KEYNOTE ARTICLE

Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity

DANIEL H. JANZEN[1], WINNIE HALLWACHS[1], PATRICK BLANDIN[2], JOHN M. BURNS[3], JEAN-MARIE CADIOU[+4], ISIDRO CHACON[5], TANYA DAPKEY[1], ANDREW R. DEANS[6], MARC E. EPSTEIN[7], BERNARDO ESPINOZA[5], JOHN G. FRANCLEMONT[+8], WILLIAM A. HABER[9], MEHRDAD HAJIBABAEI[10], JASON P. W. HALL[3], PAUL D. N. HEBERT[11], IAN D. GAULD[+12], DONALD J. HARVEY[3], AXEL HAUSMANN[13], IAN J. KITCHING[14], DON LAFONTAINE[15], JEAN-FRANÇOIS LANDRY[15], CLAUDE LEMAIRE[+16], JACQUELINE Y. MILLER[17], JAMES S. MILLER[18], LEE MILLER[+19], SCOTT E. MILLER[3], JOSE MONTERO[5], EUGENE MUNROE[+20], SUZANNE RAB GREEN[21], SUJEEVAN RATNASINGHAM[10], JOHN E. RAWLINS[22], ROBERT K. ROBBINS[3], JOSEPHINE J. RODRIGUEZ[23], RODOLPHE ROUGERIE[10], MICHAEL J. SHARKEY[24], M. ALEX SMITH[10], M. ALMA SOLIS[25], J. BOLLING SULLIVAN[26], PAUL THIAUCOURT[27], DAVID B. WAHL[28], SUSAN J. WELLER[29], JAMES B. WHITFIELD[29], KEITH R. WILLMOTT[17], D. MONTY WOOD[15], NORMAN E. WOODLEY[25] and JOHN J. WILSON[10]

[1]*Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA,* [2]*Muséum National d'Histoire Naturelle, CP 135, 57 rue Cuvier, 75005 Paris, France,* [3]*Department of Entomology, National Museum of Natural History, Smithsonian Institution, PO Box 37012, MRC 009, Washington, DC 20013-7012, USA,* [4]*Belgium,* [5]*Instituto Nacional de Biodiversidad, Apdo 22-3100, Santo Domingo de Heredia, Costa Rica,* [6]*Department of Entomology, North Carolina State University, Campus Box 7613, 2301 Gardner Hall, Raleigh, NC 27695-7613, USA,* [7]*Plant Pest Diagnostic Branch, California Department of Food & Agriculture, 3294 Meadowview Road, Sacramento, CA 95832-1448, USA,* [8]*NY, USA,* [9]*Missouri Botanical Garden, PO Box 299, St Louis, MO 63166, USA,* [10]*Canadian Centre for DNA Barcoding, Biodiversity Institute of Ontario, University of Guelph, 50 Stone Road, Guelph, ON, Canada N1G 2W1,* [11]*Department of Integrative Biology, University of Guelph, 50 Stone Road, Guelph, ON, Canada N1G 2W1,* [12]*UK,* [13]*Zoological State Collection Munich, Sektion Lepidoptera, Munchhausenstrasse 21, D-81247 Munchen, Germany,* [14]*Department of Entomology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK,* [15]*Canadian National Collection of Insects, Agriculture and Agri-Food Canada, 960 Carling Avenue, Ottawa, ON, Canada K1A 0C6,* [16]*France,* [17]*McGuire Center for Lepidoptera & Biodiversity, Florida Museum of Natural History, University of Florida, Cultural Plaza, PO Box 112710, Gainesville, FL 32611-2710, USA,* [18]*Department of Entomology,* [19]*FL, USA,* [20]*Canada,* [21]*Division of Invertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024-5192 USA,* [22]*Section of Invertebrate Zoology, Carnegie Museum of Natural History, 4400 Forbes Avenue, Pittsburgh, PA 15213-4080, USA,* [23]*Entomology, 320 Morrill Hall, 505 S. Goodwin Avenue, University of Illinois, Urbana, IL 61801, USA,* [24]*Department of Entomology, S-225 Agricultural Science Building-North, University of Kentucky, Lexington, KY 40546-0091, USA,* [25]*SEL, USDA, Smithsonian Institution, PO Box 37012, Washington, DC 20013-7012, USA,* [26]*200 Craven Street, Beaufort, NC 28516, USA,* [27]*131 Avenue de Versailles, 75016 Paris, France,* [28]*American Entomological Institute, 3005 SW 56th Avenue, Gainesville, FL 32608-5047, USA,* [29]*Bell Museum of Natural History, 10 Church Street, University of Minnesota, Minneapolis, MN 55455, USA*

Abstract

Inventory of the caterpillars, their food plants and parasitoids began in 1978 for today's Area de Conservacion Guanacaste (ACG), in northwestern Costa Rica. This complex mosaic of 120 000 ha of conserved and regenerating dry, cloud and rain forest over 0–2000 m elevation contains at least 10 000 species of non-leaf-mining caterpillars used by more than 5000 species of parasitoids. Several hundred thousand specimens of ACG-reared adult Lepidoptera and parasitoids have been intensively and extensively studied morphologically by many taxonomists, including most of the co-authors. DNA barcoding — the use of a standardized short mitochondrial DNA sequence to identify specimens and flush out undisclosed species — was added to the taxonomic identification process in 2003. Barcoding has been

Correspondence: Daniel H. Janzen, Fax: 215-898-8780, E-mail: djanzen@sas.upenn.edu
+, deceased

found to be extremely accurate during the identification of about 100 000 specimens of about 3500 morphologically defined species of adult moths, butterflies, tachinid flies, and parasitoid wasps. Less than 1% of the species have such similar barcodes that a molecularly based taxonomic identification is impossible. No specimen with a full barcode was misidentified when its barcode was compared with the barcode library. Also as expected from early trials, barcoding a series from all morphologically defined species, and correlating the morphological, ecological and barcode traits, has revealed many hundreds of overlooked presumptive species. Many but not all of these cryptic species can now be distinguished by subtle morphological and/or ecological traits previously ascribed to 'variation' or thought to be insignificant for species-level recognition. Adding DNA barcoding to the inventory has substantially improved the quality and depth of the inventory, and greatly multiplied the number of situations requiring further taxonomic work for resolution.

The problem

Humanity has had 6 million-plus years of learning, knowing, using and forgetting biodiversity. You have today a frog, leaf, hair, cockroach or fish fillet in hand, eye or mouth, and you want to know what collectively we know about that species, which may well be quite a lot. Any 10-year-old can tell you how. Just put its name into Google and click. Yes, well, where is the port on your computer, iPhone or next-generation gadget where you put the bit of it to get the species name? In short, humanity has a huge amount of biodiversity information, but just when you need to access it at your particular moment, you lack the access tag to type into the search engine. Yes, there is a taxonomic specialist, or taxonomy-tagged literature that may be able to provide the tag, but it is certain that 99.99% of the time that information resource will not be at your side. And even if it is, it takes a lifetime of vocabulary to use it. For example, the diagnosis for the Campopleginae, a 1000-species-rich subfamily of parasitic wasps is: *Propodeum short, rugose-reticulate, centrally strongly longitudinally excavate; propodeal apophyses from weak to strongly developed, often with a secondary apophysis present above insertion of hind coxa* (I.D.G.). The first triplet in the key to 105 species of Mexican oak trees: *Fruits maturing the first season ... Fruits maturing the second season ... Fruits often maturing the second season ...* (Standley 1922).

An ongoing inventory of the caterpillars and their parasitoids in a large complex Costa Rican conservation area (http://janzen.sas.upenn.edu) has this problem as much as does the Minnesota schoolchild poking at an anthill while waiting for the school bus, as does the Ugandan farmer watching small green beetles defoliate his bean plants, as does the Miami port inspector peering at the aphids in a container of broccoli.

A contribution to a solution

DNA barcoding — using a standardized short sequence of DNA as a species-level key character (Hebert *et al*. 2003; Stoeckle & Hebert 2008; http://www.barcoding.si.edu/; http://phe.rockefeller.edu/barcode/blog; Hajibabaei *et al*. 2007; CBOL 2008; Kress & Erickson 2008) — is engineering aimed at providing a pragmatic link between what we know about species (e.g. EOL at http://www.eol.org/) and what you want to know now about a specimen, any time, anywhere, cheaply (Janzen 2004b; Janzen *et al*. 2005; Wolf 2008). In other words, identify it. Like any telephone call, sometimes you learn more than you anticipated when you dial the number. And there may even occasionally be a wrong number. Yes, the social structure of humanity is based on the selective capture, release and horse-trading of information, and thus, there will be a cost to the user and a gain to the provider. But if done right, the cost will be no greater than is the cost of seeking a telephone number, a levelling of the biodiversity playing field.

DNA barcoding will offer personal real-time intimacy with wild biodiversity and what humanity knows and can find out about it. On-site real-time specimen identification has the potential for a scale of bioliteracy that is orders of magnitude finer-grained than what is possible with our traditional ways of knowing what wild species is what. And it will offer it to everyone anywhere at any time.

DNA barcoding contains an interesting semantic conundrum. Strictly speaking, to barcode a can of tuna is to put a barcode on it. DNA barcoding does not put a barcode on the specimen. Rather, it reads one that is already there. However, language being what it is, we are stuck with current usage — we barcode the ant in hand.

Will DNA barcoding replace morphology-based taxonomy as an identification tool and as a species-discovery process?

This is not a useful phrasing of a question. Will identification and species discovery be substantially improved and democratized by global application and development of DNA barcoding along with all the traditional morphological ways we recognize species? Yes. It is an easy prognosis that barcoding will rapidly replace or significantly augment morphology-based identification in many sectors and endeavours, will emerge more and more as a species-discovery tool, and will become routine for the scientific community (e.g. Kankare *et al.* 2005; Kuhlmann *et al.* 2007; Ficetola *et al.* 2008; Jurado-Rivera *et al.* 2009; King *et al.* 2008; Valentini *et al.* 2009; White *et al.* 2008; Aliabadian *et al.* 2009; Ferri *et al.* 2009) as well as the lay community. The only question is to what degree this will be an explicit, engineered and rapid change, such as the change from floppy disks on a desktop OS to wireless cloud computing, and to what degree it will emerge helter-skelter through common usage and agenda-specific public demand. Doctors — be they for humans, other animals or crop plants — are desperate for the ability to instantly know what pathogen or pathogen carrier confronts them. But the pathogen-specific barcoder and barcode library they invent will not do for the tens of millions of species on Earth.

ACG as guinea pig

Any engineering project needs mock-ups, design planning, test beds, and test pilots. In short, an evolvable guinea pig is needed to get to a usable product(s). Here we offer a sketchy report and X-ray of these past 5 years of using a massive and complex tropical biodiversity inventory of the 160 000 + hectare Area de Conservacion Guanacaste (ACG; Fig. 1) in northwestern Costa Rica (http://janzen.sas.upenn.edu) as a guinea pig for DNA barcoding. This use of the ACG came about because:

1 The Alfred P. Sloan Foundation funded two exploratory DNA barcoding workshops in March–September 2003, attended by P.D.N.H., D.H.J. and W.H., where it became evident that the personal, cheap and reusable DNA barcoder no longer needed to be thought of as the science fiction it was in the 1950s (Wolf 2008).

2 In 2003, the nascent Biodiversity Institute of Ontario (BIO) enthusiastically and successfully DNA barcoded (Hebert *et al.* 2003) hundreds of reared adults of very similar species of skipper butterflies that were previously thought to belong to the single species *Astraptes fulgerator*, so as to help clarify their species-level taxonomy as already partly visualized by their morphology and ecology (Hebert *et al.* 2004).

3 The intensive and extensive inventory of Lepidoptera and their parasitoids of Area de Conservacion Guanacaste (ACG) in northwestern Costa Rica by traditional means (find, rear, photograph, identify, voucher, database, put on web — D.H.J., W.H.) had already been in motion for 25 years, strongly complemented by the countrywide insect inventory by INBio.

4 B.I.O., and P.D.N.H., M.H., and M. Alex Smith enthusiastically accepted the offer by the ACG inventory to be a DNA barcoding guinea pig.

5 NSF and the Wege Foundation, in consort with Anne Lambert, J.D. and Nancy Turner, the Cox Family Trust, ACG, and INBio, along with substantial sweat equity by the taxasphere (the global community of taxonomic experts and all that they know) and ACG administration, agreed to cover the additional financial strain of adding intensive and thorough DNA barcoding to the specimen-as-resource side of the inventory, while BIO, the Moore Foundation, Sloan Foundation, Smithsonian Institution, NSF Biotic Surveys and Inventories, and many taxonomists and their institutions and reporters/bloggers (e.g. http://www.barcoding.si.edu/) have covered the parallel costs of the urban-laboratory and promotional side of beginning the DNA barcoding of the ACG Lepidoptera and parasitoid inventory.

