Accepted on 1 February 2008
J Zool Syst Evol Res doi: 10.1111/j.1439-0469.2008.00471.x

[1]*Laboratoire Evolution et Diversité Biologique, Université Paul Sabatier, Toulouse cedex 4, France;* [2]*Unité de Recherche, Biologie, Ecologie et Parasitologie des organismes aquatiques, Faculté des Sciences, Tunis, Tunisie;* [3]*ECOLAB, Université Paul Sabatier, Toulouse cedex 4, France*

Genetic diversity of *Ligula intestinalis* (Cestoda: Diphyllobothriidea) based on analysis of inter-simple sequence repeat markers

W. Bouzid[1, 2], S. Lek[1], M. Mace[1], O. Ben Hassine[2], R. Etienne[1], L. Legal[3] and G. Loot[1]

Abstract

In order to investigate the genetic diversity of *Ligula intestinalis* populations, nine inter-simple sequence repeat (ISSR) markers were applied to populations from nine geographical areas around the world and 10 host species. The 110 loci selected from the ISSR patterns produced revealed high variability among the analysed samples, with a polymorphism of 100% and a global coefficient of gene differentiation estimated by Nei's index (G_{ST}) of 0.776. Major genetic differentiation was found to be correlated to five broad geographical regions (Europe, China, Canada, Australia and Algeria). Nevertheless, no significant genetic variation was found among European isolates, although they originated from disparate geographical localities and/or unrelated hosts. Classical classification methods: maximum parsimony and factorial correspondence analysis were compared with an advanced statistical method: the self-organizing map (SOM). The results demonstrated that the ISSR approach is rapid and inexpensive and provides reliable markers to assess genetic diversity of *L. intestinalis*. Furthermore, SOM artificial neuronal networks are considered to provide an efficient alternative tool for mapping the genetic structures of parasite populations.

Key words: inter-simple sequence repeat – self-organizing map – genetic diversity – parasite clustering – *Ligula intestinalis* – geographic isolation

Introduction

Host–parasite systems provide useful models for studying evolutionary problems (Price 1980; De Meeus et al. 1998; Paterson and Banks 2001). However, most studies failed to show clear coevolutionary patterns, suggesting that consideration should also be given to the history of colonization and secondary patterns of dispersal simultaneously (Hoberg 1997; Wickström et al. 2003). The development of the molecular tools in the past decades has stimulated systematic researches. Nevertheless, work concerning the parasites is far from being at the level of that of the hosts, especially, as the use of the molecular data for the study of the phylogenetic relationships and the genetic characterization of the populations of parasites were largely limited to the species being of medical or economic interest (Poulin and Morand 2000). Although a fair number of genes has been used to study parasitic phylogeny (Olson and Tkach 2005), far less has been done to develop population-level markers.

In this study, we examined the genetic diversity of the parasite *Ligula intestinalis* Linnaeus, 1758, which undergoes a complex life cycle through three different hosts: a planktonic copepod (first intermediate host), a freshwater fish (second intermediate host) and a piscivorous bird, which represent the definitive host (Rosen 1918). In the parasite life cycle, the plerocercoid stage inside the fish host is the longest stage (up to 2 years) and may cause castration, modify growth rate and behaviour, and induce mortality in natural fish populations (Orr 1966; Dubinina 1980; Loot et al. 2002a; b). The parasite is more commonly encountered and sampled in the larval stage, and hence, morphological studies are not possible, as they are done on adult reproductive complexes or fully developed plerocercoids. Moreover, few number of morphological characteristics are available for species identification (Dubinina 1980).

Ligula intestinalis has been reported from a broad range of fish families, such as Cyprinidae, Catostomidae, Salmonidae or Galaxiidae (e.g. Dubinina 1980; Bean and Winfield 1992; Groves and Shields 2001; Museth 2001; Barus and Prokes 2002; Chapman et al. 2006). However, these available records show a conspicuous heterogeneity in host preference according to the geographical area studied. For instance, in South-West France, it has been demonstrated that *L. intestinalis* is highly restricted to roach populations (*Rutilus rutilus* Linnaeus, 1758), where other known potential hosts co-exist in the same locality (Loot et al. 2002a). In its adult phase, the parasite seems to be less specific to its final host (bird) (Dubinina 1980). Indeed, several families of bird eating fish are potential hosts for *Ligula* (e.g Phalacrocoracidae, Podicipedidae, Ardeidae, Laridae).

Given the extensive host and geographic ranges of this parasite, distinct strains/species are supposed to exist. Indeed, recent studies support such a view by the recovery of genetically distinct plerocercoid isolates (Olson et al. 2002; Li and Liao 2003; Logan et al. 2004). For instance, Olson et al. (2002) used the entire ITS region and the partial ribosomal 28S and found reliable genetic differences of *Ligula* populations between two sympatric fish hosts, namely roach (*R. rutilus*) and the gudgeon (*Gobio gobio* Linnaeus, 1758) in Northern Ireland. Differences in pathology owing to *Ligula* infection between these two sympatric fish hosts were also reported (Arme 1997). Facing the putative biological complexity of *L. intestinalis* plerocercoids, the lack of adequate molecular data and comprehensive sampling has been highlighted by a number of authors (Logan et al. 2004; Stefka et al. 2007).

In the present study, we explored the usefulness of inter-simple sequence repeat (ISSR) markers to determine the genetic relationship of several morphologically indistinguishable *Ligula* plerocercoid specimens. The polymerase chain reaction (PCR)–ISSR is a relatively novel technique used to screen a large part of the genome without prior knowledge of sequences. The method provides highly reproducible results and generates abundant polymorphisms in many systems. It is

proven to be efficient in distinguishing between populations and closely related species (Zietkiewicz et al. 1994; Robinson et al. 1997; Wolfe et al. 1998a,b; Hundsdoerfer and Wink 2006; Maltagliati et al. 2006). Although the approach has limitations like almost all dominant markers in terms of diallelic interpretation of generated fingerprinting (Wolfe and Liston 1998), yet it has the advantage of being technically simple to perform compared with some other molecular approaches. Furthermore, it overcomes some of the constraints of random amplified polymorphic DNA (RAPD) by using longer primers, allowing more stringent annealing temperatures. In addition, the target regions by ISSR yield higher polymorphism and reproducibility (Fang and Roose 1997; Luque et al. 2002; Wu et al. 2005). Among other advantages, the short time required to obtain results and the highly informative nature were noted (Graham et al. 1994).

This ISSR technique has been used mainly for the assessment of relationships between plant species and populations (Huang and Sun 2000; Joshi et al. 2000; Lanham et al. 2000; Treutlein et al. 2003a,b, 2005; Gobert et al. 2006). To date, its application to parasites has been limited to *Trypanosoma cruzi*, *Schistosoma mansoni*, *Leishmania braziliensis* and *Trichinella* genotypes (Oliveira et al. 1997; Gomes et al. 1998; Fonseca-Salamanca et al. 2006). To our knowledge, ISSR has not been applied to cestodes, although Oliveira et al. (1997) indicated that it should be included in the 'toolbox' for epidemiology studies, particularly for analyses of genetic variability. According to Behura (2006), the ISSR technique represents one of the most promising tools in population genetic studies and deserves increased attention (Esselman et al. 1999). However, the relevance of this approach for phylogenetic studies has been discussed, particularly when comparing genera, tribes or families (Simmons et al. 2007).

