Genetic Differentiation Among Populations of Octopus minor Based on Simple Sequence Repeats Mined from Transcriptome Data

A c e *in* Journal of Ocean University of China · September 2022

DOI: 10.1007/s11802-022-4944-z

J. Ocean Univ. China (Oceanic and Coastal Sea Research)
https://doi.org/10.1007/s11802-022-4944-z
ISSN 1672-5182, 2022 21 (5): 1265-1272

Genetic Differentiation Among Populations of *Octopus minor* Based on Simple Sequence Repeats Mined from Transcriptome Data

NAN Ze[1), 2)], XU Ran[1), 2)], HOU Chunqiang[3)], and ZHENG Xiaodong[1), 2), *]

1) , , 266003,
2) , , 266003,
3) , 300457,

Abstract (Sasaki 1920) is an important commercial cephalopod species in China. This species has declined sharply in recent years. Hence, genetic studies of are imperative to exploit and manage the wild resource. In this study, 46192 microsatellite loci were discovered in 70174 unigenes from the transcriptomic data. Among all of the simple sequence repeat (SSR) unit categories, di-nucleotide and tri-nucleotide SSRs accounted for 45.26% and 14.49%, respectively. A total of 108 SSRs were tested, of which 21 were neutral and polymorphic. Seven SSRs were selected for genetics analyses of the populations in the Bohai Sea, the Yellow Sea, and the southwest coast of the Taiwan Strait region. Significant pairwise $_{st}$ values were detected among the samples. The UPGMA tree based on genetic distances suggested that the sampling locations could be arranged in three clusters. These markers are evidence that the populations in this region may be structured, with samples from Penghu differing remarkably from those in northern China. The present study characterized genetic markers for population assessment, management, and conservation of .

Key words ; transcriptome; simple sequence repeats; genetic divergence

1 Introduction

(Sasaki 1920) is widely distributed along the coastal waters of China, the Korean Peninsula, and the Japanese archipelago (Yamamoto, 1942; Kim ., 2008). It is an important commercial fishery in China, Korea, and Japan. Many studies of in physicology (Seol ., 2007; Chen ., 2019), aquaculture (Song ., 2019), and genetics (Wang ., 2017) have been reported. However, overfishing has led to a sharp decline in the abundance of this species in recent years. Thus, it is crucial to study the genetic diversity and population structure for rational utilization of .

Simple sequence repeats (SSRs), or microsatellites are simple tandem repeat DNA sequences consisting of 1−6 bases that are widely distributed in the genome. Because of high polymorphism, neutrality, and codominance, SSRs have become an effective tool to investigate population genetics. Microsatellite markers in have been developed using the magnetic bead enrichment method (Zuo ., 2011), and population genetics analyses have been carried out using the markers developed with this method (Kang ., 2012; Gao ., 2016). However, no study has characterized microsatellites based on transcriptome data. A large number of SSRs have been identified by next-generation sequencing, providing numerous molecular markers to assess population diversity and genetic structure, which will contribute to the conservation of this species.

In the last two decades, the development of molecular tools has contributed greatly to genetic studies (Moreira ., 2011; Xu ., 2018) and the identification of cryptic species (Allcock ., 2015; Barco ., 2016; Tang ., 2020). Many studies have reported the genetic diversity and structure of using morphological diagnostic (Gao ., 2019) and molecular methods, such as mitochondrial DNA (Sun ., 2010; Xu ., 2018) and microsatellite markers (Kang ., 2012; Gao ., 2016). These studies have provided valuable information about the population genetic diversity and structure of in Chinese waters; however, little is known about the genetic divergence between the Chinese populations (particularly the South China Sea population) and other East Asian populations. The Korean Peninsula is bordered by China, so comparing the population genetics between Korean and Chinese populations would advance our

* Corresponding author. Tel: 0086-532-82032873
E-mail: xdzheng@ouc.edu.cn

understanding of the genetic diversity and structure of in East Asia. In this study, we investigated the population genetics of six geographic populations, including the Bohai Sea, the Yellow Sea, the South China Sea, and the Korean Peninsula, using seven polymorphic SSR markers. We also evaluated whether geographical distance affected the genetic structure of at different locations. This study provides a theoretical basis for the sustainable exploitation and utilization of this valuable fishery resource.

