High genetic divergence of hormogastrid earthworms (Annelida, Oligochaeta) in the central Iberian Peninsula: evolutionary and demographic implications

Marta Novo, Ana Almodóvar & Darío J. Díaz-Cosín

Submitted: 24 September 2008
Accepted: 27 January 2009
doi:10.1111/j.1463-6409.2009.00389.x

Novo, M., Almodóvar, A. & Díaz-Cosín, D. J. (2009). High genetic divergence of hormogastrid earthworms (Annelida, Oligochaeta) in the central Iberian Peninsula: evolutionary and demographic implications. — *Zoologica Scripta*, *38*, 537–552.

Hormogastridae earthworms are highly important for the functioning of the Mediterranean soil system. However, little is known about the species distribution and genetic diversity of these soil invertebrates. In the present study, the genetic differentiation and gene flow were studied among populations of hormogastrids from the central Iberian Peninsula. A 648-bp portion of the mitochondrial cytochrome c oxidase I gene was sequenced for 82 individuals from 7 localities, resulting in the identification of 38 haplotypes exclusive to localities. All of the individuals were morphologically identified as *Hormogaster elisae*, but the high genetic divergence found among populations (up to 20.20%) suggests the occurrence of more than one cryptic species within this region. Further analysis of the phylogenetic relationships revealed six different evolutionary lineages coincident with geographical location, including the two nearest populations Molar and Redueña as one evolutionary unit. From these results, at least three new species could be inferred, in addition to the morphospecies *H. elisae s.s.* Partitioning of genetic variance among populations indicated that isolation by distance was the primary agent for differentiation of the investigated hormogastrid populations. Our data suggest that the evolutionary lineages for *H. elisae s.l.* originated between the late Miocene and the early Pleistocene, but that mtDNA genealogies coalesce on a more recent scale of a few thousand years.

Corresponding author: *Marta Novo, Departamento de Zoología y Antropología Física, Facultad de Biología, Universidad Complutense de Madrid, José Antonio Nováis, 2. 28040, Madrid, Spain. E-mail: mnovo@bio.ucm.es*

Ana Almodóvar and Darío J. Díaz-Cosín, Departamento de Zoología y Antropología Física, Facultad de Biología, Universidad Complutense de Madrid, José Antonio Nováis, 2. 28040, Madrid, Spain. E-mails: aalmodovar@bio.ucm.es, dadico@bio.ucm.es

Introduction

Earthworms are a key group for the correct functioning of soil systems, since these organisms are the primary contributors in terms of soil animal biomass. Thus, they have a crucial role in soil biogeochemical cycles, with direct and indirect effects on other groups of edaphic organisms (Edwards & Bohlen 1996).

The family Hormogastridae, Michaelsen 1900 consists of four genera and 23 species of large and middle-sized earthworms, which are distributed in the western Palearctic Region (Omodeo 1998); limited almost exclusively to the countries and islands of the western Mediterranean, such as Spain, France, Italy, the North African countries, Sicily, Corsica, and Sardinia (Díaz-Cosín *et al.* 1989; Cobolli-Sbordoni *et al.* 1992). The genus *Hormogaster*, Rosa 1887 includes 19 of the 23 species with similar distribution. The Iberian Peninsula contains the highest diversity of this genus, with 17 species located in central and eastern Spain. Only one species, *Hormogaster elisae*, has been found to be endemic to central Spain. *Hormogaster elisae* was described by Álvarez (1977), with individuals first discovered in Siguero (Segovia, Spain) and subsequently found in other localities of Segovia and Madrid (Spain), (Moreno 1981; Garvín 1995; Garvín *et al.* 2002; Jesús *et al.* 2002). *Hormogaster elisae* can be much abundant in sandy soils that are poor in organic matter and prone to strong erosion and summer aridity. These soils are potentially unsuitable for most earthworm species and, therefore, the organisms cannot thrive in these environments (Hernández 2005). However, when conditions are unfavourable, due to a lack of rain and dry soil, hormogastrid earthworms can undergo paradiapause (Díaz-Cosín *et al.* 2006), and build aestivation chambers to inhabit until conditions turn favourable again. Climate change will presumably favour the expansion of

these kinds of dry, sandy soils, potentially increasing the ecological importance of hormogastrid earthworms, particularly *H. elisae*. In turn, more detailed information on the species distribution, genetic diversity, and population differentiation of this earthworm should be of great value as a basis for the analysis of future soil ecology.

To date, little is known about the phylogeny of the family Hormogastridae, other than morphological species descriptions and one allozyme study (Cobolli-Sbordoni *et al.* 1992), which analysed genetic divergence, palaeogeography, and speciation of some members of Hormogastridae. Populations from the central area of the Iberian Peninsula were not included in that study; therefore, a genetic description of these earthworm populations is presented for the first time in the present work.

Our group is currently working on the systematics and phylogeny of Hormogastridae. When studying specimens from the central area of Iberian Peninsula, individuals from every population could be clearly identified through morphological analysis as *H. elisae*. However these earthworms presented variability in some characters, such as length, diameter, and number of segments. Despite the constancy of the usual taxonomic characters studied in these earthworms, this variability raised uncertainty about the presence of some undescribed heterogeneity and prompted this investigation of the underlying genetic variability.

The objectives of this study were to determine the levels of genetic differentiation and estimate the amount of gene flow occurring between populations of hormogastrid earthworms in the central Iberian Peninsula, as well as study the possible phylogenetic relationships among these populations. Finally, we provide some information on the evolutionary and demographic histories of populations from each clade.

For this purpose, the gene sequence of the first subunit of the cytochrome c oxidase mitochondrial gene (COI) was analysed. This COI segment has been successfully used for the study of intra-specific and intra-generic relationships within annelids (Black *et al.* 1997; Bely & Wray 2004; Hurtado *et al.* 2004; Jolly *et al.* 2005; Oceguera-Figueroa *et al.* 2005; Halanych & Janosik 2006), specifically within earthworms (Pop *et al.* 2003; Chang & Chen 2005; Csuzdi *et al.* 2005; Pérez-Losada *et al.* 2005; Pop *et al.* 2005; Huang *et al.* 2007), due to a high rate of mutation, particularly in third codon positions (Halanych & Janosik 2006). Moreover, studies on various groups of animals have shown that a 650-bp fragment of this gene is generally effective as a barcode sequence, providing more than 95% species-level resolution (Hebert *et al.* 2003, 2004a, 2004b; Barrett & Hebert 2005).

Materials and methods

Sample collection and morphological study

To sample the area, 82 sexually mature clitellate individuals of hormogastrid earthworms were collected from seven localities of the central Iberian Peninsula. This area can be divided into two main geographical units: (i) Guadarrama Mountains and, (ii) Tajo Fluvial Valley (Fig. 1, Table 1). After being washed with distilled water, the individuals were preserved in absolute ethanol at –20 °C. The specimens were dissected and morphologically studied. A portion of the tegument (0.025 g) was collected from each individual and carefully cleaned under the stereomicroscope to remove soil particles and macroscopic parasites. Subsequently, tegument samples were hydrated and preserved at –80 °C until DNA extraction.

Every character contributing to Hormogastridae systematics, as described by Qiu & Bouché (1998), was studied in each individual from these seven populations (including the typical locality of the species *H. elisae*; Siguero).

DNA amplification

Total DNA was extracted from the tegument tissue samples using the DNeasy Tissue Kit (QIAGEN, IZASA, Barcelona, Spain). A segment of the COI gene was amplified using the primers LCO1490 (5′-GGTCAACAAATCATAAAGA TATTGG-3′) and HCO2198 (5′-TAAACTTCAGGGT GACCAAAAAATCA-3′) (Folmer *et al.* 1994). Polymerase chain reaction (PCR, Saiki *et al.* 1988) was performed using a Perkin Elmer 9700 thermal cycler for each individual sample. Reaction mixture contained the following: 5 μL of 10×

Table 1 *Hormogaster* spp. sampling localities in the central Iberian Peninsula. The abbreviations used in the data analysis are shown.

Locality	Abbreviation	*N*	Location*	Geographical unit	Altitude (m)	Body length (cm)
Anchuelo	ANC	9	N40°28′50.2" W3°14′33.5"	V (Henares)	780	15.45 (0.96)
Lozoyuela	LOZ	12	N40°56′51.9" W3°37′16.2"	M (upper South)	1036	7.95 (0.65)
Molar	MOL	12	N40°44′22.9" W3°33′53.1"	M (middle South)	753	14.72 (1.05)
Pardo	PAR	12	N40°31′11.0" W3°47′42.7"	V (Manzanares)	662	20.82 (1.17)
Redueña	RED	14	N40°48′46.7" W3° 36′06.2"	M (middle South)	797	12.06 (0.68)
Sevilla N.	SEV	12	N40°20′41.9" W4°00′48.9"	V (Alberche)	644	12.98 (0.96)
Siguero	SIG	11	N41°11′06.1" W3° 37′07.4"	M (upper North)	1073	22.7 (0.89)

N, number of collected individuals; V, Tajo Fluvial Valley (the tributary is indicated in parenthesis); M, Guadarrama Mountains (the exact location is indicated in parenthesis). Values in parentheses are variances of the estimates.

*The global location was calculated using a handheld global positioning system (GPS).

