Phylogeography of *Ophryotrocha labronica* (Polychaeta, Dorvilleidae) along the Italian coasts

Running head: Phylogeography of *Ophryotrocha labronica*

P. Cossu[1,4,*], **F. Maltagliati**[1], **F.G. Pannacciulli**[2], **R. Simonini**[3], **G. Massamba-N'Siala**[3], **M. Casu**[4], **C. Lardicci**[1], **D. Prevedelli**[3], **A. Castelli**[1]

[1] Dipartimento di Biologia, Via Derna 1, 56126, Pisa, Italy

[2] ENEA - Marine Environment Research Centre, PO Box 224, 19100, La Spezia, Italy

[3] Dipartimento di Scienze della Vita, Via Campi 213/d, 41125, Modena, Italy

[4] Dipartimento di Scienze della Natura e del Territorio, Via Muroni 25, 07100, Sassari, Italy

*Email: picossu@uniss.it

ABSTRACT

Species of the genus *Ophryotrocha* are a well-studied group of organisms but, despite the relatively large body of biological studies, little is known about the intra-specific patterns of genetic diversity. In the present study, we analysed the patterns of genetic variation in samples of *Ophryotrocha labronica* (Polychaeta, Dorvilleidae) collected along the Italian coasts within three regions with different thermal regimes: the Northern Adriatic Sea (NAS), the Ligurian Sea (LS), and the south/south-east Sicilian Sea (SS). A partial sequence of the cytochrome c oxidase subunit I gene (COI) was used as a genetic marker. An analysis of molecular variance (AMOVA) showed significant genetic differentiation between the NAS and the other regions. Conversely, little or no genetic structuring was found between the LS and the SS or among locations within a given region. A Bayesian phylogenetic tree and a Median-joining network provided evidence for the occurrence of two highly divergent genetic lineages characterised by a high average sequence divergence (17.2%, Kimura 2-parameter distance). The spatial patterns of genetic variation found in *O. labronica* may reflect the signature of past expansion events of the two genetic lineages. Though the high sequence divergence suggested that cryptic speciation within *O. labronica* may have occurred, other traits such as the absence of reproductive isolation, the pattern of phenotypic variation, and the habitat specificity prompted us to regard the two groups as distinct COI lineages of *O. labronica*.

KEYWORDS

Ophryotrocha labronica, COI, genetic diversity, mtDNA lineages, Mediterranean Sea

INTRODUCTION

The genus *Ophryotrocha* Claparède and Mecznikow (1869) (Annelida, Polychaeta) is a well-studied group of marine polychaetes inhabiting very diverse environments that range from polluted harbours to deep-sea sediments (see Thornhill *et al.* 2009 for a review). Many shallow-water species of *Ophryotrocha* can be easily reared in the laboratory due to their relatively short generation time and a variety of reproductive modes. These characteristics make the species of this genus ideal systems for a wide range of biological investigations, including ecological, behavioural, reproductive, developmental and toxicological studies (Thornhill *et al.* 2009 and references therein).

However, despite the relatively large body of biological studies on *Ophryotrocha* species, little is known about the intra-specific patterns of genetic diversity, though this information is necessary for understanding evolutionary and ecological processes such as dispersal, local adaptation and, ultimately, speciation. For instance, phenotypic variation along an environmental gradient or across habitats with different ecological conditions may reflect phenotypic plasticity or local adaptation. Investigating the extent to which the genetic and phenotypic variation are related may reveal whether the phenotypic variation reflect local adaptation or phenotypic plasticity (Reusch & Wood 2007). In addition, information on the species' genetic structure and the levels of gene flow among populations may shed light on the species' ability to cope with environmental changes (Zhakarov & Hellman 2008). Molecular tools are useful in delimiting species boundaries or revealing cryptic speciation. In particular, the identification of candidate cryptic species (Alonso *et al.* 2012 and references therein) may be important for the species used as models in different research fields. To tackle these issues, we used genetic

tools to analyse the patterns of intra-specific genetic diversity in one of the most well-known species of this genus, *Ophryotrocha labronica* La Greca & Bacci 1962. The species is actually subdivided into two partially reproductively isolated "sub-species" (Åkesson & Paxton 2005): *Ophryotrocha labronica pacifica* Paxton & Åkesson, 2007, which has been reported only in the Pacific Ocean, and *Ophryotrocha labronica labronica* La Greca & Bacci, 1962, which is typical of the Mediterranean and Lusitanian provinces, with isolated populations found in the Northern Red Sea and Southern Pacific. Here, we focus on *O. labronica labronica* (hereafter referred to as *O. labronica*) collected along the Italian coasts, within three distinct regions characterised by different thermal regimes (Simonini *et al.* 2010; Massamba N'Siala *et al.* 2011): a sub-continental temperate climate in the Northern Adriatic Sea (NAS), warm temperate climate in the Ligurian Sea (LS), and sub-tropical temperate climate in the south/south-east Sicilian Sea (SS).