Brief description of ACG Lepidoptera and parasitoid inventory

The project began in 1978 to provide the taxonomic platform for the ecological study of the caterpillars as major folivores in a small dry forest sector and has evolved into the inventory of today's much larger ACG (Janzen 2000; Janzen & Hallwachs 2008). By 1990, the inventory had transformed into the mission of finding 'all' (estimated) 10 000 species of non-leaf-mining wild caterpillars of ACG moths and butterflies, connecting these caterpillars to their adults, documenting at least one species of food plant for each species of caterpillar, rearing as many of their (estimated) 5000 parasitoid species as feasible, and putting all of this information into the public domain through a web site (http://janzen.sas.upenn.edu) and publications (e.g. Burns *et al.* 2008; Smith *et al.* 2008). This inventory reluctantly but explicitly refuses to be diverted into the plethora of ecological, evolutionary, behavioural, and morphological puzzles and pathways that are daily revealed by the inventory. The daily field work consists of finding wild caterpillars, rearing them, databasing, and preparing adults as museum-ready specimens. This is carried out by a team of (today) 30 Costa Rican resident parataxonomists (Janzen 2004a) working through D.H.J. and W.H. as a clearing house between them and the taxasphere (see methodology http://janzen.sas.upenn.edu).

This kind of ultra-fine-scale examination of complex tropical ecosystems has required an enormous amount of sweat-equity support from the taxasphere — the collected global array of taxonomists, collections and their knowledge
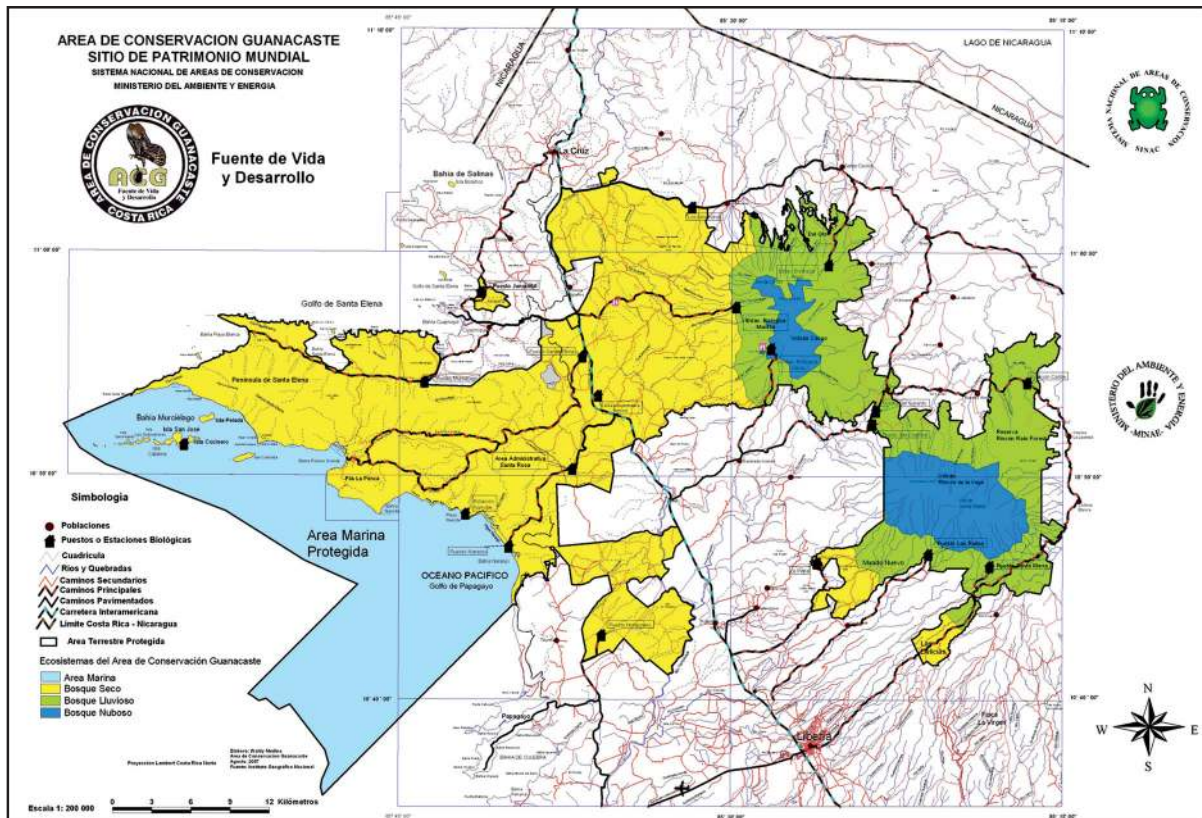
Fig. 1 Top. Area de Conservacion Guanacaste, northwestern Costa Rica with approximate locations of four major ecosystems: light blue, marine; yellow, dry forest; green, rain forest; dark blue, cloud forest. The coloured area is 77 km wide, extending from sea level to 2000 m above sea level. Bottom. Clockwise, dry forest in the dry season, cloud forest mid-day, and rain forest mid-day on an exceptionally sunny day.

in mind and print — (Janzen 1993). Some entomologically inclined members of the taxasphere are co-authors of this report. Reciprocally and simultaneously, the ACG inventory has endeavoured to support the taxasphere with information, specimens, ideas and cheerleading, and shared the taxasphere's frustration with the many small roadblocks to quick and accurate identification, and description, of the specimens that flow from the inventory. Unless this identification process is in full flower, the inventory is of minimal value to both the science and the lay communities.

By 2003, when the administration discovered the DNA barcoding approach, they realized how the approach of the ACG inventory (Hebert *et al*. 2003) could be applied to biodiversity inventories and help translate the results rapidly to society at large. By that time, the inventory already encompassed about 2500 species of morphologically characterized moths and butterflies, and their parasitoid flies and wasps. There were about 210 000 individual rearing records and 40 000 images on the project website (Janzen & Hallwachs 2008), with about 65 000 pinned, frozen or alcohol specimens deposited in museums. Approximately 70% of these rearing records are identified to species level. The remainder were, and are still, in some state of 'being identified' or 'being described' as new, by a global array of about 50 members of the taxasphere. Today, 2008, the total of rearing records is 400 000+ with other variables proportionally increased.

Mechanics of adding DNA barcoding to the ongoing inventory

A. Engineering consequences of adding DNA barcoding to an inventory

1. Day-to-day processing before DNA barcoding. Adding DNA barcoding to an ongoing inventory requires no change in the specimen-by-specimen daily mechanics of the basic inventory process in the field. A parataxonomist finds a free-living caterpillar, places it in a 4 L clear plastic bag with a branchlet of the plant on which it was found, writes date/location/name data on the bag, and takes it back to the rearing barn. If the caterpillar feeds, it remains in the inventory (prepupae and pupae are also captured). It is given a unique voucher alphanumeric code (e.g. 08-SRNP-34256; 08 is the year, SRNP are the project call letters, and 34256 is assigned that caterpillar that year) and each of the 10 rearing barns is assigned a unique set of these numbers at the beginning of the year. A flat file record — in effect an event pedigree — of finding the caterpillar is generated in FileMaker Pro at the rearing barn. The voucher code is actually for the event of finding the caterpillar and then all subsequent things that happen to it (including eventual barcoding of any adult), and in practice the voucher code of the event is also used to tag the caterpillar, body parts,

parasitoids, images and any other associated collateral. Particles of this event (images, parasitoids) are later assigned additional personal unique voucher alphanumeric codes (e.g. DHJPAR0006578 for a parasitoid, e.g. 03-SRNP-5555-DHJ376002.jpg for an image) as they become 'separated' from the event. Information as to whether the specimen or other event parts are DNA barcoded is added to the event record much later in the accumulation of this ongoing dynamic pedigree, at the time it occurs.

If the caterpillar produces an eventual adult, or parasitoid(s), that specimen is killed by freezing (moths, butterflies, some flies and wasps) or cyanide gas (some rearing stations where freezers are not accessible), and/or placing in 95% ethanol (small wasps only). If there is no usable specimen, the record (event pedigree) is nevertheless retained as collateral documentation for the images and the event. The frozen or alcohol-preserved specimen is then brought (in large multiple-month batches) to the central 'clearing house' in the Area Administrativa of Sector Santa Rosa of ACG. The field identification (to whatever taxonomic level) is reviewed record by record. D.H.J. then decides whether to discard the specimen or preserve (pin/dry/alcohol) it as a specimen for identification or other research by the taxasphere, and/or as a permanent voucher for that particular inventory record (whether barcoded or otherwise). In pre-barcoding days, specimens simply flowed to collaborating taxonomists. This was followed by years of review and study to get the specimens identified. Specimens were then deposited in museums as permanent vouchers for the inventory records and for cross-referencing future identifications. New species descriptions were published as appropriate (e.g. Gauld 1985; Lemaire 1988; Burns 1996; Miller *et al*. 1997; Schauff & Janzen 2001; Janzen *et al*. 2003; Solis *et al*. 2003; Gauld & Janzen 2004; Burns & Janzen 2005).

The caterpillars and parasitoid cocoons are routinely photographed in the field for the project web site (http://janzen.sas.upenn.edu) and photographed as historical surrogate vouchers. However, prior to barcoding only a very few select 'representative' adults were photographed, largely as taxonomic aids and to verify identifications.

2. Introduction of DNA barcoding into the inventory. When DNA barcoding was introduced into this inventory flow chart in mid 2003, the specimen passing through the taxonomic clearing-house (at the University of Pennsylvania) had one or two legs removed to be couriered to the Biodiversity Institute of Ontario (BIO) for barcoding. This voucher specimen is given a yellow label bearing a terse verse composed by J.M.B. that reads 'LEGS AWAY FOR DNA', is photographed (Lepidoptera only), and additional collateral information collected (sex, wingspan, tracking codes, collateral voucher codes). In the case of ACG inventory vouchers collected before 2004, leg removal also often occurred at the office of a collaborating taxonomist.

This DNA barcoding processing has additional cost and requires meticulous iterative attention to detail, but presents no intellectual problems other than arbitrarily deciding (for the most part) not to attempt to barcode vouchers that were collected before 1990 because of their very low chance of yielding a full DNA sequence from the COI barcode region. We also expect that fresh specimens of that species will eventually be reared again by the inventory.

Each specimen to be barcoded must be labelled with a unique voucher code. By convention, the adult reared from the caterpillar retains the event voucher code (yy-SRNP-xxxx) but each of the individual parasites reared from that caterpillar requires a second unique voucher alphanumeric (DHJPARxxxxxxx). For example, the barcode for the tachinid fly *Belvosia* Woodley05 cannot go to BOLD or GenBank, or to the world, under the same 'unique' event voucher code as is borne by its host caterpillar *Rothschildia lebeau*. The caterpillar does, however, have its own record and may even have preserved body parts and be stored under its own unique SRNP event-based voucher code. Furthermore, if, for example, three barcode-able flies of *B.* Woodley05 are reared from a single *R. lebeau* caterpillar, each requires its own unique voucher code. These parasitoid event-based records required the construction of a parallel database (DB) with many, but not all, of the fields in common with the core DB for the caterpillar collecting events.

A time-dependent complexity is also introduced to the assembly line flow of information from field to website. Each of the 10 rearing barns has its own master copy of its core database for the current year, a database in which the event records are being dynamically and sporadically updated by the parataxonomists (eclosion dates, conversion of interim names to scientific names, observations of natural history, GIS data, etc.). Copies of these unique DBs are sent to Santa Rosa on occasion, but the gatekeepers for these within-year unique DBs are the parataxonomists themselves. In December, these DBs are collected, pooled, spell-checked and logic-checked, and then added to the main project DB in the following February.

The addition of DNA barcoding has increased the need for these DBs to be rapidly reconciled. Prior to DNA barcoding, the taxonomic process (and demand for collateral data) was so slow that uses by the taxasphere generally lagged 1–4 years behind the generation of event data and specimens, and the few cases of instant within-the-year demand and feedback could be handled case-by-case. However, the assembly-line process for 20 000 to 40 000 newly reared Lepidoptera and parasitoids to be barcoded per year at BIO results in new barcodes within 12 months of the caterpillar collection event. BOLD (and subsequently GenBank) require immediate collateral data so as to gain the full information value of the barcodes being generated. Currently, the inventory maintains interim databases for this within-year information, and then integrates these databases (and their updating) with the annual databases at the end of the year as well. Consequently, the database tracking system of BOLD has to robustly tolerate empty fields that are filled later. For example, GIS coordinates for a 2008 record may not become available until March of the following year. The BOLD DB also has to be robust for numerous upgrades in collateral data as the years pass. For example, a species name field may experience as many as six-plus changes as the specimen moves from the forest to the clearing house in Santa Rosa, and then to the clearing house in Philadelphia, and then to the taxonomist to the interim deposit in the University of Pennsylvania, and then to a final deposit in a museum. And even there, names may again change as the barcoding data become integrated with morphology-based data and other museum specimen data.

