Throwing light on dark diversity of vascular plants in China: predicting the distribution of dark and threatened species under global climate change

Lili Tang[1], Runxi Wang[1], Kate S. He[2], Cong Shi[3], Tong Yang[1], Yaping Huang[1], Pufan Zheng[1] and Fuchen Shi[1]

[1] College of Life Sciences, NanKai University, Tianjin, China
[2] Department of Biological Sciences, Murray State University, Murray, KY, USA
[3] School of Environment and Energy, Shenzhen Graduate School, Peking University, Shenzhen, China

ABSTRACT

Background: As global climate change accelerates, ecologists and conservationists are increasingly investigating changes in biodiversity and predicting species distribution based on species observed at sites, but rarely consider those plant species that could potentially inhabit but are absent from these areas (i.e., the dark diversity and its distribution). Here, we estimated the dark diversity of vascular plants in China and picked up threatened dark species from the result, and applied maximum entropy (MaxEnt) model to project current and future distributions of those dark species in their potential regions (those regions that have these dark species).

Methods: We used the Beals probability index to estimate dark diversity in China based on available species distribution information and explored which environmental variables had significant impacts on dark diversity by incorporating bioclimatic data into the random forest (RF) model. We collected occurrence data of threatened dark species (*Eucommia ulmoides*, *Liriodendron chinense*, *Phoebe bournei*, *Fagus longipetiolata*, *Amentotaxus argotaenia*, and *Cathaya argyrophylla*) and related bioclimatic information that can be used to predict their distributions. In addition, we used MaxEnt modeling to project their distributions in suitable areas under future (2050 and 2070) climate change scenarios.

Results: We found that every study region's dark diversity was lower than its observed species richness. In these areas, their numbers of dark species are ranging from 0 to 215, with a generally increasing trend from western regions to the east. RF results showed that temperature variables had a more significant effect on dark diversity than those associated with precipitation. The results of MaxEnt modeling showed that most threatened dark species were climatically suitable in their potential regions from current to 2070.

Discussions: The results of this study provide the first ever dark diversity patterns concentrated in China, even though it was estimated at the provincial scale. A combination of dark diversity and MaxEnt modeling is an effective way to shed light on the species that make up the dark diversity, such as projecting the distribution of specific dark species under global climate change. Besides,

the combination of dark diversity and species distribution models (SDMs) may also be of value for ex situ conservation, ecological restoration, and species invasion prevention in the future.

INTRODUCTION

Biodiversity is a fundamental and central topic in ecology and conservation. In recent years, ecologists have been increasingly investigating the effects of climate change on biodiversity and species' distribution (*Dawson et al., 2011*; *Guillera-Arroita et al., 2015*; *Gomes et al., 2018*). Most of these studies relied on the observed species records from field sites to predict the response of species to climate change and to monitor biodiversity changes (*Cárdenas et al., 2011*; *Dawson et al., 2011*; *Turner, 2014*; *Storkey et al., 2015*; *Gray et al., 2016*). However, the observed species do not reflect the complete habitat-specific species pool of a site as some species will remain undetected due to limited sampling efforts and resources of the investigators and the timing of sampling (*Pärtel, Szava-Kovats & Zobel, 2011*). For example, rare species and species with very short lifespan can be overlooked easily. That is especially true when sampling units are large in space and sparse in time. This particular set of species that belong to a particular species pool but are not locally present, or that are absent from a community but have the potential to establish are known as "dark diversity" (*Lewis, Szava-Kovats & Pärtel, 2016*; *Pärtel, Szava-Kovats & Zobel, 2011*, *2013*).

Since dark diversity was proposed (*Pärtel, Szava-Kovats & Zobel, 2011*), there have been rapid developments regarding its estimating methods. Ellenberg indicator values (*Ellenberg et al., 1991*; *Hill et al., 1999*; *Pärtel et al., 1996*) and Beals probability index (*Beals, 1984*) are the most common methods to estimate dark diversity. In addition, SDMs are also used to complement dark diversity estimations (*De Bello et al., 2016*; *Lewis, Szava-Kovats & Pärtel, 2016*; *Ronk et al., 2016*). Ellenberg indicator values directly estimate dark diversity or species pool size, and it can only be used when there is sufficient information on the habitat requirements of species (*De Bello et al., 2016*; *Lewis, Szava-Kovats & Pärtel, 2016*; *Ronk et al., 2016*). By contrast, the Beals probability index uses the geographical distribution or co-occurrence patterns of species as a proxy of their ecological requirements (*De Bello et al., 2016*; *Lewis, Szava-Kovats & Pärtel, 2016*; *Moeslund et al., 2016*; *Riibak et al., 2015*; *Ronk et al., 2016*, *2017*; *Ronk, Szava-Kovats & Pärtel, 2015*). Moreover, comparative studies have indicated that this proxy gives accurate estimates of dark diversity for macro-scale ecological communities (*Lewis, Szava-Kovats & Pärtel, 2016*; *Ronk et al., 2016*).

To date, most research on dark diversity has been conducted in Europe (*Kasari et al., 2016*; *Lewis et al., 2017*; *Ronk, Szava-Kovats & Pärtel, 2015*), and those studies have shown that dispersal limitation and stress-tolerance play essential roles in shaping the

patterns of observed and dark diversity (*Riibak et al., 2015*). However, the impacts of future climate change are largely unknown on shaping the patterns of observed and dark diversity in those studies. Furthermore, most studies have focused on mapping dark diversity at macro-scale and comparing methods to estimate dark diversity but ignored the specific information on the identity of species that constitute dark diversity. For example, questions such as whether the dark diversity dataset includes any threatened species and what would happen to their potential distribution with global climate change have not yet been answered.

In this study, we tried to seperate threatened species from dark diversity pool and to explore how their distributions shift in the future in China using the species distribution model. We proposed three fundamental questions: (1) What is the level and distribution of dark diversity in China, and does the dark diversity include threatened species? (2) Which environmental variables affect dark species distributions the most? (3) Will the potential areas of dark (particularly threatened) plants remain suitable in the future under global climate change? To address these questions, we calculated the Beals probability index using previously complied vascular plants datasets at the macro-scale to investigate the level of dark diversity in China. We also used random forest (RF) model to assess the contribution of a number of variables to dark diversity. Lastly, we performed maximum entropy (MaxEnt) modeling to project changes in the distribution of threatened dark species under climate scenarios in the years of 2050 and 2070. The results of this study could provide novel information on dark diversity in China, as this has never been investigated before, to our best knowledge. Further, information on dark diversity at a regional scale could have significant implications on conservation efforts, including ecological restoration and invasion risk assessment.

