RESEARCH

Characterization of 12 Novel Microsatellite Markers of *Sogatella furcifera* (Hemiptera: Delphacidae) Identified From Next-Generation Sequence Data

Hwa Yeun Nam,[1] Brad Coates,[2] Kyung Seok Kim,[2] Marana Park,[1] and Joon-Ho Lee[1,3,4]

[1]Entomology Program, Department of Agricultural Biotechnology, Seoul National University, Seoul 151-921, Republic of Korea
[2]USDA–ARS, Corn Insects & Crop Genetics Research Unit, 113 Genetics Laboratory, Iowa State University, Ames, IA 50011
[3]Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul 151-921, Republic of Korea
[4]Corresponding author, e-mail: jh7lee@snu.ac.kr

Subject Editor: Sara Goodacre

J. Insect Sci. *(2015)* 15(1): 94; DOI: 10.1093/jisesa/iev069

ABSTRACT. The white-backed planthopper, *Sogatella furcifera* (Horváth) (Hemiptera: Delphacidae), is a major pest of rice and has long-range migratory behavior in Asia. Microsatellite markers (simple sequence repeats) have been widely used to determine the origins and genetic diversity of insect pests. We identified novel microsatellite loci for *S. furcifera* samples collected from Laos, Vietnam, and three localities in Bangladesh from next-generation Roche 454 pyrosequencing data. Size polymorphism at 12 microsatellite loci was verified for 40 adult individuals collected from Shinan, South Korea. The average number of alleles per locus was 7.92. The mean values of observed (H_o) and expected heterozygosities (H_E) were 0.615 and 0.757, respectively. These new microsatellite markers will be a resource for future ecological genetic studies of *S. furcifera* samples across more broad geographic regions in Asia and may assist in estimations of genetic differentiation and gene flow among populations for implementation of more effective management strategies to control this serious rice pest.

Key Words: *Sogatella furcifera*, microsatellite marker, genetic diversity

The white-backed planthopper (Horvárth) (Hemiptera: Delphacidae) is a major insect pest of rice throughout South and East Asia (Dyck and Thomas 1979). Prior to 1978, *Sogatella furcifera* caused intermittent damage in cultivated rice (Tang et al. 1996). However, as the adoption of hybrid rice varieties extended into South and East Asia, crop damage caused by *S. furcifera* also increased (Tang et al. 1998, Sogawa et al. 2009). Subsequent outbreaks of *S. furcifera* have been reported in China, Japan, and Korea at the beginning of the 21st century (Zhai and Cheng 2006, Otuka et al. 2007) and has increased concerns over increasing crop damage and reduced yields.

This species has long-range migratory patterns, which include known winter hibernation in the Indochina Peninsula followed by northward migration into Southern and Central China by way of the southwest winds from March to July (National Coordinated Research Group for Whiteback Planthopper 1981). In Korea, *S. furcifera* immigrates by atmospheric currents from mid-June to late-July. This northward movement from overwintering regions coincides with the early stages rice cultivation in nurseries, and the subsequent effects on plant health can be highly devastating (Park 1973, Kisimoto 1976, Asahina and Tsruoka 1986, Uhm et al. 1988). In many cases, however, such as in Korea, the primary overwintering *S. furcifera* source population that migrates into rice production regions remains unclear (Sogawa 1997). Therefore, identification of population genetic structure may provide the migration routes and origin source of *S. furcifera*. The population structure of *S. furcifera* has previously been investigated using variation in the mitochondrial cytochrome c oxidase subunit I (COI) sequence Mun et al. (1999) and inter-simple sequence repeat (ISSR) polymerase chain reaction (PCR) markers (Liu et al. 2010). Mun et al. (1999) described limited variation in a 823-bp fragment of COI among *S. furcifera* from four countries (China, Korea, Philippines, and Vietnam), wherein two haplotypes were identified. The only one T COI haplotype appeared to be fixed among individuals from Malaysia ($n = 4$) might be a result of sampling error (Mun et al. 1999). Analogously, Liu et al. (2010) detected little genetic variation among 47 *S. furcifera* populations using ISSRs. Microsatellite or simple sequence repeat (SSR) markers detect changes in the number of tandem repeats in arrays of one to six nucleotides at a single locus and are one of widely applied molecular tools for population genetic studies (Valdes et al. 1993, Akkaya et al. 1995, Schuler et al. 1996, Kelkar et al. 2010, Park et al. 2013). Microsatellites have been widely applied for the estimation of genetic diversity and gene flow (Kim et al. 2009, 2010), as well as migratory pattern in insects (Zhang et al. 2009, Wei et al. 2013).

In this study, we aimed to use next-generation Roche 454 pyrosequencing technology to develop the microsatellite markers (SSR) which facilitate the rapid identification of tandem repeats among sequences (Csencsics et al. 2010). In past, several studies used enrichment analysis to develop molecular markers (SSR), but there are limitation of this analysis: redundancy, multiple-process representation, and poor specificity (Squirrell et al. 2003). Also, we may not need long sequence read as achieved by Sanger sequencing and short sequence read of this technology provide sufficient opportunities for development of new application, which benefit from the particular fragment of the molecular (Morozova and Marra 2008). Roche 454 pyrosequencing technology demand shorter time and read longer length compare to other technologies, Illumina Genome Analyzer and AB SOLiD (Shendure and Ji 2008). The polymorphism and utility of these 12 microsatellite markers for future estimation of population genetic parameters were assessed by genotyping *S. furcifera* samples collected from Shinan, Korea, in 2012.

