# MOLECULAR PHYLOGENETIC ANALYSIS OF HOST USE AND BIOGEOGRAPHY WITHIN THE GENUS *RHINUSA* AND THE RELATED GENUS *GYMNETRON* (COLEOPTERA: CURCULIONIDAE)

**Gerardo Hernández Vera**

A thesis submitted for the degree of Doctor in Philosophy

at the University of East Anglia, Norwich, UK

September 2011

I dedicate this work to the memory of my father
Juan M. Hernández Santos (1947-2005) and my mother Maria del Refugio Vera Cuevas

# MOLECULAR PHYLOGENETIC ANALYSIS OF HOST USE AND BIOGEOGRAPHY WITHIN THE GENUS *RHINUSA* AND THE RELATED GENUS *GYMNETRON* (COLEOPTERA: CURCULIONIDAE)

**Gerardo Hernández Vera**

September 2011

## Abstract

Using a molecular phylogenetic approach, this thesis addresses questions surrounding the evolutionary history of endophagous parasitic weevils within the genera *Rhinusa* and *Gymnetron* (Coleoptera : Curculionidae), particularly the importance of their ecological interactions with their host plants as a major driver of diversification and shared evolutionary history. Using mitochondrial and nuclear DNA sequence data, phylogenetic analyses revealed that weevils within *Rhinusa* and *Gymnetron* exhibit phylogenetic conservatism in host use at the plant family level; however, shifts between closely related host-plants and in modes of parasitism have played an important role in the diversification of this group of weevils. Similarly, mitochondrial and nuclear DNA sequence data revealed that ecological specialization in weevils that feed, oviposit and develop within fruit capsules of particular host plant taxa can promote ecological divergence and lead to host-associated genetic differentiation, reproductive isolation, and ultimately speciation. The targeted PCR-amplification of short phylogenetically informative DNA sequences from archival samples allowed for the inference of the biogeographic origin of *Rhinusa* and *Gymnetron*, and also contributed toward the clarification of the challenging taxonomy of the group. *Rhinusa* and *Gymnetron* are not reciprocally monophyletic; they represent a complex of relatively young lineages which expanded from southern Africa into the Palaearctic during the late Miocene.

## ACKNOWLEDGEMENTS

My gratitude to Brent; an excellent person and a great supervisor from whom I received guidance and support, within and outside the work environment. I will always be grateful for his help and patience throughout this unique experience called "PhD".

Thanks to CONACyT for making this project possible, funding my studies and research at the University of East Anglia.

Without the help of collaborators, this project would not have been possible. I wish to express my gratitude to Roberto Caldara and Ivo Tosevski for their invaluable help, assistance and supply of samples and information.

I would like to thank my wife Xochitl for bearing with me in this experience; without her patience, help and unconditional support, this project would have been more difficult.

To my family; my parents Juan Manuel and Ma. Del Refugio, my brother Juan Pablo, my sisters Claudia, Viridiana, Marisol and Maricielo. Muchas gracias carnales por su apoyo. Todo mi reconocimiento y agradecimiento para el "Jefe" (dondequiera que este) y la "Jefa" por ser unos padres ejemplares que siempre nos han apoyado en todo momento.

Many thanks to Lorenza, for helping me to "settle down" in the lab and project. To the people in Brent's lab: Eugenia, Andreia, James, Conrad, Cristiana, Alicia and Isabel, thank you for your support and friendship. Thanks to Tim for his help during field-work and "research on pub-lic houses". To all the people I have met in BIO, staff and PhD students within the "Evolution, Ecology and Biodiversity" group at UEA, thank you for making all these years a nice experience!

## TABLE OF CONTENTS

# LIST OF TABLES AND FIGURES

**Chapter 5**

# Chapter 1: General Introduction

# GENERAL INTRODUCTION

## 1. 1. Plant-feeding insects as model systems in evolutionary biology

Approximately over a half of all described species are insects (Gullan & Cranston 2005). This remarkable diversity has attracted the attention of evolutionary biologists and stimulated a plethora of hypotheses that seek to explain this phenomenon. Different aspects related to morphology, behaviour and ecology have been put forward as significant variables that account for the high species-richness within this group (for a review see Mayhew 2007 and references therein). The evolution of morphological structures such as wings, an exoskeleton and external mouthparts adapted to exploit a wide range of foods, have all been invoked as key evolutionary innovations (Mayhew 2007). Among behavioural and ecological explanations, sexual selection/conflict and interactions with other organisms, especially plants, have also been proposed.

Perhaps most researchers agree on the importance of the interaction of insects with plants as one of the main drivers of diversification, where in one way or another, the vast number of different plant resources has spurred the diversification of insects that feed on them (Futuyma & Agrawal 2009; Mayhew 2007; Mitter *et al.* 1988; Rabosky 2009). It has been shown that plant-feeding lineages contain more species than their non-herbivorous sister taxa. Using a comparative approach, Mitter et al. (1988) demonstrated that herbivory has led to increased diversification rates in insects due to repeated invasions into new "adaptive zones". Similarly, using a phylogenetic approach with molecular and morphological characters, Farrell (1998) showed that within the Phytophaga, the largest and oldest radiation of plant-feeding beetles, enhanced rates of diversification are associated with a series of host shifts from

gymnosperms to angiosperms. Indeed, more than 25% of species that currently inhabit the Earth is represented by a plant-feeding insect (Odegaard 2000; Price 2002).

In seeking to explain this enhanced diversity, researchers have found plant-insect interactions as a great opportunity to investigate evolutionary processes such as adaptive radiation, ecological speciation, and coevolution. Perhaps one of the most influential essays that has inspired researchers over recent decades is Ehrlrich and Raven's study of coevolution of butterflies and plants (Ehrlich & Raven 1964). Although previous studies had highlighted the importance of insect-plant interactions (e.g. Dethier 1954; Fraenkel 1959), Ehrlrich and Raven were the first to develop this idea in a historical context. They hypothesized a scenario of "escape and radiation" coevolution between plant and insects (similar to an arms-race scenario), where in response to insect herbivory, plants may evolve chemical defences which enable them escape from their associated insects and radiate into diverse species sharing the novel defence. Under this scenario other insects eventually colonize the new plant lineages from chemically similar host-plants, thus exploiting new "empty niches" and ultimately undergoing adaptive radiation themselves. Thus, repeated iterations of these stepwise adaptive radiations (adaptations and counteradaptations) would be responsible for the high diversity observed in plant-feeding insects.

Although the ubiquity of the particular scenario envisioned by Ehrlich and Raven has been questioned (e.g. Thompson 1988), their model continues to inspire and organize research on the evolution of insect-plant interactions because it exemplifies several themes of the "new synthesis" in evolution; e.g. that diversification is driven primarily by ecological interactions (Schluter 2000). One of

the unique features that makes plant-feeding insects useful models for the study of diversification is their preference for restricted sets of host-plants. Although there are examples of highly polyphagous insects (e.g. Bernays & Minkenberg 1997; Ribeiro *et al.* 2005), most plant-feeding insects are ecologically specialized on reduced sets of host-plant taxa. This attribute represents an opportunity to study mechanisms of reproductive isolation via ecological divergence. The close association of an insect species with its host-plants may amplify the selection pressure imposed by the host and make them more susceptible to ecological divergence following a host shift for example (Funk *et al.* 2002; Mopper & Strauss 1997). A further level of ecological specialization is sometimes observed where plant-feeding insects utilize particular plant tissues for feeding and development. Developmental timing can be influenced by host-plant resources with different phenologies, thus insects from populations specialized on different host-plant resources may mature and mate at different times, leading to temporal isolation (Feder & Filchak 1999; Groman & Pellmyr 2000). Another feature related to host-plant specialization is the formation of biotypes or host-races, where genetically differentiated, sympatric populations of insects use different hosts and between which there may be appreciable, but still limited, gene flow (Dres & Mallet 2002). The continuum of populations exhibiting different degrees of reproductive isolation can be exploited advantageously for comparative analyses of speciation mechanisms.

### 1. 2. Molecular phylogenetics and the study of insect-plant associations

Studies conducted subsequent to Ehrlich and Raven's (1964) essay have focused mainly on the ecological or microevolutionary aspect of plant-insect interactions. Research has included explorations of physiological and sensorial aspects of host-

choice and adaptation to host-plants (Pierce *et al.* 1981; Vandersar & Borden 1977), the functional importance of secondary compounds (Rhoades 1985; Seigler & Price 1976), and quantitative genetic approaches to evaluate tradeoffs in host-plant adaptation (Futuyma & Philippi 1987; Hare & Kennedy 1986). However, advances in phylogenetics, and the increasing accessibility of molecular methods, has more recently contributed to an elevated interest in the original macroevolutionary processes that were the focus of Ehrlich and Raven's (1964) attention. Two methodological advances have contributed significantly to the expansion of knowledge and understanding of the organismal evolutionary history: (1) the development of molecular genetic techniques based on the Polymerase Chain Reaction (PCR), and (2) the development and refinement of analytical methods for inferring phylogenetic relationships among taxa. The first development has allowed researchers to access vast amounts of information contained within nucleic acids, and the second has facilitated the processing of this information in an evolutionary context in an efficient and meaningful way.

Mitter and Brooks (1983) were among the first researchers to capitalise upon these methodological advances for the study of insect-plant interactions, establishing the importance of combining ecological and systematic approaches and advocating the use of phylogenetic trees to infer the history of insect-plant associations (Mitter & Brooks 1983; Mitter *et al.* 1991). This combined approach enables one to address questions of when, how and why ecological associations originate in geological time, and numerous studies have emerged that have explicitly tested the macroevolutionary aspects of Ehrlich and Raven's theory (e.g. Berenbaum 1983; Farrell & Mitter 1990; Miller 1987; Mitter *et al.* 1988). Within the last 10 years a proliferation of molecular

phylogenetic information has led to a growing appreciation of the role that ecological traits have in determining phylogenetic patterns between insects and their host-plants (Clayton *et al.* 2004; Weiblen & Bush 2002). Furthermore, phylogenetic information has been incorporated in comparative methods allowing for the statistical analysis of correlations between ecological traits and the evolutionary relationships of organisms (Harvey & Pagel 1991).

The use of molecular phylogenetic analyses has also been expanded to study the composition and structure of ecological assemblages or communities over geographic areas and geological time (Ricklefs & Schluter 1993). The term "molecular biogeography" was coined by Caccone et al. (1994) and retaken by Lavin et al. (2000). This approach attempts to reconstruct the biogeographic history of taxa using a taxon cladogram (phylogenetic tree) obtained from molecular data, with the additional application of the molecular clock. Today, molecular phylogenetics is a powerful and versatile tool allowing researchers to address a broad range of questions that may in part be explained by evolutionary history. In the study of insects and their interactions with plants, phylogenetic methodology is providing researchers with increasingly powerful ways to understand how ecological interactions are moulded over evolutionary time by a combination of historical constraints and current ecology (Thompson 2002).

### 1. 3. Why studying weevils (Coleoptera : Curculionoidae)?

The order Coleoptera includes approximately 400,000 described species, representing ~40% of the total insect species described (Gullan & Cranston 2005). Within the order Coleoptera, the superfamily Curculionoidae comprises approximately 5,800 genera

and 62,000 described species of plant-feeding beetles, commonly referred to as weevils (Oberprieler *et al.* 2007). These are included in the six relatively depauperate families Nemonychidae, Anthribidae, Attelabidae, Belidae, Brentidae, and Rhynchophoridae and the considerably more diverse Curculionidae (Thompson 1992). Weevils have been described as one of the best examples of successful adaptive radiation, exploiting nearly every plant taxon and plant structure, including roots, stems, leaves, flowers, fruits, and seeds; some species also utilize dead and decaying plant material (Anderson 1995; Marvaldi *et al.* 2002; Oberprieler *et al.* 2007). Different explanations have been put forward to try and explain the remarkable diversity observed within this group of beetles, including the evolution of key morphological and ecological innovations and their associations with flowering plants. Among the morphological evolutionary innovations recognized as playing a significant role is the evolution of a rostrum or elongated mouthparts (Fig. 1). This structure, characteristic of weevils, has been hypothesized to confer two evolutionary advantages over other herbivorous beetles; the first is the ability to use it for "drilling" different plant tissues and structures, the second is the ability to use the rostrum as an oviposition tool, capable of inserting eggs inside plant tissues (Anderson 1995; Oberprieler *et al.* 2007). In turn, this ability has allowed weevil species to access new plant resources contributing to a shift in larval habits, from external plant feeders to endophytic habits. This ecological innovation is thought to have played a significant role in their diversification coupled with the evolution of flowering plants on Earth, which represented an opportunity as "new empty niches" to exploit. The interaction of weevils with angiosperms has been recognized as one of the key evolutionary events that promoted their successful adaptive radiation (Farrell 1998; McKenna *et al.* 2009; Oberprieler *et al.* 2007). However, as suggested by Oberprieler et al. (2007), rather

than a single explanation for the extraordinary diversity of weevils, the interaction of these and other factors is the more probable explanation of the diversity we see today. The emergence in cascade of all of these evolutionary innovations likely enhanced speciation rates by facilitating the colonisation of diverse new ecological niches (Fig. 2). Similarly, Hunt et al. (2007) recognize the importance of herbivory in the diversification of some beetle lineages, however, they suggest that the trait per se does not explain why beetles are so diverse. Instead, high survival of lineages and sustained diversification might be responsible for their enhanced species number. Thus, as one of the most diverse groups of organisms on Earth, weevils represent a great opportunity to study the evolutionary mechanisms responsible for such diversity.

An interesting outcome of the evolutionary history of weevils is the recurrent proliferation of related species, often differing from one another in only small ways, perhaps facilitating the reduction of ecological overlap (Marvaldi *et al.* 2002), a feature that complicates the delineation of species and taxonomic assignment. Phenotypic variation due to plasticity in response to environmental factors also contributes to the difficulty of the group in terms of classification based on morphological characters (Marvaldi *et al.* 2002; Thompson 1992). In fact, weevils have been described as "the largest outstanding problem in the higher classification of Coleoptera" (Crowson 1955; Oberprieler *et al.* 2007). With the advent of PCR-based techniques and more refined methods to infer phylogenies, molecular markers have proved to be useful tools in taxonomic and systematic studies, identifying natural groups and their relationships (Avise 2004). Thus, molecular phylogenetic techniques provide an opportunity to address taxonomic issues within weevils that might be intractable with the sole use of morphological characters.

Because weevils may be among the first enemies to consume healthy plants (Marvaldi *et al.* 2002), including many of those also utilized by humans, they have an economic impact as pests of important crops such as rice (Zou *et al.* 2004), maize (Beti *et al.* 1995), wheat (Sinha 1984), barley (Athanassiou *et al.* 2005) and cotton (Smith 1998) among others. Conversely, because weevils are frequently specialized on a reduced number of host-plants, they are also economically important as potential biological control agents against weeds and invasive plant species. They have been tested against a variety of these plants including taxa within the families Brassicaceae (Fumanal *et al.* 2004), Boraginaceae (De Clerck-Floate & Schwarzlander 2002), Proteaceae (Kluge & Gordon 2004), Polygonaceae (Lake *et al.* 2011) and Scrophulariaceae (Schat *et al.* 2011).

### 1. 4. The tribe Mecinini and the *Rhinusa*/*Gymentron* species complex

A good example of the above mentioned taxonomic difficulties encountered in the classification of weevils is found within representatives of the tribe Mecinini, included in the family Curculionidae subfamily Curculioninae. Based on morphological characters, Caldara (2001) proposed six genera within the tribe: *Mecinus, Gymnetron, Rhinusa, Rhinumiarus, Cleopomiarus and Miarus*, however, he recognizes that the systematics of the tribe and the relationships among its constituent taxa "are not unequivocal" because of the lack of sufficient shared derived traits unique to the terminal groups. *Mecinus*, *Gymnetron* and *Rhinusa* appear very closely related to each other, *Cleopomiarus* is very closely related with *Miarus*, whereas *Rhinumiarus* occupies an intermediate position (Caldara 2001). Before Caldara's (2001) taxonomic revision, *Rhinusa* had been treated as a subgenus of *Gymnetron* (e.g. Hoffmann 1958). Because of their very subtle morphological differences, sometimes it is difficult to

separate species of *Rhinusa* from *Gymnetron* and vice versa. Despite the recognition of just a few apomorphies, Caldara (2001) concluded that they should be treated as separate genera.

In this thesis, attention is focused on representatives within *Rhinusa* and *Gymnetron*, a group of weevils of economic importance as biological control agents whose endoparasitic habits on a restricted set of host plants make them suitable for the study of insect-plant interactions. The genus *Rhinusa* Stephens, 1829 comprises approximately 40 species with a Palaearctic distribution. All species live on representatives of the plant families Scrophulariaceae and Plantaginaceae (Caldara 2001; Caldara *et al.* 2010). Within the family Scrophulariaceae, *Verbascum* and *Scrophularia* are utilized as hosts, whereas within Plantaginaceae *Linaria*, *Kickxia*, *Chaenorhinum*, *Antirrhinum* and *Misopates* are known host-plants (Caldara *et al.* 2010). The genus *Gymnetron* Schönherr, 1825 includes approximately 30 species with a Palaearctic distribution and approximately 60 species from the Afrotropical region, of which 55 are known mainly from South Africa and considered to be endemic to this area (Caldara 2003). All Palaearctic species use plant species from the genus *Veronica* (Plantaginaceae) as host plants, whereas species from the Afrotropical region use different host plants within the family Scrophulariaceae, namely *Hebenstreitia*, *Sutera*, *Selago*, *Buddleja*, *Diascia*, *Nemesia*, and *Hemimeris*. Individuals have also been collected on plants of the genus *Anastrebe* in the family Stilbaceae (Caldara *et al.* 2008).

With regard to the life history of species within *Gymnetron* and *Rhinusa*, they are univoltine insects exhibiting a very close relationship with their host plants; this is

evidenced by their life cycle that is tightly linked to that of their host plants. As endophagous parasitic insects, females oviposit inside the ovaries, stems, or roots of their host plants where larvae develop and pupate. Adults emerge approximately after 45-60 days. Some species elicit plant physiological responses inducing galls, whereas others act as inquilines of the galls induced by the former ones (Gassmann & Paetel 1998; Groppe 1992; Tosevski & Gassmann 2002). Because of their close association with their host-plants and specialized parasitic habits, some species have been used as biological control agents against species of toadflax that have been introduced to Canada and the United States, and have become a problem as pests. *Linaria vulgaris* and *L. dalmatica* are weeds avoided by cattle and spread rapidly displacing useful native plants. In 1993 individuals of *Rhinusa antirrhini* were released in North America as part of a biological control programme with partially successful results due in part to behavioural variation within the species. A further study revealed the presence of cryptic species with different host-plant affiliations (see chapter 2). Other species that have been trialed as biological control agents are *R. neta* (Gassmann & Paetel 1998), and *R. thapsicola* (Tosevski & Gassmann 2002).

**1. 5. Thesis structure**

Using molecular phylogenetic methods, in this thesis I address different aspects of the evolutionary history of parasitic weevils within *Rhinusa* and *Gymnetron*; particularly their interaction with host-plants and its significance in an evolutionary context.

In chapter 2, using nuclear and mitochondrial DNA sequences I assess host-associated genetic differentiation within *Rhinusa antirrhini*. Additionally, cross-copulation experiments by collaborators in Serbia are used to assess the possibility of

reproductive isolation between weevils with different host-plant affiliation. The importance of ecological factors as drivers of diversification is discussed in the light of the results from both molecular and behavioural data. This research has been published in the journal *Molecular Ecology* in 2010.

In chapter 3, biogeographical hypotheses concerning the geographic origin of *Rhinusa* and *Gymnetron* are investigated exploiting the use of PCR-based techniques and archival specimens to augment sample numbers from different geographic origins. Using a non-destructive approach, single legs from dried weevils from an entomological collection were used as a source of DNA in addition to DNA from non-archival samples.

Chapter 4 is an investigation of phylogenetic conservatism in ecological traits across species of *Rhinusa* and *Gymnetron* and a revision of the systematics of both genera. DNA sequences from two mitochondrial and three nuclear gene fragments were used to reconstruct phylogenetic relationships and test hypotheses of conservatism in host-plant utilization and parasitic mode. Ancestral states of host plant family utilization were also reconstructed using maximum likelihood optimization.

In Chapter 5 I follow up atypical results observed during data analyses for chapter 2, where some mtDNA sequence chromatograms were found to exhibit ambiguities in several nucleotide positions in the form of double peaks. PCR assays were utilized to assess the nature of these sequence ambiguities.

A general conclusion is presented in chapter 6, where results and findings of this thesis are summarized, discussing their significance in a broader context. Future directions and further research questions are also put forward.

**Figure 1.** Detail of a long rostrum in the weevil *Curculio proboscideus* (Curculionidae: Curculioninae), the characteristic structure of this group of plant-feeding insects. (Taken from McKenna et al. 2009. Photo credit: D. McKenna).

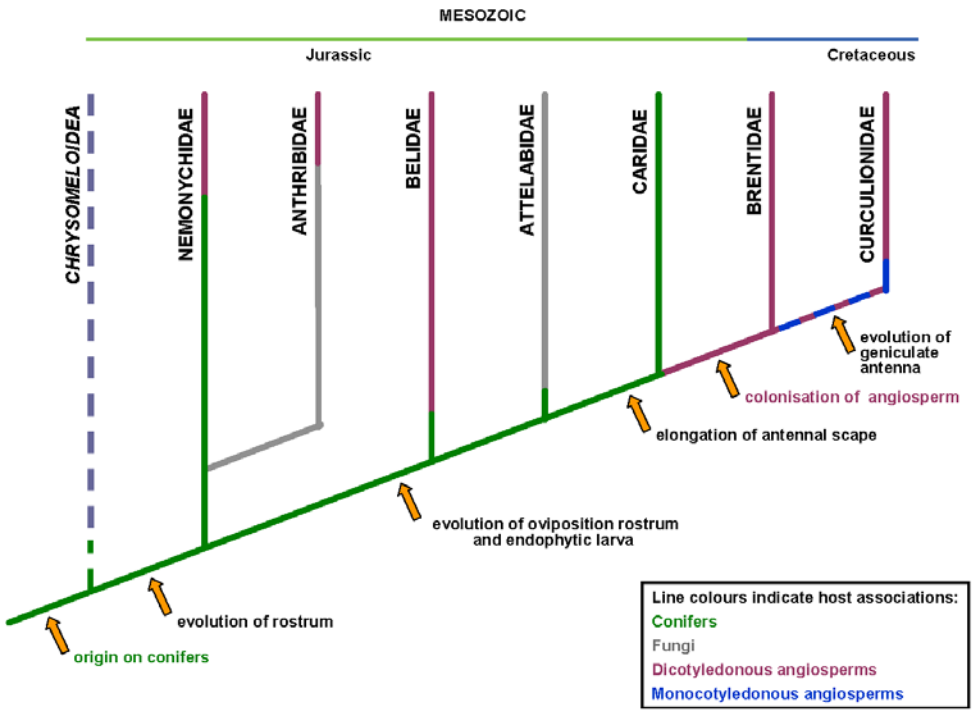


**Figure 2.** Schematic representation of key evolutionary events in the diversification of Curculionoidea (Modified from Oberprieler et al. 2007).

## REFERENCES

Anderson RS (1995) An evolutionary perspective of diversity in Curculionoidea. *Mem. Entomol. Soc. Wash.* **14**, 103-114.

Athanassiou CG, Kavallieratos NG, Economou LP, *et al.* (2005) Persistence and efficacy of three diatomaceous earth formulations against Sitophilus oryzae (Coleoptera : Curculionidae) on wheat and barley. *Journal of Economic Entomology* **98**, 1404-1412.

Avise JC (2004) *Molecular Markers, Natural History, and Evolution. 2nd edition* Sinauer Associates, Sunderland, Massachusetts.

Berenbaum M (1983) Coumarins and caterpillars - A case for coevolution. *Evolution* **37**, 163-179.

Bernays EA, Minkenberg O (1997) Insect herbivores: Different reasons for being a generalist. *Ecology* **78**, 1157-1169.

Beti JA, Phillips TW, Smalley EB (1995) Effects of maize weevils (Coleoptera, Curculionidae) on production of aflatoxin B-1 by Aspergillus flavus in stored corn. *Journal of Economic Entomology* **88**, 1776-1782.

Caccone A, Milinkovitch MC, Sbordoni V, Powell JR (1994) Molecular biogeography: using the Corsica-Sardinia microplate disjunction to calibrate mitochondrial rDNA evolutionary rates in mountain newts (Euproctus). *Journal of Evolutionary Biology* **7**, 227-245.

Caldara R (2001) Phylogenetic analysis and higher classification of the tribe Mecinini (Coleoptera: Curculionidae, Curculioninae). *Koleopterologische Rundschau* **71**, 171-203.

Caldara R (2003) Revisione dei Gymnetron della Regione Afrotropicale (Coleoptera Curculionidae). *Mem. Soc. Entomol. Ital.* **82**, 87-272.

Caldara R, Colonnelli E, Osella G (2008) Curculionidae Curculioninae: relationships between Mediterranean and southern African species (Coleoptera). *Biogeographia* **29**, 133-156.

Caldara R, Sassi D, Tosevski I (2010) Phylogeny of the weevil genus *Rhinusa* Stephens based on adult morphological characters and host plant information (Coleoptera: Curculionidae). *Zootaxa* **2627**, 39-56.

Clayton DH, Bush SE, Johnson KP (2004) Ecology of congruence: Past meets present. *Systematic Biology* **53**, 165-173.

Crowson RA (1955) *The Natural Classification of the Families of Coleoptera* N. Lloyd, London, UK (E. W. Classey: Hampton, UK, reprint 1967).

De Clerck-Floate R, Schwarzlander M (2002) Host specificity of Mogulones cruciger (Coleoptera : Curculionidae), a Biocontrol agent for houndstongue (Cynoglossum officinale), with emphasis on testing of native North American Boraginaceae. *Biocontrol Science and Technology* **12**, 293-306.

Dethier VG (1954) Evolution of feeding preferences in phytophagous insects. *Evolution* **8**, 33-54.

Dres M, Mallet J (2002) Host races in plant-feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **357**, 471-492.

Ehrlich PR, Raven PH (1964) Butterflies and plants - a study in coevolution. *Evolution* **18**, 586-608.

Farrell B, Mitter C (1990) Phylogenesis of insect plant interactions - Have Phyllobrotica leaf beetles (Chrysomelidae) and the Lamiales diversified in parallel? *Evolution* **44**, 1389-1403.

Farrell BD (1998) "Inordinate fondness" explained: Why are there so many beetles? *Science* **281**, 555-559.

Feder JL, Filchak KE (1999) It's about time: the evidence for host plant-mediated selection in the apple maggot fly, Rhagoletis pomonella, and its implications for fitness trade-offs in phytophagous insects. *Entomologia Experimentalis Et Applicata* **91**, 211-225.

Fraenkel GS (1959) Raison d'etre of secondary plant substances. *Science* **129**, 1466-1470.

Fumanal B, Martin JF, Sobhian R, Blanchet A, Bon MC (2004) Host range of Ceutorhynchus assimilis (Coleoptera : Curculionidae), a candidate for biological control of Lepidium draba (Brassicaceae) in the USA. *Biological Control* **30**, 598-607.

Funk DJ, Filchak KE, Feder JL (2002) Herbivorous insects: model systems for the comparative study of speciation ecology. *Genetica* **116**, 251-267.

Futuyma DJ, Agrawal AA (2009) Macroevolution and the biological diversity of plants and herbivores. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 18054-18061.

Futuyma DJ, Philippi TE (1987) Genetic variation and covariation in responses to host plants by Alsophila pometaria (Lepidoptera, Geometridae). *Evolution* **41**, 269-279.

Gassmann A, Paetel C (1998) Gymnetron netum (Col.; Curculionidae). A potential agent for biological control of toadflax in North America., p. 33. Centre for Agriculture and Biosciences International, Delémont, Switzerland.

Groman JD, Pellmyr O (2000) Rapid evolution and specialization following host colonization in a yucca moth. *Journal of Evolutionary Biology* **13**, 223-236.

Groppe K (1992) Gymnetron antirrhini Paykull (Col.: Curculionidae). A candidate for biological control of Dalmatian toadflax in North America, p. 22. Centre for Agriculture and Bioscience International, Delémont, Switzerland.

Gullan PJ, Cranston PS (2005) *The insects: an Ouline of Entomology. Third edition* Blackwell, Oxford, UK.

Hare JD, Kennedy GG (1986) Genetic variation in plant-insect associations - Survival of Leptinotarsa decemlineata populations on Solanum carolinense. *Evolution* **40**, 1031-1043.

Harvey PH, Pagel MD (1991) *The comparative method in evolutionary biology* Oxford University Press, Oxford.

Hoffmann A (1958) Coléoptères Curculionides (Troisième Partie). In: *Faune de France* (eds. de Beauchamp P, Chopard L), pp. 1209-1839. Paul Lechevalier, Paris.

Hunt T, Bergsten J, Levkanicova Z, *et al.* (2007) A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science* **318**, 1913-1916.

Kluge RL, Gordon AJ (2004) The fixed plot survey method for determining the host range of the flowerbud-feeding weevil Dicomada rufa, a candidate for the biological control of Hakea sericea in South Africa. *Biocontrol* **49**, 341-355.

Lake EC, Hough-Goldstein J, Shropshire KJ, D'Amico V (2011) Establishment and dispersal of the biological control weevil *Rhinoncomimus latipes* on mile-a-minute weed, *Persicaria perfoliata*. *Biological Control* **58**, 294-301.

Lavin M, Thulin M, Labat JN, Pennington RT (2000) Africa, the odd man out: Molecular biogeography of dalbergioid legumes (Fabaceae) suggests otherwise. *Systematic Botany* **25**, 449-467.

Marvaldi AE, Sequeira AS, O'Brien CW, Farrell BD (2002) Molecular and morphological phylogenetics of weevils (Coleoptera, Curculionoidea): Do niche shifts accompany diversification? *Systematic Biology* **51**, 761-785.

Mayhew PJ (2007) Why are there so many insect species? Perspectives from fossils and phylogenies. *Biological Reviews* **82**, 425-454.

McKenna DD, Sequeira AS, Marvaldi AE, Farrell BD (2009) Temporal lags and overlap in the diversification of weevils and flowering plants. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 7083-7088.

Miller JS (1987) Host-plant relationships in the Papilionidae Lepidoptera. Parallel cladogenesis or colonization? *Cladistics* **3**, 105-120.

Mitter C, Brooks DR (1983) Phylogenetic aspects of coevolution. In: *Coevolution* (eds. Futuyma DJ, Slatkin M), pp. 65-98. Sinauer, Sunderland, MA.

Mitter C, Farrell B, Futuyma DJ (1991) Phylogenetic studies of insect-plant interactions: insights into the genesis of diversity. *Trends in Ecology & Evolution* **6**, 290-293.

Mitter C, Farrell B, Wiegmann B (1988) The phylogenetic study of adaptive zones - Has phytophagy promoted insect diversification? *American Naturalist* **132**, 107-128.

Mopper S, Strauss SY (1997) Genetic Structure and Local Adaptation in Natural Insect Populations: Effects of Ecology, Life History, and Behavior. Chapman and Hall, New York; London.

Oberprieler RG, Marvaldi AE, Anderson RS (2007) Weevils, weevils, weevils everywhere. *Zootaxa*, 491-520.

Odegaard F (2000) How many species of arthropods? Erwin's estimate revised. *Biological Journal of the Linnean Society* **71**, 583-597.

Pierce AM, Borden JH, Oehlschlager AC (1981) Olfactory response to beetle-produced volatiles and host-food attractants by Oryzaephilus surinamensis and Oryzaephilus mercator (Coleoptera, Cucujidae). *Canadian Journal of Zoology-Revue Canadienne De Zoologie* **59**, 1980-1990.

Price PW (2002) Species interactions and the evolution of biodiversity. In: *Plant–Animal Interactions: An Evolutionary Approach* (eds. Herrera CM, Pellmyr O), pp. 3-25. Blackwell, Oxford, UK.

Rabosky DL (2009) Ecological limits and diversification rate: alternative paradigms to explain the variation in species richness among clades and regions. *Ecology Letters* **12**, 735-743.

Rhoades DF (1985) Offensive-defensive interactions between herbivores and plants - Their relevance in herbivore population dynamics and ecological theory. *American Naturalist* **125**, 205-238.

Ribeiro SP, Borges PAV, Gaspar C, *et al.* (2005) Canopy insect herbivores in the Azorean Laurisilva forests: key host plant species in a highly generalist insect community. *Ecography* **28**, 315-330.

Ricklefs RE, Schluter D (1993) *Species diversity in ecological communities: historical and geographical perspectives* University of Chicago Press, Chicago, IL.

Schat M, Sing SE, Peterson RKD, Menalled FD, Weaver DK (2011) Growth Inhibition of Dalmatian Toadflax, Linaria dalmatica (L.) Miller, in Response

to Herbivory by the Biological Control Agent Mecinus janthinus Germar. *Journal of Entomological Science* **46**, 232-246.

Schluter D (2000) *The ecology of adaptive radiation* Oxford University Press, Oxford.

Seigler D, Price PW (1976) Secondary compounds in plants - Primary functions. *American Naturalist* **110**, 101-105.

Sinha RN (1984) Effects of weevil (Coleoptera, Curculionidae) infestation on abiotic and biotic quality of stored wheat. *Journal of Economic Entomology* **77**, 1483-1488.

Smith JW (1998) Boll weevil eradication: Area-wide pest management. *Annals of the Entomological Society of America* **91**, 239-247.

Thompson JN (1988) Coevolution and alternative hypotheses on insect plant interactions. *Ecology* **69**, 893-895.

Thompson JN (2002) Plant-animal interactions: future directions. In: *Plant-animal interactions: an evolutionary approach* (eds. Herrera CM, Pellmyr O), pp. 236-247. Blackwell, Oxford, UK.

Thompson RT (1992) Observations on the morphology and classification of weevils (Coleoptera, Curculionoidea) with a key to major groups. *Journal of Natural History* **26**, 835-891.

Tosevski I, Gassmann A (2002) Biological control of Dalmatian and yellow toadflax, *Linaria dalmatica* and *L. vulgaris*. Centre for Agriculture and Biosciences International, Delémont, Switzerland.

Vandersar TJD, Borden JH (1977) Aspects of host selection behavior of Pissodes strobi (Coleoptera - Curculionidae) as revealed in laboratory feeding bioassays. *Canadian Journal of Zoology-Revue Canadienne De Zoologie* **55**, 405-414.

Weiblen GD, Bush GL (2002) Speciation in fig pollinators and parasites. *Molecular Ecology* **11**, 1573-1578.

Zou L, Stout MJ, Dunand RT (2004) The effects of feeding by the rice water weevil, Lissorhoptrus oryzophilus Kuschel, on the growth and yield components of rice, Oryza sativa. *Agricultural and Forest Entomology* **6**, 47-53.

# Chapter 2: Host-associated genetic differentiation in a seed parasitic weevil *Rhinusa antirrhini* (Coleoptera: Curculionidae) revealed by mitochondrial and nuclear sequence data

Note: This chapter has been published in the journal *Molecular Ecology* in 2010. Cross-copulation experiments were carried out by collaborators in Serbia.

# HOST ASSOCIATED GENETIC DIFFERENTIATION IN A SEED PARASITIC WEEVIL *Rhinusa antirrhini* (COLEOPTERA: CURCULIONIDAE) REVEALED BY MITOCHONDRIAL AND NUCLEAR SEQUENCE DATA.

## ABSTRACT

Plant feeding insects and the plants they feed upon represent an ecological association that is thought to be a key factor for the diversification of many plant feeding insects, through differential adaptation to different plant selective pressures. While a number of studies have investigated diversification of plant feeding insects above the species level, relatively less attention has been given to patterns of diversification within species, particularly those that also require plants for oviposition and subsequent larval development. In the case of plant feeding insects that also require plant tissues for the completion of their reproductive cycle through larval development, the divergent selective pressure not only acts on adults, but on the full life history of the insect. Here we focus attention on *Rhinusa antirrhini* (Curculionidae), a species of weevil broadly distributed across Europe that both feeds on, and oviposits and develops within species of the plant genus *Linaria* (Plantaginaceae). Using a combination of mtDNA (COII) and nuclear DNA (EF1-α) sequencing and copulation experiments we assess evidence for host associated genetic differentiation within *R. antirrhini*. We find substantial genetic variation within this species that is best explained by ecological specialisation on different host plant taxa. This genetic differentiation is most pronounced in the mtDNA marker, with patterns of genetic variation at the nuclear marker suggesting incomplete lineage sorting and/or gene flow between different host plant forms of *R. antirrhini*, whose origin is estimated to date to the mid-Pliocene (3.77 mya; 2.91 - 4.80 mya).

