Genetic diversity and population structure in an invasive pantropical earthworm along an altitudinal gradient

Authors

Diana Ortíz-Gamino, Luis Cunha, Esperanza Martínez-Romero, Norma Flores-Estévez and Ángel I. Ortíz-Ceballos

Affiliations

1 Instituto de Biotecnología y Ecología Aplicada (INBIOTECA), Universidad Veracruzana, Av. de las Culturas Veracruzanas No. 101 (Campus USBI) Col. Emiliano Zapata, 91090 Xalapa, Veracruz, México.

2 Organisms and Environment Division, School of Biosciences, Cardiff University, Cardiff, CF103AT, UK.

2 Embrapa Florestas, Estrada da Ribeira, Colombo, Paraná, Brazil

3 Ecología Genómica, Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Av. Universidad s/n Col. Chamilpa 62210, Cuernavaca, Morelos, México.

*Correspondence: Diana Ortíz-Gamino, Email: diana.gamino@gmail.com.
Abstract

Population genetic analyses of the invasive pantropical earthworm *P. corethrurus* populations will contribute significantly to better understand the ecology and especially the reproductive system of this species. Using 34 polymorphic ISSR markers the genetic diversity and population structure was assessed for four populations of *P. corethrurus* along an altitudinal gradient, ranging from sea level up to ~1667 meters. Nuclear markers were able distinguish two genetic clusters, probably corresponding to two distinct genetic lineages, herein defined as A and B. Clones were detected in one population (Actopan at 480 masl) and its number was lower than expected for a parthenogenetic species. Nevertheless, low levels of genetic diversity and high number of intermediary genotypes were detected among the studied *P. corethrurus* populations with no apparent population structure related to the distinct geographic regions, which may indicate that human-mediated transference is prevalent, in particular, for the lower altitude regions. Hybridisation between the two genetic clusters was tested and pointed to 11 MLGs as being later-generation hybrids (B1 introgression) mainly associated with the three lower altitude regions. Still, most of the individuals seem to belong to lineage A and only five individuals seem to belong exclusively to the lineage B. Interestingly, these parental individuals were only found present at the highest altitude site, Naolinco (1566-1667 masl), which also showed the highest values of genotypic richness. During the biological invasion, multiple introduction of different genetic lineages can provide opportunities for admixture among genetically distinct clusters. The signatures of admixture among *P. corethrurus* populations along the altitudinal gradient in Mexico may have allowed the invasion success by directly increasing fitness. ISSR markers revealed to be useful for the study of genetic variation in the invasive pantropical earthworm, *P. corethrurus*.
Keywords: Invasive earthworms, *Pontoscolex corethrurus*, asexual reproduction, altitudinal gradient.
1. Introduction

Earthworms are not only ubiquitous ecosystem engineers, but through their action the functioning of the ecosystem is sustained (Lavelle et al. 1997; Brown et al. 2001; Barros et al. 2004). Even with the importance that soils organisms have into the ground, terrestrial ecosystem seems to be the least studied (González et al. 2006; Decaëns 2010). In fact, for a long time, little recognition has been given to invasions of soil organisms (Gates 1954); however during the last 30 years belowground invasion has become highly important (Hendrix et al. 2002; Hendrix and Bohlen 2006; Craven et al. 2016; Cicconardi et al. 2017). The first records in Mexico about exotic earthworms was made between 1900-1906, lumbricids (Dendrobaena octaedra, Lumbricus Terrestris), megascoleids (Dichogaster bolauï) and rhinodrilids (Pontoscolex corethrurus) (Eisen 1900; Bedard 1912), and until now, 51 exotic species were described by classical taxonomy (Fragoso and Rojas 2014). The role involved into earthworm’s invasion is not fully understood, i.e. distribution and establishment of exotic earthworms are due to different factors acting at different time-spatial scale (Marinissen and Van den Bosch 1992; Baker et al. 2006; Brown et al. 2006; González et al. 2006; Cameron and Bayne 2009; Marichal et al. 2012; Narayanan et al. 2016). On the other side, the key of invasive earthworm’s success could be based on its molecular plasticity supported by high genomic promiscuity tightly associated with reproductive strategies (Vitturi et al. 2002; Bengtsson 2009; Dupont et al. 2012; Fernández et al. 2013; Cunha et al. 2014; Pavlíček et al. 2016).

Parthenogenesis is not uncommon in earthworms and it is usually associated with a dispersal advantage, mainly because a single propagule could be sufficient to establish a new population (Terhivuo and Saura 2006). Thus, a rapid adaptation of parthenogenetic clonal populations may be an important mechanism for the success during the colonisation event.
(González et al. 2006; Hendrix et al. 2006; Terhivuo and Saura 2006) as *Pontoscolex corethrurus*.

