

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/101404/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Cunha, Luis , Thornber, Alice, Kille, Peter , Morgan, John and Novo Rodriguez, Marta 2017. A large set of microsatellites for the highly invasive earthworm *Amyntas corticis* predicted from low coverage genomes. *Applied Soil Ecology* 119 , pp. 152-155. 10.1016/j.apsoil.2017.05.029

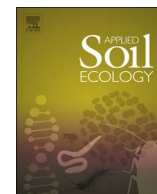
Publishers page: <https://doi.org/10.1016/j.apsoil.2017.05.029>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.





Short communication

A large set of microsatellites for the highly invasive earthworm *Amyntas corticis* predicted from low coverage genomesL. Cunha^{a,b,1}, A. Thornber^{a,1}, P. Kille^a, A.J. Morgan^{a,3}, M. Novo^{a,*,2}^a Organisms and Environment Division, School of Biosciences, Cardiff University, Cardiff, CF103AT, UK^b Embrapa Florestas, Estrada da Ribeira, Colombo, Paraná, Brazil

ARTICLE INFO

Keywords:

Earthworms
Invasive species
Microsatellite markers
Mitochondrial lineages
Bioinformatics prediction

ABSTRACT

Invasive species can significantly affect local biodiversity and create important challenges for conservation. They usually present an outstanding plasticity that permits the adaptation to the new environments. Understanding their genetic background is fundamental to better comprehend invasion dynamics and elaborate proper management plans as well to infer population and evolutionary patterns. Here, we present a reasonable set of tools for the study of a highly invasive earthworm, the megascolecid *Amyntas corticis*. We designed *in silico* a large set of primers targeting microsatellite regions (ca. 9400) from two low coverage genomes presented here. This study provides 154 high quality primer pairs targeting polymorphic repeats conserved in two *Amyntas corticis* mitochondrial lineages. From this dataset, a set of primer pairs (15) was validated by polymerase chain reaction with 86% consistent amplification, confirming the accuracy of the *in silico* prediction. Nine of the primer pairs tested were selected for population genetics and presented polymorphism in the studied populations, thus showing promising potential for future studies of this global invasive species. The nuclear markers used in this study appear to recapitulate and complement the mitochondrial relationships found in a previous study. Interestingly, all genotyped individuals showed at least one triploid *locus* profile among the tested *loci*, which may be evidence of polyploidy associated to their life history, in particular to asexual reproduction by parthenogenesis.

Biological invasions are a major concern within the present globalisation context and invasive species pose a challenge for conservation. Alien terrestrial invertebrates have significant ecological and economic impacts on ecosystem services (Vila et al., 2010) and introduced earthworms can modify soil structure and nutrient dynamics (Hendrix et al., 2008) as well as affect native populations (Hendrix et al., 2006). Genetic studies are needed in order to understand invasion patterns and create management policies. *Amyntas corticis* is a highly invasive earthworm with Eastern Palearctic origin that has shown to tolerate a wide range of environmental factors (Fragoso et al., 1999) and has been introduced worldwide (Hendrix et al., 2008). In a previous study, three mitochondrial lineages with different environmental preferences were detected within Azores Islands (Portugal), two of which were introduced and one with unknown origin (Novo et al., 2015). Here, we present a large set of microsatellite regions (9380) generated from genomic data, 154 of which are conserved across lineages and potentially polymorphic. By testing a subset, we showed that the *in silico*

prediction could work with 86% efficiency (13/15). These microsatellites would be a valuable tool for future studies on this species in order to better understand its invasion dynamics. Different lineages of invasive earthworm species are known to present different environmental preferences (Novo et al., 2015: *A. corticis* and *A. gracilis*). Therefore, genotyping should be used in order to unravel the invasion patterns in such species, since the study at species level overlooks important information. After a basic understanding of such lineages and associated environmental conditions, the gathered information could serve for prediction of future habitat colonization and even bioindication of certain ecological characteristics of the environment. This allows inclusion of mitigation/prevention measures in conservation management plans (e.g., soil import, country blacklist, etc) depending on species biological invasion potential. Furthermore, such measures are crucial for avoidance of further invasion and potential disruption of eco-functional and demographic processes on local native earthworm populations, in particular for endemic regions and delicate ecosystems.

