Molecular Phylogenetics and Evolution 94 (2016) 473–478

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Short Communication

Phylogenomic analyses of a Mediterranean earthworm family (Annelida: Hormogastridae) ☆

Marta Novo [a,b,d,*], Rosa Fernández [b], Sónia C.S. Andrade [c], Daniel F. Marchán [d], Luis Cunha [a], Darío J. Díaz Cosín [d]

[a] Cardiff School of Biosciences, Cardiff University, BIOSI 1, Museum Avenue, Cardiff CF10 3AT, UK
[b] Museum of Comparative Zoology, Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA
[c] Departamento de Genética e Biologia Evolutiva, IB-USP, São Paulo, CEP 05508-090, Brazil
[d] Departamento de Zoología y Antropología Física, Facultad de Biología, Universidad Complutense de Madrid, C/ José Antonio Nováis 2, 28040 Madrid, Spain

ARTICLE INFO

Article history:
Received 18 May 2015
Revised 19 October 2015
Accepted 24 October 2015
Available online 31 October 2015

Keywords:
Next generation sequencing
Transcriptomics
Oligochaeta
Gene conflict
Phylogeny

ABSTRACT

Earthworm taxonomy and evolutionary biology remain a challenge because of their scarce distinct morphological characters of taxonomic value, the morphological convergence by adaptation to the uniformity of the soil where they inhabit, and their high plasticity when challenged with stressful or new environmental conditions. Here we present a phylogenomic study of the family Hormogastridae, representing also the first piece of work of this type within earthworms. We included seven transcriptomes of the group representing the main lineages as previously-described, analysed in a final matrix that includes twelve earthworms and eleven outgroups. While there is a high degree of gene conflict in the generated trees that obscure some of the internal relationships, the origin of the family is well resolved: the hormogastrid *Hemigastrodrilus* appears as the most ancestral group, followed by the ailoscolecid *Ailoscolex*, therefore rejecting the validity of the family Ailoscolecidae. Our results place the origin of hormogastrids in Southern France, as previously hypothesised.

1. Introduction

Hormogastridae includes middle to large-sized earthworms, currently comprising four genera, 31 species and subspecies or varieties, all endemic to the Western Mediterranean (Omodeo and Rota, 2008) and adapted to the driest soils, thanks to their lack of dorsal pores and aestivation capacity (Diaz Cosin et al., 2006). They have been recorded in Spain, France, Italy, Maghreb, and islands such as Sicily, Corsica and Sardinia (Fig. 1) and their distribution and evolutionary relationships have been probably affected by the palaeogeographical and palaeoecological events in the Mediterranean basin (Novo et al., 2015a). Hormogastrid diversity peaks at the NE Iberian Peninsula and SE France. This area has been hypothesised as the possible ancestral area for the family in a recent biogeographical study based on ancestral area reconstructions (Novo et al., 2015a). The known diversity is currently divided into four genera: *Hormogaster* Rosa, 1887 (19 species), *Hemigastrodrilus* Bouché, 1970 (1 species), *Vignysa* Bouché, 1970 (2 species) and *Xana* Díaz Cosín et al., 1989 (1 species). Phylogenetic trees inferred by Novo et al. (2011) recovered seven well-supported clades (Fig. 1): (i) *Xana*; (ii) *Hemigastrodrilus*; (iii) *Vignysa*; (iv) *Hormogaster elisae* in the central area of the Iberian Peninsula; (v) *Hormogaster* species from the NE area of the Iberian Peninsula; (vi) *Hormogaster* species within the *H. pretiosa* group and related from Catalonia and S Sardinia; and (vii) *Hormogaster* from the Tyrrhenian area (*H. redii* and *H. samnitica*). These results highlighted the paraphyly of the genus *Hormogaster* but the relationships among the main well-supported seven clades remained equivocal. Additionally, James and Davidson (2012) proposed the placing of Ailoscolecidae, containing only the species *Ailoscolex lacteospumosus,* in the synonymy of Hormogastridae. The same authors, recovered *Hemigastrodrilus* as sister clade of Lumbricidae with molecular analyses, whereas Bouche (1970) indicated the presence of Morren glands within this genus, as it is shown in lumbricids, but not in hormogastrids (Bouche, 1972). *Hemigastrodrilus* has been placed within Hormogastridae in different positions using molecular markers (e.g. Novo et al., 2015a). As a consequence, clarifying the phylogenetic position of both *Ailoscolex* and *Hemigastrodrilus* is necessary.

In the last few years, next generation sequencing (NGS) techniques have undergone a substantial development and its

☆ This paper was edited by the Associate Editor Marcos Perez-Losada.

