DNA barcoding for identification and phylogenetic inference of (Diptera; Ceratopogonidae) pollinators of cacao

Ivan Mateo Tamayo Ceballos

Universidad del Rosario
Facultad de Ciencias Naturales
Bogotá, Colombia
2023

DNA barcoding for identification and phylogenetic inference of (Diptera; Ceratopogonidae) pollinators of cacao

IVAN MATEO TAMAYO CEBALLOS

Trabajo de grado presentado como requisito para obtener el título de:

BIÓLOGO

Director

James Edward Richardson, PhD

Co-directora

Adriana Sanchez Andrade, PhD

Tutor

Camilo Andrés Salazar, PhD

Asesora externa

Roxana Yockteng, PhD

Facultad de Ciencias Naturales
Programa de Biología
Universidad del Rosario
Bogotá, Colombia
2023

DNA barcoding for identification and phylogenetic inference of (Diptera; Ceratopogonidae) pollinators of cacao

Ivan M. Tamayo[1], Adriana Sanchez[1], Roxana Yockteng[2], Camilo A. Salazar[1] And James E. Richardson[1,3,4,5]

[1]Departamento de Biología, Facultad de Ciencias Naturales, Universidad del Rosario, Bogotá, Colombia.
[2]Corporación Colombiana de Investigación Agropecuaria, Agrosavia, Bogota, Colombia.
[3]School of Biological, Earth and Environmental Sciences, University College Cork, Cork, Ireland.
[4]Environmental Research Institute, University College Cork, Cork, Ireland.
[5]Tropical Diversity Section, Royal Botanic Garden, Edinburgh, United Kingdom.

Summary

Biodiversity loss is occurring on a large scale and the need to monitor it is becoming more and more necessary. The use of morphological techniques can be enhanced with the use of molecular tools to help resolve diversity and evolutionary history knowledge gaps. The objective of this research was to evaluate the use of the mitochondrial cytochrome *c* oxidase 1 (COI) gene as a DNA barcode marker to characterize possible pollinators of Colombian cacao crops. The study was implemented in the departments of Meta and Northern Santander, Colombia. Taxa sampled directly from flowers were analyzed to assess their diversity and phylogenetic relationships using Maximum-Likelihood (ML) and Bayesian Inference (BI). We generated sequences of approximately 656 bp for 25 Culicomorpha individuals, 13 from Meta and 12 from Northern Santander and downloaded 388 sequences of the Ceratopogonidae and Chironomidae families from GenBank. Analysis of the COI sequences reveals that our sequences were placed in three Ceratopogonidae lineages (*Forcipomyia, Dasyhelea* and *Stilobezzia*) and five unresolved lineages of Chironomidae. We also found that species that visited cacao flowers in Meta and Northern Santander plantations represented two separate guilds, which could have been influenced by the orogenic processes of the Andes Mountains. Phylogenetic reconstruction indicated that most (n=17) of our sequences were resolved in the *Forcipomyia* group. Additionally, none of our sequences were identical to any GenBank sequences, reflecting an investigative bias towards northern temperate regions and the need for more molecular studies on tropical species. Our data offer new DNA sequences of Colombian Ceratopogonidae for the development of a global inventory of pollinating species of plants of economic interest, such as cacao. The need to continue sampling tropical taxa is also suggested in order to clarify the evolutionary history of these families of flies.

Key words: DNA Barcoding, COI, Cacao, Ceratopogonidae, Phylogenies.

Resumen

La pérdida de biodiversidad está ocurriendo a gran escala y la necesidad de monitorearla es cada vez más necesaria. El uso de técnicas morfológicas se puede mejorar con el uso de herramientas moleculares para ayudar a resolver los vacíos en el conocimiento de la diversidad y la historia evolutiva de las especies. El objetivo de esta investigación fue evaluar

el uso del gen mitocondrial citocromo c oxidasa 1 (COI) como marcador de código de barras de ADN para caracterizar mosquitos como posibles polinizadores en cultivos de cacao colombianos. El estudio se implementó en los departamentos del Meta y Norte de Santander. Los taxones muestreados directamente de las flores se analizaron para evaluar su diversidad y relación filogenética utilizando la máxima verosimilitud (ML) y la inferencia bayesiana (BI). Generamos secuencias de aproximadamente 656 pb para 25 individuos Culicomorpha, 13 del Meta y 12 del Norte de Santander y descargamos 388 secuencias de las familias Ceratopogonidae y Chironomidae de GenBank. El análisis de las secuencias COI revela que nuestras secuencias se ubicaron en tres grupos de linajes de Ceratopogonidae (Forcipomyia, Dasyhelea y Stilobezzia) y cinco linajes no resueltos de Chironomidae. También encontramos que las especies que visitaron las flores de cacao en las plantaciones de Meta y Norte de Santander representaban dos grupos separados, que pueden estar influenciados por los procesos orogénicos de las montañas de los Andes. La reconstrucción filogenética indicó que la mayoría (n=17) de nuestras secuencias se resolvieron en el grupo Forcipomyia. Adicionalmente, ninguna de nuestras secuencias fue idéntica a las secuencias de GenBank, lo que refleja un sesgo de investigación para las regiones templadas del norte y la necesidad de más estudios sobre especies tropicales. Nuestros datos ofrecen nuevas secuencias moleculares de Ceratopogonidae colombianos para el desarrollo de un inventario global de especies polinizadoras de plantas de interés económico, como el cacao. Se sugiere también la necesidad de seguir muestreando taxones tropicales para esclarecer la historia evolutiva de estas familias de moscas.

Palabras clave: *Códigos de barras de ADN, COI, Cacao, Ceratopogonidae, Filogenias.*

Introduction

Ecologists and evolutionary biologists require fundamental biological information in order to adequately execute their studies. Shortfalls in various categories have been identified for which knowledge gaps exist (Hortal et al. 2015). These include the Linnean shortfall for species taxonomy, the Wallacean shortfall for species distributions, the Prestonian shortfall for species abundance and the Darwinian shortfall for evolutionary patterns. These problems are particularly acute in groups of organisms for which there are insufficient taxonomic researchers trained to tackle the scale of these shortfalls, such as fungi or insects. Surprisingly, this is even true for groups of organisms that provide important ecosystem services, such as pollinators of crop species. For these groups, a short-term solution may be to provide some form of rapid characterization that may give some indication of the degree of biodiversity that exists for a given area. These issues are particularly severe in areas of high biodiversity such as the New World Tropics. Despite the fact that there are many studies focused on analyzing animal diversity related with pollination syndromes in tropical ecosystems (Dellinger et al., 2021; Kobayashi et al., 2019; Neves et al., 2020), there are still large knowledge gaps of small-sized animals that are involved in those important ecological and evolutionary relationships. Furthermore, these small-size animals are often misidentified, leading to a confusing taxonomic status. The use of DNA barcodes may address some of these shortfalls (Gostel & Kress, 2022; Moritz & Cicero, 2004).

DNA barcoding is a powerful technique for identification and initial delimitation of species using a standardized region of DNA (*DNA Barcoding Program*, 2017; Hebert et al., 2003;

Ratnasingham & Hebert, 2007). This means that a unique region of the DNA can act as a taxonomic identifier. For animals, the cytochrome *c* oxidase I (COI) gene was established as the core of a global bio-identification system (Hebert et al., 2003) that has been used to assess the diversity, especially of Diptera species (Kumar et al., 2007; Nagy et al., 2013; Rivera & Currie, 2009; Sinclair & Gresens, 2008; Stur & Borkent, 2014; Versteirt et al., 2015). Given their use in species identification, DNA barcodes might also be used as a way of initially characterizing diversity. This approach has the potential to address the problem of recognizing complex or poorly studied groups of species in scenarios where the morphological approach alone is not sufficient or too time-consuming. Even if morphological data is still lacking, DNA barcoding allows recognition of unknown taxa (Damm et al., 2010; Kress et al., 2015). The applicability of COI also improves our understanding of the evolutionary relationships through their use in phylogenetic reconstruction. This means that multiple shortfalls could be addressed using DNA barcodes of unknown taxa and this could speed up the inventory of biological diversity. DNA barcoding allows biologists to rapidly characterize biodiversity, as a primary step in filling gaps in the knowledge of any biological group. However, it is important to recognize that barcoding has some inherent limitations (e.g., Moritz & Cicero, 2004).

