07327	97.81	730	714	1443	0.00E+00
		isotig_v2_11474	99.09	550	1	549	0.00E+00
AcZPBb	AB759550.1	isotig_v2_08245	98.40	751	673	1423	0.00E+00
		isotig_v2_20951	91.91	136	1398	1533	2.00E−47
AcZPCa	AB759551.1	isotig_v2_06483	98.58	985	319	1302	0.00E+00
AcZPCb	AB759552.1	isotig_v2_06409	98.91	1012	77	1088	0.00E+00
AcZPCc	AB759553.1	isotig_v2_06626	98.22	562	457	1018	0.00E+00
		isotig_v2_18443	98.97	387	1043	1427	0.00E+00

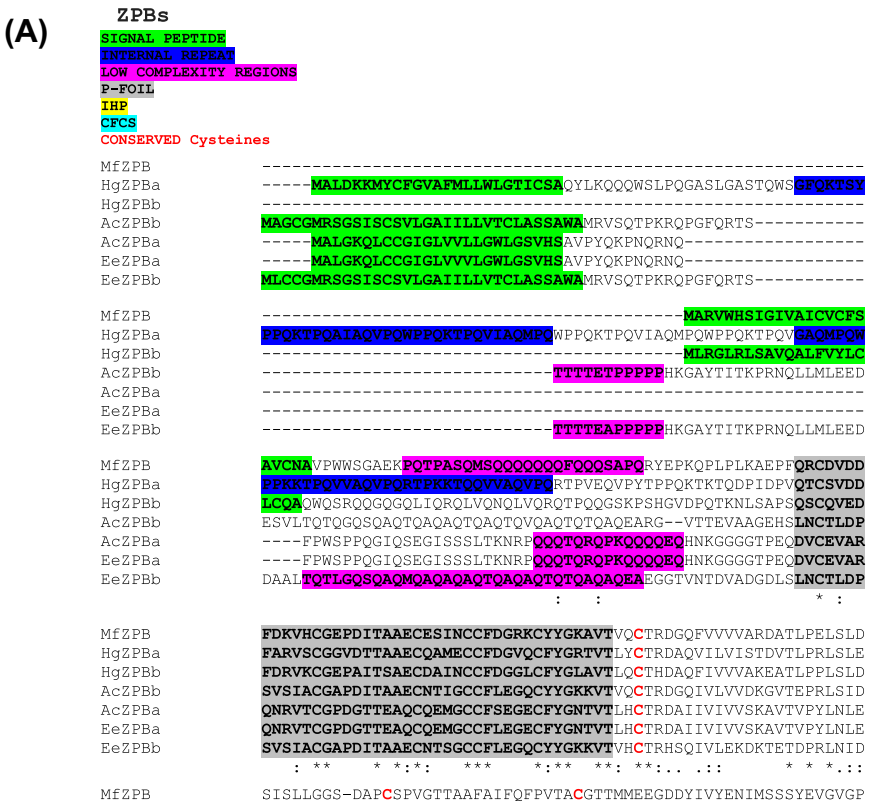

Fig. 1. ClustalW amino acid alignment of *Chanos chanos* (Mf), *Clupea pallasii* (Hg), *Engraulis japonicus* (Ac) and *Engraulis encrasicolus* (Ee) ZPBs (A) and ZPCs (B) Egg Envelope Proteins. Asterisks indicate positions that have a single, fully conserved residue. Colons indicate conservation between groups of strongly similar properties which scored more than 0.5 in the Gonnet PAM 250 matrix. Periods indicate conservation between groups of weakly similar properties, which scored 0.5 or less in the Gonnet PAM 250 matrix.

(B)

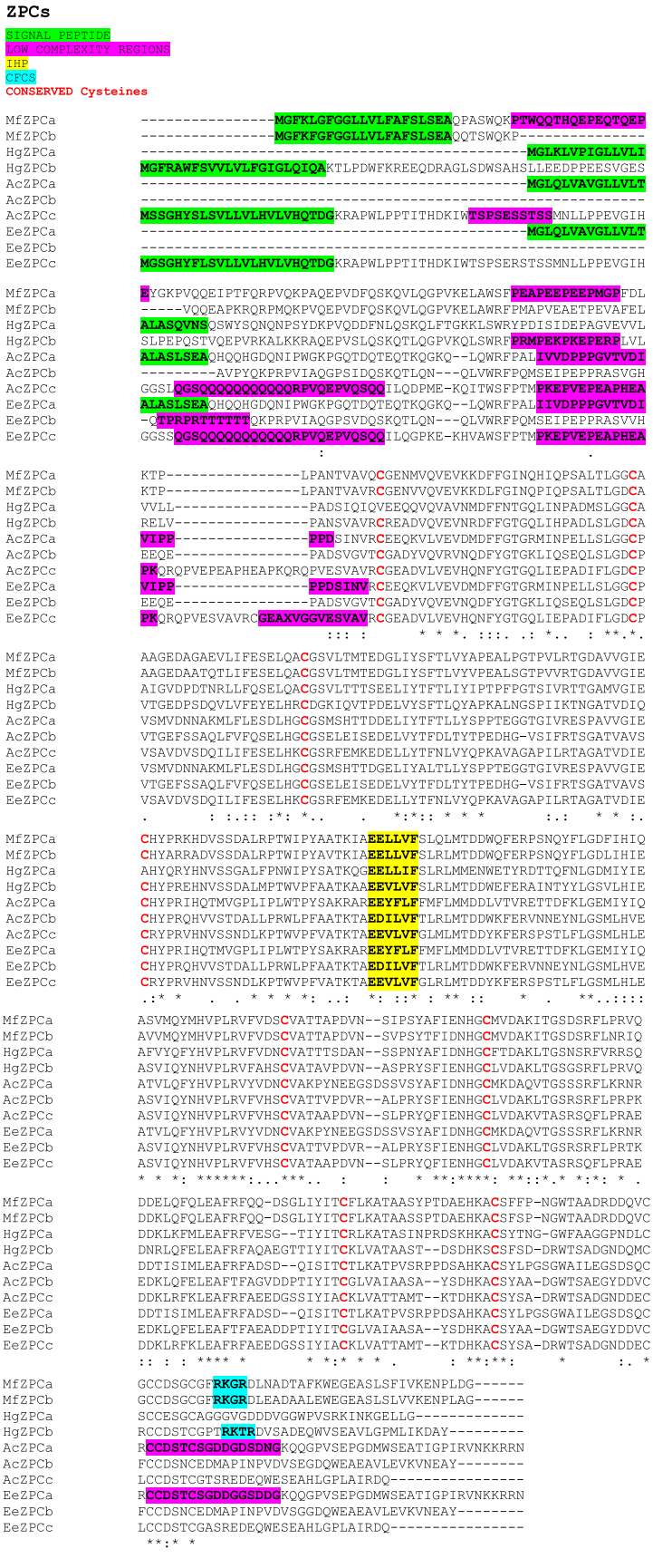

Fig. 1 (*continued*)