MATERIALS AND METHODS

Species distribution data and environmental variables

This study covered the entire area of China. We retrieved vascular plants occurrence data in each province and municipality of China from the Flora of China (http://frps.eflora.cn/) and merged data of municipalities into their nearest provinces, so we got 27 records in total (Fig. 1). In comparison with the data applied to the study by *Pärtel, Szava-Kovats & Zobel (2011)*, derived from a global map and at the scale of 100 x 100 km, although our data are at larger scale, the observed species lists are much more complete, especially for that of Northern China, which was considered difficult to access (*Kier et al., 2005*). We constructed an individual species dataset for them to show which plants were present or absent. Although the floristic data have been collected since 1959, we assume that they are still relevant because plants, particularly those in the temperate forest, respond relatively slowly to environmental changes (*Franklin et al., 2016*; *Menéndez et al., 2006*). We only included plant data at the species level, and we merged all subspecies. We also excluded all cultivated species that were not considered a natural part of the vegetation based on their life history and dispersal traits. Synonyms and conservation status such as threatened species were checked using The Plant List (http://www.theplantlist.org/)
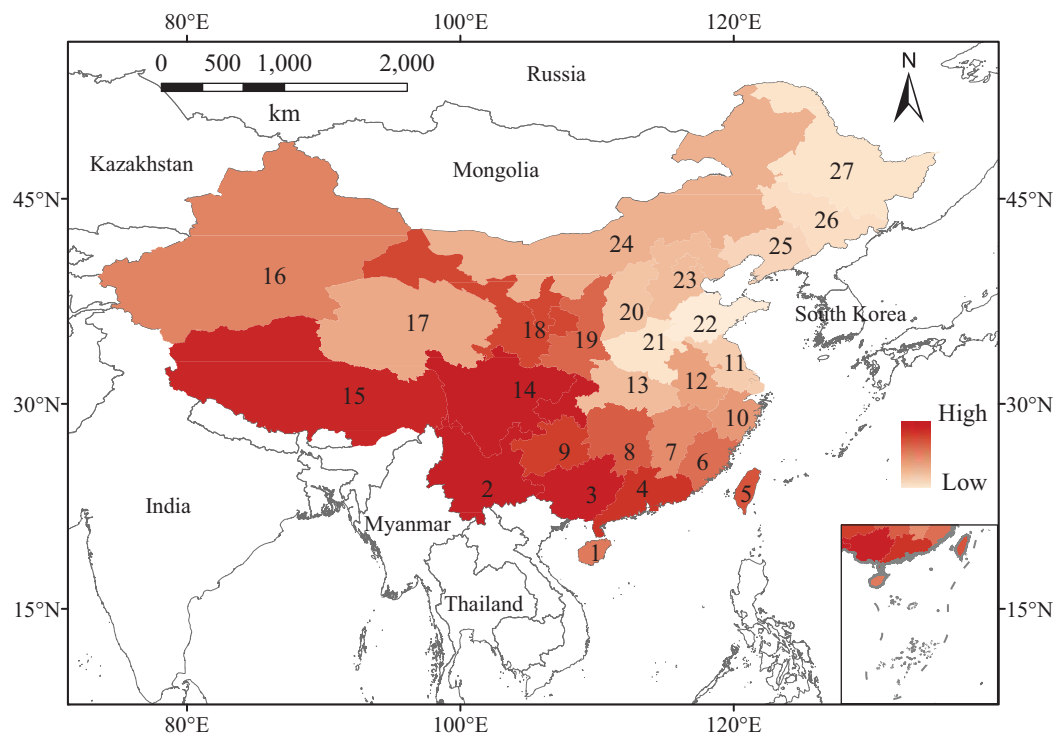
Figure 1 Map of China, showing the observed species richness for vascular plants (dark and light represent high and low richness levels, respectively). The provinces are numbered and labeled as follow: Hainan (1), Yunnan (2), Guangxi (3), Guangdong including Hongkong and Macau (4), Taiwan (5), Fujian (6), Jiangxi (7), Hunan (8), Guizhou (9), Zhejiang (10), Jiangsu including Shanghai (11), Anhui (12), Hubei (13), Sichuan including Chongqing (14), Xizang (15), Xinjiang (16), Qinghai (17), Gansu including Ningxia (18), Shaanxi (19), Shanxi (20), Henan (21), Shandong (22), Hebei including Beijing and Tianjin (23), Neimenggu (24), Liaoning (25), Jilin (26), Heilongjiang (27).

Full-size 🖼 DOI: 10.7717/peerj.6731/fig-1

and The IUCN Red List of Threatened Species (http://www.iucnredlist.org/). This observed species richness (number of species) distribution shows a latitudinal trend, increasing from north to south in China (Fig. 1).

Climatic variables, such as precipitation and temperature have a significant impact on species' distribution, particularly at large spatial scales and over long timeframes (*Beer et al., 2010*; *Kelly & Goulden, 2008*; *Loarie et al., 2009*; *Williams, Jackson & Kutzbach, 2007*). We extracted 19 bioclimatic variables (BIO1–BIO19; Table S1) in years of current, 2050 and 2070, respectively, from Worldclim (resolution, 2.5 s; www.Worldclim.org/). As for the variables in 2050 and 2070, we selected those under representative concentration pathway 4.5 (RCP4.5) and based on the CCSM4 global climate model (GLM). There are four RCPs, ranging from very high (RCP8.5, the temperature will increase to $3.7 \pm 0.7\,°C$) through to deficient (RCP2.6, the temperature will rise to $1.0 \pm 0.4\,°C$) future concentrations (*Collins et al., 2013*). Based on current status of global warming and the goal that hold average temperature increase to well below $2\,°C$ (*IPCC, 2018*), in this study, we chose RCP4.5 scenario as the future climate scenario. In addition, according to the projection by *Ying (2012)*, until the year of 2018, the CCSM4 has been showing

as the more accurate model than the other 16 GLMs under RCP4.5 scenarios, so in this study, the future climate projections are based on the CCSM4 climate model. Current and future bioclimatic projection data were used in the MaxEnt model to predict habitat suitability of six threatened dark species that will be detailed in the following sections.

Estimation of dark diversity and community completeness

To obtain the best possible estimates of dark diversity in China, we used Beals probability index (*Beals, 1984*), as recommended by *Lewis, Szava-Kovats & Pärtel (2016)*. This index can be used to estimate the probability of a species occurring in a particular region based on its co-occurrence within other regions and was calculated using the package "vegan" (*Oksanen et al., 2016*) in R ver. 3.3.1 (*R Development Core Team, 2016*).

The Beals index is defined as:

$$P_{ij} = \frac{1}{S_i - I_{ij}} \sum_{k \neq j} \frac{N_{jk} I_{ik}}{N_k}$$

where P_{ij} is the probability that species j will occur at community i, S_i is the number of species at community i, I_{ij} is the incidence (0, 1) of species j at community i. N_{jk} is the number of joint occurrence of species j and k ($k \neq j$) at community i, I_{ik} is the incidence (0, 1) of species k at community i, and N_k is the number of occurrences of species k (for further details, see *Münzbergová & Herben (2004)* and *Lewis, Szava-Kovats & Pärtel (2016)*).

The probability of occurrence varies among species, regions and depends on the frequencies of species in a particular assemblage. Each species was assigned an individual probability threshold, which was a quantile calculated from a user-defined probability. Based on the comparison of the dark diversity from different probabilities (1%, 5%, 10%) (Table S2), to explore the broader possible species pool in China, it was defined 1% in all regions here. For each area, a species was included in the dark diversity when it was absent from a target region and its occurrence probability was higher than its threshold value.

In addition, we calculated the community completeness index for each region using ln (observed richness/dark diversity) (*Pärtel, Szava-Kovats & Zobel, 2013*).

Assessing variables that affect dark diversity

We used RF to assess the contribution of bioclimatic variables toward the dark diversity. The RF technique estimates the importance of a predictive variable by evaluating the Out-of-bag (OOB) error increase (*Mutanga, Adam & Cho, 2012*). In other words, the decrease of prediction accuracy, represented as the percentage of increased mean square error (% IncMSE) when OOB data for the specific variable is switched while all the other variables remained (*Mutanga, Adam & Cho, 2012*). Here, we extracted a matrix of 3,844 spot records covering the whole area of China, with their 19 current bioclimatic data, and applied it to RF, where their dark diversity values at the regional level are treated as the response variable.

