Chapter 3

Challenges and solutions for planning and implementing large-scale biotic inventories

by

Maurice Leponce
Royal Belgian Institute of Natural Sciences
29 rue Vautier, 1000 Brussels, Belgium
Email: maurice.leponce@naturalsciences.be

Christopher Meyer
National Museum of Natural History
PO Box 37012, Washington DC 20013, USA
Email: meyerc@si.edu

Christoph L. Häuser
Museum für Naturkunde Berlin
Invalidenstr. 43, 10115 Berlin
Email: Christoph.Haeuser@mfn-berlin.de

Philippe Bouchet
Muséum National d'Histoire Naturelle
55 rue Buffon, 75005 Paris, France
Email : pbouchet@mnhn.fr

Jacques Delabie
Centro de Pesquisa do Cacau, Convênio UECS-CEPLAC
Caixa Postal 7, 45600-000 Itabuna, Bahia, Brasil
Email : jacques.delabie@gmail.com

Lee Weigt
National Museum of Natural History
PO Box 37012, Washington DC 20013, USA
Email: weigtl@si.edu

Yves Basset
Smithsonian Tropical Research Institute
Apartado 0843-03092, Balboa, Ancon, Panama City, Republic of Panama
Email: bassety@si.edu

Abstract

To date, there is still no complete or near complete information on the total biodiversity of any species-rich ecosystem around the world, even in protected areas. The benefits from biodiversity resources and healthy ecosystems are best garnered if those species and interactions are well known. Large-scale inventories can maximize the biodiversity information collected through the coordinated effort of a multidisciplinary team. Large-scale inventories may produce an overall picture of highly complex ecosystems and may be instrumental for conservation and management decisions. The taxonomic coverage of the survey may vary from all taxa present in an area (All Taxa Biodiversity Inventory, ATBI) to a selected range of them. Comprehensive biodiversity inventories basically face four kinds of challenge. First, biological challenges, as species distributions are heterogeneous in space and time. Representative results may thus only be achieved with adequate spatio-temporal replicates. Second, methodological challenges, since any sampling method provides a biased image of species composition and abundance. The use of complementary collecting methods helps to circumvent this problem. Third, taxonomical challenges, as large inventories generate an impressive amount of material to process and identify. To avoid work overload of expert taxonomists the material should be pre-processed by assistants (students, amateurs, parataxonomists, volunteers) supervised by professionals. Fourth, planning and implementation challenges, since security and legal issues, coordination of collection and processing of material, centralization of data, and follow-up of the project may not be straightforward. An ideal implementation requires an organizational structure composed of coordinators, advisors, workgroups and external partners. Comprehensive inventories typically span over several years. To keep the motivation of participants and of stakeholders the project output should include fast deliverables in addition to long-term research. Finally, the value and complementarity of large-scale inventories in terms of global biodiversity coverage and of scientific investigations may be increased by incorporating them into global networks of permanent sites.

Keywords: All Taxa Biodiversity Inventory, sampling design, project coordination, DNA barcoding.

1. Taxonomic, spatial and temporal extent of large-scale inventories

Large-scale biotic inventories differ in their size and ambition. The major factors differentiating All Taxa Biodiversity Inventories (thereinafter referred to as ATBIs) can be defined by three axes: taxonomic scope, geographic extent, and sampling frequency. Taxonomic coverage can vary from all taxa present in an area to a selected range of them, often limited by specimen size or phyletic affinity. Geographic coverage may range from an entire country, down to an island or park scale. Finally, temporal coverage may vary from an inventory at a single time slice to annual or multi-year surveys.

To date, there is still no complete or near complete information on the total biodiversity of any species-rich ecosystem around the world, even in protected areas. Microcosms, caves and other self-contained and relatively species-poor ecosystems may represent exceptions (e.g Small, 1998). In the past, the largest inventory carried out may well have been the monumental collections needed for the encyclopedia "Biologia Centrali Americana" (DuCane Godman & Salvin, 1879-1915, free digital edition available on the web at http://www.sil.si.edu/digitalcollections/bca/). During this inventory, many collectors were employed specifically to accumulate material from Mexico and Central America. Over a 36 year period, this work described over 50,000 species of animals and plants, one third of which were new.

The first ATBI (Janzen & Hallwachs, 1994) was initially planned by Daniel Janzen for the Area de Conservación Guanacaste in Costa Rica, but for financial and political reasons this endeavour changed into a survey focused on Lepidoptera, their parasites and gut micro-organisms (Janzen, 1988; Gámez *et al.*, 1997; Sharkey, 2001; White & Langdon, 2006). The concept was then applied to a temperate area in the Great Smoky Mountains National Park, USA (Nichols & Langdon, 2007).

The goal of an ATBI is to collect and disseminate useful data on all species collected in a specific area (Nichols & Langdon, 2007). In this concept, "all species" mean in fact "as many as practical", and "useful data" refer to the collection of as much collateral information as possible on species' relative abundance, distribution, natural history and ecology. Such huge data collection effort is in principle concentrated over a limited amount of time. White & Langdon (2006) calculated that a comprehensive inventory in the Great Smoky Mountain National Park would take about 150 years without an ATBI approach. Janzen & Hallwachs (1994) initially recommended a five-year period to demonstrate the desirability and usefulness of the ATBI concept without losing the momentum, the motivation of participants and, possibly, many species through local extinctions.

Currently there are only a few ongoing ATBIs. The Great Smoky Mountain National Park ATBI (referred hereafter to as Smokies ATBI), which was officially initiated in 1998, covers an area over 2000 km². Launched in 2002, the Swedish Taxonomy Initiative (STI) aims to inventory all of Sweden's multicellular organisms, approximately 50,000 species, within 20 years (Ronquist &

Gärdenfors, 2003). Between 2006 and 2011, the Moorea Biocode Project (MBP) of French Polynesia will construct a vouchered library of genetic markers and physical identifiers for every non-microbe species on the island of Moorea (134 km²), including marine, freshwater and terrestrial habitats (Check, 2006). Starting in 2007, the European Distributed Institute of Taxonomy (EDIT) has identified a series of potential ATBIs in both temperate and tropical national parks (for detailed explanations see chapter 2).

Many large-scale inventories focus on a selected range of taxa or habitats rather than all present in a specific area. This is the case of local or global projects, especially when centred around research stations in the tropics (*e.g.* Arthropod of La Selva 1991-2005: Longino & Colwell, 1997; Manaus, Reserva Ducke: Adis *et al.*, 1998; Magnusson *et al.*, 2005). These efforts can be used to spearhead more comprehensive inventories once proof of concept is demonstrated. Additionally, large naturalist expeditions such as the Royal Geographical Society expedition of 1977-1978 in Sarawak, the Royal Entomological Society expedition of 1985 in Sulawesi (project Wallace: Knight & Holloway, 1990) or EDIT's SANTO2006 project in Vanuatu (Hanbury-Tenison & Jermy, 1979; Bouchet *et al.*, 2009) all involved more than 100 scientists, many scientific programmes, lasted several months, and included a large range of taxa. Other projects such as IBISCA-Panama (Investigating the Biodiversity of Soil and Canopy Arthropods) put a strong emphasis on the collaboration between different research teams coordinated to answer common scientific questions (Basset *et al.*, 2007).

Finally large-scale inventories of selected taxa should ideally be coupled with long-term monitoring programmes. Examples of suitable locations for this long-term task include the networks of the Smithsonian Institution Global Earth Observatories (www.sigeo.si.edu/), the National Ecological Observatory Network (www. neoninc.org), long-term ecological research stations (www.lternet.edu) or Conservation International's TEAM initiative (www.teaminitiative.org/). A crucial advantage of global networks includes the collection of biodiversity information using standardized methods, which allows between-site comparisons.

2. Challenges

Large-scale biodiversity inventories are challenging in many aspects:

- Species distribution is heterogeneous in space and time. Hence, data collected during studies restricted in space and/or time may not be representative of local biodiversity. S*olution: replicate your collection*.
- Collecting, identifying and processing specimens and analyzing the information require a wide expertise and substantial coordination between project participants*. Solution: plan carefully.*
- Processing of the material collected is very time-consuming (*i.e*., costly), particularly when taxonomic coverage is wide and includes small organisms and species-rich groups (Lawton *et al.*, 1998).

- Taxonomic coverage of the project is unavoidably biased towards well-studied taxa or, at least, “non-orphan” taxa (*i.e.*, currently many species groups lack experts and this situation will get worst in the future).
- Taxonomy experts represent a scarce resource and are continuously overloaded with work (taxonomic impediment).
- The motivation of participants needs to remain focused on the project for a substantial time, typically a few years.
- Pristine habitats have virtually disappeared. The remaining undisturbed or little known habitats are generally difficult to access or are threatened by human activities. The number of suitable natural sites is therefore restricted or may be costly to access.
- Collecting and export permits (for the purpose of taxonomic studies) may be difficult to obtain for certain countries/locations.
- The colossal input and work involved is likely to slow down scientific output. Yet the project may need to rapidly demonstrate its scientific interest and deliver scientific products.