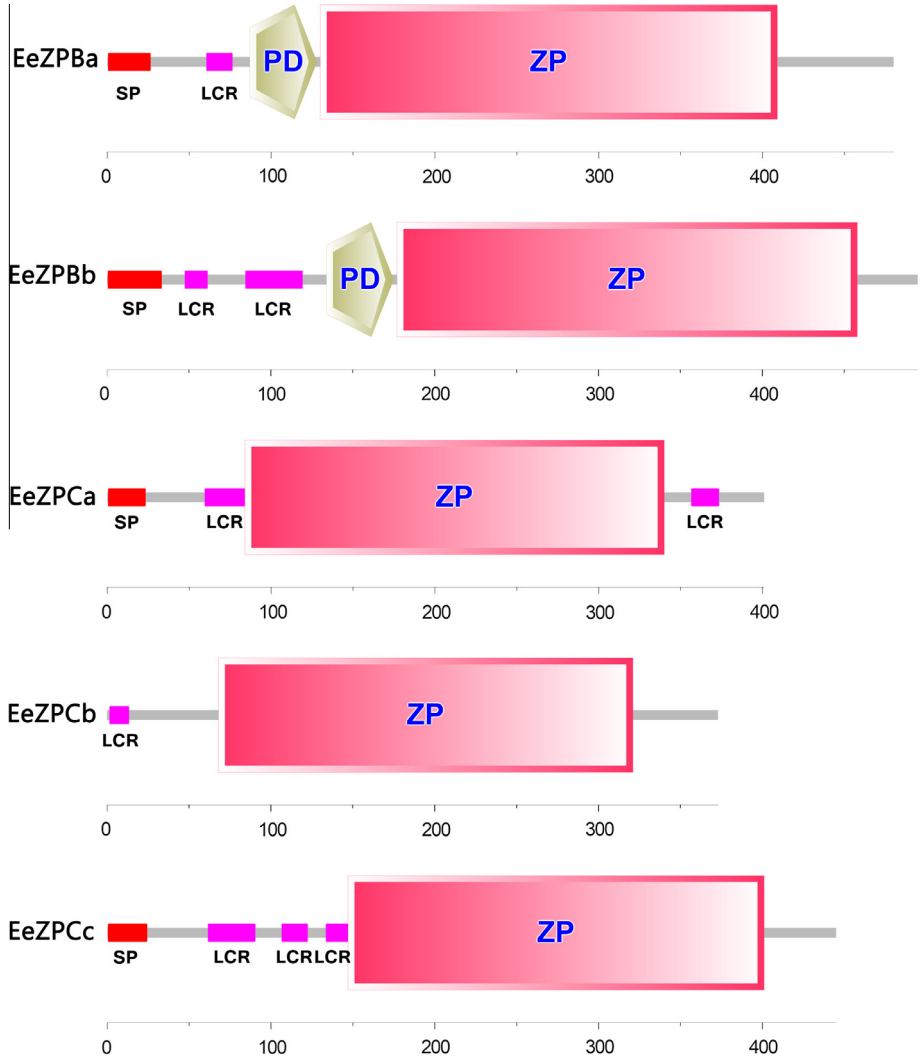

Fig. 2. Schematic representation of EeZPB and EeZPC protein sequences obtained with the SMART tool. SP = Signal Peptide; LCR = Low Complexity Region; PD = Trefoil Domain; ZP = Zona Pellucida Domain.

Table 3A
Identity among Clupeiformes zpb/zpc gene sequences. Comparisons between *Engraulis sp.* EEP isoforms are reported in bold.

	Clupeiformes zpb gene sequences			
	Aczp*ba*	Aczp*bb*	Hgzp*ba*	Hgzp*bb*
Eezp*ba*	**99.24%**	48.02%	58.68%	51.28%
Eezp*bb*	47.99%	**94.81%**	50.54%	52.16%

	Clupeiformes zpc gene sequences				
	Aczp*ca*	Aczp*cb*	Aczp*cc*	Hgzp*ca*	Hgzp*cb*
Eezp*ca*	**99.06%**	49.69%	49.16%	60.89%	49.47%
Eezp*cb*	49.95%	**97.54%**	63.84%	53.07%	69.60%
Eezp*cc*	48.37%	62.20%	**95.94%**	49.87%	65.90%

Relative abundances of the EEP mRNAs in sexually active females always showed their minimum in the F4b ("Resting") stage, and statistical significance in such sense was found for Eezp*ba*, Eezp*ca* and Eezp*cc*.

3.4. Phylogeneny

The tree extracted with the highest log likelihood (−16,780.7877) is shown in Fig. 5. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site.

Table 3B
Identity among Clupeiformes ZPB/ZPC Zona Pellucida Domains. Comparisons between *Engraulis sp.* EEP isoforms are reported in bold.

	Clupeiformes ZPB ZP domain			
	AcZPBa (280 aa)	AcZPBb (282 aa)	HgZPBa (288 aa)	HgZPBb (280 aa)
EeZPBa (280 aa)	**97.86%**	43.32%	57.14%	45.32%
EeZPBb (282 aa)	42.24%	**93.26%**	45.71%	45.68%

	Clupeiformes ZPC ZP domain				
	AcZPCa (257 aa)	AcZPCb (254 aa)	AcZPCc (255 aa)	HgZPCa (255 aa)	HgZPCb (255 aa)
EeZPCa (257 aa)	**98.83%**	43.65%	45.45%	55.29%	45.45%
EeZPCb (254 aa)	45.24%	**98.43%**	63.78%	51.98%	67.32%
EeZPCc (255 aa)	46.64%	63.39%	**99.61%**	50.99%	67.84%

All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position.

Fish EEPs were placed into two main clusters, ZPC and ZPB. Two candidates (*Oryzias latipes* ZPA and *Sparus aurata* ZPX) were chosen as outgroups and grouped separately from the two clusters. Elopomorpha, Otocephala and Euteleostei superorders are marked among branches, and the high bootstrap values support the statistical significance of the tree's conformation.

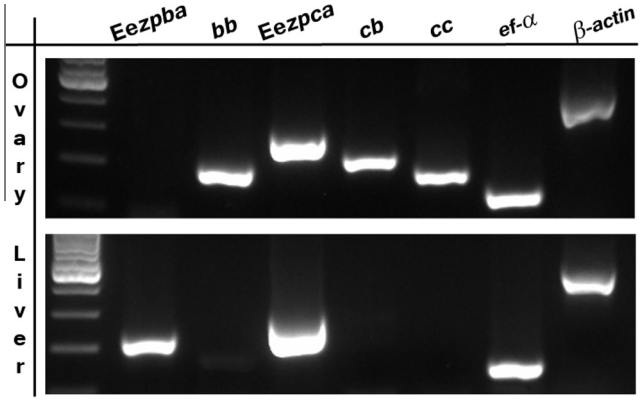

Fig. 3. Qualitative expression analyses of *Engraulis encrasicolus* EEPs in the Ovary and Liver after 35 cycles of RT-PCR amplification.

Clupea pallasii and *Engraulis sp.* sequences strongly correlated into each of the ZPC and ZPB subtrees. Gonorynchiformes representative *Chanos chanos*, which belongs to the Ostariophysi superorder, appeared more closely related to Clupeomorpha species than to other related order such as Cypriniformes.

3.5. PERMANOVA and CAP statistical analyses

The multivariate analyses of the variance performed on the complete dataset of *zp*'s relative mRNA abundance *a priori* classified into the six sexual developmental stages resulted in an overall statistically significant difference ($p = 0.001$). The pairwise test only returned statistically significant *p*-values as well (Table 4).

The CAP statistical test returned a plot (Fig. 6A) in which the individual values of each *a priori* assigned group clustered closely to its replicates, altogether forming compact clouds easily