The cacao tree (*Theobroma cacao* L.) is a species native to the Neotropical rain forest, whose agricultural importance dates back ~1,500 years (Motamayor et al., 2002). Cacao is an economically globally important species and about 5% of all chocolate production comes from Colombia (Gutiérrez García et al., 2020; Osorio-Guarín et al., 2017). The cultivation of *T. cacao* in rural communities of Colombia is a potential substitute for illicit crops and offers the possibility to generate new jobs (Lamos-Díaz et al., 2020; Rodriguez-Medina et al., 2019). However, only 2-5% of cacao flowers develop pods (Bartley, 2005; Branco et al., 2018; Frimpong et al., 2009; Moser, 2010; Vansynghel et al., 2022) and many individuals are self-incompatible. Therefore, successful fruit set is entirely dependent on pollinators (Klein et al., 2007) unless labour intensive hand pollination is undertaken. Despite their importance, little is known about the identity of cacao pollinators within its native range in the Neotropics or in other parts of the tropics in which it is cultivated. Characterization of which species pollinate cacao will allow more in-depth studies of their biology and perhaps lead to trialing new methods for improving pollination success. Mature *T. cacao* trees have a cauliflorous inflorescence specific for some insects (Fig. 1A) (Frimpong-Anin et al., 2015). Midges of the Ceratopogonidae family (known in some regions as "*Jején*") have been documented as an important visitor to cacao flowers and have been recognized as potential pollinators (Erickson et al., 1987; Frimpong et al., 2009; Kaufmann, 1975; Ronderos, 2017; Spinelli & Wolff, 2016; Vansynghel et al., 2022; Winder, 1978).

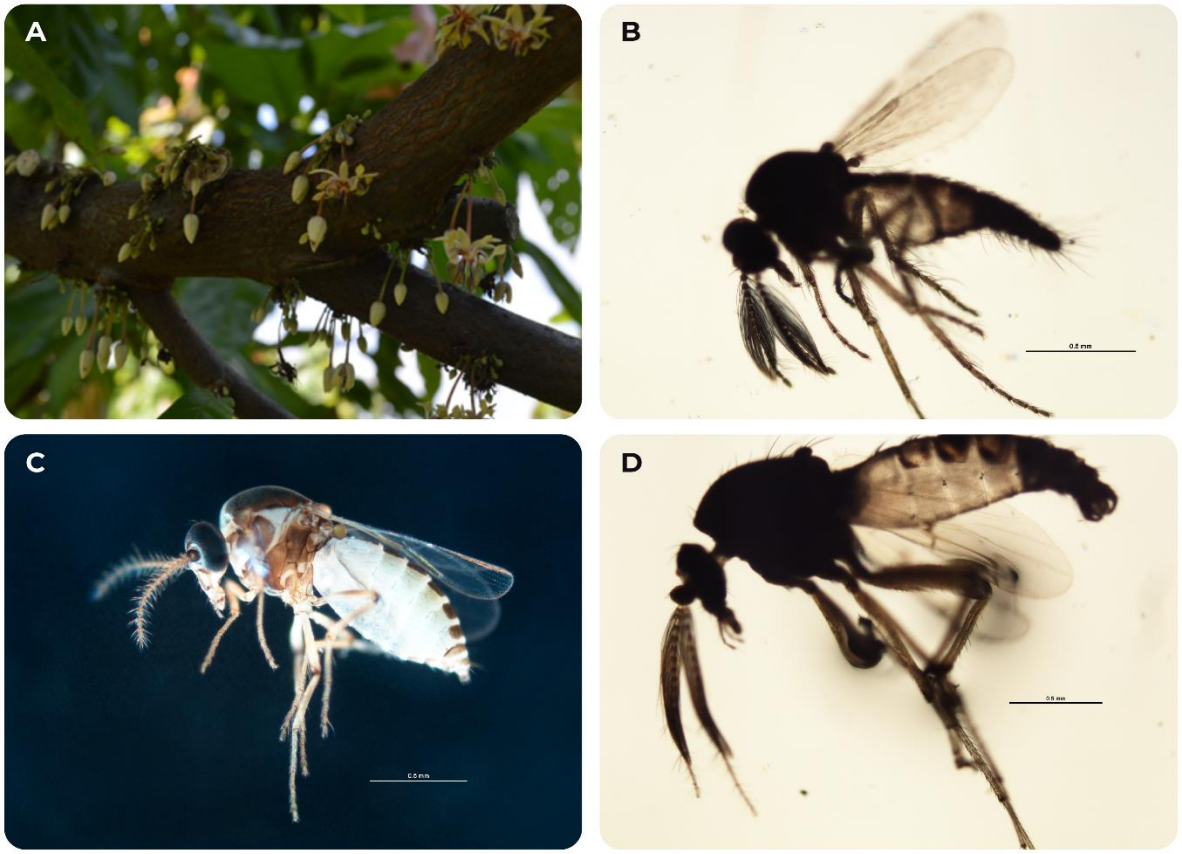

Fig. 1. A. Cauliflorous inflorescence (a cluster of flowers on a branch) of a *Theobroma cacao* tree. B. Adult female of *Forcipomyia sp.*, C. Adult female of *Dasyhelea ludingensis* and D. Adult male of *Stilobezzia diversa* (C1.8, A1.3 and bb2.1 in this study).

The Ceratopogonidae are small nematoceran dipterans (~ 2mm in length; Fig. 1B, 1C and 1D) whose life cycle and development require a particular level of humidity. Immature stages are present in aquatic and subaquatic habitats, while adult stages are common and diverse in humid terrestrial habitats (Borkent, 2001, 2014; Ronderos, 2017). Females of this family are vectors of a wide range of diseases (Carpenter et al., 2013) and are notorious as biting flies. There are ~ 6,206 recognized extant species arranged in 112 genera, belonging to the infraorder Culicomorpha and the superfamily Chironomoidea (Borkent & Dominiak, 2020; Szadziewski, 1988). Previous cladistic (Oosterbroek & Courtney, 1995) and molecular analyses (Beckenbach & Borkent, 2003) support their monophyly, with the family Chironomidae as sister. However, a well resolved phylogeny of Ceratopogonidae has not yet been published. Although this family has a wide distribution and has a large number of species, studies have been focused on northern temperate regions, leaving insufficient knowledge about its diversity in tropical areas. Previous studies of Ceratopogonidae in Colombia have focused on their morphology (La & S, 1973; Soria, 1971; Soria & Wirth, 1979), with about 235 species recognized in 28 genera (Spinelli & Wolff, 2016).

Identifications based on morphological characters are time-consuming and very few taxonomists specialize on the family. To assess the high biodiversity in tropical hotspots in the absence of extensive taxonomic knowledge, we propose an approach to investigate the

diversity of Colombian Ceratopogonidae midges using DNA barcoding. The COI gene has been shown to distinguish between species of Ceratopogonidae from northern temperate regions (Stur & Borkent, 2014), and is thus a useful marker for species identification. However, molecular studies of midges in Colombia are poorly documented with only two records in the Barcode of Life Data System (BOLD) (Ratnasingham & Hebert, 2007) and only one of which has geographic localization. In this context, the sampling of cacao flower visitors in Colombia is a huge task that has implications in conservation management projects and productivity improvement efforts. Most biodiversity hotspots, especially in tropical ecosystems, are unsampled by Ceratopogonidae taxonomists. A huge part of these ecosystems is rapidly disappearing due to anthropic disturbances, making it especially important to assess the biodiversity in these remaining areas (Borkent & Dominiak, 2020). Here, we aim to evaluate the use of the COI gene as a DNA barcode marker to characterize floral visitors to Colombian cacao crops. We also use this marker to determine whether the guild of visitors differs between two important cacao growing regions that are found on either side of the Eastern Cordillera of the Andes.