* Corresponding author at: Environmental Toxicology and Biology, Facultad de Ciencias, UNED, C/ Senda del Rey 9, 28040 Madrid, Spain.
E-mail address: mnovo@ccia.uned.es (M. Novo).

http://dx.doi.org/10.1016/j.ympev.2015.10.026
1055-7903/

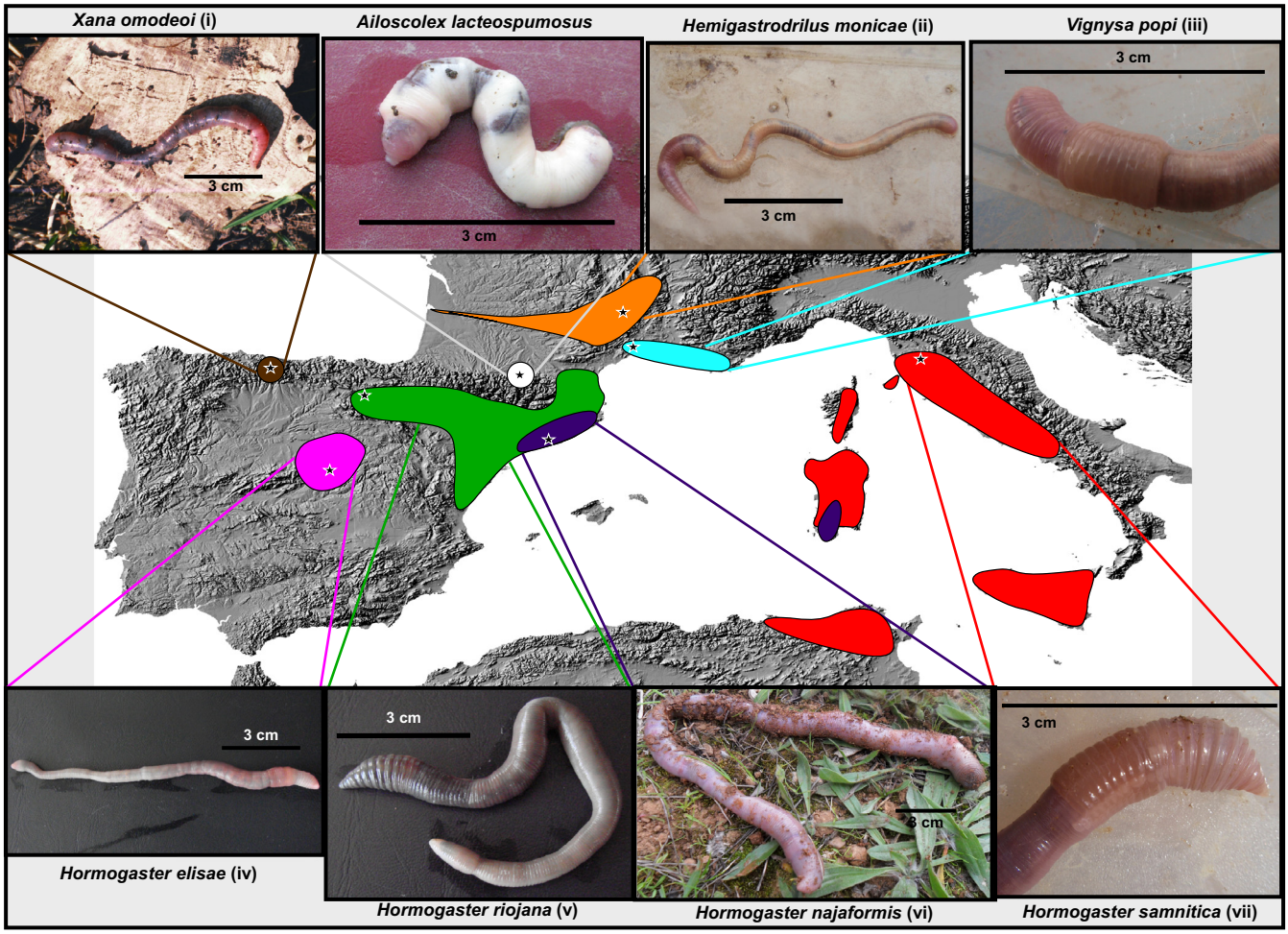

Fig. 1. Map showing the distribution of hormogastrids and the selected species for the phylogenomic analyses. Colour corresponds to that in trees by Novo et al. (2011) and to Fig. 2. A star indicates the sampling locality of the specimens. The number of clade following Novo et al. (2011), as described in the main text, is shown after the scientific name of the selected species. The species *Ailoscolex lacteospumosus* was not previously included in those phylogenetic analyses. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

application in phylogenetics of non-model organisms has become key to understand complex evolutionary scenarios that could not have been resolved before based on morphology or in a handful of genes (e.g., Fernandez et al., 2014, spiders; Laumer et al., 2015, planarians). Therefore, this emerging technique seems promising to tackle unresolved evolutionary relationships such as the case of hormogastrid phylogeny. Our intention was to shed light on the phylogenetic relationships of the family Hormogastridae (plus the putatively related Ailoscolecidae) through phylogenomics by sampling a representative of each of the main clades recovered by Novo et al. (2011) and *Ailoscolex*. This study represents the first phylogenomic investigation within earthworms.

