

INDOBIOSYS – DNA BARCODING AS A TOOL FOR THE RAPID ASSESSMENT OF HYPERDIVERSE INSECT TAXA IN INDONESIA: A STATUS REPORT

Bruno Cancian de Araujo*¹, Stefan Schmidt¹, Thomas von Rintelen², Hari Sutrisno³, Kristina von Rintelen², Rosichon Ubaidillah³, Christoph Häuser², Djunijanti Peggie³, Raden Pramesa Narakusumo³ and Michael Balke¹

¹ Staatlichen Naturwissenschaftlichen Sammlungen Bayerns, Zoologischer Staatssammlung München, Muenchhausenstr. 21, Munich, Germany

² Museum für Naturkunde Leibniz-Institut für Evolutions- und Biodiversitätsforschung, Invalidenstraße 4, Berlin, Germany

³ Zoology Division (Museum Zoologicum Bogoriense), Research Center for Biology, Indonesian Institute of Sciences, Jl. Raya Jakarta-Bogor Km 46, Cibinong, Bogor 16911, Indonesia

*Corresponding author: chalcididae@gmail.com

Received: 18 October 2017; Accepted: 4 December 2017

ABSTRACT

A status report with preliminary results for the IndoBioSys project is presented and the impact of the project results for our knowledge of the Indonesian fauna is discussed. Using the REST API available on the Barcode of Life Data System we recover 21,153 public records (3,390 BINs) from Indonesia and compare against the 21,813 records (3,580 BINs) generated by the IndoBioSys project. From all IndoBioSys BINs, 3,366 (94%) are new to Indonesia. IndoBioSys is responsible for a BIN increase of 36.5% in Lepidoptera, 62.6% in Trichoptera, 986% in Coleoptera, and 1,086% in Hymenoptera. After two years of the IndoBioSys project, the Museum Zoologicum Bogoriense became the depository institution of 51.9% of Lepidoptera records, 95.8% of Coleoptera records, 97.6% of Hymenoptera records and 59.4% of Trichoptera records for Indonesia available on Barcode of Life Data System (BOLD). Now, with 55% of all Indonesian records available on BOLD, it is the most important depository for records of Indonesian genetic biodiversity, housing more than 23,000 new voucher specimens in their collections. Before IndoBioSys, the Museum Zoologicum Bogoriense was responsible for only 9% of all records available in the Barcode of Life Data System for Indonesia, showing the importance of those pipelines in empowering the local institutions in becoming the reference depository of the local fauna.

Key words: biodiversity inventory, CO1, DNA barcode, high-throughput pipeline, metabarcoding

INTRODUCTION

Much has been written on the importance of conserving our earth's biodiversity and on the primacy of understanding that diversity by expanding taxonomic knowledge (Miller *et al.* 2016). However, the majority of species remains undescribed, and this is especially true for arthropods (Hamilton *et al.* 2010) which play a major role in ecosystem functioning and services since they comprise the overwhelming majority of species in terrestrial habitats (Kremen *et al.* 1993).

To obtain objective estimates of species diversity in a given area, it is now possible to use DNA sequences to characterise certain target taxa at the species level (Riedel *et al.* 2013), at the community level (Hendrich *et al.* 2010) or in bulk e.g., from Malaise trap samples (Yu *et al.* 2012, see also <http://biodiversitygenomics.net/projects/gmp/>).

Such high-throughput biodiversity assessment pipelines aim to document the diversity of animals, fungi, or plants in a specific area to estimate species richness and characterise its biodiversity rapidly. This information can provide the basis for taxonomic and ecological research