Adding DNA barcoding to the inventory creates a substantial engineering and budget problem for the final deposition sites of the very large number of voucher specimens. Prior to DNA barcoding, a voucher specimen deposited in a museum was viewed as having largely the classical value of being a voucher for its presence at a collecting site/time, and being an item for the morphological study of what is believed to be its species. The pre-barcoding ACG inventory only gently asked museums that deposited inventory vouchers be treated with the deference due to their being somewhat better-documented than are many museum specimens, and being more likely to be individually re-examined in the future. And, in the commonplace event that the museum collection already contained many specimens of 'that species', further deposition of inventory specimens was often rejected by a museum (Costa Rica has been heavily collected for more than a century). When barcoding was added to the inventory, the number of 'deposit-worthy' specimens greatly increased, care accorded to them in the museum deposit increased, and the interest of the taxasphere in absorbing them increased. All this is because a specimen is now the voucher for the barcode (and of course, for potential future data mining of the DNA extract from which the barcode was obtained as well) and because of the desirability of further morphological study when the barcode signals an actual or possible cryptic species. For example, were the ACG inventory to be non-barcode based, there would now be about 5000 reared specimens for 400+ species of adult ACG Hesperiidae (skipper butterflies) deposited in the National Museum of Natural History at the Smithsonian Institution, but currently there are 13 000+ specimens of 500+ species and the volume is still steadily increasing at the rate of several thousand barcoded specimens per year. However, to date none of these and following considerations have stopped or hindered the permanent deposition of all barcode vouchers in a major public museum.

All of the museum-held pre-barcoding ACG voucher specimens abruptly became potential resources for retroactive

barcoding to explore or increase sample sizes (or to add ecological and microgeographical breadth) for the barcode patterns emerging from the newly collected and routinely barcoded specimens. Many specimens of infrequently encountered species were accumulated by the inventory during its early years (ACG species collection frequencies have changed continually and episodically during the course of the inventory). Barcoding these specimens already deposited in museums introduced two engineering complications (aside from the socio-legal question of who is the owner of the specimens and its collateral information).

First, in many cases, each old voucher specimen required additional handling (photographing, de-legging, measuring, labelling) somewhere other than at the taxonomic clearing house at the University of Pennsylvania (UPenn). This problem was mostly solved by going to the specimens and conducting intensive several-day processing sessions, or by bringing the specimens *en masse* temporarily back to the UPenn clearing house for processing. The inventory is still suffering mildly on this front, although most of the older voucher specimens have now been barcode-captured (or attempted). However, the situation still remains logistically awkward because while the project does later receive the results of barcoding at the UPenn clearing house, the specimens sometimes then need re-examination but are at a distant museum. This in turn places an extra burden on the participating taxonomist or curator if the inventory staff cannot easily get to that distant museum, and/or adds months to years of delay between barcoding a specimen and collating the results with the morphology and/or curation of the specimen.

Second, since the ACG inventory began in 1978, as many as half of the older voucher specimens did not easily yield long sequences (barcodes of greater than 500 base pairs in length) or even short DNA sequences (so-called mini-barcodes, Hajibabaei *et al*. 2006). Many of the species represented by those older specimens have required recollection so as to obtain full-length barcodes, which the inventory has attempted even if there seemed to be no taxonomic confusion in their morphology-based identification. For example, the common noctuids *Azeta rhodogaster* and *Hypocala andremona* were among the first ACG noctuid moths to be reared and thus inventoried, and needed to be recollected to get their barcodes 20 years later (at which time it was discovered that *H. andremona* may well be two species, Appendix SI).

3. Consequences of barcoding feedback to the inventory. Feedback from adding DNA barcoding to an ongoing inventory creates three major engineering changes in the field operation and the handling of museum-held voucher specimens, and a fourth might be necessary for many other kinds of insect inventories.

(i) Increased sample sizes on-site during the inventory

The most glaring outcome of barcoding for the field-based part of the inventory is that a sample size of 'a few' is no longer satisfactory — even when they are good specimens, a well known species, nicely reared, and well-photographed. Even if full-length barcodes are exactly to closely mon-omorphic for 2–5 (for example) individuals (a commonplace event, see NJ trees in Appendices), there is a definite possibility that if the sample size — even at one site — is increased to 10–20 unrelated individuals, an apparent barcode polymorphism may be encountered that is simply that (or a contaminant or a pseudogene), or (very frequently in the ACG inventory) a harbinger of an overlooked species. Such an apparent polymorphism then begs for yet more sampling. For example, the low density tachinid flies *Patelloa* xanthuraDHJ02 and *Patelloa* xanthuraDHJ06 only became visible subsequent to barcoding more than 300 flies of the generalist *Patelloa xanthura* (now baptized *Patelloa* xanthuraDHJ01) (Smith *et al*. 2007). While sample-size analysis is beyond the scope of this sketchy report of ACG caterpillars and their first-pass barcoding, it is clear that even 'large' samples of 10–20 individuals will not reveal 'all' cases of sympatric or parapatric cryptic species, no matter how the sampling is structured. Just as disclosing species by morphology-based, or ecology-based, inventory is peeling back the first layer of the biodiversity onion, barcoding the same species begins to peel back a second layer. And, there may well be yet more species-level layers underneath, to say nothing of biologically interesting population level heterogeneity in barcodes.

This increase has been accompanied by much regret expressed for having discarded specimens in previous years because 'we had enough of them, and no museum wanted more of them, and/or the inventory did not have sufficient funding to receive/prepare them as museum specimens'. The consequence is that the analysis of feedback from the barcoding dictates both much more collecting of caterpillars and parasitoids of species that were previously thought to have been inventoried 'enough' for the purposes of the project, and the retention in museums of many specimens that would have been discarded pre-barcoding.

(ii) Increased retroactive sample sizes from older vouchers

When feedback from barcoding of either fresh or older specimens reveals sequence variation suggestive of the presence of cryptic species, the inventory is confronted with the logistically quick option of attempting to barcode all the old voucher specimens or the slower intensified collection of that species in the future. The latter is expensive in both time and budget, yet yields much better sequences more cheaply per specimen. There is no simple algorithm to solve this conundrum, but on numerous

occasions attempts to sequence series of older museum specimens have yielded little or no success. Additionally, such retroactive increase in sample size leads to more frequent cases of contamination because old specimens with degraded DNA occasionally yield a sequence that is derived from a contaminating scale from a fresh specimen, a scale with high quality DNA that is robustly captured by the primer (this is less frequent with parasitic wasps and flies, apparently because their hairs do not break free as readily as do Lepidoptera scales). All of these problems are exacerbated when the quest for increased sample size leads to the attempt to sequence yet other and older conspecific museum specimens derived from other studies and geographies. The situation leads to the emotion that it would be very appropriate to re-collect the world with modern technology for extraction and the analysis at hand. For example, it has led to the conclusion by D.H.J. and W.H. in 2006 to re-collect the entire fauna of adult Lepidoptera of ACG by traditional means. This ongoing project (the 2-year-old BioLep project) runs in parallel to the caterpillar inventory, but space does not permit its description here.

(iii) Increased geographical and ecological coverage

While all ACG inventory specimens are from the 'same place' (about 80×30 km), that 'same' is in fact a mosaic of rain forest, cloud forest, dry forest, and multiple intergrades (Fig. 1), all within range of insect flight of each other, and on soil ranging from ancient serpentine to modern marine and volcanic sediments, overlain by vegetation of various ages of succession following up to four centuries of European-style highly heterogeneous ranching, burning, hunting, logging and farming (Janzen 1988a, 2002). A second source of environmental heterogeneity is the extreme trophic specificity displayed by the great majority of the species of caterpillars and parasitoids (although there are a few true generalists as well; e.g. Smith *et al*. 2006–2008). The ACG inventory attempts to obtain specimens from each of these major ecological situations irrespective of adding barcoding as a tool, but when sequence variation appears, the first act is to map these on the above ecological heterogeneity as well as to examine for subtle morphological co-varying heterogeneity. The result of this analysis often leads to an increased effort in the field to get more caterpillars from the nodes of heterogeneity that correlate with the sequence polymorphism (if any be) and to spread the inventory over all ecosystem heterogeneity. For example, it may be found that a morphology-based species occurs throughout ACG dry forest and rainforest, but there are 2–3 barcode-based lumps within this species in the NJ tree (e.g. *Udranomia kikkawai* in Appendix SII), each restricted to one of these two major ecosystems. The consequence is that the number of specimens reared, retained and barcoded is further increased to thoroughly document this division.

This is to determine where these two morphs overlap and to locate yet other barcode morphs that may be 'hiding' within one of the above ranges. As a second example, a single specimen reared from an 'abnormal' food plant and barcoding differently now becomes the cause for an intense effort to get more specimens from that 'abnormal' food plant to determine if they all barcode differently, rather than simply concluding that this was a strange or 'accidental' straying from the 'normal' food plant (as is traditionally the case in morphology-based inventory).

(iv) Species-level names for species, whether morphologically defined or barcode-suggested

Irrespective of barcoding, the ACG inventory has had to evolve an interim taxonomy for its large number of morphologically defined but undescribed (or apparently undescribed) species. What has worked best for both project administration and machine-based information management are several conventions.

1 If the genus is not apparent, the project taxonomist (or biodiversity specialist) fills that field with an alphanumeric of the form 'tachWoodley06', meaning Tachinidae genus six in the ACG inventory, as perceived by taxonomist Norman Woodley (the inventory does not try to guide the entire taxasphere into one consistent interim taxonomy protocol, but tries to be internally consistent, to the degree permitted by individual taxonomists). In this spirit, 'noctJanzen01' is something for which no taxonomic labelling at the genus level for this noctuid moth has been expressed by a taxonomist.

2 At the species level, the pattern is carried forward, and therefore 'Woodley06' is the interim species-level epithet, as in '*Belvosia* Woodley06' for what is commonly called '*Belvosia* sp. 6' in more traditional labelling. By this method, we know that it is a Woodley-defined species and there are no spaces, periods or other ambiguities to confound data entry. Also, only the genus name is italicized, and that, in combination with numbers being prohibited in scientific names, tells anyone that 'Woodley06' is an interim name and not a published valid species name. If more evocative names are needed in a snarl of as yet undescribed species, as in the (now) 11 species of '*Astraptes fulgerator*' in ACG, the inventory has also used species-level interim epithets in CAPS, such as *Astraptes* HIHAMP and *Astraptes* LOHAMP (Appendix SII; Hebert *et al*. 2004). However, we resisted doing this with the hundreds of new species of small parasitic flies and wasps (Smith *et al*. 2008), and from now on, we resist doing it with the greater mass of apparently new moth and butterfly species being discovered through barcoding (e.g., Table 1).

