

Chapter 8

Citizen scientists contribution to monitor Asian biodiversity: A case study on flora of Peninsula Malaysia

Sue Han Lee¹, Hervé Goëau², Alexis Joly³, Simying Ong⁴, Yuen Peng Loh⁵, Antoine Affouard^{2,3}, Jean-Christophe Lombardo³, Julien Claude^{6,7}, and Pierre Bonnet²

¹ Swinburne University of Technology, Kuching, Sarawak, Malaysia

² AMAP, Univ Montpellier, CIRAD, CNRS, INRAE, IRD, Montpellier, France

³ INRIA Sophia-Antipolis - ZENITH team, LIRMM - UMR 5506 - CC 477, 161 rue Ada, 34095 Montpellier Cedex 5, France

⁴ Faculty of Computer Science and IT, University of Malaya, Malaysia

⁵ Faculty of Computing and Informatics (FCI), Multimedia University, Cyberjaya, Malaysia

⁶ Institut Des Sciences de L'Evolution de Montpellier, UMR CNRS/UM/IRD/EPHE, 2, Pl. E. Bataillon, cc65, 34095, Montpellier, France

⁷ Department of Biology, Faculty of Science, Chulalongkorn University, Bangkok, Thailand

Abstract

Four of the twenty five diversity hotspots cover Southeast Asia: Sundaland, Philippines, Indo-Burma and Wallacea. All these hotspots gather a large number of endemic species and ecosystems, accounting for 20% of the world's plant, animal and marine species. A better knowledge of this diversity and distribution is thus essential to enable the implementation of adapted management and protection strategies. However, as human resources and expertise to achieve this are limited, it is therefore necessary to assess the interest of a citizen mobilisation that could multiply the capacities to record and analyze this biodiversity. As it is difficult for a non-specialist to quickly and correctly identify a large number of species, the development of new methods based on automatic visual approaches should facilitate and increase citizen participation. Recent research results have shown a significant improvement in the field, however, the evaluation of such an approach to monitor flora or fauna on a large scale, in the context of high endemism, has never been done. We report here an experiment implemented at the scale of the flora of the Peninsular Malaysia, based on the Pl@ntNet citizen science platform. This study highlights the interest of Malaysian society in this type of approach, and lays the groundwork to facilitate the deployment of similar methodology in other Asian countries. **Keywords:** Artificial intelligence, Biodiversity informatics, Citizen science, Malaysian flora, Plant diversity

8.1 Introduction

The increase in world population and the resulting need for agricultural production are contributing to the growing pressure on natural resources and ecosystems. These new, fast, and greater pressures, exerted by the results of human activities, require us to rethink the way we manage and conserve biodiversity. This is in order to limit the impacts it suffers, but also to maximise its many benefits for the human subsistence. In the tropics, diversity is often one of the most threatened but also one of the less well known. Species distributions and spatio-temporal dynamics are indeed difficult to know without a good monitoring and inventory of diversity. These tasks are however complicated as they require extensive fieldwork, expertise and time.

The emergence of initiatives based on citizen science approaches to biodiversity monitoring in recent years has profoundly transformed the way by which we conduct biodiversity monitoring (Chandler et al., 2017). As proof, the five largest datasets that today contribute to the global biodiversity monitoring system (through the GBIF platform¹), are derived from extensive participatory science programs in Europe and North America. Although there has been considerable development in recent years, few citizen science initiatives are specific to tropical regions. For example, only two of the ten largest datasets of GBIF occurrences are carried by institutions based in tropical regions (respectively the Southern African Bird Atlas in Africa, and the OEH Atlas of NSW Wildlife in Australia). There is thus a big room of improvement in the development of citizen science in the tropics, and in Asia in particular.

Citizen science has considerable potential to increase contributions in both "amount of data" and "number of contributors" in tropical regions. However, this can only be done if a long-term partnership between scientists and citizens is established (Pocock et al., 2018). Such a partnership already seems to be in place for the monitoring of animals since there are four times as many occurrences of animals as plants on GBIF platform. Several initiatives in Southeast Asia have already allowed to record occurrences data, particularly for birds, starting with local initiative of BirdLife International² and with the eBird³ science platform. Other organisms received less attention regarding plants or other animals although iNaturalist⁴ or Pl@ntNet⁵ have local contributors.

The identification of plants is indeed a real challenge for specialists, especially for certain groups that are very rich in numbers of species in the tropics, and even more so for citizens. Several hundreds or even thousands of different plant species may coexist in restricted geographical areas, especially in tropical regions, making their identification and monitoring by non-specialists

¹ <https://www.gbif.org/>

² <https://www.birdlife.org/>

³ <https://ebird.org/>

⁴ A joint initiative of the California Academy of Sciences and the National Geographic Society, dedicated to manage a worldwide network of naturalists

⁵ <https://plantnet.org/en/>

even more complex. Biodiversity hotspots themselves are challenging because to be recognized as "Hotspot", these regions must gather at least 1,500 plant species and must have lost at least 70% of their original vegetation. Moreover, the use of plants in citizen science programs involving children is made more difficult by the fact that young children have little interest in plants because they are not perceived as living organisms since they move very little on the scale of human temporal perception. It is therefore essential to change this by the implementation of new approaches facilitating the identification of plants on a large scale, especially where plant biodiversity is most important.

This explains the recent investments in the development of new computer techniques (Christin et al., 2019; Wäldchen and Mäder, 2018) based on artificial intelligence to allow the identification of plants by image. The search for high-performance methods for image-based plant species identification is a recent field of research, and is of a significant importance in the development of citizen sciences platforms (Ceccaroni et al., 2019). This work is developed in particular within the framework of the LifeCLEF international scientific forum (Joly et al., 2019), whose plant identification task annually involves dozens of teams around the world, using data sets made available since 2011. Overviews of this plant identification task, which are published at the end of each annual campaign (Goëau et al., 2015, 2016, 2017, 2018, 2019, 2013b, 2011, 2012, 2014b), are used by research teams and IT companies to improve their algorithms and progress on the subject. With the recent significant progress obtained on richly illustrated temperate floras, current work is moving towards tropical floras which are much more difficult to characterize with these approaches, due to the lower volume of visual training data.

We propose in this work to present a new initiative aiming at the development and deployment of an original approach for monitoring Asian flora and more generally diversity, based on researches at the frontier between computer, life and citizen sciences. This study shows the strengths and limitations of participatory approaches for monitoring tropical floras via automated methods based on artificial intelligence. A first part of this study describes the geographical, floristic and technological contexts. It is followed by the presentation of the results, describing both the appropriation by Malaysian citizens and the data produced on the Malaysian peninsula. A third part allows the analysis and provide some perspectives regarding the potential of citizen science and machine learning before concluding.

8.2 Contexts

8.2.1 *The Malaysian flora*

Malaysia, is a Southeast Asian country, located just at the north of the equator and composed of two non-contiguous regions: the Peninsular Malaysia which is a bordering region with Thailand at the north of the

peninsula, and the Northern part of the island of Borneo (on the East side of the peninsula). The Peninsular Malaysia, which host the capital Kuala Lumpur, represents approximately 40% of the surface of the country (330,000 km² in total, at the country scale)(Corporation, 2008). Peninsular Malaysia is dominated by mountainous landscape in its Northern part, heavily forested, with a large proportion of karstic formations. East Malaysia, on the island of Borneo, has a long coastline of more than 2,600 km, and is divided between coastal regions, hills and valleys, and a mountainous landscape in its Southern part, with the Mount Kinabalu (the tallest mountain in Malaysia) located in its Eastern part. Climate is equatorial and characterised by the annual southwest (April to October) and northeast (October to February) monsoons, and as Malaysia is in the tropical cyclone belt, its coasts are sometimes subject to the heavy rainstorms (Saw, 1988).

