

BARCODING ARTHROPODS

DNA barcode accumulation curves for understudied taxa and areas

M. ALEX SMITH,* JOSE FERNANDEZ-TRIANA,* ROB ROUGHLEY† and PAUL D. N. HEBERT*

*Biodiversity Institute of Ontario, Department of Integrative Biology, University of Guelph, Guelph, Ontario, Canada N1G 2W1,

†Department of Entomology, University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2

Abstract

Frequently, the diversity of umbrella taxa is invoked to predict patterns of other, less well-known, life. However, the utility of this strategy has been questioned. We tested whether a phylogenetic diversity (PD) analysis of CO1 DNA barcodes could act as a proxy for standard methods of determining sampling efficiency within and between sites, namely that an accumulation curve of barcode diversity would be similar to curves generated using morphology or nuclear genetic markers. Using taxa at the forefront of the taxonomic impediment – parasitoid wasps (Ichneumonidae, Braconidae, Cynipidae and Diapriidae), contrasted with a taxon expected to be of low diversity (Formicidae) from an area where total diversity is expected to be low (Churchill, Manitoba), we found that barcode accumulation curves based on PD were significantly different in both slope and scale from curves generated using names based on morphological data, while curves generated using nuclear genetic data were only different in scale. We conclude that these differences clearly identify the taxonomic impediment within the strictly morphological alpha-taxonomy of these hyperdiverse insects. The absence of an asymptote within the barcode PD trend of parasitoid wasps reflects the as yet incomplete sampling of the site (and more accurately its total diversity), while the morphological analysis asymptote represents a collision with the taxonomic impediment rather than complete sampling. We conclude that a PD analysis of standardized DNA barcodes can be a transparent and reproducible triage tool for the management and conservation of species and spaces.

Keywords: Braconidae, Churchill, CO1, Cynipidae, cytochrome *c* oxidase 1, Diapriidae, Formicidae, Ichneumonidae, Manitoba, phylogenetic diversity, species abundance curve, triage

Received 22 October 2008; revision received 16 January 2009; accepted 27 January 2009

Introduction

Ongoing habitat destruction, degradation and fragmentation have resulted in escalating potentially catastrophic losses of biodiversity (Vitousek *et al.* 1997; Hoffmeister *et al.* 2005). The role of biodiversity estimates in the identification and protection of areas of overlap and endemism have traditionally been provided through the infrastructure of morphological taxonomy (Godfray & Knapp 2004). However, taxonomists are not evenly distributed throughout the entirety of biological diversity, and what expertise that

does exist tends to be biased towards larger, more 'charismatic' fauna (Renner & Häuser 2007). As a consequence, patterns of diversity within these groups are frequently invoked to predict patterns of other, less well-known, frequently smaller-sized, taxa [umbrella, flagship or surrogates (Andelman & Fagan 2000)]. Unfortunately, the utility of a strategy of surrogacy is questionable due to the radically different processes that drive patterns of diversity of different grain size (dispersal, history, etc.). Surrogacy may be most attractive to conservationists largely because of their recognition of incomplete understanding of a particular habitat or area (Andelman & Fagan 2000).

Although smaller, primarily invertebrate taxa are known to turnover at very small spatial scales (Fisher *et al.* 2000) and potentially provide a very finely grained information

Correspondence: M. Alex Smith, Fax: 519-824-5703;

E-mail: salex@uoguelph.ca

source for biodiversity assessment (Underwood & Fisher 2006) — their use as a management or conservation tool has been constricted by the taxonomic impediment (Bolton 1994; Gotelli 2004). There are simply far too many unnamed and/or unknown taxa to assure funding or management agencies that the biological currency of what is being compared between areas/times are truly the same units because of the absence of a taxonomic currency. DNA barcoding, a process whereby a single, standard locus is used as a species-level-equivalent tag (Hebert *et al.* 2003), may provide conservationists and managers with a method that permits the inclusion of smaller-sized taxa in biological surveys (Smith *et al.* 2005), both through the democratization of access to taxonomy through DNA (Janzen 2004) and through the new acceleration of alpha-taxonomy through an integrative taxonomy that includes digital information such as barcoding (Fisher & Smith 2008).

However, this does not completely exploit the potentiality of the technology, as the high-throughput capacity of a DNA pipeline remains constrained by the taxonomic impediment (increased efficiency in morphospecies recognition is only truly efficient if there are available human resources skilled in measuring and characterizing the resultant piles). One important adaptation of the capacity that a DNA barcoding approach provides to biodiversity surveys is as a standardized measure of diversity [phylogenetic diversity or PD (Faith 1994; Crozier 1997)] within and between sites (for example, see Forest *et al.* 2007). Using a single gene to estimate a phylogeny is widely accepted as unwise (Maddison 1997). However, here it is more precise to say that we are using a single-gene DNA barcode's phyletic (or lineage) diversity rather than phylogenetic (or of relating to the evolutionary history and development of a species or taxonomic grouping vs. the historical pattern of relationships between species or other groups resulting from divergence from their common ancestor) to represent the 'cloud of gene histories that ... are part of the species tree' (Maddison 1997), and we are not concluding that the mitochondrial gene tree represents the totality of the entire species tree. Indeed, while not being a phylogenetic analysis on its own, barcoding as a tool for rapid initial biodiversity assessment can lead to subsequent systematic and nomenclatorial studies to the extent that additional data support this move (Smith *et al.* 2005; Fisher & Smith 2008).

