RESEARCH ARTICLE

Botanical dispersal pathways into Malesia: The Philippines – Southeast Asian connection

P.C. van Welzen[1,2,*], S.J. Hilgen[1], J.M.A. Klink[1], R.-Y. Yu[1], M. van 't Zelfde[1,3]

Key words

dated phylogenies
migration routes
Philippines
plants
SE Asia mainland
species distribution modelling

Abstract   The Malay Archipelago, also known as Malesia, the triangle Malay Peninsula–Philippines–New Guinea, encompassing Indonesia, is tectonically one of the world's most active areas with an intricate history of amalgamation of terranes. For plants it means that ancestral species must have dispersed to the area, which presently holds about 70 % of endemic species with an estimated total of up to 45 000 species. One of the possible dispersal routes mentioned runs from the SE Asian mainland via the Philippines. This was considered likely for montane species and for species that prefer a yearly dry monsoon period. However, comparisons of numbers of species, dated phylogenies with a clade on the mainland and another in the Philippines and Species Distribution Modelling all show that most likely only few species dispersed via this northern route. Chance long distance dispersal, especially for montane species, is always possible, but for the dry monsoon preferring species, for which the Philippine connection was mentioned once as a pathway, the savannah corridor during glacial periods, running from the Malay Peninsula to Java, was likely a far more important dispersal route.

INTRODUCTION

Malesia, also indicated as the Malay Archipelago, was defined as phytogeographical area by Van Steenis (1950; see also Raes & Van Welzen 2009 for earlier definitions) based on demarcations in generic distributions at the Thai–Malay border, between Taiwan and the Philippines, and between New Guinea and Australia. It is roughly formed by the triangle Malay Peninsula–Philippines–New Guinea and encompasses Indonesia as largest country in the region. Van Welzen et al. (2005) showed in a different way that Malesia is a well-described phytogeographical area as in their sample of 6 066 indigenous species (out of an estimated 42 000 (Roos 1992) to 45 000 (Corlett & Primack 2011) species), 70 % were endemic (Van Welzen et al. 2005).

The geological history of the region is extremely complicated (Hall 2009, 2012). Most of Malesia (and parts of the SE Asian mainland) broke off from the Australian part of Gondwana as small terranes and arrived in various waves at their present position. By the late Cretaceous W Malesia was more or less in its present position and c. 45 Ma Australia, New Guinea, and the east Malesian terranes started to move north. The Philippines is a conglomerate of various areas. Part of it originated in situ, more or less SSE of the present position, due to submarine volcanic activity that created islands (c. 50–45 Ma), in this case at the edge of the Philippine Sea Plate. Palawan, and a few islands north and south of it, detached from the Chinese part

of Laurasia and started to rift eastwards to the Philippines at c. 32 Ma. Due to the northern movement of Australia and New Guinea, W Malesia turned slightly anti-clockwise and the Philippines moved north to their present position, south of Taiwan (in all reconstructions Taiwan is Laurasian and was always in its present position). The latest addition to the Philippines were islands in the Philippine–Halmahera Island arc (Hall 2009, 2012), which formed 45–25 Ma. De Boer (1995) stretches this arc even further to the east, to terranes beyond New Guinea and named it the Outer Melanesian Arc.

Hall (2009) considered that most terranes rifted completely to mainly submersed and thus had limited possibilities to migrate floral and faunal elements. This means that most of the flora must have arrived by dispersal. There are various pathways by which plants could have reached Malesia. The most important one is from the SE Asian mainland via the Malay Peninsula (e.g., Crayn et al. 2014; Fig. 1 purple arrow), either from continental Asia (e.g., Atkins et al. 2019) or rafting from Africa via India (e.g., Morley 2000). The second most important is from the east, with Australia and/or New Guinea as source (e.g., Morley 2003, Crain et al. 2014; Fig. 1 blue arrow). A third one to be investigated is a dry savannah link between Australia and the Lesser Sunda Islands/S Moluccas (Van Steenis 1979; Fig. 1 white arrow), while the fourth is a link via the Philippines (Fig. 1 yellow arrow). Within Malesia, the Philippines are part of the Wallacean area (Van Welzen et al. 2011), and often harbour species that are part of a dry monsoon period distribution pattern (e.g., Van Steenis 1979: f. 13). In this pattern, the everwet Malay Peninsula, Sumatra, and Borneo are evaded (Van Steenis 1979: f. 4, 5) and plants show distributions from the SE Asian mainland to Wallacea (Java, Philippines, Sulawesi, Lesser Sunda Islands, and Mollucas) and even Australia. Van Steenis considered the connection SE Asian mainland–Philip-

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 Institute of Environmental Sciences, Leiden University, P.O. Box 9518, 2300 RA Leiden, The Netherlands
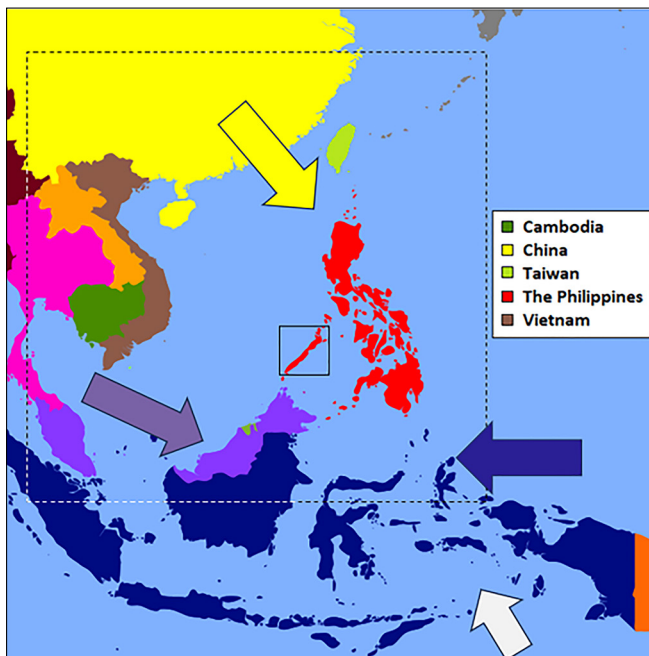* corresponging author e-mail: peter.vanwelzen@naturalis.nl

Fig. 1 Study area (0–30° latitude and 100–130° longitude) is outlined by the dashed line and consists of partly China, Taiwan, Vietnam, Cambodia, Phillippines. Palawan is located near the middle (solid black outline). Also indicated are the various (possible) pathways as discussed in the introduction (purple: SE Asia mainland via Malay Peninsula, blue: from NE Australia and New Guinea; white: Australia via S Moluccas and SE Lesser Sunda Islands; yellow: SE Asia mainland to Philippines).

pines important for dispersal of montane plants (Van Steenis 1964: f. 11) and those that prefer a dry monsoon (Van Steenis 1979: f. 4). The Philippines can be the first stepping stone into Malesia, either via rafting on Palawan, and/or dispersal via Taiwan. During the geologically recent glacial periods, with a drop of –120 m in sea levels, the connectivity between N Borneo and Palawan and NE Borneo and Mindanao increased (Morley & Flenley 1987, Hall 2013: f. 13), but also volcanic islands north of Luzon provided higher dispersal possibilities with Taiwan (Hall 2013: f. 13). One of the first to mention a potential organismal dispersal route to the Philippines from the north was Dickerson (1928). Animal evidence for the northern routes exists, for example shrew dispersal from Taiwan to Luzon via the islands of Batan and Sabtang (Esselstyn & Oliveros 2010). However, records of plant dispersal via the northern routes are relatively scarce. Wu et al. (2009) suggest that Long Distance Dispersal (LDD)-events in the genus *Euphrasia* L. (*Orobanchaceae*) could explain dispersal from the Asian mainland to Taiwan and consequently to the island of Luzon, possibly via typhoons, winter monsoons or for instance via birds. Matuszak et al. (2016) illustrate examples of possible colonization of Wallacea via Taiwan and the Philippines, although no direct evidence could be provided.

Species Distribution Modelling (SDM) has been increasingly used to model the possible viable habitat for species to better comprehend species distribution ranges (Araújo & Peterson 2012). This tool enables both the reconstruction of the historic habitat range for species and can predict future shifts in the ranges of suitable habitats. Occurrence data of species are correlated to current day climatic and soil conditions; these form the boundaries of the environmental parameters within which species can live. The areas in which species could be sustained, historically and in the future, are then conceived by projecting these parameters on the relevant estimated environmental conditions of the specified time periods.

For the historic reconstructions, species distributions are modelled for the Last Glacial Maximum, about 22 000 years ago, and

for the relatively warm period of the Mid-Holocene, about 6 000 years ago. These two periods cover the most extreme climatic conditions throughout recent geological history and for which palaeodata were available. Modelling future distributions will be based on two climate scenarios, denoted as the Representative Concentration Pathway (RCP) 2.6 and RCP 8.5. These implemented scenarios are recognized by the Intergovernmental Panel on Climate Change (IPCC 2022) and are based on different levels of emitted greenhouse gasses, which are related to the extent of implemented sustainable policies by governments and institutions on a global scale. The RCP2.6 scenario is the most positive and most sustainable development scenario, often considered to still be reversible, whilst RCP8.5 concerns the most negative scenario, with the greatest global impact and likely to be largely irreversible.

The aim of this paper is to investigate the possibility that a floral migration path was present between the SE Asian mainland and the Philippines, whereby three lines of evidence will be followed: A numerical comparison between both areas of species occurrences (present/absent) to see if taxa likely took this route, dated phylogenies with a clade in the SE Asian mainland and a clade on the Philippines to see when dispersal might have happened, and species distribution modelling to investigate if species could and can be present on both sides, in the past, present and future. The latter is added to see if former pathways remain intact as they will still be available under the influence of overpopulation and the related climate change (IPCC 2022). Shifts in plant species' ranges, both in altitude and latitude, are already discerned as a response to changes in regional climates (Walther et al. 2002, Scherrer et al. 2017, Rumpf et al. 2019).

MATERIALS AND METHODS

Several databases and data sets were prepared. All species names were checked against Plants of the World Online (POWO 2022) and for distribution modelling with its predecessor The Plant List (http://www.theplantlist.org/) via a module in R. Infraspecific taxa were reduced to the species. Only accepted names were included, while authors were adjusted where necessary. The family classification follows APG IV (APG IV 2016). The distributions were also taken from POWO (POWO 2022). Therefore, unresolved species (for which no distribution was mentioned) were deleted. Areas where the plants were introduced or where their presence was doubtful were ignored in the databases (as there is no easy way to tell when the introduction took place and if the plants were feral).

General data

The main data set (Table S1) contains all species present in the specimen database of Naturalis (L, U, WAG; for the acronyms see Thiers continuously updated) either in the Philippines or in the SE Asian mainland (here defined as the region comprising China, Vietnam, Taiwan, and Cambodia), respectively; whereby species only present in China N of c. 35° N or only W of c. 105° E were excluded (Fig. 1 dashed line) to prevent the influence of desert- and temperate species, all absent in the Malesian area. The lines also more or less coincide (therefore the use of c.) with several provinces shown in POWO (POWO 2022) as the most northern and western distributions for the species of our interest. Only the coastal countries were used as they are the most likely source of migrating species (and as there was only limited time for georeferencing. The latter is also the reason why the Ryukyu Islands were not included in the Species Distribution Modelling). Table S1 is no (complete) checklists for the area, it is only used to provide some simple overview statistics. A check of the Global Biodiversity Information Facility (GBIF; https://www.gbif.org/) for the area 0–30° N
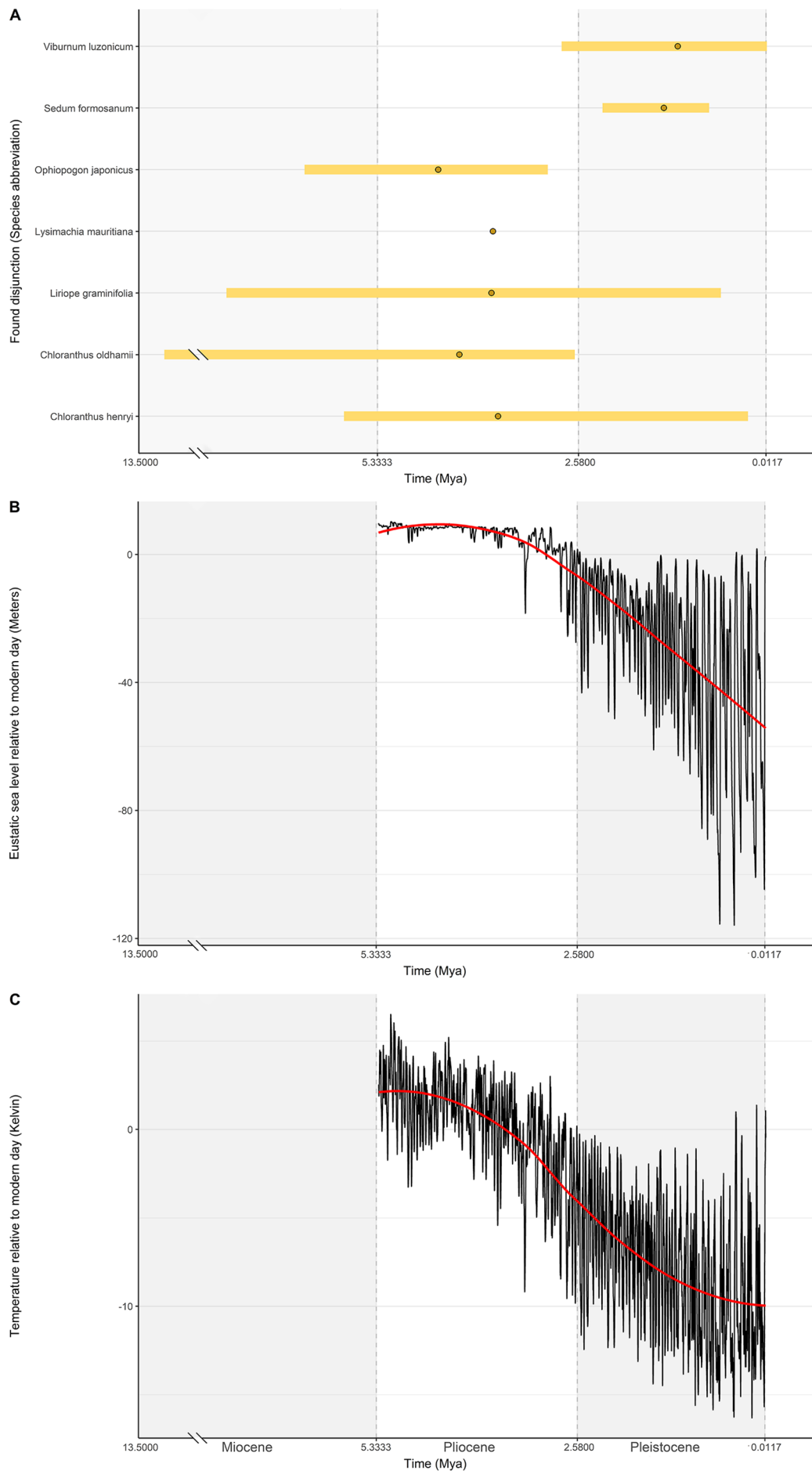
Fig. 2 A. Divergence times for various taxa indicated by a red dot and, when available, 95 %-confidence intervals with yellow beams. Species with distributions ranging further than the Philippines are indicated with a red dot. Different eras are alternately coloured, left to right: Miocene (until 5.33 Mya), Pliocene (5.33–2.58 Mya), Pleistocene (2.58–0.0117 Mya) and Holocene (0.0117–0 Mya); B. fluctuations in eustatic sea level during Pliocence, Pleistocene and Holocen; C. idem for the continental temperature on the Northern Hemisphere (data for B and C from De Boer et al. 2014).

and 100−130° E resulted in slightly more than 3 million herbarium records of flowering plants. Many of these proved to represent duplicates of material available at L, in some cases with an identification different from the one at L. Sorting out all duplicates and identifications proved too large to handle, even when pruning it down to the families treated in Flora Malesiana (Van Steenis 1950) (see Species Distribution Modelling below). We judged that in the vast majority of cases, the identification at L would be the correct one, regarding its very actively curated SE Asian collections. Furthermore, we judged the absence of a part of the specimen data would not seriously influence the outcomes related to the main dispersal patterns.

Phylogenetic data

The phylogenetic part of the study was executed mostly according to the method as used in Van den Ende et al. (2017). The species listed in Table S1 were used. Online scientific search engines were utilized to find dated molecular phylogenies containing the Philippines species/genera from the list, like Google Scholar (https://scholar.google.nl/). Used keywords were 'Species/Genus name', 'Dated Molecular Phylogeny', 'Philippines', 'Southeast Asia', and 'Taiwan'. Divergence times were noted in the list and distributional data were added when absent in the articles. Only the phylogenies containing Philippine species with a sister taxon/clade on the Asian mainland or Taiwan were kept. Species distributed over most of Malesia were excluded to eliminate species with a possibly internal-Malesian dispersal route to the Philippines; for instance if the Malay Peninsula, New Guinea, or Moluccas are inhabited too, then it is not obvious from where the ancestral species migrated. The remaining list of species was uploaded to Rstudio version 4.0.2 (Rstudio Team 2020) to plot in a Gantt chart. This Gantt chart showed the different species on the y-axis accompanied with their mean divergence times and 95 %-confidence intervals. To make the plot, the packages readxl, car, ggplot2, and tidyverse were used. The plot (Fig. 2A) was compared with the data on eustatic sea levels (Fig. 2B) and continental temperature (Fig. 2C) of the northern hemisphere based on 53 000 unique datapoints per parameter from De Boer et al. (2014) for the Pliocene (5.33−2.58 Ma) and Pleistocene (2.58−0.017 Ma) (Holocene 0.017 Ma till present added; the Miocene (to
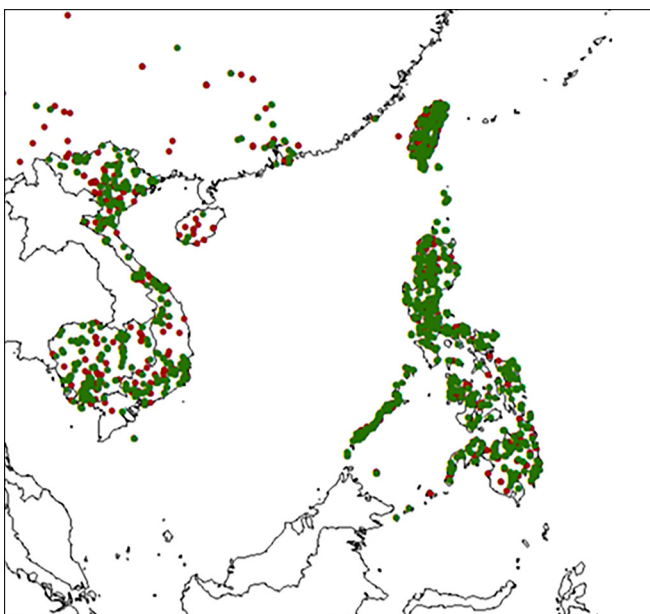

Fig. 3   Research area with specimen records shown in green and red dots; after modelling the species represented by the red dots did not deviate significantly from random or contained too few occurrences (less than 5) and were thus discarded.

Table 1   All environmental variables used in this study as abbreviation and a description of the variable. Blue rows: climatic variables; green row: elevation; yellow rows: edaphic variables. In black: all variables excluded based upon the correlation matrix or their variance inflation factor (VIF). In red: all variables included into the modelling of the species distribution.

