Exploring dimensions of biodiversity in Japanese ferns

Short running title: Biodiversity of Japanese ferns

Joel H. Nitta[1], Brent D. Mishler[2], Wataru Iwasaki[1,3], and Atsushi Ebihara[4]

[1]Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo, Japan

[2]University and Jepson Herbaria, and Department of Integrative Biology, University of California, Berkeley, CA, USA

[3]Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan

[4]Department of Botany, National Museum of Nature and Science, Tsukuba, Japan

Author for correspondence:
Joel H. Nitta
Email: joelnitta@gmail.com

Acknowledgements

Members of the Iwasaki Lab at the University of Tokyo provided comments that improved the manuscript. Shawn Laffan provided advice about CANAPE. This study was supported by JSPS KAKENHI Grant Number 16H06279.

Abstract

Aim To characterize multiple facets and drivers of biodiversity and understand how these relate to bioregions and conservation status.

Location Japan

Taxon Ferns

Methods We used data from >300,000 fern specimens in Japan to compile a community dataset of 1,240 0.2° grid-cells including 674 native, non-hybrid taxa. We combined this with a phylogenetic tree and functional trait data to analyze taxonomic, phylogenetic, and functional diversity, and modeled the distribution of biodiversity in response to environmental factors and reproductive mode. We conducted categorical analysis of neo- and paleo-endemism to identify areas of significant phylogenetic endemism. We used hierarchical clustering of taxonomic and phylogenetic distances to analyze bioregions. We compared the overlap of significantly diverse grid-cells with existing conservation zones to assess conservation status.

Results We observe three major patterns in the spatial distribution of biodiversity: that of taxonomic richness, which is highest at mid-latitudes, that of phylogenetic and functional diversity and phylogenetic endemism, which are highest in small southern islands, and that of relative phylogenetic and functional diversity, which are high at both high and low latitudes, and low at mid-latitudes. Grid-cells were grouped in to three (phylogenetic) or four (taxonomic) major bioregions. Temperature and apomixis were identified as likely drivers of patterns of taxonomic and phylogenetic diversity. Conservation status (% protected area) is generally higher than the overall rate for grid-cells with significantly high biodiversity.

Main conclusions The integrative approach used here reveals previously undetected patterns and drivers of biodiversity in the ferns of Japan. The different distributions of each diversity type reflect the major bioregions. Although peak areas of different diversity types are largely non-overlapping, each one has a relatively high protection status. Despite this, future conservation efforts must be cognizant of dangers not addressed here such as climate change and herbivory by deer.

Key words: apomixis, biodiversity, biogeography, CANAPE, ferns, Japan, phylogenetic diversity.

Introduction

Characterizing the spatial distribution of biodiversity is a major goal of evolutionary biology with two equally important aspects: to understand the processes generating it, and to conserve it. The vast majority of studies seeking to understand the spatial distribution of biodiversity focus on species richness. Hotspots, or areas of exceptionally high richness, have received particular attention, resulting in a widely-recognized set of 36 terrestrial hotspots (Myers, Mittermeier, Mittermeier, da Fonseca, & Kent, 2000; Noss, 2016) and motivating conservation strategies to preserve them (Brooks et al., 2006; Margules & Pressey, 2000). However, richness alone cannot provide a complete picture of biodiversity. All organisms are related to a greater or lesser degree by descent from a common ancestor, and these evolutionary relationships must be taken into account to obtain a full understanding of the biodiversity present in an area.

Presumably, the main reason phylogeny has not been taken into account more prominently in biodiversity studies is because data (DNA sequences), extensive spatial data, and analytic tools have only become available relatively recently. These tools and datasets now make it possible to analyze other dimensions of biodiversity, such as phylogenetic diversity and phylogenetic endemism. Better ways to characterize biological regions ("bioregions") are also available, rather than relying on ad-hoc characterizations (e.g., Laffan et al., 2016; Daru et al., 2020b; White, Dey, Mohan, Stephens, & Price, 2019). Furthermore, incorporating these two frameworks—the categorization of areas into phyloregions with analysis of over/under dispersion of biodiversity—can provide particularly powerful insights into the processes structuring biodiversity and suggest conservation priorities.

However, a comprehensive understanding of the relationships between richness and other metrics requires densely sampled data, and such datasets are rare on the regional (country) scale. The ferns of Japan are excellent model system because they have been the target of intense botanical interest for several decades and are densely sampled (reviewed in Ebihara & Nitta, 2019). One particularly valuable characteristic of this flora is the availability of data on reproductive mode (known for 500 of 721 native pteridophyte taxa excluding hybrids [69.3%]), which is likely to affect genetic diversity (Bengtsson, 2003), and thereby higher-level biodiversity (Krueger-Hadfield, Blakeslee, & Fowler, 2019).

Japan is situated along a latitudinal gradient spanning seasonal, temperate areas in the north to subtropical, mostly aseasonal islands in the south. Furthermore, it is a mountainous country with great variation in elevation on the larger islands. Plant distributions often reflect physiological adaptations to climate (e.g., precipitation and temperature). Therefore, we expect the spatial distribution of biodiversity to be determined by climatic variation in addition to reproductive mode.

Here we leverage this exceptionally rich dataset to analyze the geographic distribution of biodiversity in detail. We ask the following questions in our study system, the ferns of Japan: 1. How is biodiversity distributed? 2. How is biodiversity structured? 3. What environmental and biological factors influence biodiversity? 4. How well is biodiversity protected?

Materials and Methods

All computational analyses were carried out in R v4.0.0 (R Core Team, 2020) unless otherwise stated. The 'targets' R package was used to maintain analysis workflow (Landau, 2021).

Study Site

Our data include all ferns known to occur within the boundaries of Japan (20°N to 45°N, 122°E to 153°E). Japan consists of four main islands and thousands of smaller ones (Fig. 1). Most of the main islands have been in contact with the continent at various points in the past during periods of lower sea level, but the Ogasawara archipelago consists of oceanic islands that have never been connected to the mainland. The climate is subtropical in the south and temperate in other parts of the country. Map data were downloaded from the Geospatial Information Authority of Japan under the Creative Commons Attribution License v4.0 (https://www.gsi.go.jp/kankyochiri/gm_japan_e.html).

Occurrence data

We used a list of all non-hybrid fern specimens deposited at TNS (herbarium codes follow Thiers2021). We chose to use this over other sources (e.g., GBIF), because of its high taxonomic quality: all identifications are standardized to a common taxonomy, the Japan Green List v. 1.01 (http://www.rdplants.org/gl/). Furthermore, the overall sampling completeness of this dataset is extremely high (Fig. S1.1 in Appendix S1 in Supporting Information). The original list comprises 337,023 specimens representing 674 terminal taxa (hereafter defined as species, subspecies, forms, and varieties unless otherwise indicated). We filtered out any occurrences not within a map of standard grid squares defined by the Statistics Bureau of Japan (http://www.stat.go.jp/english/data/mesh/05.html) and removed duplicate collections (300,169 specimens, 674 taxa after filtering).

A necessary step in any analysis of biodiversity is to set the grain size used to accurately define communities. Here, one must consider that smaller grain size is needed to detect environmental effects, while larger grain size is needed to ensure adequate species sampling. We tested using 0.1°, 0.2°, 0.3°, and 0.4° grid-cells (size refers to length of side of a grid-cell) for defining communities (co-occurring species). At each grain size, species occurrences were converted to a presence-absence community matrix (a species was considered present if at least one specimen was recorded in that grid-cell). We calculated sampling redundancy (1 - richness/number of specimens) (Garcillán, Ezcurra, & Riemann, 2003) to quantify adequacy of sampling. Preliminary analysis indicated that 0.2° grid-cells are optimal for our dataset (Fig. S1.2). We filtered out any cells with redundancy < 0.1 and used this size for subsequent analyses.

Morphological trait data

We used traits originally compiled for identification of ferns and lycophytes of Japan (Ebihara, 2016, 2017). Continuous traits were measured on 10 randomly selected specimens per species. Fertile (*ie*, spore-bearing) and sterile fronds were measured separately, then averaged. Qualitative traits were scored by observing voucher specimens. All qualitative traits were scored in binary format; *eg*, a single trait with three states was formatted as three binary traits. From the original trait list, we selected only traits with putative ecological function (Table 1), and removed any that had Pearson correlation coefficient >0.6 or fewer than three observations of a given trait state.

Reproductive mode data

We used data compiled by Ebihara & Nitta (2019), which classifies each fern taxon as sexual, apomictic, mixed (both sexual and apomictic modes known from the same taxon), or unknown.

We calculated the percentage of apomictic species in each grid cell as the sum of apomictic and mixed taxa divided by total number of taxa (Fig. S1.3b).

Environmental data

We downloaded climate data at the 2.5 minute scale from the WorldClim database using the 'getData' function in the R package 'raster' (Hijmans, 2020), and extracted four variables related to our research questions as follows: mean annual temperature (bio1; hereafter "temperature"), temperature seasonality (bio4), annual precipitation (bio12; hereafter "precipitation"), and precipitation seasonality (bio15) (Fig. S1.3). We identified the intersection of the climate data at the 2.5 minute scale with the 0.2° grid-cells, then calculated the mean of each climatic variable for each 0.2° grid-cell. We calculated area as rolling mean area (km^2) across 1 degree latitudinal bands. This approach reflects the idea that the amount of area with a given climate regime enclosing each grid cell may influence species richness of that cell (Fahrig, 2013), and that using raw island areas would not be appropriate in this context as there are just four very large islands with many much smaller ones located close by. The R package 'sf' was used for GIS analyses (Pebesma, 2018).

Phylogenetic analysis

Sequences for the plastid, coding *rbcL* gene are available for ca. 97% of the Japanese fern flora (Ebihara & Nitta, 2019). However, using this dataset alone for phylogeny suffers two drawbacks: it is not densely sampled enough to resolve many internal nodes, and it lacks many lineages that have fossils available for molecular dating. Because one goal of the study is to understand the distribution of paleo- vs. neo-endemic areas (defined in units of time), we require an ultrametric tree. To obtain a maximally sampled, robust phylogenetic tree, we combined the *rbcL* data of Ebihara & Nitta (2019) (706 taxa) with a globally sampled plastid dataset including all fern sequences on GenBank for four widely-sequenced plastid genes (*atpA*, *atpB*, *rbcL*, and *rps4*; 4,492 taxa) and 58 other coding, single-copy plastid genes extracted from 123 complete fern plastomes (Nitta et al., in prep.) (4,787 OTUs total, including outgroup taxa). We conducted phylogenetic analysis with IQTREE v1.6.12 (Nguyen, Schmidt, von Haeseler, & Minh, 2015) and molecular dating with TREEPL v1.0 (Smith & O'Meara, 2012) using 27 fossil calibration points after Testo & Sundue (2016), with the exception of treating fossil *Kuylisporites* as belonging to crown *Cyathea*, not *Cyathea* + *Alsophila* (Loiseau et al., 2020). We then trimmed the tree to Japanese taxa only and used this for all further analyses.

