npg

ORIGINAL ARTICLE

High mitochondrial DNA sequence diversity in the parthenogenetic earthworm *Dendrobaena octaedra*

KE Knott and J Haimi

Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland

Apomictic parthenogens are clonal organisms with limited genetic opportunity for increasing diversity beyond mutation. However, such species can be successful and have been shown to harbor more genetic diversity than might be expected. Here we surveyed diversity of the cytochrome oxidase subunit I gene from the mitochondrial genome of the earthworm *Dendrobaena octaedra*, an apomictic parthenogen. Diversity estimates made previously from allozyme markers for this species were high, but could have been affected by a detection bias, namely variable expression of alleles in the polyploid genome. We found similarly high mtDNA diversity over three localities in Finland, each represented by two sites; either with or without anthropogenic soil metal contamination. A single haplotype was most common over all sites, and over two thirds of all haplotypes sampled were similar to it in sequence (only varying by 1–3 substitutions). However, more divergent rare haplotypes were also found in one locality, Imatra, in southeast Finland close to the Karelian Isthmus. Metal contamination in the soil did not have a significant effect on genetic diversity, although metal exposure is known to be detrimental to the worms. There was no evidence for cryptic species within *D. octaedra* and phylogenetic analyses showed some structure of lineages that may have diverged in historical glacial refugia. Other mechanisms, such as rare genetic exchange with closely related species or high mutation and dispersal rates may explain high genetic diversity in *D. octaedra*.

Heredity (2010) **105,** 341–347; doi:10.1038/hdy.2010.31;
published online 7 April 2010

Introduction

Dendrobaena octaedra is a polyploid earthworm that reproduces by apomictic parthenogenesis (Omodeo, 1955; Terhivuo and Saura, 1990). In this type of parthenogenesis eggs are formed through mitotic division, meiotic recombination cannot occur and genotypes are transmitted intact from mother to offspring. Despite the fact that offspring are genetic clones of their mothers, the genetic diversity of apomictic parthenogens can be high, allowing a potential for evolutionary response of populations to changes in environmental conditions (Maynard Smith, 1978; Schön *et al.*, 2009).

Previously, we compared clonal diversity of *D. octaedra* populations from sites with differing histories of heavy-metal contamination using allozyme markers (Haimi *et al.*, 2007). Diversity was high at all sites and also in areas with high levels of metal contamination. Although some genotypes were shared between different sites, most diversity was due to a large number of unique genotypes, represented by only one individual in the sample (63% of all genotypes). Similar results of allozyme variation in *D. octaedra* were also obtained by Terhivuo and Saura (1990). The clonal diversity of *D. octaedra* is higher than that of other parthenogenetic

earthworms: *Eiseniella tetraedra* having slightly lower diversity (Terhivuo *et al.*, 1994) and *Octolasion cyaneum* with very low diversity (Terhivuo and Saura, 2003); raising the question of how such diversity is generated and maintained.

Theoretically, high genetic diversity in apomictic parthenogens is made possible only through the generation of new genotypes by mutation. However, other mechanisms must be considered as many apomictic parthenogens have been found to not be strictly clonal. In some cases, the presence of rare males, new hybridization or polyploidization events and occasional sexual reproduction with closely related sexual species has been reported (Butlin *et al.*, 1999; Lushai and Loxdale, 2002). Indeed, recent work by Simonsen and Holmstrup (2008) provides evidence that *D. octaedra* may not be strictly clonal even though meiotic stages have never been observed in this species (Omodeo, 1955; Terhivuo and Saura, 2006). Somatic recombination also could create new genotypes, as is known for polyploid plants (Richards, 2003). However, as clonal lineages can persist to great age (Butlin, 2002), high levels of migration would allow mixing of divergent lineages, which could also explain the observed high genetic diversity of *D. octaedra* populations. Terhivuo and Saura (2006) expected that high diversity in *E. tetraedra* is due to their ability to disperse by hydrochory and hypothesized that *D. octaedra* are also highly efficient dispersers.

In addition to the possibilities just described, we must also consider a methodological explanation for our previous findings of high genetic diversity in *D. octaedra* that were based on allozyme markers (Haimi *et al.*, 2007).

Correspondence: Dr KE Knott, Department of Biological and Environmental Science, University of Jyväskylä, PO Box 35; 40014 Jyväskylä, Finland.
E-mail: emily.knott@jyu.fi

Because allozymes are surveyed at protein rather than DNA level, variation in banding patterns could be due to either simultaneous expression of multiple alleles of these genes in the polyploid genome or variable regulation of these loci. In this study, we compared mitochondrial DNA sequence diversity in a subset of the sample analyzed previously with allozymes. We chose to survey the cytochrome oxidase subunit I gene (*COI*) because of its sequence conservation and its use in DNA bar coding (Hebert *et al.*, 2003; Dupont, 2009) and in a previous study of mtDNA diversity in *D. octaedra* (Cameron *et al.*, 2008). As the mitochondrial genome does not recombine and is inherited maternally, any influence of potential unrecognized sexual reproduction should not be a concern. Also, the mtDNA haplotypes are expected to be uniform within individuals despite their polyploid nuclear genomes. So, unlike allozymes, estimates of genetic diversity from mtDNA are not affected by possible variable alleles or gene expression.

