Chapter 5
Methodological Challenges of a Megadiverse Ecosystem

G. Brehm(✉), K. Fiedler, C.L. Häuser, and H. Dalitz

5.1 Introduction: What is Megadiversity?

'Megadiversity' originated as a term in the context of biodiversity conservation in the late 1980s (Mittermeier et al. 2004). It refers to countries with an extremely high level of species richness, usually found in the tropical realm, one or two orders higher in magnitude than in most temperate zone countries. In many ways countries or areas of megadiversity coincide with the slightly longer-established concept of biodiversity hotspots (Myers et al. 2000; Brummitt and Lughadha 2003; see Chapter 2 in this volume). Unlike the concept of hotspots, megadiversity also attempts to take into account degrees of endemism, phylogenetic relatedness, and other measures of diversity applied for identifying biodiversity hotspots as opposed to pure numbers of species or taxa per unit area. More than any academic differentiation, however, the term megadiversity was recently taken up and promoted at the political level, particularly under the *Convention of Biological Diversity* (CBD), as well as the *Convention on International Trade in Endangered Species of Fauna and Flora* (CITES). Following the original meeting in Cancun, Mexico, in February 2002, 15 countries formed a group of 'Like-Minded Megadiverse Countries' (LMMC) as a forum to address the specific challenges for biodiversity conservation and sustainable use faced by countries with disproportional high levels of biodiversity ('Cancun declaration 2002'). Later joined by Australia and the USA, this informal group of countries comprises many but by no means all of the recognized global biodiversity hotspots.

The Republic of Ecuador qualifies by all standards as being megadiverse, and has also been a founding member of the LMMC group. This relatively small country (approx. 284 000 km^2) harbors an outstanding variety of habitats along pronounced elevational and wet–dry gradients. The high beta diversity along these gradients, situated within the peak of species-richness at tropical low latitudes, favors an enormous biological diversity that is rivaled by only a few other regions in the world. For example, Jørgensen and Léon-Yánez (1999) listed 15.901 vascular plant species (15 306 of them native) known to occur in Ecuador, which is more than six times the number reported for Germany with its 2682 species in a 25% larger territory (357 000 km^2). Among the various habitats in

E. Beck et al. (eds.), *Gradients in a Tropical Mountain Ecosystem of Ecuador.* *Ecological Studies 198.*

Ecuador contributing to its overall megadiverse status, the Andean forests are a significant factor which have also been identified as an area of highest priorities for biodiversity conservation (Aldrich et al. 1997; Brummitt and Lughadha 2003). Apart from the uneven spatial distribution of biodiversity, species richness is not equally distributed among taxa, and arthropods play an outstanding role (see Chapters 2, 11.3, 35 in this volume).

5.2 Methodological Challenges of Megadiversity

The scientific methodologies available for assessing and comparing megadiverse faunas and floras are essentially the same as in less diverse ecosystems. Yet, megadiversity poses three major challenges: First, the sheer number of different species makes it almost impossible to arrive at 'complete' inventories within reasonable time-scales (Novotny and Basset 2000). Second, the size and incompleteness of samples call for sophisticated analytical tools that go far beyond the simple counting of species. Third, the identification of species is critically complicated, even more so in the absence of reliable taxonomic literature and identification tools for most taxa.

The first issue is not trivial, but new methods to deal with these problems have been developed in the past decade. Mathematical procedures to estimate biodiversity burst rapidly (Colwell 2006), and empirical data as well as simulation studies now allow to select appropriate measures once a reasonable density of data has been collated (e.g. Brose and Martinez 2004). The development of 'stopping rules' allows the determination of minimum levels of sampling (Chiarucci et al. 2003). However, even the best mathematical tools are worthless unless appropriate sampling schemes are used. A freely available range of statistical estimation methods (e.g. EstimateS; Colwell 2006) can be reasonably applied only when organisms have been sampled using standardized quantitative methods with sufficient spatial and temporal replication. Even ambitious projects such as the analysis of more than 35 000 individuals of geometrid moths in the RBSF and adjacent areas (Brehm et al. 2005) still lack fully sufficient temporal and spatial replication at certain sites. DeVries et al. (1997) showed that near-complete inventories of Ecuadorian local butterfly faunas can only be achieved by large-scale sampling over many years (covering different seasons of the year), including the understorey and the canopy of the forest. It is therefore hopeless to expect meaningful richness data by sampling such habitats for a few weeks in a 'rapid assessment' style, even more so if organisms are collected qualitatively.

The second issue, i.e. to obtain meaningful comparisons between incomplete samples from diverse communities, has also been studied: Indices to calculate floral or faunal dissimilarity in such cases are now available (e.g. the NESS and CNESS family; Trueblood et al. 1994); and, by means of appropriate data transformations and sophisticated multivariate procedures (e.g. Legendre and Legendre 1998; Legendre and Gallagher 2001), it is now possible for ecologists to analyze the structure of such complex data sets.

