Phylogenetic study of some Aporrectodea species based on molecular markers

Charlène Briard, Jiang-Ping Qiu, Zhao Qi, Muriel Guernion, Daniel Cluzeau

Phylogenetic study of some *Aporrectodea* species based on molecular markers

(Oligochaeta: Lumbricidae)

Charlène Briard, Jiang-Ping Qiu, Qi Zhao, Muriel Guernion, Daniel Cluzeau

Abstract. The *Nicodrilus* genus is not accepted in the nomenclature according to ICZN; *Aporrectodea* term is used. However, according to Bouché, *Aporrectodea* and *Allolobophora* genera are not homogeneous and are supposed as polyphyletic genera. The aim is to study the phylogenetic structure of *Aporrectodea* genus in order to verify its cladistic nature and its taxonomical validity. In this work, five species, belonging to the *Aporrectodea* genus, as the most common in France are studied. First, we use usual morphological characteristics to identify each species. Species life history traits are included among morphological characteristics (*e.g.* clitellum, puberculum and pores position, Morren's gland, body size…). Then, we pursue a molecular approach on individuals sampled in France. Molecular phylogenetic analyses are based on the sequences of nuclear (rDNA 28S) and mitochondrial (COI, rDNA 16S) gene regions and performed with maximum likelihood and bayesian inference. Phylogenetic analyses revealed two deep sister clades, *Aporrectodea caliginosa* species complex in one hand, and *A. giardi*, *A. longa* and *A. nocturna* in the other hand. The status of *A. nocturna* is recognized instead of *A. longa* and *A. giardi* ones. Our study cannot confirm the validity of *Aporrectodea* genus but highlights different lineages within species such as *A. longa* and *A. caliginosa*. Markers characterization ensures a molecular genotyping in lumbricids and begins to reveal intra-specific variabilities degrees unsuspected.

Key words. Phylogeny, COI, rDNA16S, rDNA28S, *Aporrectodea*.

Introduction

Most of the available phylogenetic classifications of Lumbricidae are based on morphological studies (BOUCHÉ 1972, SIMS & GERARD 1985). Although the morphology

constitutes the base of species identification, these classifications have such particularities like a low resolution of phylogenetic relationship leading to uncertain grouping. Taking into account the important role which earthworms play as key organisms in terrestrial ecosystems functioning, the lack of knowledge in phylogenetic relationship among species can compromise interpretations in studies applied in functional ecology, biodiversity and evolution (King et al. 2008). DNA taxonomy and associated molecular tools are a good way to reveal the true level of biodiversity (Chang et al. 2009), but the use of molecular tools to study earthworm taxonomy has only recently started. Molecular markers, mitochondrial and nuclear, are thus developed (PÉREZ-LOSADA et al. 2009). Lumbricids represent a taxonomical diversified group of terrestrial Oligochaeta. Among the five families within Lumbricoidea, Lumbricidae family counts around 30 genera; some of them provide very diverse species based on their ecological functions. Molecular approaches realized for a few years on *Lumbricidae* confirmed the validity of certains genera as *Octodrilus* (POP et al. 2007), *Dendrobaena* (POP et al. 2007). However, genera as *Aporrectodea* (NOVO et al. 2010, PÉREZ-LOSADA et al. 2009) and *Allolobophora* (POP et al. 2007) remain under discussion. Molecular approaches also highlight a cryptic diversity among species such as *Allolobophora chlorotica* or *Lumbricus terrestris* (JAMES et al. 2010). Finally, in certain occasions, the lack of agreement in their ranking as diagnostic characters for taxonomic and phylogenetic purposes has led to situations in which the same species receives different names in different parts of the world and may be included in different genera depending on the classification system proposed. The problem of the use of different classification system lead to different names of species or even genera, for example the *Aporrectodea* genus which received three different names along years as *Allolobophora* and *Nicodrilus*.