2 Materials and Methods

2.1 Sample Collection and Preparation

Samples were collected from six locations, including the Bohai Sea: Tianjin (TJ), Qinhuangdao (QHD); the Yellow Sea: Dandong (DD); the southwest coast of the Korean Peninsula: Mokpo (MP), Kunsan (KS); and the southwest coast of the Taiwan Strait: Penghu (PH). The detailed sample information is shown in Fig.1 and Table 1. Genomic DNA was isolated from mantle muscle using an improved cetyl trimethyl ammonium bromide method (Winnepenninckx ., 1993).

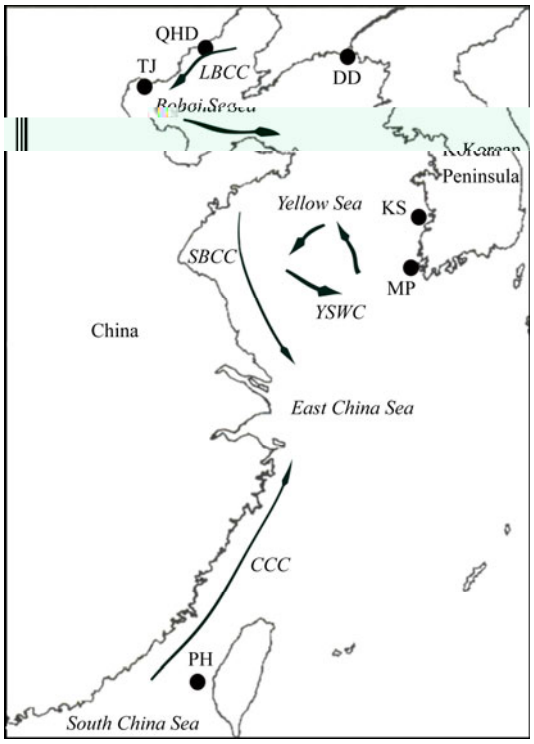

Fig.1 Map of the sampling locations and ocean currents. LBCC, Lubei Coastal Current; SBCC, Subei Coastal Current; CCC, China Coastal Current; YSWC, Yellow Sea Warm Current (Kaneko ., 2011).

2.2 SSR-Enriched Sequences

The transcriptome library was constructed, and transcriptome sequencing was performed by Gene Denovo Biotechnology Co. (Guangzhou, China) using the Illumina HiSeqTM 4000 (Xu and Zheng, 2020). MISA v2.1 (http://pgrc.ipk-gatersleben.de/misa/) was used to search all single gene clusters and identify the localization and type of microsatellites in the transcriptome following default parameters. Primers were designed using Primer3 (version 1.1.4).

Table 1 Details of the six sampling locations

Location	Abbreviation	Location	Collecting date	Sample
Tianjing	TJ	39°05′N, 117°97′E	2018.12.15	31
Qinghuangdao	QHD	39°56′N, 119°63′E	2018.10.31	28
Dandong	DD	39°85′N, 124°31′E	2018.11.15	32
Mokpo (Korea)	MP	34°48′N, 126°15′E	2018.02.02	28
Kunsan (Korea)	KS	35°99′N, 126°61′E	2019.04.28	20
Penghu	PH	23°58′N, 119°56′E	2019.08.09	14

2.3 Specific Primers for SSR Loci Screening and SSR Genotyping

A total of 108 primers were selected to test polymerphism. The polymerase chain reaction (PCR) was performed in a 10 μL volume, including 1 μL of 1× PCR Buffer (Mg^{2+} plus), 1 μL of 0.2 mmol L^{-1} dNTP mix, 0.22 μL of 1 mmol L^{-1} fluorescent label (NED, VIC, or FAM), 0.1 μL of the 1 mmol L^{-1} M13 upstream primer (F), 0.22 μL of the 1 mmol L^{-1} downstream primer (R), 0.05 μL of 0.25 U DNA polymerase, 1 μL of the 50 ng DNA template, and 6.41 μL of dH_2O. The DNA was amplified at 94℃ for 3 min, followed by 35 cycles of 94℃ for 30 s, optimal annealing temperature for 1 min, 72℃ for 75 s, and then eight cycles of 94℃ for 30 s, 53℃ for 1 min, 72℃ for 75 s, and 72℃ for 10 min. Microsatellite polymorphisms were screened using the ABI 3730xl DNA Analyzer.