Materials and methods

Specimen collection

Insects were collected from Colombian cacao farms in the Departments of Meta and Northern Santander (Table 1, Fig. 2), which are validated by The Colombian Corporation for Agricultural Research (Agrosavia, Corpoica) and The National Federation of Cacao Farmers (Fedecacao). The specimens were sampled using a rechargeable suction pump with focal observation to ensure they were all flower visitors. Different cacao genotypes were sampled per farm (Table 1). Insects were sucked from the inflorescence with a pump nozzle into a removable collection tube with a fine net at the bottom to prevent insects from escaping. We sampled for 8 h in the morning (6:00 am – 1:00 pm) and 4 h in the afternoon (2:30 – 5:30 pm) on all farms. Both the collection tube and the nozzle were detached from the main pump and cotton wool soaked in ethyl acetate was placed to kill trapped insects, as suggested by Frimpong et al. (2009). The captured insects were kept in 85% EtOH and transported in falcon tubes to Rosario University to be stored at -80 °C. Insects were separated into orders and families using taxonomic identification keys.

Table 1. Characteristics of the cacao farms.

Farm	Department	Municipality	Coordinates	Elevation (m)	Cacao genotype
La Libertad (Agrosavia)	Meta	Villavicencio	4.057878, -73.467089	341	Tame, Saravena
El Recuerdo	Meta	Granada	3.490361, -73.699944	301	TSH-565, San Vicente-41
El Palmar	Meta	Granada	3.429578, -73.745194	322	Saravena
La Ponderosa	Meta	Acacías	3.960806, -73.762306	533	Tame, Saravena
Mata de Cacao	N. Santander	El Zulia	8.018806, -72.662.833	339	TSH-565xIMC-67*

El Hoyo	N. Santander	El Zulia	8.045333, -72.686528	1014	CCN-51, ICS-95
Ins. Jorge Gaitán Durán	N. Santander	Cúcuta	7.896389, -72.508917	320	Saravena
Sucesión	N. Santander	Cúcuta	8.230722, -72.415361	80	CCN-51
Parcela #2	N. Santander	Cúcuta	8.237472, -72.403361	76	CCN-51
Las Flores	N. Santander	Cúcuta	8.251861, -72.494722	71	CCN-51, Caucasia-37
Andalucía	N. Santander	Sardinata	8.104278, -72.797417	271	CCN-51

***** *Corresponds to a hybrid genotype.*

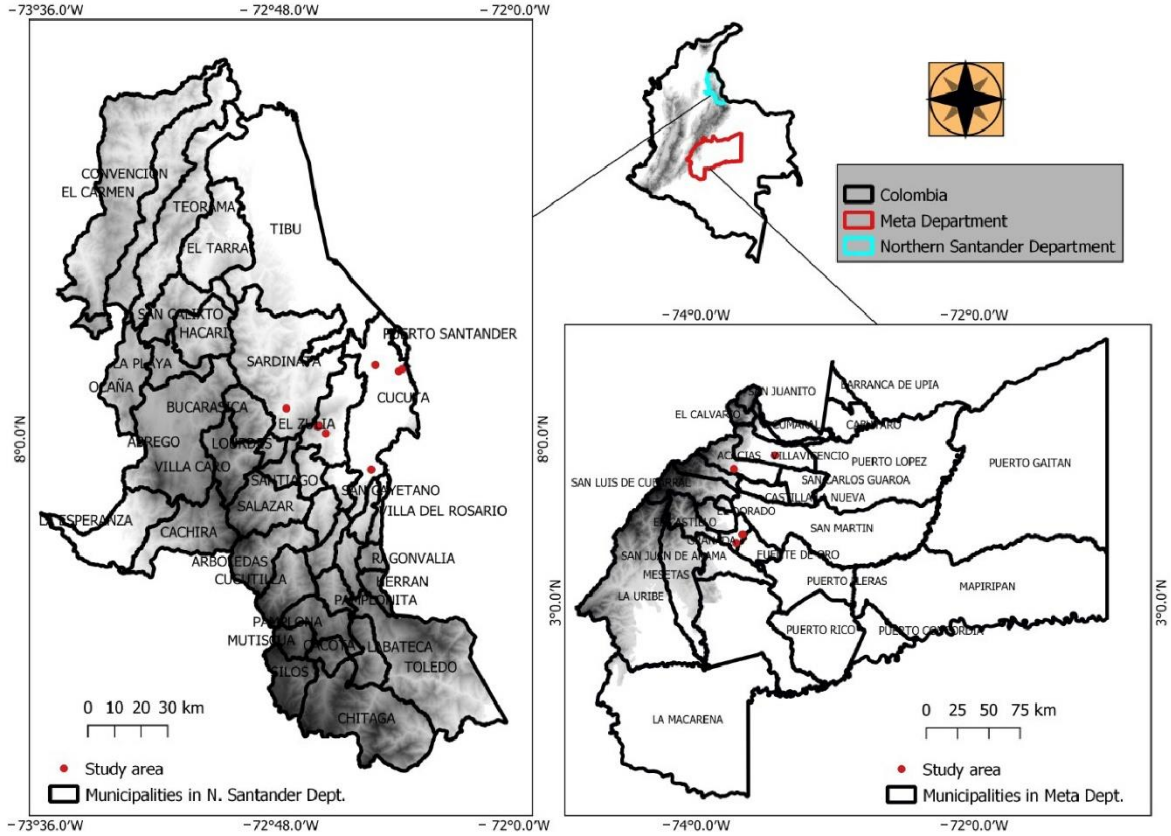

Fig. 2. Geographic location of the sampling areas of cacao crops.

DNA extraction and COI Barcoding

Total DNA was extracted from each individual specimen using the Qiagen Blood and Tissue Kit molecular DNA extraction protocol (Qiagen, Venlo, Netherlands). In some individuals we used a destructive technique, the TissueLyser II (Qiagen) while in others we used a non-destructive technique to preserve the specimens for future analysis. The extracted DNA was quantified using a NanoDrop 2000 (ThermoFisher Scientific, Waltham, MA, USA). Amplification of a ~658 bp fragment from the 5' region of the barcode region of the mitochondrial COI gene was performed by polymerase chain reaction (PCR). The DNA barcode region was targeted using the universal primers LCO1490 (forward) and HCO2198

(reverse) (Folmer et al., 1994). Two microliters of template DNA were added to a 10 µl of reaction mix, containing 6.5 µl of Taq Master Mix (5 U/µl, Qiagen), 1 µl of molecular water and 0.5 of each 10 µM primers. The PCR protocol was standardized following the methodology of Tomazatos et al. (2020), modifying the annealing temperature to 50 °C and 55 °C. All amplicons were visualized through electrophoresis on 1% agarose gel run for 45 min. Successful amplification of the barcode was indicated by the presence of a band compared with the Quantitative DNA Ladder (100-2000 bp; ThermoFisher Scientific, US). The PCR final products with 25 µl were purified by washing with ExoSAP-IT (ThermoFisher Scientific) and subsequently sent to MacroGen (South Korea) to be sequenced. The resulting electropherograms were edited in Mesquite v.3.7 (Maddison & Maddison, 2021) and the forward and reverse sequences assembled into a consensus using the bioinformatics software Geneious v.5.6 (Kearse et al., 2012).

Phylogenetic inference

Contigs were compared to sequences previously published in GenBank using the Basic Local Alignment Search Tool (BLAST) (Madden, 2003) to calculate sequence similarity. Sequences of Ceratopogonidae in GenBank were included in the phylogenetic analysis covering the species reported in this database. The matrix was built by adding sequences on the Taxonomy browser of GenBank that contain all the subfamilies of Ceratopogonidae. For the genus *Culicoides* (punkies) we added one species for each subgenus to our matrix while for the other genera, we added all the species of each subgenus available in the database. We also added the BOLD sequences of Ceratopogonidae from Colombia and sequences from the sister family Chironomidae. The sequences included in this analysis (n = 388) are listed in the supplementary material (Table S1). All sequences were aligned using MAFFT on the command line (Katoh & Standley, 2013) (*mafft –adjustdirection –globalpair –maxiterate 100000 culicomorpha.txt culicomorpha_MAFFT.nex*). The multiple alignment was then manually edited and displayed in Mesquite v.3.7.

