Supplementary

Supplementary 1. TaiBIF Dataset and EnjoyMoths

TaiBIF (http://taibif.tw/en) [1] is the Taiwan portal of GBIF which was specifically designed to broaden the GBIF network and increase the availability of local biodiversity data. Therefore, it is frequently synchronized with GBIF and includes much of the data available in GBIF (http://biota.taibif.tw/) Information available through TaiBIF [2]. (http://www.taibif.tw/en/node/2623) [1] includes: lists of local, endemic, exotic and invasive species; lists of terrestrial, aquatic and marine species; images of species; lists of experts; literature about local biodiversity topics, in both English and Mandarin; information and data about geographical and environmental features; synopses of participating or relevant institutions, organizations, projects, and locations; as well as publications available to the public [3]. Distributed Generic Information Retrieval (DiGIR) with the Darwin Core default can be used data exchange [3].

EnjoyMoths, is a citizen science project which includes 1,877 participants [3,4]. As the name implies, its main purpose is to appreciate moths, identify them, and locate them. The Facebook group was set up and is maintained by the Endemic Species Research Institute, Council of Agriculture, Taiwan [3,4]. While uploading observational data onto the Facebook page is easy, it is often difficult to transform volunteered observations, normally in the form of images and short descriptions, into scientifically useful structured data. For privacy and security reasons, the Exchangeable Image Format (EXIF) geolocation data is deleted from images uploaded to Facebook so that most images on Facebook are simply graphical records, i.e. explicit spatiotemporal data is not available for uploaded images. As such, a user generated photo description is the major source of information about the species therein [3,4]. For the EnjoyMoths project, a Natural Language Processing (NLP) method specifically designed to extract the Chinese Mandarin names of species and places from the photo description text was used to obtain scientific data.

While the Mandarin-specific NLP tools used in the EnjoyMoth project are suitable for generic functions, a number of things about Mandarin names of species and certain colloquialisms must be mentioned here for a full understanding of how the tools work. With the NLP protocols (see detail in Lin et al. [3]) and Mandarin naming conventions, matches, or partial matches, are identified within a Facebook message for some of the prefixes and species-specific names. The NLP protocol then checks for corresponding postfix matches (genus or family names). Finally, the NLP calculates confidence scores based on the length of the names and chooses scientific names with the highest scores (Eq. A.1) [3,4].

The confidence scores are ultimately a comparison of the shortened colloquial name lengths identified in the text and the potential full scientific name lengths, as given in the following equation [3,4].

$$\frac{L_{prefix} + L_{postfix}}{L_{full}}$$
(A1)

where L_{prefix} and $L_{postfix}$ are the lengths of the prefixes and postfixes, and L_{full} is the complete length of the scientific species name. While name length comparisons may seem slightly arbitrary, the method is quite effective since longer scientific names generally correspond to longer colloquial names.

In some circumstances, participants post images without text description or any mention of the species name. In these cases, other participants often attempt to identify the species and leave reply messages as post comments. When this occurs, three criteria are referenced in order to resolve the identity of a species [3,4].

- Species name frequencies throughout the post comment those with higher frequencies garner higher scores.
- Likes count of the original post or subsequent comments those with more likes garner higher scores.
- Participant prestige of commenter those with higher prestige, determined by past identification fidelity, are given higher scores.

Spatial references in Facebook posts can take the form of either specific coordinates or place names. The first step in extracting this data is simply a string matching procedure. Messages with text that correspond to a place name in the NLP database, which includes common colloquial place names, are registered as a place name. The newly registered place name is then translated into coordinate and administrative spatial information according to the Taiwan Geographic Names database [3,4], e.g., township or county.

The F-score, which treats both precision and recall with equal importance, is calculated in order to evaluate the fidelity of place and species names with the following equations [3–5]:

$$recall(name_S_{predict}) = \frac{\left|name_S_{actual} \cap name_S_{predict}\right|}{\left|name_S_{actual}\right|}$$
(A.2)

$$precision(name_S_{nlp}) = \frac{\left| name_S_{actual} \cap name_S_{predict} \right|}{\left| name_S_{predict} \right|}$$
(A.3)

$$F(name_S_{predict}) = \frac{2 \cdot precision \ (name_S_{predict}) \cdot recall \ (name_S_{predict})}{precision \ (name_S_{predict}) + recall \ (name_S_{predict})}$$
(A.4)

when name_S_{actual} is the set of species or place names that have been validated and verified by experts; and name_S_{predict} is the set of names predicted by the NLP.

