Phylogeography of the neotropical sand fly *Lutzomyia longipalpis* inferred from mitochondrial DNA sequences

J.C. Arrivillaga a,b, D.E. Norris c, M.D. Feliciangeli b, G.C. Lanzaro a,*

a Department of Pathology, University of Texas Medical Branch, 301 University Boulevard, Galveston, TX 77555-0609, USA

b Universidad De Carabobo, Fac. de Ciencias de la Salud, BIOMED, Centro Nacional de Referencia de Flebotomos, La Morita, Maracay, Apdo. 4873, Venezuela

c The W. Harry Feinstone Department of Molecular Microbiology and Immunology, The Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA

Received 8 October 2001; received in revised form 14 May 2002; accepted 8 June 2002

Abstract

Sand flies in the *Lutzomyia longipalpis* species complex include the primary vector of *Leishmania chagasi*, the etiologic agent of visceral leishmaniasis in the Neotropics. Twelve *L. longipalpis* populations from South and Central America were compared using the cytochrome *c* oxidase I (COI) gene from the mitochondrial genome. The haplotype profiles for each population revealed that the majority of sequence variation was inter-population (98%) rather than intra-population, suggesting that sequence polymorphisms at the COI locus should provide excellent characters for the study of phylogenetic relationships among populations. Phylogenetic reconstruction using distance (neighbor-joining) and maximum parsimony analysis revealed the existence of four clades among the *L. longipalpis* populations studied: (1) Laran, (2) Brazilian, (3) *cis*-Andean and (4) *trans*-Andean. We suggest that these clades represent species. A biogeographical interpretation of the molecular phylogeny suggests that the process of speciation in the *L. longipalpis* complex began in the Pliocene, from a sub-Andean–Amazonian gene pool resulting from the Andean orogeny (formation of the East Andean Cordillera). The four clades probably diverged as a result of vicariance events that occurred throughout the late Pliocene and Pleistocene. We propose and discuss several historical scenarios, based on the biogeography and historical geology of Central and South America.

Keywords: *Lutzomyia longipalpis*; mtDNA; Phylogeny; Sand flies; Vicariance

1. Introduction

*Lutzomyia longipalpis* (Diptera: Psychodidae), the principal vector of visceral leishmaniasis in the Neotropics, was first described from collections made in Brazil by Lutz & Neiva in 1912. Reviews by Forattini (1973) and Young and Duncan (1994) report *L. longipalpis* as a Neotropical species with a broad distribution from southern Mexico to northern Argentina. Although widespread, its geographic distribution is discontinuous. Its geographic range includes climatic and physiogeographic discontinuities that are associated with patterns of population divergence. There is now a consensus that *L. longipalpis* is a complex of sibling species distinguishable on the basis of male pheromone composition (Ward et al., 1988; Hamilton et al., 1996a,b,c), allozyme frequencies (Lanzaro et al., 1993; Arrivillaga, 1999; Lampo et al., 1999), interspecific male hybrid sterility (Lanzaro et al., 1993), chromosome patterns (Yin et al., 1998) and larval morphology (Arrivillaga et al., 2000, in press). However, the number of species within the complex and their geographic distributions remain unclear. Little is known about phylogenetic relationships among *L. longipalpis* populations. This knowledge is necessary for the development of models that will provide a better understanding of the current geographic distributions and processes of speciation for members of this complex. In addition, an improved understanding of the biogeography and genetics of these populations may help clarify their taxonomic status.

Mitochondrial DNA (mtDNA) has proven useful in molecular phylogenetics due to its maternal inheritance, rapid rate of divergence and lack of recombination (Avise et al., 1987; Avise, 1994; Moritz, 1994; Simon et al., 1994). Little is known about phylogenetic relationships among this group of sand flies (*Esseghir et al., 1997; Ready et al., 1997, 1998; Ishikawa et al., 1999*). However, studies of the
molecular phylogenetics of other arthropod groups suggest that sequence from region I of the mitochondrial cytochrome c oxidase I (COI) gene might be informative for clarifying Lutzomyia taxonomy at the species level (Navajas et al., 1994, 1996, 1998; Roderick and Gillespie, 1998). Here, we present a phylogeographical interpretation of L. longipalpis based on molecular phylogenetic analysis using mtDNA and available theories on the historical physical geography of Central and South America (Haffer, 1969; Howe, 1974; Keigwin, 1978; Haffer, 1981; Rod, 1981; Sykes et al., 1982; Stehli and Webb, 1985; Hoorn, 1994; Hoorn et al., 1995). Our interpretation of these data is based on historical vicariance biogeography and dispersal and refuge theories (Croizat, 1978; Rosen, 1976, 1978; Nelson, 1976; Bermingham et al., 1992; Higgs, 1994; Joseph et al., 1995; Lundberg et al., 1998). This analysis provides a coherent explanation of historical factors responsible for the genetic structure of extant populations.