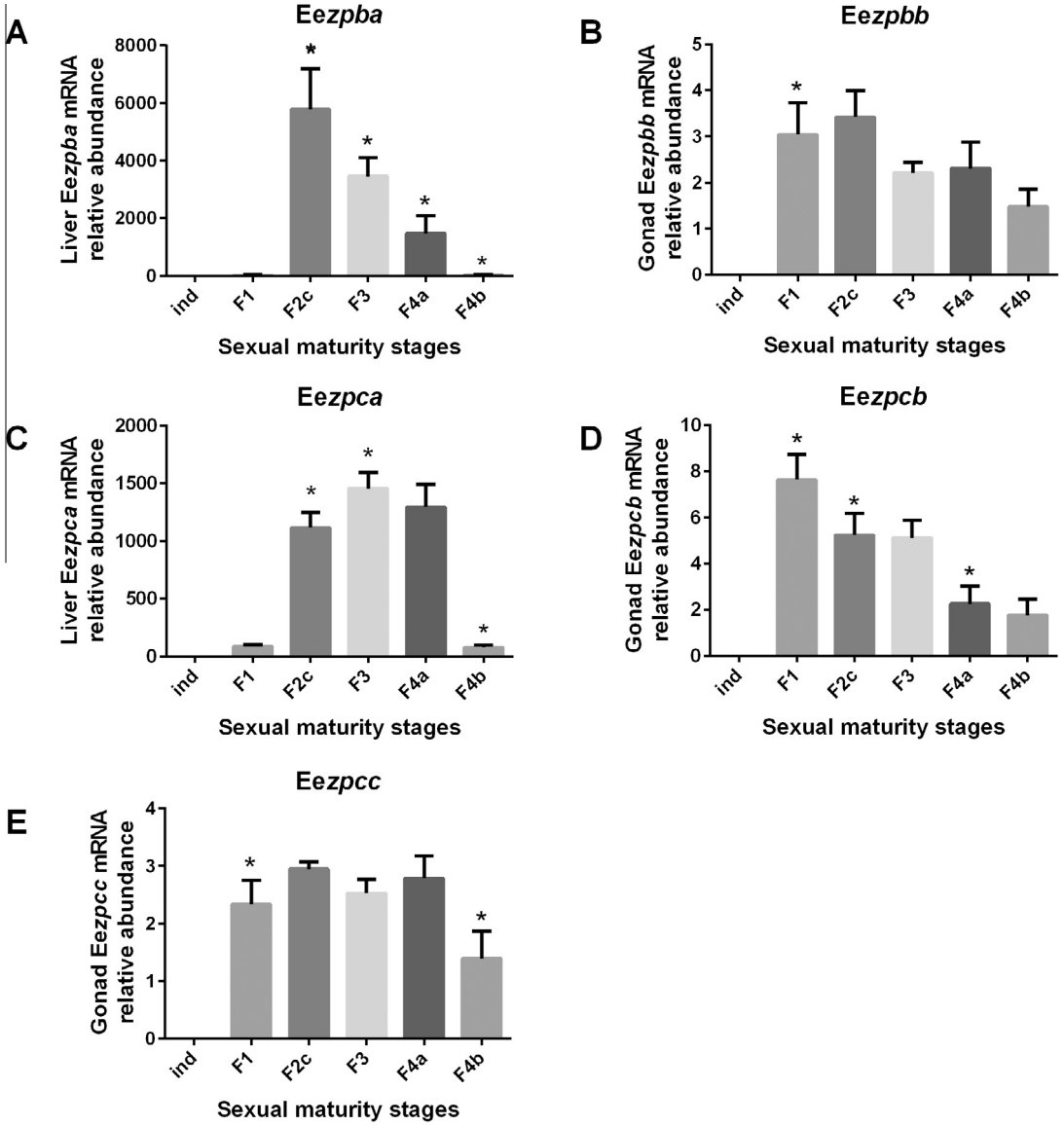

Fig. 4. Semi-quantitative estimation of EEPs mRNA abundance calculated over two housekeeping genes, *ef-α* and *β-actin* throughout *Engraulis encrasicolus* reproductive cycle. Eezpba and Eezpca were monitored in the liver, while Eezpbb, Eezpcb and Eezpcc in the ovary. Asterisks represent statistical significance ($p < 0.05$) between two consecutive sexual maturation stages, as indicated by the One-Way Anova and the *post hoc* Tukey's test.