Abbreviations	Description
Bio01	Annual mean temperature
Bio02	Mean diurnal range (mean of monthly (max temp - min temp))
Bio03	Isothermality (BIO2/BIO7) (×100)
Bio04	Temperature seasonality (standard deviation ×100)
Bio05	Max temperature of warmest month
Bio06	Min temperature of coldest month
Bio07	Temperature annual range (BIO5-BIO6)
Bio08	Mean temperature of wettest quarter
Bio09	Mean temperature of driest quarter
Bio010	Mean temperature of warmest quarter
Bio011	Mean temperature of coldest quarter
Bio012	Annual precipitation
Bio013	Precipitation of wettest month
Bio014	Precipitation of driest month
Bio015	Precipitation seasonality (coefficient of variation)
Bio016	Precipitation of wettest quarter
Bio017	Precipitation of driest quarter
Bio018	Precipitation of warmest quarter
Bio019	Precipitation of coldest quarter
Elev	Elevation
ALSA	Exchangeable aluminium (% of ECEC)
BSAT	Base saturation (% of ECEC)
BULK	Bulk density
CECC	Cation exchange capacity of clay fraction
CECS	Cation exchange capacity
CFRAG	Course fragments % (>2mm)
CLPC	Clay mass %
CNRT	C/N ratio
ECEC	Effective cation exchange capacity
ELCO	Electrical conductivity
ESP	Exchangeable sodium percentage (% of CECS)
GYPS	Gypsum content
PHAQ	pH in water
SDTO	Sand mass %
STPC	Silt mass %
TAWC	Volumetric water content (-33 to -1500 kPa cm/m^2)
TCEQ	Carbonate content
TOTC	Total carbon content
TOTN	Total nitrogen content

5.33 Ma) was ignored as it had no datapoints). A trendline was added to the climatic graphs by using the function stat_smooth() in ggplot2 (red lines in Fig. 2B, C). The option 'se' was set to 'FALSE' because of the large amount of datapoints.

Species Distribution Modelling

The research area ranges between 0−30° latitude and 100−130° longitude (Fig. 1, dashed line; slightly different from General Data as no outlines of provinces had to be followed). Occurrence data of plant species of families revised for the Flora Malesiana Project (Van Steenis 1950), thus ensuring the largest set of well-identified specimens from especially Malesia, were retrieved from the specimen database of Naturalis Biodiversity Center (L, U, WAG) for Cambodia, partially China, the Philippines, Taiwan, and Vietnam. Species names were checked against POWO (2022, see above) and with the R package Taxonstand (Cayuela et al. 2021) against ThePlantList (http://www.theplantlist.org/; superseded by https://www.worldfloraonline.org/).

The occurrence data were georeferenced in part with the co-operation of local researchers in the region (see acknowledge-

Table 2 Pearson's correlation values of all the variables (Table 1). The variables that were used in modelling are dark red, the correlations that exceeded the bounds of -0.7 and 0.7 are marked in light red; one of these was excluded. The bioclimatic variables are in blue, elevation in green and the edaphic variables in yellow.

	ALSA	BSAT	BULK	CECC	CECS	CFRAG	CLPC	CNRT	ECEC	ELCO	ESP	GYPS	PHAQ	SDTO	STPC	TAWC	TCEQ	TOTC	TOTN	
Bio01	—	-0.395	0.388	-0.287	-0.345	-0.356	-0.029	0.243	-0.319	-0.155	0.226	-0.087	0.090	0.556	0.029	0.152	-0.364	-0.164	-0.120	ALSA
Bio02	-0.138	—	0.597	0.839	0.464	-0.583	0.565	0.510	0.652	0.726	0.247	0.421	0.837	0.265	0.724	0.408	0.590	0.068	0.214	BSAT
Bio03	0.683	0.161	—	0.490	0.137	-0.969	0.613	0.792	0.351	0.480	0.510	0.293	0.929	0.842	0.691	0.494	0.098	-0.059	0.137	BULK
Bio04	-0.786	-0.131	-0.936	—	0.366	-0.472	0.516	0.301	0.81	0.513	0.169	0.366	0.697	0.119	0.663	0.368	0.357	-0.152	-0.020	CECC
Bio05	0.726	-0.206	0.100	-0.177	—	-0.316	0.69	0.598	0.443	0.279	0.060	0.207	0.350	-0.185	0.525	0.719	0.161	0.818	0.860	CECS
Bio06	0.960	-0.170	0.824	-0.898	0.534	—	-0.701	-0.886	-0.356	-0.447	-0.507	-0.284	-0.920	-0.787	-0.762	-0.668	-0.044	-0.146	-0.338	CFRAG
Bio07	-0.788	0.104	-0.915	0.967	-0.165	-0.922	—	0.715	0.688	0.339	0.337	0.274	0.690	0.164	0.704	0.661	0.019	0.334	0.486	CLPC
Bio08	0.879	-0.255	0.379	-0.513	0.785	0.773	-0.542	—	0.247	0.443	0.457	0.252	0.792	0.634	0.705	0.756	0.106	0.543	0.689	CNRT
Bio09	0.969	-0.067	0.813	-0.897	0.562	0.986	-0.893	0.768	—	0.361	0.325	0.300	0.523	-0.124	0.585	0.354	0.205	-0.102	0.014	ECEC
Bio010	0.768	-0.376	0.126	-0.211	0.972	0.597	-0.252	0.844	0.606	—	0.468	0.762	0.661	0.287	0.566	0.263	0.785	0.070	0.185	ELCO
Bio11	0.963	-0.036	0.835	-0.923	0.530	0.989	-0.912	0.765	0.993	0.570	—	0.586	0.456	0.436	0.370	0.286	0.040	0.001	0.098	ESP
Bio12	0.537	-0.270	0.676	-0.607	0.157	0.651	-0.688	0.320	0.616	0.241	0.604	—	0.400	0.135	0.343	0.196	0.298	0.029	0.098	GYPS
Bio13	0.443	-0.209	0.322	-0.417	0.197	0.473	-0.462	0.328	0.469	0.265	0.456	0.703	—	0.662	0.827	0.562	0.340	0.053	0.250	PHAQ
Bio14	0.333	-0.185	0.667	-0.463	0.010	0.473	-0.547	0.122	0.428	0.077	0.417	0.803	0.217	—	0.311	0.238	-0.017	-0.112	0.022	SDTO
Bio15	-0.226	0.272	-0.505	0.263	-0.079	-0.337	0.358	-0.045	-0.288	-0.126	-0.268	-0.550	0.127	-0.865	—	0.794	0.317	0.254	0.419	STPC
Bio16	0.458	-0.231	0.357	-0.442	0.186	0.498	-0.496	0.344	0.487	0.263	0.477	0.762	0.979	0.275	0.071	—	-0.019	0.621	0.725	TAWC
Bio17	0.342	-0.194	0.667	-0.462	0.024	0.479	-0.548	0.124	0.436	0.091	0.423	0.813	0.228	0.996	-0.872	0.285	—	0.030	0.074	TCEQ
Bio18	-0.032	-0.194	0.039	-0.032	-0.223	0.017	-0.122	0.067	-0.015	-0.112	-0.012	0.454	0.435	0.278	-0.037	0.489	0.276	—	0.976	TOTC
Bio19	0.420	-0.195	0.689	-0.523	0.082	0.550	-0.605	0.185	0.515	0.150	0.497	0.816	0.375	0.893	-0.740	0.409	0.900	0.156	—	TOTN
Elev	-0.746	0.504	-0.181	0.219	-0.917	-0.615	0.298	-0.793	-0.604	-0.967	-0.563	-0.361	-0.290	-0.224	0.292	-0.297	-0.241	0.044	-0.283	
	Bio01	Bio02	Bio03	Bio04	Bio05	Bio06	Bio07	Bio08	Bio09	Bio010	Bio11	Bio12	Bio13	Bio14	Bio15	Bio16	Bio17	Bio18	Bio19	Elev

ments) and by searching the gazetteer GeoNames (https://www.geonames.org/); while Google Maps (https://www.google.nl/maps) was primarily used to identify any discrepancies in the spelling, but also as the second source of coordinates when GeoNames did not contain a location. Records that could not be georeferenced (either because the locality could not be traced, or the area indicated was too large (more than c. 10 by 10 km, see below), or when the locality name occurred many times within the region) were discarded, just as records that were placed outside the borders of the selected countries. Lists of georeferenced data are available via the authors.

The species occurrence data were linked to five arc-minute grid cells (c. 10 by 10 km) and subsequently reduced to one record per species per grid cell, resulting in a grid of presence/absence data. The R package 'raster' (version 3.3-13; Karney 2013) was used to perform this transformation (all scripts used can be found in the Appendix). After cleaning the data, all species present in less than five grid cells were excluded from the analysis (Sheldeman & Van Zonenveld 2010; Van Proosdij et al. 2016 indicate a range of 3–13 cells depending on the extend of the distribution). The species in the final selection were divided into three categories; solely occurring on the Philippines, solely occurring in Southeast Asia, and those shared between the two regions (called shared species). The location of the records for all modelled species are shown in Fig. 3.

Table 3 The Variance Inflation Factor (VIF) for each variable (Table 1) retained after processing the correlation matrix, and the subsequent VIF for the variables before and after the removal of the variable with the highest VIF. Red outlines provide which variable were excluded based on their VIF.

a			b			c	
Variable	**VIF**		**Variable**	**VIF**		**Variable**	**VIF**
Bio01	3.207		Bio01	3.203		Bio01	3.091
Bio02	1.723		Bio02	1.723		Bio02	1.72
Bio03	2.989		Bio03	2.971		Bio03	2.937
Bio13	1.579		Bio13	1.578		Bio13	1.577
Bio17	1.938		Bio17	1.933		Bio17	1.923
Bio18	5.231		Bio18	5.230		Bio18	5.19
ALSA	4.136		ALSA	3.946		ALSA	2.135
BSAT	12.513		BSAT	12.437		CECS	1.396
CECS	15.343		CECS	1.774		ECEC	1.691
CNRT	22.592		ECEC	4.715		ESP	2.326
ECEC	4.725		ESP	3.065		GYPS	1.834
ESP	3.458		GYPS	2.170		SDTO	2.08
GYPS	2.306		SDTO	7.560		TCEQ	1.499
SDTO	18.535		TCEQ	2.648			
TCEQ	2.769						

Table 4 Numbers of families, genera and species present in SE Asian mainland (part of China, Vietnam, Taiwan, and Cambodia) and the Philippines. Unique means absent in the other area (e.g., thus unique to Philippines is not present in SE Asian mainland), it does not indicate endemism. The totals (last column) were used to calculate the percentages (between brackets).

	SE Asia mainland	Unique SE Asia mainland	Philippines	Unique Philippines	Shared	Total
No. of families	243 (95 %)	49 (19 %)	208 (81 %)	14 (5 %)	194 (75 %)	257
No. of genera	2040 (87 %)	973 (41 %)	1384 (59 %)	317 (13 %)	1067 (45 %)	2357
No. of species	8296 (68 %)	6906 (57 %)	5252 (43 %)	3861 (32 %)	1390 (11 %)	12158

Table 5 The 95th highest AUC values of the ninety-nine replicates. This AUC level was the resulting threshold that was required for models to surpass in order to significantly deviate from random.

Records	AUC	Records	AUC	Records	AUC
5	0.920	17	0.855	30	0.796
6	0.901	18	0.852	31	0.799
7	0.879	19	0.847	32	0.793
8	0.846	20	0.836	33	0.783
9	0.801	21	0.849	34	0.787
10	0.829	22	0.827	36	0.790
11	0.835	23	0.826	37	0.786
12	0.815	24	0.819	38	0.780
13	0.817	25	0.824	39	0.781
14	0.807	26	0.803	40	0.766
15	0.848	27	0.807	46	0.764
16	0.871	29	0.799	56	0.738

Fig. 4 The 95th highest AUC value of the ninety-nine replicates, per unique number of records, of the null models. The drop of AUC values between five and fifteen records has been well documented and methodically dealt with as described by Raes & Ter Steege (2007).

The current, the reconstructed historical and the predicted future climatic conditions were retrieved from WorldClim v. 2.1, the historical ones from v. 1.4 (www.worldclim.org; Table 1). Edaphic soil data were obtained from ISRIC (www.isric.org; Batjes 2012; Table 1). All data were either retrieved or transformed into a five arc-minute raster and cropped to the research area.

Many of the climatic and soil variables are derived from each other and thus highly correlated. In order to reduce overfitting (Peterson et al. 2007, Raes et al. 2014) a correlation matrix was constructed using R (Rteam*), based on which highly correlating variables were excluded, only retaining those with correlation values between -0.7 and 0.7 (Table 2). In case of ambiguity exclusions were based on the comparable work of Rutgrink et al. (2018). For the remaining variables, the variance inflation factor (VIF) was calculated. Those variables with the highest VIF were also excluded until all variables showed a VIF value below 10 (Table 3). For included and excluded variables see Table 4.

Modelling of historic conditions was done for the Mid-Holocene, about 6 000 years ago (ya), and the Last Glacial Maximum, about 22 000 ya; both with a spatial resolution of five arc minutes. The original climatic data was from CMIP5 (Coupled Model Intercomparison Project 5; Taylor et al. 2012) and was downscaled and calibrated using the WorldClim 1.4 'current' climate as a baseline. Due to the previous format of WorldClim 1.4, the degrees in temperature are stored as numbers a tenfold higher than data from WorldClim 2.1, representing an additional decimal number. Therefore, 'Annual mean temperature' and 'Mean diurnal range' were divided by ten to align with the present climatic data. Since 'Isothermality' concerns a ratio, rather than a measurement of temperature, it did not require any transformations. Future climatic conditions were produced using the climatic dataset from CMIP6 (O'Neill et al. 2016), which were downscaled and calibrated using WorldClim v. 2.1 'current' climate as a baseline. In this paper we implement the RCP2.6 and RCP8.5 pathways (IPCC 2013), the most optimistic scenario and the most extreme scenario, respectively.

The distribution models were made in MaxEnt v. 3.3.3k (https://biodiversityinformatics.amnh.org/open_source/maxent/; Phillips et al. 2005, 2006, Phillips 2006) with default settings except for: using five replicates, the addition of creating response curves, using the jackknife method for calculating the importance
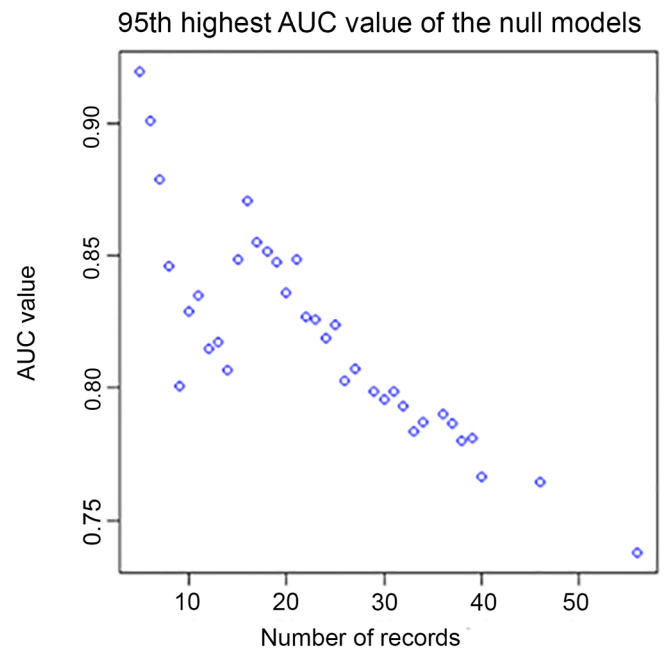
of the variables, and finally setting the threshold rule for the ten-percentile training presence. MaxEnt was selected as it had the best performance with presence-only data (Elith et al. 2010), outperforms other algorithms (Hernandez et al. 2006, Aguirre-Gutiérrez et al. 2013) and is less effected by georeference errors (Graham et al. 2008) and few presence records (Wisz et al. 2008).

The modelled species distributions were tested against random distributions via a bias-corrected null-model (Raes & Ter Steege 2007). For each number of grid cells occupied by a distribution (ranging from 5 to 56 records) 99 random distributions were made (thus each containing the collection biases) and analysed with MaxEnt under the same conditions. This resulted in 99 null-distributions per number of occupied grid cells (thus one for 5 occupied cells, one for 6, up to 56). Species deviated from random if their AUC values were in the top 5 % of the null model AUC values (Table 5; Fig. 4).

The non-random distribution models were transformed from the ascii files produced by MaxEnt into .TIFF files with ArcMap ModelBuilder in ArcGIS v. 10.6.1 (ESRI 2018). These files were turned into binary maps using the ten-percentile training presence via Python v. 2.7.1 (Van Rossum & Drake Jr 1995). In ArcMap the resulting binary maps per species were summed per region and were totalled for an overview of all species.

In order to decide whether species occurred or could occur in both areas, the number of Philippine species that were modelled to occur in grid cells on the mainland were counted and likewise for the mainland species (possibly) occurring in the Philippines. If large numbers of species have suitable habitat in the other region, yet they do not occur there, then migration is likely hindered and thus limited between the regions.

To visualize the projections into the past effectively, the rasters of the two historic species distributions were subtracted from present day distributions. The resulting areas show where the suitable habitat has decreased for a multitude of species since the historic setting in red, whilst the habitat suitability in green areas has increased. For the projections into the future, the present-day species distribution maps were subtracted from

the future distribution maps, showing the change in habitat suitability for the species that is projected to occur.

RESULTS

General data

Based on the data in Table S1 it is obvious that the SE Asian mainland and the Philippines share far more taxa at family level (75 %; Table S3) than at genus (45 %; Table S4) or species level (11 %) (Table 4; Table S1, S3). In the data set (Table S1) 20 % of the listed species only occur in the Philippines (2 444 out of 12 164 species; for exact numbers of endemic species in the Philippines see https://www.philippineplants.org/General/CDFPStats.html). On the other hand, regarding species absent from the Malay Peninsula (the more usual dispersal route to Malesia), 181 species are widespread from the SE Asian mainland to New Guinea and/or Australia including the Philippines (highlighted brown in Table S1). 444 Species (highlighted yellow in Table S1) are present in the SE Asian mainland and the Philippines, but not in the Malay Peninsula, 76 of them without a distinct pattern in Malesia. Other species, 92, show a pattern evading the everwet areas (highlighted blue in Table S1), thus are present in areas with a yearly dry monsoon. 276 Species are present in the SE Asian mainland and the Philippines only (of which 96 with a distribution of Taiwan and the Philippines; Table S2). This leaves 765 species present in the SE Asian mainland, the Philippines and at least the Malay Peninsula. It is obvious that species can extend their distribution over the SE Asian mainland and the Philippines, but they are not numerous.

Phylogenetic data

Of the list of species (Table S1) ten unique records remained. Some of these species were not on the original herbarium list, but were retrieved from dated molecular phylogenies. For details about, e.g., the publications, clades, their distributions, see Table 6. The mean divergence times, together with the 95 %-confidence intervals are visualised in Fig. 2A. Note that three of the ten taxa did not have 95 %-confidence intervals and reached further into Malesia than only the Philippines. These three taxa are therefore not visualized in the figure. Six of the ten mean divergence times date back to the middle Pliocene, two of the ten species have a mean divergence time around the Pliocene-Pleistocene border, and two species have a mean divergence time in the middle Pleistocene (n.b. some of the found species have 95 %-confidence intervals reaching deep into the Miocene). Most mean divergence times coincided with the higher eustatic sea levels (Fig. 2B) and continental temperatures (Fig. 2C).

The added climatic graphs from De Boer et al. (2014) show an overall decrease in northern hemisphere continental temperature and eustatic sea level from the middle Pliocene and onwards.

Many of the sister taxa of the selected species have relatively large distributions (the distributions of all species were added as no historical biogeographical analysis was available with node distribution reconstructions). Therefore, it is harder to interpret how these species have dispersed. Per species the potential pathway to the Philippines is discussed; these are also summarized in Table 6. Three species had a broader distribution than solely the Philippines, Taiwan and/or the SE Asian mainland and ranged to Wallacea (central Malesia). In all phylogenies used the taxa underneath the split were from the SE Asian mainland, which we translated as migration from the SE Asian mainland to the Philippines.

Lysimachia decurrens G.Forst.