We tested whether the phyletic measure of this mitochondrial DNA (mtDNA) tree is a reasonable and useful proxy of diversity information when we cannot hope to know the entire species tree on timescales needed for conservation or ecology. Calibrating such a PD approach to diversity assessment would offer a researcher the full speed of a DNA-based approach. Furthermore, it allows one to avoid the issue of species/barcode equivalence if the trends produced in proof of principle comparisons provide

nonsignificantly different patterns of identification (as in Smith *et al.* 2005). This could provide significant acceleration in terms of between-site comparisons (as in Smith *et al.* 2005; Forest *et al.* 2007), but also for within-site measures of sampling efficiency. In the case of the latter, a standard measure of efficient sampling, a species accumulation curve (Gottelli & Colwell 2001), could be augmented by the calculation of a DNA barcode accumulation curve (DBAC). Because of the standardization of the approach, inferences based on such a curve would be transparent to future researchers and directly comparable across time and space to other sampling regimes and taxa.

We tested whether PD assessments of collections made in sub-Arctic Canada (Churchill, Manitoba) in 2006 for parasitoid wasps and ants: (i) would increase in an asymptotic fashion with sampling intensity; (ii) that this accumulation would not be significantly different from the taxonomic accumulation identified using such names as could be assigned based on morphology; and (iii) further, that this barcode trend would not be significantly different from patterns using a single nuclear locus.

Supporting these hypotheses would suggest that an analysis of DNA barcode PD could be a critical triage tool for conservation biologists and managers, providing transparent and reproducible estimates of within- and between-site diversity at a rate that is unapproachable for the most taxonomically labile groups and virtually impossible for the vast majority of life where the viscosity of taxonomic information has historically discouraged the inclusion of the small and the unknown in surveys of diversity.

Methods

Field collection and subsampling

We studied two different Hymenoptera groups: ants (Formicidae) and parasitoid wasps (Braconidae, Ichneumonidae, Cynipidae and Diapriidae). Samples were collected as part of the PROBE campaign carried out between 2005–2006 (ants) and 2006–2007 (parasitoid wasps). Specimens were collected using Malaise traps and pitfall traps, checked and emptied once per week. The traps were set up at 11 locations in the vicinity of the town of Churchill.

The specimens were preserved in 100% ethanol and transported back to the Biodiversity Institute of Ontario, Guelph (ants) and the Canadian National Collection of Insects, Arachnids and Nematodes (CNC), Ottawa (parasitoid wasps), where they were removed from ethanol in no intended order — and thus sampling should approach random — although it was not designed to be a priori. The specimens were pinned or placed in gel-capsules, photographed and a single leg removed and placed in a 96-well

lysis plate. These plates were then shipped to the University of Guelph where DNA extraction, amplification and sequencing took place.

Laboratory

Standard laboratory procedures were followed at the Canadian Centre for DNA Barcoding (CCDB, see www.dnabarcoding.ca for updates to protocols). Briefly, total genomic DNA extracts were prepared from small pieces of tissue using standard protocols (Ivanova *et al.* 2006). Extracts were resuspended in 20–30 μL of dH_2O . A 658 base-pair (bp) region near the 5' terminus of the CO1 gene was amplified using primers LepF1/LepRI. In cases where a 658-bp product was not successfully generated, internal primer pairs (LepF1-C_ANTMR1), (MLepF1/LepRI) and (RonMWaspDeg_t1/LepRI) were employed to generate shorter sequences. These could be overlapped to create composite sequence (contig) or could be analysed as shorter, non-barcode length standard sequences (see Supporting Information, Appendix S1 for a complete list of primers and sources).

Polymerase chain reactions were carried out in 96-well plates in 12.5 μL reaction volumes containing: 2.5 mM MgCl_2 , 1.25 μM of each primer, 50 μM dNTPs, 10 mM Tris HCl (pH 8.3), 50 mM KCl, 10–20 ng (1–2 μL) of genomic DNA, and 0.3 U of *Taq* DNA polymerase (Platinum *Taq* DNA Polymerase; Invitrogen) using a thermocycling profile of one cycle of 2 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 36 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. Products were visualized on a 2% agarose E-Gel 96-well system (Invitrogen) and samples containing clean single bands were bidirectionally sequenced using BigDye version 3.1 on an ABI 3730xl DNA Analyser (Applied Biosystems).

