Genetic variation of the epigeic earthworm [Lumbricus castaneus] populations in urban soils of the Paris region (France) revealed using eight newly developed microsatellite markers

Lise Dupont, Maxime Pauwels, Cassandre Dume, Valentin Deschins, Helene Audusseau, Agnès Gigon, Florence Dubs, Franck Vandenbulcke

Short communication

Genetic variation of the epigeic earthworm *Lumbricus castaneus* populations in urban soils of the Paris region (France) revealed using eight newly developed microsatellite markers

Lise Dupont [1], Maxime Pauwels [2], Cassandre Dume [1,3], Valentin Deschins [1], Hélène Audusseau [1], Agnès Gigon [1], Florence Dubs [1] and Franck Vandenbulcke [3].

1. Université Paris Est Créteil (UPEC), Sorbonne universités, Paris 7, CNRS, INRA, IRD, Institut d'écologie et des sciences de l'environnement de Paris, 94010 Créteil Cedex, France

2. Université de Lille, CNRS, UMR 8198 – Unité Evolution-Ecologie-Paléontologie, F-59000 Lille, France

3. Université de Lille, EA 4515-LGCgE - Laboratoire Génie Civil et géo-Environnement, Cité scientifique, SN3, F-59655 Villeneuve d'Ascq, France

* Corresponding author. Tel : +33(0)145171664 ; e-mail address : lise.dupont@u-pec.fr

Keywords: earthworm populations, genetic diversity, inbreeding, urban area

Highlights

- Eight new microsatellite loci were developed for the earthworm *Lumbricus castaneus*
- These markers revealed an important genetic diversity in 6 populations of the Paris region
- Geographically close populations were genetically significantly different.

Abstract

Urban soils are subject to intense environmental pressures, e.g. physical disturbance, surface transformation and pollution, which greatly impact the activity of soil organisms. The epigeic earthworm *Lumbricus castaneus* (Savigny, 1826) is known to be tolerant to trace elements contamination of soils and was found to be abundant in the urban context. In order to investigate how urban environmental stressors shape the population genetic variation of this species, we developed 8 microsatellite loci from a microsatellite-enriched genomic library. Polymorphism was explored in 6 populations (153 individuals) from the Paris region. The number of alleles per locus varied from 7 to 29. The average expected heterozygosity within populations ranged from 0.596 to 0.705. These new microsatellite markers revealed a significant genetic structure at fine spatial scale, highlighting that they could be particularly useful for genetic studies of populations in urban soils.

1. Introduction

Levels of genetic variation within and among populations are determined by the joint action of mutation, selection, genetic drift, and gene flow (Hartl and Clark, 2007), which in turn operate within the historical, ecological and biological context of each species (Loveless and Hamrick, 1984). Thus, ecological preferences and life history traits can play a major role in shaping the population genetic structure of a species (Gonzalez and Zardoya, 2007). Understanding the evolution of genetic variation is a priority for biodiversity conservation and environmental biomonitoring programs, in particular because levels of genetic variation is a major determinant of the long-term maintenance of species. This is particularly true in urban environments where environmental disturbances of anthropogenic origin such as pollution, geographic isolation and habitat fragmentation can profoundly affect population genetic parameters (Johnson and Munshi-South, 2017). Despite their challenging environment, some species are able to exploit urban habitats and to maintain their populations (Takami et al., 2004). It is important for biodiversity management programs to determine the genetic background of these species evolving in urban ecosystems.

Soils are the foundation for many ecological processes and provide key ecosystem services even in urban areas where they are polluted, physically disturbed and frequently artificial (review in Pickett et al., 2011). Elevated trace element concentrations are universally reported in urban soils and sources have been associated with roadside environments, interior and exterior paint, refuse incinerators, industrial stack emissions, management and industrial waste (Pickett et al., 2011). Such contamination of soils by trace elements is known to affect exposed species, such as earthworms (Mussali-Galante et al., 2014). In particular, metal pollution is expected to profoundly alter the genetic variation within populations, in four different ways : (i) by increasing mutation rates, (ii) by directional selection on tolerant genotypes, (iii) by causing bottleneck events, and (iv) by altering migration (i.e. reducing or

increasing gene flow in relation with the level of tolerance of the individuals) (Mussali-Galante et al., 2014; van Straalen and Timmermans, 2002).