Lysimachia decurrens (*Primulaceae*; Yan et al. 2018) is compared with its sister taxon *L. silvestrii* (Pamp.) Hand.-Mazz. *Lysimachia decurrens* has a wide distribution, from SE Asia mainland and some of the Malesian islands ranging from the Philippines, downwards in southeastern direction, towards New Caledonia. *Lysimachia silvestrii* occurs in North-Central China, South-Central China and Southeast China. Presumably, *L. decurrens* originated on the Asian mainland and reached a wide distribution before spreading to the Philippines (either directly or via Taiwan) and consequently downwards to more southeastern islands. The islands where *L. decurrens* occurs all have a tropical to monsoon climate, possibly explaining the distribution of the species (Van Steenis 1979, Kottek et al. 2006).

Stemona tuberosa Lour.

Stemona tuberosa (*Stemonaceae*; Chen et al. 2021) has a sister clade of four species, *S. japonica* (Blume) Miq., *S. mairei* (H.Lév.) K.Krause, *S. parviflora* C.H.Wright, and *S. sessilifolia* (Miq.) Miq., which all occur in South-Central China, North-Central China, Southeast China and Hainan. *Stemona tuberosa* is a widespread species ranging from India and Sri Lanka to SE Asian mainland, the Philippines, Sulawesi, the Lesser Sunda Islands and New Guinea. Most probably the species has spread from the Asian mainland towards Taiwan and the Philippines, either independently or via Taiwan to the Philippines, and from there on further towards the southeast. According to Chen et al. (2021), *Stemona tuberosa* prefers a monsoon climate. This could explain the current distribution pattern towards Sulawesi, New Guinea, and the Lesser Sunda Islands, as all of these regions contain areas with a monsoon climate (Van Steenis 1979, Kottek et al. 2006).

Lysimachia capillipes Hemsl.

Lysimachia capillipes (*Primulaceae*; Yan et al. 2018) has as sister taxon *L. insignis* Hemsl., a species that occurs in South-Central China, Southeast China and Vietnam. *Lysimachia capillipes* has a distribution ranging from South-Central China and Southeast China to Taiwan, the Philippines and even further southeastwards to New Guinea. Probably, dispersal towards Taiwan and the Philippines has taken place independently or dispersal towards the Philippine islands took place via Taiwan, with a further dispersal to New Guinea.

Viburnum luzonicum Rolfe

Viburnum luzonicum (*Viburnaceae*; Landis et al. 2021) has as direct sister taxon *V. tashiroi* Nakai. A somewhat more ancestrally diverged sister clade consists of *V. fordiae* Hance and *V. formosanum* (Hance) Hayata and ancestrally to these is *V. longiradiatum* P.S.Hsu. *Viburnum longiradiatum* is limited to South-Central China, whereas *V. formosanum* is limited to Taiwan. *Viburnum fordiae* occurs in South-Central China and Southeast China and *V. tashiroi* occurs only on the Ryukyu islands (Nansei-shoto). *Viburnum luzonicum* can be found in Vietnam, South-Central China, Southeast China, the Philippines and Taiwan. Based on these distributions, *V. luzonicum* most probably dispersed from the Asian mainland (either from China or from Vietnam) to the Philippines and Taiwan independently or indirectly via Taiwan to the Philippines. Dispersal via the Ryukyu islands could also be possible (directly to the Philippines or indirectly via Taiwan). It should be noted that the found posterior probabilities were relatively low (Landis et al. 2021).

Sedum formosanum N.E.Br.

Sedum formosanum (*Crassulaceae*; Ito et al. 2017) has as sister taxa *S. sekiteiense* Yamam. and *S. tricarpum* Makino, which occur on Taiwan and Japan, respectively. *Sedum formosanum*

occurs in Taiwan, the Philippines (Batan islands) and Japan. Ito et al. (2017) suggest that dispersal may have taken place from the Asian mainland to the Ryukyu islands (Nansei-shoto) and subsequently to Taiwan, the Philippines, and Japan.

The majority of species in the used tree have a distribution limited to only Taiwan. This could imply that dispersal to the Philippines (Batan) has taken place by dispersal via Taiwan. Ito et al. (2017) state that no prominent correlation could be found between geographical distance and genetical distance between different *S. formosanum* populations, indicating that recent dispersal between the populations could have taken place. Also possible is a dispersal event from the Ryukyu islands (Nansei-shoto) to the Philippines.

Chloranthus henryi Hemsl.

Chloranthus henryi (*Chloranthaceae*; Zhang et al. 2011) has as sister taxon *C. serratus* (Thunb.) Roem. & Schult., which has a wide distribution, ranging from Vietnam in the south to Korea and the Kuril islands in the north. *Chloranthus henryi*, however, is limited to North- and South-Central China, the islands of Taiwan, Hainan and Luzon (n.b. Co's digital Flora of the Philippines (CDFP continuously updated) indicates that *C. henryi* perhaps does not appear in the Philippines, but that it is a wrong interpretation of *C. oldhamii* Solms; POWO (2022) accepts both species and lists as distributions China, Taiwan and the Philippines for *C. henryi* and Taiwan and the Philippines for *C. oldhamii*). Dispersal to Luzon has probably taken place directly from the Asian mainland or from the island of Hainan. It is also possible that indirect dispersal towards Luzon via Taiwan has taken place.

Lysimachia mauritiana Lam.

Lysimachia mauritiana (*Primulaceae*; Yan et al. 2018) has as sister taxon *L. heterogenea* Klatt, which is limited to North-Central China and South-Central China, whereas *L. mauritiana* is found on Réunion, in North-Central China and Southeast China, Manchuria, Korea and the Kazan-Retto Islands. The species is also present on Taiwan and the Batan Islands (Philippines). Presumably, *L. mauritiana* has reached the Batan islands directly from the Asian mainland or indirectly via Taiwan or the Japanese Kazan-Retto islands. However, the latter option seems unlikely given the relatively remote location of the Kazan-Retto island group.

Liriope graminifolia (L.) Baker

Liriope graminifolia (*Asparagaceae*; Wang & Yang 2018) has as sister taxa *L. platyphylla* F.T.Wang & Tang (= *L. muscari* (Decne.) L.H.Bailey) and *L. spicata* Lour. These sister taxa not only occur in countries on the Asian mainland such as Vietnam, Cambodia, Laos and Korea, but also on Taiwan, the main islands of Japan and the Ryukyu islands. *Liriope graminifolia* has more or less the same distribution but also occurs in the northern Philippines. Distribution range expansion probably occurred either directly from the Asian mainland to the northern Philippines, or indirectly via Taiwan or the Ryukyu islands.

Chloranthus oldhamii Solms

Chloranthus oldhamii (*Chloranthaceae*; Zhang et al. 2011) is sister to *C. sessilifolius* K.F.Wu, which only occurs on mainland China. *Chloranthus oldhamii* only occurs on Taiwan and the Philippines. The two most likely explanations for the distributions are dispersal directly to Taiwan and the Philippines or indirect dispersal to the Philippines via Taiwan.

Ophiopogon japonicus (Thunb.) Ker Gawl.

Ophiopogon japonicus (*Asparagaceae*; Wang & Yang 2018). The sister clade contains the species *O. bockianus* Diels,

O. bockianus var. *angustifoliatus* F.T.Wang & Tang (= *O. angustifoliatus* (F.T.Wang & Tang) S.C.Chen), *O. lancangensis* H.Li & Y.P. Yang (= *O. tienensis* F.T.Wang & Tang), *O. longibracteatus* H.Li & Y.P.Yang (= *O. intermedius* D.Don) and *O. reversus* C.C.Huang. All of these species occur on the Asian mainland with *O. intermedius* as the species with the widest distribution, ranging to India and Hainan. Of the five sister taxa, only *O. reversus* occurs on Taiwan. *Ophiopogon japonicus* occurs on the Asian mainland, Japan, Taiwan, and the Philippines. Presumably, *O. japonicus* reached the Philippines either from the Asian mainland directly, or indirectly via Taiwan or the Ryukyu islands.

SPECIES DISTRIBUTION MODELLING

Data selection

After georeferencing, validifying the nomenclature of the species with POWO and The Plant List, and excluding species that occurred in less than five grid cells, 924 species were retained representing 8 891 records. These were modelled with MaxEnt and later analysed using the null models (see next subchapter). The number of environmental variables was reduced to fifteen based on the correlation matrix (Table 2). Subsequently, the VIF factors (Table 3) resulted in also discarding 'CNRT', the carbon/nitrogen ratio (Table 3A, VIF = 22.592), and 'BSAT', the base saturation as percentage of effective cation exchange capacity (Table 3B, VIF = 12.437). A description of all the environmental variables is given in Table 1.

Null models

The null models provided an AUC value per number of records (Table 5; Fig. 4). Only 284 species from 175 genera passed the null test and were finally included as their AUC value deviated significantly from random (ending in at least the top 5 of the null models). Of these 284 species (Table 7), 79 only occurred in the Philippines, 84 only in mainland Asia and 121 in both regions (check the discussion under Species Distribution Modelling below).

Present day projections

All compiled projections of species distributions, thus for present, past and future scenarios, are shown in Fig. 5. These are the summed binary maps with the presence/absence of all species based on the ten-percentile threshold as provided by MaxEnt. The colours in Fig. 5 show intervals of twenty species, except for the first 10 species, up to 90 species per grid cell to provide the clearest picture; n.b. the last class with all species combined (column 4 in Fig. 5) extends further than a 110 species. For the present-day distributions, it is noteworthy that species from the Philippines and the SE Asian mainland are indeed projected to have suitable habitats across the South China Sea. In particular, but poorly visible in the figures due to scaling, the most northern islands of the Philippines have a suitable habitat for species from the mainland (Fig. 5: A3). Still, the overall suitability is limited to several dozens of species (31–50 in some places, to 51–70 in other areas), and since Palawan and a small part of the Philippine coast seem solely suitable for mainland species, it might likewise be the species living there, which could endure the environmental conditions on the mainland. This could mean that part of these species cannot migrate via Taiwan into the suitable habitat and was subject to rare long-distance dispersal events.

An overview of how many species were projected with a suitable habitat in the opposing region can be found in Table 8. Since the difference is highly significant, we can conclude that many species from both groups have a sustainable habitat in the opposing region, but that for many species that distribution was never realised.

Table 6 Selected species together with their tree specifications, mean divergence times, 95 %-confidence intervals when known and the native distributions of the species and their sister taxon/clade. Mean divergence times and 95 %-confidence intervals with a ± were determined by measuring and calculating the branches. Divergence times and 95 %-confidence intervals without a ± were already mentioned in the tree in the original publication. An asterisk in the last two columns indicates that the characters are mentioned on genus level.

Species	Publication	Mean divergence	95 % interval	Tree-type	Notes	Native distribution found species	Native distribution sister taxon/clade
Lysimachia decurrens G.Forst. - *L. silvestrii* (Pamp.) Hand.-Mazz.	Yan et al. (2018)	± 2.5 Ma	N/A	MCC-tree	Further dispersal	Bangladesh, South-Central China, Southeast China, East Himalaya, Java, Laos, Lesser Sunda Is., Myanmar, Nepal, New Caledonia, Philippines, Sulawesi, Taiwan, Thailand, Vanuatu, Vietnam	**silvestrii:** North-Central China, South-Central China, Southeast China
Stemona tuberosa Lour. - *S. sessilifolia* (Miq.) Miq. clade	Chen et al. (2021)	2.59 Ma	N/A	MCC-tree	Further dispersal	Assam, Bangladesh, Cambodia, South-Central China, Southeast China, Hainan, India, Laos, Lesser Sunda Is., Maluku, Myanmar, New Guinea, Philippines, Sri Lanka, Taiwan, Thailand, Vietnam	**sessilifolia clade:** North-Central China, South-Central China, Southeast China, Hainan
Lysimachia capillipes Hemsl. - *L. insignis* Hemsl.	Yan et al. (2018)	± 3.64 Ma	N/A	MCC-tree	Further dispersal	South-Central China, Southeast China, New Guinea, Philippines (North Luzon), Taiwan	**insignis:** South-Central China, Southeast China, Vietnam
Viburnum luzonicum Rolfe - *V. longiradiatum* P.S.Hsu clade	Landis et al. (2021)	± 1.22 Ma	± 0–2.81 Ma	MCC-tree	PP > 0.75	South-Central China, Southeast China, Philippines, Taiwan, Vietnam	**fordiae, formosanum, longiradiatum, tashiroi:** South-Central China, Southeast China, Nansei-shoto, Taiwan
Sedum formosanum N.E.Br. - *S. sekiteiense* Yamam./ *S. tricarpum* Makino	Ito et al. (2017)	1.41 Ma	0.79–2.25 Ma	ITS and cpDNA MCC-tree	–	Japan, Nansei-shoto, Philippines	**sekiteiense/tricarpum:** Japan, Taiwan (Batan islands), Taiwan
Chloranthus henryi Hemsl. - *C. serratus* (Thunb.) Roem. & Schult.	Zhang et al. (2011)	± 3.68 Ma	± 0.26–5.79 Ma	Bayesian uncorrelated lognormal tree		North-Central China, South-Central China, Southeast China, Hainan, Philippines (Luzon), Taiwan	**serratus:** South-Central China, Southeast China, Hainan, Japan, Korea, Kuril Is., Vietnam
Lysimachia mauritiana Lam. - *L. heterogenea* Klatt	Yan et al. (2018)	± 3.75 Ma	N/A	MCC-tree	–	North-Central China, South-Central China, Kazan-retto, Korea, Manchuria, Philippines (Batan islands), Réunion, Taiwan	**heterogenea:** South-Central China, Southeast China
Liriope graminifolia (L.) Baker - *L. spicata* Lour./*L. platyphylla* F.T.Wang & Tang	Wang & Yang (2018)	± 3.77 Ma	± 0.63–7.4 Ma	ITS Bayesian tree	Calculated mean between two nodes, *L. platyphylla* is *L. muscari* (Decne.) L.H.Bailey	North-Central China, South-Central China, Southeast China, (Northern) Philippines, Taiwan	**spicata/muscari:** Cambodia, North-Central China, South-Central China, Southeast China, Hainan, Japan, Korea, Nansei-shoto, Taiwan, Vietnam
Chloranthus oldhamii Solms - *C. sessilifolius* K.F.Wu	Zhang et al. (2011)	± 4.21 Ma	± 2.63–13.16 Ma	Bayesian un-correlated lognormal tree		Philippines, Taiwan	**sessilifolius:** South-Central China, Southeast China
Ophiopogon japonicus (Thunb.) Ker Gawl. - *O. bockianus* Diels clade	Wang & Yang (2018)	± 4.5 Ma	± 3–6.33 Ma	ITS Bayesian tree	Synonyms in sister taxa	North-Central China, South-Central China, Southeast China, Japan, Korea, Philippines, Taiwan, Vietnam	**bockianus clade:** Assam, Bangladesh, Cambodia, North-Central China, South-Central China, Southeast China, East Himalaya, West Himalaya, Hainan, India, Myanmar, Nanseishoto, Nepal, Sri Lanka, Taiwan, Thailand, Tibet, Vietnam

Table 6 (cont.)

Species	Route	Fruit type	Fruit colour	Seed size (mm)	Altitude per genus (m)
Lysimachia decurrens G.Forst. - *L. silvestrii* (Pamp.) Hand.-Mazz.	Asian mainland → Philipines (either via Taiwan or independently) → Sulawesi/Java/New Guinea/Lesser Sunda Islands	Capsule (Bentvelzen 1960)	–	0.4–2 by 0.3–1.8 by 0.3–1.5 (Oh et al. 2008)*	400–2000 (Bentvelzen 1960)
Stemona tuberosa Lour. - *S. sessilifolia* (Miq.) Miq. clade	Asian mainland → Philipines (either via Taiwan or independently) → Sulawesi/New Guinea / Lesser Sunda Islands	Capsule (Duyfjes 1993)	Green (Duyfjes 1993)	10–12 by c. 5 (Inthachub et al. 2010)	< 500 (Duyfjes 1993)
Lysimachia capillipes Hemsl. - *L. insignis* Hemsl.	South-Central China/Southeast China/Vietnam → Philippines (either via Taiwan or independently) → New Guinea	Capsule (Bentvelzen 1960)	–	0.4–2 by 0.3–1.8 by 0.3–1.5 (Oh et al. 2008)*	400–2000 (Bentvelzen 1960)
Viburnum luzonicum Rolfe - *V. longiradiatum* P.S.Hsu Fan clade	South-Central China/Southeast China/Vietnam → Philippines (via Taiwan or independently) or Asian mainland → Nansei-shoto → Philippines (either via Taiwan or independently)	Drupe (Wilkinson 1948)	Red, maturing black when ripe (Subilla & Baoanan 2021)	3.6–3.8 by 3.1–3.5 by 1.7–1.9 (Jacobs et al. 2008)	800–2000 (herbarium specimens L)
Sedum formosanum N.E.Br. - *S. sekiteiense* Yamam./ *S. tricarpum* Makino	Taiwan/Nansei-shoto → Philippines	Follicle (Choi et al. 2020)	Whitish (Van Steenis 1982)	0.4–0.5 by c. 0.1 (Choi et al. 2020)	Coast (Van Steenis 1982)
Chloranthus henryi Hemsl. - *C. serratus* (Thunb.) Roem. & Schult.	South-Central China/Southeast China → Philippines (via Taiwan or independently) or Hainan → Philippines	Drupe (Verdcourt 1984)	Mostly white* (Verdcourt 1984)	–	100–1200 (Verdcourt 1984)
Lysimachia mauritiana Lam. - *L. heterogenea* Klatt	South-Central China/Southeast China → Philippines (via Taiwan or independently) or South-Central China/Southeast China → Kazan-Retto → Philippines	Capsule (Bentvelzen 1960)	–	0.4–2 by 0.3–1.8 by 0.3–1.5 (Oh et al. 2008)*	400–2000 (Bentvelzen 1960)
Liriope graminifolia (L.) Baker - *L. spicata* Lour./*L. platyphylla* F.T.Wang & Tang	South-Central China/Southeast China/Taiwan/Ryukyu islands → Philippines (via Taiwan/Ryukyu islands or independently)	Drupe-like, Broussard (2007) and references therein	Black*, Broussard (2007) and references therein	–	575–1420 (herbarium specimens L)
Chloranthus oldhamii Solms - *C. sessilifolius* K.F.Wu	South-Central China/Southeast China → Philippines (via Taiwan or independently)	Drupe, Verdcourt (1984)	Mostly white* (Verdcourt 1984)	c. 2 diam (Cullen 2013)	100–1200 (Verdcourt 1984)
Ophiopogon japonicus (Thunb.) Ker Gawl. - *O. bockianus* Diels clade	North-Central China/South-Central China/Southeast China/ Taiwan/Ryukyu islands → Philippines (via Taiwan/Ryukyu islands or independently)	Drupe-like, Broussard (2007) and references therein	Blue*, Broussard (2007) and references therein	6–8 diam (Nesom 2010)	200–2800 (Nesom 2010)

Table 7 List of all 284 species included in this research for which the calculated AUC value (Area Under the Curve; last column) was significantly better than the relevant null model. The table provides the species name, the number of grid cells with records, the region (PP = Philippine species, AM = Southeast Asia mainland species (see Materials and Methods for definition of SE Asia mainland), PM = shared species; in yellow species that are more widespread in the models) and the AUC value of each species.

