Understanding the biogeography of a group of earthworms in the Mediterranean basin—The phylogenetic puzzle of Hormogastridae (Clitellata: Oligochaeta)

Marta Novo a,b,⁎, Ana Almodóvar a, Rosa Fernández a, Gonzalo Giribet b, Darío J. Díaz Cosín a

a Departamento de Zoología y Antropología Física, Facultad de Biología, Universidad Complutense de Madrid. C/José Antonio Nováis, 2, 28040 Madrid, Spain. Fax: +34 91 394 49 47.
b Museum of Comparative Zoology, Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA

⁎ Corresponding author at: Departamento de Zoología y Antropología Física, Facultad de Biología, Universidad Complutense de Madrid. C/José Antonio Nováis, 2, 28040 Madrid, Spain. Fax: +34 91 394 49 47.
E-mail address: mnovoa@bio.ucm.es (M. Novo).

Available online 13 June 2011

Abstract

Traditional earthworm taxonomy is hindered due to their anatomical simplicity and the plasticity of the characteristics often used for diagnosing species. Making phylogenetic inferences based on these characters is more than difficult. In this study we use molecular tools to unravel the phylogeny of the clitellate family Hormogastridae. The family includes species of large to mid-sized earthworms distributed almost exclusively in the western Mediterranean region where they play an important ecological role. We analyzed individuals from 46 locations spanning the Iberian Peninsula to Corsica and Sardinia, representing the four described genera in the family and 20 species. Molecular markers include mitochondrial regions of the cytochrome c oxidase subunit 1 gene (COI), 16S rRNA and tRNAs for Leu, Ala, and Ser, two nuclear ribosomal genes (nearly complete 18S rRNA and a fragment of 28S rRNA) and two nuclear protein-encoding genes (histones H3 and H4). Analyses of the data using different approaches corroborates monophyly of Hormogastridae, but the genus Hormogaster is paraphyletic and Hormogaster pretiosa appears polyphyletic, stressing the need for taxonomic revisionary work in the family. The genus Vignysa could represent an early offshoot in the family, although the relationships with other genera are uncertain. The genus Hemigastrodrilus is related to the Hormogaster elise complex and both are found in the Atlantic drainage of the Iberian Peninsula and France. From a biogeographic perspective Corsica and Sardinia include members of two separate hormogastrid lineages. The species located in Corsica and Northern Sardinia are related to Vignysa, whereas Hormogaster pretiosa pretiosa, from Southern Sardinia, is closely related to the Hormogaster species from the NE Iberian Peninsula. A molecular dating of the tree using the separation of the Sardinian microplate as a calibration point (at 33 MY) and assuming a model of vicariance indicates that the diversification of Hormogastridae may be ancient, ranging from 97 to 67 Ma.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

The Mediterranean basin is a hotspot of outstanding biodiversity and should have priority for conservation policies, and therefore should benefit from biogeographic and evolutionary studies (Blondel and Aronson, 1999; Myers et al., 2000). Numerous studies have centered on the biogeography of circum-Mediterranean groups (e.g., Carranza and Arnold, 2003; Santos et al., 2007; Murienne and Giribet, 2009; Ribera et al., 2010), but only a few have focused on earthworms (Annelida, Clitellata, Oligochaeta) (Novo et al., 2009, 2010a), a group of terrestrial invertebrates with a great biogeographical potential due to narrow ecological requirements and their supposed old age. Bouček (1972) suggested that hormogastrid earthworms were already present in the Iberian plate in the Late Cretaceous, occupying areas that would coincide mainly with their present distribution. The edge of the Iberian Plate underwent drastic changes that affected the Betic–Rifean range, Balearic Islands, Kabylies, Corsica, Sardinia and Calabria (see Magri et al., 2007). It is in all these areas, once forming a primitive Iberian plate, where hormogastrids should be sought for in order to understand their evolutionary and biogeographic relationships. To date, hormogastrids have been discovered in all these regions, with the exception of the Betic and Rif ranges.

Hormogastridae currently includes 22 species in four genera (plus eight subspecies or varieties) of large to mid-sized (approximately 5–90 cm in length) earthworms distributed almost exclusively in the western Mediterranean (Díaz Cosín et al., 1989; Cobolli-Sbordoni et al., 1992) where they play an important ecological role aerating the soil and recycling nutrients (Bouček,
Traditionally, earthworm systematics has been based principally on morpho-anatomical features (Bouché, 1972; Sims and Gerard, 1999). Nevertheless, their taxonomy, and specifically that of hormogastrids has remained unstable and their phylogenetic relationships are uncertain because of the low number of characters available for study (Pop et al., 2003; Álvarez, 1977). The shape and number of spermathecae have been used as taxonomic characters, but they can vary with developmental stage. Other characters used in species diagnoses, such as the extension of the clitellum or the developmental degree of tubercula pubertatis, also show ontogenetic and reproductive differences. Although morphology still has an important role for species recognition in hormogastrids (Omodeo, 1956; Bouché, 1972; Qi and Bouché, 1998), the degree of homoplasy in many characters is high, probably reflecting high levels of phenotypic plasticity.

In two recent studies, Novo et al. (2009, 2010a) detected high genetic diversity among populations of the Hormogaster elisae species complex in the central Iberian Peninsula, as well as between this species complex and other hormogastrid species. These studies questioned the status of H. elisae as a valid species and discussed its affinities within the family. In order to re-evaluate the status of H. elisae, and to test species and genus delimitations in Hormogastridae, we expanded both taxon sampling and the number of molecular markers and inferred the phylogenetic history of the family by using data from nine molecular markers. We also used geological events occurring during the evolution of the Mediterranean basin as a calibration point to provide an estimate of the age of diversification of the family.

