

Article

Genomic Traces of the Fruit Fly *Anastrepha obliqua* Associated with Its Polyphagous Nature

Elkin Aguirre-Ramirez ^{*}, Sandra Velasco-Cuervo and Nelson Toro-Perea

Grupo de Estudios Ecogenéticos y Biología Molecular, Departamento de Biología, Universidad del Valle, Cali 760032, Colombia; sandra.velasco@correounivalle.edu.co (S.V.-C.); nelson.toro@correounivalle.edu.co (N.T.-P.)

* Correspondence: elkin.aguirre@correounivalle.edu.co

Simple Summary: Individuals of a polyphagous species, such as *Anastrepha obliqua*, that plague different host plants, present differences at the genome level; these differences are associated with adaptive processes related to the exploitation of the resource and can lead to speciation in sympatry, first with the formation of host races. In this study, we used pooled RAD-seq to assess genomic differentiation and population structure in sympatric populations of *Anastrepha obliqua* that infest three different host plants. The results obtained support the effect of host plants on genomic differentiation in populations of the species *A. obliqua*. We identified a small group of candidate genes that could be under divergent selection, a product of the selective pressure mediated by the host plants.

Abstract: *Anastrepha obliqua* (Macquart) (Diptera: Tephritidae) is an important pest in the neotropical region. It is considered a polyphagous insect, meaning it infests plants of different taxonomic families and readily colonizes new host plants. The change to new hosts can lead to diversification and the formation of host races. Previous studies investigating the effect of host plants on population structure and selection in *Anastrepha obliqua* have focused on the use of data from the mitochondrial DNA sequence and microsatellite markers of nuclear DNA, and there are no analyses at the genomic level. To better understand this issue, we used a pooled restriction site-associated DNA sequencing (pooled RAD-seq) approach to assess genomic differentiation and population structure across sympatric populations of *Anastrepha obliqua* that infest three host plants—*Spondias purpurea* (red mombin), *Mangifera indica* (mango) of the family Anacardiaceae and *Averrhoa carambola* (carambola) of the family Oxalidaceae—in sympatric populations of the species *Anastrepha obliqua* of Inter-Andean Valley of the Cauca River in southwestern Colombia. Our results show genomic differentiation of populations from carambola compared to mango and red mombin populations, but the genetic structure was mainly established by geography rather than by the host plant. On the other hand, we identified 54 SNPs in 23 sequences significantly associated with the use of the host plant. Of these 23 sequences, we identified 17 candidate genes and nine protein families, of which four protein families are involved in the nutrition of these flies. Future studies should investigate the adaptive processes undergone by phytophagous insects in the Neotropics, using fruit flies as a model and state-of-the-art molecular tools.

Keywords: adaptation; population differentiation; host plant races; RAD-seq; Diptera; Tephritidae



Citation: Aguirre-Ramirez, E.; Velasco-Cuervo, S.; Toro-Perea, N. Genomic Traces of the Fruit Fly *Anastrepha obliqua* Associated with Its Polyphagous Nature. *Insects* **2021**, *12*, 1116. <https://doi.org/10.3390/insects12121116>

Received: 6 August 2021

Accepted: 27 November 2021

Published: 14 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A fundamental challenge in contemporary evolutionary biology is to understand the genetic mechanisms that determine how organisms can adapt to different environments [1]. Although there is evidence of these adaptive processes in wild conditions [2–4], the total understanding of genetic foundations in almost any system remains a challenge. In this sense, phytophagous insects have been characterized as an excellent biological model for understanding the processes involved in adaptation to new ecological environments [5–10].

This is mainly due to the close relationship between phytophagous insects and their host plants, which constitute their habitat during all or some stages of their life cycle, their oviposition site, or their food source, a relationship that leads to the selection of adaptations associated with the phenology of the plant [11]. Generally, in this ecological relationship, an optimal use of the available food resource is observed, but many times, the insect can begin to exploit other food resources around it. The opportunity to use new species of host plants can contribute to either the expansion of the distribution range of a species or lead to a process of ecological diversification with the formation of host races [12]. A host race is defined as a population of a species that is partially reproductively isolated from other conspecific populations as a direct consequence of adaptation to a specific host [13]. Conceptually, host races and species reside on different ends of a gene flow continuum, with host races representing the hypothesized incipient stage of ecological speciation and host-associated species its final product [5,14].

An interesting model of phytophagous insects with which to study the effect of host shifts has been fruit flies (Diptera: Tephritidae). This family of flies is currently composed of approximately 4500 species [15], many of which are pests of economic importance worldwide. In addition to their status as pests, tephritids include several species that have been studied by evolutionary biologists to understand the rapid divergence of species in different scenarios, such as ecological speciation in sympatry [5,9,16,17]. Among the genera of the family Tephritidae, *Anastrepha* Schiner is the most diverse and economically important in the Neotropics, with more than 300 species [18]. Few studies assess the effect of host shifts in the genus *Anastrepha* (e.g., [19–26]).

In this study, the study organism was the species *Anastrepha obliqua* (Macquart), the West Indian fruit fly. This species is distributed from northern Mexico to southern Brazil and the Caribbean, and is occasionally found in Texas and California. In addition, it is part of the taxonomic group fraterculus and is considered an important pest of 60 species in 24 plant families of tropical and subtropical areas [26,27]. In the southwest of Colombia, *A. obliqua* has been found to be infesting fruits of mango (*Mangifera indica* L., Anacardiaceae), red mombin (*Spondias purpurea* L., Anacardiaceae), carambola (*Averrhoa carambola* L., Oxalidaceae), and araza (*Eugenia stipitata* McVaugh, Myrtaceae) [24]. Additionally, *A. obliqua* has been reported to infest other fruits, both of commercial and wild interest [28]. This apparently generalist behavior suggests a process of progressive adaptation to new host plants.

Although few studies have been conducted to understand the association of *A. obliqua* with its host plants, some molecular studies have provided relevant information on this topic. For example, Aguirre-Ramirez et al. [24] found differences in the distribution of haplotype frequencies between populations infesting different host plants. Ruiz-Arce et al. [26] analyzed *A. obliqua* populations in the state of Veracruz, Mexico, from five host plant species and 52 geographic collections using data from mitochondrial DNA sequences and microsatellite nuclear DNA markers. They found significant differences between samples collected from different host plants in one locality (Apazapan), which could imply a certain degree of fidelity for the host plant (*S. purpurea* vs. *M. indica* var. Criollo) that restricts the ongoing gene flow between populations.

Traditionally, studies that seek to understand the basis of an adaptive process in nature use common garden experiments, where individuals from contrasting environments are exposed to a common laboratory environment [29] and statistical methods are used to infer the genetic differentiation underlying phenotypic traits [2]. These methods have been increasingly integrated with population-genetic tools to infer adaptation [30]. More recently, advances in genomic technologies have made it possible to identify genomic regions associated with adaptive process, for example, [9,30–37]. Among such genomic approaches are genome scan or outlier detection methods, which allow inferences about adaptive differentiation [38]. In this research, we used the sequencing of DNA fragments associated with restriction sites (RAD-seq) [39], which allows the identification of single-nucleotide polymorphisms (SNPs) throughout the genome that are involved, for example,

in the use of food resources [9]. The advantage of this technique lies in its ability to evaluate multiple SNPs from across the genome and identify regions under divergent selection.

Although members of the family Tephritidae have been used as models to understand the effect of host shift on population dynamics and the evolutionary history of populations [33,40], in the genus *Anastrepha* the genetic-molecular mechanisms are mostly unknown. To investigate this issue, we sampled sympatric populations of *A. obliqua* infesting three host plants in the Inter-Andean Valley of the Cauca River in southwestern Colombia and measured the differentiation between populations through a subset of RAD-seq data. Finally, we used three approaches to identify those significantly differentiated loci that are correlated with the use of host plants.

2. Materials and Methods

2.1. Sample Collection

Fruits were collected from three host plants species infested with *A. obliqua* larvae: mango (*M. indica*) and red mombin (*S. purpurea*) of the family Anacardiaceae, and carambola (*A. carambola*) of the family Oxalidaceae. These three fruits were sampled because:

- (i) the red mombin is a native species of America and natural host of *A. obliqua*;
- (ii) the mango, a species native to Southeast Asia introduced into Colombia in the 16th to 17th centuries [41], belonging to the same family as red mombin, and a crop of economic importance that in Colombia *A. obliqua* is its main pest;
- (iii) the carambola an introduced species around the 18th century and brought to Colombia in 1930 [42], belonging to a taxonomic family different from mango and red mombin, which could have a greater impact on the physiology of individuals, due to the composition being different from the fruits.