Materials and methods

Collection sites and methods are described in detail by Haimi *et al.* (2007). Three localities with metal contamination gradients in southern Finland were sampled, each represented by a contaminated site and an uncontaminated site (1–2 km or 4–8 km distant from the emission sources). Emission sources were a Cu-Ni smelter in Harjavalta (61°180 N, 22°090 E), a steel smelter in Imatra (61°120 N, 28°480 E) and a Zn plant in Kokkola (63°500 N, 23°080 E).

The samples had been used in allozyme analyses (see Haimi *et al.*, 2007) and then stored at −20 °C. DNA was extracted from the samples using standard chemical protocols (Gentra Puregene tissue kit; Qiagen, Hilden, Germany). Extracted DNA was somewhat degraded due to the previous sample treatment and some samples failed in subsequent DNA amplification and cycle sequencing reactions (average success rate was 65%). We obtained DNA sequence data of *COI* using a previously reported primer designed for annelids (SPE-CO1-A25; Admassu *et al.*, 2006) paired with the universal LCO 1490 primer designed by Folmer *et al.* (1994).

DNA amplification reactions were performed in a total volume of 20 µl with final concentrations: $1 \times$ magnesium-free buffer (Biotools, Madrid, Spain), 0.2 mM of each DNTP (Fermentas, St Leon-Rot, Germany), 3 mM $MgCl_2$ (Biotools), 1 µM each primer (TAG Copenhagen, Copenhagen, Denmark), 0.1 U Taq polymerase (Biotools) and approximately 0.5 ng genomic DNA. All amplification reactions were performed in 96-well plates (Axygen, Union City, CA, USA) using a Hybaid Touchdown (Thermo Fisher Scientific, Waltham, MA, USA) or Bio-Rad C1000 (Bio-Rad, Hercules, CA, USA) thermocycler fitted with hot lids and the following thermocycling protocol: initial denaturation at 94 °C for 2 min, 35 cycles of 94 °C for 15 s, 50 °C for 15 s, 72 °C for 45 s and a final extension at 72 °C for 2 min. An aliquot (3 µl) of the amplification products was separated in a 1% agarose:TA gel containing SYBR safe and visualized with a SafeImager blue light trans-illuminator (both from Invitrogen, Carlsbad, CA, USA).

Successfully amplified samples were purified with Exo-SAP enzymatic treatment and then used as templates for DNA sequencing. Cycle sequencing reactions used Applied Biosystem's BigDye v3.1 reagents following scaled-down versions of the manufacturer's protocol (16 µl). Sequencing reaction products were purified by ethanol precipitation and dissolved in 10 µl HiDi form-amide (Applied Biosystems, Foster City, CA, USA) and separated with an ABI 3130*xl* Genetic Analyzer. Applied Biosystem's Sequencing Analysis, version 5.2, program was used for base calling and sequence quality was examined. All samples were sequenced from both strands and base calling errors were checked by eye. Only high-quality sequences were retained in the analysis. Sequence data were exported and aligned in the Clustal W application in MEGA, version 4 (Tamura *et al.*, 2007), using default parameters.

Sequences were edited to equal lengths (441 bp) in MEGA, which was used also when creating different file types for use in other analysis programs. Initially, a haplotype network was generated with TCS (Clement *et al.*, 2000) using only the observed Finnish haplotypes. To produce a connected network, we fixed the connection limit at 40 steps. The Finnish haplotype data were also exported to DnaSP (Librado and Rozas, 2009) for calculation of haplotype diversity, nucleotide diversity, Tajima's *D* and pairwise mismatch distributions.

A broader data set was created containing the Finnish haplotypes as well as *D. octaedra COI* sequences published in GenBank. These included samples from Austria (DQ092895–DQ092897, Admassu *et al.*, 2006), Canada (EU035477–EU035492, Cameron *et al.*, 2008) and samples from Greenland and Denmark (FJ214234 and FJ214235). In addition, *COI* sequences from the related species *D. attemsi*, *D. clujensis* and *D. veneta* (FJ214224, FJ374778 and FJ214233, respectively) were included. Sequences from GenBank were obtained from amplifications using different primers and were different lengths than those we obtained in this study. The sequences were edited to be the same length (406 bp), which resulted in a reduction of the total number of haplotypes (that is, by shortening the sequence length, some of the previously reported haplotypes were made identical to each other and our Hap02 was made identical to Hap01). After reducing the data set in this way, 34 haplotypes of *D. octaedra* as well as sequences of outgroups *D. attemsi*, *D. clujensis* and *D. veneta* were included in phylogenetic analysis.

A test of evolutionary model using jModelTest version 0.1.1 (Posada, 2008) indicated that the general time reversible (GTR) + γ model best described the *COI* data set using the AICc whereas the Hasegawa–Kishino–Yano (HKY) + γ model was chosen using the BIC. Distances between haplotypes based on the Kimura two-parameter (K2P) + γ model were also calculated (using MEGA). Although the K2P model is not expected to be a good match to our data set, we chose to also use these distances for comparison to DNA bar-coding studies, which are typically based on K2P or even simpler models (for example, Hebert *et al.*, 2003; King *et al.*, 2008).