as well as suggesting priority areas for conservation. These pipelines are often associated with tools that allow specimens to be identified to the species or a higher taxonomic level through comparison with a database of DNA barcodes from expertly identified specimens that are deposited in a repository in case further taxonomic scrutiny is needed (Miller *et al.* 2016). Nowadays, those pipelines are often associated with barcoding tools, in order to have long-term fast specimen assessment and standard comparison against worldwide biota (see Miller *et al.* 2016). The emergence of DNA barcoding as a method to discover and differentiate species objectively and rapidly (Hebert *et al.* 2003) and its global adoption (Hebert *et al.* 2009, Miller *et al.* 2016) has provided a new tool for assessing arthropod diversity for diverse applications including the management of natural resources. Since the circumscription of species is complex and sometimes controversial (Wheeler & Meier 2000), the Barcode of Life Data System (BOLD) has established a tool to delineate species proxies (Ratnasingham & Hebert 2013) termed the Barcode Index Number (BIN) System. A well established specific algorithm cluster sequences to produce operational taxonomic units that closely correspond to species. BINs are unique in that clusters are indexed in a regimented way, so genetically identical taxa will be addressed with the same identifier (Ratnasingham & Hebert 2013).

DNA barcoding is useful for identifying specimens collected in large scale biodiversity surveys and can augment morphological taxonomy to determine species identity (Geiger *et al.* 2016). Globally, many large scale barcoding projects (Canadian Barcode of Life Network, German Barcode of Life) have produced big amount of data to populate BOLD's global database of DNA barcodes (Hendrich *et al.* 2014, Morinière *et al.* 2014, Hendrich *et al.* 2015, Schmidt *et al.* 2015, Hawlitschek *et al.* 2016, Hebert *et al.* 2016).

Based on that premise, the Indonesian Biodiversity Discovery and Information System (IndoBioSys) was established as a partnership between German and Indonesian government agencies with several goals under the umbrella of biodiversity and health, highlighting the goal of building a comprehensive barcode library for key areas in Indonesia increasing the representativeness of the Indonesian fauna in the collections of the Museum Zoologicum Bogoriense (MZB), Research Center for Biology, Indonesian Institute of Sciences (LIPI).

Even in well studied areas, DNA barcoding has uncovered many cases of overlooked, misinterpreted, cryptic, or even new species (Dincă *et al.* 2011, Hendrich *et al.* 2014, Morinière *et al.* 2014, Schmidt *et al.* 2015). The Swedish Malaise trap program has so far increased the country's insect fauna by more than 1,900 species, including several hundred species new to science (Ronquist 2010, Karlsson 2017). Identifying the estimated 40 million specimens generated by this program using morphological characters would take many biologists several decades (Geiger *et al.* 2016). For obtaining an overview of the diversity in a short time frame, especially in the tropics, modern high-throughput approaches are required.

This study aims to compare the DNA barcodes generated through IndoBioSys with existing public data available on BOLD from species collected in Indonesia. It represents the first overview in the development of a reliable DNA barcode library for Indonesia and discusses the importance of rapid biodiversity assessment pipelines to quantify the species richness of neglected, hyperdiverse taxa. This will allow Indonesia and other countries with biodiversity hotspots to better understand their biodiversity. Here, we summarise progress in the first two years of the IndoBioSys project, showing rapid growth of knowledge about the diversity of the Indonesian insect fauna. We also present and discuss the representativeness of Indonesian institutions as depositories of the local fauna before and after the IndoBioSys project.

MATERIALS AND METHODS

Specimen collecting and processing

The specimen collections for IndoBioSys project included in this analysis were performed between September of 2015 and May of 2017 at Halimun-Salak National Park area, West Java province. The specimens found as Indonesian public records on BOLD were collected between 1905 and 2016. For more details about field and lab protocol, please see Schmidt *et al.* 2015, 2017.

Data acquisition

All public records from Indonesia present in BOLD were obtained through the REST API available on the BOLD platform on 29/05/2017. We applied the “Full Data Retrieval” parameters `geo=Indonesia` and `marker=COI-5P` in order to gather all public records from Indonesia with the standard DNA barcoding marker (COI-5P). IndoBioSys data were downloaded directly from the BOLD workbench using tools available in BOLD.

Data processing

The two files that were downloaded contained information on each record including the BIN, the depository, and the GPS coordinates. The GPS data were used to generate an occurrence map of public and IndoBioSys data (Fig. 1) using Quantum GIS v. 2.8. For aesthetic reasons, this map was redrawn using GIMP 2 preserving the original information. The number of records belonging to each depository institution was evaluated and the data was separated in three categories: MZB (Museum Zoologicum Bogoriense), other institutions, and obtained from GenBank (without depository information). MZB was the only Indonesian institution represented. Two layers were generated, one with the public records only and another one that added the IndoBioSys records.

