Scientific Research Publishing

Advances in Entomology, 2026, 14(2), 159-177
https://www.scirp.org/journal/ae
ISSN Online: 2331-2017
ISSN Print: 2331-1991

Genetic Characterization of Populations of the Global Invasive Agricultural Pest *Bactrocera dorsalis* (Diptera: Tephritidae)

Toffène Diome[1], Madeleine Ivonne Mendy[1], Ablaye Faye[1], Toihir Mahadi[1], Mamecor Faye[2], Mbacké Sembène[1]

[1]Genetic Team and Population Management, Department of Animal Biology, Faculty of Sciences and Technology, Cheikh Anta Diop University, Dakar, Senegal
[2]Laboratory of Parasitology, Department of Animal Biology, Faculty of Sciences and Technology, Cheikh Anta Diop University, Dakar, Senegal
Email: toffene.diome@ucad.edu.sn

How to cite this paper: Diome, T., Mendy, M.I., Faye, A., Mahadi, T., Faye, M. and Sembène, M. (2026) Genetic Characterization of Populations of the Global Invasive Agricultural Pest *Bactrocera dorsalis* (Diptera: Tephritidae). *Advances in Entomology*, **14**, 159-177.
https://doi.org/10.4236/ae.2026.142010

Received: September 15, 2025
Accepted: April 25, 2026
Published: April 28, 2026

Abstract

Bactrocera dorsalis (Hendel, 1912), also known as the oriental fruit fly, is one of the most invasive agricultural pests worldwide. We analyzed 351 mitochondrial cytochrome oxidase I (COI) gene sequences. The genetic diversity of individuals from 19 populations across four continents (Africa, America, Asia, Oceania) was assessed. Our results revealed high haplotype diversity but low nucleotide diversity. AMOVA indicates that most of the genetic variation occurs within populations. Genetic differentiation is low, and the Mantel test showed no correlation between genetic and geographic distance. Demographic analyses suggested a recent expansion of Asian populations, whereas African and island populations exhibited signatures of founder effects or stability. Phylogenetic analysis showed two clades: one predominantly Asian and the other heterogeneous, including populations from all four continents. The existence of two clades, one of which is essentially Asian and the other heterogeneous, integrates the four continents. Overall, these findings enhance our understanding of invasion dynamics and may inform targeted management strategies against this pest.

Keywords

Bactrocera dorsalis, Genetic Diversity, Population Structure, COI

1. Introduction

Agriculture plays a central role in human survival, economic stability, and global

food security. It is estimated that agricultural production will need to increase by 60% by 2050 to feed more than 9 billion people [1]. In this context of increasing pressure on agricultural production systems, fruit crops hold a special place. In 2018, global fruit production reached 868 million tonnes [2]. Beyond their economic importance, fruits are of interest due to their potential benefits for human health. They are now recognized as functional foods, rich in fiber, vitamins, polyphenols, and anthocyanins, with antioxidant, anti-inflammatory, and protective effects against chronic diseases such as cardiovascular diseases, certain cancers, and diabetes [3] [4]. However, the production and sustainability of fruit crops are severely challenged by multiple biotic constraints, particularly insect pests. These pests represent a persistent and growing threat, intensified by global trade and the effects of climate change, thus posing a risk to food security [5] [6]. This is the case of the oriental fruit fly, *B. dorsalis* Hendel (Diptera: Tephritidae). Over the last two decades, *B. dorsalis* has become one of the most invasive and destructive insect pests. It causes substantial damage to tropical and subtropical fruits and vegetables, especially in sub-Saharan Africa and the Indian Ocean region [7] [8]. Due to its rapid adaptability to new environments and its strong dispersal capacity, *B. dorsalis* easily colonizes new habitats, spreading quickly via infested fruits or vegetables [9]. *Bactrocera dorsalis* is recognized as the dominant species at the top of the competitive hierarchy among fruit flies [10]. Indeed, it can displace and potentially drive to extinction other fruit fly species, including highly invasive ones such as *Ceratitis capitata*, *Ceratitis cosyra*, *Bactrocera tryoni*, and *Bactrocera zonata* [11]. Its high reproductive potential and contribution to massive agricultural losses make it a major threat to food security and international trade [12]. Effective monitoring and management of fruit crops are essential to ensure quality and productivity. The rapid expansion of *B. dorsalis* has prompted research into the genetic mechanisms underlying its ecological plasticity and invasive potential. Population genetic studies can help mitigate crop damage caused by *B. dorsalis* by unraveling the genetic structuring of oriental fruit fly populations. However, genetic characterization of this species at the intercontinental level has received relatively little attention, and few studies have addressed this question.

We hypothesize that the genetic diversity and differentiation of *B. dorsalis* populations are strongly influenced by substantial, yet fragmented, gene flow, particularly between regions of origin and recently colonized areas. This structuring is likely modulated by human activities and natural geographical barriers. The general objective of this study is to contribute to the genetic characterization of *B. dorsalis* populations on a global scale.

2. Materials and Methods

2.1. Study Areas

The study was conducted across 19 countries/regions on four continents: Africa: Senegal, Benin, Burkina Faso, Mali, Kenya; America: Hawaii, USA; Oceania: Papua New Guinea, French Polynesia; Asia: Thailand, Japan, Laos, Cambodia, Nepal,

Bangladesh, Malaysia, China, Sri Lanka, Philippines, and Vietnam.

2.2. Stady Population

This follows the definition criteria of [13] for an "evolutionary population". In total, 19 populations were considered.

3. Data Collection

The study focused on the cytochrome oxidase I (COI) gene and secondary data.

3.1. Interest in Cytochrome Oxidase I

COI is a subunit of the cytochrome c oxidase enzyme, a mitochondrial enzyme complex that plays a key role in the electron transport chain and in ATP production through oxidative phosphorylation. Like other mitochondrial genes, COI is commonly used as a DNA barcode for species identification, population genetics studies, and molecular evolution studies [14]-[17]. Some authors argue that the evolution of the COI gene is sufficiently rapid to differentiate closely related species and to investigate intraspecific variation [18] [19].

3.2. Sequence Acquisition

The nucleotide sequences used in this study were downloaded from the GenBank database (between December 27, 2024, and January 4, 2025), a public DNA sequence repository maintained by the National Center for Biotechnology Information (NCBI). The sequences were obtained by searching GenBank for the COI gene of *B. dorsalis*, filtering the results by country/region, and downloading them in FASTA format. For ease of analysis, the sequences were organized by geographic region, comprising 233 sequences from Malaysia, Thailand, Bangladesh, Sri Lanka, Vietnam, Philippines, China, Japan, Laos, Cambodia, and Nepal; 96 sequences from Benin, Burkina Faso, Kenya, Senegal, and Mali; 20 sequences from the Hawaii, and 22 sequences from French Polynesia and Papua New Guinea. This geographical diversity provides broad and representative coverage for subsequent analyses.

4. Genetic Analysis

4.1. Sequence Alignment

Multiple sequence alignment (MSA) is the first step in phylogenetic analysis. It is a fundamental task in biology with a wide range of applications, including homology detection, the identification of sites important for evolution [20], and phylogenetics [21]. Our sequences were aligned and refined using BioEdit software version 7.2.5 [22].