Supplementary 2. Species Information

species chosen from *EnjoyMoths* (TaiBIF, Nine were as focal species of http://www.taibif.org.tw/; The National Museum Natural Science, http://www.nmns.edu.tw/index_eng.html; Surprise Mountain Line, http://gaga.biodiv.tw/9701bx/in94.htm [1,6,7]). The F-scores of each of the nine species are all greater than 0.90, demonstrating a high level of naming precision [3].

A. egens indica inhabits low altitude areas of Taiwan. It is a widespread species and can be found from Japan to New Guinea. It is a large moth with a wingspan of only 5.4 to 6 cm. Its body and wings are shades of white to orange, while the anterior wings are dark yellow to orange with clear white veins fanning out and dark spots observed near the bottom. In contrast, the posterior wings are either white or light yellow at the bottom and varying from yellow to orange at the ends. The larvae often subsist on *Ficus retusa* plants (Taiwan National Museum of Natural Science, http://www.nmns.edu.tw/index_eng.html [8]).

A. heliconia zebrine is a member of the *Aganainae* subfamily of the *Noctuidae* family. In Taiwan, *A. heliconia zebrina* has a distribution ranging from 0 to 2000 meters above sea level. Its anterior wings are predominantly grey, having a single white line starting at the base and running to the midsection of the wing. The posterior wing is predominantly white, having broad grey edges. The larvae subsist on Moraceae and Ficus plants while adults consume rotten fruit, the nectar of flowers, and sap (TaiBIF, http://www.taibif.org.tw/; Taiwan National Museum of Natural Science, http://www.nmns.edu.tw/index_eng.html; Surprise Mountain Line, http://gaga.biodiv.tw/9701bx/in94.htmAppendix-Enjoymoth.docx [1,7,8]).

B. perclarus are normally found in low- to high-altitude areas throughout Taiwan, and are widespread throughout the island. The anterior wings are 50 to 65 mm long and situated at a

vertical angle with the body. Anterior and posterior wings are similar in color, white or yellowish white with two dark stripes. The stripes run perpendicular to the length of the wings symmetrically, almost resembling a ring. The inner ring is predominantly situated on the anterior wing with a small stripe across the base of the posterior wings and abdomen. Whereas the outer ring outlines the periphery of both wings (Taiwan National Museum of Natural Science, http://www.nmns.edu.tw/index_eng.html; The Digital Museum of Nature & Culture, http://digimuse.nmns.edu.tw; TaiBIF, http://www.taibif.org.tw/; Surprise Mountain Line, http://gaga.biodiv.tw/9701bx/in94.htm [1,7–9]).

C. magnifica are normally found in low to mid-altitude areas within Taiwan (National Museum of Natural Science, http://www.nmns.edu.tw/index eng.html; The Digital Museum of Nature & Culture, http://digimuse.nmns.edu.tw [6,9]). Its wings, head, and thorax are predominantly black with orange spots. The larvae of C. magnifica (Walker) consume bryophyte plants and adults feed on rotten fruit, sap, and the nectar of flowers (TaiBIF, http://www.taibif.org.tw/; National of Taiwan Museum Natural Science, http://www.nmns.edu.tw/index_eng.html; Surprise Mountain Line, http://gaga.biodiv.tw/9701bx/in94.htmAppendix-Enjoymoth.docx [1,7,8]). They are active throughout the year, primarily at night, although they are sometimes active during the day.

Н. flabellicornis ultima inhabit low-altitude areas of Taiwan (TaiBIF, http://www.taibif.org.tw/; National Taiwan Museum of Natural Science, http://www.nmns.edu.tw/index eng.html; The Digital Museum of Nature & Culture, http://digimuse.nmns.edu.tw [1,8,9]). The larvae consume Bischofia javanica. The anterior wings are black, while part of the posterior wings are metal blue (TaiBIF, http://taibif.tw/zh/namecode/348116; Taiwan National Museum of Natural Science, http://www.nmns.edu.tw/index eng.html; The Digital Museum of Nature & Culture, http://digimuse.nmns.edu.tw [1,8,9]). H. flabellicornis ultima are diurnal, and active in spring and autumn (The Digital Museum of Nature & Culture http://digimuse.nmns.edu.tw [9]).

