MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2013) **13**, 168–176 doi: 10.1111/1755-0998.12038

DNA barcoding and the taxonomy of Microgastrinae wasps (Hymenoptera, Braconidae): impacts after 8 years and nearly 20 000 sequences

M. ALEX SMITH,*[1] J. L. FERNÁNDEZ-TRIANA,*†[1] E. EVELEIGH,‡ J. GÓMEZ,§ C. GUCLU,¶ W. HALLWACHS,‖ P. D. N. HEBERT,* J. HRCEK,** J. T. HUBER,†† D. JANZEN,‖ P. G. MASON,† S. MILLER,‡‡ D. L. J. QUICKE,§§¶¶ J. J. RODRIGUEZ,‖‖ R. ROUGERIE,*** M. R. SHAW,††† G. VÁRKONYI,‡‡‡ D. F. WARD,§§§ J. B. WHITFIELD¶¶¶ and A. ZALDÍVAR-RIVERÓN‖‖‖‖

**Department of Integrative Biology and the Biodiversity Institute of Ontario, University of Guelph, Guelph, Ontario, Canada N1G 2W1, †Research Centre, Agriculture and Agri-Food Canada, 960 Carling Avenue, Ottawa, Ontario, Canada K1A 0C6, ‡Atlantic Forestry Centre, Natural Resources Canada, Canadian Forest Service, Fredericton, New Brunswick, Canada E3B 5P7, §El Colegio de la Frontera Sur, Barrio María Auxiliadora, San Cristóbal de Las Casas, Chiapas CP 29290, Mexico, ¶Department of Plant Protection, Faculty of Agriculture, Atatürk University, Erzurum 25240, Turkey, ‖Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA, **Faculty of Science, Czech Academy of Sciences, University of South Bohemia and Biology Center, Branisovska 31, Ceske Budejovice 37005, Czech Republic, ††Natural Resources Canada c/o Canadian National Collection of Insects, Research Centre, Agriculture and Agri-Food Canada, 960 Carling Avenue, Ottawa, Ontario, Canada K1A 0C6, ‡‡National Museum of Natural History, Smithsonian Institution, PO Box 37012, MRC 105, Washington, DC 20013-7012, USA, §§Department of Life Sciences, Imperial College London, Silwood Park Campus, Ascot, Berkshire SL5 7PY, UK, ¶¶Department of Entomology, The Natural History Museum, Cromwell Rd, London SW7 5DB, UK, ‖‖National Center for Ecological Analysis and Synthesis, University of California Santa Barbara, 735 State St. Suite 300, Santa Barbara, CA 93101, USA, ***Laboratoire ECODIV, Université de Rouen, Batiment IRESE A, Place Emile Blondel, F-76821 Mont Saint Aignan Cedex, France, †††Honorary Research Associate, National Museums of Scotland, Chambers Street, Edinburgh EH1 1JF, UK, ‡‡‡Finnish Environment Institute, Friendship Park Research Centre, Lentiirantie 342B, FI-88900 Kuhmo, Finland, §§§New Zealand Arthropod Collection, Landcare Research, Private Bag 92170, Auckland, New Zealand, ¶¶¶Department of Entomology, University of Illinois, Urbana, IL 62801, USA, ‖‖‖‖Colección Nacional de Insectos, Instituto de Biología, Universidad Nacional Autónoma de México, 3er. circuito exterior s/n, Cd. Universitaria, AP 70-233, Copilco Coyoacán, DF CP 04510, México*

Abstract

Microgastrine wasps are among the most species-rich and numerous parasitoids of caterpillars (Lepidoptera). They are often host-specific and thus are extensively used in biological control efforts and figure prominently in trophic webs. However, their extraordinary diversity coupled with the occurrence of many cryptic species produces a significant taxonomic impediment. We present and release the results of 8 years (2004–2011) of DNA barcoding microgastrine wasps. Currently they are the best represented group of parasitoid Hymenoptera in the Barcode of Life Data System (BOLD), a massive barcode storage and analysis data management site for the International Barcoding of Life (iBOL) program. There are records from more than 20 000 specimens from 75 countries, including 50 genera (90% of the known total) and more than 1700 species (as indicated by Barcode Index Numbers and 2% MOTU). We briefly discuss the importance of this DNA data set and its collateral information for future research in: (1) discovery of cryptic species and description of new taxa; (2) estimating species numbers in biodiversity inventories; (3) clarification of generic boundaries; (4) biological control programmes; (5) molecular studies of host-parasitoid biology and ecology; (6) evaluation of shifts in species distribution and phenology; and (7) fostering collaboration at national, regional and world levels. The integration of DNA barcoding with traditional morphology-based taxonomy, host records, and other data has substantially improved the accuracy of microgastrine wasp identifications and will significantly accelerate further studies on this group of parasitoids.

Keywords: cytochrome *c* oxidase, data release, deoxyribonucleic acid barcode, microgastrine, parasitoid

Received 5 January 2012; revision received 18 October 2012; accepted 22 October 2012

Correspondence: M. Alex Smith or José L. Fernández-Triana, Fax: 519-824-5703; E-mails: salex@uoguelph.ca or jftriana@uoguelph.ca

[1]These authors contributed equally to this work.

Introduction

DNA barcoding employs a short DNA sequence from a standardized gene region for the identification of specimens and species, as well as being a useful tool for the discovery of previously unrecognized provisional species that are often morphologically cryptic (Floyd *et al.* 2002; Hebert *et al.* 2003a,b). For animals, the accepted DNA barcode region is part of the mitochondrial gene, cytochrome *c* oxidase (COI or cox1). At present, the library of identified, photographed and georeferenced DNA barcodes, the Barcode of Life Data System (BOLD Ratnasingham & Hebert 2007), contains more than 1.4 million sequences from 117 thousand species. For groups where traditional taxonomic knowledge exists, the DNA barcode may be used in concert with that knowledge to streamline or supplement the identification process, but not necessarily to replace it. Barcodes have also been employed as a transparent first-pass survey of taxa where there is a paucity of other knowledge sources (Smith *et al.* 2009).

