THE PHYLOGENETIC POSITION OF CEROGLOSSUS OCHSENII GERMAIN AND CEROGLOSSUS GUERINI GERMAIN (COLEOPTERA: CARABIDAE), TWO ENDEMIC GROUND BEETLES FROM THE VALDIVIAN FOREST OF CHILE

POSICION FILOGENETICA DE CEROGLOSSUS OCHSENII GERMAIN Y CEROGLOSSUS GUERINI GERMAIN (COLEOPTERA: CARABIDAE), DOS ESCARABAJOS DE SUELO ENDEMICOS DEL BOSQUE VALDIVIANO DE CHILE

Carlos Muñoz-Ramírez 1

ABSTRACT

The phylogenetic position of two Ceroglossus species, C. ochsenii Germain and C. guerini Germain, is evaluated for the first time based on molecular data. The results of this study showed that C. ochsenii and C. guerini are closely related to C. suturalis, with the three species clustering in a strongly supported clade in agreement with previous morphological studies. Although clearly differentiated from C. suturalis, these two species show low genetic divergence between them and do not form reciprocally monophyletic clades, rising questions on their status of separated biological species. Finally, and giving the recent proliferation of subspecific descriptions within Ceroglossus, I discuss the merits of these practices and call for caution on describing new species/subspecies to avoid the inflation of diversity within the genus.

Key words: mtDNA, Phylogeny, South Chile, Systematics.

INTRODUCTION

The genus Ceroglossus Solier 1848 represents a remarkable group of flightless ground beetles endemic to the temperate forest of southern South America (Jiroux 2006). Widely distributed in forests of Central and South Chile, these carabids are highly variable intraspecifically, exhibiting different color morphs depending on the geographic area from where they are sampled. Due to this extraordinary

1 Department of Ecology and Evolutionary Biology, Museum of Zoology, University of Michigan, Ann Arbor, MI 48109-1079, USA. E-mail: carmunoz@umich.edu
color diversity and yet relatively conserved general morphology, their taxonomic study has been particularly challenging. This is illustrated by the existence of more than one hundred names that have been used in the literature since the first description of a species in 1829 (Roig-Juñet & Domínguez 2001). The most recent revisions of the genus (Jiroux 1996, 2006), however, have reduced the number of species to only eight, although many of the previously available names have been maintained for many subspecific designations. Jiroux (1996) proposed a classification for Ceroglossus consisting of four main groups or lineages, based on morphological characters: C. chilensis (Eschsoltz 1829), C. buqueti (Laporte 1834), C. darwini group (i.e. containing C. darwini (Hope 1837), C. magellanicus Gehin 1885 and C. speciosus Gerstaecker 1858), and C. suturalis group (i.e. containing the species C. suturalis (Fabricius 1775), C. ochsenii (Germain 1895), and C. guerini (Germain 1895)). This systematic hypothesis, which relies mainly on characters from the aedeagus and the presence/absence and position of carenas in the antennae of males, was partially supported later by molecular phylogenetic analyses (Okamoto et al. 2001) which, by including six out of the eight known species, recovered the C. darwini group as monophyletic. The group “suturalis”, however, was represented only by the species C. suturalis in the Okamoto et al. (2001) analysis, and the phylogenetic position of the two other species in the group, the rare C. ochsenii and C. guerini, remained untested. The study of C. darwini and C. guerini is important because these species are relatively rare and have small distribution ranges, which make them particularly vulnerable in terms of conservation (Jerez et al. 2015, Pizarro-Araya et al. 2012).

In this study, I use the mitochondrial DNA (mtDNA) marker COI to investigate the phylogenetic position of both C. ochsenii and C. guerini and to generally discuss and comment the findings in relation to previous work. Finally, I provide an opinion on the current tendency of describing new subspecies, and call for caution to avoid an overestimation of the taxonomic diversity within the genus.

MATERIAL AND METHODS

Sample collection and DNA sequencing

A total of 25 individuals from all eight species were analyzed in this study (Table 1). Three to six specimens were used per species except C. ochsenii (2) and C. guerini (1) for which only two and one specimens were available, respectively. For all species except C. ochsenii and C. guerini, individuals were chosen from geographically distant localities to better represent within-species diversity.