These plants were all found in localities of southwestern Colombia located in the north and south of the Inter-Andean Valley of the Cauca River. The reason why this region was chosen is because Aguirre-Ramirez et al. [24] reported that there was no population genetic structure of *A. obliqua* associated with geographic distance in the Inter-Andean Valley of the Cauca River. To choose the populations of fruit flies in the study, the estimated dispersal capacity of wild flies of *A. obliqua* was also taken into account, which has been reported to be approximately 151 m in radius [43].

Based on the above criteria, the north (localities Bolívar and La Unión) and south (localities Chagres, Robles and Guachinte) of the Inter-Andean Valley of the Cauca River were selected as the sampling sites (Figure 1). In each locality, fruits were collected from the three host plants, so each sample consisted of individuals of *A. obliqua* obtained from each host plant for each locality, for a total of six populations (Table 1). Obtaining adult individuals was achieved through the method of Carrejo and González [28]. Fruits were taken that showed signs of infestation, such as soft tissue and/or oviposition holes. They were deposited into a plastic container with vermiculite and covered with very fine porosity (white muslin) cloth and maintained at room temperature until the emergence of adult individuals. Each individual was morphologically identified with the taxonomic key of Caraballo [44]. The identified individuals were preserved in 96% ethanol at $-20\text{ }^{\circ}\text{C}$ until DNA extraction.

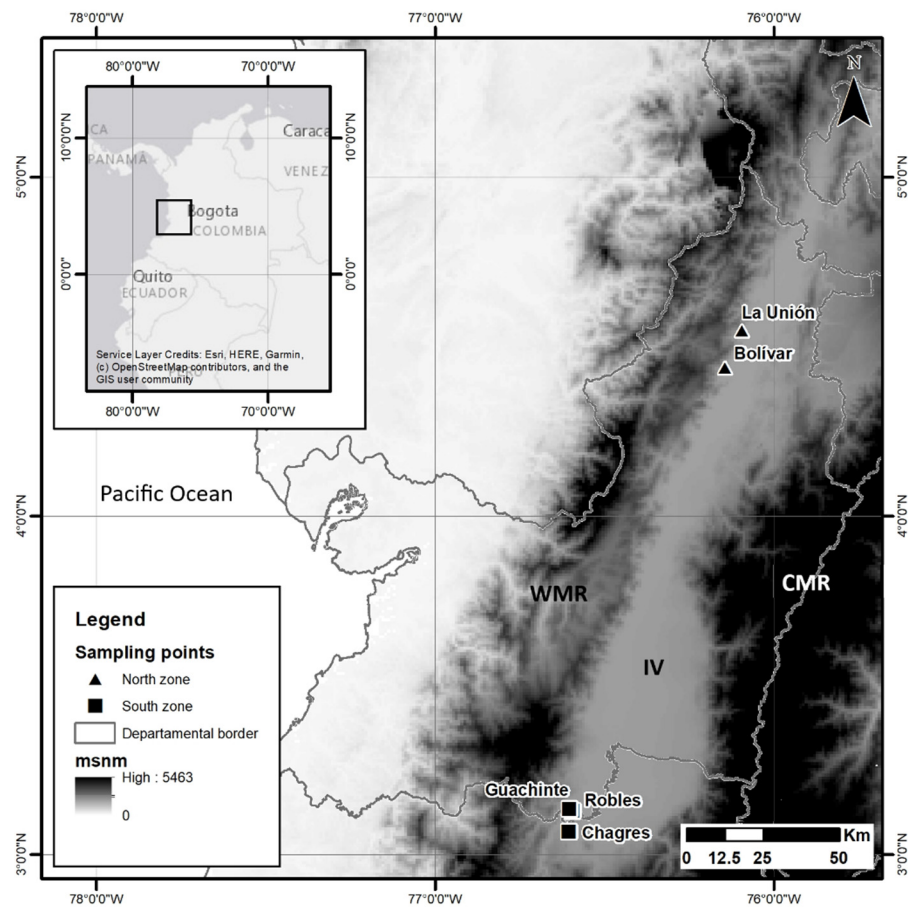


Figure 1. Geographic location of the localities where the fruits infested by the species *Anastrepha obliqua* were collected in southwestern Colombia (the north: localities Bolívar and La Unión and south: localities Chagres, Robles and Guachinte). WMR, Western Mountain Range; IV, Inter-Andean Valley of the Cauca River; CMR, Central Mountain Range.

Table 1. Sampling sites and information of the pooled DNA samples of populations of the species *Anastrepha obliqua*.

Host Plant	Locality	Code	Number of Individuals
<i>Averrhoa carambola</i>	North	CaN	26
	South	CaS	40
<i>Mangifera indica</i>	North	MN	40
	South	MS	40
<i>Spondias purpurea</i>	North	RmN	40
	South	RmS	40

2.2. DNA Extraction, RAD-Seq Libraries Construction and Sequencing

DNA was extracted from the head and legs of adult individuals of *A. obliqua* using the DNeasy Blood and Tissue kit (Qiagen, Valencia, CA, USA). DNA quality was evaluated in a 1% agarose gel. Only nondegraded DNA samples were processed and quantified in a NanoDrop ND-1000 spectrophotometer. Each sample was diluted to 20 ng/ μ L, re-quantified, and pooled by population (Table 1). Each pooled sample was quantified with a Qubit 4 fluorometer and was equalized to 20 ng/ μ L for the RAD library construction. RAD libraries and sequencing were carried out by BGI Tech Solutions, Co., Ltd. (Hong Kong, China). Briefly, DNA was fragmented by the restriction enzyme EcoRI. A P1 adapter, including a forward amplification primer, a sequencing primer and a specific barcode, was added to each of the digested DNA pools. The barcoded samples were then pooled and sheared randomly, and a P2 adapter was added to the sheared DNA fragments. DNA

with a P1 adapter was selectively enriched by PCR amplification. Finally, DNA fragments of 300–500 bp were gel purified and sequenced on the Illumina HiSeqXTen platform to generate 150 bp paired-end reads.

2.3. Data Analysis

To perform the data analysis, the POOLPARTY v.0.8 pipeline was used with some modifications. This pipeline uses best practices in the analysis of Pool-seq data [45,46]. The following is a brief description of each step of the DNA analysis.

2.3.1. Data Processing, Assembly and Alignment

To reduce the error rate towards the 3' end of Illumina reads, BBMAP v.37.93 (<http://jgi.doe.gov/data-and-tools/bb-tools/>) (accessed on 14 January 2020) was used. The raw reads were end-trimmed using a minimum Phred quality score of 20 and a minimum read length of 75 bp.

A de novo assembly was performed with RAINBOW v.2.0.4 [47], which is specifically designed to assemble contigs from RAD sequences. This analysis is not implemented in POOLPARTY, so it was necessary to add it to the pipeline. In short, it first clusters reads together that are less than 4 bp apart. These clustered reads are then recursively divided into groups representing individual allele sequences. Individual allele sequences are then assembled and merged into a final set of RAD contigs. Assembly statistics, such as assembly length, longest contig, and N50 and L50 lengths were calculated with QUASt v.5.0.2 [48].

As semi-global alignment and realignment of unmapped reads is recommended for a pooled sample [49], we used the bwa men algorithm of BWA v.07.12 [50] to map the filtered reads onto the reference RAD contigs. During alignment, SAMBLASTER v.0.1.24 was used to mark duplicate read pairs that arise during polymerase chain reaction amplification or sequencing [51]. Duplicate measurements are removed as they can create biased allele frequency estimates in downstream analyses. Finally, the bam files were sorted with PICARD TOOLS v.2.17.11 (<http://broadinstitute.github.io/picard/>) (accessed on 14 January 2020), and ambiguously mapped and unaligned reads were removed by SAMTOOLS v.1.5 [52].