Phylogenetic signal

We tested for phylogenetic signal in continuous traits using two alternative metrics, Blomberg's *K* (Blomberg, Garland Jr., & Ives, 2003) and Pagel's lambda (Pagel, 1999). *K* describes the ratio between the observed variance in a trait vs. the amount of variance expected under Brownian Motion (BM); lambda is a scaling parameter that transforms the phylogeny such that trait values fit those most closely expected under BM. Both lambda and $K = 1$ under BM; lower values indicate less signal (overdispersion) and higher values indicate more signal (clustering). We measured *K* and lambda using the 'phylosig' function in the R package 'phytools' (Revell, 2012), which also calculates significance based on a randomization test for *K* and a likelihood ratio test for lambda. We tested for phylogenetic signal in qualitative (coded as binary) traits using Fritz and Purvis' *D* (Fritz & Purvis, 2010) with the 'phylo.d' function in the R package 'caper' (Orme et al.,

2018), which also assesses significance by comparing the observed value to a null distribution of random values and a simulated distribution of values expected under BM.

Analysis of biodiversity

We measured taxonomic diversity using taxonomic richness.

We measured phylogenetic diversity (PD) using the Faith's PD (the total branch length connecting all terminal taxa in a community, including the root of the phylogeny; Faith, 1992). To detect areas with unusually long or short branches, we calculated relative phylogenetic diversity (RPD), which is a ratio of observed PD to PD measured on a tree where all branch lengths have been transformed to equal length (Mishler et al., 2014).

We measured functional diversity (FD) following the method of Petchey & Gaston (2002). We first calculated functional distances using Gower's method (Gower, 1971) on the trait data with the 'gowdis' function in the R package 'FD' (Laliberté & Legendre, 2010). We weighted continuous and qualitative traits equally, weighted each binary component trait of each qualitative trait equally, and log-transformed and scaled continuous traits prior to calculating distances. We also tested alternative weighting schemes, but these produced very similar results (data not shown). We then used the distances to build a trait dendrogram using the 'hclust' function in the 'stats' R package (R Core Team, 2020), and used the dendogram in place of a phylogenetic tree to calculate PD and RPD as described above. These functional analogs of PD and RPD are hereafter referred to as FD (functional diversity) and RFD (relative functional diversity).

We measured phylogenetic endemism (PE) as the sum of all branches at a site inversely weighted by their range size (Rosauer, Laffan, Crisp, Donnellan, & Cook, 2009).

Observed values of PD, RPD, FD, RFD, and PE are known to be related to species richness: as more taxa are drawn at random without replacement from a given set, it becomes increasingly unlikely to draw a taxon that is distantly related (or functionally dissimilar) to the others. Therefore, we determined whether the observed metrics are more or less diverse than random for a given number of taxa by conducting a randomization test. We generated 999 random communities using the "independent swap" algorithm of Gotelli (2000), which conserves richness per site and species occurrence frequencies while randomizing species identities, then compared the observed value to the distribution of random values. Statistical significance was assessed using a one-tailed test for PE or a two-tailed test for other metrics with alpha = 0.5; observed values in the extreme 5% of the null distribution were considered significant. We also calculated the standard effect size (SES) of each biodiversity metric. SES measures how extreme the observed value (*obs*) is relative to the null distribution (*null*):

$$SES = \frac{obs - mean(null)}{sd(null)}$$

We tested whether centers of endemism have an over-representation of old (paleoendemic) or new (neoendemic) lineages using categorical analysis of neo- and paleo-endemism (CANAPE), which involves measuring PE with an alternate tree where all branch lengths are equal, then comparing the ratio between raw PE and alternative PE (RPE) (Mishler et al., 2014). Low RPE indicates short range-restricted branches (neoendemism) and high RPE indicates long range-restricted branches (paleoendemism). Since all of these measures are affected by species richness, categorization of endemism is done using the rank of each observed variable relative to those calculated for the null distribution.

All measures of biodiversity and CANAPE were carried out using the 'canaper' package (Nitta, 2021), which calculates raw biodiversity metrics using functions in the 'phyloregion' package (Daru et al., 2020b) and randomizations using the 'picante' package (Kembel et al., 2010).

It is possible that endemism may be affected by species that are actually wide-ranging, but only have a small range in Japan (a "border effect"). This is especially expected for southern, sub-tropical islands, which harbor many tropical species with small ranges in Japan but wider ranges outside of the country. To partially account for this, we also calculated all endemism-related metrics using a dataset including species found only in Japan.

Spatial models

We first tested for spatial autocorrelation in biodiversity metrics (richness, SES of PD, SES of FD, SES of RPD, SES of RFD) and independent variables (environmental variables and % apomictic taxa) using Moran's I with the 'moran.mc' function in the 'spdep' package (Bivand, Pebesma, & Gómez-Rubio, 2013). In all cases, significant spatial autocorrelation was detected (Moran's I > 0, *P* value of permutation test < 0.05) (Table S1.1). Therefore, we used spatially explicit methods for modeling.

We checked for correlation between independent variables while taking into account spatial autocorrelation using a modified *t*-test with the 'modified.ttest' function in the 'SpatialPack' R package (Vallejos, Osorio, & Bevilacqua, 2020). This indicated that temperature seasonality and % apomictic taxa were correlated with mean annual temperature (Table S1.2), so for the main modeling analysis we only included the following environmental variables: mean annual temperature, precipitation, and precipitation seasonality.

We constructed a linear mixed model for each biodiversity metric dependent on the environmental variables ("environmental models"). Initial inspection of the relationship between each biodiversity metric and temperature showed a hump-shaped pattern, so we also included a quadratic term for temperature only. Richness was fit with the negative binomial response family; others were Gaussian. We fit models including a Matérn correlation matrix based on the geographic centroids of each grid cell as a random effect with the R package 'spaMM' (Rousset & Ferdy, 2014). We verified the significance of each term of the model by conducting likelihood ratio tests (LRTs) between the full model and models each with a single term of interest removed, in sequence. We also verified the significance of the full model relative to the null model (only including the Matérn correlation matrix) using LRTs.

We postulated that % apomictic taxa might explain the distribution of phylogenetic diversity better than temperature, so we also constructed a limited set of models to test this hypothesis ("reproductive models"). These models included biodiversity metrics with a phylogenetic component (SES of PD and SES of RPD) as dependent on % apomictic taxa, precipitation, and precipitation seasonality, with random spatial effects as described above. We compared the fit between environmental models and reproductive models for SES of PD and SES of RPD using conditional Akaike Information Criterion (cAIC) (Courtiol & Rousset, 2017; Vaida & Blanchard, 2005).

Analysis of bioregions

We analyzed bioregions using the framework of Daru et al. (2020a) and their R package 'phyloregion'. Briefly, this involves first calculating beta diversity (change in composition between

communities) using either species names (taxonomic bioregions) or phylogenetic distances (phylogenetic bioregions). For taxonomic bioregions, we calculated the turnover in species between sites due to species replacement (Baselga, 2012) with the Sørensen index. For phylogenetic bioregions, we calculated phylogenetic dissimilarities based on the PhyloSor index (Bryant et al., 2008). The distances are used to construct a dendrogram, which is then split into into *k* bioregions. As the optimal number of bioregions cannot be known *a-priori*, we tested values of *k* from 1 to 20.

Assessment of conservation status

We downloaded SHP files of protected areas in Japan from the Japan Ministry of the Environment (https://www.biodic.go.jp/biodiversity/activity/policy/map/map17/index.html). We only considered areas that include protection of plants, and categorized protected areas as either "high" (no human activities allowed at all) or "medium" (some economic activities allowed by permit) following (Kusumoto et al., 2017). We cropped grid-cells to only include land area, then calculated the percent of protected area of each status type within grid-cells with high levels of biodiversity as measured with PD, FD, PE, or taxonomic richness. For PD, FD, and PE, significance was assessed with a one-tailed test at alpha = 0.05 against the 999 null communities described above. For taxonomic richness, cells ranked in the top >5% were considered highly diverse.

Results

Datasets

The phylogeny included 663 taxa (97.93% of the total number of native, non-hybrid ferns) representing 98 genera and 34 families. The topology was in general agreement with other recent, large plastid fern phylogenies (Lehtonen, 2011; Schuettpelz & Pryer, 2007; Testo & Sundue, 2016) (Fig. S1.4). Ferns were recovered as monophyletic, as were all families and genera except for *Microsorum* (families follow Pteridophyte Phylogeny Group I (2016)). The age of the fern clade is estimated to be 433.6 ma.

The cleaned list of specimens included 300,169 specimens representing 674 taxa (99.56%). The distribution of collection locations was highly skewed, with > 3,000 specimens per grid-cell northwest of Tokyo in Saitama prefecture, and many fewer elsewhere (Fig. S1.5a). Despite this, sampling redundancy was generally good (> 0.3; González-Orozco, Thornhill, Knerr, Laffan, & Miller, 2014) throughout the country (mean 0.55 ± 0.17, final community data matrix; errors are SD unless mentioned otherwise) (Fig. S1.5b). The final community data matrix included 674 taxa and 1,240 grid-cells. 146 grid-cells were excluded due to low redundancy (< 0.1); these were mostly from Hokkaido and coastal areas.

After removing correlated and low-variance traits, the final trait set included one binary trait, three continuous traits, and three qualitative traits (coded as 39 binary traits) (Table 1). The trait matrix included 648 taxa (95.7%) and 43 traits, with 0.75% missing data.

Phylogenetic signal

All continuous traits showed significant phylogenetic signal, but estimates of signal strength differed between *K* and lambda: all lambda values were > 0.95, indicating evolution by BM; however, values of *K* were < 0.1, indicating less phylogenetic signal than expected under BM

(Table S1.3). 9 out of 40 of continuous traits (coded as binary) showed as much or more phylogenetic signal as expected under BM (D =< 0); others had D > 0 (Table S1.4).

Observed diversity

Taxonomic richness is lowest on the northernmost island of Hokkaido, then increases with decreasing latitude until reaching a maximum in southern Kyushu, then decreases again in smaller islands further south (Fig. 2a). Observed values of PD and FD were highly correlated with richness as expected (Fig. 2b, c), showing an initial steep slope that gradually starts to level off at higher richness (Fig. S1.6a, c). Observed values of RPD and RFD were not linearly related to richness, however, they showed high variance at low species richness, and low variance at high species richness (Fig. S1.6b, d).

Randomization tests for phylogenetic and functional diversity

Grid-cells with significantly low PD were almost all located on Honshu, the main island of Japan (Fig. 3a). A smaller number of cells had significantly high PD, and these were located mostly on smaller southern (Ryukyu, Izu, Ogasawara) islands or coastal areas (Kii Peninsula). Grid-cells with significantly low RPD were mostly observed in southern Honshu, particularly in and around Kyushu (Fig. 3b). Grid-cells with significantly high RPD are located both in the southern islands and northern Honshu and Hokkaido. Results of the randomization test for functional diversity were broadly similar to phylogenetic diversity (Fig. 3c, d).

