Molecular phylogenetics and evolution of host plant use in the Neotropical rolled leaf ‘hispine’ beetle genus *Cephaloleia* (Chevrolat) (Chrysomelidae: Cassidinae)

Duane D. McKenna *, Brian D. Farrell

Harvard University, Museum of Comparative Zoology, 26 Oxford Street, Cambridge, MA 02138, USA

Received 22 October 2004; revised 20 May 2005
Available online 27 July 2005

Abstract

Here, we report the results of a species level phylogenetic study of *Cephaloleia* beetles designed to clarify relationships and patterns of host plant taxon and tissue use among species. Our study is based on up to 2088 bp of mtDNA sequence data. Maximum parsimony, maximum likelihood, and Bayesian methods of phylogenetic inference consistently recover a monophyletic *Cephaloleia* outside of a basal clade of primarily palm feeding species (the ‘Arecaceae-feeding clade’), and *C. irregularis*. In all three analyses, the ‘Arecaceae-feeding clade’ includes *Cephaloleia* spp. with unusual morphological features, and a few species currently placed in other cassidine genera and tribes. All three analyses also recover a clade that includes all Zingiberales feeding *Cephaloleia* and most *Cephaloleia* species (the ‘Zingiberales-feeding clade’). Two notable clades are found within the ‘Zingiberales-feeding clade.’ One is comprised of beetles that normally feed only on the young rolled leaves of plants in the families Heliconiaceae and Marantaceae (the ‘Heliconiaceae & Marantaceae-feeding clade’). The other is comprised of relative host tissue generalist, primarily Zingiberales feeding species (the ‘generalist-feeding clade’). A few species in the ‘generalist-feeding clade’ utilize Cyperaceae or Poaceae as hosts. Overall, relatively basal *Cephaloleia* (e.g., the ‘Arecaceae clade’) feed on relatively basal monocots (e.g., Cyclanthaceae and Arecaceae), and relatively derived *Cephaloleia* (e.g., the ‘Zingiberales-feeding clade’) feed on relatively derived monocots (mostly in the order Zingiberales). Zingiberales feeding and specialization on young rolled Zingiberales leaves have each apparently evolved just once in *Cephaloleia*.

Keywords: Cassidinae; Cephaloleia; Chrysomelidae; Hispinae; mtDNA; Zingiberales

1. Introduction

“The plant herbivore “interface” may be the major zone of interaction responsible for generating terrestrial organic diversity (Ehrlich and Raven, 1964).” Several ecological hypotheses have been proposed to explain how plant/insect interactions affect diversification (e.g., Berenbaum, 1983; Futuyma and Moreno, 1988; Thomson, 1994), for example, as a function of the physical environment, spatial distribution of resources, competition for resources, or limitations to dispersal. The advent of modern (especially molecular) phylogenetic studies has occasioned complementary, explicitly historical approaches to the study of plant/insect diversification (Farrell, 1998, 1999; Mitter et al., 1988; Page, 1994). Nevertheless, most such studies have focused on temperate insects, especially those feeding on conifers and dicots (Farrell et al., 2001; Jordal et al., 2000; Kelley and Farrell, 1998; Kelley et al., 2000; Normark et al., 1999; Sequeira et al., 2000; Sequeira and Farrell, 2001), with comparatively little study of tropical insect herbivores or monocot associates. Here, we address this deficiency by
examining the phylogenetic relationships among monocot feeding Neotropical beetles in the genus *Cephaloleia* (Chevrolat) (Chrysomelidae: Cassidinae).

With more than 200 described species, *Cephaloleia* is one of the most species rich genera of monocot feeding leaf beetles (Staines, 1996, 2004). Members of the genus have been the subjects of ecological study for more than 25 years (Johnson, 2004; Seifert and Seifert, 1976; Strong, 1977a,b, 1982). *Cephaloleia* feed only on Monocotyledo- nae, especially the young rolled leaves of plants in the order Zingiberales, and the young folded leaves of various Arecales. Other *Cephaloleia* host plants include Bromeliaceae, Cyclanthaceae, Cyperaceae, Orchidaceae, and Poaceae (Table 1) (D. McKenna, unpublished data; D. Windsor pers. comm., 2002; Staines, 1996, 2004). Fossil evidence suggests that *Cephaloleia*-like beetles have maintained specialized interactions with their Zingiberales host plants for more than 66 Ma (Wilf et al., 2000).

2.1. Taxa

2.1.1. Taxon sampling and specimen identification

We sampled 105 taxa (98 ingroup, 7 outgroup) including 75 a priori designated *Cephaloleia* species for this study (Table 1). Central American *Cephaloleia* were sampled most thoroughly since they have been the subject of most interest in the genus, and because it was possible to identify many of them following Staines (1996). The omission of *Cephaloleia* spp. from the Atlantic coastal forests of South America is unfortunate. Efforts to obtain specimens from this region for DNA sequencing were unsuccessful. However, if morphology and host plant affiliations are reliable indicators of phylogeny, species endemic to this region are closely related to species we sampled elsewhere.

We used the descriptions and keys in Staines (1996) to identify Central American *Cephaloleia* species. Thus, the taxon sometimes called *Demotispa lata* Baly is here called *Cephaloleia lata* Baly (Staines, 1996). No comprehensive keys are available for South American *Cephaloleia*, so we primarily used museum specimens (including types), and Staines (1996) when possible, for identification. Based on field and morphological studies, several species in Staines (1996) were identified that included one or more potentially undescribed cryptic species. We included many of these a priori designated potentially unnamed taxa in this study, indicated by the Latin abbreviation cf. Species that are clearly over split are not explicitly indicated. Representatives of such taxa were included in this study (e.g., *C. championi* and *C. leuco- xantha*). Several species are referred to by their respective DNA codes because they are undescribed, or because there are no reliable taxonomic revisions available to facilitate their identification (Table 1). Many of the specimens sequenced from the Arthropods of La Selva collection (ALAS) and from the Costa Rican National Biodiversity Institute (INBIO) were annotated by Staines (Edgewater, MD), an expert on *Cephaloleia*. Borowiec (University of Wroclaw) confirmed the identification of *Pseudostilpnaspis columbica* Borowiec, and McKenna made all other identifications. Voucher specimens are housed at the Harvard University Museum of Comparative Zoology (MCZ), Cambridge, Massachusetts, USA.