2.3.2. Genome-Wide Genetic Variation and Population Genetic Structure

To characterize genome-wide patterns of genetic variation and population differentiation, we calculated the nucleotide diversity (Tajima's π) and population mutation rate (Watterson's Theta, θ_W) with POPOOLATION v.1.2.2 [53]. Because POPOOLATION is not included in POOLPARTY, it was necessary to add it to the pipeline. For this, the recommendations by Kofler [54] were followed, where (i) the bam files of each population were converted to pileup files with SAMTOOLS; (ii) the genomic regions with insertions/deletions (indels) were identified in the pileup files, and a gtf file was generated with the coordinates of these regions using the 'identify-genomic-indel-regions' command; and (iii) the gtf file was used to filter the regions surrounding the indels with the 'filter-pileup-by-gtf' command. To standardize the sequencing biases, each pileup file was filtered to a uniform coverage (minimum coverage of 20 and maximum coverage of 200) using the 'subsample-pileup' command. Finally, π and θ_W were calculated across a non-overlapping 1000 bp sliding window with the 'variance-sliding' command, taking into account a minimum minor allele count of two. Additionally, the significance for π and θ_W differences among populations was first assessed by conducting a one-way ANOVA. As part of a post-hoc analysis, populations were then assigned significance groupings using Tukey's honestly significant difference tests. One-way ANOVA and Tukey's test were performed with REAL STATISTICS v.6.8 (<http://www.real-statistics.com>) (accessed on 14 January 2020).

To determine the degree and pattern of the population genetic structure, the fixation index (F_{ST}) for each pairwise population comparison was calculated for each SNP as follows:

(i) with BCFTOOLS v.1.5 [55], an SNP call was made from the bam files of each population, and a variant call format (vcf) file was created using an SNP quality (SNPQ) of 20, a global minor allele frequency (MAF) of 0.05, and a total depth of coverage (MIDP) of 20; (ii) a mpileup file containing all populations was created with SAMTOOLS, taking into account the positions of the SNPs called with BCFTOOLS. This mpileup file was used to create a synchronized file (sync) with the 'mpileup2sync' command of POPOOLATION2 [56]. Finally, the F_{ST} for each pairwise comparison was calculated at each of the SNPs using the 'fst-sliding' command of POPOOLATION2. Confidence intervals for the F_{ST} estimates were inferred by bootstrapping 1000 times in the R package BOOTSTRAP v.2017.2 [57].

To visualize the multilocus patterns of population differentiation, a Principal Coordinates Analysis (PCoA) plot was generated using the R package LABDSV (<https://CRAN.R-project.org/package=labdsv>) (accessed on 14 January 2020) based on average F_{ST} values. Additionally, a hierarchical clustering tree was constructed [58] to evaluate the genetic relationships between the six populations analyzed. For this, the covariance matrix of population allele frequencies (Ω) was calculated with BAYPASS v.2.2 under the core model. The Ω was transformed into a correlation matrix using the 'cov2cor' function of R [59]. From the constructed correlation matrix, a distance matrix was defined using the 'hclust' function of R to perform the hierarchy grouping. Finally, the tree was plotted with the function 'plot.phylo' of the R package of APE [60].

2.3.3. Detection of Selection Footprints

Obtaining the SNPs that were likely to be differentiated as a result of selection was done following the current trend in the literature; applying multiple outlier analyses to detect potential outlier loci with increased stringency. The three methods described below were performed on the SNPs identified with POPOOLATION2. First, SNPs falling into the 99.9 percentile of the empirical distribution of each pairwise F_{ST} were identified as potentially differentiated loci following an empirical outlier detection approach [61].

Second, the identification of overly differentiated SNPs was based on the XtX statistics [62], which are analogous to the F_{ST} values but take into account the relationship between the analyzed populations and the sampling noise in pooled data (variation in the depth of the sequences between populations and SNPs) [62]. The XtX values for each SNP were estimated based on the core model implemented in BAYPASS v.2.2 [58] with the default parameters. To calibrate the XtX , a pseudo-observed dataset (POD) was simulated using the R function 'simulate_baypass'. SNPs with XtX estimates greater than the 99.9% threshold determined from POD were identified, resulting from adaptive divergence [58].

Finally, we relied on the Bayes factor (BF) calculated under the auxiliary covariate model (AUX), also implemented in BAYPASS, to identify differentiated SNPs associated with the host plants. The AUX model involves the introduction of a binary auxiliary variable to classify each locus as associated or not. This allows us to easily compute posterior inclusion probability (and BF) for each locus while explicitly accounting for multiple testing issues. For each SNP, the Bayes factor was expressed in deciban units (dB) via the transformation $10\log_{10}(\text{BF})$. As a decision rule, we then followed the popular Jeffrey's rule that quantifies the strength of evidence using the criterion of $dB > 20$ as strong evidence [58]. To detect outlier SNPs, we compared the *A. obliqua* populations within the sampled sites (north and south), separately, using core model and AUX model.

To identify candidate genes associated with the outlier SNPs found with the AUX model and at least one of the other two methods, each of the contigs associated with each outlier SNP were annotated with BLAST2GO [63]. We give priority to the AUX model since, of the three, it is the one that allows us to analyze whether a specific SNP can be associated with a given covariate (CaN vs. RmN, CaN vs. MN, RmN vs. MN, CaS vs. RmS, CaS vs. MS and, RmS vs. MS). BLAST2GO was run under default parameters using a blastx search against protein database of arthropods and only considering hits with an e-value cut-off of 1×10^{-5} . Functional annotation and gene ontology (GO) term mapping was also

performed in BLAST2GO under default parameters. Visualization of GO annotations was carried out with WEGO [64].

3. Results

3.1. Data Processing, Assembly and Alignment

A total of 258.3 million paired reads were obtained from NGS sequencing, with an average of 43 million per population. After filtering, 240.8 million paired reads were retained, with an average of 40.1 million per population (Table 2). The de novo assembly produced 43,175 contig sequences, which were used as reference sequences for the following alignment. The de novo assembly was roughly 25.5 Mb in length (this metric was computed for contigs that exceed the threshold specified in 500 bp), the longest contig was 1699 bp, and the N50 and L50 were 577 bp and 19,332 bp, respectively (Table S1). A total of 124.6 million reads were aligned to the reference sequences, and the number of aligned reads ranged from 20.1 to 21.5 million depending on the population (Table 2).

Table 2. Estimators of genetic diversity in the populations of the species *Anastrepha obliqua* of southwestern Colombia. π is the average nucleotide diversity, and θ_W is the average population mutation rate. Letters a, b and c that are found as a superscript in the values π and θ_W denote which populations are significantly different as determined by Tukey's test ($p < 0.05$). Meaning of population codes is provided in Table 1.

Population	Accessions Code	Number of Raw Reads	Number of Aligned Reads	Number of SNPs	π	θ_W
CaN	SRR13442704	43,020,774	20,493,546	488,027	0.0105 ^a	0.0112 ^b
CaS	SRR13442703	43,183,116	21,589,043	565,527	0.0108 ^b	0.0117 ^c
RmN	SRR13442702	43,003,430	20,134,297	481,206	0.0106 ^a	0.0115 ^a
RmS	SRR13442701	43,036,924	20,554,814	503,735	0.0105 ^a	0.0115 ^a
MN	SRR13442700	43,020,380	20,765,409	524,277	0.0105 ^a	0.0115 ^a
MS	SRR13442699	43,048,340	21,039,009	546,956	0.0106 ^a	0.0115 ^a

3.2. Genome-Wide Genetic Variation and Population Genetic Structure

To estimate the levels of genome-wide genetic variation, POPOOLATION identified between 481,206 and 565,527 SNPs in the populations evaluated (Table 2). The within-population genome-wide average Tajima's π and genome-wide average Watterson's θ_W ranged from 0.0105 (populations CaN, RmS and MN) to 0.0108 (population CaS) and from 0.0112 (population CaN) to 0.0117 (population CaS), respectively.