Microsatellites are commonly used to infer microevolutionary processes such as mutation, genetic drift and gene flow and provide invaluable insights into the overall levels of neutral genetic variation (i.e. neutral means that gene variants detected do not have any direct effect on fitness) within populations (Kirk and Freeland, 2011). Although microsatellite markers have been developed for eight earthworm species to date: *Lumbricus rubellus* (Harper et al., 2006), *L. terrestris* (Souleman et al., 2016; Velavan et al., 2007), *Hormogaster elisae* (Novo et al., 2008), *Allolobophora chlorotica* (Dupont et al., 2011), *Aporrectodea icterica* (Torres-Leguizamon et al., 2012), *Aporrectodea longa* (Strunk et al., 2012), *Eisenia fetida* (Somers et al., 2011) and *Amynthas cortices* (Cunha et al., 2017), they have not been used to investigate genetic changes in population exposed to trace elements yet.

Here, our main aim was to develop and characterize microsatellite loci for *L. castaneus* (Savigny, 1826), a diploid and obligatory amphimictic earthworm species. This epigeic earthworm is known to be tolerant to trace element contamination of soils (Grumiaux et al., 2015; Spurgeon and Hopkin, 1996). For instance, Spurgeon and Hopkin (1996) found that zinc was the most toxic trace element for six different earthworm species living close to smelting works at Avonmouth (southwest England) but *L. castaneus* was the most tolerant and survived in soil containing up to 3627 mg kg^{-1} of Zn. This tolerance to elevated trace element concentrations makes *L. castaneus* a good candidate for genetic studies in urban conditions, such as in the Paris region where mean concentration of Zn varies approximately between 107 and 174 mg kg^{-1} in soils of lawn and woods respectively (Foti et al., 2017). Metal-tolerant populations of *L. castaneus* situated in metal contaminated sites surrounded by non-contaminated areas provide the opportunity to investigate the initial steps in the establishment of differentiated populations under severe selection pressures and strong

demographic bottlenecks (e.g. (Mengoni et al., 2000). These populations growing in contaminated sites are thus valuable models for the study of microevolutionary processes (Macnair, 1987). In that context, our main aims were (i) to develop eight polymorphic microsatellite loci from a microsatellite-enriched genomic library and (ii) to assess their usefulness for genetic studies of *L. castaneus* populations in urban environment by examining their genetic variability in six populations of the Paris region.

2. Material and Methods

All collected specimens of *Lumbricus castaneus* were identified based on external morphology using the key of Bouché (1972). Earthworms were sampled manually in top 10 cm of soil using a solution of Allyl isothiocyanate (AITC) diluted with isopropanol (propan-2-ol, RPE grade, Carlo-Erba) to obtain a 5 g L^{-1} solution (Zaborski, 2003). Earthworms were rinsed and cleaned under tap water and stored in ethanol 96% until use.

For microsatellite-enriched genomic library construction, we maximized the genomic diversity by using eight individuals that were collected in two sites of northern Nord-Pas-de-Calais region, approximately 150 km apart (3 specimens from 50°51’46.119”N/1°50’31.765”E and 5 specimens from 50°26’56.465”N/3°2’56.558”E), and one individual that was collected in southern Poland (50°01’38.0”N/20°15’02.9”E). Total genomic DNA was isolated using the NucleoSpin ® 96 Tissue kit (Macherey Nagel, France). A stoichiometric mixture of the 9 DNAs was performed to produce a microsatellite enriched genomic library which was constructed by the Genoscreen Company (Lille) by coupling multiplex microsatellite enrichment isolation techniques with the 454 GS-FLX Titanium pyrosequencing (Malausa et al., 2011). Enrichment was performed using probes containing the following microsatellite motifs: TG, TC, AAC, AAG, AGG, ACG, ACAT and ACTC. Overall, 11875 sequences containing a single microsatellite motif were found (Accession no.

in the NCBI Short Read Archives SRP154315). This allows the identification of 664 independent microsatellite loci candidates for which amplification primer pairs were designed in silico. Among them, 40 microsatellite loci were chosen in order to diversify the repeated motif (di – tri- and tetra-nucleotides motifs), to maximize the number of repeats of the motif, and to allow multiplexing (i.e. different expected amplicon sizes).