3. Importance and implementation of large-scale inventories

3.1. Importance

The benefits from biodiversity resources and healthy ecosystems are best garnered if those species and interactions are well known. Moreover, conservation decisions and the success of those efforts can only be measured if we have a baseline of what exists. Well-integrated, large-scale inventories constitute a cost-effective way to study our biodiversity resources through coordinated collaboration between researchers. Numerous benefits can be expected from these endeavours (Janzen & Hallwachs, 1994; Sharkey, 2001; White & Langdon, 2006; Nichols & Langdon, 2007):

- **Advances in fundamental science.** The identification of species, the study of their morphological and genetic variability, and the discovery of species new to science or new to the study area allow advances in taxonomy, systematics and biogeography. Large-scale inventory sites where many species are identified and where environmental conditions are well known are also ideal locations for studying species ecological interactions (including food webs) and the functioning of whole ecosystems. Finally, new scientific approaches can arise from the confluence of ideas and methods of the various specialists involved in the collective project.
- **Advances in applied science.** Reference checklists and maps of defined areas can be used as a baseline for conservation, management and monitoring. Inventories enable assessment of the type and level of threat to which species or habitats are exposed and to update red lists. Inventories allow detection of invasive species and documentation of natural or human disturbances (habitat modification, fragmentation and isolation, or pollutants).

They also provide fundamental information that is necessary for land management, especially for protected areas. For example, an analysis of presence/absence maps and GIS layers can help determine which biotic and abiotic conditions rare or sensitive species depend upon. This information can also forecast the impact of practices such as grazing, pest control, and road or corridor building (White & Langdon, 2006). Geographic analysis of multiple species distributions can be used for protecting sensitive sites or for locating monitoring activities at the most appropriate sites. Finally, large inventories constitute a baseline for monitoring. True decreases or increases in biodiversity can be distinguished from natural variations, and the cascade effects of the disappearance of ecologically important species in the ecosystem can be studied.

- **Education.** Large-scale inventories generate a large amount of information which is useful for various segments of the population: specialists, amateurs, general public, schools, ecotourists, artists, etc. (Sharkey, 2001; Hilten *et al.*, 2006). This is especially true if the data collected are made quickly available to the public through webpages and if voucher specimens of the species collected are centralised at a single location. Ideally specialists should benefit from the tools and infrastructure supplied by the project to build interactive keys and establish a library of photos, videos, sounds or other media including DNA sequences. These electronic tools are of great help to the amateur naturalist and the general public for identifying specimens and can be used to produce field guides of local fauna and flora. Science education programs can be articulated around ATBIs and proved to be very successful in the Great Smoky Mountains National Park (Hilten *et al.*, 2006). A website (www.smokymountainseft.org) offers downloadable activities, video and interactive games to explore the biodiversity of the Park.
- **Other utilities for economy and society.** As stressed by Janzen Hallwachs (1994) “the basic goal of an ATBI is to prepare a large body of biodiversity for non-damaging use by society”. The prospecting of genes, chemicals, structures and behaviours are useful for the progress of science, art and industry. Technology can also learn from solutions found in nature for a large range of problems (*e.g.* biomechanics, biomimicry). Living samples collected during inventories can supply banks of biological material (seed, sperm, tissue), biological control centres, zoos or botanical gardens. Large-scale inventories can also stimulate local development involving the sustainable use of biodiversity resources through ecotourism, bioprospecting and sound ecosystem management.

3.2. Implementation

Large-scale inventories are characterized by advanced coordination between researchers, concentrated research effort in reference sites, wide taxonomic coverage and a diverse spectrum of biological information collected.

The choice of the reference site will depend on the scientific questions targeted, the infrastructure available and the prior commitments in conservation, research and monitoring. For ATBIs, study sites should be protected areas with a

guarantee of long-term protection and of access for inventory activities. In this perspective, the long term survival of the biodiversity contained in the site increases if the site is large and includes climatic or altitudinal gradients (buffer against climate change) and if migrations to or from surrounding habitats are made possible by the presence of buffer zones and stepping stones (Janzen & Hallwachs, 1994). Large areas also allow more replication, less impact of inventory activities and inclusion of disturbed portions of the habitat representative of various degrees of restoration or regeneration (Janzen & Hallwachs, 1994). The inclusion of anthropogenic habitats is pertinent to evaluate the impact on biodiversity of improved management or regulation (e.g. new pollutant emission rules, access restrictions, catch-limits in marine protected areas, etc.).

The choice of taxonomic coverage will depend of the aim of the project and of the taxonomic expertise available. The data collected during large-scale inventories do not simply consist of a species list. Additional information about species abundance, spatio-temporal distribution, environmental conditions and life history are needed for better predictive modelling of species diversity, distribution and response to environmental changes. Estimates of population size and rarity are necessary to appreciate the endangered status of species. Because the life cycle, distribution and abundance of organisms are tied to climate, weather data should be collected during the general inventory. Depending on the organisms studied, other useful environmental measures include: soil quality, water quality, light intensity, etc. Whenever possible environmental data are collected automatically with recording instruments. The collection of these environmental variables leads to improved predictive models, directs additional sampling and allows for further testing and refinement of those models. Any large-scale project must also disseminate knowledge and experience to a wide audience and incorporate an education and communication plan in addition to the science plan (White *et al.*, 2000; Hilten *et al.*, 2006; Parker & Bernard, 2006). Practical issues linked to the planning and logistics of large-scale inventories are developed in the next sections.

4. Management

The administrative structure depends on the size of the project but is basically composed of:

- **coordinators.** A project leader and assistants are essential. The largest projects may require a directorate with a director supervised by a National Commission (Janzen & Hallwachs, 1994). Coordinators support and integrate the work of all project participants in order to achieve a common goal and vision and ensure the circulation of information between them.
- **workgroups.** Participants in large inventories can be experts in various disciplines: field collection, taxonomy, ecology, molecular techniques, data management, statistics, etc. These experts are best organized in workgroups headed by leaders. These leaders are responsible for a particular taxa, method or task. They supervise the work of the other members of the group, train less experienced participants and are responsible for the feedback of

information to the project coordinators. Coordinators minimize redundancy in data collection, overrepresentation of popular taxa or methods, and information gaps. They also plan the actions of the workgroups. Some workgroups depend of the results of others (e.g. a botanical survey is often a preliminary to an entomological survey, a vertebrate survey comes before a survey of their parasites, pathogens or symbionts) (Janzen & Hallwachs, 1994). Taxonomic Working Groups (called "TWiGs") can be organized according to their ease of study, collection methods, or expeditions. Taxonomic coverage basically depends on the actual knowledge and expertise available for the groups encountered, their ease of identification and of collection, and species richness and abundance (Janzen & Hallwachs, 1994; Sharkey, 2001). While some ecologically important organisms are easy to inventory (*e.g.* plants, social insects, etc.) others are unlikely to be inventoried to the species level (*e.g.* rotifers, wild plant viruses, etc.). However, DNA barcoding provides new opportunities for discriminating notoriously difficult groups or cryptic species.

- **international advisory committee.** This committee discusses project planning, monitors its progress and makes recommendations to its coordinators. Success should be measured according to established benchmarks. Its competence can be related to science, education or development.
- **partners and companion structures.** Scientific partners of the project can be universities, museums, research institutes, or park administrations. Some projects are too large to be managed by any one of these partners. In this case they can be managed by a NGO (*e.g.* Conservation International for the TEAM initiative, Pro-Natura International for IBISCA projects), a private non-profit organization (*e.g.* Discover Life In America for the Great Smoky Mountains ATBI) or an international public project (*e.g*. EDIT's ATBIs) (White & Langdon, 2006). These companion structures administer and coordinate the inventories and develop resources and partnerships. Sponsors can be public or private partners. Often it is helpful to include a consortium of stakeholders or a set of local partners that have a vested interest in the heritage of the region and a sense of long-term stewardship.

5. Planning

5.1. Duration and budget

The general goals of the project must be achievable in a reasonable amount of time and include fast deliverables to maintain the motivation of stakeholders. Examples of such deliverables include dynamically updated websites, frequent progress reports, assistance to management and conservation decisions, scientific publications, and identification guides for the public.

A full-scale ATBI is a major effort that requires significant but still reasonable resources compared to the budget of the human genome project or of the 2010 football world cup (both around 3 US$ billion). To support the Smokies ATBI, approximately US$ 1.8 million has been received by individual scientists in

several grants. In addition, approximately US$ 120,000 per year has been received from local sources, which was “seed money” so researchers could leverage an additional approximately US$400,000 per year in services and sometimes funding. So far, in 11 years 6400 species have been documented in the park, 15% of them being species new to science. The Smokies ATBI database currently includes 300,000 geographic records essentials to mapping distributions and understanding ecological connections. The Moorea Biocode Project is supported by a US$5.2 million grant and EDIT’s SANTO2006 budget was € 1.2 million. However, these values do not reflect the full costs of such large-scale inventories. The salaries of the participating scientists are not included (they are covered by their institutions) and many of the costs for post-collection events such as processing of the material, databasing, storage of collections, taxonomic studies (including visits to museum’s collections), are assumed by the holding institutions. As an illustration of this, the total cost of the 2009-2010 Madagascar/Mozambique inventory organized by the French National Museum of Natural History and Pro-Natura International is ca. 4 million € of which 2,5 million € is in cash and the remainder taking the form of in-kind contributions. Considering the huge amount of biological material collected a substantial budget for these post-collection tasks must be secured to assure success. This represents a guarantee that the project will deliver a minimum scientific output within a reasonable time frame. It may also be useful to include in the budget “seed grants” that may facilitate the access to other sources of funding for participants (White & Langdon, 2006).

