Spain is a highly suitable scenario for phylogeographic studies due to its complex geological history and the variety of its environments [1]. Due to its geographical position, Spain is under the influence of the Atlantic Ocean and the Mediterranean Sea, resulting in a wide range of climates including desert, Mediterranean, Alpine and Atlantic [2] comprising two biogeographic regions: Eurosiberian and Mediterranean [3]. Some differences in presence of earthworm species between these regions have been observed, but many lumbricids (like Eiseniella tetraedra) are present in both [4]. Moreover, Spain was one of the most important Pleistocene glacial refugia in the European subcontinent [5,6] and acted in both [4]. Due to its geographical position, Spain is under the influence of the Atlantic Ocean and Mediterranean Sea, Spain has different climates, from desert to Atlantic. We sampled the parthenogenetic earthworm Eiseniella tetraedra in two different biogeographical zones in Spain, in order to study their genetic diversity and test their potential distinctiveness. Moreover, we evaluated the presence or absence of two different lineages (Eurosiberian and Mediterranean) found in other parthenogenetic earthworms such as Aporrectodea trapezoides and A. rosea. We studied the molecular markers COI, 16S and 28S. E. tetraedra presents a high diversity in Spain (one COI haplotype every two individuals were found) and no clear geographical patterns except for diffuse patterns along the Guadarrama River basin. In contrast, worldwide localities were more homogeneous with low diversity, to be confirmed with further samples. After morphological study, no correlation was found between phylogenetic relationships and the diagnostic characters for the previously described subspecies in E. tetraedra.

Keywords: Earthworm Eiseniella tetraedra Genetic variability Glaciations Population structure Phylogeography

Spain is a highly suitable scenario for phylogeographic studies due to its complex geological history and the variety of its environments [1]. Due to its geographical position, Spain is under the influence of the Atlantic Ocean and the Mediterranean Sea, resulting in a wide range of climates including desert, Mediterranean, Alpine and Atlantic [2] comprising two biogeographic regions: Eurosiberian and Mediterranean [3]. Some differences in presence of earthworm species between these regions have been observed, but many lumbricids (like Eiseniella tetraedra) are present in both [4]. Moreover, Spain was one of the most important Pleistocene glacial refugia in the European subcontinent [5,6] and acted as a species repository of northern lands [7–9]. The phylogeographic study of E. tetraedra within Spain and Europe could shed light on these processes and be useful for elaborating European settlement models.

The focus of this piece of work is Eiseniella tetraedra Savigny, 1826 [10] (Annelida, Oligochaeta), a parthenogenetic tetraploid [11] and cosmopolitan earthworm [12] with a riparian lifestyle (closely linked to edges of water). It is possible that E. tetraedra’s biology is closer to aquatic rather than terrestrial animals, due to its strict dependence on water. Together with its wide distribution and its parthenogenetic reproduction, this makes E. tetraedra an interesting candidate species for investigating biogeographic and genetic diversity patterns in soil, water and ecotone systems.

Knowledge on evolutionary processes in soil fauna is only starting to flourish. Some studies of different groups (harvestmen [13,14], oribatid mites [15], myriapods [16], earthworms [17], or caecilians [18]) have revealed shared characteristics: high genetic variability, low vagility in a particular geological time, strong population structure, and high intraspecific divergence [19]. Because of these attributes, earthworms have been proposed as good candidates for phylogeographic studies, since paleogeographic events appear to have great relevance in their current distribution [17,20].

However, this is not true for all earthworms, and examples have been found in which little or no population structure exists, like Aporrectodea icterica Savigny, 1826, Allolobophora chlorotica.
Savigny, 1826 [21,22] or the case of invasive species like Amynthas corticis Kinberg, 1867 [23]. Relative absence of population structure was also found in the semi aquatic earthworms Glyphiidrilus vangieniaensis Panha & Chanabun, 2011 and Glyphiidrilus mekongensis Panha & Chanabun, 2012. They showed isolation by distance, but not vicariance, due to river flow connecting their populations [24]. Studies of cosmopolitan, parthenogenetic earthworms have found different results: Octolasion tyrtaeum Savigny, 1826 showed high homogeneity with a low number of clones across northern Europe while Aporrectodea rosea Savigny, 1826 showed high genetic variability in the same region [25]. Moreover, studies of molecular markers found two distinct, differentiated lineages, one present in Eurosiberian region and the other one present in Mediterranean region, not only in A. rosea but also in Aporrectodea trapezoides Dugès, 1828 [26,27]. Therefore, there is the possibility that E. tetraedra follow either of these patterns, with the added uncertainty of the presence of sexual forms (which have not yet been found).

