MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2011) **11**, 786–794 doi: 10.1111/j.1755-0998.2011.03016.x

Molecular detection of trophic links in a complex insect host–parasitoid food web

JAN HRCEK,* SCOTT E. MILLER,† DONALD L. J. QUICKE‡§ and M. ALEX SMITH¶
*Faculty of Science, University of South Bohemia and Biology Center, Czech Academy of Sciences, Branisovska 31, 37005 Ceske Budejovice, Czech Republic, †National Museum of Natural History, Smithsonian Institution, PO Box 37012, MRC 105 Washington, DC 20013-7012, USA, ‡Division of Biology, Imperial College London, Silwood Park Campus, Ascot SL5 7PY, UK, §Department of Entomology, The Natural History Museum, Cromwell Rd, London SW7 5DB, UK, ¶Department of Integrative Biology and the Biodiversity Institute of Ontario, University of Guelph, Guelph, Ontario N1G 2W1, Canada

Abstract

Previously, host–parasitoid links have been unveiled almost exclusively by time-intensive rearing, while molecular methods were used only in simple agricultural host–parasitoid systems in the form of species-specific primers. Here, we present a general method for the molecular detection of these links applied to a complex caterpillar–parasitoid food web from tropical rainforest of Papua New Guinea. We DNA barcoded hosts, parasitoids and their tissue remnants and matched the sequences to our extensive library of local species. We were thus able to match 87% of host sequences and 36% of parasitoid sequences to species and infer subfamily or family in almost all cases. Our analysis affirmed 93 hitherto unknown trophic links between 37 host species from a wide range of Lepidoptera families and 46 parasitoid species from Hymenoptera and Diptera by identifying DNA sequences for both the host and the parasitoid involved in the interaction. Molecular detection proved especially useful in cases where distinguishing host species in caterpillar stage was difficult morphologically, or when the caterpillar died during rearing. We have even detected a case of extreme parasitoid specialization in a pair of *Choreutis* species that do not differ in caterpillar morphology and ecology. Using the molecular approach outlined here leads to better understanding of parasitoid host specificity, opens new possibilities for rapid surveys of food web structure and allows inference of species associations not already anticipated.

Keywords: cytochrome oxidase I, Diptera, Hymenoptera, immature stages, Lepidoptera, Papua New Guinea

Recieved 3 December 2010; revision received 3 March 2011; accepted 15 March 2011

Introduction

Molecular methods are becoming a critical tool necessary for the identification of small, morphologically uniform organisms and their biological remnants (Greenstone 2006). Until now, the molecular detection of host–parasitoid links was confined to food webs with a limited number of species, predominantly in agricultural systems (Gariepy *et al.* 2007). The most extensive such study published so far included nine parasitoid and three hyperparasitoid species of a single host species (Traugott *et al.* 2008).

In this study, our intention is to bring these methods farther into the realm of community ecology, where interactions in the tens or hundreds of species are commonplace. It is already common to use molecular methods for refining species concepts in community ecology studies, e.g. Janzen *et al.* 2009; Kaartinen *et al.* 2010; Smith *et al.* in press, 2008; so it is a logical next step to use them for mapping between-species interactions. Accurate description of host–parasitoid interactions is crucial for understanding host specificity, arguably the main parameter of host–parasitoid food webs. Knowing host specificity is important for estimates of arthropod diversity (Novotny *et al.* 2002) and explanations of its origin (Schemske *et al.* 2009). In an applied context, it is crucial for the selection of biological control agents (Miller 2007). The structure of the food web itself provides important information, as it can differ both qualitatively and quantitatively among habitats (Tylianakis *et al.* 2007), seasons (Lewis *et al.* 2002) and guilds (Novotny *et al.* 2010), and it is thus important to describe it as precisely as possible.

To date, host–parasitoid interactions have been mapped by rearing the host larvae, until either a host or a parasitoid adult emerged. Rearing provides adult specimens that are tractable for taxonomic identification and provides evidence that the parasitoid can successfully develop in the host. The main caveat of rearing is that

Correspondence: Jan Hrcek, Fax: +420 385 310 366; E-mail: janhrcek@gmail.com

host identification must be based on the immature stage. This is often difficult or impossible and commonly leads to misidentifications. Thus, many literature records on trophic interactions are unreliable (Shaw 1994). Furthermore, rearing success is commonly lower than 50%, making estimates of the rate of parasitism difficult, as parasitized and parasitoid-free larvae can differ in their rearing success. Low rearing success in combination with a naturally low rate of parasitism means that tremendous effort must be expended to get sufficiently large data sets for analysis of host–parasitoid interactions (i.e. rearing with a success of 40% and parasitism rate of 15% yields only six parasitoids of each 100 larvae sampled). DNA barcoding of adult parasitoids improves species identification in groups that have uniform morphology and greatly facilitates the identification of cryptic species (Smith *et al.* 2006, 2007, 2008). It therefore leads to better estimates of host specificity, but has not yet been used in the identification of hosts in larval stage.

Currently, there are two prevailing approaches to molecular identification of trophic links in host–parasitoid food webs and predator–prey food webs: species detection by PCR with specific primers and sequence-based detection with general or group-specific primers.

