


RESEARCH

Open Access



Whole-genome resequencing of tea grey geometrid provides insights into their population structure and adaptation to tea crops

Ruizhong Yuan^{1,2,5}, Yusi Chen^{1,2,3,5}, Xiaogui Zhou⁴, Xiaohan Shu^{1,2,3,5}, Zhaohe Lu^{1,2,5}, Pu Tang^{1,2,5}, Xiqian Ye^{1,2,5} and Zhizhi Wang^{1,2,5*} 

Abstract

Tea grey geometrid, *Ectropis grisescens* (Lepidoptera: Geometridae), is a chewing defoliator of high fecundity and gluttony, whose larvae heavily consume leaves and shoots, and consequently cause significant losses to tea crops in terms of both yield and quality. Nevertheless, a lack of population genome information has hindered previous research on the species as a tea crop pest. Here, we re-sequenced genomes of 43 tea grey geometrid individuals representing 13 distinct geographic regions and 12 individuals of its sibling species, the tea geometrid *Ectropis obliqua* (Lepidoptera: Geometridae), collected from four different sites in China. A genomic variation atlas of 627,569 single-nucleotide polymorphisms (SNPs) from whole-genome resequencing of 43 individuals was identified. Surprisingly, the populations of the tea grey geometrid (EG) are found to be assigned into two distinct subpopulations, EGA and EGB. The sample locations indicate that the subpopulation EGA is mainly distributed in the region of Zhejiang Province, where it co-occurred with tea geometrids (EO). The EGB subpopulation is distributed in the southern and central areas of China. Demographic history analysis has revealed fluctuations in population size of the tea grey geometrid over time, but following tea cultivation, the population size has started to stabilize. Meanwhile, the migration event analysis has uncovered a significant migration event between EGA and EO. Furthermore, genes (such as *P-glycoprotein* and *lactase*) under selection show possible metabolic adaptation of tea grey geometrid to the tea plants. These findings provide a broader foundation and additional resources for understanding the geometrid evolution and the genetic mechanisms underlying agriculturally important traits.

Keywords Tea grey geometrid, Genomic variation, Population genetics, Local adaptation, Evolution history

*Correspondence:

Zhizhi Wang
zzwang0730@zju.edu.cn

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Introduction

The tea, *Camellia sinensis* (Ericales: Theaceae), is a global economic crop [1–4], whose annual plantation acreage reached 3.31 million hectares in China, in 2022 (<https://www.statista.com>). As an integral component of traditional Chinese products, tea holds a significant position in the international production and trade market of China [5]. China is one of the birthplaces of the tea crop and has a history of thousands of years of tea cultivation. During the long-term evolutionary process, tea crops radiate and spread from the place of origin and are widely cultivated in more than 60 countries around the world through continuous artificial introduction and cultivation [6]. The domestication and selection of tea have resulted in different levels of genetic diversity and population differentiation, alongside the diversity level of morphological, physiological and biochemical characteristics, and sometimes biotic or abiotic stress resistance [7–9]. With many beneficial effects on human health, tea is prepared in a simple process and can be directly brewed for drinking [10, 11]. Therefore, tea products require extremely high standards of quality. At the same time, a number of variables, including pest infestation and its management, have a direct impact on the quantity and quality of tea products.

In China, tea grey geometrid, *Ectropis grisescens* (Lepidoptera: Geometridae) (EG), is the most important phyllophagous pest of tea crop [5], and seriously affects the yield and quality of tea crops and causes great economic losses to the tea production [12–19]. In addition, the tea geometrid, *Ectropis obliqua* (Lepidoptera: Geometridae) (EO), is a sibling species to the tea grey geometrid. They share a similar appearance and cause similar symptoms in tea crops [5, 20–23], but differ in geographic distribution—tea grey geometrid is extensively dispersed throughout major tea-producing areas of China while tea geometrid is primarily found near around Taihu Lake [14, 24]. The biological characteristics of the two sibling species also differ, including reproductive capacity, tolerance to virus infections, endosymbiont (*Wolbachia*) composition and virulence composition [25–30]. Besides, the interaction between geometrids and tea crops leads to reciprocal evolutionary change. For instance, the volatiles produced when tea grey geometrid larvae feed can stimulate neighbouring tea crops to emit beta-Ocimene, enhancing the ability of healthy neighbouring plants to repel adult tea grey geometrids [31]. On the other hand, the feeding of tea grey geometrid larvae induces a series of metabolic variations including flavonoids (catechin, quercetin, epigallocatechin gallate, etc.), resulting in a direct defense reaction, which is regulated by signaling pathways involving jasmonic acid, ethylene, and auxin [32, 33]. The cultivation of tea varieties

in China is mainly divided into two tea varieties, *Camellia sinensis* var. *sinensis* and *C. sinensis* var. *assamica*, which provide a good model for host-driven pest evolution. Besides, it is an essential element of integrated pest management that recognises the differences between different geographic populations of tea grey geometrid [34]. However, larger-scale genetic information analysis is still lacking.

In this study, we collected samples from major tea-growing regions in China where the tea grey geometrid has recently caused severe economic losses due to outbreaks. To elucidate the molecular mechanism behind the frequent outbreaks of *E. grisescens*, we employed next-generation sequencing technology to sequence the genomic data of the tea grey geometrid by analyzing the genetic diversity and population differentiation. We observed two distinct subpopulations (EGA and EGB) in the natural geographic distribution of the tea grey geometrid, with EGA mainly distributed in the region of Zhejiang Province (co-occurring with tea geometrids). Furthermore, we constructed a population genetic map and identified genes (such as *P-glycoprotein* and *lactase*) under selection that are possibly associated with tea grey geometrid's adaptation to tea crops. These results shed light on the evolution and genetic mechanisms of the tea grey geometrid, revealing important traits and interactions with tea crops. The genomic variation atlas produced in this study represent a comprehensive resource for future research on tea grey geometrid and related species.

Results

Population sampling and sequencing

A total of 43 tea grey geometrid individuals from 13 major geographic distribution areas were selected for whole genome resequencing, and the sample set covers a large part of the tea-growing region of China (Table S1) [14, 23]. Twelve tea geometrid samples were also sequenced as the outgroup [5, 20–23]. In total, we generated 3.31 Tb of data (11.03 billion read pairs) and mapped these data to the reference genome [13], resulting in an average sequencing depth of 34.75×. Based on the mapping results, we ultimately identified 6.27 million genetic variants in tea grey geometrid, including 627,569 high-quality single nucleotide polymorphisms (SNPs) and 131,054 small insertions and deletions (InDels). Our analysis revealed a global SNP density of one genetic variant per 1,013 bp on average. Interestingly, the intron regions accounted for 83.03% of the SNPs, while the exon region represented only 3.33% of the SNPs. Within the exon region, a total of 6,351 non-synonymous SNPs and 18,695 synonymous SNPs were identified. SNP distribution was observed across various chromosomes in the

genome, and high-density SNP regions were identified in Chr14, Chr18, Chr19, Chr23, and Chr31 (Fig. 1a). In addition, InDels density showed a similar trend as SNP density, with the highest density region presented in Chr18, suggesting comparable evolutionary and selection pressures on these genome fragments (Fig. 1b). In summary, we observed a relatively high level of whole-genome diversity within the tea grey geometrid population, indicating a high effective population size, thus creating a large genetic pool that may be useful in the future for biological control purposes [35].