Sequences were trimmed in Sequencher version 4.0.5 (Gene Codes) where no more than 25% of the sequence was trimmed until the first 25 bases contained fewer than five ambiguities. Contigs were subsequently constructed (using 90% minimum match and a minimum 20% overlap) and primer sequences trimmed (by eye) in Sequencher. Sequences contained no indels and were aligned by eye in BioEdit (Hall 1999). Sequence divergences were calculated using the Kimura 2-parameter distance model (K2P; Kimura 1980) and a neighbour-joining (NJ) tree of distances (Saitou & Nei 1987 was created using BOLD (Ratnasingham & Hebert 2007) and MEGA 3 (Kumar *et al.* 2004).

Sequences and all other specimen information are available in the project file 'Barcode Accumulation Curves' (BACAS) in the Published Projects section of the Barcode of Life Data System (www.barcodinglife.org). All sequences from the barcode region have been deposited in GenBank (FJ413054–FJ415064).

Complementary genetic analyses

For 273 wasp specimens and 100 ants, we also amplified portions of the LSU rRNA gene region (28S). Within the variable D2 region, the forward primer corresponds to positions 3549–3568 in *Drosophila melanogaster* reference sequence (GenBank M21017). Representative sequences have been deposited in GenBank: (FJ396166–FJ396437, FJ407298–FJ407397) while the primers used to generate these fragments are detailed in Appendix S1.

Analysis

Within our test of a DBAC style analysis, the unit of sampling selected was the row within the lysis plate. Sequences were labelled according to the lysis plate and well column within that plate. Thus, within each plate, there were eight sampling units that could contain a maximum of 12 specimens (Fig. 1A).

NJ trees were constructed using MEGA 3 (K2P distance, pairwise deletion) and the resultant tree was input into Conserve (Agapow & Crozier 2008). Here, sampling units (columns within plates) were chronologically added to the total phylogenetic tree, and at each iteration, the total phylogenetic diversity (Faith 1994; Crozier 1997) was calculated $\text{PD} = \sum_{k=1}^{2n-3} d_k$, where the total phylogenetic diversity is given by the total branch length spanned by the members of the queried subset, and d_k = the length of branch k in the tree. For the wasp tree, an ant sequence (HMANT002–06, CHU056–06) was used as an outgroup, and for the ant tree, no outgroup was used (Fig. 1B–D).

This iterative and additive process was used to analyse the barcode PD for all specimens collected from 2006. To this total 2006 PD, we then added the initial specimens (five plates, 433 individuals) analysed from the 2007 collection (Fig. 2) as a preliminary investigation of whether the patterns evident after a year of sampling were consistent upon the addition of one more year's sampling.

In addition to the DBAC analysis of barcode PD, we analysed the initial three wasp plates compiled from malaise traps for the diversity represented within the D2 region from 28S and compared these to the total taxonomic diversity estimated using morphological keys, and the barcode DBAC (Fig. 3).

Total diversity was determined using standard taxonomic keys for subfamily and genus (where available) and using morphospecies determined through comparison to the (CNC) specimens from Canada's north and Manitoba. Richness was also determined using 1.6% and 2% thresholds where members greater than this divergence were counted as separate, and within that divergence were counted as similar. Ninety-five per cent confidence intervals on species richness estimates [expected number of species $\tau(h) = \sum_{i=1}^s [1 - (1 - \phi_i)^h]$] from an S-by-h species–quadrat

