


RESEARCH ARTICLE

Development and characterization of microsatellite markers for population genetics of the cocoa pod borer *Conopomorpha cramerella* (Snellen) (Lepidoptera: Gracillariidae)

Marynold Purificacion¹ [✉], Roslina Binti Mohd Shah² [✉], Thierry De Meeûs^{3,4}, Saripah Binti Bakar⁵, Anisah Binti Savantil², Meriam Mohd Yusof², Divina Amalin¹, Hien Nguyen⁶, Endang Sulistyowati⁷, Aris Budiman⁷, Arni Ekayanti⁸, Jerome Niogret⁹, Sophie Ravel³, Marc J. B. Vreysen¹⁰, Adly M. M. Abd-Alla¹⁰ ^{*}

1 Biological Control Research Unit, Center for Natural Science and Environmental Research, De La Salle University, Manila, Philippines, **2** Centre for Cocoa Biotechnology Research, Malaysian Cocoa Board, Kota Kinabalu, Sabah, Malaysia, **3** Intertryp, Univ Montpellier, Cirad, IRD, Montpellier, France, **4** IRD, UMR Intertryp, Cirad, Campus International de Baillarguet, Montpellier, France, **5** Lembaga Koko Malaysia, Bagan Datuk, Perak, Malaysia, **6** Plant Protection Research Institute, Duc Thang, Hanoi, Vietnam, **7** Indonesian Coffee and Cocoa Research Institute, Jember, Indonesia, **8** Mars Cocoa Research Centre, Mars Wrigley, Sulawesi Selatan, Indonesia, **9** Mars Wrigley, Centre for Tropical Environmental & Sustainability Science, James Cook University Nguma-bada Campus, Smithfield, Australia, **10** Insect Pest Control Laboratory, Joint FAO/IAEA Centre of Nuclear Techniques in Food and Agriculture, Vienna, Austria

 These authors contributed equally to this work.

 Current address: Institute of Plant Breeding, College of Agriculture and Food Science, University of the Philippines Los Baños, Laguna, Philippines

* a.m.m.abd-alla@iaea.org



OPEN ACCESS

Citation: Purificacion M, Shah RBM, De Meeûs T, Bakar SB, Savantil AB, Yusof MM, et al. (2024) Development and characterization of microsatellite markers for population genetics of the cocoa pod borer *Conopomorpha cramerella* (Snellen) (Lepidoptera: Gracillariidae). PLoS ONE 19(4): e0297662. <https://doi.org/10.1371/journal.pone.0297662>

Editor: Jianhong Zhou, PLOS: Public Library of Science, UNITED STATES

Received: June 24, 2023

Accepted: January 9, 2024

Published: April 11, 2024

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0297662>

Copyright: © 2024 Purificacion et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its [Supporting](#)

Abstract

The cocoa pod borer (CPB) *Conopomorpha cramerella* (Snellen) (Lepidoptera: Gracillariidae) is one of the major constraints for cocoa production in South East Asia. In addition to cultural and chemical control methods, autocidal control tactics such as the Sterile Insect Technique (SIT) could be an efficient addition to the currently control strategy, however SIT implementation will depend on the population genetics of the targeted pest. The aim of the present work was to search for suitable microsatellite loci in the genome of CPB that is partially sequenced. Twelve microsatellites were initially selected and used to analyze moths collected from Indonesia, Malaysia, and the Philippines. A quality control verification process was carried out and seven microsatellites found to be suitable and efficient to distinguish differences between CPB populations from different locations. The selected microsatellites were also tested against a closely related species, i.e. the lychee fruit borer *Conopomorpha sinensis* (LFB) from Vietnam and eight loci were found to be suitable. The availability of these novel microsatellite loci will provide useful tools for the analysis of the population genetics and gene flow of these pests, to select suitable CPB strains to implement the SIT.

Information files. In addition the raw data are available in this link <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/CNRSU8>.

Funding: This work was funded by the International Atomic Energy Agency under the TC project RAS5095 and by the regular budget of the Joint FAO/IAEA Centre of Nuclear Techniques in Food and Agriculture. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

The cacao tree, *Theobroma cacao* L., belongs to the Sterculiaceae family and is one of the most widely cultivated species among the twenty-two species of the genus [1]. The International Cocoa Organization (ICCO) estimates the global market value of cocoa between 8 to 10 billion USD, based on an annual production of 4 million tons and a monthly average daily price of cocoa beans between USD 2,264 to 2,359 per ton. With the demand for finished and semi-finished cocoa products increasing yearly, cocoa production should ideally increase proportionally. However, the current worldwide production by the cocoa industry is not in line with this prediction. Farmers are producing less cocoa than world consumption, which inevitably leads to a chocolate deficit [2, 3]. The cocoa pod borer (CPB) *Conopomorpha cramerella* (Snellen) (Lepidoptera: Gracillariidae) is the major insect pest of cocoa production in South East Asia and it infests rambutan (*Nephelium lappaceum* L.), langsung (*Lansium parasiticum* (Osbeck) Sahni & Bennet) and cocoa [4]. CPB adults oviposit on the cocoa pods and the larvae enter the pods to feed on the pulp and placenta surrounding the beans that are used for chocolate production. On average, 40–60% of the cocoa beans produced in Southeast Asia are lost due to CPB and up to 80% of the crop is lost in farms in the absence of CPB management [5]. In addition, the lychee fruit borer (LFB) *Conopomorpha sinensis* Bradley (Lepidoptera: Gracillariidae), on the other hand is the most important lepidopteran borer affecting lychee and longan producing area in Southeast Asia. LFB larvae damages the fruit of lychee and longan causing up to 80% fruit decay rate [6].

Various control measures have been implemented to manage populations of these pests however they are either labor-intensive and costly (e.g., sleeving the cocoa pods with plastic bags), ineffective or viable (e.g., biological control methods such as the use of natural enemies and sex pheromones) [4] or has a negative impact on the environment (e.g., the use of chemical insecticides) [5, 7]. To reduce the amount of insecticides used, autocidal control tactics such as the sterile insect technique (SIT) could be integrated with existing control methods as part of an area-wide integrated pest management (AW-IPM) approach. However, determining the best approach for applying these strategies requires in-depth knowledge of the biology and ecology of the targeted pest populations, i.e., population size, reproductive strategies and dispersal. Information concerning the number of CPB populations present, and the degree of connectivity between them, is vital to constructing management and control policies for CPBs. For example, eradication may be possible in a series of small and discrete populations without a risk of reintroduction, whereas in geographically large populations, population reduction may be the best practicable approach [8, 9]. Such knowledge can be relatively easily assessed by the study of genetic variation of molecular markers between individuals, and within and between subsamples in space and time [10, 11].

Examples of molecular markers used for insect population studies are: mitochondrial DNA cytochrome oxidase I (mtDNA COI), coding nuclear elongation factor-1 α (EF-1 α) marker, randomly amplified polymorphic DNA (RAPD), restriction site associated DNA (RAD) and microsatellites [12, 13]. Microsatellite markers are short tandem DNA repeats of one to six nucleotides and are widely used for population genetic analysis, as they are abundant in most genomes, are highly polymorphic, Mendelian, co-dominant, and can be easily amplified by polymerase chain reaction (PCR) [14]. However, no microsatellites had been developed for any of the known *Conopomorpha* spp., and earlier attempts to analyze their population structure using COI and EF-1 α remained inconclusive [15, 16].

In the present paper, we described the development and characterization of microsatellite markers of *C. cramerella* that can be used in further population genetics studies. We also tested these markers on a closely related species, the lychee fruit borer (LFB) *Conopomorpha sinensis*, a serious pest of lychee orchards [17, 18].

Materials and methods

Insect collection and DNA extraction

Adult CPB moths were collected from Malaysia (Mal), Indonesia (Indo), and the Philippines (Phil) (Table 1). In Malaysia and Indonesia CPB were hand-caught under infected cacao branches using a cylindrical plastic bag. In the Philippines, insects were collected from infected cacao pods or adults were trapped with pheromone traps. In Vietnam, LFB individuals were collected from lychee fruits. The samples were stored in 95% ethanol (and replaced with propylene glycol before shipment) and shipped to the Insect Pest Control Laboratory (IPCL) of the Joint FAO/IAEA Centre of Nuclear Techniques in Food and Agriculture, Seibersdorf, Austria. Upon arrival, the propylene glycol was replaced with absolute ethanol and stored at -20°C.

Male and female individuals were distinguished by examining the ventrocaudal segments of the abdomen [19]. The total genomic DNA was extracted from one adult moth using Qiagen DNeasy® Blood & Tissue Kit (Qiagen Inc., Valencia, CA) according to the manufacturer's instructions. The quantity and quality of extracted DNA were assessed with a Synergy™ HI microplate reader (BioTek, Winooski, Vermont, USA).

Sequence analysis and microsatellite loci selection

The partial genome sequence of *C. cramerella* (Accession SJJU01000000, WGS SJJU01000001-SJJU01073142) available at the NCBI database was used as a reference for microsatellite primers selection. Msatcommander 1.08 [20] was used to search for di- and tri-nucleotide motifs. A total of 192 primer pairs were selected based on product size (180–300 nt) and the number of repeats (≥ 11 repeats), synthesized, and tested for microsatellite amplification with PCR using the CPB extracted DNA (S1 Table).

The 192 synthesized primer pairs were screened on the DNA extracted from three individuals from Malaysia. PCR amplifications were performed by mixing 12.5 μ L of Qiagen Taq PCR Master Mix Kit (Qiagen Inc., Valencia, CA), 10 μ L of nuclease-free H₂O (Qiagen Inc., Valencia, CA), 0.2 μ M final concentration of each forward and reverse primer and 1.5 μ L (4 ng) DNA. The PCR conditions were as follows: 94°C for 2 min; 34 cycles at 94°C for 15 s, 58°C for 15 s, and 68°C for 15 s; ending with a final extension for 5 min at 68°C. The PCR amplification was checked on 2% agarose E-gel™ stained with ethidium bromide (Invitrogen, ThermoFisher Scientific, Waltham, MA). Of 192 primer pairs, 53 successfully amplified the expected microsatellite fragments (S2 Table). The 53 primer pairs were then screened on 30 individuals of *C.*

Table 1. The locations, the geographical coordinates, and the number of adults of cacao pod borer (CPB) *Conopomorpha cramerella* and lychee fruit borer *Conopomorpha sinensis* (LFB) samples used for microsatellite development.