*P. corethrurus* is a species of earthworms that was renamed into the Rhinodrilidae family (James 2012). This species comes from South America, Guiana Shield area of the Amazon (Gates 1954; Righi 1984) and it has been suggested that its reproduction is usually parthenogenetic (Gates 1973; Buch et al. 2011). Its growth and reproduction is not limited anymore by temperature (Ortíz-Gamino et al. 2016; Narayanan et al. 2016). In parthenogenetic species such as *P. corethrurus*, the genetic variation in regions far from their natural dispersion range may be low if only one or very few individuals were colonizers (Dupont et al. 2012). In addition, there is evidence of a cryptic species complex of *P. corethrurus*, for which two deeply divergent mitochondrial lineages have been detected (Cunha et al. 2014), although using nuclear markers such divergence was not that conspicuous.

In a recent study (Ortíz-Gamino et al. 2016), it was found that the altitudinal distribution of *P. corethrurus* in the central zone of the Veracruz State is extensive and goes from sea level to around 1600 meters of altitude with an average temperature of 17 °C. In this study, it was hypothesised that populations of *P. corethrurus* along the altitudinal gradient could be distinguished into two different ecotypes, which could correspond to a different genetic background or even to the presence of cryptic species (Cunha et al. 2014; Ortíz-Gamino et al. 2016). Also in India, after just 100 years since the introduction of *P. corethrurus* is locally naturalized and shows a wide distribution along an altitudinal gradient (Narayanan et al. 2016), similar to what was found in Veracruz State, Mexico. However, as a parthenogenetic species, its dispersal seems to be strongly driven by human activities (González et al. 2006; Feijoo et al. 2007; Hendrix et al. 2008; Dupont et al. 2012).
*P. corethrurus* is considered an important ecosystem engineer (Lavelle et al. 1997; Brown et al. 2001; Domínguez and Gómez-Brandón 2009), its role should not be neglected (Lavelle et al. 1987; Pashanasi et al. 1992; Pashanasi 1994; Brown et al. 2004; Fragoso et al. 2009; Geissen et al. 2009; Huerta and Van der Wal 2012) but due to the fact that it can quickly become a dominant species in human disturbed environments, sometimes becoming the only earthworm species present (Lapied and Lavelle 2003; Feijoo et al. 2007; Uribe et al. 2012; Hernández-Castellanos et al. 2013; Zavala et al. 2013; Gutiérrez and Cardona 2014) ecological researches are necessary.

Population genetic analyses of the invasive pantropical earthworm *P. corethrurus* populations will contribute significantly to better understand the ecology and especially the reproductive system of this species (Marichal et al. 2012; Cunha et al. 2014), which might be crucial for planning successful strategies for biological invasion mitigation or for future assumptions.

In order to evaluate the clonal diversity of parthenogenetic earthworms, biochemical markers have been used like allozyme (Terhivuo and Saura 1993; Terhivuo and Saura 2006), and molecular markers like mitochondrial regions (Heethoff et al. 2004; Knott and Haimi 2010; Fernández et al. 2011; Fernández et al. 2013; Cunha et al. 2014). Several molecular methods for assaying nuclear genetic diversity have been used attempted to estimate the relationships and variability in earthworms, through AFLP (Cunha et al. 2014), RADtag (Giska et al. 2015) and RAPD, ISSR and URPs (Sharma et al. 2011) contributing significantly to unveil genetic relationships within earthworms. Inter Simple Sequence Repeats (ISSR) are useful dominant nuclear markers commonly used to study genetic variation and population structure within and among populations. ISSR have been effective in detecting certain cryptic species in mosses, such
as *Aneura pinguis* and *A. maximus* (Buczkowska et al. 2016). For *Ae. aegypti*, ISSR exhibited four unique banding profiles for each of the four genotypes (Abbot 2001). For blackfly species, ISSR-PCR is a powerful technique for fingerprinting (Dušinský et al. 2006).

The main advantages of ISSR-PCR are associated to the fact that they do not require the knowledge of genomic sequences, low cost, able to detect of high levels of polymorphism, require simple operation, can be of high stability and reproducibility, and have a rapid turnover (Wolfe and Liston 1998; Abbot 2001; Dušinský et al. 2006; Sharma et al. 2011; Ng and Tan 2015).

The objective of this work was to determine the genetic variation and population structure in four populations of *P. corethrurus* along an altitudinal gradient in the central area of the State of Veracruz using ISSR as nuclear markers. As a hypothesis, it is suggested that genetic diversity would be higher in the intermediate zone of the chosen gradient, since the extremes of the gradients are usually the most stressful environments and in the intermediary area a relatively optimal environment could be found.

**2. Material & Methods**

**2.1 Sampling Sites and Animal Collection**

The altitudinal distribution of *P. corethrurus* along the central region of Veracruz State goes from the sea level up to 1,667 masl. Four sampling points were established at each of four sites along this altitudinal gradient: Laguna Verde (LV), Actopan (AC), Ingenio La Concepción (LC) and Naolinco (NA) at 24, 480, 973–1,036 and 1,566–1,667 masl, respectively (Table 1 and Figure 1) (Ortíz-Gamino et al. 2016).
During September of 2013, 40 mature (clitellate) individuals were collected (n= 10 per site). Animals were kept in plastic boxes with moistened soil and transported to the laboratory at INBIOTECA for taxonomical/anatomy identification (Moreno and Borges 2004). Individual were rinsed in water, washed from soil particles and posteriorly fixated in 96% ethanol. Animals were conserved at -20 until further processing.