Indeed, *Nicodrilus* genus defined by BOUCHÉ in 1972 comes from the organization of the invalid genus *Allolobophora*. Based on morphological descriptions of species and he grouped them into the new *Nicodrilus* genus as more homogeneous genus for him. Since then, he distinguishes *Allolobophora sensu stricto* containing the type species *A. chlorotica*; and then *Allolobophora sensu lato* containing different taxonomic units around six species.

Another way to organize the *Allolobophora* genus was after OMODEO chose *A. chlorotica* as the type species of the genus. Since then, most of species initially identified as *Allolobo-*

phora were excluded, and the new generic name *Aporrectodea* is used. *Aporrectodea* is now the term used and known and represent around 82 species according to Fauna Europaea database (DE JONG 2010). Indeed, *Aporrectodea* term is used now considering the *Nicodrilus* genus as junior synonym of *Aporrectodea*.

This work aims to specify lumbricids taxonomy establishing phylogenetic relationship among *Aporrectodea* species. The first objective is to study the phylogenetic structure of *Aporrectodea* genus. Furthermore, we attend to verify the species status within *Aporrectodea* genus and to study the species complex of *A. caliginosa (A. caliginosa / A. tuberculata* / *A. trapezoides*). Results obtained are used to describe phylogenetic relationship among species, and discussed according to recent results presently known.

Material and methods

Sampling. A total of 41 specimens of five species of *Aporrectodea* commonly found in France were collected in different sites during spring 2009 and 2010 in France: *Aporrectodea giardi, A. longa, A. nocturna, A. trapezoides* and *A. caliginosa*. Sequences from GenBank® database of each species collected are also used to implement the dataset, as well as sequences of *Aporrectodea tuberculata, A. icterica, A. rosea* and *Allolobophora chlorotica*. The *Lumbricus* genus is added to be used as outgroup, represent by three species (*L. rubellus*, *L. castaneus* and *L. terrestris*) coming from fresh material. The specimens were described focusing on morphological characteristics described by QIU & BOUCHÉ (1998) and SIMS & GERARD (1985). Taxa names were accepted according to BLAKEMORE (2006A).

DNA extraction, polymerase chain reaction (PCR) and DNA sequencing. The total DNA was obtained from a muscular tissu section of each individual in the caudal region, avoiding the digestive tube to eliminate bacterian DNA present. DNA was extracted using the Chelex protocol according to ESTOUP ET AL. (1996), and then each entire individual is conserved in pure ethanol. The phylogenetic analysis was supported with the variation of (i) a fragment of the mitochondrial gene (631 bp) coding the first subunit of cytochrome oxidase (CO1), (ii) a fragment of the mitochondrial gene 16S (435 bp) and (iii) a fragment of the nuclear genes 28S (570 bp). Oligonucleotids used to perform the amplification by PCR (Polymerase Chain Reaction) are for

CO1 fragment, the universal primers HCO2198 and LCO1490 (FOLMER et al. 1994), for rDNA 16S fragment, the primers 16SarL (THOMAZ et al. 1996) and 16SbrH (PALUMBI et al. 1991), and for the DNA fragment 28S, the primers 28S-F1 and 28S-R1 (PÉREZ-LOSADA et al. 2009). Each PCR was conducted in a 25 µl reaction volume containing 12.5 µl of Mix [Bioline] (containing the Taq polymerase, dinucleotides dNTPs, $MgCl_2$ and tampon solution), 0.5 µl of each primer and 1.25 µl of DNA extracted. PCR was performed using, for 16S rDNA and CO1, an initial denaturation of 2 min 30 seconds at 94°C, followed by 35 cycles of 30 seconds denaturation at 94°C, 30 seconds annealing at 52°C, and at 48°C respectively for each marker, 45 seconds extension at 72°C, and a 10 min final extension at 72°C; for 18S DNA, an initial denaturation of 5 minutes at 95°C, followed by 35 cycles of 30 seconds denaturation at 95°C, 30 seconds annealing at 55°C, 1min30 extension at 72°C, and a 8 minutes final extension at 72°C; for 28S rDNA, an initial denaturation of 3 minutes at 96°C, followed by 35 cycles of 30 seconds denaturation at 95°C, 45 seconds annealing at 50°C, 1minute extension at 72°C, and a 5 minutes final extension at 72°C. Double strand sequences were obtained by direct manual sequencing of the PCR products using an automated sequencer (PE Applied Biosystems 310 Genetic Analyser, UMR 6553; plateforme de séquencage génotypage OUEST-Genopole®).