Materials and Methods

Sample Collection and DNA Extraction. Individuals of the white-backed planthopper, *S. furcifera*, were collected from nine locations in six countries: Laos, Nepal, Thailand, Vietnam, two sites in Korea, and three sites in Bangladesh (Table 1) by institutions of AFACI (Asian Food and Agriculture Cooperation Initiative) in 2012 and placed in 95% ethanol and stored at −20°C. DNA was extracted from pools of individuals (without abdomen) for each location using the QIAamp DNA Mini Kit (Qiagen, Germany). DNA samples were sent to the USDA-ARS Corn Insect & Crop Genetic Research Unit (CICGRU) in Ames, IA. Agarose gel electrophoresis determined that some DNA

Table 1. Sampling information of *S. furcifera*

Country	Sample site	Sampling date	Sample size	Coordinates
Bangladesh 1	Tarash, Sirajgonj	May 10, 2012	45	24° 23′45.0″ N; 89° 22′39.7″ E
Bangladesh 2	BRRi R/S, Gazipur	Nov. 7, 2012	60	23° 98′60.3″ N; 90° 41′14.6″ E
Bangladesh 3	Sagordi, Barisal	Sept. 27, 2012	63	22° 67′71.6″ N; 90° 36′43.4″ E
Laos	Vientiane, Phontong	Aug. 11, 2012	50	18° 30′39.0″ N; 102° 25′05.7″ E
Vietnam	Nam Dinh, Hai Loc	Sept. 28, 2012	87	20° 10′55.85″ N; 106° 20′07.79″ E

Table 2. Primer sequences used to create pooled plant hopper libraries for sequencing on the Roche 454

Primer name	Primer sequence
PriA_454M04_Eco[a]	CGTATCGCCTCCCTCGCGCCATCAGAGCACTGTAGTTCCCTACACGACGCTCTTCC
PriB_454M11_Pst[b]	CTATGCGCCTTGCCAGCCCGCTCAGTGATACGTCTTCGTAGACTGCGTACATGCAG

Underlined, gray highlighted, and black highlighted regions, respectively, correspond to the key, Midtag (4 or 11), and adapter sequences.

[a]Eco adapter primer sequence: 5′-CTCGTAGACTGCGTACC-3′, 3′-CATCTGACGCATGGTTAA-5′.

[b]Pst adapter primer sequence: 5′-CTCGTAGACTGCGTAC**A**TGCA-3, ‘3′-CATCTGACGCATGT-5′.

Table 3. Setting for multiplexing of 12 trinucleotide microsatellite markers

Multiplex	6-FAM (blue)	HEX (green)	NED (yellow)
Multiplex 1	T11	T13, T16	T5, T9, T17
Multiplex 2	T3, T4, T15	T7, T18	T8

samples had decomposed during shipment to the United States such that high molecular weight DNA library construction could not be completed. Thus, DNA samples collected of five locations in three countries (Laos, Vietnam, and three sites in Bangladesh) were determined to be of high quality and were used for Roche 454 sequencing based on the approach described by Puritz and Toonen (2013).

Roche 454 Library Preparation and Sequencing. Approximately 0.5 µg of total genomic DNA from each location was digested with 5U each of *Eco*RI and *Pst*I enzymes in a 40 µl reaction at 37°C for 4 h and then ligated to adapters *Eco*RI- and *PstI*-Ad (Vuylsteke et al. 1999) as described by Vos et al. (1995). Next digested genomic fragments from each population were PCR amplified separately using 4.0 µl of digested and ligated DNA template, 10 µM of each primer PriA_454M04_Eco and PriB_454M11_Pst (Table 2), and 1.0 U LongAmp high-fidelity DNA polymerase (New England Biolabs, Ipswich, MA) according to manufacturer's instructions. Thermocycler conditions included initial denaturation of 95°C for 2 min, then a six touchdown cycles of 95°C for 30 s, 65°C for 30 s −2°C per cycle, and 65°C for 2 min, then 32 cycles of 95°C for 30 s, 55°C for 20 s, and 65°C for 2 min. Resulting PCR products were checked on 1.5% agarose and then residual single-stranded primer digested from successful reaction products with Exonuclease I and shrimp alkaline phosphatase (New England Biolabs) at 37°C for 1 h, then inactivated at 72°C for 5 min. Reaction products were further purified using silica-based columns with the IBI Gel/PCR DNA Fragment Extraction Kit (Peosta, IA) according to manufacturer's instructions and eluted DNA quantified on an NanoDrop2000 (Thermo Scientific, Wilmington, DE). PCR products were combined in approximate equal ratios into a single pool and diluted to 2.0 ng μl^{-1}. The pool of genomic DNA representative of all geographic locations was sent to the University of Illinois, W.M, Keck Center for Comparative and Functional Genomics (Urbana, IL), High-Throughput Sequencing, from which a single barcoded library was generated. Sequence data were generated from this single library on an ~1/4 of a Roche 454 titanium plate.

Sequence Quality Filtering and Microsatellite Predictions. Three files, .sff, .fna, and .qual, were obtained from the W.M. Keck Center for Comparative and Functional Genomics. The .fna and .qual files were trimmed of quality scores (q) < 20, homopolymer stretches ≥ 6 nt, and minimum length = 100 using the script 454QC.pl from the NGSqcToolkit (Patel and Jain 2012). The resulting filtered reads were used as input for the program SciRoKo (Koefler et al. 2007) that identified mismatched microsatellite di-, tri-, tetra-, and pentanucleotide repeat motifs > 3 units in length, and a base substitution mismatch penalty of 5, seed length of 8, and max mismatches of 1. Sequence ± 300 bp of each predicted motif was obtained using a custom PERL script and output into a .fasta file. Since redundant genomic sequences likely remained among the 454 reads, highly similar sequences were assembled into contigs using CAP3 (Huang and Madan 1999). These assemble contigs and singletons containing trinucleotide repeat microsatellite repeats were extracted from this .fasta file using the script SelectSeqs.pl, and output into a separate .fasta file, and used as input for BatchPrimer3 using the SSR screening and design option (program available at http://probes.pw.usda.gov/batchprimer3/index.html).

