



The likely extinction of hundreds of palm species threatens their contributions to people and ecosystems

S. Bellot¹✉, Y. Lu¹, A. Antonelli^{1,2,3}, W. J. Baker¹, J. Dransfield¹, F. Forest¹, W. D. Kissling⁴, I. J. Leitch¹, E. Nic Lughadha¹, I. Ondo¹, S. Pironon¹, B. E. Walker¹, R. Cámara-Leret^{5,6}✉ and S. P. Bachman^{1,6}

Protecting nature's contributions to people requires accelerating extinction risk assessment and better integrating evolutionary, functional and used diversity with conservation planning. Here, we report machine learning extinction risk predictions for 1,381 palm species (Arecaceae), a plant family of high socio-economic and ecological importance. We integrate these predictions with published assessments for 508 species (covering 75% of all palm species) and we identify top-priority regions for palm conservation on the basis of their proportion of threatened evolutionarily distinct, functionally distinct and used species. Finally, we explore palm use resilience to identify non-threatened species that could potentially serve as substitutes for threatened used species by providing similar products. We estimate that over a thousand palms (56%) are probably threatened, including 185 species with documented uses. Some regions (New Guinea, Vanuatu and Vietnam) emerge as top ten priorities for conservation only after incorporating machine learning extinction risk predictions. Potential substitutes are identified for 91% of the threatened used species and regional use resilience increases with total palm richness. However, 16 threatened used species lack potential substitutes and 30 regions lack substitutes for at least one of their threatened used palm species. Overall, we show that hundreds of species of this keystone family face extinction, some of them probably irreplaceable, at least locally. This highlights the need for urgent actions to avoid major repercussions on palm-associated ecosystem processes and human livelihoods in the coming decades.

Vascular plants contribute countless services to people^{1–5}, by sustaining ecosystem functioning and through direct use for food, medicine, construction, utensils and cultural activities. Many of these contributions are threatened by global anthropogenic changes⁶. To prevent further loss of plant species and ecosystem services, we urgently need to (1) identify plant species at risk and (2) define spatial conservation priorities that capture multiple aspects of plant diversity and its contributions to humanity.

Threatened plant species are best identified using expert-validated extinction risk assessments following the methodology of the International Union for the Conservation of Nature (IUCN) Red List of Threatened Species (hereafter, the Red List). Red List assessments are a gold standard and, as such, a cornerstone of conservation planning^{7–9}. However, the Red List assessment process is time-consuming and lack of resources has so far resulted in gaps and biases in the taxonomic and spatial coverage of the Red List. For instance, only 14% of vascular plant species have been assessed (versus 80% of vertebrates¹⁰), with biases towards trees and North American species^{11–13}. A complementary and more extensive resource (ThreatSearch¹⁴) integrates Red List assessments with other kinds of published extinction risk assessments¹⁵. However, this still only covers ~30% of vascular plant species¹². Accurate and rapid extinction risk assessment approaches are therefore urgently needed to inform global conservation efforts and to facilitate regular re-assessments^{12,16}.

Two approaches for accelerating extinction risk assessments have been noted: ‘criteria explicit’ and ‘category predictive’¹⁷. The former is the automated generation of preliminary assessments based on subsets of Red List criteria^{18,19}. Such an approach has the advantage of directly implementing Red List criteria but can over-predict risk in some cases¹⁹. In the category predictive approach, extinction risk is predicted on the basis of statistical associations found between Red List categories and other variables (for example, species distribution range size), therefore only implicitly relying on the Red List criteria. This approach often uses machine learning (ML)^{20–23} and its performance is sensitive to model parametrization and dataset properties, which create uncertainty around the predictions²⁴. However, ML has been shown to perform well on a range of plant taxa and geographic scales^{21,23,25} and prediction uncertainty can be addressed by performing sensitivity analyses and by comparing predictions generated by models with different strengths and weaknesses²⁶. We hereafter refer to Red List and ThreatSearch assessments as ‘published extinction risk assessments’, to make the distinction between them and ‘extinction risk predictions’ obtained using ML approaches.

To define conservation priorities, published extinction risk assessments or predictions can be combined with measures of species evolutionary distinctiveness (how much a species contributes to the phylogenetic diversity of its clade²⁷) and functional distinctiveness (how functionally distant a species is from other species, on average²⁸). Combining multiple aspects of biodiversity into a

¹Royal Botanic Gardens, Kew, Richmond, UK. ²Department of Biological and Environmental Sciences, Gothenburg Global Biodiversity Centre, Gothenburg, Sweden. ³Department of Plant Sciences, University of Oxford, Oxford, UK. ⁴Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, Amsterdam, the Netherlands. ⁵Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland. ⁶These authors jointly supervised this work: R. Cámara-Leret, S. P. Bachman. ✉e-mail: s.bellot@kew.org; rodrigo.camaraleret@ieu.uzh.ch

single prioritization criterion has the potential to increase conservation success by capturing more biodiversity features and associated contributions to people^{29,30}. So far, however, studies combining taxonomic, functional and phylogenetic diversity have been geographically restricted^{31,32} and/or focused on vertebrates^{33–36}. As a result, there is an urgent need to extend such studies to vascular plants given their myriad contributions to people and their importance for progress towards the United Nations 2030 Sustainable Development Goals (<https://sdgs.un.org/2030agenda>) and post-2020 targets of the Global Biodiversity Framework of the Convention on Biological Diversity (CBD) (<https://www.cbd.int/conferences/post2020>).

Integrative plant conservation approaches can also benefit from the inclusion of information on plant uses (ethnobotanical knowledge). This is especially relevant because it remains unclear how well plant uses and other aspects of nature's contributions to people are captured by measures of phylogenetic and functional diversity³⁷. Plant uses can be underpinned by genetic and functional traits, resulting in correlations between uses and certain features or phylogenetic placements^{38–41}. However, plant use is also driven by plant availability and trait-independent cultural practices³⁷, thus these correlations are not always apparent⁴². In this context, integrating plant use data into plant conservation studies is needed to evaluate risks associated with these culturally important uses. Finding a relationship between uses and genetic and functional traits may in turn help to identify potential substitutes for threatened used species because phylogenetically and/or functionally closer species are likely to be better substitutes than randomly chosen species⁴¹. So far, this idea is poorly explored and plant use data remain largely untapped in large-scale conservation studies (but see recent studies focusing on New Guinea and Brazil^{43,44}), hindering the global conservation of plants that support human livelihoods.

This study integrates phylogenetic, functional and ethnobotanical information to provide a global conservation survey of palms (Arecaceae), a keystone plant family. Palms are among the most economically important plants in the world, with hundreds of wild species providing essential contributions to millions of people (for example, food, medicine and construction)^{37,45} and even supporting large-scale industries (for example, rattan products⁴⁶). They are important components of vulnerable and biodiverse ecosystems such as tropical rainforests⁴⁷. Their unique functional traits provide shelter and food for animals^{48,49} and deliver essential contributions to ecosystems^{50–52}. Palm phylogenetics, functional traits and uses are well-documented compared to many other tropical plant families^{49,53–56} but they have never been analysed together from a global conservation perspective. Currently, extinction risk assessments are published on the Red List for 797 (31%)⁵⁷ of the family's ~2,500 species^{58,59}, with palms from Africa and especially Madagascar being particularly well represented^{60,61}. This number increases to 61% when including global assessments published on ThreatSearch¹⁴. However, it decreases to 23% (Red List) and 34.5% (ThreatSearch) when only including assessments published in the last decade (hereafter 'recent'), an age above which assessments are considered outdated by the Red List¹⁶ (<https://www.iucnredlist.org/assessment/process> (accessed 16 September 2021)). The last global assessment of palms was published 25 years ago, falling well outside this time window⁶² and a recent global analysis of vascular plant Red List assessments concluded that extinction risk may be overestimated in the palm family due to representation biases (for example, in favour of threatened species and/or species from Madagascar)¹².

Here, we integrate multiple aspects of palm diversity to advance conservation planning and research by (1) quantifying levels of extinction risk among evolutionarily distinct, functionally distinct and used palm species, (2) identifying global priority regions for palm conservation that capture these three aspects and (3) exploring the resilience of palm uses. To quantify risk, we draw on

recently published extinction risk assessments of palms to train, test and evaluate 48 ML models covering a broad spectrum of strategies to address dataset gaps and imbalances (Extended Data Fig. 1, Supplementary Tables 1 and 2, Supplementary Note and Supplementary Fig. 1). We then predict the extinction risk of 1,381 palm species using the model with the highest balanced accuracy (82%; Supplementary Table 1) and we combine these predictions with 508 recently published extinction risk assessments (≤ 10 years old). This 'total evidence' approach allows us to provide a global extinction risk overview for 75% of the world's palm species and to identify priority regions for palm conservation on the basis of levels of risk among evolutionarily distinct, functionally distinct and used palm species. Finally, we develop a new approach based on phylogenetic, functional and ecological information to estimate to what degree threatened used species may be substituted by co-occurring non-threatened species for providing similar services or products (Methods). Our work has practical implications for the conservation of the economically and ecologically important palm family and provides new understanding of how integrative approaches can enhance sustainable biodiversity use.

Results

Palm diversity and uses at risk. We found that over half of the world's palm species may be threatened and that scenarios incorporating ML estimates differ from extrapolations based solely on published assessments (Fig. 1a). Our most accurate model classified 703 palm species as threatened (Supplementary Table 3). Adding these predictions to the 353 species assessed as threatened in recently published assessments, we found that 56% of the 1,889 palm species included in the total evidence approach were threatened. In contrast, 69–80% of palm species with published assessments appear to be threatened, depending on if including only recent or all published assessments (Fig. 1a and Supplementary Table 3).

We then considered threat levels among evolutionarily distinct, functionally distinct and used palms. The last included palms with at least one recorded use among ten use categories but we also separately focused on four main use categories: 'culture', 'food', 'medicine' and 'utensils, tools and construction' (Methods). According to the total evidence approach, 455 (48%) of the evolutionarily distinct, 447 (47.5%) of the functionally distinct and 185 (29%) of the used species were threatened. Percentages of threatened species were higher among species used for food and utensils, tools and construction than for culture or medicine (Fig. 1a and Supplementary Table 3). In all categories, total evidence estimates were lower than estimates based only on published assessments, especially when the latter included old assessments. The greatest difference was observed for used American species (13% versus 41–72%; Extended Data Fig. 2a and Supplementary Table 3).

Globally, palms occur in 227 out of 369 level-3 botanical countries (hereafter 'regions'; Methods). Applying the total evidence approach, the percentage of threatened species per region ranged from zero to 100%, with mean values of 21% and 9% for 'species-rich' (> 10 species) and 'species-poor' (≤ 10 species) regions, respectively (Fig. 1b and Supplementary Table 4). The percentage of threatened species per region correlated positively with the number of species in species-rich regions (Pearson correlation coefficient $r = 0.37$; $P = 0.00022$) but not in species-poor regions (Pearson's $r = 0.06$; $P = 0.5$; Extended Data Fig. 2b). The percentages of threatened species per region estimated using the total evidence approach were usually lower than those estimated only from recent published assessments, except for 11 species-poor and 29 species-rich regions (the latter including Sri Lanka, Vietnam, Sulawesi, New Guinea, Paraguay, Cuba, Brazil North, Mauritius, Democratic Republic of the Congo and Gabon), where the percentages of threatened species were 4 to 33 points higher with the total evidence approach (see red areas on map in Fig. 1c and Supplementary Table 4).

