

## RESEARCH ARTICLE

# Unraveling the hierarchical genetic structure of tea green leafhopper, *Matsumurasca onukii*, in East Asia based on SSRs and SNPs

Li Zhang<sup>1,2</sup>  | Christopher H. Dietrich<sup>3</sup> | Ye Xu<sup>1,4</sup> | Zhaofu Yang<sup>1</sup>  | Maohua Chen<sup>1</sup> | Thai H. Pham<sup>5,6</sup> | Cuong C. V. Le<sup>5</sup> | Li Qiao<sup>7</sup> | Masaya Matsumura<sup>8</sup> | Daozheng Qin<sup>1</sup> 

<sup>1</sup>Key Laboratory of Plant Protection Resources and Pest Management of the Ministry of Education, Entomological Museum, Northwest A&F University, Yangling, Shaanxi, China

<sup>2</sup>Institute of Jiangxi Oil-Tea Camellia, Jiujiang University, Jiujiang, Jiangxi, China

<sup>3</sup>Illinois Natural History Survey, Prairie Research Institute, University of Illinois, Champaign, Illinois, USA

<sup>4</sup>College of Agriculture, Jiangxi Agricultural University, Nanchang, Jiangxi, China

<sup>5</sup>Mien Trung Institute for Scientific Research, Vietnam National Museum of Nature, VAST, Hue, Vietnam

<sup>6</sup>Graduate School of Science and Technology, Vietnam Academy of Science and Technology, Hanoi, Vietnam

<sup>7</sup>College of Agronomy, Xinyang Agricultural and Forestry University, Xinyang, Henan, China

<sup>8</sup>Institute for Plant Protection, National Agriculture and Food Research Organization, Tsukuba, Ibaraki, Japan

## Correspondence

Daozheng Qin, Key Laboratory of Plant Protection Resources and Pest Management of the Ministry of Education, Entomological Museum, Northwest A&F University, Yangling, Shaanxi, China. Email: [qindaozh0426@aliyun.com](mailto:qindaozh0426@aliyun.com)

## Funding information

National Natural Science Foundation of China, Grant/Award Number: 31270689; Quanzhou City Science and Technology Program of China, Grant/Award Number: 2018C040R; Vietnam Academy of Science and Technology, Grant/Award Number: NCXS02.04/22-23

## Abstract

*Matsumurasca onukii* (Matsuda, R. (1952). *Oyo-Kontyu Tokyo*, 8(1): 19–21), one of the dominant pests in major tea production areas in Asia, currently is known to occur in Japan, Vietnam, and China, and severely threatens tea production, quality, and international trade. To elucidate the population genetic structure of this species, 1633 single nucleotide polymorphisms (SNPs) and 18 microsatellite markers (SSRs) were used to genotype samples from 27 sites representing 18 geographical populations distributed throughout the known range of the species in East Asia. Analyses of both SNPs and SSRs showed that *M. onukii* populations in Yunnan exhibit high-genetic differentiation and structure compared with the other populations. The Kagoshima (JJ) and Shizuoka (JS) populations from Japan were separated from populations from China by SNPs, but clustered with populations from Jinhua (JH), Yingde (YD), Guilin (GL), Fuzhou (FZ), Hainan (HQ), Leshan (CT), Chongqing (CY), and Zunyi (ZY) tea plantations in China and the Vietnamese Vinh Phuc (VN) population based on the SSR data. In contrast, CT, CY, ZY, and Shaanxi (SX) populations clustered together based on SNPs, but were separated by SSRs. Both marker datasets identified significant geographic differentiation among the 18 populations. Various environmental and anthropogenic factors, including geographical barriers to migration, human transport of hosts (*Camellia sinensis* [L.] O. Kuntze) and adaptation of *M. onukii* to various local climatic zones possibly account for the rapid spread of this pest in Asia. The results demonstrate that SNPs from high-throughput genotyping data can be used to reveal subtle genetic substructure at broad scales in r-strategist insects.

## KEYWORDS

*Matsumurasca onukii*, reduced-representation genome sequencing, SNPs, SSRs, subtle genetic differentiation, tea pest

## TAXONOMY CLASSIFICATION

Population genetics

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

## 1 | INTRODUCTION

Recent methodological advances in population genetic and data analysis have greatly facilitated our understanding of genetic diversity, population structure, and microevolution of insect pests, which can guide the prediction of evolutionary potential (such as environmental adaptability and population fecundity) and facilitate ecological regulation and management (such as biocontrol technology and pest behavior regulation) (Roderick, 1996; Roderick & Navajas, 2003). However, detecting subtle genetic differentiation remains a major challenge because many species, particularly r-strategists, have large effective population sizes and high gene flow due to long-distance dispersal or human-mediated movement on crop plants (Matsumoto et al., 2013; Mun et al., 1999; Shi et al., 2012). In addition to exhibiting subtle levels of genetic differentiation, such populations may display diverse patterns of genetic structure, including classical patterns of isolation by vicariance, host, regional clustering, origin, and dispersion (Du et al., 2020; Goodman et al., 2019), as well as complex patterns that lack clear spatial structuring or that may be influenced by anthropogenic factors and habitat (Jiang et al., 2020; Kareem et al., 2018; Raszick et al., 2019). Identifying subtle complex genetic patterns remains a critical objective for developing effective control measures because some pest management strategies require detailed knowledge of genetic diversity and genetic structure determination that can underpin downstream efforts to measure contemporary and historical migration, delineate groups within populations, and infer evolutionary history (Jiang et al., 2020; Raszick et al., 2019).

Tea green leafhopper, *Matsumurasca onukii* (formerly *Empoasca onukii*), is one of the dominant pests in the major tea production regions of East Asia (Chen et al., 2020; Fu, Han, & Xiao, 2014; Qin et al., 2015), the species was originally described by Matsuda (1952) as an injurious tea leafhopper in Japan. Although this species was also observed in Chinese tea plantations since the 1960s, it was widely misidentified as "*Empoasca vitis*" until recently (Kuoh & Zhang, 1988; Lv et al., 1964; Zhang, 1980; Zhao et al., 2000). Recent clarification and phylogenetic analysis of the tribe Empoascini confirmed that the Chinese and Japanese populations are conspecific and correctly named *Matsumurasca onukii* (Dworakowska, 1971, 1982; Fu, Han, & Xiao, 2014; Qin et al., 2015; Xu et al., 2021). Previous comparative morphological studies revealed obvious intraspecific variation among populations, with individuals from different populations in southern China differing in the shape of membranous flanges and lengths of spiny protuberances on the shaft of the male aedeagus (Qin et al., 2015; Xu et al., 2021). This pest currently is widely distributed and causes considerable yield loss in diverse cultivated varieties of *Camellia sinensis* (L.) O. Kuntze in Japan, Vietnam, and China that are grown under a variety of ecological conditions (Dworakowska, 1971, 1982; Qin et al., 2015). Based on these prior morphological observations, the broad distribution of the species and the differences in environmental conditions among tea-growing

regions, we hypothesized that these morphological variations reflect genetic differences among geographic populations.

Previous studies of the population structure and genetic differentiation of tea green leafhopper have used mitochondrial gene sequence data or limited numbers of microsatellite markers (Chen et al., 2015; Fu, Li, et al., 2014; Li et al., 2013; Zhang et al., 2016, 2019; Zhang et al., 2018; Zhou et al., 2014; Zhu et al., 2019). Microsatellite-based results revealed some geographical differentiation between populations, particularly between two populations in Yunnan and the other Chinese populations but provided only limited evidence of consistent population genetic structure, potentially affected by endogenous properties of the markers used (e.g., fluctuations in allele frequency and recombination rates), the quality of genetic data and the sampling methods. For the present study, we obtained more extensive data using next-generation sequencing (NGS) to more fully characterize the genetic substructure in *M. onukii* population in East Asia.

The NGS methods based on reduced-representation genome sequencing (RRGS) have become widely used in population genetic analyses, particularly genotyping-by-sequencing (GBS) and restriction-associated DNA sequencing (RAD-seq) (Andrews et al., 2016; Davey et al., 2011; Elshire et al., 2011). These methods allow rapid detection of novel nuclear markers at the minimal cost. Genome-wide SNPs can provide more power for understanding population dynamics and pest species than mitochondrial DNA or microsatellite (SSR) markers (Emerson et al., 2010; Liu et al., 2019). While SSR (e.g., high polymorphism) can provide estimates of allelic distribution across whole populations, SNPs can compensate for the drawbacks of SSRs (e.g., homoplasy, complex mutation pattern, and high prevalence of null allele) (Coates et al., 2009; Kraus et al., 2015; Morin et al., 2004). Analysis of hundreds to thousands of SNPs per population has also allowed for the use of more sophisticated tools to address questions of genetic differentiation, complex patterns, and connectivity between populations (Dowle et al., 2017; Havill et al., 2017, 2019; Ma et al., 2020).

In the present study, microsatellite markers were coupled with high-throughput sequencing approaches to obtain three goals: (i) to assess the spatial distribution of genetic variation of *M. onukii* populations in East Asia based on the microsatellite markers, followed by more detailed analysis of genetic differentiation using two types of molecular markers; (ii) to compare population relationships and levels of geographic differentiation among populations from the performances of SNPs and microsatellite markers and determine if patterns reflect morphological variation among populations; and (iii) to estimated historical migration between genetic clusters for inferring the evolutionary history of *M. onukii* populations in East Asia. This study will reveal correlated patterns of morphological variation and hierarchical genetic structure in East Asian *M. onukii* population, which will be useful in monitoring species distribution and future population dynamics. This will be enable better prediction, regulation, and management of *M. onukii* populations in different tea producing regions of East Asia.

