HELMINTHOLOGIA, 47, 1:8-19, 2010

SSR data mined from expressed sequence tags of phytoparasitic nematodes

Q. X. HUANG¹, X. Y. CHENG^{1*}, B. Y. XIE², R. M. XU¹

¹College of Life Sciences, Beijing Normal University, Beijing 100875, China, *E-mail: *Chengxy@bnu.edu.cn*; ²Institute of Vegetables and Flowers, Chinese Academy of Agriculture Sciences, Beijing 100081, China

Summary

SSR markers have become the most popular resource for studying population genetic variation in eukaryotes. However, few studies with SSR markers have been carried out with phytoparasitic nematodes. In this study a primary survey on EST-SSRs was made utilizing bioinformatics methods to derive SSRs from expressed sequence tags (ESTs) of 16 species of PPN, which belong to 7 genera and 5 families. The results showed that trinucleotide repeats were the most abundant SSRs in coding ESTs, while tetranucleotide SSRs were predominant in non-coding ESTs and genome sequences. AG/CT, AAC/GTT and (AAAN)_n motifs were predominant, and CG/GC, ACT/AGT motifs were scanty in the ESTs of most genera and species. SSRs were more abundant in non-coding ESTs than in coding ESTs. The distribution frequencies of SSR motifs in coding ESTs, non-coding ESTs and genomes are different. Our results will provide useful information for screening SSRs from each genus and species for further genetic study of phytoparastic nematodes.

Keywords: microsatellites; expressed sequence tags; plantparasitic nematodes

Introduction

Phytoparasitic nematodes (PPN) are an important group in Nematoda. Among the 20,000 species of nematodes described so far, about 3000 are parasites of plants (Castagnone-Sereno, 2002). Many of them are important plant pathogens such as root-knot, cyst, and lesion nematodes. Infestations of these nematodes pose a significant agronomic problem, causing yield losses in many important agronomical plants. PPN may cause US \$77 billion of damage or more worldwide each year (Lambert & Bekal, 2002). Genetic studies can help to understand differences in pathogenicity and the changes that may occur in nematode populations within fields to overcome plant resistance.

Simple sequence repeats (SSRs, also called microsatellites) are effective molecular markers for population genetic study, and are distributed extensively in both coding and non-coding sequences of eukaryotic genomes (Tautz & Renz, 1984). Because SSRs are highly polymorphic, easy to detect (by PCR), specific, and exhibit codominant transmission (Powell et al., 1996), they are widely used in studies of population genetics (Perera et al., 2000; Viard et al., 2001), molecular environmental genetics (Schwartz et al., 2003; Williams et al., 2003), and molecular adaptation (Saint-Laurent et al., 2003; Storz, 2002). But there have been relatively few studies on SSRs in nematodes. The most detailed information comes from studies in the freeliving nematode Caenorhabditis elegans (Haber et al., 2005) and a few animal parasitic nematodes (Criscione et al., 2007; Fisher &Viney, 1996; Grillo et al., 2006; Johnson et al., 2006; Otsen et al., 2000). Only a few PPN SSRs have been developed (De Luca et al., 2002; He et al., 2003; Plantard & Porte, 2003; Thiery & Mugniery, 2000; Zhou et al., 2007).

The rapidly expanding databases of ESTs (expressed sequence tags) provide available data for developing SSRs. Mining SSRs from ESTs through bioinformatic methods is an effective and rapid approach, and could overcome the costly, time-wasting, and complicated traditional ways of developing SSRs. Moreover, EST-SSRs have been found to be significantly more transferable across taxonomic boundaries than traditional SSRs (Sim et al., 2009; Ellis & Burke, 2007) and the levels of polymorphism of transcribed EST-SSR regions appear to be similar to those of genomic-SSR regions (Tehrani et al., 2009). As EST-SSRs generally have fewer null alleles, greater cross-species amplification, and less allelic variability than genomic SSRs, EST-SSRs have been used widely in plant genetic studies (Aggarwal et al., 2007; Park et al., 2005; Wang et al., 2007) and population genetic studies (Ellis & Burke,

Total	B. xylophilus	B. mucronatus	X. index	R. similis	P. vulnus	P. penetrans	H. schachtii	H. glycines	G. rostochiensis	G. pallida	M. arenaria	M . $paranaensis$	M. hapla	M. javanica	M. chitwoodi	M. incognita	Species	
162,487	13,340	3,193	9,351	7,380	5,812	1,920	2,818	24,444	11,851	9,020	5,057	3,710	24,452	7,587	12,218	20,334	Total ESTs	
75.20	5.69	1.38	5.22	2.88	3.10	1.09	1.34	11.98	5.26	5.06	2.34	1.72	10.51	2.94	5.15	9.53	Size (Mbp)	
16,795	452	133	545	556	585	278	270	2,670	1,342	1,448	537	433	3,176	959	1,588	1,823	No.	1
10.34	3.39	4.17	5.83	7.53	10.07	14.48	9.58	10.92	11.32	16.05	10.62	11.67	12.99	12.64	13.00	8.97	% of total	
20,055	503	142	599	598	677	432	315	3,139	1,574	2,264	639	514	3,405	1,083	2,016	2,155	anning SSI No. of SSRs	
266.68	88.41	102.59	114.67	207.28	218.55	396.75	234.62	262.03	299.30	447.67	272.95	298.96	323.98	368.00	391.36	226.11	counts/ Mbp	,
22,260	2,054	426	1,223	1,152	373	253	459	3,778	2,047	1,192	697	753	2,999	666	1,321	2,534	No. of Contigs	
39,036	4,277	1,760	3,196	2,658	1,319	162	783	4,680	2,246	3,850	1,867	851	3,524	2,255	2,338	3,270	No. of Singlets	
61,296	6,331	2,186	4,419	3,810	1,692	415	1,242	8,458	4,293	5,042	2,564	1,604	6,523	3,254	3,659	5,804	No. of NR ESTs	-
32.69	3.09	1.00	2.82	1.71	0.94	0.21	0.65	4.98	2.25	3.37	1.28	0.81	3.34	1.43	1.77	3.05	Size (Mbp)	
7,662	264	86	317	312	317	45	155	1,187	629	666	319	205	888	499	611	817	No.	
12.50	4.17	4.48	7.17	8.19	18.74	10.84	12.48	14.03	14.65	19.81	12.44	12.78	13.61	15.33	16.70	14.08	<u>K ESTs co</u> % of total	
9,841	288	105	360	340	383	50	183	1,481	780	1,731	395	256	1,036	585	864	1,004	No. of SSRs	•
301.01	93.30	104.63	127.62	198.71	407.59	232.74	283.09	297.49	346.71	514.36	308.96	315.54	310.39	409.32	488.15	329.19	counts/ Mbp	2

Ŧ
a
Ĕ
2
.—
⊳
ğ
8
ā
20
õ
č
H,
S
\mathbf{S}
R
0
e
<u> </u>
ē
d
fr
2
ы
Ξ
S
Гs
â
n
-
Б
ņ
÷
ď.
E
ď
21
It
Ξ
S_{2}
S
0
÷
16
E.
Ť
Ž
Ś
g
ğ.
ē.

