

Population genetic characterization of the Japanese oak silkmoth, *Antheraea yamamai* (Lepidoptera: Saturniidae), using novel microsatellite markers and mitochondrial DNA gene sequences

S.R. Kim^{1*}, K.Y. Kim^{1*}, J.S. Jeong², M.J. Kim², K.-H. Kim³, K.H. Choi¹ and I. Kim²

¹Department of Agricultural Biology, National Academy of Agricultural Science, Rural Development Administration, Wanju-gun, Republic of Korea

²College of Agriculture & Life Sciences, Chonnam National University, 300 Yongbong-dong, Buk-gu, Gwangju, Republic of Korea

³C & K Genomics Inc., Seoul National University, Seoul, Republic of Korea

*These authors contributed equally to this study.

Corresponding author: I. Kim

E-mail: ikkim81@chonnam.ac.kr

Genet. Mol. Res. 16 (2): gmr16029608

Received January 12, 2017

Accepted February 24, 2017

Published April 13, 2017

DOI <http://dx.doi.org/10.4238/gmr16029608>

Copyright © 2017 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution ShareAlike (CC BY-SA) 4.0 License.

ABSTRACT. The Japanese oak silkmoth, *Antheraea yamamai* Guérin-Méneville, 1861 (Lepidoptera: Saturniidae), is an important natural resource of industrial value for silk fiber production. Owing to a lack of geographic and population genetic information, systematic domestication of *An. yamamai* has not been possible yet. In this study, 10 microsatellite markers developed using next-generation sequencing and two mitochondrial DNA (mtDNA) gene sequences (*COI* and *ND4*) were used to investigate the genetic variation and geographic structure

of *An. yamamai* populations in South Korea. The two mtDNA gene sequences revealed very low total genetic variation and, consequently, low geographic variation, validating the use of more variable molecular markers. Genotyping of 76 *An. yamamai* individuals from nine localities in South Korea showed that the observed number of alleles at each locus ranged from 3 to 26, the polymorphism information content was 0.2990-0.9014, the observed and expected heterozygosities were 0.3252-0.9076 and 0.2500-0.9054, respectively, and F_{IS} was -0.654-0.520. The population-based F_{IS} , F_{ST} , R_{ST} , and global Mantel tests all suggested that the *An. yamamai* populations were overall well-interconnected, suggesting that any population can be used as a genetic source for domestication. Nevertheless, STRUCTURE analyses using microsatellite data and mtDNA sequences indicated the presence of two genetic pools in many populations, although a plausible explanation for this observation requires further studies.

Key words: Microsatellite; *Antheraea yamamai*; *COI*; *NDA*; Next-generation sequencing; Population structure

INTRODUCTION

The Japanese oak silkworm, *Antheraea yamamai* Guérin-Méneville, 1861, belongs to the lepidopteran family, Saturniidae, in the superfamily Bombycoidea, which includes the silk moths, emperor moths, sphinx moths, and their relatives. *Antheraea yamamai* is a medium to large-sized wild silkworm with a wing span of 110-152 mm. It is distributed throughout Far Eastern Russia, China, Korea, and Japan (Park et al., 1999). Phylogenetically, the family Saturniidae is closely related to the superfamily Bombycoidea, which includes the mulberry silkworm, *Bombyx mori* (Kim et al., 2014). The family Saturniidae attracted our interest because most of the larvae of this family, including that of *An. yamamai*, spin a silken cocoon, which can be used as silk fibers (Numata et al., 2015). In fact, the cocoon silk of *An. yamamai* is the most expensive silk in the world and is superior to that of the domestic silkworm in gloss, durability, and wrinkle.

For the systematic domestication of *An. yamamai*, which has industrial utility, knowledge about the genetic diversity and population genetic structure of its natural populations would be necessary. Up to now, several insect resources distributed in South Korea have been subjected to genetic surveys for diverse purposes using different molecular markers. The geographic variation and genetic diversity of the endangered tiny dragonfly (*Nannophya pygmaea* in Odonata) and dung beetle (*Copris tripartitus* in Coleoptera) were studied using mitochondrial DNA (mtDNA) sequences (Kim et al., 2007; Kang et al., 2012). The geographic variation of a pollinating bumblebee (*Bombus ignitus* in Hymenoptera) was studied using DNA barcode sequences and nuclear internal transcribed spacer 2 (Kim et al., 2009b; Oh et al., 2009; Oh et al., 2013), and its genetic structure was analyzed using microsatellite markers (Han et al., 2014). The population genetics of the predatory seven-spotted lady beetle (*Coccinella septempunctata* in Coleoptera) was studied using microsatellite markers (Kim et al., 2012). Strain-specific molecular markers for a silk-producing silkworm (*Bombyx mori* in Lepidoptera) were developed using microsatellite markers (Kim et al., 2010). With regard

to Saturniidae, several foreign species have been investigated. For example, the wild and domesticated populations of *A. assamensis* from India have been studied to estimate their genetic diversity using inter-simple sequence repeats (ISSR) and simple sequence repeats (SSR, Singh et al., 2012). In the eri silkworm, *Samia cynthia ricini*, found in Northeast India, ISSR markers have been employed to estimate the genetic structure of its unexploited, scattered populations at various altitudes, as well as the genetic interactions among its patchy subpopulations (Pradeep et al., 2011). Using limited samples, Liu et al. (2010) investigated the genetic diversity and genetic structure of *B. mori*, *A. pernyi*, and *Samia cynthia ricini* collected in China using randomly amplified polymorphic DNA markers. Li et al. (2009) sequenced a partial mtDNA *COI* from a limited number of geographic samples of *Caligula (Saturnia) japonica* collected from China. However, no population-level studies have been performed so far for any population of *An. yamamai* including those found in South Korea.

Microsatellite DNAs are SSRs that are abundantly present in both the coding and non-coding regions of all eukaryotic nuclear and some prokaryotic genomes (Tautz and Renz, 1984). The allelic hypervariability and co-dominant mode of inheritance of microsatellite DNAs render them useful for population genetic studies of closely related species (Tautz and Renz, 1984). Several techniques have been made available to facilitate screening and selection of suitable microsatellite markers, but next-generation sequencing (NGS) is regarded as one of the fastest and most cost-effective methods (Gardner et al., 2011).

In this study, we developed 10 novel microsatellite markers using NGS. These were used to genotype *An. yamamai* individuals from different locations in South Korea. We sequenced mtDNA for two partial genes (*COI* and *ND4*). Our main objectives were to unravel the extent and nature of genetic variation and population genetic structure of this species in South Korea.

MATERIAL AND METHODS

Sampling and DNA extraction

Eighty-one adult *An. yamamai* moths were collected using light traps at 13 Korean localities from July to August, 2015 (Figure 1 and **Table S1**). Approximately similar sampling efforts were used in all localities, but the sample size differed among localities due to differences in sampling success. Total DNA was extracted from the hind legs of the moths using the Wizard Genomic DNA Purification Kit (Promega, Madison, WI, USA) following the manufacturer instructions, and stored at -20°C until further use. To construct a DNA library, DNA quality and concentration were measured using a spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA).

Mitochondrial DNA amplification and sequencing

The mitochondrial (mt) *COI* gene sequence has been extensively used as a “DNA barcode” (Hebert et al., 2003). However, it has also been recommended as a useful marker for gaining an early insight into the patterning of genomic diversity within a species (Hajibabaei et al., 2007). In fact, *COI* sequences were reported to provide population genetic information applicable to a range of ecological and historical questions (Díaz-Ferguson et al., 2010). In addition to the *COI* sequence, we sequenced a partial *ND4* gene (618 bp). To amplify the 658 bp

region of the *COI* gene, polymerase chain reaction (PCR) was conducted under the following conditions: an initial denaturation step at 94°C for 7 min, followed by 35 amplification cycles (denaturation at 94°C for 1 min, annealing at 47.8°-56°C for 1 min, and extension at 72°C for 1 min), and a final extension step at 72°C for 7 min. The primers for the *COI* gene were adapted from Hajibabaei et al. (2006): LepF (5'-ATTCAACCAATCATAAAGATATTGG-3') and LepR (5'-TAAACTTCTGGATGTCCAAAAAATCA-3'). As an additional mt gene, several lepidopteran mt *ND* genes were considered based on their variability (e.g., Wan et al., 2013). After sequencing several *An. yamamai* individuals, the *ND4* gene was chosen on the basis of its amplification efficiency, sequence divergence, and number of variable sites. The primers for the amplification of the partial *ND4* gene were designed using previously published mt genome sequences of *An. yamamai* (Kim et al., 2009a): forward: 5'-TAATAGTATATACTCCCGTG-3' and reverse: 5'-GCTCATGTTGAAGCTCCTG-3'. The PCR amplification was performed using the following conditions: an initial denaturation step at 94°C for 5 min, 30 amplification cycles (denaturation at 94°C for 1 min, annealing at 48°-50°C for 1 min, and extension at 72°C for 1 min), and a final extension step at 72°C for 7 min. The PCR products were then purified using a PCR Purification Kit (Qiagen, Dusseldorf, Germany). Electrophoresis was performed in 0.5X TAE (Tris-Acetate EDTA) buffer on 0.5% agarose gels to confirm successful DNA amplification. DNA sequencing was conducted using the ABI PRISM® BigDye® Terminator v. 3.1 Cycle Sequencing Kit with an ABI 3100 Genetic Analyzer (PE Applied Biosystems, Foster City, CA, USA). All PCR products were sequenced in both directions.