Relationships between biodiversity and climate or reproductive mode

All models were significant relative to null models without any fixed effects (Table S1.5). The effect of temperature dominated (had the greatest absolute t-value) in environmental models (Fig. 4a–e), and % apomictic taxa dominated in reproductive models (Fig. 4f–g); temperature and % apomictic taxa were significant in each model where they were included, whereas other predictors (area, precipitation, precipitation seasonality) were not significant in some models. Richness showed a hump-shaped relationship with temperature, reaching maximum richness at intermediate temperatures (Fig. 5a). Phylogenetic and functional diversity showed either moderately (SES of PD, SES of FD) to strongly (SES of RPD, SES of RFD) hump-shaped relationships with temperature (Fig. 5b–e). SES of RPD showed a clear, linear decline with % apomictic taxa (Fig. 5g); a similar pattern was detected in SES of PD, but with wider confidence intervals (Fig. 5f). Conditional AIC was lower (indicating better model fit) in the reproductive model compared to the environmental model for SES of RPD and SES of PD (Table S1.6).

Categorical analysis of neo- and paleo-endemism

Nearly all grid-cells in islands south of Kyushu (Ryukyu and Ogasawara) have significant levels of phylogenetic endemism, with low (non-significant) endemism observed elsewhere (Fig. 6). The vast majority of grid-cells with significant endemism are mixed. Concentrations of paleoendemism were observed in the Ryukyu archipelago, on Okinawa, Miyako, and Iriomote Islands. Small clusters of cells containing significant endemism were also observed northwest of Tokyo (Nagano prefecture), and in Hokkaido (Fig. 6). When we ran CANAPE with a reduced dataset including only taxa absolutely restricted to Japan to account for the border effect, there

were fewer significant cells, but a similar pattern of significant endemism mostly occurring in the southern islands was observed (Fig. S1.7).

Bioregions

Analysis of bioregions grouped grid-cells into eight or five clusters (Fig. S1.8) on the basis of taxonomic or phylogenetic distances, respectively. The vast majority of grid-cells fell to a subset of these clusters: using taxonomic distances, they include Cluster 1 (Hokkaido and northern Honshu), Cluster 2 (southern Honshu), Cluster 3 (Ryukyu Islands, excluding Yakushima), and Cluster 4 (Ogasawara Islands) (Fig. 7a; Fig. S1.9a). Phylogenetic distances produced similar results, but the Ryukyu and Ogasawara Islands merged into a single cluster (Fig. 7b; Fig. S1.9b). The other much smaller clusters are likely artifacts of insufficiently low taxon richness for robust clustering, and are not discussed further.

Taxonomic clusters 3 and 4 have higher SES values of PD, RPD, FD, and RFD than clusters 1 and 2 (Fig. 8). Clusters 3 and 4 also have consistently high PE *p*-scores, whereas clusters 1 and 2 have a much wider variation. Cluster 2 has lower RPD and RFD than cluster 1. Comparison of biodiversity metrics across phylogenetic clusters was similar, except that phylogenetic cluster 3 mostly corresponds to taxonomic clusters 3 and 4 combined.

Conservation status

The percentage of protected area in grid-cells with significantly high biodiversity range from 14.7% to 33.2% total, including 2.2% to 8.6% with high status and 12.5% to 24.6% with medium status (Fig. 9). Areas with high PE were the best protected (8.6% high, 24.6% medium, 33.2% total). Cells with significantly high biodiversity generally had a greater percentage of area protected than the overall rate for Japan (2.6% high, 9.5% medium, 12.1% total); only FD was lower than the overall rate for highly protected areas.

Discussion

Distribution and drivers of biodiversity

The latitudinal hump-shaped pattern of species richness observed in ferns here (Fig. 2a) is well-documented (Ebihara & Nitta, 2019; Kusumoto et al., 2017), and similar patterns have been observed in deciduous broadleaved trees and conifers (Shiono et al., 2015), and niche-modeled distributions of woody plants (Fukaya, Kusumoto, Shiono, Fujinuma, & Kubota, 2020). Our spatial modeling analysis indicates that this is strongly linked to temperature, with maximum richness occurring at intermediate temperatures (Fig. 5a). Previous studies have identified temperature, water availability (Bickford & Laffan, 2006; Qian, Wang, Li, Xiao, & Wang, 2012), or a combination of these (Link-Pérez & Laffan, 2018; Nagalingum et al., 2015; Qian, Kessler, Deng, & Jin, 2021) to be important drivers of spatial fern diversity. A hump-shaped pattern of richness along elevation (maximum richness at intermediate elevations) has been frequently observed in ferns on mountains in tropical or sub-tropical areas (Kessler, Kluge, Hemp, & Ohlemüller, 2011). Hypotheses to explain such a pattern include purely geometrical constraints as well as biological hypothesis positing maximally favorable environmental conditions at mid-elevations (Colwell, Rahbek, & Gotelli, 2004). Another reason for the greater richness in southern Honshu compared to smaller southern islands is that southern Honshu harbors high elevational variation that may support a wider range of taxa than is present in the southern islands.

The overwhelming pattern characterizing phylogenetic and functional diversity (PD and FD) is the high diversity of southern islands (Ryukyu and Ogasawara Isl.) (Fig. 3a, c). This is almost certainly due to the presence of primarily tropical lineages that do not occur in other parts of the country; 16 genera and four families only occur south of 30.1° latitude (just south of Yakushima Isl.). The similarity in patterns of PD and FD is likely due to the moderate degree of phylogenetic signal present in the traits used in our analysis, and suggests that in ferns, phylogeny can be used as a reasonable stand-in for functional diversity. Given their high phylogenetic diversity, it is perhaps not surprising that southern islands also host high amounts of phylogenetic endemism; indeed, the vast majority of cells with significant levels of PE, as well as paleoendemic cells, are located in the southern islands. This is clearly due to the small land area of these islands, which are the only areas with a subtropical climate in Japan.

Another striking pattern revealed here is the predominance of long branches in the northern half of Honshu and in southern islands, and the predominance of short branches in the southern half of Honshu, and particularly on Kyushu (Fig. 3b). The long branches in the southern islands are likely being driven by the tropical lineages mentioned previously; those in the north may be due to refugial lineages that are distantly related to others and become more species rich at high-latitudes, such as *Equisetum* (Fig. S1.10a). The preponderance of short branches in southern Honshu may be due to the radiation of certain species-rich genera in that area such as *Deparia* (Fig. S1.10b) and *Dryopteris* (Fig. S1.10d).

Furthermore, we identified % apomictic taxa as a strong driver of SES of RPD (Fig. 5g), and to a slightly lesser extent SES of PD (Fig. 5f). Indeed, *Dryopteris* is one such genus with a high rate of apomictic taxa (Fig. S1.11), and other genera with high rates of apomixis (*Cyrtomium*, *Pteris*; Fig. S1.11) also reach greatest richness in southern Honshu (Fig. S1.10). This is consistent with the hypothesis that communities with high proportions of asexual taxa have lower genetic diversity, and is the first time to our knowledge that this has been demonstrated in ferns. We verified that this relationship is also observed when analyzed with a phylogeny with branch lengths in units of genetic change (*ie*, the output of the maximum likelihood phylogenetic analysis, before dating), a more direct measure of genetic diversity (Fig. S1.12b). While temperature is also correlated with % apomictic taxa and therefore cannot be discounted, conditional AIC indicates a better fit of the reproductive model than the environmental model for SES of RPD (Table S1.6). Ultimately, it is likely that temperature, in combination with other environmental conditions, is driving the predominance of apomictic taxa: Tanaka, Isaka, Hattori, & Sato (2014) also found that the proportion of apomictic taxa reaches a maximum at ca. 34.8 °N latitude and increases with temperature in Japan (their study was limited to Honshu and did not include the southern islands). Tanaka et al. (2014) asserted that one limiting factor affecting establishment of apomictic ferns at high latitudes or cold climates is the large cell size of apomicts, which are typically polyploid. As fern gametophytes are only a single cell-layer in thickness with no cuticle or stomata, it is expected they would be particularly susceptible to freezing damage induced by larger cells with higher water content; however, no experimental studies have tested these hypotheses.

Water limitation is thought to play an important role in promoting apomixis in ferns especially in deserts (Grusz et al., 2021), since ferns depend on water for transfer of sperm to egg, and apomictic plants would be able to reproduce in the absence of water. A similar argument has been made to assert that areas with seasonal monsoons are linked to increased rates of apomictic ferns (Liu et al., 2012; Tanaka et al., 2014), but in Japan, although the areas including high percentages of apomictic taxa are monsoonal, the degree of seasonal water limitation does

not approach that of a desert. The monsoon hypothesis is also not supported by our results, which show precipitation seasonality did not play a strong role in any model and was not correlated with % apomictic taxa.

The variation in distribution of different biodiversity metrics is reflected by the bioregions analysis, which tended to group cells with similar biodiversity values (Fig. 8). Interestingly, the bioregions identified here correspond roughly to major forest types previously categorized on an *ad-hoc* basis; Cluster 1 corresponds to the cool or subartic deciduous type, Cluster 2 to the warm evergreen type, and Clusters 3 and 4 to the subtropical rainy type (Shimizu, 2014). This suggests that ferns are useful as bioindicators in Japan. Although some studies have emphasized the importance of deep-sea straits (the Tsugaru Strait separating Hokkaido from the rest of Honshu and the Tokara Strait separating the southern islands from Kyushu) in structuring biological communities in Japan (Kubota, Hirao, Fujii, Shiono, & Kusumoto, 2014; Millien-Parra & Jaeger, 1999), these did not play a major role in structuring fern bioregions. The Tsugaru Strait had no relation to any of the bioregions, and although the Tokara Strait somewhat splits clusters 2 and 3, there is some overlap between them on the Amami-Oshima Isl. (Fig. 7). The weak effect of these straits as barriers makes sense given the presumably high dispersal abilities of ferns, which disperse via tiny spores carried on the wind. Rather, fern bioregions in Japan broadly separate along gradients of temperature and precipitation; in particular, precipitation splits taxonomic Cluster 3 (mostly the Ryukyu islands; wetter) from Cluster 4 (mostly the Ogasawara islands; drier) (Fig. S1.13).

Conservation status

Although the different facets of biodiversity have different distribution patterns, we found that grid-cells with significantly high amounts of biodiversity tend to be relatively well protected compared to the baseline protection rate, regardless of biodiversity metric (Fig. 9). One possible reason is that in the past, some ferns have been included in lists of rare and endemic plants for consideration in designation of conservation areas (A. Ebihara, pers. comm.). Another (as suggested previously) is that ferns may serve as bioindicators reflecting overall levels of biodiversity, particularly for other plant groups.