L. nobilis inhabit low to mid-altitude areas of Taiwan. They are large moths with a wingspan of 90 to 140mm. The head is light brown, while the wings have intricate banding patterns – light to dark brown bands running across them, and thin pale stripes running along the length of them. Larvae feed on a number of plant species including Pinus, Rubus, and Pteridium (TaiBIF, http://www.taibif.org.tw/; Taiwan National Museum of Natural Science, http://www.nmns.edu.tw/index_eng.html; The Digital Museum of Nature & Culture, http://digimuse.nmns.edu.tw; Surprise Mountain Line, http://gaga.biodiv.tw/9701bx/in94.htm [1,7–9]).

S. litura can be found all across Taiwan in both low and high altitude areas. Larvae feed on some economically important plants, including grains, flowers, vegetables, and fruit trees (TaiBIF, http://www.taibif.org.tw/; Taiwan National Museum of Natural Science, http://www.nmns.edu.tw/index_eng.html; The Digital Museum of Nature & Culture, http://digimuse.nmns.edu.tw [1,8,9]). It has a wingspan of 40 mm. The anterior wings are white overall, with brown edgings and two grey lines in the center (Taiwan National Museum of Nature & Culture, http://www.nmns.edu.tw/index_eng.html; The Digital Museum of National Museum of Nature & Culture, http://www.nmns.edu.tw/index_eng.html; The Digital Museum of National Museum of Nature & Culture, http://www.nmns.edu.tw/index_eng.html; The Digital Museum of Nature & Culture, http://digimuse.nmns.edu.tw [8,9]).

T. aventiaria masses on low to mid-altitude areas in Taiwan. The anterior wings are green in color, with two brown spots and same-colored edging on the front, run through by darker shaded green lines (TaiBIF, http://www.taibif.org.tw/ [1]). There is also a yellow-green colored straight line in the middle of the wings (TaiBIF, http://www.taibif.org.tw/ [1]). Larvae subsist on the Rosaceae and the Leguminosae plant families.



Figure S1. Objective function values or geometric mean of likelihood of all included samples, demonstrated (Y axis) versus the iterations (X axis) at a 0.3 cooling rate for (**a**) *Asota egens indica*, (**b**) *Asota heliconia zebrine*, (**c**) *Biston perclarus*, (**d**) *Chrysaeglia magnifica*, (**e**) *Histia flabellicornis ultima*, (**f**) *Hyposidra talaca*, (**g**) *Lebeda nobilis*, (**h**) *Spodoptera litura*, and (**i**) *Traminda aventiaria*.



Figure S2. Objective function values or geometric mean of likelihood of all included samples, demonstrated (Y axis) versus the iterations (X axis) at a 0.4 cooling rate for (**a**) *Asota egens indica*, (**b**) *Asota heliconia zebrine*, (**c**) *Biston perclarus*, (**d**) *Chrysaeglia magnifica*, (**e**) *Histia flabellicornis ultima*, (**f**) *Hyposidra talaca*, (**g**) *Lebeda nobilis*, (**h**) *Spodoptera litura*, and (**i**) *Traminda aventiaria*.









Figure S3. Observation locations in FB red (left), GBIF blue (center) and GFIB + FB_o red/blue (right); and habitat suitability distributions based on the above-mentioned datasets for each species. Note: Cooling rate is 0.3; Professionally collected data from Global Biodiversity Information Facility (GBIF); opportunistic data from Facebook (FB); opportunistic data from optimally selected Facebook dataset (FB_o).







 $0 \quad 0.125 \ 0.25 \ 0.375 \ 0.5 \ 0.625 \ 0.75 \ 0.875 \ 1$

Figure S4. Observation locations in FB red (left), GBIF blue (center) and GFIB + FB_o red/blue (right); and habitat suitability distributions based on the above-mentioned datasets for each species. Note: Cooling rate is 0.4; Professionally collected data from Global Biodiversity Information Facility (GBIF); opportunistic data from Facebook (FB); opportunistic data from optimally selected Facebook dataset (FB_o).





Figure S5. Boxplots of the AUC values derived from four models (GAM, GLM, Maxent, and SVM) of five datasets (GBIF, FB, GBIF+FB, GBIF + FB_o, and GBIG + FB_r) for nine species. Note: Cooling rate is 0.3; General Additive Model (GAM); General Linear Model (GLM); Maximum Entropy Modeling (Maxent);

and Support Vector Machine (SVM); Professionally collected data from Global Biodiversity Information Facility (GBIF); Opportunistic dataset from Facebook (FB); Professionally collected data from Global Biodiversity Information Facility plus opportunistic data from Facebook (GBIF + FB); Professionally collected data from Global Biodiversity Information Facility plus opportunistic data from optimally selected Facebook dataset (GBIF + FB_o); Professionally collected data from Global Biodiversity Information Facility plus opportunistic data from Facebook dataset (GBIF + FB_o); Professionally collected data from Global Biodiversity Information Facility plus opportunistic data from randomly selected Facebook dataset (GBIF + FB_r).





