Research Article

Development of EST-derived microsatellite markers to investigate the population structure of sparganum — the causative agent of zoonotic sparganosis

Xi Zhang, Xiu Hong, Jiang Yang Duan, Lu Lu Han, Zi Yang Hong, Peng Jiang, Zhong Quan Wang and Jing Cui

Department of Parasitology, School of Basic Medical Sciences, Zhengzhou University, Zhengzhou 450052, People's Republic of China

Abstract

The plerocercoid (sparganum) of *Spirometra erinaceieuropaei* is the main aetiological agent of human sparganosis. To improve the current knowledge on *S. erinaceieuropaei* evolution, we performed multi-locus microsatellite typing of sparganum isolates from China for the first time. All available expressed sequence tag (EST) sequences for the *Spirometra* were downloaded from the GenBank. The identification and localization of microsatellites in ESTs was accomplished by *MISA*. Based on the selected microsatellites, the genetic structure of 64 sparganum isolates collected from 11 geographical locations in southwest China were investigated through principal component analysis, STRUCTURE analysis and neighbour-joining clustering. A total of 522 non-redundant ESTs containing 915 simple sequence repeats were identified from 12 481 ESTs screened. Five primer pairs were finally selected. Using these loci, a total of 12 alleles were detected in 64 sparganum isolates. Little variability was observed within each of geographical population, especially among isolates derived from Kunming of Yunnan (YN-KM) province. Both STRUCTURE analysis and the clustering analysis supported that two genotypes existed among the sparganum isolates from southwest China. In conclusion, five microsatellite markers were successfully developed, and sparganum population was observed to harbour low genetic variation, further investigation with deeper sampling was needed to elucidate the population structure.

Introduction

The plerocercoid larvae (sparganum) of cestodes in the genus *Spirometra* are causative agent of human sparganosis that can manifest in local tissue damage, paralysis, blindness and even death (Anantaphruti *et al.*, 2011; Liu *et al.*, 2015). Spargana of all *Spirometra* species infect a broad spectrum of amphibians and reptiles as their second intermediate host (Kuchta *et al.*, 2015). The major aetiological agent of human sparganosis is *Spirometra erinaceieuropaei* which is mainly acquired by the consumption of undercooked or uncooked meat of frogs and snakes (Anantaphruti *et al.*, 2011). Although *S. erinaceieuropaei* was distributed globally, most sparganosis cases occur in eastern and southeastern Asian countries (Almeida *et al.*, 2016; Tang *et al.*, 2017). China has the largest number of sparganosis cases in the world with more than 1000 cases have been reported, scattered over almost all administrative areas, so sparganosis still poses a serious threat to human health in China (Liu *et al.*, 2015). Therefore, knowledge regarding the genetic characteristics and divergence pattern of the parasite is valuable for the control of sparganosis.

Several studies have used mitochondrial molecular markers (such as *cox*1 and *cyt*b) to study genetic diversity of sparganum isolates from several geographical locations (Okamoto *et al.*, 2007; Jeon *et al.*, 2016; Zhang *et al.*, 2016, 2017). In comparison with mitochondrial genes, the simple sequence repeats (SSRs), or microsatellites have become the molecular marker of choice for genetic analysis owing to their high level of polymorphism (Bouck and Vision, 2007). In addition, the development of microsatellites from expressed sequence tag (EST) databases has been shown to be a feasible option for obtaining high-quality genetic markers (La Rosa *et al.*, 2012). Using EST-SSRs, some remarkable advances have been made in population genetic analysis of human parasites in recent years (La Rosa *et al.*, 2012; Gelanew *et al.*, 2014; Garbarino *et al.*, 2017). For the diphyllobothriid tapeworms, Bazsalovicsová *et al.* (2018) recently developed suitable microsatellite loci for *Dibothriocephalus latus* (the broad fish tapeworm). To date, however, none microsatellite markers have ever been developed for the spirometrid tapeworms. Therefore, in this study, we attempted to identify and characterize microsatellite loci sensitive enough to genotype *S. erinaceieuropaei*, and determine their usefulness as molecular markers for exploring population structure in *S. erinaceieuropaei* from southwestern China.

CAMBRIDGE
UNIVERSITY PRESS
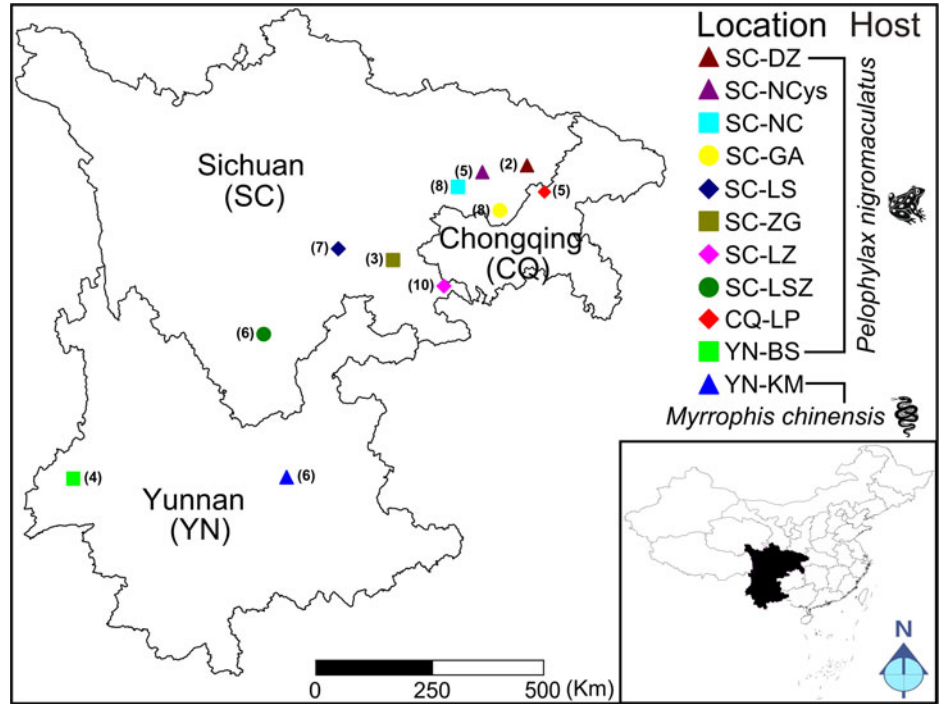
Fig. 1. Map of collection localities and host information for sparganum isolates. The number in the parenthesis indicates the sample size in each geographical location. Geographic regions in China are designated as follows: Dazhou, Sichuan (SC-DZ), Yingshan, Nanchong, Sichuan (SC-NCys), Nanchong, Sichuan (SC-NC), Guangan, Sichuan (SC-GA), Leshan, Sichuan (SC-LS), Zigong, Sichuan (SC-ZG), Luzhou, Sichuan (SC-LZ), Liangshanzhou, Sichuan (SC-LSZ), Liangping, Chongqing (CQ-LP), Baoshan, Yunnan (YN-BS) and Kunming, Yunnan (YN-KM).