## 2 | MATERIALS AND METHODS

### 2.1 | Sampling and species identification

Tea green leafhopper specimens of 27 representative populations, including 24 Chinese localities, two populations from Japan (Kagoshima and Shizuoka), and one population from Vietnam (Vinh Phuc), were collected during the seasons of tea green leafhopper occurrence from 2013 to 2020 (Table S1). The samples included specimens collected for our previous study as well as additional samples from tea producing areas of Japan (Shizuoka [JS]), Vietnam (Vinh Phuc [VN], Hainan [HQ]) and Southwest China (Mojiang [MJ], Menghai [MH], Lincang [LX], and Jingdong [JD]) (Table S1; Figure 1). All the collection sites were areas of more than 6 ha under continuous tea cultivation without weeds or trees. Adult specimens were captured by sweep net in five collecting points (10 m<sup>2</sup>) near the center of each site, spaced 100m or more apart. Specimens from all collecting sites within a location were pooled to represent one population. All these specimens were preserved in 100% ethanol and stored in the Entomological Museum, Northwest A & F University, Yangling, China (NWAUFU).

Species identifications were confirmed by examining the morphological characters of all specimens from each representative population by dissection under an OLYMPUS SZX-10 Stereoscopic Zoom Microscope. Male genitalia of each specimen was separately preserved in a small tube prefilled with glycerin and diagnostic characters (especially aedeagus characteristics) among populations were checked and compared to confirm the species identification and check for intraspecific variation.

### 2.2 | DNA extraction

Thirty male adults in each population, except for the dissected male genitalia used for species identification, were used for extracting DNA using either the CTAB method or DNeasy blood and tissue kits

(Qiagen). The quality of DNA was detected and quantified with a Nanodrop One (Thermo Scientific).

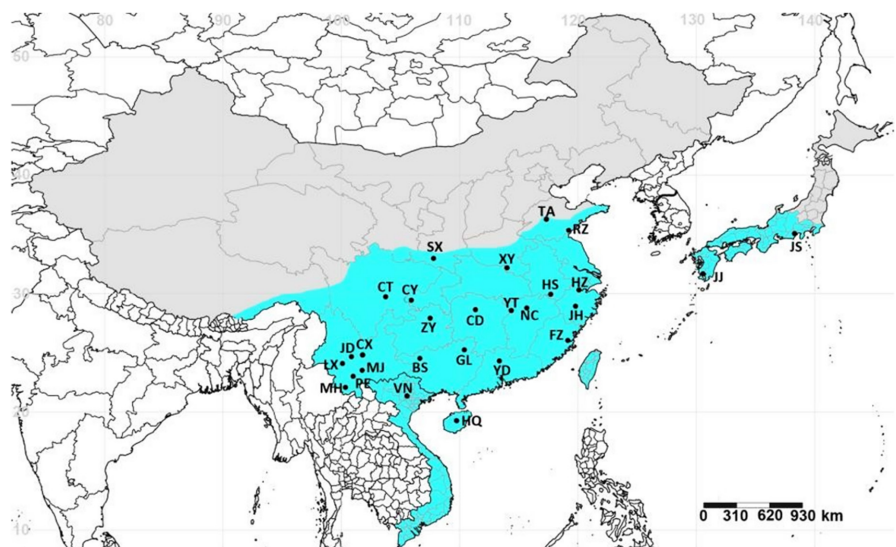
### 2.3 | Genetic structure using SSRs

#### 2.3.1 | SSRs amplification and genotyping

Individuals of seven populations (JS, VN, HQ, MJ, MH, LX, and JD) were genotyped using 18 previously developed microsatellite markers (Table S2; Zhang et al., 2016). Genotypes of the other 20 populations were obtained from previous studies (Zhang et al., 2019). All forward primers were labeled with fluorescent markers (FAM, HEX, and TAMRA). DNA from each individual was amplified for these markers in a 10 µl volume. PCR reactions followed the protocol described by Zhang et al. (2016). All the products were pooled and genotyped by automated capillary electrophoresis on an ABI 3130xl at Sangon Biotech. Microsatellite alleles were scored using the Microsatellite Plugin v.1.4 in Geneious v. 7.1.7 (Biomatters Ltd). The fragment lengths of different loci in different populations were calculated according to the molecular internal reference (Liz 500). All the genotypic data conversions were performed with GenALEx v. 6.5 (Peakall & Smouse, 2012) and CONVERT (Glaubitz, 2004).

#### 2.3.2 | Population genetic structure analysis

The frequency and signs of null alleles were investigated using FreeNA (Chapuis & Estoup, 2007) and Micro-Checker (Van Oosterhout et al., 2004). Tests for the Hardy-Weinberg Equilibrium (HWE) and Linkage disequilibrium (LD) were carried out using Genepop v. 4.2. The frequencies of alleles and allelotypes, and allelic richness (AR) were calculated using Genepop v. 4.2 (Rousset, 2008), Cervus v. 2.0 (Marshall et al., 1998), and Fstat v. 2.9.4 (Goudet, 2003), respectively. Effective number of alleles ( $N_E$ ), observed heterozygosity ( $H_O$ ), and expected heterozygosity, ( $H_E$ ) were determined using GenALEx v. 6.5.



**FIGURE 1** Geographic distribution and populations of *Matsumurasca onukii*. Population codes are listed in Table S1. SimpleMapp was used to produce a distribution map based on the geographical coordinates in Table S1. <http://www.simplemapp.net/#tabs=0>.

Structure v 2.3.4 was used to estimate the number of genetically distinct clusters (K) among the *M. onukii* populations using an admixture ancestry model with a burn-in period of  $5 \times 10^4$  iterations and  $10^6$  Markov Chain Monte Carlo (MCMC) repetitions. The range of possible clusters was set from 1 to 27 or to 18, with 20 independent runs for each K (Pritchard et al., 2000; Pritchard et al., 2010). The optimal K-values in *M. onukii* populations was estimated by posterior probability of the data (L (K)) and the ad hoc statistic ( $\Delta K$ ) using the Web server Clumpak was used to align results from replicate analyses and visualize population structure (Evanno et al., 2005; Kopelman et al., 2015; Pritchard et al., 2000; Pritchard et al., 2010). Principal component analysis (PCoA) was performed by GenALEX v. 6.5. Neighbor-Joining (NJ) tree based on Nei's genetic distance ( $D_A$ ) was constructed using POPTREE v. 2.0 (Takezaki et al., 2010), and bootstrap values were obtained using 1000 sample replicates. The analysis of molecular variance (AMOVA) based on population clustering from Structure v. 2.3.4 and  $F_{ST}$  was calculated using Arlequin v. 3.11 (Excoffier & Lischer, 2010).

## 2.4 | Genotyping by sequencing (GBS)

### 2.4.1 | Library construction for SNP sequencing

Based on the SSR population genetic analysis results, one to six populations (each population including 10 males) with significant genetic differentiation were selected from four genetic clusters. The GBS libraries were constructed and sequenced following the protocol of Elshire et al. (2011) and Onyango et al. (2016) with modifications.

*ApeKI* and 300–600bp fragments were determined by detecting genome size and its restriction sites. DNA (more than 500ng) extracted from one sample of *M. onukii* from each of the 18 populations was added to 96-well plates with 3  $\mu$ l 10 $\times$ NEB buffer and 2  $\mu$ l *ApeKI* (5 U/ $\mu$ l) (reagents from New England Biolabs) to a final volume of 30 $\mu$ l and heated in water for 2 h. Following ligation, samples in PCR-96-well plates in 40 $\mu$ l with 60 picomoles barcode DNA (5'-ACACTCTTCCCTACACGACGCTCTCCGATCTxxx and 5'-CWGyyyyA GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT) and universal adaptor (5'-CWGAGATCGGAAGAGCGTTTCAGCAGGAATG CCGAG and 5'-CTCGGC ATTCCTGCTGAACCGCTCTCCGATCT), 1mM dNTPs, 5 units T4 DNA Ligase and 4  $\mu$ l 10 $\times$ T4 DNA Ligation Buffer. The reaction system was incubated for 1 h and inactivated for 30min at 65°C. The Barcode DNA provided special identification tags for each sample, and the included universal adaptor that also provided binding sites for sequencing primers. All the DNA fragments (300–600bp) were pooled in a sequencing library after purification and recovery. Following amplification with sequencing primers, purification, and recovery, the quality of the library was assessed using an Agilent 2100 Bioanalyzer, KAPA Library Quantification Kit and Qubit 2.0. The libraries were then sequenced on an Illumina HiSeq 2500 platform using a 150bp Paired End protocol at Allwegene Technologies.

### 2.4.2 | SNP genotyping

Raw reads were obtained by Base calling, assessed for quality, and filtered by length of fragment, Q-value, and sequence alignment in databases. The identification of SNPs in Raw reads was accomplished using the Universal Network Enabled Analysis Kit (UNEAK) in TASSEL 3.0 (Bradbury et al., 2007; Lu et al., 2012, 2013). First, these reads were assigned to different individuals and populations. Following UNEAK, tags of different populations, consisting of the same reads, were used to construct network topologies. Missequenced reads, repetitive sequences, and paralogous genes were rejected in order to obtain SNPs with high accuracy. Sequencing depth and coverage, SNP calling and genotyping were processed using BWA v. 0.7.15 and SAMtools, respectively (Li et al., 2009; Li & Durbin, 2009). In population-genetic filters, the minor allele frequency was less than 0.10. Finally, SNPs with minimum allele frequency  $<0.01$  ( $Q_{20} > 95\%$ ,  $Q_{30} > 90\%$ ), coverage more than four and sequencing error rate not  $<0.05$ , were retained for subsequent population genetic analysis. Conversion from Fastq files to the desired formats was carried out using PGDSpider v. 2.0.5.1 (Lischer & Excoffier, 2012).

