

Biodiversity inventory and informatics in Southeast Asia

Campbell O. Webb · J. W. Ferry Slik · Teguh Triono

Received: 17 September 2009 / Accepted: 13 February 2010
© Springer Science+Business Media B.V. 2010

Abstract Rapidly changing land use in Southeast Asia threatens plant diversity, and reduces the time we have left to document it. Despite over 200 years of scientific plant exploration, many plant species have yet to be discovered. Moreover, we still have a very poor understanding of the distribution of known taxa in this biogeographically complex region. We review the current state of biodiversity exploration, using plants in Indonesia as an example. Traditional methods of collecting and describing species have provided a solid foundation for our understanding of plant biodiversity, but are insufficient for the pragmatic task of rapidly discovering and documenting today's biodiversity before it is gone, because general collecting expeditions tend to be infrequent, and documentation of most new species must await taxonomic revisions many years in the future. Solutions to this exploration and documentation crisis (i) could use the abundant resource of enthusiastic, networked, national biology students, (ii) should employ biodiversity informatics tools to efficiently engage both specialists and parataxonomists, and (iii) might require adoption of new types of α -taxonomy, utilizing increasingly low-cost molecular methods and high resolution photographs. We describe emerging technologies that will facilitate this

Electronic supplementary material The online version of this article (doi:[10.1007/s10531-010-9817-x](https://doi.org/10.1007/s10531-010-9817-x)) contains supplementary material, which is available to authorized users.

C. O. Webb (✉)
Arnold Arboretum of Harvard University, 22 Divinity Avenue, Cambridge, MA 02138, USA
e-mail: cwebb@oeb.harvard.edu

C. O. Webb
Herbarium Bogoriense, Bogor, Indonesia

J. W. F. Slik
Key Laboratory in Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Yunnan, China
e-mail: ferryslik@hotmail.com

T. Triono
Herbarium Bogoriense, Research Center for Biology (LIPI), Cibinong Science Center,
Cibinong, West Java, Indonesia
e-mail: tt_rekso@yahoo.com

taxonomic development. We believe that a new golden age of biodiversity exploration may be dawning, just as biodiversity itself is most threatened, and are hopeful that increasing knowledge of biodiversity will be a positive force to slow its loss.

Keywords Biodiversity crisis · Tropical forests · Species discovery · Citizen science · Parataxonomists · Digital photography · Distributed databases · Species identification · DNA barcoding · Semantic web

Introduction

Accompanying the rapid loss of biodiversity in Southeast Asia is a crisis in biodiversity knowledge. In many taxonomic groups, the delimitation of species is still unclear, and we understand little of their distributions and potential uses. While knowledge about biodiversity is not a prerequisite for protecting it, unlike boundaries, policies and policing, information about biodiversity is important. Making informed management decisions always requires some level of biodiversity data (Mace 2004), and many believe that ultimately we have a moral responsibility to know and steward the other taxa we share the planet with. Yet field inventory proceeds slowly and is uncoordinated. Traditional taxonomic revisionary activity is restricted to a few taxa. Meanwhile, as the rest of this special issue attests to, the rapid conversion of natural habitats in the region is continually reducing the time we have left to discover new species and patterns of biodiversity. Nevertheless, we are optimistic that the rate of basic biodiversity exploration and description could be increased significantly, by a combination of (i) mobilization of large numbers of biology students in the region, and (ii) the use of cutting-edge, open-source tools for biodiversity informatics. We also suspect that the adoption of new models of taxonomy will be needed, simply to cope with the backlog of unrevised specimens in herbaria and museums and with new collections.

In this essay we will briefly review the history of biodiversity exploration in the region, examine components of the biodiversity information crisis, and then look at potential solutions for increasing the rate of discovery. The biodiversity literature is full of passionate and thoughtful analyses on the need for a rejuvenation of the taxonomic foundations of biodiversity science (Wheeler 2004; Raczkowski and Wenzel 2007; Mayo et al. 2008). We aim here to bring these ideas to a more local focus. We draw our examples primarily from botanical studies in Indonesia, but our conclusions and suggestions are likely to be valid for most tropical organisms.

Biodiversity discovery in the region: botanical history

Although the oldest known tropical flora, covering ca. 80 plant species from southern China and Vietnam, was written in 304 AD by Chi Han in China, the first efforts to produce comprehensive, all inclusive floras were undertaken many centuries later after the arrival of European scientists in the region. These included Georg Eberhard Rumphius and Hendrik Adriaan van Rheede, two of the earliest botanists in Asia (middle to late seventeenth century) who were both employed by the Dutch East Indies Company and worked in the Moluccas and South India, respectively. They produced two important floras, the ‘Hortus Malabaricus,’ a 12 volume flora of Southern India that included 742 plant species, and the ‘Herbarium Amboinense’ which covered the flora of the Moluccas and included ca.

1,200 plant species. Rheedee's book formed an important reference for the Asian plant species included in Carolus Linnaeus 'Species Plantarum' that was published in 1753.

During the eighteenth and nineteenth century many expeditions were launched into remote areas of Asia to explore the geography and discover and collect 'useful' plants and animals, while simultaneously a lively international business in exotic plants and animals developed. During this period the first regional herbaria were established and started publishing local floras, usually in cooperation with large herbaria in Europe and the USA. A big problem that remains to date is that most of these floras covered political rather than biogeographic regions, resulting in fragmented taxonomic treatment of plants and inefficient use of taxonomic knowledge and resources present across the region and other parts of the world. One notable exception is the *Flora Malesiana* project, started in 1948, that covers several biogeographic regions (the Sunda Shelf, Wallacea, the Sahul shelf, and the western Pacific) including many islands and countries, with Indonesia at its core. Although contributions for this flora are produced continuously by many botanists working at several major herbaria in the region and around the world, progress is slow (8,008 of the estimated 35,000 non-orchid plant species were included up to 2001).

The term 'Malesia' was first coined by Zollinger (1857), who also proposed botanical subdivisions within the region. The history of botanical exploration in this area is described in detail in the first volume of the *Flora Malesiana* (van Steenis-Kruseman 1950). In addition to that by G.E. Rumphius, early botanical work in Malesia was conducted by Jacobus Bontius (Java), and George Joseph Kamel (Philippines; (Philippines; Frodin 1984). However, large scale botanical exploration only started late in the eighteenth and early nineteenth century, after the establishment, in 1778, of 'Bataviaasch Genootschap van Kunsten en Wetenschappen' (Batavian Society for Arts and Science), and especially the Bogor Botanical Garden (in 1817) and the Herbarium Bogoriense (in 1844). Most floristic work was initially focused on Java and Peninsular Malaysia, with exploration of the other islands gradually increasing during the nineteenth century. Major contributors during this period were C.G.C. Reinwardt, C.L. Blume, J.E. Teijsmann, J.K. Hasskarl, H. Kuhl, J.C. van Hasselt, W.S. Kurz, A. Zippelius, P.W. Korthals, F.W. Junghuhn and H. Zollinger. Blume (1827) produced the first flora of Java, followed later by the first overall attempt to cover the whole flora for the Dutch East Indies by Miquel (1859). This first attempt to cover most of the Malesian region was necessarily incomplete due to lack of sufficient collections, and the task proved too large for one man to finish. Most botanical works that followed were again focused on single islands or small political regions.