In PCR detection studies, the target gene is first sequenced from all species that are expected in the interaction and then species-specific primers are designed and tested in cross-amplification tests (King *et al.* 2008). The actual detection is carried out by a series of PCRs, each with a different species-specific primer, or in one multiplex PCR (Traugott *et al.* 2006). A successful PCR is regarded as positive detection of a species for which the primers have been designed. This approach has been commonly used in host–parasitoid systems (Tilmon *et al.* 2000; Prinsloo *et al.* 2002; Ratcliffe *et al.* 2002; Ashfaq *et al.* 2004; Weathersbee *et al.* 2004; Agusti *et al.* 2005; Gariepy *et al.* 2005; Jones *et al.* 2005). It can be very sensitive as it allows detection of the parasitoid in the host larva, including immediately after oviposition (Zhu *et al.* 2004; Traugott & Symondson 2008). The detection process is straightforward and allows processing of thousands of samples, once the specific primers have been developed (Gariepy *et al.* 2008). However, this approach does not allow recognition of unknown species, a crucial step for studying complex communities, where new species are commonly encountered in the course of the study (Janzen *et al.* 2009).

An alternative approach is sequence-based detection. Sequences of the target gene can be obtained by various ways that separate the DNA from the different organisms involved in the interaction. Clare *et al.* (2009) took advantage of macroscopic prey fragments present in bat guano and sequenced DNA extracts made from these individual remnant pieces. Another possibility is to use group-specific primers (Jarman *et al.* 2004; Deagle *et al.* 2009). Identification of the sequences is then made by comparing them to a reference library using BLAST algorithm (Ross *et al.* 2008), tree construction (Ross *et al.* 2008) or sequence similarity measures (Ratnasingham & Hebert 2007). A similar approach has also been used one trophic level down the food web for identifying host plants of insect herbivores (Jurado-Rivera *et al.* 2009).

Rougerie *et al.* (2010) have recently developed a group-specific primer approach for sequencing both the host and the parasitoid DNA barcode region from abdomens of adult parasitoids (molecular analysis of parasitoid linkages—MAPL). Their method has the same goal as our study (mapping host–parasitoid interactions), but approaches the food web from different direction, as it records hosts of a parasitoid, while the methods reported here record parasitoids of a host.

In this study, we use sequence-based identification of host and parasitoid traces to establish trophic links in a species-rich and phylogenetically diverse caterpillar–parasitoid food web from a tropical rainforest. We first obtained COI barcode sequences using general insect primers from three sets of samples: (i) reared adult parasitoids and larval host remnants collected after their emergence, (ii) freshly collected caterpillars and parasitoid larvae dissected from them and (iii) caterpillars that died during rearing and parasitoid larvae dissected from them. We then identified the sequences by matching them to our reference library of DNA barcodes built from adult host and parasitoid sequences.

In the case of dissections, the tissue of the host and the parasitoid larva cannot be separated mechanically with 100% certainty. We therefore test whether this poses a problem for the use of general insect primers, which would manifest itself in host sequences originating from parasitoid tissue samples and vice versa. We predict that such cases would be most common among small caterpillars of mining species, which are more difficult to dissect than bigger caterpillars.

Materials and methods

The field sampling took place in 2004–2007 as part of a long-term rearing programme (Novotny *et al.* 2004, 2007, 2010). The studied species included representatives of all three guilds of leaf-feeding Lepidoptera: free-living, semi-concealed (leaf-rolling and web-tying) and mining species. The caterpillars were collected by local assistants in a mosaic of primary and secondary forest surrounding Wanang and Ohu villages in Madang province and Yapsiei and Wamangu villages in East Sepik province, Papua New Guinea (PNG).

1 Parasitoids reared from caterpillars. A part of the parasitoid tissue and part of host tissue remnants were

taken for extraction, amplification and sequencing from each parasitoid rearing. Host remnants were predominantly pieces of caterpillar skin, often together with the caterpillar head, but in a few cases parts of pupal cuticle or the tissue from inside of the pupa. Samples were stored dry or in ethanol. In the latter case, the parasitoid and the host remnants were stored in the same vial.

2 Parasitoids dissected from live caterpillars collected into ethanol. Caterpillars were dissected under binocular microscope, and a tissue sample was taken for sequencing from each parasitized caterpillar and any parasitoid larvae or eggs found in, or on, the caterpillar (Santos *et al.* 2010).

3 Parasitoids dissected from caterpillars that died during rearing. Dead caterpillars were preserved in ethanol, typically <24 h after they died. They were dissected and a tissue sample was taken from each parasitized caterpillar and its parasitoids.

DNA analysis and sequence analysis

COI sequences were generated at the Biodiversity Institute of Ontario, University of Guelph using standard barcoding protocols (Ivanova *et al.* 2006). Briefly, a small amount of tissue (single leg in adult specimens) was used for the extraction of total genomic DNA using an AcroPrep™ 96 1-ml filter plate (PALL) with 3.0-μm glass fibre. DNA was eluted in 40 μl of dH_2O. Full-length COI barcodes (658 bps) were amplified using the standard insect primers LepF1/LepR1 (LepF1: 5′-ATTCAACCAATCATAAAGATATTGG-3′; LepR1 (5′-TAAACTTCTGGATGTCCAAAAAATCA-3′; (Hebert *et al.* 2004)) In cases where a full-length product was not successfully generated, internal primer pairs (LepF1/C_ANTMR1D) and (MLepF1/LepR1) (Smith *et al.* 2008) were employed to generate shorter sequences. These could be overlapped to create composite sequence (contig) or could be analysed as shorter, non-barcode-standard length sequences. Each PCR had a total volume of 12.5 μl and contained 5% trehalose (D-(+)-Trehalose dehydrate), 1.25 μl of 10× reaction buffer, 2.5 mM of $MgCl_2$, 1.25 pmol each of forward and reverse primer, 50 μM of dNTP (Promega), 0.3 U of Platinum Taq DNA polymerase (Invitrogen) and 2 μl of genomic DNA. PCRs were carried out using a thermocycling profile of one cycle of 1 min at 94 °C, five cycles of 40 s at 94 °C, 40 s at 45 °C and 1 min at 72 °C, followed by 35 cycles of 40 s at 94 °C, 40 s at 51 °C and 1 min at 72 °C, with a final step of 5 min at 72 °C. PCR products were visualized on a 2% agarose E-gel® 96-well system (Invitrogen) and were bi-directionally sequenced using BigDye v3.1 and analysed on an ABI 3730xl DNA Analyzer (Applied Biosystems). Contigs were assembled using Sequencher v 4.0.5 (Gene Codes) and were subsequently aligned by eye in Bioedit (Hall 1999) ensuring that there were no gaps or stop codons in the alignment. Sequence divergences were calculated using the K2P distance model (Kimura 1980), and a NJ tree of distances (Saitou & Nei 1987) was created to provide a graphical representation of the among-species divergences using MEGA4 (Tamura *et al.* 2007) and BOLD (Ratnasingham & Hebert 2007).

