Infection, Genetics and Evolution 64 (2018) 76–84

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Discrimination of *Anopheles* species of the Arribalzagia Series in Colombia using a multilocus approach

Natalí Álvarez[a], Giovan F. Gómez[a,b], Nelson Naranjo-Díaz[a], Margarita M. Correa[a,*]

[a] Grupo de Microbiología Molecular, Escuela de Microbiología, Universidad de Antioquia, Calle 70 No. 52-21, Medellín, Antioquia, Colombia
[b] Grupo Investigación Ciencias Forenses y de la Salud, Facultad de Derecho y Ciencias Forenses, Institución Universitaria-Tecnológico de Antioquia, Medellín, Colombia

ARTICLE INFO

Keywords:
Anopheles
Malaria
Multilocus analysis
ITS2
COI
CAD
Arribalzagia Series
Colombia

ABSTRACT

The Arribalzagia Series of the *Anopheles* Subgenus comprises morphologically similar species or members of species complexes which makes correct species identification difficult. Therefore, the aim of this work was to discriminate the morphospecies of the Arribalzagia Series present in Colombia using a multilocus approach based on ITS2, *COI* and *CAD* sequences. Specimens of the Arribalzagia Series collected at 32 localities in nine departments were allocated to seven species. Individual and concatenated Bayesian analyses showed high support for each of the species and reinforced the previous report of the Apicimacula species Complex with distribution in the Pacific Coast and northwestern Colombia. In addition, a new molecular operational taxonomic unit-MOTU was identified, herein denominated near *Anopheles peryassui*, providing support for the existence of a Peryassui species Complex. Further, the *CAD* gene, just recently used for *Anopheles* taxonomy and phylogeny, demonstrated its power in resolving phylogenetic relationships among species of the Arribalzagia Series. The divergence times for these species correspond to the early Pliocene and the Miocene. Considering the epidemiological importance of some species of the Series and their co-occurrence in malaria endemic regions of Colombia, their discrimination constitutes an important step for vector incrimination and control in the country.

1. Introduction

During 2015, Colombia reported 52,416 malaria cases (INS, 2015) and for 2016, the number of cases increased to 83,356 (INS, 2016), demonstrating that malaria continues to be an important public health concern in the country. There are currently 47 *Anopheles* species registered in Colombia (Gonzalez and Carrejo, 2009), three of which have been identified as the most important vectors of malaria, *Anopheles* (*Nyssorhynchus*) *albimanus* Wiedemann, *Anopheles* (*Nys.*) *darlingi* Root and *Anopheles* (*Nys.*) *nuneztovari* Gabaldón (Olano et al., 2001, Gutiérrez et al., 2008, 2009). At least eight species are vectors of local importance or have been detected naturally infected with *Plasmodium* spp.; and for some of these, their epidemiological importance has yet to be demonstrated (Gutiérrez et al., 2008; Naranjo-Díaz et al., 2014; Quiñones et al., 2006).

The neotropical Arribalzagia Series includes 24 species (Harbach, 2013; Root, 1922), and in Colombia, based on morphology, at least 12 of them have been recorded (Gómez et al., 2015; Gonzalez and Carrejo, 2009). Various species in this Series are problematic during morphological identification, particularly those in the informal Punctimacula Group (Wilkerson, 1991), which are morphologically very similar, both in the adult stage and in the male genitalia (Wilkerson, 1991; Wilkerson et al., 1990). This group is composed of four species, *Anopheles* (*Anopheles*) *calderoni* Wilkerson, *Anopheles* (*Ano.*) *punctimacula* Dyar & Knab, *Anopheles* (*Ano.*) *guarao* Anduze and Capdevielle and *Anopheles* (*Ano.*) *malefactor* Dyar & Knab. Similarly, *Anopheles* (*Ano.*) *apicimacula* Dyar & Knab is frequently confused with *Anopheles* (*Ano.*) *neomaculipalpus* Curry, *An. malefactor* and also with *An. punctimacula* (Cienfuegos et al., 2008; Gómez et al., 2015; Loaiza et al., 2013). Moreover, the existence of two geographically isolated lineages were detected for *An. punctimacula* in Panama (Loaiza et al., 2013), and for *An. apicimacula* in Colombia (Gómez et al., 2015). Further, some of these species are of epidemiological importance in Colombia, for example, *An. punctimacula* and *An. calderoni* were found naturally infected with *Plasmodium* spp. (Herrera and Quiñones, 1987; Naranjo-Díaz et al., 2014; Orjuela et al., 2015).

