BMC Research Notes

RESEARCH NOTE Open Access

Identification and characterization of 17 novel polynucleotide microsatellite markers in *Ocypode stimpsoni* (Decapoda: Ocypodidae)

Ha Yeun Song[1], Young Ji Choi[1], Dea-Sung Lee[2], Hyun Ki Choi[3], Kyung Min Choi[1], Kang-Yeol Yu[1], Namki Cho[4] and Seungki Lee[3*]

Abstract

Objective The ghost crab *Ocypode stimpsoni* (Decapoda) is designated as a protected marine species in Korea due to its declining population. In this study, we successfully identified 17 microsatellite markers for *O. stimpsoni* through next-generation sequencing.

Results Out of the 63 loci examined, 26 were effectively amplified in a sample of 100 individuals. These 17 loci, comprising four tri-nucleotide and 13 tetra-nucleotide repeats, exhibited a range of 4 to 26 alleles per locus (with an average of 14.1 alleles) across 100 samples from three *O. stimpsoni* populations. The mean observed and expected heterozygosities were calculated to be 0.885 and 0.836, respectively. These 17 newly identified polymorphic microsatellite loci hold valuable utility for investigating the genetic structure and diversity of this protected marine species.

Keywords *Ocypode stimpsoni*, Microsatellite loci, Genetic variability, Conservation

Introduction

The ghost crab, *Ocypode stimpsoni* Ortmann, 1897 (Ocypodidae), is a common resident of sandy beaches found along the coastlines of Korea, Japan, Taiwan and China [1–4]. Ghost crabs belonging to the *Ocypode* genus have gained widespread recognition as bio-indicator species, particularly with regard to assessing human impacts on exposed sandy beaches. They serve a significant ecological role by aiding in the cleaning of these environments through sediment ingestion in water [5–8]. Despite their utility as indicator species, research efforts on ghost crabs have been relatively limited [9, 10]. Notably, *O. stimpsoni*, which resides in Korea, has been granted protected status as a marine species and protected by a law, a designation initiated by the Ministry of Oceans and Fisheries, Korea, in 2016 in response to a rapid decrease in the population size caused by habitat loss associated with coastal development [11]. Microsatellite markers are one of the best suitable genetic markers to understand the population structure, genome variation, evolutionary process and fingerprinting purposes [12]. Therefore, it becomes imperative to delve into the genetic diversity within and among populations, shedding light on the population

*Correspondence:
Seungki Lee
metany@korea.kr
[1]Honam National Institute of Biological Resources, Mokpo 58762, Republic of Korea
[2]National Marine Biodiversity Institute of Korea, Seochun 33662, Republic of Korea
[3]National Institute of Biological Resources, Inchen 22689, Republic of Korea
[4]College of Pharmacy, Chonnam National University, Gwangju 61186, Republic of Korea

structure of this species to pave the way for effective conservation strategies.

Molecular methods are crucial for estimating genetic diversity among individuals by comparing their genotypes at polymorphic loci [13]. Among the various molecular markers available, microsatellites, or simple sequence repeats (SSRs), are popular and effective tools for assessing genetic diversity and distinguishing populations of rare and endangered species [14–16]. Over the past decade, the rapid advancement of sequencing technology, particularly NGS, has enabled the identification of a substantial number of microsatellite markers [17, 18]. Microsatellites with larger numbers of repeats tend to exhibit higher levels of polymorphism and stability, making them preferable for genetic studies as they produce clearer results compared to shorter repeat sequences [19]. In this study, we identified 17 novel polymorphic microsatellite markers for *O. stimpsoni* using the next-generation sequencing (NGS) platform Illumina MiSeq. These species-specific microsatellite markers can provide valuable information for genetic studies of this protected marine species in Korea.

Materials and methods

Sample collection and next-generation sequencing

A total of 100 *O. stimpsoni* samples were collected in April 2019 from three beaches on the Western coast of Korea: Yeonggwang ($n=40$), Taean ($n=20$), and Seocheon ($n=40$). To isolate microsatellites, high-molecular-weight DNA (≥2 µg) was extracted from the musculature tissue of each *O. stimpsoni* specimen using the DNeasy Blood & Tissue kit (QIAGEN, Hilden, Germany). Among them, two DNA samples without evident smears (the DNA concentration of 19.79 ng/ul and 19.56 ng/ul, and the total amount 1,979 ug and 1,956 ug, respectively) were used to the NGS. Subsequently, a genome library was prepared according to the manufacturer's protocol (Paired-End Library Construction) using the TruSeq Nano DNA High Throughput Library Prep kit (Illumina, San Diego, CA, USA). These DNA fragments go through the addition of a single 'A' base, and ligation of the TruSeq DNA UD Indexing adapters. The libraries were quantified using qPCR according to the qPCR Quantification Protocol Guide (KAPA Library Quantification kits for Illumina Sequencing platforms) and qualified using the Agilent Technologies 4200 TapeStation D1000 screentape (Agilent technologies).