* Corresponding author.

E-mail address: martanovorodriguez@gmail.com (M. Novo).¹ The first two authors contributed equally to this work.² Current address: Environmental Toxicology and Biology Group, Dpto Física Matemática y de Fluidos. Senda del Rey 9, Facultad de Ciencias, UNED, 28040, Madrid, Spain.³ This author was deceased during the preparation of this manuscript.

Microsatellites allow the study of genetic diversity and its spatial structure at a finer scale than mitochondrial markers in earthworms (Novo et al., 2010), for whose distribution environmental soil micro-variables are known to be of great importance (Joschko et al., 2006; Palm et al., 2013). They also permit the evaluation of admixture of lineages and ploidy level of specimens (e.g. Hänfling et al., 2005; do Prado et al., 2017).

We assembled genomic information of *A. corticis* from two of the lineages identified in Novo et al. (2015) in Azores (Portugal), namely lineages A and C. Both type specimens used in this study were obtained from pasture fields (Lineage A: 37.76604, –25.62494; Lineage C, 37.76369, –25.53497) located on São Miguel Island (Azores archipelago, Portugal), and the tissue sample was deposited in Cardiff University. DNA was sheared, Illumina adapters were ligated, libraries were controlled for quality, normalized, pooled and ran on a HiSeq 2500, in rapid run mode (paired-end 150 bp). Contigs used for the microsatellite mining were *de novo* assembled using SOAPdenovo v.2 (Luo et al., 2012) using short-reads. These draft genomes are available from the European Nucleotide Archive (ENA) under the study number: PRJEB20839. Lineage A genome assembly (sample accession: ERS1735863) yielded a 1.2 Gb total base pair reconstruction, with a n50 of ~600 bp and ~3 million contigs. Lineage C assembly (sample accession: ERS1735864) showed a smaller size reconstruction yielding a 0.85 Gb of total base pair reconstruction, n50 of ~400 bp and ~2.9 million contigs. The genome draft assembly reconstruction sizes were higher 2.1 times for lineage A and 1.5 times for lineage C when compared to the estimated genome size using flow cytometry measurements of coelomic cells (570 Mb, data not shown). Higher *in silico* genome size than estimated haplome is not uncommon and is usually associated to high genome heterozygosity that makes the assembly process more fragmented (Zwarycz et al., 2016). Microsatellite regions were identified and primers were designed using PrimerPipeline package (<http://www.scrufster.com/primerpipeline/> [last date accessed: 10.03.17]) as implemented in Macdonald et al. (2016) which incorporates the microsatellite mining tool, MISA (Varshney et al., 2002) and Primer3 (Untergasser et al., 2012). Tables with the recovered microsatellite regions for the two studied lineages (4677 for lineage A, and 4703 for lineage C) and potential primers are provided in Supplementary files 1 and 2. Conserved cross-lineage primers were assessed *in silico* using iPCRESS (European Bioinformatics Institute, Exonerate package, version 1.2.0) using primers developed in lineage A as input primers to test in lineage C draft genome. This provided 3038 cross-lineage primers pairs (with no base mismatch allowed). Posterior analysis was restricted to primer pairs that would amplify a unique sequence in a unique genomic contig, conserved between the genomes (a total of 159 primer pairs). *In silico* evaluation (difference between amplifiable products in genomes of both lineages) showed that 97% of these primer pairs could potentially amplify polymorphic loci (154 primer pairs, Supplementary file 3; sequences from both lineages are shown in Supplementary files 4 for lineage A and 5 for lineage C). From this dataset, 15 primer pairs were chosen for assessment of polymorphism in loci, using polymerase chain reaction. Out of the 15 primer pairs, 13 showed consistent amplification (two showed inconsistent multiple band patterns) and polymorphism in preliminary analyses of five specimens representing three mitochondrial lineages and including the individuals from which the low coverage genomes were sequenced. Nine primer pairs were further amplified for a larger set of specimens (see below). The forward primers were tailed with the M13 sequence (TGTAACAACGACGCCAGT) at their 5' end and the universal fluorescent-labelled M13 primer was added to the reaction as well (6FAM). See description of the method in Schuelke (2000).

