



## Original article

# Macroecological inferences on soil fauna through comparative niche modeling: The case of Hormogastridae (Annelida, Oligochaeta)



Daniel F. Marchán <sup>a,\*</sup>, Pablo Refoyo <sup>a,1</sup>, Rosa Fernández <sup>b</sup>, Marta Novo <sup>c</sup>, Irene de Sosa <sup>a</sup>, Darío J. Díaz Cosín <sup>a</sup>

<sup>a</sup> Departamento de Zoología y Antropología Física, Facultad de Biología, Universidad Complutense de Madrid, C/José Antonio Nováis 2, 28040 Madrid, Spain

<sup>b</sup> Museum of Comparative Zoology, Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA

<sup>c</sup> Environmental Toxicology and Biology, Facultad de Ciencias UNED, C/Senda del Rey 9, 28040 Madrid, Spain

## ARTICLE INFO

## Article history:

Received 25 January 2016

Received in revised form

12 May 2016

Accepted 17 May 2016

## Keywords:

Ecological niche models

ENMtools

Macroecological inferences

Soil fauna

Hormogastridae

## ABSTRACT

Ecological Niche Modeling (ENM) through MaxEnt and quantitative comparison techniques using ENMtools could facilitate ecological inferences in problematic soil dwelling taxa. Despite its ecological relevance in the Western Mediterranean basin, the ecology of the endemic family Hormogastridae (Annelida, Oligochaeta) is poorly known. Applying this comparative approach to the main clades of Hormogastridae would allow a better understanding of their ecological preferences and differences. One hundred twenty-four occurrence data belonging to four clades within this earthworm family were used as input to infer separate MaxEnt models, including seven predictor variables. Niche breadth, niche overlap and identity tests were calculated in ENMtools; a spatial Principal Components Analysis (sPCA) was performed to contrast with the realized niches. The highly suitable predicted ranges varied in their ability to reflect the known distribution of the clades. The different analyses pointed towards different ecological preferences and significant ecological divergence in the four above-mentioned clades. These results are an example of wide-scale ecological inferences for soil fauna made possible by this promising methodology, and show how ecological characterization of relevant taxonomic units could be a useful support for systematic revisions.

© 2016 Elsevier Masson SAS. All rights reserved.

## 1. Introduction

Macroecological studies comparing the ecological preferences of different soil taxa are almost absent from the literature (but see Ref. [1]): this is not necessarily a case of a lack of interest for this community, but most probably related to the difficulty of their study [2].

One approach which has facilitated the ecological inferences in these problematic groups is Ecological Niche Modeling (ENM), with MaxEnt [3] standing out among the different methodologies due to

its high performance when including presence data only. It has been implemented in several groups (ground beetles [4], termites [5], millipedes [6]) including earthworms [7]: used MaxEnt to study the effect of large-scale ecological variables in the distribution of *Hormogaster elisae*, corroborating its high predictive power and the ability to reflect accurately its soil preferences.

Additionally, the implementation of several existing indices and statistical tests in the software ENMtools [8] has allowed the quantitative comparison of Ecological Niche Models (ENMs) between related species, including niche overlap, niche breadth and testing for statistically significant differentiation. Some recent studies have proven the usefulness of these methodologies to answer diverse biological questions, applying them to different animal groups. For example [9], found ecological niche differentiation in two cryptic beetle species, using the fact as support for their status as valid species. Ref. [10] studied niche overlap and niche breadth in three cryptic bat species complexes, as part of their research on how environmental factors and ecological interactions influenced their speciation. However, these promising

**Abbreviations:** ENM, Ecological Niche Modeling; sPCA, Spatial Principal Component Analysis; ROC-AUC, Receiver Operating Characteristic-Area Under the Curve, AUC in shorter form; TRANGE, Mean Diurnal Range; ISOTHER, Isothermality; TCOLD, Mean Temperature of the Coldest Quarter; PRDRY, Precipitation of Driest Month; LITHO, Lithology; VEGET, Land cover; ANTHRO, Human influence; HGI, Highest gain in isolation; HDGO, Highest decrease in gain when omitted.

\* Corresponding author.

E-mail address: [danifermch@gmail.com](mailto:danifermch@gmail.com) (D.F. Marchán).

<sup>1</sup> Daniel F. Marchán and Pablo Refoyo contributed equally.

and insightful methodologies have been scarcely used to address similar questions in soil animals (and especially earthworms).

Several authors have justified the utility of both modeling distributional patterns and ecological niche characteristics above the species level (e.g., [11–13]) when individual species have scarce occurrence data.

This approach appears promising for a relatively obscure earthworm family, the Hormogastridae Michaelsen 1900. It is the second most diverse earthworm family in the western Palearctic region, after the Lumbricidae Claus 1876. They play an important ecological role as deep-burrowing endogeics in the western Mediterranean basin [14], processing great amounts of soils expelled as casts [15]. In some places, such as Sardinia, they were shown to be dominant in abundance in earthworm communities [16], being adapted to drought and impoverished soils [17,18]. To understand better the role of hormogastrids in soil ecosystems, it is necessary to comprehend their ecological preferences, adaptations and response to environmental variables. However, those have been scarcely studied, mainly focusing on one particular species: *Hormogaster elisae* Álvarez 1977. The larger body size and associated slower reproductive rate [19] of most species in the family, together with their scattered distribution and difficult capture (their deep burrowing requires intensive digging efforts), have discouraged their laboratory and field research. The only work on their ecological preferences [20] corroborated the presence of most hormogastrids in soils with low nutrients content, with a preference for soils more basic and fine-textured than the ones observed for *H. elisae*.