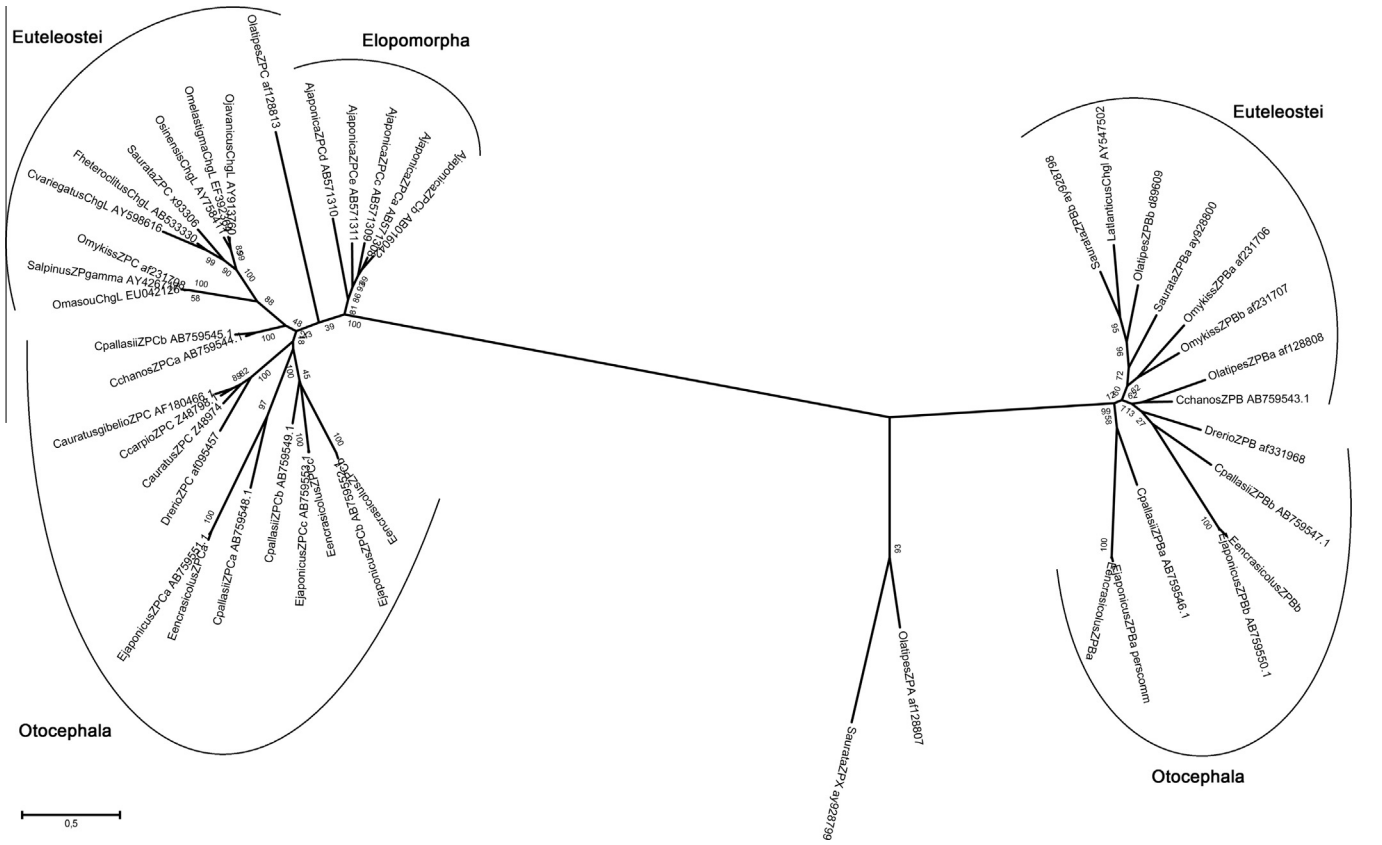

Fig. 5. Egg envelope proteins sequence similarities shown in a phylogenetic tree build with Mega v.6. GenBank accession numbers are: *Anguilla japonica* ZPCa (AB571308); *Anguilla japonica* ZPCb (AB016042); *Anguilla japonica* ZPCc (AB571309); *Anguilla japonica* ZPCd (AB571310); *Anguilla japonica* ZPCe (AB571311); *Danio rerio* ZPC (AF095457); *Carassius auratus* ZPC (Z48974); *Cyprinus carpio* ZPC (Z48798.1); *Carassius auratus gibelio* ZPC (AF180466.1); *Clupea pallasii* ZPCb (AB759545.1); *Engraulis japonicus* ZPCb (AB759552.1); *Engraulis encrasicolus* ZPCb (KU518810); *Engraulis japonicus* ZPCc (AB759553.1); *Engraulis encrasicolus* ZPCc (KU518811); *Chanos chanos* ZPCa (AB759544.1); *Clupea pallasii* ZPCb (AB759549.1); *Clupea pallasii* ZPCa (AB759548.1); *Engraulis encrasicolus* ZPCa (KU518809); *Engraulis japonicus* ZPCa (AB759551.1); *Oryzias latipes* ZPC (AF128813); *Cyprinodon variegatus* ChgL (AY598616); *Fundulus heteroclitus* ChgL (AB533330); *Sparus aurata* ZPC (X93306); *Oryzias sinensis* ChgL (AY758411); *Oryzias melastigma* ChgL (EF392364); *Oryzias javanicus* ChgL (AY913760); *Oncorhynchus mykiss* ZPC (AF231708); *Oncorhynchus masou* ChgL (EU042126); *Salvelinus alpinus* ZPgamma (AY426717); *Oryzias latipes* ZPA (AF128807); *Sparus aurata* ZPX (AY928799); *Oryzias latipes* ZPBa (AF128808); *Chanos chanos* ZPB (AB759543.1); *Engraulis encrasicolus* ZPBb (KU518808); *Engraulis japonicus* ZPBb (AB759550.1); *Clupea pallasii* ZPBb (AB759547.1); *Danio rerio* ZPB (AF331968); *Engraulis encrasicolus* ZPBa (KU518807); *Engraulis japonicus* ZPBa (personal communication by Prof. Yasumasu, Faculty of Science and Technology, Sophia University, Tokyo); *Clupea pallasii* ZPBa (AB759546.1); *Oncorhynchus mykiss* ZPBa (AF231706); *Oncorhynchus mykiss* ZPBb (AF231707); *Sparus aurata* ZPBa (AY928800); *Oryzias latipes* ZPBb (D89609); *Sparus aurata* ZPBb (AY928798); *Liparis atlanticus* Chgl (AY547502). Relative branch lengths indicate rates of evolution along a particular branch. Bootstrap values are reported at the nodes.

discernible from the others. F2c, F3 and F4a clouds allocated nearer to each other than to the Ind, F1 and F4b.

The "leave-one-out" allocation of observations to groups demonstrated a 100% correctness in four out of five cases, while one replicate firstly identified as belonging to stage F2c was re-classified F3 by the algorithm. Therefore, the total misclassification error rate equaled 4.76% and the correctness 95.24%.

The tests were repeated with five distinct data sets (Fig. 6B–F) each removed from one variable at a time (i.e. one EEP isoform's expression data). Strong significance was found: the PERMANOVA main tests' *p*-value was always 0.001, whereas pairwise's are reported in Table 4. CAP total correctness were 95.24%, 95.24%, 90.48%, 85.71%, 95.24% when Ee*zpba*, Ee*zpbb*, Ee*zpca*, Ee*zpcb* and Ee*zpcc* were respectively removed from the data set.

The plot resulting after Ee*zpcb* removal (Fig. 6E) is the only one in which the clouds intersected noticeably. In particular, F3 crossed F4a (the specific PERMANOVA pairwise *p*-value was 0.045, see Table 4), whereas F1 cloud approached F4b's.

Table 4
PERMANOVA pairwise test's *p*-values reported for all six data sets analyzed.

	PERMANOVA pairwise *p*-values					
	All EEPs	w/o EeZPBa	w/o EeZPBb	w/o EeZPCa	w/o EeZPCb	w/o EeZPCc
ind–F1	0.002	0.001	0.001	0.001	0.001	0.001
F1–F2c	0.001	0.001	0.001	0.003	0.001	0.001
F2c–F3	0.038	0.025	0.038	0.046	0.027	0.031
F3–F4a	0.023	0.031	0.011	0.009	0.045	0.009
F4a–F4b	0.001	0.003	0.001	0.001	0.001	0.002

4. Discussion

The Clupeiformes order comprises 7 worldwide distributed families, more than 80 genera and over 400 species (Nelson, 1994), including herrings, sardines, pilchards, sprats, shads, anchovies and wolf-herrings, which altogether contribute to more than a quarter of the world fish catch (FAO, 1985). In 2003, the Clupeiformes' capture production of marine and freshwater world fisheries was 19,000,000 tons, a value much higher than any other single systematic group of fishes (Lavoué et al., 2007). Such economic importance is not supported by a deep understanding of their reproductive process from a physiological perspective.

To date, studies on Clupeiformes' egg envelope proteins, fundamental to zonagenesis and therefore for ensuring proper oocyte formation and viability, have been carried out just in the Japanese

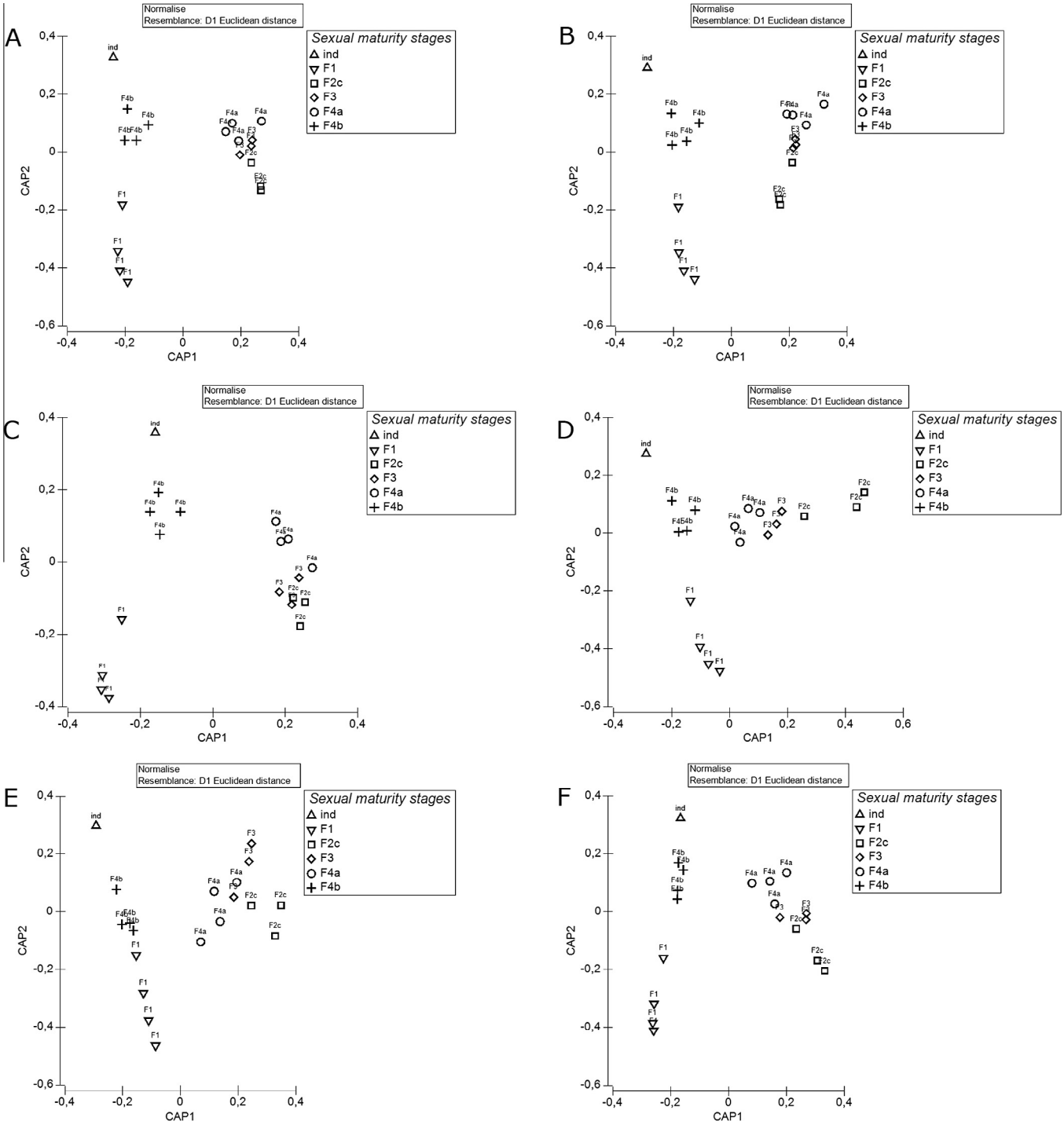

Fig. 6. CAP plots showing the spatial distribution of the *a priori* assigned staging groups. Statistics were run on 6 different data sets, the first one including all five *zp* genes' expression values (A), the remaining five subtracted from Ee*zpba* (B), Ee*zpbb* (C), Ee*zpca* (D), Ee*zpcb* (E) and Ee*zpcc* (F), respectively.

anchovy *Engraulis japonicus* (whose distribution spans from the Sea of Japan to the Java Sea) and the Pacific herring *Clupea pallasii* (a north temperate species, reported in the Artic Sea and the Pacific Ocean at both western and eastern coasts). In this study, we focused on the cosmopolitan *Engraulis* genus and specifically on the European anchovy *Engraulis encrasicolus*. Such species has a geographically wide distribution, as it has been reported in the Eastern, North and Central Atlantic Ocean, in the whole Mediterranean and Black and Azov seas and southward along the African west coast. It is one of the chief targets of a well consolidated fishery, globally accounting for 406.115 tons in year 2013, as to the latest information available (http://www.fao.org/fishery/).