Malaysia flora is characterized by a dense evergreen rainforest, which covers two-fifths of the peninsula and some two-thirds of Borneo states. The various soil types and elevation produce distinctive vegetation zones from the coastal line up to the mountain forests in higher areas. This rich flora is estimated to over 15,500 vascular plants, with an high percentage of endemism (United Nations Environment Programme, 2004). This explains why its recognized as one of the megadiverse country in the world (Von Rintelen et al., 2017). Malaysian biodiversity is part of the Sundaland biodiversity hotspot, the second richest one at the world scale (Myers et al., 2000). Peninsular Malaysia has around 8,300 species of vascular plants (with a species endemism rate of about 30%), while Malaysian part of Borneo has around 12,000 species (Saw et al., 2010). This indigenous flora makes part of the habitat of a large variety of animal life, including emblematic species such as elephants, tigers, Malayan gaurs, Sumatran rhinoceroses, orangutans and the famous proboscis monkeys. A project dedicated to put online the updated checklist of the flora of Malaysia is in progress (Saw and Chung, 2015), and should soon complement the global effort towards the World Flora Online project (Loizeau and Jackson, 2017).

In order to rationalize resources for this study and to present consistent results, we only focused on the Peninsular Malaysia, due to the volume of data available and the range of the 3G connections, compared to Malaysian part of Borneo. All the methodology and analyses which are provided in the next part of the article, are then focus on that part of Malaysia.

8.2.2 *The Pl@ntNet platform*

The citizen science platform Pl@ntNet¹ was chosen for this study. This platform, which was initiated more than ten years ago, uses web and mobile interfaces to identify plants by an automatic visual identification engine. Initially developed for the flora of the Western Europe, it was later adapted to other regions in America, Africa, Oceania, and finally in Asia. As for other similar programs, identification performance depends of the volume and

¹ <https://plantnet.org/>

quality of the visual data used during training phase of the deep learning model of the identification engine. The increasing volume of participants and data produced explains the increase in recent years of the confidence score automatically computed for the model's predictions, on the images submitted by the user community. Figure 1 shows a continuous progression of this score over the last three years.

The identification service is updated monthly on the basis of new data produced, shared and validated by the network of participants (i.e users who have created an account and become members of the community) (Affouard et al., 2017). Pl@ntNet mainly concerns wild plants (i.e., plants that propagate spontaneously in the natural environment), but also cultivated plants (of agronomic and horticultural interest) and associated weeds. The visibility and exploitation of this platform has accelerated since February 2013, after its deployment on mobile devices (iOS in 2013 (Goëau et al., 2013a), and Android in 2014 (Goëau et al., 2014a)). Since 2013, the number of daily users has doubled each year, reaching more than 500,000 users per day during peaks in 2020. In total, more than 33 million people (of which 2.8 million have created a user account) have used the application worldwide.

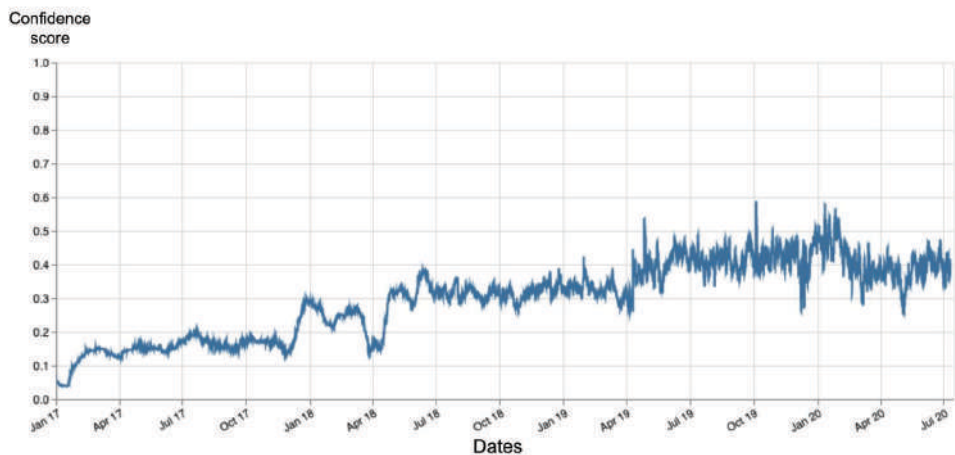


Figure 1. Confidence score of the deep learning plant species identification model.

8.3 Methods of implementation

In order to implement the adaptation of the Pl@ntNet platform to the flora of the Peninsular Malaysia, the eight following steps have been established:

1. The concept of a new flora on Pl@ntNet platform allows the full adaptation of the web and mobile interfaces to a dedicated species list. For the purpose of this study, a collaboration has thus been established with Kew Garden, in order to import the checklist of the flora of Peninsular Malaysia managed within the World Checklist of Selected Plant Families platform (WCSP, 2020) in the Pl@ntNet

form. The full list of valid plant species names with their taxonomic hierarchy has been imported.

2. As the selection of a flora of interest for an app user is automatically done based on its geolocation, an expert-based shape file for the geographical area including the Peninsular Malaysia has been drawn and implemented, in order to provide such service for Malaysian users.
3. Even if the English language is widely used in Malaysia, the app interfaces were translated in Malay with the help of scientists from the University of Malaya (Kuala Lumpur), and in Chinese language with Chinese partners. This has ensured a larger access of the app, for most of the Malay linguistic communities.
4. An aggregation of the collaboratively validated botanical observations on the Pl@ntNet platform, matching the Peninsular Malaysia checklist has been operated. This has allowed to obtain a significant part of that checklist illustrated, even if only a small number of people has used Pl@ntNet in Malaysia before the official launch of that flora.
5. In the same time, an aggregation of the vernacular names managed on the platform, in English, Malay and Chinese was done, in order to display to the users that vernacular names aside the scientific latin names. This step was particularly important to respond to most of the non-scientist users, who often need the common vernacular names to easily remember them, and share information with others participants.
6. This illustrated checklist has then been used during the training phase of the Pl@ntNet deep learning model. This step has contributed to design a beta version of the app for the Malaysian flora. At this stage, specific Malaysian partners who had a Pl@ntNet account were able to test the platform, exploring the data, identifying species already illustrated in the training database, contributing on illustrated or unillustrated species.
7. A field trip was then organized in order to share the results of this work at this stage with partners, and to produce complementary visual botanical observations on unillustrated species of the species checklist.
8. The public launch of the project, was organized in December 2019, during a scientific seminar of one of the co-author (LSH) that was held at the University of Malaya. This event has provided the opportunity to share with students and scientists, the main outcomes

of the project, and the key information needed to correctly use the Pl@ntNet platform in Malaysia.

8.4 Results

8.4.1 Coverage of the Malaysian flora

After seven months of the official launch of the adaptation of the Pl@ntNet platform to the Malaysian flora, the platform covers 1,205 species, belonging to 733 genera and 165 plant families. These data can be explored at the following url: <https://identify.plantnet.org/explo/malaysia/>, on the web version of the platform. The total number of illustrated species represents less than 15% of the full species checklist of the Malaysian peninsula, 40% of the Genera and 68% of the families. The illustrated species checklist is provided with 23,693 vernacular names in one of the 156 different languages which are managed on the platform. The number of vernacular names in English, Malay and Chinese, and the proportion of the illustrated species with these vernacular names are provided in Table 1.

