Ecological Entomology

Ecological Entomology (2010), **35** (Suppl. 1), 18–32 DOI: 10.1111/j.1365-2311.2009.01145.x

OVERVIEW

DNA profiling of host–herbivore interactions in tropical forests

SARA PINZÓN NAVARRO,[1,2] JOSÉ A. JURADO-RIVERA,[3] JESÚS GÓMEZ-ZURITA,[4] CHRISTOPHER H. C. LYAL[1] and ALFRIED P. VOGLER[1,2] [1]Department of Entomology, Natural History Museum, London, U.K., [2]Division of Biology, Imperial College London, Ascot, U.K., [3]Laboratori de Genètica, Dpt. Biologia, Universitat de les Illes Balears, Palma de Mallorca, Balearic Islands, Spain and [4]Filogenia y Sistemática Animal, Institut de Biologia Evolutiva (CSIC-UPF), Barcelona, Spain

Abstract. 1. The diversity of insects in tropical forests remains poorly known, in particular regarding the critical feeding associations of herbivores, which are thought to drive species richness in these ecosystems.

2. Host records remain elusive and traditionally require labour-intensive feeding trials. A recent approach analyses plant DNA ingested by herbivorous insects; direct PCR amplification from DNA extracts from weevils (Curculionoidea) using chloroplast (*trnL* intron) primers was successful in 41 of 115 cases, resulting in 40 different sequences.

3. The resulting *trnL* intron sequences were identified against public databases to various hierarchical levels based on their position in phylogenetic trees and shown to be members of 26 plant families from different major groups of angiosperms.

4. Among the *trnL* intron sequences, seven pairs or triplets of close relatives (0–2 bp difference) were found which may represent intraspecific variation in the respective host plants.

5. Molecular clock calibrations of mitochondrial *cox1* sequences of weevils established great distances of lineages obtained (all splits estimated >20 Mya). Distant taxa were found to feed on the same or similar hosts in some cases, showing low evolutionary conservation of host associations among deeper levels.

6. The technique provides a new means of studying species diversity and plant–herbivore interactions in tropical forests, and removes the constraints of the need for actual observations of feeding in ecological and evolutionary studies.

Key words. Molecular identification, insect–plant interactions, curculionoidea, mtDNA, cpDNA.

Introduction

A means of rapidly and effectively assessing insect–plant feeding associations is urgently needed. The huge species richness of insects in tropical forests is widely attributed to their interactions with an equally puzzling diversity of angiosperm host plants (Janzen, 1970; Farrell, 1998). The major hypotheses either invoke the complex interactions of the insect herbivores and plants, e.g. due to high host specificity and niche partitioning associated with plant defence systems (Coley & Barone, 1996; Novotny *et al*., 2006) or the correlation with host plant phylogenetic diversity, which is highest in tropical regions (Coley & Barone, 1996; Novotny *et al*., 2006). A further possible factor ratcheting up total species richness in the tropics is turnover (beta-diversity) in host associations among local assemblages (Lewinsohn & Roslin, 2008). In all cases, testing these hypotheses of tropical insect–host plant diversification requires solid data on host associations and diet breadth. Such host records are equally important in estimations of the magnitude of species richness

Correspondence: Alfried P. Vogler, Department of Entomology, Natural History Museum, London SW7 5BD, U.K. E-mail: apv@nhm.ac.uk

Conflicts of interest: the authors have declared no conflicts of interest.

itself, which rely heavily on extrapolations from the better-known plant diversity (e.g. Erwin, 1982; Odegaard, 2000).

The methodological approach adopted for establishing feeding associations may affect the inferences drawn about host specificity and herbivore species richness. Diverse approaches have been used to obtain data on host associations and host preferences of insects, but all of them are time-consuming and have various limitations. Classical methods include observations of host use either *in situ* (Barone, 1998, 2000) or in laboratory tests (De Boer & Hanson, 1984; Lee & Bernays, 1988; Barone, 1998; Novotny *et al*., 2002; Novotny *et al*., 2006; Dyer *et al*., 2007), transplantation experiments (Eichhorn *et al*., 2008) or behavioural tests by exposure to plant volatiles (Schneider, 1957; Chen & Fadamiro, 2007; Fernandez *et al*., 2007). Other studies have attempted the direct identification of the feeding source, either through morphological analysis of the gut content (Otte & Joern, 1976; Fry *et al*., 1978), diet plant tissue-specific staining techniques (Schlein & Jacobson, 1999), or diet plant isotope analysis from gut contents (Post, 2002). Most analyses of insect host specificity in rainforests have been much less detailed. Early studies did not test feeding directly but used capture sites (individual rainforest trees) to establish host associations (e.g. Erwin, 1982). Studies based on direct feeding trials generally led to estimates of tighter host specificity and generally better founded data on host associations than simple presence on a host plant (Lewinsohn & Roslin, 2008). This would make a strong case for feeding studies to establish the true nature of herbivore interactions, but the huge expense in manpower required (e.g. Weiblen *et al*., 2006) is prohibitive in most cases. In addition, feeding studies of this kind usually concentrate on common herbivore species (Barone, 2000; Novotny *et al*., 2007), although most species (of plants and herbivores) in tropical forest assemblages are rare (Novotny & Basset, 2000), while artificial breeding conditions may alter insect behaviour and therefore result in inaccurate conclusions on host breadth. A further problem is that the taxonomic uncertainty in hyperdiverse insect groups due to numerous unnamed species which requires that specimens rather than names must be cross-checked among samples, which in turn adds great difficulties to analyses of herbivore data among independent studies (Lewinsohn *et al*., 2005). Finally, where species delimitation is incomplete and relies on preliminary morphospecies approaches, the finer details of host associations, including recently diverged races or species, may be missed altogether (Hebert *et al*., 2004; Condon *et al*., 2008).

DNA-based techniques can potentially solve the dual problem of imprecise insect taxonomy and incomplete host plant data in a single step. Specimens of folivorous leaf beetles have been shown to contain a 'molecular record' of their feeding source in the form of ingested plant material. Consequently, host plant DNA can be PCR amplified from a standard whole-body DNA extraction of herbivorous insects and identified against existing taxonomic DNA databases (Jurado-Rivera *et al*., 2009). This same DNA extraction is used to amplify diagnostic insect DNA fragments for a sequence-based identification of the herbivore. This procedure opens up entirely new approaches to the study of tropical herbivore–host plant interactions overcoming the existing limitations, as it both provides a sensitive means of comparison and potentially obviates the need for taxonomic identification to species or below of either plant or insect. If applied on a large scale, the technique promises to provide the elusive host data for complete herbivore communities in tropical forests and elsewhere, including rare species currently left out of the analyses. The approach has been shown to perform in practice, through comparison of hosts suggested by DNA analysis and hosts observed in the field (Jurado-Rivera *et al*., 2009).

Jurado-Rivera *et al*. (2009) applied their analysis to a group of Australian leaf beetles (Chrysomelidae), with the specific purpose of establishing herbivore–host plant co-evolution in the subfamily Chrysomelinae. Using the *trnL* (UAA) intron chloroplast marker (Taberlet *et al*., 1991) as a plant 'DNA barcode' (Taberlet *et al*., 2007), PCR amplification was achieved readily from freshly caught specimens. The resulting sequences were searched against a database of over 67 000 *trnL* intron sequences available at that time in GenBank and permitted identification to genus or tribe in most cases. PCR amplification was obtained in nearly every sample tested and revealed only a single host plant except in a small proportion (10 out of 76) of species (Jurado-Rivera *et al*., 2009). The latter could be separated with standard cloning techniques and resulted in successful identification of multiple hosts. A separate study obtained DNA from the gut of plant feeding individuals from various insect orders and readily produced authentic sequences of the *rbcL* locus (Matheson *et al*., 2007).

The Australian chrysomelids used for host identification were mainly taken from comparatively species poor plant assemblages in subtropical sclerophyll forest. At most study sites an inventory of local flowering plants was available and provided a means of confirming or refining the identification where the *trnL* sequence was inconclusive, either due to insufficient sequence variation or poor taxon coverage in the database (Jurado-Rivera *et al*., 2009). Host identifications in tropical rain forests constitute a greater challenge because of higher overall diversity of the assemblage, less complete floristic information at local sites, and highly variable database coverage among major host plant lineages. Here, we describe the results of a trial prior to any large-scale attempts, based on a set of weevils (Curculionoidea) reared from fruits and seeds of known hosts. DNA from beetle extractions was used for PCR amplification using plant *trnL* intron markers. In addition, we applied the procedure to specimens collected in flight intercept traps or by hand without the context of a host plant. Most of them were obtained from the tropical forest of Barro Colorado Island (BCI), Panama, whose plant composition is comparatively well known (Croat, 1978). The specimens were used to test factors critical for DNA-based surveys of host association, including (1) the proportion of individuals producing (single or multiple) plant sequences, and (2) the confidence of identifications achievable with this marker given the level of sequence variation and the richness of the existing reference database. Phylogenetic inferences from both insect and host plant sequences provided additional information on their level of co-evolutionary associations, despite the limited sampling available in this study.