The Braconidae: Agathidinae sequences analysed in this study display 1-bp deletions that are frequently found in this subfamily. If unrecognized, these 1-bp deletions place the alignment out of frame and result in stop codons—an accepted signature of a pseudogene or NUMT. However, the 1-bp deletions are present in specimens from widely divergent localities and specimens from provisional species that display host-specific ecologies and are also retrieved using multiple primer combinations (Smith unpublished). Perhaps, while gene product may be internally deleted, it is largely functional, corrected possibly by RNA editing (Russell & Beckenbach 2008) via an analogous system for -1 programmed translational frameshifting. Further work, beyond the scope of this investigation, is clearly required on this subfamily to determine whether this actually is a pseudogene or not. In this study, we treat the Agathidinae sequences that included the deletions as CO1 marker sequences.

All sequences from this project, as well as five sequences closest to each of them from our host and parasitoid reference libraries, have been deposited in both Genbank and BOLD databases (Table S1, Supporting information).

The sequences were identified to putative species using the BOLD-IDS tool, searching through all previously identified or characterized COI barcodes within the BOLD database and placing them in a distance tree using the neighbour-joining (NJ) algorithm applied to Kimura 2-parameter corrected distances, as implemented in BOLD (Ratnasingham & Hebert 2007). This tool gives three categories of matches based on sequence divergence measure—strict (up to 1%), loose (between 1% and 5%) and distant (greater than 5%). We took species-level identifications only from the strict category. We then identified the rest of the sequences by searching through the whole BOLD database with the BOLD-IDS tool and constructing a NJ tree from 99 most similar sequences plus the query sequence. We took as the identification the narrowest taxonomic category shared by five sequences that were closest to our query sequence on the tree, skipping sequences from our query data set.

As a way of confirming the identifications, we have also built a combined NJ tree from all query and reference sequences.

The barcode library of PNG Lepidoptera has been built as part of our rearing effort since 1994 and currently contains over 9000 sequences from over 1000 identified species. General field and laboratory methods are described by Miller *et al*. 2003 and Craft *et al*. 2010.

Identifications have been based on extensive dissection of genitalia and reference to type specimens. Three new generic combinations in Lepidoptera are validated by publication here so they can be used for identifications: Lymantriidae, *Arna galactopis* (Turner) and *Somena alba* (Bethune-Baker); Thyrididae, *Collinsa acutalis* (Walker). See taxonomic notes on Lepidoptera hosts in Appendix S1(Supporting information) for details on these and other Lepidoptera species.

The barcode library of PNG parasitoids is less extensive and contains ~400 sequences from ~130 identified species and further ~1200 sequences from adults identified only to family or subfamily. The family and subfamily identifications were made by JH, while the species identifications were made by specialists on each group, whom we list, with thanks, in the Acknowledgements. Only two parasitoid species referred to in this study have been formally described (Diptera: Tachinidae: *Argyrophylax solomonica* and *Paradrino laevicula*), the other species are putative assessments by the respective taxonomists.

Results

We isolated the DNA and attempted sequencing of the mitochondrial COI barcode region for 392 samples, approximately equally distributed between (i) reared parasitoids with (ii) host remnants from their rearing and (iii) caterpillars with (iv) parasitoid larvae dissected from them.

We obtained DNA barcodes from 313 samples (80%), with the highest success rate for caterpillar tissue (100%) and the lowest success rate for host remnants associated with reared parasitoids (51%).

Forty-five sequences (14.5%) were not in the higher taxa we expected, including 33 sequences (11%), which were not of the intended half of the host–parasitoid relationship, but they matched the sequence of the sample from the other side of the interaction pair whenever we had the opportunity to compare them (25 cases). Ten sequences (3%), sampled as ectoparasitoid larvae or pupae, belonged in fact to Cecidomyiidae and Phoridae flies. In two cases (0.6%), we sequenced a Tachinidae primary parasitoid of a Chalcidoidea hyperparasitoid instead of the Lepidoptera primary host from the host remnants.

A majority of all nontarget sequences came from <1-cm-long miner caterpillars and their parasitoids (23 cases; 51%) and host remnants from rearings of braconid subfamilies Macrocentrinae (seven cases; 16%) and Agathidinae (four cases; 9%). Host remnant samples stored in a vial together with the parasitoid often produced cross-interaction sequences (37% from 41 samples).

A sample of 11 dead caterpillars that were dissected after unsuccessful rearing had comparable sequencing success (100% vs. 100%) and cross-amplification rate (9% vs. 13%) to the caterpillars that were preserved fresh, directly after collection. The same was true for parasitoid larvae dissected from these caterpillars (100% vs. 92% and 28% vs. 18%, respectively).

