Molecular phylogenetic analysis of archival tissue reveals the origin of a disjunct southern African–Palaearctic weevil radiation

Gerardo Hernández-Vera¹, Roberto Caldara², Ivo Toševski³,⁴ and Brent C. Emerson¹,⁵*

¹School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, UK, ²via Lorenteggio 37, 20146, Milan, Italy, ³CABI Europe – Switzerland, 2800, Delémont, Switzerland, ⁴Department of Plant Pests, Institute for Plant Protection and Environment, Zemun, Serbia, ⁵Island Ecology and Evolution Research Group, IPNA-CSIC, 38206 La Laguna, Tenerife, Canary Islands, Spain

ABSTRACT

Aim We test three alternative hypotheses for the disjunct Mediterranean–southern African distribution of endophagous weevils within the genera Rhinusa and Gymnetron (Coleoptera: Curculionidae): (1) a Palaearctic origin with dispersal to southern Africa; (2) a southern African origin with dispersal to the Palaearctic; and (3) a widespread ancestral distribution fragmented by vicariance. Divergence times are estimated to provide an approximate temporal framework for the evolution of the group and to evaluate potential palaeogeographical scenarios.

Location Southern Africa, the Mediterranean region, the Palaearctic and eastern Africa.

Methods Freshly collected and dry, pinned samples of weevils were used as a source of DNA. Prior genetic information was used to identify short phylogenetically informative amplicons within the 16S ribosomal RNA gene (16S). Phylogenetic reconstructions using Bayesian and maximum likelihood analyses of mitochondrial and nuclear DNA sequence data and molecular dating techniques were used to infer the biogeographical history of Rhinusa and Gymnetron species. A statistical approach to dispersal–vicariance analysis (s-DIVA) was used to further assess biogeographical hypotheses.

Results Successful polymerase chain reaction amplification of targeted short 16S DNA sequences (150 bp) from dry, pinned specimens provided for increased species sampling of Rhinusa and Gymnetron by 230%, greatly expanding species representation from southern Africa. Phylogenetic reconstructions and s-DIVA analyses support a southern African origin for Rhinusa and Gymnetron species. Divergence time estimates suggest southern African and Palaearctic lineages diverged c. 11.6–7.4 Ma.

Main conclusions Rhinusa and Gymnetron represent a complex of lineages with a shared evolutionary history of range expansions from southern Africa into the Palaearctic. Our results support a late Miocene vicariance scenario, most likely as a result of repeated desertification. The use of prior genetic information to identify short phylogenetically informative amplicons offers a useful approach for molecular phylogenetic analyses incorporating archival material.

Keywords Archival tissue, Curculionidae, disjunct distribution, dispersal, Palaearctic, phylogeny, southern Africa, vicariance, weevil.
INTRODUCTION

Examples of Mediterranean–southern African disjunct distributions are perhaps mostly known in plants, given the floristic affinities shared by these two regions, which comprise two of the five mediterranean-type floras of the world (Cowling et al., 1996). Previous studies have proposed long-distance dispersal as the major cause of this disjunction (Thorne, 1972; Raven, 1973) and, with some exceptions (e.g. McGuire & Kron, 2005), the most widely accepted hypothesis is a southern African origin with dispersal to the north through an East African corridor (Caujapé-Castells et al., 2001; Coleman et al., 2003; Calviño et al., 2006; del Hoyo et al., 2009). Interestingly, there are very few studies on plant-feeding insects that exhibit similar disjunct distributions, perhaps tracking the ancestral distribution of their host plants (but see Mey, 2006, and Kirk-Spriggs & McGregor, 2009, for examples in Lepidoptera and Diptera, respectively). Within Coleoptera, proposed explanations for south African–Mediterranean disjunct distributions favour a vicariance scenario (Biondi & D’Alessandro, 2008; Bologna et al., 2008), acknowledging the possible ecological connections between the southern African and Mediterranean regions in the past via ‘arid corridors’, as originally described by Balinsky (1962), which are thought to have appeared in eastern Africa prior to the end of the Miocene (Verdcourt, 1969; Goldblatt, 1978; Jürgens, 1997).

To further investigate disjunct Mediterranean–southern African insect distributions, we focus attention on weevil species of the closely related genera Rhinusa Stephens, 1829 and Gymnetron Schoenherr, 1825, within the tribe Mecinini (Curculionidae: Curculioninae). Adults of this species-rich group range in size from 2 to 5 mm, with representatives in both the Palaearctic and Afrotropical regions. Species of both genera are endophagous parasites whose larvae feed and develop within tissues of plants in the families Scrophulariaceae and Plantaginaceae as host plants (see below). Thirty-two species are thought to have appeared in eastern Africa prior to the end of the Miocene (Verdcourt, 1969; Goldblatt, 1978; Jürgens, 1997).

Thus, using both archival specimens and freshly collected samples we test three alternative hypotheses for the disjunct Mediterranean–southern African distribution of Gymnetron species: (1) a Palaearctic origin with dispersal to southern Africa; (2) a southern African origin with dispersal to the Palaearctic; and (3) a widespread ancestral distribution fragmented by vicariance. Divergence times are estimated to provide an approximate temporal framework for the evolution of the group and evaluate potential palaeogeographical scenarios.