Mainland species

Species	Filled grid cells	Region(s)	Region(s) modelled	AUC
Acalypha australis	5	MA	MA	0.992
Acer albopurpurascens	6	MA	PM	0.984
Afzelia xylocarpa	7	MA	PM	0.912
Alangium handelii	9	MA	PM	0.861
Antidesma fordii	5	MA	PM	0.977
Ardisia cornudentata	5	MA	PM	0.947
Ardisia quinquegona	7	MA	PM	0.932
Asparagus cochinchinens	5	MA	MA	0.976
Bauhinia cardinalis	5	MA	PM	0.953
Bridelia balansae	9	MA	PM	0.911
Callicarpa pedunculata	5	MA	PM	0.985
Calycopteris floribunda	5	MA	PM	0.979
Canarium album	7	MA	PM	0.916
Canarium subulatum	6	MA	PM	0.922
Capparis acutifolia	8	MA	PM	0.857
Capparis cantoniensis	5	MA	PM	0.938
Castanopsis indica	9	MA	PM	0.872
Celtis formosana	5	MA	MA	0.938
Chaetocarpus castanocar	7	MA	PM	0.909
Clerodendrum cyrtophyllu	8	MA	PM	0.922
Clinopodium umbrosum	5	MA	MA	0.944
Cnestis palala	6	MA	PM	0.913
Coccinia grandis	8	MA	PM	0.863
Combretum quadrangular	6	MA	PM	0.965
Combretum trifoliatum	6	MA	PM	0.917
Corydalis tashiroi	5	MA	MA	0.951
Dillenia hookeri	6	MA	PM	0.916
Dipterocarpus intricatus	6	MA	PM	0.907
Elaeagnus oldhamii	7	MA	PM	0.924
Eriobotrya bengalensis	12	MA	PM	0.884
Eurya gnaphalocarpa	5	MA	MA	0.985
Ficus hispida	13	MA	PM	0.859
Ficus ischnopoda	6	MA	PM	0.961
Holarrhena pubescens	10	MA	PM	0.866
Hopea odorata	7	MA	PM	0.904
Hymenocardia punctata	6	MA	PM	0.963
Kibatalia macrophylla	5	MA	MA	0.941
Korthalsella japonica	5	MA	PM	0.956
Lagerstroemia floribunda	8	MA	PM	0.932
Maesa japonica	7	MA	PM	0.944
Maesa tenera	7	MA	MA	0.935
Magnolia compressa	5	MA	MA	0.963
Mallotus apelta	10	MA	PM	0.834
Mallotus barbatus	9	MA	PM	0.902
Mallotus metcalfianus	6	MA	PM	0.973
Microtropis fokienensis	6	MA	PM	0.940
Ochna integerrima	10	MA	PM	0.880
Pantadenia adenanthera	7	MA	PM	0.889
Peltophorum dasyrachis	14	MA	PM	0.885
Pilogyne bodinieri	9	MA	PM	0.872
Prunus fordiana	6	MA	PM	0.926
Prunus phaeosticta	6	MA	PM	0.968
Pterocarpus macrocarpus	12	MA	PM	0.931
Quercus glauca	5	MA	MA	0.955
Rhynchospora longisetis	5	MA	PM	0.938
Rosa laevigata	5	MA	PM	0.954
Rubus alceaefolius	6	MA	PM	0.918
Rubus cochinchinensis	8	MA	PM	0.861
Rubus ellipticus	6	MA	PM	0.957
Rubus leucanthus	9	MA	PM	0.866
Saraca declinata	7	MA	PM	0.955
Saurauia tristyla	5	MA	PM	0.920
Schleichera oleosa	7	MA	PM	0.980
Scolopia oldhamii	5	MA	MA	0.982
Shorea roxburghii	6	MA	PM	0.951
Shorea siamensis	7	MA	PM	0.954
Shorea thorelii	6	MA	PM	0.911
Sindora siamensis	5	MA	PM	0.935
Sorbus corymbifera	8	MA	PM	0.883
Spatholobus parviflorus	8	MA	PM	0.948
Strychnos nux-vomica	9	MA	PM	0.930
Styrax formosanus	5	MA	MA	0.977
Styrax suberifolius	6	MA	PM	0.927
Suregada multiflora	11	MA	PM	0.912
Symplocos anomala	8	MA	PM	0.872
Symplocos paniculata	7	MA	PM	0.885
Tabernaemontana bovina	10	MA	PM	0.847
Terminalia alata	6	MA	PM	0.945
Terminalia chebula	8	MA	PM	0.874
Terminalia corticosa	5	MA	PM	0.935
Torilis japonica	5	MA	PM	0.969
Triadica sebifera	9	MA	PM	0.941
Walsura robusta	7	MA	PM	0.926
Willughbeia edulis	9	MA	PM	0.862

Widespread (SE Asia and Philippines)

Species	Filled grid cells	Region(s)	Region(s) modelled	AUC
Acacia concinna	9	PM	PM	0.849
Acalypha kotoensis	7	PM	PM	0.923
Aglaia lawii	27	PM	PM	0.876
Alchornea rugosa	25	PM	PM	0.915
Allophylus cobbe	46	PM	PM	0.853
Alstonia scholaris	9	PM	PM	0.811
Amphineurion marginatum	8	PM	PM	0.854
Antidesma bunius	21	PM	PM	0.884
Antidesma ghaesembilla	37	PM	PM	0.854
Antidesma montanum	56	PM	PM	0.899
Aphanamixis polystachya	23	PM	PM	0.923
Archidendron clypearia	31	PM	PM	0.859
Barringtonia acutangula	14	PM	PM	0.841
Barringtonia racemosa	10	PM	PM	0.874
Bauhinia malabarica	9	PM	PM	0.863
Bischofia javanica	23	PM	PM	0.877
Bothriospermum tenellum	7	PM	PM	0.913
Breynia vitis-idaea	12	PM	PM	0.848
Bridelia stipularis	11	PM	PM	0.854
Bruguiera cylindrica	5	PM	PM	0.930
Buchanania arborescens	38	PM	PM	0.908
Buchanania lucida	5	PM	PM	0.925
Carallia brachiata	21	PM	PM	0.905
Casearia grewiifolia	19	PM	PM	0.851
Connarus semidecandrus	5	PM	PM	0.941
Crotalaria retusa	10	PM	PM	0.842
Crotalaria trichotoma	6	PM	PM	0.924
Croton argyratus	11	PM	PM	0.866
Crypteronia paniculata	10	PM	PM	0.829
Cyperus compactus	14	PM	PM	0.851
Cyperus kyllingia	11	PM	PM	0.841
Cyperus laxus	13	PM	PM	0.840
Desmodium microphyllum	8	PM	PM	0.847
Desmodium velutinum	9	PM	PM	0.915
Dipterocarpus grandiflorus	12	PM	PM	0.816
Dysoxylum alliaceum	8	PM	PM	0.850
Endocomia macrocoma	11	PM	PM	0.854
Euphorbia atoto	10	PM	PM	0.877
Eurya acuminata	12	PM	PM	0.888
Fibraurea tinctoria	5	PM	PM	0.928
Ficus ampelas	39	PM	PM	0.905
Ficus benguetensis	30	PM	PM	0.913
Ficus benjamina	26	PM	PM	0.891
Ficus caulocarpa	9	PM	PM	0.811
Ficus cumingii	37	PM	PM	0.896
Ficus fistulosa	32	PM	PM	0.892
Ficus heteropleura	31	PM	PM	0.918
Ficus microcarpa	26	PM	PM	0.857
Ficus pedunculosa	13	PM	PM	0.820
Ficus ruficaulis	25	PM	PM	0.896
Ficus septica	37	PM	PM	0.917
Ficus subulata	29	PM	PM	0.876
Ficus sumatrana	20	PM	PM	0.964
Ficus tinctoria	14	PM	PM	0.893
Ficus variegata	29	PM	PM	0.860
Ficus virens	6	PM	PM	0.938
Ficus virgata	9	PM	PM	0.887
Fimbristylis dichotoma	27	PM	PM	0.836
Fimbristylis miliacea	15	PM	PM	0.873
Fimbristylis tetragona	8	PM	PM	0.945
Flagellaria indica	12	PM	PM	0.919
Flemingia strobilifera	22	PM	PM	0.910
Flueggea virosa	34	PM	PM	0.864
Gaultheria leucocarpa	9	PM	PM	0.857
Gmelina philippensis	9	PM	PM	0.806
Gomphia serrata	6	PM	PM	0.940
Guioa pleuropteris	21	PM	PM	0.912
Gynostemma pentaphyllu	20	PM	PM	0.867
Harpullia arborea	25	PM	PM	0.945
Heynea trijuga	9	PM	PM	0.884
Hiptage benghalensis	10	PM	PM	0.845
Homonoia riparia	34	PM	PM	0.841
Horsfieldia irya	5	PM	PM	0.974
Hypolytrum nemorum	20	PM	PM	0.909
Hyptis capitata	9	PM	PM	0.818
Indigofera zollingeriana	11	PM	PM	0.853
Leea guineensis	29	PM	PM	0.849
Leea indica	23	PM	PM	0.835
Leea philippinensis	13	PM	PM	0.848

Table 7 (cont.)

Species	Filled grid cells	Region(s)	Region(s) modelled	AUC
Lepisanthes rubiginosa	25	PM	PM	0.842
Lepisanthes tetraphylla	20	PM	PM	0.885
Lonicera acuminata	6	PM	PM	0.934
Macaranga tanarius	29	PM	PM	0.872
Macroptilium lathyroides	5	PM	PM	0.923
Magnolia liliifera	19	PM	PM	0.910
Mallotus mollissimus	40	PM	PM	0.932
Mallotus philippensis	38	PM	PM	0.819
Mallotus repandus	24	PM	PM	0.846
Mallotus resinosus	14	PM	PM	0.895
Mallotus tiliifolius	13	PM	PM	0.831
Melanolepis multiglandulo	31	PM	PM	0.867
Meliosma pinnata	22	PM	PM	0.851
Merremia tridentata	6	PM	PM	0.947
Merremia umbellata	6	PM	PM	0.942
Microdesmis caseariifolia	15	PM	PM	0.909
Millettia pinnata	10	PM	PM	0.856
Mischocarpus pentapetalu	22	PM	PM	0.898
Osmelia philippina	12	PM	PM	0.842
Parameria laevigata	19	PM	PM	0.883
Phylacium bracteosum	7	PM	PM	0.886
Pleomele angustifolia	5	PM	PM	0.960
Premna tomentosa	14	PM	PM	0.817
Prunus grisea	14	PM	PM	0.820
Pterocarpus indicus	24	PM	PM	0.907
Pueraria phaseoloides	12	PM	PM	0.889
Quisqualis indica	15	PM	PM	0.865
Rubus rolfei	5	PM	PM	0.960
Sandoricum koetjape	11	PM	PM	0.857
Scurrula ferruginea	6	PM	PM	0.920
Scutellaria indica	8	PM	PM	0.935
Semecarpus cuneiformis	25	PM	PM	0.908
Shorea guiso	11	PM	PM	0.861
Siphonodon celastrineus	18	PM	PM	0.911
Strophanthus caudatus	10	PM	PM	0.849
Symplocos cochinchinens	40	PM	PM	0.837
Trema cannabina	9	PM	PM	0.856
Uraria lagopodioides	12	PM	PM	0.836
Viburnum luzonicum	21	PM	PM	0.884
Viburnum odoratissimum	9	PM	PM	0.820
Viscum ovalifolium	9	PM	PM	0.814
Wikstroemia indica	9	PM	PM	0.841

Philippine species

Species	Filled grid cells	Region(s)	Region(s) modelled	AUC
Acalypha amentacea	33	PP	PM	0.951
Anisoptera thurifera	9	PP	PM	0.813
Antidesma edule	8	PP	PM	0.863
Antidesma microcarpum	9	PP	PM	0.801
Antidesma tomentosum	19	PP	PM	0.949
Aporosa banahaensis	8	PP	PM	0.872
Aporosa sphaeridiophora	12	PP	PM	0.879
Aporosa symplocifolia	9	PP	PM	0.904
Archidendron scutiferum	6	PP	PM	0.901
Baccaurea philippinensis	14	PP	PM	0.864
Baccaurea tetrandra	22	PP	PM	0.896
Balakata luzonica	6	PP	PM	0.933
Breynia cernua	29	PP	PM	0.932
Buchanania insignis	8	PP	PM	0.908
Buchanania microphylla	6	PP	PM	0.932
Buchanania nitida	9	PP	PM	0.920

Species	Filled grid cells	Region(s)	Region(s) modelled	AUC
Canarium hirsutum	12	PP	PM	0.836
Casearia fuliginosa	11	PP	PM	0.886
Clerodendrum minahassa	15	PP	PM	0.900
Clethra canescens	18	PP	PM	0.886
Codiaeum luzonicum	7	PP	PM	0.895
Croton leiophyllus	7	PP	PM	0.960
Dichapetalum gelonioides	6	PP	PM	0.913
Discocalyx cybianthoides	8	PP	PM	0.851
Dracontomelon dao	14	PP	PM	0.902
Dysoxylum arborescens	7	PP	PM	0.935
Dysoxylum gaudichaudian	5	PP	PM	0.958
Euonymus javanicus	18	PP	PM	0.876
Ficus callophylla	18	PP	PM	0.879
Ficus carpenteriana	10	PP	PM	0.909
Ficus chrysolepis	11	PP	PM	0.862
Ficus fiskei	10	PP	PM	0.856
Ficus nota	13	PP	PM	0.827
Ficus pisifera	27	PP	PM	0.955
Ficus pseudopalma	12	PP	PM	0.876
Ficus satterthwaitei	13	PP	PM	0.856
Ficus ulmifolia	11	PP	PM	0.884
Ficus uniglandulosa	5	PP	PP	0.995
Gironniera celtidifolia	12	PP	PM	0.944
Glyptopetalum euphlebiun	5	PP	PM	0.927
Guioa koelreuteria	30	PP	PM	0.923
Gymnacranthera farquhar	8	PP	PM	0.863
Hiptage luzonica	9	PP	PM	0.869
Homalanthus macradeniu	15	PP	PM	0.937
Homalanthus populneus	31	PP	PM	0.962
Hopea philippinensis	10	PP	PM	0.866
Horsfieldia ardisiifolia	5	PP	PM	0.962
Hoya odorata	5	PP	PM	0.936
Knema glomerata	23	PP	PM	0.859
Knema korthalsii	5	PP	PM	0.951
Koordersiodendron pinnat	10	PP	PM	0.835
Leea aculeata	18	PP	PM	0.872
Lepidopetalum perrottetii	18	PP	PM	0.887
Lithocarpus coopertus	11	PP	PM	0.861
Macaranga bicolor	14	PP	PM	0.870
Macaranga hispida	16	PP	PM	0.879
Mangifera altissima	8	PP	PM	0.858
Merremia peltata	5	PP	PM	0.922
Myristica simiarum	7	PP	PM	0.894
Ochrosia glomerata	14	PP	PM	0.874
Octomeles sumatrana	5	PP	PM	0.932
Ormosia calavensis	12	PP	PM	0.858
Paracroton pendulus	13	PP	PM	0.889
Parartocarpus venenosus	9	PP	PM	0.844
Pentaphragma grandifloru	9	PP	PM	0.844
Pittosporum resiniferum	13	PP	PM	0.830
Prunus marsupialis	14	PP	PM	0.949
Radermachera pinnata	5	PP	PM	0.932
Scolopia luzonensis	10	PP	PM	0.832
Shorea polysperma	12	PP	PM	0.933
Suregada glomerulata	16	PP	PM	0.899
Swintonia foxworthyi	6	PP	PM	0.990
Trigonachras cultrata	5	PP	PM	0.928
Trigonopleura dubia	6	PP	PM	0.992
Turpinia borneensis	9	PP	PM	0.868
Vaccinium cumingianum	8	PP	PM	0.866
Vatica mangachapoi	12	PP	PM	0.865
Vavaea amicorum	9	PP	PM	0.909
Wikstroemia ovata	9	PP	PM	0.820

Historic projections

From the Last Glacial Maximum (Fig. 6 row 4) to the present, the suitability of habitat greatly increases (green) for the modelled species. The decrease (red) of habitat suitability mostly reflects the loss of land area due to rising sea levels, but additionally there is a distinct decrease of habitat suitability in the centre of Borneo and the Malay Peninsula. This change primarily affects the species shared between the mainland and the Philippines, whilst this effect is discernible to a lesser extent for species from the Philippines.

Since the Mid-Holocene (Fig. 6 row 3), there is another visible decrease of habitat suitability on Borneo and to its east for shared species mostly, yet in this projection, the decrease envelops the whole island. There are many other patches of decreasing habitat suitability; for shared species and those from the mainland, there is a visible decrease along the Chinese coast; the habitat suitability for shared species additionally decreases in the southern part of Cambodia and Vietnam; and finally, the northern part of the Philippines shows a decrease for Philippine and shared species.

Future projections

Overall, the future projections (Fig. 6 row 1 & 2) show that, concerning plant species, climate change tends to increase the area suitable to harbour a larger number of species. This increase is mainly found on Borneo and Mainland Southeast Asia. The group that is most negatively affected are the species shared between the Philippines and the mainland. For these species, there is a decrease in habitat suitability both in southern Cambodia, near the coast of southeast Vietnam, and the central to northern part of the Philippines including Palawan. Yet, this is only clearly visible in the more extreme RCP8.5 scenario.
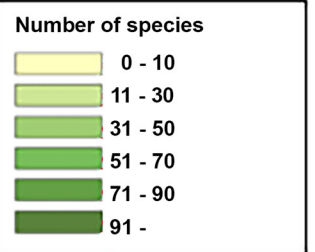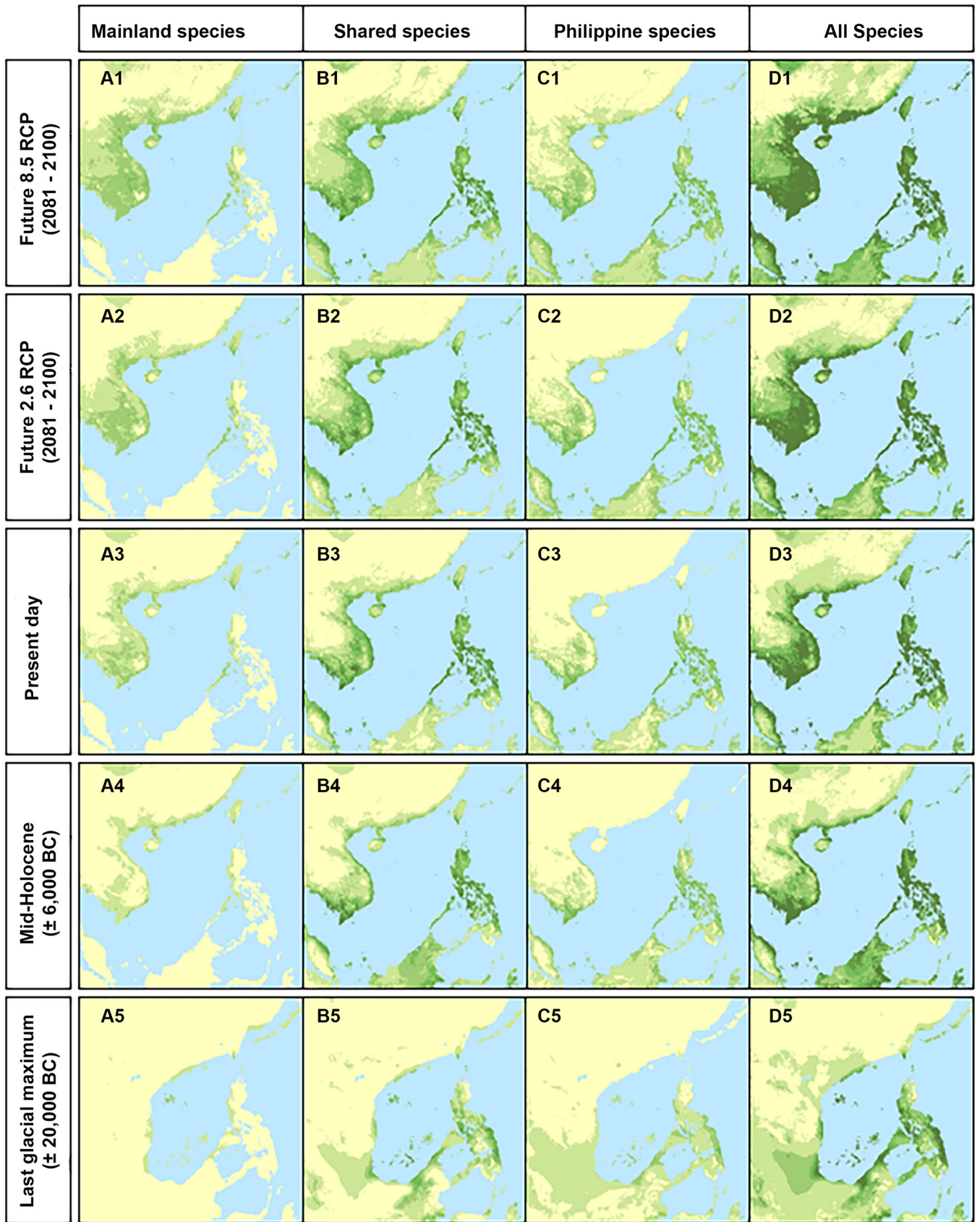
Fig. 5 Modelled stacked species distributions. The shades of yellow/green, see legend, show the numbers of species per grid cell. Row 1: Future projection using the environmental conditions based on the Representative Concentration Pathway 8.5. Row 2: Future projection using the environmental conditions based on the Representative Concentration Pathway 2.6. Row 3: Present day species distributions. Row 4: Past projection using the environmental conditions of the Mid-Holocene period (± 6.000 BC). Row 5: Past projection using the environmental conditions of the Last Glacial Maximum (± 20.000 BC). Column A: Species solely found on Mainland Southeast Asia plus China and absent from the Philippines. Column B): Species both found on the mainland and on the Philippines. Column C: Species solely found on the Philippines and absent on the mainland. Column D: All species used in the species distribution models.