Figure S6. Boxplots of the AUC values derived from four models (GAM, GLM, Maxent, and SVM) of five datasets (GBIF, FB, GBIF + FB, GBIF + FB_o, and GBIG + FB_r) for nine species. Note: Cooling rate is 0.4; General Additive Model (GAM); General Linear Model (GLM); Maximum Entropy Modeling (Maxent);

and Support Vector Machine (SVM); Professionally collected data from Global Biodiversity Information Facility (GBIF); Opportunistic dataset from Facebook (FB); Professionally collected data from Global Biodiversity Information Facility plus opportunistic data from Facebook (GBIF + FB); Professionally collected data from Global Biodiversity Information Facility plus opportunistic data from optimally selected Facebook dataset (GBIF + FB_o); Professionally collected data from Global Biodiversity Information Facility plus opportunistic data from Facebook dataset (GBIF + FB_o); Professionally collected data from Global Biodiversity Information Facility plus opportunistic data from randomly selected Facebook dataset (GBIF + FB_r).

	Data source	GBIF	FB	FB_o	p value	KW pair test
	sp 1	0.68	0.63	1	< 0.001	GBIF-FB_o vs FB-FB_o
	sp 2	0.68	0.74	0.93	< 0.01	GBIF-FB_o
	sp 3	0.86	0.72	1	< 0.001	FB-FB_o
	sp 4	0.82	0.55	1	< 0.001	FB-FB_o
Forest	sp 5	0.69	0.42	1	< 0.001	GBIF-FB vs FB-FB_o
	sp 6	0.56	0.74	1	< 0.001	GBIF-FB_o
	sp 7	0.81	0.76	0.8	0.85	
	sp 8	0.61	0.49	0.9	< 0.01	FB- FB_o
	sp 9	0.82	0.61	0.75	0.06	
	sp 1	0.29	0.25	0.07	< 0.001	GBIF-FB_o vs FB-FB_o
WTRSHD	sp 2	0.27	0.14	0.07	< 0.001	GBIF-FB vs GBIF-FB_o
	sp 3	0.32	0.21	0.11	< 0.001	GBIF-FB vs GBIF-FB_o
	sp 4	0.24	0.19	0.06	< 0.01	GBIF-FB_o
	sp 5	0.39	0.25	0.16	< 0.001	GBIF-FB vs GBIF-FB_o
	sp 6	0.36	0.18	0.1	< 0.001	GBIF-FB vs GBIF-FB_o
	sp 7	0.22	0.18	0.05	< 0.01	GBIF-C vs FB-FB_0
	sp 8	0.43	0.19	0.09	< 0.001	GBIF-FB vs GBIF-FB_0
	sp 9	0.24	0.18	0.05	< 0.001	GBIF-FB_o vs FB-FB_o
	sp 1	0.28	0.12	0.05	< 0.001	GBIF-FB vs GBIF-FB_o vs FB-FB_o
	sp 2	0.21	0.08	0.05	< 0.001	GBIF-FB vs GBIF-FB_o
	sp 3	0.28	0.12	0.09	< 0.001	GBIF-FB vs GBIF-FB_o
	sp 4	0.25	0.09	0.03	< 0.001	GBIF-FB vs GBIF-FB_0
DI CITV	sp 5	0.24	0.08	0.1	< 0.001	GBIF-FB vs GBIF-FB_0
D2_CITT	sp 6	0.26	0.1	0.08	< 0.001	GBIF-FB vs GBIF-FB_0
	sp 7	0.25	0.11	0.05	< 0.001	GBIF-FB vs GBIF-FB_o
	sp 8	0.22	0.1	0.07	<0.01	GBIF-FB vs GBIF-FB_o
	sp 9	0.25	0.09	0.03	< 0.001	GBIF-FB vs GBIF-FB_o
	sp 1	0.12	0.11	0.08	0.36	
	sp 2	0.09	0.12	0.09	0.06	
	sp 3	0.12	0.11	0.07	0.16	
D2_ROAD	sp 4	0.13	0.1	0.09	0.71	
1	sp 5	0.08	0.09	0.08	0.93	
	sp 6	0.1	0.14	0.07	0.1	
	sp 7	0.15	0.14	0.08	0.44	