Several tree reconstruction methods were used to investigate phylogenetic relationships of the sample including: neighbor-joining analysis in MEGA (NJ), maximum likelihood analysis with PHYML (ML; Guindon and Gascuel, 2003) and Bayesian analysis with MrBayes (BA; Ronquist and Huelsenbeck, 2003). In ML and BA analyses, both GTR + γ and HKY + γ models were used with γ-shaped parameters estimated during the analysis. Because MEGA does not allow either of these models, we performed the analysis under a Tamura-Nei (TrN) + γ model. Here, the γ-shaped

parameter was set to 0.25, similar to its estimate in jModelTest. Bootstrapping was performed in NJ (1000 replicates) and ML (100 replicates) analyses. We also calculated GTR + γ and HKY + γ distance matrices for the data using TreeFinder (Jobb *et al.*, 2004) and input these matrices to MEGA for NJ analysis.

Results

A total of 24 *COI* haplotypes were detected in our sample of *D. octaedra* (named Hap01–Hap24; GenBank GU248222–GU248245; see Supplementary Material). The majority of the haplotypes were similar (differing by only a few bases), whereas a few rare haplotypes were

quite distinct, differing by more than 25 substitutions. No clustering of haplotypes from contaminated sites compared to those from uncontaminated sites was observed in our haplotype network (Figure 1). There was also no difference in the numbers of haplotypes (*t*-tests, all nonsignificant) nor in the distribution of haplotypes (*G*-tests, all nonsignificant) when comparing contaminated and uncontaminated sites within localities (Table 1). Hap01 was represented at all localities (both contaminated and uncontaminated sites) and was the most prevalent haplotype, representing close to half of the total number of successfully sequenced individuals (53 of 118; Supplementary Material). The majority (18 of 24) of the haplotypes were unique (found in only one
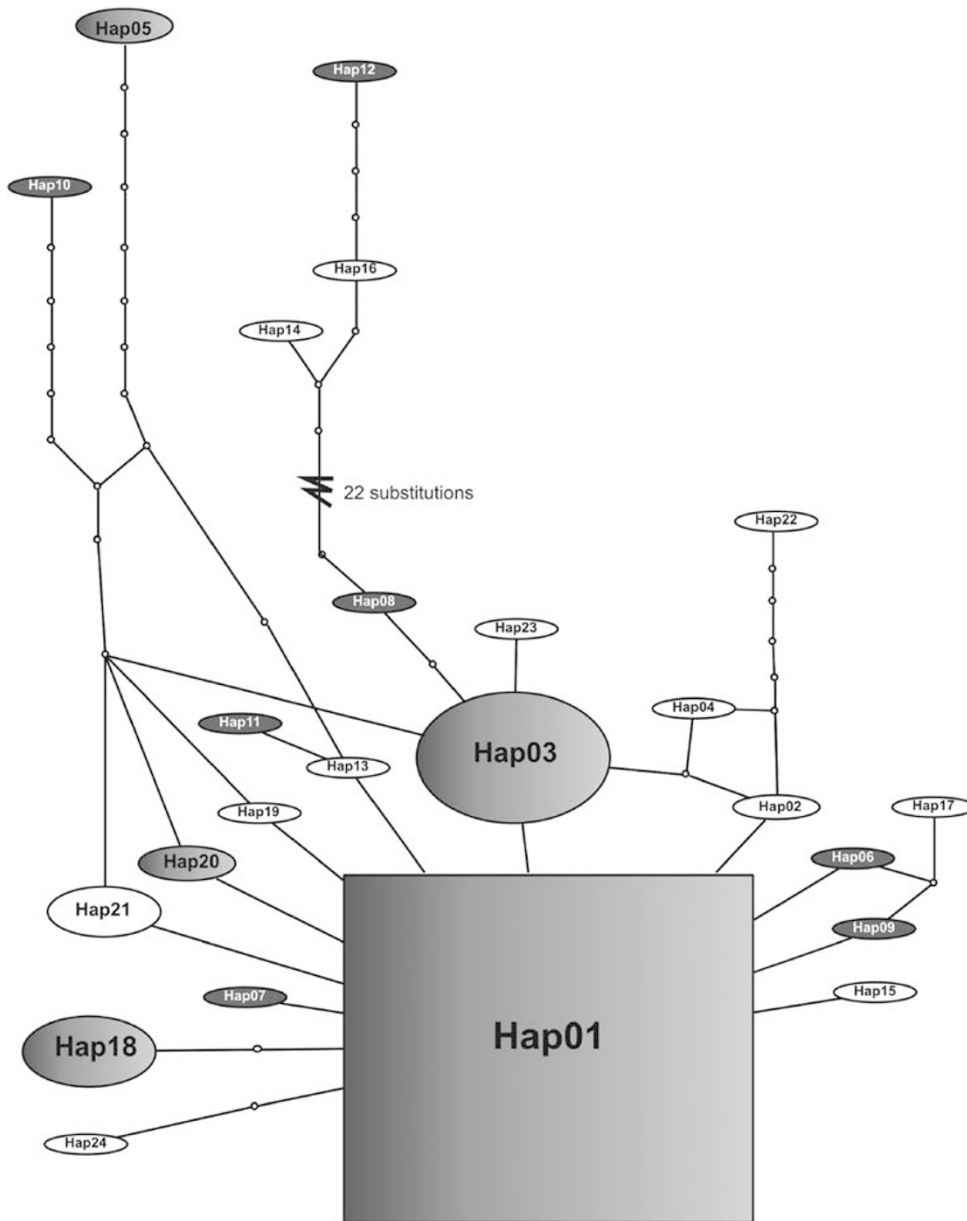