Microgastrinae wasps (Hymenoptera: Braconidae) are among the most species rich and numerous parasitoids of caterpillars (Lepidoptera), often being host-specific and thus extensively used in biological control efforts (Whitfield 1995, 1997). They are also very relevant in trophic web ecology (e.g. Smith *et al.* 2008, 2011). There are currently about 2000 described species (Yu *et al.* 2012) but, based on a recent study that used an array of focal study faunas to provide maximum and minimum boundaries for revised estimates of species richness (Rodriguez *et al.* 2012) we feel that a good estimate of species richness for Microgastrinae is likely 8–20 times the number of currently described species. The extraordinary species richness of the group, coupled with the occurrence of many cryptic species (e.g. Smith *et al.* 2008) produces a significant taxonomic impediment that hinders the advancement of their study.

The impediment is that production of species descriptions and scientific names that meet the requirements of the International Code of Zoological Nomenclature (ICZN) is much slower than desired. The problem is exacerbated for groups like Microgastrinae where a large number of DNA barcodes have been produced. In Microgastrinae, formal descriptions lag even farther behind because of the higher number of species discovered as a result of barcode production. The reverse situation is, of course, far more common—most species (of any taxon) are already described morphologically because morphological descriptions have been published for over 250 years (since Linnaeus in 1758, for animals), but most of these species do not, and may never, have barcodes produced. Until provisional species identified with barcodes are formally recognized with descriptions and valid scientific names that permit them to enter the taxonomic literature for use and discussion, they are 'invisible' to taxonomy. This greatly hinders discussion about them in ecological or biological studies and literature. Only when specimens with barcodes have formal species names will it be possible to extract information associated with them meaningfully from the scientific literature using key words, i.e. species names.

In this study, we release sequences, collateral information summary statistics and data from barcoding more than 20 000 microgastrine specimens during the last 8 years (2004–2011). As part of this release, we provide context to the novel data released here by comparing it to data already public as part of a policy of rapid data release (i.e. the Fort Lauderdale Principles; Wellcome Trust 2011; Schindel *et al.* 2011) and discuss the potential uses of the DNA barcode library assembled in BOLD, the barcode storage and analysis data management (Ratnasingham & Hebert 2007).

Methods

The word 'species' as used in this study is one used traditionally in insect taxonomy. If members of a 'species' are distinctive enough in some (usually morphological) trait(s), so as to be consistently distinguishable from other 'species' the assumption is that they are likely to be found to be relatively reproductively isolated from other similar 'species' if biological studies could be executed.

DNA extracts were prepared from single legs, or abdomens, using a glass-fibre protocol (Ivanova *et al.* 2006). Extracts were re-suspended in 30 μL of dH2O, and a 658-bp region near the 5′ terminus of the COI gene was amplified using primers (LepF1–LepR1) following standard protocols (Smith *et al.* 2008). Composite sequences were generated using internal primers when initial amplification was not successful (Smith *et al.* 2008). Primer information for individual sequences can be retrieved from BOLD using the Process IDs detailed in Appendix S1, but primers are as detailed in Smith *et al.* (2008). Sequence divergences were calculated using the K2P distance model and a NJ tree of K2P distance was created to provide a graphic representation of the among-species divergences. Full details of methodology are given in Smith *et al.* (2008).

All sequence data are publicly available in BOLD (www.barcodinglife.org) and can be examined and/or downloaded by searching within the Published Projects section (The records in this dataset can be retrieved from the public Data Portal using the following permanent URL: dx.doi.org/10.5883/DATASET-ASMIC1). Collection information and specimen accessions (BOLD and GenBank) for all sequences are listed in Appendix S1. The classification of Microgastrinae used here is the same

system as used in BOLD, and follows Mason (1981) and Whitfield (1995, 1997), Whitfield *et al.* (2002).

The data presented in this study come from different sources, institutions and research projects (Table 1). The following acronyms are used:

BIO – Biodiversity Institute of Ontario, Guelph, Ontario, Canada.

CAS – California Academy of Sciences, San Francisco, California, United States.

CNC – Canadian National Collection of Insects, Ottawa, Canada.

EMET – Entomology Museum, Faculty of Agriculture, Atatürk University, Erzurum, Turkey.

INHS – Illinois Natural History Survey, Champaign, Illinois, United States.

INBio – Instituto Nacional de Biodiversidad, Santo Domingo, Costa Rica.

HIC – Hymenoptera Institute Collection, University of Kentucky, Lexington, Kentucky, United States.

NARI- National Agriculture Research Institute, Papua New Guinea.

NHRS – Swedish Museum of Natural History, Stockholm, Sweden.

NMNH – National Museum of Natural History, the Smithsonian Institution, Washington DC, United States.

RMNH – Nationaal Natuurhistorische Museum (Naturalis), Leiden, The Netherlands.

UNAM – Universidad Nacional Autónoma de México, México D.F., México.

NZAC – New Zealand Arthropod Collection, Auckland, New Zealand.

Some results have already been published or are currently in press, representing a few hundred microgastrine barcodes (e.g. Smith *et al.* 2008, 2009; Fernández-Triana 2010; Hrcek *et al.* 2011; Rougerie *et al.* 2011; Quicke *et al.* 2012). However, the great majority of the sequences and related information presented here have not previously been made public (Appendix S1–S2).

Table 1 Snapshot of the barcoding of Microgastrinae braconids worldwide

Category	Count	
Specimens (sampled for barcodes)	27 103	
Specimens with sequences	18 730	
Specimens with barcode-complaint sequences	17 496	(*)
Diversity measured as MOTU	2221	(**)
MOTU named to species level in BOLD	1665	(**)

Data from BOLD (http://www.boldsystems.org/views/taxbrowser.php?taxid=2099) as of December, 2011. (*) – Barcode compliant sequences are those with >500 base pairs with less than 1% of base ambiguities. (**) – Species as estimated by Barcode Index Numbers (BINs) and 2% MOTU (jMOTU – Jones *et al.* 2011)).