### 2.4.3 | Population diversity, population structure and SNP phylogeography

The observed heterozygosity ( $H_o$ ), expected heterozygosity ( $H_e$ ), and nucleotide diversity ( $\pi$ ) were determined by VCFtools v.4.0 (Danecek et al., 2011). Population structure of *M. onukii* was inferred using three methods for the GBS datasets. The number of subgroups (K) in random mating populations was estimated by FASTStructure with a Bayesian assignment test, and the optimal K was chosen based on analyses with K from 1 to 20 replicates for cross-validation, with clustering pattern generated by the online software Structure Plot v. 2.0. We also implemented the best model (determined by JMODELTest v. 2.1.7) to reconstruct population relationships using maximum-likelihood (ML) with 100 bootstrap replicates in MEGA v. 7.0.14 (Kumar et al., 2016), setting *Empoasca. (s.str.) vitis* as the outgroup. The final tree was visualized and beautified using iTOL v. 4.0 (Letunic & Bork, 2019; <http://itol.ebi.de>); PCoA was performed using SNPRelate in R (Zheng et al., 2012). To confirm the results in FASTStructure, AMOVA,  $F_{ST}$  and statistical significance with 1000 re-sampled replicates were performed in Arlequin v. 3.5 (Excoffier & Lischer, 2010).

## 2.5 | Isolation by distance analysis

Isolation by distance (IBD) analyses were conducted to test whether genetic differentiation increased with geographic distance. The correlation between the linearized pairwise  $F_{ST}$  matrix ( $F_{ST}/[1-F_{ST}]$ ) and the logarithm of geographic distance matrix (Log km) were estimated using a Mantel test with 999 permutations



in GenALEx v. 6.5. The geographic distance between each pair of populations was estimated by Geographic Distance Matrix Generator v. 1.2.3 (Ersts, 2014).

## 2.6 | Historical gene flow estimates

Given high polymorphism and mutation rates, microsatellite markers were used to estimate population size and migration parameters among clusters obtained from the results of population structure analysis. The immigration rate ( $M$ ) per generation and the mutation-scaled population size ( $\theta$ ) were analyzed using the Bayesian inference with full model in MIGRATE v. 3.7.2 (Beerli, 2006). The effective population size ( $N_e$ ) of each population was calculated with  $\theta = 4N_e\mu$  ( $\mu$  is mutation rate,  $10^{-4}$  per locus per generation for microsatellites) (Whittaker et al., 2003). Effective number of migrants per generation ( $N_{em}$ ) was calculated as  $N_{em} = \theta M/4$ . The first run was estimated from  $F_{ST}$ , and the other three runs were started with  $\theta$  and  $M$  from the previous run. During the Bayesian search, one long chain with four independent replicates was used for each run with 10,000,000 generations after an initial burn in of 100,000 interactions. Heated chain was set to: 1.0, 1.5, 3.0, and 1000,000.

## 3 | RESULTS

### 3.1 | Morphological variation

The morphological comparisons among specimens from different populations revealed that the aedeagus of the Yunnan populations (CX, PE, MJ, MH, LX, and JD) of *M. onukii* exhibits substantial

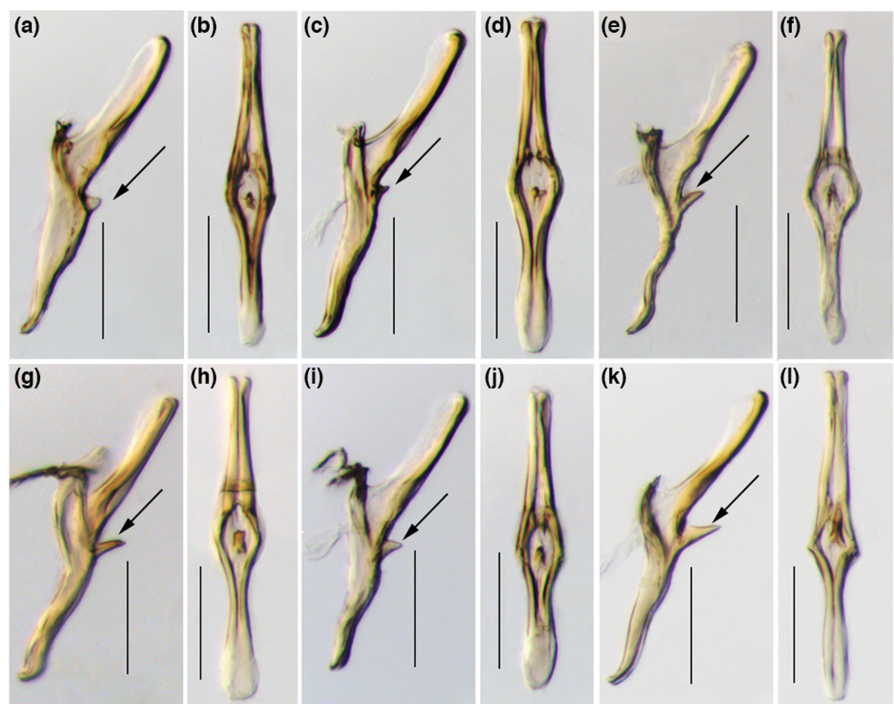
variation in the shape of the dorsal membranous flanges and the length of ventro-basal spiny protuberances of the aedeagal shaft (Figure 2); in contrast, such variation was not observed in males from the other populations from China and populations from Japan and Vietnam. Genetic data were obtained from individuals representing these different morphological variants (spiny protuberances and no protuberances on dorsal aedeagal shaft).

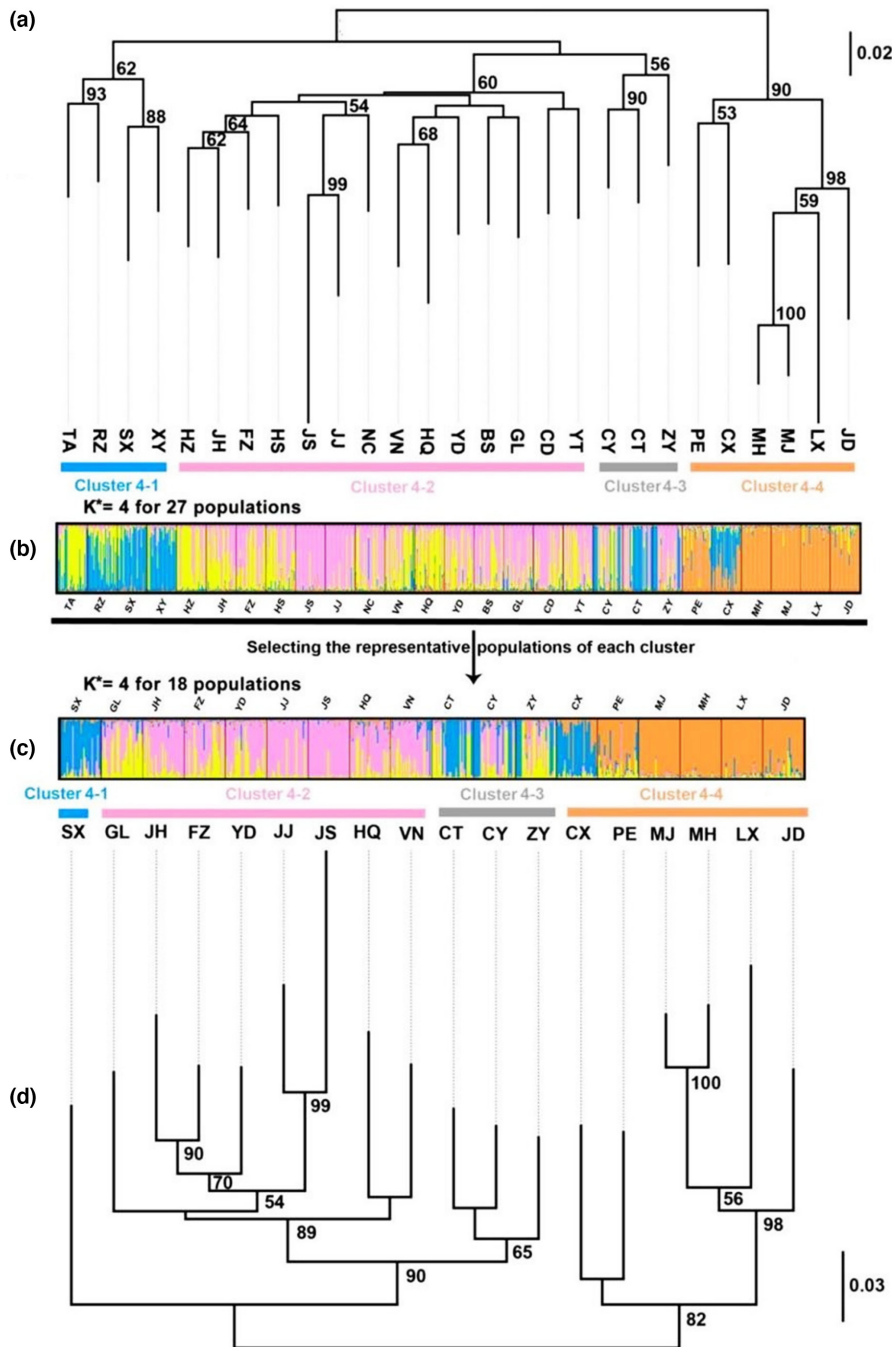
### 3.2 | Genetic diversity and genetic structure revealing by SSRs

The HWE test showed that 18 microsatellite markers were suitable for further genetic analysis, with low-null allele frequencies, ranging from 0.009 to 0.081. The observed populations of *M. onukii* showed high-genetic diversity in all tea production areas, as measured by the number of effective alleles (4.11–5.93), expected heterozygosity (0.665–0.778) and AR (7.244–10.060). Higher genetic diversity was found across *M. onukii* populations in China than in Japan (China/Japan;  $N_E$ : 5.11/4.28,  $p = .036$ ; AR: 8.815/7.748,  $p = .040$ ;  $H_E$ : 0.741/0.696,  $p = .020$ ) (Table S1).