Ophryotrocha labronica is a small-sized, gonochoristic polychaete (4 mm of maximum length) with semi-continuous reproduction and direct development, typically found in harbours and brackish-water environments (Paxton & Åkesson 2007). Despite its very limited dispersal capabilities, the species displays a worldwide, patchy geographical distribution (Simonini et al. 2009; Prevedelli *et al.* 2005). Long-distance dispersal due to human-mediated transport or other natural vectors (e.g., drifting and/or rafting) may be responsible for this distribution (Åkesson & Paxton 2005), though the overall causative factors are still debated (Simonini *et al.* 2010). *Ophryotrocha labronica* is by far the most common species of this genus along the Italian coasts, where its distribution is not affected by the climate regime and/or the type of fouling community in which it occurs (Simonini *et al.* 2010). Moreover, the species shows a complex pattern of phenotypic variation

that varies at both the local and regional scale (Massamba N'Siala *et al.* 2011). According to this study the heterogeneity of abiotic conditions within harbour habitats may have promoted the differentiation of isolated patches through local adaptation or random genetic drift. On the other hand, the environmental factors acting at a larger spatial scale, such as the thermal regime, may have shaped the observed pattern of phenotypic variation on the regional scale. Indeed, the life-history traits varied among the three regions (Massamba N'Siala *et al.* 2011).

The aim of this study was to analyse the pattern of spatial genetic structure within and among the aforementioned regions. In particular, we aimed to assess whether local adaptation, random genetic drift or population history were the main drivers of the species' genetic structure, if any was discovered. The patterns of genetic variation of *O. labronica* were examined using a partial sequence of the mitochondrial cytochrome c oxidase subunit I (COI) gene. This marker has proven effective in elucidating the phylogeographical structure and the underlying evolutionary history in polychaetes (e.g., Virgilio *et al.* 2009), as well as highlighting cryptic speciation (e.g., Audzjionyte *et al.* 2008; Carr *et al.* 2011).

MATERIALS AND METHODS

Sampling design

Samples of *Ophryotrocha labronica* were collected along the Italian coast following the procedures reported in Prevedelli *et al.* (2005). We adopted the sampling design described in Massamba N'Siala *et al.* (2011); four harbours within each region, and two locations within each harbour to check for small- scale genetic heterogeneity. However, only samples collected in 2007 were used in this study to avoid the potential confounding effects due to temporal genetic variability. Hence, it

was possible sampling two locations only at one harbour per region. The locations used in this study were: Grado site 3 (GRC), Chioggia site 1 (CHA), Ravenna site 1 (RVA) and site 2 (RVB) and Rimini site 1 (RIA) for NAS; Genova site 1 (GEA) and site 2 (GEB), La Spezia site 1 (LSA), Viareggio site 2 (VIB) and Livorno site 1 (LIA) for LS; Catania site 3 (CTC), Siracusa site 1 (SIA), Porto Palo di Capo Passero site 1 (PPA), Gela site 1 (GLA) and site 2 (GLB) for SS (Fig. 1).

DNA extraction

The specimens were stored in absolute ethanol at -20 °C until DNA extraction. Whole genomic DNA was extracted from 150 specimens (10 individuals per sample) using the following procedure: Specimens were washed with sterile water to remove residual ethanol and placed in 1.5 ml tubes containing 60 µl of deionised sterile water (Sigma). After this osmotic shock, the specimens were boiled for 3 min at 100 °C in a thermal block. The tubes were then vortexed for 30 s and successively centrifuged at 3000 rpm for 2 min. The genomic DNA was stored at –20 °C until the genetic analyses were performed.

DNA sequencing and alignment

A partial sequence of the cytochrome c oxidase subunit 1 gene (COI) was amplified by Polymerase Chain Reaction (PCR) using the universal primers LCO1490 (5′-ggtcaacaaatcataaagatattgg-3′) and HCO2198 (5′-taaacttcagggtgaccaaaaaatca-3) (Folmer *et al.* 1994). The amplifications were carried out in a 20 µl reaction containing ~30 ng of genomic DNA, 0.5 U of Euroclone *Taq* DNA polymerase, 1× *Taq* Buffer, 2.5 mM $MgCl_2$, 200 µM of each dNTP and 0.2 µM of each primer. The PCR reactions were performed on a MJ PTC 200 thermal cycler

120 with an initial denaturing step at 94 °C for 2 min, 35 amplification cycles (94 °C for
121 30 s, 50 °C for 60 s, 72 °C for 90 s) and a final elongation step at 72 °C for 7 min.
122 The PCR products were directly sequenced by the BMR Genomics sequencing
123 service (Padova, Italy).
124
125 Data analysis
126 The sequences were aligned using the algorithm CLUSTAL W implemented
127 in MEGA 5 (Tamura *et al.* 2011) with the default settings and were checked
128 manually. The levels of genetic polymorphism for each location and region were
129 estimated by the number of haplotypes (N_h), haplotype diversity (h) and nucleotide
130 diversity (π) using DNAsp 5.10.01 (Librado & Rozas 2009). The Mann-Whitney *U*-
131 test was used to test for differences in the haplotype and nucleotide diversity values
132 between the three regions.
133 Hierarchical analysis of molecular variance (AMOVA) was used to quantify
134 the partitioning of genetic variance within and among the localities and among the
135 three regions. We used the approach implemented in the software Arlequin 3.5.1.3
136 (Excoffier & Lischer 2010). The Tamura & Nei (1993) genetic distances with gamma
137 distribution ($\alpha = 0.346$) were used according to the most likely model of sequence
138 evolution (TPM3uf+G) estimated with the J-ModelTest 0.1 (Posada 2008) using the
139 default options. Likelihood scores were computed for each of the 88 models under
140 the maximum likelihood framework and then the best fitting model was selected by
141 the Akaike information criterion (AIC). We also computed pairwise Φ_{ST} values
142 among all location pairs. The significance of the variance components and pairwise
143 Φ_{ST} values were assessed by a permutation test with 10,000 replicates; where