## 2.1. INTRODUCTION

With approximately 4,600 genera and 51,000 described plant feeding species (Oberprieler *et al.* 2007), the family Curculionidae, commonly referred to as weevils, constitutes the largest family in the animal kingdom based on the number of recognised species (Marvaldi *et al.* 2002; Thompson 1992). This successful adaptive radiation has been linked to the origin of angiosperms, the evolution of a rostrum and shifts in larval feeding habits (Marvaldi *et al.* 2002; Oberprieler *et al.* 2007). An additional likely driver of this radiation is the specialisation of many weevil species onto one or a few host plant species. Specific host association, although not characteristic of all weevil species, is a distinctive and recurrent feature across many genera within the Curculionidae. However, host specificity within weevils has received relatively little attention at the molecular level, particularly within species (but see Barat *et al.* 2008; Downie *et al.* 2008; Erney *et al.* 1996 for examples). This is perhaps surprising given the economic importance of weevils both as agricultural pests and biocontrol agents.

Here we assess evidence for host specialisation and reproductive isolation within a weevil species across multiple host plant taxa over a wide geographic range. Our focal species is *Rhinusa antirrhini*, a univoltine fruit feeding weevil described from Sweden and subsequently reported from across Europe and western Russia (Caldara 2004). The genus *Rhinusa* is comprised of approximately 45 species distributed across the Palaearctic region (Caldara 2001) with life histories that involve endogenous parasitism, with larvae developing and feeding exclusively inside either

fruit capsules, stems or roots, and inducing galls in some cases (Caldara 2001; Gassmann & Paetel 1998; Groppe 1992).

In North America, *R. antirrhini* is thought to have been accidentally introduced from its native Eurasia in the early 1900s (De Clerck-Floate & Harris 2002), using *Linaria vulgaris* as primary host. A South Macedonian population of *R. antirrhini* collected from *Linaria dalmatica ssp. macedonica* was deliberately introduced in Canada in 1993 for biological control of *L. dalmatica* (De Clerck-Floate & Harris 2002). Both the adventive strain and the introduced strain of *R. antirrhini* have shown only limited population increase and success in the reduction of *Linaria* populations in North America (De Clerck-Floate & Harris 2002; McClay & De Clerck-Floate 2002). This may in part be due to behavioural variation within the species, as experimental observations for *R. antirrhini* collected from *L. vulgaris* and *L. dalmatica* indicate that individuals of *R. antirrhini* exhibit strong preference for oviposition on the host plant species from which they are collected (Toševski, unpublished data). However, the heritability of this behaviour is unknown. It has long been suggested that experience in the natal habitat can play an important role in shaping habitat preferences of dispersing animals, and particularly of phytophagous insects (Davis & Stamps 2004 and references therein). Several theories, namely the Hopkins principle (Dethier 1954), the neo-Hopkins principle (Jaenike 1983) and the chemical legacy hypothesis (Corbet 1985) have been proposed as non-genetic explanations for this phenomenon. However, it is likely that multiple factors underlie the process (Barron 2001). In a study of host preferences in a granary weevil, Rietdorf and Steidle (2002), conclude that larval and early adult experience as well as genetic predisposition, can determine host preference. Thus, conditioning,

genetics and selection could all be involved in the formation of a preference for a new host and eventually in the formation of a host race (Barron 2001). Indeed, the close association between endophagous insects and their host plants may amplify the selection pressure imposed by the host (Mopper 1996), thus making endophagous insects particularly susceptible to ecological divergence following a host shift (Berlocher & Feder 2002; Funk *et al.* 2002; Mopper & Strauss 1997).

The occurrence of genetically distinct host forms in other phytophagous insect groups has been documented in detail (e.g. Feder *et al.* 1988; Sword *et al.* 2005; Via 1999; Via *et al.* 2000; Waring *et al.* 1990), suggesting that ecological specialization has played an important role in their diversification and speciation. A recent assessment of host associated differentiation within a community of insects utilising goldenrods (Stireman *et al.* 2005) shows that this could be a recurrent phenomenon, contributing significantly as a mode of speciation in phytophagous insects. However, despite the great diversity of insects and their hosts, evolutionary studies of intra-specific patterns of host use are few in number. Our aims are to: (1) assess levels of genetic variation within *R. antirrhini*; (2) evaluate to what extent the distribution of this variation corresponds to host plant; (3) assess evidence for reproductive barriers between weevils sampled from two different host plant species and (4) infer the timing of origin of any observed host entities within *R. antirrhini*. The host genus *Linaria* (Plantaginaceae, formerly Scrophulariaceae) has its centre of origin in the Mediterranean region and comprises approximately 150 species with a holarctic distribution (Hong 1983; Sutton 1988).We have sampled *R. antirrhini* from the four known host species of *Linaria* (within sections Linaria and Speciosae), including several recognised subspecies within two of these. *Linaria genistifolia*, the

type species for section Speciosae, is naturally distributed in central and eastern Europe, extending across through the Asian regions of Turkey, Kazakstan and central Asia to west China (Sutton 1988). We have sampled from five subspecies: *Linaria genistifolia genistifolia, L. genistifolia sofiana, L. genistifolia linifolia, L. genistifolia confertiflora* and *L. genistifolia artvinensis*. *Linaria dalmatica* is naturally distributed in southern Europe (mainly the Balkan peninsula) and south west Asia (Sutton 1988) and we have sampled from the subspecies *L. dalmatica macedonica*. We have also sampled from *L. rubioides*, an endemic toadflax from the Balkan region and *L. vulgaris*, the type species of section Linaria, native to western and eastern Europe, but widely naturalized in temperate regions (Sutton 1988). Within this last species we have sampled from multiple locations, from the United Kingdom through to Russia to provide some estimate of the role of geography in structuring genetic variation. Additionally we have sampled *R. florum*, an ecologically and phylogenetically related species that uses *L. genistifolia genistifolia* as a host in south-eastern Europe (Caldara 2008). To achieve aims 1-2 and 4 we use DNA sequence variation for the mitochondrial COII gene and the nuclear EF1-α gene. The mitochondrial COII gene is a powerful marker for the discrimination of evolutionary divergence of host-plant choice for oviposition because of its female inheritance combined with its high mutation rate and small effective population size (0.25) relative to the nuclear genome. The contrasting biparental inheritance, lower mutation rate and larger effective population size for the EF1-α gene are expected to result in less discriminatory power for intraspecific differentiation, but will complement an analysis of mtDNA variation. To attain our third aim we have undertaken cross copulation experiments between weevils sampled from *Linaria*

*vulgaris* and *Linaria genistifolia genistifolia* and we evaluate the implications of these results in the light of our molecular data.

## 2.2. MATERIALS AND METHODS

### 2.2.1. Insect sampling for molecular analysis

We sampled 93 individuals of *Rhinusa antirrhini* collected over a broad range within the species distribution from 8 different host-plant taxa of *Linaria* (Figure 1 and Table 1) plus two individuals of the related species *R. florum* collected on *L. genistifolia genistifolia*. Specimens were labelled, placed individually in 96% ethanol and stored at 4°C until DNA extraction. *Rhinusa griseohirta* was sampled from *Antirrhinum graniticum* and used as an outgroup.

### 2.2.2. DNA extraction, PCR and sequencing reactions

Individual weevils were punctured through the abdomen and total genomic DNA was extracted using the QIAGEN DNeasy extraction Kit (QIAGEN) following the manufacturer's instructions. After DNA extraction, weevils were placed again in 96% ethanol and maintained at 4°C as vouchers. A fragment of between 758-782 bp of the COII gene was amplified using the primers TL2-J-3038 (5'-TAATATGGCAGATTAGTGCATTGGA) (Emerson *et al.* 2000) and TK-N 3782 (5'-GAGACCATTACTTGCTTTCAGTCATCT) (EVA-Harrison Laboratory, Cornell University, Ithaca, NY, USA). Primers EF1-Bf (5'-AGAACGTGAACGTGGTATCA) and EF1-Br (5'-

CTTGGAGTCACCAGCTACATAACC) were used to amplify a fragment of between 877-897 bp of the EF1-α gene.

Polymerase chain reactions (PCR) contained $NH_4$ buffer (1x), 2.5 - 3.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.1 - 0.5 µM of each primer and 0.5 U of Taq polymerase (Bioline) in a 25µL final volume. PCR cycles were carried out using the following thermal profile for COII: 95°C for 3 min, 33 cycles at 95°C for 1 min, annealing temperatures between 48-58°C for 1 min, 72°C for 1 min, and a final extension at 72°C for 3 min. For EF1-α a touchdown profile was used (Don *et al.* 1991): 94°C for 1 min 30 s, 14 cycles at 94°C for 30 s, 62°C for 1 min, 72°C for 1 min, decreasing the annealing temperature by 2 degrees every 2 cycles down to 50°C, then 24 cycles at 94°C for 30 s, 48°C for 1 min, 72°C for 1 min, and 72°C for 7 min as a final extension. PCR products were cleaned using the QIAquick PCR clean-up kit (QIAGEN), and a PerkinElmer BigDye terminator reaction protocol was followed to generate sequences in a PerkinElmer ABI3700 automated sequencer using the same primers for amplification reactions. Sequences for COII were obtained with the forward primer only whereas the EF1-α fragment was sequenced in both directions.

### 2.2.3. Sequence alignment and haplotype reconstruction

COII sequences were aligned by eye using BioEdit version 7.0.9 (Hall 1999). EF1-α forward and reverse sequences were assembled as contigs using BioEdit version 7.0.9 (Hall 1999) and Lasergene Seqman version 6 (DNASTAR, Inc.), and automatically aligned using the CLUSTAL method with further manual alignment. EF1-α haplotypes from heterozygous individuals were reconstructed with PHASE

version 2.1.1 (Stephens *et al.* 2001) in order to identify their gametic phases. Using a Bayesian framework, this computational method has proven its accuracy in a variety of tests using both simulated and empirical data (Harrigan *et al.* 2008; Stephens & Donnelly 2003; Xu *et al.* 2002), thus avoiding costly and time-consuming PCR-product cloning procedures. Different methods to detect recombination in EF1-α were applied: RDP method (Martin & Rybicki 2000), Bootscanning (Salminen *et al.* 1995), GENECONV (Padidam *et al.* 1999), Maximum Chi-Square (Posada & Crandall 2001a; Smith 1992), Chimaera (Posada & Crandall 2001a) and Sister Scanning (Gibbs *et al.* 2000), which have been implemented in RDP3 (Martin *et al.* 2005).

### 2.2.4. Evolutionary tree construction

Neighbour-joining trees were generated in PAUP* version 4.0b10 (Swofford 2002) using the model of nucleotide substitution that best fits the data, determined with MODELTEST version 3.7(Posada & Crandall 1998). One thousand bootstrap replicates using the NJ search were performed to assess branch support in the resulting tree topology. Maximum-parsimony tree construction was also performed with PAUP*. One hundred replicates of a heuristic search were performed with an initial random stepwise addition of sequences and tree bisection–reconnection branch swapping. Branch support was estimated from 1000 replicates of a bootstrap search. Bayesian analyses were also carried out with the program MRBAYES version 3.1 (Huelsenbeck & Ronquist 2001; Ronquist & Huelsenbeck 2003). The settings were two simultaneous runs (each with two Markov chains) of the Markov chain Monte Carlo (MCMC) for three million generations, sampling every 100 generations, a heating parameter value of 0.20 and a ‘burn-in’ of 25%, using the

general time reversible model (GTR + Γ + I) of sequence evolution with priors set to the default values. Summaries of 65% (19,500 samples) of the sampled parameter values and sampled trees were obtained, as well as a majority-rule consensus tree with posterior probabilities for each bipartition.

### 2.2.5. Haplotype network construction

Although evolutionary gene trees may be informative at the intraspecific level, relationships resulting from intrinsic processes of population dynamics (e.g. persistence of ancestral haplotypes, multifurcations, recombination and horizontal transfer) are better visualized in reticulated graphs or networks (Cassens *et al.* 2005; Huson & Bryant 2006; Posada & Crandall 2001b). For a given taxon these processes are expected to be more acute for nuclear genes that evolve more slowly with a larger effective population size than mtDNA genes. Gene genealogies were inferred using two approaches for haplotype network construction. Median-joining networks (Bandelt *et al.* 1999) were calculated with the program NETWORK version 4.5.1.0 (www.fluxusengineering.com) keeping the parameter $\varepsilon = 0$. This method starts with minimum spanning trees combined within a single network and then, to reduce tree length, median vectors (consensus sequences) are added. Such vectors can be interpreted as possibly extant unsampled sequences or extinct ancestral sequences (Bandelt *et al.* 1999). In addition, TCS version 1.21 (Clement *et al.* 2000) was employed to infer haplotype networks using statistical parsimony (Templeton *et al.* 1992) with a confidence limit of 95%.

### 2.2.6. Genetic structure

To assess the roles of host plant and geography in the structuring of genetic variation within *R. antirrhini* two separate analyses of molecular variance (AMOVA) (Excoffier *et al.* 1992) were performed for both the nuclear and mitochondrial data using ARLEQUIN version 3.1 (Schneider *et al.* 2000). In the first analysis, the haplotypes were grouped by host plant, while for the second analysis haplotypes were grouped into three geographic regions: Western Europe, the Balkan region and Eastern Europe. A pairwise distance matrix was generated and used in the AMOVA with 1023 permutations as a significance test ($\alpha = 0.05$). Also, estimates of $\Phi$ statistics (F-statistic analogs) were calculated overall for all host-associated populations and pairwise between host-associated populations to assess the degree of genetic differentiation among them, testing statistical significance with 1023 permutations ($\alpha = 0.05$).

### 2.2.7. Estimation of divergence times

In the absence of geological and/or fossil calibration points to estimate divergence times we have taken a Bayesian approach using a generalized clock to estimate the age of the most recent common ancestor (mrca), and divergence times within *R. antirrhini*. Mitochondrial rates have been proposed for arthropods in the range of 1.2% to 4.96% pairwise divergence per million years (my) (Brower 1994; Caccone & Sbordoni 2001; Desalle *et al.* 1987; Wares 2001). A comparative rate estimate of the mtDNA COII gene across the Pancrustacea, including 15 hexapod orders, has demonstrated the Coleoptera rate to approximate the mean rate across the Pancrustacea (Cicconardi *et al.* 2010). We therefore apply a mean rate estimate of 3.05% pairwise divergence per million years for our analyses using BEAST version 1.4.8 (Drummond & Rambaut 2007). Unlike nonparametric rate smoothing

(Sanderson 1997), penalised likelihood (Sanderson 2002) or the relaxed Bayesian approach as implemented in Multidivtime (Thorne & Kishino 2002) BEAST does not assume rate autocorrelation which may systematically distort branch lengths, reducing the ratio of deep to shallow nodes (Hugall & Lee 2004; Martin *et al.* 2004). Instead, BEAST accommodates among-branch rate variation by allowing each branch to draw its rate from a discretized lognormal distribution, whose shape is estimated as part of the analysis (Drummond *et al.* 2006). For our analyses we used an uncorrelated lognormal relaxed molecular-clock model in BEAST with the average number of substitutions per site across the tree averaged to be 1.525 per million years, but with rates for individual branches unconstrained. Initial substitution model parameter values were selected according to the results of MODELTEST version 3.7 (Posada & Crandall 1998), with unconstrained prior distributions. A separate demographic model of constant population size was applied, in the form of a coalescent prior, to clades conforming to genetic variation within host associated lineages, and a Yule tree prior was used for the basal branches connecting these. Input files were generated with BEAUTI version 1.4.8 (Rambaut & Drummond 2007). Two runs consisting of 100,000,000 generations each and sampling every 500 generations were performed and combined, checking sampling, mixing and convergence to a stationary distribution.

### 2.2.8. Cross copulation experiments

No choice copulation trials were conducted to assess the potential for reproductive isolation between beetles sampled from two different host plants, *Linaria vulgaris* and *L. genistifolia genistifolia*. A total of 271 weevils (138 females + 133 males) collected from seeds capsules of *L. vulgaris* and 191 weevils (86 females + 105

males) from *L. g. genistifolia* were sampled at the beginning of September in 2007. Males and females were separated and set up in 4 net-cages (30x25x45cm) planted with their original host plants for hibernation. Weevils emerged from hibernation in the following May and two sets of copulation trials were established (Table 9). Pairs of weevils were placed in small plastic vials (40x10mm) and fed twice a day with toadflax flowers (corresponding to the female's original host) for 5-7 days to record matings, after which the male was removed. Females were then transferred into plastic cylinders (40 x 9 cm) containing flowering branches inside isolated field mesh-cages (200x200x240cm) and monitored daily for oviposition on their original host plants. Flowering branches exposed to oviposition were collected after 30 days. All seed capsules were inspected and dissected recording offspring numbers (larvae, pupae and adults). Five replicate pairs of same plant control matings were also conducted.

## 2.3. RESULTS

### 2.3.1. Mitochondrial COII analyses

The final alignment of the COII sequences consisted of 696 bp, with a total of 202 (29%) polymorphic nucleotides of which 195 are parsimony informative. Thirty-seven different haplotypes were identified within *R. antirrhini* and these are available from GenBank under accession numbers HM007201–HM007237. Under the Akaike information criterion MODELTEST analyses revealed the transversion substitution model with invariant sites and rate heterogeneity across sites to best describe the pattern of sequence variation within this fragment, and this model was

employed for the estimation of pairwise genetic distances and a neighbour-joining tree. The maximum ingroup genetic distance was 19% (15% uncorrected) and all three phylogenetic analyses revealed significant genetic structure associated with host plant use (Figure 2). Seven mitochondrial lineages are clearly defined with bootstrap support values higher than 90%, and all but one of these is associated with a single host plant taxon. Lineage four is comprised of weevils collected from 3 different host plant taxa; *Linaria genistifolia genistifolia, L. genistifolia sofiana,* and *L. dalmatica macedonica*. An analysis of molecular variance revealed that 97% of the observed genetic variation in the mtDNA sequence data is explained by host plant use with 1.6% due to variation among populations within host plant groups and 1.4% accounts for variation within populations (Table 2). Estimated pairwise Φst values show significant genetic differentiation between weevils sampled from different host plant taxa (Table 3), with 26 of 28 pairwise comparisons yielding high and significant levels of differentiation. When haplotypes are grouped by major geographic region, these explain 52% of the genetic variation (Table 4).

### 2.3.2. EF1-α analyses

DNA sequencing yielded a total of 39 EF1-α haplotypes from the 83 ingroup weevils, with length variation due to insertions and deletions ranging from 776-796 nucleotides. Thirty-six individuals were heterozygous and haplotypes were inferred manually from forward and reverse sequence chromatograms following similar approaches as those reported by Flot *et al.*(2006) and Peters *et al.* (2007), and with PHASE, employing a 95% significance threshold. Thus 12 individuals were excluded from further analyses due to the inability to assign their haplotype state by

either method. All 40 sequences are available from Genbank under accession numbers HM007238–HM007277.

From the 798 bp alignment (including alignment gaps) of EF1-α haplotypes, there were a total of 48 polymorphic nucleotides (6%) of which 47 are parsimony informative. MODELTEST analyses revealed the general time reversible substitution model with invariant sites and rate heterogeneity across sites to best describe the pattern of sequence variation within EF1-α, and this model was employed for the parameterisation of further analyses. The maximum ingroup genetic distance was 2.42% (uncorrected) and all three phylogenetic analyses described a less evident pattern of host-associated genetic structure (data not shown). To further explore patterns of relatedness among the nuclear sequences, median joining and statistical parsimony networks were constructed. Apart from some minor differences, both approaches describe the same relationships among the 40 different haplotypes. The median joining network contained a single ambiguous connection (Figure 3). No reticulations were found in the parsimony network (Figure 4), but haplotypes corresponding to the outgroup *Rhinusa griseohirta* and *R. florum*, were not connected to the main network at the 95% parsimony connection limit. When the connection limit was reduced to 89% the haplotypes for *Rhinusa griseohirta* and *R. florum* were connected as in the median-joining network.

AMOVA analyses detected significant genetic variation associated with host plant taxon. Approximately 55% of the variation at the EF1-α locus is explained by host-plant use, with variation among populations within a host plant taxon accounting for less than 7%, while 39% of the observed genetic variation is found within

populations (Table 5). Estimated pairwise Φst values reveal significant genetic differentiation between weevils sampled from different host plant taxa (Table 6) for 23 of 28 pairwise comparisons. When haplotypes are grouped by major geographic region, these explain only 26% of the genetic variation, with more variation (41%) explained by differences within populations (Table 7).

### 2.3.3. Estimation of divergence times

For the COII gene tree the coefficient of variation of rates among branches was statistically different from zero (4.95, 95% HPD: 3.96-5.82), indicating variation in rates among branches and departure from a constant clock model. The covariance was not significantly different from zero (0.04, 95% HPD: -0.03-0.14), indicating there is no strong evidence of autocorrelation of rates in the phylogeny. Age estimates were made for various nodes within the tree (Figure 2 and Table 8). Due to low sample sizes we do not present age estimates for the mrca for each of lineages 2 and 7, but in both cases these were not significantly different from zero. The *R. antirrhini* species complex is estimated to have diverged from its sister lineage, *R. griseohirta*, approximately 4.26 mya (million years ago) (95% HPD: 3.26-5.79), with initial diversification within *R. antirrhini* commencing 3.77 mya (95% HPD: 2.91-4.80). With the exception of lineages 2 and 7, the five remaining mtDNA lineages exhibit intra-lineage variation significantly older than 130 kya (thousand years ago), this being the lower 95% HPD value for mtDNA lineage 3 associated with *L. genistifolia linifolia*. Age estimates were produced for 2 additional internal nodes within the tree. The divergence between mtDNA lineages 5 and 6 associated with *L. g. confertiflora* and *L. g. artivensis* in Turkey is estimated to have occurred 1.33 mya (95% HPD: 0.82-1.83). This clade is in turn estimated to have shared a

mrca with mtDNA lineage 4, associated with *L. g. genistifolia, L. d. macedonica,* and *L. g. sofiana* in the Balkans, 1.83 mya (95% HPD: 01.32-2.55).

### 2.3.4. Cross copulation experiments

Results of cross copulation experiments are presented in Table 9. A total of 40 mating pairs were established for female *Rhinusa antirrhini* from *Linaria vulgaris* with male *R. antirrhini* from *L. genistifolia genistifolia.* Forty-five mating pairs were established for female *R. antirrhini* from *L. g. genistifolia* with male *R. antirrhini* from *L. vulgaris*. Five mating pairs each were established for weevils sampled from *L. vulgaris* and *L. g. genistifolia*. Matings were observed for all 10 control crosses, with all pairs producing offspring, and averages of 83 offspring resulting from matings of *L. vulgaris* origin and 53 offspring from matings of *L. g. genistifolia* origin. In contrast significantly reduced reproductive performance was observed for the test-crosses. Copulation was observed for only 35% of mating pairs between females collected from *L. vulgaris* and males from *L. g. genistifolia* ($\chi^2$=7.697, $P$< 0.01, df = 1) and 44% of mating pairs between females from *L. g. genistifolia* and males from *L. vulgaris* ($\chi^2$ = 5.555, $P$< 0.05, df = 1). From these observed matings only 6 females (44% of mating pairs) produced offspring for the former ($\chi^2$= 4.935, $P$< 0.05, df = 1), and only 1 (5% of mating pairs) for the latter ($\chi^2$= 19.79, $P$< 0.01, df = 1) (Table 9).

## 2.4. DISCUSSION

### 2.4.1. Genetic diversity and structure within *Rhinusa antirrhini*

Analyses of both mitochondrial and nuclear DNA sequences reveal high levels of genetic variation and host-associated genetic structure among sampled populations of *R. antirrhini* on different host plants. An average sequence divergence of 9.2% for the mtDNA COII gene exceeds values that have been observed within other weevil species complexes (e.g. Barat *et al.* 2008; Erney *et al.* 1996; Langor & Sperling 1997; Normark 1996). The comparatively lower genetic variation observed within the EF1-α sequences (average sequence divergence 0.9%) is consistent with the low levels of divergence expected for intraspecific nuclear data (Zhang & Hewitt 2003). However, it is at the same time much higher than the 0.2 – 0.3 % divergence in EF1- α sequences reported for gorse weevils using different hosts (Barat *et al.* 2008).

Phylogenetic analyses of mtDNA COII sequences reveal clear genetic structure with six mitochondrial lineages within *R. antirrhini* associated with different taxa of *Linaria*. There is no pronounced phylogenetic signal within the EF1-α sequence data, but results from AMOVA analysis reveal structuring of genetic variation among host plant taxa. The distinct geographic distributions of the different *Linaria* taxa mean that conclusions of host plant associated genetic differentiation could be confounded by geographic effects, but three lines of evidence argue against this. First, AMOVA analyses of both gene regions reveal that host plant, not geography, offers greater explanation for the structuring of genetic variation. Second, within the most geographically widespread taxa, *L. vulgaris*, there is little evidence for

geographic structuring of genetic variation among individuals of *R. antirrhini*. A single individual from the more eastern Russian location does suggest the possibility of a phylogeographic east-west division. However, in comparison to the genetic differentiation observed between *R. antirrhini* sampled from different *Linaria* taxa, samples from *L. vulgaris* from the Balkans through to Western Europe show remarkably little differentiation among mtDNA COII haplotypes. Both these lines of evidence offer indirect support for a host plant effect over a geographic effect. The third and more direct line of evidence for host plant effect comes from weevils collected on different plants growing sympatrically. *Rhinusa antirrhini* were sampled from both *L. vulgaris* (3 individuals) and *L. genistifolia sofiana* (3 individuals) growing sympatrically on Mt Rila in Bulgaria (Figure 1). Similarly *R. antirrhini* were sampled from both *L. vulgaris* (4 individuals) and *L. rubioides* (3 individuals) growing sympatrically in Mokra Gora in Serbia (Figure 1). In both cases the correspondence between host plant taxa and mtDNA haplogroup is maintained (Figure 2). Although not sampled sympatrically, both *L. genistifolia artvinensis* and *L. genistifolia linifolia* were each sampled from two locations in Turkey approximately 33 km apart. Correspondence between host plant taxa and mtDNA haplogroup is observed, despite closer geographic proximity between different plant taxon sampling sites than between same plant taxon sampling sites (Figure 1).

Taken together our data argue for ecological divergence, with different resource use being the driving agent for genetic differentiation within *Rhinusa antirrhini*. All but one mtDNA lineage is associated with a single taxon of *Linaria*, with lineage 4 being the exception. Lineage 4 includes individuals sampled within the Balkan

region from *Linaria genistifolia genistifolia*, *L. genistifolia sofiana* and *L. dalmatica macedonica*, suggesting generalist oviposition behaviour, although two other plausible explanations cannot be rejected. It may be that there is ecological divergence within this mtDNA lineage, but of more recent origin, beyond the resolving power of mitochondrial DNA. There is also some doubt surrounding the taxonomic distinction of the three host taxa of lineage 4 (Chater *et al.* 1972), and it may be that the genetic similarity of weevils across these three *Linaria* taxa reflects this.

### 2.4.2. Incomplete lineage sorting or gene flow?

The less evident pattern of host-associated genetic structure revealed by phylogenetic analyses of the EF1-α sequences maybe a consequence of incomplete lineage sorting due to the higher effective population size and lower mutation rate of this nuclear sequence, compared to those of mitochondrial sequences (Desalle *et al.* 1987; Monteiro & Pierce 2001; Moriyama & Powell 1997). Alternatively it could be that gene flow between weevils adapted to different *Linaria* taxa has contributed to the less pronounced pattern of genetic differentiation among *Linaria* taxa compared to mitochondrial DNA. This must be considered plausible as it has been suggested that host mediated selection can maintain the genetic distinction of host races even in the face of moderate gene flow (Feder *et al.* 1997; Filchak *et al.* 1999). Both incomplete lineage sorting and gene flow may well have contributed to the evolutionary history of *Rhinusa antirrhini*, and a consideration of the phylogenetic relationships of the allelic variation at the EF1-α locus does not favour one above the other. Considering that in a phylogenetic network interior haplotypes are older than those found at the tips (Crandall & Templeton 1993; Posada & Crandall 2001b),

shared haplotypes H17 and H21 would argue for recent gene flow, whereas shared haplotypes H2, H11 and H14 suggest retained ancestral polymorphism (Figure 3).

Results from the cross copulation experiments indicate that reproductive barriers between weevils with different host affiliation have evolved. Compared to control crosses, significantly fewer observed copulations and offspring were produced between weevils from different host plants, indicating that both pre-mating and post-mating reproductive barriers are operating. In addition to this, previous experimental observations (Toševski, unpublished data) reveal *Rhinusa antirrhini* to exhibit both feeding and oviposition preference for the host plant from which they are collected, thus mating among individuals with the same host preference should be more likely than mating among individuals with different host affiliation (Craig *et al.* 1993; Feder *et al.* 1994; Via 1999). However, our cross copulation experiments also demonstrate that reproductive isolation between host associated mtDNA lineages is not complete. We recognise that the "no choice" nature of our experiments represents an extreme situation, and that they are limited to identifying the generation of, but not the fitness of, F1 progeny. But given that the *R. antirrhini* complex as a whole has an estimated evolutionary history of 3.77 myr (Table 8), in the light of our results several considerations suggest gene flow to be plausible within the history of this complex. First, our cross copulation experiments involved two of the more divergent host associated mtDNA lineages, representing the earliest divergence event associated with the root age of 3.77 myr. This provides a substantial amount of time for introgression, particularly so during early divergence. Second, most species diverged from each other more recently than the test-cross pair, meaning genetic incompatibilities for hybridisation are likely to be less. Third,

while contemporary gene flow would be limited to contemporary sympatry, the dramatic climatic changes within Europe over the estimated 3.77 myr history of the group may have facilitated past sympatry of other *Linaria* taxa, and gene flow among associated *R. antirrhini* lineages, that are presently allopatric.

One interesting consideration is that if gene flow between host-adapted weevils has featured within the evolutionary history of this group, it has not resulted in the disruption of the relationship between mtDNA lineages and the different *Linaria* taxa. This in itself may be seen as evidence against a history involving gene flow, but if host choice for oviposition were sex-linked, as shown for some Lepidoptera (Janz 1998; Scriber *et al.* 1991), or perhaps sex-influenced, as observed in some tephritid fly host races (Craig *et al.* 2001), such a pattern could be maintained in the face of gene flow between host forms.

### 2.4.3. Divergence times

Our divergence time estimates suggest that *Rhinusa antirrhini* diverged from *R. griseohirta* approximately 4.26 mya (95% HPD: 3.26-5.79), and began to diversify approximately 3.77 mya (95% HPD: 2.91-4.80) in the middle Pliocene, with the most recent divergence event between host-associated lineages estimated to have occurred approximately 1.33 mya (95% HPD: 0.82-1.83). Several mtDNA lineages were sampled densely enough to permit the estimation of the age of the coalescence to the mrca (Table 8), with the youngest of these estimated to be approximately 810 kya (thousand years ago, HPD: 0.13-1.85 mya) indicating intra-lineage diversification dating back to the mid-Pleistocene. Extreme morphological similarity of the six host associated lineages of *R. antirrhini* has been maintained over the 3.77

myr period of adaptation to different *Linaria* taxa. However, the phylogenetic placement of the morphologically distinct *R. florum* (both rostrum and genitalia) within the *R. antirhini* complex does indicate some morphological change within the group over the last 3.77 mya. It is interesting that *R. florum* is one of only two taxa sharing a common host plant (*L. genistifolia genistifolia* is also used by lineage 4), suggesting a possible history of character displacement, as has been noted in other closely related sympatric plant feeding insects (Jordal *et al.* 2006).

## 2.5. CONCLUSION

Within plant-feeding insects, host plant specialization has long been suggested to have facilitated insect diversification (Berlocher & Feder 2002; Bush 1969; Ehrlich & Raven 1964). Endophagous insects in particular appear to be more susceptible to selective pressures imposed by their host plants due to their more intimate interaction with the host (Funk *et al.* 2002; Mopper 1996). Host plant selective pressure is likely to be further enhanced when endophagy is accompanied by insect-elicited plant physiological responses for successful larval development. Despite this recognition, studies of host-mediated selection in endophagous insects are scarce. Here we demonstrate host-associated genetic differentiation within a parasitic weevil whose life history is tightly linked to that of its host plant due to its ecological specialization of feeding, ovipositing, and developing within the fruit capsules of particular *Linaria* taxa. Our results provide strong evidence that ecological divergence by specializing on different resources is driving genetic differentiation within *R.antirrhini*. Given the morphologically cryptic nature of host associated

lineages within *R. antirrhini*, we advocate caution when interpreting shared morphology of insect herbivores across different plant taxa as evidence of an ecologically generalist life history. In addition to the important task of quantifying biodiversity, this point is of particular relevance when focal taxa are potential pests, or biocontrol agents as in the case for *R. antirrhini.*

**Acknowledgements**

This research was funded by Consejo Nacional de Ciencia y Tecnologia (CONACYT) Mexico, and in part by Grant 143006B (The Ministry of Science and Technological Development of the Republic of Serbia) within the framework of a biological control program against exotic toadflaxes in North America. We thank the Wyoming Biological Control Steering Committee, the Ministry of Forests and Range, British Columbia Provincial Government, USDA-APHIS-CPHST, USDA Forest Service through the Montana State University, and the California Department of Food and Agriculture who supported this programme in 2008 and 2009. We gratefully acknowledge the support of Dr De Clerck-Floate (AAFC, Lethbridge, Canada) and Dr Andrew Norton (Colorado State University, USA), the coordinators of the toadflax consortium in North America.

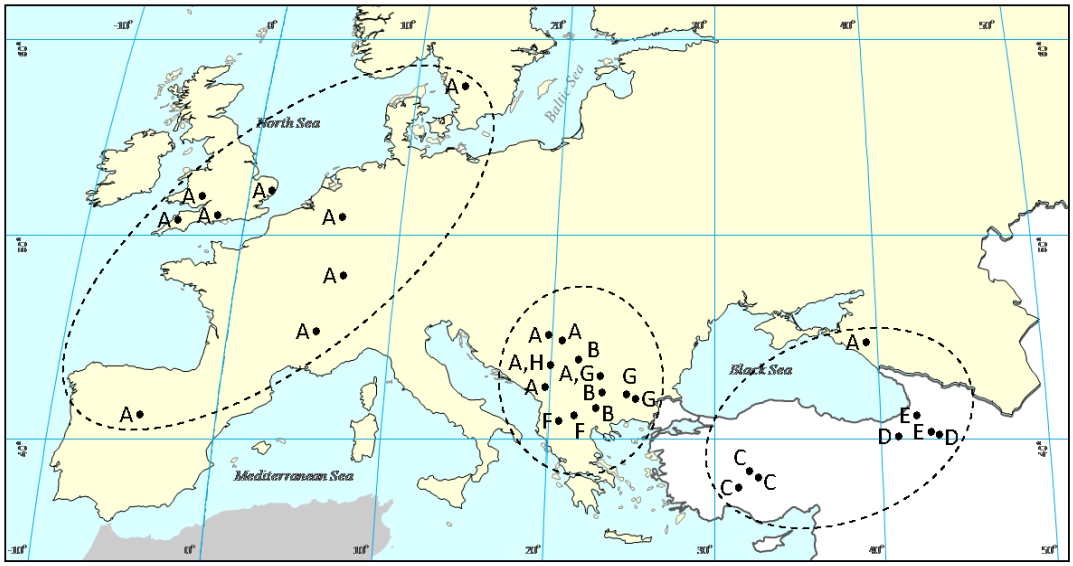


**Figure 1.** Sampling sites of weevils and associated host plants: (A) *Linaria vulgaris*, (B) *L. genistifolia genistifolia*, (C) *L. genistifolia confertiflora*, (D) *L. genistifolia linifolia*, (E) *L. genistifolia artvinensis*, (F) *L. dalmatica macedonica*, (G) *L. genistifolia sofiana*, (H) *L. rubioides*. Ovals indicate geographic regions (see text).