The general time reversible model with unequal rates and unequal base frequencies plus empirical base frequencies and invariable site, plus discrete gamma model for rate heterogeneity across sites (GTR+F+I+G4), was identified as the most appropriate of molecular evolution using the IQ-TREE model search algorithm (Minh et al., 2020; Nguyen et al., 2015). The nucleotide substitution model in the alignment was based on the Akaike Information Criterion (AIC), Corrected Akaike information criterion (AICc) and Bayesian Information Criterion (BIC) scores.

To assess the phylogenetic relationships among our unknown Colombian taxa and all of the previously reported species, a Maximum Likelihood tree (ML) was constructed with IQ-TREE (Nguyen et al., 2015) using as outgroups the families Simuliidae and Thaumaleidae (Table S1). Branch support was tested using approximate likelihood ratio test (SH-aLRT) (Guindon et al., 2010), local bootstrap support (Adachi & Hasegawa, 1996), approximate Bayes test (Anisimova et al., 2011) and UF bootstrap (Hoang et al., 2018) with 100,000 replicates for each. We also used a Bayesian Inference (BI) approach to build a consensus tree

and Posterior Probability (PP) for branch support, using MrBayes v.3.2.6 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003). For the BI, fifty million generations of MCMC trees were run in four chains, one cold and three heated, samples were taken every 1000^{th} with burn-in of 25% to inform the Bayesian posterior probabilities. Each run was examined using Tracer v.1.7.2 (Rambaut et al., 2018) to visualize the sampled parameters values and determine if there was convergence between the chains. Identification of highly supported clades on the phylogeny (i.e., UF bootstrap and Bayesian Posterior Probabilities [PP] ≥ 95%) allows recognition of monophyletic groups.

Results

We generated COI gene data for 20 Ceratopogonidae specimens; 12 from Meta Department (A1.3, A1.4, A1.5, A2.3, C1.1, C1.4, C1.5, C1.6, C1.8, C1.9, C2.4 and D1.2). and eight from Northern Santander Department (aa3.2, aa3.3, aa3.4, aa4.3, bb2.1, bb3.8, cc2.3, dd1.2). Interestingly, we also identified and generated the barcode for five specimens of the sister family Chironomidae of which one is from Meta (C1.2) and four are from Northern Santander (bb3.5, bb3.7, bb3.10 and cc1.4). All barcodes generated a sequence coverage of 656 bp. Overall, the multiple alignment (n = 413 sequences, Table S1) had 271 invariable sites, 332 parsimony informative sites and 53 singleton sites. The average percentage of A, C, G and T in the matrix was 27.44%, 17.92%, 16.33% and 38.31% respectively.

Using the ML approach based on the best substitution model with 100,000 replicates of UF bootstrap, indicates that the family Ceratopogonidae forms a non-monophyletic group as *Jenkinshelea* sp. grouped with the sister family (see Fig. S1 for all branch supports, with the families Simuliidae and Thaumaleidae as outgroups). In the Ceratopogonidae phylogeny, our Colombian sequences were grouped in 10 different clades with high node support for most of them (UF bootstrap ≥ 95%, Fig. 3). In the Chironomidae phylogeny, the Colombian sequences were grouped with low support into different clades (Fig. S1). There are three main lineages identified in the Ceratopogonidae phylogeny that include Colombian taxa (*Dasyhelea*, *Forcipomyia* and *Stilobezzia*, Figs. 3, 4). The ML tree for those lineages reconstructs the *Dasyhelea* clade as monophyletic and both *Forcipomyia* and *Stilobezzia* as paraphyletic. The UF bootstrap indicated that the A1.3 sequence shares the same ancestor as *Dasyhelea ludingensis* with 99% support (Fig 3). Even though *Forcipomyia* corresponds to the most diverse group and includes a large number of the Colombian sequences, there are uncertainties as to which could be the lineage that shares the most recent common ancestor for most of them. For instance, the sequences A2.3, C1.1, C1.5, aa4.3; C1.6, bb3.2; and C1.8, bb3.8 are grouped in strong clades, with 95-100% support, but there is uncertainty about which *Forcipomyia* species are more closely related (Fig. 3). Additionally, the ML tree indicated that the A1.4 and bb2.1 sequences share the same ancestor as *Stilobezzia diversa*, with 97% UF bootstrap support.

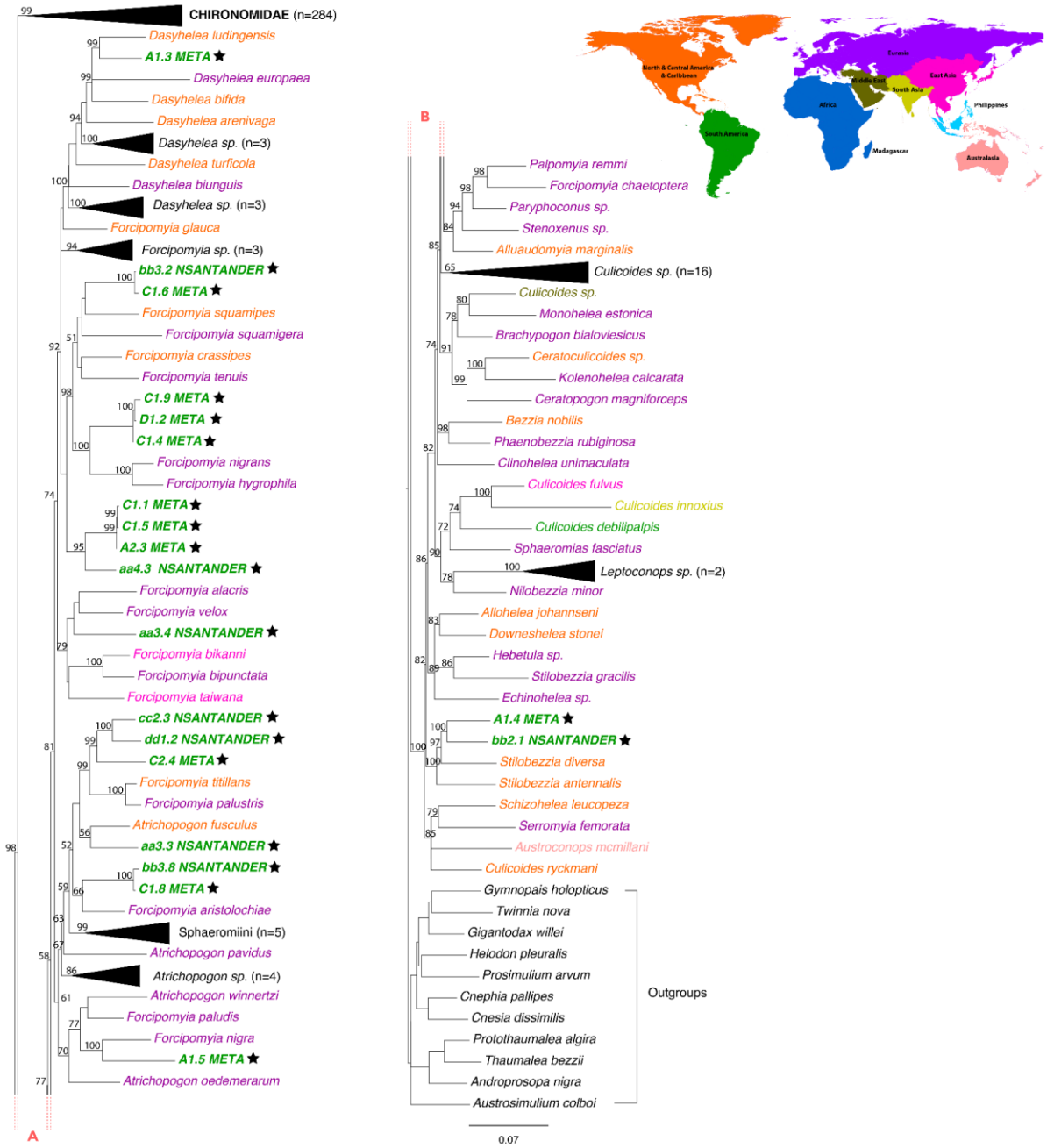

Fig. 3. Maximum likelihood (ML) tree based on 413 COI sequences of the Ceratopogonidae and its sister family species, including the sequences of the Colombian taxa. Numbers above branches represent UF bootstrap support values (only values ≥ 50% are shown) based on 100,000 replicates. The colors match the map above which represent the geographic location of those sequences. Stars represent new sequences for Colombian Ceratopogonidae. Species of the families Simullidae and Thaumaleidae represent the outgroups.

