

# Development and Characterization of Novel Microsatellite Markers for a Dominant Parasitoid (*Stethynium Empoasca*) in Tea Plantations Using High-throughput Sequencing

**Jinyu Li**

Institute of Applied Ecology

**Longqing Shi**

Institute of Rice

**Jie Chen**

Institute of Applied Ecology

**Minsheng You** (✉ [msyou@fafu.edu.cn](mailto:msyou@fafu.edu.cn))

Institute of Applied Ecology

**Shijun You**

Institute of Applied Ecology

---

## Short report

**Keywords:** parasitoid, biological control, microsatellites, high-throughput sequencing, tea plantation

**DOI:** <https://doi.org/10.21203/rs.3.rs-65383/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

# Abstract

**Background:** *Stethynium empoasca* is an egg parasitoid of the predominant tea pest, and is regarded as the most promising candidate for both augmentative and conservative biological control in tea plantations. However, little is presently known about its biology, ecology, and genetics.

**Methods:** Novel microsatellite markers were developed for *S. empoasca* from transcriptome sequences generated using high-throughput sequencing. The identified markers were then validated and characterized using 98 individuals from five geographically distinct populations through the tests of Hardy–Weinberg equilibrium and linkage disequilibrium as well as the analyses of genetic diversity, genetic differentiation, and gene flow.

**Results:** A total of 54,520 microsatellites were identified from 117Mb clean sequences. By assessing with five geographical populations, a total of 18 loci were demonstrated to be polymorphic, stable, and repetitive. The genetic variation over the 18 loci varied with allele number ranging from 2 to 7, polymorphism information content from 0.234 to 0.634; and observed and expected heterozygosity from 0.166 to 0.686 and 0.276 to 0.753, respectively. Meanwhile, the 18 loci showed a low frequency of null alleles ( $0 - 0.107$ ), and the inbreeding coefficient revealed a range from  $-0.143$  to  $0.173$ , with the exception of loci Ste36 ( $0.4076$ ). Based on analysis of these 18 loci, the assessed populations showed low to moderate levels of genetic differentiation, in which individuals clearly grouped into two clusters. And the recent dispersal rates between the geographically distinct populations were identified to be much lower ( $0.011 - 0.250$ ) than the individual gene exchange rates within a population ( $0.683 - 0.939$ ).

**Conclusion:** The identified 18 microsatellite markers could reveal a pattern of spatial structure and gene flow in *S. empoasca* populations according to geographic variability. This work provides an important basis for future studies on how these markers can be used in studies of the biology, genetics, and ecology of this important parasitoid. The findings can further provide important information for the development of biological control strategies in tea plantations. Additionally, our study reaffirms the importance and efficiency of high-throughput sequencing in microsatellite marker development for non-model species lacking reference genome information.

## 1. Background

Biological control is an important pest management strategy in agriculture, and its application has increasingly grown, with the public becoming more knowledgeable and less tolerant of the use of pest controls that are potentially harmful to people and the environment [1]. In tea plantations, the tea green leafhopper *Empoasca onukii* is the predominant pest, causing 15%–50% loss in the annual yield [2]. At present, control strategies for *E. onukii* have mainly relied on the use of synthetic insecticides, and efficient biological control strategies are still lacking. Field surveys have demonstrated that the solitary egg parasitoid *Stethynium empoasca* Subba Rao is the most common natural enemy of *E. onukii*, with a high rate of parasitism (up to 30%) [3, 4]. As an egg parasitoid, *S. empoasca* tend to offer more effective biological control as they can immediately prevent further herbivory, in contrast to most larvae parasitoids, that commonly keep the host alive for prolonging nutrient assimilation [5]. Thus, *S. empoasca* can be the most promising candidate for augmentative biological control, and also an object for conservation biological control. However, little is currently known about its biology, ecology, and genetics. Knowledge about the biological characteristics, ecological preference, genetic variation, and gene flow of parasitoid populations is needed in order to increase their adoption and success as biological control agents [6], as well as to favor the development of an effective habitat management strategy for parasitoid conservation [7].

The lack of research on *S. empoasca* may largely be due to the fact that it is small in size ( $597.15-637.23\mu\text{m}$ ) [8], and traditional observation methods could face challenges when studying such a minute parasitoid. Alternatively, molecular markers are capable of detecting minute quantities of parasitoid DNA. In fact, molecular marker-based techniques have been widely used for evaluating the genetic diversity and phylogenetic relationships of natural enemy populations, and are increasingly being recognized as valuable tools in facilitating ecological studies (e.g., population dynamics and dispersal) on small-sized parasitoids [6, 9]. Among the available markers, microsatellites are widely used owing to their various desirable traits, such as high mutation rate and polymorphism, co-dominant inheritance, and genome-wide distribution, and their use in analysis produces results with good reproducibility.

The traditional approach of microsatellite development using enriched libraries is time-consuming and labor-intensive [10]. Hence, high-throughput sequencing-based approaches can be a good alternative. The advantages of this technology are that it makes the development of novel microsatellite markers efficient and, importantly, it also makes it possible to investigate and characterize microsatellites at a genome-wide scale [11].

In this study, we aimed to identify and characterize microsatellites for *S. empoasca* from transcriptome sequences. This is the first studying reporting on novel microsatellites for *S. empoasca*. These developed markers will facilitate further biological, ecological, and genetic studies of *S. empoasca*.

## 2. Methods And Materials

### 2.1 Sample collection and DNA extraction

A total of 98 *S. empoasca* adult females were collected from five sampling sites in China (Table 2; Fig. 2), and were used for the identification of the polymorphic loci and in the performance of population genetic analyses. Genomic DNA were individually extracted with a TIANamp Micro DNA Kit (TIANGEN, Beijing, China) following the manufacturer's instructions. Additionally, 50 adult females collected from FZ were used for the initial test of selected primer pairs with genomic DNA, also extracted using the TIANamp Micro DNA Kit. Prior to use, these samples were stored in absolute ethanol and were frozen at  $-80^{\circ}\text{C}$ .

### 2.2 Sequencing, microsatellites mining, and primer design

A total of 100 *S. empoasca* adult females collected from FZ were used to generate the transcriptome sequencing libraries using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA). After clustering on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumina), the library

preparations were sequenced on an Illumina HiSeq platform. The generated transcriptome sequences were assembled using Trinity [12] with the default parameters.

The microsatellites were identified from the transcriptome sequences using MISA [13]. The microsatellite search criteria were as follows: at least 10, 6, 5, 5, and 5 repeats for the mono-, di-, tri-, tetra-, penta-, and hexa nucleotide motifs, respectively. The remaining parameters were used at default settings. An optimal primer set for each microsatellite was designed using Primer3 with default settings.