We used the BOLD-IDS tool for sequence identification (Table S2, Supporting information) and were able to identify 87% of the host and 36% of the parasitoid sequences to species. All of the tight matches were based on other PNG sequences from the ongoing rearing campaign stored in BOLD. Almost all (98%) parasitoid sequences could be identified at least to family, thanks mostly to partially identified sequences in our reference library.

The combined tree of all query and reference sequences produced the same identifications as the BOLD-IDS tool, identifying only one more host species (*Melanocercops* sp. mine005b).

Identified tissue samples spanned several families of Macrolepidoptera (Arctiidae, Geometridae, Limacodidae, Lycaenidae, Lymantriidae, Noctuidae) and Microlepidoptera (Choreutidae, Crambidae, Gelechiidae, Gracillariidae, Pyralidae, Thyrididae, Tortricidae), as well as main taxa of their parasitoids (Hymenoptera: Braconidae, Ichneumonidae, Chalcidoidea, Bethylidae and Diptera: Tachinidae). We recorded 93 associations between 37 host species and 46 parasitoid species (Fig. 1) by combining sequence matches from each host–parasitoid pair.

We examined in detail parasitoids of two host species (*Choreutis* cf. *anthorma* and *Choreutis basalis*) whose adults are distinctive, but the caterpillars feed on the same host plants and are not distinguishable morphologically (Fig. 2). We have detected a specific parasitoid of *C. basalis* within this pair of hosts, despite the caterpillars being much rarer in our sample than *C.* cf. *anthorma*. We also detected two cases of multiparasitism while dissecting the caterpillars. In one case, both parasitoid larvae were from the same species of Tachinidae (Fig. 2, letter A), but had 0.92% different sequences. In the other case, each parasitoid larva belonged to a different tachinid species (Fig. 2, letter B), while one of them was unusually located inside the caterpillar's head capsule.

Discussion

Sequencing success and nontarget sequences

All materials in our study were collected at remote field sites in PNG and were therefore not always stored in the

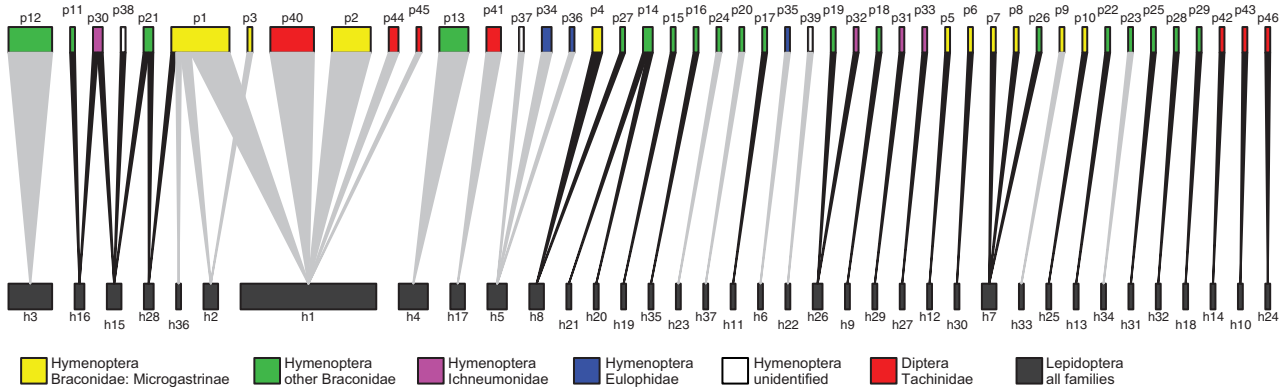

Fig. 1 Trophic links detected in the food web by barcoding reared parasitoids and their host remnants (black links) and parasitoid larvae dissected from caterpillars (grey links). Only the trophic links where we successfully sequenced both the host and the parasitoid are shown. The figure was made with bipartite package (Dormann *et al.* 2009) in R (R Development Core Team, 2009). Lepidoptera hosts: Choreutidae: *Choreutis* cf. *anthorma* (h1), *Choreutis basalis* (h2). Gracillariidae: *Melanocercops* sp. mine005b (h3), *Spulerina* sp. 1 (h4), *Stomphastis* n. sp. (mine015) (h5), *Neolithocolletis pentadesma* (h6). Crambidae: *Glyphodes margaritaria* (h7), *Parotis hilaralis* (h8), *Glyphodes* nr. *stolalis* (h9), *Meekiaria* sp. CRAM151 (h10), *Meekiaria* sp. BOLD:AAA3374 (h11), *Parotis marginata* complex (h12), *Piletocera* sp. CRAM094 (h13), *Talanga excelsalis* (h14). Tortricidae: *Adoxophyes thoracica* (h15), *Adoxophyes* sp. 3 (h16). Lycaenidae: *Philiris helena* (h17), *Hypochrysops chrysargyrus* (h18). Noctuidae: *Earias flavida* (h19), *Etanna brunnea* (h20), *Giaura* sp. 1 (h21), *Tamba* sp. (h22). Lymantriidae: *Arctornis* nr. *intacta* (h23), *Arna* sp. nr. *galactopis* (h24), *Olene* sp. nr. *mendosa* (h25). Thyrididae: *Addaea pusilla* (h26), *Collinsa acutalis* (h27). Gelechiidae: *Dichomeris* sp. XXXX068 (h28). Geometridae: *Aeolochroma perfulvata* (h29), *Hypomecis infaustaria* (h30). Arctiidae (h31). Limacodidae (h32). Pyralidae: *Unadophanes trissomita* (h33). Unknown family (h34–h37). Hymenoptera parasitoids: Braconidae: Microgastrinae: *Apanteles* Whitfield17 (p1), *Dolichogenidea* Whitfield01 (p2), *Dolichogenidea* Whitfield02 (p3), unknown genus (p4–p10). Braconidae: Agathidinae: *Camptothlipsis* Sharkey02 (p11), unknown genus (p12–p16). Braconidae: Cardiochilinae: *Schoenlandella* Whitfield01 (p17). Braconidae: Cheloninae: *Chelonus* Yu01 (p18), unknown genus (p19). Braconidae: Meteorinae (p20). Braconidae: Orgilinae (p21–p23). Braconidae: Rogadinae: *Aleoides* Quicke01 (p24), *Canalirogas* Quicke01 (p25), *Colastomion* Quicke01 (p26), *Colastomion* Quicke03 (p27), *Spinaria* Quicke01 (p28), unknown genus (p29). Ichneumonidae (p30–p33). Eulophidae (p34–p36). Unknown family (p37–p39). Diptera parasitoids: Tachinidae: *Actia* Shima01 n. sp. nr *cinerea* (p40), *Paradrino laevicula* (p41), *Argyrophylax solomonica* (p42), *Argyrophylax* Shima01 n. sp. (p43), unknown genus (p44–p46).