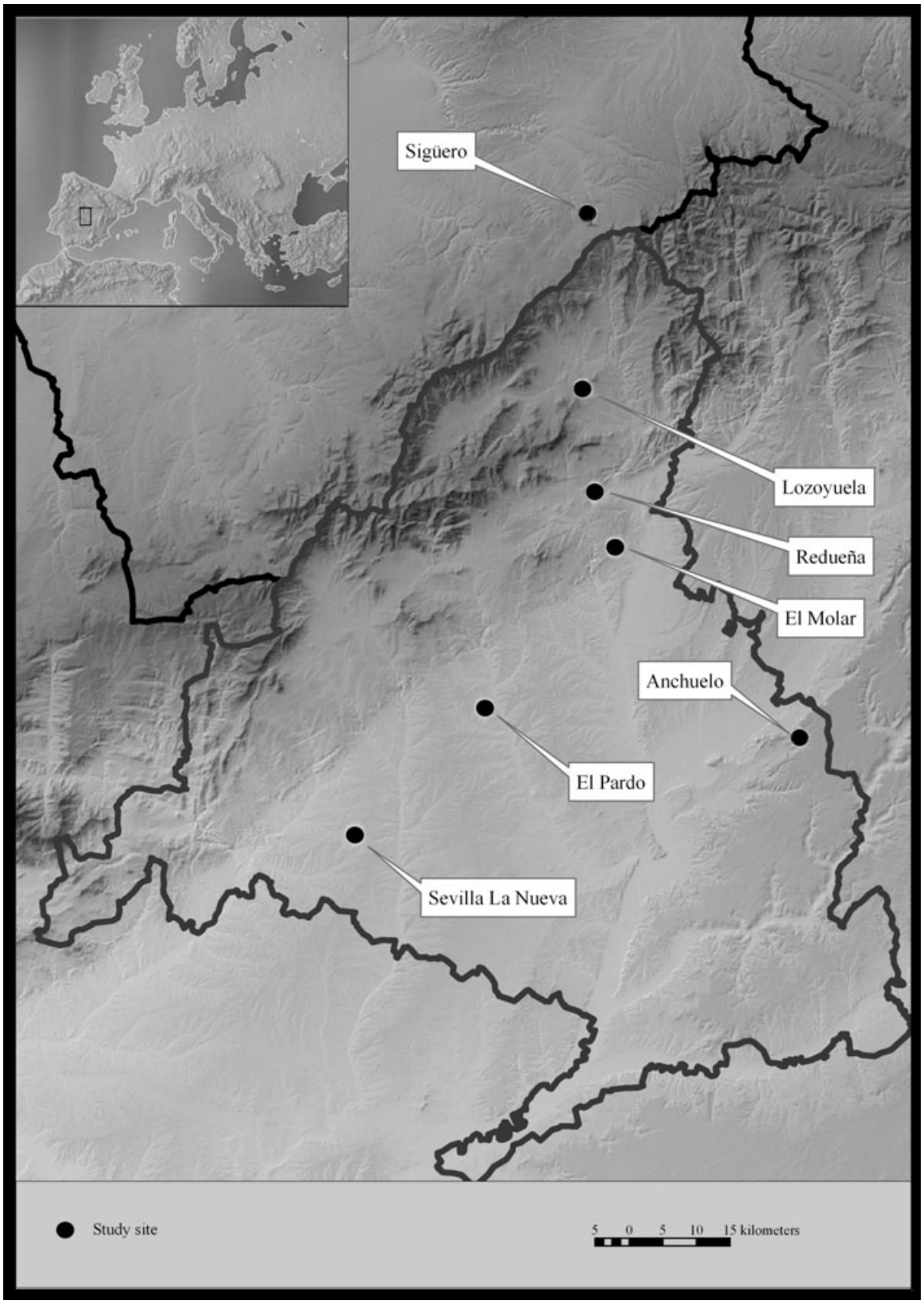

Fig. 1 Map showing study sites for Hormogastridae in central Iberian Peninsula.

Mg-free PCR buffer (Biotools), 3.5 μL of 50 mM $MgCl_2$, 2 μL of 10 mM dNTPs, 1 μL of each 20 μM primer, 1.25 U of Taq DNA polymerase, 1 μL of genomic DNA template, and sterile H_2O to a final volume of 50 μL. In some cases, 6.25 μL of 1 μg/μL BSA (bovine serum albumin) was added to reactions, when initial PCR reactions were unsuccessful. Humic acids found in soil are known PCR inhibitors (Tsai & Olson 1992), and earthworm samples containing these soil components can be problematic for PCR. In such cases, the use of such additives is preferable to long and expensive DNA purifying methods (Kreader 1996). Among those additives, BSA proved useful in preventing PCR inhibition across a wide range of conditions (Höss *et al.* 1992; Akane *et al.* 1993; Höss & Pääbo 1993; Romanowski *et al.* 1993). The PCR profile used in this study was 96 °C (3 min), 40 cycles of [95 °C (30 s), 50 °C (45 s), 72 °C (1 min)], with a final extension of 5 min at 72 °C. PCR products were resolved by 1% agarose gel electrophoresis, visualized by ethidium bromide fluorescence and purified using the BIOCLEAN DNA Purification Columns kit (Biotools, Madrid, Spain). Sequencing was performed by the Genomic Unit, Scientific Park of Madrid (Spain). Sequences were then compared with known earthworm sequences in GenBank using the BLAST search algorithm (Altschul *et al.* 1997), and aligned with CLUSTALX (Thompson *et al.* 1997).

Data analysis

Inter- and intra-population variation. Estimates of variability were computed with ARLEQUIN version 2.000 (Schneider *et al.* 2000). We expressed variation within populations as haplotype diversity (*H*), nucleotide diversity (π), and number of segregating sites (*S*). Mean genetic differentiation among and within populations was estimated using Kimura's (1980) 2-parameter model of sequence evolution. Pairwise Φ_{ST} was calculated, taking into account the variation in haplotype frequencies among populations and genetic distance based on nucleotide variation.

Distribution of genetic variance within a hierarchical structure of population organization was expressed as Φ statistics in a nested analysis of molecular variance (AMOVA; Weir & Cockerham 1984; Excoffier *et al.* 1992). Statistical confidence in variance estimates was determined by comparing the observed Φ statistics against a distribution of estimates generated from 1000 permutations of data (Excoffier *et al.* 1992).

A pattern of isolation by distance was tested with a Mantel (1967) test, which correlates the matrix of genetic distance between localities (Kimura 2-parameter for this study) and geographical distance (straight line between study sites for this study). The test was performed in the program IBD (Jensen *et al.* 2005) and the significance of matrices correlation was evaluated comparing the Mantel test statistic Z, for which random distributions were obtained with 10 000 permutations.

A haplotype network was constructed using the statistical parsimony procedure (Templeton *et al.* 1992; Crandall *et al.* 1994) and a 200-step connection limit with TCS version 1.12 (Clement *et al.* 2000). This method estimates the unrooted tree and provides a 95% plausible set for all sequence type linkages within the unrooted tree. The reconstruction of this type of network has been recommended for intra-specific phylogenetic studies (Posada & Crandall 2001), due to the production of reticulate relationships at the population level which are not represented by traditional phylogenetic methods based on a bifurcating tree. For those phylogenetic trees, all haplotypes are located in tips or terminal nodes, assuming that ancestral haplotypes are extinct. However, taking into account coalescence theory, ancestral haplotypes may be the most frequent in the population and will tend to have a lot of mutational connections with younger haplotypes (Watterson & Guess 1977; Crandall & Templeton 1993).

Phylogenetic relationships. Neighbor-Joining (NJ), Maximum Parsimony (MP), Maximum-Likelihood (ML), and Bayesian analysis (BI) were used for phylogenetic reconstruction, including two different species within the genus *Hormogaster*: *H. riojana* Qiu & Bouché (1998) (from Alesanco, La Rioja, North of Spain) and *H. pretiosa* Michaelsen (1899) (from Cervera del Maestre, Castellón de la Plana, East of Spain). Sequences for these analyses were obtained in our laboratory following the above protocol.

NJ, MP, and ML were performed using the computer programme PAUP 4.0b3 (Swofford 2002). Likelihood ratio tests performed with MODELTEST 3.0 (Posada & Crandall 1998) indicated model Hasegawa–Kishino–Yano (Hasegawa *et al.* 1985), with a gamma distribution and invariable sites (HKY + I + G) as the best-fitted model for the data. This model was used in the phylogeny estimation for the NJ, ML and BI approaches. Both ML and MP included heuristic searches involving tree bisection and reconnection (TBR) branch swapping with 10 random stepwise additions of taxa. Bootstrap analysis (Felsenstein 1985) was used to estimate support for the resulting topologies, with 1000 replicates used for NJ and MP and 100 replicates used for ML analysis. Bayesian phylogeny estimation was performed using the program MRBAYES v.3.1.2 (Ronquist & Huelsenbeck 2003). Clade support was estimated utilizing a Markov chain Monte Carlo (MCMC) algorithm. Parameters in MRBAYES were set to four million generations and 8000 trees were sampled for every 500th generation, using the default random tree option to initiate the analysis. The analysis was performed twice and all sample points prior to the plateau phase were discarded as 'burn in'. The remaining trees were combined to find the maximum *a posteriori* probability estimate of phylogeny.

Nucleotide sequences were translated into amino acids with DnaSP v.4.0 (Rozas *et al.* 2003), using the *Lumbricus terrestris* mitochondrial genetic code (Boore & Brown 1995). Phylogenetic

analyses were subsequently performed with this amino acid-based data set.

Historic demography. Population genetic parameters were determined from the genetic data to infer demographic history of the populations. Historic demographic expansions were detected by the examination of frequency distributions of pairwise differences in sequences (mismatch distributions) within populations (Slatkin & Hudson 1991; Rogers & Harpending 1992). Concordance of our data with the distribution underlying the sudden-expansion model of Rogers (1995) was assessed by means of a least squares approach (Schneider & Excoffier 1999) implemented with ARLEQUIN. For distributions that did not differ significantly ($P > 0.05$) from the expectations of the sudden-expansion model, τ (an estimate of the mode of the mismatch distribution) was estimated, which is an index of time since expansion expressed in units of mutational time (Slatkin & Hudson 1991). Confidence intervals for τ were estimated from 10 000 bootstrap replicates.

