Modeling Hotspots of Plant Diversity in New Guinea

JAMIE D. HOOVER[1*], SUNIL KUMAR[2], SHELLEY A. JAMES[3], STEPHEN J. LEISZ [4], MELINDA LAITURI[5]

[1]*Graduate Degree Program in Ecology, Colorado State University, Fort Collins, CO, USA*
[2]*Natural Resource Ecology Laboratory, Colorado State University, Fort Collins, CO, USA*
[3]*iDigBio, Florida Museum of Natural History, Gainesville FL, USA*
[4]*Department of Anthropology, Colorado State University, Fort Collins, CO, USA*
[5]*Department of Ecosystem Science and Sustainability, Colorado State University,
Fort Collins, CO, USA*

Abstract: New Guinea is widely known for rich biodiversity. This study provides a foundation for understanding vascular and non-vascular plant distributions at the genus taxonomic level. Analyses objectively and quantitatively showed collection density and biases at 50 km spatial resolution and predict genus richness at 1 km spatial resolution. To model the distribution of 1354 genera presence-only data and the MaxEnt niche model were used. Results showed many cells with fewer than 5 unique genera whereas other cells had upwards of 600 unique genera. Model performance achieved a mean test AUC score of 0.7. The most influential environmental variables were elevation, slope, and temperature annual range. Predicted genus richness increased from south to north across the island and west to east along the central mountain range. Models predicted higher richness in topographically complex regions at higher elevations. Comparatively, lower elevations with homogenous topography had larger tracts of high genus richness. Modeling genus level data supplied baseline information about plant distributions, yet some genera are more speciose than others, so the full scope of richness or endemism may not be captured. Results can be used to prioritize sampling needs, support conservation strategies, compare genus diversity to other regions, and discuss biogeography principles.

Key words: Ecological niche modeling, land Change, MaxEnt, Papua New Guinea, plant richness, Western Papua and Papua (Indonesia).

Introduction

New Guinea is estimated to harbor 5–10% of the world's biodiversity in only 0.5% of earth's land area (Mittermeier *et al.* 2003; Supriatna *et al.* 1999). For plant biodiversity, New Guinea ranks second to Amazonia and this equates roughly to 17,000 unique species, 10,200 of which are thought to be endemic (Mittermeier *et al.* 2003). While it is difficult to deny the diversity of the biota in New Guinea, the evidence to support the high rates of diversity and endemism are not based on comprehensive taxonomically vouchered collections. This is especially the case for embryophyta or vascular and non-vascular terrestrial plants, the focus of this study (hereafter referred to as terrestrial plants). Estimates of diversity have been based on expert opinion (Vollering *et al.* 2015), and endemism rates for terrestrial plants have been estimated using the richness of taxonomic groups, such as orchids and ferns (Supriatna *et al.* 1999). However, in Ecuador, Mandl *et al.* (2010) showed that epiphytic plant diversity differs from other terrestrial plant diversity due to differing environmental requirements. Other more systematic approaches have used topographic and climatic data to identify unique biogeographical environments where high diversity is likely to occur (Heads 2006; Vollering *et al.* 2015). Phylogenetic molecular techniques for identifying dispersal and speciation for the tropical South Pacific have been conducted but are limited to a handful of species and higher level taxa and likewise, New Guinea is poorly represented (Keppel *et al.* 2009).

Corresponding Author; e-mail: jamie.hoover11@gmail.com

Table 1. Environmental variables used in the model.

Predictor	Description	Source
ALT+	Altitude from digital elevation model	
BIO4+	Temperature seasonality (standard deviation *100)	
BIO7+	Temperature annual range (Max T. of warmest month - Min T. of coldest month)	BioClim http://www.worldclim.org/
BIO12+	Annual precipitation	
BIO15	Precipitation seasonality (Coefficient of variation)	
BIO18	Precipitation of warmest quarter	
pH	pH of water in soil at 10 cm depth	
BD	Bulk Density: ratio of soil mass to soil volume at 10 cm depth	
CEC	Cation exchange capacity at 10 cm depth	
Clay+	Fraction of clay by weight at 10 cm depth	ISRIC World Soil http://www.isric.org/
CF	Coarse fragments >2mm in volumetric percent at 10 cm depth	
OC	Organic carbon at 10 cm depth	
Silt	Fraction of silt by weight at 10 cm depth	
Exposure+	Topographic exposure	A 3x3 cell mean was calculated on the ALT layer; the difference between the ALT and 3x3 mean layers is calculated.
Slope+	Slope in degrees	Calculated using the ALT layer
Eastness+	Sine of aspect	Aspect calculated using the ALT layer; sine or cosine is calculated
Northness+	Cosine of aspect	
Correlation	Linear dependency of EVI on adjacent pixels	
evenness	Evenness of EVI	Global habitat heterogeneity http://www.earthenv.org/texture.html
Uniform	Orderliness of EVI	
Variance+	Dispersion of EVI combinations between adjacent pixels	

+ indicates the variables used when occurrence points are between 10 and 25.

There are two widely accepted explanations for the high biological diversity in New Guinea. First, island biogeography theory states that islands with the richest species diversity are larger in area, have higher elevations, and are more proximate to source areas (Brooks *et al.* 2006; Cronk *et al.* 2005; Heads 2001; Keppel *et al.* 2009; Neall & Trewick 2008; Roos *et al.* 2004; Vollering *et al.* 2015). The island of New Guinea fits all of these characteristics as it is the largest in the Pacific, has the highest mean elevation (highest point at 4884 m in Southeast Asia and Oceania), and is proximal to many source areas, such as Southeast Asia, Australia, and multiple island archipelagos across Malesia, Micronesia, and Polynesia. Second, the other theory that supports the high rates of diversity on the island of New Guinea is the tectonic history. The northern portion of the island formed by the accretion of 32 distinct terrains, each with unique origins, history, and biota (Heads 2001, 2006; Hill & Hall 2003). The southern portion of New Guinea is the northern reach of the Australian Craton (Baldwin *et al.* 2012; Hill & Hall 2003). Therefore, the processes of dispersal and vicariance are believed to largely influence patterns of plant distributions (Cronk *et al.* 2005; Heads 2009; Keppel *et al.* 2009).

New Guinea is understudied compared to other tropical areas (Heads 2001, 2006; Keppel *et al.* 2009; Vollering *et al.* 2015) and even the more systematic approaches and sound theories lack adequate taxonomic catalogues to verify or comprehensively assess the distribution of biota and richness therein (Roos *et al.* 2004). Tropical forests rarely have complete catalogues of biota because these ecosystems have high species
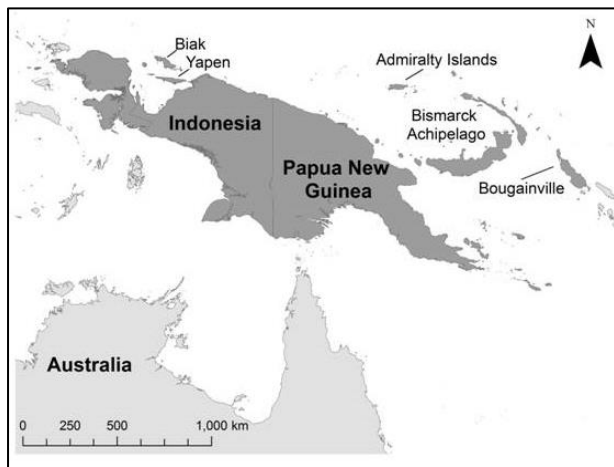
Fig. 1. The island of New Guinea is occupied by two countries, Indonesia to the west and Papua New Guinea to the east. Included in this study are the Indonesian islands Biak and Yapen, and Papua New Guinea's Admiralty Islands and Bismarck Archipelago. Also part of the study is the autonomous island of Bougainville, which is part of the Solomon Archipelago. The projection is Albers Equal Area projection, WGS84.

richness and surveying efforts are laborious, expensive, and spatially biased. In New Guinea, survey efforts are spatially biased in multiple ways. First, there are more specimens collected in areas that are easier to access (near towns, rivers, and roads). This is exacerbated in New Guinea as travel on the island is greatly limited due to the lack of infrastructure. Second, collection densities in New Guinea increase from west to east and from south to north (Takeuchi 2007). Even though Indonesia is known for high rates of biodiversity, the full scope of diversity is unknown because the Indonesian territories in New Guinea are severely under sampled. Although collections are low, overall the number of specimens collected in Papua New Guinea (PNG) is over 30 times greater than Indonesia's collections in New Guinea. This can be easily visualized online at biodiversity data websites such as Global Biodiversity Information Facility (GBIF; http://www.gbif.org/) and iDigBio (https://www.idigbio.org/). Third, survey efforts increase with elevation, and this is especially the case in the highland areas of PNG (Takeuchi 2007). Fourth, it is a challenge to gain land access to study biodiversity or collect specimens because land is under customary land tenure and foreigners are viewed as untrustworthy. Even though researchers approach land managers with transparent intentions, government agencies and resource extraction companies have had a long history of corruption and illegal operations and this history has caused distrust of all types of surveying (PNG specific, A. Allison, personal comm.). Last, the biological surveys that have occurred in recent decades are rapid biological assessments (RAP surveys), which are conducted over a short period of time, cover small areas, are often in response to pending resource extraction or development (e.g., dams and mining; Katovai *et al.* 2015), and are often published in grey literature (Leisz *et al.* 2000; Mack & Alonso 2000; McGavin 2009; Richards & Gamui 2011).