Table 8   Comparison and statistical testing between the regions where the species occur presently and the areas that seemed to be suitable after modelling. Only the species present in one region are taken in consideration (this summarizes table 7). X^2: the chi-squared values as calculated for the difference between the observation and the expectation if chances were equal. The p-value is based on 1 degree of freedom (df = 1; 10.828 in table for probabilities of the chi-square).

Mainland species					
Modelled suitable habitat					
Mainland	Both	Total	Expected	X^2	p-value
12	72	84	42	42.857	>0.001

Philippine species					
Modelled suitable habitat					
Philippines	Both	Total	Expected	X^2	p-value
1	78	79	39.5	75.051	>0.001

Table 9   Contributions of the variables per regional group to the contemporary distribution models: the Kruskal-Wallis comparison (Kruskal & Wallis 1952) with the Bonferroni (1936) adjustment for multiple group comparison; MA-PM: comparison between species from the SE Asian mainland and the shared species group; MA-PP: comparison between species from the mainland and species from the Philippines; PM-PP: Comparison between species from the Philippines and the shared species group. Blue variables are the bioclimatic variables, yellow variables the edaphic variables. Green cells indicate a significant difference, while red cells indicate a non-significant difference.

| Variable	Kruskal Wallis	MA-PM	MA-PP	PM-PP
Annual mean T	0.5541	0.54	0.54	0.54
Mean diurnal range	1.786e-08	0.00387	3.8e-08	0.00011
Isothermality	7.652e-08	0.00028	6.1e-09	0.07716
Precipitation wettest month	3.21e-10	4.0e-09	6.4e-08	0.83
Precipitation driest quarter	0.007057	0.012	0.012	0.330
Precipitation warmest quarter	2.616e-12	6.6e-08	1.1e-10	0.002
Exchangeable aluminium	7.432e-08	0.016	1.2e-07	9.7e-05
Cation exchange cap	6.517e-12	1.0e-10	4.2e-09	0.97
Effective cation exchange	0.006676	0.0054	0.1504	0.1504
Exchangeable sodium %	0.007734	0.0427	0.0083	0.1997
Gypsum content	0.000861	0.4565	0.0013	0.0056
Sand mass	0.1241	0.52	0.15	0.15
Carbonate content	0.01527	0.208	0.018	0.065
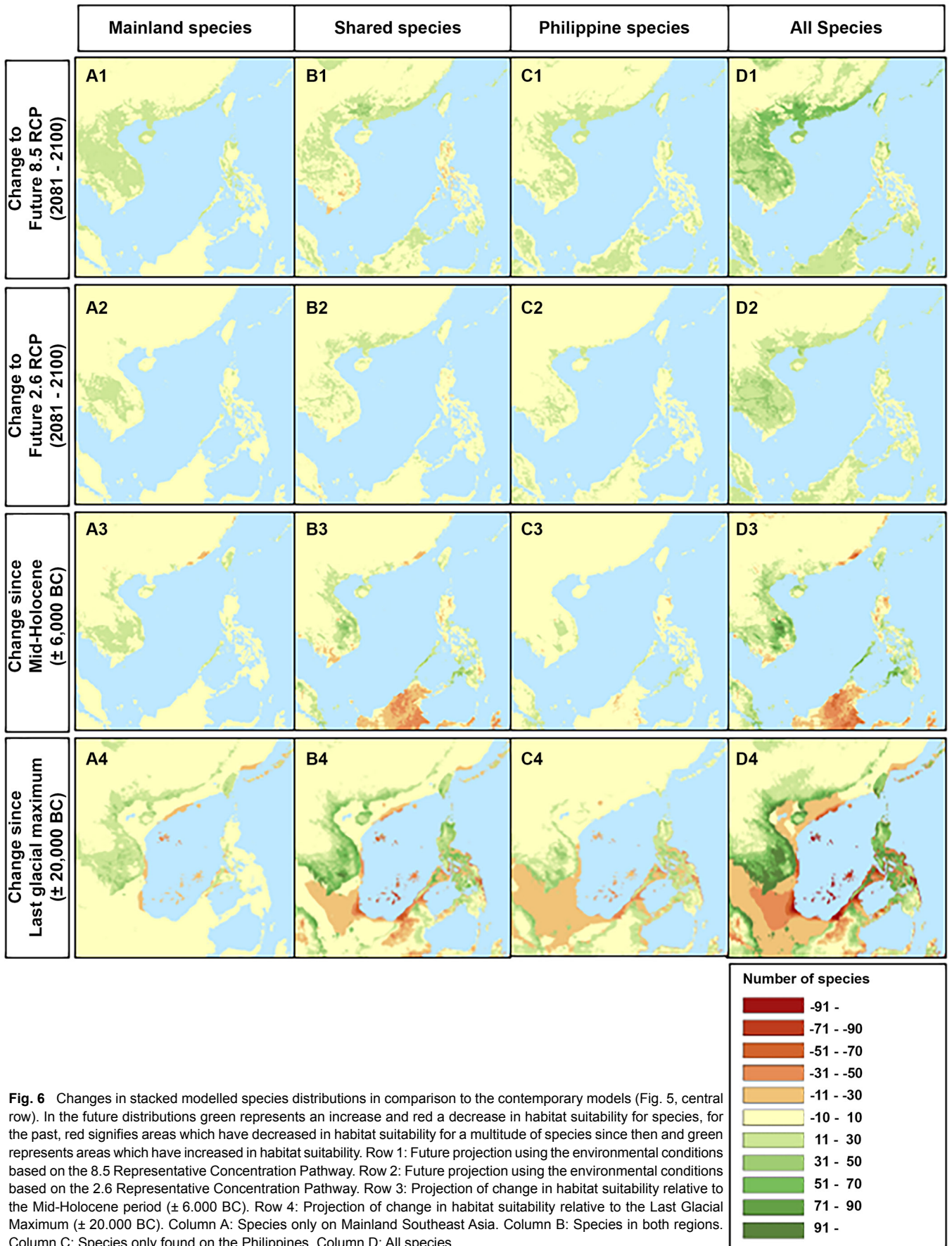
Fig. 6 Changes in stacked modelled species distributions in comparison to the contemporary models (Fig. 5, central row). In the future distributions green represents an increase and red a decrease in habitat suitability for species, for the past, red signifies areas which have decreased in habitat suitability for a multitude of species since then and green represents areas which have increased in habitat suitability. Row 1: Future projection using the environmental conditions based on the 8.5 Representative Concentration Pathway. Row 2: Future projection using the environmental conditions based on the 2.6 Representative Concentration Pathway. Row 3: Projection of change in habitat suitability relative to the Mid-Holocene period (± 6.000 BC). Row 4: Projection of change in habitat suitability relative to the Last Glacial Maximum (± 20.000 BC). Column A: Species only on Mainland Southeast Asia. Column B: Species in both regions. Column C: Species only found on the Philippines. Column D: All species.
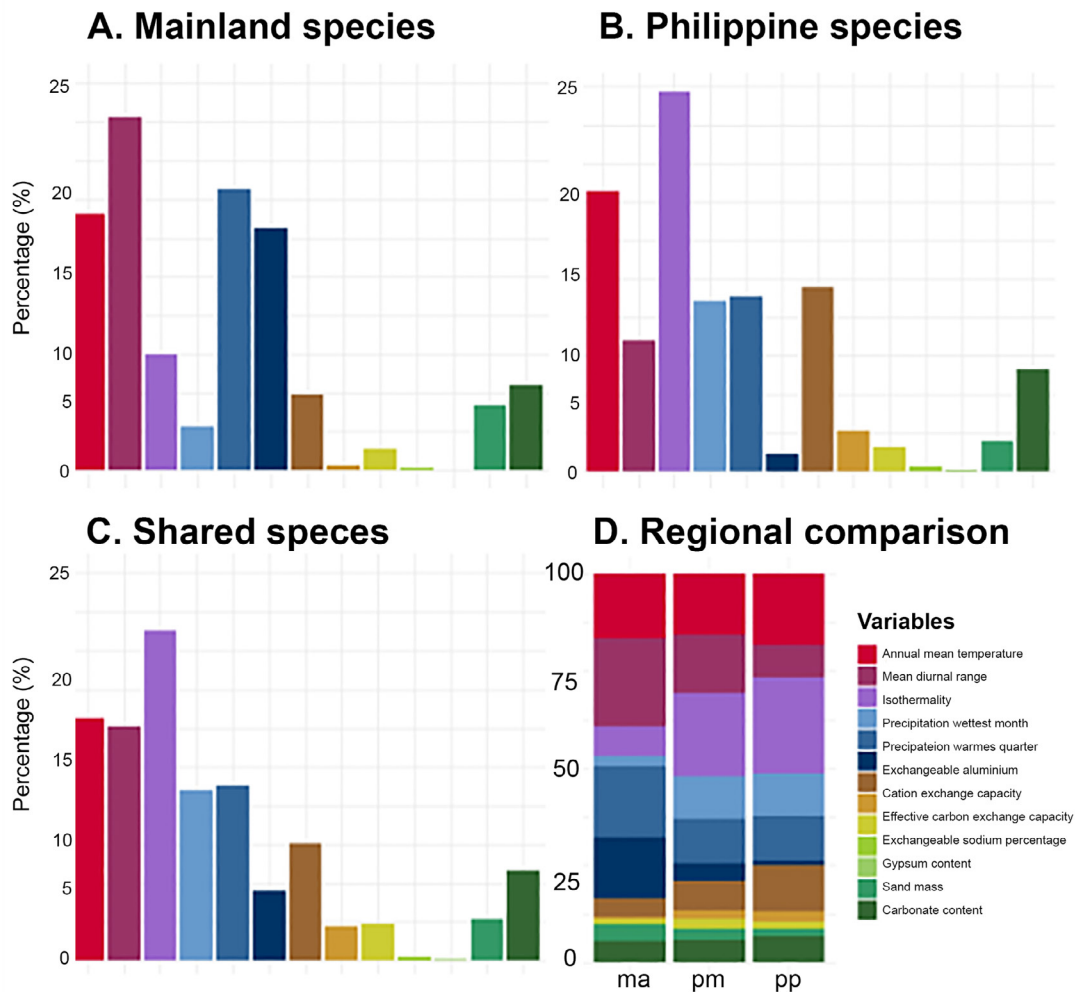
Fig. 7 The average contribution per variable for modelling the species of each regional group in percentages. A. The average contribution for all variables concerning species only occurring on Mainland Southeast Asia; B. the average contribution for all variables concerning species only occurring in the Philippines; C. the average contribution for all variables concerning species both occurring on the mainland and in the Philippines; D. the contributions as stacked percentages for all three regional groups; ma = SE Asian mainland species, pm = shared species, pp = Philippine species.

Influence of the variables

The results from MaxEnt show that the contribution of the climatic variables in the species distribution models outweigh those of the edaphic variables (Fig. 7). For all species distribution models, the climatic variables contribute to 79.097 % of the models, whilst the edaphic variables contribute only 20.903 % to the distributions. In the overall modelling, it is solely the 'Exchangeable aluminium' that has a similar importance comparable to the lower contributing climatic variables, namely the 'Precipitation of the wettest month' and 'Precipitation of warmest quarter'. For region specific groups, 'Sand mass' and 'Carbonate content' outweigh some climatic variables, yet their contributions are less than 5 % and 10 %, respectively, across the regions.

Table 9 shows the results of the Kruskal-Wallis method (Kruskal & Wallis 1952) with the Bonferroni adjustment (Bonferroni 1936), which was used to discern any significant differences between the contributions of each variable between the regions. The most interesting results are the relatively small number of significant differences between species that are shared between the mainland and the Philippines, and those only occurring in the Philippines. Notable is also the fact that the 'Annual Mean Temperature' does not differ greatly in its contribution to modelling the species distributions, but it is the second most important factor for explaining the distributions.

DISCUSSION

The distribution of the records are uneven (Fig. 3) as China shows a very limited number of records compared to the other countries. This bias results in distributions being more dominated by the Philippine data, because the Philippines are smaller in surface area and more homogeneous climatologically and because of the many more records held in L. The same applies to Mindanao, the southernmost Philippine island of the larger islands, rich in vegetation. Here less collecting occurred due to religious uproar, hence, less dark green colouring in Fig. 5. The bias in collections has certainly influenced the modelling, showing less possibilities for SE Asian mainland species in the Philippines (degree of influence unknown because including the GBIF data set, see Material and Methods, was too complicated), but with the general data, which uses POWO (POWO 2022) distributions, it is obvious that not many species occur on the SE Asian mainland and the Philippines.

General data

Table S1 provides a rough overview of the flora of the SE Asian mainland (Cambodia, part of China, Taiwan, Vietnam) and of the flora of the Philippines. The purpose is to see if it is likely that taxa dispersed from the SE Asian mainland to the Philippines and from there further into Malesia. The higher resemblance in families and genera (Table 4) shows that most migration occurred on ancestral levels and thus not recently. This is supported by the high numbers of endemic species

(here simply regarded as newly evolved species, but they can also be relics of wider distributions) in the Philippines, which show that a high level of local speciation was present. There are not many species distributions with a SE Asian mainland–Philippine connection without presence in the Malay Peninsula, 625 species, of which 181 are widespread up to New Guinea/Australia and may even have originated there. The remaining 444 may indicate a pathway between the SE Asian mainland and the Philippines (both ways), though many are restricted to Taiwan and the Philippines. The latter show that a floral exchange between these two areas occurred regularly, but it is unknown in which period(s). There are species distributions with presence in the SE Asian mainland, the Malay Peninsula and the Philippines (no other areas); it may then well be that these species dispersed to the Philippines and the Malay Peninsula independently (or if originating in either of the two areas, spread via the SE Asian mainland to the other area). The southern route, via the Malay Peninsula (Fig. 1 purple arrow), is more likely than via the Philippines, as there are more similar climates, either everwet or the savannah corridor (see below), then between the Philippines and the often colder and drier SE Asian mainland.

Phylogenetic data

Of all thousands of species checked, only ten species (as part of dated phylogenies) remained that showed a northern route towards the Philippines. More than 100 records (not shown), which also contained Philippine species, did not show evidence for a northern route, but likely showed a southern dispersal route to the Philippines via other Malesian islands. It is obvious that many more phylogenetic and dated biogeographic analyses are needed. Unfortunately, due to the low number of cases no general conclusion can be drawn, but a few possibilities can be discussed.

All mean divergence times for the species are of an age of 4.5 My or younger (Table 6; Fig. 2A). Therefore, dispersal via rafting on Palawan is unlikely as Palawan reached close proximity to Borneo and the rest of the Philippines around 15 Ma (Blackburn et al. 2010, and references therein; Brown et al. 2013, 2016).

Most mean divergence times were in the Pliocene during high water levels (Fig. 2B), only the species at the boundary of the Pliocene-Pleistocene and during the Pleistocene may have profited from lower sea levels in their dispersal, because the sea barriers were narrower. Nonetheless, sea barriers always had to be crossed (Heany 1985), likely via stepping stones, for instance the Batan and Babuyan Islands between Taiwan and the Philippines. The winter monsoon that comes from the northeast brings relatively dry and cool air towards the Philippines annually and can potentially transport diaspores over large distances (Wu et al. 2009). Additionally, Passey et al. (2009) suggest that the winter monsoon was stronger between 4 and 2 Mya, which coincides with the majority of the found mean divergence times.

Table 6 also shows the various fruit types and fruit colours. The fruits of most species are small drupes and likely eaten and transported over perhaps longer distances by birds and bats (Dardick & Callahan 2014). The bird dispersal syndrome (Van der Pijl 1982) shows that birds prefer fruits with red colours (often contrasting with dark seeds) and they may eat the fruits of *Viburnum* and the bat dispersal syndrome (Van der Pijl 1982) indicates that bats may like the white fruits of *Chloranthus* (contrasting with the dark night), which leaves a question mark for *Liriope* and *Ophiopogon* (black and blue, respectively). The arillate seeds of *Stemona tuberosa* are known to be dispersed by hornets. After the capsule has dehisced, seeds remain connected to the plant until hornets bite them off and transport

them to another location (Chen et al. 2017). Although it seems unlikely that hornets have dispersed seeds directly across the Luzon strait, potentially they could have used the islands within the strait as stepping-stones. Most of the species occur on higher altitudes (Table 6), making wind, stronger on higher altitudes, more likely as a means of long distance dispersal, conforming to the montane plant dispersal route via Taiwan, Luzon and Mindoro as mentioned in Van Steenis (1964: one of the tracks in f. 11). Also, the Batan and Babuyan Islands are volcanic islands and have high peaks (up to 1 000 and 1 100 m, respectively). Thus, chance long distance dispersal is always a possibility in especially the dispersal of montane species.

Ito et al. (2017) suggest that *Sedum*, growing on the seacoast, can be dispersed via sea currents, only the current in the Luzon strait is westwards directed (making dispersal from the SE Asian mainland and Taiwan unlikely), while the Kuroshio sea current moves north towards the Ryukyu islands, away from the Philippines. The Kuroshi sea current gained its present position c. 3 Ma (Gallagher et al. 2015), which is twice as old as the divergence moment in the *Sedum* phylogeny (Table 6).

Species Distribution Modelling

SDMs only indicate where the abiotic conditions are suitable for species. The influence of biotic factors (e.g., diaspore type and size, herbivores, pollinators, dispersers) and especially human land use pose their restrictions on the distributions and especially on the unlimited dispersal that modelling software allows. Thus the actual distributions are likely smaller and the chances of dispersal are much lower and/or over shorter distances due to human-made obstacles (e.g., roads), and will continue to decrease with increasing human influences. Overestimations in the models become more relevant when projecting species distributions in the past or future, then changes in the distributions, especially extensions of areas, are likely exaggerated, and can even hide actual decreases in distributions. The fact that especially climate influences the models, and far less the soil variables (Fig. 7), justifies the predictions for future and past models with adjusted climate variables.

The large number of records will incorporate several errors, either in the georeferenced locations or in the taxonomic nomenclature, moreover many of the records are slightly antiquated. Similar with Table 7, due to the inability to find coordinates for all specimens, it happens that species are not listed for areas where they occur (e.g., *Acalypha australis* L. absent from the Philippines). The large numbers of taxa and the ten-percentile threshold as it is included in MaxEnt, and as used in this study, should reduce the impact of these artefacts on the final distributions. Also, the stacking of the individual species distributions should mitigate the occasional incorrect data.

Currently, there is a large debate on whether research should standardize the use of ensemble model algorithms rather than single model methods as used here. Though ensemble models tend to be slightly more reliable, the quality difference in results seems limited (Kaky et al. 2020). Moreover, Kamworapan & Surussavadee (2019) show that large ensembles of algorithms could actually be worse due to the inclusion of less accurate models. Therefore, it was opted to solely implement MaxEnt, which is generally considered the best program for presence-only data (Elith et al. 2006).