Figure 1 Haplotype network for 24 cytochrome oxidase subunit I gene (*COI*) haplotypes obtained from *Dendrobaena octaedra* sampled in Finland (Hap01–Hap24). Rectangle represents presumed ancestral sequence. Size of shapes indicates relative proportion of each haplotype in the sample. Unfilled shapes indicate haplotypes sampled from uncontaminated sites only; solid gray shapes indicate haplotypes sampled from contaminated sites only; shapes with a gradation of coloration indicate haplotypes found in both. Small unlabeled circles represent unsampled intermediate haplotypes. Branches between shapes indicate one substitution except where noted.

Table 1 Numbers of individuals successfully sequenced, haplotypes found and diversity statistics for contaminated (C) and uncontaminated (U) sites at three localities in Finland (see text for more information)

Locality/site	Imatra C	Imatra U	Harjavalta C	Harjavalta U	Kokkola C	Kokkola U	Total
Number of individuals	19	22	20	18	20	19	118
Number of haplotypes	10	8	3	3	4	8	24
Haplotype diversity	0.917	0.526*	0.689	0.242	0.689	0.813	0.714
Nucleotide diversity	0.02033	0.01378	0.00307	0.00057	0.00266	0.00507	0.00719
Tajima's D	−1.36464	−1.70060	1.60144	−1.49796	0.11864	−1.64060	−2.18387**

*Outside 95% confidence intervals produced by coalescent simulation; **$P < 0.01$.
Total represents results for the combined data set.

individual of one site). In Imatra and Harjavalta, uncontaminated sites were dominated by Hap01 only, whereas the nearby contaminated sites and both contaminated and uncontaminated sites in Kokkola showed a more even distribution with two or three common haplotypes (Supplementary Material). Owing to these differences in distribution, a significant difference among the three localities was observed (R × C test of independence with G-test; $G_{adj} = 79.98$, $P < 0.01$).

Haplotype diversity (0.714) was high. All sampled sites except the uncontaminated site in Imatra had haplotype diversity values that fell within 95% confidence intervals of empirical distributions created with coalescent simulation in DnaSP (Table 1). In total, nucleotide diversity, π, equaled 0.00719. Nucleotide diversity values also fell within 95% confidence intervals of empirical distributions, but the values differed among the localities (Table 1). Highest values were observed in Imatra (both contaminated and uncontaminated sites) due to the fact that the most divergent haplotypes were found at this site.

When each site was analyzed separately, Tajima's D statistics were all nonsignificant (Table 1). Most of these were negative, however, and a significant Tajima's D was found for the total data set (all sites combined). Mismatch distributions of the sample data from most sites are consistent with that expected for a stable population with constant growth. Exceptionally, sites in Imatra seemed to indicate secondary contact (Figure 2).

The parameters estimated for the GTR + γ and the HKY + γ models (chosen by AICc and BIC, respectively) are reported in the Supplementary Material. Because phylogenetic analysis with both models gave identical or nearly identical results, we describe only those obtained using the GTR + γ model. Use of the TrN + γ model rather than the more complex models in NJ analysis did not affect the results considerably, as evidenced by NJ analysis of GTR + γ and HKY + γ distance matrices (see Supplementary Material).

Phylogenetic analyses using all three tree reconstruction methods were largely congruent, each resulting in 3–4 groups, most representing clades (see A–D; Figure 3). *D. clujensis* and *D. veneta* were clearly outgroups that clustered together, whereas *D. attemsi* was always included in the ingroup, most closely with our Hap10 and one haplotype from Austria. The Hap10 and *D. attemsi* sequences were identical as was one sequence from Canada (H9 in Cameron *et al.*, 2008), which had been removed from the data set. The sequence from Austria only differed from Hap10 because of 4 bp missing data at the 5′ end and the two haplotypes may in fact be identical. The clade containing *D. attemsi* (B)
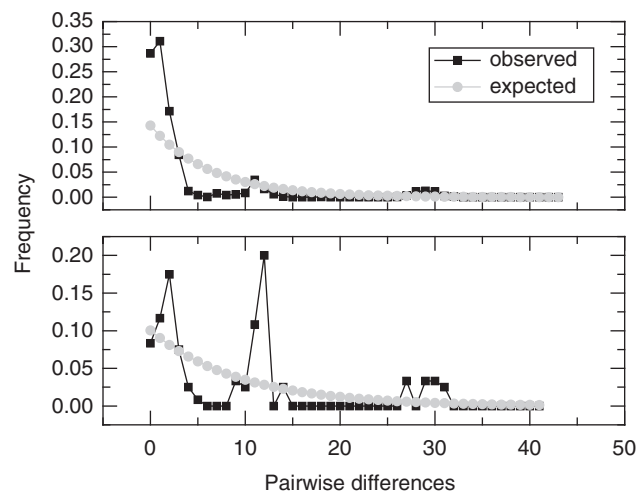