Results and discussion

Currently, Microgastrinae is the group of parasitic Hymenoptera with the most geographically and taxonomically comprehensive DNA barcoding library in BOLD (Tables 1 and 2). Barcodes have been produced over the past 8 years from over 20 000 specimens, representing 50 genera (90% of the total known), and from 75 countries. The only described genera lacking DNA barcodes are *Austrocotesia*, *Exulonyx*, *Miropotes*, *Napamus* and *Semionis*.

The data set contains more than 1985 species (Fig. 1), as determined by using a 2% distance cut-off used for the determination of Molecular Operational Taxonomic Units (or MOTU – Jones *et al.* 2011) – the Barcode Index Numbers (BINs) (*sensu* http://www.barcodinglife.org/index.php/Public_BarcodeIndexNumber_Home). The decision of what threshold to apply is important, but arbitrary. No one threshold captures all species concepts or operational criteria. We have selected 2% as this has historically captured many species defined without molecular means. Close to 20% of those species represent actual described species. Two-thirds of the species have interim names, for ease of information management, until further taxonomic research assigns them a formal (binomial) name. For example, *Apanteles* Rodriguez10, refers to species 10 of *Apanteles* in Costa Rica, as identified by Rodriguez in host-specificity research (Smith *et al.* 2008) (see Fig. 2a and b). The rest of the species are identified only to generic level— or just to subfamily in the case of specimens deemed to represent new genera. Most of these species (63%) are represented by only one or two samples.

A massive DNA library such as this has already had an important impact in the study of Microgastrinae (e.g.

Table 2 Major sources of barcoded specimens up to December, 2011. See Methods for institutional acronyms

Country	Number of specimens	Major donor/repository institutions
Costa Rica	12 274	INBio, NMNH, CNC, INHS
Canada	4736	CNC, BIO
United States	1038	CNC, BIO
Papua New Guinea	921	NARI, NMNH
Thailand	739	HIC, CNC
New Zealand	647	NZAC, CNC
Mexico	602	UNAM
United Arab Emirates and Yemen	493	RMNH, CNC
Turkey	423	EMET
Sweden	421	NHRS, CNC
Republic of the Congo	350	HIC, CNC
Madagascar	267	CAS, CNC

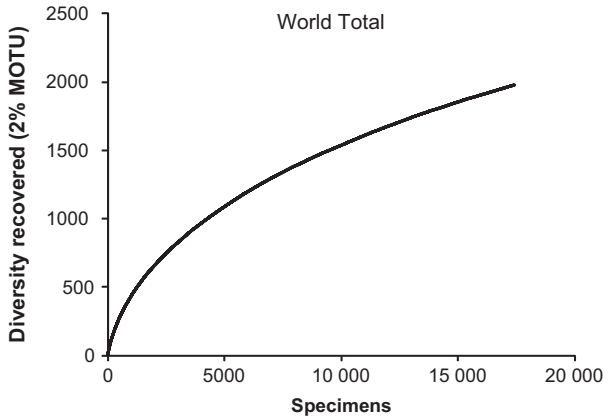

Fig. 1 Accumulation curve of barcode MOTU (2%) for all specimens with barcode sequences of more than 500 base pairs. Data from BOLD, October 31, 2011.

Smith *et al.* 2008, 2009; Fernández-Triana 2010; Fernández-Triana *et al.* 2011; Janzen & Hallwachs 2011; Janzen *et al.* 2009, 2011; Rougerie *et al.* 2011; Wilson *et al.* 2012). We anticipate that integration of DNA barcoding data with traditional taxonomy, host records and other data will play an even more significant role in the near future (Page *et al.* 2005). Below we examine some of these potential applications by presenting how we have used them with published datasets.

Discovery of cryptic species and description of new taxa

Microgastrine wasps have a high number of cryptic, morphologically similar species (e.g. Kankare *et al.*2005; Smith *et al.* 2008). Morphology-based taxonomic studies alone were unable to resolve this problem. DNA barcoding has revealed dramatic increases in species numbers from both tropical (Smith *et al.* 2008) and extra-tropical (Fernández-Triana 2010) biomes. These barcode discontinuities (e.g. Appendix S2) have been found to correlate with wasp biology and ecology, geographical distribution, and traits revealed *a posteriori* by careful and detailed morphological study. As in other zoological groups, barcodes have been incorporated into the description of new species of Microgastrinae (e.g. Grinter *et al.* 2009; Fernández-Triana 2010). In these studies adding barcodes to the array of other traits used to discover and describe species not only substantially improved the accuracy of identifications but also accelerated the study of this taxonomically difficult and hyper-diverse taxon (e.g. Whitfield *et al.* 2012; Wilson *et al.*2012).

Estimation of species numbers in biodiversity inventories

The application of DNA barcoding to biodiversity inventories is related to the previous topic. Several such