Species	Country of origin	Sampling location	Latitude	Longitude	Collected moths		
					Females	Males	Total
<i>C. cramerella</i>	Indonesia	Tarengge, Sulawesi	-2.51029	120.80766	12	12	24
	Malaysia	Bagan Datuk	3.97197	100.753445	12	12	24
	Philippines	Kabacan, Cotabato	7.14839	124.85229	12	12	24
<i>C. sinensis</i>	Vietnam	Luc Ngan, Bac Giang province	02363299	0665936	11	11	22

<https://doi.org/10.1371/journal.pone.0297662.t001>

cramerella from three locations (10 individuals per location), and 12 primer pairs were selected based on consistent amplification across all individuals and expected fragment size (Table 2).

Genotyping of microsatellite loci

A total of 72 individuals of CPB from three locations (Table 1) were used for microsatellite characterization. Microsatellite amplifications were performed by using a 12.5 μ L Qiagen *Taq* PCR Master Mix Kit (Qiagen Inc., Valencia, CA), mixed with 9.6 μ L nuclease-free H₂O (Qiagen Inc., Valencia, CA), 0.016 μ M forward primer with M13-adaptor (5'-CACGACGTTGTAA AACGAC-3'), 0.2 μ M reverse primer, 0.2 μ M M13 adaptor labeled with fluorescent dye (6-FAM for Cpb14, Cpb84 and Cpb136, HEX for Cpb55, Cpb122 and Cpb160, ATTO 550 for Cpb54, Cpb112 and Cpb139 or ATTO 565 for Cpb62, Cpb135 and Cpb190) and 1.5 μ L diluted DNA (~ 4 ng/ μ L). We used the Platinum[™] II Hot Start PCR 2X Master Mix (Thermo Fisher Scientific, Waltham, MA) instead of Qiagen *Taq* PCR Master Mix for Viet samples due to its higher sensitivity. Amplifications were done using 34 cycles and annealing temperatures given in Table 2. PCR amplicon products were resolved in 4% agarose E-gel[™] stained with ethidium bromide (Invitrogen, Thermo Fisher Scientific, Waltham, MA) and genotyped using ABI 3500XL Genetic Analyzer (Applied Biosystems[™], Foster City, CA) with a GeneScan[™] 600 LIZ[™] dye size standard (Thermo Fisher Scientific, Waltham, MA). Allele calling was performed

Table 2. Primer details from Msatcommander and optimized annealing temperature of 12 polymorphic microsatellites of *Conopomorpha cramerella*.

Microsatellite	Primer sequence (5' - 3')	Repeat motif and number	Product size, bp	Annealing temp, °C
Cpb14*	F: CTGTAGAGTGCGGAGTGTCG	(CCG) ₁₉	264	60
	R: TTTGCTCGCTGTTAGGTCGG			
Cpb54	F: GGTAAGAGTTGCCGAATGGC	(AC) ₁₄	264	55
	R: TCGCGGGAATAAGGGCAC			
Cpb55*	F: TGAATAAGCACCTCTCACG	(CCG) ₁₄	273	55
	R: ATAGCCCAGAACCACCTTC			
Cpb62	F: CAGAACACAGATCGTTGCC	(AC) ₁₃	213	61
	R: CATGGCGATGAAAGTGATTGC			
Cpb84	F: CACACAGCTAAGCGAACACC	(AC) ₁₂	189	61
	R: CTTCACCGCTCATCACCTGC			
Cpb112	F: TCGGCCGTCTCGAGATATTC	(AC) ₁₂	288	60
	R: CTCAGAAATGGTGACCCGTG			
Cpb122	F: AAGCAAATGTACCCGACCC	(AC) ₁₁	189	55
	R: GCCGGGCACTTTACTTGATG			
Cpb135*	F: TGTAATCGGCCCACTTCCTC	(CCG) ₁₁	214	56
	R: TCGGAGATGGATCGTGTCTG			
Cpb136 [#]	F: TGTAATCGGCCCACTTCCTC	(AC) ₁₁	218	55
	R: TCGGACGTTGTTACTTTCG			
Cpb139	F: GTCATTTACCGACGACTATGG	(CCG) ₁₁	228	55
	R: AACCCACCGATTCCAGAGAG			
Cpb160	F: GTTGACGTGACCCATATGCG	(AC) ₁₁	257	55
	R: TCGGATAGCGTTTCGAGTGG			
Cpb190	F: CTGTTGTGAGCCGTTCTCTG	(AC) ₁₁	284	60
	R: CTCACACATCCTGGCGAATG			

*: Loci excluded from quality testing due to amplification failures.

[#]: Msatcommander prediction indicated these loci as tri-nucleotides however genmapper alleles indicated di-nucleotides.

<https://doi.org/10.1371/journal.pone.0297662.t002>

using GeneMapper 4.1 software (Applied Biosystems™, Foster City, CA) to obtain a co-dominant allele matrix from the raw data. The genetic data were formatted for Create 1.37 [21] to convert the datasets into the required formats according to needs.

In Lepidoptera, sex determinism consists of female heterogamety, with a majority of species displaying the WZ/ZZ (female/male) mechanism [22]. Following this, any polymorphic locus located in one of the two sex chromosomes should display strong discrepancies of heterozygosity between females and males. In absence of recombination for the W chromosome, we also expect an accumulation of divergent mutations at any locus located on a sex chromosome between males and females. To check for chromosomal location, we have measured the heterozygote deficit with Weir and Cockerham's (Weir and Cockerham 1984) unbiased estimator of Wright's [23] F_{IS} in males and females. To limit the number of tests, we compared males and females for the three loci (Cpb14, Cpb112 and Cpb122) displaying the biggest F_{IS} differences with a Wilcoxon signed rank test for paired data, the pairing unit being the allele. These tests were undertaken with R-commander (Fox, 2005, 2007) for R (R-core Team). For genetic divergence, we used Weir and Cockerham's unbiased estimator of Wright's F_{ST} measured between females and males within each country and for each locus, and tested the significance of this divergence with the G -based randomization test (Goudet et al 1996), after 10,000 randomizations of individuals between subsamples. These computations were undertaken with Fstat 2.9.4 [24]. We averaged F_{ST} s across countries and combined the corresponding p -values with the generalized binomial procedure with MultiTest [25]. False discovery rate across the 11 loci was then taken into account with the Benjamini and Hochberg procedure [26] with the command "p.adjust" of R.

Detection of outlier individuals in the different subsamples

To check if we have outliers in CPB and LFB samples, we first used the Bayesian clustering algorithm of STRUCTURE 2.3.4 [27]. For this analysis data were converted with Convert [28]. We used a burn-in period of 5,000 and 50,000 Markov chain Monte Carlo (MCMC) iterations, asked for a number of clusters K between 1 and 10, with 10 iterations. The number of genetic clusters was determined following the Delta- K method [29] using Structure Harvester [30]. Lastly, to visualize the genetic relationships between individuals, a genetic distance matrix was computed using Cavalli-Sforza and Edwards' chord distance (D_{CSE}) [31], corrected for null alleles with the INA procedure with FreeNA [32]. This matrix was used to build a neighbor-joining tree (NJTree) [33] with Mega 7 software [34] as recommended by Takezaki and Nei [35]. This version of MEGA indeed allow to import the resulting tree into an object that is editable with a drawing/presentation program (Impress, PowerPoint, etc.), while more recent versions only create an image that cannot be modified.

Quality assessment of microsatellite loci

The following analyses were undertaken on the two species separately. The G -based tests for linkage disequilibrium (LD) between each pair of loci [36] were conducted with 10000 randomizations, over all subsamples. This produced as many non-independent tests as there are locus pairs. The series of p -values was thus adjusted using the Benjamin and Yekutieli (BY) false discovery rate (FDR) procedure [37] in RStudio 2021.09.2 [38]. Proportions of significant tests in the different sites were compared with a Fisher exact test under rcmdr.

The deviation from panmictic expectations of genotypic frequencies within subsamples (Wright's F_{IS}), and subdivision (Wright's F_{ST}) were estimated using Weir and Cockerham's unbiased estimators [39]. The significance of deviations from panmixia and of subdivision were tested with 10000 permutations of alleles between individuals within subsamples and of

individuals between subsamples, respectively. For F_{IS} , tests are one-sided (heterozygote deficits) by default. We computed two-sided p -values by doubling the p -value in case of positive F_{IS} , or by doubling $(1-p)$ -value otherwise. For subdivision, tests were kept one-sided. The confidence intervals were computed after 5,000 bootstraps over loci. Wright's F_{IT} was also estimated for short alleles dominance (SAD) analysis (see below).

In case of significant heterozygote deficits, we used the approaches described in several papers [40–43]. For null alleles, the ratio of standard errors of F_{IS} over F_{ST} ($r_{FIS/FST}$) was computed, a value above 2 being a signature of probable null alleles. We also measured the correlation between F_{IS} and F_{ST} , and between F_{IS} and missing data (N_{Blanks} : putative null homozygotes), which are expected to be positive in case of null alleles. These correlations were measured and tested with one-sided Spearman's rank correlation test with `rcmdr`. We also computed the slope, intercept and determination coefficient of the regression $F_{IS} \sim N_{Blanks}$. Expected frequencies of null alleles followed the expectation-maximization step (EM) algorithm with `FreeNA` and were compared to observed ones. Stuttering signatures were checked with the recently developed methods of De Meeûs and Noûs [43] and corrected by pooling alleles with one to two base differences, following rules defined in [41]. The presence of SAD was checked using a one-sided (negative correlation) Spearman's rank correlation test between F_{IT} and allele size [42]. In case of doubt, we also undertook the regression approach $F_{IS} \sim \text{Allele size}$ weighted by $p_i(1-p_i)$, where p_i is the frequency of allele of size i [44]. The goodness of fit of expected null homozygotes and observed missing data was tested with a one-sided exact binomial test with `RStudio 2021.09.2` [38] (command "`binom.test`" with the alternative = "less"). In all cases, independent test series were adjusted with the Benjamini and Hochberg (BH) procedure [26] with `RStudio` (S1 File).

Population genetics structure of *Conopomorpha* samples

These analyses (subdivision, population size, number of immigrants, immigration rates and dispersal distance) were undertaken for each species separately, and after cleaning datasets from outlier loci or individuals. Because of the presence of null alleles, we used `FreeNA` to measure subdivision with the excluding null alleles (ENA) correction for null alleles, after recoding missing data as homozygotes for the null allele (999999) when relevant [32]. We computed 95% confidence intervals (CI) with 5000 bootstraps over loci. We corrected this estimate for excess of polymorphism using Meirmans' method with `RecodeData` [45] to compute the maximum possible divergence F_{ST_max} with `Fstat`. The standardized estimate was then computed as F_{ST_FreeNA}/F_{ST_max} . For the record, we did not use Meirmans and Hedrick's G_{ST} [46], because it cannot correct for the presence of null alleles. Significance was assessed with the G -based permutation test with `Fstat` and false discovery rate correction was undertaken with Benjamini and Yekutieli (Benjamini and Yekutieli) (BY) correction with the command "`p.adjust`" of R.