The phylogeny of Hormogastridae has been clarified in the last years using molecular evidence: after [21–23] it has been divided in 9 clades. *Xana*, *Vignysa*, *Hemigastrodrilus* and *Ailoscolex* had already been described as independent genus, but the other five remain artificially joined in the catch-all genus *Hormogaster*. Four of these latter clades (informally termed Central Iberian, Northeastern Iberian, Tyrrhenian, and Disjunct) have wide geographic ranges and a high number of known populations, which makes them suitable for an in-depth macroecological study. Understanding their ecological preferences and the differences in their ecological niches would be helpful as an additional support for their future definition as new genera in the taxonomic revision of the family.

The main aim of this work is to obtain macroecological inferences for the four main clades of *Hormogaster* through comparative niche modeling as an example of the potential of this methodology. We used all the available geo-referenced presence locations to obtain the ENMs for each of the clades, and ENMtools and niche space visualization to perform quantitative comparisons. Our objectives were: i) to predict the distribution of the main clades of *Hormogaster* in their home range; ii) to find the environmental variables with a stronger influence in their distribution; iii) to study the overlap, breadth and statistical differentiation of their ecological niches; and iv) to compare their potential and realized niches.

This study could be potentially useful as a base for wide-scale ecological inferences in other groups of earthworms and soil fauna, a key element on most ecosystems around the world.

## 2. Materials and methods

### 2.1. Training data

One hundred twenty-four presence localities were used to train the models (detailed in Suppl. Material 1): 44 for the Central Iberia clade, 29 for the Tyrrhenian clade, 30 for the Northeastern Iberia clade and 21 for the Disjunct clade. All the presence data were obtained in sampling campaigns by the authors, ensuring high reliability; they also constitute a good representation of the known

ranges of the species, defined after more than a century of field works by other researchers in the Mediterranean.

The Central Iberia clade corresponds to the *H. elisae* morpho-species, which comprises at least five cryptic lineages [24]. [7] included several new populations, which considerably widened its known range; their genetic variability is currently being researched.

The Northeastern Iberia clade comprises a high number of closely related species (see Suppl. Material 1) with high morphological variability in their diagnostic characters [20]. They inhabit Northeastern Spain and a small region of Southeastern France, with most of their diversity located in Catalonia (Spain).

The Disjunct clade includes the Sardinian populations of *Hormogaster pretiosa* Michaelsen 1889 (a taxonomically problematic species into which other unrelated species were wrongfully assigned [21]) confined to the southwestern part of the island, *Hormogaster najiformis* Qiu & Bouché 1998 and *H. regina* Rota 2016 from Catalonia (Spain) and an assembly of undescribed related forms in the latter region.

The Tyrrhenian clade includes *Hormogaster redii* Rosa, 1887, *Hormogaster samnitica* Cognetti, 1914 and their subspecies [25] found deep genetic divergence pointing to them being composed of cryptic lineages. They are distributed around the Tyrrhenian Sea, occupying most of Sardinia, Corsica, Tuscan Archipelago, the Tyrrhenian side of Italy from Tuscany to Naples, Sicily and a small area between northern Algeria and Tunisia [16].

### 2.2. Environmental variables

The large-scale variables potentially relevant for the biology of Hormogastridae were chosen as predictor variables to model its distribution, as described below.

Four bioclimatic variables were selected from Worldclim (<http://www.worldclim.org/> accessed 12/05/2016): Mean Diurnal Range-BIO2 (**TRANGE**) and Isothermality-BIO 3 (**ISOTHER**) are suitable to represent the influence of extreme temperature variation (both daily and across the year) on earthworm distribution in the Mediterranean region. Mean Temperature of the Coldest Quarter-BIO11 (**TCOLD**) was chosen to reflect the impact of soil freezing and low temperatures on the activity of earthworms. Precipitation of Driest Month-BIO 14 (**PRDRY**) is likely to reflect the availability of water in the soil across the year (an essential requirement for earthworms) and the severity of drought periods.

As topographical variable we selected Lithology -PAR-MAT-DOM2, Second level code for the dominant parent material of the STU from the European Soil Database Raster Library 1 km × 1 km ([http://eussoils.jrc.ec.europa.eu/ESDB\\_Archive/ESDB\\_data\\_1k\\_raster\\_intro/ESDB\\_1k\\_raster\\_data\\_intro.html](http://eussoils.jrc.ec.europa.eu/ESDB_Archive/ESDB_data_1k_raster_intro/ESDB_1k_raster_data_intro.html) accessed 12/05/2016)- (**LITHO**). Lithology is likely to influence indirectly Hormogastridae ecology through a wide range of correlated variables, including the structure and biochemical characteristics of soils.

CORINE 2006 Land Cover (version 12/2009: <http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2006-clc2006-100-m-version-12-2009> accessed 12/05/2016) – (**VEGET**) was chosen to incorporate information about vegetation and land use, which are widely known to influence earthworm distribution (e.g. [26]).

As large, deep burrowing species as most Hormogastridae are affected negatively by human disturbance [27], the 'Human footprint' data set -representing the human influence on land surface [28] – (**ANTHRO**) was selected to include the effect of anthropic activities on habitat suitability.

These variables were the same that successfully predicted the distribution of *H. elisae* (and relevant soil characteristics) in Ref. [7]. Precipitation of the coldest quarter was replaced by Mean

Temperature of the Coldest Quarter to represent the inactivity period in the coldest months.

After testing for collinearity, no significant correlation above 0.8 was found between the predictor variables, thus none was removed.

### 2.3. Ecological niche modeling

The following parameters of the model as implemented in MaxEnt were considered by default: a maximum number of 500 iterations, a convergence-threshold limit of 0.00001, 10,000 points as number maximum of background points and regularization multiplier equal to 1, as recommended by Ref. [3]. Background points were randomly selected from the home range of the family: Iberian Peninsula, Southern France, Corsica, Sardinia, Italy and Sicily. Northern Africa was not included due to the lack of data for some variables. Due to the gaps in the knowledge of earthworm fine-scale distribution, sampling bias and dispersal ability we chose the background point selection method which requires fewer assumptions over other more restrictive approaches.

