Plant taxonomy: cooperation, innovation, productivity

Peter C. van Welzen[1,2*], Eva Kristinawati Putri[1,2,3], and Roxali Bijmoer[2]

[1] Naturalis Biodiversity Center, P.O. Box 9517, 2300 RA Leiden, The Netherlands
[2] Institute of Biology Leiden, Leiden University, P.O. Box 9505, 2300 RA Leiden, The Netherlands
[3] Biology Study Program, Universitas Negeri Surabaya, Indonesia

Abstract. Floristic taxonomic research, looking for constant differences between species, is a very time-consuming and slow process. The flora of the Malay Archipelago is estimated to have 45,000 vascular plant species and the Flora Malesiana project that describes them has completed c. 1/3rd in 75 years. Presently, the flora is heavily threatened because of land use and climate change. We have to speed up the flora writing to understand what will be threatened and how to counter the threats. Innovation using Next Generation Sequencing of DNA to find the species in combination with Image Recognition might help to increase speed. Another aspect is cooperation, more international scientists should be involved, also students and citizens via citizen science projects. Citizens can help to gather information (observations and collecting). Once the flora is better known, monitoring of the flora by regular inventories (done by non-scientists in the Netherlands) helps to see which plants are threatened or become plagues and causes can then be investigated. Monitoring will help to change energy consumption (no more fossil energy), agriculture, industry, and living: improve our climate, and save the planet.

1 Introduction

It is now obvious that the enormous numbers and activities of humans threaten the existence of more and more plants and animals due to increased land use for housing, industry, and food and climate change. Erroneously enough, we often do not know yet what we are threatening. Especially, in tropical countries the knowledge of plants and animals is limited.

In this article, we will focus on the Flora Malesiana Project, initiated in the 1950s by van Steenis [1] to describe the flora of the Malay Archipelago, including Indonesia. The Malesian flora has an estimated number of 42,000 [2] to 45,000 [3] species of ferns, gymnosperms, and flowering plants (of which 70% only occurs in this area [4] and nowhere else). We will sometimes contrast Flora Malesiana with the flora of the Netherlands [5], with an estimated number of 1,500 species. The Flora Malesiana project

* Corresponding author: peter.vanwelzen@naturalis.nl

is presently coordinated by Daniel Thomas in Singapore. He also maintains the website (https://floramalesiana.org/new/) with much more information than described here.

Most systematic work on plants is done in herbariums, research institutes specialized in storing, maintaining, and using collections of dried plants. The herbarium technique, mounting dried plants on a thick paper sheet together with a label containing information (collector, place of collection, colours, etc.), was developed in c. 1550 in Italy. The first herbariums were herbals, books in which the plants were mounted for their beauty, not for scientific work (Figure 1).

The shift to loose sheets was brought about by Linnaeus, in the 1700s (Fig. 2). Presently, 3100 herbariums store about 390 million specimens of plants [6]. These herbariums are the major source of all new species that are yearly described.

Fig. 1. A. 'En Tibi' Herbarium (1558), a book herbal. **B.** page with on right the oldest dried tomato in the world.

What does it entail to write a Flora? A Flora is a critical overview of a certain area, usually a country, for which plant species are described, classified in a hierarchical system (genera, families, etc.) that models the evolution and is made recognizable via tables with questions (identification keys). Often notes about the distribution, ecology, uses, and local names are added. For the Flora Malesiana project, the whole flora is re-evaluated. Why? Systematic work is not only describing new species but also evaluating the work of others, especially if many different authors described species in the course of time. Quite often it appears that many names are in fact synonyms, either because the area was once occupied by Western forces, each independently describing the flora (e.g., Spain and the USA in the Philippines, the Netherlands in Indonesia, England in Malaysia, France in Indochina, Portugal in Timor Este, Germany in Papua New Guinea) or researchers only saw forms on one island and were not aware of the complete variation in a species (see how variable our own species is). While re-evaluating a systematist tries to find morphological differences between species, which are constant, and always present between all members of those species. Unfortunately, characteristics that work for one group, are often useless in other groups. This makes systematic work tedious and extremely slow.

Fig. 2. Herbarium specimens mounted on loose sheets. Labels contain hardly any information. A nice touch is the cut-out printed vase to cover the cut branches.

Fig. 3. Sample dried and treated with alcohol after collecting, brown and with degraded DNA. Labels with more information.