Motif	Meloidogyne	Heterodera	Radopholus	Pratylenchus	Xiphinema	Globodera	Bursaphelenchu
)/GT	83 (7.1) [17.2]	23 (4.1) [11.3]	28 (16.4) [60.9]	19 (16.4) [47.5]	11 (3.9) [37.9]	111 (19.8) [18.6]	13 (3.2) [20.3]
i/CT	298 (25.5) [61.7]	164 (29.2) [80.8]	16 (9.4) [34.8]	17 (9.4) [42.5]	8 (2.8) [27.6]	322 (57.4) [54.0]	35 (8.6) [54.7]
/AT	90 (7.7) [18.6]	10(1.8)[4.9]	1 (0.6) [2.2]	$4\ (0.6)\ [10.0]$	8 (2.8) [27.6]	122 (21.7) [20.5]	15 (3.7) [23.4]
/GC	12 (1.0) [2.5]	6 (1.1) [3.0]	1 (0.6) [2.2]	0(0.6)[0.0]	2 (0.7) [6.9]	41 (7.3) [6.9]	$1\ (0.2)\ [1.6]$
C/GTT	873 (74.8) [40.3]	152 (27.0) [16.6]	26 (15.2) [16.1]	35 (20.5) [14.9]	15 (5.3) [8.4]	189 (33.7) [20.4]	39 (9.5) [17.2]
G/CTT	333 (28.5) [15.4]	88 (15.7) [9.6]	19 (11.1) [11.8]	28 (16.4) [11.9]	21 (7.4) [11.7]	70 (12.5) [7.6]	83 (20.3) [36.6]
T/ATT	494 (42.3) [22.8]	57 (10.1) [6.2]	22 (12.9) [13.7]	13 (7.6) [5.5]	16 (5.7) [8.9]	62 (11.0) [6.7]	4 (1.0) [1.8]
C/GGT	139 (11.9) [6.4]	65 (11.6) [7.1]	8 (4.7) [5.0]	19 (11.1) [8.1]	41 (14.5) [22.9]	75 (13.4) [8.1]	32 (7.8) [14.1]
G/CGT	3(0.3)[0.1]	67 (11.9) [7.3]	7 (4.1) [4.3]	4 (2.3) [1.7]	12 (4.3) [6.7]	67 (11.9) [7.2]	2 (0.5) [0.9]
T/AGT	32 (2.7) [1.5]	0 (0.0) [0.0]	0(0.0)[0.0]	1(0.6)[0.4]	1(0.4)[0.6]	5 (0.9) [0.5]	4 (1.0) [1.8]
C/GCT	42 (3.6) [1.9]	183 (32.6) [20.0]	33 (19.3) [20.5]	26 (15.2) [11.1]	26 (9.2) [14.5]	153 (27.3) [16.5]	13 (3.2) [5.7]
G/CCT	107 (9.2) [4.9]	188 (33.5) [20.6]	23 (13.5) [14.3]	70 (40.9) [29.8]	13 (4.6) [7.3]	140 (24.9) [15.1]	15 (3.7) [6.6]
C/GAT	137 (11.7) [6.3]	60 (10.7) [6.6]	16 (9.4) [9.9]	28 (16.4) [11.9]	22 (7.8) [12.3]	66 (11.8) [7.1]	23 (5.6) [10.1]
	7 (0.6) [0.3]	53 (9.4) [5.8]	7 (4.1) [4.3]	11 (6.4) [4.7]	12 (4.3) [6.7]	98 (17.5) [10.6]	12 (2.9) [5.3]

Table 2. Total number and frequency of different repeat types in ESTs of 7 PPN genera*

Average distance (kb)

Total SSRs

Hexa-

Penta-

 $\begin{array}{c} 4,140 \ (354.6) \\ 1,664 \ (296.1) \end{array}$

115 (9.9) [2.8] 56 (10.0) [3.4] 9 (5.3) [2.6]

1,34 2(114.9) [32.4]

2,167 (185.6) [52.3] 913 (162.5) [54.9] 161 (94.1) [47.4]

483 (41.4) [11.7] 203 (36.1) [12.2] 46 (26.9) [13.5]

Meloidogyne

Genus

Ē

Ē

Tetra-

468 (83.3) [28.1]

340 (198.7) 433 (375.1) 360 (127.7)

 $\begin{array}{c} 33 \left(2.8 \right) \left[0.8 \right] \\ 24 \left(4.3 \right) \left[1.4 \right] \\ 0 \left(0.0 \right) \left[0.0 \right] \\ 4 \left(3.5 \right) \left[0.9 \right] \\ 45 \left(16.0 \right) \left[12.5 \right] \\ 70 \left(12.5 \right) \left[2.8 \right] \end{array}$

2.825.035.037.847.8410.41

3.32

9,841 (301.01)

 $\begin{array}{c} 4 \\ 180 \\ (5.5) \\ 188 \\ 1.8 \\$

393 (96.1)

2,511 (447.2)

 $\begin{array}{c} 17 (14.7) [3.9] \\ 27 (9.6) [7.5] \\ 191 (34.0) [7.6] \\ 19 (4.6) [4.8] \\ 434 (13.3) [4.4] \end{array}$

124 (72.5) [36.5] 137 (118.7) [31.6] 80 (28.4) [22.2] 729 (129.8) [29.0] 79 (19.3) [20.1]

> 235 (203.6) [54.3] 179 (63.5) [49.7] 925 (164.7) [36.8] 227 (55.5) [57.8] 4,807 (147.0) [48.8]

40 (34.7) [9.2] 29 (10.3) [8.1] 596 (106.1) [23.7]

Pratylenchus

Xiphinema Globodera

Heterodera Radopholus 64 (15.6) [16.3] 1,461 (44.7) [14.8]

Bursaphelenchus

Total

2,959 (90.5) [30.1]

2007; Slate et al., 2007; Wang et al., 2006; Kim et al., 2008). Furthermore, as EST-SSRs are derived from coding sequences, they could be involved in regulating gene expression (Kunzler et al., 1995; Moxon & Wills, 1999), and the varieties of repeat units might reflect gene expression (Wells & Warren, 1998).

For the reasons discussed above, a preliminary survey of EST-SSRs was made based on currently available PPN EST data. Similarities and differences among EST-SSRs are summarized. The data provide useful information for screening SSRs in further study of PPN genetics.

Materials and methods

Sequence data sources

ESTs of all species studied were downloaded from GenBank (ftp://ftp.ncbi.nih.gov/blast/db, data on October 16, 2009). As shown in Table 1, a total of 162,487 ESTs were selected from 16 PPN species, belonging to 7 genera and 5 families. The species are listed as follows: 6 species of Meloidogynidae Results [Meloidogyne incognita (Kofoid & White); M. chitwoodi Golden, O'Bannon, Santo & Finley; M. javanica (Treub); M. hapla Chitwood; M. paranaensis Carneiro, Carneiro, Abrantes, Santos & Almeida; and *M. arenaria* (Neal)]; 4 species of Heteroderidae [Heterodera glycines Ichinohe, H. schachtii Schmidt, Globodera pallida (Stone), and G. rostochiensis (Wollenweber)]; 3 species of Pratylenchidae [Radopholus similes (Cobb), Pratylenchus penetrans (Cobb), and P. vulnus Allen & Jensen]: 1 species of Longidoridae [Xiphinema index Thorne & Allen]; 2 species of Aphelenchoididae [Bursaphelenchus xvlophilus (Steiner & Buhrer) and B. mucronatus Mamiya & Enda]. In addition, the genome sequences of Meloidogyne inkognita (Abad et al., 2008) (version WS205, ~82.1 Mb) and *M. hapla* (Opperman et al., 2008) (version WS205, ~53.0 Mb) were also downloaded from wormbase (ftp://ftp.wormbase.org/pub/ wormbase/) for a comparison.

Purifying EST data

The stand-alone TRIMEST program of EMBOSS package (Rice et al., 2000) was used for removing the poly A/T ends of the EST sequences. The program was run using the following settings: (i) minimum length of a poly-A tail (integer) = 4; (ii) number of contiguous mismatches allowed in a tail (integer) = 1; (iii) write the reverse complement Pratylenchus (375.1/Mbp), Meloidogyne (354.6/Mbp), when poly-T is removed (yes); and (iv) remove poly-T tails Heterodera (296.1/Mbp), Radopholus (198.7/Mbp), and at the 5' end of the sequence (yes). Vector DNA was screened by CrossMatch (Phil Green, unpublished) using default parameters. Vector sequences for comparisons were obtained from GenBank (ftp://ftp.ncbi.nih.gov/blast/db/ FASTA/vector.gz).

EST assembly

The purified ESTs of each species were assembled by running DNA sequence assembly program CAP3 (Huang & Madan, 1999) using default parameters. Redundant ESTs were assembled into 'contigs', and unique ESTs were recognized as 'singletons'. Thus, non-redundant EST (NR ESTs) datasets were built.

Classification of coding ESTs and non-coding ESTs

All the NR ESTs were classified into 'coding' and 'noncoding' groups by the following steps: (i) Translated into amino acids in all six open reading frames (ORFs) by using the GETORF program of EMBOSS package (Rice et al., 2000) (minimum size 150 nt); (ii) ESTs generating an ORF of no less than 50 amino acids were recognized as 'coding' ESTs; and the remaining ESTs were classified to be 'non-coding' ESTs.