Plant collecting continued and was enhanced by the establishment of forestry services in the early twentieth century, and new collecting expeditions on most islands in the region. The growing numbers of collections and increasing international cooperation between taxonomic institutions and botanists finally resulted in a new effort to produce a flora for the whole Malesian region, with the launch, in 1948, of the *Flora Malesiana* and its accompanying *Flora Malesiana Bulletin*. However, the slow pace of progress of the *Flora Malesiana* has recently lead to a renewed focus on local floras bound by political boundaries, such as the Tree Flora of Sabah and Sarawak, and the Flora of Malaysia. Meanwhile, the Malesian region remains patchily explored, with a strong research and collection focus on the west (Java, Sumatra, Borneo and Peninsular Malaysia) but limited attention to central and eastern Malesia. Although the *Flora Malesiana* region includes some of the most densely populated and fastest developing nations in the world, the Asian tropical flora remains one of the least studied and least known.

The crisis in biodiversity inventory

There are many species yet to discover

Current estimates for the total number of extant species range from five to over 30 million (May 1988; Mace et al. 2005), with arthropod estimates alone ranging from 4.8 to 31.0 million (Erwin 1982; Novotny et al. 2002). With only ca. 1.75 million species described (Mace et al. 2005), of which ca. 350,000 are plant species (Wortley and Scotland 2004), we clearly are a long way from being able to circumscribe the planet's biota. For some well-known groups (e.g., mammals and birds) finding a new species is newsworthy, while for others (e.g., beetles), a single collection event can turn up hundreds of undescribed species.

The simplest way to assess the state of our current knowledge is to observe the number of new species published in recent revisions. As an example, in 28 recent revisions in the journal *Blumea* (see online appendix) the number of new species published ranged from zero out of 15 species (Gardner 2006) to 20 out of 54 (37%; de Wilde and Duyfjes 2006), with a median of 10%. However, this figure is likely to be an underestimate of new taxa discovered during the revision process, since many taxonomists publish their new species separately from (and before) the revisions themselves.

Estimating the number of *unknown* species in a group proceeds by extrapolating from known numbers of species, under some set of assumptions. Some methods use ratios of subsets to whole (by area or taxon; Stork 1997; Adamowicz and Purvis 2005), while others use extrapolation from samples (Erwin 1982; Novotny et al. 2002), or extrapolation from declining rates of new-species description (e.g., Giam et al. 2010, this issue). Using the latter approach, Dolphin and Quicke (2001) predicted an increase of 100% to 200% in species numbers for Braconidae wasps, and Meier and Dikow (2004) predicted a 36–41% increase for *Euscelidia* flies.

Another approach to estimating the total number of taxa in a taxonomic group is to use general nonparametric estimators of classes in a population (Chao 1984). If we assume that the total set of specimens collected for a taxonomic group over time, that are reviewed during revisionary work, are a random subset of all individuals of this group in the field, then we can estimate the total number of species (S_{tot}) based on the number of species known only from their type (f_1 ; 'singletons') and species with only two collections (f_2 ; 'doubletons'), using a simple formula: $S_{\text{tot}} = S_{\text{known}} + f_1^2/2 f_2$ (Chao 1984), or a revised version $S_{\text{tot}} = S_{\text{known}} + f_1 (f_1 - 1)/2 (f_2 + 1)$, which works when there are no doubletons (Chao 2006). Groups with many undescribed species will have a higher proportion of 'type-only' species ('singletons') than well-sampled groups, for a given number of collections.

We applied this method to recent plant genus (and subgenus and section) revisions for the Malesian region published in *Blumea*, which encourages authors to publish full lists of all specimens examined in making the revision. Of the 68 cases we found (see online Appendix A), 23 higher taxa had some singleton, or 'type-only' species. Fifty-five of the 68 taxa were predicted to be 'complete,' i.e., to have no undiscovered taxa, while 13 were predicted to have from 0.5 to 6 undiscovered species, representing 1–24% of their current richness. Overall, the 68 groups analyzed contained 532 revised species, and were predicted to have 45 (9%) undiscovered species among them. We found a strong positive relationship between the number of known species and the number of predicted unknown species ($r^2 = 0.4104$, $P < 0.0001$), indicating not surprisingly that in our sample, diverse clades tend to have more unknown taxa in them than clades with few species. This suggests

that because it is mainly the smaller, lower-diversity taxonomic groups that have been tackled first (Roos 2003), we can expect the true number of unknown species to be far higher than the 9% in this small sample of taxa. High type-only rates can be found in revisions of larger genera (e.g., ca. 50% in *Croton* of Madagascar, B. van Ee, pers. comm.), and in pre-revision checklists of diverse regions, although synonymy rates can be very high in unrevised groups (e.g., 4,861 synonym names for 1,006 species in F. Slik's Southeast Asia tree checklist;¹ see also Scotland and Wortley 2003). Additionally, collecting in poorly sampled areas can still turn up startlingly high rates of new species: a recent trip to Gag island in West Papua collected 14 palm species of which nine were new, one even being in a new genus (*C. Heatubun*, pers. comm.). Higher plants are relatively well explored compared to many other taxonomic groups, and it is likely that the majority of, for example, fungi species in Southeast Asia have yet to be discovered (Hawksworth and Rossman 1997).

We don't have much time left

This special issue is organized around the premise that the biodiversity of Southeast Asia is seriously threatened (e.g., Bickford et al. 2010; Wilcove and Koh 2010; Peh 2010), and numerous other papers attest to this fact (e.g., Sodhi et al. 2006). Indonesia, perhaps the country with the highest total biodiversity in the world (marine and terrestrial combined), was estimated to have lost 72% of original vegetation cover up to 1997 (Bryant et al. 1997). While there is great uncertainty in current species extinction rates (May et al. 1995) and it is notoriously difficult to prove the extinction of a particular species, there are habitats in Southeast Asia that either simply do not exist any more or are drastically reduced in area (Sodhi et al. 2006), and there can be little doubt that many species are going extinct before we ever record them, and that this situation is getting worse rather than better. While there may remain alive a few individuals of taxa restricted to these habitats, they represent the 'living dead' (Janzen 1986) and the eventual extinction of their species will occur shortly (but see Collar 1998). Collecting trips to these small remnants of forest in areas where all other forest cover has gone should be a high priority. Habitats under particular threat are forests on alluvial soils on Borneo and Sumatra, many lowland forest types in the Philippines (Sohmer 1996), and any lowland forest on Java. Nevertheless, these remnants will already have lost many species that are sensitive to local environmental stresses (Sodhi et al. 2009), and trips to the larger areas of intact habitat are clearly needed to find the majority of undetected species. In some parts of Southeast Asia we probably have only a decade or two to record many taxa before they are gone forever.

Current biodiversity inventory activity

Despite the general concern that biodiversity activities do not attract a level of funding commensurate with the scale of the biodiversity crisis, it is likely that national funding for taxonomic work has been steadily climbing in most countries in the region. In recent years this is partly due to inputs from large international initiatives (e.g., Global Environment Facility and Convention on Biological Diversity, CBD). The number of plant collections from Indonesia arriving each year at the Leiden Herbarium has increased during the last two centuries (Fig. 1), and Herbarium Bogoriense itself (BO) is currently receiving ca. 1,500 Indonesian collections a year. Activity by foreign systematists has generally been

¹ <http://www.asianplant.net/>.