For the species endemic to the Philippines (Fig. 5: C3) there are also suitable habitats in the more coastal areas of southern Vietnam and Cambodia, areas the plants never dispersed too. Likewise, the plants endemic to SE Asian mainland (Fig. 5: A3) could have found suitable habitats on Palawan and SW Luzon. All of these are the more everwet areas. Table 7, and tested significantly different in Table 8, shows that most of these

endemic species get models in which they occur at both sides of the South China Sea. Thus, many species cannot easily disperse over the South China Sea, also not during glacial periods (Fig. 6, row 5) when landmasses were at a maximum. Only the species present at both sides of the South China Sea (Fig. 5: B3) colour more areas dark green, also on Taiwan; these are likely the species that are adapted to a yearly dry monsoon.

The General Data showed that there are a few species shared between Taiwan and the Philippines. For the endemic Philippines species Taiwan is generally not part of the model (Fig. 5: C3) except for the southernmost point of it (where they do not occur in reality). For species endemic to the mainland, Taiwan is more often part of the model, shared with the wetter southern areas of the mainland. The models of the widespread species most often include Taiwan, something which is likely true for the species distributions in reality (see Table S1). During the glacial maximum, when distances across water were shortest, the species modelled barely showed any overlap between the northern Philippines and Taiwan; seemingly the climatic conditions for dispersal were too poor (Fig. 7, 8; bottom rows).

In the future, higher temperatures and more rain are expected, which will create an everwet climate in more northern regions. As a result models based on RCP 2.6 and 8.5 mainly indicate an increase in biodiversity, especially on the mainland (Fig. 8: D1, D2) with an exception for a few small areas in southern Vietnam. The highest losses in biodiversity will be in the species at both sides of the China Sea (shared species), under RCP 2.6 mainly on Luzon, the most northern large island of the Philippines, and under RCP 8.5 on Luzon and Palawan in the Philippines, and along the coast of S Vietnam and Cambodia. Under both scenarios it seems likely that many species will migrate northwards. However, this will be a slow process and likely increased human land use may have a more direct impact on species distributions, thus regional efforts might be best directed to counter this threat first.

Van Steenis (1979) considered the SE Asian mainland–Philippine connection especially important for dispersal of species preferring a yearly dry monsoon. However, all data above show that some species may have dispersed from SE Asian mainland to the Philippines (or vice versa), but not particularly all species that prefer a dry monsoon. Unknown to Van Steenis then, is the present knowledge of the occurrence of a savannah corridor, running from the Gulf of Thailand to the Java Sea during glacial periods when the Sunda Shelf became dry (thus uniting the Malay Peninsula, Sumatra, Borneo and NE Java into one area). This corridor with a much dryer climate is the direct result of this large land mass, comparable with Australia, dry in the middle, wet at the edges. The savannah corridor was likely a better dispersal route to the south for taxa occurring north of the Malay-Thai border (border of the Malesian area) at present and used to a dry period per year (prolonging with increasing latitudes), as no sea barriers had to be crossed and because it had an excellent climate for periodically dry monsoon loving species (e.g., see Cannon et al. 2009, Slik et al. 2011 for reconstructions).

CONCLUSIONS

Van Steenis (1964, 1979) expected that the Philippines would be a gateway for montane plants and for species preferring a yearly dry monsoon to disperse from the SE Asian mainland to Malesia. However, the patterns found indicate that it seems to have happened only in a limited number of cases. The dispersal of species in the first group could have occurred, as Van Steenis (1964) described, by long-distance dispersal between Asia and the Philippines. Research on the origin of Mt Kinabalu

(N Borneo; Merckx et al. 2015) confirms this (next to recruitment from local lowland ancestral species). For dispersal of the latter group, the savannah corridor during glacial periods was likely far more important as no sea barriers had to be crossed.

In the future, especially under the worst climate scenario, RCP 8.5, the coastal areas of China and Taiwan will become more suitable for many species due to more rain and higher temperatures, but as the weather will become extremer and human influence is still increasing, it is unlikely that plant species will find a suitable (long distance) dispersal pathway between the Philippines and the SE Asian mainland.

REFERENCES

Aguirre-Gutiérrez J, Carvalheiro LG, Polce C, et al. 2013. Fit-for-purpose: species distribution model performance depends on evaluation criteria – Dutch hoverflies as case study. PloS ONE 8(5): 363708. https://doi:10.1371/journal.pone.0063708.

APG IV. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Botanical Journal of the Linnean Society 181(1): 1–20. https://doi.org/10.1111/boj.12385.

Araújo MB, Peterson AT. 2012. Uses and misuses of bioclimatic envelope modelling. Ecology 93: 1527–1539.

Atkins HJ, Bramley GLC, Johnson MA, et al. 2019. A molecular phylogeny of Southeast Asian Cyrtandra (Gesneriaceae) supports an emerging paradigm for Malesian plant biogeography. Frontiers of Biogeography 12(1): e44814.

Batjes NH. 2012. ISRIC-WISE derived soil properties on a 5 by 5 arc-minutes global grid (ver. 1.2). ISRIC report 2012/01. ISRIC-World Soil Information, Wageningen.

Bentvelzen PAJ. 1960. Primulaceae. In: Van Steenis CGGJ (ed), Flora Malesiana Ser. I, 6: 173–192. Noordhoff-Kolff, Batavia.

Blackburn DC, Bickford DP, Diesmos AC, et al. 2010. An ancient origin for the enigmatic flat-headed frogs (Bombinatoridae: Barbourula) from the Islands of Southeast Asia. PLoS ONE 5: e12090.

Bonferroni CE. 1936. Teoria statistica delle classi e calcolo delle probabilit. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze 8: 3–62.

Broussard MC. 2007. A horticultural study of Liriope and Ophiopogon: nomenclature, morphology, and culture. Louisiana State University Doctoral Dissertations, 4002. https://digitalcommons.lsu.edu/gradschool_dissertations/4002 [last checked Aug. 2022].

Brown RM, Siler CD, Oliveros CH, et al. 2013. Evolutionary processes of diversification in a model island archipelago. Annual Review of Ecology, Evolution, and Systematics 44: 411–435.

Brown RM, Su Y-C, Barger B, et al. 2016. Phylogeny of the island archipelago frog genus Sanguirana: Another endemic Philippine radiation that diversified 'Out-of-Palawan'. Molecular Phylogenetics and Evolution 94: 531–536.

Cannon CH, Morley RJ, Bush ABG. 2009. The current ferugial rainforests of Sundaland are unrepresentative of their biogeographic past and highly vulnerable to disturbance. Proceedings of the National Academy of Sciences of the United States of America 106(27): 11188–11193.

Cayuela L, Macarro I, Stein A, et al. 2021. Automated standardization of taxonomic names and removal of orthographic errors in plant species names using 'The Plant List' website. Package version 2.4. https://CRAN.R-project.org/package=Taxonstand; succeeded by WorldFlora package.

Chen G, Wang Z-W, Qin Y, et al. 2017. Seed dispersal by hornets: An insect-plant interaction. Journal of Integrative Plant Biology 59: 792–796.

Chen Y-S, Zeng C-X, Muellner-Riehl AN, et al. 2021. Invertebrate-mediated dispersal plays an important role in shaping the current distribution of a herbaceous monocot. Journal of Biogeography 48: 1101–1111.

Choi SS, Kim J, Kim CH. 2020. The taxonomic entity and distribution of Korean Sedum formosanum (Crassulaceae) revealed in 133 years. Korean Journal of Plant Taxonomy 50: 377–384.

Co's digital Flora of the Philippines (CDFP). Continuously updated. https://www.philippineplants.org/index.html [last checked 7 Aug. 2023].

Corlett RT, Primack RB. 2011. Tropical rain forests: An ecological and bio-geographical comparison, ed. 2. Wiley-Blackwell, Chichester.

Crayn DM, Costion C, Harrington MG. 2014. The Sahul-Sunda floristic exchange: dated molecular phylogenies document Cenozoic intercontinental dispersal dynamics. Journal of Biogeography 42: 11–24. https://doi. org/10.1111/jbi.12405.

Cullen J. 2013. 757. Chloranthus oldhamii, Chloranthaceae. Curtis's Botanical Magazine 30: 68–75.

Dardick C, Callahan AM. 2014. Evolution of the fruit endocarp: molecular mechanisms underlying adaptions in seed protection and dispersal strategies. Frontiers in Plant Science 5: 284

De Boer AJ. 1995. Island and cicadas adrift in the West Pacific. Biogeographic patterns related to plate tectonics. Tijdschrift voor Entomologie 138: 169–244.

De Boer B, Lourens LJ, Van de Wal RSW. 2014. Persistent 400,000-year variability of Antarctic ice volume and the carbon cycle is revealed throughout the plio-pleistocene. Nature communications 5: 2999.

Dickerson RE. 1928. Distribution of life in the Philippines. Monographs of the Bureau of Science, Manila 2: 1–322.

Duyfjes BEE. 1993. Stemonaceae. In: De Wilde WJJO (ed), Flora Malesiana Ser. I, 11: 399–409. Rijksherbarium/Hortus Botanicus, Leiden.

Elith J, Graham CH, Anderson RP, et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29:129–151.

Elith J, Phillips SJ, Hastie T, et al. 2010. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 7: 43–57.

ESRI. 2018. ArcGIS Desktop: Release 10.6.1. Environmental Systems Research Institute, Redlands, California (https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview [last checked Sept. 2022].

Esselstyn JA, Oliveros CH. 2010. Colonization of the Philippines from Taiwan: a multi-locus test of the biogeographic and phylogenetic relationships of isolated populations of shrews. Journal of Biogeography 37: 1504–1514.

Gallagher SJ, Kitamura A, Iryu Y, et al. 2015. The Pliocene to recent history of the Kuroshio and Tsushima currents: a multi-proxy approach. Progress in Earth and Planetary Science 2(17): 1–23.

Graham CH, Elith J, Hijmans RJ, et al. 2008. The influence of spatial errors in species occurrence data used in distribution models. Journal of Applied Ecology 45: 239–247.

Hall R. 2009. Southeast Asia's changing palaeogeography. Blumea 54: 148–161. https://doi.org/10.3767/000651909X475941.

Hall R. 2012. Sundaland and Wallacea: Geology, plate tectonics and palaeogeography. In: Gower D, Johnson K, Richardson J, et al. (eds), Biotic Evolution and Environmental Change in Southeast Asia (Systematics Association Special Volume Serie): 32–78. Cambridge University Press, Cambridge.

Hall R. 2013. The palaeogeography of Sundaland and Wallaca since the Later Jurassic. Journal of Limnology 72(s2): 117.

Heaney LR. 1985. Zoogeographic evidence for middle and late Pleistocene land bridges to the Philippine islands. Modern Quaternary Research in Southeast Asia 9: 127–143.

Hernandez PA, Graham CH, Master LL, et al. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29: 773–785.

Inthachub P, Vajrodaya S, Duyfjes BEE. 2010. Census of Stemona (Stemonaceae) in Thailand. Blumea 55: 143–152.

IPCC - Intergovernmental Panel on Climate Change. 2013. Contribution of working group I to the fifth assessment report of the Intergovernmental Panel on Climate Change. In: Stocker TF, Qin D, Plattner GK, et al. (eds), Climate Change 2013. The Physical Science Basis. Cambridge University Press, Cambridge.

IPCC - Intergovernmental Panel on Climate Change. 2022. In: Pörtner H-O, Roberts DC, Tignor, et al. (eds), Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, New York. https://doi.org/10.1017/9781009325844. https://www.ipcc.ch/report/ar6/wg2/ [last visited Aug. 2022].

Ito T, Yu C-C, Nakamura K, et al. 2017. Unique parallel radiations of high-mountainous species of the genus Sedum (Crassulaceae) on the continental island of Taiwan. Molecular Phylogenetics and Evolution 113: 9–22.

Jacobs B, Donoghue MJ, Bouman F, et al. 2008. Evolution and phylogenetic importance of endocarp and seed characters in Viburnum (Adoxaceae). International Journal of Plant Sciences 169: 409–431.

Kaky E, Nolana V, Alatawia A, et al. 2020. A comparison between Ensemble and MaxEnt species distribution modelling approaches for conservation: A case study with Egyptian medicinal plants. Ecological Informatics 60: 101150.

Kamworapan S, Surussavadee C. 2019. Evaluation of CMIP5 Global Climate Models for simulating climatological temperature and precipitation for Southeast Asia. Advances in Meteorology 2019: 1067365.

Karney CFF. 2013. Algorithms for geodesics. Journal of Geodesy 87: 43–55.

Kottek M, Grieser J, Beck C, et al. 2006. World Map of the Köppen-Geiger climate classification updated. Meteorologische Zeitschrift 15: 259–263.

Kruskal WH, Wallis WA. 1952. Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association 47: 583–621.

Landis MJ, Eaton DAR, Clement WL, et al. 2021. Joint phylogenetic estimation of geographic movements and biome shifts during the global diversification of Viburnum. Systems Biology 70: 67–85.

Matuszak S, Muellner-Riehl AN, Sun H, et al. 2016. Dispersal routes between biodiversity hotspots in Asia: the case of the mountain genus Tripterospermum (Gentianinae, Gentianaceae) and its close relatives. Journal of Biogeography 43: 580–590.

Merckx VSFT, Hendriks KP, Beentjes KK, et al. 2015. Evolution of endemism on a young tropical mountain. Nature 524: 347–350.

Morley RJ. 2000. Origin and evolution of tropical rain forests. Wiley, Chichester, New York.

Morley RJ. 2003. Interplate disperdal paths for megathermal angiosperms. Perspectives in Plant Ecology, Evolution and Systematics 6(1–2): 5–20.

Morley RJ, Flenley JR. 1987. Late Cainozoic vegetational and environmental changes in the Malay Archipelago. In: Whitmore TC (ed), Biogeographical evolution of the Malay Archipelago: 50–59. Clarendon Press, Oxford.

Nesom GL. 2010. Overview of Liriope and Ophiopogon (Ruscaceae) naturalized and commonly cultivated in the USA. Phytoneuron 56: 1–31.

O'Neill BC, Tebaldi C, Van Vuuren DP, et al. 2016. The scenario model intercomparison project (ScenarioMIP) for CMIP6. Geoscientific Model Development 9: 3461–3482.

Oh I-C, Anderberg A-I, Schönenberger J, et al. 2008. Comparative seed morphology and character evolution in the genus Lysimachia (Myrsinaceae) and related taxa. Plant Systematics and Evolutions 271: 177–197.

Passey BH, Ayliffe LK, Kaakinen A, et al. 2009. Strengthened East Asian summer monsoons during a period of high-latitude warmth? Isotopic evidence from Mio-Pliocene fossil mammals and soil carbonates from northern China. Earth and Planetary Science Letters 277: 443–452.

Peterson AT, Papeş M, Eaton M. 2007. Transferability and model evaluation in ecological niche modelling: a comparison of GARP and Maxent. Ecography 30: 550–560.

Phillips SJ. 2006. A brief tutorial on Maxent. AT&T Research (https://biodiversityinformatics.amnh.org/open_source/maxent/Maxent_tutorial2017.pdf [last checked Sept. 2022).

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190: 231–259.

Phillips SJ, Dudik M, Schapire RE. 2005. Maxent software for species distribution modelling. http://www.cs.princeton.edu/~schapire/maxent/, nowadays: http://rob.schapire.net/ [checked Feb. 2023]

POWO - Plants of the World Online. 2022. Facilitated by the Royal Botanic Gardens, Kew. http://www.plantsoftheworldonline.org/ [last checked March 2022].

Raes N, Cannon CH, Hijmans RJ, et al. 2014. Historical distribution of Sundaland's Dipterocarp rainforests at Quaternary glacial maxima. Proceedings of the National Academy of Sciences USA 111: 16790–16795.

Raes N, Ter Steege H. 2007. A null-model for significance testing of presence-only species distribution models. Ecography 30: 727–736.

Raes N, Van Welzen PC. 2009. The demarcation and internal division of Flora Malesiana: 1857 – present. Blumea 54: 6–8. https://doi.org/10.3767/000651909X475888.

Roos MC. 1992. State of affairs regarding Flora Malesiana: progress in revision work and publication schedule. Flora Malesiana Bulletin 11: 133–142.

Rstudio Team. 2020. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA. https://posit.co/download/rstudio-desktop/ [last checked Aug. 2022].

Rumpf SB, Hülber K, Zimmermann NE, et al. 2019. Elevational rear edges shifted at least as much as leading edges over the last century. Global Ecology and Biogeography 28: 533–543.

Rutgrink AL, Visser M, Van Welzen PC. 2018. Differences between the floras of the North and South Moluccas (Indonesia). Journal of Systematics and Evolution 56: 652–662.

Scherrer D, Massy S, Meier S, et al. 2017. Assessing and predicting shifts in mountain forest composition across 25 years of climate change. Diversity and Distributions 23: 517–528.

Sheldeman X, Van Zonneveld M. 2010. Training manual on spatial analysis of plant diversity and distribution. Biodiversity International, Rome.

Slik JWF, Aiba SI, Bastian M, et al. 2011. Soils on exposed Sunda Shelf shaped biogeographic patterns in the equatorial forests of Southeast Asia. Proceedings of the National Academy of Sciences of the United States of America 108(30): 12343–12347.

Subilla MS, Baoanan ZG. 2021. Viburnum luzonicum Rolfe Viburnaceae Raf. In: Franco FM (ed), Ethnobotany of the Mountain Regions of Southeast Asia: 1107–1113. Springer Nature Switzerland AG, Cham.

Taylor KE, Stouffer RJ, Meehl GA. 2012. An overview of CMIP5 and the experiment design. Bulletin of the American Meteorological Society 93: 485–498

Thiers B. Continuously updated. Index Herbariorum: A global directory of public Herbaria and associated Staff. New York Botanical Garden's Virtual Herbarium, New York. Available from http://sweetgum.nybg.org/science/ih/ [last checked 22 Aug. 2022].

Van den Ende C, White LT, Van Welzen PC. 2017. The existence and break-up of the Antarctic land bridge as indicated by both amphi-Pacific distributions and tectonics. Gondwana Research 44: 219–227.

Van der Pijl L. 1982. Principles of dispersal in higher plants. Springer-Verlag, Berlin, Heidelberg, New York.

Van Proosdij ASJ, Sosef MSM, Wieringa JJ, et al. 2016. Minimum required number of specimen records to develop accurate species distribution models. Ecography 39: 542–552.

Van Rossum G, Drake Jr FL. 1995. Python reference manual. Centrum voor Wiskunde en Informatica Amsterdam.

Van Steenis CGGJ. 1950. The delimitation of Malaysia and its main plant geographical divisions. In: Van Steenis CGGJ (ed), Flora Malesiana Ser. I, 1: lxx–lxxv. Noordhoff Kolff n.v., Djakarta.

Van Steenis CGGJ. 1964. Plant geography of the Mountain Flora of Mt Kinabalu. Proceedings of the Royal Society of London, Ser. B, Biological Sciences 161: 7–38.

Van Steenis CGGJ. 1979. Plant-geography of east Malesia. Botanical Journal of the Linnean Society 79: 97–178.

Van Steenis CGGJ. 1982. Addenda, corrigenda et emendanda. In: Van Steenis CGGJ (ed), Flora Malesiana Ser. I, 9: 553–573. Nijhoff/Junk, The Hague.

Van Welzen PC, Parnell JAN, Slik JWF. 2011. Wallace's line and plant distributions: two or three phytogeographical areas and where to group Java? Biological Journal of the Linnean Society 103: 531–545.

Van Welzen PC, Slik JWF, Alahuhta J. 2005. Plant distribution patterns and plate tectonics in Malesia. In: Fries I, Balslev H (eds), Plant diversity and complexity patterns: Local, regional and global dimensions. Biologiske Skrifter 55: 199–217.

Verdcourt B. 1984. Chloranthaceae. In: Van Steenis CGGJ, De Wilde WJJO (eds), Flora Malesiana Ser. I, 10: 123–144. Noordhoff-Kolff, Batavia.

Walther GR, Post E, Convey P, et al. 2002. Ecological responses to recent climate change. Nature 416: 389–395.

Wang G-Y, Yang Y-P. 2018. Hypothesizing the origin, migration routes, and distribution patterns of Ophiopogon (Asparagaceae) in East and Southeast Asia. Journal of Systematics and Evolution 56: 194–201.