Random forest is a machine learning approach based on classification and regression trees (CART; *Breiman, 2001*). The RF model-building process is similar to that of CART; only it combines numerous independent trees to reach a final decision (for further details, see *Wiesmeier et al. (2011)* and *Da Silva Chagas et al. (2016)*). We estimated the relative importance of each variable in the RF model using OOB randomly selected data. The mean square error (MSE) was calculated as:

$$\text{MSE}_{\text{OBB}} = \frac{1}{N} \sum_{i=1}^{n} \left(Z_i - \hat{Z}_i^{\text{OBB}} \right)^2$$

where Z_i is the measured value of the variable and $\hat{Z}_i^{\text{OBB}}$ is the average of all OBB predictions. The MSE_{OBB} is normalized as it depends on the unit of the response variable.

Random forest modeling was implemented by R ver. 3.3.1 (*R Development Core Team, 2016*), using the "RF" package (*Liaw & Wiener, 2002*). The number of trees (ntree) was 500 and the number of randomly selected predictor variables at each node (mtry) was three.

Predicting the distribution of rare dark species with future climate change

We used MaxEnt modeling (Version 3.4.1) to predict changes in the potential areas of six rare dark plant species under global climate change. MaxEnt is a widely used species distribution algorithm, and many studies have compared it with other SDMs to confirm its predictive ability (*Elith & Graham, 2009*; *Elith, Kearney & Phillips, 2010*; *Tognelli et al., 2009*; *Williams et al., 2009*). It is used to predict the geographic distributions of species based on incomplete information of species (presence-only datasets) occurrence data and environmental variables by calculating the MaxEnt of species distribution (*Phillips, Anderson & Schapire, 2006*; *Phillips & Dudik, 2008*).

In this study, we chose the near threatened, vulnerable, and endemic species: *Eucommia ulmoides, Liriodendron chinense, Phoebe bournei, Fagus longipetiolata, Amentotaxus argotaenia*, and *Cathaya argyrophylla* from the threatened dark species based on available data and species' distribution range: all species have a limited distribution and a small dataset of occurrence sites. In total, we found 187 occurrence sites for these species in relevant references, local flora, and the flora of China, which we included in the MaxEnt model (*A. argotaenia*: 30 sites, *C. argyrophylla*: 22 sites, *E. ulmoides*: 32 sites, *L. chinense*: 35 sites, *F. longipetiolata*: 29 sites, *P. bournei*: 39 sites) (Fig. S1; Table 1). According to the records of the references, geographic coordinate data are within an accuracy of about three km. The 19 bioclimatic variables were filtered out some less important ones by using RF model and removed a few high correlated variables by using correlation coefficient. The correlation coefficient was implemented in R ver. 3.3.1 (*R Development Core Team, 2016*), using the "raster" package (*Hijmans, 2018*), the cut-off threshold is 0.80.

In our model, we split the dataset into two parts: 75% of the occurrence data were used as training set, while the remaining 25% was used as a test set to evaluate the strength of the model. In this model, we set the 10,000 points as the max number of background sampling entire China. As for the small size of our data, we tuned the regularization

Table 1 Threatened species among the dark diversity in China.

	Species	Threatened level	Potential area	Life form	Life span	Habitat types	[i]Data source for MaxEnt	Year of data source
1	*Amentotaxus argotaenia*	NT	Anhuai	Tree	Perennial	Forest	1–6	2000, 2001, 2004, 2007, 2014, 2017
2	*Eucommia ulmoides*	NT	Jiangsu (Shanghai)	Tree	Perennial	Forest	7–10	2013, 2014, 2016
3	*Liriodendron chinense*	NT	Jiangsu (Shanghai) & Guangdong (Hongkong, Macao)	Tree	Perennial	Forest	3, 11–18	2003, 2007, 2011, 2014, 2016, 2017
4	*Phoebe bournei*	NT	Anhuai & Jiangsu (Shanghai)	Tree	Perennial	Forest	19	2012
5	*Cathaya argyrophylla*	VU	Hubei	Tree	Perennial	Forest	20–22	1994, 2006, 2016
6	*Fagus longipetiolata*	VU	Jiangsu (Shanghai)	Tree	Perennial	Forest	23, 24	1997, 2008
7	*Cycas taiwaniana*	EN	Guizhou & Hainan	Tree	Perennial	Forest		
8	*Dendrobium officinale*	CR	Guizhou	Tree	Perennial	Forest		

Note:

IUCN, International Union for the Conservation of Nature threat levels; NT, near threatened; VU, vulnerable; EN, endangered; CR, critically endangered. Potential area was estimated by dark diversity model.

[i] Details can be found in the Supplemental Information.

multipliers (0.1, 0.25, 0.35, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 2.25, 2.5, 2.75, and 3) in MaxEnt model, to find out the best one with the smallest corrected Akaike information criterion (AICc) value (*Morales, Fernández & Baca-González, 2017*). The AICc was calculated by ENMTOOLS (version 1.4.4) (*Warren & Seifert, 2011*). The area under the receiver operating characteristic curve (AUC) and omission error (minimum training presence as the logistic threshold) was used to quantify the strength of the current scenario models. AUC values range from zero to one, with the values of ≤0.5 indicating that the model performs worse than a random model and values close to 1 indicating that the model performs better than a random model (*Swets, 1988*). In this study, we considered values of 0.5–0.6 to represent no discrimination, 0.6–0.7 as unaccepted, 0.7–0.8 as accepted, 0.8–0.9 as excellent, 0.9–1 as outstanding and 1 as perfect (*Mischler et al., 2012*; *Phillips, Anderson & Schapire, 2006*). Finally, we used the Jackknife test to evaluate which variable made the greatest contribution to the distribution of each of these six species under the current climatic scenario.

RESULTS

Dark diversity and community completeness in China

Different regions in China show various dark diversities, ranging from 0 to 215. In comparison with their observed local richness (from 1,170 to 11,760), their dark diversities are smaller. Meanwhile, the observed richness showed a latitudinal gradient change, increasing from north to south (Fig. 1), while dark diversity exhibited no latitudinal trend, and the values declined from east to west, except for Qinghai. The highest levels of dark diversity were concentrated in three regions: Qinghai (215), the southeast of the northern China plain (Shandong, 159; Henan, 155; Jiangsu including Shanghai, 143), and Heilongjiang (142) (Fig. 2A).