Table 1. Frequency and distribution of different types of SSRs identified in the *Spirometra* ESTs (after considering sequence complementarities of the repeat motifs)

| Repeat motif | Number of repeat units | | | | | | | | | | | | | Total repeats |
	3	4	5	6	7	8	9	10	11	12	13	14	≥15	
A/T	–	–	–	–	–	–	–	28	21	20	24	13	114	220
C/G	–	–	–	–	–	–	–	64	39	22	16	12	54	207
AC/GT	–	85	19	6	2	1	–	–	–	1	–	1	–	115
AG/CT	–	85	15	3	5	4	–	1	–	1	1	–	17	132
AT/TA	–	19	2	–	–	1	–	–	1	–	–	–	1	24
CG/GC	–	14	1	–	–	–	–	–	–	–	–	–	–	15
AAC/GTT	–	8	–	–	–	–	–	–	–	–	–	–	–	9
AAG/CTT	–	19	7	–	–	–	–	–	–	–	–	1	–	27
AAT/ATT	–	3	1	–	–	–	1	–	–	–	–	–	–	5
ACC/GGT	–	9	2	1	–	–	–	–	–	–	–	–	–	12
ACG/CGT	–	4	3	1	–	–	–	–	–	–	–	–	–	8
ACT/AGT	–	–	–	2	–	1	1	–	–	–	–	–	–	4
AGC/CTG	–	12	3	4	1	1	–	–	–	–	–	–	–	21
AGG/CCT	–	22	6	3	4	–	1	–	–	–	–	–	–	36
ATC/ATG	–	9	1	1	–	–	–	–	–	–	–	–	–	11
CCG/CGG	–	10	5	2	2	–	1	–	–	–	–	–	–	20
AAAC/GTTT	1	–	–	–	–	–	–	–	–	–	–	–	–	1
AAAG/CTTT	3	–	–	–	–	–	–	–	–	–	–	–	–	3
AAAT/ATTT	3	–	–	–	–	–	–	–	–	–	–	–	–	3
AACC/GGTT	2	–	–	–	–	–	–	–	–	–	–	–	–	2
AACG/CGTT	2	–	–	–	–	–	–	–	–	–	–	–	–	2
AAGC/CTTG	1	–	–	–	–	–	–	–	–	–	–	–	–	1
AAGG/CCTT	1	–	–	–	–	–	–	–	–	–	–	–	–	1
ACAG/CTGT	3	–	–	–	–	–	–	–	–	–	–	–	–	3
ACCT/AGGT	–	1	–	–	–	–	–	–	–	–	–	–	–	1

(*Continued*)

Table 1. (*Continued.*)

Repeat motif	Number of repeat units													Total repeats
	3	4	5	6	7	8	9	10	11	12	13	14	⩾15	
ACGC/CGTG	4	–	1	–	–	–	–	–	–	–	–	–	–	5
ACGG/CCGT	1	–	–	–	–	–	–	–	–	–	–	–	–	1
ACTG/AGTC	1	–	–	–	–	–	–	–	–	–	–	–	–	1
AGAT/ATCT	1	2	–	–	–	–	–	–	–	–	–	–	–	3
AGCC/CTGG	1	–	–	–	–	–	–	–	–	–	–	–	–	1
AGGC/CCTG	3	–	–	–	–	–	–	–	–	–	–	–	–	3
AGGG/CCCT	6	–	–	–	–	–	–	–	–	–	–	–	–	6
ATCC/ATGG	2	–	–	–	–	–	–	–	–	–	–	–	–	2
ATGC/ATGC	2	–	–	–	–	–	–	–	–	–	–	–	–	2
AAAAT/ATTTT	1	–	–	–	–	–	–	–	–	–	–	–	–	1
AAATG/ATTTC	1	–	–	–	–	–	–	–	–	–	–	–	–	1
ACACG/CGTGT	1	–	–	–	–	–	–	–	–	–	–	–	–	1
AGCCG/CGGCT	1	–	–	–	–	–	–	–	–	–	–	–	–	1
AGGCG/CCTCG	–	1	–	–	–	–	–	–	–	–	–	–	–	1
ACCGCC/CGGTGG	1	–	–	–	–	–	–	–	–	–	–	–	–	1
AGGCGG/CCGCCT	2	–	–	–	–	–	–	–	–	–	–	–	–	2
N (MNR)	–	–	–	–	–	–	–	92	60	42	40	25	168	427
NN (DNR)	–	203	37	9	7	6	–	1	1	2	1	1	18	286
NNN (TNR)	–	96	28	14	7	2	4	–	–	–	–	1	1	153
NNNN (TTNR)	37	3	1	–	–	–	–	–	–	–	–	–	–	41
NNNNN (PNR)	4	1	–	–	–	–	–	–	–	–	–	–	–	5
NNNNNN (HNR)	3	–	–	–	–	–	–	–	–	–	–	–	–	3