Table 1. Attributes of the sampling sites along an altitudinal gradient in Veracruz State, México.

<table>
<thead>
<tr>
<th>Sampling Site</th>
<th>Altitude (masl)</th>
<th>Climate</th>
<th>Soil's texture</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Clay (%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Silt (%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sand (%)</td>
</tr>
</tbody>
</table>

Figure 1. Sampling sites along an altitudinal gradient in center region of Veracruz state, Mexico. LV, Laguna verde (24 masl), AC, Actopan (480 masl), LC, Ingenio La Concepción (973-1036 masl) and NA, Naolinco (1566-1667 masl). The digital elevation model was created using data provided by Instituto Nacional de Estadística y Geografía, México.
<table>
<thead>
<tr>
<th>Location</th>
<th>Code</th>
<th>Climate</th>
<th>Temperature 1</th>
<th>Temperature 2</th>
<th>Temperature 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laguna Verde (LV)</td>
<td>24</td>
<td>Aw1(w)(i')g</td>
<td>25.6-27.6</td>
<td>21.8-34.4</td>
<td>40.0-50.6</td>
</tr>
<tr>
<td>Actopan (AC)</td>
<td>480</td>
<td>Aw0(w)(i')gw&quot;</td>
<td>12.0-13.6</td>
<td>28.8-35.8</td>
<td>50.6-59.2</td>
</tr>
<tr>
<td>Ingenio La Concepción (LC)</td>
<td>973-1036</td>
<td>(A)Ca(f)(i')gw&quot;</td>
<td>25.6-27.6</td>
<td>21.8-34.4</td>
<td>40.0-50.6</td>
</tr>
<tr>
<td>Naolinco (NA)</td>
<td>1566-1667</td>
<td>Cb (fm)(i')gw&quot;</td>
<td>12.0-13.6</td>
<td>28.8-35.8</td>
<td>50.6-59.2</td>
</tr>
</tbody>
</table>

Notes:
Climate: Aw1(w)(i')g and Aw0(w)(i')gw" are used for warm and sub-humid climate; (A)Ca(f)(i')gw" for warm and humid and Cb (fm)(i')gw" for wet and semi-humid climate. For more details (average temperature, evapotranspiration, total annual precipitation, etc.) see Ortiz-Gamino et al. (2016).

2.2 DNA isolation and quantification

Tail wall tissue was used for extraction of genomic DNA. Total DNA was extracted using the DNeasy Blood & Tissue kit (Qiagen) following the manufacturer’s instructions. DNA was checked for quality by gel electrophoresis and quantified using a spectrophotometer (ND-2000, Nanodrop Technologies, Wilmington, DE).

2.3 ISSR protocol

Forty Pontoscolex corethrurus individuals (n = 10 per site) were used for ISSR screening. This screen used 5 ISSR primers (Table 2), which were previously reported to produce polymorphic and reproducible DNA fingerprints from Eudrilus eugeniae and E. fetida (Sharma et al. 2011). Each PCR reaction contained 1X PCR buffer, 2 mM MgCl₂, dNTPs at 100 µM each, primers at 0.8 µM each, 1.5 U DNA Taq polymerase, 1ug/ul BSA and 30 ng template DNA. The PCR reaction mix was adjusted to a final volume of 10 µL with water. PCR amplifications were performed in an Eppendorf Mastercycler. Conditions included an initial step of 3 min at 95°C, followed by 35 cycles of denaturation at 95°C for 30 seconds, annealing at primer-specific temperature (Table 2) for 30 seconds, and elongation at 72°C for 1 min. A final extension was performed at 72°C for 10 min. PCR products were visualized in 2% agarose gels (with ethidium
bromide at 1ul/ml). The initial screening used a total of 18 primers but only primers that were polymorphic and reproducible were chosen (Table 2).

Table 2. Primers used for ISSR amplification.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Ta (°C)</th>
<th>Maximum number of bands</th>
<th>Estimated size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>840</td>
<td>GAGAGAGAGAGAGAGAYT</td>
<td>59.5</td>
<td>8</td>
<td>2000-200</td>
</tr>
<tr>
<td>834</td>
<td>AGAGAGAGAGAGAGAGYT</td>
<td>61</td>
<td>7</td>
<td>2000-300</td>
</tr>
<tr>
<td>866</td>
<td>CTCCTCCTCCTCCTCCTC</td>
<td>70</td>
<td>6</td>
<td>2000-400</td>
</tr>
<tr>
<td>810</td>
<td>GAGAGAGAGAGAGAGAT</td>
<td>52.4</td>
<td>6</td>
<td>2000-400</td>
</tr>
<tr>
<td>807</td>
<td>AGAGAGAGAGAGAGAT</td>
<td>54.4</td>
<td>7</td>
<td>2000-300</td>
</tr>
</tbody>
</table>

2.4 Data analysis

The amplified DNA fragments were transformed into a binary matrix (1 = presence, 0 = absence). A multilocus genotype (MLG) was constructed for each individual by combining data of single ISSR fingerprints using the procedure available in the package POPPR (Kamvar et al. 2014). Isolates with the same MLG were considered clones, and some analyses were conducted for the original and clone-corrected dataset (Gramaje et al. 2014).