The ability to identify distribution patterns is interesting theoretically to the scientific community but also can be used for land-use planning and management and conservation strategies (Anderson 2013; de Barros Ferraz *et al.* 2012; Heads 2001). Information on the spatial patterns of terrestrial plant species richness in New Guinea is not available, and it is urgently needed to address threats to biodiversity due to habitat losses via resource extraction and development (logging, mining, fiber, and oil palm), which have cleared or degraded approximately 30% of forests across New Guinea and the surrounding islands (Abood *et al.* 2015; Bryan & Shearman 2015; Shearman *et al.* 2009). Higher rates of forest losses are observed in areas that are more easily accessed such as coastal lowlands and islands. In PNG over 43% of coastal lowland forests have been cleared at least once between 1972 and 2014 (Shearman *et al.* 2009, 2015). Another major aspect of forest loss across New Guinea is logging regulations, which are often violated. For example, repeat harvests are scheduled in too short of a time scale (e.g., 15 years instead of 35 years), illegal logging occurs, and industries expand outside set boundaries (Bryan & Shearman 2015). Across Indonesia, around 55% of resource extractions occurred outside of set boundaries (Abood *et al.* 2015).

Although the biological knowledge of the island is far from complete, recent interest in understanding the spatial distribution of biota has been ignited with efforts to amalgamate and digitize specimen data from herbaria, museums, and private collections into online databases. These databases along with ecological niche models (ENMs; also called species distribution models (SDMs)) have become a valuable tool in biogeographic research. ENMs are based on the fundamental and realized niche concepts and approximate a species' distribution using
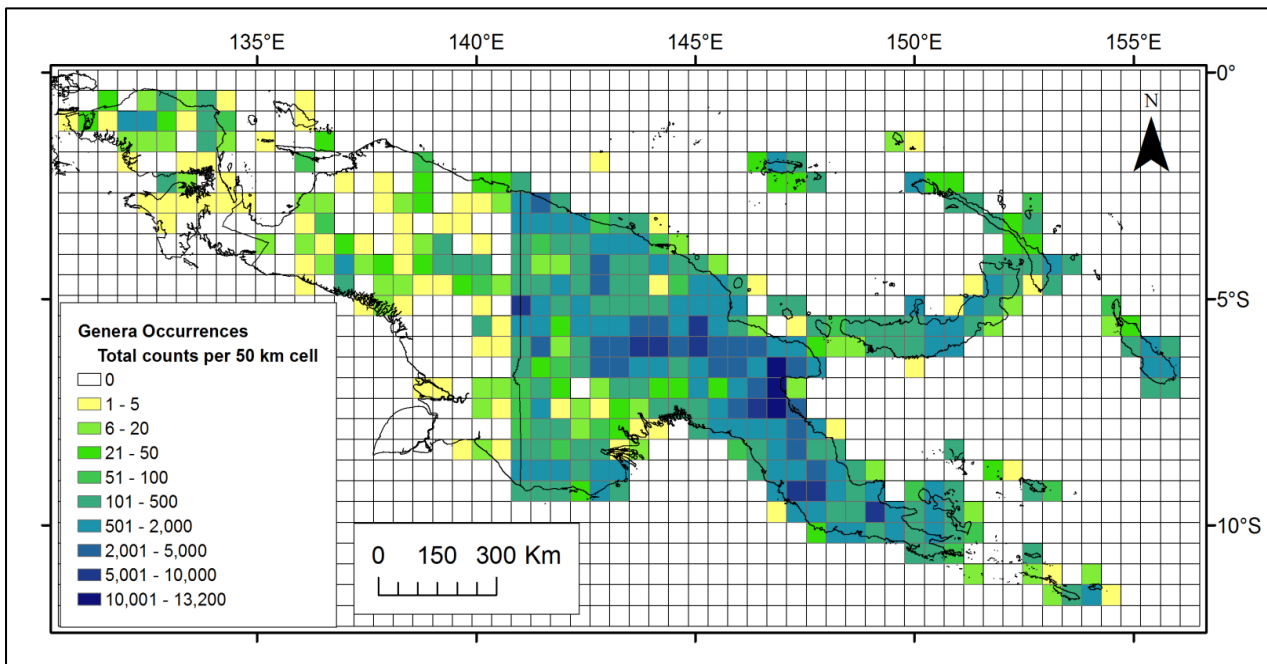
Fig. 2. The total number of specimens collected per 50 km grid cell shows collection efforts. A single cell can have one or more of the same genus. The projection is Albers Equal Area projection, WGS84.

occurrence data and environmental conditions (e.g., climate, topographic; Peterson *et al.* 2011).

To date there have been few attempts that systematically and objectively assessed terrestrial plant distribution (Heads 2001; Roos *et al.* 2004; Vollering *et al.* 2015) and none to date have used all available occurrence data. In this study our aim was to map the distribution of terrestrial plants at the genus taxonomic level using a maximum entropy model or MaxEnt (Phillips *et al.* 2006). The specific goals of this study were to: 1) identify sampling intensity and sampling bias; 2) identify the most influential abiotic drivers associated with terrestrial plant distributions; 3) identify the regions of New Guinea that are likely to harbor high terrestrial plant richness; and 4) discuss the implications of threatened habitat and biodiversity losses due to resource development and land use changes.

Study site

This study was conducted on the island of New Guinea, which is politically divided into the Republic of Indonesia to the west and the Independent Nation of Papua New Guinea (PNG) to the east. Many of the surrounding islands were also included in this study and some of the major island groups are the Bismarck Archipelago and Admiralty Islands of PNG, Biak and Yapen of Indonesia, and the autonomous island nation

Bougainville, which is part of the Solomon Archipelago (Fig. 1).

The total landmass of New Guinea is 786,000 km², excluding the surrounding island archipelagos. New Guinea is the second largest island in the world and tallest landmass in the south Pacific, which is geographically inclusive of Southeast Asia, Australia, and New Zealand. Elevation ranges from sea level to 4884 m and is typically divided into coastal lowlands (0-1000 m), lower montane (1000–2800 m), and upper montane (2800–4900) (Bryan & Shearman 2008). Temperatures are on average 28 °C at sea level, inland and mountain areas average 26 °C and higher elevations are 23 °C. The temperature variability ranges between 6.8 and 14.6 °C, with the greatest degree changes in the lower and upper montane zones.

Precipitation varies greatly across the island from 970 mm to 7500 mm per year. Glaciers still exist on peaks in the upper montane zone, but the snowlines have been rapidly retreating in the past century (Hope 2014).

New Guinea is composed of three distinct geologic formations, the Stable Platform, Fold Belt, and Mobile Belt. The Stable Platform is a continuation of the Australian Craton and the Fold Belt is the northern edge of this Craton (Heads 2006; Hill & Hall 2003). The Fold Belt or central mountain range spans east-west across New
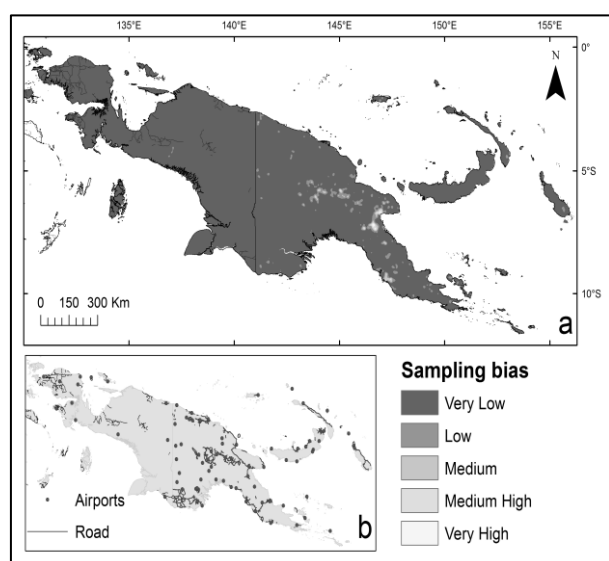
Fig. 3. (a) Sampling bias created using kernel density estimate at a 10 km resolution with some locations identified. (b) Roads and airports show the influence of infrastructure on sampling bias. The projection is Albers Equal Area projection, WGS84.

Guinea and was the result of fold and thrust deformations from arc-continent collisions (Baldwin *et al.* 2012; Hill & Hall 2003; Polhemus & Polhemus 1998). The Mobile belt was created over the past 40 Myr from a series of 32 island arcs, some composite, that accreted to the Fold Belt (Heads 2001, 2006; Hill & Hall 2003). The Bismarck Archipelago is in route to collide with New Guinea in the next 10 Myr (Polhemus & Polhemus 1998).

Methods

Occurrence data

All georeferenced specimen occurrence records were combined from the PNG Plant database (Conn *et al.* 2004), Herbarium Pacificum, Bernice P. Bishop Museum (Bishop Museum 2015), and Global Biodiversity Information Facility (GBIF) data portal (GBIF 2015). Generic taxonomy was updated based on Angiosperm Phylogeny Group (APG) IV classification (Chase *et al.* 2016). To maximize the number of unique occurrences and ensure data quality the genus taxonomic level was used, as the species level data had many inconsistencies and too few occurrences per species.

The original dataset contained around 3,000 unique genera with over 100,000 specimens. All occurrences with incomplete location information, missing or incorrect taxon names were removed

from the dataset. Any occurrence records of cultivated or introduced taxa in New Guinea were removed from the dataset. Of the remaining occurrences, 36% of the genera were not used in the study because there were fewer than 10 specimens.

Duplicates were removed using occurrence identification numbers and location information. If multiple records of the same genus were found in the same 1 km^2 grid cell, only a single record was included. To account for spatially auto-correlated occurrence points and avoid model over fitting, all points were spatially filtered at 5 km. Spatial filtering also ensured that the test and training data were independent when cross-validation evaluation techniques were used (Boria *et al.* 2014; de Oliveira *et al.* 2014; Radosavljevic & Anderson 2014; Sidder *et al.* 2016; Veloz 2009). After spatial filtering, genera with fewer than 10 occurrences constituted 7% of the dataset and were not included because there were too few occurrences for a general model (Austin 2002; Bell & Schlaepfer 2016). These genera are listed in Supplementary Material 3. The final dataset contained 1,354 genera with 85,481 occurrence points. There were around 5,000 occurrence points in Indonesia and 80,000 points in PNG. Supplementary Material 1, provides a table of the genera used in this study and the number of occurrences. The genera that lacked adequate occurrences are also in the Supplementary Material 2, so that future surveys can focus on data deficient genera.