Various molecular approaches have been used for problematic species identification. The nuclear Internal Transcribed Spacer 2-ITS2 region has shown great potential for *Anopheles* species discrimination (Cienfuegos et al., 2008, 2011; Zapata et al., 2007). Also, a fragment of the gene encoding for the mitochondrial Cytochrome C Oxidase subunit I (*COI*) was proposed as the barcode for mosquito identification

* Corresponding author.
E-mail address: margarita.correao@udea.edu.co (M.M. Correa).

https://doi.org/10.1016/j.meegid.2018.06.018
Received 1 March 2018; Received in revised form 6 June 2018; Accepted 15 June 2018
Available online 18 June 2018

(Cywinska et al., 2006; Hebert et al., 2003). Although, these two markers have demonstrated potential in molecular taxonomy, some reports of *COI* introgression (Gómez et al., 2015; Rona et al., 2010; Walton et al., 2000), and the presence of insertions, deletions and incomplete concerted evolution on ITS2 are challenging (Bourke et al., 2013; Motoki et al., 2011; Sallum et al., 2008). In addition, the nuclear single copy *CAD* gene, which encodes for a trifunctional protein associated with three enzymatic activities involved in the *de novo* biosynthesis of pyrimidines (Moulton and Wiegmann, 2004), has been used as a marker to resolve phylogenetic relationships in Diptera, with adequate resolution and at different hierarchical levels (Moulton and Wiegmann, 2004; Winterton et al., 2007). The *CAD* gene is linked to the X chromosome and presents low heterozygosity (Moulton and Wiegmann, 2004). *CAD* has corroborated phylogenetic relationships within 29 species of flies of the Eremoneura taxa, previously infered based on morphology and 28S rDNA sequences. *CAD* was also used to establish phylogenetic relationships within the small family of small headed flies Acroceridae, resulting in better performance than 16S, 28S and *COI* markers (Winterton et al., 2007). In the *Anopheles* genus, *CAD* was used to confirm species and decipher phylogenetic relationships in the *Nyssorhynchus* subgenus (Foster et al., 2013). Also, the more variable 3' *CAD* region was recently used for *Anopheles* genetic population studies in Colombia, and together with *COI* showed genetic differentiation of both, *An. darlingi* and *An. nuneztovari*, at both sides of the Andean mountains (Naranjo-Díaz et al., 2016a, 2016b).

Because of the potential limitations in the use of a single marker, a multilocus approach is strongly recommended to reconstruct evolutionary relationships, infer new linages within species and reveal species diversity within systems (Bourke et al., 2013; Phunngam et al., 2017). As such, a multilocus approach applying *COI*, ITS2 and the nuclear *white* gene helped to define the taxonomic status and to clarify phylogenetic relationships among the closely related species of the Strodei Subgroup (Bourke et al., 2013). Also, concatenated *COI*, *CAD* and *white* gene analysis performed better than single gene analysis for species confirmation within the *Nyssorhynchus* subgenus, and allowed the recognition of new species and identification of species within complexes (Foster et al., 2013). Considering the importance of the correct identification of *Anopheles* species and the power of a multilocus approach for species discrimination, this study applied a strategy that included three markers of different origin, mitochondrial (*COI*) and nuclear (*CAD* and ITS2), for the discrimination of species of the Arribalzagia Series in Colombia and to more precisely define their phylogenetic relationships.

2. Materials and methods

2.1. Mosquito processing

Mosquitoes included in this work were previously collected at 32 localities in nine Colombian departments (Table 1, Fig. 1). Collection details are in Altamiranda-Saavedra et al. (2017) and Naranjo-Díaz et al. (2016b). A total of 94 mosquitoes belonging to species of the Arribalzagia Series were analyzed. In addition, sequences available from GenBank, corresponding to Colombian specimens of *An. apicimacula*, *An. calderoni*, *An. malefactor*, *Anopheles* (*Ano.*) *mattogrossensis* Lutz & Neiva *An. neomaculipalpus*, *Anopheles* (*Ano.*) *peryassui* Dyar & Knab, *An. punctimacula*, were included in some of the analyses (Ahumada et al., 2016; Gómez et al., 2015) (Table 1). Adult mosquitoes were identified by morphological characters using available taxonomic keys (Gonzalez and Carrejo, 2009; Wilkerson et al., 1990). Genomic DNA extraction of individual specimens was performed using a previously standardized salt precipitation protocol (Rosero et al., 2010). Species assignment was performed by comparison of sequences to those in Genbank and The Barcode of Life Database (BOLD). The polymerase chain reaction- restriction fragment length polymorphism-PCR-RFLP-ITS2 strategy (Cienfuegos et al., 2008, 2011; Zapata et al., 2007) was used to separate *An. apicimacula* lineages.

