Journal of Economic Entomology, 109(1), 2016, 366–374
doi: 10.1093/jee/tov258
Advance Access Publication Date: 13 September 2015
Research article

DNA Barcoding of Gypsy Moths From China (Lepidoptera: Erebidae) Reveals New Haplotypes and Divergence Patterns Within Gypsy Moth Subspecies

Fang Chen,[1] Youqing Luo,[1] Melody A. Keena,[2] Ying Wu,[1] Peng Wu,[3] and Juan Shi[1,4]

[1]Beijing Key Laboratory for Forest Pest Control, Forestry College, Beijing Forestry University, 35 Qinghua East Road, Beijing 100083, China (chen-fang866@163.com; youqingluo@126.com; 704658127@qq.com; shi_juan@263.net), [2]Northern Research Station, Forest Service, United States Department of Agriculture, 51 Mill Pond Road, Hamden, CT 06492 (mkeena@fs.fed.us), [3]Beijing Entry-Exit Inspection and Quarantine Bureau, 6 Tianshuiyuan Street, Beijing 100026, China (619049202@qq.com), and [4]Corresponding author, e-mail: shi_juan@263.net

Received 10 April 2015; Accepted 14 August 2015

Abstract

The gypsy moth from Asia (two subspecies) is considered a greater threat to North America than European gypsy moth, because of a broader host range and females being capable of flight. Variation within and among gypsy moths from China (nine locations), one of the native countries of Asian gypsy moth, were compared using DNA barcode sequences (658 bp of mtDNA cytochrome c oxidase subunit 1 [COI] sequence), together with two restriction site mtDNA markers (NIaIII and BamHI in COI), which is the standard system used to distinguish European gypsy moths from Asian gypsy moths. Relatedness of these populations to gypsy moths from seven other world areas was also examined. The restriction site markers showed that two Chinese populations had both Asian and European haplotypes. DNA barcode sequence divergence between the Asian populations and the European populations was three times greater than the variation within each group. Using Bayesian and parsimonious network analyses, nine previously unknown barcode haplotypes were documented from China and a single haplotype was found to be shared by 55% of the Chinese and some Far Eastern Russian and Japanese individuals. Some gypsy moths from two Chinese populations showed genetic affinity with mtDNA haplotypes from Siberia, Russia, suggesting there could be a cryptic new subspecies in *Lymantria dispar* (L.) or human-aided movement of moths between these two locations at an earlier point in time. The previously unknown haplotype patterns may complicate efforts to identify Asian gypsy moth introductions and require changes in monitoring and exclusion programs.

Key words: gypsy moth, COI, NB system, DNA barcoding, haplotype

The gypsy moth, *Lymantria dispar* (L.) (Lepidoptera: Erebidae), is a serious quarantine pest with a broad host range, including both deciduous and coniferous trees (Pogue and Schaefer 2007). Asian gypsy moth, including *Lymantria dispar asiatica* Vnukovskij, *Lymantria dispar japonica* Motschulsky, *Lymantria umbrosa* (Butler), *Lymantria albescens* Hori and Umeno, and *Lymantria postalba* Inoue, and the European gypsy moth, *Lymantria dispar dispar* L., are recognized on the basis of differences in their distribution and in the females' flight capabilities. European gypsy moth was accidentally introduced to the United States of America from France in 1869, where it has continued to disperse ever since (Forbush and Fernald 1986). Asian gypsy moth is primarily found in Asia, but females that are capable of flight also occur in parts of Europe (Keena et al. 2008). The *asiatica* and *japonica* subspecies have long been considered to pose more significant global threats, owing to their broader range of host species (Baranchikov and Montgomery 2009) and to the superior flight ability of females (Keena et al. 2008). Gypsy moths from different geographic origins exhibit different behavioral and biological traits that can influence their invasive ability (i.e., their ability to spread and establish in new areas). Despite these differences, morphological variation between the Asian gypsy moth and the European gypsy moth is minimal, making the identification of the geographic origins of populations difficult. Therefore, molecular markers based both on mitochondrial DNA and on nuclear DNA have been used to identify the geographic origins of gypsy moth populations (Pfeifer et al. 1995; Garner and Slavicek 1996; Bogdanowicz et al. 1997, 2000; Schreiber et al. 1997; Reineke and Zebitz 1999; Koshio et al. 2002).

Increasing international trade and tourism in recent years has facilitated gypsy moths' invasion into areas where their influence on

Table 1. Details of the gypsy moth populations analyzed in this study (ordered by longitude from east to west)

Population	Country	Collection site	No. of individuals in analysis[a]	Latitude	Longitude	Collection date[b]
JN	Japan	Nagoya, Honshu	11 (7)	35° 8′59″ N	137° 4′48″ E	Mar., 1996 (Nov., 2013)
JT	Japan	Hachioji-shi, Tokyo	13 (10)	35° 37′48″ N	139° 13′48″ E	Aug., 2011 (May, 2013)
RM	Russia	Mineralni, Primorski	14 (9)	44° 6′6″ N	133° 9′2″ E	Aug., 1992 (Nov., 2013)
CR	China	Harbin, Heilongjiang	13 (7)	45° 46′45″ N	126° 36′38″ E	Aug., 2012
CL	China	Sandeli, Liaoning	24 (6)	41° 30′33″ N	122° 22′27″ E	Sept., 2011
CX	China	Arxan, Inner Mongolia	10 (7)	47° 10′26″ N	119° 56′44″ E	June, 2011
CS	China	Lianyungang, Jiangsu	29 (10)	34° 44′16″ N	119° 22′ 45″ E	July 2011
CH	China	Longhua, Hebei	24 (10)	41° 37′8″ N	118° 14′49″ E	Mar., 2012
CB	China	Yanzikou, Beijng	30 (9)	40° 19′21″ N	116° 09′11″ E	Aug., 2011
CA	China	Liu'an, Anhui	30 (9)	31° 30′34″ N	115° 53′39″ E	Aug., 2011
CJ	China	Jining, Inner Mongolia	9 (6)	40° 36′47″ N	112° 07′30″ E	Aug., 2011
CG	China	Xifeng, Guizhou	20 (10)	27° 11′40″ N	106° 49′53″ E	Nov., 2011
RBI	Russia	Bellk, Krasnoyarsk	15 (10)	54° 17′59″ N	91° 10′48″ E	Dec., 1992 (Nov., 2013)
GK	Greece	Kavala, Macedonia	13 (9)	41° 0′0″ N	24° 15′0″ E	Feb., 1997 (Nov., 2013)
LJ	Lithuania	Juodkrante, Kuzsin Nezijos	16 (5)	55° 18′36″ N	21° 3′36″ E	Aug., 1994 (Nov., 2013)
UC	United States	Bethany, New Haven County	17 (10)	41° 15′0″ N	73° 0′0″ W	Mar., 1994 (Nov., 2013)

[a] Number outside of parentheses is the amount of samples collected and used in "NB system" analysis, and number in the parentheses is the amount of COI sequences from each population used in sequence analysis.