2. Materials and methods

2.1. Sample collections

Lutzomyia longipalpis adults (N = 343) were analyzed from 12 field localities in Central and South America (Table 1 and Fig. 1). Populations were selected based on previous isozyme studies that suggested these populations represent three or more species (Lanzaro et al., 1993; Mutebi et al., 1998, 1999; Yin et al., 1998; Arrivillaga et al., 2000, in press). Adult flies were collected in Colombia, Costa Rica, Brazil, Honduras and Venezuela using aspirators from domestic animals (chickens, cattle, pigs or dogs) or from resting sites near animal enclosures, and with Centers for Disease Control (CDC, Atlanta, USA) light traps.

DNA was extracted from individual flies using previously described methods (Bender et al., 1983). A ∼540 bp fragment of the cytochrome COI gene was PCR amplified as a single product using previously published universal primers: CI-J-1632 (+): 5′-TGATCAAATTATAAAT-3′ and CL-N-2191 (−): 5′-GGTAAAATATATATAACTCT-3′ and amplification conditions originally described by Simon et al. (1994) and later modified by Kambhamumpi and Smith (1995). The PCR products were purified using QIAquick PCR purification kits to remove excess primers, nucleotides and polymers. Each sample was screened for polymorphism by single strand conformation polymorphism (SSCP) analysis. Electrophoresis for SSCP was performed on gels composed of 0.6× TBE (1× TBE = 53 mM Trizma base), 53 mM boric acid, 1.5 mM EDTA pH 8.0, 7.5% acrylamide, and 0.21% N,N,N′,N′-methylenebisacrylamide. Following PCR amplification, 7 μl of product was removed to a 500 μl tube containing 9 μl of denaturing loading mix (20 mM NaOH, 80% formamide, 0.05% bromophenol blue and 0.05% xylene cyanol). The tube was tapped to mix the contents, spun briefly in a microcentrifuge, heated to 95 °C for 3 min, and plunged into ice for at least 10 min. From this cooled mixture, 16 μl was loaded onto the SSCP gel. Electrophoresis was carried out on vertical slab gels (16 cm × 18 cm) and run in refrigerated (4 °C), circulating 0.5× TBE electrode buffer. Amperage was maintained at 20 mA until the xylene cyanol dye had migrated to the bottom of the gel. Gels were then silver stained, dried, scored and documented. Polymorphism was detected by the following procedure: first, putative haplotypes present in each local population were identified by visual inspection of SSCP gels so that a haplotype profile could be established for each population. Unique haplotypes in each population profile were