To estimate the levels of genome-wide genetic differentiation between populations, a total of 13,566 SNPs that satisfied our filtering criteria (see methods) were subsampled across all populations with POPOOLATION2. The average pairwise F_{ST} across all SNPs were similar between the populations. The values varied between 0.0105 (MS vs. CaS) and 0.0162 (CaN vs. CaS). Slightly higher F_{ST} values were obtained between the CaN population and the other populations (Table 3). This result was consistent with the differentiation pattern observed in the PCoA plot, where the CaN population was the most differentiated from the other populations along the first axis (PCoA1), which explained 55.24% of the genome-wide genetic variation. (Figure 2). Additionally, this axis revealed a geographically ordered pattern where it is observed that the southern populations (CaS, MS and RmS) present a higher degree of genomic relationship with each other than with the northern populations (CaN, MN and RmN). The second axis (PCoA2) explained 12.59% of the variation (Figure 2) and pooled the CaS, RmN, and MS populations, while CaN, RmS, and MN were not pooled with any other population. The reconstruction of the genetic relationships between the populations, based on the hierarchical clustering tree (Figure 3), revealed a differentiation pattern similar to that delineated by the first axis of the PCoA. According to this tree, the MS, RmS, and CaS populations were more closely related to each other and comprised a group separated from the other populations. On the other hand, the CaN and MN populations were grouped, while RmN was separated from them as a basal population for all the others. We can say, then, from the PCoA and the hierarchical

grouping tree that the pattern of differentiation of the analyzed populations of the species *A. obliqua* in southwestern Colombia is established mainly by geography.

Table 3. Average pairwise F_{ST} values among the six populations of the species *Anastrepha obliqua* of southwestern Colombia based on the 13,566 SNPs identified using POPOOLATION2 (lower diagonal) and 95% confidence intervals (upper diagonal). Meaning of population codes is provided in Table 1.

Population	CaN	CaS	RmN	RmS	MN	MS
CaN	-	(0.0159–0.0166)	(0.0158–0.0164)	(0.0155–0.0161)	(0.0154–0.0160)	(0.0153–0.0159)
CaS	0.0162	-	(0.0106–0.0111)	(0.0105–0.0110)	(0.0109–0.0116)	(0.0103–0.0108)
RmN	0.0161	0.0109	-	(0.0117–0.0123)	(0.0112–0.0118)	(0.0108–0.0113)
RmS	0.0158	0.0107	0.0120	-	(0.0109–0.0115)	(0.0107–0.0112)
MN	0.0157	0.0113	0.0115	0.0112	-	(0.0112–0.0118)
MS	0.0156	0.0105	0.0111	0.0110	0.0115	-

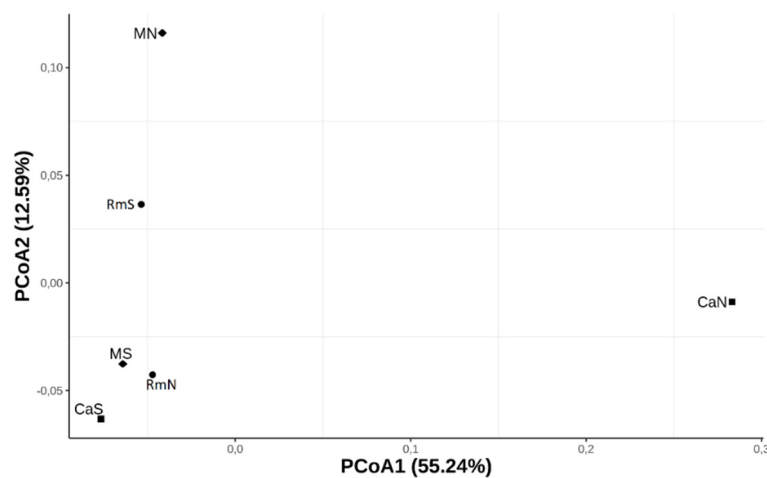


Figure 2. Principal coordinates analysis (PCoA). Plot of the overall average pairwise F_{ST} values of 13,566 SNPs among the six populations of the species *Anastrepha obliqua* of southwestern Colombia.

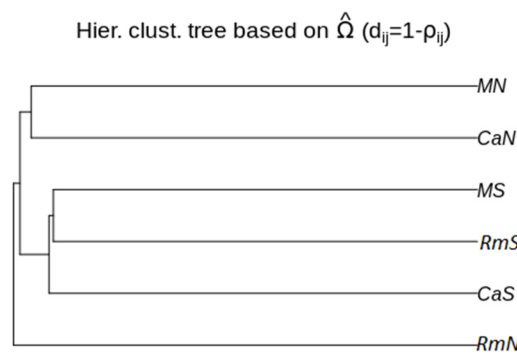


Figure 3. Hierarchical clustering tree showing the genetic relationships among the six populations of the species *Anastrepha obliqua* of southwestern Colombia.

3.3. Detection and Annotation of SNPs under Selection

Using the empirical outlier detection approach, 14 SNPs (in 3 contigs) were found at least once in the top 0.1% of pairwise F_{ST} values ($F_{ST} > 0.0912$, Figure S1), which were therefore possible candidates for divergent selection. Additionally, we detected significant selection signatures based on the differences in allele frequencies between populations within the sampling sites (north and south), separately, using two different and robust Bayesian approaches. First, a genomic exploration was performed for differentiated SNPs based on the XfX statistic. (the decision criterion to identify SNPs under divergent selection was $XfX > 6.63$, Figure S2). In total, 210 SNPs were detected in 79 contigs. Second, we

performed an association analysis to detect SNPs that could be under selection and could be correlated with host plants using the AUX model, also implemented in BAYPASS, with which we identified 77 SNPs ($dB > 20$) in 44 contigs. To our surprise, some overly differentiated SNPs (high values of the XtX statistic) were not associated with the use of the host plant (dB below expectations), which indicated the presence of other selective pressures. From the results of all three methods, 233 SNPs were identified in 113 contigs. Of these, 54 SNPs in 23 contigs were detected by the AUX model and at least one of the other two methods (Figure 4). The distribution of these 54 SNPs in the population comparisons performed (CaN vs. RmN, CaN vs. MN, and RmN vs. MN for the north; CaS vs. RmS, CaS vs. MS, and RmS vs. MS for the south) showed a higher proportion of SNPs that were associated with host plants in the northern region than in the southern region, suggesting a greater effect of host plant on northern populations of *A. obliqua* (Table S2). The SNPs in the north were different from the SNPs found in the south (Table S2). Four SNPs were recurrent in the north when we compared the MN population with respect to the CaN and RmN populations. In the south, we only found one recurring SNP, between the MS population and the other populations (CaS and RmS) (Table S2).

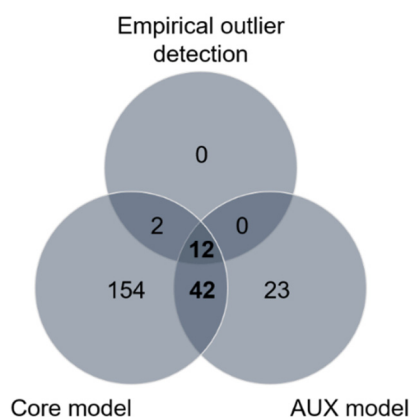


Figure 4. Summary of selection footprint detection using different (empirical, core model and AUX model) methods for the six populations of the species *Anastrepha obliqua* of southwestern Colombia. SNPs identified with AUX model and at least one of the other two methods are considered putative candidate SNPs under host-related divergent selection.

Of the 23 contigs that contained SNPs that were shown to be correlated with the use of host plants, we identified 17 contigs that presented homologous sequences with the NCBI arthropod protein database in a BLASTx search (Table S3). Of the 17 contigs that managed to be annotated, six could be associated with a GO term (Figure S3). Gene ontology (GO) annotations revealed several loci could be assigned molecular functions of catalytic activity and binding processes and loci that could be related to cellular components, especially membrane components (Figure S3).

We also determined the protein families related to the identified sequences. For this, we used the InterProScan tool of the BLASTO2GO program to retrieve information about the protein families (Figure 5).