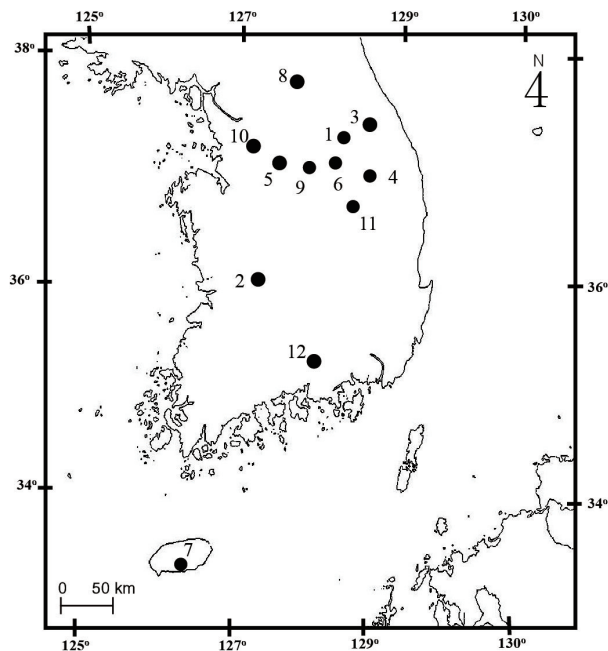


Figure 1. Sampling location of *Antheraea yamamai* in Korea. General locality names are as follows: 1. Chongil-myeon, Gangwon Province; 2. Mt. Unjang, Jeollabuk Province; 3. Nae-myeon, Gangwon Province; 4. Jungdong-myeon, Gangwon Province; 5. Mt. Mangbong, Gyeonggi Province; 6. Mt. Baekwoon, Chungcheongbuk Province; 7. Sanghyo-dong, Jeju Province; 8. Sanae-myeon, Gangwon Province; 9. Mt. Mireuk, Gangwon Province; 10. Jukjeon-dong, Gyeonggi Province; 11. Hyoja-myeon, Gyeongsangbuk Province; and 12. Macheon-myeon, Gyeongsangnam Province.

The sequences of both DNA strands from each individual were aligned using the ClustalW2 program (Larkin et al., 2007; <http://www.ebi.ac.uk/Tools/msa/clustalw2>), to obtain a finalized individual gene sequence. Different haplotypes were chosen for all individuals differing by one or more nucleotides by performing unordered pairwise comparisons among sequences, using PAUP v. 4.0b (Swofford, 2001). Haplotype designations were applied to the new sequences as they were discovered (i.e., AYBAR01, AYBAR02, AYBAR03, and so forth for *COI*; AYND401, AYND402, and AYND403 for *ND4*; and AYCNO1, AYCNO2, AYCNO3, and so forth for the concatenated sequences of both genes).

Development of microsatellite markers and genotyping

Approximately 200 ng purified genomic DNA was sheared to approximately 550-bp fragments using a Covaris S220 (Covaris, Woburn, MA, USA). The sheared genomic DNA was processed to produce a paired-end library using the TrueSeqnano DNA Library Kit (Illumina, San Diego, CA, USA). The size and concentration of the prepared library were confirmed using the Agilent 2100 Bioanalyzer system (Agilent Technologies, Santa Clara, CA, USA) and quantitative PCR-based KAPA library quantification kit (KAPA Biosystems, Wilmington, MA, USA), respectively. DNA sequencing of the prepared library was conducted on the Miseq (Illumina) with 250-bp paired-end reads. Assembly routines were performed using the CLC Genomics Workbench software v. 7 (CLC Bio, Aarhus, Denmark) and De Bruijn graph algorithm. An assembly mapping was obtained after removal of the Illumina adapters and low-quality sequences using the CLC trimmer function (default limit = 0.05). The fraction length and sequence similarities between DNA reads were set at the maximum stringencies of 0.50 and 0.80, respectively, whereas the minimum contig length was set to 70 bp.

The microsatellite sequences with 2-6 repeat motifs were searched using the Msatcommander program (Faircloth, 2008). Fifty candidate microsatellite loci were validated by PCR using primers designed in Primer 3 (Rozen and Skaletsky, 2000). The primers were designed with their lengths ranging from 20-26 nucleotides, annealing temperatures ranging from 49°-58°C, and product sizes ranging from 100-300 bp. Of the 50 candidate loci, 10 were eventually selected. The PCR parameters and GenBank accession Nos of the 10 loci are listed in Table 1.

Table 1. Ten microsatellite markers developed from *Antheraea yamamai*.

Marker name	Repeat motif	Primer sequence (5'-3')	Annealing temperature (°C)	Size (bp)	GenBank accession No.
10545	(AAT) ₁₂	GGGCTACAAAGAAATCTTAT	53	263	KJ735682
		GTCGTCTTGAAATGTTTTA			
10908	(AAT) ₁₅	CTGAGTACCAGATTAAAACA	53	123	KJ735684
		ATACCTGGTTAAAAACAAT			
12519	(ATT) ₁₂	GCATTCATTAGACAAATACCA	53	272	KJ735688
		CGTCATGGAGAGAAATATATC			
57897	(AC) ₁₂	TCGATCTGGTTTATCTTTTTAA	53	103	KJ735719
		ATGTCTGGAATTAACTCATC			
74763	(GAT) ₁₅	GATATACGAGTAGAGGAAGA	53	204	KM582134
		TAAAAACCCAAGACATACAA			
8575	(AT) ₂₀	GTTCAGTCGGTCATATTAAT	53	223	KM582137
		CGCTAAAAACAATAACAGAA			
47017	(AG) ₄₆	CTAAATTTGTCTTTACCAGC	53	234	KM582149
		TATATCGCCAGTTATATTGG			
30993	(ATC) ₃₃	GCAGTTTATTGAATAATGCT	53	177	KM582156
		ATATAGCCCATAGTCTTCTT			
123729	(GGT) ₃₀	TGGAATTCACCAAGATAATT	53	209	KM582158
		TATATAGCCTATAGCCTTCC			
50112	(AGT) ₃₀	CACGACTTTAGTTAGTACTT	53	245	KM582159
		ATAAGGGTTATACCAAGCTA			

For genotyping, one of each primer pair (Gencube, Seoul, Korea) was labeled with 6-FAM fluorescent dyes. The PCR was performed in a 25- μ L reaction volume containing 30 ng DNA, 1X PCR buffer [50 mM KCl, 10 mM Tris-HCl (pH 8.8), 1.5 mM MgCl₂], 2.5 mM dNTPs, 200 nM each primer, and 1 U Pure Speed PFU DNA polymerase (Smarteome, Jeonju, Korea) in an ABI 2720 Thermo cycler (Applied Biosystems). The following PCR cycling conditions were used: an initial denaturation step at 95°C for 3 min, 30 amplification cycles (denaturation at 94°C for 30 s, annealing at 53°C for 30 s, and extension at 72°C for 1 min), and a final extension step at 72°C for 5 min (Table 1). After the PCR, 0.2 μ L PCR product was mixed with 9.8 μ L Hi-Di Formamide (Applied Biosystems) and 0.2 μ L Liz-500 standard size (Applied Biosystems). The mixture was then denatured at 95°C for 5 min, placed on ice, and separated on an ABI 3730xl sequencer (Applied Biosystems). Allele sizes and genotypes were analyzed using GeneMapper® v. 4.1 (Applied Biosystems).