Amplification trials, tests of polymorphism and subsequent analyses of genetic structure were performed on six populations from the Paris region. Four of these populations, namely the Buttes-Chaumont park (BCH), the Butte du chapeau rouge square (BCR), the Belleville park (BEL), and the Compans square (COM) are geographically close to each other (less than 2 kilometers) and situated in the 19^{th} and 20^{th} districts northeast of Paris. The remaining two populations are from the park of Pierre-Fitte in Villeneuve-le Roi city (VLR 1 and VLR2), approximately 16 km away from Paris (Table 1). The population from Buttes-Chaumont Park (BCH) was sampled over two consecutive years (2015 and 2016). Genomic DNA was extracted using the same methodology as previously described. Tests of amplification patterns were carried out on a panel of 8 individuals from these six populations. After basic PCR amplifications, PCR products were checked using gel electrophoresis in 2% agarose gels. Of the 40 primer pairs, 28 had ambiguous banding patterns. The 12 markers that proved to be monolocus were analysed separately using labelled forward primers on an ABI 3130 xl Genetic Analyzer (Applied Biosystems, genomic platform of IMRB, Mondor Institute, Créteil). Fragment lengths were manually checked on chromatograms to detect inconsistencies and genotypes were scored against the GeneScan-500 Liz Size Standard (Applied Biosystems) using GeneMapper 5 software (Applied Biosystems).

Of the 12 primer pairs, 4 were discarded because of ambiguous banding patterns or monomorphism. The eight microsatellite markers (Table 2 and Table S1 of the supplementary data) showing both consistent amplification results and genetic variation among tested

individuals were used to develop one multiplex set to be carried out in 12.5 µl reactions using 10 ng of DNA and the Qiagen ® Multiplex Kit according to the manufacturer's protocol.

In order to quantify genotyping errors, 23 individuals of the VLR2 population were genotyped twice (genotypes were produced from two different DNA extracts from the same individuals). Error rates were quantified by computing the mean error rate per allele (e_a) and the mean error rate per locus (e_l, Pompanon et al., 2005). Moreover, null allele frequencies were estimated using the software FreeNa (Chapuis and Estoup, 2007) with a number of replicates fixed to 1000.

From the entire data set (153 individuals), the number of alleles, the observed and expected heterozygosities and the estimator of the inbreeding coefficient (F_{is}) were calculated using the program Genetix V. 4.05 (Belkhir et al., 2004). The allelic richness standardized for sample size (A_r N = 14) was computed using the program Fstat V. 2.9.3.2 (Goudet, 2000). Exact test for genotypic disequilibrium and deviations from Hardy-Weinberg equilibrium were computed using Genepop V. 4.4 (Rousset, 2008). In order to investigate the importance of genetic exchanges among populations, exact tests of allelic differentiation were carried out using Genepop V. 4.4. To adjust for multiple comparisons, sequential Bonferroni correction was used. Moreover, Weir and Cockerham's (1984) estimator of the fixation index F_{st} was calculated with Genepop V. 4.4. The program BAPS V. 6 (Corander and Marttinen, 2006; Corander et al., 2008) was used to carry out an analysis of spatial clustering of groups. We ran 10 replicates for k = 6 where k is the maximum number of genetically divergent groups (populations).

The 658 bp fragment of the COI gene proposed as a standard DNA barcode for animals (Hebert et al., 2003) was sequenced using LCO1490F and HCO2192R primers (Folmer et al., 1994) for six individuals used for the genomic library (DNA of 3 individuals could not be amplified) and for a sub-sample of each study population (54 individuals, details

in Table 1, GenBank Accession Numbers MH684623 – MH684682). DNA sequencing was performed in both directions by Eurofins Genomics company and sequences were manually aligned using the BioEdit program (Hall, 1999). Taxonomy of the samples was checked using the identification engine of BOLD (Barcode of Life Data Systems – https//www.boldsystems.org/). Using the software MEGA6, we constructed a phylogenetic tree by neighbour joining with p-distance and 1000 bootstraps. To root the tree, we used a sequence of *Lumbricus herculeus* (GenBank Accession Number MH638308). Moreover, as Porco et al. (2018) revealed the existence of three lineages within the morphospecies *L. castaneus*, one sequence per lineage was added to the dataset (GenBank Accession Numbers FJ937284, MF121706, GU206163). The neighbor-joining phylogenetic tree revealed that all the study specimens belong to the L1 lineage (Fig S1, supplementary data). Cross-amplification of the new *L. castaneus* microsatellites across mitochondrial lineages could thus not be tested.