5.2. Sampling design

The task of documenting the diversity and distribution of species in a given area faces three kinds of challenges: biological, methodological and taxonomical.

5.2.1. Biological challenges

The distribution of species is heterogeneous in space and time (Fig. 1). The scale of this heterogeneity depends of the organism studied. To obtain a representative biodiversity inventory, spatial and temporal replicates are therefore necessary.

Spatial replicates are best conducted through a stratified sampling encompassing various spatial scales, with replications at each scale (Table 1, iBOL Barcoding Biotas Working Group, 2009). The rationale behind this approach is that the distribution of organisms is often related to the distribution of their resources (*i.e.* food, nesting sites). Stratified sampling allows measuring the diversity partitioned within the habitat studied. Vertical sampling is of particular importance in multi-layered habitats such as forest canopies or soils (Basset *et al.*, 2003a; André *et al.*, 2002). For example, a study in Gabon indicated that, for a particular time period, forest strata explained a higher fraction of variance in the distribution of species of insect herbivores than location *per se* in the forest or diel activity (73%, 19% and 8% of the variance explained, respectively; Basset *et al.*, 2001).

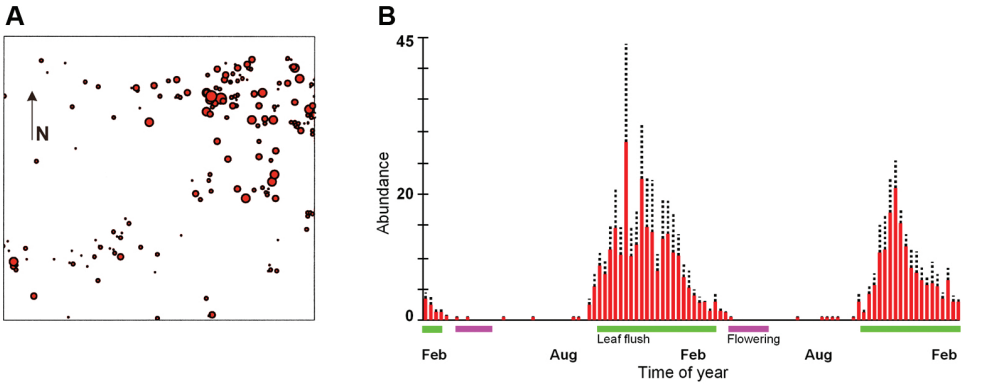

Fig. 1. Illustration of the spatio-temporal heterogeneity of species distribution. A. Variability of the spatial distribution of a highly generalist leafhopper, *Soosiulus fabricii* Metcalf (Hemiptera: Cicadellidae), in a plot of 1 km^2 in Guyana. The size of the bubbles is proportional to the mean abundance of specimens collected at each station. B. Variability of the abundance of an insect group associated to the resource availability in a tropical rainforest. Weekly number of Chrysomelidae (Coleoptera, herbivore) collected with respect to host leaf flush and flowering (solid bars represent means and broken lines standard errors). (Source: Basset, 1991, 2000).

Temporal replicates should be conducted at least during one year to document the seasonal variation of species distribution, relative abundance, and all life stages of the organisms present. If time is restricted, the priority may be to collect during a period of high biological activity (but background information is required to achieve this). Nycthemeral cycles should be particularly taken into account when sampling organisms whose activity is much affected by light or temperature.

A structured sampling approach (such as the stratified sampling presented in the previous paragraph) allows quantification of biodiversity and therefore statistical comparisons among sampling units, sites, or seasons. However not all taxa are reliably sampled by sampling devices (e.g. traps) used in a structured approach. Therefore a complementary approach is traditional sampling conducted by experienced taxonomists who rely on tacit knowledge of their target taxa to effectively locate them. “Bio-blitzes” that bring together large numbers of experts and volunteers are sometimes organized during a short period of time to collect a large amount of specimens of the target taxa (Nichols & Langdon, 2007). It should be noted that some sampling protocols are a mixture of the structured and traditional approaches (*e.g.* termites which are collected by visual search along transects: see Jones & Eggleton, 2000; Roisin & Leponce, 2004). In general comparison of the results obtained from the traditional and structured sampling approaches gives some indications on the completeness of the inventory (Nichols & Langdon, 2007).

Habitats	Vertical strata	Microhabitats
Forest	Canopy	Leaves
		Flowers & fruits
		Bark
		Epiphytes
		…
	Understorey	cf. canopy
	Ground surface	Leaf litter
		Dead wood
		…
	Soil	Humus
		Roots
		…
Ecotone forest/grassland	cf. forest	
Grassland	cf. forest ground and soil	

Table 1. Example of stratified sampling in a hypothetical simplified landscape composed of two terrestrial habitats: a forest and a grassland. The number of subdivisions is non exhaustive and depends of both the habitat characteristics and the type of organisms targeted. For example the microhabitat scale presented here is relevant for arthropods but not for plants.

The identification of immature stages of animals or of plants at a period of the year during which they do not show any useful characteristics (flowers, fruits, leaves) is often problematic. DNA barcoding techniques are becoming increasingly efficient and affordable to solve this problem (Janzen *et al.*, 2005; Hajibabaei *et al.*, 2005; Kress *et al.*, 2005; Schlick-Steiner *et al.*, 2010).

5.2.2. Methodological challenges

In many cases inventories will only collect a fraction of the species present in the landscape because of problems of catchability. Some species are difficult to collect because they are geographically, temporarily or even methodologically rare (Novotny & Basset, 2000; Longino *et al.*, 2002; Novotny *et al.*, 2007). Some habitats such as the canopy or the soil are notoriously difficult to sample and require specialized techniques (André *et al.*, 2002; Basset *et al.*, 2003b; Basset *et al.*, 2007). In practice this results in incomplete surveys and biased samples due to undersampling (Coddington *et al.*, 2009), two common traits of any ATBI. Sampling protocols must be developed and adjusted to mitigate these effects. Completeness and bias of the survey can be easily evaluated by analyzing the data matrix (taxa by sample) with a popular freeware called EstimateS (Colwell, 1994).

- **Evaluation of sampling completeness.** Assessing sampling completeness during data collection helps to assess the cost-effectiveness of the inventory and to decide when to stop collecting. Sampling completeness can be evaluated by calculating rarefaction curves plotting the number of species that are statistically expected to be found after collecting a given number of samples (or individuals). Sample-based curves are convenient to assess the number of samples required to reach a given level of inventory completeness (Fig. 2A). A common problem encountered with species rich taxa is that the rarefaction curve does not reach a plateau even after a considerable sampling effort. A useful tip to evaluate if the rarefaction curves approaches a plateau is to use a logarithmic scale for the abscissa and to see if the number of rare species (singletons) decrease at the end of the inventory (Fig. 2B) (Longino, 2000; Longino *et al.*, 2002). Individual-based rarefaction curves are suitable to compare taxon richness among assemblages (Gotelli &Colwell, 2001). When the sampling is incomplete, various techniques exist to estimate the total number of species in the assemblage: parametric, non-parametric and curve-fitting methods (Colwell & Coddington, 1994; Chazdon *et al.*, 1998; Longino *et al.*, 2002; Walther & Moore, 2005). A simple method such as the non-parametric Chao1 or Chao2 is useful to estimate the total species richness (Fig. 2A). These estimators are directly calculated with EstimateS, are conservative (give a minimal value) and can be coupled with the calculation of the number of samples needed to obtain the total number of species that they predict (Chao *et al.*, 2009).
- **Evaluation of sampling bias.** Typically a collection method collects only a fraction of the species present in the assemblage (Fig. 2C) and gives a biased image of the true relative occurrence/abundance of species (Fig. 2D). A solution to mitigate these effects is to multiply the collection methods and focus on the most effective, simple, cheap and complementary techniques.