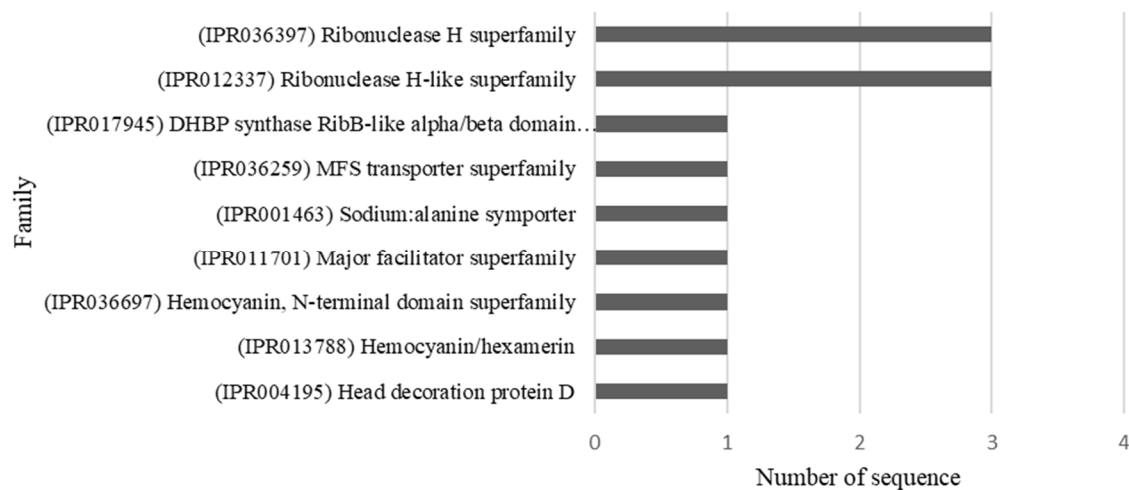


Figure 5. Families of proteins associated with the sequences that contain SNPs that showed signs of being differentially associated to host plants of *Anastrepha obliqua* analyzed.

4. Discussion

In this study, we used a pooled RAD-seq approach to evaluate genomic differences between six wild populations of the species *Anastrepha obliqua* associated with the use of host plant. To date, genetic studies that have provided information on the association of *A. obliqua* with its host plants have focussed on the use of data from mitochondrial DNA sequences and microsatellite markers of nuclear DNA [24,26], and there are no analyses at the genomic level in the species *A. obliqua*. This study is the first to provide information at this level on the population differentiation of *A. obliqua* associated with its host plants. The most interesting findings of this work are as follows: (i) the analysis of the genetic differentiation between the populations yielded low values of F_{ST} based on the genomic SNPs examined. However, (ii) a subset of them showed signs of adaptive differentiation, suggesting the presence of selective pressures in these populations. Of the SNPs identified as being putatively adaptive 54 were associated with host plant specificity. Finally, (iii) some of the adaptive genomic regions linked with host plant specificity are associated with protein families involved in nutrition.

4.1. Genome-Wide Genetic Variation and Population Genetic Structure

The π and θ_W between the analyzed populations were very similar, with a variation of 0.0105 to 0.0108 for π and from 0.0112 to 0.0117 for θ_W . However, we found evidence, with Tukey's test, of significant differences between the patterns of genomic variation in the populations, for both π and for θ_W . According to our analysis, the differences found between the patterns genomic variation were more pronounced between the populations of *A. obliqua* from carambola fruits than between mango and red mombin populations (Table 2). We did not find significant differences between the patterns of genetic variation of *A. obliqua* populations that infest mango and red mombin. On the other hand, the degree of genomic differentiation was generally low (F_{ST} ranged from 0.0105 to 0.0162) among the populations of *A. obliqua* of different hosts plants species. This result is similar to that found by Bakovic et al. [9] when estimating genomic differentiation in *Rhagoletis cerasi* populations from 2494 SNPs. This fruit fly uses two species of host plants, *Prunus* spp. and *Lonicera* spp., which can be found in sympatry. These researchers revealed that the genomic differentiation between the populations of this species was low (F_{ST} ranged from 0.004 to 0.143), especially when they compared populations of different host plants found in sympatry. Based on the average pairwise F_{ST} values obtained, we can suggest that the populations analyzed in this study show low genetic differentiation. The level of genetic differentiation observed is likely due to that ongoing genetic flow persists between the populations or due to a recent descent of the populations.

The population of *A. obliqua* with the greatest degree of genomic differentiation from the others was CaN, since its average pairwise F_{ST} values were higher. This result is consistent with the differentiation inferred from the patterns of genetic variation of the populations. What could have caused this? Studies of the chemical composition of fruits of different plant species have shown that they can vary widely in both nutrients and secondary metabolites, which could be toxic to fruit flies. For example, Oroño et al. [22] examined the genetic structure of three populations of *A. fraterculus* associated with three species of host plants that were in sympatry and differed in chemical composition, with the aim of testing the hypothesis of sympatric speciation mediated by chemical substances. The researchers found that the adaptive processes related to the plant chemistry of the host plants seem to have promoted genetic differentiation between the populations of *A. fraterculus*, since the individuals analyzed were pooled according to the host plants and, in addition, extremely low gene flow was discovered between populations. In this sense, it is possible that the chemical composition of the fruits of carambola plants, which belong to the family Oxalidaceae, exerts selective pressure on the larvae that feed on the carambola pulp, promoting their genomic differentiation more than the red mombin and mango plants, which are part of the family Anacardiaceae. This differentiable pattern was also found when the bacterial community was analyzed in populations that infest the same three host plants (red mombin, mango, and carambola) in the same geographic region analysed by us [65]. These researchers found that composition of the bacterial community of *A. obliqua* of carambola fruits was significantly different from that of the same fruit fly in mango and red mombin. On the other hand, Aguirre-Ramirez et al. [24] reported the existence of genetic differentiation in *A. obliqua* populations that were associated with different host plants. Particularly for carambola fruits, these authors reported a lower haplotype diversity than was found on mango and red mombin fruits. Because the number of individuals in the CaN population is lower than that of the other populations (see Table 1), we suggest that future research consider the possibility of corroborating the interpretations made here about the CaN population. This will enable researchers to understand if the genetic differentiation of the CaN population due to the low number of individuals sampled, and whether that with a higher sample number the within-population genetic differentiation will be reduced.

In general, we did not observe a genetic structure associated with the host plant. The first axis (PCoA1) of the principal coordinate analysis and the hierarchical clustering tree suggests the possibility of a genetic relationship among the populations of *A. obliqua* that were in sympatry sharing the same geographic region, be it the northern region or the southern region of the Inter-Andean Valley of the Cauca River. This finding could have two possible explanations. First, and according to Hernández et al. [43], the average dispersion distance of wild flies of the species *A. obliqua* is approximately 151 m. We estimate that the distance between the northern and southern localities sampled here is approximately 160 km. Therefore, the mobility of this species, from one region to another, could be limited by its capacity of dispersion, which might explain the closer genetic relationship between the populations of the same region. However, when looking at the characteristics of the Inter-Andean Valley of the Cauca River, we can find a continuum of fruit trees that host *A. obliqua* from north to south, which could favor the gene flow from one region to another. Therefore, the dispersal capacity of *A. obliqua* would not be the only factor that would explain the closer genetic relationship between populations of the same region. The second explanation would be related to the settlement of these populations in different regions, where flies can find their host plants available without having to travel long distances. Our observations differ from the results obtained by Aguirre-Ramirez et al. [24], who did not find a population genetic structure of *A. obliqua* associated with geographic distance between populations from the same geographic sites analyzed in this study. However, it should be noted that our two studies used different molecular techniques. In Aguirre-Ramirez et al. [24], the analyses were performed on the molecular data obtained from two

mitochondrial genes (COI and ND6), while in this work, we used the pooled RAD-seq technique to detect divergence patterns throughout the genome.

It should be noted that the main objective of this research was to assess genomic differentiation and population structure in sympatric populations of *A. obliqua* that infest three different host plants. Additionally, we tried an isolation by distance test as described in [66], but it was not possible since our experimental design did not allow it. To perform an isolation by distance test it is necessary to determine the geographical distance and genetic distance between the populations analyzed; with these variables it is possible to estimate the regression equation. However, we have genetic differences between the different populations but we only have a single reference of geographic distance between all populations, which corresponds to the geographic distance that we have estimated from north to south. We consider that the results obtained from the PCoA and clustering tree could guide new research works where the isolation by distance test is considered and compared with results obtained here. According to Doellman et al. [67], in the case of relatively high levels of gene flow, it would be expected that populations associated with host plants would generally be genetically pooled according to geography, with a subset of the genome that would show differentiation related to host plant. Our results agree with the patterns observed by Doellman et al. [67] and with the findings of Ruiz-Arce et al. [26], who worked with populations of the species *A. obliqua* in Veracruz, Mexico. They mention that the genetic structure of populations of the species *A. obliqua* in this locality is determined more by geography than by the host plant.