Sampling intensity and biases

To identify the spatial distribution of collection efforts across New Guinea and surrounding archipelagos a 50 km grid was created. The occurrence data were counted per grid cell in two ways. First, all occurrences were counted to show overall sampling efforts per 50 km grid cell. The second method counted the number of unique genera or genus richness per grid cell. The spatial biases for sampling efforts was created using Gaussian kernel density estimate tool from the SDM Toolbox (Brown 2014).

Environmental data

Environmental data from three different sources were used and these included 19 bioclimatic and elevation variables from the World Clim dataset (Hijmans *et al.* 2005), global habitat heterogeneity (GHH; Tuanmu & Jetz 2015), and soil data from the ISRIC (ISRIC 2015; Table 1).
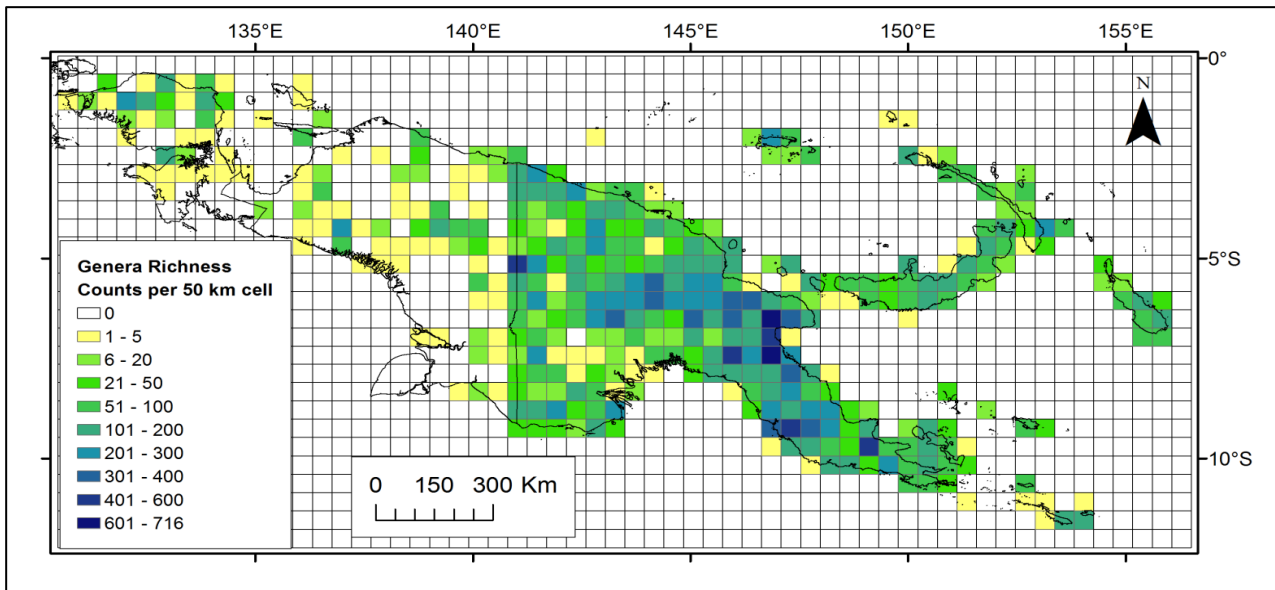
Fig. 4. Genus richness shows the number of unique genera collected per 50 km grid cell. The projection is Albers Equal Area projection, WGS84.

Multiple variables were generated from the altitude data including slope (in degrees), aspect, and topographic exposure. The GHH data were all based on texture features of the enhanced vegetation index (EVI) and aimed to quantify spatial heterogeneity (Tuanmu & Jetz 2015). The northness and eastness variables were derived from the cosine and sine transformation of the aspect, respectively. Topographic exposure was calculated using the difference between the altitude layer and a transformed altitude raster where a 3 x 3 neighborhood mean was applied. All environmental data were continuous variables and had a spatial resolution of 1 km².

For this study, the Pearson correlation coefficient (r) was used to account for multicollinearity among environmental variables (Dormann et al. 2013). If two variables were highly collinear ($|r| > 0.75$) one was removed and the variable retained was the one that was perceived to be more ecologically influential to terrestrial plants. The number of environmental predictor variables used in the modeling was reduced to 21 (Table 1). All 21 variables were considered when the occurrence counts were greater than 25. Ten environmental variables were used for the genera with occurrences fewer than 25 so to not over or under predict the distribution. The 10 environmental variables selected were the ones that were directly measured (e.g., altitude and temperature) and least correlated (Table 1). All

environmental and occurrence data were projected to an equal area projection (Cylindrical Equal Area Conic, Datum WGS84).

Model calibration and validation

The maximum entropy model or MaxEnt (version 3.3.3; Phillips et al. 2006) was used to map the distribution of terrestrial plants in New Guinea. Of the current models available, MaxEnt was the top choice for this study for multiple reasons. First, MaxEnt uses presences-only data. Second, it generally outperforms other niche models (Evangelista et al. 2008). Third, it has performed well with small sample sizes (Wisz et al. 2008) and was found to be suitable for our dataset as some of the genera have a minimum of 10 occurrence records. Last, MaxEnt can be used to run models for thousands of species at a time.

In general, default settings were used, and when this was not the case, the changes are described below. The dataset was split into two groups of occurrences, between 10 and 25 (group 1) and greater than 25 (group 2). This was done so that a different set of variables could be used in the MaxEnt model; e.g., fewer number of variables for group 1 and higher number for group 2. This was done specifically for Feature selection and the number of iterations. Auto Features was selected for all genera unless the genera had too few occurrences and in such case the Linear (L) and Quadratic (Q) to L, Q and Product (P), L to L, Q, and

hinge threshold defaults were retained in the experimental tab. The number of iterations was set to 10 for 10-fold cross-validation to test model accuracy. The number of background points was left at the default value of 10,000 because this relates to the overall extent of the study area and is appropriate for New Guinea. The background points were not randomly assigned but adjusted to account for to the sampling bias (Elith *et al.* 2011; Syfert *et al.* 2013). Although there may be datasets collected in New Guinea with non-bias sampling strategies, the data are from multiple different sources and all were treated as biased. The bias surface was created using a kernel density estimate in the SDM Toolbox (Brown 2014), and it was used to constrain background samples so that there was similar bias between the occurrence and background points. This essentially canceled out the bias within the model (Phillips *et al.* 2009). Fade-by-clamping was selected, as predictions were made where clamping occurred, resulting in more accurate predictions (Owens *et al.* 2013).

Binary map creation

To minimize an over fit model a 5th percentile sensitivity threshold was calculated for each genus and applied to the average occurrence probability outputs from MaxEnt. The occurrence data points were used to identify the 5th percentile value. If the 5th percentile value landed between two points, the value was rounded to the nearest integer or point and this point value was used as the 5th percentile sensitivity threshold. This value was then used to create binary maps of presence-absence. For each occurrence probability map, the cell values lower than the 5th percentile value were converted to 0 (species absence) and those higher were converted to 1 (species presence). All of the binary maps were summed to create a map that showed genus richness.

Analyzing model results

To evaluate model performance, the area under the receiver operating characteristic (ROC) curve (AUC) and test sensitivity was used. The AUC is the probability that a randomly selected presence site is ranked above a randomly selected absence site and is a quantitative assessment of performance because it is independent of a chosen threshold. AUC values greater than 0.75 indicate that the model is able to accurately predict test points (Phillips & Dudík 2008) and values greater than 0.9
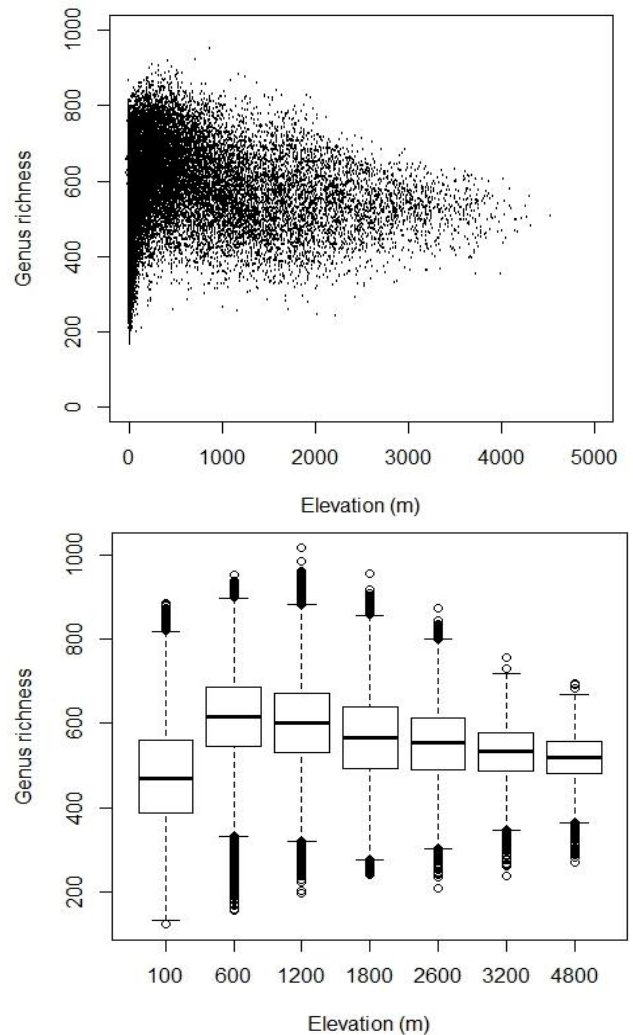
Fig. 5. Generic richness in relation to elevation. The line across each box represent the mean number of genera per elevstion category.

are considered very good (La Manna *et al.* 2011). By contrast, AUC scores lower than 0.5 indicate a worse than random predicted distribution. We reviewed each genus with a low AUC score (< 0.5) and the genera with greater than 50 occurrences were retained in the model. We felt that occurrences greater than 50 were representative distributions of each of the genera and that the lower AUC scores more likely corresponded to a more widely distributed genus (Elith *et al.* 2006; Raes & ter Steege 2007) than a poorly fit model. We report the mean AUC in our results; AUC scores for all individual genera are provided in Supplementary Material 1.