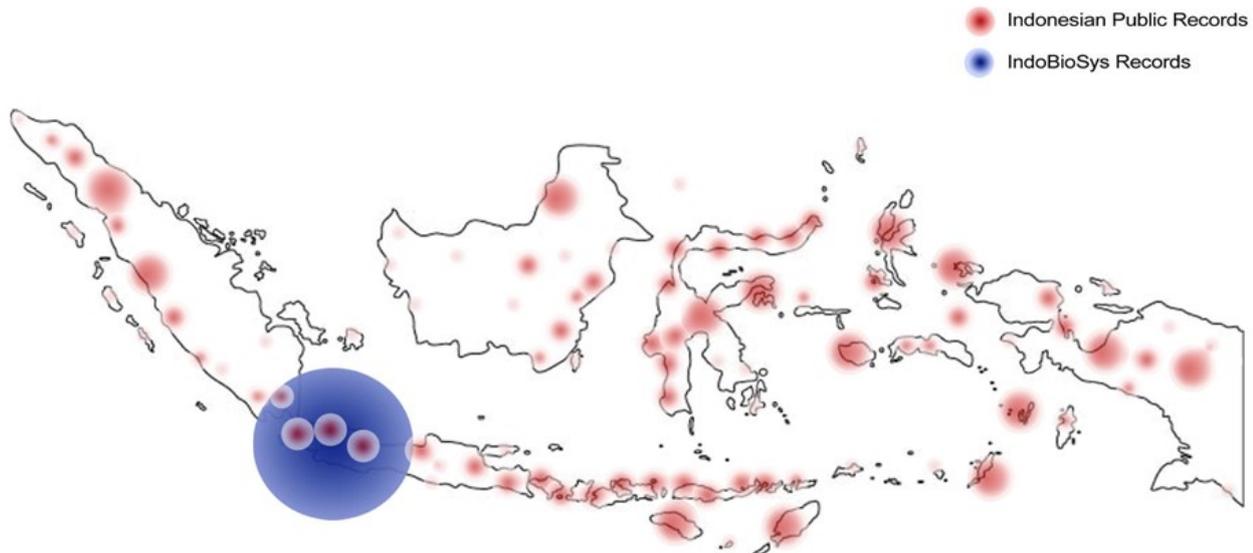


Figure 1. The center of each circle indicates the geospatial location of DNA barcode records, and their diameter indicates the relative abundance of Indonesian records available on BOLD (red) or IndoBioSys (blue).

For comparison between public and IndoBioSys data, duplicate BINs were removed from each of the two sources, public and IndoBioSys, were removed using Microsoft Excel, in order to have only one representative of each BIN. The direct comparison of the number of BINs and shared BINs between these two sources was made using Microsoft Excel. The number of BINs shared by the two sources was evaluated and after that, the shared BINs were subtracted from IndoBioSys list in order to ascertain the contribution of the project for Indonesian records. For practical reasons, only BINs from four orders of insects were compared (Coleoptera, Lepidoptera, Hymenoptera and Trichoptera).

RESULTS

We found 21,153 public records from Indonesia in BOLD, corresponding to 5,676 BINs. From those BINs, 3,390 are from one of the four insect orders selected for the comparison, i.e. 103 BINs of Coleoptera, 3,093 of Lepidoptera, 103 of Hymenoptera and 91 of Trichoptera. At the time of the search the IndoBioSys project had 21,813 records and 3,580 BINs, representing 3,322 BINs of the selected orders, being 1,016 BINs of Coleoptera, 1,130 of Lepidoptera, 1,119 of Hymenoptera and 57 of Trichoptera. The distribution of all 42,336 records over Indonesia is shown in Fig 1.

