A new cryptic lineage for the earthworm species *Aporrectodea caliginosa* (Lumbricidae)

David Porco[1], Anne Zangerlé[2], Sylvie Hermant[1] & Guy Colling[1]

[1] Luxembourg National Museum of Natural History, 25, rue Münster, L-2160 Luxembourg · david.porco.gm@gmail.com https://orcid.org/0000-0003-1914-1973 · sylvie.hermant@mnhn.lu · guy.colling@mnhn.lu https://orcid.org/0000-0002-8994-9540

[2] ASTA, 16, route d'Esch, L-1470 Luxembourg · anne.zangerle@asta.etat.lu

Porco, D., A. Zangerlé, S. Hermant & G. Colling, 2023. A new cryptic lineage for the earthworm species *Aporrectodea caliginosa* (Lumbricidae). *Bulletin de la Société des naturalistes luxembourgeois* 125 : 119–124. https://doi.org/10.59513/snl.2023.125.119

Published online 10 November 2023 (ISSN 2716-750X), printed volume 12 December 2023 (ISSN 0304-9620).

Abstract. A new cryptic lineage of the complex *A. caliginosa* was found in the grasslands of Luxembourg. Its genetic divergence was of the same level than the one found for the other COI lineages of the complex. The distribution of the pairwise distance comparisons of the DNA barcodes within the complex showed that it could potentially experience gene flow among its components, including this new lineage. This exemplifies that diversity in Lumbricidae can still be uncovered in the well-prospected European zone.

Key words: earthworms, Lumbricidae, DNA barcoding, COI, cryptic diversity.

1. Introduction

Biodiversity is sustaining a major decline in most ecosystems and a significant part of it may be threatened without having been formally described (Scheffers et al. 2012). This lack of knowledge of biological diversity is particularly critical in soil invertebrates (Cameron et al. 2018). In earthworm communities, thanks to the generalization of the DNA barcoding use, it was shown that this lack of knowledge encompasses a significant part of the diversity of the group even in a well-sampled area such as Europe (Porco et al. 2018). This is all the more surprising given that earthworms are among the best-studied soil invertebrates due to their ecological importance. Indeed, these soil ecosystem engineers (Lavelle 1988, Jones et al. 1994) represent the most abundant animal biomass in terrestrial ecosystems (Lavelle & Spain 2001). Thus, their quantitative impact on soil functioning is considerable: positive impact on soil functions such as soil structure, water regulation, nutrient cycling, primary production and remediation (Blouin et al. 2013, van Groenigen et al. 2014). Under the same pedoclimatic conditions, land use and agricultural practices are the main drivers shaping earthworm communities (Riley et al. 2008, Pelosi et al. 2009).

In the habitat mosaics represented by agroecosystems, grasslands are one of the most important elements for earthworm density and diversity (e.g. Rutgers et al. 2009). Here, we sampled both a permanent grassland and an adjacent crop field for *Aporrectodea caliginosa* specimens. This species was specifically targeted as it was already known to be a complex of cryptic lineages (Pérez-Losada et al. 2009, Porco et al. 2013). This work was intended as a pilot study for the survey of Lumbricidae diversity through DNA barcodes in Luxembourg.

2. Material and methods

2.1. Sampling

The sampling took place in October 2020 in a permanent grassland (lat: 49.767993, long: 5.968183, WGS84) and an adjacent crop field (lat: 49.767309, long: 5.968423, WGS84) in Useldange (Luxembourg).

At each site, *Aporrectodea caliginosa* specimens were collected on three points by a combination of mustard extraction (i.e. application of a mustard solution on the soil surface) and a hand sorting of soil blocks of 25 × 25 × 25 cm. Collected specimens were freeze-killed, subsampled for a piece of muscle tissue (1 mm^2) and preserved in liquid nitrogen.