The utility of the different loci as genetic markers was tested in a panel of 34 individuals coming from different populations and mitochondrial lineages collected mainly in São Miguel Island (Azores, Portugal) but also in Spain, US and China (see Novo et al., 2015; Supplementary file 6 and Fig. 1A). DNA was extracted from 25 mg of

collected tissue samples using the DNeasy® Blood & Tissue kit (Qiagen) adding 70 µl of Elution Buffer twice. PCR amplifications were performed in a Veriti® 96-Well Thermal Cycler (Applied Biosystems) with 10–20 ng of genomic DNA. Reactions included a final volume of 20 µl and 2X GoTaq® Reaction Buffer, 2.5 mM MgCl₂, 0.5 mM dNTPs, 0.5 µM of M-13 universal primers labelled with a fluorescent dye attached to the 5' end, 0.25 µM of M13-tagged forward primer, 0.5 µM of reverse primer and 1U of GoTaq® Polymerase (Promega). Two different cycles (A and B) were used for PCR depending on primers and sample (Supplementary file 6), and included 94 °C (5 min), 30 cycles of [94 °C (30 s); 61 °C (15 s) for A and 58 °C (30 s) for B; 72 °C (15 s)], 10 cycles of [94 °C (15 s); 53 °C (15 s); 72 °C (15 s)] with a final extension of 10 min at 72 °C for A and 7 min for B. Amplified products were first checked on a 1.5% agarose gel and then analysed on the ABI 3730xl DNA sequencer (Applied Biosystems), following the manufacturer's instructions. Resolved PCR products were precisely sized using Genescan Rox 500 size standard and scored in Peak Scanner™ Software 2 (Applied Biosystems). In order to account for genotyping errors (Pompanon et al., 2005), the samples were analysed twice in the sequencer using two different dilutions. Scoring of alleles is shown in Supplementary file 6. Details of the nine optimised markers are shown in Table 1. Number of alleles varied between two and sixteen and number of genotypes between two and ten. All the individuals presented three alleles at least in one locus, and therefore this suggests that all the genotyped individuals were triploid. This evidence of polyploidy supports the idea that the genotyped individuals of this species reproduce asexually by parthenogenesis. This mode of reproduction presents fitness advantages that may promote colonization and maintenance of stable populations in new environments due to the capacity of a single organism to establish an entire population (Hughes, 1989; Terhivuo and Saura, 2006; D & az Cos & n et al., 2011). Indeed, Novo et al. (2015) suggested a higher invasiveness of *A. corticis* in Azores when compared to *A. gracilis* (diploid with sexual reproduction). The analysis of such polyploid data is challenging but GenoDive (Meirman and Van Tienderen, 2004) allows for a correction of the unknown dosage of alleles. This correction uses a maximum likelihood method based on random mating within populations, modified from De Silva et al. (2005). Heterozygosity was calculated per locus and is shown in Table 1. Since individuals from outside Azores were just included for testing the amplification success, and the numbers were low, we focused for further analyses just on the Azores individuals. In order to test genetic structure and compare it with the information given by mitochondrial markers (Fig. 1B) we performed a Principal Component Analysis (PCA) in GenoDive. The results revealed that genetic structure was not dependent on population origin but on mitochondrial lineage (Fig. 1C). G'st (Nei, 1987) were also calculated in order to show differentiation among those lineages, resulting in values of 0.394 between lineages A and B; 0.409 between A and C; 0.266 between B and C. This was further explored with a Bayesian analysis of population structure performed in Structure v2.3.4 (Pritchard et al., 2000). An admixture model was used allowing alpha and lambda to be inferred from the data. Structure was independently run 20 times, testing values of clusters (K) between 1 and 7 and using 100,000 steps of burn-in for the MCMC algorithm and 100,000 steps of data collection phase. Evanno et al. (2005) method was implemented in Structure Harvester v0.6.94 (Earl and vonHoldt, 2012) to infer the best value of K, which was K = 3. Clusters showed that the microsatellite data divided the individuals per mitochondrial lineage independent of population. Several individuals from the lineage B and one individual from lineage A showed admixture with the other two lineages (Fig. 1D).