MATERIALS AND METHODS

Taxon sampling

Our sampling strategy was to obtain broad geographical and taxonomic coverage of representatives within Gymnetron and Rhinusa, using freshly collected samples where possible, and augmenting our sampling for material that was harder to obtain using archival specimens (see Appendix S1 in Supporting Information). Sixty-seven specimens were collected in the field, placed in 96–100% ethanol, and stored at 4 °C until DNA extraction. Forty-two dry, pinned specimens were used from the personal entomological collection of R. Caldara for DNA extraction from a single leg (see below). Thirty-two

INTRODUCTION

Examples of Mediterranean–southern African disjunct distributions are perhaps mostly known in plants, given the floristic affinities shared by these two regions, which comprise two of the five mediterranean-type floras of the world (Cowling et al., 1996). Previous studies have proposed long-distance dispersal as the major cause of this disjunction (Thorne, 1972; Raven, 1973) and, with some exceptions (e.g. McGuire & Kron, 2005), the most widely accepted hypothesis is a southern African origin with dispersal to the north through an East African corridor (Caujapé-Castells et al., 2001; Coleman et al., 2003; Calviño et al., 2006; del Hoyo et al., 2009). Interestingly, there are very few studies on plant-feeding insects that exhibit similar disjunct distributions, perhaps tracking the ancestral distribution of their host plants (but see Mey, 2006, and Kirk-Spriggs & McGregor, 2009, for examples in Lepidoptera and Diptera, respectively). Within Coleoptera, proposed explanations for south African–Mediterranean disjunct distributions favour a vicariance scenario (Biondi & D’Alessandro, 2008; Bologna et al., 2008), acknowledging the possible ecological connections between the southern African and Mediterranean regions in the past via ‘arid corridors’, as originally described by Balinsky (1962), which are thought to have appeared in eastern Africa prior to the end of the Miocene (Verdcourt, 1969; Goldblatt, 1978; Jürgens, 1997).

To further investigate disjunct Mediterranean–southern African insect distributions, we focus attention on weevil species of the closely related genera Rhinusa Stephens, 1829 and Gymnetron Schoenherr, 1825, within the tribe Mecinini (Curculionidae: Curculioninae). Adults of this species-rich group range in size from 2 to 5 mm, with representatives in both the Palaearctic and Afrotropical regions. Species of both genera are endophagous parasites whose larvae feed and develop within tissues of plants in the families Scrophulariaceae and Plantaginaceae as host plants (see below). Thirty-two species are thought to have appeared in eastern Africa prior to the end of the Miocene (Verdcourt, 1969; Goldblatt, 1978; Jürgens, 1997).

Thus, using both archival specimens and freshly collected samples we test three alternative hypotheses for the disjunct Mediterranean–southern African distribution of Gymnetron species: (1) a Palaearctic origin with dispersal to southern Africa; (2) a southern African origin with dispersal to the Palaearctic; and (3) a widespread ancestral distribution fragmented by vicariance. Divergence times are estimated to provide an approximate temporal framework for the evolution of the group and evaluate potential palaeogeographical scenarios.

MATERIALS AND METHODS

Taxon sampling

Our sampling strategy was to obtain broad geographical and taxonomic coverage of representatives within Gymnetron and Rhinusa, using freshly collected samples where possible, and augmenting our sampling for material that was harder to obtain using archival specimens (see Appendix S1 in Supporting Information). Sixty-seven specimens were collected in the field, placed in 96–100% ethanol, and stored at 4 °C until DNA extraction. Forty-two dry, pinned specimens were used from the personal entomological collection of R. Caldara for DNA extraction from a single leg (see below). Thirty-two

INTRODUCTION

Examples of Mediterranean–southern African disjunct distributions are perhaps mostly known in plants, given the floristic affinities shared by these two regions, which comprise two of the five mediterranean-type floras of the world (Cowling et al., 1996). Previous studies have proposed long-distance dispersal as the major cause of this disjunction (Thorne, 1972; Raven, 1973) and, with some exceptions (e.g. McGuire & Kron, 2005), the most widely accepted hypothesis is a southern African origin with dispersal to the north through an East African corridor (Caujapé-Castells et al., 2001; Coleman et al., 2003; Calviño et al., 2006; del Hoyo et al., 2009). Interestingly, there are very few studies on plant-feeding insects that exhibit similar disjunct distributions, perhaps tracking the ancestral distribution of their host plants (but see Mey, 2006, and Kirk-Spriggs & McGregor, 2009, for examples in Lepidoptera and Diptera, respectively). Within Coleoptera, proposed explanations for south African–Mediterranean disjunct distributions favour a vicariance scenario (Biondi & D’Alessandro, 2008; Bologna et al., 2008), acknowledging the possible ecological connections between the southern African and Mediterranean regions in the past via ‘arid corridors’, as originally described by Balinsky (1962), which are thought to have appeared in eastern Africa prior to the end of the Miocene (Verdcourt, 1969; Goldblatt, 1978; Jürgens, 1997).

To further investigate disjunct Mediterranean–southern African insect distributions, we focus attention on weevil species of the closely related genera Rhinusa Stephens, 1829 and Gymnetron Schoenherr, 1825, within the tribe Mecinini (Curculionidae: Curculioninae). Adults of this species-rich group range in size from 2 to 5 mm, with representatives in both the Palaearctic and Afrotropical regions. Species of both genera are endophagous parasites whose larvae feed and develop within tissues of plants in the families Scrophulariaceae and Plantaginaceae as host plants (see below). Thirty-two species are thought to have appeared in eastern Africa prior to the end of the Miocene (Verdcourt, 1969; Goldblatt, 1978; Jürgens, 1997).