We acknowledge that some of the genus distributions may not be accurate as the occurrence data may not represent the realized niche (e.g. sink-source populations, biased, low number of
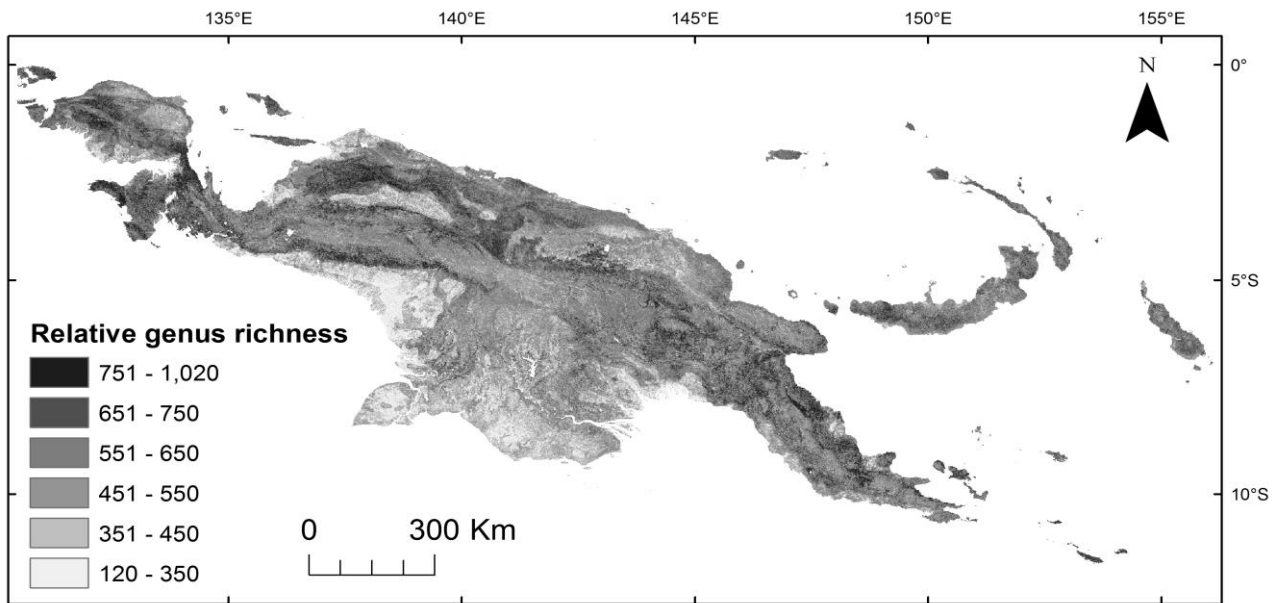
Fig. 6. The number of genera predicted to occur across New Guinea and the surrounding islands. This map is the sum of binary occurrence maps using the 5th percentile sensitivity threshold for 1354 genera. Darker colors indicate areas with higher predicted richness (1 km spatial resolution). The projection is Albers Equal Area projection, WGS84.

occurrences, time since collection). Likewise, the generalized model parameters may miss unique environments where a genus could occur. To improve our distribution modeling efforts and test the assumption that the distributions are driven by environmental parameters, we used Raes & ter Steege (2007) null-model approach. While running MaxEnt 999 times for each taxa is valid when the number of different taxa is reasonable low, it is computationally exhaustive for over 1300 genera. Also, statistically comparing null-model AUC scores to our test AUC scores did not improve model performance or predictive power and we did not want to exclude additional taxa from the study due to significant differences in AUC scores. However, we did compare null-model results for four genera with narrow to wide ranging distributions (*Nothofagus, Rhododendron, Alstonia,* and *Acaena*). We found that the AUC scores were higher than the Null AUC scores for all except *Alstonia*, which had nearly equal scores. These comparisons are available in the Supplementary Material 1. We hope that other researchers collect more occurrence data in different locations in New Guinea to validate or refute this baseline information in the future.

Results

The sampling intensity (Fig. 2) shows the number of genera collected per 50 km cell. Much of the Indonesian side has not been sampled, or at the very least, voucher specimen collections have yet to be digitized and data mobilized. Also, many cells that contain occurrences had five or fewer specimens (yellow). Sampling efforts on the Eastern half of New Guinea showed that a majority of the cells contained less than 500 collections per 50 km cell, and although this is substantially higher than the western half of New Guinea, it is still quite low.

Figure 3 shows genus richness or the number of different genera accounted for in each 50 km cell. The retention of a single genus for each cell does not account for the number of different species that were present but it provided a relative idea of the diversity of genera collected in each area. In Indonesia, the majority of cells had five or fewer genera collected and only seven cells had more than 100 genera. PNG had a larger number of collected samples overall and therefore the number of genera represented was greater.

While the total number of collected specimens alluded to a more comprehensive sampling effort, many of the 50 km cells with higher genus counts in Figs. 2 and 3 were subject to sampling bias (Fig. 4). Due to the very low sampling effort across all of Indonesia, biases were virtually nonexistent. In PNG sampling biases were higher along the roads and near areas with larger populations (towns and the Highlands region), but these areas are relative to the areas around them, that are also very low.
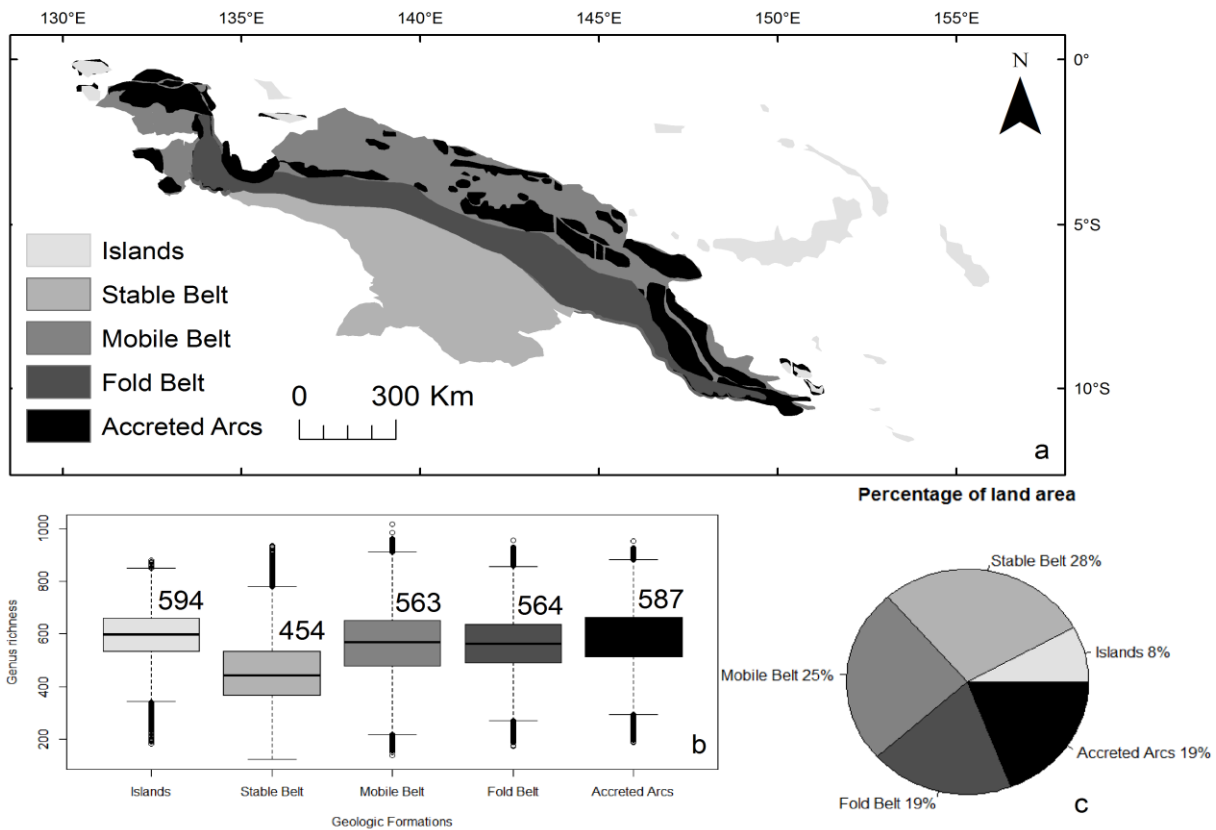
Fig. 7. (a) Major Geologic formations, (b) the predicted genus richness per geologic formation, and (c) percentage of land area each formation covers in the study area. The predicted richness box plot shows the mean number of genera per formation with outliers. The projection is Albers Equal Area projection, WGS84.

Because sampling efforts along the coast, along rivers, and near airports have occurred, low sampling biases were observed (cyan). However, these show up only because the areas around these locations had fewer, if any occurrences.

