

RESEARCH ARTICLE

A COI Nonsynonymous Mutation as Diagnostic Tool for Intraspecific Discrimination in the European Anchovy *Engraulis encrasicolus* (Linnaeus)

Anna Maria Pappalardo, Concetta Federico, Giorgio Sabella, Salvatore Saccone, Venera Ferrito*

Department of Biological, Geological and Environmental Sciences, Sec. Animal Biology "M. La Greca", University of Catania, Catania, Italy

* vferrito@unict.it



OPEN ACCESS

Citation: Pappalardo AM, Federico C, Sabella G, Saccone S, Ferrito V (2015) A COI Nonsynonymous Mutation as Diagnostic Tool for Intraspecific Discrimination in the European Anchovy *Engraulis encrasicolus* (Linnaeus). PLoS ONE 10(11): e0143297. doi:10.1371/journal.pone.0143297

Editor: Yong-Gang Yao, Kunming Institute of Zoology, Chinese Academy of Sciences, CHINA

Received: April 24, 2015

Accepted: November 3, 2015

Published: November 24, 2015

Copyright: © 2015 Pappalardo et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

The European anchovy, *Engraulis encrasicolus*, is currently one of the principal target species for commercial fisheries in Europe. In this study, the mitochondrial Control Region (CR) and the Cytochrome Oxidase I (COI) mitochondrial gene were analyzed in 74 specimens of *E. encrasicolus* from four localities in the central Mediterranean. In both populations, the two markers revealed the presence of two main haplogroups, A and B, already detected in previous investigations of different classes of molecular markers. Both CR and COI markers consistently identified two haplogroups. The COI sequence analysis identified a non-synonymous transversion (T to G) at position 116 of the translated sequence, resulting in an amino acid change. All COI sequences of haplogroup A had an amino acid sequence with alanine in this position, while serine was present in the same position in haplogroup B. The two haplogroups A and B were also discriminated by the variable number of TACA elements at the 5'-end of the mitochondrial CR. The selection tests applied to the COI dataset revealed that codon 116 was not under positive selection, that seven amino acid changes were under purifying selection, and that two amino acids were under episodic positive selection.

Introduction

The European anchovy, *Engraulis encrasicolus* (Linnaeus, 1758), is a clupeoid pelagic species widely distributed in the Mediterranean, Black and Azov Seas, Eastern Atlantic coastline from Norway to Angola [1, 2], and around the tip of southern Africa [3, 4]. Its biology has received particular attention due to the commercial interest of its fisheries [5 – 8]. In the last twenty years a number of scientific studies have focused on the detection of genetic population structure in this commercially important species, through morphological and molecular analyses. Molecular studies described a complex genetic structure of *E. encrasicolus* using allozymes,

mitochondrial DNA (mtDNA) and nuclear markers [9–14]. In the Atlantic Ocean and the Mediterranean Basin, ten genetically differentiated European anchovy populations have been identified by surveys of the variability of the mitochondrial control region with implications for the management of fishery stocks [15, 16]. In particular, mtDNA, SNPs and nuclear intron markers have all supported the presence of two sympatric clades, named phylads or haplogroups A and B. The distribution of these haplogroups differs among natural populations but shows a constant proportion in each population, over time [15, 17–19]. Clade B prevails at the northern and southern high latitudes with frequencies decreasing towards the tropics, whereas clade A is present with higher frequencies in the tropical and subtropical areas [19]. Studies by Magoulas et al. [17, 18] and Grant [4] pointed out the role of climate change during Pleistocene glaciations, suggesting that the two phylads evolved in isolation in an Atlantic refugium and in the Mediterranean, respectively. It has also been suggested that the two mtDNA clades can also occur in sympatry, and that temperature seems to have shaped their relative frequency, as a result of range expansion and post-glacial secondary mixing [18, 19]. European anchovy population dynamics and distribution patterns are known to be strongly dependent on the environment. Thus, this species has been considered an ideal organism to study both the adaptive behaviour of small pelagic fish under different environmental conditions [19], and to understand the effect of quaternary climatic fluctuations on the distribution of marine organisms [14]. More specifically, studies by Silva et al. [14] demonstrated that the high mobile nature of anchovies allowed them to escape the adverse temperatures during the last glacial maximum without losing genetic diversity. Further, Silva et al. [19] detected positive selection in a single codon of the mitochondrial cytochrome b (cytb) gene in clade B, geographically correlated with environmental temperature. In the Mediterranean basin, within a latitudinal range of about 10° (45°–35°N) the annual average sea surface temperature goes from 15°C, in the northern Adriatic to 20°C in the Strait of Sicily or Sicilian Channel between the Sicilian and the North African coasts [20]. Moreover, the main water masses present in this channel, connecting the western and the eastern basins of the Mediterranean Sea, are characterized by a complex multi-scale thermohaline circulation, driven by ocean currents, wind effects, and mesoscale activity. All these environmental features are known to modify the temperature regime of the surface waters in the Sicilian anchovy habitat [21, 22]. Although the Sicilian Channel can be considered a key zone in the Mediterranean basin for the European anchovy population structure, few genetic studies [23] have been carried out in this area that was identified by Garcia Lafuente et al. [24] as one of the main spawning area of *E. encrasicolus*. Based on the considerations above, in this study the genetic structure of four populations of European anchovy from the Tyrrhenian, Ionian and Adriatic seas and the Sicilian Channel was investigated through the sequence analysis of two mitochondrial markers already used to detect A and B haplogroups in *E. encrasicolus* [16, 25], namely the noncoding Control Region (CR) and the coding Cytochrome Oxidase I (COI). The CR has been often used to infer intraspecific genetic structure and population demographic history [26–32]. Partial sequences of the mitochondrial Cytochrome c Oxidase I (COI) gene, a highly conserved, bioenergetic gene encoding for protein subunits of the respiratory chain [33], have been used to investigate the phylogeny of several animal taxa, including fishes, and as barcode sequences [34–39]. The aims are: i) to explore the nucleotide sequences of the two mitochondrial markers to unveil the molecular traits discriminating the two haplogroups of *E. encrasicolus*; ii) starting from the main conclusion of Silva et al. [19], and considering that a thermal cline is present between the Adriatic and the Sicilian Channel [20], to explore whether selection acted on the COI gene, using tests of recombination and selection based on different models of evolution.