### 2.3 Primer testing and polymorphism detection

In total, 50 primer pairs were selected and synthesized for validation. The primer effectiveness and polymorphism were initially tested using the mixed genomic DNA of 50 *S. empoasca* adult females collected from FZ. The amplifications were prepared using a Golden Easy PCR System (TIANGEN, Beijing, China) in a final volume of 10  $\mu$ L, consisting of 5  $\mu$ L Master Mix, 0.5  $\mu$ L forward primer, 0.5  $\mu$ L reverse primer, 0.5  $\mu$ L (12.5ng) template DNA, and 3.5  $\mu$ L ddH<sub>2</sub>O. The amplifications were performed in a BIO-RAD C1000 TOUCH™ thermal cycler under the following conditions: 4 min at 94°C; 30 cycles of 30s at 94°C, 30s at 56°C, and 30s at 72°C; with a final extension at 72°C for 7 min. The PCR products were visualized using electrophoresis on a 2% agarose gel for the primary screening of the primers, and those primers that successfully produced positive PCR amplifications were considered to be valid.

A total of 37 primers, selected in previous steps, were further used to test for polymorphism and 98 female adults from FZ, AX, and WYS were characterized (Table 2). The forward primers that were used were randomly tagged with one of the four universal primers (Tail A: GCCTCCCTCGCGCCA; Tail B: GCCTTGCCAGCCGC; Tail C: CAGGACCAGGCTACCGTG; Tail D: CGGAGAGCCGAGAGGTG) [14] at the 5' end, and the reverse primers were the same as previously used. Additionally, the corresponding universal primers labeled with fluorescent dye (Tail A—FAM; Tail B—TAMRA; Tail C—HEX; Tail D—ROX) were also used. Six or seven primers, tagged unique universal primers or different allele size ranges, were pooled into one group and co-amplified using multiplex PCR. The amplifications were prepared using a Multiplex PCR Kit (TIANGEN, Beijing, China) in a final volume of 10  $\mu$ L containing 1  $\mu$ L Multi HotStar Buffer, 0.8  $\mu$ L (40  $\mu$ M) dNTPs, 0.2  $\mu$ L Multi HotStar, 0.48  $\mu$ L or 0.56  $\mu$ L primer mix (0.02  $\mu$ L for each forward primer and 0.02  $\mu$ L for the corresponding tail, and 0.04  $\mu$ L for reverse primer), and 6.02  $\mu$ L or 5.94  $\mu$ L ddH<sub>2</sub>O. The amplifications were performed in a BIO-RAD C1000 TOUCH™ thermal cycler under the following conditions: 15 min at 95°C; 40 cycles of 30s at 94°C, 90s at 58°C, and 90s at 72°C; with a final extension at 72°C for 10 min. The fluorescently labeled PCR products were analyzed on an ABI 3730XL capillary sequencer (Applied Biosystems, USA) using the GeneScan 500 LIZ size standard (Applied Biosystems, USA) to score the fragment size. The genotypes and allele sizes were then visualized using GENEMAPPER 4.0 (Applied Biosystems, USA) and further checked manually. In the case of failed amplifications, reactions were repeated twice to obtain results in the case of occasional technical mistakes. Finally, 18 microsatellite loci were found to be polymorphic, and their genotypic data were selected for further analyses of the genetic characteristics of the population from which they corresponded to, such as in terms of genetic diversity and population structure.

### 2.4 Statistical analysis

The linkage disequilibrium (LD) between each pair of loci at each population and the Hardy–Weinberg equilibrium (HWE) for each locus at each population were tested using GENEPOP 4.7.0 [15]. Null allele frequencies were estimated for each locus within each population using the expectation maximization algorithm in FreeNA [16] with 1000 bootstrap replicates. Population genetic diversity indices, such as the total number of alleles ( $A_T$ ) and the unbiased expected heterozygosity ( $H_E$ ) [17] were estimated in GENCLONE 2.0 [18], while the observed heterozygosity ( $H_o$ ) and polymorphism information content (PIC) were obtained using the macros Microsatellite Tools [19]. Given the effect of sample size on diversity statistics, the standardized number of alleles ( $A_S$ ) for a minimum sample size of nine diploid individuals was calculated in GENCLONE using the rarefaction approach. The allelic richness ( $A_R$ ) and allelic richness of private alleles ( $P_{AP}$ ) were also computed on a minimum sample size of nine diploid individuals using the rarefaction method in HP-RARE 1.1 [20]. Additionally, the inbreeding coefficient ( $F_{IS}$ ) for each population was reported using FSTAT 2.9.3 [21].

Population differentiation was first investigated by calculating Nei's genetic distance [17, 22] and Provesti's genetic distance of paired populations using the R package 'adegenet' 2.0.0 [23]. The pairwise  $F_{ST}$  values estimated in ARLEQUIN 3.5.2 [24] and the corrected pairwise  $F_{ST}$  values computed using the ENA correction method [16] in FreeNA were compared in order to examine the effect of null alleles on the population differentiation analyses. The population structures were then explored using three independent programs: STRUCTURE 2.3.4 [25], Discriminant Analysis of Principal Components (DAPC) [26], and TESS3r [27, 28]. STRUCTURE was run using the admixture model with correlated allele frequencies and no prior location information. The clustering test was performed using 30 independent runs for each  $K$  value (from 1 to 10) and 200,000 Markov chain Monte Carlo (MCMC) iterations after a burn-in of 100,000 iterations for each run. The delta  $K$  method [29] was used to determine the optimal value of  $K$ . The membership coefficient matrices (Q-matrices) of the inferred optimal  $K$  values were combined using CLUMPP 1.1.2 [30], and then visualized using DISTRUCT 1.1 [31]. The model-free analysis, DAPC, was performed to interpret and visualize the genetic correlation among individuals defined by sampling sites by retaining 200 principal components and all of the linear discriminants. The analysis was finished in R package 'adegenet' 2.0.0. As a spatially explicit clustering method, TESS3r was performed by incorporating the specific geographic coordinates with the genotype data. A similar range of  $K$  values (from 1 to 10) was tested with 30 independent runs for each  $K$ , and the default cross-validation criterion was used to evaluate the model performance. Finally, the data were further analyzed to evaluate the contribution of the geographic distance on *S. empoasca* genetic differentiation among populations using GLMM in the R package 'stats' [32] and a simple Mantel test in the R package 'ecodist' [33]. A significant result would provide direct evidence for a pattern of isolation by geographic distance (IBD). The geographic distances were computed as the shortest distance between two sample sites using the "Vincenty (ellipsoid)" method [34] in R package 'GEOSPHERE' 1.5-10 [35], while the genetic dissimilarity was developed using pairwise  $F_{ST}$ , Nei's, and Provesti's genetic distance.

BAYESASS 3.0.4 [36] was used to estimate the dispersal rates over a few generations of populations. Ten runs were conducted to combine the results with 20,000,000 MCMC interactions, sampled every 100 interactions after a burn-in of 2,000,000 for each run.