Microsatellite discovery and primer screening

De novo genome assembly was conducted using the IDBA-UD assembler v1.1.1 [20, 21]. To identify SSRs among the assembled contigs (68,969), the microsatellite identification tool (MISA) software, a Perl SSR motif-scanning tool (http://pgrc.ipk-gatersleben.de/misa/), was employed. Sequences longer than 300 bp were subjected to screening using MSATCOMMANDER 1.08 [22], with a requirement of at least five repeats for di-, tri-, or tetranucleotide repeat motifs. Primers were designed for suitable singleton sequences utilizing the Primer3 software [23]. The primer design criteria comprised a GC content within the range of 45–55%, an optimum annealing temperature of 54–60 °C, a PCR product length of 150–500 bp, and adherence to default parameters for primer self-complementarities and quality criteria.

PCR amplification and genotyping

All 63 primer pairs, derived from suitable sequences, underwent preliminary testing for PCR amplification consistency. The initial screening employed four *O. stimpsoni* individuals collected in Seocheon, Korea. PCR amplification reactions were conducted in 15-µL volumes using a SimpliAmp™ Thermal Cycler (Thermo Fisher Scientific, Waltham, MA, USA). The reaction mix included 0.5 U of Ex Taq DNA polymerase (Takara Biomedical, Inc., Shiga, Japan), 1× PCR buffer, 0.2 mM dNTPs, 10 pmol of each primer, and approximately 30 ng of template DNA. The forward primer from each pair was 5′-end-labeled with 6-FAM, VIC, NED, or PET (Applied Biosystems Inc., Foster City, CA, USA). The PCR program was initiated with 10 min at 95 °C, followed by 35 cycles consisting of 20 s at 95 °C, 40 s at 59 °C, and 1 min at 72 °C, concluding with a 3-min final extension at 72 °C. Amplification success was determined by the presence of a visible band upon visualization on a 2.0% agarose gel. Subsequently, the 26 selected polymorphic loci were assessed in all the collected individuals. To enhance genotyping efficiency, fluorescently labeled PCR products with three or four markers of varying sizes were combined in a single capillary. Capillary electrophoresis on an ABI PRISM 3130 xL Genetic Analyzer (Applied Biosystems) was used for the separation of PCR products, alongside a molecular size marker (GENESCAN 500 LIZ, Applied Biosystems). GENOTYPER version 4.1 (Applied Biosystems, Foster City, CA) was employed for the final determination of allele sizes in the fluorescent DNA fragments.

Genetic analysis

CERVUS version 3.03 [24] was employed to compute various genetic parameters, including the number of alleles (N_A), observed heterozygosity (*Ho*), expected heterozygosity (*He*), and polymorphic information content (*PIC*). To assess deviations from the Hardy–Weinberg equilibrium (HWE) [25] and linkage disequilibrium (LD), which indicate the extent of distortion from the independent segregation of loci, GENEPOP version 4.2 [26] was utilized. Significance levels were adjusted for multiple tests through the sequential Bonferroni correction

[27]. Heterozygote excess was further tested using Genepop (H1: excess of heterozygotes). The total *p* value was obtained from a global test for deviation from Hardy-Weinberg equilibrium (H1: heterozygote excess), with significance levels adjusted for multiple comparisons using the sequential Bonferroni method [28]. Bottleneck events were evaluated using the BOTTLENECK 1.2.02 program [29]. To predict the probability of presence of null alleles and large allele dropout, Microchecker v2.2.3 was used with 95% confidence intervals and 1000 randomizations, while the presence of null alleles was estimated using Oosterhout's estimator [30].

Results and discussion

The raw sequence data obtained from two *O. stimpsoni* samples comprised a total of 1,470,273,091 bp, with a quality score of Q30=91.69%, distributed across 5,706,546 Illumina sequencing reads. The final assembly yielded 68,969 contigs, with a cumulative length of 54,165,912 bp and an average contig length of 785 bp. The N50 scaffold size was 748 bp, and the GC content was 38.7% (Table 1). Among the 238,776 identified SSR sequences, tri-nucleotide repeats were the most common, accounting for 66.08% (157,788) of the total, followed by di- (25.69%), tetra- (7.1%), penta- (0.9%), and hexa-nucleotide repeats (0.1%).