Genomic DNA was extracted from legs using the QIAGEN DNeasy Tissue Kit (QIAGEN Inc., Chatsworth CA) following the manufacturer protocol. For the only two dry specimens, one from C. ochsenii and one from C. guerini, the complete head and pronotum were used in the extraction procedure to ensure obtaining enough amplifiable DNA. A portion of the mtDNA COI gene (~680 bp), was amplified for all individuals using the universal primers LCO1490 5′-GGTCAACAAATCTAAAGATATTGG-3′ and HCO2198 5′-TAAACTTCAGGGTGACCAAAAATCA-3′ (Folmer et al. 1994). Each PCR reaction contained 1μl of extracted DNA, 2 μl of 10x buffer, 1.5μl of MgCl2, 1μl of 10mM dNTPs, 0.4 μl of 1% BSA, 0.8μl of each primer (10μM), 0.06μl of Tag DNA polymerase (Invitrogen, USA), and ddH2O to make a total of 25μl reaction. A standard PCR profile with one-minute duration for each step, a total of 35 cycles, and a final extension of 10 minutes at 72°C was followed. The annealing temperature was 52°C. PCR products were sequenced on an ABI Model 3730 XL sequencer by the Sequencing Core, University of Michigan, USA. All sequences were deposited in GenBank, with accession numbers KT997732, KT997737, KT997740, KT997744–45, KT997747, KT997750, KT997752–54, KT997760–
Table 1: Locality data and voucher specimens used in this study.
Tabla 1: Localidades y especímenes usados en este estudio.

<table>
<thead>
<tr>
<th>Locality</th>
<th>Latitude</th>
<th>Longitude</th>
<th>C. chilensis</th>
<th>C. buqueti</th>
<th>C. darwini group</th>
<th>C. suturalis group</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Radal-Siete Tazas</td>
<td>-35.4756</td>
<td>-70.9938</td>
<td>ST.Cc.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Retiro</td>
<td>-36.0908</td>
<td>-71.7834</td>
<td>Ret.Cc.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Pucón</td>
<td>-39.3508</td>
<td>-71.968</td>
<td>Puc.Cc.02</td>
<td>Puc.Cb.02</td>
<td>Puc.Cm.02</td>
<td></td>
</tr>
<tr>
<td>7. Máfí</td>
<td>-39.7028</td>
<td>-72.9198</td>
<td></td>
<td></td>
<td>Maf.Cg.01</td>
<td></td>
</tr>
<tr>
<td>8. Parque Alerce Costero</td>
<td>-40.2063</td>
<td>-73.4129</td>
<td>Ale3.Cc.01</td>
<td>Ale2.Cd.01</td>
<td>Ale3.Cc.01</td>
<td>Ale2.Cd.01</td>
</tr>
<tr>
<td>10. Ancud, Chiloé</td>
<td>-41.882</td>
<td>-73.8799</td>
<td></td>
<td></td>
<td></td>
<td>Chil1.Csp.01-03</td>
</tr>
<tr>
<td>11. Puntra, Chiloé</td>
<td>-42.1195</td>
<td>-73.8066</td>
<td>Chil2.Cc.01</td>
<td></td>
<td></td>
<td>Chil5.Cs.01</td>
</tr>
<tr>
<td>12. Cucao, Chiloé</td>
<td>-42.6481</td>
<td>-74.0653</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Chaitén</td>
<td>-42.9097</td>
<td>-72.7074</td>
<td>Cb.Chai.01</td>
<td>Cb.Chai.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Aysén</td>
<td>-45.1327</td>
<td>-73.0154</td>
<td></td>
<td>Ays.Cb.01</td>
<td></td>
<td>AysRM.Cs.01</td>
</tr>
<tr>
<td>15. Isla Navarino</td>
<td>-54.948</td>
<td>-67.646</td>
<td></td>
<td></td>
<td></td>
<td>Nav.Cs.01</td>
</tr>
</tbody>
</table>

Sequence editing and Phylogenetic analyses

Chromatograms were edited in CodonCode Aligner version 3.0.3 and then imported in Bioedit 7.2.5 (Hall et al. 2011). Sequences were then aligned using the ClustalW algorithm implemented in Bioedit and checked via amino acid coding in MEGA 6.0 (Tamura et al. 2013) to test for unexpected frame shift errors or stop codons. Previous to the phylogenetic analyses, the Xia test (Xia et al. 2003) was conducted to evaluate the degree of sequence saturation, and therefore, the utility of the marker for phylogenetic reconstruction.