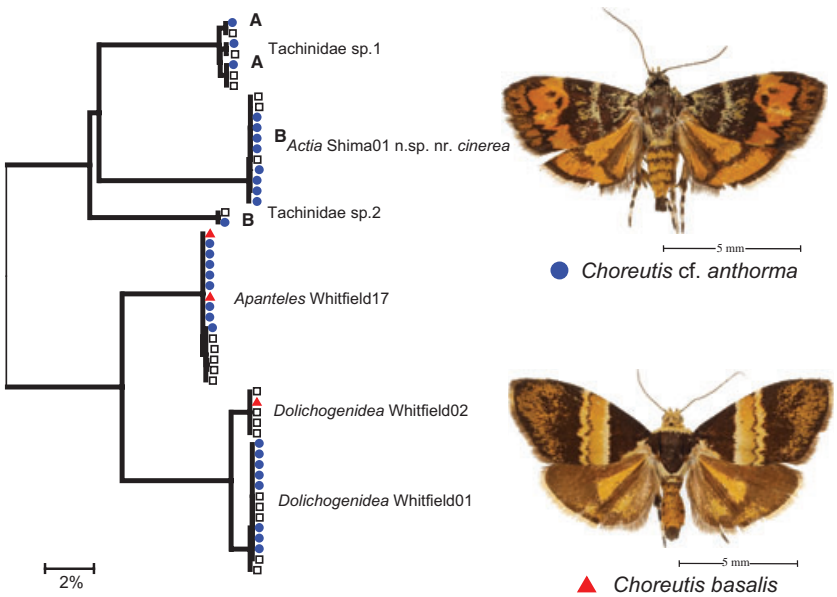

Fig. 2 NJ tree of parasitoid larvae dissected from caterpillars of *Choreutis* cf. *anthorma* (blue dots) and *Choreutis basalis* (red triangles). Reference sequences (reared adult parasitoids) are shown as white squares. Pairs of parasitoid larvae found in the same host individual are marked with the same letter (A, B). The scale bar shows sequence divergence.

best conditions for DNA preservation. Despite that, sequencing success was good for all but the host remnants.

All nontarget sequences most likely arose during the tissue sampling and were caused by incorrect identification of the tissue fragments or cross-amplification of host DNA from the parasitoid sample and vice versa. We anticipated cross-amplification as we used general primers on tissue samples, which cannot be altogether separated mechanically. Cross-amplification proved to be an important issue especially in barcoding parasitoids dissected from small mining caterpillars. In miners, we have also mistakenly barcoded several cecydomiid pupae, confusing them for ectoparasitoids. Our method is therefore of limited use for parasitoids of miners and other tiny insects, or small developmental stages like eggs, where an approach based on group-specific primers might be more appropriate. Alternatively, a mature barcode library could be used to construct a microarray that would then be used to screen very small hosts (Hajibabaei *et al.* 2007). However, cross-amplification was rare in the dissections of caterpillars bigger than 1 cm, permitting efficient use of the method in leaf-rolling and free-living caterpillars.

Cross-amplification was also common for host remnants stored in ethanol together with the parasitoid, particularly from rearings of parasitoids from braconid subfamilies Agathidinae and Macrocentrinae. This may have been caused by mistakenly sampling parasitoid larval skin remnants instead of caterpillar skin, or the samples could have been inundated with DNA from the parasitoid (Shokralla *et al.* 2010). The latter problem could be solved by storing the host remnants in separate vial or by using group-specific primers for targeting the host DNA (Rougerie *et al.* 2010).

Very good barcoding success and low rate of cross-amplification were recorded from another type of degraded tissue: caterpillars that died during rearing and parasitoids dissected from them. This makes it possible to map host–parasitoid interactions in the large proportion of caterpillars that die in rearing surveys.

Sequence identification

Both ways of identifying the sequences (BOLD-IDS and tree of all query and reference sequences) produced the same identifications. The only species identified in addition by the latter method could not have been picked up by BOLD-IDS, because the reference sequences for it are shorter than 300 bp (BOLD-IDS works with reference sequences at least 500-bp long). However, we feel confident about the identification, as we know the host plant from which the sample was collected and this miner is host specific to it.