Population structure and phylogenetic analysis of tea grey geometrid

To reveal the natural population relationships among various geographical populations of tea grey geometrid, we utilized the SNP dataset for further analysis. In general, population structure analysis (Fig. 2a, Fig. S1, Table S2), principal component analysis (PCA) (Fig. 2b, Table S3) and identity by state (IBS) genetic distance (Fig. 2c,

Table S4) supported the divergent of the tea grey geometrid samples. Samples from Zhejiang Province and Henan Province were assigned EGA ($n=17$), while individuals from other regions were named EGB ($n=26$). Phylogenetic analysis also showed all the tea grey geometrid samples were divided into two clusters (Fig. 2d). These results suggest that the populations of tea grey geometrid in the primary tea regions of China have differential into two subpopulations, one of which is mainly distributed in Zhejiang Province, while the other is evenly distributed throughout other areas (Table S1). Obviously, the population diversification would be a crucial factor to cause a higher genetic variation in tea grey geometrid.

Demographic history and migration event analysis

To investigate the factors leading to the divergence among the geometrid wild lineage (EGA, EGB and EO), we generated historical effective population size change curves (Fig. 3a). The demographic history of tea grey geometrid suggested a fluctuation in the estimated



Fig. 1 The density of SNPs/InDels of tea grey geometrid. **a** The density of SNPs in 31 chromosomes of tea grey geometrid. **b** The density of InDels of 31 chromosomes of tea grey geometrid

effective population size (N_e), characterized by a sequential contraction and expansion over time. After the last glacial period, the population sizes of EGA and EGB remained relatively stable, fluctuating between approximately 1,000 and 10,000. However, following the advent of tea cultivation [36], the effective population size of both subpopulations began to level off. During recent evolution, EGB has shown a similar diminutive effective population size to EGA. Genetic drift emerging within small subpopulations has been a potential contributing factor in the natural evolution of these populations. In addition, the population size change of EO exhibited a contrasting trend to that of EGA and EGB. In contrast, EO demonstrated persistent fluctuations since the last glacial period, suggesting that the effective population size of EO is vulnerable to external factors.

To explore possible genetic exchange, we performed a TreeMix analysis to infer population differentiation and mix model. The optimum value of the migration edge (m) ranging from 1 to 10 was evaluated. When delta m value

reached the smallest, the optimal m was determined to be 1. Further, the maximum likelihood tree and reside heat map were conducted with $m=1$ (Fig. S2-3). This value points to a significant migration event among EGA, EGB and EO. Interestingly, the migration event occurs between the population EG1 and EO, of which the former was found in Hangzhou, Zhejiang Province, and the latter was distributed around Taihu Lake (Fig. 3b) [14, 24]. The genetic drift parameter between EG1 and EO was found to be less than 0.002, indicating a resemblance in genetic structure between these two populations.

Selection signal as a potential force for forming subpopulation

We performed selection signal detection in all the EG samples based on population differentiation monitoring and locus frequency monitoring. The nucleotide polymorphism indices (θ_π) of the two subpopulations showed some differences (Table S5), proportional to the genotypic diversity within a single population, with EGA

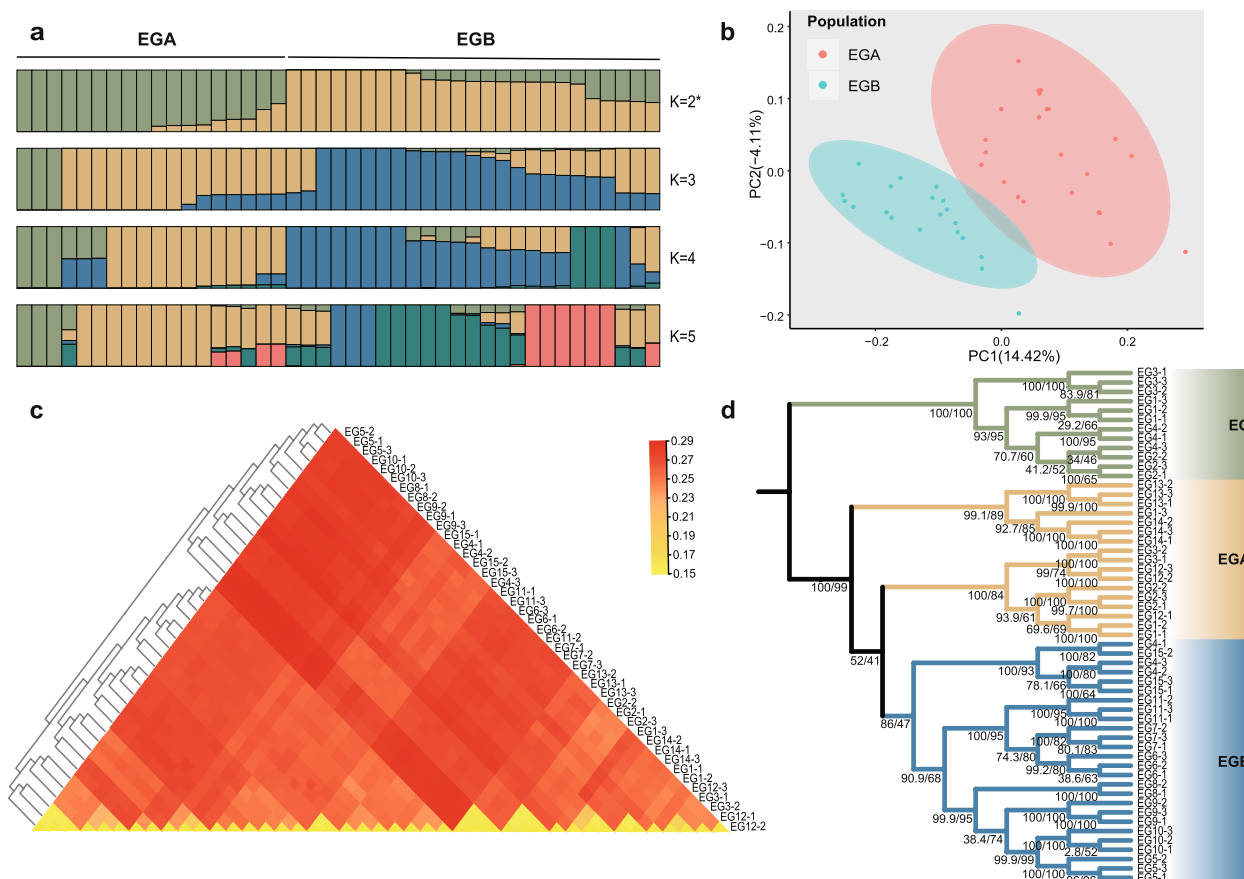


Fig. 2 Population structure of tea grey geometrid. **a** Admixture plot that indicates population structures for K from 2 to 5. “*” means that K=2 is the optimal K value. **b** PCA plot analysis of tea grey geometrid samples. The confidence interval was set to 95%. **c** IBS genetic distance of tea grey geometrid samples. **d** The phylogenetic analysis of all the samples using SNP data. The phylogenetic was constructed by the maximum likelihood (ML) method and SNP data of tea grey geometrid samples was used as outgroup (EO). The bootstrap/alert values are indicated on the node