Table 1. Numbers and percentages of common names provided for the three main languages used in Malaysia, on Pl@Net platform for the Flora of Malaysia.

Languages	Nb of common names	Perc. of species common names
English	7,926	69
Malay	322	19
Chinese	210	16

225,697 images illustrate the species checklist, with an average of 187 images per species. The highest represented type of views are flowers (with an average of 78 images per species), followed by leaves (76 img./sp.), growth forms (13 img./sp.), fruits (11 img./sp.) and stems (4 img./sp.). As the performances of the identification search engine are strongly correlated to the volume of training data, these volumes illustrate the types of views which are the most suitable to use to identify plant species at this stage of progress of the platform. The Top 5 families, in term of number of illustrated species are the Fabaceae (119 sp., with 11,112 images), the Poaceae (87 sp., with 12,007 images), the Asteraceae (46 sp. with 53,794 images), the Lamiaceae (42 sp., with 7,859 images), and the Orchidaceae (39 sp., 1,323 images). As we can see, this dataset is highly imbalanced, some species being illustrated by several thousands of images, while most of the others have only few.

The economic importance, based on the GRIN (Germplasm Resources Information Network) platform¹, is provided for 583 different species. As we can see in Table 2, the most represented ones are Medicinal and environmental uses, weed characterization, and uses for material and food, with more than 100 species each. The fact that 48% of the illustrated species has been recognized to

¹ <https://www.ars-grin.gov/>

have an economical impact, illustrated that a strong initial focus has been done to record such important plant species in the platform. This can be explained by their larger accessibility by citizens and professionals, who managed them during their spare time or daily working activities (such as farmers or foresters for example).

The conservation status for 339 species is also provided. The distribution of these status, based on the IUCN red list criteria, is shown in Table 3. This table illustrates that most of the species are "least concern", which is also the status of most of the evaluated plant species at the world level (see IUCN red list website¹, for detailed information).

By looking at the most requested species with the Pl@ntNet identification search engine, we see that the most often searched plants are of horticultural, agricultural or medicinal use (Table 4, columns A and B). However, limiting the analysis to the species which are only in two floras of the Pl@ntNet platform (and not in anymore floras related to other tropical regions), we obtain the most remarkable requested plants of this region (Table 4, columns C and D). Such list illustrates the potential for monitoring rare plant species with this approach at a large geographical scale. Among these ten species, seven of them are illustrated by a small number (few dozens) of occurrences with coordinates on the GBIF platform, which illustrate the strong benefit of collecting any new georeferenced records on that ones. Figure 2, provides the maps of occurrences displayed on the Pl@ntNet platform, for four of these species, in blue the Pl@ntNet occurrences, in yellow the GBIF ones.

Table 2. Numbers of species recorded on Pl@Net platform for the Flora of Malaysia, according to the GRIN Economic importance criteria.

Economic importance criteria	Nb of Species
Medicines	285
environmental uses	277
weeds	222
materials	154
food	137
poisons	90
Forage	67
food additives	39
gene sources	32
Fuels	30
Alternative host	23
Host	14
Social uses	9
Bee plants	8
Pesticide	5

¹ <https://www.iucnredlist.org/search>

Table 3. Numbers of species recorded per IUCN criteria on Pl@Net platform for the Flora of Malaysia.

IUCN conservation status	Nb of Species
CR - Critically Endangered	5
EN - Endangered	8
VU - Vulnerable	7
LR/cd - Lower Risk: Conservation Dependent	2
LR/nt - Lower Risk: Near Threatened	4
LR/lc - Lower Risk: Least Concern	9
NT - Near Threatened	8
LC - Least Concern	290
DD - Data Deficient	6

Table 4. List of the most requested plant species for the Flora of Malaysia on the Pl@ntNet platform. Columns A and B, for species related to any floras on the Pl@ntNet platform. Columns C and D, for species related to only one other flora of the platform. Species are ordered by decreasing number of sent identification requests.

(A) Top position	(B) Species names	(C) Top position	(D) Species names
1	Dieffenbachia seguine	1	Scindapsus pictus
2	Ficus elastica	2	Musa × paradisiaca
3	Dracaena fragrans	3	Hoya kerrii
4	Canna indica	4	Solanum trilobatum
5	Ocimum basilicum	5	Dillenia suffruticosa
6	Epipremnum aureum	6	Rhaphidophora tetrasperma
7	Mangifera indica	7	Ficus maclellandii
8	Syngonium podophyllum	8	Phrynium pubinerve
9	Tabernaemontana divaricata	9	Clinacanthus nutans
10	Epipremnum pinnatum	10	Nepenthes × neglecta

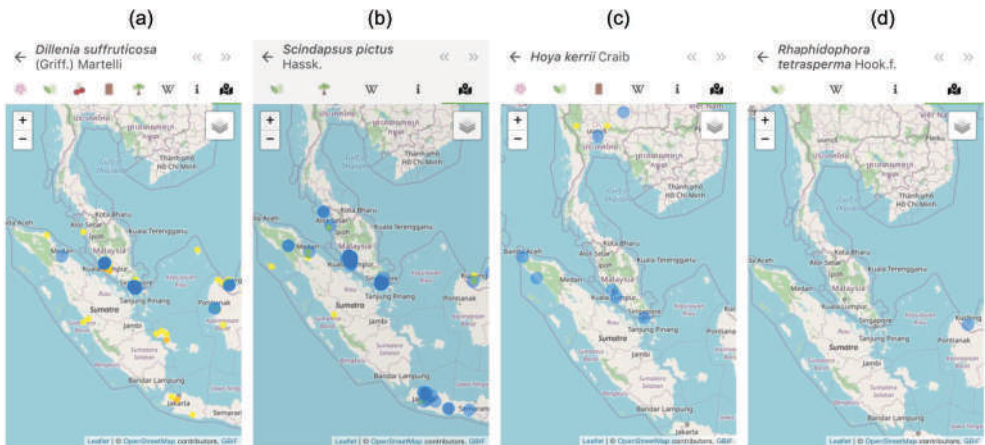


Figure 2. Mobiles interface providing occurrences managed on Pl@ntNet (in blue) and GBIF (in yellow) for four requested species on the Flora of Malaysia.

8.4.2 A fully adapted citizen science platform

Pl@ntNet framework offers 3 main front-ends, an Android app (the most used one, accessible on Google Play¹), an iOS app (available on Apple Store²) and a web interface³. All of them, freely accessible, can be used without user account. They allow to submit from one up to four pictures of the same plant, in order to get a list of the most probable species in return. The list of candidate species predicted for a searched plant observation is displayed in decreasing order of confidence and each species is illustrated by the most similar picture of that species. If the user selects a species of the result, the other retrieved images are displayed by decreasing visual quality score (grouped by types of views, such as: flower, fruit, leaf, stem, habit). At this stage, the user can access to richer information about the species such as all the available illustrations in the dataset (allowing to refine the identification), map of the species occurrences, or descriptive web pages providing useful complementary information about the plants. Each species page is linked to the GBIF web website, allowing to explore data aggregated by all research institutions and citizen sciences initiatives shared on this worldwide platform. Invasive species descriptions are available throughout the link the Invasive species compendium⁴ of the CABI⁵, and useful plants are described throughout the Prosea encyclopedia⁶ hosted on the Pl@ntUse website. Illustration of 4 interfaces of the mobile app are provided in Figure 3.