Thus, using both archival specimens and freshly collected samples we test three alternative hypotheses for the disjunct Mediterranean–southern African distribution of Gymnetron species: (1) a Palaearctic origin with dispersal to southern Africa; (2) a southern African origin with dispersal to the Palaearctic; and (3) a widespread ancestral distribution fragmented by vicariance. Divergence times are estimated to provide an approximate temporal framework for the evolution of the group and evaluate potential palaeogeographical scenarios.

MATERIALS AND METHODS

Taxon sampling

Our sampling strategy was to obtain broad geographical and taxonomic coverage of representatives within Gymnetron and Rhinusa, using freshly collected samples where possible, and augmenting our sampling for material that was harder to obtain using archival specimens (see Appendix S1 in Supporting Information). Sixty-seven specimens were collected in the field, placed in 96–100% ethanol, and stored at 4 °C until DNA extraction. Forty-two dry, pinned specimens were used from the personal entomological collection of R. Caldara for DNA extraction from a single leg (see below). Thirty-two
species of *Rhinusa* were sampled (18 field-collected and 14 archival), representing c. 80% of recognized species, including representatives from the three main taxonomic groups proposed by Caldara *et al.* (2010). Twenty-five species of *Gymnetron* were sampled (6 field-collected and 20 archival; one species – *G. villosulum* – was sampled both fresh and archival), representing 8 of the 13 Afrotropical groups proposed by Caldara (2003) and 11 species from the Palaearctic region. One specimen of the genus *Cleopomniarius* was also included, which is the only genus within the tribe Mecinini apart from *Gymnetron* to have representatives in both the Palaearctic and Afrotropical regions (Caldara *et al.*, 2008).

**DNA extraction**

Field-sampled weevils were punctured through the abdomen and total genomic DNA was extracted from each individual using the DNeasy extraction kit (Qiagen, Hilden, Germany), following the manufacturer’s protocol. After DNA extraction, insects were placed again in 96% ethanol and kept at 4 °C, before subsequent preparation as voucher specimens within the collection of R.C. When using archival specimens, destruction of the sample for DNA extraction has always been a concern, especially for small specimens (Thomas, 1994; Whitfield & Cameron, 1994), and for arthropods, several methods have been proposed to minimize damage (Gilbert *et al.*, 2007; Rowley *et al.*, 2007). We use a non-destructive approach using a single leg for DNA extraction as suggested by Gilbert *et al.* (2007). Dry, pinned insects were first left overnight in a humid chamber in order to reduce exoskeleton brittleness and allow manipulation of the appendages. A single posterior leg was then removed under a dissecting microscope and placed in 200 μL of DNeasy extraction buffer for overnight incubation at 56 °C, followed by DNA isolation as for non-archival samples. DNA-extracted appendages were recovered and remounted with their corresponding specimen on a new entomological card (Appendix S3; Fig. S1).

**PCR amplification and sequencing reactions**

Two mitochondrial (cytochrome *c* oxidase subunit II (*COII*) and 16S ribosomal RNA (*16S*)) and three nuclear gene fragments (elongation factor-1α (*EF-1x*), arginine kinase (*AK*) and 18S ribosomal RNA (*18S*)) were used. The primers used for each gene are described in Appendix S3. For all loci, polymerase chain reactions (PCR) were performed with BioTaq DNA polymerase (Bioline, London, UK) with 1× NH₄ buffer, 3.5–5.0 mM MgCl₂, 0.2 mM each dNTP, 0.2–0.4 μM each primer, 0.5 U of Taq polymerase and 1–5 μL of DNA template in a 25-μL final volume. PCR cycles were carried out using the following thermal profile for the *COII* and *AK* gene fragments: 95 °C for 3 min, 33 and 37 cycles respectively at 95 °C for 1 min, 48 °C for 1 min, 72 °C for 1 min, and a final extension at 72 °C for 3 min. For *EF-1x*, 16S and 18S genes, two different touchdown profiles (Don *et al.*, 1991) were used. For *EF-1x* and 16S, the programme comprised 94 °C for 1 min 30 s, 10 cycles at 94 °C for 45 s, 58 °C for 1 min, 72 °C for 1 min, decreasing the annealing temperature by 1 °C every cycle, then 35 cycles at 94 °C for 45 s, 48 °C for 1 min, 72 °C for 1 min, and 72 °C for 3 min as a final extension. For the 18S fragment, the programme comprised 95 °C for 3 min, 8 cycles at 94 °C, 54 °C and 72 °C for 1 min at each temperature, decreasing the annealing temperature by 1 °C every cycle, then 28 cycles at 94 °C, 46 °C and 72 °C for 1 min at each temperature and a final extension at 72 °C for 2 min. Sequences were generated with a PerkinElmer ABI3700 automated sequencer, using the BigDye terminator reaction protocol (v3.1 PerkinElmer, Warrington, UK) in a 10-μL final volume. For the *COII*, 16S, *AK* and 18S gene fragments, sequences were obtained with the forward primer only, except in a few cases which required sequencing in both directions because of low-quality chromatograms. All of the *EF-1x* sequences were obtained with forward and reverse primers because of the presence of an intron region with indels.