Gy091_Lv_Bonn_Ger
707_Lv_MkG_Serb
Gy986_Lv_Lez_Serb
Gy988_Lv_Vrc_Serb
Gy987_Lv_Lez_Serb
Gy657_Lv_Bonn_Germ
Gy090_Lv_Bonn_Ger
Gy647_Lv_LeP_France
Gy089_Lv_ND_France
Gy088_Lv_ND_France
Gy1182_Lv_Mor_Swed
771_Lv_MkG_Serb
Gy890_Lv_Gra_Mont
Gy642_Lv_Bas_Switz
722_Lv_MtRl_Bulg
723_Lv_MtRl_Bulg
Gy653_Lv_Gro_Mont
Gy652_Lv_Gro_Mont
705_Lv_MkG_Serb
Rav7_Lv_MkG_Serb
Gy643_Lv_Bas_Switz
Gy887_Lv_Bas_Switz
310_Lv_MtRl_Bulg
Gy1045_Lv_StH_UK
Gy1185_Lv_StH_UK
Gy915_Lv_Avell_Spain
Gy648_Lv_Avell_Spain
Gy1047_Lv_Tin_UK
Gy656_Lv_Bonn_Germ
Gy1180_Lv_Mor_Swed
Gy885_Lv_StH_UK
Gy1188_Lv_Tin_UK
Gy989_Lv_Tin_UK
Gy1179_Lv_Mor_Swed
Gy1041_Lv_Tin_UK
Gy1181_Lv_Mor_Swed
Gy886_Lv_Sto_UK
Gy1048_Lv_Sto_UK
Gy1189_Lv_Sto_UK
Gy1044_Lv_StH_UK
Gy1049_Lv_Sto_UK
Gy888_Lv_Sw_UK
Gy1043_Hot_Lv_Sto_UK
Gy1046_Lv_Tin_UK
Gy1183_Het_Lv_Mor_Swed
Gy1178_Lv_Kras_Russ
Gy916_Lr_MkG_Serb
Ran4_Lr_MkG_Serb
Gy917_Lr_MkG_Serb
Gy944f_Lg_Ale_Serb
Gy960f_Lg_Ale_Serb
Gy881_Lg_Ale_Serb
Gy985_Lg_Fur_Mace
309_Lg_sofi_MtRl_Bulg
Gy1016cf_Ldmac
Gy1166_Lg_sofi_Plo_Bulg
Gy093_Ldm_Pril_Mace
Gy908cf_Ldmac_Ohrid_Mace
Gy910cf_Ldmac_Ohrid_Mace
Gy1024_Lg_MtRl_Bulg
Gy907cf_Ldmac_Ohrid_Mace
716_Lg_sofi_MtRl_Bulg
Gy1165_Lg_sofi_Plo_Bulg
226_l_Lg_sofi_Har_Bulg
Gy909cf_Ldmac_Ohrid_Mace
Gy1150_Lg_cnfl_Kon_Turk
Gy1155_Lg_cnfl_Kon_Turk
Gy1153_Lg_cnfl_Bey_Turk
Gy1154_Lg_cnfl_Kon_Turk
228_l_Lg_cnfl_Bey_Turk
Gy1158_Lg_cnfl_Kon_Turk
Gy1156_Lg_cnfl_Kon_Turk
Gy1157_Lg_cnfl_Kon_Turk
Gy1159_Lg_cnfl_Kon_Turk
Gy1151_Het_Lg_cnfl_Kon_Turk
Gy1152_Lg_cnfl_Bey_Turk
Gy1161_Lg_artv_Art_Turk
Gy1162_Lg_artv_Art_Turk
Gy1170_Lg_artv_Bal_Turk
Gy1169_Lg_artv_Bal_Turk
Gy1175_Lg_artv_Bal_Turk
Gy1172_Lg_artv_Bal_Turk
Gy1173_Lg_artv_Bal_Turk
304_l_Lg_artv_Art_Turk
Gy1167_Lg_artv_Bal_Turk
Gy1176_Lg_artv_Bal_Turk
301_l_Lg_artv_Bal_Turk
Gy1171_Lg_artv_Bal_Turk
Gy1174_Lg_artv_Bal_Turk
Gy1168_Lg_artv_Bal_Turk
Gy1160_Lg_cf_Iin_Kar_Turk
314_l_Lg_cf_Iin_Kar_Turk
Gy1164_Lg_Iinl_Clm_Turk
312_l_Lg_Iinl_Clm_Turk
305_l_Lg_Iinl_Clm_Turk
Gy831_Ag_Sev_Spain
Gy_832_Ag_Sev_Spain
I
II
III
IV
V
VI
VII
VIII
IX
1
2
3
4
5
6
7
1
100 / 100
1
100 / 100
1
100 / 100
0.8
-- / --
--
-- / 75
1
91 / 99
1
100 / 100
1
96 / 95
1
100 / 100
1
100 / 100

**Figure 2.** Bayesian phylogenetic tree inferred from the 696 bp COII fragment of *Rhinusa antirrhini* on different host plants. Bayesian posterior probabilities ≥ 0.8 are shown above branches, maximum parsimony and neighbour joining bootstrap support values are indicated below branches in that order; values lower than 70% are omitted. Roman numerals refer to estimated divergence times given in Table 8. Arabic numerals refer to host plant-associated lineages. 1 = *Linaria vulgaris*, 2 = *L. rubioides*, 3 = *L. genistifolia linifolia*, 4 = *L. genistifolia genistifolia* / *L. genistifolia sofiana* / *L. dalmatica macedonica*, 5 = *L. genistifolia confertiflora*, 6 = *L. genistifolia artvinensis*, 7 = *Rhinusa florum* on *L. genistifolia genistifolia*.

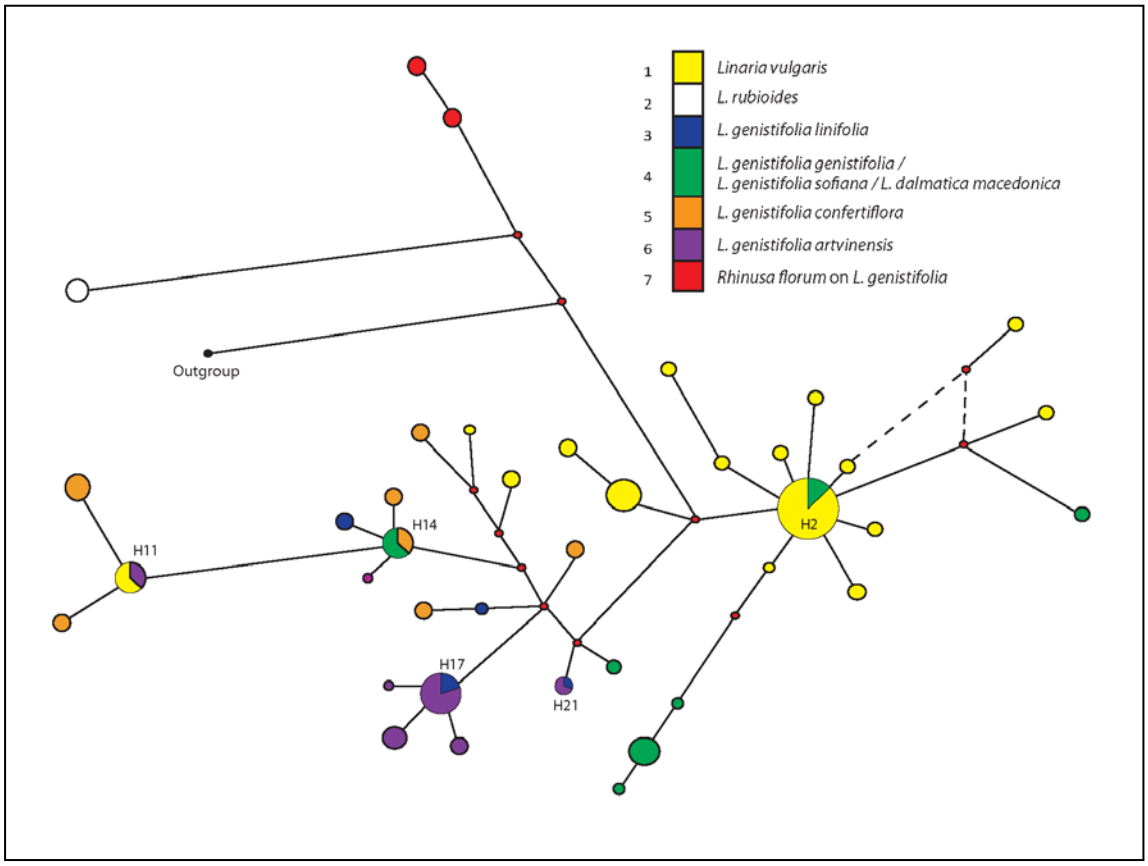


**Figure 3.** Median joining network obtained from EF1-α sequences of *R. antirrhini*. Circle sizes are proportional to haplotype frequency. The numbers and colours correspond to the mitochondrial lineages associated with different host plants as indicated in Figure 2. EF1-α haplotypes H2, H21, H17, H14 and H11 are associated with more than one mitochondrial lineage. Ambiguous connections are indicated with dotted lines, and red dot vertices are median vectors representing missing haplotypes.

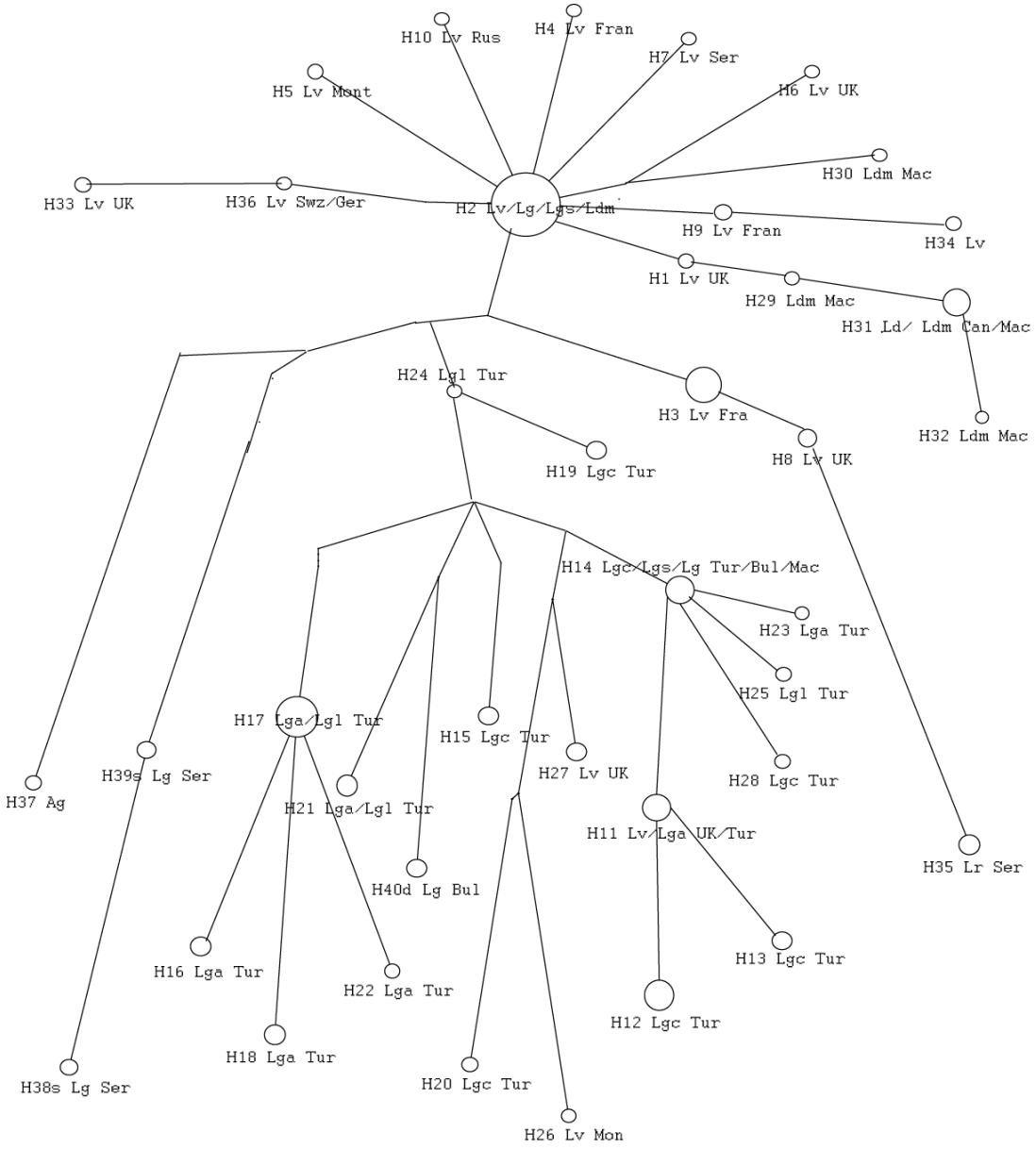


**Figure 4.** Statistical parsimony network obtained from EF1-α DNA sequences of *Rhinusa antirrhini*. Circle sizes are proportional to haplotype frequency. Haplotypes are numbered H1 – H40.

**Table 1.** List of specimens used with host plant affiliations and origin. All samples are *Rhinusa antirrhini* except where indicated. Accessions marked with an asterisk were excluded from EF1-α data analyses because of unresolved gametic phase (< 1 probability) of heterozygote sequences.

| Accession Code | Host Plant | Location / Country |
|---|---|---|
| Gy831 (*R. griseohirta*) | *Antirrhinum graniticum* | Sevilla, Spain |
| Gy832 *(R. griseohirta)* | *Antirrhinum graniticum* | Sevilla, Spain |
| Gy093 | *L. dalmatica macedonica* | Prilep, Macedonia |
| Gy907 | *L. dalmatica macedonica* | Ohrid, Macedonia |
| Gy908 | *L. dalmatica macedonica* | Ohrid, Macedonia |
| Gy909 | *L. dalmatica macedonica* | Ohrid, Macedonia |
| Gy1016 | *L. dalmatica macedonica* | Prilep, Macedonia |
| Gy910 | *L. dalmatica macedonica* | Ohrid, Macedonia |
| Gy985 | *L. genistifolia genistifiolia* | Furka, Macedonia |
| Gy944 *(R. florum)* | *L. genistifolia genistifiolia* | Aleksinac, Serbia |
| Gy960 *(R. florum)* | *L. genistifolia genistifiolia* | Aleksinac, Serbia |
| Gy881 * | *L. genistifolia genistifiolia* | Aleksinac, Serbia |
| Gy1161 | *L. genistifolia artvinensis* | Artvin, Turkey |
| Gy1162 | *L. genistifolia artvinensis* | Artvin, Turkey |
| 304-I | *L. genistifolia artvinensis* | Artvin, Turkey |
| Gy1167 * | *L. genistifolia artvinensis* | Balcesme, Turkey |
| Gy1168 | *L. genistifolia artvinensis* | Balcesme, Turkey |
| Gy1169 | *L. genistifolia artvinensis* | Balcesme, Turkey |
| Gy1170 | *L. genistifolia artvinensis* | Balcesme, Turkey |
| Gy1171 | *L. genistifolia artvinensis* | Balcesme, Turkey |
| Gy1172 * | *L. genistifolia artvinensis* | Balcesme, Turkey |
| Gy1173 | *L. genistifolia artvinensis* | Balcesme, Turkey |
| Gy1174 | *L. genistifolia artvinensis* | Balcesme, Turkey |
| Gy1175 | *L. genistifolia artvinensis* | Balcesme, Turkey |
| Gy1176 | *L. genistifolia artvinensis* | Balcesme, Turkey |
| 301-I | *L. genistifolia artvinensis* | Balcesme, Turkey |

| | | |
|---|---|---|
| Gy1150 | *L. genistifolia confertiflora* | Konya, Turkey |
| Gy1151 | *L. genistifolia confertiflora* | Konya, Turkey |
| Gy1152 * | *L. genistifolia confertiflora* | Beysehir, Turkey |
| Gy1153 | *L. genistifolia confertiflora* | Beysehir, Turkey |
| 228-I | *L. genistifolia confertiflora* | Beysehir, Turkey |
| Gy1154 | *L. genistifolia confertiflora* | Cilanayayla, Turkey |
| Gy1155 | *L. genistifolia confertiflora* | Cilanayayla, Turkey |
| Gy1156 | *L. genistifolia confertiflora* | Cilanayayla, Turkey |
| Gy1157 | *L. genistifolia confertiflora* | Cilanayayla, Turkey |
| Gy1158 | *L. genistifolia confertiflora* | Cilanayayla, Turkey |
| Gy1159 | *L. genistifolia confertiflora* | Cilanayayla, Turkey |
| Gy1160 * | *L. genistifolia linifolia* | Kars, Turkey |
| 314-I | *L. genistifolia linifolia* | Kars, Turkey |
| Gy1164 * | *L. genistifolia linifolia* | Cimil, Turkey |
| 305-I | *L. genistifolia linifolia* | Cimil, Turkey |
| 312-I | *L. genistifolia linifolia* | Cimil, Turkey |
| Gy1165 | *L. genistifolia sofiana* | Plovdiv, Bulgaria |
| Gy1166 | *L. genistifolia sofiana* | Plovdiv, Bulgaria |
| 226-I * | *L. genistifolia sofiana* | Harmanli, Bulgaria |
| 309-I | *L. genistifolia sofiana* | Mt. Rila, Bulgaria |
| 716-I | *L. genistifolia sofiana* | Mt. Rila, Bulgaria |
| Gy1024 | *L. genistifolia sofiana* | Mt. Rila, Bulgaria |
| Rar4 | *L. rubioides* | Mokra Gora, Serbia |
| Gy917 | *L. rubioides* | Mokra Gora, Serbia |
| Gy916 | *L. rubioides* | Mokra Gora, Serbia |
| Gy1043 | *L. vulgaris* | Stonehenge, UK |
| Gy1041 | *L. vulgaris* | Tintagel, UK |
| Gy647 | *L. vulgaris* | Le Pra, France |
| Gy089 | *L. vulgaris* | Notre Dame, France |
| Gy986 | *L. vulgaris* | Lezimir, Serbia |
| Gy652 | *L. vulgaris* | Gracanica, Montenegro |
| Gy885 | *L. vulgaris* | Sutton Hoo, UK |
| Gy887 | *L. vulgaris* | Basel, Switzerland |

| | | |
|---|---|---|
| Gy888 | *L. vulgaris* | Swansea, UK |
| Gy890 | *L. vulgaris* | Gracanica, Montenegro |
| Gy987 | *L. vulgaris* | Lezimir, Serbia |
| Gy988 * | *L. vulgaris* | Vrcin, Serbia |
| Gy090 | *L. vulgaris* | Bonn, Germany |
| Gy989 | *L. vulgaris* | Tintagel, UK |
| Gy1045 | *L. vulgaris* | Sutton Hoo, UK |
| Gy1046 | *L. vulgaris* | Tintagel, UK |
| Gy657 | *L. vulgaris* | Bonn, Germany |
| Gy1048 | *L. vulgaris* | Stonehenge, UK |
| Gy1049 | *L. vulgaris* | Stonehenge, UK |
| Gy088 | *L. vulgaris* | Notre Dame, France |
| Gy1044 | *L. vulgaris* | Sutton Hoo, UK |
| Gy915 * | *L. vulgaris* | Avellaneda, Spain |
| Gy653 | *L. vulgaris* | Gracanica, Montenegro |
| Gy886 | *L. vulgaris* | Stonehenge, UK |
| Gy648 * | *L. vulgaris* | Avellaneda, Spain |
| Gy1047 | *L. vulgaris* | Tintagel, UK |
| Gy091 | *L. vulgaris* | Bonn, Germany |
| Gy642 | *L. vulgaris* | Basel, Switzerland |
| Gy643 | *L. vulgaris* | Basel, Switzerland |
| Gy656 | *L. vulgaris* | Bonn, Germany |
| Gy1178 | *L. vulgaris* | Krasnodar, Russia |
| Gy1179 | *L. vulgaris* | Morlanda, Sweden |
| Gy1180 | *L. vulgaris* | Morlanda, Sweden |
| Gy1181 | *L. vulgaris* | Morlanda, Sweden |
| Gy1182 | *L. vulgaris* | Morlanda, Sweden |
| Gy1185 | *L. vulgaris* | Sutton Hoo, UK |
| Gy1188 * | *L. vulgaris* | Tintagel, UK |
| Gy1189 * | *L. vulgaris* | Stonehenge, UK |
| Gy1183 | *L. vulgaris* | Morlanda, Sweden |
| 310-I | *L. vulgaris* | Mt. Rila, Bulgaria |
| 722-I | *L. vulgaris* | Mt. Rila, Bulgaria |

| | | |
|---|---|---|
| 723-I | *L. vulgaris* | Mt. Rila, Bulgaria |
| Rav7 | *L. vulgaris* | Mokra Gora, Serbia |
| 707-I | *L. vulgaris* | Mokra Gora, Serbia |
| 705-I | *L. vulgaris* | Mokra Gora, Serbia |
| 771-I | *L. vulgaris* | Mokra Gora, Serbia |

**Table 2.** Results of the COII analysis of molecular variance (AMOVA) for *Rhinusa antirrhini* populations grouped by host plant taxa.

| Source of variation | d. f. | Sum of squares | Variance components | Percentage of variation | Φ statistics |
|---|---|---|---|---|---|
| Among groups (host plants) | 7 | 2758.534 | 41.65632 | 96.92 | 0.970 |
| Among populations within groups | 20 | 54.975 | 0.71389 | 1.66 | 0.539 |
| Within populations | 65 | 39.577 | 0.60888 | 1.42 | 0.985 |

**Table 3**. Pairwise Φst values between host-associated populations of *Rhinusa antirrhini* estimated from the COII sequences. All values are significant at the 0.05 level except where indicated.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | 0 | | | | | | | |
| 2 | 0.9709 | 0 | | | | | | |
| 3 | 0.9695 | 0.9586 | 0 | | | | | |
| 4 | 0.9581 | 0.9838 | 0.9857 | 0 | | | | |
| 5 | 0.9649 | 0.9922 | 0.9948 | 0.9803 | 0 | | | |
| 6 | 0.9683 | 0.9912 | 0.9944 | 0.9861 | 1.0000 | 0 | | |
| 7 | 0.9647 | 0.9866 | 0.9896 | 0.9695 | 0.9931* | 0.5384 | 0 | |
| 8 | 0.9676 | 0.9906 | 0.9940 | 0.9844 | 1.0000 | 0.0000* | 0.4736 | 0 |

1: *Linaria vulgaris,* 2: *L. genistifolia artvinensis,* 3: *L. genistifolia confertiflora,* 4: *L. genistifolia linifolia,* 5: *L. rubioides,* 6: *L. dalmatica macedonica,* 7: *L. genistifolia genistifolia,* 8: *L. genistifolia sofiana.* *Not significant.

**Table 4**. Results of the COII analysis of molecular variance (AMOVA) for *Rhinusa antirrhini* populations grouped by geographic regions.

| Source of variation | d. f. | Sum of squares | Variance components | Percentage of variation | Φ statistics |
|---|---|---|---|---|---|
| Among groups (geographic regions) | 2 | 1458.855 | 20.05482 | 52.03 | 0.520 |
| Among populations within groups | 22 | 1066.621 | 17.31126 | 46.39 | 0.967 |
| Within populations | 55 | 32.994 | 0.59989 | 1.58 | 0.984 |

**Table 5**. Results of the EF1-α analysis of molecular variance (AMOVA) for *Rhinusa antirrhini* populations grouped by host plant taxa.

| Source of variation | d. f. | Sum of squares | Variance components | Percentage of variation | Φ statistics |
|---|---|---|---|---|---|
| Among groups (host plant) | 7 | 398.216 | 3.47427 | 54.6 | 0.545 |
| Among populations within groups | 15 | 72.865 | 0.41199 | 6.47 | 0.142 |
| Within populations | 123 | 304.994 | 2.47963 | 38.95 | 0.610 |

**Table 6**. Pairwise Φst values between host-associated populations of *Rhinusa antirrhini* estimated from the EF1-α sequences. All values are significant at the 0.05 level except where indicated.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1 | 0 | | | | | | | |
| 2 | 0.6318 | 0.0000 | | | | | | |
| 3 | 0.5765 | 0.4330 | 0.0000 | | | | | |
| 4 | 0.5489 | 0.1699* | 0.2482 | 0 | | | | |
| 5 | 0.4030 | 0.4227 | 0.1557* | 0.0809* | 0 | | | |
| 6 | 0.2921 | 0.3997 | 0.2259 | 0.0809* | -0.0909* | 0 | | |
| 7 | 0.7768 | 0.8432 | 0.7352 | 0.8426 | 0.8988 | 0.8915 | 0 | |
| 8 | 0.4402 | 0.6871 | 0.5779 | 0.6144 | 0.5747 | 0.5441 | 0.8233 | 0 |

1: *Linaria vulgaris*, 2: *L. genistifolia artvinensis*, 3: *L. genistifolia confertiflora*, 4: *L. genistifolia linifolia*, 5: *L. genistifolia sofiana*, 6: *L. genistifolia genistifolia*, 7: *L. rubioides*, 8: *L. dalmatica macedonica*. *Not significant.

**Table 7.** Results of the EF1-α analysis of molecular variance (AMOVA) for *Rhinusa antirrhini* populations grouped by geographic region.

| Source of variation | d. f. | Sum of squares | Variance components | Percentage of variation | Φ statistics |
|---|---|---|---|---|---|
| Among groups (geographic regions) | 2 | 200.566 | 1.57873 | 26.09 | 0.261 |
| Among populations within groups | 19 | 266.940 | 1.98314 | 32.78 | 0.443 |
| Within populations | 124 | 308.569 | 2.48846 | 41.13 | 0.588 |

**Table 8.** Estimated times to the most recent common ancestor (mrca) for host-associated mitochondrial lineages expressed as mean values with 95% highest posterior density intervals.

| Node number | | Mean value (my) | 95% HPD |
|---|---|---|---|
| I | root of the tree | 4.26 | 3.26-5.79 |
| II | mrca of ingroup | 3.77 | 2.91-4.80 |
| III | mrca of lineage 1 (*L. vulgaris*) | 1.73 | 1.00-2.48 |
| IV | mrca of lineage 3 (*L. genistifolia linifolia*) | 0.81 | 0.13-1.85 |
| V | mrca of lineage 4 (*L.dalmatica macedonica / L. genistifolia genistifolia / L. g. sofiana*) | 0.816 | 0.38-1.27 |
| VI | mrca of lineage 5 (*L. g. confertiflora*) | 0.93 | 0.44-1.45 |
| VII | mrca of lineage 6 (*L. g. artvinensis*) | 0.89 | 0.47-1.29 |
| VIII | mrca of lineages 5 and 6 | 1.33 | 0.82-1.83 |
| IX | mrca of lineages 4, 5 and 6 | 1.83 | 1.32-2.55 |

**Table 9.** Results of cross copulation experiments among *Rhinusa antirrhini* individuals sampled from *Linaria vulgaris* (V) and *Linaria genistifolia genistifolia* (G).

| Cross copulation experiment type | Number of replicate pairs | Number of observed copulation pairs | Number of pairs producing offspring | Mean number of offspring per offspring producing female (±SD) |
|---|---|---|---|---|
| [1]V-V control) | 5 | 5 (100%[2]) | 5 (100%[3]) | **82.6±27.6** |
| G-G (control) | 5 | 5 (100%) | 5 (100%) | **52.4±14.4** |
| V-G | 40 | 14 (35%) | 6 (43%) | **30.4±15.8** |
| G-V | 45 | 20 (44%) | 1 (5%) | **23.0** |

[1]First letter represents origin of female, second, origin of male.

[2]Expressed as the percentage of replicate pairs

[3]Expressed as the percentage of pairs observed copulating

REFERENCES

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37-48.

Barat M, Tarayre M, Atlan A (2008) Genetic divergence and ecological specialisation of seed weevils (Exapion spp.) on gorses (Ulex spp.). *Ecological Entomology* **33**, 328-336.

Barron AB (2001) The life and death of Hopkins' host-selection principle. *Journal of Insect Behavior* **14**, 725-737.

Berlocher SH, Feder JL (2002) Sympatric speciation in phytophagous insects: Moving beyond controversy? *Annual Review of Entomology* **47**, 773-815.

Brower AVZ (1994) Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial-dna evolution.

*Proceedings of the National Academy of Sciences of the United States of America* **91**, 6491-6495.

Bush GL (1969) Sympatric host race formation and speciation in frugivorous flies of genus Rhagoletis (Diptera, Tephritidae). *Evolution* **23**, 237-251.

Caccone A, Sbordoni V (2001) Molecular biogeography of cave life: A study using mitochondrial DNA from Bathysciine beetles. *Evolution* **55**, 122-130.

Caldara R (2001) Phylogenetic analysis and higher classification of the tribe Mecinini (Coleoptera: Curculionidae, Curculioninae). *Koleopterologische Rundschau* **71**, 171-203.

Caldara R (2004) Fauna Europea: Curculionidae. In: Alonso-Zarazaga, MA (ed.) (2004) Fauna Europea: Coleoptera 1. Fauna Europaea version 2.0, http://www.faunaeur.org/ [update 10.12.2009].

Caldara R (2008) On the taxonomy and nomenclature of some mecinini (Coleoptera, Curculionidae). *Fragmenta Entomologica* **40**, 125-137.

Cassens I, Mardulyn P, Milinkovitch M (2005) Evaluating Intraspecific "Network" Construction Methods Using Simulated Sequence Data: Do Existing Algorithms Outperform the Global Maximum Parsimony Approach? *Systematic Biology* **54**, 363-372.

Chater AO, Valdes B, Webb DA (1972) Linaria Miller. In: *Flora Europeae, Vol. 3* (eds. Tutin TG, Heywood VH, Burges NA*, et al.*), pp. 226-236. Cambridge University Press, Cambridge, MA.

Cicconardi F, Nardi F, Emerson BC, Frati F, Fanciulli PP (2010) Deep phylogeographic divisions and long-term persistence of forest invertebrates in the North-Western Mediterranean basin. *Molecular Ecology* **19**, 386-400.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**, 1657-1659.

Corbet SA (1985) Insect Chemosensory Responses - A chemical legacy hypothesis. *Ecological Entomology* **10**, 143-153.

Craig TP, Horner JD, Itami JK (2001) Genetics, experience, and host-plant preference in Eurosta solidaginis: Implications for host shifts and speciation. *Evolution* **55**, 773-782.

Craig TP, Itami JK, Abrahamson WG, Horner JD (1993) Behavioral evidence for host-race formation in Eurosta solidaginis. *Evolution* **47**, 1696-1710.

Crandall KA, Templeton AR (1993) Empirical Tests of Some Predictions From Coalescent Theory With Applications to Intraspecific Phylogeny Reconstruction. *Genetics* **134**, 959-969.

Davis JM, Stamps JA (2004) The effect of natal experience on habitat preferences. *Trends in Ecology & Evolution* **19**, 411-416.

De Clerck-Floate RA, Harris P (2002) Linaria dalmatica (L.) Miller, Dalmatian toadflax (Scrophulariaceae). In: *Biological Control Programmes in Canada, 1981–2000* (eds. Mason PG, Huber JT), pp. 368-374. CABI Publishing, Wallingford, UK.

Desalle R, Freedman T, Prager EM, Wilson AC (1987) Tempo and mode of sequence evolution in mitochondrial-DNA of hawaiian Drosophila. *Journal of Molecular Evolution* **26**, 157-164.

Dethier VG (1954) Evolution of feeding preferences in phytophagous insects. *Evolution* **8**, 33-54.

Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS (1991) Touchdown PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research* **19**, 4008-4008.

Downie DA, Donaldson JS, Oberprieler RG (2008) Molecular systematics and evolution in an African cycad-weevil interaction: Amorphocerini (Coleoptera : Curculionidae : Molytinae) weevils on Encephalartos. *Molecular Phylogenetics and Evolution* **47**, 102-116.

Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *Plos Biology* **4**, 699-710.

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *Bmc Evolutionary Biology* **7**.

Ehrlich PR, Raven PH (1964) Butterflies and plants - a study in coevolution. *Evolution* **18**, 586-608.

Emerson BC, Oromi P, Hewitt GM (2000) Interpreting colonization of the Calathus (Coleoptera : Carabidae) on the Canary Islands and Madeira through the application of the parametric bootstrap. *Evolution* **54**, 2081-2090.

Erney SJ, Pruess KP, Danielson SD, Powers TO (1996) Molecular differentiation of alfalfa weevil strains (Coleoptera: Curculionidae). *Annals of the Entomological Society of America* **89**, 804-811.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. *Genetics* **131**, 479-491.

Feder JL, Chilcote CA, Bush GL (1988) Genetic differentiation between sympatric host races of the apple maggot fly Rhagoletis pomonella. *Nature* **336**, 61-64.

Feder JL, Opp SB, Wlazlo B*, et al.* (1994) Host fidelity is an effective premating barrier between sympatric races of the apple maggot fly. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 7990-7994.

Feder JL, Roethele JB, Wlazlo B, Berlocher SH (1997) Selective maintenance of allozyme differences among sympatric host races of the apple maggot fly. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 11417-11421.

Filchak KE, Feder JL, Roethele JB, Stolz U (1999) A field test for host-plant dependent selection on larvae of the apple maggot fly, Rhagoletis pomonella. *Evolution* **53**, 187-200.

Flot JF, Tillier A, Samadi S, Tillier S (2006) Phase determination from direct sequencing of length-variable DNA regions. *Molecular Ecology Notes* **6**, 627-630.

Funk DJ, Filchak KE, Feder JL (2002) Herbivorous insects: model systems for the comparative study of speciation ecology. *Genetica* **116**, 251-267.

Gassmann A, Paetel C (1998) Gymnetron netum (Col.; Curculionidae). A potential agent for biological control of toadflax in North America., p. 33. Centre for Agriculture and Biosciences International, Delémont, Switzerland.

Gibbs MJ, Armstrong JS, Gibbs AJ (2000) Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* **16**, 573-582.

Groppe K (1992) Gymnetron antirrhini Paykull (Col.: Curculionidae). A candidate for biological control of Dalmatian toadflax in North America, p. 22. Centre for Agriculture and Bioscience International, Delémont, Switzerland.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucl. Acids Symp. Ser.* **41**, 95-98.

Harrigan RJ, Mazza ME, Sorenson MD (2008) Computation vs. cloning: evaluation of two methods for haplotype determination. *Molecular Ecology Resources* **8**, 1239-1248.

Hong DY (1983) The distribution of Scrophulariaceae in the holarctic with special reference to the floristic relationships between eastern Asia and eastern North-America. *Annals of the Missouri Botanical Garden* **70**, 701-712.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755.

Hugall AF, Lee MSY (2004) Molecular claims of Gondwanan age for Australian agamid lizards are untenable. *Molecular Biology and Evolution* **21**, 2102-2110.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**, 254-267.

Jaenike J (1983) Induction of host preference in Drosophila melanogaster. *Oecologia* **58**, 320-325.

Janz N (1998) Sex-linked inheritance of host-plant specialization in a polyphagous butterfly. *Proceedings of the Royal Society of London Series B-Biological Sciences* **265**, 1675-1678.

Jordal BH, Emerson BC, Hewitt GM (2006) Apparent 'sympatric' speciation in ecologically similar herbivorous beetles facilitated by multiple colonizations of an island. *Molecular Ecology* **15**, 2935-2947.

Langor DW, Sperling FAH (1997) Mitochondrial DNA sequence divergence in weevils of the Pissodes strobi species complex (Coleoptera: Curculionidae). *Insect Molecular Biology* **6**, 255-265.

Martin AP, Costello EK, Meyer AF, Nemergut DR, Schmidt SK (2004) The rate and pattern of cladogenesis in microbes. *Evolution* **58**, 946-955.

Martin D, Rybicki E (2000) RDP: detection of recombination amongst aligned sequences. *Bioinformatics* **16**, 562-563.

Martin DP, Williamson C, Posada D (2005) RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* **21**, 260-262.

Marvaldi AE, Sequeira AS, O'Brien CW, Farrell BD (2002) Molecular and morphological phylogenetics of weevils (Coleoptera, Curculionoidea): Do niche shifts accompany diversification? *Systematic Biology* **51**, 761-785.

McClay AS, De Clerck-Floate RA (2002) Linaria vulgaris Miller, yellow toadflax (Scrophulariaceae). In: *Biological Control Programmes in Canada, 1981–2000.* (eds. Mason PG, Huber JT), pp. 375–382. CABI Publishing, Wallingford, UK.

Monteiro A, Pierce NE (2001) Phylogeny of Bicyclus (Lepidoptera : Nymphalidae) inferred from COI, COII, and EF-1 alpha gene sequences. *Molecular Phylogenetics and Evolution* **18**, 264-281.

Mopper S (1996) Adaptive genetic structure in phytophagous insect populations. *Trends in Ecology & Evolution* **11**, 235-238.

Mopper S, Strauss SY (1997) Genetic Structure and Local Adaptation in Natural Insect Populations: Effects of Ecology, Life History, and Behavior. Chapman and Hall, New York; London.

Moriyama EN, Powell JR (1997) Synonymous substitution rates in Drosophila: Mitochondrial versus nuclear genes. *Journal of Molecular Evolution* **45**, 378-391.

Normark BB (1996) Phylogeny and evolution of parthenogenetic weevils of the Aramigus tessellatus species complex (Coleoptera: Curculionidae: Naupactini): Evidence from mitochondrial DNA sequences. *Evolution* **50**, 734-745.

Oberprieler RG, Marvaldi AE, Anderson RS (2007) Weevils, weevils, weevils everywhere. *Zootaxa*, 491-520.

Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. *Virology* **265**, 218-225.

Peters JL, Zhuravlev Y, Fefelov I, Logie A, Omland KE (2007) Nuclear loci and coalescent methods support ancient hybridization as cause of mitochondrial paraphyly between gadwall and falcated duck (Anas spp.). *Evolution* **61**, 1992-2006.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817-818.

Posada D, Crandall KA (2001a) Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 13757-13762.

Posada D, Crandall KA (2001b) Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology & Evolution* **16**, 37-45.

Rambaut A, Drummond AJ (2007) LogCombiner, version 1.4.8.

Rietdorf K, Steidle JLM (2002) Was Hopkins right? Influence of larval and early adult experience on the olfactory response in the granary weevil Sitophilus granarius (Coleoptera, Curculionidae). *Physiological Entomology* **27**, 223-227.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574.

Salminen MO, Carr JK, Burke DS, McCutchan FE (1995) Identification of breakpoints in intergenotypic recombinants of HIV type-1 by bootscanning. *Aids Research and Human Retroviruses* **11**, 1423-1425.