The genotypes of the samples were analyzed at seven polymorphic microsatellite loci: OMS41, OMS44, OMS51, OMS78, OMS80, OMS96, and OMS99 (Table 1).

2.4 Statistical Analysis

Micro-Checker 2.2.3 (Oosterhout ., 2004) was used to inspect the null alleles. GenALEx v.6.5 (Peakall ., 2012) was used to calculate the Hardy–Weinberg equilibrium (-HWE), the number of alleles per locus (a), observed heterozygosity (o), expected heterozygosity (e), the st values, and Nei's genetic distance. Cervus v.3.0.3 (Marshall ., 1998) was used to estimate the polymorphism information content (PIC). The neutrality of the polymorphic loci was analyzed using the Ewens–Watterson test in POPGENE v.1.32. Multilocus analysis of molecular variance (AMOVA) and pairwise st, values were explored with ARLEQUIN v.3.5 (Excoffier and Lischer, 2010). A Mantel test implemented in Genepop (https://genepop.curtin.edu.au/) was performed to test the isolation of distance (IBD) model by correlating geographic dis-

tance to genetic distance ($_{st}$/(1− $_{st}$); Rousset, 1997). An unweighted pair group method with arithmetic mean (UPGMA) tree was constructed based on the genetic distances among the samples from 6 locations using MEGA v.6.0.

3 Results

3.1 Characterization of the Genic Microsatellites

A total of 46192 SSR loci were discovered in 70174 unigenes from the transcriptome, and the frequency of SSR occurrence in the unigenes was 65.82%. The best represented microsatellite categories were di-nucleotide (45.26%) and mono-nucleotide (39.40%), followed by tri-nucleotide SSRs (14.94%), while tetra-nucleotide and penta-nucleotide SSRs comprised < 1%. Among the dinucleotide repeats, the most abundant repeat motif was AT/AT (22.41%), followed by AC/GT (16.49%) (Table 2). Repeating units between 4 and 14 were detected in the SSRs of the transcriptome, accounting for 99.97% of the total number, and the highest proportion of repeated times (including mono-nucleotide and penta-nucleotide) was six (31.89%).

3.2 Specific Primers for SSR Loci Screening

In this study, 21 of the 108 analyzed markers (19.44%) were polymorphic (showed in Table 3). The amplicon sequences were deposited in the GenBank database (accession numbers: KX061842−KX061864). The number of alleles per loci ranged from 3 to 12, with an average of 5.6. Three loci (OMS78, OMS64, and OMS34) deviated significantly from the -HWE after a Bonferroni correction (<0.05). $_{o}$ ranged from 0.267 to 0.941 (mean = 0.614). $_{e}$ ranged from 0.242 to 0.868 (mean = 0.619). The PIC ranged from 0.231 to 0.854 (mean = 0.572). The Ewens-Watterson neutral test showed that 21 microsatellite loci were located within the 95% confidence interval (Obs. F > L95), indicating the neutrality of these polymorphic markers.