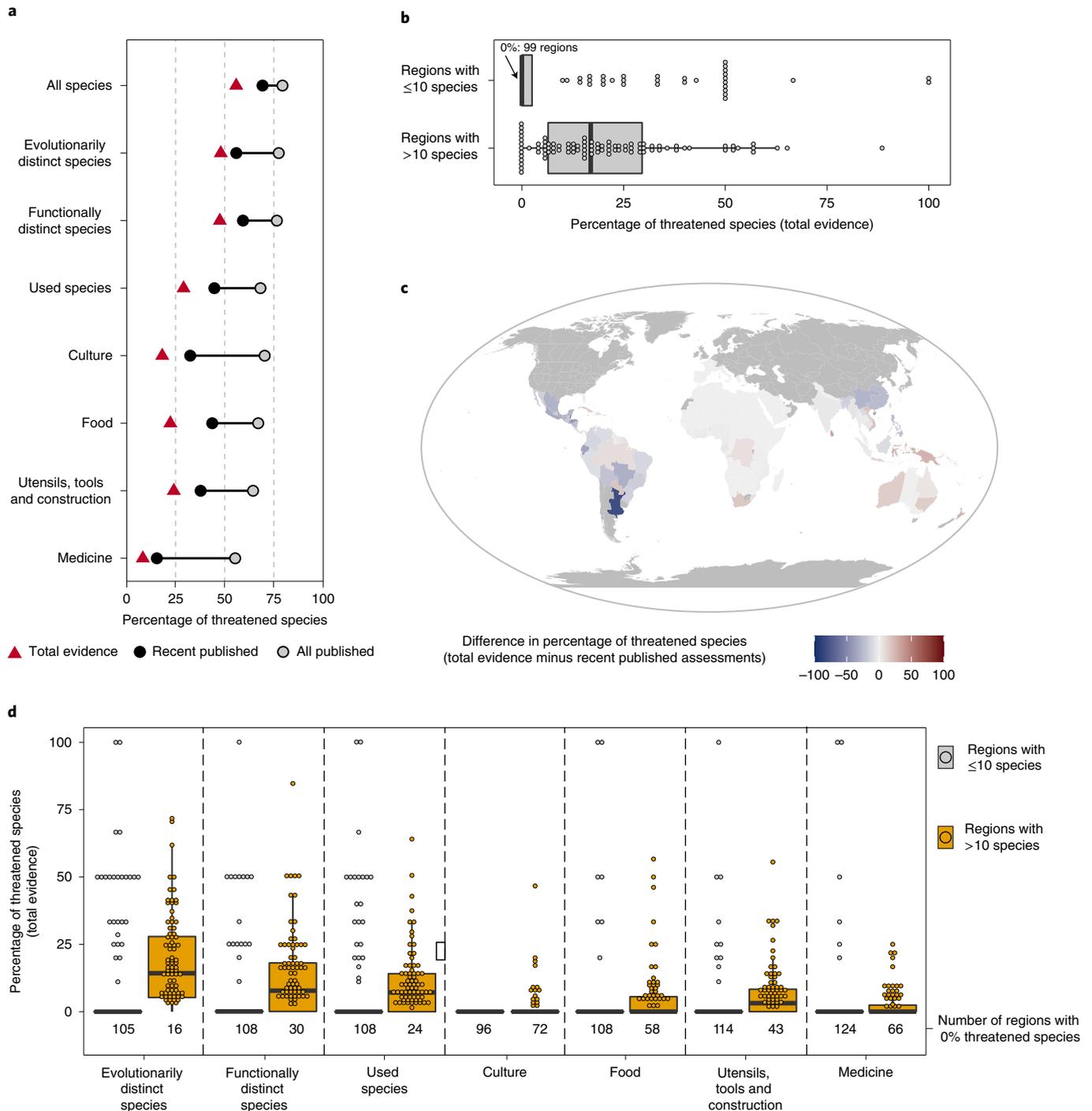


Fig. 1 | Percentages of threatened palm species in different diversity and use categories. a, Comparison between global estimates obtained with or without ML predictions. Triangles, estimates obtained by combining recent published assessments and ML predictions from the most accurate model (total evidence approach). Circles, extrapolations based on published assessments only: grey, all published assessments; black, only recent (that is ≤ 10 years old) published assessments. **b**, Regional percentages of threatened species according to the total evidence approach. **c**, Difference between regional percentages of threatened species obtained with and without ML predictions. **d**, Regional variation in percentages of threatened species, for each diversity and use category, under the total evidence approach. For all boxplots in **b** and **d**, each dot represents one region with at least one palm species of the category considered, the bold line represents the median value, the box spans values from the first to the third quartile and the lines outside the box extend until the smallest and largest values, no further than 1.5 times the distance between the first and third quartiles. To help visualization when many regions have 0% of threatened species, dots for these regions are not shown but instead the number of regions with 0% of threatened species is indicated.

The percentage of threatened palm species per region varied between the different measures of diversity and between use categories (Fig. 1d and Supplementary Table 4). When considering

species-rich regions only, the median regional percentage of threatened species among evolutionarily distinct species was higher than for functionally distinct and used species. Furthermore, median

regional percentages of threatened species in the utensils, tools and construction and food categories were higher than in the culture and medicine categories (Fig. 1d). There was no such pattern in species-poor regions, as most had no threatened species, regardless of the diversity measure or use category considered (Fig. 1d). Despite these variations, regional percentages of threatened species among evolutionarily distinct, functionally distinct and used species were significantly positively correlated (Pearson's $r=0.58-0.84$ depending on the categories tested; $P < 2.2 \times 10^{-16}$; Extended Data Fig. 2b), reflecting the fact that 80% of threatened used species were evolutionarily and/or functionally distinct (Supplementary Table 5). However, variations between categories were still sufficient to result in different region rankings based on the percentage of threatened species in each category (Extended Data Fig. 3).

Priority regions for palm conservation and research. Regional variation in extinction risk among evolutionarily distinct, functionally distinct and used species suggested that basing conservation prioritization on a single diversity measure or use category may miss risks associated with other categories. To identify priority regions for palm conservation while accounting for this variation, we therefore scored regions on the basis of their proportion of threatened species among evolutionarily distinct and/or functionally distinct and/or used species (hereafter referred to as 'species of interest'; Fig. 2a). Under the total evidence approach, there were 25 regions where $\geq 40\%$ of species of interest were threatened, including ten species-rich regions: Madagascar (199 palm species in total), New Guinea (280 species), the Philippines (130 species), Hawaii (34 species), Borneo (291 species), Jamaica (12 species), Vietnam (113 species), Vanuatu (21 species), New Caledonia (43 species), Sulawesi (62 species; Fig. 2a,b and Supplementary Table 4). In addition to these 'top-priority' regions for conservation, 36 lower priority regions (including 25 species-rich regions) had 20–39% of their species of interest potentially threatened, while the remaining 164 regions (including 58 species-rich regions) had $< 20\%$ of their species of interest potentially threatened (Fig. 2a,b and Supplementary Table 4). All species-rich regions identified as top priorities with the total evidence approach were also top priorities when considering only recent and/or all published assessments (Supplementary Table 4). However, the ranking of the regions differed between approaches, with New Guinea, Vietnam and Vanuatu appearing among the top ten priorities only when applying the total evidence approach (Extended Data Fig. 3).

To explore if priority regions for conservation were also among the least well-studied regions, the number of species lacking extinction risk information under the total evidence approach was calculated for each region (Fig. 2b and Supplementary Table 4). This highlighted priorities for research that were not among the top priorities for conservation, namely Peninsular Malaysia, Cuba, Sumatra, Thailand and India, which all had ≥ 20 species without extinction risk information (Fig. 2b). There was a weak but significant positive correlation between the number of species lacking extinction risk information and the percentage of threatened species of interest (Pearson's $r=0.25$; $P=0.00014$), highlighting that some priority regions for conservation were also priorities for research. This was the case of New Guinea, Borneo, the Philippines and Sulawesi (Fig. 2b).

Potential alternatives to threatened used species. To explore the resilience of palm uses, we calculated the number of species that could serve as a potential alternative for each threatened used species. A species was considered a potential alternative if it was from the same region and biome as the threatened used species and if it was significantly phylogenetically and/or functionally close to it (Methods and Discussion). Importantly, this exploration of the replaceability of used species only addresses their direct use by

people, without necessarily capturing their other contributions or importance for ecosystem functioning. The global replaceability of threatened used species appeared high, with a median of 8 potential alternatives and 1–84 potential alternatives being identified for 91% of the species (Fig. 3a and Supplementary Table 5). However, 77 (42%) threatened used species had only up to 5 alternatives identified, including 16 (9%) that completely lacked alternatives (Fig. 3a). Species that could be substituted usually had alternatives in all their regions of occurrences, except for nine species (5%) that had alternatives in some but not all their regions of occurrence (Extended Data Figure 4a). Utensils, tools and construction had the highest number of threatened species for which no potential alternative could be found (11 species) but this category and culture appeared generally more resilient than food and medicine on the basis of median numbers of potential alternatives per threatened used species (Extended Data Figure 4b). The global potential of non-threatened species for serving as substitutes for threatened used species tended to be restricted, suggesting a lack of use redundancy among palm species. Indeed, median and maximal numbers of substituted species per non-threatened species were only 1 and 33, respectively, and 268 (32%) species could not qualify as potential substitute for any threatened used species (Fig. 3b). Species with the highest replaceability were among those with intermediate stem volumes and smallest fruits, mostly in subfamilies Ceroxyloideae and Calamoideae (dark red, Extended Data Figure 4c). As expected from an approach relying on phylogenetic and functional distances, species with high potential for serving as substitute tended to occupy the same area of the trait space as the most replaceable threatened used species and to cluster with them in the palm phylogeny (dark blue, Extended Data Figure 4c).

There were 92 regions with threatened used species and their median number of potential alternatives per threatened used species varied from 0 to 63 (Fig. 3c and Supplementary Table 4). There was a positive correlation between median number of alternatives and total number of palm species in the region (Pearson's $r=0.84$; $P=2.2 \times 10^{-16}$; Extended Data Figure 5). Consistently, regions with the lowest median numbers of alternatives were spread across the tropics and subtropics, while regions with the highest median numbers of alternatives were concentrated in palm-rich areas such as South-East Asia and North-East South America (Fig. 3c). In two-thirds of the regions, potential alternatives could be identified for all threatened used species, while 30 regions (including 18 species-rich regions) lacked alternatives for some or all of their threatened used species. (Fig. 3c, Extended Data Fig. 5 and Supplementary Table 4). Among species-rich regions identified as conservation priorities in the previous section, most had a low number of threatened used species (< 5) and low median numbers of alternatives (< 20 ; Fig. 3c and Supplementary Table 4). The Philippines and Madagascar stood out as high-risk and low-resilience priority regions with 41 and 25 threatened used species and median numbers of alternatives of only 10 and 2, respectively. In contrast, Borneo had only 12 threatened used species but the highest resilience, while New Guinea appeared as a high-risk and intermediate-resilience region, with 25 threatened used species and a median number of alternatives of 19 (Fig. 3c).

Discussion

Our multidimensional assessment of the extinction risk faced by palms and their contributions to people shows that (1) over 1,000 palm species may be threatened with extinction, (2) most priority regions for palm conservation and research are in South-East Asia and the Pacific and (3) alternatives (based on morphological similarity and relatedness) may be available for most threatened used species, albeit with regional variations in use resilience.

ML predictions allowed us to obtain extinction risk information for almost three times more species than when using only published

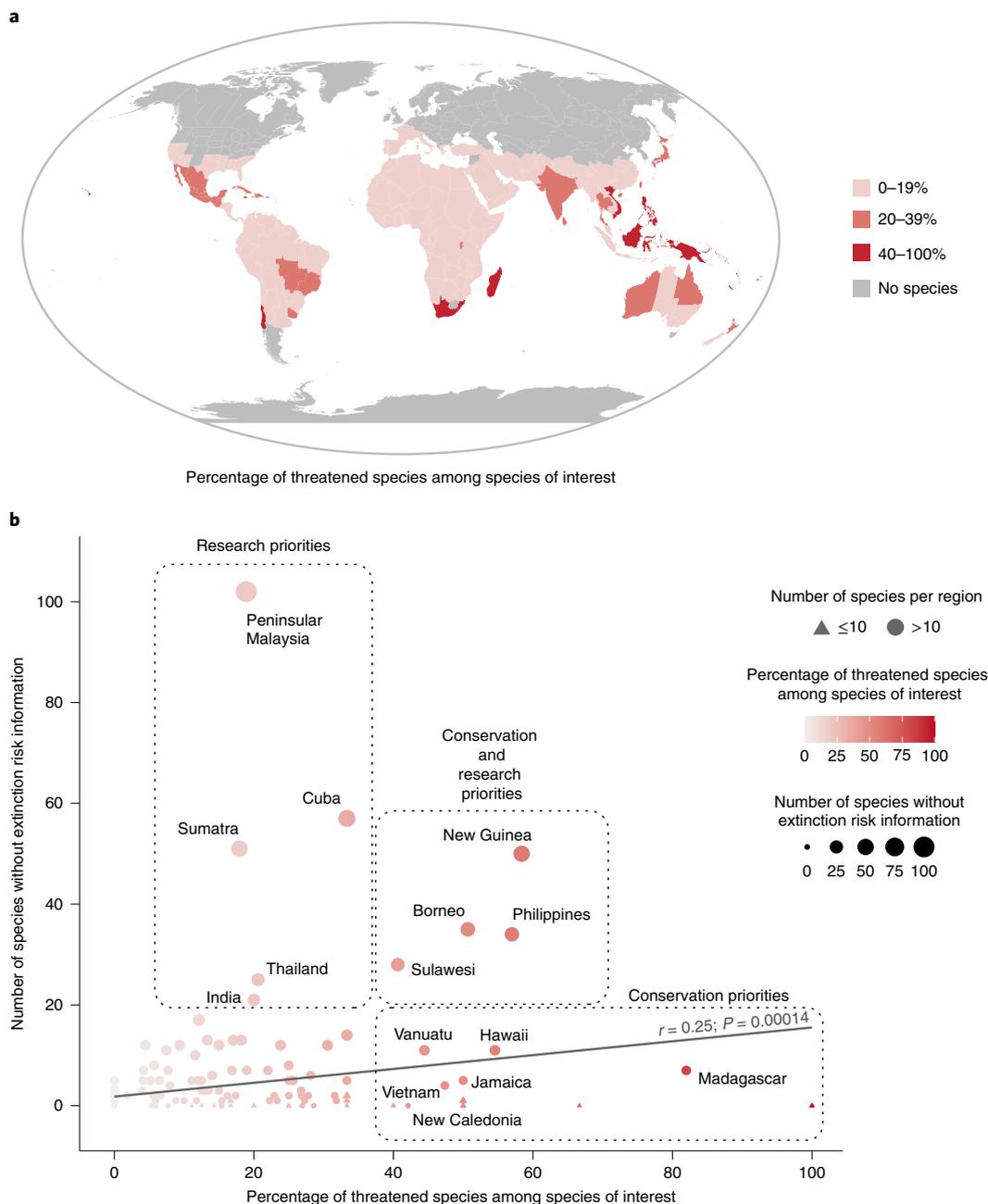


Fig. 2 | Priority regions for palm conservation and research. **a**, Regional percentages of threatened species among species of interest. Percentages are based on recent (that is ≤ 10 year) published assessments and ML predictions (total evidence approach). Species of interest are evolutionarily distinct and/or functionally distinct and/or used by humans. **b**, Relationship between regional lack of extinction risk information and level of extinction risk among species of interest. Regions identified here as conservation priorities have $\geq 40\%$ of their species of interest that are threatened, while research priorities have ≥ 20 species without extinction risk information under the total evidence approach. Only priority regions with >10 species are named (details on other regions can be found in Supplementary Table 4). All regions with at least one palm species of interest are represented ($n = 225$). The association between the variables was measured using Pearson's product moment correlation coefficient in a two-sided Pearson's correlation test. r is Pearson's correlation coefficient and P is the associated P value.