Ten replicates were run for each model. Each time, a 30% of the sample records were randomly removed without replacement (through the subsample option) to be used as test points in order to measure the quality of the model, and the 70% of records were used to build the model (e.g. Ref. [29]). The final model was constructed with the average of the replicates.

The ROC-AUC (Receiver Operating Characteristic- Area Under the Curve) technique has been implemented in MaxEnt to analyze the goodness of fit of the analysis in contrast to other models of evaluation, since it avoids the problem of selecting threshold values [30]. It is capable of measuring the model ability to discriminate between the sites of species presence from the areas of absence [31–34].

### 2.4. Niche analysis

The statistics **niche overlap** [35] –which reflects the similarity between the niche models- and **niche breadth** [36] were calculated using ENMTools 1.3.

**Niche overlap** analyses were applied to pairwise comparisons of all clades, based on the values of two indexes, Schoener's *D* [37] and Hellinger's *I* [35]. Low values of these statistics indicate little overlap, while values close to 1 mean great similarity.

Levin's B1 (inverse concentration) and B2 (uncertainty) [38] were obtained for each clade as a measure of **niche breadth**: higher values of these indexes indicate a broader niche.

The **identity test** (or test of niche equivalency) was run in ENMtools to test for statistically significant differences between the obtained ENMs. A null distribution is produced by pooling together all occurrence data from two clades and randomly separating them in two sets, generating two ENMs and obtaining their overlap values (with 100 separate replicates). If the observed *D* and *I* overlap values are lower than the confidence interval of the null distribution, the null hypothesis of both niches being equivalent is rejected.

We decided not to employ the background test (or test of niche similarity), which relies in the *a priori* definition of the available space for the different studied clades. According to [8] the delimitation of the available background is a critical aspect of the test, and we lacked enough biological or geographical justifications to define it with confidence.

### 2.5. Realized niche visualization

Spatial principal components analysis (sPCA) was used

following [10] to visualize the ecological niches of the different clades. The values of all the environmental variables were extracted for each presence record, and a PCA was performed in STATISTICA 7 (StatSoft Inc, Tulsa, OK, USA). The resulting factor scores for each record were imported into ArcGIS 10.0 as x, y coordinates. Minimum convex polygons were drawn to delimit each clade's realized niche (the actually occupied environmental space) and their individual areas and the area of overlap between the niches were measured.

## 3. Results

### 3.1. Ecological niche models

All models showed high predictive power, with good to very good average AUC values [39]: Tyrrhenian 0.951, Central Iberia 0.990, Northeastern Iberia 0.856, Disjunct 0.960. The geographical representation of the predicted suitability values is shown in Fig. 1. The predicted highly suitable areas roughly matched the known extent of their ranges for the Central and Northeastern Iberia clades; for the Disjunct and Tyrrhenian clades the areas were respectively wider and narrower than expected according to literature [16]. A few isolated areas outside the known range of Hormogastridae were also predicted as highly suitable.

The relative contributions of the variables to each model are shown in Fig. 2; the variables with highest relative contribution, highest gain in isolation-HGI (the one which improves the model the most when the rest are removed) and highest decrease in gain when omitted-HDGO (the one which worsen the model the most when removed) for each model are shown in Table 1. The preferred classes for the two categorical variables (lithology-LITHO and land cover-VEGET) are shown in Table 2. In all of these cases important differences were found between the clades.

### 3.2. Niche analysis

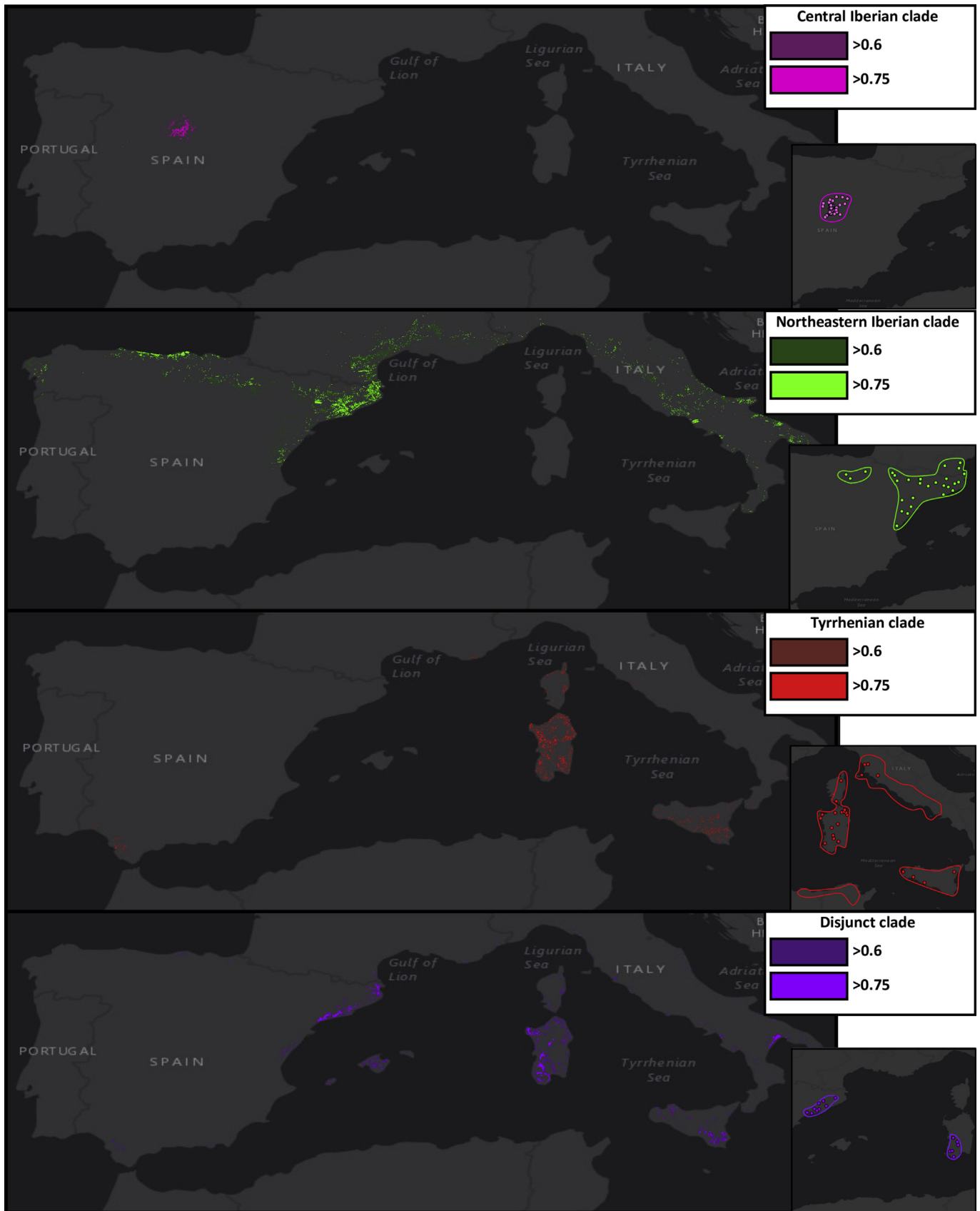
The Northeastern Iberia clade showed the broadest niche (indicated by the highest value of B1 and B2: 0.36 and 0.95 respectively), while the Central Iberia clade showed the narrowest (B1 and B2: 0.03 and 0.79). The most similar niches were Tyrrhenian and Disjunct, which is indicated by the highest values of niche overlap (*I*: 0.85, *D*: 0.60). Northeastern Iberia was more similar to them (*I*: 0.50–0.70, *D*: 0.24–0.42) than Central Iberia, which showed the lowest overlap with the rest (*I*: 0.14–0.25, *D*: 0.04–0.08).