Statistical analysis of mitochondrial DNA

The within-population haplotype diversity (h) and nucleotide diversity (π) were estimated using Arlequin v. 3.5 (Excoffier and Lischer, 2010), considering only populations with more than two individuals. Consequently, among the 12 populations, nine populations, which comprised a total of 76 individuals, were subjected to analysis. However, only five populations (locality 2, Mt. Unjang; 3, Nae-myeon; 5, Mt. Mangbong; 8, Sanae-myeon; and 12, Macheon-myeon) contained more than two concatenated sequence-based haplotypes, suitable for genetic diversity evaluation. The maximum sequence divergence within each locality was obtained by extracting the within-locality estimates of the unrooted pairwise distances from PAUP (Swofford, 2001). The genetic distance and per-generation female migration rate were estimated from subroutines in Arlequin v. 3.5 (Excoffier and Lischer, 2010). To estimate the per-generation female migration rate, Nm (the product of the effective population size, N_e , and the migration rate, m), pairwise genetic distances (F_{ST}) were obtained based upon the equilibrium relationship proposed by Excoffier et al. (1992):

$$F_{ST} = 1 / (2Nm + 1) \quad (\text{Equation 1})$$

Permutation tests of significant differentiation among all pairs of localities (1000 bootstraps) were performed using the approach described by Excoffier et al. (1992). The distances between the DNA sequences were calculated using the Kimura 2-parameter method (Kimura, 1980).

For the phylogenetic analyses, well-aligned conserved blocks from each of the *COI* and *ND4* gene sequences were selected using GBlocks 0.91b (Castresana, 2000), with the minimum block length set to 10 and no gap positions allowed and concatenated. For the Bayesian inference (BI) and maximum-likelihood (ML) methods, we used the GTR + GAMMA + I model, which was selected by comparing AIC scores (Akaike, 1974) using Modeltest v. 3.7 (Posada and Crandall, 1998). The BI and ML methods were conducted using MrBayes v. 3.2.2 (Huelsenbeck and Ronquist, 2001) and RAxML-HPC2 on XSEDE v. 8.0.24 (Stamatakis, 2006), respectively, which were implemented on the CIPRES Portal v. 3.1 (Miller et al., 2010). For the BI analysis, two independent runs of four incrementally heated Markov chain Monte Carlo (MCMC) algorithms (one cold chain and three hot chains) were simultaneously run for one million generations. Tree sampling was conducted every 100

generations and the first 25% of the sampled trees were discarded as burn-in. An average of the split frequencies under 0.01 was the criterion we used to decide that convergence had been reached among the two independent runs. The confidence values of the BI tree are presented as the Bayesian posterior probabilities. The congeneric species *A. pernyi* (Liu et al., 2008) was included as outgroup. The generated trees were viewed with FigTree v. 1.4.2 (tree.bio.ed.ac.uk/software/figtree). When intraspecific mtDNA sequences are used for phylogenetic analysis, the haplotype relationships are often incompletely resolved due to the presence of polytomies. In this case, the haplotype relationships, resulting from the intrinsic dynamics of population processes, can sometimes be better illustrated using other methods, such as various kinds of reticulate networks. We therefore employed the median-joining algorithm incorporated in SplitsTree 4.14.3 (<http://www.splitstree.org>; Huson and Bryant, 2006) to further illustrate the relationships of the *An. yamamai* haplotypes.

Statistical analysis of microsatellite markers

For each locus, the observed number of alleles, expected heterozygosity (H_E), observed heterozygosity (H_O), and polymorphic information content (PIC; Botstein et al., 1980) were calculated using GENEPOP 4.0 (Rousset, 2008). The allelic richness standardized for variation in sample size was calculated using FSTAT 2.9.3.2 (Goudet, 2001). The deviation of genotypic frequencies from the Hardy-Weinberg equilibrium (HWE) was tested in GENEPOP using the Markov-chain approach modified from Guo and Thompson (1992) with 10,000 steps of dememorization and iteration. F_{IS} , which measures the deficiency of heterozygosity due to non-random mating, was estimated for each locus and each population using GenAlEx 6.5 (Peakall and Smouse, 2012). The genotypes at the 10 microsatellite loci were tested for linkage disequilibrium (LD) between all pairs of loci using GENEPOP. For both the HWE and LD tests, the 95% significance level was adjusted using Bonferroni correction.

Arlequin v. 3.5 (Excoffier and Lischer, 2010) was used to calculate F_{ST} (Weir and Cockerham, 1984) and R_{ST} (Slatkin, 1995) between all pairs of populations that contained more than two individuals. Consequently, nine populations were included in these analyses. The F_{ST} was estimated based on the infinite allele model of mutation (Weir and Cockerham, 1984), whereas R_{ST} was estimated based on the sum of squared size difference, assuming a stepwise mutation process (Slatkin, 1995). The significance of the F_{ST} between all pairs of populations was obtained using Fisher's exact test based on 10,000 permutations, and adjusted using the Bonferroni correction.

In order to identify the true number of populations (clusters) and assign individuals to each cluster to infer population structure, an MCMC algorithm-based clustering approach implemented in STRUCTURE v. 2.3.1 (Pritchard et al., 2000) was used. This program assigns individuals to a given number populations (K) probabilistically based on their multi-locus genotypes. An admixture model with correlated allele frequencies was used, and the K -value was set from 1-10. Ten independent runs were performed for each K -value, with a burn-in period of 10,000 iterations, followed by 50,000 iterations for data collection. Structure Harvester v. 0.6.8 (Earl and vonHoldth, 2012) was used to visualize the structure results. A principal coordinates analysis (PCoA) via covariance with standardization of the individual F_{ST} was performed to detect and plot the relationships between individuals belonging to different populations and among populations using GenAlEx v. 6.5 with default

parameters (Peakall and Smouse, 2012). A neighbor-joining (NJ) analysis was performed to infer the relationships among populations using uncorrected F_{ST} . Bootstrap resampling (10,000 replicates) was performed to test the robustness of the tree using the web version of the POPTREEW software (Takezaki et al., 2014).

Mantel tests of the correlations between the genetic and geographical distance matrices were performed using two sets of F_{ST} ; pairwise $F_{ST}/1 - F_{ST}$ and $R_{ST}/1 - R_{ST}$ (Rousset, 1997), using the Web Service IBDWS program v. 3.11 (Jensen et al., 2005). The geographical distances were measured using the coordinates between all location pairs, and were \log_{10} -transformed prior to all analyses. The significance of the Mantel test was assessed by 30,000 randomizations of the F_{ST} matrix.

RESULTS

Mitochondrial DNA-based analyses

The mtDNA analyses of the 79 individuals revealed five haplotypes (AYBAR01-AYBAR05) of the 658-bp *COI* gene sequence and three haplotypes (AYND401-AYND403) of the 618-bp *ND4* gene sequence, with a maximum sequence divergence of 0.46% (3 bp) and 0.32% (2 bp), respectively (data not shown). When *COI* and *ND4* were concatenated (1276 bp), a total of six haplotypes (AYCN01-AYCN06) were identified, with a maximum sequence divergence ranging from 1 (0.08) to 4 bp (0.31%), resulting in a low sequence divergence just as the individual gene sequences (**Table S2**). For simplicity, the analyses are focused on the concatenated sequences.

The both the BI and ML phylogenetic analyses performed to investigate the relationships and divergence among the *An. yamamai* haplotypes strongly supported the monophyly of the *An. yamamai* haplotypes separated from the outgroup species (Figure 2A). Three haplotypes (AYCN02, AYCNO4, and AYCNO5) formed one distinguishable subgroup, with a substantially high nodal support (BI: 0.98; ML: 90%), but the remaining three haplotypes (AYCN01, AYCNO3, and AYCNO6) did not form a clear group. Subsequent network analysis separated the remaining three haplotypes into another distinguishable group, indicating the presence of two slight but distant groups of *An. yamamai* in South Korea (named AYCNO1, AYCNO3, and AYCNO6 as group A and AYCNO2, AYCNO4, and AYCNO5 as group B; Figure 2B). These two groups showed a minimum of only one nucleotide difference between them (**Table S2**). Six localities were comprised of members of only group A, whereas a single locality (locality 8) was comprised of only group B members. On the other hand, the remaining six localities contained members of both groups A and B (**Table S1** and Figure 1), indicating no obvious distribution pattern between groups A and B.

The estimation of the within-locality diversity indicated moderate-to-low h and π per population, ranging from 0.5 to 0.67886 and 0.000392 to 0.000644, respectively (Table 2). The F_{ST} and Nm between the pairs of localities (only including those with more than two haplotypes per population) showed no locality pair with statistically significant F_{ST} , indicating a genetic interrelation of all *An. yamamai* populations in South Korea (**Table S3**). Collectively, the mtDNA-based population genetic analyses indicated that the *An. yamamai* populations in South Korea tend to show very low diversity and are well-interconnected.