Table 2. Primer sets used to amplify microsatellites identified in sparganum isolates

Locus	GenBank accession no.[a]	Primer set	T_m	Repeat motif
C07	HS516371	CCGGATTCGCCAGGTGAG	60.2	$(ATG)_5$
		GAGGCCGATCAGTCGCTC	60.0	
I03	HS517370	CCCGGATCCAAACTACCCC	59.5	$(GAT)_4$
		AACAGTGCTGGGATCGCC	60.0	
M15	HS515019	AAAACTTCCCTTGCGCGC	59.7	$(AGTC)_3$
		GCCACCAGTACCACCCAC	60.0	
M20	HS518254	TTTCGCCTCCTGCACCAG	60.0	$(CAG)_4$
		ACTGGCGAGCATGCAGAG	60.1	
N11	HS514256	CGCTAAGCCGGTTGTTGC	59.8	$(CGA)_5$
		CGCCAGGTCTCCGTGAAG	60.1	

T_m, melting temperature. The subscript '*n*' means a variable number of repeats.
[a]Accession number of the EST sequences from which the primers were designed.

Materials and methods

Sparganum isolates

Total 64 sparganum isolates originating from 11 geographic regions of southwestern China were collected. The detailed information about isolates and hosts is shown in Fig. 1 and Supplementary Table S1. The spargana of *S. erinaceieuropaei* were collected from infected frogs (*Pelophylax nigromaculatus*) and snakes (*Myrrophis chinensis*) according to the methods described in Wei *et al.* (2015). Each tested sparganum was isolated from a single frog or snake, and then subjected to the analysis.

Marker development

For each non-redundant SSR–EST, primer pairs that flanked the microsatellite and that spanned no greater than 300 bp were

Table 3. General features of the genetic variability of sparganum isolates tested

| Populations (N)	Parameter	Microsatellite locus				
		C07	I03	M15	M20	N11
CQ-LP (5)	A	1	1	2	2	2
	H_o	0	0	0.6000	0.6000	0.6000
	H_e	0	0	0.4667	0.4667	0.4667
YN-BS (4)	A	1	1	1	1	2
	H_o	0	0	0	0	0.2500
	H_e	0	0	0	0	0.2500
YN-KM (6)	A	1	1	1	1	1
	H_o	0	0	0	0	0
	H_e	0	0	0	0	0
SC-GA (8)	A	1	1	2	2	2
	H_o	0	0	0.3750	0.3750	0.3750
	H_e	0	0	0.3250	0.3250	0.3250
SC-LZ (10)	A	2	2	2	3	3
	H_o	0.1000	0.1000	0.9000	0.9000	0.9000
	H_e	0.2684	0.2684	0.5211	0.5737	0.5737
SC-NC (8)	A	1	1	2	2	2
	H_o	0	0	0.2500	0.2500	0.2500
	H_e	0	0	0.2333	0.2333	0.2333
SC-DZ (2)	A	1	1	2	2	2
	H_o	0	0	0.50000	0.50000	0.50000
	H_e	0	0	0.50000	0.50000	0.50000
SC-LSZ (6)	A	2	2	2	2	2
	H_o	0.1667	0.1667	0.8333	0.8333	0.8333
	H_e	0.1667	0.1667	0.5303	0.5303	0.5303
SC-LS (7)	A	2	1	2	2	2
	H_o	0.1429	0	0.8571	0.7143	0.7143
	H_e	0.3626	0	0.5275	0.4945	0.4945
SC-ZG (3)	A	1	1	2	2	2
	H_o	0	0	0.6667	0.6667	0.6667
	H_e	0	0	0.5333	0.5333	0.5333
SC-NCys (5)	A	1	1	2	2	2
	H_o	0	0	0.8000	0.8000	0.8000
	H_e	0	0	0.5333	0.5333	0.5333

N, number of individuals analysed; A, number of alleles; H_o, observed heterozygosity; H_e, expected heterozygosity.

designed by using the Primer3 program (https://sourceforge.net/projects/primer3/). The screening of specific primers for *S. erinaceieuropaei* was performed in NCBI_Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). Total genomic DNA was extracted from plerocercoid larvae using the EasyPure Genomic DNA Kit (Transgen, China) following the manufacturer's protocol. Polymerase chain reaction (PCR) amplifications of DNA were performed in 25 μL using a premixed 2× EasyTaq PCR Super Mix (Transgen, China). The amplification was carried out as follows: an initial denaturation at 94 °C for 5 min, then 94 °C for 30 s (denaturation); 50–60 °C for 30–60 s (annealing); 72 °C for 30–60 s (extension) for 32 cycles, followed by a final extension at 72 °C for 5 min. The PCR products used for genotyping were separated by electrophoresis on 8% non-denaturing polyacrylamide gels with a voltage of 100 V lasting 30 min and visualized *via* silver-staining. The samples which displayed more than a single band were considered to contain putative polymorphic sites and were further characterized.