144 necessary, the false discovery rate method (FDR) was used to correct for multiple
145 testing (Benjamini & Yekutieli 2001).
146 The phylogenetic relationships among the observed haplotypes were inferred
147 using a Bayesian approach. Two published COI sequences of *Ophryotrocha japonica*
148 were used as outgroups (GenBank accession number EF464541 and GQ415478). We
149 chose this species as 1) it is not too phylogenetically divergent from *O. labronica*,
150 and 2) being the second most common species along Italian coasts (Simonini et al.
151 2010), we want to rule out even the most little chance of mtDNA introgression
152 between the two species. The COI dataset was partitioned according to codon
153 position, and for each of the resulting datasets, the most likely model of sequence
154 evolution was estimated using J-ModelTest 0.1 (Posada 2008). Bayesian
155 phylogenetic inference was carried out using MrBayes 3.1.2 (Ronquist &
156 Huelsenbeck 2003). We used a partitioned model to estimate the parameters
157 independently for each codon position with the following settings: nst = 1, rates =
158 gamma for the first codon position, according to the model F81+G; nst = 2, rates =
159 equal for the second codon position, according to the selected HKY model; nst = 6,
160 rates = equal for the third codon position, according to selected TIM3ef model. The
161 option UNLINK was used to estimate the parameters independently for each
162 partition. With the other options set to the default values, two independent runs, each
163 of four Metropolis-coupled MCMC chains (one cold and three heated chains), were
164 run simultaneously to optimise the search in tree-space. The convergence of runs was
165 assessed by checking that the average standard deviation of the split frequencies
166 reached and stabilised at values less than 0.01 (Huelsenbeck & Ronquist 2001). Trees
167 were sampled every 500 generations in each of the two independent runs, resulting in

168 a sample of 10,000 trees for each run (5,000,000 generations). The first 25% of
169 sampled trees were discarded for each run yielding a total of 15,000 trees.
170 The genetic relationships among haplotypes were also analysed by a Median-
171 Joning (MJ) network using the software Network 4.6.1.1 (Bandelt et al. 1999). We
172 used the default options, as changing parameters did not improve the network.
173 Finally, we used the Kimura 2-parameter (K2P) model to estimate the average
174 sequence divergence within and among inferred clades. The analyses were carried
175 out in Mega 5 (Tamura *et al.* 2011).
176 Historical population dynamics, such as past demographic expansions and/or
177 bottlenecks, were inferred by comparing the observed mismatch distribution of the
178 DNA substitution pairwise differences to a model of sudden population expansion
179 (Rogers & Harpending 1992). Such distributions are unimodal when populations
180 have experienced a recent expansion and multimodal at demographic equilibrium or
181 when populations are significantly subdivided. Departures from the demographic
182 expansion model were tested by comparing the sum of squared deviations (SSD)
183 between the observed and the expected mismatch distribution under such a model.
184 These analyses were carried out using Arlequin 3.5.1.3, and the significance of was
185 assessed with 10,000 bootstrap replicates.
186 Tajima's (1989) *D* neutrality test was used to infer departures from
187 equilibrium. The significance was assessed by 10,000 permutations in Arlequin
188 3.5.1.3. Significant negative values are expected in cases of recent population
189 expansion, population contraction or selective sweeps, whereas positive values are
190 expected under balancing selection, population subdivision or recent bottlenecks
191 (Soriano *et al.* 2008 and references therein). Moreover, to account for bias due to
192 small sample size, Ramos-Onsins & Rozas' (2002) R_2 statistic, which is considered