Table S1. The average values of 14 environmental variables and habitat suitability indices in three data sources, GBIF, FB, and FB_o for nine species; the Kruskal–Wallis (KW) test p value among the three datasets and KW pair tests between every two datasets; cooling rate is 0.5.

	sp 8	0.09	0.11	0.07	0.54	
	sp 9	0.13	0.11	0.08	0.44	
	sp 1	0.18	0.11	0.07	0.3	
	sp 2	0.15	0.1	0.08	0.59	
	sp 3	0.18	0.12	0.1	0.21	
	sp 4	0.17	0.08	0.06	< 0.01	GBIF-FB vs GBIF-FB_0
ELEV	sp 5	0.06	0.06	0.04	0.45	
	sp 6	0.11	0.11	0.08	0.17	
	sp 7	0.23	0.12	0.11	< 0.05	GBIF-FB
	sp 8	0.08	0.08	0.06	0.53	
	sp 9	0.17	0.1	0.08	0.07	
	sp 1	0.53	0.48	0.61	< 0.01	FB-FB_o
	sp 2	0.44	0.48	0.52	0.12	
	sp 3	0.53	0.46	0.61	< 0.001	FB-FB_o
	sp 4	0.5	0.37	0.42	< 0.05	GBIF-FB
HUMIDIT Y	sp 5	0.46	0.3	0.48	< 0.001	GBIF-FB
	sp 6	0.43	0.47	0.51	0.27	
	sp 7	0.5	0.47	0.42	0.54	
	sp 8	0.34	0.42	0.52	< 0.05	GBIF-FB_o
	sp 9	0.5	0.43	0.41	0.17	
	sp 1	0.23	0.29	0.38	< 0.001	GBIF-FB_o vs FB-FB_o
	sp 2	0.23	0.34	0.37	< 0.001	GBIF-FB vs GBIF-FB_o
	sp 3	0.21	0.28	0.31	< 0.01	GBIF-FB vs GBIF-FB_o
	sp 4	0.26	0.3	0.32	0.19	
P_PC1	sp 5	0.18	0.27	0.3	< 0.01	GBIF-FB vs GBIF-FB_o
	sp 6	0.2	0.31	0.33	< 0.001	GBIF-FB vs GBIF-FB_o
	sp 7	0.29	0.31	0.39	< 0.01	GBIF-FB_o vs FB-FB_o
	sp 8	0.16	0.24	0.27	< 0.001	GBIF-FB vs GBIF-FB_o
	sp 9	0.26	0.29	0.39	< 0.001	GBIF-FB_o vs FB-FB_o
	sp 1	0.74	0.71	0.6	< 0.001	GBIF-FB_o vs FB-FB_o
	sp 2	0.73	0.66	0.61	< 0.001	GBIF-FB vs GBIF-FB_o
	sp 3	0.79	0.72	0.72	0.05	
	sp 4	0.72	0.68	0.65	0.18	
P_PC2	sp 5	0.74	0.69	0.69	0.06	
	sp 6	0.75	0.69	0.68	0.07	
	sp 7	0.7	0.69	0.59	< 0.01	GBIF-FB_o vs FB-FB_o
	sp 8	0.77	0.72	0.7	< 0.05	GBIF-FB_o
	sp 9	0.72	0.69	0.58	< 0.001	GBIF-FB_o vs FB-FB_o