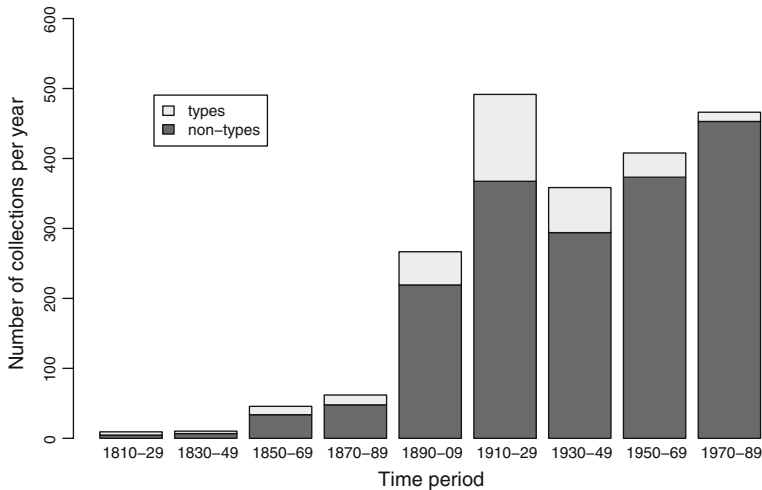


Fig. 1 Indonesian collections arriving at Leiden herbarium, 1800–1989. Using the online collection database of the Nationaal Herbarium Nederland, Leiden University Branch, we counted the number of collections and types from Indonesia in each year from 1800 to 1989, for a standardized subset of taxa that were databased in a comparable way: specimens that were sent on loans to other herbaria, and some targeted digitization projects (see online Appendix B)

declining, and the increase is due to collections by local botanists. Despite this apparently hopeful trend, the question remains whether the ongoing activity is sufficient to visit and collect in the areas that are most threatened, and are least explored. In Indonesia, current plant collecting is either (i) by small numbers of foreign systematists on small expeditions (ca. 10–15 trips a year), (ii) from a few larger expeditions, either funded by government institutions (e.g., recent BO expeditions to West Papua) or NGOs (e.g., the Conservation International expeditions to Mamberamo), or (iii) a product of long-term collaborative biodiversity projects (e.g., the STORMA project in Central Sulawesi). Specialists in particular taxa, both national and foreign, probably visit the widest range of sites, but usually make few general collections. Overall, a relatively few number of sites are being targeted for general collecting. The two highest diversity regions in Indonesia, Kalimantan (on Borneo) and Papua (on New Guinea), are definitely not being visited at a frequency proportionate to the rate of forest degradation. Kalimantan, while better collected historically than Papua (but less than Sabah and Sarawak, Raes and Steege 2007), has the additional constraint that due to the supra-annual mast flowering that dominates the phenology of plants in the lowlands (Appanah 1993), fertile material is only available for many taxa once every five or more years.

While the rate of deposition of specimens has increased, the rate of deposition of types per year has been steadily declining since the early 1900s (Fig. 1), and the number of types recorded per collection has been declining since 1800, indicating that increasingly more collecting effort is needed to find new species. However, the most recent rate estimate, of one new species per 40 collections (see Fig. 1: 1970–1989), while lower than the one-in-two rate in the early 1800s, is still amazingly high, indicating in a different way the large number of new species yet to be found.

This decline in types-per-collection may also indicate a reduction in revisionary activity, even as the absolute rate of specimens collected increases. The ‘easy’ plant groups

Table 1 Specimen records server through GBIF for the countries of Southeast Asia and New Guinea (as of 5 September 2009)

	Total records	Georef. recs.	Species
Indonesia	458,042	107,845	51,911
Philippines	399,809	113,799	37,034
Papua New Guinea	494,500	266,371	29,756
Malaysia	130,304	35,478	19,548
Thailand	149,358	36,743	16,463
Viet Nam	104,937	30,135	12,380
Myanmar	39,477	17,022	6,594
Singapore	13,399	4,507	4,452
Lao PDR	36,118	7,920	3,331
Cambodia	9,990	1,122	2,482
Brunei Darussalam	6,228	629	1,746

have largely been tackled by the *Flora Malesiana* project, leaving the large, complex genera in the large, complex families (Roos 2003). Even if the rate of species revision were to continue as it has since the 1950s, it would take at least another 125 years for *Flora Malesiana* to be completed (Roos 2003). Globally, Stork (1997) similarly calculated that even using the lowest estimate of species diversity of three million, it would take another 90–120 years to describe the planet's biodiversity. Yet, in many countries (especially Western ones) the number of active 'alpha-taxonomists' is declining (Lee 2000; Hyde 2003), as younger systematists are both drawn into other sub-disciplines of taxonomy, and unable to find education and apprenticeship in traditional descriptive taxonomy (the 'taxonomic impediment,' Wheeler 2004; but see de Carvalho et al. 2005). Locally, we are pleased to note that the number of active (publishing) taxonomists in the Indonesian government's Research Center for Biology is increasing each year (currently ca. 15 botanists and 20 zoologists). However, this group can still only tackle a small number of taxonomic groups, given the staggering biological diversity of the country.

Taxonomic information is of little value if it is not accessible. The internet revolution has fundamentally changed the methods and approach of biodiversity inventory work (Mayo et al. 2008). Many herbaria and museums now serve their collections data via the DiGIR² or TAPIR³ transfer protocols and the clearinghouse of the Global Biodiversity Information Facility⁴ (GBIF) network, which now offers data for over one and a half million collections originally made in Southeast Asia (Table 1). However, almost all of these data are served by institutions outside the region, representing their own collections; as of 5 September 2009, only one institution in the region was acting as a data provider to GBIF: Raffles Museum of Biodiversity Research (RMBR) in Singapore (13,477 records). There are a number of significant internet-accessible databases we are aware of (e.g., the type specimen databases of Herbarium Bogoriense and Museum Zoologi Bogor,⁵ Indonesian Invasive Plants database,⁶ the Indonesian Marine Conservation Database,⁷ the

² <http://digir.net/>.

³ <http://www.tdwg.org/activities/tapir/specification>.

⁴ <http://www.gbif.org/>.

⁵ <http://ibis.biologi.lipi.go.id/>.

⁶ <http://www.biotrop.org/database.php>.

⁷ <http://www.konservasi-laut.net/>.

RMBR moss databases,⁸ and the collections of the Singapore herbarium⁹) but these are searchable only via a web-browser interface, thus limiting the rate of access to the underlying data.

Increasing the rate of biodiversity discovery and description

We believe a vast increase in biodiversity exploration is necessary if we are to even begin to document species and their distribution in the region before they become extinct. We also believe that the process of discovering and documenting biodiversity could be a major positive force for conservation (Novacek 2008), by engaging much larger segments of the population of Southeast Asia and publicizing what biodiversity is and why the region should be very proud of, and careful with, its species and habitats. Three components of a solution to the biodiversity inventory crisis are (i) more collecting, in more places, by parataxonomists and citizen scientists, (ii) development of biodiversity informatics tools and local expertise in these tools, and (iii) joining the global revolution in how descriptive taxonomy is done, as a pragmatic answer to the overwhelming quantity of data.