Model performance

The 10-fold cross-validation test AUC (AUC$_{cv}$) scores ranged between 0.42 and 0.99 with a mean of 0.7. There were 83 genera with AUC$_{cv}$ scores lower than the 0.50 threshold. The genera with an AUC$_{cv}$ lower than 0.50 and greater than 50 occurrences totaled 21 and achieved a mean AUC$_{cv}$ of 0.48. The number of occurrences was not correlated to the AUC$_{cv}$ score (Supplementary Material 1). A list of the genera with low AUC$_{cv}$ scores that were not included in the analysis is provided in Supplementary Material 2. Elevation, slope, and temperature annual range (BIO7) ranked, in order, as the most influential environmental and climate variables in the model. The average contribution of each environmental variable is provided in the Supplementary Material 1. Fig. 5 shows genus

richness in relation to elevation. Genus richness was greatest at elevations between 100 and 600 m and slightly decreased as elevation increased. Elevations between 0–100 m covers the largest area, and has the lowest in generic richness.

The relative, predicted genus richness for New Guinea and the surrounding islands is shown in Fig. 6. Across the study, the predicted number of genera per 1 km cell ranged between 120 and 1020, where the total number of genera possible was 1354. Darker colors show regions with higher predicted genus richness, whereas lighter colors show lower predicted richness. Across New Guinea there was higher variation in predicted richness, yet in general, the northern two-thirds of New Guinea showed higher predicted richness than the southern third.

Regardless of area, the generic richness across the different geologic land forms was similar (Fig. 7). The Islands achieved the highest predicted richness with a mean of 594 and were smallest in area (8% of land area). Accreted Arcs closely followed the Islands for predicted genus richness
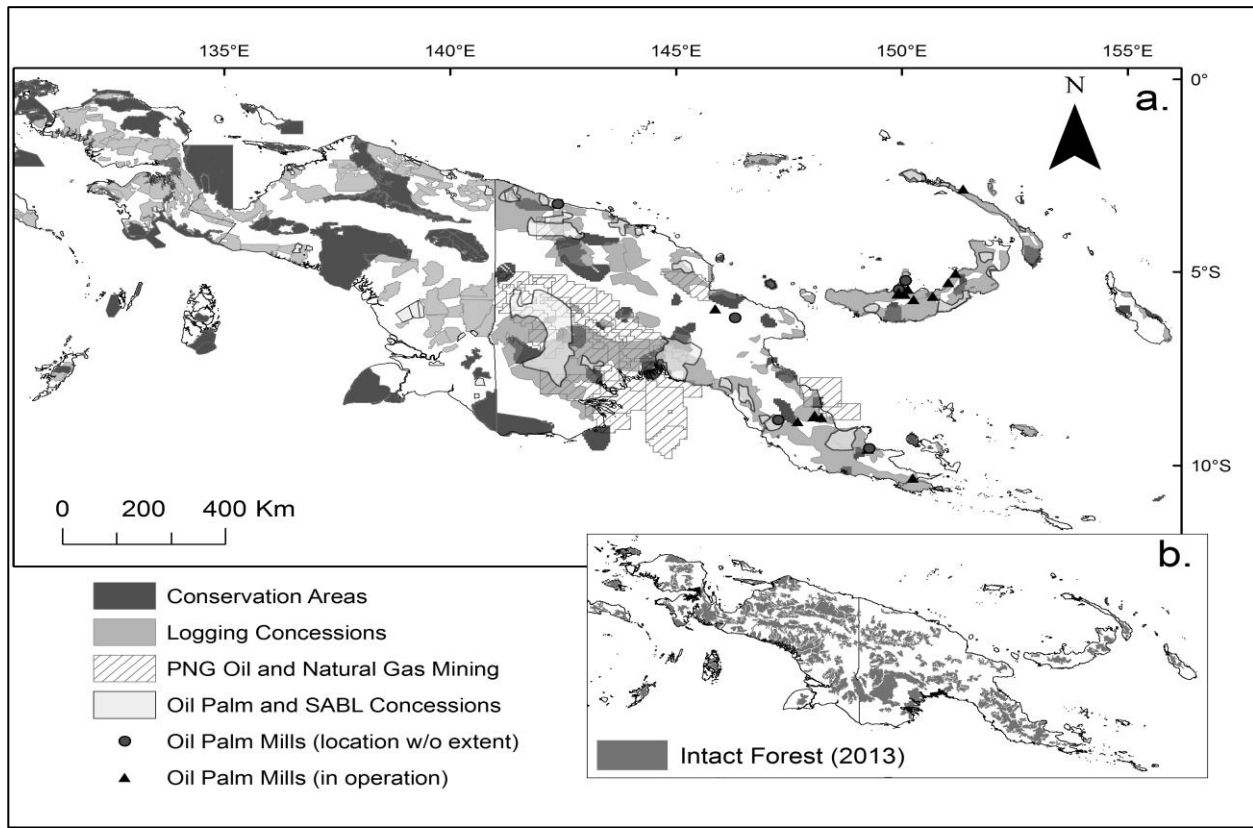
Fig. 8. New Guinea (a) conservation areas and resource extraction and (b) intact forests (Potapov *et al.* 2008). For PNG, oil palm area data were derived from Nelson *et al.* (2014) and Bryan and Shearman (2015) and most areas were simplified into composites; logging concessions (Shearman *et al.* 2009); conservation areas (IUCN and UNEP-WCMC 2016), and oil and natural gas (World Resources Institute 2016). For Indonesia spatial data were derived for oil palm, logging, (Ministry of Forestry 2010, ESRI ArcGIS online data 2016), and conservation areas (IUCN and UNEP-WCMC 2016). The projection is Albers Equal Area projection, WGS84.

with a mean of 587, and covered 19% of the land area. The Mobile Belt and Fold Belt were similar in (563 and 564, respectively) predicted genus richness but the Mobile Belt had slightly more land area at 25% compared to 19%. The Stable Belt had the lowest predicted richness with a mean of 454 and was largest in area covering 28% of the study area.

Conservation implications

Our results show predicted genus richness without consideration to land-use and land-cover (LULC) changes and such changes would influence plant distribution and community composition. The five major contributors to deforestation and land degradation were logging, subsistence agriculture, and fiber, mining and oil palm development (Abood *et al.* 2015; Bryan & Shearman 2015; Shearman *et al.* 2009). In PNG Special Agricultural and Business Leases (SABL) are designated for industrial agricultural activities, such as oil palm

development (Nelson *et al.* 2014). We provided a conservation areas and resource extraction map with data from multiple sources (Fig. 8a) and using data from Potapov *et al.* (2008) we provided a comparison of how much intact primary forest and habitat has remained unaltered (Fig. 8b). Conservation areas are loosely defined as land under some type of protection or conservation and can include community-based Wildlife Management Area (PNG specific), marine reserves, hunting reserves, national parks, and internationally recognized conservation areas (IUCN & UNEP-WCMC 2016).

Discussion

Our study provides a foundation for terrestrial plant distributions at the genera taxonomic level across New Guinea and for the surrounding islands. These results objectively and quantitatively show

collection density and spatial sampling biases and predict patterns of genus richness at the finest spatial resolution (1 km). The collection density and sampling bias maps provide guidance for future sampling strategies. However, there may be areas that have been sampled that are yet to be digitized and georeferenced, and these data may fill some gaps. Likewise, there are collections and data that have not been released to the public that could also be informative to a wide audience and for other efforts such as this. Biogeography in this region of the world is complex and is a result of a combination of abiotic and biotic drivers that influence terrestrial plant distributions and richness. The predicted patterns of genus richness conform to and differ from previous observations and theories, and we acknowledge that there is much work to be done to confirm or refute our analyses and observations.

Influence of sampling bias and density

After spatial biases and filtering, around 1083 genera had to be excluded from the analyses due to fewer than 10 occurrences. Approximately 23% or 252 of these are genera with only a single available specimen record. Specimen occurrence data eliminated from studies due the lack of presence points are not limited to New Guinea as similarities are observed for the African continent. Africa has one of the longest sampling histories, yet Stropp *et al.* (2016) identified that 31% of species in their dataset contained only one specimen. The lack of numerous collections for single taxon greatly inhibits using the modeling framework to predict distributions and inhibits our understanding of the ecology and community structure of the region.

The number of genera found in the highest sampled areas of New Guinea (Fig. 3) show that over 1000 different genera are present. Although collection biases are typically a negative aspect of distribution modeling, we can use the cells with the highest genera richness and ecological principles of the tropics to posit that the potential genera abundance across the study area may be similar. While we cannot assume that all regions in New Guinea have 1000 different genera, tropical areas typically harbour high taxonomic abundance compared to other biomes, such as grasslands. Grasslands have hundreds of plant species but are dominated by only a few, and the abundance of the non-dominant species is quite low.

In general, tropical forests lack dominant species and instead have a larger number of different taxa. ENMs are unable to predict abundance, yet they can identify environments that are more suitable for a larger number of genera according to occurrence data. Because the regions that are more comprehensively sampled are not predicted to have particularly high generic richness, it is unknown if under sampled regions would have a similar amount of generic richness or even similar communities. In this study, we lacked the data to assess locations that were progressively distant from well sampled areas, and therefore, cannot test beta-diversity or if the principle of distance-decay is applicable (Condit *et al.* 2016; Nekola & White 1999; Tobler 1970). However, the first law of geography and island biogeography theory support that the species assemblages in communities that are closer in proximity have more similarities than those more distant and this is influenced by environmental gradients and dispersal limitations (Nekola & White 1999; Stropp *et al.* 2016). This was shown by Condit *et al.* (2016) in Panama and Amazonia (Peru and Ecuador), where specie similarities declined rapidly with distance. In Panama, only 1–15% of the species were similar for plots separated by 50 km and in Amazonia only 30–40% of species were similar for plots separated by 100 km. There is a greater decline in species similarities found in New Guinea by Katovai *et al.* (2015) who found that across a 13 km transect species composition similarities ranged between 4% and 18%. From these results, Katovai *et al.* (2015) proposed that beta diversity may be higher than expected in New Guinea due to the diverse terrain that exits across the island. Overall, diversity in the tropics changes with distance and thus community composition in one location may not be similar across a larger extent, even if environmental variables are similar.