It is the third issue which still causes the largest concern, i.e. proper and reliable sorting of samples and identification of species in cases where the flora, or fauna, is extremely rich, poorly known, or even partly undescribed. This task is often the most time-consuming in tropical community ecology. While whole organisms or samples are collected in a broad range of organisms, distance-sampling methods are applied for many groups of vertebrates. These methods require an excellent expertise that often needs to be trained over months or, more realistically, over years of fieldwork. For example, quantitative bird surveys in tropical forests require the knowledge of a vast number of similar bird voices (C. Rahbek, personal communication). Much expertise needs to be gathered in order to get 'the look' for certain patterns in other groups: Where does the well camouflaged caterpillar or the singing bush cricket sit? Where do the liverwort or *Piper* species grow? Where is the best place to net bats? In which trees do the tanagers feed or breed? A good field biologist will quickly learn, but the amount of knowledge about the many species in megadiverse communities often seems endless. While tropical fieldwork poses various problems, subsequent laboratory work on field samples is challenging in a different manner and might consume even more time. For example, during three studies of moth diversity along environmental gradients in southern Ecuador (Brehm 2002; Süßenbach 2003; Hilt 2005) the total time devoted to field work was in the order of 6–10 months each, whereas processing, sorting and identification of samples required double the time and more, even when assisted in handling the samples. Moreover, in the absence of valid literature, identification entailed repeated visits to large research collections held at natural history museums in various countries (see below). All these steps required substantial resources, not least those invested by collection curators to give advice. Therefore, any means to speed up the process of identification will be of significance for progress in tropical biodiversity research. Moreover, it is important to make the data and experiences accumulated during all larger projects internationally available and free of charge. Then, scientists who continue with similar work can build upon the knowledge that has been gained so far, even at other sites or under different institutional coverage.

5.3 Taxonomic Identification of a Diverse Insect Group: a Case Study

A bottleneck for handling megadiverse situations is the problem of obtaining reliable species identifications. Brehm et al. (Chapter 2 in this volume) and Krell (2004) discuss why proper identification is required and why the exclusive use of morphospecies is an inappropriate shortcut. Comprehensive identification literature is usually not available for most groups of arthropods and other species-rich taxa. In the case of geometrid moths, for example, most original species descriptions date from the late 1800s and early 1900s, the 'golden age' of taxonomy (Gaston et al. 1995). Unfortunately, standards for species descriptions then were usually rather low as seen from today's perspective. The species were often described in a few sentences without examination of important morphological diagnostic characters, illustrations or comparisons with other species. As a consequence, such original descriptions are usually a very poor identification tool.

The most reliable and often only way to identify geometrid species is by comparison of specimens or digital images (Fig. 5.1) with type material deposited in natural history museums (e.g. Brehm 2002; Brehm et al. 2005; Chapter 2 in this volume). However, such an approach is time-consuming and costly since much material needs to be examined in many museums. For this work, hundreds of pre-sorted morphospecies needed to be compared with thousands of specimens in different museum collections: finally, a total of 1266 species were sorted, 63% of which were assigned to known species (Brehm et al. 2005). Since the most relevant museums were visited (most twice), this percentage actually indicates that ca. one-third of the species are likely to be new to science and need to be described taxonomically. Fortunately, type specimens of neotropical geometrid moths are rather concentrated in a few museums [the most important being the Natural History Museum (London), the United States Museum of Natural History (Washington D.C.), the Zoologische Staatssammlung (Munich), the American Museum of Natural History (New York), the Senckenberg Museum (Frankfurt), the Humboldt Museum (Berlin)]. Species identification becomes of course more difficult for those taxa in which type material is scattered in many museums around the globe.

Preliminary identification would greatly be enhanced if digital images of all type material deposited in museums became available in the Internet, since this allows one to compare samples quickly and cheaply, and to prove existing identifications. Although such projects are on their way for some insect taxa, such as butterflies (Lamas et al. 2000; Häuser et al. 2004), the vast majority of Neotropical insect type material is still not accessible via online databases. Though undoubtedly extremely useful, putting digital images of type material on the web is not a solution for all problems: Such images cannot replace proper identification literature written by experienced taxonomists. Moreover, good identification literature usually does not illustrate the often odd and worn types but proper specimens belonging to the same species. High standards in identification literature as set in the European fauna (e.g. Hausmann 2001) are still a far vision for most tropical Lepidoptera and insects in general, particularly when considering: (a) the recent retirement (without replacement) of taxonomists working on neotropical Geometridae, and (b) the general neglect of taxonomy as a fundamental biological discipline.