Sanderson MJ (1997) A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution* **14**, 1218-1231.

Sanderson MJ (2002) Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. *Molecular Biology and Evolution* **19**, 101-109.

Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN, version 3.1: *a Software for Population Genetic Data Analysis.* Genetics and Biometry Laboratory, University of Geneva. Switzerland.

Scriber JM, Giebink BL, Snider D (1991) Reciprocal latitudinal clines in oviposition behavior of Papilio glaucus and P. canadensis across the Great-Lakes hybrid zone - possible sex-linkage of oviposition preferences. *Oecologia* **87**, 360-368.

Smith JM (1992) Analyzing the mosaic structure of genes. *Journal of Molecular Evolution* **34**, 126-129.

Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics* **73**, 1162-1169.

Stephens M, Smith NJ, Donnelly P (2001) A New Statistical Method for Haplotype Reconstruction from Population Data. *The American Journal of Human Genetics* **68**, 978-989.

Stireman JO, Nason JD, Heard SB (2005) Host-associated genetic differentiation in phytophagous insects: General phenomenon or isolated exceptions? Evidence from a goldenrod-insect community. *Evolution* **59**, 2573-2587.

Sutton DA (1988) *A revision of the tribe Antirrhineae.* Oxford University Press., London & Oxford.

Swofford DL (2002) Paup*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, MA.

Sword GA, Joern A, Senior LB (2005) Host plant-associated genetic differentiation in the snakeweed grasshopper, Hesperotettix viridis (Orthoptera : Acrididae). *Molecular Ecology* **14**, 2197-2205.

Templeton AR, Crandall KA, Sing CF (1992) A cladistic-analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA-sequence data. III. Cladogram estimation. *Genetics* **132**, 619-633.

Thompson RT (1992) Observations on the morphology and classification of weevils (Coleoptera, Curculionoidea) with a key to major groups. *Journal of Natural History* **26**, 835-891.

Thorne JL, Kishino H (2002) Divergence time and evolutionary rate estimation with multilocus data. *Systematic Biology* **51**, 689-702.

Via S (1999) Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice. *Evolution* **53**, 1446-1457.

Via S, Bouck AC, Skillman S (2000) Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in the parental environments. *Evolution* **54**, 1626-1637.

Wares JP (2001) Patterns of speciation inferred from mitochondrial DNA in North American Chthamalus (Cirripedia : Balanomorpha : Chthamaloidea). *Molecular Phylogenetics and Evolution* **18**, 104-116.

Waring GL, Abrahamson WG, Howard DJ (1990) Genetic differentiation among host-associated populations of the gallmaker Eurosta solidaginis (Diptera, Tephritidae). *Evolution* **44**, 1648-1655.

Xu CF, Lewis K, Cantone KL*, et al.* (2002) Effectiveness of computational methods in haplotype prediction. *Human Genetics* **110**, 148-156.

Zhang D-X, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. *Molecular Ecology* **12**, 563-584.

# Chapter 3: Molecular phylogenetic analysis of archival tissue reveals the origin of a disjunct Southern African – Palaearctic weevil radiation

# MOLECULAR PHYLOGENETIC ANALYSIS OF ARCHIVAL TISSUE REVEALS THE ORIGIN OF A DISJUNCT SOUTHERN AFRICAN – PALAEARCTIC WEEVIL RADIATION.

## ABSTRACT

As the ability to obtain DNA sequence data for phylogenetic analysis becomes less demanding with improved technology, researchers are increasingly finding sample acquisition to be a limiting factor for the maximisation of taxonomic and geographic coverage for molecular phylogenetic analyses. Such sampling limitations are problematic as they are frequently expected to be biased against species that are naturally rare, perhaps even extinct, or logistically very difficult to collect. Archival collections such as those of museums provide a wealth of material in this context, providing a resource for the augmentation of phylogenetic analyses that might otherwise suffer from sampling issues. It has been shown that high quantities of fragmented DNA can be obtained from archival material, and that sequences can be recovered using conventional PCR protocols. While this has been exploited to some degree for addressing taxonomic issues and the analysis of intraspecific variation, the utilization of archival material for interspecific phylogenetic analysis has not been addressed. Using a group of endophagous parasitic weevils to test biogeographic hypotheses for South African and Palaearctic disjunct distributions we demonstrate that the targeted amplification of short phylogenetically informative amplicons (SPIAs) from archival samples can provide sufficient data for their incorporation into molecular phylogenetic analyses.

## 3.1. INTRODUCTION

Recent advances have demonstrated the long-term persistence of short fragments of DNA in environmental samples. Recovery of DNA sequences from cave sediments, ice cores, permafrost and water samples is providing valuable information for assessing changes in species composition over time, reconstructing past flora and fauna, and the analysis of biodiversity (Ficetola et al., 2008; Hofreiter et al., 2003; Willerslev et al., 2007; Willerslev et al., 2003). These results suggest the possibility of reliably amplifying short fragments of DNA in tissues that are typically considered non-ideal for PCR-based analyses, such as archival specimens. It has been shown that high quantities of fragmented DNA, of up to only a few hundred base pairs in length, can be found in museum samples such as dried and pinned insects and formaldehyde-preserved animal tissues (Zimmermann et al., 2008). Sequences from such archival material can be recovered using conventional PCR protocols, providing useful information for studies of conservation and population genetics (Wandeler et al., 2007 and references therein). Similarly, the recent development and successful use of mini barcode amplicons (100-250 bp) for species identification and biodiversity surveys (Hajibabaei et al., 2006; Meusnier et al., 2008), offers promise in this direction.

Even though archival material, such as that contained in museum collections, represents a potential wealth of genetic information, as archival material frequently encompasses difficult to collect, rare, or even extinct species, the use of museum specimens to extract genetic information is underutilized, with a considerable bias toward taxonomic studies (e.g. Tosevski et al., in press) and few examples

addressing specific evolutionary questions (Wandeler et al., 2007). To date, most evolutionary studies have focussed within species, and have exploited DNA from subfossils (Campos et al., 2010; Gilbert et al., 2008; Hanni et al., 1994; Ritchie et al., 2004; Shapiro et al., 2004). Few studies have sought to obtain sufficient genetic information from non-ideal tissues for higher-level phylogenetic analyses (but see Huynen et al., 2003; Shapiro et al., 2002; Thomas et al., 1989).

Here we demonstrate a reliable, efficient, and non-destructive approach to obtain phylogenetically informative DNA sequence data from dried preserved insect material. Our approach differs from recent barcoding approaches that rely on the universality of primers for amplification across broad taxonomic diversity. We use genetic data sampled from non-archival material as prior information for the design of primers to amplify short phylogenetically informative amplicons (SPIAs) (<100 bp) for the incorporation of archival material in molecular phylogenetic analyses. Our approach is to maximise the number of phylogenetically informative sites, while minimising amplicon length, and optimising cross-species amplification. Using both archival specimens and freshly collected samples, we undertake a molecular phylogenetic analysis to test between competing biogeographical hypotheses of vicariance and dispersal within a group of endophagous parasitic weevils (Coleoptera: Curculionidae) exhibiting a Mediterranean-Southern African disjunct distribution. Examples of Mediterranean-Southern African disjunct distributions are perhaps better 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 (Raven, 1973; Thorne, 1972), and with some

exceptions (e.g. McGuire and Kron, 2005), the most widely accepted hypothesis is a South African origin with dispersal to the north through an East African corridor (Calvino et al., 2006; Caujape-Castells et al., 2001; Coleman et al., 2003; 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 and McGregor (2009) for examples in Lepidoptera and Diptera, respectively. In a taxonomic revision of chrysomelid beetles, Biondi and D'Alessandro (2008) conclude that, given the remarkable morphological similarities, groups with South African and Mediterranean distributions could in fact represent monophyletic units, suggesting ecological connections between both areas in the past, with long-distance dispersal events being very unlikely. Similarly, Bologna et al. (2008) revised the systematics and biogeography of *Actenodia* beetles in the family Meloidae with representatives distributed in the Mediterranean and Southern Africa, suggesting an ancient Miocene distribution of the genus from southern and eastern Africa to the north, including Arabia with further colonization of Mediterranean lands. In both cases, proposed explanations for the observed disjunct distribution favour a vicariance scenario acknowledging the possible ecological connections between the Mediterranean and South African regions in the past via "arid corridors", as originally described by Balinsky (1962), which are thought to appeared in eastern Africa since the end of the Miocene (Goldblatt, 1978; Jürgens, 1997; Verdcourt, 1969).

To further investigate disjunct Mediterranean - Southern African insect distributions we focus attention on species of the closely related genera *Rhinusa* and *Gymnetron*, within the tribe Mecinini (Curculionidae: Curculioninae), a species-rich group of parasitic weevils with representatives in both the Palaearctic and Afrotropical regions. Species within both genera are endophagous parasites whose larvae feed and develop within tissues of plant species within the families Scrophulariaceae and Plantaginaceae. The genus *Rhinusa* comprises approximately 40 species following a Palaearctic distribution (Caldara, 2001) feeding on representatives within the plant families Scrophulariaceae (*Verbascum* and *Scrophularia*) and Plantaginaceae (*Linaria*, *Kickxia*, *Chaenorhinum*, *Antirrhinum* and *Misopates*) (Caldara et al., 2010). The genus is taxonomically challenging (Hernandez-Vera et al., 2010), with few morphologically informative characters, (Caldara et al., 2010) and was until recently considered a subgenus within *Gymnetron* (Caldara, 2001). *Gymnetron* includes approximately 30 species with a Palaearctic distribution and approximately 60 species from the Afrotropical region, of which 55 are known mainly from South Africa and considered to be endemic to this area (Caldara, 2003). All Palaearctic species use plant species from the genus *Veronica* (Plantaginaceae) as host plants, whereas representatives from the Afrotropical region use different host plants within the family Scrophulariaceae (*Hebenstreitia*, *Sutera*, *Selago*, *Buddleja*, *Diascia*, *Nemesia*, *Hemimeris*) and the genus *Anastrebe* in the family Stilbaceae (Caldara et al., 2008). In a recent revision of the relationships between Mediterranean and Southern African species within the subfamily Curculioninae, Caldara et al. (2008) conclude that species from the Palaearctic and Afrotropical regions appear very closely related with only a few subtle morphological differences, and suggest probable dispersal routes in the past via the Nile river valley. However, they

recognize that it is not possible to distinguish between hypotheses of dispersal or vicariance, nor whether the genus may have originated in southern Africa, or in the Mediterranean region.

To test among competing biogeographic hypotheses for disjunct Mediterranean and Southern African insect distributions we use sequence data from two mitochondrial (cytochrome c oxidase subunit II [COII] and 16S) and three nuclear (arginine kinase [AK], 18S and elongation factor-1α [EF1-α]) genomic regions to reconstruct phylogenetic relationships of representatives within *Gymnetron* and *Rhinusa*. We maximise our taxonomic and geographic coverage by exploiting the use of archival specimens and the targeted amplification of SPIAs of the mitochondrial DNA (mtDNA) 16S gene. We test three alternative hypotheses for the disjunct Mediterranean- southern Africa 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 paleogeographic scenarios.

## 3.2. MATERIALS AND METHODS

### 3.2.1. 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 more difficult to obtain material using archival specimens (Table 1). 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-one species of *Rhinusa* were sampled (18 field collected and 14 archival; one species, *R. dieckmanni* was sampled both fresh and archivally), representing 77.5% 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, *R. villosulum* was sampled both fresh and archivally), representing eight of the 13 Afrotropical groups proposed by Caldara (2003) and 11 species from the Palaearctic region. One specimen of the genus *Cleopomiarus* was also included, which along with *Gymnetron* is the only other genus within the tribe Mecinini having representatives in both the Palaearctic and Afrotropical regions (Caldara et al., 2008).

### 3.2.2. 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), following the manufacturer's protocol. After DNA extraction, insects were placed again in 96% ethanol and kept at 4° C as voucher specimens. When using archival specimens, destruction of the sample for DNA extraction has always been a concern (Thomas, 1994; Whitfield and 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 only a single leg for DNA extraction, with

no damage to the appendage. 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 dissection 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 (Fig. 1). Full precautions were taken to rule out any possible contamination following the recommendations of Wanderler et al. (2007).

### 3.2.3. PCR amplification and sequencing reactions

Two mitochondrial and three nuclear gene fragments were utilized; cytochrome c oxidase subunit II (COII), 16S, elongation factor-1α (EF-1α), arginine kinase (AK) and the nuclear 18S rDNA (18S). Primers used for each gene are described in Table 2. For all loci, polymerase chain reactions (PCR) were performed with BIOTAQ DNA polymerase (Bioline) with $NH_4$ buffer (1x), 3.5 – 5.0 mM $MgCl_2$, 0.2mM of each dNTP, 0.2 - 0.4 µM of each primer, 0.5 U of taq polymerase and 1-5 µl of DNA template in a 25µl 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 EF1-α, 16S and 18S genes, two different touchdown profiles (Don et al., 1991) were used. For EF1-α and 16S, 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 one degree 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, 95°C for 3 min, 8 cycles at 94, 54 and 72°C for 1 min in each temperature,

decreasing the annealing temperature by one degree every cycle, then 28 cycles at 94, 46 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) in a 10 µl final volume. For all gene fragments, sequences were obtained with the forward primer, with sequences also generated with the reverse primer for the EF1-α fragment.

### 3.2.4. Phylogenetic analyses

All sequences were automatically aligned using the CLUSTAL W algorithm as implemented in BioEdit version 7.0.9 (Hall, 1999) with further manual alignment. For EF-1α, 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 v.4.0.1 (Tamura et al., 2007). Phylogenetic analyses were performed for individual partitions, and for a concatenated data set of nuclear and mitochondrial sequences. For the latter, analyses were performed using two data sets. The first one consisted of ingroup taxa with complete sampling of the 5 gene partitions. The second one 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 v3.1 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) using the substitution models selected by jModelTest 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 (5 partitions) we also used two independent runs but with four Markov chains each (8 chains in total), to optimize convergence for large datasets as suggested by Ronquist et al. (2005), and 20 million generations. Maximum likelihood analyses were performed with the parallel version of RAxML 7.0.4 (Stamatakis, 2006) for the 5 gene-partition data set 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.

### 3.2.5. Assessing the phylogenetic placement of archival marterial sampled for SPIAs

Phylogenetic placement of the archival taxa sampled for SPIAs was assessed with two approaches. As a first approach we performed a Bayesian analysis of non-archival specimens with the concatenated 5 gene partitions with nucleotide data for selected taxa, representing a range of divergences from sister lineages, reduced to

the SPIAs 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. If two hypotheses are equally likely a priori, then the Bayes factor quantifies the relative support of the competing hypotheses given the observed data (Kass and Raftery, 1995; Suchard et al., 2005). The harmonic means of sampled likelihoods were estimated using the *sump* command in MrBayes and Bayes factors were estimated as twice the difference in the natural log of the harmonic mean of model likelihoods of each model (2Δln HML). Values were interpreted according to the guidelines proposed by Kass and Raftery (1995).

### 3.2.6. Estimation of divergence times

In the absence of geological or fossil calibration points to estimate divergence times we take a Bayesian approach using a relaxed molecular clock to estimate the age of the most recent common ancestor (mrca) of nodes of interest representing divergences between South African and Palaearctic lineages. We apply 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 used then as prior information for rate calibration of the SPIAs. All analyses were run for 40 million generations and sampling frequency of 4000 generations, with a substitution model for the SPIA partition selected according to the results of jModeltest.

## 3.3. RESULTS

The five gene fragments were amplified and sequenced for 56 of the 67 field collected samples. The ten 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 5 gene partitions, only 5 samples present some missing sequence data due to poor read quality in some regions of the chromatograms (see Table 3). Sequence variation within each gene partition is detailed in Table 3. The combined 5-gene data set contained 3943 nucleotides of which 882 sites were variable and 19% were parsimony informative.

### 3.3.1. 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 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 maximise the broad taxonomic utility of these primers, including the following genera: *Curculio*, *Anthonomus*, *Eutoxus*, *Brachonyx*, *Tychius*. We also included *Tribolium,* from the family Tenebrionidae. For one of the 16S SPIAs the forward PCR primer included an M13 adaptor oligonucleotide (see Table 2), 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, 45,

72°C for 20 s, and a final extension at 72°C for 5 min. 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 (Table 1). The remaining 37 samples yielded readable sequence chromatograms for both SPIAs.

### 3.3.2. 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 (data not shown). Bayesian and ML analyses of the concatenated data set of the 5 gene partitions, excluding the 16S SPIAs, recovered the same robustly supported phylogeny (Fig. 2), with phylogenetic relationships among *Rhinusa* species broadly in agreement with the taxonomy of *Rhinusa* based on morphological characters (Caldara et al., 2010). However, neither *Rhinusa* nor *Gymnetron* are monophyletic. *Gymnetron piceum* is placed in group F as a basal lineage to group E, which includes *R. tetra, R. asellus, R. verbasci and R. bipistulata*. *Gymnetron melanarium*, *G. veronicae* and *G. villosulum* are monophyletic and form a sister group to the aforementioned groups E and F. *Cleopomiarus meridionalis* is placed basally in the tree as a sister lineage to the clade of *Gymnetron* and *Rhinusa* species (Fig. 2).

Bayesian and ML analyses including the 16S SPIAs within the 5-gene alignment resulted in identical phylogenetic placements of the archival specimens (Fig. 3), with only two minor exceptions: *G. agile* and the lineage comprising *R. depressa* and *G. bipartitum* are placed basally within clade E in the ML analysis. Despite the

limited amount of nucleotide data (150 bp), relatively high posterior probability (PP) and bootstrap support values (BS) were observed for the placement of some archival specimens. Two South African taxa, *Gymnetron perrinae* and *G. bisignatum*, are united as sister taxa with PP = 1 and BS = 93. *Rhinusa exigua* is placed as a sister lineage to the clade of *R. antirrhini* /*R. dieckmanni*/*R. florum* with PP = 0.85 and BS = 77. *Rhinusa brisouti* is placed as a sister lineage to *R. linariae* with PP = 0.90 and BS = 73. *Rhinusa moroderi* is placed as a sister lineage to *R. tetra* with PP = 0.98 and BS = 82. *Rhinusa comosa* and *R. propecomosa* are placed in a monophyletic group with *R. verbasci* and *R. cf. verbasci* with PP = 1 and BS = 96. *Gymnetron minimum and G. pauxillum* are united as sister species with PP = 1 and BS = 99. Finally, *G. aequale* and *G. veronicae* are united as sister taxa with PP = 1 and BS = 100. Twenty three archival specimens are placed within clades A, B, E, F and G described by the analysis of concatenated data set of five genes without archival specimens (Fig. 2), with the remaining 14 samples, plus the ten samples of *G. rotundicolle* and the sample of *G. rostellum*, placed outside of lineages and clades defined in Figure 2.

A Bayesian analysis of the concatenated 5-gene alignment without archival samples was performed with sequences of *R. linariae, R. vestita* and *G. piceum* trimmed to represent the 16S SPIAs, to examine how known placements might be influenced by limited sequence representation. In the case of *R. vestita* and *R. linariae*, representing moderate and intermediate divergences, respectively, from their sister lineages (Fig. 2), they were both assigned to their correct clades, but with some slight topological changes. A single branch change places *R. linariae* as a sister lineage to *R. griseohirta* within clade A (Fig. 4). Two branch arrangements result in

a more derived phylogenetic position for *R. vestita* within clade B (Fig. 4). In the more extreme case of divergence from a sister lineage, the phylogenetic placement of *G. piceum* was substantially altered (Fig. 4). These results indicate that while approximate phylogenetic placement can be achieved with limited SPIA data, exact placements may not be reliably inferred. Thus our phylogenetic results clearly support deep genetic divergences across South African taxa, however, specific relationships of South African lineages to Palaearctic lineages and the basal divergences of several South African lineages require further assessment. To achieve this we carried out Bayes factor tests to compare harmonic means of likelihood values of (i) models with enforced monophyly for each of nodes I – V (Fig. 3) that represent the mrcas of South African and Palearctic lineages, and (ii) a model enforcing the monophyly of clades A – J and the species *G. agile*, *G. bipartitum* and *R. depressa*, excluding all remaining South African species, against models without the constraint of monophyly. Applying the guidelines of Kass and Raftery (1995), we obtained very strong support for the monophyly of nodes II – IV and the clade comprising groups A – J and the species *G. agile*, *G. bipartitum* and *R. depressa* (ln Bayes factor > 10), moderate support was found for the monophyly of node I (2 < ln Bayes factor < 10) and no evidence of monophyly for node V was found (ln Bayes factor < 0).

### 3.3.3. Divergence time estimates

BEAST analyses calibrated with the evolutionary rate for the mtDNA COII gene generated a mean estimate for the age of the mrca of the *Gymnetron* and *Rhinusa* species of 26.7 mya (million years ago) with a 95% HPD interval of 21.2 - 32.5 myr (million years). This age estimate was used as prior information to calibrate the

corresponding node of the tree comprising both archival and non-archival samples (Fig. 3) to estimate the approximate divergence times of South African *Gymnetron* lineages from Palaearctic *Gymnetron* and *Rhinusa* lineages with 16S SPIA sequence data. We imposed a normal distribution for the nodal age with a mean of 26.7 myr and a standard deviation of 7 myr allowing a 95% probability distribution for sampling of 15 – 35 myr, an interval that includes the 95% HPD estimated with mtDNA COII sequence data. We also incorporated prior information for the 16S SPIA substitution rate. We estimated a range of 16S SPIA rates by fixing the mrca of *Gymnetron* and *Rhinusa* to be (i) 21.2 myr, representing the lower 2.5% posterior probability value estimated from mtDNA COII, and (ii) 32.5 myr, representing the upper 2.5% posterior probability value. Based on the rate estimates 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, incorporating our estimated range within the 95% confidence interval of the probability distribution. Ages for four nodes corresponding to divergences between South African and Palaearctic lineages (Fig. 3) were estimated and these are presented in Table 4. The age of node V was not estimated given that monophyly for that node was not supported by the Bayes factor test. The results suggest that Palaearctic and South-African lineages diverged in the late Miocene (~ 11.6 – 7.4 mya). Very similar ages were estimated for nodes II and III with means of 11.6 and 11.0 myr respectively, and 95% HPD intervals between 4.1 – 20.1. Likewise, similar ages were estimated for nodes I and IV with means of 7.4 and 8.5 myr respectively, and 95% HPD intervals between 1.1 – 16.0 (Table 4).

## 3.4. DISCUSSION

The inclusion of missing entries in phylogenetic data matrices has long been considered problematic, and is generally avoided due to concerns surrounding unresolved or inaccurate phylogenetic relationships (see Kearney and Clark, 2003 for a review). However, it has also been shown that adding incomplete taxa (i.e. taxa with missing characters) to data sets can provide data capable of testing phylogenetic hypotheses (Burleigh et al., 2009; Kearney, 2002), and in some instances even lead to increased phylogenetic resolution (Wiens, 1998; Wiens, 2005). Recent empirical investigations (Fulton and Strobeck, 2006; Wiens, 2005) and computer simulation studies (Philippe et al., 2004; Wiens, 2003) have also suggested that, rather than the amount or proportion of missing data, a crucial factor is whether or not the characters available for a taxon are sufficiently informative phylogenetically. Here we demonstrate that the geographic and taxonomic scope of phylogenetic studies can be augmented with archival specimens by the targeted amplification of phylogenetically informative DNA regions.

### 3.4.1. Phylogenetic analyses and assignment of archival specimens

Archival material such as that contained in museum collections represents a potentially vast repository of genetic information, as this material frequently encompasses difficult to collect, rare, or even extinct species. However, the exploitation of these resources for the extraction of genetic information remains underutilized and limited in scope, as most studies to date have been taxonomy-oriented. Here we illustrate how short DNA sequences from archival material can provide useful phylogenetic information to address specific evolutionary questions.

Our results indicate that taxa with relatively few characters can be approximately placed within a phylogeny of reference and provide valuable information for testing biogeographical hypotheses. Analyses of *Rhinusa* and *Gymnetron* species sampled for all 5 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 within the data matrix, Bayesian and ML analyses recovered essentially the same tree topology, with most archival samples falling within clades defined by the complete 5-gene matrix. High support values for the placement of several archival specimens, obtained in both Bayesian and ML analyses, suggest that in some cases phylogenetic placement can be achieved with a high degree of confidence using SPIA's. However, for more divergent lineages within a particular clade, phylogenetic placement may be less reliably inferred, as revealed by results of the Bayesian analyses with sequences for selected taxa reduced to the 150 nucleotides of the 16S SPIA's. Thus for nodes of relevance for our biogeographic hypotheses that lack high support from both Bayesian and maximum likelihood analyses, we have applied Bayes factor analyses to test hypotheses of monophyly against the alternative hypothesis of non-monophyly.

### 3.4.2. Historical biogeography of *Rhinusa* and *Gymnetron*

Our results support a shared evolutionary history between *Gymnetron* and *Rhinusa* species from South Africa and the Palaearctic. South African taxa are characterised by deep genetic divergences, several lineages of which form monophyletic groups

with Palaearctic *Rhinusa* lineages. Bayes factor analyses support a phylogenetic topology with basal lineages of South African origin. Following the reasoning of Bremer (1992), who developed a procedure for estimating ancestral areas from topological information of area cladograms, the basal placement of three South African lineages within the phylogeny (i: *G. gossypinus*, ii: *G. bisignatum* + *G. perinea*, iii: *G. buddleiae*) supports South Africa as the ancestral area from where extant *Gymnetron* and *Rhinusa* species diversity is derived.

It has previously been suggested that one of the more likely opportunities for large-scale range expansions into or from Africa probably occurred during the early-mid Miocene boundary, (~ 17 mya) when a land connection formed between Europe and Africa after the closure of the Tethys Sea (Levyns, 1964; McGuire and Kron, 2005). Our estimated divergence times between Palaearctic and South African lineages suggest that divergences post-date the early-mid Miocene, occurring within the late Miocene (~ 11.6 – 7.4 mya). These divergence times are consistent with an "arid corridor" connecting southern, eastern and north-eastern African areas and thought to have appeared in eastern Africa at the end of the Miocene (Goldblatt, 1978; Jürgens, 1997; Verdcourt, 1969). Further support for the east African arid corridor scenario comes from the presence of a few *Gymnetron* species (<5) in eastern Africa (Caldara et al., 2008), consistent with more northerly African relicts of an ancestral range stretching north from South Africa through east Africa. The land connection between the African and Asian plates would have facilitated range expansion further north into the Mediterranean and European regions where new vegetation zones represented the opportunity to exploit new niches. Climate-mediated vicariance events would have acted to fragment a once continuous range as conditions became

progressively drier since the mid-Miocene (Axelrod and Raven, 1978), with periods of greater aridity in the Late Miocene (~ 6 Ma), the Pliocene (~ 3 Ma), and near the Pliocene‑Pleistocene boundary (< 2 Ma) (Bobe, 2006).

Vicariance scenarios have previously been proposed 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. Molecular phylogenetic data for *Gymnetron* and *Rhinusa* provide quantitative support for a model of climate-mediated expansion of southern African lineages into the Palaearctic in the late Miocene, followed by climate-mediated vicariance. It should be noted that our divergence date estimates may overestimate the timing of the expansion of southern African lineages into the Palaearctic, due to incomplete species sampling and species extinctions. However, the broad consistency across our four age estimates argues compellingly for a late Miocene expansion of southern African *Gymnetron* lineages into the Palaearctic, a scenario that can be further evaluated with molecular analyses of other similarly disjunct invertebrate groups.

## 3.5. CONCLUSION

While it is clearly desirable to fully sample all species for all genes for molecular phylogenetic analysis, this is seldom going to be achieved due to logistical

constraints, and sampling biases will inevitably tend toward rare, extinct, or otherwise difficult to collect species. Here we have demonstrated that the targeted amplification of short phylogenetically informative amplicons can provide researchers with the ability to take advantage of archival material to augment sampling for molecular phylogenetic analyses. Using two phylogenetically informative amplicons from the mitochondrial 16S gene we have been able to increase our species sampling of *Rhinusa* and *Gymnetron* by 220%, greatly expanding species representation from South Africa. The results support a southern African origin for the group, with its range subsequently extending through eastern Africa into the Palaearctic in the late Miocene. While we have focussed on conserved sites for primer design within the 16S gene, advances in the understanding of primer design and enhanced functionality (Regier and Shi, 2005) mean that our strategy can be easily extended to protein coding genes as well, greatly enhancing the range of genes that can be exploited for the development of SPIAs.

**Acknowledgements**

This research was funded by Consejo Nacional de Ciencia y Tecnología (CONACYT) Mexico. 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. Thanks to James Kitson for assisting with removal and remounting of weevil legs on entomological cards and the image for Figure 1.

**Table 1**. Field-sampled and archival specimens used in this study.

| Accession Code | Species | Locality |
|---|---|---|
| Gy085 | *Rhinusa tetra* | Spain |
| Gy091 | *Rhinusa antirrhini* | Germany |
| Gy135 | *Rhinusa vestita* | Portugal |
| Gy495 | *R. vestita* | France |
| Gy638 | *R. linariae* | Switzerland |
| Gy640 | *R. neta* | Switzerland |
| Gy649 | *R. asellus* | Spain |
| Gy658 | *R. asellus* | Germany |
| Gy831 | *R. griseohirta* | Spain |
| Gy832 | *R. griseohirta* | Spain |
| Gy849 | *R. vestita* | Spain |
| Gy868 | *R. brondelii* | Serbia |
| Gy871 | *R. pilosa* | Romania |
| Gy872 | *R. collina* | Serbia |
| Gy881 | *R. antirrhini* | Serbia |
| Gy886 | *R. antirrhini* | UK |
| Gy894 | *R. linariae* | Serbia |
| Gy896 | *R. brondelii* | Serbia |
| Gy897 | *R. pilosa* | Serbia |
| Gy909 | *R. cf. antirrhini* | Macedonia |
| Gy912 | *R. neta* | France |
| Gy915 | *R. antirrhini* | Spain |
| Gy916 | *R. antirrhini* | Serbia |
| Gy917 | *R. antirrhini* | Serbia |
| Gy924 | *R. bipustulata* | Italy |
| Gy935 | *R. neta* | Macedonia |
| Gy938 | *R. neta* | Serbia |
| Gy944 | *R. florum* | Serbia |
| Gy949 | *R. thapsicola* | Romania |
| Gy959 | *R. bipustulata* | Italy |
| Gy984 | *R. tetra* | Serbia |
| Gy986 | *R. antirrhini* | Serbia |
| Gy996 | *R. tetra* | Serbia |
| Gy1024 | *R. dieckmanni* | Bulgaria |
| 283t1 | *R. canescens* | Italy |
| 229t1 | *Gymnetron piceum* | S Africa |
| 229t2 | *Gymnetron piceum* | S Africa |
| 350t1 | *R. cf. neta sp. nov.* | Turkey |
| 387t1 | *R. florum* | Macedonia |
| 413t1 | *R. verbasci* | Macedonia |
| 430t1 | *R. cf. linariae sp. nov.* | Turkey |
| 484t1 | *R. verbasci cf. tetra* | Macedonia |

| | | |
|---|---|---|
| 442t2 | *R. cf. neta sp. nov.* | Turkey |
| 303i | *R. collina* | Turkey |
| 401i | *R. thapsicola* | Serbia |
| 423i | *R. canescens* | Italy |
| 596i | *R. linariae* | Turkey |
| 631i | *R. melas* | Serbia |
| 632i | *R. melas* | Serbia |
| 644i | *R. florum* | Turkey |
| Gy975 | *G. melanarium* | UK |
| Gy976 | *G. veronicae* | Italy |
| Gy977 | *G. villosulum* | Italy |
| Gy1138 | *G. rotundicolle* | Italy |
| Gy1139 | *G. rotundicolle* | Italy |
| Gy1140 | *G. rotundicolle* | Italy |
| Gy1146 | *G. rotundicolle* | Italy |
| Gy1145 | *G. rotundicolle* | Italy |
| Gy1144 | *G. rotundicolle* | Italy |
| Gy1143 | *G. rotundicolle* | Italy |
| Gy1142 | *G. rotundicolle* | Italy |
| Gy1141 | *G. rotundicolle* | Italy |
| Gy1137 | *G. rotundicolle* | Italy |
| Gy1147 | *G. rostellum* | Italy |
| Gy 967 | *Cleopomiarus meridionalis* | Italy |
| § Gy1008 | *Ceutorhynchus scrobicollis* | Germany |
| § Gy1018 | *Tychius junceus* | Spain |
| | ARCHIVAL SAMPLES | |
| Gy1036 | *G. gossypinus* | S Africa |
| Gy603 | *R. herbarum* | N. A. |
| Gy604 | *R. moroderi* | N. A. |
| Gy605 | *G. villosulum* | N. A. |
| Gy1032 | *R. comosa* | N. A. |
| Gy1034 | *R. herbarum* | France |
| Gy1037 | *G. perrinae* | S Africa |
| Gy1149 | *R. scrophulariae sp. nov.* | U A Emirates |
| Gy1148 | *R. scrophulariae sp. nov.* | U A Emirates |
| T1 | *G. bipartitum* | S Africa |
| T3 | *R. mauritii* | Morocco |
| T4 | *R. depressa* | Morocco |
| T5 | *G. vittipene* | Armenia |
| T6 | *R. propecomosa* | Turkey |
| T2 | *G. bisignatum* | S Africa |
| 6D | *G. bisignatum* | S Africa |
| 2D | *R. exigua* | Turkey |
| 3D | *G. histrix* | S Africa |
| 4D | *G. histrix* | S Africa |

| | | |
|---|---|---|
| 5D | *G. agile* | S Africa |
| 7D | *G. buddleiae* | S Africa |
| 9D | *G. minimum* | S Africa |
| 10D | *G. pauxillum* | S Africa |
| 11D | *G. quadratum* | S Africa |
| 13D | *G. aequale* | Turkey |
| 14D | *G. desbrochersi* | Russia |
| 15D | *G. linkei* | Turkey |
| 16D | *G. stimulosum* | Czech Republic |
| 17D | *G. tibiellum* | Italy |
| 18D | *G. vittipene* | Armenia |
| 19D | *R. algirica* | Tunisia |
| 20D | *R. brisouti* | Kazakhstan |
| 21D | *R. emmrichi* | Turkmenistan |
| 23D | *R. mateui* | Tunisia |
| T14 | *R. exigua* | Turkey |
| Gy1033 | *R. moroderi* | N. A. |
| Gy974 | *G. clepsydra* | S Africa |
| *Gy1035 | *G. oxistomoides* | South Africa |
| *1D | *R. dieckmanni* | Bulgaria |
| *8D | *G. bipartitum* | South Africa |
| *12D | *G. simulator* | South Africa |
| *22D | *R. littorea* | France |

§ Outgroup sequences.

* Samples with comparatively weak PCR products that did not generate readable sequence chromatograms.

**Table 2.** Primers used in the amplification of nuclear and mitochondrial gene fragments: arginine kinase (AK), 18S, elongation factor-1α (EF1- α), Cytochrome Oxidase II (COII) and 16S.

| Gene | Name of primer | Sequence |
|---|---|---|
| AK | ArgKforB2 | 5'-GAYTCCGGWATYGGWATCTAYGCTCC (f) |
| | ArgKrevB1 | 5'-TCNGTRAGRCCCATWCGTCTC (r) |
| | ArgKG1f | 5'-ATYGGWATCTAYGCTCCYGAYGC (f) |
| | ArgKG1r | 5'-GCCCATWCGTCTCTTRTTRGAAAT (r) |
| 18S | 18S5' | 5'-GACAACCTGGTTGATCCTGCCAGT (f) |
| | 18Sb0.5 | 5' –GTTTCAGCTTTGCAACCAT (r) |
| | 18Sa1.0 | 5' –GGTGAAATTCTTGGACCGTC (f) |
| | 18S3'I | 5' –CACCTACGGAAACCTTGTTACGAC (r) |
| EF1- α | EF1-Bf | 5'-AGAACGTGAACGTGGTATCA (f) |
| | EF-Br | 5'-CTTGGAGTCACCAGCTACATAACC (r) |
| COII | TL2-J-3038 | 5'-TAATATGGCAGATTAGTGCATTGGA (f) |
| | TK-N 3782 | 5'-GAGACCATTACTTGCTTTCAGTCATCT (r) |
| 16S | 16Sar | 5'-CGCCTGTTTATCAAAAACAT (f) |
| | 16Sbr | 5'-CCGGTCTGAACTCAGATCACGT (r) |
| | 16S_7bp_FGer | 5'-[GTAAAACGACGGCCAGT]AATMATTAGTTTYTTAATT (f) |
| | 16S_7bp_RGer | 5'-TAYAGGGTCTTCTCGTCTT (r) |
| | 16S_48bpF1 | 5'-CGAGAAGACCCTATAGAGTTT (f) |
| | 16S_48bpR1 | 5'-TCAATCACCCCAAYYAAAT (r) |

(f) = Forward reading, (r) = Reverse reading. [ ] = M13 adaptor sequence.