Vollering *et al.* (2015) suggested that the higher orchid richness observed in eastern New Guinea is not favored by higher collection densities because environmental conditions of occurrences are well represented and spatial biases were accounted for when modeling. We agree with this on an east-west basis for New Guinea because we also found that predicted generic richness increased west to east along the central mountain range. However, on a north-south gradient we speculate that low collection counts may influence differences in communities, as sampling intensities are very low and the environmental characteristics and geologic histories are markedly different. For instance, there is a large region of lower genus richness (genus predictions ranging between 120 and 400) which

coincides with the area with the lowest sampling intensities (Figs. 2 and 3). Similarly, the Stable Belt has fewer pockets of high genus richness compared to the Fold and Mobile Belts. We posit that the Stable Belt may be in, and of itself, unique or taxonomically similar to northern Australia, as both are part of the Australian Craton. Although these landmasses are separated by higher sea levels today, they have been united twice in the past 120,000 years during the glacial maxima when sea levels dropped 90 m. Therefore, vicariance also may explain the lower or different genus richness in the Mobile Belt. Vicariance is when a species exists in an area and then through continental drift, sea level changes, or mountain formations the taxa are separated into two locations and over time, and speciation occurs. Biota along the mountain range of the Fold Belt may share taxonomic lineage with biota found in the Stable Belt, but many have adapted to higher elevation environments. The Fold Belt also separates the Stable Belt from the Mobile Belt, and thereby the interactions between communities are minimized. As in the past, some of these isolated islands will shift towards and eventually accrete. The Mobile Belt biota also may differ from the Stable Belt because it has experienced various island accretion events and with each event, different taxa were in tow. Overall, due to the low number of samples, different geologic histories and biogeographical processes, and mountain barrier splitting New Guinea, we question whether the model performs adequately for Stable Belt. More occurrence records are needed across the Stable belt to confirm this hypothesis.

Geologic drivers and environmental variables

The geologic history, topography, and the location of New Guinea are believed to be the main drivers of plant distribution and richness. New Guinea sits at the crossroads of Southeast Asia, Australia, and many Pacific Islands, it is both at the receiving end and acts as a source area for dispersal events. Since most of these Pacific Islands east of New Guinea (Solomon Islands, Bismarck Archipelago, Fiji, Vanuatu, Samoa, Tonga) were formed from volcanic activity and tectonic plate shifts, colonization of taxa occurs from long and short distance dispersal events. New Guinea is believed to be a primary source of biota for many of the Pacific islands and the farther an island is from New Guinea, the fewer genera are present (Keppel et al. 2009). Dispersal events are continuous, yet much is left to chance and the resilience of the

traveling disperser. The populations that make it to these islands are genetic subset of the larger population and are isolated for long periods of time so speciation often results. As in the past, these isolated islands shift towards and will eventually accrete to the northern coast of New Guinea. As these islands move closer to one another dispersal is facilitated by proximity, in a stepping stone fashion or transported by carriers such as birds, bats, or humans (Boivin et al. 2016; Keppel et al. 2009).

Areas of higher predicted richness are not consistent across all areas outlined as accreted islands. We observed that it is not the accreted land that harbours the highest richness but the margins or collision zones between these accreted terrains and the Mobile Belt. As an island moves toward the north coast of New Guinea and begins the accretion process, a collision zone forms. Collision zones or successor basins overlap terrain boundaries and help to constrain the time of accretion. Successor basins can begin as submerged alluvial sediments that either dry out as ocean inlets close or are pushed above sea level from continued plate movements. Our results suggest that the interiors of nearly all basins are associated with lower richness, except the Bintuni Basin (130–135°E and 1–4°S) and have higher predicted richness. However, for the remaining basins higher predicted richness occurs outside the borders of the basins. For example, there are two large successor basins with locations centered at 137°E and 3°S (Meervlakte Basin) and 142°E and 4°S (Sepik Basin), that show genus richness to be low within the basin and higher outside the basin. We posit that this is what causes higher genus richness across the Mobile Belt (Fig. 7). In theory, collision zones are areas where the rates of species interactions and dispersal is the greatest, yet many of these regions have yet to be identified or investigated as regions with potentially high richness.

Successor basins and the surrounding areas with high genus richness are topographically homogenous areas (see Supplementary Material 1 for topographic heterogeneity map). Topographic heterogeneity and unique abiotic environments are often used as proxy data to identify regions of higher diversity rates because there are more opportunities for niche partitioning. We found the opposite to be true in some regions of New Guinea, where topographic homogeneity was associated with higher genus richness. This is supported by

Allouche *et al.* (2012) who showed that environmental heterogeneity has a unimodal response rather than a positive effect on species richness. They suggest that richness is more dependent on available area than a diverse environment. This seems to be the case in our study area, as the larger, more homogenous areas (e.g. Fig. 6 at approximately 137°E and 2.5°S) are associated with higher richness. Comparatively, we find that areas with higher environmental heterogeneity are smaller in area and tend to occur at higher elevations and where elevation gradients rapidly change. It is this response that causes slope to be the second most influential environmental variables.

The other type of basin is a foreland basin, and these occur adjacent and parallel to mountain belts and are formed through mountain belt growth and lithosphere flexion and stretching. The foreland basin (Mapenduma) is located between 135°E–140°E and 5°S–5.5°S and in this case higher predicted richness occurs along the northern edge of the basin where the mountain range begins (Mapenduma anticline). This conforms to the relationship between heterogeneous environment and higher richness.

Land-cover changes

New Guinea remains one of the last high-biodiversity wilderness areas (Mittermeier *et al.* 2003) and the loss of wilderness and forest cover is occurring at a more rapid rate than the Amazonia (Hansen *et al.* 2013). In PNG rates of forest loss between 2002 and 2014 was 0.49% per year. When forest loss was limited to accessible forests, the rates were higher at 0.61% per year. While the lack of successful conservation areas is an issue (Bryan & Shearman 2015), more attention should be drawn to illegal resource extraction, the disregard for regulations and laws, and the transparency of land leases and concessions for customary land managers (Nelson *et al.* 2014). The rates of forest loss in Indonesia are similar to those in PNG but have increased since 2000. It is estimated that Indonesia is losing 1% of primary forests per year (Miettinen *et al.* 2011). Approximately 30% of Indonesian forests in New Guinea have been degraded or deforested via industrial concessions (oil palm, logging, fiber, mixed concessions) (Abood *et al.* 2015). Across the whole of Indonesia, 41% of forests are under some type of preservation, however, Abood *et al.* (2015) identified that over 55% if Indonesian deforestation has occurred

outside industrial concessions and regulations are weakly enforced. It is unknown how much of the boundary violations occur next to preserved land in Indonesia. The comparison of Figs. 8a and 8b shows that intact forests have not always remained unscathed because resource extraction activities overlap many areas.

One should note that many of the logging concession data sources are outdated for both Indonesia and PNG. Although logging has been and will continue to be a major threat to forests in New Guinea, oil palm is resulting in forest changes quite rapidly and this is especially so across Southeast Asia (Dislich *et al.* 2016). It is estimated that oil palm accounts for 3.4% of deforestation in Indonesia and 3.0% of deforestation in PNG (Abood *et al.* 2015; Bryan & Shearman 2015). However, for many areas oil palm spatial data were not available and this is shown in PNG with dots and triangles in Fig. 8a to represent oil palm concessions and mills locations instead of geographic extents (Nelson *et al.* 2014). Likewise, Indonesian lacks adequate spatial oil palm data, as the areas devoted to it are much fewer in number and smaller in area compared to PNG.

The ecological and social impacts of oil palm were recently comprehensively addressed in a review by Dislich *et al.* (2016). Oil palm development in peat swamp forests, which constitutes 21% of concessions across the nation of Indonesia, result in long-term greenhouse gas emissions, flooding, salinization of freshwater, and high fire risk (Abood *et al.* 2015; Dislich *et al.* 2016). Slightly different ecological impacts influence the SABL land that is designated for oil palm in PNG because these areas are often used for unsustainable logging even though contracts are explicitly for industrial agricultural development. Twelve percent of PNG land area is designated as SABLs and concession boundaries are often disputed, overlap with customary tenured territories or other concessions, and do not inform or seek consent from landowners (Nelson *et al.* 2014).

Methodological limitations and considerations

Selecting the genus taxonomic level for occurrence data improved data quality for this study. Species level data were littered with issues that included a large number species with fewer than 10 occurrences, numerous data entry errors (e.g. misspelling, incorrect species identifications

according to genus listed) and missing information (e.g. coordinates). The genus level data may not fully capture the richness or endemism because some genera are more speciose than others. It is also likely that many of the rarer genera were excluded from the model because there were either too few collection points initially or after biases were accounted for the occurrence dropped below 10.

ENMs are based on the assumption that taxa are in equilibrium with the climatic envelope in which they are present, and absent in unsuitable climates. This translates to ENMs assuming the fundamental niche, or all of the locations where the species could exist. The realized niche is where the taxa actually occur. However, taxa found in the realized niche could be source or sink populations and not represent the true niche of the taxa. Due to the history of island accretion and mountain orogeny there have been relatively rapid changes in environmental gradients, which has assisted dispersal and created unique community assemblages. This violates dispersal limitations and shifts plant communities to exist in unsuitable climates for a short time periods. For example, a portion of a coastal community could be uplifted to an alpine environment over a short period of time (e.g., one Myr), and while some taxa in the alpine environment will go extinct and others will persist (Heads 2006; Trigas *et al.* 2013).