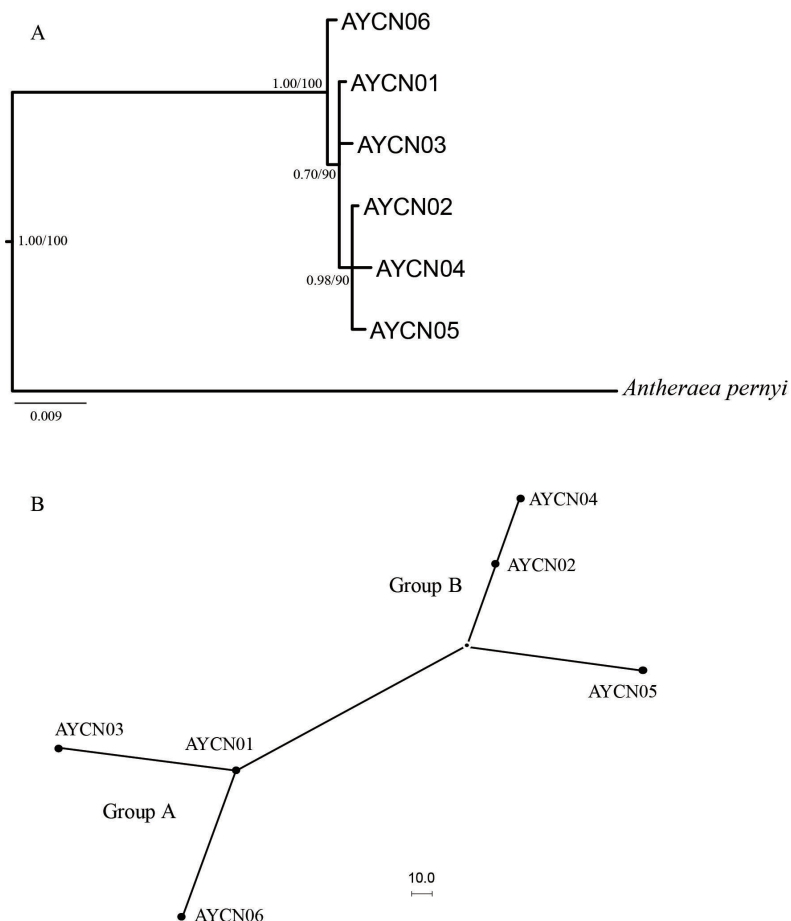


Figure 2. Phylogeny of *Antheraea pernyi* haplotypes constructed using concatenated sequences of the *COI* and *ND4* genes. **A.** Bayesian inference tree. The first numbers at each node indicates the Bayesian posterior probability and the second value indicates the bootstrap percentage of 1000 pseudoreplicates by maximum likelihood analysis. The scale bar indicates the number of substitutions per site. *Antheraea pernyi* was used as an outgroup in order to root the tree. **B.** Median-joining networks. The branch lengths represent the number of character-state changes occurring on that branch. The hypothetical haplotype that was not found in this study is marked with a dot (·).

Table 2. Within-locality diversity estimates obtained from the concatenated sequences of *COI* + *ND4* of *Antheraea yamamai*.

Locality	SS ^a	NH ^b	<i>h</i> ^c	NP ^d	MSD ^e (%)	MPD ^f	π ^g
2. Mt. Unjang	4	2	0.5000	1	0.08	0.500000	0.000392
3. Nae-myeon	8	3	0.6786	2	0.16	0.821429	0.000644
5. Mt. Mangbong	8	3	0.6071	2	0.24	0.678571	0.000532
8. Sanae-myeon	12	3	0.5303	2	0.16	0.651515	0.000511
12. Macheon-myeon	15	3	0.5143	2	0.16	0.552381	0.000433

^aSample size, ^bnumber of haplotypes, ^chaplotype diversity, ^dnumber of polymorphic sites, ^emaximum sequence divergence, ^fmean number of pairwise differences, ^gnucleotide diversity.

Development of microsatellite markers and genotyping

We used NGS to obtain sequence information about *An. yamamai*. An Illumina paired-end library with a mean insert size of 556 bp was made using one adult individual. The Illumina Mi-Seq platform generated 250 bp 2X paired-end reads from a sequencing library, resulting in a total of 40,182,580 reads, 149,317 contigs longer than 200 bp, and an average contig length of 2848 bp (Table S4). Tetranucleotide repeats were the most abundant class of microsatellites (21,501 regions) detected in the *An. yamamai* genome, followed by dinucleotide (14,639 regions) and trinucleotide (10,983 regions) repeats (Figure 3A). The most frequent tetranucleotide repeat in the *An. yamamai* genome was CTGT (21.1%), followed by ACAG (20.9%), AAAC (1.8%), and GTTT (1.7%) (Figure 3B).

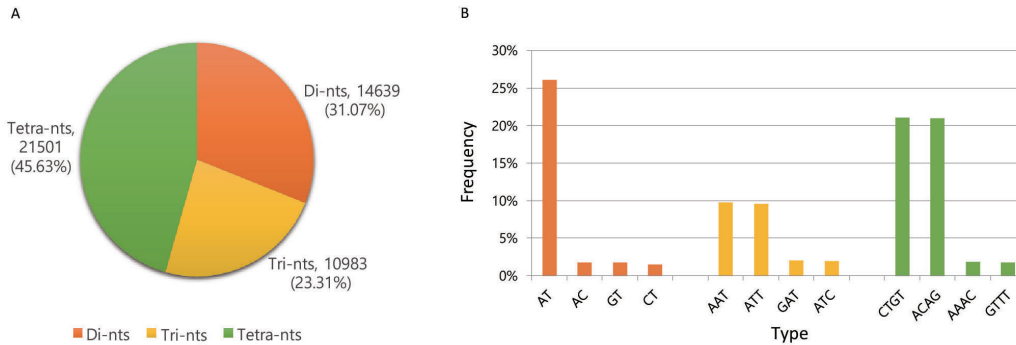


Figure 3. Distribution of microsatellite DNA. **A.** Distribution of di- and tetranucleotide microsatellites on contigs with a minimum of 10X coverage. **B.** Distribution of the most frequent repeat motifs with a minimum of 10X coverage.

We initially selected 50 candidate microsatellites for testing the availability of primer sites, amplification efficiency, degree of polymorphism, and specificity for target loci. Of these, ten markers were eventually selected and used for subsequent genotyping (Table 1). The availability of these ten microsatellite markers ranged from 0.87 to 1.00, with an average of 0.98, indicating an overall high genotyping success (Table 3). For each locus, the allele numbers ranged from three (locus 10545) to 26 (locus 47017), with a mean number of 9.1. The major allele frequency ranged from 0.2095 (locus 47017) to 0.8092 (locus 10545). Consistent with the highest allele number, the genotype number (55) was also found to be the highest in locus 47017. The lowest genotype number (4) was found for locus 10545 and the average genotype number was 16.6. The H_o and H_e values ranged from 0.250 to 1.00 (mean = 0.6443) and 0.3252 to 0.9076 (mean = 0.6593), respectively, indicating a somewhat low H_o . Positive F_{IS} values, which indicate a heterozygote deficiency, were observed in three loci. No locus was found to be significantly different from HWE after applying the Bonferroni correction ($P = 0.05/10 \leq 0.005$). The tests of genotypic LD showed no significant allele associations among the 10 loci after applying the Bonferroni correction, suggesting that all loci can be considered independent markers.

Table 3. Characteristics of ten microsatellite loci in *Antheraea yamamai*.

Locus	N	N_A	Availability ^a	MAF	AR	No. genotype	H_E	H_O	PIC	F_{IS}	HWE ^b (P value)
10545	76	3.0	1.00	0.8092	1.5433	4.0	0.3252	0.2500	0.2990	-0.063	0.806
10908	76	4.0	1.00	0.5658	2.0241	5.0	0.5325	0.8421	0.4377	-0.621	0.375
12519	76	3.0	0.99	0.5333	2.3677	6.0	0.6052	0.6133	0.5367	-0.654	0.079
57897	76	6.0	1.00	0.4737	3.3964	6.0	0.6673	1.0000	0.6129	-0.053	0.720
74763	76	6.0	0.96	0.5616	2.2336	9.0	0.5901	0.2603	0.5280	-0.089	0.733
8575	76	8.0	0.96	0.3562	2.2151	13.0	0.7206	0.6575	0.6750	0.520	0.210
47017	76	26.0	0.97	0.2095	2.4337	55.0	0.9076	0.9054	0.9014	0.165	0.408
30993	76	15.0	1.00	0.4145	2.9396	31.0	0.7758	0.7895	0.7562	0.084	0.334
123729	76	12.0	0.87	0.2879	2.5748	26.0	0.8073	0.6515	0.7834	-0.038	0.425
50112	76	8.0	1.00	0.4934	2.8269	11.0	0.6614	0.4737	0.6106	-0.117	0.508
Mean	76	9.1	0.98	0.4705	2.4555	16.6	0.6593	0.6443	0.6141		

N, number of tested individuals; N_A , number of observed alleles; MAF, major allele frequency; AR, allele richness; and PIC, polymorphic information contents. ^aAvailability is defined as $1 - Obs / n$, where *Obs* is the number of observations and *n* is the number of individuals sampled. ^bSignificant deviation from Hardy-Weinberg equilibrium after Bonferroni correction.