Fig. 4. Quickly dried collection, still nicely green and likely with good DNA. Label with good information about data (colors, place, habitat, etc.).

Since the start of Flora Malesiana, in more than 70 years ago, only about a third of the Flora is finished, and even in the Netherlands, an update of the Dutch Flora takes always much more time than anticipated. Unfortunately, the threats to the Flora develop much faster, which means that gaining knowledge of the Flora should also increase at a much faster pace. Another problem of all human-induced changes is also the introduction of many new species into areas, willingly or accidentally, of which some grow into plagues and are termed invasive species. Learning to recognize them is also important. Here we will show ideas how to speed up Flora's work, through innovation and by cooperation. We will also discuss on how the results of systematic work can be used in improving our planet.

2 Innovation

If only looking at morphology makes progress in Flora's description too slow, then which other possibilities do we have? We already used DNA, the molecular sequences, of plants and animals for decades. Still, plants often present problems. One of them is caused by the speed with which they are dried; if this preservation takes too long, then plant DNA is degraded and broken up into smaller parts. This is often the case with plants collected in the tropics. In the field, it is too humid to dry the plants properly and to keep them dry. Therefore, they were often placed between newspapers, piled in a plastic bag, soaked with alcohol and sealed. Proper drying can then wait for months. However, the amount of alcohol (expensive, heavy to carry) is generally much too low to quickly kill the plant cells. Usually, these plants are dark brown when dry (Fig. 3); while specimens that are quickly dried tend to remain green (Fig. 4). Besides collecting the specimen, modern collecting usually also puts a leaf fragment on silica gel in a small plastic bag, thus drying it quickly and making it suitable for DNA analysis. Another problem with plants is that their secondary metabolites sometimes interfere with DNA extraction and multiplication. Then the only answer is tweaking with methods in the lab to reduce the influence of these metabolites.

Herbaria possesses many plants that are poorly dried and lack silica samples. Their DNA is usually corrupted and the plants were either not used for molecular work, or if used, usually generated meager results. However, new molecular techniques, next generation sequencing, produced a method that might be very useful for all dried plants, Hyb-Seq (Hybrid Sequencing) [7]. This technique breaks the DNA strands up into small parts (which already happened due to the poor drying of many specimens). Many short DNA strands are targeted and then multiplied via PCR, after which they are sequenced; in the next step, a lot of computational work is needed to unite the short sequences into longer gene sequences. The main idea is that specimens that belong to the same species will group together once an evolutionary model (phylogenetic tree) has been made of the sequence data (Fig. 5). This will hopefully shorten the time needed to revise species. Presently, it takes time to learn the Hyb-Seq method. (Just a side remark, if an identification is missing then molecular sequences, especially regions like ITS, rbcL and trnL, are often used as so-called DNA-barcodes and compared with huge databases such as GenBank (https://www.ncbi.nlm.nih.gov/genbank/) and BOLD (Barcode of Life Data system; https://www.boldsystems.org/) to see what the identification might be. Many SE Asian species are missing in these databases, thus often this method only shows to which family and perhaps which genus the specimen belongs).

The next step is to place the remaining dried specimens on the piles of specimens created by the phylogenetic tree. That can be done by using morphological characters, but it is likely faster to use Image Recognition, an Artificial Intelligence (AI) technique. In the Netherlands, ObsIdentify (https://waarneming.nl/apps/obsidentify/) is an app (Fig. 6) with which you can make a photo of a plant or animal that the software can identify via Image Recognition (IR); it also gives a chance of how reliable the identification is. Photos are added to the learning photos, see below, after confirmation of the identification by an expert, also as GSMs have their GPS – Global Positioning Software – usually active, the locality of the plant or animal is automatically known.

Image Recognition software needs photos of the objects to train what it has to look for. For this, photos of the specimens used for the molecular analysis can be used. Many herbaria have all their specimens photographed (and generally online), just like Naturalis Biodiversity Center (Bioportal: https://bioportal.naturalis.nl/). The files with photos of the remaining herbarium specimens of a genus or an area can then be presented to the IR software, after which the plants that were recognized can be added to the piles created by the molecular research. These piles can then be used to find morphological differences, which is likely much faster than starting to do so from scratch as the plants are (hopefully) already organised in piles. The piles, if different enough, can be recognized as species.

https://doi.org/10.1051/e3sconf/202451303013

If the combination of next-generation sequencing and image recognition works then we should end up with well-circumscribed species in a faster way. The techniques are not new, the innovation part is to combine both techniques.