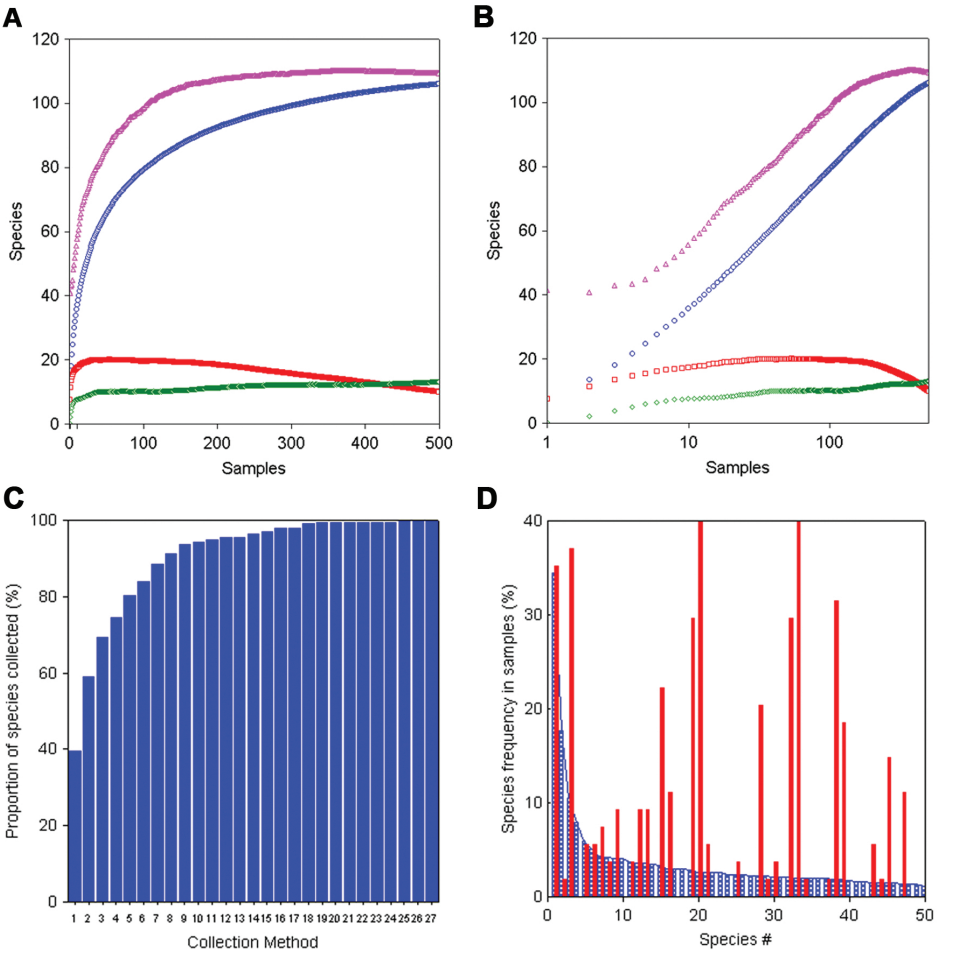

Fig. 2. Evaluation of inventory completeness of a species assemblage (A and B), of efficiency – in terms of fraction of species present collected- (C) and representativeness – in terms of species relative frequency- (D) of sampling methods used. A. Sample-based rarefaction curve *(i.e.* randomized species accumulation curve) allowing to assess sampling completeness. If the survey approaches near-completion the curve of singletons (*i.e.* rare species, represented in the sampling by a single individual, red curve) decreases and the rarefaction curve tends to reach a plateau. Associated with the decrease of singletons, the number of doubletons (*i.e.* rare species, represented by only 2 individuals, green curve) increases. B. These trends are more visible when a logarithmic scale is used for the abscissa. C. Cumulative proportion of species collected by more or less complementary sampling methods illustrating the fact that each method collects only a fraction of the local assemblage. For example method no. 1 collects 39% of the species present and 60% when combined with method no. 2. D. Comparison of the best approximation of each true species frequency (value ranging from 0.1 to 100.0%, calculated on the basis of the 27 collection methods from Fig. 2C, blue stripped bar) with the value obtained by a single method (red bars) (first 50 samples ranked by decreasing true frequency shown). For example, method no. 1 seems to give a representative value of the frequency of the most common species (no. 1), but vastly underestimates the frequency of species no. 2. (Datasets presented in Delabie *et al.*, 2000, 2007).

5.2.3. Taxonomical challenges

Large inventories generate an impressive amount of material to identify. At the same time the number of professional taxonomists working in museums and universities is declining (Hopkins & Freckleton, 2002; Godfray, 2002; Miller *et al.*, 2004; Leather, 2009). A solution to relieve the work of the expert taxonomist is to rely on assistants who can sort, prepare and morphosort specimens. Specimens collected by selective methods (*e.g.* plants collected by botanists) are directly chosen in the field by the dedicated workgroup(s). The pre-processing of the material collected by non-selective methods (*e.g.* mass collection with entomological traps) is best conducted on site when workgroup leaders are all present during a collection episode (see 5.4.7). This allows them to supervise directly the sorting and pre-identification of the material to higher taxa levels and its dispatching to the appropriate taxonomic workgroup (Fig. 3).

Four categories of assistants can be distinguished: biology students, amateur taxonomists, parataxonomists and volunteers. Depending of the circumstances the inventory can employ one or several of these categories. Assistants must be trained, supervised and rewarded for their activity.

Biology students can find opportunities to gain professional experience and establish contact with a large network of professionals. Grants for a master or PhD thesis will be more easily obtained if ecological or evolutionary hypotheses are tested in addition to the purely taxonomic work. Enthusiastic secondary level students can also find an opportunity to have a work experience.

Experienced **amateur taxonomists** constitute an important workforce in temperate latitudes. They are highly motivated. Usually they expect doubles of the specimens for their private collection. Such situation requires that they subscribe to the general terms of involvement of the participants project (see 5.4.1). In some instances, a naturalist association can be a satellite structure supporting the inventory work. In this case a contract has to be signed between the association and the inventory project to guarantee the release of the data and specimens.

A **parataxonomist** is “a resident, field-based, biodiversity inventory specialist who is largely on-the-job trained out of the rural work force and makes a career of providing specimens and their natural history information to the taxasphere, and therefore to a multitude of users across society” (Janzen, 2004). The use of parataxonomists has proved to be very successful in a number of projects (Basset *et al.*, 2000, 2004; Janzen, 2004). Because they live near the study site, they are a potential source of natural history information and can easily be in charge of further field sampling and monitoring.

Volunteers are “citizen scientists” happy to collaborate on a scientific project and who have variable degrees of taxonomic expertise or interests. Other (non-taxonomic) skills they have can also be very valuable for the project (experience in databasing, web page development, illustration, photography, fund raising, administration, outreach, etc.). Volunteers have been a major assistance to the Smokies ATBI since it began (White & Langdon, 2006).

Fig. 3. Collection and processing of specimens. A. Collection of samples (here suspended soil in the tree canopy). B. Extraction of the fauna from the sample (Berlese-Tullgren apparatus). C. An assistant sorts the material extracted into major taxa. D. The corresponding subsamples are dispatched to taxonomic workgroups. E. Each taxonomic workgroup leader organizes the identification. F. The databasing of the information related to its taxa of interest. Images from the IBISCA-Panama project. (Pictures by H.-P. Aberlenc, S. Ribeiro, R. Le Guen / Panacoco).

5.3. Preparation phase

During the preparation phase, background information about the study site is collated and made available electronically to the project participants. This includes biological, physical, sociological, historical, and administrative information (Janzen & Hallwachs, 1994). More global information, for example existing general sources of information on the fauna, flora or habitats of the region are added too. Existing maps, aerial or satellite images, and GIS layers are of special interest to pre-select study plots. Weather and soil data are particularly important to plan ecological studies. Legal data are needed about the local regulations related to the collection and exportation of biological materials. Part of the information may not be readily available under a published form. Interviews of residents or neighbours can provide useful historical information about the presence of organisms and about natural (hurricanes, floods, and landslides) or human disturbances (*e.g.* previous land use) that occurred at the study site. Land owners and local authorities must be contacted to obtain all required authorizations and also to secure support from local communities.

Prior to the start of the project, a pilot study allows adjustment of sampling protocols (*e.g.* according to the habitat heterogeneity or phenology of the organisms), validation of plot locations inside the study site, and trial runs of database systems. During the preparation phase, priority surveys can be initiated (*e.g.* botanic surveys in study plots). In cases where DNA barcoding is also included, careful plans should be made before and during its execution to minimize genetic degradation (see chapter 7 and appendix I).

5.4. Execution phase

Once budgets have been secured, protocols have been defined and tested, and background information has been accumulated, the execution phase can start. Experts in various disciplines are then invited to participate to the project. A key to the collective success of the project is that participants adhere to some rules.

5.4.1. Terms of engagement of participants

Participants must agree to follow the ‘rules of the club’ which basically are:

- To minimize environmental impact: perturbations associated with collecting, observing or sampling the site biodiversity should be kept as close as possible to natural level (*i.e.* should not induce unusual levels of population fluctuations) (Janzen & Hallwachs, 1994). Interferences with the organization of local human communities should be reduced too.
- To accept the logistical, financial and security constraints: *i.e.* to support as agreed part or the totality of costs related to food, accommodation, transport, lab or other infrastructure and to respect the conditions of access to these infrastructures.

- To facilitate collegial activities: *i.e.* to collaborate to the coordination of field and laboratory activities, to share information and material with other participants.
- To accept the responsibility of delivering data and specimens according to the schedule agreed with the project coordinators. This is sometimes a prerequisite for the reimbursement of part of their expenses by the project organization. The data must be provided in a format compatible with the collective database. Specimens should be deposited in major museums to ensure their long-term conservation and accessibility. Ideally voucher specimens should be available in the form of a reference collection accessible locally, near the study site, and globally in the form of an interactive image database.

5.4.2. Central database

The scientific impact of the inventory clearly depends on the cooperation between participants and on the sharing of the data collected. The knowledge of the concurrent distributions of multiple taxa in an ecosystem is of great value in terms of conservation and for understanding the biology and ecology of the organisms present. The central database (Fig. 4) is supervised by an administrator who is responsible for the data integrity of the whole project. This is only possible if all participants use the same data organizational schema (*e.g*. collecting event, higher taxa, etc.). This implies an exchange of information between the database administrator who has to circulate standardized data fields and identifiers, and the survey participants who have to provide the basic information.