Population analysis using microsatellite markers

The allelic patterns across populations were investigated using multiple approaches (Figure 4). The mean total number of alleles per population was highest in locality 12 (Macheon-myeon) at 5.6 (SD = 1.454), and the lowest in locality 2 (Mt. Unjang) at 2.60 (SD = 0.221). A limited allelic diversity was found in locality 2 probably because of the small sample size (four individuals) at this locality. Excluding this locality, no obvious differences were detected in the mean total number of alleles among populations. Six of the nine populations showed the presence of private alleles, with their numbers ranging from 0.1 (SD = 0.1; locality 5) to 0.7 (SD = 0.335; locality 12), but they were not statistically different from each other. The within-population gene diversity, which corresponds to the H_E in diploid data, ranged from 0.648 (SD = 0.050; locality 6, Mt. Mangbong) to 0.503 (Figure 4; SD = 0.059; locality 2, Mt. Unjang).

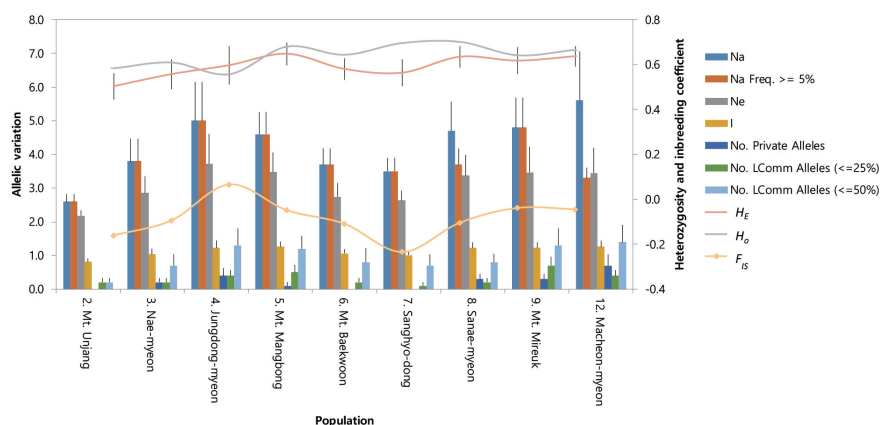


Figure 4. Mean allelic patterns across nine populations based on 10 loci *Antheraea yamamai*. N_A , number of different alleles; N_A (Freq. $\geq 5\%$), number of alleles with frequency greater than 5%; N_E , number of effective alleles; I , Shannon's information index; No. Private Alleles, number of alleles unique to a single population; No. LComm Alleles ($\leq 25\%$), number of locally common alleles occurring in 25% or less of the populations; No. LComm Alleles ($\leq 50\%$), number of locally common alleles occurring in 50% or less of the populations; H_E , expected heterozygosity; H_O , observed heterozygosity; and F_{IS} , inbreeding coefficient. Vertical bars represent the standard error.

Examination of the likelihood scores from the 10 replicate runs across K -values from 1-10 indicated that the optimal K -value was 2, suggesting the presence of two genetic groups (Figure 5). The assignment results from $K = 2$ showed that all sampled individuals exhibited admixture between the two gene-pools with roughly equal contributions from both gene-pools. This admixture, however, was found in all populations, showing no correspondence between the geographical and gene-pool assignments.

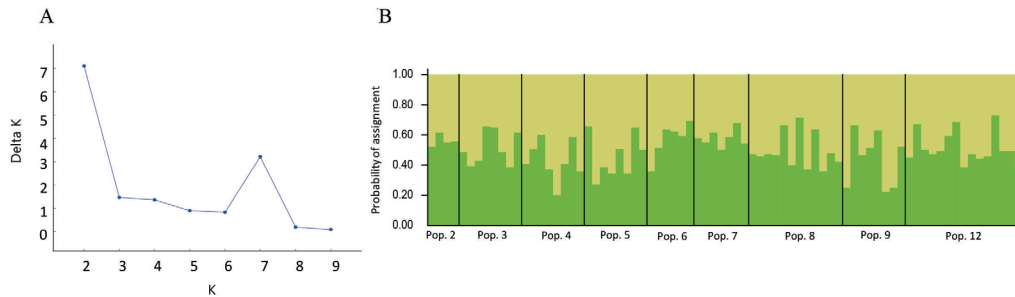


Figure 5. Clustering analysis of the multilocus microsatellite data of *Antheraea yamamai* performed using STRUCTURE. (A) Plot of Delta K calculated with the formula $\Delta K = \text{mean}(|L''(K)|) / \text{sd}(L(K))$, $N = 76$. (B) Bar plot of estimated membership of each individual in $K = 2$ clusters. Black bars separate the nine sampled populations. Different colors represent different gene-pools. Pop. 2, Mt. Unjang; Pop. 3, Nae-myeon; Pop. 4, Jungdong-myeon; Pop. 5, Mt. Mangbong; Pop. 6, Mt. Baekwoon; Pop. 7, Sanghyo-dong; Pop. 8, Sanae-myeon; Pop. 9, Mt. Mireuk; and Pop. 12, Macheon-myeon.

A PCoA analysis based on the first two principal coordinates was performed to investigate the population patterns based on the genetic distances among populations (Figure 6). The first two components explained 69.75% of the total variation. The first component, which accounted for 41.66% of the variation, roughly explained the divergence of localities 2 (Mt. Unjang) and 3 (Nae-myeon), but no other obvious sub-cluster was observed. On the other hand, the individual-based PCoA analysis revealed no discernible group, and even individuals from localities 3 and 4 did not form any inclusive group (data not shown).

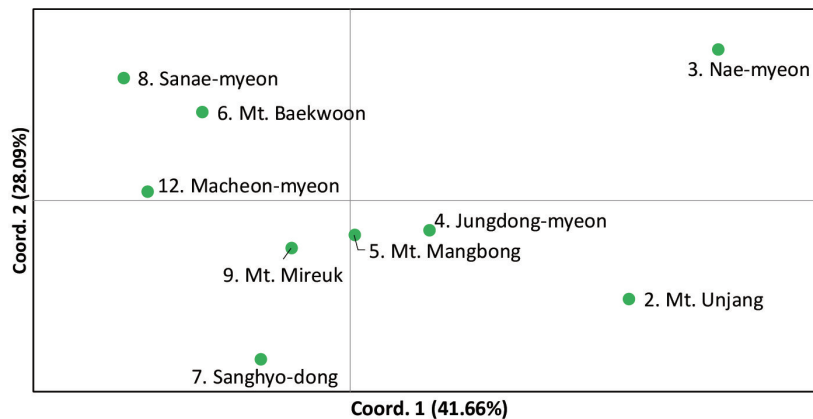


Figure 6. Principal coordinate analysis (PCoA) of populations showing allelic variance at ten microsatellite loci. The percent variation explained by the 1st and 2nd axes are indicated.

The analysis of the F_{ST} and R_{ST} values between populations showed statistically significant differences between several population pairs (Table 4). In the F_{ST} analysis, a statistically significant divergence ($P \leq 0.05$) was most often observed in locality 3 (Nae-myeon) five times, in locality 8 (Sanae-myeon) four times, and in localities 5 (Mt. Mangbong) and 9 (Mt. Mireuk) three times. This indicates the presence of a certain level of divergence in a few populations, but no population was completely isolated from the others. In the R_{ST} analysis, a relatively small number of population pairs showed significant divergence (10 from F_{ST} and eight from R_{ST} among 36 comparisons). Both localities 3 and 9 were most often significantly divergent from the other localities four times. In both analyses, two population pairs (between localities 2 and 5 and 3 and 8) were most commonly found to be significantly divergent. The NJ analysis showed a slight divergence of localities 2 and 3, forming a separate group. However, the distance of this group from the remaining localities was very low, with a threshold value of 0.006, indicating only a slight divergence of the two populations (Figure 7). The global Mantel test revealed no significant correlation between the $F_{ST}/1 - F_{ST}$ genetic distance and geographical distance matrices ($r = 0.2402$, $P = 0.1587$; Figure 8A). Similarly, no significant correlation was observed between the $R_{ST}/1 - R_{ST}$ genetic distance and geographical distances ($r = -0.0701$, $P = 0.5908$; Figure 8B).

Table 4. Analysis of genetic differentiation between pairs of *Antheraea yamamai* populations.