Table 2 Frequency of di-, tri-, and tetra-nucleotide repeat motifs in the transcriptome (–, not available)

Repeat type	Repeat motif	Repeat number								Total	Percent (%)
		5	6	7	8	9	10	11	>11		
Di-nucleotide	AT/AT	–	3076	2369	2316	2009	536	48	–	10354	22.41
	AC/GT	–	2809	1869	1354	1040	467	81	1	7621	16.49
	AG/CT	–	1120	712	410	353	249	65	4	2913	6.30
	CG/CG	–	17	–	2	–	–	–	–	19	0.04
Tri-nucleotide	AAT/ATT	1462	857	276	5	–	–	–	–	2600	5.62
	ATC/ATG	695	332	102	4	–	–	–	–	1133	2.45
	AAC/GTT	560	204	50	3	–	–	–	–	817	1.76
	ACC/GGT	404	170	42	3	–	–	–	–	620	1.34
	ACT/AGT	264	190	86	5	–	–	–	–	545	1.18
	AGC/CTG	253	53	27	2	–	–	–	–	335	0.72
	AAG/CTT	133	52	46	2	–	–	–	–	233	0.50
	AGG/CCT	147	56	24	5	–	–	–	–	232	0.50
	Others	125	38	14	1	1	1	–	–	180	0.39
Tetra-nucleotide	ACAT/ATGT	–	127	4	–	–	–	–	1	132	0.28
	AGAT/ATCT	–	53	4	–	–	–	–	–	57	0.12
	Others	–	170	22	1	1	–	–	–	194	0.42
	Total	4043	9324	5647	4113	3405	1253	194	6		

Table 3 Basic genetic information of the 21 microsatellite primers

Locus	GenBank accession no.	Repeat motif	Primer sequences (5'−3')	Ta (℃)	Size (bp)		$_{o}$	$_{e}$	-HWE	PIC	Ewens-Watterson test		
											Obs. F	L95*	U95*
OMS78	MN565062	(AG)10n(GAA)6	F:CACTGACCATCTCCCTTCGT R:AACGCGAAGTTCCTCACACT	55	208–217	4	0.267	0.673	0.000*	0.611	0.3272	0.3083	0.8728
OMS80	MN565063	(TTG)6	F:CCCACCTCTACATTGGATCG R:CACTGCACCGCCTTCTTAAC	55	194–206	5	0.581	0.614	0.451	0.553	0.3855	0.2586	0.7903
OMS44	MN565064	(AC)9	F:CATGCTTGGTGCAGAACAAC R:TGAAGATGAGGCTGTGGAAG	52	240–252	4	0.267	0.242	0.994	0.231	0.7583	0.3061	0.8433
OMS41	MN565065	(GA)9	F:TGGATGCTGCTGTTCTTCAG R:CCTCAGTGCATTCCAGCTTC	52	176–208	8	0.839	0.754	0.989	0.718	0.2461	0.1738	0.5447
OMS61	MN565066	(AT)9n(ACA)6n(TTA)7	F:GGTGGATATCGTCTGCACAGT R:GAACATGTGATTGCCAAACC	52	233–244	6	0.552	0.543	0.992	0.501	0.4572	0.2176	0.6968
OMS64	MN565067	(TAG)6	F:AGACAATGGGTGGTGGAGAG R:CACCAGCTGATTGGCAAAG	52	184–208	8	0.724	0.756	0.000*	0.724	0.2438	0.1724	0.5410
OMS34	MN565068	(CA)9n(TAG)6	F:CAGCAGCTGCAACAACAGAT R:TCCTGCCATCACCTATACCA	52	266–275	5	0.607	0.601	0.000*	0.552	0.3986	0.2526	0.7717
OMS50	MN565069	(CTG)7	F:GCGTGCCACTGGGGTAAC R:TGATTGGTGATCAGTAAAATTCCA	52	195–201	3	0.483	0.496	0.379	0.432	0.5036	0.3585	0.9328

( )

( )