2.1.2. Outgroup choice

Phylogenetic relationships within and between cassidine tribes remain mostly unclear or unknown, so we selected outgroup taxa from several tribes: *Alurnus ornatus* Baly (Alurnini), *Chelobasis perplexa* Baly (Arescini), *Demotispa* sp. 175 (*Cephaloleia*), *Imatidium* cf. *rufiventre* Boheman (Imatidini), *Prosopodonta limburga* Baly (Prosopodontini), and *Pseudostilpnaspis columbica*
Table 1

Specimen information

<table>
<thead>
<tr>
<th>Species</th>
<th>DNA code</th>
<th>COIa</th>
<th>tRNAleua</th>
<th>COIIa</th>
<th>Cyt b</th>
<th>Plant collected from</th>
<th>Collector</th>
<th>Date</th>
<th>Locality</th>
</tr>
</thead>
<tbody>
<tr>
<td>adusta Uhmann</td>
<td>057</td>
<td>1–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>1–449</td>
<td><em>Heliconia</em> sp. (Heliconiaceae)</td>
<td>DM</td>
<td>15 July 2002</td>
<td>Carate, Costa Rica</td>
</tr>
<tr>
<td>alternans Waterh.</td>
<td>082</td>
<td>277–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>0</td>
<td><em>Heliconia</em> sp.</td>
<td>DW</td>
<td>23 August 1997</td>
<td>Lita, Ecuador</td>
</tr>
<tr>
<td>bella Baly</td>
<td>017</td>
<td>1–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>0</td>
<td><em>Heliconia imbricata</em></td>
<td>DM</td>
<td>16 July 2002</td>
<td>Corcovado NP, Costa Rica</td>
</tr>
<tr>
<td>belti Baly</td>
<td>095</td>
<td>1–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>21–449</td>
<td><em>Heliconia</em> sp.</td>
<td>DW</td>
<td>12 April 2003</td>
<td>La Fortuna, Panama</td>
</tr>
<tr>
<td>belti Baly</td>
<td>117</td>
<td>894–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>0</td>
<td><em>Heliconia</em> sp.</td>
<td>DM</td>
<td>23–30 June 2001</td>
<td>Braulio Carrillo NP, Costa Rica</td>
</tr>
<tr>
<td>cf. belti Baly</td>
<td>134</td>
<td>859–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>0</td>
<td><em>Heliconia</em> sp.</td>
<td>DM</td>
<td>May 2002</td>
<td>Monteverde, Costa Rica</td>
</tr>
<tr>
<td>congneri Baly</td>
<td>141</td>
<td>856–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>0</td>
<td><em>Heliconia secunda</em></td>
<td>DM</td>
<td>January 2004</td>
<td>Braulio Carrillo NP, Costa Rica</td>
</tr>
<tr>
<td>cyanec Staines</td>
<td>098</td>
<td>859–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>25–381</td>
<td>Unknown</td>
<td>DW</td>
<td>10 April 2003</td>
<td>La Fortuna, Panama</td>
</tr>
<tr>
<td>deficiens Uhmann</td>
<td>001</td>
<td>1–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>1–449</td>
<td><em>Costus</em> sp. (Costaceae)</td>
<td>DM</td>
<td>2–6 April 2003</td>
<td>OTS La Selva, Costa Rica</td>
</tr>
<tr>
<td>deficiens Uhmann</td>
<td>150</td>
<td>859–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>0</td>
<td><em>Costus</em> sp.</td>
<td>DM</td>
<td>11 March 2003</td>
<td>La Virgen, Costa Rica</td>
</tr>
<tr>
<td>deficiens Uhmann</td>
<td>156</td>
<td>858–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>0</td>
<td><em>Costus</em> sp.</td>
<td>ALAS</td>
<td>8 April 2003</td>
<td>Southeast of La Virgen, Costa Rica</td>
</tr>
<tr>
<td>dilatocollis Baly</td>
<td>031</td>
<td>532–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>0</td>
<td><em>Calathea</em> sp.</td>
<td>DM</td>
<td>October 2002</td>
<td>Darien Prov., Panama</td>
</tr>
<tr>
<td>cf. dilatocollis Baly</td>
<td>092</td>
<td>286–953, 1006–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>0</td>
<td><em>Renealmia</em> sp.</td>
<td>DW</td>
<td>16 August 1997</td>
<td>Yasuni NP, Ecuador</td>
</tr>
<tr>
<td>distincta Baly</td>
<td>053</td>
<td>1–1044</td>
<td>1045–1102</td>
<td>1103–1599</td>
<td>0</td>
<td><em>Heliconia</em> sp.</td>
<td>DM</td>
<td>October 2002</td>
<td>Old Gamboa Rd., Panama</td>
</tr>
<tr>
<td>cf. dorsalis Baly</td>
<td>119</td>
<td>891–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>0</td>
<td><em>Renealmia</em> sp.</td>
<td>DW</td>
<td>7 August 1997</td>
<td>Ecuador</td>
</tr>
<tr>
<td>erichsonii Baly</td>
<td>008</td>
<td>1–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>1–449</td>
<td><em>Calathea gymnocarpa</em></td>
<td>DM</td>
<td>2–6 April 2003</td>
<td>OTS La Selva, Costa Rica</td>
</tr>
<tr>
<td>erichsonii Baly</td>
<td>040</td>
<td>1–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>1–449</td>
<td><em>Calathea</em> sp.</td>
<td>DM</td>
<td>16 July 2002</td>
<td>Corcovado NP, Costa Rica</td>
</tr>
<tr>
<td>cf. erichsonii Baly</td>
<td>010</td>
<td>286–1044</td>
<td>1045–1102</td>
<td>1103–1631</td>
<td>1–449</td>
<td><em>Calathea</em> sp. or <em>Calathea</em> sp.</td>
<td>SR</td>
<td>15 January 2003</td>
<td>La Virginia, Risaralda, Colombia</td>
</tr>
<tr>
<td>exigua Uhmann</td>
<td>023</td>
<td>274–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>1–449</td>
<td><em>Calanthes</em> bipartitus (Cyclanthaceae)</td>
<td>DM</td>
<td>October 2002</td>
<td>Fort Sherman, Panama</td>
</tr>
</tbody>
</table>