More collecting, done differently

The use of trained technicians, often known as ‘parataxonomists,’ has long been offered as a solution to the low absolute numbers of taxonomists (Oliver and Beattie 1993; Basset et al. 2000; Pfeiffer and Uril 2003; Baraloto et al. 2007). Indeed, almost every collector has used local assistants, who often become masters in taxon identification and collection methods, and can then operate independently for far lower cost than the professional. As an example, our current biodiversity inventory work at the Gunung Palung National Park (West Kalimantan, Indonesia) is executed mainly by a team comprising a park ranger, a local university graduate, and local villagers, who make high quality physical collections and document plants with numerous digital photographs. However, parataxonomists, with a direct relationship to a professional biologist, will always be too few to become the ‘movement’ that is necessary to document the region’s biodiversity. What is needed is to engage the general public in ‘citizen science’ activities where non-professionals collect and contribute data to serious scientific activities (Cohn 2008; Couvet et al. 2008). Most notably, amateur bird watchers have contributed greatly to bird distribution databases in the US and Europe (e.g., eBird¹⁰ and the Avian Knowledge Network, Kelling et al. 2009), and elsewhere (Greenwood 2007), and numerous other projects are beginning to utilize this vast human resource for serious science (e.g., National Phenology Network,¹¹ Mushroom Observer,¹² The Great Sunflower Project,¹³ FLORON,¹⁴ Reefcheck,¹⁵ and numerous local

⁸ <http://rmbr.nus.edu.sg/databases/>.

⁹ <http://dps.plants.ox.ac.uk/bol/singapore>.

¹⁰ <http://ebird.org/>, and <http://www.birds.cornell.edu/netcommunity/citisci/>.

¹¹ <http://www.usanpn.org/>.

¹² <http://mushroomobserver.org/>.

¹³ <http://www.greatsunflower.org/>.

¹⁴ <http://floron.nl/>.

¹⁵ <http://reefcheck.org/>.

‘Bioblitz’ surveys¹⁶). Some may consider it unlikely that many in the Southeast Asian population would be interested, or have the luxury of the time or funds to participate in biodiversity citizen science, but this view ignores the strong attractions of having an excuse to spend time in Nature, contributing to a larger project that one can believe in, and the liberation and empowerment that scientific thinking itself can offer (Nanda 1997). Indeed, citizen science has been taking place for many years in Singapore (Wee and Subaraj 2009), Malaysia (with the Malaysian Nature Society) and Indonesia (with activities of the Indonesian Ornithological Union, and Carnivorous Plant Community).

We believe however that the much larger, untapped resource is school and university students. An example in a lower-income country is a successful ant survey project in South Africa that has involved over 3,000 high school students (Braschler 2009). In our experience, university students studying biology in Indonesia are immensely enthusiastic about their subject and often travel, at their own expense, to participate in small surveys in conservation areas. They are eager to collect high quality data, but most are unaware that the data they already collect, with a bit more coordination and standardization, could be integrated into global databases. We foresee a loose network of active university students, identifying important areas to visit via internet resources, sampling, taking digital photographs, collecting metadata, and making standard collections (including DNA vouchers). In our experience, many students have made physical collections as part of their course and dissertation work, but these specimens seldom make their way to herbaria and museums, more often because of lack of knowledge that they will be welcomed than for reasons of cost. The data these students collect could then be almost immediately entered into community biodiversity databases.

There are a number of important issues to consider when working with non-specialists:

Coordination—Managing the activities of a network of citizen scientists is important to reduce redundancy in activities and to target those activities to the most important sites. The internet greatly facilitates this coordination. Indeed, we believe that we are arriving at a key moment for biodiversity inventory because it is only in the last few years that the majority of university students have come online, even in the more remote parts of the region. A good web-portal will be required to share and present data.

Personal and social motivation—Participants must feel like they are contributing, increasing their own knowledge and experience, and rising in the estimation of their peers. A well-designed web portal for coordination and data uploading should contain elements borrowed from social networking sites (e.g., Facebook), so that participants can see what others are doing, how they rate against others, and can contribute to others via messaging and offering e.g., new taxon identifications.

Quality control—Some professionals may recoil at the thought of using amateur-collected data, and data will probably vary widely in quality from professional-level to error-ridden (Cohn 2008). The goal should always be to raise the quality with education, but filtering of data simply becomes another informatics component (Caruana et al. 2006). Recording and displaying participant activity and prior quality can be used to contribute to a public reputation, and can in turn be used as a quality filter (van House 2002; Udell 2008; Lawrence 2009).

Retention—The ubiquitous abundance structure of ecological communities (a few common species; most rare) means that a collector encounters personally new species increasingly infrequently, and the ability to disregard common, previously collected

¹⁶ See links at: <http://citizensci.com/>.

species can only be gained with experience. Thus keeping participants through the years will increase the efficiency of a community collecting program. Fortunately, the longer a person has been collecting, the more likely they are to be motivated by a personal desire to find new taxa and to want to keep active.

Funding—There is no doubt that many interested students and older citizens in the countries of Southeast Asia will be unable to participate in a community biodiversity science initiative due to a lack of funds to meet even minor travel costs. Securing funding, ideally from within the region's countries, will definitely be important. But the real costs are relatively low and we hope that government and non-government funding sources will recognize the great benefit of such programs for education and conservation. All governments within the region are party to the Global Taxonomy Initiative of the CBD, and are already spending large sums on biodiversity research, some small part of which would fund a lot of community science. Additional funding would also be needed for the herbaria and museums to handle the large increase in specimen input from a successful citizen science initiative.

Another sector of the population that it is vital to encourage to contribute data are people whose job it already is to spend time looking at plants and animals, but who would not describe themselves as taxonomists: the foresters, tree cruisers, ecologists and conservation workers. These professionals often have a better field knowledge of the flora or fauna of an area than professional systematists, and when they encounter a species they do not know, this event should be recorded. It is quite possible that many of the trees in ecological plots that are never identified beyond morphospecies, despite the attention of good field botanists, actually represent new species.

Biodiversity informatics in Southeast Asia

'Biodiversity informatics' (BI) is "an emerging field that applies information management tools to the management and analysis of species occurrence, taxonomic character, and image data" (Johnson 2007). In this information-rich age, it is inherently exciting, and has developed into a field rich with ideas and promise in just a few years (Soberón and Peterson 2004; Page 2006; Guralnick and Hill 2009). Behind much of the excitement is a vision of finally understanding the mysteries of the biosphere, as all the world's biodiversity data are linked together and analyzed. It is often driven by an 'if we build it they will use it' attitude, which is, however, no different than the way that many now-indispensable web technologies started. Despite the rapid and deep adoption of cutting-edge information and communication technology (ICT) in the region, and the large numbers of young programmers being produced, little of the rapid development in BI is happening in Southeast Asia. We see this simply as an incidental failure of interested parties to connect rather than as indicative of a potential lack of interest or funding. In fact, the great majority of BI uses free and open source tools, and with laptop ownership increasing, there are no significant resource barriers to developing a vibrant local BI community. We will now focus on various aspects of BI that are important for increasing biodiversity inventory efficiency in this region, using as an example a developing biodiversity informatics platform for plants of Gunung Palung National Park.¹⁷

¹⁷ <http://phylodiversity.net/xmalesia/>.