The relationship $\tau = 2ut$ was used to approximate time of expansion in generations (t) for *H. elisae* populations, where u is the mutation rate per sequence, per generation, and t is the time since expansion. The value of u was calculated from $u = 2\mu k$, where μ is the mutation rate per nucleotide and k is the number of nucleotides of the analysed fragment. Mutation rates for earthworm COI have been reported to be 3.5% per million years, using sequences from *Metaphire yuhsi* (Chang & Chen 2005), and considering the known geological events that could have resulted in the isolation of clades. *Hormogaster elisae* requires approximately 1 year to attain full maturity (clitellum) (Díaz-Cosín *et al.*, in press), and thus, this was the generation time used for calculations. Given the fact that mismatch distributions have been found to be very conservative (Ramos-Onsins & Rozas 2002), and in order to have a wider view of the evolutionary scenario, deviations from a neutral Wright–Fisher model were also tested for consistency, with population expansion by means of other statistics, such as Fu's *FS*, Fu and Li's *F**, Fu and Li's *D**, and Tajima's *D*, calculated using DnaSP 4.10 (Rozas *et al.* 2003).

Results

Morphological features

All of the 82 hormogastrid specimens collected in central area of the Iberian Peninsula clearly belong to the species *H. elisae*, since these earthworms present constancy in a set of characters that strongly separate this species from the remaining species of the genus *Hormogaster*. These characters are clitellum position in (13)14(15)–26(27)28, tubercula pubertatis position in 22(23)–25(26), the presence of two spermathecae pairs in 9 and 10, with the pair in 10 being much larger, and the existence of a typhlosole with 5 lamellae within a minimum of 30 or 40 segments. Nevertheless, significant differences were found between the specimens, when length was analysed (ANOVA, $F_{6.46} = 38.096$, $P < 0.001$). The largest specimens of *H. elisae* were found in Siguero ($x = 22.7$ cm), and the shortest in Lozoyuela ($x = 7.95$ cm) (Table 1). *Post hoc* comparisons demonstrated that three size groups could be established: (i) Siguero, Pardo, (ii) Redueña, Sevilla la Nueva, Molar, Anchuelo, and (iii) Lozoyuela.

In spite of these differences, this morphological study leads to the conclusion that all specimens from the seven different populations correspond to the current morphospecies *H. elisae*, because these organisms exhibited constant key characters used in earthworm taxonomy.

Sequence variation

A 648-bp fragment of the COI gene, corresponding to the positions 65–712 of *L. terrestris* (Boore & Brown 1995), was amplified. No deletion or insertion of bases was observed. A total of 211 polymorphic sites and 295 mutations were detected, resulting in 200 informative sites for parsimony analysis. Furthermore, 34 of the polymorphic sites (16.11% of variable positions) were at the first codon position, 2 (0.95%) were at the second, and 175 (82.94%) occurred at the third codon position. The average A + T content of the *H. elisae* COI gene was 59.8%.

A total of 38 haplotypes were found among the 82 individuals of *H. elisae s.l.* collected from seven populations along the central area of the Iberian Peninsula, indicating a high degree of polymorphism. All the sequences have been deposited in the GenBank database, with the accession numbers EF653868–EF653905 for the current morphospecies *H. elisae* (Appendix I), and EF653906 and EF653907 for *H. riojana* and *H. pretiosa*, respectively. The complete information on haplotype frequencies for each sampling site is shown in Appendix I, and the values of haplotype and nucleotide diversity for each locality are shown in Table 2. All the haplotypes were exclusive to the locality of collection, that is, there were no shared haplotypes among localities. The number of haplotypes per location ranged from three to seven (Table 2). The global values of haplotype and nucleotide diversity were $H = 0.97$ and $\pi = 0.131$, respectively.

Inter- and intra-population variation

Genetic divergence values are shown in Table 3. Divergence values with *H. riojana* and *H. pretiosa* are included to provide a reference point.

Divergence values of nominal *H. elisae* among localities are generally high, oscillating between 5.75% (between Molar and Redueña) and 20.2% (between Molar and Sevilla la Nueva). Notably, genetic divergence values between some localities are similar or even higher than values observed between morphologically defined species. Thus, putative *H. elisae* from the Pardo population and *H. riojana* have a divergence value

Table 2 Measures of intra-population variability for morphoespecies *H. elisae* in the central Iberian Peninsula based on the COI gene. Values in parentheses are variances of the estimates. K2P genetic divergence is expressed in the percentage of changes.

Locality	*N*	Number of haplotypes	Haplotype diversity (*H*)	Nucleotide diversity (π)	Segregation sites (*S*)	Ts	Tv	Mean K2P genetic divergence
Anchuelo	9	6	0.92 (0.07)	0.003 (0.002)	5	4	1	0.27
Lozoyuela	12	5	0.83 (0.07)	0.016 (0.009)	23	19	4	1.58
Molar	12	7	0.91 (0.06)	0.010 (0.006)	21	17	4	1.03
Redueña	14	5	0.70 (0.10)	0.038 (0.020)	49	43	6	3.84
Pardo	12	5	0.76 (0.09)	0.002 (0.002)	5	5	0	0.25
Sevilla N.	12	3	0.67 (0.09)	0.002 (0.001)	3	2	1	0.17
Siguero	11	7	0.82 (0.12)	0.009 (0.005)	15	15	0	0.92

N, number of individuals per sampling site; Ts, number of transitions; Tv, number of transversions.

Table 3 Measures of population differentiation for nominal *H. elisae* in the central Iberian Peninsula obtained from COI data. Values above the diagonal are pairwise Φ_{ST}, and those below the diagonal indicate mean K2P genetic divergence (in percentage of changes) between pairs of populations. Reference species are included for genetic divergence.

	Anchuelo	Molar	Redueña	Lozoyuela	Siguero	Pardo	Sevilla N.	*H. riojana*
Anchuelo		0.96	0.86	0.94	0.96	0.99	0.99	
Molar	16.9		0.69	0.91	0.94	0.97	0.97	
Redueña	15.0	5.7		0.80	0.83	0.89	0.89	
Lozoyuela	17.3	13.1	10.9		0.93	0.96	0.96	
Siguero	15.2	15.2	12.9	17.3		0.97	0.97	
Pardo	18.3	19.0	17.9	19.7	18.0		0.98	
Sevilla N.	17.1	20.2	16.7	18.5	18.3	11.4		
H.riojana	19.7	20.9	19.8	22.1	20.9	17.4	18.8	
H.pretiosa	21.9	22.1	20.9	21.5	23.2	21.0	23.2	15.8

Table 4 AMOVA results for morphospecies *H. elisae* populations based on COI gene.

Source of variation	d.f.	Percentage of variation	Φ_{ST}
Among populations	6	92.59	0.926 ($P < 0.0001$)
Within populations	75	7.41	

d.f., degrees of freedom.

of 17.45%, and the genetic divergence between *H. riojana* and *H. pretiosa* is 15.83%, whereas Sevilla la Nueva and El Pardo have higher values of genetic divergence in comparison to the other populations. Genetic variability within populations was observed (Table 2), mainly in Redueña, with a high divergence among haplotypes (3.84%).

AMOVA results demonstrated that most (92.59 %) of the genetic variation is due to differences among populations (Table 4). The Φ_{ST} value was high, indicating significant population genetic structure. All pairwise comparisons of Φ_{ST} among populations were significant (even after Bonferroni correction; critical significance level of $\alpha = 0.0024$, Table 3) and significantly high, which indicates a lack of gene flow among the populations.

Mantel test revealed the presence of a positive correlation between genetic and geographical distances among populations ($r = 0.5328$, $P = 0.0219$, $r^2 = 0.284$), which was higher when logarithmic transformation of geographical distance was included in the analysis ($r = 0.6969$, $P = 0.0086$, $r^2 = 0.486$).

The haplotype network obtained with TCS is shown in Fig. 2. The quantity of mutational steps allowed by a statistical parsimony level of 90% was 16 and based on this analysis, each population formed an exclusive network, which was not connected to other population networks. In order to visualize connections between populations, a connection limit of 200 steps was defined for the analysis, since non-parsimonious connections can also be considered (Templeton *et al.* 1992). Thus, the number of mutational steps is shown in Fig. 2 between those haplotypes that differ by more than 16 positions. The population of Redueña has one frequent haplotype (RED5) that presents a high degree of difference (41 mutational steps) from the rest of haplotypes found in that locality. This difference is of similar magnitude to the difference among haplotypes from different localities (43 steps between RED5 and haplotypes of Molar). Except for Redueña, the remainder populations present haplotypes in parsimonious connected networks. Mountain localities present more dispersed networks