4.2. Detection and Annotation of SNPs under Selection

We do not rule out the participation of host plants in the differentiation of populations, since we identified a subset of SNPs that presented differences associated with the host plant. Research on insects phytophages that manage to form host races has helped to understand how the accumulation of small changes in the genome that show some association with the use of the host plant can act in combination to reduce the effective rate of gene flow and promote the formation of host races and subsequent speciation [68]. For this reason, the subset of outlier SNPs associated with host plants that we identified could be considered evidence of the diversification processes mediated by host plants in the species *A. obliqua*.

We identified 54 SNPs that showed significant differentiation associated with host plant specificity. BLAST search queries and the annotation of the SNP-containing 23 RAD contigs allowed us to identify 17 candidate genes and nine protein families. Among the most interesting annotated protein families, the MFS transporter superfamily and major facilitator superfamily stands out; they are related to the digestive system of herbivorous insects [11] and are involved in the hydrolysis of sucrose and absorption of nutrients [69]. These protein superfamilies could be a key factor that facilitates the adaptation and survival of insects in response to new diets [70]. Additionally, the Hemocyanin, N-terminal domain superfamily and Hemocyanin/hexamerin, which are related to the storage of nutrients, specifically amino acids, for the synthesis of proteins in adults were identified [71]. We also identified the Ribonuclease H superfamily and the Ribonuclease H-like superfamily, which are associated with numerous fundamental processes, including DNA replication and repair, homologous recombination, transposition and RNA interference [72]. Finally, we identified a DHBP synthase RibB-like alpha/beta domain superfamily; this superfamily is related to the enzyme DHBP synthase RibB that catalyzes the conversion of D-ribulose 5-phosphate to formate and 3,4-dihydroxy-2-butanone 4-phosphate [73], and is also related to the YrdC family of hypothetical proteins that are widely distributed in eukaryotes and prokaryotes. YrdC is predicted to be an rRNA maturation factor, as deletions in its gene lead to immature ribosomal 30S subunits and, consequently, fewer translating ribosomes [74].

5. Conclusions

In conclusion, we present evidence that supports the effect of host plants on genomic differentiation in populations of the species *A. obliqua*. We cannot say with certainty that this species presents host races. However, our results support the hypothesis that the species *A. obliqua* could be in an incipient stage of population differentiation mediated in part by its host plants. Our results add to a growing list of possible examples of ongoing ecological divergence. We identified a small group of candidate genes that could be under divergent selection; a product of the pressure mediated by the host plants. We believe that in future research, an analysis based on the sequencing of the complete genome of individuals of the species *A. obliqua* would allow us to detect more important genomic regions that support the adaptation of *A. obliqua* to their host plants. Our study is a first step to better understand the population genomics of an important pest of fruit crops, *A. obliqua*.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/insects12121116/s1>, Figure S1: Distribution of the average pairwise F_{ST} values in the 13,566 SNPs obtained with POPOOLATION2, Figure S2: Genomic exploration for loci with a high degree of differentiation according to the XtX statistics estimated under the central BAYPASS model. The dashed line represents the significance threshold of 0.1%, Figure S3: GO classification of the six contigs with SNPs correlated with the use of the host plant, Table S1: Assembly statistics, such as number of contigs at certain lengths, total assembly length, N50 and L50 lengths, and number of unknown bases per kilobase (# N's per 100 kbp) are reported for *A. obliqua*, Table S2: BAYPASS analysis using the AUX model to detect SNPs associated with use of host plants: *Mangifera indica* (mango), *Spondias purpurea* (red mombin) and *Averrhoa carambola* (carambola) in six populations of the species *Anastrepha obliqua* located in the north and south of southwestern, Table S3: BLAST and GO annotation of 23 contigs that contain SNPs that showed signs of being differentially associated to host plants of *A. obliqua* analysed.

Author Contributions: E.A.-R., S.V.-C. and N.T.-P. conceived and designed research. E.A.-R. and S.V.-C. conducted field and laboratory work including data collection and curation. E.A.-R. analysis of data. E.A.-R. wrote original draft of the manuscript. E.A.-R., S.V.-C. and N.T.-P. advised on experimental design and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was internally funded by Universidad del Valle with the project entitled "Análisis de la huella genómica en la mosca de la fruta *Anastrepha obliqua*, asociada con su condición de insecto polífago" with information code 71157.

Data Availability Statement: The raw sequence data have been deposited in NCBI's Sequence Read Archive, and accession code are listed in Table 2.

Acknowledgments: Thanks to Lorena Ruiz Montoya and Diana Carolina López for the critical reading of this manuscript which helped to improve it.

Conflicts of Interest: The authors declare they have no conflict of interest.

References

1. Guo, B.; Li, Z.; Merilä, J. Population genomic evidence for adaptive differentiation in the Baltic Sea herring. *Mol. Ecol.* **2016**, *25*, 2833–2852. [[CrossRef](#)]
2. Hereford, J. A quantitative survey of local adaptation and fitness trade-offs. *Am. Nat.* **2009**, *173*, 579–588. [[CrossRef](#)]
3. Blanquart, F.; Kaltz, O.; Nuismer, S.L.; Gandon, S. A practical guide to measuring local adaptation. *Ecol. Lett.* **2013**, *16*, 1195–1205. [[CrossRef](#)]
4. Leinonen, T.; McCairns, R.J.S.; O'Hara, R.B.; Merilä, J. Q(ST)-F(ST) comparisons: Evolutionary and ecological insights from genomic heterogeneity. *Nat. Rev. Genet.* **2013**, *14*, 179–190. [[CrossRef](#)] [[PubMed](#)]
5. Berlocher, S.H.; Feder, J.L. Sympatric speciation in phytophagous insects: Moving beyond controversy? *Annu. Rev. Entomol.* **2002**, *47*, 773–815. [[CrossRef](#)] [[PubMed](#)]
6. Drès, M.; Mallet, J. Host races in plant-feeding insects and their importance in sympatric speciation. *Philos. Trans. R. Soc. B Biol. Sci.* **2002**, *357*, 471–492. [[CrossRef](#)]

