



# Unearthing the historical biogeography of Mediterranean earthworms (Annelida: Hormogastridae)

Marta Novo<sup>1,2\*</sup>, Rosa Fernández<sup>1</sup>, Daniel Fernández Marchán<sup>3</sup>, Dolores Trigo<sup>3</sup>, Darío J. Díaz Cosín<sup>3</sup> and Gonzalo Giribet<sup>1</sup>

<sup>1</sup>Department of Organismic and Evolutionary Biology, Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138, USA, <sup>2</sup>Cardiff School of Biosciences, Cardiff University, Cardiff, UK, <sup>3</sup>Departamento de Zoología y Antropología Física, Facultad de Biología, Universidad Complutense de Madrid, 28040 Madrid, Spain

## ABSTRACT

**Aim** The genetic diversity and distribution of earthworm species in the Mediterranean Basin has been influenced by their low vagility, as well as by the basin's complex geological and climatic history. Within this context, our objective was to evaluate the evolutionary history of hormogastrid earthworms by exploring their phylogeny, reconstructing ancestral areas and identifying potential vicariance and dispersal events.

**Location** Western Mediterranean region, encompassing the geographical range of the family Hormogastridae (four genera, 30 species and subspecies).

**Methods** We sampled the easternmost area of the hormogastrid range and integrated 606 new sequences with all the molecular data available from previous studies on the westernmost area. Mitochondrial and nuclear markers were amplified and sequenced, and the hormogastrid phylogeny (using Bayesian and likelihood methods) and networks were investigated. Ancestral-area reconstructions were implemented in RASP.

**Results** Ancestral-area reconstructions provided similar results for alternative phylogenetic hypotheses, placing the origin of Hormogastridae between southern France and the north-eastern Iberian Peninsula, and highlighting north–central Sardinia as a key ancestral area for diversification of eastern Hormogastridae (*Hormogaster redii* and *H. samnitica*). Multiple vicariance and dispersal events were detected, indicating a complex evolutionary history. Sardinian populations of *H. samnitica* and *H. redii* were not monophyletic but were related to populations from other areas, suggesting that Sardinia is a more complex biogeographical region than anticipated. Our results identified a phylogenetic relationship between the southern Sardinian fauna and that of Sicily, as well as relationships between the fauna of north-eastern Sardinia and those of the Italian Peninsula and the Tuscan Archipelago.

**Main conclusions** Our results suggest a complex evolutionary history for hormogastrid earthworms, showing the possibility of many dispersal and vicariance events and multiple faunal interchanges between land-masses, which may reflect the biogeographical complexity of the Mediterranean Basin.

## Keywords

Ancestral area, Annelida, earthworms, Mediterranean, Messinian salinity crisis, microplates, vicariance.

\*Correspondence and current address: Marta Novo, Cardiff School of Biosciences, Cardiff University, BIOSI 1, Museum Avenue, Cardiff CF10 3TL, UK.  
E-mail: novom@cf.ac.uk

## INTRODUCTION

The Mediterranean Basin has been the scene of a turbulent geological and climatic history that has shaped species' genetic diversity and distributions (e.g. Magri *et al.*, 2007;

Bidegaray-Batista & Arnedo, 2011). Tectonics in the Mediterranean region during the Oligocene resulted in vicariance events, which are often cited as promoters of speciation (Cunningham & Collins, 1998). Multiple contacts and discontinuities of land-masses in the region during the

Messinian salinity crisis (MSC; up to 5.33 Ma) and other climatic vicissitudes have affected the present species distributions in complex ways, altering their diversification and gene flow (Ketmaier *et al.*, 2006; Zangari *et al.*, 2006; Gentile *et al.*, 2010; Salvo *et al.*, 2010; Bidegaray-Batista & Arnedo, 2011). Consequently, this well-defined biogeographical region, with its unique fauna and an abundance of information on its past geological history (Ketmaier & Caccone, 2013), is an ideal setting for evolutionary research.

Of particular interest is the geomorphology and history of Sardinia, the south-western area of which has unique vegetation, geolithological, geomorphological, palaeogeographical and bioclimatic characteristics (Bacchetta *et al.*, 2007) due to its separation from the rest of the island by a sea arm, which was later filled with alluvial deposits that now form the Campidano plain (Omodeo & Rota, 2008). In addition, intense tectonic activity caused the separation of its eastern and western areas by shallow seas (Andeweg, 2002). Finally, Gallura in the north-east has been shown to represent a zoogeographical singularity that is isolated from the rest of the island through habitat fragmentation (Rota, 1992).

The faunal colonization of Corsica and Sardinia is believed to have occurred in three phases (see examples in Ketmaier & Caccone, 2013). First, the detachment of the Corso-Sardinian microplate from the Iberian Peninsula (pre-Miocene) was responsible for a subsequent origin of endemic species derived from the Iberian ancestors. Second, the MSC permitted secondary contact and interchange of species across a largely drained Mediterranean Basin and, third, further connections during the Quaternary ice ages may have allowed further contact and dispersal.

The study of species' distributions and diversity in this area can aid in the understanding of this complex evolutionary setting. Earthworms have been useful in biogeographical studies, and several studies indicate that their distributions have been largely shaped by palaeogeographical and palaeoecological events as a result of their low vagility (e.g. Michaelsen, 1922; Omodeo, 1963, 2000; Buckley *et al.*, 2011; Novo *et al.*, 2011), although this may not be the case in synanthropic species with near-cosmopolitan distributions (e.g. Gates, 1966; Bouché, 1972). James (2004) postulated that Earth history linked to human activity might have affected earthworm distributions, highlighting the need to consider such potential influences where possible. In the particular case of hormogastrids, given their endemism, endogeic ecology, and their near absence in human-transformed areas, it seems reasonable to hypothesize that their present distribution reflects ancient geological events rather than being the result of human activity.

Earthworms of the family Hormogastridae are distributed in the western Mediterranean region (Cobolli Sbordoni *et al.*, 1992), with three species occupying the easternmost area of the family's range: *Hormogaster redii* Rosa, 1887 (and subspecies *H. redii gigantea* Michaelsen, 1918 and *H. redii insularis* Bouché, 1970), *H. samnitica* Cognetti, 1914 (and subspecies *H. samnitica lirapora* Bouché, 1970) and *H. pretiosa* Michaelsen, 1899 (Omodeo & Rota, 2008). (Although we follow the

current taxonomy for the group, including many subspecies, this is no indication that we accept or condone the use of these subspecific taxa.) *Hormogaster redii* shows remarkable ecological breadth, inhabiting a range of habitats from sclerophyllous forests, pastures and coarse granitic sand to sandy beaches above the shoreline (Omodeo & Rota, 2008; M.N., pers. obs.). Owing to these characteristics, *H. redii* has been described as one of the dominant members of the earthworm community of Sardinia and lesser islands (Omodeo & Rota, 2008). Furthermore, although allopatry seems to be the rule for hormogastrids (Qiu & Bouché, 1998; Novo *et al.*, 2011), *H. redii* is an exception, coexisting with either *H. samnitica* or *H. pretiosa* in the eastern area of the family's distribution.

The main purpose of the present study was therefore to explore the biogeographical history of hormogastrid earthworms, with an emphasis on the easternmost area of their distribution. For this, we studied the phylogenetic relationships and inferred ancestral areas as well as possible dispersal and vicariance events related to the complex Mediterranean geological and climatic history.