Population genetic analysis

The population genetic parameters including allelic richness (A), observed heterozygosity (H_o) and expected heterozygosity (H_e) were calculated in GenePop v4.2 (Rousset, 2008). The pairwise F_{st} values between populations were calculated in Arlequin v3.5 (Excoffier and Lischer, 2010) to estimate the levels of genetic

Principal Coordinates (PCoA)
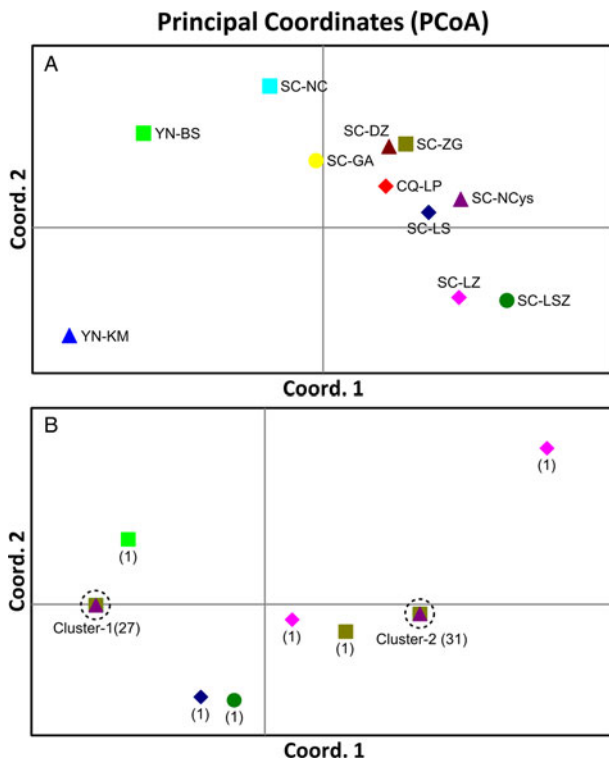
Fig. 2. (A) PCA describing the relationships of the 11 sparganum populations studied based on the F_{st} values calculated using five microsatellite markers. (B) PCA describing the relationships of 64 individuals (based on the covariance values calculated using five microsatellite markers). The number in the parenthesis indicates the individual numbers included in each cluster. For the isolate code, see Fig. 1.

differentiation among the populations and among the loci. In order to summarize the genetic relationships among the isolates and among the individuals, the principal coordinate analysis (PCA) was performed in GenAlEx v6.5 software (Smouse et al., 2015). The structure of the sparganum populations was determined in STRUCTURE v2.3.4 (Hubisz et al., 2009) through K-means partitional clustering and the admixture model (Falush et al., 2003). Twenty independent runs were conducted for each of $K = 1–11$ using a length of burn-in of 10^5 and 10^6 replicates of Bayesian Monte-Carlo Markov Chain sampling. The average membership coefficients for the 20 simulation runs of a given K value were generated using CLUMPP v1.1.2 (Jakobsson and Rosenberg, 2007) and a graphical presentation of the average membership coefficients for each isolate was generated in Microsoft Excel. The most appropriate number of K was calculated using STRUCTURE HARVESTER, web version (Earl and vonHoldt, 2012). Allelic data of EST–SSR were also used to confirm the genetic relationships among the tested genotypes by cluster analysis. The allelic data were transformed to binary mode using scores 1/0 for presence/absence of allele respectively, as described in Aggarwal et al. (2007). The clustering based on the neighbour-joining (NJ) method was performed in PAUP*4b10 (Swofford, 2003). In addition, in order to compare the phylogenetic patterns generated by microsatellites and mitochondrial markers, the clustering analysis of these sparganum samples was also performed using the mitochondrial cytochrome c oxidase subunit 1 (cox1) sequences. The cox1 sequences of isolates from CQ-LP, SC-NC, SC-NCys, SC-GA, SC-LZ, YN-BS and YN-KM were derived using the method described in Zhang et al. (2016). A total of 18 newly added sequences representing samples from SC-DZ, SC-LS, SC-ZG and SC-LSZ were sequenced in this study using the method described in Zhang et al. (2016). The

phylogenetic analysis (NJ method) was also performed in PAUP*4b10.

## Results

Development of functional EST–SSR markers

A total of 12 481 ESTs of the plerocercoid were obtained from the GenBank database. After a redundancy analysis to reduce over-estimation, the cluster analysis revealed 267 SSR–ESTs as singletons and 2059 SSR–ESTs into 255 clusters. As a result, 522 non-redundant ESTs were identified that contained a total of 915 SSRs (Table 1). Analysis of SSR motifs in the non-redundant SSR–ESTs revealed 144 (27.59%) ESTs that contained more than one SSR. Of the total 915 SSRs, 427 (46.67%) were mononucleotide repeats (MNRs), 286 (31.26%) were di-nucleotide repeats (DNRs), 153 (16.72%) were tri-nucleotide repeats (TNRs), 41 (4.48%) were tetra-nucleotide repeats (TTNRs), 5 (0.55%) were penta-nucleotide repeats (PNRs) and 3 (0.33%) were hexa-nucleotide repeats (HNRs). Among the 286 DNRs, AG motif was the most common (46.15%) followed by the AC (40.21%) motif, whereas AT and CG motifs were uncommon (8.39 and 5.25%). Similarly, among the TNRs, the motif AGG was the most common (23.53%) followed by the motifs AAG (17.65%), AGC (13.73%) and CCG (13.07%) whereas the motif ACT was the least common (2.61%). However, the TTNRs, PNRs or HNRs were found in insignificant numbers (<5.36%).

The 522 non-redundant SSR–ESTs were used for primer designing. In which, primers could be designed successfully for 246 ESTs. The remainders were inappropriate for primer designing due to short unique domains flanking the microsatellite core. Of the 246 potential SSR–ESTs, 104 were excluded for containing MNRs as the SSR core. Of the remaining 142 ESTs, only 81 were finally used for primer designing under the criteria of amplifying fragments shorter than 300 bp. Sixteen of these 81 SSR–ESTs varied among sparganum isolates employed in heteroduplex analysis, and the variation in only five of these derived from repeat variation. These five loci (C07, I03, M15, M20 and N11) were selected for use in the genetic diversity analysis (Table 2). TTNR was found in one locus (M15), the other four loci bore TNRs.