Materials and Methods

Sampling, DNA extraction and PCR amplification

A total of 74 adult individuals of morphologically identified *E. encrasicolus* (deposited as vouchers at the Department of Biological, Geological and Environmental Science, Section of Animal Biology, in Catania, Italy) from four populations were collected in the central Mediterranean from 2012 to 2013 (Table 1).

Fishes were obtained from artisanal fish landing sites, at night. Fish samples were collected off Porticello (38°05'05"N-13°32'47"E, Tyrrhenian Sea), Riposto (37°43'10"N-15°15'36"E, Ionian Sea), Porto Empedocle (37°14'59"N-13°31'02"E, Sicilian Channel) and Ancona (43°46'47"N-14°09'37"E, Adriatic Sea). Fishes were transferred to 95% ethanol before further processing in the laboratory. No experimentation with animals was performed. No other ethical issues applied to the present research. Total genomic DNA was extracted from muscle tissue (25–30 mg) using the DNeasy tissue kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. All PCR amplifications were carried out in 25 µl using approximately 50 ng of the isolated DNA as a template. In addition, each PCR reaction contained 16Taq DNA polymerase buffer (supplied by the respective Taq DNA polymerase manufacturer), 1.5–2 mM of MgCl₂, 200 mM of each dNTP, 10 pmols of each primer and 0.5 U of Taq DNA polymerase (Platinum Taq DNA polymerase, Invitrogen). Amplification and sequencing of the mtDNA COI gene was performed using the primers FishF1 and FishR1 described in Ward et al. [38]. Thermal cycles involved an initial denaturing step of 2 min at 94°C, followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 52°C for 45 s and extension at 72°C for 1 min. Negative controls were included in all PCR runs to control for cross-contamination. Double-stranded products were checked by agarose gel electrophoresis, purified with the Qiaquick PCR purification kit (Qiagen) and subsequently sequenced in the forward and reverse directions. For PCR amplification of mitochondrial CR, two primers (ENG_FW: 5' – TGTA AAC GACGGCCAGTTTCTAAAGTTAAACTACCCTCT; ENG_REV1: 5' – CAGGAAACAGCTATGAC TTAAGTGAACGCTCGGCATGG–3') specific to European anchovy were designed from the species' complete mtDNA sequence.

The PCR conditions were as follows: initial denaturation at 94°C for 5 min, followed by denaturation at 94°C for 45 sec, annealing at 55°C (45 sec) and extension at 72°C (1 min) repeated for 35 cycles and with a final extension step at 72°C for 5 min. Negative controls were included in all PCR runs to control for cross-contamination. Double-stranded products were checked by agarose gel electrophoresis and purified with the Qiaquick PCR purification kit (Qiagen). PCR products were subsequently sequenced in the forward and reverse direction by Genechron (<http://www.genechron.it/index.php/sanger-sequencing>) using an ABI Prism 3100 automated sequencer (Applied Biosystems) and M13 forward and M13 reverse primers for sequencing.

Data analysis

The chromatograms obtained were assembled and checked by eye. Edited sequences were aligned using the default settings in ClustalX 2.0 software [40] and the alignment was manually revised in BioEdit (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>). The number of polymorphic sites and genetic diversity indexes (number of haplotypes, haplotype diversity and nucleotide diversity) were calculated using DNAsp 5.1 [41] (Table 1). An analysis of molecular variance (AMOVA), implemented in ARLEQUIN version 3.5.2.1 [42] was applied to the distance matrix to estimate variance components and Φ_{st} values. A haplotype network was built using the median joining (MJ) algorithm [43] in the program Network 4.6.1.3. (Fluxus-

Table 1. Diversity measures for the collecting sites of *Engraulis encrasicolus* for CR and COI: number of sequences (*n*), number of haplotypes (*N_h*), haplotype diversity (*h*), nucleotide diversity (π) and relative standard deviation (SD).

Location	Code	Date of collection	<i>n</i>	Control region			Cytochrome Oxidase I		
				<i>N_h</i>	<i>h</i> ± SD	π ± SD (10 ⁻²)	<i>N_h</i>	<i>h</i> ± SD	π ± SD (10 ⁻²)
Sicilian Channel	SIC	October 2012	18	11	0.941 (0.033)	0.025 (0.002)	12	0.948 (0.033)	0.012 (0.002)
Ionian Sea	ION	September 2013	20	13	0.947 (0.032)	0.029 (0.002)	10	0.837 (0.076)	0.012 (0.001)
Adriatic Sea	ADR	October 2013	18	11	0.908 (0.051)	0.023 (0.004)	11	0.817 (0.095)	0.008 (0.002)
Tyrrhenian Sea	TYR	March 2014	18	9	0.908 (0.039)	0.029 (0.003)	10	0.941 (0.029)	0.009 (0.001)
				overall 0.98 (0.005)			overall 0.93 (0.024)		

doi:10.1371/journal.pone.0143297.t001

engineering.com). Nucleotide diagnostics (NDs) [44, 45] for each European anchovy haplogroup were detected after the editing of sequence alignments using CLC Sequence Viewer 7.6 (www.clcbio.com). To correctly assign each sequence to haplogroup A or B, we used the mtDNA CR sequences of the population in the Canary Islands where 100% of the individuals belong to clade A, as a reference (GenBank accession numbers: HQ215641, JQ595031, JQ595040, JQ595059, JQ595109, JQ595102, JQ595187-JQ595216) [16]. After alignment, the mtDNA CR sequences that clustered with those of the Canary Islands were considered to belong to clade A.