For the public records, the MZB held vouchers of 2,018 specimens with COI barcodes available through BOLD while all other institutions held 15,387 vouchers, and 3,748 vouchers recorded from GenBank (NCBI), without specific depository information (Fig. 2). The MZB is the depository for all 21,813 voucher specimens in the IndoBioSys project. In terms of local collection

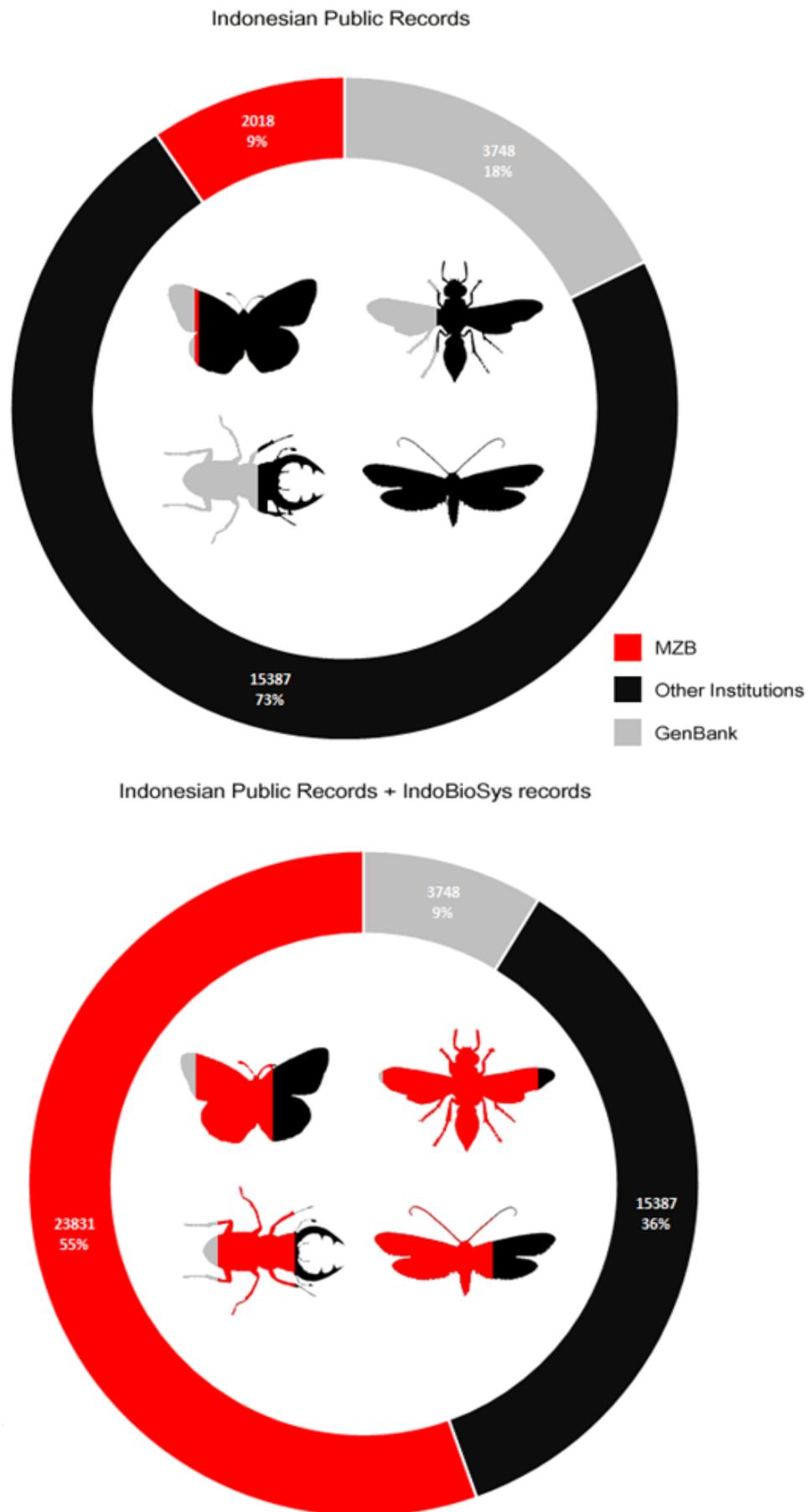


Figure 2. Relevance of the MZB as depository for Indonesian barcoding vouchers before and after the IndoBioSys project. The circular graph corresponds to the total number of records for all groups of organisms in BOLD. The graphs inside illustrate the representativeness for the four groups highlighted in the present study.