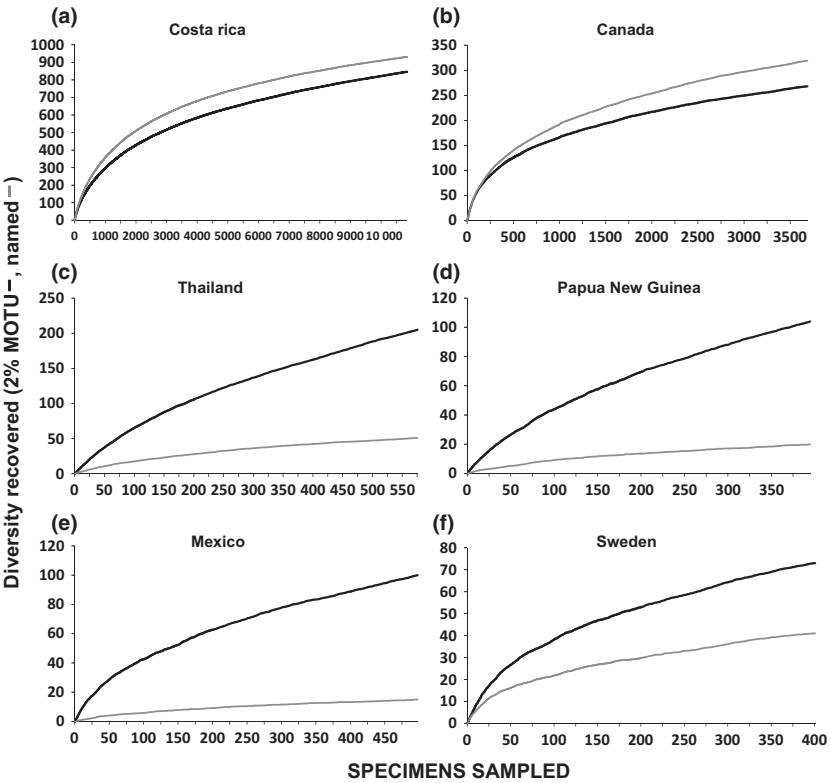

Fig. 2 Diversity accumulation curves using Barcode MOTU (BOLD BINs and 2% MOTU) in black and named diversity in grey. All data retrieved from BOLD on December, 2011. (a) Costa Rica, (b) Canada and Alaska, (c) Thailand, (d) Papua New Guinea, (e) Mexico, (f) Sweden. Note cases, such as in Canada and Costa Rica where the active working standards include interim names, (for ease of information management until further taxonomic research assigns them a formal (binomial) name.) based on barcode clusters and ecological divisions (such as host species). In cases such as these, the numbers of MOTU are surpassed by the number of named specimens exactly because the naming includes the barcode data supplemented with observed ecological divisions.

inventories, using DNA barcode molecular operational taxonomic units (MOTU – Floyd *et al.*2002) as a proxy to estimate species richness, are currently underway for the Arabian Peninsula, Costa Rica, Madagascar, Mexico, New Zealand, Sweden and Thailand. The BIN system employed by BOLD utilises such an approach and here we specifically use a 2% MOTU as implemented by the program jMOTU (Jones *et al.* 2011). In short: if two specimens yield sequences that are similar within some defined cut-off threshold, they are assigned to the same MOTU (Blaxter 2004). This is similar to the operational taxonomic unit (OTU) that has been used in the past in taxonomy based on morphology, biology, ecology, etc. The use of MOTUs was analysed in detail by Blaxter & Floyd (2003) and Blaxter (2004), and has also been used in a discussion of parasitoid wasp diversity by Smith *et al.* (2009).

It might be argued that the delineation of diversity using mitochondrial DNA MOTUs is controversial, in that the criterion used to separate them is not defined or specified. However, it is no more controversial than crude sorting of morphospecies as OTUs, and it is actually much better than OTUs for identifying cryptic species. Several approaches have been proposed to discriminate species using barcoding thresholds (Blaxter & Floyd 2003; Hebert *et al.* 2003a,b; Page *et al.* 2005), and we recognize that there can never be a single percentage discontinuity that is appropriate for species discrimination across most or all species, any more than there can be a single degree of morphological discontinuity for species discrimination. It is, however, worth noting that, even with only a range of values for barcoding MOTUs being usable (and that in combination with other traits), DNA barcoding still separates accurately more than 95% of the species of microgastrines in the published studies available (Smith *et al.* 2008, 2009; Fernández-Triana 2010; Fernández-Triana *et al.* 2011) as well as in the data we release here.

When barcode diversity, measured as MOTU, is used to generate species accumulation curves (Fig. 2), the results show a much higher diversity than is currently known with morphological OTUs. This situation occurs in each region of the planet that has been studied (Rodriguez *et al.* 2012) (apparent exceptions in Costa Rica and Canada (Fig. 2) occur due to the standard working procedures with these collections whereby interim names are generated based on barcode clusters and ecological divisions (such as host species). In these localities, the named specimens tally higher than MOTU diversity exactly because the naming included molecular data).

Generic boundaries in Microgastrinae

Although COI is not the ideal gene to unravel higher level phylogeny, and single genetic markers are clearly not as efficient as multiple markers (Whitfield *et al.* 2002; Quicke *et al.* 2012), the extensive library on barcoding available for the Microgastrinae offers further information about boundaries of some genera that have been controversial (e.g. Mason 1981; Austin & Dangerfield 1992; Whitfield 1995, 1997; Van Achterberg 2003; Banks & Whitfield 2006; Yu *et al.* 2012). Our study releases barcoding data for four pairs of genera considered either closely related or treated as synonyms by several authors (for examples examine the names current as of submission in Appendix S2 or alternatively, the reader can re-make the calculations on the Barcode of Life Data System).

The results suggest that *Apanteles* and *Dolichogenidea* are indeed different genera, with a small number of species 'out of place' that will need further morphological study to be resolved; the barcode data support the generic division proposed by Mason (1981). The limits of *Glyptapanteles* and *Protapanteles* as well as those of *Choeras* and *Sathon* are poorly defined (using barcodes as well as morphology) and more study will be needed to clarify their status and/or limits. For *Microplitis* and *Snellenius*, the results, though not conclusive, suggest that *Snellenius* is just a derived group within *Microplitis* and might therefore better be treated as a synonym of that genus (as suggested by Austin & Dangerfield 1992).

Further exploration of these problems requires the use of more genes and more detailed morphological study combined into an integrated phylogenetic study. Although a complete examination of this is beyond the scope of this study, it illustrates the potential of DNA barcoding to explore taxonomic boundaries within genera.

Biological control programs

The potential of DNA barcoding to improve biological control programs is large, particularly with regard to more accurate species identification and a better understanding of the boundaries among parasitoid species. Past failures in pest control have often been due to the introduction or release of an incorrectly identified wasp species (Huber *et al.* 2001). Misinterpreting host range would result in potential candidate species being removed from consideration because the host range was considered to be too broad (when it was later shown that, in fact, several cryptic species existed, each with a narrower host range e.g. Smith *et al.* (2011)). We present here two examples of the potential of DNA barcoding to improve existing biological control programmes in Canada using microgastrine wasps.