assessments. These ML predictions were essential to provide a global view of extinction risk that was less biased than when extrapolating from published assessments alone (Fig. 1a,c). Furthermore, ML allowed us to correct the bias of the Red List towards threatened species¹², which for palms was most pronounced in the Americas. The greater magnitude of this bias when considering assessments >10 years old (Fig. 1a and Extended Data Fig. 2a) probably results

from data accumulation and guidelines development over the years, which enable better manual assessments (and predictions) today than a decade ago⁵⁷.

Policy-relevant, our results and data contribute to accelerating Red List assessments by enabling us to (1) rapidly Red List as Least Concern⁶³ the species identified as non-threatened by our models and (2) prioritize the manual Red Listing of the 703 species listed

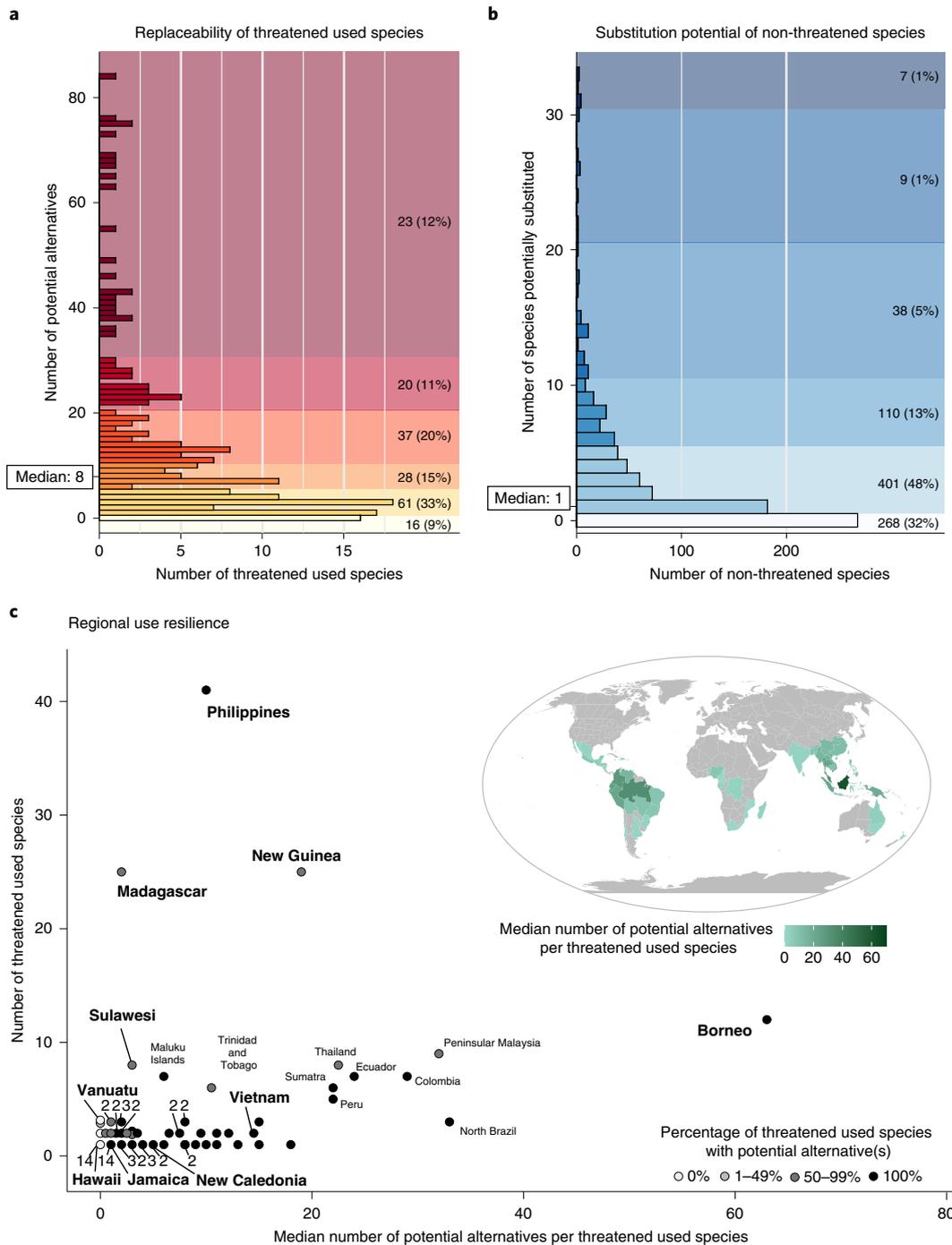


Fig. 3 | Availability of potential alternatives for threatened used palm species. a, Global replaceability of threatened used species. The replaceability of a species is defined as the number of potential alternatives identified for that species. **b**, Global substitution potential of non-threatened species. The potential for substitution of a species is defined as the number of threatened used species that may be substituted by that species. **c**, Resilience of threatened uses across regions. Use resilience is assessed by looking at the median number of potential alternative species per threatened used species for the region and by looking at the percentage of threatened used species with at least one potential alternative in the region. Each dot is a region with at least one threatened used palm species. Numbers indicate the number of identical dots overlapping. Region names are indicated in regular font for regions with >5 threatened used species and in bold for species-rich (>10 species) regions identified as top priorities for conservation in this study (Fig. 2).

as threatened by our most accurate model. Moreover, as further data and published assessments become available, our workflow can be re-used to update palm extinction risk predictions and to re-evaluate conservation priorities.

While ML extinction risk predictions have their limitations^{19,24,26}, our tests suggest that the results presented in this study are robust

to geographical or taxonomic biases in the ML training and test datasets and to the omission of extinction risk predictors (Methods, Supplementary Note and Supplementary Fig. 1). The main limitation in this study comes from the number of occurrence data points used. Indeed, our ML predictions revealed hundreds of likely threatened species, in some cases based on as few as one occurrence

record (Supplementary Table 5). On the one hand, this lack of data may have biased predictions towards higher probabilities of being threatened. On the other hand, many such poorly documented species predicted to be threatened may be truly rare and/or threatened and should therefore at least be considered priorities for research (Supplementary Note and Supplementary Fig. 2). The collection and curation of additional palm occurrence data will be essential to further improve the accuracy of ML predictions and to publish assessments for these species. Meanwhile, our lists of potentially threatened species include information on the underpinning data (Supplementary Table 5), thereby paving the way for developing research and conservation strategies so that palms can continue to provide ecosystem services and underpin livelihoods.

Integrating extinction risk evidence with phylogenetic, functional and ethnobotanical data enabled us to identify global geographic priorities for palm conservation and research that account for multiple aspects of palm diversity (Fig. 2). Even when using the regional percentage rather than the number of threatened species of interest as a prioritization criterion, palm-rich regions such as New Guinea, Madagascar and Borneo emerge as priorities. This is probably due to a combination of factors including the high species diversity and small range size of many palms (and other plants)⁶⁴ in these regions, combined with high pressures for land use change in some of them^{65–67}. Although some of the priority regions we identified were already suspected to be important for palm research and/or conservation⁶², this study identified New Guinea, Vietnam and Vanuatu as newly emerging among the top ten priority regions and provides a much-needed update for palm research and conservation projects globally. Importantly, by focusing on percentages rather than numbers of threatened species of interest, we shed light on 15 species-poor regions in which $\geq 40\%$ palm species of interest are probably threatened (Fig. 2 and Supplementary Table 4). Although palm species are not numerous in these regions, single palm species can be of high importance locally⁶² and our results will underpin further investigation about the threats they face in these regions.

Global and regional proportions of threatened used species were relatively low (Fig. 1). However, this remains concerning because many species and uses may be threatened locally even if they are not threatened globally⁶¹ and because many plant uses are not easily replaceable. For instance, species recorded as belonging to the same use category (for example, medicine) may have different and non-interchangeable uses (for example, for different ailments⁶⁸). Even co-occurring species providing apparently identical contributions to people may not be interchangeable, for example, if they can be used at different times of the year^{69,70}. On the other hand, many different species are used for similar purposes throughout the world^{54,71}, suggesting that some threatened used species may be replaceable by others. Our approach to evaluate regional use resilience combines phylogenetic, functional and ecological information and specifically addresses replaceability. However, it will need to be benchmarked against data on the transferability of uses and used species between communities and between regions, notably in the presence of fine-scale ecological heterogeneity. This will allow to identify the most suitable phylogenetic and functional distance thresholds and to select additional factors for classifying species as potential alternatives. Such additional drivers of species potential for substitution may include cultural priorities and practices, knowledge exchange between communities and regions, the ecology of threatened species and their potential alternatives, and the resilience of the latter to climate change⁷². By excluding these factors, our estimates of used species replaceability are probably too optimistic, as suggested by our finding that 91% of threatened used species may have alternatives (Fig. 3a). Yet, about a third of the regions comprising threatened used species appear to lack alternatives for some or all of them, highlighting instances of potentially low use resilience even under such a conservative approach

(Fig. 3c and Extended Data Figure 5). Furthermore, it is important to remember that many threatened used species are also functionally and/or evolutionarily distinct⁴¹, so their loss could negatively impact ecosystems and humanity, even if they were found to be replaceable in terms of their direct use by people. Mitigating or adapting to the loss of used species will be best achieved by their users themselves, so that local priorities and knowledge can be fully accounted for. Our list of candidate alternative species per region (Supplementary Table 6) set an optimistic baseline that can underpin community-led benchmarking and conservation actions, for example, to assess whether alternative species are already locally used or whether they are as useful as predicted.

On a broader scale, the methods outlined in this paper urgently need to be tested on other plant groups which also contribute to supporting ecosystem functioning and people's well-being and livelihoods. This will enable the rapid identification of a greater diversity of species most at risk and the development of effective and culturally relevant conservation strategies to enhance ecosystem health and the sustainable use of plants globally.

Methods

Machine learning predictions of extinction risk. *Species sampling and cleaning of spatial occurrence data.* We aimed to sample all the 2,510 palm species recognized at the time of the study⁷³. However, since predictors for species conservation status can be obtained more precisely from occurrence data than from species presence/absence records at the region level, our machine learning analyses only included species with at least one valid occurrence record. The few palm species known only from cultivation were kept in the dataset as they represent a negligible fraction of all species. Global spatial occurrence data for 7,469 palm names with a Global Biodiversity Information Facility (GBIF) key out of the 7,570 published palm names⁷³ were sourced from GBIF (derived dataset GBIF.org <https://doi.org/10.15468/dd.at82kf>) using the R package *rgbio* v.0.9.9 (ref. ⁷⁴). Another 14,169 occurrence data points were obtained from herbarium specimen records from the database of the Royal Botanic Gardens, Kew (UK) and 106 from the database of the Naturalis Biodiversity Center, Leiden (the Netherlands). Each occurrence point was assigned to one of the 2,510 accepted palm species names⁷³, or discarded, depending on whether the name associated with the occurrence point in GBIF could be unambiguously matched to an accepted name. Occurrence records were cleaned on the basis of the GBIF coordinate issue flags and using the R package *CoordinateCleaner* v.1.0–7 (ref. ⁷⁵). Obvious issues such as wrong coordinate signs were corrected, while coordinates falling into marine areas, cities, province or country centroids or biodiversity institutions were removed. Coordinates with zero values, an uncertainty > 100 km, inconsistent with country assignment, falling outside the reported native distribution range of the species, considered extinct in the wild (both following Plants Of The World Online⁷⁶) or recorded before 1945 (when the precision of geolocalization devices was poor) were also removed, following recommendations from the authors of *CoordinateCleaner* (https://ropensci.github.io/CoordinateCleaner/articles/Cleaning_GBIF_data_with_CoordinateCleaner.html). Duplicated occurrence records were omitted. In total, 1,820 species (72.5%) had at least one clean occurrence and could thus be used in the ML analyses (Extended Data Fig. 1). Cleaned occurrence data can be found in Supplementary Table 7. Additional R packages used for cleaning occurrence data points included *devtools* v.1.13.5, *tidyverse* v.1.3.1, *countrycode* v.1.1, *maps* v.3.3.0, *maptools* v.0.9-2, *rworldmap* v.1.3-6 and *sp* v.1.2-7 (refs. ^{77–83}).