Traditional statistical methods, such as the unweighted pair-group method with arithmetic averages (UPGMA) is most of the time used for reconstructions based on multilocus DNA markers (Mace et al. 1999; Koopman et al. 2001). Nevertheless, this method has the bias of being sample-order dependent and can have the limitation, when applied to evolutionary processes, that there is no search for an optimal tree. Therefore, alternative methods are being increasingly sought or used in order to achieve more confidence/reliability in data processing and results. Among these, artificial neural networks (ANN), and in particular, Kohonen self-organizing maps (SOM) have been used extensively in biological research for pattern recognition and may provide an alternative to conventional statistical methods because they detect non-linear relationships, allow the visualization of complex data and remain robust despite experimental variation. Phylogenetic reconstruction, classification of proteins and genomic analysis are a few applications of SOM in molecular biology. Ecological applications for SOM are also being explored for classification and modelling of populations and ecosystems (Lek 2000; Park et al. 2003). To our knowledge, only a few recent studies have used SOM for genetic data, particularly for multilocus molecular analyses (Giraudel et al. 2000; Ruanet et al. 2005; Zhao et al. 2005; Roux et al. 2007).

The aims of the present study were: (i) to assess the genetic variability of *L. intestinalis* plerocercoids from different hosts and geographical regions; (ii) to support or to invalidate the existence of a complex of species represented by *L. intestinalis*; and (iii) to assess the usefulness of a recent, non-linear statistical approach, the SOM, as an alternative clustering method to the traditional approaches of maximum parsimony (MP) and factorial correspondence analysis (FCA).

Materials and Methods

DNA extraction

In order to assess the genetic diversity of *L. intestinalis* populations, a total of 159 specimens were collected from a range of countries and from different fish species. Details of location and host fish range are given in Table 1. Genomic DNA was extracted by small-scale sodium-dodecyl-sulphate (SDS)-proteinase K digestion of 10–20 mg of *Ligula* larval tissue (either ethanol fixed or fresh/frozen) and column-purification (DNA Clean-Up kit; Promega, Madison, WI, USA).

ISSR–PCR amplification

The PCR was carried out in a volume of 20 µl consisting of 20 ng of genomic DNA template, 2 µl of $MgCl_2$ (15 mM), 2 µl of dNTP (10 mM), 2 µl of primer (50 µM), 0.25 U Taq DNA polymerase (M166A; Promega) and 2 µl of Taq DNA polymerase buffer (10 mM Tris-HCl, pH 9.0, 50 mM KCl and 0.1% Triton® X-100). Cycling was performed in a T3 thermocycler Biometra with the following assay conditions: 4 min at 94°C, 39 cycles of 45 s at 94°C, 45 s at the annealing temperature depending on the primer used (see Table S1 in the supporting information), 2 min at 72°C and a final extension of 10 min at 72°C.

In this study, we tested a range of nine primers (Table S1). The annealing temperature of each primer was optimized by testing several temperatures between 50 and 60°C using a Mastercycler Gradient thermocycler (Eppendorf, Hamburg, Germany). The selection of optimal annealing temperatures is central to achieving optimum profile quality and reproducibility of ISSR fingerprints.

In order to test the reproducibility of ISSR, the patterns of 15 *L. intestinalis* DNA samples were tested three times on three different days using the same primers. Electrophoresis was performed on 2% agarose gels in 1 × TAE buffer (40mM Tris–NaOH, 20mM acetate, 1 mM EDTA). The gels sized 9 × 11 cm (Apelex, Massy, France) were stained with ethidium bromide and the run time (about 3 h) was kept constant between gels at 110 V. Digital images of banding patterns were captured over ultraviolet light using a camera supported by BIO-capt 97.03 (Vilber Lourmat, Marne-la-vallée, France).

Analyses of molecular data

Only bands which displayed reproducibility on multiple independent runs were considered. The intensity of the bands was not taken into account for general scoring. They were scored qualitatively as present (1) or absent (0), and both polymorphic and monomorphic bands were considered in the final data set. Binary data from all retained primers were gathered to perform the analysis with samples on the rows and markers on the columns.

We used the resultant presence/absence matrix, under the Hardy–Weinberg equilibrium to calculate the percentage of polymorphic bands (P), Shannon's information index (i) (Lewontin 1972), Nei's total gene diversity (Ht), Nei's gene diversity within each population (He), Nei's gene diversity among populations (Hs) and the coefficient of gene differentiation (G_{ST}). The two latter parameters were examined for the five populations of the European group, including samples from Tunisia. Nei's genetic identity (I) and genetic distance (D) were examined for all pairwise comparisons between populations. All these parameters were calculated using the program POPGENE version 1.32 (Yeh et al. 1997).

Genetic differentiation between and among populations was assessed by analysis of molecular variance analysis (AMOVA) using the software package Arlequin 3.000 (Schneider et al. 2000). AMOVA was carried out using a matrix of Euclidean distances with 10 000 permutations. Analyses were conducted in a hierarchical framework; populations from the same geographical origin were defined as specimens belonging to the same group (e.g. all samples from Australia were gathered in the same group regardless of their sampling locality

Table 1. Specimens of *Ligula intestinalis* collected from various host species and countries for the present study

Country of origin	Collection locality	Host	Coordinates
France	Lavernose (9)	*Rutilus rutilus* Linnaeus, 1758 (Cyprinidae)	43°23'N, 1°15'E
	Pareloup (8)	*R. rutilus*	44°12'N, 2°44'E
	Muret (8)	*R. rutilus*	43°27'N, 1°19'E
	Creteil (9)	*R. rutilus*	48°46'N, 2° 28'E
Germany	Müggelsee (6)	*R. rutilus*	52°26'N, 13°37'E
Czech Republic	Nové Mlýny (8)	*Abramis brama* Linnaeus, 1758 (Cyprinidae)	48°51'N, 16°43'E
	Lipno (6)	*R. rutilus*	48°37'N, 14°12'E
	Zelivka (7)	*R. rutilus*	49°16'N, 17°28'E
Russia	Rybinsk (9)	*A. brama*	58°10'N, 38°32'E
Tunisia	Sidi Salem (23)	*Rutilus rubilio* Bonaparte, 1837 (Cyprinidae)	36°40'N, 9°25'E
	Sidi Salem (6)	*Scardinius erythrophtalmus* Linnaeus, 1758 (Cyprinidae)	
Algeria	Hamiz (7)	*Barbus* sp. Cuvier and Cloquet, 1816 (Cyprinidae)	28º00'N, 3º00'E
Canada	Dumbo (22)	*Semotilus atromaculatus* Mitchill, 1818 (Cyprinidae)	46°06'N, 74°02'W
	Dalpec (1)	*Coulsius plumbeus* Agassiz, 1850 (Cyprinidae)	46°06'N, 74°02'W
China	Tian'e Zhou (22)	*Neosalanx taihuensis* Chen, 1956 (Salangidae)	29°43'N, 112°24'E
Australia	Goodga (1)	*Galaxias truttaceus* Cuvier and Valenciennes, 1846 (Galaxiidae)	34°57'S, 118° 05'E
	Goodga (1)	*Galaxias maculatus* Jenyns, 1842 (Galaxiidae)	
	Moates (3)	*G. truttaceus*	34°57'S, 118° 6'E
	Moates (4)	*G. maculatus*	