SSR mining

The identification and localization of SSRs were achieved using SciRoKo (Kofler et al., 2007), searching for the SSRs containing motifs from 2 to 6 bp (di-, tri-, tetra-, penta-and hexanucleotide), no less than 5, 4, 3, 3, and 3 repeat times, respectively. SSRs in the pre-assembled ESTs, non-redundant ESTs, coding ESTs, non-coding ESTs and genome sequences were derived respectively.

Frequency of SSRs in ESTs of PPN

A total of 162,487 ESTs (total length 75.20 Mbp) of 16 PPN species were selected for SSR mining in this study. After removing low-quality sequences (poly A/T ends, vector sequences), SSRs were mined from these pre-assembled ESTs. As shown in Table 1, 20,055 SSR loci were recognized from 16,795 ESTs (10.34 % of the ESTs), with an average of 266.68 SSR loci per Mbp. By running the CAP3 program, 61,296 non-redundant ESTs (including 22,260 contigs and 39,036 singlets, total length 32.69 Mbp) were assembled. A total of 9,841 SSR loci were mined from 7,662 NR ESTs (12.50 % of total NR ESTs). An average of 301.01 SSR loci was derived per Mbp.

Although a bias could be induced by the numbers of ESTs in each species, no clear relationship was observed between the richness of SSRs and the number of ESTs from Table 1. Distribution frequencies of EST-SSRs were obviously different among PPN genera and species. Among them, G. pallida had the most abundant EST-SSRs (514.36/Mbp), and B. xylophilus had the fewest (93.30/Mbp) (Table 1). As shown in Table 2, Globodera had the highest density (447.2/Mbp), followed in order by Xiphinema (127.7/Mbp). Bursaphelenchus had the lowest density (96.1/Mbp). Within a genus, the distribution frequencies of EST-SSRs were also obviously distinct among species (Table 1). In the genus Meloidogvne, M. chitwoodi had an extra high density (488.15/Mbp), much higher than the other species (308.96-409.32/Mbp). Similarly, the frequencies of EST-SSRs were also quite different between G. pallida and G. rostochiensis (514.36/Mbp vs. 346.71/Mbp), and between P. penetrans and P. vulnus (232.74/Mbp vs. 407.59/Mbp). The differences among species of Heterodera and Bursaphelenchus, however, were not so evident.

Table 4. Frequency of tetranucleotide and pentanucleotide repeats in 7 PPN genera*

Motif	Globodera	Bursaphelenchus	Meloidogyne	Heterodera	Pratylenchus	Xiphinema	Radopholus
Tetra-	AAAT(22.62)	AAAT(9.05)	AAAT(64.92)	AATT(17.25)	ATCC(30.32)	AAAG(2.48)	AATG(11.69)
	AAAC(13.71)	AAAG(1.96)	AATT(26.81)	AAAT(14.22)	AAAT(19.92)	AATG(2.48)	AAAT(11.1)
	AAAG(13.53)	AAAC(1.47)	AAAC(7.02)	AATG(11.38)	AATG(12.13)	AAAC(2.13)	AAAC(10.52)
	AGGG(9.97)	AACC(0.73)	AAAG(4.2)	AAAC(5.87)	AAGG(11.21)	AACC(2.13)	AATC(7.60)
_	AATG(9.44)	AATC(0.73)	AATG(2.4)	AAAG(5.69)	AAAC(11.26)	ACCC(1.77)	ATCC(4.68)
Penta-	AAAAG(4.63)	CCCGG(1.22)	AAAAT(4.11)	AAAAT(1.42)	ACCCC(1.73)	AAAAT(1.42)	AAAAT(1.17)
	AAAAC(4.27)	AATAT(0.73)	AAATT(2.14)	AAAAC(1.24)	AAAAG(1.73)	AATTC(1.06)	AAAAG(0.58)
	AGGGG(4.1)	AAATT(0.49)	AAAAC(0.94)	AAATT(0.71)	AAATT(1.73)	AGAGG(0.71)	AAATT(0.58)
	AAAAT(2.67) *Number	AAACC(0.49) in parentheses show	AAAAG(0.77)	AAATC(0.53) Rs (counts/Mbp)	AATAC(1.73) Only predomination	AAGTG(0.71) nt motifs are liste	AAAGG(0.58)

Types of EST-SSRs in PPN

Distribution frequencies of different repeat types of SSRs in PPN genera are shown in Table 2. Trinucleotide repeats were the most abundant (48.8 % of total), followed by tetranucleotide repeats (30.1 %) and dinucleotide repeats (14.8 %). Pentanucleotide and hexanucleotide repeats were less abundant (4.4 % and 1.8 % respectively). The relationship between the number and the repeat times is shown in Figure 1. It was obvious that the numbers of SSRs decreased with increasing repeat times in all types, except for dinucleotide SSRs of nine repeat times. This indicated that SSRs with the fewer repeat times were more abundant. Distribution frequencies of different SSR motifs in PPN

are shown in Table 3 and Table 4. Among dinucleotide repeats, AG/CT were prevailing motifs in most genera, except that AC/GT motifs were abundant in *Radopholus*, *Pratylenchus* and *Xiphinema*. However, CG/GC motifs are few in all genera.

For trinucleotide repeats, the most predominant SSRs in each genus were different, e.g., AAC/GTT motifs in *Meloidogyne* and *Globodera*, AGG/CCT in *Heterodera* and *Pratylenchus*, AGC/GCT in *Radopholus* and *Heterodera*, ACC/GGT in *Xiphinema*, and AAG/CTT in *Bursaphelenchus*. The AAC/GTT motifs were the most widely abundant in all genera, and ACT/AGT motifs were scanty in most genera.