Figure 2 Mismatch distributions: (top) showing a close match to expectations for a stable population with constant growth, observed for the total data set (shown) and most of the sample sites; (bottom) showing deviation from expectations characteristic of secondary contact as observed for the contaminated site in Imatra (similar results, not shown, were observed for the uncontaminated site in Imatra).

was grouped with the largest clade (A) in NJ and ML analyses, but with low bootstrap support in the ML analysis. Within clade A there was very little divergence among the haplotypes, and here the most common Finnish (Hap01) and Canadian (H3CANADA) haplotypes were found, as well as other haplotypes from Austria, Denmark and Greenland. The most divergent haplotypes in the Finnish sample (Hap12, Hap14 and Hap16) always grouped together and in most cases this clade also grouped with the most divergent haplotype in the Canadian sample: H1CANADA (group D; a clade in BA and ML analyses but not in NJ analysis; but see also Supplementary Material). Clade C (two haplotypes from Canada and one from Finland) was either in a polytomy with the other clades (BA and ML) or more closely aligned with clades A and B (NJ), however this grouping may be an artifact of the NJ tree reconstruction method.

Discussion

Diversity in *COI* sequences
We found high mtDNA haplotype diversity (0.714) in the *COI* gene sequenced from *D. octaedra* from three localities in Finland with metal contamination gradients. The
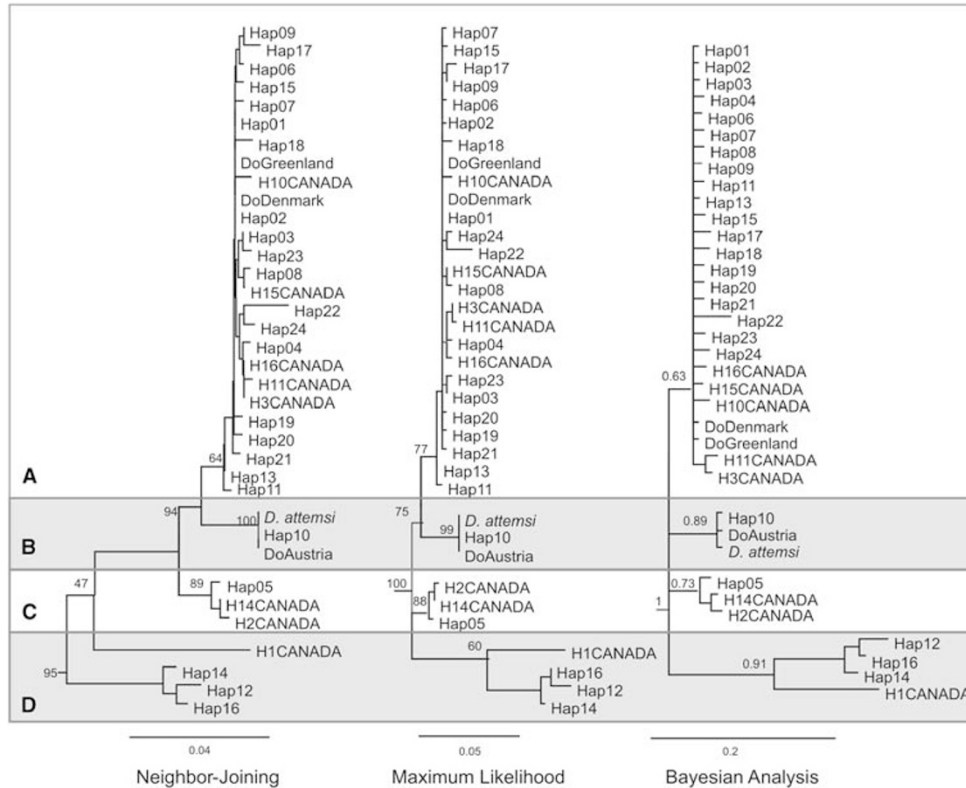
Figure 3 Results from phylogenetic analyses. Trees obtained using neighbor-joining, maximum likelihood and Bayesian analyses. Outgroups are removed for clarity. Groups A–D, are enclosed by boxes alternating with shading and labeled at left. Finnish haplotypes are labeled Hap01–Hap24. Canadian haplotypes not identical to Finnish haplotypes (see text) retain numbering as in Cameron *et al.* (2008), but with the designation 'CANADA' following the haplotype code, for example, 'H1CANADA'. Other haplotypes are named 'Do' followed by name of sample origin, for example, 'DoGreenland'. Clade support values listed at nodes are bootstrap values for NJ (1000 replicates) and ML (100 replicates) and posterior probabilities for BA.