We had to come up with arbitrary rules for identifying sequences that did not have strict matches in the reference library. We kept these rules conservative, as we experienced family-level jumps in less-conservative measures as the reference library grew. The rule of looking at five nearest reference sequences worked well for our data set, but we recommend also verification based on species biology and tree topology.

The technical details of making the identification are important to get right, but the single most important issue in producing reliable identifications is the quality and size of the reference library. The focal community is very species rich (Novotny *et al.* 2010), but the analysis benefited strongly from the long-term rearing effort in the study area—about 9000 sequences from 1000 species in the reference library were enough to identify almost all host sequences.

An important advantage of sequence-based identification is the possibility to build the library progressively—anything that was not identified now can be identified later when more sequences from identified specimens are available. It also provides provisional identification for species not present in the library. The BOLD database (Ratnasingham & Hebert 2007) works as an efficient workbench and depository for the sequence library, as it accommodates both detailed data about the samples and the identification engine. Additionally, the identifications and sequences from the PNG project can be easily integrated and compared with other regional and global queries of both the hosts and the parasitoids (Holloway *et al.* 2009; Janzen *et al.* 2009).

Mapping host–parasitoid interactions

Both methods for mapping host–parasitoid links presented in this study work for a broad range of Lepidoptera hosts and their parasitoids (Fig. 1), but each method has a distinct set of constraints and advantages.

Barcoding host remnants together with the reared parasitoid. The disadvantage of this method is that it requires successful rearing, the advantage that it permits precise identification of both the host and the parasitoid. Such accuracy is crucial for avoiding erroneous records of association, which are common in the literature (Shaw 1994). The method is useful in any situation where rearing unambiguously associates the parasitoid with its host, while the host is difficult to identify in the developmental stage (egg, larva, pupa) in which it was collected. The rearing is useful also because it proves that the parasitoid is able to complete development in the given host.

Barcoding host larvae and parasitoid larvae dissected from them. This method requires host dissection, which can

be difficult in small individuals. It is also likely to be less sensitive than specific PCR methods in the detection of parasitoids in early developmental stages as small parasitoids can be missed during the dissection. The method can be particularly useful for concealed hosts (e.g. wood-boring, fruit feeding and gall-making insects) and their parasitoids, because it is difficult to establish direct links through rearing for these insects. The method also opens new possibilities for quick food web surveys (i.e. Santos *et al.* 2010), where time or facilities needed for rearing are not available, while keeping species-level resolution, until now achievable only by rearing. Furthermore, it can be used to study the hosts that died during rearing, thus significantly increasing sample size for rearing campaigns and improving estimates of parasitism rates and food web structure.

While dissection is imprecise for very small hosts such as miners, it can uncover unexpected mortality agents (parasitic nematodes or pathogens), as well as cases of multiparasitism and hyperparasitism. We have recorded two cases of multiparasitism, but it is not clear what would be the fate of the two parasitoid larvae from one host in natural conditions—one, none or both of them could successfully finish development. The interactions recovered using this method have to be considered as provisional, because it does not prove that the parasitoid can successfully develop on the host.

The methods presented here nicely complement the method of Rougerie *et al.* (2010) for mapping host–parasitoid interactions. Their approach of sequencing host DNA from the guts of adult parasitoid is unique, because it allows recording hosts of a parasitoid without rearing. Therefore, it has great potential to uncover hosts of the many adult parasitoids that are often encountered in nature by sweeping or malaise trapping, but their hosts are unknown.

The sampling unit for the methods presented here is a host larva, which is either dissected or reared. Our methods are therefore likely to be more useful for host–parasitoid food web studies that rely on sampling of the host larvae for reliable quantification of the interactions.

Barcoding studies have recently found numerous cryptic species in parasitoid communities and documented their higher than anticipated host specialization (Smith *et al.* 2006, 2007, 2008). Our results support this trend towards a previously unappreciated high degree of host specialization. We have detected two *Dolichogenidea* parasitoids, each specific for one of the two *Choreutis* species (Fig. 2). Such precise description of the interaction would not be possible by conventional techniques, as the host caterpillars do not differ in morphology or ecology. Rearing studies would thus produce two adult moth species and two parasitoid species, with little information on the host specificity of the parasitoids.

We wish to emphasize the usefulness of these methods for the study of complex food webs with numerous unknown species. These food webs are too complex for methods based on species-specific primers, as these are not able to detect unknown or unexpected species (Gariepy *et al.* 2007). However, nothing restrains the use of the methods described here in simple food webs typical for biological control applications. Moreover, these methods are easy to use, as they do not require development and testing of a new set of primers for each studied system. As sequencing costs continue to decline, the approach will become increasingly affordable, permitting the methods described here to be employed in large-scale sampling campaigns to unravel host–parasitoid interactions in unprecedented detail.