Phylogenetic analysis. Nucleotide sequences from each gene region were aligned using ClustalW program (version 1.7, THOMPSON et al. 1994). MrAIC program (version 1.4.2, NYLANDER 2004) was used to select the best fitted model of sequences evolution for each gene partition under the Akaike Information Criterion (AIC, AKAIKE 1973) and the Bayesian Information Criterion (BIC, SCHWARZ 1978). Gene regions were analyzed both in combination as a single dataset and as multiple concatenated partitions. Maximum Likelihood (ML) analysis of the concatenated dataset (3 partitions) was performed in RAxML (STAMATAKIS 2006) using 1000 RA, under the general time reversible model of sequences evolution and a gamma distribution in height categories (GTR+Γ). Clade support under the ML approach was assessed using the non-parametric bootstrap procedure (FELSENSTEIN 1985) with 1000 bootstrap replicates. The Consense program from Phylip (version 3.6) (FELSENSTEIN 2004) was used to build the consensus tree from the 1000 bootstrap replicates. The dataset was also analysed using bayesian methods coupled with Markov chain Monte Carlo (MCMC) inference as implemented in MrBayes v. 3.1.2 (RONQUIST & HUELSENBECK 2003). Four chains of MCMC analyses were run and each Markov chain was started from a random tree and run for 2 million generations, sampling every 100 gen-

erations. The first 2000 trees were taken away from final analysis corresponding to the burn-in period.

Results

According to the concatenated dataset analysis, the phylogenetic tree shows six taxonomic units with all a strong support, posterior probabilities above 0.9 (Fig. 1). The tree topology is not different between the ML and BI analyses methods. The first taxonomic unit contains most of the individuals initially identified as *A. giardi* (AG). *Aporrectodea giardi* appear to be monophyletic (Fig. 1-A; pp=1.00, bp=99.8). The second taxonomic unit contains all of our *A. nocturna* (AN) individuals with a probability equal to 1 (Fig. 1-B1; pp=1.00, bp=99.1), but also contains the *A. longa* species from GenBank (Fig 1-B2; pp=1.00, bp=79.5). Concerning the *A. longa* (AL) species, our specimens are close-related but not with the *A. longa* from GenBank (Fig. 1C; pp=1.00, bp=98.3). This problem can come from a misidentification of the specimen either by us or from GenBank specimen.

