Using species distribution models to assess the importance of Egypt's Protected Areas for the conservation of medicinal plants

Emad Kaky & Francis Gilbert*

School of Life Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK

*Author for correspondence

Abstract

Human activities affect the distribution and abundance of plants, with impacts on ecosystem services and human well-being; it is thus vital that a network of Protected Areas is capable of conserving plants that are useful. Using the species distribution (SDM) model algorithm MaxEnt, we tested whether Egypt's network of Protected Areas performs well in conserving the region's important medicinal plant species. We constructed individual SDMs for each species, and then combined the models into a single 'species-richness' layer, which we then compared to the distribution of the existing Protected Areas. Temperature was the most important of eleven predictor variables used to build the SDMs. Assuming the SDM's prediction of suitable habitat was accurate and corresponded to the occurrence each of the medicinal plant species, then on average species richness was significantly higher within than outside the Protected Areas. Based on our findings, Egypt's Protected Areas are effective at conserving its medicinal plants.

Key words: MaxEnt, Egypt, ecosystem services, conservation.

Introduction

Human activities are having a strong impact on plant abundance and distribution, with consequent effects on ecosystem services and human well-being (Klein *et al.*, 2008). This growing effect of human activities on biodiversity (Chapin *et al.,* 2000) creates an urgent need to understand the elements that determine the distribution and abundance of plant in order to enhance their conservation (Dubuis *et al.*, 2011). The identification of species-rich regions and those where geographically limited species co-occur can optimise the creation of Protected Areas (Bojórquez-Tapia *et al.*, 1995).

Medicinal plants are one of the most important elements of biodiversity around the world (Klein *et al.*, 2008; Okigbo *et al.*, 2008) because of their role in ecosystem services such as healthcare, cultural value and heritage, local economics and human well-being, especially in poor areas (Klein *et al.*, 2008; Okigbo *et al.*, 2008). Conserving and protecting these kinds of species is vital, including improving knowledge about the important ecological requirements of medicinal plants, and raising awareness among all stakeholders to protect this heritage. Consequently, conservation planning and effective management is important in protecting the most threatened species in order to avoid declines in the diversity of medicinal plants.

Species distribution models (SDMs) can predict the geographic distribution of individual species using locality data and ecological variables as predictors (Franklin, 2009). While occurrence records can be harvested from museums/herbaria, published reports, and original fieldwork, accurately identifying whether a species is truly absent is exceedingly difficult. To address this challenge, several SDM algorithms have been designed to employ only positive presence data (Phillips *et al.*, 2006). One such SDM algorithm, MaxEnt, has been shown to be one of the most effective tools for accurately predicting species distributions (Elith *et al.*, 2006). SDMs using MaxEnt offer a valuable tool for creating general patterns of species richness without needing to analyse the specific quality or precision of the predictions for every individual species (Pineda and Lobo, 2009). Several studies have added together the models for individual species to create maps of species richness, the approach we adopt here. For example, Ortega-Huerta and Peterson (2004) added the individual maps of 285 bird and 114 mammal species of part of Mexico to create a map of species richness; Newbold *et al.* (2009) and Pineda and Lobo (2009) used the same approach for Egyptian mammals and butterflies, and Mexican amphibians respectively, as did de Pous *et al.* (2011) on Moroccan reptiles. It is exciting that the same approach can be used to project into the future under climate change (Distler *et al.*, 2015), as we have also done (Kaky & Gilbert, in prep.). Ideally the maps of predicted species richness should be validated using independent data (Pineda and Lobo, 2009). Such species-richness maps make it possible to distinguish hotspots of species richness (Newbold *et al.*, 2010), and hence to select feasible regions for conservation relatively objectively (Pressey *et al.*, 1993). This is a powerful tool to help build conservation efforts or anticipate the future of biodiversity under worldwide climate change (Algar *et al.*, 2009; Distler *et al.*, 2015).

The climatic predictors used in our SDMs should be very suitable for plants. The physiological toleration hypothesis suggests that plant species richness is most elevated in warm and/or wet environments because a more extensive range of functions can persevere under such circumstances (Spasojevic *et al.*, 2014). For instance, Hawkins *et al.* (2003) found that a measure of the balance between energy and water nearly always described spatial difference in species richness better than other environmental variables. In warm regions of the tropics and subtropics, the most robust predictors are typically water variables, while water/energy variables (for plants) or energy predictors (for animals) predominate in high latitudes (Hawkins *et al.*, 2003).

PAs currently cover about 12% of the terrestrial surface of the earth (Seiferling *et al.*, 2012), while those that have been declared in Egypt cover 15% of the total land area (El-Gabbas *et al.*, 2016). The 30 Egyptian Pas were all established since 1983, based on the recommendations of experts familiar with Egyptian biodiversity (Newbold *et al.*, 2009). An obvious issue is the extent to which these PAs are capable of conserving Egypt’s fauna and flora: a basic requirement is that they contain a high proportion of the biodiversity of the country. Thus ideally there should be higher species richness within the PAs than outside them. Several studies have measured this: for example, Sciberras *et al.* (2013) showed that the density and biomass of fish and invertebrates inside partially protected areas was higher than in unprotected areas; Newbold *et al.* (2009) and Lee *et al.* (2007) found that species richness inside PAs was higher than outside, but others found the reverse (Pawar *et al.*, 2007; Traba *et al.*, 2007). Human activities are one of the main reasons for declines both inside and especially outside PAs: thus forest cover decreased between 1980 and 2001 in areas surrounding most tropical PAs (Defries *et al.*, 2005), and one might anticipate similar declines in the fauna. The active management of PAs needs many more such comparisons to guide management decisions (Linkie *et al.*, 2006).

Our objective is therefore to assess the role of the network of Egyptian PAs in conserving medicinal plants by comparing their diversity within and just outside each PA, averaging this difference across all the PAs. We did this by predicting the distribution of each species using SDMs, and summing together all the SDMs to create two kinds of species-richness maps (by either using or not using thresholds to binarize the predicted habitat suitabilities). We then use these maps to assess the predicted species richness inside and outside Egypt's PAs.

Methods

We used data for 121 medicinal plant species of the Egyptian flora. The occurrence data for these species were collated by the BioMAP project (http://www.biomapegypt.org/), a project run from Cairo in 2004–2008 and funded by Italian Debt Swap. The data are presence-only records collected from different sources (i.e. literature, herbarium, and field work). To avoid inaccurate predictions, we deleted species with fewer than ten records to avoid overfitting (Baldwin, 2009), species with more than ten but spatially very restricted records, and the one species whose SDM had a mean AUC less than 0.7 (Franklin, 2009). We ended up with 114 species of Egyptian medicinal plants, with 14396 point records.