### 3. Results And Discussion

#### 3.1 Microsatellites in transcriptome sequences

In total, 44,388,868 raw reads of paired-end (PE) sequences were generated from transcriptome libraries constructed by the Illumina HiSeq platform. Removal of adapter and poly-N sequences as well as low-quality reads resulted in 6.51 GB of sequence, including 43,375,184 clean reads. After assembly, a total of 116,907,081 bp sequences, including 68,134 unigenes, were produced for microsatellite mining. The unigenes had a mean size of 1716 bp and an N50 of 2676 bp, which indicates they are much longer than those of previous studies on *Parantica sita* (unigenes = 886 bp) [37], *Dorcus hopei* (N50 = 1218 bp) [38] and *Carposina sasakii* (N50 = 286 bp) [10]. This difference may be explained by the coverage of the sequencing reads and the method of assembly.

About 40% (27,172) of the assembled genes were found to contain microsatellite sequences, resulting in a total of 54,520 candidate microsatellites being identified. The total relative abundance was 466.35 loci/Mb, which is higher than that of *Frankliniella occidentalis* (322.01 loci/Mb) [11] detected in the genome and *Pachypeltis micranthus* (40.47 loci/Mb) [39] in the transcriptomic expressed sequence tags, as well as for *Aphis glycines* (139.01 loci/Mb) [40] as identified in transcript sequences. However, it is unclear whether this difference is species-specific difference or due to the different methods used for sequencing and assembly. Nevertheless, our results indicate that microsatellites are abundant in *S. empoasca* and have great potential as a tool for further biological, ecological, and population studies. The identified microsatellites contained 40,021 (73.41%) mononucleotides, 8393 (15.39%) dinucleotide, 5764 (10.57%) trinucleotide, 265 (0.49%) tetranucleotide, 17 (0.03%) pentanucleotide, and 60 (0.11%) hexanucleotide motifs. After mononucleotide, dinucleotide motifs were the most abundant, followed by trinucleotide and tetranucleotide (Table S1). The quantity of microsatellites decreased as the corresponding motif repeats increased (Fig. 1). These findings are in agreement with previous studies on other insects, like *F. occidentalis* [11], *C. sasakii* [10], and *A. glycines* [40]. The 15 most abundant microsatellite motifs accounted for 92.75% of all of the detected microsatellites, which were (A)<sub>n</sub>, (T)<sub>n</sub>, (AT)<sub>n</sub>, (TA)<sub>n</sub>, (AAT)<sub>n</sub>, (TTA)<sub>n</sub>, (ATT)<sub>n</sub>, (TAA)<sub>n</sub>, (ATA)<sub>n</sub>, (TAT)<sub>n</sub>, (TC)<sub>n</sub>, (AG)<sub>n</sub>, (GA)<sub>n</sub>, (CAG)<sub>n</sub>, and (TG)<sub>n</sub>.

#### 3.2 Development and performance of microsatellite loci

The Primer3 [41, 42] generated a possible 47,490 primers for analysis of the detected microsatellites, of which 50 primer pairs (39 trinucleotide, 7 tetranucleotide, 2 pentanucleotide, and 2 complex repeat types) were randomly selected for the validation test. In the test, 13 primer pairs did not produce any visible amplification and 10 failed to produce stable amplification, while 9 loci were monomorphic and the amplification of 18 showed they were polymorphic (Table 1).

Through the assessment of five natural *S. empoasca* populations (N=98; XC, QCC, TF, FZ, and AX; Table 2 and Fig. 2), the average frequency of the null alleles for the 18 polymorphic loci was determined to range from 0 to 0.107 (Table 1), yielding little effect on the estimation of the genetic diversity and structure [16]. Among the 18 loci, 38 of the 765 locus–locus pairs in a population showed linkage disequilibrium (LD;  $P < 0.05$ ; Table S2), and 7 of the 153 locus pairs across all populations (Ste24–Ste47, Ste24–Ste10, Ste49–Ste50, Ste26–Ste21, Ste47–Ste44, Ste26–Ste44, and Ste9–Ste44) exhibited linkage disequilibrium ( $P < 0.05$ ; Table S2). However, none of the locus pairs were consistently significant in all five populations. Four of the 95 loci–population pairs deviated from Hardy–Weinberg equilibrium (HWE;  $P < 0.05$ ; Table S3), and two (Ste26 and Ste36) of the loci across all populations deviated from HWE ( $P < 0.05$ ; Table 1). However, a global HWE test across all loci and populations showed non-significant deviation ( $P = 0.5992$ ), and the test on each of the populations based on all loci also indicated non-significant deviation ( $P > 0.05$ ; Table 2). The presence of LD and the deviation from HWE are not necessarily due to the characteristics of the loci [10]. They can be caused by a variety of reasons, such as the Wahlund effect [43], the structure of the subpopulations or occurrence of a bottleneck [9, 44], or simply the biological properties of the species [10].

The genetic variation over the 18 polymorphic loci inferred based on four of the populations varied with allele number ( $A_T$ ), ranging from 2 to 7 (mean of 4.06); average allelic richness ( $A_R$ ; for nine specimens of each population), from 2 to 4.17; and observed ( $H_O$ ) and expected ( $H_E$ ) heterozygosity from 0.166 to 0.686 and 0.276 to 0.753, respectively (Table 1). These seem to be species-specific results, as both similar (e.g., *C. sasakii*) [10] and comparatively different (e.g., *F. occidentalis*, *Trichinella spiralis*) [11] [45] data have previously been reported in other studies. The loci with polymorphism information content (PIC) values of  $< 0.25$  and  $> 0.5$  are commonly considered to indicate low- and high polymorphism, respectively [9]. The PIC values of our study loci ranged from 0.250 to 0.634, except for the loci of Ste47 (0.234) and Ste42 (0.243; Table 1), indicating a medium-to-high level of polymorphism. The inbreeding coefficient ( $F_{IS}$ ) revealed a range from  $-0.143$  to 0.173, with the exception of loci Ste36 (0.4076, Table 1). The high  $F_{IS}$  in Ste36 might be due to the low  $H_O$  (0.166; Table 1) [46].