When it is discovered that a morphologically defined species is made up of several (initially) barcode-defined

Table 1 Tally of results of barcoding morphologically defined species in the inventory of 19 families of Area de Conservacion Guanacaste Lepidoptera, Tachinidae, and six genera of microgastrine Braconidae through 2007

Family	Original number of morphology-based species	Number of morphology-based species split up when barcoded	Number of barcode lumps from split species	Number of split out lumps having morphological or ecological correlates	Number of lumps among these putative species	Number of split out lumps lacking correlates	Number of lumps among these putative species	Number of cases of confusion (species)	Number of certain species	Estimated maximum number of species	Number of species that can ID by barcode	Total number of caterpillars reared or parasitoid rearings
Arctiidae	214	30	69	14	31	16	38	1 (2), 1 (2)	231	253	249	9516
Sphingidae	119	19	42	6	14	13	28	0	127	142	142	16 147
Saturniidae	76	13	28	9	20	4	8	0	87	91	91	18 003
Hesperiidae	413	57	169	23	89	34	80	1 (3)	479	525	522	47 280
Nymphalidae	212	28	61	12	24	16	37	1 (2)	224	245	243	18 181
Papilionidae	24	3	7	2	4	1	3	0	26	28	28	1424
Pieridae	31	3	6	1	2	2	4	0	32	34	34	2632
Noctuidae	654	66	171	55	139	11	32	0	738	759	759	24 081
Geometridae	283	30	76	16	39	14	37	0	306	329	329	5951
Hedylidae	5	2	5	2	5	0	0	0	8	8	8	220
Limacodidae	54	13	34	1	2	12	32	0	55	75	75	2656
Lymantriidae	10	2	5	1	2	1	3	0	11	13	13	899
Dalceridae	11	0	0	0	0	0	0	0	11	11	11	111
Bombycidae	41	2	4	1	2	1	2	0	42	44	44	2167
Notodontidae	241	37	92	20	51	17	41	0	272	296	296	22 846
Crambidae	273	27	60	14	32	13	28	0	291	306	306	17 297
Megalopygidae	23	1	4	1	4	0	0	0	26	26	26	663
Riodinidae	81	6	14	1	2	5	12	0	82	89	89	4433
Lycaenidae	61	1	2	0	0	1	2	0	61	62	62	875
Total Lepidoptera	2810	340	848	179	462	161	387	4	3098	3325	3316	195 382
Microgastrine braconid wasps (6 genera)	171							1 (2)		313	311	2597 barcoded
Belvosia flies	20							0		32	32	1728
Tachinidae flies	499							1 (2) 1 (2)		720	716	9671
Grand total	3500							7 (15)		4390	4375	

species (either correlated with other traits or suspected of being so, Table 1), in order not to abandon the 'parent' species' concept and nomenclature, an additional terminology is layered over that described above. Thus, the notodontid moth *Dunama mexicana* becomes *Dunama* mexicanaDHJ01, *Dunama* mexicanaDHJ02, *Dunama* mexicanaDHJ03 and *Dunama* mexicanaDHJ04, for four barcode lumps in the NJ tree (e.g. Appendix SIII). Of course, at this point, it is not known if either or any of these entities are actually *Dunama mexicana*, and correlation with traits other than the barcode are required to feel comfortable with the hypothesis that '*D. mexicana*' in ACG is at least four species (as an aside, the four have now been found to have different genitalia, B.S.). This naming system is also applied to interim names as well, such as in the ichneumonid parasitoid wasp *Hyposoter* INB-42DHJ01, *Hyposoter* INB-42DHJ02, *Hyposoter* INB-42DHJ03, etc. (Appendix SIV).

Barcoding has also revealed an interesting tangle in the taxasphere tradition of recording the identifier of a given specimen. If a morphology-based taxonomist identifies the moth as *Dunama mexicana*, and the inventory identifies the same specimen as *Dunama* mexicanaDHJ02, who is the identifier of the specimen? Actually, both are, although dual identifiers (as opposed to a two-member team) do not find a comfortable home in current taxonomic databases. For the moment, the inventory records the taxonomist as the identifier of the specimens examined by the taxonomist (despite that the DHJ02 was assigned by the inventory), and records members of the inventory team as identifiers of the specimens that have been determined by barcode or visual inspection, since they used the barcode and/or collateral morphological data that they had in hand to make the identification. The inventory explicitly recognizes that this is an area of gradient between one tradition and another.

Barcoding also intersects with a quite different component of species identification or delimitation, that of 'subspecies'. It has been the experience of the project that when a morphological entity was formally described as a subspecies (usually meaning that it differs only 'slightly' from another subspecies of the same species, and often was believed to be allopatric to it), thereby generating a category with many definitions for many people, barcoding and close attention to both morphological/ecological traits and the barcodes very often leads to the conclusion that rather than there being (for example) two subspecies in ACG, there are more simply, two species. This means for the ACG inventory that we predict that in fact the two formerly labelled subspecies are not freely interbreeding when in parapatry or sympatry. This is, of course, a working hypothesis and in some cases the two 'subspecies' will be found to simply blur back together when the two populations encounter each other. In the ACG case, its populations all exist in sympatry or parapatry, but the larger question is whether to use the name for the 'Central American subspecies'

or use the species name, ignoring the geographical segregates that have been tagged with subspecies epithets by others. The inventory hypothesizes that the ACG, Costa Rican, or Central American subspecies will most usually be found to be actually a different species from the Mexican or South American subspecies (plural), and therefore, usually makes the decision to elevate the subspecies name applied to Costa Rica or Central America to species rank in the project databases. *Prepona laertes demodice* (Fig. 3) thus becomes *Prepona demodice* in the inventory (and then becomes *P.* demodiceDHJ01 and *P.* demodiceDHJ02 when it is found by barcoding and caterpillar food plant that it is in fact two sympatric species in ACG, see below).

All of these actions result in the ACG inventory with DNA barcoding greatly increasing the number of specimens that it retains and processes, as well as increasing the number of species encountered, as compared with an inventory based purely on morphologically defined species.

B. Philosophical and scientific results of adding DNA barcoding to an inventory

Prior to DNA barcoding, efforts to identify, and therefore link the inventory specimens to collective knowledge about them, was almost entirely done by morphological comparisons with curated collections or literature, and through direct assistance from a taxonomist. There is great heterogeneity in the degree of taxonomic attention and intensity (as well as collection thoroughness) that has been applied to different higher taxa among the very species-rich Neotropical Lepidoptera. Below, we start from the fundamental base of enormous but heterogeneous effort by morphological taxonomists and discuss what barcoding (along with collateral rearing data) has added to it in this inventory, and especially the act of flagging specimens to be examined more closely morphologically. This is not a referendum on success of morphological taxonomy. Of course, much of what barcoding has exposed would probably have been discovered morphologically or by genitalia, but barcoding greatly increases the efficiency of where to allocate this scarce resource. While not the purpose of barcoding, the extremely good match of barcode-identified groups to morphologically identified groups is at once a statement that morphology works extremely well for both identification and species discovery — if the user has the training and knowledge to apply it.

1. ACG barcodes reliably identify morphologically defined species within ACG. To date, the ACG inventory has barcoded about 3500 morphologically defined species of moths, butterflies, and parasitoid flies and wasps (Table 1). These are combinations of species that already bear morphology-based scientific names, and species with interim names assigned based on morphological characters that await replacement by scientific names. In this set, there are 4 instances of

Lepidoptera, 2 of Tachinidae and 1 of microgastrine braconids where either 2 or 3 morphologically defined species have been found to have apparently indistinguishable barcodes (Table 1). In each of these cases, morphological traits (or *n* = 1, ecological traits) reliably distinguish among the members of the barcode-equal group, and no group's barcode is confusable with any other inventory barcodes (see also NJ trees in Appendices SI–SX).

Put another way, for 3500 morphology-based species of insects in three orders in ACG, a 654-base pair COI barcode yields 99.57% success in identifying species. The 15 species that are confusable are only confusable with 1 or 2 close relatives, rather than with any of the other species. In other words, if a single full barcode sequence from any one of those 3500 species is dropped on the project NJ tree (based on more than 50 000 sequences), it invariably lands in the lump already blessed with that morphology-based name (and for 15 species, it lands in a lump blessed with two or three names). If we include all 4375 barcode-defined lumps in the NJ trees (3500 morphology-based species prior to barcoding, plus those lumps that were found to have other correlated traits after barcoding, plus those lumps that remain distinguishable only by barcode), the success is 99.68%. The inventory can 'identify' +99% of its OTUs that are morphology-based and barcode-based for the cost of getting the barcode and comparing it with the growing barcode library. If there are hybrids hiding in the analysis to date, in each case encountered except for one, they apparently have a barcode that matches the morphology and other traits of the parent normally associated with that barcode (and are therefore indistinguishable from nonhybrids).

There are no indications that this general rate of identification success with these three orders of insects will be significantly altered with a larger sample size of species or with more specimens from ACG. However, this does not mean that the inventory stops barcoding specimens of morphology-based species that have now been 'inventoried', because further barcoding continues to expose hidden species-level heterogeneity and clarify whether slight variation in barcodes within what appears to be one population is a signal that it is two or more.

The overall conclusion that barcodes work very well for identifying the members of this species-rich insect fauna does however have complications, caveats and exceptions that can be explored in many ways. A few of these are lightly explored below so as to illustrate some of the salient features of the inventory topography, but each merits far more examination than space allows here.

2. An ongoing illustrative case study: Phoebis argante (Pieridae). As soon as barcoding was applied to the inventory, it became apparent that it exposes overlooked species (e.g. Hebert *et al*. 2004; Smith *et al*. 2006–2008, Burns *et al*. 2008). The large yellow pierid butterfly *Phoebis argante* (Fig. 2) is

an ongoing example. *P. argante* is a well-known butterfly throughout the neotropics (DeVries 1987). The ACG inventory has reared it more than 350 times from caterpillars found on *Zygia longifolia* and six species of *Inga* (Fabaceae) in ACG dry forest, rain forest and intergrades. When barcoded, however, two very distinct lumps of barcodes appeared in the NJ tree of ACG Pieridae, each being so far apart that a similar species (*Phoebis virgo*) falls between them (Appendix SV). The two barcode lumps, *P.* arganteDHJ01 and *P.* arganteDHJ02, do not correlate with food plant, season or ecosystem. The caterpillars of both have been collected from the same plant at the same time, although the caterpillar of *P.* arganteDHJ01 is about three times more commonly found than is that of *P.* arganteDHJ02. The genitalia of the two species appear to be the same at the level of scrutiny normally accorded to pierid genitalia (and were first decreed by an experienced taxonomist to be 'the same') but on closer study display slight but consistent differences (Fig. 2). Likewise, when sorted by barcode, the fine details of the wing colour pattern of both sexes became evident as a 100% reliable method for distinguishing these two species that are contained within ACG '*P. argante*' [e.g. note the difference in the pattern of black on the outer margin of the male forewing (Fig. 2)]. Prior to barcoding, these segregating wing traits had been viewed by the inventory and experienced taxonomists as variation not indicating a species-level dichotomy, and indeed both cryptic species are somewhat variable in these traits. UV light-visible wing reflectance patterns, long studied in Pieridae and interspecifically discriminatory (Silberglied & Taylor 1973; Allyn & Downey 1977; Rutowski & Macedonia 2008) are not different between the males of the two species of *P. argante* but do differ on the undersides of the females in ACG. Both species occur throughout Costa Rica, as based on morphological inspection of 66 net-caught adults in the INBio national inventory (IC). However, in contrast to ACG inventory rearing records, *P.* arganteDHJ01 is the low-density species among these net-collected free-flying adults.

Which, if either, of the two species of *P. argante* in ACG is actually *P. argante*? *P. argante* was described in 1775 from a Brazilian specimen and several so-called subspecies of *P. argante* have been described from Caribbean Islands, Peru, 'America', and Mexico (Lamas 2004). All of these are candidates to match one, both, or none of the species in ACG and Costa Rica. However, a direct quote from a 1929 analysis of *Phoebis argante* Neotropical taxonomy (Brown 1929: 12) is relevant:

'Two forms of the mainland race occur: the nymotypical *argante* and Cramer's *hersilia*. In studying a series of about two hundred specimens from over the entire range of this insect, it became apparent at once that *hersilia* is the dominant tropical form and *argante* the dominant form in the north and south. Form *hersilia* is

Fig. 2 Left panels. *Phoebis* arganteDHJ01. Right panels. *Phoebis* arganteDHJ02. Male adults above, left harp of dissected male genitalia in centre, female adults below.

gradually separating itself from the *argante* type and forming a purely tropical race that will be flanked on the north and south by forms similar to our present *argante*. In this case the difference in forms is apparent in the males — the marginal row of black dots having become confluent and having formed a black band in *hersilia*. However, only about forty per cent of the specimens from the northern part of South America show a complete band, but only one to two per cent show no trace of this band. There is a transitional group of about seventy per cent of all the males in the collections I have examined. The form *hersilia* ranges from Honduras to Bolivia — almost the entire range of the mainland race.'

If Brown had had our ACG barcode results, there is little doubt that he would have concluded that *P.* arganteDHJ01 is *P. argante* and elevated *P.* arganteDHJ02 to *P. hersilia*.

However, when the holotypes behind these two names are examined in detail with a full barcode analysis of '*P. argante*' over all of its range, yet additional barcode groups may be found that correlate with morphological, geographical, ecological and/or behavioural traits, suggesting that *P. argante* is yet more than two species on the mainland.

Of the 31 morphology-based species of Pieridae barcoded to date in the ACG inventory, three have split into two barcode lumps each (Table 1; Appendix SV), but to date, no other traits have been found to correlate with the barcode lumps of the other two species. These lumps now are flagged and will be treated to yet more intensive rearing and collateral data capture, as well as scrutiny of other genes, a treatment bestowed on all such cases in the inventory.

Now multiply the example of *P. argante* by hundreds for an image of the taxonomic tangles unearthed by barcoding the ACG inventory.