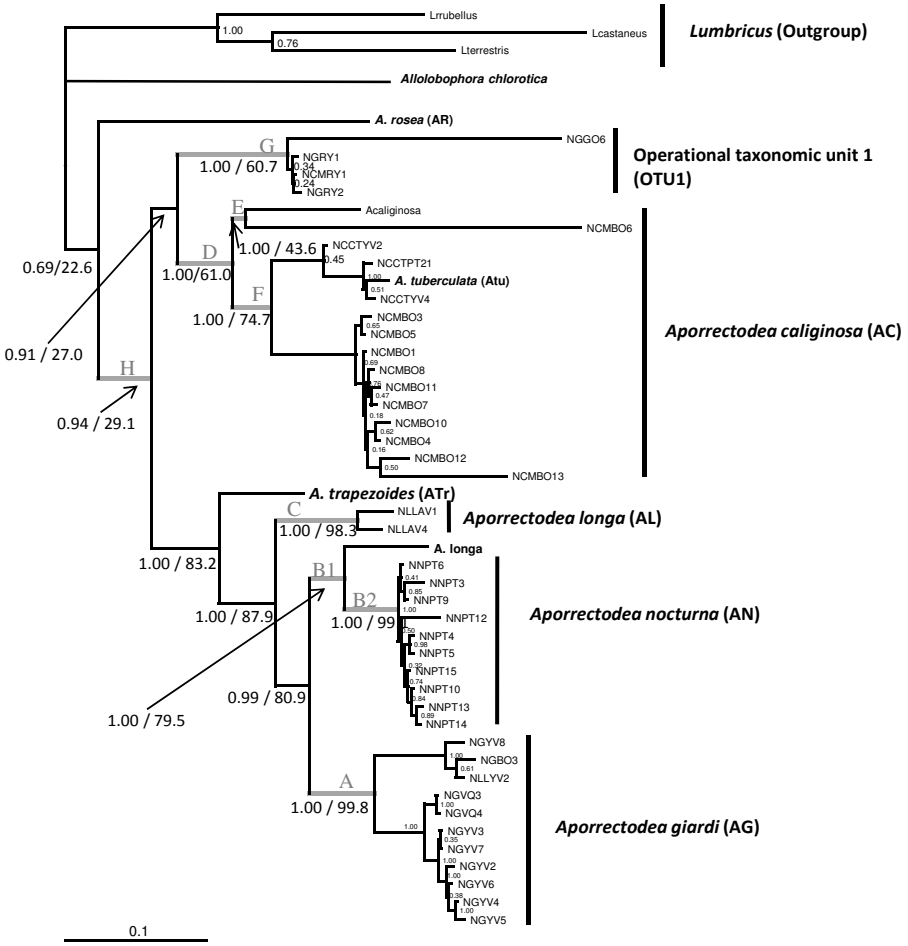

Fig. 1. Phylogenetic relationship obtained from Bayesian Inference analysis on 3 genes concatenated (CO1, 16S & 28S; 1636 nt). Numbers at nodes refer, from left to right respectively, to posterior probabilities in Bayesian analysis and to bootstrap percentages in ML analysis (not shown within species). Dataset was analyzed with the GTR + G model applied to each partition.

The *A. caliginosa group* (AC), also well supported (Fig. 1-D; pp=1.00, bp=61.0) is clearly separated from ATr, AL and AG species. But in this group, we can clearly see two well supported taxonomic units.

The first one contains our *A. caliginosa* but also one *A. tuberculata* (ATu) (Fig. 1-E; pp=1.00, bp=65.8). The identification on morphological characters is quite difficult in order to distinguish these two species. A wrong identification cannot be rejected but the close relationship between them shows that maybe we have two subspecies and not two species. The data need to be completed by sequences from others *A. tuberculata* species. The second taxonomic unit also contains what we identified as *A. caliginosa* (Fig. 1-F). But this group is clearly a new lineage of *A. caliginosa* with a strong support (pp=1.00, bp=74.7).

The well-supported and close-related clade to "*A. caliginosa group*" contains four individuals identified as different species (OTU1, Fig. 1-G; pp=1.00, bp=60.7): three *A. giardi* which are not close-related with the others *A. giardi*, but also they are close related to *A. caliginosa* species with a strong support only by the Bayesian analysis (Fig. 1-H; pp=0.91).

Phylogenetic analyses reveal two deep sister clade: *A. caliginosa* group is clearly separated from ATr, AL, AN and AG. AG is sister of AN, then both are sister of AL. Although species relationship can be described, different lineages seem to be revealed within each species as well.