Effective population sizes (N_e) were estimated with five different methods. The first method (N_{eFIS}) used a recent F_{IS} based method where $N_e = -1/(2F_{IS}) - F_{IS}/[2(1+F_{IS})]$ [43]. Because some loci with null alleles and heterozygote deficits could not be used, we averaged the values obtained across loci displaying an heterozygote excess. The second method (LDN_e) was the LD based method [47], adjusted for missing data [48], where only alleles with frequency above 0.05 were used and random mating assumed. The third method ($CoAN_e$) corresponded to the co-ancestry method [49]. These two methods were computed with `NeEstimator 2.1` [50]. The fourth method (IN_e) was computed with `ESTIM 1.2` updated from [51] which uses the one and two locus identity probabilities based effective population size [52, 53]. The last one is the sibship frequency based effective population size (SN_e) [54], assuming polygamy and inbreeding, computed with `Colony 2.0.6.8` [55]. All methods but IN_e assume isolated subsamples,

while INe assumes an Island model of migration. Non-isolation should tend to generate under-estimates for small subpopulations with rather small immigration rates, at least for $N_e F_{IS}$, and LDN_e , while over-estimates could be expected with $CoAN_e$ and SN_e . For IN_e , alternative population structure as isolation by distance should produce over-estimates as well. Nevertheless, these biases are not expected to change the order of magnitude of estimates. For each method, we computed the average and minimum and maximum values. We then computed the unweighted grand average across methods of these three quantities. This hopefully ensured a range of values that contained real values.

The effective number of immigrants exchanged between subsamples was then calculated: globally, assuming an infinite island model, with $N_e m = (1 - F_{ST_FreeNA}) / (4F_{ST_FreeNA})$ (e.g. De Meeûs 2021 [56], page 52, equation 24, assuming $m \gg u$). Geographic distances between all pairs of subsamples (D_{geo}) were computed from the GPS coordinates in decimal degrees using the geosphere package [57] of R. Finally, we estimated the immigration rates ($m = N_e m / N_e$) and dispersal distances ($\delta = m \times D_{geo}$) where D_{geo} is the geographic distance, either between two subsamples or the average between all subsamples.

Inclusivity in global research. Additional information regarding the ethical, cultural, and scientific considerations specific to inclusivity in global research is included in the [S2 File](#).

Results

Markers development and validation

The Msatcommander search for the di- and tri nucleotides motifs in the genome of the CPB showed a total of 37,532 motifs and 35,076 pairs of primers. Out of those, 6,346 primer pairs were found in duplicates in the genome and therefore excluded from the analysis. Combining the unique primer pairs with the motifs using vlookup in Excel produced 28,730 primer pairs of which 9,968 primer pairs amplified di-nucleotides repeats and 18,762 amplified trinucleotide repeats. Of these, 7,155 primer pairs produced PCR products ranging between 180 and 300 nucleotides of which 203 primers pairs amplified motifs with ≥ 11 repeats. A total of 192 primer pairs were then synthesized and tested by PCR with a pooled DNA extract from three CPB adults from Malaysia ([S1 Table](#)).

Of the 192 primer pairs, 53 primers amplified a PCR product with the expected size ([S2 Table](#)). Testing these 53 primer pairs on the DNA of LFB collected from Vietnam indicated that only 29 primer pairs amplified the expected PCR product ([S2 Table](#)). Out of the 29 primer pairs that amplified the expected PCR product for CPB and LFB, 12 primer pairs showed heterogeneity between tested individuals and were therefore used for the initial population genetics study. One of these loci (Cpb136) displayed more than 50% amplification failures and was also disregarded.

Data obtained with the remaining 11 loci are presented in [S3 Table](#). With the 11 remaining loci, none of the F_{IS} comparisons between females and males for CPB and LFB showed a significant value (smallest p -value > 0.09094). No locus provided a significant genetic divergence between genders (minimum p -value > 0.1661) ([S4 and S5 Tables](#)). We thus assumed all these 11 loci to be autosomal.

Detection of outlier individuals in the different subsamples

Most individuals were assigned with very high probabilities (> 0.99) to either LFB in Vietnam or CPB in the other sampling locations ([Fig 1](#)). However, two individuals did not follow this rule, i.e., individual 77 from Vietnam clearly belonged to CPB, and individual 53 from Indonesia could represent a hybrid between the two species.

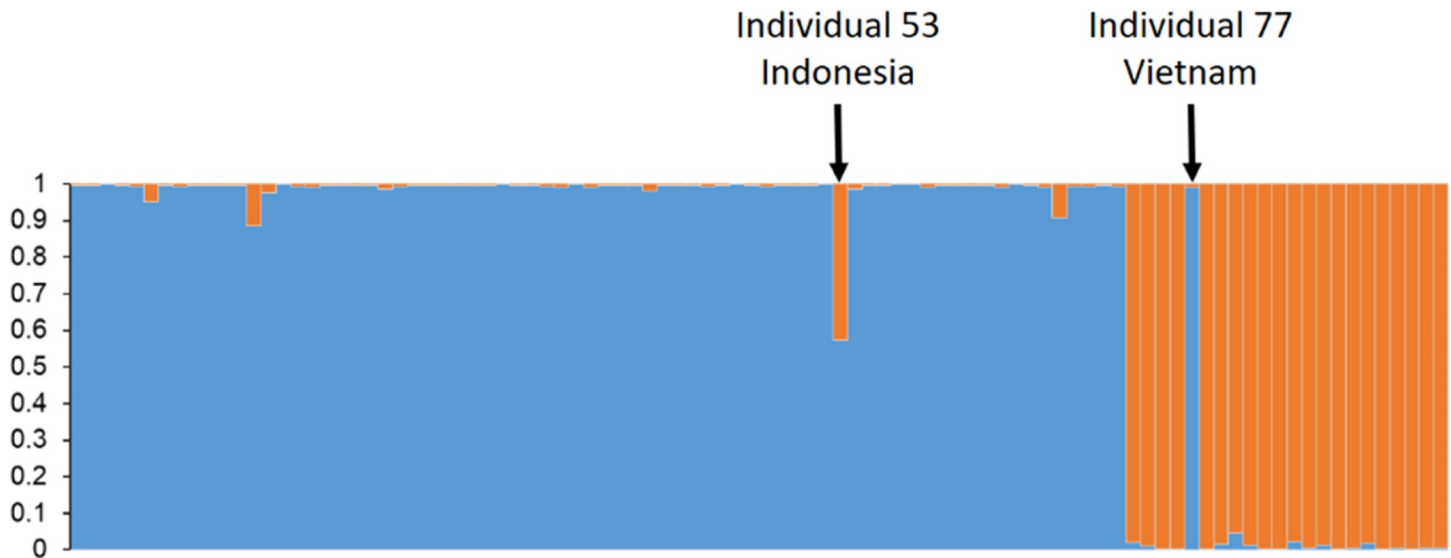


Fig 1. Assignment probabilities of each individual of *Conopomorpha cramerella* and *C. sinensis* to the two optimal clusters obtained with STRUCTURE and STRUCTURE Harvester. Results are presented for the optimal $K = 2$.

<https://doi.org/10.1371/journal.pone.0297662.g001>

The NJTree analysis did not confirm the hybrid status of individual 53, but confirmed that individual 77, sampled from lychee fruits from Vietnam belongs to *C. cramerella* (Fig 2), obviously represented a misplaced individual. Whatever the cause (immigrant from a cacao plantation, mislabelling or manipulation error), we removed individual 77 from further investigations. Only one locus appeared diagnostic between the two species, locus Cpb62, with alleles 191 and 211 that are only found in LFB, while alleles 185, 205, 217 to 233 (with one repeat increment), 237 and 257, were only found in CPB.

Quality assessment of microsatellite loci for *C. cramerella*

Proportions of significant LD tests did not significantly vary across subsamples (p -value = 0.1538). Over all subsamples, eight locus pairs displayed a significant LD (13%). Three locus pairs stayed significant after BY correction: Cpb55 x Cpb54, Cpb135 x Cpb160 and Cpb139 x Cpb160. The loci involved in these three significant pairs will thus need special attention.

We observed a highly significant and highly variable F_{IS} ($F_{IS} = 0.285$, p -value > 0.0002, 95% CI = [-0.021, 0.261]) (Fig 3). The ratio $r_{FIS/FST} = 8$ suggested that null alleles may have explained a part of this deficit. The correlation between F_{IS} and F_{ST} was positive but not significant ($\rho = 0.4727$, p -value = 0.0728), and the correlation between F_{IS} and N_{Blanks} was negative. However, there was quite a good agreement between expected and observed missing data, since only locus Cpb135 displayed a significant deviation (p -value = 0.02665), which did not stay significant after BH correction. Obviously, null alleles cannot explain all heterozygote deficits. Only a single SAD correlation test was significant, for locus Cpb160 (p -value = 0.04575), which was not confirmed by the regression test (one sided p -value = 0.2877). Finally, no loci displayed a significant stuttering (all p -values > 0.13).

Given that some loci display negative values, as expected in dioecious species [58–60], and other loci heterozygote deficits, locus specific causes, unlike a Wahlund effect, probably better explain our data. This was confirmed by the substantially variable and highly significant $F_{ST} = 0.043$ in 95% CI = [0.027, 0.06], with several loci displaying very small or very high values that

