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’-ggcaaaatcataaatattgg-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 MgCl₂, 200 μM of each dNTP and 0.2 μM of each primer. The PCR reactions were performed on a MJ PTC 200 thermal cycler.
with an initial denaturing step at 94 °C for 2 min, 35 amplification cycles (94 °C for
30 s, 50 °C for 60 s, 72 °C for 90 s) and a final elongation step at 72 °C for 7 min.
The PCR products were directly sequenced by the BMR Genomics sequencing
service (Padova, Italy).

Data analysis

The sequences were aligned using the algorithm CLUSTAL W implemented
in MEGA 5 (Tamura et al. 2011) with the default settings and were checked
manually. The levels of genetic polymorphism for each location and region were
estimated by the number of haplotypes \(N_h\), haplotype diversity \(h\) and nucleotide
diversity \(\pi\) using DNAsp 5.10.01 (Librado & Rozas 2009). The Mann-Whitney \(U\)-
test was used to test for differences in the haplotype and nucleotide diversity values
between the three regions.

Hierarchical analysis of molecular variance (AMOVA) was used to quantify
the partitioning of genetic variance within and among the localities and among the
three regions. We used the approach implemented in the software Arlequin 3.5.1.3
(Excoffier & Lischer 2010). The Tamura & Nei (1993) genetic distances with gamma
distribution (\(\alpha = 0.346\)) were used according to the most likely model of sequence
evolution (TPM3uf+G) estimated with the J-ModelTest 0.1 (Posada 2008) using the
default options. Likelihood scores were computed for each of the 88 models under
the maximum likelihood framework and then the best fitting model was selected by
the Akaike information criterion (AIC). We also computed pairwise \(\Phi_{ST}\) values
among all location pairs. The significance of the variance components and pairwise
\(\Phi_{ST}\) values were assessed by a permutation test with 10,000 replicates; where
necessary, the false discovery rate method (FDR) was used to correct for multiple testing (Benjamini & Yekutieli 2001).

The phylogenetic relationships among the observed haplotypes were inferred using a Bayesian approach. Two published COI sequences of *Ophryotrocha japonica* were used as outgroups (GenBank accession number EF464541 and GQ415478). We chose this species as 1) it is not too phylogenetically divergent from *O. labronica*, and 2) being the second most common species along Italian coasts (Simonini et al. 2010), we want to rule out even the most little chance of mtDNA introgression between the two species. The COI dataset was partitioned according to codon position, and for each of the resulting datasets, the most likely model of sequence evolution was estimated using J-ModelTest 0.1 (Posada 2008). Bayesian phylogenetic inference was carried out using MrBayes 3.1.2 (Ronquist & Huelsenbeck 2003). We used a partitioned model to estimate the parameters independently for each codon position with the following settings: nst = 1, rates = gamma for the first codon position, according to the model F81+G; nst = 2, rates = equal for the second codon position, according to the selected HKY model; nst = 6, rates = equal for the third codon position, according to selected TIM3ef model. The option UNLINK was used to estimate the parameters independently for each partition. With the other options set to the default values, two independent runs, each of four Metropolis-coupled MCMC chains (one cold and three heated chains), were run simultaneously to optimise the search in tree-space. The convergence of runs was assessed by checking that the average standard deviation of the split frequencies reached and stabilised at values less than 0.01 (Huelsenbeck & Ronquist 2001). Trees were sampled every 500 generations in each of the two independent runs, resulting in
a sample of 10,000 trees for each run (5,000,000 generations). The first 25% of
sampled trees were discarded for each run yielding a total of 15,000 trees.
The genetic relationships among haplotypes were also analysed by a Median-
Joning (MJ) network using the software Network 4.6.1.1 (Bandelt et al. 1999). We
used the default options, as changing parameters did not improve the network.
Finally, we used the Kimura 2-parameter (K2P) model to estimate the average
sequence divergence within and among inferred clades. The analyses were carried
out in Mega 5 (Tamura et al. 2011).
Historical population dynamics, such as past demographic expansions and/or
bottlenecks, were inferred by comparing the observed mismatch distribution of the
DNA substitution pairwise differences to a model of sudden population expansion
(Rogers & Harpending 1992). Such distributions are unimodal when populations
have experienced a recent expansion and multimodal at demographic equilibrium or
when populations are significantly subdivided. Departures from the demographic
expansion model were tested by comparing the sum of squared deviations (SSD)
between the observed and the expected mismatch distribution under such a model.
These analyses were carried out using Arlequin 3.5.1.3, and the significance of was
assessed with 10,000 bootstrap replicates.
Tajima’s (1989) \( D \) neutrality test was used to infer departures from
equilibrium. The significance was assessed by 10,000 permutations in Arlequin
3.5.1.3. Significant negative values are expected in cases of recent population
expansion, population contraction or selective sweeps, whereas positive values are
expected under balancing selection, population subdivision or recent bottlenecks
(Soriano et al. 2008 and references therein). Moreover, to account for bias due to
small sample size, Ramos-Ornsins & Rozas’ (2002) \( R^2 \) statistic, which is considered
more powerful in such cases, was performed using DNAsp 5.10.01 (Librado & Rozas 2009) to test for selective neutrality and/or population expansion. The significance was tested by coalescent simulations with 10,000 permutations.