Despite the good news that ferns enjoy relatively strong protection status in Japan, we must bear in mind that there are other threats posed that conservation zones cannot ameliorate, the most obvious being climate change. One other such threat is herbivory by native Japanese deer, *Cervus nippon*. Populations of these deer have increased rapidly since the 1970s due to a combination of climate change and a decrease in human hunters and natural predators (Takatsuki, 2009). At dense populations, the deer can prevent forest regeneration and result in plant communities dominated by a limited number of species unpalatable to deer (Takatsuki, 2009), although lower population sizes may promote plant diversity as per the intermediate disturbance hypothesis (Nishizawa, Tatsumi, Kitagawa, & Mori, 2016). They have caused extensive damage even within national parks (Tokita, 2006), which have the highest protection status. Although various efforts to prevent deer herbivory have been deployed such as culling and exclusion fences (Takatsuki, 2009), much work remains to be done to save vulnerable plant populations.

Limits of the study and future outlook

There are a number of caveats that must be kept in mind when interpreting these results, but these also suggest avenues for future research. First, the data include specimens collected over

a wide time period, so they do not necessarily represent the current distribution of ferns in Japan. Furthermore, the collections are not evenly collected throughout the country, but rather some areas have many more specimens than others (Fig. S1.5). Indeed, one promising avenue of research may be to take advantage of the most densely sampled areas to investigate changes in abundance over time.

The extreme evolutionary distinctiveness of southern islands is both a major feature of the Japanese fern flora and a challenge for analysis. The tropical lineages occurring in these islands are likely responsible for the comparatively low (*ie*, clustered) values of PD observed on Honshu (Fig. 3a). If a more restricted null model was used that did not consider all taxa capable of dispersing anywhere within the study area, different results would likely be obtained for the randomization test (Mishler et al., 2020). However, ferns are known to have high dispersal abilities (Barrington, 1993; Moran & Smith, 2001; Tryon, 1970). In our dataset, the taxon with the greatest latitudinal distribution, *Coniogramme intermedia* Hieron., spans nearly the entire region (24.2341°N to 45.4583°N), and 17 taxa have latitudinal distributions spanning the Tokara strait in the south (29°N) to the Tsugaru strait in the north (41.5°N). Therefore, it seems reasonable to assume that ferns are capable of dispersing across the entire study area.

One element of biodiversity that we did not consider here is hybrids. The Japanese fern flora includes 373 hybrid taxa. Although hybrids are often thought of as evolutionary “dead-ends”, in some cases they are able to interbreed with diploids and thus contribute to diversification (Barrington, Haufler, & Werth, 1989). Future studies should consider the distribution of hybrids in Japan and their possible effects on biodiversity. Methods for calculating PD would need to be modified to take into account multiple lineages contributing to one terminal taxon.

Finally, a long under-appreciated aspect of fern ecology is the role of the gametophyte. Unlike seed plants, ferns have gametophytes that are capable of growing independently from the sporophyte. Futhermore, the two stages of the life cycle have vastly different morphology and ecophysiology, and may even occur over partially to completely disjunct ranges (reviewed in Pinson, Chambers, Nitta, Kuo, & Sessa, 2017). Here, as in the vast majority of fern ecology studies, we consider only the sporophyte stage. However, recent studies in Japan (Ebihara et al., 2019, 2013) and elsewhere (Nitta, Meyer, Taputuarai, & Davis, 2017) have revealed different patterns in the distribution of gametophytes and sporophytes. The comprehensive molecular sampling of sporophytes in Japan makes this area ideal for conducting high-throughput DNA sequencing analyses to compare patterns of biodiversity between life stages in ferns across large spatial scales.

Trait	Significance	Reference
Indusium (b)	Protects developing spores	Poppinga, Haushahn, Warnke, Masselter, & Speck (2015)
Frond width (c)	Scales with overall plant size	Creese, Lee, & Sack (2011)
No. pinna pairs (c)	Less divided fronds have lower rates of evapotranspiration	Kluge & Kessler (2007)
Stipe length (c)	Shorter lengths minimize resistance	Watkins Jr., Holbrook, & Zwieniecki (2010)
Frond margin (q)	Thermoregulation	Nicotra et al. (2011)
Frond shape (q)	Less divided fronds have lower rates of evapotranspiration	Kluge & Kessler (2007)
Frond texture (q)	Thicker leaves have lower rates of evapotranspiration	Givnish (1987)

Table 1. Fern traits used in this study. Letters in parentheses indicate trait type: b, binary; c, continuous; q, qualitative. Qualitative traits were coded as binary for analysis (see Materials and Methods).

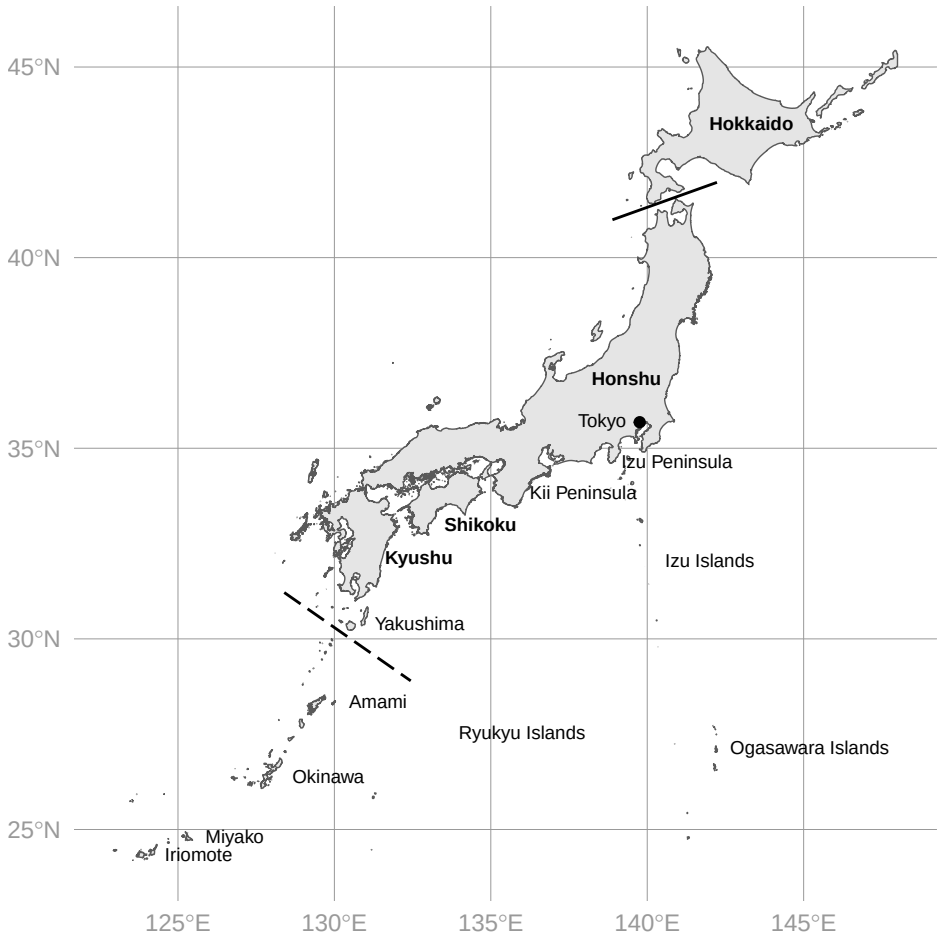

Fig. 1. Map of Japan showing names of places mentioned in the text. Names of the four main islands in bold. The Ryukyu Islands include Yakushima, Amami, Okinawa, Miyako, and Iriomote. Solid line shows location of Tsugaru Strait; dashed line shows Tokara Strait.

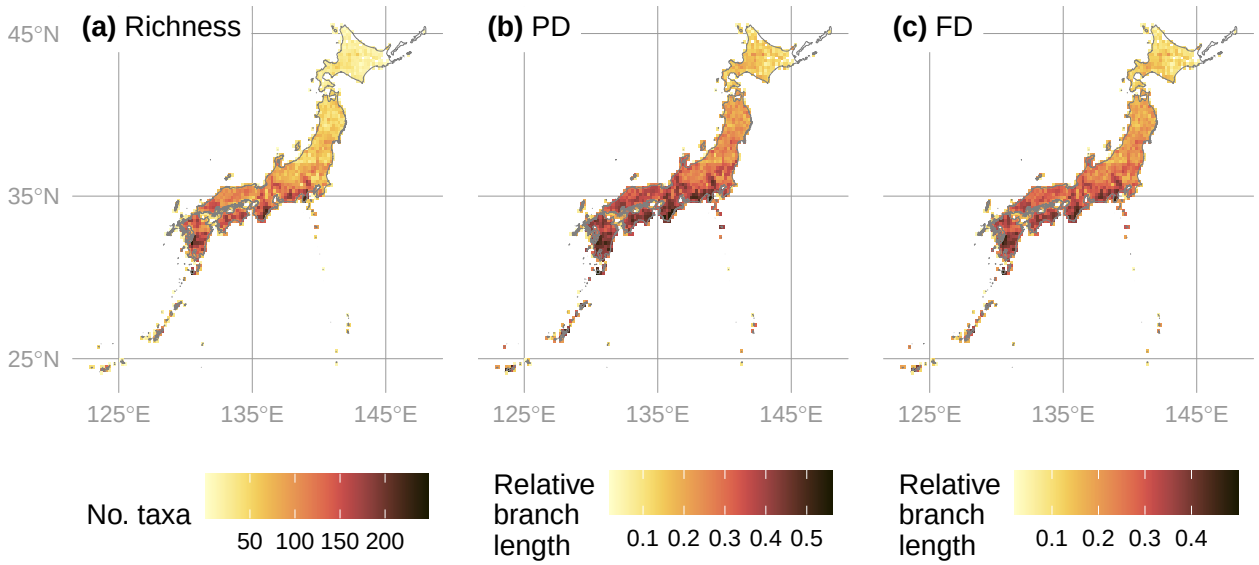

Fig. 2. (a) Raw taxonomic diversity (b), phylogenetic diversity, and (c) functional diversity of the ferns of Japan. For (b) and (c), raw branchlengths were transformed to relative values by dividing each branchlength by the sum of all branchlengths.