Out of a total of 13,195 unique sequences containing pure/compound microsatellite regions and primer-designable flanking regions, sequences were chosen for primer design based on specific criteria. These criteria included a minimum of seven tri- or tetra-nucleotide repeat motifs, sequences of sufficient length (over 100 bp), and unique sequences flanking the microsatellite array with a minimum of 100 bases. Subsequently, 63 microsatellite loci were selected for further screening, consisting of 32 tri- and 31 tetra-nucleotide repeats. In the initial assessment, 26 loci (8 tri- and 18 tetra-nucleotide repeats) were successfully amplified, while the remaining 37 primers did not amplify the desired PCR products. Among the 26 loci, 7 displayed faint or inconsistent bands, possibly due to nonspecific PCR amplification. Further screening revealed that 17 (58.3%) of these loci were polymorphic, while two loci were monomorphic.

Table 1 Summary of illumina MiSeq sequencing

Description	
Total number of bases	1,470.27 Mb
Average read length	257 nucleotides
Number of reads	5,706,546
Number of contigs	68,969
Total contigs	54,165,912 nucleotides
Average contig read length	785 nucleotides
Max. contig length/Min. contig length	16,572/310 nucleotides
N50	784 nucleotides

Among the 17 polymorphic microsatellite loci identified in this study, four were tri-nucleotide repeats, while the remaining 13 were tetra-nucleotide repeats. Table 2 summarizes the names of the primers, their sequences, repeat motifs, fluorescent labels, and GenBank accession numbers for the 17 new microsatellite loci and Table 3 shows comparison of characteristics of microsatellite loci among three populations. A total of 100 *O. stimpsoni* individuals were screened for variation using the 17 new microsatellites. A BLAST program homology search confirmed that none of the 17 sequences exhibited similarity to any sequences in GenBank.

A total of 240 alleles were identified across the 17 loci, with the N_A per locus ranging from four at OS55 to 26 at OS11, resulting in an average N_A of 14.11 (Table 2). The *Ho* varied from 0.734 at OS46 to 0.99 at OS56, with a mean *Ho* of 0.885. Additionally, the *He* varied from 0.616 at OS55 to 0.936 at OS11, with an average of 0.836 (Table 2). All 17 loci exhibited a high level of *PIC* (>0.5), and most loci featured rare alleles with a frequency of <5%. No genotyping errors or allele dropouts due to stuttering were detected in any of the 17 loci. The likelihood of poor DNA quality influencing the results was minimal, as samples that failed to amplify upon retesting were excluded. Deviations from HWE after the application of sequential Bonferroni correction ($p<0.05/(17-\text{rank}+1)$) were observed in 14 loci (OS10, OS11, OS12, OS13, OS34, OS36, OS41, OS43, OS46, OS47, OS54, OS55, OS59 and OS63) (Table 2), and in comparison among populations, 6 loci (OS10, OS43, OS47, OS54, OS59 and OS63) were found in Yeonggwang ($n=40$), one locus (OS43) in Taean ($n=20$) and 10 loci (OS10, OS11, OS12, OS13, OS36, OS42, OS43, OS46, OS55 and OS59) in Seocheon ($n=40$) (Table 3). Significant ($p<0.05/(17-\text{rank}+1)$) heterozygote excess is found for 4 loci (OS34, OS36, OS54 and OS55) out of 14 loci detected significant deviations from HWE (Table 2) and the remaining 10 loci (OS10, OS11, OS12, OS13, OS41, OS43, OS46, OS47, OS59 and OS63) had *p* values between 0.012 and 1.000, indicating HWE for heterozygote excess (Table 2). In comparison among populations, significant heterozygote excess is found for 4 loci (OS42, OS54, OS56 and OS59) in Yeonggwang, 2 loci (OS41 and OS43) in Taean and 2 loci (OS36 and OS55) in Seocheon, repectively (Table 3). The reasons for the findings are not clear, but we suggest that departures from Hardy–Weinberg proportions can be due to copy number variation, inbreeding or population substructure, migration, natural selection, genotyping error, genetic drift and mutation. Ultimately, deviations from Hardy–Weinberg proportions highlight the biological complexity of natural populations [31–34].