The community completeness indexes showed a similar distribution as their observed richness, almost fully complete in the west communities (except for Qinghai Province),
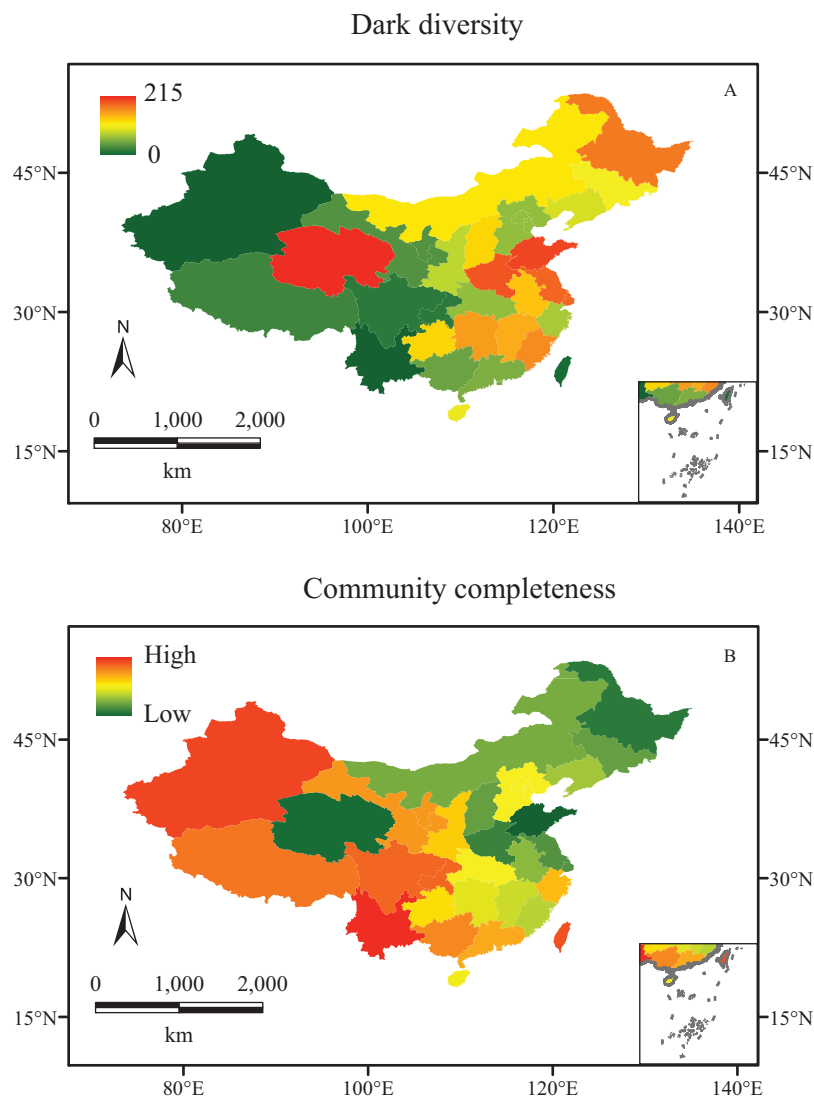
Figure 2 The biodiversity of vascular plant species in China. (A) Dark diversity, and (B) community completeness (ln (observed richness/dark diversity)). Red and green indicate high and low values, respectively. Full-size 🖼 DOI: 10.7717/peerj.6731/fig-2

and less complete in the east. In addition, the completeness indexes were lower in the northeast than those in the southeast of China (Fig. 2B).

We identified eight threatened species among the dark diversity, representing a range of threat levels (Table 1).

Explanatory variables for dark species

The relative importance of the 19 bioclimatic variables regarding dark diversity in China, as assessed using RF, is shown in Fig. 3. We found that variables associated with temperature showed more importance than those associated with precipitation (Fig. 3). However, together these variables only explained 71.72% of the variance, thus few unknown factors remain. The most important variables were Temperature annual range (BIO7) and Temperature seasonality (BIO4), Mean temperature of warmest quarter
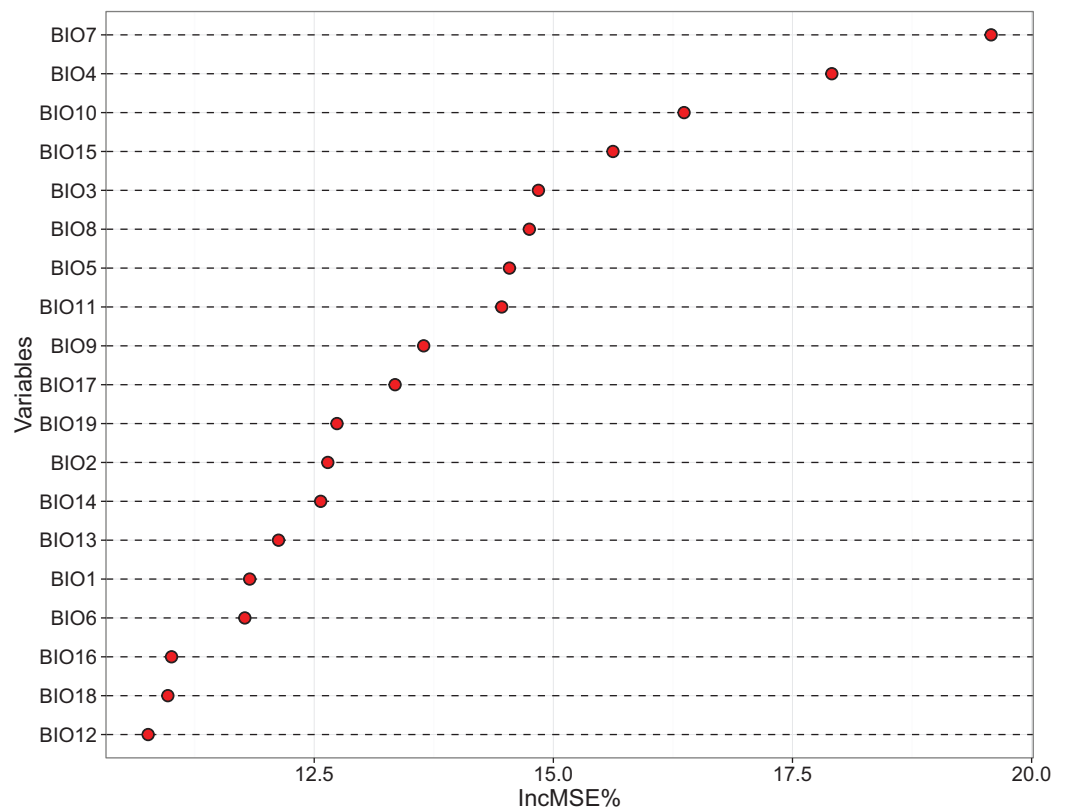
Figure 3 Variables importance derived from random forest models showed by increase in MSE (%). BIO 1–BIO 11 are the variables associated with temperature, while BIO 12–BIO 19 are the variables associated with precipitation. In this figure, most of temperature variables are gain the high % IncMSE.

Full-size ⤢ DOI: 10.7717/peerj.6731/fig-3

(BIO10), and Precipitation seasonality (BIO15). Other significant variables were Isothermality (BIO3), Mean temperature of wettest quarter (BIO8), Max temperature of warmest month (BIO5), Mean temperature of coldest quarter (BIO11), Mean temperature of driest quarter (BIO9), Precipitation of driest quarter (BIO17), Precipitation of coldest quarter (BIO19), Mean diurnal range (BIO2), and Precipitation of driest month (BIO14). These variables were filtered to use in MaxEnt projection firstly.

Then according to the results of cross-correlation (Fig. S2), we selected seven bioclimatic variables used in the MaxEnt model for current climate scenario (Fig. S2), and the contribution of them differed between six species (Table 2). For these six species, BIO14 in current bioclimatic scenarios was the most important variable when projected their distribution, BIO5 also showed its significance to project the species distribution (Table 2). For the 2050 and 2070 climate scenarios, we selected six bioclimatic variables used in MaxEnt (Fig. S2).

Distribution of dark threatened species in the future

All the models under current bioclimatic scenarios were considered outstanding according to their omission error, AUC and AICc values (Table 2), suggesting that they could accurately predict species distributions.

Table 2 Summary of the contribution of the bioclimatic variables used in the MaxEnt model and the omission rate, AICc, and AUC values for the model.