3. Results and Discussion

In total eight polymorphic microsatellite markers were developed for *L. castaneus.* None of these loci showed significant linkage disequilibrium. Over the whole data set (N = 153), allele number per locus ranged from 7 to 29 (Table 2). Null alleles are likely to occur at all loci but not at the same frequency according to the population. Null allele frequencies ranged from 0 to 0.293 (Table 2). The presence of null alleles in population genetics studies using microsatellite markers have been frequently reported for a wide range of taxa (Dakin and Avise, 2004), and in earthworms in particular (Dupont et al., 2011; Harper et al., 2006; Novo et al., 2008; Souleman et al., 2016; Velavan et al., 2007). Several statistical corrections for null allele are available in population genetics software to overcome this problem (e.g. Chapuis and Estoup, 2007). Apart from the risk of null alleles, the reliability of the new

189 microsatellite markers was good. Indeed, the risk of genotyping error was null for the loci
190 LC02, LC05, LC10, LC27 and LC36. The higher values of mean error rate per allele and
191 mean error rate per locus were obtained for the locus LC33 (0.05 and 0.10 respectively).

192 We further evaluated the suitability of the use of these microsatellites for genetic
193 studies of populations exposed to environmental stressors by studying the genetic diversity
194 and the genetic structure of the six populations we sampled in the Paris region. Summary
195 statistics describing the genetic variation within populations are given Table 1. Significant
196 departure from Hardy-Weinberg equilibrium caused by a deficiency in heterozygotes was
197 observed in all sites with F_{is} estimator ranging from 0.132 to 0.366 (Table 1). This
198 heterozygote deficit could be due to null alleles, but also to a Wahlund effect (i.e. cryptic
199 population structure within site) or to inbreeding. Additional population genetic studies within
200 each site are necessary to further interpret these results. The allelic richness and the expected
201 heterozygosity were found to vary slightly depending on the population (from 6.06 to 7.70
202 and 0.596 to 0.705, respectively). The highest values of the genetic diversity estimators were
203 observed in both populations of Villeneuve-le-Roi (VLR1 and VLR2) which were sampled in
204 an urban wasteland of 8 ha. The sample from the Compans square (COM) was the population
205 with the lowest expected heterozygosity and the second lowest allelic richness (Table 1).
206 Because of the spatial isolation and the restricted size of the Compans square (2444 m^2), we
207 suggest that the low level of genetic diversity observed in COM may be attributed to
208 important genetic drift effects due to small effective population size.

209 Among populations, a low but significant genetic structure was revealed at the level of
210 the whole study (F_{st} = 0.033, P<0.001). As expected, the BCH15 and BCH16 samples,
211 collected at exactly the same locality of the Buttes-Chaumont Park but over two consecutive
212 years, were not genetically differentiated and, thus, are indeed from the same population
213 (Table 3). The highest pairwise F_{st} values were obtained between the COM population and all

214 the other populations (Table 3). This important genetic differentiation of the COM population
215 confirms the suspected founder effect indicated by the low level of genetic diversity observed,
216 and is probably the result of genetic drift. This singularity was confirmed by the spatial
217 clustering of groups that revealed 3 clearly differentiated genetic clusters with one of the
218 cluster corresponding only to COM (Fig. 1). The populations of the Buttes-Chaumont and
219 Butte-du-Chapeau-Rouge Parks, both situated in the 19th district northeast of Paris, clustered
220 as expected with respect to their geographical proximity (Fig. 1). For the same reason, the
221 cluster found for both populations from Villeneuve-le-Roi (VLR1 and VLR2) was predictable
222 but the fact that the Belleville population (BEL) also belongs to this third cluster was
223 surprising. This unexpected pattern of genetic structure highlights the probable role of human
224 transfer of earthworms between sites as had been already shown for instance for the endogeic
225 species *Allolobophora chlorotica* and *Aporrectodea icterica* (Dupont et al., 2015; Dupont et
226 al., 2017; Torres-Leguizamon et al., 2014).

227 To conclude, the eight newly developed microsatellite markers for *L. castaneus* were
228 highly polymorphic and allowed to differentiate populations even at a fine scale (< 2 km) in
229 an urban context. Therefore, we believe that they will be of great utility to investigate the
230 consequences of environmental stressors, such as trace elements contamination, on neutral
231 genetic variation of *L. castaneus* populations.