[b] Date in parentheses is when the individuals were reared in laboratory for this study.

the economic benefits and ecological stability of the forest ecosystems are difficult to predict (deWaard et al. 2010). Since 2008, the North America Plant Protection Organization (NAPPO) has adopted specific phytosanitary measurements for ships and their cargo that come from or have visited areas infested with Asian gypsy moth to prevent its introduction (NAPPO 2009). This regulation has drawn worldwide attention to Asian gypsy moth which has females flying to lights in ports or industrial areas that are adjacent to forests containing their host trees. Because China is in the native range of Asian gypsy moth, it initiated a port inspection program to minimize the risk of Asian gypsy moth introduction to Asian-gypsy-moth-free countries and incursion of Asian gypsy moth from other native areas, such as Russia and Japan. In the past, there has been limited investigation of both the genetic variation within gypsy moths from China and between gypsy moths from China and those from other world areas.

An mtDNA haplotype identification system developed by Bogdanowicz is based on the presence of cytochrome c oxidase subunit 1 (COI) sequence at two restriction sites, BamHI and NlaIII, which is a simple molecular method that has become standard for identifying the origins of gypsy moth populations (Bogdanowicz et al. 1993, Keena et al. 2008). Previous studies showed that the presence of the COI sequence required for the DNA to be cut by the BamHI (B+) only occurs in Asian gypsy moth from Far East Asia, with moth populations from Siberia, Europe, and the United States lacking the appropriate COI sequence at BamHI (B−). The NlaIII restriction site is more specific, with the absence of cut sequence at NlaIII (N−) only occurring in North America, while all populations from Far East Asia, Siberia, and Europe have the corresponding cut site (N+). On the basis of these differences, an analysis based on COI sequence divergence was developed to verify the existence of the European gypsy moth lineage and to investigate if the divergence between the European and Asian lineages is relatively ancient (Harrison and Odell 1989).

DNA barcoding based on sequence diversity of standardized gene regions provides an alternative approach for invasive species detection and discovery of new ones (Hajibabaei 2006). In recent years, the development of DNA barcoding has provided a more powerful molecular tool for understanding divergence patterns and relationships among many taxa (Armstrong and Ball 2005, Floyd et al. 2010), particularly Lepidopteran species (Ball and Armstrong 2006; Hajibabaei 2006; Hebert et al. 2003a, 2009). A previous study has utilized DNA barcoding techniques to detect the geographic origins of gypsy moths globally, with very high reliability and resolution (deWaard et al. 2010). In this study, therefore, DNA barcoding techniques together with the standard "NB system" were utilized to identify the haplotypes present in gypsy moth populations in China, and to compare them with those found in gypsy moths from other world areas, which is an important information for biosecurity management.

Materials and Methods

Sample Collection and Rearing

We analyzed a total of 288 specimens of *L. dispar* sourced from 16 sites: nine sites in China, two sites in Japan, two sites in Russia, two sites in Europe, and one site in the United States (Table 1; Fig. 1). Permission for native sample collection was authorized by Forest Protection Station of National Forestry Bureau of China, and the collections at each location were overseen by staff of the local forestry bureau. The nonnative populations were imported from the United States Department of Agriculture Forest Service, Northern Research Station, and authorized by the General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, under the requirement that the moths will only be used in scientific research and not be released.

Nine adult moths (6 male, 3 female) from Jining, Inner Mongolia, were captured by hand in the field with the permission of the Forestry Bureau of Inner Mongolia (no endangered or protected species were involved). All the other moths were brought into the laboratory as egg masses and reared to mature larvae for this work. Prior to use, larvae were starved for 24 h, and then frozen at −80°C.

DNA Extraction

DNA was extracted from ~30 to 50 mg of body tissue from each individual with the Insect gDNA Miniprep Kit (BIOMIGA, Beijing,

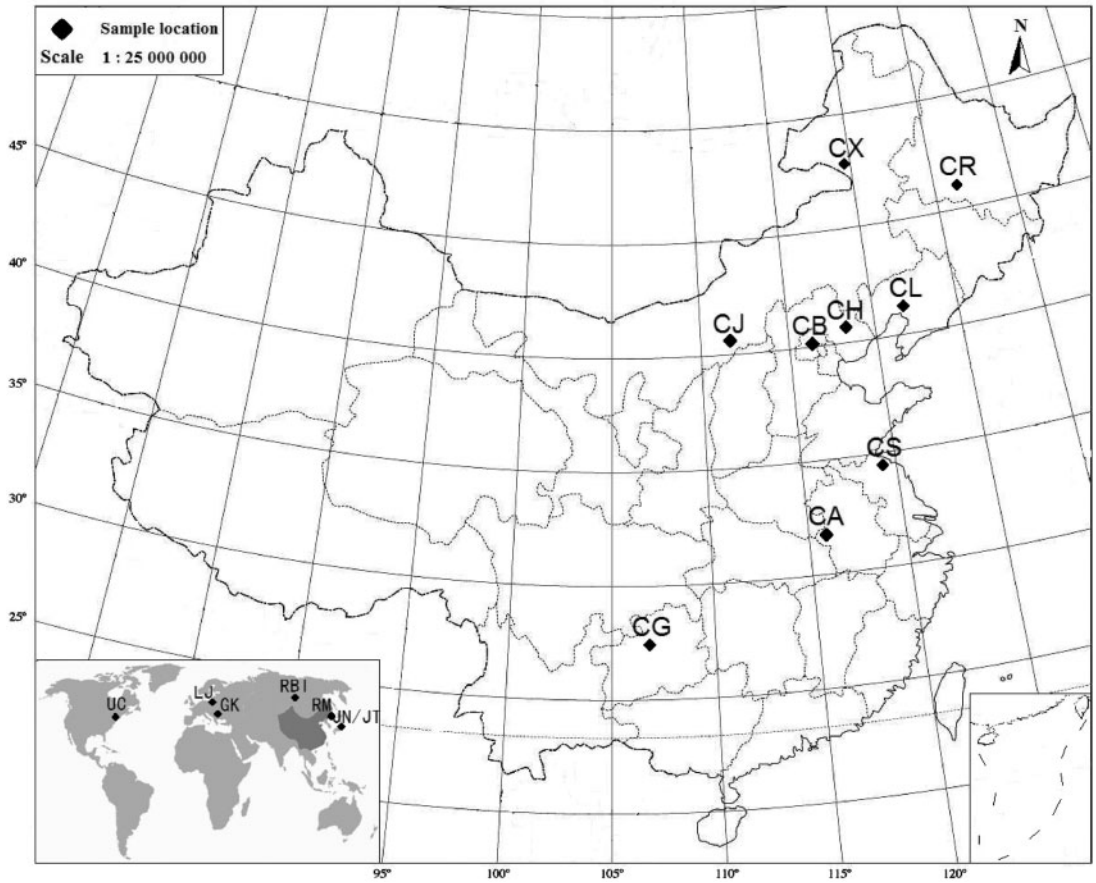

Fig. 1 Location of *L. dispar* sampling sites. Population labels follow those in Table 1.