The environmental variables used in this study were 23 predictors, 19 of them (Bio layers) downloaded from the WorldClim v1.4 dataset at resolution of 2.5 arc-minutes (http://www.worldclim.org/bioclim) (Hijmans et al., 2005) (Table 1). Normalized Difference Vegetation Index (NDVI) data for seven years (2004 to 2010) were downloaded from the Spot Vegetation website (http://free.vgt.vito.be/) and used to create two layers: maximum NDVI (Max_NDVI), and the difference between the Minimum and Maximum NDVI values (NDVI_differences). A further environment layer was a habitat layer, derived from the Biomap project, which divided Egypt's terrain into eleven classes ("sea, littoral coastal land, cultivated land, sand dune, wadi, metamorphic rock, igneous rock, gravels, serir sand sheets, sabkhas and sedimentary rocks") (for more detail, see (Newbold *et al.*, 2009). Altitude data were downloaded from http://www.cgiar-csi.org/data/elevation and the resolution rescaled from 90 m to be 2.5 arc-minutes (see (El-Gabbas *et al.*, 2016). Eleven of the 23 environmental variables (see Table 1) remained for use after 12 were removed based on collinearity analysis using the Variance Inflation Factor, implemented in R v2.15 (the 'car' package: R Development Core Team 2012).

We used Maximum Entropy (MaxEnt) version 3.3.3k (Phillips et al., 2006) (downloaded from: http://www.cs.princeton.edu/~schapire/maxent/) to run the models, choosing a set of options (i.e. feature classes QPT, 10000 background points, 1000 iterations, cross-validation with 10 replications, 10% training presence threshold, and logistic output format) to create both 'probability' (i.e. raw values of habitat suitability) and 'binary' (predicted 'suitable'/'unsuitable' via thresholding) maps. MaxEnt performance is good with presence-only data and small numbers of records (Elith *et al.*, 2006; Franklin, 2009), and its performance is good in comparison with other algorithms (Elith *et al.*, 2006). The options were chosen after exhaustive runs with different option combinations (of feature classes, number of background points, number of iterations and regularization values) to obtain the best models. Two statistics were used to evaluate the accuracy of each model, the AUC, and the true skill statistic (TSS) (Allouche *et al.*, 2006). TSS values lie between -1 and +1: close to +1 indicates perfect performance, while close to zero or less than zero point to model performance no better than random (details, see Allouche et al. 2006). (For details of each SDM, see Supplementary Table S1.)

The relative importance of the environmental predictors can be determined in three ways by Maxent (percent contribution, permutation importance, jacknife: (Phillips *et al.*, 2006)). Care is needed when there are high correlations between variables, but pre-screening variables for collinearity (as we have done) minimises this problem. Here we used permutation importance to determine the importance of the environmental predictors, calculated by permuting the values of each predictor and calculating the resulting reduction in the training AUC: a large reduction shows that the model is influenced by that predictor. The values are standardized to a percentage (Phillips *et al.*, 2006).

We created two kinds of maps of the distribution of species richness. The first was the ‘probability’ map, made manually by obtaining the average of the replicate ascii files obtained from Maxent for each species, and then adding all the species layers together using the ‘raster calculator’ of ArcGIS10.2.2. This map was then rescaled to fit the same range as the second type, the ‘binary’ map, which is the product of adding together the binary maps for each species. The binary map converts each pixel value of the MaxEnt output (a continuous value between 0 and 1) into binary data (predicted suitable/ unsuitable) by choosing a threshold rule (see Liu *et al.*, 2005). We chose the "10% training presence" as our threshold rule (El-Gabbas *et al.*, 2016), which produced a binary map for each of the 10 replicates for each species. Subsequently we produced a single consensus binary map for each species by allocating ‘suitable’ to a pixel that had ‘suitable’ values in more than 50% of the model runs (i.e. >5 replicates). Then we added together all the species maps to create the ‘binary map’ for species richness.

Finally we compared the species richness inside and outside PAs. First we chose at random 2000 pixels from the map. A 50-km buffer zone was created around each PA, and the random pixels that lay within each PA and within each buffer zone identified. The mean species richness for the random pixels within each PA ('within') and within its buffer zone ('outside') created paired values inside and outside each PA. We then compared the average difference (within - outside) using a paired t-test.

Results

In terms of mean AUC values, all models showed good performance (range 0.802 to 0.989) (mean = 0.901 ± 0.0036), as do the TSS scores (mean TSS across all species 0.63 ± 0.01). The lowest mean AUC value was recorded for *Pluchea dioscoridis* and the highest for *Solanum elaeagnifolium* (for more details see Supplementary Table S1). High mean AUC values were not just limited to species with few records, since there were several species with large numbers of records which achieved very good performance. There were 12 species with mean AUC values of 0.80 – 0.85, 38 species between 0.85 – 0.90, 55 species between 0.90 – 0.95, and 10 with very high AUC between 0.95 – 1 (Fig. 1). There was no significant correlation between the mean AUC and the number of records used in the model (n=114, r=-0.052, P>0.05). In general, for the 10 replicates for each species there were not big differences between the AUC values for each run. The standard deviations ranged between 0.011 and 0.291, the smallest for *Lavandula pubescens* and the highest for *Herniaria hirsuta*. There were five species with a standard deviation between 0.2 – 0.3, 14 species between 0.2 – 0.1, and the rest (96 species) less than 0.1 (Table S1).

Environmental predictors that achieved highest permutation importance through all the modelled species, and the maximum contribution to the final models, were Bio6 (the minimum temperature of the coldest month), altitude, Bio3 (isothermality, the ratio of the mean monthly temperature range [max – min] and the maximum annual temperature range), Bio8 (the mean temperature of the wettest quarter), and Bio4 (temperature seasonality, the SD of monthly temperature). There were six variables with low permutation importance: Bio13 (precipitation of the wettest month) Bio15 (precipitation seasonality, the CV of monthly precipitation), habitat, Bio9 (mean temperature of the driest quarter), differences between maximum and minimum NDVI, and maximum NDVI (Fig. 2). Across all species, Bio6 was the highest for 36 species, followed by altitude (highest for 19 species), Bio3 and Bio8 (16 species) and Bio9 (see Fig. 3). Sometimes Bio15, Bio13, habitat and Bio9 achieved the highest mean permutation importance, but clearly these were not normally the most influential predictor.

In general, the occurrence locations (Fig 4) match well with both types of species richness maps (Fig 5 A & B). Species richness is highest from the southwest to the northeast, especially North and South Sinai, along the Mediterranean coast, and scattered areas of the Nile Delta. The probability species richness map (Fig. 5A) shows that the highest predicted species richness is situated in south Sinai, especially the area around St Katherine to Sharm El-Sheikh, to the Aqaba Gulf from Sharm El-Sheikh through Dahab to Taba, around El-Tur, some scattered locations between Abu Zneima to Suez, some scattered locations in North Sinai around Gebel Yillaq, El-Hassana, Gebel El-Hallal, Gebel El-Maghara, and some small areas on the border between Egypt and Israel, especially east of Gebel El-Hallal. All locations along the Mediterranean Sea coast from Rafah to Port Said are also suggested to have high species richness, especially from around Lake Bardawil to Mersa Martruh, and inland from Alexandria to Wadi El-Natrun (Supplementary Fig. S1).