Decentralization and standards

The keys to efficiently delivering biodiversity data over the internet are standardization (using a machine-readable data format) and automation (allowing other machines to access that data via a 'web service'). An effective example of this procedure is the use of the DiGIR protocol to transfer data between a querier (e.g., the GBIF clearinghouse) and a data server (e.g., a museum collections database). Structured data are most commonly represented today in XML (Extensible Markup Language), a simple notation that labels each datum in a hierarchical manner. To standardize the structure of the XML file, and validate instances of the data, a schema is developed.¹⁸ A machine can then easily request data, and parse the data objects into the appropriate fields of its internal data representation (often in a Standard Query Language, SQL, compliant database). The Taxonomic Data Working Group (TDWG¹⁹) maintains and develops these standards for biodiversity data interchange. However, XML documents can also easily be translated between formats using XSLT (XML Stylesheet Transformations), should it be desirable to use a custom schema.

Armed with this technology, the decentralization of data becomes effective, and advantageous. Rather than a single national database holding all records, with the risk of data being withheld and the difficulty of data update, individual institutions can manage and serve the data for which they are responsible, delivering these data via standard transmission protocols, i.e., using the appropriate 'cyberinfrastructure' (Graham et al. 2008).

The actual data that we need to record and transmit include the standard information about collections (where, when, by who), but increasingly also digital images (or at least their Universal Resources Identifiers, URIs), morphological character, DNA sequence, growth and phenology data, and, importantly, the legal and intellectual property metadata for each collection.²⁰ Storing taxonomic name information is a more complex problem than might first be thought, because of the multiple names applied to a collection through time and the multiple synonyms for most taxa. The keys are (i) to store a determination history for each specimen or individual, and (ii) to internally store as little information about the taxonomic hierarchy of this name as possible (e.g., family), instead querying dynamic taxonomic databases over the web for the latest taxonomic placement of a species name.

Image and morphology data collection

The digital revolution offers biologists working in remote areas access to international literature and collections. The Biodiversity Heritage Library project²¹ is working to make primary taxonomic literature available to all. Many museums and herbaria are working to provide digital images of their collections. While a digital image can never be as data-rich as a physical specimen, because micro-structures and internal anatomy are hidden, a comprehensive set of high resolution images can offer access to the full set of diagnostic characters for many taxa, and can even be sufficient for taxon description. There is great value in taking good photographs of plants in the field, even if a physical collections cannot

¹⁸ For example, the DiGIR schema: <http://digir.net/schema/protocol/2003/1.0/digir.xsd>.

¹⁹ <http://www.tdwg.org/>.

²⁰ See, e.g., the schema at: <http://phylodiversity.net/xmalesia/schema/xmalesia.rnc>.

²¹ <http://www.biodiversitylibrary.org/>.

be made. However, the quality of images is key, as is comprehensive coverage of morphology (Baskauf and Kirchoff 2008). For our Gunung Palung flora project we are taking 10–20 images of each fresh plant: e.g., bark slash, whole twig, twig tip, twig surface, stipules, whole leaf (above and below), close-up of leaf base underside and petiole, inflorescence, flower (or fruit) from distal, side, and proximal angles, and a transverse and longitudinal section. We are also re-photographing the plants once dry, so that they can be more directly compared to herbarium material. Similarly, asking the collectors to score the specimens for a small (5–10), standard set of diagnostic morphological characters can assist greatly in the retrieval of digital collections for matching with an unknown collection.

Social networking and crowdsourcing

The explosion in popularity of Facebook and Twitter indicates the attraction of making social connections over the web, no matter how shallow. It is very likely that a community biodiversity inventory project that incorporates aspects of internet social networking will be more successful than one that does not. What we probably need is a ‘Facebook for biodiversity.’ Key elements are easy communication, ability to observe what peers are up to, challenges (“what plant is this?”), online measures of status or reputation (“how many collections and identifications have you made?”). As an example, our Gunung Palung flora project has had high levels of participation by invited systematists in its ‘collection alert’ facility. The systematist receives an automated email when a new collection has been added that has a tentative identification that matches his or her interests. The systematist can then visit the link in the email, see the images, and send back an identification. In this way, we have discovered one probable new species and several important range extensions.

While developing specialized software such as Facebook is an expensive proposition, existing general-purpose platforms can be adapted to create a social networking biodiversity platform. We can use, for example, a ‘content management system’ such as Drupal,²² or a Wiki platform, such as MediaWiki²³ (as used by Wikipedia).

Another recent social phenomenon on the web is ‘crowdsourcing,’ where problems are offered to the internet community, and taken up by enough people to be solved, usually for free. Suitable problems are ones that computers still find hard to solve, involving creativity or visual processing. A good example is the use of CAPTCHAs (the mini tests on many web forms that determine if the user is a human) to digitize hard-to-read text words from old manuscripts.²⁴ There are many challenges in biodiversity informatics that could benefit from crowdsourcing solutions, in particular the matching of taxon images during identification; important steps have been made in the automation of this process (Weeks and Gaston 1997; Kress 2004), but the human brain is still far more effective at comparing images. We can envision sharing out the labor-intensive step of sorting sample collections (e.g., insects from a Malaise trap), and even of the matching of morphotypes during the species delimitation phase. The challenge in designing a crowdsourcing solution is clearly to make it interesting or important enough to motivate participation. Even if an effective free crowdsourcing solution cannot be offered, a system involving small payments may work. A major comparative advantage for Southeast Asia is its huge number of available,

²² <http://drupal.org/>.

²³ <http://www.mediawiki.org/>.

²⁴ <http://recaptcha.net/>.

relatively low cost workers; there are many biology graduates in the region who would value a chance to make some money doing meaningful, biology-related work.

The Semantic Web

While the model of databases serving data to other databases in common structured formats has been powerful (e.g., a server speaking to GBIF with the DiGIR protocol), there is one more step to truly ‘free’ the data: it is to place the data in a universal format that could be understood by all users and machines. This format already exists and is the *lingua franca* of the ‘Semantic Web’ (Berners-Lee and Fischetti 1999), which is a vision of the internet as a domain of machine-readable knowledge rather than just text streams that can only be searched for substrings. Data in the Semantic Web (SW) exist as information triples: [SomethinghasPropertyPropertyValue], with each part of the triple having a unique identifier (which may be a web address). For example, [<http://museum.org/coll/13245botany:hasLeafInsertionterms:Opposite>]. The `hasProperty`, or verb, is part of a logical knowledge structure, or ontology, so that its meaning is understandable to humans and machines alike. Such a statement is now ‘free’ and can live in any database of knowledge statements anywhere; the triple is usually encoded in an XML format known as RDF (Resource Description Framework), and stored in an RDF ‘triple-store’ database. The RDF also includes metadata such as who made this knowledge statement and when. We can then imagine meta-databases (e.g., Tummarello and Morbidoni 2008) of all the world’s (triple-encoded) knowledge which can then be queried (via the SPARQL query language), e.g.: “What Somethings, that areIndividualPlants, have ever been recorded to have a property `hasLeafInsertion` with value `Opposite`?” A series of URIs will be returned, pointing to the digital representations of these plants, wherever they are.