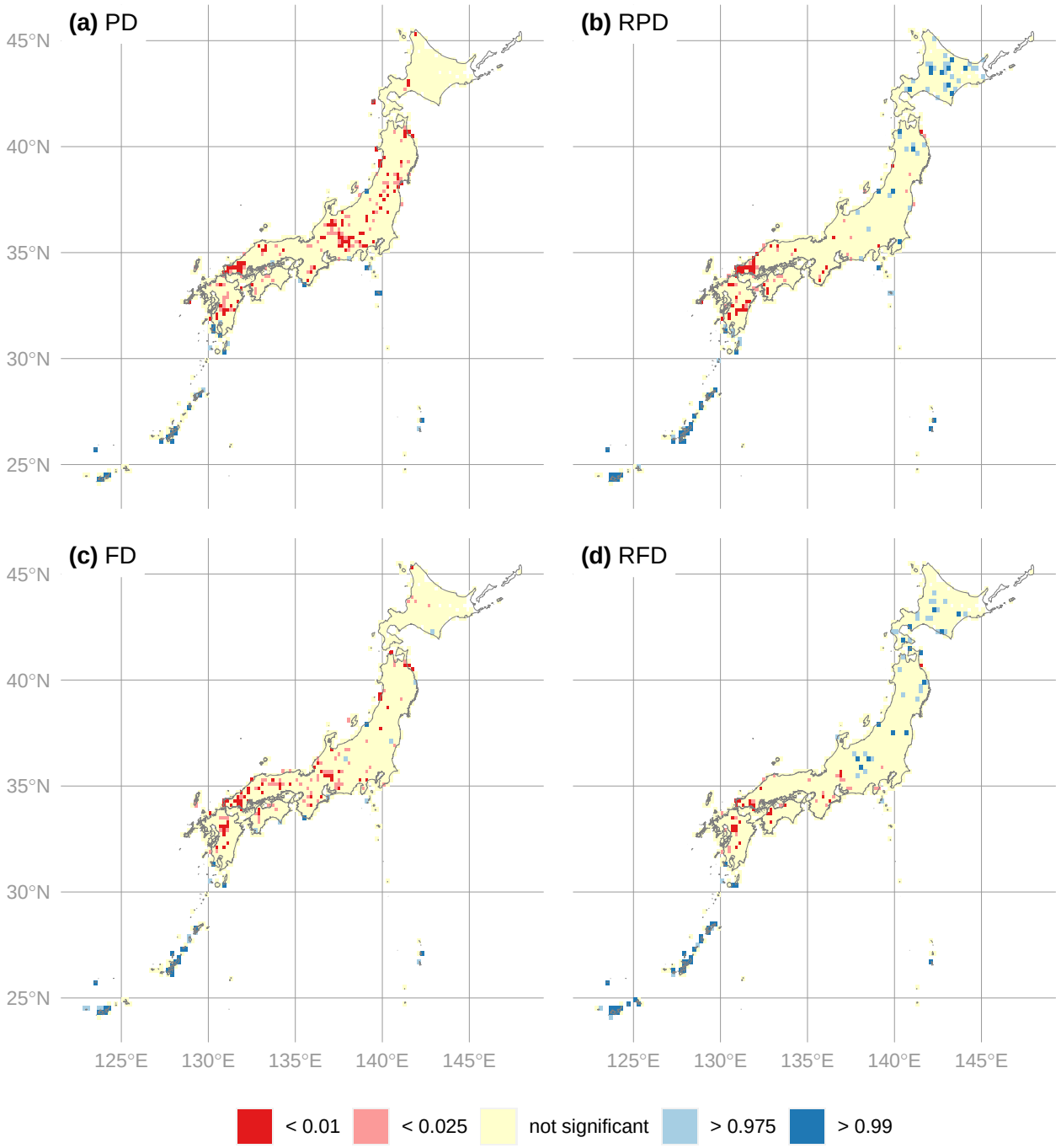

Fig. 3. Results of randomization test for phylogenetic and functional diversity of the ferns of Japan. (a), Phylogenetic diversity (PD). (b), Relative phylogenetic diversity (RPD). (c), Functional diversity (FD). (d), relative functional diversity (RFD). For each metric, raw values were compared to a null distribution of 999 random values, and the *P*-value calculated using a two-tailed test. For details, see Materials and Methods.

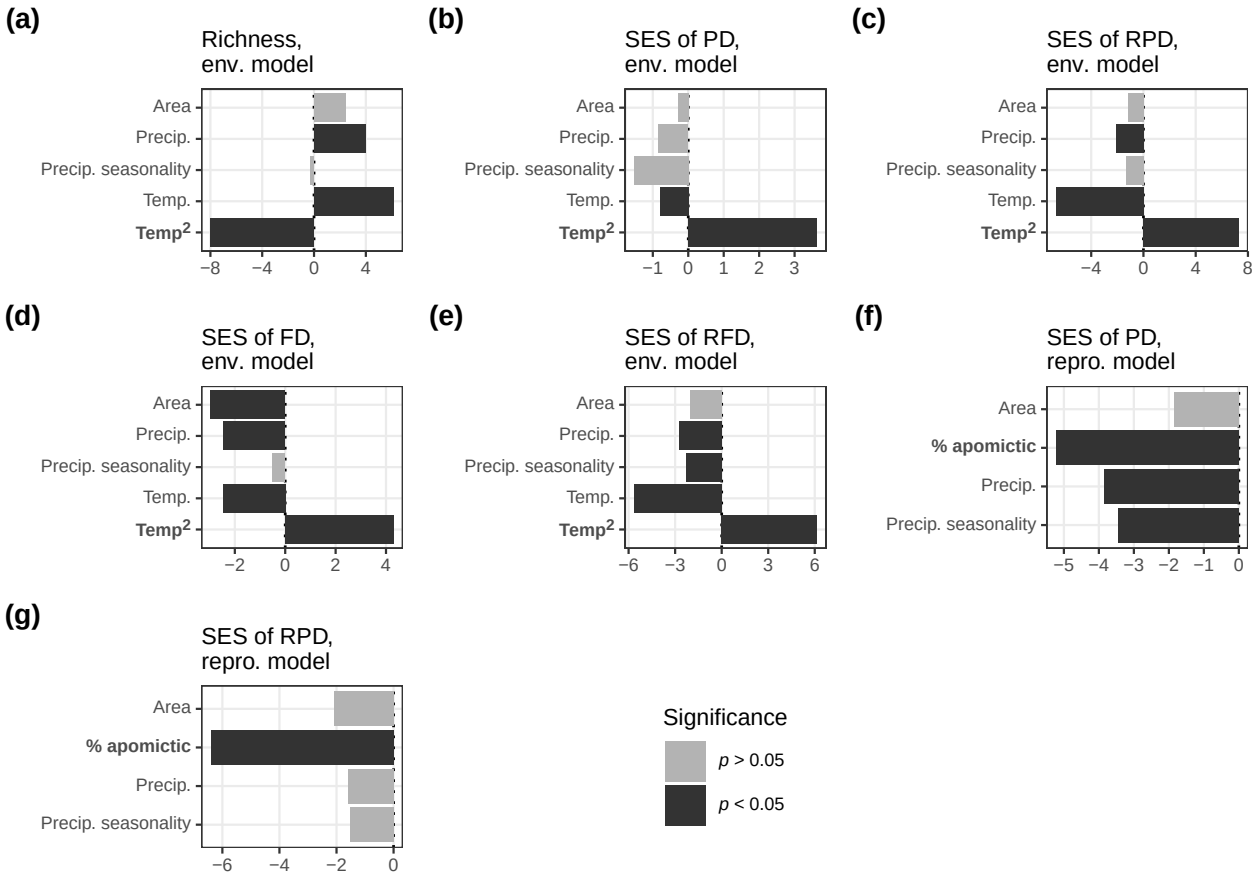

Fig. 4. Regression coefficient *t*-values (coefficient divided by standard error) of general mixed models. (a)–(e), environmental models (models including temperature); (f) and (g), reproductive models (models including % apomictic taxa). Statistical significance assessed with a likelihood ratio test (LRT) between the full model and a model with the focal independent variable removed (null model); *p*-value indicates probability of no difference between the full and null model. Name of response variable and model type indicated above each subplot. Independent variable with greatest absolute *t*-value in bold for each model.

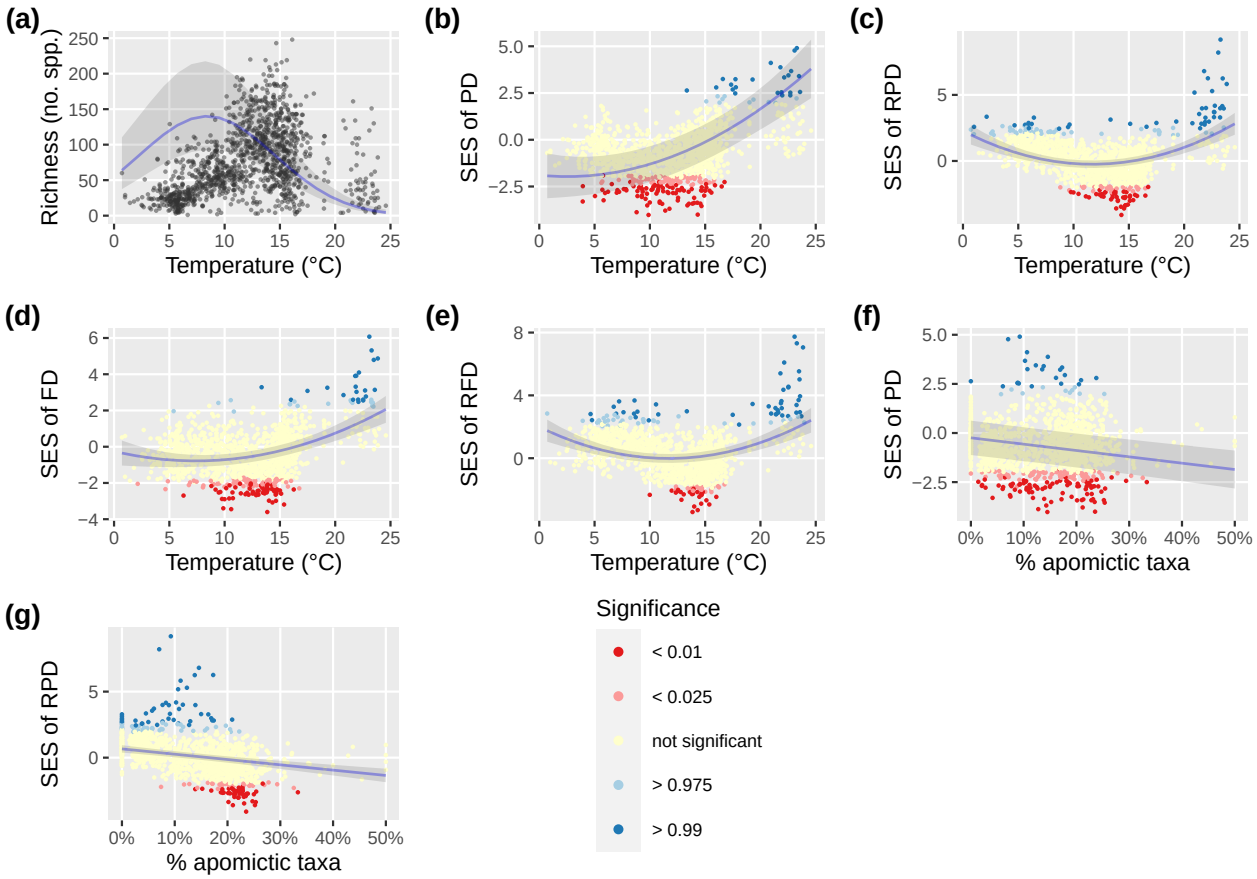

Fig. 5. Relationship between biodiversity metrics and selected predictor variables. (a)–(e), environmental models (models including temperature); (f) and (g), reproductive models (models including % apomictic taxa). Ribbon shows 95% confidence interval of model. Line fit with the focal predictor variable while averaging over other predictors. For phylogenetic diversity (PD), relative PD (RPD), functional diversity (FD), and relative FD (RFD), the standard effect size (SES) was calculated by comparing raw values to a null distribution of 999 random values, and significance (*P*-value) calculated using a two-tailed test (see Materials and Methods).