**Table 3.** Summary statistics for the five sequenced genes and the 16S SPIAs.

| | COII | 16S | 16S SPIAs | AK | EF1-α | 18S |
|---|---|---|---|---|---|---|
| Aligned base pairs | 697[a] | 307 | 150 | 715[b] | 647[c] | 1577 |
| Conserved sites | 300 | 176 | 72 | 537 | 504 | 1540 |
| Variable sites | 397 | 130 | 78 | 178 | 143 | 37 |
| Parsimony informative | 372 (53.4%) | 104 (33.9%) | 67 (44.6%) | 132 (18.5%) | 118 (18.2%) | 30 (2%) |

[a] 682 for Gy1018

[b] 423 for Gy976, 382 for Gy977

[c] 418 for Gy975 and Gy976, 628 for Gy967

**Table 4.** Estimated times to the most recent common ancestor (tmrca) for nodes indicating splits of South African – Palaearctic lineages.

| Node number | Mean (My) | 95% HPD intervals |
|---|---|---|
| I. | 7.4 | 1.1 – 15.9 |
| II. | 11.6 | 4.1 – 20.1 |
| III. | 11.0 | 4.1 – 19.1 |
| IV. | 8.5 | 2.6 – 16.0 |

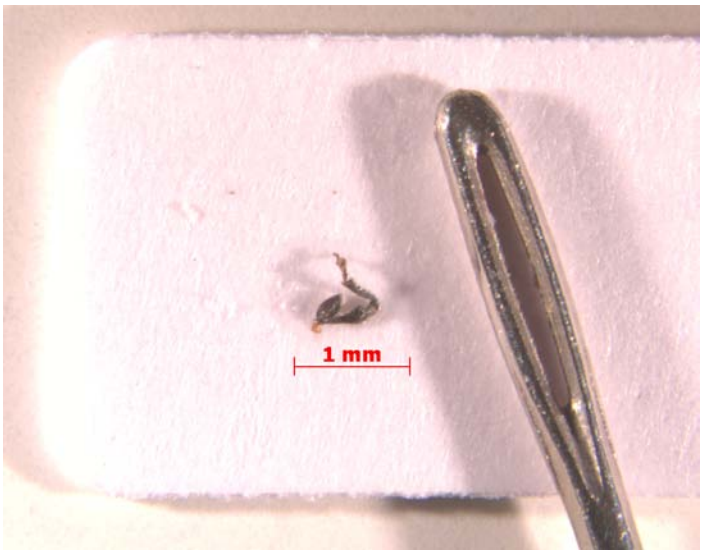


**Figure 1.** Leg of a dry pinned weevil remounted on a new entomological card after DNA extraction.

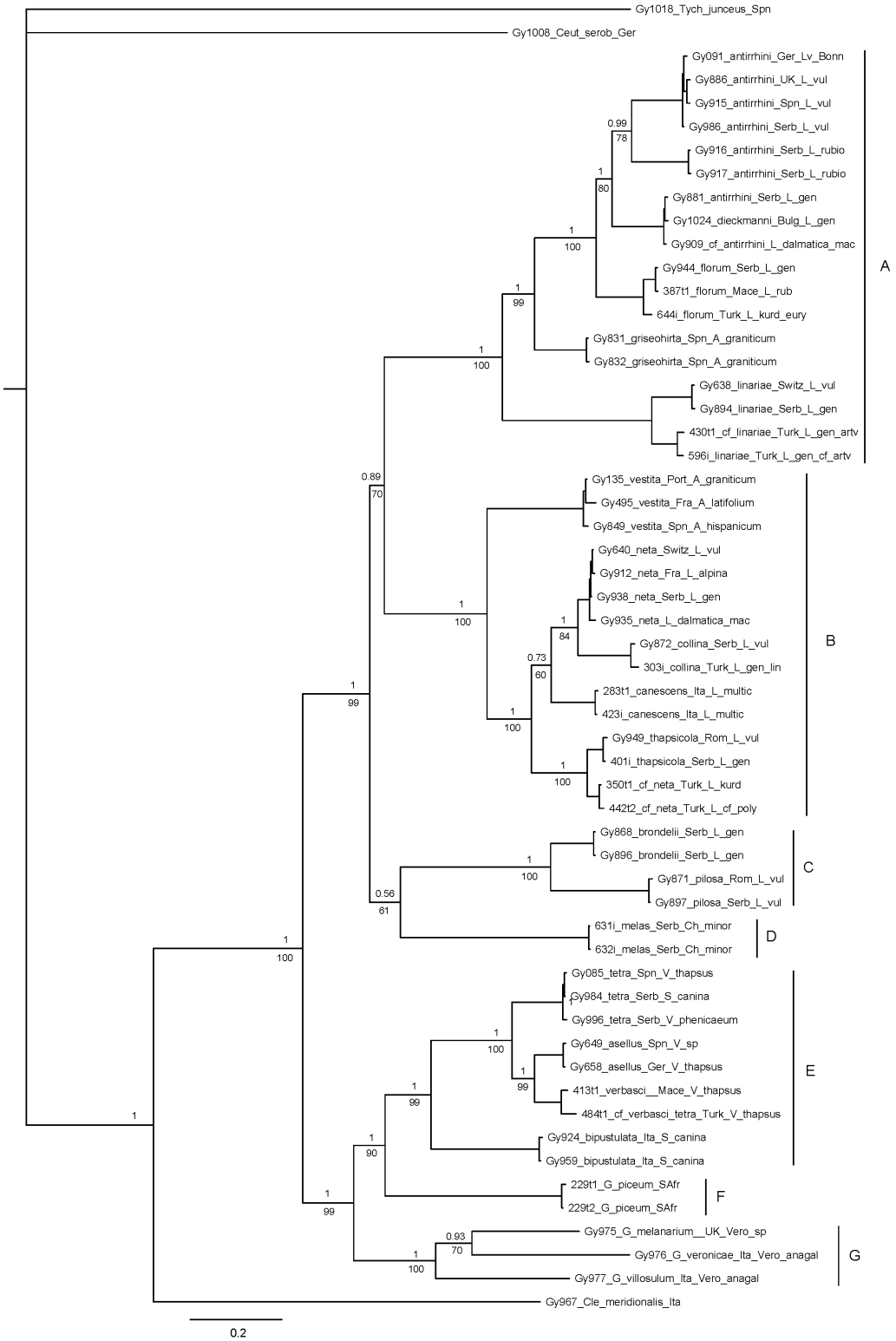
Gy1018_Tych_junceus_Spn
Gy1008_Ceut_serob_Ger
Gy091_antirrhini_Ger_Lv_Bonn
Gy886_antirrhini_UK_L_vul
Gy915_antirrhini_Spn_L_vul
Gy986_antirrhini_Serb_L_vul
Gy916_antirrhini_Serb_L_rubio
Gy917_antirrhini_Serb_L_rubio
Gy881_antirrhini_Serb_L_gen
Gy1024_dieckmanni_Bulg_L_gen
Gy909_cf_antirrhini_L_dalmatica_mac
Gy944_florum_Serb_L_gen
387t1_florum_Mace_L_rub
644i_florum_Turk_L_kurd_eury
Gy831_griseohirta_Spn_A_graniticum
Gy832_griseohirta_Spn_A_graniticum
Gy638_linariae_Switz_L_vul
Gy894_linariae_Serb_L_gen
430t1_cf_linariae_Turk_L_gen_artv
596i_linariae_Turk_L_gen_cf_artv
A
Gy135_vestita_Port_A_graniticum
Gy495_vestita_Fra_A_latifolium
Gy849_vestita_Spn_A_hispanicum
Gy640_neta_Switz_L_vul
Gy912_neta_Fra_L_alpina
Gy938_neta_Serb_L_gen
Gy935_neta_L_dalmatica_mac
Gy872_collina_Serb_L_vul
303i_collina_Turk_L_gen_lin
283t1_canescens_lta_L_multic
423i_canescens_lta_L_multic
Gy949_thapsicola_Rom_L_vul
401i_thapsicola_Serb_L_gen
350t1_cf_neta_Turk_L_kurd
442t2_cf_neta_Turk_L_cf_poly
B
Gy868_brondelii_Serb_L_gen
Gy896_brondelii_Serb_L_gen
Gy871_pilosa_Rom_L_vul
Gy897_pilosa_Serb_L_vul
C
631i_melas_Serb_Ch_minor
632i_melas_Serb_Ch_minor
D
Gy085_tetra_Spn_V_thapsus
Gy984_tetra_Serb_S_canina
Gy996_tetra_Serb_V_phenicaeum
Gy649_asellus_Spn_V_sp
Gy658_asellus_Ger_V_thapsus
413t1_verbasci__Mace_V_thapsus
484t1_cf_verbasci_tetra_Turk_V_thapsus
Gy924_bipustulata_lta_S_canina
Gy959_bipustulata_lta_S_canina
E
229t1_G_piceum_SAfr
229t2_G_piceum_SAfr
F
Gy975_G_melanarium__UK_Vero_sp
Gy976_G_veronicae_lta_Vero_anagal
Gy977_G_villosulum_lta_Vero_anagal
G
Gy967_Cle_meridionalis_lta
0.2

**Figure 2.** Bayesian phylogenetic tree inferred from a concatenated data set of 5 DNA sequence partitions (cytochrome c oxidase subunit II [COII], 16S, elongation factor-1α [EF-1α], arginine kinase [AK] and the nuclear 18S rDNA [18S] ) comprising 3,943 bp. Bayesian posterior probabilities and ML bootstrap values are shown above and below branches, respectively. Seven higher order lineages and well-supported clades are labelled A – G.

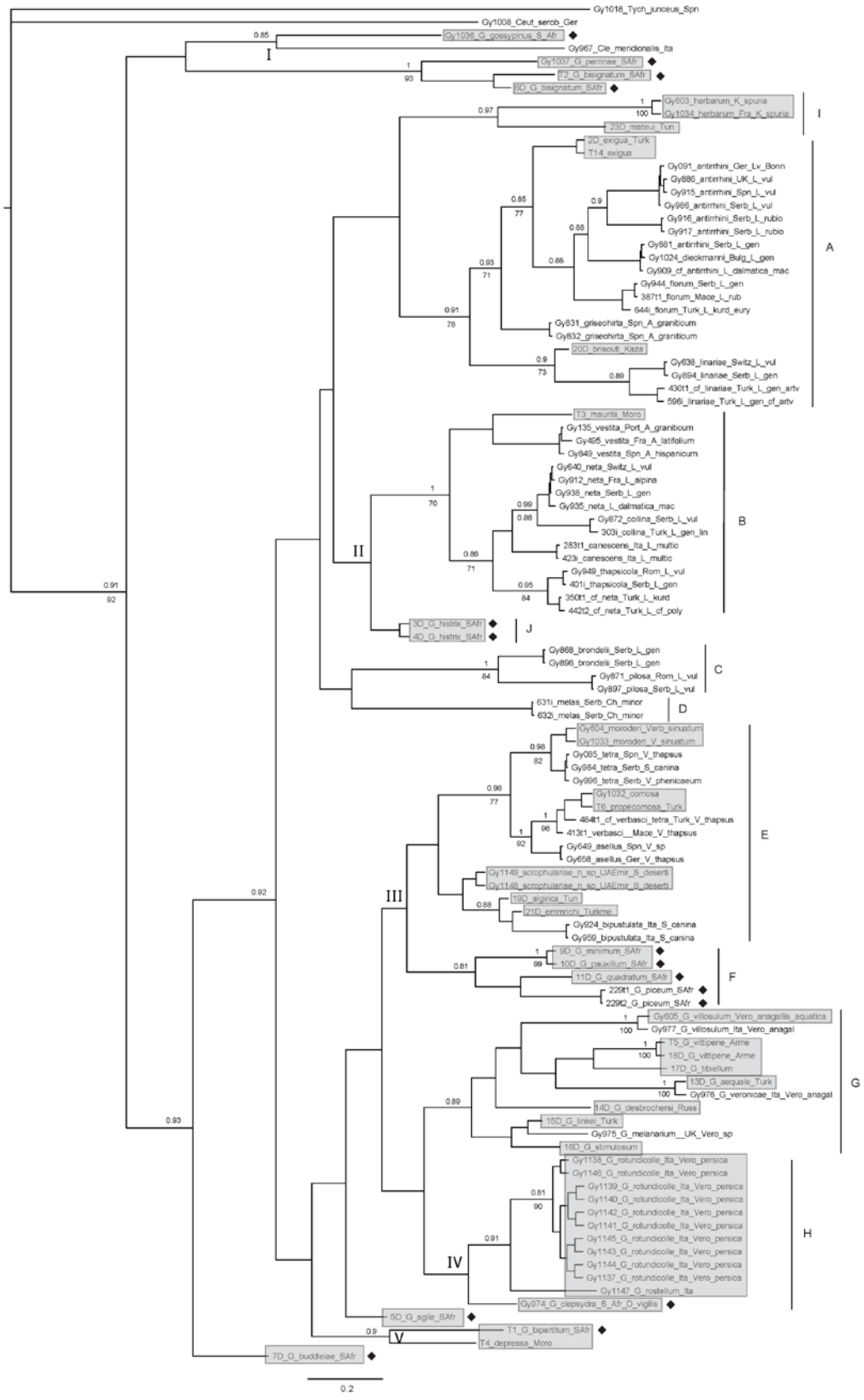
Gy1018_Tych_junceus_Spn
Gy1008_Ceut_sercb_Ger
Gy1036_G_gossypinus_S_Afr
Gy967_Cle_meridionalis_Ita
Gy1037_G_perrinae_SAfr
T2_G_bisignatum_SAfr
6D_G_bisignatum_SAfr
Gy803_herbarum_K_spuria
Gy1034_herbarum_Fra_K_spuria
23D_mateui_Tun
2D_exigua_Turk
T14_exigua
Gy091_antirrhini_Ger_Lv_Bonn
Gy886_antirrhini_UK_L_vul
Gy915_antirrhini_Spn_L_vul
Gy986_antirrhini_Serb_L_vul
Gy916_antirrhini_Serb_L_rubio
Gy917_antirrhini_Serb_L_rubio
Gy881_antirrhini_Serb_L_gen
Gy1024_dieckmanni_Bulg_L_gen
Gy909_cf_antirrhini_L_dalmatica_mac
Gy944_florum_Serb_L_gen
387t1_florum_Mace_L_rub
644i_florum_Turk_L_kurd_eury
Gy831_griseohirta_Spn_A_graniticum
Gy832_griseohirta_Spn_A_graniticum
20D_briscuti_Kaza
Gy638_linariae_Switz_L_vul
Gy894_linariae_Serb_L_gen
430t1_cf_linariae_Turk_L_gen_artv
596i_linariae_Turk_L_gen_cf_artv
T3_mauritii_Moro
Gy135_vestita_Port_A_graniticum
Gy495_vestita_Fra_A_latifolium
Gy849_vestita_Spn_A_hispanicum
Gy640_neta_Switz_L_vul
Gy912_neta_Fra_L_alpina
Gy938_neta_Serb_L_gen
Gy935_neta_L_dalmatica_mac
Gy672_collina_Serb_L_vul
303i_collina_Turk_L_gen_lin
283t1_canescens_Ita_L_multic
423i_canescens_Ita_L_multic
Gy949_thapsicola_Rom_L_vul
401i_thapsicola_Serb_L_gen
350t1_cf_neta_Turk_L_kurd
442t2_cf_neta_Turk_L_cf_poly
3D_G_histrix_SAfr
4D_G_histrix_SAfr
Gy868_brondelii_Serb_L_gen
Gy896_brondelii_Serb_L_gen
Gy871_pilosa_Rom_L_vul
Gy897_pilosa_Serb_L_vul
631i_melas_Serb_Ch_minor
632i_melas_Serb_Ch_minor
Gy604_moroderi_Verb_sinuatum
Gy1033_moroderi_V_sinuatum
Gy085_tetra_Spn_V_thapsus
Gy984_tetra_Serb_S_canina
Gy996_tetra_Serb_V_phenicaeum
Gy1032_comosa
T6_propecomosa_Turk
484t1_cf_verbasci_tetra_Turk_V_thapsus
413t1_verbasci__Mace_V_thapsus
Gy649_asellus_Spn_V_sp
Gy658_asellus_Ger_V_thapsus
Gy1149_scrophulariae_n_sp_UAEmir_S_deserti
Gy1148_scrophulariae_n_sp_UAEmir_S_deserti
19D_algirica_Tun
21D_emmrichi_Turkme
Gy924_bipustulata_Ita_S_canina
Gy959_bipustulata_Ita_S_canina
9D_G_minimum_SAfr
10D_G_pauxillum_SAfr
11D_G_quadratum_SAfr
229t1_G_piceum_SAfr
229t2_G_piceum_SAfr
Gy605_G_villosulum_Vero_anagallis_aquatica
Gy977_G_villosulum_Ita_Vero_anagal
T5_G_vittipene_Arme
18D_G_vittipene_Arme
17D_G_tibellum
13D_G_aequale_Turk
Gy976_G_veronicae_Ita_Vero_anagal
14D_G_desbrochersi_Russ
15D_G_linkei_Turk
Gy975_G_melanarium__UK_Vero_sp
16D_G_stimulosum
Gy1138_G_rotundicolle_Ita_Vero_persica
Gy1146_G_rotundicolle_Ita_Vero_persica
Gy1139_G_rotundicolle_Ita_Vero_persica
Gy1140_G_rotundicolle_Ita_Vero_persica
Gy1142_G_rotundicolle_Ita_Vero_persica
Gy1141_G_rotundicolle_Ita_Vero_persica
Gy1145_G_rotundicolle_Ita_Vero_persica
Gy1143_G_rotundicolle_Ita_Vero_persica
Gy1144_G_rotundicolle_Ita_Vero_persica
Gy1137_G_rotundicolle_Ita_Vero_persica
Gy1147_G_rostellum_Ita
Gy974_G_clepsydra_S_Afr_D_vigilis
5D_G_agile_SAfr
T1_G_bipartitum_SAfr
T4_depressa_Moro
7D_G_buddleiae_SAfr
I
II
III
IV
V
A
B
C
D
E
F
G
H
J
0.2

**Figure 3.** Bayesian phylogenetic tree inferred from a concatenated data set of 5 gene DNA sequences comprising 3,943 bp from 54 ingroup taxa plus 37 archival specimens and 11 degraded samples with only 150 bp of nucleotide data corresponding to the 16S SPIAs (highlighted in grey). Specimens with a South African origin are indicated with a black diamond. Bayesian posterior probabilities ≥ 0.8 and ML bootstrap values ≥ 70% are shown above and below branches, respectively. For clarity, support values at terminal taxa were omitted except when these imply archival specimens with significant values. Major groups are indicated (A – J). Roman numerals refer to estimated divergence times given in Table 4.

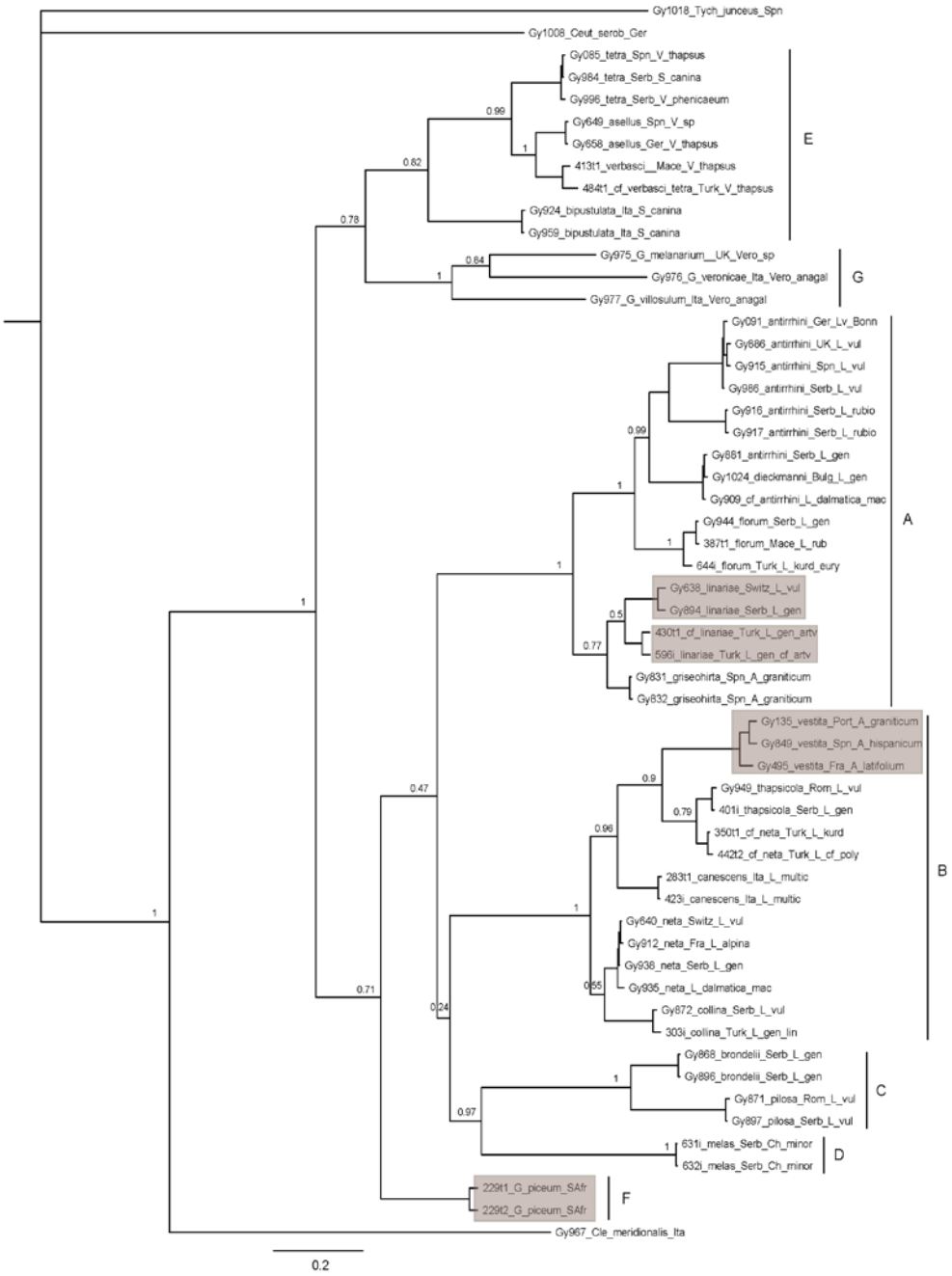


**Figure 4.** Bayesian phylogenetic tree inferred from a concatenated data set of 5 gene DNA sequences (cytochrome c oxidase subunit II [COII], 16S, elongation factor-1α [EF-1α], arginine kinase [AK] and the nuclear 18S rDNA [18S] ) comprising 3,943 bp. Nucleotide data for highlighted taxa were reduced to 150 bp corresponding to the 16S SPIAs. Posterior probabilities are shown above branches and major groups are indicated (A – G).

## REFERENCES

Audisio, P., A. De Biase, A. H. Kirk-Spriggs, A. R. Cline, M. Trizzino, G. Antonini, and E. Mancini. 2008. Molecular biogeography of Mediterranean and southern African disjunctions as exemplified by pollen beetles of the Meligethes planiusculus species-group and related taxa (Coleoptera: Nitidulidae; Meligethinae. Biogeographia 29:45-65.

Axelrod, D. I., and P. H. Raven. 1978. Late Cretaceous and Tertiary vegetation history of Africa*in* Biogeography and ecology of Southern Africa (M. J. A. Werger, ed.) Dr W. Junk bv, The Hague, The Netherlands.

Balinsky, B. I. 1962. Patterns of animal distribution of the African continent. Ann. Cape Province Mus. 2:299–310.

Biondi, M., and P. D'Alessandro. 2008. Taxonomical revision of the Longitarsus capensis species-group: An example of Mediterranean-southern African disjunct distributions (Coleoptera: Chrysomelidae). European Journal of Entomology 105:719-736.

Bobe, R. 2006. The evolution of arid ecosystems in eastern Africa. Journal of Arid Environments 66:564-584.

Bologna, M. A., A. Di Giulio, and M. Pitzalis. 2008. Systematics and biogeography of the genus Actenodia (Coleoptera : Meloidae : Mylabrini). Systematic Entomology 33:319-360.

Bremer, K. 1992. Ancestral areas - A cladistic reinterpretation of the center of origin concept. Systematic Biology 41:436-445.

Burleigh, J. G., K. W. Hilu, and D. E. Soltis. 2009. Inferring phylogenies with incomplete data sets: a 5-gene, 567-taxon analysis of angiosperms. Bmc Evolutionary Biology 9.

Caldara, R. 2001. Phylogenetic analysis and higher classification of the tribe Mecinini (Coleoptera: Curculionidae, Curculioninae). Koleopterologische Rundschau 71:171-203.

Caldara, R. 2003. Revisione dei Gymnetron della Regione Afrotropicale (Coleoptera Curculionidae). Mem. Soc. Entomol. Ital. 82:87-272.

Caldara, R., E. Colonnelli, and G. Osella. 2008. Curculionidae Curculioninae: relationships between Mediterranean and southern African species (Coleoptera). Biogeographia 29:133-156.

Caldara, R., D. Sassi, and I. Tosevski. 2010. Phylogeny of the weevil genus *Rhinusa* Stephens based on adult morphological characters and host plant information (Coleoptera: Curculionidae). Zootaxa 2627:39-56.

Calvino, C. I., P. M. Tilney, B. E. van Wyk, and S. R. Downie. 2006. A molecular phylogenetic study of southern African Apiaceae. American Journal of Botany 93:1828-1847.

Campos, P. F., E. Willerslev, A. Sher, L. Orlando, E. Axelsson, A. Tikhonov, K. Aaris-Sorensen, A. D. Greenwood, R. D. Kahlke, P. Kosintsev, T. Krakhmalnaya, T. Kuznetsova, P. Lemey, R. MacPhee, C. A. Norris, K. Shepherd, M. A. Suchard, G. D. Zazula, B. Shapiro, and M. T. P. Gilbert. 2010. Ancient DNA analyses exclude humans as the driving force behind late Pleistocene musk ox (Ovibos moschatus) population dynamics. Proceedings of the National Academy of Sciences of the United States of America 107:5675-5680.

Caujape-Castells, J., R. K. Jansen, N. Membrives, J. Pedrola-Monfort, J. M. Montserrat, and A. Ardanuy. 2001. Historical biogeography of Androcymbium Willd. (Colchicaceae) in Africa: evidence from cpDNA RFLPs. Botanical Journal of the Linnean Society 136:379-392.

Cicconardi, F., F. Nardi, B. C. Emerson, F. Frati, and P. P. Fanciulli. 2010. Deep phylogeographic divisions and long-term persistence of forest invertebrates in the North-Western Mediterranean basin. Molecular Ecology 19:386-400.

Coleman, M., A. Liston, J. W. Kadereit, and R. J. Abbott. 2003. Repeat intercontinental dispersal and Pleistocene speciation in disjunct Mediterranean and desert Senecio (Asteraceae). American Journal of Botany 90:1446-1454.

Cowling, R. M., P. W. Rundel, B. B. Lamont, M. K. Arroyo, and M. Arianoutsou. 1996. Plant diversity in Mediterranean-climate regions. Trends in Ecology & Evolution 11:362-366.

del Hoyo, A., J. L. Garcia-Marin, and J. Pedrola-Monfort. 2009. Temporal and spatial diversification of the African disjunct genus Androcymbium (Colchicaceae). Molecular Phylogenetics and Evolution 53:848-861.

Don, R. H., P. T. Cox, B. J. Wainwright, K. Baker, and J. S. Mattick. 1991. Touchdown PCR to circumvent spurious priming during gene amplification. Nucleic Acids Research 19:4008-4008.

Ficetola, G. F., C. Miaud, F. Pompanon, and P. Taberlet. 2008. Species detection using environmental DNA from water samples. Biology Letters 4:423-425.

Fulton, T. L., and C. Strobeck. 2006. Molecular phylogeny of the Arctoidea (Carnivora): Effect of missing data on supertree and supermatrix analyses of multiple gene data sets. Molecular Phylogenetics and Evolution 41:165-181.

Gilbert, M. T. P., D. I. Drautz, A. M. Lesk, S. Y. W. Ho, J. Qi, A. Ratan, C. H. Hsu, A. Sher, L. Dalen, A. Gotherstrom, L. P. Tomsho, S. Rendulic, M. Packard, P. F. Campos, T. V. Kuznetsova, F. Shidlovskiy, A. Tikhonov, E. Willerslev, P. Iacumin, B. Buigues, P. G. P. Ericson, M. Germonpre, P. Kosintsev, V. Nikolaev, M. Nowak-Kemp, J. R. Knight, G. P. Irzyk, C. S. Perbost, K. M. Fredrikson, T. T. Harkins, S. Sheridan, W. Miller, and S. C. Schuster. 2008. Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. Proceedings of the National Academy of Sciences of the United States of America 105:8327-8332.

Gilbert, M. T. P., W. Moore, L. Melchior, and M. Worobey. 2007. DNA Extraction from Dry Museum Beetles without Conferring External Morphological Damage. PLoS ONE 2.

Goldblatt, P. 1978. Analysis of the flora of southern Africa: its characteristics, relationships, and origins. Annals of the Missouri Botanical Garden 65:369-436.

Hajibabaei, M., M. A. Smith, D. H. Janzen, J. J. Rodriguez, J. B. Whitfield, and P. D. N. Hebert. 2006. A minimalist barcode can identify a specimen whose DNA is degraded. Molecular Ecology Notes 6:959-964.

Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucl. Acids Symp. Ser. 41:95-98.

Hanni, C., V. Laudet, D. Stehelin, and P. Taberlet. 1994. Tracking the origins of the cave bear (Ursus-spelaeus) by mitochondrial-DNA sequencing. Proceedings of the National Academy of Sciences of the United States of America 91:12336-12340.

Hernandez-Vera, G., M. Mitrovic, J. Jovic, I. Tosevski, R. Caldara, A. Gassmann, and B. C. Emerson. 2010. Host-associated genetic differentiation in a seed parasitic weevil Rhinusa antirrhini (Coleptera: Curculionidae) revealed by mitochondrial and nuclear sequence data. Molecular Ecology 19:2286-2300.

Hofreiter, M., J. I. Mead, P. Martin, and H. N. Poinar. 2003. Molecular caving. Current Biology 13:R693-R695.

Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754-755.

Huynen, L., C. D. Millar, R. P. Scofield, and D. M. Lambert. 2003. Nuclear DNA sequences detect species limits in ancient moa. Nature 425:175-178.

Jürgens, N. 1997. Floristic biodiversity and history of African arid regions. Biodiversity and Conservation 6:495-514.

Kass, R. E., and A. E. Raftery. 1995. Bayes Factors. Journal of the American Statistical Association 90:773-795.

Kearney, M. 2002. Fragmentary taxa, missing data, and ambiguity: Mistaken assumptions and conclusions. Systematic Biology 51:369-381.

Kearney, M., and J. M. Clark. 2003. Problems due to missing data in phylogenetic analyses including fossils: A critical review. Journal of Vertebrate Paleontology 23:263-274.

Kirk-Spriggs, A. H., and G. McGregor. 2009. Disjunctions in the Diptera (Insecta) fauna of the Mediterranean Province and southern Africa and a discussion of biogeographical considerations. Transactions of the Royal Society of South Africa 64:32 - 52.

Levyns, M. R. 1964. Migrations and origin of the Cape Flora. Transactions of the Royal Society of South Africa 37:85-107.

McGuire, A. F., and K. A. Kron. 2005. Phylogenetic relationships of European and African Ericas. International Journal of Plant Sciences 166:311-318.

Meusnier, I., G. A. C. Singer, J. F. Landry, D. A. Hickey, P. D. N. Hebert, and M. Hajibabaei. 2008. A universal DNA mini-barcode for biodiversity analysis. Bmc Genomics 9.

Mey, W. 2006. Revision of the genus Phyllobrostis Staudinger, 1859 (Lepidoptera, Lyonetiidae). Deutsche Entomologische Zeitschrift 53:114-147.

Philippe, H., E. A. Snell, E. Bapteste, P. Lopez, P. W. H. Holland, and D. Casane. 2004. Phylogenomics of eukaryotes: Impact of missing data on large alignments. Molecular Biology and Evolution 21:1740-1752.

Raven, P. H. 1973. The evolution of Mediterranean floras. Pages 213–224 *in* Mediterranean Type Ecosystems: Origin and Structure (F. di Castri, and H. A. Mooney, eds.). Springer-Verlag, New York, USA.

Regier, J. C., and D. Shi. 2005. Increased yield of PCR product from degenerate primers with nondegenerate, nonhomologous 5 ' tails. Biotechniques 38:34-38.

Ritchie, P. A., C. D. Millar, G. C. Gibb, C. Baroni, and D. M. Lambert. 2004. Ancient DNA enables timing of the Pleistocene origin and Holocene expansion of two Adelie penguin lineages in Antarctica. Molecular Biology and Evolution 21:240-248.

Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.

Ronquist, F., J. P. Huelsenbeck, and P. van der Mark. 2005. MrBayes 3.1 Manual, draft 5/ 26/2005.

Rowley, D. L., J. A. Coddington, M. W. Gates, A. L. Norrbom, R. A. Ochoa, N. J. Vandenberg, and M. H. Greenstone. 2007. Vouchering DNA-barcoded specimens: test of a nondestructive extraction protocol for terrestrial arthropods. Molecular Ecology Notes 7:915-924.

Shapiro, B., A. J. Drummond, A. Rambaut, M. C. Wilson, P. E. Matheus, A. V. Sher, O. G. Pybus, M. T. P. Gilbert, I. Barnes, J. Binladen, E. Willerslev, A. J. Hansen, G. F. Baryshnikov, J. A. Burns, S. Davydov, J. C. Driver, D. G. Froese, C. R. Harington, G. Keddie, P. Kosintsev, M. L. Kunz, L. D. Martin, R. O. Stephenson, J. Storer, R. Tedford, S. Zimov, and A. Cooper. 2004. Rise and fall of the Beringian steppe bison. Science 306:1561-1565.

Shapiro, B., D. Sibthorpe, A. Rambaut, J. Austin, G. M. Wragg, O. R. P. Bininda-Emonds, P. L. M. Lee, and A. Cooper. 2002. Flight of the dodo. Science 295:1683-1683.

Stamatakis, A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688-2690.

Stamatakis, A., P. Hoover, and J. Rougemont. 2008. A Rapid Bootstrap Algorithm for the RAxML Web Servers. Systematic Biology 57:758-771.

Suchard, M. A., R. E. Weiss, and J. S. Sinsheimer. 2005. Models for estimating Bayes factors with applications to phylogeny and tests of monophyly. Biometrics 61:665-673.

Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24:1596-1599.

Thomas, R. H. 1994. Molecules, museums and vouchers. Trends in Ecology & Evolution 9:413-414.

Thomas, R. H., W. Schaffner, A. C. Wilson, and S. Paabo. 1989. DNA phylogeny of the extinct marsupial wolf. Nature 340:465-467.

Thorne, R. F. 1972. Major disjunctions in the geographic ranges of seed plants. Quarterly Review of Biology 47:365-411.

Tosevski, I., R. Caldara, J. Jovic, G. Hernandez-Vera, C. Baviera, A. Gassmann, and B. C. Emerson. in press. Morphological, molecular and biological evidence reveal two cryptic species in Mecinus janthinus Germar (Coleoptera, Curculionidae), a successful biological control agent of Dalmatian toadflax, Linaria dalmatica (Lamiales, Plantaginaceae).

Verdcourt, B. 1969. The arid corridor between the north-east and south-west areas of Africa. Palaeoecol. Afr. 4:140–144.

Wandeler, P., P. E. A. Hoeck, and L. F. Keller. 2007. Back to the future: museum specimens in population genetics. Trends in Ecology & Evolution 22:634-642.

Whitfield, J. B., and S. A. Cameron. 1994. Museum policies concerning specimen loans for molecular systematic research. Molecular Phylogenetics and Evolution 3:268-270.

Wiens, J. J. 1998. Does adding characters with missing data increase or decrease phylogenetic accuracy? Systematic Biology 47:625-640.

Wiens, J. J. 2003. Missing data, incomplete taxa, and phylogenetic accuracy. Systematic Biology 52:528-538.

Wiens, J. J. 2005. Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? Systematic Biology 54:731-742.

Willerslev, E., E. Cappellini, W. Boomsma, R. Nielsen, M. B. Hebsgaard, T. B. Brand, M. Hofreiter, M. Bunce, H. N. Poinar, D. Dahl-Jensen, S. Johnsen, J. P. Steffensen, O. Bennike, J. L. Schwenninger, R. Nathan, S. Armitage, C. J. de Hoog, V. Alfimov, M. Christl, J. Beer, R. Muscheler, J. Barker, M. Sharp, K. E. H. Penkman, J. Haile, P. Taberlet, M. T. P. Gilbert, A. Casoli, E. Campani, and M. J. Collins. 2007. Ancient biomolecules from deep ice cores reveal a forested Southern Greenland. Science 317:111-114.

Willerslev, E., A. J. Hansen, J. Binladen, T. B. Brand, M. T. P. Gilbert, B. Shapiro, M. Bunce, C. Wiuf, D. A. Gilichinsky, and A. Cooper. 2003. Diverse plant and animal genetic records from Holocene and Pleistocene sediments. Science 300:791-795.