Tests of recombination and selection

To test the presence of mitochondrial COI recombinants, we used the Genetic Algorithms for Recombination Detection (GARD) [46] (HyPhy package, accessed at www.datamonkey.org). To test the presence of selection on COI, the Z-test [47] was performed in MEGA v.6 [48]. To determine the site-specific selection pressures acting on the COI gene, the single-likelihood ancestor counting (SLAC), fixed-effects likelihood (FEL), internal FEL (IFEL), fast unconstrained Bayesian approximation (FUBAR) and mixed effects model of evolution (MEME) methods were applied. Sites with *p*-values less than 0.05 for SLAC, FEL, IFEL and MEME, as well as sites with posterior probability of more than 0.9 for FUBAR were all considered as being under selection [19]. To visualize the structural position of positively selected sites, the secondary structure of the barcode region was predicted using the PSIPRED method [49, 50]. The 3-dimensional structure was drawn with Jmol (accessed at <http://www.jmol.org>).

The 217 COI amino acid sequence examined in this study is homologous to the 18–234 fragment of the full bovine COI protein sequence (<http://c.expasy.org/uniprot/P00396#seq>).

Results

Mitochondrial CR analysis and NDs

For the CR, after alignment, a 577 bp fragment was analyzed. We obtained 74 sequences defining 44 haplotypes (S1 File), with a total of 76 polymorphic sites including 18 singletons and 58 parsimoniously informative sites. No shared haplotypes were found among populations. High values of haplotype and nucleotide diversity indices were found in Ionian Sea and Sicilian Channel populations (Table 1). AMOVA revealed overall low but statistically significant genetic structuring of the analyzed samples ($\Phi_{st} = 0.11$ $p < 0.0001$ +/- 0.0000). The majority of the variance was distributed within (89.50%) and not among (10.50%) samples (S2 File). Overall haplotype diversity of mitochondrial CR (0.98) was much higher than COI (Table 1). The MJ network showed the clear splitting of the control region haplotypes into two groups including samples from all populations (Fig 1).