The R package POPPR (Kamvar et al. 2014) was used to calculate the dissimilarity distance matrices and to generate a minimum spanning network from the matrices. To assess the possible evolutionary relationships among MLGs, a minimum spanning network was constructed using the genotypes of each sampling location. Bootstrapping was performed with the R package POPPR (Kamvar et al. 2014) with 1000 bootstrap resampling. Genotypic diversity, genetic richness and the evenness index adapted from the Simpson index of genotypic diversity were calculated for each population using the R package POPPR (Kamvar et al. 2014). The function ‘rarecurve’ from the R package VEGAN (Oksanen et al. 2007) was used to generate rarefaction
curves representing species richness. Curves were calculated to determine if the sampling intensity was adequate to detect most of the *P. corethrurus*’ MGLs present in each population. Additionally, calculation for the minimum number of loci able to distinguish all MGLs was also performed. Because sample size varied among populations we employed rarefaction to explore the effect of sample size on observed species richness.

2.5 Genetic differentiation, structure and clustering analyses

The estimation of the genetic variance for all MLGs was evaluated by analysis of molecular variance (AMOVA), using the GenAlEx v.6.5 software (Peakall and Smouse 2006). It was calculated as PhiPT (analogue of F$_{ST}$ fixation index) for all populations and within populations (Peakall and Smouse 2006) relative to the total variance. Significance was obtained by using 9999 permutations, and confidence interval at 95%, by 10,000 re-samplings. For this analysis, only single copies of the different genotypes were used to give identical weight to MLGs. The Mantel test was used to evaluate whether there was a correlation between the matrix of genetic differentiation between the pairs of MLGs and the spatial distance matrix between the populations, using Arlequin v.3.5 (Excoffier and Lischer 2010). The existence of association between *P. corethrurus* individuals was assessed initially using a Principal Coordinates Analysis (PCA), implemented in GenAlEx v. 6.51 (Peakall and Smouse 2006). As PCA is independent of any genetic hypotheses it is suitable for the analysis of partially clonal species. Additionally, Unweighted Pair Group Method with Arithmetic Mean (UPGMA) dendrograms were also created using the POPPR package in R (R core team 2004) Bootstrapping was performed with the R package PVCLUST with 10,000 bootstrap resamplings (Suzuki and Shimodaira 2006).

Populations structure was explored using the Bayesian clustering method implemented in STRUCTURE v.2.3.4 (Pritchard et al. 2000) and distance based approach, a Discriminant
analysis of principal components (DAPC) implemented in Adegenet R package (Jombart et al. 2010). STRUCTURE v.2.3.4 (Pritchard et al. 2000) was used to identify the number of genetic clusters within the dataset, and to assign individuals to the defined clusters using an admixture model. For this, 15 independent runs for each K (1 - 8) to confirm consistency. A burn-in period of 100,000 iterations and further 200,000 MCMC generations. To determine the most likely value of K, we used the “BestK” implemented in CLUMPAK (Evanno et al. 2005), which uses ΔK method of Evanno et al. (2005) Results from the 15 replicate runs were combined using CLUMPAK online version (Evanno et al. 2005). Since DAPC allows assessment of structure of populations without making assumptions of panmixia makes it appropriated to populations that may show partial clonality. Moreover, this technique provides a robust alternative to compare to the Bayesian clustering methods like the one implemented in STRUCTURE (Pritchard et al. 2000). Relative dissimilarity distances were calculated according to the index of association (Brown et al. 1980; Smith et al. 1993). The approach returns a distance reflecting a ratio of the number of observed differences by the number of possible differences.

The hybridisation status of individuals according to the Bayesian genetic clusters defined in Structure (defined as putative Lineage A and Lineage B) was further investigated using NEWHYBRIDS v1.1 (Anderson and Thompson 2002) which also uses a Bayesian assignment, by implementing a multilocus allele frequency model-based approach. This approach performs clustering of MGLs without any a priori knowledge of parental allele frequencies, and also has the advantage of specifically assuming a mixture of parental and several hybrid and parental classes, such as the genotypes corresponding to parental lineages, F1, F2 (F1 x F1) and later generation or introgressive hybrids B1 (e.g. parental 1 x F1) and B2 (e.g. parental 2 x F1). Individual posterior probabilities belonging to each hybrid category were estimated using the
MCMC method in a Bayesian framework using Jeffreys-type priors and a burn-in period of 100000 iterations followed by 50000 sweeps from the posterior distribution sampling (Anderson and Thompson 2002). Linkage disequilibrium as an indication of random mating was calculated and tested for significance with 1,000 randomizations using the R package POPPR (R core team 2004) measures of gametic disequilibrium tested were the index of association ($I_A$) (Brown et al. 1980; Smith et al. 1993), and a standardized alternative of the $I_A$ ($\bar{r}_d$) (Agapow and Burt 2001). The null hypothesis for this test is that there is random association among alleles at different loci and $I_A=0$; the null hypothesis for random mating is rejected where $I_A>0$.