At the 10 loci among 17 loci, the *Ho* was higher than the *He*. The higher *Ho* than *He* at most loci in a population suggests that this population may have recently

Table 2 Characteristics of the 17 microsatellite loci identified for *Ocypode stimpsoni*

Locus (GenBank accession no.)	Primer sequence (5′-3′)	No. indiv PCR success	Repeat motif	Size range (bp)	N_A	Ho	He	PIC	P_{HWE}	*P*	Null present
OS10	F: 6fam-ACTACTGCTACTACTACCGT	78	(TAT)17	188–291	16	0.897	0.899	0.884	0.000*	0.781	no
(MF461429)	R: CCCCTGATAACCTGTCGACG										
OS11	F: hex-GGCGTTATTAGCACTGCTGC	97	(ACT)11	103–148	26	0.825	0.936	0.927	0.000*	1.000	no
(MF461431)	R: TGTTCCTTTCCTTTTTCACTTT										
OS12	F: ned-GGCCAGCACAGGTAGAGAAA	96	(AGT)21	209–281	24	0.896	0.928	0.918	0.000*	0.457	no
(MF461432)	R: AGCAGCATCAGGAACAACAA										
OS13	F: 6fam-AGCAGCAGTAATAGTAGTA GCAGT	100	(CAG)8	209–278	18	0.910	0.850	0.833	0.000*	0.317	no
(MF461433)	R: GTGTTTCCTTCGCTCAGTGC										
OS34	F: pet-TGTCTGTCTGTCTTTTGGTCT	100	(TGTC)11	120–168	14	0.960	0.825	0.802	0.005*	0.000*	no
(MF461434)	R: TGACCCGAGGCTTAAAACCC										
OS36	F: hex-CCCTAACCCCCTCTCTGTCT-	100	(CTAT)8	142–186	11	0.950	0.773	0.736	0.000*	0.000*	no
(MF461435)	R: AACGTGGCAATGCATAACCG										
OS41	F: 6fam-GGAACTCTCTCTCAGCAC AGG-	100	(TATT)7	206–234	8	0.920	0.778	0.743	0.000*	0.012	no
(MF461436)	R: GAAACACCTGTGCAGCAGTG										
OS42	F: 6fam-TTCCTCTCTTCAGCAGCACC	99	(TCCT)7	162–186	7	0.788	0.783	0.746	0.263	0.054	no
(MF461438)	R: TGGGGATGACAAGAGAGCTG										
OS43	F: pet-CTTACGAAGGGGAGAGCGAG	100	(AGAC)8	225–281	14	0.850	0.873	0.855	0.000*	0.594	no
(MF461440)	R: TGTAATCTACCGTGCCCGAG										
OS46	F: pet-GGAAGGCAGGTATGGAGAGC	94	(AGGA)7	270–338	17	0.734	0.890	0.876	0.000*	0.998	no
(MF461442)	R: AATCGAAACCAAGCCCTCGT										
OS47	F: ned-CGGCGGGTGATTGTAGCTA	100	(GTTA)7	269–297	8	0.900	0.770	0.731	0.000*	0.096	no
(MF461443)	R: GAGCTTTGTCAAGAAGCTGCA										
OS54	F: 6fam-TTGCGACTCCAGAAGGTCAC	99	(ATAC)10	207–235	8	0.919	0.789	0.756	0.004*	0.000*	no
(MF461444)	R: GCTCCAAGGGCAGAGGTATT										
OS55	F: hex-TGGTGGGGATTCGAATAGCG	76	(TAAA)7	192–208	4	0.868	0.616	0.533	0.000*	0.000*	no
(MF461445)	R: TGCACCATCCACCCTCATTT										
OS56	F: hex-ACCACCCATTCGTCATGTGT	100	(CATA)8	343–393	14	0.990	0.861	0.84	0.025	0.000*	no
(MF461447)	R: GATGATGGACGGGTCGGTTA										
OS57	F: pet-GGTCAGGACGGTAATGGCAT	100	(GTAT)9	280–328	14	0.890	0.865	0.846	0.019	0.866	no
(MF461449)	R: ACGATGAAAACGGCAAAAGTG										
OS59	F: ned-CTGACCTGCTGGCTGGTAAA	96	(GGCT)8	113–149	19	0.979	0.854	0.833	0.000*	0.042	no
(MF461450)	R: CACCCCAGCTCAAAGACTCA										
OS63	F: ned-CGCAACCTACACAACAGCTG	88	(TAAA)7	228–296	18	0.773	0.935	0.925	0.000*	1.000	yes
(MF461453)	R: GAGTGCTAGGTAGACATGCACA										(0.054)
Avg		95.47			14.11	0.885	0.836	0.810	0.102	0.365	