Of it, we have characterized the egg envelope proteins by means of a multidisciplinary approach that included bioinformatics and molecular methodologies, and used phylogenetic analyses to draw a picture about their evolution with respect to other 19 teleost species, both marine and freshwater, belonging to the Euteleostei, Elopomorpha and Otocephala clades. With the exception of EeZPCb, we succeeded to obtain their full-length coding DNA sequence and describe typical inter-taxa evolutionary conserved

domains, as well as identify their primary site of transcription and relatively quantify their mRNA levels in the predominant tissue of expression. The overall aim was to provide, for the first time, the scientific community with an insight into the European anchovy's egg envelope proteins as well as with an expanded knowledge about such aspect of the Clupeiformes order and the Otocephala superorder.

Despite the two Clupeiformes species studied to date (Sano et al., 2010) are characterized by different spawning and hatching strategies (Kawaguchi et al., 2009), our findings are in agreement with their scenarios. In *Engraulis encrasicolus* too we found two classes of ZP-domain containing egg envelope protein genes, *zpb* and *zpc*. They respectively grouped two and three isoforms encoding for what we named EeZPBa, EeZPBb, EeZPCa, EeZPCb and EeZPCc glycoproteins. They display a good conservation with EEP transcripts of other Clupeiformes species, as reported in Table 3A.

Identifying multiple isoforms for each of the EEP clusters should be not a reason of surprise, since many authors already obtained similar results, stating that teleost fish appear to have an expanded family of ZPB and ZPC genes (Kanamori et al., 2003; Mold et al., 2001; Smith et al., 2005). Supposedly, this is due to an evolutionary polyploidization event that served as the starting point from which differences have started to accumulate.

The protein structure of the different ZPs is very similar (Babin et al., 2007). Indeed, all translated sequences retrieved from the *Engraulis encrasicolus* transcriptome and extended by means of either standard or RACE-PCRs contained conserved modules, the most important of which was the ZP domain (Figs. 1A, B and 2). The common presence of the ZP domain in filament- or matrix-forming proteins reflects its fundamental role in protein polymerization. Such homologous region is a ~260/280-amino acid polymerization module containing 8–12 conserved cysteine residues (Fig. 1A and B) involved in intra-molecular disulfide bonds (Darie et al., 2004). It is localized close to the C terminus of many secreted eukaryotic glycoproteins (Jovine et al., 2002) that play key roles in development, hearing, immunity, and cancer (e.g. Jovine et al., 2004). The general ZP domain features were confirmed by our findings and, more precisely, consisted in 280, 282, 257, 254 and 255 residues for Ee*zpba*, Ee*zpbb*, Ee*zpca*, Ee*zpcb* and Ee*zpcc*, respectively. As to Table 3B, the percentages of amino acid similarity among Clupeiformes ZP domains never accounted for less than 42%, showing the highest degree of homology of 99.61% for ZPCs within the *Engraulis* genus.

The number of aligned C residues were also consistent with previous findings and equaled 8 (Ee*zpc*s) and 12 (Ee*zpb*s). One Internal Hydrophobic Patch necessary to the formation of β-strands (Jovine et al., 2004) and the prevention from premature polymerization was found in all five translated proteins (Fig. 1A and B). IHP residues of both ZP clusters appear to be very much conserved among Clupeiformes. In fact, each position, if not identical, was largely represented by amino acids showing strongly similar properties (which scored more than 0.5 in the Gonnet PAM 250 matrix), as indicated by the colon punctuation mark. Noteworthy, IHP positions of each cluster's isoforms were fully conserved between the *Engraulis* genus.

A so-called trefoil domain was detected in Ee*zpba* and Ee*zpbb* proteins but not in the ZPC cluster (identified as the PD mark in Fig. 2). This is in accordance with many reviews (e.g. Litscher and Wassarman, 2007). Despite the uncertainty about its role, it was speculated to confer resistance to proteolytic degradation (Bork, 1993; Suemori et al., 1991).

We were unable to clearly identify neither the External Hydrophobic Patch (EHP) motif nor the C-terminal Trans-Membrane Domain (TMD), consistently with other Clupeiformes members. In fact, the large majority of teleost species, especially those synthesizing ZP proteins in the liver (Sugiyama and Iuchi, 2000), doesn't display such membrane-anchoring region (Kanamori et al., 2003; Yang and Hedrick, 1997) even though, according to Jovine et al. (2005), some can instead have short extensions located C-terminal to the ZP domain, which would anyway be lost either before or during protein polymerization because of a proteolytic cleavage at conserved sites (CFCS – Consensus Furin Cleavage Site).

An exception to teleost lacking TMDs is represented by both the Japanese (Sano et al., 2010) and European (Mazzeo et al., 2012) eels. Likely, this suggests that the Elopomorpha superorder, to which the *Anguilla spp.* belongs, possessed such feature, and allows the scientific community to speculate about the reasons of its loss during evolution. Indeed, the phylogenetic tree that was inferred from various teleost ZP protein sequences distinctly separate the three superorders (Fig. 5). Considering the above-mentioned ZP protein structure similarity, it is conceivable that such difference may depend, if not exclusively, also on the TMD region. Probably, the TMD is unnecessary to Euteleost fishes, since the corresponding genes are expressed in the liver and the secreted proteins transported to the ovary. However, the ovarian-synthesized Otocephala ZP proteins lack a TMD as well (Kanamori, 2000), suggesting that further studies should address such topic more thoroughly.

In accordance with previously studied Clupeiformes members (Sano et al., 2013) and as reported by Kanamori et al. (2003), the majority of *E. encrasicolus* ZP proteins have either ovarian- or liver-specific expression (Fig. 3). A stage 3 female was employed to study the matter, as we wanted to avoid any *zp* isoforms underrepresentation at either expression sites. Ee*zpba*, *cb* and *cc* have an exclusive expression site, while Ee*zpbb* and *ca* are mainly transcribed in one tissue (ovary and liver, respectively) but also slightly in the other. The European anchovy EEPs expression pattern very much resembles those of *C. pallasii* (Hg) and *E. japonicus* (Ac) (Sano et al., 2013). Yet, we found Ee*zpca* mRNA to be present in both ovary and liver, similarly to Hg*zpca* but not Ac*zpca*. Such observation was reported also in rainbow trout (Oppen-Berntsen et al., 1992a), especially for the *zpc* gene (Hyllner et al., 2001), and several other fish species (Babin et al., 2007). Given the genetic similarity of the Japanese and European anchovies, it is unlikely that the same signal in the *Engraulis* genus members studied so far ceased to be expressed in just one of the two species' transcription sites. Rather, we believe that the observed difference lies in the method of investigation used by Sano et al. (2013), and, more precisely in the number of PCR cycles set for the experiment (24/28 against 35 cycles), which were not sufficient to amplify ovarian Ac*zpca* detectably. Nevertheless, we cannot exclude that the differences in the results might depend on a genetic divergence at primordial stage, given the geographical distance and reproductive isolation to which they are subjected.