(continued on next page)
<table>
<thead>
<tr>
<th>Species</th>
<th>DNA code</th>
<th>COI&lt;sup&gt;a&lt;/sup&gt;</th>
<th>tRNA&lt;sub&gt;leu&lt;/sub&gt;&lt;sup&gt;a&lt;/sup&gt;</th>
<th>COII&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Cyt b&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Plant collected from</th>
<th>Collector</th>
<th>Date</th>
<th>Locality</th>
</tr>
</thead>
<tbody>
<tr>
<td>cf. fenestrata Weise</td>
<td>2146</td>
<td>2589–1024</td>
<td>0</td>
<td>0</td>
<td>1–440</td>
<td>Pleiostachya pruinosa</td>
<td>DM</td>
<td>15 July 2002</td>
<td>Puerto Jimenez–Carate Rd., Costa Rica</td>
</tr>
<tr>
<td>flava Uhmann</td>
<td>065</td>
<td>10–1044</td>
<td>1045–1102</td>
<td>q</td>
<td>0</td>
<td>Unknown</td>
<td>DM</td>
<td>25 April 2002</td>
<td>Monteverde, Costa Rica</td>
</tr>
<tr>
<td>cf. falcicollis Weise</td>
<td>049</td>
<td>1–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>0</td>
<td>Heliconia sp.</td>
<td>DM</td>
<td>16 July 2002</td>
<td>Corcovado NP, Costa Rica</td>
</tr>
<tr>
<td>fulvilineata Baly</td>
<td>126</td>
<td>1–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>0</td>
<td>Unknown</td>
<td>ALAS</td>
<td>17 April 2003</td>
<td>Southeast of La Virgen, Costa Rica</td>
</tr>
<tr>
<td>gilvipes Uhmann</td>
<td>186</td>
<td>856–1044</td>
<td>1045–1102</td>
<td>1103–1637</td>
<td>0</td>
<td>Unknown</td>
<td>INBIO</td>
<td>22 March–9 April 2002</td>
<td>Vara Blanca, Costa Rica</td>
</tr>
<tr>
<td>heliconiae Uhmann</td>
<td>097</td>
<td>1–1044</td>
<td>1045–1102</td>
<td>1103–1615</td>
<td>0</td>
<td>Calathea sp.</td>
<td>KN</td>
<td>October 2001</td>
<td>La Fortuna, Panama</td>
</tr>
<tr>
<td>hisstro Guérin</td>
<td>011</td>
<td>286–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>53–449</td>
<td>Heliconia sp. or Calathea sp.</td>
<td>SR</td>
<td>5 January 2003</td>
<td>La Virginia, Risaralda, Colombia</td>
</tr>
<tr>
<td>lata Baly</td>
<td>005</td>
<td>1–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>1–449</td>
<td>Chamaedorea tepejilote (Arecaceae)</td>
<td>DM</td>
<td>2–6 April 2003</td>
<td>OTS La Selva, Costa Rica</td>
</tr>
<tr>
<td>marginella Uhmann</td>
<td>096</td>
<td>1–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>0</td>
<td>Heliconia sp.</td>
<td>DM</td>
<td>12 April 2003</td>
<td>La Fortuna, Panama</td>
</tr>
<tr>
<td>mauliki Uhmann</td>
<td>064</td>
<td>1–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>0</td>
<td>Unknown</td>
<td>KN</td>
<td>17 April 2003</td>
<td>OTS La Selva, Costa Rica</td>
</tr>
<tr>
<td>metallescens Baly</td>
<td>182</td>
<td>1006–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>0</td>
<td>Unknown</td>
<td>KN</td>
<td>17 April 2003</td>
<td>Southeast of La Virgen, Costa Rica</td>
</tr>
<tr>
<td>nigricornis (Fab.)</td>
<td>143</td>
<td>902–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>0</td>
<td>Unknown</td>
<td>KN</td>
<td>17 April 2003</td>
<td>Southeast of La Virgen, Costa Rica</td>
</tr>
<tr>
<td>nigropicta Baly</td>
<td>2316</td>
<td>259–991</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>Heliconia latispatha</td>
<td>DM</td>
<td>1–10 August 2001</td>
<td>Pipeline Rd., Panama</td>
</tr>
<tr>
<td>partita Weise</td>
<td>020</td>
<td>17–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>0</td>
<td>Heliconia latispatha</td>
<td>DM</td>
<td>1–10 August 2001</td>
<td>Pipeline Rd., Panama</td>
</tr>
<tr>
<td>partita Weise</td>
<td>077</td>
<td>1–1044</td>
<td>1045–1102</td>
<td>1103–1605</td>
<td>0</td>
<td>Heliconia latispatha</td>
<td>DM</td>
<td>3 August 2002</td>
<td>Old Gamboa Rd., Panama</td>
</tr>
<tr>
<td>placida Baly</td>
<td>055</td>
<td>1–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>12–371</td>
<td>Heliconia sp. or Calathea sp.</td>
<td>SR</td>
<td>5 January 2003</td>
<td>La Virginia, Risaralda, Colombia</td>
</tr>
<tr>
<td>pretiosa Baly</td>
<td>009</td>
<td>1–1044</td>
<td>1045–1102</td>
<td>1103–1645</td>
<td>1–449</td>
<td>Heliconia sp.</td>
<td>SR</td>
<td>5 January 2003</td>
<td>La Virginia, Risaralda, Colombia</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Collector(s)</th>
<th>Plant identifications</th>
<th>COI position</th>
<th>COII position</th>
<th>Outgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arthropods of La Selva Project (ALAS), D. McKenna (DM), Costa Rican Institute for Biodiversity (INBIO), S. Ramirez, Harvard University Museum of Comparative Zoology (SR), M. Braby, R. Eastwood, and K. Nishida, Harvard University Museum of Comparative Zoology (MB/RE/KN), and D. Windsor, STRI (DW). Plant identifications were provided by the collector(s) or were made by DM.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a COI position 1 is located at position 1470 in the Drosophila yakuba mtDNA (Clary and Wolstenholme, 1985). We typically amplified a single fragment (s1859–a3661) containing consecutively, the last part of COI, the entire tRNALeu, and the first part of COII. Therefore, these gene regions are reported here using a single range of numbers 1–1645. The number of nucleotides reported for tRNA-Leu includes the six nucleotide ambiguously aligned region that was excluded from all analyses.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b Cyt b position 1 is located at position 10,923 in the D. yakuba mtDNA (Clary and Wolstenholme, 1985).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Collector(s)</th>
<th>Plant identifications</th>
<th>COI position</th>
<th>COII position</th>
<th>Outgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arthropods of La Selva Project (ALAS), D. McKenna (DM), Costa Rican Institute for Biodiversity (INBIO), S. Ramirez, Harvard University Museum of Comparative Zoology (SR), M. Braby, R. Eastwood, and K. Nishida, Harvard University Museum of Comparative Zoology (MB/RE/KN), and D. Windsor, STRI (DW). Plant identifications were provided by the collector(s) or were made by DM.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a COI position 1 is located at position 1470 in the Drosophila yakuba mtDNA (Clary and Wolstenholme, 1985). We typically amplified a single fragment (s1859–a3661) containing consecutively, the last part of COI, the entire tRNALeu, and the first part of COII. Therefore, these gene regions are reported here using a single range of numbers 1–1645. The number of nucleotides reported for tRNA-Leu includes the six nucleotide ambiguously aligned region that was excluded from all analyses.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b Cyt b position 1 is located at position 10,923 in the D. yakuba mtDNA (Clary and Wolstenholme, 1985).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Collector(s)</th>
<th>Plant identifications</th>
<th>COI position</th>
<th>COII position</th>
<th>Outgroups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arthropods of La Selva Project (ALAS), D. McKenna (DM), Costa Rican Institute for Biodiversity (INBIO), S. Ramirez, Harvard University Museum of Comparative Zoology (SR), M. Braby, R. Eastwood, and K. Nishida, Harvard University Museum of Comparative Zoology (MB/RE/KN), and D. Windsor, STRI (DW). Plant identifications were provided by the collector(s) or were made by DM.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a COI position 1 is located at position 1470 in the Drosophila yakuba mtDNA (Clary and Wolstenholme, 1985). We typically amplified a single fragment (s1859–a3661) containing consecutively, the last part of COI, the entire tRNALeu, and the first part of COII. Therefore, these gene regions are reported here using a single range of numbers 1–1645. The number of nucleotides reported for tRNA-Leu includes the six nucleotide ambiguously aligned region that was excluded from all analyses.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b Cyt b position 1 is located at position 10,923 in the D. yakuba mtDNA (Clary and Wolstenholme, 1985).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.2. Laboratory procedures