No indiv PCR success, Number of individuals of PCR success; N_A, number of alleles per locus; *Ho*, observed heterozygosity; *He*, expected heterozygosity; *PIC*, polymorphism information content; P_{HWE}, *P* values of Hardy Weinberg equilibrium test; *P*, the probability for rejecting Hardy-Weinberg equilibrium when H1 = Heterozygote excess using Genepop (Global Hardy-Weinberg tests); Avg, average. * Indicates significant difference after the application of Holm's Sequential Bonferroni's correction ($p < 0.05/(17-\text{rank}+1)$). The presence of null alleles was investigated using MICROCHECKER v2.2.3 with 1000 randomizations with 95% confidence intervals and 1000 randomizations

experienced a genetic bottleneck [35]. Populations that have experienced a recent bottleneck are predicted to show an excess of heterozygosity because allelic loss occurs more faster than loss of genic heterozygosity during a bottleneck [36]. The genetic bottleneck analysis results under the Infinite Alleles Model (IAM) and Two-Phase Mutation Model (TPM) suggested that all populations (Yeonggwang, Taean and Seocheon) recently experienced bottleneck events. We detected a significant heterozygosity excess ($P < 0.05$) for Sign test, standardized differences test, and Wilcoxon's signed rank tests), indicating a contemporary population bottleneck. Heterozygosity excess as an indication of genetic bottleneck was detected under the IAM at all populations (Table 4). Under the TPM (a combination of the IAM 30% and SMM 70%), heterozygosity excess was also detected at all populations, except for the sign test for Yeonggwang (Table 4). Heterozygosity excess was detected only the Standardized differences test for Yeonggwang under the Stepwise Mutation Model (SMM) (Table 4). All three

Table 3 Comparison of characteristics of *Ocypode stimpsoni* microsatellite loci among three populations

Locus	Yeonggwang ($n=40$)							Taean ($n=20$)							Seocheon ($n=40$)						
	No indiv PCR success	N_A	Ho	He	PIC	P_{HWE} (P)	Null present	No indiv PCR success	N_A	Ho	He	PIC	P_{HWE} (P)	Null present	No indiv PCR success	N_A	Ho	He	PIC	P_{HWE} (P)	Null present
OS10	36	15	0.944	0.899	0.876	0.000* (0.973)	no	20	12	0.900	0.909	0.876	0.398 (0.687)	no	22	13	0.818	0.919	0.889	0.000* (0.420)	no
OS11	37	13	0.946	0.902	0.880	0.076 (0.091)	no	20	10	0.950	0.881	0.844	0.172 (0.629)	Yes (0.150)	40	25	0.650	0.961	0.947	0.000* (1.000)	no
OS12	36	17	1.000	0.907	0.885	0.014 (0.035)	no	20	11	0.850	0.890	0.854	0.876 (0.159)	Yes (0.080)	40	20	0.825	0.903	0.885	0.000* (0.933)	no
OS13	40	9	0.975	0.786	0.750	0.263 (0.923)	no	20	8	0.950	0.774	0.729	0.682 (0.005)	no	40	15	0.825	0.892	0.87	0.001* (0.078)	no
OS34	40	10	0.950	0.772	0.739	0.138 (0.023)	no	20	9	1.000	0.832	0.789	0.096 (0.015)	no	40	9	0.950	0.822	0.788	0.624 (0.008)	no
OS36	40	10	0.975	0.797	0.760	0.144 (0.009)	no	20	6	1.000	0.745	0.679	0.049 (0.011)	no	40	7	0.900	0.770	0.723	0.003* (0.000*)	no
OS41	40	8	0.900	0.789	0.751	0.023 (0.216)	no	20	8	0.800	0.785	0.743	0.137 (0.003*)	no	40	4	1.000	0.738	0.677	0.024 (0.192)	no
OS42	40	6	1.000	0.818	0.780	0.035 (0.000*)	no	19	7	0.947	0.817	0.766	0.870 (0.016)	Yes (0.197)	40	6	0.500	0.666	0.604	0.000* (0.956)	no
OS43	40	14	0.875	0.869	0.844	0.000* (0.632)	no	20	9	0.750	0.872	0.833	0.000* (0.000*)	no	40	14	0.875	0.869	0.844	0.000* (0.845)	no
OS46	36	11	0.944	0.838	0.811	0.682 (0.565)	no	20	9	0.800	0.841	0.798	0.225 (0.024)	Yes (0.241)	38	13	0.500	0.869	0.842	0.000* (1.000)	no
OS47	40	6	0.875	0.727	0.636	0.000* (0.886)	no	20	6	0.850	0.776	0.716	0.426 (0.024)	no	40	8	0.950	0.783	0.740	0.045 (0.058)	no
OS54	40	8	1.000	0.798	0.758	0.001* (0.000*)	no	20	8	1.000	0.810	0.767	0.139 (0.008)	no	39	7	0.795	0.752	0.710	0.742 (0.571)	no
OS55	36	4	0.833	0.615	0.526	0.150 (0.028)	no	20	4	0.900	0.638	0.545	0.318 (0.065)	no	20	4	0.900	0.624	0.529	0.000* (0.000*)	no
OS56	40	11	1.000	0.863	0.834	0.444 (0.000*)	no	20	9	1.000	0.865	0.824	0.090 (0.029)	no	40	11	0.975	0.864	0.835	0.832 (0.006)	no
OS57	40	12	0.975	0.853	0.824	0.062 (0.929)	no	20	9	1.000	0.859	0.819	0.388 (0.148)	no	40	12	0.750	0.874	0.848	0.064 (0.750)	no
OS59	36	7	1.000	0.815	0.776	0.000* (0.000*)	no	20	6	1.000	0.735	0.671	0.266 (0.019)	no	40	15	0.950	0.900	0.879	0.000* (0.842)	no
OS63	34	16	0.824	0.937	0.918	0.000* (1.000)	Yes (0.054)	19	8	0.947	0.878	0.837	0.126 (0.601)	no	35	17	0.629	0.915	0.895	0.033 (0.764)	Yes (0.279)
Avg	38.29	10.41	0.942	0.822	0.785	0.184 (0.371)		19.88	8.17	0.920	0.818	0.770	0.328 (0.143)		37.29	11.76	0.811	0.830	0.794	0.337 (0.495)	