The Identity test (or test of niche equivalency) showed significant differences between all the niches by rejecting the null hypothesis of niche identity with very high confidence ( $\alpha = 0.01$ ) both for *I* and *D* statistics.

### 3.3. Realized niche analysis

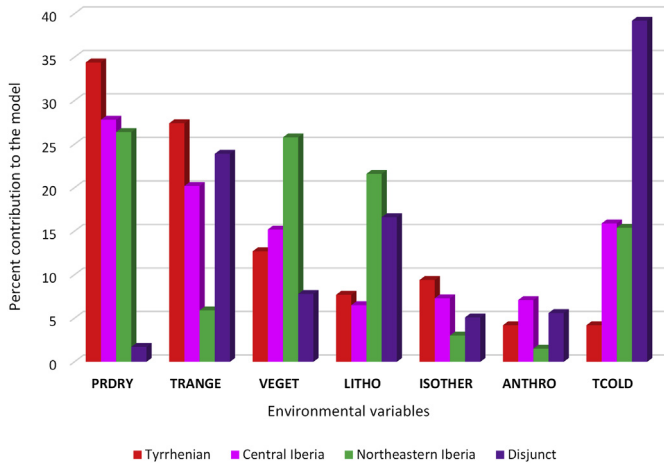
The first two PCA factors chosen for the realized niche representation explained 55.21% of the variance. The first factor was highly and positively correlated to Mean Temperature of the Coldest Quarter-(TCOLD). Isothermality (ISOTHER) and Mean Diurnal Range (TRANGE) were negatively correlated to the second factor.

The Central Iberia niche is highly divergent from the Tyrrhenian and Disjunct niches, which in turn are spatially close (and show high mutual overlap percentages). The Northeastern Iberia niche, with the largest area, widely overlaps with the other three niches, with the highest percentages corresponding to Central Iberia and Tyrrhenian niches (Fig. 3).



**Fig. 1.** Habitat suitability maps for the four clades in this study. Only values above 0.6 (darker shade) and 0.75 (lighter shade) are shown. The small maps show the occurrence data for each clade, and the colored outline shows the known range of the clade. Central Iberian clade: pink; Northeastern Iberian clade: green; Tyrrhenian clade: red; Disjunct clade: purple. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)





**Fig. 2.** Percent contribution of the environmental variables to the Ecological Niche Models (ENMs). Mean Diurnal Range-(TRANGE), Isothermality-(ISOTHER), Mean Temperature of the Coldest Quarter-(TCOLD), Precipitation of Driest Month-(PRDRY), Lithology – (LITHO), Land cover-(VEGET), Human influence-(ANTHRO).

**Table 1**

Variables with highest relative contribution, highest gain in isolation (HGI) and highest decrease in gain when omitted (HDGO) in the habitat suitability models of each clade. Mean Diurnal Range-(TRANGE), Mean Temperature of the Coldest Quarter-(TCOLD), Precipitation of Driest Month-(PRDRY), Lithology – (LITHO), Land cover-(VEGET).

|                      | Highest contribution                       | HGI   | HDGO  |
|----------------------|--|-------|-------|
| Tyrrhenian           | PRDRY (34.4%) TRANGE (27.4%) VEGET (12.7%) | TCOLD | VEGET |
| Central Iberian      | PRDRY (27.8%) TRANGE (20.2%) TCOLD (15.9%) | PRDRY | PRDRY |
| Northeastern Iberian | PRDRY (26.4%) VEGET (25.8%) LITHO (21.6%)  | PRDRY | PRDRY |
| Disjunct             | TCOLD (39.2%) TRANGE (23.9%) LITHO (16.6%) | TCOLD | TCOLD |