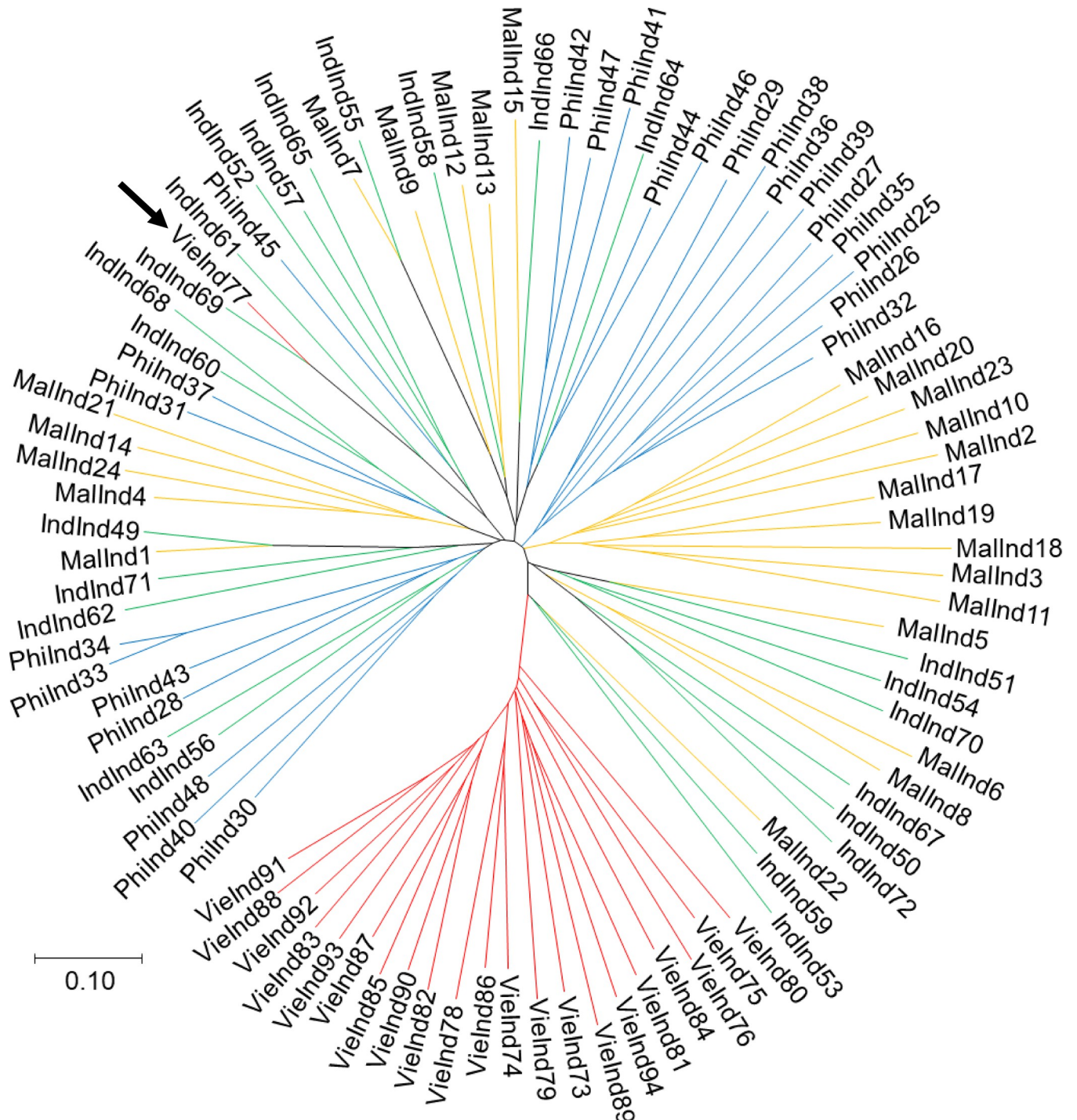


Fig 2. Neighbor-joining tree based on Cavalli-Sforza and Edward's chord distance between individuals of the two species of *Conopomorpha*, and corrected for null alleles. Individual names are composed of the three first letters of the site of origin (Ind = Indonesia, Mal = Malaysia, Phi = Philippines and Vie = Vietnam), followed by the individual number. Sites are also indicated by different colors of the branches (Indonesia: green; Malaysia: Yellow; Philippines: Blue; and Vietnam: red).

<https://doi.org/10.1371/journal.pone.0297662.g002>

ranged outside the 95% CI: loci Cpb14, Cpb55, Cpb62, Cpb84, Cpb135, Cpb139 and Cpb190 (Fig 4). Moreover, three locus pairs were found in significant LD and display very extreme F_{IS}

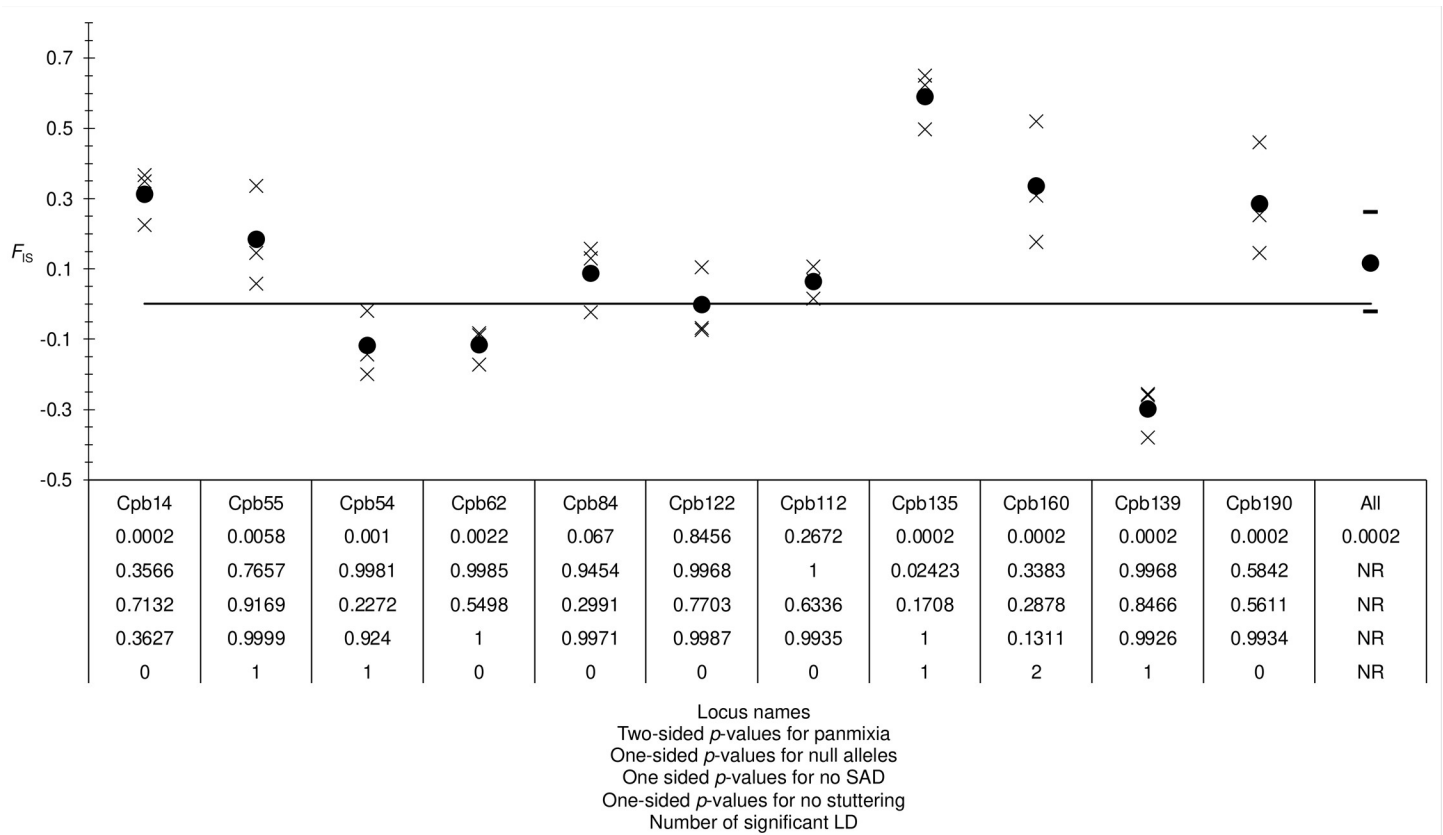


Fig 3. Variation of F_{IS} across the 11 loci for *Conopomorpha cramerella* subsamples, and over all loci (All). The 95% CI of bootstraps over loci is also represented (dashes), and the p -values for no subdivision can be seen under locus names. Under these p -values, result of tests for null alleles, SAD, stuttering and LD were also inserted.

<https://doi.org/10.1371/journal.pone.0297662.g003>

(Fig 3). Locus Cpb139 was an outlier for both F_{IS} and F_{ST} and was involved in a significant LD pair. Locus Cpb14 displayed a negative F_{ST} and was also an outlier for F_{IS} . Cpb55 was an outlier for F_{ST} and in significant LD, and Cpb135 was an outlier for F_{ST} , F_{IS} and for LD. All these complementary observations suggest that these loci are under selection of some sort (homogenizing or disruptive, depending on the locus) and we decided to eliminate these loci in order to limit as much as possible all biases of this kind.

The selected seven CPB microsatellite loci (Cpb54, Cpb62, Cpb84, Cpb112, Cpb122, Cpb160 and Cpb190) were characterized and the results indicated that the number of alleles ranged from 12 (Cpb190) to 19 (Cpb84) with an average of 16 alleles per locus. The total genetic diversity (H_t) ranged from 0.85 (Cpb190) to 0.896 (Cpb84) with an average of 0.869 (Table 3). Wright’s F -statistics were $F_{IS} = 0.072$, p -value = 0.0002, 95% CI = [-0.042, 0.197]), and $F_{ST} = 0.044$, p -value = 0.0001, 95% CI = [0.026, 0.067]) (S1 Fig). The correlation between the F_{IS} values and the number of missing genotypes (N_{Blanks}) displayed a negative slope (Fig 5). Cpb62 and Cpb122 had a negative F_{IS} with 2 and 9 missing data respectively. Obviously, these missing data are not null homozygotes since they should display 0 null homozygotes, hence 0 missing data due to null alleles. If we ignore these loci from the regression, the slope becomes positive and the correlation between F_{IS} N_{Blanks} is now significant ($\rho = 0.8660$, p -value = 0.0288) (S2 Fig). Moreover, we noticed that the averages of null frequency estimates (S6 Table) are far from the 0.20 threshold described in Séré et al. [61], and that null alleles explained 80% of F_{IS} variation, with an intercept (F_{IS} with no null alleles) $F_{IS,0} = -0.03$.