China) according to the manufacturer's instructions. The concentration and purity of the extracted genomic DNA (gDNA) was determined with a Thermo Scientific NanoDrop 2000 Spectrophotometer V1.0 (Gene Company Limited, Hong Kong, China). A260/A280 values between 1.8 and 2.0 were considered to be of sufficient quality. Extracted DNA (50 ng/μl) was stored at −20°C.

Detection of the Geographic Origins by Restriction Site Analysis: "NB System"

All individuals in Table 1 were included in this analysis. The COI sequence was amplified with primers A2191 (5′ CCC GGT AAA ATT AAA ATA TAA ACT TC3′) and S1859 (5′ GGA ICI GGA TGA ACW GTT TAY CCI CC3′) (Bogdanowicz et al. 1993). All PCR amplifications were carried out in a 25 μl volume, containing 75 ng of template DNA, 2.5 μl 10× PCR buffer (TaKaRa Bio, Beijing, China), 0.2 μM of A2191 (SBS Genetech Co., Ltd, Beijing, China), 0.7 μM of S1859 (SBS Genetech Co., Ltd), 200 μM of dNTP (TaKaRa Bio), 3.0 mM of $MgCl_2$ (TaKaRa Bio), 0.6 U Taq DNA Polymerase (TaKaRa Bio), and ddH_2O to the final volume. Replicates and negative controls were included to ensure the consistency of results. The amplification reaction was performed in a DNA Engine Thermal Cycler PTC-200 (BIO-RAD, Shanghai, China), with the following cycle conditions: 95°C for 1 min, 35 cycles of 95°C for 15 s, annealing at 55°C for 30 s, and 72°C for 90 s, and a final extension time of 5 min at 72°C.

A 4 μl volume of the PCR product was incubated at 37°C with 2.0 U of BamHI (NEW ENGLAND BioLabs, Beijing, China) or 5.0 U of NlaIII (NEW ENGLAND BioLabs) in a reaction volume of 20 μl, for 3 h for BamHI or 1.5 h for NlaIII. Following this, 5 μl of each PCR product (digested with either BamHI or NlaIII)was added to 1 μl of 6 × loading buffer (TaKaRa Bio) and loaded in a 1.5% agarose gel (Ffgene CO., LTD. Spain) for electrophoresis. After running at 110 V for 30 min, the gel was stained in an ethidium bromide buffer for 3 min, rinsed with distilled water, and photographed in a gel documentation machine (Gene Company Limited, Hong Kong, China). A DL 2000 marker (Takara Bio) was run as a standard alongside samples.

Detection of the Geographic Origins and Divergence Patterns by DNA Barcoding

Ten representative individuals from each population, except nine from CJ, totaling 159 samples, were sequenced for the 658-bp DNA barcode region used in previous gypsy moth work (deWaard et al. 2010). The 5′ end of the mitochondrial COI gene was amplified with primers: Lep-F1 (5′ ATT CAA CCA ATC ATA AAA GAT AT 3′) and Lep-R1 (TAA ACT TCT GGA TGT CCA AAA A3′) using standard PCR amplification conditions (Hebert, et al. 2004). PCR products were visualized on a 1% agarose gel and cycle sequenced with the Lep-F1 primer. Sequences were obtained by ABI 3730xl DNA Analyzer with ABI BigDye terminator v3.1 Cycle Sequencing Kit (Applied Biosystem). Sequences were edited to remove ambiguous bases and primer sequences using Chromas 1.62 (McCarthy 1996).

To ensure that all available COI sequences of gypsy moths from China were included in the analyses, nine sequences collected in 1992 originating from China were obtained from GenBank

(GenBank HM775692, HM775691, HM775690, HM775689, HM775688, HM775687, HM775685, HM775684, HM775686). To evaluate how the current work compares with those already published, another nine sequences from prior studies were downloaded from GenBank: two from North America (GenBank GU091222, HM775564), four from Europe (GenBank HM775655, HM775600, HM775659, and HM775587) and three from Asia outside of China (GenBank HM775604, HM775582, and KF746253). COI sequences of *L. albescens*, *L. obfuscate*, and *L. umbrosa* were downloaded from the GenBank and used as outgroups (GenBank KF746255, HM775825, and HM775854, respectively). All the sequences were aligned by using ClustalW (Thompson et al. 1994) in the software BioEdit (Hall 1999). Sequence divergence was calculated using Kimura 2-parameter (K2P) to allow comparison of our data with other DNA barcoding studies which commonly use this distance model.

Phylogenetic trees were constructed using Bayesian analysis (MrBayes 3.2.3; Ronquist and Huelsenbeck 2003). The evolutionary model chosen based on the best fit to the data was a Tamura-3-parameter substitution model (Tamura 1992) with Gamma distributed rate variation across sites after running a model test in MEGA 6 (Tamura et al. 2013). Bayesian analysis used the default priors and consisted of two concurrent runs, with four chains, each with 5,000,000 generations, sampled every 1,000 generations. The consensus trees were obtained after a burnin of the first 2% of sampled trees. To construct the tree showing the genetic relationships among Chinese samples, 83 COI sequences originating from China were included (74 individuals from this study and nine from GenBank), while the tree comparing worldwide samples included 152 COI sequences (134 from this study and 18 from GenBank).

To visualize the connections and number of mutational differences between haplotypes, a network based on the statistical parsimony method of Templeton (Templeton et al. 1992) was constructed, using the software TCS 1.21 (Clement et al. 2000) with a 90% relative confidence limit. This method has been extensively applied to nucleotide sequence data to infer population-level genealogies when divergences are low (Gerber et al. 1996, Gómez-Zurita et al. 2000).