Acknowledgements

We thank New Guinea Binatang Research Center parataxonomists, village assistants, Vojtech Novotny, Lauren Helgen, Karolyn Darrow, Kristyna Hrckova, Toko Pagi, David Pollock and Legi Sam for technical assistance and general support. Lepidoptera taxonomy was assisted by Jeremy Holloway, Jadranka Rota, Michael Shaffer, Tosio Kumata, Issei Ohshima and others acknowledged in (Craft *et al.* 2010). Parasitoids were identified by: Kees van Achterberg (Macrocentrinae), Celso Oliveira Azevedo (Bethylidae), Yves Braet (Orgilinae), Michael W. Gates (Chalcidoidea), Ian D. Gauld (Ichneumonidae), Donald L.J. Quicke (Rogadinae), Michael J. Sharkey (Agathidinae), Hiroshi Shima (Tachinidae), David Wahl (Ichneumonidae), James B. Whitfield (Microgastrinae, Cardiochilinae) and Dicky Sick Ki Yu (Cheloninae). This study is based on work supported by the US National Science Foundation (DEB 0841885, 0816749, 0515678), Czech Science Foundation (206/08/H044, 206/09/0115 and P505/10/0673), Czech Ministry of Education (LC06073, ME9082 and MSM6007665801) and Grant Agency of the University of South Bohemia (GAJU 136/2010/P). The barcode library of local fauna was built from NSF-funded research since 1995 in collaboration with Vojtech Novotny, George Weiblen and Yves Basset. We thank Paul Hebert and all of our colleagues at the Biodiversity Institute of Ontario for their assistance and support. This Project was funded by the Government of Canada through Genome Canada and the Ontario Genomics Institute (2008-0GI-ICI-03).

References

Agusti N, Bourguet D, Spataro T *et al.* (2005) Detection, identification and geographical distribution of European corn borer larval parasitoids using molecular markers. *Molecular Ecology*, **14**, 3267–3274.

Ashfaq M, Braun L, Hegedus D, Erlandson M (2004) Estimating parasitism levels in *Lygus* spp. (Hemiptera: Miridae) field populations using standard and molecular techniques. *Biocontrol Science and Technology*, **14**, 731–735.

Clare EL, Fraser EE, Braid HE, Fenton MB, Hebert PDN (2009) Species on the menu of a generalist predator, the eastern red bat (*Lasiurus borealis*): using a molecular approach to detect arthropod prey. *Molecular Ecology*, **18**, 2532–2542.

Craft KJ, Pauls SU, Darrow K *et al.* (2010) Population genetics of ecological communities with DNA barcodes: an example from New Guinea

Lepidoptera. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 5041–5046.

Deagle BE, Kirkwood R, Jarman SN (2009) Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. *Molecular Ecology*, **18**, 2022–2038.

Dormann CF, Fründ J, Blüthgen N, Gruber B (2009) Indices, graphs and null models: analyzing bipartite ecological networks. *The Open Ecology Journal*, **2**, 7–24.

Gariepy TD, Kuhlmann U, Haye T, Gillott C, Erlandson M (2005) A single-step multiplex PCR assay for the detection of European *Peristenus* spp., parasitoids of *Lygus* spp. *Biocontrol Science and Technology*, **15**, 481–495.

Gariepy TD, Kuhlmann U, Gillott C, Erlandson M (2007) Parasitoids, predators and PCR: the use of diagnostic molecular markers in biological control of Arthropods. *Journal of Applied Entomology*, **131**, 225–240.

Gariepy T, Kuhlmann U, Gillott C, Erlandson M (2008) A large-scale comparison of conventional and molecular methods for the evaluation of host-parasitoid associations in non-target risk-assessment studies. *Journal of Applied Ecology*, **45**, 708–715.

Greenstone MH (2006) Molecular methods for assessing insect parasitism. *Bulletin of Entomological Research*, **96**, 1–13.

Hajibabaei M, Singer GAC, Clare EL, Hebert PDN (2007) Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring. *BMC Biology*, **5**, 24.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Hebert PDN, Penton EH, Burns JM, Janzen DH, 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.

Holloway J, Miller S, Pollock D, Helgen L, Darrow K (2009) GONGED (Geometridae of New Guinea Electronic Database): a progress report on development of an online facility of images. *Spixiana*, **32**, 122–123.

Ivanova NV, Dewaard JR, Hebert PDN (2006) An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes*, **6**, 998–1002.

Janzen DH, Hallwachs W, Blandin P *et al.* (2009) Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. *Molecular Ecology Resources*, **9**, 1–26.

Jarman SN, Deagle BE, Gales NJ (2004) Group-specific polymerase chain reaction for DNA-based analysis of species diversity and identity in dietary samples. *Molecular Ecology*, **13**, 1313–1322.

Jones DB, Giles KL, Chen Y, Shufran KA (2005) Estimation of hymenopteran parasitism in cereal aphids by using molecular markers. *Journal of Economic Entomology*, **98**, 217–221.

Jurado-Rivera JA, Vogler AP, Reid CAM, Petitpierre E, Gomez-Zurita J (2009) DNA barcoding insect-host plant associations. *Proceedings of the Royal Society B-Biological Sciences*, **276**, 639–648.

Kaartinen R, Stone GN, Hearn J, Lohse K, Roslin T (2010) Revealing secret liaisons: DNA barcoding changes our understanding of food webs. *Ecological Entomology*, **35**, 623–638.

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111–120.

King RA, Read DS, Traugott M, Symondson WOC (2008) Molecular analysis of predation: a review of best practice for DNA-based approaches. *Molecular Ecology*, **17**, 947–963.

Lewis OT, Memmott J, Lasalle J *et al.* (2002) Structure of a diverse tropical forest insect-parasitoid community. *Journal of Animal Ecology*, **71**, 855–873.

Miller SE (2007) DNA barcoding and the renaissance of taxonomy. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 4775–4776.

Miller SE, Novotny V, Basset Y (2003) Studies on New Guinea moths. I. Introduction (Lepidoptera). *Proceedings of the Entomological Society of Washington*, **105**, 1034–1042.

Novotny V, Basset Y, Miller SE *et al.* (2002) Low host specificity of herbivorous insects in a tropical forest. *Nature*, **416**, 841–844.

Novotny V, Miller SE, Leps J *et al.* (2004) No tree an island: the plant-caterpillar food web of a secondary rain forest in New Guinea. *Ecology Letters*, **7**, 1090–1100.