4.2. Basic Parameters of Population Genetic Diversity

The parameters of genetic variability can be considered as the identity profile of our dataset. The number of sites (N), the number of polymorphic sites, the number of parsimony-informative sites, the number of singleton sites, and the number

of nucleotide differences (k) were determined using DnaSP v6.12.03 software [23]. MEGA software version 11.0.13 [24] was used to calculate the nucleotide frequency, the types of substitutions—namely synonymous mutations (Ks) and non-synonymous mutations (Kns)—the nature of mutations (percentage of transitions and transversions), the total number of mutations (Eta), and the mutation rate (R), defined as the ratio between the percentage of transitions and the percentage of transversions.

4.3. Genetic Diversity Indices

Standard indices of genetic diversity were estimated using DnaSP v6.12.03 [23]. These include haplotypic diversity (Hd) and nucleotide diversity (Pi). The former defines the probability that two randomly sampled genes from a population are identical [25], taking into account the number of individuals and the frequency of haplotypes. Thus, the higher the Hd value, the greater the likelihood of observing different haplotypes between two randomly chosen individuals. The latter (Pi) is also a measure of genetic diversity, but it additionally reflects the actual degree of difference between the sequences analyzed. A signal of a stable population with a large effective size or admixture generally reflects demographic stability rather than a bottleneck. In contrast, a severe and prolonged bottleneck is typically associated with reduced haplotype diversity and altered nucleotide diversity patterns. However, high Hd coupled with low Pi would suggest rapid population growth from a small ancestral population, where insufficient time has elapsed to accumulate substantial sequence divergence among haplotypes. In contrast, a transient bottleneck in a large ancestral population would be reflected by low Hd and high Pi [26].

4.4. Differentiation and Genetic Distance

The estimation of differentiation and genetic distance was based on two indices: the genetic differentiation index (FST) [27] and genetic distance. Genetic distance within and between populations was computed using MEGA v11.0.13 [24]. Pairwise values between populations were estimated with Arlequin v3.5.2.2 [28].

4.5. Correlation Test between Genetic and Geographic Distance

The Mantel test [29] was applied to assess the relationship between genetic differentiation among populations and their geographic distribution. Analyses were conducted in R software version 3.4.3 [30] using the packages readxl [31], vegan [32], and ggplot2 [33]. The alternative hypothesis states that genetic and geographic distances are correlated. When the p-value is below the 0.05 significance threshold, the null hypothesis is rejected; otherwise, it is retained.

4.6. AMOVA Test

The analysis of molecular variance (AMOVA) implemented in Arlequin version 3.5.1.3 [28] was conducted to estimate the distribution of genetic diversity across three hierarchical levels within the populations. This approach partitions the total

genetic variance into components attributable to differences among individuals within populations and among populations.

4.7. Demogenetic Test

Demogenetic tests assess the selective neutrality and equilibrium of populations. Among these, Tajima's D and Fu's Fs are particularly sensitive to population expansion, often yielding strongly negative values. Analyses were performed using Arlequin version 3.5.1.3 [28].

Mismatch distribution analysis was conducted using DnaSp software under the null hypothesis of demographic expansion. A unimodal distribution indicates a population undergoing demographic expansion, whereas a multimodal distribution suggests demographic stability [34]. The latter is further evaluated using two indices: SSD (sum of squared deviations) and Raggedness (Rag irregularity index), both calculated with Arlequin version 3.5.1.3 [28].

4.8. Analysis of Phylogenetic Relationships

4.8.1. Haplotype Network

Haplotype networks employ the median-joining method to illustrate the relationships among different haplotypes. A minimum haplotype network consists of nodes (circles) and branches (links) connecting them. Each node represents a haplotype, with its size proportional to the haplotype's frequency in the dataset [16]. The haplotype network was inferred using the TCS method [35], as implemented in PopART software [36].

4.8.2. Phylogenetic Reconstructions

Phylogenetic trees were constructed using MEGA v11.0.13 software [24] to assess the phylogenetic relationships among individuals with two approaches: the Neighbor-Joining method (NJ), which relies on genetic distances, and the Maximum Likelihood (ML) method, which evaluates all possible evolutionary histories that could have generated the observed dataset. Branch support was assessed with 1000 bootstrap replicates. Bayesian inference was performed using MrBayes version 3.2.5 software [37], based on posterior probability calculations. This analysis was run for 1,000,000 generations with 4 Markov chains, sampling every 100 generations. The GTR model (six substitution types) without rate variation among sites was applied. Tree visualization was conducted using FigTree version 1.4.2 [38]. For all analyses, a COI sequence from *Ceratitis capitata* was used as the outgroup.

5. Results and Discussion

5.1. Results

5.1.1. Genetic Variability

Our dataset comprises a total of 352 nucleotide sequences, each consisting of 609 sites. Of these sites, 487 are monomorphic, and 119 are polymorphic, including 33 non-informative variable sites and 86 informative variable sites. The total num-

ber of mutations (Eta) is 129. The mean number of nucleotide differences (k) was estimated at 7.33. All results are summarized in Table 1.

Table 1. Basic parameters of genetic diversity.

Sample Size (N)	351
Number of Sites (N)	609
Invariable Sites (Monomorphic Sites)	487
Variable Sites (Polymorphic Sites)	119
Number of Singleton Variable Sites (Non-Informative Variable Sites)	33
Number of Parsimony Informative sites	86
Total Number of Mutations (Eta)	129
Average Number of Nucleotide Differences (k)	7.33
Number of Haplotypes (h)	133

The number of observed substitutions shows a higher percentage of transitions (89.3%) than transversions (10.58%), with a ratio of 8.44 (Table 2). Among the transitions, substitutions between purines are predominant, accounting for 46.3%, while transitions between pyrimidines represent 43% of all mutations. Although transitions are markedly dominant, transversions exhibit an asymmetric distribution. The substitution of guanine by thymine is clearly dominant, representing 17% of transversions. Other substitutions, such as adenine to cytosine (1.06%) and thymine to adenine or cytosine (1.60% each), were also detected. Reverse transversions remain rare.

Table 2. Shows the percentage of transitions and transversions.

	A	T	C	G
A	-	1.87	1.06	15
T	1.60	-	15.5	0.76
C	1.60	27.50	-	0.76
G	31.30	17	1.06	-

5.1.2. Genetic Diversity Indices

Nucleotide diversity (Pi) and haplotype diversity (Hd) were estimated for each population as well as for the overall dataset (Table 3). The number of haplotypes (h) observed across different geographic populations ranged from 1 to 17, with an average of 8.8 per population. The overall Hd value was 0.964 ± 0.005, while the overall Pi value was 0.012 ± 0.001. In general, the populations exhibited moderate (0.337) to high (1.000) haplotype diversity, with the exception of French Polynesia. By contrast, nucleotide diversity was relatively low, ranging from 0.000 ± 0.000 to 0.044 ± 0.003. Nearly half of the populations (Thailand, Laos, Cambodia, Bang-

ladesh, Malaysia, China, Sri Lanka, Vietnam, and Nepal) showed Hd values above 0.90 and Pi values between 0.006 and 0.044. Sri Lanka displayed the highest nucleotide diversity (Pi = 0.044) despite an Hd value of 0.932. In contrast, French Polynesia showed null values for both Hd and Pi.