2. Materials and methods

2.1. Sampling and morphological study

 Mature individuals, representing 20 of the 22 described species of Hormogastridae, were collected from 30 localities in the Iberian Peninsula, France (including Corsica) and Sardinia, including the type locality of each known species. Additional data from the study of Novo et al. (2010a) in Central Spain were included, making for a total of 46 localities (Tables S1 and S2, Fig. 1). We included an immature specimen collected near the shore of the Ter river (approximately 9 km from Girona, Spain) that was analyzed for allozymes in Cobolli-Bordoni et al. (1992), and identified as Hormogaster pretiosa cf. hispanica (HPA). Amplification of specimens from Elba, Sardinia, and France, also provided by P. Omodeo, was unsuccessful.

The type locality of Hormogaster reddii reddii (Chilarza [GH], Sardinia, Italy) (Rosa, 1887), the type species of the family, was sampled for this study. In addition an individual of this species from Alghero (AL, Sardinia, Italy) was kindly provided by Maguolidà Ricoy. A new species, whose diagnosis will be presented (M. Novo, work in progress), was collected at Sant Joan de les Abadesses (SAN, Girona, Spain). The type locality of Hormogaster multilamella (Etxauri, Navarra, Spain) was visited several times but no specimens were obtained. Amplification of the target regions from a formalin-fixed specimen kindly provided by El Museo de Zoología de la Universidad de Navarra was also unsuccessful. Specimens collected in Talarn (TAL, Lleida, Spain) do not match the morphological description of Hormogaster leidiana, and we refer to them as Hormogaster sp. These two species, H. multilamella and H. leidiana, are thus the only described hormogastrids missing from this study. However, the identity of these two species is questionable, as Qi and Bouché (1998) based their morphological descriptions on a single adult specimen for H. multilamella and one adult and one juvenile for H. leidiana, which could explain the difficulties in collecting the specimens in the first case and the mismatch morphology in the second.

All individuals were washed in distilled water and preserved in 96% EtOH at −20 °C for subsequent molecular work. A portion of the integument (ca. 25 mg) was carefully cleaned under a stereomicroscope to remove soil particles and associated fauna. Subsequently, integument samples were hydrated and preserved at −80 °C until DNA extraction.

2.2. DNA extraction, gene amplification and sequencing

Total genomic DNA was extracted from a tissue sample from the integument using the DNeasy Tissue Kit (Qiagen) eluting twice with 70 μl of buffer. Molecular markers included mitochondrial regions of the cytochrome c oxidase subunit I gene (COI), 16S rRNA gene and tRNA Leu, Ala, and Ser, two nuclear ribosomal genes (complete 18S rRNA and a fragment of 28S rRNA) and two nuclear protein-encoding genes (histone H3 and histone H4) (see Table 1 for primer information). We also tried amplifying the Internal Transcribed Spacer gene (ITS) in two fragments with the primers: 5′-CAG GAT TCC CTA AAC GAC TCG ATG ATT C-3′ for ITS1 and 5′-TGA TCG CCT AGT TTA TCC GCC GAA G-3′ (Carraza, 1997) and 5.8Srev: 5′-GGC TTA GAC GAT CAG TAT GAA TCC CA-3′ for ITS2. However, no amplification was obtained, probably reflecting high levels of phenotypic plasticity. Successful amplification was obtained for ITS1 and 5.8Sfor: 5′-GAA TCG CTA TCG AGT TTA TCA GAC C-3′ and 28Srev: 5′-GTT AGT TTC TTT TCC TCC GCC GGT T-3′ for ITS2, but only ITS1 was consistently amplified, and only in two of the populations studied (CER: Hormogaster sp. and GRA: Hormogaster eserana). Only two nucleotide changes were detected among them.

Polymerase chain reactions (PCR) (50 μl) included 0.6 μl of template DNA, 1 μl of each primer, 200 μM of dNTPs (Invitrogen), 1X PCR buffer containing 1.5 mM MgCl₂ (Applied Biosystems,) and 1.25 units of AmpliTaq DNA polymerase (Applied Biosystems). DMSO (1 μl) was included for the 28S rRNA reaction. PCR reactions were carried out using a GeneAmp Multicycler Ep gradient (Eppendorf), and included an initial denaturation step (5 min at 95 °C) followed by 35 or 40 cycles including denaturation at 95 °C for 30 s, annealing (ranging from 49 °C to 60 °C) for 30 s or 1 min, and extension at 72 °C for 1 min, with a final extension step at 72 °C for 7 min. PCR products were resolved by 1% agarose gel electrophoresis, visualized by ethidium bromide fluorescence and purified using MultiScreen PCR Plates and a Vacuum Manifold (Millipore). The purified products were sequenced directly with the same primers pairs used for amplification. Each sequence reaction contained a total volume of 10 μl including 3 μl of the PCR product, 0.32 μM of one of the PCR primer pairs, 0.25 μl of ABI BigDye 5X sequencing buffer and 0.5 μl of ABI BigDye Terminator (Applied Biosystems). The sequencing reactions involved an initial denaturation step (5 min at 95 °C) and 30 cycles (95 °C for 10 s, 50 °C for 5 s, and 60 °C for 4 min). The BigDye-labeled PCR products were cleaned using Performa DTR Plates (Edge Biosystems), the sequence reaction products were then analyzed using an ABI Prism 3730xl Genetic Analyzer (Applied Biosystems).

Chromatograms were visualized in Sequencer v.4.7 (Gene Codes Corporation) to obtain the assembled sequences. These were compared against the GenBank database with the BLAST algorithm (Altschul et al., 1997). All new sequences have been deposited in GenBank (Tables S1 and S2).