2.2.1. DNA procedures

Most specimens used in this study were collected in the field from host plants, but a few had been previously pinned and dried (Table 1). Total genomic DNA was extracted from the abdomen, legs (1–3), or the entire specimen, using a QIAquick DNeasy Tissue Kit (Qiagen, Valencia, CA). We used 50 μL volume PCRs comprised of 37.35 μL water, 5 μL of 5× buffer (Qiagen Inc.), 0.4 μL of 10 mM dNTP (Qiagen Inc.), 1 μL of each 10 mM primer, 3 μL of 2.5 mM MgCl₂ (Qiagen Inc.), 0.25 μL of Taq DNA Polymerase (Qiagen Inc.), and approximately 100 ng of genomic DNA template per 25 μL PCR volume.

To amplify COI, tRNA-Leu, and COII, we primarily used the following PCR program: (1) an initial denaturation of 5 min at 94 °C; (2) 30 s at 94 °C denaturation, 30 s at 49 °C annealing, and 1:30 min at 72 °C extension (40 ×); and (3) a final extension of 5 min at 72 °C. We occasionally used a program that differed from the above by having an initial denaturation of 1 min at 94 °C, and 30 s at 47 °C annealing. To amplify Cyt b, we primarily used the following program: (1) an initial denaturation of 5 min at 94 °C; (2) 30 s at 94 °C denaturation, 1:30 min at 58 °C annealing, and 1 min at 72 °C extension (2 ×); (3) 1 min each at 56 °C annealing and 72 °C extension (repeated twice at 2 °C annealing increments from 58–44 °C); (4) 1 min each at 42 °C annealing, and 72 °C extension (×18); and (5) a single final extension of 5 min at 72 °C. Amplified PCR products were cleaned with a QIAquick PCR Purification Kit (Qiagen Inc.), or were gel purified using a Qiagen QIAquick Gel Purification Kit.