Locus	GenBank accession no.	Repeat motif	Primer sequences (5' – 3')	Ta (℃)	Size (bp)		o	e	-HWE	PIC	Ewens-Watterson test		
											Obs. F	L95*	U95*
OMS15	MN565070	(TC)9	F:TGGACATGATGGTGCTTTTG R:CCATCATCATCGTCATCAGC	58	184–190	5	0.621	0.554	0.658	0.506	0.4459	0.2604	0.7776
OMS28	MN565071	(TA)9n(GAT)6n(ATT)6	F:GCATCAGCTGGAACAGAACC R:CTGTCATTGAGCCAATGGTG	58	196–199	2	0.581	0.494	0.431	0.387	0.5062	0.3678	0.9370
OMS99	MN565072	(ACG)6	F:ACCAACCGTCAGTCGAAGAG R:GGAAGAATGATGCCGTTCAC	55	197–209	5	0.724	0.693	0.903	0.636	0.3068	0.2551	0.7771
OMS32	MN565073	(TAC)6	F:CATGGATGAGCAGATCCTGA R:CCTTGCAGAGATCCATAACCA	52	216–228	5	0.941	0.693	0.490	0.611	0.3351	0.2483	0.7700
OMS45	MN565074	(TGG)6n(ATTT)5	F:CACTCCATTCTTAGCACACAGC R:TGTTGCAACGGTAGCTTGTAG	52	211–221	6	0.429	0.679	0.047	0.641	0.3209	0.2132	0.6644
OMS106	MN565075	(TAC)6	F:CAGCTTTGCCATTGCTACC R:AGTGGTGGTGGAGGCAGTAG	52	187–208	6	0.567	0.599	0.747	0.557	0.4006	0.2206	0.7033
OMS51	MN565076	(TAG)6	F:TCATGATCATTGCCATAACAG R:CTCCATTCACGAGACACACC	52	199–211	5	0.517	0.498	0.555	0.472	0.5024	0.2479	0.7788
OMS11	MN565077	(TTG)6	F:CCTAGGTTAGGGGCCTTGAT R:ATTGTTCCCAGGCTTCCTTC	58	184–205	9	0.724	0.728	0.131	0.687	0.2717	0.1552	0.4893
OMS84	MN565078	(GACG)6	F:CAACGCTCTCGTTGAAGATG R:CCGTTGTTCGTTCTGTTTGA	55	179–207	12	0.935	0.868	0.863	0.854	0.1322	0.1212	0.3476
OMS37	MN565079	(TTC)6	F:GGTCGTCATTACCGAACTGG R:GCAGGTAGAGGAGGTTGTGG	52	194–212	6	0.931	0.721	0.157	0.675	0.2788	0.2182	0.6968
OMS107	MN565080	(TGT)6	F:TGTTGTGATTGGCACCACTT R:CATCAGCAACAGCATCGTCT	55	208–223	6	0.742	0.686	0.445	0.652	0.3137	0.2300	0.7102
OMS77	MN565081	(TAA)7n(AAT)6n(TTA)6	F:GCACACACTTCAGGAACACG R:TATGCAACACAGGGTGGTGT	55	195–204	4	0.286	0.470	0.019	0.426	0.5300	0.3029	0.8642
OMS96	MN565082	(ATG)6n(TGG)6	F:TCCGTCGGACAGAATTATCC R:ACCACCACCACCAATACCAT	55	247–256	4	0.577	0.638	0.068	0.593	0.3617	0.2870	0.8543

Notes: Ta, annealing temperature; o, observed heterozygosity; e, expected heterozygosity; PIC, polymorphism information content. * Significant deviations from -HWE after Bonferroni correction (<0.05; =21).