Table 3. Number of haplotypes per population (H), haplotype diversity (Hd) (mean ± SD), and nucleotide diversity (Pi) (mean ± SD) of the studied populations.

Population	N	H	Hd	Pi	Population	N	H	Hd	Pi
Senegal	20	3	0.468 ± 0.005	0.005 ± 0.001	Laos	20	18	0.989 ± 0.019	0.009 ± 0.000
Benin	20	6	0.716 ± 0.087	0.005 ± 0.001	Cambodia	20	17	0.979 ± 0.024	0.007 ± 0.001
Burkina Faso	20	5	0.600 ± 0.101	0.013 ± 0.006	Nepal	9	8	1 ± 0.052	0.008 ± 0.001
Mali	16	2	0.542 ± 0.098	0.006 ± 0.001	Bangladesh	20	15	0.979 ± 0.021	0.008 ± 0.000
Kenya	20	2	0.337 ± 0.110	0.004 ± 0.001	Malaysia	20	15	0.968 ± 0.025	0.010 ± 0.000
Hawaii	20	3	0.542 ± 0.076	0.002 ± 0.001	China	20	14	0.963 ± 0.028	0.007 ± 0.000
Papua NG	20	3	0.468 ± 0.104	0.002 ± 0.000	Sri Lanka	20	12	0.932 ± 0.035	0.044 ± 0.003
French Polynesia	2	1	0.000 ± 0.000	0.000 ± 0.000	Vietnam	20	17	0.995 ± 0.018	0.007 ± 0.000
Thailand	20	14	0.968 ± 0.022	0.007 ± 0.000	Pop. Total	351		0.964 ± 0.005	0.012 ± 0.001
Philippines	20	6	0.768 ± 0.062	0.005 ± 0.001					
Japan	4	2	0.500 ± 0.265	0.001 ± 0.000					

5.1.3. Genetic Structure

1) Genetic Differentiation

According to the genetic differentiation (FST) values among the twenty populations studied (Table 4), pairwise FST values ranged from 0.00 to 0.77. The lowest genetic differentiation (FST) was observed between Bur/Sen, Chi, Hawaii/Pol, and Jap/Pol (0.000), whereas the highest significant genetic differentiation (FST = 0.77) occurred between Jap/Pap. Overall, a statistically significant differentiation among populations was detected ($p < 0.001$).

High FST values are represented in dark blue, whereas low FST values are shown in light blue.

2) Genetic Distance

The analysis of genetic structure across the five groups (Table 5) revealed low within-group genetic distances, with the lowest observed in Oceania (0.0031) and the highest within-group value recorded in Africa (0.0149). This analysis also indicated low between-group genetic distances. The lowest inter-group genetic distance was observed between America and Oceania (0.0072), whereas the highest was found between America and Asia (0.0134). Overall, these results suggest that the populations are relatively homogeneous, as evidenced by the consistently low within- and between-continent genetic distances.

Table 4. Heatmap illustrating pairwise FST values among populations.

Po	Ben	Bur	Cam	Chi	Haw	Jap	Ken	Lao	Mal	Mala	Nep	Pap	Phi	Pol	Sen	Sri	Tha	Vie
Ban	0.11	0.18	0.05	0.01	0.47	0.39	0.22	0.02	0.15	0.15	0.04	0.26	0.36	0.30	0.28	0.31	0.00	0.01
Ben		0.24	0.04	0.16	0.66	0.61	0.21	0.18	0.34	0.34	0.15	0.54	0.57	0.56	0.38	0.34	0.19	0.17
Bur			0.24	0.22	0.46	0.37	0.22	0.16	0.21	0.21	0.20	0.23	0.38	0.28	0.00	0.28	0.16	0.18
Cam				0.09	0.58	0.49	0.20	0.29	0.26	0.26	0.04	0.45	0.48	0.43	0.37	0.32	0.10	0.08
Chi					0.52	0.42	0.28	0.06	0.19	0.19	0.09	0.33	0.41	0.34	0.33	0.33	0.04	0.03
Haw						0.12	0.71	0.38	0.41	0.41	0.60	0.70	0.68	0.00	0.55	0.41	0.49	0.49
Jap							0.69	0.29	0.32	0.32	0.49	0.77	0.66	0.00	0.62	0.26	0.42	0.40
Ken								0.23	0.37	0.37	0.20	0.60	0.61	0.67	0.37	0.35	0.27	0.25
Lao									0.10	0.10	0.06	0.23	0.37	0.18	0.26	0.31	0.01	0.00
Mal										0.00	0.21	0.23	0.19	0.22	0.31	0.34	0.12	0.12
Mala											0.21	0.23	0.19	0.22	0.31	0.34	0.12	0.12
Nep												0.46	0.44	0.40	0.37	0.25	0.08	0.05
Pap													0.50	0.76	0.40	0.38	0.19	0.28
Phi														0.63	0.53	0.39	0.35	0.41
Pol															0.59	0.15	0.34	0.32
Sen																0.37	0.27	0.29
Sri																	0.32	0.33
Tha																		0.01

Table 5. Genetic distance intra- and inter-continental.

Genetic Distance						
Intra-Continental			**Inter-Continental**			
Group		**SD**	**Africa**	**Asia**	**Oceania**	**America**
Africa	0.0149	0.0040		0.0130	0.0083	0.0120
Asia	0.0084	0.0112	0.0130		0.0105	0.0134
Oceania	0.0031	0.0015	0.0083	0.0105		0.0072
America	0.0035		0.0120	0.0134	0.0072	

Analysis of genetic distances between populations revealed low divergence, with values ranging from 0.007 to 0.040 (Table 6). A marked genetic divergence was observed between the Sri Lankan population and those of the other sites, whereas the remaining populations exhibited substantial genetic proximity.

5.1.4. Correlation between Genetic and Geographic Distance

The correlation between the genetic and geographic distance matrices was assessed using the Mantel test (Figure 1). The results show that pairwise correlations between geographic and genetic distances were not significant (standardized Mantel statistic $r = -0.051$; $p = 0.423$). Thus, the null hypothesis of no correlation cannot be

rejected. These findings indicate that the two matrices do not display a statistically significant relationship, suggesting an absence of isolation by distance (IBD).

Table 6. Heat maps illustrate inter-population genetic distance values.