Discussion and conclusion

The phylogenetic structure of *Aporrectodea* genus is studied focusing on five species commonly found in France. ML and BMCMC phylogenetic analyses based on three different mitochondrial and nuclear genes revealed two deep sister clades. One is composed of *A. caliginosa*, *A. tuberculata*, and another taxonomic unit containing *A. giardi* specimens (OTU1); and another is composed of *A. trapezoides, A. nocturna*, *A. longa* and *A. giardi*. *A. caliginosa*, *A. tuberculata*, and *A. nocturna* formed monophyletic assemblages, but *A. giardi* and *A. longa* resulted paraphyletic. The *A. caliginosa group* (AC) is clearly separated from ATr, AL and AG species as it has been already shown by previous studies of *A. caliginosa*

species complex (PÉREZ-LOSADA et al. 2009). Results show that the phylogenetic relationship between species is difficult to assess because of the non-verified status of species. Indeed, results clearly show that *A. nocturna* species is well recognized, only if we ignore the *A.longa* specimen coming from GenBank which belongs to the same clade as our *A. nocturna* specimens. In this case, a wrong identification could be possible. In the other hand, *A. giardi, A. longa* and *A. caliginosa* are revealed as polyphyletic. Although the species status of *A. longa* has been widely accepted because of its morphological differences, molecular analyses revealed the paraphyly of this species (PÉREZ-LOSADA et al. 2009). *A. longa* is also taxonomically considered the closest species to *A. trapezoides* and *A. nocturna* (BLAKEMORE 2006b, GATES 1972). Molecular trees seem to support this relationship. *A. giardi* is composed by two subclades, which are not related by species location. Thus, concerning *A. giardi* and *A. longa* species, others specimens must be added to confirm our grouping. The species complex *A. caliginosa* (*A. caliginosa* / *A. tuberculata* / *A. trapezoides*) is not resolved yet. Three subclades are revealed, close-related to OTU1 and two of them are well-supported (Fig. 1E, 1F). Those different lineages observed are not related to species locations because specimens mixed in the same group come from different locations. *A. tuberculata* and *A. trapezoides* specimens should be added to visualize their positions within the *A. caliginosa species complex*. More populations from different locations should also be studied (only three populations are shown here).

The OTU1 contains four individuals identified as different species (Fig. 1-G; pp=1.00, bp=60.7): three *A. giardi* which are not close-related with the others *A. giardi*, and one *A. caliginosa*. OTU1 is close-related to *A. caliginosa* species with a strong support only by the Bayesian analysis (Fig. 1-H; pp=0.91). This OTU1 can come from a misidentification of specimens but *A. giardi* and *A. caliginosa* are morphologically easy to distinguish. It is difficult to confuse knowing that those species belong to different ecological categories; their pigmentation and size are also clearly different. Molecular mistakes cannot be rejected or phenomema as length branch attraction can be assumed during the phylogenetic analyses. This grouping remains unresolved and need more focus to understand it.

The integrative approach of species delimitation can greatly be helpful to species identification (PUILLANDRE et al. 2009, ROE & SPERLING 2007). Actually, in order to really study the

variability within a species, the strategy should be first to increase the number of specimens by species as well as the number of locations where specimens are collected. Then, microsatellites markers could be used instead as they are usually used for intraspecific level study as well as barcoding techniques. Indeed, very highly divergent lineages within earthworm species were reported analyzing mitochondrial COI and 16S genes as well; for example within *Aporrectodea longa*, *A. rosea*, *Allolobophora chlorotica* and *Lumbricus rubellus* (KING et al. 2008) or *Lumbricus terrestris* (JAMES et al. 2010). The existence of multiple cryptic species within these taxa is suggested. Our results support this pattern, hence suggesting an unprecedented diversity within Lumbricidae earthworms (KING et al. 2008).

Finally, our study cannot confirm the validity of *Aporrectodea* genus. Indeed, only five species are studied here. Others species should be added (as *A. tuberculata, A. velox, A. gognus* also present in France) in order to have the majority of species and obtain a global view of this genus. This study also needs to include others *Aporrectodea* species as *A. icterica, A. handlirshi, A. jassyensis, A. limicola*, less common in France. Some of them were collected this year and this work is expected to be improved.