Choice of extinction risk predictors. The classification of species into different extinction risk categories in the Red List is based on population size, trends in population sizes (for example, loss of habitat or declines due to species exploitation), species range size (for example, restricted and fragmented) and habitat quality (for example, impact of pests and invasive species)³⁷. Predictors providing information on a species' range size, namely extent of occurrence (EOO) and area of occupancy (AOO), were selected as they are more readily available than those relating to declines or population size. Two additional range-based metrics, i.e. number of subpopulations and number of locations (definitions¹⁸ provided in Supplementary Table 8), were included as they can be useful predictors, even if not explicitly aligned with IUCN criteria. Coarse-scale distribution data have also been shown to be useful predictors of extinction risk for plants²¹, we therefore also included the number of level-3 botanical countries occupied. Level-3 botanical countries are biogeographical units defined by the World Geographical Scheme for Recording Plant Distributions⁸⁴ to reflect political country boundaries while taking into account botanical tradition and botanical heterogeneity within and between political countries⁸⁵. For simplicity, we refer to them as 'regions' hereafter and in the main text. Detailed data on habitat quality and species exploitation are rarely available in a form that can be used in automated extinction risk assessments. However, the influence of habitat quality on a species is likely to be less important

if the species has a large ecological amplitude and forest species like most palms are more likely to be subject to population size decline if they occur in areas strongly impacted by humans and especially deforestation. Accordingly, we included eight further predictors in our analyses, that is climatic amplitude in terms of temperature, climatic amplitude in terms of precipitation, average temperature seasonality, average precipitation seasonality, number of ecoregions⁸⁶ occupied (defined in Supplementary Table 8), human impact, human density and forest loss. All 13 predictors are described with their sources in Supplementary Table 8.

Species value calculations for each predictor. Species EOO and AOO were calculated from the occurrence data using the R package rCAT v.0.1.6 (ref. ⁸⁷) or the R package ConR v.1.3 (ref. ¹⁸) when species only had two occurrence data points available. EOO could not be estimated for species with one occurrence point, while AOO could be calculated for all species analysed. Numbers of ecoregions and of regions occupied were obtained from the source datasets^{84,88,89} (Supplementary Table 8) using custom R scripts relying on the R packages sf v.0.9, doParallel v.1.0.16, foreach v.1.5, httr v.1.4.2, jsonlite v.1.7.2 and progress v.1.2.2 (refs. ^{81,90–94}). The Human Footprint Index (used to estimate human impact), human population density, temperature seasonality, precipitation seasonality, minimal temperature of the coldest month, maximal temperature of the warmest month, average precipitation of the driest month and average precipitation of the wettest month were obtained from global raster layers^{95,96} (Supplementary Table 8) and derived at 10 min resolution to match the precision uncertainty in the coordinates of occurrence records, while reducing computational burden. For each species, we extracted predictor values at each occurrence location and averaged them to obtain one value per species. The temperature and precipitation data were used to calculate temperature amplitude (average minimal temperature of the coldest month/ average maximal temperature of the warmest month) and precipitation amplitude (average precipitation of driest month/ average precipitation of wettest month), respectively. These indices account for cases where seasonality is low at each occurrence point for a species but high when considering all occurrence records. For forest loss, values assigned to species represented the proportion of species occurrence points found in areas that experienced forest loss between 2001 and 2018⁹⁷. The predictors were rescaled to range between 0 and 1 to improve the estimation of the models' parameters. All data geoprocessing was performed using the R packages raster v.3.1–5, gdalUtils v.2.0.3.2, rgdal v.1.5–8, maps v.3.3.0, dplyr v.1.0.7, tidyr v.1.1.4, tidyverse v.1.3.1 and plyr v.1.8.6 (refs. ^{80,82,98–103}).

Delimitation of training and test subsets. We obtained all available extinction risk assessments for palms first from the Red List¹⁰⁴ and then from the global assessments collated in the ThreatSearch database of Botanic Gardens Conservation International¹⁰⁵ when no assessment was available in the Red List. ThreatSearch assessments were only considered global if they were specified as such (assessments of unknown scope were not considered to be global). Species with assessments made before 2008 were considered unassessed to ensure an up-to-date training set and 'data deficient' (DD) species were also considered unassessed. Among the 1,820 species with occurrence data, 439 assessed species had a non-DD extinction risk assessment from 2008 or later (321 from the Red List and 118 from ThreatSearch) and were available to train or test (see below) the models used to predict the extinction risk of the remaining 1,381 unassessed species with occurrence data. However, assessed species were heavily biased towards the genus *Dyopsis* from Madagascar⁶⁰. We therefore randomly subsampled *Dyopsis* species to balance their proportion in the assessed group with their proportion in the unassessed group, resulting in only two *Dyopsis* species left in the assessed group. After removing most *Dyopsis*, the taxonomic and geographic representation of the assessed and unassessed groups became more similar, although a small degree of geographic imbalance remained (Supplementary Fig. 2). The resulting group of 300 assessed species comprised 130 'non-threatened' species and 170 'threatened' species. The non-threatened category included species classified as 'least concern' (LC) in the Red List or as not threatened in ThreatSearch^{104,105}. The threatened category included species classified in the Red List as critically endangered (CR), endangered (EN), vulnerable (VU) or near threatened (NT) or in ThreatSearch as threatened, near threatened or possibly threatened (Supplementary Table 5). These steps were performed using the R packages rredlist v.0.7.0, stringr v.1.4.0 and stringi v.1.7.5 (refs. ^{106–108}).

To evaluate model performance, we divided this dataset of 300 representative assessed species into a training set comprising 225 species (75%) and a test set comprising the 75 (25%) remaining species. The training set was used for model parameterization, while the test set was only used to assess model performance, independent of the model parameterization process. To increase its representativeness, the test set was built iteratively by first randomly choosing 15 assessed species and adding 60 assessed species sequentially so that each added species would be as dissimilar as possible to the species already present in the test set, based on their extinction risk predictor values (Supplementary Table 5). This was done using the function maxDissim from the R package caret v.6.0 (ref. ¹⁰⁹) with the default settings after preliminary tests indicated no effect of changing these settings on the representativeness of the test set. Details of the datasets are provided in Extended Data Fig. 1.

Addressing missing or biased data and correlated predictors. Three extinction risk predictors had missing values for some species: EOO (390 species), human footprint (4 species) and forest loss (9 species). The knnImputation function of the R package DMwR v.0.4.1 (ref. ¹¹⁰) was used to fill missing values by averaging the values of the species' five nearest neighbours in terms of extinction risk predictors. Extinction risk predictors showed various degrees of correlation between each other (predictor redundancy), so ML analyses were run once without considering correlation among predictors and once after removing the predictors with a correlation coefficient >0.75 by using the 'cutoff' option of the preProcess function in the caret package¹⁰⁹.

To account for the imbalance of the extinction risk categories in the training set, each ML analysis was performed first without any resampling and then repeated once with downsampling of the majority extinction risk category (down), once with upsampling of the minority category (up) and once with a method synthesizing new data for the minority category using the synthetic minority oversampling technique (smote)¹¹¹.

Representation biases in the training and test sets were evaluated by plotting histograms of each extinction risk predictor for each dataset (unassessed, training and test) and calculating their degree of intersection with the function intersect.dist of the R package HistogramTools¹¹². All intersections were high (≥ 0.81 and ≥ 0.95 for the intersection between the unassessed and test data) but visual inspection of the histogram overlaps revealed that human footprint, human population density and the four temperature and precipitation predictors had some parts of their distributions under-represented in both the training and test sets (Supplementary Fig. 2). We therefore trained ML models either including or excluding these predictors.

Taken together, these sensitivity analyses represented 16 different models for a given ML method (2 (with/without strongly correlated predictors) \times 4 (resampling strategy) \times 2 (with/without predictors with representation biases)). We combined these with three different ML methods (see below), giving a total of 48 models. A summary of the approach is provided in Extended Data Fig. 1. In addition to the above-cited R packages, evaluating and visualizing the representativeness, imbalance and redundancy of the datasets also relied on lattice v.0.20–40, ggplot2 v.3.3, gridExtra v.2.3, mosaic v.1.6.0, proxy v.0.4–23, plyr v.1.8.6, scales v.1.1 and UBL v.0.0.6 (refs. ^{98,113–119}).

Choice of machine learning method and model tuning. We first used a random forest algorithm to build the predictive models using the R package randomForest¹²⁰ through implementation in the R package caret¹⁰⁹. This method is hereafter referred to as the 'RF method'. We fitted each model on the training set using ten repeats of tenfold cross-validations and calculated the average Kappa¹²¹ for the model on the basis of the training data. This was repeated many times to tune the 'mtry' parameter to find out how many extinction risk predictors used to split a node in the tree gave the best Kappa for a given model. Kappa was chosen because it has previously been shown to perform better than other metrics for unbalanced datasets (definition below). All entire numbers between one and the number of independent extinction risk predictors available were tested (up to 13, depending on whether predictors with shifted representation or high correlation were included; see above).

During training and testing of the models, binary classification of the species into threatened and non-threatened was based on their estimated probability of being threatened, with a probability threshold of 0.5 between classes. However, this approach may be problematic with unbalanced training sets because the probabilities may be skewed towards the most represented class. For each trained random forest model, we therefore also estimated what probability threshold allowed us to maximize both specificity and sensitivity (see below). This method is hereafter referred to as the 'RFt method'. This was done by modifying the original RF method so that the class probability threshold was considered as a parameter to tune (among 20 ranging regularly from 0 to 1), in addition to mtry. The parameters were tuned by looking for the mtry and threshold combination conferring both the highest specificity and sensitivity and comparing them to the theoretical perfect specificity and sensitivity values of 1. This additional performance indicator was called distance to perfect model (DPM, below). This approach was adapted from an example described in the caret manual (section 13.8 of <https://topepo.github.io/caret/index.html>).

To compare very different methods, we also trained a neural network with a single internal layer using the same cross-validation approach as for the random forest. This method is hereafter referred to as the 'NN method'. The number of neurons in the internal layer (size parameter) and the weight decay were tuned by finding the size and decay values maximizing Kappa, among the following values—size: 1, 2, 3, 5, 7, 9, 11, 13; decay: 0.005, 0.01, 0.05, 0.1, 0.5, 1, 2, 5, 8, 10. This was performed using the R package nnet v.7.3–13 (ref. ¹²²) through implementation in caret. Each network was run for 1,000 iterations, which allowed all runs to reach convergence.

All 48 random forest and neural network models are listed in Supplementary Table 1, together with their dataset and parameter specifications. Additional packages required to perform the model training included gdata v.2.18.0, ranger v.0.12.1, e1071 v.1.7–3 and RANN v.2.6.1 (refs. ^{123–126}).

Model performance and prediction choice. To select the model to use for our final analyses, the performance of the 48 models was assessed on the test set by using each model to predict the extinction risk of the test species, comparing the predictions to the published (observed) extinction risks and calculating nine performance indicators: (1) area under the receiver operating characteristic curve (AUC)—how well a model distinguishes threatened species from non-threatened ones (ranges between 0 and 1, with 1 indicating perfect discriminatory power); (2) sensitivity, the percentage of threatened species correctly classified; (3) specificity, the percentage of non-threatened species correctly classified; (4) DPM, measuring how far the model is from a perfect classification where sensitivity = specificity = 1; (5) accuracy, the percentage of correct classifications; (6) balanced accuracy, calculated by averaging sensitivity and specificity; (7) Cohen's Kappa coefficient, comparing the accuracy of the model relative to that of a random classification based only on class frequencies (ranges between -1 and 1, with 1 indicating a perfect classifier); (8) precision for the positive class (hereafter 'precision')—the percentage of species predicted to be threatened being indeed threatened; and (9) precision for the negative class (hereafter 'negative precision')—calculating the percentage of species predicted as non-threatened being indeed non-threatened. In addition, we also attributed a weight to each model on the basis of its balanced accuracy and binarily coded predictions as 0 when the species was predicted to be non-threatened and 1 when it was predicted to be threatened. We then estimated a weighted average binary prediction across models and estimated the performance of this averaging method using the above indicators.