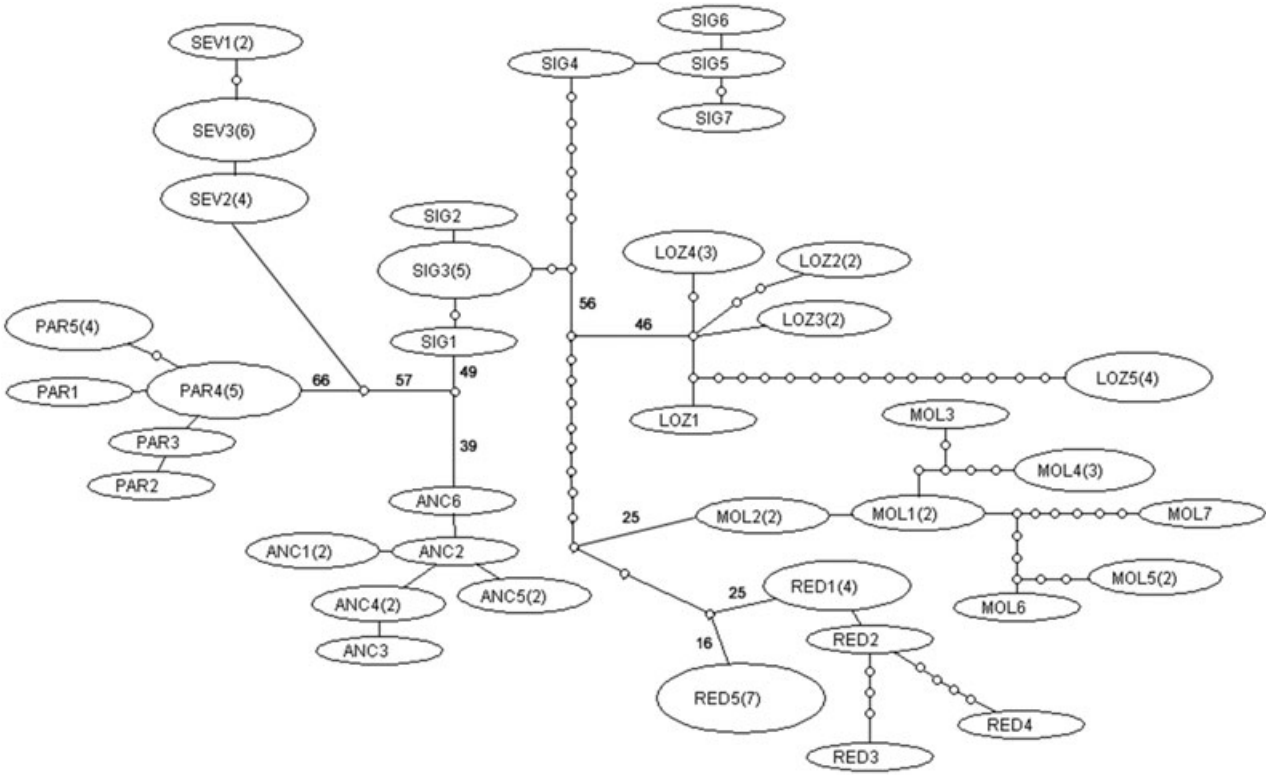

Fig. 2 Reconstructed TCS network of the 38 COI haplotypes of morphospecies *H. elisae* in the studied locations. Connection limit: 200 steps. Each haplotype is named with the abbreviation of its locality (see Table 1) followed by its number. Oval size depends on haplotype frequency, which is shown in brakets in those haplotypes that include more than one individual. Empty circles represent intermediate haplotypes (unsampled). Each branch represents one mutational step except for the cases in which the number of steps exceeds 16 (parsimony level of 90%). In those, the number of steps is shown. Branch length is meaningless.

than valley localities, which is a signature of different values of nucleotide diversity that are higher in mountainous sample sites.

Phylogenetic relationships

Four methods of phylogenetic inference were performed: Bayesian inference (BI), Maximum Parsimony (MP), Maximum Likelihood (ML), and Neighbor-Joining (NJ). All generated trees produced congruent topologies. The different populations and phylogenetic relationships were generally well-supported by high bootstrap and posterior probability values, whereas some haplotype relationships within populations differed among analyses. The ML tree (log likelihood: –3496.43935) is shown in Fig. 3, and bootstrap and posterior probability values obtained in coincident associations of MP, NJ, and BI are indicated. According to all of the phylogenetic analyses, populations of the morphospecies *H. elisae* in the central area of the Iberian Peninsula presented two big groups, with populations of Pardo and Sevilla la Nueva clearly separated from the other populations. Subclades can be then identified: (i) Molar and Redueña, (ii) Lozoyuela, (iii) Siguero, (iv) Anchuelo, (v) Pardo, and (vi) Sevilla la Nueva, most with bootstrap values of 100% and posterior probability values of 1.00. Within subclade 1, there were differences in the topology of the ML and BI trees when compared to the NJ and MP trees. The ML and BI trees demonstrated that haplotypes of Redueña formed a paraphyletic group, with haplotype RED5 outside of the population set, whereas the NJ and MP trees presented Redueña as a monophyletic population. Nevertheless, the Molar population was monophyletic in all cases, exhibiting 100% of bootstrap support and 1.00 of posterior probability. Thus, the phylogenetic analyses revealed six highly supported monophyletic groups, congruent with the geographical location of haplotypes, presenting Molar and Redueña (with a geographical distance of only 8.5 km) as one evolutionary unit.

Translated sequences contained 215 amino acids of 645 included nucleotide positions. Most of the substitutions observed were synonymous, and thus, the amino acid sequences of all specimens exhibited few differences. There were only nine positions with different amino acids; three were informative and the remainder were singletons. Phylogenetic trees obtained

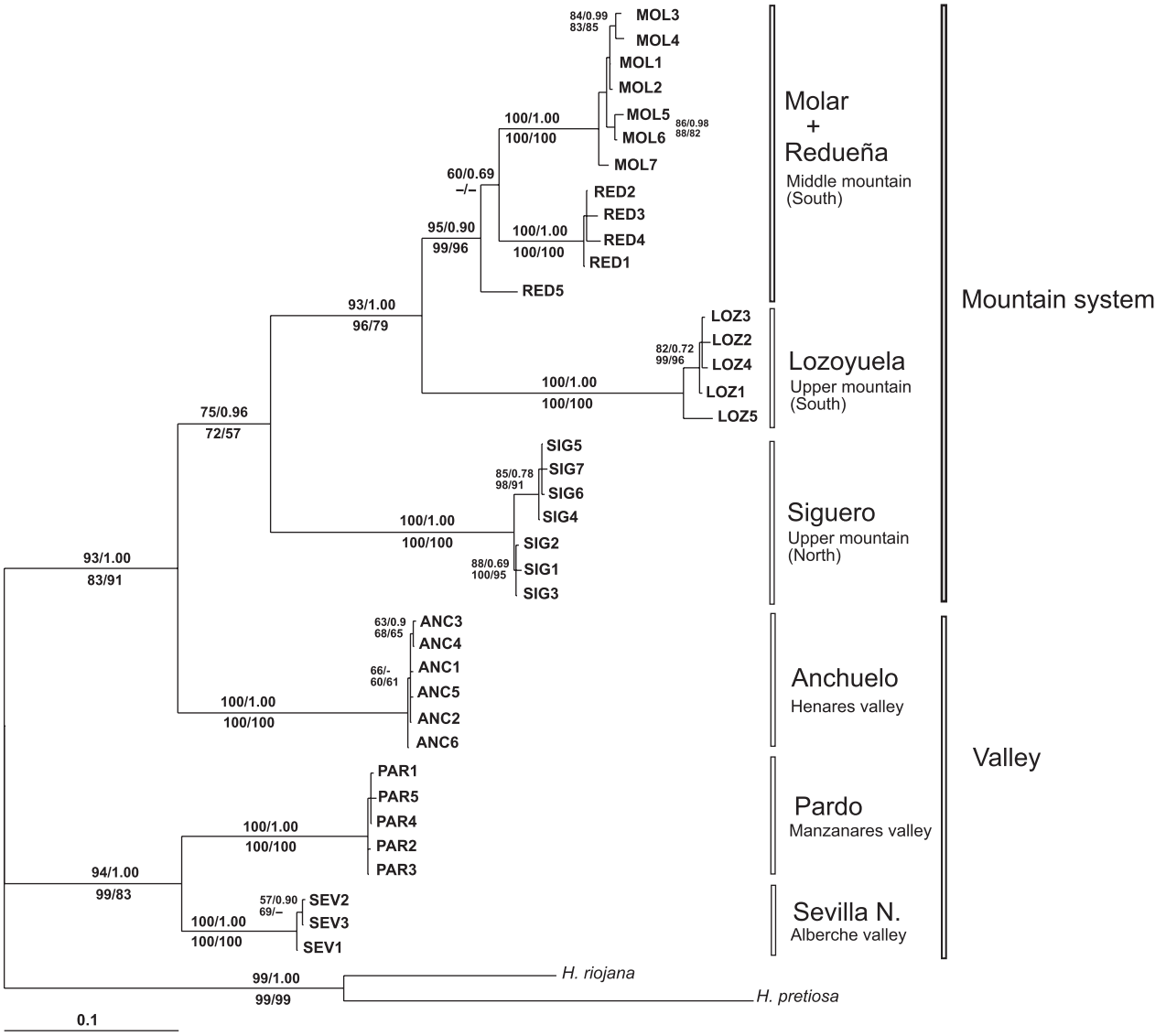

Fig. 3 Maximum likelihood (ML) tree (log likelihood –3496.43935, HKY + I + G model of sequence evolution) based on the analysis of a 648-bp fragment of COI gene in morphospecies *H. elisae*. Bootstrap values (100 pseudoreplicates) are shown above branches, when they are > 50%. Posterior probability values of coincident associations given by Bayesian analysis (BI) are shown also above branches when they are > 0.50 (ML/BI). Neighbor-Joining (NJ) and Maximum Parsimony (MP) bootstrap support values (1000 pseudoreplicates) are shown below branches (NJ/MP). Each haplotype is named with the abbreviation of its locality (see Table 1) followed by its number.

with amino acid sequences had poor bootstrap support values, but yielded some interesting information. Haplotypes from Molar, Redueña, and Lozoyuela were clustered together and mixed, presenting similar amino acid sequences, whereas the remaining clades are similar to those obtained by DNA-based trees (not shown).