193 more powerful in such cases, was performed using DNAsp 5.10.01 (Librado &
194 Rozas 2009) to test for selective neutrality and/or population expansion. The
195 significance was tested by coalescent simulations with 10,000 permutations.
196
197 RESULTS
198 Patterns of genetic variation within and among regions
199 The successful sequencing of the 619 base pair fragment of the COI gene in
200 135 out of 150 specimens of *Ophryotrocha labronica* distinguished 40 haplotypes
201 (GenBank Accession numbers: KF305775 to KF305814). These haplotypes derived
202 from 123 polymorphic sites (19.9%), of which 105 were parsimony informative.
203 Most polymorphisms were due to single nucleotide changes (114 of the 123 sites)
204 and 14 mutations produced amino acid replacements (non-synonymous changes).
205 Thirty-three of the 40 haplotypes were location-private, 29 of which were
206 represented by a single individual. By contrast, the three most common haplotypes,
207 which were shared between different sampling sites, were present in 81 of the 135
208 specimens (Supporting information, Table S1). These haplotypes showed a disjoined
209 geographical distribution; haplotypes 1 and 2 were present only in the NAS, whereas
210 haplotype 13 (observed in 53 of the 135 specimens) occurred in all localities of the
211 SS and the LS, with the exception of Viareggio (VIB).
212 The mean nucleotide diversity across all locations was $\pi = 0.008 \pm 0.001$,
213 ranging from $\pi = 0.000$ to 0.082 ± 0.014 (Table 1). The samples from the NAS
214 showed a mean nucleotide diversity ($\pi = 0.007 \pm 0.001$) that was lower than the
215 values from the SS and LS ($\pi = 0.033 \pm 0.012$ and $\pi = 0.056 \pm 0.010$, respectively);
216 however, these values were not significantly different (Mann-Whitney *U*-test; NAS
217 *vs*. SS, p = 0.600; NAS *vs*. LS, p = 0.600; LS *vs*. SS, p = 0.917). The higher mean

218 nucleotide diversities found in the LS and SS were due to some outlier samples that
219 showed nucleotide diversities remarkably higher than those observed elsewhere:
220 Genova 1 (GEA), Viareggio (VIB) and Livorno (LIA) in the LS and Porto Palo di
221 Capo Passero (PPA) in the SS.

222 Similarly, the values of haplotype diversity were not significantly different
223 between the three regions (Mann-Whitney *U*-test; NAS *vs*. SS, p = 0.151; NAS *vs*.
224 LS, p = 0.690; LS *vs*. SS, p = 0.841). The lowest value was observed in the SS and
225 was mainly due to the presence of a single haplotype at Gela site 1 (GLA). The
226 highest haplotype diversity was observed in the VIB sample from the LS (h = 0.933
227 ± 0.077).

228 Significant genetic differentiation, which accounted for 63.1% of the total
229 variation (Table 2), was observed among the three regions (AMOVA, Φ_{CT} = 0.631, p
230 < 0.001). Such genetic structure was almost entirely due to genetic divergence
231 between the NAS and the other regions, as evidenced by pairwise Φ_{ST} values (Table
232 3). Indeed, 39 of the 44 comparisons that showed significant genetic differentiation
233 after correction for multiple testing involved comparisons between the Adriatic and
234 non-Adriatic samples. Two samples, PPA (IS) and VIB (LS), were not genetically
235 divergent from any of the Adriatic samples. Furthermore, the LIA sample (LS) was
236 not genetically divergent from the northernmost sample from the NAS (GRC). No
237 other clear patterns of genetic structure emerged (Table 3). Non-significant genetic
238 divergence was found in most comparisons between the LS and SS, as well as within
239 each region (50 of the 55 location pairwise F_{ST} values). Exceptions to this
240 observation involved the VIB and PPA samples (Table 3).

241

242 Phylogenetic analyses

243 The Bayesian phylogenetic tree evidenced two well-supported, reciprocally
244 monophyletic clades (Fig. 2), hereafter referred to as haplogroups A (HG-A) and B
245 (HG-B). The deep genetic divergence between HG-A and HG-B, as compared to
246 genetic variation within haplogroups, is highlighted by the MJ network (supporting
247 information, Fig. S1). The two haplogroups showed a different, though partially
248 overlapping, geographical distribution (Fig. 1 and Fig. S1). HG-A haplotypes were
249 present in all three of the regions, whereas HG-B haplotypes were absent in the NAS.
250 Conversely, HG-B haplotypes were dominant at most sampling sites in the LS and
251 SS regions. At localities where both types occurred sympatrically, HG-B
252 predominated at the GEA and LIA sites (LS), whereas HG-A was the most prominent
253 haplogroup at the VIB (LS) and PPA (IS) sites.
254 A total of 90 mutations (83 of which were fixed) distinguished the two
255 haplogroups, resulting in an average Kimura 2-Parameter (K2P) distance of 17.2%.
256 The divergence between the haplogroups was 18.3 and 41.9 times larger than the
257 average sequence divergence within HG-A (K2P = 0.94%) or HG-B (K2P = 0.41%),
258 respectively. HG-A displayed a deeper, more structured phylogeny than HG-B (Fig. 2
259 and Fig. S1). Most haplotypes were restricted to a single region, and only 3 of the 18
260 haplotypes were shared between two regions. Nonetheless, HG-A did not show a
261 clear-cut geographical structuring; the haplotypes did not cluster into sub-clades
262 corresponding sharply to the three regions. Instead, haplotypes from different regions
263 may be as genealogically related as haplotypes from the same region. Conversely, no
264 genetic structuring was found within the HG-B haplogroup, for which the network
265 revealed a star-like phylogeny (Fig. S1).
266
267 Demographic history

The two haplogroups showed different mismatch distributions of pairwise DNA differences (supporting information, Fig. S2); both of which, however, fit the model of demographic expansion (goodness-of-fit test, Table 4). Conversely, HG-A and HG-B evidenced discordant signatures of departure from equilibrium models according to the neutrality tests (Table 4). Only the HG-B haplotypes departed from an equilibrium model; the outcome of the neutrality tests were consistent with the results of the mismatch distributions, as significant and negative values of Tajima's *D* and low values of the R_2 statistic were found (Table 4).