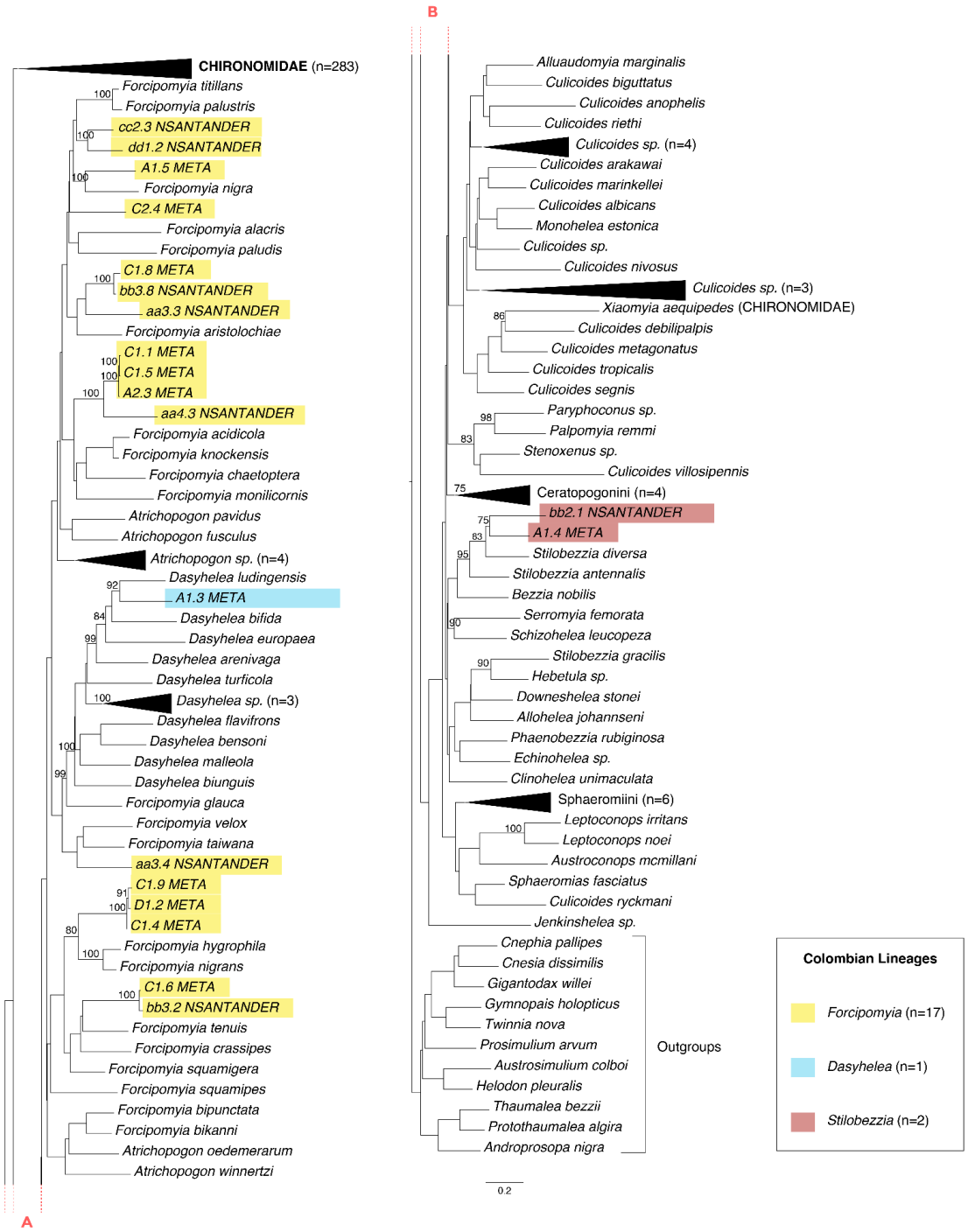

Fig. 4. Bayesian consensus tree based on 413 COI sequences of the Ceratopogonidae and its sister family species, including the sequences of the Colombian taxa. Numbers above the nodes represent Bayesian posterior probability (PP) (only values ≥ 75%). The families Simullidae and Thaumaleidae correspond to the outgroups.

In the Bayesian Inference (BI), the Effective Sampling Size (ESS) values for all parameters were above 200. The consensus tree (Fig. S2 for all branch supports) indicates that the Chironomidae family forms a paraphyletic group with one species more closely related to *Culicoides* (Ceratopogonideae). The Ceratopogonidae family forms a monophyletic group (PP

= 0.90). Overall, the tree based on BI produced a similar topology with the ML tree. There is a well-supported (PP = 0.92) sister group relationship between the sequence A1.3 and *Dasyhelea ludingensis* (Fig. 4). There are also uncertainties about the most recent common ancestor for most *Forcipomyia* lineages as with the ML tree. The BI tree reconstructs well-supported internal nodes (PP = 0.95) in the *Stilobezzia* clade that contain Colombian sequences. Nevertheless, there are two sequences (aa3.3. and aa3.4) that did not group clearly with any lineage in both phylogenetic trees (Figs. 3, 4).

Lastly, the sequences bb3.5, bb3.7, bb3.10 from Northern Santander were resolved in the same clade in both analyzes (Figs. S1, S2). However, even though they are mixed in the *Tanytarsus* clade of the Chironomidae family, there is no clear support to a sister group. Likewise, the sequences cc1.4 from Northern Santander and C1.2 from Meta are not clearly placed in the phylogeny (Figs. S1, S2).

Discussion

We found that there are three main groups of Ceratopogonidae lineages that visit cacao flowers in our study sites (Figs. 3, 4). We also found that the guild of visitors in Meta and Northern Santander were different, indicating that cacao likely relies on different pollinators in different parts of its native and cultivated range. As expected, we found that the taxa we sampled from cacao flowers fall in lineages of *Forcipomyia*, but we also found associations with species in *Dasyhelea* and *Stilobezzia*. For example, the ML and BI phylogenies place the Colombian sequence A1.3, with high support, as sister to *Dasyhelea ludingensis*. It should be noted that in our phylogeny *Dasyhelea* is nested within a paraphyletic *Forcipomyia* and thus might ultimately be considered as congeneric with the latter. The sequences of the Colombian taxa A1.4 and bb2.1 were grouped together with *Stilobezzia diversa*. Given that these two species are found in northern temperate regions, and that many of our sequences are sisters to species from those regions, it is this likely that our phylogenies reflect the lack of sampling in tropical regions.

Naturally, the distribution of many organisms depends, in part, on its biogeographic history. There is evidence of ancient Ceratopogonidae taxa from Lebanese amber (Borkent, 2001; Choufani et al., 2015; Pielowska-Ceranowska et al., 2022; Szadziewski, 2008; Szadziewski & Arillo, 2003) that suggest a Laurasian origin. Indeed, many relict genera of this family (i.e., *Astroconops, Leptoconops, Metahelea*) were found from fossil records in the Northern Hemisphere (Szadziewski, 2008). Furthermore, the phylogenetic relationships shown here indicates possible long-distance transoceanic dispersal, post-vicariance. Beckenbach and Borkent (2003) estimate the age of *Forcipomyia*, a genus with a near cosmopolitan distribution, to have diverged from its nearest ancestor ca. 65 million years ago, long after many continental land masses had diverged through continental drift. Molecular evidence (Krosch et al., 2011; Page et al., 2005; Rowe et al., 2010) also suggests that post-vicariance radiation and Eocene dispersal have been important in driving diversification of Gondwana taxa, justifying the great diversity of Neotropical groups. Studies have also shown that some organisms can disperse, in the form of “aeroplankton”, among separated continents and islands by active or passive dispersal through wind currents (Bourguignon et al., 2018;

Freeman, 1945; Holzapfel & Harrell, 1968). This phenomenon was observed in some Ceratopogonidae groups (i.e., *Forcipomyia* and *Dasyhelea*) and could explain the ability to disperse over greater distances (Borkent, 1991; Eagles et al., 2012, 2014; Peck, 1994; Szadziewski, 2008). However, the assumption of a transoceanic dispersion for most of the Colombian lineages needs to be tested.