Complementary functionalities are provided on the web version of the platform such as (i) the access to all vernacular names in all languages managed by the platform (while it is restricted in the language of the app selected by the user on the mobile version), (ii) graphs showing the phenological trends of the species for leaf, flowers and fruits, based on platform occurrences, (iii) access to the exploring space of the user data (allowing him to export all his data in csv file).

8.4.3 Platform uses

Since the launch of the adaptation of the Pl@ntNet platform to the Malaysian flora, we see a growing number of contributors. Indeed, Figure 4 shows that 240 people have contributed to this project, by sharing illustrated botanical observations, during the first seven months of the year 2020. Even if this number of contributors is quite high, we can see when it is compared to Figure 5 that it's only a small part of the full number of users who have used the application without sharing observations. The Figure 5 first shows that more than 6,000 people were active early in January 2020. This illustrates that the

¹ <http://bit.ly/1K4D1eU>

² <http://apple.co/2cMtWgu>

³ <https://identify.plantnet.org/>

⁴ <https://www.cabi.org/ISC>

⁵ Commonwealth Agricultural Bureaux International

⁶ https://uses.plantnet-project.org/en/Introduction_to_PROSEA_on_Pl@ntUse

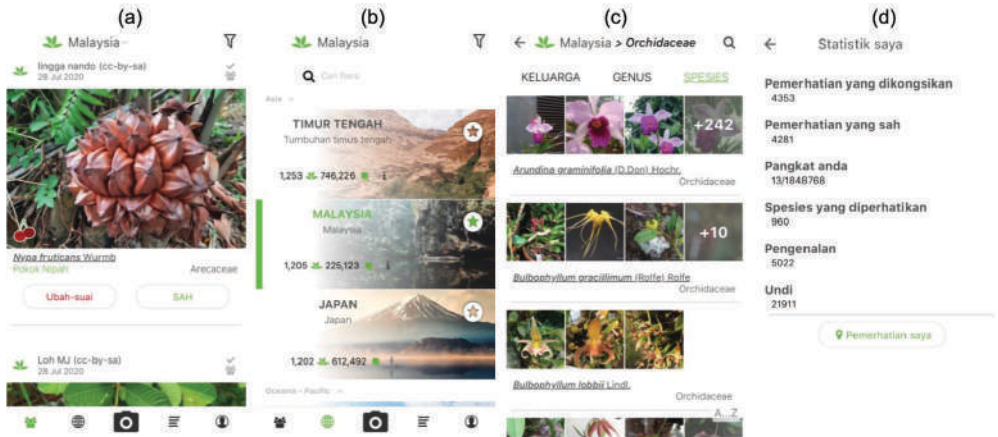


Figure 3. Screenshots of the mobile interfaces of the Pl@ntNet app translated in Malay language, illustrating : (a) the data stream of the user contributions, (b) the different floras accessible in Asia on the platform, (c) taxonomic explorer on the orchids species for the Flora of Malaysia, (d), personal statistics of the user.

visibility and use of that platform was quite high, even before its adaptation to the Malaysian flora. Secondly, we can see a peak of use the 15th March 2020 (most probably related to a communication in the national medias) and a progressively growing number of active users reaching more than 9,700 users at the end of July 2020. During the same period 55,283 user sessions have been recorded, which allows to estimate at more than 5, the average number of usage sessions per user. This volume of use explains why Pl@ntNet android app, is in the top 200 of the educational free apps on the Google play store, since the 1st of June 2020 (source : Sensor Tower).

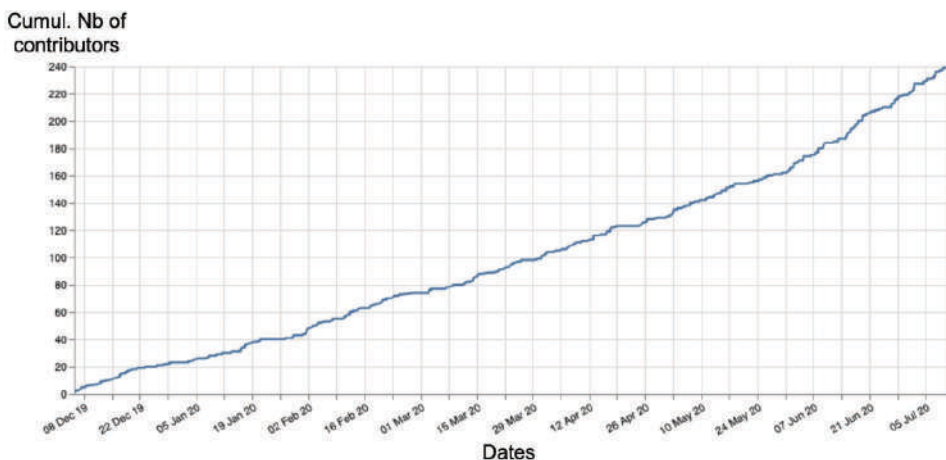


Figure 4. Cumulative number of contributors from Peninsular Malaysia since December 2019.

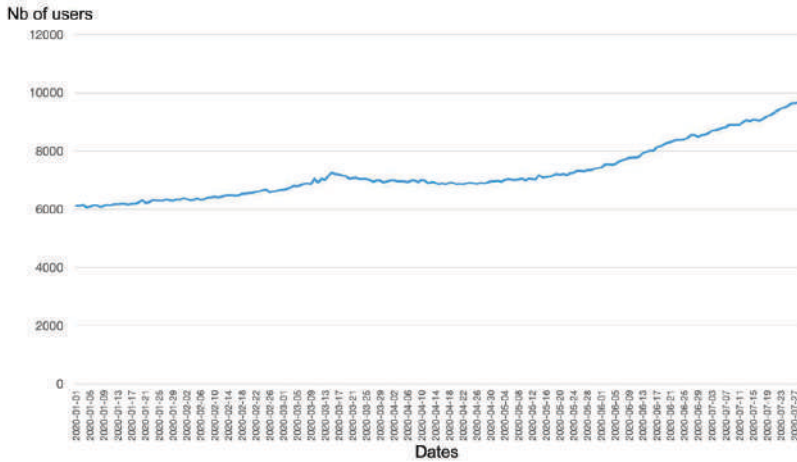


Figure 5. Daily number of active users from Peninsular Malaysia, since January 2020.

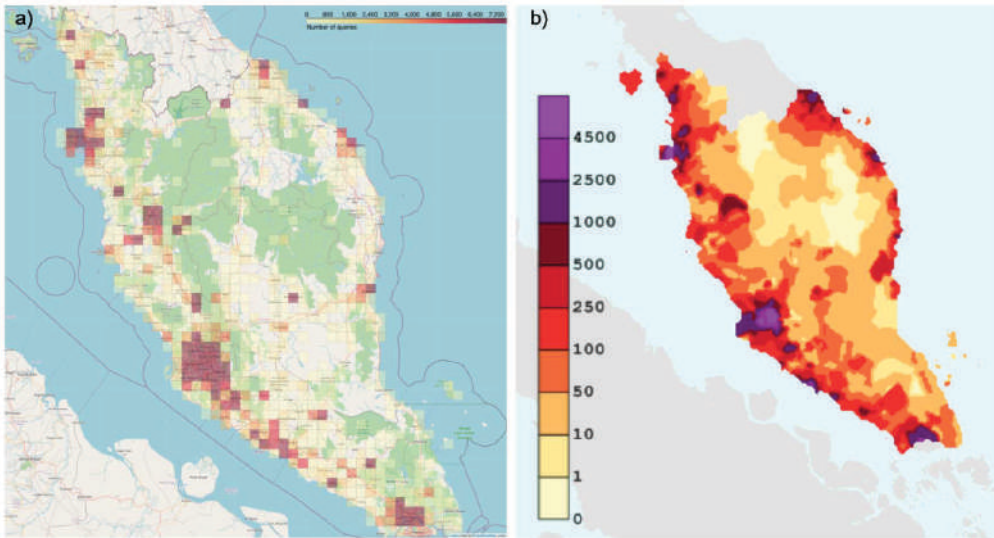


Figure 6. (a) Map of the identification requests sent on the Pl@ntNet platform, from Peninsular Malaysia, (b) Estimated population density, shown as persons per km², from the Socioeconomic Data and Applications Center.