DISCUSSION

In this study, we analysed the patterns of intra-specific genetic variation in samples of *Ophryotrocha labronica* from three distinct regions of the Mediterranean Sea: the Northern Adriatic Sea (NAS), Ligurian Sea (LS) and south/south-east Sicilian Sea (SS). We found significant genetic differentiation between the NAS and the other two regions, whilst shallow or no genetic structuring was found within each region or between the LS and SS. This genetic pattern is strongly characterised by the occurrence of two highly divergent COI lineages, here referred to as haplogroups A (HG-A) and B (HG-B), which display different frequencies and geographical distributions across the three regions.

Patterns of genetic diversity

Estimates of within-location genetic diversity (Table 1) were comparable to those found in other polychaete species (Virgilio *et al.* 2009 and references therein). Additionally, following Grant & Bowen's (1998) considerations about historical demographic inference, only six of the 15 locations of *O. labronica* in the present

study displayed values of haplotype and nucleotide diversity that were consistent with the occurrence of a recent bottleneck (GLA location) or a rapid expansion after a genetic bottleneck (GEB, LSA, CTC, SIA and GLB locations). The marked temporal and spatial fluctuations in the density of *O. labronica* populations (Prevedelli *et al.* 2005; Åkesson & Paxton 2005) may account for these results. At larger spatial scales, the absence of a significant variation in haplotype and nucleotide diversity across the regions suggests that different environmental conditions have little influence on the amount of genetic variation within these locations. This result may reflect the adaptive plasticity of *O. labronica* to the different environmental conditions that characterise each region. Indeed, populations of species with a limited adaptive potential may undergo a depletion of genetic variability if they are not well adapted to environmental conditions, as suggested in the copepod *Tigriopus californicus* along a thermal gradient (Kelly *et al.* 2012).

Alternatively, genetic variation may be maintained by local adaptation driven by evolution under spatially divergent selection, which requires a set of phenotypes with maximised fitness in a given habitat type and limited gene flow (Kawecki & Ebert 2004). In *O. labronica,* life-history traits related to fitness highlighted a significant differentiation between the LS and the other two regions, the SS and NAS (Massamba N'Siala *et al.* 2011). Conversely, hierarchical AMOVA (Table 2) and pairwise Φ_{ST} values (Table 4) evidenced a genetic architecture that was partially different from the pattern of phenotypic variation, showing a remarkable genetic divergence between the NAS and the other regions. Overall, these results suggest that either phenotypic plasticity is driving the adaptation to different environmental regimes or the patterns of COI variation and the life-history traits (Massamba N'Siala *et al.* 2011) are not related. Indeed, the patterns of genetic variation may

reflect genetic drift and/or historical vicariant events, rather than local adaptation (Kawecki & Ebert 2004).

Phylogeographical patterns

The Adriatic Sea represents a region of phylogeographical discontinuity within the Mediterranean Sea (Patarnello *et al.* 2007). Here, historical and present-day barriers to dispersal might have promoted genetic divergence in the Adriatic populations of many invertebrate species (Virgilio *et al.* 2009 and references therein). However, the picture may be more complex in *O. labronica*. First, the two main genetic lineages (HG-A and HG-B) evidenced by the phylogenetic tree (Fig. 2) and network (Fig. S1) did not show a completely separate geographical distribution, but were sympatric at some locations in the LS and SS regions (Fig. 1). However, sympatric divergence is unlikely to occur in *O. labronica:* cross-breeding experiments showed that individuals collected from the same locations used in this study, including those in which HG-A and HG-B co-occur, were interfertile (Massamba N'Siala *et al.* 2011).

Taking into account the high level of average sequence divergence between HG-A and HG-B, allopatric divergence followed by secondary contact is a more likely explanation than sympatric divergence. Indeed, in geographically isolated populations, mutations may accumulate for a long evolutionary time before reproductive barriers arise, if they ever arise (Norris & Hull 2011). As suggested by Virgilio *et al.* (2009), concerning the Adriatic and non-Adriatic haplotypes of the polychaete *Hediste diversicolor* in the Mediterranean, the divergence may be older than the re-flooding of the northern and central Adriatic basins subsequent to the end of the last ice age (ca. 18,000 years ago). Such a scenario is consistent with our

results on *O. labronica*, as *Ophryotrocha* species may easily survive in ballast waters and in fouling communities on ship hulls (Åkesson & Paxton 2005); hence, vessel-mediated dispersal may promote the secondary contacts of geographically isolated lineages. The chance of dispersal by ship transport, even over long distances, is supported by Simonini *et al*. (2009) observations of *Ophryotrocha japonica* and *Ophryotrocha diadema,* which were recently introduced into the Mediterranean through vessel-mediated dispersal, and by Åkesson & Paxton's (2005) findings on the population of *O. labronica* recorded in the Sydney harbour (Australia), which may have reached the Pacific Ocean through the same mechanism.