<table>
<thead>
<tr>
<th>Population</th>
<th>Location</th>
<th>Country</th>
<th>Year</th>
<th>Method capture</th>
<th>Geographical region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liberia</td>
<td>10° 38′ 18″N, 85° 26′ 32″W</td>
<td>Costa Rica</td>
<td>August 1997</td>
<td>Manual aspirator</td>
<td>Mountain system, Central America</td>
</tr>
<tr>
<td>Panama</td>
<td>15° 25′ 35″N, 87° 20′ 1″W</td>
<td>Honduras</td>
<td>July 1995</td>
<td>Manual aspirator</td>
<td>Mountain system, Central America</td>
</tr>
<tr>
<td>Carara</td>
<td>9° 59′ 4″N, 60° 55′W</td>
<td>Venezuela</td>
<td>1996–1999</td>
<td>CDC trap</td>
<td>Loran depression (flood Andean Cordillera) limited by northern by Caranica mountain system, Falcon basin</td>
</tr>
<tr>
<td>Bucaramanga</td>
<td>7° 08′ 48″N, 73° 09′W</td>
<td>Colombia</td>
<td>June 1993</td>
<td>Manual aspirator</td>
<td>East Andean Cordillera, middle Magdalena River</td>
</tr>
<tr>
<td>Neira</td>
<td>12° 56′ 5″N, 75° 18′W</td>
<td>Colombia</td>
<td>June 1993</td>
<td>Manual aspirator</td>
<td>East Andean Cordillera, upper Magdalena River</td>
</tr>
<tr>
<td>Ronima</td>
<td>5° 30′ 6″N, 60° 40′W</td>
<td>Brazil</td>
<td>1999</td>
<td>Manual aspirator</td>
<td>Guayana Shield, Pakaraima Mountains</td>
</tr>
<tr>
<td>Santarem</td>
<td>2° 26′ 38″N, 54° 41′W</td>
<td>Brazil</td>
<td>August 1997</td>
<td>Manual aspirator</td>
<td>Amazon River</td>
</tr>
<tr>
<td>Salvatierra</td>
<td>0° 46′ 1′S, 48° 31′W</td>
<td>Brazil</td>
<td>August 1997</td>
<td>Manual aspirator</td>
<td>Delta Amazon River</td>
</tr>
<tr>
<td>Batamani</td>
<td>4° 20′ 5″S, 38° 53′W</td>
<td>Brazil</td>
<td>August 1997</td>
<td>Manual aspirator</td>
<td>Sao Paulo Brazilian</td>
</tr>
<tr>
<td>Jacobina</td>
<td>16° 11′ 5″S, 40° 31′W</td>
<td>Brazil</td>
<td>August 1997</td>
<td>Manual aspirator</td>
<td>Sao Paulo basin</td>
</tr>
<tr>
<td>Lapinha</td>
<td>15° 19′ 03″S, 43° 57′W</td>
<td>Brazil</td>
<td>November 1997</td>
<td>Manual aspirator</td>
<td>Brazilian Shield, Highlands</td>
</tr>
</tbody>
</table>
sequenced using standard ABI 377 automated sequencing techniques. Cycle sequencing products were run over a G-50 Sephadex matrix in Centri-Sep spin columns to remove excess primers, dye, nucleotides, and polymerase. The reliability of SSCP for identifying haplotypes was evaluated by running samples, taken from different populations but having identical haplotypes (determined by sequencing), on a single gel. In every case, the SSCP patterns were identical. Conversely, samples known to have different haplotypes, determined by sequencing, presented unique SSCP banding phenotypes when run side by side on a single gel.

Sequences were initially aligned using Clustal V (Higgins et al., 1992) and manually adjusted for obvious misalignments. Gaps were treated as missing data. All phylogenetic analyses were completed using PAUP 4.0b2 (Swofford, 1999). Neighbor-joining (NJ) analysis (Saito and Nei, 1987) was completed under the assumptions of the Kimura 2-parameter (K2P) model. Maximum parsimony (MP) analysis was completed with equally weighted characters and re-weighting (Farris, 1969) employing the consistency index (CI) by heuristic search using the tree bisection-reconnection (TBR) branch-swapping algorithm with stepwise addition at random using 100 replications each. A strict consensus was calculated from the most parsimonious trees and confidence values (Felsenstein, 1985) were estimated by bootstrapping (1000 replications), after deleting uninformative characters (Carpenter, 1996). Nucleotide composition, pairwise base differences and genetic distance (K2P) were calculated using PAUP 4.0b2. An MP phylogram was constructed to illustrate the length of each branch where branch length represents the number of character state changes necessary to support the relationships among taxa in the tree.