The genetic diversity over *S. empoasca* populations, estimated based on the 18 loci with varied  $A_T$ , ranged from 85 to 99,  $A_S$  (the standardized number of alleles for a minimum sample size of nine diploid individuals) from 77.547 to 85.682,  $A_R$  from 2.76 to 2.95,  $H_O$  from 0.424 to 0.476,  $H_E$  from 0.448 to 0.522, and  $P_{AR}$  from 0.03 to 0.17 (Table 2). The populations showed a low level of  $F_{IS}$ , with values ranging from 0.038 to 0.057 (Table 2), indicating a low chance of the study individual coming from the same egg brood and, thus, suggesting the effectiveness of the sampling strategy and the analyzed results. Because of the low frequency of null alleles, the pairwise  $F_{ST}$  estimation yielded similar results both with and without using the *ENA* correction method, with values ranging from 0.0106 (TF–AX) to 0.099 (FZ–QCC; Fig. 3a). Comparatively higher values ( $> 0.5$ ) were detected between FZ and other populations, while lower values were detected between TF and others. Consistently, the Nei's and Provesti's genetic distance among populations ranged from 0.034 to 0.150 and 0.131 to 0.264, respectively, with the largest values between FZ and other populations, and the lowest between TF and others (Fig. 3b). These findings suggest a moderate differentiation between FZ and other populations, and a low differentiation between TF and others [47]. This conclusion was further confirmed by the population structure analyses using both spatially and non-spatially clustering models, as well as a model-free program. The three methods consistently and clearly divided the 98 individuals into two distinct clusters, with one cluster containing samples from FZ and another composed of individuals from the other sites (Fig. 4 and S1). Meanwhile, gene flow analysis provided additional and more direct evidence. Multiple runs in BayesAss consistently indicated that the recent dispersal rates between the geographically distinct populations were much lower (0.011–0.250) than the individual gene exchange rates within a

population (0.683–0.939; Fig. 5). This suggests that *S. empoasca* tends undergo gene exchange within populations with little gene flow occurring between. The TF population showed a somewhat different pattern, with some gene flow towards XC, QLL, and AX, although it still had relatively low rates of 0.248, 0.250, and 0.199, respectively (Fig. 5). The distinct spatial structure and limited gene flow between the geographically distinct populations was expected, as small insects are commonly known to lack a strong natural ability to disperse [9, 48, 49]. Unexpectedly, both the logit link generalized linear model (GLMM) and simple Mantel analysis indicated that the geographic distance was not significantly correlated with the pairwise  $F_{ST}$ , Provesti's genetic distance, and Nei's genetic distance (Table S4), suggesting a lack of isolation by distance (IBD) pattern in *S. empoasca*. This might be caused by the insufficient number of populations examined or by the relevant effect of human activity. Nevertheless, a pattern of genetic differentiation and gene flow was found based on the 18 loci in this study, indicating that these microsatellite markers could be used to discriminate geographic populations as well as to facilitate other genetic and ecology studies on *S. empoasca*.

## 4. Conclusions

In this study, we developed and characterized 18 microsatellite markers for a parasitoid of the predominant tea pest from the transcriptome sequences generated using high-throughput sequencing. By assessing 98 individuals from five geographic ally distinct populations, 18 loci were shown to be polymorphic, and were verified to be able to reveal a pattern of spatial structure and gene flow in *S. empoasca* populations. This work provides an important basis for future studies on how these markers can be used in studies of the biology, genetics, and ecology of this important parasitoid. These findings can further provide important information for the development of biological control strategies in tea plantations. Additionally, our study reaffirms the importance and efficiency of high-throughput sequencing in microsatellite marker development for non-model species without a reference genome.

## Declarations

**Author Contributions:** JYL and LQS conceived and designed the research. JYL and JC conducted experiments. JYL and SJY contributed new reagents and/or analytical tools. JYL analyzed data. JYL, SJY and MSY wrote and revised the manuscript. All authors read and approved the manuscript.

**Acknowledgements:** We are grateful to Liwei Han and Haifang Hefor their help with the sample collections. Funding for this research was provided by the National Key Research and Development Program of China (2019YFD1002100, 2016YFE0102100), and Development and Reform Commission Agricultural 'Five New' Program of Fujian, China (Minfa Reform Agriculture [2017]410).

**Funding:** This research was funded by the National Key R & D Program of China (grant no. 2019YFD1002100 and 2016YFE0102100); the Development and Reform Commission of Fujian Province, China (Minfa Reform Agriculture grant no. [2017]410).

**Availability of data and materials:** Data for this study are available at: to be completed after manuscript is accepted for publication.

**Ethics approval and consent to participate:** Not applicable.

**Consent for publication:** Not applicable.

**Competing interests:** The authors declare that they have no competing interest.

## References

1. Hajek AE, Eilenberg J. Natural enemies: an introduction to biological control. 2nd edn. Cambridge: Cambridge University Press; 2018.
2. Chen LL, Yuan P, Pozsgai G, Chen P, Zhu H, You MS. The impact of cover crops on the predatory mite *Anystis baccarum* (Acari, Anystidae) and the leafhopper pest *Empoasca onukii* (Hemiptera, Cicadellidae) in a tea plantation. *Pest Manage Sci.* 2019;75:3371-80.
3. Mao YX, Zou W, Ma XH, Gao MQ, Lin NQ. The egg parasitoids of the small green leafhopper, *Empoasca vitis*, and their population dynamics. *Chin Bull Entomol.* 2008;45:472-4. (in Chinese with English abstract)
4. Li HL, Lin NQ, Guo JX, Wang SJ, Wang QS, Zeng MS, et al. Effects on small green leafhopper and its natural enemy Mymarid in tea plantations with intercropping of green manure plants. *Chin J Biol Control.* 2016;32:50-4. (in Chinese with English abstract)
5. Perović DJ, Gámez-Virués S, Landis DA, Wäckers F, Gurr GM, Wratten SD, et al. Managing biological control services through multi-trophic trait interactions: review and guidelines for implementation at local and landscape scales. *Biol Rev.* 2018;93:306-21.
6. Garipey T, Kuhlmann U, Gillott C, Erlandson M. Parasitoids, predators and PCR: the use of diagnostic molecular markers in biological control of Arthropods. *J Appl Entomol.* 2007;131:225-40.
7. Lavandero B, Figueroa CC, Franck P, Mendez A. Estimating gene flow between refuges and crops: a case study of the biological control of *Eriosoma lanigerum* by *Aphelinus mali* in apple orchards. *PLoS One.* 2011;6:e26694.
8. Li HL, Lin NQ. Studies on the biology of the egg parasitoids of tea leafhopper, *Empoasca vitis* (Göthe). *J Tea Sci.* 2019;28:407-13. (in Chinese with English abstract)
9. Liu T, Ke F, You S, Chen W, He W, You M. Isolation and Characterization of Microsatellite Loci for *Cotesia plutellae* (Hymenoptera: Braconidae). *Insects.* 2017;8:63.
10. Wang YZ, Cao LJ, Zhu JY, Wei SJ. Development and characterization of novel microsatellite markers for the peach fruit moth *Carposina sasakii* (Lepidoptera: Carposinidae) using next-generation sequencing. *Int J Mol Sci.* 2016;17:362.