In the binary richness map (Fig. 5B) the highest species richness is predicted to be located in north-eastern Egypt, especially in Sinai from the north to the mountain areas of the south, in the north particularly at Gebel Yillaq, El-Hassana, Gebel El-Hallal, Gebel El-Maghara, all the border between Egypt and Israel, the coastal regions of the Mediterranean Sea from Rafah to Port Said, and south of Gebel Yillaq and El-Hassana. In the south the highest predicted species richness is the area from St Katherine to Sharm El-Sheikh, then the entire border along the Aqaba Gulf and along the other side from St Katherine to El-Tur, and to Suez along the Red Sea. The highest predicted species richness is north of Suez to Ismailia, east and west of Ismailia, Greater Cairo, the Mediterranean Sea coast from Lake Manzala to Sallum, north of Wadi El-Natrun, Ain Sukhna, Gebel El-Gallala El-Bahariya, and from

Ras Zaafarana south to Ras Gharib, then from Ras Gharib to Hurghada, with some scattered locations at Gebel El-Gallala El-Qibliya. There are also some scattered areas between Mersa Alam to Berenice, and south of Halayeb (Supplementary Fig. S1).

The predicted species richness was significantly higher inside PAs than outside for both the binary map (paired t = 14.8, df = 24, p<0.001) (Figure 6A) and for the probability map (paired t = 9.9, df = 24, p<0.001) (Figure 6B).

Discussion

The most important result of this study was that the predicted species richness of medicinal plants was higher inside Egypt's PAs than outside, implying that the Protected Areas have been well located to implement the conservation of these important deliverers of a valuable ecosystem sevice.

Overall model performances were good in terms of the mean AUC scores. There are some studies which have recently criticized the use of AUC as an indicator for model accuracy (Austin, 2007; Lobo *et al.*, 2008), because of its bias caused by species with narrow ranges (Lobo *et al.*, 2008). Getting high AUC values it is easy when there are relatively few records (Jiménez-Valverde *et al.*, 2008; Lobo *et al.*, 2008), and therefore it is worth using other criteria such as the True Skill Statistic, although many recent studies still use AUC alone e.g. (Warren and Seifert, 2011; Beauregard and de Blois, 2014). When there is agreement between both validation methods, then we can assume good model performance (Beauregard and de Blois, 2014). In our data there was no significant correlation between the mean AUC values and the number of records, and hence we believe that sample size did not affect model performance (Elith *et al.*, 2006; de Pous *et al.*, 2011). Some other studies have achieved good model performance with large sample sizes (Kadmon *et al.*, 2003; Hernandez *et al.*, 2006), as we did.

In SDM studies, selecting appropriate environmental variables is very important because climate predictors are assumed to determine the distribution of species; a current topic of research is the extent to which biotic interactions affect distributions, but there is no consensus about how to allow for this (Wisz *et al.*, 2013). Robust models are produced by choosing the right predictors and modelling approach (Elith and Leathwick, 2009), which are then useful in conservation analysis (Austin, 2007; Araújo and Peterson, 2012). The most significant environmental variables in our study were the minimum temperature of the coldest month, followed by altitude; these make ecological sense in that temperature and elevation should predict much of the distribution of plant species in Egypt. (Newbold *et al.*, 2009) found that temperature was the major predictor of the distributions of Egyptian butterflies, again making perfect ecological sense. Some variables did not have much of an effect on species distributions (e.g. habitat, and NDVI): neither of the NDVI predictors provided useful information on Egyptian plant distributions. Some studies have found NDVI important (Anderson *et al.*, 2006), while some have not (El-Gabbas *et al.*, 2016). Most of Egypt is hyper-arid with extremely low NDVI values, so it is not surprising that NDVI is poor as a predictor. Habitat was not a powerful predictor either, perhaps related to its correlation with other predictors (e.g. altitude).

The predictions showed that the main hotspots of plant species richness are found in South Sinai, extending around the northern part of Egypt: this pattern occurs in both probability and binary species-richness maps. Similar studies on Egyptian animal taxa (Gilbert and Zalat, 2008; Basuony *et al.*, 2010; Leach *et al.*, 2013; El-Gabbas *et al.*, 2016) found high levels of predicted species richness around greater Cairo. This may be the result of spatial bias in the records, particularly of mammals. In the plant dataset, recent more systematic collecting has been undertaken in Sinai, and hence there is a different spatial bias in the data. However, the gradient from south-west to north-east in plant species richness is undoubtedly correct. The physiologically optimal temperature for most plants is between 10-35 ºC (Berry and Bjorkman, 1980), much more present in the north than in the south, although desert plants live in much higher temperature (Berry and Bjorkman, 1980) and most Egyptian habitats are deserts of one kind or another (90% of the land). Most areas in Egypt receive very much less than 80 mm precipitation annually, while the northern coastal areas can receive the highest recorded levels of up to 180-200 mm (El-Nahrawy, 2011) (albeit meagre by world standards).

Plant species richness for both the probability and binary maps was significantly higher inside Protected Areas than outside, as Newbold *et al.* (2009) found for Egyptian butterflies and mammals. Thus despite their relatively recent establishment, the locations of Egypt's PAs were well chosen. Sciberras *et al.* (2013) for marine reserves and Lee *et al.* (2007) for Sulawesi also found higher biomass inside PAs than outside, but other studies on Indian herpetofauna (Pawar *et al.*, 2007; Traba *et al.*, 2007) have found the converse, and some have found no differences (e.g. Joppa *et al.* (2008) showed that the vegetation inside and outside PAs in both the Amazon and Congo was very similar). Obviously PAs are generally established in places known to have high biodiversity, and the Egyptian PAs, although relatively new, were chosen carefully with expert knowledge (Newbold *et al.*, 2009). Alternatively, for older reserves, effective ecosystem management inside PAs could be one of the reasons for their high biodiversity (Thomas and Gillingham, 2015).

About 12% of global terrestrial habitat is covered by PAs, but many of them fail to protect biodiversity and ecological processes (Seiferling *et al.*, 2012). One of the main reasons for that failure is human activity changing the vegetation inside PAs and the areas around them (Defries *et al.*, 2005). It is important to sustain habitat heterogeneity within PAs and the surrounding areas to enable good management (Oliver *et al.*, 2010). There is clear evidence that forest cover has decreased from 1980 to 2001 in the areas neighbouring PAs in tropical regions. High human population densities and land-use isolate PAs from their surroundings (Joppa *et al.*, 2009).