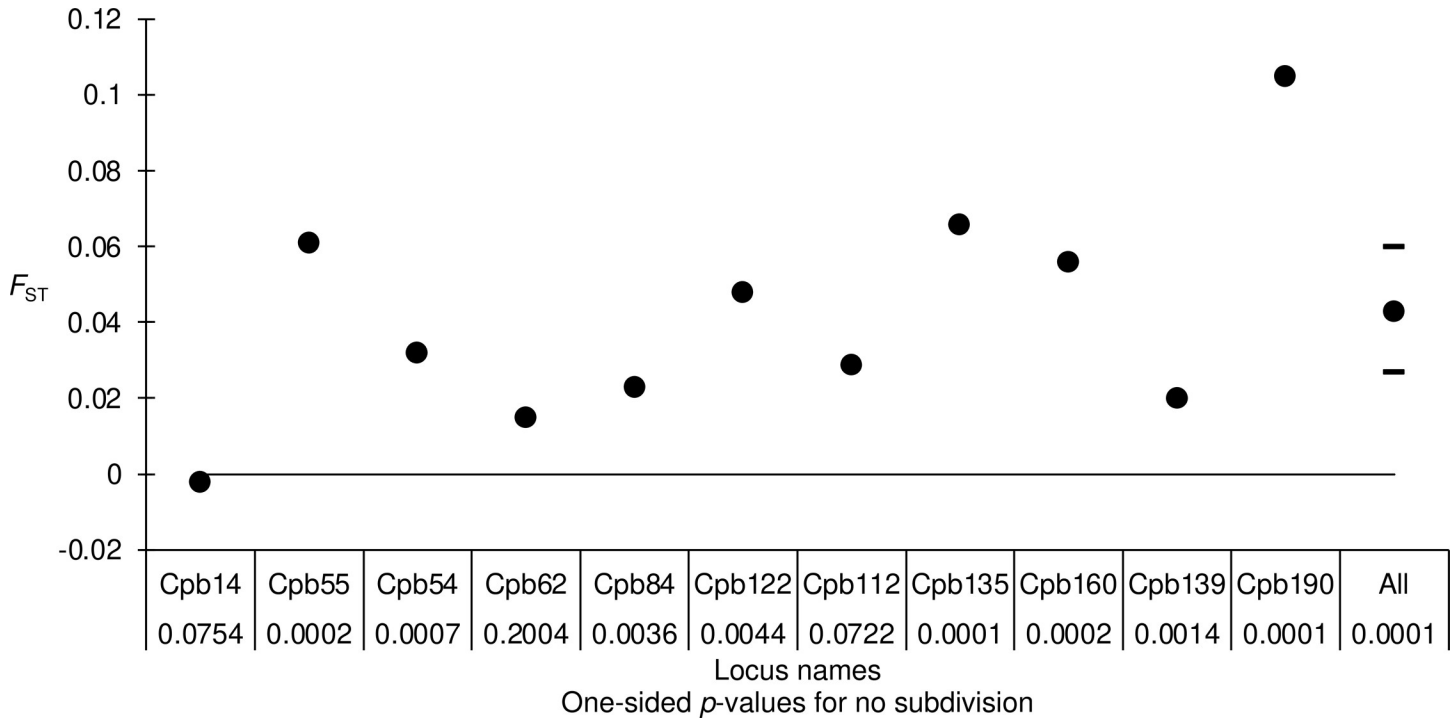


Fig 4. Variation of F_{ST} across the 11 loci for *Conopomorpha cramerella* subsamples, and over all loci (All). The 95% CI of bootstraps over loci is also represented (dashes), and the p -values for no subdivision can be seen under locus names.

<https://doi.org/10.1371/journal.pone.0297662.g004>

Population structure of *C. cramerella* at the seven selected loci

For this analysis, we kept loci Cpb54, Cpb62, Cpb84, Cpb122, Cpb112, Cpb160 and Cpb190, with 0, 2, 0, 9, 0, 1, and 1 missing data respectively, and recoded missing data as homozygous for the null allele (999999) for only loci Cpb160 and Cpb190. Globally, $F_{ST} = 0.044$ in 95% CI = [0.026, 0.067] was highly significant (p -value < 0.0001) and its variation mainly explained by null alleles ($R^2 = 0.67$ for the regression of F_{ST} and N_{Blanks}) (S2 Fig). After correction for null alleles and excess of polymorphism, we obtained $F_{ST_FreeNA} = 0.2629$ in 95% CI = [0.1681, 0.3768], which translated in $N_e m = 0.7$ in 95%, CI = [0.4, 1.2] individuals per generation. Given that this pest has a one month generation time all year long [62], and given the large distances between the subsamples (between 1159 and 2693 km), this potentially represents a

Table 3. Characterization of the seven (7) polymorphic microsatellite loci of cacao pod borer *Conopomorpha cramerella*.

Locus	Repeat motif and number	% Amplification	Size range (bp)*	N_A	H_o	H_s	H_t
Cpb54	(AC) ₁₄	100	131–281	18	0.972	0.869	0.889
Cpb62	(AC) ₁₃	97.22	185–257	13	0.957	0.858	0.867
Cpb84	(AC) ₁₂	100	197–245	19	0.806	0.882	0.896
Cpb122	(AC) ₁₁	87.5	184–226	18	0.854	0.846	0.877
Cpb112	(AC) ₁₂	100	184–310	16	0.778	0.83	0.847
Cpb160	(AC) ₁₁	98.61	264–302	13	0.547	0.826	0.859
Cpb190	(AC) ₁₁	98.61	281–307	12	0.565	0.789	0.85

N_A = number of alleles, H_o = observed heterozygosity, H_s = expected heterozygosity within samples, H_t = expected heterozygosity between samples

* Allele size includes the M13 adapter tail (19 bp).

<https://doi.org/10.1371/journal.pone.0297662.t003>

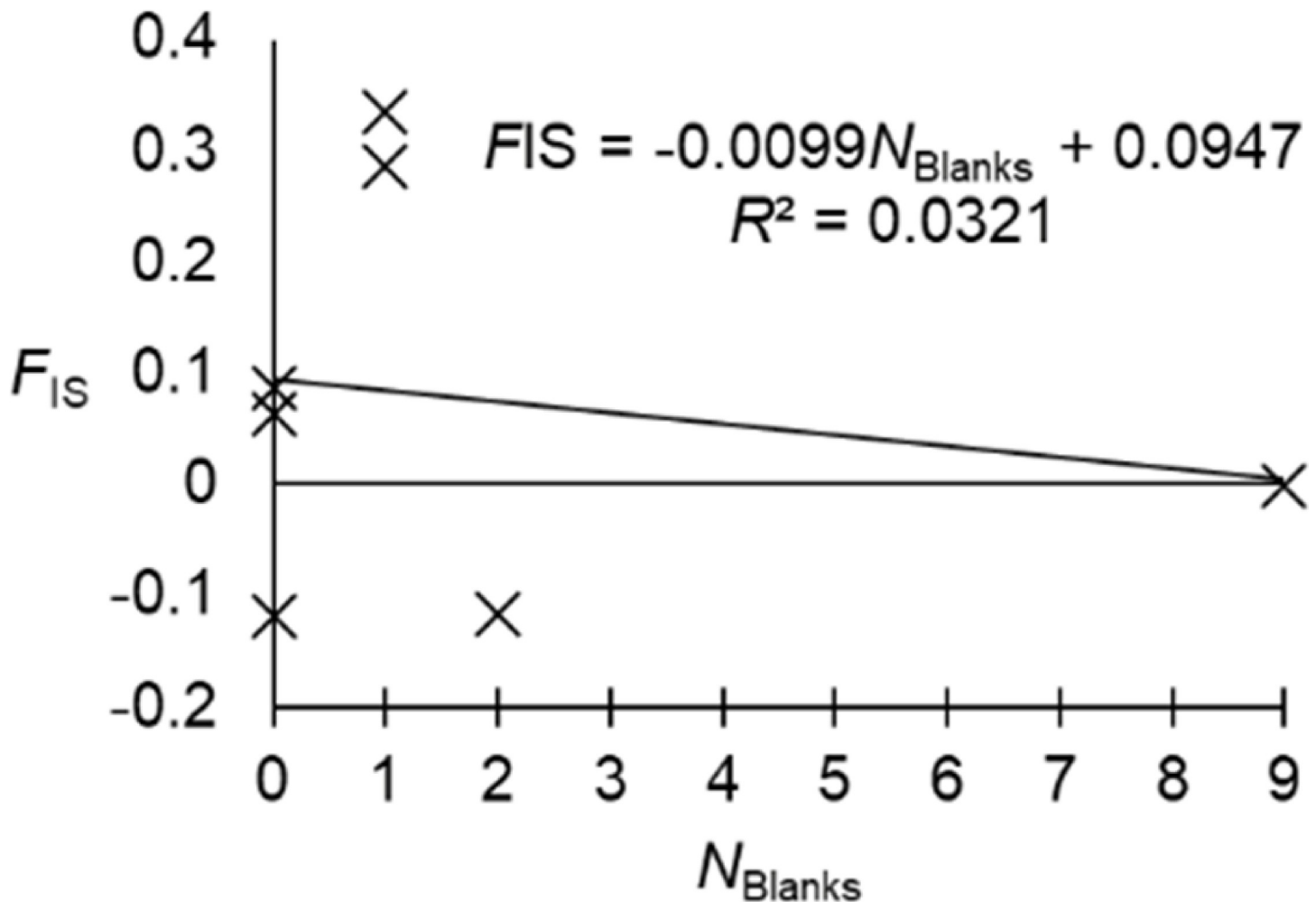


Fig 5. Regression between F_{IS} and number of missing genotypes (N_{Blanks}) for the selected seven loci of *Conopomorpha cramerella* after exclusion of putatively loci under selection.

<https://doi.org/10.1371/journal.pone.0297662.g005>

substantial exchange of 8 in 95% CI = [1, 14] individuals per year between the three countries. Effective population sizes were highly variable across methods and subsamples: average $N_e = 2128$ in minimax = [45, 4228] (Fig 6). This translated into similar dispersal estimates assuming an Island model, with average distances between subsamples, or computed from paired subsamples. On average $\delta \approx 1$ km per generation, with an important window of possible values: i.e. between 100 m and 60 km per generation (Fig 7).

Quality assessment of microsatellite loci for *C. sinensis*

This analysis was carried out with the data from Vietnam, after exclusion of individual 77, which appeared to belong to CPB (see above). No significant LD was detected. We observed a highly significant heterozygote deficit, which was quite variable across loci ($F_{IS} = 0.124$, p -value = 0.0001, 95% CI = [0.017, 0.242]) (Fig 8). No F_{ST} or F_{IT} were available (only a single subsample). SAD tests were undertaken with the weighted regression of F_{IS} only. No evidence of SAD or stuttering could be found, and null alleles explained the data rather well. More missing data were observed as compared to those expected if null alleles explained all heterozygote deficits (all p -values > 0.7), and the correlation between F_{IS} and missing data was significant ($\rho =$