Among the five EEPs herein presented, we were unable to completely characterize Ee*zpcb*'s open reading frame. A thorough search of *in silico* data was performed, after which several primer combinations were tested in the 5′RACE-PCR technique, without succeeding in obtaining the initiation codon. The same difficulties were faced by Sano et al. (2013), who eventually concluded it is a minor component of the EEPs set, together with ZPBb and ZPCc.

Egg Envelope Proteins mRNA abundance levels were relatively quantified over two housekeeping genes throughout the European anchovy's sexual reproductive cycle (Fig. 4). To our knowledge, this is the first characterization of their expression patterns. These results and their discussion, though, should be addressed in relation with the species' reproductive strategy, according to which females have an asynchronous, indeterminate and batch-spawning ovarian development (Murua and Saborido-Rey, 2003).

The highest EEP mRNA relative abundances were generally found in the F2c and F3 stages, in which the majority of oocytes

in the ovaries were histologically seen to be in an advanced maturation stage (data not shown). However, many female specific genes are known to be subjected to a tightly regulated expression, even at the earliest stage of oogenesis, and some of them consist of egg envelope protein mRNAs (Kanamori, 2000). This statement could explain the high levels of Ee*zpbb*, Ee*zpcb* and Ee*zpcc* we found at the F1 stage. These isoforms were referred to by Sano et al. (2013) as minor components of the vitelline envelope in the Japanese anchovy. It could be interesting to deepen whether the importance of the EEPs contribution could be negatively correlated to their temporal extent of expression.

The phylogeny of a set of Teleost ZPBs and ZPCs homologues containing forty-five egg envelope protein sequences of twenty different species was inferred through a phylogenetic tree and statistically analyzed with the Bootstrap method, with bootstrap replication set at 1000.

Engraulis sp. ZPBs and ZPCs were highly related to each other, as indicated by the similar length of the branches grouping the different EEP gene clusters. The resulting subgroup set itself into the Clupeiformes area. As hinted at earlier, the three studied Clupeiformes species occupy distinct ecosystems, which marginally overlap just for the Pacific herring and the Japanese anchovy in the Sea of Japan, and are therefore subjected to discrete biotic and abiotic factors exerting different evolutionary pressures. Nonetheless, the monophyly of the order, reported by Lavoué et al. (2007), despite less extensively, was also demonstrated EEPs-wise.

Clupeiformes arranged closely to Cypriniformes representatives (*Danio rerio*, *Carassius auratus*, *Cyprinus carpio*), as demonstrated firstly by Ishiguro et al. (2003) and confirmed thereafter by Lavoué et al. (2007).

Interestingly, Gonorynchiformes representative *Chanos chanos*, which belongs to the Ostariophysi superorder, appeared more closely related to Clupeomorpha species than to other known-related orders such as Cypriniformes. Such finding was already discussed (Saitoh et al., 2003); our data provide further support to the hypothesis that Gonorynchiformes and Clupeiformes form a monophyletic group.

Eventually, we propose an innovative, objective and reliable method for confirming and/or inferring the sexual maturity stages classification. We have based such method on *Engraulis encrasicolus*, but its strength relies on its interspecificity, in the sense that it is broadly applicable to potentially every bony fish, given the availability of EEPs mRNA relative abundances to be used as input data.

The rationale behind the technique was to statistically analyze the expression data of the five *zp* genes with a multivariate analyses of the variance and a constrained ordination method that would indicate the spatial distribution of the *a priori* assigned classification, in order to deduce the stage of the reproductive cycle an *Engraulis encrasicolus* specimen is at. In other words, the genes codifying for the various isoforms of the egg envelope proteins were used as a proxy to allow a more objectively staging of the European anchovy.

Results from both the PERMANOVA main and pairwise tests indicated the existence of strong statistical significance among the *zp* genes' expression values of *a priori* assigned sexual stages, therefore demonstrating the feasibility of such method (Table 4). In particular, the *p*-values of the comparisons between consecutive stages (ind–F1, F1–F2c, F2c–F3, F3–F4a, F4a–F4b) were equal to 0.002, 0.001, 0.038, 0.023, 0.001, respectively, indicating the differences in the EEP genes' relative mRNA abundance at discrete sexual stages.

The Canonical Analyses of Principal coordinates confirmed the situation already highlighted by the PERMANOVA. An approach similar to the "leave-one-out" procedure (Anderson and Willis, 2003; Lachenbruch and Mickey, 1968) was used. We were able to define whether different *zp* gene isoforms have a higher/lower impact in altering both the significance among assigned groups and the spatial distribution in the CAP plot by subtracting each one at a time and running the analyses again. The various plots described more or less stable scenarios: the differences were always significant, but what changed was the spatial distribution in terms of proximity among groups. F2c, F3 and F4a (i.e. the reproductively active stages) always clustered close to each other and separately from ind, F1 and F4b (i.e. the sexually inactive groups).

When Ee*zpca* was detracted from the data set to be analyzed, reproductively active vs. inactive stages were less distanced, but the total correctness of the allocation of observations to groups equaled 90.48%. Furthermore, two specimens firstly classified as F2c and F4a were both corrected to F3.

Instead, when Ee*zpcb* was removed, despite significance remained, the spatial distribution of plotted F3–F4a groups overlapped and the separation between F1–F4b clouds decreased. In this case, total correctness lowered to 85.71%. This suggests the importance of *zpcb* for allowing a proper stage allocation and indicates this isoform as possible marker of the reproductive stage.

The q-PCR expression analyses revealed Ee*zpba* to be one of the two main proteins of the European anchovy egg envelope. Nevertheless, neither the significance *p*-values nor the allocation correctness after the removal of Ee*zpba* from the data set submitted to PERMANOVA and CAP analyses changed. In other words, such protein-codifying gene isoform has marginal role in resolving the reproductive stage and is not essential to ensuring the reliability of the model.

We strongly believe that such method could be of great help, considering the Clupeiformes' high economic value and the need of an accurate identification of the sexual maturity stages for proper stock management purposes.

As a conclusion, the study of vertebrate reproductive proteins within closely related species, although fundamental for assessing the evolutionary processes underlying speciation, has been lacking until recently (Turner and Hoekstra, 2008). We have herein presented data processed from a multidisciplinary approach that can contribute to expand the current knowledge on one of the most scientifically important and commercially valuable Teleost group, the Clupeiformes, by specifically focusing on the European anchovy, *Engraulis encrasicolus*. In addition, we have proposed a statistics-based method for classifying bony fish sexual stages, the rationale of which could be applied to a broader range of species. In such way, the subjectivity from which most of the currently employed staging procedures suffer, could be easily overcome.

Conflicts of interests

The authors declare no conflict of interests.

Acknowledgements

We thank Prof. Shigeki Yasumasu of the Faculty of Science and Technology, Sophia University, Tokyo, for providing us with the unpublished Ac*zpba* nucleotide sequence which was used as query in the transcriptome and for the precious email exchange regarding ZPCb destiny evolutionary-wise.

We thank PhD student Jorge Langa of the University of the Basque Country UPV/EHU, Leioa, for the help provided with the *E. encrasicolus* dataset (Montes et al., 2013) and homology searches.

We thank PhD student Valentina Notarstefano of Università Politecnica delle Marche for her assistance with the PRIMER software.

We thank Dr. Claudio Vasapollo for his statistical advices.

We thank also the captain and crew on board the R/V Dallaporta during the research cruise MEDIAS GSA17 and GSA18, alongside all the researchers and the technical personnel involved in the cruise of 2014.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2016.04.006.

References

Anderson, M.J., 2006. A new method for non-parametric multivariate analysis of variance. Aust. Ecol. 26, 32–46.

Anderson, M.J., Gorley, R.N., Clarke, K.R., 2008. PERMANOVA+ for PRIMER: Guide to Software and Statistical Methods. PRIMER-E, Plymouth, UK, 214pp.

Anderson, M.J., Willis, T.J., 2003. Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. Ecology 84, 511–525. http://dx.doi.org/10.1890/0012-9658(2003)084[0511:CAOPCA]2.0.CO;2.

Babin, P., Cerdà, J., Lubzens, E., 2007. The Fish Oocyte: From Basic Studies to Biotechnological Applications. Springer Netherlands, pp. 1–508.