We have presented a large set (ca. 9400) of microsatellite regions predicted from genomic data coming from two different mitochondrial lineages, a subset of potentially polymorphic regions common in both lineages (154) and the associated primers to use them as molecular markers. We have verified that the *in silico* prediction is accurate since 13 out of the 15 primer pairs tested yielded consistent amplification in

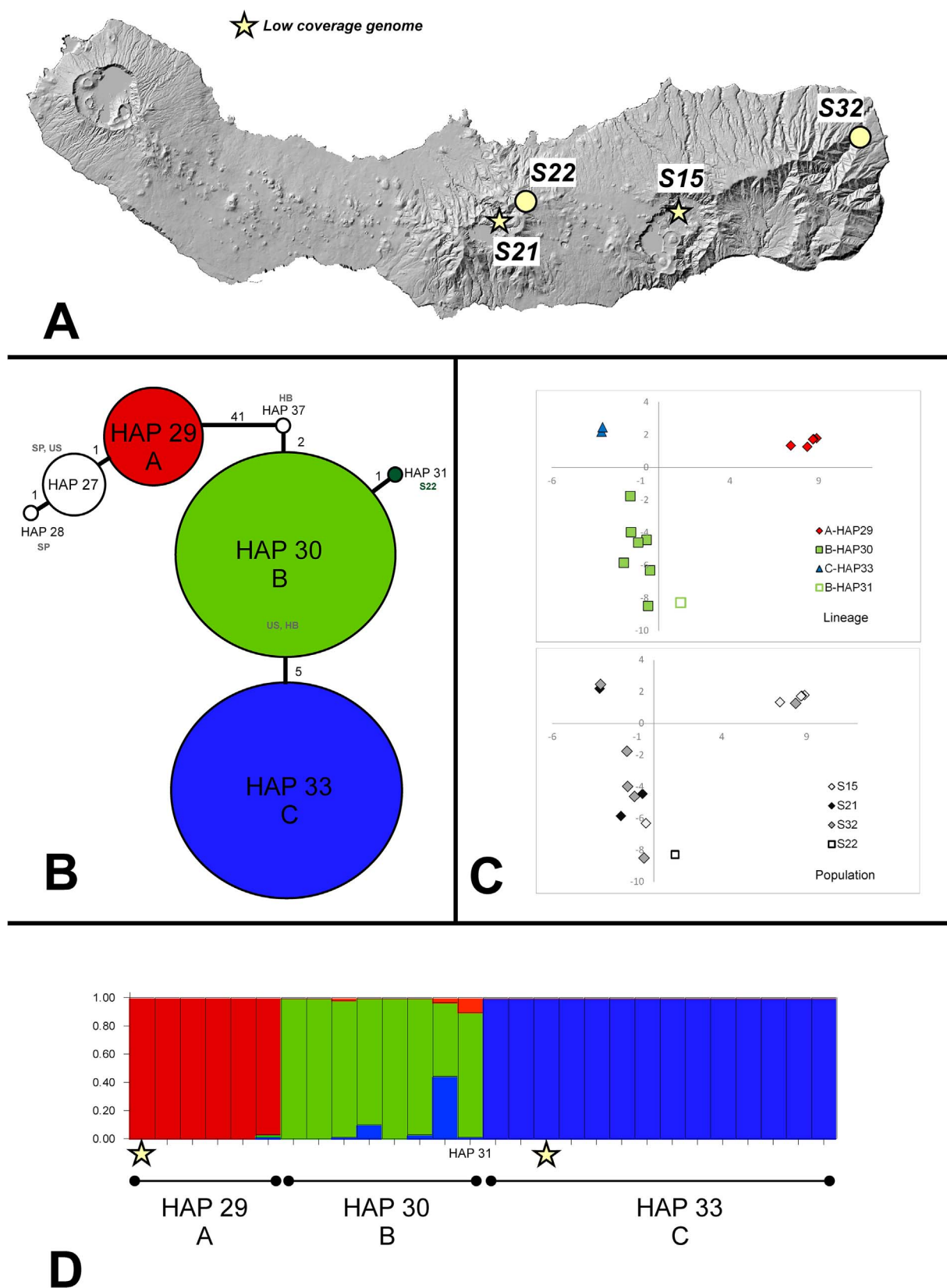


Fig. 1. Results of the analyses for microsatellites of *Amynthes corticis*. **1A:** Sampled populations within São Miguel island (Azores, Portugal). See Supplementary file 6 for number of individuals and tested specimens abroad. The stars represent the populations from which genomic information we designed the primers (lineages A and C). **1B:** Haplotype network from mitochondrial data modified from Novo et al. (2015). Individuals from the coloured haplotypes A, B and C (and HAP-31) were tested for the presented microsatellite regions. **1C:** Results for the PCA analysis performed in GenoDive with the new microsatellite data after correction for allele dosage. The individuals are differentiated by mitochondrial lineage (above) and by population (below). **1D:** Results of analysis in Structure with the new microsatellite data. Each genotype is represented by a vertical bar.

Table 1