Table 4 genetic diversity indices

Location		Locus							
		41	44	51	78	80	96	99	Mean
TJ	a	8	4	5	4	5	4	5	5.0
	-HWE	0.718	0.231	0.472	**0.000***	0.451	0.593	0.636	–
	e	0.754	0.242	0.498	0.673	0.581	0.638	0.724	0.587
	o	0.839	0.267	0.517	0.267	0.470	0.577	0.760	0.528
QHD	a	6	3	5	5	6	6	4	5.0
	-HWE	0.061	0.752	0.742	**0.000***	0.306	0.008	0.046	–
	e	0.795	0.310	0.474	0.747	0.478	0.591	0.693	0.584
	o	0.714	0.360	0.517	0.321	0.448	0.483	0.724	0.509
DD	a	7	4	6	4	4	4	3	4.5
	-HWE	0.964	**0.000***	**0.003***	**0.004***	0.156	**0.005***	0.393	–
	e	0.634	0.400	0.427	0.663	0.643	0.617	0.760	0.592
	o	0.700	0.283	0.310	0.316	0.650	0.280	0.640	0.425
KS	a	6	3	5	5	3	3	4	4.1
	-HWE	0.256	0.145	0.926	0.021	0.572	**0.001***	0.831	–
	e	0.709	0.469	0.583	0.684	0.406	0.558	0.611	0.574
	o	0.850	0.294	0.765	0.417	0.500	0.333	0.750	0.558
MP	a	9	4	4	4	4	3	4	4.6
	-HWE	0.278	0.046	0.706	**0.000***	0.888	0.023	0.706	–
	e	0.806	0.477	0.446	0.668	0.516	0.589	0.614	0.588
	o	0.680	0.308	0.560	0.393	0.536	0.821	0.773	0.581
PH	a	6	7	5	5	2	5	5	5.0
	-HWE	0.099	**0.000***	0.103	0.212	0.501	0.694	**0.002***	–
	e	0.393	0.442	0.684	0.447	0.492	0.683	0.260	0.483
	o	0.364	0.426	0.500	0.462	0.375	0.615	0.214	0.392

Notes: Ta, annealing temperature; o, observed heterozygosity; e, expected heterozygosity; R, allelic size range; * Significant deviations from the -HWE after Bonferroni correction (<0.05; =7); –, not available.

3.3 Genetic Diversity and Structure Among the Samples

Table 4 summarizes the genetic diversity indices of 7 microsatellite loci from the six sampling locations. The number of alleles per locus varied from 5 (at OMS99) to 11 (at OMS11). Observed heterozygosity values ranged from 0.241 to 0.850, and expected heterozygosity values ranged from 0.242 to 0.860. Among all loci, the average number of alleles for each location varied from 4.1 to 5.0. The lowest average number of alleles was detected in the DD (4.1), whereas TJ, QHD, and PH showed an average of five alleles per location. The average observed and expected heterozygosity value per location ranged from 0.392 (PH) to 0.581 (MP), and from 0.483 (PH) to 0.592 (DD), respectively. No linkage disequilibrium was detected in the locus pairs, suggesting that the loci can be treated as independent variables. After the Bonferroni correction, ten of 42 locus-location combinations significantly deviated from -HWE (<0.05), among which OMS78 showed a deviation at all four locations.

Table 5 lists the Nei's genetic distance (Dc, above the diagonal) and pairwise $_{st}$ values (below the diagonal) in pairwise comparisons at different locations. Pairwise $_{st}$ values ranged from 0.014 to 0.174, with relatively high values being detected between PH and the other locations (Table 5). Similarly, the genetic distances between PH and the other locations were much larger than the pairwise comparisons between the other locations. These results were further confirmed by a UPGMA dendrogram, which showed three distinct clades: a single clade for PH, one for the Korean samples (MP and KS), and one for the northern Chinese samples (TJ, QHD, and DD) (Fig.2).

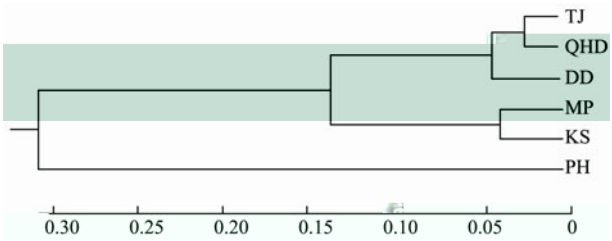

Fig.2 UPGMA tree based on matrices of the pairwise Nei's genetic distances of the microsatellites.

Table 5 Nei's genetic distance (Dc, above the diagonal) and pairwise $_{st}$ values (below the diagonal) among the six locations

	TJ	QHD	DD	KS	MP	PH
TJ	–	0.044	0.079	0.253	0.294	0.629
QHD	0.014	–	0.085	0.232	0.300	0.614
DD	0.026	0.027	–	0.270	0.286	0.693
KS	0.071*	0.064*	0.074*	–	0.073	0.582
MP	0.078*	0.079*	0.078*	0.023	–	0.568
PH	0.157*	0.159*	0.174*	0.158*	0.152*	–

Note: *Significant difference after the Bonferroni correction.