Results

Using the "NB system" we identified two haplotypes among all the Chinese gypsy moths sampled. The majority of individuals (87%) exhibited the N+B+ mtDNA haplotype, characteristic of gypsy moths from Far East Asia. However, all individuals from Guizhou (CG, 20 moths) and additional four from Anhui (CA) and one from Inner Mongolia (CX) exhibited the N+B- haplotype (Table 2). The N+B- haplotype was previously found only in European gypsy moth individuals from Europe or Asian gypsy moth individuals from Siberian Russia (Bogdanowicz et al. 1993, Keena et al. 2008).

One hundred thirty-four COI barcode sequences representing 16 populations from six countries were used in the sequence analysis (Table 1). An average of 8.3 sequences per population was obtained (range = 5 to 10 sequences per population). The final COI sequence fragment was 658 bp after the sequences were aligned and primer sequences were deleted. There was no evidence of insertions or deletions in the DNA sequences, nor were stop codons found when translating amino acids, suggesting that pseudogenes were absent.

The nucleotide frequencies were as follows: T, 39.4; C, 15.9; A, 30.0; and G, 14.7%. The GC content of the COI gene was low (30%) and similar in value to that obtained in molecular studies of other Lepidopteran species (31%) (Hebert et al. 2003a). In particular, GC frequency was extremely low at the third codon position (11%), but much higher at the first and second codon positions (40%). The overall average pairwise nucleotide sequence divergence using K2P was 0.005. COI sequence divergence ranged from 0 to 1.5% (mean = 0.3%) within the populations of Asian origin, while among populations from Europe and North America, sequence divergence ranged from 0 to 0.8% (mean = 0.3%). COI sequence divergence between the Asian populations and the European populations ranged from 0.2 to 1.7% (mean = 0.9%). Twenty-nine substitutions were detected in the 658 bp COI fragment across the 143 individuals (134 from this study and nine from GenBank; Table 3), of which 89% were transitions and 11% were transversions.

Table 2. Results of the "NB system" analysis showing MtDNA COI gene haplotypes for the *L. dispar* samples

Population	No. of individuals in each haplotype based on MtDNA COI "NB system"		
	N+B+	N+B-	N-B-
JN	11		
JT	13		
RM	14		
CR	13		
CL	24		
CX	9	1	
CS	29		
CH	24		
CB	30		
CA	26	4	
CJ	9		
CG		20	
RBI		15	
GK		13	
LJ		16	
UC			17

The consensus tree constructed in MrBayes using the Chinese samples showed 14 haplotypes in gypsy moth from China, 12 from this study, and two additional from GenBank (Fig. 2). Most of the individuals, despite being from sites that are geographically distant or being collected more than two decades apart, shared a common haplotype. None of the individuals identified as the European subspecies by the "NB system" had the dominant haplotype. Additionally, one population from Guizhou (CG), in southern China, together with one individual from Liaoning (CL), in northeastern China, formed a separate unique cluster.

The consensus tree based on COI sequences of worldwide gypsy moths indicated delineation between most of the European and Asian gypsy moth groups. Most of the moths originating from Asia clustered into one group with eight populations from China, two populations from Japan, and one population from Russia. A second cluster contained the European gypsy moth populations from United States, Greece, Lithuania and the Asian gypsy moth population from Siberian Russia (RBI; Fig. 3). The RBI population is from the east of the Ural Mountains and the females are capable of full flight, so it has been classified as Asian gypsy moth; however, both the "NB" haplotype and the barcoding results show it shares the European gypsy moth haplotype as has been previously documented (Keena et al. 2008). A third cluster was formed by the same moths from Guizhou and Liaoning, China, which had formed a unique cluster when only the Chinese individuals were used in the analysis. Further, the four individuals from Anhui, China, which were classified as the European subspecies by the "NB system," clustered as an independent branch in the Asian group.

Table 3. Variable sites of mtDNA COI haplotypes in *L. dispar*, 134 individuals from this study, and 9 COI sequences from GenBank of gypsy moth from China were included

H	Subspecies included	*N*	40	50	59	104	196	202	208	217	223	225	235	238	268	300	343	368	410	506	513	526	552	560	574	622	625	628	631	641	646
1	*L. dispar asiatica /japonica*	61(5)	A	C	C	G	G	A	A	A	T	A	C	C	A	G	G	G	G	G	G	A	C	G	A	A	A	A	G	C	C
2	*L. dispar japonica*	1	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	T	•	•	•	•	•	•
3	*L. dispar asiatica*	1	•	•	•	•	•	•	G	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	A	•	•
4	*L. dispar asiatica*	4	•	•	•	•	A	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
5	*L. dispar asiatica*	1	•	•	A	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
6	*L. dispar asiatica*	1 (1)	G	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
7	*L. dispar asiatica*	1	•	•	•	•	•	•	G	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
8	*L. dispar asiatica*	1	•	•	A	•	•	•	•	•	•	•	•	T	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
9	*L. dispar asiatica*	9	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	G	•	•	•	•
10	*L. dispar asiatica*	4	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	T
11	*L. dispar asiatica*	1	•	•	•	•	•	•	•	•	•	•	•	•	•	A	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
12	*L. dispar asiatica*	1	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	G	•	•	•	•	•
13	*L. dispar asiatica*	1	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	C	•	•	•	•	•	•	•	•	•	•	•	•
14	*L. dispar asiatica*	1	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	C	•	A	G	•	•	•	•	•	•	•	T	•
15	*L. dispar asiatica*	1	•	•	•	•	•	•	•	•	•	•	•	•	•	A	•	•	•	•	•	•	•	A	•	•	•	•	•	•	•
16	*L. dispar asiatica*	(1)	•	•	•	•	•	•	•	•	•	•	T	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
17	*L. dispar asiatica*	(2)	•	•	•	•	•	•	•	•	•	•	•	•	G	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
18	*L. dispar asiatica*	1	G	•	•	•	A	•	•	G	A	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
19	*L. dispar asiatica*	1	G	•	•	•	A	•	•	G	•	T	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
20	*L. dispar asiatica*	9	G	•	•	•	A	•	•	G	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	A	•	•
21	*L. dispar asiatica*	5	G	•	•	•	A	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	A	•	T
22	*L. dispar asiatica*	5	G	•	•	A	A	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	A	•	T
23	*L. dispar dispar*	4	G	•	•	•	A	•	•	•	•	•	•	•	•	•	•	A	•	A	•	•	•	•	•	•	•	G	A	•	T
24	*L. dispar dispar*	6	G	•	•	•	A	•	•	•	•	•	•	•	•	•	•	A	•	•	•	•	•	•	•	•	•	G	A	•	T
25	*L. dispar dispar*	5	G	T	•	•	A	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	G	A	•	T
26	*L. dispar dispar*	3	G	•	•	•	A	G	•	•	•	•	•	•	•	•	C	•	•	•	•	•	•	•	•	•	•	•	A	•	T
27	*L. dispar dispar*	5	G	•	•	•	A	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	G	A	•	T
28	*L. dispar dispar*	1	G	•	•	•	A	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	A	•	•	•	•	G	A	•	T

H represents haplotype, *N* defines the number of individuals in this analysis belonging to that haplotype, and numbers of the 9 sequences from GenBank are showed in the parentheses. Digital numbers above indicate the nucleotide position corresponding to the 658 bp COI sequence. Nucleotides identical to the first haplotype are indicated by a dot.