### 3. Results

#### 3.1 Genotypic diversity, genetic richness and evenness index

Table 3. Genetic parameters of genetic variation in *Pontoscolex corethrurus* along an altitudinal gradient in central Veracruz state, México.

<table>
<thead>
<tr>
<th>Sampling Site</th>
<th>N</th>
<th>MLG</th>
<th>eMLG</th>
<th>Pb</th>
<th>Tb</th>
<th>H</th>
<th>G</th>
<th>$E_S$</th>
<th>$H_{exp}$</th>
<th>Ia</th>
<th>$\bar{r}_D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LV</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>20</td>
<td>34</td>
<td>2.20</td>
<td>9</td>
<td>1</td>
<td>0.30</td>
<td>1.98*</td>
<td>0.10*</td>
</tr>
<tr>
<td>AT</td>
<td>10</td>
<td>8</td>
<td>8</td>
<td>22</td>
<td>34</td>
<td>2.03</td>
<td>7.14</td>
<td>0.93</td>
<td>0.30</td>
<td>3.23*</td>
<td>0.15*</td>
</tr>
<tr>
<td>LC</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>27</td>
<td>34</td>
<td>1.79</td>
<td>6</td>
<td>1</td>
<td>0.39</td>
<td>1.16*</td>
<td>0.04*</td>
</tr>
<tr>
<td>N</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>24</td>
<td>34</td>
<td>2.30</td>
<td>10</td>
<td>1</td>
<td>0.29</td>
<td>4.08*</td>
<td>0.18*</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>35</td>
<td>33</td>
<td>9.85</td>
<td>34</td>
<td>34</td>
<td>3.48</td>
<td>31.41</td>
<td>0.97</td>
<td>0.40</td>
<td>1.65</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Notes:

- N, Number of individuals observed; MLG, Number of multilocus genotypes observed; Pb, polymorphic bands; Tb, total number of bands present H, Shannon-Wiener Index of MLG diversity; G, Stoddart and Taylor’s Index of MLG diversity (Stoddart and Taylor 1988); Lambda, Simpson’s Index; $E_S$, Evenness, (Pielou 1975; Ludwig and Reynolds 1988; Grünwald et al. 2003); $H_{exp}$, Nei’s unbiased gene diversity (Nei 1978); Ia, the index of association (Brown et al. 1980; Smith et al. 1993); $\bar{r}_D$, the standardized index of association ($\bar{r}_d$) (Agapow and Burt 2001); * shows significant value with $p < 0.01$

All populations showed high genotypic diversity, but without significant differences between the four populations. The five ISSR primers generated a total of 33 MLGs among the 35 *P.*
corthrurus individuals. Two Actopan MLG were observed twice in that population. The rest of 31 MLGs were observed only once. Overall, MLG diversity (H) varied across populations defined by different altitudes in Veracruz (Table 3). Evenness values were higher in Laguna Verde (24 masl), La Concepcion (973-1036 masl) and Naolinco (1566-1667 masl), and this is due to the fact that all genotypes found were unique, compared with the population from Actopan (480 masl), that showed repeated genotypes (0.93). Nei’s unbiased gene diversity (Hexp) values varied from the highest found in La Concepcion (0.39) and to a lowest in Naolinco (0.29). According to the rarefaction curves (Figure 2A), Naolinco had a higher number of samples and observed MGLs relative to the other populations. Comparison of rarefaction curves showed significant differences in diversity among these populations with Naolinco population being more diverse than the others, and with La Concepcion having a lower level of MLG diversity. Interestingly, the minimum number of loci to define the total number of MLGs found reached a plateau after 18 loci (Figure 2B).
Figure 2. Rarefaction curve of expected number of multilocus genotypes captured per individual sampled (A), and a multilocus genotype accumulation curve according to the number of loci sampled (B).

3.2 Genetic differentiation

The AMOVA results (Table 4), showed 25% of the variation among populations and 75% within populations. The overall PhiPT was 0.25 (with a PhiPT_max of 0.686 and a Phi’PT of 0.362) with an associated permutation p-value <0.001.