Sequencing was performed using ABI PRISM BigDye Terminator Cycle Sequencing Kits (versions 3.0 and 3.1) (Applied Biosystems). The same primers were used for amplification and sequencing (Table 2). We designed some primers specifically for use in the genus Cephaloleia (using Oligo Primer Analysis Software version 4.05 (Long Lake, MN)), but in most cases the universal primers worked most reliably across taxa. All sequencing was performed on an ABI PRISM 3100 Genetic Analyzer.

2.2.2. DNA sequence data

We targeted an approximately 1800-base pair (bp) fragment (primers s1859–a3661) for amplification that included (1) part of the mtDNA COI gene, (2) the entire tRNA-Leu, and (3) a portion of the mtDNA COII gene (Table 2). We separately amplified an approximately 450 bp fragment of the mtDNA Cyt b gene (primers CB1, CytB B.1, CBdms, or CB1c, to CB2). We included taxa in our analyses even when they were represented by only a subset of the potential total sequence data (Table 1). This approach is supported by simulation studies which have shown that even highly incomplete taxa can be accurately placed in combined analyses with sufficient phylogenetically informative characters (Wiens, 1998, 2003a,b). Sequence for Criceris duodecimpunctata (L.) was obtained from GenBank (Accession No. AF467886 (Stewart and Beckenbach, 2003)). All DNA sequences were obtained from GenBank (Clary and Wolstenholme, 1985).
Table 3
Characteristics of the mtDNA regions sequenced, including the number of taxa in each analysis, the total number of characters (excluding the six
nucleotide ambiguous region excluded from analyses), the number of parsimony informative and variable characters, and the number of equally par-
simonious trees

<table>
<thead>
<tr>
<th>Gene</th>
<th>Total number of taxa</th>
<th>Total number of characters</th>
<th>Parsimony informative characters</th>
<th>Variable characters</th>
<th>MP tree length</th>
<th>Number of equally parsimonious trees</th>
</tr>
</thead>
<tbody>
<tr>
<td>COI</td>
<td>105</td>
<td>1044</td>
<td>473</td>
<td>539</td>
<td>4455</td>
<td>&gt;50,000</td>
</tr>
<tr>
<td>tRNA-Leu</td>
<td>101</td>
<td>52</td>
<td>16</td>
<td>25</td>
<td>91</td>
<td>&gt;50,000</td>
</tr>
<tr>
<td>COII</td>
<td>100</td>
<td>543</td>
<td>298</td>
<td>335</td>
<td>3418</td>
<td>27</td>
</tr>
<tr>
<td>Cyt b</td>
<td>57</td>
<td>449</td>
<td>233</td>
<td>247</td>
<td>1944</td>
<td>4</td>
</tr>
<tr>
<td>Combined</td>
<td>105</td>
<td>2088</td>
<td>1020</td>
<td>1156</td>
<td>10,400</td>
<td>162</td>
</tr>
</tbody>
</table>

were deposited in GenBank under Accession Nos. DQ026066–DQ026225.

2.2.3. Alignment of nucleotide sequences and preliminary sequence analysis

Protein coding sequences were aligned by eye using Sequencher 3.1.1 (GeneCodes Corporation, 1999) and viewed using MacClade 4.03 (Maddison and Maddison, 2001). A six-nucleotide region at the 3’ end of tRNA-Leu could not be unambiguously aligned and was excluded from subsequent analyses. Phylogenetic analyses were based on the remaining 2088 bp of aligned nucleotide data. Gaps were treated as missing data in all analyses.

We explored potential incongruence among the four mtDNA fragments using the incongruence length difference (ILD) test (Farris et al., 1994) (100 replications) implemented as the partition homogeneity test in PAUP*4.03b10 (PAUP) (Swofford, 2001). Uninformative sites were excluded from analysis, and we limited branch swapping to 1000 trees per replicate (Lee, 2001). The ILD test identified significant incongruence among the four data partitions (P = 0.01). However, we view this result with some skepticism due to differences in taxon sampling, missing data, and numbers of variable and phylogenetically informative characters among data partitions (Table 3) (Dowton and Austin, 2002). Further, empirical (Yoder et al., 2001) and simulated data (Barker and Lutzoni, 2002) have shown that the ILD test can fail to allow combination of data partitions when they should be combined. Finally, the regions sequenced form a single linkage group (mtDNA). Therefore, despite a significant ILD test, we combined the four mtDNA partitions and analyzed them together in subsequent analyses.

2.3. Phylogenetic analyses

Initial phylogenetic analyses were conducted using MP criteria in PAUP. Equally weighted heuristic tree searches were performed on the combined data using 1000 random sequence additions and tree bisection-reconnection (TBR) branch swapping. The parsimony ratchet procedure (Nixon, 1999) was then implemented in PAUP using 200 replicates and repeated with 10–25% weighted characters using batch files generated by PAUP
PRat version 1 (Sikes and Lewis, 2001). The MP tree(s) generated from the parsimony ratchet procedure were then used to start another equal weights heuristic tree search in PAUP. Nodal support was evaluated with 1000 non-parametric bootstrap pseudoreplicates, using a simple addition sequence of taxa and TBR branch swapping (Felsenstein, 1985; Hillis and Bull, 1993) in PAUP. Bremer support values were obtained using a command file of constraint trees generated by TreeRot version 2 (Sorenson, 1999).