Performance indicator values for all models are provided in Supplementary Table 1 and show that, although all three methods (RF, RfT and NN) performed well, RF and RfT models performed more similarly to each other than to NN models, regardless of the performance indicator considered. No model scored the highest for all indicators, reflecting trade-offs between sensitivity and specificity. Within a given ML method, performance was more affected by resampling strategy and extinction risk predictor representativeness than by predictor redundancy due to the existence of correlations between predictors (Supplementary Table 1). The model with the highest balanced accuracy (82%) was used to predict the extinction risk of the 1,381 unassessed species. This model was a random forest with optimized class probability threshold and an upsampling strategy, from which correlated and shifted predictors were removed (Supplementary Table 1). Details of the limitations of the different models tested and their robustness to data imbalances or shortages are provided in the Supplementary Note.

Drivers of wrong predictions and predictor importance. We used the framework of Shapley additive explanations (SHAP)¹²⁷ to interrogate the behaviour of our selected model. We calculated SHAPs for all test set predictions made by the model, using the implementation in the R package *fastshap* v.0.0.7 (ref. 128). We used these SHAPs to generate explanations for overall model behaviour by calculating predictor importance as the mean absolute SHAP value of each predictor across all test set predictions. We also visualized the distribution of SHAP values to compare the partial dependence of test set predictions on each extinction risk predictor. We used explanations of individual predictions to highlight the prediction pathway of species that the model predicted incorrectly. For comparison, we also calculated the predictor importance with *vip* v.0.3.2 (ref. 129) by randomly shuffling the values of each predictor in turn and calculating the resulting decrease in accuracy of the test set predictions. We repeated this process 1,000 times per predictor and reported the importance as the mean decrease in accuracy. Although the permutation-based importance should be consistent with the SHAP-based importance, permutation-based predictor importance gives an indication of each predictor's contribution to the accuracy of a model, while SHAP-based importance indicates the average contribution of each predictor to the predicted values themselves. The results and implications of these analyses are provided in Supplementary Note and Supplementary Fig. 1. Additional R packages used to visualize model performance and behaviour included *glm2* v.1.2.1, *reshape2* v.1.4.4, *pROC* v.1.18.0, *ggplot2* v.3.3, *gplots* v.3.0.3, *here* v.1.0.1, *readr* v.1.3.1, *readxl* v.1.3.1, *purrr* v.0.3.4, *ggforce* v.0.3.1, *patchwork* v.1.0.0, *glue* v.1.3.1, *writexl* v.1.2, *scales* v.1.1, *plyr* v.1.8.6, *dplyr* v.1.0.7, *stringr* v.1.4.0, *tidyr* v.1.1.4, *ggfortify* v.0.4.8, *randomForest* v.4.6-14 and *ggpubr* v.0.2.5 (refs. 98–100,107,114,118,130–144).

Identification of conservation priorities. Species use and evolutionary and functional distinctiveness. We classified palm species as being 'of interest' (Results) on the basis of trait data retrieved from *PalmTraits*⁴⁹ and of use data retrieved from the World Checklist of Useful Plant Species⁵⁴ and the literature^{37,145,146}. A species was considered of interest if it was used by humans and/or had a higher evolutionary and/or functional distinctiveness than the median distinctiveness. Species were considered to be used by humans if they had at least one use recorded among the following categories: 'animal food', 'culture', 'environmental', 'fuel', 'gene sources', 'food', 'medicine', 'toxic', 'utensils, tools and construction' and 'other uses' (category matching between the different data sources is described in Supplementary Table 9). This category scheme was adopted because it generally follows a widely used classification¹⁴⁷ for palm uses while accounting for the fact that the utensils and tools and construction categories used in that classification were merged in some of our data sources (Supplementary Table 9).

To calculate species evolutionary distinctiveness²⁷ (ED) the *ed.calc* function of the R package *caper* v.1.0.1 (ref. 148) was applied to 750 species-level palm-wide phylogenetic trees. These trees were obtained from a study⁵⁶ presenting multiple Bayesian phylogenetic analyses of all palm species described at the time relying on a compilation of morphological, genetic and taxonomic data. The trees we used correspond to the post-burnin posterior distribution of trees generated by their 'unconstrained' analysis with a taxonomic backbone following modifications by ref. 59. The ED value of each species was then obtained by averaging the values obtained for that species in the 750 trees. Ninety-three species recognized by our taxonomic backbone were absent from the trees—even after using the World Checklist of Selected Plant Families⁷³ to identify synonyms—and therefore lacked ED values. A few species from the original trees were considered synonyms in our taxonomic backbone, so their ED values were averaged (this happened 38 times, involving mostly two but up to seven synonyms). While these two problems could bias the ED values of the species remaining in the tree, this bias is likely to be limited because species lacking ED values were spread throughout the phylogenetic trees and represented only 3.7% of the total diversity of the family. We therefore consider that our ED calculations are of sufficient precision to be used for our purpose, which relies on splitting ED values into two categories (higher versus lower than the median) rather than relying on exact ED values.

To calculate functional distinctiveness²⁸ (*sensu ref. 28* (p. 1,366) 'average functional distance of a species to the other species in the community', with the community here being the global palm species diversity) the 'distinctiveness' function of the R package *funrar*²⁸ v.1.4.0 was used with the Gower distance option. The traits extracted from the *PalmTraits*⁴⁹ database comprised: maximum stem height, maximum stem diameter, maximum leaf number, maximum blade length, average fruit length, average fruit width, presence of spines, growth form coded as all possible combinations between climbing and/or acaulescent and/or erect, stem habit coded as all possible combinations between solitary and/or clustering and vegetation stratum coded as all possible combinations between understorey and/or canopy. All continuous trait variables were log-transformed and rescaled to range between 0 and 1. There were 85 species with no trait data at all and 435 species lacking more than three traits. When excluding the former 85 species, the trait data had 21% of missing values. These missing data were estimated by averaging the values from their 11 nearest neighbours in the trait space using the *kNN* and *weightedMean* functions of the R packages *VIM* v.6.1.1 and *laeken* v.0.5.2 (refs. 149,150). For categorical traits, missing values were obtained by taking the most represented category among the 11 neighbours. Traits for the 85 species with only missing values were not imputed and these species were therefore not considered to be functionally distinct.

Proportions of threatened species and priority regions. Four datasets were produced to explore the influence of using ML predictions on estimates of the proportions of threatened species and on conservation prioritization (Extended Data Fig. 1). The first dataset (all published) only contained extinction risk information for species with published assessments on the Red List or on ThreatSearch as described above. The second dataset (recently published) was a subset of the all published dataset including only extinction risk information for species with assessments published from 2008 onwards, thereby following the IUCN guidelines that >10-year-old assessments need updating¹⁶ (<https://www.iucnredlist.org/assessment/process> accessed 16 September 2021). The third dataset (total evidence) combined the extinction risk information from the recently published dataset with the most accurate ML extinction risk predictions for the species with spatial occurrence data that were either unassessed, assessed as DD or with old assessments. A fourth dataset was produced by combining the 'total evidence' dataset with information on the extinction risk of an additional 521 species for which we did not have spatial occurrence data or recent published assessments (Extended Data Fig. 1). This information was gathered from old published assessments for 296 of species or produced as a quick 'expert guess' based on our palm expertise for the remaining 225 species. This dataset was only used as an attempt to provide an even more complete view of palms at risk accounting for very poorly documented taxa (Supplementary Note). To remain conservative, this last dataset was not used in the prioritization or use resilience analyses and proportions of threatened species estimated from it are not discussed in the main text. For each dataset, the proportions of threatened species in the evolutionarily distinct, functionally distinct and used species (distinguishing four use categories; Results) were calculated for the world and separately for the America (longitude $\in (-180^\circ, -25^\circ)$), Africa/ West Asia (longitude $\in (-25^\circ, 68^\circ)$) and East Asia/ Pacific (longitude $\in (68^\circ, 180^\circ)$) regions. This splitting scheme was chosen because there are no palm species that have native ranges spanning more than one of these areas⁶. The same analysis was then performed at the region level (for each botanical country⁵⁴, see above) on the all published, recently published and total evidence datasets.

In addition to separating proportions of threatened species among evolutionarily distinct, functionally distinct or used species, we also calculated proportions of threatened species among species fitting in at least one of these categories (species of interest). This was calculated for each region and used as a criterion for conservation prioritization (regions with higher proportions of threatened species among species of interest were considered higher priorities).

Using proportions instead of number of threatened species of interest allowed us to account for regions where few but valuable species may be threatened and it did not prevent species-rich regions from being captured in the top priorities (Results). However, for comparison, we repeated the ranking with only regions comprising more than ten palm species. These analyses were performed on the all published, recently published and total evidence datasets, to assess if priority regions changed when using ML predictions compared to published assessments only. To estimate the robustness of the results to the fact that some species had their EOO imputed (previous section), proportions of threatened species and region ranks were also inferred on the basis of the total evidence dataset after excluding species that had their EOO imputed. The results are presented in Supplementary Tables 3 and 4 and compared to the main results in the Supplementary Note and Supplementary Fig. 3.

Potential alternatives for threatened used species. *Rationale.* Traits and genes underpin plant uses to a certain extent³⁷, so data on plant morphology and phylogenetic placement may help to predict if a species could be used for the same use as another species. Although phylogenetic signal in plant use may vary from non-existent to very strong depending on the uses or the taxa considered, accumulating evidence suggests that plant uses are phylogenetically clustered to some extent^{38–42,151}, including in the palm family³⁷. Our data corroborated these findings, as seven out of ten use categories, including the four most represented (food, utensils, tools and construction, medicine and culture), showed some degree of phylogenetic signal based on Fritz and Purvis's *D* statistic tests¹⁵². The tests were applied to the 750 above-mentioned palm phylogenetic trees obtained from ref. ⁵⁶ and to a maximum clade credibility consensus of these trees (Supplementary Data) obtained with Tree Annotator (part of BEAST v.1.10.2; ref. ¹⁵³) and results are provided in Supplementary Tables 9 and 10. The relationship between plant uses and functional traits is more difficult to characterize but previous work on palms has shown that some uses are correlated to morphological traits³⁷.

As a first attempt to explore how these theoretical and empirical relationships can help in predicting regional use resilience, we used functional trait data, phylogenetic information and ethnobotanical information to identify non-threatened species that may be suitable substitutes for a threatened species occurring in the same region. To account for the fact that species may not be easily moved across different ecological settings, we further restricted the search for potential alternatives to species occurring in the same biome as the threatened used species to replace. Species were therefore considered as potential alternatives only if they fulfilled the following four conditions: (1) being assessed or predicted as non-threatened, (2) occurring in the same region as the species to replace, (3) being known to occur in at least one of the biomes occupied by the species to replace and (4) being significantly close to the species to replace in terms of phylogenetic and/or functional distance. Being significantly phylogenetically or functionally close in this context was understood as having a phylogenetic or functional distance between the species that was inferior or equal to the median of the pairwise distances between species in the considered biome of the considered region minus the standard deviation of these distances. We chose to use local thresholds defined for each region and biome combination for two reasons. First, they allow to account for the fact that species from one biome may not easily be found or grown in another biome. Second, using local phylogenetic and functional distance thresholds instead of global thresholds better reflects the reality of communities who need to find alternatives among the species that are available in their region/biome and who may therefore choose an alternative on the basis of its similarity to the species to replace, regardless of if more similar alternatives occur elsewhere. The search for alternatives was done once using phylogenetic distances and once using functional distances and the union of both lists of alternatives thereby identified was used because it provided a more conservative (optimistic) view of region use resilience and species replaceability across uses. Moreover, more stringent thresholds than the median distance minus standard deviation could be experimented and would result in smaller lists of potential alternative species and lower regional use resilience estimates. We chose to use a threshold providing a relatively optimistic view to obtain a baseline from which more pessimistic scenarios can be envisioned and because larger lists of potential alternatives may facilitate the identification of realistic alternatives at the community level. To estimate the robustness of these lists to the fact that some species had their EOO imputed (see above), the search for potential alternatives, estimations of species replaceability and regional use resilience analyses were also performed based on the total evidence dataset after excluding species that had their EOO imputed. The results are presented in Supplementary Tables 4 and 5 and compared to the main results in the Supplementary Note.