Analysis was conducted using several different out-group strategies, as follows: (a) without outgroups, (b) using Phlebotomus (Phlebotomus) papatasii from Saudi Arabia as outgroup and Lutzomyia evansi (Vernacularum group) from Costa Rica as sister in-group, (c) using L. gomezi from Venezuela and L. renei from Brazil as sister groups (both in the sub-group Lutzomyia), (d) using P. papatasii, L. evansi and L. gomezi as outgroups and (e) repeating b, c and d but enforcing monophyly of the L. longipalpis complex by using the Curarigua haplotypes as outgroups.
<table>
<thead>
<tr>
<th>Marker</th>
<th>Frequency in Population 1</th>
<th>Frequency in Population 2</th>
<th>Difference</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mark1</td>
<td>0.25</td>
<td>0.30</td>
<td>0.05</td>
<td>0.23</td>
</tr>
<tr>
<td>Mark2</td>
<td>0.10</td>
<td>0.15</td>
<td>0.05</td>
<td>0.45</td>
</tr>
<tr>
<td>Mark3</td>
<td>0.30</td>
<td>0.25</td>
<td>0.05</td>
<td>0.60</td>
</tr>
<tr>
<td>Mark4</td>
<td>0.40</td>
<td>0.45</td>
<td>0.05</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Note: The table shows the frequencies of different markers in two populations, along with the calculated difference and P-value for each marker.
18. haplotypes were observed among 343 individu-
als from 12 sites examined. Sample size per site, distribu-
tion and frequencies of haplotypes are described in Table 2. Relationships among haplotypes, described as pairwise K2P distances, are presented in Table 3. Variation at the COI locus was low within populations, but high between populations (Table 2, Figs. 2–4). From a total of 18 haplotypes, seven were unique to a single popula-
tion (Table 2). Five populations contained more than one haplotype, but none were shared with other populations. Analysis of patterns of sequence divergence among the pop-
ulations studied support the existence of at least four species in the taxon now recognized as \( L. \) \( longipalpis \). We de-
scribe these as follows: species A (Laran Clade), species B (Brazilian Clade), species C (cis-Andean Clade) and species D (trans-Andean Clade). These results are consistent with earlier studies based on isozyme analysis (Lanzaro et al., 1993, 1998; Mutebi et al., 1998, 1999; Lampo et al., 1999). 3.1. Phylogenetic analyses Phylogenetic trees, generated by NJ and MP, had iden-
tical topologies and branches were supported by similar bootstrap values. Likewise, MP analysis using heuristic search and branch and bound methods resulted in trees with identical topology and with similar bootstrap values. Tree topology and bootstrap values resulting from analyses with the different outgroup arrangements described above were equivalent. A analysis using the Curarigua haplotypes as outgroups revealed that this population represents an internal-basal clade within the ingroup. We conclude that the heuristic search method is adequate for the illustration of phylogenetic relationships among \( L. \) \( longipalpis \) haplo-
types and employ this method with \( P. \) \( papatasi \) as outgroup and \( L. \) \( evansi \) and \( L. \) \( gomezi \) as sister groups. The strict consensus tree (Fig. 2) derived from the three equally most parsimonious trees (length = 299 steps, CI = 0.75) indicate the existence of three monophyletic lineages supported by bootstrap values >80%: (I) Laran (Curarigua), (II) Andean (Pavana, Liberia, Colombia, Tru-
jillo, Bucaramanga, Neiva, Roraima) and (III) Brazilian (Lapinha, Salvaterra, Santarem, Jacobina, Batutiré). The spatial distribution of populations contained within each clade is illustrated in Fig. 1. The Laran Clade is represented by a single population collected in Curarigua, Venezuela. The three haplotypes ob-
served in this population were nearly identical (percentage nucleotide (nt) divergence = 0.01–0.02), but very different from haplotypes present in the other populations (Table 3 and Fig. 3). The data suggest that the Laran Clade is relict-
ual. There is strong support by both NJ and MP analyses to place the Laran Clade in a basal position in the phylogenetic tree (Fig. 2). Analysis of variable sites reveal two additional groups (Fig. 2): the Andean and Brazilian Clades. Within the Andean Clade are two internal clades, a trans-Andean Clade, represented by Central American and Andean hap-
lotypes and a cis-Andean Clade represented by Andean and Amazonian haplotypes. These are not as strongly supported as the three main clades, with bootstrap values slightly less than 80%. Analysis of mtDNA haplotypes from the Laran (29 apo-
morphic characters) and Brazilian Clades (11 apomorphic characters) suggest they are monophyletic and strongly support phylogenetic species status for each. We refer to the Laran Clade as species B, which was recently recog-
nized taxonomically as \( L. \) \( pseudolongipalpis \) Arrivillaga and Feliciangeli (2001). The Brazilian Clade we refer to as species A, and will probably be designated the nominal species in the \( L. \) \( longipalpis \) complex when the taxonomy of the group is formalized. The third major clade resolved in our analysis we refer to as the Andean Clade, which we subdivide into a cis-Andean Clade (Colombia and Andean Venezuela) and trans-Andean Clade (Central America). The occurrence, in sympatry, of closely related haplotypes among samples from Bucaramanga and Trujillo do not support monophyly of the trans- and cis-Andean Clades, as suggested by Brower (1999). However, in cross-mating studies Lanzaro et al. (1993) demonstrated post-mating re-
productive barriers, in the form of hybrid male sterility, in reciprocal crosses between members of these two clades. In addition, we have found morphological characters that can be used to distinguish members of these two clades (Arriv-
illaga, unpublished). We refer to the cis-Andean Clade as species C and the trans-Andean Clade as species D. There is currently insufficient knowledge of the molecular evolution of this group of sand flies to permit a meaningful estimate of divergence times based on sequence data. In the following discussion, we have attempted to develop several alternative theories to explain diversification of taxa within the \( L. \) \( longipalpis \) complex, based on analyses of available geological data, and phylogenetic analysis of mtDNA COI sequence polymorphism. 4. Vicariance hypothesis Divergence of the Laran Clade from an ancestral North Andean–Amazonian gene pool was the first step toward speciation of \( L. \) \( longipalpis \) in northern, sub-Andean South America. This is reflected in the mitochondrial haplotypes that exist within this clade, which, under this scheme, rep-
resent ancestral haplotypes (10.91% nt divergence, 5.37% nt are synapomorphies). Specimens from the Laran Clade were collected in a depression in the Chivacoa Gap that lies over the Bocono Fault, at the town of Curarigua. Isolation of the Chivacoa Gap area began with tectonic movement in the late Miocene and continued until the major uplift of the East Andean Cordillera (Rod, 1981; Hooen et al., 1995). The degree of divergence between this and the remaining clades (Figs. 2 and 3, Table 4) suggests a long period of
Table 3
Kimura 2-parameter (K2P) pairwise genetic distances among 18 cytosome c oxidase subunit I (COI) mtDNA haplotypes from members of the L. longipalpis species complex