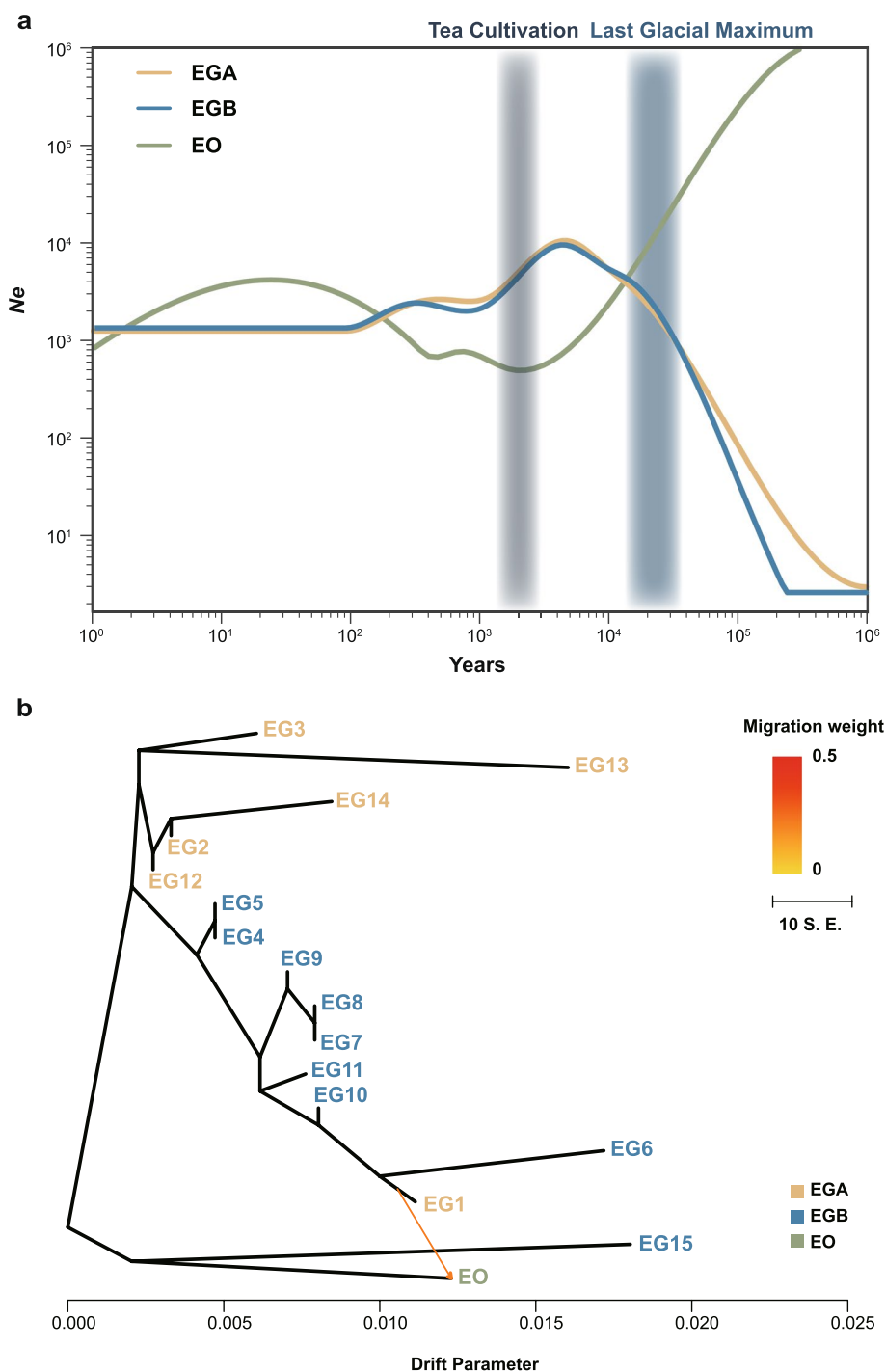


Fig. 3 Demographic history and migration event analysis. **a** The demographic analysis of tea grey geometrid and tea geometrid. Tea cultivation period and last glacial maximum period were marked as bar in plot with slight blue and weight blue, respectively. **b** The migration event in EGA, EGB and EO. Arrow means migration event orientation in different subpopulation and population. The color range of arrows means standard error (S. E.) value of migration weight

and EGB of the wild line being closer. The inter-population genetic differentiation index (F_{ST}) was calculated to obtain F_{ST} values [37], and the F_{ST} value between 0.25

and 1 for a single genetic window indicates a high genetic differentiation between populations (Fig. 4a, Table S6). Chr15 contains the highest number of SNP sites in the

top 5% of the F_{ST} windows, while no SNP was found in exon regions. We computed the genetic diversity (θ_{π}) and Tajima's D value (Table S7) on Chr15. According to the results of the π value shown in Fig. S4, we observed that regions near the both ends of chromosome 15 (the region 0~15 Mb and the region 25~35 Mb) exhibited a higher value of π , indicating increased genetic diversity towards the chromosome ends. As for Tajima's D value (Fig. S5), we observed that the majority of sites on chromosome 15 display positive values, indicating a higher frequency of intermediate mutations in the population. This could be attributed to influences such as population expansion, positive selection, or rare variants.

A selection of 5% F_{ST} windows yielded 222 genes (Table S8) and a total of 348 SNPs were found in their exon region. To understand the function of the selected genes, GO enrichment and KEGG pathway were performed. Functional annotation of these 222 candidate genes was conducted by EggNOG [38] (Table S8). Generally, the GO term of these genes mainly focused on the cellular process, biological regulation, regulation of biological process, etc. (Fig. 4b, Fig. S6), while the KEGG pathway

mainly focused on transcriptional regulation, protein processing and modification, etc. (Fig. 4c). Of note, several exon-occurring nonsynonymous SNPs were found in genes belonging to ABC transporters and carbohydrate metabolism. These genes includes two P-glycoprotein genes (*pgp-2*: *Egri011147.1* and *Egri010743.1*) from ABC transporter family, and two lactase genes (*LCT*: *Egri001325.1* and *Egri002194.1*) from glycoside hydrolase superfamily, which functions are related to digestion and xenobiotic metabolism [39, 40], such as the metabolism of catechins in geometrid moths [41]. In *Egri011147.1*, an SNP (chr18:166944, C to T) on the 13th base of the open reading frame (ORF) causes the encoded amino acid to change from leucine (L) to phenylalanine (F), and an SNP (chr18:159413, G to C) on the 223rd base of *Egri010743.1* ORF makes the encoded amino acid alanine (A) change to proline (P) (Fig. 4d). In addition, one amino acid change from serine (S) to proline (P) due to an SNP (chr05:28328974, T to C) on the 85th base of the ORF of *Egri001325.1* and leucine (L) to serine (S) owing to an SNP (chr26:10859177, T to C) on the 53rd base of *Egri002194.1* ORF (Fig. 4e). Furthermore, paralogous