Bakun, A., 2010. Linking climate to population variability in marine ecosystems characterized by non-simple dynamics: conceptual templates and schematic constructs. J. Mar. Syst. 79, 361–373. http://dx.doi.org/10.1016/j.jmarsys.2008.12.008.

Basilone, G., Guisande, C., Patti, B., Mazzola, S., Cuttitta, a., Bonanno, a., Vergara, a.R., Maneiro, I., 2006. Effect of habitat conditions on reproduction of the European anchovy (*Engraulis encrasicolus*) in the Strait of Sicily. Fish. Oceanogr. 15, 271–280. http://dx.doi.org/10.1111/j.1365-2419.2005.00391.x.

Berg, A.H., Westerlund, L., Olsson, P.E., 2004. Regulation of Arctic char (*Salvelinus alpinus*) egg shell proteins and vitellogenin during reproduction and in response to 17β-estradiol and cortisol. Gen. Comp. Endocrinol. 135, 276–285. http://dx.doi.org/10.1016/j.ygcen.2003.10.004.

Bork, P., 1993. A trefoil domain in the major rabbit zona pellucida protein. Protein Sci. 2, 669–670. http://dx.doi.org/10.1002/pro.5560020417.

Borme, D., Tirelli, V., Brandt, S.B., Fonda Umani, S., Arneri, E., 2009. Diet of *Engraulis encrasicolus* in the northern Adriatic Sea (Mediterranean): ontogenetic changes and feeding selectivity. Mar. Ecol. Prog. Ser. 392, 193–209. http://dx.doi.org/10.3354/meps08214.

Chang, Y.S., Wang, S.C., Tsao, C.C., Huang, F.L., 1996. Molecular cloning, structural analysis, and expression of carp ZP3 gene. Mol. Reprod. Dev. 44, 295–304. http://dx.doi.org/10.1002/(SICI)1098-2795(199607)44:3<295::AID-MRD3>3.0.CO;2-H.

Conner, S.J., Hughes, D.C., 2003. Analysis of fish ZP1/ZPB homologous genes – evidence for both genome duplication and species-specific amplification models of evolution. Reproduction 126, 347–352. http://dx.doi.org/10.1530/reprod/126.3.347.

Darie, C.C., Biniossek, M.L., Jovine, L., Litscher, E.S., Wassarman, P.M., 2004. Structural characterization of fish egg vitelline envelope proteins by mass spectrometry. Biochemistry 43, 7459–7478. http://dx.doi.org/10.1021/bi0495937.

Dumont, J.N., Brummett, A.R., 1985. Egg envelopes in vertebrates. Dev. Biol. (N.Y. 1985) 1, 235–288.

Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5, 113. http://dx.doi.org/10.1186/1471-2105-5-113.

FAO, 1985. FAO species catalogue. Vol. 7. Clupeoid Fishes of the World (Suborder Clupeoidei). An Annotated and Illustrated Catalogue of the Herrings, Sardines, Pilchards, Sprats, Shads, Anchovies and Wolf-herrings. Part 1. Chirocentridae, Clupeidae & Pristigasteridae.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L.L., Tate, J., Punta, M., 2014. Pfam: the protein families database. Nucleic Acids Res. 42, D222–D230. http://dx.doi.org/10.1093/nar/gkt1223.

Goudet, G., Mugnier, S., Callebaut, I., Monget, P., 2008. Phylogenetic analysis and identification of pseudogenes reveal a progressive loss of zona pellucida genes during evolution of vertebrates. Biol. Reprod. 78, 796–806. http://dx.doi.org/10.1095/biolreprod.107.064568.

Hall, B.G., 2013. Building phylogenetic trees from molecular data with MEGA. Mol. Biol. Evol. 30, 1229–1235. http://dx.doi.org/10.1093/molbev/mst012.

Hart, N.H., 1990. Fertilization in teleost fishes: mechanisms of sperm-egg interactions. Int. Rev. Cytol. 121, 1–66. http://dx.doi.org/10.1016/S0074-7696(08)60658-0.

Hedrick, J.L., 2008. Anuran and pig egg zona pellucida glycoproteins in fertilization and early development. Int. J. Dev. Biol. http://dx.doi.org/10.1387/ijdb.082580jh.

Hyllner, S.J., Westerlund, L., Olsson, P.E., Schopen, A., 2001. Cloning of rainbow trout egg envelope proteins: members of a unique group of structural proteins. Biol. Reprod. 64, 805–811. http://dx.doi.org/10.1095/biolreprod64.3.805.

Ishiguro, N.B., Miya, M., Nishida, M., 2003. Basal euteleostean relationships: a mitogenomic perspective on the phylogenetic reality of the "Protacanthopterygii". Mol. Phylogenet. Evol. 27, 476–488. http://dx.doi.org/10.1016/S1055-7903(02)00418-9.

Jovine, L., Darie, C.C., Litscher, E.S., Wassarman, P.M., 2005. Zona pellucida domain proteins. Annu. Rev. Biochem. 74, 83–114. http://dx.doi.org/10.1146/annurev.biochem.74.082803.133039.

Jovine, L., Qi, H., Williams, Z., Litscher, E., Wassarman, P.M., 2002. The ZP domain is a conserved module for polymerization of extracellular proteins. Nat. Cell Biol. 4, 457–461. http://dx.doi.org/10.1038/ncb802.

Jovine, L., Qi, H., Williams, Z., Litscher, E.S., Wassarman, P.M., 2004. A duplicated motif controls assembly of zona pellucida domain proteins. Proc. Natl. Acad. Sci. USA 101, 5922–5927. http://dx.doi.org/10.1073/pnas.0401600101.

Kanamori, A., 2000. Systematic identification of genes expressed during early oogenesis in medaka. Mol. Reprod. Dev. 55, 31–36. http://dx.doi.org/10.1002/(SICI)1098-2795(200001)55:1<31::AID-MRD5>3.0.CO;2-7.

Kanamori, A., Naruse, K., Mitani, H., Shima, A., Hori, H., 2003. Genomic organization of ZP domain containing egg envelope genes in medaka (*Oryzias latipes*). Gene 305, 35–45. http://dx.doi.org/10.1016/S0378-1119(02)01211-8.

Kawaguchi, M., Fujita, H., Yoshizaki, N., Hiroi, J., Okouchi, H., Nagakura, Y., Noda, T., Watanabe, S., Katayama, S., Iwamuro, S., Nishida, M., Iuchi, I., Yasumasu, S., 2009. Different hatching strategies in embryos of two species, pacific herring *Clupea pallasii* and japanese anchovy *Engraulis japonicus*, that belong to the same order Clupeiformes, and their environmental adaptation. J. Exp. Zool. Part B Mol. Dev. Evol. 312, 95–107. http://dx.doi.org/10.1002/jez.b.21247.

Kudo, S., 2000. Enzymes responsible for the bactericidal effect in extracts of vitelline and fertilisation envelopes of rainbow trout eggs. Zygote 8, 257–265.

Lachenbruch, P.A., Mickey, M.R., 1968. Estimation of error rates in discriminant analysis. Technometrics 10, 1–11. http://dx.doi.org/10.1080/00401706.1968.10490530.

Lavoué, S., Miya, M., Saitoh, K., Ishiguro, N.B., Nishida, M., 2007. Phylogenetic relationships among anchovies, sardines, herrings and their relatives (Clupeiformes), inferred from whole mitogenome sequences. Mol. Phylogenet. Evol. 43, 1096–1105. http://dx.doi.org/10.1016/j.ympev.2006.09.018.

Leonori, I., De Felice, A., Campanella, F., Biagiotti, I., Canduci, G., 2011. Assessment of small pelagic fish biomass in the Adriatic Sea by means of acoustic methodology. In: Brugnoli, E., Cavaretta, G., Mazzola, S., Trincardi, F., Ravaioli, M., Santoleri, R. (Eds.), Marine Research at CNR, pp. 2019–2029.

Leonori, I., Ticina, V., De Felice, A., Vidjak, O., Grubišić, L., Pallaoro, A., 2012. Comparisons of two research vessels' properties in the acoustic surveys of small pelagic fish. Acta Adriat. 53, 389–398.