Numbers within brackets correspond to the population sampling size per locality.

or host species). The AMOVA hierarchy consisted of five groups (Europe, China, Algeria, Canada and Australia). Each group was composed of one population, whereas the European group was composed of five populations (France, Czech Republic, Germany, Russia and Tunisia). The Tunisian population was included in this group because of a European origin of this population (Losse et al. 1991) and according to a close similarity in their ISSR pattern with populations of European group. However, we also carried out AMOVA analysis where Tunisian population was treated as one independent group.

Parsimony-based analyses

All analyses were performed on Paup 4.0b10 (Swofford 2001). Starting trees were obtained by stepwise addition. We checked that random addition of taxa did not lead to alternate more parsimonious trees. All heuristic searches for optimal trees were carried out by tree bissection–reconnection (TBR) branch swapping with option MULPARS in effect.

Goloboff's method was applied (constant $k = 4$ and 500 replicates) (Goloboff 1993). In this non-iterative approach to homoplasy-based weighted parsimony, a tree is sought that maximizes the sum of the weights of individual sites. Consensus tree was obtained on a majority rule with a 50% limit.

Bootstrap values are calculated under the same criteria including Goloboff correction. Only 100 repeats were performed as the calculation was time consuming with such parameters (72 h for 100 repeats).

Tree was displayed using TreeView 1.5 software (Page 1996). The binary data from the ISSR fingerprinting was also subjected to FCA implemented in R software (Ihaka and Gentleman 1996).

SOM

A fundamentally new method of genetic polymorphism estimation using ANN technologies, namely SOM, was employed. Selected articles describe the details of the SOM algorithm (Chon et al. 1996; Kohonen 2001) and its applications (Lek 2000; Park et al. 2003); functions used herein are from the SOM toolbox (http://www.cis.hut.fi/projects/somtoolbox/) operating in a Matlab environment (MathWorks 2001). This powerful and adaptive ANN method uses an unsupervised learning algorithm that is efficient in modelling complex non-linear relationships. The aim of SOM is to perform a non-linear projection of the multi-dimensional data space onto 2D space called the Kohonen map (Kohonen 1982, 2001). It has the properties of neighbourhood preservation of the input data. The algorithm consists of two layers:

1. An input layer formed by a set of units called neurons associated with a vector (x) represented by the samples of the data, previously randomly mixed. There are as many neurons in this layer as elements in the samples.
2. An output layer formed by a set of neurons represented most of the time by a hexagonal grid (i.e. the Kohonen map).

The two layers are connected by weights (W) between neurons of the output layer and neurons of the input layer. The distance between the weights and input vectors is computed using Euclidian distance. The neuron that has competed with all neurons of the output layer for which the distance was minimal is the winner. This winning neuron called the 'best matching unit' (BMU) is selected with its neighbours, whose size is defined with a neighbourhood function, and is updated during the learning process.

The learning process trains the network to pattern the input vectors and is stopped, usually when weight vectors stabilize or when the number of iterations are completed (Kohonen 2001).

The SOM learning rule is summarized as follows:

$$W_{ij}(t+1) = W_{ij}(t) + \alpha(t).h_{jc}(t)[x_i(t) - W_{ij}(t)]$$

$W_{ij}(t)$: weight between neuron i in the input layer and neuron j in the output layer in an iteration time t

$\alpha(t)$: learning-rate factor which is a decreasing function of the iteration time t

h_{jc}: a function that defines the neighbourhood size of the wining neuron c

$x_i(t)$: vector chosen as an input vector at iteration time t

In this study, SOM was trained with different number of output neurons in order to select the optimum Kohonen map to classify the samples. A topographic error is computed for each map as an indicator of the topology preservation. SOM preserves the neighbourhood so that the samples classified in the same output neuron are considered similar. Moreover, the samples that are neighbours on the map are also expected to be more similar to each other.

The selected Kohonen map was then subdivided into different groups by using a hierarchical cluster analysis. To do this, the weight vectors associated with each output neuron of the trained Kohonen map were used.

Finally, to test whether there is a statistically significant difference between the groups defined by the hierarchical cluster analysis, an analysis of similarity (ANOSIM) (Clarke 1993) was performed between each pair of groups. This test involves a dissimilarity matrix in which the distances have been converted to ranks such that the smallest distance has a rank (r) of 1. The dissimilarity matrix was computed by using data on the presence or absence of 110 loci.

The ANOSIM statistics R is based on the difference of mean ranks between groups (rB) and within groups (rW):

$$R = (\mathrm{rB} - \mathrm{rW})/(N(N-1)/4)$$

N = total number of individuals

The ANOSIM statistics R range from 0 to 1. A 0 indicates that there is no difference between groups (i.e. the null hypothesis), while a 1 indicates that all samples within groups are more similar to one another than any samples from different groups (Clarke 1993). The statistical significance of observed R is assessed by permuting the grouping vector (1000 permutations) to obtain the empirical distribution of R under the null hypothesis (Ihaka and Gentleman 1996).

Results

Genetic structure and diversity

A PCR-based ISSR analysis was performed using nine primers; only four primers yielded polymorphism and defined profiles (or phenotypes, which we refer to as 'genotype' henceforth) (Table S1). A total of 110 reliable bands were recorded for the 159 *Ligula* samples. The size of the bands displayed ranged from 250 to 1500 bp (Fig. 1). The four primers provided different patterns and number of bands but gave almost the same percentage of polymorphism. The number of scored bands were 24, 26, 29 and 31 fragments for WB$(GACA)_4$, $(GACA)_4$WB, $(ACA)_5$BDB and BDB $(ACA)_5$, respectively.