	3	4	5	6	7	8	9	10	13
2. Mt. Unjang	0	0.07963	0.26293	0.20224*	0.59644	0.42355	0.27169*	0.05664*	0.40257*
3. Nae-myeon	0.01714	0	0.05108	0.03827	0.19596	0.06374	0.03775*	-0.02352	0.09619
4. Jungdong-myeon	-0.00081	0.04435	0	-0.03094	0.06747	-0.00574	0.01143	0.00144	0.03246
5. Mt. Mangbong	0.06613*	0.06126*	0.05191*	0	0.18886	0.08242	0.07965	-0.03219	0.13749
6. Mt. Baekwoon	0.00707	0.05629*	-0.00034	0.01248	0	0.04118	0.08076*	0.2044	0.0804
7. Sanghyo-dong	-0.00596	0.03591	-0.00893	0.01013	-0.02134	0	0.00394*	0.08142	0.01443*
8. Sanae-myeon	0.04197	0.04005*	0.01367	0.01178	-0.00386	-0.00241	0	0.05298	-0.01082
9. Mt. Mireuk	0.01791	0.08081*	0.03761*	0.0446*	0.00172	0.00444	0.03941	0	0.11452
12. Macheon-myeon	0.0152	0.07406*	0.01771	0.01	0.01124	-0.00503	0.00883	0.02979*	0

* $P < 0.05$. Below diagonal, F_{ST} ; and above diagonal, R_{ST} .

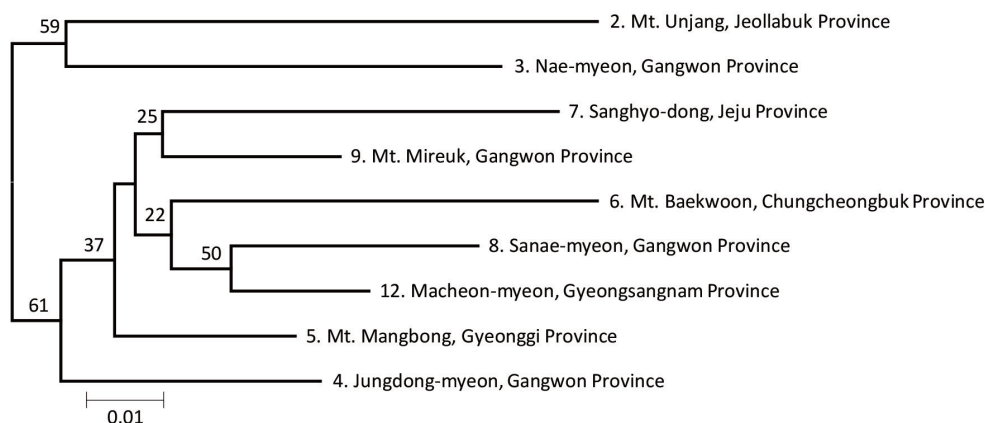


Figure 7. Neighbor-joining tree using F_{ST} distance among populations (Latter, 1972) from multilocus microsatellite data. Bootstrap values and a distance scale are provided on the tree.

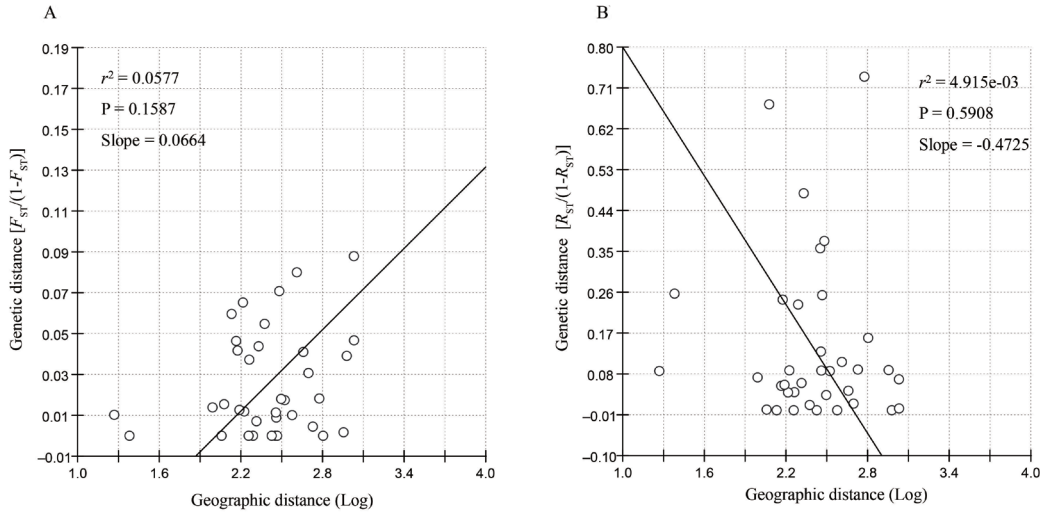


Figure 8. Isolation by distance analysis performed by plotting (A) $F_{ST}/1 - F_{ST}$ and (B) $R_{ST}/1 - R_{ST}$ against the log10-transformed geographic distance. Coefficients of determination (r^2), P-value, and regression slopes are given in the figure.

DISCUSSION

An. yamamai is an important insect resource for the agricultural industry in Asia, including South Korea, but no population genetic information is available for its domestic or foreign populations (Liu et al., 2010; Pradeep et al., 2011; Singh et al., 2012). Inferring the population genetic structure of this species would be important for systematic domestication and subsequent cocoon production, along with the traditional perspectives to understand the evolutionary process of the populations. The regional patterns of the population structure and genetic diversity distribution based on microsatellite markers and mtDNA sequences both suggested a limited population structure in South Korea.

Overall low mtDNA diversity

The concatenated mtDNA sequences of the *COI* and *ND4* genes (1276 bp) revealed very low genetic diversity. Only six haplotypes were obtained with a maximum sequence divergence of 0.31% (4 bp). It was difficult to find comparable data obtained under similar experimental conditions. Several previous studies have reported the magnitude of genetic diversity of several insect species found in South Korea using partial *COI* gene sequences (e.g., Kang et al., 2012). When these results were considered, the maximum sequence divergence of the *COI* gene in *An. yamamai* (0.46%, 3 bp) was one of the lower ones. The others ranged from 0.15 (pear psylla in Hemiptera; Kang et al., 2012) to 11.0% (oriental mole cricket in Orthoptera; Kim et al., 2007). Similarly, the h of *An. yamamai*, ranging from 0.5000 to 0.6071 in *COI* (0.5 to 0.67886 in combination), was one of the lower ones (Table 2). The value of h was 0.5022-0.8372 in the Asian cavity-nesting honey bee (*Apis cerana*), 0.5556-0.8444 in the endangered dragonfly (*Nannophya pygmaea*), 0.7111-0.9286 in the swallowtail butterfly (*Papilio xuthus*), and 0.5333-0.8929 in the cabbage butterfly (*Pieris rapae*) (Kim et al., 2007;

Jeong et al., 2009; Lee et al., 2016). Furthermore, in *An. yamamai*, the π was moderate-to-low, ranging from 0.000760 to 0.001194 in *COI* (0.000392 to 0.000644 in combination), whereas the value of π was 0.00101-0.00337 in *Papilio xuthus* and 0.00129-0.00641 in *Pieris rapae* (Jeong et al., 2009), showing a several-fold smaller value of π in *An. yamamai* compared to the other species.

Several mechanisms might be responsible for this low genetic diversity (Hedrick, 1974). Among them, population size is of primary interest because a sustained reduction in population size might result in loss of diversity. Many factors govern population size, but no obvious reason can be identified in the case of *An. yamamai*. For example, the larvae of *An. yamamai* feed on several host species (e.g., several oak species) (Park et al., 1999). These are distributed throughout and are common in the mountainous regions of South Korea (Park et al., 1999). No serious reduction in the population size of the host plants has been reported. Furthermore, certain natural enemies (e.g., parasitoid flies, wasps, and birds; http://tpittaway.tripod.com/silk/a_yam.htm; last visited on December 12, 2016) might cause a population reduction, but no data on their impact on the population size of *An. yamamai* is available. However, a heavy loss of the congeneric species *A. mylitta*, *A. proylei*, and *A. roylei* by a stink bug *Canthecona furcellata* (Hemiptera: Pentatomidae) has been reported from India (Sen et al., 1971; Singh et al., 1992; Bardagade and Gathalkar, 2016). Our field observations allow us to speculate that *An. yamamai* is not an “abundant” species because several night hours of trapping effort at most localities most often only resulted in a few captured individuals. However, direct inferences of the population size from trapping effort might not necessarily be accurate. Thus, further studies on the population dynamics of *An. yamamai* would be necessary.