Historic demography

Mismatch distributions were tested for each of the six evolutionary units generated by phylogenetic analysis. Distributions for Anchuelo, Pardo, Sevilla la Nueva, and Siguero did not differ significantly ($P > 0.05$) from expectations under the sudden-expansion model and, therefore, were suitable for analysis of demographic patterns. Distributions for the Molar-Redueña and Lozoyuela clades differed significantly from the sudden-expansion model ($P < 0.05$). These two evolutionary units exhibited a bimodal distribution, which was consistent with population size constancy (Fig. 4). In addition, the value of the four statistical tests (Fu's *FS*, Fu and Li's *F**, Fu and Li's *D**, and Tajima's *D*) did not deviate

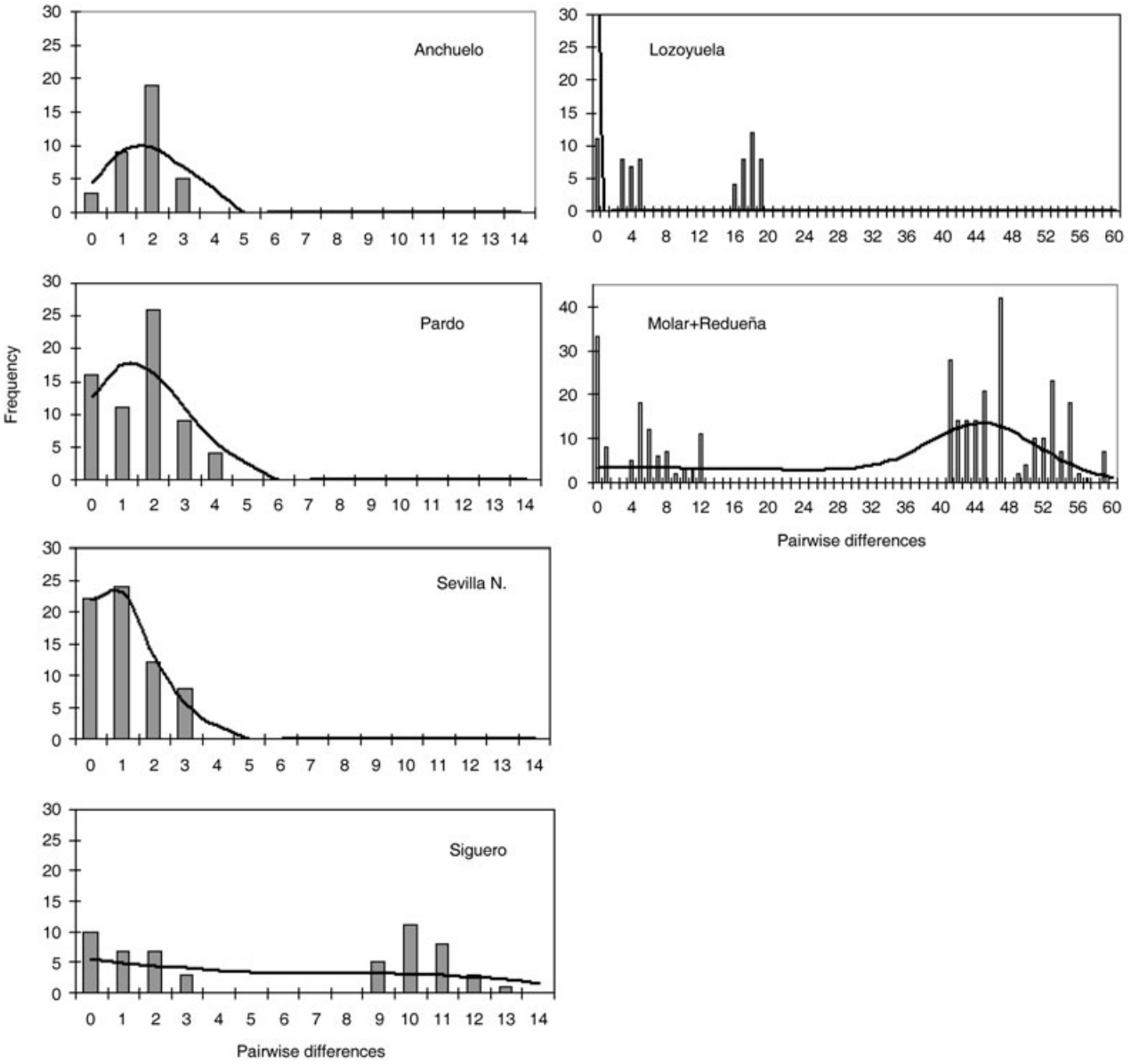

Fig. 4 Mismatch distributions for populations of nominal *H. elisae*, based on 648-bp segment of COI gene. The abcissa represents number of pairwise differences and the ordinate represents the number of observations. The vertical bars are the observed distribution of mismatches and the smoothed line represents the expected distribution under the sudden-expansion model of Rogers (1995) as modified by Schneider and Excoffier (1999).

significantly from the neutral Wright–Fisher model for studied populations. However, signatures of historic bottlenecks were detected for Anchuelo, whose values of Fu's *FS* were negative and significant (–2.58288, $P = 0.015$).

Anchuelo, Pardo and Sevilla la Nueva showed evidence of recent expansions, whereas the Siguero expansion was older (Table 5). Estimates of τ indicated that the relative age of expansion for the Siguero population was seven times older than the expansion of *H. elisae* in the three remaining populations (mean value of $\tau = 1.7$). The 95% confidence intervals of τ overlapped in comparisons among Anchuelo, Pardo and Sevilla la Nueva, which indicated similar times since expansion and showed unimodal mismatch distributions consistent with population growth. The mismatch distribution for Siguero exhibited two distinct waves, which we attributed to temporally distinct expansions (Fig. 4). Our estimates of the time of expansion for Anchuelo, Pardo, and Sevilla la Nueva were between approximately 8200 and 23 700 years ago,

Table 5 Population genetic parameters estimated from COI sequence data by pairwise haplotype mismatch analysis under a sudden-expansion model. Lozoyuela and Molar-Redueña clades are not included, since these clades did not meet the assumptions of the model. Based on the population genetic parameters, timing of population expansion was calculated, assuming one generation per year and a mutation rate of 3.5% per million years.

Parameters	Anchuelo	Pardo	Sevilla N	Siguero
τ	2.076	2.150	0.744	11.793
$\theta_{0(\text{per sequence})}$	0.002	0.000	0.448	0.004
$\theta_{1(\text{per sequence})}$	99 999	7.447	99 999	9.058
P (Sim. SSD ≥ Obs. SSD)*	0.08	0.22	0.91	0.26
Time of expansion (years)	22 900		23 700	8200
130 000				

whereas our estimate for Siguero was approximately 130 000 years ago (Table 5). Thus, expansion of the Siguero clade (located in the mountainous area) occurred earlier in the expansion of the central Iberian Peninsula populations, while expansions in the Anchuelo, Pardo and Sevilla la Nueva clades (located in the valley area) were more recent with similar timings.

Discussion

Sequence variation and morphological features

Applicability of COI sequence data to infer population divergence and genetic structure has been previously proven, due to the high mutation rate of this sequence (Halanych & Janosik 2006) making COI a very useful gene in phylogenetic studies, such as the present one. Most of the observed substitutions in the sequences generated from all *H. elisae* specimens, occurred in the expected third position. Because of the degenerate genetic code, these substitutions were primarily synonymous. The reported A + T content of 59.8% is similar to of *L. terrestris*, as determined by the study of Boore & Brown (1995), who observed an A + T content of 61.6%, which was considered one of the lowest among invertebrate mitochondrial genomes.

Despite the high genetic divergences found among the studied populations as further described herein, a morphological constancy was found that led to the conclusion that all studied individuals belonged to the morphospecies *H. elisae.* These results could indicate discordant patterns of morphological and molecular evolution or a morphological stasis, which could be common in non-visual invertebrates, such as those residing in turbid aquatic systems or soil environments, where chemical signalling may play a more important role than morphology in sexual selection (Lee & Frost 2002).

Inter- and intra-population variation

This study revealed a high degree of heterogeneity in the genetic structure of hormogastrids in the central Iberian Peninsula. Significant genetic subdivision and high genetic diversity was evident and supported by the high and significant pairwise Φ_{ST} between populations (Table 3). Such genetic subdivision could have arisen from restricted gene flow between the populations. This is supported by AMOVA results, which suggest the presence of a defined population structure, demonstrating that most of the found genetic variability is due to differences among localities and a minimum portion is due to variation within populations. Moreover, different populations in the haplotype network are not connected if parsimony limits are considered.

Results have demonstrated that there is a pattern of isolation by distance (IBD, Wright 1943), and thus, more geographically distant populations present higher genetic divergences. At equilibrium, under dispersal and genetic drift, the IBD pattern is revealed by a positive and significant correlation between genetic differentiation and geographical distances (Slatkin 1993; Rousset 1997). We also found this in the present study. This isolation by distance pattern may be the result of historical events. In fact, there is growing evidence to suggest that many species have not yet reached migration-drift equilibrium and that observed patterns of genetic differentiation reflect population history rather than current levels of gene flow (e.g. Latta & Mitton 1999; Pogson *et al.* 2001; Turgeon & Bernatchez 2001). There is no available information regarding the dispersion ability of hormogastrid species. However, endogeic earthworms are expected to have a limited dispersion capacity and the data reveal that the populations are highly fragmented.