	sp 1	0.49	0.42	0.4	< 0.01	GBIF-FB vs GBIF-FB_o
	sp 2	0.45	0.38	0.37	< 0.001	GBIF-FB vs GBIF-FB_o
	sp 3	0.41	0.4	0.36	< 0.05	FB-FB_o
	sp 4	0.49	0.39	0.37	< 0.001	GBIF-FB vs GBIF-FB_o
P_PC3	sp 5	0.49	0.4	0.39	< 0.001	GBIF-FB vs GBIF-FB_o
	sp 6	0.51	0.37	0.35	< 0.001	GBIF-FB vs GBIF-FB_o
	sp 7	0.49	0.38	0.41	< 0.01	GBIF-FB
	sp 8	0.5	0.39	0.39	< 0.001	GBIF-FB vs GBIF-FB_o
	sp 9	0.49	0.39	0.38	< 0.001	GBIF-FB vs GBIF-FB_o
	sp 1	0.39	0.41	0.39	< 0.05	
	sp 2	0.4	0.41	0.41	< 0.05	GBIF-FB
	sp 3	0.37	0.39	0.36	< 0.01	FB-FB_o
	sp 4	0.39	0.42	0.4	< 0.05	GBIF-FB
P_PC4	sp 5	0.42	0.42	0.4	0.24	
	sp 6	0.42	0.39	0.39	0.83	
	sp 7	0.37	0.39	0.39	0.21	
	sp 8	0.4	0.42	0.39	0.11	
	sp 9	0.39	0.4	0.41	0.05	
	sp 1	0.64	0.73	0.72	0.08	
	sp 2	0.68	0.72	0.71	0.51	
	sp 3	0.69	0.71	0.71	0.74	
	sp 4	0.65	0.75	0.74	< 0.05	GBIF-FB
T_PC1	sp 5	0.79	0.78	0.77	0.13	
	sp 6	0.72	0.71	0.73	< 0.05	GBIF-FB
	sp 7	0.59	0.71	0.66	< 0.05	
	sp 8	0.81	0.74	0.76	< 0.001	GBIF-FB vs GBIF-FB_o
	sp 9	0.65	0.73	0.7	0.1	
	sp 1	0.55	0.54	0.59	< 0.05	FB-FB_o
	sp 2	0.56	0.58	0.6	< 0.05	GBIF-FB_o
	sp 3	0.54	0.56	0.56	0.31	
	sp 4	0.6	0.56	0.6	0.16	
T_PC2	sp 5	0.45	0.52	0.54	< 0.001	GBIF-FB vs GBIF-FB_0
	sp 6	0.47	0.57	0.57	< 0.001	GBIF-FB vs GBIF-FB_0
	sp 7	0.64	0.58	0.66	< 0.01	FB-FB_o
	sp 8	0.46	0.55	0.55	< 0.001	GBIF-FB vs GBIF-FB_0
	sp 9	0.6	0.56	0.62	< 0.05	FB-FB_o
	sp 1	0.62	0.6	0.68	< 0.001	GBIF-FB_o vs FB-FB_o
T_PC3	sp 2	0.6	0.63	0.67	< 0.001	GBIF-FB vs GBIF-FB_o
	sp 3	0.6	0.61	0.67	< 0.05	FB-FB_o

	sp 4	0.61	0.62	0.67	< 0.05	GBIF-FB_o
	sp 5	0.61	0.62	0.67	0.24	
	sp 6	0.62	0.62	0.68	< 0.05	GBIF-FB_o
	sp 7	0.62	0.62	0.69	0.08	
	sp 8	0.63	0.61	0.69	< 0.05	FB-FB_o
	sp 9	0.61	0.63	0.67	0.1	
	sp 1	0.77	0.71	0.71	< 0.01	GBIF-FB vs GBIF-FB_0
	sp 2	0.75	0.72	0.75	0.16	GBIF-FB vs GBIF-FB_o
	sp 3	0.83	0.75	0.75	< 0.001	GBIF-FB
	sp 4	0.78	0.71	0.73	< 0.01	GBIF-FB
NDVI	sp 5	0.77	0.65	0.73	< 0.001	
	sp 6	0.71	0.75	0.71	0.35	
	sp 7	0.78	0.74	0.77	0.33	
	sp 8	0.7	0.68	0.7	0.96	
	sp 9	0.78	0.74	0.78	0.18	GBIF-FB vs GBIF-FB_0
	sp 1	0.51	0.35	0.52	< 0.001	GBIF-FB vs FB-FB_o
	sp 2	0.51	0.52	0.64	< 0.001	GBIF-FB_o vs FB-FB_o
	sp 3	0.53	0.36	0.61	< 0.001	GBIF-FB vs FB-FB_o
	sp 4	0.53	0.38	0.61	< 0.001	GBIF-FB vs FB-FB_o
HSI	sp 5	0.51	0.25	0.53	< 0.001	GBIF-FB vs FB-FB_o
	sp 6	0.5	0.34	0.45	< 0.001	GBIF-FB vs FB-FB_o
	sp 7	0.52	0.39	0.64	< 0.001	FB-FB_o
	sp 8	0.52	0.29	0.52	< 0.001	GBIF-FB vs FB-FB_o
	sp 9	0.53	0.39	0.56	< 0.001	GBIF-FB vs FB-FB_o

Note: Asota egens indica (sp1); Asota heliconia zebrine (sp2); Biston perclarus (sp3); Chrysaeglia magnifica (sp4); Histia flabellicornis ultima (sp5); Hyposidra talaca (sp6); Lebeda nobilis (sp7); Spodoptera litura (sp8); Traminda aventiaria (sp9); Professionally collected data from Global Biodiversity Information Facility (GBIF); Opportunistic dataset from Facebook (FB); Professionally collected data from Global Biodiversity Information Facility plus opportunistic data from optimally selected Facebook dataset (GBIF + FB_o).