Broadly, this involves the following categories:

- **Collecting events.** This information includes the “where” and “when” components. Even if planned ahead during the general sampling design (*e.g.* plots or transects), participants must provide or validate some information (*e.g.* time and date, collector, method, habitat description, and other additional information such as images).
- **Specimens.** This set of information pertains to “what”. If the central database does not already include a complete list of higher taxa (taxa that ranks above than the species level), participants should provide their own list of taxa of interest. Participants must also provide the list of taxa that they identified in their samples. If several taxonomists identify specimens of the same higher taxa, the taxonomic workgroup leader must standardize the same system of codes (*i.e.* taxon identifier). Additional specimen-based data include who identified the specimen, the basis of the ID, refined location (height, depth), microhabitat, associated specimen, etc. If no species name is readily assignable, some system of morphospecies designation should be adopted.

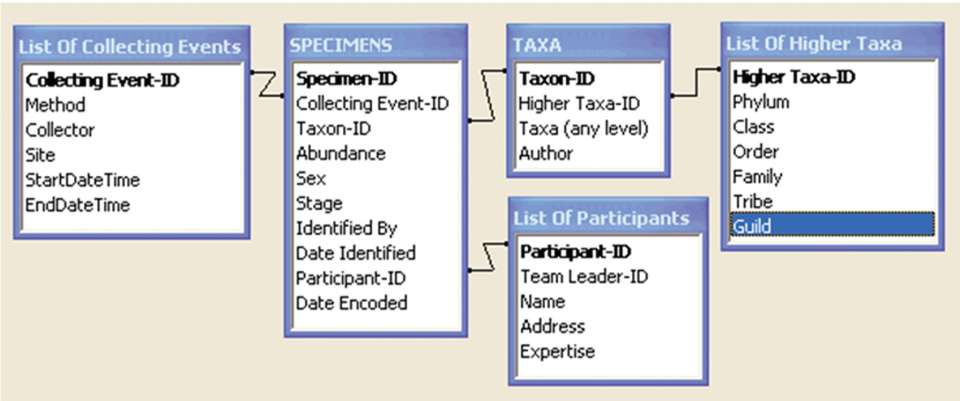

Fig. 4. Simplified structure of the five main tables constituting the core of a central database. Each table is equivalent to a spreadsheet. Fields (column headers) are listed in each box. Each record has a unique identifier (fields in bold ended by "-ID") ensuring an unambiguous relationship with the other tables and avoiding information duplication. Tables are linked with "one to many" relationships symbolised by connecting lines (e.g. several specimens belonging to different taxa can be found during a single collecting event). Three tables, those starting with "List Of", contain data common to all participants and which are used as entries for combo boxes. This allows all participants to use the same identifier for collecting events, participant names and higher taxa. Individual participants input data in the "Specimens" table and in the "Taxa" table. Other tables with additional information on study sites, environmental data, etc. can be added to the system.

It should be noted that the use of imposed codes for the whole project does not preclude the participants to use in parallel their own coding system. The database should be designed to handle participants' own collector's codes. Furthermore besides the standardized basic information about collecting events, taxa and specimens presented above, participants must be free to add in the centralized database additional information specifically relevant to their target taxa.

Even if the central database is web-accessible and can be downloaded, it is sometimes more practical for participants to encode the information in a local file saved on their own computer. Usually this is done by downloading a template under the form of a worksheet or database (preferably in an open source format). Once the data input is completed, participants can upload their file which is merged to the central database by the database administrator. To facilitate a wide dissemination of the biodiversity information, the central database can ultimately be provided to GBIF (Global Biodiversity Information Facility).

The database must also be GIS-interrelational and the use of a GPS device to georeference the observations in the field must be encouraged (see chapter 4). The mapping of environmental data and of other factors such as plot accessibility can be a very useful organizational and analysis tool. Plots can be selected along environmental gradients or to maximize the return of information per unit effort (White & Langdon, 2006). Maps of predicted species distribution can also be inferred from these data.

5.4.3. Labelling: standardization of data coding

Correct labelling of specimens is of prime importance. A label with a misspelled code or with unreadable information because its ink faded becomes unusable. If possible, it is recommended that good quality labels are printed for participants before field work starts and plan this item in the overall budget (see example at Fig. 5A). Labels with a unique identifier (*e.g.* an alphanumeric code) for each collecting event will serve as reference for the whole project. Participants carry a series of labels to the field and are encouraged to add to the sample another label with an alias identifier corresponding to their own coding system and with details on the collecting event (*e.g.* method, site, date, etc.).

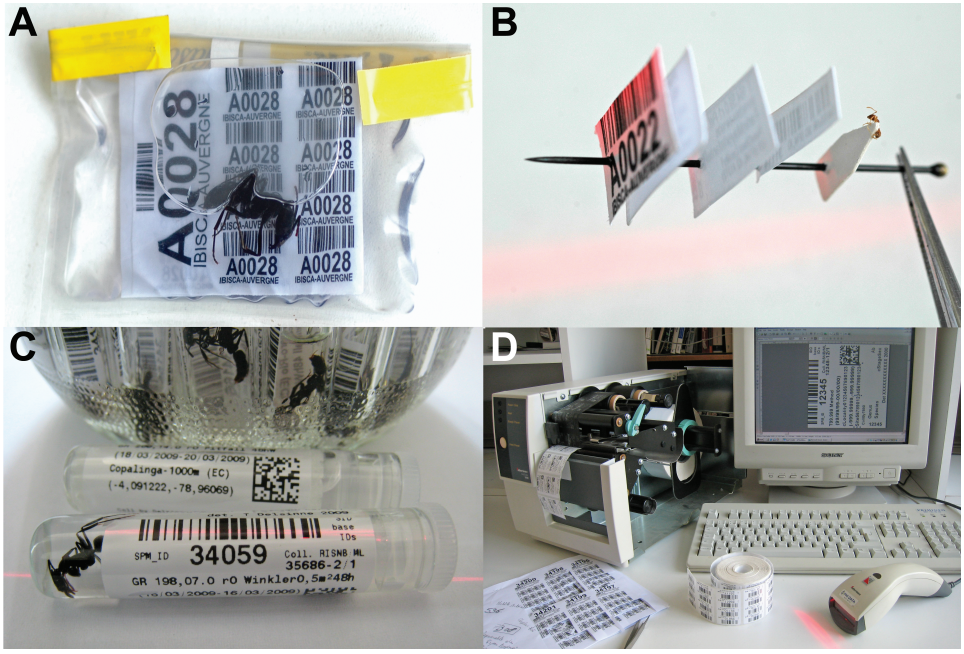

Fig. 5. Labelling of specimens with optical barcodes facilitates specimen management. A. Sampling bag (whirlpak) containing a series of identical labels. When sorting the sample it is then easy to add a label to each subsample stored in dry (B) or wet (C) condition. Two types of barcode are useable: unidimensional (stripes) or bidimensional (mosaic). D. A thermal label printer (on the left) is more costly than a regular laser printer but allows printing long-lasting labels required for long-term storage. Barcodes are generated by a specialized software. The barcode scanner (on the right) is connected in parallel to the computer keyboard. (Pictures by M. Leponce).

The addition of optical barcodes on the label speeds up the work of data encoding without errors (*e.g.* acquisition of sample, specimen or species codes). The barcode scanner is connected to the computer. No additional software is required since the scanner signal enters directly into the keyboard input. Unidimensional barcodes are the most commonly used, are generated by cheap or even free software and are readable with scanners costing around 250 €.

Bidimensional (2D) barcodes (dot matrices) allow storing a higher density of information but require a slightly more expensive scanner for reading them. Symbology code 128 has been used with success to generate small barcodes suitable for entomology (see Fig. 5B). Labels can be printed on 120 g paper with a laser printer but will peel off in alcohol over time. The ink of inkjet printers is often dissolved in the preservative. The best solution, but more expensive (several hundred to thousands Euros), is to use a thermal printer burning a special solvent-resistant ink on a thin sheet of polyester (Fig. 5C,D). The only disadvantage is that it is difficult to write on those polyester labels. Additional data (like sex, colour, length, etc.) is easier to write on archive paper with a pencil.

5.4.4. On-site management

Project coordinators assist and supervise the work of participants on site. They organize meetings with the participants and ensure that field data collection follows the general sampling protocol and is smoothly coordinated. Assistants dedicated to special techniques (DNA sampling, parasite collection, photography, etc.) accompany the collectors to the field. Managers are charged with planning all logistical aspects associated with the project (equipment, food, transport and accommodation of participants, base camps). Administrative constraints should be kept as light as possible to allow the participants to concentrate on their research.

5.4.5. Legal issues, collecting and export permits

Participants must respect all local regulations, decrees, laws and traditions. In particular they must ensure that they possess all the necessary permits for specimen collection and exportation. For some large projects such as the SANTO2006 expedition and the Moorea Biocode Project, collective, or umbrella permits are obtained. This implies that a single institution (*e.g.* a museum) or consortium may be responsible for all the material collected. These institutions may require that some or most material be deposited in their collections. In other cases each participant must ask a permit for its own material. The collections associated to the inventory are then spread among various institutions. Project coordinators should know the permitting and exportation rules and provide the required information to participants. Special attention should be given to the use of certain techniques or species. In particular, mass-collection is sometimes forbidden (*e.g*. rotenone stations, tree fogging), sensitive areas are often off-limits (*e.g*. small mountain peaks, popular dive sites), rare species are usually protected (endangered plants and animals), and the transportation of specimens in hazardous (flammable) fluids is regulated.