2.2. Molecular analysis

DNA was extracted from the muscle tissue subsamples from five *Aporrectodea caliginosa* specimens with the DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's protocol. The standard DNA barcode region of COI (mitochondrial gene coding for the cytochrome oxidase subunit 1) was amplified using primers LepF1 and LepR1 (Hebert et al. 2004). The amplification reaction contained 7.5 µl of GoTaq mastermix (Promega), 0.3µM of each primer, 1 µl of template DNA and was completed with water to 12 µl. The temperature program was 95 °C for 3 min followed by 35 cycles of denaturation at 95 °C for 30 s, annealing at 60 °C for 30 s, extension at 72 °C for 60 s and a final extension at 72 °C for 5 min. The amplification success was checked on a 2% agarose gel. Amplicons were purified with AMPure XP (Agencourt) and sequenced in both directions with Big Dye terminator V3.1 (Big Dye terminator). A second purification with CleanSEQ (Agencourt) eliminated the excess dye-terminators. The sequencing was run on a Sanger sequencer ABI 3730 XL. Sequences were assembled and aligned with Geneious 8.1.9 (Biomatters). Sequences are publicly available in the BOLD dataset DS-ACALUXL4 and on Genbank (accessions: OQ025632-OQ025636).

2.3. Molecular distance analysis

The sequences produced for *A. caliginosa* from Luxembourg were analyzed with those of others cryptic lineages (L1, L2 and L3) of this species and two other species of the genus *A. cupulifera* and *A. longa* (Porco et al. 2013, Porco et al. 2018) (Genbank accessions: GU013819-GU013824, HM417933, HM417936, HM417946, HM417951, JQ908771-JQ908773, JQ908775, JQ908779, JQ908780, JQ908782, JQ908783, JQ908800-JQ908802, JQ908806, JQ908807, JQ908825, JQ908828, JQ908833-JQ908836, JQ908847, JQ908890-JQ908897, JQ908902, JQ908903, JQ908909, JQ908911, JQ908913, JQ908921, JQ908922, JQ908942, JQ908944-JQ908947, MF121697, MF121722, MF121725, MF121755, MF121763, MF121771, MF121775, MW080722-MW080725, MW080728, MW080729). Distance analyses were performed with MEGA7 (Kumar et al. 2016), using a Neighbor-Joining (Saitou & Nei 1987) algorithm with the Kimura-2 parameter model (K2P – Kimura 1980) to estimate genetic distances. The robustness of nodes was evaluated through bootstrap re-analysis of 1000 pseudoreplicates. The tree was replotted using R package ggtree (Xu et al. 2022).

2.4. Distribution of pairwise comparisons and modes analysis

The distribution of the pairwise distances produced from the sequences of the Luxembourgish specimens *of A. caliginosa* was analysed in order to classify them into the different modes (i.e local frequency maxima in the distribution of pairwise distance comparisons) that could be related to a putative reproductive isolation status through the comparison with previously published datasets (Porco et al. 2018). A mixture of normal distributions was fitted through the Expectation-Maximization algorithm implemented in the R package 'mixtools' (Benaglia et al. 2009). The fitting, based on the distribution of the density of probabilities, was cross-validated by comparing the Bayesian Information Criterion (BIC) values for mixtures including a lesser number of modes than the one defined a priori as the optimal expectation.

3. Results

3.1. DNA barcode distances

A clear barcode gap was found in this dataset with a mean intraspecific divergence of 0.8% and a mean interspecific divergence of 18.2% (Table 1, Fig. 1).

The sequences of the *A. caliginosa* specimens collected in Luxembourg exhibited a genetic

Table 1. Intraspecific/lineages and Interspecific/lineages K2P distances.