Our work constitutes a preliminary study since additionally information brought by others markers could be necessary to confirm our results on three markers analyses. The study on species complex *Aporrectodea caliginosa* (PÉREZ-LOSADA et al. 2009) does not count less than six molecular markers (mtDNA and nuDNA). CO2, 12S rDNA, 28S rDNA and 16S rDNA were used to try to light taxonomic problems on this complex, unresolved with molecular markers as 16S rDNA and COI used (POP et al. 2007) until now. This kind of study underlines the importance of using multiloci data to delimitate earthworm species and to estimate phylogenetic relationship. Markers characterization ensures a molecular genotyping in lumbricids (CHANG et al. 2009) and begins to reveal intra-specific variability degrees unsuspected, which result in some cases, from cryptic speciation, defined as morphological similar species but genetically distinct (RICHARD et al. 2010, KING et al. 2008, NOVO et al. 2010, JAMES et al. 2010).

References

AKAIKE, H. (1973): Maximum Likelihood Identification of Gaussian Autoregressive Moving Average Models. – Biometrika 60: 255-265.

BLAKEMORE, R. J. (2006a): A Series of Searchable Texts on Earthworm Biodiversity, Ecology and Systematics from Various Regions of the World – 2nd Edition and Supplement (2006)**.** In: N. KANEKO & M.T. ITO (Eds). CD-ROM Publication. COE Soil Ecology Research Group, Yokohama National University, Japan. Online: http://bio-eco.eis.ynu.ac.jp/eng/database/earthworm/ [2011].

BLAKEMORE, R. J. (2006b): Cosmopolitan Earthworms - An Eco-Taxonomic Guide to the Peregrine Species of the World. 2nd ed. VermEcology. – Japan. 600 pp.

BOUCHE, -M. B. (1972): Lombriciens de France, Ecologie et Systématique. (1 ed.). – Paris: I.N.R.A.

CHANG, C. H., ROUGERIE, R. & CHEN, J. H. (2009): Identifying earthworms through DNA barcodes: Pitfalls and promise. – Pedobiologia 52: 171-180.

DE JONG, Y. S. D. M. (Ed.) (2010): Fauna Europaea. version 2.4. –www.faunaeur.org. – Ecosystems 7: 672-685.

ESTOUP, -A., LARGIADER, C. R., PERROT, E. & CHOURROUT, D. (1996): Rapid one-tube DNA extraction for reliable PCR detection of fish polymorphic markers and transgenes. – Molecular Marine Biology and Biotechnology 5: 295-298.

FELSENSTEIN, J. (1985): Confidence limits on phylogenies: an approach unsing the bootstrap. – Evolution 39: 783-791.

FELSENSTEIN, J. (2004): PHYLIP (Phylogeny Inference Package) version 3.6. – Seattle: Department of Genome Sciences, University of Washington.

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. – Molecular Marine Biology and Biotechnology 3: 294-299.

GATES, G. E. (1972): Contributions to North American earthworms (Annelida: Oligochaeta). No. 3: towards a revision of the earthworm family Lumbricidae IV. The trapezoides species group. – Bull. Tall Timbers Res. Station 12: 1-146.

JAMES, S. W., PORCO, D., DECAËNS, T., RICHARD, B., ROUGERIE, R. & ERSEUS, C. (2010): DNA Barcoding Reveals Cryptic Diversity in *Lumbricus terrestris* L., 1758 (Clitellata): Resurrection of *L. herculeus* (Savigny, 1826). – PLoS ONE 5(12): e15629. doi:10.1371/journal.pone.0015629.

KING, R. A., TIBBLE, A. L. & SYMONDSON, W. O. C. (2008): Opening a can of worms: unprecedented sympatric cryptic diversity within British lumbricid earthworms. – Molecular Ecology 17: 4684-4698.