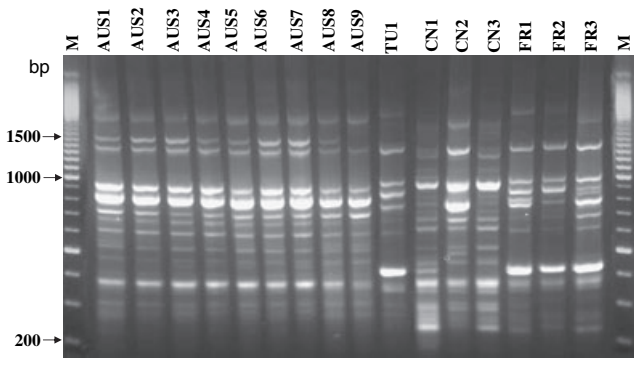

Fig. 1. Inter-simple sequence repeat amplification pattern obtained for DNA of various populations of *Ligula intestinalis* using $(GACA)_4$WB primer. AUS1–AUS9 refer to *Ligula* populations from Australia, TU1 from Tunisia, CN1–CN3 from China and FR1–FR3 from France. M refers to molecular weight marker

The total polymorphism (P) scored between populations from different geographical regions was 100%, whereas less polymorphism was detected among individuals within each group. Individuals from China showed the highest polymorphism (55.45%), whereas individuals from Germany showed the lowest rate (4.55%) (see Table S2 in the supporting information). The total gene diversity (Ht) was 0.293 ± 0.019, gene diversity within populations (Hs) was 0.065 ± 0.003 and the global coefficient of gene differentiation (G_{ST}) was 0.776. For the European group, the global coefficient of gene differentiation (G_{ST}) was 0.230. Nei's gene diversity within each population (He) and Shannon's index (i) are summarized in Table S2.

Nei's genetic identity (I) ranged from 0.525 to 0.999 (see Table S3 in the supporting information). Genetic distances (D) among specimens from European and Tunisian origin were minimal (Table S3). These specimens belong however to different locations and/or different host species (Table S1).

The distances in the European-Tunisian group varied from 0.0002 between French and Czech populations to 0.0294 between German and Russian ones (Table S3). Genetic distances were high between European, Chinese, Algerian, Australian and Canadian populations. The Chinese population had the greatest genetic distances from all other countries.

The hierarchical AMOVA analysis provided additional evidence for the geographical genetic structure between European, Chinese, Algerian, Australian and Canadian groups. Highly significant genetic differences were detected among these five groups (83.06%, D.F. = 4, p = 0.00891). A very low variability was expressed among populations within groups (0.41%, D.F. = 4, p = 0.00129). Finally, a low variation within populations was found (16.53%, D.F. = 150, $p = 10^{-5}$). AMOVA analysis considering Tunisian population as one separate group from Europe was not significant among groups (78.79%, D.F. = 5, p = 0.03119).

Clustering

Relationships displayed by MP clustering and FCA were concordant and showed clearly structured relationships among groups of *Ligula* specimens. Five main clusters were defined for the 10 populations included in this study (Fig. 2a,b). The MP tree showed a divergence of groups, with a clear differentiation between analysed populations from China, Canada, Australia, and Algeria. The tree grouped all the five European populations in the same clade with Tunisian individuals also belonging to this group.

The MP showed a well-defined clustering of the European group, whereas the Chinese one seems to be paraphyletic and structured as an important radiation.

Data on the presence or absence of 110 loci were presented to the input layer of the SOM (as input data set). A Kohonen map with 63 output neurons (i.e. a map of 9×7 neurons) was used and provided the best classification resolution of the studied samples (Fig. 3a) with low value of final topographic error (0.001). A hierarchical clustering analysis applied on the Kohonen map allowed the identification of five groups on the map. These groups were concordant with the geographic location of the samples, i.e. Europe-Tunisia, Australia, Algeria, Canada and China (Fig. 3b). The ANOSIM test computed for all pairwise comparisons revealed that these five groups differed significantly $p < 0.001$ (see Table S4 in the supporting

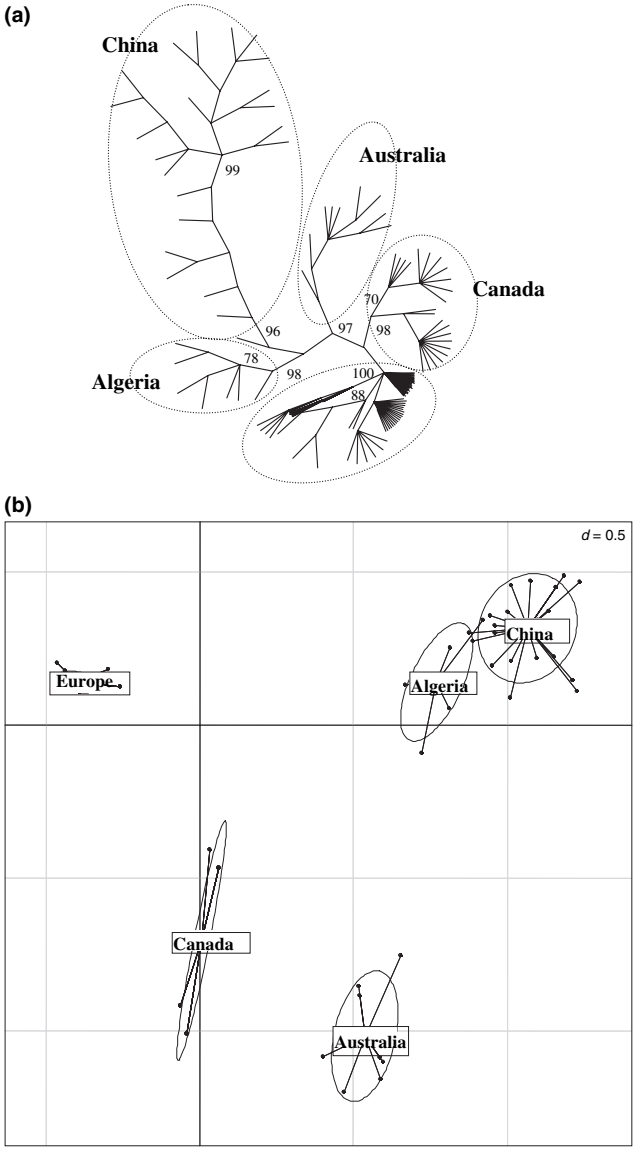

Fig. 2. (a) Consensus tree (see Materials and Methods section) based on 100 sub-equal parsimony trees with a goloboff correction Gk = 4. Numbers on branches are corresponding to bootstrap values when >65%). (b) Scatter plots of 159 *Ligula* individuals based on the first and the second components of factorial correspondence analysis using inter-simple sequence repeat data