2. Materials and methods

2.1. Taxon sampling

Transcriptome reads from *Hormogaster samnitica* and *H. elisae* were already available and retrieved from Riesgo et al. (2012) and Novo et al. (2013) respectively. We sampled six new transcriptomes (five hormogastrids plus *Ailoscolex*) covering the remaining clades from Novo et al. (2011) (see Fig. 1). We selected earthworms from different families as close outgroups: *Pontoscolex corethrurus* (Rhinodrilidae); *Eisenia fetida* and *Lumbricus rubellus* (Lumbricidae) and *Amynthas gracilis* (Megascolecidae). Their transcriptomes are part of different studies, some of which are still under preparation and their unassembled reads were used (Supplementary File 1). Finally, we added eleven more taxa as distant outgroups, including sipunculans, polychaetes, molluscs and nemerteans whose assembled transcriptomes were available (Lemer et al., 2015; Weigert et al., 2014).

Sampling localities and coordinates for the earthworm specimens are shown in Supplementary File 1. Leftover tissues are deposited in the Oligochaete Cryo collection of the Departamento de Zoología y Antropología Física, Universidad Complutense de Madrid (DZAF, UCM), Spain.

2.2. Library construction and transcript sequencing

Earthworms were preserved in RNA*later*® in the field. DNA was extracted and individuals were genotyped using cytochrome C oxidase subunit I (COI) primers LCO1490 and HCO2198 (Folmer et al., 1994). Sequences were compared to reference specimens previously dissected and taxonomically identified (see protocol and reference sequences in Novo et al. (2011)). For *Ailoscolex lacteospumosus*, seven molecular regions were amplified in order to further explore its relationship with the previously-sequenced hormogastrids and perform a topology test (see below): mitochondrial COI, 16S rRNA and tRNA Leu, Ala, and Ser (16S t-RNAs), one nuclear ribosomal gene (a fragment of 28S rRNA) and one nuclear

protein-encoding gene (histone H3) following Novo et al. (2011). GenBank accession numbers for *A. lacteospumosus* sequences are KP944199–KP944207.

Once the identity of the specimens was verified, RNA was extracted (body parts indicated in Supplementary File 1) by combining Trizol extraction with column purification (RNeasy mini kit QIAGEN), following the adapted protocol for earthworms in Novo et al. (2015b). RNA integrity was verified using Bioanalyzer (RNA Nano Chip) and concentration determined using NanoDrop (Thermo Scientific). Samples were prepared for cDNA library construction by BaseClear (www.baseclear.com) following the Tru-Seq (Illumina) protocol. In total six libraries were multiplexed for sequencing on Illumina Hi-Seq 2000 using a 100 bp paired end protocol. Fastq sequence files were generated using the Illumina Casava v. 1.8.2 pipeline. The data was initially checked for base quality and filtered for data passing the Illumina Chastity default parameters. Sequences containing adapters and/or PhiX control signal were removed with a filtering protocol by BaseClear. FastQC v. 0.10.0 was used for further quality assessment and to check that the average Phred values were above 33. All reads generated for this study are deposited in the European Nucleotide Archive (ENA) within the study PRJEB8761 (Supplementary File 1).

2.3. Sequence processing and orthology prediction

De novo transcriptome assemblies were performed using Trinity v. r2013-08-14 (Haas et al., 2013) with default parameters except for – path_reinforcement_distance, set to 50. CD-HIT-EST was used for redundancy reduction (Fu et al., 2012) in the raw assemblies (98% global similarity). Afterwards, identification of candidate Open Reading Frames (ORFs) within the transcripts was performed in TransDecoder (Haas et al., 2013) and predicted peptides were further filtered to select only one peptide per putative unigene, by choosing the longest ORF per Trinity subcomponent with a custom Python script (Fernandez et al., 2014). Predicted ORFs were assigned to orthologous groups using the Orthologous MAtrix (OMA v.099u) stand-alone algorithm (Altenhoff et al., 2011).