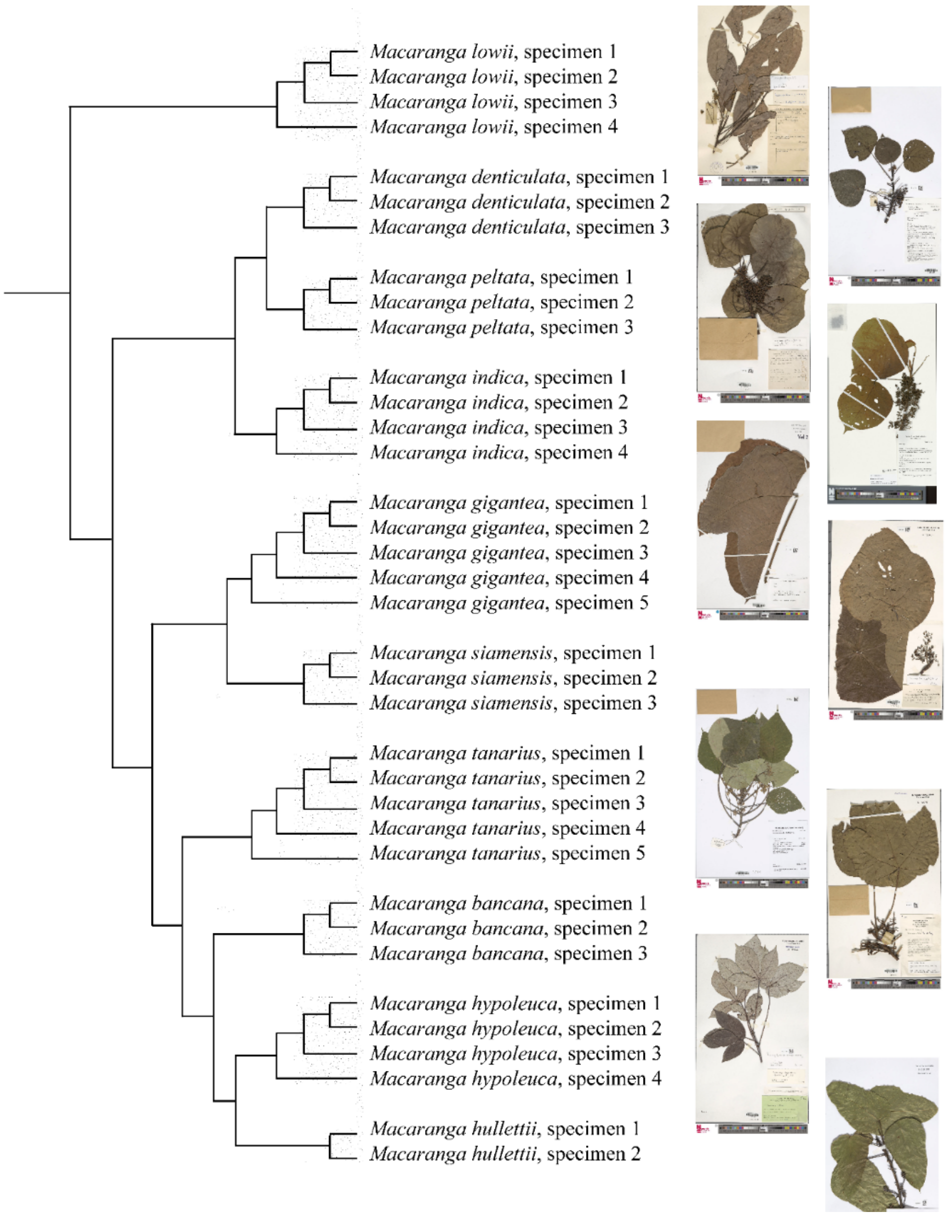

Fig. 5. Example phylogenetic tree of *Macaranga* (Euphorbiaceae) whereby the specimen group in species after DNA analysis via Hyb-Seq.

 https://doi.org/10.1051/e3sconf/202451303013

Figure 6. Logo of the ObsIdentify app as presented in the App Store and Google Play.

3 Cooperation and Visibility

New technology might help to speed up flora writing, but more cooperation will help too. A nice example is the way the Flora of Thailand is handled. Especially for large families (somehow publication in floras is always per family) not only international groups of researchers are invited to join the project, but also biology students interested in plant taxonomy. One person coordinates the revision of the family, but internship students can be supervised by various researchers. Unfortunately, the number of taxonomists is decreasing worldwide, making cooperation more difficult.