Diversity analysis and genetic relationship

Total 64 natural sparganum isolates from 11 geographical locations of southwest China were genotyped at each of five selected loci. A total of 12 alleles were observed from 64 individuals. The loci of M20 and N11 identified three alleles, two alleles were observed under C07, I03 and M15 (average, 2.4 alleles per locus). Using the C07 only isolates from SC-LZ, SC-LSZ and SC-LS identified two alleles, others were fixed for a single allele (Table 3). With the I03, each geographical population identified only one allele with the exception of larvae from SC-LZ and SC-LSZ which identified two alleles. Under the loci of M15, M20 and N11, most of populations were observed more than one allele. However, specimens from YN-KM just identified one allele under all five loci. The overall observed heterozygosity (H_o) and expected heterozygosity (H_e) ranged from 0 to 0.9000 and 0 to 0.5737 respectively.

Analysing isolates from each geographical location as a distinct subpopulation enabled pairwise estimates of F_{st} values which are used as a measure of genetic differentiation between populations (Supplementary Table S2). With the exception of F_{st} values between YN-BS, YN-KM and several subpopulations from SC (such as SC-LZ, SC-LSZ, SC-LS and SC-NCys) above 0.25 ($P < 0.05$), most of the F_{st} values were very low (below 0.15) and statistically insignificant ($P > 0.05$). Principal component analysis was
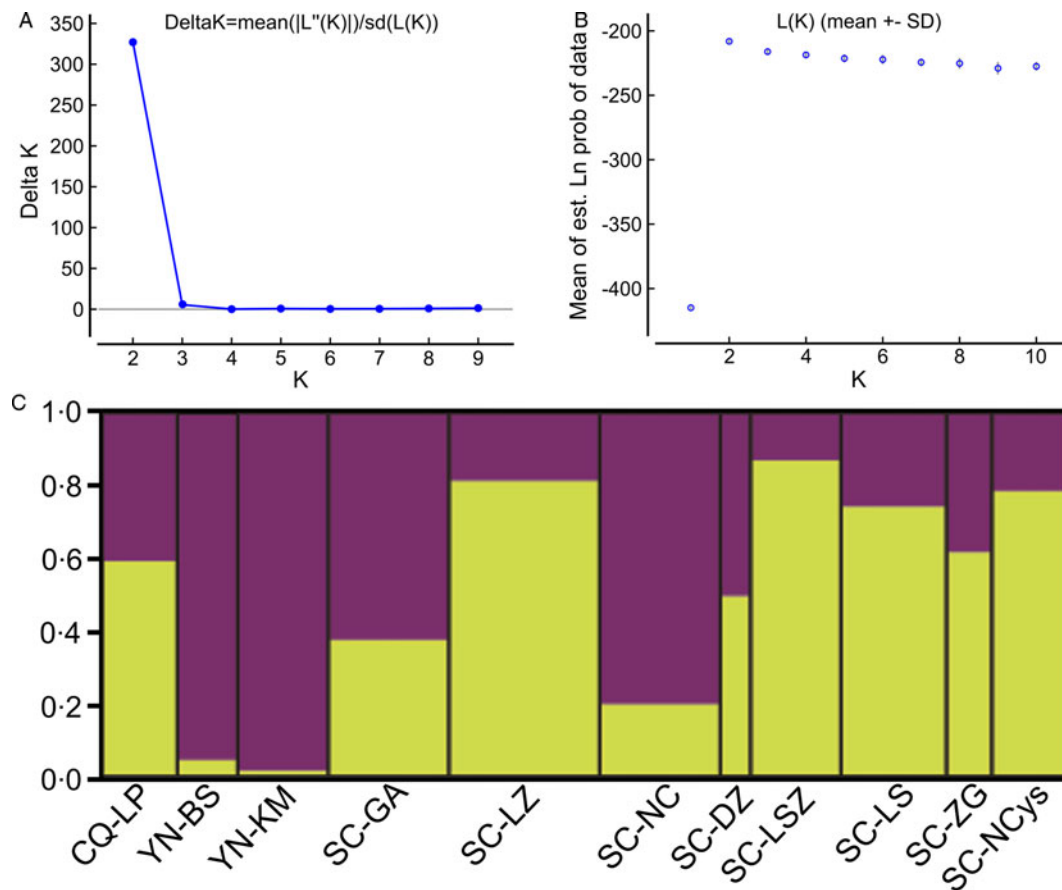
Fig. 3. Estimated population structure of sparganum isolates in China as inferred by STRUCTURE software on the basis of the data of five microsatellite markers obtained for 64 isolates from 11 geographical locations. (A) Plot of the mean posterior probability (Ln $P(D)$) values per clusters (K), based on 20 replicates per K, generated by the STRUCTURE software, and (B) delta-K analysis of Ln $P(K)$. (C) STRUCTURE plots grouped by the Q-matrix (estimated membership coefficient for each sample) showing the distribution of genetic variation at $K = 2$. Each strain is represented by a vertical line, which is partitioned into the coloured segments that represent the parasite estimated membership fractions in K. The same colour indicates that the isolates belong to the same group. Different colours for the same isolate indicate the percentage of the genotype shared with each group. For the isolate code, see Fig. 1.