	Ben	Bur	Cam	Chi	Haw	Jap	Ken	Lao	Mala	Mal	Nep	Pap	Phi	Pol	Sen	Sri	Tha	Vie
Ban	0.008	0.013	0.008	0.008	0.011	0.011	0.008	0.009	0.011	0.009	0.008	0.007	0.010	0.010	0.010	0.037	0.008	0.008
Ben		0.013	0.007	0.008	0.012	0.012	0.006	0.009	0.012	0.007	0.007	0.008	0.012	0.011	0.009	0.036	0.008	0.008
Bur			0.014	0.014	0.015	0.016	0.011	0.014	0.015	0.011	0.013	0.010	0.014	0.015	0.009	0.039	0.013	0.013
Cam				0.008	0.012	0.012	0.007	0.009	0.012	0.008	0.007	0.009	0.011	0.011	0.010	0.036	0.008	0.008
Chi					0.011	0.011	0.008	0.009	0.011	0.009	0.008	0.008	0.011	0.010	0.010	0.038	0.008	0.008
Haw						0.003	0.011	0.010	0.011	0.011	0.011	0.008	0.011	0.002	0.011	0.038	0.010	0.009
Jap							0.011	0.010	0.011	0.012	0.011	0.009	0.012	0.001	0.012	0.039	0.010	0.009
Ken								0.009	0.011	0.005	0.006	0.007	0.011	0.010	0.007	0.036	0.008	0.007
Lao									0.011	0.009	0.008	0.007	0.011	0.009	0.010	0.038	0.008	0.008
Mala										0.011	0.011	0.008	0.009	0.010	0.011	0.040	0.010	0.010
Mal											0.007	0.007	0.011	0.011	0.007	0.036	0.008	0.008
Nep												0.007	0.010	0.010	0.009	0.036	0.008	0.007
Pap													0.006	0.008	0.006	0.036	0.006	0.006
Phi														0.011	0.010	0.038	0.009	0.010
Pol															0.011	0.038	0.009	0.008
Sen																0.038	0.009	0.009
Sri																	0.036	0.037
Tha																		0.007

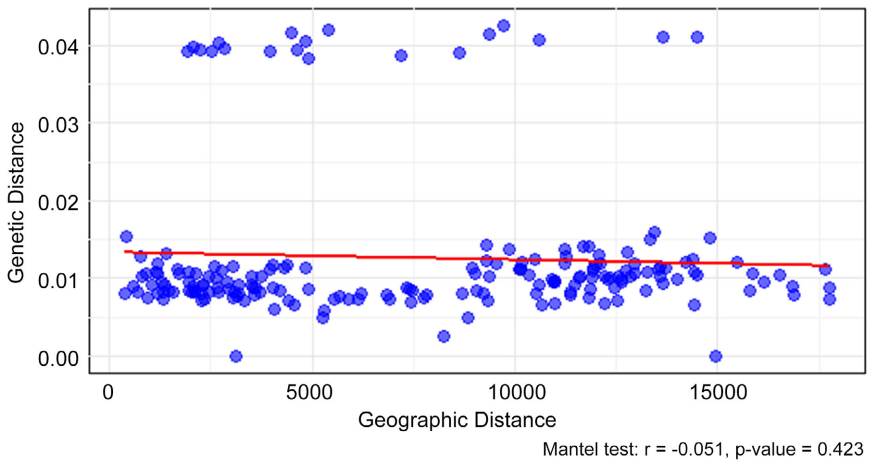

Figure 1. Mantel test. Each point is labeled with the corresponding subpopulation pair and its respective geographic and genetic distance values.

5.1.5. Analysis of Variance

The analysis of molecular variance (AMOVA) of *B. dorsalis* populations revealed

that only 5.56% of the genetic variation was partitioned among groups, whereas 25.96% was attributed to variation among populations within groups (Table 7). In contrast, the majority of genetic variation (68.48%) was found within populations. Moreover, AMOVA indicated a high and significant level of genetic differentiation among populations (FST = 0.315; $p < 0.001$).

Table 7. Analysis of molecular variance.

Source of Variation	Degrees of Freedom (DF)	Sum of Squares (SS)	Variance Components	Percentage of Variation	
Distribution of Molecular Variance across Biogeographic Regions					
Among Groups	3	112.546	0.22272 Va	5.56	FST = 0.315
Among Populations within Groups	16	332.667	1.04088 Vb	25.96	
Within Populations	335	919.672	2.74529 Vc	68.48	
Total	354	1364.885	4.00889		

5.1.6. Demographic Evolution

Mismatch distribution analyses revealed multimodal and irregular patterns for all population groups, indicative of demographic equilibrium or population stability, except for the Asian population group, which exhibited a unimodal distribution (Figure 2). Tajima's D test yielded mostly negative and non-significant values across groups, with the exception of the African populations, which showed significant Tajima's D values. In contrast, Fu's Fs test displayed opposite trends, with most populations showing non-significant positive deviation from neutrality, except for the Oceanian population, which exhibited a significant negative value. Similarly, demographic indices (SSD and Rag) were non-significant for most regions, with notable exceptions in African populations (SSD = 0.153; $p = 0.010$; Rag = 0.286, $p = 0.000$) and America (SSD = 0.500, $p = 0.000$) (Table 8).

Table 8. Past demographic statistics and neutrality tests.

Geographic Groups	Africa	America	Asia	Oceania	Total
Tajima's D					
Tajima's D	−1.734	−1.002	−1.745	−0.460	−1.235
p-Value	0.012	0.177	0.010	0.360	0.140
Fu's Fs					
Fu's Fs	4.527	2.450	−24.384	1.764	−3.910
p-Value	0.910	0.880	0.000	0.656	0.656
SSD Index					
SSD	0.153	0.500	0.002	0.108	0.200
p-Value	0.010	0.000	0.160	0.140	0.077
Rag Index					
Rag	0.286	0.620	0.010	0.288	0.301
p-Value	0.000	0.910	0.600	0.160	0.417