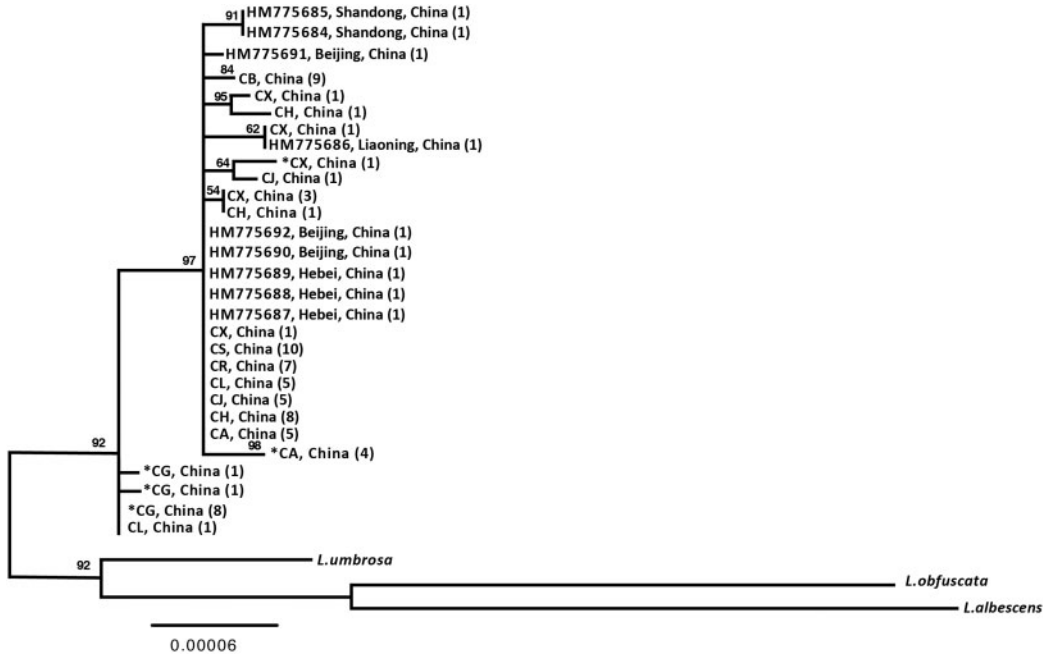

Fig. 2 Consensus tree of barcodes for gypsy moths from China using Bayesian analysis. Number of individuals from the same population that shared the same haplotype is shown the parentheses. The support values for clades are shown above the branches. Asterisks indicate individuals identified as European gypsy moth with the "NB system." Eighty-three COI sequences originating from China were included (74 individuals from this study and nine from GenBank). *L. albescens*, *L. obfuscate*, and *L. umbrosa* were included as outgroups. Population labels follow those in Table 1.