ENMs also do not integrate taxa range limitations (biotic and environmental), traits (biotic interactions, dispersal type, pollination type, lifespan (short or long lived)), or intra-species competition and this influences the predictive performance (Hanspach *et al.* 2010). In part, this is an issue of scale as the predicted distributions use climatic and environmental variables that are at regional and continental scales, whereas biotic interactions and competition are at a local scale (Austin 2002; Kumar *et al.* 2015).

The predictive performance (AUC scores) ranged among genera. Low predictive performance was observed for taxa that have a large range (low specialization) because there are fewer contrasts among the occurrence locations (Evangelista *et al.* 2008). We found this to be the case for *Ficus* and *Syzygium*, which have more than 700 occurrence points and achieved AUC scores of approximately 0.5. Similarly, highly specialized taxa do not perform well in ENM models, as the environmental conditions in which they exist are localized. A low AUC score could be caused by a narrow or wide

ranging genus but due to sampling bias and the limited spatial distribution of collections, it is unknown which is the case.

Land-use and land-cover (LULC) changes are also important to understanding terrestrial plant distributions and potential changes. Much of the biased sampling across the New Guinea occurs near airports, the coast, and populated areas. The taxa in these regions may be influenced by human induced LULC changes and thereby viable habitat may shrink or seed sources may be reduced and ultimately influence the long-term survival of certain taxa. The city of Lae, PNG, has some of the highest collection densities because it is home to the national herbarium. However, it is also densely populated and subject to LULC impacts, such impacts may cause certain species to perish.

Conclusion

It is extraordinarily difficult to tease apart the nuances and drivers of diversity in New Guinea because it is necessary to examine the ecology and evolutionary biology throughout geographic space and geologic time. While relationships can be drawn to support or refute nearly every theory concerning the biodiversity in New Guinea, such conclusions will not be adequate until there are ample collection data in which to do so and a greater understanding of biological and environmental interactions.

As suspected, we identified many areas with high genus richness in regions of high elevation and topographically heterogeneous locations. What differs from previous expectations is that we also found areas of high genus richness at low elevations, in regions that are topographically homogeneous. The difference between these two results is the area that each covers. At higher elevations and in transition zones, where topography is more complex, there are numerous smaller areas with higher richness. Comparatively, lower elevations are associated with more homogenous topography and have larger tracts of predicted genus richness. The environmental variables that most influenced these results are elevation, slope, and temperature annual range.

The geologic history is an important driver of genus richness and accreted islands often are the focus of diversity. Our results suggest that more focus should be drawn to the regions between these accretions (successor basins) as they offer ample space for niche partitioning and show many areas of predicted high genus richness. Sampling

strategies can be approached in a few ways, but any additions to the occurrence database are welcomed. Sampling efforts could focus on specific genera that have low overall occurrences or on regions that are poorly sampled. Sampling could also be focused in regions with high or low predicted richness to assess our results. Review maps and Supplemental Materials for regions and genera to focus on, as there are ample opportunities whatever avenue chosen.

The results can be used to prioritize sampling needs, support conservation strategies, compare genus diversity to other regions of the world, and discuss principles and drivers of biogeography. There are ample avenues identified for future work throughout this text, most of which cite the need for increased sampling efforts and data quality improvements. Identifying the most current LULC trends will assist in improving the success of current conservation areas and prioritizing new conservation strategies. To do this, finer resolution remote sensing data (≤ 30 m) should be paired with data from various sources, such as, government sanctioned concessions, small-scale resource extractions, illegal concessions and operations, Food and Agricultural Organization (FAO) data, and land-manager land-use. Collaborating with land-managers and communities to thwart resource development and incentivize preservation is also vital.

Acknowledgements

References

Abood, S. A., J. S. H. Lee, Z. Burivalova, J. Garcia-Ulloa & L. P. Koh. 2015. Relative contributions of the logging, fiber, oil palm, and mining industries to forest loss in Indonesia. *Conservation Letters* **8**: 58–67.

Allouche, O., M. Kalyuzhny, G. Moreno-Rueda, M. Pizarro & R. Kadmon. 2012. Area-heterogeneity tradeoff and the diversity of ecological communities. *Proceedings of the National Academy of Sciences*, U.S.A. **109**: 17495–17500.

Anderson, R. P. 2013. A framework for using niche models to estimate impacts of climate change on species distributions. *Annals of the New York Academy of Sciences* **1297**: 8–28.

Austin, M. P. 2002. Spatial prediction of species distribution: An interface between ecological theory and statistical modelling. *Ecological Modelling* **157**: 101–118.

Bishop Museum 2015. Bernice Pauhi Bishop Museum: Home page 2015). www.bishopmuseum.org (accessed on 5 July, 2015).

Baldwin, S. L., P. G. Fitzgerald & L. E. Webb. 2012. Tectonics of the New Guinea region. *Annual Reviews of Earth and Planetary Sciences* **40**: 495–520.

de Barros Ferraz, K. M. P. M., S. F. de Barros Feraz, R. C. de Paula & C. Breitenmoser. 2012. Species distribution modeling for conservation purposes. *Natureza & Conservacao* **10**: 214–220.

Bell, D. M. & D. R. Schlaepfer. 2016. On the dangers of model complexity without ecological justification in species distribution modeling. *Ecological Modelling* **330**: 50–59.

Boivin, N. L., M. A. Zeder, D. Q. Fuller, A. Crowther, G. Larson, J. M. Erlandson, T. Denham & M. D. Petraglia. 2016. Ecological consequences of human niche construction: Examining long-term anthropogenic shaping of global species distributions. *Proceedings of the National Academy of Sciences*, U.S.A. **113**: 6388–6396.

Boria, R. A., L. E. Olson, S. M. Goodman & R. P. Anderson. 2014. Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling* **275**: 73–77.

Brooks, T. M., R. A. Mittermeier, G. A. B. da Fonseca, J. Gerlach, M. Hoffmann, J. F. Lamoreux, C. G. Mittermeier, J. D. Pilgrim & A. S. L. Rodrigues. 2006. Global biodiversity conservation priorities. *Science* **313**: 58–61.

Brown, J. L. 2014. SDMtoolbox: A python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *Methods in Ecology and Evolution* **5**: 694–700.

Bryan, J. E. & P. Shearman. 2015. *The State of the Forests of Papua New Guinea 2014: Measuring Change Over the period 2002-2014*. University of Papua New Guinea. Port Moresby, PNG.

Bryan, J. E. & P. L. Shearman. 2008. *PNG Resource Information System Handbook*. 3rd edition.

University of Papua New Guinea.

Chase, M. W., M. J. M. Christenhusz, M. F. Fay, J. W. Byng, W. S. Judd, D. E. Soltis, D. J. Mabberley, *et al.* 2016. An update of the Angiosperm Phylogeny Group classification for the orders & families of flowering plants: APG IV. Botanical *Journal of the Linnean Society* **181**: 1–20.

Condit, R., N. Pitman, E. G. Leigh Jr. J. Chave, J. Terborgh, R. B. Foster, P. Núñez S. Aguilar, R. Valencia, G. Villa, H. C. Muller-landau, E. Losos, S. P. Hubbell, 2016. Beta-diversity in tropical forest trees. *Science* **295**: 666–669.

Conn, B. J., L. L. Lee & R. (2004+). (http://www. pngplants. org/PNGdatabase. Kiapranis. 2004. PNGplants database: Plant collections from Papua New Guinea. http://www.pngplants.org/PNGdatabase.

Cronk, Q. C. B., M. Kiehn, W. L. Wagner & J. F. Smith. 2005. Evolution of Cyrtandra (Gesneriaceae) in the Pacific Ocean: The origin of a supertramp clade. *American Journal of Botany* **92**: 1017–1024.

Dislich, C., A. C. Keyel, J. Salecker, Y. Kisel, K. M. Meyer, M. Auliya, A. D. Barnes, *et al.* 2016. A review of the ecosystem functions in oil palm plantations, using forests as a reference system. *Biological Reviews of the Cambridge Philosophical Society*: **92**: 1539–1569.

Dormann, C. F., J. Elith, S. Bacher, C. Buchmann, G. Carl, G. Carré, J. R. G. Marquéz, *et al.* 2013. Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**: 027–046.

Elith, J., C. H. Graham, S. Ferrier, A. Guisan, R. P. Anderson, M. Dudı, R. J. Hijmans, *et al.* 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**: 129–151.

Elith, J., S. J. Phillips, T. Hastie, M. Dudík, Y. E. Chee & C. J. Yates. 2011. A statistical explanation of MaxEnt for ecologists. *Diversity & Distributions* **17**: 43–57.

ESRI ArcGIS online data. 2016. *Enviornmental Systems Resesarch Institute: ArcGIS Desktop: Release 10.3.* Redlands, CA.

Evangelista, P. H., S. Kumar, T. J. Stohlgren, C. S. Jarnevich, A. W. Crall, J. B. Norman & D. T. Barnett. 2008. Modelling invasion for a habitat generalist and a specialist plant species. *Diversity & Distributions* **14**: 808–817.

GBIF. 2015. *Papua New Guinea and Indonesia occurance-only data.* Global Biodiversity Information Facility.

Hansen, M. C., P. V Potapov, R. Moore, M. Hancher, S. A. Turubanova & A. Tyukavina. 2013. High-resolution global maps of 21st-centruy forest cover change.

Science **342**: 850–853.

Hanspach, J., I. Kuhn, S. Pompe & S. Klotz. 2010. Predictive performance of plant species distribution models depends on species traits. *Perspectives in Plant Ecology, Evolution and Systematics* **12**: 219–225.

Heads, M. 2001. Regional patterns of biodiversity in New Guinea plants. *Botanical Journal* of the *Linnean Society* **136**: 67–73.

Heads, M. 2006. Biogeography, ecology and tectonics in New Guinea. *Journal of Biogeography* **33**: 957–958.

Heads, M. 2009. Dispersal and vicariance. pp: 947–950. *Molecular Panbiogeography of the Tropics.* University of California Press, Berkely, CA, USA.