No indiv PCR success, Number of individuals of PCR success; N_A, number of alleles per locus; *Ho*, observed heterozygosity; *He*, expected heterozygosity; *PIC*, polymorphism information content; P_{HWE}, *P* values of Hardy Weinberg equilibrium test; *P*, the probability for rejecting Hardy-Weinberg equilibrium when H1 = Heterozygote excess using Genepop (Global Hardy-Weinberg tests; Avg, average. * Indicates significant difference after the application of Holm's Sequential Bonferroni's correction ($p<0.05/(17-\text{rank}+1)$). The presence of null alleles was investigated using MICROCHECKER v2.2.3 with 1000 randomizations with 95% confidence intervals and 1000 randomizations

Table 4 Bottleneck analysis for three populations of *Ocypode Stimpsoni*

Test/Models		Yeonggwang ($n=40$)			Taean ($n=20$)			Seocheon ($n=40$)		
		IAM	TPM	SMM	IAM	TPM	SMM	IAM	TPM	SMM
Sign test (number of loci with heterozygosity excess)	Exp	10.23	10.14	10.05	10.13	10.18	9.96	10.24	10.05	9.99
	Hex	17	13	8	16	15	13	17	15	7
	Probability	**0.00018**	0.11973	0.22092	**0.00187**	**0.01199**	0.10341	**0.00018**	**0.01025**	0.11053
Standardized differences test	T2	3.746	1.812	-1.906	4.125	2.699	0.601	4.199	2.393	-1.087
	Probability	**0.00009**	**0.03500**	**0.02831**	**0.00002**	**0.00348**	0.27404	**0.00001**	**0.00835**	0.13852
Wilcoxon test (One tail for H deficiency)	Probability	1.00000	0.99605	0.12175	1.00000	0.99966	0.92728	1.00000	0.99973	0.17644
(One tail for H excess)	Probability	**0.00000**	**0.00467**	0.88766	**0.00001**	**0.00042**	0.07969	**0.00000**	**0.00034**	0.83553
(Two tail for H excess and deficiency)	Probability	**0.00001**	**0.00934**	0.24350	**0.00002**	**0.00084**	0.15938	**0.00001**	**0.00067**	0.35287
Allele frequency distribution		Normal L-shaped			Normal L-shaped			Normal L-shaped		

Estimation based on 1,000 replications, Exp: Expected number of loci with heterozygosity excess; Hex: Heterozygosity excess; IAM, infinite alleles model; TPM, 2-phase model (a combination of the IAM 30% and SMM 70%); SMM, stepwise mutation model. Significance at $P<0.05$, with significant bottleneck effect indicated in bold

populations exhibited no significant shift in distribution, maintaining a normal L-shape, as shown by the allele frequency distribution test. The strict SMM is the most conservative model for evaluating a significant excess of heterozygosity due to bottlenecks, because if the locus evolves under the strict SMM, there can be situations where this heterozygosity excess is not observed [29, 37]. Taken together all the results suggest that serious demographic bottlenecks have most probably not occurred.