**SPIAs primer design, PCR amplification and sequencing**

Based on the alignment of gene partitions from non-archival samples, we identified two adjacent variable regions of 95 bp and 55 bp within the mitochondrial 16S gene, flanked by comparatively conserved motifs. We designed primers spanning each of these amplicons, also including representative sequences from GenBank within the family Curculionidae to maximize the broad taxonomic utility of these primers, including the following genera: *Anthonomus*, *Brachonyx*, *Curtuio*, *Eutoxus* and *Tychius*. We also included *Tribolium* from the family Tenebrionidae. For one of the 16S SPIAs, the forward PCR primer included an M13 adaptor oligonucleotide (Appendix S3), and the adaptor was used as a sequencing primer to improve sequence read quality immediately after the 3’ end of the primer. The two 16S amplicons were PCR amplified with the following conditions: 95 °C for 5 min; 40 cycles at 95 °C, 45 °C and 72 °C for 20 s; and a final extension at 72 °C for 5 min. To control for contamination, full precautions were taken; these included the use of a PCR hood workstation with a 15-minute UV-light sterilization step, dedicated PCR pipettes, filtered tips, and the inclusion of multiple negative controls. All short 16S amplicons were sequenced in both directions, and in some cases resequenced from an independent PCR reaction to corroborate accuracy.

**Phylogenetic analyses**

All sequences were automatically aligned using the ClustalW algorithm, as implemented in BioEdit 7.0.9 (Hall, 1999), with further manual alignment. For *EF-1x*, forward and reverse sequences were first assembled as contigs, and the intron region removed due to the inability to align it unambiguously. Sequence properties for each individual
gene partition were assessed using mega 4.0.1 (Tamura et al., 2007). Phylogenetic analyses were performed for individual partitions, and for a concatenated alignment of nuclear and mitochondrial sequences. For the latter, analyses were performed using two data sets. The first data set consisted of ingroup taxa with complete sampling of the five gene partitions. The second data set consisted of the first data set with the addition of archival specimens sampled for the SPIAs. Bayesian analyses were performed with the parallel version of MrBayes 3.1 (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003) using the substitution models selected by jModelTest (Posada, 2008) for each partition, with priors set to the default values as recommended by Ronquist et al. (2005). Settings for each gene partition in the individual analyses were as follows: two simultaneous runs (each with two chains) of the Markov chain Monte Carlo (MCMC) for five million generations, with a sampling frequency of 100 generations, a heating parameter value of 0.02–0.05 (decreased from the default value 0.2 to improve swapping of states between the heated and cold chains) and a relative burn-in of 25%. Summaries from the stationary distribution of the sampled parameter values and sampled trees were obtained, as well as a majority-rule consensus tree with posterior probabilities for each bipartition. For concatenated sequence analyses (five gene partitions) with and without archival specimens, we also used two independent runs but with four Markov chains each, to optimize convergence for large data sets, as suggested by Ronquist et al. (2005), and 20 million generations, sampling every 1000 generations. To determine stationarity of the MCMC chains, variation in log-likelihood scores was examined graphically using the program Tracer 1.5 (Rambaut & Drummond, 2009); the final standard deviation of split frequencies was ≤ 0.01 in all analyses. Maximum likelihood (ML) analyses were performed with the parallel version of RAXML 7.0.4 (Stamatakis, 2006) for the data set of five gene partitions with and without archival specimens sampled for the SPIAs. A partitioned model was used where each genetic marker was assigned a separate GTR+I+G model. One thousand heuristic searches were executed using the default settings to find the ML tree. Branch support was estimated from 1000 replicates using the standard bootstrap procedure as implemented in RAXML (Stamatakis et al., 2008). Both Bayesian and ML analyses were conducted on the High Performance Computing Cluster at the University of East Anglia.

Assessing the phylogenetic placement of archival material sampled for SPIAs

Phylogenetic placement of the archival taxa sampled for SPIAs was assessed with two approaches. First, we performed a Bayesian analysis of non-archival specimens with the concatenated alignment of five gene partitions. Nucleotide data for selected taxa representing a range of divergences from sister lineages were reduced or ‘trimmed’ to the SPIA partitions. We then compared the phylogenetic placement of these character-reduced taxa within the phylogeny with their placement in analyses using the full set of characters. As a second approach, Bayes factors were used to compare the likelihoods of models with enforced monophyly at key nodes of interest including archival specimens sampled for the SPIAs, against models without such constraints.

Estimation of divergence times

In the absence of geological or fossil calibration points to estimate divergence times, we took a Bayesian approach as implemented in BEAST 1.4.8 (Drummond & Rambaut, 2007) using a relaxed molecular clock to estimate the age of the most recent common ancestor (MRCA) of nodes of interest representing divergences between southern African and Palaeartic lineages. We applied a mean pancrustacean COII rate estimate of 3.05% pairwise divergence per million years, based on previous work showing the mean substitution rate of Coleoptera to approximate the mean rate across the Pancrustacea (Cicconardi et al., 2010). The estimated age of the root of the COII tree was then used as prior information for rate calibration of the SPIAs. The validity of applying the COII mean substitution rate to the 16S gene was assessed by comparison of the uncorrected pairwise distances of both genes which are closely correlated (r = 0.74) (see Appendix S2: Fig. S2).

Reconstruction of ancestral ranges

To further evaluate the biogeographical patterns observed in the phylogeny and infer ancestral ranges, we used statistical dispersal–vicariance analysis as implemented in the software s-diva (Yu et al., 2010). This program incorporates the methods of Nylander et al. (2008) and Harris & Xiang (2009) to account for phylogenetic uncertainty – and uncertainty in dispersal–vicariance analysis (DIVA) optimization – using a nonparametric empirical Bayesian approach. Statistical support for ancestral range reconstructions is determined by averaging the frequencies of an ancestral range at a node over a set of trees (Yu et al., 2010). We used the 20,000 trees sampled from the Bayesian analysis of the data set of five gene partitions including archival specimens with 5000 trees discarded as burn-in. The number of maximum areas was set to two, reflecting the ingroup’s number of defined areas of distribution – (A) southern Africa and (B) Palaeartic.