2.3. Data analysis

2.3.1. Phylogenetic relationships

Sequences of the polychaete Arenicola marina and other clitellates were included as outgroups. These sequences were kindly provided by the Protostome AToL project (PI G. Giribet), downloaded from GenBank (see Table S3), or generated de novo (histone H4 for Aporrectodea trapezoides).

Data analyses were based on two approaches, a traditional two-step approach (multiple sequence alignment + phylogenetic analysis) and a single-step approach with direct optimization (Wheeler,
For direct optimization we used parsimony as the optimality criterion, as implemented in the program POY v.4.1.2 (Varón et al., 2010). Tree searches were conducted by a combination of random addition sequences with subtree pruning and regrafting (SPR) and tree bisection and reconnection (TBR) branch swapping followed by multiple rounds of tree fusing (Goloboff, 1999, Fig. 1. Sampling localities. The abbreviation of the species found in each locality is indicated (see in Table S1 the complete name of species and localities). Colors match those used in the phylogenetic trees. ORD, HPA y VIL are shown with an asterisk since they form the clade used for tree calibration (see phylogenetic trees). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Sequences of primers used to amplify and sequence gene fragments in this study.

<table>
<thead>
<tr>
<th>Amplified region (size in bp.)</th>
<th>Primer name</th>
<th>Primer sequence (5’–3’)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>COI (648)</td>
<td>LCO1490</td>
<td>GGTCAACAAATCATGACATATTGG</td>
<td>Folmer et al. (1994)</td>
</tr>
<tr>
<td></td>
<td>HCO2198</td>
<td>TAAAACATTCACTAGTTGCCG</td>
<td></td>
</tr>
<tr>
<td>16S-tRNAs (759–793)</td>
<td>16S-tRNA-Leu-Ala-Ser-Leu-LumbF2</td>
<td>GCACTGTTTTAACAAAAACATTGC</td>
<td>Pérez-Losada et al. (2009)</td>
</tr>
<tr>
<td></td>
<td>Ho_16sRa:</td>
<td>GCACATCTCGCACTGCTTG</td>
<td>Novo et al. (2010a)</td>
</tr>
<tr>
<td>18S (1701)</td>
<td>1F</td>
<td>TACCTGTCACCTGCCAGTAG</td>
<td>Giribet et al. (1996), Whiting et al. (1997)</td>
</tr>
<tr>
<td></td>
<td>3F</td>
<td>GAATACCCCGCGCTGCTGG</td>
<td></td>
</tr>
<tr>
<td></td>
<td>18Sa:2.0</td>
<td>GAGTTGCAAAGCTGAAAC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9R</td>
<td>GATCATCAGGCTCACTAC</td>
<td></td>
</tr>
<tr>
<td>28S (738–739)</td>
<td>28S-F1</td>
<td>GAGTACCCTGAAACCCTCTAG</td>
<td>Pérez-Losada et al. (2009)</td>
</tr>
<tr>
<td></td>
<td>28S-R1</td>
<td>GTCATCCGCAAAGGCTC</td>
<td></td>
</tr>
<tr>
<td>H3 (328)</td>
<td>H3a F</td>
<td>ATGCCCTGTACCCAGCACACVG</td>
<td>Colgan et al. (1998)</td>
</tr>
<tr>
<td></td>
<td>H3a R</td>
<td>ATAATCTGCGATATGGTCGC</td>
<td></td>
</tr>
<tr>
<td>H4 (183)</td>
<td>H4F-2S</td>
<td>TSCGGGAAATCAGGGGATCATAC</td>
<td>Pineau et al. (2005)</td>
</tr>
<tr>
<td></td>
<td>H4F-2er</td>
<td>CKYTTACGGTCAAGCCAGTTCTC</td>
<td></td>
</tr>
</tbody>
</table>

1996). For direct optimization we used parsimony as the optimality criterion, as implemented in the program POY v.4.1.2 (Varón et al., 2010). Tree searches were conducted by a combination of random addition sequences with subtree pruning and regrafting (SPR) and tree bisection and reconnection (TBR) branch swapping followed by multiple rounds of tree fusing (Goloboff, 1999,
The protein-encoding genes were “pre-aligned” as their alignments were trivial due to lack of length variation. We undertook a sensitivity analysis of 10 parameter sets varying the contributions of indels and base transformations (Wheeler, 1995) and used an index of character congruence for selecting the optimal parameter set. For the sensitivity analyses data were analyzed in combination and for each independent partition. Nodal support was estimated by bootstrap resampling (Felsenstein, 1985) with 100 pseudoreplicates.

For the two-step approach sequences of each individual gene were aligned in CLUSTALX v. 2.0.12 (Thompson et al., 1997) with default settings and concatenated with PHYUTILITY (Smith and Dunn, 2008). JModelTest v. 0.1.1 (Posada, 2008) was used to select the best-fit evolutionary model using the Akaïke information criterion (AIC; Akaïke, 1973). The phylogenetic hypothesis was then estimated with MRBAYES v. 3.1.2 (Ronquist and Huelsenbeck, 2003). In the concatenated matrix unlinked nucleotide substitution models selected by JModelTest were specified for each gene fragment and the nucleotide substitution estimates were allowed to vary independently between each partition. Parameters in MRBAYES were set to two million generations and 4000 trees were sampled for every 500th generation, using the default random tree option to initiate the analysis. The analysis was performed twice and all sample points prior to the plateau phase were discarded as burn-in. The remaining trees were combined to find the maximum a posteriori probability estimate of phylogeny. Because we detected a saturation pattern in third codon positions of COI and H3 using DAMBE (Xia and Xie, 2001), a second phylogenetic tree was also constructed without these positions. We also explored the results after an alignment with MAFFT v6 (Katoh et al., 2005; online server: http://mafft.cbrc.jp/alignment/server/). Finally, trees were inferred with mitochondrial and nuclear fragments separate using the same settings as indicated for the general analysis, to test for possible conflicting signal among these two sets of genes.