Preliminary depictions of genetic structure, as revealed by Structure v. 2.3.4, correspond with branches on the NJ tree (Figure 3a,b). Although  $\Delta K$  had peaks in  $K = 2$  and  $K = 4$ ,  $L(K)$  was relatively constant at  $K = 4$ . Furthermore, individuals were strongly assigned to  $K = 4$ . East Asian populations of *M. onukii* formed four main genetic clusters: Cluster 4–1 includes four populations from Central and East China, viz. Xinyang (XY), Rizhao (RZ), Taian (TA), and Hanzhong (SX) populations; Cluster 4–2 consists of 14 populations from South China, Japan, and Vietnam, viz. Jinhua (JH), Hangzhou (HZ), Huangshan (HS), Nanchang (NC), Yichun

**FIGURE 2** Morphological variations of aedeagus of *Matsumurasca onukii* from different tea production regions of Yunnan Province. (a) Chuxiong (CX) population, aedeagus; (b) Chuxiong (CX) population; (c) Pu'er (PE) population; (d) Pu'er (PE) population; (e) Mojiang (MJ) population; (f) Mojiang (MJ) population; (g) Lincang (LX) population; (h) Lincang (LX) population; (i) Jingdong (JD) population; (j) Jingdong (JD) population; (k) Menghai (MH) population; (l) Menghai (MH) population. (a,c,e,g,i,k) aedeagus, left lateral view; (b,d,f,h,i,j) aedeagus, ventral view. Scale bars: (a–l) = 0.1 mm. Arrows display variations in the shape of the dorsal membranous flanges and the length of ventro-basal spiny protuberances of the aedeagal shaft.





**FIGURE 3** Bayesian clustering results and neighbor-joining (NJ) tree of *Matsumurasca onukii* populations based on 18 microsatellite markers. Numbers on nodes represent bootstrap support values (values <50% not shown). The colors indicate the major clusters inferred by Bayesian clustering analysis when  $K = 4$ , except in Cluster 4–3 mixed genetic components are indicated by gray. Asterisk represent the optimal  $K = 4$ . (a,d): NJ tree for 27 populations and 18 populations; (b,c): Bayesian clustering results for 27 populations and 18 populations.

**TABLE 1** Analysis of molecular variance (AMOVA) of 18 *Matsumurasca onukii* populations based on the microsatellite markers

Cluster	Source of variation	df	Sum of squares	Variance components	Percentage of variation (%)	Fixation indices ( $p < .0001$ )
K = 4	Among clusters	3	201.039	0.19071 Va	2.78	FCT = 0.028
	Among populations within clusters	14	303.325	0.25423 Vb	3.71	FSC = 0.038
	Within population	1062	6809.550	6.41201 Vc	93.51	FST = 0.065
	Total	1079	7313.914	6.85695		

(YT), Chengde (CD), Yingde (YD), Guilin (GL), Baise (BS), Fuzhou (FZ), Hainan (HQ), Kagoshima (JJ), Shizuoka (JS), and Vinh Phuc (VN) populations; Cluster 4–3 includes three populations from Southwest China, viz. Leshan (CT), Chongqing (CY), and Zunyi (ZY)

populations; Cluster 4–4 includes six populations from Southwest China, viz. Chuxiong (CX), Pu'er (PE), Mojiang (MJ), Menghai (MH), Lincang (LX), and Jingdong (JD) populations. The six populations of *M. onukii* from the Southwest tea production area in China

(CX, JD, MH, PE, LX, and MJ) were also clustered (Figure S1a,b) in the principal coordinate analysis. The AMOVA revealed that most of the variance was due to variation within populations (93.51%,  $F_{ST} = 0.065$ ;  $p < .0001$ ); although a significant fraction of the total variation was also due to variation between clusters (2.78%,  $F_{CT} = 0.028$ ;  $p < .0001$ ) and between populations (3.71%,  $F_{SC} = 0.038$ ;  $p < .0001$ ) (Table 1). Low-level genetic differentiation was detected between populations in Cluster 4-1, Cluster 4-2 and Cluster 4-3, whereas there were moderate to high levels of genetic differentiation between Yunnan (Southwest China) populations and other populations of *M. onukii* throughout the tea plantations in East Asia ( $F_{ST} > 0.050$ ) (Figure S2).

### 3.3 | Genetic diversity revealed by SNPs

Based on the genetic structure and genetic differentiation analyses of 27 populations of *M. onukii* in East Asia using microsatellite data, 18 populations were chosen for GBS analyses. After quality control, a final GBS dataset of 1633 SNPs representing 180 genotypes of *M. onukii* was obtained for further genetic analysis. On an average, the sequence depth of populations ranged from 5.85 to 10.84. The observed heterozygosity (range 0.231–0.364), the expected heterozygosity (range 0.210–0.291) and nucleotide diversity (range 0.198–0.275) indicated that higher genetic diversity is present in populations of *M. onukii* from the Southwest tea production areas in China ( $H_0$ :  $t$  test:  $t = -2.665$ ,  $df = 13$ ,  $p = .019$ ;  $H_E$ :  $t$  test:  $t = -2.671$ ,  $df = 13$ ,  $p = .040$ ). Among all 18 populations, the genetic diversity of populations of *M. onukii* in China was higher than those in Japan (China/Japan;  $\pi$ :  $t$  test:  $t = 3.702$ ,  $df = 15$ ,  $p = .002$ ) and Vietnam ( $H_0$ :  $t$  test:  $t = 5.560$ ,  $df = 15$ ,  $p = .000$ ;  $H_E$ :  $t$  test:  $t = 5.473$ ,  $df = 15$ ,  $p = .000$ ;  $\pi$ :  $t$  test:  $t = 3.141$ ,  $df = 15$ ,  $p = .007$ ) (Table S1).

### 3.4 | Population structure and SNP phylogeography

The PCoA and ML analyses yielded three genetic clusters (Figure S3; Figure 4a): Cluster 3-1 contains all of Clusters 4-1 and 4-3 and part of Cluster 4-2 from the microsatellite-based analysis: Hanzhong (SX), Jinhua (JH), Yingde (YD), Guilin (GL), Fuzhou (FZ), Hainan (HQ), Leshan (CT), Chongqing (CY), Zunyi (ZY), and Vinh Phuc (VN) populations. Cluster 3-2 is identical to Cluster 4-4 from the microsatellite-based analysis and consists of Chuxiong (CX), Pu'er (PE), Mojiang (MJ), Menghai (MH), Lincang (LX), and Jingdong (JD) populations. Cluster 3-3 contains only the two Japanese populations: Kagoshima (JJ), Shizuoka (JS). Japanese populations grouped with Cluster 4-2 comprising populations from central China and Vietnam in the microsatellite-based analysis although the former grouped together on a single, relatively long branch in the NJ tree. The optimal K for the SNP data inferred from FASTStructure ranged from 1 to 18 (Figure 4b), and

the lowest cross-validation error (CV) was obtained for  $K = 3$ . This clustering condition, supported by AMOVA, revealed that populations were genetically more differentiated across the three clusters (13.38%,  $F_{CT} = 0.134$ ;  $p < .0001$ ) than between populations (4.32%,  $F_{CT} = 0.050$ ;  $p < .0001$ ) (Figure 4; Table 2). Pairwise  $F_{ST}$  ranged from 0.00039 to 0.362 for SNPs. Both datasets indicate that six populations of *M. onukii* from the Southwest tea production area in China are significantly differentiated from populations in all other sites, particularly populations in Japan (Figure S4).

### 3.5 | Isolation by distance analysis

The traditional Mantel test identified significant geographic differentiation among 18 populations in both marker datasets (SNPs:  $R_{XY} = .461$ ,  $p < .001$ ; SSR:  $R_{XY} = .426$ ,  $p < .000$ ) (Figure S5). The IBD signal within populations from Southwest China was considerable when based on SNPs ( $R_{XY} = .647$ ,  $p = .002$ ), but not significant based on microsatellite markers ( $R_{XY} = .319$ ,  $p = .061$ ). IBD tests revealed no significant geographic differentiation in genetic clusters (Cluster 3-1:  $R_{XY} = .220$ ,  $p = .196$ ) with weak genetic differentiation and small distribution (Figure S6).