Demographic history

The lack of genetic structure observed within each lineage of *O. labronica* and within each region, as well as the non-significant genetic divergence between the LS and SS, contradicts the expectation of deep genetic structuring in species without pelagic larvae. Along with the contribution to gene flow provided by human-mediated dispersal, drifting, or rafting (e.g., Simonini *et al.* 2010 and references therein), other processes such as bottlenecks, expansions and recent colonisations may have contributed to the observed patterns. Patarnello *et al.* (2007) showed that recent expansions and/or bottlenecks may explain the lack of genetic structure in several Mediterranean species.

The demographic history of the HG-B haplotypes is consistent with a recent bottleneck/expansion scenario, with a clear unimodal mismatch distribution (Fig. S2e, f and g). Furthermore, the concordant results of all neutrality tests (Table 4) point to population growth rather than other alternative scenarios (Ramos-Onsin & Rozas 2002; Soriano *et al.* 2008). The picture is more complex for the HG-A

368 haplotypes; notwithstanding a multimodal pattern (Fig. S2a, b, c and d), the observed
369 mismatch distributions still fit a model of demographic expansion according to the
370 goodness-of-fit test, whereas none of the neutrality tests showed significant
371 departures from an equilibrium model (Table 4). However, the observed pattern does
372 not necessarily imply a picture of demographic stability; multimodal mismatch
373 distributions combined with slightly negative, but non-significant, Tajima's *D* values
374 may reflect mild population expansions combined with heterogeneity of mutation
375 rates (Aris-Brosou & Excoffier 1996).
376 Alternatively, these results fit the outcome observed in the range expansion
377 model, in which genetically structured populations exchange genes (Ray *et al.* 2003).
378 Under these conditions, the mismatch distributions display multimodal patterns and
379 the neutrality tests may fail to reject departure from equilibrium. Furthermore, as
380 expected under this model for moderate migration rates (Excoffier 2004), the
381 multimodal mismatch distribution of the HG-A haplotypes showed a peak close to
382 the origin. Patarnello *et al.* (2007) suggested that similar patterns of mismatch
383 distributions reflect multiple expansion events. Moreover, expansions after a
384 population size reduction may determine the lack of genetic structure observed in
385 some species, even in the absence of gene flow. Indeed, fluctuations in abundance
386 over time and variance in reproductive success may lead to shallow genetic
387 structuring in marine species (Grant & Bowen 1998). This picture fits with the
388 metapopulation scenario invoked for *O. labronica*, an opportunistic species
389 subdivided into many local populations that proliferate under favourable conditions
390 and may rapidly go extinct to be replaced by new populations (Åkesson & Paxton
391 2005; Prevedelli *et al.* 2005).
392

393 Genetic differentiation between haplogroups

394 Our results raise the question whether the two haplogroups may be the result
395 of cryptic speciation. The occurrence of cryptic species is rather common within the
396 polychaetes (Carr *et al.* 2011 and references therein). The average sequence
397 divergence between the two lineages detected in the present study (K2P distance =
398 17.2%) was comparable to the values of COI divergence found among polychaete
399 cryptic species. For instance, the COI sequencing of 1876 specimens from 142
400 morphologically identified species of Canadian polychaetes yielded a total of 333
401 divergent lineages, with a mean K2P distance between the genetic clusters that
402 averaged 16.5% (Carr *et al.* 2011). In the ragworm *Hediste diversicolor*, the mtDNA
403 sequence divergence among distinct lineages and/or across different geographical
404 regions ranged between 4.5 and 8% (Virgilio *et al.* 2009). At least two of these
405 lineages are likely to be cryptic species, consistent with the results of studies that
406 used nuclear molecular markers (Audzijonyte *et al.* 2008). Moreover, the mean
407 sequence divergence between the haplogroups of *O. labronica* was 18 and 43 times
408 higher than within-haplogroup variation for HG-A and HG-B (K2P distance = 0.94%
409 and 0.41%), respectively. These ratios exceeded the threshold used for the COI
410 barcoding gap, by which inter-specific variation is expected to be at least ten times
411 larger than intra-specific variation (Hebert *et al.* 2004).
412 Although COI proved its effectiveness for DNA barcoding and species identification
413 in polychaetes (Carr *et al.* 2011), caution is needed when inferring cryptic speciation
414 in *O. labronica*. Cryptic polychaete species inferred by COI barcoding usually
415 showed differences in other traits, such as those related to life- history, ecology,
416 habitat specificity and/or the presence of reproductive isolation (Carr *et al.* 2011 and
417 references therein). The individuals of the two haplogroups found in *O. labronica*

were neither reproductively isolated, nor did they show habitat specificity. Moreover, the patterns of mtDNA and life-history trait variation were poorly correlated, and the geographical distribution of the haplogroups could not be straightforwardly linked to ecological differences among the three regions. Hence, given the absence of such lines of evidence and the comparison of mtDNA data with nuclear markers, following the recommendations of Bickford *et al.* (2007), HG-A and HG-B should be regarded as distinct mtDNA lineages of the species *O. labronica*.