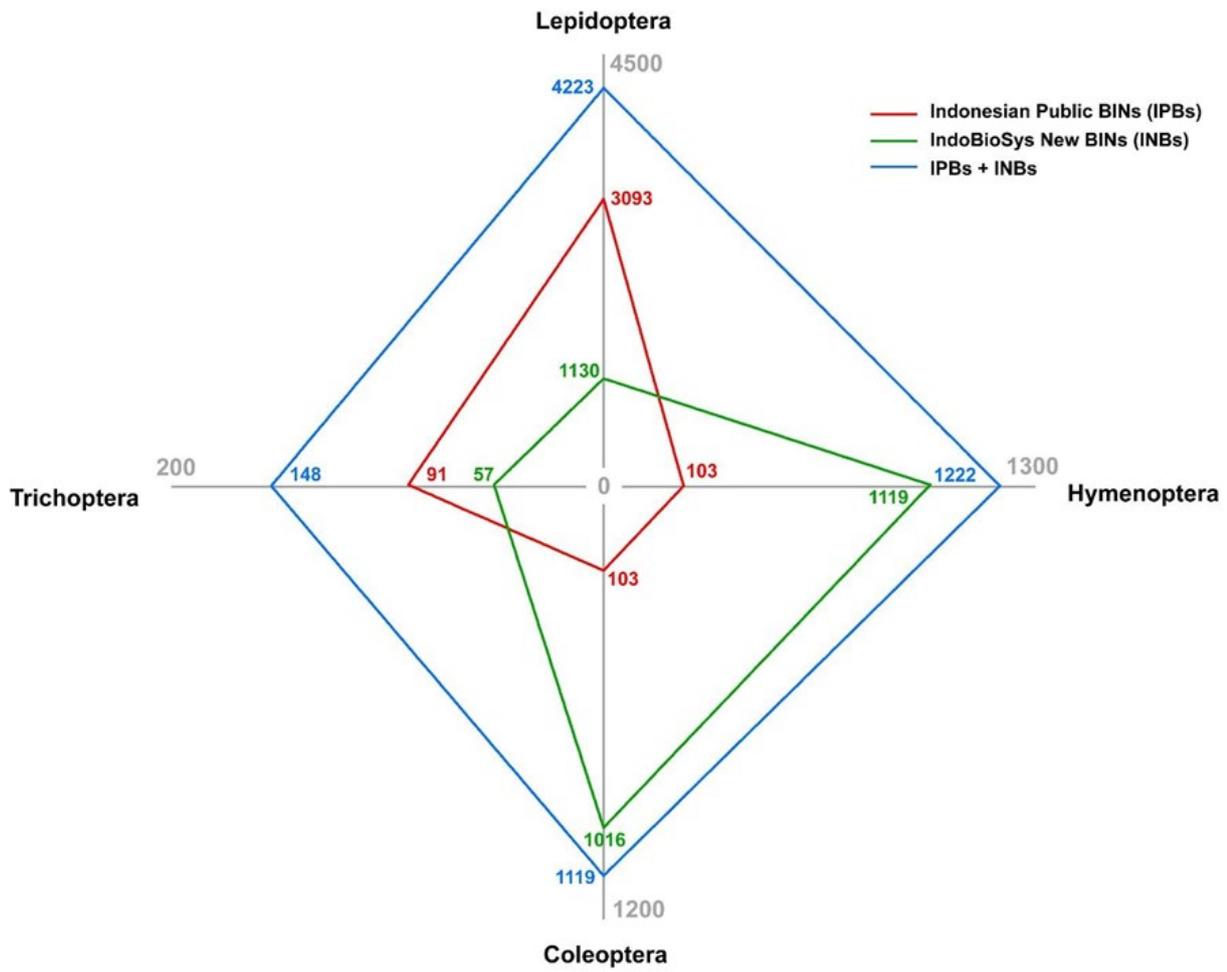


Figure 3. Indonesian BINs available for four insect orders before and after the IndoBioSys project.