11. Cao LJ, Li ZM, Wang ZH, Zhu L, Gong YJ, Chen M, et al. Bulk development and stringent selection of microsatellite markers in the western flower thrips *Frankliniella occidentalis*. *Sci Rep*. 2016;6:1-8.
12. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol*. 2011;29:644.
13. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. *Bioinformatics*. 2017;33:2583-5.
14. Blacket M, Robin C, Good R, Lee S, Miller A. Universal primers for fluorescent labelling of PCR fragments—an efficient and cost-effective approach to genotyping by fluorescence. *Mol Ecol Resour*. 2012;12:456-63.
15. Rousset F. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol Ecol Resour*. 2008;8:103-6.
16. Chapuis M-P, Estoup A. Microsatellite null alleles and estimation of population differentiation. *Mol Biol Evol*. 2007;24:621-31.
17. Nei M. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*. 1978;89:583-90.
18. Arnaud-Haond S, Belkhir K. GENCLONE: a computer program to analyse genotypic data, test for clonality and describe spatial clonal organization. *Mol Ecol Notes*. 2007;7:15-7.
19. Park S. Trypanotolerance in West African cattle and the population genetic effects of selection. University of Dublin. Ph. D. Thesis (In Prep.); 2001.
20. Kalinowski ST. HP-RARE 1.0: a computer program for performing rarefaction on measures of allelic richness. *Mol Ecol Notes*. 2005;5:187-9.
21. Goudet J. FSTAT, a program to estimate and test gene diversity and fixation indices. 2001. <http://www2.unil.ch/popgen/softwares/fstat.Htm>. Accessed 03 Jan 2020.
22. Nei M. Genetic distance between populations. *Am Nat*. 1972;106:283-92.
23. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*. 2008;24:1403-5.
24. Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. 2010;10:564-7.
25. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155:945-59.
26. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet*. 2010;11:94.
27. Caye K, Jay F, Michel O, François O. Fast inference of individual admixture coefficients using geographic data. *Ann Appl Stat*. 2018;12:586-608.
28. Caye K, Deist TM, Martins H, Michel O, François O. TESS3: fast inference of spatial population structure and genome scans for selection. *Mol Ecol Resour*. 2016;16:540-8.
29. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol*. 2005;14:2611-20.
30. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*. 2007;23:1801-6.
31. Rosenberg NA. DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes*. 2004;4:137-8.
32. R core team. The R Stats Package version 3.6.2. 2020. <https://www.rdocumentation.org/packages/stats/versions/3.6.2>. Accessed 08 Aug 2020.
33. Goslee SC, Urban DL. The ecodist package for dissimilarity-based analysis of ecological data. *J Stat Softw*. 2007;22:1-19.
34. Vincenty T. Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Surv Rev*. 1975;23:88-93.
35. Hijmans RJ. Introduction to the "geosphere" package. 2012. <https://cran.r-project.org/web/packages/geosphere/>. Accessed 03 Jan 2020
36. Wilson GA, Rannala B. Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*. 2003;163:1177-91.
37. Hu P, Huang CL, Luo MX, Hsu YF, Wang RJ. Development and characterization of novel microsatellite markers in chestnut tiger butterfly *Parantica sita* (Lepidoptera: Nymphalidae) using next-generation sequencing. *Appl Entomol Zool*. 2020:1-6.
38. Kang TH, Han SH, Park SJ. Development of seven microsatellite markers using next generation sequencing for the conservation on the Korean population of *Dorcus hopei* (E. Saunders, 1854) (Coleoptera, Lucanidae). *Int J Mol Sci*. 2015;16:21330-41.
39. Zhu JY, Ze SZ, Cao LJ, Yang B. Development of microsatellite markers for the plant bug, *Pachypeltis micranthus* (Hemiptera: Miridae). *Appl Entomol Zool*. 2016;51:327-31.
40. Bai X, Zhang W, Orantes L, Jun T-H, Mittapalli O, Mian MR, et al. Combining next-generation sequencing strategies for rapid molecular resource development from an invasive aphid species, *Aphis glycines*. *PloS one*. 2010;5:e11370.
41. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3—new capabilities and interfaces. *Nucleic Acids Res*. 2012;40:e115-e.
42. Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. *Bioinformatics*. 2007;23:1289-91.
43. Dharmarajan G, Beatty WS, Rhodes Jr OE. Heterozygote deficiencies caused by a Wahlund effect: dispelling unfounded expectations. *J Wildlife Manage*. 2013;77:226-34.
44. Arunkumar K, Kifayathullah L, Nagaraju J. Microsatellite markers for the Indian golden silkworm, *Antheraea assama* (Saturniidae: Lepidoptera). *Mol Ecol Resour*. 2009;9:268-70.
45. Li TT, Tang B, Bai X, Wang XL, Luo XN, Yan HB, et al. Development of genome-wide polymorphic microsatellite markers for *Trichinella spiralis*. *Parasites Vectors*. 2020;13:58.
46. Hamarsheh O, Karakuş M, Azmi K, Jaouadi K, Yaghoobi-Ershadi MR, Krüger A, et al. Development of polymorphic EST microsatellite markers for the sand fly, *Phlebotomus papatasi* (Diptera: Psychodidae). *Parasites Vectors*. 2018;11:1-5.

47. Wriht S: Evolution and the Genetics of Population, Variability within and among Natural Populations. Chicago: The University of Chicago Press; 1978.
48. Lo N, Montagu A, Noack A, Nahrung H, Wei H, Eldridge M, et al. Population genetics of the Australian eucalypt pest *Thaumastocoris peregrinus*: evidence for a recent invasion of Sydney. J Pest Sci. 2019;92:201-12.
49. Wei SJ, Zhou Y, Fan XL, Hoffmann AA, Cao LJ, Chen XX, et al. Different genetic structures revealed resident populations of a specialist parasitoid wasp in contrast to its migratory host. Ecol Evol. 2017;7:5400-9.