Species	Model	GBIF	FB	GBIF + FB	GBIF + FB_o	GBIF + FB_r
	GAM	0.89	0.98	0.97	0.98	0.92
Asota egens i	GLM	0.51	0.87	0.89	0.84	0.79
ndica	Maxent	0.31	0.78	0.79	0.65	0.61
	SVM	0.68	0.83	0.81	0.87	0.74
	GAM	0.95	0.98	0.98	0.98	0.97
Asota heliconia	GLM	0.80	0.81	0.88	0.88	0.84
zebrina	Maxent	0.62	0.60	0.79	0.74	0.72
	SVM	0.79	0.86	0.83	0.85	0.81
	GAM	0.90	0.98	0.98	0.97	0.92
Rister and sure	GLM	0.42	0.85	0.87	0.74	0.71
bision perciurus	Maxent	0.27	0.70	0.72	0.47	0.46
	SVM	0.71	0.83	0.81	0.84	0.74
	GAM	0.91	0.95	0.96	0.95	0.90
Chrysaeglia magnifica	GLM	0.54	0.67	0.77	0.71	0.65
	Maxent	0.30	0.41	0.57	0.42	0.37
	SVM	0.69	0.79	0.77	0.82	0.70
Histia flabellicornis ultima	GAM	0.96	0.95	0.97	0.97	0.95
	GLM	0.72	0.71	0.86	0.76	0.74
	Maxent	0.50	0.45	0.69	0.54	0.49
	SVM	0.83	0.71	0.81	0.86	0.82
	GAM	0.88	0.95	0.95	0.96	0.88
	GLM	0.46	0.74	0.79	0.71	0.63
1199051010 101000	Maxent	0.32	0.47	0.63	0.48	0.43
	SVM	0.64	0.83	0.80	0.80	0.73
	GAM	0.88	0.96	0.96	0.96	0.87
Labada nobilis	GLM	0.28	0.71	0.80	0.62	0.61
Lebeuu noonis	Maxent	0.25	0.46	0.55	0.38	0.35
	SVM	0.68	0.82	0.81	0.83	0.73
	GAM	0.91	0.96	0.95	0.95	0.88
Cnodontara litura	GLM	0.34	0.70	0.72	0.58	0.53
Spoaoptera litura	Maxent	0.20	0.41	0.51	0.34	0.29
	SVM	0.72	0.76	0.72	0.75	0.62
Traminda	GAM	0.91	0.96	0.97	0.96	0.92
aventiaria	GLM	0.56	0.69	0.81	0.72	0.69

Table S2. Explained variation by the first PCA component of the 1000 species distributions derived from four SDMs based on five datasets at a 0.3 cooling rate.

	Maxent	0.31	0.39	0.60	0.44	0.40
	SVM	0.7	0.79	0.77	0.81	0.72
Averag	e	0.62	0.76	0.81	0.75	0.70

Note: Cooling rate is 0.3; General Additive Model (GAM); General Linear Model (GLM); Maximum Entropy Modeling (Maxent); and Support Vector Machine (SVM); Professionally collected data from Global Biodiversity Information Facility (GBIF); Opportunistic dataset from Facebook (FB); Professionally collected data from Global Biodiversity Information Facility plus opportunistic data from Facebook (GBIF + FB); Professionally collected data from Global Biodiversity Information Facility Information Facility plus opportunistic data from Global Biodiversity Information Facility plus opportunistic data from Facility Plus