The evolution of methods for modeling data has had an impact on the options available for biologists to share and access their data. SQL offers rigor for internal relational databases, XML allows structured data exchange among machines, and RDF now offers universal communication of concepts. We feel that adopting SW technology is important for reaching the most parties with our data. While many SW tools are developed for computer scientists, the Semantic MediaWiki²⁵ (SMW) and Semantic Forms²⁶ extensions to the popular MediaWiki platform offer the ability to annotate data easily with semantic meaning, in a shared Wiki environment. SMW provides for the import of namespaces, meaning that vocabulary for classes and properties can be imported from external ontologies, allowing in turn the export of universally interpretable RDF triples.

Training

As noted above, development in BI is not well represented in the region, but we feel that this reflects primarily a lack of awareness within the regional systematics community of the power, relative ease of development, and excitement that BI offers. There are already several centers for *molecular* bioinformatics in the region²⁷ (see also Tongsim et al. 2008; Zeti et al. 2009). Bringing programmers together with taxonomists in workshops, and encouraging participants to solve local biodiversity data problems, would be an important

²⁵ <http://semantic-mediawiki.org/>.

²⁶ http://www.mediawiki.org/wiki/Extension:Semantic_Forms.

²⁷ <http://www.bic.nus.edu.sg/>, <http://www.eijkman.go.id/Research/Bioinformatics>.

start, but perhaps more effective will be invitations to regional programmer-biologists to collaborate directly on exciting open source BI projects. Offering university courses in BI will also be important.

From inventory to species description

Beyond making more collections and sharing those collections digitally, the regional systematics community must address the vast backlog of historical collections that have not yet been considered in a monograph. Mountains of ‘indets’ exist in all museums and herbaria. New solutions are needed to the problem of species delimitation and description, not because traditional approaches are faulty—they have provided us with a taxonomic and phylogenetic framework that is solid, and nearly two million named species—but because the time available and urgency of describing the remaining millions of species pragmatically precludes a business-as-usual strategy. The need to develop a new ‘e-taxonomy’ is now widely recognized (Mayo et al. 2008), and a number of initiatives exist to develop online tools to speed taxonomic work (e.g. CATE,²⁸ EDIT,²⁹ Scratchpads,³⁰ and PBI sites³¹), most based around the idea of a ‘unitary taxonomy,’ a web resource where taxonomic information is collated and served (Godfray 2002; Scoble 2004; Clark et al. 2009). However, the problem remains of how to organize and describe the thousands of clades of tropical organisms that are not, and will not be, the focus of unitary taxonomy projects. Most of the current initiatives, while using the web creatively to manage and share data, still require academic taxonomists to make decisions and craft taxonomies, but there are just not enough professional taxonomists to do the work.

We do not yet know exactly what this new approach will look like, but key elements will probably include (i) substituting parataxonomists, and perhaps even the wider public of citizen scientists, to perform much of the ‘grunt work’ of taxonomy (character scoring, specimen sorting, digital curation; see above), (ii) using computational approaches to organize, delimit and classify taxa, based on morphological matrices, image recognition, and DNA sequence data (Hibbett et al. 2005), and (iii) accepting a different model of species description, that could live in parallel with the crafting of traditional Latin binomial species.

Molecular data are already the dominant method for reconstructing phylogenetic relationships of individuals, and current barcoding initiatives (Chase et al. 2005) are demonstrating that standardized markers can expose hitherto undetected taxa (Janzen et al. 2009), and can give up to 75% resolution among plant species (CBOL Plant Working Group 2009). This level may not yet be sufficient for taxonomic needs (Meier et al. 2006), although next generation sequencing technologies should offer a large number of variable markers which will help improve resolution, even in recently evolved radiations. The current global barcoding initiatives (such as TreeBoL, the tree barcode-of-life), should lead to the formation of data sharing networks in Southeast Asia, and could be useful vehicles for promoting new approaches to inventory and taxonomy in the region, both among scientists and the general public. The technological needs for DNA barcoding are relatively low and easily within the reach of 10–100s of labs in every country of Southeast Asia; for

²⁸ <http://www.cate-project.org/>.

²⁹ <http://www.editwebrevisions.info/>.

³⁰ <http://scratchpads.eu/>.

³¹ e.g., <http://www.nhm.ac.uk/solanaceaesource/>, <http://silurus.acnatsci.org/>.

example, the Gunung Palung flora project referred to above is proceeding with DNA barcode work at the molecular lab in Herbarium Bogoriense.

Most taxonomists are probably in agreement, however, that morphology and molecules must both be used in species delimitation (Mayo et al. 2008). What may be required to speed up the process of species delimitation and description is the acceptance of a more flexible model of taxonomy itself. We envision a system where automated DNA and morphological analyses cluster new specimens and backlogged indets into *species-like* groups. Where morphological and DNA data exist, type specimens should be included in analyses as anchors and links to existing literature (acknowledging that molecular analysis of some type specimens may not be permitted). These groups (sets of specimens) would be checked by trained parataxonomists and professionals for obvious clustering errors (e.g., excessive lumping), who also, time permitting, would map as many of the groups as possible to binomial species using keys and museum collections. Automated phylogeny generation would indicate which sets of specimens were most closely related (cf. PhyLoTA, Sanderson et al. 2008), and parataxonomists would then seek diagnostic characters useful for field identification, thus creating efficient, guided descriptions rather than exhaustive ones. Unnamed groups would be given useful, unique identifiers, which could be used to refer to a particular set of specimens at a particular time. The whole taxonomy would then given a date (or version number) and released, along with the keys, images and DNA. The process would be repeated as new material became available. The key element of such a semi-automated taxonomy generator is that ever-changing, inter-specimen distances are the primary data in allocating objects to temporary classes, rather than pre-specified fixed classes, in which the objects are placed. This alternative taxonomy process would proceed in parallel to the necessary work of assigning Latin binomials, which could be performed on a linked unitary taxonomy platform. We would thus be able to circumscribe the full diversity of collections far sooner, and would generate data that would be fundamentally useful in the painstaking process of hand-crafted alpha-taxonomy.

A sense of promise

Despite the grim outlook for many of the species in Southeast Asia, and the current insufficient level of collecting in the region, we feel there are realistic solutions for increasing species discovery and description activity. We see the signs of a convergence between the availability of enthusiastic students, their increasing online connectivity, and the development of tools to help share knowledge efficiently. The best news is that these tools are free, and the main barrier need not be a financial one. Instead the main limitations are lack of coordination and awareness, and perhaps partially of an openness to change. We also believe that one of the keys to effective conservation is bringing the reality of biodiversity home to the general population of an area. As the Indonesian saying goes: *Tidak kenal maka tak sayang*, “You can’t love what you don’t know.”