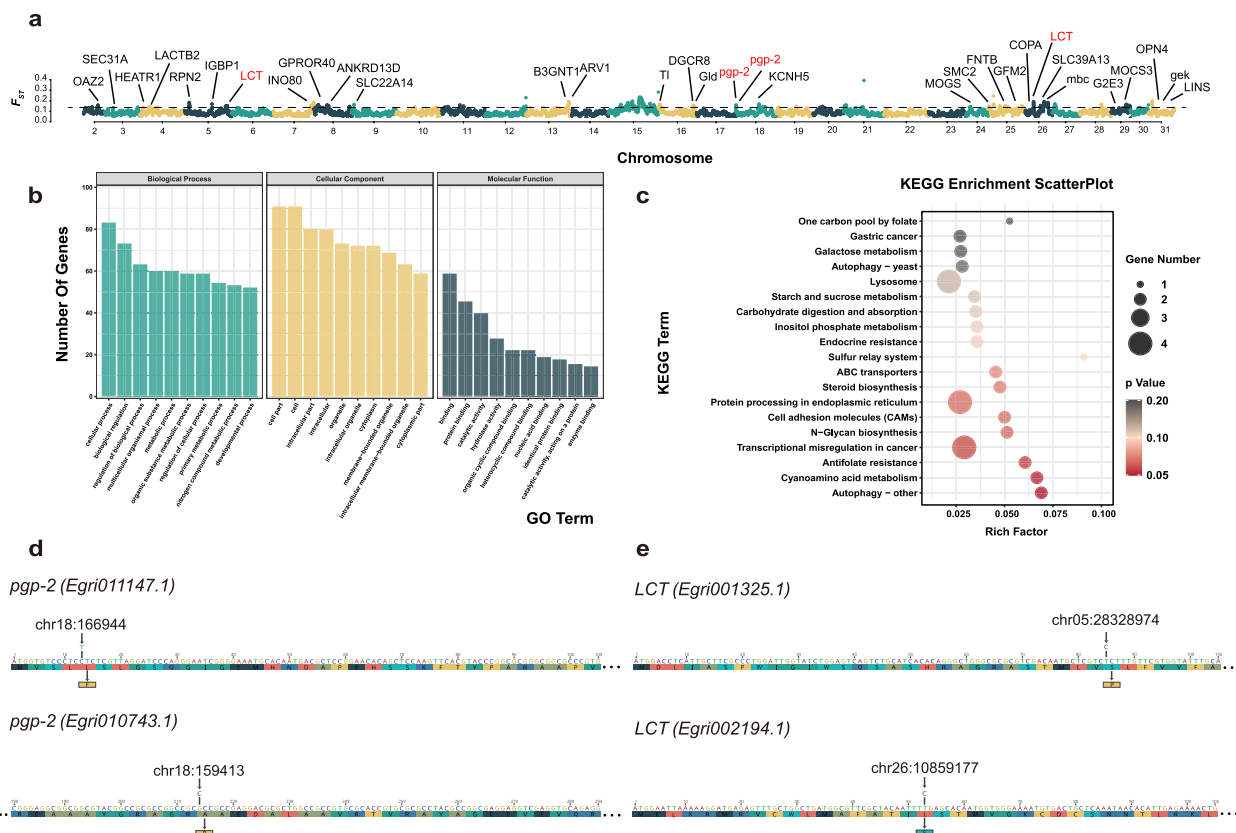


Fig. 4 Sweep selection of SNPs in the tea grey geometrid. **a** F_{ST} values. Window = 5000 kb, 5% as candidate genes, and >0.25 as key windows that influence relationship between each subpopulation. **b** GO enrichment of 222 candidate genes. **c** KEGG pathway enrichment of 222 candidate genes. **d** The SNP in two *pgp-2* genes. **e** The SNP in two *LCT* genes

genes of *pgp-2* and *LCT* were screened from the genome of the tea grey geometrid for natural selection analysis. Phylogenetic trees were constructed, and the dN/dS ratio was calculated (Fig. S7-8, Table S9-10). Within *pgp-2*, the dN/dS ratio was 0.292 at base 13 and 0.373 at base 223. Both ratios exceeded the overall average of 0.279. Regarding *LCT*, the dN/dS ratio was 0.408 at base 85 and base 53. Both ratios were higher than the mean value (0.297). These findings suggested a likelihood of natural selection occurring at these loci. Thus, we propose that the diversification of tea grey geometrid subsets might be influenced by the interplay between tea geometrid and tea crops, mediated by the interaction of *pgp-2* and *LCT*.

Discussion

In this study, the population structure, evolutionary patterns, and historical population changes of the wild lineage of the tea grey geometrid across different geographic populations were inferred. It was concluded that migration event at molecular level between tea grey geometrid and tea geometrid in the same geographic area played a significant role in subpopulation differentiation of wild lineage. The selective genes are likely associated with crucial physiological and biochemical processes that facilitate the insect's adaptation to their ecological niche within the tea plantation ecosystem. This study illuminated genomic characteristics and interactions with tea crops that have contributed to evolution and genetic mechanisms of the tea grey geometrid.

Tea crops mainly grow in central and southern China [42]. Based on the results of this study, tea grey geometrid in central and southern China diverged into two subpopulations, namely EGA and EGB. As a result, EGA is mainly distributed in Zhejiang Province (EG1, EG2 and EG3), while EGB is found in other regions. It is worth noting that the tea grey geometrid population in Nanyang, Henan Province (EG3) belongs to the EGA and shares a genetic origin with the population in Zhejiang Province, suggesting that EG3 may have originated from Zhejiang Province and spread to the region through commercial activities or tea crop transportation. The EGB largely corresponded to the Jiangnan Tea Area and the South China Tea Area.

As *E. obliqua* co-occurred in tea crops, we added it to the demographic history and migration event analysis. After the last glacial maximum, both EGA and EGB populations showed smaller fluctuations in their effective populations, indicating that temperature was the main factor influencing their population size before the last glacial maximum. Once humans started tea cultivation, tea became an important economic crop, leading to tea-dominated agricultural ecosystems [36]. It is evident that following the domestication of tea

crops, the effective population sizes of EGA and EGB have reached a stable state, closely associated with the emergence of tea-dominated agricultural ecosystems. On the other hand, the less distributed EO continued to fluctuate, suggesting that tea plants might also have an impact on the population size of these two species. This association is intricately connected to the widespread growth of tea cultivation and the relatively low proportion of wild tea crops. Besides, the resistance for different insecticides [43], development and reproduction [28, 29], and toxicity difference of microbiota [25–27] of the two sibling species also suggests EG is adapted well to domesticated tea crops. Consequently, the geographical distribution of tea grey geometrids is much larger compared to tea geometrids. However, the adaptability of EG and EO to the cultivation of tea needs further studies. Gene flow occurs between the subpopulation EGA and the sibling species EO. One possible factor is that EGA and EO are sympatric distributed in some areas which allows the sibling species to share similar habitats and therefore increase the likelihood of hybridization in the field [12], though it is demonstrated that the asymmetric mate of EG and EO produce sterile offspring. Corresponding to their hybrid results, the gene flow from EG to EO may have side effects on the EO population leading to narrower distribution.