performed on populations and then on individual larva to explore how variation was partitioned in the sample. Applying the PCA algorithm to the multi-locus F_{st} values (Fig. 2A) separated YN-KM isolates from remaining isolates. Besides, the populations of YN-BS and SC-NC were distributed more separately. Analysing genetic distances among the 64 larvae by PCA showed two main clusters: cluster-1 and cluster-2 (Fig. 2B). Cluster-1 included individuals from SC-NCys (1 individual), SC-ZG (1), SC-LS (1), SC-DZ (1), SC-NC (6), SC-LZ (1), SC-GA (5), YN-KM (6), YN-BS (3) and CQ-LP (2). Cluster-2 contained samples from SC-NCys (4), SC-ZG (1), SC-LS (5), SC-LSZ (5), SC-DZ (1), SC-NC (2), SC-LZ (7), SC-GA (3) and CQ-LP (3). The Bayesian model-based clustering analysis implemented in STRUCTURE indicated that the sparganum isolates from southwest China could be organized derived from the admixture of two ancestral clusters ($K = 2$) based on the results of both the maximum ΔK and maximum $L(K)$ (Fig. 3A and B). The STRUCTURE analysis did not divide the 11 populations according to their geographical origin. In contrast, each population showed a mixed membership to the inferred clusters. Isolates from YN-KM and a sample from YN-BS revealed one prominent genotype, whereas the remaining isolates indicated two main genotypes (Fig. 3C). In consistent with the STRUCTURE analysis, the NJ tree based on the allelic data also revealed two main clusters (Fig. 4). In addition, the two clusters revealed in the phylogenetic tree using the mitochondrial $cox1$ data too (Supplementary Fig. S1). The main patterns of trees generated by the two types of data (allelic data and $cox1$ data) were generally identical with

the exception of two samples from Liangping of Chongqing of China (CQ-LP2 and CQ-LP3).

Discussion

The larval forms (procercoid/plerocercoid) of *Spirometra* species can parasitize in human causing sparganosis. Although four valid species of spirometrid tapeworms have been described worldwide (Kuchta and Scholz, 2017), only *S. erinaceieuropaei* affects humans in China leading to certain symptoms (Liu *et al.*, 2015). It is still unknown whether there are other human infective species or genotypes of *S. erinaceieuropaei* in China. Here, we developed five EST-derived microsatellite markers, and performed a genetic variation analysis of 64 larvae of *Spirometra* collecting from 11 distinct geographical locations in China to determine the usefulness of these loci as genetic markers for examining population structure in *Spirometra* species.

EST databases are an attractive source of EST-based microsatellites for the population genetic analysis (Ellis and Burke, 2007; Bouck and Vision, 2007). However, the frequency and abundance of microsatellites can be highly variable depending on the dataset size (Varshney *et al.*, 2005). The spirometrid tapeworms as neglected parasites for a long time, the corresponding dataset of ESTs is relatively small. Moreover, besides the *S. erinaceieuropaei*, there is almost no EST-dataset for other *Spirometra* species. Although SSRs from *S. erinaceieuropaei* were detected in a panel of 12 481 ESTs, many potential markers had been
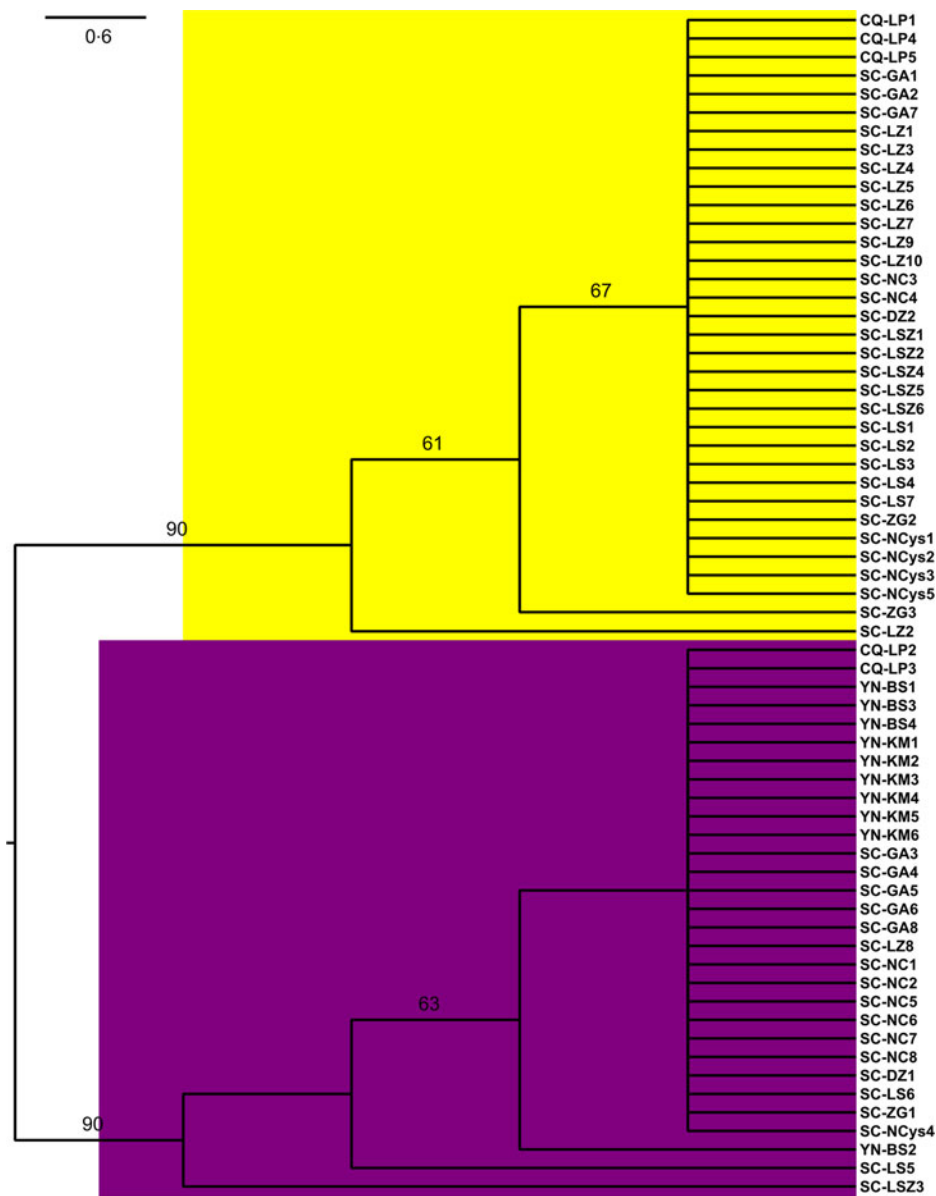
Fig. 4. The unrooted NJ tree based on five microsatellite data of 64 sparganum isolates. Numbers above branches represent the bootstrap values. Only bootstrap values above 60 are shown. For the isolate code, see Fig. 1.

excluded because they did not comply with one or more of criteria described in La Rosa *et al.* (2012). Finally, only five survived our criteria for inclusion.