## MATERIALS AND METHODS

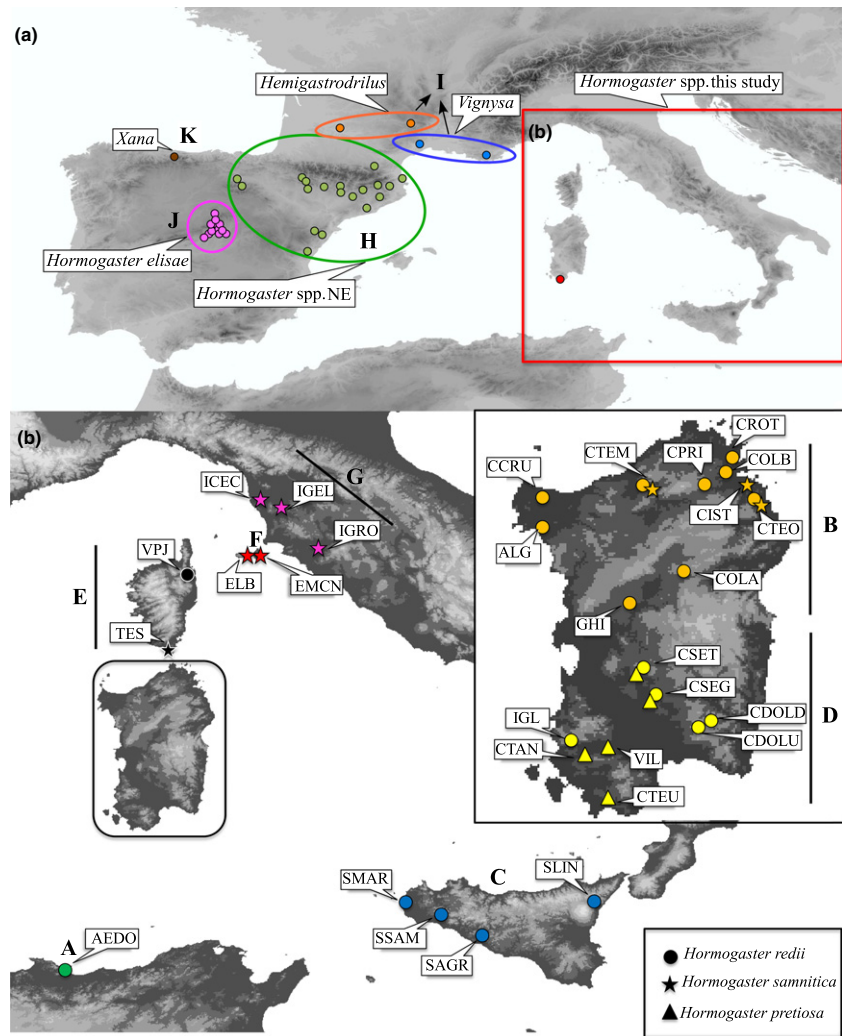
### Sampling and morphological study

The easternmost distribution area of Hormogastridae was sampled intensely: 24 localities yielded 192 mature individuals (Fig. 1, and see Appendix S1 in Supporting Information) from the Italian Peninsula, Sardinia, Sicily, Elba, Corsica and Algeria. Intense sampling effort in the Balearic Islands (Mallorca, Menorca and Ibiza) yielded no hormogastrid specimens, as documented also by other authors (Pérez-Losada *et al.*, 2011). Further collecting attempts were made at many sampling sites in the areas indicated by Omodeo (1956) and Omodeo & Rota (2008) in the Italian Peninsula, but were unsuccessful, probably because of the suboptimal climate conditions of the sampling year. Additional data presented in Novo *et al.* (2010, 2011, 2012) along the westernmost distribution area of hormogastrids were included for the general tree and ancestral area analyses, including six populations already collected by these authors in the area of emphasis of this paper (Fig. 1, Appendix S1).

All individuals were hand-collected, washed in distilled water and preserved in *c.* 96% ethanol. A portion of the integument (*c.* 25 mg) was cleaned under a stereo microscope to remove soil particles and preserved at  $-20^{\circ}\text{C}$  until DNA extraction. All specimens were dissected and examined morphologically for their taxonomic identification following Qiu & Bouché (1998) and the literature describing hormogastrid species in this area (see Novo, 2010, for full references).

### DNA extraction, gene amplification and sequencing

Total genomic DNA was extracted from the integument tissue sample using the DNeasy Tissue Kit (Qiagen, Valencia, CA, USA), eluting twice with 70  $\mu\text{L}$  of buffer. Molecular



**Figure 1** (a) Sampling localities for hormogastrid earthworms in previous studies (see Novo *et al.*, 2011). (b) Sampling localities for the present study in the eastern area of their distribution. Each species is represented by a different symbol. GPS coordinates and the number of individuals analysed are shown in Appendix S1. Geographical areas used for the ancestral range analyses are indicated by different letters.

markers included the mitochondrial genes cytochrome *c* oxidase subunit I (COI), 16S rRNA and Leu, Ala and Ser tRNAs, two nuclear ribosomal genes (complete 18S rRNA and a fragment of 28S rRNA) and two nuclear protein-encoding genes (histones H3 and H4). Primer sequences, polymerase chain reactions (PCR) and sequencing reactions are as in Novo *et al.* (2011). Chromatograms were visualized and assembled in SEQUENCHER 5.0 (Gene Codes Corp., Ann Arbor, MI, USA). All new sequences have been deposited in GenBank (Appendix S1).

### Unravelling the phylogenetic relationships

Two data sets were analysed. First, a data set combining the new data with all the previously available hormogastrid information ('large' data set hereafter; see Fig. 1) was examined to evaluate the position of the new samples in the broader phylogeny of the family. A concatenated matrix of 4558 bp was used for the analyses. Given that individuals from the same locality always clustered together, a single individual was used to represent each sampling site. The same outgroups as in Novo *et al.* (2011) were used. The second data set ('small' data

set hereafter) consisted of the subset of hormogastrid samples collected in the eastern area, targeting *H. samnitica* and *H. redii*. This data set included all individuals, but was restricted to the more variable genes (COI, 16S rRNA and histones H3 and H4), resulting in a concatenated matrix of 1965 bp. *Hormogaster pretiosa* was used as an outgroup.

Sequences for each individual gene were aligned in MAFFT 6 (Katoh, 2008; online server: <http://mafft.cbrc.jp/alignment/server/>) and CLUSTALX 2.0.12 (Thompson *et al.*, 1997) with default settings and concatenated with PHYUTILITY (Smith & Dunn, 2008). jMODELTEST 0.1.1 (Posada, 2008) was used to select the best-fitting evolutionary model under the Akaike information criterion (AIC; Akaike, 1973) for each individual gene in each data set: GTR+I+ $\Gamma$  for COI, 16S rRNA and 18S rRNA; GTR+ $\Gamma$  for 28S rRNA; HKY+I+ $\Gamma$  for histone H3 (HKY+I for the small data set); and HKY+I for histone H4 (GTR+I for the small data set). A Bayesian inference (BI) phylogeny of the partitioned data was estimated with MRBAYES 3.1.2 (Ronquist & Huelsenbeck, 2003).

In the concatenated matrices, unlinked nucleotide substitution models selected by jMODELTEST were specified for each gene fragment, and the nucleotide substitution estimates were

allowed to vary independently in each partition. Parameters in MRBAYES were set to 10 million generations and 10,000 trees were sampled every 1000th generation, using the default random-tree option to initiate the analysis. The analysis was run twice and 20% of the trees were discarded as burn-in after checking for convergence with TRACER 1.5 (Rambaut & Drummond, 2007). The remaining trees were combined to find the maximum posterior probability estimate. Maximum-likelihood (ML) analyses were performed with RAXML 7.2.7 (Stamatakis, 2006) as implemented in the CIPRES Science Gateway 3.3 (<http://www.phylo.org/index.php/portal/>) using a partitioned GTR+I+ $\Gamma$  model of sequence evolution and estimating the support for the resulting topologies by 100 rapid bootstrap replicates (Stamatakis *et al.*, 2008).