3.4 Analysis of Molecular Variance and IBD

AMOVA indicated that 12% of the total genetic variation occurred among the populations in the six sampling locations; 26% was attributed to variation among individuals and 63% occurred within individuals (Table 6). According to the IBD analysis, the genetic and geographic distances were highly positively correlated, and the genetic distances explained 80% of the total variance (^{2}=0.7906, =0.001), as shown in Fig.3.

Table 6 Analysis of molecular variance (AMOVA) of genetic differentiation in (–, not available)

Source of variation		S.S	M.S	Est. Var.	% of variance
Among pops	5	89.577	17.915	0.296	12
Among individual	148	428.306	2.894	0.650	26
Within individual	154	245.500	1.594	1.594	63
Total	307	763.383	–	2.540	100

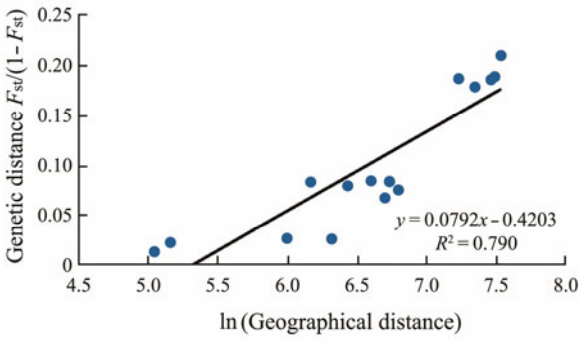

Fig.3 Scatter plot of the genetic and geographical distances for the pairwise location comparisons.

4 Discussion

A total of 46192 SSR loci in was discovered in 70174 unigenes, and the frequency of SSR occurrence in the unigenes was 65.82%, which was much higher than that reported in (48.70%) (Guan ., 2018). Mono-nucleotide and di-nucleotide SSRs accounted for 39.40% and 45.26%, respectively. Similarly, di-nucleotides are highly over-represented in the genome (Wang ., 2018). Repeating units between 4 and 14 occurred in the SSRs, accounting for 99.97% of the total, and the highest number of re-

peats was six (31.89%). The number of markers decreased as the number of repeated units and the number of repetitions increased, which was also observed by Sun (2017) and Shang . (2019). The markers developed in this study were highly polymorphic, as 76% of the PIC was > 0.5. These highly polymorphic markers can be applied to subsequent population genetics studies.

We investigated the genetic structure of from six geographic locations using seven microsatellite DNA markers. The average number of alleles in the different geographic populations ranged from 4.1 to 5.0, which was slightly lower than the number reported by other studies (Kang ., 2012; Gao ., 2016). This difference can be explained in two ways: 1) variations among the different SSR markers used in the studies; and 2) differences in the diversity and abundance of the different sampling locations. Ten of the 42 locus-location combinations significantly deviated from -HWE after a Bonferroni correction (< 0.05). Kang (2012) and Gao (2016) reported that 23 of 56 and 20 of 80 locus-location combinations deviated from HWE, respectively. In addition, similar results have been reported in other cephalopods (Cabranes ., 2008; Doubleday ., 2009). This deviation may be due to the small population or sample size, inbreeding, or the presence of null alleles, as indicated by other studies (Shaw ., 1999; Perez-Losada ., 2002; Kang ., 2012).

The UPGMA tree based on Nei's genetic distance suggests that samples could be arranged in three clusters. The Korean samples formed a sister clade to the samples from the Bohai Sea and the Yellow Sea, while PH samples from the South China Sea were separated from the other samples. PH had the largest geographical distance to the other locations and the samples from PH presented the highest genetic distance compared to the samples from the other locations. Moreover, the Korean samples (MP and KS) were also different from those of the three northern Chinese samples (TJ, QHD, and DD). Similarly, the $_{st}$ values between PH and the other locations were the highest among all the comparisons. The Korean samples also had a high $_{st}$ value compared to the samples from northern China, but the $_{st}$ and the genetic distance values between samples from the two Korean locations were relatively low. Taken together, our results indicate the divergence in the populations containing the northern Chinese samples (TJ, QHD, and DD), the Korean samples (MP and KS), and the southern Chinese sample (PH). Our results are consistent with the findings in other studies using mitochondrial and morphological markers, which reported large divergences between northern and southern populations (Xu ., 2018; Gao ., 2019). The population divergences between the Korean and northern Chinese samples were also revealed by Kang . (2012).