Microsatellite Marker Development. To verify the utility of microsatellite markers, we used the *S. furcifera* samples (n = 40) which were collected from Shinan, South Korea in 2012. Genomic DNA was extracted from the whole body of individuals without abdomen using the AccuPrep DNA Extraction Kit (BIONEER). Amplification of microsatellite maker was conducted in total 10 µl reaction volume to genotype the samples: 4.9 µl distilled water, 1.0 µl 10X PCR buffer, 1.0 µl 10 mM dNTP Mixture, 0.5 µl of each primer, 0.1 µl of Taq polymerase, and 2.0 µl template DNA. The PCR was performed under the following conditions: initial denaturation for 4 min at 94°C, followed by 35 cycles of 94°C for 30 s, annealing at 61°C for 30 s, 72°C for 40 s, a final extension was performed at 72°C for 15 min. The forward primer was labeled with fluorescent dye (HEX, 6FAM, or NED dyes) and multiplexed in two groups (Table 3). Amplified PCR mixtures were separated electrophoresed on an ABI Prim 3730 XL DNA Analyzer (Applied Biosystems Inc.) with the GENESCAN-500 (Rox) size standard. The genotype data were analyzed by using GeneMapper version 3.7 (Applied Biosystems Inc.).

GenAlEx 6.501 (Peakall and Smouse 2006) was used to calculate the three measure of genetic diversity: the number of alleles (A) per locus, observed heterozygosity (H_o), and heterozygosity expected under Hardy–Weinberg assumptions for each locus (H_E). The inbreeding coefficient at each locus, F_{IS}, imply a deficiency of heterozygotes compared with that expected under Hardy–Weinberg equilibrium (HWE). HWE was tested for each locus in the population by using GENEPOP version 4.2.1 (Raymond and Rousset 1995). The possible occurrence of null alleles was tested using the program MICROCHECKER (Oosterhout et al. 2004). This program estimates the frequency of null alleles and modifies the allele and genotype

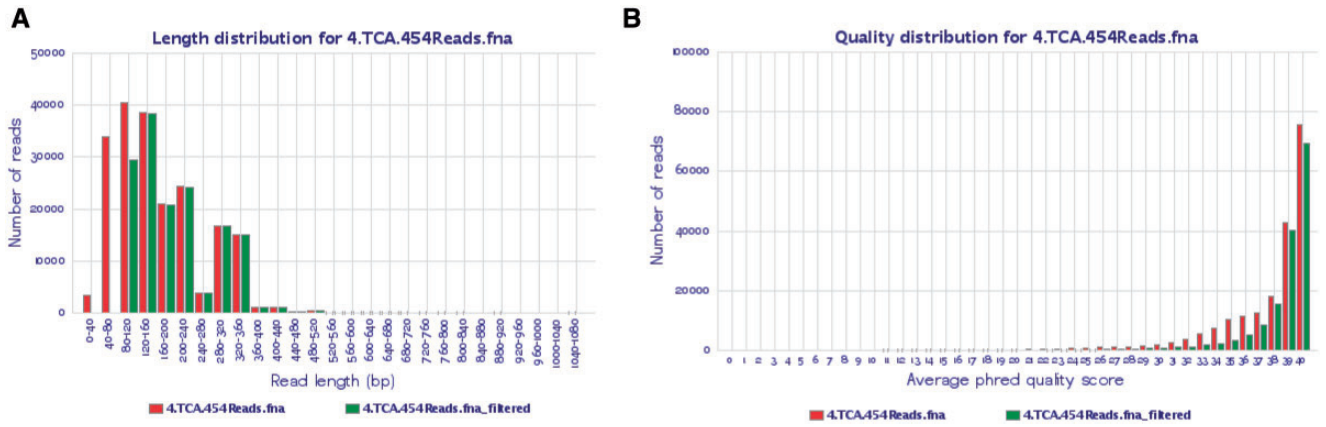

Fig. 1. Roche 454 read length (A) and quality statistics (B) for raw and filtered sequence data.

Table 4. Summary statistics for microsatellite motifs predicted from filtered *S. furcifera* 454 read data

Motif	Counts	Average_length	Average_mismatches	Counts/Mb
AG	57	33.6	1.05	2.51
AC	18	17.3	0.22	0.79
AT	14	15.9	0.00	0.62
AAG	291	24.6	0.31	12.82
AGG	21	25.3	1.05	0.93
AGC	9	31.1	1.00	0.40
ACG	9	17.7	0.00	0.40
AAC	7	23.7	1.29	0.31
ATC	3	20.0	0.33	0.13
AAGG	41	47.7	5.88	1.81
AAAT	12	21.8	0.08	0.53
ACTC	11	16.6	0.00	0.48
ACGT	5	74.0	0.80	0.22
AAGT	3	17.0	0.00	0.13
ACGTC	112	16.0	0.00	4.94
AACGT	4	53.3	4.00	0.18
AGCGC	3	22.7	1.00	0.13

frequencies of the amplified alleles. We used an admixture model and correlated allele frequency among populations to define the ancestry of individuals. The population structure of *S. furcifera* in Shinan was examined by the Bayesian clustering procedure implemented in the STRUCTURE 2.3.3 (Pritchard et al. 2000). The number of Markov chain Monte Carlo replications was set to 200,000 iterations after a burn-in period of 100,000. Likelihood values, Ln Pr(X/K), was identified for each of the 10 STRUCTURE runs at K values from 1 to 10. Also, the mean posterior probability for each K and delta K analysis of LnP(D) was calculated.

Results

PCR-amplified libraries were successfully PCR amplified from pools *S. furcifera* individuals and showed approximate size range of 200–500 bp (results not shown) and were used for Roche 454 sequencing. A total of 32.8 million nucleotides were obtained from Roche 454 sequencing of 199,878 raw reads (N50 = 299 bases; minimum length = 40; maximum length = 1,053), whereas after filtering the read count was reduced to 150,895 containing 29.8 million nucleotides (N50 = 339; minimum length = 100; maximum length = 677; mean q = 38.57; Fig. 1). SciRoKo output identified a total of 636 filtered Roche 454 fragments containing 17 unique repeat motifs, wherein a majority (n = 291) contained $(AAG)_n$ repeats (Table 4, Supp File 1 [online only]). The number of repeat units in each array had a mean of 8.2 ± 4.1 (Supp File 1 [online only]). Because of potential redundancy among independent reads, highly similar sequences were merged into contigs. The CAP3 assembly of microsatellite containing sequences putatively representative of those from the same genetic locus resulting in merger of 506 sequences being merged into 37 contigs (13.7 ± 25.6 sequences and 253 ± 99 bp per contig; Supp File 2 [online only]). Thirty-nine microsatellite containing sequences were unique and were not assembled into contigs (singletons). Redundancy among next-generation sequencing (NGS) reads have been used to predict allelic variation at a microsatellite locus in prior studies (Vukosavljev et al. 2015), clustering assumes that similar sequences are derived from homologous genetic loci and can oftentimes lead to spurious results (see Discussion).Regardless, these fragments in this study containing putative microsatellite motifs ranged in size from 100 to 780 bp (mean = 220.7 ± 80 bp; Supp File 3 [online only]). Filtering of microsatellites further to only those with trinucleotide motifs was conducted using BatchPrimer3 and resulted in oligonucleotide primer pairs being successfully designed for 22 loci with tri-nucleotide repeat motifs (Supp Table 1 [online only]).