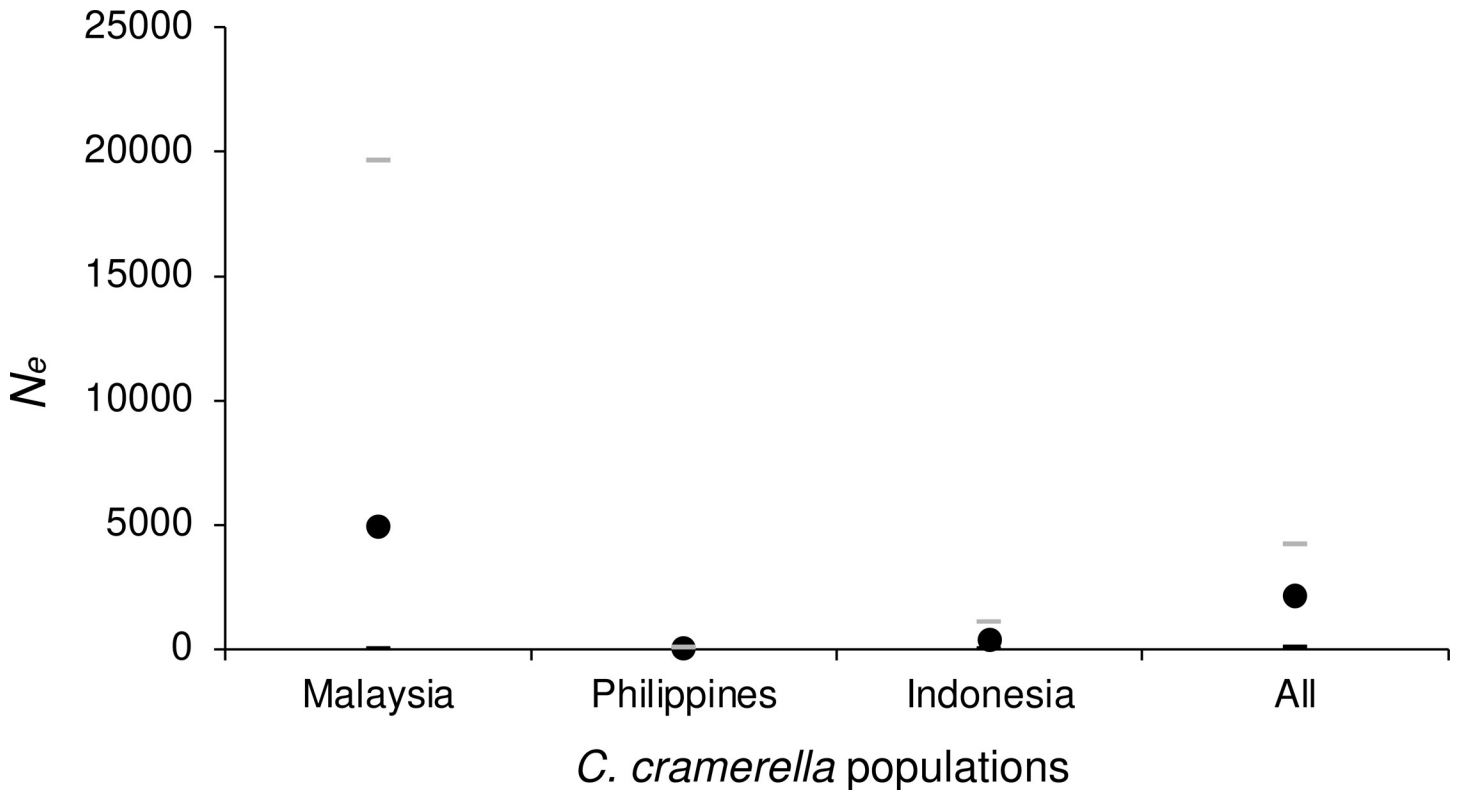


Fig 6. Effective population sizes of the cocoa pod borer *Conopomorpha cramerella* populations using different methods. [•]: Average population size, [—]: Upper and lower limits.

<https://doi.org/10.1371/journal.pone.0297662.g006>

0.5861, p -value = 0.0290), and the corresponding regression model explained 48% of F_{IS} variation (S3 Fig). On the other hand, using null allele frequency estimates, the average was also far from the threshold and null alleles explained 75% of F_{IS} variation with $F_{IS,0} = -0.031$ (S6 Table). It can also be noted that locus Cpb62 was almost monomorphic ($H_S = 0.048$).

Population genetics structure of *C. sinensis* on the eleven loci

With a single subsample (Vietnam), we could only estimate effective population size. To obtain minimum and maximum values with LDN_e , $CoAN_e$, IN_e and SN_e , we kept the 95% CI values

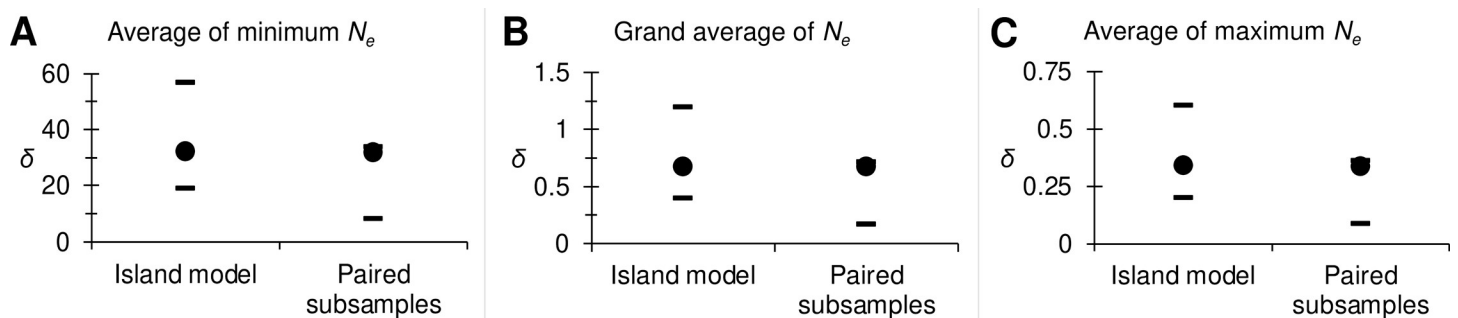


Fig 7. Dispersal estimates (δ) assuming an Island model and paired subsamples based on different population size (N_e) estimates: (A) minimum, (B) grand average and (C) maximum.

<https://doi.org/10.1371/journal.pone.0297662.g007>

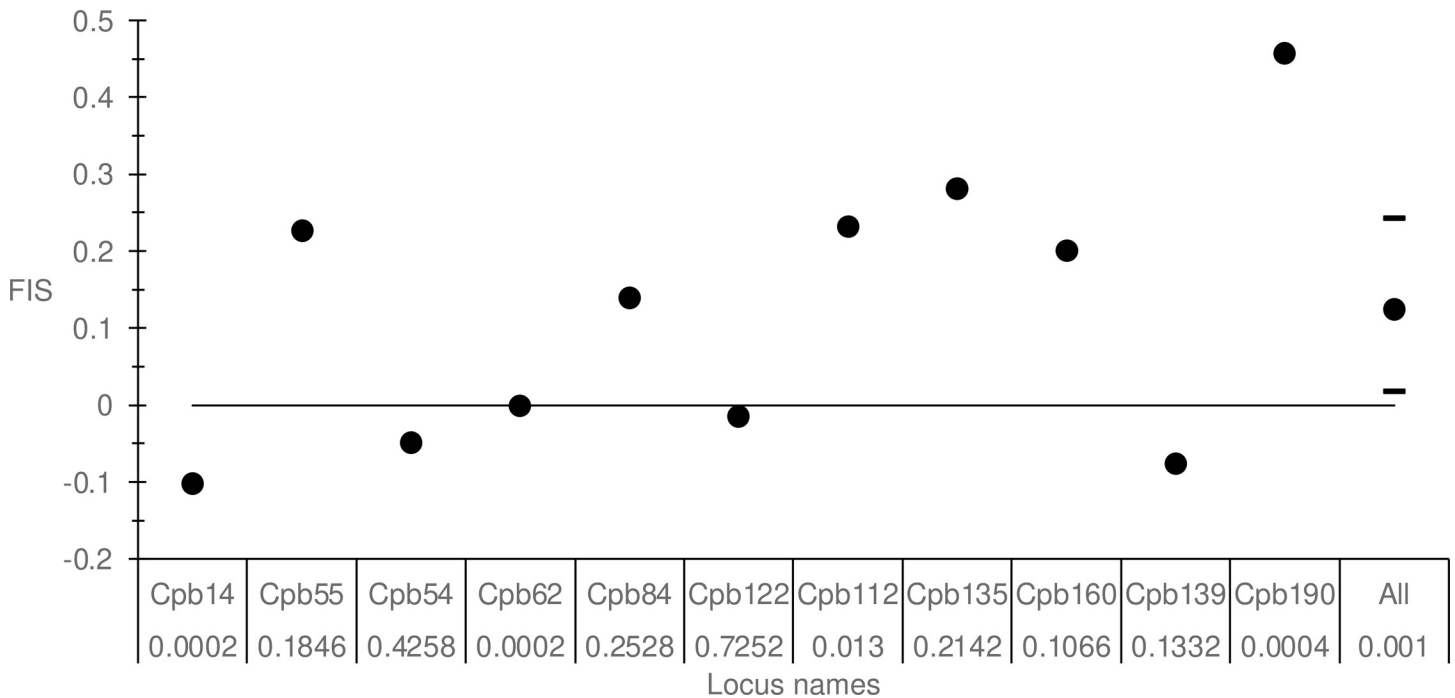


Fig 8. Quality assessment of the cocoa pod borer *Conopomorpha cramerella* microsatellite loci on lychee fruit borer *Conopomorpha sinensis* samples. [•]: F_{IS} per locus, [-]: Upper and lower limits.

<https://doi.org/10.1371/journal.pone.0297662.g008>

provided by the output of the corresponding software. We observed on average relatively smaller values, with $N_e \approx 50$ in $\text{minimax} = [16, 191]$, but in view that the overlap with CPB populations was substantial, nothing conclusive could be extrapolated.

Discussion

The goal of this study was to develop new microsatellite markers that could effectively be used to explore the genetic differences between different populations of the CPB and test the selected microsatellites on a closely related species, the LFB. To the best of our knowledge, there were no microsatellite markers available for these insects, and therefore, no available information on the population's genetic structure of these pests. The availability of these microsatellite markers will allow analyzing the population genetics of this species and therefore assess the level of gene flow between infested countries in Southeastern Asia. This information would be of paramount importance for trading barriers and for the development of AW-IPM strategies that could potentially include the SIT to manage various populations of this pest.

The availability of the partial genome sequence of the CPB enabled us to search for potential microsatellite markers. The search for microsatellite markers indicated that more tri-nucleotides than di-nucleotides markers were found, which is contradictory to the theory that the microsatellite abundance decreases with the increase of motif repeat number and repeat length [63]. This might be due to the incomplete genome of the CPB that was used for the analysis, and the sequence that is missing might have contained more di-nucleotides sequences. On the other hand, our result agrees with the data of Liu et al. [64] who also obtained more frequent tri-nucleotides than di-nucleotides repeats in the hoverfly, *Eupeodes corollae*.

Marker development allowed to pre-select 11 loci that worked well on the two species, CPB and LFB. Following the quality control tests, 7 microsatellite markers can be used for the study

of the CPB populations, and 11 markers for the LFB wherein the 7 markers that were selected for CPB are also usable in the LFB. Most loci were affected by null alleles, while SAD and stuttering were not evidenced at any loci. Locus Cpb62 was found diagnostic between the two species and will be useful as additional tool in case of uncertain morphological determinations.