2.4. Phylogenomic analyses

We performed the phylogenetic analyses with seven matrices: (1) a large matrix with 23 taxa and 2320 orthogroups (50% gene occupancy, with genes present in at least 12 taxa); (2) a small matrix with 23 taxa and 355 orthogroups (75% gene occupancy, with genes present in at least 17 taxa). The next three matrices were built based on evolutionary rate, in order to account for rate of molecular evolution effect. Percent pairwise identity was employed as a proxy and calculated in Geneious v. 8.0.4 for the large matrix orthogroups: (3) 200 slowest-evolving genes, (4) 200 genes with intermediate evolutionary rate, (5) 200 fastest-evolving genes. Finally, we performed a matrix reduction approach with MAtrix Reduction (MARE v.0.1.2-rc; Meyer, 2011), which estimates informativeness of every orthogroup based on weighted geometry quartette mapping (Nieselt-Struwe and von Haeseler, 2001). Therefore two final matrices were generated: (6) Reduction of matrix (1), which resulted in the retention of 1927 orthogroups in 15 taxa and (7) Reduction of matrix (2), which resulted in the retention of 337 orthogroups in 17 taxa.

Orthogroups were aligned with MUSCLE v 3.6 (Edgar, 2004). To increase the signal-to-noise ratio and improve the discriminatory power of phylogenetic methods, we applied a probabilistic character masking with ZORRO (Wu et al., 2012) to account for alignment uncertainty, ran using default parameters. We discarded the positions assigned a confidence score below a threshold of 5 with a custom Python script prior to concatenation (using Phyutility 2.6; Smith and Dunn, 2008) and subsequent phylogenomic analyses.

Maximum likelihood (ML) inference was conducted with RAxML 7.7.5 (Stamatakis, 2006). PROTGAMMALG4X was selected as the best model of amino acid substitution. Best-scoring ML trees were inferred for each gene under the selected model from 100 replicates of parsimony starting trees. Bootstrap values were estimated with 100 replicates under the rapid bootstrapping algorithm. Bayesian analysis was conducted with Exabayes version 1.3.1 with default options (Aberer et al., 2014). Two runs were run in parallel with a parsimony starting tree. The analysis was parallelized over 32 nodes. By default, ExaBayes computes the average standard deviation of split frequencies (asdsf) every 5000 generations and stops the analysis once the asdsf is better than 5%. In order to reduce memory consumption, the memory versus runtime trade-off was set to 3 and the SEV-technique for gap columns on large alignments was enabled.

2.5. Congruence assessment and topology test

In order to investigate incongruence between individual gene trees, we inferred gene trees for each OMA group included in our supermatrices. For each aligned, ZORRO masked OMA group, best-scoring ML trees were inferred for each gene under the PROTGAMMALG4X model as implemented in RAxML 7.7.5 (Stamatakis, 2006). To assess potential conflict between these gene trees, we followed a supernetwork approach using SuperQ v.1.1 (Grunewald et al., 2013). This method permits visualisation of predominant intergenic conflict by building a supernetwork after decomposing all gene trees into quartets and assigning edge lengths based on quartet frequencies followed by the Gurobi optimization, and the filter parameter set as 0.1. SplitsTree v.4.13.1 (Huson and Bryant, 2006) was then used for network visualisation.

To test whether the main topologies recovered from the phylogenomic analyses were compatible with the one obtained from the molecular markers used in Novo et al. (2015a) with a wider species sampling, the sequences obtained for *A. lacteospumosus* were included into the alignment from the latter. Maximum likelihood (ML) analyses were performed with RAxML-HPC v.8.1.11 (Stamatakis, 2014) as implemented in the CIPRES Science Gateway 3.3 (http://www.phylo.org/index.php/portal/) using a partitioned GTR + I + Γ model of sequence evolution and estimating the support for the resulting topologies by 100 rapid bootstrap replicates (Stamatakis et al., 2008). The analysis was repeated including five different sets of constraints based on the results of the phylogenomic analyses. The trees obtained were compared using the test by Shimodaira and Hasegawa (1999) implemented in RAxML-HPC v.8.1.11 in CIPRES Science Gateway. We statistically tested the hypothesis from James and Davidson (2012) that *Hemigastrodrilus* is the sister group of Lumbricidae by comparing to our ML tree using the SOWH test in SOWHAT v.035 (Church et al., 2015) specifying a constraint tree and the WAG + Gamma model on Matrix (6), the main topology shown in Fig. 2.