haplotype diversity was similar to that observed for clonal diversity in our own and previous studies of *D. octaedra* in Finland using allozyme markers (average 0.76, Terhivuo and Saura, 2006; average 0.66, Haimi *et al.*, 2007). High allozyme clonal diversity was found also for *D. octaedra* populations in Sweden, Estonia and Poland (Terhivuo and Saura, 2006; Simonsen *et al.*, 2008), whereas populations in Greenland had somewhat lower clonal diversity (Liland Hansen *et al.*, 2006). Our study results indicate that high clonal diversity estimates obtained in previous studies are not an artifact of variable allozyme expression or sexual reproduction as high diversity is also reflected in mtDNA.

However, mtDNA haplotype diversity of *COI* from *D. octaedra* collected from multiple sites in Alberta, Canada was lower than that observed here, and included some populations represented by a single haplotype (diversity at the population level ranged from 0 to 0.64; Cameron *et al.*, 2008). It is important to note that our estimates of haplotype diversity were based on a smaller-sized portion of the *COI* gene, but a larger sample size per population than those of Cameron *et al.* (2008), and thus, they may not be directly comparable. Lower diversity values for the Canadian populations are expected, due to the limited and relatively recent introductions of *D. octaedra* to the Canadian forests. Our phylogenetic results confirm the conclusions of

Cameron *et al.* (2008) that diversity in Canadian populations is likely due to multiple independent introductions or introduction of a diverse sample of *D. octaedra*.

Identical *COI* haplotypes were found among the Finnish samples and from all the countries represented by the sequences obtained from GenBank. Indeed, the most common haplotype observed in Finland (Hap01) was identical with the Danish, two of the Austrian and four of the Canadian haplotypes (H5, H6, H7 and H13 in Cameron *et al.*, 2008) after comparing the same length of sequence. Even the most divergent haplotypes were found in both Finland and Canada (Canadian H9 = Finnish Hap10; others shown in Figure 3). However, the most common haplotype in Canada (H3) was not found in our samples.

Effects of metal contamination
As we found in our previous study (Haimi *et al.*, 2007), there was no effect of soil metal contamination on genetic diversity of *D. octaedra*. Contaminated and uncontaminated sites had similar high haplotype diversity and distribution, and similar levels of nucleotide diversity. Haplotype diversity ranged considerably among the sites (0.24–0.92, Table 1), but in most cases these values fell within 95% confidence intervals of the empirical distribution obtained by coalescent simulation. Only the

uncontaminated site at Imatra was outside this confidence interval. The Harjavalta sites had the lowest clonal diversities in our previous study and soil metal concentrations were also highest there, but worms from Imatra had highest metal body burdens (Haimi *et al.*, 2007). Regardless, there is little support for a detrimental effect of soil contamination on genetic diversity of *D. octaedra* and our data seem to indicate the opposite (with uncontaminated sites being somewhat less diverse than contaminated ones). Simonsen *et al.* (2008) also did not find strong evidence for an effect of soil metal contamination on genetic diversity of *D. octaedra* in their study of contamination gradients in Poland and Sweden.

These results are somewhat surprising as metal contamination is known to have detrimental effects on earthworm population densities and biodiversity (Bengtsson *et al.*, 1992; Spurgeon and Hopkin, 1999). However, due to decreasing emissions at least in our study localities, metal concentrations are lower in the litter layer, where *D. octaedra* is found, than deeper in the soil (Lukkari *et al.*, 2004). More significantly, *D. octaedra* is known to reduce its exposure to metals in contaminated environments through avoidance behavior (Lukkari and Haimi, 2005).

Cryptic species in *D. octaedra*?

A recent focus on applying molecular genetic tools to study earthworm populations (for example, Heethoff *et al.*, 2004; Pérez-Losada *et al.*, 2005; King *et al.*, 2008; Dupont, 2009) has raised the possibility that cryptic species may be more common within the Lumbricidae than previously thought. Indeed, King *et al.* (2008) identified at least four independent lineages within *Allolobophora chlorotica* alone, as well as potentially multiple lineages of *Aporrectodea longa*, *Apo. rosea* and *Lumbricus rubellus* from only a small sample of worms from the UK. Cryptic divergent lineages have been identified with sequence data even in the parthenogenetic lumbricid *O. tyrtaeum* (Heethoff *et al.*, 2004). Identification of cryptic diversity is complicated by similar morphologies observed among lumbricids and calls for further use of molecular tools in identifying biodiversity (Dupont, 2009). DNA bar-coding efforts have indicated that such tools are useful in this regard, with between-species divergences in the range of 13–18% for earthworms (Huang *et al.*, 2007; King *et al.*, 2008; Chang *et al.*, 2009).

We found that most *D. octaedra* COI haplotypes in Finland and elsewhere were very similar with few substitutions and that these cluster together in phylogenetic analyses (Figures 1 and 3, clade A). More divergent haplotypes were rare, but present in Finland, Austria and Canada. These haplotypes were basal to or in a polytomy with the more common haplotypes (Figure 3, groups B–D). Regardless of their divergence relative to clade A, the entire data set had divergence of only 3% (see Supplementary Material), which is within expectations of within species divergence in lumbricids (see King *et al.*, 2008).