By processing PCR amplification, primarily we examined electrophoresis to inspect the size range of the markers. Moreover, we used GeneMapper version 3.7 (Applied Biosystems Inc.) program and verified the allele size and the real peak in the size range of the 21 microsatellite marker. As a result, we selected proper 12 microsatellite markers (Table 5), WBPH_T3, WBPH_T4, WBPH_T5, WBPH_T7, WBPH_T8, WBPH_T9, WBPH_T11, WBPH_T13, WBPH_ T15, WBPH_T16, WBPH_T17, and WBPH_T18.

These 12 microsatellite loci were genotyped for 40 individual specimens and samples were collected from Shinan, Korea, in 2012. All microsatellite markers were polymorphic, with the alleles per locus ranging from 5 to 11 (mean = 7.92). The mean H_o and H_E values were 0.615 (0.225–0.875) and 0.757 (0.558–0.852), respectively (Table 5). The F_{IS} ranged from −0.5582 in WBPH_T4 to 0.7261 in WBPH_T8, with a mean of 0.182 across loci. Most of the loci showed significant deviation from HWE in the direction of heterozygote deficiency, and only one loci WBPH_T18 had no significant deviation. Through the program MICROCHECKER, WBPH_T3 (0.1187), WBPH_T5 (0.1939), WBPH_T8 (0.349), WBPH_T13 (0.1838), WBPH_T15 (0.1769), and WBPH_T16 (0.1345) appeared to harbor possible null alleles; however, further verification including other random mating populations is required since the program MICROCHECKER estimates null alleles for the population assuming the population under HWE. Structure 2.3.3 software was applied to determine the population structure of 40 *S. furcifera* sampled from Shinan in 2012.

In addition, likelihood values and delta K analysis of LnP(D) have been examined (Evanno et al. 2005) (Fig. 2). The highest likelihood values in all runs were determined for K = 1, implying that the *S. furcifera* from Shinan constitute a single genetic cluster. Therefore, the 40 *S. furcifera* specimen collected from Shinan where the occurrence of the species was reported in Korea have genetically similar ancestry with no distinct genetic structuring among them.

Table 5. Twelve proper microsatellite markers for *S. furcifera*

Locus	Primer sequence (5′–3′)	Repeat motif	Mean of *N*	No. alleles	Size range (bp)	H_o	H_E	*D*	F_{IS}
WBPH_T3	F: (6-FAM)- CGACAGCACGTACTCCTGCTT R: ACACGACGCTCTTCCTCCTTC	GAG	40	8	208–250	0.575	0.739	*	0.2337
WBPH_T4	F: (6-FAM)- GGAAGAAACGGATGGAATTACG R: ACGACGCTCTTCCTCCTCATC	AGA	40	5	103–156	0.875	0.558	*	−0.5582
WBPH_T5	F: (NED)- TTCCAATCCTGCTTACAGTCCAA R: GCGTACATGCAGTGGACAGAT	TTC	40	7	200–250	0.400	0.655	*	0.4000
WBPH_T7	F: (HEX)- CCCTCTTCTCTCGCCCTCT R: GTCGTGCTGAGGCTCGTC	GAC	40	8	69–126	0.675	0.788	*	0.1563
WBPH_T8	F: (NED)- TCAGCCAGAGCTGTAGAATCAA R: CAGCGTCTCTGTCCATTCG	AGA	40	10	100–137	0.225	0.804	*	0.7261
WBPH_T9	F: (NED)- GCCGCCCAGTTCTGTAAAGTC R: CTGATGCTGCCGCTGTTGT	GCA	40	9	56–99	0.875	0.852	*	−0.0149
WBPH_T11	F: (6-FAM)- CTAAAACGCTCGCGTCTGC R: GCTCAGTCAGTGATACGTCTTCG	GAT	40	9	295–359	0.625	0.703	*	0.1232
WBPH_T13	F: (HEX)- GCCTCCTCTGCTGTTGAGAAA R: CATTGGCCATCTTGGTGACTG	GAA	40	7	356–386	0.500	0.786	*	0.3604
WBPH_T15	F: (6-FAM)- GCGCGCGCATATATACAGTTG R: AAGCGACGCAAGTGACGATAA	CGT	40	8	157–211	0.500	0.786	*	0.3747
WBPH_T16	F: (HEX)- GGGTACACCGTTCGAGTCGTT R: CCGCTCAGTCAGTGATACGC	CGT	40	10	202–252	0.600	0.833	*	0.2914
WBPH_T17	F: (NED)- TCCTGAGGCACGCTAACTGAC R: CTTGTGCGTGGGTCATGAGAT	ATC	40	8	340–377	0.700	0.807	*	0.1445
WBPH_T18	F: (HEX)- GTGCGAAGGGAAATGCAGAAG R: TTCTCCATCGCATCTCTTGTTCT	GAA	40	6	183–225	0.825	0.774	NS	−0.0528
Across loci				7.92		0.615	0.757	*	0.1820

Microsatellite primer sequences with fluorescent labeled dyes, repeat motifs, mean of individuals (*N*), number of alleles (*A*), size of PCR products in base pairs (bp), expected heteozygosity (H_E), observed heterozygosity (H_o), *D,* deviation from HWE; NS, not significant.