### Biogeographical reconstructions

Ultrametric trees were generated with BEAST 1.7 (Drummond *et al.*, 2012) for the large data set, using either the topology of the Bayesian phylogeny or that of the ML phylogeny as a constraint prior. We used relative timings by calibrating the trees with an age of 1 for the root for Hormogastridae. BEAST trees were generated after alignment of the large data set with MAFFT, using unlinked GTR+I+ $\Gamma$  models for each partition. The analysis was conducted under a Yule diversification model and an uncorrelated lognormal relaxed clock. For each alternative topology, 10 parallel runs were specified, each of which included 10 million generations, sampling every thousandth generation. Tree and log files were combined in LOGCOMBINER 1.7 by resampling at lower frequency (every 10,000th generation) and the results were visualized in TRACER 1.5 (Rambaut & Drummond, 2007). The final trees were generated by TREEANNOTATOR 1.7 with a burn-in of 2000.

Ultrametric trees generated from the two constrained priors were used for investigating the biogeographical reconstructions under two different approaches. Ancestral areas were inferred in RASP 2.1 (Ali *et al.*, 2012; Yu *et al.*, 2014), using areas coded as follows (Fig. 1): A, Africa; B, north and central Sardinia; C, Sicily; D, southern Sardinia; E, Corsica; F, Elba; G, Italian Peninsula; H, north-eastern Iberian Peninsula; I, southern France; J, central Iberian Peninsula; and K, Asturias, northern Spain. We used a Bayesian approach, with two runs of 10 million generations, sampling frequency of 1000, and burn-in of 2500. We also tested an ML approach through the dispersal–extinction–cladogenesis model (DEC; Ree *et al.*, 2005).

We initially considered dating the phylogeny but chose not to because no earthworm fossils are available and the only calibration points would be biogeographical. Using them would imply the assumption of vicariance, which has been criticized before (e.g. Kodandaramaiah, 2011). Moreover, there is controversy about using fixed rates for the evolution of this family (Novo *et al.*, 2012) and the only published rates for other earthworms were also calculated assuming vicariance.

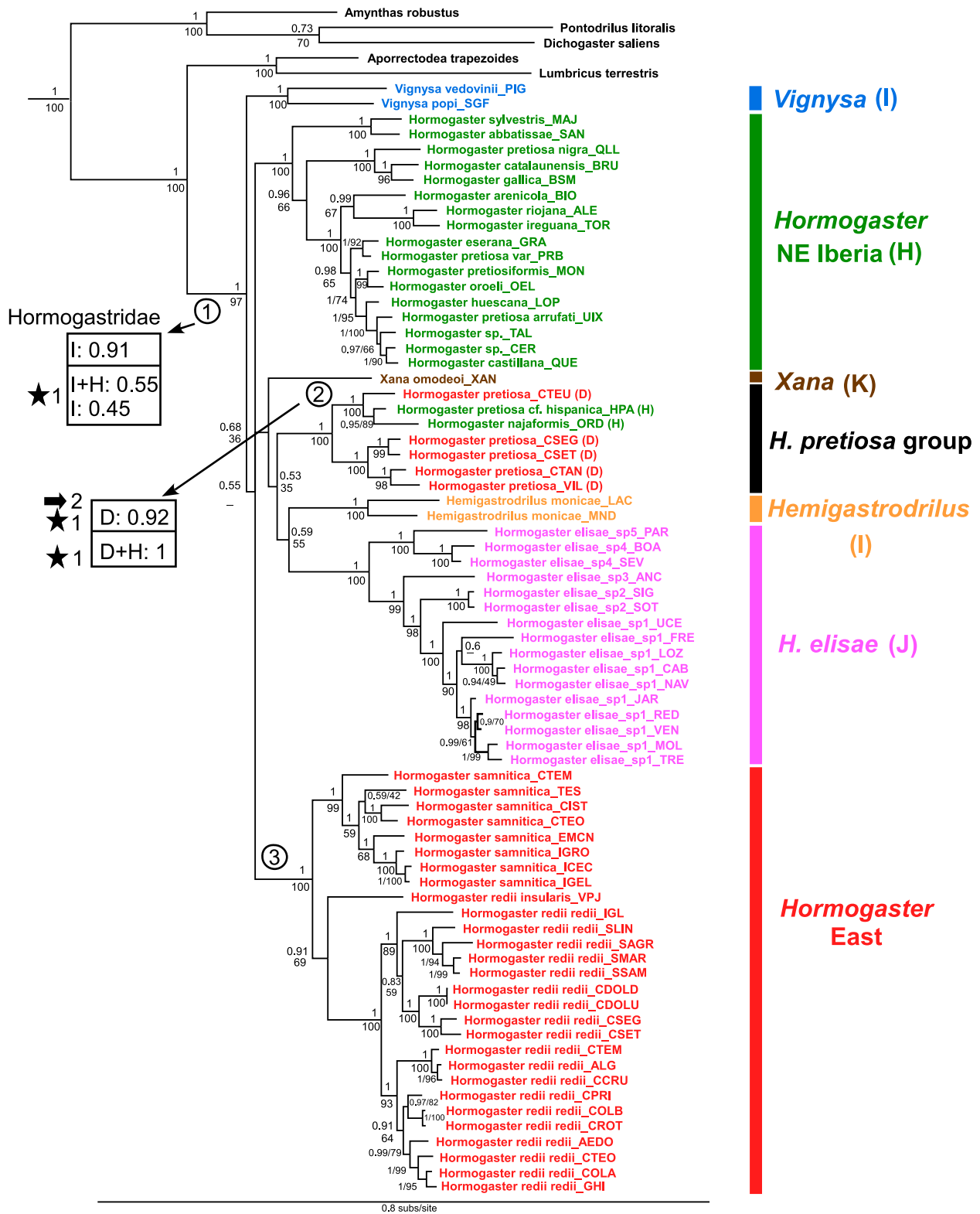
### Gene networks

Haplotype networks were constructed for the most variable genes (COI, 16S rRNA and histones H3 and H4) for *H. samnitica* and *H. redii*. A statistical parsimony procedure (Templeton *et al.*, 1992; Crandall *et al.*, 1994) with a 95% connection threshold was used in TCS 1.2.1 (Clement *et al.*, 2000) for the histones. Mitochondrial genes showed large divergence, most of the populations being unconnected, and these sequences were thus analysed using SPLITS TREE 4.10 (Huson & Bryant, 2006). Default settings were used to construct a neighbour-net with uncorrected pairwise (*p*)-distances.