## Tables

<b>Table 1</b> Characteristics of 18 polymorphic microsatellite loci validated in 98 <i>Stethynium empoasca</i> individuals from five geographic populations.												
locus	Dye	Repeat Motif	Primer Sequence	$N_A$	Size Range (bp)	$HWE$	Null	PIC	$H_O$	$H_E$	$A_R$	$F_{IS}$
Ste49	5'FAM	(AAC)8	F: GCCACCTTTCCACTTGCGATG R: GAAAATGGGCACGACGACAC	4	240-258	0.970	0.001	0.467	0.575	0.578	2.951	-0.022
Ste4	5'TAMRA	(ACT)12	F: TCGCGTCAAACCTGCTTAT R: ACAAAGTGACTTGTGCGCATCA	5	222-237	0.408	0.047	0.505	0.480	0.668	4.141	0.150
Ste3	5'HEX	(ATA)8	F: TCAATTTCTTCTCGTTGCACAATTCT R: AATCCAATACGCGTGCCAAT	4	257-272	0.571	0.027	0.490	0.603	0.630	3.148	-0.079
Ste42	5'FAM	(ATAAA)6	F: TGGTTTTCGCTAGTTTCGGATTG R: CGCTCACAAAATCTGAGTGGG	4	276-291	0.978	0.020	0.243	0.248	0.312	2.791	0.071
Ste24	5'TAMRA	(AAT)8	F: TGCTCTTTAGTCGTCAGGAACA R: AGCGCAGTTATTGTAGCGAA	4	235-244	0.386	0.043	0.501	0.505	0.619	3.346	0.169
Ste47	5'HEX	(ATAA)5	F: AAACCATGGGATAGAGGCAC R: CTCGCTGAGGCAATGCTAGA	3	173-181	1.000	0.000	0.234	0.308	0.276	2.053	-0.106
Ste32	5'ROX	(GAT)8	F: AGCAAAGAGGCCAGTCACTC R: CCAGAAATGCACATCCGATTCC	3	250-259	0.671	0.047	0.341	0.376	0.462	2.092	0.173
Ste10	5'FAM	(CAA)11	F: ACGATCTTCAAAGGGTTCGACT R: ACCTGATTGGCAGCTCTCTG	2	173-176	0.731	0.042	0.371	0.439	0.502	2.000	0.168
Ste8	5'TAMRA	(ATT)8	F: TGAGAATGCGCTGCCAAAAA R: ACAGCTGTACATCAATATAACGT	3	270-282	0.664	0.032	0.313	0.329	0.408	2.828	0.048
Ste26	5'HEX	(AAT)8	F: GGGTCGCGCCAAATATTTTC R: AGGTGAGCCGTTTTAGGACG	4	192-201	0.002*	0.050	0.421	0.500	0.555	2.809	0.103
Ste50	5'ROX	(ATT)8	F: AGTTGTACGCAGGACCTTT R: GACTGGCATTGTTTTCCGCA	7	237-261	0.388	0.035	0.499	0.539	0.577	4.109	0.007
Ste21	5'FAM	(AAT)8	F: TCAATGTATATGTACAACGTATATGCA R: AGAGGGAATAACGCGTGGTT	4	232-241	0.132	0.058	0.489	0.506	0.582	3.266	0.170
<b>Table 1 Cont.</b>												
locus	Dye	Repeat Motif	Primer Sequence	$N_A$	Size Range (bp)	$HWE$	Null	PIC	$H_O$	$H_E$	$A_R$	$F_{IS}$
Ste9	5'TAMRA	(AT)7(ATA)5	F: ACTTGCCTTACCAGCTGAGT R: ACATGCCTGAAAATGTTGTCTTCA	4	238-247	0.829	0.024	0.377	0.424	0.473	2.895	0.071
Ste44	5'HEX	(ATT)8	F: ACGGGGATTTGGCAATGAGT R: CTAAGCCACCGTACAAGG	6	238-262	0.977	0.000	0.385	0.468	0.453	3.251	-0.045
Ste14	5'ROX	(ATA)7	F: AAAACAGCAGAACCGAGGGG R: ACGTTGAACCGTAGCGCATA	4	234-243	0.814	0.000	0.401	0.547	0.512	3.012	-0.143
Ste19	5'TAMRA	(AAT)10	F: CATGTGGTGTCTTTAGGAATGATTC R: ACTTGAACAATGTTCTGCCTAA	3	212-221	0.111	0.062	0.383	0.398	0.526	2.639	0.148
Ste36	5'TAMRA	(AAT)8	F: TGCATTATTCACGCCTGATGA R: CGAACGCCAGCAAATGTGAT	3	146-152	0.009*	0.107	0.250	0.166	0.313	2.362	0.408
Ste35	5'HEX	(AAT)8	F: GCCTGTAGCGGGTTGATTA R: CGAAAACAGGCAGCATACCG	6	239-269	0.612	0.028	0.634	0.686	0.753	4.170	0.039

F: forward primer; R: reverse primer;  $N_A$ : number of alleles;  $HWE$ : exact  $p$ -value of Hardy-Weinberg Equilibrium ( $*p < 0.05$ ); Null: frequency of null allele; PIC: polymorphism information content;  $H_O$ : observed heterozygosity;  $H_E$ : expected heterozygosity;  $A_R$ : average allelic richness (for 9 specimens per sample);  $F_{IS}$ : inbreeding coefficient.

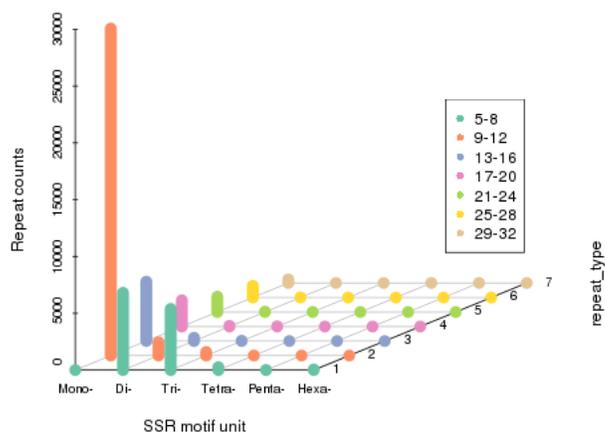
**Table 2** Statistics for *Stethynium empoasca* populations screened with the 18 polymorphic microsatellite loci.

Population	Longitude	Latitude	Sample size	$H_{WE}$	$H_O$	$H_E$	$A_T$	$A_S$	$A_R$	$P_{AR}$	$F_{IS}$
XC	117.915436	27.638767	15	0.956	0.424	0.448	85.000	81.455	2.760	0.110	0.057
QL	117.955289	27.609211	15	0.893	0.430	0.459	80.000	77.547	2.700	0.030	0.064
TF	120.388308	27.145455	25	0.106	0.446	0.485	99.000	85.682	2.900	0.170	0.070
FZ	119.246884	26.087619	30	0.050	0.476	0.522	93.000	76.138	2.950	0.170	0.089
AX	117.8739722	25.00241667	13	0.840	0.474	0.494	95.000	95.000	2.880	0.120	0.038

$H_{WE}$ : exact  $p$ -value of Hardy-Weinberg Equilibrium ( $*p < 0.05$ );  $H_O$ : observed heterozygosity;  $H_E$ : expected heterozygosity;  $A_T$ : total number of alleles;  $A_S$ : standardized total number of alleles (for 9 specimens);  $A_R$ : average allelic richness (for 9 specimens);  $P_{AR}$ : private allelic richness;  $F_{IS}$ : inbreeding coefficient. Refer to Fig. 2 for the sample sites of the populations.