The average sequence divergence among studied populations is surprisingly high. Similarly, unusually high values were reported in a study by Field *et al.* (2007), for *L. terrestris* within a relatively small spatial scale. Nevertheless, the ND2 and ND4 genes were sequenced in that study, which evolve faster than COI. Therefore, the obtained average sequence divergence is still very high. To understand this magnitude, the value has been compared among studied populations (so-called *H. elisae*), and between other accepted and recognizable species of the same genus.

Genetic divergence between morphologically defined *H. elisae* and other *Hormogaster* species is lower than the divergence between some of the analysed populations within the focus species. This indicates that there is more than one species within *H. elisae s.l.* Genetic divergence between *H. elisae* from El Pardo and *H. riojana* is 17.45%, and the divergence between *H. riojana* and *H. pretiosa* is lower, at 15.83%. Therefore, it seems natural to think that to identify two different species, their divergence values should be higher than those previously cited. Thus, the existence of different species could be considered in some localities sampled for this study. For instance, new species could be proposed for the populations at El Pardo and Sevilla la Nueva, since these sample populations have a genetic divergence of 16.75% and 20.2% with all other populations, respectively, but there is only 11% divergence between the

two populations. Additionally, the genetic divergence between Siguero and the remainder populations is high (greater than 13%), which also occurs for Anchuelo (greater than 15%), whereas Redueña, Lozoyuela, and Molar could be grouped, since these populations exhibit lower divergence values. According to these divergence values, three new species could be proposed besides the morphoespecies *H. elisae*, originally described in Siguero, which is the type locality. Nevertheless, the divergence value between defined species could be underestimated, as the entire diversity present in *H. riojana* and *H. pretiosa* may not have been represented in the analysis of few individuals. Additionally, *H. elisae* may possibly be subject to higher substitution rates.

In order to establish well-founded conclusions, other genetic divergence values found in similar studies of the same gene in other earthworm genus were compared. Pérez-Losada *et al.* (2005), found genetic divergences from 31% to 56% (corrected values) among different species within the genus *Eisenia*, whereas intra-specific divergences among very distant populations are less than 0.57%, which was determined from analysis of *E. andrei* samples from Brazil, Ireland, and Spain. Only one higher intra-specific divergence value of 18% was determined, and the possible existence of a new, different species in that existing population was suggested, with the additional use of a phylogenetic study.

Chang & Chen (2005) found intra-specific genetic divergences (corrected values) from 0.2% to 10% in *Amynthas yuhsi*, and from 2.9% to 10.3% in *A. formosae* individuals. Both species had been considered synonymous before that study, and separation was proposed based on the determined genetic divergence (12.6%) that was higher than the divergence values among different species of the genus *Metaphire (*ranging from 10.2% to 25.5%).

As an example of other annelid groups, Oceguera-Figueroa *et al.* (2005) found 4.2–11.9% intra-specific genetic divergence in leeches, which were similar values to the inter-specific divergence of 14.2%, suggesting the existence of cryptic species. Bely & Wray (2004) investigated naidid Oligochaeta and estimated intra-specific divergence values from 0.2% to 11%, inter-specific values from 5% to 17%, and inter-genus values from 12% to 21%. These authors indicated that species presenting 11% intra-specific genetic divergence could be different cryptic species.

Estimated values of genetic divergence are variable, depending on the taxa and there is no consensus regarding the establishment of the species limit. Nevertheless, it seems that the authors become suspicious about the existence of a new species when divergence values around 10% or 11% are found. If that limit was considered in the present study, five new species in addition to the original *H. elisae* could be inferred from our findings, as only populations from Redueña and Molar would be classified as a single species.

The molecular clock for COI in earthworms was calibrated by Chang & Chen (2005), using sequences from *Metaphire yuhsi* and considering the known geological events that could have caused the isolation of clades. A divergence rate of 3.5% per million years and base pair was suggested. That is higher than the estimated rate for this gene in other animals (2% in general, Avise 2000; 2.25% in mites, Salomone *et al.* 2002; and 2.3% in arthropods, Brower 1994). However, Chevaldonné *et al.* (2002), estimated a rate of 0.5% divergence per million years and base pair for COI in other annelids (hydrothermal Vestimentifera). The molecular clock rate of 3.5% per million years and base pair was chosen for COI divergence time estimates, because these rates can significantly vary from one animal group to another (Avise 1994), and this rate is the only available reference in earthworms.

Thus, the studied populations may have divergence times within the range of 1.64 million years between Redueña and Molar (Pleistocene), and 5.77 million years between Sevilla la Nueva and Molar (late Miocene, Messinian). The Iberian Peninsula was a refuge during the Quaternary glaciations, which could explain such timings and the high genetic divergence values (Taberlet *et al.* 1998). The East area of the Iberian Peninsula held as a xerophilous bush reserve (Pleguezuelos 1997), which represents the preferred habitat of this earthworm genus. Therefore, most of the *Hormogaster* species in Spain are distributed along the Mediterranean coastline.

Phylogenetic relationships

The phylogenetic hypothesis presented herein support genetic divergence data, as every population appears to be monophyletic, except for Redueña in ML and BI analyses where it is presented as paraphyletic group. Haplotype position within the trees is congruent with geographical location, which confirms the results of the haplotype network, with each population clearly delineated. Redueña paraphyletism is further corroborated by the great difference between one of the haplotypes from this group with the rest. This can also be observed in the network, where the haplotype is not connected when parsimony limits are considered. Redueña haplotypes should be considered with El Molar to be a monophyletic group.

Thus, the area contained six genetically divergent clades that do not overlap in distribution, and each particular location tended to be heavily dominated by a single clade. Populations from the Siguero and Lozoyuela clades are distributed in the northern and southern sides from upper areas of the Guadarrama Mountains, respectively. However, populations from the Molar-Redueña clade dominated the middle zones. The three remaining clades have appeared on different fluvial valleys from the studied region.

Wiens & Penkrot (2002), proposed a method for delimiting species based on phylogenetic trees. They suggested that

haplotypes from one locality that did not cluster (i.e. appear to be mixed with haplotypes from other localities) indicated potential evidence of gene flow, and thus, all the included haplotypes should be classified as the same species. The different haplotypes of nominal *H. elisae* are clustered according to respective localities, with the exception of Redueña and Molar. Therefore, six different species could be defined in the studied area, coincident with the six discovered evolutionary lineages. However, the amino acid tree reveals mixture of sequences from Redueña, Molar, and Lozoyuela, which indicates that at least at that level these three populations could be included just in one group.

It is difficult to take a clear decision about the limit of species within the studied populations of nominal *H. elisae* from our findings. To clarify this, the whole distribution area could be explored in a continuous manner to determine that the found genetic discontinuities truly delimit cryptic species and are not artifacts of discontinuous sampling. This investigation is currently being performed on the Hormogastridae family as a whole.

Size groups do not exactly coincide with the more genetically-related populations. Differences in size could be due to evolution, but could also be related to environmental conditions or a combination of both factors. Thus, investigations of soil characteristics in each locality, as well as of the growth of individuals from different populations in the same soil are necessary and may illuminate these pronounced size differences.

Historic demography

In a qualitative assessment of demographics, Anchuelo, Pardo, and Sevilla la Nueva clades from the valley populations were determined to have high levels of haplotype diversity H and low levels of nucleotide diversity π, a signature of rapid demographic expansion from a small effective population size, retaining new mutations (Avise *et al.* 1984; Watterson 1984), and often related to episodes of climate oscillations. However, the remaining clades from the Guadarrama Mountains have a relatively large H and large π, a characteristic of a stable population (haplotypes that are more dispersed in the network). The clade Molar-Redueña from the southern side of the Guadarrama Mountains middle area, which have few highly divergent haplotypes and exhibited separation of the haplotype RED5 in both the haplotype network as well as ML and BI trees, could exhibit a secondary contact process between previously differentiated allopatric lineages.

The interior clade in the haplotype network, and consequently the most ancient population, seems to be Siguero, which is the type locality of *H. elisae.* The dispersal could occur from one of two Siguero lineages, but from two separated migrant groups, as exhibited by the two-waved Mismatch distribution. One group contained migrants towards mountain populations (Molar-Redueña and Lozoyuela), and a second group dispersed to the other valley populations (El Pardo, Sevilla la Nueva and Anchuelo). In actuality, El Pardo and Sevilla la Nueva are localized to the occidental limit of the distribution area of Hormogastridae, and thus, these populations likely resulted from posterior colonization events. Nevertheless, more extensive sampling of the whole distribution area of Hormogastridae might reveal additional levels of heterogeneity and population genetic structure that will allow a better understanding of the colonization and evolutionary processes of this earthworm.

These clades are thought to have originated in the late Miocene or early Pleistocene, but their mtDNA genealogies coalesce on a more recent scale of around a few thousand years. The estimate of expansion time from the two geographical regions studied (upper mountain and valleys) resulted in different dates. Accordingly, population expansion in fluvial valleys (Anchuelo, Pardo y Sevilla la Nueva) by the late Pleistocene is suggested, prior to the most recent glaciations (18 000 years ago), when the region was already configured and the climate was well-established (Blondel & Aronson 1999).

In addition to the mismatch distribution based in the pairwise sequence differences, there are other statistics based on the mutation (segregating site) frequencies and haplotype distribution, which have been suggested to be appropriate and powerful for detecting population growth events (Ramos-Onsins & Rozas 2002). Clade expansion for the whole area under investigation is not well-supported by these statistics, while only Fu's FS is significant in the Anchuelo clade. Therefore, different timings and intensities of the expansion or other evolutionary processes, such as genetic hitchhiking, could have resulted in these different results.