Fig.1 Distribution of different repeat units of SSRs in ESTs of 7 PPN genera

paranaensis M. arenaria 4.69	paranaensis M. arenaria 4.69 G. pallida 27.93	paranaensis M. arenaria 4.69 G. pallida 27.93 G. 7.56	paranaensis M. arenaria 4.69 G. pallida 27.93 G. 7.56 rostochiensis	paranaensis M. arenaria 4.69 G. pallida 27.93 G. 7.56 rostochiensis H. glycines 4.42	paranaensis M. arenaria 4.69 G. pallida 27.93 G. 7.56 rostochiensis H. glycines 4.42 H. schachtii 0.00	paranaensis M. arenaria 4.69 G. pallida 27.93 G. 7.56 rostochiensis H. glycines 4.42 H. schachtii 0.00 P. penetrans 0.00	paranaensis M. arenaria 4.69 G. pallida 27.93 G. rostochiensis H. glycines 4.42 H. schachtii 0.00 P. penetrans 0.00 P. vulnus 20.22	paranaensisM. arenaria4.69G. pallida27.93G.7.56rostochiensis7.56H. glycines4.42H. schachtii0.00P. penetrans0.00P. vulnus20.22R. similis9.35	paranaensisparanaensisM. arenaria4.69G. pallida27.93G. rostochiensisH. glycinesH. glycinesH. schachtii0.00P. penetrans0.00P. vulnus20.22R. similis9.35X. index3.90	paranaensisM. arenaria4.69G. pallida27.93G.7.56rostochiensis7.56H. glycines4.42H. schachtii0.00P. penetrans0.00P. vulnus20.22R. similis9.35X. index3.90B. mucronatus2.99	paranaensisparanaensisM. arenaria4.69G. pallida27.93G. rostochiensisH. glycinesH. glycinesP. penetrans0.00P. vulnus20.22R. similis9.35X. index9.35X. index3.90B. mucronatus2.99B. xylophilus3.24Table 6. Total number and	paranaensisparanaensisM. arenaria4.69G. pallida27.93G. pallida7.56rostochiensisH. glycines4.42H. schachtii0.00P. penetrans0.00P. vulnus20.22R. similis9.35X. index9.35B. mucronatus2.99B. xylophilus3.24Table 6. Total number and	paranaensisparanaensisM. arenaria4.69G. pallida27.93G. rostochiensisH. glycines7.56rostochiensisH. schachtii0.00P. penetrans0.00P. vulnus20.22R. similis9.35X. index3.90B. mucronatus2.99B. xylophilus3.24Table 6. Total number andTypeCodingI	paranaensis M. arenaria 4.69 G. pallida 27.93 G. rostochiensis H. glycines 4.42 H. schachtii 0.00 P. penetrans 0.00 P. vulnus 20.22 R. similis 9.35 X. index 3.90 B. mucronatus 2.99 B. mucronatus 2.99 B. xylophilus 3.24 Table 6. Total number and Type Coding 1,349 Di- 1,349 [14.4]	$ paranaensis \\ paranaensis \\ M. arenaria 4.69 \\ G. pallida 27.93 \\ G. 7.56 \\ rossochiensis \\ H. glycines 4.42 \\ H. schachtii 0.000 \\ P. penetrans 0.000 \\ P. vulmus 20.22 \\ R. similis 9.35 \\ X. index 3.90 \\ B. mucronatus 2.99 \\ B. mucronatus 3.24 \\ \hline \\ Table 6. Total number and \\ \hline \\ Type Coding ESTs 1 \\ I.,349 \\ Di- [14.4] \\ 4.706 \\ Tri- (147.14) 10 \\ [50.1] \\ $	paranaensis M. arenaria 4.69 G. pallida 27.93 G. pallida 27.93 G. 7.56 rostochiensis 4.42 H. gdycines 4.42 H. schachtii 0.00 P. penetrans 0.00 P. vulnus 20.22 R. similis 9.35 X. index 3.90 B. mucronatus 2.99 B. mucronatus 2.99 B. mucronatus 2.99 B. mucronatus 2.99 B. mucronatus 1.00 F. vulnus 1.00 P. vulnus 1.00 [50.1] 2.748 Tetra- (85.92) 21 Vita 1.00 [20.2] 21 Vita 1.00 [20.2] P. vulnus 1	$ \begin{array}{c} paramensis \\ paramensis \\ M. arenaria \\ 4.69 \\ G. pallida \\ 27.93 \\ G. \\ 7.56 \\ rossochiensis \\ H. glycines \\ 4.42 \\ H. schachtii \\ 0.00 \\ P. penetrans \\ 0.00 \\ P. penetrans \\ 0.00 \\ P. vulnus \\ 20.22 \\ R. similis \\ 9.35 \\ X. index \\ 3.90 \\ B. mucronatus \\ 20.22 \\ R. similis \\ 9.35 \\ X. index \\ 3.90 \\ B. mucronatus \\ 2.99 \\ B. xylophilus \\ 3.24 \\ \hline \\ Table 6. Total number and \\ \hline \\ Type \\ Coding \\ ESTs \\ 1 \\ 1 \\ STs \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ $	paramaensis 4.69 M. arenaria 4.69 G. pallida 27.93 G. pallida 7.56 rostochiensis 4.42 H. glycines 0.00 P. penetrans 0.00 P. vulnus 20.22 R. similis 9.35 X. index 3.90 B. mucronatus 2.99 B. xylophilus 3.24 Table 6. Total number and 1 Type Coding ESTs 1 Di- (12.18) 11 [14.4] 4,706 10 Tri- (147.14) 10 [50.1] 2.748 21 Penta- (85.92) 21 Penta- 415 (12.98) 1 Hexa- 178 (5.57) 1
83.79 34	17.78 3.1		30.13 2.0	21.66 0.0	1000	23.27 0.0	23.27 0.0 12.77 4.2	23.27 0.0 12.77 4.2 16.36 0.0	23.27 0.0 12.77 4.2 16.36 0.0 2.84 2.8	23.27 0.0 12.77 4.2 16.36 0.0 2.84 2.8 3.99 1.9	23.27 0.1 12.77 4.2 16.36 0.0 2.84 2.8 3.99 1.9 10.04 4.2 1 frequency of SS	23.27 0.1 12.77 4.2 16.36 0.0 2.84 2.8 3.99 1.9 10.04 4.2 1 frequency of SS	2.3.27 12.77 16.36 2.84 3.99 10.04 4.2 3.99 1.9 1.9 1.9 1.9 1.9 1.9 1.9 1.9 1.9	23.27 12.77 16.36 2.84 3.99 10.04 4.2 10.04 4.2 10.04 4.2 10.04 4.2 110.04 4.2 110.04 4.2 110.04 4.2 110.04 1.2 1.2 1.2 1.2 1.2 1.2 1.2 1.2 1.2 1.2	23.27 12.77 16.36 2.84 3.99 10.04 4.2 10.04 4.2 1.9 1.9 1.9 1.9 1.9 1.9 1.9 1.9 1.9 1.9	23.27 12.77 16.36 2.84 3.99 10.04 4.2 10.04 4.2 10.04 4.2 10.04 4.2 11 10.04 4.2 11 10.04 4.2 11 12 11 12 157.66) [25 12 11 (142.18) [22 147]	23.27 12.77 16.36 2.84 3.99 1.004 4.2 10.04 4.2 1.9 1.9 1.9 1.9 1.004 4.2 1.9 1.9 1.9 1.9 1.9 1.9 1.9 1.9 1.9 1.9	$\begin{array}{ccccccc} 2.3.27 & 0.0\\ 12.77 & 4.2\\ 16.36 & 0.0\\ 2.84 & 2.8\\ 3.99 & 1.9\\ 10.04 & 4.2\\ 10.04 & 4.2\\ 10.04 & 4.2\\ 11 & (1.04 & 4.2\\ 11 & (1.04 & 4.2\\ 11 & (1.04 & 4.2\\ 12 & (1.04 & 1.2)\\ 12 & (1.04 & 1$
17 11.89	0.00		1.21	0.00	0.00	0.00	0.00	17 N N 1	0./ F	0.00	R motifs derived	P9 0.00 21 0.00 R motifs derived	R motifs derived Ts Coding	R motifs derived R motifs derived Ts Coding 2] 98 (32.9	P9 0.00 21 0.00 R motifs derived Ts Coding 2] 98 (32.9 7] 557 (187.	$\begin{array}{c} \begin{array}{c} \begin{array}{c} \begin{array}{c} \begin{array}{c} \begin{array}{c} \begin{array}{c} \begin{array}{c} $	$\begin{array}{c} \begin{array}{c} \begin{array}{c} \begin{array}{c} \begin{array}{c} \begin{array}{c} \begin{array}{c} \begin{array}{c} $	R motifs derived Ts Coding [2] 98 (32.9 [4] 239 (80.3 [4] 239 (80.3 [1] 27 (9.0 [1] 27 (9.0
30.01	39.12		26.92	27.85	37.24	28.73	15.20	5.32		9.96	9.96 9.39 from non-red	9.96 9.39 from non-red	9.96 9.39 from non-red	9.96 9.39 from non-redu g ESTs 5) [10.5]	9.96 9.39 from non-redu g ESTs 5) [10.5] 5) [10.5]	9.96 9.39 from non-redu <u>3</u> ESTs 5) [10.5] 5) [10.5] 26) [59.8] 26) [59.8]	9.96 9.39 from non-redu g ESTs 5) [10.5] 5) [10.5] 26) [59.8] 26) [59.8] 26) [25.7]	9.96 9.39 from non-redu g ESTs 5) [10.5] 5) [10.5] 26) [59.8] 26) [59.8] 35) [25.7] 8) [2.9]
14.56	9.33		14.86	21.66	18.62	25.54	11.10	777	1.44	7.44 21.92	7.44 21.92 19.76 undant ESTs, J	7.44 21.92 19.76 M. inco	7.44 21.92 19.76 <i>M. inco</i> Non-coding	7.44 21.92 19.76 M. <i>inco</i> Non-coding 32 (424.17)	21.92 19.76 undant ESTs, J <i>M. inco</i> Non-coding 32 (424.17) 24 (318.13)	1.44 21.92 19.76 <i>M. inco</i> Non-coding 32 (424.17) 24 (318.13) 16 (212.09)	1.44 21.92 19.76 <i>M. inco.</i> Non-coding 32 (424.17) 32 (424.17) 24 (318.13) 16 (212.09) 1 (13.26)	1.44 21.92 19.76 <i>M. inco.</i> Non-coding 32 (424.17) 32 (424.17) 32 (424.17) 16 (212.09) 16 (212.09) 1 (13.26) [
12.48	8.89		9.64	13.92	4.76	11.71	12.86	5.67	0 0 0	0.00	0.97 predicted co	0.00 0.97 oredicted co gnita	0.00 0.97 predicted co gnita ESTs	0.97 0.97 gnita ESTs [43.8]	0.97 0.97 gnita ESTs [43.8] [32.9] 8	0.97 0.97 gnita ESTs [43.8] [43.8] [32.9] § [32.9] §	<u>oredicted co</u> gnita [43.8] [32.9] { [21.9] 1 [1.4]	0.97 0.97 0.97 0.97 0.97 0.97 0.97 0.97
15.15	10.67		11.85	9.28	0.00	20.22	4.68	14.53		9.96	9.96 7.13 ding and non-c	9.96 7.13 ding and non-c	9.96 7.13 ding and non-c Genom	9.96 7.13 ding and non-e Genom 3,099 (37.75)	9.96 7.13 ding and non-e Genom 3,099 (37.75) 3,294 (101.03	9.96 7.13 Genom 3,099 (37.75) 8,294 (101.03 2,330 (150.19	9.96 7.13 Genom 3,099 (37.75) 3,294 (101.03 8,294 (101.03 8,294 (150.19 1,845 (22.47	9.96 7.13 Genom 3,099 (37.75) 3,294 (101.03 8,294 (101.03 2,330 (150.19 1,845 (22.47 1,726 (21.02
9.21	16.00		12.86	4.64	0.00	4.26	4.09	4.25		0.00	0.00 0.00 oding ESTgr	0.00 0.00 oding ESTgr	0.00 0.00 oding ESTgr	0.00 0.00 oding ESTgr e [11.4]	0.00 0.00 oding ESTgr e [11.4] [11.4] [30.4] 2	oding ESTgrc e [11.4] [30.4] 2 [45.2] 3	0.00 0.00 0.00 0.00 0.00 0.00 [11.4] [11.4] [11.4] [11.4] [11.4] [11.4] [11.4] [11.4] [11.4] [11.4] [11.4] [11.4] [11.4] [11.4] [11.4] [11.4] [11.4] [11.6]	oding ESTgr e [11.4] [30.4] 2 [6.8] [6.3]
0.00	1.78		0.00	0.00	0.00	0.00	0.00	0.19		0.00	0.00 0.97 ups, and genc	0.00 0.97 oups, and genc	0.00 0.97 nups, and gene Coding ES	0.00 0.97 oups, and genc Coding ES 77 (23.92)	0.00 0.97 oups, and genc Coding ES 77 (23.92) 145 (138.22)	0.00 0.97 Coding Es 77 (23.92) 145 (138.22)	0.00 0.97 Coding ES 77 (23.92) 78 (117.41) 35 (10.87)	0.00 0.97 0.97 Coding ES 77 (23.92) 77 (23.92) 145 (138.22) 145 (117.41) 35 (10.87) 8 (2.48) [C
23.18	33.34		32.14	35.58	27.93	21.28	19.29	9.22	5.98		2.27 me sequenc	2.27 me sequenc	2.27 me sequenc 5Ts N	2.27 me sequenc STs N [8.2] 2	2.27 me sequenc 3Ts N [8.2] 2 [47.2] 1	<u>2.27</u> me sequenc 3Ts N [8.2] 2 [47.2] 1 [40.1] 4	<u>2.27</u> me sequenc <u>3.75</u> N [47.2] 17 [40.1] 4; [40.1] 4;	<u>2.27</u> me sequenc 3Ts N [8.2] 2 [8.2] 2 [40.1] 4 [40.1] 4 [3.7] .8]
21.69	29.78		34.55	24.75	37.24	65.98	13.44	4.61	100	4.98	4.98 3.24 es of two plant para	4.98 3.24 es of two plant pare M. hapla	4.98 3.24 as of two plant para <i>M. hapla</i> on-coding ESTs	4.98 3.24 3: of two plant para M. hapla on-coding ESTs 7 (228.22) [29.0]	4.98 3.24 s of two plant para M. hapla on-coding ESTs 7 (228.22) [29.0] 7 (143.70) [18.3]	4.98 3.24 3.24 M. hapla M. hapla on-coding ESTs 7 (228.22) [29.0] 7 (143.70) [18.3] 3 (363.47) [46.2]	4.98 3.24 M. hapla on-coding ESTs 7 (228.22) [29.0] 7 (143.70) [18.3] 3 (363.47) [46.2] 5 (42.26) [5.4]	4.98 3.24 3.24 M. hapla on-coding ESTs 7 (228.22) [29.0] 7 (143.70) [18.3] 7 (143.70) [18.3] 5 (363.47) [46.2] 5 (42.26) [5.4] 1 (8.45) [1.1]
12.18	11.11		11.65	3.09	18.62	25.54	9.35	7.80	6.98		5.18 asitic nematc	5.18 asitic nemato	5.18 tsitic nemato Gen	5.18 tsitic nemato Gen 1,7 (33.65	5.18 stiic nemato Gen (33.65 (110.55	5.18 Gen 1,7 (33.65 (110.55 (110.55 (282.1)	5.18 Gen 1,7 (33.65 (110.55 (110.55 (282.1) (33.57	5.18 stiic nemato Gen 1,7 (33.65 (110.55 (110.55 (1282.1) (282.1) (33.57) (33.57)
13.97	22.67		9.84	6.19	0.00	11.71	4.09	4.25	1.99)) ,	3.24 des*	3.24 .des*	3.24 des*	3.24 des* ome 84 [7.2]	3.24 des* ome)[7.2])[7.2] (61 (61 (23.7]	3.24 des* ome)[7.2] 84)[7.2] 0[7.2])[23.7])[23.7] 0[60.5]	<u>3.24</u> ome)[7.2])[7.2])[23.7])[23.7])[26.5])[60.5])[60.5]	<u>3.24</u> des* <u>anne</u> (7.2] (7.2] (61 (61 (61.5] (60.5] (60.5] (60.5] (7.2] (7.2]