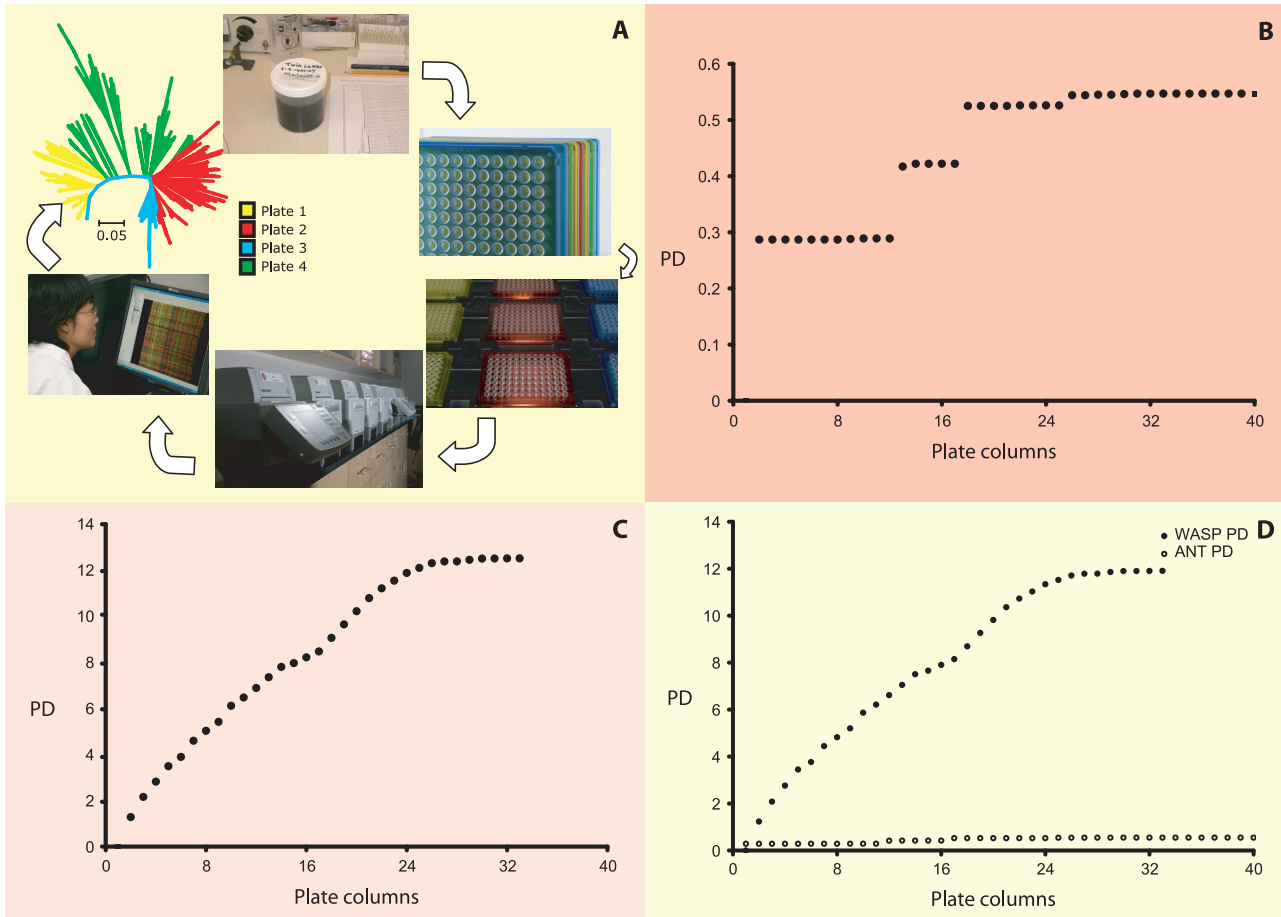


Fig. 1 DNA barcode accumulation curves for Hymenoptera of Churchill, Manitoba. (A) The flow of analysis: Malaise and pitfall trap specimens, randomly sorted into lysis plates (where the column became the unit of analysis), where amplified by polymerase chain reaction and the barcode region was sequenced and then analysed through a tree diagram of distances (NJ, K2P) where the addition of diversity within plates (crudely represented in this simplified example as four colours), and rows within plates, were calculated to create the scaling of diversity represented in the sample. The trends observed for ants (B), parasitoid wasps (C) and the two trends combined to scale (D) are shown.

incidence matrix where the *i*th species has the same probability (ϕ_i) of being present in each sampling unit Mao Tau (Colwell *et al.* 2004). Number of species expected in the pooled samples, given the empirical data per sampling unit (column in lysis plates), were calculated using EstimateS (Colwell 2006) (Figs 3 and 4).

Results

Morphology

Wasps came primarily from four families (in order of abundance), Ichneumonidae, Braconidae, Diapriidae and Cynipidae. See Appendix S1 for a complete listing of specimen taxonomy, collection information and molecular accessions.

Molecular

The first five lysis plates of ants had 94% barcode amplification success (442/470) and 100/188 D2B sequences for the two plates amplified (52.6%; no effort was made during this study to optimize the 28S amplifications conditions).

The initial four lysis plates of wasps had an average 89.6% success rate of barcode amplification (337 CO1 sequences/376 specimens), while the first three plates had 96.8% amplification efficiency for D2B (273 sequences/282 specimens).

The mtDNA sequences from the Hymenoptera, as expected, are heavily AT-biased (ant mtDNA average AT content 72.31, SD = 2.497; wasp mtDNA average AT content 70.00, SD = 7.5). This AT bias leads to third codon saturation. As noted elsewhere (Smith *et al.* 2008), there is