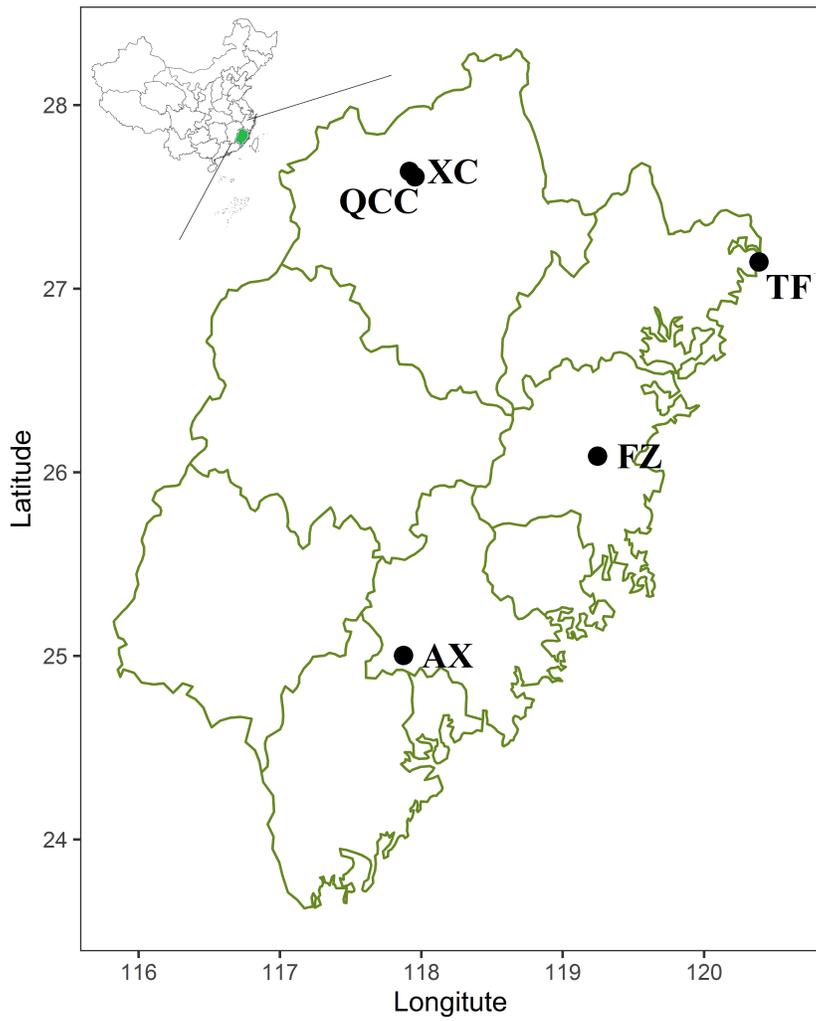
## Figures

**Distribution of SSR Motifs**

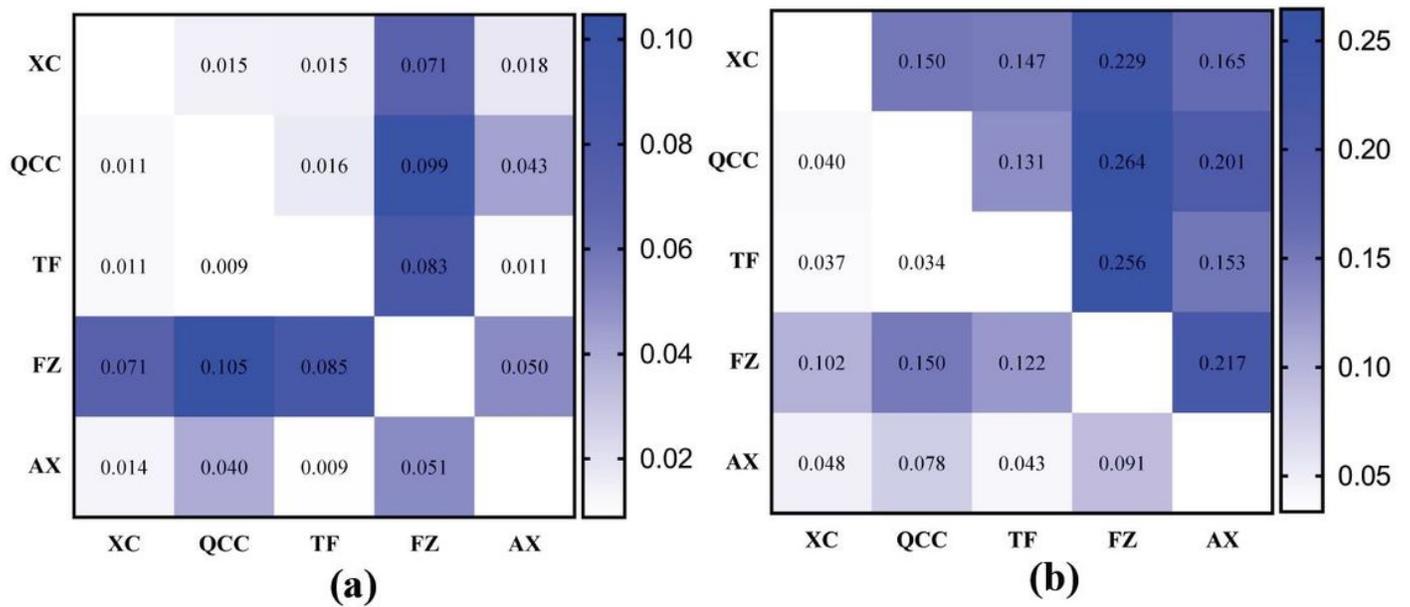


**Figure 1**

Counts of different microsatellite (SSR) motif types with different repeat type in *Stethynium empoasca*.