3. Barcodes of ACG specimens agree with those of traditional morphologically defined conspecifics from a wider geographical range. The ACG inventory, and the subsequent intensive addition of barcoding to it, deliberately and explicitly does not itself attempt to spread the action and analysis over a larger geographical area. However, nothing encountered to date implies that the barcode derived from an ACG morphologically defined species will fall within the barcode lump created by specimens of a different morphologically defined species from some other area. For example, when the barcodes of the 119 species of morphologically defined ACG Sphingidae (each being unambiguously and long-recognized as different at the species level) are dropped into a NJ tree of the extensively barcoded Sphingidae of the world, a global sample that contains specimens of all of the ACG morphologically defined species from multiple points in Central and South America as well as many other congeners, no ACG species' barcode falls into the barcode cluster of a different morphologically defined species from elsewhere (R. Rougerie, personal observation). Doing this with other families may flush out some exceptions, but each will require careful development to be certain that the two apparently barcode-confused morphology-based species really are two different species. Just within ACG there are dramatic cases of barcodes confirming that what was thought to be two species are indeed just one (Fig. 10, see below).

4. DNA barcoding of the inventory specimens finds that many morphologically defined species are composites of overlooked species. The example of *P. argante* offered above is commonplace among the ACG specimens. Barcoding has disclosed numerous similar examples that probably flag species complexes that may be more clearly revealed with more haphazard or directed sampling of both specimens and collateral information. For example, when 2810 morphologically defined ACG species of Lepidoptera were barcoded, 340 of them (12%) broke up into two or more lumps of barcodes in the NJ tree, for a total of 848 lumps in the NJ tree (Table 1). Fifty-five per cent of these barcode lumps have already been found to have morphological, ecological, and/or micro-ecogeographical correlates that support the hypothesis that they are indeed actual overlooked species. The other 45% remain in the category of 'perhaps'. These are a mix of cases created by pseudogenes (see below), laboratory analysis glitches, true barcode within-population polymorphisms, and species in which collateral traits are present but are somehow out of sight. For example, to date, we have not conducted a systematic survey of genitalia across the morphologically defined species that contain two or more barcode lumps. Such a survey (which of course could have also been done prior to barcoding) will undoubtedly reveal morphological collateral that strongly support the hypothesis that a barcode polymorphism within a superficially morphologically defined species

is flagging cryptic species. All project barcodes have been subject to scrutiny (and removal) for pseudogenes through routine BOLD protocols (Hebert *et al*. 2003, 2004; Ratnasingham & Hebert 2007, http://phe.rockefeller.edu/barcode/blog/). The laboratory analysis glitches (cases of contamination, controversial reads of trace files, incomplete barcodes) are routinely searched for and removed both by the inventory analysis of NJ trees and BOLD. We posit that easily 90% of the 45% 'perhaps' are a combination of cases where morphological, ecological and micro-ecogeographical correlates have been overlooked. Rarely will we find that what appear to be within-population barcode polymorphisms are due to the contemporary fusion of previously diverging populations. In all of these 'perhaps' cases, we have been careful not to baptize different barcode lumps within a morphologically defined species as being several interim barcode species unless there is at least some suggestion other than the barcode polymorphism that there are more than one species, and even when so baptized, it is simply a flag to remind to scrutinize the situation more closely.

This 'species prospecting by barcoding' is adding 5–15% overlooked (cryptic and usually undescribed) species to a well-studied fauna of large moths and butterflies (e.g. Hebert *et al*. 2004; Burns *et al*. 2008), and as many as 100% overlooked (cryptic and almost invariably undescribed) species to the little-studied fauna of morphologically and visually similar small wasps and flies (e.g. Smith *et al*. 2006–2008). The case of *Phoebis argante* described above illustrates resolution of barcode heterogeneity through morphology. An example of resolution through food plants is the case of *Prepona demodice* (Nymphalidae) mentioned earlier. This is a well-known gaudy butterfly, is often seen at fermenting fruit baits in dry forest and rain forest — e.g. search Flikr for *Prepona laertes* — http://www.flickr.com/search/?q = prepona + laertes. Its caterpillars have been found feeding on Chrysobalanaceae and Fabaceae several hundred times in ACG, and were thought to be well understood. However, when the many reared adults were barcoded, they fell into two distinct groups (Appendix SVI). One group was reared from caterpillars found only on Chrysobalanaceae and the other from those found only on Fabaceae. The adults are indistinguishable by facies (Fig. 3) but have not yet had their genitalia compared. Both *P. demodice* caterpillars co-occur in ACG dry forest and rain forest. *P.* demodiceDHJ01 and *P.* demodiceDHJ02 are unambiguously two species, each specialized on a different group of food plants. Whether either matches the holotype of *P. demodice* remains to be determined; this 'species' has had at least 100 different scientific names applied to this apparent morphologically based species since it was described from South America in 1824 (see Lamas 2004). One or both or none of the ACG '*P. demodice*' could match the holotypes of *P. laertes octavia* and *P. laertes demodice*. Were it not for barcoding, the inventory

Fig. 3 Four pairs of overlooked Area de Conservacion Guanacaste (ACG) species flushed out of the inventory by barcoding. Top to bottom: left, *Prepona* demodiceDHJ01 (caterpillars eat Chrysobalanaceae); right, *Prepona* demodiceDHJ02 (caterpillars eat Fabaceae); left, *Cocytius* luciferDHJ01 (caterpillars found in ACG dry forest); right, *Cocytius* luciferDHJ02 (caterpillars found in ACG rain forest); left, *Udranomia* kikkawaiDHJ01 (caterpillars found in ACG dry forest); right, *Udranomia* kikkawaiDHJ02 (caterpillars found in ACG rain forest); left, *Eacles* imperialisDHJ01 (caterpillars found in ACG rain forest); right, *Eacles* imperialisDHJ02 (caterpillars found in ACG dry forest).

would have continued to assume that the caterpillars of *P. demodice* eat plants in two quite different families, in the same place at the same time. This degree of unresolved ecological specificity is scattered across the inventory in other taxa (e.g. Janzen 2003).

Barcode groups within a morphologically defined apparently single species occasionally segregate by microgeography within ACG. Temperature, rainfall, and elevation all differ across ACG transects from the dry forest ecosystem to the cloud forest ecosystem to the rain forest ecosystem (Fig. 1). *Cocytius lucifer* (Fig. 3) is a large and well-known sphingid moth (Sphingidae) that migrates seasonally throughout Costa Rica, and back and forth between the ACG dry forest ecosystem and rain forest ecosystem (Janzen 1988b). Morphologically defined, it is unambiguously one species by both facies and genitalia (Thierry Vaglia, in lit). However, almost all the caterpillars of *C. lucifer*DHJ01 have been found in ACG dry forest, and all the caterpillars of *C. lucifer*DHJ02 (Appendix SVII) have been found in ACG

rain forest a few kilometres away. All feed on Annonaceae, but the rain forest caterpillars have a darker pattern than do the nearly pattern-free dry forest caterpillars. As a first pass hypothesis, *C. lucifer*DHJ02 is a rain-forest resident species, while *C. lucifer*DHJ01 is a dry-forest breeding species that migrates to the rain forest in the dry season, and occasionally breeds there as well (as based on a single *C. lucifer*DHJ01 caterpillar found about 20 km east of the ACG dry forest).

A second example is *Udranomia kikkawai* (Fig. 3), a small and extremely host-specific species of skipper butterfly (Hesperiidae). Its caterpillar eats only very young foliage of Ochnaceae in ACG dry forest and rain forest (Janzen & Hallwachs 2008). When barcoded (Appendix SII), *U. kikkawai*DHJ02 was found to occur exclusively in the rain forest, and *U. kikkawai*DHJ01 and *U. kikkawai*DHJ03 to be exclusively dry-forest animals. Intense scrutiny of facies and genitalia of these three presumptive species of skippers shows no difference among them in these traits. This

contrasts with three other ACG hesperiid-described species pairs where there are facies and, in two out of three cases, strong distinctive genitalic differences that correlate with their microgeographical distributions and barcode segregations (Burns *et al*. 2007). The large yellow and very well-known saturniid moth *Eacles imperialis* (Saturniidae) (Fig. 3) ranges from Canada to Argentina (Lemaire 1988) and its caterpillar feeds on hundreds of plant species (e.g. Janzen 2003). It is viewed as an extreme generalist in ecological tolerance. The barcodes of the ACG rain forest *E.* imperialisDHJ01 differ from those of the ACG dry forest *E.* imperialisDHJ02 by 8% (52 nucleotides) (Appendix SVIII) but the two are indistinguishable by facies at the level of scrutiny normally accorded large showy and slightly variable tropical saturniid moths (the genitalia have not yet been compared among *E. imperialis* barcode lumps). In the transition zone between ACG dry forest and rain forest, both adult males have been collected from the same light trap and caterpillars of both reared from the same patch of forest; they are ecologically parapatric and coexist in the ecotone between the dry forest and rain forest. Interestingly, the barcode of *E.* imperialisDHJ02 differs from *E. imperialis* barcodes from the population that extends from Canada to South Carolina by 5% (P.D.N.H. and R.R.); it is an easy prediction that there are at least three species within this part of what has been called *E. imperialis*. How these match up with the 10-plus subspecies that have been described within *E. imperialis* (Lemaire 1988) is a matter for future research.

5. The smaller the individuals of a species, the more likely that barcodes will flush out overlooked species. When barcoded, higher ACG taxa made up of small similar species display a substantially greater frequency of cryptic species than do large showy butterflies and moths, paralleling the trend recently encountered in Australian beetles (Stork *et al*. 2008). As the inventory moves to quite small non-leaf-mining moths (Elachistidae, Gelechiidae, Tortricidae, Crambidae, Pyralidae, etc.), this becomes quite evident. To some degree, this is a consequence of less taxonomic attention and intensity (many apparent species have never had their genitalia examined). It also is a consequence of small Lepidoptera that do not rely on species-specific appearance for mating and predator avoidance. This anthropo-biological phenomenon has already been well documented with ACG parasitoid tachinid flies, braconid wasps, and ichneumonid wasps. When the 20 species of extremely similar black and yellow (bumblebee mimic) ACG *Belvosia* tachinids (all reared, all extremely host-specific) were barcoded, they became at least 32 species (Smith *et al*. 2006). When we barcoded 16 other morphologically defined species of other extremely generalist genera of ACG-reared tachinids — all looking vaguely like various sizes of houseflies with a few big black ones added in — they were found to be at least 73 species (Smith *et al*. 2007). These

generalists were found to include an overlooked array of specialists. Much of their 'generalist' trait disappeared when barcodes were added to the morphological and host traits used to hypothesize the existence of a species. However, 9 of the 16 generalist species (now defined by both morphology, barcode and other genetic markers) remained legitimate generalists, although with shorter lists of host species (and see Hulcr *et al*. 2007).

When the minute (2–4 mm long) microgastrine wasps (Braconidae) that parasitize all families of caterpillars of ACG macromoths and many micromoths were barcoded, we found almost twice as many species of wasps as had been recognized by morphology-based species recognition; furthermore, the barcodes corroborated the ecological observation that the wasps are extremely host-specific (Table 1 and see Smith *et al*. 2008). There were found to be almost no generalists among hundreds of species. A telling example among quite small Ichneumonidae is the undescribed common species *Hyposoter* INB-42 that ranges from Mexico to Costa Rica (I.D.G.). It has been reared over 800 times from second to fourth instar caterpillars of 43 species of ACG Hesperiidae and in all ACG ecosystems and intergrades (Janzen & Hallwachs 2008). Many other species of ACG Ichneumonidae parasitizing caterpillars appear to have this kind of generalism within a higher taxon of host. However, when barcoded, *H.* INB-42 was found to be at least eight species in ACG, each restricted to a distinctive subset of the 43 species of Hesperiidae. This mirrors the case of the microgastrine braconid *Apanteles 'leucostigmus'*, which also attacks at least 46 species of ACG Hesperiidae caterpillars and turned out to be at least 32 species of extreme specialists when barcoded and the barcode lumps matched with caterpillar species (Smith *et al*. 2008).

6. Why there are many look-alike species with different barcodes (and matching different ecology) in one place. When similar species are encountered, it is commonplace to think of them as having only recently separated evolutionarily from each other, as being 'closely related', as being 'sibling species'. However, this broad-sweep barcode survey of ACG Lepidoptera, Diptera and Hymenoptera brings to mind two other processes that result in species being 'similar' (at least in the eyes of a large diurnal mammal that often needs glasses and a microscope to 'look' at an insect) quite irrespective of how long they have been on separate evolutionary trajectories.