The genetic diversity in *S. erinaceieuropaei* was first described by Okamoto *et al.* (2007), who detected the level of nucleotide variation in the partial sequence of mitochondrial *cox*1 gene of *S. erinaceieuropaei* collected from Asia, Australia and New Zealand, and concluded that the genetic difference within *S. erinaceieuropaei* was <2.6%. Furthermore, polymorphisms in spargana isolated in different intermediate hosts have been observed using different molecular markers (Wei *et al.*, 2015; Almeida *et al.*, 2016; Jeon *et al.*, 2016; Zhang *et al.*, 2016). Recently, several complete mitochondrial genomes of *Spirometra* have been sequenced and mitogenomic comparative analysis have been made (Liu *et al.*, 2012; Eom *et al.*, 2015; Zhang *et al.*, 2017), which greatly improved the genetic diversity study of *Spirometra* species. As described above, in contrast to the analysis using single or multiple mitochondrial markers, microsatellites can be used for a more powerful exploration of genetic diversity (La Rosa *et al.*, 2012, 2018). Here, we performed a first analysis of sparganum population variability using five microsatellite markers. For these sparganum isolates, little variability was observed within each of geographical population. This was especially evident

among isolates derived from YN-KM (only fixation of one allele). The pairwise F_{st} was used to measure the population differentiation. According to Balloux and Lugon-Moulin (2002), F_{st} value within the range 0–0.05 indicates little genetic differentiation, 0.05–0.15 moderate differentiation, 0.15–0.25 great differentiation and >0.25 very great genetic differentiation. In this study most of the F_{st} values were lower than 0.15 with the exception of F_{st} values between YN isolates and several subpopulations from SC, which were higher than 0.25, indicating very great genetic differentiation, although the interpretation of F_{st} must be qualified in sampling errors or very high genetic differentiation (Steinauer *et al.*, 2010; Zhang *et al.*, 2018). Both STRUCTURE analysis and clustering analysis supported that two genotypes existed among the Chinese sparganum populations, this is in agreement with the analysis based on mitochondrial molecular markers (Zhang *et al.*, 2016). It will be valuable to use the same specimens analysed in previous study and investigate if microsatellites show the same pattern as mitochondrial genes. Therefore, the NJ analysis using the mitochondrial *cox*1 sequences of the same samples selected here was also performed. Interestingly, both of the phylogenetic trees generated by allelic data and *cox*1 data revealed two clusters. This may confirm the presence of these two lineages. Considering that the distribution of *Spirometra* species is

cosmopolitan, the sampling amounts in this study are too small to reflect the true pattern of the genetic structure of spirometrid tapeworms. Therefore, we need for further investigation with deeper sampling (especially in the areas where there are cases of sparganosis) to elucidate the population structure.

In conclusion, the current study describes the effort to develop useful microsatellite markers for plerocercoid larvae of spirometrid tapeworms using available EST databases of *S. erinaceieuropaei*. Of all 12 481 ESTs screened, 522 non-redundant ESTs were identified that contained a total of 915 SSRs. Finally, five loci were selected for use in the next genetic diversity analysis. Using these loci, a total of 12 alleles were detected in 64 sparganum isolates collected from 11 geographical locations. Little variability was observed within each of geographical population, especially among isolates derived from YN-KM. Both STRUCTURE analysis and clustering analysis supported that two genotypes existed among the sparganum isolates from southwest China. To conclude, the five designed and validated microsatellite loci promise future application in population genetic studies on *S. erinaceieuropaei*. We look forward to the application of such markers for tracing and evaluating the genetic diversity of defined zoonotic outbreaks causing by spirometrid tapeworms.

References

Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V and Singh L (2007) Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. *Theoretical and Applied Genetics* **114**, 359–372.

Almeida GG, Coscarelli D, Melo MN, Melo AL and Pinto HA (2016) Molecular identification of *Spirometra* spp. (Cestoda: Diphyllobothriidae) in some wild animals from Brazil. *Parasitology International* **65**, 428–431.

Anantaphruti MT, Nawa Y and Vanvanitchai Y (2011) Human sparganosis in Thailand: an overview. *Acta Tropica* **118**, 171–176.

Balloux F and Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. *Molecular Ecology* **11**, 155–165.

Bazsalovicsova E, Koleničová A, Králova-Hromadová I, Minarik G, Šoltys K, Kuchta R and Štefka J (2018) Development of microsatellite loci in zoonotic tapeworm *Dibothriocephalus latus* (Linnaeus, 1758), Lühe, 1899 (syn. *Diphyllobothrium latum*) using microsatellite library screening. *Molecular and Biochemical Parasitology* **225**, 1–3.

Bouck A and Vision T (2007) The molecular ecologist's guide to expressed sequence tags. *Molecular Ecology* **16**, 907–924.

Earl DA and vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**, 359–361.

Ellis JR and Burke JM (2007) EST-SSRs as a resource for population genetic analyses. *Heredity* **99**, 125–132.