Bayesian Markov Chain Monte Carlo methods were also used to estimate phylogeny in the program MrBayes version 3.0b4 (MrBayes) (Huelsenbeck and Ronquist, 2001; Larget and Simon, 1999; Rannala and Yang, 1996). The simplest best-fit substitution model was selected for each data partition with Modeltest version 3.5 (Posada and Crandall, 1998) using hierarchical likelihood ratio tests (LRT) and Akaike information criterion (AIC) (Posada and Crandall, 2001). Both methods selected the GTR+I+G model for COI. For tRNA-Leu, the LRT selected the F81+I+G model, and the AIC selected the TVMef+I+G model. For COII, both methods selected the GTR+I+G model, and for Cyt b the LRT selected the TVM+I+G model, and AIC selected the GTR+I+G model. For the combined data set, both methods selected the GTR+I+G model.

We ran four separate analyses in MrBayes using the GTR+I+G model, and starting with random trees generated by the program defaults. We allowed MrBayes to estimate parameter values separately for each data partition (COI, tRNA-Leu, COII, and Cyt b). Three heated and one cold chain was used in all analyses. We ran each analysis for 1.0 x 10^6 generations, sampling every 100 generations. A single additional analysis was run for 2.0 x 10^6 generations. We evaluated the log likelihood scores from each of the five runs to see if and when stationarity was reached and to evaluate convergence of log likelihood scores across runs. To avoid overrepresentation of trees from a single run, the trees obtained from the 2.0 x 10^6 generation analysis were used only for diagnosing convergence of log likelihood scores and stationarity across runs. We discarded all samples preceding stationarity as a “burn in.” The post “burn in” trees saved from each of the four, 1.0 x 10^6 generation
runs were combined and used to generate a 50% majority rule consensus tree in PAUP, and a 95% credible set of trees.

Phylogeny was also estimated using ML as implemented in PAUP. We applied a successive approximations approach, similar to that of Lin et al. (2004), to search for the ML tree (Swofford et al., 1996). A heuristic search with TBR branch swapping was started using the Bayesian consensus as the starting tree, the GTR+I+G substitution model selected by Modeltest, and the initial parameter estimates obtained from the Bayesian consensus. The ML parameters were optimized for each of the iterations on the new tree, and the search was repeated with the optimized parameters fixed.

3. Results

3.1. Sequence alignment

The combined mtDNA data set comprised a total of 2094 nucleotide sites, including a maximum of 1044 sites from the 3' end of COI (including a single three nucleotide insert at the 3' end of COI at position 1033 in the alignment, present only in the outgroup Chelobasis perplexa), the complete tRNA-Leu (maximum 58 nucleotide sites total; 52 sites excluding those whose alignment was ambiguous), a maximum 543 nucleotide sites from the 5' end of COII, and a maximum 449 nucleotide sites from near the middle of Cyt b (Table 1). Of the 2088 total nucleotide sites included in analyses, 1156 sites (55%) were variable and 1020 sites (49%) were parsimony informative (Table 3).

3.2. Maximum parsimony phylogenetic analyses

Maximum parsimony analysis yielded 162 trees of length 10,400 (Table 3). The shortest trees found by the parsimony ratchet were also of length 10,400. When the parsimony ratchet trees were used to start an MP heuristic search, the same 162 trees were recovered as when starting with random trees. Overall, relationships are well resolved in the MP strict consensus (Fig. 1). The MP tree recovers a monophyletic Cephaloleia but without robust bootstrap support (<50%), as long as Demotispa sp. 175, and Pseudostilpnspsis columbica are included. Most species relationships are resolved and supported by moderate (>75%) to high (>95%) bootstrap values. Higher-level relationships are generally supported by low bootstrap (<50%) and moderate Bremer support values. Summary statistics resulting from separate MP analysis of the four regions sequenced, and from the combined data set are provided in Table 3. Cephaloleia belti, C. deficiens, C. cf. erichsonii, C. gilvipes, and C. trivittata are polyphyletic in the MP tree. Cephaloleia dilatata and C. erichsonii are paraphyletic. The unexpected placement of C. semivittata in a clade bounded by C. reventazonica and C. pulchella is not well supported (<50% bootstrap value). Demotispa sp. 175 and Pseudostilpspsis columbica are recovered in a basal clade with several unique Cephaloleia spp. Cephaloleia irregularis is resolved in a position that is basal to all other Cephaloleia.

3.3. Model selection and Bayesian inference of phylogeny

The best-fit substitution model for the combined data set selected with the LRT and AIC in Modeltest was the GTR+I+G (log likelihood = 43610.348). The parameter values estimated by Modeltest were A=C: 0.29, A=G: 8.88, A=T: 0.40, C=G: 1.81, C=T: 4.76, and G=T: 1.0. Estimated base composition was A = 0.39, C = 0.16, G = 0.04, T = 0.41, the proportion of invariant sites = 0.38, and the x values of the γ shape distribution = 0.49. The log likelihood scores from each run converged on approximately the same value and were stable after approximately 3.0 × 10^7 generations. The first 5000 trees from each run were discarded as a conservative “burn in.” The remaining 5000 trees from each of the four 1.0 × 10^6 generation runs were pooled for a total of 20,000 trees and used to generate a 50% majority rule consensus tree (Fig. 2). The 95% credible set of post “burn in” trees contained 10,075 trees.

The Bayesian majority rule consensus tree recovers most of the same clades as the MP tree (Fig. 2). Forty-four nodes have ≥70% bootstrap support and ≥0.95 BPP. Several taxa differ in placement in the two trees, most notably, C. dilaticollis, C. cf. dilaticollis, C. irregularis, C. cf. pulchella, C. semivittata, and Imatidium cf. ruficentre. Bootstrap values for most of these placements are low, and Bayesian posterior probabilities (BPP) mostly non-significant (<0.95). The MP and Bayesian trees otherwise differ primarily in resolution and the arrangement of tips within major clades. The Bayesian tree recovered the same polyphyletic and paraphyletic taxa as the MP tree.