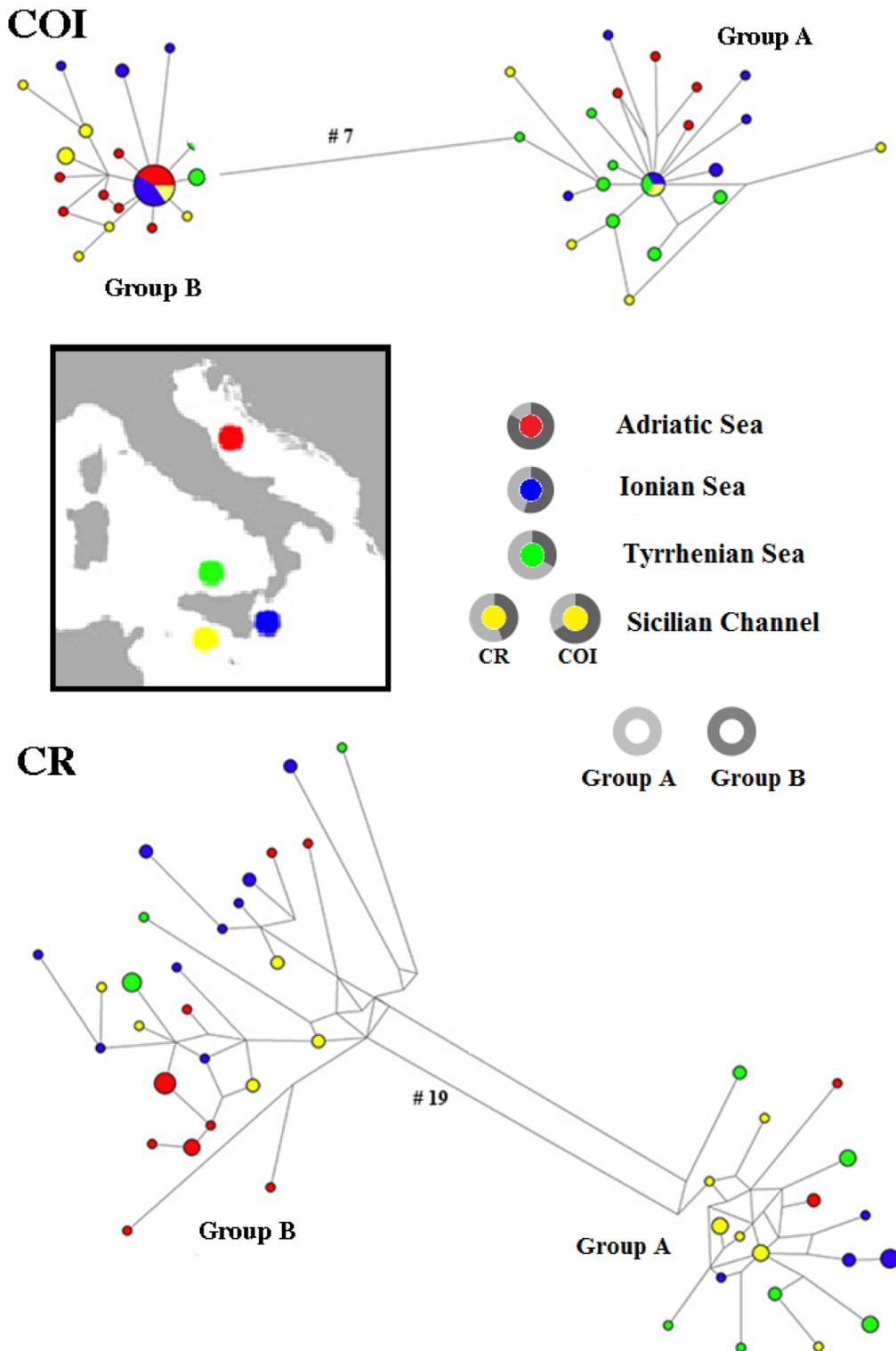


Fig 1. Median-joining network for the COI and CR of *Engraulis encrasicolus*. The area of the circles is proportional to each haplotype frequency. Colors refer to the region in which haplotypes were found. In the case where haplotypes are shared among regions, pie charts show the relative haplotype frequency in each region. Grey circles represent mtDNA CR and COI clades A (light grey) and B (dark grey) proportions. Only in the Sicilian Channel the relative proportion of the two clades is different for the two markers.

doi:10.1371/journal.pone.0143297.g001

The inclusion of mtDNA *CR* sequences of the Canary Islands [16] in the dataset (not shown) allowed the identification of haplogroup A. Notably, four or five 5'-TACA-3' repeats were found at the 5' end of the *CR*, at positions 82, 120, 124, 140 and 160. More specifically, no TACA repeats at position 124 were found in haplotypes of haplogroup A, while 65% of haplotypes in haplogroup B had TACA repeats in this position. The TACA at position 140 was not found in 96% of the haplotype sequences in haplogroup B, while all haplotypes in haplogroup A had TACA repeats in this position. A total of four NDs, including three transversions (positions: 37, 222 and 467) and one transition at position 333, discriminated haplogroups A and B (S2 File).

COI sequence analysis, NDs

Unambiguously aligned sequences were obtained for 651 bp of *COI* sequence from 74 tissue samples of *E. encrasicolus*. All sequences were deposited in GenBank under the accession numbers reported in S1 File. No insertions, deletions or stop codons were observed. The lack of stop codons is consistent with all amplified sequences being functional mitochondrial *COI* sequences, along with the fact that all amplified sequences were of the same length (651 bp). This suggests that NUMTs (nuclear DNA sequences originating from mitochondrial DNA sequences) were not sequenced (vertebrate NUMTs are generally smaller than 600 bp) [51]. A total of 57 variable nucleotide sites with 28 parsimony-informative sites defined 39 distinct anchovy haplotypes (S1 File). Two haplotypes were found in different sampled regions: H5 in the Adriatic Sea, Sicilian Channel and Ionian Sea; and H18 in the Tyrrhenian Sea, Sicilian Channel and Ionian Sea. High values of haplotype diversity were found in the Sicilian Channel and Tyrrhenian Sea samples (Table 1). AMOVA revealed overall low but statistically significant genetic structuring of the analyzed samples ($\Phi_{st} = 0.11$ $p < 0.007$ +/- 0.0023). The majority of the variance was distributed within (88.88%) and not among (11.12%) samples (S3 File). The MJ network showed the clear splitting of the *COI* haplotypes into two groups (A and B) including samples from all populations (Fig 1).

A total of six nucleotide diagnostics (NDs) for haplogroup discrimination of *E. encrasicolus* (group A and group B) are identified in the S4 File. Four of these are transitions (positions: 344, 358, 421 and 454) and two are transversions (positions: 347 and 367). All substitutions are silent with the exception of the G to T transversion at position 347, which changes the amino acid from alanine (A) to serine (S) at position 116 in 58% of the sequences (Fig 2).

All amino acid sequences of haplogroup A have an alanine at position 116, and the serine is found in the same position in all sequences of haplogroup B (S5 File).

Tests of recombination and selection

No evidence for recombination was found in the *COI* dataset with GARD. The Z-test rejected the null hypothesis of strict neutrality ($d_{NON-SYNONYMOUS} = d_{SYNONYMOUS}$; $p = 0,000$) in favor of the alternative hypothesis of positive selection ($d_N > d_S$) in clade B. Seven amino acids were under purifying selection and two under positive selection (codons 11 and 121). Episodic positive selection was detected with the MEME method for codon 11 and 121 in clade B (Table 2) and with MEME and IFEL methods for codon 121 in clade A.

In haplogroup A histidine replaced methionine at codon 11 and proline replaced histidine at codon 121. The amino acids in the "Folmer region" (the standard barcode region) form five helices (<http://c.expasy.org/uniprot/P00396#seq>). Based on the prediction of the protein secondary structure, codon 116 is located in the third coil toward the inter-membrane space. Codons 11 and 121 under episodic positive selection are located in the first helix and in the third coil, respectively (Fig 3).

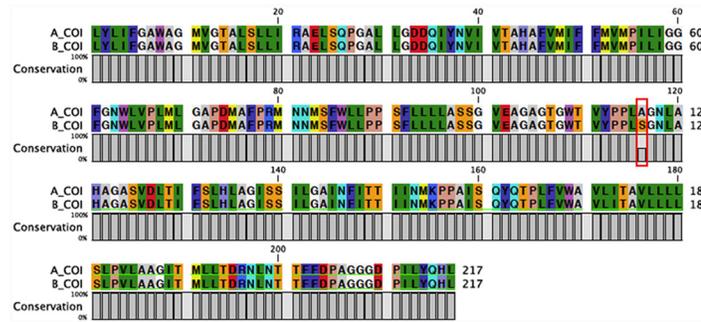


Fig 2. COI amino acid aligned sequences of *Engraulis encrasicolus*. The red box indicates the amino acid change. Serine (S) is found in haplogroup B and alanine (A) in haplogroup A.

doi:10.1371/journal.pone.0143297.g002

Comparison between CR and COI results

The two divergent haplogroups detected with both mitochondrial markers were separated by 7 point mutations in the COI network and by 19 point mutations in the CR network. The percentage of divergence between the two clades was 4.4% for CR and 2% for COI. Substantial agreement exists between CR and COI in the assignment of each individual to the same haplogroup for all populations except for Sicilian Channel (Fig 1). Overall, 46% of the CR haplotypes and 42% of the COI haplotypes were included in clade A. More specifically, clade B prevails in the Adriatic and in the Ionian samples with both markers.