Species	Model	GBIF	FB	GBIF + FB	GBIF + FB_o	GBIF + FB_r
	GAM	0.90	0.98	0.98	0.97	0.91
Asota egens	GLM	0.53	0.86	0.89	0.84	0.75
indica	Maxent	0.31	0.77	0.79	0.62	0.57
	SVM	0.68	0.82	0.82	0.87	0.72
	GAM	0.95	0.98	0.98	0.98	0.97
Asota heliconia	GLM	0.79	0.81	0.88	0.88	0.85
zebrina	Maxent	0.62	0.59	0.78	0.77	0.73
	SVM	0.80	0.85	0.83	0.85	0.81
	GAM	0.89	0.98	0.98	0.97	0.91
Biston norclarus	GLM	0.43	0.85	0.87	0.71	0.70
Dision perciurus	Maxent	0.28	0.70	0.73	0.43	0.41
	SVM	0.72	0.82	0.81	0.83	0.73
	GAM	0.92	0.96	0.96	0.96	0.91
Chrysaeglia	GLM	0.54	0.68	0.78	0.70	0.65
magnifica	Maxent	0.29	0.42	0.58	0.43	0.37
	SVM	0.69	0.78	0.78	0.81	0.72
Histia	GAM	0.95	0.95	0.97	0.97	0.95
	GLM	0.69	0.67	0.86	0.74	0.73
ultima	Maxent	0.48	0.45	0.70	0.53	0.48
шити	SVM	0.83	0.70	0.81	0.86	0.81
	GAM	0.88	0.95	0.95	0.95	0.88
Humosidra talaca	GLM	0.46	0.72	0.78	0.71	0.66
1199051010 101000	Maxent	0.34	0.47	0.62	0.49	0.46
	SVM	0.66	0.82	0.80	0.81	0.73
	GAM	0.87	0.96	0.95	0.96	0.87
Lahada nohilis	GLM	0.29	0.71	0.79	0.67	0.62
Lebeaa nobiiis	Maxent	0.24	0.45	0.55	0.41	0.32
	SVM	0.66	0.81	0.80	0.81	0.71
	GAM	0.91	0.95	0.95	0.95	0.88
Snodontara litura	GLM	0.38	0.70	0.74	0.59	0.47
Spoaoptera litura	Maxent	0.21	0.43	0.53	0.36	0.28
	SVM	0.75	0.73	0.73	0.75	0.63
Traminda	GAM	0.91	0.96	0.97	0.96	0.91
aventiaria	GLM	0.54	0.67	0.79	0.73	0.68

Table S3. Explained variation by the first PCA component of the 1000 species distributions derived from four SDMs based on five datasets at a 0.4 cooling rate.

	Maxent	0.30	0.40	0.59	0.45	0.41
	SVM	0.69	0.78	0.77	0.82	0.71
Averag	ge	0.62	0.75	0.81	0.75	0.69

Note: Cooling rate is 0.4; General Additive Model (GAM); General Linear Model (GLM); Maximum Entropy Modeling (Maxent); and Support Vector Machine (SVM); Professionally collected data from Global Biodiversity Information Facility (GBIF); Opportunistic dataset from Facebook (FB); Professionally collected data from Global Biodiversity Information Facility plus opportunistic data from Facebook (GBIF+FB); Professionally collected data from Global Biodiversity Information Facility plus opportunistic data from Facebook dataset (GBIF + FB_o); Professionally collected data from Global Biodiversity Information Facility plus opportunistic data from randomly selected Facebook dataset (GBIF + FB r).

References

- 1. TaiBIF. Avalilable online: <u>http://taibif.tw/en</u>.
- 2. GBIF. Avalilable online: <u>http://biota.taibif.tw/</u>
- Lin, Y.-P.; Deng, D.; Lin, W.-C.; Lemmens, R.; Crossm, N.D.; Henle, K.; Schmeller, D.S. Uncertainty analysis of crowd-sourced and professionally collected field data used in species distribution models of taiwanese moths. *Biological Conservation* 2015, 181, 102-110; doi: org/10.1016/j.biocon.2014.11.012.
- Deng, D.-P.; Chuang, T.-R.; Shao, K.-T.; Mai, G.-S.; Lin, T.-E.; Lemmens, R.; Hsu, C.-H.; Lin, H.-H.; Kraak, M.-J. Using social media for collaborative species identification and occurrence: Issues, methods, and tools. *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information* 2012, 22-29; doi: 10.1145/2442952.2442957.
- 5. Thessen, A.E.; Cui, H.; Mozzherin, D. Applications of natural language processing in biodiversity science. *Advances in Bioinformatics* **2012**, 2012, 1-17; doi: org/10.1155/2012/391574.
- 6. National museum of natural science. Avalilable online: http://www.nmns.edu.tw/index_eng.html
- 7. Surprise mountain line. Avalilable online: <u>http://gaga.biodiv.tw/9701bx/in94.htmAppendix-Enjoymoth.docx</u>
- 8. Taiwan national museum of natural science. Avalilable online: <u>http://www.nmns.edu.tw/index_eng.html</u>
- 9. The digital museum of nature & culture. Available online: <u>http://digimuse.nmns.edu.tw</u>