Although these organisms may disperse across great masses of water, little is known about their ability to disperse over mountains at great elevations. In this study, we sampled cacao crops in two departments of Colombia, separated by the Andes Mountains. Our results suggest that species that visited cacao flowers in Meta and Northern Santander represented two separate guilds (Figs. 3, 4). The Andes uplift may have driven speciation events of some taxa in the past, resulting in the genetic differences that are accumulating between each other. Nevertheless, it is possible that those “aeroplankton " taxa are mixing between these two areas separated by the Andes and that is the reason why some lineages group together. This issue needs to be clarified in further research and with additional molecular markers.

Our analyses indicate that many genera are not monophyletic, suggesting that much systematics research needs to be done to better resolve relationships, define taxa and understand patterns of evolution. The precise relationships of most of the Colombian taxa, despite being resolved in *Forcipomyia* lineages, are doubtful. These uncertainties reflect the low sample coverage of Ceratopogonidae taxa in tropical regions, especially in Colombian ecosystems. The patterns in the relationships of Colombian taxa and their high degree of genetic differentiation from other sampled taxa (evident in the branch lengths; Figs. 3, 4), suggest the possible occurrence of species unknown to science. A great example is the sequence aa3.3. and the sequence aa3.4 that did not group closely with any lineage in both phylogenies.

Previous studies have mentioned that individuals of the genus *Forcipomyia* act as pollinators of *Theobroma cacao* (Erickson et al., 1987; Frimpong et al., 2009; Kaufmann, 1975; Ronderos, 2017; Spinelli & Wolff, 2016; Vansynghel et al., 2022; Winder, 1978). Our results are consistent with those reports. However, all previous research on these pollinators has focused on ecological and morphological descriptions. Here we present an evolutionary perspective on unknown sequences of Ceratopogonidae taxa associated with cacao flowers. Additionally, some of the visitors (20% of our samples) are representatives of the family Chironomidae, which is sister to Ceratopogonidae. Interestingly, sequences in the Chironomidae are not clearly placed (low branch support; Figs S1, S2). This could also be due to lack of a more comprehensive sampling in the tropical regions.

In this study, we used DNA barcoding and phylogenetic inference to assess shortfalls in biodiversity data. We generated the first DNA sequences of Ceratopogonidae and Chironomidae species involved in interactions with cacao flowers in Colombian crops. Although this group of organisms is poorly studied in the tropics, there have been studies on their taxonomy and species have been described based on their morphology (Borkent and Spinelli, 2007). However, there are not enough trained taxonomists to identify species using morphological characters. One possible solution would be to generate DNA barcode reference

libraries for species that have been described, using DNA from the type specimens of those species. This approach has been advocated in plants that are difficult to identify because the fertile material that is required for identification is not available (Sánchez-C. et al., 2022). A DNA barcode reference library with sequence data generated from type specimens would allow us to match the barcode sequences of unidentified material with the type specimen's sequence and thus the name of the species. It would also highlight sequences that do not match with any in the reference library and may therefore represent new species that require further morphological investigation to describe them formally.

We highlight the usefulness of DNA barcoding in contributing, in a rapid way, to characterize the biological diversity of groups that are difficult to identify. Once we discover, recognize, and delineate the species, clarifying their evolutionary relationships (i.e., addressing the Darwinian Shortfall) might also contribute to improving our understanding of their biology and ecology. In this way we can facilitate the development of a global inventory of pollinator species for economically important crops, such as cacao, which might lead to an improved understanding of their biology and better pollination success.

Acknowledgments

We thank the Alejandro Angel Escobar Foundation (*FAAE*) *"Colombia Biodiversa"* program who financed the project. Additionally, we thank The Colombian Corporation for Agricultural Research (Agrosavia, Corpoica) for the materials for the midge samples and The National Federation of Cacao Farmers (Fedecacao) for granting permission to visit plantations. We are also grateful to the Faculty of Natural Sciences at the Rosario University, for academic and financial support.

References

Adachi, J., & Hasegawa, M. (1996). Model of amino acid substitution in proteins encoded by mitochondrial DNA. *Journal of Molecular Evolution*, *42*(4), 459-468. https://doi.org/10.1007/BF02498640

Anisimova, M., Gil, M., Dufayard, J.-F., Dessimoz, C. & Gascuel, O. (2011). Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Systematic Biology*, *60*(5), 685-699. https://doi.org/10.1093/sysbio/syr041

Bartley, B. G. D. (Ed.). (2005). *The genetic diversity of cacao and its utilization*. CABI. https://doi.org/10.1079/9780851996196.0000

Beckenbach, A. T., & Borkent, A. (2003). Molecular analysis of the biting midges (Diptera: Ceratopogonidae), based on mitochondrial cytochrome oxidase subunit 2. *Molecular Phylogenetics and Evolution*, *27*(1), 21-35. https://doi.org/10.1016/s1055-7903(02)00395-0

Borkent, A. (1991). The Ceratopogonidae (Diptera) of the Galápagos Islands, Ecuador with a discussion of their phylogenetic relationships and zoogeographic origins. *Insect Systematics & Evolution*, *22*(1), 97-122. https://doi.org/10.1163/187631291X00336

Borkent, A. (2001). Leptoconops (Diptera: Ceratopogonidae), the Earliest Extant Lineage of Biting Midge, Discovered in 120–122 Million-Year-Old Lebanese Amber. *American Museum Novitates*, *2001*(3328), 1-12. https://doi.org/10.1206/0003-0082(2001)328<0001:LDCTEE>2.0.CO;2

Borkent, A. (2014). The pupae of the biting midges of the world (Diptera: Ceratopogonidae), with a generic key and analysis of the phylogenetic relationships between genera. *Zootaxa*, *3879*, 1-327. https://doi.org/10.11646/zootaxa.3879.1.1

Borkent, A. & G. R. Spinelli. (2007). Neotropical Ceratopogonidae (Diptera: Insecta). In: Adis, J., J. R. Arias, G. Rueda-Delgado & K. M. Wattzen (Eds): Aquatic Biodiversity in Latin America (ABLA). Vol 4. Pensoft Publishers, Sofia-Moscow, 198 pp. *Revista de La Sociedad Entomológica Argentina*, *67*(1-2), Art. 1-2. https://www.biotaxa.org/RSEA/article/view/31038

Borkent, A., & Dominiak, P. (2020). Catalog of the Biting Midges of the World (Diptera: Ceratopogonidae). *Zootaxa*, *4787*(1), zootaxa.4787.1.1. https://doi.org/10.11646/zootaxa.4787.1.1

Bourguignon, T., Tang, Q., Ho, S. Y. W., Juna, F., Wang, Z., Arab, D. A., Cameron, S. L., Walker, J., Rentz, D., Evans, T. A., & Lo, N. (2018). Transoceanic Dispersal and Plate Tectonics Shaped Global Cockroach Distributions: Evidence from Mitochondrial Phylogenomics. *Molecular Biology and Evolution*, *35*(4), 970-983. https://doi.org/10.1093/molbev/msy013

Branco, S. M. de J., Silva, D. V. da, Lopes, U. V. & Corrêa, R. X. (2018). Characterization of the Sexual Self- and Cross-Compatibility in Genotypes of Cacao. *American Journal of Plant Sciences*, *9*(9), Art. 9. https://doi.org/10.4236/ajps.2018.99131

Bridgemohan, P., Singh, K., Cazoe, E., Perry, G., Mohamed, A., & Bridgemohan, R. S. (2017). Cacao floral phenology and pollination: Implications for productivity in Caribbean Islands. *Journal of Plant Breeding and Crop Science*, *9*(7), 106-117. https://doi.org/10.5897/JPBCS2016.0598

Carpenter, S., Groschup, M. H., Garros, C., Felippe-Bauer, M. L., & Purse, B. V. (2013). Culicoides biting midges, arboviruses and public health in Europe. *Antiviral Research*, *100*(1), 102-113. https://doi.org/10.1016/j.antiviral.2013.07.020