In conclusion, the positions of Egypt's PAs appear to have been well chosen to maximise their potential effectiveness in conserving plant diversity, and their potential ability to preserve at least one important ecosystem service, that deriving from medicinal plants. A second conclusion we can draw is that species distribution modelling is an appropriate approach to measuring patterns of species richness in countries where information is sparse, and records may be the only available data. The models can predict new suitable locations for species that have not been surveyed very well (Franklin, 2009), helping to save time and costs. Thus SDMs represent a very useful tool to help plan the conservation process and suggest the locations of new PAs in such countries.

Acknowledgements

We thank the Higher Committee for Education Development in Iraq (HCED) for funding this work; and Ahmed El-Gabbas, Tim Newbold and Katie Leach for advice. We would like to thank two anonymous referees and an Associate Editor (Dr David Eldridge) for their constructive and helpful comments on a previous version of the manuscript.

References

Algar, A. C., Kharouba, H. M., Young, E. R. & Kerr, J. T. 2009. Predicting the future of species diversity: macroecological theory, climate change, and direct tests of alternative forecasting methods. *Ecography,* 32**,** 22-33.

Allouche, O., Tsoar, A. & Kadmon, R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology,* 43**,** 1223-1232.

Anderson, R. P., Peterson, A. T. & Egbert, S. L. 2006. Vegetation-Index Models Predict Areas Vulnerable To Purple Loosestrife (Lythrum Salicaria) Invasion In Kansas. *The Southwestern Naturalist,* 51**,** 471–480.

Araújo, M. B. & Peterson, A. T. 2012. Uses and misuses of bioclimatic envelope modeling. *Ecology,* 93**,** 1527–1539.

Austin, M. 2007. Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling,* 200**,** 1-19.

Baldwin, R. A. 2009. Use of Maximum Entropy Modeling in Wildlife Research. *Entropy,* 11**,** 854-866.

Basuony, M. I., Gilbert, F. & Zalat, S. 2010. Mammals of Egypt : atlas, red data listing & conservation. BioMAP & CultNat**,** EEAA & Bibliotheca Alexandrina, Cairo.

Beauregard, F. & de Blois, S. 2014. Beyond a climate-centric view of plant distribution: edaphic variables add value to distribution models. *PLoS One,* 9**,** e92642.

Berry, J. & Bjorkman, O. 1980. Photosynthetic Response And Adaptation To Temperature In Higher Plants. *Ann. Rev. Plant Physiol,* 31**,** 491-543.

Bojórquez-Tapia, L. A., Azuara, I., Ezcurra, E. & Flores-Villela, O. 1995. Identifying Conservation Priorities in Mexico Through Geographic Information Systems and Modeling. *Ecological Applications,* 5.

Chapin, F. S., Zavaleta, E. S., Eviner, V. T., Naylor, R. L. *et al.* 2000. Consequences of changing biodiversity. *Nature,* 405**,** 234-242.

de Pous, P., Beukema, W., Weterings, M., Dümmer, I. *et al.* 2011. Area prioritization and performance evaluation of the conservation area network for the Moroccan herpetofauna: a preliminary assessment. *Biodiversity and Conservation,* 20**,** 89-118.

Defries, R., Hansen, A., Newton, A. C. & Hansen, M. C. 2005. Increasing Isolation of Protected Areas in Tropical Forests over the past Twenty Years. *Ecological Applications,* 15**,** 19-26.

Distler, T., Schuetz, J. G., Velásquez-Tibatá, J., Langham, G. M. *et al.* 2015. Stacked species distribution models and macroecological models provide congruent projections of avian species richness under climate change. *Journal of Biogeography,* 42**,** 976-988.

Dubuis, A., Pottier, J., Rion, V., Pellissier, L. *et al.* 2011. Predicting spatial patterns of plant species richness: a comparison of direct macroecological and species stacking modelling approaches. *Diversity and Distributions,* 17**,** 1122-1131.

El-Gabbas, A., Baha El Din, S., Zalat, S. & Gilbert, F. 2016. Conserving Egypt's reptiles under climate change. *Journal of Arid Environments,* 127**,** 211-221.

El-Nahrawy, M. A. 2011. Country pasture/forage resource profiles. *Food and Agriculture Organization of the United Nations (FAO)*.

Elith, J., Graham, C. H., Anderson, R. P., Dudı´k, M. *et al.* 2006. Novel methods improve prediction of species’ distributions from occurrence data. *Ecography,* 29**,** 129-151.

Elith, J. & Leathwick, J. R. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics,* 40**,** 677-697.

Franklin, J. 2009. *Mapping species distributions: spatial inference and prediction,* Cambridge University Press, Cambridge, UK.

Gilbert, F. & Zalat, S. 2008. The Butterflies of Egypt: Atlas, Red Data listing & Conservation. BioMAP, EEAA, Cairo.

Hawkins, B. A., Field, R., Cornell, H. V., Currie, D. J. *et al.* 2003. Energy, water, and broad-scale geographic patterns of species richness. *Ecology,* 84.

Hernandez, P. A., Graham, C. H., Master, L. L. & Albert, D. L. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography,* 29**,** 773-785.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. *et al.* 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology,* 25**,** 1965-1978.

Jiménez-Valverde, A., Lobo, J. M. & Hortal, J. 2008. Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and Distributions,* 14**,** 885-890.

Joppa, L. N., Loarie, S. R. & Pimm, S. L. 2008. On the protection of "protected areas". *Proc Natl Acad Sci U S A,* 105**,** 6673-8.

Joppa, L. N., Loarie, S. R. & Pimm, S. L. 2009. On Population Growth Near Protected Areas. *PLoS One,* 4**,** e4279.

Kadmon, R., Farber, O. & Danin, A. 2003. A systematic analysis of factors affecting the performance of climatic envelope models. *Ecological Applications,* 13**,** 853-867.

Klein, J. A., Harte, J. & Zhao, X.-Q. 2008. Decline in Medicinal and Forage Species with Warming is Mediated by Plant Traits on the Tibetan Plateau. *Ecosystems,* 11**,** 775-789.

Leach, K., Zalat, S. & Gilbert, F. 2013. Egypt's Protected Area network under future climate change. *Biological Conservation,* 159**,** 490-500.

Lee, T. M., Sodhi, N. S. & Prawiradilaga, D. M. 2007. The Importance Of Protected Areas For The Forest And Endemic Avifauna Of Sulawesi (Indonesia). *Ecological Applications,,* 17**,** 1727-1741.

Linkie, M., Chapron, G., Martyr, D. J., Holden, J. *et al.* 2006. Assessing the viability of tiger subpopulations in a fragmented landscape. *Journal of Applied Ecology,* 43**,** 576-586.

Liu, C., Berry, P. M., Dawson, T. P. & Pearson, R. G. 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography,* 28**,** 385-393.

Lobo, J. M., Jiménez-Valverde, A. & Real, R. 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography,* 17**,** 145-151.

Newbold, T., Gilbert, F., Zalat, S., El-Gabbas, A. *et al.* 2009. Climate-based models of spatial patterns of species richness in Egypt's butterfly and mammal fauna. *Journal of Biogeography,* 36**,** 2085-2095.