First, throughout the tropics mimicry is far more widespread than is generally realized, leading to the feeling 'if there are any non-mimics, please identify yourself'. Two dramatic examples are offered by barcoded ACG Hesperiidae. The facies displayed by the *Astraptes 'fulgerator'* that broke up into 10 species in ACG when barcoded (Hebert *et al*. 2004), and now 11 species, is commonly thought of as a 'showy butterfly' and not much else. However, when

Fig. 4 Twenty-eight of the species in the blue–white–black mimicry ring of pyrgine and hesperiine ACG Hesperiidae (see Appendix XI for names and voucher codes), with the first 11 text-wise being species of male "*Astraptes fulgerator*" (e.g. Hebert *et al*. 2004 for the first 10, the 11th, *Astraptes* ENTA, found since). Compare with Figs 5 and 6.

seen in the context of a complex tropical habitat and the neotropics as a whole, it is clear that the *A. fulgerator* facies is one that 'works' and has been converged on by many lineages (Fig. 4). This facies has likely been in the tropics for

the many millions of years that there have been visually orienting (potentially) butterfly-eating birds. In short, when one of these lineages splits, and thus their barcodes begin to diverge, whatever selection is favouring the *A. fulgerator*

Fig. 5 The 28 last instar caterpillars of the matching adults in Fig. 4, each in the same cell as its adult (see Appendix XI for names and voucher codes). Compare with Fig. 6.

colour pattern still favours it; the blue, white and black pattern — widespread in the neotropics — is probably an ostentatious signal that says to a bird 'don't bother to try, I am way too fast for you' (D.H.J.). If it works, wear it. At the same time, the divergent lineages have caterpillars with interspecifically different food plant, crypsis and mimesis ecologies, resulting in dramatically different caterpillar colours despite the very similar adults (compare Fig. 4 with Fig. 5). Exactly the same case is offered by the blue, white and black pattern — but arrayed a quite different way

across the wings — of at least 12 other extremely similar mimetic species of ACG large Hesperiidae (in the genera *Phocides, Elbella, Parelbella, Jemadia*). Other parallel cases in ACG are offered by *Adelpha* (Nymphalidae) and its mimics, *Urbanus* (Hesperiidae) and its mimics, *Parides* (Papilionidae) and its mimics, etc. Where there are large mimicry complexes with long-term partnerships, it is quite reasonable for species to be separated long enough to have strongly different barcodes (Fig. 6) yet retain extremely similar appearance through membership in these large and presumably generally effective mimicry systems.

Second, as mentioned earlier, the smaller the insect, the less selection there is favouring highly visible morphological traits — the kind used by large mammalian taxonomists — that are strongly different among evolutionary lineages. Tiny parasitic wasps offer a dramatic example. ACG habitats are seething with them, as evidenced by Malaise trap catches and their high frequency as parasitoids of ACG caterpillars, yet their differences are close to invisible to the uninitiated and even to the specialist. Different species don't 'look' different (Fig. 7) because there is minimal selective pressure for them to do so. However, they can have wildly different interspecific host-searching and host-defence-tolerance traits, and the different hosts and barcodes to go along with them — as evidenced by both Smith *et al.* 2008 and the accumulating records for all parasitoids in the ACG inventory primary data base (Janzen & Hallwachs 2008).

These observations do need, however, to be accompanied by the reminder that when something appears to be one species morphologically, and then is found to break up into two or more barcode lumps in an NJ tree, it is worthwhile to search for overlooked morphological traits (as in the example of *Phoebis argante* above). When inconspicuous differentiating morphological traits are located, it has often been the experience of the inventory that indeed, observant earlier taxonomists had long ago noticed these traits, or found it easy to incorporate them into their diagnoses of species recognized long before either the inventory or barcoding. Figure 8 displays four examples that emphasize how subtle can be the difference in appearance of recognized species that have substantially different barcodes.

1 In Costa Rica, the satyrine butterflies *Taygetis laches* and *Taygetis thamyra* (Nymphalidae) have long been hidden under the name *Taygetis andromeda* (DeVries 1987), but when *T. andromeda* was barcoded in ACG, it was found to contain two lumps of barcodes about 3% different from each other (Appendix SVI). L.M. and J.M. quickly realized then that the ACG *T. andromeda* was really *T. laches* and the other lump was *T. thamyra*, a South American species not previously noticed in Central America. Once it is realized that these two totally sympatric species occur in ACG, they can be distinguished by the pattern on the underside of the hind wing and forewing shape (Fig. 8), as well as by their barcodes.

2 The moth *Perigonia ilus/lusca* (Sphingidae) (Fig. 8) has long been argued over as to whether it was one species, two subspecies, or whatever (D'Abrera 1986; Kitching & Cadiou 2000). Rearing records, differences in degree of yellow, size, and as pointed out by Jean-Marie Cadiou, the hue of the underside of the wings, long ago convinced the inventory that it is two species in ACG dry forest, and this was reinforced by a taxonomic decision (Haxaire 1996). When barcoded, *P. ilus* and *P. lusca* were found to differ by about 2.5% in their barcodes (Appendix SVII) and ACG specimens are easily identified by facies (Fig. 8). However, the inventory has just discovered that '*P. ilus*' as morphologically defined contains a second ACG rain forest entity, differing by about 2% (Appendix SVII) (and 1 km) in its barcode from the dry forest *P. ilus*, as based on specimens taken with a light trap. When the barcodes of these two species were then compared with the many *Perigonia* barcodes accumulated in BOLD from throughout the neotropics, it became apparent that *P. ilus*DHJ01 ranges from Guatemala to Argentina, and *P. ilus*DHJ02 ranges from Guatemala to Venezuela. As with the *Phoebis argante* case above, the question then becomes which, if either, matches the type specimen of *P. ilus* and if it is worth the time and effort to describe yet another species (or two) of *Perigonia*. To show how complex this can be, *P. ilus* was described by Boisduval, and therefore there is no holotype, only a type series, an array of specimens from Mexico and Honduras that could easily contain both species of *P. ilus* (Jean Haxaire, in lit).

3 It is well known that the papilionid butterfly *Heraclides autocles* (Fig. 8; also known as *Papilio thoas* and *P. thoas autocles*) caterpillars eat Piperaceae while those of *Heraclides cresphontes* (also known as *Papilio cresphontes*) eat Rutaceae. Nonetheless, these two species have been viewed as impossible to distinguish without rearing or close examination of their genitalia (DeVries 1987), and the confusion was enhanced by the DeVries (1987) field guide to these butterflies in Costa Rica figuring a look-alike species (*Heraclides paeon*) in place of *H. cresphontes* (Brown 1988). However, joining ACG rearing data with barcode data — they are about 5% different (Appendix SIX) — corroborates a subtle and variable wing pattern character mentioned by Tyler *et al.* (1994), that *H. autocles* has four yellow spots on the lower outer margin of the forewing (like that of *H. paeon* figured in DeVries (1987)), while *H. cresphontes* has three (occasionally a trace of a fourth), as well as a different quality to the yellow of its upperside pattern.

4 The moth *Pachydota saduca* (Arctiidae) (Fig. 8) is common at ACG light traps and more than 300 *P. saduca* caterpillars have been reared from ACG cloud forest to rain forest over 30 years. It is so well understood that many reared specimens were discarded after it was thoroughly
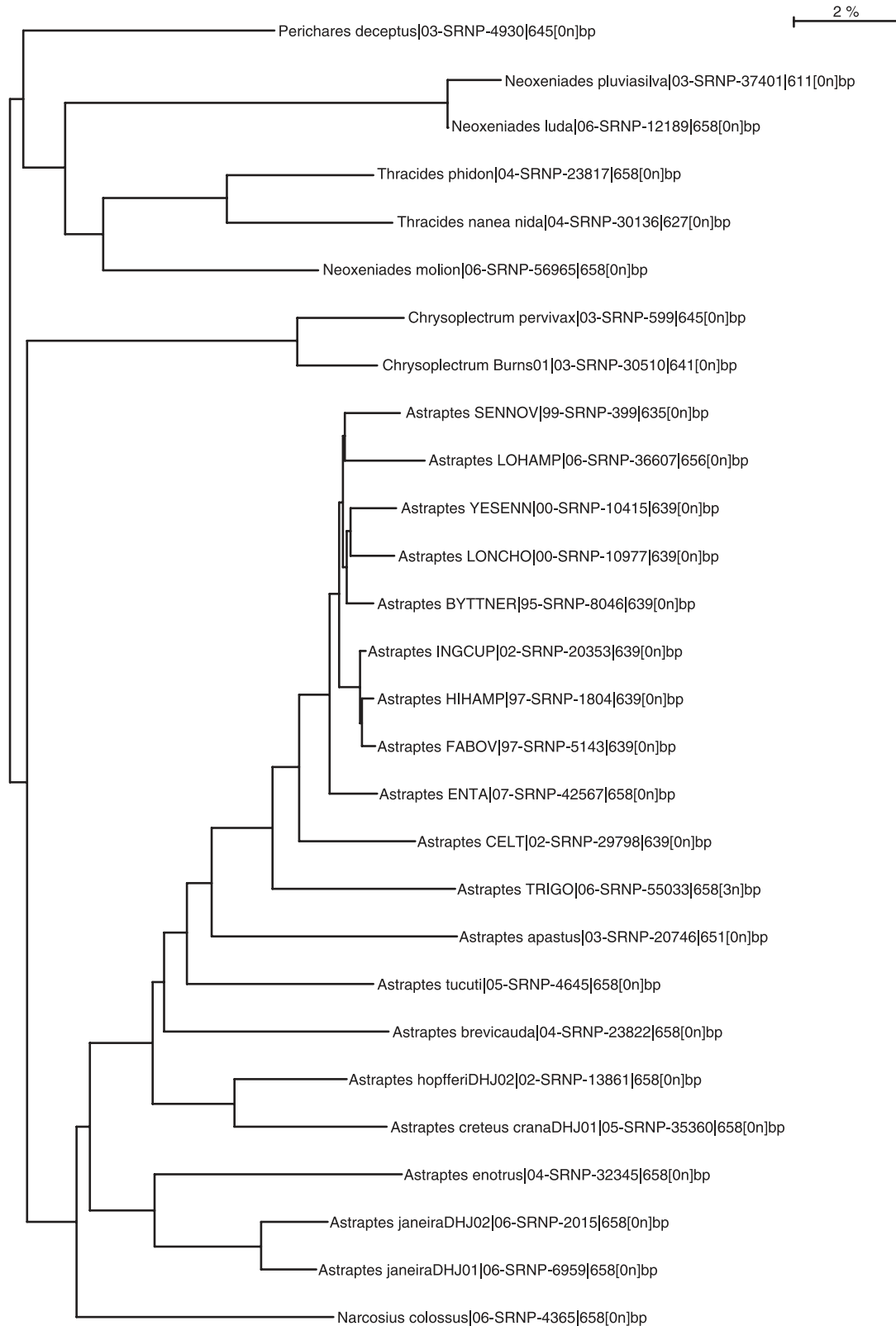
Fig. 6 Raw NJ tree from BOLD for the 28 species of Hesperiidae in Figs 4 and 5, as based on a single 'representative' barcode from each species, where 'representative' means haphazardly selected from the lump of barcodes for that species in the Area de Conservacion Guanacaste inventory.
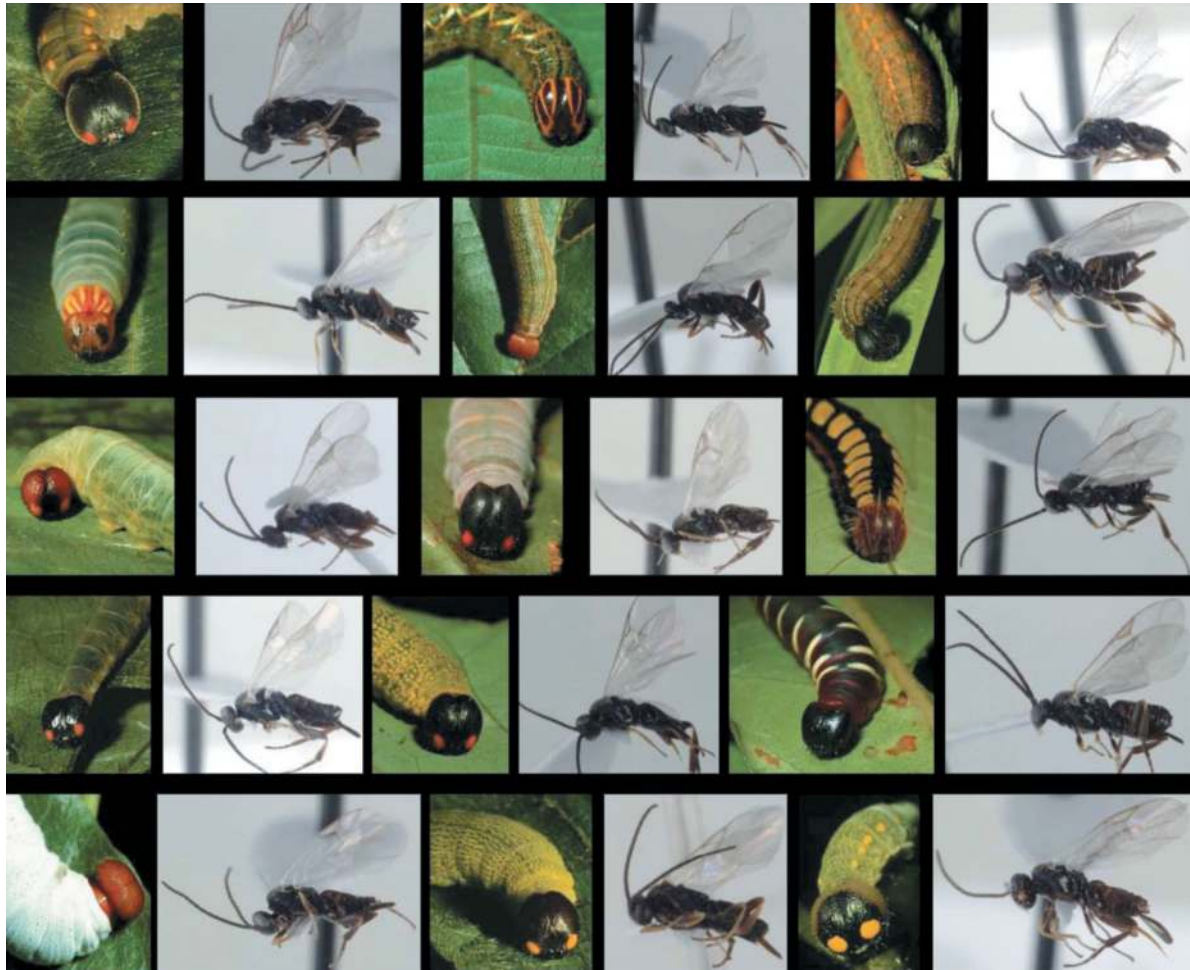
Fig. 7 Fifteen of the 32 presumptive Area de Conservacion Guanacaste species of '*Apanteles leucostigmus*' that are exceedingly similar morphologically but oligophagous host-specific to different species of caterpillar hosts and have different DNA barcodes, paired with their host caterpillars (see Smith *et al.* 2008).