The bertha armyworm, *Mamestra configurata* Walker (Noctuidae) is a major pest of canola in Canada (Mason *et al.* 1998). As part of the efforts to control this caterpillar, Canada has been conducting studies for the possible

importation and release of the microgastrine *Microplitis mediator* (Haliday) from Europe (Mason *et al.* 2001). Preliminary analysis of the specimens for which there are barcoding sequences available shows that the European specimens are identical to the barcodes of a native Nearctic species, *M. varicolor* Viereck, a wasp that has not been reported as attacking *Mamestra*. It also seems that there are more species and/or confusion in the taxonomy of the specimens involved (see Appendix S2). Furthermore, review of the literature suggests that *M. mediator* is associated with >40 host species in three Lepidopteran families (Mason and Fernández-Triana unpublished data). Based on our experience, this combination of barcode division associated with host-species isolation supports the hypothesis that *M. mediator* actually comprises multiple cryptic species. Barcoding data can be used as a comparatively rapid and easy first approach that brings such information to the attention of taxonomists, which can help further efforts in characterizing the species. This particular example calls attention to the urgent need to unravel this complex. Although it is theoretically important to clarify species boundaries, pragmatically we need to determine the necessity of importing the European species.

Recently, a research programme in Ottawa, Canada, accidentally discovered what appears to be a new species of *Cotesia* that attacks the diamondback moth *Plutella xylostella* (L.), a major pest of cruciferous crops (Sarfraz *et al.* 2005). A preliminary morphological study of the specimens (Fernández-Triana unpublished data) revealed that they do not belong to any of the eight species of *Cotesia* already recorded parasitizing *P. xylostella* (Yu *et al.* 2012). Barcodes were obtained for 20 specimens (named as *Cotesia* jft03 in Appendices S1–S2) and the results show that they do not represent any of the 200+ species of *Cotesia* already barcoded and available for comparison in BOLD (see Appendix S2). The wasp is naturally occurring in several localities in Ontario and Manitoba, Canada (Fernández-Triana unpublished data); if it is a native species then it represents a species with great potential to be used in biological control efforts against *P. xylostella* (Mason unpublished data)— further work in Europe and North America should help determine this potential.

Molecular studies of host-parasitoid biology and ecology

The spruce budworm and relatives (*Choristoneura* spp., Tortricidae) are among the most important forest pests in North America (Huber *et al.* 1996). Among the 230 species of parasitoids reported parasitizing the caterpillars were 50 species of Braconidae, half of them microgastrines (Fernández-Triana & Huber 2010). Although Mason (1974) had studied the Microgastrinae from conifer-feeding *Choristoneura* and described three new species, Fernández-Triana & Huber (2010) suggested that more species were involved. Using an integrated taxonomic approach (i.e. morphology, host records and barcoding), Fernández-Triana (2010) showed that *Apanteles* was represented by a complex of species and described a further two new species. Barcoding data was recently used to reinterpret the extensive data available on spruce budworm food webs in eastern Canada (Eveleigh *et al.* 2007; Smith *et al.* 2011). Interestingly, while DNA barcode identification of the 'nodes' of this food web did result in an increased compartmentalisation of the total web, the tested measures of ecosystem function were strengthened using barcoding MOTUs (Smith *et al.* 2011). Yet, in spite of the amount and scope of data in those studies, gaps still exist in our understanding of pest-parasitoid relationships within this complex. The use of barcoding data revealed a confusion of names and misidentifications, even though they were made by experienced taxonomists, as well as potentially more undescribed species (see Appendix S2 searching for *Apanteles*). The pattern that emerged strongly suggests that we are confusing and misidentifying the parasitoid species, which, if not corrected, may have a negative effect on better understanding the ecological interrelationships of pest *Choristoneura* spp. and their microgastrine parasitoids. It should also be noted that the genus *Choristoneura* comprises a cryptic species complex by itself, which has been similarly studied by an integration of morphology and mitochondrial DNA (Lumley & Sperling 2010).

An exciting new technique that has become recently available offers a tremendous possibility to increase what we know about host-parasitoid biology. It has been shown that host DNA can be extracted from the gut of the adult parasitoid wasp, allowing the host identification to be obtained from wild-caught adults (Rougerie *et al.* 2011). This technique has been called Molecular Analysis of Parasitoid Linkages (MAPL). Although MAPL cannot substitute for the practice of rearing host caterpillars to see what species are attacking them, i.e. the traditional way of obtaining host information, it may provide an easier, faster and cheaper way to gather data when hosts have never been recorded for a parasitoid species, or where hosts are difficult to find. The reverse situation, where hosts are collected and parasitoid DNA (if present) within them is identified (e.g. Hrcek *et al.* 2011), may allow parasitism rates in economically important pests to be determined more accurately and perhaps more quickly. Combining both approaches has the potential of revolutionizing studies of caterpillar-parasitoid trophic webs as well as biological control efforts.

The MAPL approach was proposed after obtaining data from three wasp species (two of which were microgastrines). However, it has never been tested on a large scale, owing to the lack of an *a priori* comprehensive

library of host DNA with which to match the sequences obtained from the parasitoid gut. In that regard, comprehensive sympatric host and parasitoid barcode libraries from one place, such as those being amassed for Area de Conservación Guanacaste in north-western Costa Rica (Janzen *et al.* 2009, 2011) or in Papua New Guinea (Hrcek *et al.* 2011) will make such large-scale analyses possible.

The massive DNA barcode libraries already available for Lepidoptera caterpillars (e.g. Hebert *et al.* 2009) and Microgastrinae wasps (data released in this study), likely ensure that they will become the first model groups of host/parasitoid associations where the MAPL approach could be tried extensively.

Evaluation of shifts in species distribution and phenology

DNA barcode surveys of other museum collections and comparisons to contemporary collecting events will similarly produce rich opportunities for the documentation of faunal shifts. Barcoding can also add scientific value to standard museum specimens, as the information they contain is revealed through molecular analyses that place the specimens in a population context.