Table 7. Frequency of di- and tri- and tetranucleotide SSR motifs derived from non-redundant ESTs, predicted coding and non-coding ESTgroups, and genome sequences of two plant parasitic nematodes*

Modif	ND PCT.	Coding BCTs	Mon anding DCT ₆		M. incognita			M. hapla	
THOM	SICE NN	COULING ENTS	ST CEI BIIIDOD-HONI	Coding ESTs	Non-coding ESTs	Genome	Coding ESTs	Non-coding ESTs	Genome
AC/GT	8.44	8.41	9.85	16.84	0.00	7.14	1.86	00.0	2.85
AG/CT	26.67	24.61	119.65	9.09	362.50	18.48	13.35	208.33	12.73
AT/AT	7.65	7.22	26.75	4.04	0.00	12.02	8.39	0.00	17.73
CG/GC	1.93	1.94	0.00	3.03	0.00	0.11	0.00	0.00	0.34
AAC/GTT	40.65	40.58	43.64	100.67	162.50	42.27	45.96	58.33	34.69
AAG/CTT	19.64	19.6	21.12	28.28	62.50	13.28	25.78	25.00	14.98
AAT/ATT	20.43	20.1	35.19	20.88	50.00	24.48	37.27	41.67	43.78
ACC/GGT	11.59	11.6	11.26	10.44	0.00	4.79	11.49	0.00	4.92
ACG/CGT	4.96	5.07	0.00	1.01	0.00	0.32	0.00	0.00	0.28
ACT/AGT	1.32	1.31	0.00	0.00	0.00	1.22	2.48	0.00	0.64
AGC/GCT	14.56	14.73	7.04	5.05	0.00	1.72	2.48	0.00	1.58
AGG/CCT	17.01	17.17	9.85	11.45	0.00	7.59	4.04	0.00	4.06
AGT/ACT	10.77	10.76	11.26	7.74	0.00	5.19	8.07	0.00	5.43
CCG/CGG	6.12	6.22	0.00	1.35	0.00	0.18	0.00	0.00	0.19
AAAT/ATTT	32.06	29.2	160.48	33.00	112.50	83.04	74.53	225.00	182.15
AATT/AATT	13.67	12.38	71.79	20.54	87.50	31.9	22.98	125.00	70.62
AAAC/GTTT	7.13	7.13	7.04	3.03	0.00	8.86	5.90	0.00	8.94
		*Dat	a refer to counts/Mbp; O	nly the predomina	ant tetranucleotide repea	at motifs were	shown.		

As shown in Table 4, AAAT/ATTT was the predominant of tetranucleotide SSR, and AAAAT/ATTTT, AAAAG/CTTTT motifs were most abundant among the few pentanucleotide SSRs.