There was no evidence of linkage disequilibrium. Microchecker identified that the large allele dropout was not found and the OS63 locus showed the possibility of null allele presence. The selected 17 markers amplified well with an average of 95.4% of 100 individuals in PCR reaction (Table 2). In comparison among populations, the PCR amplification rates were 95.7% in Yeonggwang ($n=40$), 99.4% in Taean ($n=20$) and 93.2% in Seocheon ($n=40$) (Table 3). These results strongly supported the effectiveness of these novel microsatellite markers with a good quality of detection, high polymorphism, and low frequency of null alleles.

These 17 highly polymorphic microsatellite markers hold significant potential for genetic diversity studies, conservation genetics research, and the implementation of effective management strategies for the protection of the *O. stimpsoni* marine species in Korea.

Limitations

One limitation of this study is the number of published microsatellite markers available for analysis may be insufficient for achieving optimal performance in certain genetic studies.

Acknowledgements

For help in obtaining samples, we thank (in addition to coauthors) the Dr. Min SeopKim, National Marine Biodiversity Institute of Korea.

Author contributions

H.Y. Song wrote the manuscript and analysed the data, Y.J. Choi carried out the experiment and analysed the data, D.-S. Lee planned the experiments, H.K. Choi contributed to sample preparation, K.M. Choi supported this projet and revised the manuscript, K.-Y. Yu revised the manuscript, N. Cho revised the manuscript and S. Lee design of the work and substantively revised the manuscript.

Funding

This work was supported by Korea Environment Industry & Technology Institute (KEITI) through project to make multi-ministerial national biological research resources more advanced program, funded by Korea Ministry of Environment (MOE) (RS-2023-00230404).

Data availability

Sequence data that support the findings of this study have been deposited in the NCBI, and their accession numbers are provided within the manuscript.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 13 December 2023 / Accepted: 18 December 2024
Published online: 30 December 2024