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

Hill, K. C. & R. Hall. 2003. Mesozoic–Cenozoic evolution of Australia's New Guinea margin in a west Pacific context. *Geological Society of America Special Papers* **372**: 265–289.

Hope, G. 2014. The sensitivity of the high mountain ecosystems of New Guinea to climatic change and anthropogenic impact. *Arctic, Antarctic, and Alpine Research* **46**: 777–786.

ISRIC-World Soil Information. 2013. Soil Grids: an automated system for global soil mapping, Accessed Jan. 2015. http://www.isric.org/.

IUCN and UNEP-WCMC. 2016. The World Database on Protected Areas (WDPA), Accessed May 2016. www.protectedplanet.net.

Katovai, E., D. D. Katovai, W. Edwards & W. F. Laurance. 2015. Forest structure, plant diversity and local endemism in a highly varied New Guinea landscape. *Tropical Conservation Science* **8**: 284–300.

Keppel, G., A. J. Lowe & H. P. Possingham. 2009. Changing perspectives on the biogeography of the tropical South Pacific: Influences of dispersal, vicariance and extinction. *Journal of Biogeography* **36**: 1035–1054.

Kumar, S., E. G. LeBrun, T. J. Stohlgren, J. A. Stabach, D. L. McDonald, D. H. Oi & J. S. LaPolla. 2015. Evidence of niche shift and global invasion potential of the Tawny Crazy ant, *Nylanderia fulva. Ecology and Evolution* **5**: 4628–4641.

Leisz, S. J., J. B. Burnett & A. Allison. 2000. Consensus Report Climate Change and Biodiversity in Melanesia: What do we know? https://www.researchgate.net/publication/239931053_Consensus_Report_Climate_Change_and_Biodiversity_in_Melanesia_What_do_we_know.

Mack, A. & L. E. Alonso. 2000. *A Biological Assessment of*

the Wapoga River area of Northwestern Irian Jaya, Indonesia. Conservation International.

Mandl, N., M. Lehnert, M. Kessler & S. R. Gradstein. 2010. A comparison of alpha and beta diversity patterns of ferns, bryophytes and macrolichens in tropical montane forests of southern Ecuador. *Biodiversity and Conservation* **19**: 2359–2369.

La Manna, L., S. D. Matteucci & T. Kitzberger. 2011. Modelling Phytophthora disease risk in *Austrocedrus chilensis* forests of Patagonia. *European Journal of Forest Research* **131**: 323–337.

McGavin, G. 2009. *Scientific Expidition to Mount Bosavi, Southern Highlands, Papua New Guinea.* Report. Oxford University Museum of Natural History and the Department of Zoology, Oxford University, UK. http://downloads.bbc.co.uk/springwatch/llotv_finalreport_20090907.pdf

Miettinen, J., C. Shi & S. C. Liew. 2011. Deforestation rates in insular Southeast Asia between 2000 and 2010. *Global Change Biology* **17**: 2261–2270.

Ministry of Forestry. 2010. *Pemantauan Hutan di Indonesia (Forest Monitoring in Indonesia). Overview of the forest monitoring activities by the Ministry of Forestry of Indonesia.* (http://www.unredd.net/index.php?option=com_docman&task=doc_download.

Mittermeier, R. A. C. G. Mittermeier, T. M. Brooks, J. D. Pilgrim, W. R. Konstant, G. A. da Fonseca & C. Kormos. 2003. Wilderness and biodiversity conservation. *Proceedings of the National Academy of Sciences, U.S.A.* **100**: 10309–10313.

Neall, V. E. & S. A. Trewick. 2008. The age and origin of the Pacific islands: a geological overview. *Philological Transaction of Royal Society*: *Biological Sciences* **363**: 3293–3308.

Nekola, J. C., & P. S. White. 1999. The distance decay of similarity in biogeography and ecology. *Journal of Biogeography* **26**: 867–878.

Nelson, P. N., J. Gabriel, C. Filer, M. Banabas, J. A. Sayer, G. N. Curry, G. Koczberski, and O. Venter. 2014. Oil palm and deforestation in Papua New Guinea. *Conservation Letters* **7**: 188–195.

de Oliveira, G., T. F. Rangel, M. S. Lima-Ribeiro, L. C. Terribile & J. A. F. Diniz-Filho. 2014. Evaluating, partitioning, and mapping the spatial autocorrelation component in ecological niche modeling: A new approach based on environmentally equidistant records. *Ecography* **37**: 637–647.

Owens, H. L., L. P. Campbell, L. L. Dornak, E. E. Saupe, N. Barve, J. Soberon, K. Ingenloff, A. Lira-Noriega, C. M. Hensz, C. E. Myers & A. T. Peterson. 2013. Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecological Modelling* **263**: 10–18.

Peterson, A. T., J. Soberon, R. G. Pearson, R. P. Anderson, M. Martinez-Meyer, Nakamura & M. Bastos Araujo. 2011. *Ecological Niches and Geographic Distributions*. Princeton University Press, Princeton, NJ.

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

Phillips, S. J. & M. Dudık. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* **31**:161–175.

Phillips, S. J., M. Dudík, J. Elith, C. H. Graham, A. Lehmann, J. Leathwick & S. Ferrier. 2009. Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications* **19**: 181–197.

Polhemus, D. & J. T. Polhemus. 1998. Assembling New Guinea: 40 million years of island arc accretion as indicated by the distributions of aquatic Heteroptera (Insecta). *Biogeography and Geological Evolution of SE Asia*: 327–340.

Potapov, P., A. Yaroshenko, S. Turubanova, M. Dubinin, L. Laestadius, C. Thies, D. Aksenov *et al.* 2008. Mapping the world's intact forest landscapes by remote sensing. *Ecology and Society* **13**: 51:16.

Radosavljevic, A. & R. P. Anderson. 2014. Making better Maxent models of species distributions: Complexity, overfitting and evaluation. *Journal of Biogeography* **41**: 629–643.

Richards, S. J. & B. G. Gamui. 2011. *Rapid Biological Assessments of the Nakanai Mountains and the Upper Strickland Basin: Surveying the Biodiversity of Papua New Guinea's Sublime Karst Environments.* RAP Bulletin Biological Assessment.

Roos, M. C., P. J. A. Keßler, S. R. Gradstein & P. Baas. 2004. Species diversity and endemism of five major Malesian islands: diversity-area relationships. *Journal of Biogeography* **31**: 1893–1908.

Shearman, P. L., J. Ash, B. Mackey, J. E. Bryan & B. Lokes. 2009. Forest conversion and degradation in Papua New Guinea 1972–2002. *Biotropica* **41**: 379–390.

Shearman, P. L. & J. Bryan. 2011. A bioregional analysis of the distribution of rainforest cover, deforestation and degradation in Papua New Guinea. *Austral Ecology* **36**: 9–24.

Sidder, A. M., S. Kumar, M. Laituri & J. S. Sibold. 2016. Using spatiotemporal correlative niche models for evaluating the effects of climate change on mountain pine beetle. *Ecosphere* **7**:12.

Stropp, J., R. Ladle, A. C. M. Malhado, J. Hortal, J. Gaffuri, W. Temperley, J. O. Skøien & P. Mayaux.

2016. Mapping ignorance: 300 years of knowledge of flowering plants in Africa. *Global Ecology* and *Biogeography* **25**: 1085–1093.

Supriatna, J., Y. de Fretes, A. Mack, Y. P. Yeager, S. Olivieri, J. B. Burnett, I. Wijayanto, S. Suryadi & A. Suryadi. 1999. *The Irian Jaya Biodiversity Conservation Priority- Setting Workshop.* Washington, DC.

Syfert, M. M., M. J. Smith & D. A. Coomes. 2013. The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. *PLoS ONE* **8**: e55158

Takeuchi, W. 2007. *Vascular Plants of the Kaijende Highlands, Papua New Guinea: Taxonomic and Vegetation Survey*, RAP Bulletin Biological Assessment: A Rapid Biodiversity Assessment of the Kaijende Highlands, Enga Province, Papua New Guinea. Conservation International.

Tobler, W. R. 1970. A Computer movie simulating urban growth in the Detroit region supplement. *Economic Geography* **46**: 234–240.

Trigas, P., M. Panitsa & S. Tsiftsis. 2013. Elevational gradient of vascular plant species richness and endemism in crete. The effect of post-isolation mountain uplift on a continental island system. *PLoS*

Tuanmu, M. N. & W. Jetz. 2015. A global, remote sensing-based characterization of terrestrial habitat heterogeneity for biodiversity and ecosystem modelling. *Global Ecology* and *Biogeography* **24**: 1329–1339.

Veloz, S. D. 2009. Spatially autocorrelated sampling falsely inflates measures of accuracy for presence-only niche models. *Journal of Biogeography* **36**: 2290–2299.

Vollering, J., A. Schuiteman, E. de Vogel, R. van Vugt & N. Raes. 2015. Phytogeography of New Guinean orchids: Patterns of species richness and turnover. *Journal of Biogeography* **43:** 204–214.

Wisz, M. S., R. J. Hijmans, J. Li, A. T. Peterson, C. H. Graham, A. Guisan, J. Elith, *et al.* 2008. Effects of sample size on the performance of species distribution models. *Diversity and Distributions* **14**: 763–773.

World Resources Institute. 2016. *Papua New Guinea Exploratory Oil and Natural Gas Mining Concessions.* http://pubs.wri.org/pubs_dataset.cfm?PubID=3874.

Supplementary Information

Additional Supporting information may be found in the online version of this article:

Supplementary Material 1: List of genera used in analysis with Test AUC scores and additional Figs.

Supplementary Material 2: List of genera with low AUC scores and fewer than 50 occurrences.

Supplementary Material 3: List of genera with too few occurrences to use in the model.