Materials and methods

Taxon sampling and plant 'barcodes' from ingested DNA

Weevils were collected during April–November 2006 on Barro Colorado Island, Panama, in flight intercept traps and manually. Flight intercept traps were attached to 12 trees from different species in Barro Colorado Island Monument at an approximate height of 25–30 m from the ground. Two traps were also set in San Lorenzo National Park at similar heights. A few specimens were obtained by manual collecting at Parque Nacional Altos de Campana and Fortuna Dam Area (both Republic of Panama) and La Selva Biological Station (Costa Rica). In addition, several individuals were reared from seeds of known host plants, to test the success of the DNA-based host identification. Local experts identified the sampled plants, using morphological characters of fruits and leaves. Specimens of weevils were identified to the lowest taxonomic level possible using morphological characters and comparisons to the Natural History Museum collection. The species included here were members of three families of Curculionoidea: Brentidae, Curculionidae (six subfamilies) and Dryophthoridae.

A total of 115 individuals were selected for genomic DNA extraction (Table S1). The head and thorax of each individual were used for the extraction with a Promega 96-well plate kit. After extraction, beetles were prepared for morphological identification; vouchers will remain in the collections at Universidad de Panama and the Natural History Museum, London. To establish phylogenetic affinities of the beetles, a fragment of the 3′ end of the cytochrome oxidase I (*cox1*) gene was amplified using primers C1-J-2183 (Jerry) and TL2-N-3014 (Pat) (Simon *et al*., 1994) and resulted in 600–782 bp of sequence per individual. The same DNA template was used for PCR amplification of the plastid *trnL* intron using the plant-specific primers c-A49325 and d-B49863 (Taberlet *et al*., 1991; Jurado-Rivera *et al*., 2009). Sequencing was from both strands with the same primers used for PCR amplification, using BigDye 2.1 and an ABI PRISM3730 automated sequencer (Applied Biosystems, Foster City, CA, USA). Sequence chromatograms were assembled and edited with Sequencher 4.6 (Gene Codes Corp., Ann Harbor, Michigan) (Table S2).

Phylogenetic analysis and host species identification

Phylogenetic trees from *cox1* sequences of weevils were obtained using maximum likelihood searches with RAxML 7.0.4 (Stamatakis, 2006) using two partitions (first and second codon positions together vs. third codon positions) and 1000 bootstrap replicates were performed (Stamatakis *et al*., 2008) under a $\mathrm{GTR} + \mathrm{I} + \Gamma$ substitution model fixed in RAxML. Representatives of Brentidae were specified as outgroup during the tree search, as this family is closely related, but clearly separated, from the Curculionidae (Hunt *et al*., 2007; McKenna *et al*., 2009).

Node ages were calculated using an uncorrelated lognormal relaxed clock in BEAST 1.4.8 (Drummond & Rambaut, 2007). The brentid species were constrained as a monophyletic clade and 2.3% divergence My^{-1} (Brower, 1994) was used as a fixed substitution rate. Two independent runs of 100 million generations (sampling every 10 000th generation) were done, using the $\mathrm{GTR} + \mathrm{I} + \Gamma$ model and a Yule prior under default parameters for all other settings. Convergence and mixing of MCMC chains were estimated by trace plots using Tracer 1.4.1 (Drummond & Rambaut, 2007), ensuring stationarity. Trees were summarised in Tree Annotator 1.4.8, keeping trees after burn-in of 100 000 generations.

Identification of *trnL* sequences was performed individually through similarity BLAST searches against GenBank (Altschul *et al*., 1990), and subsequent phylogenetic analysis of the query (diet) sequence together with the 100 top hits. The resulting topology was annotated for the higher taxa included at the level of genus, tribe and family. The position of the query in this tree was used to identify the unknown sequence as member of the lowest hierarchical group into which it was included with posterior probability >0.7 in Bayesian analysis (Jurado-Rivera *et al*., 2009). Prior to tree reconstruction, each query and its top hits, plus outgroup gymnosperm sequences of *Cycas siamensis* (AY651841), *Gingko biloba* (AY145323) and *Pseudotsuga menziensii* (AF327589) were aligned with MAFFT 5.0 using default parameters for the L-INS-I strategy (Katoh *et al*., 2005). Phylogenetic trees were obtained with MrBayes 3.1.2 (Ronquist & Huelsenbeck, 2003) using the $\mathrm{GTR} + \mathrm{I} + \Gamma$ model, as chosen by Modeltest 3.7 (Posada & Crandall, 1998). Trees were obtained after two independent runs of three MCMC chains with varying numbers of generations, with a starting number of 3 million generations, sampling trees every 100th generation. Runs that converged only after 10 million generations were re-run using one cold and two incrementally heated Markov chains ($\lambda = 0.1$) and sampling every 1000 steps. Burn-in generations were calculated and excluded from the analysis as before and all-compatible consensus trees were obtained from the retained trees.

Results

Diversity of weevils

Out of 781 nucleotide positions sequenced for the mitochondrial *cox1*, 454 positions were variable and 389 informative. The sample included 13, 99 and 3 individuals of Brentidae, Curculionidae, and Dryophthoridae, respectively, but more detailed morphological identification to species level proved difficult despite access to excellent reference collections. However, identification to genus or subfamily revealed a great taxonomic diversity. Only the genus *Conotrachelus* was represented by several species, including three species (four individuals) reared from fruits and seeds of plants of known identity. A maximum likelihood tree obtained from the *cox1* sequences was generally consistent with the currently accepted classification (Alonso Zarazaga & Lyal, 1999), except for the paraphyly of the most widely sampled subfamily Cryptorhynchinae, and some inconsistencies at the tribal (Hylobiini) and genus (*Teramocerus*, *Zygops*, *Eubulus*) level (Fig. 1). Preliminary analysis of the sequences showed that in

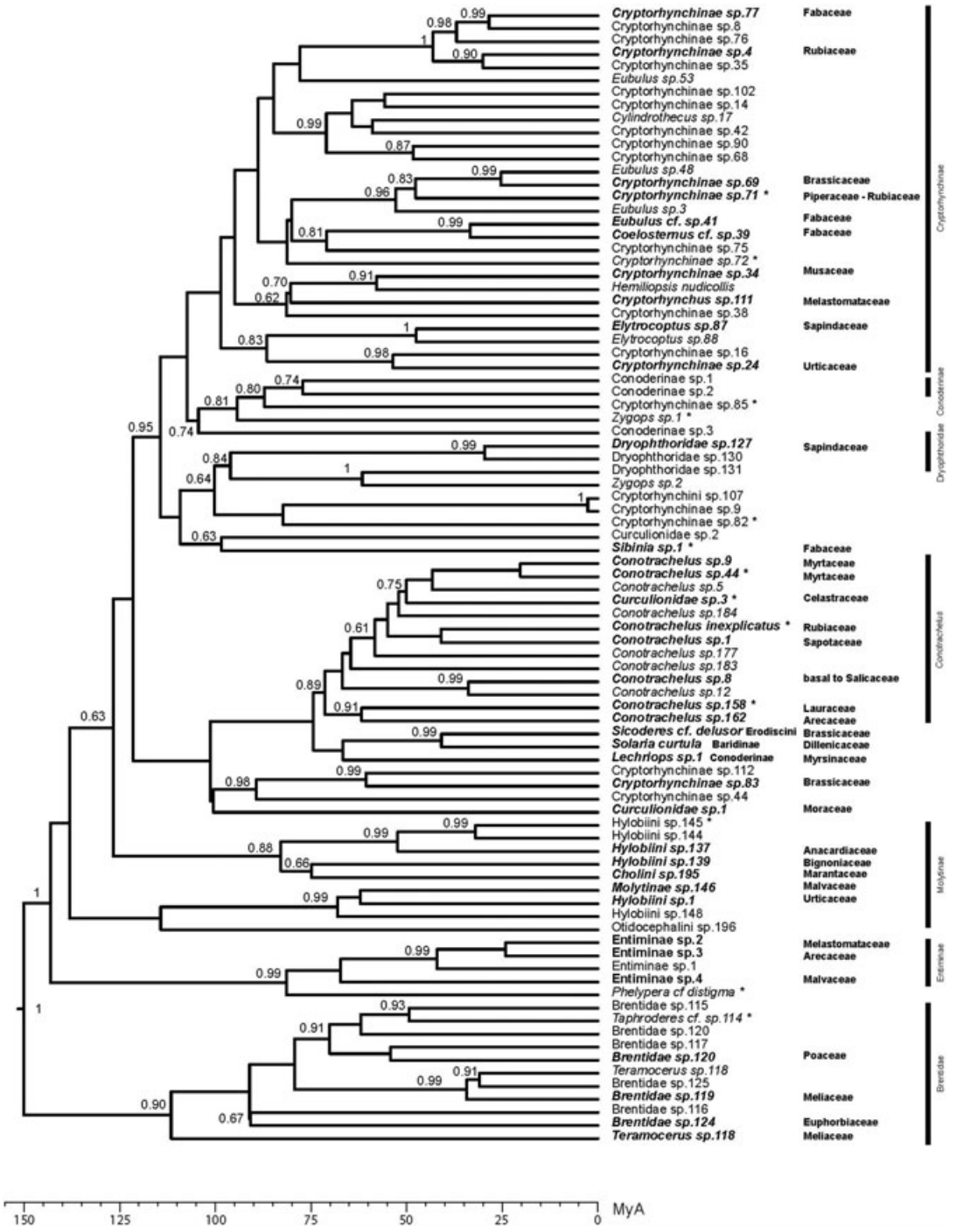

Fig. 1. Time-calibrated phylogenetic tree of weevils tested for *trnL* intron amplifications showing the major families and subfamilies. Species for which *trnL* sequences were obtained are mapped with their inferred plant families (bold terminals). Terminals marked with an asterisk (*) each represent a cluster of closely related sequences presumably representing members of the same species, including *Conotrachelus inexplicatus* (two individuals), *Conotrachelus* sp. 44 (two individuals), *Conotrachelus* sp. 158 (three individuals), Cryptorhynchinae sp. 71 (five individuals), Cryptorhynchinae sp. 72 (two individuals), Cryptorhynchinae sp. 82 (two individuals), Cryptorhynchinae sp. 85 (three individuals), Curculionidae sp. 3 (two individuals), Hylobiini sp. 145 (three individuals), *Phelypera* cf. *distigma* (five individuals), *Sibinia* sp. 1(two individuals), *Taphroderes* cf. sp. 114 (two individuals), and *Zygops* sp. 1(three individuals). Multiple host records were obtained only for Cryptorhynchinae sp. 71 (Piperaceae and Rubiaceae) and Curculionidae sp. 3 (both individuals feeding on the same host, Celastraceae). Support values on internodes represent posterior probabilities of the Bayesian analysis.