**Table 2**

Preferred lithologies and land cover classes for the four clades, obtained from the habitat suitability models.

|                      |            |  |
|----------------------|------------|--|
| Northeastern Iberian | Lithology  | Calcareous rocks, fluvial clays, silts and loams   |
|                      | Land cover | Permanently irrigated land, vineyards, moors and heatland, transitional woodland-shrub   |
| Disjunct             | Lithology  | Pyroclastic rocks, acid regional metamorphic rocks, unconsolidated deposits  |
|                      | Land cover | Permanently irrigated land, vineyards, fruit tree/berry plantations, annual crops associated with permanent crops, land principally occupied by agriculture with significant areas of natural vegetation, and natural grasslands |
| Tyrrhenian           | Lithology  | Pyroclastic rocks, acid regional metamorphic rocks, unconsolidated deposits, marine and estuarine clays and silts, residual and redeposited clays from calcareous rocks  |
|                      | Land cover | Natural grasslands and sclerophyllous vegetation   |
| Central Iberian      | Lithology  | Consolidated-clastic-sedimentary rocks, arenites and acid to intermediate plutonic rocks   |
|                      | Land cover | Natural grasslands and sclerophyllous vegetation   |

## 4. Discussion

### 4.1. Predicted suitable ranges

While the habitat suitability predictions mostly adjusted to our knowledge of hormogastrid earthworm distribution, some highly suitable areas were highlighted where they have never been found after intensive sampling (but being deep burrowers with a patchy distribution leaves the possibility of false negatives [7]). According to [36], absence from some part of the predicted niche suggest the action of unidentified factors excluding the species from these locations, such as competition or dispersal limitation. A combination of paleogeographical events [16,21,26] and competition [40] with other earthworms (mainly the family Lumbricidae) could explain some of the most remarkable absences. The Northeastern Iberia clade has not been found in the highly suitable zone of Cantabria and the Basque Country: this region was covered by the sea until

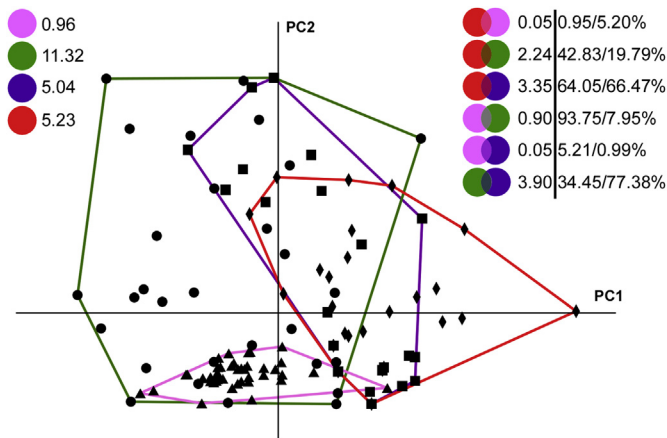
the Upper Tertiary, then it was likely colonized from Aquitaine (France) by its dominant lumbricid fauna, with *Lumbricus friendi* and the giant *Scherotheca* sp among them [41]. Something similar could explain the absence of the Disjunct clade in the Balearic Islands. All their area except the main mountain ranges in Majorca got submersed during episodes in the Oligocene and Miocene, and the available environment was likely already inhabited by another big-sized endogeic lumbricid genus, *Postandrilus* sp. [42–44].

Other highly suitable areas in Galicia, Portugal and Andalucía are too far from the main range of Hormogastridae to have been colonized during its evolutionary history. It is well known that endogeic earthworms have poor active dispersal capabilities, as [45] found for *H. elisae*, which hardly moves to find a mating partner. Passive dispersal is still possible for earthworms, so geographical barriers or extinctions through their long evolutionary history could also explain their absence.

A specific case of this overprediction was the Disjunct clade predicted suitable range, which was wider than expected. These earthworms are absent from Northern Sardinia, Corsica, the Tuscan Archipelago, continental Italy and Sicily (thoroughly sampled areas

as reviewed by Ref. [16]), but a few highly suitable areas were found in them. Competitive exclusion from suitable habitats by the overlapping Tyrrhenian clade seems unlikely, as they have been found in sympatry (showing vertical niche segregation with *H. redii* in Sardinian localities –*pers. obs.*). One possible reason for this phenomenon is the important influence of tectonic events in the dispersion of these animals [25,46]: the absence of the Disjunct clade from the microplates that drifted eastwards would explain their absence from areas that the Tyrrhenian clade managed to reach.

The other main finding was the inferior fitting of the predicted suitable range of the Tyrrhenian clade to our previous knowledge. Ref. [16] found a strong presence of *H. redii* and *H. samnitica* in continental Italy, while our model showed few highly suitable areas. The low predictive performance could be explained by the scarce number of Italian occurrences included in the training data despite the strong sampling effort.



**Fig. 3.** Realized niches of the four hormogastrid clades obtained in the spatial PCA analysis. The points from each clade are shown as follows: Central Iberia – triangles, Eastern Iberia – circles, Disjunct – squares, Tyrrhenian – diamonds. The areas of the realized niches are shown in the top left corner, and the overlap between them (in absolute value and percentage of their respective area) in the top right corner. Central Iberian clade: pink; Northeastern Iberian clade: green; Tyrrhenian clade: red; Disjunct clade: purple. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Overall, our results confirmed the suitability of ecological niche modeling at above-species level for soil-dwelling fauna, as seen in Ref. [47]; but also highlighted the importance of comprehensive sampling through the known range, and the influence of the specific characteristics of the clades in its performance.