References

1. Lucrezi S, Schlacher TA, Walker S. Monitoring human impacts on sandy shore ecosystems: a test of ghost crabs (*Ocypode* spp.) as biological indicators on an urban beach. Environ Monit Assess. 2009;152:413–24.
2. Karleskint G, Turner RK, Small J. Intertidal communities. Introduction to marine biology. Cengage Learning. 2009;3rd ed:356–411.
3. Hughes RN, Hughes DJ, Smith IP. The ecology of ghost crabs. Oceanogr Mar Biol Annu Rev. 2014;52:201–56.
4. Sakai K, Türkay M. Revision of the genus *Ocypode* with the description of a new genus, Hoplocypode (Crustacea: Decapoda: Brachyura). Mem Queensl Mus Nat. 2013;56(2):665–793.
5. Barros F. Ghost crabs as a tool for rapid assessment of human impacts on exposed sandy beaches. Biol Conserv. 2001;97:399–404.
6. Hobbs CH, Landry CB, Perry JE. Assessing anthropogenic and natural impacts on ghost crabs (*Ocypode quadrata*) at Cape Hatteras National Seashore, North Carolina. J Coast Res. 2008;246:1450–8.
7. Schlacher TA, Lucrezi S, Connolly RM, Peterson CH, Gilby BL, Maslo B, Maslo B, Olds AD, Walker SJ, Leon JX, Huijbers CM, Weston MA, Turra A, Hyndes GA, Holt RA, Schoeman DS. Human threats to sandy beaches: a meta-analysis of ghost crabs illustrates global anthropogenic impacts. Estuar Coast Shelf Sci. 2016;169:56–73.
8. Gül MR, Griffen BD. Impacts of human disturbance on ghost crab burrow morphology and distribution on sandy shores. PLoS ONE. 2018;13:e0209977. https://doi.org/10.1371/journal.pone.0209977
9. Imafuku M, Habu E, Nakajima H. Analysis of waving and sound-production display in the ghost crab, *Ocypode Stimpsoni*. Mar Freshw Behav Physiol. 2001;34:197–211.
10. Tsai JR, Lin HC. A shift to an ion regulatory role by gills of a semiterrestrial crab, *Ocypode Stimpsoni*. Zool Stud. 2012;51:606–18.
11. Conservation and Management of Marine Ecosystems Act. 2022. Available online: https://elaw.klri.re.kr/eng_mobile/viewer.do?hseq=61419&type=part&key=39 (accessed on 6 August 2024).
12. Abdul-Muneer PM. Application of microsatellite markers in conservation genetics and fisheries management: recent advances in population structure analysis and conservation strategies. Genet Res Int. 2014;691759. https://doi.org/10.1155/2014/691759
13. Avise JC. Molecular markers, natural history and evolution. Sunderland, Massachusetts: Sinauer Associates; 2004.
14. Ortego J, Aguirre MP, Noguerales V, Cordero PJ. Consequences of extensive habitat fragmentation in landscape-level patterns of genetic diversity and structure in the Mediterranean Esparto grasshopper. Evol Appl. 2015;8:621–32.
15. Cleary KA, Waits LP, Hohenlohe PA. Development and characterization of fourteen novel microsatellite markers for the chestnut short-tailed fruit bat (*Carollia castanea*), and cross-amplification to related species. PeerJ. 2016;4:e2465.
16. Szczecińska M, Sramko G, Wołosz K, Sawicki J. Genetic diversity and population structure of the rare and endangered plant species Pulsatilla patens (L.) Mill in East Central Europe. PLoS ONE. 2016;11:e0151730. https://doi.org/10.1371/journal.pone.0151730
17. Taheri S, Abdullah TL, Rafii MY, Harikrishna JA, Werbrouck SPO, Teo CH, Sahebi M, Azizi P. De novo assembly of transcriptomes, mining, and development of novel EST-SSR markers in *Curcuma alismatifolia* (Zingiberaceae family) through Illumina sequencing. Sci Rep. 2019;9:3047. https://doi.org/10.1038/s41598-019-39944-2
18. Liu Q, Lu Z, He W, Li F, Chen W, Li C, Chao Z, Tian E. Development and characterization of 16 novel microsatellite markers by transcriptome sequencing for *Angelica Dahurica* and test for cross-species amplification. BMC Plant Biol. 2020;20:152–60.
19. Lindqvist AK, Magnusson PK, Balciuniene J, Wadelius C, Lindholm E, Alarcón-Riquelme ME, Gyllensten UB. Chromosome-specific panels of tri- and tetranucleotide microsatellite markers for multiplex fluorescent detection and automated genotyping: evaluation of their utility in pathology and forensics. Genome Res. 1996;6:1170–6.
20. Peng Y, Leung HC, Yiu SM, Chin FY. IDBA- a practical iterative de Bruijn graph de novo assembler. RECOMB, LNBI. Berger B, editor 6044. 2010:426–440.
21. Peng Y, Leung HC, Yiu SM, Chin FY. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics. 2012;28:1420–8.
22. Faircloth BC. MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. Mol Ecol Resour. 2008;8:92–4.
23. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. Primer3-new capabilities and interfaces. Nucleic Acids Res. 2012;40:e115.
24. Kalinowski ST, Taper ML, Marshall TC. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol. 2007;16:1099–106.
25. Holm S. A simple sequential rejective multiple test procedure. Scand J Stat. 1979;6:65–70.
26. Rousset F. GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol Ecol Resour. 2008;8:103–6.
27. Rice WR. Analyzing tables of statistical tests. Evolution. 1989;43:223–5.
28. Raymond M, Rousset F. GENEPOP (Version 1.2): population genetics software for exact tests and ecumenicism. J Hered. 1995;86:248–9.
29. Piry S, Luikart G, Cornuet JM. BOTTLENECK: a computer program for detecting recent reductions in the effective population size using allele frequency data. J Heredity. 1999;90:502–3.
30. Van Oosterhout C, Hutchinson WF, Wills DP, Shipley P. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes. 2004;4:535–8. https://doi.org/10.1111/j.1471-8286.2004.00684.x
31. Lee S, Kasif S, Weng Z, Cantor CR. Quantitative analysis of single nucleotide polymorphisms within copy number variation. PLoS ONE. 2008;3:e3906. https://doi.org/10.1371/journal.pone.0003906
32. Wang J, Shete S. Testing departure from Hardy-Weinberg proportions. Methods Mol Biol. 2012;850:77–102. https://doi.org/10.1007/978-1-61779-555-8_6
33. Lachance J. Hardy-Weinberg equilibrium and random mating. In R. M. Kliman, editor. Encyclopedia of Evolutionary Biology. 2016:208–211.
34. Graffelman J, Jain D, Weir B. A genome-wide study of hardy–Weinberg equilibrium with next generation sequence data. Hum Genet. 2017;136:727–41.
35. Cornuet JM, Luikart G. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics. 1997;144:2001–14.
36. Luikart G, Cornuet JM. Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. Conserv Biol. 1998;12:228–37.
37. Sharma R, Kumar B, Arora R, Ahlawat S, Mishra AK, Tantia MS. Genetic diversity estimates point to immediate efforts for conserving the endangered Tibetan sheep of India. Meta Gene. 2016;8:14–20.

Publisher's note