#	Species	Intraspecific	Interspecific					
			1	2	3	4	5	6
1	*A. cupulifera*	0.1						
2	*A. longa*	1.5	23.2					
3	*A. caliginosa* L1	1.5	19.0	22.3				
4	*A. caliginosa* L2	1.1	19.9	18.3	15.5			
5	*A. caliginosa* L3	0.6	21.9	18.2	17.2	11.9		
6	*A. caliginosa* L4	0.0	20.8	18.1	16.9	13.4	15.7	

divergence ranging from 13.4% to 16.9% with the other cryptic lineages from the species complex. These genetic divergence values are similar to those found within the rest of the cryptic *A. caliginosa* lineages (mean divergence without Luxembourgish specimens 14.9%).

3.2. Analysis of modes in pairwise distance comparisons

The multimodal analysis of the distribution of pairwise distances showed the occurrence of four of the five modes that were previously described in Lumbricidae communities (Porco et al. 2018 - Fig. 2). Apart from the mode M1, that only encompasses the intraspecifc/lineage pairwise distances, three other modes that comprise interspecific/lineages pairwise distances were recovered: M3, M4 and M5 (Fig. 2). More information on these modes and their definition in a broader taxonomic context can be found in Porco et al. 2018. Here is given a brief summary of each mode potential properties and content: a) M3 contains pairwise comparisons for lineages experiencing gene flow (e.g. *Allolobophora chlorotica* L1/L2/L3), thus, other lineages for which distance pairwise comparisons are comprised in this mode could be suspected of gene flow or incomplete lineage sorting events (e.g. *A. caliginosa* L1/L2 and L2/L3); b) M4 comprises pairwise comparisons from

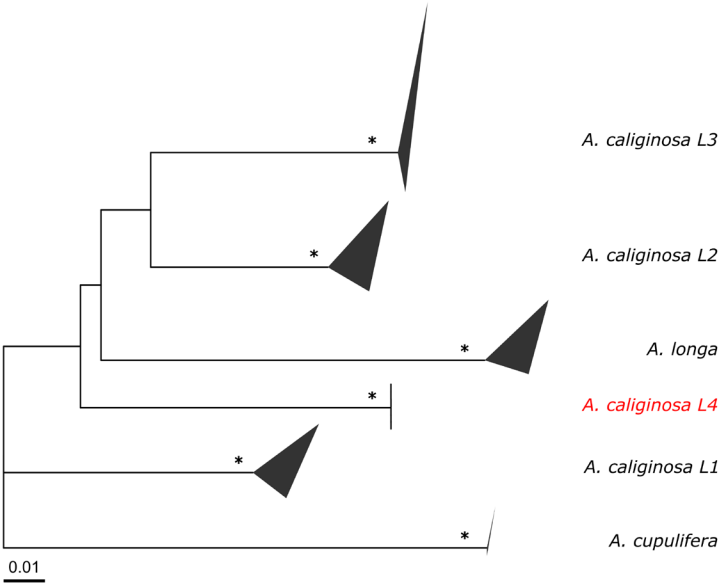


Fig. 1. Neighbor joining tree of COI K2P distances among the four lineages of *A. caliginosa* and two other Aporrectodea species. Each cluster was symbolized by a triangle: the upper and lower sides of the triangle represent respectively the maximal and the minimal genetic distance within the cluster. Bootstrap supports ≥99% are marked with an asterisk.