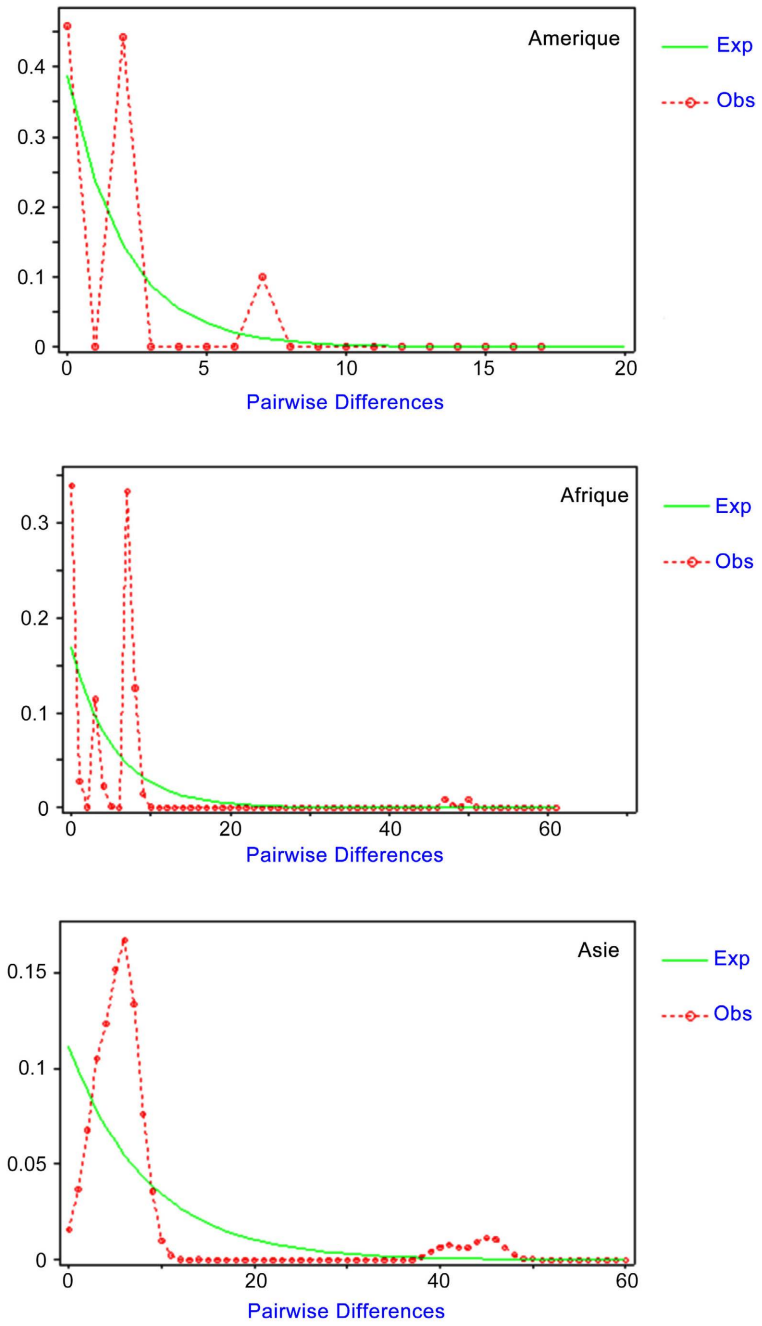

Figure 2. Observed (red line) and simulated (green dotted line) mismatch distribution analysis within populations of *B. dorsalis*.

5.1.7. Analysis of Phylogenetic Relationships

1) Haplotype Networks

Haplotype networks provide insights into the relationships among haplotypes. Figure 3 shows the haplotype network (TCS) constructed from mtDNA COI sequences. No clear geographic pattern was detected in the distribution of the observed haplotypes across distinct populations. The network revealed 133 haplotypes separated by 1 to 35 mutational steps. Of these, 41 haplotypes were shared

across different locations. The most widespread haplotype was Hap_17, detected in 40 individuals from 5 African populations. The second most frequent was Hap_16, found in 38 individuals from 5 African populations. The third most frequent haplotypes were Hap_6 and Hap_14, each occurring in 17 individuals across several populations.

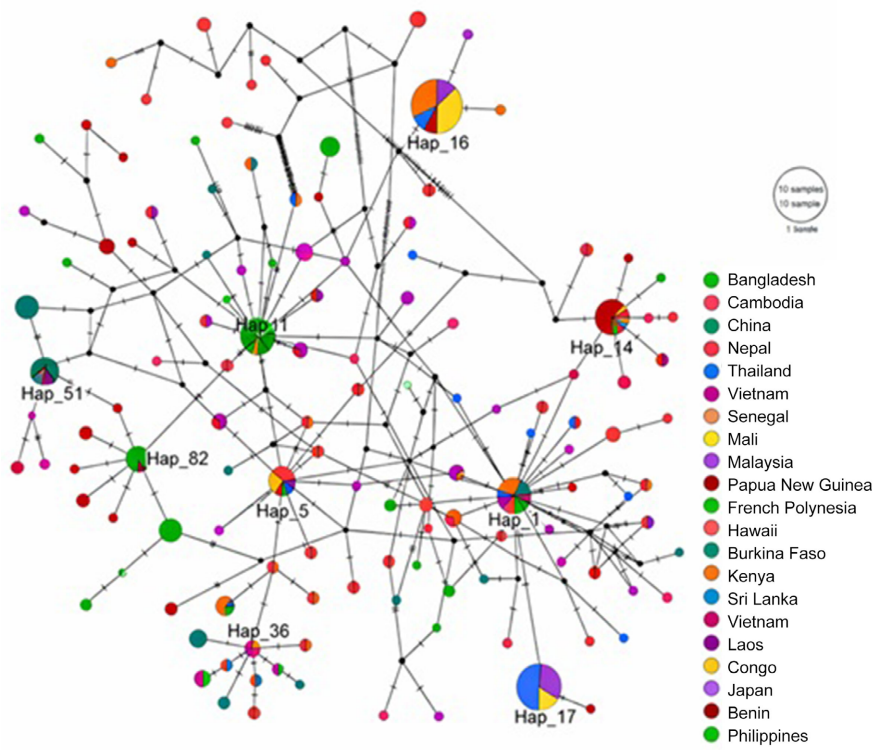

Figure 3. TCS network of *B. dorsalis* based on COI haplotypes. Circle size is proportional to the number of individuals carrying each haplotype.

2) Phylogenetic Reconstructions

Phylogenetic tree analyses showed strong statistical support for the cables, with generally high bootstrap values (above 50%) across all three methods. The haplotype phylogenetic tree constructed using Neighbor-Joining and Maximum Likelihood methods (**Figure 4**), revealed the formation of two clades. Clade C1 was supported by a bootstrap value of 100% in the Maximum Likelihood method. Clade C2 received strong support, with a bootstrap value of 100% in both NJ and ML analyses and included the outgroup *Ceratitis capitata* along with haplotypes from Asia and Africa. The tree topology obtained from the Maximum Likelihood method was largely consistent with that of the Neighbor-Joining method (**Figure 4**).

5.2. Discussion

The objective of this study is to genetically characterize *B. dorsalis* populations on a global scale. Analyses of genetic diversity revealed high haplotype diversity (Hd = 0.964 ± 0.005), coupled with low nucleotide diversity (Pi = 0.012 ± 0.001) within