Zimmermann, J., M. Hajibabaei, D. C. Blackburn, J. Hanken, E. Cantin, J. Posfai, and T. C. Evans. 2008. DNA damage in preserved specimens and tissue samples: a molecular assessment. Frontiers in Zoology 5:18.

# Chapter 4: Host associations and molecular systematics within parasitic weevils of *Rhinusa* and *Gymnetron* (Coleoptera : Curculionidae)

# HOST ASSOCIATIONS AND MOLECULAR SYSTEMATICS WITHIN PARASITIC WEEVILS OF THE GENUS *Rhinusa* AND *Gymnetron* (COLEOPTERA : CURCULIONIDAE)

## ABSTRACT

The study of insect-host plant associations has long been a focus of attention by biologists because they represent a useful arena in which one can test different evolutionary hypotheses. Here, we investigate the evolution of host use in *Rhinusa* and *Gymnetron*, two closely related groups of weevils parasitizing plant species within the families Scrophulariaceae and Plantaginaceae. The use of plant tissues by *Rhinusa* and *Gymnetron* for the completion of their reproductive cycle suggests an intimate relationship with their hosts, thus representing an excellent opportunity to study insect-host plant associations in the context of speciation and diversification. DNA sequences from two mitochondrial (COII and 16S) and three nuclear gene fragments (AK, EF1-α and 18S) were used to reconstruct phylogenetic relationships and test hypotheses of conservatism in host-plant utilization and parasitic mode. Ancestral states of host plant family utilization were reconstructed using maximum likelihood optimization. The results indicate that host utilization is a conserved trait at the plant family level, but less so at the genus level for both host utilization and mode of parasitism. The labile quality in these two dimensions of host-plant use explain in part the successful diversification of this group of weevils, particularly the fine partitioning of ecological resources by utilizing different plant tissue/organs within single host species. The inferred phylogenetic relationships are largely in agreement with the morphology-based taxonomy of the group; however *Gymnetron* and *Rhinusa* are not reciprocally monophyletic.

## 4.1. INTRODUCTION

Comprising approximately 5,800 genera and more than 60,000 described species (McKenna *et al.* 2009), weevils (Coleoptera : Curculionoidae) have been described as one of the most successful adaptive radiations on Earth (Mayr 1963; McKenna *et al.* 2009). The evolution of a rostrum, shifts in larval feeding habits and co-evolutionary relationships with flowering plants have been proposed as likely explanations for this diversity (Marvaldi *et al.* 2002; McKenna *et al.* 2009; Oberprieler *et al.* 2007). Using every plant part and nearly every plant taxon (Anderson 1995; McKenna *et al.* 2009), weevils exhibit a close relationship with their host-plants and, as is frequently observed within many other plant-feeding insect groups (Futuyma *et al.* 1993; Jaenike 1990; Janz & Nylin 1998), specialization on one or a few closely related host-plant species is a recurrent phenomenon (Marvaldi *et al.* 2002). In some cases the degree of weevil specialization is reflected in life history attributes such as endophagy, in which larvae not only feed but also develop inside a great variety of plant structures. Although there are several phylogenetic analyses of host use in insects, most of these have focused on external feeders such as Lepidoptera (e. g. Ehrlich & Raven 1964; Janz & Nylin 1998; Miller 1987; Wahlberg 2001), chrysomelid leaf beetles (e. g. Becerra & Venable 1999; Farrell & Mitter 1990; Farrell 1998; Kölsch & Pedersen 2008), and Hemiptera (e. g. Percy 2003; Percy *et al.* 2004). In contrast, relatively less attention has been given to endophagous insects (but see Crespi *et al.* 1998; Kergoat *et al.* 2007; Lopez-Vaamonde *et al.* 2003; Nyman *et al.* 2000).

Here, we focus attention on the evolution of host plant use among species from the genera *Rhinusa* and *Gymnetron*. Both genera represent endophagous parasitic weevils (Curculionidae) whose larvae feed and develop within tissues of plant species in the families Scrophulariaceae and Plantaginaceae. The genus *Rhinusa* comprises approximately 40 species following a Palaearctic distribution (Caldara 2001) and feeding on species within the plant genera *Verbascum* and *Scrophularia* (Scrophulariaceae) and *Linaria, Kickxia, Chaenorhinum, Antirrhinum* and *Misopates* (Plantaginaceae) (Caldara *et al.* 2010). *Gymnetron* includes approximately 30 species with a Palaearctic distribution and approximately 60 species from the Afrotropical region, of which 55 are known mainly from South Africa and considered to be endemic to this area (Caldara 2003). All Palaearctic *Gymnetron* use species from the genus *Veronica* (Plantaginaceae) as host plants, whereas representatives from the Afrotropical region use different host plants within the family Scrophulariaceae (*Hebenstreitia, Sutera, Selago, Buddleja, Diascia, Nemesia, Hemimeris*) and the genus *Anastrebe* in the family Stilbaceae (Caldara *et al.* 2008). Recent molecular phylogenetic analyses have revealed *Rhinusa* and *Gymnetron* not to be reciprocally monophyletic, with *Rhinusa* being a paraphyletic assemblage derived from within *Gymnetron* (Chapter 3).

The use of plant tissues by *Rhinusa* and *Gymnetron* for the completion of their reproductive cycle suggests an intimate relationship with their hosts, thus representing an excellent opportunity to study insect-host plant associations in the context of speciation and diversification. At the intraspecific level mitochondrial and nuclear DNA sequence data has revealed cryptic host-associated diversity within the species *Rhinusa antirrhini* (Hernandez-Vera *et al.* 2010). Six mitochondrial lineages

were found to be associated with species and subspecies within the genus *Linaria*, suggesting host specialization as a likely driver for diversification. Caldara et al. (2010) suggest that *Rhinusa* species typically exhibit host conservatism at the plant family level and perhaps at the plant genus level for some *Rhinusa* species. Similarly, *Gymnetron* species appear to show host conservatism at the plant family level, as Palaearctic species have been reported to parasitize only species of the genus *Veronica* within the family Plantaginaceae, whereas Afrotropical species utilize representatives within the family Scrophulariaceae (Caldara *et al.* 2008). However, the extent to which host plant use is phylogenetically conserved across species within both genera remains to be explored quantitatively. In addition to exploiting a range of host plant species, *Rhinusa* and *Gymnetron* also exploit a range of host plant tissues, including ovaries, stems, roots, and galls induced by other species from the same genera. As an example, *R. antirrhini, R. linariae*, and *R. pilosa* use the same host plant species, *Linaria vulgaris*, with each species exploiting a different plant structure. *Rhinusa antirrhini* feeds and develops inside fruit capsules, whereas *R. linariae* and *R. pilosa* are both gall inducers, the former utilizes roots and the latter stems. A further level of host use is present with two other species, *R. collina* and *R. eversmanni*, acting as inquilines of the galls induced by *R. linariae* and *R. pilosa* respectively. Thus, it seems reasonable to hypothesize that this high degree of specialization by exploiting different plant resources has contributed to the diversification of the group.

Using mitochondrial and nuclear DNA sequence data, we reconstruct phylogenetic relationships of representative taxa within *Rhinusa* and *Gymnetron*, with a specific focus on species of *Rhinusa*, with the following four aims: (1) assess taxonomic

conservatism in host-plant use at the plant family and genus level across species; (2) determine whether shifts in parasitic mode have contributed to diversification of the group; (3) estimate ancestral states of host utilization and modes of parasitism; and (4) examine the generality of host specificity within *R. neta*, *R. antirrhini* and *R. vestita*, three seed parasitic species widely distributed in Europe parasitizing a range of plant species within the genera *Linaria* (*R. neta*, *R. antirrhini*) and *Antirrhinum* (*R. vestita*). Additionally, the systematics of *Rhinusa* is revised, testing the monophyly of morphologically-based taxonomic groups proposed by Caldara et al. (2010).

## 4.2. MATERIALS AND METHODS

### 4.2.1. Taxon sampling and host-plant information

For this study, the set of samples used in Chapter 3 (65 and 38 individuals of *Rhinusa* and *Gymnetron* respectively, of which 48 were sequenced for the 16S SPIA's only) was used to address aims 1 – 3 regarding host-plant use and the systematics of *Rhinusa*. To achieve aim 4, a subset of samples of the former dataset including only individuals of *R. neta* (4), *R. vestita* (3) and *R. antirrhini* (8) was expanded to increase the number of individuals and host plant records for these species. A total of 31, 32 and 38 individuals of *R. neta*, *R. vestita*, and *R. antirrhini* respectively, were sampled. Host plant associations were determined by direct observation of emerging weevils from host plants in the field, records from published literature (Caldara 2001; Caldara 2003; Caldara *et al.* 2010), personal communications of two collaborators with extensive field-work experience, and

non-published studies of host plant use and host preferences in which individuals were reared and monitored through adult emergences (Gassmann & Paetel 1998; Groppe 1992; Tosevski *et al.* 2005).

### 4.2.2. DNA extraction, PCR amplification and sequencing reactions

The newly incorporated samples of *R. neta*, *R. vestita*, and *R. antirrhini* were punctured through the abdomen and total genomic DNA was extracted from each weevil according to the protocol included in the DNeasy extraction Kit (QIAGEN). After DNA extraction, specimens were placed again in 96% ethanol and kept at 4°C as voucher specimens. Samples for this dataset were PCR-amplified and sequenced only for a subset of the genes from chapter 3 (COII and AK genes) as the sequence information provided by these genes is sufficient to recover the same phylogenetic relationships recovered with the 5 gene partition data set. Amplification conditions are as in Chapter 3.

### 4.2.3. Sequences alignment and phylogenetic analyses

The newly obtained COII and AK sequences were automatically aligned using the CLUSTAL W algorithm as implemented in BioEdit version 7.0.9 (Hall 1999) with further manual alignment. COII and AK sequences were combined into a dataset of concatenated sequences partitioned by gene, 696 bp for COII and 715 bp for AK. Phylogenetic analyses were performed for both, the COII+AK dataset and the 5-gene partition dataset from chapter 3. Bayesian analyses were performed with the parallel version of MRBAYES v3.1 (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 the COII+AK dataset were as follows: two simultaneous runs with two Markov chains each (4 chains in total) for 10 million generations sampling every 1000 generations, a heating parameter value of 0.02 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. The same settings were used for the 5 gene partition dataset except that the number of chains and generations was increased to optimize convergence for large datasets as suggested by Ronquist et al. (2005). Two independent runs with four Markov chains each (8 chains in total) and 20 million generations were applied. Maximum likelihood analyses were performed with the parallel version of RAxML 7.0.4 (Stamatakis 2006). For both datasets, 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.

### 4.2.4. Assessing phylogenetic conservatism of host plant use and mode of parasitism

To assess phylogenetic conservatism of host–plant use at the family and genus level, Bayes factors were used to compare the likelihoods of models where constraints of monophyly were imposed on species that utilize host species from the same plant family and species that utilize host plant species from the same genus, against

models with no topological constraints enforced. Host-plant families and genera are presented in Table 1. Similarly, conservatism of mode of parasitism was assessed by imposing constraints of monophyly on weevil species that exhibit the same parasitic mode, and the likelihoods of these models were compared against an inferred phylogeny without any constraints. The following modes of parasitism were considered according to the plant tissue/structure used for feeding and developing: i) seed capsule, ii) stem, iii) root gall inducer, iv) stem gall inducer, v) root gall inquiline, vi) stem gall inquilines (Table 2). This approach is similar to that followed by Borghuis et al. (2009) and Kato et al. (2010) where different hypotheses concerning conservatism of ecological traits in beetles were tested. Within a Bayesian framework, Bayes factors have been adopted as a useful hypothesis-testing tool (Nylander *et al.* 2004; Sperling *et al.* 2009; Suchard *et al.* 2001; Suchard *et al.* 2005). If two hypotheses are equally likely a priori, then the Bayes factor quantifies the relative support of the competing hypotheses given the observed data (Kass & Raftery 1995; Suchard *et al.* 2005). The harmonic means of sampled likelihoods were estimated using the *sump* command in MrBayes and Bayes factors were estimated as twice the difference in the natural log of the harmonic mean of model likelihoods of each model ($2\Delta\ln$ HML). Values were interpreted according to the guidelines proposed by Kass and Raftery (1995).

### 4.2.5. Ancestral state estimation of host plant use and mode of parasitism

Ancestral states of host use were estimated under a maximum likelihood (ML) framework using Mesquite version 2.74 (Maddison & Maddison 2010). As opposed to maximum parsimony optimisations, likelihood-based optimisations take into account branch lengths (if branch lengths are time-proportional more changes are

expected on long branches) and they allow the assessment of uncertainty in ancestral state reconstruction (Pagel 1999; Schluter *et al.* 1997). Host plant utilization was categorized at the family level into the two known host plant families used by this group of weevils, Scrophulariaceae and Plantaginaceae, as recently circumscribed (Albach *et al.* 2005; Olmstead *et al.* 2001). Ancestral state reconstructions for host plant utilization at the genus level and mode of parasitism were not performed given that the Bayes factor tests support phylogenetic conservatism only in two cases for each of these ecological traits (see Results). Thus, we cannot reasonably use the phylogeny to reconstruct the evolution of traits that have no relationship to the tree, as the reconstruction would be highly uncertain (Schluter *et al.* 1997).

### 4.2.6. Assessing monophyly of taxonomic groups

The same approach for the assessment of phylogenetic conservatism of host-plant use and mode of parasitism was applied to assess the monophyly of taxonomic groups based on morphological characters. The likelihoods of models with constraints of monophyly imposed on taxonomic groups of weevils as proposed by Caldara et al. (2010) were each compared against a model with no topological constraint enforced. This approach was used particularly to test those groups with low or not support from Bayesian and ML analyses, as these support values can be affected by the presence of short DNA sequences within the data matrix (chapter 2). Ten species groups belonging to three more inclusive species assemblages have been proposed. The first assemblage (A) includes nine species within the *Rhinusa bipustulata* and *R. tetra* groups, the second assemblage (B) includes 14 species within the *R. antirrhini* and *R. linariae* groups, whereas the third assemblage (C)

comprises 14 species included in six groups: *R. pilosa*, *R. herbarum*, *R. neta*, *R. vestita*, *R. mauritii* and *R. melas*.

## 4.3. RESULTS

### 4.3.1. Phylogenetic conservatism of host plant use and mode of parasitism

Constrained models hypothesizing phylogenetic conservatism of host plant family use provide a better fit than the model without constraints. According to the guidelines of Kass and Raftery (1995), there is strong support for the constrained models on groups of weevils affiliated to Scrophulariaceae and Plantaginaceae (ln Bayes factor > 10 in both cases). Phylogenetic conservatism of host plant use at the genus level is supported in two instances only. Constrained models for species of weevil affiliated to the genera *Diascia* and *Veronica* exhibit ln Bayes factors > 10. Results of the Bayes factor tests for phylogenetic conservatism of host-plant use are summarized in Table 1. Similarly, support for phylogenetic conservatism of mode of parasitism was found in two cases only. Models with constraints of monophyly for species of weevil parasitizing host plant stems, and species that are inquilines of stem galls, are significantly better than the model without constraints (ln Bayes factor > 10) (Table 2).

### 4.3.2. Ancestral character states

Ancestral state reconstruction of host plant utilization at the family level is shown in Figure 1. Results from the maximum likelihood character optimization provide significant support for the family Scrophulariaceae as the ancestral character state of

host-plant use by *Gymnetron* and *Rhinusa*. Proportional likelihoods for this character state are equal or higher than 95% in two of the most basal nodes of the tree (Fig. 1).

### 4.3.3. Molecular phylogenetic analysis of *Rhinusa neta, R. antirrhini* and *R. vestita*

To assess host-associated genetic structure at the intraspecific level, phylogenetic relationships were reconstructed for the data set including the seed parasitic weevils *R. neta*, *R. antirrhini* and *R. vestita*. Bayesian and ML analyses recovered three well supported monophyletic groups according to weevil species (PP = 1, BS > 95). In accordance with previous results (Hernandez-Vera *et al.* 2010) a clear pattern of host-associated genetic structure is observed for *R. antirrhini* with four monophyletic groups associated with different host plants; 1) *Linaria vulgaris*, 2) *L. rubioides*, 3) *L. genistifolia artvinensis*, and 4) the species complex *L. genistifolia* / *L. dalmatica macedonica*. A fifth monophyletic group includes samples affiliated with three different host plant taxa: *L. genistifolia polyclada*, *L. kurdica* and *L. grandiflora* (Fig. 2). Within the *R. vestita* clade (A), there is no evident pattern of host-associated genetic structure; two monophyletic groups were recovered, one with moderate support values (PP $\geq$ 0.8, BS $\geq$ 70) which includes samples associated with the host plants *Antirrhinum latifolium* and *A. majus pseudomajus*, and the second one with high support values (PP=1.0, BS=98) includes samples associated with the host plants *A. latifolium, A. majus striatum* and *A. boissieri*. In a more basal position within the *R. vestita* clade, there is an assemblage of samples associated with the host plants *A. boissieri*, *A. hispanicum*, *A. molle*, *A. majus pseudomajus*, *A. majus cirrhigerum*, *A. graniticum*, and *A. litigiosum*. Within the *R. neta* clade (B),

two monophyletic groups with high support values (PP = 1.0, BS > 90) were recovered; one includes samples associated with the host plants *Linaria dalmatica*, *L. genistifolia*, *L. vulgaris*, *L. alpina*, *L. rubioides*, and *L. angustissima*, and the second one includes three samples with uncertain taxonomic identification based on morphological characters; these individuals are associated with the host plant species *L. kurdica* (group 6 within clade B) (Fig. 2).

### 4.3.4. Molecular systematics

Both Bayesian and ML analyses recovered essentially the same phylogenetic relationships, largely in agreement with the morphology-based taxonomic groups proposed by Caldara et al. (2010), except that *Gymnetron* and *Rhinusa* are not reciprocally monophyletic (Fig. 3). The results support the monophyly of the major *Rhinusa* assemblages and the monophyly of the constituent species groupings within them as defined by Caldara et al. (2010). Low support values were obtained from Bayesian and ML analyses for clade A, however, results from the Bayes factor test strongly support the monophyly of this group (ln Bayes factor > 10). Six *Gymnetron* species (*G. bipartitum*, *G. minimum*, *G. pauxillum*, *G. quadratum*, *G. piceum* and *G. agile*) are placed as basal lineages to this assemblage (Fig. 3). Monophyly of this more inclusive clade is also supported by Bayes factor tests as reported in chapter 3. Relatively high support values were obtained for assemblage B, posterior probability (PP) = 0.92 and bootstrap value (BS) = 75. *Rhinusa herbarum* and *R. mateui*, are placed as a basal lineage to this assemblage although with no support from Bayesian and ML analyses (see below). Initially, assemblage C was not recovered as monophyletic from Bayesian and ML analyses. Three of the 6 species groups included in this assemblage based on morphological characters fall outside this

clade, although with very low support values (PP < 0.7, BS < 60). The *R. herbarum* group (including *R. herbarum* and *R. mateui*) was placed as a basal lineage to assemblage B, whereas the *R. pilosa* (including *R. brondelii* and *R. pilosa*) and *R. melas* groups form a monophyletic group placed as a sister lineage to assemblages B and C (Fig. 3). However, when the species groups comprising this assemblage (according to Caldara et al. [2010]) were constrained to be monophyletic, the Bayes factor test strongly supports its monophyly. A fourth major assemblage D, including only *Gymnetron* species, was recovered as monophyletic and sister to assemblage A (ln Bayes factor > 10).

Likewise, results support the monophyly of the species groups comprising the four major assemblages. Within assemblage A, the *R. tetra* group is supported as monophyletic with a PP = 0.97 and BS = 71, comprising the species *R. moroderi*, *R. tetra*, *R. comosa*, *R. verbasci*, and *R. asellus*. Low support values from Bayesian and ML analyses were obtained for the *R. bipustulata* group, however the Bayes factor test support the monophyly of the group (ln Bayes factor > 10) which includes the species *R. bipustulata*, *R. scrophulariae*, *R. algirica*, and *R. emmrichi* (the three latter archival specimens). Assemblage B includes the *R. antirrhini* (PP= 0.93, BS = 72) and *R. linariae* (PP= 0.90, BS = 75) groups. The former includes the species *R. antirrhini*, *R. dieckmanii*, *R. florum*, and *R. exigua* forming a monophyletic group with moderate support values (PP = 0.85, BS = 78), and *R. griseohirta*, which is placed as a basal lineage to this clade. The *R. linariae* group (PP = 0.9, BS = 75) includes the sister species *R. linariae* and *R. brisouti*, the latter placed as a more divergent lineage. Within assemblage C, the *R. neta* and *R. vestita* groups form a monophyletic clade with high support from the Bayesian analysis (PP = 0.99) and

moderate support from the ML analysis (BS = 70). The *R. neta* group includes the species *R. canescens, R. collina, R. neta, R. eversmanni* and *R. mauritii*. Moderate and very low support values were obtained for this group from Bayesian and ML analyses (PP = 0.80, BS < 70), however, results from the Bayes factor test strongly support monophyly (ln Bayes factor > 10). Similarly, the *R. vestita* group which includes the species *R. vestita* and *R. depressa*, exhibits very low support values from Bayesian and ML analyses but the monophyly of this group is supported by the Bayes factor test (ln Bayes factor > 10). *Gymnetron histrix* is placed as a more divergent sister lineage to the *R. neta* and *R. vestita* groups. As previously mentioned, the *R. herbarum, R. pilosa* and *R. melas* groups were not recovered as monophyletic within assemblage C, however constraining these groups to this major assemblage significantly improved the likelihood of the model. The *R. herbarum* group includes the species *R. herbarum* and *R. mateui* (PP = 0.96, BS < 70). The *R. pilosa* group (PP = 1, BS = 96) includes the sister species *R. brondelii* and *R. pilosa*, placed as a sister lineage to the monospecific group *R. melas*. Assemblage D includes two monophyletic species groups, the *G. melanarium* (PP = 0.88, ln Bayes factor > 10) and *G. rotundicolle* groups (ln Bayes factor > 10). The first one includes nine *Gymnetron* species (*G. stimulosum, G. melanarium, G. linkei, G. desbrochersi, G. veroniceae, G. aequale, G. tibielum, G. vitipenne, G. villosulum*) with unresolved basal relationships and high support values (PP = 1.0, BS = 100) for three terminal clades. The second one includes three *Gymnetron* species; *G. rotundicolle* and *G. rostellum* are placed as sister lineages and *G. clepsydra* is placed as a more divergent basal lineage within this group.

## 4.4. DISCUSION

### 4.4.1. Phylogenetic conservatism of host-plant use

The results provide evidence of phylogenetic conservatism of host plant use at the family level, whereas at the genus level only two instances of conservatism are supported. Since the publication of Ehrlich and Raven's (1964) influential paper, it has been recognized that host-plant chemistry (secondary compounds) and insect dietary tolerances play a significant role in shaping insect-host plant associations. Although examples of generalist plant-feeding insects exist (e.g. Bernays & Minkenberg 1997; Ribeiro *et al.* 2005), the most commonly observed pattern is that most species are restricted to a few closely related plant taxa (Futuyma 1983; Jaenike 1990; Schoonhoven *et al.* 2005) where related insects tend to feed/specialize on related groups of plants (Ehrlich & Raven 1964; Janz & Nylin 1998). However, although chemical similarity of plant taxa is largely correlated with their phylogenetic relationships, this correlation is not perfect and it has been shown that host shifts of related insects are more strongly correlated with plant chemistry than plant phylogeny. For instance, in a study of the leaf beetle genus *Blepharida*, Becerra (1997), demonstrated that historical patterns of host shifts strongly correspond to the pattern of their host chemical similarities, concluding that host plant chemistry has a greater influence in the evolution of host use than host plant phylogeny in this group of beetles. Similarly, in a study of evolutionary history of host-plant use in butterflies in the tribe Melitaeini, Wahlberg (2001) found evidence that plant chemistry is a more conservative trait than plant taxonomy. Thus, the absence of phylogenetic conservatism in host use at the genus level in *Rhinusa* and

*Gymnetron* species might be explained by potentially frequent host shifts given the chemical similarities in the secondary compounds of their host plants at this taxonomic level.

In contrast at a higher taxonomic level such as plant family, plants may exhibit important differences in their secondary compounds, making host shifts more difficult to occur due to the inability of insects to metabolize significantly different secondary compounds. This pattern has been observed in different groups of insects including Lepidoptera (Janz & Nylin 1998), and Coleoptera (Becerra & Venable 1999) among others. In a compilation of phylogenetic studies of phytophagous insects, Winkler and Mitter (2008) show that approximately only 8% of speciation events included a host shift to a different plant family. Nonetheless, there is also evidence of host shifts between more distantly related plants; however, these cases usually involve host plant families bearing similar secondary compounds (e. g. Berenbaum 2001; Zakharov *et al.* 2004). Other possible causes of "major" host shifts in phytophagous insects have been proposed, including feeding strategies (Nyman *et al.* 2006) and properties of plant taxa and/or communities (Mitter & Farrell 1991 and references therein).

### 4.4.2. Phylogenetic conservatism of mode of parasitism

According to the Bayes factor tests only two modes of parasitism exhibit phylogenetic conservatism; weevils parasitizing stems and inquiline weevils using stem galls produced by another weevil. The absence of a more general pattern suggests that the mode of parasitism does not impose significant constraints on host plant utilization, representing a labile ecological trait. This result might seem

counterintuitive given the intimate relationship of this group of weevils with their hosts, feeding and developing inside plant tissues, however similar findings have been reported for other insects exhibiting close relationships with their hosts. For instance, Cook et al. (2002) found evidence that shifts between host-plant organs can occur at a significantly greater rate than shifts between host oak sections in oak gallwasps. Similarly, in a phylogenetic study, Joy et al. (2007) showed that gall-inducing flies within the genus *Asphondylia* have undergone numerous shifts between different plant organs of the same host-plant species. In both cases, the partitioning of ecological resources has been put forward as a likely driver for diversification.

In weevils, it has been shown that diversification has been accompanied by niche shifts in host-plant associations and larval habits (Marvaldi *et al.* 2002). Our results indicate that mode of parasitism is not a conservative trait and shifts in parasitic mode have contributed to the diversification of this group of weevils. Different modes of parasitism provide alternative ecological axis along which *Gymnetron* and *Rhinusa* can diversify and reduce potential interspecific competition. A clear example of this is illustrated by five species of *Rhinusa* exploiting different resources within the same host plant species *Linaria vulgaris*. *Rhinusa antirrhini* feeds and develops inside fruit capsules, whereas *R. linariae* and *R. pilosa* are both gall inducers, the former utilizes roots and the latter stems. A further level of resource partitioning is present with two other species, *R. collina* and *R. eversmanni*, acting as inquilines of the galls induced by *R. linariae* and *R. pilosa* respectively.

### 4.4.3. Ancestral character states

Results from the maximum likelihood character optimization suggest that the ancestral condition for host utilization at the plant family level in *Rhinusa* and *Gymnetron*, was a specialist of host plants within the family Scrophulariaceae. This result is consistent with the hypothesis put forward in chapter 3 of a South African origin for this species complex, given that the predominant concentration of genera within this plant family is distributed in the southern hemisphere, particularly Africa (Olmstead *et al.* 2001). Also in accordance with this is the age of the family, which is older than Plantaginaceae. In a recent comprehensive analysis of divergence times across the angiosperms, Bell et al. (2010) estimate the age of the family to be 51 – 53 My whereas the estimated age for Plantaginaceae is 42 – 46 My. Thus it seems reasonable to think of a scenario where weevils initially specialized on plants within Scrophulariaceae and then shifted to a new set of plants closely related to the ancestral ones (Albach *et al.* 2005), a colonization event which very likely facilitated the diversification of the group, exploiting and adapting to the newly opened ecological niches.

### 4.4.4. Phylogenetic analysis of *Rhinusa neta, R. antirrhini* and *R. vestita*

In chapter 2, it was demonstrated that weevils with a high degree of ecological specialization by feeding and developing inside fruit capsules may experience ecological divergence and host-associated genetic differentiation. Here, the generality of this process was assessed by using other weevil species exhibiting these ecological characteristics. Bayesian and ML analyses recovered three well supported clades corresponding to the three *Rhinusa* species. In agreement with previous results (Hernandez-Vera *et al.* 2010), the *R. antirrhini* clade exhibits a clear pattern of host-associated genetic differentiation. Clades 1 – 4 were recovered as

monophyletic groups with very high support values (PP=1.0, BS=100), corresponding to 4 of the 7 host-associated mitochondrial lineages described in chapter 2. Clade 5 also exhibits very high support values (PP=1.0, BS=98), and similar to lineage 4, includes samples associated with three different host plant taxa. However, more samples will be necessary to determine if this is a single lineage exhibiting a more generalist behaviour or represents several lineages, each one associated with a single host plant taxon. Within the *R. neta* clade the well supported divergent group of samples associated with the host plant *L. kurdica* (clade 6) suggests cryptic diversity associated to host plant use, however, this lineage could also represent a different species given the uncertain taxonomic identity of the samples within this clade. In contrast, there is no clear evidence of phylogenetic structure associated with host plant use within the *R. vestita* clade. Two monophyletic groups were recovered within this clade, one with high support values (PP = 1.0, BS = 98) and the other one with moderate support values (PP=8.0, BS=70), however these are not correlated with different host plant taxa; most of the samples within these groups share the same host *Antirrhinum latifolium*. Although further studies with expanded sampling and more variable markers will be necessary, one clear example of host-associated genetic differentiation out of three different species suggest that host-associated genetic differentiation might be a recurrent phenomenon within endophagous parasitic weevils exhibiting an intimate relationship with their hosts.

### 4.4.5. Molecular systematics

In his taxonomic revision of the tribe Mecinini, Caldara (2001) acknowledges the challenging systematics of the group given the few morphologically informative

characters present. Despite the recognition of just a few subtle apomorphies between *Gymnetron* and *Rhinusa*, he concludes that they should be considered as separate genera. Although largely in agreement with the proposed taxonomy for *Rhinusa* based on morphological characters (Caldara *et al.* 2010), our results indicate that *Gymnetron* and *Rhinusa* are not reciprocally monophyletic; *Gymnetron* species exhibit deep genetic divergences but they form monophyletic groups with *Rhinusa* lineages.

Our results support the monophyly of both the species groups and the more inclusive major assemblages as proposed by Caldara et al. (2010), however, some differences were found, particularly within assemblage C, a clade which is weakly supported as monophyletic in the morphology-based taxonomy. Initially, the Bayesian and ML analyses did not recover this assemblage as monophyletic, however when the species groups comprising this major clade were enforced to be monophyletic, the likelihood of the model increased significantly. Thus, the results support the monophyly of the assemblage, comprised by the six species groups as previously proposed by Caldara et al. (2010): *R. pilosa*, *R. herbarum*, *R. neta*, *R. vestita*, *R. mauritii* and *R. melas*. With regard to the last three species groups, Caldara et al. (2010) consider the phylogenetic placement of their constituent species as difficult to ascertain. The *R. vestita* group includes the species *R. vestita* and *R. depressa*, whereas the *R. mauritii* and *R. melas* groups, both are monospecific groups, i. e. composed of the single species which gives the name to the group. *Rhinusa vestita*, and *R. depressa* are considered as closely related with significant morphological similarities; *R. mauritii* is placed as a sister lineage to the *R. vestita* and *R. neta* groups, whereas *R. melas* is placed as sister to the *R. mauritii* + *R. vestita* + *R. neta* clade. Our results support the monophyly of the *R. vestita* group, with *R. depressa*

and *R. vestita* as sister lineages and *R. mauritii* as sister lineage to the clade including *R. eversmanni* + *R. cf. neta* within the *R. neta* group. *Rhinusa melas* is placed as a sister lineage to the *R. pilosa* group. *Gymnetron histrix* is placed as a sister lineage to the *R. neta* and *R. vestita* groups; this clade is supported as monophyletic by the Bayes factor test.

Low support values were obtained for assemblage A, however its monophyly is supported by the Bayes factor test. This assemblage is composed of the *R. tetra* and *R. bipustulata* groups. In agreement with the morphology-based taxonomy, it is noteworthy the placement of archival samples within these two species groups. *Rhinusa moroderi* is placed as a sister lineage to *R. tetra* and *R. comosa* is placed as a sister lineage to *R. verbasci* within the *R. tetra* group. Similarly, archival specimens of *Rhinusa scrophulariae*, *R. algirica* and *R. emmirichi* are correctly placed within the *R. bipustulata* group. Monophyly of the more inclusive clade comprising assemblage A and the six *Gymnetron* species placed as sister lineages to the former is also supported by the Bayes factor test.

Within assemblage B the concordance with the morphology-based taxonomy is also noteworthy for the placement of the archival sample *R. brisouti* as a sister lineage to *R. linariae*. Both species form the monophyletic group *R. linariae* with relatively high support (PP = 0.9, BS = 75). *Rhinusa exigua* represents a newly described species from Turkey (Caldara & Korotyaev 2010), therefore it was not included in Caldara's taxonomic revision of the genus. Based on morphological characters, Caldara and Korotyaev (2010), describe the species as very closely related to *R. antirrhini*. Our results place this species as a sister lineage to the *R. antirrhini* + *R. florum* clade; moderate support values were obtain for the monophyly of these three lineages (PP = 0.85, BS = 78).

Assemblage D, including only *Gymnetron* species, was supported as monophyletic, very low support values were obtained from both Bayesian and ML analyses, however constraining the species within this assemblage to be monophyletic, the likelihood of the tree was increased significantly (ln Bayes factor > 10). Similarly, the Bayes factor test support the monophyly of the two species groups within the assemblage.

## 4.5. CONCLUSION

Our results suggest a scenario where weevils are restricted to large sets of plants (family level) most likely because of substantial differences in their chemical composition, however, a more dynamic process occurs at lower taxonomic levels, where plant taxa exhibit similar secondary compounds and potentially allow for more frequent host shifts. Similarly, the mode of parasitism represents a labile ecological trait, adding another important dimension of host use: a finer partitioning of resources by utilizing different plant tissue/organs within single host species, a phenomenon which would explain in part the successful diversification of this group of weevils. Host-associated genetic differentiation might be a recurrent phenomenon within weevils exhibiting an intimate relationship with their hosts.

The use of molecular data allowed for the resolution of phylogenetic uncertainties in a challenging group in terms of taxonomic classification given the few morphologically informative characters present. The use of short DNA sequences (16S SPIAs) from archival specimens allowed for the augmention of the scope of the study proving useful in placing samples within the phylogeny with a high degree of confidence in some instances. Although *Gymnetron* species exhibit deep genetic

divergences, *Rhinusa* and *Gymnetron* represent lineages with a shared evolutionary history.