Data used to search for potential alternatives. Phylogenetic distances were calculated as the averaged sum of the branch lengths linking each pair of species in a sample of 100 trees chosen randomly from the above-mentioned 750 palm phylogenetic trees from ref. ⁵⁶. This was done with the function 'cophenetic.phylo' from the R package ape v.5.0 (ref. ¹⁵⁴) and the trees sampled are listed in the scripts (Data availability). For species not in the tree, the distance was obtained as the median distance between congeneric species. Only two species not in the tree and with no congeneric species did not have distances and had to be excluded from

the calculations: *Sabinaria magnifica* Galeano & R. Bernal and *Wallaceodoxa raja-ampat* Heatubun & W.J. Baker. Functional distances were calculated on the basis of three traits that were found to be associated with palm use in a previous study: maximum leaf (blade) length, fruit volume and stem volume ($\pi \times r \times r \times h$; with *r* being the stem radius, derived from maximum stem diameter and *h* being the maximum stem height)³⁷. The traits were obtained from the trait data matrix with imputed missing values used above for the calculation of functional distinctiveness values and the function `compute_dist_matrix` from the R package `funrar` v.1.4.1 (ref. ²⁸) was used to calculate the distance between species in this three-dimensional trait space. The above-mentioned 85 species that could not have their traits imputed were then added to the distance matrix and the distances between these species and the rest of the species were obtained as the median distances between congeneric species. Phylogenetic and functional distances were log-transformed and rescaled to range between 0 and 1. The biome(s) occupied by each species was obtained from the World Checklist of Vascular Plants¹⁵⁵ accessed in February 2021 and consisted in a categorical variable with six categories ('desert or dry shrubland', 'montane tropical', 'seasonally dry tropical', 'subtropical', 'temperate' and 'wet tropical'). Two species (*Acoelorrhaphes wrightii* and *Dyopsis declivium*) lacked biome data and had to be excluded from the analysis. Trait and biome data are provided in Supplementary Table 5.

Maps and graphs presenting all above results were obtained using the R packages `ggplotify` v.0.0.4, `aplot` v.0.0.3, `stringr` v.1.4.0, `ggrepel` v.0.8.1, `GGally` v.1.4.0, `hrbrthemes` v.0.6.0, `rgdal` v.1.5-8, `scales` v.1.1, `reshape2` v.1.4.4, `dplyr` v.1.0.7, `gridExtra` v.2.3, `ggpubr` v.0.2.5, `ggstance` v.0.3.3, `ggtree` v.2.4.1, `hash` v.2.2.6.1, `ggplot2` v.3.3 and `plyr` v.1.8.6 (refs. ^{98,100,101,107,115,118,131,144,156–164}). All analyses were performed with R v.4.0.2 in RStudio^{165,166}.

Limitations and guidelines for interpreting the results. Traits and genes are not the only drivers of species use, so we may have over- or under-estimated the availability of potential alternatives in some cases. Ideally, the search for alternatives should consider cultural practices and preferences but these data are currently not available at a global scale. In addition, the occurrence of the species to replace and its potential substitute in different subregions or different ecological conditions could be an obstacle to their interchangeability. If these differences were not captured by the use of biome data, the availability of potential alternative species may have been overestimated. Another concern may be that we do not know if a species reported to be used somewhere is (or could be) used in the same way throughout its distribution range. However, to our knowledge, there is no reason to assume that a species could not be used in a certain way in a region just because its use there is not yet known. In fact, communities are constantly experimenting with new species, as evidenced by the widespread use of non-native species in local pharmacopoeias⁷¹ and by a higher likelihood of naturalization in plants with economic value³⁹. Finally, there may be cases where extremely functionally and phylogenetically distant species could successfully be used as substitutes of each other in some regions. These will be missed by the search for alternative species as implemented here, so our results should not be interpreted as evidence that species that were not identified as potential alternatives cannot be useful. Overall, our results illustrate a potential for replacement that will have to be ground checked and discussed with communities. Until then, they remain useful as an (optimistic) estimate of regional resilience of palm uses.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data necessary to perform the analyses are provided as Supplementary Tables or on Zenodo (<https://zenodo.org/>)¹⁶⁷. Source data are provided with this paper.

Code availability

All scripts necessary to perform the analyses are provided on Zenodo (<https://zenodo.org/>)¹⁶⁷.

Received: 15 November 2021; Accepted: 24 July 2022;

Published online: 26 September 2022

References

1. Isbell, F. et al. High plant diversity is needed to maintain ecosystem services. *Nature* **477**, 199–202 (2011).
2. van der Sande, M. T. et al. Biodiversity in species, traits, and structure determines carbon stocks and uptake in tropical forests. *Biotropica* **49**, 593–603 (2017).
3. Grace, O. M. et al. Plant power: opportunities and challenges for meeting sustainable energy needs from the plant and fungal kingdoms. *Plants People Planet* **2**, 446–462 (2020).
4. Howes, M. J. R. et al. Molecules from nature: reconciling biodiversity conservation and global healthcare imperatives for sustainable use of medicinal plants and fungi. *Plants People Planet* **2**, 463–481 (2020).

5. Ulian, T. et al. Unlocking plant resources to support food security and promote sustainable agriculture. *Plants People Planet* **2**, 421–445 (2020).
6. Brondizio, E., Diaz, S., Settele, J. & Ngo, H. T. (eds) Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on biodiversity and ecosystem services. *Zenodo* <https://doi.org/10.5281/zenodo.3831673> (2019).
7. Bennun, L. et al. The value of the IUCN Red List for business decision-making. *Conserv. Lett.* **11**, e12353 (2018).
8. Betts, J. et al. A framework for evaluating the impact of the IUCN Red List of threatened species. *Conserv. Biol.* **34**, 632–643 (2020).
9. Maira, L. et al. Achieving international species conservation targets: closing the gap between top-down and bottom-up approaches. *Conserv. Soc.* **19**, 25–33 (2021).
10. *IUCN Red List version 2022-2: Table 1a* (IUCN, 2022); <https://www.iucnredlist.org/resources/summary-statistics#Figure2>
11. Rivers, M. The global tree assessment—red listing the world's trees. *BGJournal* **14**, 16–19 (2017).
12. Nic Lughadha, E. et al. Extinction risk and threats to plants and fungi. *Plants People Planet* **2**, 389–408 (2020).
13. Silva, S. V. et al. Global estimation and mapping of the conservation status of tree species using artificial intelligence. *Front. Plant Sci.* **13**, 839792 (2022).
14. *ThreatSearch Online Database* (Botanic Gardens Conservation International, accessed 12 October 2021); https://tools.bgci.org/threat_search.php
15. Bachman, S. P., Nic Lughadha, E. M. & Rivers, M. C. Quantifying progress toward a conservation assessment for all plants. *Conserv. Biol.* **32**, 516–524 (2018).
16. Rondinini, C., Di Marco, M., Visconti, P., Butchart, S. H. M. & Boitani, L. Update or outdated: long-term viability of the IUCN Red List. *Conserv. Lett.* **7**, 126–130 (2014).
17. Cazalis, V. et al. Bridging the research–implementation gap in IUCN Red List assessments. *Trends Ecol. Evol.* **37**, 359–370 (2022).
18. Dauby, G. et al. ConR: an R package to assist large-scale multispecies preliminary conservation assessments using distribution data. *Ecol. Evol.* **7**, 11292–11303 (2017).
19. Stévant, T. et al. A third of the tropical African flora is potentially threatened with extinction. *Sci. Adv.* **5**, eaax9444 (2019).
20. Bland, L. M., Collen, B., Orme, C. D. L. & Bielby, J. Predicting the conservation status of data-deficient species. *Conserv. Biol.* **29**, 250–259 (2015).
21. Darrah, S. E., Bland, L. M., Bachman, S. P., Clubbe, C. P. & Trias-Blasi, A. Using coarse-scale species distribution data to predict extinction risk in plants. *Divers. Distrib.* **23**, 435–447 (2017).
22. Pelletier, T. A., Carstens, B. C., Tank, D. C., Sullivan, J. & Espíndola, A. Predicting plant conservation priorities on a global scale. *Proc. Natl Acad. Sci. USA* **115**, 13027–13032 (2018).
23. Zizka, A., Silvestro, D., Vitt, P. & Knight, T. M. Automated conservation assessment of the orchid family with deep learning. *Conserv. Biol.* **35**, 897–908 (2021).
24. Walker, B. E., Leão, T. C. C., Bachman, S. P., Bolam, F. C. & Nic Lughadha, E. Caution needed when predicting species threat status for conservation prioritization on a global scale. *Front. Plant Sci.* **11**, 520 (2020).
25. Lughadha, E. N. et al. The use and misuse of herbarium specimens in evaluating plant extinction risks. *Philos. Trans. R. Soc. B* **374**, 20170402 (2019).
26. Walker, B. E., Leão, T. C. C., Bachman, S. P., Lucas, E. & Nic Lughadha, E. M. Evidence-based guidelines for developing automated assessment methods. Preprint at <https://ecoevorxiv.org/zxq6s/> (2021).
27. Isaac, N. J. B., Turvey, S. T., Collen, B., Waterman, C. & Baillie, J. E. M. Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PLoS ONE* **2**, e296 (2007).
28. Grenié, M., Denelle, P., Tucker, C. M., Munoz, F. & Violle, C. funrar: an R package to characterize functional rarity. *Divers. Distrib.* **23**, 1365–1371 (2017).
29. Lindegren, M., Holt, B. G., MacKenzie, B. R. & Rahbek, C. A global mismatch in the protection of multiple marine biodiversity components and ecosystem services. *Sci. Rep.* **8**, 4099 (2018).
30. Pollock, L. J. et al. Protecting biodiversity (in all its complexity): new models and methods. *Trends Ecol. Evol.* **35**, 1119–1128 (2020).
31. Arnan, X., Cerdá, X. & Retana, J. Relationships among taxonomic, functional, and phylogenetic ant diversity across the biogeographic regions of Europe. *Ecography* **40**, 448–457 (2017).
32. Wong, J. S. Y. et al. Comparing patterns of taxonomic, functional and phylogenetic diversity in reef coral communities. *Coral Reefs* **37**, 737–750 (2018).
33. Devictor, V. et al. Spatial mismatch and congruence between taxonomic, phylogenetic and functional diversity: the need for integrative conservation strategies in a changing world. *Ecol. Lett.* **13**, 1030–1040 (2010).
34. Brum, F. T. et al. Global priorities for conservation across multiple dimensions of mammalian diversity. *Proc. Natl Acad. Sci. USA* **114**, 7641–7646 (2017).
35. Pollock, L. J., Thuiller, W. & Jetz, W. Large conservation gains possible for global biodiversity facets. *Nature* **546**, 141–144 (2017).
36. Strassburg, B. B. N. et al. Global priority areas for ecosystem restoration. *Nature* **586**, 724–729 (2020).
37. Cámara-Leret, R. et al. Fundamental species traits explain provisioning services of tropical American palms. *Nat. Plants* **3**, 16220 (2017).
38. Salsis-Lagoudakis, C. H. et al. Phylogenies reveal predictive power of traditional medicine bioprospecting. *Proc. Natl Acad. Sci. USA* **109**, 15835–15840 (2012).
39. van Kleunen, M. et al. Economic use of plants is key to their naturalization success. *Nat. Commun.* **11**, 3201 (2020).
40. Molina-Venegas, R., Rodríguez, M., Pardo-de-Santayana, M., Ronquillo, C. & Mabblerley, D. J. Maximum levels of global phylogenetic diversity efficiently capture plant services for humankind. *Nat. Ecol. Evol.* **5**, 583–588 (2021).
41. Molina-Venegas, R. Conserving evolutionarily distinct species is critical to safeguard human well-being. *Sci. Rep.* **11**, 24187 (2021).
42. Zaman, W. et al. Predicting potential medicinal plants with phylogenetic topology: inspiration from the research of traditional Chinese medicine. *J. Ethnopharmacol.* **281**, 114515 (2021).
43. Cámara-Leret, R. et al. Climate change threatens New Guinea's biocultural heritage. *Sci. Adv.* **5**, eaaz1455 (2019).
44. Lima, V. P. et al. Climate change threatens native potential agroforestry plant species in Brazil. *Sci. Rep.* **12**, 2267 (2022).
45. Johnson, D. V. *Tropical Palms 2010 Revision Non-Wood Forest Products* 10 (FAO, 2010).
46. Johnson, D. V. & Sunderland, T. C. H. *Rattan Glossary and Compendium Glossary with Emphasis on Africa Non-Wood Forest Products* 16 (FAO, 2004).
47. Ter Steege, H. et al. Hyperdominance in the Amazonian tree flora. *Science* **342**, 1243092 (2013).
48. Zona, S. & Henderson, A. A review of animal-mediated seed dispersal of palms. *Selbyana* **11**, 6–21 (1989).
49. Kissling, W. D. et al. PalmTraits 1.0, a species-level functional trait database of palms worldwide. *Sci. Data* **6**, 178 (2019).
50. Tomlinson, P. B. The uniqueness of palms. *Bot. J. Linn. Soc.* **151**, 5–14 (2006).
51. Diaz, S. et al. The global spectrum of plant form and function. *Nature* **529**, 167–171 (2016).
52. Muscarella, R. et al. The global abundance of tree palms. *Glob. Ecol. Biogeogr.* **29**, 1495–1514 (2020).
53. Dransfield, J. et al. *Genera Palmarum: The Evolution and Classification of Palms* (Kew Publishing, 2008).
54. Diazgranados, M. et al. *World Checklist of Useful Plant Species* (Royal Botanic Gardens, Kew, 2020).
55. Couvreur, T. L. P. & Baker, W. J. Tropical rain forest evolution: palms as a model group. *BMC Biol.* **11**, 2–5 (2013).
56. Faurby, S., Eisehardt, W. L., Baker, W. J. & Svenning, J. Molecular phylogenetics and evolution: an all-evidence species-level supertree for the palms (Arecaceae). *Mol. Phylogenet. Evol.* **100**, 57–69 (2016).
57. *The IUCN Red List of Threatened Species Version 2021-2* (IUCN, accessed 12 October 2021); <https://www.iucnredlist.org>
58. Baker, W. J. & Dransfield, J. Beyond general Palmarum: progress and prospects in palm systematics. *Bot. J. Linn. Soc.* **182**, 207–233 (2016).
59. Henderson, A. A revision of *Calamus* (Arecaceae, Calamoideae, Calameae, Calaminae). *Phytotaxa* <https://doi.org/10.11646/phytotaxa.445.1.1> (2020).
60. Rakotoarinivo, M., Dransfield, J., Bachman, S. P., Moat, J. & Baker, W. J. Comprehensive red list assessment reveals exceptionally high extinction risk to Madagascar palms. *PLoS ONE* **9**, e103684 (2014).
61. Cosiaux, A. et al. Low extinction risk for an important plant resource: conservation assessments of continental African palms (Arecaceae/Palmae). *Biol. Conserv.* **221**, 323–333 (2018).
62. Johnson, D. & UICN/SSC Palm Specialist Group (eds) *Palms, Their Conservation and Sustained Utilization—Status Survey and Conservation Action Plan* (Union Internationale pour la Conservation de la Nature et des Ressources, 1996).
63. Bachman, S., Walker, B. E., Barrios, S., Copeland, A. & Moat, J. Rapid least concern: towards automating red list assessments. *Biodivers. Data J.* **8**, e47018 (2020).
64. Enquist, B. J. et al. The commonness of rarity: global and future distribution of rarity across land plants. *Sci. Adv.* <https://doi.org/10.1126/sciadv.aaz0414> (2019).
65. Vieilledent, G. et al. Combining global tree cover loss data with historical national forest cover maps to look at six decades of deforestation and forest fragmentation in Madagascar. *Biol. Conserv.* **222**, 189–197 (2018).