Newbold, T., Reader, T., El-Gabbas, A., Berg, W. *et al.* 2010. Testing the accuracy of species distribution models using species records from a new field survey. *Oikos,* 119**,** 1326-1334.

Okigbo, R. N., Eme, U. E. & Ogbogu, S. 2008. Biodiversity and conservation of medicinal and aromatic plants in Africa. *Biotechnology and Molecular Biology Reviews,* 3 127-134.

Oliver, T., Roy, D. B., Hill, J. K., Brereton, T. *et al.* 2010. Heterogeneous landscapes promote population stability. *Ecol Lett,* 13**,** 473-84.

Ortega-Huerta, M. A. & Peterson, A. T. 2004. Modelling spatial patterns of biodiversity for conservation prioritization in North-eastern Mexico. *Diversity and Distributions,* 10**,** 39–54.

Pawar, S., Koo, M. S., Kelley, C., Ahmed, M. F. *et al.* 2007. Conservation assessment and prioritization of areas in Northeast India: Priorities for amphibians and reptiles. *Biological Conservation,* 136**,** 346-361.

Phillips, S. J., Anderson, R. P. & Schapire, R. E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling,* 190**,** 231-259.

Pineda, E. & Lobo, J. M. 2009. Assessing the accuracy of species distribution models to predict amphibian species richness patterns. *Journal of Animal Ecology,* 78**,** 182–190.

Pressey, R. L., Humphries, C. J., Margules, C. R., Vane-Wright, R. I. *et al.* 1993. Beyond Opportunism: Key Principles for Systematic Reserve Selection. *Trends In Ecology & Evolution,* 8**,** 124-128.

Sciberras, M., Jenkins, S. R., Kaiser, M. J., Hawkins, S. J. *et al.* 2013. Evaluating the biological effectiveness of fully and partially protected marine areas. *Environmental Evidence,* 2**,** 1-31.

Seiferling, I. S., Proulx, R., Peres-Neto, P. R., Fahrig, L. *et al.* 2012. Measuring protected-area isolation and correlations of isolation with land-use intensity and protection status. *Conserv Biol,* 26**,** 610-8.

Spasojevic, M. J., Grace, J. B., Harrison, S., Damschen, E. I. *et al.* 2014. Functional diversity supports the physiological tolerance hypothesis for plant species richness along climatic gradients. *Journal of Ecology,* 102**,** 447-455.

Thomas, C. d. & Gillingham, p. k. 2015. The performance of rotected reas for biodiversity under climate change. *Biological Journal of the Linnean Society,* 115**,** 718–730.

Traba, J., García de la Morena, E. L., Morales, M. B. & Suárez, F. 2007. Determining high value areas for steppe birds in Spain: hot spots, complementarity and the efficiency of protected areas. *Biodiversity and Conservation,* 16**,** 3255-3275.

Warren, D. L. & Seifert, S. N. 2011. Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications,* 21**,** 335–342.

Wisz, M. S., Pottier, J., Kissling, W. D., Pellissier, L. *et al.* 2013. The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biol Rev Camb Philos Soc,* 88**,** 15-30.

Table 1: Environmental variables used to build the models (The highlighted one thrown), after applying Variance Inflation Factor (VIF) to reduce the collinearity.

BIO1	Annual Mean Temperature
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO3	Isothermality (BIO2/BIO7) (* 100)
BIO4	Temperature Seasonality (standard deviation *100)
BIO5	Max Temperature of Warmest Month
BIO6	Min Temperature of Coldest Month
BIO7	Temperature Annual Range (BIO5-BIO6)
BIO8	Mean Temperature of Wettest Quarter
BIO9	Mean Temperature of Driest Quarter
BIO10	Mean Temperature of Warmest Quarter
BIO11	Mean Temperature of Coldest Quarter
BIO12	Annual Precipitation
BIO13	Precipitation of Wettest Month
BIO14	Precipitation of Driest Month
BIO15	Precipitation Seasonality (Coefficient of Variation)
BIO16	Precipitation of Wettest Quarter
BIO17	Precipitation of Driest Quarter
BIO18	Precipitation of Warmest Quarter

BIO19	Precipitation of Coldest Quarter
Altitude	Altitude
Habitat	Habitat
NDVI_Max	NDVI maximum value
NDVI_Difference	Absolute difference between the highest and lowest NDVI values

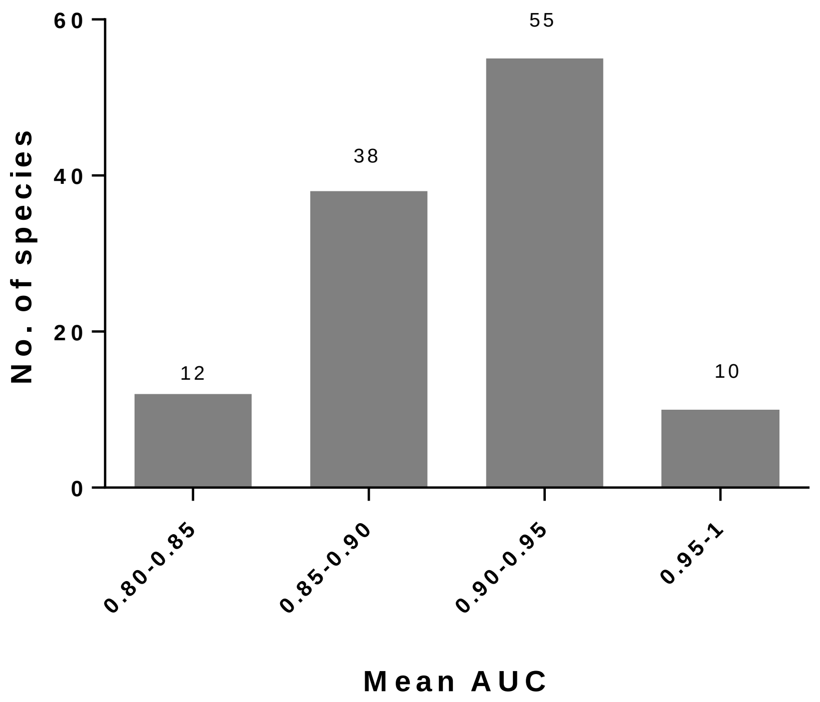


Figure 1: Frequency distribution of the mean AUC values achieved in the distribution models of plant species.