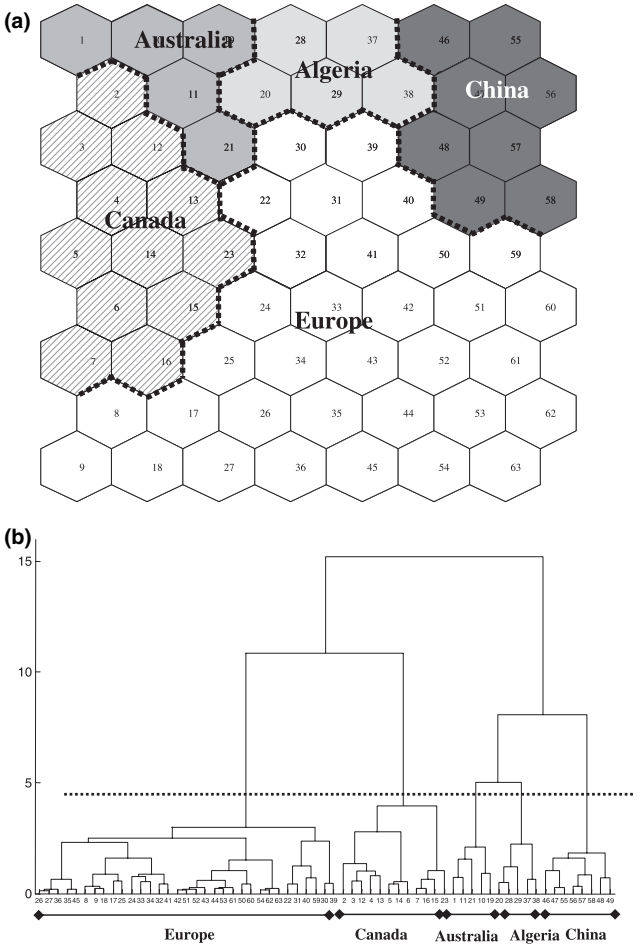

Fig. 3. Classification of the *Ligula intestinalis* populations based on the self-organizing map (SOM) analysis. (a) The patterned SOM map. The 159 samples analysed were ordinated into the map and were classified into five main clusters using a hierarchical cluster as shown in (b). Five clusters are marked off with bold lines. Individuals belonging to the same cluster are grouped in the same or neighbouring cells. (b) Hierarchical cluster analysis of the Kohonen map. Interrupted line shows the level of significance of the clustering

information). This confirms the genetically high differences between these groups.

According to the hierarchical cluster analysis, the samples from Europe and Tunisia were grouped into the same cluster. These samples could be further divided into two groups (Fig. 3b). However, the ANOSIM test revealed that these two groups did not differ significantly ($R = 0.02$, $p > 0.05$).

Discussion

The five geographical groups inferred from MP, FCA and SOM methods based on ISSR markers indicated a strong genetic discrimination between the analysed specimens. This finding was not surprising regarding the broad geographical range represented. The isolation of populations owing to glacial barriers has often been proposed to explain genetic differentiation between large geographical areas (Hewitt 2000). Additionally, such patterns could be maintained by isolation by distance.

Genetic variation among populations ($G_{ST} = 0.776$) and the level of polymorphism (100%) raise the evidence of the existence of distinct genetic isolates of *L. intestinalis* and support the presence of a complex of closely related species. Indeed, ISSR–PCR approach was shown to be suitable to study inter-specific variation in very close species (Luque et al. 2002). The present data demonstrated low intra-population genetic variability within the groups. This could be attributed to a recent origin of *Ligula* lineages and short differentiation times as explained previously by Li et al. (2000), even though ISSR consists mainly of non-coding DNA where fixation rates of mutation are higher than in coding regions (Kimura 1980). Low genetic variability could also be explained by the self-fertilization of the Cestoda, as described for numerous hermaphroditic parasites (Brown et al. 2001).

In addition, there were no significant genetic differences among the individuals from Europe. Hence, no genetic distinction was found between samples from the same geographic origin but from different hosts (e.g. samples from Czech Republic from bream and roach were grouped together) and no differentiation was found between samples belonging to the same host but from different European origins (e.g. samples from roach from France formed the same cluster with samples from Czech Republic). A possible explanation, which was also suggested for *Digramma* (a sister genus of *Ligula*) (Li and Liao 2003), for this low genetic variability is the migration of the definitive hosts of *Ligula* (e.g. *Larus ridibundus*, *Ardea cinerea* and *Phalacrocorax carbo*). Indeed, it was recently accepted that migrating birds may transport organisms across vast distances (McCoy et al. 2003; Gittenberger et al. 2006). For instance, Figuerola et al. (2005) showed that known waterfowl migratory movements are strongly linked to gene flow of aquatic invertebrates. In addition, promising studies suggested that vector-transmitted parasites, such as malarial parasites, might serve as useful geographical signals to track bird migration (Webster et al. 2002). In our perspective, bird migration could partly explain why European *L. intestinalis* specimens are indistinguishable from each other.

Although groups formed by classification methods seem to result from geographical isolation, the present results provide no evidence as to whether the resultant structure of the *Ligula* populations is host specific and/or geography dependent. Moreover, the analyses of ISSR data could not lead to a detailed picture of the evolutionary history of *L. intestinalis* populations. Possible strict co-speciation between parasites and their hosts and/or historical parasite shifts from one host lineage to another, were also recognized to explain differences in genetic patterns (Wickström 2003). This proposal could be tested by analyses of gene sequences and/or microsatellite which may overcome the ISSR limits and allow a genotypic analysis. More samples are also needed to precisely account for the relatively unequal distribution and restriction of *Ligula* populations to some host species in a given geographical area.

To our knowledge, this study is the first attempt to apply ISSR markers to cestodes. The study shows that ISSR can be applied to fresh/frozen or ethanol-fixed material. There was reproducibility in banding patterns; band intensity only decreased when the DNA amounts were limiting, but no new bands were displayed when DNA amounts were increased. This contrasts the situation with the RAPD technique, where there can be problems with the reproducibility of banding patterns, also relative to DNA amounts used in the PCR. From the present results, the ISSR approach employing the primers $WB(GACA)_4$, $(GACA)_4WB$, $(ACA)_5BDB$ and $BDB(ACA)_5$ emerged as a cost-effective tool to assess the population level of polymorphism and diversity among the analysed specimens. Other workers (Oliveira et al. 1997; Gomes et al. 1998; Fonseca-Salamanca et al. 2006) have recommended the $[CA]_nRY$ primer for studying genetic variability in parasites, because CA repeats are considered an abundant nucleotide combination in eukaryotic genomes (Tautz and Renz 1984) and are present in high copy number. These di-nucleotide motifs produce highly informative DNA fingerprinting when targeted with an anchored CA-repeat primer (Oliveira et al. 1997). None of the previous studies however have tested a primer with a tetra-nucleotide motif. The present results show that GACA 5' or 3' anchored primer gave an excellent resolution, whereas the di-nucleotide CA-repeat primer gave relatively smeary bands and irreproducible patterns. It is shown that an anchored $[GACA]_n$ primer should be considered in addition to $[CA]_nRY$ primer to study parasite genetic variability.