*Significant deviations from Hardy–Weinberg expectations ($P < 0.05$) and inbreeding (F_{IS}) are shown.

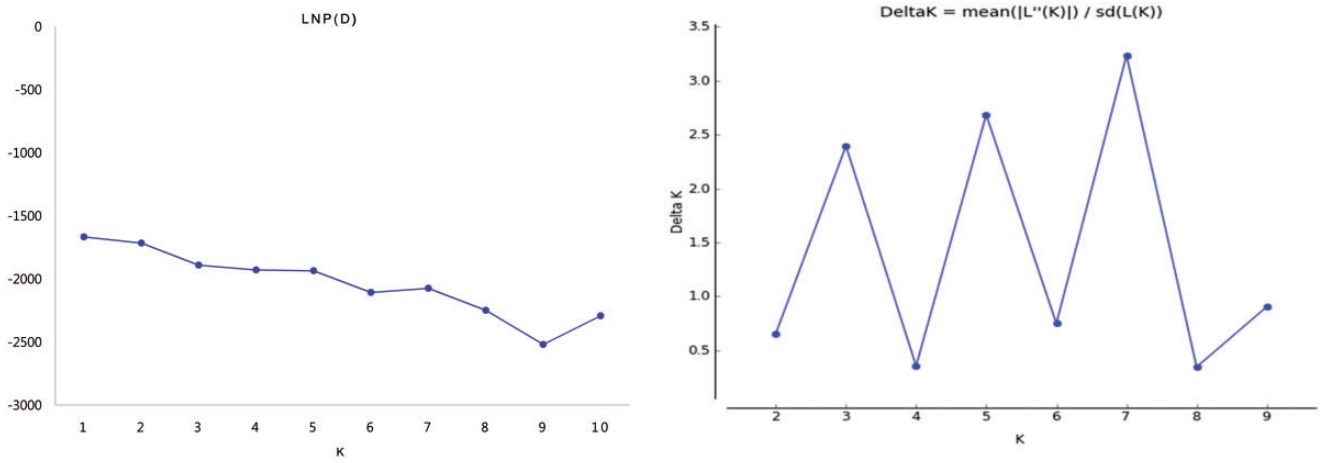

Fig. 2. Plot of mean posterior probability (LnP(D)) values per clusters (*K*), based on 10 iterations per *K*, from STRUCTURE analyses (Pritchard et al. 2000) and delta *K* analysis of LnP(D) (Evanno et al. 2005) to estimate the genetic structure of the 40 *S. furcifera* specimens sampled from Shinan in 2012.

Discussion

The isolation of microsatellite loci from genomic DNA sources had previously been laborious and time-consuming due to methods requiring enrichment, cloning, plasmid purification, and Sanger sequencing (Nunome et al. 2006, Techen et al. 2010), which reduced the number of candidate loci that could be feasibly identified. In contrast, NGS methods have increased the throughput and cost effectiveness of reduced representation library sequencing and has allowed the rapid identification of tandem repeats among sequences in genomic and expressed sequence tag libraries (Csencsics et al. 2010, Agunbiade et al. 2012). In this study, we identified 636 candidate microsatellite repeats from Roche 454 read data. Although the proportion of fragments containing tandem repeats was low (636 of 150,895 filtered reads = 0.42%), output from mass sequencing platforms offer sufficient raw data in which to discover microsatellite sequence elements. Microsatellite repeats within these Roch 454 did not contain GC-rich repeats such as $(CG)_n$ or $(CGG)_n$ arrays, which may be due to our methods of subsampling the *S. furcifera* genome from *Pst*I and *Eco*RI sites. Specifically, microsatellites repeats of a particular nucleotide composition may not be randomly distributed in a genome, such that our methods used for library construction biased the repeats captured within the resulting sequence data. Whole-genome shotgun sequencing approaches may prove to be more appropriate to obtain a more random sampling of loci but was not investigated in this study. These discoveries let to the development of molecular genetic markers for 12 microsatellite loci and have been successfully applied estimates of polymorphism at these loci to obtain population genetics parameters for 40 *S. furcifera* samples collected in Korea. All markers were developed from loci containing trinucleotide repeats, which, although generally more infrequent compared with dinucleotide repeat microsatellite loci (Schug et al. 1998), tend to produce less *Taq* polymerase-induced stutter bands that increases accuracy of downstream genotyping (Edwards et al. 1991). Eleven out of 12 microsatellite loci indicated significant

heterozygote deficiencies in Shinan population, and MICROCHECKER analysis showed the signs of "null alleles" in six microsatellite loci (WBPH_T3, WBPH_T5, WBPH_T8, WBPH_T13, WBPH_T15, and WBPH_T16). "Null alleles" are not uncommon among microsatellite markers and is generally considered to be caused by the failure of PCR to amplify one or more allele in a population due to mutations in the regions of genomic DNA that flank the tandem repeat, such that oligonucleotide primer binding does not occur. Alternatively, the observed heterozygote deficiencies might possibly owing to nonrandom sampling or actual increased frequency of homozygote genotypes in the population. Therefore, additional studies are required to estimate H_o among a broader sampling of populations and will indicate if these six markers can be applied for population genetics studies after analytical methods are used to correct for null alleles (Chapuis and Estoup 2007, Kim et al. 2009).

Although molecular marker-based *S. furcifera* population genetic studies have been previously conducted, the current set of microsatellite markers comparatively offer distinct advantages including codominant alleles, scalability, and increased assay reproducibility. By comparison, maternally inherited mitochondrial haplotypes allow only the estimation of female movements, and direct Sanger sequencing methods such as those used by Mun et al. (1999) are not cost effective for large-scale population genetic studies. Additionally, mitochondrial markers have recently been shown incapable of detecting regional variation among *S. furcifera* populations (Matsumoto et al. 2013). ISSR markers were previously applied for estimation of genetic differentiation among *S. furcifera* populations (Liu et al. 2010). Although ISSR methods offer a relatively easy and inexpensive access to molecular genetic study (Cichorz et al. 2014), yet several researchers doubts about their reproducibility, ascendancy, and homology (Hansen et al. 1998). More recently, 21 microsatellite markers were developed from *S. furcifera* expressed sequence tag resources and were used in preliminary genotyping of samples from China (Sun et al. 2014).