RESULTS

The five gene fragments were amplified and sequenced for 56 of the 67 field-collected samples (see Appendix S1 for GenBank accession numbers). The 10 samples of G. rotundicolle and the sample of G. rostellum consistently failed to amplify for any gene fragment, and were subsequently found to have been stored in denatured ethanol. They were consequently amplified using the same protocol applied to archival
samples. Across the 56 individuals sequenced for the five gene partitions, only five samples present some missing sequence data due to poor read quality in some regions of the chromatograms. Sequence variation within each gene partition is detailed in Table 1. The combined data of five gene partitions contained 3943 nucleotides, of which 885 sites were variable and 19% were parsimony-informative. The full DNA sequence data matrix including SPIA sequences is available in TreeBase (http://treebase.org/, accession number S13611).

**SPIA primer design, PCR amplification and sequencing**

Successful PCR amplification was achieved for both SPIAs for all 42 archival samples and the 11 samples in denatured ethanol. Five archival samples yielded comparatively weak PCR products, and did not generate readable sequence chromatograms. The remaining 37 samples yielded readable sequence chromatograms for both SPIAs (Appendix S1), increasing our species sampling with the addition of 11 and 17 taxa from *Rhinusa* and *Gymnetron*, respectively, and greatly expanding species representation from southern Africa.

**Phylogenetic analyses and assignment of archival specimens**

Analyses of individual gene partitions essentially recovered the same tree topology, but with different degrees of phylogenetic resolution. Bayesian and ML analyses of the data set of five gene partitions, without archival samples, recovered the same robustly supported phylogeny (Fig. 1), revealing neither *Rhinusa* nor *Gymnetron* to be monophyletic, but with phylogenetic relationships among *Rhinusa* species broadly in agreement with the taxonomy of *Rhinusa* based on morphological characters. The same phylogenetic relationships among non-archival samples were derived from Bayesian and ML analyses when the 16S SPIAs were included within the alignment of the five gene partitions (Fig. 2), with the single exception of *G. melanarium* being placed basally within clade G in the Bayesian analysis. Bayesian and ML analyses with the inclusion of the 16S SPIAs also resulted in the identical phylogenetic placements of the archival specimens (Fig. 2), with the single exception of clade I being placed as the sister group to clade L in the ML analysis.

Despite the limited amount of nucleotide data (150 bp), relatively high posterior probability (PP) and/or bootstrap support (BS) values were observed for the placement of some archival specimens. Two southern African taxa, *Gymnetron perriniae* and *G. bisignatum*, are united as sister taxa with PP = 1 and BS = 93%. *Rhinusa exigua* is placed as the sister lineage to the clade of *R. antirrhini*/R. florum* with PP = 0.85 and BS = 78%. *Rhinusa brisouti* is placed as the sister lineage to *R. linariae* with PP = 0.90 and BS = 75%. *Rhinusa moroderi* is placed as the sister lineage to *R. tetra* with PP = 0.98 and BS = 85%. *Rhinusa comosa* is placed as the sister lineage to *R. verbasci* with PP = 1 and BS = 98%. *Gymnetron minimum* and *G. pauciillum* are united as sister species with PP = 1 and BS = 100%. Twenty-four archival specimens are placed within clades A, B, E, F and G (Fig. 2), which were also inferred by the analysis of the five-gene data set without archival specimens (Fig. 1). The remaining 13 samples, plus the 10 samples of *G. rotundicole* and the sample of *G. rostellum*, were placed outside the lineages and clades defined in Fig. 1.

To examine how placements might be influenced by limited nucleotide data, a Bayesian analysis of the concatenated alignment of the five genes without archival samples was performed with the sequences of *R. linariae*, *R. vestita* and *G. piceum* trimmed to represent the 16S SPIAs. In the case of *R. linariae* and *R. vestita*, representing moderate and intermediate divergences, respectively, from their sister lineages, they were both assigned to their correct clades, but with some slight topological changes. A single branch arrangement places *R. linariae* as a sister lineage to *R. griscohirta* within clade A (see Appendix S2: Fig. S3). Two branch arrangements result in a more derived phylogenetic position for *R. vestita* within clade B. In the more extreme case of divergence from a sister lineage, the phylogenetic placement of *G. piceum* was substantially altered (Fig. S3). These results indicate that while approximate phylogenetic placement can be achieved with limited SPIA data, exact placements might not be reliably inferred. Thus, our phylogenetic results clearly support deep genetic divergences across southern African taxa, although the specific relationships of southern African lineages to Palaearctic lineages and

<table>
<thead>
<tr>
<th></th>
<th>COII</th>
<th>16S</th>
<th>16S SPIAs</th>
<th>AK</th>
<th>EF-1α</th>
<th>18S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aligned base pairs</td>
<td>697*</td>
<td>307</td>
<td>150</td>
<td>715†</td>
<td>647‡</td>
<td>1577</td>
</tr>
<tr>
<td>Variable sites</td>
<td>397</td>
<td>130</td>
<td>78</td>
<td>178</td>
<td>143</td>
<td>37</td>
</tr>
<tr>
<td>Parsimony informative</td>
<td>372 (53.4%)</td>
<td>104 (33.9%)</td>
<td>67 (44.6%)</td>
<td>132 (18.5%)</td>
<td>118 (18.2%)</td>
<td>30 (2%)</td>
</tr>
</tbody>
</table>

*682 bp for Gy1018.
†423 bp for Gy976; 382 bp for Gy977.
‡418 bp for Gy975 and Gy976; 628 bp for Gy967.
Figure 1 Bayesian phylogenetic tree of southern African and Palaearctic lineages of endophagous weevils within the genera Rhinusa and Gymnetron inferred from DNA sequences from five genes (COII, 16S, EF-1α, AK and 18S) comprising 3943 bp. Bayesian posterior probabilities and maximum likelihood bootstrap values are shown above and below branches, respectively. Seven higher-order lineages and well-supported clades are labelled A–G. Species of Gymnetron are highlighted in bold.