5.4.6. Security

Procedures in case of emergencies must be planned. General safety instructions should be given to project participants going out to the field (see for example Langdon & MacCulloch, 2004). If field operations are based from a research

station, all participants should be briefed by station staff about standard safety issues. Trained personnel (*e.g*. physician, nurse, etc.) should be available especially when conducting expeditions in remote areas, and an emergency plan should be submitted prior to operations that include the location of nearest hospital, decompression chamber, etc. Field participants should carry telecommunication equipment (*e.g*. mobile phone, walkie talkie, satellite phone, satellite beacon) and a first aid kit. A registry must be kept at the field base where participants indicate for each day their planned activity, location, estimated departure and return time. Such registry is also necessary to allow managers to organize the transportation and to provide the equipment and personnel (e.g. boats, climbing gears and tree climbers) needed by each research team. Trails and hazards (*e.g*. unapparent traps) should be well marked. Specific precautions associated with the handling of dangerous organisms or fixatives or with sampling in "extremes" habitats (*e.g*. caves) are described in the relevant chapters of this manual.

5.4.7. On-site processing and dispatching of material

Specimens captured with mass-collection methods have to be pre-processed as soon as possible, ideally on site (*e.g*. in a field research station). This work can be done by assistants supervised by senior taxonomists. During this process samples are divided in subsamples, based on taxonomic groups (Fig. 3), and are dispatched to taxonomic workgroup leaders in charge. Each leader defines who may be the appropriate taxonomist(s) for a finer identification, sends him the material and is responsible for the return of information to the central database. The number of specimens in each subsample is counted (or estimated roughly in the case of huge numbers) and encoded in the central database together with the name of the workgroup leader in charge. The same kind of information is also encoded for taxon-specific collection methods to keep a complete record of the material collected. When the database administrator receives datasets with identified specimens and merges them to the central database he must adjust the total numbers of identified and unidentified specimens per sample. If no experts are at hand for some taxonomic groups, the related specimens can be grouped as "residual material" and kept for later study or advertised on the clearing house web page (see above).

5.4.8. Incentives and follow-up

After the field work, progress reports are sent by workgroups to coordinators, and meetings are organized among workgroup leaders in order to follow a common agenda leading to several collective publications (*e.g*. book, special issue of a journal, publication in a high profile journal). Workshops are useful to review the overall progress of the inventory, address the concerns of stakeholders, define resource needs, promote consensus and reassess priorities and objectives. A substantial budget – at least the same amount than for field work – must be secured to stimulate the completion of the work.

5.4.9. Monitoring

Monitoring involves the repeated collection of long-term biodiversity data to evaluate changes in populations and environmental conditions. It can be used as an early warning system of changes in ecosystem functioning or to evaluate management actions. Monitoring targets certain taxa with the use of specific, non-intruding protocols. This activity is out of the scope of the present chapter and we refer the reader to more specialized references concerning that matter (*e.g.* Comiskey *et al.*, 1999; Yoccoz *et al.*, 2001; Schmeller, 2008; Nielsen *et al.*, 2009).

6. Conclusions

Large-scale species inventories and especially ATBIs are an effective way to increase our knowledge of the diversity of life on our planet. They are successful by creating synergies among the participants and allow an overall picture of complex ecosystems, something that would be impossible to obtain with smaller projects. In terms of science, long-term and representative biodiversity datasets are of great impact. Comprehensive inventories valorise the role of biological diversity in the functioning of ecosystems and the fundamental role of taxonomists. They are also instrumental for conservation and management decisions and contribute to raise public awareness about the need of conserving biodiversity. However the task is so huge that such endeavour requires careful planning. Resources to conduct ATBIs are limited, especially the taxonomic workforce itself. Lessons learned from the ongoing Smokies ATBI (Langdon *et al.*, 2006) show in particular that data management and data quality assurance are absolutely critical, funding must be secured to secure taxonomic assistance, bureaucratic burden must be reduced, over-collection of specimens must be avoided, the right person must be matched with the right position (organization of workgroups), participants must be well treated (infrastructure and logistics), and everyone must be involved in keeping costs down.

Biodiversity inventories can become never ending tasks. It is therefore important to keep the motivation of participants and of stakeholders by carefully planning the project output. The strategy must include pilot studies with fast deliverables in addition to long-term studies. A continuously updated website and database is probably a very good portal to show the results and dynamics of the project.

Taxonomic work (identifications, descriptions, revisions) usually takes time but preliminary results such as images of specimens or DNA sequences can be made publicly available quickly. Good visibility of the project is certainly also important to maintain the interest of the sponsors. New tools and approaches remain to be developed to increase the inventory efficiency. One could think of more automation in the tracking, documentation, storage and retrieval of specimens (*e.g.* increased use of standardized optical barcodes, semi-automated 3D image capture of specimens) or of new techniques to access environments difficult to reach (deep soil, canopy, etc.). The workforce should also be increased with the development of new taxonomic centres that could process efficiently all the material collected. The development of DNA barcoding and the

reduction of its cost will certainly open new opportunities to conduct inventories and monitoring. However, specimens or their body parts will always need to be collected in the first place and this represents the main bottleneck to appreciate the true dimensions of biodiversity on Earth (May, 2004). The accessibility of biodiversity data should be increased, following initiatives such as those of the Global Biodiversity Information Facility (GBIF), Encyclopedia of Life (EoL), Consortium for the Barcode of Life (CBOL), etc. Finally, the complementarity of ATBIs in terms of global biodiversity coverage and of scientific questions tackled should be increased by incorporating them into a global network. A move in this direction is done with the ATBI alliance which aims to expand the Smokies ATBI model to other protected areas in the U.S. and to provide the connectivity between local ATBI efforts (Langdon *et al.*, 2006; Hetrick *et al.*, 2007). The network of permanent plots of the Smithsonian Institution Global Earth Observatories aims at long-term monitoring of tropical and temperate forests. The network is in fast expansion and currently includes 34 sites in 20 countries. Collaboration with other global networks could be the next step for EDIT's ATBIs+M.

7. Acknowledgments

JD acknowledges his research grant by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brasil. CPM acknowledges funding from the Gordon and Betty Moore Foundation and EDIT for participation in the ATBI+M workshop. Roger Le Guen, Panacoco, provided free use of his photos. Isabelle Bachy & Yves Laurent, RBINS, helped to edit the figures and text. We thank an anonymous referee for useful suggestions.

8. References

ADIS, J., DA FONSECA, C.R.V & MARTIUS, C. 1998. Mecanismos que mantém a diversidade tropical - um projeto de pesquisa bilateral teuto-brasileiro. *Acta Amazônica* 28: 205-215.

ANDRE, H.M., DUCARME, X. & LEBRUN, P. 2002. Soil biodiversity : myth, reality or conning ? *Oikos* 96: 3-24.

BASSET, Y. 1991. Leaf production of an overstorey rainforest tree and its effects on the temporal distribution of associated insect herbivores. *Oecologia* 88: 211-219.

BASSET, Y. 2000. Insect herbivores foraging on seedlings in an unlogged rain forest in Guyana: spatial and temporal considerations. *Studies on Neotropical Fauna and Environmen*t 35: 115-129.

BASSET, Y., NOVOTNY, V., MILLER, S.E. & PYLE, R. 2000. Quantifying biodiversity: experience with parataxonomists and digital photography in New Guinea and Guyana. *Bioscience* 50: 899-908.

BASSET, Y., ABERLENC, H.-P., BARRIOS, H., CURLETTI, G., BERANGER, J.-M., VESCO, J.-P., CAUSSE, P., HAUG, A., HENNION, A.-S., LESOBRE, L., MARQUES, F. & O'MEARA, R. 2001. Stratification and diel activity of arthropods in a lowland rain

forest in Gabon. *Biological Journal of the Linnean Society* 72: 585-607.

BASSET, Y., NOVOTNY, V., MILLER, S.E. & KITCHING, R. 2003a. Arthropods of tropical forests. Spatio-temporal dynamics and resource use in the canopy. Cambridge University Press, xvi + 474 pp.

BASSET, Y., HORLYCK, V. & WRIGHT, S.J. (Eds) 2003b. *Studying Forest Canopies from Above: The International Canopy Crane Network*. Smithsonian Tropical Research Institute and UNEP, Panama City, 196 pp.

BASSET, Y., NOVOTNY, V., MILLER, S.E., WEIBLEN, G.D., MISSA, O. & STEWARD, A.J.A. 2004. Conservation and biological monitoring of tropical forests: the role of parataxonomists. *Journal of Applied Ecology* 41: 163-174.