7. Matsubayashi, K.W.; Ohshima, I.; Nosil, P. Ecological speciation in phytophagous insects. *Entomol. Exp. Appl.* **2010**, *134*, 1–27. [[CrossRef](#)]
8. Forbes, A.A.; Devine, S.N.; Hippee, A.C.; Tvedte, E.S.; Ward, A.K.G.; Widmayer, H.A.; Wilson, C.J. Revisiting the particular role of host shifts in initiating insect speciation. *Evolution* **2017**, *71*, 1126–1137. [[CrossRef](#)]
9. Bakovic, V.; Schuler, H.; Schebeck, M.; Feder, J.L.; Stauffer, C.; Ragland, G.J. Host plant-related genomic differentiation in the European cherry fruit fly, *Rhagoletis cerasi*. *Mol. Ecol.* **2019**, *28*, 4648–4666. [[CrossRef](#)]
10. Jousset, E.; Elias, M. Testing host-plant driven speciation in phytophagous insects: A phylogenetic perspective. *arXiv* **2019**, arXiv:1910.09510.
11. Simon, J.C.; D'alengon, E.; Guy, E.; Jacquin-Joly, E.; Jaquière, J.; Nouhaud, P.; Peccoud, J.; Sugio, A.; Streiff, R. Genomics of adaptation to host-plants in herbivorous insects. *Brief. Funct. Genom.* **2015**, *14*, 413–423. [[CrossRef](#)]
12. Dambroski, H.R.; Feder, J.L. Host plant and latitude-related diapause variation in *Rhagoletis pomonella*: A test for multifaceted life history adaptation on different stages of diapause development. *J. Evol. Biol.* **2007**, *20*, 2101–2112. [[CrossRef](#)] [[PubMed](#)]
13. Feder, J.L.; Filchak, K.E. It's about time: The evidence for host plant-mediated selection in the apple maggot fly, *Rhagoletis pomonella*, and its implications for fitness trade-offs in phytophagous insects. *Entomol. Exp. Appl.* **1999**, *91*, 211–225. [[CrossRef](#)]
14. Feder, J.L.; Egan, S.P.; Nosil, P. The genomics of speciation-with-gene-flow. *Trends Genet.* **2012**, *28*, 342–350. [[CrossRef](#)] [[PubMed](#)]
15. Rull, J.; Tadeo, E.; Lasa, R.; Díaz-Fleischer, F.; Arredondo, J.; Aluja, M. Reproductive compatibility among Mexican populations of *Anastrepha obliqua*: Theoretical and management implications. *J. Appl. Entomol.* **2018**, *142*, 667–678. [[CrossRef](#)]
16. Schwarz, D.; Matta, B.M.; Shakir-Botteri, N.L.; McPheron, B.A. Host shift to an invasive plant triggers rapid animal hybrid speciation. *Nature* **2005**, *436*, 546–549. [[CrossRef](#)]
17. Cha, D.H.; Olsson, S.B.; Yee, W.L.; Goughnour, R.B.; Hood, G.R.; Mattsson, M.; Schwarz, D.; Feder, J.L.; Linn, C.E. Identification of host fruit volatiles from Snowberry (*Symphoricarpos albus*), Attractive to *Rhagoletis zephyria* flies from the Western United States. *J. Chem. Ecol.* **2017**, *43*, 188–197. [[CrossRef](#)]
18. Troya, H.; Norrbom, A.L.; Bolanos Pineda, J.E. Two new species of *Anastrepha* (Diptera: Tephritidae) from Ecuador. *Zootaxa* **2020**, *4820*, 366–372. [[CrossRef](#)]
19. Aluja, M.; Birke, A. Habitat use by adults of *Anastrepha obliqua* (Diptera: Tephritidae) in a mixed mango and tropical plum orchard. *Ann. Entomol. Soc. Am.* **1993**, *86*, 799–812. [[CrossRef](#)]
20. Camargo, C.; Odell, E.; Jirón, L.F. Interspecific interactions and host preference of *Anastrepha obliqua* and *Ceratitis capitata* (Diptera: Tephritidae), two pests of mango in Central America. *Fla. Entomol.* **1996**, *79*, 266–268. [[CrossRef](#)]
21. Aluja, M.; Pérez-Staples, D.; Macías-Ordóñez, R.; Piñero, J.; McPheron, B.; Hernández-Ortiz, V. Nonhost status of Citrus sinensis cultivar valencia and C. paradisi cultivar ruby red to Mexican *Anastrepha fraterculus* (Diptera: Tephritidae). *J. Econ. Entomol.* **2003**, *96*, 1693–1703. [[CrossRef](#)]
22. Oroño, L.; Paulin, L.; Alberti, A.C.; Hilal, M.; Ovruski, S.; Vilardi, J.C.; Rull, J.; Aluja, M. Effect of host plant chemistry on genetic differentiation and reduction of gene flow among *Anastrepha fraterculus* (Diptera: Tephritidae) populations exploiting sympatric, synchronic hosts. *Environ. Entomol.* **2013**, *42*, 790–798. [[CrossRef](#)] [[PubMed](#)]
23. Gómez-Cendra, P.V.; Paulin, L.E.; Orono, L.; Ovruski, S.M.; Vilardi, J.C. Morphometric differentiation among *Anastrepha fraterculus* (Diptera: Tephritidae) exploiting sympatric alternate hosts. *Environ. Entomol.* **2016**, *45*, 508–517. [[CrossRef](#)]
24. Aguirre-Ramirez, E.J.; Velasco-Cuervo, S.M.; Gallo-Franco, J.J.; González, R.; Carrejo, N.S.; Toro-Perea, N. Genetic diversity and population structure of *Anastrepha obliqua* in southwestern Colombia. *Entomol. Exp. Appl.* **2017**, *164*, 291–304. [[CrossRef](#)]
25. Hernández, E.; Ruiz-Montoya, L.; Toledo, J.; Montoya, P.; Liedo, P.; Aceituno-Medina, M.; Perales, H. A comparison of sexual competitiveness and demographic traits of *Anastrepha obliqua* (Macquart) (Diptera: Tephritidae) among fruit-associated populations. *Bull. Entomol. Res.* **2019**, *109*, 333–341. [[CrossRef](#)]
26. Ruiz-Arce, R.; Islam, M.-S.; Aluja, M.; McPheron, B.A. Genetic variation in *Anastrepha obliqua* (Diptera: Tephritidae) in a highly diverse tropical environment in the Mexican state of Veracruz. *J. Econ. Entomol.* **2019**, *112*, 2952–2965. [[CrossRef](#)] [[PubMed](#)]
27. Santos, R.P.D.; Silva, J.G.; Miranda, E.A. The past and current potential distribution of the fruit fly *Anastrepha obliqua* (Diptera: Tephritidae) in South America. *Neotrop. Entomol.* **2019**, *49*, 284–291. [[CrossRef](#)]
28. Carrejo, N.; González, R. Lista preliminar de las moscas de la fruta del genero *Anastrepha* (Dip.: Tephritidae) en el departamento del Valle del Cauca Cali, Colombia. *Boletín Mus. Entomol. Univ. Val.* **1994**, *2*, 85–93.
29. Conover, D.; Schultz, E. Phenotypic similarity and the evolutionary significance of countergradient variation. *Trends Ecol. Evol.* **1995**, *10*, 248–252. [[CrossRef](#)]
30. Kremer, A.; Le Corre, V. Decoupling of differentiation between traits and their underlying genes in response to divergent selection. *Heredity* **2012**, *108*, 375–385. [[CrossRef](#)]
31. Comeault, A.A.; Soria-Carrasco, V.; Gompert, Z.; Farkas, T.E.; Buerkle, C.A.; Parchman, T.L.; Nosil, P. Genome-wide association mapping of phenotypic traits subject to a range of intensities of natural selection in *Timema cristinae*. *Am. Nat.* **2014**, *183*, 711–727. [[CrossRef](#)]
32. Ellegren, H. Genome sequencing and population genomics in non-model organisms. *Trends Ecol. Evol.* **2014**, *29*, 51–63. [[CrossRef](#)] [[PubMed](#)]
33. Egan, S.P.; Ragland, G.J.; Assour, L.; Powell, T.H.Q.; Hood, G.R.; Emrich, S.; Nosil, P.; Feder, J.L. Experimental evidence of genome-wide impact of ecological selection during early stages of speciation-with-gene-flow. *Ecol. Lett.* **2015**, *18*, 817–825. [[CrossRef](#)]