Acknowledgements

This work was funded by the French National Research Agency (ANR) as a part of the project Ecoville No. ANR-14-CE22-0021, by the town council of Paris through the Paris2030 calls for projects (Soil biodiversity in green spaces of Paris project) and by the Ile de France region through the “Partenariats institutions-citoyens pour la recherche et l’innovation” (Picri) calls for projects (ReFUJ project). The development of the microsatellite-enriched genomic library was funded by the “Institut de Recherches Pluridisciplinaires en Sciences de l'Environnement” (IREPSE, Lille 1 University).

References

Belkhir, K., Borsa, P., Goudet, J., Chikhi, L., Bonhomme, F., 2004. GENETIX 4.05, logiciel sous Windows pour la génétique des populations. Laboratoire Génome, Population, Interactions, CNRS UMR 5000, Université Montpellier II, Montpellier (France).

Bouché, M.B., 1972. Lombriciens de France. Ecologie et systématique. INRA, Paris.

Chapuis, M.P., Estoup, A., 2007. Microsatellite null alleles and estimation of population differentiation. Mol. Biol. Evol. 24, 621-631.

Corander, J., Marttinen, P., 2006. Bayesian identification of admixture events using multilocus molecular markers. Mol. Ecol. 15, 2833-2843.

Corander, J., Sirén, J., Arjas, E., 2008. Bayesian spatial modelling of genetic population structure. CompStat 23, 111-129.

Cunha, L., Thornber, A., Kille, P., Morgan, A.J., Novo, M., 2017. A large set of microsatellites for the highly invasive earthworm *Amynthas corticis* predicted from low coverage genomes. Appl. Soil Ecol. 119, 152-155.

Dakin, E.E., Avise, J.C., 2004. Microsatellite null alleles in parentage analysis. Heredity 93, 504-509.

Dupont, L., Grésille, Y., Richard, B., Decaëns, T., Mathieu, J., 2015. Fine-scale spatial genetic structure and dispersal constraints in two earthworm species. Biol. J. Linn. Soc. 114, 335-347.

Dupont, L., Lazrek, F., Porco, D., King, R.A., Rougerie, R., Symondson, W.O.C., Livet, A., Richard, B., Decaens, T., Butt, K.R., Mathieu, J., 2011. New insight into the genetic structure of the *Allolobophora chlorotica* aggregate in Europe using microsatellite and mitochondrial data. Pedobiologia 54, 217-224.

Dupont, L., Torres-Leguizamon, M., Rene-Corail, P., Mathieu, J., 2017. Landscape features impact connectivity between soil populations: a comparative study of gene flow in earthworms. Mol. Ecol. 26, 3128-3140.

Folmer, O., Black, M., Hoeh, W., Lutz, R., Vrijenhoek, R., 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Bio. & Biotech. 3, 294-299.

Foti, L., Dubs, F., Gignoux, J., Lata, J.-C., Lerch, T.Z., Mathieu, J., Nold, F., Nunan, N., Raynaud, X., Abbadie, L., Barot, S., 2017. Trace element concentrations along a gradient of urban pressure in forest and lawn soils of the Paris region (France). Sci. Total Environ. 598, 938-948.

Gonzalez, E.G., Zardoya, R., 2007. Relative role of life-history traits and historical factors in shaping genetic population structure of sardines (*Sardina pilchardus*). BMC Evol. Biol. 7, 197.

Goudet, J., 2000. FSTAT, a program to estimate and test gene diversities and fixation indices, version 2.9.1 ed. Available from https://www2.unil.ch/popgen/softwares/fstat.htm. (last access date 06/11/18)

Grumiaux, F., Demuynck, S., Pernin, C., Lepretre, A., 2015. Earthworm populations of highly metal-contaminated soils restored by fly ash-aided phytostabilisation. Ecotox. Environ. Safe. 113, 183-190.

Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41, 95-98.

Harper, G.L., Cesarini, S., Casey, S.P., Morgan, A.J., Kille, P., Bruford, M.W., 2006. Microsatellite markers for the earthworm *Lumbricus rubellus*. Mol. Ecol. Notes 6, 325-327.

Hartl, D.L., Clark, A.G., 2007. Principles of population genetics, 4 ed. Sinauer, Sunderland, MA.