3.4. ML phylogenetic analyses implemented in PAUP

One tree with a log likelihood of 43188.323 resulted from the ML analysis of the combined data set in PAUP (applying the GTR+I+G model) after branch swapping with exceedingly little improvement (<0.0005%) of log likelihood score from 43209.203 (Bayesian consensus tree) (Fig. 1). The parameter values estimated for our final ML tree were A=C: 0.35, A=G: 8.97, A=T: 0.48, C=G: 1.77, C=T: 5.83, and G=T: 1.0. Estimated base composition was A = 0.40, C = 0.15, G = 0.04, T = 0.41, the proportion of variable sites = 0.37, and the x values of the γ shape distribution = 0.46.
Fig. 1. Strict consensus of 162 equally parsimonious trees of length 10,400 steps generated by PAUP from the combined data set with equal weights. Bootstrap and Bremer support values are shown above and below branches, respectively. Bootstrap values <50% are not shown.
Fig. 2. Fifty percent majority rule consensus tree generated from the 20,000 trees retained from the four separate $1.0 \times 10^6$ generation Bayesian analyses of the combined data set based on the GTR+I+G substitution model in MrBayes. Numbers above branches are posterior probability values of the nodes (e.g., $1 = 100\%$, $0.95 = 95\%$, etc.). Values <0.5 are not shown.
Fig. 3. Maximum likelihood phylogram (log likelihood 43188.323) generated in PAUP from analysis of the combined data set based on the GTR+I+G substitution model.
The ML tree, like the Bayesian consensus, recovers a monophyletic *Cephaloleia* as long as *Demotispa* sp. 175, *Imatidium* cf. *ruftiventre*, and *Pseudostilpnaspis columbica* are included (Fig. 3). The ML tree recovered the same polyphyletic and paraphyletic taxa as the MP tree and the Bayesian consensus. Differences between the ML tree and the MP tree are mostly the same as those for the Bayesian tree and the MP tree (see above), except for the placement of the clade containing *C. cf. erichsonii*, *C. mauliki*, *C. nigricornis*, *C. placida*, and *C. cf. pulchella*.

3.5. Patterns of host tissue and host taxon usage

In all three analyses, all Zingiberales feeders and most *Cephaloleia* species form a single large derived clade (<50% bootstrap, 0.96 BPP) (the ‘Zingiberales-feeding clade’). Notable within the ‘Zingiberales-feeding clade’ is a clade of beetles that normally feed as adults only on the leaf rolls of plants in the tropical families Heliconiaceae and Marantaceae (the ‘Heliconiaceae & Marantaceae-feeding clade’). Bootstrap support for the ‘Heliconiaceae & Marantaceae-feeding clade’ is low (<50%), and BPP non-significant (0.63). Sister to the Heliconiaceae & Marantaceae-feeding clade is a clade of relative generalist Zingiberales feeders (the ‘generalist-feeding clade’) with low bootstrap support (<50%), and non-significant BPP (0.79) (Fig. 2). Most species in the ‘generalist-feeding clade’ feed on Zingiberales in the families Costaceae, Heliconiaceae, Marantaceae, and Zingiberaeaceae. *Cephaloleia ruficollis*, *C. tenella*, and *C. uhmanii* feed on Cyperaceae or Poaceae. A basal clade we refer to as the ‘Arecaceae-feeding clade’ (bootstrap value < 50%, BPP = 1.0) includes all Arecaceae feeding *Cephaloleia* and *C. exigua* (a Cyclanthaceae feeder). None of these species are known to feed on Zingiberales. The ‘Arecaceae-feeding clade’ includes several unusual *Cephaloleia* spp., and species currently placed in other cassidine genera in the tribes Cephaloleiini (*Demotispa* sp. 175) and Imatidini (*Imatidium* cf. *ruftiventre* and *Pseudostilpnaspis columbica*).

4. Discussion

4.1. Cephaloleia phylogeny

None of the 162 MP trees, the ML tree, or the 20,000 trees used to generate the Bayesian consensus tree, recovers a monophyletic *Cephaloleia*. We are not surprised that *Cephaloleia* is monophyletic only with the inclusion of *Demotispa* sp. 175 (Cephaloleiini), *Pseudostilpnaspis columbica* (Imatidini), and *Imatidium* cf. *ruftiventre* (Imatidini) (resolved within *Cephaloleia* only in the MP and Bayesian consensus trees), because a large number of *Demotispa* (and members of several other genera of Cephaloleiini), and at least a few *Pseudostilpnaspis* (and other Imatidini) feed on palms, much like most *Cephaloleia* spp. placed in the ‘Arecaceae-feeding clade’ in our analyses. However, relationships between basal Cephaloleiini and Imatidini remain little known. We are surprised that *Chelobasis perplexa* (Arescini) may be only distantly related to *Cephaloleia*. *Chelobasis* spp. (and other very closely related Arescini in the genera *Arescus*, *Nympharescus*, and *Xenarescus*) often co-occur with *Cephaloleia* spp. in the leaf rolls of plants in the family Heliconiaceae.