Details of the microsatellite markers for *Amyntas corticis* that were further tested in 34 individuals and polymorphic. NG, number of genotypes obtained per locus; NA, number of alleles per locus; NAind, number of alleles per individual; Ht, expected total heterozygosity; H't, corrected total heterozygosity for a bias that stems from sampling a limited number of populations. Whereas GenoDive corrects the allele frequencies for the unknown dosage of polyploid data, Ho would not be based on the corrected frequencies and therefore not shown.

Locus	Repeat motif	Primer sequence (5'–3')	Size range (bp)	NA	NAind	NG	Ht	H't	ENA Accession no.
AC7	(TTAA) ₃	F: GCAGCTCCAACCTGGGTTTG R: CCGTTAAACGCTGGAACCG	268–282	6	1–3	5	0.720685	0.719659	LT714131
AC8	(GA) ₅	F: CCGCCTGCGTTTATTTCGAAG R: CTCCTGTCGACAGGTGGAACC	239–266	6	1–2	7	0.693149	0.697175	LT714132
AC10	(CAG) ₃ (GCCC) ₃	F: CTCGAGACGACGCTGATTCA R: CGTTTCCTTTCAAGGTGGCG	237–241	2	1	2	0.431957	0.460404	LT714133
AC20	(TC) ₂₃	F: ACATTTGGCGTCTTTGTCTCTCA R: AGGGAGACGCCGCTTCTATA	283–305	8	1–2	6	0.758759	0.770657	LT714134
AC21	(AC) ₂₂	F: AGCACCAACAAATGCACACA R: TGGTACTTGGCGTTTCTGGA	271–293	9	2–3	8	0.857476	0.861242	LT714135
AC22	(GA) ₂₇	F: TCGGGTTTACGAAGGAAGA R: AAGAGTGACAGATAGGTGAATAACA	281–345	9	1–3	6	0.797996	0.806006	LT714136
AC24	(AG) ₂₁	F: ATCTCCGCTCTCGAACGAGA R: GCACACGGTCCACAAAGAT	271–291	6	1–3	6	0.804999	0.818324	LT714137
AC28	(ATC) ₁₃	F: TGGACTGTCAATCTCCACATGA R: AGTCATGCTTGGAGGAACCG	210–320	15	1–3	10	0.934463	0.953177	LT714138
AC29	(TTGA) ₁₄	F: TGCCAGAGCATGACTTGTCA R: ACAGTCACAGAGGATGTCAGA	200–310	16	1–3	7	0.909789	0.920824	LT714139