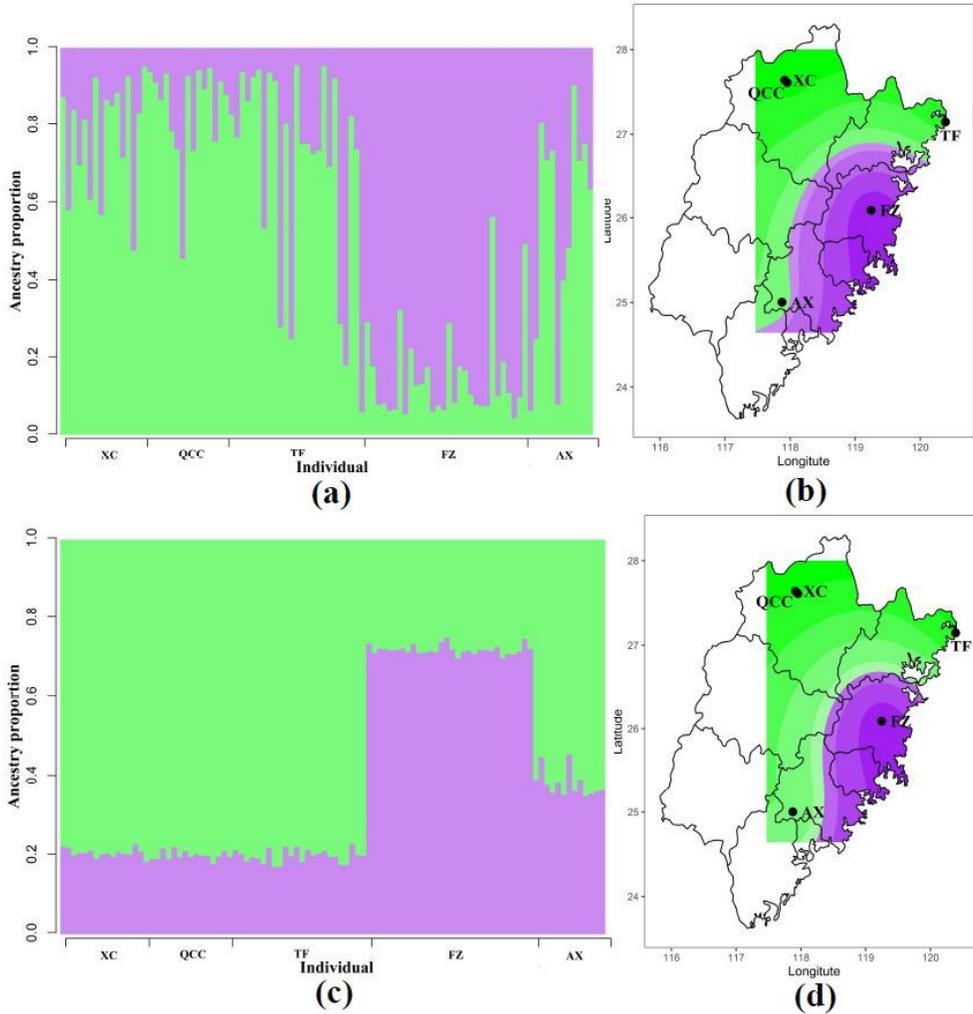


**Figure 2**  
 Sampling sites (or locations) of the five *Stethynium empoasca* populations used to access the 18 microsatellite loci in this study. Refer to Table 2 for the coordinates of the populations' sample sites.



**Figure 3**

Pairwise  $F_{ST}$  and genetic distance among the five populations of *Stethynium empoasca* estimated based on the 18 microsatellite loci. (a) above diagonal: pairwise  $F_{ST}$  estimated using the ENA correction method; below diagonal: pairwise  $F_{ST}$  estimated without using the ENA correction method. (b) above diagonal: Provesti's genetic distance; below diagonal: Nei's genetic distance. Refer to Table 2 and Fig. 2 for the population information.



**Figure 4**  
 Population genetic structure ( $K=2$ ) inferred by non-spatial (STRUCTURE; a, b) and spatial (TESS3r; c, d) clustering approaches based on 18 microsatellite loci. a, c: Individual's admixture assignments; b, d: Geographical patterns of the population differentiation. Refer to Table 2 and Fig. 2 for the population information.

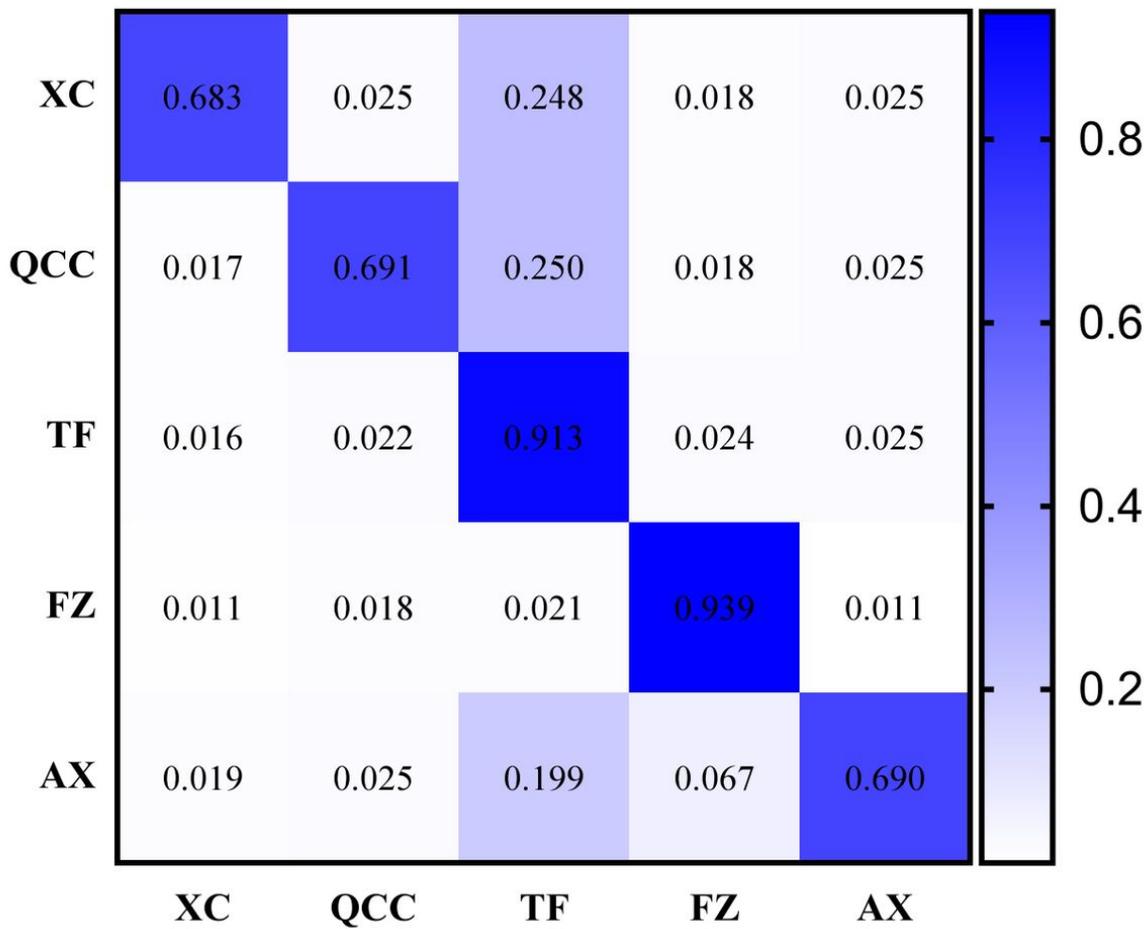


Figure 5

Dispersal rates over recent generation among *Stethynium empoasca* populations estimated by BayeAss based on the 18 microsatellite loci. Source populations are listed in horizontal coordinates and recipient populations in vertical coordinates. The values on the diagonal are self-recruitment rates for each of the populations. Refer to Table 2 and Fig. 2 for the population information.

### Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SumplymentaryfilesLietal20200824.zip](#)