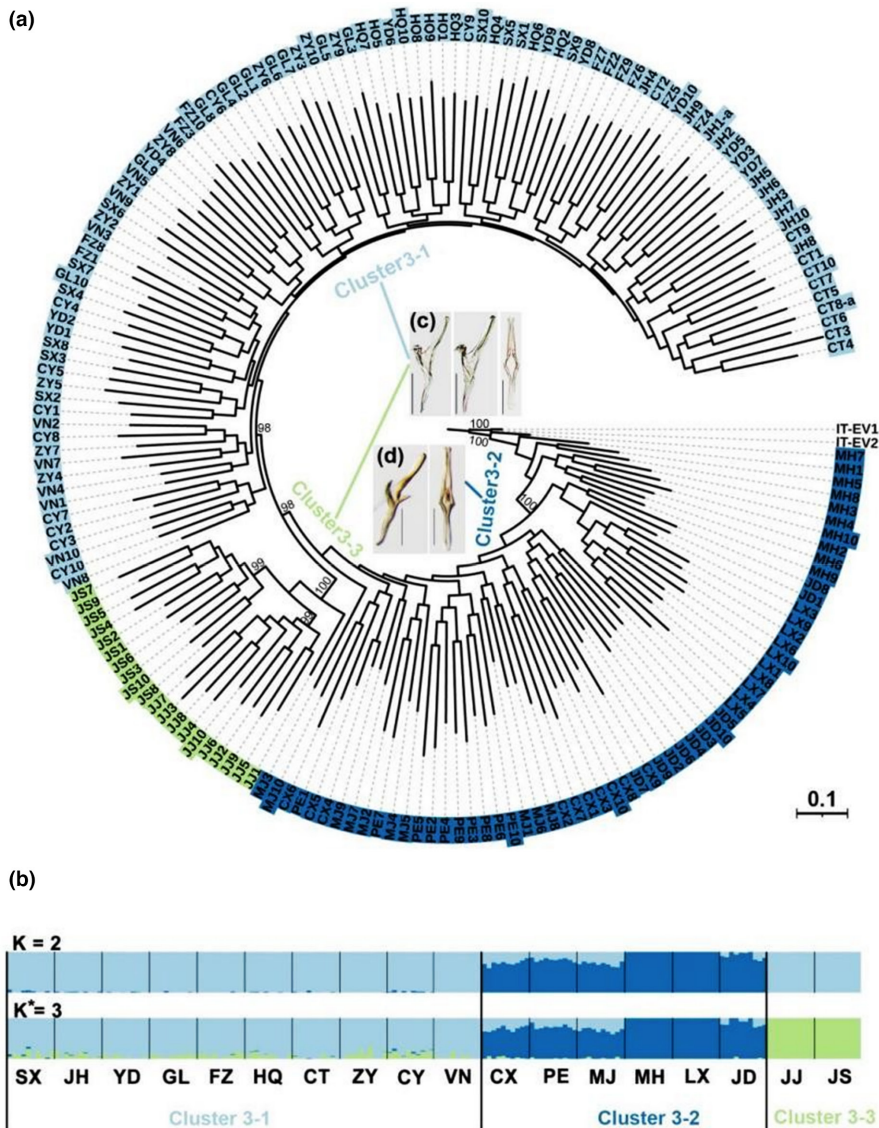
### 3.6 | Historical gene flow estimates

Given the differences in results of genetic structure analyses based on SNPs versus microsatellites, estimates of gene flow among population clusters identified by both analyses were made. In order to clarify the relationship between populations in China and Vietnam, the VN population was also separated for gene flow estimates. Effective cluster size calculated with population size parameter  $\theta$ , ranged from 563 to 2053. Frequent exchange of migrants per generation was inferred to occur among clusters (Figure 5). High levels of bidirectional migration were inferred between Clusters 3-1 and 3-2 ( $Nem > 20$ ). Asymmetric migration was found among clusters in China and between clusters in China, Japan, and Vietnam, with effective number of migrants per generation ( $Nem$ ) from Cluster 3-1 and Cluster 3-2 to SX (34.29 and 11.37) higher than numbers of migrants in the opposite direction (11.16 and 6.33). CT/CY/ZY populations displayed asymmetric migration from Cluster 3-1 (65.13), Cluster 3-2 (20.06) and SX populations (13.02);  $Nem$  from Cluster 3-1 and Cluster 3-2 to Cluster 3-3 were 32.73 and 15.83; and immigrants to the VN population were mainly from Cluster 3-1 (19.10). When SX population or CT/CY/ZY populations were included in Cluster 3-1 based on the results by SNP dataset, obvious asymmetric migration was also detected (Figure S7).

## 4 | DISCUSSION

Accurate assessment of genetic structure is a major aim of pest molecular ecology, as it can reveal patterns of association among





**FIGURE 4** Maximum-likelihood (ML) tree (a) and Bayesian clustering result (b) for 180 samples of *Matsumurasca onukii* based on SNPs. Numbers on nodes represent bootstrap support values (values <50 not shown). The three clusters yielded by Bayesian clustering analysis are shown. Asterisk represent the optimal  $K = 3$ . (c) Populations in Cluster 3-1 and Cluster 3-3, aedeagus with no protuberances on left lateral view and ventral view; (d) populations in Cluster 3-2, aedeagus with spiny protuberances on left lateral view and ventral view.

**TABLE 2** The AMOVA of 18 *Matsumurasca onukii* populations based on reduced-representation genome sequencing

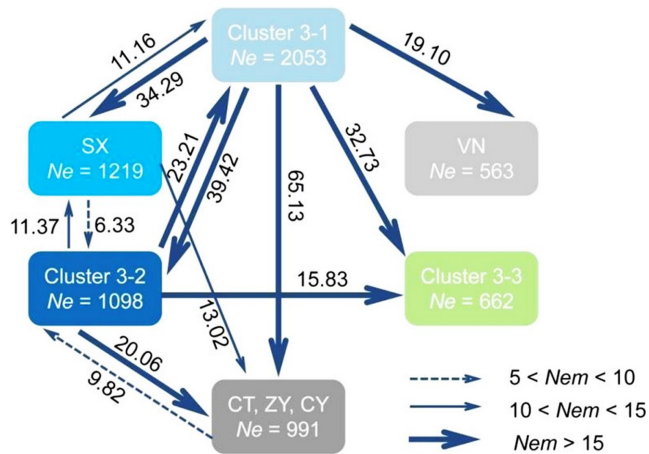
Cluster	Source of variation	df	Sum of squares	Variance components	Percentage of variation (%)	Fixation indices ( $p < .0001$ )
K = 3	Among clusters	2	4312.070	35.64395 Va	13.38	FCT = 0.134
	Among populations within clusters	15	5013.497	11.49930 Vb	4.32	FSC = 0.050
	Within population	162	35516.900	219.24012 Vc	81.75	FST = 0.177
	Total	179	44842.467	266.38337		

populations and contribute toward ecology-based management methods. In this study, we compared assessments of population structure in *M. onukii* based on two different sets of markers, SNPs and SSRs. Our SNP dataset confirmed some genetic structural patterns in *M. onukii* revealed by prior studies based on microsatellites (Zhang et al., 2019) but also revealed some differences in genetic diversity, relatedness, population structure and geographic differentiation.

#### 4.1 | Hierarchical pattern of *M. onukii* genetic structure based on microsatellites versus SNPs

As in the prior analysis based on microsatellites, analysis of SNPs showed that populations in the Yunnan tea production area of China are genetically differentiated from the other populations (Zhang et al., 2016, 2019). Nevertheless, the present analysis based on SNPs revealed an additional genetic cluster consisting only of the





**FIGURE 5** Maximum likelihood estimates for gene flow among clusters using SSRs. Cluster 3-1 excepted for Hanzhong (SX), Vinh Phuc (VN), Leshan (CT), Chongqing (CY), and Zunyi (ZY) populations (included Jinhua (JH), Yingde (YD), Guilin (GL), Fuzhou (FZ), and Hainan (HQ) populations), Cluster 3-2 (Yunnan populations included Chuxiong (CX), Pu'er (PE), Mojiang (MJ), Menghai (MH), Lincang (LX), and Jingdong (JD) populations), Cluster 3-3 (included Kagoshima (JJ) and Shizuoka (JS) populations) and Hanzhong (SX), Vinh Phuc (VN) populations, and Leshan (CT)/Chongqing (CY)/Zunyi (ZY) populations.  $N_e$  inside circles is effective population size (i.e.,  $\theta = 4 N_e \mu$ , where  $\theta$  is the population size parameter and  $\mu$  is mutation rate) and  $N_{em}$ : Effective number of migrants per generation (i.e.,  $N_{em} = \theta M/4$ , where  $M$  is scaled migration rate per generation) for *Matsumurasca onukii*. Arrows indicate the gene flow direction among clusters.  $N_{em} < 5$  and no significant asymmetrical effective migrants not shown.

two Japanese populations (Cluster 3-3 in Figure 4a) which was not recognized by Zhang et al. (2016, 2019), and also uncovered finer-scale geographic differentiation. The findings are consistent with other studies on various species, such as *Laodelphax striatellus*, *Plasmodium vivax*, *Carassius carassius*, and so on (Fola et al., 2020; Jeffries et al., 2016; Liu et al., 2019), where SNPs outperformed microsatellite markers in accuracy and clarifying finer population structure.