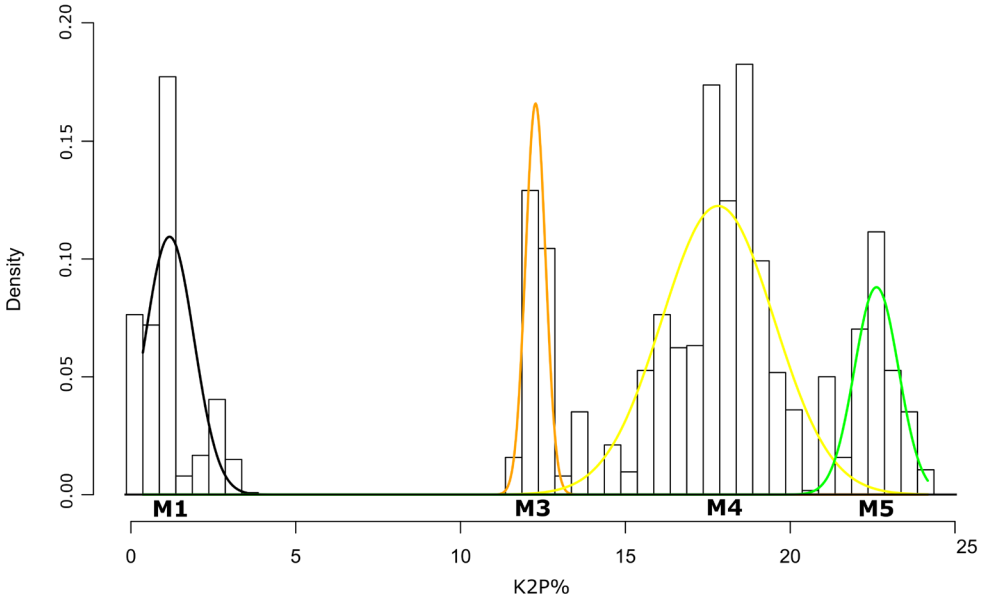


Fig. 2. Normal distributions fitted on the distribution of the pairwise comparisons of K2P distances, delineating the different modes. Four normal distributions were fitted thus delineating four modes in the global distribution of pairwise comparisons. Line colors of the normal distributions feature the different modes that are also annotated under the x-axis (M1-5). The density of frequencies was used on y-axis. This value is proportional to frequencies values reported to the amplitude of K2P distance bins in which pairwise comparisons were classified (i.e. 0.5%). The shape of distribution is unchanged but this transformation allows the use of a probability density function in order to fit normal distributions.

well defined species such as *L terrestris* /*L herculeus* but also from several cryptic lineages (e.g. *A. caliginosa* L1/L3) - no or low level of gene flow can be expected among the member of this mode (see the *A. chlorotica* L4 case (Dupont et al. 2011, Porco et al. 2018)); c) M5 contains pairwise comparisons only for well discriminated species for both morphology and DNA barcode in which no gene flow was shown or suspected yet). No pairwise distance was classified in the mode M2 (see Porco et al. 2018).

Within these three modes, all pairwise comparisons within the *A. caliginosa* complex were found in Mode 3 and 4 when the comparisons involving at least one of the external species to the complex ranged into the mode M5 (Fig. 3).

The cross-validation through BIC values confirmed that the mixture models comprising the amount of modes retrieved were more consistent than those comprising a lesser number of modes (which exhibited higher BIC values).

4. Discussion

The *A. caliginosa* complex was uncovered by Pérez-Losada et al. (2009) and further explored and formalized with DNA barcodes by Porco et al. (2013). Here, we showed that a new member of the *A. caliginosa* complex dwells in Luxembourg grasslands, and it was found as genetically distinct from other lineages of the complex. We will name L4 following the previous nomenclature applied to the complex (Porco et al. 2013). This finding supports that this common and highly studied group of Lumbricidae can still present levels of diversity that have not yet been recorded. The lineage was not found in the adjacent crop field, which might suggest an ecological constraint on its occurrence or its density in disturbed and resource-poor habitat such as crop fields. This will be tested in further collection events.

Pairwise comparisons between the cryptic lineages of the *A. caliginosa* complex ranged in two of the modes, M3 and M4, that were