Figure 2 Bayesian phylogenetic tree of southern African and Palaearctic lineages of endophagous weevils within the genera Rhinusa and Gymnetron inferred from (1) DNA sequences from five genes (COII, 16S, EF-1α, AK and 18S) comprising 3943 bp for 54 ingroup samples and two outgroup taxa, and (2) 150 bp of nucleotide data corresponding to the 16S short phylogenetically informative amplicons (SPIAs) sampled from 37 archival specimens and 11 degraded specimens (highlighted in grey). Species of Gymnetron are highlighted in bold. Green branches indicate species from southern Africa. Bayesian posterior probabilities ≥ 0.8 and maximum likelihood bootstrap values ≥ 70% are shown above and below branches, respectively. Major groups are labelled A–L. Roman numerals indicate nodes for which divergence times are estimated between southern African and Palaearctic lineages (see Table 2). Pie charts indicate relative probabilities of alternative ancestral ranges mapped onto nodes of interest, (A) southern Africa, (B) Palaearctic. Branches for archival samples associated with no mutational events have been assigned zero length.
Biogeography of a southern African–Palaearctic weevil radiation
the basal divergences of several southern African lineages require further assessment. To achieve this, we carried out Bayes factor tests to compare the harmonic means of likelihood values of (1) models with enforced monophyly for each of nodes I–IV (Fig. 2) that represent the MRCA of southern African and Palaearctic lineages, and (2) a model enforcing the monophyly of clades A–L, excluding all remaining southern African species, against models without the constraint of monophyly. Applying the guidelines of Kass & Raftery (1995), we obtained very strong support for the monophyly of nodes II–IV, and the clade comprising groups A–L (in Bayes factor > 10), while only moderate support was obtained for the monophyly of node I (2 < ln Bayes factor < 10).

Divergence time estimates

BEAST analyses calibrated with the evolutionary rate for the mtDNA COI gene generated a mean estimate of 26.7 million years ago (Ma) for the MRCA of Gymnetron and Rhinusa, with a 95% highest posterior density (HPD) interval of 21.2–32.5 Ma. This age estimate was used as prior information to calibrate the corresponding node of the tree containing both archival and non-archival samples (Fig. 2), to estimate the approximate divergence times of southern African Gymnetron lineages from Palaearctic Gymnetron and Rhinusa lineages with 16S SPIA sequence data. We imposed a normal distribution for the root age with a mean of 26.7 Ma and a standard deviation of 7 Myr, allowing a 95% probability distribution for sampling of 15–35 Ma, an interval that includes the 95% HPD estimated with mtDNA COI sequence data. Prior information for the 16S SPIA substitution rate was also incorporated. We estimated a range of 16S SPIA rates by fixing the MRCA of Gymnetron and Rhinusa to be (1) 21.2 Ma, representing the lower 2.5% posterior probability value estimated from mtDNA COI data, and (2) 32.5 Ma, representing the upper 2.5% posterior probability value. Based on the resulting mean rate estimates from these calibrations of 0.0047 and 0.018 substitutions per site per million years, we applied a normal distribution for the SPIA mutation rate with a mean of 0.01 and a standard deviation of 0.005, thus incorporating our estimated range within the 95% HPD interval. Normal distributions were chosen as prior parametric distributions because they allow bidirectional uncertainty in the calibration estimates (Ho & Phillips, 2009). Ages for four nodes corresponding to divergences between southern African and Palaearctic lineages were estimated (Fig. 2, Table 2), yielding broad HPDs, reflecting the broad prior root age and substitution-rate distributions that we allowed the analysis to sample from. However, the mean age estimates for all four divergence events range between 11.6 and 7.4 Ma, suggesting a late Miocene divergence between Palaearctic and southern African lineages.

Reconstruction of ancestral ranges

The relative probabilities of alternative ancestral ranges mapped onto the two most basal nodes of the phylogenetic tree that included the archival specimens suggest a southern African origin for Rhinusa and Gymnetron, with values higher than 60% (Fig. 2). Subsequent vicariant events across ancestral ranges that extended to include both southern Africa and the Palaearctic were inferred for nodes representing the MRCA of southern African and Palaearctic lineages (nodes I–IV; Fig. 2).