NOVO, M., ALMODOVAR, -A., FERNANDEZ, R., TRIGO, D. & DIAZ COSIN, D. J. (2010): Cryptic speciation of hormogastrid earthworms revealed by mitochondrial and nuclear data. – Molecular Phylogenetics and Evolution 56: 507-512.

NYLANDER, J. A. A. (ED.) (2004): *MrAIC.pl*. – Uppsala University: Evolutionary Biology Centre.

OMODEO, P. (1956): Contributo alla revisione dei Lumbricidae. – Arch. zool. ital. 41: 129-213.

PALUMBI, S. R., MARTIN, A., ROMANO, S., MCMILLAN, W. O., STICE, L. & GRABOWSKI,G. (1991): The Simple Fool's Guide to PCR. Version 2. – Honolulu: University of Hawaii.

PEREZ-LOSADA, M., RICOY, M., MARSHALL, J. C. & DOMINGUEZ, J. (2009): Phylogenetic assessment of the earthworm *Aporrectodea caliginosa* species complex (Oligochaeta: Lumbricidae) based on mitochondrial and nuclear DNA sequences. – Mol. Phylogenet. Evol. 52: 293-302.

POP, A. A., CECH, G., WINK, M., CSUZDI, CS. & POP, V. V. (2006): Application of 16S, 18S rDNA and COI sequences in the molecular systematics of the earthworm family Lumbricidae (Annelida, Oligochaeta). In: 8th International Symposium on Earthworm Ecology (ISEE8), pp. S43-S52. – Cracow, Poland.

PUILLANDRE, N., BAYLAC, M., BOISSELIER, M. C., CRUAUD, C. & SAMADI, S. (2009): An integrative approach to species delimitation in Benthomangelia (Mollusca: Conoidea). – Biological Journal of the Linnean Society 96: 696-708.

QIU, J. P. & BOUCHÉ, M.B. (1998): Révision des taxons supraspécifiques de Lumbricoidea. – Doc. Pedozoologiques et Integrologiques 3: 179-216.

RICHARD, B., DECAËNS, T., ROUGERIE, R., JAMES, S. W., PORCO, D. & HEBERT, P. D. N. (2010): Re-integrating earthworm juveniles into soil biodiversity studies: species identification through DNA barcoding. – Molecular Ecology Resources 10: 606-614.

ROE, A. D. & SPERLING, F. A. H. (2007): Population structure and species boundaries delimitation of cryptic *Dioryctria* moths-: an integrative approach. – Mol. Ecol. 16: 3617–3633.

RONQUIST, F. & HUELSENBECK, J. P. (2003): MrBayes 3: Bayesian phylogenetic inference under mixed models. – Bioinformatics 19: 1572-1574.

SCHWARZ, G. (1978): Estimating the Dimension of a Model. – Annals of Statistics 6: 461-464.

SIMS, R. W. & GERARD, B. M. (1985): Earthworms. – London: Linnean Society of London & The Estuarine and Brackish-Water Sciences Association.

STAMATAKIS, A. (2006): RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. – Bioinformatics 22: 2688-2690.

THOMAZ, D., GUILLER, A. & CLARKE, B. (1996): Extreme divergence of mitochondrial DNA within species of pulmonate land snails. – Proceedings of the Royal Society of London Series B-Biological Sciences 263: 363-368.

THOMPSON, J. D., HIGGINS, D. G. & GIBSON, T. J. (1994): Clustal-W - Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. – Nucleic Acids Research 22: 4673-4680.

Author's adresses: C. Briard, Q. Zhao, M. Guernion, L. Rougé, A. Bellido, D. Cluzeau, Université Rennes 1, UMR CNRS EcoBio, Observatoire de Rennes, Station Biologique, 35380 Paimpont, France. — J.-P. Qiu, Shanghai Jiao Tong University, Agricultural and Biological School, Resources and Environment Department, Shanghai, China. – E-mail contact: charlene.briard@univ-rennes1.fr.