34. Dupuis, J.R.; Sim, S.B.; San Jose, M.; Leblanc, L.; Hoassain, M.A.; Rubinoff, D.; Geib, S.M. Population genomics and comparisons of selective signatures in two invasions of melon fly, *Bactrocera cucurbitae* (Diptera: Tephritidae). *Biol. Invasions* **2018**, *20*, 1211–1228. [[CrossRef](#)]
35. Nouhaud, P.; Gautier, M.; Gouin, A.; Jaquiéry, J.; Peccoud, J.; Legeai, F.; Mieuze, L.; Smadja, C.M.; Lemaitre, C.; Vitalis, R.; et al. Identifying genomic hotspots of differentiation and candidate genes involved in the adaptive divergence of pea aphid host races. *Mol. Ecol.* **2018**, *27*, 3287–3300. [[CrossRef](#)] [[PubMed](#)]
36. Wachi, N.; Matsubayashi, K.W.; Maeto, K. Application of next-generation sequencing to the study of non-model insects. *Entomol. Sci.* **2018**, *21*, 3–11. [[CrossRef](#)]
37. Koch, J.B.; Dupuis, J.R.; Ouedraogo, N.; Geib, S.M.; Follett, P.A.; Price, D.K. Population genomic and phenotype diversity of invasive *Drosophila suzukii* in Hawai'i. *Biol. Invasions* **2020**, *22*, 1753–1770. [[CrossRef](#)]
38. Guo, B.; DeFaveri, J.; Sotelo, G.; Nair, A.; Merilä, J. Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks. *BMC Biol.* **2015**, *13*, 19. [[CrossRef](#)]
39. Baird, N.A.; Etter, P.D.; Atwood, T.S.; Currey, M.C.; Shiver, A.L.; Lewis, Z.A.; Selker, E.U.; Cresko, W.A.; Johnson, E.A. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* **2008**, *3*, e3376. [[CrossRef](#)]
40. Ragland, G.J.; Doellman, M.M.; Meyers, P.; Hood, G.R.; Egan, S.P.; Powell, T.H.Q.; Hahn, D.A.; Nosil, P.; Feder, J.L. A test of genomic modularity among life history adaptations promoting speciation- with-gene-flow. *Int. J. Lab. Hematol.* **2017**, *26*, 3926–3942. [[CrossRef](#)]
41. Ramirez, F.; Davenport, T.L. *Mangoes in Colombia*; Valavi, S.G., Rajmohan, K., Govil, J.N., Peter, K.V., Thottappilly, G., Eds.; Studium Press LLC: Houston, TX, USA, 2012; ISBN 1-933699-94-9.
42. Mateus, D.; Arias, M.; Orduz, J. El cultivo de carambolo (*Averrhoa carambola* L.) y su comportamiento en el piedemonte del Meta. *Rev. Colomb. Cienc. Hortícolas* **2015**, *9*, 135–148. [[CrossRef](#)]
43. Hernández, E.; Orozco, D.; Breceda, S.F.; Domínguez, J. Dispersal and longevity of wild and mass-reared *Anastrepha Ludens* and *Anastrepha obliqua* (Diptera: Tephritidae). *Fla. Entomol.* **2007**, *90*, 123–135. [[CrossRef](#)]
44. Caraballo, J. Diagnóstico y clave pictórica para las especies del género *Anastrepha* Schiner, 1868 (Diptera: Tephritidae) de importancia económica en Venezuela. *Entomotropica* **2007**, *16*, 157–164.
45. Micheletti, S.J.; Narum, S.R. Utility of pooled sequencing for association mapping in nonmodel organisms. *Mol. Ecol. Resour.* **2018**, *18*, 825–837. [[CrossRef](#)] [[PubMed](#)]
46. Schlötterer, C.; Tobler, R.; Kofler, R.; Nolte, V. Sequencing pools of individuals-mining genome-wide polymorphism data without big funding. *Nat. Rev. Genet.* **2014**, *15*, 749–763. [[CrossRef](#)]
47. Chong, Z.; Ruan, J.; Wu, C.I. Rainbow: An integrated tool for efficient clustering and assembling RAD-seq reads. *Bioinformatics* **2012**, *28*, 2732–2737. [[CrossRef](#)]
48. Mikheenko, A.; Prijbelski, A.; Saveliev, V.; Antipov, D.; Gurevich, A. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* **2018**, *34*, i142–i150. [[CrossRef](#)]
49. Davey, J.W.; Cezard, T.; Fuentes-Utrilla, P.; Eland, C.; Gharbi, K.; Blaxter, M.L. Special features of RAD Sequencing data: Implications for genotyping. *Mol. Ecol.* **2013**, *22*, 3151–3164. [[CrossRef](#)]
50. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)]
51. Faust, G.G.; Hall, I.M. SAMBLASTER: Fast duplicate marking and structural variant read extraction. *Bioinformatics* **2014**, *30*, 2503–2505. [[CrossRef](#)]
52. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)] [[PubMed](#)]
53. Kofler, R.; Orozco-terWengel, P.; de Maio, N.; Pandey, R.V.; Nolte, V.; Futschik, A.; Kosiol, C.; Schlötterer, C. Popoolation: A toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS ONE* **2011**, *6*, e15925. [[CrossRef](#)] [[PubMed](#)]
54. Kofler, R. Population Genomics with Sequenced Pools (Pool-Seq). Available online: https://www.kofler.or.at/bioinformatic/wp-content/uploads/2018/07/pooledAnalysis_part1.pdf (accessed on 19 January 2020).
55. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **2011**, *27*, 2987–2993. [[CrossRef](#)] [[PubMed](#)]
56. Kofler, R.; Pandey, R.V.; Schlötterer, C. PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* **2011**, *27*, 3435–3436. [[CrossRef](#)]
57. Efron, B.; Tibshirani, R.J. *An Introduction to the Bootstrap*; Chapman & Hall: Boca Raton, FL, USA, 1994; ISBN 9780412042317.
58. Gautier, M. Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics* **2015**, *201*, 1555–1579. [[CrossRef](#)]
59. Team, R.C. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Australia, 2019.
60. Paradis, E.; Claude, J.; Strimmer, K. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **2004**, *20*, 289–290. [[CrossRef](#)]
61. Akey, J.M.; Ruhe, A.L.; Akey, D.T.; Wong, A.K.; Connelly, C.F.; Madeoy, J.; Nicholas, T.J.; Neff, M.W. Tracking footprints of artificial selection in the dog genome. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 1160–1165. [[CrossRef](#)]

62. Günther, T.; Coop, G. Robust identification of local adaptation from allele frequencies. *Genetics* **2013**, *195*, 205–220. [[CrossRef](#)]
63. Conesa, A.; Götz, S.; García-gómez, J.M.; Terol, J.; Talón, M.; Genómica, D.; Valenciano, I.; Agrarias, D.I.; Valencia, U.P. De Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **2005**, *21*, 3674–3676. [[CrossRef](#)]
64. Ye, J.; Fang, L.; Zheng, H.; Zhang, Y.; Chen, J.; Zhang, Z.; Wang, J.; Li, S.; Li, R.; Bolund, L.; et al. WEGO: A web tool for plotting GO annotations. *Nucleic Acids Res.* **2006**, *34*, 293–297. [[CrossRef](#)]
65. Gallo-Franco, J.J.; Toro-Perea, N. Variations in the bacterial communities in *Anastrepha obliqua* (Diptera: Tephritidae) according to the insect life stage and host plant. *Curr. Microbiol.* **2020**, *77*, 1283–1291. [[CrossRef](#)]
66. Kahnt, B.; Theodorou, P.; Soro, A.; Hollens-Kuhr, H.; Kuhlmann, M.; Pauw, A.; Paxton, R.J. Small and genetically highly structured populations in a long-legged bee, *Rediviva longimanus*, as inferred by pooled RAD-seq. *BMC Evol. Biol.* **2018**, *18*, 196. [[CrossRef](#)] [[PubMed](#)]
67. Doellman, M.M.; Ragland, G.J.; Hood, G.R.; Meyers, P.J.; Egan, S.P.; Powell, T.H.Q.; Lazorchak, P.; Glover, M.M.; Tait, C.; Schuler, H.; et al. Genomic differentiation during speciation-with-gene-flow: Comparing geographic and host-related variation in divergent life history adaptation in rhauletis pomonella. *Genes* **2018**, *9*, 262. [[CrossRef](#)]
68. Doellman, M.M.; Feder, J.L. Genomic transitions during host race and species formation. *Curr. Opin. Insect Sci.* **2019**, *31*, 84–92. [[CrossRef](#)] [[PubMed](#)]
69. Ye, X.D.; Su, Y.L.; Zhao, Q.Y.; Xia, W.Q.; Liu, S.S.; Wang, X.W. Transcriptomic analyses reveal the adaptive features and biological differences of guts from two invasive whitefly species. *BMC Genom.* **2014**, *15*, 512. [[CrossRef](#)] [[PubMed](#)]
70. Dermauw, W.; Wybouw, N.; Rombauts, S.; Menten, B.; Vontas, J.; Grbić, M.; Clark, R.M.; Feyereisen, R.; Van Leeuwen, T. A link between host plant adaptation and pesticide resistance in the polyphagous spider mite *Tetranychus urticae*. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 113–122. [[CrossRef](#)]
71. Bauer, M.; Katzenberger, J.D.; Hamm, A.C.; Bonaus, M.; Zinke, I.; Jaekel, J.; Pankratz, M.J. Purine and folate metabolism as a potential target of sex-specific nutrient allocation in *Drosophila* and its implication for lifespan-reproduction tradeoff. *Physiol. Genom.* **2006**, *25*, 393–404. [[CrossRef](#)]
72. Majorek, K.A.; Dunin-Horkawicz, S.; Steczkiewicz, K.; Muszewska, A.; Nowotny, M.; Ginalska, K.; Bujnicki, J.M. The RNase H-like superfamily: New members, comparative structural analysis and evolutionary classification. *Nucleic Acids Res.* **2014**, *42*, 4160–4179. [[CrossRef](#)]
73. Richter, G.; Krieger, C.; Volk, R.; Kis, K.; Ritz, H.; Götze, E.; Bacher, A. Biosynthesis of riboflavin: 3,4-dihydroxy-2-butanone-4-phosphate synthase. *Methods Enzymol.* **1997**, *280*, 374–382. [[CrossRef](#)]
74. Kaczanowska, M.; Rydén-Aulin, M. The YrdC protein—A putative ribosome maturation factor. *Biochim. Biophys. Acta-Genet. Struct. Expr.* **2005**, *1727*, 87–96. [[CrossRef](#)]