Table 2. Positively and negatively selected sites in the cytochrome oxidase I gene estimated by FUBAR, SLAC, IFEL, FEL and MEME models.

Codon Selection type	11 positive	42 purifying	82 purifying	104 purifying	113 purifying	121 positive	150 purifying	155 purifying	211 purifying
all	FUBAR	##		#	#		##		##
	SLAC								*
	IFEL								
	FEL		*	*		*		*	*
	MEME	***					**		
Clade A	FUBAR	##				##			##
	SLAC								
	IFEL						*		
	FEL					*			*
	MEME						**		
Clade B	FUBAR	##		#	#		##		##
	SLAC								
	IFEL								
	FEL		*	*		*		*	*
	MEME	***					**		

#bpp ≥0.9;
 ##bpp ≥0.95;
 ###bpp ≥0.99
 *p < 0.05;
 ** p <0.01;
 ***p <0.001

doi:10.1371/journal.pone.0143297.t002

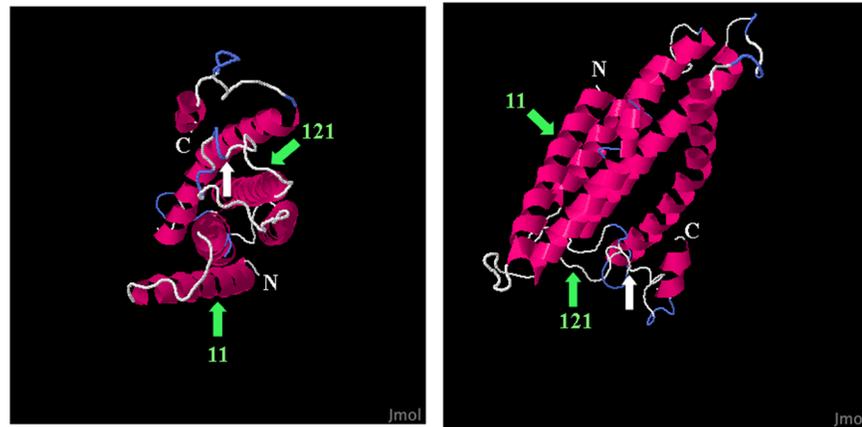


Fig 3. 3-dimensional structure of the *COI* barcode region and structural location of positively selected sites (green arrows). On the right, lateral view with the amino terminal tail (N) facing towards the mitochondrial matrix surface and the carboxy terminal tail (C) facing towards the inter-membrane space surface. On the left, a top view. White arrow point to the neutral non-synonymous change at position 116 allowing for haplogroup discrimination.

doi:10.1371/journal.pone.0143297.g003

Discussion

Both mitochondrial molecular markers identified two main haplogroups which were identified as the haplogroups A and B, by comparison with some mtDNA CR sequences from the Canary Islands [16]. Both revealed low but significant genetic structuring with Φ_{st} values congruent with those detected in other studies [18]. The average *COI* haplotype diversity ($h = 0.93$) is lower than for *CR* ($h = 0.98$). These results indicate the effectiveness of *COI* sequence variation in detecting genetic structure in *E. encrasicolus* and other teleosts such as swordfish and lanternfish species [52, 53]. The distribution of TACA repeats in the *CR* allowed to discriminate the two haplogroups of European anchovy; a similar result was obtained in *Xiphias gladius* where TACA repeats discriminated geographical stocks of this species [52]. A relevant result was the detection of a non-synonymous substitution at the site 347 of the *COI* sequence. It should be noted that the substitution involved an aliphatic and apolar amino acid (alanine) and a hydroxylated and polar amino acid (serine), which are functionally different. This is contrary to the evidence that substitutions in the *COI* amino acid sequence are known to be frequent among physicochemically similar amino acids [54]. The relevance of this finding should be assessed given that the 655 bp region of *COI* codes for 218 amino acids and tends to be strongly conserved due to functional constraints on amino acid substitutions. Studies by Ward and Holmes [54] on the nucleotide and amino acid variability of *COI* in fishes demonstrated that the ratio of non-synonymous to synonymous substitutions is much lower than one, indicating that this gene is subject to strong purifying selection. In exploring the role of the mutation and selection in mitochondrial protein-coding genes, Castellana et al. [33] noted the strong influence of purifying selection on all mitochondrial genes and especially on *COI*, *COII*, *COIII* and *Cytb*, with crucial functions in the respiratory chain and thus strongly preserved. However, the frequently discussed role of positive selection in thermal adaptation and aerobic performance has also been demonstrated for these genes in invertebrates and vertebrates [55–57], making the implicit assumption of neutrality of mitochondrial markers no longer valid [58]. The European anchovy has been studied in this respect by Silva et al. [19], which detected positive selection in a single codon of the mitochondrial *Cytb* gene in the clade B strongly correlated with the temperature. The authors discuss the functional role of the non-synonymous

substitution that they identified, suggesting that it may have an impact on the overall metabolic performance of clade B. The tests of selection applied to our *COI* dataset indicated positive selection only for clade B through the Z test. Seven codons were found to be under purifying selection and two (codons 11 and 121) under episodic positive selection. However, codon 116, located near codon 121 on the third coil, was not among the codons under selection. Codons under episodic positive selection experience purifying selection for the majority of their evolution, interspersed with bursts of positive selection that may occur only in restricted lineages [59]. Mutations at such sites may experience transient positive selection, followed by purifying selection to maintain the change, and could play a key role in adaptive evolution [59, 60]. Molecular footprints of episodic positive selection have been underestimated until now and the MEME method shows a great power to unveil sites under this model of selection. Tomasco & Lessa [57] found two sites in both *COII* and *Cytb*, using MEME, showing footprints of episodic positive selection, likely functionally relevant.