Patterns resulting from SOM-clustering methods were congruent with those obtained from analyses using MP and FCA. Analysis by the SOM method allowed the definition of five main geographically differentiated groups (Figs 2a,b and 3). The number of clusters used to subdivide the SOM map was defined *a priori* according to the geographical origins of the samples. This limitation of the SOM method was overcome using ANOSIM. This test showed that the differences between clusters were significant, and hence, demonstrates the statistical robustness of the SOM analysis. The major advantage of the SOM clustering compared with the usual genetic clustering method is the short running time of the program. Moreover, SOM is an unsupervised learning algorithm able to preserve the topology of the clusters, well suited to complex ecological data analysis. It is proposed that the SOM could be an alternative tool to standard genetic classification programs when complemented by ANOSIM tests, which are much more relevant than the classical Bootstrap analysis used classically in phylogenetic studies (Davidson and MacKinnon 1999).

In conclusion, it is proposed that the ISSR approach provides valuable markers to discriminate genetic distant *Ligula* specimens and SOM shows considerable promise as a rapid method to map the genetic structure of parasite populations, thus complementing conventional methods.

Acknowledgements

The authors would like to thank J. Stefka, T. Sholz, A. Arab, S. Thabet, N. Haddaoui, S. Frank and A. Lymbery for providing parts of *Ligula* samples and Z.J. Shao, C. Shinn and P. Winterton for English correction of the manuscript.

Résumé

Diversité génétique du parasite Ligula intestinalis *(Cestoda: Diphyllobothriidea) basée sur l'analyse des marqueurs des répétitions intermicrosatellites (ISSR)*

Dans cette étude, nous avons analysé la variabilité génétique de différentes populations de *Ligula intestinalis*. Pour cela, 9 marqueurs des répétitions intermicrosatellites (ISSR) ont été selectionnés et appliqués à des échantillons collectés dans 9 régions géographiques du monde et dans 10 espèces hôtes. Les patrons ISSR ont généré 110 loci et montrent une forte variabilité avec un polymorphisme de 100% et un coefficient global de différenciation de gène estimé par l'indice de Nei (G_{ST}) de 0.776. Nous trouvons une forte différentiation génétique entre 5 groupes qui correspondent à 5 larges régions géographiques (Europe, Chine, Canada, Australie et Algerie). Néanmoins aucune variation génétique significative n'a été retrouvée parmi les échantillons Européens, bien qu'ils soient originaires de différentes localités géographiques et/ou de différents hôtes. Les méthodes de classification classique : Maximum de parcimonie (MP) et l'Analyse Factorielle de Correspondance (AFC) ont été comparées avec une méthode statistique récente : les cartes auto-organisatrices (SOM). Les résultats montrent que l'approche ISSR que nous avons utilisée fournit des marqueurs fiables capables de discriminer entre différentes populations de *Ligula*. Les SOM représentent un outil alternatif efficace pour établir la structure génétique des populations de parasites.

References

Arme C (1997) Ligulosis in two cyprinid hosts: *Rutilus rutilus* and *Gobio gobio*. Helminthologia **34:**191–196.

Barus V, Prokes M (2002) Length and weight of *Ligula intestinalis* plerocercoids (Cestoda) parasitizing adult cyprinid fishes (Cyprinidae): a comparative analysis. Helminthologia **39:**29–34.

Bean CC, Winfield IJ (1992) Influences of the tapeworm *Ligula intestinalis* (L.) on the spatial distributions of juvenile roach *Rutilus rutilus* (L.) and gudgeon *Gobio gobio* (L.) in Lough Neagh, Northern Ireland. Neth J Zool **42:**416–429.

Behura SK (2006) Molecular marker systems in insects: current trends and future avenues. Mol Ecol **15:**3087–3113.

Brown SP, Renaud F, Guegan JF, Thomas F (2001) Evolution of trophic transmission in parasites: the need to reach a mating place? J Evol Biol **14:**815–820.

Chapman A, Hobbs RP, Morgan DL, Gill HS (2006) Helminth parasitism of *Galaxias maculatus* (Jenyns 1842) in southwestern Australia. Ecol Freshw Fish **15:**559–564.

Chon TS, Park YS, Moon KH, Cha EY (1996) Patternizing communities by using an artificial neural network. Ecol Model **90:**69–78.

Clarke KR (1993) Non-parametric multivariate analyses of changes in community structure. Aust J Ecol **18:**117–143.

Davidson R, MacKinnon JG (1999) The size distortion of bootstrap tests. Econ Theory **15:**361–376.

De Meeus T, Michalakis Y, Renaud F (1998) Santa Rosalia revisited: or why are there so many kinds of parasites in 'The Garden of Earthly Delights'? Parasitol Today **14:**10–13.

Dubinina MN (1980) Tapeworms (Cestoda, Ligulidae) of the Fauna of the USSR. Amerind Publishing Co. Pvt. Ltd., New Delhi, p. 320.

Esselman EJ, Jianqiang L, Crawford DJ, Windus JL, Wolfe AD (1999) Clonal diversity in the rare *Calamagrostis porteri ssp*. insperata (Poaceae): comparative results for allozymes and random amplified polymorphic DNA (RAPD) and intersimple sequence repeat (ISSR) markers. Mol Ecol **8:**443–451.

Fang DQ, Roose ML (1997) Identification of closely related *Citrus cultivars* with inter-simple sequence repeat markers. Theor Appl Genet **95:**408–417.

Figuerola J, Green AJ, Michot TC (2005) Invertebrate eggs can fly: evidence of waterfowl-mediated gene flow in aquatic invertebrates. Am Nat **165:**274–280.

Fonseca-Salamanca F, Nogal-Ruiz JJ, Benito C, Camacho MV, Martínez-Fernández AR (2006) Molecular characterization of *Trichinella* genotypes by inter-simple sequence repeat polymerase chain reaction (ISSR-PCR). J Parasitol **92:**606–610.

Giraudel JL, Aurelle D, Berrebi P, Lek S (2000) Application of the self-organizing mapping and fuzzy clustering to microsatellite data: how to detect genetic structure in brown trout (*Salmo trutta*) populations. In: Lek S, Guégan JF (eds), Artificial Neuronal Networks. Springer-Verlag, Berlin, pp 187–202.

Gittenberger E, Groenenberg DSJ, Kokshoorn B, Preece RC (2006) Biogeography: molecular trails from hitch-hiking snails. Nature **409:**439–409.

Gobert V, Moja S, Taberlet P, Wink M (2006) Heterogeneity of three molecular data partition phylogenies of mints related to *M*. x *piperita* (*Mentha*; Lamiaceae). Plant Biol **8:**470–485.

Goloboff PA (1993) Estimating character weights during tree search. Cladistics **9:**83–91.

Gomes ML, Macedo AM, Pena SD, Chiari E (1998) Genetic relationships between *Trypanosoma cruzi* strains isolated from chronic chagasic patients in southern Brazil as revealed by RAPD and SSR-PCR analysis. Acta Trop **69:**99–109.

Graham J, McNicol RJ, Greig K, Van De Ven WTG (1994) Identification of red raspberry cultivars and an assessment of their relatedness using fingerprints produced by random primers. J Hortic Sci **69:**123–130.