Phylogenetic relationships were estimated by maximum likelihood (ML) and Bayesian inference (BI) using the best-fit model selected by JModeltest (Posada 2008) under the Bayesian information criterion (BIC). JModeltest identified HKY+I+G as the best-fit model with the following parameter estimates: Lset Base = (0.3030 0.1709 0.1464) Nst = 2 Tratio = 5.8132 Rates = gamma Shape=1.6760 Ncat=4 Pinvar=0.6880. The ML tree was reconstructed using the software PAUP (Swofford 2002). The tree was estimated via a heuristic search with five random additions of taxa and TBR branch swapping. Node support was assessed by 1000 bootstrap replicates using a heuristic search with 5 random additions of taxa and TBR branch swapping. It is known that distant outgroups (relative to the ingroup) might not be suitable for rooting some phylogenetic trees (Graham et al. 2002). In this...
case, *Ceroglossus* diverged from its closest relatives (e.g. the lineage containing *Carabus* and *Callosoma*) about 50 my ago (Andújar et al. 2012), which certainly makes the rooting by outgroup more challenging. Here, the tree was rooted via the midpoint method (after testing for the molecular clock assumption in PAUP) given that preliminary attempts of rooting via outgroup yielded problematic results (e.g. the root commonly fell among intraspecific clades).

The BI analysis was performed in MrBayes 3.1.2 (Ronquist & Huelsenbeck 2003) and posterior probabilities resulting from this analysis were added as branch support to the ML tree. Two independent runs were run to check for convergence of parameters. Four chains were used for phylogeny estimation, starting the analysis with a random tree and running it for 5,000,000 generations, sampling every 1000 trees and discarding the initial 20% as burning using the ‘sumt’ command. Convergence of parameters from the two independent runs was assessed by the average standard deviation of split frequencies (i.e. values below 0.01) and the potential scale reduction factor (PSRF; values should approach 1). Topological convergence was evaluated by using the plotting tool compare implemented in the online program AWTY (Nylander et al. 2008). Once convergence was confirmed, the two runs were combined to obtain a total of 8002 trees.

**RESULTS**

Xia’s test showed no substitution saturation (Iss=0.3; Iss.c=0.73; p=0.000), which indicates the molecular data is useful for phylogenetic inference. Phylogenetic analyses showed four major clades in agreement with the four-species hypothesis proposed by Jiroux (1996) on the basis of morphological characters (ML best-tree score= 2607.362; Fig. 1, clades A-D): *C. chilensis* (clade A), *C. darwini* group (clade B), *C. buqueti* (clade C), and *C. suturalis* group (clade D). *C. guerini* appeared more closely related with *C. suturalis* than to any other *Ceroglossus* species in agreement with the taxonomic scheme of Jiroux (1996). Support for all main clades was generally high, indicated either by high bootstrap values and/or high Bayesian posterior probabilities. A few other internal clades were also recovered with high support (subclades a1–a3, b1–b2, c1–c2 and d2). In addition, some internal subclades matched more or less some species boundaries. Subclade b1 corresponded to the species *C. magellanicus*,
Figure 2: Maximum likelihood (likelihood score \((-\ln L) = 2613\)) phylogenetic reconstruction of *Ceroglossus* species. Branch support is indicated by the values above the branches (Bayesian posterior probability/ML bootstrap). Only values equal or above 0.5/50 are shown. See Table 1 for details about tip labels.

Figure 2: Árbol filogenético estimado para las especies de *Ceroglossus* mediante máxima verosimilitud (ML). El soporte de rama se indica sobre éstas (probabilidad a posteriori de Bayes/ML bootstrap). Sólo se muestran los valores mayores o iguales a 0.5/50. Ver Tabla 1 para mayores detalles sobre las etiquetas de las ramas terminales.

while clade d1 corresponded to the species *C. suturalis*. Subclade b2, on the other hand, contained both *C. darwini* and *C. speciosus*, with individuals from the latter being grouped within a weakly supported subclade. These species were not represented by reciprocally monophyletic clades, showing paraphyletic relationships. The individual Ale2.Cd.01 from Parque Alerce Costero (identified as *C. darwini ugartei* following Jiroux 2006) was more closely related to the species *C. magellanicus* than to the species *C. darwini*.

Within the clade D, that contains the *C. suturalis* group, the relationship between the three species was not resolved, and even though the species *C. suturalis* was recovered as a monophyletic group, *C. guerini* and *C. ochsenii* were not recovered as reciprocally monophyletic
groups. In other words, one individual from *C. ochsenii* (Ale4.C0.01; Table 1) appeared more closely related to *C. guerini* than to the other individual of *C. ochsenii* (Ale.CO.01).

**DISCUSSION**

The results support the taxonomic scheme of Jiroux (1996) that considered the existence of four major groups within the genus *Ceroglossus*. Please consider the relationships between the four major lineages with caution as the rooting was conducted by the midpoint rooting method because outgroup rooting resulted problematic due to the high evolutionary distance separating outgroup and ingroup (Andújat et al. 2012). However, relationships within each main lineages, which is the main purpose of this study, should not be affected by this issue. Here I focus on relationships within main lineages rather than between main lineages. As proposed by Jiroux (1996, 2006), *C. ochsenii* and *C. guerini* belong to the same group that contains *C. suturalis* to which they are sister species. However, *C. ochsenii* and *C. guerini* do not formed reciprocally monophyletic clades, indicating low genetic divergence between these species.