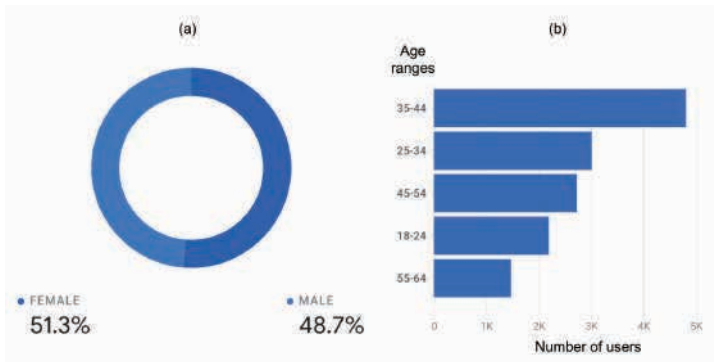


Figure 7. User profile of the Pl@ntNet platform in Malaysia, (a) Percentage of users per gender balance, (b) Number of users per age ranges.

As we can see in Figure 6, the geographical distribution of the identification requests sent from Peninsular Malaysia (Figure 6 (a)) seems to be highly correlated to the human population density (Figure 6 (b)). The highest volumes of identification requests are located around Alor Setar, Seberang Perai, Ipoh, Kuala Lumpur, Malacca City and Johor Bahru from the North to the South of the West coast, and Kota Bharu, Kuala Terengganu, Kuantan on the East coast. Another visible geographical pattern visible on Figure 6 (a) is the decreasing volume of identification requests with the distance to main roads. This can contribute to explain why natural reserve on the West coast are much more prospected by Pl@ntNet user community than natural reserves of the Central part of the country (such as the Taman Negara National park, Gunung Stong State Park and the Bang Lang national park). More surprisingly, users were also recorded on Tioman and Pemanggil islands, on the South East coast of the Peninsula, even if they have a very small populations density.

In Malaysia, more than 90% Pl@ntNet mobile app users use it from an Android smartphone, while this percentage is of 80% at the world scale. Another noticeable difference of the user profile in Malaysia, is the fact the platform is more used by women (with 51.3%) than men, while it is the opposite at the world scale (Figure 7 (a)). Figure 7 (b) shows that a significant part of the users are young people. The age range of the 18-24 years in particular, represents 12% of users, while it is of only 5% in France, where the project has been initially launched.

8.4.4 Usage in nearby areas

In the aim to evaluate the impact of the deployment of this new flora on the platform, the cumulative number of queries sent has been computed for Peninsular Malaysia, East Malaysia (i.e Sarawak and Sabah states), and Thailand in Figure 8. We can see that there is one order of magnitude between Peninsular Malaysia and Thailand, and two orders of magnitude with East Malaysia. Nevertheless, this Figure also illustrates the interest of the people in these nearby areas, even if the citizen science platform is not yet fully adapted to response to their needs and their region. Figure 9, which shows the geographical distribution of these identification requests, allows to see that the platform usage is not only concentrated in urban areas, but also in rural ones. The coverage of the Thai territory in particular, is quite important, which suggests that an adaptation of the platform to this territory would allow a significant participation of that population.

8.5 Difficulties and challenges

8.5.1 Engagement and recognition of participants

Due to the recent deployment of the Pl@ntNet citizen science platform on Malaysian flora, it is not yet possible to get strong conclusions on the

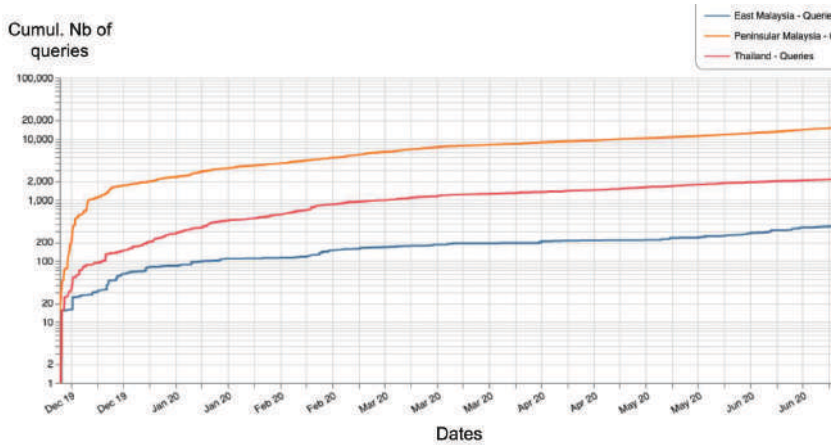


Figure 8. (a) Cumulative number of identification requests sent from Peninsular Malaysia (orange), East Malaysia (blue), and Thailand (red).

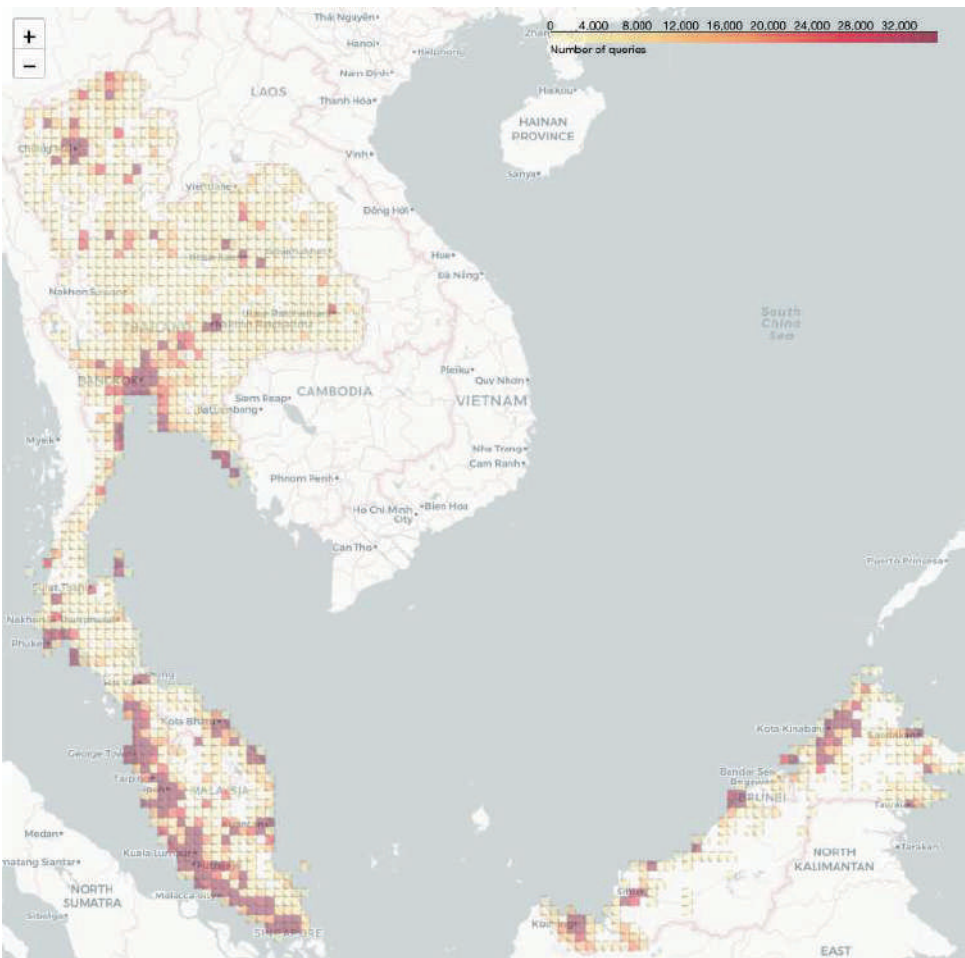


Figure 9. Map of the identification requests sent from Thailand, Peninsular Malaysia and East Malaysia.