#### 4.2. Ecological characterization and niche differentiation

The differences in the most relevant environmental variables for each clade hint at differential ecological preferences (as seen in Ref. [36]). Precipitation showed a strong influence on Tyrrhenian, Northeastern and Central Iberia clades but their response differed (as shown by response curves, [Suppl. Material 2](#)): the Tyrrhenian clade showed a preference for the driest climate, Northeastern Iberia showing increased probability of presence in wetter habitats and the Central Iberia earthworms in between.

Meanwhile, low temperature affected Disjunct clade distribution significantly. There is some empirical evidence for the differential effect of these variables. For example, earthworms of Northeastern Iberia and Disjunct clades have been found in the same location in a humid but cold month; the former were active but the latter were aestivating. Conversely, Disjunct clade earthworms were found active in drier soils under warmer conditions (*pers. obs.*).

Even when its influence was moderate compared to climatic variables, the clades showed different responses to human influence, with the Central Iberia and Disjunct clades being the most affected.

Land cover heavily influenced the Northeastern Iberia clade, as lithology did for them and the Disjunct clade. The preferred land uses and lithological classes found for the four clades constitute a useful preliminary description of the habitats these earthworms select.

While it can be argued that small scale soil variables should be considered to study the ecological preferences and differences between these soil-dwelling taxa, to our knowledge this is not a concern. On one hand [7] showed strong correlation between large-scale environmental variables and the most relevant soil variables affecting *H. elisae* distribution: this supports our studied variables covering to some extent the variability of lower-scale variables. On

the other hand, preference for local soil conditions should be expected to be less conserved across clades than large-scale environmental variables, thus being less suitable for macroecological analysis at above-species level.

Together with the significant differences between the ecological niches shown by the identity test, these results highlight the ecological divergence between the studied clades; in a similar way [47] highlighted the ecological divergence of the four main clades of the soil-dwelling mite harvestmen (Cyphophthalmi). It does not only reinforce their biological relevance as evolutionary entities; it also provides additional characters to the diagnosis of the future generic system in Hormogastridae: altitudinal preference, selected habitats and lithologies, and susceptibility to drought and cold periods being the most remarkable.

According to the niche conservatism hypothesis, higher niche similarity should point to closer phylogenetic affinities between the clades. Interestingly, Tyrrhenian and Disjunct have been hypothesized to be sister clades based on morphology (sharing multiple spermathecae) and phylogenomic analysis [22]. Unfortunately, the scarce occurrence data for individual species of Hormogastridae hinders any attempt of correlating niche similarity and phylogenetic relatedness at a finer scale, but it could be more easily implemented to other groups of soil fauna with more widespread species.

#### 4.3. Niche breadth and overlap: predicted vs realized niche

The predictions of niche breadth from Levin's index and sPCA areas were in agreement. This suggests there is a good correspondence between the predicted and realized niches of the four clades.

It is worth noting that field observations indicate *H. redii* has an exceptionally wide ecological valence [16], which was not reflected in a highest niche breadth of its (Tyrrhenian) clade. This indicates analyzing above-species level clades could sometimes mask the ecological particularities of some of their members.

The high overlap between Tyrrhenian and Disjunct realized niches is concordant with the overlap shown by *D* and *I* indexes. The high overlap between Northeastern Iberia and the other realized niches is also in agreement with these indexes, except for Northeastern-Central, which is surprisingly lower according to the latter.

These results show sPCA as a valuable complementary analysis in ecological niche studies, which provides additional insight in aspects like niche filling.

## 5. Conclusions

MaxEnt allowed predicting with varying accuracy the highly suitable range of the main Hormogastridae clades; some exceptions constitute an interesting starting point for hypothesis on which factors shaped their distribution (dispersal ability, biotic interactions and paleogeographical events among them).

We found evidence for significant divergence in the ecological niches of the studied clades. The analyses also provided information about their differential ecological preferences. In the same way that ecological characters have been proposed to reinforce species delimitation, comparative studies at higher taxonomic levels such as this one appear suitable to characterize and delimit putative genera (or other taxonomic categories).

The Northeastern Iberia clade niche was the broadest, widely overlapping with the rest, while the Tyrrhenian clade was narrower than previously thought. This kind of preliminary results, combined with the reconstruction of the family's divergence, constitute a promising starting point to study their macroecology and macroevolution.

The integration of powerful, objective tools for phylogenetic and macroecological inference will lead to a deeper understanding of the processes operating at wide scale on soil fauna.

## Acknowledgments

The authors would like to thank the two anonymous reviewers for their contribution in improving this manuscript. This research was funded by the project CGL2013-42908-P from the Spanish Government. D.F.M was supported by a fellowship granted by the Universidad Complutense de Madrid. R. F. was supported by a fellowship from the Fundación Ramón Areces. M.N was supported by a Marie Curie Fellowship under EU FP7 program; Project IEF-GA-2012-329690 and a Postdoctoral Fellowship from the Spanish Government.

## Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.ejsobi.2016.05.003>.