66. Gaveau, D. L. A. et al. Rise and fall of forest loss and industrial plantations in Borneo (2000–2017). *Conserv. Lett.* **12**, e12622 (2019).
67. Gamoga, G., Turia, R., Abe, H., Haraguchi, M. & Iuda, O. The forest extent in 2015 and the drivers of forest change between 2000 and 2015 in Papua New Guinea: deforestation and forest degradation in Papua New Guinea. *Case Stud. Environ.* **5**, 1442018 (2021).
68. Cámara-Leret, R. & Bascompte, J. Language extinction triggers the loss of unique medicinal knowledge. *Proc. Natl Acad. Sci. USA* **118**, e2103683118 (2021).
69. Henderson, A., Fischer, B., Scariot, A., Whitaker Pacheco, M. A. & Pardini, R. Flowering phenology of a palm community in a central Amazon forest. *Brittonia* **52**, 149–159 (2000).
70. Olivares, I. & Galeano, G. Leaf and inflorescence production of the wine palm (*Attalea butyracea*) in the dry Magdalena river valley, Colombia. *Caldasia* **35**, 37–48 (2013).
71. Voeks, R. A. Disturbance pharmacopoeias: medicine and myth from the humid tropics. *Ann. Assoc. Am. Geogr.* **94**, 868–888 (2004).
72. Pironon, S. et al. Potential adaptive strategies for 29 sub-Saharan crops under future climate change. *Nat. Clim. Change* **9**, 758–763 (2019).
73. Govaerts, R., Dransfield, J., Zona, S. & Henderson, A. *World Checklist of Areaceae* (Royal Botanic Gardens, Kew, accessed 1 March 2018); <http://wmsp.science.kew.org/>
74. Chamberlain, S. et al. rgbif: Interface to the Global Biodiversity Information Facility API. R package version 3.6.0 (2021).
75. Zizka, A. et al. CoordinateCleaner: standardized cleaning of occurrence records from biological collection databases. *Methods Ecol. Evol.* **10**, 744–751 (2019).
76. *Plants of the World Online* (Royal Botanic Gardens, Kew, accessed 1 March 2018); <http://www.plantsoftheworldonline.org/>
77. South, A. rworldmap v.1.3-6: Mapping global data (2016).
78. Bivand, R. et al. maptools v.0.9-2: Tools for handling spatial objects (2017).
79. Arel-Bundock, V., Enevoldsen, N. & Yetman, C. countrycode: an R package to convert country names and country codes. *J. Open Source Softw.* **3**, 848 (2018).
80. Becker, R. A., Wilks, A. R., Brownrigg, R., Minka, T. P. & Deckmyn, A. maps v.3.3.0: Draw geographical maps (2018).
81. Pebesma, E. et al. sp v.1.2-7: Classes and methods for spatial data (2018).
82. Wickham, H. et al. Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
83. Wickham, H., Hester, J. & Chang, W. devtools v.1.13.5: Tools to make developing R packages easier (2018).
84. *World Geographic Scheme for Recording Plant Distributions Standard* (TDWG, 2001); <http://www.tdwg.org/standards/109>
85. Brummitt, R. K. *World Geographical Scheme for Recording Plant Distributions* (Hunt Institute for Botanical Documentation, 2001).
86. Olson, D. M. et al. Terrestrial ecoregions of the world: a new map of life on Earth. *Bioscience* **51**, 933–938 (2001).
87. Moat, J. & Bachman, S. P. rCAT v.0.1.6: Conservation assessment tools (2017).
88. Dinerstein, E. et al. An ecoregion-based approach to protecting half the terrestrial realm. *Bioscience* **67**, 534–545 (2017).
89. *Plants of the World Online* (Royal Botanic Gardens, Kew, accessed 10 June 2020); <http://www.plantsoftheworldonline.org/>
90. Csárdi, G. & FitzJohn, R. progress v.1.2.2: Terminal progress bars (2019).
91. Microsoft Corporation & Weston, S. doParallel: Foreach parallel adaptor for the 'parallel' package. R package version 1.0.16 (2020).
92. Microsoft Corporation & Weston, S. foreach: Provides foreach looping construct. R package version 1.5.0 (2020).
93. Ooms, J., Lang, D. T. & Hilal, J. jsonlite v.1.7.2: A simple and robust JSON parser and generator for R (2020).
94. Wickham, H. httr v.1.4.2: Tools for working with URLs and HTTP (2020).
95. *Global Human Footprint (Geographic)*, v2 (1995–2004) (SEDAC, accessed 14 May 2018); <https://doi.org/10.7927/H4M61H5F>
96. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **37**, 4302–4315 (2017).
97. Hansen, M. C. et al. High-resolution global maps of 21st-century forest cover change. *Science* **342**, 850–853 (2013).
98. Wickham, H. plyr v.1.8.6: Tools for splitting, applying and combining data (2021).
99. Wickham, H. & RStudio. tidy v.1.1.4: Tidy messy data (2021).
100. Wickham, H., François, R., Henry, L. & Müller, K. dplyr v.1.0.7: A grammar of data manipulation (2021).
101. Bivand, R. et al. rgdal v.1.5-8: Bindings for the 'geospatial' data abstraction library (2020).
102. Greenberg, J. A. & Mattiuzzi, M. gdalUtils v.2.0.3.2: Wrappers for the Geospatial data Abstraction Library (GDAL) utilities (2020).
103. Hijmans, R. J. et al. raster v.3.1-5: Geographic data analysis and modeling (2020).
104. *The IUCN Red List of Threatened Species* (IUCN, accessed 22 March 2018); <https://www.iucnredlist.org/>
105. *ThreatSearch Online Database* (Botanic Gardens Conservation International, accessed 1 March 2018); https://tools.bgci.org/threat_search.php
106. Chamberlain, S., ROpenSci & Salmon, M. rredlist: 'IUCN' Red List client (2020).
107. Wickham, H. stringr v.1.4.0: Simple, consistent wrappers for common string operations (2019).
108. Gagolewski, M. & Tartanus, B. stringi v.1.7.5: Character string processing facilities (2021).
109. Kuhn, M. caret: Classification and regression training. R package version 6.0-86 (2020).
110. Torgo, L. *Data Mining with R, Learning with Case Studies* (Chapman and Hall/CRC, 2010).
111. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, P. SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357 (2020).
112. Stokely, M. HistogramTools: Utility functions for R histograms. R package version 0.3.2 (2015).
113. Sarkar, D. et al. lattice v.0.20-40: Trellis graphics for R (2020).
114. Wickham, H. *ggplot2 Elegant Graphics for Data Analysis* (Springer, 2016).
115. Auguie, B. & Antonov, A. gridExtra v.2.3: Miscellaneous functions for 'grid' graphics (2017).
116. Pruim, R., Kaplan, D. T. & Horton, N. J. mosaic v.1.6.0: Project MOSAIC statistics and mathematics teaching utilities (2020).
117. Meyer, D. & Buchta, C. proxy v.0.4-23: Distance and similarity measures (2019).
118. Wickham, H. & Seidel, D. scales v.1.1: Scale functions for visualization (2019).
119. Branco, P., Ribeiro, R. & Torgo, L. UBL v.0.0.6: An implementation of re-sampling approaches to utility-based learning for both classification and regression tasks (2017).
120. Liaw, A. & Wiener, M. Classification and regression by randomForest. *R News* **2**, 18–22 (2002).
121. Cohen, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **20**, 37–46 (1960).
122. Ripley, B. & Venables, W. nnet v.7.3-13: Feed-forward neural networks and multinomial log-linear models (2020).
123. Warnes, G. R. et al. gdata v.2.18.0: Various R programming tools for data manipulation (2017).
124. Wright, M. N., Wager, S. & Probst, P. ranger v.0.12.1: A fast implementation of random forests (2020).
125. Arya, S., Mount, D., Kemp, S. E. & Jefferis, G. RANN v.2.6.1: Fast nearest neighbour search (wraps ANN Library) using L2 metric (2019).
126. Meyer, D. et al. e1071 v.1.7-3: Misc Functions of the Department of Statistics, Probability Theory Group (formerly: E1071), TU Wien (2019).
127. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **30**, 4765–4774 (2017).
128. Greenwell, B. fastshap v.0.0.7: Fast approximate Shapley values (2021).
129. Greenwell, B. vip v.0.3.2: Variable importance plots (2020).
130. Donoghoe, M. W. glm2 v.1.2.1: Fitting generalized linear models (2018).
131. Wickham, H. reshape2 v.1.4.4: Flexibly reshape data: a reboot of the reshape package (2020).
132. Robin, X. et al. pROC v.1.18.0: Display and analyze ROC curves (2020).
133. Warnes, G. R. et al. gplots v.3.0.3: Various R programming tools for plotting data (2019).
134. Müller, K. & Bryan, J. here v.1.0.1: A simpler way to find your files (2017).
135. Wickham, H., Hester, J., François, R., Jylänki, J. & Jørgensen, M. readr v.1.3.1: Read rectangular text data (2018).
136. Wickham, H. et al. readxl v.1.3.1: Read Excel files (2019).
137. Henry, L. & Wickham, H. purrr v.0.3.4: Functional programming tools (2020).
138. Lin Pedersen, T. ggforce v.0.3.1: Accelerating 'ggplot2' (2019).
139. Lin Pedersen, T. patchwork v.1.0.0: The composer of plots (2019).
140. Hester, J. glue v.1.3.1: Interpreted string literals (2019).
141. Ooms, J. & McNamara, J. writexl v.1.2: Export data frames to Excel 'xlsx' format (2019).
142. Horikoshi, M. et al. ggfortify v.0.4.8: Data visualization tools for statistical analysis results (2019).
143. Liaw, A. randomForest v.4.6-14: Breiman and Cutler's random forests for classification and regression (2018).
144. Kassambara, A. ggpubr v.0.2.5: 'ggplot2' based publication ready plots (2020).
145. Gruca, M., Blach-Overgaard, A. & Balslev, H. African palm ethno-medicine. *J. Ethnopharmacol.* **165**, 227–237 (2015).
146. Cámara-Leret, R. & Dennehy, Z. Indigenous knowledge of New Guinea's useful plants: a review. *Econ. Bot.* **73**, 405–415 (2019).
147. Macia, M. J. et al. Palm uses in Northwestern South America: a quantitative review. *Bot. Rev.* **77**, 462–570 (2011).