DNA barcoding offers the possibility of comparing specimens from historical collections with freshly collected material, allowing for evaluation of shifts in place-specific species composition through time. This has been done for a single locality in sub-Arctic Canada (Smith *et al.* 2009; Fernández-Triana *et al.* 2011; see below), but has the potential to be done elsewhere and also to be applied on a larger scale, wherever historical collections are available.

In Canada, a legacy of scientific investigation in Churchill, Manitoba, dating back to the early 1930s and continuing in the current Northern Insect Survey (NIS), has opened the possibility of species diversity comparisons over several decades. The NIS was active throughout the 1940s and 1950s when Canadian scientists made extensive insect collections, deposited in the CNC in Ottawa, Ontario (Fernández-Triana *et al.* 2009). The impressive diversity of parasitoid Hymenoptera was immediately apparent to these researchers, as stated by McClure (1943) '*The number of species of parasitic Hymenoptera was great and they were present in all types of terrestrial habitats*' and Freeman & Twinn (1954) '*… as the scanty material so far obtained from this coast indicates a rich fauna of parasitic Hymenoptera*'. In a high-latitude locality such as Churchill, where climate change is anticipated to accelerate the arrival and departure of many species, using these diverse holdings to compare and contrast the early to mid 20th century to the early 21st century appeared ideal. However, McClure (1943) stated most specimens had only a family or genus level identification as '*Identification of these beyond genus was nearly impossible because of lack of associated host relationships*' and the specimens are still unnamed to species, for the most part. Thus, although known to be diverse and while being held in a major, scientifically accessible collection, the Churchill specimens are a good example of the taxonomic impediment that prevented their use as biological sentinels for a changing climate. Successfully obtaining DNA fragments, often even smaller than the standard barcoding locus (Hajibabaei *et al.* 2006), from these historic collections has permitted the comparison of specimens from this locality through time. Using DNA barcodes from microgastrine specimens collected between 2005 and 2007 (Smith *et al.* 2009) the contemporary diversity was compared to what was documented within Churchill and the surrounding north between 1930 and 1960. The results showed very little overlap in species between the two collection periods and, furthermore, showed a much larger regional affinity between the south and the contemporary collections contrasting clearly with the north and the historic collections (Fernández-Triana *et al.* 2011). DNA barcoding of the contemporary and the historic collections at this one locality allowed identification of a dramatic biological shift, despite a taxonomic impediment.

Fostering collaboration at national, regional and world level

The BOLD database provides a unique environment for sharing data across projects, and has already catalysed many analyses that would not have happened in the traditional environment—where it was not easy to access data from specimens other than those stored by the researcher or collaborating institutions. Examples include published studies on Costa Rican microgastrines (Smith *et al.* 2008; Grinter *et al.* 2009; Janzen *et al.* 2011), where specialists from several institutions have analysed specimens after barcodes became available; and also Canada (Fernández-Triana 2010; Fernández-Triana *et al.* 2011; Smith *et al.* 2011) where several institutions shared information on barcoded specimens. And there are currently a number of projects underway (e.g. Table 2) which exemplify those approaches.

Collaboration will also be of critical importance to fill the current gaps in geographical coverage, species not barcoded yet, and type specimens – whenever possible.

Conclusions

The examples we present here illustrate the great potential that DNA barcoding has for answering questions in a variety of areas relating to parasitoid taxonomy, biology and ecology. Although the taxonomic impediment still

exists (too few taxonomists for the enormous number of species that need formal description, especially in parasitic wasps, which make up by far the greatest proportion of Hymenoptera; Huber (2009)), information from barcoding greatly helps to resolve the impediment and will possibly accelerate the process of species description. Barcoding combined with morphologically based classical taxonomy, can accelerate the identification of morphologically cryptic species within species complexes. This, in turn, will hopefully encourage taxonomists to revisit those complexes for more detailed study of their morphology, resulting in the possible discovery of unnoticed or unappreciated characters or differences in features that previously might have been overlooked or treated as intraspecific variation. Over time, integrating morphological and barcode information sources should enlighten each other reciprocally and together provide a stronger base for a better understanding of parasitoid taxonomy and diversity in general.

Acknowledgements

JFT is greatly indebted to all curators that kindly sent extensive samples of Microgastrinae and allowed some of them to be barcoded: Kees van Achterberg (the Netherlands), Antonius van Harten (UAE), Mark Shaw (UK), Michael Sharkey and Robert Zuparko (US), the Sweddish Malaise Trap Project (Sweden) and many Canadian institutions. JBW was supported by NSF grants DEB 07-17402 and 1020510. DHJ supported by NSF DEB 0515699 and DHJ and WH emphatically and gratefully acknowledge the support of the parataxonomist team in finding and rearing the caterpillars and their parasites from Area de Conservación Guanacaste, Costa Rica and the Guanacaste Dry Forest Conservation Fund and the Wedge Foundation for funding portions of the research. MAS was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant. JR was supported by a National Center for Ecological Analysis and Synthesis postdoctoral fellowship and NSF Doctoral Dissertation Enhancement Program award OISE-0809175. DFW and JFT were supported by The "Defining New Zealand's Land Biota" programme (MSI), and an International Mobility Fund grant. AZR was supported by grants given by CONACyT (Red del Código de Barras de la Vida; proyecto SEP-Ciencia Básica convocatoria 2008). EE gratefully acknowledges the support of Natural Resources Canada, Canadian Forest Service. Laboratory analyses on sequences generated since 2009 were funded by the Government of Canada through Genome Canada and the Ontario Genomics Institute (2008-0GI-ICI-03). For those specimens collected in Canadian National Parks by the Biodiversity Institute of Ontario's BIObus program we gratefully acknowledge the support from Parks Canada (permit NAP-2008-1636). The Papua New Guinea work, in collaboration with Vojtech Novotny, George Weiblen and Yves Basset, was 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).

References

Austin AD, Dangerfield PC (1992) Synopsis of Australasian microgastrinae (Hymenoptera: Braconidae), with a key to genera and description of new taxa. *Invertebrate Taxonomy*, **6**, 1–76.