<table>
<thead>
<tr>
<th>Haplotype name</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>P. puparii (1)</td>
<td>0.172</td>
<td>0.199</td>
<td>0.173</td>
<td>0.158</td>
<td>0.159</td>
<td>0.175</td>
<td>0.182</td>
<td>0.2</td>
<td>0.169</td>
<td>0.169</td>
<td>0.18</td>
<td>0.153</td>
<td>0.181</td>
<td>0.164</td>
<td>0.168</td>
<td>0.171</td>
<td>0.176</td>
<td>0.176</td>
<td></td>
</tr>
<tr>
<td>L. evansi (2)</td>
<td>0.155</td>
<td>0.149</td>
<td>0.153</td>
<td>0.152</td>
<td>0.165</td>
<td>0.177</td>
<td>0.203</td>
<td>0.17</td>
<td>0.164</td>
<td>0.167</td>
<td>0.157</td>
<td>0.177</td>
<td>0.161</td>
<td>0.151</td>
<td>0.153</td>
<td>0.157</td>
<td>0.163</td>
<td>0.166</td>
<td></td>
</tr>
<tr>
<td>Cururupu 2 (3)</td>
<td>0.009</td>
<td>0.013</td>
<td>0.11</td>
<td>0.177</td>
<td>0.124</td>
<td>0.156</td>
<td>0.118</td>
<td>0.114</td>
<td>0.13</td>
<td>0.125</td>
<td>0.141</td>
<td>0.125</td>
<td>0.123</td>
<td>0.118</td>
<td>0.128</td>
<td>0.132</td>
<td>0.122</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cururupu 6 (4)</td>
<td>0.019</td>
<td>0.108</td>
<td>0.116</td>
<td>0.124</td>
<td>0.155</td>
<td>0.119</td>
<td>0.115</td>
<td>0.125</td>
<td>0.125</td>
<td>0.139</td>
<td>0.126</td>
<td>0.122</td>
<td>0.113</td>
<td>0.125</td>
<td>0.131</td>
<td>0.123</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cururupu 10 (5)</td>
<td>0.109</td>
<td>0.116</td>
<td>0.122</td>
<td>0.157</td>
<td>0.111</td>
<td>0.105</td>
<td>0.125</td>
<td>0.124</td>
<td>0.14</td>
<td>0.126</td>
<td>0.124</td>
<td>0.119</td>
<td>0.125</td>
<td>0.129</td>
<td>0.113</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lapiilha 53 (6)</td>
<td>0.087</td>
<td>0.07</td>
<td>0.104</td>
<td>0.05</td>
<td>0.052</td>
<td>0.098</td>
<td>0.086</td>
<td>0.116</td>
<td>0.092</td>
<td>0.078</td>
<td>0.075</td>
<td>0.089</td>
<td>0.117</td>
<td>0.094</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salvarriana (7)</td>
<td>0.04</td>
<td>0.07</td>
<td>0.029</td>
<td>0.019</td>
<td>0.115</td>
<td>0.097</td>
<td>0.139</td>
<td>0.101</td>
<td>0.091</td>
<td>0.066</td>
<td>0.096</td>
<td>0.135</td>
<td>0.117</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Santarem 46 (8)</td>
<td>0.006</td>
<td>0.05</td>
<td>0.042</td>
<td>0.126</td>
<td>0.113</td>
<td>0.144</td>
<td>0.117</td>
<td>0.101</td>
<td>0.112</td>
<td>0.112</td>
<td>0.144</td>
<td>0.134</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Santarem 49 (9)</td>
<td>0.085</td>
<td>0.078</td>
<td>0.156</td>
<td>0.136</td>
<td>0.167</td>
<td>0.13</td>
<td>0.12</td>
<td>0.133</td>
<td>0.132</td>
<td>0.136</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jacobina (10)</td>
<td>0.009</td>
<td>0.122</td>
<td>0.086</td>
<td>0.145</td>
<td>0.094</td>
<td>0.094</td>
<td>0.104</td>
<td>0.103</td>
<td>0.134</td>
<td>0.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beitute (11)</td>
<td>0.12</td>
<td>0.092</td>
<td>0.141</td>
<td>0.09</td>
<td>0.096</td>
<td>0.103</td>
<td>0.097</td>
<td>0.13</td>
<td>0.106</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trujillo 10 (12)</td>
<td>0.074</td>
<td>0.104</td>
<td>0.078</td>
<td>0.062</td>
<td>0.065</td>
<td>0.061</td>
<td>0.104</td>
<td>0.094</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trujillo 12 (13)</td>
<td>0.103</td>
<td>0.025</td>
<td>0.034</td>
<td>0.043</td>
<td>0.041</td>
<td>0.098</td>
<td>0.079</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacaraimanga 90 (14)</td>
<td>0.101</td>
<td>0.095</td>
<td>0.084</td>
<td>0.096</td>
<td>0.081</td>
<td>0.046</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacaraimanga 90 (15)</td>
<td>0.04</td>
<td>0.047</td>
<td>0.017</td>
<td>0.098</td>
<td>0.077</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roraima 1 (16)</td>
<td>0.015</td>
<td>0.017</td>
<td>0.09</td>
<td>0.071</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roraima 2 (17)</td>
<td>0.025</td>
<td>0.08</td>
<td>0.064</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neiva 25 (18)</td>
<td>0.097</td>
<td>0.076</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liberia 63 (19)</td>
<td>0.044</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pacana 1 (20)</td>
<td>0.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Two additional sequences are included: P. puparii (1rogroup) and L. evansi (1rogroup).*
Fig. 2. Rooted phylogenetic tree (MP) produced by PAUP 4.0b2 based on sequence for the mitochondrial COI locus. For the COI sequences, the strict consensus of 10 trees is shown based on 131 synapomorphic characters. Consistency index = 0.75 in 299 steps. Symbols: (●), (○), (△), (□) and (○) correspond to locations indicated on the map in Fig. 1.
Fig. 3. Maximum Parsimony phylogram indicating the number of nucleotide changes among haplotypes based on branch lengths from MP strict consensus tree illustrated in Fig. 2.