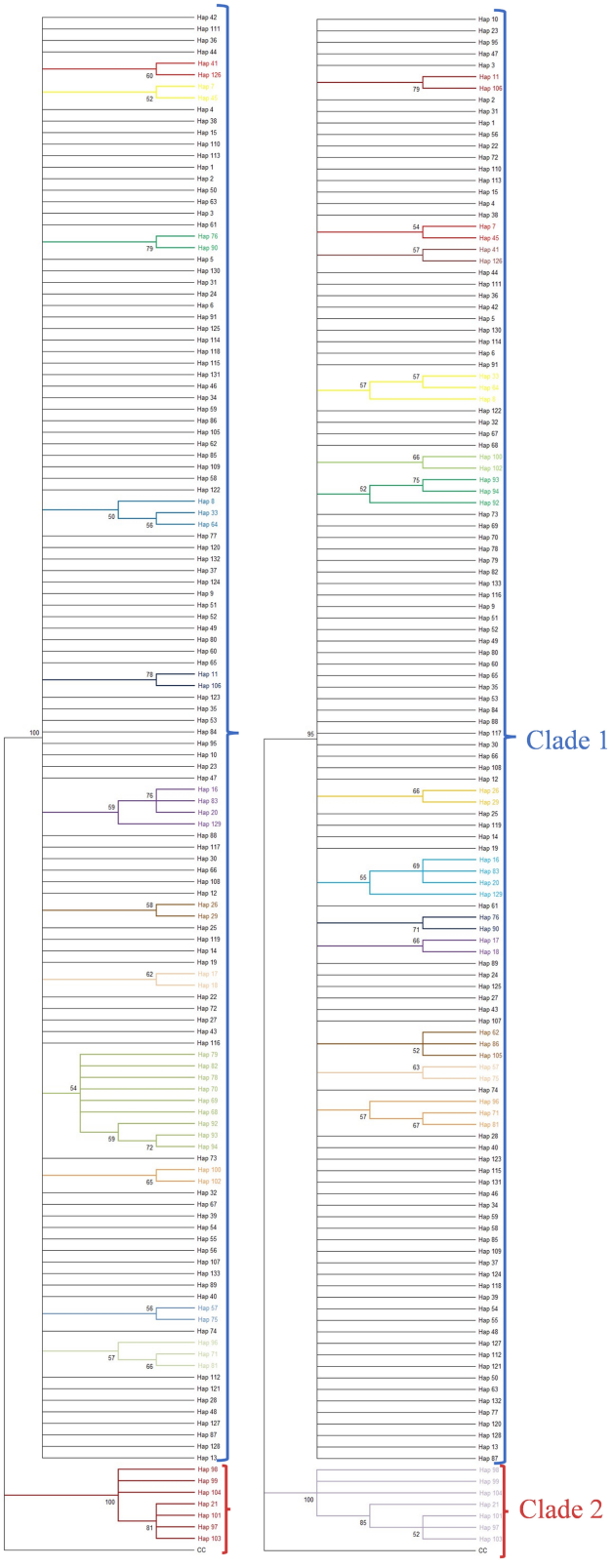

Figure 4. Haplotype phylogenetic trees reconstructed using the Neighbor-Joining method (right) and the Maximum Likelihood method (left). Subclades are counted from top to bottom in ascending order.

the overall population. Elevated levels of genetic diversity indicate strong viability and adaptability, potentially enhancing fitness under changing environmental conditions. This pattern was observed across all Asian populations, with the exception of those from Japan. These findings support the hypothesis of an Asian origin of *B. dorsalis*, more specifically in Southeast Asia, a hypothesis initially proposed by [12] and later confirmed by [39], with genetic diversification likely promoted by a broad range of tropical hosts and favorable ecological conditions.

The low genetic diversity observed in Japan may be the result of a bottleneck. Japan has maintained the eradication status of the oriental fruit fly since 1986, although occasional captures still occur within restricted areas and over short periods on the small islands of southwestern Japan [40]. This has led to a substantial reduction in population size, resulting in decreased genetic diversity. Hawaii, Africa, and Oceania also exhibit reduced genetic diversity, characterized by a limited number of haplotypes. These findings are consistent with those of [41], who reported that African and Hawaiian populations both display markedly lower molecular diversity than Asia as a whole. Such a reduction in diversity may be attributed to a severe founder effect, whereby only a small number of individuals from the source region established populations in these areas, with initially limited expansion following introduction, although some African populations may currently be undergoing early-stage expansion. Analysis of genetic differentiation (FST) and genetic distance among *B. dorsalis* populations revealed weak overall genetic structuring across the species distribution. In Asia, the genetic structure of *B. dorsalis* is weakly supported [9] [39], although it reveals several regional patterns, particularly in Japan and the Philippines. This contrast suggests that Asian populations of *B. dorsalis* are genetically weakly structured due to extensive gene flow. Nevertheless, our findings also highlight more pronounced regional structuring patterns in Japan and the Philippines, implying that certain geographical, ecological, or behavioral barriers may restrict gene flow in these areas.

The genetic proximity between Japan, French Polynesia, and Hawaii, despite their wide geographical separation, calls for the exploration of specific hypotheses. This proximity suggests the existence of shared invasion routes or common sources. One hypothesis involves large-scale natural dispersal, as documented between the Philippines, China, and the southern islands of Japan, where typhoons, high-pressure ridges, and tropical depressions have been identified as potential vectors of species dispersal [40]. Moreover, previous studies have shown that Hawaiian populations of *Bactrocera dorsalis* likely originated from East Asia, particularly coastal areas of China, which may explain their genetic affinity with Japanese populations [9]. The colonization of French Polynesia, in turn, appears to have been initiated by individuals originating from Southeast Asia or Hawaii. Beyond genetic proximity, haplotype analysis reveals a shared haplotype (Hap_51) among Japan, French Polynesia, and Hawaii, further supporting a common origin. The low intracontinental genetic distances observed for Asia (0.0084), Oceania (0.0031), and Hawaii (0.0035) confirm regional genetic homogeneity, likely maintained by continuous exchanges

within the regions (e.g., fruit trade, human migrations). The Mantel test (r = −0.051, p = 0.423) showed no significant correlation between geographical and genetic distances, thereby rejecting the isolation by distance (IBD) hypothesis, which predicts that genetic differentiation increases with geographic distance due to restricted gene flow. Similar findings have been reported for *B. dorsalis* [39] and other highly mobile species such as *Tribolium castaneum* [42], *Bactrocera zonata* [43], and *Chlorops oryzae* [44]. This pattern can be explained by multiple factors, including anthropogenic dispersal and repeated introductions of genetically similar populations in different regions. Analysis of molecular variance (AMOVA) indicated that most of the genetic variation was partitioned within populations (68.48%), rather than among populations within the same geographic group (25.96%) or among geographical groups (5.56%). This distribution suggests that genetic diversity is predominantly intra-populational, which is typical of species with high dispersal capacity and a panmictic regional structure, as previously observed in *B. dorsalis* [9] [17] [38] [45]-[47], in *Ceratitis capitata*, and *Chlorops oryzae* [43] [48].

Demogenetic tests (Tajima's D, Fu's Fs) and mismatch distribution analysis revealed contrasting demographic dynamics across regions. In Asia, the significantly negative values observed (D = −1.745; p = 0.010; Fs = −24.384; p = 0.000) indicate an excess of rare mutations, a typical signature of recent demographic expansion. This hypothesis is supported by mismatch distribution results, including a non-significant SSD, high haplotype diversity, and low nucleotide diversity, all consistent with the expansion model. These findings align with previous studies [9] [44], which also reported population expansion in this region. In Africa, Tajima's D was likewise significantly negative (−1.734, p = 0.012), suggesting demographic expansion, but this signal was not corroborated by other tests. Fu's Fs was detected and was non-significant (Fs = 4.527, p = 0.910), and mismatch distribution analysis indicated a poor fit to the expansion model (SSD = 0.153, p = 0.010; Rag = 0.286, p = 0.000). No clear demographic signal of expansion was detected in the Americas and Oceania. Tajima's D values (D = −1.002; p = 0.177 for the Americas; D = −0.460; p = 0.360 for Oceania) and Fu's Fs values (Fs = 2.450; p = 0.880 for America; Fs = 1.764; p = 0.656 for Oceania) suggest demographic equilibrium or only moderate expansion. The mismatch distribution curves were multimodal, a pattern generally consistent with stable populations. However, in the Americas, the SSD was significant (SSD = 0.500; p = 0.000), indicating a poor fit to the expansion model and reinforcing the hypothesis of demographic stability or complexity. Conversely, Oceania exhibited a non-significant SSD (SSD = 0.108; p = 0.140), suggesting a moderate, though statistically inconclusive, fit to the expansion model. These results indicate that populations in these two regions likely stem from recent introductions but have not yet accumulated the genetic signature of rapid demographic growth, unlike the patterns observed in Asia and Africa. Phylogenetic trees reconstructed using Bayesian inference, Neighbor-Joining, and maximum likelihood further support the existence of distinct genetic clades, some homogeneous (primarily Asian), and others heterogeneous, encom-