Only studies based on enzymatic polymorphisms have been conducted for E. tetraedra [28,29], which showed high clonal variability with no clones shared between Sweden, Finland and Russia [30], yet they found shared clones between Northern Norway and Southern Finland with no clear distribution patterns [28]. These works had two additional interesting outcomes. Firstly, morphological variability showed no correlation to the enzyme patterns in the studied populations. Secondly, a higher clonal diversity was found in the lower course compared to that upstream in the studied rivers [30], due to its dispersion through water. This pattern has also been observed in other clonal and flightless invertebrates [31,32].

The morphological and anatomical simplicity of soil dwelling animals, such as earthworms, has limited the establishment of a robust taxonomy. It remains anchored, to some degree, in subjective criteria of each author. Within this context, integrative approaches including molecular information are becoming more popular to solve the phylogenetic position of conflictive taxa [33]. Morphological variability in Eiseniella tetraedra and its taxonomy are an example of a field open to such research. At least three subspecies have been described [10,34,35] the position of the male pores being the diagnostic character. Due to the high polymorphism of this character in different populations (pers. observ.), the correlation of morphological and genetic variability of specimens collected in different localities could support or contradict the use of this character to differentiate subspecies.

With this background, E. tetraedra constitutes a suitable model for testing not only different hypotheses about riparian and soil fauna but also for testing some animal distribution patterns in Europe. The aim of this work is the study of the genetic structure of E. tetraedra in two different biogeographical zones (Eurosiberian and Mediterranean) with the following objectives: 1. To assess whether this species shares the general traits of most soil fauna (high genetic variability and strong population structure). 2. To test whether different lineages are found in European and Mediterranean areas as previously shown in other species. 3. To study the distribution of genetic variability along a river basin to test whether E. tetraedra follows described patterns for other riparian animals. 4. To evaluate the taxonomic subspecies division based on male pore position, by correlating genetic and morphological variability.

1. Material and methods

1.1. Sampling and morphological studies

We collected specimens from thirty localities from the Spain: 14 from the Northwestern area and 16 from the central region (Fig. 1, see geographical coordinates in Supplementary Tables 1 and 2). For comparative purposes, all sequences of the mitochondrial cytochrome c oxidase subunit I gene (COI) belonging to Eiseniella tetraedra were downloaded from the BOLD system web (Supplementary Table 3). Moreover, we collected one more population in Wales, United Kingdom.

All individuals were hand collected, washed in distilled water, fixed in 96% EtOH and preserved at -20 °C. A fragment of the body wall from the posterior end was separated for genetic analysis. Later, morphological studies and dissections were performed, focusing on: length, dry weight, clitellum position, tubercula pubertatis, male pore position, seminal vesicles number and position. Presence of iridescence in spermathecae and male funnels indicates presence of sperm [36]. In order to test the existence of sexual specimens in this supposed strictly parthenogenetic species, presence or absence of iridescence of spermathecae and male funnels was studied.

1.2. DNA extraction, gene amplification and sequencing

Total genomic DNA was extracted using the DNeasy Blood and Tissue Kit (QIAGEN). Furthermore, two mitochondrial markers (COI and 16S + tRNAs Leu, Ala and Ser) and one nuclear marker (28S) were amplified.

For COI (632 bp) we used the primers and PCR conditions used by Pop et al. [37]. For 16S-tRNAs (775 bp) and for 28S (806 bp) primers and PCR conditions from Fernández et al. [27] were used. PCR products were sequenced by Macrogen Europe Inc. (Holland).

1.3. Data analysis

Sequences were aligned in MAFFT v.7 [38] and concatenated in Bioedit [39]. Haplotype of each gene were retrieved by DNAsp v.5 [40]. Phylogenetic trees based on the concatenated sequence (2270 bp) were built through Maximum Likelihood using the software RaxML v7.0.3 [41] and Bayesian Inference with MrBayes v3.1.2 [42] both implemented in the Cypress Science Gateway [43]. GTR+Γ+I was chosen by jModelTest2 [44] as the best-fit substitution model. Bayesian analysis consisted of two parallels runs of 10 millions of generations and 20% of the trees were discarded as burn-in. For the ML analysis rapid bootstrapping was conducted including 1000 replicates. Hormogaster elisae Alvarez, 1977, Lumbricus rubellus Hoffmeyser, 1843, Dendrobaena subtile Rossa, 1893, Eisenia oliveri Bouché, 1854 and Prosectolokus biaraculatus Bouché, 1972 were chosen as outgroups and their sequences were retrieved from GenBank (Supplementary Table 4). Furthermore, uncorrected pairwise distances for 16S-tRNAs and COI were computed in Arlequin v.3.5 [45].