Eom KS, Park H, Lee D, Choe S, Kim KH and Jeon HK (2015) Mitochondrial genome sequences of *Spirometra erinaceieuropaei* and *S. decipiens* (cestoidea: Diphyllobothriidae). *Korean Journal of Parasitology* **53**, 455–463.

Excoffier L and Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564–567.

Falush D, Stephens M and Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* **164**, 1567–1587.

Garbarino C, Interisano M, Chiatante A, Marucci G, Merli E, Arrigoni N, Cammi G, Ricchi M, Tonanzi D, Tamba M, La Rosa G and Pozio E (2017) *Trichinella spiralis* a new alien parasite in Italy and the increased risk of infection for domestic and wild swine. *Veterinary Parasitology* **246**, 1–4.

Gelanew T, Hailu A, Schőnian G, Lewis MD, Miles MA and Yeo M (2014) Multilocus sequence and microsatellite identification of intra-specific hybrids and ancestor-like donors among natural Ethiopian isolates of *Leishmania donovani*. *International Journal for Parasitology* **44**, 751–757.

Hubisz MJ, Falush D, Stephens M and Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* **9**, 1322–1332.

Jakobsson M and Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics (Oxford, England)* **23**, 1801–1806.

Jeon HK, Park H, Lee D, Choe S, Kim KH, Sohn WM and Eom KS (2016) Genetic identification of *Spirometra decipiens* plerocercoids in terrestrial snakes from Korea and China. *Korean Journal of Parasitology* **54**, 181–185.

Kuchta R and Scholz T (2017) Diphyllobothriidea. In Caira JN and Jensen J (eds), *Tapeworms from Vertebrate Bowels of the Earth 2008–2017*. Lawrence, Kansas, USA: University of Kansas, Natural History Museum, Special Publication No. 25, pp. 167–189.

Kuchta R, Scholz T, Brabec J and Wicht B (2015) Chapter 16. *Diphyllobothrium, Diplogonoporus and Spirometra*. In Xiao L, Ryan U and Feng F (eds), *Biology of Foodborne Parasites*. Section III Important Foodborne Helminths, FL, USA: CRC Press, pp. 299–326.

La Rosa G, Marucci G, Rosenthal BM and Pozio E (2012) Development of a single larva microsatellite analysis to investigate the population structure of *Trichinella spiralis*. *Infection, Genetics and Evolution* **12**, 369–376.

La Rosa G, Calero-Bernal R, Pérez-Martín JE, Tonanzi D, Galati F, Serrano-Aguilera FJ, Rosenthal BM and Pozio E (2018) Rare but evolutionarily consequential outcrossing in a highly inbred zoonotic parasite. *International Journal for Parasitology* **48**, 543–553.

Liu GH, Li C, Yuan J, Zhou DH, Xiong RC, Lin RQ, Zou FC and Zhu XQ (2012) Characterization of the complete mitochondrial genome sequence of *Spirometra erinaceieuropaei* (Cestoda: Diphyllobothriidae) from China. *International Journal of Biological Sciences* **8**, 640–649.

Liu Q, Li MW, Wang ZD, Zhao GH and Zhu XQ (2015) Human sparganosis, a neglected food borne zoonosis. *Lancet Infectious Diseases* **15**, 1226–1235.

Okamoto M, Iseto C, Shibahara T, Sato MO, Wandra T, Craig PS and Ito A (2007) Intraspecific variation of *Spirometra erinaceieuropaei* and phylogenetic relationship between *Spirometra* and *Diphyllobothrium* inferred from mitochondrial CO1 gene sequences. *Parasitology International* **56**, 235–238.

Rousset F (2008) Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* **8**, 103–106.

Smouse PE, Whitehead MR and Peakall R (2015) An informational diversity framework, illustrated with sexually deceptive orchids in early stages of speciation. *Molecular Ecology Resources* **15**, 1375–1384.

Steinauer ML, Blouin MS and Criscione CD (2010) Applying evolutionary genetics to schistosome epidemiology. *Infection, Genetics and Evolution* **10**, 433–443.

Swofford DL (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.0b10. Sinauer Associates, Sunderland, MA.

Tang TH, Wong SS, Lai CK, Poon RW, Chan HS, Wu TC, Cheung YF, Poon TL, Tsang YP, Tang WL and Wu AK (2017) Molecular

identification of *Spirometra erinaceieuropaei* tapeworm in cases of human sparganosis, Hong Kong. *Emerging Infectious Diseases* **23**, 665–668.

Varshney RK, Graner A and Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology* **23**, 48–55.

Wei T, Zhang X, Cui J, Liu LN, Jiang P and Wang ZQ (2015) Levels of sparganum infections and phylogenetic analysis of the tapeworm *Spirometra erinaceieuropaei* sparganum in wild frogs from Henan Province in central China. *Journal of Helminthology* **89**, 433–438.

Zhang X, Wang H, Cui J, Jiang P, Lin ML, Zhang YL, Liu RD and Wang ZQ (2016) The phylogenetic diversity of *Spirometra erinaceieuropaei* isolates from southwest China revealed by multi genes. *Acta Tropica* **156**, 108–114.

Zhang X, Duan JY, Shi YL, Jiang P, Zeng DJ, Wang ZQ and Cui J (2017) Comparative mitochondrial genomics among *Spirometra* (cestoda: Diphyllobothriidae) and the molecular phylogeny of related tapeworms. *Molecular Phylogenetics and Evolution* **117**, 75–82.

Zhang X, Shi YL, Han LL, Xiong C, Yi SQ, Jiang P, Wang ZX, Shen JL, Cui J and Wang ZQ (2018) Population structure analysis of the neglected parasite *Thelazia callipaeda* revealed high genetic diversity in Eastern Asia isolates. *PLoS Neglected Tropical Diseases* **12**, e0006165.