Very weak population structure based on microsatellite markers

The primer pairs for the 10 microsatellite markers successfully amplified the target DNA in *An. yamamai*. Compared with mtDNA sequences, microsatellite markers detect a higher level of diversity and are useful for studies of population genetic structure. In fact, the overall genetic variability for each population, represented by Shannon’s information index, was rather high with an average of 1.122 (Figure 4). For example, a microsatellite analysis of an alpine butterfly, *Colias behrii* (Lepidoptera: Pieridae), collected from 18 sites in Nevada showed a Shannon’s information index ranging from 0.521 to 0.763 (Schoville et al., 2012). The high value of Shannon’s information index validates the effectiveness of microsatellite loci for performing studies of population genetic structure.

Our microsatellite data indicate that the *An. yamamai* populations in South Korea do not appear to be small, isolated populations, which would facilitate inbreeding. When heterozygote deficiencies are prevalent across loci, these findings are often ascribed to a variety of factors including inbreeding and substructuring (Lade et al., 1996). Our F_{IS} data indicated positive F_{IS} values, which can be interpreted as a signal of heterozygote deficiency in three loci (Table 3). However, the population-level analysis showed positive F_{IS} values only in one population (locality 4; Figure 4), indicating no overall heterozygote deficiency in the South Korean *An. yamamai* populations. Although population subdivision could also contribute to heterozygote deficiency, the pairwise F_{ST} and R_{ST} values revealed significant differentiation ($P < 0.05$) only in a limited number of population pairs (10 and 8 among 36 F_{ST} and R_{ST} comparisons, respectively). Thus, no population was obviously differentiated from the rest

(Table 1). This result is also reflected in the population-level NJ tree, which shows that only two populations (localities 2 and 3) were divergent from the remaining populations. However, even this divergence was marginal with low nodal support for the group and with low genetic distance (Figure 7). Similarly, the global Mantel test based on both $F_{ST}/1 - F_{ST}$ ($r = 0.2402$, $P = 0.1587$) and $R_{ST}/1 - R_{ST}$ ($r = -0.0701$, $P = 0.5908$) revealed no significant correlation (Figure 8).

A previous study on the relationships among population isolation, heterozygosity, and inbreeding depression on a butterfly species (*Melitaea cinxia*) indicated that small, isolated populations were subjected to inbreeding, presented decreased heterozygosity and an increased risk of extinction (Saccheri et al., 1998). Contrary to *Melitaea cinxia*, the current population genetic analyses of *An. yamamai* appear to suggest an intermediate-to-high gene flow, which facilitates interconnection among populations.

Considering the nearly absent geographic subdivision, the results of the STRUCTURE analysis revealed an optimal K -value of 2, indicating a division of genetic variation into two groups (Figure 5). However, the two clusters were distributed across all populations at a similar frequency, indicating that the genetic structure of *An. yamamai* is not related to its geographic distribution in South Korea. Interestingly, the mtDNA-based data also supported the presence of the two slight but discernible clusters without obvious geographic distinction (Figure 2B). Considering that the two gene clusters are widely distributed in South Korea, the event that caused this gene clustering might have occurred long ago with enough time to allow the admixture of the two gene clusters throughout South Korea. However, the exact cause for such a gene clustering event is not known.

In summary, our field observations allowed us to speculate that *An. yamamai* is not an “abundant” species. However, most of our data, such as the deficiency of heterozygosity only in limited loci, significant F_{ST} and R_{ST} only in a few population comparisons, and only a slight separation of a few populations from the others, obviously indicate that the *An. yamamai* populations in South Korea comprise a large panmictic unit. Thus, the sampling effort for the domestication of *An. yamamai* as a genetic resource might focus on any of the regions where *An. yamamai* is found, excluding the little collected populations (e.g., locality 2).

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by the Rural Development Administration, Republic of Korea (grant #PJ010442).

REFERENCES

- Akaike H (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19: 716-723. <http://dx.doi.org/10.1109/TAC.1974.1100705>
- Bardagade DD and Gathalkar GB (2016). First predation record of *Canthecona furcellata* (Wolff.) (Hemiptera: Pentatomidae) on spinning stage silkworm *Antheraea mylitta* (Drury). *Entomol. Res.* 46: 236-245. <http://dx.doi.org/10.1111/1748-5967.12169>
- Botstein D, White RL, Skolnick M and Davis RW (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32: 314-331.

- Castresana J (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17: 540-552. <http://dx.doi.org/10.1093/oxfordjournals.molbev.a026334>
- Díaz-Ferguson E, Robinson JD, Silliman B and Wares JP (2010). Comparative phylogeography of North American Atlantic salt marsh communities. *Estuaries Coasts* 33: 828-839. <http://dx.doi.org/10.1007/s12237-009-9220-6>
- Earl DA and vonHoldth BM (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 4: 359-361. <http://dx.doi.org/10.1007/s12686-011-9548-7>
- Excoffier L and Lischer HE (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10: 564-567. <http://dx.doi.org/10.1111/j.1755-0998.2010.02847.x>
- Excoffier L, Smouse PE and Quattro JM (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491.
- Faircloth BC (2008). msatcommander: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Mol. Ecol. Resour.* 8: 92-94. <http://dx.doi.org/10.1111/j.1471-8286.2007.01884.x>
- Gardner MG, Fitch AJ, Bertozzi T and Lowe AJ (2011). Rise of the machines--recommendations for ecologists when using next generation sequencing for microsatellite development. *Mol. Ecol. Resour.* 11: 1093-1101. <http://dx.doi.org/10.1111/j.1755-0998.2011.03037.x>
- Goudet J (2001). *FSTAT*: A program to estimate and test gene diversities and fixation indices, version 2.9.3. <http://www2.unil.ch/popgen/softwares/fstat.htm>. Accessed November 26, 2016.
- Guo SW and Thompson EA (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48: 361-372. <http://dx.doi.org/10.2307/2532296>
- Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, et al. (2006). DNA barcodes distinguish species of tropical Lepidoptera. *Proc. Natl. Acad. Sci. USA* 103: 968-971. <http://dx.doi.org/10.1073/pnas.0510466103>
- Hajibabaei M, Singer GA, Hebert PD and Hickey DA (2007). DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends Genet.* 23: 167-172. <http://dx.doi.org/10.1016/j.tig.2007.02.001>
- Han T, Park H, Park IG, Yoon HJ, et al. (2014). Genetic structure of Korean populations of bumblebees *Bombus ignitus* (Hymenoptera: Apidae) as revealed by microsatellite markers. *Entomol. Res.* 44: 262-270. <http://dx.doi.org/10.1111/1748-5967.12077>
- Hebert PD, Cywinska A, Ball SL and deWaard JR (2003). Biological identifications through DNA barcodes. *Proc. Biol. Sci.* 270: 313-321. <http://dx.doi.org/10.1098/rspb.2002.2218>
- Hedrick PW (1974). Genetic variation in a heterogeneous environment. I. Temporal heterogeneity and the absolute dominance model. *Genetics* 78: 757-770.
- Huelsenbeck JP and Ronquist F (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754-755. <http://dx.doi.org/10.1093/bioinformatics/17.8.754>
- Huson DH and Bryant D (2006). Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23: 254-267. <http://dx.doi.org/10.1093/molbev/msj030>
- Jensen JL, Bohonak AJ and Kelley ST (2005). Isolation by distance, web service. *BMC Genet.* 6: 13. <http://dx.doi.org/10.1186/1471-2156-6-13>
- Jeong HC, Kim JA, Im HH, Jeong HU, et al. (2009). Mitochondrial DNA sequence variation of the swallowtail butterfly, *Papilio xuthus*, and the cabbage butterfly, *Pieris rapae*. *Biochem. Genet.* 47: 165-178. <http://dx.doi.org/10.1007/s10528-008-9214-2>
- Kang AR, Kim K-G, Park JW and Kim I (2012). Genetic diversity of the dung beetle, *Copris tripartitus* (Coleoptera: Scarabaeidae), that is endangered in Korea. *Entomol. Res.* 42: 247-261. <http://dx.doi.org/10.1111/j.1748-5967.2012.00470.x>
- Kim KG, Jang SK, Park DW, Hong MY, et al. (2007). Mitochondrial DNA sequence variation of the tiny dragonfly, *Nannophya pygmaea* (Odonata: Libellulidae). *Int. J. Indust. Entomol.* 15: 47-58.
- Kim KY, Kang PD, Lee KG, Oh HK, et al. (2010). Microsatellite analysis of the silkworm strains (*Bombyx mori*): high variability and potential markers for strain identification. *Genes Genomics* 32: 532-543. <http://dx.doi.org/10.1007/s13258-010-0066-x>
- Kim SR, Kim MI, Hong MY, Kim KY, et al. (2009a). The complete mitogenome sequence of the Japanese oak silkworm, *Antheraea yamamai* (Lepidoptera: Saturniidae). *Mol. Biol. Rep.* 36: 1871-1880. <http://dx.doi.org/10.1007/s11033-008-9393-2>
- Kim MJ, Kim K-H, Kim KY, Cho Y, et al. (2012). Development and characterization of nine polymorphic microsatellite markers in the seven-spotted lady beetle, *Coccinella septempunctata* (Coleoptera: Coccinellidae). *Afr. J. Biotechnol.* 11: 16731-16736.
- Kim MJ, Yoon HJ, Im HH, Jeong HU, et al. (2009b). Mitochondrial DNA sequence variation of the bumblebee, *Bombus ardens* (Hymenoptera: Apidae). *J. Asia Pac. Entomol.* 12: 133-139. <http://dx.doi.org/10.1016/j.aspen.2009.02.003>