Acknowledgements Webb would like to thank M. Sanderson, T. Pennington, M. Donoghue, S. Renner, C. Cannon, J. Macklin, L. Lohman and C. Dick for stimulating discussions about these issues, the Government of Indonesia, (MenRisTek FRP, LIPI-PPB, PHKA, and Balai TN Gunung Palung) for assistance and permissions in Indonesia, the Arnold Arboretum and US National Science Foundation for funding, H. Yanto and E. Setiawan for field assistance, and G. Dean and R. Asmarayani for lab assistance. We are grateful to the Leiden herbarium for access to collections data. Finally, we thank the editors of this special issue for the invitation to participate and two anonymous reviewers for their comments.

References

- Adamowicz SJ, Purvis A (2005) How many branchiopod crustacean species are there? Quantifying the components of underestimation. *Glob Ecol Biogeogr* 14(5):455–468
- Appanah S (1993) Mass flowering of dipterocarp forests in the aseasonal tropics. *J Biosci* 18(4):457–474
- Baraloto C, Ferreira E, Rockwell C, Walthier F (2007) Limitations and applications of parataxonomy for community forest management in southwestern Amazonia. *Ethnobot Res Appl* 5:77–84
- Baskauf SJ, Kirchoff BK (2008) Digital plant images as specimens: toward standards for photographing living plants. *Vulpia* 7:16–30
- Basset Y, Novotny V, Miller SE, Pyle R (2000) Quantifying biodiversity: experience with parataxonomists and digital photography in Papua New Guinea and Guyana. *BioScience* 50:899–908
- Berners-Lee T, Fischetti M (1999) *Weaving the Web*. Harper, San Francisco
- Bickford D, Howard SD, Ng DJJ, Sheridan JA (2010) Impacts of climate change on the amphibians and reptiles of Southeast Asia. *Biodivers Conserv*. doi:10.1007/s10531-010-9782-4
- Blume CL (1827) *Bijdragen tot de flora van Nederlandsch Indie*. Ter Lands Drukkerij, Batavia, Indonesia
- Braschler B (2009) Successfully implementing a citizen-scientist approach to insect monitoring in a resource-poor country. *BioScience* 59(2):103–104
- Bryant D, Nielsen D, Tangle L (1997) *The last frontier forests: ecosystems & economies on the edge*. Technical report, World Resources Institute, Forest Frontiers Initiative, Washington, DC
- Caruana R, Elhawary M, Munson A, Riedewald M, Sorokina D, Fink D, Hochachka W, Kelling S (2006) Mining citizen science data to predict prevalence of wild bird species. In: *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp 909–915
- CBOL Plant Working Group (2009) A DNA barcode for land plants. *Proc Natl Acad Sci* 106(31):12794–12797
- Chao A (1984) Nonparametric estimation of the number of classes in a population. *Scand J Stat* 11: 265–270
- Chao A (2006) Species estimation and applications. In: Kotz S, Read CB, Balakrishnan N, Vidakovic B (eds) *Encyclopedia of statistical sciences*. Wiley, New York, pp 7907–7916
- Chase MW, Salamin N, Wilkinson M, Dunwell JM, Kesanakurthi RP, Haidar N, Savolainen V (2005) Land plants and DNA barcodes: short-term and long-term goals. *Phil Trans R Soc B* 360(1462):1889–1895
- Clark B, Godfray H, Kitching I, Mayo S, Scoble M (2009) Taxonomy as an eScience. *Phil Trans A* 367(1890):953–966
- Cohn J (2008) Citizen science: can volunteers do real research? *BioScience* 58(3):192–197
- Collar N (1998) Extinction by assumption; or, the Romeo Error on Cebu. *Oryx* 32(4):239–244
- Couvet D, Jiguet F, Julliard R, Levrel H, Teyssedre A (2008) Enhancing citizen contributions to biodiversity science and public policy. *Interdiscip Sci Rev* 33(1):95–103
- de Carvalho M, Bockmann F, Amorim D, de Vivo M, de Toledo-Piza M, Menezes N, de Figueiredo J, Castro R, Gill A, McEachran J et al (2005) Revisiting the taxonomic impediment. *Science* 307(5708):353–353
- de Wilde WJJ, Duyfjes BEE (2006) Redefinition of *Zehneria* and four new related genera (Cucurbitaceae), with an enumeration of the Australasian and Pacific species. *Blumea* 51:1–88
- Dolphin K, Quicke DLJ (2001) Estimating the global species richness of an incompletely described taxon: an example using parasitoid wasps (Hymenoptera: Braconidae). *Biol J Linn Soc* 73(3):279–286
- Erwin TL (1982) Tropical forests: their richness in Coleoptera and other species. *Coleopt Bull* 36:74–75
- Frodin DG (1984) *Guide to the Standard Floras of the World*. Cambridge University Press, Cambridge, UK
- Gardner RO (2006) Piper (Piperaceae) in the Philippine islands: the climbing species. *Blumea* 51:569–586
- Giam X, Ng TH, Yap VB, Tan HTW (2010) The extent of undiscovered species in Southeast Asia. *Biodivers Conserv*. doi:10.1007/s10531-010-9792-2
- Godfray HJ (2002) Challenges for taxonomy. *Nature* 417:17–19
- Graham J, Simpson A, Crall A, Jarnevich C, Newman G, Stohlgren T (2008) Vision of a cyberinfrastructure for nonnative, invasive species management. *BioScience* 58(3):263–268
- Greenwood JJD (2007) Citizens, science and bird conservation. *J Ornithol* 148:S77–S124
- Guralnick R, Hill A (2009) Biodiversity informatics: automated approaches for documenting global biodiversity patterns and processes. *Bioinformatics* 25(4):421–428
- Hawksworth DL, Rossman AY (1997) Where are all the undescribed fungi? *Phytopathology* 87:888–891
- Hibbett DS, Nilsson RH, Snyder M, Fonseca M, Costanzo J, Shonfeld M (2005) Automated phylogenetic taxonomy: an example in the Homobasidiomycetes (mushroom-forming fungi). *Syst Biol* 54:660–668
- Hyde K (2003) Mycology in the future in the Asia-Pacific region. *Fungal Divers* 13:59–68
- Janzen DH (1986) Blurry catastrophes. *Oikos* 47:1–2
- Janzen DH, Hallwachs W, Blandin P, Burns JM, Cadiou J, Chacon I, Dapkey T, Deans AR, Epstein ME, Espinoza B, Franclemont JG, Haber WA, Hajibabaei M, Hall JPW, Hebert PDN, Gauld ID,