participants engagement. Nevertheless, we can already see some limitations in the users participation. The number of active contributors, who produce and share high quality plant observations is quite low, compare to the full number of users in Malaysia. The choice of the development team to keep the platform open, even for unregistered users, contribute to the large visibility of this initiative in the country. Nevertheless this community of users (registered and unregistered ones) is probably much more diverse than iNaturalist community for which the user registration is mandatory.

The long term user engagement has been recognized as an important limitation in citizen science projects. In the Pl@ntNet platform, in which we can see an increasing number of registered users since the beginning of the initiative, the improvement of the identification performances is one of the key feature which contributes to keep participants motivated and active. As their contributions directly enrich the identification service from which they benefit, they can easily see the results of their investment on a quite short period. Nevertheless others incentive mechanisms have to be established in order to stimulate contributors motivation on the long term. In this aspect, a good recognition of the participation is important, this is why actual efforts are conducted to improve such recognition and visibility among contributors. The recent share of two Pl@ntNet datasets (Affouard et al., 2020a,b) on Gbif platform contributes to that recognition, by increasing contributors visibility up to the largest worldwide initiative in Biodiversity informatics.

8.5.2 *Rural vs urban sampling*

Geographical distribution of the identification requests (Figure 6) and list of the most requested species (Table 4), clearly show that, at present time, Pl@ntNet platform is largely used to record urban species and plants related to human activities. This can be very useful for many purposes (such as horticultural species management, food and medicinal uses, etc.), even for biodiversity conservation. Indeed early detection of invasive species in a new territory can allow quick and efficient management practices. Citizen sciences platforms contributes a lot to this type of monitoring (Chandler et al., 2017; Gallo and Waitt, 2011; Giovos et al., 2019, Johnson et al., 2020), and Pl@ntNet has already been used for that purpose (Botella et al., 2018). Nevertheless, increasing the monitoring of rare and rural species with such approach is much more challenging. As people in remote areas have less access to new technologies (such as recent mobile smartphones, high speed internet, etc.), citizen science biodiversity monitoring platforms have to invest in low data consumption applications, and adapted to a wide variety of devices. In order to response to this need, Pl@ntNet team actually experiments technologies dedicated to off-line use of the identification search engine. Such new functionality could contribute to increase the volume of records in isolated areas, such as in dense evergreen forests, or high mountainous environments, which are quite frequent in Malaysia. This is also very important

to limit the geographical bias in the monitoring network, and to ensure a more homogeneous geographical coverage of the observations made. The presence of such biases in Pl@ntNet data requires the use of appropriate statistical methods to identify them and reduce their impacts for the development of accurate species distribution models (Botella et al., 2020). The implementation of incentive mechanisms, favouring the collection of data on poorly observed species or in poorly surveyed areas, could also reduce such biases.

8.5.3 *Digital data on tropical flora*

Due to the unbalanced available data between temperate and tropical region, the development of automated visual methods for plant species identification in the tropics is much more difficult. Indeed, the volume of visual data being much more low and endemism and diversity being higher, accuracy of automated approaches in such contexts is lower. In order to invest that problem, the PlantCLEF task of the LifeCLEF international scientific and technological benchmark has been moved in 2019 to tropical flora (Goëau et al., 2019). In 2020, a new task dedicated to experiment the use of natural history digitized herbarium collections for the identification of field plant images was set up (Joly et al., 2020). Such challenge could help to take advantage of the hundreds of thousands of herbarium specimens already collected and digitized on the tropical flora of Southeast Asia, to improve accuracy of the automated plant species identifications services.

8.6 *Opportunities and perspectives*

The appropriation by young people in Malaysia of this type of technology (Figure 7b) represents a major asset for training in botany, ecology and conservation biology. By reducing the difficulty of access to species identification methods, the sharing of data and knowledge on species is highly facilitated. The transmission of expertise between generations and between specialists/non-specialists should be able to reach unequalled scales. The country's technological development also represents a tremendous advantage for the expansion of the pool of participants. Indeed, the development of telecommunications services in Malaysia (MCMC, 2019), which is accompanied by an increase in the number of mobile connections and a reduction in access costs, facilitates the use of platforms such as Pl@ntNet. The biodiversity conservation community needs to take advantage of these technological developments by taking ownership of such tools and integrating them in a complementary workflow. This type of approach should thus allow a better and quicker interaction between the producers of field data and those who analyse them. In this perspective, the improvement of data quality revision methods is an important step. It is indeed critical that the participants have quick feedback on their contributions, while ensuring that the most expert people of the flora (specialists of a region or a taxonomic group), are not overwhelmed by the

volume of data to be corrected. A good distribution of data requiring validation is therefore important to put in place. Such mechanisms, based on bayesian approaches for assigning botanical validation between users, are thus being experimented within Pl@ntNet.

The protected area network in Peninsular Malaysia covers 2.98 million hectares, managed by various agencies including the Federal Department of Wildlife and National Parks, and the state forestry departments, under the Ministry of Energy and Natural resources. These fairly large areas required a large number of people to monitor and manage them properly. For this reason, the support of civil society members through a citizen science platform such a Pl@ntNet could be useful for early detection of important events (e.g arrival of a new invasive species, flowering of a rare one, etc.) on a management perspective. It can be also important to evaluate the impact of restoration policies in allowing plant successions to be identified. As already illustrated in Europe and Africa (Bonnet et al., 2020), with the implementation of "Pl@ntNet micro-projects" (i.e a contextualized version of the Pl@ntNet platform), Malaysian private and national parks and conservatories could thus benefit from the use of the Pl@ntNet platform to strengthen their capacity to record and manage plant biodiversity based on the basis of citizen investment.

The proven use of the Pl@ntNet platform in the bordering regions of this study (Figure 9) suggests a significant interest in this type of approach. As Malaysia shares a large part of its flora with neighbouring countries (Thailand and Indonesia in particular), international cooperation through this type of platform would allow a significant increase of its impact for the whole region. Feedback on European flora could thus be beneficial to facilitate and accelerate such an approach in Southeast Asia.

The recent and important development of Computer vision in Agriculture (Kalantidis et al., 2020) can lead the development of new form of agricultural practices, less impactful for the environment and the biodiversity. As the weed removal task is one of the main consuming task of chemical products in Agriculture, the early and wide automated identification of weed species in agricultural fields can contribute to significantly reduce negative impact of agrochemical industry. Experiments in this regard have been conducted on autonomous electric weeding robots (Champ et al., 2020). If their acquisition cost is reduced, their wider use could represent an important benefit for the entire biodiversity conservation community.