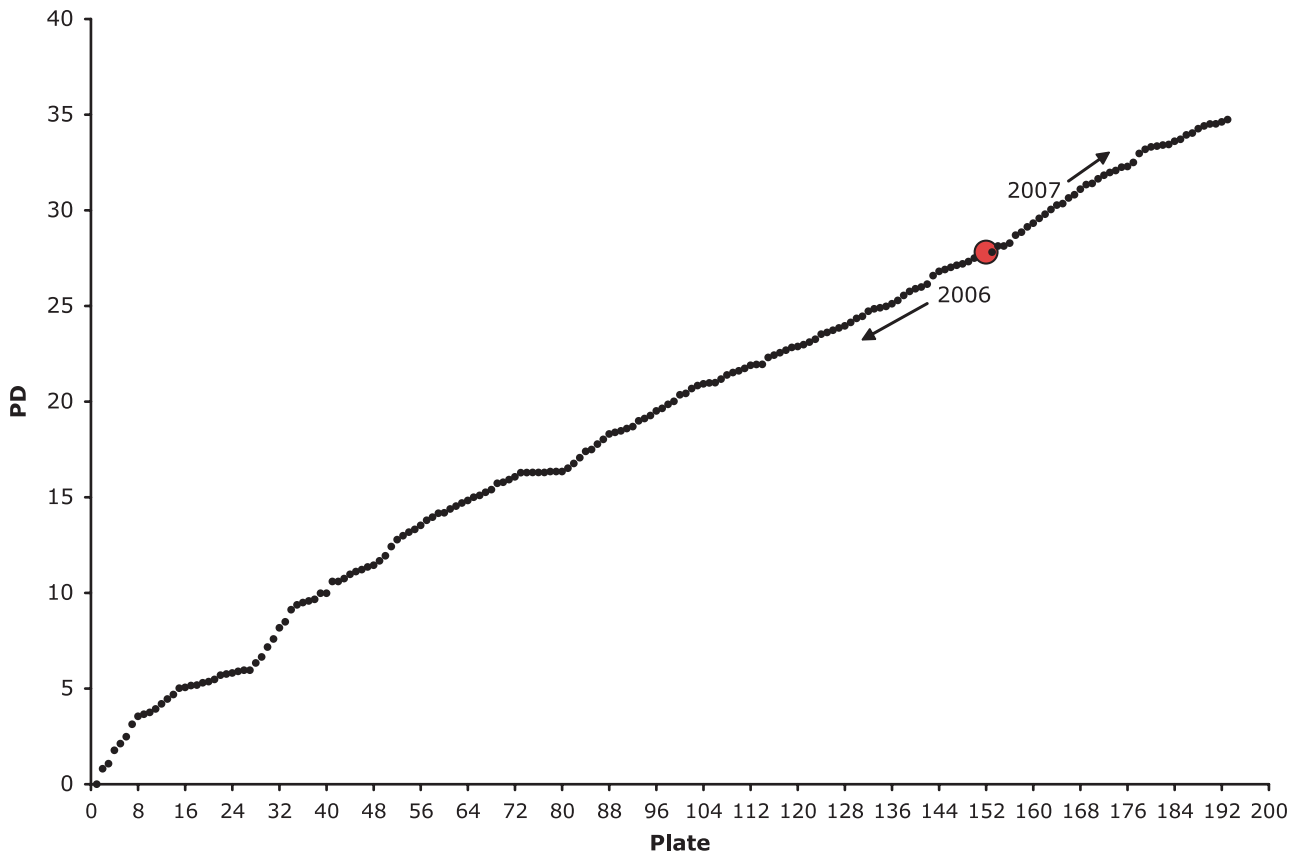


Fig. 2 DNA barcode accumulation curve for parasitoid wasps through the entire 2006 collection combined with the initial samples from 2007. The red point marks the addition of the first samples from 2007. The curve does not approach an asymptote.

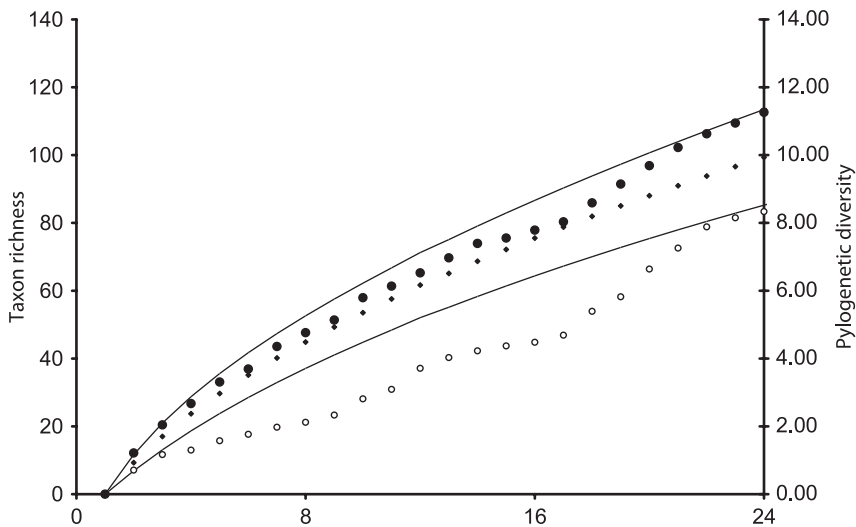


Fig. 3 Taxon richness [Sobs (Mao Tau) with 95% CIs], small diamonds; Barcode PD, filled circles; nuclear PD (28S-D2), empty circles. The three results are different in scale, but not in trend.

a poly-T region within these parasitoid wasps between 7 bp and 14 bp long that can cause difficulty in sequencing.

The barcode PD accumulation curves of ants and wasps are significantly different Kruskal–Wallis test (chi-square approximation = 52.868, d.f. = 1, $P \leq 0.000$; Fig. 1D).

The wasp barcode PD and LSU PD accumulation curves are significantly different in scale Kruskal–Wallis test (chi-square approximation = 7.170, d.f. = 1, $P = 0.007$; Fig. 3).