Choufani, J., El-Halabi, W., Azar, D., & Nel, A. (2015). First fossil insect from Lower Cretaceous Lebanese amber in Syria (Diptera: Ceratopogonidae). *Cretaceous Research*, *54*, 106-116. https://doi.org/10.1016/j.cretres.2014.12.006

Damm, S., Schierwater, B. & Hadrys, H. (2010). An integrative approach to species discovery in odonates: From character-based DNA barcoding to ecology. *Molecular Ecology*, *19*(18), 3881-3893. https://doi.org/10.1111/j.1365-294X.2010.04720.x

DNA Barcoding Program. (2017, agosto 22). ForestGEO. https://forestgeo.si.edu/research-programs/dna-barcoding-program

Dellinger, A. S., Pérez-Barrales, R., Michelangeli, F. A., Penneys, D. S., Fernández-Fernández, D. M., & Schönenberger, J. (2021). Low bee visitation rates explain pollinator shifts to vertebrates in tropical mountains. *New Phytologist*, *231*(2), 864-877. https://doi.org/10.1111/nph.17390

Eagles, D., Deveson, T., Walker, P. J., Zalucki, M. P., & Durr, P. (2012). Evaluation of long-distance dispersal of Culicoides midges into northern Australia using a migration model. *Medical and Veterinary Entomology*, *26*(3), 334-340. https://doi.org/10.1111/j.1365-2915.2011.01005.x

Eagles, D., Melville, L., Weir, R., Davis, S., Bellis, G., Zalucki, M. P., Walker, P. J., & Durr, P. A. (2014). Long-distance aerial dispersal modelling of Culicoides biting midges: Case studies of incursions into Australia. *BMC Veterinary Research*, *10*(1), 135. https://doi.org/10.1186/1746-6148-10-135

Erickson, B. J., Young, A. M., Strand, M. A., & Erickson, E. H. (1987). Pollination biology of *Theobroma* and *Herrania* (Sterculiaceae). *International Journal of Tropical Insect Science*, *8*(3), 301-310. https://doi.org/10.1017/S1742758400005282

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, *3*(5), 294-299.

Freeman, J. A. (1945). Studies in the Distribution of Insects by Aerial Currents. *Journal of Animal Ecology*, *14*(2), 128-154. https://doi.org/10.2307/1389

Frimpong, E. A., Gordon, I., Kwapong, P. K., & Gemmill-Herren, B. (2009). Dynamics of cacao pollination: Tools and applications for surveying and monitoring cacao pollinators. *International Journal of Tropical Insect Science*, *29*(2), 62-69. https://doi.org/10.1017/S1742758409990117

Frimpong-Anin, K., Bosu, P., Adjaloo, M., Braimah, H. & Oduro, W. (2015). *Some Facts About Cacao Pollination*. Global Pollination Project-Ghana. ISBN: 9964-3-9352-0

Gostel, M. R., & Kress, W. J. (2022). The Expanding Role of DNA Barcodes: Indispensable Tools for Ecology, Evolution, and Conservation. *Diversity*, *14*(3), 213. https://doi.org/10.3390/d14030213

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the

performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307-321. https://doi.org/10.1093/sysbio/syq010

Gutiérrez García, G. A., Gutiérrez-Montes, I., Hernández Núñez, H. E., Suárez Salazar, J. C., & Casanoves, F. (2020). Relevance of local knowledge in decision-making and rural innovation: A methodological proposal for leveraging participation of Colombian cacao producers. *Journal of Rural Studies*, *75*, 119-124. https://doi.org/10.1016/j.jrurstud.2020.01.012

Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1512), 313-321. https://doi.org/10.1098/rspb.2002.2218

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, *35*(2), 518-522. https://doi.org/10.1093/molbev/msx281

Holzapfel, E., & Harrell, J. (1968). *Transoceanic dispersal studies of insects.* https://www.semanticscholar.org/paper/TRANSOCEANIC-DISPERSAL-STUDIES-OF-INSECTS-Holzapfel-Harrell/5e5d6cffb83ae68da4ac989c56eb6684d6c001be

Hortal, J., de Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M. & Ladle, R. J. (2015). Seven shortfalls that beset large-scale knowledge of biodiversity. *Annual Review of Ecology, Evolution, and Systematics*, *46*(1), 523-549. https://doi.org/10.1146/annurev-ecolsys-112414-054400

Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics (Oxford, England)*, *17*(8), 754-755. https://doi.org/10.1093/bioinformatics/17.8.754

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772-780. https://doi.org/10.1093/molbev/mst010

Kaufmann, T. (1975). Studies on the ecology and biology of a cacao pollinator, *Forcipomyia squamipennis* I. & M. (Diptera, Ceratopogonidae) in Ghana. *Bulletin of Entomological Research*, *65*(2), 263-268. https://doi.org/10.1017/S0007485300005940

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647-1649. https://doi.org/10.1093/bioinformatics/bts199

Klein, A.-M., Vaissière, B. E., Cane, J. H., Steffan-Dewenter, I., Cunningham, S. A., Kremen, C., & Tscharntke, T. (2007). Importance of pollinators in changing landscapes for world

crops. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1608), 303-313. https://doi.org/10.1098/rspb.2006.3721

Kobayashi, S., Denda, T., Placksanoi, J., Waengsothorn, S., Aryuthaka, C., Panha, S., & Izawa, M. (2019). The pollination system of the widely distributed mammal-pollinated *Mucuna macrocarpa* (Fabaceae) in the tropics. *Ecology and Evolution*, *9*(11), 6276-6286. https://doi.org/10.1002/ece3.5201

Kress, W. J., García-Robledo, C., Uriarte, M., & Erickson, D. L. (2015). DNA barcodes for ecology, evolution, and conservation. *Trends in Ecology & Evolution*, *30*(1), 25-35. https://doi.org/10.1016/j.tree.2014.10.008

Krosch, M. N., Baker, A. M., Mather, P. B., & Cranston, P. S. (2011). Systematics and biogeography of the Gondwanan Orthocladiinae (Diptera: Chironomidae). *Molecular Phylogenetics and Evolution*, *59*(2), 458-468. https://doi.org/10.1016/j.ympev.2011.03.003

Kumar, N. P., Rajavel, A. R., Natarajan, R., & Jambulingam, P. (2007). DNA Barcodes Can Distinguish Species of Indian Mosquitoes (Diptera: Culicidae). *Journal of Medical Entomology*, *44*(1), 01-07. https://doi.org/10.1093/jmedent/41.5.01

La, C. L. J. de, & S, S. V. (1973). Study on the fluctuation of cacao pollination by midges *Forcipomyia* sp. (Diptera, Ceratopogonidae), in Palmira, Valle, Colombia. [Spanish]. *Acta Agronomica*. https://scholar.google.com/scholar_lookup?title=%5BStudy+on+the+fluctuation+of+cacao+pollination+by+midges+Forcipomyia+sp.+%28Diptera%2C+Ceratopogonidae%29%2C+in+Palmira%2C+Valle%2C+Colombia%5D.+%5BSpanish%5D&author=Cruz+L+J.+de+la&publication_year=1973

Lamos-Díaz, H., Puentes-Garzón, D. E. & Zarate-Caicedo, D. A. (2020). Comparison between machine learning models for yield forecast in cacao crops in Santander, Colombia. *Revista Facultad de Ingeniería*, *29*(54), Art. 54. https://doi.org/10.19053/01211129.v29.n54.2020.10853

Madden, T. (2003). The BLAST Sequence Analysis Tool. En *The NCBI Handbook [Internet]*. National Center for Biotechnology Information (US). https://www.ncbi.nlm.nih.gov/books/NBK21097/

Maddison, W., & Maddison, D. (2009). MESQUITE: A modular system for evolutionary analysis. 11.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, *37*(5), 1530-1534. https://doi.org/10.1093/molbev/msaa015

Moritz, C., & Cicero, C. (2004). DNA Barcoding: Promise and Pitfalls. *PLoS Biology*, *2*(10), e354. https://doi.org/10.1371/journal.pbio.0020354