Species	Contribution to MaxEnt models (%)							Betamultiplier	AICc	Fractional predicted area	Omission rate	AUC
	BIO2	BIO3	BIO5	BIO7	BIO11	BIO14	BIO15					
Amentotaxus argotaenia	0.02	0	21.09	9.23	6.74	62.16	0.77	0.5	1,741.1	0.175	0	0.971
Cathaya argyrophylla	1.86	2.25	30.12	1.63	4.8	59.33	0	2.25	603.46	0.036	0	0.995
Eucommia ulmoides	0.75	6.53	7.26	13.95	38.8	30.65	2.05	2	966.64	0.172	0	0.94
Fagus longipetiolata	0.41	1.24	30.2	1.01	1.45	65.14	0.55	1	826.84	0.076	0.286	0.956
Liriodendron chinense	6.84	0	13.56	1.97	12.92	64.45	0.26	1.5	1,033.54	0.198	0	0.948
Phoebe bournei	0.72	0.01	8.28	8.09	0.28	81.16	1.46	0.75	1,122.42	0.07	0.222	0.953

Note:
According to omission rate and AUC values, all the models were considered outstanding. For six trees, BIO14 was the most important variables when projected their distributions under current climatic scenarios.

By overlaying the distribution results from MaxEnt with those from Dark Diversity, we could find out that most of these dark plant potential regions were also their climatically suitable areas from current to 2070 in MaxEnt modeling. However, Jiangsu province was estimated having dark species *F. longipetiolata* while MaxEnt models predicted it not climatically suitable for *F. longipetiolata* either under current or future climatic scenarios (Fig. 4A).

In addition, for these six species, the projected climatic suitability in their potential regions would change under future climatic scenarios. Our MaxEnt modeling results showed that most species' suitable areas would firstly increase by 2050, and then contract by 2070 (Fig. 4). However, *E. ulmoides* suitable area in Jiangsu province would shrink by 2050 and then expand a little by 2070 (Fig. 4A). We believed these shifts probably related to BIO14 decline from current to 2050, and BIO19 increase from current to 2070, and similar associations have also been found in other species (*Bueno et al., 2017*; *Deb et al., 2017*; *Noulekoun et al., 2017*; *Wu et al., 2017*).

DISCUSSION

Unlike most other studies to date, we estimated dark diversity using data at the provincial scale instead of at spatial grid (*Lewis, Szava-Kovats & Pärtel, 2016*; *Ronk, Szava-Kovats & Pärtel, 2015*). We do consider that a provincial level study is rather coarse in terms of its spatial scales, but still, the results of this study provide the first ever dark diversity patterns concentrated in China. A more robust study can be carried out when data at finer scales become available. Moreover, our result showed a similar distribution compared with the previous global scale study, where data at spatial grid were applied (*Pärtel, Szava-Kovats & Zobel, 2011*). Thirdly, gridded data at a finer scale may not always be able to bring more accuracy to dark diversity estimation. In the study by *De Bello et al. (2016)*, they merged plot-level species lists into 18 habitats at a larger scale and produced habitat-level estimates of dark diversity estimates. They concluded that habitat-level analysis had a higher similarity with expert estimates than that in the grid (*De Bello et al., 2016*).

In this study, the results combination of dark diversity and Maxent helped six threatened species to find out their possibly suitable areas presently and in the future,
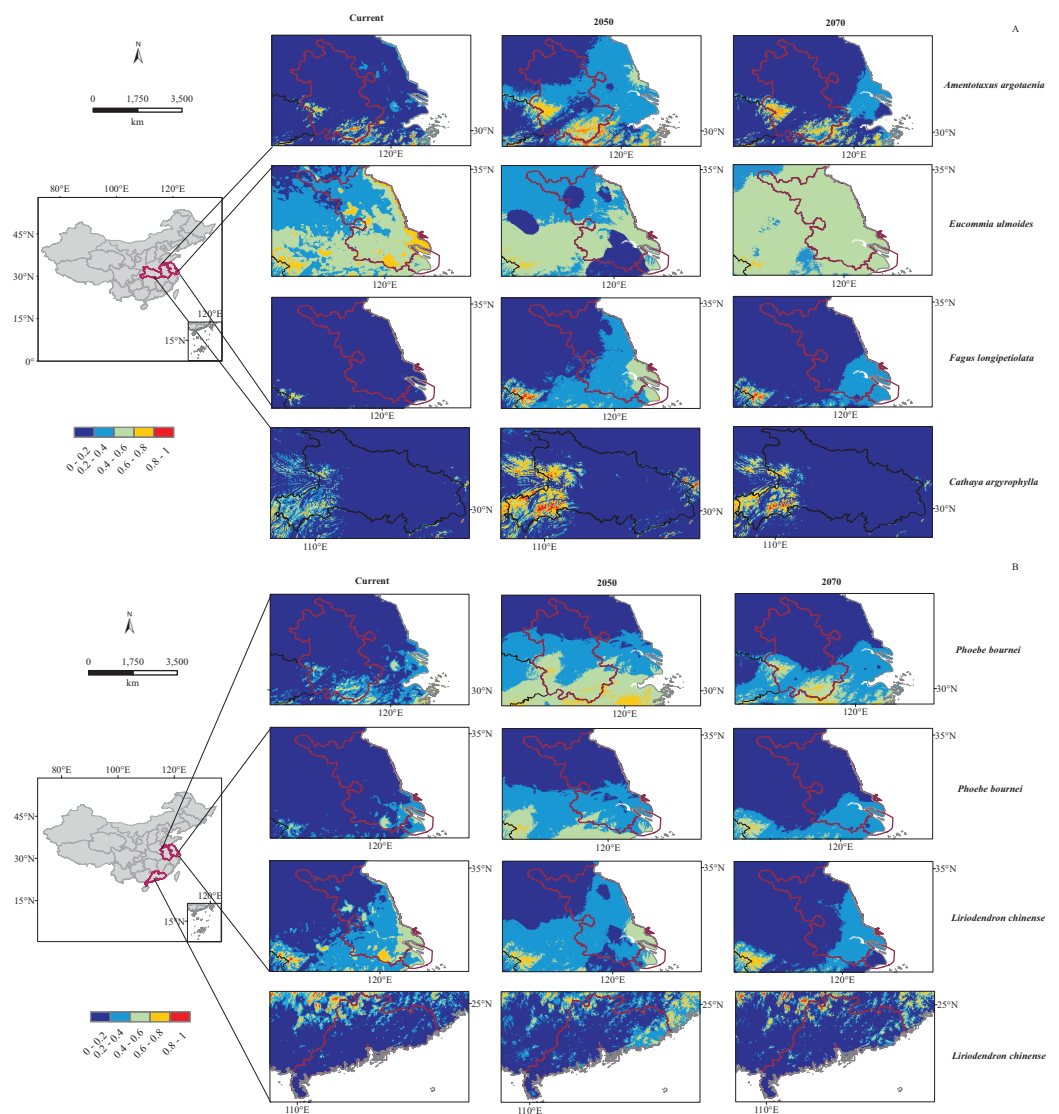
Figure 4 Predicted distribution of six dark species under current and future (2050, 2070) bioclimatic scenarios in potential regions. (A) For *Amentotaxus argotaenia*, *Eucommia ulmoides*, *Cathaya argyrophylla*, *Fagus longipetiolata*. (B) For *Liriodendron chinense* and *Phoebe bournei*. Red and blue indicate a high and low probability of occurrence, respectively. Full-size ☑ DOI: 10.7717/peerj.6731/fig-4

and we believe it is valuable for ex situ conversation activities. Currently, the most common approach, that is, used in ex situ conservation is to collect rare species and grow them in plantations or resource nurseries and to maintain their seed banks (*Hawkins, Sharrock & Havens, 2008*). However, this requires the reestablishment of plants, which currently has a low success rate globally (*Polak & Saltz, 2011*; *Sheean, Manning & Lindenmayer, 2012*). MaxEnt modeling has already been used in species reintroduction and ecological restoration studies (*Angelieri et al., 2016*; *Ardestani et al., 2015*; *Hiers et al., 2012*; *Remya, Ramachandran & Jayakumar, 2015*; *Yang et al., 2013*), but its combination with dark diversity modeling will allow more suitable species and broader habitat to be found for species conservation. In spite of that, this combination can be applied to research

on species reintroduction, ecological restoration and preventing exotic invasion. For example, possibly suitable areas predication can reduce labor and financial costs and increase the success rate of reintroductions.