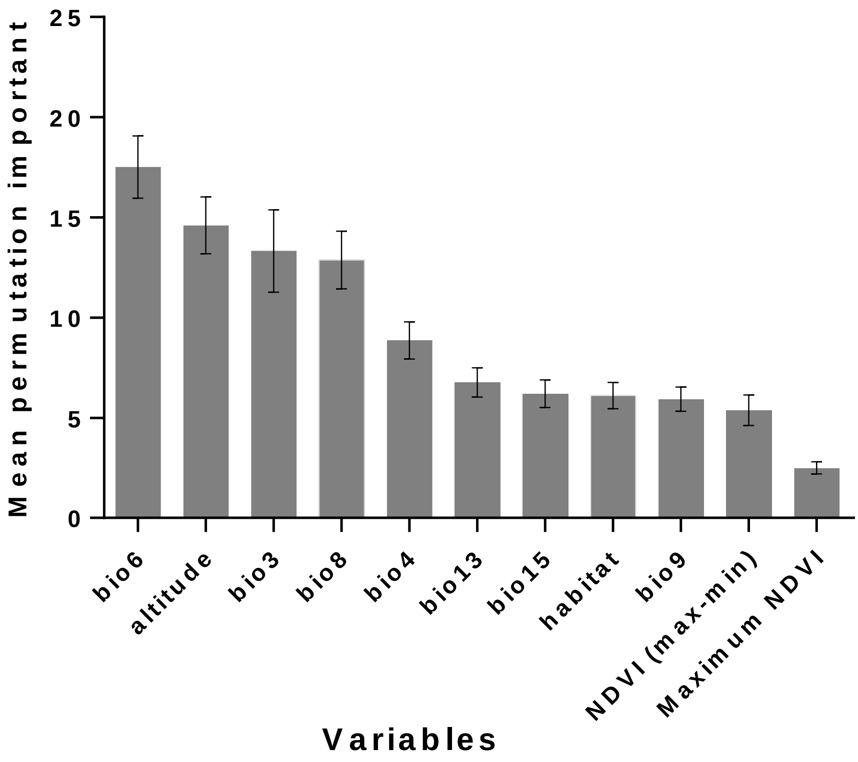


Figure 2: Contribution to the final species distribution models made by each environmental predictor, illustrated by the mean permutation importance.

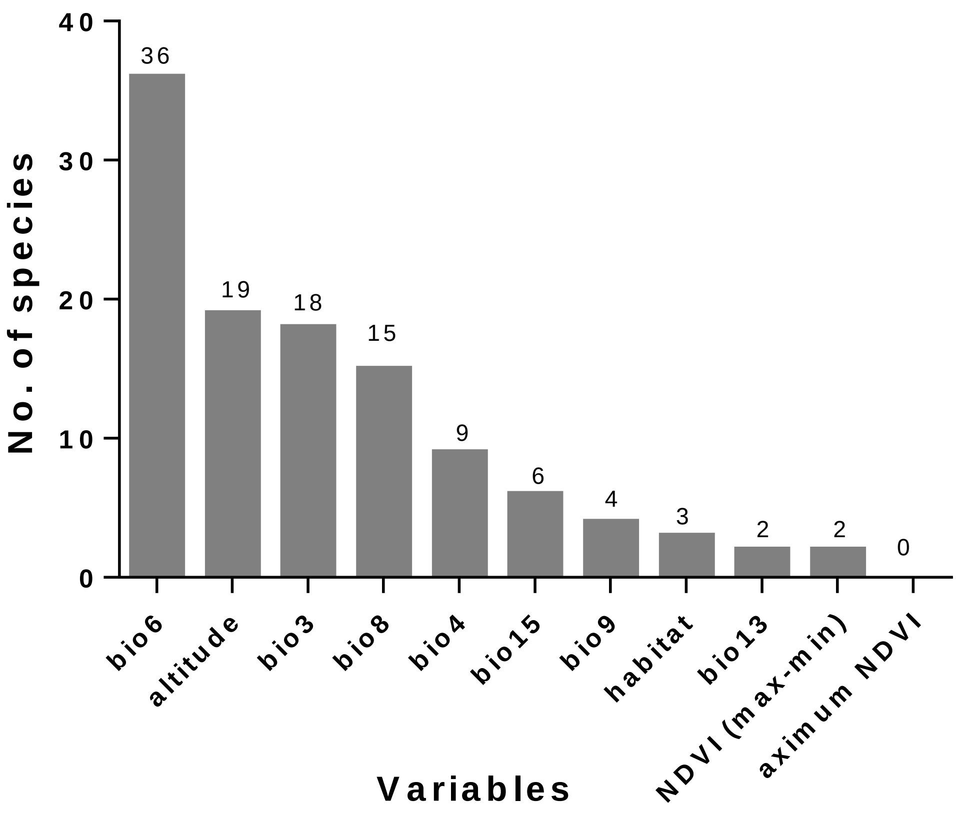


Figure 3: Numbers of species where particular variables were the best predictor in species distribution models. One of the variables (ndvi_max) included in modelling was never the best predictor.

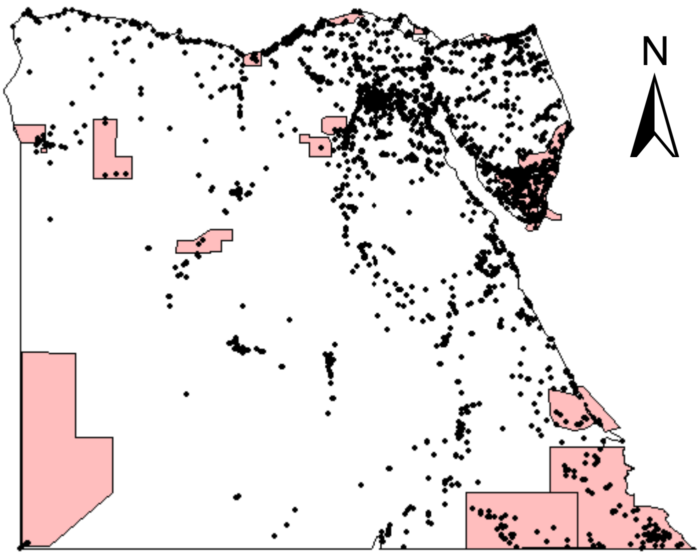


Figure 4: Locations which Egyptian plants were sampled (square circle), and protected areas (PAs) of Egypt (pink shading).