We observed a relatively high level of whole-genome diversity within the tea grey geometrid genome population. The order Lepidoptera exhibits high genomic diversity [44–46], which may be a key reason for the significant number of SNPs occurring abundantly in coding regions. SNPs and InDels show a similar distribution pattern across chromosomes, with clustered occurrences on several chromosomes (Fig. 1b-c). This suggests that these chromosomal regions may harbor genes that contribute to intraspecific diversity and have great potential for exploration in the adaptation of tea grey geometrid to different tea cultivars.

From the enrichment analysis of GO and KEGG, it was observed that genes associated with subpopulation differentiation are predominantly found in metabolic pathways. The identified genes are primarily involved in the cellular process, biological regulation, regulation of biological process, etc. Additionally, the KEGG pathway analysis highlights transcriptional regulation, protein processing and modification, etc. The species of tea plants varied among samples, which are featured with diverse classes of metabolites [32, 47], which play an important role in the defense of tea plants against tea grey geometrid. For instance, the defense pathway involved in jasmonic acid and ethylene production in the tea plant is induced after the infestation of EG and

catechin compounds accumulated, which, in turn, reduces the fitness of EG larvae [32]. Though EG can digest these metabolites, the activity of certain digestive enzyme, such as *LCT*, is inhibited by catechins [48]. Thus, the exon-occurring nonsynonymous SNPs in *LCT* possibly present an evolutionary way to adapt to domesticated tea crops. Collectively, the findings suggest that the selective genes are likely involved in important physiological and biochemical processes that facilitate the insect's adaptation to its specific ecological niche within the tea plantation ecosystem.

The interaction between tea crops and tea grey geometrid is a natural phenomenon that involves considerations of adaptability and symbiosis. Of course, human activities can also profoundly affect the interaction between tea grey geometrid and tea crops, for example, long-term irrational use of chemical pesticides has led to an increase in resistance within the field populations of tea grey geometrids, further enhancing the likelihood of genetic variation occurring in these geometrids [49]. This could potentially have a significant impact on the yield and quality improvement of tea crops. Additionally, tea grey geometrids that escape during inspection and quarantine procedures may further spread within a larger geographical range through human activities. With a better understanding of the genetic adaptation of tea grey geometrid to tea crops, more effective strategies for pest management and crop protection can be developed, supporting sustainable agricultural practices [35].

Materials and methods

DNA sample preparation and sequencing

Tea grey geometrid samples were obtained from 13 different tea-producing areas and tea geometrid samples were obtained from four different tea-producing sites. All the samples were subjected to standard identification based on their distinct morphological characteristics and cytochrome c oxidase I (COI) barcodes information (Fig. S9). Subsequently, all the samples were stored in 100% ethanol at -80°C before genomic DNA extraction.

The genomic DNA was isolated using FastPure Cell/Tissue DNA Isolation Mini Kit (Vazyme Biotech Co., Ltd, Nanjing, China) according to the manufacturer's instructions. The VAHTS Universal Plus DNA Library Prep Kit for Illumina (Vazyme Biotech Co., Ltd, Nanjing, China) was used for individual genomic DNA libraries using $\sim 1\ \mu\text{g}$ genomic DNA as input from each sample. The constructed libraries were sequenced on the Illumina NovaSeq sequencing platform (GrandOmics, Wuhan, China) and paired-end 150 bp reads were generated.

Variation calling and annotation

Illumina paired-end sequenced raw reads for mapping were filtered using Fastp v0.20.1 [50] preprocessor to remove low-quality reads, adaptors, and reads containing poly-N using default parameters, and then clean reads were mapped to *E. griseocens* (ASM1756216v1, GCA_017562165.1) reference genome [13] using BWA v5.0.1 MEM module [51] with default parameters. SAMtools v1.12 software was used to convert mapping results into the BAM format and sort them. Duplicated reads were filtered with the Picard package v2.26.6 (<https://broadinstitute.github.io/picard/>). After BWA alignment, the reads around InDels were realigned, and realignment was performed with Genome Analysis Toolkit (GATK) v4.1.8.1 [52] in two steps. The first step used the RealignerTargetCreator package to identify regions where realignment was needed, and the second step used Indel-Realigner to realign the regions found in the first step, which produced a realigned BAM file for each accession.

The variation detection followed the best practice workflow recommended by GATK. In brief, the variants were called for each accession by the GATK Haplotype-Caller. Sentieon v20211201 (<https://www.sentieon.com>) was used to perform joint genotyping on the gVCF files v0.1.17 [53] software, resulting in a comprehensive union of genetic variations. In the filtering step, (a) retain the variants with a minimum quality score of 30 ($-\text{minQ } 30$), (b) only keep sites with two alleles ($-\text{min-alleles } 2$ $-\text{max-alleles } 2$), (c) retain sites with at least four alleles carried by each site ($-\text{mac } 4$), (d) remove sites with a missing rate exceeding 15% the maximum allowed missing rate is 95% ($-\text{max-missing } 0.95$).

Sites with a minimum read depth of 3 were kept, and an error correction script was applied to further improve the data quality. The sex chromosome was not considered. InDels and SNPs with no bi-allelic, $>40\%$ missing calls and $\text{MAF} < 0.005$ were removed, which yielded the basic set. SNPs with $\text{MAF} < 0.05$ were further removed for phylogenetic tree structure, IBS calculation, LD decay, PCA and population structure analyses.

SNPs and InDels annotation were performed according to the tea grey geometrid genome using the package ANNOVAR v2015-12-14 [54]. The coverage of each accession against each chromosome of the grapevine genome was counted based on an aligned BAM file using SAMtools v1.12 software. SNP density, InDel density, and total genetic diversity across each chromosome were counted with a 100 kb sliding window using VCFtools v0.1.17 [53] software.

Phylogeny construction and population structure analysis

For all individuals, we first used PLINK v1.90b6.26 [55] with the parameters “ $-\text{indep } -\text{pairphase } 100\ 10\ 0.2$ ” to

determine a pruned SNP set to be used in the population structure analysis. In this way, we used 1,432,873 SNPs for the phylogeny and population analysis. The IBS genetic distance matrix was calculated to quantify the correlation between individuals using the “-distance 1-ibs flat-missing” parameter in PLINK. For the principal component analysis (PCA), we used PLINK with default parameters to extract the top 10 principal components (PCs). The top three PCs were plotted using package ggplot2 v3.4.2 [56] in R v4.2.0 (<https://www.R-project.org>). *E. obliqua* was used as the outgroup to root the phylogenetic tree. Maximum likelihood estimation of individual ancestries was performed using ADMIXTURE v1.3.0 [57] with the parameters “-cv -j4” for multiple repeats with different random seeds, and plotted in R. We used the web tool iTOLs (<https://itol.embl.de/>) to color the phylogenetic tree. We defined three phylogenetic groups according to the phylogenetic tree.

Inference of population demographic history

For the demographic analysis, we used SMC++ [58] v1.15.4. For SMC++, we selected more representative individuals from the three phylogenetic groups. *E. obliqua* was selected as the comparison species, and we used SNPable (<http://lh3lh3.users.sourceforge.net/snpable.shtml>), MSMC mappability, and BEDTools [59] v2.27.1 to prepare the input file for SMC++. In this analysis, the timepoints are specified as 1,000~1,000,000, the number of knots for the spline is 15, and the mutation rate is 2e-8.