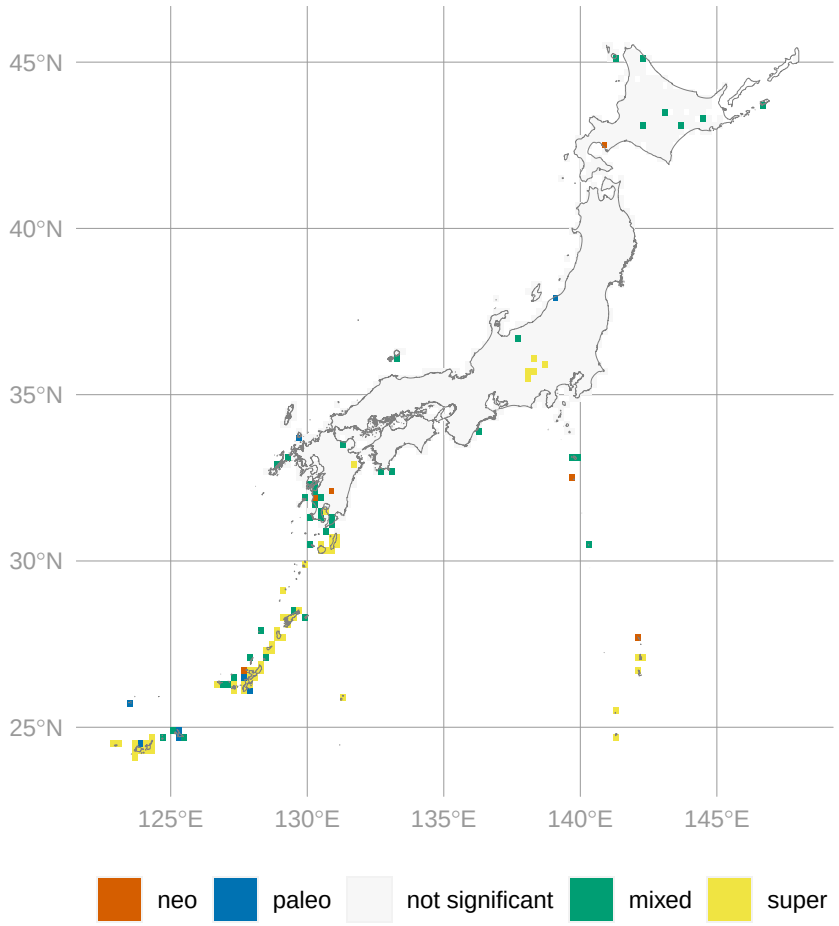

Fig. 6. Phylogenetic endemism of the ferns of Japan measured using CANAPE (categorical analysis of neo- and paleo-endemism). 'paleo' indicates areas with an overabundance of rare long branches; 'neo' indicates areas with an overabundance of rare short branches; 'mixed' indicates significantly endemic sites that have neither an overabundance of short nor long branches; 'super' indicates highly significantly endemic sites that have neither an overabundance of short nor long branches. For details, see Materials and Methods.

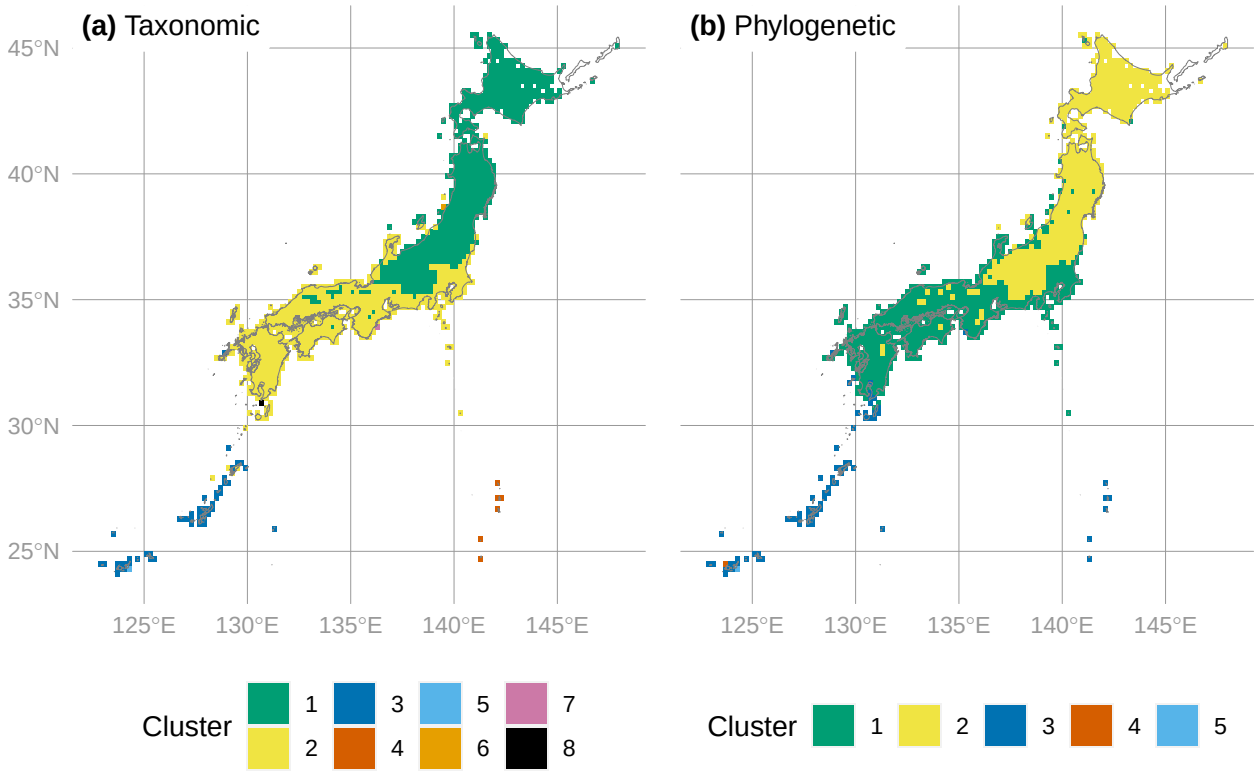

Fig. 7. Bioregions of the ferns of Japan. (a), taxonomic bioregions. (b), phylogenetic bioregions. Bioregions determined by clustering taxonomic (Sørensen) or phylogenetic (PhyloSor) distances into *k* = 8 or 7 groups, respectively.

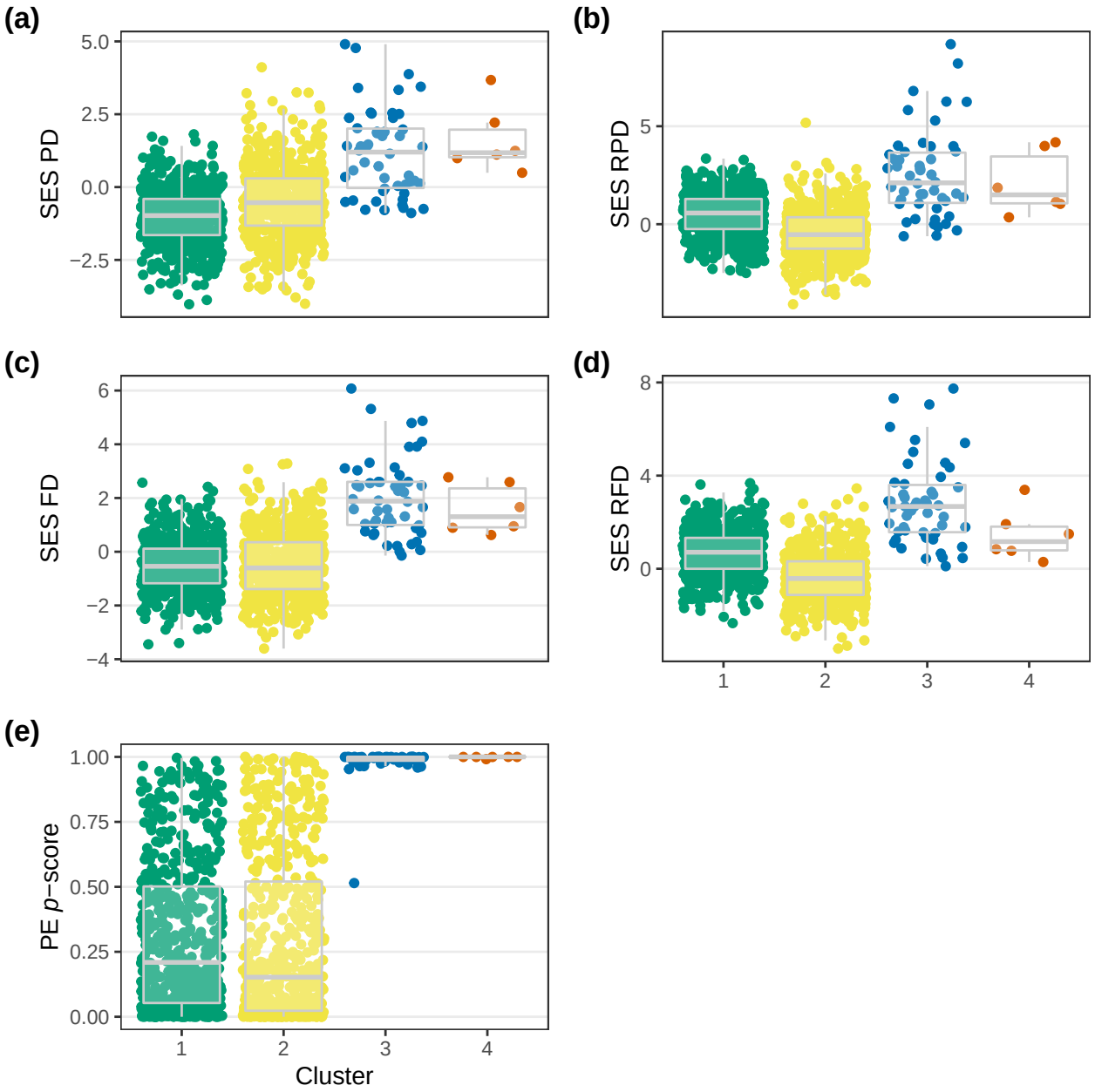

Fig. 8. Phylogenetic and morphological diversity of the ferns of Japan by taxonomic bioregion. (a), Standard effect size (SES) of phylogenetic diversity (PD). (b), SES of relative PD (RPD). (c), SES of functional diversity (FD). (d), SES of relative FD (RFD). (e), *p*-score of observed phylogenetic endemism (PE) relative to 999 random communities. Only top four clusters with the most sites shown.