## References

- [1] T. Decaëns, Macroecological patterns in soil communities, *Glob. Ecol. Biogeogr.* 19 (3) (2010) 287–302.
- [2] J. Smith, S. Potts, P. Eggleton, Evaluating the efficiency of sampling methods in assessing soil macrofauna communities in arable systems, *Eur. J. Soil Biol.* 44 (3) (2008) 271–276.
- [3] S.J. Phillips, R.P. Anderson, R.E. Schapire, Maximum entropy modeling of species geographic distributions, *Ecol. Model.* 190 (2006) 231–259.
- [4] P.H. Crawford, B.W. Hoagland, Using species distribution models to guide conservation at the state level: the endangered American burying beetle (*Nicrophorus americanus*) in Oklahoma, *J. Insect Conserv.* 14 (5) (2010) 511–521.
- [5] D.S. Maynard, T.W. Crowther, J.R. King, R.J. Warren, M.A. Bradford, Temperate forest termites: ecology, biogeography, and ecosystem impacts, *Ecol. Entomol.* 40 (3) (2015) 199–210.
- [6] P.E. Marek, W.A. Shear, J.E. Bond, A redescription of the leggiest animal, the millipede *Illacme plenipes*, with notes on its natural history and biogeography (Diplopoda, Siphonophorida, Siphonorhinidae), *ZooKeys* 241 (2012) 77.
- [7] D.F. Marchán, P. Refoyo, M. Novo, R. Fernández, D. Trigo, D.J. Díaz Cosín, Predicting soil micro-variables and the distribution of an endogeic earthworm species through a model based on large-scale variables, *Soil Biol. Biochem.* 81 (2015) 124–127, <http://dx.doi.org/10.1016/j.soilbio.2014.10.023>.
- [8] D.L. Warren, R.E. Glor, M. Turelli, ENMTools: a toolbox for comparative studies of environmental niche models, *Ecography* 33 (3) (2010) 607–611, <http://dx.doi.org/10.1111/j.1600-0587.2009.06142.x>.
- [9] O. Hawlitschek, N. Porch, L. Hendrich, M. Balke, Ecological niche modelling and ndna sequencing support a new, morphologically cryptic beetle species unveiled by DNA barcoding, *PLoS One* 6 (2) (2011), <http://dx.doi.org/10.1371/journal.pone.0016662>.
- [10] H. Santos, J. Juste, C. Ibáñez, J.M. Palmeirim, R. Godinho, F. Amorim, P. Albes, H. Costa, O. De Paz, G. Pérez-Suárez, S. Martínez-Alos, G. Jones, H. Rebelo, Influences of ecology and biogeography on shaping the distributions of cryptic species: three bat tales in Iberia, *Biol. J. Linn. Soc.* 112 (1) (2014) 150–162, <http://dx.doi.org/10.1111/bj.12247>.
- [11] J. Heino, J. Soiminen, Are higher taxa adequate surrogates for species-level assemblage patterns and species richness in stream organisms? *Biol. Conserv.* 137 (2007) 78–89.
- [12] E.A. Hadly, P.A. Spaeth, C. Li, Niche conservatism above the species level, *P. Natl. Acad. Sci. U. S. A.* 106 (2009) 19707–19714.
- [13] J.A.F. Diniz, P. De Marco, B.A. Hawkins, Defying the curse of ignorance: perspectives in insect macroecology and conservation biogeography, *Insect Conserv. Diver* 3 (2010) 172–179.
- [14] M.B. Bouché, *Lombriciens de France, Écologie et Systématique*, INRA, Paris, 1972, p. 671.
- [15] D.J. Díaz Cosín, R.P. Moro, J.V. Valle, M.H. Garvín, D. Trigo, J.B. Jesús, Producción de heces de *Hormogaster elisae* en diferentes tipos de cultivos en laboratorio, *Bol. R. Soc. Esp. Hist. Nat. Sec. Biol.* 92 (1996) 177–184.
- [16] P. Omodeo, E. Rota, Earthworm diversity and land evolution in three Mediterranean districts, *Proc. Calif. Acad. Sci.* 59 (5) (2008) 65–83.
- [17] P. Hernández, Estudio de la distribución horizontal de *Hormogaster elisae* (Oligochaeta, Hormogastridae) en el Molar y de los factores que le regulan, Doctoral dissertation, Universidad Complutense de Madrid, 2005.
- [18] D.J. Díaz Cosín, M.P. Ruiz, M. Ramajo, M. Gutiérrez, Is the aestivation of the earthworm *Hormogaster elisae* a paradiapause? *Invertebr. Biol.* 125 (3) (2006) 250–255, <http://dx.doi.org/10.1111/j.1744-7410.2006.00057.x>.
- [19] K.L. Lee, *Earthworms. Their Ecology and Relationships with Soils and Land Use*, Academic Press, Sidney, 1985, 411 Pages.
- [20] M. Novo, A. Almodóvar, R. Fernández, D. Trigo, D.J. Díaz Cosín, G. Giribet, Appearances can be deceptive: different diversification patterns within a group of mediterranean earthworms (Oligochaeta, Hormogastridae), *Mol. Ecol.* 21 (2012) 3776–3793, <http://dx.doi.org/10.1111/j.1365-294X.2012.05648.x>.
- [21] M. Novo, A. Almodóvar, R. Fernández, G. Giribet, D.J. Díaz Cosín, Understanding the biogeography of a group of earthworms in the Mediterranean basin—The phylogenetic puzzle of Hormogastridae (Clitellata: Oligochaeta), *Mol. Phylogenet. Evol.* 61 (2011) 125–135.
- [22] M. Novo, R. Fernández, S.C.S. Andrade, D.F. Marchán, L. Cunha, D.J. Díaz Cosín, Phylogenomic analyses of a Mediterranean earthworm family (Annelida: Hormogastridae), *Mol. Phylogenet. Evol.* 94B (2016) 473–478.
- [23] D.F. Marchán, R. Fernández, M. Novo, D.J. Díaz Cosín, New light into the hormogastrid riddle: morphological and molecular description of *Hormogaster joseantonioi* sp. n. (Annelida, Clitellata, Hormogastridae), *ZooKeys* 414 (2014) 1–17, <http://dx.doi.org/10.3897/zookeys.414.7665>.
- [24] M. Novo, A. Almodóvar, D.J. Díaz Cosín, High genetic divergence of hormogastrid earthworms (Annelida, Oligochaeta) in the central Iberian Peninsula: evolutionary and demographic implications, *Zool. Scr.* 38 (5) (2009) 537–552.