**Table 1.** Results of Bayes factors tests to assess phylogenetic conservatism of host-plant use at the family and genus level. Log-Bayes factors are calculated as twice the difference in the natural log of the harmonic mean of model likelihoods of constrained and unconstrained models (2Δln HML). Values were interpreted according to the guidelines proposed by Kass and Raftery (1995): 0-2 weak evidence, 2-6 positive evidence, 6-10 strong evidence, >10 very strong evidence.

| Host-plant family and genus | Harmonic mean | log-Bayes Factor |
|---|---|---|
| Scrophulariaceae | -20936.53 | 126.24 |
| *Diascia* | -21124.64 | 22.58 |
| *Scrophularia* | -21243.19 | -214.52 |
| *Selago* | -21156.97 | -42.08 |
| *Verbascum* | -21157.92 | -43.98 |
| | | |
| Plantaginaceae | -20991.29 | 16.72 |
| *Antirrhinum* | -21431.74 | -591.62 |
| *Chaenorrhinum* | -21136.32 | -0.78 |
| *Kickxia* | -21136.33 | -0.8 |
| *Linaria* | -21417.37 | -509.66 |
| *Veronica* | -21126.47 | 18.92 |

**Table 2.** Results of Bayes factors tests to assess phylogenetic conservatism of mode of parasitism. Log-Bayes factors are calculated as twice the difference in the natural log of the harmonic mean of model likelihoods of constrained and unconstrained models (2Δln HML). Values were interpreted according to the guidelines proposed by Kass and Raftery (1995): 0-2 weak evidence, 2-6 positive evidence, 6-10 strong evidence, >10 very strong evidence.

| Mode of parasitism | Harmonic mean | log-Bayes Factor |
|---|---|---|
| Seeds | -22219.77 | -2167.68 |
| Stems | -21088.32 | 95.22 |
| Root gall inducer | -21151.39 | -30.92 |
| Stem gall inducer | -21137.75 | -3.64 |
| Root + Stem gall inducers | -21233.73 | -195.6 |
| Inquiline of stem galls | -21126.94 | 17.98 |
| Inquiline of root galls | -21149.26 | -26.66 |
| Constraint on both, inquilines of root + stem galls | -21269.73 | -267.6 |

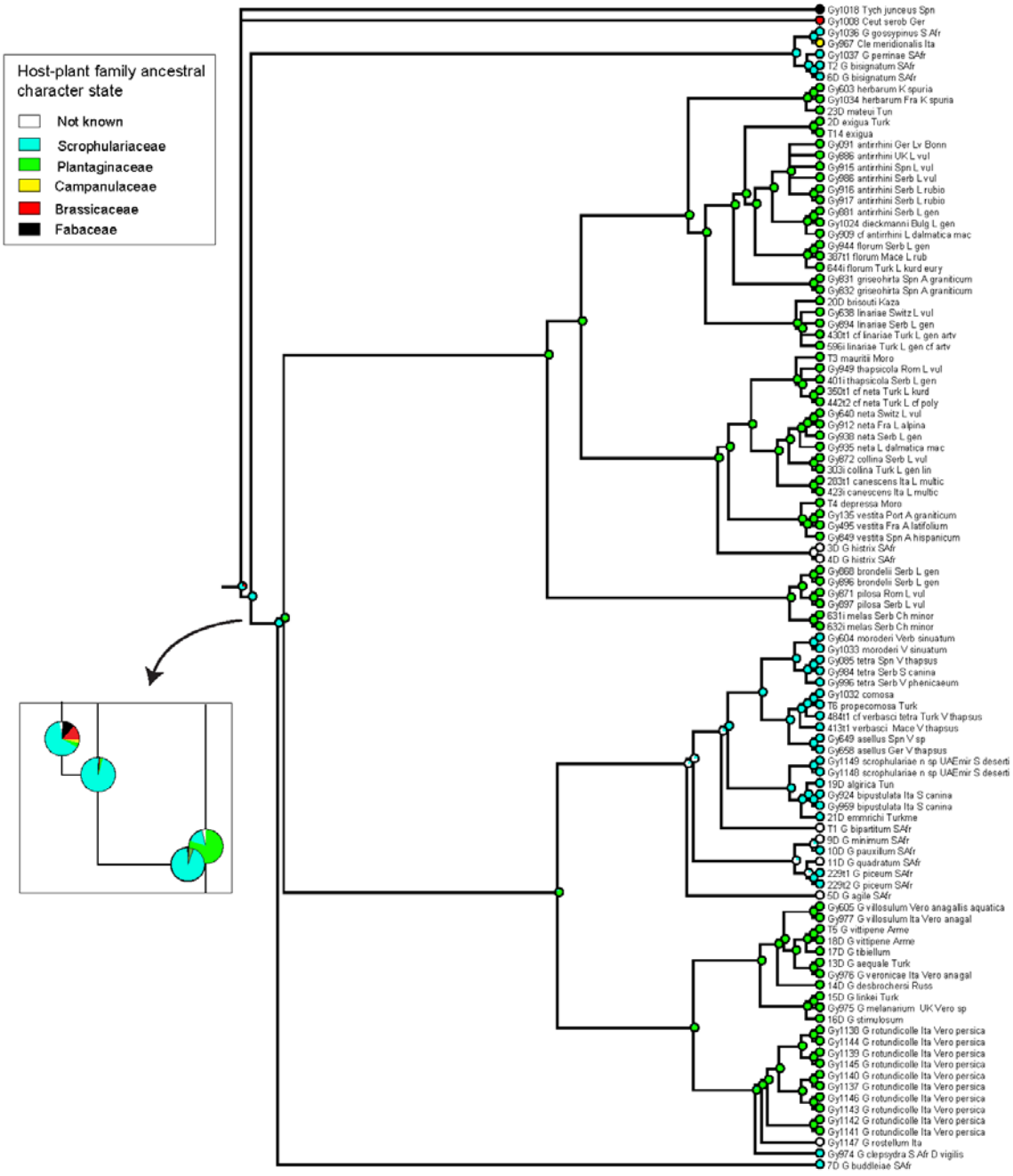


**Figure 1.** Ancestral character reconstruction for the evolution of host utilization at the plant family level in *Rhinusa* and *Gymnetron* species based on a Bayesian majority rule consensus tree using maximum likelihood optimization. Pie graphs represent proportional likelihoods of character states.

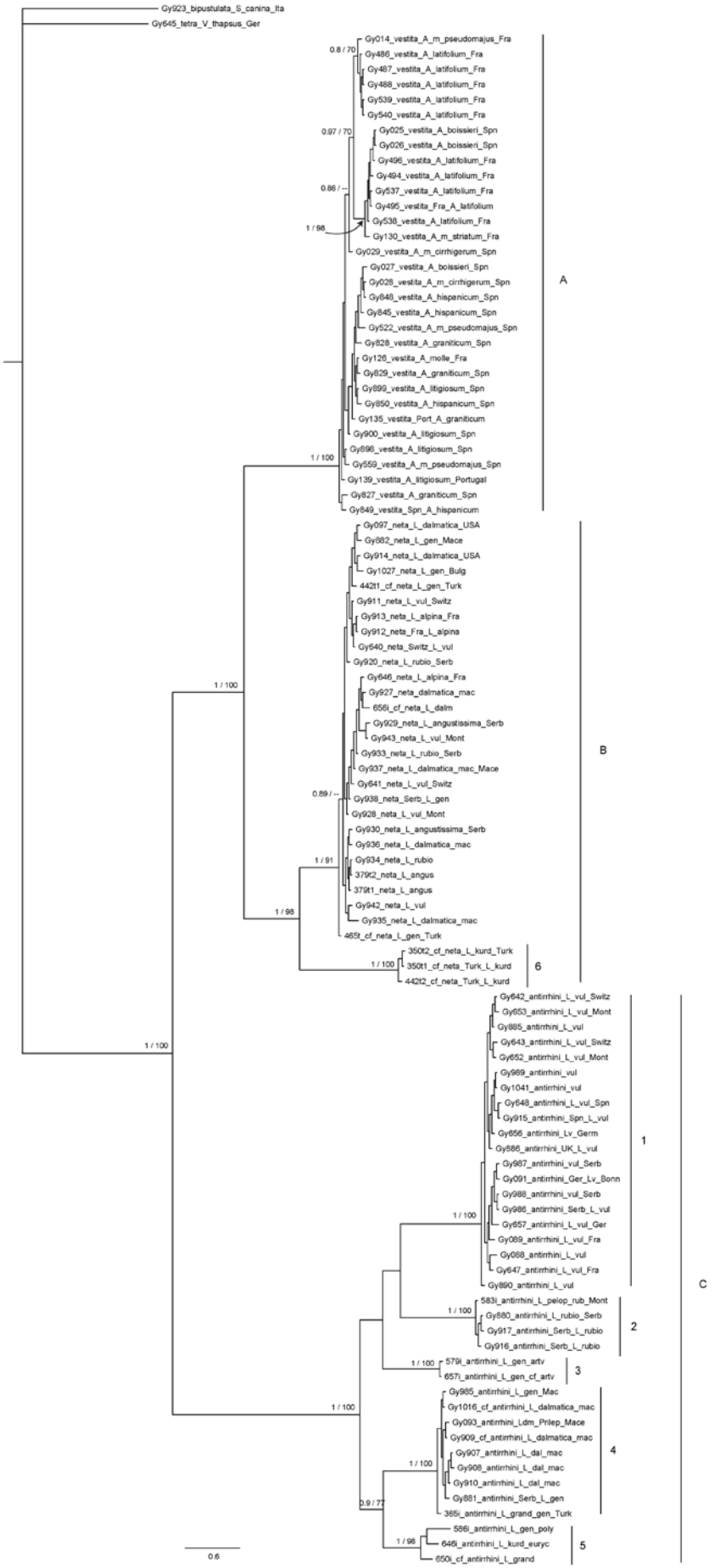
Gy923_bipustulata_S_canina_Ita
Gy645_tetra_V_thapsus_Ger
Gy014_vestita_A_m_pseudomajus_Fra
Gy486_vestita_A_latifolium_Fra
Gy487_vestita_A_latifolium_Fra
Gy488_vestita_A_latifolium_Fra
Gy539_vestita_A_latifolium_Fra
Gy540_vestita_A_latifolium_Fra
Gy025_vestita_A_boissieri_Spn
Gy026_vestita_A_boissieri_Spn
Gy496_vestita_A_latifolium_Fra
Gy494_vestita_A_latifolium_Fra
Gy537_vestita_A_latifolium_Fra
Gy495_vestita_Fra_A_latifolium
Gy538_vestita_A_latifolium_Fra
Gy130_vestita_A_m_striatum_Fra
Gy029_vestita_A_m_cirrhigerum_Spn
Gy027_vestita_A_boissieri_Spn
Gy028_vestita_A_m_cirrhigerum_Spn
Gy648_vestita_A_hispanicum_Spn
Gy845_vestita_A_hispanicum_Spn
Gy522_vestita_A_m_pseudomajus_Spn
Gy828_vestita_A_graniticum_Spn
Gy126_vestita_A_molle_Fra
Gy829_vestita_A_graniticum_Spn
Gy899_vestita_A_litigiosum_Spn
Gy850_vestita_A_hispanicum_Spn
Gy135_vestita_Port_A_graniticum
Gy900_vestita_A_litigiosum_Spn
Gy696_vestita_A_litigiosum_Spn
Gy559_vestita_A_m_pseudomajus_Spn
Gy139_vestita_A_litigiosum_Portugal
Gy827_vestita_A_graniticum_Spn
Gy849_vestita_Spn_A_hispanicum
A
Gy097_neta_L_dalmatica_USA
Gy882_neta_L_gen_Mace
Gy914_neta_L_dalmatica_USA
Gy1027_neta_L_gen_Bulg
442t1_cf_neta_L_gen_Turk
Gy911_neta_L_vul_Switz
Gy913_neta_L_alpina_Fra
Gy912_neta_Fra_L_alpina
Gy640_neta_Switz_L_vul
Gy920_neta_L_rubio_Serb
Gy646_neta_L_alpina_Fra
Gy927_neta_dalmatica_mac
656i_cf_neta_L_dalm
Gy929_neta_L_angustissima_Serb
Gy943_neta_L_vul_Mont
Gy933_neta_L_rubio_Serb
Gy937_neta_L_dalmatica_mac_Mace
Gy641_neta_L_vul_Switz
Gy938_neta_Serb_L_gen
Gy928_neta_L_vul_Mont
Gy930_neta_L_angustissima_Serb
Gy936_neta_L_dalmatica_mac
Gy934_neta_L_rubio
379t2_neta_L_angus
379t1_neta_L_angus
Gy942_neta_L_vul
Gy935_neta_L_dalmatica_mac
465t_cf_neta_L_gen_Turk
B
350t2_cf_neta_L_kurd_Turk
350t1_cf_neta_Turk_L_kurd
442t2_cf_neta_Turk_L_kurd
6
Gy642_antirrhini_L_vul_Switz
Gy653_antirrhini_L_vul_Mont
Gy885_antirrhini_L_vul
Gy643_antirrhini_L_vul_Switz
Gy652_antirrhini_L_vul_Mont
Gy989_antirrhini_vul
Gy1041_antirrhini_vul
Gy648_antirrhini_L_vul_Spn
Gy915_antirrhini_Spn_L_vul
Gy656_antirrhini_Lv_Germ
Gy886_antirrhini_UK_L_vul
Gy987_antirrhini_vul_Serb
Gy091_antirrhini_Ger_Lv_Bonn
Gy988_antirrhini_vul_Serb
Gy986_antirrhini_Serb_L_vul
Gy657_antirrhini_L_vul_Ger
Gy089_antirrhini_L_vul_Fra
Gy088_antirrhini_L_vul
Gy647_antirrhini_L_vul_Fra
Gy890_antirrhini_L_vul
1
583i_antirrhini_L_pelop_rub_Mont
Gy880_antirrhini_L_rubio_Serb
Gy917_antirrhini_Serb_L_rubio
Gy916_antirrhini_Serb_L_rubio
2
579i_antirrhini_L_gen_artv
657i_antirrhini_L_gen_cf_artv
3
Gy985_antirrhini_L_gen_Mac
Gy1016_cf_antirrhini_L_dalmatica_mac
Gy093_antirrhini_Ldm_Prilep_Mace
Gy909_cf_antirrhini_L_dalmatica_mac
Gy907_antirrhini_L_dal_mac
Gy908_antirrhini_L_dal_mac
Gy910_antirrhini_L_dal_mac
Gy881_antirrhini_Serb_L_gen
365i_antirrhini_L_grand_gen_Turk
4
586i_antirrhini_L_gen_poly
646i_antirrhini_L_kurd_euryc
650i_cf_antirrhini_L_grand
5
C
0.8 / 70
0.97 / 70
0.86 / --
1 / 98
1 / 100
1 / 100
0.89 / --
1 / 91
1 / 98
1 / 100
1 / 100
1 / 100
1 / 100
1 / 100
1 / 100
0.9 / 77
1 / 98
0.6

**Figure 2.** Bayesian phylogenetic tree inferred from a concatenated data set of 2 gene DNA sequences (COII and AK) comprising 1411 bp from 101 individuals of three species of *Rhinusa*; *R. vestita* (A), *R. neta* (B), and *R. antirrhini* (C). Bayesian posterior probabilities≥ 0.8 and ML bootstrap values ≥ 70% are shown above branches in that order. Numerals indicate host-associated lineages (see text).

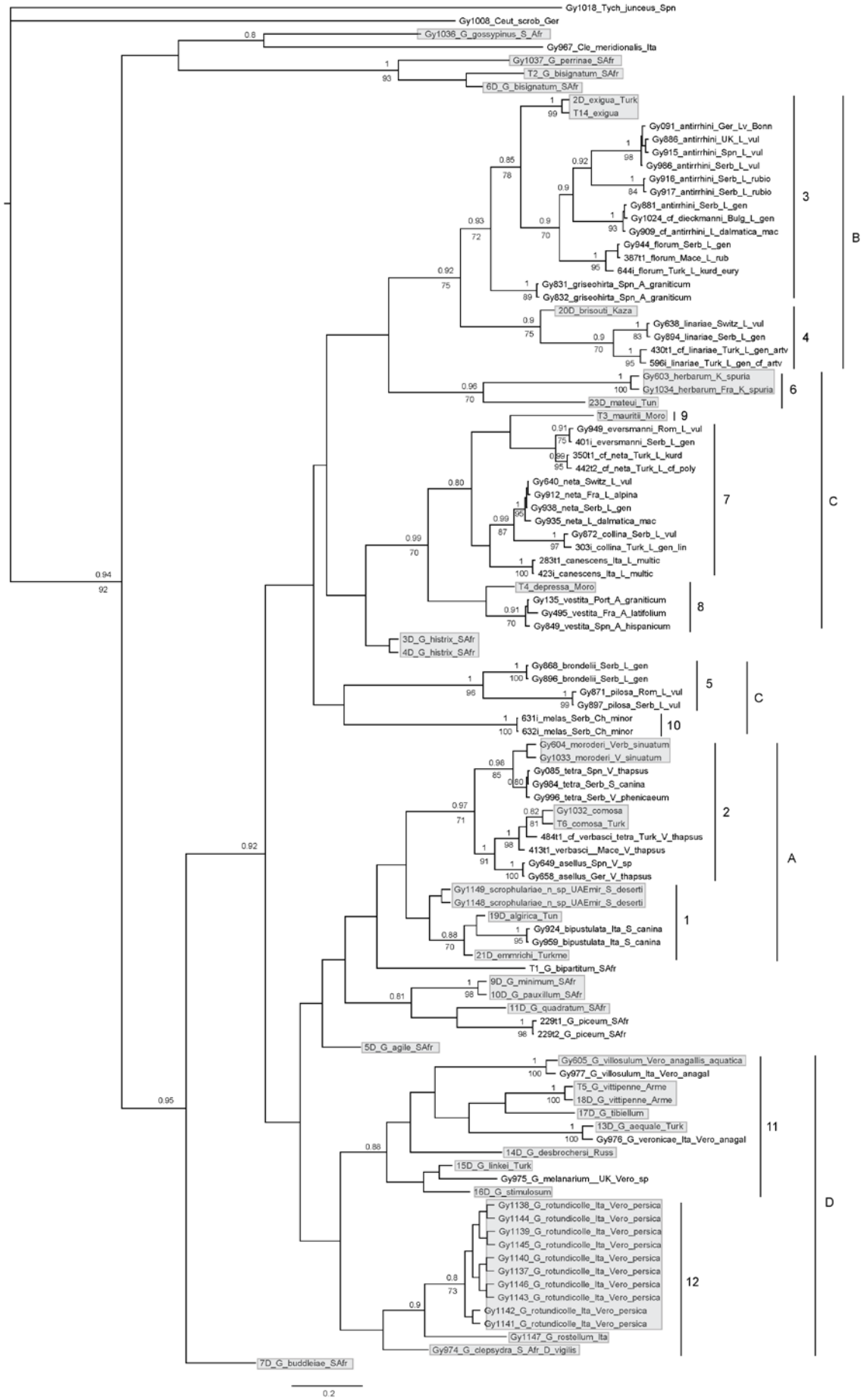
Gy1018_Tych_junceus_Spn
Gy1008_Ceut_scrob_Ger
Gy1036_G_gossypinus_S_Afr
Gy967_Cle_meridionalis_Ita
Gy1037_G_perrinae_SAfr
T2_G_bisignatum_SAfr
6D_G_bisignatum_SAfr
2D_exigua_Turk
T14_exigua
Gy091_antirrhini_Ger_Lv_Bonn
Gy886_antirrhini_UK_L_vul
Gy915_antirrhini_Spn_L_vul
Gy986_antirrhini_Serb_L_vul
Gy916_antirrhini_Serb_L_rubio
Gy917_antirrhini_Serb_L_rubio
Gy881_antirrhini_Serb_L_gen
Gy1024_cf_dieckmanni_Bulg_L_gen
Gy909_cf_antirrhini_L_dalmatica_mac
Gy944_florum_Serb_L_gen
387t1_florum_Mace_L_rub
644i_florum_Turk_L_kurd_eury
Gy831_griseohirta_Spn_A_graniticum
Gy832_griseohirta_Spn_A_graniticum
20D_brisouti_Kaza
Gy638_linariae_Switz_L_vul
Gy894_linariae_Serb_L_gen
430t1_cf_linariae_Turk_L_gen_artv
596i_linariae_Turk_L_gen_cf_artv
Gy603_herbarum_K_spuria
Gy1034_herbarum_Fra_K_spuria
23D_mateui_Tun
T3_mauritii_Moro
Gy949_eversmanni_Rom_L_vul
401i_eversmanni_Serb_L_gen
350t1_cf_neta_Turk_L_kurd
442t2_cf_neta_Turk_L_cf_poly
Gy640_neta_Switz_L_vul
Gy912_neta_Fra_L_alpina
Gy938_neta_Serb_L_gen
Gy935_neta_L_dalmatica_mac
Gy872_collina_Serb_L_vul
303i_collina_Turk_L_gen_lin
283t1_canescens_Ita_L_multic
423i_canescens_Ita_L_multic
T4_depressa_Moro
Gy135_vestita_Port_A_graniticum
Gy495_vestita_Fra_A_latifolium
Gy849_vestita_Spn_A_hispanicum
3D_G_histrix_SAfr
4D_G_histrix_SAfr
Gy868_brondelii_Serb_L_gen
Gy896_brondelii_Serb_L_gen
Gy871_pilosa_Rom_L_vul
Gy897_pilosa_Serb_L_vul
631i_melas_Serb_Ch_minor
632i_melas_Serb_Ch_minor
Gy604_moroderi_Verb_sinuatum
Gy1033_moroderi_V_sinuatum
Gy085_tetra_Spn_V_thapsus
Gy984_tetra_Serb_S_canina
Gy996_tetra_Serb_V_phenicaeum
Gy1032_comosa
T6_comosa_Turk
484t1_cf_verbasci_tetra_Turk_V_thapsus
413t1_verbasci__Mace_V_thapsus
Gy649_asellus_Spn_V_sp
Gy658_asellus_Ger_V_thapsus
Gy1149_scrophulariae_n_sp_UAEmir_S_deserti
Gy1148_scrophulariae_n_sp_UAEmir_S_deserti
19D_algirica_Tun
Gy924_bipustulata_Ita_S_canina
Gy959_bipustulata_Ita_S_canina
21D_emmrichi_Turkme
T1_G_bipartitum_SAfr
9D_G_minimum_SAfr
10D_G_pauxillum_SAfr
11D_G_quadratum_SAfr
229t1_G_piceum_SAfr
229t2_G_piceum_SAfr
5D_G_agile_SAfr
Gy605_G_villosulum_Vero_anagallis_aquatica
Gy977_G_villosulum_Ita_Vero_anagal
T5_G_vittipenne_Arme
18D_G_vittipenne_Arme
17D_G_tibiellum
13D_G_aequale_Turk
Gy976_G_veronicae_Ita_Vero_anagal
14D_G_desbrochersi_Russ
15D_G_linkei_Turk
Gy975_G_melanarium__UK_Vero_sp
16D_G_stimulosum
Gy1138_G_rotundicolle_Ita_Vero_persica
Gy1144_G_rotundicolle_Ita_Vero_persica
Gy1139_G_rotundicolle_Ita_Vero_persica
Gy1145_G_rotundicolle_Ita_Vero_persica
Gy1140_G_rotundicolle_Ita_Vero_persica
Gy1137_G_rotundicolle_Ita_Vero_persica
Gy1146_G_rotundicolle_Ita_Vero_persica
Gy1143_G_rotundicolle_Ita_Vero_persica
Gy1142_G_rotundicolle_Ita_Vero_persica
Gy1141_G_rotundicolle_Ita_Vero_persica
Gy1147_G_rostellum_Ita
Gy974_G_clepsydra_S_Afr_D_vigilis
7D_G_buddleiae_SAfr
0.2
1
2
3
4
5
6
7
8
9
10
11
12
A
B
C
D

**Figure 3.** Bayesian phylogenetic tree inferred from a concatenated data set of 5 gene DNA sequences (COII, 16S, AK, EF1-α, and 18S) comprising 3,943 bp from 54 ingroup taxa plus 37 archival specimens and 11 degraded samples with only 150 bp of nucleotide data corresponding to the 16S SPIAs (highlighted in grey) as described in chapter 3. Bayesian posterior probabilities≥ 0.8 and ML bootstrap values ≥ 70% are shown above and below branches, respectively. Letters (A – C) and numerals (1 – 10) refer to major assemblages and species groups, respectively, as proposed by Caldara et al. (2010), except assemblage D and its constituent species groups (11 – 12). 1) *Rhinusa bipustulata* group, 2) *R. tetra* group, 3) *R. antirrhini* group, 4) *R. linariae* group, 5) *R. pilosa* group, 6) *R. herbarum* group, 7) *R. neta* group, 8) *R. vestita* group, 9) *R. mauritii* group, 10) *R. melas* group, 11) *Gymnetron melanarium* group, 12) *G. rotundicolle* group.

## REFERENCES

Albach DC, Meudt HM, Oxelman B (2005) Piecing together the "new" Plantaginaceae. *American Journal of Botany* **92**, 297-315.

Anderson RS (1995) An evolutionary perspective of diversity in Curculionoidea. *Mem. Entomol. Soc. Wash.* **14**, 103-114.

Becerra JX (1997) Insects on plants: Macroevolutionary chemical trends in host use. *Science* **276**, 253-256.

Becerra JX, Venable DL (1999) Macroevolution of insect-plant associations: The relevance of host biogeography to host affiliation. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 12626-12631.

Bell CD, Soltis DE, Soltis PS (2010) The age and diversification of the angiosperms re-revisited. *American Journal of Botany* **97**, 1296-1303.

Berenbaum MR (2001) Chemical mediation of coevolution: Phylogenetic evidence for Apiaceae and associates. *Annals of the Missouri Botanical Garden* **88**, 45-59.

Bernays EA, Minkenberg O (1997) Insect herbivores: Different reasons for being a generalist. *Ecology* **78**, 1157-1169.

Borghuis A, van Groenendael J, Madsen O, Ouborg J (2009) Phylogenetic analyses of the leaf beetle genus Galerucella: Evidence for host switching at speciation? *Molecular Phylogenetics and Evolution* **53**, 361-367.

Caldara R (2001) Phylogenetic analysis and higher classification of the tribe Mecinini (Coleoptera: Curculionidae, Curculioninae). *Koleopterologische Rundschau* **71**, 171-203.

Caldara R (2003) Revisione dei Gymnetron della Regione Afrotropicale (Coleoptera Curculionidae). *Mem. Soc. Entomol. Ital.* **82**, 87-272.

Caldara R, Colonnelli E, Osella G (2008) Curculionidae Curculioninae: relationships between Mediterranean and southern African species (Coleoptera). *Biogeographia* **29**, 133-156.

Caldara R, Korotyaev BA (2010) Description of *Rhinusa exigua* sp.n. from Turkey (Coleoptera: Curculionidae: Curculioninae). *Koleopterologische Rundschau* **80**, 205-208.

Caldara R, Sassi D, Tosevski I (2010) Phylogeny of the weevil genus *Rhinusa* Stephens based on adult morphological characters and host plant information (Coleoptera: Curculionidae). *Zootaxa* **2627**, 39-56.

Cook JM, Rokas A, Pagel M, Stone GN (2002) Evolutionary shifts between host oak sections and host-plant organs in Andricus gallwasps. *Evolution* **56**, 1821-1830.

Crespi BJ, Carmean DA, Mound LA, Worobey M, Morris D (1998) Phylogenetics of social behavior in Australian gall-forming thrips: Evidence from mitochondrial DNA sequence, adult morphology and behavior, and gall morphology. *Molecular Phylogenetics and Evolution* **9**, 163-180.

Ehrlich PR, Raven PH (1964) Butterflies and plants - a study in coevolution. *Evolution* **18**, 586-608.

Farrell B, Mitter C (1990) Phylogenesis of insect plant interactions - Have Phyllobrotica leaf beetles (Chrysomelidae) and the Lamiales diversified in parallel? *Evolution* **44**, 1389-1403.

Farrell BD (1998) "Inordinate fondness" explained: Why are there so many beetles? *Science* **281**, 555-559.

Futuyma DJ (1983) Evolutionary interactions among herbivorous insects and plants. In: *Coevolution* (eds. Futuyma DJ, Slatkin M), pp. 207-231. Sinauer Associates, Sunderland, MA.

Futuyma DJ, Keese MC, Scheffer SJ (1993) Genetic constraints and the phylogeny of insect-plant associations - responses of *Ophraella communa* (Coleoptera, Chrysomelidae) to host plants of its congeners. *Evolution* **47**, 888-905.

Gassmann A, Paetel C (1998) Gymnetron netum (Col.; Curculionidae). A potential agent for biological control of toadflax in North America., p. 33. Centre for Agriculture and Biosciences International, Delémont, Switzerland.

Groppe K (1992) Gymnetron antirrhini Paykull (Col.: Curculionidae). A candidate for biological control of Dalmatian toadflax in North America, p. 22. Centre for Agriculture and Bioscience International, Delémont, Switzerland.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucl. Acids Symp. Ser.* **41**, 95-98.

Hernandez-Vera G, Mitrovic M, Jovic J*, et al.* (2010) Host-associated genetic differentiation in a seed parasitic weevil Rhinusa antirrhini (Coleptera: Curculionidae) revealed by mitochondrial and nuclear sequence data. *Molecular Ecology* **19**, 2286-2300.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755.

Jaenike J (1990) Host specialization in phytophagous insects. *Annual Review of Ecology and Systematics* **21**, 243-273.

Janz N, Nylin S (1998) Butterflies and plants: A phylogenetic study. *Evolution* **52**, 486-502.

Joy JB, Crespi BJ (2007) Adaptive radiation of gall-inducing insects within a single host-plant species. *Evolution* **61**, 784-795.

Kass RE, Raftery AE (1995) Bayes Factors. *Journal of the American Statistical Association* **90**, 773-795.

Kato T, Bonet A, Yoshitake H*, et al.* (2010) Evolution of host utilization patterns in the seed beetle genus Mimosestes Bridwell (Coleoptera: Chrysomelidae: Bruchinae). *Molecular Phylogenetics and Evolution* **55**, 816-832.

Kergoat GJ, Silvain JF, Delobel A, Tuda M, Anton KW (2007) Defining the limits of taxonomic conservatism in host-plant use for phytophagous insects: Molecular systematics and evolution of host-plant associations in the seed-beetle genus Bruchus Linnaeus (Coleoptera : Chrysomelidae : Bruchinae). *Molecular Phylogenetics and Evolution* **43**, 251-269.

Kölsch G, Pedersen BV (2008) Molecular phylogeny of reed beetles (Col., Chrysomelidae, Donaciinae): The signature of ecological specialization and geographical isolation. *Molecular Phylogenetics and Evolution* **48**, 936-952.

Lopez-Vaamonde C, Charles H, Godfray HCJ, Cook JM (2003) Evolutionary dynamics of host-plant use in a genus of leaf-mining moths. *Evolution* **57**, 1804-1821.

Maddison WP, Maddison DR (2010) Mesquite: a modular system for evolutionary analysis. Version 2.74 http://mesquiteproject.org.

Marvaldi AE, Sequeira AS, O'Brien CW, Farrell BD (2002) Molecular and morphological phylogenetics of weevils (Coleoptera, Curculionoidea): Do niche shifts accompany diversification? *Systematic Biology* **51**, 761-785.

Mayr E (1963) *Populations, species, and evolution* Harvard Univ. Press, Cambridge, Massachusetts.

McKenna DD, Sequeira AS, Marvaldi AE, Farrell BD (2009) Temporal lags and overlap in the diversification of weevils and flowering plants. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 7083-7088.

Miller JS (1987) HOST-PLANT RELATIONSHIPS IN THE PAPILIONIDAE LEPIDOPTERA PARALLEL CLADOGENESIS OR COLONIZATION? *Cladistics* **3**, 105-120.

Mitter C, Farrell B (1991) Macroevolutionary aspects of insect-plant relationships. In: *Insect/plant interactions* (ed. Bernays EA), pp. 35-78. CRC Press, Boca Raton, FL.

Nylander JAA, Ronquist F, Huelsenbeck JP, Nieves-Aldrey JL (2004) Bayesian phylogenetic analysis of combined data. *Systematic Biology* **53**, 47-67.

Nyman T, Farrell BD, Zinovjev AG, Vikberg V (2006) Larval habits, host-plant associations, and speciation in nematine sawflies (Hymenoptera : Tenthredinidae). *Evolution* **60**, 1622-1637.

Nyman T, Widmer A, Roininen H (2000) Evolution of gall morphology and host-plant relationships in willow-feeding sawflies (Hymenoptera : Tenthredinidae). *Evolution* **54**, 526-533.

Oberprieler RG, Marvaldi AE, Anderson RS (2007) Weevils, weevils, weevils everywhere. *Zootaxa*, 491-520.

Olmstead RG, DePamphilis CW, Wolfe AD*, et al.* (2001) Disintegration of the Scrophulariaceae. *American Journal of Botany* **88**, 348-361.

Pagel M (1999) The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Systematic Biology* **48**, 612-622.

Percy DM (2003) Radiation, diversity, and host-plant interactions among island and continental legume-feeding psyllids. *Evolution* **57**, 2540-2556.

Percy DM, Page RDM, Cronk QCB (2004) Plant-insect interactions: Double-dating associated insect and plant lineages reveals asynchronous radiations. *Systematic Biology* **53**, 120-127.

Posada D (2008) jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution* **25**, 1253-1256.

Ribeiro SP, Borges PAV, Gaspar C*, et al.* (2005) Canopy insect herbivores in the Azorean Laurisilva forests: key host plant species in a highly generalist insect community. *Ecography* **28**, 315-330.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574.

Ronquist F, Huelsenbeck JP, van der Mark P (2005) MrBayes 3.1 Manual, draft 5/ 26/2005.

Schluter D, Price T, Mooers AO, Ludwig D (1997) Likelihood of ancestor states in adaptive radiation. *Evolution* **51**, 1699-1711.

Schoonhoven L, van Loon J, Dicke M (2005) *Insect-Plant Biology*, 2nd Ed. edn. Oxford Univ Press, Oxford.

Sperling EA, Peterson KJ, Pisani D (2009) Phylogenetic-Signal Dissection of Nuclear Housekeeping Genes Supports the Paraphyly of Sponges and the Monophyly of Eumetazoa. *Molecular Biology and Evolution* **26**, 2261-2274.

Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690.

Stamatakis A, Hoover P, Rougemont J (2008) A Rapid Bootstrap Algorithm for the RAxML Web Servers. *Systematic Biology* **57**, 758-771.

Suchard MA, Weiss RE, Sinsheimer JS (2001) Bayesian selection of continuous-time Markov chain evolutionary models. *Molecular Biology and Evolution* **18**, 1001-1013.

Suchard MA, Weiss RE, Sinsheimer JS (2005) Models for estimating Bayes factors with applications to phylogeny and tests of monophyly. *Biometrics* **61**, 665-673.

Tosevski I, Gassmann A, Desancic M (2005) Biological control of Dalmatian and yellow toadflaxes, *Linaria dalmatica* and *L. vulgaris*., p. 1. Centre for Agriculture and Biosciences International, Delémont., Switzerland.

Wahlberg N (2001) The phylogenetics and biochemistry of host-plant specialization in Melitaeine butterflies (Lepidoptera : Nymphalidae). *Evolution* **55**, 522-537.

Winkler IS, Mitter C (2008) The phylogenetic dimension of insect/plant interactions: A summary of recent evidence. In: *Specialization, Speciation, and Radiation: The Evolutionary Biology of Herbivorous Insects* (ed. Tilmon K), pp. 240-263. Univ California Press, Berkeley.

Zakharov EV, Caterino MS, Sperling FAH (2004) Molecular phylogeny, historical biogeography, and divergence time estimates for swallowtail butterflies of the genus Papilio (Lepidoptera : Papilionidae). *Systematic Biology* **53**, 193-215.

# Chapter 5: Assessing the nature of mtDNA sequence ambiguities in the seed parasitic weevil *R. antirrhini* (Coleoptera : Curculionidae)

# ASSESSING THE NATURE OF mtDNA SEQUENCE AMBIGUITIES IN THE SEED PARASITIC WEEVIL *Rhinusa antirrhini* (COLEOPTERA : CURCULIONIDAE).

## ABSTRACT

Mitochondrial DNA (mtDNA) sequence ambiguities can be the result of laboratory contamination or natural biological phenomena such as the co-occurrence of two or more genetically distinct mtDNA genomes within the organelle (heteroplasmy) or nuclear-mitochondrial insertions (NUMT). In the course of data analyses for chapter 2, several mtDNA chromatogram sequences from seed parasitic weevils of *Rhinusa antirrhini* were found to exhibit ambiguous sites (double peaks). Using PCR assays, the nature of these sequence ambiguities was assessed. The results discard the possibility of cross-contamination between samples and provide evidence of gene exchange between divergent mitochondrial lineages either through heteroplasmy or a nuclear-mitochondrial insertion (NUMT).

## 5.1. INTRODUCTION

Since the advent of PCR based techniques, mitochondrial DNA (mtDNA) has long been the genetic marker of choice to infer the evolutionary and demographic histories of both populations and species (Avise *et al.* 1987). A large number of studies have employed this molecule because of a series of experimental and biological advantages that it possesses (Avise 1986). Experimentally, mitochondrial DNA is easy to obtain and amplify because of its high copy number per cell, it exhibits very short intergenic regions with no introns, and variable regions are frequently flanked by more conserved regions which facilitate the design of primers (Gissi *et al.* 2008; Harrison 1989; Moritz *et al.* 1987). In terms of biological properties, the higher mean mutation rate of mtDNA relative to nuclear DNA provides the opportunity to capture phylogenetic signal over short time frames. Additionally, the mtDNA genome behaves as a single non-recombining locus due to its maternal inheritance (Birky 1995), and it is also considered to evolve in a nearly neutral fashion, due to the evolutionary constraints on the protein coding genes that are involved in basic metabolic functions (Avise 1986; Moritz *et al.* 1987).

More recently many of the assumed properties described above have been questioned (e.g. Galtier *et al.* 2009; White *et al.* 2008). Among these concerns is a technical issue associated with the use of the mtDNA molecule in PCR-based techniques, specifically the amplification of mitochondrial genes that have been inserted into the nuclear genome (Bensasson *et al.* 2001; Zhang & Hewitt 1996). The presence of mitochondrial pseudogenes or nuclear-mitochondrial sequences (NUMTs) (Lopez *et al.* 1994) has been known for at least four decades and has been reported in a wide variety of organisms including bacteria, fungi, arthropods, and animals (see Bensasson *et al.* 2001 for a thorough review). Although exact details of

the mechanisms involved still remain largely unknown, recent evidence suggests that this phenomenon is an ongoing evolutionary process and incorporation of mitochondrial fragments into the nuclear genome occurs via non-homologous end joining at double-strand breaks after degradation of abnormal mitochondria (Hazkani-Covo *et al.* 2010; Henze & Martin 2001). Accidental co-amplification of these nuclear-mitochondrial DNA fragments can produce misleading and sometimes undetected incorrect results for estimates of phylogenetic relationships (Arctander 1995; Williams & Knowlton 2001), measures of biodiversity (Song *et al.* 2008), and studies of human diseases associated with mtDNA mutations (Yao *et al.* 2008). However, in some cases rather than being a nuisance they have been used in evolutionary studies as molecular markers, and useful tools to study rates and patterns of evolution in nuclear and mitochondrial DNA (Bensasson *et al.* 2001 and references therein).