The wasp barcode 1.6% MOTU and morphological diversity accumulation curves as calculated using EstimateS

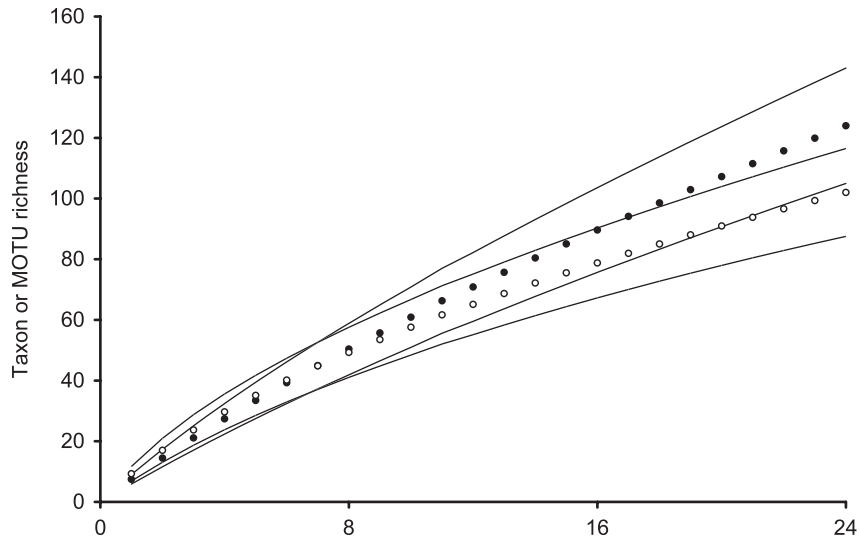


Fig. 4 Comparing barcode MOTU and taxonomic richness using EstimateS (Colwell 2006). Filled circles are barcode MOTU (at 1.6%), and empty circles are taxon richness values for Sobs (Mao Tau) with 95% CIs plotted. The likely eventual divergence of these trends represents a specific manifestation of the taxonomic impediment where names/keys/experts for specimens do not exist. While clearly not statistically significant at $P < 0.05$ (overlap of 95% confidence intervals) the trends do appear to be diverging near the end of the third sampling plate. Qualitatively, this interpretation is supported by the observation that after 2 years of sampling, we estimate there to be a thousand species of parasitoid wasp apparent through barcoding while concomitant estimates, using traditional taxonomy, have named approximately 300 species.

are not significantly different ($F = 0.655$, d.f. = 1, $P = 0.423$; Fig. 4).

For a total list of specimens analysed, CO1 sequence length, collection information, relevant taxonomy at time of submission and GenBank Accessions, please see Appendix S1. The project is ongoing and therefore taxonomic names will be updated as taxonomic decisions are made on the BOLD database within the projects described above.

Discussion

Phylogenetic diversity and DNA barcoding

We sampled two broad taxonomic groups of Hymenoptera, one expected to be more diverse than the other, both from a region where total diversity was not expected to be great. We tested whether an accumulation curve of standardized DNA barcode diversity would facilitate the determination of sampling efficiency at that site, and whether such a barcode accumulation curve was different from estimates generated using standard morphology or nuclear DNA sequences.

We found that the wasp DBAC was significantly different from the accumulation curve generated for ants and the curve generated using a single nuclear marker (LSU-D2). Expected taxon richness (Mao Tau) for 1.6% MOTU and morphology were not significantly different, although the trends evidently diverge near the end of fourth plates' sam-

pling. This final pattern is significant not as it represents different trends of molecular vs. morphologic evolution (though it may), but rather it is significant as it is a direct reflection of the taxonomic impediment restricting the use of invertebrates in large-scale sampling programmes even in areas expected to be of low diversity. DNA barcode accumulation curves for parasitoid wasps in Churchill, Manitoba, are not significantly different from curves generated using a nuclear marker in slope, only in scale, as one would expect based on the slower rates of molecular evolution.

These findings suggest to us that barcode accumulation curves provide a method through which taxonomically unknown and small life may be included in large-scale surveys of biodiversity. Indeed, barcode accumulation curves may be the only manner in which we can harness the predictive power of small invertebrate life for diversity surveys.

We have used barcode sequence data as a direct comparison to species as defined using a more traditional morphological framework. However, one important perspective of our findings is that the utility of a DNA barcoding accumulation curve based on estimates of PD is not dependent (can in fact be separate) from assignment of sequences to species. While the primary goal of DNA barcoding is the eventual creation of a global library of publicly accessible sequences and their associated meta-data, the transparency, objectivity and reproducibility of this process allows it to be used independently from species assignment. This

use of phylogenetic information has proven useful for estimating site diversity and complementarity for plants (and was significantly different from predictions made using species richness; Forest *et al.* 2007), and for ecosystem health (Maherali & Klironomos 2007).