2.2. PCR design and standardization for CAD amplification

To standardize the *CAD* PCR, the DNAs from specimens of four species of the Arribalzagia Series previously confirmed by *COI* were used and included, *An. calderoni*, *An. punctimacula*, *An. malefactor* and *An. neomaculipalus*. A *CAD* gene fragment was amplified by nested PCR using degenerated primers 581-F and 1098-R (Moulton and Wiegmann, 2004), and the amplification conditions optimized in the lab. The PCR reaction contained: 0.2 µM of 581-F and 1098-R oligonucleotides, 20 ng/µl of genomic DNA, 1 U of polymerase (Taq polymerase Bioline), 0.2 mM dNTPs, 0.5 µg/µl BSA, 1.5 mM MgCl2, 1 × buffer and nuclease free water to complete 25 µl. PCR conditions were: initial denaturation at 94 °C for 3 min, 35 cycles at 94 °C for 30 s, alignment at 50 °C for 20 s, elongation at 72 °C for 1 min, and a final extension at 72 °C for 8 min. A PCR product of 1600 bp was obtained. In the second nested PCR, the primers 581-F and 843-R were used, and the conditions were as for the first nested PCR. A PCR product of 1040 bp was obtained and sequenced in both directions. Sequences were aligned and edited using Geneious 8.1.8 (Kearse et al., 2012). A sequence similarity search was performed using the Basic Local Alignment Search Tool-BLAST (Altschul et al., 1990). The obtained *CAD* sequences were aligned with the entire *CAD* gene and with the *CAD* region of *Anopheles* (*Cellia*) *gambiae* Patton (Mongin et al., 2004; Wiegmann et al., 2011), using the MUSCLE algorithm (Edgar, 2004), in Geneious 8.1.8.

For primer design, the Primer 3 software (Untergasser et al., 2012) was used and the resulting oligonucleotides were revised with OligoAnalyzer version 3.1. The primers designed for the *CAD* PCR were *CAD*-F (GAGTAYGTCGAGCARGTGATMAAG) and *CAD*-R (CAGTAGTC-KAGGCTSGGCTCRAAG); they amplified an 839 bp fragment that corresponds to the CPS region on the second *CAD* gene exon. The PCR reaction mix to amplify *CAD* with the newly designed primers consisted of 0.15 µM oligonucleotides *CAD*-F and *CAD*-R, 20 ng/µl DNA, 1 U/reaction thermostable polymerase, 0.2 mM of dNTPs, 0.5 µg/µl BSA, 2 mM MgCl2, 1 X buffer and nuclease free water. The PCR conditions consisted of an initial denaturation step at 94 °C for 3 min, 35 cycles at 94 °C for 30 s, alignment at 59.7 °C for 20 s, elongation at 72 °C for 30 s, and a final extension step at 72 °C for 8 min. Amplified products were sequenced in both directions. *CAD* sequences were edited and aligned in Geneious 8.1.8, using the MUSCLE algorithm, and synonymous and non-synonymous mutations were determined using DnaSP v. 5 (Librado and Rozas, 2009). An analysis performed in the program PHASE v2.1 (Stephens et al., 2001) showed *CAD* homozygosity in all species evaluated.

2.3. ITS2 and COI workflow

The 5′ *COI* barcode fragment was amplified using universal primers LCO and HCO (Folmer et al., 1994), and conditions standardized by Gómez et al. (2013). The ITS2 region was amplified using primers and conditions standardized by Zapata et al. (2007). *COI* and ITS2 amplification products were sequenced in both directions. Sequences were aligned and edited using the MUSCLE algorithm in Geneious 8.1.8 and similarity searches were performed using BLAST. *COI* sequences were compared to those available on the BOLD platform. ITS2 sequences were annotated based on 5.8S and 28S conserved sequences available in GenBank for species of the Arribalzagia Series (KF698892, KF698888, KF698896, KF698897, KF698907, KM262760, KF698904).

2.4. Genetic distance analysis

Intra and interspecific pairwise genetic distances for each marker were evaluated using MEGA v. 7.0.18 (Kumar et al., 2016; Tamura and Nei, 1993), under the bootstrap method for ITS2 and *CAD*, and the Kimura 2- parameter model (K2P) for *COI* (Hebert et al., 2003). The