Table 4
Summary of genetic distances (K2P) among the four *L. longipalpis* clades revealed in this study

<table>
<thead>
<tr>
<th>Clade</th>
<th>Nanaimo (N = 2)</th>
<th>trans-Andean (N = 4)</th>
<th>cis-Andean (N = 4)</th>
<th>Brazilian (N = 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique MP Tree</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Distances are based on sequence from the mitochondrial cytochrome c oxidase locus. The number of populations analyzed within each clade is designated by *N*. Standard deviations are in parentheses.
Fig. 4. Variable sites in a ~540 bp fragment of the mtDNA COI region from L. forcipatus. Sequence analysis revealed 18 haplotypes from 12 populations. The insert includes a sequence for haplotype Carigua10 (* variable sites).
geological quiescence following this vicariant event. Isolation of the Chivacoa Gap area was maintained throughout this period by four geographical barriers (Pliocene ≈ 5.45 million years): (1) to the north by incursion of the Caribbean Sea into the Falcon Basin, (2) to the west by the incipient Andean Cordillera, (3) to the east by the incipient Coastal Cordillera and perhaps most importantly, (4) to the south by the Baul Arch, which formed a barrier between the basin of the ancient Eastern Cordillera and the lowlands southeast of the rising Andes (Higgs, 1994).

Divergence of the Brazilian Clade (9.8% nt divergence, 2.7% nt are synapomorphies) and Andean Clades (9.47% nt divergence, 2.59% are synapomorphies) were more recent. These data suggest the existence of an ancestral, sub-Andean–Amazonian gene pool prior to vicariant events in the Ploceine–Pleistocene. Genetic uniformity in this gene pool was presumably maintained by gene flow among populations in the Andes, Amazon Basin and Guyana Shield followed by colonization of Central America in the late Pliocene–early Pleistocene (Croizat, 1981; Marshall, 1988). Repeated incursions of the Caribbean Sea and changes in drainage patterns of the Orinoco River, from the early Oligocene to the late Miocene allowed further differentiation of the sub-Andean–Amazonian gene pool from the Laran Clade in northern South America. Complete isolation of these two major Clades probably occurred in the Pleistocene, with the final uplift of the East Andean Cordillera and subsequent extension of the Amazon River to the Atlantic Ocean (Hoorn et al., 1995) as the principal vicariant events.