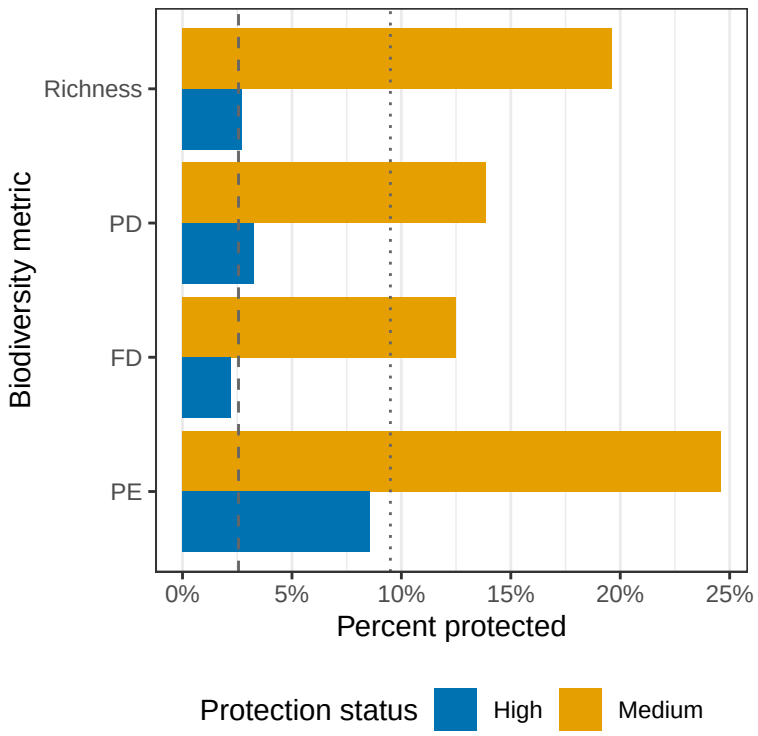

Fig. 9. Percent of land area with protected status for grid cells with significantly high biodiversity by at least one metric. Biodiversity metrics include phylogenetic diversity (PD), functional diversity (FD), and phylogenetic endemism (PE). Significance of biodiversity metrics assessed by a one-tailed test comparing observed values to a null distribution of 999 random values (see Materials and Methods). Vertical lines indicate total percent area protected across Japan regardless of degree of biodiversity; dashed, high protection level (2.6%); dotted, medium protection level (9.5%).

Data availability statement

Code to replicate all analyses, figures, and this manuscript are available at https://github.com/joelnitta/japan_ferns_biogeo. A Docker image to run the code is available at https://hub.docker.com/repository/docker/joelnitta/japan_ferns_biogeo. Data files, including the DNA alignment, phylogenetic tree, and community data matrix, are openly available from Dryad at http://doi.org/ADD_DOI_HERE_WHEN_ASSIGNED.

References

Barrington, D. S. (1993). Ecological and historical factors in fern biogeography. *Journal of Biogeography*, *20*(3), 275–279.

Barrington, D. S., Haufler, C. H., & Werth, C. R. (1989). Hybridization, reticulation, and species concepts in the ferns. *American Fern Journal*, *79*, 55–64.

Baselga, A. (2012). The relationship between species replacement, dissimilarity derived from nestedness, and nestedness: Species replacement and nestedness. *Global Ecology and Biogeography*, *21*(12), 1223–1232. https://doi.org/10/gddc72

Bengtsson, B. O. (2003). Genetic variation in organisms with sexual and asexual reproduction. *Journal of Evolutionary Biology*, *16*(2), 189–199. https://doi.org/10/bf52w9

Bickford, S. A., & Laffan, S. W. (2006). Multi-extent analysis of the relationship between pteridophyte species richness and climate. *Global Ecology and Biogeography*, *0*(0), 588–601. https://doi.org/10.1111/j.1466-822X.2006.00250.x

Bivand, R. S., Pebesma, E., & Gómez-Rubio, V. (2013). Applied spatial data analysis with R: Second edition. In *Applied spatial data analysis with r: Second edition*. New York: Springer. https://doi.org/10.1007/978-1-4614-7618-4

Blomberg, S. P., Garland Jr., T., & Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: Behavioral traits are more labile. *Evolution*, *57*(4), 717–745. https://doi.org/10.1111/j.0014-3820.2003.tb00285.x

Brooks, T. M., Mittermeier, R. A., Fonseca, G. A. B. da, Gerlach, J., Hoffmann, M., Lamoreux, J. F., … Rodrigues, A. S. L. (2006). Global Biodiversity Conservation Priorities. *Science*, *313*(5783), 58–61. https://doi.org/10/b84nbv

Bryant, J., Lamanna, C., Morlon, H., Kerkhoff, A., Enquist, B., & Green, J. (2008). Microbes on mountainsides: contrasting elevational patterns of bacterial and plant diversity. *Proceedings of the National Academy of Sciences*, *105*(Supplement 1), 11505–11511.

Colwell, R. K., Rahbek, C., & Gotelli, N. J. (2004). The mid-domain effect and species richness patterns: What have we learned so far? *The American Naturalist*, *163*(3), E1–E23.

Courtiol, A., & Rousset, F. (2017). Modelling isoscapes using mixed models. *bioRxiv*, 207662. https://doi.org/10/gj769c

Creese, C., Lee, A., & Sack, L. (2011). Drivers of morphological diversity and distribution in the Hawaiian fern flora: Trait associations with size, growth form, and environment. *American Journal of Botany*, *98*(6), 956–966.

Daru, B. H., Farooq, H., Antonelli, A., & Faurby, S. (2020a). Endemism patterns are scale dependent. *Nature Communications*, *11*(1), 2115. https://doi.org/10/ggvnhf

Daru, B. H., Karunarathne, P., & Schliep, K. (2020b). phyloregion: R package for biogeographical regionalization and macroecology. *Methods in Ecology and Evolution*, *11*(11), 1483–1491. https://doi.org/10.1111/2041-210x.13478

Ebihara, A. (2016). *The Standard of Ferns and Lycophytes in Japan, Vol. 1*. Tokyo: Gakken Plus.

Ebihara, A. (2017). *The Standard of Ferns and Lycophytes in Japan, Vol. 2*. Tokyo: Gakken Plus.

Ebihara, A., & Nitta, J. H. (2019). An update and reassessment of fern and lycophyte diversity data in the Japanese Archipelago. *Journal of Plant Research*, *132*(6), 723–738. https://doi.org/10.1007/s10265-019-01137-3

Ebihara, A., Nitta, J. H., Matsumoto, Y., Fukazawa, Y., Kurihara, M., Yokote, H., … Azakami, O. (2019). Growth dynamics of independent gametophytes of *Pleurosoriopsis makinoi* (Polypodiaceae). *Bulletin of the National Museum of Nature and Science, Series B (Botany), Tokyo*, *45*(2), 77–86.

Ebihara, A., Yamaoka, A., Mizukami, N., Sakoda, A., Nitta, J. H., & Imaichi, R. (2013). A survey of the fern gametophyte flora of Japan: Frequent independent occurrences of noncordiform

gametophytes. *American Journal of Botany*, *100*(4), 735–743. https://doi.org/10.3732/ajb.1200555

Fahrig, L. (2013). Rethinking patch size and isolation effects: the habitat amount hypothesis. *Journal of Biogeography*, *40*(9), 1649–1663. https://doi.org/10/ggk3qt

Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biological Conservation*, *61*(1), 1–10. https://doi.org/10.1016/0006-3207(92)91201-3

Fritz, S. A., & Purvis, A. (2010). Selectivity in mammalian extinction risk and threat types: A new measure of phylogenetic signal strength in binary traits. *Conservation Biology*, *24*(4), 1042–1051. https://doi.org/10.1111/j.1523-1739.2010.01455.x

Fukaya, K., Kusumoto, B., Shiono, T., Fujinuma, J., & Kubota, Y. (2020). Integrating multiple sources of ecological data to unveil macroscale species abundance. *Nature Communications*, *11*(1), 1695. https://doi.org/10/ggztv3

Garcillán, P. P., Ezcurra, E., & Riemann, H. (2003). Distribution and species richness of woody dryland legumes in Baja California, Mexico. *Journal of Vegetation Science*, *14*(4), 475–486. https://doi.org/10/bgh9z9

Givnish, T. J. (1987). Comparative studies of leaf form: assessing the relative roles of selective pressures and phylogenetic constraints. *New Phytologist*, 131–160.

González-Orozco, C. E., Thornhill, A. H., Knerr, N., Laffan, S., & Miller, J. T. (2014). Biogeographical regions and phytogeography of the eucalypts. *Diversity and Distributions*, *20*(1), 46–58. https://doi.org/10/f5mddz

Gotelli, N. J. (2000). Null model analysis of species co-occurrence patterns. *Ecology*, *81*(9), 2606–2621. https://doi.org/10.2307/177478

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, *27*(4), 857–871. https://doi.org/10.2307/2528823

Grusz, A. L., Windham, M. D., Picard, K. T., Pryer, K. M., Schuettpelz, E., & Haufler, C. H. (2021). A drought-driven model for the evolution of obligate apomixis in ferns: evidence from pellaeids (Pteridaceae). *American Journal of Botany*, *108*(2), 263–283. https://doi.org/10/gjc84t

Hijmans, R. J. (2020). Raster: Geographic data analysis and modeling. Retrieved from https://CRAN.R-project.org/package=raster

Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., … Webb, C. O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, *26*(11), 1463–1464. https://doi.org/10.1093/bioinformatics/btq166

Kessler, M., Kluge, J., Hemp, A., & Ohlemüller, R. (2011). A global comparative analysis of elevational species richness patterns of ferns. *Global Ecology and Biogeography*, *20*, 868–880. https://doi.org/10.1111/j.1466-8238.2011.00653.x

Kluge, J., & Kessler, M. (2007). Morphological characteristics of fern assemblages along an elevational gradient: Patterns and causes. *Ecotropica.*, *13*(1), 27–43.