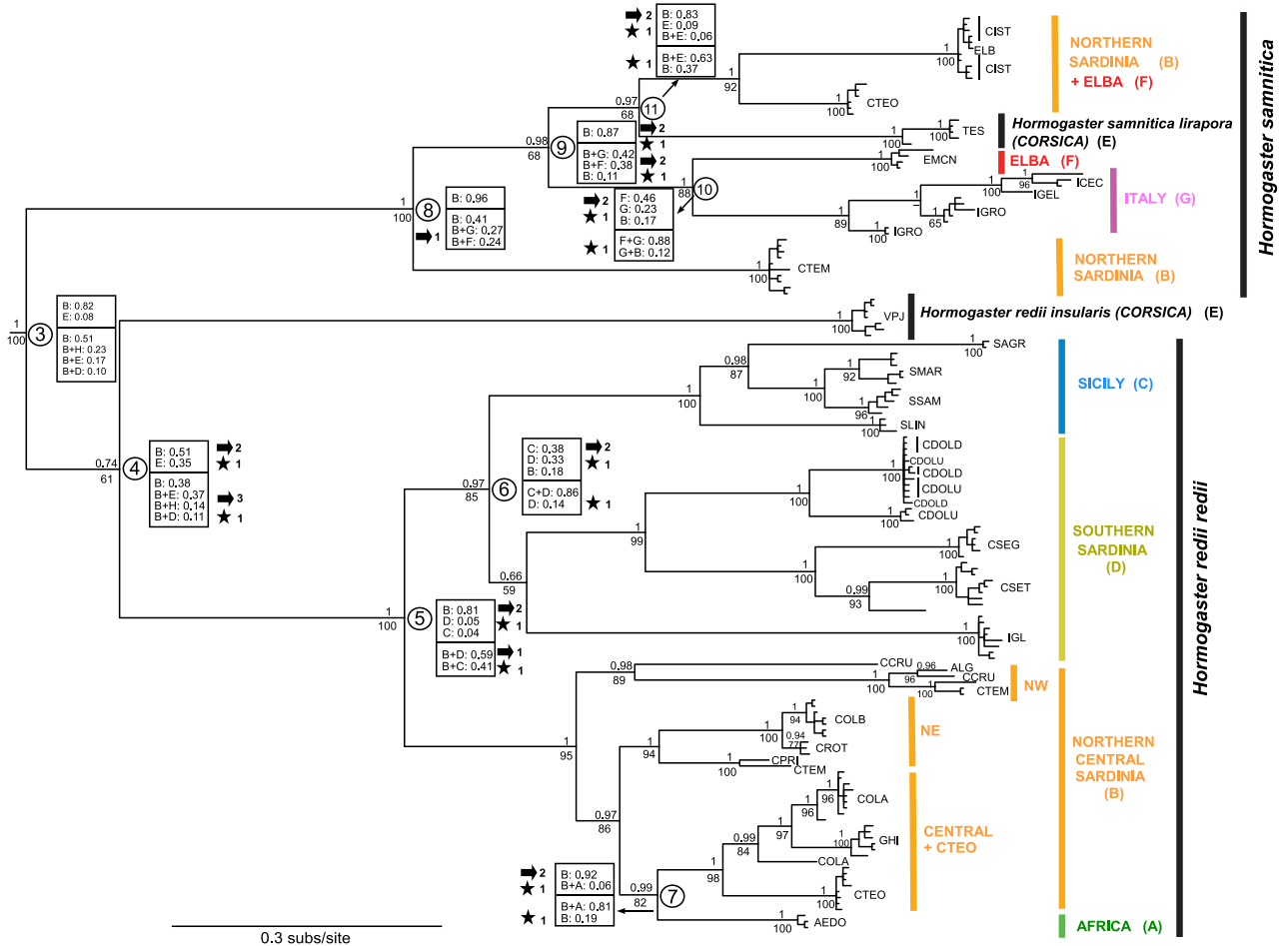
## RESULTS

### Phylogenetic relationships

The Bayesian tree for the large data set (Fig. 2) recovered multiple broadly recognized clades: *Vignysa*, *Hemigastrodrius*, *Hormogaster elisae*, *Hormogaster* from the north-eastern Iberian Peninsula, the *Hormogaster pretiosa* group, and easternmost *Hormogaster*. These clades are consistent with previous results of hormogastrid phylogeny (e.g. Novo *et al.*, 2011), but their inter-relationships remain unstable even after the addition of more populations. Differences also exist between the MAFFT BI tree and the trees produced from the CLUSTALX alignment and/or ML analysis (bootstrap values for the congruent relationships after ML with MAFFT are shown in Fig. 2). Therefore, whereas the placement of *Hormogaster pretiosa* from southern Sardinia closest to *H. pretiosa* from the Iberian Peninsula is supported by all analyses, the sister clade of the other hormogastrids in the eastern area of the family's distribution (*H. samnitica* and *H. redii*) remains ambiguous. *Vignysa* was the sister clade of those species in the ML tree (with low support; tree not shown) but BI showed *Vignysa* to be the sister group to the rest of the family (again without support; Fig. 2). In contrast, phylogenetic analyses of the small data set (including only hormogastrids from the easternmost area) resulted in a highly congruent topology among alignment strategies and phylogenetic methodologies. The Bayesian hypothesis after alignment with MAFFT (Fig. 3) placed *Hormogaster redii insularis* from Corsica as the sister group to *H. redii redii*. The Sardinian populations were not monophyletic; the southern populations were closely related to the Sicilian ones whereas the northern and central populations were paraphyletic with respect to the African specimens. Interestingly, all the individuals from each population clustered together, with the exception of Tempio (population CTEM) in northern Sardinia, whose individuals were most closely related either to those from the north-eastern area (one individual) or those from the north-western area. Analyses of *H. samnitica* revealed that individuals from Tempio formed the sister clade to the remaining populations. Again, Sardinian specimens did not form a monophyletic group, with the north-easternmost populations (specifically Istanta, CIST) being



**Figure 2** Bayesian consensus tree based on the large data set for hormogastrids after MAFFT alignment. Posterior probability and bootstrap support values are shown. Dashes indicate different relationships in the maximum likelihood topology. Some outgroups were omitted for better visualization. Geographical areas used for biogeographical reconstruction analyses are shown in parentheses (see Fig. 1) and circled numbers indicate the nodes of interest (see Table 1). Results from Bayesian ancestral-area reconstruction (above) and dispersal–extinction–cladogenesis (DEC) analyses (below) are indicated at nodes where applicable and probabilities of each area are shown. Number of dispersal (arrow) and vicariance (star) events found by the reconstructions are presented. Nodes and ancestral areas of the *Hormogaster* East clade (node 3) are shown in Fig. 3.



**Figure 3** Bayesian tree based on the small data set for hormogastrids. Posterior probability and bootstrap support values are shown. Dashes indicate different relationships in the maximum likelihood topology. Some outgroups were omitted for better visualization. Geographical areas used for biogeographical reconstruction analyses are shown in parentheses (see Fig. 1) and circled numbers indicate the nodes of interest (see Table 1). Results from Bayesian ancestral-area reconstruction (above) and dispersal–extinction–cladogenesis (DEC) analyses (below) are indicated at nodes where applicable and probabilities of each area are shown. Number of dispersal (arrow) and vicariance (star) events found by the reconstructions are presented. Support values for within-population relationships are not shown for sake of clarity.

most closely related to some Elba specimens (population ELB). The other individuals from Elba (EMCN) clustered with specimens from the Italian Peninsula.

**Biogeographical reconstructions**

The alternative ancestral reconstruction analyses in RASP showed similar results for the two tested topologies (ancestral areas and probabilities from Bayesian, ML and DEC analyses are shown in Table 1; see also Figs 2 & 3). In total, Bayesian analyses detected 27 dispersals and 14 vicariance events, and DEC analyses found 18 and 12, respectively (for specific results for selected nodes, see Table 1). All analyses agree on an origin of Hormogastridae between southern France and the north-eastern Iberian Peninsula. North–central Sardinia seems to have been a key area for diversification of easternmost Hormogastridae (*H. redii* and *H. samnitica*).