Surprisingly, the *COI* sequence from *D. attemsi* was identical to *D. octaedra* haplotypes and fell well within the *D. octaedra* clade. *D. attemsi* and *D. octaedra* are morphologically superficially similar, but nonetheless distinguishable (Rota and Erséus, 1997). The *D. attemsi*

sample from which the *COI* data were obtained was collected from Washington, USA (see GenBank record FJ214224) and could represent an error of species identification. Such errors are not uncommon (for example, Chang *et al.*, 2009). Previous authors have questioned validity of recognizing *D. attemsi* as a separate species (see discussion and references in Rota and Erséus, 1997), but according to recent molecular phylogenies it should be more closely related to *D. clujensis* rather than *D. octaedra* (Csuzdi *et al.*, 2006). *D. attemsi* is not known from Finland, and it was only first noted in Scandinavia (that is, Stockholm, Sweden) in the last decade (Rota and Erséus, 1997).

The structure observed in our phylogenetic analyses is more likely due to persistent divergent lineages in *D. octaedra* rather than representing cryptic species. However, it is interesting to note that all of the most divergent haplotypes obtained in our study (Hap05, 10, 12, 14 and 16) were found only in Imatra (both contaminated and uncontaminated sites). Imatra is situated in the southeast, near the border of Finland and Russia, and close to the Karelian Isthmus, the postglacial dispersal pathway of both flora and fauna to Finland (Terhivuo and Saura, 1990, 1996). It is possible that the 'divergent' haplotypes observed here are more common in Russian populations that have recently migrated or otherwise been introduced to southeastern Finland. Indeed this may be the case as mismatch distributions of the Imatra populations were consistent with secondary contact (Figure 2). The higher π values observed for the Imatra sites were also likely the result of admixed populations and contributed to the significant Tajima's *D* value for the total sample. Negative *D* represents an excess of rare mutations, which can be used to infer purifying selection or a selective sweep and can indicate an admixed population (Tajima, 1989; Stajich and Hahn, 2005).

Conclusions

We observed high mtDNA diversity in the apomictic parthenogen *D. octaedra*, similar to that observed for allozyme markers. Although clonal, these earthworms are genetically diverse. Several different mechanisms could account for high diversity in *D. octaedra*. These results suggest that our previous estimates of diversity made from allozyme markers were not due to variable expression of alleles in the polyploid genome or sexual reproduction. Other possibilities, such as genetic exchange with closely related species, high mutation and dispersal rates, and frequency-dependent selection among lineages, have yet to be explored. Sequence diversity at *COI* is well within that expected for a single species, and likely *D. octaedra* does not contain cryptic taxa. Rare, divergent haplotypes found in our data set are more likely due to either multiple historical polyploidization events or divergence of historical lineages isolated in glacial refugia.

Conflict of interest

The authors declare no conflict of interest.

References

Admassu B, Juen A, Traugott M (2006). Earthworm primers for DNA-based gut content analysis and their cross-reactivity in a multi-species system. *Soil Biol Biochem* **38**: 1308–1315.

Bengtsson G, Ek H, Rundgren S (1992). Evolutionary response of earthworms to long-term metal exposure. *Oikos* **63**: 289–297.

Butlin RK (2002). Evolution of sex: the costs and benefits of sex: new insights from old asexual lineages. *Nat Rev Genet* **3**: 311–317.

Butlin RK, Schön I, Martens K (1999). Origin, age, and diversity of clones. *J Evol Biol* **12**: 1020–1022.

Cameron EK, Bayne EM, Coltman DW (2008). Genetic structure of invasive earthworms *Dendrobaena octaedra* in the boreal forest of Alberta: insights into introduction mechanisms. *Mol Ecol* **17**: 1189–1197.

Chang C-H, Rougerie R, Chen J-H (2009). Identifying earthworms through DNA barcodes: pitfalls and promise. *Pedobiologia* **52**: 171–180.

Clement M, Posada D, Crandall K (2000). TCS: a computer program to estimate gene genealogies. *Mol Ecol* **9**: 1657–1660.

Csuzdi CS, Pop AA, Pop VV, Zicsi A, Wink M (2006). Revision of the *Dendrobaena alpina* (Rosa) species group (Oligochaeta, Lumbricidae). In: Pop AA, Pop VV (eds). *Advances in Earthworm Taxonomy II. Proceedings IOTM2*. Cluj University Press: Romania. pp 119–128.

Dupont L (2009). Perspectives on the application of molecular genetics to earthworm ecology. *Pedobiologia* **52**: 191–205.

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 Biol Biotechnol* **3**: 294–299.

Guindon S, Gascuel O (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Sys Biol* **52**: 696–704.

Haimi J, Knott KE, Yliniemi I (2007). Does metal contamination affect clonal diversity of the parthenogenetic earthworm *Dendrobaena octaedra*? *Eur J Soil Biol* **43**: S268–S272.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003). Biological identifications through DNA barcodes. *Proc R Soc London B Biol Sci* **270**: 313–321.