Several complications have been encountered in past efforts to develop molecular microsatellite markers of *S. furcifera*. First, Zhang et al. (2013) observed that tandem repeats between the two *S. furcifera* individuals were identical, and only a few nucleotide variations were revealed, and moreover that the flanking sequences of this repeat region between the two *S. furcifera* individuals were exactly same except for one site difference, which suggests that *S. furcifera* population may have low overall levels of molecular diversity. Similar results were observed in this study from studying sequence data alone. Specifically, the CAP3-assembled contigs showed little copy number variation among microsatellite repeat arrays, with the exceptions of Contig 11 which showed $(TCT)_{9–11}$ repeat, Contig 17 with a $(GAA)_{8–11}$ repeat, and Contig 31 with a $(TCT)_{9–11}$ repeat (Supp File 2 [online only]). All other mutations were within microsatellite flanking sequences and provided empirical evidence that our data could detect variation at 3 of 37 loci for which redundant sequences had been obtained. Although prior studies have used redundancy among NGS reads to predict allelic variation at a microsatellite locus (Vukosavljev et al. 2015), clustering of sequences makes assumptions that all similar sequences are derived from one homologous genetic loci. Quite to the contrary, it has been shown that microsatellite repeats are often located within highly repetitive genome regions wherein flanking sequences tend to be highly similar (Megle'cz et al. 2004, 2007; Zhang 2004; Van't Hof et al. 2007). This phenomenon may be partially influenced by both integration of transposable DNA elements into microsatellites (Coates et al. 2010) and transposition of microsatellite repeats within transposons themselves (Coates 2015; Coates et al. 2011, 2012). This suggests that microsatellite sequences that are highly prevalent among raw nonnormalized sequence reads may indeed be multicopy within a genome. Second, a *S. fucifera* repeat unit had also been previously shown to be within the genome of the small brown planthopper, *Laodelphax striatellus*, which belong to different genera of Delphacidae, and these shared microsatellite sequences were almost identical except for mutations or single nucleotide deletions in a few repeat units (Zhang et al. 2014). Therefore, such applications aimed to estimate allelic variation may be flawed due to specious clustering of nonhomologous sequence data. Indeed, validation for variation at a single locus tends to be accurate using empirical PCR-amplified microsatellite marker data (Agunbiade et al. 2012) and was the tactic used in our work.

These 12 new microsatellite markers will facilitate the study of the gene flow and migration route of *S. furcifera* in Asia. For example, these microsatellites could be suitable to elucidate invasion routes of the insects form China to Korea using the Approximate Bayesian Computation method as was done to manifest frequent and in progress introduction of western corn rootworm from North to Europe (Miller et al. 2005). These markers can be used to the study migration and genetic structure of insect genetic structure data of migration insect. For example, Llewellyn et al. (2003) used microsatellite to study genetic variability of the grain aphid, *Sitobion avenae*, in Britain relate to climate and clonal fluctuation. These markers are also suitable to explain migration routes of insects from Korea and other countries. These markers can characterize the migration patterns, gene flow, and genetic connectivity among geographic populations of *S. furcifera* in Asia. Successful results from this examination will contribute to map effectively strategies for the migration route and origin of this species.

Supplementary Data

Supplementary data are available at *Journal of Insect Science* online.

Acknowledgments

This research was supported by grants from Rural Development Administration in Korea (PJ00922907) and was partially supported by the Brain Korea Plus. Computational, and bioinformatic support for sequence data analysis was provided by the United States Department of Agriculture, 565 Agricultural 566 Research Service (USDA-ARS; CRIS Project 3625-22000-017-00), and the Iowa Agriculture 567 and Home Economics Experiment Station, Ames, IA (Project 3543).

References Cited

Agunbiade, T. A., B. S. Coates, K. S. Kim, D. Forgacs, V. M. Margam, L. L. Murdock, M. N. Ba, L. C. Binso-Dabire, I. Baoua, M. F. Ishiyaku, M. Tamò, and B. R. Pittendrigh. 2012. The spatial genetic differentiation of the legume pod borer, *Maruca vitrata* F. (Lepidoptera: Crambidae) populations in West Africa. Bull. Entomol. Res. 102: 589–599.

Akkaya, M. S., R. C. Shoemaker, J. E. Specht, A. A. Bhagwat, and P. B. Cregan. 1995. Integration of simple sequence repeat and markers into a soybean linkage map. Crop Sci. 35: 1439–1445.

Asahina, S., and T. Tsruoka. 1986. Record of the insects which visited a weather ship located at ocean weather station Tango on the Pacific. Konchu 36: 190–202.

Chapuis, M. P., and A. Estoup. 2007. Microsatellite null alleles and estimation of population differentiation. Mol. Biol. Evol. 3: 621–631.

Cichorz, S., M. Gośka, and A. Litwiniec. 2014. *Miscanthus*: genetic diversity and genotype identification using ISSR and RAPD markers. Mol. Biotechnol. 56: 1–14.

Coates, B. S. 2015. Horizontal transfer of a non-autonomous *Helitron* among insect and viral genomes. BMC Genomics 16: 137.

Coates, B. S., D. V. Sumerford, R. L. Hellmich, and L. C. Lewis. 2010. A *Helitron*-like transposon superfamily from Lepidoptera disrupts $(GAAA)_n$ microsatellites and is responsible for flanking sequence similarity within a microsatellite family. J. Mol. Evol. 70: 278–288.

Coates, B. S., J. A. Kroemer, D. V. Sumerford, and R. L. Hellmich. 2011. A novel class of miniature inverted repeat transposable elements (MITEs) that contain hitchhiking $(GTCY)_n$ microsatellites. Insect Mol. Biol. 20: 15–27.

Coates, B. S., R. L. Hellmich, D. M. Grant, and C. A. Abel. 2012. Mobilizing the genome of Lepidoptera through novel sequence gains and end creation by non-autonomous *Lep*1 *Helitrons*. DNA Res. 19: 11–21.