DISCUSSION

Phylogenetic analyses and assignment of archival specimens

Analyses of Rhinusa and Gymnetron species sampled for all five gene partitions resulted in a robust phylogeny, recovering four monophyletic groups largely in agreement with a proposed taxonomy based on adult morphological characters (Caldara et al., 2010). The single major difference from conventional taxonomic classification is that Gymnetron and Rhinusa are not reciprocally monophyletic. When the 16S SPIAs from archival specimens were included in the data matrix, Bayesian and ML analyses recovered essentially the same tree topology, with most archival samples falling within clades defined by the complete matrix of five genes, largely in agreement with prior taxonomic expectations (Caldara, 2003, 2008; Caldara et al., 2010). High support values for the placement of several archival specimens, obtained in both Bayesian and ML analyses, suggest that phylogenetic placement can in some cases be achieved with a high degree of confidence using SPIAs. This is consistent with previous findings that show there is no relationship between character completeness of a taxon, and the level of support for its placement within a phylogeny (Wiens et al., 2005). However, for more divergent archival lineages within a particular clade, phylogenetic placement may be less reliably inferred, as revealed by our Bayesian analyses with sequences for selected taxa reduced to the 150 nucleotides of the 16S SPIAs. Thus, for nodes of relevance to our biogeographical hypotheses that lack high support from both Bayesian posterior probabilities and ML bootstrap values, we have applied Bayes factor analyses to test hypotheses of monophyly, and applied statistical

Table 2 Estimated times to the most recent common ancestor (TMRCA) for nodes indicating divergence of southern African and Palaearctic lineages of endophagous weevils within the genera Rhinusa and Gymnetron.

<table>
<thead>
<tr>
<th>Node number</th>
<th>Mean (Ma)</th>
<th>95% HPD intervals (Ma)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>7.4</td>
<td>1.1–15.9</td>
</tr>
<tr>
<td>II</td>
<td>11.6</td>
<td>4.2–20.1</td>
</tr>
<tr>
<td>III</td>
<td>11.0</td>
<td>4.1–19.1</td>
</tr>
<tr>
<td>IV</td>
<td>8.5</td>
<td>2.7–16.0</td>
</tr>
</tbody>
</table>

HPD, highest posterior density.
dispersal–vicariance analyses, allowing us to take phylo-
genetic uncertainty into account.

Although several authors have suggested that including
taxa with a high proportion of missing characters is poten-
tially problematic for phylogenetic reconstruction (Huelsen-
beck, 1991; Wilkinson, 1993; Kearney, 2002; Hartmann & Vision, 2008), our results are largely in agreement with evi-
dence suggesting that as long as taxa with missing characters
bear sufficient phylogenetically informative data, and the
overall number of characters included in the analysis is not
unusually low (< 100), highly incomplete taxa can be accu-
rate placed within a phylogeny (Wiens & Morrill, 2011;
and references therein). It has previously been suggested that
parametric methods, such as ML and Bayesian analyses,
perform well in placing taxa with missing characters onto
individual branches with high support values by extrapola-
ting topological and branch length inferences from one group
of characters to another (Philippe et al., 2004; Fulton & Stro-
beck, 2006; Wiens & Moen, 2008; Wolsan & Sato, 2010).

Using a sequence data matrix of 3943 bp as a ‘scaffold’ on
which the missing-character taxa could be placed (Wiens et al., 2005; Wiens, 2006), we have shown that the geo-
graphical and taxonomic scope of phylogenetic studies can be aug-
mented with archival specimens by the targeted amplification
of short phylogenetically informative DNA regions.

Historical biogeography of Rhinusa and Gymnetron

Southern African taxa are characterized by deep genetic
divergences, several lineages of which form monophyletic
groups with Palaearctic Rhinusa and Gymnetron lineages. Ba-
yes factor analyses support a phylogenetic topology with
basal lineages of southern African origin. Following the rea-
soning of Bremer (1992) and Hausdorf (1998), plesiomor-
phic areas in a cladogram are more likely to be part of the
ancestral range than apomorphic areas; likewise, areas repre-
sented on more than one branch have a higher probability of
being part of the ancestral area than areas less represented.
Thus, the basal placement of three southern African lineages
within the phylogeny (G. gossypinus, G. bisignatum
+ G. perrinae, and G. buddleiae) supports southern Africa as
the ancestral area from where extant Gymnetron and Rhinusa
species diversity is derived. Congruent with this hypothesis,
results from the s-diva analysis show higher relative proba-
bilities for southern Africa as the ancestral range at the two
most basal nodes in the phylogeny (Fig. 2).

One of the more likely opportunities for large-scale range
expansions from Africa into the Palaearctic, or the Palaearctic
into Africa, probably occurred during the early–mid Miocene
boundary (c. 17 Ma), when a land connection formed between
Europe and Africa after the closure of the Tethys Sea (Levys, 1964; McGuire & Kron, 2005). Several authors have suggested
the existence of geographical and ecological routes connecting
southern and eastern Africa with the Mediterranean region in
the past that would have facilitated range expansions of plants
and animals between the Palaearctic and Afro tropical regions.