Our results suggest that microsatellites and SNPs may capture somewhat different aspects of population structure in *M. onukii* although the reasons for these differences remain to be explored. Microsatellite markers tend to exhibit higher levels of homoplasy, null alleles, and usually reveal a large number of alleles at a locus, resulting in underestimated genetic differentiation between subpopulations with high-genetic distance (e.g., Chinese populations and Japanese populations) and identify population structure between populations with similar levels of genetic diversity, which is consistent to many prior studies (Campos et al., 2017; Fola et al., 2020). In contrast, SNPs have a fixed number of allelic states (Elshire et al., 2011), and SNPs with slower mutation rates may also facilitate inferring higher level of genetic differentiation and provide additional discrimination among genetic clusters (Coates et al., 2009; Morin et al., 2004). This may explain why the two Japanese populations formed their own separate cluster in the analysis based on

SNPs (Cluster 3-3 in Figure 4a) while these populations grouped within a large cluster including populations from central China and Vietnam in the microsatellite-based analysis (Cluster 4-2 in Figure 3d). Similarly, the CT, CY, ZY, and SX populations were clustered with Cluster 4-1 and Cluster 4-3 based on microsatellite data (Figure 3c,d), while they were clustered with JH, YD, GL, FZ, HQ, and VN populations in Cluster 3-1 by the SNPs data (Figure 4a,b).

Our results imply that different mutation rates of the two types of markers might reflect different stages of population evolutionary history, which can lead to differences in estimated population structure. Microsatellite markers, with higher mutation rates, tend to provide insights mostly on recent divergences by compared to SNPs (Sun et al., 2009). Transfer of famous tea varieties among regions has a long history in China (Ruan, 2019) and presumably resulted in occasional transport of tea seedlings with eggs and adults of *M. onukii* over long distances between tea production areas. Frequent migration or human-mediated exchange between and within populations may have yielded similar genetic structure of *M. onukii* populations in most of tea production areas of China, except for the Southwest (particularly in Yunnan Province). The "introduction from south to north" mainly occurred in recent decades, during which major tea planting areas were established in North China (An et al., 2014). Southwest China (included Yungui Plateau and Sichuan-Chongqing region) has been confirmed as the original area of tea production (Chen, 2009; Chen et al., 2019; Yao et al., 2008). At present, transfer of fresh tea leaves among some famous tea planting areas is occurring frequently, especially in the Southwest tea areas bordering on the other tea areas (e.g., Sichuan, Chongqing, Guizhou Provinces). Differences observed among *M. onukii* populations may have resulted from occasional population bottlenecks and selection based on adaptation to local environmental conditions as *M. onukii* populations migrate along with their host plants, accounting for similar genetic diversity and structure in Cluster 4-1 and higher genetic diversity and hybrid genetic structure in Cluster 4-3 shown by microsatellites, whereas SNPs detected different scale geographic population structure that may be a result of reduced gene flow due to geographic barriers present for a long time (e.g., between Southwest China and the remaining populations and between China and Japan) (Figure 5).

## 4.2 | Correlations between morphological and geographic distance and genetic variation

Although we detected no significant genetic substructure in most Chinese tea areas, our analyses of two types of markers revealed special genetic structure and high-genetic differentiation in *M. onukii* populations from the Southwest tea areas, especially among the populations from Yunnan tea plantations that also exhibit morphological variation in the aedeagus. Such correlation between genetic and morphological variation contrasts with a previous study of *Asymmetrasca decedens*, in which no genetic differentiation was found between individuals of two morphotypes based on differences in the shape of the aedeagus, (Loukas & Drosopoulos, 1992).

The CX, PE, MJ, JD, LX, and MH populations of *M. onukii* were collected from cultivated and wild Yunnan big leaf tea in relatively isolated environments, while the other populations of China and Vietnam were from fewer varieties planted over larger areas.

Our analysis using two markers also revealed that the Kagoshima (JJ) and Shizuoka (JS) populations from Japan and the Vinh Phuc (VN) population from Vietnam are closely related to *M. onukii* populations in China. Not surprisingly, the VN population is more similar to the Chinese populations than are the two populations from Japan (Figures 3b and 4a). Japanese populations also showed relatively high levels of genetic differentiation and lower genetic diversity compared with the Chinese populations, indicating that gene flow between Chinese and Japanese populations is very low. Genetic drift and natural selection on the limited genetic resources of *M. onukii* in Japan may have yielded low-genetic diversity and high differentiation compared with the mainland populations (Kaundun et al., 2000; Wachira et al., 2001; Yao et al., 2008). In contrast, high-genetic diversity and less genetic variation were revealed in the Vietnamese population as expected, given the absence of significant geographical barriers, similar host resources, and frequency of gene exchange between VN population and JH, YD, GL, FZ, HQ, CT, CY, and ZY populations (Figures 3b and 4a; Figures S2 and S4). As mentioned earlier, genetic structure of *M. onukii* may be affected by variation in tea germplasm resources and habitat among tea plantations.

### 4.3 | Genetic connectivity and population dynamics

Analyses of population structure and phylogeography based on the SNPs reveal significant genetic differentiation between *M. onukii* populations in Yunnan (Cluster 3-2), Japan (Cluster 3-3) and the other tea plantations of China and Vietnam. Significant asymmetrical historical migration from Cluster 3-1 and Cluster 3-2 to Cluster 3-3 is supported by non-overlapping 95% confidence intervals (Figure 5; Figure S7). Previous studies based on mitochondrial markers indicated that few private haplotypes are present in the Yunnan populations (Zhou et al., 2014; Zhu et al., 2019), and the putative primitive haplotype was also present in populations from Yunnan and neighboring Sichuan, Guizhou, and Chongqing Provinces (Chen et al., 2015; Li et al., 2013). According to genetic diversity, genetic structure characteristics, historical gene flow of *M. onukii* populations revealed by this study and historical expansion of tea cultivation (Takeo et al., 1992), we hypothesize that the tea green leafhopper in China spread from Yunnan or the southwest populations to the other tea areas in Jiangnan, South China and, ultimately, to Japan with domestication and cultivation (the historical sources), and then returned back to the origin by recent events (such as fresh tea processing and tea germplasm exchange) forming a new diffusion source pattern. As a result, the leafhopper was most likely introduced to Japan tea plantations through these two diffusion sources from China (Figure 5; Figure S7). We found no evidence of

frequent, long-distance directional dispersal, as has been observed in some other species of Empoascini (e.g., the potato leafhopper, *Empoasca fabae*) in *M. onukii*. Instead, our results suggest that natural or human-assisted dispersal occurs among major tea growing areas of China but not frequently enough to prevent maintenance of population-level differences between major tea areas and high levels of genetic variability within some populations.

Additional genetic substructure in *M. onukii* populations throughout East Asia was further elucidated in this study by SNPs, and high levels of genetic differentiation were observed between populations in different clusters. But our analyses failed to reveal fine-scale dispersal routes between populations. Incorporation of additional SNPs into future studies by more efficient genotyping with the *M. onukii* reference genome (ASM1883171v1) in Genome Warehouse may allow for finer-scale analysis of gene flow among populations (Malinsky et al., 2021; Pickrell & Pritchard, 2012).

## 5 | CONCLUSION

Recent advances in sequencing technologies can provide genomic data for accurately identifying patterns of genetic structure with effective samples. The accurate genetic structure will be obtained when the gene markers and samples are chosen reasonably. In this study, representative populations of *M. onukii* were chosen by the genetic structure based on microsatellite markers, which GBS-derived SNPs and microsatellite markers all captured multiscale patterns of genetic structure. Three clusters were discovered in East Asia by SNPs, performed better than the microsatellite markers at fine spatial scales. General migration route showed that there are two diffusion sources including Yunnan and the other tea plantations of China, respectively, from where introduced to Japan tea plantations. Our findings indicate that the host plants (e.g., special tea germplasm resources) and nature geographical condition (e.g., geographical barriers) may be significant involved to the genetic variation of *M. onukii* between different Clusters.

Therefore, to illuminate fine-scale dispersal routes of *M. onukii*, further study should encompass more SNPs and additional samples from various host plants within *Camellia* spp. at fine scales (such as within the southwest tea planting area). Higher-resolution study of gene flow and its relation to geography/host associations will help elucidate the microevolution of adaptability in *M. onukii* populations in the East Asia.

### AUTHOR CONTRIBUTIONS

**Li Zhang:** Conceptualization (equal); formal analysis (lead); investigation (lead); resources (equal); software (equal); writing – original draft (lead); writing – review and editing (equal). **Christopher H. Dietrich:** Conceptualization (supporting); writing – original draft (supporting); writing – review and editing (equal). **Ye Xu:** Investigation (supporting); writing – original draft (supporting); writing – review and editing (equal). **Zhaofu Yang:** Software (equal); writing – review and editing (equal). **Maohua**

**Chen:** Conceptualization (supporting); methodology (lead); writing – review and editing (equal). **Thái H. Pham:** Funding acquisition (supporting); resources (supporting); writing – review and editing (supporting). **Cuong C. V. Le:** Resources (supporting); writing – review and editing (supporting). **Li Qiao:** Resources (supporting); writing – review and editing (supporting). **Masaya Matsumura:** Resources (supporting); writing – review and editing (supporting). **Daozheng Qin:** Conceptualization (equal); funding acquisition (lead); investigation (supporting); resources (lead); writing – original draft (supporting); writing – review and editing (equal).

## ACKNOWLEDGMENTS

This study was funded by the National Natural Science Foundation of China (31270689), Quanzhou City Science and Technology Program of China (2018C040R), and Vietnam Academy of Science and Technology (VAST) (NCXS02.04/22-23).

## CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest.

## DATA AVAILABILITY STATEMENT

Individual genotype data of *M. onukii* are available on Dryad (<https://datadryad.org/stash/share/muyHpreBGUaO-Kanv0KS5L6n7i4p-wgJ6KfgZ6qqFBM>). The raw Illumina reads are archived at the NCBI Sequence Read Archive (SRA) under BioProject ID PRJNA799739.