Wilkinson AM. 1948. Floral anatomy and morphology of some species of the genus Viburnum of the Caprifoliaceae. American Journal of Botany 35: 455–465.

Wisz MS, Hijmans RJ, Li J, et al. 2008. Effects of sample size on the performance of species distribution models. Diversity and Distributions 14: 763–773.

Wu M-J, Huang T-C, Huang S-F. 2009. Phylogenetic biogeography of Euphrasia section Malesianae (Orobanchaceae) in Taiwan and Malesia. Blumea 54: 242–247.

Yan H-F, Zhang C-Y, Anderberg AA, et al. 2018. What explains high plant richness in East Asia? Time and diversification in the tribe Lysimachieae (Primulaceae). New Phytologist 219: 436–448.

Zhang Q, Antonelli A, Feild TS, et al. 2011. Revisiting taxonomy, morphological evolution, and fosil calibration strategies in Chloranthaceae. Journal of Systematics and Evolution 49: 315–329.

Supplementary material

All to be found on https://repository.naturalis.nl/pub/800821.

Table S1 List of all species (checked against POWO 2022) present in L, U, WAG in SE Asia mainland (part of China, Vietnam, Taiwan, Cambodia) and the Philippines.

Table S2 List of all species (checked against POWO 2022) present in L, U, WAG in SE Asia mainland (part of China, Vietnam, Taiwan, Cambodia) and only the Philippines (no other Malesian areas).

Table S3 List of all families (checked against APG IV 2016) present in L, U, WAG in SE Asia mainland (part of China, Vietnam, Taiwan, Cambodia) and the Philippines.

Table S4 List of all genera (checked against POWO 2022) present in L, U, WAG in SE Asia mainland (part of China, Vietnam, Taiwan, Cambodia) and the Philippines.

Appendix Scripts used during Species Distribution Modelling.

1. Cropping and selecting the environmental variables in R

```
## bioclimatic variable selection

setwd("D:/Universiteit_Leiden/Stage_Naturalis/Data")
getwd()

#install.packages("sp")
#install.packages("rgdal")
#install.packages("raster")
#install.packages("sf")
#install.packages("usdm")

library(sp)
library(rgdal)
library(raster)
library(biomod2)
library(sf)
library(usdm)
```

```
# Loading present data
rst.namesB <- list.files ("./expl_var/present/wc2.1_5m_bio/", all.files = FALSE)
rst.fldB <- list.files("./expl_var/present/wc2.1_5m_bio/", all.files = FALSE, full.
names = T)
rst.fldB
stk.presentB <- stack(rst.fldB)

rst.namesE <- list.files ("./expl_var/present/wc2.1_5m_elev/", all.files = FALSE)
rst.fldE <- list.files("./expl_var/present/wc2.1_5m_elev/", all.files = FALSE, full.
names = T)
rst.fldE
stk.presentE <- stack(rst.fldE)

stk.present <- raster::stack(stk.presentB, stk.presentE)

# processing present data

names(stk.present)
names(stk.present) <- c("Bio01", "Bio02","Bio03","Bio04", "Bio05","Bio06",
"Bio07","Bio08","Bio09",
                "Bio010","Bio11","Bio12","Bio13", "Bio14","Bio15","Bio16",
"Bio17","Bio18", "Bio19", "Elev")

res(stk.present)

# Loading future data
#> r = raster("./GcrfPicture.tif")
#als je een stacked tiff wil binnenhalen dan doe je dit met Stack(name)

# path 2.6
stk.future26 = stack("./expl_var/future/wc2.1_5m_bioc_CNRM-CM6-1_
ssp126_2081-2100/wc2.1_5m_bioc_CNRM-CM6-1_ssp126_2081-2100.tif")
stk.future26
#class      : RasterStack
#dimensions : 720, 960, 691200, 4  (nrow, ncol, ncell, nlayers)

stk.fut.26 <- raster::stack(stk.future26, stk.presentE)

names(stk.fut.26)
names(stk.fut.26) <- c("Bio01", "Bio02","Bio03","Bio04", "Bio05","Bio06",
"Bio07","Bio08","Bio09",
                "Bio010","Bio11","Bio12","Bio13", "Bio14","Bio15","Bio16",
"Bio17","Bio18", "Bio19", "Elev")

# path 8.5
stk.future85 = stack("./expl_var/future/wc2.1_5m_bioc_CNRM-CM6-1_
ssp585_2081-2100/wc2.1_5m_bioc_CNRM-CM6-1-ssp585_2081-2100.tif")
stk.future85
#class      : RasterStack
#dimensions : 720, 960, 691200, 4  (nrow, ncol, ncell, nlayers)

stk.fut.85 <- raster::stack(stk.future85, stk.presentE)

names(stk.fut.85)
names(stk.fut.85) <- c("Bio01", "Bio02","Bio03","Bio04", "Bio05","Bio06",
"Bio07","Bio08","Bio09",
                "Bio010","Bio11","Bio12","Bio13", "Bio14","Bio15","Bio16",
"Bio17","Bio18", "Bio19", "Elev")
```

```
# Loading past data
# Holocene
stk.past.holo <- list.files ("./expl_var/past/cnrm-cm5_midholocone_5m/", all.
files = FALSE)
rst.fld.holo <- list.files("./expl_var/past/cnrm-cm5_midholocone_5m/", all.
```

```
files = FALSE, full.names = T)[c(1,12:19,2:11)]
rst.fld.holo
stk.past.holo <- stack(rst.fld.holo)

stk.past.holo
stk.past.E <- stk.presentE
stk.past.E <- crop(stk.past.E,extent(stk.past.holo))

stk.past.holoE <- raster::stack(stk.past.holo, stk.past.E)
stk.past.holoE

names(stk.past.holoE)
names(stk.past.holoE) <- c("Bio01", "Bio02","Bio03","Bio04", "Bio05","Bio06",
"Bio07","Bio08","Bio09",
                "Bio010","Bio11","Bio12","Bio13", "Bio14","Bio15","Bio16",
"Bio17","Bio18", "Bio19", "Elev")

# Glacial maximum
stk.past.glac <- list.files ("./expl_var/past/ccsm4_glacialmaximum_5m/", all.
files = FALSE)
rst.fld.glac <- list.files("./expl_var/past/ccsm4_glacialmaximum_5m/", all.
files = FALSE, full.names = T)[c(1,12:19,2:11)]
rst.fld.glac
stk.past.glac <- stack(rst.fld.glac)

stk.past.glacE <- raster::stack(stk.past.glac, stk.past.E)
stk.past.glacE

names(stk.past.glacE)
names(stk.past.glacE) <- c("Bio01", "Bio02","Bio03","Bio04", "Bio05","Bio06",
"Bio07","Bio08","Bio09",
                "Bio010","Bio11","Bio12","Bio13", "Bio14","Bio15","Bio16",
"Bio17","Bio18", "Bio19", "Elev")
```

```
# Loading soil data
# Loading soil data
rst.names_edaphic <- list.files ("./expl_var/present/soil_layers/tiff_layers/", all.
files = FALSE)
rst.fld_edaphic<- list.files("./expl_var/present/soil_layers/tiff_layers/", all.
files = FALSE, full.names = T)
rst.fld_edaphic
stk.edaphic <- stack(rst.fld_edaphic)

# processing soil data
names(stk.edaphic)
res(stk.edaphic)
plot(stk.edaphic)
```

```
# Bbox with coordinates file
range <- read.csv("./range.csv", header = T, sep = ";")
range_shp <- range
coordinates(range_shp) <- ~longitude+latitude

rr <- raster(range_shp)
crs(rr)
crs(rr) <- "+proj = longlat +ellps = WGS84 +datum = WGS84 +no_defs"
crs(rr)
w <- wkt (rr)
w
cat(w,"/n")

bbox <- extent(range_shp)

plot(stk.present$Bio01, ext = bbox)
sp_shp <- plot(bbox, col = 'blue', add = T)
PLOT <- plot(sp_shp, add = T, pch = 19, col = 'red',)

grid(PLOT, nx = NULL, col = "lightgray")

# cropping present data
stk.present.AOI.crop <- crop(stk.present,bbox)
par(mfrow = c(1,2)) #sets plotting area to a 1 line 2 columns set up
plot(stk.present$Bio01,main = "Original extent")
plot(stk.present.AOI.crop$Bio01, main  = "Cropped extent")
par(mfrow = c(1,1))

# cropping future data
stk.fut.26.AOI.crop <- crop(stk.fut.26,bbox)
par(mfrow = c(1,2)) #sets plotting area to a 1 line 2 columns set up
plot(stk.fut.26$Bio01,main = "Original extent 26")
plot(stk.fut.26.AOI.crop$Bio01, main  = "Cropped extent 26")
par(mfrow = c(1,1))
```

```
stk.fut.85.AOI.crop <- crop(stk.fut.85,bbox)
par(mfrow = c(1,2)) #sets plotting area to a 1 line 2 columns set up
plot(stk.fut.85$Bio01,main = "Original extent 85")
plot(stk.fut.85.AOI.crop$Bio01, main  = "Cropped extent 85")
par(mfrow = c(1,1))

# cropping past data
stk.past.holo.AOI.crop <- crop(stk.past.holoE,bbox)
par(mfrow = c(1,2)) #sets plotting area to a 1 line 2 columns set up
plot(stk.past.holoE$Bio01,main = "Original extent H")
plot(stk.past.holo.AOI.crop$Bio01, main  = "Cropped extent Holo")
par(mfrow = c(1,1))

stk.past.glac.AOI.crop <- crop(stk.past.glacE,bbox)
par(mfrow = c(1,2)) #sets plotting area to a 1 line 2 columns set up
plot(stk.past.glacE$Bio01,main = "Original extent")
plot(stk.past.glac.AOI.crop$Bio01, main  = "Cropped extent Glac")
par(mfrow = c(1,1))

# cropping edaphic data
stk.edaphic.AOI.crop <- crop(stk.edaphic,bbox)
par(mfrow = c(1,2)) #sets plotting area to a 1 line 2 columns set up
plot(stk.edaphic$gyps,main = "Original extent")
plot(stk.edaphic.AOI.crop$gyps, main = "Cropped extent")
par(mfrow = c(1,1))
```

```
### Autocorrelation testing ###

## Pairwise testing
#convert cropped raster to a dataframe
stk.present_bee_cor <- na.omit(as.data.frame(stk.present_bee))

# store Pearson correlation in a matrix
cor.tab <- cor(stk.present_bee_cor)
# make csv
write.csv2(cor.tab,"CorrelationTable_AOI_bee.csv")

### VIF test for multicollinearity ###

#edited data files

head(stk.present_bee_cor)
stk.present_bee_cor_sel <- stk.present_bee_cor[,c("Bio01", "Bio05",
"Bio09","Bio010", "Bio12", "Bio13", "alsa", "bsat", "cecs", "cnrt", "ecec", "esp",
"gyps", "SDTO", "tceq")]
vifcor(stk.present_bee_cor_sel, maxobservations = nrow(stk.present_bee_
cor_sel))

#var large: bsat, cecs, cnrt, sdto are large (cnrt largest removal)

stk.present_bee_cor_sel2 <- stk.present_bee_cor[,c("Bio01", "Bio05",
"Bio09","Bio010", "Bio12", "Bio13", "alsa", "bsat", "cecs",  "ecec", "esp", "gyps",
"SDTO", "tceq")]
vifcor(stk.present_bee_cor_sel2, maxobservations = nrow(stk.present_bee_
cor_sel2))

# var bsat >10 (removal)

stk.present_bee_cor_sel3 <- stk.present_bee_cor[,c("Bio01", "Bio05",
"Bio09","Bio010", "Bio12", "Bio13", "alsa", "cecs", "ecec", "esp", "gyps", "SDTO",
"tceq")]
vifcor(stk.present_bee_cor_sel3, maxobservations = nrow(stk.present_bee_
cor_sel3))

# final naming
stk.present_bee_sel <- stk.present_bee_cor[,c("Bio01", "Bio05", "Bio09","Bio010",
"Bio12", "Bio13", "alsa", "cecs", "ecec", "esp", "gyps", "SDTO", "tceq")]
```

```
# Stack AOI bio files with AOI soil (based on variables after multicollinearity
testing plus VIF)

# present
```

```
stk.present_bee <- raster::stack(stk.present.AOI.crop, stk.edaphic.AOI.crop)
names(stk.present_bee)

# future
stk.fut.26_bee <- raster::stack(stk.fut.26.AOI.crop, stk.edaphic.AOI.crop)
names(stk.fut.26_bee)
stk.fut.26_bee_cor <- na.omit(as.data.frame(stk.fut.26_bee))
stk.fut.26_bee_sel <- stk.fut.26_bee_cor[,c("Bio01", "Bio05", "Bio09","Bio010",
"Bio12", "Bio13", "alsa", "cecs", "ecec", "esp", "gyps", "SDTO", "tceq")]

stk.fut.85_bee <- raster::stack(stk.fut.85.AOI.crop, stk.edaphic.AOI.crop)
names(stk.fut.26_bee)
stk.fut.85_bee_cor <- na.omit(as.data.frame(stk.fut.85_bee))
stk.fut.85_bee_sel <- stk.fut.85_bee_cor[,c("Bio01", "Bio05", "Bio09","Bio010",
"Bio12", "Bio13", "alsa", "cecs", "ecec", "esp", "gyps", "SDTO", "tceq")]

# past
stk.past.holo_bee <- raster::stack(stk.past.holo.AOI.crop, stk.edaphic.AOI.crop)
names(stk.past.holo_bee)
stk.past.holo_bee_cor <- na.omit(as.data.frame(stk.past.holo_bee))
stk.past.holo_bee_sel <- stk.past.holo_bee_cor[,c("Bio01", "Bio05",
"Bio09","Bio010", "Bio12", "Bio13", "alsa", "cecs", "ecec", "esp", "gyps", "SDTO",
"tceq")]

stk.past.glac_bee <- raster::stack(stk.past.glac.AOI.crop, stk.edaphic.AOI.crop)
names(stk.past.glac_bee)
stk.past.glac_bee_cor <- na.omit(as.data.frame(stk.past.glac_bee))
stk.past.glac_bee_sel <- stk.past.glac_bee_cor[,c("Bio01", "Bio05",
"Bio09","Bio010","Bio12", "Bio13", "alsa", "cecs", "ecec", "esp", "gyps", "SDTO",
"tceq")]

# dataframes into rasters

stk.present_bee_sel <- as.raster(stk.present_bee_sel)
stk.fut.26_bee_sel
stk.fut.85_bee_sel
stk.past.holo_bee_sel
stk.past.glac_bee_sel
```

```
# save files
writeRaster(stk.present_bee,
        "./expl_var/maxent_ascii/present_ascii/stk.present_bee_sel.asc",
        overwrite = T,
        bylayer = T,
        suffix = "names")
#future
writeRaster(stk.fut.26_bee,
        "./expl_var/maxent_ascii/future_ascii/26_ascii/stk.fut.26_bee_sel.asc",
        overwrite = T,
        bylayer = T,
        suffix = "names")

writeRaster(stk.fut.85_bee,
        "./expl_var/maxent_ascii//future_ascii/85_ascii//stk.fut.85_bee_sel.asc",
        overwrite = T,
        bylayer = T,
        suffix = "names")
#past
writeRaster(stk.past.holo_bee,
        "./expl_var/maxent_ascii/past_ascii/holo_ascii/stk.past.holo_bee_sel.asc",
        overwrite = T,
        bylayer = T,
        suffix = "names")

writeRaster(stk.past.glac_bee,
        "./expl_var/maxent_ascii/past_ascii/glac_ascii/stk.past.glac_bee_sel.asc",
        overwrite = T,
        bylayer = T,
        suffix = "names")
```

....

2. Creating binary maps with the ten percentile threshold in Python

```
import sys, os, string

os.chdir("C:\\Program Files (x86)\\ArcGIS\\Desktop10.6\\arcpy\\arcpy")

print(sys.path)

sys.path.append("C:\\Program Files (x86)\\ArcGIS\\Desktop10.6\\arcpy")
sys.path.append("C:\\Program Files (x86)\\ArcGIS\\Desktop10.6\\ArcToolbox\\
Scripts")
sys.path.append("C:\\Program Files (x86)\\ArcGIS\\Desktop10.6\\bin")

print(sys.path)
```

```
import arcpy

# Load species_threshold file

report = open('D:\\Universiteit_Leiden\\Stage_naturalis\\Data\\maxent_results\\
results_tiffs_regions\\ma_tifs\\reports\\report_ma_present.txt', 'w') #rapport van
handelingen in het script plus errors
###########

species_threshold = open ("D:\\Universiteit_Leiden\\Stage_Naturalis\\Data\\
maxent_results\\results_tiffs_regions\\ma_tifs\\ma_thresholds_edit.txt", 'r') #
vervangen met species_threshold en dus overige namen
###########
```

```
try:
    # Check out Spatial Analyst extension license
    arcpy.CheckOutExtension("Spatial")

    # Load required toolboxes...
    arcpy.AddToolbox("C:\\Program Files (x86)\\ArcGIS\\Desktop10.6\\ArcToolbox\\
Toolboxes\\Spatial Analyst Tools.tbx") #adjust file location***
    arcpy.AddToolbox("C:\\Program Files (x86)\\ArcGIS\\Desktop10.6\\ArcToolbox\\
Toolboxes\\Conversion Tools.tbx")

    TrueValue = 1 # start optellen van binary maps
    FalseValue = 0

    SomExpressionAll_A1 = ""

    NAll = 0

    OutRasterSomAll_A1 = "D:\\Universiteit_Leiden\\Stage_naturalis\\Data\\max-
ent_results\\results_tiffs_regions\\ma_tifs\\sum_maps\\ma_present_sum.tif" #
wegschrijven van binary optelsom
    ############

    for Record in species_threshold.readlines(): # inlezen soort voor soort

        NAll = NAll + 1
        print Record #print van resultaten naar scherm
        Recordvar = Record.split(";") #opdelen tot array
        Random = Recordvar [0]
        SpeciesId = Recordvar[1] #benoemen van kolommen
        ScientificName = Recordvar[2] #
        Threshold = Recordvar[3]
        Region = Recordvar [4]
        ############

        print SpeciesId, ScientificName, Threshold

        report.write("\n" + SpeciesId + " - " + ScientificName + " - " + Threshold + "\n")

        # Set local variables # aanpassen van file path (_avg.tif)

        InRasterA1 = "D:\\Universiteit_Leiden\\Stage_naturalis\\Data\\maxent_re-
sults\\results_tiffs_regions\\ma_tifs\\present\\" + str(ScientificName) + "_avg.tif"
# opslaan tif file voor gebruik (input)
        ############

        OutRasterA1 = "D:\\Universiteit_Leiden\\Stage_naturalis\\Data\\maxent_re-
sults\\results_tiffs_regions\\ma_tifs\\bin_maps\\present\\" + str(ScientificName)
+ "_bin.tif" # output tif file  (individuele binaire kaarten)
        ############
        if NAll  = =  1:
            SomExpressionAll_A1 = SomExpressionAll_A1 + "D:\\Universiteit_Leiden\\
Stage_naturalis\\Data\\maxent_results\\results_tiffs_regions\\ma_tifs\\bin_maps\\
present\\" + str(ScientificName) + "_bin.tif"
            ############
        else:
            SomExpressionAll_A1 = SomExpressionAll_A1 + " + " + "D:\\Univer-
siteit_Leiden\\Stage_naturalis\\Data\\maxent_results\\results_tiffs_regions\\
ma_tifs\\bin_maps\\present\\" + str(ScientificName) + "_bin.tif"
            ############

        arcpy.rasterStatistics = "STATISTICS 1 1"

        print (TrueValue)
        print(FalseValue)
        print(Threshold)
        print(InRasterA1)

        # Alternative way of con # cell>threshold
        inExpressionA1 = "con(" + InRasterA1 + ' > ' + str(Threshold) + ', ' +
str(TrueValue) + ', ' + str(FalseValue) + ")"

        print inExpressionA1
        report.write(inExpressionA1 + "\n")
        report.write(InRasterA1 + "\n")
        report.write(OutRasterA1 + "\n")

        arcpy.SingleOutputMapAlgebra_sa(inExpressionA1, OutRasterA1) # kan
aangepast zijn ondertussen # de werkelijke binaire kaart

    # Volgende soort aanroepen bovenin

    # map om output weg te schrijven (workspace geeft de folder aan) gesom-
meerde kaarten
    arcpy.env.Workspace = "D:\\Universiteit_Leiden\\Stage_naturalis\\Data\\max-
ent_results\\results_tiffs_regions\\ma_tifs\\bin_maps\\present\\"
    ############
    print (SomExpressionAll_A1)
    print (OutRasterSomAll_A1)
    arcpy.SingleOutputMapAlgebra_sa(SomExpressionAll_A1, OutRaster-
SomAll_A1)

    print NAll
    report.write("\n" + str(NAll) + " species\n")
    print NAll, " species"

    report.write("\nBoolean van planten verliep zonder problemen ")
    print "Klaar"
    report.close()
    species_threshold.close()
except:
    # If an error occurred while running a tool, then print the messages.
    print arcpy.GetMessages()
    report.write("Errormessages: " + "\\n")
    report.write(arcpy.GetMessages(2))
    report.write("\\n")
    report.close()
    species_threshold.close()
```