- Harvey DJ, Hausmann A, Kitching IJ, Lafontaine D, Landry J, Lemaire C, Miller JY, Miller JS, Miller L, Miller SE, Montero J, Munroe E, Green SR, Ratnasingham S, Rawlins JE, Robbins RK, Rodriguez JJ, Rougerie R, Sharkey MJ, Smith MA, Solis MA, Sullivan JB, Thiaucourt P, Wahl DB, Weller SJ, Whitfield JB, Willmott KR, Wood DM, Woodley NE, Wilson JJ (2009) Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. *Mol Ecol Resour* 9(s1):1–26
- Johnson NF (2007) Biodiversity informatics. *Annu Rev Entomol* 52:421–438
- Kelling S, Hochachka W, Fink D, Riedewald M, Caruana R, Ballard G, Hooker G (2009) Data-intensive science: a new paradigm for biodiversity studies. *BioScience* 59(7):613–620
- Kress W (2004) Paper floras: how long will they last? A review of Flowering Plants of the Neotropics. *Am J Bot* 91(12):2124–2127
- Lawrence A (2009) The first cuckoo in winter: phenology, recording, credibility and meaning in Britain. *Glob Environ Change* 19(2):173–179. Traditional Peoples and Climate Change
- Lee M (2000) A worrying systematic decline. *Trends Ecol Evol* 15(8):346–348
- Mace G (2004) The role of taxonomy in species conservation. *Philos Trans* 359(1444):711–719
- Mace G, Masundire H, Baillie J et al (2005) Biodiversity. In: Hassan RM, Scholes R, Ash N (eds) *Ecosystems and human well-Being: current state and trends. Findings of the Condition and Trends Working Group. Millennium Ecosystem Assessment, Chap. 4.* Island Press, Washington, DC
- May RM (1988) How many species are there on earth? *Science* 247:1441–1449
- May R, Lawton J, Stork N (1995) Assessing extinction rates. In: Lawton JH, May RM, Maurer BA (eds) *Extinction rates.* Oxford University Press, Oxford, pp 1–24
- Mayo SJ, Allkin R, Baker W, Blagoderov V, Brake I, Clark B, Govaerts R, Godfray C, Haigh A, Hand R, Harman K, Jackson M, Kilian N, Kirkup DW, Kitching I, Knapp S, Lewis GP, Malcolm P, von Raab-Straube E, Roberts DM, Scoble M, Simpson DA, Smith C, Smith V, Villalba S, Walley L, Wilkin P (2008) Alpha e-taxonomy: responses from the systematics community to the biodiversity crisis. *Kew Bull* 63:1–16
- Meier R, Dikow T (2004) Significance of specimen databases from taxonomic revisions for estimating and mapping the global species diversity of invertebrates and repatriating reliable specimen data. *Conserv Biol* 18:478–488
- Meier R, Shiyang K, Vaidya G, Ng PKL (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Syst Biol* 55:715–728
- Miquel FAW (1859) *Flora van Nederlandsch Indie.* C. G. van der Post, Amsterdam, The Netherlands
- Nanda M (1997) Against social destruction of science: cautionary tales from the Third World. *Monthly Review*
- Novacek M (2008) Engaging the public in biodiversity issues. *Proc Natl Acad Sci* 105(Suppl 1):11571–11578
- Novotny V, Basset Y, Miller SE, Weiblen GD, Bremer B, Cizek L, Drozd P (2002) Low host specificity of herbivorous insects in a tropical forest. *Nature* 416(6883): 841–844
- Oliver I, Beattie AJ (1993) A possible method for the rapid assessment of biodiversity. *Conserv Biol* 7:562–568
- Page RDM (2006) Taxonomic names, metadata, and the semantic web. *Biodivers Inform* 3:1–15
- Peh KSH (2010) Invasive species in Southeast Asia: the knowledge so far. *Biodivers Conserv* (this volume). doi:10.1007/s10531-009-9755-7
- Pfeiffer J, Uril Y (2003) The role of indigenous parataxonomists in botanical inventory: from Herbarium Amboinense to Herbarium Floresense. *Telopea* 10:61–72
- Raczkowski J, Wenzel J (2007) Biodiversity studies and their foundation in taxonomic scholarship. *Bio-Science* 57(11):974–979
- Raes N, Steege H (2007) A null-model for significance testing of presence-only species distribution models. *Ecography* 30(5):727–736
- Roos M (2003) Flora Malesiana 1991–2001: what has been achieved: revitalisation, momentum? What next? *Telopea* 10:1–10
- Sanderson M, Boss D, Chen D, Cranston K, Wehe A (2008) The PhyLoTA browser: processing GenBank for molecular phylogenetics research. *Syst Biol* 57(3):335–346
- Scoble MJ (2004) Unitary or unified taxonomy? *Phil Trans R Soc B* 359(1444):699–710
- Scotland R, Wortley A (2003) How many species of seed plants are there? *Taxon* 52:101–104
- Soberón J, Peterson T (2004) Biodiversity informatics: managing and applying primary biodiversity data. *Phil Trans R Soc Lond Ser B* 359:689–698
- Sodhi NS, Brooks TM, Koh LP, Acciaioli G, Erb M, Tan AK-J, Curran LM, Brosius P, Lee TM, Patlis JM, Gumal M, Lee RJ (2006) Biodiversity and human livelihood crises in the Malay archipelago. *Conserv Biol* 20:1811–1813

- Sodhi NS, Lee TM, Koh LP, Brook BW (2009) A meta-analysis of the impact of anthropogenic forest disturbance on Southeast Asia's biotas. *Biotropica* 41:103–109
- Sohmer SH (1996) The urgency of documenting plant diversity: the Flora of the Philippines Project — a last chance to study the plant diversity of the Philippines. In: Stuessy TF, Sohmer SH (eds) *Sampling the green world*. Columbia University Press, New York, pp. 28–40
- Stork N (1997) Measuring global biodiversity and its decline. In: Reaka-Kudla M, Wilson D, Wilson E (eds) *Biodiversity II: Understanding and Protecting our Biological Resources*. National Academy Press, Washington, DC, pp 41–68
- Tongsima W, Tongsima S, Palittapongarnpim P (2008), Outlook on Thailand's genomics and computational biology research and development. *PLoS Comput Biol* 4(7):e1000115
- Tummarello G, Morbidoni C (2008) The DBin platform: A complete environment for Semantic Web Communities. *J Web Semant* 6(4):257–265. *Semantic Web Challenge 2006/2007*
- Udell J (2008) A conversation with Janis Dickinson about citizen science. Internet Blog, URL: <http://blog.jonudell.net/2008/05/05/a-conversation-with-janis-dickinson-about-citizen-science/>
- van House NA (2002) Digital libraries and practices of trust: networked biodiversity information. *Soc Epistemol* 16:99–114
- van Steenis-Kruseman MJ (1950) *Cyclopaedia of collectors*. *Flora Males Ser* 1(1):1–639
- Wee YC, Subaraj R (2009) Citizen science and the gathering of ornithological data in Singapore. *Nat Singap* 2:27–30
- Weeks P, Gaston K (1997) Image analysis, neural networks, and the taxonomic impediment to biodiversity studies. *Biodivers Conserv* 6(2):263–274
- Wheeler QD (2004) Taxonomic triage and the poverty of phylogeny. *Phil Trans R Soc Lond Ser B* 359(1444):571–583
- Wilcove DS, Koh LP (2010) Addressing the threats to biodiversity from oil palm agriculture. *Biodivers Conserv*. doi:10.1007/s10531-009-9760-x
- Wortley AH, Scotland RW (2004), Synonymy, sampling and seed plant numbers. *Taxon* 53: 478–480
- Zeti AMH, Shamsir MS, Tajul-Arifin K, Merican AF, Mohamed R, Nathan S, Mahadi NM, Napis S, Tan TW (2009) Bioinformatics in Malaysia: hope, initiative, effort, reality, and challenges. *PLoS Comput Biol* 5(8):e1000457
- Zollinger H (1857) Over het begrip en den omvang eener Flora Malesiana. *Natuurkundig Tijdschrift voor Nederlandsch-Indie* 13:293–322