Heethoff M, Etzold K, Scheu S (2004). Mitochondrial COII sequences indicate that the parthenogenetic earthworm *Octolasion tyrtaeum* (Savigny 1826) constitutes two lineages differing in body size and genotype. *Pedobiologia* **48**: 9–13.

Huang J, Xu Q, Sun ZJ, Tang GL, Su ZY (2007). Identifying earthworms through DNA barcodes. *Pedobiologia* **51**: 301–309.

Jobb G, von Haeseler A, Strimmer K (2004). TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol* **4**: 18; doi:10.1186/1471–2148–4–18.

King RA, Tibble AL, Symondson WOC (2008). Opening a can of worms: unprecedented sympatric cryptic diversity within British lumbricid earthworms. *Mol Ecol* **17**: 4684–4698.

Librado P, Rozas J (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451+.

Liland Hansen P, Holmstrup M, Bayley M, Simonsen V (2006). Low genetic variation for *Dendrobaena octaedra* from Greenland compared to populations from Europe and North America: refuge or selection? *Pedobiologia* **50**: 225–234.

Lukkari T, Taavitsainen M, Väisänen A, Haimi J (2004). Effects of heavy metals on earthworms along contamination gradients in organic rich soils. *Ecotoxicol Environ Saf* **59**: 340–348.

Lukkari T, Haimi J (2005). Avoidance of Cu and Zn contaminated soil by three ecologically different earthworm species. *Ecotoxicol Environ Saf* **62**: 35–41.

Lushai G, Loxdale HD (2002). The biological improbability of a clone. *Genet Res* **79**: 1–9.

Maynard Smith J (1978). *The Evolution of Sex*. Cambridge University Press: Cambridge.

Omodeo P (1955). Cariologia dei Lumbricidae, II Contributo. *Caryologia* **8**: 135–178.

Pérez-Losada M, Eiroa J, Mato S, Domínguez J (2005). Phylogenetic species delimitation of the earthworms *Eisenia fetida* (Savigny, 1826) and *Eisenia andrei* Bouche', 1972 (Oligochaeta, Lumbricidae) based on mitochondrial and nuclear DNA sequences. *Pedobiologia* **49**: 317–324.

Posada D (2008). jModelTest: phylogenetic model averaging. *Mol Biol Evol* **25**: 1253–1256.

Richards AJ (2003). Apomixis in flowering plants: an overview. *Phil Trans R Soc London B* **358**: 1085–1093.

Ronquist F, Huelsenbeck JP (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.

Rota E, Erséus C (1997). First record of *Dendrobaena attemsi* (Michaelsen) (Oligochaeta, Lumbricidae) in Scandinavia, with a critical review of its morphological variation, taxonomic relationships and geographical range. *Ann Zool Fenn* **34**: 89–104.

Schön I, Martens K, van Dijk P (eds) (2009). *Lost Sex. The Evolutionary Biology of Parthenogenesis*. Springer: Berlin.

Simonsen V, Holmstrup M (2008). Deviation from apomictic reproduction in *Dendrobaena octaedra*? *Hereditas* **145**: 212–214.

Simonsen V, Laskowski R, Bayley M, Holmstrup M (2008). Low impact of metal pollution in the earthworm *Dendrobaena octaedra* measured by allozymes. *Pedobiologia* **52**: 51–60.

Spurgeon DJ, Hopkin SP (1999). Seasonal variation in the abundance, biomass and biodiversity of earthworms in soils contaminated with metal emissions from a primary smelting works. *J Appl Ecol* **36**: 173–183.

Stajich JE, Hahn MW (2005). Disentangling the effects of demography and selection in human history. *Mol Biol Evol* **22**: 63–73.

Tajima F (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.

Tamura K, Dudley J, Nei M, Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**: 1596–1599.

Terhivuo J, Saura A (1990). Allozyme variation in parthenogenetic *Dendrobaena octaedra* (Oligochaeta: Lumbricidae) populations of eastern Fennoscandia. *Pedobiologia* **34**: 113–139.

Terhivuo J, Saura A (1996). Clone pool structure and morphometric variation in endogeic and epigeic North-European parthenogenetic earthworms (Oligochaeta: Lumbricidae). *Pedobiologia* **40**: 226–239.

Terhivuo J, Saura A (2003). Low clonal diversity and morphometrics in the parthenogenetic earthworm *Octolasion cyaneum* (Sav.). *Pedobiologia* **47**: 434–439.

Terhivuo J, Saura A (2006). Dispersal and clonal diversity of North-European parthenogenetic earthworms. *Biol Invasions* **8**: 1205–1218.

Terhivuo J, Saura A, Hongell K (1994). Genetic and morphological variation in the parthenogenetic earthworm *Eiseniella tetraedra* (Sav.) (Oligochaeta: Lumbricidae) from South Finland and North Norway. *Pedobiologia* **38**: 81–96.

Supplementary Information accompanies the paper on Heredity website (http://www.nature.com/hdy)