preliminary analyses. Nine of them were further amplified for populations within Azores and some individuals from abroad. We anticipate this could be a very valuable tool for the study of evolutionary and invasion processes of the earthworm *A. corticis*, which has shown great success in new environments when introduced (Novo et al., 2015).

Acknowledgements

Novo M was supported by a Marie Curie Fellowship FP7-PEOPLE-2012-IEF GA-329690 and a Postdoctoral Fellowship (FPDI-2016-16407) from the Spanish Government and Thornber A by a CUROP studentship from Cardiff University. Cunha L was supported by a EU Marie Curie fellowship – MSCA-IF-2014-GF-660378. This study was funded by the grant NERCNE/1026022/1 and we are very grateful to the team in that Project that helped collecting the specimens.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.apsoil.2017.05.029>.

References

- Díaz Cosín, D.J., Novo, M., Fernández, R., 2011. Reproduction of earthworms: sexual selection and parthenogenesis. In: Karaka, A. (Ed.), *Biology of Earthworms*. Springer, Berlin, pp. 69–86.
- De Silva, H.N., Hall, A.J., Rikkerink, E., Fraser, L.G., 2005. Estimation of allele frequencies in polyploids under certain patterns of inheritance. *Heredity* 95, 327–334.
- Earl, D.A., vonHoldt, B.M., 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4, 359–361.
- Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14, 2611–2620.
- Fragoso, C., Kanyonyo, J., Moreno, A., Senapati, B.K., Blanchart, E., Rodríguez, C., 1999. A survey of tropical earthworms: taxonomy, biogeography and environmental plasticity. In: Lavelle, P., Brussaard, L., Hendrix, P.F. (Eds.), *Earthworm Management in Tropical Agroecosystems*. CAB International, Wallingford, U.K., pp. 1–26.
- Hänfling, B., Bolton, P., Harley, M., Carvalho, G.R., 2005. A molecular approach to detect hybridisation between crucian carp (*Carassius carassius*) and non-indigenous carp species (*Carassius spp.* and *Cyprinus carpio*). *Freshwater Biol.* 50, 403–417.
- Hendrix, P.F., Baker, G.H., Callahan, M.A., Damoff, G.A., Fragoso, C., Gonzalez, G., James, S.W., Lachnicht, S.L., Winsome, T., Zou, X., 2006. Invasion of exotic earthworms into ecosystems inhabited by native earthworms. *Biol. Invasions* 8, 1287–1300.
- Hendrix, P.F., Callahan, M.A., Drake, J.M., Huang, C.Y., James, S.W., Snyder, B.A., Zhang, W.X., 2008. Pandora's box contained bait: the global problem of introduced earthworms. *Ann. Rev. Ecol. Syst.* 39, 593–613.
- Hughes, R.N., 1989. *A Functional Biology of Clonal Animals*. Chapman and Hall London, New York, pp. 331 (xii).
- Joschko, M., Fox, C.A., Lentzsch, P., Kiesel, J., Hierold, W., Krück, S., Timmer, J., 2006. Spatial analysis of earthworm biodiversity at the regional scale. *Agric. Ecosyst. Environ.* 112, 367–380.
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu, S.M., Peng, S., Xiaoqian, Z., Li, u G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T.W., Wang, J., 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* 1, 18.
- Macdonald, H.C., Cunha, L., Bruford, M.W., 2016. Development of genomic resources for four potential environmental bioindicator species: *Isoperla grammica*, *Amphinemura sulcirostris*, *Oniscus asellus* and *Baetis rhodani*. *bioRxiv*. <http://bioRxiv.org/content/early/2016/03/29/046227>.