representativeness, MZB was previously held only 9% of all public barcode records from Indonesia, which increased to 55% after the addition of the IndoBioSys records. In the four target taxa of this study, MZB previously had only 11 records for Lepidoptera and none for Coleoptera, Hymenoptera and Trichoptera. After adding the IndoBioSys records, the MZB now holds 3,661 Lepidoptera records, 9,644 Coleoptera records, 7,447 Hymenoptera records and 400 Trichoptera records.

The two sources that were evaluated shared only 6% of BINs, meaning that 94% of IndoBioSys BINs are new barcodes for Indonesia. Considering the four insect groups selected for comparison, the IndoBioSys project is responsible for BIN increases of 36.5% in Lepidoptera, 62.6% in Trichoptera, 986% in Coleoptera, and 1,086% in Hymenoptera (Fig. 3). All records and associated sequences referred in this paper will be available through the Barcode of Life Database (www.boldsystems.org) and the project website (www.indobiosys.org).

DISCUSSION

Progress toward barcode inventories in megadiverse tropical areas is highlighted by the 94% of new BINs for Indonesia that were generated by the IndoBioSys project. More than 95% of the IndoBioSys records were from only one area (Mount Halimun-Salak National Park in West Java) meaning that one project based in a single area was able to drastically increase knowledge about the Indonesian fauna in less than two years. It shows how important and urgent high-performance biodiversity assessment pipelines are in order to uncover and quantify biodiversity of neglected hyperdiverse taxa in the tropics, especially in areas that are threatened. Halimun-Salak National Park has been threatened by rapid forest degradation. The forest canopy density decreased by a total of 7,587.18 ha (6.69% of the total area) from 2003 to 2011 (Carolyn *et al.* 2013). Therefore, IndoBioSys plays an important role in estimating the true species richness within the Halimun-Salak National Park despite ongoing forest degradation.

Before the IndoBioSys project started, the megadiverse orders Coleoptera and Hymenoptera were represented in BOLD by only 103 records each from Indonesia, a miniscule number considering the tremendous diversity of those groups in the tropics. Even for well-studied groups as Lepidoptera, a substantial increase in the number of BINs was achieved.

As intended by the Convention on Biological Diversity, in particular the Nagoya Protocol, those pipelines promote non-commercial research that brings benefit to the local institutional capacity as cooperation and contribution in research and training as well as through scientific information relevant to biological inventories and taxonomic studies. Furthermore, these pipelines are also important for empowering the local institutions to become the reference repository of the local fauna. Before IndoBioSys started, MZB was responsible for only 9% of all records available

in BOLD. Considering the four taxa highlighted, MZB had only 0.32% of the total records of Lepidoptera and no records of Coleoptera, Hymenoptera and Trichoptera. After two years of the IndoBioSys project, the MZB is the depository of 51.9% of Lepidoptera records, 95.8% of Coleoptera records, 97.6% of Hymenoptera records and 59.4% of Trichoptera records for Indonesia. Now, with 55% of all Indonesian barcode records available on BOLD, MZB has become one of the most important depositories for Indonesian DNA barcoding vouchers, housing more than 23,000 new records in their insect collections. Those are high-quality records with complete collection data, DNA extracts, CO1 sequences, good quality images, and with the voucher specimens quickly dried, mounted, labeled and stored in the Indonesian national archive of the archipelago's animal diversity. All available MZB data collections will be incorporated into the Indonesia Biodiversity facility (InaBIF) as an accessible biodiversity database.