Banks JC, Whitfield JB (2006) Dissecting the ancient rapid radiation of microgastrine wasp genera using additional nuclear genes. *Molecular Phylogenetics and Evolution*, **41**, 690–703.

Blaxter ML (2004) The promise of a DNA taxonomy. *Philosophical Transactions of the Royal Society of London B*, **359**, 669–679.

Blaxter ML, Floyd R (2003) Molecular taxonomics for biodiversity surveys: already a reality. *Trends in Ecology and Evolution*, **18**, 268–269.

Eveleigh ES, McCann KS, McCarthy PC *et al.* (2007) Fluctuations in density of an outbreak species drive diversity cascades in food webs. *Proceedings of the National Academy of Science, USA*, **104**, 16976–16981.

Fernández-Triana J (2010) Eight new species and an annotated checklist of Microgastrinae (Hymenoptera: Braconidae) from Canada and Alaska. *Zookeys*, **63**, 1–53.

Fernández-Triana J, Huber JT (2010) The braconid parasitoids (Hymenoptera: Braconidae) of Nearctic *Choristoneura* species (Lepidoptera: Tortricidae), with a summary of other parasitoid families attacking *Choristoneura*. *The Canadian Entomologist*, **142**, 295–343.

Fernández-Triana J, Boudreault C, Goulet H (2009) Revisiting the northern insect survey: preliminary results for microgastrinae wasps (Hymenoptera: Braconidae). *Newsletter of the Biological Survey of Canada (Terrestrial Arthropods)*, **28**, 21–25.

Fernández-Triana J, Smith MA, Boudreault C *et al.* (2011) A poorly known high-latitude parasitoid wasp community: unexpected diversity and dramatic changes through time. *PLoS ONE*, **6**, e23719.

Floyd R, Eyualem A, Papert A, Blaxter ML (2002) Molecular barcodes for soil nematode identification. *Molecular Ecology*, **11**, 839–850.

Freeman T, Twinn C (1954) Present trends and future needs of entomological research in Northern Canada. *Arctic*, **1954**, 275–283.

Grinter CC, Whitfield JB, Connahs H, Dyer LA, Hallwachs W, Janzen DH (2009) A key to new world *Distatrix* mason (Hymenoptera: Braconidae), with descriptions of six new reared Neotropical species. *Journal of Insect Science*, **9**, 1–16.

Hajibabaei M, Smith MA, Janzen DH, Rodriguez JJ, Whitfield JB, Hebert PDN (2006) A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes*, **6**, 959–964.

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003a) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B*, **270**, 313–321.

Hebert PDN, Ratnasingham S, deWaard JR (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B*, **270**(Suppl. 1), S96–S99.

Hebert PDN, deWaard JR, Landry JF (2009) DNA barcodes for 1/1000 of the animal kingdom. *Biology Letters*, **6**, 359–362.

Hrcek J, Miller SE, Quicke DLJ, Smith MA (2011) Molecular detection of trophic links in a complex insect host–parasitoid food web. *Molecular Ecology Resources*, **11**, 786–794.

Huber JT (2009) Biodiversity of Hymenoptera. In: *Insect Biodiversity: Science and Society* (eds Foottit RG & Adler PH), pp. 303–323. Wiley-Blackwell, Chichester, UK. 632 pp.

Huber JT, Eveleigh ES, Pollock S, McCarthy PC (1996) The chalcidoid parasitoids and hyperparasitoids (Hymenoptera: Chalcidoidea) of *Choristoneura* species (Lepidoptera: Tortricidae) in America north of Mexico. *The Canadian Entomologist*, **128**, 1167–1220.

Huber JT, Darbyshire S, Bissett J, Foottit RG (2001) Taxonomy and Biological Control. In: *Biological Control Programmes in Canada 1981–2000* (eds Mason PG & Huber JT), pp. 14–22. CABI Publishing, Wallingford, UK.

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 (2011) Joining inventory by parataxonomists with DNA barcoding of a large complex tropical conserved wildland in northwestern Costa Rica. *PLoS ONE*, **6**, e18123.

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.

Janzen DH, Hallwachs W, Blandin P *et al.* (2011) Joining inventory by parataxonomists with DNA barcoding of a large complex tropical conserved wildland in northwestern Costa Rica. *PLoS ONE*, **6**, e18123.

Jones M, Ghoorah A, Blaxter M (2011) jMOTU and taxonerator: turning DNA barcode sequences into annotated operational taxonomic units. *PLoS ONE*, **6**, e19259.

Kankare M, Stefanescu C, Van Nouhuys , Shaw MR (2005) Host specialization by *Cotesia* wasps (Hymenoptera: Braconidae) parasitizing species-rich Melitaeini (Lepidoptera: Nymphalidae) communities in north-eastern Spain. *Biological Journal of the Linnean Society*, **86**, 45–65.

Lumley LM, Sperling FAH (2010) Integrating morphology and mitochondrial DNA for species delimitation within the spruce budworm (*Choristoneura fumiferana*) cryptic species complex (Lepidoptera: Tortricidae). *Systematic Entomology*, **35**, 416–428.

Mason WRM (1974) The *Apanteles* species (Hymenoptera: Braconidae attacking Lepidoptera in the micro-habitat of the spruce budworm (Lepidoptera: Tortricidae). *The Canadian Entomologist*, **106**, 1087–1102.

Mason WRM (1981) The polyphyletic nature of *Apanteles* Foerster (Hymenoptera: Braconidae): a phylogeny and reclassification of Microgastrinae. *Memoirs of the Entomological Society of Canada*, **115**, 1–147.

Mason PG, Arthur AP, Olfert OO, Erlandson MA (1998) The bertha armyworm (*Mamestra configurata* Walker: Lepidoptera: Noctuidae) in Western Canada. *The Canadian Entomologist*, **130**, 321–336.