**Gene networks**

Networks for both mitochondrial genes represented the same relationships and proportionally the same distances and therefore only the SPLITSTREE network for COI is presented. For *H. redii* (Fig. 4), Corsica was again identified as the most distinct population, as shown by the mitochondrial genes. In the case of the histones, southern Sardinia and Sicily were equally distant. These areas are highly differentiated, although some populations share haplotypes for histone H4. Iglesias (IGL) in Sardinia and Linguaglossa (SLIN) were revealed as interesting populations that may have constituted the connection between the two islands. Also, the relationship of the African population with the central area of Sardinia is confirmed here. Tempio (CTEM) presented haplotypes in different groups, as shown by the COI network and in the Bayesian tree (Fig. 3). In the case of *H. samnitica* (Fig. 5),

**Table 1** Ancestral areas for the clades of interest in hormogastrid phylogeny (see Figs 1–3) recovered from Bayesian inference topology. Ancestral areas were recovered in *RASP* by Bayesian and dispersal–extinction–cladogenesis (DEC) analyses. The number of dispersal (Disp) and vicariance (Vic) events recovered by each analysis is shown. The ML topology differed only at Node 1, which the Bayesian reconstruction showed an origin of H ( $P = 0.74$ ) and I ( $P = 0.07$ ) and DEC recovered I+H ( $P = 1$ ). Prob, probability values associated with the inference of the area and the dispersal and vicariance events hypothesis (0, lowest; 1, highest).

Node	Bayesian				DEC					
	Area	Prob	Disp	Vic	Prob	Area	Prob	Disp	Vic	Prob
1	I	0.91	0	0	0.53	I + H	0.55	1	0	0.55
						I	0.45			
2	D	0.92	2	1	0.88	D + H	1.00	0	1	1.00
3	B	0.82	0	0	0.40	B	0.51	0	0	0.08
	E	0.08				B + H	0.23			
						B + E	0.17			
						B + D	0.10			
4	B	0.51	2	1	0.41	B	0.38	3	1	0.22
	E	0.35				B + E	0.37			
						B + H	0.14			
						B + D	0.11			
5	B	0.81	2	1	0.30	B + D	0.59	1	1	0.51
	D	0.05				B + C	0.41			
	C	0.04								
6	C	0.38	2	1	0.36	D + C	0.86	0	1	0.86
	D	0.32				D	0.14			
	B	0.18								
7	B	0.92	2	1	0.90	B + A	0.81	0	1	0.81
	B+ A	0.06				B	0.19			
8	B	0.96	0	0	0.83	B	0.41	1	0	0.17
						B + G	0.27			
						B + F	0.24			
9	B	0.87	2	1	0.33	B + G	0.42	2	1	0.23
						B + F	0.38			
						B	0.11			
10	F	0.46	2	1	0.45	F + G	0.88	0	1	0.88
	G	0.23				G + B	0.12			
	B	0.17								
11	B	0.83	2	1	0.82	B + E	0.63	0	1	0.63
	E	0.09				B	0.37			
	B + E	0.06								

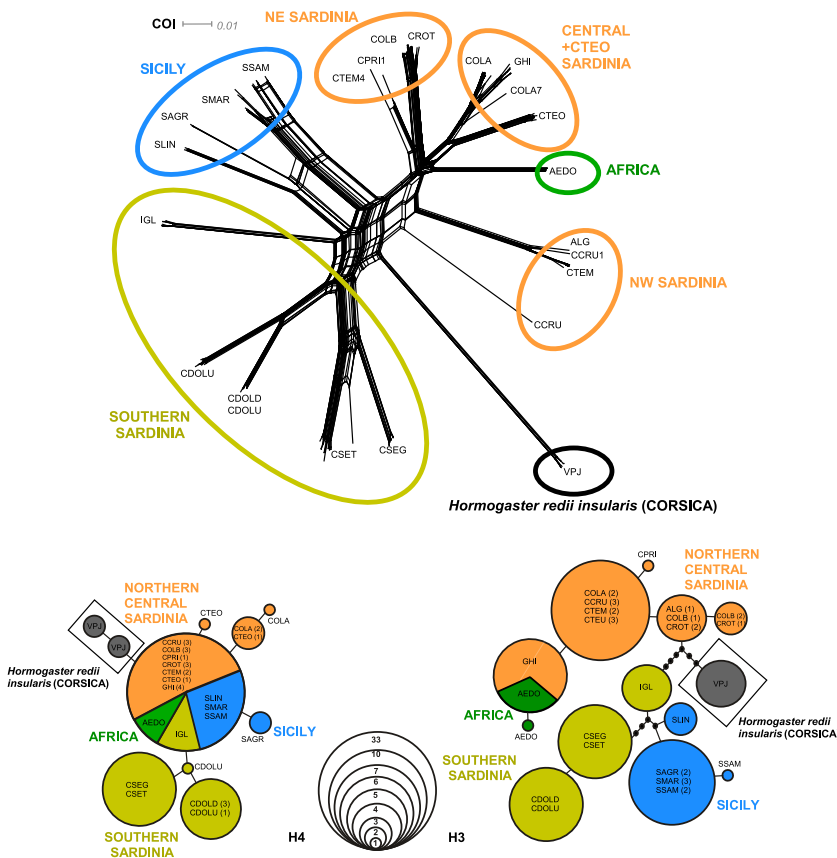
the heterogeneity of the Sardinian populations was again clear, and samples from Italy, Elba (EMCN) and Corsica were closely related, although mitochondrial genes (see also Fig. 3) showed Corsica as being close to some of the Sardinian populations. Samples from Elba collected by P. Omodeo (ELB) clustered with samples from Istana (CIST) in north-eastern Sardinia according to the mitochondrial genes, and sequences for the nuclear genes were not available for this population because PCR amplification was unsuccessful. All individuals of *H. samnitica* shared the same haplotype for histone H4.

## DISCUSSION

The study of the distributions and diversification patterns of animals that ride on continental plates is particularly interesting in the context of a geologically complex setting such as the Mediterranean Basin. Earthworms, particularly those with a Mediterranean distribution, have been previously

highlighted in the context of classical zoology (Michaelsen, 1922; Omodeo & Rota, 2008) and more recently in molecular studies (e.g. Cobolli Sbordoni *et al.*, 1992; Buckley *et al.*, 2011; Novo *et al.*, 2011) as models for biogeographical study. Our results suggest a complex evolutionary history for hormogastrid earthworms, showing the possibility of many dispersal and vicariance events that reflect the complexity of this biogeographical scenario. *Hormogaster pretiosa*, whose distribution in Sardinia is limited to the south-west, is closely related to the Iberian morphotypes of *H. pretiosa*, suggesting an ancient relationship and subsequent cladogenesis during plate movement. Bayesian and DEC models both detect a vicariance event in this node (and in other splits that may be due to plate separation; see below). Bayesian analyses postulate an additional earlier dispersal event, which is compatible with this scenario.

Other terrestrial animals and plants have been shown to follow the fate of the Mediterranean microplates (e.g. Magri *et al.*, 2007; Mansion *et al.*, 2008; Muriene *et al.*, 2010;



**Figure 4** Networks for *Hormogaster redii* populations. Top, COI network constructed with SPLITSTREE4. Code for localities is shown at the tips. Branch lengths are proportional to genetic distance. Bottom, TCS networks based on histones. The size of circles reflects haplotype frequency as shown by the legend. Each branch represents one mutational step; each dot represents an unsampled haplotype. When the proportion of individuals differs among populations, it is indicated by their number. Scale bar refers to uncorrected pairwise distance.

Bidegaray-Batista & Arnedo, 2011), and Ketmaier & Caccone (2013) suggest in their review of organisms with peri-Tyrrhenian distributions that diversification in the area was mainly driven by vicariance. The fact that no hormogastrids have been found in the Balearic Islands reinforces the idea that the homeland of this family excludes the Baetic–Rifan system, as suggested by Omodeo & Rota (2008).