Csencsics D., B. Sabine, and H. Rolf. 2010. Cost-effective, species-specific microsatellite development for the endangered Dwarf Bulrush (*Typha minima*) using next-generation sequencing technology. J. Hered. 101: 789–793.

Dyck, V. A., and B. Thomas. 1979. Brown planthopper: threat to rice production in Asia, pp. 3–17. International Rice Research Institute. Manila, Philippines.

Edwards, A., A. Civitello, H. A. Hammond, and C. T. Caskey. 1991. DNA typing and genetic mapping with trimeric and tetrameric tandom repeats. Am. J. Hum. Genet. 49: 746–756.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14: 2611–2620.

Hansen, M., C. Halldén, and T. Säll. 1998. Error rates and polymorphism frequencies for three RAPD protocols. Plant Mol. Biol. Rep. 16: 139–139.

Huang, X., and A. Madan. 1999. CAP3: a DNA sequence assembly program. Genome Res. 9: 868–877.

Kelkar, Y. D., N. Strubczewski, S. E. Hile, F. Chiaromonte, K. A. Eckert, and K. D. Makova. 2010. What is a microsatellite: a computational and experimental definition based upon repeat mutational behavior at A/T and GT/AC repeats. Genome Biol. Evol. 2: 620–635.

Kim, K. S., M. J. Bagley, B. S. Coates, R. L. Hellmich, and T. W. Sappington. 2009. Spatial and temporal genetic analyses show high gene flow among European corn Borer (Lepidoptera: Crambidae) populations across the central U.S. Corn Belt. Environ. Entomol. 38: 1312–1323.

Kim, K. S., G. D. Jones, J. K. Westbrook, and T. W. Sappington. 2010. Multidisciplinary fingerprints: forensic reconstruction of an insect reinvation. J. R. Soc Interface 7: 677–686.

Kisimoto, R. 1976. Synoptic weather condition including long-distance immigration of planthopper, *Sogatella furcifera* Horvath and *Nilaparvata lugens* Stål. Ecol. Entomol. 1: 95–109.

Koefler, R., C. Schlotterer, and T. Lelley. 2007. SciRoKo: a now tool for whole genome microsatellite search and investigation. Bioinformatics 23: 1683–1685.

Liu, J. N., F. R. Gui, and Z. Y. Li. 2010. Genetic diversity of the planthopper, *Sogatella furcifera* in the Greater Mekong Subregion detected by inter-simple sequence repeats (ISSR) Markers. J. Insect Sci. 10: 1–14.

Llewellyn, K. S., H. D. Loxdale, R. Harrington, C. P. Brookes, S. J. Clark, and P. Sunnucks. 2003. Migration and genetic structure of the grain aphid (*Sitobion avenae*) in Britain related to climate and fluctuation as revealed using microsatellites. Mol. Ecol. 12: 21–34.

Matsumoto, Y., M. Matsumura, S. Sanada-Morimura, Y. Hirai, Y. Sato, and H. Noda. 2013. Mitochondrial *cox* sequences of *Nilaparvata lugens* and *Sogatella furcifera* (Hemiptera, Delphacidae): low specificity among Asian planthopper populations. Bull. Entomol. Res. 103: 382–392.

Meglécz, E., F. Pétenian, E. Danchin, D', A. C. Acier, J. Y. Rasplus, and E. Faure. 2004. High similarity between flanking regions of different microsatellites detected within each of two species of Lepidoptera: *Parnassius apollo* and *Euphydryas aurinia*. Mol. Ecol. 13: 1693–1700.

Meglécz, E., S. J. Anderson, D. Bourguet, R. Butcher, A. Caldas, A. Cassel-Lundhagen, A. C. d'Acier, D. A. Dawson, N. Faure, C. Fauvelot, et al. 2007. Microsatellite flanking region similarities among different loci within insect species. Insect Mol. Biol. 16: 175–185.

Miller, N., A. Estoup, S. Toepfer, D. Bourguet, L. Lapchin, S. Derridj, K. S. Kim, P. Reynaud, L. Furlan, and T. Guillemaud. 2005. Multiple transatlantic introductions of the western corn rootworm. Science 310: 992.

Morozova, O., and M. A. Marra. 2008. Applications of next-generation sequencing technologies in functional genomics. Genomics 92.5: 255–264.

Mun, J. H., Y. H. Song, K. L. Heong, and G. K. Roderick. 1999. Genetic variation among Asian populations of rice planthoppers, *Nilaparvata lugens* and *Sogatella furcifera* (Hemiptera: Delphacidae): mitochondrial DNA sequences. Bull. Entomol. Res. 89: 245–253.

Nunome, T., S. Negoro, K. Miyatake, H. Yamaguchi, and H. Fukuoka. 2006. A protocol for the construction of microsatellite enriched genomic library. Plant Mol. Biol. Rep. 24: 305–312.

National Coordinated Research Group for Whiteback Planthopper. 1981. Studies on the migration of Whiteback planthopper (*Sogatella furcifera*). Scientia Agriculture Sinica 14: 25–31.

Oosterhout, C. V., W. F. Hutchinson, D. P. M. Wills, and P. Shipley. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol. Ecol. Notes 4: 535–538.

Otuka, A., M. Matsumura, and T. Watanabe. 2007. Recent occurrence of rice planthoppers in East Asian countries. Plant Prot. Sci. 61: 249–253.

Park, J. S. 1973. Studies on the recent occurrence tendency of major insect pest on rice plant, pp. 91–102. Symposium in Plant Environment research. Office of Rural Development, Korea.

Park, M., K. S. Kim, and J. H. Lee. 2013. Genetic structure of *Lycorma delicatula* (Hemiptera: Fulgoridae) populations in Korea: implication for invasion processes in heterogeneous landscapes. Bull. Entomol. Res. 103: 414–424.

Patel, R. K., and M. Jain. 2012. NGS QC toolkit: a toolkit for quality control of next generation sequencing data. PLoS One 7: e30619.

Peakall, R., and P. E. Smouse. 2006. GenAlex 6: Genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6: 288–295.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945–959.