Transparency across space and time

The Churchill PROBE (www.polarbarcoding.org) campaign is not a unique all-taxa endeavour. For instance, the Moorea BIOCODE project (moorea.berkeley.edu/biocode) is designed to genetically document the entirety of the diversity of life within the South Pacific island (Check 2006). Ultimately, the goal of campaigns such as Churchill and Moorea is the same: to catalogue the diversity of life within their regions of interest while facilitating the extraction of the entire ancillary and meta-data associated with each of these sequenced specimens. However, in advance of completing these ongoing surveys, and before placing formalized names on all specimens, a PD approach to barcode diversity offers results that can be replicated, tested and are compared across sites. The act of making long-term biological surveys comparable across space and time would permit a level of standardization that should encourage increased support for conservation and preservation. Of course, ultimately the goal is to accrue much more complete lines of ecological and evolutionary evidence to substantiate our hypotheses erected using barcode data; critically however, barcode data allow us to formulate these hypotheses without waiting for the ultimate taxonomic names to be attached to specimens and collection events and saving both critical time and resources.

Iterative and integrative taxonomy

As is to be expected in any first-pass identification through a morass of samples and diversity (Smith *et al.* 2008), we encountered several cases of apparent (and/or initial) morphological crypsis. Here, a single name was subsequently discovered to harbour multiple, diverse barcode lineages. These lineages were then subjected to further morphological examination (although a complete alpha-taxonomic description has not yet been finished for any of the provisional species described here; see further discussion of this point below) and amplification of an independent nuclear marker (here we used the variable D2 region of 28S). As an example of this process, we present the chronology of the iterative decisions we made for one little known subfamily. Initially, based on an initial morphological examination, we thought there to be a single new provisional species of *Orthocentrus* (Orthocentrinae, Ichneumonidae — dubbed spJFT02). However, we found 10 distinct barcode lineages within this name.

Further testing affirmed that these 10 barcode lineages were directly correlated with 10 divergent nuclear lineages. Upon physical re-examination of the *Orthocentrus* spJFT02 specimens, we (J.F-T.) placed them into 10 new provisional species. Sufficient morphological variation was there to be analysed only once barcoding had called attention to variation worthy of taxonomic consideration. This type of reciprocal illumination is one of the most immediate and powerful applications of barcoding technology.

Conclusions

Our proposition is rooted in the recognition that names are one of the last characteristics that we will know about a region's terrestrial arthropod fauna. Without this characteristic measure of diversity, it is extremely unlikely that we will ever be able to accurately estimate how well we have sampled the invertebrate diversity of an area, especially a community of morphologically character-depauperate species such as parasitoid wasps. With this realization underpinning our efforts to characterize a region's invertebrate fauna, it is clear that we need to accelerate the process of community sampling if we are to understand how well we have sampled the community, let alone include these diverse insect groups in our survey. Phylogenetic diversity analysis of standardized DNA barcodes has provided us with clear, transparent and reproducible estimates of the diversity of these insect families. The estimates suggest that we have not yet neared a sampling asymptote of diversity for parasitoid wasps, but reached that asymptote within approximately 100 ants. In addition, we were able to test whether the results of the PD analysis of mtDNA barcodes were significantly different from estimates generated using morphologically derived names, or a nuclear marker. In each case, the results were different in scale, but not in trend.

In uncovering the diversity of parasitoid life in this northern region, we would prefer to invoke a long list of ecological and host associations that would facilitate the process of recognizing these barcode groups as separate species. However, unlike specimens accrued through extremely special, long-term rearing programmes (e.g. Area de Conservacion de Guanacaste; Smith *et al.* 2008), this kind of intimate information is extremely unlikely to ever be available for these specimens. To survey regions using these taxa, but lacking such a cloud of meta-data, we need an approach that is not as exposed to this impediment of collection and taxonomy. An important associated conclusion is that although it has been recognized that an approach merging barcoding and field studies will greatly refine known arthropod diversity (Leather *et al.* 2008), these conclusions were targeted towards the soil and canopies of tropical rainforest. Our work suggests that benefits accrued from integrating barcoding and field

studies will not be restricted to the tropics and will certainly be evident even in the near-north.

Biologists studying microbial and unicellular life have long recognized the critical role that a phylogenetic approach can bring to surveying life (Crozier *et al.* 1999); we feel that a PD analysis of standardized DNA barcodes can offer the same benefits to eukaryotic life. A direct phylogenetic assessment will permit researchers to avoid the taxonomic impediment that affects many groups of terrestrial arthropods and will provide a transparent and comparable snapshot of the evolutionary potential of a set of sites. In the long run, while the barcode information will be a portion of the integrative taxonomic framework that seeks to name all specimens in a region, in the short to near term, the information can be used to answer questions of biodiversity assessment, conservation and estimation of sampling efficiency without waiting for the necessarily slower process of taxonomy. In parallel, the process provides a wealth of new material for the taxonomic enterprise among taxa that have often been neglected due to their perceived taxonomic opacity. Maximizing preserved evolutionary history, as represented by distance within phylogenetic trees using a standardized DNA barcode, provides a triage tool advantageous to both ecologists and conservationists.