Mason PG, Erlandson MA, Youngs BJ (2001) Effects of parasitism by *Banchus flavescens* (Hymenoptera: Ichneumonidae) and *Microplitis mediator* (Hymenoptera: Braconidae) on the bertha armyworm, *Mamestra configurata* (Lepidoptera: Noctuidae). *Journal of Hymenoptera Research*, **10**, 81–90.

McClure HE (1943) Aspection in the biotic communities of the Churchill area, Manitoba. *Ecological Monographs*, **13**, 1–35.

Page TJ, Choy SC, Hughes JM (2005) The taxonomic feedback loop: symbiosis of morphology and molecules. *Biology Letters*, **1**, 139–142.

Quicke DLJ, Smith MA, Janzen DH *et al.* (2012) Utility of the DNA barcoding gene fragment for parasitic wasp phylogeny (Hymenoptera: Ichneumonoidea): data release and a new measure of taxonomic congruence. *Molecular Ecology Resources*, **12**, 676–685.

Ratnasingham S, Hebert PDN (2007) BOLD: the barcode of life data system (www.barcodinglife.org). *Molecular Ecology Notes*, **7**, 355364.

Rodriguez JJ, Fernández-Triana J, Smith MA *et al.* (2012) Extrapolations from field studies and known faunas converge on dramatically increased estimates of global microgastrine parasitoid wasp species. *Insect Conservation and Diversity*. In press. doi: 10.1111/icad.12003

Rougerie R, Smith MA, Fernández-Triana JL, Lopez-Vaamonde C, Ratnasingham S, Hebert PDN (2011) Molecular analysis of parasitoid linkages (MAPL): gut contents of adult parasitoid wasps reveal larval host. *Molecular Ecology*, **20**, 179–186.

Sarfraz M, Keddie AB, Dosdall LM (2005) Biological control of the diamondback moth, Plutella xylostella: a review. *Biocontrol Science and Technology*, **15**, 763–789.

Schindel D, Stoeckle M, Milensky C *et al.* (2011) Project description: DNA barcodes of bird species in the national museum of natural history, Smithsonian Institution, USA. *ZooKeys*, **152**, 87–91.

Smith MA, Rodríguez 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, US*, **105**, 12359–12364.

Smith MA, Fernández-Triana JL, Roughley R, Hebert PDN (2009) DNA barcode accumulation curves for understudied taxa and areas. *Molecular Ecology Resources*, **9**(Suppl. 1), 208–216.

Smith MA, Eveleigh ES, McCann KS, Merilo MT, McCarthy PC, van Rooyen KI (2011) Barcoding a quantified food web: crypsis, concepts, ecology and hypotheses. *PLoS ONE*, **6**, e14424.

Van Achterberg C (2003) Western Palaearctic genera of the subfamily Microgastrinae: a reappraisal of the generic and tribal division (Hymenoptera: Braconidae). In: *Parasitic Wasps: Evolution, Systematics, Biodiversity and Biological Control* (eds Melika G & Thuróczy G), pp. 19–35. Agroinform, Budapest, Hungary.

Wellcome Trust (2011) Sharing data from large-scale biological research projects: a system of tripartite responsibility; Report of a meeting organized by the Wellcome Trust and held on 14–15 January 2003 at Fort Lauderdale, USA", http://www.genome.gov/Pages/Research/WellcomeReport0303.pdf.

Whitfield JB (1995) Annotated checklist of the Microgastrinae of North America north of Mexico (Hymenoptera: Braconidae). *Journal of the Kansas Entomological Society*, **68**, 245–262.

Whitfield JB (1997) Microgastrinae. In: *Manual of the New World Genera of the Family Braconidae (Hymenoptera)* (eds Wharton R, Marsh P & Sharkey MJ), vol. 1, pp. 333–366. Special Publication of the International Society of Hymenopterists, Washington.

Whitfield JB, Mardulyn P, Austin AD, Dowton M (2002) Phylogenetic relationships among microgastrine braconid wasp genera based on data from the 16S, COI and 28S genes and morphology. *Systematic Entomology*, **27**, 337–359.

Whitfield JB, Fernández-Triana J, Janzen DH, Hallwachs W, Smith MA, Cardinal S (2012) *Mariapanteles* (Hymenoptera: Braconidae), a new Neotropical genus of microgastrine parasitoid wasp discovered through biodiversity inventory. *ZooKeys*, **208**, 61–80.

Wilson JS, Forister ML, Dyer LA *et al.* (2012) Host conservatism, host shifts and diversification across three trophic levels in two Neotropical forests. *Journal of Evolutionary Biology*, **25**, 532–546.

Yu DSK, van Achterberg C, Horstmann K (2012) Taxapad 2012, Ichneumonoidea 2011. Database on flash-drive. www.taxapad.com, Ottawa, Ontario, Canada.

Conceived and designed the experiments: M.A.S., J.L.F-T. Performed the experiments: M.A.S., J.L.F-T. Analyzed the data: M.A.S., J.L.F-T. Contributed reagents/materials/analysis tools: M.A.S., J.L.F-T., E.S.E., J.G., C.G., W.H., P.D.N.H., J.H., J.T.H., D.H.J., P.G.M., S.M., D.L.J.Q., J.J.R., R.R., M.R.S., G.V.A., D.F.W., J.B.W., A.Z. Wrote the paper: M.A.S., J.L.F-T., E.S.E., J.G., C.G., W.H., P.D.N.H., J.H., J.T.H., D.H.J., P.G.M., S.M., D.L.J.Q., J.J.R., R.R., M.R.S., G.V.A., D.F.W., J.B.W., and A.Z.

Data accessibility

The contents of Appendix S1 comprise an Excel file that documents collection information, current level of identification at time of submission and various accessions for each individual specimen included in the data release.

Supporting information

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

Appendix S1 Excel file documenting collection information, current level of identification at time of submission and various accessions for all the specimens from the data release.

Appendix S2 A Neighbor-joining tree made with 18 518 sequences from specimens that have sequences with greater than 500 bp. Taxonomy evident in tip labels represent the taxonomic state as of October 2012 and will inevitably change as careful and detailed morphological study continues.