Puritz, J. B., and R. J. Toonen. 2013. Next-generation sequencing for high-throughput molecular ecology: a step-by-step protocol for targeted multilocus genotyping by pyrosequencing. Microsatellites Methods Mol. Biol. 1006: 89–99.

Raymond, M., and F. Rousset. 1995. GENEPOP version 4.0: population genetics software for exact tests and ecumenicism. J. Hered. 86: 248–249.

Schug, M. D., K. A. Wetterstrand, M. S. Gaudette, R. H. Lim, C. M. Hutter, and C. F. Aquadro. 1998. The distribution and frequency of microsatellite loci in *Drosophila melanogaster*. Mol. Ecol. 7: 57–70.

Schuler, G. D., M. S. Boguski, E. A. Stewart, L. D. Stein, G. Gyapay, K. Rice, R. E. White, P. Rodriguez-Tomé, A. Aggarwal, E. Bajorek, S. Bentolila, et al. 1996. A gene map of the human genome. Science 274: 540–546.

Shendure, J., and H. Ji. 2008. Next-generation DNA sequencing. Nat. Biotechnol. 26: 1135–1145.

Sogawa, K. 1997. The monsoon-dependent migrations of rice planthoppers in East Asia, pp. 217–230. China National Rice Research Institute.

Sogawa, K., G. Liu, and Q. Qiang. 2009. Planthoppers: new threats to the sustainability of intensive rice production systems in Asia: prevalence of whitebacked planthoppers in Chinese hybrid rice and whitebacked planthopper resistance in Chinese japonica rice, pp. 257–280. International Rice Research Institute. Metro Manila, Philippines.

Squirrell, J., P. M. Hollingsworth, M. Woodhead, J. Russell, A. J. Lowe, M. Gibby, and W. Powell. 2003. How much effort is required to isolate nuclear microsatellites from plants? Mol. Ecol. 12: 1339–1348.

Sun, J. T., X. Y. Jiang, M. M. Wang, and X. Y. Hong. 2014. Development of microsatellite markers for, and a preliminary population genetic analysis of, the white-backed planthopper. Bull. Entomol. Res. 104: 765–773.

Tang, J. Y., B. H. Hu, and J. Q. Wang. 1996. Outbreak analysis of rice migratory pests in China and management strategies recommended. Acta Ecologica Sinica 16: 167–173.

Tang, Q. Y., G. W. Hu, J. Tang, and Y. Hu. 1998. Relationship between outbreak frequency of *Sogatella furcifera* (Horváth) and grown area of hybrid rice. J. Southwest Agric. Univ. 20: 456–459.

Techen, N., R. S. Arias, N. C. Glynn, Z. Pan, I. A. Khan, and B. E. Scheffler. 2010. Optimized construction of microsatellite-enriched libraries. Mol. Ecol. Resour. 10: 508–515.

Uhm, K. B., J. S. Park, Y. I. Lee, K. M. Choi, M. H. Lee, and J. O. Lee. 1988. Relationship between some weather conditions and immigration of the brown planthopper, *Nilaparvata lugens* (Stål). Korean J. Appl. Entomol. 27: 200–210.

Valdes, A. M., M. Slatkin, and N. B. Freimer. 1993. Allele frequencies at microsatellite loci—the stepwise mutation model revisited. Genetics 133: 737–749.

Van't Hof, A. E., P. M. Brakefield, I. J. Saccheri, and B. J. Zwaan. 2007. Evolutionary dynamics of multilocus microsatellite arrangements in the genome of the butterfly *Bicyclus anynana*, with implications for other Lepidoptera. Heredity 98: 320–328.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. Van de Lee, M. Hornes, A. Friters, J. Pot, J. Paleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23: 4407–4414.

Vuylsteke, M., R. Mank, R. Antonise, E. Bastiaans, M. L. Senior, C. W. Stuber, A. E. Melchinger, T. Lubberstedt, X. C. Xia, P. Stam, et al. 1999. Two high-density AFLP linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers. Theor. Appl. Genet. 99: 921–935.

Vukosavljev, M., G. D. Esselink, W. P. C. van't Westende, P. Cox, R. G. F. Visser, P. Arens, and M. J. M. Smulder. 2015. Efficient development of highly polymorphic microsatellite markers based on polymorphic repeats in transcriptome sequences of multiple individuals. Mol. Ecol. Res. 15: 17–27.

Wei, S.-J., B.-C. Shi, Y.-J. Gong, G.-H. Jin, X.-X. Chen, and X.-F. Meng. 2013. Genetic Structure and Demographic History Reveal Migration of the Diamondback Moth *Plutella xylostella* (Lepidoptera: Plutellidae) from the Southern to Northern Regions of China. PLoS One 8: e59654.

Zhai, B. P., and J. A. Cheng. 2006. The conference summary of workshop on the two primary migratory pests of rice, rice planthopper and rice leaf roller, in 2006. Chin. Bull. Entomol. 43: 585–588.

Zhang, D. X. 2004. Lepidopteran microsatellite DNA: redundant but promising. Trends Ecol. Evol. 19: 507–509.

Zhang, D. X., L. N. Yan, Y. J. Ji, G. M. Hewitt, and Z. S. Huang. 2009. Unexpected relationships of substructured populations in Chinese *Locusta migratoria*. BMC Evol. Biol. 9: 144.

Zhang, K. J., W. C. Zhu, X. Rong, Y. K. Zhang, X. L. Ding, J. Liu, D. S. Chen, Y. Du, and X. Y. Hong. 2013. The complete mitochondrial genomes of two rice planthoppers, *Nilaparvata lugens* and *Laodelphax striatellus*: conserved genome rearrangement in Delphacidae and discovery of new characteristics of atp8 and tRNA genes. BMC Genomics 14: 417.

Zhang, K. J., W. C. Zhu, X. Rong, J. Liu, X. L. Ding, and X. Y. Hong. 2014. The complete mitochondrial genome sequence of *Sogatella furcifera* (Horváth) and a comparative mitogenomic analysis of three predominant rice planthoppers. Gene 533: 100–109.

Received 5 November 2014; accepted 10 June 2015.