BASSET, Y., CORBARA, B., BARRIOS, H., CUÉNOUD, PH., LEPONCE, M., ABERLENC, H.P., BAIL, J., BITO, D., BRIDLE, J.R., CASTAÑO-MENESES, G., CIZEK, L., CORNEJO, A., CURLETTI, G., DELABIE, J.H.C., DEJEAN, A., DIDHAM, R.K., DUFRÊNE, M., FAGAN, L.L., FLOREN, A., FRAME, D., HALLÉ, F., HARDY, O.J., HERNANDEZ, A., KITCHING, R.L., LEWINSOHN, T.M., LEWIS, O.T., MANUMBOR, M., MEDIANERO, E., MISSA, O., MITCHELL, A.M., MOGIA, M., NOVOTNY, V., ØDEGAARD, F., GAMA DE OLIVEIRA, E., ORIVEL, J., OZANNE, C.M.P., PASCAL, O., PINZÓN, S., RAPP, M., RIBEIRO, S.P., ROISIN, Y., ROSLIN, T., ROUBIK, D.W., SAMANIEGO, M., SCHMIDL, J., SØRENSEN, L.L., TISHECHKIN, A., VAN OSSELAER, C. & WINCHESTER, N.N. 2007. IBISCA-Panama, a large-scale study of arthropod beta-diversity and vertical stratification in a lowland rainforest: rationale, study sites and field protocols. *Bulletin de l'Institut Royal des Sciences Naturelles de Belgique, Entomologie* 77: 39-69.

BOUCHET, P., LE GUYADER, H. & PASCAL, O. 2009. The SANTO 2006 Global Biodiversity Survey: an attempt to reconcile the pace of taxonomy and conservation. *Zoosystema* 31: 401-406.

CHAO, A., COLWELL, R.K., LIN, C.W. & GOTELLI, N.J. 2009. Sufficient sampling for asymptotic minimum species richness estimators. *Ecology* 90: 1125-1133.

CHAZDON, R.L., COLWELL, R.K., DENSLOW, J.S. & GUARIGUATA, M.R. 1998. Statistical methods for estimating species richness of woody regeneration in primary and secondary rain forests of NE Costa Rica. *In* Dallmeier, F. & Comiskey, J.A. (Eds). Forest biodiversity research, monitoring and modeling: conceptual background & Old World case studies. Parthenon Publishing, Paris: 285-309.

CHECK, E. 2006. Treasure island: pinning down a model ecosystem. *Nature* 439: 378-379.

COLWELL, R.K. 1994-present. EstimateS: statistical estimation of species richness and shared species from samples. [Persistent URL: http://purl.oclc.org/estimates.]

COLWELL, R.K & CODDINGTON, J.A. 1994. Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions of the Royal Society of London, Series B* 345: 101-118.

CODDINGTON, J.A., AGNARSSON, I., MILLER, J.A., KUNTNER, M. & HORMIGA, G. 2009. Undersampling bias: the null hypothesis for singleton species in tropical arthropod surveys. *Journal of Animal Ecology* 78: 573-584.

COMISKEY, J.A., DALLMEIER, F. & ALONSO, A. 1999. Framework assessment and monitoring of biodiversity. *In*: LEVIN, S. (Ed.) *Encyclopedia of Biodiversity*. Princeton, NJ: 63-73.

DELABIE, J.H.C., FISHER, B.L., MAJER, J.D., WRIGHT, I.W. 2000. Sampling effort and choice of methods. *In*: AGOSTI, D.; MAJER, J. ALONSO, L.; SCHULTZ, T. (Eds). *Ants. Standard methods for measuring and monitoring biodiversity*. Smithsonian Institution Press: 145-154.

DELABIE, J.H.C., JAHYNY, B., DO NASCIMENTO, I.C., MARIANO, C.S.F., LACAU, S., CAMPIOLO, S., PHILPOTT, S.M. & LEPONCE, M. 2007. Contribution of cocoa plantations to the conservation of native ants (Insecta : Hymenoptera : Formicidae) with a special emphasis on the Atlantic Forest fauna of southern Bahia, Brazil. *Biodiversity and Conservation* 16: 2359-2384.

DUCANE GODMAN, F. & SALVIN, O. (Eds) 1879-1915. *Biologia Centrali-Americana (Prospectus): Zoology, Botany and Archæology*. 63 vol. published R.H. Porter, London.

GOTELLI, N.J. & COLWELL, R.K. 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* 4: 379-391.

GÁMEZ, R., LOVEJOY, T., SOLORZANO & R., JANZEN, D. 1997. Costa Rican All-Taxa Survey. *Science* 277:18-19.

GODFRAY, H.C.J. 2002. Challenges for taxonomy. *Nature* 417: 17-19.

HAJIBABAEI, M., DEWAARD, J. R., IVANOVA, N.V., RATNASINGHAM, S., DOOH, R.T., KIRK, S.L., MACKIE, P.M. & HEBERT, P.D.N. 2005. Critical factors for assembling a high volume of DNA barcodes. *Philosophical Transactions of the Royal Society of London*, *Series B* 360: 1959-1967.

HANBURY-TENISON, R . & JERMY, A.C. 1979. The RGS expedition to Gunung Mulu Sarawak. *Geographical Journal* 145: 175-191.

HETRICK, S., KILLEFFER, T., & HILTEN, J. (Eds). 2007. *Biodiversity data interaction for an alliance of all taxa biodiversity inventories: a workshop to design methods, standards, and practices for data-sharing at the national level.* Proceedings of a Workshop. 04-05 December, 2006. Gatlinburg, Tennessee, USA. 44 pp.

HILTEN, J., COX, J., DOURSON, D., GROSSNICKLE, H., PIERCE, J., SACHS, S., SUPER, P. & WETZEL, M. 2006. Science Education Programs with the ATBI. *The George Wright Forum* 23: 37-44.

HOPKINS, G.W. & FRECKLETON, R.P. 2002. Declines in the numbers of amateur and professional taxonomists: implications for conservation. *Animal Conservation* 5: 245-249.

IBOL BARCODING BIOTAS WORKING GROUP 2009. How to barcode biotas. A manual for planning. Biodiversity. Institute of Ontario, 40 pp.

JANZEN, D.H. 1988. Guanacaste National Park: tropical ecological and biocultural restoration. In: *Rehabilitating damaged ecosystems*, vol. 11. Cairns, J., Jr. (Ed), CRC Press, Boca Raton, Florida: 143-192.

JANZEN, D.H. & HALLWACHS, W. 1994. All Taxa Biodiversity Inventory (ATBI) of terrestrial systems: a generic protocol for preparing wildland biodiversity for non-damaging use. Report of a National Science Foundation Workshop, 16-18 April 1993, Philadelphia, Pennsylvania, USA, 141 p.

JANZEN, D.H. 2004. Setting up tropical biodiversity for conservation through non-damaging use: participation by parataxonomists *Journal of Applied Ecology* 41: 181-187.

JANZEN, D.H., HAJIBABAEI, M., BURNS, J., HALLWACHS, W., REMIGIO, E. & HEBERT, P. 2005. Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. *Philosophical Transactions of the Royal Society of London, Series B* 360: 1835-1845.

JONES, D.T. & EGGLETON, P. 2000. Sampling termite assemblages in tropical forests: testing a rapid biodiversity assessment protocol. *Journal of Applied Ecology* 37: 191-203.

KNIGHT, W.J. & HOLLOWAY, J.D. (Eds) 1990. *Insects and the rain forests of south east Asia (Wallacea)*. Royal Entomological Society of London, 343 pp.

KRESS, W. J., WURDACK, K. J., ZIMMER, E. A., WEIGT, L. A. & JANZEN, D. H. 2005. Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences* 102: 8369-8374.

LANGDON, K. & MACCULLOCH, H. 2004. Backcountry Safety Guide. *Inventory and Monitoring Branch, Great Smoky Mountains National Park.* URL: http://www.dlia.org/atbi/science/guidelines_safety.shtml

LANGDON, K., PARKER, C. & NICHOLS, B. 2006. Scientific findings, success stories, lessons learned, and an alliance of ATBIs. *The George Wright Forum* 23: 52-60.

LAWTON, J.H., BIGNELL, D.E., BOLTON, B., BLOEMERS, G.F., EGGLETON, P., HAMMOND, P.M., HODDA, M., HOLT, R.D., LARSEN, T.B., MAWDSLEY, N.A., STORK, N.E., SRIVASTAVA, D.S & WATT, A.D. 1998. Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest. *Nature* 391: 72-76.

LEATHER, S.R. 2009. Taxonomic chauvinism threatens the future of entomology. *The Biologist* 56: 10-13.

LONGINO, J.T. & COLWELL, R.K. 1997. Biodiversity assessment using structured inventory: capturing the ant fauna of a tropical rainforest. *Ecological Applications* 7: 1263-1277.

LONGINO, J.T. 2000. What to do with the data? *In*: Agosti, D., Majer, J., Alonso, L. & Schultz, T. (Eds). *Ants. Standard methods for measuring and monitoring biodiversity?* Smithsonian Institution Press: 186-203.

LONGINO, J.T., CODDINGTON, J. & COLWELL, R.K. 2002. The ant fauna of a tropical rain forest: estimating species richness in three different ways. *Ecology* 83: 689-702.

MAGNUSSON, W. E., LIMA, A. P., LUIZÃO, R., LUIZÃO, F., COSTA, F. R. C., VOLKMER DE CASTILHO, C. & KINUPP, V.F. 2005. RAPELD: a modification of the Gentry method for biodiversity surveys in long-term ecological research sites. *Biota Neotropica* 5: 1-6.

MAY, R.M. 2004. Tomorrow's taxonomy: collecting new species in the field will remain the rate-limiting step. *Philosophical Transactions of the Royal Society of London, Series B* 359: 733-734.