several cases individuals clustered closely together, suggesting that they were members of a single species (Hebert & Gregory, 2005; Pons *et al*., 2006). This was supported by their morphological similarity in each case where adults were available. Using only one individual per sequence cluster reduced the total from 115 to 87 individuals in the phylogenetic analysis of the weevils (see legend to Fig. 1 for details). Yet, beyond these intraspecific clusters the long terminal branches in the clock-constrained tree suggested a great phylogenetic divergence in this set of specimens.

Diversity of host-plant sequences

PCR amplification with the *trnL* intron primers on the 115 extracts producing *cox1* sequences was successful in 41 cases (35.6%). Failure of amplification did not show any apparent bias across the weevil groups represented, although was mostly associated with individuals collected in flight intercept traps, suggesting that host DNA in these individuals is not optimal for amplification. Sequencing in each of the 41 positive cases resulted in clear sequence reads that were confirmed as authentic plant DNA sequences by high GenBank matches with angiosperm *trnL* sequences. The sequences obtained ranged in length from 379 to 609 bp (including terminal missing data) and showed several long indels that were generally synapomorphic for groups at the family level. After alignment the data matrix included 854 positions, of which 514 were variable and 356 parsimony informative. Average divergence (uncorrected *p* distances) between any pair of *trnL* sequences was 17.5%, ranging from 0 (sequences obtained from two specimens of *Conotrachelus* sp. 44 reared from *Eugenia galalonensis*) to 31% (obtained from Brentidae sp. 120 associated with Poaceae and Brentidae sp. 119 feeding on Meliaceae). Several other host sequences differed by 1–2 bp, i.e. less than 1% (Table 2).

Genetic distances between the *trnL* sequences and their closest GenBank hit ranged from 0 to 10.7% (K2P corrected distances) (Table 1). The phylogenetic position for each query sequence relative to those available from GenBank could be established to various levels of accuracy, as taxon coverage and marker informativeness were not universally good. Phylogenetic analyses of each *trnL* sequence together with their respective top 100 GenBank entries recovered these sequences as members of well-established clades. The lowest taxonomic level to which the sequences could be assigned was family, tribe, and genus level in 5, 14, and 23 cases, respectively (Table 1). In most of these searches focal taxa were monophyletic, demonstrating the power of the *trnL* locus to place sequences according to established taxonomic groupings. Where the taxonomic groups were not recovered as monophyletic, as in the genus *Acacia* (Fabaceae), their paraphyly had already been established previously (Maslin *et al*., 2003; Miller *et al*., 2003) (Fig. 2). The *trnL* intron based inference was consistent with the morphological host-plant identification in the reared specimens, although due to the incomplete host database the corroboration was to family level only in several cases (Table 1).

We were specifically interested in the utility of the *trnL* intron to discriminate between close relatives. The literature on DNA barcoding has already established that different species may exhibit identical sequences (Lahaye *et al*., 2008; Hollingsworth *et al*., 2009). This was confirmed here, e.g. in reference sequences of the family Fabaceae the GenBank entries for up to six species were identical. In our sample, only a single case of the weevil *Sibinia* sp. 1 reared from fruits that were morphologically identified as *Mimosa pigra* L. showed complete identity with a GenBank entry, also designated as *M. pigra*. The closest relative in GenBank, *M. tweedieana* was different by two base pairs (Fig. 2), indicating the high discriminatory power of this locus and the precision of the DNA-based identification in the genus *Mimosa*. In others, e.g. the closely related *Conotrachelus* sp. 44 and *C*. sp. 9 reared from *Eugenia nesiotica* Standl. and *E. galalonensis* (Wright), respectively, host sequences were distinguishable by nine nucleotide changes, while the two records from *E. galalonensis* showed identical sequences. However, two specimens of Curculionidae sp. 3 that were reared from fruits of the same species (an unclassified species in the family Celastraceae) differed by two base pairs, suggesting intraspecific variation of the *trnL* intron. Similarly, several wild-caught individuals (i.e. of unknown host plant) produced closely related sequences also (1 or 2 bp divergences; Table 2). These slightly divergent sequences may reflect variation within a single species, or indeed constitute evidence for feeding on closely related (sympatric) species that differ slightly at this locus.

Phylogenetic structuring of host use and floristic implications

In total, the set of 41 sequences could be ascribed to 26 plant families. All *trnL* sequences were used to build a phylogenetic tree of the host plants, which showed general agreement with known angiosperm relationships (Bremer *et al*., 2003), including the basal split of monocots, magnoliids, and all other angiosperms, the asterids, and groups within the eurosids I and II (which both were split in two subclades) (Fig. 3). When mapped on the tree of curculionoids, the host-plant tree showed a high level of incongruence indicating a general lack of host conservation at deeper levels (Fig. 4). Nearer to the tips, there was a small clade of *Conotrachelus* sp. 44 and *Conotrachelus* sp. 9 that was congruent (Fig. 4), but major shifts in host association were evident even at the species level, as the two representatives of Cryptorhynchinae sp. 71 fed on the distant families Piperaceae and Rubiaceae. Vice versa, where closely related plant sequences were obtained from two or three weevils, these were highly divergent (Table 2) and their mean level of divergence did not differ greatly from the average distances of all weevils (Table 2).

Host plants inferred from the *trnL* intron sequences were encountered that had not been reported from the collecting site. For example, *M. pigra* is not known from BCI, although *M. tenuiflora* (Willd.) Poir. does occur in the area. The latter species was included in the analysis but was positioned in a distant clade, confirming the presence of *M. pigra* (or a very closely related species that is indistinguishable in *trnL* intron) existing in BCI. The same applies to *Conotrachelus* sp. 1, whose host sequence matched a GenBank entry attributed

Table 1. Specimens from the superfamily Curculionoidea with *trnL* inferred hosts.