RESULTS

Patterns of genetic variation within and among regions

The successful sequencing of the 619 base pair fragment of the COI gene in 135 out of 150 specimens of *Ophryotrocha labronica* distinguished 40 haplotypes (GenBank Accession numbers: KF305775 to KF305814). These haplotypes derived from 123 polymorphic sites (19.9%), of which 105 were parsimony informative. Most polymorphisms were due to single nucleotide changes (114 of the 123 sites) and 14 mutations produced amino acid replacements (non-synonymous changes). Thirty-three of the 40 haplotypes were location-private, 29 of which were represented by a single individual. By contrast, the three most common haplotypes, which were shared between different sampling sites, were present in 81 of the 135 specimens (Supporting information, Table S1). These haplotypes showed a disjoined geographical distribution; haplotypes 1 and 2 were present only in the NAS, whereas haplotype 13 (observed in 53 of the 135 specimens) occurred in all localities of the SS and the LS, with the exception of Viareggio (VIB).

The mean nucleotide diversity across all locations was $\pi = 0.008 \pm 0.001$, ranging from $\pi = 0.000$ to $0.082 \pm 0.014$ (Table 1). The samples from the NAS showed a mean nucleotide diversity ($\pi = 0.007 \pm 0.001$) that was lower than the values from the SS and LS ($\pi = 0.033 \pm 0.012$ and $\pi = 0.056 \pm 0.010$, respectively); however, these values were not significantly different (Mann-Whitney $U$-test; NAS vs. SS, $p = 0.600$; NAS vs. LS, $p = 0.600$; LS vs. SS, $p = 0.917$). The higher mean
nucleotide diversities found in the LS and SS were due to some outlier samples that
showed nucleotide diversities remarkably higher than those observed elsewhere:
Genova 1 (GEA), Viareggio (VIB) and Livorno (LIA) in the LS and Porto Palo di
Capo Passero (PPA) in the SS.

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

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

Phylogenetic analyses
The Bayesian phylogenetic tree evidenced two well-supported, reciprocally monophyletic clades (Fig. 2), hereafter referred to as haplogroups A (HG-A) and B (HG-B). The deep genetic divergence between HG-A and HG-B, as compared to genetic variation within haplogroups, is highlighted by the MJ network (supporting information, Fig. S1). The two haplogroups showed a different, though partially overlapping, geographical distribution (Fig. 1 and Fig. S1). HG-A haplotypes were present in all three of the regions, whereas HG-B haplotypes were absent in the NAS. Conversely, HG-B haplotypes were dominant at most sampling sites in the LS and SS regions. At localities where both types occurred sympatrically, HG-B predominated at the GEA and LIA sites (LS), whereas HG-A was the most prominent haplogroup at the VIB (LS) and PPA (IS) sites.