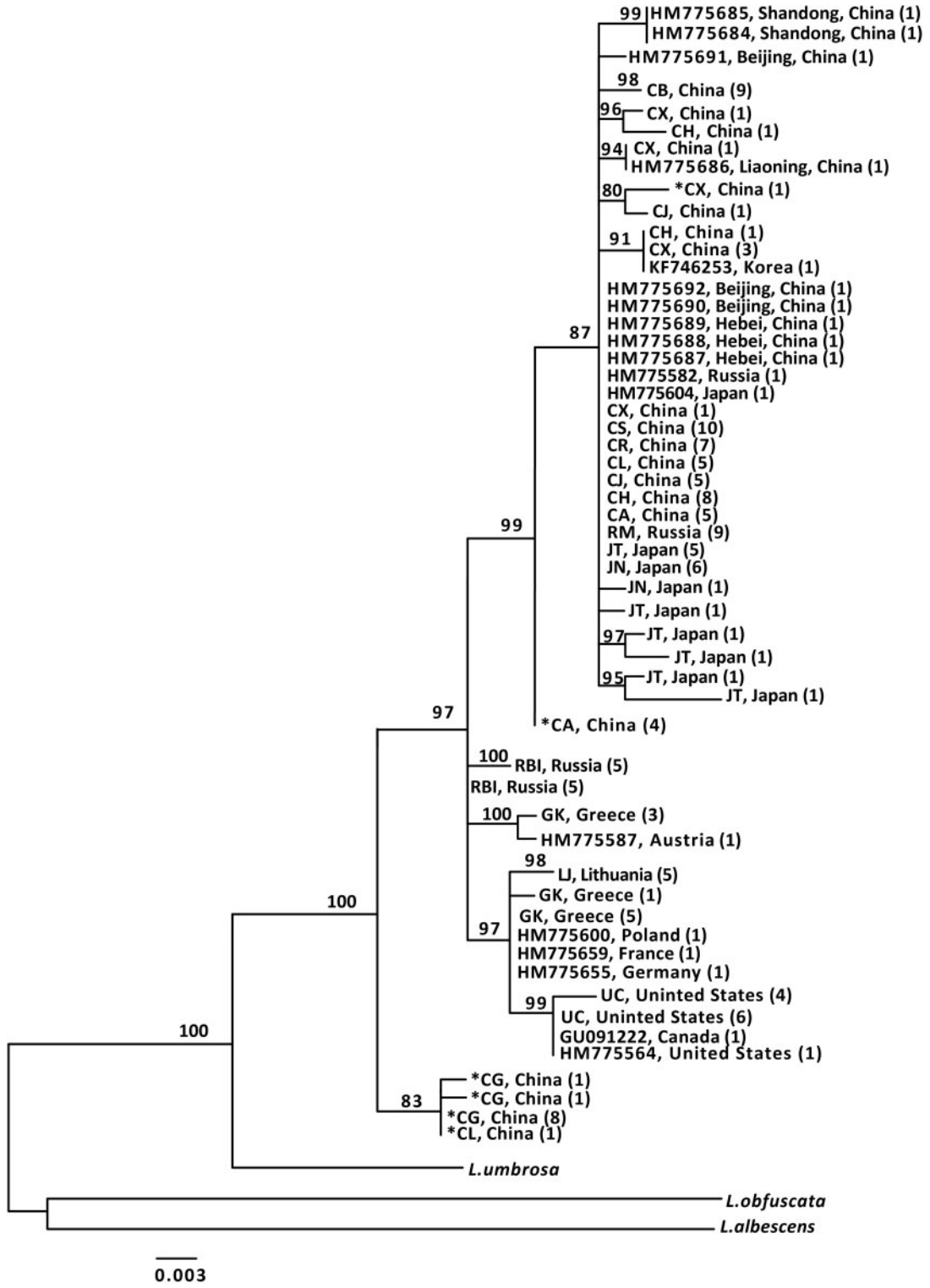

Fig. 3 Consensus tree of barcodes for worldwide gypsy moths using Bayesian analysis. Number of individuals from the same population that shared the same haplotype is shown the parentheses. The support values for clades are shown above the branches. Asterisks indicated the individuals who were identified as European gypsy moth with the "NB system." One hundred fifty-two COI sequences were included (134 from this study and 18 from GenBank). *L. albescens, L. obfuscate,* and *L. umbrosa* information were included as outgroups. Population labels follow those in Table 1.

A parsimonious network of the 28 haplotypes was constructed using TCS 1.21 (Fig. 4). The network contains two groupings of haplotypes: one of Asian origin and one that is primarily made up of moths of European or United States origin. The Asian group contained 17 unique haplotypes separated by one to six mutations and consisted of individuals from China, the Far East of Russia, and Japan. The European group included samples from China, Siberian Russia, Greece, Lithuania, and the United States, which include 11 unique haplotypes differing from each other by one to seven mutations. A haplotype from Anhui, China, served as the intermediate

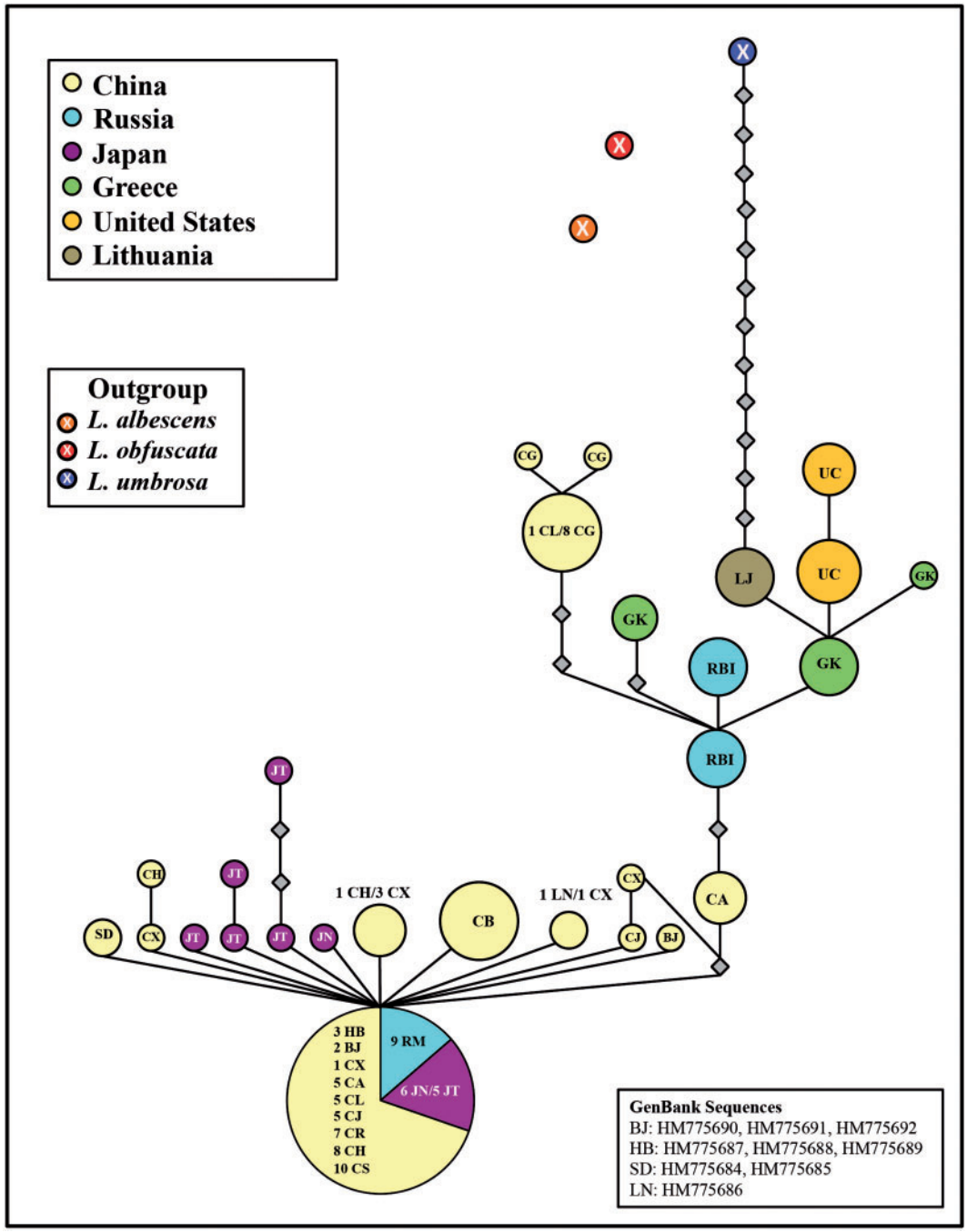

Fig. 4 Parsimonious joining network (TCS 1.21) based on a relative connection level of 90%. One hundred thirty-four COI sequences from this study and nine downloaded from GenBank originating from China were included. The size of each circle is proportional to the frequency of the haplotype. Gray diamonds define mutation steps separating the observed haplotypes and different colors characterize the geographic origin of the sample. Population labels follow those in Table 1 or are assigned to GenBank sequences (see inset).

haplotype between the Asian and European groups. Samples from two Chinese populations, CL and CG that had formed a separate cluster in the Bayesian trees, were connected to a haplotype from RBI in the European group. One of the outgroup species, *L. umbrosa*, was distantly connected (13 mutations) to the network through a haplotype from gypsy moths originating in Lithuania.