					Host inferred				
Subfamily	BMNH	Species	Site	Collection method	Family	Lowest level	PP	% K2P	Hypothesis based on local records and plant systematic literature
Brentidae	747250	Brentidae sp. 119	BCI	Manual	Meliaceae	Sister to *Malleastrum* sp.	0.54	0.8	*Trichilia* spp. (9 spp.)
	747443	Brentidae sp. 120	BCI	Manual	Poaceae	Sister to *Agrostis capillaris*	0.92	0	*Agrostis* spp. (4 spp.)
	747323	Brentidae sp. 124	BCI	Manual	Euphorbiaceae	Croton clade	1	1.4	*Croton billbergianus*
Trachelizinae	747322	*Teramocerus* sp. 118	BCI	Manual	Meliaceae	Sister to *Malleastrum* sp.	0.65	0.6	*Trichilia* spp. (9 spp.)
Dryophthoridae	796478	Dryophthoridae sp. 127	BCI	FIT	Sapindaceae	Sister to *Serjania altissima*	0.95	1.6	*Serjania* spp. (9 spp.)
Curculionidae	793058	Curculionidae sp. 1	SEL	Manual	Moraceae	*Brosimum* clade	0.99	1.8	*Dorstenia* (2 spp.) or *Brosimum* (3 spp.)
	796412	Curculionidae sp. 2	BCI	Manual	Theaceae	Sister to *Camellia japonica*		0.6	*Gordonia* (2 spp.)
	826571	Curculionidae sp. 3*	CAM	Reared	Celastraceae	Sister to *Loeseneriella africana*		1.9	Celastraceae
	826574	Curculionidae sp. 3*	CAM	Reared	Celastraceae	Sister to *Loeseneriella africana*		1.6	Celastraceae
Baridinae	793067	*Solaria curtula*	SEL	Reared	Sister to Dilleniaceae	Sister to *Tetracera asiatica*	1	5.2	Sister to Dilleniaceae
Conoderinae	826540	*Lechriops* sp. 1	FOR	Reared	Myrsinaceae	Sister to *Ardisia speciosa*	0.83	1.6	*Ardisia* spp.
Cryptorhynchinae	747260	*Coelosternus* cf. sp. 39	BCI	Manual	Fabaceae: Mimosoideae	Sister to *Albizia*		1.1	Tribe Ingeae
	747622	Cryptorhynchinae sp. 4	BCI	FIT	Rubiaceae	Sister to *Faramea multiflora*	1	2.9	*Faramea* (10 spp.) or *Coussarea* (4 spp.)
	747286	Cryptorhynchinae sp. 24	BCI	Manual	Urticaceae	*Pilea ternifolia, P. melastomoides* clade	1	3.2	*Pilea* spp.
	796424	Cryptorhynchinae sp. 34	BCI	Manual	Musaceae	Basal to *Musa* clade	0.98	0.2	*Musa sapientum*
	747320	Cryptorhynchinae sp. 69	BCI	Manual	Brassicaceae	*Arabis alpina* clade		0.6	*Cardamine flexuosa*
	826691	Cryptorhynchinae sp. 71	BCI	Manual	Rubiaceae	Basal to *Hallea* and *Mitragyna*		0.3	Naucleeae species
	826713	Cryptorhynchinae sp. 71	BCI	Manual	Piperaceae	*Piper aequale, P. urophyllum* clade	1	2	*Piper aequale* clade
	826712	Cryptorhynchinae sp. 77	BCI	Manual	Fabaceae	Ingeae tribe		1.3	Tribe Ingeae
	747370	Cryptorhynchinae sp. 83	BCI	Manual	Brassicaceae	Sister to *Arabis alpina*		0.3	*Cardamine flexuosa*
	796489	*Cryptorhynchus* sp. 111	BCI	Manual	Melastomataceae	Sister to *Adelobotrys*	0.96	2.5	Merianieae, Miconieae, Bertolonieae or Blakeeae
	747380	*Elytrocoptus* sp. 87	BCI	Manual	Sapindaceae	Basal to *Matayba* and *Scyphonychium*		1.2	*Allophylus, Paullinia, Sapindus, Thinouia*
	747261	*Eubulus* cf. sp. 41	BCI	Manual	Fabaceae: Mimosoideae	Sister to Albizia		1.4	Tribe Ingeae
Curculioninae	826065	*Sibinia* sp.*	BCI	Reared	Fabaceae: Mimosoideae	Basal to *Mimosa pigra* and *M. tweedieana*	0.95	0.1	*Mimosa pellita* var. *pellita* (=*Mimosa pigra*)
	747414	*Sicoderus* cf. *delusor*	BCI	Manual	Brassicaceae	Sister to *Arabis alpina*		0.9	*Cardamine flexuosa*
Entiminae	796343	Entiminae sp. 2	CAM	Manual	Melastomataceae	Basal to Adelobotrys, Melastoma and Tibouchina	1	6.5	Memecylaceae, Crypteroniaceae, Alzateaceae, Rhynchocalycaceae, Penaeaceae, or Oliniaceae
	796276	Entiminae sp. 3	FOR	Manual	Asteraceae	Asteroideae	1	0.7	Asteroideae
	747283	Entiminae sp. 4	BCI	Manual	Malvaceae	Bombacoidae clade	0.9	1	*Cavanillensia, Pseudobombax, Ceiba, Gyranthera*

Table 1. *Continued*

Subfamily	BMNH	Species	Site	Collection method	Host inferred				
					Family	Lowest level	PP	% K2P	Hypothesis based on local records and plant systematic literature
Molytinae	793141	Cholini sp. 195	SEL	Manual	Marantaceae	Basal to *Ataenidia, Marantocloa* and *Phacelophrynium*	0.95	2.9	Marantaceae
	826663	*Conotrachelus inexplicatus* Faust	BCI	FIT	Rubiaceae	Basal to *Calycophyllum*		1	Cinchonoideae
	793105	*Conotrachelus* sp. 1†	SEL	Manual	Sapotaceae	Sister to *Pouteria vernicosa*		0	*Pouteria* spp.
	826553	*Conotrachelus* sp. 8∗	BCI	Reared	Basal to Salicaceae	Basal to Salicaceae	0.7	1.7	Lacistemataceae, Turneraceae, Passifloraceae
	826118	*Conotrachelus* sp. 9∗	BCI	Reared	Myrtaceae	Sister to *Leptospermum scoparium*		2.4	Chamelaucieae, Lindsayomyrteae, Leptospermeae, Syncarpieae and Myrteae
	825983	*Conotrachelus* sp. 44∗	BCI	Reared	Myrtaceae	Sister to *Leptospermum scoparium*		3.6	Chamelaucieae, Lindsayomyrteae, Leptospermeae, Syncarpieae and Myrteae
	825984	*Conotrachelus* sp. 44∗	BCI	Reared	Myrtaceae	Sister to *Leptospermum scoparium*		3.6	Chamelaucieae, Lindsayomyrteae, Leptospermeae, Syncarpieae and Myrteae
	796425	*Conotrachelus* sp. 158	BCI	Manual	Lauraceae	Sister to *Persea americana*		1.1	sister to *Persea americana*
	796287	*Conotrachelus* sp. 162	BCI	Manual	Arecaceae	Sister to *Satakentia liukiuensis*		0	Arecoideae
	793143	Hylobiini sp. 1	SEL	Manual	Urticaceae	Basal to *Cecropia*	1	6.1	Urticaceae
	747365	Hylobiini sp. 137	BCI	Manual	Anacardiaceae	Sister to *Anacardium occidental* E	1	1	*Anacardium excelsum, A. occidentale* or *Mangifera indica*
	796415	Hylobiini sp. 139	BCI	Manual	Bignoniaceae	Sister to *Arrabidaea pubescens*	1	0.6	*Arrabidaea* clade
	747437	Molytinae sp. 146	BCI	Manual	Malvaceae	Sister to *Chorisia speciosa*		0.7	*Cavanillensia, Pseudobombax, Ceiba, Gyranthera*

Samples from Panama [Barro Colorado Island (BCI), Fortuna Dam Area (FOR), Altos de Campana National Park (CAM) and San Lorenzo National Park (SHE)] and Costa Rica [La Selva Biological Station (SEL)] were collected by flight intercept traps (FIT), manual collection and control specimens reared from host plants. The host associations based on rearing are as follows: *Conotrachelus* sp. 8 from Flacourtiaceae, *Conotrachelus* sp. 9 and *C*. sp. 44 were reared from Myrtaceae (*Eugenia nesiotica* and *E. galalonensis*, respectively), Curculionidae sp. 3 from Celestraceae, *Lechriops* sp. 1 from *Ardisia* sp. (Myrsinaceae), *Sibinia* sp. from *Mimosa pellita* = *Mimosa pigra* (Fabaceae: Mimosoideae), *Solaria curtula* from *Davilla nitida* (Dilleniaceae). †*Conotrachelus* sp. 1 was collected wandering on fruits of *Inga alba*. All specimens used for extraction were adults except seven larvae that were reared from fruits (*). K2P = genetic distance calculated with the Kimura 2-Parameter model. PP = Posterior probability of query sequence to the nearest identified node. BMNH = Sample number of specimens in the British Museum of Natural History.