Genome scanning for selective sweep signals

We calculated the genetic diversity (π) using PLINK with the parameters “-window-pi 100,000”. We used PLINK to calculate the inter-population genetic differentiation index (F_{ST}) of different phylogenetic groups in 100 kb windows [37]. We used PLINK to calculate Tajima's D in 100-kb non-overlapping windows with parameter “-TajimaD 100000”.

We also performed a genetic differentiation and polymorphism levels (θ_{π} , pairwise nucleotide variation as a measure of variability) based cross approach to investigate the selection signals across the whole genome. A 100 kb sliding window with a 10 kb step approach was applied to quantify F_{ST} and θ_{π} by using VCFtools v0.1.17 [53] software. The candidates that meet both top 5% of the two values were selected as selective signals.

Migration events by TreeMix analysis

To infer migration events among the eight species, we used TreeMix v1.13 [60] to construct an ML tree with bighorn as the root using the “-noss” option to turn off the sample size correction, a window size (-K) of 500 SNPs (approximately 609 kb in this study) to account

for the impact of LD, which is more than the average LD length of ~150 kb observed in sheep7. Blocks with 500 SNPs were resampled, and 100 bootstrap replications were performed. We constructed the ML trees with 0–11 migration events and corresponding residuals. The proportions of explained variance for the migration numbers were calculated using in-house scripts.

Positive selection

To detect positive selection sites in *pgp-2* and *LCT*, a total of 9 *pgp-2* and 26 *LCT* were aligned using Multiple Alignment using Fast Fourier Transform (MAFFT) v7.480 software [61], respectively. This alignment helped to identify similarities and differences among the genes. Subsequently, dN/dS values were calculated using the site model in the CodeML tools of PAML v4.9j [62], employing the Bayes Empirical Bayes (BEB) method.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1007/s44297-024-00026-z>.

Additional file 1: Fig. S1. The optimal K values from cross-validation (CV) error test for the analysis of Admixture. Fig. S2. The optimum value of the migration edge (m) for TreeMix analysis. Fig. S3. The residual plot of TreeMix analysis. Fig. S4. The π value of SNPs in chromosome 15. Fig. S5. The Tajima's D value of SNPs in chromosome 15. Fig. S6. GO enrichment of 222 candidate genes. Fig. S7. The phylogenetic tree of *pgp-2* genes. Fig. S8. The phylogenetic tree of *LCT*. Fig. S9. The phylogenetic tree of *COI* genes.

Additional file 2: Table S1. The site information of geometrid samples. Table S2. The statistic of CV error value. Table S3. The IBS matrix data. Table S4. The 20 vector quantities in PCA analysis. Table S5. The π value in 30 chromosomes. Table S6. The F_{ST} -value of 30 chromosomes. Table S7. The Tajima's D value of 30 chromosomes. Table S8. The function annotation and list of 222 candidate genes. Table S9. The Bayes Empirical Bayes (BEB) probabilities of all aligned sites in *pgp-2*. Table S10. The Bayes Empirical Bayes (BEB) probabilities of all aligned sites in *LCT* genes.

Acknowledgements

We thank Enhui Shen for providing technical assistance and Xiuqin Ge for collecting and rearing the tea gray geometrid. We also thank the other members of the Chen laboratory for critical comments to improve the manuscript.

Authors' contributions

Conceptualization and supervision, Zhizhi Wang; Software, Ruizhong Yuan, Xiaohan Shu & Zhaohe Lu; Investigation, Ruizhong Yuan, Yusi Chen, Xiaogui Zhou & Zhizhi Wang; Writing – Original Draft Preparation, Ruizhong Yuan; Writing – Review & Editing, Ruizhong Yuan, Pu Tang & Zhizhi Wang; Visualization, Ruizhong Yuan & Xiqian Ye; Funding Acquisition, Zhizhi Wang.

Funding

This research was funded by the National Key Research and Development Plan (Grant No.2019YFD1002101), the Guangdong Laboratory of Lingnan Modern Agriculture Project (Grant No. NT2021003), the Key Research and Development Program of Zhejiang Province (Grant No. 2021C02045) and the Natural Science Foundation of China (Grant Nos. U22A20485 and 32272607).

Availability of data and materials

Raw sequencing data that support the findings of this study can be found in the NCBI SRA database under the BioProject (PRJNA983079) with the accession numbers SRR25047583, SRR25047584, SRR25047585, SRR25047586,

SRR25047587, SRR25047588, SRR25047589, SRR25047590, SRR25047591, SRR25047592, SRR25047593, SRR25047594, SRR25047595, SRR25047596, SRR25047597, SRR25047598, SRR25047599, SRR25047600, SRR25047601, SRR25047602, SRR25047603, SRR25047604, SRR25047605, SRR25047606, SRR25047607, SRR25047608, SRR25047609, SRR25047610, SRR25047611, SRR25047612, SRR25047613, SRR25047614, SRR25047615, SRR25047616, SRR25047617, SRR25047618, SRR25047619, SRR25047620, SRR25047621, SRR25047622, SRR25047623, SRR25047624, SRR25047625, SRR25047626, SRR25047627, SRR25047628, SRR25047629, SRR25047630, SRR25047631, SRR25047632, SRR25047633, SRR25047634, SRR25047635, SRR25047636. Additional data such as raw image files and in-house scripts that support this study are available from the first authors upon request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Institute of Insect Sciences, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310058, China. ²State Key Lab of Rice Biology, Ministry of Agriculture Key Lab of Molecular Biology of Crop Pathogens and Insects, and Zhejiang Provincial Key Laboratory of Biology of Crop Pathogens and Insects, Zhejiang University, Hangzhou 310058, China. ³Hainan Institute, Zhejiang University, Sanya 572025, China. ⁴Key Laboratory of Tea Biology and Resources Utilization, Ministry of Agriculture, National Center for Tea Improvement, Tea Research Institute, Chinese Academy of Agricultural Sciences, Hangzhou 310008, China. ⁵Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou 510642, China.