A maximum likelihood (ML) analysis was performed with PhyML v. 2.4.4 (Guindon and Gascuel, 2003) using simultaneous NNI to estimate tree topology. The best-fit evolutionary model selected by jModelTest for the concatenated sequences was specified and the program optimized α and I values. Clade support was evaluated by 1000 replicates of non-parametric bootstrapping (Felsenstein, 1985).

2.3.2. Estimating divergence times

All the phylogenetic trees were calibrated by using the separation between H. pretiosa pretiosa from Sardinia (VIL) and the continental species clustering in the same clade, H. pretiosa cf. hispanica collected in the shore of the river Ter (HPA) and Hormogaster najiformis (ORD), assuming a divergence time for these lineages going back to the separation of the Corso-Sardinian microplate from continental Europe. This is justified due to the low dispersal ability of earthworms in general (Lighthart and Peek, 1997; Hale et al., 2005) and of hormogasters in particular (Novo et al., 2010b; Omodeo and Rota, 2008). This low vagility is best explained by the endogeic nature of hormogasters. As a calibration point we use the most current estimates for the separation of the occidental Mediterranean microplates (33 Ma; Schettino and Turco, 2006). It is of course well understood that this minimum age only applies if there has been no transoceanic dispersal of hormogaster species, an assumption that we accept given the biology of the animals of choice.

To our knowledge, there are no similar substitution rate estimates in earthworms that could be used for comparison to our data. Buckley et al. (2011) constructed an ultrametric tree with BEAST for New Zealand megascolecid, including 16S rRNA and 28S rRNA but their age estimation was relative due to the lack of a calibration point. Also, Chang et al. (2008) offered a rough estimate for the COI rate of a megascolecid species complex, but this is not comparable to our estimate of the combined mitochondrial genes.

Clockness of the data (constant substitution rate) was tested by means of a likelihood-ratio test in jModelTest. The Bayesian tree and the evolutionary model parameters employed were introduced in PaupUP v. 4.0b10 (Calendini and Martin, 2005) and its likelihood values were estimated with and without the assumption of a molecular clock. The comparison of these values in jModelTest was significant, indicating the lack of a molecular clock (the substitution rates vary along the branches of the tree). Therefore r8s v. 1.71 (Sanderson, 1997, 2006) was used for estimating divergence times. Initially the truncated Newton (TN) algorithm was used to do a cross-validation with eight smoothing factors. Divergence times were then estimated using the penalized likelihood (PL) method and the optimal smoothing value and fixing the age of the separation between continental “H. pretiosa” (see below) and the Sardinian population (where this species was originally described by Michaelis, 1899) at 33 Ma. The analyses were then repeated with both the Bayesian and maximum likelihood trees, including and excluding outgroups. Another analysis was conducted with the implied alignment (Wheeler, 2003; Gribet, 2005) from POY under the optimal parameter set (221). The implied alignment was then introduced in PAUP*v. 4.0b3 (Swofford, 2002) to estimate branch lengths and finally the divergence times were calculated with r8s.

We finally obtained an ultrametric tree by means of Bayesian estimations as implemented in BEAST v.1.6.1. (Drummond and Rambaut, 2007), including the same evolutionary models as in the phylogenetic analyses for each gene partition. The well-supported node of all the earthworms was constrained to be monophyletic and the analysis was performed under an uncorrelated lognormal relaxed clock and a Yule diversification model. The age of the calibration clade was set to a normal distribution with mean of 33 MY and a 95% confidence interval. BEAST was run for 50 millions of generations, saving trees at every 5000th generation. The “burnin” was set to 2000 in TreAnnotator v.1.6.1. after visualizing the results with Tracer v.1.5. (Drummond and Rambaut, 2007).

3. Results

3.1. Morphological features

The morphological characters from the individuals collected in Cervera del Maestre (CER, Castellón, Spain), match those described by Zici (1970) for H. pretiosa from the same locality. These characters, however, do not match the original description by Michaelis (1899) and emended by Omodeo (1956). Therefore we use the name Hormogaster sp. for these specimens for the reasons outlined below. Figs. 2 and 3 illustrate important morphological features of the species included in this study.

3.2. Gene diversity and phylogenetic relationships

Values of genetic variability for the Hormogasteridae data set are shown in Table 2. The mitochondrial COI and 16S-trNA were the most variable regions, showing a mean percentage of uncorrected pairwise divergence between 14.5% and 17.4%. The nuclear genes were less variable, showing a mean percentage of uncorrected pairwise divergence of around 3% for histones (being H3 more variable than H4) and around 2% for 28S rRNA. The least variable gene was 18S rRNA, showing a mean percentage of uncorrected pairwise divergence of only 0.23%. Within the
Hormogastridae data set indels were only found in the 16S-tRNA and 28S rRNA genes.

In the parsimony analysis with direct optimization the combination of parameter sets that minimized the incongruence among partitions was “221” (Table 3). By applying this optimal combination only one tree of 10,933 weighted steps was found (Fig. 4).

The multiple sequence alignment of the concatenated sequence data used for the two-step phylogenetic analyses was 4547 bp in length. Likelihood ratio tests performed with jModelTest indicated that the best-fit models of sequence evolution are GTR + I + G for the concatenated data set, as well as for all individual partitions, with the exception of histone H4, which adjusts to a HKY + I model.

The 50% majority-rule consensus tree generated by MRBAYES for this alignment is shown in Fig. 5 (BI). The topology was identical whether the third codon positions for COI and histone H3 were excluded or not, but posterior probability values varied slightly (not shown). The maximum likelihood tree for all data for the aligned data set (lnL = −37,467.335496) is also shown in Fig. 5 (ML). The analyses after alignment of sequences with MAFFT showed similar results, with the previously well-supported clades being the same.