3. Results

3.1. Transcriptomes assembly and orthology assignment

A summary of the statistics is shown in Supplementary File 1. Statistics for the remaining 11 outgroups can be seen in Lemer et al. (2015), from which those assemblies were retrieved. For earthworms, N50 ranged from 840 in *L. rubellus* (an outgroup) to 2134 in *A. lacteospumosus*. Regarding the 23 taxa, the number of

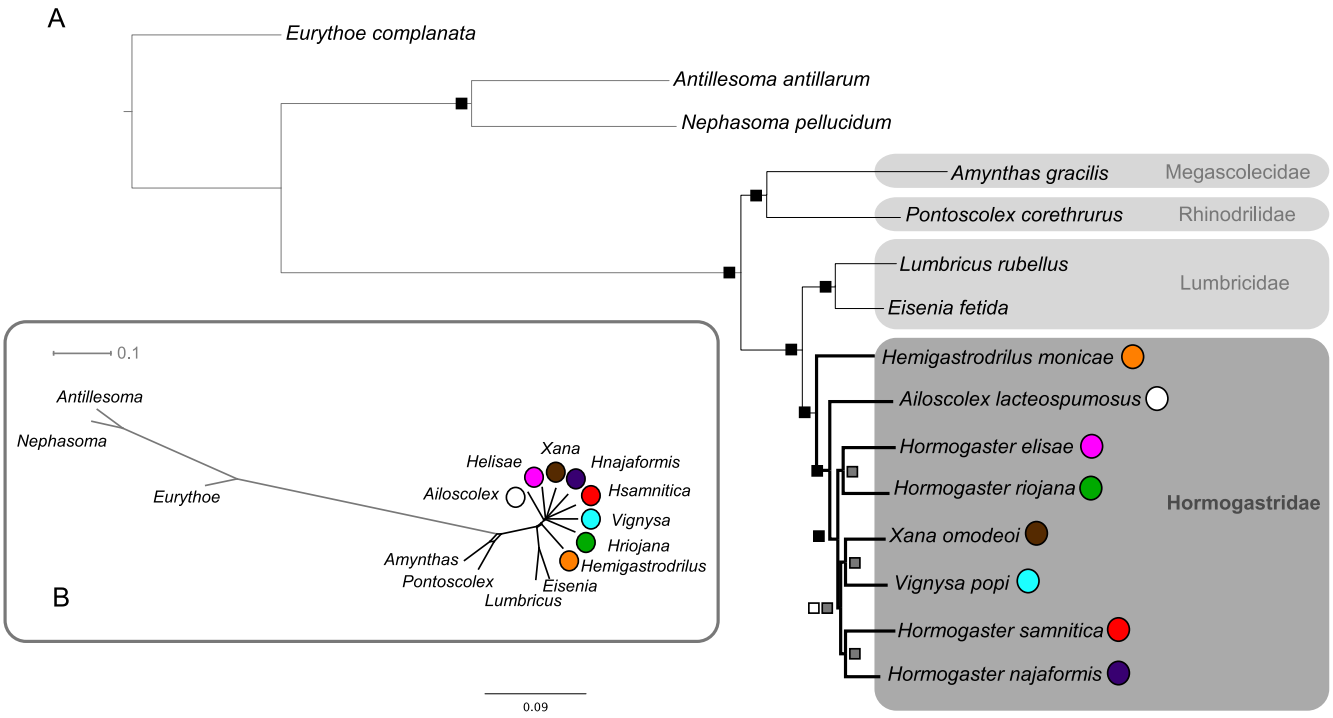

Fig. 2. (A) Phylogenetic hypothesis after analyses of the big matrix reduced with MARE (6) in RAxML ($\ln L = -4{,}169{,}347.56$). Support values for the ML analysis of the big matrix (1), the small matrix (2) and the reduced small matrix with MARE (7) are also implicit in the figure and represented with squares, since topology was similar. Black squares indicate maximum support of the node in the four analyses. Grey squares indicate support <60 in all the analyses and the white square represents the incongruence in matrix (1) for that node, whose analyses show the clades of *H. elisae* + *H. riojana* and *H. samnitica* + *H. najaformis* as sister clades. (B) SuperQ network derived from ML gene trees in the reduced matrix (6), the same one whose tree is represented within this figure, including 1927 orthogroups and 15 taxa.

peptides retained per taxon after Open Reading Frame (ORF) prediction and isoform filtering, was 1251–38,355 (16,246–38,355 for earthworms, Supplementary File 2) and the orthology assignments by OMA grouped those in a total of 100,081 orthogroups. From those, we selected 2320 orthogroups (50% gene occupancy) and 355 orthogroups (75% gene occupancy) as the base for all the analyses (see Section 2). Length of the matrices analysed is shown in Supplementary File 2. The gene coverage within hormogastrids had a maximum of missing data of about 25% for *H. samnitica* (Supplementary File 3, panel A). When applying the MAtrix Reduction approach this went to maximums of 26% for *H. samnitica* in matrix (6) (panel B) and 12% for *H. samnitica* in matrix (7) (panel C). Values of missing data per taxon are shown in Supplementary File 2.