ACKNOWLEDGMENTS

The authors thank Dr. Rachel Vicente (UFES, Brazil) for helping with Quantum GIS. Further thanks go to Tri Siswo Rahardjo for permission to access Halimun-Salak National Park; to Dr. Witjaksono for the permission of the work in Research Center for Biology, to Erniwati (MZB researcher), Dr. Andreas Florean (University Würzburg); Giyanto, Sarino, Darmawan (MZB Entomology technicians), for technical help in the fieldwork and for laboratory work. We thank the Ministry of Research and Higher Education of the Republic of Indonesia for providing a foreign research permit (number 2B/TKPIPA/E5/Dit.KI/II/2016). The IndoBioSys project is funded by the Bundesministerium für Bildung und Forschung within the bilateral "Biodiversity and Health" funding programme (Project numbers: 16GW0111K, 16GW0112) and funded by DIPA PUSLIT Biologi LIPI 2015-2016.

REFERENCES

- Carolyn, R.D., D.P.T. Baskoro & L.B. Prasetyo 2013. Analisis degradasi untuk penyusunan arahan strategi pengendaliannya di Taman Nasional Gunung Halimun-Salak Provinsi Jawa Barat. *Globe* **15**(1): 39–47. <<http://jurnal.big.go.id/index.php/GL/article/viewFile/70/67>>
- Dincă, V., V.A. Lukhtanov, G. Talavera & R. Vila 2011. Unexpected layers of cryptic diversity in wood white *Leptidea* butterflies. *Nature Communications* **2**: 324. <<http://www.nature.com/articles/ncomms1329>>
- Geiger, M. F., J.J. Astrin, T. Borsch, U. Burkhardt, P. Grobe, R. Hand, A. Hausmann, K. Hohberg, L. Krogmann, M. Lutz, C. Monje, B. Misof, J. Morinière, K. Müller, S. Pietsch, D. Quandt, B. Rulik, M. Scholler, W. Traunspurger, G. Haszprunar & W. Wägele 2016. How to tackle the molecular species inventory for an industrialized nation—lessons from the first phase of the German Barcode of Life initiative GBOL (2012–2015). *Genome* **59**(9): 661–670. <<http://www.nrcresearchpress.com/doi/abs/10.1139/gen-2015-0185#.WedUqluCybg>>