Frequencies of dinucleotide and trinucleotide SSR motifs in each species are listed in Table 5. The distribution of dinucleotide repeat types was basically consistent among species. AG/CT motifs were predominant and CG/GC motifs were scarce in all species except for G. pallida (11.89/Mbp). The distribution tendency of trinucleotide repeats in the species within a genus was basically consistent. AAC/GTT motifs were more abundant than other SSR motifs in species of Meloidogyne, except that AAT/ATT motifs were more predominant in M. chitwoodi (117.52/Mbp). ACG/CGT and CCG/CGG motifs were scanty in most species of Meloidogyne. In Globodera, AAC/GTT were the most abundant motifs, and ACT/AGT motifs were rare. CCG/CGG motifs, however, were abundant in Globodera. AGC/GCT and AGG/CCT motifs were predominant in Heterodera. AGG/ CCT and AAC/GTT were predominant motifs in the two species of Pratylenchus. In R. similis, AGC/GCT motifs were abundant. While no ACT/AGT motifs were found in the above three genera. ACC/GGT motifs were predominant motif in X. index. The distribution tendency of SSR motifs was very similar in the two species of Bursaphelenchus, with AAG/CTT motifs predominant and ACG/CGT absent.

SSRs distributed in coding and non-coding ESTs

By the use of the GETORF program in the EMBOSS package, NR ESTs were classified into 'coding' and 'noncoding' ESTs according to the minimum length of 50 amino acids of the translated sequences encoded in all 6 ORFs. A total of 56,468 coding ESTs (92.12 % of total) and 4,828 non-coding ESTs (7.88 % of total) were identified.

A total of 9,369 SSRs in coding ESTs and 445 SSRs in noncoding ESTs were derived. The abundance of SSRs in noncoding ESTs was much higher than that in coding ESTs (626.42/Mbp vs. 293.79/Mbp). Trinucleotide motifs were predominant in coding ESTs (50.1 % of total), while tetranucleotide motifs were more abundant than trinucleotide in non-coding ESTs (47.4% vs. 22.7 %). Dinucleotide motifs were abundant in non-coding ESTs (Table 6). Of these, AG/CT, AAAT/ATTT and AATT/AATT motifs were greatly increased. The frequency of trinucleotide motifs changed little in non-coding ESTs. The frequencies of SSRs in coding ESTs were similar to those in NR ESTs (Table 7).

A comparison of EST-SSRs and genomic SSRs

Of the PPN species used in this study, whole genome data for *M. incognita* and *M. hapla* are available now. We made a survey of SSRs in the two nematodes genomes and compared them with the SSRs detected in ESTs. A total of 27,294 SSRs were detected in the *M. incognita* genome (1/3.01kb) and 24,739 SSRs were detected in the *M. hapla* genome (1/2.14kb). The SSR types are listed in Table 6. Unlike the SSRs in NR ESTs, tetranucleotide SSRs were more extensively distributed in both the PPN genomes (45.2 % in *M. incognita* and 60.5 % in *M. hapla*). There were followed by tri-, bi- and pentanucleotide SSRs in that order. The distribution frequency of motifs is shown in Table 7. In both genomes, AT-rich SSRs (AT/AT, AAT/ATT, AAAT/ATTT and AATT/AATT) prevailed, while CG-rich SSRs (CG/GC, ACG/CGT and CCG/CGG) were rare. AG/CT and AAC/GTT were also abundant SSRs in the two genomes.

Comparing the distribution of SSRs in genomes with those in NR ESTs of the two species, it was shown that there are many differences between them (Table 6 and 7). This is because the EST data are from only small parts of the genomes, and many SSRs exist in the non-coding sequences.

Discussion

Similarity and diversity among the EST-SSRs in PPN

Based on our data, many and varied EST-SSRs occur in PPN. Although a bias maybe exist owing to differences in numbers of ESTs among species and genera, we still can discern some generalities. The similarities and differences in type and distribution frequencies of EST-SSRs motifs among PPN genera and species were correlated with the known relationships among the taxa. For closely related taxa, more similarities occurred and fewer differences. For example, differences among genera were more evident than differences among species within a genus. The genera *Heterodera* and *Globodera* are both cyst nematodes and fewer differences between them were found. For sister species *B. mucronatus* and *B. xylophilus*, the frequencies and types of EST-SSRs were very similar, despite the fact that *B. xylophilus* had more ESTs (Table 1).

Hancock (1996) reported that the entire genome would accept simple sequences repeats in a concerted manner when the genome size increased. In our study, this was not exactly the case. The *M. incognita* genome (~82.1Mb) is larger than the *M. hapla* genome (~53.0Mb) and there were more SSRs in the former than in the latter (27, 294 vs. 24,739). However, the frequency of SSRs in the former was obviously lower than in the latter (332.48 vs 466.63/Mbp) (Table 6). Table 8 lists the published genome sizes of PPN and the abundance of EST-SSR. Based on the present data it is difficult to verify whether the diversity of EST-SSRs in PPN was related to genome size or to characteristics related to parasitism of the PPN species. This could be clarified in future studies that included more genomic sequences of PPN.

Comparison of SSRs distributed in different data sets

ESTs are mainly composed of coding sequences, while the largest part of genome is composed of non-coding regions. Because the selection pressures acting on coding and non-coding sequences are different, it has been suggested that the diversity of SSRs in ESTs might be different from that in other genome regions (Chakraborty *et al.*, 1997; Hancock, 1995). Our results supported this suggestion. We compared SSRs derived from coding ESTs, non-coding ESTs and whole genomes. The distribution frequencies of SSR motifs are very different. Non-coding ESTs (UTRs,

snus gy <i>ne</i>	Species M. incognita M. chitwoodi	Genome size (10 ⁹ bp) ^[48] ~0.08 ^[31]	SSR distribution rate (/Mbp) 332.48 488.15	Main reproductive mode ^[1] parthenogenesis (gamogenesis during environmental stress)	Host range wide	Resistant nematode stages no	Infected part root	Sexual dimorphism obvious	Parasitic mode ^[2] sedentary endoparasitic
	M. javanıca M. hapla M. paranaensis M. arenaria	0.05 ~0.05 ^[37] 	409.32 466.63 315.54 308.96						
	G. pallida G. rostochiensis		514.36 346.71	gamogenesis	narrow	yes	root	obvious	sedentary endoparasitic
	H. glycines H. schachtii	0.06 0.06	297.49 283.09						
S	P. penetrans P. vulnus R. similis		232.74 407.59 198.71	gamogenesis (with parthenogenetic capability)	wide	оп	root	not obvious	migratory endoparasitic
schus	B. mucronatus B. xylophilus	0.03	104.63 93.30	gamogenesis	pine tree	yes	stem	not obvious	migratory endoparasitic
	X. index	0.11	127.62	parthenogenesis	narrow	no	root	not obvious	ectoparasitic

f16 ppN 3 4 SCD ¢ Table 8 I ist mainly) are much richer in SSR loci than are coding ESTs and genomes (Table 6). Tetranucleotide SSRs were the most abundant type in non-coding ESTs and genomes, while trinucleotide SSRs were most frequently located in coding ESTs (Table 7). The results are consistent with SSR distribution in coding and non-coding sequences of *C. elegans* and other animals (Toth *et al.*, 2000).

Comparison of EST-SSRs in PPN with C. elegans and other eukaryotes

Our results showed that the average abundance of SSRs in NR ESTs of PPN (1/3.32kb) was much less than that of the coding sequences of *C. elegans* (1/1.38kb) (Li *et al.*, 2004). But if the mononucleotide repeats are included (\geq 10bp), the average abundance of SSRs in NR ESTs of PPN could be 1/1.35 kb (data not shown), a figure more similar to that of the coding sequences of *C. elegans*.

In PPN, as in other studies, AG/CT motifs were the most predominant dinucleotide types and GC/CG repeats were extremely rare. As reported, AG/CT motifs were abundant in C. elegans (Li et al., 2004) and in the exons of mammalia and embryophyta (Hancock, 1995), whereas, GC/CG repeats were rare in many animal genomes (Katti et al., 2001). AAC/GTT were the most abundant motifs in PPN, the same as reported in the studies of other eukaryotic genome sequences (Toth et al., 2000; Katti et al., 2001). Interestingly, ACT/AGT motifs were rare in most PPN genera. This may be attributed to the presence of a stop codon (UGA) in transcription (Toth et al., 2000). For tetranucleotide SSRs, AAAT/ATTT motifs were the prevailing tetranucleotide repeats in coding ESTs, non-coding ESTs and genomes (Table 7). Even though we could not absolutely exclude a false reading of poly-A tails by machine, we accepted it as a fact, because the same finding was reported in other studies, viz. that (AAAN)_n SSRs prevailed in exons of C. elegans, human, fungi, and embryophyta (Toth et al., 2000; Katti et al., 2001; Subramanian et al., 2003).