A total of 90 mutations (83 of which were fixed) distinguished the two haplogroups, resulting in an average Kimura 2-Parameter (K2P) distance of 17.2%. The divergence between the haplogroups was 18.3 and 41.9 times larger than the average sequence divergence within HG-A (K2P = 0.94%) or HG-B (K2P = 0.41%), respectively. HG-A displayed a deeper, more structured phylogeny than HG-B (Fig. 2 and Fig. S1). Most haplotypes were restricted to a single region, and only 3 of the 18 haplotypes were shared between two regions. Nonetheless, HG-A did not show a clear-cut geographical structuring; the haplotypes did not cluster into sub-clades corresponding sharply to the three regions. Instead, haplotypes from different regions may be as genealogically related as haplotypes from the same region. Conversely, no genetic structuring was found within the HG-B haplogroup, for which the network revealed a star-like phylogeny (Fig. S1).

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 $\Phi_{ST}$ 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
haplotypes; notwithstanding a multimodal pattern (Fig. S2a, b, c and d), the observed mismatch distributions still fit a model of demographic expansion according to the goodness-of-fit test, whereas none of the neutrality tests showed significant departures from an equilibrium model (Table 4). However, the observed pattern does not necessarily imply a picture of demographic stability; multimodal mismatch distributions combined with slightly negative, but non-significant, Tajima’s $D$ values may reflect mild population expansions combined with heterogeneity of mutation rates (Aris-Brosou & Excoffier 1996).

Alternatively, these results fit the outcome observed in the range expansion model, in which genetically structured populations exchange genes (Ray et al. 2003). Under these conditions, the mismatch distributions display multimodal patterns and the neutrality tests may fail to reject departure from equilibrium. Furthermore, as expected under this model for moderate migration rates (Excoffier 2004), the multimodal mismatch distribution of the HG-A haplotypes showed a peak close to the origin. Patarnello et al. (2007) suggested that similar patterns of mismatch distributions reflect multiple expansion events. Moreover, expansions after a population size reduction may determine the lack of genetic structure observed in some species, even in the absence of gene flow. Indeed, fluctuations in abundance over time and variance in reproductive success may lead to shallow genetic structuring in marine species (Grant & Bowen 1998). This picture fits with the metapopulation scenario invoked for $O. \text{labronica}$, an opportunistic species subdivided into many local populations that proliferate under favourable conditions and may rapidly go extinct to be replaced by new populations (Åkesson & Paxton 2005; Prevedelli et al. 2005).
Our results raise the question whether the two haplogroups may be the result of cryptic speciation. The occurrence of cryptic species is rather common within the polychaetes (Carr et al. 2011 and references therein). The average sequence divergence between the two lineages detected in the present study (K2P distance = 17.2%) was comparable to the values of COI divergence found among polychaete cryptic species. For instance, the COI sequencing of 1876 specimens from 142 morphologically identified species of Canadian polychaetes yielded a total of 333 divergent lineages, with a mean K2P distance between the genetic clusters that averaged 16.5% (Carr et al. 2011). In the ragworm Hediste diversicolor, the mtDNA sequence divergence among distinct lineages and/or across different geographical regions ranged between 4.5 and 8% (Virgilio et al. 2009). At least two of these lineages are likely to be cryptic species, consistent with the results of studies that used nuclear molecular markers (Audzijonyte et al. 2008). Moreover, the mean sequence divergence between the haplogroups of O. labronica was 18 and 43 times higher than within-haplogroup variation for HG-A and HG-B (K2P distance = 0.94% and 0.41%), respectively. These ratios exceeded the threshold used for the COI barcoding gap, by which inter-specific variation is expected to be at least ten times larger than intra-specific variation (Hebert et al. 2004).

Although COI proved its effectiveness for DNA barcoding and species identification in polychaetes (Carr et al. 2011), caution is needed when inferring cryptic speciation in O. labronica. Cryptic polychaete species inferred by COI barcoding usually showed differences in other traits, such as those related to life-history, ecology, habitat specificity and/or the presence of reproductive isolation (Carr et al. 2011 and 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'Università 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**


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); $\delta$: 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 $\Phi_{ST}$ 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 (Type 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)