Table 4. Analysis of molecular variance (AMOVA)

<table>
<thead>
<tr>
<th>Level of variation</th>
<th>d.f.</th>
<th>SS</th>
<th>MS</th>
<th>Est. Var.</th>
<th>Proportion (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among Populations</td>
<td>3</td>
<td>59.079</td>
<td>19.693</td>
<td>1.760</td>
<td>25%</td>
</tr>
<tr>
<td>Within Populations</td>
<td>29</td>
<td>154.497</td>
<td>5.327</td>
<td>5.327</td>
<td>75%</td>
</tr>
<tr>
<td>Total</td>
<td>32</td>
<td>213.576</td>
<td>7.088</td>
<td></td>
<td>100%</td>
</tr>
</tbody>
</table>
### 3.3 Clustering and structure analysis

The principal component analysis (PCA) showed that, although some individuals cluster together per site (e.g. animals in Naolinco), most of the individuals scattered in a non-uniform clustering (Figure 3A). Axis 1 and 2 of the PCA accounted for 26% and 15% of total genetic variability. The global minimum spanning network showed that all populations have MLGs that are closely related (Figure 3B). The comparison of the Euclidian genetic distance matrix with the matrix of geographic distances using the Mantel test showed that there is no correlation between these two matrices suggesting that data in the genetic distance matrix is not explained by the geographic positioning of the populations. The optimal UPGMA tree for the individuals and by populations is presented with bootstrapped values (Figure 4A and 4B respectively). In both analyses, the PCA and dendrogram there is no clear discrimination related to the geographic distances. Still, La Concepcion and Laguna Verde seem to be closely associated although geographically distant.
Figure 3. A Principal Component Analysis (A) and a Minimum Spanning Network (B) showing the relationship among multilocus genotypes (MLGs) observed along an altitudinal gradient in central Veracruz State, Mexico in four distinct populations. Each node signifies a different MLG. Node sizes and colours match to the number of individuals and population membership, respectively. Edge thickness and colour are proportional to absolute genetic distance. Edge lengths are arbitrary.
Figure 4. UPGMA dendrogram of genetic distance among MGLs (A) and among Populations (B) observed along an altitudinal gradient in central Veracruz State, Mexico. Only values with support greater than 70% using 1000 bootstrap samples are shown.

Figure 5. The estimated population structure with a summary plot of the estimates of Q based on the ISSR data observed along four populations along an altitudinal gradient in central Veracruz State, Mexico. Each individual is
represented by a vertical line, which is partitioned into $K$ coloured segments representing the individual's estimated membership fractions in $K$ clusters ($\%$).

The Bayesian analysis of population structure estimated two distinct genetic clusters ($K = 2$, $\text{Ln } P(D) = -557.89 \pm 0.31$, s.d.) distributed along the four geographic populations (Fig. 5). Similarly, as the PCA and dendrogram, there is no apparent structuring according to the geographic location, however, the population at Naolinco seems to belong mostly to cluster 2 while Laguna Verde to cluster 1, intermediary altitude sites, Actopan and La Concepcion seem to be strongly admixed, having a more heterogenous proportion of each cluster among individuals. Apparently, two distinct genetic lineages are present among the populations, and will be herein defined as Lineage A (mostly present in Laguna Verde) and Lineage B (mostly present in Naolinco). The DAPC analysis (Figure 6) confirmed the relation found on UPGMA dendrogram where La Concepcion and Laguna Verde, although geographically distant they seem to be closely related in terms of structuring, potentially associated with the heterogenous proportions of each genetic cluster as defined in the Bayesian analysis, while Actopan is detached (Lineage A) as in the Bayesian analysis, which also could explain the detachment of Naolinco population (Lineage B). Notable, both clone-corrected ($n = 33$) and uncorrected data ($n = 35$) reject the hypothesis of no linkage among markers as the probabilities (both $I_A$ and $r_d$ show significant $p$-values, see Table 3), these results support asexuality in populations (Supplementary Figure 1). Interestingly, the Bayesian clustering according putative hybridisation status (Figure 7) showed 12 MLGs as being later-generation hybrids B1 (Lineage A x F1 introgression) mainly associated with the three lower altitude regions. In general, individuals belonging to parental Lineage A dominated (16 individuals) followed by later-generation B1 (12 individuals).
Figure 6. Genetic structure using the ISSR data in 35 *Pontoscolex corethrurus* individuals based on discriminant analysis of principal components (DAPC). Proportion of Eigenvalues in discriminant analysis is given on bottom left graphic and bottom right shows the PCA Eigenvalues, where the first 12 significant principal components are highlighted in black.
Figure 7. Classification of *P. corethrurus* individuals according to a Bayesian assignment algorithm implemented in NEWHYBRIDS (Anderson and Thompson, 2002) to detect gene flow. Each unit represents an individual corresponding to parental lineages (Lineage A and Lineage B), F1 generation, F2 (F1 x F1) and later generation or introgressive hybrids B1 (Lineage A x F1) and B2 (e.g. Lineage B x F1).