Discussion

The observed genetic relationships based on COI sequences revealed previously unknown haplotypes and new clustering patterns. Compared to all the public COI data in GenBank by the time this manuscript was written, nine previously unknown haplotypes were discovered, making this the most comprehensive sampling of gypsy moth from China to date. Thus, there is considerable haplotype variation in Chinese gypsy moths and there maybe even more haplotypes that exist, but just have not been sampled. Among all samples collected from a broad geographic range in China, ~55% individuals shared a single haplotype, which was also found in samples collected from Japan and Far Eastern Russia. Novel haplotypes were found for individuals from seven of the nine populations, indicating

great genetic variability exists among gypsy moths from China. Additionally, the population from Beijing, China (CB), is the only one with all samples sharing one haplotype, and this haplotype is different from the two previously reported haplotypes found in samples collected in Beijing in 1992. Furthermore, when querying GenBank with our new sequences, one haplotype found in three individuals from Inner Mongolia (CX) and one individual from Hebei (CH), matched a haplotype from Korea reported in January 2014 (Fig. 3).

Gypsy moths from Siberia (RBI) were classified as Asian gypsy moth not only because that it is from east of the Ural Mountains but also its females have strong flight capability (Keena et al. 2008). However, RBI had a European mtDNA haplotype and grouped with gypsy moths from Greece, Lithuania, and the United States in the network analysis, suggesting the Ural Mountains is not an appropriate geographic border for separating Asian gypsy moth from European gypsy moth. In addition, all the individuals from the Guizhou (CG) and one individual from Liaoning (CL) separated themselves from the major Asian gypsy moth haplotypes, forming a unique cluster connected to Siberia (RBI) in the European cluster. This also re-emphasizes the concern raised previously that geographic populations with and without female flight cannot be reliably distinguished using these mtDNA markers (Pogue and Schaefer 2007, Keena et al. 2008).

The average K2P distance within each subspecies, *asiatica* and *dispar*, in this study was 0.3% which is close to the value previously reported for sequence divergences within Lepidopteran families (0.25%; Hebert et al. 2003b). The K2P value between subspecies, however, was much greater (0.9%) than that observed within subspecies. This result reveals the degree of genetic divergence that exists between the Asian gypsy moth and the European gypsy moth, and validates the general use of COI sequences to distinguish the two subspecies.

One advantage of the "NB system" is that it targets a shorter COI gene sequence and thus, provides a rapid initial assessment on the geographic origins of the gypsy moth. However, because of the limited resolution of the two restriction sites, this method has a high probability of misclassification. Indeed, five samples in this study exhibited the N+B- haplotype of *L. dispar dispar* but, in fact, clustered most closely with the *L. dispar asiatica* when the full barcode was used. Similarly, one individual from Liaoning (CL) was classified as Asian gypsy moth by the "NB system" (exhibiting an N+B+ haplotype), but clustered with Guizhou (CG) in a separate cluster from the rest of the Asian gypsy moths. These results confirm that the COI barcode approach offers greater accuracy and reliability for classifying gypsy moth to subspecies than the "NB system" (Rivera and Currie 2009, Derocles et al. 2012).

The COI sequence analysis based on both MrBayes and TCS strongly suggest that gypsy moths from Guizhou (CG) and one from Liaoning (CL) have genetic affinity with mtDNA haplotypes from Siberia, Russia (RBI), and could even be a cryptic new subspecies in *Lymantria dispar*. Female gypsy moths from Siberia are known to be capable of strong flight (even over mountains) (Mikkola 1971). Based on previous work on female flight behavior of gypsy moths from China, from the same populations using flight-mills (Yang 2013), the Guizhou (CG) and Liaoning (CL) populations showed the strongest flight potential, both in duration and distance; 1-d-old females could fly up to 7.5 km and 6.63 km for CG and CL, respectively. However, natural dispersal between Siberia and the two locations in China where a similar haplotype was found is unlikely due to sheer distance alone (~3, 000 km between the locations). In addition, Guizhou (CG) is unlikely to exchange genetic material with other populations, because it is geographically isolated. Guizhou is located in the eastern part of the Yungui Plateau, an area averaging about 1,100 m in altitude; deep valleys and rugged plateaus constitute about 87% of the total area. These geographic features could form a natural barrier to prevent gene flow between gypsy moths in this area and those in other areas. Therefore, human aided movement likely occurred at an earlier point in time since the Chinese haplotypes are three or four mutations distant from the Siberian ones. The direction of that movement, however, cannot be decisively determined with the information available and should be addressed in future investigations.

In contrast to deWaard (deWaard et al. 2010), our study using these 16 populations of gypsy moths was unable to distinguishing gypsy moths subspecies based on DNA barcode clusters. Given the high genetic variability in Asian gypsy moth, some nuclear genetic markers may be needed to provide the needed high-resolution for identification between gypsy moth subspecies and populations with female flight capability. For quarantine purposes, more samples from a wider geographic distribution are needed to work out a comprehensive map of the haplotypes of gypsy moth worldwide. The recognition of the haplotypes is necessary to distinguish and prevent the introduction of genotypes to areas where they are not already present. An incursion of a new haplotype to a novel habitat would have unknown consequences some of which could be negative. These new findings of greater haplotype variation in China and the genetic affinity of some of gypsy moth from China to those in Siberia may complicate efforts to identify the sources of introductions of gypsy moths, and require changes in monitoring and exclusion programs.

Data Accessibility

The sequences reported from this publication have been submitted to the Barcode of Life Data Systems database (http://www.boldsystems.org/) and assigned in the project "Gypsymoth-BJFU" (project code GMCF).

Acknowledgments

We are grateful to J.Y. Yu (Guizhou Academy of Forestry), Z.C. Xu (Beijing Forestry University), and M. Takeda (Kobe University) for their kind help in sample collection. We also wish to thank Y. Wu, H.J. Yu, P. Shi, and Y. Li for their participation in rearing moths in the lab and P.F. Li and N. Havill for their assistance with genetic analysis. This study was funded by Beijing Higher Education Young Elite Teacher Project (YETP0740), the National Natural Science Foundation of China (31170613), the Fundamental Research for the Central Universities (BLYJ201306), and the China Scholarship Council (CSC).

References Cited

Armstrong, K., and S. Ball. 2005. DNA barcodes for biosecurity: invasive species identification. Philos. Trans. R. Soc. Biol. Sci. 360: 1813–1823.

Ball, S. L., and K. F. Armstrong. 2006. DNA barcodes for insect pest identification: a test case with tussock moths (Lepidoptera: Lymantriidae). Can. J. For. Res. 36: 337–350.

Baranchikov, Y., and M. Montgomery. 2009. Comparison of the polyphagy of gypsy moths from different contients. Trans. Saint Petersburg For. Eng. Acad. 183: 40–46.

Bogdanowicz, S., W. Wallner, J. Bell, T. Odell, and R. Harrison. 1993. Asian gypsy moths (Lepidoptera: Lymantriidae) in North America: evidence from molecular data. Ann. Entomol. Soc. Am. 86: 710–715.