## BENEFIT-SHARING STATEMENT

Benefits from this research accrue from the sharing of our data and results on public databases as described earlier. No specific permits were required for this study. Tea green leafhopper, *M. onukii*, is an agricultural pest, not an endangered or protected species. All the samples were collected in open tea plantations and not from any national parks or protected areas.

## ORCID

Li Zhang  <https://orcid.org/0000-0002-5081-194X>

Zhaofu Yang  <https://orcid.org/0000-0001-8391-1907>

Daozheng Qin  <https://orcid.org/0000-0001-8219-6753>

## REFERENCES

- An, B. Y., Guo, J. Z., Han, P. B., & Sun, G. X. (2014). Discussion on controlling disease, pests and grass in organic tea garden of Jiangbei tea area. *China Tea*, 36(9), 24–25.
- Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, 17(2), 81–92.
- Beerli, P. (2006). Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics*, 22(3), 341–345.
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23, 2633–2635.
- Campos, M., Conn, J. E., Alonso, D. P., Vinetz, J. M., Emerson, K. J., & Ribolla, P. E. M. (2017). Microgeographical structure in the major neotropical malaria vector *Anopheles darlingi* using microsatellites and SNP markers. *Parasites & Vectors*, 10, 76.
- Chapuis, M. P., & Estoup, A. (2007). Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, 24, 621–631.
- Chen, D. J., Ma, C. L., & Chen, L. (2019). Research on tea germplasm resources of China in the past 40 years. *China Tea*, 41(6), 1–5.
- Chen, S. C., Wang, X. Q., Peng, P., Hu, X., Duan, X. F., & Lin, Q. (2015). Genetic differentiation of *Empoasca vitis* (Göethe) (Hemiptera: Cicadellidae) among eleven populations based on 16S rRNA sequence. *Southwest China Journal of Agricultural Sciences*, 28(2), 584–590.
- Chen, Z. (2009). Guizhou tea tree species resources and tea industrial economic development analysis. *Seed*, 28(11), 74–76.
- Chen, Z. M., Cai, X. M., Zhou, L., Bian, L., & Luo, Z. X. (2020). Developments on tea plant pest control in past 40 years in China. *China Tea*, 42(1), 1–8.
- Coates, B. S., Sumerford, D. V., Miller, N. J., Kim, K. S., Sappington, T. W., Siegfried, B. D., & Lewis, L. C. (2009). Comparative performance of single nucleotide polymorphism and microsatellite markers for population genetic analysis. *Journal of Heredity*, 100(5), 556–564.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, B., Lunter, G., Marth, G., Sherry, S., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158.
- Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, 12(7), 499–510.
- Dowle, E. J., Bracewell, R. R., Pfrender, M. E., Mock, K. E., Bentz, B. J., & Ragland, G. J. (2017). Reproductive isolation and environmental adaptation shape the phylogeography of mountain pine beetle (*Dendroctonus ponderosae*). *Molecular Ecology*, 26(21), 6071–6084.
- Du, Z. Y., Wu, Y. F., Chen, Z., Cao, L. M., Ishikawa, T., Kamitani, S., Sota, T., Song, F., Tian, L., Cai, W. Z., & Li, H. (2020). Global phylogeography and invasion history of the spotted lanternfly revealed by mitochondrial phylogenomics. *Evolutionary Applications*, 00, 1–16. <https://doi.org/10.1111/eva.13170>
- Dworakowska, I. (1971). *Dayus takagii* sp. N. and some other Empoascini (Auchenorrhyncha: Cicadellidae: Typhlocybinae). *Bulletin de l'Academie Polonaise des Science. Serie des Sciences Biologiques*, 19, 501–509.
- Dworakowska, I. (1982). Empoascini of Japan, Korea and north-east part of China (Homoptera, Auchenorrhyncha, Cicadellidae, Typhlocybinae). *Reichenbachia*, 20, 33–57.
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, 6(5), e19379.
- Emerson, K. J., Merz, C. R., Catchen, J. M., Hohenlohe, P. A., Cresko, W. A., Bradshaw, W. E., & Holzapfel, C. M. (2010). Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, 107(37), 16196–16200.
- Ersts, P. J. (2014). Geographic Distance Matrix Generator (version 1.2.3) American Museum of Natural History, Center for Biodiversity and Conservation. [http://biodiversityinformatics.amnh.org/opne\\_source/gdmg](http://biodiversityinformatics.amnh.org/opne_source/gdmg)
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology*, 14, 2611–2620.
- Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and windows. *Molecular Ecology Resources*, 10, 564–567.
- Fola, A. A., Kattenberg, E., Razook, Z., Lautu-Gumal, D., Lee, S., Mehra, S., Bahlo, M., Kazura, J., Robinson, L. J., Laman, M., Mueller, L., &

- Barry, A. E. (2020). SNP barcodes provide higher resolution than microsatellite markers to measure *plasmodium vivax* population genetics. *Malaria Journal*, 19(1), 375.
- Fu, J. Y., Han, B. Y., & Xiao, Q. (2014). Mitochondrial *COI* and 16sRNA evidence for a single species hypothesis of *E. vitis*, *J. formosana* and *E. onukii* in East Asia. *PLoS One*, 9(12), e115259.
- Fu, J. Y., Li, L., Yuan, Z. J., Tang, M. J., & Xiao, Q. (2014). Molecular phylogenetic analysis of tea green leafhopper (*Empoasca vitis*) based on mtDNA *COI* sequences. *Journal of Tea Science*, 34(6), 601–608.
- Glaubitz, J. C. (2004). CONVERT: A user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Molecular Ecology Notes*, 4(2), 309–310.
- Goodman, K. R., Prost, S., Bi, K., Brewer, M. S., & Gillespie, R. G. (2019). Host and geography together drive early adaptive radiation of Hawaiian planthoppers. *Molecular Ecology*, 28(19), 4513–4528.
- Goudet, J. (2003). FSTAT (version 2.9.4), a program to estimate and test population genetics parameters. <http://www.unil.ch/popgen/softwares/fstat.html>
- Havill, N. P., Cognato, A. I., Del-Val, E. K., Rabaglia, R. J., & Garrick, R. C. (2019). New molecular tools for *Dendroctonus frontalis* (coleoptera: Curculionidae: Scolytinae) reveal an east-west genetic subdivision of early Pleistocene origin. *Insect Systematics and Diversity*, 3(2), 1–14.
- Havill, N. P., Elkinton, J. S., Andersen, J. C., Hagen, S. B., Broadley, H. J., Boettner, G. J., & Caccone, A. (2017). Asymmetric hybridization between non-native winter moth, *Operophtera brumata* (lepidoptera: Geometridae), and native Bruce spanworm, *Operophtera bruceata*, in the northeastern United States, assessed with novel microsatellites and SNPs. *Bulletin of Entomological Research*, 107(2), 241–250.
- Jeffries, D. L., Copp, G. H., Lawson Handley, L., Olsén, K. H., Sayer, C. D., & Hänfling, B. (2016). Comparing RADseq and microsatellites to infer complex phylogeographic patterns, an empirical perspective in the crucian carp, *Carassius carassius*. *Molecular Ecology*, 25(13), 2997–3018.
- Jiang, S., Zhu, K. X., Han, L., Chen, C., Wang, M. M., & Wang, X. Y. (2020). Genetic variation and phylogeographic structure of *Laodelphax striatellus* in China based on microsatellite markers. *Journal of Applied Entomology*, 145(4), 336–347.
- Kareem, A. A., Port, G. R., & Wolff, K. (2018). Population structure of the glasshouse whitefly *Trialeurodes vaporariorum* (Hemiptera: Aleyrodidae) shows multiple introductions to the U.K. *Agricultural and Forest Entomology*, 21(2), 139–148.
- Kaundun, S. S., Zhyvoloup, A., & Park, Y. G. (2000). Evaluation of the genetic diversity among elite tea (*Camellia sinensis* var. *sinensis*) accessions using RAPD markers. *Euphytica*, 115, 7–16.
- Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpark: A program for identifying clustering modes and packaging population structure inferences across *K*. *Molecular Ecology Resources*, 15(5), 1179–1191.
- Kraus, R. H., Vonholdt, B. M., Cocchiararo, B., Harms, V., Bayerl, H., Kühn, R., Förster, D. W., Fickel, J., Roos, C., & Nowak, C. (2015). A single-nucleotide polymorphism-based approach for rapid and cost-effective genetic wolf monitoring in Europe based on noninvasively collected samples. *Molecular Ecology Resources*, 15(2), 295–305.
- Kumar, S., Stecher, S., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33(7), 1870–1874.
- Kuoh, C. L., & Zhang, H. G. (1988). Research on the cicadellid species damaging Chinese tea (I). *Journal of Tea Business*, 1, 15–18.
- Letunic, I., & Bork, P. (2019). Interactive tree of life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Research*, 47(W1), W256–W259. <https://doi.org/10.1093/nar/gkz239>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, 25(14), 1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079.
- Li, L., Fu, J. Y., & Xiao, Q. (2013). Sequence analysis of the mtDNA gene and genetic differentiation in geographic populations of *Empoasca vitis*. *Journal Applied Entomology*, 50(3), 675–685.
- Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, 28(2), 298–299.
- Liu, Y., Chen, L., Duan, X. Z., Zhao, D. S., Sun, J. T., & Hong, X. Y. (2019). Genome-wide single nucleotide polymorphisms are robust in resolving fine-scale population genetic structure of the small brown planthopper, *Laodelphax striatellus* (Fallén) (Hemiptera: Delphacidae). *Molecular Entomology*, 112(5), 2362–2368.
- Loukas, M., & Drosopoulos, S. (1992). Population genetic studies of leafhopper (*Empoasca*) species. *Entomologia Experimentalis et Applicata*, 63(1), 71–79.
- Lu, F., Glaubitz, J., Harriman, J., Casstevens, T., Elshire, R. 2012. TASSEL 3.0 universal network enabled analysis kit (UNEAK) pipeline documentation. <https://bytebucket.org/tasseladmin/tassel-5-source/wiki/docs/TasselPipelineUNEAK.pdf>
- Lu, F., Lipka, A. E., Glaubitz, J., Elshire, R., Cherney, J. H., Casler, M. D., Buckler, E. S., & Costich, D. E. (2013). Switchgrass genomic diversity, ploidy, and evolution: Novel insights from a network-based SNP discovery protocol. *PLoS Genetics*, 9(1), e1003215.
- Lv, W. M., Chen, X., & Luo, Q. R. (1964). Research on occurrence and control of *Empoasca flavescens*. *Journal of Tea Science*, 1, 45–55.
- Ma, L., Cao, L. J., Hoffmann, A. A., Gong, Y. J., Chen, J. C., Chen, H. S., Wang, X. B., Zeng, A. P., Wei, S. J., & Zhou, Z. S. (2020). Rapid and strong population genetic differentiation and genomic signatures of climatic adaptation in an invasive mealybug. *Diversity & Distributions*, 26(5), 610–622.
- Malinsky, M., Matschiner, M., & Svardal, H. (2021). Dsuite-fast D-statistics and related admixture evidence from VCF files. *Molecular Ecology Resources*, 21, 584–595.
- Marshall, T. C., Slate, J., Kruuk, L. E. B., & Pemberton, J. M. (1998). Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, 7, 639–655.
- Matsuda, R. (1952). A new green *Empoasca* leafhopper injurious to the tea shrub in Japan. *Oyo-Kontyu Tokyo*, 8(1), 19–21.
- Matsumoto, Y., Matsumura, M., Sanada-Morimura, S., Hirai, Y., Sato, Y., & Noda, H. (2013). Mitochondrial COX sequences of *Nilaparvata lugens* and *Sogatella furcifera* (Hemiptera, Delphacidae): Low specificity among Asian planthopper populations. *Bulletin of Entomological Research*, 103(4), 382–392.
- Morin, P. A., Luikart, G., & Wayne, R. K. (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, 19(4), 208–216.
- Mun, J. H., Song, Y. H., Heong, K. L., & Roderick, G. K. (1999). Genetic variation among Asian populations of rice planthoppers, *Nilaparvata lugens* and *Sogatella furcifera* (Hemiptera: Delphacidae): Mitochondrial DNA sequences. *Bulletin of Entomological Research*, 89(3), 245–253.
- Onyango, M. G., Aitken, N. C., Jack, C., Chuah, A., Oguya, J., Djikeng, A., Kemp, S., Bellis, G. A., Nicholas, A., Walker, P. J., & Duchemin, J. B. (2016). Genotyping of whole genome amplified reduced representation libraries reveals a cryptic population of *Culicoides brevitarsis* in the northern territory, Australia. *BMC Genomics*, 17, 769.
- Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: Genetic analysis in excel. Population genetic software for teaching and research – An update. *Bioinformatics*, 28(19), 2537–2539.
- Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, 8(11), e1002967.



- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.
- Pritchard, J. K., Wen, X. Q., Falush, D. (2010). Documentation for structure software: Version 2.3, 2 February 2010. [https://web.stanford.edu/group/pritchardlab/structure\\_software/release\\_versions/v2.3.4/html/structure.html](https://web.stanford.edu/group/pritchardlab/structure_software/release_versions/v2.3.4/html/structure.html)
- Qin, D. Z., Zhang, L., Xiao, Q., Dietrich, C. H., & Matsumura, M. (2015). Clarification of the identity of the tea green leafhopper based on morphological comparison between Chinese and Japanese specimens. *PLoS One*, *10*(9), e0139202.
- Raszick, T. J., Suh, C. P., Dickens, C. M., & Sword, G. A. (2019). Genome-wide markers reveal temporal instability of local population genetic structure in the cotton fleahopper, *Pseudatomoscelis seriatus* (Hemiptera: Miridae). *Pest Management Science*, *76*(1), 324–332.
- Roderick, G. K. (1996). Geographic structure of insect population: Gene flow, phylogeography, and their uses. *Annual Review of Entomology*, *41*, 325–352.
- Roderick, G. K., & Navajas, M. (2003). Genes in new environments: Genetics and evolution in biological control. *Nature Reviews Genetics*, *4*(11), 889–899.
- Rousset, F. (2008). Genepop'007: A complete re-implementation of the genepop software for windows and Linux. *Molecular Ecology Resources*, *8*(1), 103–106.
- Ruan, J. Y. (2019). Developments on tea plant cultivation in past 40 years in China. *China Tea*, *41*(7), 1–7.
- Shi, W., Kerdelhué, C., & Ye, H. (2012). Genetic structure and inferences on potential source areas for *Bactrocera dorsalis* (Hendel) based on mitochondrial and microsatellite markers. *PLoS One*, *7*(5), e37083.
- Sun, J. X., Mullikin, J. C., Patterson, N., & Reich, D. E. (2009). Microsatellites are molecular clocks that support accurate inferences about history. *Molecular Biology and Evolution*, *26*(5), 1017–1027.
- Takeo, T., You, X. Q., Wang, H. F., Kinukasa, H., Li, M. J., Chen, Q. K., & Wang, H. S. (1992). One speculation the origin and dispersion of tea plant in China—One speculation based on the chemotaxonomy by using the content-ratio of terpene-alcohols found in tea aroma composition. *Journal of Tea Science*, *12*, 81–86.
- Takezaki, N., Nei, M., & Tamura, K. (2010). POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with windows interface. *Molecular Biology and Evolution*, *27*(4), 747–752.
- Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., & Shipley, P. (2004). Micro-checker: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, *4*, 535–538.
- Wachira, F. N., Tanaka, J., & Takeda, Y. (2001). Genetic variation and differentiation in tea (*Camellia sinensis*) germplasm revealed by RAPD and AFLP variation. *Journal of Horticultural Science & Biotechnology*, *76*, 557–563.
- Whittaker, J. C., Harbord, R. M., Boxall, N., Mackay, I., Dawson, G., & Sibly, R. M. (2003). Likelihood-based estimation of microsatellite mutation rates. *Genetics*, *164*(2), 781–787.
- Xu, Y., Dietrich, C. H., Zhang, Y. L., Dmitriev, D. A., Zhang, L., Wang, Y. M., Lu, S. H., & Qin, D. Z. (2021). Phylogeny of the tribe Empoasini (hemiptera: Cicadellidae: Typhlocybinae) based on morphological characteristics, with reclassification of the *Empoasca* generic group. *Systematic Entomology*, *46*, 266–286.
- Yao, M. Z., Chen, L., & Liang, Y. R. (2008). Genetic diversity among tea cultivars from China, Japan and Kenya revealed by ISSR markers and its implication for parental selection in tea breeding programmes. *Plant Breeding*, *127*(2), 166–172.
- Zhang, J. W. (1980). The eighth lecture, cicadellids. *Tea Communication*, *4*, 63–65.
- Zhang, L., Dietrich, C. H., & Qin, D. Z. (2016). Microsatellite markers from tea green leafhopper *Empoasca (Matsumurasca) onukii*: A powerful tool for studying genetic structure in tea plantations. *BMC Genetics*, *17*, 112.
- Zhang, L., Wang, F. P., Qiao, L., Dietrich, C. H., Matsumura, M., & Qin, D. Z. (2019). Population structure and genetic differentiation of tea green leafhopper, *Empoasca (Matsumurasca) onukii*, in China based on microsatellite markers. *Scientific Reports*, *9*, 1202.
- Zhang, Y. J., Chen, W. L., Yang, L., Chen, X. S., Dong, M. S., & Li, X. (2018). Sequence analysis of mtDNA gene and genetic differentiation in geographic populations of *Empoasca onukii* in Southwest (Guizhou Province) of China. *International Journal of Agricultural and Biology*, *20*(2), 404–414.
- Zhao, D. X., Chen, Z. M., & Cheng, J. A. (2000). Belongingness of tea leafhopper dominant species. *Journal of Tea Science*, *20*, 101–104.
- Zheng, X. W., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328.
- Zhou, N. N., Wang, M. X., Cui, L., Pan, C., Zhang, X. T., & Han, B. Y. (2014). Genetic variation of *Empoasca vitis* (Göthe) (Hemiptera: Cicadellidae) among different geographical populations based on mtDNA *COI* complete sequence. *Acta Ecologica Sinica*, *34*(23), 6879–6889.
- Zhu, Y., Li, J., Pang, H. Y., Wang, Z., Du, N., Xiong, S. L., & Cai, T. (2019). Genetic differentiation of 11 geographical populations of *Empoasca vitis* (Göethe) (Hemiptera: Cicadellidae) based on mtDNA *cytb* gene. *Biotic Resources*, *41*(3), 228–235.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Zhang, L., Dietrich, C. H., Xu, Y., Yang, Z., Chen, M., Pham, T. H., Le, C. C. V., Qiao, L., Matsumura, M., & Qin, D. (2022). Unraveling the hierarchical genetic structure of tea green leafhopper, *Matsumurasca onukii*, in East Asia based on SSRs and SNPs. *Ecology and Evolution*, *12*, e9377. <https://doi.org/10.1002/ece3.9377>