All trees share common features such as the monophyletic nature of Hormogastridae (support is low in the direct optimization analysis), the paraphyly of Hormogaster, which includes the genera Xana and Hemigastrodrilus in all trees and Vignysa in some of the analyses, or the polyphyly of H. pretiosa. Many of the identified clades are also shared among analyses. The H. elisae species complex constitutes a lineage generally related to Hemigastrodrilus (MP and ML), but no other Hormogaster species appear in this clade. The Hormogaster species from the NE Iberian Peninsula and S France form another stable clade, with two exceptions, H. najaformis (ORD) and the H. pretiosa cf. hispanica specimen from the Ter river, both forming a clade with the population of H. pretiosa pretiosa from Sardinia (VIL) and with Xana omodeoi. A clade formed by H. redii and Hormogaster samnitica lirapora from Corsica and Sardinia sometimes appears related to Vignysa (BI and ML analyses). However, this latter genus appears to have diverged earlier from the rest of the family in the direct optimization analysis (Fig. 4).

The trees built with mitochondrial and nuclear genes separately (Fig. S1) are generally congruent with the trees obtained from the combined analysis—our preferred hypothesis—but there are differences worth highlighting. When mitochondrial genes were analyzed alone, lumbricids nest within homogastrids, forming a clade with Vignysa. In the tree based on nuclear sequence data alone, Xana and Vignysa form a clade with high support values, a result not obtained in other analyses. These genes also support a subdivision of the clade that contains most of the species from the NE Iberian Peninsula and S France.

Fig. 2. Morphological features of the different genera of hormogastrids. Xana omodeoi (A–C), specimens (A) spermathecae from one side (B) and typhosole (C). Vignysa (D–G), D: Spermathecae from one side of V. popi; E: Anterior part of a V. popi specimen collected in Saint Gely du Fesc (France); F: Blood vesicles of V. popi; G: Pair of spermathecae of V. vedovinii. Hormogaster (H and I), Anterior part of a Hormogaster redi insularis specimen (H: Volpajola, Corsica, France) and a Hormogaster samnitica lirapora specimen (I: Col de la Testa, Corsica, France). Hemigastrodrilus monicae (J–M), J: Fememine pores (pointed by the arrows), located in the ventral area; K: Specimens of He. monicae collected in Cardaye (France); L: Morren gland; M: Spermathecae from one side.
3.3. Divergence time estimates

When the separation of the Western Mediterranean microplates was used as a calibration point, both the maximum likelihood and the Bayesian analyses (Fig. 5) show concordant dates and the inclusion or exclusion of the outgroups had no effect on the dates. The calibration of the tree from direct optimization (Fig. 4) has some differences due to the topological disagreement mostly of Vignysa, but the main dates coincide. Fig. 6 illustrates the results of the tree generated with BEAST, that showed a later diversification date than alternative methods. Hormogastrids seem to have started diversifying at 97–67 Ma (this range shows the values of the different methods), and the clade formed by the Hormogaster species from the NE Iberian Peninsula and S France is dated at 56–46 Ma. Perhaps most interesting is the old age of the H. elisae species complex, dated at 42–37 Ma. The recovered substitution rates were similar among the analyses. As an example, the values from the analyses with r8s and the Bayesian tree were 0.28% (sd. 0.16%) substitutions per site per million years for the concatenated data set, 0.52% (sd. 0.20%) for the mitochondrial genes and 0.017% (sd. 1.822 e-10) for the nuclear genes.

4. Discussion

All phylogenetic analyses of the combined data show the monophyly of the family Hormogastridae, which is well supported in the probabilistic two-step analyses, as well as in the direct optimization approach, albeit with low support values in the latter.

Table 2

<table>
<thead>
<tr>
<th></th>
<th>COI</th>
<th>16S-1RNA</th>
<th>H3</th>
<th>H4</th>
<th>28S</th>
<th>18S</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>45</td>
<td>46</td>
<td>46 (92)</td>
<td>44 (88)</td>
<td>45</td>
<td>46</td>
</tr>
<tr>
<td>NHAP</td>
<td>45</td>
<td>46</td>
<td>44</td>
<td>24</td>
<td>29</td>
<td>19</td>
</tr>
<tr>
<td>H</td>
<td>1 (0.0047)</td>
<td>1 (0.00145)</td>
<td>0.98 (0.0032)</td>
<td>0.91 (0.0188)</td>
<td>0.93 (0.03)</td>
<td>0.90 (0.03)</td>
</tr>
<tr>
<td>π</td>
<td>0.174 (0.085)</td>
<td>0.144 (0.070)</td>
<td>0.036 (0.018)</td>
<td>0.027 (0.014)</td>
<td>0.021 (0.011)</td>
<td>0.002 (0.001)</td>
</tr>
<tr>
<td>Sites</td>
<td>648</td>
<td>806</td>
<td>328</td>
<td>183</td>
<td>739</td>
<td>1701</td>
</tr>
<tr>
<td>Gaps</td>
<td>0</td>
<td>77</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Mutations</td>
<td>498</td>
<td>532</td>
<td>86</td>
<td>34</td>
<td>89</td>
<td>22</td>
</tr>
<tr>
<td>S</td>
<td>262</td>
<td>366</td>
<td>64</td>
<td>26</td>
<td>73</td>
<td>21</td>
</tr>
<tr>
<td>Parsimony IS</td>
<td>250</td>
<td>225</td>
<td>63</td>
<td>26</td>
<td>49</td>
<td>21</td>
</tr>
<tr>
<td>Ts</td>
<td>350</td>
<td>334</td>
<td>53</td>
<td>21</td>
<td>50</td>
<td>19</td>
</tr>
<tr>
<td>Tv</td>
<td>148</td>
<td>198</td>
<td>33</td>
<td>13</td>
<td>39</td>
<td>3</td>
</tr>
<tr>
<td>Dp (N)</td>
<td>113.04 (49.44)</td>
<td>116.45 (50.89)</td>
<td>11.87 (5.42)</td>
<td>4.88 (2.40)</td>
<td>15.58 (7.09)</td>
<td>3.87 (1.98)</td>
</tr>
<tr>
<td>Dp (%)</td>
<td>17.44 (7.63)</td>
<td>14.45 (6.31)</td>
<td>3.62 (1.65)</td>
<td>2.67 (1.31)</td>
<td>2.11 (0.96)</td>
<td>0.23 (0.12)</td>
</tr>
</tbody>
</table>