The Flora Malesiana project has described the flora of the Malay Archipelago already for decades, but so far, only very few families have been contributed by the countries in the region. There is not a real close friendship between various countries and most countries give priority to describing their national flora (Flora of the Malay Peninsula, Tree Flora of Sabah and Sarawak, Flora of Singapore, etc.). However, cooperation between these countries can also be organised differently. Each country can for instance revise the national species in one family, after which a coordinator will combine the various contributions into one treatment. Presently, Indonesia is revising the Indonesian species of the family Pandanaceae (screw palms), known for their delicious smell and green dye, often used in cooking. If other countries do the same, then the coordinator in Switzerland can create one treatment.

A group that can also help are the citizens. Many of them can be interested in collecting plants or to make observations. In the Netherlands, there are various citizen science projects. A simple one, to make people aware of the biodiversity in their neighbourhood, is the ‘street plant project’, learning people to appreciate the weeds growing in the city. Examples of observations are counting projects, like bees, birds and sea shells. During a day per year people, for instance, identify and count the bees, bumble bees, hover flies and wasps in their garden for c. 30 minutes. They can use a form with pictures of the insects for this purpose (Figure 7). The results can be sent by cell phone, which not only gives quick results but because of the GPS function of the phones, the place of occurrence is also noted. The use of the ObsIdentify app already contributed to species new to the Dutch flora [5].

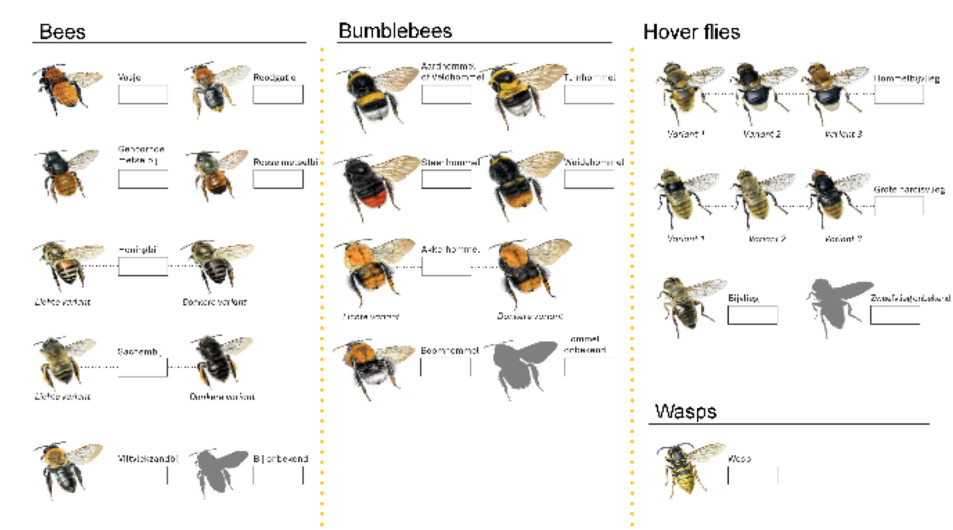

Fig. 7. Form to be used with the national bee count in the Netherlands. The pictures help to identify the insects, the squares can be used for the counting.

The counting and the quick results, visible via the internet, motivate people to cooperate with these projects. And that is another problem for floras. Flora Malesiana treats family by family [8]. In the beginning, the small families were done and results were soon available. Now the large and larger ones are left and they sometimes take decades to finish. In the meantime, revisions of the various genera are published in various journals but are not yet visible as a whole. Thus the 1/3rd of the FM is in fact an understatement as much is already done, only not published as an edition of the flora. When a family is finally published, it appears as a book with lots of text, few drawings and perhaps some photos [9]. Not appealing to most people. Therefore, the visibility and attractiveness of FM have to be increased. This can be done via the internet, which has the potential to include lots of photos and update the text when changes are published, and small contributions can already be added. For instance, if Indonesia finishes the Pandanaceae, then it can already be shown on the internet and can be used; it will also help the researchers revise in the other countries. With enough photos, an identification app can be developed. The software already exists, and the photos of the various species can be used to learn the software to recognize them. Once identified, the app can transfer the user to other information, like a description, distribution map, ecological preferences, uses, vernacular names, etc. The higher the visibility, the more use, the more appreciation of nature, and the more help in maintaining and protecting biodiversity. Books are often expensive, while the internet is cheaper and thus it can reach a much larger public.