The IBD analysis in Fig.3 shows a strong correlation between genetic distance and geographical distance, accounting for about 80% of the total divergence. Therefore, our results indicate that these divergence results can be mainly explained by the theory of genetic and geographical distances, which hypothesizes that the effect of gene flow can lead to genetic similarities among locations with small geographical distances; therefore, geographical proximity may reflect the high correlation between genetic distance and geographical distance (Scribner ., 1986). However, other effects, such as ecological habitat and ocean currents, can also contribute to population divergence. For example, Kang . (2012) suggested that divergence of the Korean population may be due to ecological differences in the habitats between the western muddy coast and the southern rocky areas. Many studies have shown that ocean currents play a crucial role in the population genetic differentiation of benthic marine organisms (Doubleday ., 2009; Zhan ., 2009; Ni ., 2011; Gao ., 2016). The Yellow Sea Warm Current flows through the west coast of the Korean Peninsula into the northwestern coast of China during April–August. However, this current is relatively weak compared with the coastal waters of China and rarely reaches the inside of the Gulf of Bohai Sea (Pang and Kim, 1998). The duration of the benthic planktonic larval stage of is relatively short (Zheng ., 2014). We inferred that geographic distance is one of the main factors affecting the dispersal of Although other factors might simultaneously affect genetic differentiation, the IBD model explained about 80% of the variance of the six geographical

Barco, A., Raupach, M. J., Laakmann, S., Neumann, H., and Knebelsberger, T., 2016. Identification of North Sea mollusks with DNA barcoding. , **16** (1): 288-297.

Cabranes, C., Fernandez-Rueda, P., and Martinez, J. L., 2008. Genetic structure of around the Iberian Peninsula and Canary Islands as indicated by microsatellite DNA variation. , **65**: 12-16.

Chen, Z. W., Xu, R., Nan, Z., and Zheng, X. D., 2019. Determination of the median lethal concentration and acute toxicity of under different concentrations of ammonia and nitrogen stress. , **50** (6): 1361-1370 (in Chinese with English abstract).

Doubleday, Z. A., Semmens, J. M., Smolenski, A. J., and Shaw, P. W., 2009. Microsatellite DNA markers and morphometrics reveal a complex population structure in a merobenthic species ( ) in south-east Australia and New Zealand. , **156** (6): 1183-1192.

Excoffier, L., and Lischer, H. E. L., 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. , **10**: 564-567.

Gao, X., Xu, R., Zhang, Z., and Zheng, X., 2019. Morphological variation analysis of in the coastal waters of China. , **43** (7): 1593-1602 (in Chinese with English abstract).

Gao, X., Zheng, X., Bo, Q., and Qi, L., 2016. Population genetics of the common long-armed octopus (Sasaki, 1920) (Cephalopoda: Octopoda) in Chinese waters based on microsatellite analysis. , **66**: 129-136.

Yamamoto, T., 1942. On the ecology of (Sasaki), with special reference to its breeding habits. , **12**: 9-20.

Zhan, A. B., Hu, J. J., Hu, X. L., Zhou, Z. C., Hui, M., Wang, S., ., 2009. Fine-scale population genetic structure of Zhikong scallop ( ): Do local marine currents drive geographical differentiation? , **11** (2): 223-235.

Zheng, X. D., Qian, Y. S., Liu, C., and Li, Q., 2014. . In: Iglesias, P., ., eds., Springer, New York, 415-426.

Zuo, Z. R., Zheng, X. D., Yang, Y., and Li, Q., 2011. Development and characterization of 12 polymorphic microsatellite loci in (Sasaki, 1920). , **3**: 489-491.

(Edited by Qiu Yantao)