Moser, G. (2010). Experimental evidence for stronger cacao yield limitation by pollination than by plant resources. *Perspectives in Plant Ecology Evolution and Systematics*. https://www.academia.edu/5105088/Experimental_evidence_for_stronger_cacao_yield_limitation_by_pollination_than_by_plant_resources

Motamayor, J. C., Risterucci, A. M., Lopez, P. A., Ortiz, C. F., Moreno, A., & Lanaud, C. (2002). Cacao domestication I: The origin of the cacao cultivated by the Mayas. *Heredity*, *89*(5), Art. 5. https://doi.org/10.1038/sj.hdy.6800156

Nagy, Z. T., Sonet, G., Mortelmans, J., Vandewynkel, C. & Grootaert, P. (2013). Using DNA barcodes for assessing diversity in the family Hybotidae (Diptera, Empidoidea). *ZooKeys*, *365*, 263-278. https://doi.org/10.3897/zookeys.365.6070

Neves, B., Kessous, I. M., Moura, R. L., Couto, D. R., Zanella, C. M., Antonelli, A., Bacon, C. D., Salgueiro, F., & Costa, A. F. (2020). Pollinators drive floral evolution in an Atlantic Forest genus. *AoB PLANTS*, *12*(5), plaa046. https://doi.org/10.1093/aobpla/plaa046

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268-274. https://doi.org/10.1093/molbev/msu300

Oosterbroek, P., & Courtney, G. (1995). Phylogeny of the nematocerous families of Diptera (Insecta). *Zoological Journal of the Linnean Society*, *115*(3), 267-311. https://doi.org/10.1006/zjls.1995.0080

Osorio-Guarín, J. A., Berdugo-Cely, J., Coronado, R. A., Zapata, Y. P., Quintero, C., Gallego-Sánchez, G., & Yockteng, R. (2017). Colombia a source of cacao genetic diversity as revealed by the population structure analysis of germplasm bank of *Theobroma cacao* L. *Frontiers in Plant Science*, *8*. https://www.frontiersin.org/article/10.3389/fpls.2017.01994

Page, T. J., Baker, A. M., Cook, B. D., & Hughes, J. M. (2005). Historical transoceanic dispersal of a freshwater shrimp: The colonization of the South Pacific by the genus *Paratya* (Atyidae). *Journal of Biogeography*, *32*(4), 581-593. https://doi.org/10.1111/j.1365-2699.2004.01226.x

Peck, S. B. (1994). Aerial Dispersal of Insects Between and to Islands in the Galápagos Archipelago, Ecuador. *Annals of the Entomological Society of America*, *87*(2), 218-224. https://doi.org/10.1093/aesa/87.2.218

Pielowska-Ceranowska, A., Azar, D., & Szwedo, J. (2022). New fossil biting midge from the Lower Cretaceous amber of Lebanon (Diptera: Ceratopogonidae). *Cretaceous Research*, *129*, 105004. https://doi.org/10.1016/j.cretres.2021.105004

Ratnasingham, S., & Hebert, P. D. N. (2007). bold: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*, *7*(3), 355-364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Rivera, J., & Currie, D. C. (2009). Identification of Nearctic black flies using DNA barcodes (Diptera: Simuliidae). *Molecular Ecology Resources*, *9*(s1), 224-236. https://doi.org/10.1111/j.1755-0998.2009.02648.x

Rodriguez-Medina, C., Arana, A. C., Sounigo, O., Argout, X., Alvarado, G. A., & Yockteng, R. (2019). Cacao breeding in Colombia, past, present and future. *Breeding Science*, *69*(3), 373-382. https://doi.org/10.1270/jsbbs.19011

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics (Oxford, England)*, *19*(12), 1572-1574. https://doi.org/10.1093/bioinformatics/btg180

Rowe, D. L., Dunn, K. A., Adkins, R. M., & Honeycutt, R. L. (2010). Molecular clocks keep dispersal hypothesis afloat: Evidence for trans-Atlantic rafting by rodents. *Journal of Biogeography*, *37*(2), 305-324. https://doi.org/10.1111/j.1365-2699.2009.02190.x

Sánchez-C., D., Richardson, J. E., Hart, M., Serrano, J., Cárdenas, D., Gonzalez, M. A., & Cortés-B., R. (2022). A plea to DNA barcode type specimens: An example from Micropholis (Sapotaceae). *TAXON*, *71*(1), 154-167. https://doi.org/10.1002/tax.12652

Sinclair, C. S., & Gresens, S. E. (2008). Discrimination of *Cricotopus* species (Diptera: Chironomidae) by DNA barcoding. *Bulletin of Entomological Research*, *98*(6), 555-563. https://doi.org/10.1017/S0007485308005865

Soria, S. de J. (1971). La polinizacion del cacao por las mosquitas *Forcipomyia* spp. (Diptera, cera topoconidae) en Palmira, Colombia. *Acta Agronómica*, *21*(2), Art. 2.

Soria, S. de J., & Wirth, W. W. (1979). Ceratopogonid midges (Diptera: Nematocera) collected from cacao flowers in Palmira, Colombia: an account of their pollinating abilities. *Revista Theobroma*. https://scholar.google.com/scholar_lookup?title=Ceratopogonid+midges+%28Diptera%3A+Nematocera%29+collected+from+cacao+flowers+in+Palmira%2C+Colombia%3A+an+account+of+their+pollinating+abilities&author=Soria%2C+S.+de+J.&publication_year=1979

Spinelli, G. A., & Wolff, M. (2016). Family Ceratogopogonidae. *Zootaxa*, *4122*(1), 98-141. https://doi.org/10.11646/zootaxa.4122.1.15

Stur, E. & Borkent, A. (2014). When DNA barcoding and morphology mesh: Ceratopogonidae diversity in Finnmark, Norway. *ZooKeys*, *463*, 95-131. https://doi.org/10.3897/zookeys.463.7964

Szadziewski, R. (1988). Biting midges (Diptera, Ceratopogonidae) from Baltic amber. *Polish Journal of Entomology*, *58*, 3-283.

Szadziewski, R., & Arillo, A. (2003). The oldest fossil record of the extant subgenus Leptoconops [Leptoconops] [Diptera: Ceratopogonidae]. *Acta Zoologica Cracoviensia*.

https://www.semanticscholar.org/paper/The-oldest-fossil-record-of-the-extant-subgenus-Szadziewski-Arillo/ce395a2575178b5822ea8745e48755844cad143b

Szadziewski, R. (2008). Age and recent distribution of extant genera of Ceratopogonidae (Diptera) present in the fossil record. *Alavesia*, *2*, 87-99.

Tomazatos, A., Jöst, H., Schulze, J., Spînu, M., Schmidt-Chanasit, J., Cadar, D., & Lühken, R. (2020). Blood-meal analysis of Culicoides (Diptera: Ceratopogonidae) reveals a broad host range and new species records for Romania. *Parasites & Vectors*, *13*(1), 79. https://doi.org/10.1186/s13071-020-3938-1

Vansynghel, J., Ocampo-Ariza, C., Maas, B., Martin, E. A., Thomas, E., Hanf-Dressler, T., Schumacher, N.-C., Ulloque-Samatelo, C., Tscharntke, T., & Steffan-Dewenter, I. (2022). Cacao flower visitation: Low pollen deposition, low fruit set and dominance of herbivores. *Ecological Solutions and Evidence*, *3*(2), e12140. https://doi.org/10.1002/2688-8319.12140

Versteirt, V., Nagy, Z. T., Roelants, P., Denis, L., Breman, F. C., Damiens, D., Dekoninck, W., Backeljau, T., Coosemans, M., & Van Bortel, W. (2015). Identification of Belgian mosquito species (Diptera: Culicidae) by DNA barcoding. *Molecular Ecology Resources*, *15*(2), 449-457. https://doi.org/10.1111/1755-0998.12318

Winder, J. A. (1977). Field observations on Ceratopogonidae and other Diptera: *Nematocera* associated with cacao flowers in Brazil. *Bulletin of Entomological Research*, *67*(1), 57-63. https://doi.org/10.1017/S0007485300010890

Winder, J. A. (1978). Cacao Flower Diptera; Their identity, pollinating activity and breeding sites. *PNAS*, *24*(1), 5-18. https://doi.org/10.1080/09670877809414251