Future study of EST-SSRs in PPN

Although EST-SSR diversity does not fully represent the SSR diversity of the whole genome, it has been verified that SSR markers from either coding or non-coding sequence could produce very similar estimates of population differentiation (Woodhead et al., 2005). In addition, EST-SSRs could be more conserved among species in the same lineage, allowing cross-species transferability and a lower frequency of null alleles; and they could more likely appear as a single copy in the genome, alleviating the multiple-band problem (Liewlaksaneevanawin et al., 2004). Thus, EST-SSRs markers could be applied widely in the study of PPN population genetics. Although we have not provided the information about regions flanking the microsatellites, the information about SSR motif types and density in each genus and species can be used for SSRs screening and further study. With more EST and genome data, more information about PPN SSRs will be available, including types and frequency, mechanisms for developing SSRs, roles in regulating gene expressions and evolution, etc. We hope in future studies, people can identify SSRs related to genes involved in parasitism and pathogenicity of phytoparasitic nematodes.

Acknowledgments

We are very grateful to Prof. V. Ferris for her generous help in revising this paper. We also thank the anonymous reviewer for providing good suggestions for analysis coding and noncoding ESTs, and thank Dr. Lin Kui, Dr. Zhu Zhen-Dong, Ms. Guo Jie, Ms. Wu Xiao-Mei, Mr. Wang Yun-Sheng, Prof. He Qing, Mr. Zhu Xiao-Peng, Prof. Chen Run-Sheng for their valuable suggestions and help during the process of data analysis. This research is supported by the National Basic Research and Development Plan (2009CB119200), the National Natural Science Foundation of China (No. 30570288), the National Scientific and Technical Supporting Programs (2006BAD08A1 9116) and the HSCC of Beijing Normal University.

References

ABAD, P., GOUZY, J., AURY, J. M., CASTAGNONE-SERENO, P., DANCHIN, E. G. J., DELEURY, E., PERFUS-BARBEOCH, L., ANTHOUARD, V., ARTIGUENAVE, F., BLOK, V. C., CAIL-LAUD, M. C., COUTINHO, P. M., DASILVA, C., DE LUCA, F., DEAU, F., ESQUIBET, M., FLUTRE, T., GOLDSTONE, J. V., HAMAMOUCH, N., HEWEZI, T., JAILLON, O., JUBIN, C., LEONETTI, P., MAGLIANO, M., MAIER, T. R., MARKOV, G. V., MCVEIGH, P., PESOLE, G., POULAIN, J., ROBINSON-RECHAVI, M., SALLET, E., SEGURENS, B., STEINBACH, D., TYTGAT, T., UGARTE, E., VAN GHELDER, C., VERONICO, P., BAUM, T. J., BLAXTER, M., BLEVE-ZACHEO, T., DAVIS, E. L., EWBANK, J. J., FAVERY, B., GRENIER, E., HENRISSAT, B., JONES, J. T., LAUDET, V., MAULE, A. G., QUESNEVILLE, H., ROSSO, M. N., SCHIEX, T., SMANT, G., WEISSENBACH, J.WINCKER, P. (2008): Genome sequence of the metazoan plant-parasitic nematode Meloidogyne incognita. Nat. Biotechnol., 26(8): 909 - 915. DOI: 10.1038/nbt.1482

AGGARWAL, R. K., HENDRE, P. S., VARSHNEY, R. K., BHAT, P. R., KRISHNAKUMAR, V.SINGH, L. (2007): Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. *Theor. Appl. Genet.*, 114(2): 359 – 372. DOI: 10.1007/s00122-006-0440-x

CASTAGNONE-SERENO, P. (2002): Genetic variability of nematodes: a threat to the durability of plant resistance genes? *Euphytica*, 124(2): 193 – 199

CHAKRABORTY, R., KIMMEL, M., STIVERS, D. N., DAVI-SON, L. J., DEKA, R. (1997): Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proc. Natl. Acad. Sci. U. S. A.*, 94(3): 1041 – 1046

CRISCIONE, C. D., ANDERSON, J. D., RABY, K., SUDIMACK, D., SUBEDI, J., RAI, D. R., UPADHAYAY, R. P., JHA, B., WILLIAMS-BLANGERO, S., ANDERSON, T. J. C. (2007): Microsatellite markers for the human nematode parasite *Ascaris lumbricoides*: Development and assessment of utility. J. Parasitol., 93(3): 704 – 708. DOI: 10.1645/GE-1058R.1

DE LUCA, F., REYES, A., VERONICO, P., DI VITO, M., LAM-BERTI, F., DE GIORGI, C. (2002): Characterization of the

(GAAA) microsatellite region in the plant parasitic nematode *Meloidogyne artiellia*. *Gene*, 293(1 – 2): 191 – 198. DOI: 10.1016/S0378-1119(02)00756-4

ELLIS, J. R., BURKE, J. M. (2007): EST-SSRs as a resource for population genetic analyses. *Heredity*, 99: 125 – 132

FISHER, M. C., VINEY, M. E. (1996): Microsatellites of the parasitic nematode *Strongyloides ratti. Mol. Biochem. Parasitol.*, 80(2): 221 – 224. DOI: 10.1016/0166-6851(96)02699-0

GREGORY, T. R. (2007): Animal Genome Size Database., http://www.genomesize.com.

GRILLO, V., JACKSON, F., GILLEARD, J. S. (2006): Characterisation of *Teladorsagia circumcincta* microsatellites and their development as population genetic markers. *Mol. Biochem. Parasitol.*, 148(2): 181 – 189. DOI: 10.1016/j.molbiopara.2006.03.014

HABER, M., SCHUNGEL, M., PUTZ, A., MULLER, S., HASERT, B., SCHULENBURG, H. (2005): Evolutionary history of *Caenorhabditis elegans* inferred from microsatellites: Evidence for spatial and temporal genetic differentiation and the occurrence of outbreeding. *Mol. Biol. Evol.*, 22(1): 160 – 173. DOI: 10.1093/molbev/msh264

HANCOCK, J. M. (1995): The contribution of slippage-like processes to genome evolution. *J. Mol. Evol.*, 41(6): 1038 – 1047. DOI: 10.1007/BF00173185

HANCOCK, J. M. (1996): Simple sequences and the expanding genome. *BioEssays*, 18(5): 421 – 425. DOI: 10.1002/bies.950180512

HE, Y., LI, H. M., BROWN, D. J. F., LAMBERTI, F., MOENS, M. (2003): Isolation and characterisation of microsatellites for *Xiphinema index* using degenerate oligonucleotide primed PCR. *Nematology*, 5: 809 – 819

HUANG, X. Q., MADAN, A. (1999): CAP3: A DNA sequence assembly program. *Genome Res.*, 9(9): 868 – 877. DOI: 10.1101/gr.9.9.868

JOHNSON, P. C. D., WEBSTER, L. M. I., ADAM, A., BUCK-LAND, R., DAWSON, D. A., KELLER, L. F. (2006): Abundant variation in microsatellites of the parasitic nematode *Trichostrongylus tenuis* and linkage to a tandem repeat. *Mol. Biochem. Parasitol.*, 148(2): 210 – 218. DOI: 10.1016/j.molbiopara.2006.04.011

KATTI, M. V., RANJEKAR, P. K., GUPTA, V. S. (2001): Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Mol. Biol. Evol.*, 18(7): 1161 – 1167

KIM, K. S., RATCLIFFE, S. T., FRENCH, B. W., LIU, L., SAPPINGTON, T. W. (2008): Utility of EST-derived SSRs as population genetics markers in a beetle. *J. Hered.*, 99(2): 112 – 124. DOI: 10.1093/jhered/esm104

KOFLER, R., SCHLOTTERER, C., LELLEY, T. (2007): SciRoKo: a new tool for whole genome microsatellite search and investigation. *Bioinformatics*, 23(13): 1683 – 1685. DOI: 10.1093/bioinformatics/btm157

KUNZLER, P., MATSUO, K., SCHAFFNER, W. (1995):