A COI based haplotype network was constructed using TCS version 1.21 [46] with statistical parsimony and a connection limit of 95%.

We also obtained haplotype and nucleotide diversities of the localities from Guadarrama river basin (in order to compare the different parts of the river) and of each lineage retrieved from phylogenetic analysis by DNAsp v.5 [40].

Finally, statistical analyses of morphological data (Kruskal-Wallis test) were conducted in Statistica v.7. In order to test whether genetic and morphological data were congruent and thus the taxonomic value of morphology in this case.

2. Results

A total of 113 haplotypes were identified among 271 sequences for COI gene, 40 haplotypes within 58 sequences for 16S-tRNAs and 3 haplotypes within 28 sequences for 28S.
2.1. Phylogenetic analysis

Both Maximum Likelihood and Bayesian approaches presented trees with congruent topology (Fig. 2). Haplotypes nested in six well supported lineages (A to F). Lineages A, B, E and F showed higher genetic diversity (with 18.18%, 19.93%, 19.69% and 22.27% of total haplotypes respectively) than lineages C and D (with 9.09% and 11.36% of haplotypes). As shown by the short branch lengths of the tree and reticulated structure of haplotype networks (Fig. 2) lineages presented a high internal homogeneity. In contrast, the different lineages showed deep divergences with long branches.

2.2. Haplotype and lineage distribution

Studied localities showed high haplotypic variability. Most lineages, except B and C, appeared in both regions (Eurosiberian and Mediterranean), lacking a pronounced geographical structure. In Spain, lineage B is exclusively Mediterranean and C Eurosiberian. The rest of the sequences of the world (RSW), were retrieved only in lineages B and E and showed rather low haplotypic variability. The number of sequences/specimens included in each lineage is: A, 33; B, 98; C, 13; D, 18; E, 68; F, 45.

As haplotype networks show (Fig. 2) regional haplotypes were found, either region-specific (Eurosiberian or continental Mediterranean) or even locality-specific. On the other hand, interregional haplotypes (shared between both regions) were found as well.

Fig. 3 shows the distribution of the lineages in the studied regions (maintaining the color code of Fig. 2). Both localities with presence of a single lineage and localities with several lineages were found. In the central area of Spain (Fig. 3C) the lineages are differentially distributed along the River Guadarrama basin: some lineages predominate in the upper course, while others have a stronger presence in the middle and lower course. Fig. 3D shows a higher lineage diversity in Spain and Wales, while the rest of localities around the world are very homogeneous. Lineage B is distributed in the western localities of Europe and North America (with as many as 90 sequences), and lineage E was found in the eastern localities of Europe (Austria) and Oceania (New Zealand).

2.3. Genetic diversity

As it shown in Table 1, genetic distances based on 16S-tRNAs were lower than those based on COI, due to 16S-tRNAs being a more preserved region of the mitochondrial genome. A certain degree of variability within lineages was observed, but divergence between them was remarkably higher. Values between lineages were considerably lower than those found between the outgroups, although some of them (specially lineage A) are near (or within) the ambiguous gap between intraspecific and interspecific divergence in earthworms proposed by Chang and James [47] –9 to 15%.

Haplotype and nucleotide diversities were high at all localities from the Guadarrama river basin (Table 2). The highest diversity was found in the middle course, while the lowest diversity was found in the lower course.
High haplotypic but low nucleotide diversity values were found within lineages, meaning that haplotypes within each lineage were abundant but very similar to each other (Table 3).

2.4. Morphological studies

Collected individuals showed wide variability in the studied morphological characters. In terms of internal characters, spermathecae and male funnels were never iridescent, even being absent in many individuals.

Length, weight and number of segments also showed a great degree of variability, but no significant differences were found between lineages (Supplementary Figs. 1–3). Morphological data is presented in Supplementary Table 5.

Regarding the position of the male pore, we found identical haplotypes with different state of this character.

3. Discussion

Phylogenetic analysis and pairwise distances showed six divergent, internally homogeneous lineages of *Eiseniella tetraedra*. This pattern suggests that not enough time would have elapsed for the differentiation of the haplotypes in each lineage due to regular bottlenecks caused by a constant founder effect or selective sweeps.