Hebert, P.D.N., Cywinska, A., Ball, S.L., Dewaard, J.R., 2003. Biological identifications through DNA barcodes. P. Roy. Soc. B. - Biol. Sci. 270, 313-321.

Johnson, M.T.J., Munshi-South, J., 2017. Evolution of life in urban environments. Science 358.

Kirk, H., Freeland, J.R., 2011. Applications and Implications of Neutral versus Non-neutral Markers in Molecular Ecology. Int. J. Mol. Sci. 12, 3966-3988.

Loveless, M.D., Hamrick, J.L., 1984. Ecological determinants of genetic structure in plant populations Annu. Rev. Ecol. Syst. 15, 65-95.

Macnair, M.R., 1987. Heavy metal tolerance in plants: a model evolutionary system. Trends Ecol. Evol. 2, 354–359.

Malausa, T., Gilles, A., Meglecz, E., Blanquart, H., Duthoy, S., Costedoat, C., Dubut, V., Pech, N., Castagnone-Sereno, P., Délye, C., Feau, N., Frey, P., Gauthier, P., Guillemau, T., Hazard, L., Le Corre, V., Lung-Escarmant, B., Malé, P.-J.G., Ferreira, S., Martin, J.F., 2011. High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. Mol. Ecol. Resour. 11, 638-644.

Mengoni, A., Gonnelli, C., Galardi, F., Gabbrielli, R., Bazzicalupo, M., 2000. Genetic diversity and heavy metal tolerance in populations of *Silene paradoxa* L. (Caryophyllaceae): a random amplified polymorphic DNA analysis. Mol. Ecol. 9, 1319-1324.

Mussali-Galante, P., Tovar-Sanchez, E., Valverde, M., Rojas, E., 2014. Genetic structure and diversity of animal populations exposed to metal pollution, in: Whitacre, D. (Ed.), Reviews of environmental contamination and toxicology. Springer International Publishing Switzerland, pp. 79-106.

Novo, M., Velavan, T.P., Almodovar, A., Schulenburg, H., Diaz-Cosin, D.J., Michiels, N.K., 2008. Microsatellite markers for the drought-resistant earthworm *Hormogaster elisae*. Mol. Ecol. Resour. 8, 901-903.

Pickett, S.T.A., Cadenasso, M.L., Grove, J.M., Boone, C.G., Groffman, P.M., Irwin, E., Kaushal, S.S., Marshall, V., McGrath, B.P., Nilon, C.H., Pouyat, R.V., Szlavecz, K., Troy, A., Warren, P., 2011. Urban ecological systems: Scientific foundations and a decade of progress. J. Environ. Manage. 92, 331-362.

Pompanon, F., Bonin, A., Bellemain, E., Taberlet, P., 2005. Genotyping errors: Causes, consequences and solutions. Nat. Rev. Genet. 6, 847-859.

Porco, D., Chang, C.-H., Dupont, L., James, S., Richard, B., Decaëns, T., 2018. A reference library of DNA barcodes for the earthworms from Upper Normandy: Biodiversity assessment, new records, potential cases of cryptic diversity and ongoing speciation. Appl. Soil. Ecol. 124, 362-371.

Rousset, F., 2008. GENEPOP ' 007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol. Ecol. Resour. 8, 103-106.

Somers, C.M., Neudorf, K., Jones, K.L., Lance, S.L., 2011. Novel microsatellite loci for the compost earthworm *Eisenia fetida*: A genetic comparison of three North American vermiculture stocks. Pedobiologia 54, 111-117.

Souleman, D., Grumiaux, F., Frerot, H., Vandenbulcke, F., Pauwels, M., 2016. Isolation and characterization of eight polymorphic microsatellites markers for the earthworm *Lumbricus terrestris*. Eur. J. Soil Biol. 74, 76-80.

Spurgeon, D.J., Hopkin, S.P., 1996. The effects of metal contamination on earthworm populations around a smelting works: Quantifying species effects. Appl. Soil. Ecol. 4, 147-160.

Strunk, H., Hochkirch, A., Veith, M., Hankeln, T., Emmerling, C., 2012. Isolation and characterization of eleven polymorphic microsatellite markers for the earthworm *Aporrectodea longa* (Ude). Eur. J. Soil Biol. 48, 56-58.