MILLER, S.E., KRESS, W.J. & SAMPER, C.K. 2004. Crisis for biodiversity collections. *Science* 303: 310.

NICHOLS, B.J. & LANGDON, K.R. 2007. The Smokies All Taxa Biodiversity Inventory: History and progress. *Southeastern Naturalist*. Special Issue 1: 27-34.

NIELSEN, S.E., HAUGHLAND, D.L., BAYNE, E. & SCHIECK, J. 2009. Capacity of large-scale, long-term biodiversity monitoring programmes to detect trends in species prevalence. *Biodiversity and Conservation* 18: 2961-2978

NOVOTNY, V. & BASSET, Y. 2000. Rare species in communities of tropical insect herbivores: pondering the mystery of singletons. *Oikos* 89: 564-572.

NOVOTNY, V., MILLER, S.E., HULCR, J., DREW, R.A.I., BASSET, Y., JANDA, M., SETLIFF, G.P., DARROW, K., STEWART, A.J.A, AUGA, J., ISUA, B., MOLEM, K., MANUMBOR, M., TAMTIAI, E., MOGIA, M. & WEIBLEN, G.D. 2007. Low beta diversity of herbivorous insects in tropical forests. *Nature* 448: 692-695.

PARKER, C. & BERNARD, E. 2006. The science approach to the Smokies ATBI. *The George Wright Forum* 23: 26-36.

ROISIN, Y. & LEPONCE, M. 2004. Characterizing termite assemblages in fragmented forests: a test case in the Argentinian Chaco. *Austral Ecology* 29: 637-646.

RONQUIST, F. & GÄRDENFORS, U. 2003. Taxonomy and biodiversity inventories: time to deliver. *Trends in Ecology and Evolution* 18: 269-270.

SCHLICK-STEINER, B., STEINER, F.M., SEIFERT, B., STAUFFER, C., CHRISTIAN, E. & CROZIER, R.H. 2010. Integrative taxonomy: a multisource approach to exploring biodiversity. *Annual Review of Entomology* 55: 421-438.

SCHMELLER, D.S. 2008. European species and habitat monitoring: where are we now? *Biodiversity and Conservation* 17: 3321-3326.

SHARKEY, M.J. 2001. The All Taxa Biological Inventory of the Great Smoky Mountains National Park. *Florida Entomologist* 84: 556-564.

SMALL A., ADEY A. & SPOON D. 1998. Are current estimates of coral reef biodiversity too low? The view through the window of a microcosm. *Atoll Research Bulletin* 458: 1-20.

WALTHER, B.A. & MOORE, J.L. 2005. The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography* 28: 815-829.

WHITE, P., MORSE, J., HARRIS, F., LANGDON, K., LOWE, R., NICHOLS, B., PARKER, C., PICKERING, J. & SHARKEY, M. 2000. The Science Plan for the All Taxa Biodiversity Inventory in Great Smoky Mountains National Park, North Carolina and Tennessee.
On-line at www.dlia.org/atbi/quarterly_newsletter/pdfs/science_plan.pdf.

WHITE, P. & LANGDON, K. 2006. The ATBI in the Smokies: an overview. *The George Wright Forum* 23: 18-25.

YOCCOZ, N.G., NICHOLS, J.D. & BOULINIER, T. 2001 Monitoring of biological diversity in space and time. *Trends in Ecology and Evolution* 16: 446-453.

9. Appendices

Appendix 1. New barcoders checklist

Some questions to which you should have specific answers before starting, and questions for iterative re-visiting as you go along.

General/Roles

- What organisms am I going to Barcode? What particular set of problems do these organisms present? Has anyone already done or actively is doing this? How many samples do I foresee collecting/processing?
- What do I already have and what do I need to do so?
- How much is this going to cost?
- Have I arranged for permits: collecting, export, import? Who is doing so if not me?
- Who (what person, institution or country) owns the intellectual rights to the specimens and to the information (barcode and collateral) associated with the project?
- Do you have the political/social/permit power to donate the specimens and/or their information to the recipient (GenBank, Guelph, Smithsonian or your museum, etc.)?
- What kind of condition will they be captured in, maintained in, transported in, vouchered in?
- Who are your points of contact for permits, vouchering, taxonomic identification, DNA extraction, extraction bio-banking, Polymerase Chain Reaction (PCR), DNA sequencing, data quality control (QC), data management, etc.?
- Who is primarily planning the project you are doing or contributing to?
- Which costs should I anticipate at various stages of the analytical process?
- Do I fully grasp the implications and differences between doing barcoding to simply build the overall/global barcode library, and doing barcoding for both

this purpose and species discovery (both in simply new species and in cryptic species)?

- What will I do when my sequencing lab runs out of funding in the middle of my project?
- Who is going to write the publication(s)? Who is going to co-author the work, in what order? Who is going to pay the reprint and page charges?

Vouchers/Taxonomy

- What collection/institution is going to receive my voucher specimens, what care will they get?
- Who is going to pay the bill for the storage/curation/subsequent identification and re-identification of my voucher specimens, why, when and for what reciprocal gain?
- How many vouchers per species is the receiver willing to take in, and just for barcode vouchers or also for exploratory biosystematics where warranted (*e.g.* 5-10 might be fine to establish a barcode library, but 100's may be necessary for exploring variation and cryptic species).
- Who is going to actually identify (first pass), re-identify (second pass), re-re-identify (n pass) my voucher specimens, and why should they care or bother (How am I going to compensate them)?

Metadata

- Who is collecting the metadata (GPS, photo, measurements, etc)? What metadata do I need to (minimum) or want to (optimally) collect?
- Do I have a digital camera and GPS unit? How do I keep the photos linked with the specimens?
- Who will receive and store and curate the images that are collateral data for the voucher specimens (and sometimes, the only voucher specimen that there is)?
- Do I understand to take high quality (though not necessarily beautiful) voucher specimen images that display the important identification traits (if possible) of the voucher specimen?
- Am I planning to have an individual metadata record in a standard database (DB) for every voucher specimen and collecting event? Is this DB website friendly and what website will house (and curate?) this DB, for what reason and with what caveats?
- Where will an electronic hard copy (frozen) version of this DB be deposited for long-term permanent storage, but at intervals replaced with an updated version (and who will do it)?
- Do I understand the difference between an event-based DB and a specimen-based DB?

Specimen collection and sub-sampling

- Who is actually collecting the specimens? Who is collecting the tissues for lab analysis?
- How are they collecting them? Where (if needed) are they getting the training, supplies, materials, kits and instructions to do so?
- What portion(s) of my animals am I going to take? Is this compatible with success in the lab and with subsequent morphological examination of the voucher? How do I remove my compatible tissues from the shell/skeleton/body part, etc.?
- What kind of container am I putting the tissue into? With what preservative? How am I transporting these back to the lab? How do I avoid cross-contamination?
- Where am I bio-banking the (leftover) tissues?
- Who is guiding/proofing/fact-checking the field operations as they happen, and then by what mechanism will the barcoding results be fed back to the DB that contains the voucher collateral information, both to correct errors and to update the field identifications?

Laboratory

- Where/who is performing my DNA extractions? Are they archivable? Where will they be bio-banked?
- What protocol should I use for DNA extractions? Which sub-sampling procedure should I use to avoid cross-contamination? How much tissue do I need for DNA extraction? Do my organisms present any difficulties for DNA extractions? How do I do quality control (QC) for DNA extractions?
- Who is doing the PCR? Am I using Cytochrome Oxidase I (COI)? Does COI work for my organisms? What primers should I use? Are there any potential PCR obstacles from my organisms?
- How do I do QC for PCR? How do I check for contamination?
- Who is doing the DNA sequencing reactions? What do I need to provide them? How do I do QC on my DNA sequences? How do I know if my DNA sequences are good? Correct?
- How should I label my DNA sequencing reactions so that the chromatograms are easy to upload to BoLD/GenBank with my data?
- Who is going to, and WHY (and who pays his costs) manage the iterative process of my getting back neighbour joining (NJ) trees of sequences for my vouchers, studying them, and sending comments for corrections and elaborations back to the person/system that provided the NJ trees, and go

round and round with this? Who and why will then search for corroborative nuclear sequences when appropriate/warranted?

Data Handling

- How should I manage all my data? How and when do I submit specimen data to BoLD?
- What problems can I anticipate and avoid?
- When I discover errors or updates in my voucher specimen collateral information DB, how does this modification arrive at the target DBs such as GenBank, BoLD, etc.
- What do I do when a taxonomist out there disagrees with the name that I submitted for a barcode voucher, either at the species level or a higher taxon?
- Who owns which portions of your datastream from the field to GenBank or other final repository?
- What do I do if there is no taxonomist or taxonomic process willing to do the basic taxonomic process on my voucher specimens?
- Where do I turn for help?

Appendix 2. Some links and contacts

- CBOL: http://www.barcoding.si.edu
- BOLD: http://www.barcodinglife.com
- Some leading labs contacts: Lee Weigt (weigtl@si.edu); Chris Meyer (meyerc@si.edu); Amy Driskell (driskella@si.edu); Robyn Cowan (r.cowan@rbgkew.org.uk); Natalia Ivanova (nivanova@uoguelph.ca); Mehrdad Hajibabaei (mhajibab@uoguelph.ca).