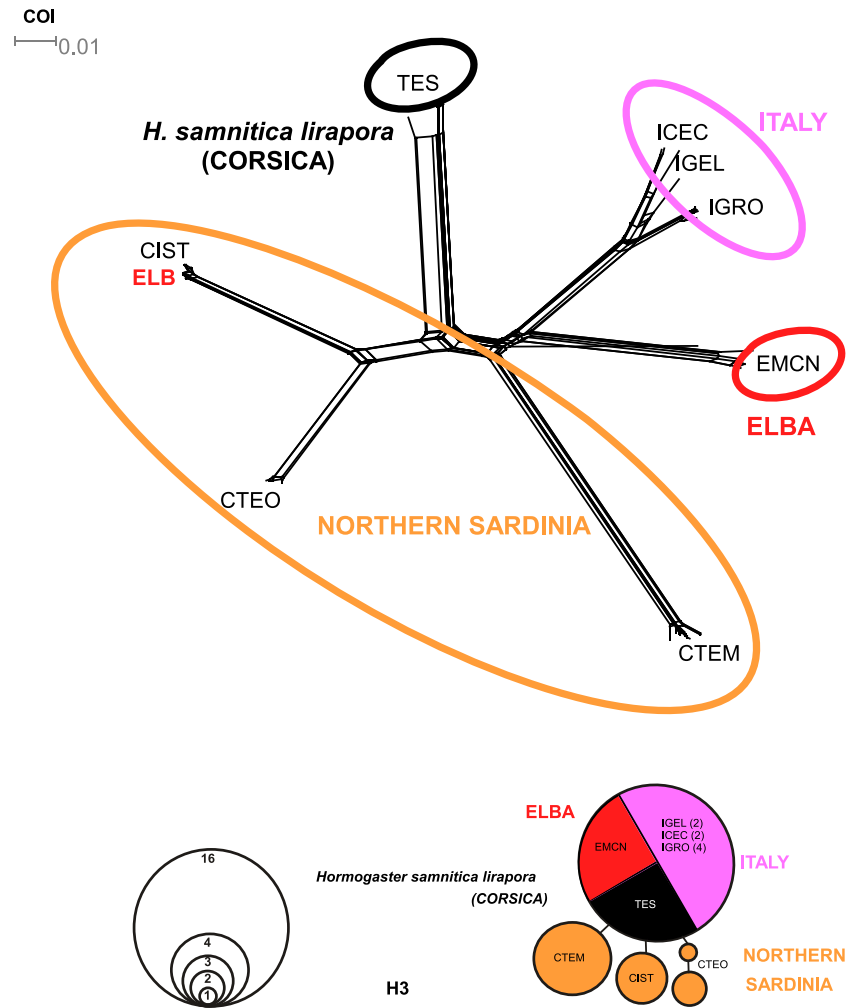
The origins of the *H. redii* and *H. samnitica* groups (*Hormogaster* East; clade 6 in Fig. 2) remain unclear and although they could be linked to Provence (France), as suggested by Omodeo & Rota (2008) and Novo *et al.* (2011), our results leave the matter unsettled. However, the biogeographical reconstructions suggest an origin near southern France and north–central Sardinia for this clade, as well as for the subsequent split between *H. redii insularis* and *H. redii redii*, the diversification of the latter and the origin of *H. samnitica*.

*Hormogaster samnitica* is limited to north-eastern Sardinia (and Maddalena Archipelago), southern Corsica, the Tuscan Archipelago (including Elba) and the Tuscan mainland (Omodeo & Rota, 2008). Our results highlight Sardinia as a complex area for this species as shown initially by the high haplotypic diversity. The population that shows the most ancient connection with the rest of the *H. samnitica* group is the one from north–central Sardinia (CTEM). This population is outside the area of Gallura, where the remaining sampled populations from Sardinia were collected for this species. Gallura has been isolated from the rest of Sardinia

through habitat degradation, and was recognized as a unique zoogeographical area by Rota (1992), who suggested that most of Sardinia is uninhabitable by *H. samnitica* because of geological changes that affected the island (uplift, subsidence and faulting); Gallura therefore retains characteristics of the Miocene western Mediterranean flora. This population (CTEM) could therefore be a remnant from the past, and the ancestral area for *H. samnitica* seems to have been located near it. A similar isolation pattern is shown in some groups of beetles (Jeannel, 1942, 1961). In the histone H4 network, the differentiation of both populations from north-eastern Sardinia (CTEO and CIST) is also evident, agreeing with the results from *H. redii* (see below), where the CTEO population is more closely related to populations in the central area than those in the north-east. It is remarkable to see the congruence of our tree with Rota (1992), where it is stated that the faunas from Maddalena and northern Sardinia (represented here by CIST and CTEO) are more similar to that of the Tuscan Archipelago (here, ELB) than to that of Corsica (here, TES). The remaining areas of Sardinia would contain the most distant fauna (here, CTEM).

It seems that the fauna in this area may have been influenced by the complex series of connections and retractions of land-masses. Around 24 Ma, Corsica was connected to the mainland, although later land-bridges with the mainland and with Sardinia before the MSC have also been suggested (Andeweg, 2002). Moreover, sea-level oscillations continued





**Figure 5** Networks for *Hormogaster samnitica*. Top, COI network constructed with SPLITSTREE4. Code for localities is shown at the tips. Branch lengths are proportional to genetic distance. Bottom, rcs networks based on histones. The size of circles reflects haplotype frequency as shown by the legend. When the proportion of individuals differs among populations, it is indicated by their number. No individuals from ELB are included in the histone H3 network because of amplification problems. Scale bar refers to uncorrected pairwise distance.

from the Miocene until well into the Pleistocene (5.7–0.23 Ma) creating other land connections between the Sardo-Corsican System and the mainland, via Elba (Weiss & Ferrand, 2007). This scenario has given multiple opportunities for this species to disperse and diversify, as shown by our analyses. North–central Sardinia appeared as the origin of *H. samnitica* in the biogeographical reconstructions for the RASP analyses. We find two different situations regarding the Elba populations. On the one hand, the population we collected in eastern Elba coalesces with Italy. In this case, Elba could have given rise to the Italian populations according to the ancestral reconstruction analyses, although with low probability. On the other hand, the specimens collected by P. Omodeo (in the western area, ELB) are closely related to earthworms from north-eastern Sardinia (CIST) (Figs 3 & 5). Although we could not amplify regions other than COI for these specimens, the analyses showed that this divergence is relatively recent. It could be attributed either to late land connections or to some sort of human-mediated dispersal. Omodeo & Rota (2008) suggested migration of fauna towards Corsica and Sardinia during the Pleistocene ice age, but they rejected the opposite route, as supported here. This would highlight the outstanding resistance of hormogastrids

to the climatic vicissitudes that permitted their survival and interchange between formerly connected land-masses.

*Hormogaster redii* is more widespread than *H. samnitica*. It is the dominant hormogastrid species in Sardinia and it is present in Sicily, North Africa (from Tunisia to the Edough promontory in Algeria), the Tyrrhenian side of Italy (from south of the river Arno to 40° S) and the northern tip of Corsica (Omodeo & Rota, 2008), from where *H. redii insularis* was described. The separation of this population (VPJ) from the remainder occurred soon after the diversification of the easternmost hormogastrid species (*H. samnitica* and *H. redii*). *Hormogaster redii* is another example of the biogeographical complexity of Sardinia. Populations from southern Sardinia cluster with individuals from Sicily. Prior to the opening of the Mediterranean, Calabria was adjacent to Sardinia (Álvarez *et al.*, 1974) and afterwards, it constituted part of what is now Sicily, and specifically the eastern area of Sicily where the population (SLIN) is located. This population seems to represent the connection between the two islands, as shown by the networks (Fig. 4) and by its position in the trees, where it is presented as the sister group of the other populations in Sicily. Both RASP analyses detected a vicariance event (and dispersal with lower

probability) and show Sicily and southern Sardinia as the ancestral area for this clade.