- Hamilton, A. J., Y. Basset, K.K. Benke, P.S. Grimbacher, S.E. Miller, V. Novotný, G.A. Samuelson, N.E. Stork, G.D. Weiblen & J.D.L. Yen 2010. Quantifying uncertainty in estimation of tropical arthropod species richness. *The American Naturalist* **176**(1): 90–95. < <http://www.journals.uchicago.edu/doi/10.1086/652998>>
- Hawlitsek, O., J. Morinière, G.U.C. Lehmann, A.W. Lehmann, M. Kropf, A. Dunz, F. Glaw, M. Detcharoen, S. Schmidt, A. Hausmann, N.U. Szucsich, S.A. Caetano-Wyler & G. Haszprunar 2016. DNA barcoding of crickets, katydids and grasshoppers (Orthoptera) from Central Europe with focus on Austria, Germany and Switzerland. *Molecular Ecology Resources*. DOI:10.1111/1755-0998.12638. < <http://onlinelibrary.wiley.com/doi/10.1111/1755-0998.12638/full>>
- Hebert, P. D. N., J.R. de Waard, J. –F. Landry 2009. DNA barcodes for 1/1000 of the animal kingdom. *Biology Letters* **6**(3): 359–362. < <http://rsbl.royalsocietypublishing.org/content/6/3/359>>
- Hebert, P. D. N., S. Ratnasingham & J.R. deWaard 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B: Biological Sciences* **270**: S96–S99. <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1698023/pdf/12952648.pdf>>
- Hebert, P.D.N., S. Ratnasingham, E.V. Zakharov, A.C. Telfer, V. Levesque-Beaudin, M.A. Milton, S. Pedersen, P. Jannetta & J.R. deWaard 2016. Counting animal species with DNA barcodes: Canadian insects. *Philosophical Transactions of the Royal Society B* **371**: 1–10, 20150333. <<http://rstb.royalsocietypublishing.org/content/371/1702/20150333>>
- Hendrich L, Pons J, Ribera I, Balke M. 2010. Mitochondrial Cox1 Sequence Data Reliably Uncover Patterns of Insect Diversity But Suffer from High Lineage–Idiosyncratic Error Rates. *PLoS ONE* **5**(12): e14448. < <https://doi.org/10.1371/journal.pone.0014448>>
- Hendrich, L., J. Morinière, G. Haszprunar, P.D.N. Hebert, A. Hausmann, F. Kohler & M. Balke 2015. A comprehensive DNA barcode database for Central European beetles with a focus on Germany: adding more than 3500 identified species in BOLD. *Molecular Ecology Resources* **15**(4): 795–818. < <http://onlinelibrary.wiley.com/doi/10.1111/1755-0998.12354/full>>
- Karlsson, D. 2017. The Swedish Malaise Trap Project 2.0. *Hamuli*. **8**(1): 6–7.
- Kremen, C., R. Colwell, T. Erwin, D. Murphy, R. Noss & M. Sanjayan 1993. Terrestrial Arthropod Assemblages: Their Use in Conservation Planning. *Conservation Biology* **7**(4): 796–808. < https://entomology.si.edu/StaffPages/Erwin/T's%20updated%20pub%20PDFs%2010Jan2014/111_1993_TerrestrialArthropodAssemblages_Conservation.pdf>
- Miller, S.E., A. Hausmann, W. Hallwachs & D.H. Janzen 2016. Advancing taxonomy and bioinventories with DNA barcodes. *Philosophical Transactions of the Royal Society B: Biological Sciences*. **371**(1702): 20150339. <<http://rstb.royalsocietypublishing.org/content/royptb/371/1702/20150339.full.pdf>>
- Morinière, J., L. Hendrich, A. Hausmann, P. Hebert, G. Haszprunar & A. Gruppe 2014. Barcoding Fauna Bavarica: 78% of the Neuropterida Fauna Barcoded!. *PLoS ONE* **9**(10): e109719. < <http://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0109719&type=printable>>
- Ratnasingham, S. & P.D.N. Hebert 2007. BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Molecular Ecology Notes* **7**: 355–364. < <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1890991/>>
- Ratnasingham, S. & P.D.N. Hebert 2013. A DNA-based registry for all animal species: the Barcode Index Number (BIN) system. *PLoS ONE* **8**(7): e66213. < <http://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0066213&type=printable>>
- Riedel, A., K. Sagata, Y.R. Suhardjono, R. Tänzler & M. Balke 2013. Integrative taxonomy on the fast track - towards more sustainability in biodiversity research. *Frontiers in Zoologie* **10**(1): 15. <<https://frontiersinzoology.biomedcentral.com/track/pdf/10.1186/1742-9994-10-15 site=frontiersinzoology.biomedcentral.com>>

- Ronquist, F. 2010. 250 Years of Swedish Taxonomy. In: Polaszek A. (ed.), *Systema Naturae 250 - The Linnaean Ark*. CRC Press. 239–248. < <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5136679/>>
- Schmidt, S., C. Schmid-Egger, J. Morinière, G. Haszprunar & P. Hebert 2015. DNA barcoding largely supports 250 years of classical taxonomy: identifications for Central European bees (Hymenoptera, Apoidea partim). *Molecular Ecology Resources* **15**(4): 985–1000. < <http://onlinelibrary.wiley.com/doi/10.1111/1755-0998.12363/full>>
- Schmidt, O., A. Hausmann, B. Cancian de Araujo, H. Sutrisno, D. Peggie & S. Schmidt 2017. A streamlined collecting and preparation protocol for DNA barcoding of Lepidoptera as part of large-scale rapid biodiversity assessment projects, exemplified by the Indonesian Biodiversity Discovery and Information System (IndoBioSys). *Biodiversity Data Journal* **5**: e20006.<<https://doi.org/10.3897/BDJ.5.e20006>>
- Wheeler, Q. & R. Meier 2000. *Species Concepts and Phylogenetic Theory: a Debate*. Columbia University Press.
- Yu, D.W., Y. Ji, B.C. Emerson, X. Wang, C. Ye, C. Yang & Z. Ding 2012. Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution* **3**: 613–623.