4 Monitoring

The primary purpose of floras is to aid in the identification (naming) of plants and to provide basic information on morphology and, for instance, distribution [10]. However, its use can be taken a step further by starting to monitor the plants identified. Monitoring results in knowledge of the quality of species in their surroundings, are they disappearing, coping, or becoming a pest? Monitoring makes obvious which species are threatened by land use and climate change, or which species start to flourish. The cause of the changes can then be investigated, is it the use of pesticides, or fertilizer, the natural habitat becoming too small, the length or extremeness of the dry period, flooding, etc.? Some of the causes will be difficult to change with countermeasures, others are easier. Once countermeasures have been taken (e.g., less or no more pesticides, etc.) then the monitoring can be used to see whether or not the measures are effective.

Ideally one monitors the complete flora, all species. However, in countries with a very rich flora, like in the Malesian region, that is impossible. Then one should focus on indicator species, species that are typical for a certain surrounding (e.g., beach, primary rainforest, heath forest, freshwater swamp forest, etc.) and monitor those.

In the Netherlands, a foundation, Floron (Floristic Research of the Netherlands, Fig. 8), coordinates and organises yearly inventories of the whole country, mainly via volunteers (citizen science again!). Volunteers make inventories of the plant species per square kilometre, again with their phone, and of some indicator species even estimates of their numbers. The results are shown on the website (https://www.floron.nl/, in Dutch) and are important sources of data for nature conservation clubs, managers of natural areas and the authorities.

Fig. 8. Logo of Floron, a foundation for floristic research in the Netherlands.

In the Netherlands, because of all the increased interest in nature and its threats, especially by younger people, started to result in a much higher awareness of biodiversity and even legislation that favors biodiversity. Some examples: Most building projects stopped in the Netherlands due to a court ruling that the amounts of nitrogen have to decrease based on European legislation. Many people stop with using fossil energy sources and switch to solar and wind energy by placing solar panels on their roofs and/or financially participating in windmill projects. Also, other alternatives for fuel, like hydrogen are being developed. It is or should be obvious that we have to counter the damage we are inflicting on the earth. It does not matter what you do, how small or how big, but: Do something (Fig. 9).

Fig. 9. Help save the earth!

5 Conclusion

Describing the flora of the Malay Archipelago can likely speed up by using new techniques (Hyb-Seq, Image Recognition) and more collaboration (citizen science, students). The flora has to become more visible, updatable, and freely accessible via the internet and apps. The results should be used for monitoring the flora to evaluate the effects of land use, climate change and measures to counter these.

References

1. C. G. G. J. Van Steenis, Flora Malesiana ser. 1, vol. 1 (Noordhoff-Kolff N.V., Djakarta, 1950)
2. M. C. Roos, Flora Malesiana Bulletin **11** (1992)
3. R. T. Corlett, R. B. Primack, Tropical Rain Forests: An Ecological and Biogeographical Comparison, ed. 2. (Wiley-Blackwell, Chichester, 2011)
4. P. C. van Welzen, J. W. F. Slik, J. Alahuhta, *Plant distribution patterns and plate tectonics in Malesia. In Plant diversity and complexity patterns: local, regional and global dimensions*, in Proceedings of an International Symposium held at the Royal Danish Academy of Sciences and Letters, 25-28 May, Denmark (2003)
5. H. Duistermaat, Heukels' Flora van Nederland, 24th ed. (Noorhoff Uitgevers B.V., Groningen, Utrecht, 2020)
6. B. Thiers, Continuously updated. Index Herbariorum. A global directory of public herbaria and associated staff (New York Botanical Garden's Virtual Herbarium, New York, 1998)
7. T. Andermann, M. F. Torres Jiménez, P. Matos-Maravi, et al., Frontiers in Genetics **10**, 1407 (2020)
8. L.G. Saw, E. Soepadmo, Flora Malesiana Bulletin **13**, 2 (2002)
9. M. H. Alford, Brittonia **73**, 4 (2021)
10. T. Y. Petrisheva, I. N. Usacheva, A. L. Kravchenko, A. V. Polteva, T. A. Masina, IOP Conference Series: Earth and Environmental Science **981**, 4 (2022)