3.2. Phylogeny of Hormogastridae and gene conflict

All of the analyses agree on the monophyly of Hormogastridae and include *Ailoscolex* within the family. They also agree that *Hemigastrodrilus* is the sister group to the rest of the family, followed by *Ailoscolex*. Finally, a clade including the four *Hormogaster* species plus *Xana* and *Vignysa* is highly supported. The two latter are represented as sister groups in all four analyses but with low support as well as the pairs *H. elisae* with *H. riojana* and *H. samnitica* with *H. najaformis*. The only difference found in the topologies of the four main analyses is that all but the complete big matrix (1) show the paraphyly of *Hormogaster*. The tree recovered by matrix (1) shows *Hormogaster* as monophyletic but still with a very low bootstrap value (42) that seems inconclusive. The ML hypothesis recovered for matrix (6) (i.e. Reduced matrix with MARE) is shown in Fig. 2, where the support for the rest of the main matrices (1) (2) and (7) is indicated. The BI analysis of matrices (1), (2), (6) and (7) never converged after several months of analyses, and therefore the results are not reported.

The topologies of matrices (3), (4) and (5) (ie, genes with different molecular rates) are shown in Supplementary File 4. The percent of pairwise identity within the conserved genes (3) was 89.7–99.8%; within the medium rate genes (4) was 76–78.4%, and within the fast genes (5) 49.5–64.4%. These trees presented again Hormogastridae as monophyletic, *Hemigastrodrilus* as the basal group and *Ailoscolex* included within the family and as sister to the remaining species. In addition, *Hormogaster* resulted paraphyletic and *Xana* and *Vignysa* appeared as sister clades. Some differences were recovered by conserved genes (see Supplementary File 4).

The supernetworks for the seven matrices analysed revealed the presence of intergene conflict as shown by their star-like topology within hormogastrids. The supernetwork for matrix (6) is shown in Fig. 2B and serves as illustration on how all of them look like. Supplementary File 4 includes the remaining supernetworks for each of the trees represented there.

The topologies compared in the SH test were: (i) *He. monicae* as the basal group followed by *A. lacteospumosus* as sister group of the remainder, as the most commonly recovered topology; (ii) topology presented in Fig. 2 and recovered by the main matrices; (iii) topology of the tree based on the fastest evolving genes; (iv) topology from the tree based on the middle evolving genes; (v) from the tree based on the more conserved genes (iii, iv and v are shown in Supplementary File 4). The unconstrained ML topology ($\ln L = -29123.998629$) was not found to be significantly better than the first four constrained hypotheses ($\ln L = -29141.553649/-29155.589733/-29148.007321/-29148.558373$). However, the topology derived from the most conserved genes was found to be significantly worse ($\ln L = -29160.529340$ for best suboptimal tree; $\alpha = 0.05$). Enforcing *Hemigastrodrilus* as the sister group of Lumbricidae showed as a significantly worse result (SOWH test: $n = 1000$, Delta-likelihood 2022.607, $p < 0.001$) than our most likely tree (Fig. 2).

4. Discussion

The results of this study represent, to our understanding, the first phylogenomic analysis within earthworms, including six new transcriptomes and following various approaches of matrix construction. The monophyly of the family Hormogastridae as well as the basal position of *Hemigastrodrilus* was consistent and well-supported in all the data sets. Moreover, *Ailoscolex*, proved to be included within the family, being sister clade to the group formed by *Hormogaster*, *Xana* and *Vignysa*. *Hemigastrodrilus* has been shown in low-supported different positions using mitochondrial and nuclear molecular markers (e.g. Novo et al., 2015a), even appearing outside hormogastrids as sister clade of Lumbricidae (James and Davidson, 2012), as mentioned above. In this context, our consistent results and the topology tests performed allow us to say that the base of the hormogastrid phylogeny is solved, with *Hemigastrodrilus* as the sister clade of the rest of the family members. Moreover, as previously proposed by James and Davidson (2012), *Ailoscolex* should be, with no doubt, included within Hormogastridae. The most apparent morphological characteristic of *Ailoscolex* is the lack of dorsal pores, which is an important character for hormogastrids. In addition, the anterior position of the clitellum, absence of Morren gland, presence of two anterior gizzards and multilamellar typhlosole fits within the characteristics of the family (Bouché, 1969; Bouche, 1972). Nevertheless, the masculine pore is in a more posterior position, which may be an apomorphy of this species. *Ailoscolex* is shown to be sister to *Hormogaster*, *Xana* and *Vignysa* with high support. The placement of these two genera (*Hemigastrodrilus* and *Ailoscolex*), distributed in France and the Pyrenees (see Fig. 1) in the base of the family, support the hypothesis that the origin of the family may have been located between Northeastern Spain and Southern France (Novo et al., 2015a). Both genera present a short clitellum when compared with the remaining hormogastrids (covering 9–11 segments), as well as two pairs of globular spermathecae that resemble those in lumbricids (Bouche, 1970, 1972; Bouché, 1969). Another consistent result but with lower support values is the fact that *Xana* and *Vignysa* are recovered as sister clades. Both genera present only two pairs of gizzards, whereas the remaining genera present three of them (except for *Ailoscolex* Bouche, 1969, 1972). *Hormogaster* genus is shown as paraphyletic in most of the analyses (except for conserved genes and also for the big complete matrix (1)), which agrees with results recovered by previous molecular phylogenies (Novo et al., 2011, 2015a).