There are two possible explanations for the low divergence between *C. ochsenii* and *C. guerini*. One possibility is that these species have diverged very recently, so species still maintain some ancestral polymorphism (i.e. it has not been completely sorted by genetic drift). The other possibility is that these entities have not completed the speciation processes, and the lack of reciprocally monophyletic clades reflects ongoing or very recent gene flow. Morphological differences between *C. ochsenii* and *C. guerini* are well established in the literature (Jiroux 1996, 2006) and seem to support the former hypothesis. Following this argument, the lack of reciprocally monophyletic clades would be caused by the presence of incomplete lineage sorting (ILS), which can be common in recently diverged taxa that exhibit relatively large population sizes. However, these species are rare and have small geographic distributions, suggesting they have relatively small population sizes. Under small population sizes, the presence of ILS is less likely, as the higher magnitude of genetic drift in small population would accelerate the formation of reciprocally monophyletic clades. It is possible then, that the lack of monophyly observed in these species may reflect ongoing gene flow instead of ILS, and the observed morphological differences respond to phenotypic variation within species rather than differentiation between species. If that is the case, the validity of these entities as two separate biological species should be re-examined in the light of new evidence.

Low divergence was also found between *Ceroglossus darwini* and *Ceroglossus speciosus*. Previous phylogenetic work (Okamoto et al. 2001) included only one specimen of *C. speciosus*, so patterns of reciprocal monophyly (or the lack of it) could not be assessed. This species also have small geographic range (i.e. Ancud and surrounding areas) so questions remain whether the low divergence reflects recent speciation, recent gene flow, or both. A pattern that was consistent with previous molecular work is the placement of *C. darwini ugariei* into the clade of *C. magellanicus*. The inclusion of this subspecies within *C. darwini* was acknowledged as problematic by Jiroux (2006) due to the presence of characters from both species. The phylogenetic results from this study and from Okamoto et al. (2001) clearly show this subspecies belong to *C. magellanicus*, so amendments should be made to update the *Ceroglossus* taxonomy.

**Comments on the phenotypic diversity within Ceroglossus**

Although there is a high phenotypic diversity within some *Ceroglossus* species, this work does not provide a basis for distinguishing diversity at a finer taxonomic scale (i.e. subspecies or color morphs). Indeed, the lack of reciprocal monophyly found between some species suggests that making claims about the
validity of some species designations can be challenging even for species with clearly distinguishable phenotypes. Then, making the case for different taxonomic entities below the level of species should be done carefully, after examining different types of evidence (e.g. DNA and Morphology) (Dayrat 2005, Padial et al. 2010). Unfortunately, there has been a proliferation of work describing new subspecies in the recent literature, which do not rely on multiple sources of evidence and lack the use of clear biological criteria. This practice may distort the amount of diversity that actually exists within these taxa, and create problems related with the stability of its taxonomy. I am not against the practice of describing subspecies, as long as it relies on biologically informed criteria and it is supported by multiple types of evidence (e.g. ecological, behavioral, morphological, geographical, or molecular). Describing new subspecies without the appropriate support can artificially inflate the actual diversity within this group and have a negative impact on studies that build on information about taxonomic diversity (e.g. ecological, biogeographical, and conservation studies). Further studies are needed to understand the mechanisms promoting phenotypic diversity in Ceroglossus, and future work should focus on testing the actual diversity within the genus. Until then, researchers should refrain from describing new species/subspecies without making reference to clearly defined criteria and analyzing multiple, independent types of evidence.

ACKNOWLEDGMENTS

I want to thank the Corporación Nacional Forestal (CONAF) for allowing sampling within several National Parks (Authorization number: 024/2012; see Table 1). I also appreciate the material donated by Álvaro Zúñiga, Dusan Boric, Gustavo Valenzuela, Mauricio Cid, Christian Muñoz, Laurent Rataj, and Dr. Viviane Jerez. I am especially grateful to Álvaro Zuñiga, Christian Muñoz, Alfonso Jara, Luis Correa, Jorge Avaria, Pablo Fuentes, and many others for their selfless assistance in the field and appreciate the comments and feedback provided by two anonymous reviewers that greatly improved the manuscript.

REFERENCES


(Recibido: 10 agosto 2015; Aceptado 3 diciembre 2015; Publicado: 30 diciembre 2015)
Manejado por Viviane Jerez.