Pathological, Physiological, and Evolutionary Aspects of Short Unstable DNA Repeats in the Human Genome. *Biol. Chem. Hoppe-Seyler*, 376(4): 201 – 211

LAMBERT, K., BEKAL, S. (2002): Introduction to Plant-Parasitic Nematodes. *The Plant Health Instructor*. DOI: 10.1094/PHI-I-2002-1218-01

LI, B., XIA, Q., LU, C., ZHOU, Z., XIANG, Z. (2004): Analysis on frequency and density of microsatellites in coding sequences of several eukaryotic genomes. *Genomics Proteomics Bioinformatics*, 2(1): 24 - 31

LIEWLAKSANEEYANAWIN, C., RITLAND, C. E., EL-KASSA-BY, Y. A., RITLAND, K. (2004): Single-copy, species-transferable microsatellite markers developed from loblolly pine ESTs. *Theor. Appl. Genet.*, 109(2): 361 – 369. DOI: 10.1007/s00122-004-1635-7

MOXON, E. R., WILLS, C. (1999): DNA microsatellites: Agents of evolution? *Scientific American*, 280(1): 94 – 99 OPPERMAN, C. H., BIRD, D. M., WILLIAMSON, V. M., ROK-HSAR, D. S., BURKE, M., COHN, J., CROMER, J., DIENER, S., GAJAN, J., GRAHAM, S., HOUFEK, T. D., LIU, Q., MITROS, T., SCHAFF, J., SCHAFFER, R., SCHOLL, E., SOSINSKI, B. R., THOMAS, V. P., WINDHAM, E. (2008): Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism. *Proc. Natl. Acad. Sci. U. S. A.*, 105(39): 14802 – 14807. DOI: 10.1073/pnas.0805946105

OTSEN, M., PLAS, M. E., LENSTRA, J. A., ROOS, M. H., HOEKSTRA, R. (2000): Microsatellite diversity of isolates of the parasitic nematode *Haemonchus contortus*. *Mol. Biochem. Parasitol.*, 110(1): 69 – 77. DOI: 10.1016/S016 6-6851(00)00257-7

PARK, Y. H., ALABADY, M. S., ULLOA, M., SICKLER, B., WILKINS, T. A., YU, J., STELLY, D. M., KOHEL, R. J., EL-SHIHY, O. M.CANTRELL, R. G. (2005): Genetic mapping of new cotton fiber loci using EST-derived microsatellites in an interspecific recombinant inbred line cotton population. *Mol. Genet. Genomics*, 274(4): 428 – 441. DOI: 10.1007/ s00438-005-0037-0

PERERA, L., RUSSELL, J. R., PROVAN, J., POWELL, W. (2000): Use of microsatellite DNA markers to investigate the level of genetic diversity and population genetic structure of coconut (*Cocos nucifera* L.). *Genome*, 43(1): 15 - 21

PLANTARD, O., PORTE, C. (2003): Isolation and characterization of microsatellite loci in the sugar beet cyst nematode *Heterodera schachtii. Mol. Ecol. Notes*, 3(1): 139 – 141. DOI: 10.1046/j.1471-8286.2003.00383.x

POWELL, W., MACHRAY, G. C., PROVEN, J. (1996): Polymorphism revealed by simple sequence repeats. *Trends Plant Sci.*, 1: 215 – 222. DOI: 10.1016/1360-1385(96)86898-1

RICE, P., LONGDEN, I., BLEASBY, A. (2000): EMBOSS: The European molecular biology open software suite. *Trends Genet.*, 16(6): 276 – 277. DOI: 10.1016/S0168-95 25(00)02024-2

SAINT-LAURENT, R., LEGAULT, M., BERNATCHEZ, L. (2003): Divergent selection maintains adaptive differentiation despite high gene flow between sympatric rainbow smelt ecotypes (*Osmerus mordax* Mitchill). *Mol. Ecol.*, 12(2): 315 – 330. DOI: 10.1046/j.1365-294X.2003.01735.x

SCHWARTZ, M. K., MILLS, L. S., ORTEGA, Y., RUGGIERO,

L. F., ALLENDORF, F. W. (2003): Landscape location affects genetic variation of Canada lynx (*Lynx canadensis*). *Mol. Ecol.*, 12(7): 1807 – 1816. DOI: 10.1046/j.1365-294X.2003.01878.x

SIM, S. C., YU, J. K., JO, Y. K., SORRELLS, M. E., JUNG, G. (2009): Transferability of cereal EST-SSR markers to ryegrass. *Genome*, 52(5): 431 – 437. DOI: 10.1139/G09-019

SLATE, J., HALE, M. C., BIRKHEAD, T. R. (2007): Simple sequence repeats in zebra finch (*Taeniopygia guttata*) expressed sequence tags: a new resource for evolutionary genetic studies of passerines. *BMC Genomics*, 8: 52. DOI: 10.1186/1471-2164-8-52

STORZ, J. F. (2002): Contrasting patterns of divergence in quantitative traits and neutral DNA markers: analysis of clinal variation. *Mol. Ecol.*, 11(12): 2537-2551. DOI: 10.10 46/j.1365-294X.2002.01636.x

SUBRAMANIAN, S., MISHRA, R. K., SINGH, L. (2003): Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. *Genome Biol.*, 4(2): R13. DOI: 10.1186/gb-2003-4-2-r13

TAUTZ, D., RENZ, M. (1984): Simple sequence repeats are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res.*, 12: 4127 – 4138

TEHRANI, M. S., MARDI, M., SAHEBI, J., CATALAN, P., DIAZ-PEREZ, A. (2009). Genetic diversity and structure among Iranian tall fescue populations based on genomic-SSR and EST-SSR marker analysis. *Plant Syst. Evol.*, 282(1–2): 57–70. DOI: 10.1007/s00606-009-0207-3.

THIERY, M., MUGNIERY, D. (2000): Microsatellite loci in the phytoparasitic nematode *Globodera*. *Genome*, 43(1): 160 – 165. DOI: 10.1139/gen-43-1-160

TOTH, G., GASPARI, Z., JURKA, J. (2000): Microsatellites in different eukaryotic genomes: Survey and analysis. *Ge*-

nome Res., 10(7): 967 – 981

VIARD, F., EL-KASSABY, Y. A.RITLAND, K. (2001): Diversity and genetic structure in populations of *Pseudotsuga menziesii* (Pinaceae) at chloroplast microsatellite loci. *Genome*, 44(3): 336 – 344. DOI: 10.1139/gen-44-3-336

WANG, H. Y., WEI, Y. M., YAN, Z. H., ZHENG, Y. L. (2007): EST-SSR DNA polymorphism in durum wheat (*Triticum durum* L.) collections. *J. Appl. Genet.*, 48(1): 35 – 42

WANG, M. L., MOSJIDIS, J. A., MORRIS, J. B., DEAN, R. E., JENKINS, T. M.PEDERSON, G. A. (2006): Genetic diversity of *Crotalaria germplasm* assessed through phylogenetic analysis of EST-SSR markers. *Genome*, 49(6): 707 – 715. DOI: 10.1139/G06-027

WELLS, R. D., WARREN, S. T. (1998): Genetic instabilities and neurological diseases. Elsevier/Academic Press, San Diego.

WILLIAMS, B. L., BRAWN, J. D., PAIGE, K. N. (2003): Landscape scale genetic effects of habitat fragmentation on a high gene flow species: *Speyeria idalia* (Nymphalidae). *Mol. Ecol.*, 12(1): 11 – 20. DOI: 10.1046/j.1365-294X.20 03.01700.x

WOODHEAD, M., RUSSELL, J., SQUIRRELL, J., HOLLINGS-WORTH, P. M., MACKENZIE, K., GIBBY, M., POWELL, W. (2005): Comparative analysis of population genetic structure in *Athyrium distentifolium* (Pteridophyta) using AFLPs and SSRs from anonymous and transcribed gene regions. *Mol. Ecol.*, 14(6): 1681 – 1695. DOI: 10.1111/j.13 65-294X.2005.02543.x

ZHOU, Z., SAKAUE, D., WU, B., HOGETSU, T. (2007): Genetic structure of populations of the pinewood nematode *Bursaphelenchus xylophilus*, the pathogen of pine wilt disease, between and within pine forests. *Nematology*, 97(3): 304 - 310

RECEIVED NOVEMBER 28, 2008

ACCEPTED OCTOBER 20, 2009