3. Processing MaxEnt results in R

```
### statistical analysis maxent results

setwd("D:/Universiteit_Leiden/Stage_Naturalis/Data/maxent_results")
getwd()

#install.packages("ggfortify")
library(devtools)
#install_github("vqv/ggbiplot")
library(ggbiplot)
library(ggplot2)
library(ggfortify)
library(dplyr)
library(mvnormtest)
library(raster)
library(tidyr)
library(ggpubr)
library(tidyverse)

----------------------------------------------------------------------------------------------------

masterfile <- read.csv("masterfile_avg_with_region.csv", sep = ";")

master_edit <- masterfile %>%
  rename(
    Annual.mean.temperature = Bio01.contribution ,
    Mean.diurnal.range = Bio02.contribution,
    Isothermality = Bio03.contribution,
    Precipitation.of.wettest.month = Bio13.contribution,
    Precipitation.of.driest.quarter = Bio17.contribution,
    Precipitation.of.warmest.quarter =  Bio18.contribution,
    Exchangeable.aluminium = alsa.contribution,
    Cation.exchange.capacity = cecs.contribution,
    Effective.cation.exchange.capacity = ecec.contribution,
    Exchangeable.sodium.percentage = esp.contribution,
    Gypsum.content = gyps.contribution,
    Sand.mass = sdto.contribution,
    Carbonate.content = tceq.contribution)

order <- c("Annual.mean.temperature","Mean.diurnal.range" ,"Isothermality"
,"Precipitation.of.wettest.month" ,
        "Precipitation.of.driest.quarter","Precipitation.of.warmest.quarter" ,"Ex-
changeable.aluminium","Cation.exchange.capacity" ,
    "Effective.cation.exchange.capacity","Exchangeable.sodium.percentage"
,"Gypsum.content","Sand.mass","Carbonate.content")

----------------------------------------------------------------------------------------------------

# Kruskal.Wallis
kruskal.test(Annual.mean.temperature ~ Region, data = master_edit)
pairwise.wilcox.test(master_edit$Annual.mean.temperature, master_
edit$Region,
        p.adjust.method = "BH")

kruskal.test(Mean.diurnal.range ~ Region, data = master_edit)
pairwise.wilcox.test(master_edit$Mean.diurnal.range, master_edit$Region,
        p.adjust.method = "BH")

kruskal.test(Isothermality ~ Region, data = master_edit)
pairwise.wilcox.test(master_edit$Isothermality, master_edit$Region,
        p.adjust.method = "BH")

kruskal.test(Precipitation.of.wettest.month ~ Region, data = master_edit)
pairwise.wilcox.test(master_edit$Precipitation.of.wettest.month, master_
edit$Region,
        p.adjust.method = "BH")

kruskal.test(Precipitation.of.driest.quarter ~ Region, data = master_edit)
```

```
pairwise.wilcox.test(master_edit$Precipitation.of.driest.quarter, master_
edit$Region,
        p.adjust.method = "BH")
kruskal.test(Precipitation.of.warmest.quarter ~ Region, data = master_edit)
pairwise.wilcox.test(master_edit$Precipitation.of.warmest.quarter, master_
edit$Region,
        p.adjust.method = "BH")
kruskal.test(Exchangeable.aluminium ~ Region, data = master_edit)
pairwise.wilcox.test(master_edit$Exchangeable.aluminium, master_edit$Region,
        p.adjust.method = "BH")
kruskal.test(Cation.exchange.capacity ~ Region, data = master_edit)
pairwise.wilcox.test(master_edit$Cation.exchange.capacity, master_edit$Region,
        p.adjust.method = "BH")
kruskal.test(Effective.cation.exchange.capacity ~ Region, data = master_edit)
pairwise.wilcox.test(master_edit$Effective.cation.exchange.capacity, mas-
ter_edit$Region,
        p.adjust.method = "BH")
kruskal.test(Exchangeable.sodium.percentage ~ Region, data = master_edit)
pairwise.wilcox.test(master_edit$Exchangeable.sodium.percentage, mas-
ter_edit$Region,
        p.adjust.method = "BH")
kruskal.test(Gypsum.content ~ Region, data = master_edit)
pairwise.wilcox.test(master_edit$Gypsum.content, master_edit$Region,
        p.adjust.method = "BH")
kruskal.test(Sand.mass ~ Region, data = master_edit)
pairwise.wilcox.test(master_edit$Sand.mass, master_edit$Region,
        p.adjust.method = "BH")

---------------------------------------------------------------------------------------------------

## load mean contribution file
mean_contr <- read.csv("contr_means.csv", sep = ";")
## replace names in column
mean_contr$variable <- gsub("Bio01.contribution", "Annual.mean.temperature",
mean_contr$variable)
mean_contr$variable <- gsub("Bio02.contribution", "Mean.diurnal.range",
mean_contr$variable)
mean_contr$variable <- gsub("Bio03.contribution", "Isothermality", mean_
contr$variable)
mean_contr$variable <- gsub("Bio13.contribution", "Precipitation.of.wettest.
month", mean_contr$variable)
mean_contr$variable <- gsub("Bio17.contribution", "Precipitation.of.driest.quar-
ter", mean_contr$variable)
mean_contr$variable <- gsub("Bio18.contribution", "Precipitation.of.warmest.
quarter", mean_contr$variable)
mean_contr$variable <- gsub("alsa.contribution", "Exchangeable.aluminium",
mean_contr$variable)
mean_contr$variable <- gsub("cecs.contribution", "Cation.exchange.capacity",
mean_contr$variable)
mean_contr$variable <- gsub("ecec.contribution", "Effective.cation.exchange.
capacity", mean_contr$variable)
mean_contr$variable <- gsub("esp.contribution", "Exchangeable.sodium.percent-
age", mean_contr$variable)
mean_contr$variable <- gsub("gyps.contribution", "Gypsum.content", mean_
contr$variable)
mean_contr$variable <- gsub("sdto.contribution", "Sand.mass", mean_
contr$variable)
mean_contr$variable <- gsub("tceq.contribution", "Carbonate.content", mean_
contr$variable)
order2 <- c("Annual.mean.temperature","Mean.diurnal.range" ,"Isothermality"
,"Precipitation.of.wettest.month",
        "Precipitation.of.driest.quarter","Precipitation.of.warmest.quarter" ,"Ex-
changeable.aluminium","Cation.exchange.capacity",
        "Effective.cation.exchange.capacity","Exchangeable.sodium.percentage"
,"Gypsum.content","Sand.mass","Carbonate.content",
        "Annual.mean.temperature","Mean.diurnal.range" ,"Isothermality" ,"Pre-
cipitation.of.wettest.month" ,
        "Precipitation.of.driest.quarter","Precipitation.of.warmest.quarter" ,"Ex-
changeable.aluminium","Cation.exchange.capacity",
        "Effective.cation.exchange.capacity","Exchangeable.sodium.percentage"
,"Gypsum.content","Sand.mass","Carbonate.content",
        "Annual.mean.temperature","Mean.diurnal.range" ,"Isothermality" ,"Pre-
cipitation.of.wettest.month",
        "Precipitation.of.driest.quarter","Precipitation.of.warmest.quarter" ,"Ex-
changeable.aluminium","Cation.exchange.capacity",
        "Effective.cation.exchange.capacity","Exchangeable.sodium.percentage"
,"Gypsum.content","Sand.mass","Carbonate.content")

---------------------------------------------------------------------------------------------------

# change to percentage contribution
mean_contr$mean <- mean_contr$mean/10000
mean_contr$variable <- factor(mean_contr$variable,
        levels = c("Annual.mean.temperature","Mean.diurnal.range" ,"Iso-
thermality" ,"Precipitation.of.wettest.month",
        "Precipitation.of.driest.quarter","Precipitation.of.warmest.
quarter" ,"Exchangeable.aluminium","Cation.exchange.capacity",
        "Effective.cation.exchange.capacity","Exchangeable.sodium.
percentage" ,"Gypsum.content","Sand.mass","Carbonate.content"),
        ordered = TRUE)
ggplot(data = mean_contr, aes(x = region, y = mean, fill = variable)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values = c("Annual.mean.temperature" = "#CC0033","Mean.
diurnal.range" = "#993366","Isothermality" = "#9966CC",
        "Precipitation.of.wettest.month" = "#6699CC","Precipitation.
of.driest.quarter" = "#336699",
        "Precipitation.of.warmest.quarter" = "#003366","Exchangeable.
aluminium" = "#996633",
        "Cation.exchange.capacity" = "#CC9933","Effective.cation.
exchange.capacity" = "#CCCC33",
        "Exchangeable.sodium.percentage" = "#99CC33" ,"Gypsum.
content" = "#99CC66","Sand.mass" = "#339966",
        "Carbonate.content" = "#336633"), breaks = c("Annual.mean.
temperature","Mean.diurnal.range" ,"Isothermality",
        "Precipitation.of.wettest.month" ,"Precipitation.of.driest.
quarter","Precipitation.of.warmest.quarter",
        "Exchangeable.aluminium","Cation.exchange.capacity" ,"Ef-
fective.cation.exchange.capacity",
        "Exchangeable.sodium.percentage" ,"Gypsum.content","Sand.
mass","Carbonate.content")) +
  labs(title = "Contribution in Percentage per Region",x = "Region", y = "Percent-
age (%)", fill = "Variables") +
  theme_minimal() +
  theme(text = element_text(size = 16)) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        axis.text.y = element_text(angle = 45))
all_regions_plot <- ggplot(data = mean_contr, aes(x = region, y = mean, fill = vari-
able)) +
  geom_bar(stat = "identity", position = position_dodge())+
  scale_fill_manual(values = c("Annual.mean.temperature" = "#CC0033","Mean.
diurnal.range" = "#993366","Isothermality" = "#9966CC",
        "Precipitation.of.wettest.month" = "#6699CC","Precipitation.
of.driest.quarter" = "#336699",
        "Precipitation.of.warmest.quarter" = "#003366","Exchangeable.
aluminium" = "#996633",
        "Cation.exchange.capacity" = "#CC9933","Effective.cation.
exchange.capacity" = "#CCCC33",
        "Exchangeable.sodium.percentage" = "#99CC33" ,"Gypsum.
content" = "#99CC66","Sand.mass" = "#339966",
        "Carbonate.content" = "#336633"), breaks = c("Annual.mean.
temperature","Mean.diurnal.range" ,"Isothermality",
        "Precipitation.of.wettest.month" ,"Precipitation.of.driest.
quarter","Precipitation.of.warmest.quarter" ,
        "Exchangeable.aluminium","Cation.exchange.capacity" ,"Ef-
fective.cation.exchange.capacity",
        "Exchangeable.sodium.percentage" ,"Gypsum.content","Sand.
mass","Carbonate.content")) +
  labs(title = "Contribution in Percentage per Region",x = "Region", y = "Percent-
age (%)")+  theme_minimal()
all_regions_plot
### Seperate plots for percentages
mean_contr_pp <- subset(mean_contr, region = = "pp")
variables_pp <- ggplot(data = mean_contr_pp, aes(x = variable, y = mean,
fill = variable)) +
  ylim(0,25) +
  scale_x_discrete(limits = order) +
  geom_bar(stat = "identity", position = position_dodge())+
  scale_fill_manual(values = c("Annual.mean.temperature" = "#CC0033","Mean.
diurnal.range" = "#993366","Isothermality" = "#9966CC",
        "Precipitation.of.wettest.month" = "#6699CC","Precipitation.
of.driest.quarter" = "#336699",
        "Precipitation.of.warmest.quarter" = "#003366","Exchangeable.
aluminium" = "#996633",
        "Cation.exchange.capacity" = "#CC9933","Effective.cation.
exchange.capacity" = "#CCCC33",
        "Exchangeable.sodium.percentage" = "#99CC33" ,"Gypsum.
content" = "#99CC66","Sand.mass" = "#339966",
        "Carbonate.content" = "#336633"), breaks = c("Annual.mean.
temperature","Mean.diurnal.range" ,"Isothermality" ,
        "Precipitation.of.wettest.month" ,"Precipitation.of.driest.
```

```
quarter","Precipitation.of.warmest.quarter" ,
            "Exchangeable.aluminium","Cation.exchange.capacity" ,"Ef-
fective.cation.exchange.capacity",
            "Exchangeable.sodium.percentage" ,"Gypsum.content","Sand.
mass","Carbonate.content")) +
  labs(title = "Contribution in Percentage per Region",x = "The Philippines",
y = "Percentage (%)")+
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
     axis.text.y = element_text(angle = 45))

variables_pp

-------------------------------------------------------------------------------

mean_contr_pm <- subset(mean_contr, region  = =  "pm")

variables_pm <- ggplot(data = mean_contr_pm, aes(x = variable, y = mean,
fill = variable)) +
  ylim(0,25) +
  scale_x_discrete(limits = order) +
  geom_bar(stat = "identity", position = position_dodge())+
  scale_fill_manual(values = c("Annual.mean.temperature" = "#CC0033","Mean.
diurnal.range" = "#993366","Isothermality" = "#9966CC",
            "Precipitation.of.wettest.month" = "#6699CC","Precipitation.
of.driest.quarter" = "#336699",
            "Precipitation.of.warmest.quarter" = "#003366","Exchangeable.
aluminium" = "#996633",
            "Cation.exchange.capacity" = "#CC9933","Effective.cation.
exchange.capacity" = "#CCCC33",
            "Exchangeable.sodium.percentage" = "#99CC33" ,"Gypsum.
content" = "#99CC66","Sand.mass" = "#339966",
            "Carbonate.content" = "#336633"), breaks = c("Annual.mean.
temperature","Mean.diurnal.range" ,"Isothermality" ,
            "Precipitation.of.wettest.month" ,"Precipitation.of.driest.
quarter","Precipitation.of.warmest.quarter" ,
            "Exchangeable.aluminium","Cation.exchange.capacity" ,"Ef-
fective.cation.exchange.capacity",
            "Exchangeable.sodium.percentage" ,"Gypsum.content","Sand.
mass","Carbonate.content")) +
  labs(title = "Contribution in Percentage per Region",x = "Shared Species",
y = "Percentage (%)")+
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
     axis.text.y = element_text(angle = 45))

variables_pm

mean_contr_ma <- subset(mean_contr, region  = =  "ma")

variables_ma <- ggplot(data = mean_contr_ma, aes(x = variable, y = mean,
fill = variable)) +
  ylim(0,25) +
  scale_x_discrete(limits = order) +
  geom_bar(stat = "identity", position = position_dodge())+
  scale_fill_manual(values = c("Annual.mean.temperature" = "#CC0033","Mean.
diurnal.range" = "#993366","Isothermality" = "#9966CC",
            "Precipitation.of.wettest.month" = "#6699CC","Precipitation.
of.driest.quarter" = "#336699",
            "Precipitation.of.warmest.quarter" = "#003366","Exchangeable.
aluminium" = "#996633",
            "Cation.exchange.capacity" = "#CC9933","Effective.cation.
exchange.capacity" = "#CCCC33",
            "Exchangeable.sodium.percentage" = "#99CC33" ,"Gypsum.
content" = "#99CC66","Sand.mass" = "#339966",
            "Carbonate.content" = "#336633"), breaks = c("Annual.mean.
temperature","Mean.diurnal.range" ,"Isothermality" ,
            "Precipitation.of.wettest.month" ,"Precipitation.of.driest.
quarter","Precipitation.of.warmest.quarter" ,
            "Exchangeable.aluminium","Cation.exchange.capacity" ,"Ef-
fective.cation.exchange.capacity",
```

```
            "Exchangeable.sodium.percentage" ,"Gypsum.content","Sand.
mass","Carbonate.content")) +
  labs(title = "Contribution in Percentage per Region",x = "Mainland Asia",
y = "Percentage (%)")+
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
     axis.text.y = element_text(angle = 45))

variables_ma

-------------------------------------------------------------------------------

# average of the three means
library(data.table)
tot <- mean_contr %>% dplyr::select("variable", "mean")
variable_tot_list <- rbindlist(list(tot))[,lapply(.SD,mean), list(variable)]

variables_tot <- ggplot(data = variable_tot_list, aes(x = variable, y = mean,
fill = variable)) +
  ylim(0,25) +
  scale_x_discrete(limits = order) +
  geom_bar(stat = "identity", position = position_dodge())+
  labs(title = "Contribution in Percentage",x = "Bioclimatic Variables", y = "Per-
centage (%)")+
  scale_fill_manual(values = c("Annual.mean.temperature" = "#CC0033","Mean.
diurnal.range" = "#993366","Isothermality" = "#9966CC",
            "Precipitation.of.wettest.month" = "#6699CC","Precipitation.
of.driest.quarter" = "#336699",
            "Precipitation.of.warmest.quarter" = "#003366","Exchangeable.
aluminium" = "#996633",
            "Cation.exchange.capacity" = "#CC9933","Effective.cation.
exchange.capacity" = "#CCCC33",
            "Exchangeable.sodium.percentage" = "#99CC33" ,"Gypsum.
content" = "#99CC66","Sand.mass" = "#339966",
            "Carbonate.content" = "#336633"), breaks = c("Annual.mean.
temperature","Mean.diurnal.range" ,"Isothermality" ,
            "Precipitation.of.wettest.month" ,"Precipitation.of.driest.
quarter","Precipitation.of.warmest.quarter" ,
            "Exchangeable.aluminium","Cation.exchange.capacity" ,"Ef-
fective.cation.exchange.capacity",
            "Exchangeable.sodium.percentage" ,"Gypsum.content","Sand.
mass","Carbonate.content")) +
  theme_minimal()+
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
            axis.text.y = element_text(angle = 45))

variables_tot

with(variable_tot_list, sum(mean[variable  = =  "Annual.mean.temperature" |
variable  = =  "Mean.diurnal.range" |
            variable  = =  "Isothermality" | variable  = = "Precipitation.
of.wettest.month" |
            variable  = =  "Precipitation.of.driest.quarter" | vari-
able  = = "Precipitation.of.warmest.quarter"]))

with(variable_tot_list, sum(mean[variable  = =  "Exchangeable.aluminium" | vari-
able  = = "Cation.exchange.capacity" |
            variable  = =  "Effective.cation.exchange.capacity" | vari-
able  = = "Exchangeable.sodium.percentage" |
            variable  = =  "Gypsum.content" | variable  = = "Sand.mass"|
variable  = = "Carbonate.content"]))

my_comparisons <- list( c("Isothermality", "Gypsum.content"), c("Annual.mean.
temperature", "Precipitation.of.driest.quarter"), c("Exchangeable.aluminium",
"Gypsum.content") )

ggboxplot(ToothGrowth, x = "dose", y = "len",
     color = "dose", palette = "jco")+
  stat_compare_means(comparisons = my_comparisons)+ # Add pairwise
comparisons p-value
  stat_compare_means(label.y = 50)
```