Fig. 3. Spermathecae of different described species of Hormogaster. A: H. pretiosa nigra, collected in Quillan (France); B: H. redii insularis, collected in Volpajola (Corsica, France); C: H. riojana, collected in Alesanco (Spain); D: H. elisae, collected in Siguero (Spain); E: H. samnitica lirapora, collected in Col de la Testa (Corsica, France).

130 M. Novo et al. / Molecular Phylogenetics and Evolution 61 (2011) 125–135
The only significant discrepancy among analyses seems to be due to possible conflict between the nuclear and mitochondrial genes, as the latter group lumbricids with Vignysa. In a similar case by Briones et al. (2009) it was concluded that 16S rRNA and COI have limited discriminatory power above the genus level. Nevertheless, and in contrast to our combined approach, Jamieson (1988) suggested non-monophyly of Hormogastridae based on morphological characters, suggesting that Vignysa was related to lumbricids whereas Hemigastrodrilus and Hormogaster were more closely related to microchaetids and glossoscolecids. Future analyses including representatives of these additional earthworm families should be able to fully test Jamieson’s phylogenetic scheme.

The geological scenario found at the moment when Hormogastridae may have started its diversification (ca. 97–67 Ma) suffices for explaining their current distribution without the need to invoke oceanic dispersal. This date range agrees with the late Cretaceous to the beginning of the Tertiary postulated by Bouché (1972) for the origin of the Lumbricoida families, and suggests that the current distribution of the different genera was probably achieved during the early Cenozoic. Omodeo and Rota (2008) however, use 180 million years for the split of two megadrile genera (see Omodeo, 2000). This was based on estimates of the divergence time between congeneric species of both sides of the Atlantic (Omodeo, 2000), but it is not applicable for the present study because substitution rates in earthworms are highly variable even among members of the same family (Novo, 2010). Of course, our

Table 3
Tree lengths for the different partitions analyzed and congruence values (ILD) for the combined analysis of the six loci with different parameter set combinations (left column). The first numeral used in the parameter set corresponds to the ratio between indel/transversion and the following two numbers correspond to the ratio between transversion/transition; for example, 111 corresponds to equal weights and 121 corresponds to a indel/transversion ratio of 1 and to a transversion/transition ratio of 2:1, thus, the indels and transversions have a cost of 2, and transitions have a cost of 1. (For a list of the specific step matrices that this involves see Giribet et al., 2002).

<table>
<thead>
<tr>
<th>COI</th>
<th>16S-rRNA</th>
<th>H3</th>
<th>H4</th>
<th>28S</th>
<th>18S</th>
<th>All</th>
<th>ILD</th>
</tr>
</thead>
<tbody>
<tr>
<td>111</td>
<td>2927</td>
<td>2210</td>
<td>324</td>
<td>57</td>
<td>651</td>
<td>637</td>
<td>6980</td>
</tr>
<tr>
<td>121</td>
<td>4059</td>
<td>3332</td>
<td>479</td>
<td>81</td>
<td>1000</td>
<td>964</td>
<td>10150</td>
</tr>
<tr>
<td>211</td>
<td>2927</td>
<td>2457</td>
<td>324</td>
<td>57</td>
<td>737</td>
<td>732</td>
<td>7415</td>
</tr>
<tr>
<td>221</td>
<td>4059</td>
<td>3771</td>
<td>478</td>
<td>81</td>
<td>1154</td>
<td>1152</td>
<td>10933</td>
</tr>
<tr>
<td>241</td>
<td>6273</td>
<td>6353</td>
<td>779</td>
<td>127</td>
<td>1970</td>
<td>1970</td>
<td>17872</td>
</tr>
<tr>
<td>411</td>
<td>2927</td>
<td>2811</td>
<td>324</td>
<td>57</td>
<td>839</td>
<td>905</td>
<td>9606</td>
</tr>
<tr>
<td>421</td>
<td>4059</td>
<td>4457</td>
<td>479</td>
<td>81</td>
<td>1358</td>
<td>1495</td>
<td>12197</td>
</tr>
<tr>
<td>441</td>
<td>6273</td>
<td>7700</td>
<td>779</td>
<td>127</td>
<td>2383</td>
<td>2624</td>
<td>20396</td>
</tr>
<tr>
<td>3221</td>
<td>2927</td>
<td>2761</td>
<td>324</td>
<td>57</td>
<td>810</td>
<td>797</td>
<td>7798</td>
</tr>
</tbody>
</table>

The only significant discrepancy among analyses seems to be due to possible conflict between the nuclear and mitochondrial genes, as the latter group lumbricids with Vignysa. In a similar case by Briones et al. (2009) it was concluded that 16S rRNA and COI have limited discriminatory power above the genus level. Nevertheless, and in contrast to our combined approach, Jamieson (1988) suggested non-monophyly of Hormogastridae based on morphological characters, suggesting that Vignysa was related to lumbricids whereas Hemigastrodrilus and Hormogaster were more closely related to microchaetids and glossoscolecids. Future analyses including representatives of these additional earthworm families should be able to fully test Jamieson’s phylogenetic scheme.