Conclusion

Hormogastrid earthworms were sampled in the central area of the Iberian Peninsula and every collected specimen corresponded morphologically to the species *H. elisae*. However, mitochondrial data exhibited a very high genetic divergence among studied populations, which indicates that the morphospecies *H. elisae* may be a complex of at least four cryptic species. The first would include Molar, Redueña and Lozoyuela localities, the second would be comprised of Pardo and Sevilla la Nueva, Anchuelo would represent the third, and Siguero (i.e. the locality where *H. elisae* was originally described) would be the fourth. Nevertheless, six different evolutionary lineages were observed, and therefore, further investigations are needed to confirm if these found lineages could represent new species.

These results may indicate that the morphology-based systematics in hormogastrids is unreliable or that a morphological stasis is occurring in these organisms, and therefore, molecular biology is essential to understand the variability and detect cryptic species within this family.

Acknowledgements

We would like to thank Carlos Bayón, Rosa Fernández, Mónica Gutiérrez, Juan B. Jesús, Noa Novo, Juan José Molina, Marta Ramajo, and Dolores Trigo for assisting us during sample collection, Daniel Ayllón and Sheila Leal for technical support, and Louis Bernatchez and Hinrich Schulenburg for valuable comments on previous manuscript versions. M. N. was supported by a FPU grant from the Spanish Government. This research was funded by project CGL2007-60715/BOS from the Spanish Government.

References

Akane, A., Shiono, H., Matsubara, K., Nakamura, H., Hasegawa, M. & Kagawa, M. (1993). Purification of forensic specimens for the polymerase chain reaction (PCR) analysis. *Journal of Forensic Science*, *38*, 691–701.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, *25*, 3389–3402.

Álvarez, J. (1977). *Los oligoquetos terrícolas de la Península Ibérica*. Phd Thesis. Universidad Complutense de Madrid, Madrid Spain.

Avise, J. C. (1994). *Molecular Markers, Natural History and Evolution*. London, UK: Chapman & Hall.

Avise, J. C. (2000). *Phylogeography: the History and Formation of Species*. Cambridge, UK: Harvard University Press.

Avise, J. C., Neigel, J. E. & Arnold, J. (1984). Demographic influences on mitochondrial DNA lineage survivorship in animal populations. *Journal of Molecular Evolution*, *20*, 99–105.

Barrett, R. D. H. & Hebert, P. D. N. (2005). Identifying spiders through DNA barcodes. *Canadian Journal of Zoology*, *83*, 481–491.

Bely, A. E. & Wray, G. A. (2004). Molecular phylogeny of naidid worms (Annelida: Clitellata) based on cyrochrome oxidase I. *Molecular Phylogenetics and Evolution*, *30*, 50–63.

Black, M. B., Halanych, K. M., Maas, P. A. Y., Hoeh, W. R., Hashimoto, J., Desbruyères, D., Lutz, R. A. & Vrijenhoek, C. (1997). Molecular systematics of vestimentiferan tubeworms from hidrothermal vents and cold-water seeps. *Marine Biology*, *130*, 141–149.

Blondel, J. & Aronson, J. (1999). *Biology and Wildlife of the Mediterranean Region*. Oxford: Oxford University Press.

Boore, J. L. & Brown, W. M. (1995). Complete sequence of the mitochondrial DNA of the annelid worm *Lumbricus terrestris*. *Genetics*, *141*, 305–319.

Brower, A. V. Z. (1994). Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings National Academy of Science USA*, *91*, 6491–6495.

Chang, C. H. & Chen, J. H. (2005). Taxonomic status and intraspecific phylogeography of two sibling species of *Metaphire* (Oligochaeta: Megascolecidae) in Taiwan. *Pedobiologia*, *49*, 591–600.

Chevaldonné, P., Jollivet, D., Desbruyères, D., Lutz, R. A. & Vrijenhoek, C. (2002). Sister-species of eastern Pacific hydrotermal vent worms (Ampharetidae, Alvinellidae, Vestimentifera) provide new mitochondrial COI clock calibration. *Les Cahiers de Biologie Marine*, *43*, 367–370.

Clement, M., Posada, D. & Crandall, K. A. (2000). TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, *9*, 1657–1659.

Cobolli-Sbordoni, M., de Matthaeis, E., Alonzi, A. *et al*. (1992). Speciation, genetic divergence and paleogeography in the Hormogastridae. *Soil Biology and Biochemistry*, *24*, 1213–1221.

Crandall, K. A. & Templeton, A. R. (1993). Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics*, *134*, 959–969.

Crandall, K. A., Templeton, A. R. & Sing, C. F. (1994). Intraspecific cladogram estimation: problems and solutions. In Scotland, R. W., Siebert, D. J. & Williams, D. M. (Eds) *Models in Phylogeny Reconstruction* (pp. 273–297). Oxford: Clarendon Press.

Csuzdi, C., Pop, A. A., Pop, V. V., Wink, M. & Zicsi, A. (2005). Revision of the *Dendrobaena alpina* (Rosa, 1884) species group (Oligochaeta, Lumbricidae) by morphological and molecular methods. In V. V. Pop & A. A. Pop (Eds) *Advances in Earthworm Taxonomy II*. Cluj-Napoca: Cluj University Press.

Díaz-Cosín, D. J., Briones, M. J. I. & Trigo, D. (1989). Descripción de una nueva especie de lombriz de tierra, *Xana omodeoi* (Hormogastridae, Oligochaeta) y sus implicaciones en la división de los Hormogastridae. *Revue d'Écologie et Biologie Du Sology*, *26*, 225–231.

Díaz-Cosín, D. J., Ruiz, M. P., Ramajo, M. & Gutiérrez, M. (2006). Is the aestivation of the earthworm *Hormogaster elisae* a paradiapause? *Invertebrate Biology*, *125*, 250–255.

Díaz-Cosín, D. J., Hernández, P., Trigo, D., Fernández, R. & Novo, M. (in press). Algunos aspectos del ciclo biológico del endemismo ibérico, *Hormogaster elisae* Álvarez, 1977 (Oligochaeta, Hormogastridae), en cultivos de laboratorio. *Boletín de la Real Sociedad Española de Historia Natural (Sección Biológica)*.

Edwards, C. A. & Bohlen, P. J. (1996). Biology and Ecology of Earthworms. 3rd edn. London: Chapman & Hall.

Excoffier, L., Smouse, P. E. & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, *131*, 479–491.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, *39*, 783–791.

Field, S. G., Lange, M., Schulenburg, H., Velavan, T. P. & Michiels, N. K. (2007). Genetic diversity and parasite defense in a fragmented urban metapopulation of earthworms. *Animal Conservation*, *10*, 162–175.

Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxydase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, *3*, 294–299.

Garvín, M. H. (1995). Efecto de la lombriz de tierra *Hormogaster elisae* Álvarez, 1977 (Oligochaeta, Hormogastridae) sobre la mineralización del carbono y el nitrógeno del suelo en sistemas cerrados. *Tesina de Licenciatura*. Universidad Complutense de Madrid.

Garvín, M. H., Trigo, D., Hernández, P., Ruíz, M. P. & Díaz-Cosín, D. J. (2002). Interactions of *Hormogaster elisae* (*Oligochaeta*, *Hormagastridae*) with other Earthworms Species from Redueña (Madrid, Spain). *Applied Soil Ecology*, *20*, 163–169.

Halanych, K. M. & Janosik, A. M. (2006). A review of molecular markers used for Annelid phylogenetics. *Integrative and Comparative Biology*, *46*, 533–543.

Hasegawa, M., Kishino, H. & Yano, K. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, *22*, 160–174.

Hebert, P. D. N., Cywinska, A., Ball, S. L. & deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of*

the Royal Society of London Series B, Biological Sciences, *270*, 313–321.

Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H. & Hallwachs, W. (2004a). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences USA*, *101*, 14812–14817.

Hebert, P. D. N., Stoeckle, M. Y., Zemlak, T. S. & Francis, C. M. (2004b). Identification of birds though DNA barcodes. *Public Library of Science Biology*, *2*, 1657–1663.

Hernández, P. (2005). Estudio de la distribución horizontal de *Hormogaster elisae* (Oligochaeta, Hormogastridae) en el Molar y de los factores que la regulan. Phd Thesis, Universidad Complutense de Madrid, Madrid, Spain.

Höss, M. & Pääbo, S. (1993). DNA extraction from Pleistocene bones by a silica-based purification method. *Nucleic Acids Research*, *21*, 3913–3914.

Höss, M., Kohn, M., Pääbo, S., Knauer, F. & Schröder, W. (1992). Excrement analysis by PCR. *Nature*, *359*, 199.

Huang, J., Xu, Q., Jun Sun, Z., Lan Tang, G. & You Su, Z. (2007). Identifying earthworms through DNA barcodes. *Pedobiologia*, *51*, 301–309.

Hurtado, L. A., Lutz, R. A. & Vrijenhoek, R. C. (2004). Distinct patterns of genetic differentiation among annelids of eastern Pacific hydrotermal vents. *Molecular Ecology*, *13*, 2603–2615.

Jensen, J. L., Bohonak, A. J. & Kelley, S. T. (2005). Isolation by distance, web service. *BMC Genetics*, *6*, 13–18. *Available via* http://ibdws.sdsu.edu/.

Jesús, J. B., Fernández, B. & Gutiérrez, M. (2002). Lombrices de tierra de la Comunidad de Madrid (España). II. Géneros *Eisenia*, *Eiseniella*, *Eiseniona*, *Lumbricus*, *Octodrilus*, *Ostolasion*, *Hormogaster*, *Microscolex*, *Eukerria y Ocnerodrilus* (Annelida, Oligochaeta). *Boletín de la Real Sociedad Española de Historia Natural (Sección Biológica)*, *97*(1–4), 61–69.