Takami, Y., Koshio, C., Ishii, M., Fujii, H., Hidaka, T., Shimizu, I., 2004. Genetic diversity and structure of urban populations of Pieris butterflies assessed using amplified fragment length polymorphism. Mol. Ecol. 13, 245-258.

Torres-Leguizamon, M., Mathieu, J., Decaëns, T., Dupont, L., 2014. Genetic structure of earthworm populations at a regional scale: inferences from mitochondrial and microsatellite molecular markers in *Aporrectodea icterica* (Savigny 1826). Plos One 9, e101597.

Torres-Leguizamon, M., Mathieu, J., Livet, A., Dupont, L., 2012. Isolation of polymorphic microsatellite markers in *Aporrectodea icterica* (Savigny 1826). Soil Biol. Biochem. 51, 16-19.

van Straalen, N.M., Timmermans, M.J.T.N., 2002. Genetic variation in toxicant-stressed populations: An evaluation of the "genetic erosion" hypothesis. Hum Ecol Risk Assess 8, 983-1002.

Velavan, T.P., Schulenburg, H., Michiels, N., 2007. Development and characterization of novel microsatellite markers for the common earthworm (*Lumbricus terrestris* L). Mol. Ecol. Notes 7, 1060-1062.

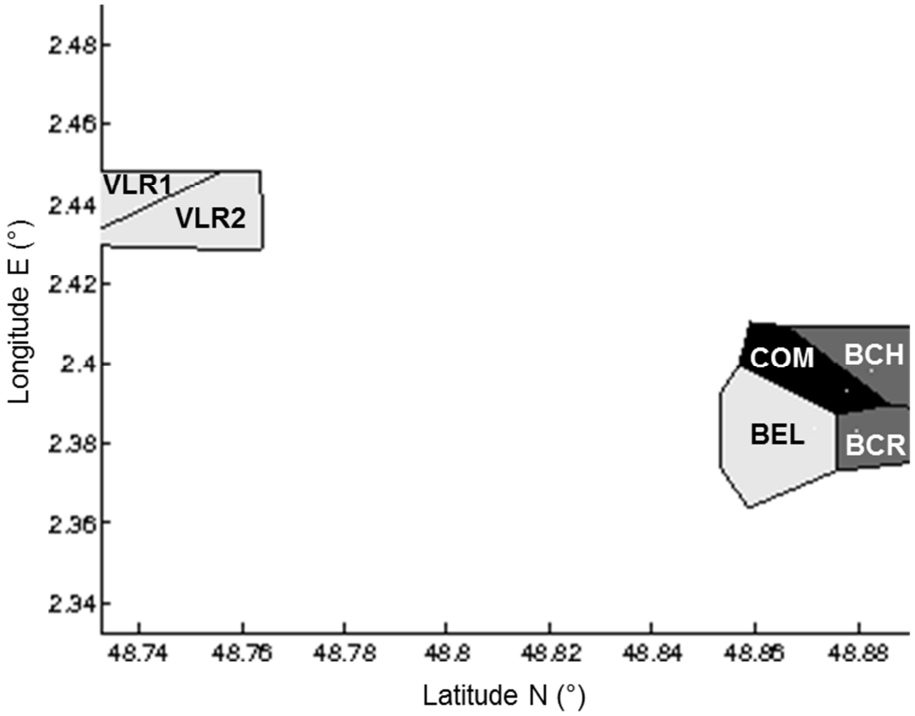
Weir, B.S., Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure. Evolution 38, 1358-1370.

Zaborski, E.R., 2003. Allyl isothiocyanate: an alternative chemical expellant for sampling earthworms. Appl. Soil Ecol. 22, 87-95.

Figure legends

Figure 1

Spatial clustering analysis using a Voronoi tessellation in BAPS. The optimal partition value was for K = 3. Different greyscales represent the three genetic clusters. Each cell of the tessellation corresponds to the physical neighbourhood of an observed data point.


2.48
2.46
2.44
2.42
2.4
2.38
2.36
2.34
Longitude E (°)
VLR1
VLR2
COM
BCH
BEL
BCR
48.74
48.76
48.78
48.8
48.82
48.84
48.86
48.88
Latitude N (°)

350

351

352 **Table 1:** Characteristics of the six study samples, with geographical location (locality, latitude and longitude), year of collection (date), sample
353 size (N), number of COI sequences (N_{seq}), allelic richness after rarefaction to a population size of 14 (A_r), observed heterozygosity (H_o), expected
354 heterozygosity (H_e), and fixation index F_{is} which was always associated to a highly significant probability of the exact test for deviations from
355 Hardy Weinberg expectations (*).