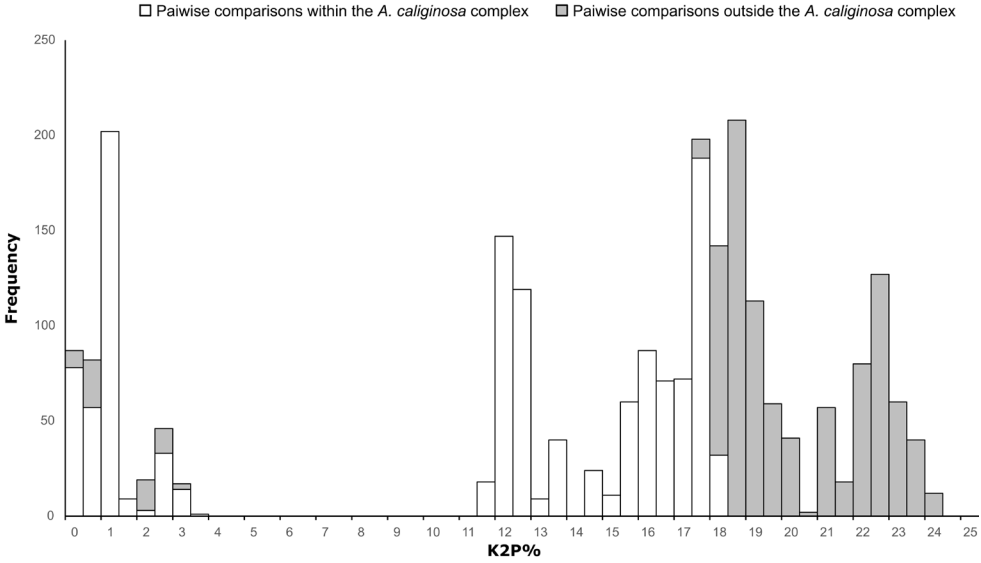


Fig. 3. Distribution of frequencies of the pairwise comparisons of K2P distances showing the range of comparisons involving *A. caliginosa* lineages alone (white) and those involving other species (Gray).

previously shown to contain lineages still experiencing gene flow through an incomplete achievement of reproductive isolation (Dupont et al. 2016, Porco et al. 2018). More specifically, *Allolobophora chlorotica* is another complex comprising four COI lineages, exhibiting the same range of interlineages genetic distances (i.e. modes M3 and M4) and which was extensively shown to present various levels of hybridization capability among its COI lineages (Dupont et al. 2016, Dupont et al. 2022). This may suggest that the same incomplete reproductive isolation could occur among the cryptic lineages of *A. caliginosa* as well.

5. Conclusion

A new COI cryptic lineage for the *A. caliginosa* complex was found in Luxembourg grasslands. The inclusion of this new lineage into the multimodal analysis of pairwise comparisons of genetic distances allowed to further hypothesize on the genetic isolation among these lineages. It also demonstrates the need for the systematic DNA barcoding of the Lumbricidae communities in Europe where species-level genetic variations are yet to be described.

Acknowledgments

We are grateful to two anonymous reviewers who suggested improvements of the manuscript. We thank Claude Meisch for the editorial work and Paul Diederich for the page layout.

References

Benaglia T., D. Chauveau, D.R. Hunter & D.S. Young, 2009. mixtools: An R Package for Analyzing Finite Mixture Models. *Journal of Statistical Software* 32: 1-29.

Blouin M., M.E. Hodson, E.A. Delgado, G. Baker, L. Brussaard, K.R. Butt, J. Dai, L. Dendooven, G. Peres, J.E. Tondoh, D. Cluzeau & J.J. Brun, 2013. A review of earthworm impact on soil function and ecosystem services. *European Journal of Soil Science* 64: 161-182.

Cameron E.K., I.S. Martins, P. Lavelle, J. Mathieu, L. Tedersoo, F. Gottschall, C.A. Guerra, J. Hines, G. Patoine, J. Siebert, M. Winter, S. Cesarz, M. Delgado-Baquerizo, O. Ferlian, N. Fierer, H. Kreft, T.E. Lovejoy, L. Montanarella, A. Orgiazzi, H.M. Pereira, H.R.P. Phillips, J. Settele, D.H. Wall & N. Eisenhauer, 2018. Global gaps in soil biodiversity data. *Nature Ecology & Evolution* 2: 1042-1043.