ACKNOWLEDGEMENTS

This research was supported by the national PRIN 2005 project (prot. 2005050952), funded by the Italian *Ministero dell'Istruzione*, dell'*Universita e* della *Ricerca*. Genetic analyses were carried out at the Molecular Ecology laboratory of the ENEA, Marine Environment Research Centre Santa Teresa, la Spezia, Italy.

LITERATURE CITED

Åkesson B., Paxton H. (2005) Biogeography and incipient speciation in *O. labronica* (Polychaeta, Dorvilleidae). *Marine Biology Research*, **1**, 27–139.

Alonso R., Crawford A.J., Bermingham E. (2012) Molecular phylogeny of an endemic radiation of Cuban toads (Bufonidae: Peltophryne) based on mitochondrial and nuclear genes. *Journal of Biogeography*, **39**, 434–451.

Aris-Brosou S., Excoffier L. (1996) The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. *Molecular Biology and Evolution*, **13**, 494–504.

Audzijonyte A., Ovcarenko I., Bastrop R., Väinölä R. (2008) Two cryptic species of the *Hediste diversicolor* group (Polychaeta, Nereididae) in the Baltic Sea, with

443 mitochondrial signatures of different population histories. *Marine Biology*, **155**,
444 599–612.
445 Bandelt H.J., Forster P., Rohl A. (1999) Median-joining networks for inferring
446 intraspecific phylogenies. *Molecular Biology and Evolution,* **16**: 37–48.
447 Benjamini Y., Yekutieli D. (2001) The control of the false discovery rate in multiple
448 testing under dependency. *The Annals of Statistics*, **29**, 1165–1188.
449 Bickford D., Lohman D.J., Sodhi N.S., Ng P.K.L., Meier R., Winker K., Ingram
450 K.K., Das I. (2007) Cryptic species as window on diversity and conservation.
451 *Trends in Ecology and Evolution*, **22,** 148–155.
452
453 Carr C.M., Hardy S.M., Brown T.M., Macdonald T.A., Hebert P.D. (2011) A Tri-
454 Oceanic perspective: DNA barcoding reveals geographic structure and cryptic
455 diversity in Canadian polychaetes. *PLoS One*, **6**, e22232.
456
457 Excoffier L. (2004) Patterns of DNA sequence diversity and genetic structure after a
458 range expansion: lessons from the infinite-island model. *Molecular Ecology*, **13**,
459 853–864.
460 Excoffier L., Lischer H.E.L. (2010) Arlequin suite ver 3.5: A new series of programs
461 to perform population genetics analyses under Linux and Windows. *Molecular*
462 *Ecology Resources*, **10**, 564–567.
463 Folmer O., Black M., Hoeh W., Lutz R., Vrijenhoek R. (1994) DNA primers for
464 amplification of mitochondrial cytochrome c oxidase subunit I from diverse
465 metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294–
466 296.

467 Grant W.A.S., Bowen B.W. (1998) Shallow population histories in deep evolutionary
468 lineages of marine fishes: insights from sardines and anchovies and lessons for
469 conservation. *Journal of Heredity*, **89**, 415–426.
470 Hebert P.N., Stoeckle M.Y., Zemlak T.S., Francis C.M. (2004) Identification of birds
471 through DNA barcodes. *PLoS Biology*, **2**,1657–1663.
472 Huelsenbeck J.P., Ronquist F. (2001) MRBAYES: Bayesian inference of phylogeny.
473 *Bioinformatics*, **17**, 754–755.
474
475 Kawecki T.D., Ebert D. (2004) Conceptual issues in local adaptation. *Ecology*
476 *Letters*, **7**, 1225–1241.
477 Kelly M.W., Sanford E., Grosberg R.K. (2012) Limited potential for adaptation to
478 climate change in a broadly distributed marine crustacean. *Proceeding of the*
479 *Royal Society B*, **279**, 349–356.
480 Librado P., Rozas J. (2009) DnaSP v5: A software for comprehensive analysis of
481 DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.
482 Massamba-N'Siala G., Simonini R., Cossu P., Maltagliati F., Castelli A., Prevedelli
483 D. (2011) Life-history and demographic spatial variation in Mediterranean
484 populations of the opportunistic polychaete *Ophryotrocha labronica* (Polychaeta,
485 Dorvilleidae). *Marine Biology*, **158**, 1523–1535.
486 Norris R.D., Hull P.M. (2012) The temporal dimension of marine speciation.
487 *Evolutionary Ecology*, **26**, 393–415.
488 Patarnello T., Volckaert F.A.M.J., Castilho R. (2007) Pillars of hercules: is the
489 Atlantic-Mediterranean transition a phylogeographical break? *Molecular*
490 *Ecology*, **16**, 4426–4444.