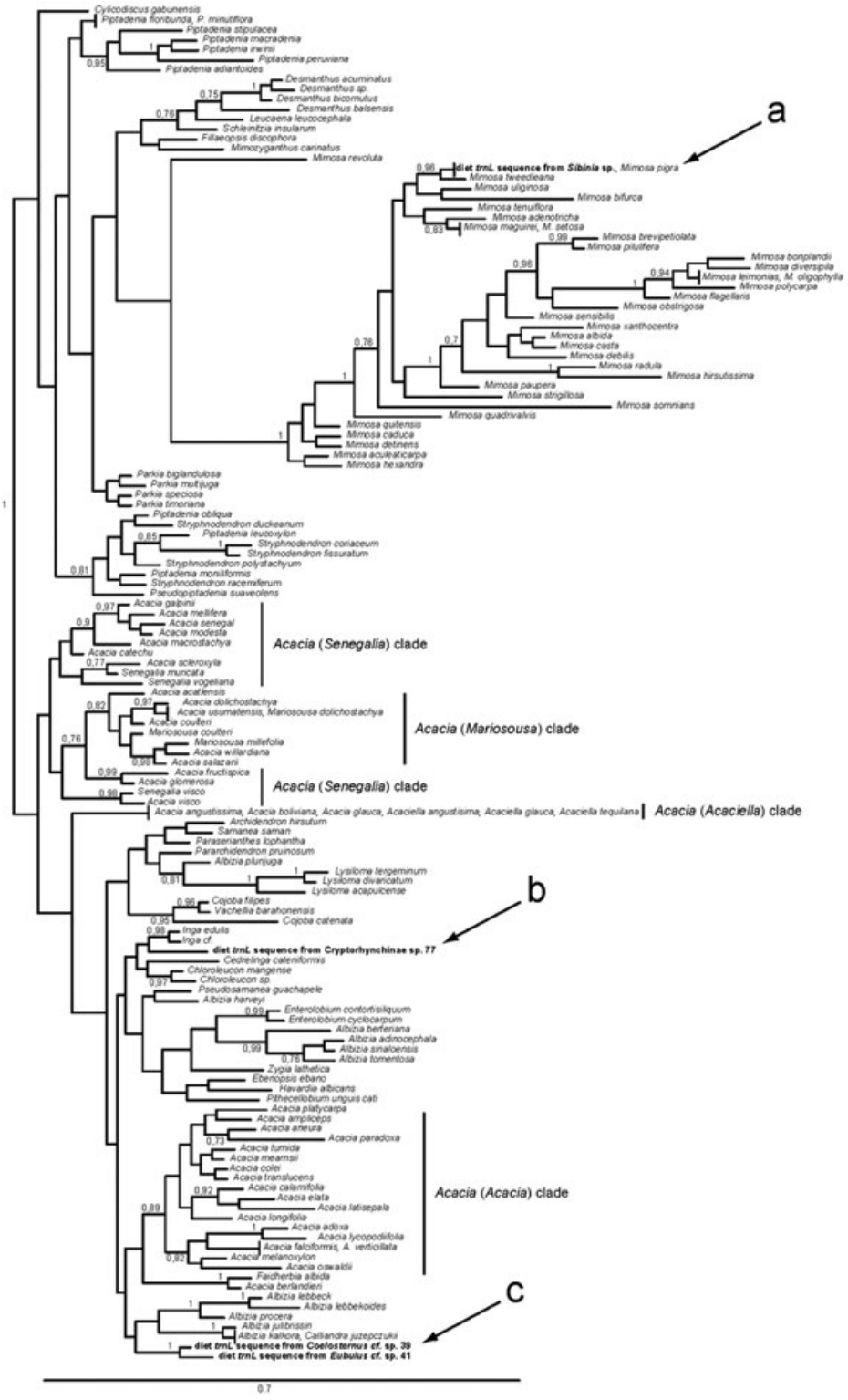

Fig. 2. An example of phylogenetic analyses used for identification of host species. A set of *trnL* intron sequences of Fabaceae obtained as the top hits on GenBank against four queries extracted from beetle tissue was aligned and subjected to Bayesian tree searches. Major clades of Fabaceae, including established grouping that contradict current genus names, are marked. The species of weevils from which the sequences were obtained are: (a) *Sibinia* sp. feeding on *Mimosa pigra*, (b) Cryptorhynchinae sp. 77 on Ingeae, and (c) *Coelosternus* cf. sp. 39 and *Eubulus* cf. sp. 41 feeding on Ingeae.

Table 2. Pairs of closely related host sequences, identified according to phylogenetic analysis of closely related GenBank records, and their level of divergence in the *trnL* intron marker.

Identification	bp difference (uncorr. *P*)	Herbivore 1	Herbivore 2	Uncorr. *P* (*cox1*)	Locality
Sister to *Malleastrum* sp.	1 (0.00174)	Brentidae sp. 119	*Teramocerus* sp. 118	0.13263	BCI
Bombacoideae clade	1 (0.00153)	Entiminae sp. 4	Molytinae sp. 146	0.17029	BCI
Celastraceae	2 (0.00337)	Curculionidae sp. 3∗	Curculionidae sp. 3*	0	CAM
Sister to *Albizia*	2 (0.00348)	*Coelosternus* cf. sp. 39	*Eubulus* cf. sp. 41	0.13008	BCI
Sister to *Leptospermum scoparium*	8 (0.01427)	*Conotrachelus* sp. 9∗	*Conotrachelus* sp. 44∗	0.1402	BCI
Sister to *Leptospermum scoparium*	0	*Conotrachelus* sp. 44∗	*Conotrachelus* sp. 44∗	0	BCI
Sister to *Leptospermum scoparium*	9 (0.01586)	*Conotrachelus* sp. 9	*Conotrachelus* sp. 44∗	0.13974	BCI
Sister to *Arabis alpina*	2 (0.00510)	Cryptorhynchinae sp. 69	Cryptorhynchinae sp. 83	0.22667	BCI
Sister to *Arabis alpina*	2 (0.00510)	Cryptorhynchinae sp. 83	*Sicoderus* cf. *delusor*	0.21284	BCI
Arabis alpina clade	2 (0.00512)	*Sicoderus* cf. *delusor*	Cryptorhynchinae sp. 69	0.16987	BCI

Also listed are the two respective herbivores and their uncorrected sequence divergence. Reared specimens are marked by an asterisk
BCI, Barro Colorado Island; CAM, Altos de Campana National Park.

to *Pouteria vernicosa* T.D. Pennington. This specimen was collected in La Selva, Costa Rica where species plant lists show no record of this species but report 11 other species of the genus *Pouteria*. Likewise, the genus *Pilea* (Urticaceae), the inferred host of Cryptorhynchinae sp. 24, has not been reported for BCI, although it is known from the surrounding Panama Canal Area. The host of Brentidae sp. 120 was inferred to be a close relative of the genera *Agrostis*, *Calamagrostis*, or *Poa* (Poaceae), none of which has been cited as occurring at BCI, although they are known from the wider area.

Discussion

The diversity of insect herbivores remains a major challenge to the understanding of species richness and functioning of tropical forests. However, the difficulty in the establishment of host associations hampers the study of plant–herbivore interactions and their role in promoting tropical species richness. A recent review concluded: 'For establishing feeding association, we see no viable alternatives to experimental feeding trials or direct feeding and rearing records' (Lewinsohn & Roslin, 2008). The proof of DNA-based plant identification from herbivore tissue (Jurado-Rivera *et al*., 2009) now provides a novel method with great potential for the study of rainforests, as the most complex plant–herbivore assemblages on Earth, that will resolve the long-standing questions about the factors promoting species diversity. Although based on a very small sample of herbivores, the current study demonstrates the great potential of this procedure for determining host associations and ultimately diet breadth of tropical insects.

Identification success and phylogenetic information content of trnL

Accurate identification of the host is affected by the completeness of the reference DNA database, as well as the discriminatory power of the locus used for sequencing. We chose the *trnL* intron mainly because it has the highest level of coverage in GenBank among potential barcoding markers; the universality of PCR primers; its good PCR success with degraded DNA (Taberlet *et al*., 2007); and its power in phylogenetic analysis across a range of hierarchical levels (Bremer *et al*., 2002; Shaw *et al*., 2005). The latter is due to the presence of conserved and highly variable regions and the relative ease of alignment. Whereas no comparisons were conducted with other potential markers (Hollingsworth *et al*., 2009), the *trnL* intron sequences were sufficient to discriminate among all individuals found in these forests, with the single exception of two individuals of *Conotrachelus* sp. 44 which were reared from a single host species. The *trnL* locus discriminated between congenerics based on several nucleotide changes (e.g. *E. nesiotica* and *E. galalonensis*; Table 2), while there was no intraspecific variation (*E. galalonensis*–*E. galalonensis*). However, discrimination at the species level was not universal, e.g. in the genus *Acacia* where up to six species are indistinguishable using this marker (see Fig. 2). In turn, we observed *trnL* variation *within* a morphologically defined host species in the two host records for Curculionidae sp. 3 (both reared from an unidentified species of Celastraceae). This intraspecific variation for the locus is not unprecedented and has been reported previously (Taberlet *et al*., 2007; Tsai *et al*., 2008). However, it does mean that in examples of 1- or 2-bp differences in *trnL* sequences we do not know whether this is associated with intraspecific or interspecific variation without being able to identify the plants in another way. Hence, as there is no clear 'barcoding gap' as that seen in mtDNA of animals (Meyer & Paulay, 2005), the use of *trnL* intron to resolve taxonomic identifications near the species level remains limited (see also Chase *et al*., 2007). Possibly this problem can be overcome with other chloroplast markers exhibiting faster substitution rates (Lahaye *et al*., 2008) or a set of markers (Tsai *et al*., 2008; Hollingsworth *et al*., 2009). However, the discrimination of host races and very closely related isolates (e.g. Hebert *et al*., 2004) may remain problematic with the use of chloroplast markers.

The problem of low interspecific variation was relevant for identifications in the current sample only in a few cases, as coverage in GenBank was generally not sufficient to provide