Jolly, M. T., Jollivet, D., Gentil, F., Thiébaut, E. & Viard, F. (2005). Sharp genetic break between Atlantic and English Channel populations of the polychaete *Pectinaria koreni*, along the North coast of France. *Heredity*, *94*, 23–32.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions that compare studies of nucleotide sequences. *Journal of Molecular Evolution*, *16*, 11–120.

Kreader, C. A. (1996). Relief of amplification inhibition in PCR with Bovine Serum Albumin or T4 Gene 32 Protein. *Applied and Environmental Microbiology*, *62*(3), 1102–1106.

Latta, R. G. & Mitton, J. B. (1999). Historical separation and present gene flow through a zone of secondary contact in ponderosa pine. *Evolution*, *53*, 769–776.

Lee, C. E. & Frost, B. W. (2002). Morphological stasis in the *Eurytemora affinis* species complex (Copepoda: Temoridae). *Hydrobiologia*, *480*, 111–128.

Mantel, N. (1967). The detection of disease clustering and generalized regression approach. *Cancer Research*, *27*, 209–220.

Michaelsen, W. (1899). Revision der Kinberg'schen Oligochaeten-Type. *Öfversigt Akademiens Förhandlingar*, *Stockholm*, *56*, 413–448.

Michaelsen, W. (1900). Oligochaeta. *Das Tierreich*, *10*, 1–575.

oreno, A. G. (1981). Estudio de algunas poblaciones de lombrices de tierra (Annelida: Oligochaeta: Lumbricidae, Megascolecidae y Glossoscolecidae) de los alrededores de Madrid. Phd Thesis, Universidad Complutense de Madrid, Madrid, Spain.

Oceguera-Figueroa, A., León-Regagnon, V. & Siddall, M. E. (2005). Phylogeny and revision of *Erpobdelliformes* (Annelida, Arhynchobdellida) from Mexico based on nuclear and mitochondrial gene sequences. *Revista Mexicana de Biodiversidad*, *76*, 191–198.

Omodeo, P. (1998). History of *Clitellata*. *Italian Journal of Zoology*, *65*, 51–73.

Pérez-Losada, M., Eiroa, J., Mato, S. & Domínguez, J. (2005). Phylogenetic species delimitation of the earthworms *Eisenia fetida* (Savigny 1826) and *Eisenia andrei* Bouché, 1972 (Oligochaeta, Lumbricidae) based on mitochondrial and nuclear DNA sequences. *Pedobiologia*, *49*, 317–324.

Pleguezuelos, J. M. (1997). *Distribución y Biogeografía de Los Anfibios y Reptiles en España y Portugal*. Granada: Universidad de Granada.

Pogson, G. H., Taggart, C. T., Mesa, K. A. & Boutilie, R. R. G. (2001). Isolation by distance in the atlantic cod, *Gadus morhua*, at large and small geographic scales. *Evolution*, *55*, 131–146.

Pop, A. A., Wink, M. & Pop, V. V. (2003). Use of 18S, 16S rDNA and cytochrome c oxidase sequences in earthworm taxonomy (Oligochaeta, Lumbricidae). *Pedobiologia*, *47*, 428–433.

Pop, A. A., Csuzdi, C., Wink, M. & Pop, V. V. (2005). An attempt to reconstruct the molecular phylogeny of the genus *Allolobophora* Eisen, 1874 (sensu lato, Pop, 1941) using 16S rDNA and COI sequences (Oligochaeta, Lumbricidae). In V. V. Pop & A. A. Pop (Eds) *Advances in Earthworm Taxonomy II*. Cluj-Napoca: Cluj University Press.

Posada, D. & Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics*, *14*, 817–818.

Posada, D. & Crandall, K. A. (2001). Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology and Evolution*, *16*, 37–45.

Qiu, J. P. & Bouché, M. B. (1998). Contribution to the taxonomy of Hormogastridae (Annelida: Oligochaeta) with description of new species from Spain. *Documents Pedozoologiques et Integrologiques*, *4*(15), 164–177.

Ramos-Onsis, S. E. & Rozas, J. (2002). Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, *19*, 2092–2100.

Rogers, A. R. (1995). Genetic evidence for a Pleistocene population explosion. *Evolution*, *49*, 608–615.

Rogers, A. R. & Harpending, H. (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, *9*, 552–569.

Romanowski, G., Lorenz, M. G. & Wackernagel, W. (1993). Use of polymerase chain reaction and electroporation of *Escherichia coli* to monitor the persistence of extracellular plasmid DNA introduced into natural soils. *Applied and Environmental Microbiology*, *59*, 3438–3446.

Ronquist, F. & Huelsenbeck, J. P. (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*, 1572–1574.

Rosa, D. (1887). *Bolletino Dei Musei Di Zoologia Ed Anatomia Comparata Della Università Di Torino*, *2*(32), 1.

Rousset, F. (1997). Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics*, *145*, 1219–1228.

Rozas, J., Sánchez-del-Barrio, J. C., Messeguer, X. & Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, *19*, 2496–2497.

Saiki, R. K., Gelfand, D. H. & Stoffel, S. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, *239*, 487–491.

Salomone, N., Emerson, B. C., Hewitt, G. M. & Bernini, F. (2002). Phylogenetic relationships among the Canary Island *Steganacaridae* (Acari, Oribatida) inferred from mitochondrial DNA sequence data. *Molecular Ecology*, *11*, 79–89.

Schneider, S. & Excoffier, L. (1999). Estimation of demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: Application to human mitochondrial DNA. *Genetics*, *152*, 1079–1089.

Schneider, S., Roessli, D. & Excoffier, L. (2000). Arlequin ver. 2.000: a software for population genetics data analysis. *Genetics and Biometry Laboratory*. Switzerland: University of Geneva.

Slatkin, M. (1993). Isolation by distance in equilibrium and nonequilibrium populations. *Evolution*, *47*, 264–279.

Slatkin, M. & Hudson, R. R. (1991). Pairwise comparison of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, *129*, 555–562.

Swofford, D. L. (2002). *Paup*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4. Sunderland, Massachusetts: Sinauer Associates.

Taberlet, P., Fumagalli, L., Wust-Saucy, A. G. & Cosson, J. F. (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, *7*, 453–464.

Templeton, A. R., Crandall, K. A. & Sing, C. F. (1992). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data III. Cladogram estimation. *Genetics*, *132*, 619–633.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, *24*, 4876–4882.

Tsai, Y. L. & Olson, B. H. (1992). Rapid method for separation of bacterial DNA from humic substances in sediments for polymerase chain reaction. *Applied and Environmental Microbiology*, *58*, 2292–2229.

Turgeon, J. & Bernatchez, L. (2001). Clinal variation at microsatellite loci reveals historical secondary intergradation between glacial races of *Coregonus artedi* (Teleostei: Coregoninae). *Evolution*, *55*, 2274–2286.

Watterson, G. A. (1984). Allele frequencies after a bottleneck. *Theoretical Population Biology*, *26*, 387–407.

Watterson, G. A. & Guess, H. A. (1977). Is the most frequent allele the oldest? *Theoretical Population Biology*, *11*, 141–160.

Weir, B. S. & Cockerham, C. C. (1984). Estimating F-Statistics for the analysis of population structure. *Evolution*, *38*, 1358–1370.

Wiens, J. J. & Penkrot, T. A. (2002). Delimitig species using DNA and morphological variation and discordant species limits in spiny lizards (*Sceloporus*)'. *Systematic Biology*, *51*(1), 69–91.

Wright, S. (1943). Isolation by distance. *Genetics*, *28*, 114–113.

Appendix I Frequencies of COI sequence haplotypes in *H. elisae* and GenBank Accession numbers.

Haplotype	Accession number	Anchuelo N9	Lozoyuela N12	Molar N12	Redueña N14	Pardo N12	Sevilla N12	Siguero N11
ANC1	EF653870	0.222 (2)						
ANC2	EF653871	0.111 (1)						
ANC3	EF653868	0.111 (1)						
ANC4	EF653869	0.222 (2)						
ANC5	EF653872	0.222 (2)						
ANC6	EF653873	0.111 (1)						
LOZ1	EF653888		0.0833 (1)					
LOZ2	EF653889		0.167 (2)					
LOZ3	EF653886		0.167 (2)					
LOZ4	EF653887		0.25 (3)					
LOZ5	EF653890		0.333 (4)					
MOL1	EF653874			0.167 (2)				
MOL2	EF653875			0.167 (2)				
MOL3	EF653876			0.0833 (1)				
MOL4	EF653877			0.25 (3)				
MOL5	EF653878			0.167 (2)				
MOL6	EF653879			0.0833 (1)				
MOL7	EF653880			0.0833 (1)				
RED1	EF653883				0.286 (4)			
RED2	EF653881				0.0714 (1)			
RED3	EF653884				0.0714 (1)			
RED4	EF653882				0.0714 (1)			
RED5	EF653885				0.5 (7)			
PAR1	EF653898					0.0833 (1)		
PAR2	EF653900					0.0833 (1)		
PAR3	EF653901					0.0833 (1)		
PAR4	EF653899					0.417 (5)		
PAR5	EF653902					0.333 (4)		
SEV1	EF653905						0.167 (2)	
SEV2	EF653903						0.333 (4)	
SEV3	EF653904						0.5 (6)	
SIG1	EF653893							0.0909 (1)
SIG2	EF653891							0.0909 (1)
SIG3	EF653892							0.455 (5)
SIG4	EF653894							0.0909 (1)
SIG5	EF653895							0.0909 (1)
SIG6	EF653896							0.0909 (1)
SIG7	EF653897							0.0909 (1)

N indicates number of sampled individuals in each locality. Number of individuals that presented each haplotype is shown after frequency values (in parentheses).