Locality	**Latitude N (°)**	**Longitude E (°)**	**Sample**	**Date**	***N***	N_{seq}	$\boldsymbol{A_r}$	$\boldsymbol{H_o}$	$\boldsymbol{H_e}$	$\boldsymbol{F_{is}}$
Buttes-Chaumont park- Paris	48.879978	2.383333	BCH15	2015	19	19	6.06	0.526	0.603	0.132*
			BCH16	2016	25	0	6.64	0.487	0.651	0.257*
Belleville park- Paris	48.871901	2.383527	BEL	2016	26	2	7.26	0.463	0.643	0.271*
Butte du chapeau rouge square - Paris	48.882608	2.398003	BCR	2016	24	2	7.05	0.466	0.646	0.284*
Compans square - Paris	48.877923	2.393220	COM	2016	18	2	6.13	0.424	0.596	0.298*
Site 1 – Pierre-Fitte –Villeneuve - le - Roi	48.7399292	2.4391225	VLR1	2015	18	18	7.56	0.450	0.705	0.366*
Site 2 – Pierre-Fitte –Villeneuve - le - Roi	48.7406666	2.4379152	VLR2	2016	23	11	7.70	0.570	0.676	0.160*

356

357

358 **Table 2:** Characteristics of eight microsatellite loci isolated from *Lumbricus castaneus*, with locus name, repeat array, primer sequences (F:
359 forward primer, R: reverse primer), allele size range (ASR), total number of allele over the whole data set (N_{tall}), range of the estimation of null
360 allele frequency in the study samples (Null), mean error rate per allele (e_a) and mean error rate per locus (e_l).

Locus ID	Repeat array	Primer sequence (5' – 3')	ASR (bp)	N_{tall}	Null	e_a	e_l
LC02	$(TC)_{10}$	F : AT565-GTTATCGGTGCCTTCCATGT R : GTGACCATCCTTTGATTGCC	093 - 115	7	0.000 – 0.195	0.00	0.00
LC05	$(GT)_8$	F : FAM-ATACGTGTCTGGAAGGGTGG R : CTCCGTTGTACCGCTGTGTA	120 - 170	7	0.000 – 0.135	0.00	0.00
LC10	$(AATC)_{10}$	F : AT550-GAAGCCATACTGCCACTGGT R : GCTTTACTTGTTGCCATTCTGTT	141 - 208	19	0.029 – 0.208	0.00	0.00
LC16	$(GACA)_8$	F : AT565-AACAACCGAAAACTGCAAGTC R : CCTCAAGGCAAGCTCAGGTA	225 - 344	29	0.053 – 0.293	0.01	0.05
LC18	$(TCAA)_{11}$	F : FAM-GAGGCCAAAACCCATCACTA R : TGTCTTTCAGGGCAGAAGTG	223 - 286	22	0.021 – 0.253	0.01	0.05
LC27	$(CT)_5$	F : AT550-GCCATTTTGTCATTCGGTCT R : AACATAACGCAACACCCACA	237 - 266	13	0.000 – 0.139	0.00	0.00
LC33	$(TTA)_7$	F : YAKIMAYELLOW-CTCAGAATTGACGTTGGCAG R : ATCATGCATGGACAGCGTT	110 - 132	10	0.063 – 0.290	0.05	0.10
LC36	$(GAAT)_6$	F : YAKIMAYELLOW-GACGTAACGCAATGTGATGG R : TCCAGGCAGAGTTATTTCGG	151 - 215	19	0.000 – 0.264	0.00	0.00

361

362 **Table 3** Pairwise multi-loci F_{ST} estimates between different populations. Significant exact

363 tests of allelic differentiation after sequential Bonferroni correction are indicated in bold.

BCH16	0.004					
BEL	**0.054**	**0.037**				
BCR	**0.025**	**0.018**	**0.014**			
COM	**0.069**	**0.069**	**0.050**	**0.040**		
VLR1	**0.030**	**0.021**	**0.016**	**0.007**	**0.039**	
VLR2	**0.048**	**0.036**	**0.031**	**0.032**	**0.072**	**0.004**
	BCH15	BCH16	BEL	BCR	COM	VLR1

364

365