The geological scenario found at the moment when Hormogastridae may have started its diversification (ca. 97–67 Ma) suffices for explaining their current distribution without the need to invoke oceanic dispersal. This date range agrees with the late Cretaceous to the beginning of the Tertiary postulated by Bouché (1972) for the origin of the Lumbricoida families, and suggests that the current distribution of the different genera was probably achieved during the early Cenozoic. Omodeo and Rota (2008) however, use 180 million years for the split of two megadrile genera (see Omodeo, 2000). This was based on estimates of the divergence time between congeneric species of both sides of the Atlantic (Omodeo, 2000), but it is not applicable for the present study because substitution rates in earthworms are highly variable even among members of the same family (Novo, 2010). Of course, our
estimates are contingent to our assumption that the cladogenetic event leading to \( H. \) pretiosa pretiosa from Sardinia (VIL) and the continental specimens of \( H. \) was due to a vicariant event during the separation of the Corso-Sardinian microplate from continental Europe. If this were the case, the cladogenetic event could still be older than the geologic event used for the calibration, and therefore our estimate for the divergence of the family would appear younger than it was.

During the late Cretaceous the Iberian Peninsula was divided into an Eastern and a Western area separated by water (Andeweg, 2002), perhaps explaining the current absence of hormogastrids in the westernmost area of the Iberian Peninsula. The Betic range was focused on the type localities of all the known species but there are several undescribed hormogastrid species in regions of suitable habitat that remain unexplored. In the present work the sampling focused on the type localities of all the known species but there are ample areas in the Iberian Peninsula where earthworms have not been studied. In addition, even though we expected to find \( H. \) gallica in Sant Joan de les Abadesses (Girona), a new species was found instead (M. Novo work in progress). Likewise, in Talarn (Lleida) the specimens collected do not match the description of \( H. \) lleidana, reported from this locality by Qiu and Bouché (1998). These findings indicate that a greater sampling effort may be required to have a better representation of the diversity of this earthworm family.

The central area of the Iberian Peninsula has already been surveyed extensively (Novo et al., 2010a) and a speciation event is responsible for explaining the diversity of \( H. \) that has persisted in this region. \( H. \) is the only morphospecies and \( Vignysa \) (Murienne and Giribet, 2009).

According to the present analyses, the genus \( H. \) is paraphyletic, as it includes the members of several other hormogastrid genera. A denser sampling of \( Vignysa \) and \( H. \) (there are no additional known populations of \( Xana \) is necessary to further test these relationships. Moreover there are probably several undescribed hormogastrid species in regions of suitable habitat that remain unexplored. In the present work the sampling focused on the type localities of all the known species but there are ample areas in the Iberian Peninsula where earthworms have not been studied. In addition, even though we expected to find \( H. \) in Sant Joan de les Abadesses (Girona), a new species was found instead (M. Novo work in progress). Likewise, in Talarn (Lleida) the specimens collected do not match the description of \( H. \) lleidana, reported from this locality by Qiu and Bouché (1998). These findings indicate that a greater sampling effort may be required to have a better representation of the diversity of this earthworm family.

The central area of the Iberian Peninsula has already been surveyed extensively (Novo et al., 2010a) and a speciation event is responsible for explaining the diversity of \( H. \) that has persisted in this region. \( H. \) is the only morphospecies and \( Vignysa \) (Murienne and Giribet, 2009).

According to the present analyses, the genus \( H. \) is paraphyletic, as it includes the members of several other hormogastrid genera. A denser sampling of \( Vignysa \) and \( H. \) (there are no additional known populations of \( Xana \) is necessary to further test these relationships. Moreover there are probably several undescribed hormogastrid species in regions of suitable habitat that remain unexplored. In the present work the sampling focused on the type localities of all the known species but there are ample areas in the Iberian Peninsula where earthworms have not been studied. In addition, even though we expected to find \( H. \) in Sant Joan de les Abadesses (Girona), a new species was found instead (M. Novo work in progress). Likewise, in Talarn (Lleida) the specimens collected do not match the description of \( H. \) lleidana, reported from this locality by Qiu and Bouché (1998). These findings indicate that a greater sampling effort may be required to have a better representation of the diversity of this earthworm family.

The central area of the Iberian Peninsula has already been surveyed extensively (Novo et al., 2010a) and a speciation event is responsible for explaining the diversity of \( H. \) that has persisted in this region. \( H. \) is the only morphospecies and \( Vignysa \) (Murienne and Giribet, 2009).

According to the present analyses, the genus \( H. \) is paraphyletic, as it includes the members of several other hormogastrid genera. A denser sampling of \( Vignysa \) and \( H. \) (there are no additional known populations of \( Xana \) is necessary to further test these relationships. Moreover there are probably several undescribed hormogastrid species in regions of suitable habitat that remain unexplored. In the present work the sampling focused on the type localities of all the known species but there are ample areas in the Iberian Peninsula where earthworms have not been studied. In addition, even though we expected to find \( H. \) in Sant Joan de les Abadesses (Girona), a new species was found instead (M. Novo work in progress). Likewise, in Talarn (Lleida) the specimens collected do not match the description of \( H. \) lleidana, reported from this locality by Qiu and Bouché (1998). These findings indicate that a greater sampling effort may be required to have a better representation of the diversity of this earthworm family.