passing populations from all four regions. The presence of mixed clades confirms that several populations from different continents share common haplotypes, suggesting introduction events from shared genetic sources. For instance, clade C2 (NJ and ML) includes the outgroup (*Ceratitis capitata*) together with haplotypes from Asia (Sri Lanka) and Africa (Burkina Faso), suggesting that these haplotypes represent conserved ancestral lineages, originating from early divergences within the group, and reflecting historical migrations between these regions. Indeed, the African invasion likely originates from South Asia [41], corroborating earlier findings [40] [41]. The reconstructed trees also displayed high bootstrap values, underscoring the robustness of the identified clades. Nonetheless, topological inconsistencies among methods (Bayesian vs. NJ vs. ML) highlight the limitations of relying on a single mitochondrial marker, underscoring the need to incorporate nuclear markers (e.g., microsatellites, SNPs) to refine demographic and phylogenetic inferences.

In African populations, the significant negative Tajima's D suggests an excess of rare haplotypes that may reflect recent demographic expansion or purifying selection. However, the multimodal mismatch distribution and non-supportive Fu's Fs indicate that this expansion signal is weak or recent. Together, these results suggest a complex demographic history combining founder effects followed by limited or incipient expansion.

6. Conclusion

This study provides a global overview of the genetic diversity and population structure of *Bactrocera dorsalis* based on mitochondrial COI sequences. The species exhibits high haplotype diversity but low nucleotide diversity, with most variation occurring within populations. Phylogenetic and demographic analyses support an Asian origin followed by multiple introductions into Africa, Oceania, and America, likely accompanied by founder effects and varying degrees of demographic expansion. These findings improve our understanding of the invasion dynamics of *B. dorsalis* and may contribute to the development of more targeted and effective pest management strategies. Future studies integrating nuclear markers are recommended.

Acknowledgements

We thank the Genetics and Population Management team for their daily support.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Alexandratos, N. and Bruinsma, J. (2012) World Agriculture towards 2030/2050: The 2012 Revision (ESA Working Paper No. 12-03). FAO. https://www.fao.org/3/ap106e/ap106e.pdf

[2] FAO (2021) Fruits et légumes-Éléments essentiels de ton alimentation. FAO. https://doi.org/10.4060/cb2395fr

[3] Fernandes, A.S., Ferreira-Pêgo, C. and Costa, J.G. (2023) Functional Foods for Health: The Antioxidant and Anti-Inflammatory Role of Fruits, Vegetables and Culinary Herbs. *Foods*, **12**, Article No. 2742. https://doi.org/10.3390/foods12142742

[4] Cruz, M.A.A.S., Coimbra, P.P.S., Araújo-Lima, C.F., Freitas-Silva, O. and Teodoro, A.J. (2024) Hybrid Fruits for Improving Health—A Comprehensive Review. *Foods*, **13**, Article No. 219. https://doi.org/10.3390/foods13020219

[5] IPPC (2021) Scientific Review of the Impact of Climate Change on Plant Pests. FAO.

[6] Horrocks, K.J., Romeis, J. and Collatz, J. (2025) Environmental Impacts of Agricultural Pest Insects: Five Case Studies Reveal Overlooked Impact Mechanisms and Specify Knowledge Gaps. *NeoBiota*, **104**, 251-279. https://doi.org/10.3897/neobiota.104.158217

[7] Drew, R.A.I. and Hancock, D.L. (1999) Phylogeny of the Tribe Dacini (Dacinae) Based on Morphological, Distributional, and Biological Data. *Systematic Entomology*, **24**, 231-252.

[8] Schutze, M.K., Aketarawong, N., Amornsak, W., Armstrong, K.F., Augustinos, A.A., Barr, N., *et al.* (2015) The Oriental Fruit Fly, *Bactrocera dorsalis* (Hendel), Revisited: Species Limits and Systematics. *Zoologica Scripta*, **44**, 477-490.

[9] Wan, X., Liu, Y. and Zhang, B. (2012) Invasion History of the Oriental Fruit Fly, *Bactrocera dorsalis*, in the Pacific-Asia Region: Two Main Invasion Routes. *PLOS ONE*, **7**, e36176. https://doi.org/10.1371/journal.pone.0036176

[10] Clarke, A.R. (2019) Biology and Management of Bactrocera and Related Fruit Flies. CABI.

[11] Duyck, P., David, P. and Quilici, S. (2004) A Review of Relationships between Interspecific Competition and Invasions in Fruit Flies (Diptera: Tephritidae). *Ecological Entomology*, **29**, 511-520. https://doi.org/10.1111/j.0307-6946.2004.00638.x

[12] Drew, R.A.I., Tsuruta, K. and White, I.M. (2005) A New Species of Pest Fruit Fly (Diptera: Tephritidae: Dacinae) from Sri Lanka and Africa. *African Entomology*, **13**, 149-154.

[13] Hartl, D.L. and Clark, A.G. (2007) Principles of Population Genetics. 4th Edition, Sinauer Associates.

[14] Cameron, S.L. (2014) How to Sequence and Annotate Insect Mitochondrial Genomes for Systematic and Comparative Genomics Research. *Systematic Entomology*, **39**, 400-411. https://doi.org/10.1111/syen.12071

[15] Rodrigues, M.S., Morelli, K.A. and Jansen, A.M. (2017) Cytochrome C Oxidase Subunit 1 Gene as a DNA Barcode for Discriminating Trypanosoma Cruzi DTUs and Closely Related Species. *Parasites & Vectors*, **10**, Article No. 488. https://doi.org/10.1186/s13071-017-2457-1

[16] Ndiaye, M.R. (2018) Évolution phylogéographique et structure génétique des populations de *Sitophilus zeamais*. Thèse de Doctorat, Université Cheikh Anta Diop de Dakar, 151 p. https://hal.science/tel-02116356

[17] Kim, S., Park, J.S., Kim, M.J., Kim, K., Kim, S. and Kim, I. (2021) Complete Mitochondrial Genome of the Highly Fecund *Bombyx mori* Linnaeus, 1758 (Lepidoptera: Bombycidae) Strain Jam 146. *Mitochondrial DNA Part B: Resources*, **6**, 2278-2280. https://doi.org/10.1080/23802359.2021.1920860

[18] Lin, X., Stur, E. and Ekrem, T. (2015) Exploring Genetic Divergence in a Species-Rich Insect Genus Using 2790 DNA Barcodes. *PLOS ONE*, **10**, e0138993.