- Kim MJ, Wang AR, Park JS and Kim I (2014). Complete mitochondrial genomes of five skippers (Lepidoptera: Hesperidae) and phylogenetic reconstruction of Lepidoptera. *Gene* 549: 97-112. <http://dx.doi.org/10.1016/j.gene.2014.07.052>
- Kimura DK (1980). Likelihood methods for the von Bertalanffy growth curve. *Fish Bull.* 77: 765-776.
- Lade JA, Murray ND, Marks CA and Robinson NA (1996). Microsatellite differentiation between Phillip Island and mainland Australian populations of the red fox *Vulpes vulpes*. *Mol. Ecol.* 5: 81-87. <http://dx.doi.org/10.1111/j.1365-294X.1996.tb00293.x>
- Larkin MA, Blackshields G, Brown NP, Chenna R, et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947-2948. <http://dx.doi.org/10.1093/bioinformatics/btm404>
- Latter BD (1972). Selection in finite populations with multiple alleles. 3. Genetic divergence with centripetal selection and mutation. *Genetics* 70: 475-490.
- Lee JY, Wang AR, Choi YS, Thapa R, et al. (2016). Mitochondrial DNA variations in Korean *Apis cerana* (Hymenoptera: Apidae) and development of another potential marker. *Apidologie (Celle)* 47: 123-134. <http://dx.doi.org/10.1007/s13592-015-0381-y>
- Li Y, Yang B, Wang H, Xia R, et al. (2009). Mitochondrial DNA analysis reveals a low nucleotide diversity of *Caligula japonica* in China. *Afr. J. Biotechnol.* 8: 2707-2712.
- Liu Y, Li Y, Pan M, Dai F, et al. (2008). The complete mitochondrial genome of the Chinese oak silkworm, *Antheraea pernyi* (Lepidoptera: Saturniidae). *Acta Biochim. Biophys. Sin. (Shanghai)* 40: 693-703. <http://dx.doi.org/10.1093/abbs/40.8.693>
- Liu YQ, Qin L, Li YP, Wang H, et al. (2010). Comparative genetic diversity and genetic structure of three Chinese silkworm species *Bombyx mori* L. (Lepidoptera: Bombycidae), *Antheraea pernyi* Guérin-Meneville and *Samia cynthia ricini* Donovan (Lepidoptera: Saturniidae). *Neotrop. Entomol.* 39: 967-976. <http://dx.doi.org/10.1590/S1519-566X2010000600019>
- Miller MA, Pfeiffer W and Schwartz T (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the gateway computing environments workshop (GCE). New Orleans, 1-8.
- Numata K, Sato R, Yazawa K, Hikima T, et al. (2015). Crystal structure and physical properties of *Antheraea yamamai* silk fibers: Long poly(alanine) sequences are partially in the crystalline region. *Polymer (Guildf)* 77: 87-94. <http://dx.doi.org/10.1016/j.polymer.2015.09.025>
- Oh HK, Yoon HJ, Kim MJ, Jeong HU, et al. (2009). ITS2 ribosomal DNA sequence variation of the bumblebee, *Bombus ardens* (Hymenoptera: Apidae). *Genes Genomics* 31: 293-303. <http://dx.doi.org/10.1007/BF03191202>
- Oh HK, Yoon HJ, Lee JY, Park JS, et al. (2013). Population genetic structure of the bumblebee, *Bombus ignitus* (Hymenoptera: Apidae), based on mitochondrial COI gene and nuclear ribosomal ITS2 sequences. *Int. J. Indust. Entomol.* 27: 142-158. <http://dx.doi.org/10.7852/ijie.2013.27.1.142>
- Park KT, Kim SS, Tshistjakov YA and Kwon YD (1999). Illustrated catalogue of moths in Korea I (Sphingidae, Bombycoidea, Notodontidae). In: *Insects of Korea [Series 4]* (Park KT, ed.). Jungghaeng-Sa, Seoul.
- Peakall R and Smouse PE (2012). GenAEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics* 28: 2537-2539. <http://dx.doi.org/10.1093/bioinformatics/bts460>
- Posada D and Crandall KA (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817-818. <http://dx.doi.org/10.1093/bioinformatics/14.9.817>
- Pradeep AR, Jingade AH, Singh CK, Awasthi AK, et al. (2011). Genetic analysis of scattered populations of the Indian eri silkworm, *Samia cynthia ricini* Donovan: Differentiation of subpopulations. *Genet. Mol. Biol.* 34: 502-510. <http://dx.doi.org/10.1590/S1415-47572011005000033>
- Pritchard JK, Stephens M and Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.
- Rousset F (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145: 1219-1228.
- Rousset F (2008). genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Resour.* 8: 103-106. <http://dx.doi.org/10.1111/j.1471-8286.2007.01931.x>
- Rozen S and Skaletsky H (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* 132: 365-386.
- Saccheri I, Kuussaari M, Kankare M, Vikman P, et al. (1998). Inbreeding and extinction in a butterfly metapopulation. *Nature* 392: 491-494. <http://dx.doi.org/10.1038/33136>
- Schoville SD, Lam AW and Roderick GK (2012). A range-wide genetic bottleneck overwhelms contemporary landscape factors and local abundance in shaping genetic patterns of an alpine butterfly (Lepidoptera: Pieridae: *Colias behrii*). *Mol. Ecol.* 21: 4242-4256. <http://dx.doi.org/10.1111/j.1365-294X.2012.05696.x>
- Sen SK, Jolly MS and Jammy TR (1971). Biology and life cycle of *Canthecona furcellata* Wolff (Hemiptera: Pentatomidae), predator of tasar silkworm *Antheraea mylitta* Drury. *Indian J. Seric.* 10: 53-56.

- Singh RN, Bajpayee CM, Jayaswal J and Thangavelu K (1992). Perspective of biological control in tasar culture. *Indian Silk* 31: 48-50.
- Singh YT, Mazumdar-Leighton S, Saikia M, Pant P, et al. (2012). Genetic variation within native populations of endemic silkmoth *Antheraea assamensis* (Helfer) from Northeast India indicates need for *in situ* conservation. *PLoS One* 7: e49972. <http://dx.doi.org/10.1371/journal.pone.0049972>
- Slatkin M (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139: 457-462.
- Stamatakis A (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688-2690. <http://dx.doi.org/10.1093/bioinformatics/btl446>
- Swofford DL (2001). PAUP* 4.0: phylogenetic analysis using parsimony (*and other methods). Version 4b2. Sinauer Associates, Sunderland.
- Takezaki N, Nei M and Tamura K (2014). POPTREEW: web version of POPTREE for constructing population trees from allele frequency data and computing some other quantities. *Mol. Biol. Evol.* 31: 1622-1624. <http://dx.doi.org/10.1093/molbev/msu093>
- Tautz D and Renz M (1984). Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res.* 12: 4127-4138. <http://dx.doi.org/10.1093/nar/12.10.4127>
- Wan X, Kim MJ and Kim I (2013). Description of new mitochondrial genomes (*Spodoptera litura*, Noctuoidea and *Cnaphalocrocis medinalis*, Pyraloidea) and phylogenetic reconstruction of Lepidoptera with the comment on optimization schemes. *Mol. Biol. Rep.* 40: 6333-6349. <http://dx.doi.org/10.1007/s11033-013-2748-3>
- Weir BS and Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358-1370. <http://dx.doi.org/10.2307/2408641>

Supplementary material

Table S1. A list of samples with sequence information of *Antheraea yamamai*.

Table S2. Pairwise comparisons between pairs of *Antheraea yamamai* haplotypes obtained from the concatenated sequences of *COI* + *ND4*.

Table S3. Fixation indices (*F_{ST}*) and migration rate (*N_m*) between pairs of *Antheraea yamamai* haplotypes obtained from the concatenated sequences of *COI* + *ND4*.

Table S4. Summary of Illumina Mi-Seq paired-end (2X250) read sequence data and *de novo* assembly of *Antheraea yamamai* genome.