Bogdanowicz, S., V. Mastro, D. Prasher, and R. Harrison. 1997. Microsatellite DNA variation among Asian and North American gypsy moths (Lepidoptera: Lymantriidae). Ann. Entomol. Soc. Am. 90: 768–775.

Bogdanowicz, S., P. Schaefer, and R. Harrison. 2000. Mitochondrial DNA variation among worldwide populations of gypsy moths, *Lymantria dispar*. Mol. Phylogenet. Evol. 15: 487–495.

Clement, M., D. Posada, and K. A. Crandall. 2000. TCS: a computer program to estimate gene genealogy. Mol. Ecol. 9: 1657–1659.

Derocles, S. A., A. LE Ralec, M. Plantegenest, B. Chaubet, C. Cruaud, A. Cruaud, and J. Y. Rasplus. 2012. Identification of molecular markers for DNA barcoding in the Aphidiinae (Hym. Braconidae). Mol. Ecol. Resour. 12: 197–208.

deWaard, J. R., A. Mitchell, M. A. Keena, D. Gopurenko, L. M. Boykin, K. F. Armstrong, M. G. Pogue, J. Lima, R. Floyd, R. H. Hanner, et al. 2010. Towards a Global Barcode Library for Lymantria (Lepidoptera: Lymantriinae) Tussock Moths of Biosecurity Concern. PLoS ONE 5: e14280.

Floyd, R., J. Lima, J. deWaard, L. Humble, and R. Hanner. 2010. Common goals: policy implications of DNA barcoding as a protocol for identification of arthropod pests. Biol. Invasions. 12: 2947–2954.

Forbush, E. H., and C. H. Fernald. 1896. The gypsy moth, p. 495. *Porthetria dispar* (Linn.). Wright & Potter, Boston.

Garner, K., and J. Slavicek. 1996. Identification and characterization of a RAPD-PCR marker for distinguishing Asian and North American gypsy moths. Insect. Mol. Biol. 5: 81–91.

Gerber, A. S., and A. R. Templeton. 1996. Population sized and within-deme movement of *Trimerotropis saxatilis* (Acrididae), a grasshopper with a fragmented distribution. Oecologia 105: 343–350.

Gómez-Zurita, J., E. Petitpierre, and C. Juan. 2000. Nested cladistic analysis, phylogeography and speciation in the *Timarcha goettingensis* complex (Coleoptera, Chrysomelidae). Mol. Ecol. 9: 557–570.

Hajibabaei, M. 2006. DNA barcodes distinguish species of tropical Lepidoptera. P. N. A. S. 103: 968–971.

Hall, T. A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic acids symposium series. pp. 95–98.

Harrison, R. G., and T. M. ODell. 1989. Mitochondrial DNA as a tracer of gypsy moth origins, pp. 265–273 *In* W. E. Wallner and K. A. McManus (Eds.), Proceedings, Lymantriidae: A comparison of features of New and Old World tussock moths. US Department of Agriculture, General Technical Report NE123.

Hebert, P.D.N. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. P. N. A. S. 101: 14812–14817.

Hebert, P.D.N., A. Cywinska, and S. L. Ball. 2003a. Biological identifications through DNA barcodes. Proc. R. Soc. Lond. B. Biol. 270: 313–321.

Hebert, P.D.N., S. Ratnasingham, and J. R. deWaard. 2003b. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc. R. Soc. Lond. B. Biol. 270: S96–S99.

Hebert, P.D.N., J. R. deWaard, and J. F. Landry. 2009. DNA barcodes for 1/1000 of the animal kingdom. Biol. Lett. 6: 359–362.

Keena, M. A., M. J. Côté, P. S. Grinberg, and W. E. Wallner. 2008. World distribution of female flight and genetic variation in *Lymantria dispar* (Lepidoptera: Lymantriidae). Environ. Entomol. 37: 636–649.

Koshio, C., M. Tomishima, K. Shimizu, H.-S. Kim, and O. Takenaka. 2002. Microsatellites in the gypsy moth, *Lymantria dispar* L. (Lepidoptera: Lymantriidae). Appl. Entomol. and Zool. 37: 309–312.

McCarthy, C. 1996. Chromas version 1.45. School of Health science, Griffifth University, Gold Coast Campus, Queensland, Australia.

Mikkola, K. 1971. Migratory habit of *Lymantria dispar* (Lepidoptera, Lymantriidae) adults of continental Eurasia in the light of a flight to Finland. Acta. Entomol. Fenn. 28: 107–120.

(NAPPO) North America Plant Protection Organization. 2009. Guidelines for Regulating the Movement of Ships and Cargo from Areas Infested with the Asian Gypsy Moth. Regional Standard for Phytosanitary Measures No. 33, NAPPO, Canada.

Pfeifer, T. A., L. M. Humble, M. Ring, and T. A. Grigliatti. 1995. Characterization of gypsy moth populations and related species using a nuclear DNA marker. Can. Entomol. 127: 49–58.

Pogue, M. G., and P. W. Schaefer. 2007. A review of selected species of Lymantria Hübner [1819] including three new species (Lepidoptera: Noctuidae: Lymantriinae). USDA Forest Health Technology Enterprise Team. FHTET-2006-07.

Reineke, A., and C. P. Zebitz. 1999. Suitability of polymerase chain reaction-based approaches for identification of different gypsy moth (Lepidoptera: Lymantriidae) genotypes in central Europe. Ann. Entomol. Soc. Am. 92: 737–741.

Rivera, J., and D. C. Currie. 2009. Identification of Nearctic black flies using DNA barcodes (Diptera: Simuliidae). Mol. Ecol. Resour. 9: 224–236.

Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574.

Schreiber, D. E., K. J. Garner, and J. M. Slavicek. 1997. Identification of three randomly amplified polymorphic dna-polymerase chain reaction markers for distinguishing Asian and North American Gypsy Moths (Lepidoptera: Lymantriidae). USDA Forest Service / UNL Faculty Publications. Paper 126.

Tamura, K. 1992. Estimation of the number of nucletide substituions when there are strong transition-transverstion and G+C-content biases. Mol. Biol. Evol. 9: 678–687.

Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30: 2725–2729.

Templeton, A. R., K. A. Crandall, and C. F. Sing. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restricion endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics 132: 619–633.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic. Acids. Res. 22: 4673–4680.

Yang, F. 2013. The study on flight ability among different geographic populations of Asian Gypsy Moth, *Lymantria dispar* in China. M. S. dissertation, Beijing Forestry Univeristy, Beijing, China.