At the same time, we have noticed that sometimes the results from two models may have a little contradiction, and we think the reason could be different types of data subject they were using or the different theories they were based on. For example, in our study, *F. longipetiolata* is a dark species in Jiangsu province while MaxEnt showed low climatic suitability for it there (Fig. 4). In addition, MaxEnt was only trained with climatic variables, but some other variables not covered in this study, such as anthropogenic activities, invasive species, atmospheric carbon dioxide, downward radiation, and evolutionary history may also affect the species distribution (*Cardinale et al., 2012*; *Doblas-Miranda et al., 2017*; *Woodward & Kelly, 2008*). Moreover, as a present-background (present-only) based model, MaxEnt can only estimate the relative likelihood and probability, not probability of occurrence, which can be estimated with presence-absence and occupancy-detection models (*Gomes et al., 2018*; *Guillera-Arroita et al., 2015*), and this may explain the differences from the result of dark diversity. Moreover, the use of one climate model scenario maybe also a potential limitation.

CONCLUSIONS

Dark diversity is a new and useful concept to indicate how many species hidden behind its observed species in a community (*Pärtel, Szava-Kovats & Zobel, 2011*, *2013*). Like other studies, we mapped the results of dark diversity in China, which shows that more and more plants could be potentially distributed and restored in the east of China. In spite of that, we combined the results of dark diversity and MaxEnt, and predicted the climatically suitable areas presently and in the future for six threatened species, which can be valuable for their conversation. This combination can also make a contribution to restoration efforts and invasion risk assessment of alien species.

ACKNOWLEDGEMENTS

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Lili Tang conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, approved the final draft.
- Runxi Wang contributed reagents/materials/analysis tools, authored or reviewed drafts of the paper.
- Kate S. He authored or reviewed drafts of the paper.
- Cong Shi authored or reviewed drafts of the paper.
- Tong Yang authored or reviewed drafts of the paper.
- Yaping Huang authored or reviewed drafts of the paper.
- Pufan Zheng authored or reviewed drafts of the paper.
- Fuchen Shi conceived and designed the experiments, approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The raw measurements are available in the Supplemental Files.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.6731#supplemental-information.

REFERENCES

Angelieri CCS, Adams-Hosking C, De Barros Ferraz KMPM, De Souza MP, McAlpine CA. 2016. Using species distribution models to predict potential landscape restoration effects on puma conservation. *PLOS ONE* **11(1)**:e0145232 DOI 10.1371/journal.pone.0145232.

Ardestani EG, Tarkesh M, Bassiri M, Vahabi MR. 2015. Potential habitat modeling for reintroduction of three native plant species in central Iran. *Journal of Arid Land* **7(3)**:381–390 DOI 10.1007/s40333-014-0050-4.

Beals EW. 1984. Bray-Curtis ordination: an effective strategy for analysis of multivariate ecological data. *Advances in Ecological Research* **14**:1–55.

Beer C, Reichstein M, Tomelleri E, Ciais P, Jung M, Carvalhais N, Rodenbeck C, Arain MA, Baldocchi D, Bonan GB, Bondeau A, Cescatti A, Lasslop G, Lindroth A, Lomas M, Luyssaert S, Margolis H, Oleson KW, Roupsard O, Veenendaal E, Viovy N, Williams C, Woodward FI, Papale D. 2010. Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. *Science* **329(5993)**:834–838 DOI 10.1126/science.1184984.

Breiman L. 2001. Random forests. *Machine Learning* **45(1)**:5–32.

Bueno ML, Pennington RT, Dexter KG, Kamino LHY, Pontara V, Neves DM, Ratter JA, De Oliveira-Filho AT. 2017. Effects of quaternary climatic fluctuations on the distribution of Neotropical savanna tree species. *Ecography* **40(3)**:403–414 DOI 10.1111/ecog.01860.

Cárdenas ML, Gosling WD, Sherlock SC, Poole I, Pennington RT, Mothes P. 2011. The response of vegetation on the andean flank in western amazonia to pleistocene climate change. *Science* **331(6020)**:1055–1058 DOI 10.1126/science.1197947.

Cardinale BJ, Duffy JE, Gonzalez A, Hooper DU, Perrings C, Venail P, Narwani A, Mace GM, Tilman D, Wardle DA, Kinzig AP, Daily GC, Loreau M, Grace JB, Larigauderie A, Srivastava DS, Naeem S. 2012. Biodiversity loss and its impact on humanity. *Nature* **486(7401)**:59–67 DOI 10.1038/nature11148.

Collins M, Knutti R, Arblaster J, Dufresne J-L, Fichefet T, Friedlingstein P, Gao X, Gutowski WJ, Johns T, Krinner G, Shongwe M, Tebaldi C, Weaver AJ, Wehner M. 2013. Chapter 12—long-term climate change: projections, commitments and irreversibility. In: Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM, eds. *Climate Change 2013: The Physical Science Basis. IPCC Working Group I Contribution to AR5 IPCC*. Cambridge: Cambridge University.

Da Silva Chagas C, De Carvalho Junior W, Bhering SB, Filho BC. 2016. Spatial prediction of soil surface texture in a semiarid region using random forest and multiple linear regressions. *CATENA* **139**:232–240 DOI 10.1016/j.catena.2016.01.001.

Dawson TP, Jackson ST, House JI, Prentice IC, Mace GM. 2011. Beyond predictions: biodiversity conservation in a changing climate. *Science* **332(6025)**:53–58 DOI 10.1126/science.1200303.

De Bello F, Fibich P, Zelený D, Kopecký M, Mudrák Ořej, Chytrý M, Pyšek P, Wild J, Michalcová D, Sádlo Jří, Šmilauer P, Lepš J, Pärtel M. 2016. Measuring size and composition of species pools: a comparison of dark diversity estimates. *Ecology and Evolution* **6(12)**:4088–4101 DOI 10.1002/ece3.2169.

Deb JC, Phinn S, Butt N, McAlpine CA. 2017. The impact of climate change on the distribution of two threatened Dipterocarp trees. *Ecology and Evolution* **7(7)**:2238–2248 DOI 10.1002/ece3.2846.

Doblas-Miranda E, Alonso R, Arnan X, Bermejo V, Brotons L, De las Heras J, Estiarte M, Hódar JA, Llorens P, Lloret F, López-Serrano FR, Martínez-Vilalta J, Moya D, Peñuelas J, Pino J, Rodrigo A, Roura-Pascual N, Valladares F, Vilà M, Zamora R, Retana J. 2017. A review of the combination among global change factors in forests, shrublands and pastures of the mediterranean region: beyond drought effects. *Global and Planetary Change* **148**:42–54 DOI 10.1016/j.gloplacha.2016.11.012.

Elith J, Graham CH. 2009. Do they? How do they? Why do they differ? On finding reasons for differing performances of species distribution models. *Ecography* **32(1)**:66–77 DOI 10.1111/j.1600-0587.2008.05505.x.

Elith J, Kearney M, Phillips S. 2010. The art of modelling range-shifting species. *Methods in Ecology and Evolution* **1(4)**:330–342 DOI 10.1111/j.2041-210X.2010.00036.x.