The seemingly promising phylogenomic approach failed to resolve many of the nodes within Hormogastridae and the remaining relationships remain equivocal. The test performed by constraining the sequences from Novo et al. (2015a) to the different recovered topologies by phylogenomic analyses showed that all hypotheses, except for the tree recovered by conserved genes, are congruent with the widely-used molecular markers, with the main topology shown in Fig. 2 having a slightly lower ln*L* value. We found intergene conflict in all the different partitions analysed. Strong intergene conflict has been identified in annelids before (Andrade et al., 2015) and some other invertebrate taxa (e.g., Sharma et al., 2014). Several reasons for gene conflict have been indicated, including ancient origins of the group and rapid diversification (Sharma et al., 2014), which could be plausible for Hormogastridae (Novo et al., 2011) as well as different rates of evolution of different lineages (Sharma et al., 2014), again detected within this family (Novo et al., 2012). Further analyses (particularly extending the taxon sampling) are needed to elucidate the main causes of the observed conflict and to fully understand the evolution of this family.

Acknowledgments

We would like to thank Daniel Chana, Vanessa González, Horácio Montenegro, and Daniel Pass for bioinformatics advice and assistance, Sarah Lemer for providing us with outgroups assemblies, Sam James for information for locating *Ailoscolex*, Gonzalo Giribet for his advice and reviewers who helped improve a first version of this manuscript. We are also thankful to Joseph Ryan for his invaluable assistance with SOWHAT package. Marta Novo was supported by a Marie Curie Fellowship under EU FP7 program; Project IEF-GA-2012-329690. This study was funded by the Spanish Government; Projects CGL2013-42908-P and CGL2010/16032. The computations in this paper were run on the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard University and on the Biotecnologia Animal Laboratory from Prof. Luiz L. Coutinho, at ESALQ-USP. We are very grateful to Peter Kille and the team in Project NERC NE/I026022/1 for letting us use their *Amynthas* data and to the Ecoworm consortium financed by NERC that generated the data for *Lumbricus* (www.earthworms.og).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2015.10.026.

References

Aberer, A.J., Kobert, K., Stamatakis, A., 2014. ExaBayes: massively parallel bayesian tree inference for the whole-genome era. Mol. Biol. Evol. 31, 2553–2556.
Altenhoff, A.M., Schneider, A., Gonnet, G.H., Dessimoz, C., 2011. OMA 2011: orthology inference among 1000 complete genomes. Nucleic Acids Res. 39, D289–D294.
Andrade, S., Novo, M., Kawauchi, G.Y., Worsaae, K., Pleijel, F., Giribet, G., 2015. Articulating the "archiannelids": a phylogenomic approach to annelid relationships with emphasis on meiofaunal taxa. Mol. Biol. Evol. http://dx.doi.org/10.1093/molbev/msv157.
Bouche, M., 1970. Remarques sur quelques Lumbricina de France et consequences de la decouverte des noveaux taxons Vignysinae (Subfam. nov) et Diporodrilodae (Fam. nov.). Pedobiologia 10, 246–256.
Bouché, M., 1969. Ailoscolex lacteospumosus, n.gen, n.sp. Un ver de terre aux caractères morphologiques et biologuiques remarquables (Oligochaeta, Ailoscolecida, nov. fam.). Revue d Ecologie et de Biologie du Sol 4, 525–531.
Bouche, M.B., 1972. Lombriciens de France: ecologie et systematique. INRA, Paris.
Church, S.H., Ryan, J.F., Dunn, C.W., 2015. Automation and evaluation of the SOWH test for phylogenetic topologies with SOWHAT. Syst. Biol. http://dx.doi.org/10.1093/sysbio/syv055.
Diaz Cosin, D.J., Ruiz, M.P., Ramajo, M., Gutierrez, M., 2006. Is the aestivation of the earthworm Hormogaster elisae a paradiapause? Invertebr. Biol. 125, 250–255.
Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.
Fernandez, R., Laumer, C.E., Vahtera, V., Libro, S., Kaluziak, S., Sharma, P.P., Perez-Porro, A.R., Edgecombe, G.D., Giribet, G., 2014. Evaluating topological conflict in centipede phylogeny using transcriptomic data sets. Mol. Biol. Evol. 31, 1500–1513.
Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Biol. Biotechnol. 3, 294–299.
Fu, L.M., Niu, B.F., Zhu, Z.W., Wu, S.T., Li, W.Z., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28, 3150–3152.
Grunewald, S., Spillner, A., Bastkowski, S., Bogershausen, A., Moulton, V., 2013. SuperQ: Computing Supernetworks from Quartets. Ieee-Acm Trans. Comput. Biol. Bioinform. 10, 151–160.
Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., Leduc, R.D., Friedman, N., Regev, A., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494–1512.
Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267.