Evidence supporting the presence of an ‘arid corridor’ con-
necting southern, eastern and north-eastern African areas prior
to the end of the Miocene has been documented for many
plant groups (Verdcourt, 1969; Goldblatt, 1978; Jürgens, 1997;
Caujapé-Castells et al., 2001; Coleman et al., 2003; Thiv et al.,
2011) and animals (Balinsky, 1962; La Greca, 1990; Freitag & Robinson, 1993; Herron et al., 2005) exhibiting disjunct
distributions. Severance of this corridor has been linked to late
Miocene uplift and associated rifting in Africa and major cli-
matic changes during the late Miocene (Chorowicz, 2005;
Schuster et al., 2006). Both our age estimates and s-diva anal-
yses are consistent with this scenario. Although our analyses
yielded broad HPD intervals, reflecting the broad prior root
age and substitution rate distributions employed, mean esti-
mated divergence times between Palaearctic and southern
African lineages (nodes I–IV; Fig. 2) suggest that divergences
most probably occurred within the late Miocene (c. 11.6–
7.4 Ma). s-diva analysis inferred vicariant events and an
ancestral range including both southern Africa and the Palaear-
colic regions at nodes representing the MRCA of southern
African and Palaearctic lineages (nodes I–IV; Fig. 2). Further
support for the east African arid corridor hypothesis comes
from the presence of several Gymnetron species in eastern
Africa (Caldara et al., 2008), suggesting more northerly African
relicts of an ancestral range once stretching north from south-
ern Africa to East Africa. The land connection between the
African and Asian plates would have facilitated range expan-
sion further north into the Mediterranean and European
regions, with new vegetation zones representing the opportu-
nity to exploit new niches. Climate-mediated vicariant events
would have acted to fragment a once-continuous range as con-
ditions became progressively drier after the mid-Miocene
(Axelrod & Raven, 1978), with periods of greater aridity, par-
sicularly for the eastern region of Africa, in the late Miocene
c. 6 Ma), the Pliocene (c. 3 Ma), and near the Pliocene–Pleis-
tocene boundary (< 2 Ma) (deMenocal, 1995; Bobe, 2006).

The importance of these climatic changes as vicariant events
has been demonstrated for different groups of animals (Leakey
et al., 1996; Douady et al., 2003) and plants (del Hoyo et al.,
2009; Thiv et al., 2010), including examples of mid-to-late
Miocene radiations in the genera Nemesia (Datson et al., 2008)
and Linaria (Fernández-Mazuecos & Vargas, 2011), two host
plant genera utilized by Gymnetron and Rhinusa, respectively.

Within Coleoptera, our results are in accordance with previ-
ously proposed vicariance scenarios for Mediterranean–southern
African disjunct distributions of beetles in the families Meloidae
(Bologna et al., 2008) and Nitidulidae (Audisio et al., 2008). For
both these groups, repeated desertification phenomena since the
Miocene and through the Pliocene and Pleistocene are suggested
to have fragmented probable ancestral distributions extending
between European–Mediterranean and eastern/southern African
areas. It should be noted that our divergence-date estimates may
overestimate the timing of divergence between southern African
and Palaearctic lineages, due to incomplete species sampling and
species extinctions. Additionally, our analyses have incorporated
broad parametric prior distributions to account for uncertainty
both in tree root age, and in molecular evolutionary rate, yielding wide HPD intervals for our estimated divergence times. However, the broad consistency across our four mean age estimates, together with ancestral range reconstruction, argue for a southern African origin of species diversity within Gymnetron and Rhinusa, with subsequent range expansion into the Palaearctic followed by late Miocene vicariance and divergence between southern African Gymnetron lineages and lineages of Gymnetron and Rhinusa in the Palaearctic.

CONCLUSIONS

Rhinusa and Gymnetron are not monophyletic with respect to each other, and present multiple lineages that are disjunct between southern Africa and the Palaearctic. We were able to test biogeographical hypotheses for the origin of their southern African–Palaearctic disjunct distribution by incorporating many species only available as archival material, using prior genetic information to identify short phylogenetically informative amplexics for their approximate phylogenetic placement. Molecular phylogenetic data provide support for a southern African origin, with the earliest divergence within the group estimated to date back to c. 27 Ma, and subsequent range expansion into the Palaearctic. The timing of the northward range expansion of the group into the Palaearctic is not discernible from the tree topology, but would have been facilitated by a Miocene ‘arid corridor’, suggested to have connected southern, eastern and north-eastern African areas. Molecular data permit estimation of the approximate timings of four vicariant events between Palaearctic and southern African lineages, which are consistent with an inferred late Miocene disruption of the arid corridor as a consequence of geological and climatic changes within this period.

ACKNOWLEDGEMENTS

This research was funded by the Mexican Consejo Nacional de Ciencia y Tecnología (CONACYT) and in part by Grant III43001 (Ministry of Education and Science of the Republic of Serbia). Bayesian and maximum-likelihood analyses were carried out on the High Performance Computing Cluster supported by the Research Computing Service at the University of East Anglia. We are grateful to Mark Carine and three anonymous referees for useful suggestions to improve the manuscript.

REFERENCES


SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1 Field-sampled and archival specimens used in this study with GenBank accession numbers except for samples whose DNA was amplified for the short 16S amplicons only.

Appendix S2 Supplementary figures. Leg of Gymnetron villosulum remounted on an entomological card after DNA extraction (Fig. S1), graph showing correlation between uncorrected pairwise distances of 16S rRNA and COII mitochondrial genes (Fig. S2), and Bayesian phylogenetic tree inferred from DNA sequences from five genes (Fig. S3).

Appendix S3 List of primers used in PCR and sequencing reactions.

BIOSKETCH

Gerardo Hernández-Vera recently obtained his PhD at the University of East Anglia, UK, and is interested in molecular phylogenetic methods applied to studies of evolutionary ecology, biogeography, and speciation and diversification processes. Roberto Caldara, Ivo Tosevski and Brent Emerson have been collaborating for several years, focusing on the taxonomy and evolutionary history of weevils within the tribe Mecinini (Curculionidae: Curculioninae).

Author contributions: B.E. conceived and led the project; R.C. and I.T. provided material and assisted with the manuscript and data collection; G.H.-V. analysed the data and wrote the manuscript; B.E. assisted with the manuscript and contributed as senior author.

Editor: Mark Carine