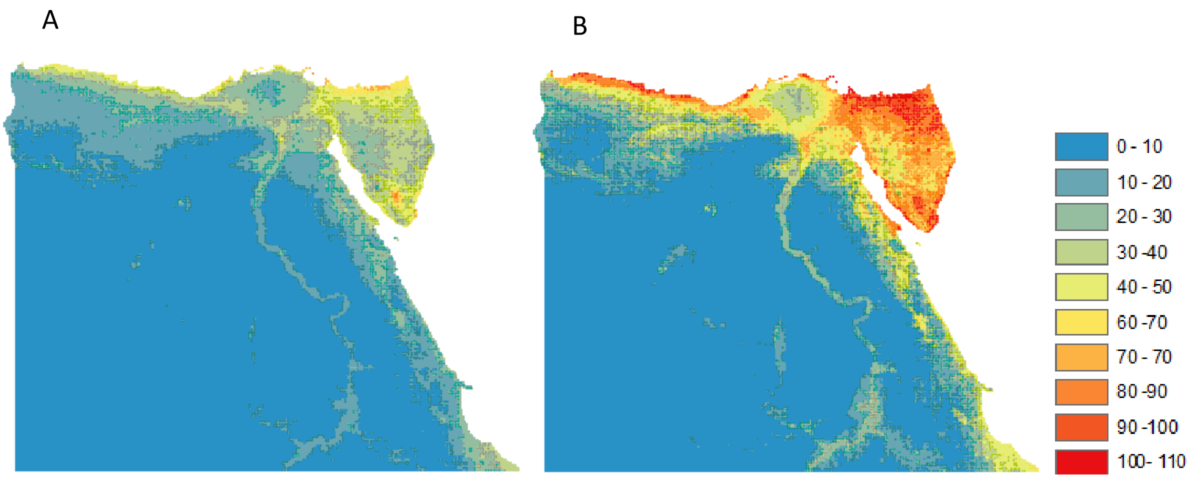


Figure 5: Species richness for predicting current distributions: (A) probability richness map resulting from summing all individual species probability maps then rescaled to the same range as that of the binary map; (B) binary richness map, produced from adding all individual species thresholded maps. The colours ranged from blue to red, which blue indicate for low species richness and red indicated for high species richness.

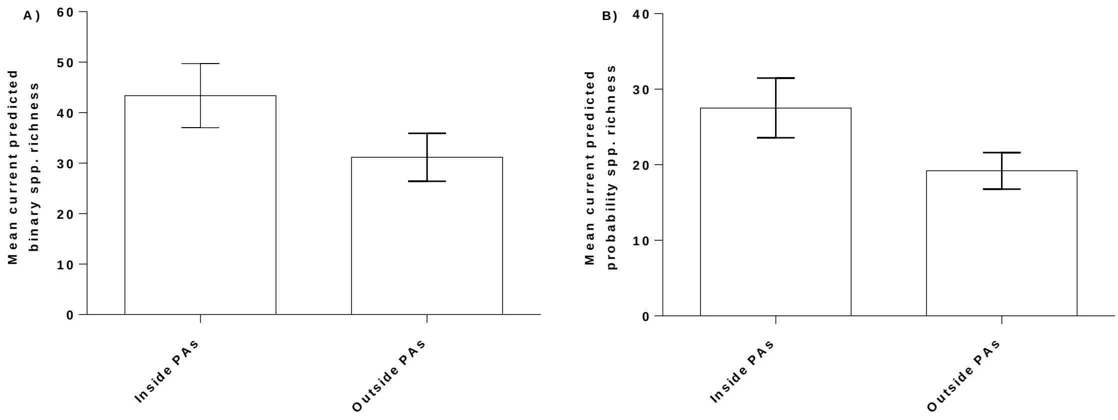


Figure 6: A) Comparison of binary predicted species richness (mean ±SE) between inside and outside protected areas (PAs), expected by adding all individual species distribution models; B) Comparison of probability predicted species richness (mean ±SE) between inside and outside protected areas.

Supplementary

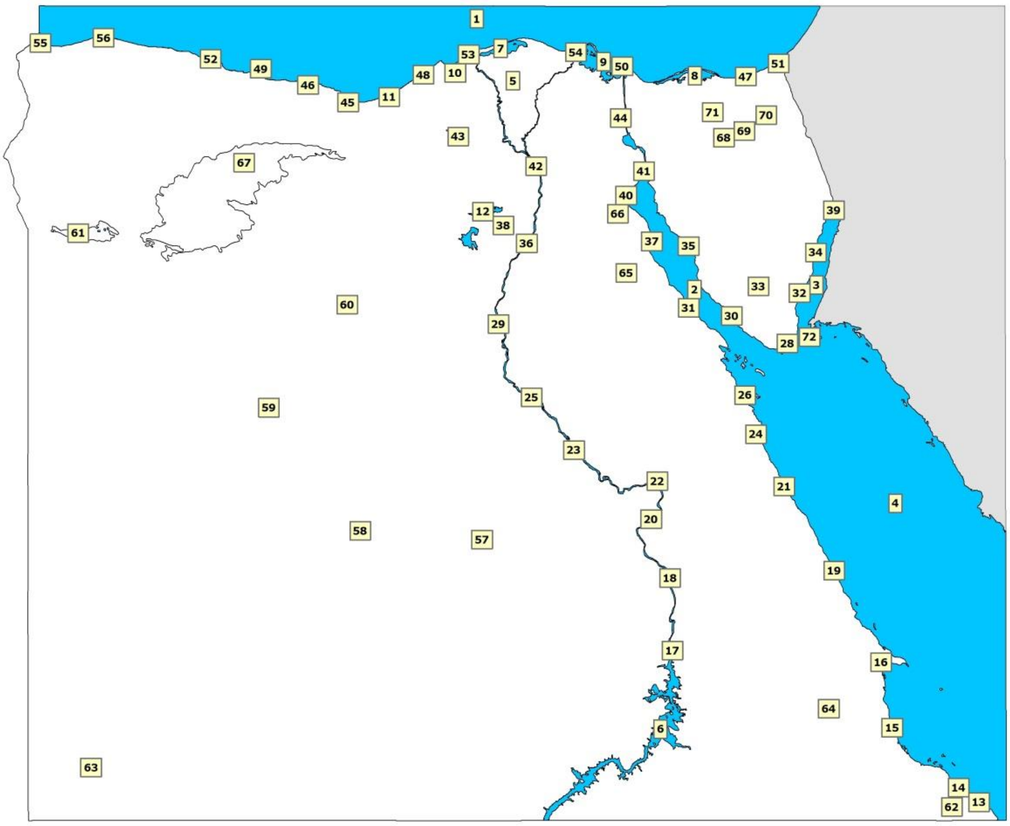
Table S1: The plant species used for species distribution models, showing the model fit in terms of the mean and standard deviation of the AUC values of the 10 replicates.