Ellenberg H, Weber H, Dull R, Wirth V, Werner W, Paulissen D. 1991. Zeigwerte von Pflanzen in MittelEuropa. *Scripta Geobot* **18**:1–24.

Franklin J, Serra-Diaz JM, Syphard AD, Regan HM. 2016. Global change and terrestrial plant community dynamics. *Proceedings of the National Academy of Sciences of the United States of America* **113(14)**:3725–3734 DOI 10.1073/pnas.1519911113.

Gomes VHF, IJff SD, Raes N, Amaral ILão, Salomão RP, De Souza Coelho L, De Almeida Matos FD, Castilho CV, De Andrade Lima Filho D, López DC, Guevara JE,

Magnusson WE, Phillips OL, Wittmann F, De Jesus Veiga Carim M, Martins MP, Irume MV, Sabatier D, Molino J-F, Bánki OS, Da Silva Guimarães JR, Pitman NCA, Piedade MTF, Mendoza AM, Luize BG, Venticinque EM, De Leão Novo EMM, Vargas PN, Silva TSF, Manzatto AG, Terborgh J, Reis NFC, Montero JC, Casula KR, Marimon BS, Marimon B-H, Coronado ENH, Feldpausch TR, Duque A, Zartman CE, Arboleda NC, Killeen TJ, Mostacedo B, Vasquez R, Schöngart J, Assis RL, Medeiros MB, Simon MF, Andrade A, Laurance WF, Camargo JL, Demarchi LO, Laurance SGW, De Sousa Farias E, Nascimento HEM, Revilla JDC, Quaresma A, Costa FRC, Vieira ICG, Cintra BBL, Castellanos H, Brienen R, Stevenson PR, Feitosa Y, Duivenvoorden JF, Aymard CGA, Mogollón HF, Targhetta N, Comiskey JA, Vicentini A, Lopes A, Damasco G, Dávila N, García-Villacorta R, Levis C, Schietti J, Souza P, Emilio T, Alonso A, Neill D, Dallmeier F, Ferreira LV, Araujo-Murakami A, Praia D, Do Amaral DD, Carvalho FA, De Souza FC, Feeley K, Arroyo L, Pansonato MP, Gribel R, Villa B, Licona JC, Fine PVA, Cerón C, Baraloto C, Jimenez EM, Stropp J, Engel J, Silveira M, Mora MCP, Petronelli P, Maas P, Thomas-Caesar R, Henkel TW, Daly D, Paredes MR, Baker TR, Fuentes A, Peres CA, Chave J, Pena JLM, Dexter KG, Silman MR, Jørgensen PM, Pennington T, Di Fiore A, Valverde FC, Phillips JF, Rivas-Torres G, Von Hildebrand P, Van Andel TR, Ruschel AR, Prieto A, Rudas A, Hoffman B, Vela CIA, Barbosa EM, Zent EL, Gonzales GPG, Doza HPD, De Andrade Miranda IP, Guillaumet J-L, Pinto LFM, De Matos Bonates LC, Silva N, Gómez RZ, Zent S, Gonzales T, Vos VA, Malhi Y, Oliveira AA, Cano A, Albuquerque BW, Vriesendorp C, Correa DF, Torre EV, Van Der Heijden G, Ramirez-Angulo H, Ramos JF, Young KR, Rocha M, Nascimento MT, Medina MNU, Tirado M, Wang O, Sierra R, Torres-Lezama A, Mendoza C, Ferreira C, Baider C, Villarroel D, Balslev H, Mesones I, Giraldo LEU, Casas LF, Reategui MAA, Linares-Palomino R, Zagt R, Cárdenas S, Farfan-Rios W, Sampaio AF, Pauletto D, Sandoval EHV, Arevalo FR, Huamantupa-Chuquimaco I, Garcia-Cabrera K, Hernandez L, Gamarra LV, Alexiades MN, Pansini S, Cuenca WP, Milliken W, Ricardo J, Lopez-Gonzalez G, Pos E, Ter Steege H. 2018. Species distribution modelling: contrasting presence-only models with plot abundance data. *Scientific Reports* **8(1)**:1003 DOI 10.1038/s41598-017-18927-1.

Gray CL, Hill SLL, Newbold T, Hudson LN, Börger L, Contu S, Hoskins AJ, Ferrier S, Purvis A, Scharlemann JPW. 2016. Local biodiversity is higher inside than outside terrestrial protected areas worldwide. *Nature Communications* **7(1)**:12306 DOI 10.1038/ncomms12306.

Guillera-Arroita G, Lahoz-Monfort JJ, Elith J, Gordon A, Kujala H, Lentini PE, McCarthy MA, Tingley R, Wintle BA. 2015. Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecology and Biogeography* **24(3)**:276–292 DOI 10.1111/geb.12268.

Hawkins B, Sharrock S, Havens K. 2008. *Plants and climate change: which future?* Richmond: Botanic Gardens Conservation International.

Hiers JK, Mitchell RJ, Barnett A, Walters JR, Mack M, Williams B, Sutter R. 2012. The dynamic reference concept: measuring restoration success in a rapidly changing no-analogue future. *Ecological Restoration* **30(1)**:27–36 DOI 10.3368/er.30.1.27.

Hill MO, Mountford J, Roy D, Bunce RGH. 1999. *Ellenberg's indicator values for British plants. ECOFACT Volume 2 Technical Annex*. Huntingdon: Institute of Terrestrial Ecology, 46.

IPCC. 2018. Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. *Available at https://www.ipcc.ch/sr15/*.

Kasari L, Saar L, De Bello F, Takkis K, Helm A. 2016. Hybrid ecosystems can contribute to local biodiversity conservation. *Biodiversity and Conservation* **25(14)**:3023–3041 DOI 10.1007/s10531-016-1218-3.

Kelly AE, Goulden ML. 2008. Rapid shifts in plant distribution with recent climate change. *Proceedings of the National Academy of Sciences of the United States of America* **105(33)**:11823–11826 DOI 10.1073/pnas.0802891105.

Kier G, Mutke J, Dinerstein E, Ricketts TH, Küper W, Kreft H, Barthlott W. 2005. Global patterns of plant diversity and floristic knowledge. *Journal of Biogeography* **32(7)**:1107–1116 DOI 10.1111/j.1365-2699.2005.01272.x.

Lewis RJ, De Bello F, Bennett JA, Fibich P, Finerty GE, Götzenberger L, Hiiesalu I, Kasari L, Lepš J, Májeková M, Mudrák Ořej, Riibak K, Ronk A, Rychtecká T, Vitová A, Pärtel M. 2017. Applying the dark diversity concept to nature conservation. *Conservation Biology: The Journal of the Society for Conservation Biology* **31(1)**:40–47 DOI 10.1111/cobi.12723.

Lewis RJ, Szava-Kovats R, Pärtel M. 2016. Estimating dark diversity and species pools: an empirical assessment of two methods. *Methods in Ecology and Evolution* **7(1)**:104–113 DOI 10.1111/2041-210X.12443.

Liaw A, Wiener M. 2002. Classification and regression by randomForest. *R News* **2(3)**:18–22.

Loarie SR, Duffy PB, Hamilton H, Asner GP, Field CB, Ackerly DD. 2009. The velocity of climate change. *Nature* **462(7276)**:1052–1055 DOI 10.1038/nature08649.

Menéndez R, Megías AG, Hill JK, Braschler B, Willis SG, Collingham Y, Fox R, Roy DB, Thomas CD. 2006. Species richness changes lag behind climate change. *Proceedings of the Royal Society B: Biological Sciences* **273(1593)**:1465–1470 DOI 10.1098/rspb.2006.3484.