Groves KL, Shields BA (2001) Observations on the Plerocercoid stage of the tapeworm *Ligula* in three species of fish from the Lower Crooked River of Central Oregon. J Aquat Anim Health **13:**285–289.

Hewitt G (2000) Post-glacial re-colonization of European biota. Biol J Linn Soc **68:**87–112.

Hoberg EP (1997) Phylogeny and historical reconstruction: host–parasite systems as keystones in biogeography and ecology. In: Reaka-Kudla ML, Wilson DE, Wilson EO (eds), Biodiversity II: Understanding and Protecting Our Biology Resources. Joseph Henry Press, Washington, pp 243–261.

Huang JC, Sun M (2000) Genetic diversity and relationships of sweetpotato and its wild relatives in Ipomoea series Batatas (Convolvulaceae) as revealed by inter-simple sequence repeat (ISSR) and restriction analysis of chloroplast DNA. Theor Appl Genet **100:**1050–1060.

Hundsdoerfer AK, Wink M (2006) Incongruence of morphology and genetic markers in *Hyles tithymali* (Lepidoptera: Sphingidae) from the Canary Islands. J Zool Syst Evol Res **44:**316–322.

Ihaka R, Gentleman R (1996) R: a language for data analysis and graphics. J Comput Graph Stat **5:**299–314.

Joshi SP, Gupta VS, Aggarwal RK, Ranjekar PK, Brar DS (2000) Genetic diversity and phylogenetic relationship as revealed by inter simple sequence repeat (ISSR) polymorphism in the genus *Oryza*. Theor Appl Genet **100:**1311–1320.

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol **16:**111–120.

Kohonen T (1982) Self-organized formation of topologically correct feature maps. Biol Cybern **43:**59–69.

Kohonen T (2001) Self-Organizing Maps, 3rd edn. Springer-Verlag, Heidelberg.

Koopman WJM, Zevenbergen MJ, Van den Berg RG (2001) Species relationships in *Lactuca* S.L. (Lactuceae, Asteraceae) inferred from AFLP fingerprints1. Am J Bot **88:**1881–1887.

Lanham PG, Korycinska A, Brennan RM (2000) Genetic diversity within a secondary gene pool for *Ribes nigrum* L. revealed by RAPD and ISSR markers. J Hortic Sci Biotech **75:**371–375.

Lek S (2000) Artificial Neuronal Networks: Application to Ecology and Evolution. Springer, Berlin.

Lewontin RC (1972) The apportionment of human diversity. In: Dobzhansky T, Hecht MK, Steere WC (eds), Evolutionary Biology 6. Appleton-Century-Crofts, New York, pp 381–398.

Li J, Liao X (2003) The taxonomic status of *Digramma* (Pseudophyllidae: Ligulidae) inferred from DNA sequences. J Parasitol **89:**792–799.

Li J, Liao X, Yang H (2000) Molecular characterization of a parasitic tapeworm (*Ligula*) based on DNA sequences from formalin-fixed specimens. Biochem Genet **38:**309–322.

Logan FJ, Horák A, Štefka J, Aydogdu A, Scholz T (2004) The phylogeny of diphyllobothriid tapeworms (Cestoda: Pseudophyllidea) based on ITS-2 rDNA sequences. Parasitol Res **94:**10–15.

Loot G, Poulin R, Lek S, Guegan JF (2002a) The differential effects of *Ligula intestinalis* (L.) plerocercoids on host growth in three natural populations of roach (*Rutilus rutilus*). Ecol Freshw Fish **11:**168–177.

Loot G, Aulagnier S, Lek S, Thomas F, Guegan JF (2002b) Experimental demonstration of a behavioural modification in a cyprinid fish, *Rutilus rutilus* (L.), induced by a parasite, *Ligula intestinalis* (L.). Can J Zool **80:**738–744.

Losse GF, Nau W, Winter M (1991) Le développement de la pêche en eau douce dans le nord de la Tunisie Projet de la coopération technique. Utilisation des barrages pour la pisciculture. CGP/GTZ, Tunis, Tunisie, p. 418.

Luque C, Legal L, Staudter H, Gers C, Wink M (2002) Brief report ISSR (inter simple sequence repeats) as genetic markers in Noctuids (Lepidoptera). Hereditas **136:**251–253.

Mace ES, Gebhardt CG, Lester RN (1999) AFLP analysis of genetic relationships in the tribe Datureae (Solanaceae). Theor Appl Genet **99:**634–641.

Maltagliati F, Lai T, Casu M, Valdesalici S, Castelli A (2006) Identification of endangered Mediterranean cyprinodontiform fish by means of DNA inter-simple sequence repeats (ISSRs). Biochem Syst Ecol **34:**626–634.

MathWorks (2001) Laboratory of Information and Computer Science, Helsinki University of Technology (Version 6.1.0.450 Release 12.1). The MathWorks, Inc. Available from http://www.cis.hut.fi/projects/somtoolbox.