We have not evaluated the functional implications of the amino acid substitutions found in the *COI* gene, all involving residues with dissimilar properties. However, amino acid substitutions not involved regions known to have highly conserved, important functions (e.g. d-pathway or proton pathway or binding the heme or cytochrome c interactions) [54]. We could only argue that subtle modifications may alter interactions between proteins and their stability, modifying the performance of oxidative phosphorylation processes under different environmental selective pressures [19, 56, 61]. The above results are consistent with the main conclusions of previous investigations aimed to demonstrate the high responsiveness of European anchovy to the changing environment [14, 19] and the strong influence of contemporary and historical processes in shaping the distribution and the genetic structure of this small pelagic species [18].

Our sequence analyses revealed the molecular traits of the mitochondrial *CR* and *COI* fragments useful to discriminate haplogroups of *E. encrasicolus*. The application of the MEME method unveiled that two sites of *COI* were under episodic positive selection. These results add to the emerging data from the most recent surveys on adaptive selection of mitochondrial DNA, suggesting that episodic positive selection affects a large proportion of sites, also in a highly conserved gene such as *COI*.

Supporting Information

S1 File. List of *CR* and *COI* haplotypes from *E. encrasicolus* samples.

(XLSX)

S2 File. Variable nucleotide sites in the 575–577 bp sequences of the European anchovy *CR* examined in this work. Marked haplotypes belong to haplogroup A.

(PDF)

S3 File. Analysis of Molecular Variance (AMOVA) calculated for the populations examined in this work with Arlequin ver 3.5.2.1.

(XLSX)

S4 File. Variable nucleotide sites in the 651 bp sequences of the European anchovy *COI* gene examined in this work. Marked haplotypes belong to haplogroup A. The non-synonymous mutation highlighted in red is diagnostic for discrimination of A and B haplogroups.

(PDF)

S5 File. Variable codon sites in the 217 amino acids of the European anchovy *COI* gene examined in this work. The codon 116 highlighted in red is diagnostic for discrimination of

A and B haplogroups. The codons 11 and 121 highlighted in grey are under episodic positive selection.

(PDF)

Acknowledgments

We are grateful to the referees for useful suggestions and we thank Rita Castilho for analytical advice and comments. We are also grateful to prof. Adriana Canapa (Università Politecnica delle Marche, Dip. di Scienze della vita e dell'ambiente) who provided Adriatic samples and doctor Angelo Sardella who provided samples from Tyrrhenian Sea.

Author Contributions

Conceived and designed the experiments: AMP VF. Performed the experiments: AMP CF VF. Analyzed the data: AMP VF. Contributed reagents/materials/analysis tools: VF GS SS. Wrote the paper: AMP VF.

References

1. Whitehead PJP, Nelson G, Wongratana T. Clupeoid Fishes of the World (Suborder Clupeoidei). FAO Fisheries Synopsis No. 125, Vol. 7. Part 2 –Engraulidae. FAO, Rome; 1988.
2. Whitehead PJP. Engraulidae. In: Quero J.-C., Karrer C., Post A., Saldanha L. (Eds.), Checklist of the Fishes of the Eastern Tropical Atlantic (CLOFETA), vol. 1 UNESCO, Paris; 1990. pp. 228–229
3. Grant WS, Bowen BW. Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. *J Hered.* 1998; 89: 415–426.
4. Grant WS. A second look at mitochondrial DNA variability in European anchovy (*Engraulis encrasicolus*): assessing models of population structure and the black sea isolation hypothesis. *Genetica.* 2005; 125: 293–309. PMID: [16247701](#)
5. Tudela S, Palomera I, Quílez G. Feeding of anchovy *Engraulis encrasicolus* larvae in the north-west Mediterranean. *J Mar Biol Ass UK.* 2002; 82: 349–350.
6. Basilone G, Guisande C, Patti B, Mazzola S, Cuttitta A, Bonanno A, et al. Effect of habitat conditions on reproduction of the European anchovy (*Engraulis encrasicolus*) in the Strait of Sicily. *Fish Oceanogr.* 2006; 15: 271–280.
7. Raab K, Llope M, Nagelkerke LAJ, Rijnsdorp AD, Teal LR, Licandro P, et al. Influence of temperature and food availability on juvenile European anchovy *Engraulis encrasicolus* at its northern boundary. *Mar Ecol Prog Ser.* 2013; 488: 233–245.
8. Mhlongo N, Yemane D, Hendricks M, Van der Lingen CD. Have the spawning habitat preferences of anchovy (*Engraulis encrasicolus*) and sardine (*Sardinops sagax*) in the southern Benguela changed in recent years? *Fish Oceanogr.* 2015; 24 (Suppl. 1): 1–14.
9. Ivanova PP, Dobrovolov IS. Population genetic structure on European anchovy (*Engraulis encrasicolus*, Linnaeus, 1758) (Osteichthyes: Engraulidae) from Mediterranean Basin and Atlantic Ocean. *Acta Adriat.* 2006; 47(1): 13–22.
10. Kristoffersen JB, Magoulas A. Population structure of anchovy *Engraulis encrasicolus* L. in the Mediterranean Sea inferred from multiple methods. *Fish Res.* 2008; 91: 187–195.
11. Sanz N, García-Marín JL, Viñas J, Roldan M, Pla C. Spawning groups of European anchovy: population structure and management implications. *ICES J Mar Sci.* 2008; 65: 1635–1644.
12. Erdoğan Z, Turan C, Koç HT. Morphologic and Allozyme Analyses of European anchovy (*Engraulis encrasicolus* (L. 1758)) in the Black, Marmara and Aegean Seas. *Acta Adriat.* 2009; 50 (1): 77–90.
13. Borrell YJ, Piñera JA, Prado Sanchez, Blanco G. Mitochondrial DNA and microsatellite genetic differentiation in the European anchovy (*Engraulis encrasicolus*). *ICES J Mar Sci.* 2012; 69: 1357–13.
14. Silva G, Horne JB, Castilho R. Anchovies go north and west without losing diversity: post-glacial range expansions in a small pelagic fish. *J Biogeogr.* 2014; 41, 1171–1182.
15. Zarranaindia I, Iriondo M, Albaina A, Padro Miguel A, Manzano C, Grant WS, et al. Multiple SNP markers reveal fine-scale population and deep phylogeographic structure in european anchovy (*Engraulis encrasicolus* L.). *PLoS ONE.* 2012; 7: e42201. doi: [10.1371/journal.pone.0042201](#) PMID: [22860082](#)