Received: 14 December 2023 Revised: 11 March 2024 Accepted: 11 April 2024

Published online: 01 June 2024

References

- Ye GY, Xiao Q, Chen M, Chen XX, Yuan ZJ, Stanley DW, et al. Tea: Biological control of insect and mite pests in China. *Biol Control*. 2014;68:73–91. <https://doi.org/10.1016/j.biocontrol.2013.06.013>.
- Drew L. Making tea. *Nature*. 2019;566(7742):S2–4. <https://doi.org/10.1038/d41586-019-00395-4>.
- Wei CL, Yang H, Wang SB, Zhao J, Liu C, Gao LP, et al. Draft genome sequence of *Camellia sinensis* var. *sinensis* provides insights into the evolution of the tea genome and tea quality. *Proc Natl Acad Sci USA*. 2018;115(18):E4151–8. <https://doi.org/10.1073/pnas.1719622115>.
- Xia EH, Zhang HB, Sheng J, Li K, Zhang QJ, Kim C, et al. The tea tree genome provides insights into tea flavor and independent evolution of caffeine biosynthesis. *Mol Plant*. 2017;10(6):866–77. <https://doi.org/10.1016/j.molp.2017.04.002>.
- Zhang GH, Yuan ZJ, Zhang CX, Yin KS, Tang MJ, Guo HW, et al. Detecting deep divergence in seventeen populations of tea geometrid (*Ectopis obliqua* Prout) in China by COI mtDNA and cross-breeding. *PLoS One*. 2014;9(6). <https://doi.org/10.1371/journal.pone.0099373>.
- Hazra A, Dasgupta N, Sengupta C, Bera B, Das S. Tea: a worthwhile, popular beverage crop since time immemorial. *Agron Crops*. 2019;507–31. https://doi.org/10.1007/978-981-32-9151-5_22.
- Yu X, Xiao J, Chen S, Yu Y, Ma J, Lin Y, et al. Metabolite signatures of diverse *Camellia sinensis* tea populations. *Nat Commun*. 2020;11(1). <https://doi.org/10.1038/s41467-020-19441-1>.
- Xia E, Tong W, Hou Y, An Y, Chen L, Wu Q, et al. The reference genome of tea plant and resequencing of 81 diverse accessions provide insights into its genome evolution and adaptation. *Mol Plant*. 2020;13(7):1013–26. <https://doi.org/10.1016/j.molp.2020.04.010>.
- Wang X, Feng H, Chang Y, Ma C, Wang L, Hao X, et al. Population sequencing enhances understanding of tea plant evolution. *Nat Commun*. 2020;11(1). <https://doi.org/10.1038/s41467-020-18228-8>.
- Samanta S. Potential bioactive components and health promotional benefits of tea (*Camellia sinensis*). *J Am Nutr Assoc*. 2022;41(1):65–93. <https://doi.org/10.1080/07315724.2020.1827082>.
- Brody H. Tea. *Nature*. 2019;566(7742):S1. <https://doi.org/10.1038/d41586-019-00394-5>.
- Wang ZB, Bai JH, Liu YJ, Li H, Zhan S, Xiao Q. Transcriptomic analysis reveals insect hormone biosynthesis pathway involved in desynchronized development phenomenon in hybridized sibling species of tea geometrids (*Ectopis griseascens* and *Ectopis obliqua*). *Insects*. 2019;10(11). <https://doi.org/10.3390/insects10110381>.
- Pan Y, Fang G, Wang Z, Cao Y, Liu Y, Li G, et al. Chromosome-level genome reference and genome editing of the tea geometrid. *Mol Ecol Resour*. 2021;21(6):2034–49. <https://doi.org/10.1111/1755-0998.13385>.
- Li ZQ, Cai XM, Luo ZX, Bian L, Xin ZJ, Liu Y, et al. Geographical distribution of *Ectopis griseascens* (Lepidoptera: Geometridae) and *Ectopis obliqua* in China and description of an efficient identification method. *J Econ Entomol*. 2019;112(1):277–83. <https://doi.org/10.1093/jeet/toy358>.
- Hazarika LK, Bhuyan M, Hazarika BN. Insect pests of tea and their management. *Annu Rev Entomol*. 2009;54:267–84. <https://doi.org/10.1146/annurev.ento.53.103106.093359>.
- Sinu PA, Mandal P, Banerjee D, Mallick S, Talukdar T, Pathak SK. Moth pests collected in light traps of tea plantations in North East India: species composition, seasonality and effect of habitat type. *Curr Sci*. 2013;104(5):646–51.
- Wang ZB, Ma T, Mao TF, Guo HW, Zhou XG, Wen XJ, et al. Application technology of the sex pheromone of the tea geometrid *Ectopis griseascens* (Lepidoptera: Geometridae). *Int J Pest Manag*. 2018;64(4):372–8. <https://doi.org/10.1080/09670874.2018.1447170>.
- Wang HF, Ma T, Xiao Q, Cao PR, Chen X, Wen YZ, et al. Pupation behaviors and emergence successes of *Ectopis griseascens* (Lepidoptera: Geometridae) in response to different substrate types and moisture contents. *Environ Entomol*. 2017;46(6):1365–73. <https://doi.org/10.1093/ee/nvx168>.
- Sun L, Mao TF, Zhang YX, Wu JJ, Bai JH, Zhang YN, et al. Characterization of candidate odorant-binding proteins and chemosensory proteins in the tea geometrid *Ectopis obliqua* Prout (Lepidoptera: Geometridae). *Arch Insect Biochem Physiol*. 2017;94(4). <https://doi.org/10.1002/arch.21383>.
- Luo ZX, Li ZQ, Cai XM, Bian L, Chen ZM. Evidence of premating isolation between two sibling moths: *Ectopis griseascens* and *Ectopis obliqua* (Lepidoptera: Geometridae). *J Econ Entomol*. 2017;110(6):2364–70. <https://doi.org/10.1093/jeet/tox216>.
- Zhang GH, Yuan ZJ, Yin KS, Fu JY, Tang MJ, Xiao Q. Asymmetrical reproductive interference between two sibling species of tea looper: *Ectopis griseascens* and *Ectopis obliqua*. *Bull Entomol Res*. 2016;1–8. <https://doi.org/10.1017/S0007485316000602>.
- Cheng ZQ, Li JQ, Ye XH, Li F, He K. Comparative analysis of the draft genome of two sibling tea geometrids, *Ectopis griseascens* and *Ectopis obliqua*. *J Plant Prot*. 2022;59(4):726–40.
- Zhou QY, Zhao SQ, Zhu JY, Li FD, Tong W, Liu SR, et al. Transcriptomic analyses reveal a systemic defense role of the uninfested adjacent leaf in tea plant (*Camellia sinensis*) attacked by tea geometrids (*Ectopis obliqua*). *Genomics*. 2020;112(5):3658–67. <https://doi.org/10.1016/j.ygeno.2020.03.007>.
- Jiang N, Liu S-X, Xue D-Y, Tang M-J, Xiao Q, Han H-X. External morphology and molecular identification of two tea geometrid moth from southern China. *Chinese J Appl Entomol*. 2014;51(4):987–1002. <https://doi.org/10.7679/j.issn.2095-1353.2014.118>.
- Wang ZB, Li H, Zhou XG, Tang MJ, Sun L, Zhan S, et al. Comparative characterization of microbiota between the sibling species of tea geometrid moth *Ectopis obliqua* Prout and *E. griseascens* Warren. *Bull Entomol Res*. 2020;110(6):684–93. <https://doi.org/10.1017/s0007485320000164>.
- Zhang XX, Mei Y, Li H, Tang MJ, He K, Xiao Q. Larval-transcriptome dynamics of *Ectopis griseascens* reveals differences in virulence mechanism between two EcobNPV strains. *Insects*. 2022;13(12). <https://doi.org/10.3390/insects13121088>.
- Li H, Tang M, Guo H, Wang Z, Xiao Q. Toxicity difference of EoNPV of two sibling species of tea geometrids. *Acta Agric Zhejiangensis*. 2020;32(8):1415–9.