Several apps such as Pl@ntNet or iNaturalist, based automated visual identification methods are now available for free. These tools are obvious allies in conservation management while ecosystems are facing to rapid changes. To be operational these platforms need massive data but this study show that in just a few months, it is possible to gather a significantly large number of new observation in various locations. In addition to automated vision-based methods, the development of sound-based programs to identify insect, birds,

frogs or bats still need more development and experts to establish an operational reference line for Southeast Asia. Such automated sound recognition could be a good complement as sonogram can be recorded without targeting a specific need and also because several animals are difficult to observe or to photograph. For instance, application like FrogID¹ produced by the Australian Museum could certainly be quickly adapted to Southeast Asia but may need to establish a good initial baseline to be operational in the region. Automatic bird song recognition applications, could also be a good complement to eBird reports that are based on sightings or photographs

8.7 *Implications for policies*

Pl@ntNet implementation in Europe has contributed with its computational services to develop new environmental education activities. Such activities, organized in formal and informal courses, can be of great interest to increase nature perception and recognition of the biodiversity contribution to human life style by the youngest people. In Malaysia, as we can already see an interest of the youngest generations to this platform, we can hope a larger appropriation of it by the teachers and educators communities. Such appropriation could facilitate students engagement in Ecology and Biological training, due to the mobile and multimedia nature of the Pl@ntNet platform. This appropriation can also stimulate community oriented conservation management and local development as it can be rapidly used to produce diversity inventories without the need of experts on sites. Bioblitzes, such as those sometimes initiated by NGOs or local authorities, can be punctual events that allow for a rapid inventory of a given geographical area if there are enough participants. Such events can easily be initiated by teachers, universities and museums.

Education is far from being the only area where Pl@ntNet and similar participative taxonomy applications could contribute in Malaysia. Several Sustainable Development Goals (SDG) related to "Life on Land", recognized by the United Nation, could take benefit of the development of that platform in Asia. The sustainable use of terrestrial ecosystems and their services (SDG 15.1) could be increased by a better recognition of the society of the individual contribution of the species to our daily life (as mentioned before). Due to the volume of the georeferenced observations of plants and animals produced by the network of users, Pl@ntNet and applications regarding animal diversity can significantly contribute to refine geographical delimitation of the regional and national protected areas (contributing in that aspect to the SDG 15.5). With the support of Policy makers, two others SDGs could take a large advantage of the platform development: SDG 15.7 wich is related to the elimination of the trafficking of the protected species, and SDG 15.8, which is focused on the control of the invasive alien species. In both cases, field actors in charge

¹ <https://www.frogid.net.au/>

of the application of the regional and national policies related to these two SDGs, could take advantage of the accuracy of automated identification system to quickly detect protected and invasive species, in order to take appropriate decision for their control.

8.8 Conclusion

This paper describes the methodology implemented to adapt a citizen science platform for plant biodiversity monitoring at a national scale in Asia. The use of a mobile app combined with artificial intelligence to assist the identification of plants made it easy to involve users and contributors, interested to enlarge their botanical knowledge. Even if the observations produced, are not uniformly distributed on the territory, they can be of real interest for threatened and invasive species monitoring, and to a lesser extent to monitor species for which geographical information are much less available at large scale. The platform implementation of Pl@ntNet in Malaysia has shown difference between the community of users here, compared to European users (for which the proportion of men and older users is more important). Even if it is difficult to draw conclusion on the impact of its implementation, due to its recent deployment, we can already see a large participation and an increasing use of it. This encourage us to push its adaptation in other Asian countries, in order to ensure the widest benefit as possible.

Acknowledgements

This work would not have been possible without the participation of the thousands of Pl@ntNet users in Malaysia, who made this initiative possible. This research project was supported by Agropolis Fondation, Numev, Cemeb, #DigitAG, under the reference ID 1604-019 through the "Investissements d'avenir" programme (Labex Agro:ANR-10-LABX-0001-01). This work was also supported by the French National Research Agency under the Investments for the Future Program, referred as ANR-16-CONV-0004. Finally, this project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No° 863463 (Cos4Cloud project). Finally, authors would like to thank Agropolis Fondation which has supported the development of Pl@ntNet platform since 2009. For the Malaysian counterpart, the authors wish to express their gratitude to Dr Yong Kien Thai from the University of Malaya. Thanks to his support, images of 45 species of Malaysian flora were successfully captured in Rimba Ilmu for the Pl@ntNet Malaysian Flora project. The authors would also like to thank Neuron AI for their helps in translating the description of the Pl@ntNet application on Google play.

References

- Affouard, A., Joly, A., and Bonnet, P. (2020a). Pl@ntnet automatically identified occurrences. pl@ntnet. occurrence dataset. <https://doi.org/10.15468/mma2ec> accessed via gbif.org. (accessed 1 August 2020).
- Affouard, A., Joly, A., and Bonnet, P. (2020b). Pl@ntnet observations. pl@ntnet. Occurrence dataset. <https://doi.org/10.15468/gtebaa> accessed via gbif.org. (accessed 1 August 2020).
- Affouard, A., Lombardo, J. C., Goeau, H., Bonnet, P., and Joly, A. (2017). Pl@ntnet app in the era of deep learning. 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Workshop Track Proceedings.
- Bonnet, P., Joly, A., Faton, J. M., Brown, S., Kimiti, D., Deneu, B., ... and Munoz, F. (2020). How citizen scientists contribute to monitor protected areas thanks to automatic plant identification tools. *Ecological Solutions and Evidence*, 1(2), e12023.
- Botella, C., Joly, A., Bonnet, P., Monestiez, P., and Munoz, F. (2018). Species distribution modeling based on the automated identification of citizen observations. *Applications in Plant Sciences*, 6(2), e1029.
- Botella, C., Joly, A., Monestiez, P., Bonnet, P., and Munoz, F. (2020). Bias in presence-only niche models related to sampling effort and species niches: Lessons for background point selection. *PLoS One*, 15(5), e0232078.
- Ceccaroni, L., Bibby, J., Roger, E., Flemons, P., Michael, K., Fagan, L., and Oliver, J. L. (2019). Opportunities and risks for citizen science in the age of artificial intelligence. *Citizen Science: Theory and Practice*, 4(1), 1-14.
- Champ, J., Mora-Fallas, A., Goëau, H., Mata-Montero, E., Bonnet, P., and Joly, A. (2020). Instance segmentation for the fine detection of crop and weed plants by precision agricultural robots. *Applications in Plant Sciences*, 8(7), e11373.
- Chandler, M., See, L., Copas, K., Bonde, A. M., López, B. C., Danielsen, F., Legind, J. K., Masinde, S., Miller-Rushing, A. J., Newman, G., et al. (2017). Contribution of citizen science towards international biodiversity monitoring. *Biological Conservation*, 213, 280–294.
- Christin, S., Hervet, E., and Lecomte, N. (2019). Applications for deep learning in ecology. *Methods in Ecology and Evolution*, 10(10), 1632–1644.
- Corporation, M. C. (2008). *World and Its Peoples: Malaysia, Philippines, Singapore, and Brunei*.
- Gallo, T. and Waite, D. (2011). Creating a successful citizen science model to detect and report invasive species. *BioScience*, 61(6), 459–465.
- Giovos, I., Kleitou, P., Poursanidis, D., Batjakas, I., Bernardi, G., Crocetta, F., ... and Katsanevakis, S. (2019). Citizen-science for monitoring marine invasions and stimulating public engagement: a case project from the eastern Mediterranean. *Biological Invasions*, 21, 3707–3721.
- Goëau, H., Bonnet, P., and Joly, A. (2015). LifeCLEF Plant Identification Task 2015. In *CLEF: Conference and Labs of the Evaluation Forum*, volume CEUR Workshop Proceedings of Working Notes of CLEF 2015 - Conference and Labs of the Evaluation forum, Toulouse, France.
- Goëau, H., Bonnet, P., and Joly, A. (2016). Plant Identification in an Open-world (LifeCLEF 2016). In *CLEF: Conference and Labs of the Evaluation Forum*, volume CEUR Workshop Proceedings.
- Goëau, H., Bonnet, P., and Joly, A. (2017). Plant identification based on noisy web data: the amazing performance of deep learning (LifeCLEF 2017). In *CLEF: Conference and Labs of the Evaluation Forum*, volume CEUR Workshop Proceedings, Dublin, Ireland.
- Goëau, H., Bonnet, P., and Joly, A. (2018). Overview of ExpertLifeCLEF 2018: how far automated identification systems are from the best experts?. In *CLEF - Conference and Labs of the Evaluation Forum*, volume CEUR Workshop Proceedings, Avignon, France.
- Goëau, H., Bonnet, P., and Joly, A. (2019). Overview of LifeCLEF Plant Identification Task 2019: diving into Data Deficient Tropical Countries. In *CLEF 2019 - Conference and Labs of the Evaluation Forum*, volume 2380.
- Goëau, H., Bonnet, P., Joly, A., Affouard, A., Bakić, V., Barbe, J., Dufour-Kowalski, S., Selmi, S., Itheri, Y., Vignau, C., Barthelemy, D., and Boujemaa, N. (2014a). *Pl@ntNet Mobile 2014: Android port*