Novotny V, Miller SE, Hulcr J *et al.* (2007) Low beta diversity of herbivorous insects in tropical forests. *Nature*, **448**, 692–695.

Novotny V, Miller SE, Baje L *et al.* (2010) Guild-specific patterns of species richness and host specialization in plant-herbivore food webs from a tropical forest. *Journal of Animal Ecology*, **79**, 1193–1203.

Prinsloo G, Chen Y, Giles KL, Greenstone MH (2002) Release and recovery in South Africa of the exotic aphid parasitoid *Aphelinus hordei* verified by the polymerase chain reaction. *BioControl*, **47**, 127–136.

R Development Core Team (2009) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Ratcliffe ST, Robertson HM, Jones CJ, Bollero GA, Weinzierl RA (2002) Assessment of parasitism of house fly and stable fly (Diptera: Muscidae) pupae by pteromalid (Hymenoptera: Pteromalidae) parasitoids using a polymerase chain reaction assay. *Journal of Medical Entomology*, **39**, 52–60.

Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System (http://www.barcodinglife.org). *Molecular Ecology Notes*, **7**, 355–364.

Ross HA, Murugan S, Li WLS (2008) Testing the reliability of genetic methods of species identification via simulation. *Systematic Biology*, **57**, 216–230.

Rougerie R, Smith MA, Fernandez-Triana J *et al.* (2010) Molecular analysis of parasitoid linkages (MAPL): gut contents of adult parasitoid wasps reveal larval host. *Molecular Ecology*, **20**, 179–186.

Russell RD, Beckenbach AT (2008) Recoding of translation in turtle mitochondrial genomes: programmed frameshift mutations and evidence of a modified genetic code. *Journal of Molecular Evolution*, **67**, 682–695.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406–425.

Santos AMC, Besnard G, Quicke DLJ (2011) Applying DNA barcoding for the study of geographical variation in host-parasitoid interactions. *Molecular Ecology Resources*, **11**, 46–59.

Schemske DW, Mittelbach GG, Cornell HV, Sobel JM, Roy K (2009) Is there a latitudinal gradient in the importance of biotic interactions? *Annual Review of Ecology Evolution and Systematics*, **40**, 245–269.

Shaw MR (1994) Parasitoid host ranges. In: *Parasitoid Community Ecology* (eds Hawkins BA, Sheehan W), pp. 111–144. Oxford University Press, Oxford.

Shokralla S, Singer GAC, Hajibabaei M (2010) Direct PCR amplification and sequencing of specimens' DNA from preservative ethanol. *BioTechniques*, **48**, 233–234.

Smith MA, Woodley NE, Janzen DH, Hallwachs W, Hebert PDN (2006) DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 3657–3662.

Smith MA, Wood DM, Janzen DH, Hallwachs W, Hebert PDN (2007) DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (Diptera, Tachinidae) are not all generalists. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 4967–4972.

Smith MA, Rodriguez JJ, Whitfield JB *et al.* (2008) Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology and collections. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 12359–12364.

Smith M, Eveleigh E, McCann K *et al.* (in press) Barcoding a quantified food web: crypsis, concepts, ecology and hypotheses. *PLoS ONE*.

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

Tilmon KJ, Danforth BN, Day WH, Hoffmann MP (2000) Determining parasitoid species composition in a host population: a molecular approach. *Annals of the Entomological Society of America*, **93**, 640–647.

Traugott M, Symondson WOC (2008) Molecular analysis of predation on parasitized hosts. *Bulletin of Entomological Research*, **98**, 223–231.

Traugott M, Zangerl P, Juen A, Schallhart N, Pfiffner L (2006) Detecting key parasitoids of lepidopteran pests by multiplex PCR. *Biological Control*, **39**, 39–46.

Traugott M, Bell JR, Broad GR *et al.* (2008) Endoparasitism in cereal aphids: molecular analysis of a whole parasitoid community. *Molecular Ecology*, **17**, 3928–3938.

Tylianakis JM, Tscharntke T, Lewis OT (2007) Habitat modification alters the structure of tropical host-parasitoid food webs. *Nature*, **445**, 202–205.

Weathersbee AA, Shufran KA, Panchal TD, Dang PM, Evans GA (2004) Detection and differentiation of parasitoids (Hymenoptera: Aphidiidae and Aphelinidae) of the brown citrus aphid (Homoptera: Aphididae): species-specific polymerase chain reaction amplification of 18S rDNA. *Annals of the Entomological Society of America*, **97**, 286–292.

Zhu YC, Riddick EN, Williams L *et al.* (2004) Potential of detection and identification of nymphal parasitoids (Hymenoptera: Braconidae) of *Lygus* bugs (Heteroptera: Miridae) by using polymerase chain reaction and ITS2 sequence analysis techniques. *Annals of the Entomological Society of America*, **97**, 743–752.

Data accessibility

DNA sequences: Genbank and BOLD accession numbers are listed in Table S1.
Data sets: DRYAD entry doi:10.5061/dryad.8720.

Supporting Information

Additional supporting information may be found in the online version of this article.

Table S1 Details of all records in the dataset and relevant reference records.

Table S2 Tight matches for dataset query sequences retreived from BOLD-IDS interface, 22 November 2010.

Appendix S1 Taxonomic notes on Lepidoptera hosts.

Note added in proof: Subsequent work on the PNG and Australian barcode libraries has allowed one more host sequence (h35 in Fig. 1; BOLD: ASPN909-09; Genbank: JF271414) to be identified as *Chrysothyridia invertalis* (Crambidae: Spilomelinae).

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