Geological evidence suggests that alterations in central South America occurred gradually, firstly, with the development of the Amazon River (late Miocene) and later with alteration of its drainage patterns, from the western Andes to northeastern Amazonia, between the late and post Pleistocene. (Rasanei et al., 1987; Smith and Patton, 1993; Hoorn et al., 1995; daSilva and Patton, 1998). These events led to cladogenesis of the Brazilian and Andean Clades and are reflected in the high degree of genetic differentiation among populations in the Brazilian Clade, leading to the occurrence of at least four closely related population groups with 5.25% nt divergence within this clade (Fig. 2). Phylogenetic analysis (MP) reveals close relationships among populations in the Brazilian highlands (Lapinha Cave), the Amazon (Salvatera and Santarem), San Francisco Basin (Jocobina) and Brazilian scrubland (Batutiri). These results are concordant with patterns of divergence in the nuclear genome (isozyme variation) among the same Brazilian populations, with the exception of the population at Roraima, that was not included in the isozyme studies (Mukhopadhyay et al., 1998; Mutebi et al., 1999; Azevedo et al., 2000). These authors concluded that populations in Brazil represent a single species, but with relatively high levels of genetic divergence among local populations.

The Andean Clade is complex and comprised of two lineages with a bootstrap value ≈78% (Fig. 2), the cis- and trans-Andean Clades. The occurrence in sympathy of haplotypes with close relationships to those from both the cis- and trans-Andean Clades (Bucaramanga 86 and 90, and Trujillo 10 and 12, Fig. 2) suggests past introgression into Central America by gene flow across the Magdalena River Basin during the late Ploceine–Pleistocene (Hammen and Gonzalez, 1964; Wellman, 1970; Howe, 1974; Rod, 1981; Hoorn et al., 1995). However, nt divergence is highly significant (10.08-7.4%) among the sympatric haplotypes. This suggests a lack of contemporary gene flow and biological species status for these two clades, but possibly only incipient phylogenetic species status. The notion that the cis- and trans-Andean Clades represent biological species is supported by isozyme studies of these populations that revealed levels of genetic distance consistent with species in the genus Lutzomyia (Lanzaro and Warburg, 1995; Lanzaro et al., 1998; Mutebi et al., 1998). In addition, crosses made between individuals belonging to the two clades resulted in sterile male progeny (Lanzaro et al., 1993).

Strong clustering of samples from Neiva (Andes) and Roraima (Guyana Shield) populations (cis-Andean Clade) suggest recent gene flow (3% nt divergence) between these geographically isolated populations (Croizat, 1981). Paleogeographical and sedimentological data reveal that sedimentary rock (non-marine formation) around Neiva in the southwestern portion of the Central Andean Cordillera (Colombia) is continuous with the Guyana Shield, on which Roraima is located, suggesting these regions were connected before the final uplift of the East Andean Cordillera in the late Pleistocene (Howe, 1974). This geological evidence may explain the close relationship between the Neiva and Roraima populations.

5. Dispersal hypothesis

The current distribution of L. longipalpis, includes southern Mexico, Central America, Colombia, Venezuela (north of the Orinoco River), Brazil (south of the Amazon River), Paraguay, Bolivia and Argentina. These populations may have derived from an ancestral Amazonian gene pool that dispersed across Brazil, to the Andes and finally to Central America. Dispersal would have occurred after the final uplift of the East Andean Cordillera in the early Pleistocene. Cladogenesis resulting in the four contemporary clades may have been the consequence of adaptation to local environmental conditions and isolation/drift. Gaps in the contemporary geographic distribution of L. longipalpis and large genetic discontinuities between population groups are the result of regional extinction, for example in the area from the Orinoco River and continuing south to the Amazon River where L. longipalpis currently is not found. Early geographic changes resulting in the isolation of the Chivacoa Gap region and divergence of the population there resulted in the Laran Clade. The existence of sympatric haplotypes in the Andes (cis- and trans-Andean Clades) also could be
explained by dispersal, with multiple episodes of migration as ecological conditions permitted range expansion of local populations (later Pliocene–early Pleistocene). However, the fact that *L. longipalpis* is a poor flyer, with maximum daily flight range estimated at 1000 m (Alexander, 1987; Morrison et al., 1993), argues against large-scale dispersal as a mode of cladogenesis in this group.