Effective population sizes varied from 50 to 4200 individuals for CPB from Indonesia, Malaysia and Philippines, and between 20 and 200 for LFB from Vietnam. Whether these differences are significant will require more subsamples of both species. Dispersal distances could be estimated from subdivision measures between sites occupied by the cocoa moth populations. Using this limited sampling effort, it seemed that this species can disperse over quite long distances, i.e. on average 17 km per generation, and with minimum and maximum distances of 9 and 31 km per generation, respectively. In view that a generation lasts only one month, dispersal over a twelve month period can be very large. Despite the potential large dispersal capabilities of CPB, no hybrids with LFB could be detected in areas where both crops are cultivated in close proximity. However, the large dispersal capacity observed in our analysis might also be due to human induced involvement of infested fruit with larvae between countries/islands. This however will require more sampling of both species in closely located cacao and lychee plantations.

The results indicate the effectiveness of the selected loci to explore differences between CPB populations. A previous study analyzed the genetic structure of the CPB from Malaysia using mitochondrial DNA markers, and indicated that this species might have been exposed to a bottleneck event [15]. The availability of microsatellite loci is expected to provide more thorough and in-depth analysis that might explore more genetic differences in the different populations. Determining the level of gene flow between CPB populations in neighbouring countries might have economic consequences with respect to trading regulations and might help to develop sustainable AW-IPM strategies at the regional level.

Supporting information

S1 Table. Selected primers of the cocoa pod borer *Conopomorpha cramerella* microsatellites. The primers were selected using msatcommander software with the partial genome available in the database, primers predicted to amplify a product length ranged between 180–300 nucleotides with minimum 11 repeat of each motif were selected.
(XLSX)

S2 Table. List of primers successfully amplify PCR the expected PCR product with the DNA samples of the cocoa pod borer *Conopomorpha cramerella* and the litchi fruit borer *Conopomorpha sinensis*. *: primers did not amplify PCR product in litchi fruit borer *Conopomorpha sinensis* samples collected from Vietnam.
(XLSX)

S3 Table. List of genotypes (alleles) per locus for the cocoa pod borer *Conopomorpha cramerella* and the lychee fruit borer *Conopomorpha sinensis*. Dye-labeled PCR were analysed by fragment analyzer and the resulted data were read with Genemapper.
(CSV)

S4 Table. Comparison of Weir and Cockerham's unbiased estimator of Wright's F_{IS} in males and females of using Wilcoxon signed rank test.
(XLSX)

S5 Table. Weir and Cockerham's unbiased estimator of Wright's F_{ST} measured between females and males within each country and for each locus, and G-based randomization

test for significant of divergence.

(XLSX)

S6 Table. Null allele frequency estimates.

(XLSX)

S1 Fig. (A) Wright's F_{IS} with p-values for panmixia, null alleles, SAD, stuttering and null alleles and (B) F_{ST} with p-values for subdivision of the seven selected microsatellite loci of cocoa pod borer *Conopomorpha cramerella*. [•]: F_{IS} or F_{ST} per locus, [x]: F_{IS} per population [-]: 95% CI.

(TIF)

S2 Fig. Correlation between the F_{IS} and F_{ST} values and the missing genotypes (N_{Blanks}) of cocoa pod borer *Conopomorpha cramerella*. The correlation shown is after assuming that the missing data for Cpb122 and Cpb62 are not null homozygotes.

(TIF)

S3 Fig. Correlation between the F_{IS} values and the missing genotypes (N_{Blanks}) of lychee fruit borer *Conopomorpha sinensis*.

(TIF)

S1 File. R Markdowns.

(DOCX)

S2 File. Inclusivity in global research document.

(PDF)

Acknowledgments

This study was financed by the International Atomic Energy Agency (IAEA), Austria.

Author Contributions

Conceptualization: Marc J. B. Vreysen, Adly M. M. Abd-Alla.

Data curation: Marynold Purificacion, Roslina Binti Mohd Shah, Saripah Binti Bakar, Anisah Binti Savantil, Meriam Mohd Yusof, Divina Amalin, Hien Nguyen, Endang Sulistyowati, Aris Budiman, Arni Ekayanti, Jerome Niogret, Adly M. M. Abd-Alla.

Formal analysis: Marynold Purificacion, Thierry De Meeûs, Sophie Ravel, Adly M. M. Abd-Alla.

Funding acquisition: Marc J. B. Vreysen.

Investigation: Marynold Purificacion, Roslina Binti Mohd Shah.

Methodology: Marynold Purificacion, Roslina Binti Mohd Shah, Sophie Ravel, Adly M. M. Abd-Alla.

Project administration: Marc J. B. Vreysen.

Resources: Saripah Binti Bakar, Anisah Binti Savantil, Meriam Mohd Yusof, Divina Amalin, Hien Nguyen, Endang Sulistyowati, Aris Budiman, Arni Ekayanti, Jerome Niogret, Adly M. M. Abd-Alla.

Software: Sophie Ravel, Adly M. M. Abd-Alla.

Validation: Thierry De Meeûs, Adly M. M. Abd-Alla.

Visualization: Marynold Purificacion, Thierry De Meeûs, Adly M. M. Abd-Alla.

Writing – original draft: Adly M. M. Abd-Alla.

Writing – review & editing: Roslina Binti Mohd Shah, Saripah Binti Bakar, Anisah Binti Savantil, Meriam Mohd Yusof, Divina Amalin, Hien Nguyen, Endang Sulistyowati, Aris Budiman, Arni Ekayanti, Jerome Niogret, Sophie Ravel, Marc J. B. Vreysen.

References

- Japar A, Yazik NM, Ramba H. Evaluation of international cocoa clones at CRDC Madai, Sabah. *Malays Cocoa J.* 2021; 13: 39–43. Available: <https://www.koko.gov.my/emcj/ViewPDF.aspx?ID=30>
- Hebbar PK. Cacao Diseases: A Global Perspective from an Industry Point of View. *Phytopathology*®. 2007; 97: 1658–1663. <https://doi.org/10.1094/PHYTO-97-12-1658> PMID: 18943730
- Wahyudi T. The world scenario of cocoa production and consumption. In 2nd International Plantation Industry Conference and Exhibition (IPiCEX), UiTM Shah Alam, Selangor, Malaysia 2008 Nov (pp. 18–21). Selangor, Malaysia; 2008. pp. 18–21. Available: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=22c32822d0d8f82983e8175cf6c08bc0df9b9ebd>
- Niogret J, Ekayanti A, Ingram K, Lambert S, Kendra PE, Alborn H, et al. Development and behavioral ecology of *Conopomorpha cramerella* (Lepidoptera: Gracillariidae). 2019. 2019; 102: 382–387. <https://doi.org/10.1653/024.102.0214>
- Niogret J, Ekayanti A, Kendra PE, Ingram K, Lambert S, Epsky ND, et al. Host preferences of the cocoa pod borer, *Conopomorpha cramerella*, the main threat to cocoa production in Southeast Asia. *Entomol Exp Appl.* 2020; 168: 221–227. <https://doi.org/10.1111/eea.12882>
- Chang H, Guo J, Li M, Gao Y, Wang S, Wang X, et al. Comparative genome and phylogenetic analysis revealed the complex mitochondrial genome and phylogenetic position of *Conopomorpha sinensis* Bradley. *Sci Rep.* 2023; 13: 4989. <https://doi.org/10.1038/s41598-023-30570-7> PMID: 36973296
- Zhang A, Kuang LF, Maisin N, Karumuru B, Hall DR, Virdiana I, et al. Activity Evaluation of Cocoa Pod Borer Sex Pheromone in Cacao Fields. *Environ Entomol.* 2008; 37: 719–724. [https://doi.org/10.1603/0046-225x\(2008\)37\[719:aeocpb\]2.0.co;2\[719:AEOCPB\]2.0.CO;2](https://doi.org/10.1603/0046-225x(2008)37[719:aeocpb]2.0.co;2[719:AEOCPB]2.0.CO;2) PMID: 18559177
- Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J. Biodiversity hotspots for conservation priorities. *Nature.* 2000; 403: 853–858. <https://doi.org/10.1038/35002501> PMID: 10706275
- Spencer P, Woolnough A. Size should matter: Distribution and genetic considerations for pest animal management. *Ecol Manag Restor.* 2005; 5: 231–234. <https://doi.org/10.1111/j.1442-8903.2004.209-9.x>
- De Meeûs T, McCoy KD, Prugnolle F, Chevillon C, Durand P, Hurtrez-Bousses S, et al. Population genetics and molecular epidemiology or how to “de ’busquer la be ’te.” *Infect Genet Evol.* 2007; 7: 308–332. <https://doi.org/10.1016/j.meegid.2006.07.003> PMID: 16949350
- Gooding RH, Krafur ES. Tsetse genetics: Contributions to biology, systematics, and control of tsetse flies. *Annu Rev Entomol.* 2005; 50: 101–123. <https://doi.org/10.1146/annurev.ento.50.071803.130443> PMID: 15355235
- Behura S. Molecular marker systems in insects: current trends and future avenues. *Mol Ecol.* 2006; 15: 3087–113. <https://doi.org/10.1111/j.1365-294X.2006.03014.x> PMID: 16968257
- Gopurenko D, Gillespie PS, Minana R, Reynolds OL. DNA barcode identification of *Conopomorpha cramerella* (Snellen, 1904) (Lepidoptera: Gracillariidae) and other moths affecting cacao in Papua New Guinea. *Austral Entomol.* 2021; 60: 598–609. <https://doi.org/10.1111/aen.12559>
- Field D, Wills C. Long, polymorphic microsatellites in simple organisms. *Proc R Soc B Biol Sci.* 1996; 263: 209–15. <https://doi.org/10.1098/rspb.1996.0033> PMID: 8728984
- Shapiro LH, Scheffer SJ, Maisin N, Lambert S, Purung H, Sulistyowati E, et al. *Conopomorpha cramerella* (Lepidoptera: Gracillariidae) in the Malay Archipelago: genetic signature of a bottlenecked population?. 2008 Sep 1; 101(5):930–8. 2008; 101: 930–938. [https://doi.org/10.1603/0013-8746\(2008\)101\[930:cclgit\]2.0.co;2](https://doi.org/10.1603/0013-8746(2008)101[930:cclgit]2.0.co;2)
- Srivastava K, Choudhary JS, Patel RK, Reddy PVR, Nath V. Identification and phylogenetic analysis of fruit borer species of litchi using DNA barcode sequences. *Indian J Hortic.* 2018; 75: 415. <https://doi.org/10.5958/0974-0112.2018.00071.3>
- Meng X, Hu J, Li Y, Dai J, Guo M, Ouyang G. The preference choices of *Conopomorpha sinensis* Bradley (Lepidoptera: Gracillariidae) for litchi based on its host surface characteristics and volatiles. *Sci Rep.* 2018; 8: 2013. <https://doi.org/10.1038/s41598-018-20383-4> PMID: 29386547