Dupont L., H. Audusseau, D. Porco & K.R. Butt, 2022. Mitonuclear discordance and patterns of reproductive isolation in a complex of simultaneously hermaphroditic species, the

Allolobophora chlorotica case study. *Journal of Evolutionary Biology* 35: 831-843.

Dupont L., F. Lazrek, D. Porco, R.A. King, R. Rougerie, W.O.C. Symondson, A. Livet, B. Richard, T. Decaëns, K.R. Butt & J. Mathieu, 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., D. Porco, W.O.C. Symondson & V. Roy, 2016. Hybridization relics complicate barcode-based identification of species in earthworms. *Molecular Ecology Resources* 16: 883-894.

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

Jones C.G., J.H. Lawton & M. Shachak, 1994. Organisms as ecosystem engineers. *Oikos* 69: 373-386.

Kimura M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide-sequences. *Journal of Molecular Evolution* 16: 111-120.

Kumar S., G. Stecher & K. Tamura, 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution* 33: 1870-1874.

Lavelle P., 1988. EARTHWORM ACTIVITIES AND THE SOIL SYSTEM. *Biology and Fertility of Soils* 6: 237-251.

Lavelle P. & A.V. Spain (2001) Soil ecology. Kluwer Academic Publishers, i-xxvii, 1-654 pp.

Pelosi C., M. Bertrand, Y. Capowiez, H. Boizard & J. Roger-Estrade, 2009. Earthworm collection from agricultural fields: Comparisons of selected expellants in presence/absence of hand-sorting. *European Journal of Soil Biology* 45: 176-183.

Pérez-Losada M., M. Ricoy, J.C. Marshall & J. Dominguez, 2009. Phylogenetic assessment of the earthworm *Aporrectodea caliginosa* species complex (Oligochaeta: Lumbricidae) based on mitochondrial and nuclear DNA sequences. *Molecular Phylogenetics and Evolution* 52: 293-302.

Porco D., T. Decaëns, L. Deharveng, S. James, D. Skarżyński, C. Erséus, K. Butt, B. Richard & P.N. Hebert, 2013. Biological invasions in soil: DNA barcoding as a monitoring tool in a multiple taxa survey targeting European earthworms and springtails in North America. *Biological Invasions* 15: 899-910.

Porco D., C.H. Chang, L. Dupont, S. James, B. Richard & T. Decaëns, 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. *Applied Soil Ecology* 124: 362-371.

Riley H., R. Pommeresche, R. Eltun, S. Hansen & A. Korsaeth, 2008. Soil structure, organic matter and earthworm activity in a comparison of cropping systems with contrasting tillage, rotations, fertilizer levels and manure use. *Agriculture Ecosystems & Environment* 124: 275-284.

Rutgers M., A.J. Schouten, J. Bloem, N. van Eekeren, R.G.M. de Goede, G. Akkerhuis, A. van der Wal, C. Mulder, L. Brussaard & A.M. Breure, 2009. Biological measurements in a nationwide soil monitoring network. *European Journal of Soil Science* 60: 820-832.

Saitou N. & M. Nei, 1987. The neighbor-joining method - a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4: 406-425.

Scheffers B.R., L.N. Joppa, S.L. Pimm & W.F. Laurance, 2012. What we know and don't know about Earth's missing biodiversity. *Trends in Ecology & Evolution* 27: 501-510.

van Groenigen J.W., I.M. Lubbers, H.M.J. Vos, G.G. Brown, G.B. De Deyn & K.J. van Groenigen, 2014. Earthworms increase plant production: a meta-analysis. *Scientific Reports* 4: 6365.

Xu S., L. Li, X. Luo, M. Chen, W. Tang, L. Zhan, Z. Dai, T.T. Lam, Y. Guan & G. Yu, 2022. Ggtree: A serialized data object for visualization of a phylogenetic tree and annotation data. *iMeta*: e56.