28. Bai J, Tang M, Yin K, Wang Z, Xiao Q. Differential biological characteristics between closely related tea geometrid species, *Ectropis obliqua* and *Ectropis griseascens*. *Acta Agric Zhejiangensis*. 2018;30(5):797–803.
29. Ge F, Shi F, Cheng C, Chen L. Development and reproduction of *Ectropis griseascens* and *Ectropis obliqua* based on the two-sex life table. *J Tea Sci*. 2022;42(6):828–38.
30. Wang ZB, Liu YJ, Bai JH, Zhang XX, Xiao Q. Wolbachia regulates the egg hatching of hybrids of *Ectropis obliqua* and *E. griseascens* (Lepidoptera: Geometridae). *Acta Entomol Sinica*. 2022;65(8):999–1009. <https://doi.org/10.16380/j.kcxb.2022.08.008>.
31. Jing TT, Qian XN, Du WK, Gao T, Li DF, Guo DY, et al. Herbivore-induced volatiles influence moth preference by increasing the beta-Ocimene emission of neighbouring tea plants. *Plant Cell Environ*. 2021;44(11):3667–80. <https://doi.org/10.1111/pce.14174>.
32. Li XW, Zhang J, Lin SB, Xing YX, Zhang X, Ye M, et al. (+)-Catechin, epicatechin and epigallocatechin are important inducible defensive compounds against *Ectropis griseascens* in tea plants. *Plant Cell Environ*. 2022;45(2):496–511. <https://doi.org/10.1111/pce.14216>.
33. Jing TT, Du WK, Qian XA, Wang K, Luo LX, Zhang XY, et al. UGT89AC1-mediated quercetin glucosylation is induced upon herbivore damage and enhances *Camellia sinensis* resistance to insect feeding. *Plant Cell Environ*. 2023. <https://doi.org/10.1111/pce.14751>.
34. Jing SL, Zheng BX, Qiao L, Zhang K, Chen GY, Guo SQ, et al. Characterization of eleven microsatellite markers for the tea geometrid *Ectropis obliqua* (Lepidoptera: Geometridae). *Appl Entomol Zool*. 2016;51(4):673–6. <https://doi.org/10.1007/s13355-016-0442-3>.
35. Chen LL, Yuan P, You MS, Pozsgai G, Ma X, Zhu HP, et al. Cover crops enhance natural enemies while help suppressing pests in a tea plantation. *Ann Entomol Soc Am*. 2019;112(4):348–55. <https://doi.org/10.1093/aesa/say050>.
36. Zhang W, Rong J, Wei C, Gao L, Chen J. Domestication origin and spread of cultivated tea plants. *Biodivers Sci*. 2018;26(4):357–72. <https://doi.org/10.17520/biods.2018006>.
37. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population-structure. *Evolution*. 1984;38(6):1358–70. <https://doi.org/10.2307/2408641>.
38. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol*. 2021;38(12):5825–9. <https://doi.org/10.1093/molbev/msab293>.
39. Schlegel A, Stainier DYR. Lessons from “lower” organisms: what worms, flies, and zebrafish can teach us about human energy metabolism. *PLoS Genet*. 2007;3(11):2037–48. <https://doi.org/10.1371/journal.pgen.0030199>.
40. Dermauw W, Van Leeuwen T. The ABC gene family in arthropods: Comparative genomics and role in insecticide transport and resistance. *Insect Biochem Mol Biol*. 2014;45:89–110. <https://doi.org/10.1016/j.ibmb.2013.11.001>.
41. Ran W, Zhang J, Zhang X, Lin S, Sun X. Infestation of *Ectropis obliqua* affects the catechin metabolism in tea plants. *J Tea Sci*. 2018;38(2):133–9.
42. Wang P, Yu F. The geographic distribution, diversity and utilization of wild tea *Camellias* in China. *J Tea Sci*. 2002;22(2):105.
43. Chen YS, Zhou XG, Zeng WJ, Xiao Q, Wang ZZ, Chen XX. Resistance monitoring of two tea geometrid moths (*Ectropis obliqua* and *E. griseascens*) to five frequently used insecticides in different tea plantations. *J Environ Entomol*. 2022;45(4):1103–10.
44. Edger PP, Heidel-Fischer HM, Bekaert M, Rota J, Gloeckner G, Platts AE, et al. The butterfly plant arms-race escalated by gene and genome duplications. *Proc Natl Acad Sci USA*. 2015;112(27):8362–6. <https://doi.org/10.1073/pnas.1503926112>.
45. Höök L, Näsvalk K, Vila R, Wiklund C, Backström N. High-density linkage maps and chromosome level genome assemblies unveil direction and frequency of extensive structural rearrangements in wood white butterflies (*Leptidea* spp.). *Chromosome Res*. 2023;31(1). <https://doi.org/10.1007/s10577-023-09713-z>.
46. Mitter C, Davis DR, Cummings MP. Phylogeny and evolution of lepidoptera. *Annu Rev Entomol*. 2017;62:265–83.
47. Yu XM, Xiao JJ, Chen S, Yu Y, Ma JQ, Lin YZ, et al. Metabolite signatures of diverse *Camellia sinensis* tea populations. *Nat Commun*. 2020;11(1). <https://doi.org/10.1038/s41467-020-19441-1>.
48. Williamson G. Possible effects of dietary polyphenols on sugar absorption and digestion. *Mol Nutr Food Res*. 2013;57(1):48–57. <https://doi.org/10.1002/mnfr.201200511>.
49. Liang C, Apostolides Z, Zong-Mao C. Global tea breeding. achievements, challenges and perspectives. Tea Research Institute of the Chinese Academy of Agricultural Sciences, National Center for Tea Improvement, Hangzhou, China.: Tea Research Institute of the Chinese Academy of Agricultural Sciences, National Center for Tea Improvement, Hangzhou, China. 2012.
50. Chen SF, Zhou YQ, Chen YR, Gu J. Fastp: an ultra-fast all-in-one FASTQ pre-processor. *Bioinformatics*. 2018;34(17):884–90. <https://doi.org/10.1093/bioinformatics/bty560>.
51. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324>.
52. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43(5):491–8. <https://doi.org/10.1038/ng.806>.
53. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156–8. <https://doi.org/10.1093/bioinformatics/btr330>.
54. Wang K, Li MY, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16). <https://doi.org/10.1093/nar/gkq603>.
55. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Human Genet*. 2007;81(3):559–75. <https://doi.org/10.1086/519795>.
56. Ginestet C. Ggplot2: elegant graphics for data analysis. *J R Stat Soc Ser A Stat Soc*. 2011;174:245. https://doi.org/10.1111/j.1467-985X.2010.00676_9.x.
57. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19(9):1655–64. <https://doi.org/10.1101/gr.094052.109>.
58. Terhorst J, Kamm JA, Song YS. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat Genet*. 2017;49(2):303–9. <https://doi.org/10.1038/ng.3748>.
59. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2. <https://doi.org/10.1093/bioinformatics/btq033>.
60. Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet*. 2012;8(11). <https://doi.org/10.1371/journal.pgen.1002967>.
61. Katoh K, Standley DM. MAFFT: multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–80. <https://doi.org/10.1093/molbev/mst010>.
62. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007;24(8):1586–91. <https://doi.org/10.1093/molbev/msm088>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.