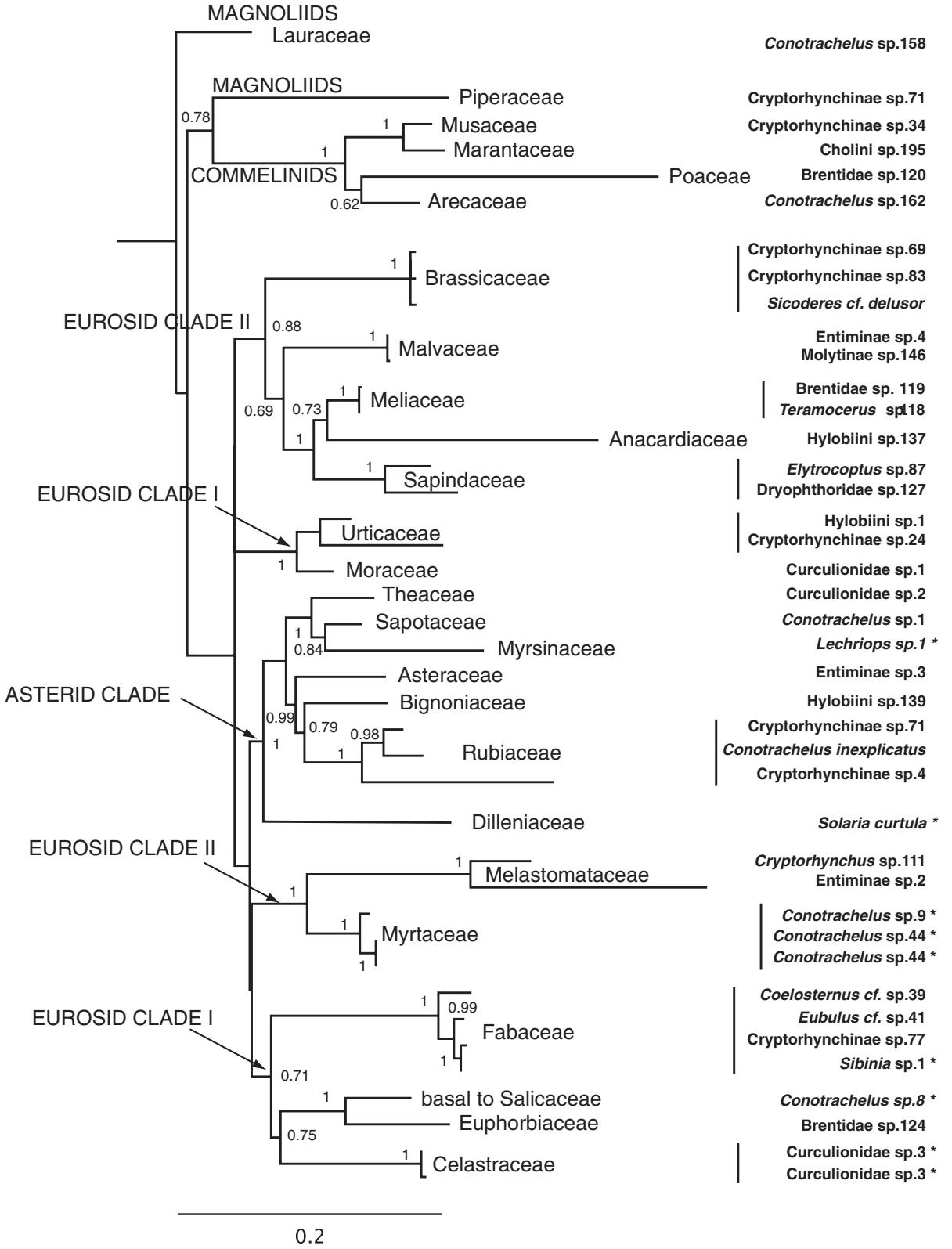

Fig. 3. Maximum likelihood tree based on *trnL* intron sequences retrieved from weevil tissue. Outgroups used (not shown) were *Cycas siamensis*, *Gynko biloba*, and *Pseudotsuga menziensii*. Terminals correspond to inferred plant families and names on the right represent the studied beetle taxa, including reared specimens in larval stages (*). Support values are shown for nodes with PP > 60 only. PP = posterior probabilities.

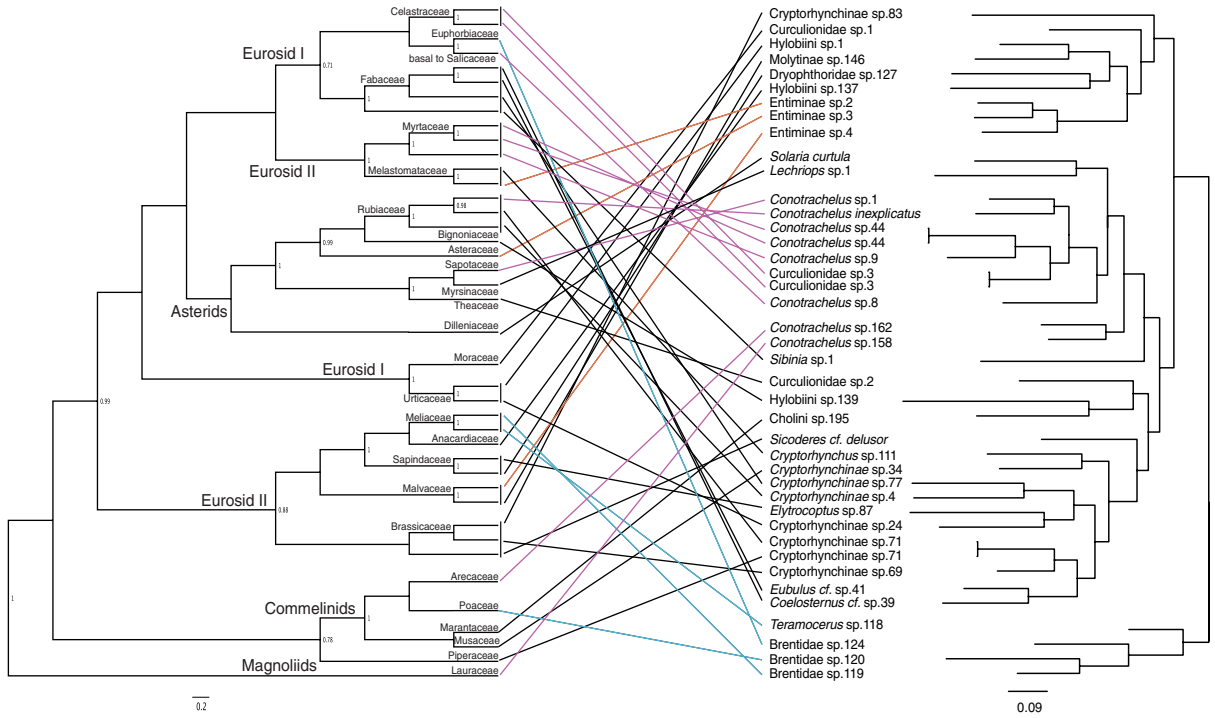

Fig. 4. Tanglegram of the host-plant tree inferred from *trnL* intron and herbivore tree from *cox1* sequences. Species of *Conotrachelus* (pink lines), subfamily Entiminae (orange lines) and Brentidae (blue lines) show broad host ranges. Species of the genus *Conotrachelus* (pink lines), the subfamily Entiminae (orange lines) and the family Brentidae (blue lines) showing broad host ranges.

matches or near matches to sequences obtained here. Instead, phylogenetic inferences relative to the top 100 GenBank entries had to be used to place the query sequence. Frequently, host sequences could only be assigned to a plant family or subfamily, in particular when applying our rather stringent criteria for identification (based on the Bayesian inference) as a group member that required high support levels of a subtending node. The *trnL* intron marker proved surprisingly powerful for phylogenetic reconstruction of relationships of host plant lineages over a wide hierarchical range. The locus separated major groups of flowering plants (Fig. 3) and established genus-level relationships in accordance with recent DNA-based studies (Fig. 2). This was important to provide accurate high-level identifications (Jurado-Rivera *et al*., 2009), which were confirmed where these inferences were applied to reared specimens of known host associations (Table 1). With greater taxonomic coverage, the *trnL* locus provided increased precision of identification, as in the case of the densely sampled genus *Mimosa* that permitted species-level identification (Fig. 2). At the current state of the databases, low taxon coverage of *trnL* intron sequences, rather than short fragment length or limited sequence variability, had the greatest impact on identification success.

Host identification will gradually become more accurate, as database content accumulates from studies of host use (Jurado-Rivera *et al*., 2009), DNA barcoding (Taberlet *et al*., 2007), and phylogenetic analysis (Bremer *et al*., 2002; Shaw *et al*., 2005). Eventually, even species-level questions may be addressed with greater accuracy once DNA-based analyses of species boundaries and intraspecific variation have been conducted. In addition, a combined set of cpDNA and other markers (including single-copy nuclear markers) will be amplifiable from the herbivore tissue also. In addition to improving species discrimination, such multiple-marker system will also address discrimination for those lineages with poor taxon coverage in the *trn*L intron. Therefore, a narrow focus on levels of variation in developing DNA barcodes (Lahaye *et al*., 2008) should not ignore the need for phylogenetic power that ultimately will put sequences in the context of others. Phylogenetics will be important for identification of taxa not represented in the database as well as for evolutionary analyses of ecological, biogeographic, behavioural and other data that may be associated with the individuals from which the sequences were obtained, or for co-evolutionary analyses of plants and their herbivores or pathogens (also see Jurado-Rivera *et al*., 2009).

Composition of the sample and its relevance for larger DNA-based surveys of host plants

The *trnL* intron sequence provided a feeding record for a particular individual. While such individual records do not directly investigate diet breadth, as feeding records build up, they will provide the spectrum of potential host plants from which diet breadth can be derived. Quantitative sampling will also indicate food preferences within the host range.

The DNA largely conveys information on the most recent feeding episode. It is known that ingested DNA decreases exponentially in abundance over a period of 4–50 h (Chen *et al*., 2000; Greenstone *et al*., 2007; King *et al*., 2008) and therefore obtaining an individual at different times during the life cycle or feeding cycle might have resulted in a different host association. This is also of significance for trap catches, as the amplification of plant DNA from trap catches failed frequently in this study. It is possible (although not tested) that the plant DNA in the insect gut degrades over the period (sometimes of several days) that the insect spends in the comparatively weak preservative in the trap.

The current study demonstrated the interest of findings from the technique. First, from each of our samples only a single sequence per individual was obtained, i.e. there was no intra-individual variation and hence these feeding records were unequivocal for that individual without the need for cloning of PCR products. Second, weevils exploited a broad range of host-plant families showing a high trophic diversity across most major groups of angiosperms. While some host sequences were closely related (Table 2) and might indeed be from the same host species, the sequence variation still indicates a diversity of host trees and populations. Third, host records may be highly variable for a species when more than one individual was available for analyses (e.g. results for Cryptorhynchinae sp. 71), but host conservation derived from high sequence similarity is typically found (e.g. *Conotrachelus* sp. 44 and its sister species *Conotrachelus* sp. 9, or Curculionidae sp. 3). Fourth, host specificity was not phylogenetically conserved in the studied group of weevils, as sister taxa usually feed on different plant families, and generally very distant ones phylogenetically, and reciprocally with similar host plants being used by highly divergent herbivores (Fig. 3).