James, S.W., Davidson, S.K., 2012. Molecular phylogeny of earthworms (Annelida: Crassiclitellata) based on 28S, 18S and 16S gene sequences. Inver. Syst. 26, 213–229.

Laumer, C., Hejnol, A., Giribet, G., 2015. Nuclear genomic signals of the 'microturbellarian' roots of platyhelminth evolutionary innovation. eLife 4, e05503.

Lemer, S., Kawauchi, G.Y., Andrade, S.C.S., González, V.L., Boyle, M.J., Giribet, G., 2015. Re-evaluating the phylogeny of Sipuncula through transcriptomics. Mol. Phylogenet. Evol. 83, 174–183.

Meyer, B., 2011. MARE v. 0.1.2-rc: MAtrix REduction – a Tool to Select Optimized Data Subsets from Supermatrices for Phylogenetic Inference. See <http://mare.zfmk.de>.

Nieselt-Struwe, K., von Haeseler, A., 2001. Quartet-mapping, a generalization of the likelihood-mapping procedure. Mol. Biol. Evol. 18, 1204–1219.

Novo, M., Almodovar, A., Fernandez, R., Giribet, G., Diaz-Cosin, D.J., 2011. Understanding the biogeography of a group of earthworms in the Mediterranean basin – the phylogenetic puzzle of Hormogastridae (Clitellata: Oligochaeta). Mol. Phylogenet. Evol. 61, 125–135.

Novo, M., Almodovar, A., Fernandez, R., Trigo, D., Diaz-Cosin, D.J., Giribet, G., 2012. Appearances can be deceptive: different diversification patterns within a group of Mediterranean earthworms (Oligochaeta, Hormogastridae). Mol. Ecol. 21, 3776–3793.

Novo, M., Fernandez, R., Fernández Marchán, D., Trigo, D., Diaz-Cosin, D.J., Giribet, G., 2015a. Unearthing the historical biogeography of Mediterranean earthworms (Annelida: Hormogastridae). J. Biogeogr. 42, 751–762.

Novo, M., Lahive, E., Díez Ortiz, M., Matzke, M., Morgan, A.J., Spurgeon, D.J., Svedsen, C., Kille, P., 2015b. Different routes, same pathways: Molecular mechanisms under silver ion and nanoparticle exposures in the soil sentinel *Eisenia fetida*. Environ. Pollut. 205, 385–393.

Novo, M., Riesgo, A., Fernandez-Guerra, A., Giribet, G., 2013. Pheromone evolution, reproductive genes, and comparative transcriptomics in Mediterranean earthworms (Annelida, Oligochaeta, Hormogastridae). Mol. Biol. Evol. 30, 1614–1629.

Omodeo, P., Rota, E., 2008. Earthworm diversity and land evolution in three Mediterranean districts. Proc. California Acad. Sci. 59, 65–83.

Riesgo, A., Andrade, S.C.S., Sharma, P.P., Novo, M., Perez-Porro, A.R., Vahtera, V., Gonzalez, V.L., Kawauchi, G.Y., Giribet, G., 2012. Comparative description of ten transcriptomes of newly sequenced invertebrates and efficiency estimation of genomic sampling in non-model taxa. Front. Zool. 9.

Sharma, P.P., Kaluziak, S.T., Perez-Porro, A.R., Gonzalez, V.L., Hormiga, G., Wheeler, W.C., Giribet, G., 2014. Phylogenomic interrogation of arachnida reveals systemic conflicts in phylogenetic signal. Mol. Biol. Evol. 31, 2963–2984.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.

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-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst. Biol. 57, 758–771.

Weigert, A., Helm, C., Meyer, M., Nickel, B., Arendt, D., Hausdorf, B., Santos, S.R., Halanych, K.M., Purschke, G., Bleidorn, C., Struck, T.H., 2014. Illuminating the base of the annelid tree using transcriptomics. Mol. Biol. Evol. 31, 1391–1401.

Wu, M.T., Chatterji, S., Eisen, J.A., 2012. Accounting for alignment uncertainty in phylogenomics. PLoS ONE 7.