4. Discussion

The factors that are playing a role in the earthworms’ invasion processes are soil texture and chemical parameters (Marichal et al. 2012) but also other ecological variables (Hendrix and Bohlen 2006; Craven et al. 2016) and features inherent to the species (González et al. 2006; Cameron et al. 2008; Marichal et al. 2012). Among factors that are known to influence genetic structure in earthworms such as the slow dispersion capacity, life history and reproduction mode, and the environmental filters, which may contribute for population isolation, may favour the intra-specific relationships as strong factor acting on population structure (Baker 1998; Novo et al. 2010; Fernández et al. 2013). Additionally, in tropical regions temperature may also be imposing range restrictions in species complexes (Janzen 1967) which build up the assumption that temperature in tropics may promote isolation of earthworm’s species.

Genetic variability in asexual organisms may be high (Terhivuo and Saura 2006; Bengtsson 2009), however for *P. corethrurus*, even within its proposed endemic range shows
low genetic variation (Dupont et al. 2012). These results are the first showing genetic diversity and population structure of four *P. corethrurus* populations located along an altitudinal gradient through ISSR markers including when in Sharma et al. (2011) the best nuclear markers with *E. eugeniae* and *E. fetida* were URPs which showed 98.3% of polymorphism. Our hypothesis was expect to find higher genetic diversity in intermediary areas resulting from selection and more extreme environmental conditions, however, movement of MLGs along the altitudinal gradient seem more prevalent and more conspicuous, in particular along the lower altitude regions (Laguna Verde, Actopan and La Concepcion). This could be only explained by an intense human-mediated transference through road networks (Ortíz-Gamino et al. 2016) or probably associated with agriculture and other human activities (Lapied and Lavelle 2003; Baker et al. 2006; González et al. 2006; Feijoo et al. 2007; Dupont et al. 2012).

Clustering analysis (PCA and dendrogram based on genetic distances and bootstrapping) identified a huge admixture among all populations, but was possible identify at least two divergent and well-differentiated genetic clusters, potentially corresponding to different clonal lineages (Lineage A and Lineage B). Also in French Guiana, 80 individuals of *P. corethrurus* were grouped into two major clusters (Dupont et al. 2012), but this pattern was not found in Sao Miguel Island because the PCA constructed using AFLP markers was spread rather than clustered (Cunha et al. 2014). Here in Naolinco (1566-1667 masl) population showed the best distinguishable cluster of individuals. The unique Naolinco signature may be associated with the contrasting environmental conditions found at such higher altitude, which may be acting as an environmental filter over this population. Possibly, at the highest altitude found in Naolinco, and although, MGLs predominantly result from successive introductions, the temperature dispersion (year average of 17 °C) may be the main barrier to MGLs until now (Ortíz-Gamino et al. 2016),
the Lineage B could be adapted to lower temperatures. Nevertheless, under a global change scenario with intensive land-use change and alarming global warming (Jiménez and Decaëns 2000; Eisenhauer et al. 2014; Gutiérrez and Cardona 2014) this barrier could decrease and benefit invasive pantropical earthworm’s species. Otherwise, the lineage A seems spread since the sea level until LC (973-1036 masl) which could indicate a wide ecological tolerance and probably well-adapted to warm temperatures. The factors that are shaping this dispersion pattern in clonal lineages are multivariable; some examples have been documented, in Octolasion tyrtaeum some widespread and restricted clonal lineages have been identified (Terhivuo and Saura 1993) also showed that one single haplotype was widespread and found in all sampling stations. In Aporrectodea trapezoides, the clonal lineages seem to tend to stay close to their original areas, which can indicate some level of local adaptation and/or strong interspecific relations (Fernández et al. 2013). But for peregrine species, the evidence indicates that human activities are strongly shaping dispersion patterns and MLG introduction through accidental transfer (Cameron et al. 2008; Dupont et al. 2012; Ortíz-Gamino et al. 2016).

The AMOVA analyses showed that most genetic diversity (75%) was found within populations and only 25% among populations. Similar results were found by Cameron et al. (2008) in the earthworm Dendrobaena octaedra from Alberta, Canada. Cunha et al. (2014), using AFLP markers on P. corethrurus found that most of the variability was also found within populations and only a few proportion was associated between populations. The two distinguishable genetic clusters revealed here by the Bayesian analysis would then correspond to an established asexual lineage intermixed with individuals derived at least from two different asexual lineages (Gramaje et al. 2014). Indeed, multiple introductions of different genetic lineages is highly probable, there are at least three well-established P. corethrurus lineages, the
original from South America (Dupont et al. 2012), other that probably has 100 years in India (Narayanan et al. 2016), and the oldest one described until now in Sao Miguel Island in the Azores Archipelago (Portugal) with ~200 years but even there the 28 haplotypes that were found, all belonging to a single genetic lineage (Cunha et al. 2014). The accidental transfer of P. corethrurus between Mexico and these three sites (French Guiana, India and Azores) can happen (Cuatrecasas 1964; Schultes 1984).