The central area of the Iberian Peninsula has already been surveyed extensively (Novo et al., 2010a) and a speciation event is responsible for explaining the diversity of \( H. \) that has persisted in this region. \( H. \) is the only morphospecies and \( Vignysa \) (Murienne and Giribet, 2009).

According to the present analyses, the genus \( H. \) is paraphyletic, as it includes the members of several other hormogastrid genera. A denser sampling of \( Vignysa \) and \( H. \) (there are no additional known populations of \( Xana \) is necessary to further test these relationships. Moreover there are probably several undescribed hormogastrid species in regions of suitable habitat that remain unexplored. In the present work the sampling focused on the type localities of all the known species but there are ample areas in the Iberian Peninsula where earthworms have not been studied. In addition, even though we expected to find \( H. \) in Sant Joan de les Abadesses (Girona), a new species was found instead (M. Novo work in progress). Likewise, in Talarn (Lleida) the specimens collected do not match the description of \( H. \) lleidana, reported from this locality by Qiu and Bouché (1998). These findings indicate that a greater sampling effort may be required to have a better representation of the diversity of this earthworm family.
Fig. 6. Ultrametric tree obtained with BEAST using the combined data. The estimated dates (MY, below the bars) and their 95% confidence bars are shown for the main clades in the same colors as in previous figures. Posterior probability values for these clades are shown above bars. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

It can also be concluded that the hormogastrid species present in Corsica and Sardinia have received two different genetic influences. One includes H. redi and H. sanmittica liparosa, related to Vignysa in some analyses, but of unresolved position (appears basal in the direct optimization analysis, unresolved in the Bayesian analysis and derived in the maximum likelihood analysis). It is plausible that in this first lineage cryptic species exist, as the genetic divergences found between populations they assigned to different species status to H. pretiosa hispanica cf. hispanica, as understood nowadays, is polyphyletic, suggesting the need of revisionary taxonomy in the group (authors’ work in progress). Cobolli-Sbordoni et al. (1992) pointed out that H. pretiosa hispanica may belong to a different species, as they show deep genetic divergences as well as important differences in the number and position of spermathecae. The type specimen of H. pretiosa pretiosa is from Sardinia and many of the NE Iberian specimens formerly identified as H. pretiosa may belong to a different species, as they show deep genetic divergences as well as important differences in the number and position of spermathecae. Omodeo and Rota (2008) indicated that Cobolli-Sbordoni et al. (1992) had already proposed the name Hormogaster hispanica for the populations from Catalonia, but this is inaccurate, as these authors simply stated that the high divergence values found between populations they assigned to H. pretiosa suggested the subdivision of this morphospecies into three sibling species, but they did not assign species status to these three lineages—which include H. pretiosa cf. hispanica. Later, Rota (1993) described a new species for the third lineage, H. gallica, but never assigned species status to H. pretiosa hispanica.

The phylogenetic hypothesis generated under direct optimization is perhaps the most credible from a biogeographical point of view, in spite of the low bootstrap support values. This hypothesis shows Vignysa as the sister clade to the remaining species of the family. Following this tree, two clades differentiate afterwards, one present in Corsica and the Northern area of Sardinia and another including the Atlantic lineage inclusive of Hemigastrodrilus and H. elisae and the lineage present in the NE Iberian Peninsula and S France. The latter clade includes two groups, one with H. sanmittica liparosa and Vignysa vedovinii, multiple in Sardinia (VIL) and Xana, and another with the remainder of the Hormogaster species.

Xana is again challenging morphological reasoning, as its position in the trees implies multiple origins of the third gizzard or a regression to the two-gizzard state in one occasion. Xana inhabits soils that are very different from those of other hormogastrids (Novo, 2010). The higher content of clay in these soils could explain the regression of this character to a two-gizzard state, as Xana has no need for such an exhaustive grinding of the soil as required in a sandier soil. Therefore it is plausible that selective pressure has contributed to eliminating the third gizzard in Xana. Nonetheless, as discussed above, Xana and Vignysa do form a clade in the analyses using only conserved nuclear genes, and this correlates with the apomorphic presence of two gizzards in both genera. In addition to the number of gizzards, other morphological characters also show homoplasy; the shape of the spermathecae, tubular in H. elisae and in Vignysa popi, multiple in H. najaformis and H. sanmittica liparosa, or multilocular in some Hormogaster species of the NE Iberian clade and Vignysa vedovinii. The results of this work have important implications for the taxonomy of the family, evidencing that it should be revised and that generic designations must be tested phylogenetically. In addi-
tion, species definitions are also deficient in some cases, as in H. elisa and H. pretiosa, calling perhaps for combined diagnosis of species using molecules in addition to morphology (see for example Edgecombe and Giribet, 2008; Cook et al., 2010). There are still lots of mysteries around hormogastrid systematics and the information at hand is only sufficient to delimit the backbone of this phylogenetic and biogeographic puzzle. Additional sampling effort and exploration of new geographic regions in the Western Mediterranean basin are necessary in order to fill in the missing pieces.

Acknowledgments

We are indebted to Carlos Bayón, Pablo Cabo, Inés García, Noa Nuevo and Juan José Molina, for assisting with fieldwork, Maigualida Ricoy and Pietro Omodeo for sharing additional samples and Daniel Ayllón and Sheila Leal for technical support. The members from Giribet lab offered help for lab work and data analyses, especially Ron Clouse, Sebastián Vélez and Prashant Sharma. Two anonymous reviewers provided comments that helped to improve this manuscript. M.N. was supported by a FPI Grant from the Spanish Government and a Grant from the Fundación Caja Madrid. Additional funding came from the Museum of Comparative Zoology and from the US National Science Foundation (#EF-0531757: ATOL: Collaborative Proposal: Assembling the Proteostome Tree of Life).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2011.05.018.

References


Author's personal copy