The most basal divergence was between lineage A and the rest of lineages. These results suggest that lineage A could represent an ancestral group with a stable demographic history due to its high nucleotide and haplotype diversity [48].

The geographic distribution of lineages (all present in Eurosiberian and Mediterranean areas except B and C) and the dispersed geographical distribution of haplotypes within lineages supports the absence of strong population structure and geographical structure, despite some lineages showing predominance in certain regions. However, clearly defined Eurosiberian and Mediterranean lineages as in *Aporrectodea rosea* and *A. trapezoides* [26,27] were not found. The most extreme case of this is lineage C. It presents haplotypes from Northwestern Spain and from Wales, which could be explained by human-mediated dispersal. The historical connection between these areas (i.e. the Britonia Breton enclave in northern Lugo [49]) supports this possibility.

Lineages B and E included specimens of the rest of the world (RSW), in addition to specimens from Spain. It is known that recent glaciations could have led to the extinction of most of the North European populations of earthworms [50]. A possible explanation could be a recolonization from both Spanish regions (Eurosiberian and Mediterranean), which acted as glacial shelters, to Europe. A recolonization from other refugia (like Ibalic or Balcanic peninsulas) is also probable, as seen in other groups of animals and plants [5]. Including individuals from those refugia in future studies is necessary in order to provide a wider view of the biogeography of this species in Europe. *E. tetraedra* has expanded from glacial shelters to Scandinavia in approximately 15,000 years, which indicates a not exceedingly low vagility. While such active colonization seems unlikely, it could be explained by passive dispersion...
along rivers or by human mediated transport (as suggested for Sparganophillus tamesis Benham, 1892 [51]). The distribution of the Lineages B and E in the westernmost and easternmost localities of the studied area suggest two independent colonization events from the highly diverse Spain or other refugia. These findings suggest a certain geographical ancestral structure modified by climate and dispersive history of the species.

The majority of haplotypes found were regional or even unique in each locality, as previously found by Terhivuo and Saura [52] using allozymes. Interregional haplotypes were also found, shared between Northwestern and central area of Spain. Moreover, haplotypes (based in COI) shared between Northwestern Spain and Wales were found. Terhivuo et al. [29] didn’t find shared clones between Russia, Finland and Sweden but they found shared clones between north Norway and south Finland, separated by about 1200 km [28].

Table 1

<table>
<thead>
<tr>
<th>COI/16S-tRNAs</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.97</td>
<td>2.22</td>
<td>3.21</td>
<td>3.45</td>
<td>2.52</td>
<td>2.77</td>
</tr>
<tr>
<td>B</td>
<td>9.79</td>
<td>2.67</td>
<td>2.3</td>
<td>2.06</td>
<td>2.79</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>9.71</td>
<td>4.93</td>
<td>0.58</td>
<td>2.64</td>
<td>3.04</td>
<td>4.06</td>
</tr>
<tr>
<td>D</td>
<td>10.25</td>
<td>5.53</td>
<td>5.2</td>
<td>1.65</td>
<td>2.84</td>
<td>3.66</td>
</tr>
<tr>
<td>E</td>
<td>7.94</td>
<td>8.13</td>
<td>8.41</td>
<td>9.42</td>
<td>0.47</td>
<td>2.9</td>
</tr>
<tr>
<td>F</td>
<td>7.75</td>
<td>8.19</td>
<td>7.88</td>
<td>8.86</td>
<td>5.86</td>
<td>0.84</td>
</tr>
</tbody>
</table>

Table 2

<table>
<thead>
<tr>
<th></th>
<th>Number of samples</th>
<th>Number of haplotypes</th>
<th>H diversity</th>
<th>I diversity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper course</td>
<td>49</td>
<td>29</td>
<td>0.82</td>
<td>0.11458</td>
</tr>
<tr>
<td>Middle course</td>
<td>22</td>
<td>19</td>
<td>0.97</td>
<td>0.20719</td>
</tr>
<tr>
<td>Lower course</td>
<td>21</td>
<td>8</td>
<td>0.59</td>
<td>0.06364</td>
</tr>
</tbody>
</table>