The differentiation of the populations in the south-western area of Sardinia and the unique history of the populations within the island are not surprising given its active geomorphological history. The south-western corner has been regarded as a different biogeographical region by some (Bacchetta *et al.*, 2007), and was separated from the rest of Sardinia by a sea arm that later filled with alluvial deposits (Omodeo & Rota, 2008). This may explain the different genetic composition of *H. redii* in this area, closest to Sicily, and also the endemism of *H. pretiosa*, limited to south-western Sardinia and the north-eastern Iberian Peninsula.

The other main lineage within *H. redii* is the one composed of specimens from north-central Sardinia as well as the African population. The split of this African sample from populations in central and north-eastern Sardinia is difficult to comprehend, but it could be related to the MSC episode, although our sampling is not optimal for testing such a hypothesis. A colonization route through Sicily and Tunisia has already been proposed by Omodeo & Rota (2008). We were unable to collect specimens of *H. redii* from Italian continental populations or Tunisia, which could have been useful for detecting additional connections during the MSC.

The split between the north-western populations and the north-eastern and central Sardinian populations could be related to the separation *c.* 24 Ma of eastern and western areas of Sardinia by the Sardinian Rift (Andeweg, 2002). An outstanding example of this separation is the population CTEM (in north-central Sardinia). Most of its individuals were included within the north-western group, but one clustered with the north-eastern group. This could indicate secondary contact between lineages within this locality. We also recovered the isolation of the north-eastern populations within the area of Gallura (CPRI, CROT, COLB), considered a relictual area because of its physical characteristics and vegetation (Rota, 1992), as highlighted above.

This study represents a good example of how earthworms can be used for tracing the complex biogeographical history of the Mediterranean region. Despite the lack of fossils to conduct proper molecular dating analyses, different sources of evidence indicate that hormogastrids are undoubtedly linked to the geological fate of the land-masses they inhabit, serving as models to understand general biogeographical patterns in soil fauna. Quoting Bidegaray-Batista & Arnedo (2011), earthworms seem to be 'gone with the plate'.

## ACKNOWLEDGEMENTS

We would like to express our gratitude to Daniel Chana and Hans Pech for their assistance with fieldwork, Pietro Omodeo for providing specimens from Elba and the Ter River and his guidance in the sampling, Kamel Bazri for providing African samples, Pierfrancesco Sechi for his advice on BEAST analyses and John Morgan for proofreading the manuscript. We also acknowledge two anonymous referees and Sam

James for their constructive criticism that helped to improve earlier versions of this manuscript; Christine Palmer is acknowledged for language editing. M.N. was supported by a grant from the Fundación Caja Madrid and a Marie Curie Fellowship. R.F. was supported by a Postdoctoral Fellowship from Fundación Ramón Areces. This research was funded by project CGL2010-16032 from the Spanish Government.

## REFERENCES

- Akaike, H. (1973) Information theory and an extension of the maximum likelihood principle. *Proceedings of the Second International Symposium on Information Theory* (ed. by B.N. Petrov and F. Csaki), pp. 267–281. Akademiai Kiado, Budapest.
- Ali, S.S., Yu, Y., Pfosser, M. & Wetschnig, W. (2012) Inferences of biogeographical histories within subfamily Hyacinthoideae using S-DIVA and Bayesian binary MCMC analysis implemented in RASP (Reconstruct Ancestral State in Phylogenies). *Annals of Botany*, **109**, 95–107.
- Álvarez, W., Coccozza, T. & Wezel, F.C. (1974) Fragmentation of the Alpine orogenic belt by microplate dispersal. *Nature*, **248**, 309–314.
- Andeweg, B. (2002) *Cenozoic tectonic evolution of the Iberian Peninsula: effects and causes of changing stress fields*. PhD Thesis, Vrije Universiteit Amsterdam, Amsterdam.
- Bacchetta, G., Mandis, G. & Pontecorvo, C. (2007) Contribution to the knowledge of the endemic vascular flora of Sulcis (SW Sardinia - Italy). *Bocconea*, **21**, 155–166.
- Bidegaray-Batista, L. & Arnedo, M.A. (2011) Gone with the plate: the opening of the Western Mediterranean basin drove the diversification of ground-dweller spiders. *BMC Evolutionary Biology*, **11**, 317.
- Bouché, M.B. (1972) *Lombriciens de France: écologie et systématique*. INRA, Paris.
- Buckley, T.R., James, S., Allwood, J., Bartlam, S., Howitt, R. & Prada, D. (2011) Phylogenetic analysis of New Zealand earthworms (Oligochaeta: Megascolecidae) reveals ancient clades and cryptic taxonomic diversity. *Molecular Phylogenetics and Evolution*, **58**, 85–96.
- Clement, M., Posada, D. & Crandall, K.A. (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.
- Cobolli Sbordoni, M., De Mattheis, E., Alonzi, A., Mattocchia, M., Omodeo, P. & Rota, E. (1992) Speciation, genetic divergence and palaeogeography in the Hormogastridae. *Soil Biology and Biochemistry*, **24**, 1213–1221.
- Crandall, K.A., Templeton, A.R. & Sing, C.F. (1994) Intra-specific cladogram estimation: problems and solutions. *Models in phylogeny reconstruction* (ed. by R.W. Scotland, D.J. Siebert and D.M. Williams), pp. 273–297. Clarendon Press, Oxford, UK.
- Cunningham, C.W. & Collins, T. (1998) Beyond area relationships: extinction and recolonization in marine molecular biogeography. *Molecular approaches to ecology and*