18. Schulte MJ, Martin K, Sauerborn J. Biology and control of the fruit borer, *Conopomorpha sinensis* Bradley on litchi (*Litchi chinensis* Sonn.) in northern Thailand. *Insect Sci.* 2007; 14: 525–529. <https://doi.org/10.1111/j.1744-7917.2007.00182.x>
19. Posada FJ, Virdiana I, Navies M, Pava-Ripoll M, Hebbar P. Sexual Dimorphism of Pupae and Adults of the Cocoa Pod Borer, *Conopomorpha cramerella*. *J Insect Sci.* 2011; 11: 1–8. <https://doi.org/10.1673/031.011.5201> PMID: 21861656
20. Faircloth BC. msatcommander: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Mol Ecol Resour.* 2008; 8: 92–94. <https://doi.org/10.1111/j.1471-8286.2007.01884.x> PMID: 21585724
21. Coombs JA, Letcher BH, Nislow KH. create: a software to create input files from diploid genotypic data for 52 genetic software programs. *Mol Ecol Resour.* 2008; 8: 578–580. <https://doi.org/10.1111/j.1471-8286.2007.02036.x> PMID: 21585837
22. Traut W, Marec F. Sex Chromosome Differentiation in Some Species of Lepidoptera (Insecta). *Chromosome Res.* 1997; 5: 283–291. <https://doi.org/10.1023/B:CHRO.0000038758.08263.c3:CHRO.0000038758.08263.c3> PMID: 9292232
23. Wright S. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution.* 1965; 19: 395–420. <https://doi.org/10.1111/j.1558-5646.1965.tb01731.x>
24. Goudet J. Fstat (ver. 2.9.4), a program to estimate and test population genetics parameters. Available at <http://www.t-de-meeus.fr/Programs/Fstat294.zip>, Updated from Goudet (1995). 2003. Available: <http://www.t-de-meeus.fr/Programs/Fstat294.zip>
25. De Meeùs T, Guégan J-F, Teriokhin AT. MultiTest V.1.2, a program to binomially combine independent tests and performance comparison with other related methods on proportional data. *BMC Bioinformatics.* 2009; 10: 443. <https://doi.org/10.1186/1471-2105-10-443> PMID: 20030807
26. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B Methodol.* 1995; 57: 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
27. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000; 155: 945–959. <https://doi.org/10.1093/genetics/155.2.945> PMID: 10835412
28. Glaubitz JC. convert: A user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Mol Ecol Notes.* 2004; 4: 309–310. <https://doi.org/10.1111/j.1471-8286.2004.00597.x>
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: 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x> PMID: 15969739
30. Earl DA, Vonholdt B. Structure Harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour.* 2012;4.
31. Cavalli-Sforza LL, Edwards AW. Phylogenetic analysis. Models and estimation procedures. *Am J Hum Genet.* 1967; 19: 233–257. PMID: 6026583
32. Chapuis M-P, Estoup A, Chapuis MP, Estoup A. Microsatellite null alleles and estimation of population differentiation. *Mol Bio Evol* 24: 621–631. *Mol Biol Evol.* 2007; 24: 621–31. <https://doi.org/10.1093/molbev/msl191> PMID: 17150975
33. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 1987; 4: 406–425. <https://doi.org/10.1093/oxfordjournals.molbev.a040454> PMID: 3447015
34. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol.* 2016; 33: 1870–1874. <https://doi.org/10.1093/molbev/msw054> PMID: 27004904
35. Takezaki N, Nei M. Genetic Distances and Reconstruction of Phylogenetic Trees From Microsatellite DNA. *Genetics.* 1996; 144: 389–399. <https://doi.org/10.1093/genetics/144.1.389> PMID: 8878702
36. Goudet J, Raymond M, De Meeùs T, Rousset F. Testing differentiation in diploid populations. *Genetics.* 1996; 144: 1933–1940. <https://doi.org/10.1093/genetics/144.4.1933> PMID: 8978076
37. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. *Ann Stat.* 2001; 29: 1165–1188. <https://doi.org/10.1214/aos/1013699998>
38. R Core Team. R: A language and environment for statistical computing. R Found Stat Comput Vienna Austria. 2021. Available: <http://www.R-project.org>
39. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evol Int J Org Evol.* 1984; 38: 1358–1370. <https://doi.org/10.1111/j.1558-5646.1984.tb05657.x> PMID: 28563791
40. De Meeùs T. Revisiting FIS, FST, Wahlund Effects, and Null Alleles. *J Hered.* 2018; 109: 446–456. <https://doi.org/10.1093/jhered/esx106> PMID: 29165594

41. De Meeûs T, Chan CT, Ludwig JM, Tsao JI, Patel J, Bhagatwala J, et al. Deceptive combined effects of short allele dominance and stuttering: an example with *Ixodes scapularis*, the main vector of Lyme disease in the U.S.A. *Peer Community J*. 2021; 622373. <https://doi.org/10.24072/pcjournal.34>
42. Manangwa O, De Meeûs T, Grébaud P, Ségard A, Byamungu M, Ravel S. Detecting Wahlund effects together with amplification problems: cryptic species, null alleles and short allele dominance in *Glossina pallidipes* populations from Tanzania. *Mol Ecol Resour*. 2019; 19: 757–772. <https://doi.org/10.1111/1755-0998.12989> PMID: 30615304
43. De Meeûs T, Noûs C. A new and almost perfectly accurate approximation of the eigenvalue effective population size of a dioecious population: comparisons with other estimates and detailed proofs. *Peer Community J*. 2023;3. <https://doi.org/10.24072/pcjournal.280>
44. De Meeûs T, Humair PF, Grunau C, Delaye C, Renaud F. Non-Mendelian transmission of alleles at microsatellite loci: an example in *Ixode ricinus*, the vector of Lyme disease. *Int J Parasitol*. 2004; 34: 943–950. <https://doi.org/10.1016/j.ijpara.2004.04.006> PMID: 15217733
45. Meirmans P. Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evol Int J Org Evol*. 2006; 60: 2399–402. <https://doi.org/10.1111/j.0014-3820.2006.tb01874.x> PMID: 17236430
46. Meirmans PG, Hedrick PW. Assessing population structure: F_{ST} and related measures. *Mol Ecol Resour*. 2011; 11: 5–18. <https://doi.org/10.1111/j.1755-0998.2010.02927.x> PMID: 21429096
47. Waples RS. A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci*. *Conserv Genet*. 2006; 7: 167. <https://doi.org/10.1007/s10592-005-9100-y>
48. Peel D, Waples RS, Macbeth GM, Do C, Ovenden JR. Accounting for missing data in the estimation of contemporary genetic effective population size (N_e). *Mol Ecol Resour*. 2013; 13: 243–253. <https://doi.org/10.1111/1755-0998.12049> PMID: 23280157
49. Nomura T. Estimation of effective number of breeders from molecular coancestry of single cohort sample. *Evol Appl*. 2008; 1: 462–474. <https://doi.org/10.1111/j.1752-4571.2008.00015.x> PMID: 25567728
50. Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. *Mol Ecol Resour*. 2014; 14: 209–214. <https://doi.org/10.1111/1755-0998.12157> PMID: 23992227
51. Vitalis R, Couvet D. ESTIM 1.0: A computer program to infer population parameters from one- And two-locus gene identity probabilities. *Mol Ecol Notes*. 2001; 1: 354–356. <https://doi.org/10.1046/j.1471-8278.2001.00086.x>
52. Vitalis R, Couvet D. Estimation of Effective Population Size and Migration Rate From One- and Two-Locus Identity Measures. *Genetics*. 2001; 157: 911–925. <https://doi.org/10.1093/genetics/157.2.911> PMID: 11157007
53. Vitalis R, Couvet D. Two-locus identity probabilities and identity disequilibrium in a partially selfing subdivided population. *Genet Res*. 2001; 77: 67–81. <https://doi.org/10.1017/s0016672300004833> PMID: 11279832
54. Wang J. A new method for estimating effective population sizes from a single sample of multilocus genotypes. *Mol Ecol*. 2009; 18: 2148–2164. <https://doi.org/10.1111/j.1365-294X.2009.04175.x> PMID: 19389175
55. Jones OR, Wang J. COLONY: a program for parentage and sibship inference from multilocus genotype data. *Mol Ecol Resour*. 2010; 10: 551–555. <https://doi.org/10.1111/j.1755-0998.2009.02787.x> PMID: 21565056
56. De Meeûs T. Initiation à la génétique des populations naturelles: application aux parasites et à leurs vecteurs. IRD Éditions; 2021. Available: <https://books.google.com.ph/books?id=q-lxzgEACAAJ>
57. Hijmans R. Geosphere: Spherical Trigonometry_. R package version 1.5–18,. 2022. Available: <https://CRAN.R-project.org/package=geosphere>
58. De Meeûs T, Noûs C. A simple procedure to detect, test for the presence of stuttering, and cure stuttered data with spreadsheet programs. *Peer Community J*. 2022;2. <https://doi.org/10.24072/pcjournal.165>
59. Pudovkin AI, Zaykin DV, Hedgecock D. On the Potential for Estimating the Effective Number of Breeders From Heterozygote-Excess in Progeny. *Genetics*. 1996; 144: 383–387. <https://doi.org/10.1093/genetics/144.1.383> PMID: 8878701
60. Balloux F. Heterozygote excess in small populations and the heterozygote-excess effective population size. *Evol Int J Org Evol*. 2004; 58: 1891–1900. <https://doi.org/10.1111/j.0014-3820.2004.tb00477.x> PMID: 15521449
61. Séré M, Thévenon S, Belem AMG, De Meeûs T. Comparison of different genetic distances to test isolation by distance between populations. *Heredity*. 2017; 119: 55–63. <https://doi.org/10.1038/hdy.2017.26> PMID: 28537571

62. Day RK. Effect of cocoa pod borer, *Conopomorpha cramerella*, on cocoa yield and quality in Sabah, Malaysia. *Crop Prot.* 1989; 8: 332–339. [https://doi.org/10.1016/0261-2194\(89\)90052-5](https://doi.org/10.1016/0261-2194(89)90052-5)
63. Wang H-L, Yang J, Boykin LM, Zhao Q-Y, Wang Y-J, Liu S-S, et al. Developing conversed microsatellite markers and their implications in evolutionary analysis of the *Bemisia tabaci* complex. *Sci Rep.* 2014; 4: 6351. <https://doi.org/10.1038/srep06351> PMID: 25220501
64. Liu M, Wang X, Ma L, Cao L, Liu H, Pu D, et al. Genome-wide developed microsatellites reveal a weak population differentiation in the hoverfly *Eupeodes corollae* (Diptera: Syrphidae) across China. *PloS One.* 2019; 14: e0215888. <https://doi.org/10.1371/journal.pone.0215888> PMID: 31557189