- [25] M. Novo, R. Fernández, D.F. Marchán, D. Trigo, D.J. Díaz Cosín, G. Giribet, Unearthing the historical biogeography of Mediterranean earthworms (Annelida: Hormogastridae), *J. Biogeogr.* 42 (2015) 751–762.
- [26] A.M.L. Lindahl, I.G. Dubus, N.J. Jarvis, Site classification to predict the abundance of the deep-burrowing earthworm L. Vadose Zone J. 8 (4) (2009) 911–915.
- [27] M.G. Paoletti, The role of earthworms for assessment of sustainability and as bioindicators, *Agriculture, Ecosyst. Environ.* 74 (1) (1999) 137–155.
- [28] E.W. Sanderson, M. Jaiteh, M.A. Levy, K.H. Redford, A.V. Wannebo, G. Woolmer, The Human Footprint and the Last of the Wild: the human footprint is a global map of human influence on the land surface, which suggests that human beings are stewards of nature, whether we like it or not, *BioScience* 52 (10) (2002) 891–904.
- [29] J.L. Tellería, T. Santos, P. Refoyo, J. Muñoz, Use of ring recoveries to predict habitat suitability in small passerines, *Distrib. Divers* 18 (11) (2012) 1130–1138.
- [30] A. Lehmann, J.M. Overton, J.R. Leathwick, GRASP: generalized regression analysis and spatial prediction, *Ecol. Model.* 157 (2002) 189–207.
- [31] A.H. Fielding, J.F. Bell, A review of methods for the assessment of prediction errors in conservation presence/absence models, *Environ. Conserv.* 24 (1997) 38–49.
- [32] R. Engler, A. Guisan, L. Rechsteiner, An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data, *J. Appl. Ecol.* 41 (2004) 263–274.
- [33] J. Elith, C.H. Graham, R.P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. Hijmans, F. Huettmann, J.R. Leathwick, A. Lehmann, L. Jin, L. Lohmann, B.A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. McC. Overton, A.T. Peterson, S.J. Phillips, Novel methods improve prediction of species distributions from occurrence data, *Ecography* 29 (2006) 129–151.
- [34] J. Elith, S.J. Phillips, T. Hastie, M. Dudík, Y.E. Chee, C.J. Yates, A statistical explanation of MaxEnt for ecologists, *Divers. Distrib.* 17 (2010) 43–57.
- [35] D.L. Warren, R.E. Glor, M. Turelli, Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution, *Evolution* 62 (11) (2008) 2868–2883, <http://dx.doi.org/10.1111/j.1558-5646.2008.00482.x>.
- [36] T. Nakazato, D.L. Warren, L.C. Moyle, Ecological and geographic modes of species divergence in wild tomatoes, *Am. J. Bot.* 97 (4) (2010) 680–693, <http://dx.doi.org/10.3732/ajb.0900216>.
- [37] T.W. Schoener, The Anolis lizards of Bimini: resource partitioning in a complex fauna, *Ecology* 49 (1968) 704–726.
- [38] R. Levin, *Evolution in Changing Environments: Some Theoretical Explorations*, Princeton University Press, 1968.
- [39] J.A. Swets, Measuring the accuracy of diagnostic systems, *Science* 240 (1988) 1285–1293.
- [40] M. Ramajo, Relaciones interespecíficas de lombrices de tierra en una parcela de El Molar (Madrid), Doctoral dissertation, Universidad Complutense de Madrid, 2009.
- [41] P. López Echezarreta, Cambios en las sinusias de lumbricidos (Oligochaeta, Lumbricidae) en las plantaciones de coníferas exóticas en el País Vasco (Guipúzcoa), *Munibe* 2 (1980) 695–766.
- [42] E. Ramos, L. Cabrera, H.W. Hagemann, W. Pickel, I. Zamarreño, Palaeogene lacustrine record in Mallorca (NW Mediterranean, Spain): depositional, palaeogeographic and palaeoclimatic implications for the ancient south-eastern Iberian margin, *Palaeogeogr. Palaeoclimatol.* 172 (2001) 1–37, [http://dx.doi.org/10.1016/S0031-0182\(01\)00277-2](http://dx.doi.org/10.1016/S0031-0182(01)00277-2).
- [43] M.M. Bauzá-Ribot, D. Jaume, J.J. Fornós, C. Juan, J. Pons, Islands beneath islands: phylogeography of a groundwater amphipod crustacean in the Balearic archipelago, *BMC Evol. Biol.* 11 (1) (2011) 221, <http://dx.doi.org/10.1186/1471-2148-11-221>.
- [44] M. Pérez-Losada, J.W. Breinholt, P.G. Porto, M. Aira, J. Domínguez, An earthworm riddle: systematics and phylogeography of the Spanish lumbricid postandrius, *PLoS One* 6 (11) (2011), <http://dx.doi.org/10.1371/journal.pone.0028153>.
- [45] M. Novo, A. Almodóvar, R. Fernández, M. Gutiérrez, D.J. Díaz Cosín, Mate choice of an endogeic earthworm revealed by microsatellite markers, *Pedobiologia* 53 (6) (2010) 375–379.

- [46] R. Fernández, A. Almodóvar, M. Novo, M. Gutiérrez, D.J. Díaz Cosín, Earthworms, good indicators for palaeogeographical studies? Testing the genetic structure and demographic history in the peregrine earthworm *Aporrectodea trapezoides* (Dugès, 1828) in southern Europe, *Soil Biol. Biochem.* 58 (2013) 127–135.
- [47] G. Giribet, P.P. Sharma, L.R. Benavides, S.L. Boyer, R.M. Clouse, B.L. de Bivort, D. Dimitrov, G.Y. Kawauchi, J. Murienne, P.J. Schwendinger, Evolutionary and biogeographical history of an ancient and global group of arachnids (Arachnida: Opiliones: Cyphophthalmi) with a new taxonomic arrangement, *Biol. J. Linn. Soc.* 105 (1) (2012) 92–130, <http://dx.doi.org/10.1111/j.1095-8312.2011.01774.x>.