Mischler P, Kearney M, McCarroll JC, Scholte RG, Vounatsou P, Malone JB. 2012. Environmental and socio-economic risk modelling for Chagas disease in Bolivia. *Geospatial Health* **6(3)**:59–66 DOI 10.4081/gh.2012.123.

Moeslund JE, Brunbjerg AK, Clausen KK, Dalby L, Fløjgaard C, Juel A, Lenoir J. 2016. Dark diversity illuminates the dim side of conservation and restoration. *biorxiv preprint* DOI 10.1101/057315.

Morales NS, Fernández IC, Baca-González V. 2017. MaxEnt's parameter configuration and small samples: are we paying attention to recommendations? A systematic review. *PeerJ* **5(2)**:e3093 DOI 10.7717/peerj.3093.

Münzbergová Z, Herben T. 2004. Identification of suitable unoccupied habitats in metapopulation studies using co-occurrence of species. *Oikos* **105(2)**:408–414 DOI 10.1111/j.0030-1299.2004.13017.x.

Mutanga O, Adam E, Cho MA. 2012. High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm. *International Journal of Applied Earth Observation and Geoinformation* **18**:399–406.

Noulekoun F, Chude S, Zenebe A, Birhane E. 2017. Climate change impacts on *Faidherbia albida* (Delile) A. Chev. distribution in dry lands of Ethiopia. *African Journal of Ecology* **55(2)**:233–243 DOI 10.1111/aje.12345.

Oksanen J, Guillaume Blanchet F, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H. 2016. *vegan: community ecology package*. R package version: 2.4-1. *Available at http://cran.r-project.org/package=vegan*.

Pärtel M, Szava-Kovats R, Zobel M. 2011. Dark diversity: shedding light on absent species. *Trends in Ecology & Evolution* **26(3)**:124–128 DOI 10.1016/j.tree.2010.12.004.

Pärtel M, Szava-Kovats R, Zobel M. 2013. Community completeness: linking local and dark diversity within the species pool concept. *Folia Geobotanica* **48(3)**:307–317 DOI 10.1007/s12224-013-9169-x.

Pärtel M, Zobel M, Zobel K, Van Der Maarel E. 1996. The species pool and its relation to species richness: evidence from Estonian plant communities. *Oikos* **75(1)**:111–117 DOI 10.2307/3546327.

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190(3–4)**:231–259 DOI 10.1016/j.ecolmodel.2005.03.026.

Phillips SJ, Dudik M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31(2)**:161–175 DOI 10.1111/j.0906-7590.2008.5203.x.

Polak T, Saltz D. 2011. Reintroduction as an ecosystem restoration technique. *Conservation Biology* **25(3)**:424 DOI 10.1111/j.1523-1739.2011.01669.x.

R Development Core Team. 2016. *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing. *Available at https://www.r-project.org*.

Remya K, Ramachandran A, Jayakumar S. 2015. Predicting the current and future suitable habitat distribution of *Myristica dactyloides* Gaertn. using MaxEnt model in the Eastern Ghats, India. *Ecological Engineering* **82**:184–188 DOI 10.1016/j.ecoleng.2015.04.053.

Riibak K, Reitalu T, Tamme R, Helm A, Gerhold P, Znamenskiy S, Bengtsson K, Rosén E, Prentice HC, Pärtel M. 2015. Dark diversity in dry calcareous grasslands is determined by dispersal ability and stress-tolerance. *Ecography* **38(7)**:713–721 DOI 10.1111/ecog.01312.

Hijmans RJ. 2018. *raster: geographic data analysis and modeling.* R package version 2.8-4. *Available at https://CRAN.R-project.org/package=raster*.

Ronk A, De Bello F, Fibich P, Pärtel M. 2016. Large-scale dark diversity estimates: new perspectives with combined methods. *Ecology and Evolution* **6(17)**:6266–6281 DOI 10.1002/ece3.2371.

Ronk A, Szava-Kovats R, Pärtel M. 2015. Applying the dark diversity concept to plants at the European scale. *Ecography* **38(10)**:1015–1025 DOI 10.1111/ecog.01236.

Ronk A, Szava-Kovats R, Zobel M, Pärtel M. 2017. Observed and dark diversity of alien plant species in Europe: estimating future invasion risk. *Biodiversity and Conservation* **26(4)**:899–916 DOI 10.1007/s10531-016-1278-4.

Sheean VA, Manning AD, Lindenmayer DB. 2012. An assessment of scientific approaches towards species relocations in Australia. *Austral Ecology* **37(2)**:204–215 DOI 10.1111/j.1442-9993.2011.02264.x.

Storkey J, Macdonald AJ, Poulton PR, Scott T, Köhler IH, Schnyder H, Goulding KWT, Crawley MJ. 2015. Grassland biodiversity bounces back from long-term nitrogen addition. *Nature* **528(7582)**:401–404 DOI 10.1038/nature16444.

Swets J. 1988. Measuring the accuracy of diagnostic systems. *Science* **240(4857)**:1285–1293 DOI 10.1126/science.3287615.

Tognelli MF, Roig-JuÑEnt SA, Marvaldi AE, Flores GE, Lobo JM. 2009. An evaluation of methods for modelling distribution of Patagonian insects. *Revista Chilena de Historia Natural* **82(3)**:347–360 DOI 10.4067/S0716-078X2009000300003.

Turner W. 2014. Sensing biodiversity. *Science* **346(6207)**:301–302 DOI 10.1126/science.1256014.

Warren DL, Seifert SN. 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications* **21(2)**:335–342 DOI 10.1890/10-1171.1.

Wiesmeier M, Barthold F, Blank B, Kögel-Knabner I. 2011. Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. *Plant and Soil* **340(1–2)**:7–24 DOI 10.1007/s11104-010-0425-z.

Williams JW, Jackson ST, Kutzbach JE. 2007. Projected distributions of novel and disappearing climates by 2100 AD. *Proceedings of the National Academy of Sciences of the United States of America* **104(14)**:5738–5742 DOI 10.1073/pnas.0606292104.

Williams JN, Seo C, Thorne J, Nelson JK, Erwin S, O'Brien JM, Schwartz MW. 2009. Using species distribution models to predict new occurrences for rare plants. *Diversity and Distributions* **15(4)**:565–576 DOI 10.1111/j.1472-4642.2009.00567.x.

Woodward FI, Kelly CK. 2008. Responses of global plant diversity capacity to changes in carbon dioxide concentration and climate. *Ecology Letters* **11(11)**:1229–1237 DOI 10.1111/j.1461-0248.2008.01240.x.

Wu XY, Dong SK, Liu SL, Su XK, Han YH, Shi JB, Zhang Y, Zhao ZZ, Sha W, Zhang X, Gao F, Xu DH. 2017. Predicting the shift of threatened ungulates' habitats with climate change in Altun Mountain National Nature Reserve of the Northwestern Qinghai-Tibetan Plateau. *Climatic Change* **142(3–4)**:331–344 DOI 10.1007/s10584-017-1939-7.

Yang X-Q, Kushwaha SPS, Saran S, Xu J, Roy PS. 2013. Maxent modeling for predicting the potential distribution of medicinal plant, *Justicia adhatoda* L. in Lesser Himalayan foothills. *Ecological Engineering* **51**:83–87 DOI 10.1016/j.ecoleng.2012.12.004.

Ying Z. 2012. Projections of 2.0°C warming over the globe and China under RCP4.5. *Atmospheric and Oceanic Science Letters* **5(6)**:514–520 DOI 10.1080/16742834.2012.11447047.