16. Viñas J, Sanz N, Peñarrubia L, Araguas RM, Garcia-Marin JL, Roldan MI, et al. Genetic population structure of European anchovy in the Mediterranean Sea and the Northeast Atlantic Ocean using sequence analysis of the mitochondrial DNA control region. *ICES J Mar Sci.* 2014; 71(2), 391–397.
17. Magoulas A, Tsimenides N, Zouros E. Mitochondrial DNA phylogeny and the reconstruction of the population history of a species: the case of the European anchovy (*Engraulis encrasicolus*). *Mol Biol Evol.* 1996; 13: 178–190. PMID: [8583890](#)
18. Magoulas A, Castilho R, Caetano S, Marcato S, Patarnello T. Mitochondrial DNA reveals a mosaic pattern of phylogeographical structure in Atlantic and Mediterranean populations of anchovy (*Engraulis encrasicolus*). *Mol Phyl Evol.* 2006; 39: 734–746.
19. Silva G, Lima FP, Martel P, Castilho R. Thermal adaptation and clinal mitochondrial DNA variation of European anchovy. *Proc R Soc B* 2014; 281: 1093.
20. Shaltout M, Omstedt A. Recent sea surface temperature trends and future scenarios for the Mediterranean Sea. *Oceanologia* 2014; 56 (3): 411–443.
21. Bonanno A, Placenti F, Basilone G, Mifsud R, Genovese S, Patti B, et al. Variability of water mass properties in the Strait of Sicily in summer period of 1998–2013. *Ocean Sci.* 2014; 10: 759–770.
22. Falcini F, Palatella L, Cuttitta A, Buongiorno Nardelli B, Lacorata G, Lanotte AS, et al. The role of hydrodynamic processes of anchovy eggs and larvae distribution in the Sicily Channel (Mediterranean Sea): a case study for the 2004 data set. *Plos One* 2015; 10(4): e0123213. doi: [10.1371/journal.pone.0123213](#) PMID: [25915489](#)
23. Cuttitta A, Patti B, Maggio T, Quinci EM, Pappalardo AM, Ferrito V, et al. Larval population structure of *Engraulis encrasicolus* in the Strait of Sicily as revealed by morphometric and genetic analysis. *Fish Oceanogr.* 2015; 24(2): 135–149.
24. Garcia Lafuente J, Garcia A, Mazzola S, Quintanilla L, Delgado A, Cuttitta A, et al. Hydrographic phenomena influencing early life stages of the Sicilian Channel anchovy. *Fish Oceanogr.* 2002; 11: 31–44.
25. Pappalardo AM, Ferrito V. A COI-Bar-RFLP strategy for the rapid detection of *Engraulis encrasicolus* in processed anchovy products. *Food Control* 2015; 57: 385–392.
26. Avise JC. *Phylogeography: The History and Formation of Species.* 2000; Cambridge: Harvard University Press
27. Zardoya R, Castilho R, Grande C, Favre-Krey L, Caetano S, Marcato S, et al. Differential population structuring of two closely related fish species, the mackerel (*Scomber scombrus*) and the chub mackerel (*Scomber japonicus*), in the Mediterranean Sea. *Mol Ecol* 2004; 13: 1785–1798. PMID: [15189203](#)
28. Pappalardo AM, Ferrito V, Messina A, Guarino F, Patarnello T, De Pinto V, et al. Genetic structure of the killifish *Aphanius fasciatus*, Nardo 1827 (Teleostei, Cyprinodontidae) results of mitochondrial DNA analysis. *J Fish Biol* 2008; 72 (Supplement X): 1–20.
29. Robalo JI, Castilho R, Francisco SM, Almada F, Knutsen H, Jorde PE, et al. Northern refugia and recent expansion in the North Sea: the case of the wrasse *Symphodus melops* (Linnaeus, 1758). *Ecol Evol* 2011; 2: 153–64.
30. Ferrito V, Pappalardo AM, Canapa A, Barucca M, Doadrio I, Olmo E, et al. Mitochondrial phylogeography of the killifish *Aphanius fasciatus* (Teleostei, Cyprinodontidae) reveals highly divergent Mediterranean populations. *Mar Biol* 2013; 160 (12): 3193–3208.
31. Fruciano C, Pappalardo AM, Tigano C, Ferrito V. Phylogeographical relationships of Sicilian brown trout and the effect of genetic introgression on morphospace occupation. *Biol J Linn Soc* 2013; 112: 387–398.
32. Pappalardo AM, Gonzalez EG, Tigano C, Doadrio I, Ferrito V. Comparative pattern of genetic structure in two Mediterranean killifishes (*Aphanius fasciatus* and *A. iberus*) inferred from both mitochondrial and nuclear data. *J Fish Biol.* 2015; 87: 69–87. doi: [10.1111/jfb.12693](#) PMID: [26010329](#)
33. Castellana S, Vicario S, Saccone C. Evolutionary pattern of mitochondrial genome in metazoa: exploring the role of mutation and selection in mitochondrial protein-coding genes. *Genome Biol Evol.* 2011; 3: 1067–1079.
34. Hebert PDN, Cywinska A, Ball S, deWaard J. Biological identifications through DNA barcodes. *P Roy Soc Lond B Bio.* 2003; 270: 313–321.
35. Hebert PDN, Ratnasingham S, deWaard J. Barcoding animal life: cytochrome c oxidase subunit 1 divergence, among closely related species. *P Roy Soc Lond B Bio.* 2003; (Suppl:) 270: S96–S99.
36. Hebert PDN, Stoeckle MY, Zemplak TS, Francis CM. Identification of birdsthrough DNA barcodes. *PLOS Biol.* 2004; 2: e312. PMID: [15455034](#)
37. Hajibabaei M, Smith MA, Janzen DH, Rodriguez JJ, Whitfield JB, Hebert PDN. A minimalist barcode can identify a specimen whose DNA is degraded. *Mol Ecol Notes.* 2006; 6: 959–964.

38. Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN. DNA barcoding Australia's fish species. *Phil. Trans. Royal Soc. B-Biol. Sci.* 2005; 360: 1847–1857.
39. Pappalardo A, Ferrito V. DNA barcoding species identification unveils mislabeling of processed flatfish products in southern Italy markets. *Fish Res.* 2015; 164: 153–158.
40. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007; 23: 2947–2948. PMID: [17846036](#)
41. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics.* 2009; 25: 1451–1452. doi: [10.1093/bioinformatics/btp187](#) PMID: [19346325](#)
42. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Res.* 2010; 10: 564–567.
43. Bandelt HJ, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol.* 1999; 16: 37–48. PMID: [10331250](#)
44. Sarkar IN, Thornton JW, Planet PJ, Desalle R. An automated phylogenetic key for classifying homeoboxes. *Mol Phylogenet Evol.* 2002; 24: 388–399. PMID: [12220982](#)
45. Wong EHK, Shivji MS, Hanner RH. Identifying shark with DNA barcode: assessing the utility of a nucleotide diagnostic approach. *Mol Ecol Res.* 2009; 9 (Suppl. 1): 243–256.
46. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SDW. GARD: a genetic algorithm for recombination detection. *Bioinformatics.* 2006; 22: 3096–3098. PMID: [17110367](#)
47. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and non-synonymous nucleotide substitutions. *Mol. Biol. Evol.* 1986; 3: 418–426. PMID: [3444411](#)
48. Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol.* 2013; 30: 2725–2729. doi: [10.1093/molbev/mst197](#) PMID: [24132122](#)
49. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol.* 1999; 292: 195–202. PMID: [10493868](#)
50. Buchan DWA, Minnici F, Nugent TCO, Bryson K, Jones DT. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nuc Acids Res.* 2013; 41(W1): W340–W348.
51. Zhang DX, Hewitt GM. Nuclear integrations: challenges for mitochondrial DNA markers. *Trends Ecol Evol.* 1996; 11: 247–251. PMID: [21237827](#)
52. Pappalardo AM, Guarino F, Reina S, Messina A, De Pinto V. Geographically widespread swordfish barcode stock identification: a case study of its application. *PLoS One.* 2011; 6(10): e25516. doi: [10.1371/journal.pone.0025516](#) PMID: [22039414](#)
53. Pappalardo AM, Cuttitta A, Sardella A, Musco M, Maggio T, Patti B, et al. DNA barcoding and COI sequence variation in Mediterranean lanternfishes larvae. *Hydrobiologia.* 2015; 749: 155–167.
54. Ward R, Holmes BH. An analysis of nucleotide and amino acid variability in the barcode region of cytochrome c oxidase I (coxI) in fishes. *Mol Ecol.* 2007; 7: 899–907.
55. Blier PU, Dufresne F, Burton RS. Natural selection and the evolution of mtDNA-encoded peptides: evidence for intergenomic co-adaptation. *TRENDS in Genetics.* 2001; 17: 400–406. PMID: [11418221](#)
56. Dalziel AC, Moyes CD, Fredriksson E, Loughheed SC. Molecular evolution of Cytochrome c Oxidase in high performance fish (Teleostei: Scombroidei). *J Mol Evol.* 2006; 62: 319–321. PMID: [16477525](#)
57. Tomasco IH, Lessa EP. Two mitochondrial genes under episodic positive selection in subterranean octodontoid rodents. *Gene.* 2014; 534: 371–378. doi: [10.1016/j.gene.2013.09.097](#) PMID: [24113079](#)
58. Rand DM. The units of selection of mitochondrial DNA. *Annu Rev Ecol Syst.* 2001; 32: 425–448.
59. Murrel B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* 2012; 8(7): e1002764. doi: [10.1371/journal.pgen.1002764](#) PMID: [22807683](#)
60. Grueber CE, Wallis GP, Jamieson IG. Episodic positive selection in the evolution of avian toll-like receptor innate immunity genes. *PLoS One.* 2014; 9(3): e89632. doi: [10.1371/journal.pone.0089632](#) PMID: [24595315](#)
61. Grant WS, Spies IB, Canino MF. Biogeographic evidence for selection on mitochondrial DNA in North Pacific walleye Pollock *Theragra chalcogramma*. *J Hered.* 2006; 97 (6): 571–580. PMID: [17038421](#)