https://doi.org/10.1371/journal.pone.0138993

[19] Allio, R., Donega, S., Galtier, N. and Nabholz, B. (2017) Large Variation in the Ratio of Mitochondrial to Nuclear Mutation Rate across Animals: Implications for Genetic Diversity and the Use of Mitochondrial DNA as a Molecular Marker. *Molecular Biology and Evolution*, **34**, 2762-2772. https://doi.org/10.1093/molbev/msx197

[20] Katsonis, P. and Lichtarge, O. (2014) A Formal Perturbation Equation between Genotype and Phenotype Determines the Evolutionary Action of Protein-Coding Variations on Fitness. *Genome Research*, **24**, 2050-2058. https://doi.org/10.1101/gr.176214.114

[21] Katoh, K. and Standley, D.M. (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, **30**, 772-780. https://doi.org/10.1093/molbev/mst010

[22] Hall, T.A. (1999) BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95-98

[23] Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E., *et al.* (2017) DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Molecular Biology and Evolution*, **34**, 3299-3302. https://doi.org/10.1093/molbev/msx248

[24] Tamura, K., Stecher, G. and Kumar, S. (2021) MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, **38**, 3022-3027. https://doi.org/10.1093/molbev/msab120

[25] Nei, M. (1996) Phylogenetic Analysis in Molecular Evolutionary Genetics. *Annual Review of Genetics*, **30**, 371-403. https://doi.org/10.1146/annurev.genet.30.1.371

[26] Nei, M. (1977) *f*-Statistics and Analysis of Gene Diversity in Subdivided Populations. *Annals of Human Genetics*, **41**, 225-233. https://doi.org/10.1111/j.1469-1809.1977.tb01918.x

[27] Weir, B.S. and Hill, W.G. (2002) Estimating F-Statistics. *Annual Review of Genetics*, **36**, 721-750. https://doi.org/10.1146/annurev.genet.36.050802.093940

[28] Excoffier, L. and Lischer, H.E.L. (2010) Arlequin Suite Ver 3.5: A New Series of Programs to Perform Population Genetics Analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564-567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

[29] Mantel, N. (1967) The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer Research*, **27**, 209-220.

[30] R Core Team (2017) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing.

[31] Wickham, H. and Bryan, J. (2023) Readxl: Read Excel Files. R Package Version 1.4.3.

[32] Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., *et al.* (2025) Vegan: Community Ecology Package (Version 2.7-1). CRAN.

[33] Wickham, H. (2016) Ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag. https://ggplot2.tidyverse.org

[34] Harpending, H.C., Sherry, S.T., Rogers, A.R. and Stoneking, M. (1993) The Genetic Structure of Ancient Human Populations. *Current Anthropology*, **34**, 483-496. https://doi.org/10.1086/204195

[35] Clement, M., Posada, D. and Crandall, K.A. (2000) TCS: A Computer Program to Estimate Gene Genealogies. *Molecular Ecology*, **9**, 1657-1659. https://doi.org/10.1046/j.1365-294x.2000.01020.x

[36] Leigh, J.W. and Bryant, D. (2015) Popart: Full-Feature Software for Haplotype Network Construction. *Methods in Ecology and Evolution*, **6**, 1110-1116. https://doi.org/10.1111/2041-210x.12410

[37] Huelsenbeck, J.P. and Ronquist, F. (2001) MrBayes: Bayesian Inference of Phylogenetic Trees. *Bioinformatics*, **17**, 754-755. https://doi.org/10.1093/bioinformatics/17.8.754

[38] Rambaut, A. and Drummond, A.J. (2012) FigTree, a Graphical Viewer of Phylogenetic Trees (v1.4.2).

[39] Shi, W., Kerdelhué, C. and Ye, H. (2012) Genetic Structure and Inferences on Potential Source Areas for *Bactrocera dorsalis* (Hendel) Based on Mitochondrial and Microsatellite Markers. *PLOS ONE*, **7**, e37083. https://doi.org/10.1371/journal.pone.0037083

[40] Otuka, A., Hu, Z. and Lu, Z. (2016) Migration Sources and Pathways of the Pest Species *Sogatella furcifera* (Horváth) in Topologically Complex Regions like Yunnan, China, and Adjacent Montane Areas. *PLOS ONE*, **11**, e0156693.

[41] Qin, Y., Krosch, M.N., Schutze, M.K., Zhang, Y., Wang, X., Prabhakar, C.S., *et al.* (2018) Population Structure of a Global Agricultural Invasive Pest, *Bactrocera dorsalis* (Diptera: Tephritidae). *Evolutionary Applications*, **11**, 1990-2003. https://doi.org/10.1111/eva.12701

[42] Diome, T. (2014) Biodémographie et diversité génétique des populations de *Tribolium castaneum* Herbst (Coleoptera, Tenebrionidae) ravageur des grains de mil au Sénégal. Thèse de doctorat, Université Cheikh Anta Diop de Dakar, 94 p.

[43] Choudhary, J.S., Naaz, N., Lemtur, M., Das, B., Singh, A.K., Bhatt, B.P., *et al.* (2017) Genetic Analysis of *Bactrocera zonata* (Diptera: Tephritidae) Populations from India Based on *cox*1 and *nad*1 Gene Sequences. *Mitochondrial DNA Part A*, **29**, 727-736. https://doi.org/10.1080/24701394.2017.1350952

[44] Zhou, Y., Li, X., Wu, J. and Zhang, X. (2020) Genetic Differentiation in the Rice Pest Chlorops oryzae. *Journal of Applied Entomology*, **144**, 165-175.

[45] Aketarawong, N., Bonizzoni, M., Thanaphum, S., Gomulski, L.M., Gasperi, G., Malacrida, A.R., *et al.* (2007) Inferences on the Population Structure and Colonization Process of the Invasive Oriental Fruit Fly, *Bactrocera dorsalis* (Hendel). *Molecular Ecology*, **16**, 3522-3532. https://doi.org/10.1111/j.1365-294x.2007.03409.x

[46] Keita, N.D. (2018) Caractérisation morphométrique génétique de Bactrocera dorsalis et traitements biocides à base d'*Allium sativum* et *Gymnanthemum amygdalinum*. Thèse de Doctorat, Université Cheikh Anta Diop de Dakar, 151 p.

[47] Afroz, S., Noman, M.S., Zhang, Y., Qin, Y., Chowdhury, S.M.K.H. and Li, Z. (2021) Population Genetic Structure of *Bactrocera dorsalis* Based on Cox1 Sequences from Bangladesh and Neighboring Countries. *Journal of Asia-Pacific Entomology*, **24**, 182-190. https://doi.org/10.1016/j.aspen.2021.02.011

[48] Deschepper, P., Todd, T.N., Virgilio, M., De Meyer, M., Barr, N.B. and Ruiz-Arce, R. (2021) Looking at the Big Picture: Worldwide Population Structure and Range Expansion of the Cosmopolitan Pest *Ceratitis capitata* (Diptera, Tephritidae). *Biological Invasions*, **23**, 3529-3543. https://doi.org/10.1007/s10530-021-02595-4