- Meirmans, P.G., Van Tienderen, P.H., 2004. GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Mol. Ecol. Notes* 4, 792–794.
- Nei, M., 1987. *Molecular Evolutionary Genetics*. New York. Columbia University Press.
- Novo, M., Almodóvar, A., Fernández, R., Gutiérrez, M., Díaz Cosín, D.J., 2010. Mate choice of an endogamic earthworm revealed by microsatellite markers. *Pedobiologia* 53, 375–379.
- Novo, M., Cunha, L., Maceda-veiga, A., Talavera, J.A., Hodson, M.E., Spurgeon, D., Bruford, M.W., Morgan, A.J., Kille, P., 2015. Multiple introductions and environmental factors affecting the establishment of invasive species on a volcanic island. *Soil Biol. Biochem.* 85, 89–100.
- Palm, J., van Schaik, N.L., Schröder, B., 2013. Modelling distribution patterns of anecic, epigeic and endogamic earthworms at catchment-scale in agro-ecosystems. *Pedobiologia* 56, 23–31.
- Pompanon, F., Bonin, A., Bellemain, E., Taberlet, P., 2005. Genotyping errors: causes: consequences and solutions. *Nat. Rev. Genet.* 6, 847–859.
- Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Schuelke, M., 2000. An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotechnol.* 18, 233–234.
- Terhivuo, J., Saura, A., 2006. Dispersal and clonal diversity of North-European parthenogenetic earthworms. *Biol. Invasions* 8, 1205–1218.
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M., Rozen, S.G., 2012. Primer3—new capabilities and interfaces. *Nucleic Acids Res.* 40, e115.
- Varshney, R.K., Thiel, T., Stein, N., Langridge, P., Graner, A., 2002. *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cell. Mol. Biol. Lett.* 7, 537–546.
- Vila, M., Basnou, C., Pysek, P., Josefsson, M., Genovesi, P., Gollasch, S., Nentwig, W., Olenin, S., Roques, A., Roy, D., Hulme, P.E., Andriopoulos, P., Arianoutsou, M., Augustin, S., Bacher, S., Bazos, I., Bretagnolle, F., Chiron, F., Clergeau, P., Cochar, P.O., Cocquemont, C., Coeur d'Acier, A., David, M., Delipetrou, P., Desprez-Loustau, M.L., Didziulis, V., Dorkeld, F., Essl, F., Galil, B.S., Gasquez, J., Georgiadi, K., Hejda, M., Jarosik, V., Kark, S., Kokkoris, I., Kuhn, I., Lambdon, P.W., Lopez-Vaamonde, C., Marcer, A., Migeon, A., McLoughlin, M., Minchin, D., Navajas, M., Panov, V.E., Pascal, M., Pergl, J., Perglova, I., Pino, J., Poboljsaj, K., Rabitsch, W., Rasplus, J.Y., Sauvard, D., Scalera, R., Sedlacek, O., Shirley, S., Winter, M., Yannitsaros, A., Yart, A., Zagatti, P., Zikos, A., Partners, D., 2010. How well do we understand the impacts of alien species on ecosystem services? A pan-European, cross-taxa assessment. *Front. Ecol. Environ.* 8, 135–144.
- Zwarycz, A.S., Nossa, C.W., Putnam, N.H., Ryan, J.F., 2016. Timing and scope of genomic expansion within Annelida: evidence from homeoboxes in the genome of the earthworm *Eisenia fetida*. *Genome Biol. Evol.* 8, 271–281.
- do Prado, F.D., Fernández-Cebrián, R., Hashimoto, D.T., Senhorini, J.A., Foresti, F., Martínez, P., Porto-Foresti, P., 2017. Hybridization and genetic introgression patterns between two South American catfish along their sympatric distribution range. *Hydrobiologia* 788, 319–343.