Krueger-Hadfield, S. A., Blakeslee, A. M. H., & Fowler, A. E. (2019). Incorporating ploidy diversity into ecological and community genetics. *Journal of Phycology*, *55*(6), 1198–1207. https://doi.org/10/gjppkc

Kubota, Y., Hirao, T., Fujii, S. J., Shiono, T., & Kusumoto, B. (2014). Beta diversity of woody plants in the Japanese archipelago: The roles of geohistorical and ecological processes. *Journal of Biogeography*, *41*, 1267–1276. https://doi.org/10.1111/jbi.12290

Kusumoto, B., Shiono, T., Konoshima, M., Yoshimoto, A., Tanaka, T., & Kubota, Y. (2017). How well are biodiversity drivers reflected in protected areas? A representativeness assessment of the geohistorical gradients that shaped endemic flora in Japan. *Ecological Research*, *32*(3), 299–311. https://doi.org/10/f96h8f

Laffan, S. W., Rosauer, D. F., Virgilio, G. D., Miller, J. T., Knerr, N., Thornhill, A. H., … Gonz, C. E. (2016). *Range-weighted metrics of species and phylogenetic turnover can better resolve biogeographic transition zones*. 580–588. https://doi.org/10.1111/2041-210X.12513

Laliberté, E., & Legendre, P. (2010). A distance-based framework for measuring functional diversity from multiple traits. *Ecology*, *91*(1), 299–305.

Landau, W. M. (2021). The targets R package: a dynamic Make-like function-oriented pipeline toolkit for reproducibility and high-performance computing. *Journal of Open Source Software*, *6*(57), 2959. https://doi.org/10.21105/joss.02959

Lehtonen, S. (2011). Towards resolving the complete fern tree of life. *PLoS ONE*, *6*(10), e24851. https://doi.org/https://doi.org/10.1371/journal.pone.0024851

Link-Pérez, M. A., & Laffan, S. W. (2018). Fern and lycophyte diversity in the Pacific Northwest: Patterns and predictors. *Journal of Systematics and Evolution*. https://doi.org/10.1111/jse.12430

Liu, H.-M., Dyer, R., Guo, Z., Meng, Z., Li, J., & Schneider, H. (2012). The evolutionary dynamics of apomixis in ferns: a case study from polystichoid ferns. *Journal of Botany*, *2012*(2), 1–11.

Loiseau, O., Weigand, A., Noben, S., Rolland, J., Silvestro, D., Kessler, M., … Salamin, N. (2020). Slowly but surely: Gradual diversification and phenotypic evolution in the hyperdiverse tree fern family Cyatheaceae. *Annals of Botany*, *125*(1), 93–103. https://doi.org/10.1093/aob/mcz145

Margules, C. R., & Pressey, R. L. (2000). Systematic conservation planning. *Nature*, *405*(6783), 243–253. https://doi.org/10/dn4xdx

Millien-Parra, V., & Jaeger, J.-J. (1999). Island biogeography of the Japanese terrestrial mammal assemblages: an example of a relict fauna. *Journal of Biogeography*, *26*(5), 959–972. https://doi.org/10/dqc4qp

Mishler, B. D., Guralnick, R., Soltis, P. S., Smith, S. A., Soltis, D. E., Barve, N., … Laffan, S. W. (2020). Spatial phylogenetics of the North American flora. *Journal of Systematics and Evolution*, jse.12590. https://doi.org/10/gg3h86

Mishler, B. D., Knerr, N., González-Orozco, C. E., Thornhill, A. H., Laffan, S. W., & Miller, J. T. (2014). Phylogenetic measures of biodiversity and neo- and paleo-endemism in Australian Acacia. *Nature Communications*, *5*, 4473. https://doi.org/10.1038/ncomms5473

Moran, R., & Smith, A. R. (2001). Phytogeographic relationships between neotropical and African-Madagascan pteridophytes. *Brittonia*, *53*(2), 304–351.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, *403*(6772), 853–858. https://doi.org/10.1038/35002501

Nagalingum, N. S., Knerr, N., Laffan, S. W., González-Orozco, C. E., Thornhill, A. H., Miller, J. T., & Mishler, B. D. (2015). Continental scale patterns and predictors of fern richness and phylogenetic diversity. *Frontiers in Genetics*, *6*(April), 132. https://doi.org/10.3389/fgene.2015.00132

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Nicotra, A. B., Leigh, A., Boyce, C. K., Jones, C. S., Niklas, K. J., Royer, D. L., & Tsukaya, H. (2011). The evolution and functional significance of leaf shape in the angiosperms. *Functional Plant Biology*, *38*(7), 535. https://doi.org/10/dt32jd

Nishizawa, K., Tatsumi, S., Kitagawa, R., & Mori, A. S. (2016). Deer herbivory affects the functional diversity of forest floor plants via changes in competition-mediated assembly rules. *Ecological Research*, *31*(4), 569–578. https://doi.org/10/f8t3zh

Nitta, J. H. (2021). canaper (Version 0.0.0.9000). Zenodo. https://doi.org/10.5281/ZENODO.5094033

Nitta, J. H., Meyer, J.-Y., Taputuarai, R., & Davis, C. C. (2017). Life cycle matters: DNA barcoding reveals contrasting community structure between fern sporophytes and gametophytes. *Ecological Monographs*, *87*(2), 278–296. https://doi.org/10.1002/ecm.1246

Noss, R. (2016). Announcing the world's 36th biodiversity hotspot: The North American coastal plain. Retrieved August 26, 2021, from https://www.cepf.net/node/4422

Orme, D., Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S., Isaac, N., & Pearse, W. (2018). Caper: Comparative analyses of phylogenetics and evolution in r. Version 1.0.1. Retrieved from https://CRAN.R-project.org/package=caper

Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*, *401*(6756), 877–884. https://doi.org/10.1038/44766

Pebesma, E. (2018). Simple features for R: Standardized support for spatial vector data. *The R Journal*, *10*(1), 439. https://doi.org/10/gf2ztt

Petchey, O. L., & Gaston, K. J. (2002). Functional diversity (FD), species richness and community composition. *Ecology Letters*, *5*(3), 402–411. https://doi.org/10/frgvzn

Pinson, J. B., Chambers, S. M., Nitta, J. H., Kuo, L.-Y., & Sessa, E. B. (2017). The separation of generations: Biology and biogeography of long-lived sporophyteless fern gametophytes. *International Journal of Plant Sciences*, *178*(1), 1–18. https://doi.org/10.1086/688773

Poppinga, S., Haushahn, T., Warnke, M., Masselter, T., & Speck, T. (2015). Sporangium exposure and spore release in the peruvian maidenhair fern (Adiantum peruvianum, Pteridaceae). *PLoS ONE*, *10*(10), 1–20. https://doi.org/10.1371/journal.pone.0138495

Pteridophyte Phylogeny Group I. (2016). A community-derived classification for extant lycophytes and ferns. *Journal of Systematics and Evolution*, *54*(6), 563–603. https://doi.org/10.1111/jse.12229

Qian, H., Kessler, M., Deng, T., & Jin, Y. (2021). Patterns and drivers of phylogenetic structure of pteridophytes in China. *Global Ecology and Biogeography*, geb.13349. https://doi.org/10/gk6mzf

Qian, H., Wang, S. L., Li, Y. L., Xiao, M., & Wang, X. L. (2012). Disentangling the relative effects of ambient energy, water availability, and energy-water balance on pteridophyte species richness at a landscape scale in China. *Plant Ecology*, *213*(5), 749–756. https://doi.org/DOI%2010.1007/s11258-012-0038-0

R Core Team. (2020). R: A language and environment for statistical computing. Version 4.0.2. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, *3*(2), 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x

Rosauer, D. F., Laffan, S. W., Crisp, M., Donnellan, S., & Cook, L. (2009). Phylogenetic endemism: a new approach for identifying geographical concentrations of evolutionary history. *Molecular Ecology*, *18*(19), 4061–4072.

Rousset, F., & Ferdy, J. B. (2014). Testing environmental and genetic effects in the presence of spatial autocorrelation. *Ecography*, *37*(8), 781–790. https://doi.org/10.1111/ecog.00566

Schuettpelz, E., & Pryer, K. M. (2007). Fern phylogeny inferred from 400 leptosporangiate species and three plastid genes. *Taxon*, *56*(4), 1037–1050. https://doi.org/10.2307/25065903

Shimizu, Y. (2014). Process of the formation of Japanese forest and typification of vegetation zone : From an East Asian viewpoint. *Chiikigaku-Kenkyu (in Japanese)*, (27), 19–75.

Shiono, T., Kusumoto, B., Maeshiro, R., Fujii, S. J., Götzenberger, L., de Bello, F., & Kubota, Y. (2015). Climatic drivers of trait assembly in woody plants in Japan. *Journal of Biogeography*, *42*(6), 1176–1186. https://doi.org/10.1111/jbi.12503

Smith, S. A., & O'Meara, B. C. (2012). treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics*, *28*(20), 2689–2690. https://doi.org/10.1093/bioinformatics/bts492

Takatsuki, S. (2009). Effects of sika deer on vegetation in Japan: A review. *Biological Conservation*, *142*(9), 1922–1929. https://doi.org/10/dh3bxn

Tanaka, T., Isaka, Y., Hattori, M., & Sato, T. (2014). Ecological and phylogenetic approaches for diversification of apogamous ferns in Japan. *Plant Systematics and Evolution*, *300*(9), 2041–2050. https://doi.org/10.1007/s00606-014-1036-6

Testo, W. L., & Sundue, M. A. (2016). A 4000-species dataset provides new insight into the evolution of ferns. *Molecular Phylogenetics and Evolution*, *105*, 200–211. https://doi.org/10.1016/j.ympev.2016.09.003

Tokita, K. (2006). Shizen kouen ni okeru shika mondai (The problem of deer in natural parks). In Japanese. In T. Yumoto & H. Matsuda (Eds.), *Sekai-isan wo shika ga kuu: shika to mori no seitaigaku (Deer eat world heritage sites: Ecology of deer and forests). In Japanese.* (pp. 20–37). Tokyo: Bunichi Sogo Shuppan. Retrieved from https://ci.nii.ac.jp/naid/10029109218

Tryon, R. (1970). Development and evolution of fern floras of oceanic islands. *Biotropica*, *2*(2), 76–84.

Vaida, F., & Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika*, *92*(2), 351–370. https://doi.org/10/cstbq8

Vallejos, R., Osorio, F., & Bevilacqua, M. (2020). *Spatial relationships between two georeferenced variables: With applications in r*. New York: Springer. Retrieved from http://srb2gv.mat.utfsm.cl/

Watkins Jr., J. E., Holbrook, N. M., & Zwieniecki, M. (2010). Hydraulic properties of fern sporophytes: Consequences for ecological and evolutionary diversification. *American Journal of Botany*, *97*(12), 2007–2019.

White, A. E., Dey, K. K., Mohan, D., Stephens, M., & Price, T. D. (2019). Regional influences on community structure across the tropical-temperate divide. *Nature Communications*, *10*(1), 2646. https://doi.org/10.1038/s41467-019-10253-6

Biosketch

Joel H. Nitta's research focuses on patterns and processes of diversification in ferns. For more information, see https://joelnitta.com.

Author contributions: J.H.N., B.M., and A.E. conceived the ideas; A.E. provided the data; J.H.N. analyzed the data; W.I provided resources; J.H.N. wrote the manuscript with input from all co-authors.