barcoded and found to be quite barcode monomorphic. However, in 2007 a single specimen of *P. saduca* barcoded dramatically differently from its hundreds of conspecifics. It was first assumed that this was a contaminant or pseudogene; however, when compared with the entire BOLD Lepidoptera database, it was found to be a very normal arctiid barcode. Re-examination of the morphology of this specimen by a taxonomist discovered that it was in fact *Pachydota rosenbergi* (Fig. 8), quite distinguishable by its genitalia (and as discovered later, by its caterpillar being white instead of the black basic colour of *P. saduca*). This led to a thorough barcoding of all retained '*P. saduca*' specimens, with the discovery of two more specimens of *P. rosenbergi* in the inventory (Appendix SX), and the realization that others were very likely discarded before 2007. A yet more inconvenient realization is that all future '*P. saduca*' need to be barcoded (cheap) or have their genitalia examined (expensive),

since they cannot be reliably distinguished at a glance (unless the caterpillar colour was recorded).

7. Unexpected overlooked species. If a long list of ACG inventory morphology-based species is presented to a taxonomist quite familiar with this list, and the taxonomist is asked to predict which ones will be found to be made up of cryptic species clusters when barcoded, many of the complex species complexes will be predicted because they have the subtly variable traits mentioned above, or have long been the subject of discussions as to what the variation might mean. However, it has been our experience that another category of complexity is not predicted. This category contains the species that happen to have a conspicuous morphological identifier trait, a trait that leads to easy keying and easy identification of the species when doing rapid sorting of large samples. Because they are so easy to identify, they often have not received intense

Fig. 8 Four pairs of Area de Conservacion Guanacaste species that had morphology-based names applied before barcoding, but can be easily distinguished with barcodes as well as by careful attention to details of morphology (facies). Top to bottom: left, *Taygetis laches*; right, *Taygetis thamyra*; left, *Perigonia lusca*; right: *Perigonia* ilusDHJ01; left, *Heraclides autocles*; right, *Heraclides cresphontes*; left, *Pachydota rosenbergi*; right, *Pachydota saduca*.

scrutiny — 'everyone knows that species'. The three (at least) species of brilliantly black-and-white killer whales (LeDuc *et al*. 2008) offer a mammalian example. The brilliant blue *Prepona demodice* described above is a butterfly example, while *Apanteles leucostigmus* is a tiny braconid wasp example — it has a whitish stigma on an otherwise transparent wing with black veins, is quite small, and distinctively lacking in other morphologically useful traits — and turned into 32 presumed species when barcoded (Smith *et al*. 2008). At least five common species of large ACG sphingid moths appear to fit into this category — *Pachylia ficus*, *Eumorpha satellitia*, *Protambulyx strigilis*, *Aleuron chloroptera*, and *Xylophanes porcus*, as well as the *Cocytius lucifer* in Fig. 3. The large and conspicuous ichneumonid wasp, *Creagrura nigripes*, with a distinctive black tip to its wing, that ranges from Canada to Argentina (I.D.G.), turned out to be three ACG species when barcoded, each parasitizing a distinctive and related set of

hesperiine skipper caterpillars. Another ACG ichneumonid, *Cubus validus*, with a similar range and distinctively yellow-and-black-ringed abdomen and amber wings, well known to indiscriminately attack crambid moth caterpillars in their rolled leaves (I.D.G.), was found to be eight ACG species when barcoded, each attacking a distinctive set of species of leaf-rolling crambids.

8. Apparently dissimilar species with very similar barcodes. A taxonomist tends to view as very dissimilar those species that display very different traits from the viewpoint of the large diurnal mammal that we are. When a pair of such species has very similar barcodes, it conflicts with the concept that morphological dissimilarity takes a long time to evolve. A striking ACG case of dissimilar species with very similar, or identical, barcodes is that of *Adelpha melanthe* and *Adelpha pseudaethalia* (Nymphalidae) (Fig. 9), their barcodes differ by only 0–1 base pairs (Appendix SVI) yet

Fig. 9 Three pairs of Area de Conservacion Guanacaste species with very different adult appearance but very similar barcodes as well as similar caterpillar and pupal stages, and the same caterpillar food plants. Top to bottom: left, *Adelpha melanthe*; right, *Adelpha pseudaethalia*; left, *Ormetica temperata*; right, *Ormetica guapisa*; left, *Astraptes tucuti*; right, *Urbanus pronta*.

the adults have extremely different colour patterns but very similar larvae and pupae (Willmott 2003). The answer to this conundrum very likely is that all of the many tens of species of *Adelpha* are members of large Batesian, Mullerian and (probably) Mertensian mimicry rings. It is likely that when the ancestor of this pair split into two, there was intense selection on one lineage to hop into a different mimicry ring, something that could be accomplished with very little mutation and cost very little evolutionary time, not enough time for barcode differences to accumulate. The same may be said of the arctiids *Ormetica temperata* and *O. guapisa* (Fig. 9; Appendix SX), which have such indistinguishable barcodes that it was initially believed that it was simply a polymorphic species (such as *Calodesma maculifrons*, Fig. 10). A somewhat similar case is that of *Astraptes tucuti* and *Urbanus pronta*. These two species of Hesperiidae have 6% different barcodes but nevertheless *A. tucuti* positions in the portion of the NJ tree that contains various species of *Astraptes* (Appendix SII) and not among the many other species of *Urbanus*. These two species have very similar caterpillars, pupae, and food plants, but differ

strongly in adult appearance (Fig. 9). Furthermore, each species is part of a huge mimicry ring of Neotropical, blue, white and black Hesperiidae (Fig. 4) or long-tailed brown Hesperiidae (*Urbanus, Polythrix, Typhedanus, Chioides, Aguna*, etc.).

9. Matching males and females, and polymorphs, with barcodes. An obvious application of DNA barcoding — and a capability beyond what genitalic comparison provides — is matching males and females of highly dimorphic species (e.g. Hulcr *et al.* 2007). While ACG Lepidoptera, parasitic flies and parasitic wasps have not generally presented exceptional difficulty in matching males with females, barcoding the inventory specimens has uncovered a few spectacular cases. The arctiid moths *Loxophlebia flavipicta* (all males in the INBio collection) and *Loxophlebia egregia* (all females in the INBio collections) (Fig. 10) were both described by William Schaus in 1911 as two Costa Rican species. Both species were reared by the caterpillar inventory but they were encountered as solitary caterpillars years apart and given their different names by a very good

Fig. 10 Upper two pairs: very different males and females described as different species, but found to be one via DNA barcoding. Upper — left, *Loxophlebia flavipicta* (female) formerly known as *Loxophlebia egregia*; right, *Loxophlebia flavipicta* (male), Arctiidae; left, *Dysschema perplexa* (female) formerly known as *Dysschema guapa*; right, *Dysschema perplexa* (male), Arctiidae. Lower two pairs: within-sex polymorphisms identified via barcoding. Lower — left, *Calodesma maculifrons* (yellow morph); right, *Calodesma maculifrons* (dark morph, known previously as *Calodesma melanochroia*), Arctiidae; left: *Heraclides tolmides* female, one morph; right, *Heraclides tolmides* female, the other morph (previously thought by the inventory to be the female of *Heraclides rhodostictus*), Papilionidae.

arctiid taxonomist. When they were barcoded and found to have identical barcodes, contamination was suspected. However, the caterpillar images and food plants were then found to be identical, confirming the signal from the accumulating lump of arctiid barcodes (Appendix SX). A glance at the previously unassociated male and female of *Dysschema perplexa* (Fig. 10), and its name, makes it obvious why the ACG inventory was very happy to find that they have the same barcodes. Six new species of ACG campoplegine ichneumonid parasitic wasps were about to be described (I.D.G.) when barcoding showed that they were but three new species; interestingly, all three have such similar barcodes that they make a single cluster of ichneumonid lumps in the NJ tree (Appendix SIV). This suggests that a separate genus might be appropriate for them, a genus based on their having very dimorphic sexes as well as other shared traits. Among the ACG-reared tachinids, there are many cases where classical morphological mysteries as to which males go with which females have been solved by rearing both sexes from the same individual caterpillar, or barcode matching males with females, or both.

Within-sex polymorphisms are also conveniently cleared up with barcode data. The arctiid *Calodesma maculifrons* has long been known as a different species from *Calodesma melanochroia* (Fig. 10), but their identical barcodes lent credence to the ACG inventory conclusion that they are but one species — both 'species' were reared on three different occasions from what appeared to be sibling groups of wild-caught caterpillars. The inventory searched long and hard for the male to match one of what turned out to be two female morphs of *Heraclides tolmides* (Papilionidae, see Tyler *et al.* 1994) (Fig. 10) — until it was discovered that the two female morphs had not only the same caterpillar and food plant, but also the same barcode.

10. Future of barcoded specimen extracts. The DNA extracts from ACG specimens are held under a Material Transfer Agreement by the Biodiversity Institute of Ontario (BIO)

and treated with the same care and long-term preservation, with anticipated genomic exploration as desired and appropriate, as are all other barcode samples at BOLD (see Ratnasingham & Hebert 2007). As indicated earlier, the voucher specimens from which these samples are derived are all deposited as permanent vouchers in the natural history collections appropriate to the particular taxon, as collaboratively determined between the inventory and BIO.

In closing

This integration of DNA barcoding with the ongoing ACG inventory is very much a work in progress. It is hard to think of a new technology that is not feared at the time of its introduction (D.H.J. remembers when hand-held calculators were banned from classrooms because 'they would impair our ability to think'). The experience of the ACG inventory is that DNA barcoding is on the way to becoming an essential tool for field identification, and for disclosing another layer of biodiversity beyond that which is revealed by traditional methods (and see recent examples in other organisms: starfish, Vogler *et al.* 2008; earthworms, King *et al.* 2008; spiders, Bond & Stockman 2008; killer whales, LeDuc *et al.* 2008; giraffes, Brown *et al.* 2007). Simultaneously, barcoding reaffirms and clarifies the power of comparative morphology in taxonomy, both for identifying and delimiting species. Barcoding and morphological approaches legitimize each other, just as genes and morphology legitimize each other in phylogenetic studies.