All three methods of analysis identify the same polyphyletic taxa (*C. belti, C. deficiens, C. cf. erichsonii, C. gilvipes, and C. trivittata*), and the same paraphyletic taxa (*C. dilaticollis* and *C. erichsonii*). All polyphyletic taxa are the result of cryptic species. The two polyphyletic taxa require further study. Moderate to high-bootstrap values and significant BPPs occur primarily near the tips of the tree. Where topologies differ, bootstrap values are generally low, and BPPs non-significant. A few deep nodes in the Bayesian consensus tree have relatively high BPPs (e.g., the ‘Zingiberales-feeding clade’), but lack robust bootstrap support. Bayesian posterior probabilities should be interpreted cautiously in such situations. Suzuki et al. (2002) showed that when analyzing concatenated DNA sequences, BPPs could be “excessively liberal.” Further, it should be noted that the Bayesian consensus and the ML tree are not entirely independent estimates of phylogeny, especially given our approach to the ML analysis (see Section 2 for details).

The conflicting placement of the clade containing *C. cf. erichsonii*, *C. mauliki*, *C. nigricornis*, *C. placida*, and *C. cf. pulchella* in the ML versus the MP and Bayesian trees is intriguing. All of the above species with known host plants feed on Zingiberales leaf rolls, suggesting that the most likely of potential affiliations is with the ‘Heliconiaceae and Marantaceae-feeding clade’ as in the MP and Bayesian consensus trees. The positions of *C. dilaticollis*, and *C. cf. dilaticollis* are likewise perplexing. Placement at the base of the ‘generalist-feeding clade’ as in the ML and Bayesian trees seems to make the most sense in light of the fact that both species are host tissue generalists on Zingiberales (especially Marantaceae).

4.2. Evolution of host taxon and host tissue use

Overall, the evolution of host taxon usage in *Cephaloleia* is remarkably consistent with current concepts of monocot phylogeny (Janssen and Bremer, 2004). For example, relatively basal *Cephaloleia* (members of the ‘Arecaceae-feeding clade’) feed on the relatively most basal host plant families (e.g., Cyclanthaceae and Arecales). Zingiberales are considered derived with respect to the aforementioned plant families, and this is reflected in the *Cephaloleia* phylogeny by placement of the ‘Zingiberales-feeding clade’ in a derived position relative to the ‘Arecaceae-feeding clade.’ Host taxon usage in the ‘Heliconiaceae & Marantaceae-feeding clade’ is limited to plants in the families Heliconiaceae and Marantaceae.
The ‘Heliconiaceae & Marantaceae-feeding clade’ contains the only Cephaloleia species that feed exclusively in the rolled leaves of Zingiberales. While patterns of host plant taxa and tissue use are not definitively known for all species included in our study, 100% of species in the ‘Heliconiaceae & Marantaceae-feeding clade’ with known hosts feed exclusively in the rolled leaves of Heliconiaceae and Marantaceae as adults. The vast majority of larvae are also limited to leaf rolls. Many species in the ‘generalist-feeding clade’ feed on Zingiberales; however, in contrast to the tissue specialized ‘Heliconiaceae & Marantaceae-feeding clade,’ the larvae and adults of Zingiberales feeding species in the ‘generalist-feeding clade’ utilize multiple host plant parts, including inflorescence bracts, petioles, and leaf rolls. None of the included species feed exclusively on leaf rolls. Most beetles in the ‘Arecaceae-feeding clade’ feed primarily on the immature folded leaves of their hosts, none of which are Zingiberales. Thus, association with Zingiberales, and specialization on Zingiberales leaf rolls have each apparently evolved just once in Cephaloleia.

4.3. Phylogenetic relationships among Cephaloleia species and implications for taxonomy

Whereas the ‘Arecaceae-feeding clade’ is clearly divergent from other Cephaloleia, the relationships among members of the ‘Arecaceae-feeding clade,’ and between members of the ‘Arecaceae-feeding clade’ and putative outgroups, remain somewhat unclear. Resolution is also limited within the ‘Zingiberales-feeding clade.’ Ongoing research, including sequencing of nuclear protein coding genes, reducing the amount of missing data, and additional taxon sampling (including Cephaloleia-like taxa in other genera such as Aslamidium) may help clarify some of these issues, and may provide more resolution and better nodal support, especially for deeper divergences.

Given the relative concordance of morphological and molecular genetic concepts of the genus Cephaloleia, we forego making any formal taxonomic changes until sequence data from additional taxa and from the nuclear genome are available to corroborate polyphyly of especially, the ‘Arecaceae-feeding clade.’ Nevertheless, the results of our work demonstrate novel relationships among Cephaloleia spp., and between Cephaloleia and select other Cassidinae, and provide a glimpse into the evolution of patterns of host plant taxon and tissue use in the genus.

Acknowledgments

We are indebted to the Organization for Tropical Studies (OTS), the ALAS Project, INBIO, D. Windsor and the Smithsonian Tropical Research Institute (STRI), J. Vandermeer (University of Michigan), J. Maes (Museo Entomológico, León, Nicaragua), and countless others, for logistic support for fieldwork in Central America and/or for providing specimens used in this study. Special thanks to K. McKenna who assisted with fieldwork, and C. Staines, E. Riley, and D. Windsor for enlightening discussions on “cassidoid Hispinae” and for providing information on host plant affinities. Don Windsor, L. Borowiec, and S. Boyce provided useful comments on the manuscript. The manuscript was further improved by the comments of five anonymous reviewers. This study was partially supported by an A.W. Mellon foundation OTS-STRI exploration award, a graduate research fellowship from OTS, a fellowship from the United States Environmental Protection Agency STAR program, a grant from the David Rockefeller Center for Latin American Studies at Harvard University, a grant from the Putnam Fund for Harvard University MCZ Expeditions, and by the Harvard University Department of Organismic and Evolutionary Biology.

References


Sorensen, M.D., 1999. TreeRot, version 2. Boston University, Boston, MA.


Strong, D.R., 1977a. Rolled-leaf hispine beetles (Chrysomelidae) and their Zingiberaleas host plants in Middle America. Biotropica 9, 156–169.