- evolution (ed. by R. DeSalle and B. Schierwater), pp. 297–322. Birkhäuser, Basel, Switzerland.
- Drummond, A.J., Suchard, M.A., Xie, D. & Rambaut, A. (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.
- Gates, G.E. (1966) Requiem – for megadrile utopias. A contribution toward the understanding of the earthworm fauna of North America. *Proceedings of the Biological Society of Washington*, **79**, 239–254.
- Gentile, G., Campanaro, A., Carosi, M., Sbordoni, V. & Argano, R. (2010) Phylogeography of *Helleria brevicornis* Ebner 1868 (Crustacea, Oniscidea): old and recent differentiations of an ancient lineage. *Molecular Phylogenetics and Evolution*, **54**, 640–646.
- Huson, D.H. & Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.
- James, S.W. (2004) Planetary processes and their interactions with earthworm distributions and ecology. *Earthworm ecology*, 2nd edn (ed. by C.A. Edwards), pp. 53–62. CRC Press, Boca Raton, FL.
- Jeannel, R. (1942) *La genèse des faunes terrestres*. Presses Universitaires de France, Paris.
- Jeannel, R. (1961) Quelques données sur le peuplement de la Corse et de la Sardaigne. *Le Peuplement de Îles Méditerranéennes et le Problème de l'Insularité. XCIV Colloques Internationaux du Centre National de la Recherche Scientifique, Banyuls-sur-Mer*. Editions CNRS, Paris.
- Katoh, T. (2008) Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics*, **9**, 286–298.
- Ketmaier, V. & Caccone, A. (2013) Twenty years of molecular biogeography in the West Mediterranean islands of Corsica and Sardinia: lessons learnt and future prospects. *Current progress in biological research* (ed. by M. Silva-Opps), pp. 71–93. Intech, Rijeka, Croatia. doi:10.5772/55458.
- Ketmaier, V., Giusti, F. & Caccone, A. (2006) Molecular phylogeny and historical biogeography of the land snail genus *Solatopupa* (Pulmonata) in the peri-Tyrrhenian area. *Molecular Phylogenetics and Evolution*, **39**, 439–451.
- Kodandaramaiah, U. (2011) Tectonic calibrations in molecular dating. *Current Zoology*, **57**, 116–124.
- Magri, D., Fineschi, S., Bellarosa, R., Buonamici, A., Sebastiani, F., Schirone, B., Simeone, M.C. & Vendramin, G.G. (2007) The distribution of *Quercus suber* chloroplast haplotypes matches the palaeogeographical history of the western Mediterranean. *Molecular Ecology*, **16**, 5259–5266.
- Mansion, G., Rosenbaum, G., Schoenenberger, N., Bacchetta, G., Rosselló, J.A. & Conti, E. (2008) Phylogenetic analysis informed by geological history supports multiple, sequential invasions of the Mediterranean Basin by the angiosperm family Araceae. *Systematic Biology*, **57**, 269–285.
- Michaelsen, W. (1922) Die Verbreitung der Oligochaeten im Lichte der Wegenerischen Theorie der Kontinenten-Verschiebung und andere Fragen zur Stammesgeschichte und Verbreitung dieser Tiergruppe. *Verhandlungen des Vereins für Naturwissenschaftliche Unterhaltung zu Hamburg*, **29**, 45–79.
- Murienne, J., Karaman, I. & Giribet, G. (2010) Explosive evolution of an ancient group of Cyphophthalmi (Arachnida: Opiliones) in the Balkan Peninsula. *Journal of Biogeography*, **37**, 90–102.
- Novo, M. (2010) *Algunos aspectos de la biología evolutiva de la familia Hormogastridae Michaelsen, 1900: filogenia, filogeografía y selección sexual*. PhD Thesis. Universidad Complutense de Madrid, Madrid.
- Novo, M., Almodóvar, A., Fernández, R., Trigo, D. & Díaz Cosín, D.J. (2010) Cryptic speciation of hormogastrid earthworms revealed by mitochondrial and nuclear data. *Molecular Phylogenetics and Evolution*, **56**, 507–512.
- Novo, M., Almodóvar, A., Fernández, R., Giribet, G. & Díaz Cosín, D.J. (2011) Understanding the biogeography of a group of earthworms in the Mediterranean basin – the phylogenetic puzzle of Hormogastridae (Clitellata: Oligochaeta). *Molecular Phylogenetics and Evolution*, **61**, 125–135.
- Novo, M., Almodóvar, A., Fernández, R., Trigo, D., Díaz Cosín, D.J. & Giribet, G. (2012) Appearances can be deceptive: different diversification patterns within a group of Mediterranean earthworms (Oligochaeta: Hormogastridae). *Molecular Ecology*, **21**, 3776–3793.
- Omodeo, P. (1956) Sistematica e distribuzione geografica degli Hormogastrinae (Oligocheti). *Archivio Botanico e Biogeografico Italiano*, **32**, 159–177.
- Omodeo, P. (1963) Distribution of the terricolous oligochaetes on the two shores of the Atlantic. *North Atlantic biota and their history* (ed. by A. Love and D. Love), pp. 127–151. Pergamon Press, New York.
- Omodeo, P. (2000) Evolution and biogeography of megadriles (Annelida, Clitellata). *Italian Journal of Zoology*, **67**, 179–201.
- Omodeo, P. & Rota, E. (2008) Earthworm diversity and land evolution in three Mediterranean districts. *Proceedings of the California Academy of Sciences* (Suppl. 1), **59**, 65–83.
- Pérez-Losada, M., Breinholt, J.W., Porto, P.G., Aira, M. & Domínguez, J. (2011) An earthworm riddle: systematics and phylogeography of the Spanish lumbricid *Postandrilus*. *PLoS ONE*, **6**, e28153.
- Posada, D. (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253–1256.
- Qiu, J.P. & Bouché, M.B. (1998) Contribution to the taxonomy of Hormogastridae (Annelida: Oligochaeta) with description of new species from Spain. *Documents Pédozoologiques et Intégréologiques*, **4**, 164–177.
- Rambaut, A. & Drummond, A.J. (2007) *Tracer v1.4*. University of Edinburgh, Edinburgh, UK. Available at: <http://beast.bio.ed.ac.uk/Tracer>.

- Ree, R.H., Moore, B.R., Webb, C.O. & Donoghue, M.J. (2005) A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. *Evolution*, **5**, 2299–2311.
- Ronquist, F. & Huelsenbeck, J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.
- Rota, E. (1992) New data on the earthworm fauna of Gallura (Northeast Sardinia). *Soil Biology and Biochemistry*, **24**, 1383–1388.
- Salvo, G., Ho, S.Y.W., Rosenbaum, G., Ree, R. & Conti, E. (2010) Tracing the temporal and spatial origins of island endemics in the Mediterranean region: a case study from the citrus family (*Ruta* L., Rutaceae). *Systematic Biology*, **59**, 705–722.
- Smith, S.A. & Dunn, C.W. (2008) Phyutility: a phyloinformatics tool for trees, alignments, and molecular data. *Bioinformatics*, **24**, 715–716.
- Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, **22**, 2688–2690.
- Stamatakis, A., Hoover, P. & Rougemont, J. (2008) A rapid bootstrap algorithm for the RAxML Web servers. *Systematic Biology*, **57**, 758–771.
- Templeton, A.R., Crandall, K.A. & Sing, C.F. (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data III. Cladogram estimation. *Genetics*, **132**, 619–633.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. & Higgins, D.G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **24**, 4876–4882.
- Weiss, S. & Ferrand, N. (eds) (2007) *Phylogeography of Southern European refugia: evolutionary perspectives on the origins and conservation of European biodiversity*. Springer, Berlin.
- Yu, Y., Harris, A.J. & He, X.-J. (2014) *RASP (Reconstruct Ancestral State in Phylogenies)*. Sichuan University, Chengdu, China. Available at: <http://mnh.scu.edu.cn/soft/blog/RASP>.
- Zangari, F., Cimmaruta, R. & Nascetti, G. (2006) Genetic relationships of the western Mediterranean painted frogs based on allozymes and mitochondrial markers: evolutionary and taxonomic inferences (Amphibia, Anura, Discoglossidae). *Biological Journal of the Linnean Society*, **87**, 515–536.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

**Appendix S1** Sampled localities, number of individuals sequenced and GenBank accession numbers.

## BIOSKETCHES

**Marta Novo** is currently a Marie Curie Fellow at Cardiff University (UK), and previously worked at the Museum of Comparative Zoology (Harvard University, USA), where this work was conducted. She is interested in the evolutionary biology of invertebrates, particularly their phylogeny, phylogeography, reproduction and adaptation to extreme environments.

The Research Group of Soil Zoology (Complutense University of Madrid) focuses on the study of earthworms through a combination of taxonomic, ecological and molecular approaches, and the Giribet Lab (Harvard University) focuses on the systematics, evolution and biogeography of invertebrates using morphological and molecular analyses.

Author contributions: M.N. and D.J.D.C. conceived the research; M.N. and R.F. collected specimens; M.N. and D.T. generated the data; M.N., D.F.M., and R.F. analysed the data; all authors discussed results; M.N. led the writing, which was revised by all authors.

---

Editor: Melodie McGeoch