In addition to nuclear copies, the presence of multiple DNA sequences of mtDNA origin within a PCR product may also find explanation from heteroplasmy. The traditional paradigm is that, in the majority of animals, the inheritance of mtDNA occurs uniparentally through the maternal germline, and therefore, each individual possess only one mtDNA haplotype (Ballard & Whitlock 2004). However, increasing evidence shows that more than one type of mtDNA can be present within a cell or individual, and examples can be found in a range of organisms including protists, fungi, plants, vertebrates and humans (Barr *et al.* 2005; Bromham *et al.* 2003; Kmiec *et al.* 2006). Heteroplasmy can result from point mutations, insertions or deletions, intra-mitochondrial duplications or perhaps more commonly, paternal mtDNA entering the egg cytoplasm at fertilization, a phenomenon referred to as “paternal leakage” (Ballard & Whitlock 2004; White *et al.* 2008). The mechanisms

preventing paternal mtDNA transmission vary across organisms and details are not fully understood yet (Birky 1995; Birky 2001); however it has been suggested that the frequency of paternal mtDNA leakage can be higher in interspecific crosses than at the intraspecific level because the molecular recognition system which destroy paternal mitochondria may be relaxed in such crosses (Rokas *et al.* 2003).

Perhaps because of the tacit acceptance of the aforementioned "standard" paradigm regarding the inheritance and biological properties of the mtDNA, researchers are likely to dismiss the idea of heteroplasmy or NUMTs when ambiguous sites are encountered during examination of mtDNA sequences unless there is a specific interest in these topics. In this chapter, I follow up atypical results observed during data analyses for chapter 2, where several mtDNA sequence chromatograms from the seed parasitic weevil *Rhinusa antirrhini* were found to exhibit ambiguities in several nucleotide positions in the form of double peaks. Sequences from the cytochrome c oxidase subunit II gene (COII) revealed the presence of host-associated mitochondrial lineages within *R. antirrhini*. Across all sampled individuals, five of them collected at different localities in England and Sweden presented double peaks at nucleotide positions associated with the discrimination of different host-associated plant lineages. Using PCR assays, the aim of this study is to assess whether the observed sequence ambiguities are the result of either (i) sample cross-contamination, or (ii) the manifestation of one of the biological phenomena of NUMT's or heteroplasmy. The implications of NUMTs or heteroplasmy are discussed in the context of potential gene flow between divergent host plant races of *R. antirrhini*.

## 5.2. MATERIALS AND METHODS

### 5.2.1. Taxon sampling

Five COII sequences from chapter 2 were previously observed to exhibit nucleotide ambiguities. The sequences correspond to samples of *R. antirrhini* collected on *Linaria vulgaris* from four localities in England (Tintagel, Stonehenge, Sutton Hoo, and Swansea), and one locality in Sweden (Morlanda). Further sampling from the localities in England and Sweden was undertaken to exclude sample contamination and assess the frequency of occurrence of the described sequence ambiguity. Fifty-three individuals of *R. antirrhini* were sampled from the above mentioned localities in England, and 1 individual from the locality in Sweden; all collected on the host plant species *Linaria vulgaris*.

### 5.2.2. DNA extraction, PCR amplification and sequencing reactions

The fifty-four newly-sampled weevils were punctured through the abdomen and total genomic DNA was extracted using the QIAGEN DNeasy extraction Kit (QIAGEN) following the manufacturer's instructions. After DNA extraction, weevils were placed again in 96% ethanol and maintained at 4°C as vouchers. To verify the atypical mtDNA sequences observed during data analyses for chapter 2, DNA from all of the samples previously collected at the localities above mentioned (20 individuals) was re-amplified for the cytochrome c oxidase subunit II gene (COII) along with the new set of 54 samples. A fragment of 697 bp from the COII mitochondrial gene was amplified using the primers TL2-J-3038 (5'-TAATATGGCAGATTAGTGCATTGGA) (Emerson *et al.* 2000) and TK-N 3782 (5'-GAGACCATTACTTGCTTTCAGTCATCT) (EVA-Harrison Laboratory,

Cornell University, Ithaca, NY, USA). Polymerase chain reactions (PCR) contained $NH_4$ buffer (1x), 2.5 - 3.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.2 - 0.4 µM of each primer, 0.5 U of Taq polymerase (Bioline) and 1-3 µl of DNA template in a 25µL final volume. PCR cycles were carried out using the following thermal profile: 95°C for 3 min, 33 cycles at 95°C for 1 min, annealing temperatures between 48-58°C for 1 min, 72°C for 1 min, and a final extension at 72°C for 3 min. Sequences were generated with a PerkinElmer ABI3700 automated sequencer, using the forward primer and a BigDye terminator reaction protocol (v3.1 PerkinElmer) in a 10 µl final volume.

### 5.2.3. Assessing the nature of mtDNA sequence ambiguities

Two PCR assays were carried out to assess the possibility of either a NUMT or heteroplasmy as the cause of mtDNA COII sequence ambiguities. The first assay involved the amplification of another mitochondrial gene, the ribosomal 16S, which is physically located at least 4,000 bp away from the COII gene within the mitochondrial genome. It has been shown that on average, nuclear-mitochondrial insertions tend to be short (~100 – 300 bp) (Richly & Leister 2004). Thus the rationale is that if we assume that the size of the translocated mtDNA fragment is not large enough as to encompass both the COII and 16S genes, then we would expect 16S sequences with no ambiguities. A 500 bp fragment of the 16S gene was amplified from three of the samples collected in England exhibiting double peaks; Gy1046 (Tintagel), Gy888 (Swansea), and Gy1224 (Stonehenge) using the primers 16Sar 5'-CGCCTGTTTATCAAAAACAT, and 16Sbr 5'-

CCGGTCTGAACTCAGATCACGT. PCR reactions and thermal cycles were carried out as previously described for the COII gene.

The second PCR assay consisted of amplifying the COII gene fragment using a series of dilutions of the DNA template. Assuming a lower ratio of nuclear DNA to mtDNA, given that cells contain several hundreds of mitochondria, each containing itself several mtDNA molecules (Sorenson & Quinn 1998), by diluting the template DNA we expect less copies of a putative NUMT to be amplified in PCR reactions. Therefore, with progressive dilutions, a decrease in the relative height of the putative NUMT double peaks should be observed in the chromatograms. Similar approaches of PCR amplification on diluted DNA have been applied to test for the presence of NUMTs in different organisms such as crustaceans (Calvignac *et al.* 2011), birds (Kidd & Friesen 1998), and a range of tissues from vertebrates and invertebrates (Ibarguchi *et al.* 2006). Four different dilutions were applied to DNA from two samples exhibiting ambiguous sequences, Gy1223 and Gy1224; dilutions used were 1:10, 1:50, 1:100, and 1:500. PCR amplification and sequencing reactions were performed as previously described. All sequences were automatically aligned using the ClustalW algorithm as implemented in BioEdit version 7.0.9 (Hall 1999) with further manual alignment.

## 5.3. RESULTS

### 5.3.1. Initial COII PCR amplification and sequencing reactions

Re-amplification of the 20 samples previously analyzed in chapter 2 confirmed the presence of COII sequence ambiguities in the same five samples of *R. antirrhini*; four collected from different localities in England, Gy1046 Tintagel, Gy1043

Stonehenge, Gy1044 Sutton Hoo, Gy888 Swansea, and one from Sweden: Gy1183 Morlanda. Five new instances of sequence ambiguities were detected from the set of 54 newly-collected samples, three collected from Stonehenge, (Gy1223, Gy1224 and Gy1248), one from Sutton Hoo (Gy1234), and one from Morlanda, Sweden (Gy1184).

Comparing the sequences against the full alignment from chapter 2 it was found that the observed sequence ambiguities correspond to diagnostic sites defining two host-associated mtDNA lineages. All of the samples exhibit one sequence corresponding to the mtDNA lineage associated with their host plant *L. vulgaris* (lineage 1 in chapter 2) and the other represents a new sequence (hereafter refer to as lineage 8), differing by 6 nucleotides from sequences of lineage 4 (chapter 2), associated with the host plants *L. genistifolia genistifolia*, *L. genistifolia sofiana*, and *L. dalmatica macedonica* (hereafter refer to as the *L. genistifolia/L. dalmatica* species complex) (Fig. 1). Interestingly, three individuals from the set of newly-collected samples were found to be homoplasmic for the sequence of lineage 8. These correspond to samples Gy1230 and Gy1190 from Stonehenge and Gy1187 from Tintagel, all collected on the host plant *L. vulgaris*.

### 5.3.2. 16S and COII PCR assays

The three samples amplified for the 16S gene fragment, Gy1046 (Tintagel), Gy888 (Swansea), and Gy1224 (Stonehenge) exhibited double peaks in their sequence chromatograms. Similar to the results from the initial COII PCR reactions, it was found that the observed nucleotide ambiguities correspond to diagnostic sites defining two host-associated mtDNA lineages. One sequence corresponds to the

mtDNA lineage associated with their host plant *L. vulgaris* (L1) and the other one is the same as sequences from *R. antirrhini* samples within lineage 4 (chapter 2) (Fig. 2). Results of the COII PCR dilution assay show that, chromatogram double peaks from samples Gy1223 and Gy1224 decreased in height with higher dilutions of DNA template (Fig. 3). These attenuated double peaks correspond to the sequence of lineage 1, associated with the host *L. vulgaris*.

## 5.4. DISCUSSION

Re-amplification and sequencing of a subset of samples gave the same results as previously reported. Additionally, independent DNA extractions from new individuals revealed another five instances of COII chromatograms with ambiguities corresponding to the same polymorphic sites. These results clearly rule out the possibility of cross-contamination between samples and provide evidence that the observed mitochondrial sequence ambiguities are the result of either a NUMT or heteroplasmy. Chromatogram sequences from both COII and 16S mitochondrial gene fragments exhibited nucleotide ambiguities (double peaks) at diagnostic sites defining host-associated mtDNA lineages. After discarding the possibility of contamination, two possible explanations can be put forward: (a) the polymorphic individuals have more than one copy of mtDNA (heteroplasmy), or (b) there is a considerably large NUMT which encompasses both the COII and the 16S mitochondrial genes. Results of the PCR dilution assay, show that chromatogram double peaks decrease in height with higher dilutions of DNA template, a result consistent with the hypothesis of a nuclear mtDNA insertion given the lower ratio of nuclear DNA to mtDNA (Sorenson & Quinn 1998). This result implies that the size

of the NUMT is at least 5,000 bp, encompassing both, the COII and the 16S amplified gene fragments. It has been suggested that, on average, NUMT's tend to be short (~100 – 300 bp) because typically they undergo fragmentation before nuclear integration or progressive deletion once in the nucleus (Pamilo *et al.* 2007; Richly & Leister 2004). However, transpositions of large fragments of mtDNA have been found in the nuclear genome of a number of organisms including, plants (620 Kb) (Stupar *et al.* 2001), Lizards (>7.6 Kb) (Podnar *et al.* 2007), humans (14.65 Kb) (Mourier *et al.* 2001), felines (7.9 Kb) (Lopez *et al.* 1994), and insects (3.5 Kb) (Gellissen *et al.* 1983).

An alternative explanation for the presence of double peaks in both COII and 16S chromatograms is that of heteroplasmy. It has been reported that heteroplasmy is a dynamic phenomenon with quantitative changes in the proportion of heteroplasmic variants and even rapid genomic shifts from a state of heteroplasmy to homoplasmy (Kmiec *et al.* 2006; Millar *et al.* 2008). Thus, the decrease in height of double peaks in the PCR dilution assay could be the result of low numbers of a different copy of mtDNA present in the organelle. A heteroplasmic state can be reached through different phenomena such as mutations, or intra-mitochondrial duplications, however, in animals the most frequently reported cause of this condition is paternal leakage, i.e. paternal mtDNA entering the egg cytoplasm at fertilization (White *et al.* 2008). This scenario appears plausible given that homoplasmic individuals exhibiting the COII sequence corresponding to lineage 8 were found at two localities shared with heteroplasmic individuals. These homoplasmic individuals correspond to samples Gy1230 and Gy1190 collected from Stonehenge and Gy1187 collected from Tintagel in England. Thus, the geographic proximity could facilitate gene

exchange between populations of weevils exhibiting divergent mitochondrial lineages and increase the possibility of mitochondrial introgression.

Further studies will be necessary to clarify the nature of the observed mtDNA ambiguities, however, an interesting aspect of this study is the detection of gene exchange between divergent mitochondrial lineages. This result is consistent with findings from Chapter 2, where it is shown that while reproductive barriers have evolved between host-associated mitochondrial lineages of *R. antirrhini*, reproductive isolation is not complete. Results from cross copulation experiments in Chapter 2 indicate that despite reduced frequency and offspring numbers, copulation between divergent mitochondrial lineages can occur. However, even though all samples in this study were collected on the host plant *L. vulgaris*, the mtDNA sequence of lineage 8 is closer to lineage 4 differing by 6 nucleotides. As shown in chapter 2, lineage 4 was the only mtDNA lineage found to be associated with more than one host plant taxa. This lineage includes individuals collected on *L. genistifolia genistifolia*, *L. genistifolia sofiana* and *L. dalmatica macedonica* from the Balkan region suggesting a more generalist oviposition behaviour. Therefore it is possible that lineage 8 represents a previously unsampled lineage closely related to lineage 4 and whose representatives exhibit a similar generalist behaviour. The plant species *L. genistifolia* has not been reported for England but there are records for *L. dalmatica*; a closely related species introduced as an ornamental plant (Clapham *et al.* 1987; Stace 1997). Thus, the possibility of having introduced *R. antirrhini* lineages associated with host plants within the *L. dalmatica/L. genistifolia* species complex seems very likely. Taxonomic doubts have surrounded this species complex in part because of the relatively ease to form hybrids, and until recently, *L. dalmatica* had been considered a subspecies of *L. genistifolia* (Chater *et al.* 1972;

Sutton 1988). Clapham et al. (1987) report the occurrence of hybrids of *L. vulgaris* and *L. dalmatica* in England, therefore another possibility is that homoplasmic individuals exhibiting lineage 8 were in fact collected from hybrid host plants.

## 5.5. CONCLUSION

Even though further studies will be necessary to clarify the nature of the observed mtDNA sequence ambiguities, the results clearly ruled out the possibility of cross-contamination between samples and provide evidence of gene exchange between divergent mitochondrial lineages of *R. antirrhini*. One of these represent a previously unsampled lineage, however, a more complete sampling will be needed to determine its host plant affiliation. Mitochondrial DNA sequence ambiguities should not be dismissed as contamination or PCR artefacts as these could be the result of biological phenomena providing valuable insights into the evolutionary history of an organism.

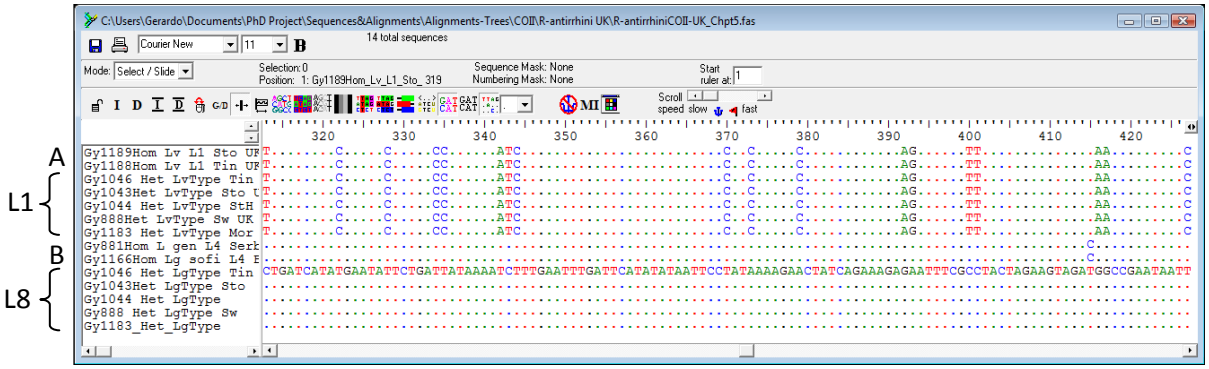


**Figure 1.** COII mitochondrial sequences corresponding to lineages 1 (L1) and 8 (L8) inferred from *Rhinusa antirrhini* samples exhibiting nucleotide ambiguities (double peaks) in their sequences. Letters A and B indicate sequences from L1 and L4 homoplasmic *R. antirrhini* individuals associated with the host plants *Linaria vulgaris*, and the *L. genistifolia/L. dalmatica* species complex respectively.

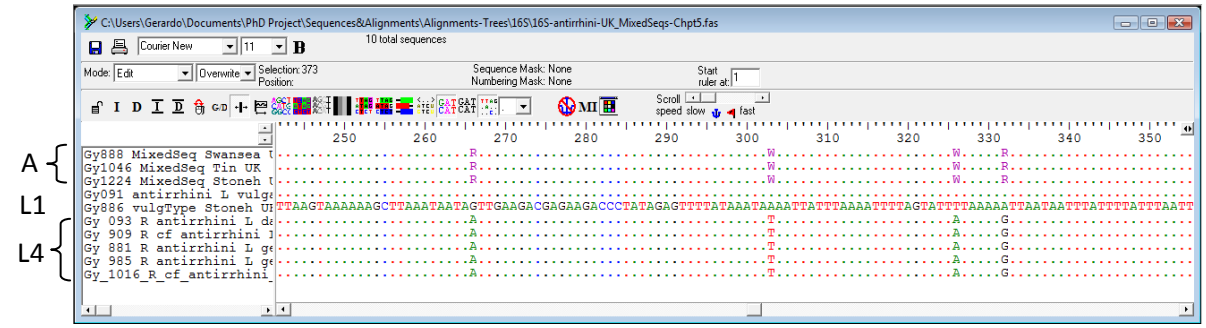


**Figure 2.** Detail of 16S mitochondrial sequences of *Rhinusa antirrhini* samples with nucleotide ambiguities (A) matching diagnostic sites of *R. antirrhini* sequences associated with host plants within the *L. genistifolia/L. dalmatica* species complex (L4) (lineage 4 in chapter 2). L1 indicates sequences from homoplasmic *R. antirrhini* individuals associated with *L. vulgaris* (lineage 1).

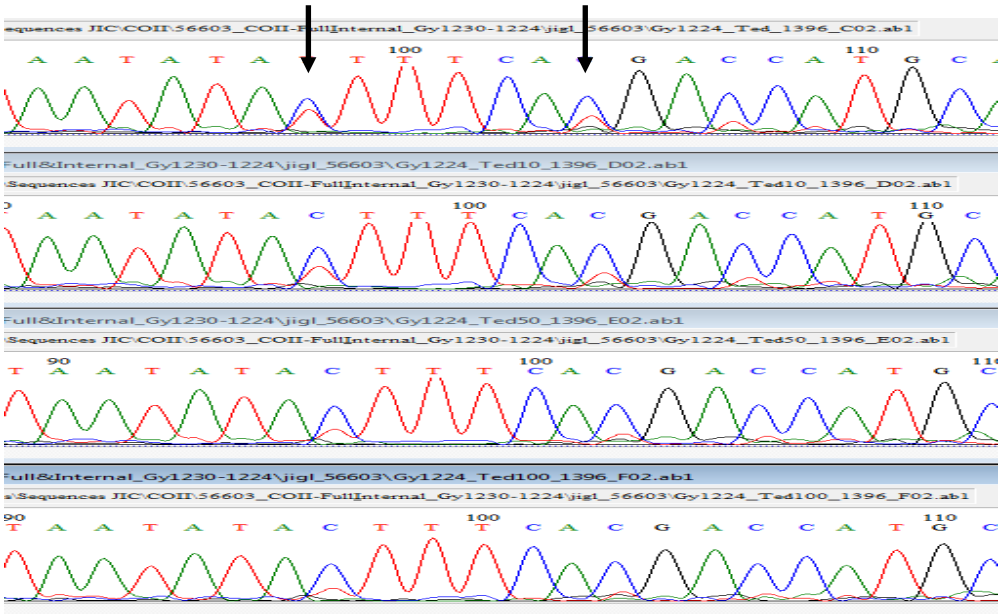

**Figure 3.** Detail of COII chromatogram sequences from PCR reactions using diluted DNA of sample Gy1224 *R. antirrhini*, collected from Stonehenge, England on the host plant *Linaria vulgaris*. From top to bottom dilutions used were 1:10, 1:50, 1:100, and 1:500. Arrows indicate decrease in relative height of double peaks at two ambiguous sites.

# REFERENCES


Arctander P (1995) Comparison of a mitochondrial gene and a corresponding nuclear pseudogene. *Proceedings of the Royal Society of London Series B-Biological Sciences* **262**, 13-19.

Avise JC (1986) Mitochondrial-DNA and the evolutionary genetics of higher animals. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **312**, 325-342.

Avise JC, Arnold J, Ball RM*, et al.* (1987) Intraspecific phylogeography: The mitochondrial-DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* **18**, 489-522.

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology* **13**, 729-744.

Barr CM, Neiman M, Taylor DR (2005) Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytologist* **168**, 39-50.

Bensasson D, Zhang DX, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology & Evolution* **16**, 314-321.

Birky CW (1995) Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 11331-11338.

Birky CW (2001) The inheritance of genes in mitochondria and chloroplasts: Laws, mechanisms, and models. *Annual Review of Genetics* **35**, 125-148.

Bromham L, Eyre-Walker A, Smith NH, Smith JM (2003) Mitochondrial Steve: paternal inheritance of mitochondria in humans. *Trends in Ecology & Evolution* **18**, 2-4.

Calvignac S, Konecny L, Malard F, Douady CJ (2011) Preventing the pollution of mitochondrial datasets with nuclear mitochondrial paralogs (numts). *Mitochondrion* **11**, 246-254.

Chater AO, Valdes B, Webb DA (1972) Linaria Miller. In: *Flora Europeae, Vol. 3* (eds. Tutin TG, Heywood VH, Burges NA*, et al.*), pp. 226-236. Cambridge University Press, Cambridge, MA.

Clapham AR, Tutin TG, Moore DM (1987) *Flora of the British Isles* Cambridge University Press, Cambridge.

Emerson BC, Oromi P, Hewitt GM (2000) Interpreting colonization of the Calathus (Coleoptera : Carabidae) on the Canary Islands and Madeira through the application of the parametric bootstrap. *Evolution* **54**, 2081-2090.

Galtier N, Nabholz B, Glemin S, Hurst GDD (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology* **18**, 4541-4550.

Gellissen G, Bradfield JY, White BN, Wyatt GR (1983) Mitochondrial-DNA sequences in the nuclear genome of a locust. *Nature* **301**, 631-634.

Gissi C, Iannelli F, Pesole G (2008) Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity* **101**, 301-320.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucl. Acids Symp. Ser.* **41**, 95-98.

Harrison RG (1989) Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends in Ecology & Evolution* **4**, 6-11.

Hazkani-Covo E, Zeller RM, Martin W (2010) Molecular Poltergeists: Mitochondrial DNA Copies (numts) in Sequenced Nuclear Genomes. *Plos Genetics* **6**.

Henze K, Martin W (2001) How do mitochondrial genes get into the nucleus? *Trends in Genetics* **17**, 383-387.

Ibarguchi G, Friesen VL, Lougheed SC (2006) Defeating numts: Semi-pure mitochondrial DNA from eggs and simple purification methods for field-collected wildlife tissues. *Genome* **49**, 1438-1450.

Kidd MG, Friesen VL (1998) Sequence variation in the guillemot (Alcidae : Cepphus) mitochondrial control region and its nuclear homolog. *Molecular Biology and Evolution* **15**, 61-70.

Kmiec B, Woloszynska M, Janska H (2006) Heteroplasmy as a common state of mitochondrial genetic information in plants and animals. *Current Genetics* **50**, 149-159.

Lopez JV, Yuhki N, Masuda R, Modi W, Obrien SJ (1994) NUMT, a recent transfer and tandem amplification of mitochondrial-DNA to the nuclear genome of the domestic cat. *Journal of Molecular Evolution* **39**, 174-190.

Millar CD, Dodd A, Anderson J*, et al.* (2008) Mutation and Evolutionary Rates in Adelie Penguins from the Antarctic. *Plos Genetics* **4**.

Moritz C, Dowling TE, Brown WM (1987) Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annual Review of Ecology and Systematics* **18**, 269-292.

Mourier T, Hansen AJ, Willerslev E, Arctander P (2001) The human genome project reveals a continuous, transfer of large mitochondrial fragments to the nucleus. *Molecular Biology and Evolution* **18**, 1833-1837.

Pamilo P, Viljakainen L, Vihavainen A (2007) Exceptionally high density of NUMTs in the honeybee genome. *Molecular Biology and Evolution* **24**, 1340-1346.

Podnar M, Haring E, Pinsker W, Mayer W (2007) Unusual origin of a nuclear pseudogene in the italian wall lizard: Intergenomic and interspecific transfer of a large section of the mitochondrial genome in the genus Podarcis (Lacertidae). *Journal of Molecular Evolution* **64**, 308-320.

Richly E, Leister D (2004) NUMTs in sequenced eukaryotic genomes. *Molecular Biology and Evolution* **21**, 1081-1084.

Rokas A, Ladoukakis E, Zouros E (2003) Animal mitochondrial DNA recombination revisited. *Trends in Ecology & Evolution* **18**, 411-417.

Song H, Buhay JE, Whiting MF, Crandall KA (2008) Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 13486-13491.

Sorenson MD, Quinn TW (1998) Numts: A challenge for avian systematics and population biology. *Auk* **115**, 214-221.

Stace C (1997) *New flora of the British Isles. Second edition*, Second edn. Cambridge University Press, Cambridge.

Stupar RM, Lilly JW, Town CD*, et al.* (2001) Complex mtDNA constitutes an approximate 620-kb insertion on Arabidopsis thaliana chromosome 2: implication of potential

sequencing errors caused by large-unit repeats. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 5099-5103.

Sutton DA (1988) *A revision of the tribe Antirrhineae.* Oxford University Press., London & Oxford.

White DJ, Wolff JN, Pierson M, Gemmell NJ (2008) Revealing the hidden complexities of mtDNA inheritance. *Molecular Ecology* **17**, 4925-4942.

Williams ST, Knowlton N (2001) Mitochondrial pseudogenes are pervasive and often insidious in the snapping shrimp genus Alpheus. *Molecular Biology and Evolution* **18**, 1484-1493.

Yao YG, Kong QP, Salas A, Bandelt HJ (2008) Pseudomitochondrial genome haunts disease studies. *Journal of Medical Genetics* **45**, 769-772.

Zhang DX, Hewitt GM (1996) Nuclear integrations: Challenges for mitochondrial DNA markers. *Trends in Ecology & Evolution* **11**, 247-251.

# Chapter 6: Concluding remarks and future prospects

# CONCLUDING REMARKS AND FUTURE PROSPECTS

Using a molecular phylogenetic approach, this thesis has identified questions and provided important answers concerning the evolutionary history of endophagous parasitic weevils within the genera *Rhinusa* and *Gymnetron*, specifically, the importance of their ecological interactions with their host plants as a major driver of diversification and their shared evolutionary history and geographic origin. Consistent with most of the reports on phylogenetic conservatism in host-plant use by phytophagous insects (Winkler & Mitter 2008), weevils within *Rhinusa* and *Gymnetron* are associated with a broad range of plant species. They exhibit phylogenetic conservatism in host use at the plant family level, most likely because of substantial differences in their chemical composition, however, a more dynamic process occurs at lower host taxonomic levels, where plant taxa exhibit similar secondary compounds and potentially allow for more frequent host shifts. Thus, while "major" host shifts (i.e. between different host-plant families) may be uncommon, shifts between closely related host-plants and in modes of parasitism have played an important role in the diversification of this group of weevils.

Although different modes of parasitism within a related group of taxa can be seen as an expression of ecological trait lability over geological time, evolutionary lineages within groups of related taxa may exhibit trait conservatism with ecological specialization on particular plant tissue/organs within restricted sets of host species. This phenomenon adds another important dimension of host use, as finer partitioning of resources facilitates the co-existence of multiple lineages sharing the same set of host-plants but within different adaptive zones or ecological niches. Such ecological divergence can lead to genetic differentiation, reproductive isolation and, ultimately, speciation (Hernandez-Vera *et al.* 2010). Results from chapter 2

represent an important contribution towards a better appreciation of this phenomenon, providing evidence that the degree by which two lineages diverge ecologically can be strong enough to overcome the homogenizing effects of gene flow. Weevils collected on different host-plants growing in sympatry exhibited host-associated genetic differentiation.

Before this thesis, *Rhinusa* and *Gymnetron* where considered to be reciprocally monophyletic groups. Furthermore, the South African-Mediterranean disjunct distribution exhibited by *Gymnetron* species represented an intriguing issue regarding their evolutionary origin (Caldara *et al.* 2008). Results of this thesis reveal that *Rhinusa* and *Gymnetron* represent a complex of lineages with a shared evolutionary history, providing evidence that South Africa represents the ancestral area from where these lineages started to diversify. Consistent with explanations put forward for similar disjunct distributions in other groups of beetles (Audisio *et al.* 2008; Bologna *et al.* 2008), the results support a late Miocene vicariance scenario, most likely as a result of repeated desertification phenomena.

In the context of advancing methodology and providing a sampling strategy that can benefit the community of molecular phylogeneticists, this thesis also represents an important contribution. The power and utility of PCR-based techniques in combination with advanced phylogenetic methods is evidenced by the results obtained from chapters 3 and 4. The targeted amplification of short phylogenetically informative amplicons can provide researchers with the ability to take advantage of archival material to augment sampling for molecular phylogenetic analyses. Even though archival material, represents a potential wealth of genetic information, as archival material frequently encompasses difficult to collect, rare, or even extinct species, the use of museum specimens to extract genetic information is

underutilized (Wandeler *et al.* 2007). In this thesis, the increase of species representation by using dry-pinned samples from an entomological collection, allowed for the inference of the biogeographic origin of *Rhinusa* and *Gymnetron*, and also contributed toward the clarification of the challenging taxonomy of the group.

This thesis work has provided me with a greater appreciation of the power and utility of molecular phylogenetics, and in part this appreciation has come from a lot of troubleshooting in laboratory work. It is now widely appreciated that a variety of technical issues associated with the use of mtDNA in PCR-based techniques may arise within such studies. The presence of nucleotide ambiguities in mtDNA sequences can be the result of either laboratory contamination or natural biological processes such has heteroplasmy or mitochondrial pseudogenes inserted into the nuclear genome (NUMT's). In any case, co-amplification of these spurious mtDNA fragments can produce misleading and sometimes undetected incorrect results. Perhaps because of the tacit acceptance of the "standard" paradigm regarding the inheritance and biological properties of the mtDNA, researchers are likely to dismiss the idea of heteroplasmy or NUMTs when ambiguous sites are encountered during examination of mtDNA sequences. However, as demonstrated in chapter 5, when they are identified as the result of these biological processes they can provide valuable insights into the evolutionary history of an organism. After discarding the possibility of cross-contamination between samples, the results provided evidence of gene exchange between divergent mitochondrial lineages of *R. antirrhini.*

Phylogenetic analyses of host use in plant-feeding insects have become a focus of interest in recent years. However, most of these studies have focused on external feeders (but see Barat *et al.* 2008; Downie *et al.* 2008; Erney *et al.* 1996 for

examples), neglecting the opportunity that endophagous insects represent for the study of insect-plant interactions given the more intimate association with their hosts, and therefore their increased susceptibility to disruptive selection pressures. Furthermore, empirical work on host-associated reproductive isolation is still rather limited and has mostly occurred since the late 1980s (Funk *et al.* 2002). In this thesis I have taken advantage of the biological and ecological characteristics of endophagous parasitic weevils whose life cycle is intimately linked to that of their host-plants, with larvae feeding and developing inside plant tissues. With the help of collaborators, I have complemented molecular phylogenetic analyses with behavioural data to assess the importance of ecological divergence in promoting reproductive isolation and restricted gene flow. Although further studies will be necessary to specifically test it, taken together, the results of this thesis suggest a scenario of adaptive radiation, whereby relatively young lineages of weevils are in the process of diversifying within the "speciation continuum" (Dres & Mallet 2002). Subtle morphological variations and partial reproductive isolation between host-associated lineages may be a signature of an ongoing invasion into new adaptive zones with underutilized niches as theory predicts (Schluter 2000). The importance of environmental and ecological factors in shaping the diversity of organisms has always been an intuitive idea since the publication of Darwin's book "The origin of the species". Today, there is a growing appreciation of the important role of ecological divergence in promoting speciation (Schluter 2000). In this context, this thesis reinforces the notion that ecological interactions are an important mechanism for the generation of biological diversity.

*Future prospects*

While this thesis has provided answers to some interesting questions surrounding the evolutionary history of parasitic weevils within the genera *Rhinusa* and *Gymnetron*, other questions that warrant investigation have arisen. For instance, results from chapter 2 suggest incomplete lineage sorting and/or gene flow between different host-associated lineages of *R. antirrhini*. Taking advantage of increasingly refined analytic methods such as coalescent-based models under a Bayesian framework (Hey 2010; Nielsen & Wakeley 2001; Pinho & Hey 2010), it would be interesting to quantitatively assess the contribution of both phenomena. As revealed by the results from chapters 3 and 4, *Rhinusa* and *Gymnetron* are not reciprocally monophyletic groups as previously suggested. Because of the taxonomic uncertainties surrounding the relationships between members of the tribe Mecinini, a phylogenetic analysis of the tribe including samples with different geographic origins would contribute not only to clarify the systematics of the group, but also to have a more complete picture of their evolutionary history. Unfortunately, little is known about the life histories of many taxa within this group of weevils. Future investigations into their life cycles, modes of parasitism, and interactions with other weevils and insect taxa will contribute to a more accurate representation of the role of ecological factors in generating biodiversity. Although results from chapter 5 suggest a nuclear-mitochondrial introgression as the cause for the observed mtDNA sequence ambiguities, further studies will be necessary to clarify the nature of these. Whether this is the result of a NUMT or heteroplasmy, the system offers the opportunity to address interesting questions regarding gene exchange between divergent mitochondrial lineages of plant parasitic insects associated with different host plant species.

# REFERENCES

Audisio P, De Biase A, Kirk-Spriggs AH*, et al.* (2008) Molecular biogeography of Mediterranean and southern African disjunctions as exemplified by pollen beetles of the Meligethes planiusculus species-group and related taxa (Coleoptera: Nitidulidae; Meligethinae. *Biogeographia* **29**, 45-65.

Barat M, Tarayre M, Atlan A (2008) Genetic divergence and ecological specialisation of seed weevils (Exapion spp.) on gorses (Ulex spp.). *Ecological Entomology* **33**, 328-336.

Bologna MA, Di Giulio A, Pitzalis M (2008) Systematics and biogeography of the genus Actenodia (Coleoptera : Meloidae : Mylabrini). *Systematic Entomology* **33**, 319-360.

Caldara R, Colonnelli E, Osella G (2008) Curculionidae Curculioninae: relationships between Mediterranean and southern African species (Coleoptera). *Biogeographia* **29**, 133-156.

Downie DA, Donaldson JS, Oberprieler RG (2008) Molecular systematics and evolution in an African cycad-weevil interaction: Amorphocerini (Coleoptera : Curculionidae : Molytinae) weevils on Encephalartos. *Molecular Phylogenetics and Evolution* **47**, 102-116.

Dres M, Mallet J (2002) Host races in plant-feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **357**, 471-492.

Erney SJ, Pruess KP, Danielson SD, Powers TO (1996) Molecular differentiation of alfalfa weevil strains (Coleoptera: Curculionidae). *Annals of the Entomological Society of America* **89**, 804-811.

Funk DJ, Filchak KE, Feder JL (2002) Herbivorous insects: model systems for the comparative study of speciation ecology. *Genetica* **116**, 251-267.

Hernandez-Vera G, Mitrovic M, Jovic J*, et al.* (2010) Host-associated genetic differentiation in a seed parasitic weevil Rhinusa antirrhini (Coleptera: Curculionidae) revealed by mitochondrial and nuclear sequence data. *Molecular Ecology* **19**, 2286-2300.

Hey J (2010) The Divergence of Chimpanzee Species and Subspecies as Revealed in Multipopulation Isolation-with-Migration Analyses. *Molecular Biology and Evolution* **27**, 921-933.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: A Markov chain Monte Carlo approach. *Genetics* **158**, 885-896.

Pinho C, Hey J (2010) Divergence with Gene Flow: Models and Data. In: *Annual Review of Ecology, Evolution, and Systematics, Vol 41*, pp. 215-230.

Schluter D (2000) *The ecology of adaptive radiation* Oxford University Press, Oxford.

Wandeler P, Hoeck PEA, Keller LF (2007) Back to the future: museum specimens in population genetics. *Trends in Ecology & Evolution* **22**, 634-642.

Winkler IS, Mitter C (2008) The phylogenetic dimension of insect/plant interactions: A summary of recent evidence. In: *Specialization, Speciation, and Radiation: The Evolutionary Biology of Herbivorous Insects* (ed. Tilmon K), pp. 240-263. Univ California Press, Berkeley.