How much these host records reflect the diet breadth of these herbivore species, and to what extent this demonstrates divergent larval and adult host preferences and diet breadth remains to be assessed based on the study of further specimens. In this respect, currently larval host associations are hypothesised to be more restricted than those of adults (Novotny & Basset, 2005). While our ecological knowledge is limited regarding the breadth of adult versus larval hosts and variation among subsequent food intakes, DNA-based feeding records will be able to build an increasingly complete image of feeding behaviour from comprehensive sampling of different stages representing particular species across seasonal and geographical ranges. The cases encountered here illustrate extremes in the spectrum of host specificities, which may be narrow in some species, and (phylogenetically) broad in others (Symons & Beccaloni, 1999). Increased taxonomic coverage and larger sample sizes per species will make it possible to quantify host specificity in the future, from monophagous to oligophagous or polyphagous and from evolutionarily conserved to labile, while also taking into account local differences in response to beta-diversity of host plant assemblages (e.g. Condon *et al*., 2008).

Finally, the DNA-based feeding records may reveal discrepancies in collecting site and actual food plants. For example, our sample of *Conotrachelus* sp. 1 was obtained from an individual walking on fruits of *Inga alba* (Fabaceae), a frequent host for this genus, but the DNA-based inference of feeding source was the distantly related asterid Sapotaceae (also a known host of *Conotrachelus*). In addition, the comparison of species encountered at a site may be compared with floristic inventories, which already pointed to the presence of unreported plant species in an area. The host records can potentially be used to discriminate between host trees of different genotypes at the population level, providing a measure of dispersal of the insects between rain forest sites.

Phylogenetic information content of cox1 and coevolutionary relationships

The use of *cox1* in identification and species delimitation in insects is well established (Hebert *et al*., 2004; Pons *et al*., 2006). Levels of sequence divergence in *cox1* near the species level were much higher than in the *trnL* intron, and detected intraspecific variation as clusters of closely related sequences. The *cox1* marker was similarly powerful in separate deep-level groups (families and subfamilies) in Curculionoidea (Fig. 1) and resolved relationships within those, e.g. recovering the basal branching pattern in *Conotrachelus.* Sequence divergences in mitochondrial *cox1* and chloroplast *trnL* were more similar at deeper levels, presumably because mtDNA is affected by saturation of nucleotide variation. This may also compromise its power at basal levels of the tree, e.g. resulting in the failure to recover Dryophthoridae outside of the Curculionidae [which is the main incongruence of the tree with the existing classification, although a multi-gene data set also found this position for Dryophthoridae (Hunt *et al*., 2007; McKenna *et al*., 2009)]. However, the precise resolution of basal relationships of Curculionoidea were not of great concern for the current study, as coevolutionary analyses (below) were mostly affected by host switches nearer the tips of the trees. Therefore, more critical than adding markers (which potentially could be obtained from existing phylogenetic datasets) the power of the current analyses would be improved with denser taxonomic sampling.

Virtually all of the terminal branches of the curculionid tree were very long relative to the internal branches, indicating that clade diversity was not captured to any degree of completion. According to the dating procedure all pairs of sister taxa were separated by >20 Mya. The absolute calibration using the universal 2.3% divergence My^{-1} is problematic and its utility needs to be confirmed for dating older nodes, while confidence intervals may be large given the slow convergence of the MCMC chain. It is noteworthy, however, that the basal split of Brentidae and Curculionidae estimated to 150 Mya was in almost perfect agreement with a recent age estimation of this node considering multiple genes and fossil calibration across the Curculionoidea (McKenna *et al*., 2009). Therefore, deep branches of the tree obtained here, e.g. resolving the basal lineages in Cryptorhynchinae, are ancient and probably represent many thousands of species of that subfamily missing from the trees. Likewise, the genus *Conotrachelus* represented here by just eight species includes

more than 1200 described from the Neotropics and many more undescribed species. Inevitably the incomplete taxon sampling limits the evolutionary analysis of character changes, e.g. to study the rate of host switches and conservation within clades. However, it is clear that host plant–herbivore interactions in the rainforest assemblage are evolutionarily fluid. Although the estimates for appearance of the earliest crown group angiosperms at >140 Mya (Friis *et al*., 2006) are similar to the estimated age of the weevils, host associations are unlikely to be ancient, and host shifts and lags in diversification of herbivores are well established (Lopez-Vaamonde *et al*., 2006; Gómez-Zurita *et al*., 2007; Hunt *et al*., 2007; McKenna *et al*., 2009). However, much denser taxon sampling is required to assess the frequency and step size of host shifts.

Conclusions

Ingested chloroplast DNA that is obtained from the insect specimens using standard extraction protocols provides the information on host records necessary for evolutionary and ecological analyses of herbivore–host plant interactions (Jurado-Rivera *et al*., 2009). The current study assesses the utility of this method when applied to a highly complex rainforest assemblage. Evident limitations for host identification are the insufficient local host database and the lack of clearly known species limits in the hosts, i.e. the *trnL* sequences do not always provide an unequivocal host record. While future increase of the database and use of additional chloroplast markers will improve the accuracy, some apparent limitations of host plant inferences are independent of the study method. Feeding studies of tropical forest assemblages (e.g. Barone, 2000; Novotny *et al*., 2002) to date have not questioned taxonomy and species limits of host plants, nor assessed population differentiation and geographic turnover. Due to their greater resolution, DNA-based analyses can contribute vital information on host populations and spatial differentiation of host use. The technique therefore permits the reinvestigation of pertinent hypotheses explaining tropical forest diversity, e.g. regarding the greater trophic specialisation and higher species diversity in the tropics (Coley & Barone, 1996), density-dependent factors to maintain high diversity (Janzen, 1970), or the correlation of herbivore diversity with phylogenetic diversity of host plants (Novotny *et al*., 2006). Solid data on the host-plant use and host specificity are a necessity for testing any of these hypotheses.

Acknowledgements

We are grateful to Anna Papadopoulou for her important input to this work. We thank Osvaldo Calderon and Salomon Aguilar for their help with botanical information and plant identifications. We also thank Hugh Loxdale and Michael Claridge for suggesting the topic of our symposium contribution. We are grateful to the STRI staff who helped with facilities and ANAM collection permits to S. P. N. We thank Dr Richard Thompson for his help with weevil identification. This work has been supported by the studentships from SENACYT-Panama, CONACYT-Mexico (196749) and the European Union-Alban (E05D058103PA) to S. P. N.

J. J. R. received support from the SYNTHESYS grant (GB-TAF-4394) of the European Community Research Infrastructure and J.G-Z. from the Spanish "Fundación BBVA" (project >IN2-BST-Nicaragua).

References

Alonso Zarazaga, M.A. & Lyal, C.H.C. (1999) *A World Catalogue of families and genera of Curculionoidea (Insecta: Coleoptera) (excepting Scolytidae and Platypodidae). Entomopraxis*, Barcelona, Spain.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Barone, J.A. (1998) Host-specificity of folivorous insects in a moist tropical forest. *Journal of Animal Ecology*, **67**, 400–409.

Barone, J.A. (2000) Comparison of herbivores and herbivory in the canopy and understory for two tropical tree species. *Biotropica*, **32**, 307–317.

Bremer, B., Bremer, K., Chase, M.W., Reveal, J.L., Soltis, D.E.Soltis, P.S. *et al*. (2003) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society*, **141**, 399–436.

Bremer, B., Bremer, K., Heidari, N., Erixon, P., Olmstead, R.G., Anderberg, A.A. *et al*. (2002) Phylogenetics of asterids based on 3 coding and 3 non-coding chloroplast DNA markers and the utility of non-coding DNA at higher taxonomic levels. *Molecular Phylogenetics and Evolution*, **24**, 274–301.

Brower, A.V.Z. (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.

Chase, M.W., Cowan, R.S., Hollingsworth, P.M., van den Berg, C., Madrinan, S., Petersen, G. *et al*. (2007) A proposal for a standardised protocol to barcode all land plants. *Taxon*, **56**, 295–299.

Chen, L. & Fadamiro, H.Y. (2007) Differential electroantennogram response of females and males of two parasitoid species to host-related green leaf volatiles and inducible compounds. *Bulletin of Entomological Research*, **97**, 515–522.

Chen, Y., Giles, K.L., Payton, M.E. & Greenstone, M.H. (2000) Identifying key cereal aphid predators by molecular gut analysis. *Molecular Ecology*, **9**, 1887–1898.

Coley, P.D. & Barone, J.A. (1996) Herbivory and plant defenses in tropical forests. *Annual Review of Ecology and Systematics*, **27**, 305–335.

Condon, M.A., Scheffer, S.J., Lewis, M.L. & Swensen, S.M. (2008) Hidden neotropical diversity: greater than the sum of its parts. *Science*, **320**, 928–931.

Croat, T.B. (1978) *Flora of Barro Colorado Island*. Stanford University Press, Stanford, California.