Fig. 3. Lineage distribution in A: the localities from Northwestern Spain, B: A micro-scale study in Carnota, A Coruña, Galicia, Spain from localities 1 to 8 (area indicated by a square in A), C: Lineages distribution in the central area of Spain, D: Lineages distribution in all studied areas. Colors used are the same as in Fig. 2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
In the present study, two opposite patterns of genetic diversity were found. On one hand, localities with representatives from only one lineage or even only one haplotype were found. On the other hand we found heterogeneous localities, with representatives from two or more lineages and/or a great variety of haplotypes. This result suggests at least two possible interpretations: competitive exclusion between lineages (or haplotypes), and the sequential arrival of founder populations followed by the coexistence of lineages and/or haplotypes. Terhivuo and Saura [53] thought that the absence of interregional clones in a country is due to dispersive history. Possibility of different colonization events in each locality suggests complex evolutionary histories in each population, following a random pattern [30]. The co-existence of both diversity patterns in geographically close populations, together with the long-range (more than 2000 km) presence of shared haplotypes, give as a result an unexpectedly ‘messy’ phylogeography for *Eiseniella tetraedra*.

A very small subset of the *E. tetraedra* genetic variability was represented, with the available data, in the areas outside Spain and Wales (RWS): only 15 haplotypes in 90 sequences from 8 different countries, compared to the total 129 haplotypes in 271 sequences. This pattern is similar to the one described by Mathieu and Davies [50] in France based on species diversity instead of genetic diversity: according to the authors this latitudinal gradient of diversity would be caused by the post-glaciation dispersive history of the different lineages.

In terms of haplotype and nucleotide diversity in the Guadarrama river basin in the central area of Spain, localities didn’t seem to follow the dispersive pathway found by Terhivuo and Saura [53] in Ume and Vindel rivers (Sweden) where the lower course presented the highest genetic diversity. Instead, in our study, the middle course of the Guadarrama river basin showed the highest variability being the lower course the less diverse. Environmental conditions in the lower course are expected to be less suitable for *E. tetraedra* (like intermittent water availability in the tributary streams, use of water for irrigated crops and other human consumption, and presence of contamination, which will be tested in further studies), which could negatively affect their genetic variability. Thus, the middle course would be more representative of the lower course in said study, accumulating haplotypes from upstream which would not thrive downstream due to the unfavorable conditions.

Uniparental and parthenogenetic reproduction of *Eiseniella tetraedra* were inferred by Gavrilov [54] and the chromosomal study performed by Muldal [55] confirmed said condition. Morphological studies showed anomalies in structures related to this kind of reproduction, like variability in number and absence of iridescence in spermathecae, reduction or absence of male funnels, and male pores in different positions. All these variations were found in other studies [12,28,54,56,57]. Some of these works found the presence of spermatothecae (structures associated with biparental reproduction), but they were absent from our samples.

Male pore position was used as a taxonomical character separating different subspecies [12,56,57]. According to our results, subspecies division based in male pore position does not appear to have a phylogenetic base, as no correlation between this character and the relationships recovered by the three molecular markers studied was found.

Further studies will shed light on the complete phylogeography of this species in whole Spain. Therefore, we want to realize one “*Eiseniella tetraedra project*” with as much populations all over the world as we can, trying to understand the distribution of this cosmopolitan species (found even in Australia [58]). We invite anyone interested in participate in this project.

4. Conclusions

Some of the initial questions can be answered. The high genetic variability and population structure were variable between populations. *Eiseniella tetraedra* showed a great genetic variability in the two studied areas in Spain with no clear patterns of population structure as a result of two opposite trends (monohaplotypic and polihaplotypic populations in both areas). There were some genetic differences between populations of *E. tetraedra* from the Eurosiberian and Mediterranean studied regions. We found six deeply divergent lineages, all of them represented in both zones of Spain except lineage C, which was only found in Northwestern Spain and Wales, and lineage B, only present in central area of Spain and worldwide localities. The samples from the rest of the world showed limited variability. The studied samples from the rest of the world showed limited variability, which can be biased by the number of samples. We suggested a recolonization from Spain to Europe even though other possible refugia such as Italian and Balcanic Peninsulas should be explored with further sampling.

An apparent dispersion pattern was detected along the Guadarrama river basin, with lineages upstream and middle/low stream being different.

Finally, actual subspecies subdivision based on position of male pores didn’t show correlation with genetic data.

Acknowledgements

This research was funded by project CGL2013–42908–P from the Spanish Government. We would like to thank John Morgan for a delightful sampling day in Wales. IS was hired by the project. MN was supported by a Postdoctoral Fellowship, FDPI-2016–16407, from the Spanish Government. DFM was supported by a Predoctoral Fellowship grant by Universidad Complutense de Madrid, Spain.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.ejsobi.2016.11.006.

References