By adding barcodes and their collateral to a diversity of morphology, ecology, and micro-ecogeographical collateral, the inventory becomes more certain, more exploratory, more revealing and perhaps most important of all, more possible to pass to future generations. Such a passing on is critical for both scientific understanding and for increasing the chance that society will permit wild biodiversity to survive. The day is coming when there will be a hand-held barcorder (Janzen 2004b; Janzen *et al.* 2005) for everyone to use as their linkage between the wild world and what humanity knows about it. The simple fact is that just 5 years of trial and error with DNA barcoding has enormously improved the ongoing ACG inventory of moths, butterflies, and parasitic flies and wasps. Simultaneously, it has exposed a horrendous taxonomic and nomenclatorial problem of morphologically defined species that contain apparently distinct phylogenetic lineages of uncertain relationship to the ancient type specimens for those species.

Acknowledgements

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

Aliabadian M, Kaboli M, Nijman V, Vences, M (2009) Molecular identification of birds: performance of distance-based DNA barcoding in three genes to delimit parapatric species. *PLOS One*, **4**(1): e4119. doc10.1371/journal.pone.0004119.

Allyn AC, Downey JC (1977) Observations on male U-V reflectance and scale ultrastructure in *Phoebis* (Pieridae). *Bulletin of the Allyn Museum*, **42**, 1–20.

Bond JE, Stockman AK (2008) An integrative method for delimiting cohesion species: finding the population-species interface in a group of California trapdoor spiders with extreme genetic divergence and geographic structuring. *Systematic Biology*, **57**(4), 628–646.

Brown FM (1929) A revision of the genus *Phoebis* (Lepidoptera). *American Museum Novitates*, **368**, 1–22.

Brown KS (1988) Book review. The butterflies of Costa Rica and their natural history. *Journal of the Lepidopterists' Society*, **42**, 240–244.

Brown DM, Brenneman RA, Koepfli K-P *et al.* (2007) Extensive population genetic structure in the giraffe. *Biomed Central Biology*, **5**, 57.

Burns JM (1996) Genitalia and the proper genus: *Codatractus* gets *mysie* and *uvydixa* — in a compact *cyda* group — as well as a *hyster*ectomy, while *Cephise* gets part of *Polythrix* (Hesperiidae: Pyrginae). *Journal of the Lepidopterists' Society*, **50**, 173–216.

Burns JM, Janzen DH (2005) Pan-Neotropical genus *Venada* (Hesperiidae: Pyrginae) is not monotypic: Four new species occur on one volcano in the Area de Conservacion Guanacaste, Costa Rica. *Journal of the Lepidopterists' Society*, **59**, 19–34.

Burns JM, Janzen DH, Hajibabaei M, Hallwachs W, Hebert PDN (2007) DNA barcodes of closely related (but morphologically and ecologically distinct) species of skipper butterflies (Hesperiidae) can differ by only one to three nucleotides. *Journal of the Lepidopterists' Society*, **61**, 138–153.

Burns JM, Janzen DH, Hajibabaei M, Hallwachs W, Hebert PDN (2008) DNA barcodes and cryptic species of skipper butterflies in the genus *Perichares* in Area de Conservacion Guanacaste, Costa Rica. *Proceedings of the National Academy of Sciences, USA*, **105**, 6350–6355.

CBOL (2008) http://barcoding.si.edu.Newsletter #2.

D'Abrera B (1986) *Sphingidae Mundi. Hawk Moths of the World*. E. W. Classey, Faringdon, Oxon, UK.

DeVries PJ (1987) The butterflies of Costa Rica and their natural history. *Papilionidae, Pieridae, Nymphalidae*. Princeton University Press, Princeton, New Jersey.

Ferri E, Barbuto M, Bain O *et al.* (2009) Integrated taxonomy. traditional approach and DNA barcoding for the identification of filaroid worms and related parasites (Nematoda). *Frontiers in Zoology*, **6**, 1–27.

Ficetola GF, Miaud C, Pompanon F, Taberlet P (2008) Species detection using environmental DNA from water samples. *Biology Letters*, **4**, 423–425.

Gauld ID (1985) The phylogeny, classification and evolution of parasitic wasps of the subfamily Ophioninae (Ichneumonidae). *Bulletin of the British Museum (Natural History) (Entomology)*, **51**, 61–185.

Gauld ID, Janzen DH (2004) The systematics and biology of the Costa Rican species of parasitic wasps in the *Thyreodon* genus-group (Hymenoptera: Ichneumonidae). *Zoological Journal of the Linnean Society*, **141**, 297–351.

Hajibabaei M, Singer GAC, Hebert PDN, Hickey DA (2007) DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends in Genetics* **23**, 167–172.

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.

Haxaire J (1996) Les genres *Pachygonidia* Fletcher, *Nyceryx* Boisduval et *Perigonia* Herrich-Schaffer en Guyane francaise (Lepidoptera, Sphingidae). *Lambillionea*, **96**, 342–350.

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

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, USA*, **101**, 14812–14817.

Hulcr J, Miller SE, Setliff GP *et al.* (2007) DNA barcoding confirms polyphagy in a generalist moth, *Hormon mermerodes* (Lepidoptera: Tortricidae). *Molecular Ecology Notes*, **7**, 549–557.

Janzen DH (1988a) Guanacaste National Park: tropical ecological and biocultural restoration. In. *Rehabilitating Damaged Ecosystems* (ed. JJ Cairns), pp. 143–192. Vol. II, CRC Press, Boca Raton, Florida.

Janzen DH (1988b) The migrant moths of Guanacaste. *Orion Nature Quarterly*, **7**, 38–41.

Janzen DH (1993) Taxonomy: universal and essential infrastructure for development and management of tropical wildland biodiversity. In: *Proceedings of the Norway/UNEP Expert Conference on Biodiversity, Trondheim, Norway* (eds Sandlund OT, Schei PJ), pp. 100–113. NINA, Trondheim, Norway.

Janzen DH (2000) Costa Rica's Area de Conservación Guanacaste: a long march to survival through non-damaging biodevelopment. *Biodiversity*, **1**(2), 7–20.

Janzen DH (2002) Tropical dry forest: Area de Conservación Guanacaste, northwestern Costa Rica. In: *Handbook of Ecological Restoration, Vol. 2, Restoration in Practice* (eds Perrow MR, Davy AJ), pp. 559–583. Cambridge University Press, Cambridge, UK.

Janzen DH (2003) How polyphagous are Costa Rican dry forest saturniid caterpillars? In: *Arthropods of Tropical Forests. Spatiotemporal Dynamics and Resource Use in the Canopy* (eds Basset Y, Novotny V, Miller SE, Kitching RL), pp. 369–379. Cambridge University Press, Cambridge, UK.

Janzen DH (2004a) Setting up tropical biodiversity for conservation through non-damaging use: participation by parataxonomists. *Journal of Applied Ecology*, **41**, 181–187.

Janzen DH (2004b) Now is the time. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **359**, 731–732.

Janzen DH, Hajibabaei M, Burns JM, Hallwachs W, Remigio E, Hebert PDN (2005) Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**(1462), 1835–1846.

Janzen DH, Hallwachs W (2008) *Event-Based Database of Caterpillars, their Food Plants, and their Parasitoids in Area de Conservacion Guanacaste, Northwestern Costa Rica* (http://janzen.sas.upenn.edu).

Janzen DH, Walker AK, Whitfield JB, Delvare G, Gauld ID (2003) Host-specificity and hyperparasitoids of three new Costa Rican species of *Microplitis* Foerster (Hymenoptera: Braconidae: Microgastrinae), parasitoids of sphingid caterpillars. *Journal of Hymenoptera Research*, **12**(1), 42–76.

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**(1567), 639–648.

Kankare M, Stefanescu C, Van Houhuys S, Shaw MR (2005) Molecular phylogeny of *Cotesia* Cameron, 1891 (Insecta: Hymenoptera: Braconidae: Microgastrinae) parasitoids associated with Melitaeini butterflies (Insecta: Lepidoptera: Nymphalidae: Melitaeini). *Journal of the Linnean Society*, **86**, 45–65.

King RA, Tibble AL, Symondson WOc (2008) Opening a can of worms: unprecedented sympatric cryptic diversity within British lumbricid earthworms. *Molecular Ecology*, **17**, 4684–4698.

Kitching IJ, Cadiou J-M (2000) *Hawkmoths of the World: An Annotated and Illustrated Revisionary Checklist (Lepidoptera: Sphingidae)*. Cornell University Press, Ithaca, New York.

Kress WJ, Erickson Dl (2008) DNA barcoding—a windfall for tropical biology? *Biotropica*, **40**, 405–408.

Kuhlmann M, Else GR, Dawson A, Quicke DLj (2007) Molecular, biogeographical and phenological evidence for the existence of three western European sibling species in the *Colletes succinctus* group (Hymenoptera: Apidae). *Organisms, Diversity and Evolution*, **7**, 155–165.

Lamas G ed. (2004) *Atlas of Neotropical Lepidoptera. Checklist: Part 4A. Hesperoidea—Papilionoidea in Atlas Neotrop Lepid 5A*. Association for Tropical Lepidoptera, Gainesville, Florida.

LeDuc RG, Robertson KM, Pitman RL (2008) Mitochondrial sequence divergence among Antarctic killer whale ecotypes is consistent with multiple species. *Biology Letters*, **4**, 426–429.

Lemaire C (1988) *The Saturniidae of America. Ceratocampinae*. Museo Nacional de Costa Rica, San Jose, Costa Rica.

Miller JS, Janzen DH, Franclemont JG (1997) New species of *Euhapigioides*, new genus, and *Hapigiodes* in Hapigiini, new tribe, from Costa Rica, with notes on their life history and immatures (Lepidoptera: Notodontidae). *Tropical Lepidoptera*, **8**(2), 81–99.

Ratnasingham S, Hebert PDN (2007) BOLD: the barcode of life data system (www.barcodinglife.org). *Molecular Ecology Notes*, **7**, 355–364.

Rutowski RL, Macedonia JM (2008) Limits on the wavelengths represented in ultraviolet images of Lepidoptera. *Journal of the Lepidopterists' Society*, **62**, 133–137.

Schauff ME, Janzen DH (2001) Taxonomy and ecology of Costa Rican *Euplectrus* (Hymenoptera: Eulophidae), parasitoids of caterpillars (Lepidoptera). *Journal of Hymenoptera Research*, **10**(2), 181–230.

Silberglied RE, Taylor OR (1973) Ultraviolet differences between the sulfur butterflies *Colias eurytheme* and *C. philodice*, and a possible isolating mechanism. *Nature*, **241**, 406–408.

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.

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

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, USA*, **103**, 3657–3662.

Solis MA, Sullivan JJ, Janzen DH (2003) *Cromarcha stroudagnesia*, a new chrysaugine species boring in shoots of *Tabebuia ochracea* (Bignoniaceae) in a Costa Rica dry forest (Lepidoptera: Pyralidae: Chrysauginae). *Tropical Lepidoptera*, **11**, 33–39.

Standley PC (1922) Trees and shrubs of Mexico. Contributions from the United States National Herbarium. *Smithsonian Institution*, **23**, 171–515.

Stoeckle MY, Hebert PDN (2008) *Barcode of Life*. Scientific American, October 2008: 82–88.

Stork NE, Grimbacher PS, Storey R, Oberprieler RG, Reid C, Slipinski SA (2008) What determines whether a species of insect is described? Evidence from a study of tropical forest beetles. *Insect Conservation and Diversity*, **1**, 114–119.

Tyler HA, Brown KS Jr, Wilson KH (1994) Swallowtail butterflies of the Americas. *A Study in Biological Dynamics, Ecological Diversity, Biosystematics, and Conservation*. Gainesville Scientific Publishers, Gainesville, Florida.

Valentini A, Pompanon F, Taberlet P (2009) DNA barcoding for ecologists. *Tree*, **24**, 110–117.

Vogler C, Benzie J, Lessios H, Barber P, Worheide G (2008) A threat to coral reefs multiplied? Four species of crown-of-thorns starfish. *Biology Letters*, **4**, 696–699.

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

Willmott KR (2003) *The Genus Adelpha: Its Systematics, Biology and Biogeography (Lepidoptera: Nymphalidae: Limenitidini)*. Gainesville Scientific Publishers, Gainesville, Florida.

Wolf G (2008) Barcode of life. *Wired*, **16**(10) (October), 200–212.

Supporting information

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

Appendix SI NJ tree of Noctuidae

Appendix SII NJ tree of Hesperiidae

Appendix SIII NJ tree of Notodontidae

Appendix SIV NJ tree of Campopleginae, Ichneumonidae

Appendix SV NJ tree of Pieridae

Appendix SVI NJ tree of Nymphalidae

Appendix SVII NJ tree of Sphingidae

Appendix SVIII NJ tree of Saturniidae

Appendix SIX NJ tree of Papilionidae

Appendix SX NJ tree of Arctiidae

Appendix SXI Identification and voucher codes of the 28 specimens in Figs 4–5

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.