De Boer, G. & Hanson, F.E. (1984) Foodplant selection and induction of feeding preference among host and non-host plants in larvae of the Tobacco Hornworm *Manduca sexta*. *Entomologia Experimentalis et Applicata*, **35**, 177–193.

Drummond, A.J. & Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *MBC Evolutionary Biology*, **7**: 214. doi:10.1186/1471-2148-7-214.

Dyer, L.A., Singer, M.S., Lill, J.T., Stireman, J.O., Gentry, G.L., Marquis, R.J. *et al*. (2007) Host specificity of Lepidoptera in tropical and temperate forests. *Nature*, **448**, 696–699.

Eichhorn, M.P., Compton, S.G. & Hartley, S.E. (2008) The influence of soil type on rain forest insect herbivore communities. *Biotropica*, **40**, 707–713.

Erwin, T.L. (1982) Tropical forests: their richness in Coleoptera and other arthropod species. *Coleopterists Bulletin*, **36**, 74–75.

Farrell, B.D. (1998) 'Inordinate fondness' explained: why are there so many beetles? *Science*, **281**, 555–559.

Fernandez, P.C., Meiners, T., Bjorkman, C. & Hilker, M. (2007) Electrophysiological responses of the blue willow leaf beetle, *Phratora vulgatissima*, to volatiles of different *Salix viminalis* genotypes. *Entomologia Experimentalis et Applicata*, **125**, 157–164.

Friis, E.M., Pedersen, K.R. & Crane, P.R. (2006) Cretaceous angiosperin flowers: innovation and evolution in plant reproduction. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **232**, 251–293.

Fry, B., Joern, A. & Parker, P.L. (1978) Grasshopper food web analysis: use of carbon isotope ratios to examine feeding relationships among terrestrial herbivores. *Ecology*, **59**, 498–506.

Greenstone, M.H., Rowley, D.L., Weber, D.C., Payton, M.E. & Hawthorne, D.J. (2007) Feeding mode and prey detectability half-lives in molecular gut-content analysis: an example with two predators of the Colorado potato beetle. *Bulletin of Entomological Research*, **97**, 201–209.

Gómez-Zurita, J., Hunt, T., Kopliku, F. & Vogler, A.P. (2007) Recalibrating the tree of Leaf Beetles (Chrysomelidae): independent diversification of angiosperms and their insect herbivores. *PLoS ONE*, **2**, e360.

Hebert, P.D.N. & Gregory, T.R. (2005) The promise of DNA barcoding for taxonomy. *Systematic Biology*, **54**, 852–859.

Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H. & Hallwachs, W. (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 14812–14817.

Hollingsworth, P.M., Forrest, L.L., Spouge, J.L., Hajibabaei, M., Ratnasingham, S., van der Bank, M. *et al*. (2009) A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 12794–12797.

Hunt, T., Bergsten, J., Levkanicova, Z., Papadopoulou, A., John, O.S., Wild, R.R. *et al*. (2007) A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science*, **318**, 1913–1916.

Janzen, D.H. (1970) Herbivores and number of tree species in tropical forests. *American Naturalist*, **104**, 501–528.

Jurado-Rivera, J.A., Vogler, A.P., Reid, C.A.M., Petitpierre, E. & Gómez-Zurita, J. (2009) DNA barcoding insect–hostplant associations. *Proceedings of the Royal Society Series B*, **276**, 639–648.

Katoh, K., Kuma, K., Toh, H. & Miyata, T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research*, **33**, 511–518.

King, R.A., Read, D.S., Traugott, M. & Symondson, W.O.C. (2008) Molecular analysis of predation: a review of best practice for DNA-based approaches. *Molecular Ecology*, **17**, 947–963.

Lahaye, R., van der Bank, M., Bogarin, D., Warner, J., Pupulin, F., Gigot, G. *et al*. (2008) DNA barcoding the floras of biodiversity hotspots. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 2923–2928.

Lee, J.C.J.C. & Bernays, E.A.E.A. (1988) Declining acceptability of a food plant for the polyphagous grasshopper *Schistocerca americana* –the role of food aversion learning. *Physiological Entomology*, **13**, 291–301.

Lewinsohn, T.M. & Roslin, T. (2008) Four ways towards tropical herbivore megadiversity. *Ecology Letters*, **11**, 398–416.

Lewinsohn, T.M., Novotny, V. & Basset, Y. (2005) Insects on plants: diversity of herbivore assemblages revisited. *Annual Review of Ecology, Evolution and Systematics*, **36**, 597–620.

Lopez-Vaamonde, C., Wikstrom, N., Labandeira, C., Godfray, H.C.J., Goodman, S.J. & Cook, J.M. (2006) Fossil-calibrated molecular phylogenies reveal that leaf-mining moths radiated millions of years after their host plants. *Journal of Evolutionary Biology*, **19**, 1314–1326.

Maslin, B.R., Miller, J.T. & Seigler, D.S. (2003) Overview of the generic status of *Acacia* (Leguminosae: Mimosoideae). *Australian Systematic Botany*, **16**, 1–18.

Matheson, C.D., Muller, G.C., Junnila, A., Vernon, K., Hausmann, A., Miller, M.A. *et al*. (2007) A PCR method for detection of plant meals from the guts of insects. *Organisms Diversity and Evolution*, **7**, 294–303.

McKenna, D.D., Sequeira, A.S., Marvaldi, A.E. & Farrell, B.D. (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.

Meyer, C.P. & Paulay, G. (2005) DNA barcoding: Error rates based on comprehensive sampling. *PLoS Biology*, **3**, 2229–2238.

Miller, J.T., Andrew, R. & Bayer, R.J. (2003) Molecular phylogenetics of the Australian acacias of subg. Phyllodineae (Fabaceae: Mimosoideae) based on the *trn*K intron. *Australian Journal of Botany*, **51**, 167–177.

Novotny, V. & Basset, Y. (2000) Rare species in communities of tropical insect herbivores: pondering the mystery of singletons. *Oikos*, **89**, 564–572.

Novotny, V. & Basset, Y. (2005) Review–host specificity of insect herbivores in tropical forests. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 1083–1090.

Novotny, V., Basset, Y., Miller, S.E., Weiblen, G.D., Bremer, B., Cizek, L.L. *et al*. (2002) Low host specificity of herbivorous insects in a tropical forest. *Nature*, **416**, 841–844.

Novotny, V., Drozd, P., Miller, S.E., Kulfan, M., Janda, M., Basset, Y. *et al*. (2006) Why are there so many species of herbivorous insects in tropical rainforests? *Science*, **313**, 1115–1118.

Novotny, V., Miller, S.E., Hulcr, J., Drew, R.A.I., Basset, Y., Janda, M.M. *et al*. (2007) Low beta diversity of herbivorous insects in tropical forests. *Nature*, **448**, 692–695.

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

Otte, D. & Joern, A. (1976) On feeding patterns in desert grasshoppers and the evolution of specialized diets. *Proceedings of the Academy of Natural Sciences of Philadelphia*, **128**, 89–126.

Pons, J., Barraclough, T.G., Gomez-Zurita, J., Cardoso, A., Duran, D.P., Hazell, S. *et al*. (2006) Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, **55**, 595–609.

Posada, D. & Crandall, K.A. (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Post, D.M. (2002) The long and short of food-chain length. *Trends in Ecology & Evolution*, **17**, 269–277.

Ronquist, F. & Huelsenbeck, J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Schlein, Y. & Jacobson, R.L. (1999) Sugar meals and longevity of the sandfly *Phlebotomus papatasi* in an arid focus of *Leishmania major* in the Jordan Valley. *Medical and Veterinary Entomology*, **13**, 65–71.

Schneider, D. (1957) Elektrophysiologische untersuchungen von chemo- und mechanorezetoren der antenne des seidenspinners, *Bombyx mori* L. *Zeitschrift fur Vergleichende Physiologie*, **40**, 8–41.

Shaw, J., Lickey, E.B., Beck, J.T., Farmer, S.B., Liu, W.S., Miller, J.J. *et al.* (2005) The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany*, **92**, 142–166.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. & Flook, P. (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, **87**, 651–701.

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.

Symons, F.B. & Beccaloni, G.W. (1999) Phylogenetic indices for measuring the diet breadths of phytophagous insects. *Oecologia*, **119**, 427–434.

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A. *et al*. (2007) Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Research*, **35**, Article No.: e14.10.1093/nar/gk1938.

Taberlet, P., Gielly, L., Pautou, G. & Bouvet, J. (1991) Universal primers for amplification of 3 noncoding regions of chloroplast DNA. *Plant Molecular Biology*, **17**, 1105–1109.

Tsai, L.C., Wang, J.C., Hsieh, H.M., Liu, K.L., Linacre, A. & Lee, J.C.I. (2008) Bidens identification using the noncoding regions of chloroplast genome and nuclear ribosomal DNA. *Forensic Science International: Genetics*, **2**, 35–40.

Weiblen, G.D., Webb, C.O., Novotny, V., Basset, Y. & Miller, S.E. (2006) Phylogenetic dispersion of host use in a tropical insect herbivore community. *Ecology*, **87**, S62–S75.

Accepted 11 October 2009

Supporting Information

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

Table S1. Specimens from the superfamily Curculionoidea tested for *trnL* intron, not yielding good PCR products.

Table S2. Voucher specimens and accession numbers of weevils *cox1* and plant *trnl* intron.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.