Letunic, I., Doerks, T., Bork, P., 2015. SMART: recent updates, new developments and status in 2015. Nucleic Acids Res. 43, D257–D260. http://dx.doi.org/10.1093/nar/gku949.

Litscher, E.S., Wassarman, P.M., 2007. Egg extracellular coat proteins: from fish to mammals. Histol. Histopathol. 22, 337–347.

Madden, T., 2003. The BLAST sequence analysis tool. In: The NCBI Handbook. Bethesda, pp. 1–18.

Mazzeo, I., Peñaranda, D.S., Gallego, V., Hildahl, J., Nourizadeh-Lillabadi, R., Asturiano, J.F., Pérez, L., Weltzien, F.a., 2012. Variations in the gene expression of zona pellucida proteins, zpb and zpc, in female European eel (*Anguilla anguilla*) during induced sexual maturation. Gen. Comp. Endocrinol. 178, 338–346. http://dx.doi.org/10.1016/j.ygcen.2012.06.003.

MEDIAS, 2012. MEDIAS handbook. Common Protocol for the Pan-MEditerranean Acoustic Survey (MEDIAS), March 2012.

Medits Working Group, 2012. International bottom trawl survey in the Mediterranean Instruction manual Version 6.

Miccoli, A., Gioacchini, G., Maradonna, F., Benato, F., Skobo, T., Carnevali, O., 2015. Beneficial bacteria affect *Danio rerio* development by the modulation of maternal factors involved in autophagic, apoptotic and dorsalizing processes. Cell. Physiol. Biochem. 35, 1706–1718. http://dx.doi.org/10.1159/000373983.

Mold, D.E., Kim, I.F., Tsai, C.M., Lee, D., Chang, C.Y., Huang, R.C.C., 2001. Cluster of genes encoding the major egg envelope protein of zebrafish. Mol. Reprod. Dev. 58, 4–14. http://dx.doi.org/10.1002/1098-2795(200101)58:1<4::AID-MRD2>3.0.CO;2-P.

Montes, I., Conklin, D., Albaina, A., Creer, S., Carvalho, G.R., Santos, M., Estonba, A., 2013. SNP discovery in European anchovy (*Engraulis encrasicolus*, L) by high-throughput transcriptome and genome sequencing. PLoS ONE 8, e70051. http://dx.doi.org/10.1371/journal.pone.0070051.

Murua, H., Saborido-Rey, F., 2003. Female reproductive strategies of marine fish species of the north Atlantic. J. Northwest Atl. Fish. Sci. 33, 23–31. http://dx.doi.org/10.2960/J.v33.a2.

Nelson, J.S., 1994. Fishes of the world. J. Mar. Biol. Assoc. U. K. 74, 987. http://dx.doi.org/10.1017/S0025315400044192.

Oppen-Berntsen, D.O., Gram-Jensen, E., Walther, B.T., 1992a. Zona radiata proteins are synthesized by rainbow trout (*Oncorhynchus mykiss*) hepatocytes in response to oestradiol-17β. J. Endocrinol. 135, 293–302. http://dx.doi.org/10.1677/joe.0.1350293.

Oppen-Berntsen, D.O., Hyllner, S.J., Haux, C., Helvik, J.V., Walther, B.T., 1992b. Eggshell zona radiata-proteins from cod (*Gadus morhua*): extra-ovarian origin and induction by estradiol-17 beta. Int. J. Dev. Biol. 36, 247–254.

Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8, 785–786. http://dx.doi.org/10.1038/nmeth.1701.

Podolsky, R.D., 2002. Fertilization ecology of egg coats: physical versus chemical contributions to fertilization success of free-spawned eggs. J. Exp. Biol. 205, 1657–1668.

Sabatés, A., Zaragoza, N., Grau, C., Salat, J., 2008. Vertical distribution of early developmental stages in two coexisting clupeoid species, *Sardinella aurita* and *Engraulis encrasicolus*. Mar. Ecol. Prog. Ser. 364, 169–180. http://dx.doi.org/10.3354/meps07461.

Saitoh, K., Miya, M., Inoue, J.G., Ishiguro, N.B., Nishida, M., 2003. Mitochondrial genomics of ostariophysan fishes: perspectives on phylogeny and biogeography. J. Mol. Evol. 56, 464–472. http://dx.doi.org/10.1007/s00239-002-2417-y.

Sano, K., Kawaguchi, M., Watanabe, S., Nagakura, Y., Hiraki, T., Yasumasu, S., 2013. Inferring the evolution of teleostean zp genes based on their sites of expression. J. Exp. Zool. Part B Mol. Dev. Evol. 320, 332–343. http://dx.doi.org/10.1002/jez.b.22507.

Sano, K., Kawaguchi, M., Yoshikawa, M., Iuchi, I., Yasumasu, S., 2010. Evolution of the teleostean zona pellucida gene inferred from the egg envelope protein genes of the Japanese eel, *Anguilla japonica*. FEBS J. 277, 4674–4684. http://dx.doi.org/10.1111/j.1742-4658.2010.07874.x.

Schismenou, E., Somarakis, S., Thorsen, A., Kjesbu, O.S., 2012. Dynamics of de novo vitellogenesis in fish with indeterminate fecundity: an application of oocyte packing density theory to European anchovy, *Engraulis encrasicolus*. Mar. Biol. 159, 757–768. http://dx.doi.org/10.1007/s00227-011-1852-y.

Schultz, J., Milpetz, F., Bork, P., Ponting, C.P., 1998. SMART, a simple modular architecture research tool: identification of signaling domains. Proc. Natl. Acad. Sci. USA 95, 5857–5864.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539. http://dx.doi.org/10.1038/msb.2011.75.

Smith, J., Paton, I.R., Hughes, D.C., Burt, D.W., 2005. Isolation and mapping the chicken zona pellucida genes: an insight into the evolution of orthologous genes in different species. Mol. Reprod. Dev. 70, 133–145. http://dx.doi.org/10.1002/mrd.20197.

Spargo, S.C., Hope, R.M., 2003. Evolution and nomenclature of the zona pellucida gene family. Biol. Reprod. 68, 358–362. http://dx.doi.org/10.1095/biolreprod.102.008086.

Suemori, S., Lynch-Devaney, K., Podolsky, D.K., 1991. Identification and characterization of rat intestinal trefoil factor: tissue- and cell-specific member of the trefoil protein family. Proc. Natl. Acad. Sci. 88, 11017–11021. http://dx.doi.org/10.1073/pnas.88.24.11017.

Sugiyama, H., Iuchi, I., 2000. Molecular structure and hardening of egg envelope in fish. Recent Res. Dev. Comp. Biochem. Physiol. 1, 139–161.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729. http://dx.doi.org/10.1093/molbev/mst197.

Turner, L.M., Hoekstra, H.E., 2008. Causes and consequences of the evolution of reproductive proteins. Int. J. Dev. Biol.

Whelan, S., Goldman, N., 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol. Biol. Evol. 18, 691–699. http://dx.doi.org/10.1093/oxfordjournals.molbev.a003851.

Yamagami, K., Hamazaki, T.S., Yasumasut, S., Masuda, K., Luchi, I., 1992. Molecular and cellular basis of formation, hardening, and breakdown of the egg envelope in fish. Int. Rev. Cytol. 136, 51–92. http://dx.doi.org/10.1016/S0074-7696(08)62050-1.

Yang, J.C., Hedrick, J.L., 1997. CDNA cloning and sequence analysis of the *Xenopus laevis* egg envelope glycoprotein gp43. Dev. Growth Differ. 39, 457–467. http://dx.doi.org/10.1046/j.1440-169X.1997.t01-3-00007.x.

Yue, H.-M., Cao, H., Chen, X.-H., Ye, H., Li, C.-J., Du, H., 2014. Molecular characterization of the cDNAs of two zona pellucida genes in the Chinese sturgeon, *Acipenser sinensis* Gray, 1835. J. Appl. Ichthyol. 30, 1273–1281. http://dx.doi.org/10.1111/jai.12586.