6. Forest refugia hypothesis

Under this scenario clades comprising the *L. longipalpis* complex resulted from climatic fluctuations that caused periodic geographical fragmentation and the creation of savanna replacing humid forest. This led to the divergence, in allopatry, of Brazilian, northwestern Venezuelan and Andean—Central American population groups. Divergence of lineages occurred after the final uplift of the East Andean Cordillera. This event resulted in an ecological change of what was previously an extensive humid forested area to the production of an extensive arid/semi-arid savanna (late Pleistocene–Quaternary). The preferred habitat for contemporary populations of *L. longipalpis* s.l. is dry tropical forest and open savanna (Forattini, 1973), however, it is likely that ancestral populations were associated with wet forest habitats. The large degree of genetic differentiation between populations across the Andes do not support a recent divergence time, suggesting that the refugia model does not provide an adequate explanation for the evolution of the clades under study here. Eiseighr et al. (1997) suggests that diversification of closely related sand fly species in the genus Phlebotomus occurred in the Quaternary period, based on 1–2.5% nt divergence in the mitochondrial cytochrome b locus, whereas our values range between 9.47 and 10.97% nt divergence among three clades in the cytochrome c locus. However, climatic changes during the late Pleistocene may have led to changes in the distribution of plant communities at lower elevations, creating refugia that contributed to the formation of *L. longipalpis* clades.

Under the refugia hypothesis, the origin of the Laran Clade might be explained by the existence of isolated rain forest refugia in what are now Lara, Falcon and Yaracuy states in Venezuela. These rainforest “islands” were isolated by a large area of contiguous semi-arid grasslands in Lara state that would have restricted dispersal of a forest dwelling *L. longipalpis* ancestor between the Andes Mountains and Coastal Cordillera (Huber, 1986). In the Pleistocene, Guyanan and Brazilian forest refugia may have permitted the survival and differentiation resulting in a Brazilian Clade. However, the current Darien Forest between Panama and Colombia does not appear to be a barrier separating populations in Central America from those in Colombia and Venezuela. The existence of haplotypes in the Colombian and Venezuelan Andes that are closely related to Central American populations, suggests past secondary contact between two principal refugia: the Darien Forest (western Colombia and Panama) and the Choco Forest and/or Andalucia Pass in southern Colombia.

7. Conclusion

Analysis of nucleotide diversity in the mitochondrial genome of *L. longipalpis* revealed the existence of four clades in this taxon. These are well resolved by phylogenetic reconstruction (Fig. 2) and separated by high levels of genetic divergence (Table 4, Fig. 3). Three of these clearly represent phylogenetic species, supported with bootstrap values >80% (Fig. 2). The Andean Clade contains two internal clades, cis- and trans-Andean, with only slightly weaker bootstrap support (75–78%). The deep phylogenetic discontinuity and large geographic gap separating the Laran and Brazilian clades suggests that their most recent ancestor was a widespread sub-Andean–Amazonian gene pool. The large genetic distance between the Laran, Andean and Brazilian Clades supports vicariance as the most significant force in their evolution. The occurrence, in sympatry, of haplotypes belonging to the cis- and trans-Andean Clades at Bucaramanga and Trujillo, suggest these clades are not monophyletic. Their history likely involved a period of isolation, followed by a period of dispersal across the boundary separating the two. The lack of identical (in state) haplotypes shared by the two suggests that there is no contemporary gene flow between them, with isolation resulting from vicariant microevens. This is further supported by the demonstration of post-mating reproductive isolation, in the form of hybrid male sterility, in laboratory crosses between members of the cis- and trans-Andean Clades (Lanzaro et al., 1993).

Acknowledgements

We thank the following individuals for assisting in the collecting of the insects used in this study: Drs. John-Paul Matebi and Ed Rowton (USA); Bruce Alexander, Elizabeth Rangel, Italo Sherlock, Adelson Souza, Jose Wellington and Jeffrey Shaw (Brazil); Carlos Ponce (Honduras); Marco Herrero (Costa Rica); Milagros Oviedo (Venezuela), James Montoya (Colombia). We thank Juan Carlos Navarro (UCV-Venezuela) for advice in the phylogenetic analysis, Roger Perez-Hernandez (UCV-Venezuela), Eduardo Domínguez (UNT-Argentina); Professor Machado-Allison (UCV-Venezuela) and Frederic Triplet (UTMB) for helpful revisions to earlier versions of the manuscript. This research was supported, in part, by a grant from the John D. and Catherine T. MacArthur Foundation program entitled “Molecular Biology of Parasite Vectors” and by grant AI39540 from the National Institutes of Health (USA) to GCL and a fellowship from TDR-WHO to JCA.
References


Hammen, T., Gotakazi, E., 1964. A pollen diagram from the quaternary of the savanna Borgota (Colombia) and its significance for the geology of the Northern Andes. Geologie en Mijnbouw 43, 113–117.