Paxton H., Åkesson B. (2007) Redescription of *Ophryotrocha puerilis* and *O. labronica* (Annelida, Dorvilleidae). *Marine Biology Research*, **3**, 3–19.

Posada D. (2008) jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution*, **25**, 1253–1256.

Prevedelli D., Massamba-N'Siala G., Simonini R. (2005) The seasonal dynamics of six species of Dorvilleidae (Polychaeta) in the harbour of La Spezia (Italy). *Marine Ecology*, **26**,286–293.

Ramos-Onsins S.E., Rozas J.E. (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**, 2092–2100.

Ray N., Currat M., Excoffier L. (2003) Intra-deme molecular diversity in spatially expanding populations. *Molecular Biology and Evolution*, **20**, 76–86.

Reusch T.B.H., Wood T.E. (2007) Molecular ecology of global change. *Molecular Ecology*, **16**, 3973–3992.

Rogers A.R., Harpending H. (1992) population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.

Ronquist F., Huelsenbeck J.P. (2003) mrbayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Simonini, R., Massamba-N'Siala, G., Prandi, V., Prevedelli, D. (2009) Distribution of the genus *Ophryotrocha* (Polychaeta) in Italy: new records and comments on the biogeography of Mediterranean species. *Vie et Milieu*, **59**, 79–88.

Simonini R., Grandi V., Massamba-N'Siala G., Martino M.P., Castelli A., Prevedelli D. (2010) Diversity, habitat affinities and diet of *Ophryotrocha* species

(Polychaeta, Dorvilleidae) living in Mediterranean harbour habitats. *Vie et Milieu*, **60**, 27–38.

Soriano A.R., Ramos-Onsins S.E., Rozas J., Calafell F., Navarro A. (2008) Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. *Genetics*, **179**, 555–567.

Tajima F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–589.

Tamura K., Nei M. (1993) Estimations of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512–526.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M., Kumar S. (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Thornhill D.J., Dahlgren T.G., Halanych K. (2009) The evolution and ecology of Ophryotrocha (Dorvilleidae, Eunicida). In: D.H. Shain (Eds). *Annelids as Model Systems in the Biological Sciences*. John Wiley and Sons: 242–256.

Virgilio M., Fauvelot C., Costantini F., Abbiati M., Backeljau T. (2009) Phylogeography of the common ragworm *Hediste diversicolor* (Polychaeta: Nereididae) reveals cryptic diversity and multiple colonization events across its distribution. *Molecular Ecology*, **18**, 1980–1994.

Zakharov E.V., Hellmann J.J. (2008) Genetic differentiation across a latitudinal gradient in two co-occurring butterfly species: revealing population differences in a context of climate change. *Molecular Ecology*, **17**, 189–208.

TABLE LEGENDS

Table 1: Estimates of genetic diversity. *N*: sample size; N_h: number of haplotypes; *N*p: number of polymorphic sites; *h*: haplotype diversity (± standard deviation); ð: nucleotide diversity (± standard deviation)

Table 2: Results of the hierarchical analysis of molecular variance (AMOVA) with groups corresponding to: Northern Adriatic Sea (NAS); Ligurian Sea (LS); south/south-east Sicilian Sea (SS).

Table 3: Pairwise Φ_{ST} values between samples. Values that were significant at $P < 0.05$ after the correction for multiple testing are in bold. The false discovery rate method (FDR) according to Benjamini and Yekutieli (2001) was applied to detect the false positives (Tipe I error). Label codes are reported in Table 1

Table 4: Goodness-of-fit tests for mismatch distributions and neutrality tests for the two main lineages and the three regions considered in this study. Goodness-of-fit test: SSD, sums of squared deviations. Neutrality tests: *D*, Tajima's (1989) neutrality test; R_2, Ramos-Onsins & Rozas' (2002) statistics. Significant values are outlined in bold

FIGURE CAPTIONS

Figure 1: a) Geographical location of sampling sites. GRC: Grado (45°40'N, 13°23'E); CHA: Chioggia (45°13', 12°16'E); RVA and RVB: Ravenna (44°28'N, 12°15'E); RIA: Rimini (44°05'N, 12°33'E); GEA and GEB: Genova (44°24'N, 8°54'E); LSA: La Spezia (44°04'N, 9°49'E); VIB: Viareggio (43°51'N, 10°14'E); LIA: Livorno (43°33'N, 10°18'E); CTC: Catania (37°29'N, 15°05'E); SIA: Siracusa (37°03'N, 15°17'E); PPA: Porto Palo di Capo Passero (36°40'N, 15°07'E); GLA and GLB: Gela (37°04'N, 14°03'E). The piecharts depict the Geographical distribution of the two COI lineages (HG-A and HG-B) at each sampling location

Figure 2: Bayesian tree showing the phylogenetic relationships among COI haplotypes . The two COI lineages are denoted with HG-A (Haplogroup A) and HG-B (Haplogroup B). Tree was rooted using two sequences of *Ophryotrocha japonica* as outgroups. Only nodes supported by posterior probabilities ≥95% are reported. Localities in which were found haplotypes and their frequency are reported in Table S1 (supplementary material)

SSSS