McCoy KD, Boulinier T, Tirard C, Michalakis Y (2003) Host-dependent genetic structure of parasite populations: differential dispersal of seabird tick host races. Evolution **57:**288–296.
Museth J (2001) Effects of *Ligula intestinalis* on habitat use, predation risk and catchability in European minnows. J Fish Biol **59:**1070–1080.
Oliveira RP, Macedo AM, Chiari E, Pena SD (1997) An alternative approach to evaluating the intraspecific genetic variability of parasites. Parasitol Today **13:**196–200.
Olson PD, Tkach VV (2005) Advances and trends in the molecular systematics of the parasitic platyhelminthes. Adv Parasitol **60:**165–243.
Olson PD, Littlewood DTJ, Griffiths D, Kennedy CR, Arme C (2002) Evidence for the co-existence of separate strains or species of *Ligula* in Lough Neagh, Northern Ireland. J Helminthol **76:**171–174.
Orr TSC (1966) Spawning behaviour of Rudd, *Scardinius erythrophthalmus* infested with Plerocercoids of *Ligula intestinalis*. Nature **12:**736.
Page RDM (1996) Treeview: an application to display phylogenetic trees on personal computers. Comp Appl Biosci **12:**357–358.
Park YS, Cereghino R, Compin A, Lek S (2003) Applications of artificial neural networks for patterning and predicting aquatic insect species richness in running waters. Ecol Model **160:**265–280.
Paterson AM, Banks J (2001) Analytical approaches to measuring cospeciation of host and parasites: through a glass, darkly. Int J Parasitol **31:**1012–1022.
Poulin R, Morand S (2000) The diversity of parasites. Q Rev Biol **75:**277–293.
Price PW (1980) Evolutionary Biology of Parasites. Princeton University Press, Princeton, New Jersey, USA.
Robinson WA, Liston A, Doescher PS, Svejcar T (1997) Using ISSR markers to quantify clonal vs. sexual reproduction in *Festuca idahoensis* (Poaceae). Am J Bot **84:**89.
Rosen F (1918) Recherches sur le développement des cestodes. I. Le cycle évolutif des Bothriocéphales. Bull de la Soc Neuchâtel de Sci Nat **43:**241–300.
Roux O, Gevrey M, Arvanitakis L, Gers C, Bordat D, Legal L (2007) ISSR-PCR: tool for discrimination and genetic structure analysis of *Plutella xylostella* populations native to different geographical areas. Mol Phylogenet Evol **43:**240–250.
Ruanet VV, Kochieva EZ, Ryzhova NN (2005) Kohonen network study of the results of RAPD and ISSR analyses of genomic polymorphism in the genus *Capsicum* L. Rus J Genet **41:**202–210.
Schneider S, Roessli D, Excoffier L (2000) A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.
Simmons MP, Zhang LB, Webb CT, Muller K (2007) A penalty of using anonymous dominant markers (AFLPs, ISSRs, and RAPDs) for phylogenetic inference. Mol Phylogenet Evol **42:**528–542.
Stefka J, Gilleard JS, Grillo V, Hypsa V (2007) Isolation and characterization of microsatellite loci in the tapeworm *Ligula intestinalis* (Cestoda: Pseudophyllidea). Mol Ecol Notes. doi: 10.1111/j.1471-8286.2007.01704-X.
Swofford DL (2001) PAUP* Phylogenetic analysis using parsimony (*and other methods). Sinauer, Sunderland, MA (Version 4.0b10).
Tautz D, Renz M (1984) Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res **12:**4127–4138.
Treutlein J, Smith GF, van Wyk BE, Wink M (2003a) Evidence for the polyphyly of Haworthia (Asphodelaceae Subfamily Alooideae; Asparagales) inferred from nucleotide sequences of rbcL, matK, ITS1 and genomic fingerprinting with ISSR-PCR. Plant Biol **5:**513–521.
Treutlein J, Smith GF, van Wyk BE, Wink M (2003b) Phylogenetic relationships in Asphodelaceae (subfamily Alooideae) inferred from chloroplast DNA sequences (rbcL, matK) and from genomic fingerprinting (ISSR). Taxon **52:**193–207.
Treutlein J, Vorster P, Wink M (2005) Molecular relationships in Encephalartos (Zamiaceae, Cycadales) based on nucleotide sequences of nuclear ITS 1&2, rbcL, and Genomic ISSR Fingerprinting. Plant biol **7:**79–90.
Webster MS, Marra PP, Haig SM, Bensch S, Holmes RT (2002) Links between worlds: unravelling migratory connectivity. Trends Ecol Evol **17:**76–83.
Wickström LM (2003) Phylogeny, phyletic coevolution and phylogeography of anoplocephaline cestodes in mammals. Finnish Forest Research Institute, Vantaa Research Centre, Finland.
Wickström LM, Haukisalmi V, Varis S, Hantula J, Fedorov VB, Henttonen H (2003) Phylogeography of the circumpolar *Paranoplocephala arctica* species complex (Cestoda: Anoplocephalidae) parasitizing collared lemmings (*Dicrostonyx* spp.). Mol Ecol **12:**3359–3371.
Wolfe AD, Liston A (1998) Contributions of PCR-based methods to plant systematics and evolutionary biology. In: Soltis DE, Soltis PS, Doyle JJ (eds), Molecular Systematics of Plants II, DNA Sequencing. Kluwer Academic Publishers, New York, USA, pp 43–86.
Wolfe AD, Xiang QY, Kephart SR (1998a) Assessing hybridization in natural populations of Penstemon (Scrophulariaceae) using hypervariable intersimple sequence repeat (ISSR) bands. Mol Ecol **7:**1107–1125.
Wolfe AD, Xiang QY, Kephart SR (1998b) Diploid hybrid speciation in Penstemon (Scrophulariaceae). Proc Natl Acad Sci USA **95:**5112–5115.
Wu W, Zheng YL, Chen L, Wei YM, Yang RW, Yan ZH (2005) Evaluation of genetic relationships in the genus *Houttuynia* Thunb. in China based on RAPD and ISSR markers. Biochem Syst Ecol **33:**1141–1157.
Yeh FC, Yang RC, Boyle TBJ, Ye ZH, Mao JX (1997) POPGENE, The User-Friendly Shareware for Population Genetic Analysis. Molecular Biology and Biotechnology center, University of Alberta, Canada (1.32).
Zhao N, Ai W, Shao Z, Zhu B, Brosse S, Chang J (2005) Microsatellites assessment of Chinese sturgeon (*Acipenser sinensis* Gray) genetic variability. J Appl Ichthyol **21:**7–13.
Zietkiewicz E, Rafalski A, Labuda D (1994) Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. Genomics **20:**176–183.

Authors' addresses: Wafa Bouzid, Laboratoire Evolution et Diversité Biologique, UMR CNRS-UPS 5174, Université Paul Sabatier, 118 route de Narbonne, F-31062 Toulouse cedex 4, France. E-mail: bouzid@cict.fr; Sovan Lek, Laboratoire Evolution et Diversité Biologique, UMR CNRS-UPS 5174, Université Paul Sabatier, 118 route de Narbonne, F-31062 Toulouse cedex 4, France. E-mail: lek@cict.fr; Matthias Mace, Laboratoire Evolution et Diversité Biologique, UMR CNRS-UPS 5174, Université Paul Sabatier, 118 route de Narbonne, F-31062 Toulouse cedex 4, France. E-mail: Matthias.Mace@wanadoo.fr; Oum Kalthoum Ben Hassine, Unité de Recherche, Biologie, Ecologie et Parasitologie des Organismes Aquatiques, Faculté des Sciences, 1060 Tunis, Tunisie. E-mail: kalthoum.benhassine@gmail.com; Roselyne Etienne, Laboratoire Evolution et Diversité Biologique, UMR CNRS-UPS 5174, Université Paul Sabatier, 118 route de Narbonne, F-31062 Toulouse cedex 4, France. E-mail: etienne@cict.fr; Luc Legal, ECOLAB, UMR CNRS-UPS 5245, Université Paul Sabatier, 118 Route de Narbonne, F-31062 Toulouse cedex 4, France. E-mail: legal@cict.fr; Géraldine Loot, Laboratoire Evolution et Diversité Biologique, UMR CNRS-UPS 5174, Université Paul Sabatier, 118 route de Narbonne, F-31062 Toulouse cedex 4, France. E-mail: loot@cict.fr

Supporting information

Additional supporting information may be found in the online version of this article:

Table S1. Oligonucleotide primers used in the polymerase chain reaction-based inter-simple sequence repeat and their annealing temperatures

Table S2. Details of genetic variability for the 10 *Ligula intestinalis* populations studied

Table S3. Nei's measures of genetic identity (*i*) (above diagonal) and genetic distance (*D*) (below diagonal) for all pairwise comparisons between populations

Table S4. Results of the ANOSIM test[1] for all pairwise comparisons between five groups

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.