- and new features. In ICMR: International Conference on Multimedia Retrieval.
- Goëau, H., Bonnet, P., Joly, A., Bakić, V., Barbe, J., Selmi, S., Carré, J., Barthélémy, D., Boujemaa, N., Molino, J. F., Duché, G., and Perronet, A. (2013a). Pl@ntNet Mobile App. In ACM Multimedia Conference, MM '13, October 21-25, 2013.
- Goëau, H., Bonnet, P., Joly, A., Bakić, V., Barthélémy, D., Boujemaa, N., and Molino, J.-F. (2013b). The ImageCLEF 2013 Plant Identification Task. In CLEF - Conference and Labs of the Evaluation Forum, CEUR Workshop Proceedings, Valencia, Spain.
- Goëau, H., Bonnet, P., Joly, A., Boujemaa, N., Barthélémy, D., Molino, J.-F., Birnbaum, P., Mouysset, E., and Picard, M. (2011). The ImageCLEF 2011 plant images classification task. In CLEF - Conference and Labs of the Evaluation Forum, Amsterdam, Netherlands.
- Goëau, H., Bonnet, P., Joly, A., Yahiaoui, I., Barthélémy, D., Boujemaa, N., and Molino, J. F. (2012). The ImageCLEF 2012 Plant Identification Task. In CLEF'2012: Conference and Labs of the Evaluation Forum, Rome, Italy.
- Goëau, H., Joly, A., Bonnet, P., Selmi, S., Molino, J.-F., Barthélémy, D., and Boujemaa, N. (2014b). LifeCLEF Plant Identification Task 2014. In Cappellato, L., Ferro, N., Halvey, M., and Kraaij, W., editors, CLEF: Conference and Labs of the Evaluation Forum, volume CEUR Workshop Proceedings of Working Notes for CLEF 2014 Conference, pp. 598–615.
- Johnson, B. A., Mader, A. D., Dasgupta, R., and Kumar, P. (2020). Citizen science and invasive alien species: An analysis of citizen science initiatives using information and communications technology (ICT) to collect invasive alien species observations. *Global Ecology and Conservation*, 21, e00812.
- Joly, A., Goëau, H., Botella, C., Kahl, S., Servajean, M., Glotin, H., Bonnet, P., Planqué, R., Robert-Stöter, F., Vellinga, W.-P., et al. (2019). Overview of lifeclef 2019: Identification of amazonian plants, south and north american birds, and niche prediction. In International Conference of the Cross-Language Evaluation Forum for European Languages, pp. 387–401.
- Joly, A., Goëau, H., Kahl, S., Botella, C., Ruiz De Castaneda, R., Glotin, H., ... and Müller, H. (2020). Lifeclef 2020 teaser: Biodiversity identification and prediction challenges. In European Conference on Information Retrieval.
- Kalantidis, Y., Sevilla-Lara, L., Mwebaze, E., Machuve, D., Alemohammad, H., and Guereña, D. (2020). Proceedings of the iclr workshop on computer vision for agriculture (cv4a) 2020. arXiv:2004.11051.
- Loizeau, P.-A. and Jackson, P. W. (2017). World flora online mid-term update. *Annals of the Missouri Botanical Garden*, 102(2):341–346.
- MCMC (2019). Connectivity, key to digital transformation. industry performance report 2018. In Malaysian Communications and Multimedia Commission. <https://www.mcmc.gov.my/skmmgovmy/media/General/pdf/Industry-Performance-Report-2018.pdf>.
- Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A., and Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403(6772), 853–858.
- Pocock, M. J., Chandler, M., Bonney, R., Thornhill, I., Albin, A., August, T., ... and Danielsen, F. (2018). A vision for global biodiversity monitoring with citizen science. In *Advances in Ecological Research* (pp. 169-223). Academic Press.
- Saw, L., Chua, L., Suhaida, M., Yong, W., and Hamidah, M. (2010). Conservation of some rare and endangered plants from peninsular Malaysia. *Kew Bulletin*, 65(4), 681–689.
- Saw, L. and Chung, R. (2015). The flora of Malaysia projects. *Rodriguésia*, 66(4), 947–960.
- Saw, S. H. (1988). *The Population of Peninsular Malaysia*. NUS Press.
- United Nations Environment Programme, W. C. M. (2004). Known species of mammals, plants, and breeding birds. The United Nations Environment Program (UNEP) and the World Conservation Management Centre (WCMC).
- Von Rintelen, K., Arida, E., and Häuser, C. (2017). A review of biodiversity-related issues and challenges in megadiverse Indonesia and other southeast Asian countries. *Research Ideas and Outcomes*, 3, e20860.
- Wäldchen, J. and Mäder, P. (2018). Machine learning for image based species identification. *Methods in*

Ecology and Evolution, 9(11), 2216–2225.

WCSP (2020). World checklist of selected plant families. facilitated by the royal botanic gardens, Kew.

In Published in: <http://wcsp.science.kew.org/> (accessed 29 July 2020).



Chula
Chulalongkorn University



UNIVERSITÉ
DE MONTPELLIER



Institut de Recherche
pour le Développement
FRANCE

On the Edge of the Sixth Mass Extinction in Biodiversity Hotspots:

*Facts, needs, solutions and opportunities
in Thailand and adjacent countries*

Editors:

Julien Claude and Noppadon Kitana

Book title:
On the edge of the sixth mass extinction in biodiversity hotspots:
Facts, needs, solutions and opportunities in Thailand and adjacent
countries

Editors:
Julien Claude and Noppadon Kitana

Typesetting by:
Kasidit Rison and Rachata Maneein

ISBN xxx-xxxx

September 2024

@2024 Chulalongkorn University

This book is published under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0). This license requires that reusers give credit to the creator. It allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, for noncommercial purposes only. If others modify or adapt the material, they must license the modified material under identical terms.

For more information about our licenses, please visit
<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Published by:



Center of Learning Network for the Region (CLNR),
Chulalongkorn University, Phayathai Road, Bangkok 10330, Thailand

Printed at CU Press, Bangkok, Thailand