The standardized index of association (τd) supported the hypothesis for clonal population structure due the widespread of several intermediary and similar MLGs along the altitudinal gradient, but also by the linkage disequilibrium tests, where the null hypothesis of random mating was rejected for all populations. Nevertheless, the interpretation of the results requires caution because it is difficult to demonstrate the presence of linkage disequilibrium with small sample sizes (Hagenblad et al. 2006; Du et al. 2007; Gramaje et al. 2014). The high levels of genotypic diversity found in this study could be expected because there is evidence of high diversity found in populations of P. corethrurus using AFLPs that may be indicative of polyploidy (Cunha et al. 2014), but also the results are due because the evolutionary rate of change within repeat regions compared with other nuclear regions like AFLPs, RAPD or URPs (Sharma et al. 2011; Gramaje et al. 2014; Seyedimoradi and Talebi 2014; Ng and Tan 2015). Even with several studies showing high percentage of polymorphic loci using ISSR (Abbot 2001; Dušinský et al. 2006; Luan et al. 2006; Sharma et al. 2011; Seyedimoradi and Talebi 2014; Ng and Tan 2015; Buczkowska et al. 2016). Our decision to use ISSR markers was supported by their potential of discrimination among geographically more or less isolated populations (Dušinský et al. 2006), the capacity of ISSR to distinguish among cryptic species (Buczkowska et al. 2016), but also ISSR amplify only nuclear regions of eukaryote (Seyedimoradi and Talebi
2014) and as an endogeic earthworm, *P. corethrurus* is commonly associated with Bacteroidetes phylum (Bernard et al. 2012).

Finally, sexual reproduction is a rare event in *P. corethrurus*, is widely accepted that its reproduction occurs mainly by parthenogenesis (Gates 1973). In earthworms, parthenogenesis was associated with polyploidy as well to high levels of DNA methylation (Regev et al. 1998). Therefore, is also plausible that methylation may be fostering the epigenetic control of phenotypic plasticity, which is crucial for colonisation success (Stürzenbaum et al. 2009). Assuming this hypothesis, *P. corethrurus* response to temperate environments (Ortíz-Gamino et al. 2016) may be sustained by asexual reproduction, polyploidy and epigenetic control.

5. Conclusion

Over the last hundred years *P. corethrurus* has been shown to be involved with human disturbance in many regions worldwide. We exposed the capability of ISSR to show high genetic diversity in four populations of the invasive earthworm *P. corethrurus* along an altitudinal gradient in tropical areas. These results revealed the existence of at least two well-differentiated genetic clusters, potentially corresponding to different clonal lineages. Continued genotyping of *P. corethrurus* individuals from same and other populations will be necessary to confirm the results of genotype diversity, to track the movement and diversification of the lineages and to identify new dominant genotypes or newly introduced lineages. Our evidence spotlight the necessity of more studies about the ecology of the invasive species, because its establishment is so complex, other parameters could help to understanding why *P. corethrurus* populations occur in some site and not in others (Marichal et al. 2012) but we emphasize in particular about reproductive behaviour of this animal, but this also applies for any peregrine species (Baker et al.
Furthermore, a comprehensive taxonomic revision of *P. corethrurus* cryptic species complex (Cunha et al. 2014) is needed in order to assess the precise taxonomic placement and rank. This would be helpful to dissect the within variability found in our studied populations and precisely define any lineages present.

**Acknowledgements**

We would like to thank to Wilberth Poot Poot for the support and help at INBIOTECa, Ernesto Ormeño for his help at Ciencias Genómicas and to Rogelio Lara and Katia Martínez for their help during fieldwork. Also, thankful to Luis Alberto Lara and Dulce Galván for suggestions and comments on the manuscript. Thankful to Alberto Ortíz Brito for proving the sampling map. Diana Ortíz-Gamino was supported by a scholarship by Consejo Nacional de Ciencia y Tecnología (CONACyT, Mexico: No. 251818). Luis Cunha was supported by a EU Marie Curie fellowship - MSCA-IF-2014-GF-660378.

**References**


Eisen G (1900) Researches in the American Oligocheta, with special reference to those of the Pacific coast and adjacent islands. Proceedings of California Academy of Science. 2:85–276


Hendrix PF, Bohlen PJ (2002) Exotic Earthworm Invasions in North America: Ecological and Policy Implications Expanding global commerce may be increasing the likelihood of exotic earthworm invasions, which could have negative implications for soil processes, other animal and plant species, and importation of certain pathogens. Bioscience 52(9): 801-811.


Ortiz-Gamino D, Pérez-Rodríguez P, Ortiz-Ceballos AI (2016) Invasion of the tropical earthworm Pontoscolex corethrurus (Rhinodrilidae, Oligochaeta) in temperate grasslands. PeerJ. 4, e2572. doi.org/10.7717/peerj.2572


Pashanasi B, Melendez G, Szott L, Lavelle P (1992) Effect of inoculation with the endogeic earthworm Pontoscolex corethrurus (Glossoscolecidiae) on N availability, soil microbial biomass


Vitturi R, Colomba MS, Pirrone AM, Mandrioli M (2002) rDNA (18S–28S and 5S) colocalization and linkage between ribosomal genes and (TTAGGG)(n) telomeric sequence in