Acknowledgements

We thank Torbjørn Ekrem and Elisabeth Stur (for setting up and maintaining the traps in 2006), Dirk Steinke, Jayme Sones and MD. A. Hannan (for specimen preparation collation, validation and curation), Kate Crosby and Anibal Castillo (for diligent, thorough and thoughtful laboratory assistance), Rick Turner (for photographs), and Alex Borisenko, Robert Hanner, Mehrdad Hajibabaei, Natalia Ivanova, Sujeevan Ratnasingham, Rodolphe Rougerie, Justin Schonfeld, Dirk Steinke and Xin Zhou, (for thoughtful comments on an earlier draft of this manuscript). This research was supported by an NSERC International Polar Year Grant, the Canadian Barcode of Life Research Network from Genome Canada through the Ontario Genomics Institute and other sponsors listed at polarbarcoding.org.

Conflict of interest statement

The authors have no conflict of interest to declare and note that the funders of this research had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

- Agapow P-M, Crozier RH (2008) Conserve IV, www.agapow.net/software/conserv.
- Andelman SJ, Fagan WF (2000) Umbrellas and flagships: efficient conservation surrogates or expensive mistakes? *Proceedings of the National Academy of Sciences, USA*, **97**, 5954–5959.
- Bolton B (1994) *Identification Guide to the Ant Genera of the World*. Harvard University Press, Cambridge, Massachusetts.
- Check E (2006) Treasure island: pinning down a model ecosystem. *Nature*, **439**, 378–379.
- Colwell RK (2006) EstimateS: statistical estimation of species richness and shared species from samples. Persistent URL: purl.oclc.org/estimates.
- Colwell RK, Mao CX, Chang J (2004) Interpolating, extrapolating and comparing incidence-based species accumulation curves. *Ecology*, **85**, 2717–2727.
- Crozier RH (1997) Preserving the information content of species: genetic diversity, phylogeny and conservation worth. *Annual Review of Ecology and Systematics*, **28**, 243–268.
- Crozier RH, Agapow PM, Pedersen K (1999) Towards complete biodiversity assessment: an evaluation of the subterranean bacterial communities in the Oklo region of the sole surviving natural nuclear reactor. *FEMS Microbial Ecology*, **28**, 325–334.
- Faith DP (1994) Phylogenetic pattern and the quantification of organismal biodiversity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **345**, 45–58.
- Fisher BL, Smith MA (2008) A revision of Malagasy species of *Anochetus* Mayr and *Odontomachus* Latreille (Hymenoptera: Formicidae). *PLoS ONE*, **3**, e1787.
- Fisher BL, Malsch AKF, Gadagkar R *et al.* (2000) Applying the ALL protocol: selected case studies. In: *Ants: Standard Methods for Measuring and Monitoring Biodiversity* (eds Agosti D, Majer JD, Tennant de Alonso L, Schultz T), pp. 207–214. Smithsonian Institution, Washington D.C.
- Forest F, Grenyer R, Rouget M *et al.* (2007) Preserving the evolutionary potential of floras in biodiversity hotspots. *Nature*, **445**, 757–760.
- Godfray HCJ, Knapp S (2004) Introduction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**.
- Gotelli NJ (2004) A taxonomic wish-list for community ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**, 585–597.
- Gottelli N, Colwell RK (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, **4**, 379–391.
- 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, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Science*, **270**, 313–321.
- Hoffmeister TS, Vet LEM, Biere A, Holsinger K, Filser J (2005) Ecological and evolutionary consequences of biological invasion and habitat fragmentation. *Ecosystems*, **8**, 657–667.
- 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 (2004) Now is the time. *Philosophical Transactions of the Royal Society B: Biological Science*, **359**, 731–732.
- 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.
- Kumar SK, Tamura K, Nei M (2004) MEGA 3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics*, **5**, 150–163.
- Leather SR, Basset Y, Hawkins BA (2008) Insect conservation: finding the way forward. *Insect Conservation and Diversity*, **1**, 67–69.

- Maddison WP (1997) Gene trees in species trees. *Systematic Biology*, **46**, 523–536.
- Maherali H, Klironomos JN (2007) Influence of phylogeny on fungal community assembly and ecosystem functioning. *Science*, **316**, 1746–1748.
- Ratnasingham S, Hebert PDN (2007) BOLD: the Barcode of Life Data System. *Molecular Ecology Notes*, **7**, 355–364.
- Renner SC, Häuser CL (2007) How many taxonomists are there?
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenies. *Molecular Biology and Evolution*, **4**, 406–425.
- Smith MA, Fisher BL, Hebert PDN (2005) DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Philosophical Transactions of the Royal Society B: Biological Science*, **360**, 1825–1834.
- 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, USA*, **105**, 12359–12364.
- Underwood EC, Fisher BL (2006) The role of ants in conservation monitoring: If, when and how. *Biological Conservation*, **132**, 166–182.
- Vitousek PM, Mooney HA, Lubchenco J, Melillo J (1997) Human domination of Earth's ecosystem. *Science*, **277**, 494–499.

Supporting Information

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

Appendix S1 Information regarding specimens, CO1 and 28S accession data, primers used for sequencing and amplification, EstimateS matrices and exemplar NJ trees used to parameterize Conserve.

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