Species	Mean AUC	AUC SD	Number of records
Acacia pachyceras	0.982	0.018	11
Acacia tortilis	0.89	0.031	242
Achillea fragrantissima	0.926	0.035	239
Achillea santolina	0.98	0.029	13
Adonis dentata	0.972	0.035	33
Aerva javanica	0.911	0.026	175
Agathophora alopecuroides	0.891	0.074	26
Alhagi graecorum	0.806	0.067	100
Anabasis articulata	0.924	0.032	126
Anagallis arvensis	0.889	0.084	119
Anastatica hierochuntica	0.841	0.13	31
Andrachne aspera	0.941	0.049	31
Artemisia judaica	0.93	0.031	684
Artemisia monosperma	0.946	0.025	115
Asclepias sinaica	0.973	0.044	31
Asparagus stipularis	0.987	0.016	20
Atriplex halimus	0.898	0.055	125
Avena barbata	0.914	0.112	22
Ballota undulata	0.929	0.061	124
Bassia muricata	0.853	0.074	83
Calendula arvensis	0.895	0.052	66
Calotropis procera	0.858	0.065	228
Capparis spinosa	0.859	0.065	241
Chenopodium album	0.926	0.027	166
Chenopodium murale	0.88	0.027	319
Chiliadenus montanus	0.91	0.081	88
Citrullus colocynthis	0.879	0.047	168
Cleome amblyocarpa	0.916	0.044	71
Colutea istria	0.91	0.107	24
Cornulaca monacantha	0.914	0.053	93
Cymbopogon schoenanthus	0.835	0.291	15
Cynodon dactylon	0.88	0.045	240
Deverra tortuosa	0.94	0.021	141

Deverra triradiata	0.832	0.096	64
Diplotaxis acris	0.916	0.05	60
Diplotaxis erucoides	0.898	0.061	12
Diplotaxis harra	0.911	0.039	147
Echinops spinosus	0.915	0.024	213
Ephedra alata	0.85	0.087	46
Eruca sativa	0.938	0.038	96
Euphorbia peplis	0.96	0.049	13
Euphorbia retusa	0.885	0.059	114
Fagonia arabica	0.91	0.025	294
Fagonia glutinosa	0.902	0.077	125
Fagonia mollis	0.924	0.062	698
Farsetia aegyptia	0.906	0.043	114
Globularia arabica	0.923	0.063	51
Gypsophila capillaris	0.885	0.067	59
Halocnemum strobilaceum	0.897	0.134	53
Haloxylon salicornicum	0.943	0.026	242
Haloxylon scoparium	0.882	0.172	36
Haplophyllum tuberculatum	0.889	0.046	73
Heliotropium arbainese	0.877	0.05	146
Herniaria hirsuta	0.846	0.287	13
Hyoscyamus muticus	0.877	0.043	143
Imperata cylindrica	0.835	0.054	109
Iphiona mucronata	0.904	0.089	53
Juncus rigidus	0.855	0.042	235
Lavandula pubescens	0.986	0.011	15
Lycium shawii	0.905	0.04	161
Malva parviflora	0.866	0.053	165
Melilotus indicus	0.906	0.04	239
Mesembryanthemum crystallinum	0.934	0.091	38
Mesembryanthemum forsskaolii	0.833	0.246	13
Mesembryanthemum nodiflorum	0.912	0.059	79
Moltkiopsis ciliata	0.931	0.03	181
Nitraria retusa	0.92	0.045	100
Noaea mucronata	0.86	0.149	47
Ochradenus baccatus	0.898	0.045	137
Orobanche cernua	0.909	0.086	42

Pancratium sickenbergeri	0.93	0.144	42
Panicum turgidum	0.906	0.048	164
Paronychia arabica	0.888	0.058	169
Paronychia argentea	0.867	0.215	17
Peganum harmala	0.913	0.055	138
Pergularia tomentosa	0.85	0.053	362
Phoenix dactylifera	0.826	0.078	64
Phragmites australis	0.858	0.061	128
Plantago afra	0.908	0.081	43
Plantago ovata	0.904	0.043	36
Pluchea dioscoridis	0.802	0.145	62
Polycarpaea repens	0.895	0.04	172
Polycarpon succulentum	0.908	0.099	94
Pulicaria undulata	0.89	0.035	271
Reaumuria hirtella	0.948	0.023	109
Reichardia tingitana	0.902	0.044	120
Retama raetam	0.915	0.037	261
Salvia aegyptiaca	0.911	0.058	69
Salvia lanigera	0.877	0.107	44
Senecio glaucus	0.914	0.05	159
Seriphidium herba-album	0.94	0.045	267
Silene succulenta	0.857	0.202	25
Silene villosa	0.908	0.14	51
Sisymbrium irio	0.88	0.05	56
Solanum elaeagnifolium	0.989	0.014	24
Solanum nigrum	0.883	0.041	199
Stachys aegyptiaca	0.95	0.022	168
Stipagrostis scoparia	0.93	0.061	49
Tamarix aphylla	0.823	0.122	86
Tamarix nilotica	0.853	0.025	332
Tephrosia purpurea	0.881	0.068	78
Teucrium leucocladum	0.943	0.071	66
Teucrium polium	0.868	0.13	133
Thymelaea hirsuta	0.92	0.092	61
Tribulus terrestris	0.869	0.067	65
Trifolium resupinatum	0.93	0.043	126
Trigonella stellata	0.922	0.031	62

Urginea maritima	0.918	0.185	24
Urtica urens	0.919	0.092	34
Vicia sativa	0.851	0.121	68
Zilla spinosa	0.919	0.036	632
Zygophyllum album	0.92	0.034	115
Zygophyllum coccineum	0.87	0.039	488
Zygophyllum dumosum	0.971	0.019	27



1	The Mediterranean Sea	**25**	Assiut	**49**	Ras El-Hekma
2	The Suez Gulf	**26**	Hurghada	**50**	Port-Said
3	The Aqaba Gulf	**27**	Ras Mohamed	**51**	Rafah
4	The Red Sea	**28**	Sharm El-Sheikh	**52**	Mersa Matruh
5	The Nile Delta	**29**	El-Minia	**53**	Rosetta
6	Lake Nasser	**30**	El-Tur	**54**	Damietta
7	Lake Brullus	**31**	Ras Gharib	**55**	Sallum
8	Lake Bardawil	**32**	Dahab	**56**	Sidi Barrani
9	Lake Manzala	**33**	Saint-Katherine	**57**	Kharga oasis
10	Lake Idku	**34**	Nuweiba	**58**	Dakhla oasis
11	Lake Mariut	**35**	Abu Zneima	**59**	Farafra oasis
12	Lake Qarun	**36**	Beni Suef	**60**	Bahariya oasis
13	Halayeb	**37**	Ras Zaafarana	**61**	Siwa oasis
14	Abu Ramad	**38**	Fayoum	**62**	Gebel Elba area
15	Al-Shalatein	**39**	Taba	**63**	El-Gilf El-Kebir
16	Berenice	**40**	Ain Sukhna	**64**	Gebel Abraq area
17	Aswan	**41**	Suez	**65**	Gebel El-Gallala El-Qibliya
18	Edfu	**42**	The greater Cairo	**66**	Gebel El-Gallala El-Bahariya

19	Mersa Alam	**43**	Wadi El-Natrun	**67**	Qattara Depression
20	Luxor	**44**	Ismailia	**68**	Gebel Yillaq
21	El-Quseir	**45**	El-Alamein	**69**	El-Hassana
22	Qena	**46**	El-Dabaa	**70**	Gebel El-Hallal
23	Sohag	**47**	El-Arish	**71**	Gebel El-Maghara
24	Safaga	**48**	Alexandria	**72**	Tiran & Sanafir islands

Figure S1 : Egypt's political border and all cities and geographical regions mentioned in this study (El-Gabbas *et al.*, 2016).