5.4 Imaging of Plants and Moths

High-resolution digital images of specimens and samples of many groups of organisms can nowadays easily be obtained, stored, and administered (Basset et al. 2000; for best practice, see Häuser et al. 2005). Relational databases allow easy data access, data exchange, and the publication of images and data on the Internet. Identification tools can be incorporated in such databases and are available in the

Fig. 5.1 Sixteen (out of 102 species collected in the RBSF and adjacent areas) belonging to the very species-rich geometrid genus *Eois* (*amarillada*, *angulata*, *antiopata*, *azafranata*, *basaliata*, *binaria*, *biradiata*, *borrata*, *burla*, *camptographata*, *chrysocraspedata*, *ciocolatina*, *cobardata*, *cogitata*, *contraversa*, *encina*; Brehm et al. 2005). No identification literature is available for most tropical insects, including *Eois*. Digital images (three megapixels, Nikon Coolpix 990) of the upper- and undersides of more than 1200 species of geometrid moths were taken and used for identification in museums and for documentation. The *Eois* images are provided on the Internet (http://www.personal.uni-jena.de/~b6brgu2). *Bars* 10 mm. Images taken by G. Brehm

field for local researchers. During the course of our research in Ecuador, fresh plant specimens were scanned soon after collection. The difference between such fresh specimens and flattened, brown herbarium specimens often was considerable. Pigment colors in moths and butterflies usually fade less rapidly than in plants if the specimens are protected from light.

Appropriately spread Lepidoptera specimens display most parts of their wing patterns (Fig. 5.1), and moths can usually be photographed months or even years after collection without a significant loss of quality. Images taken in the field can provide additional and highly valuable data, e.g. about behavior. However, such images are much more difficult to obtain than images of pinned specimens. Unlike the small and secretive moths, sedentary organisms such as plants can relatively easily be photographed in the field. Combined field and herbarium images provide a rather complete visual representation of the respective individual. Close-ups of flowers or fruits provide important characters in high resolution. It is well known that images are generally easier to memorize than pure text and 'dry' sets of characters. They effectively help scientists to recognize their target organisms in the field.

5.5 Data Processing and Accessibility

In species-rich regions, fieldwork results in a huge number of collected specimens and associated meta data. Each collected specimen must be labeled with data such as country, region, habitat, elevation, geographical position, time of collection, and collector. The specimen should also be labeled with its individual database number and the assigned scientific name. All these meta data are also stored in a database together with further information (e.g. morphological characters, where and when identified, etc.) and digital images. It is evident that databases are much more efficient for the storage and analysis of data than spreadsheet software. Moreover, typing efforts and the risk of misspellings are reduced. A rather complete standard for the design of taxonomic databases is available from the Taxonomic Database Working Group (TDWG; for a description of the standards, see: www.bgbm.org/TDWG/CODATA/Schema/).

Biodiversity informatics is a rapidly growing field for the data analysis of large collections stored in natural history museums. However, many of the initiatives are related to specimen data that are already stored in collections. Data availability is often restricted to experts and/or local networks, and not all tools are intended for fieldwork at a relatively low level of taxonomic knowledge. In research networks such as in the RBSF, data accessibility between all participating groups is of crucial importance. Site-specific conditions of a research area and available facilities define how data can be shared among the researchers. If the Internet can be accessed, a web-based database is most appropriate (e.g. as provided by the Arthropods of La Selva project, available at: http://viceroy.eeb.uconn.edu/ALAS/ALAS.html). However, in smaller research stations in tropical areas that lack permanent Internet access, a local version could be the best way to share information.

5.6 'Visual Plants' – a Biodiversity Database for Non-Specialists

'Visual Plants' is a database that specializes in storing information on plants. The program can be installed on local computers and is used in the research network at the RBSF. It is also available on the Internet (www.visualplants.de). The main intention of the database is the presentation of images together with their meta data. Attached characters describe the specimen to allow a pre-selection of species based on vegetative characters, such as life form, leaf arrangement, or flower color.

A set-based query tool allows the interactive minimization of the number of specimens or field images stored in the database through the selection of character statements. Although the system is not a complete identification system, it is a useful tool, particularly for non-specialists and where other sources for identification (floras, etc.) are absent. It publishes specimens collected in the field, even if they are not identified at species level, since this might ease the discovery of new species. Hopefully, the database will be expanded to taxa other than plants in the future to provide a multi-tool for preliminary species identification of a broad range of organisms in the field.

5.7 Conclusions

Megadiverse ecosystems such as tropical montane forests in Ecuador pose major methodological challenges with their species-richness, incomplete sampling, and difficulties in the identification of species. Digital images taken from the type specimens as well as from fresh specimens can greatly ease identification. The availability of additional information, e.g. morphological characters, is essential, preferably organized in a database. However, such tools still need to be developed for most groups of organisms.