148. Orme, D. et al. caper: Comparative analyses of phylogenetics and evolution in R. R package version 1.0.1 <https://cran.r-project.org/package=caper> (2018).
149. Kowarik, A. & Templ, M. Imputation with the R package VIM. *J. Stat. Softw.* **74**, 1–16 (2016).
150. Alfons, A. & Templ, M. Estimation of social exclusion indicators from complex surveys: the R package laeken. *J. Stat. Softw.* **54**, 1–25 (2013).
151. Milliken, W., Walker, B. E., Howes, M. J. R., Forest, F. & Nic Lughadha, E. Plants used traditionally as antimalarials in Latin America: mining the tree of life for potential new medicines. *J. Ethnopharmacol.* **279**, 114221 (2021).
152. Fritz, S. A. & Purvis, A. Selectivity in mammalian extinction risk and threat types: a new measure of phylogenetic signal strength in binary traits. *Conserv. Biol.* **24**, 1042–1051 (2010).
153. Suchard, M. A. et al. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**, vey016 (2018).
154. Paradis, E. & Schliep, K. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
155. Govaerts, R., Nic Lughadha, E., Black, N., Turner, R. & Paton, A. The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. *Sci. Data* **8**, 215 (2021).
156. Yu, G. ggplotify v.0.0.4: Convert plot to 'grob' or 'ggplot' object (2019).
157. Yu, G. aplot v.0.0.3: Decorate a 'ggplot' with associated information (2020).
158. Slowikowski, K. et al. ggrepel v.0.8.1: Automatically position non-overlapping text labels with 'ggplot2' (2019).
159. Schloerke, B. et al. GGally v.1.4.0: Extension to 'ggplot2' (2018).
160. Rubis, B. et al. hrbrthemes v.0.6.0: Additional themes, theme components and utilities for 'ggplot2' (2019).
161. Henry, L., Wickham, H. & Chang, W. ggstance v.0.3.3: Horizontal 'ggplot2' components (2019).
162. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. Y. Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
163. Brown, C. hash v.2.2.6.1: Full feature implementation of hash/associated arrays/dictionaries (2019).
164. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, 2016).
165. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2020).
166. RStudio Team. *RStudio: Integrated Development for R* (RStudio, 2021).
167. Bellot, S. et al. Workflow and code used to perform palm extinction risk and regional palm use resilience analyses. *Zenodo* <https://doi.org/10.5281/zenodo.6678122> (2022).

Acknowledgements

We are grateful to Naturalis Biodiversity Center (Leiden, the Netherlands) and Kew Herbarium (United Kingdom) for sharing their geographic coordinates databases,

to O. A. Pérez-Escobar for insightful discussion and to T. Couvreur for insightful discussion and critical reading of an earlier version of the manuscript. A.A. is supported by funding from the Swedish Research Council and the Royal Botanic Gardens, Kew. W.D.K. acknowledges funding of palm research from the Netherlands Organisation for Scientific Research (grant no. 824.15.007).

Author contributions

S.B. and S.P.B. conceived the study, with input from Y.L., R.C.-L. and W.J.B. Y.L. compiled most of the data, did preliminary random forest ML analyses and wrote part of an early draft, with input and training from S.B. and S.P.B. S.B. compiled traits and use data with help from R.C.-L., I.O. and S.P. I.O. and S.P.B. provided and ran scripts to retrieve numbers of ecoregions and TDWG3 regions. J.D. provided a preliminary guess of the extinction risk of ~400 species without occurrence data or extinction risk assessment. Y.L. and S.B. cleaned the data. S.B. did all final random forest analyses, all neural network analyses and all analyses using the predictions, except the SHAP analysis, done by B.E.W., who also wrote the corresponding Supplementary Note and methods sections. R.C.-L., F.F., I.O., S.P.B. and S.P. advised on some analyses. S.B. wrote most of the final manuscript, with extensive input from S.P.B. and R.C.-L. and contributions from all authors. A.A., W.J.B., R.C.-L., J.D., F.F., W.D.K., I.J.L., E.N.L., I.O. and S.P. provided transformative feedback.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41559-022-01858-0>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41559-022-01858-0>.

Correspondence and requests for materials should be addressed to S. Bellot or R. Cámara-Leret.

Peer review information *Nature Ecology & Evolution* thanks Tinde van AnDEL, Rafael Molina-Venegas and Danilo Neves for their contribution to the peer review of this work. Peer reviewer reports are available.

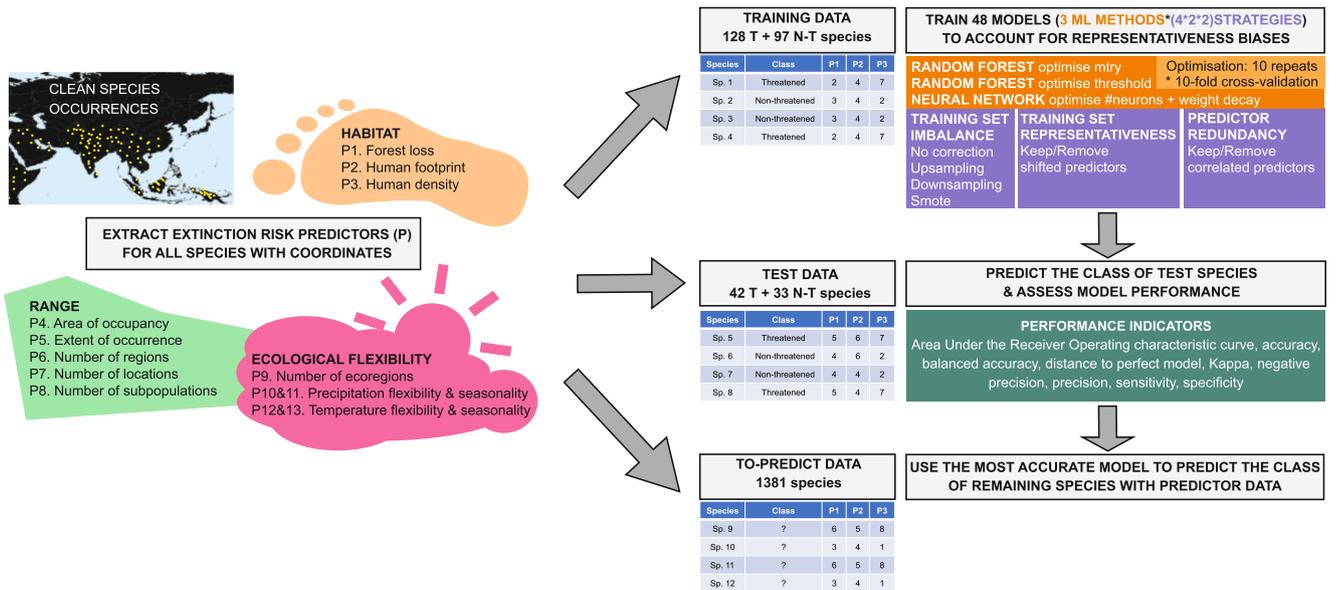
Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2022

a



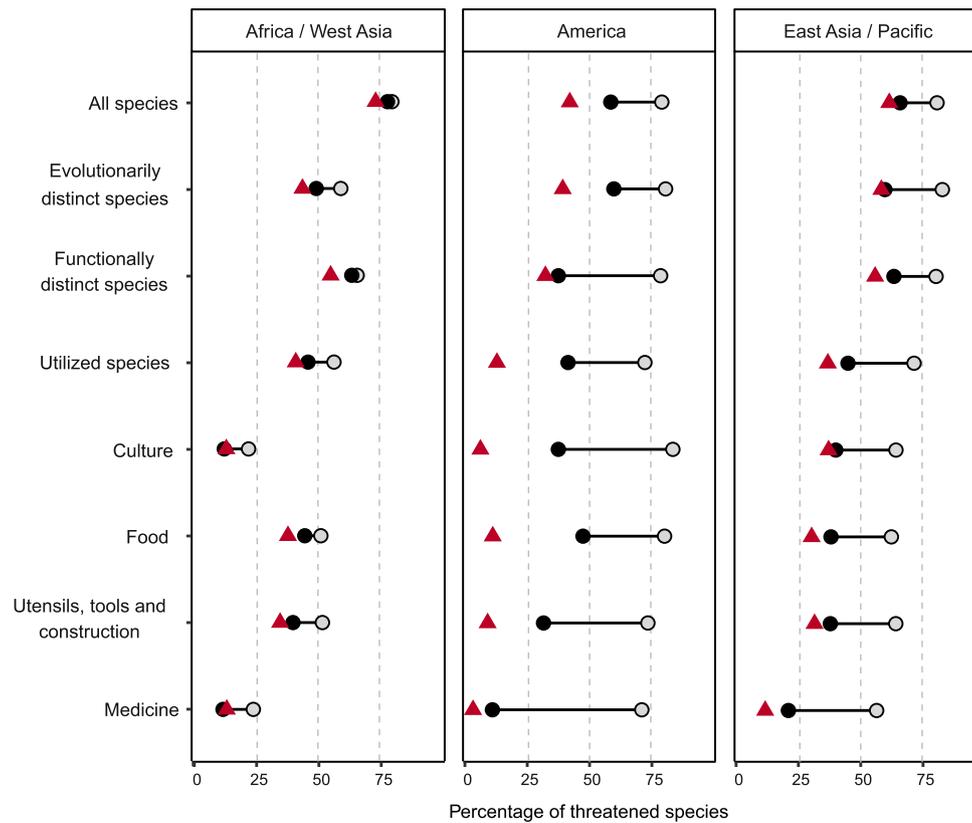
b

All species (2510)											
Assessed after 2008 (508)			Assessed before 2008 (627)			Non-assessed (1375)					
Representative (300)	Dyopsis (139)	No occurrence (69)	No occurrence* (226)	≥1 occurrence (401)	≥1 occurrence (980)		No occurrence (395)				
Use for this study											
Training (225)		Test (75)		No machine learning			To be predicted			No machine learning	
T (128)	N-T (97)	T (39)	N-T (36)								
Total Evidence dataset											
Recently published (508)			Not Included		Predicted (1381)				Not Included		
Other datasets											
All published (1134)					Not Included						
Recently published (508)			Not Included								
Published (733)			Predicted (1381)				Expert guess (296)		NI (99)		

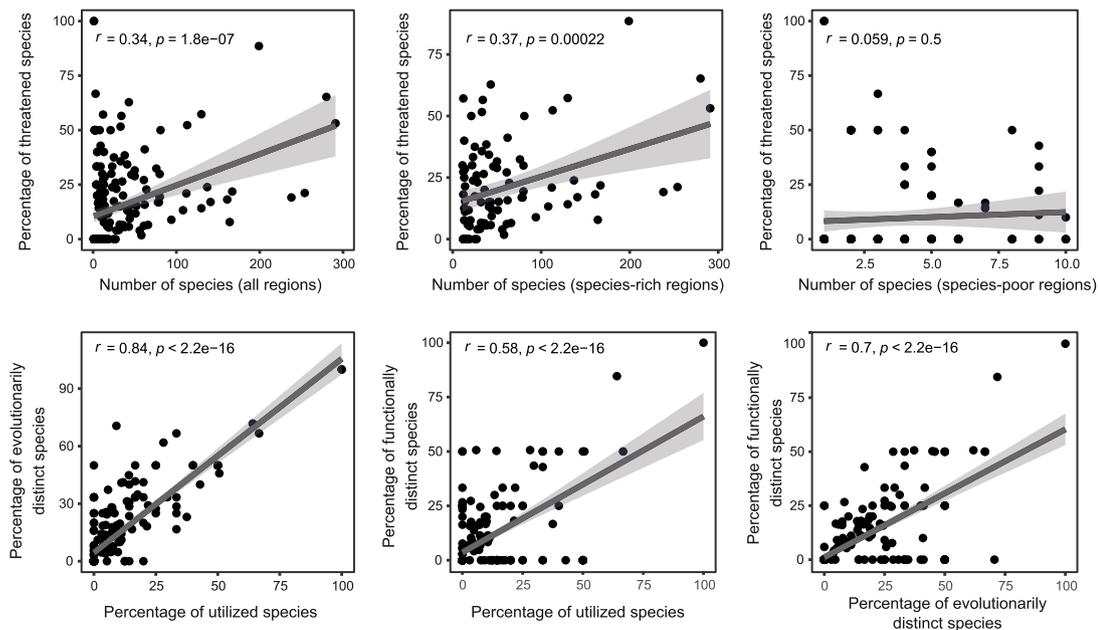
*One of these species (*Hyophorbe americana*) was excluded from downstream analyses because it is considered extinct in the wild.

Extended Data Fig. 1 | Study design. a. Model training and testing. b. Datasets. NI: not included; N-T: non-threatened; T: threatened. See Methods for further details.

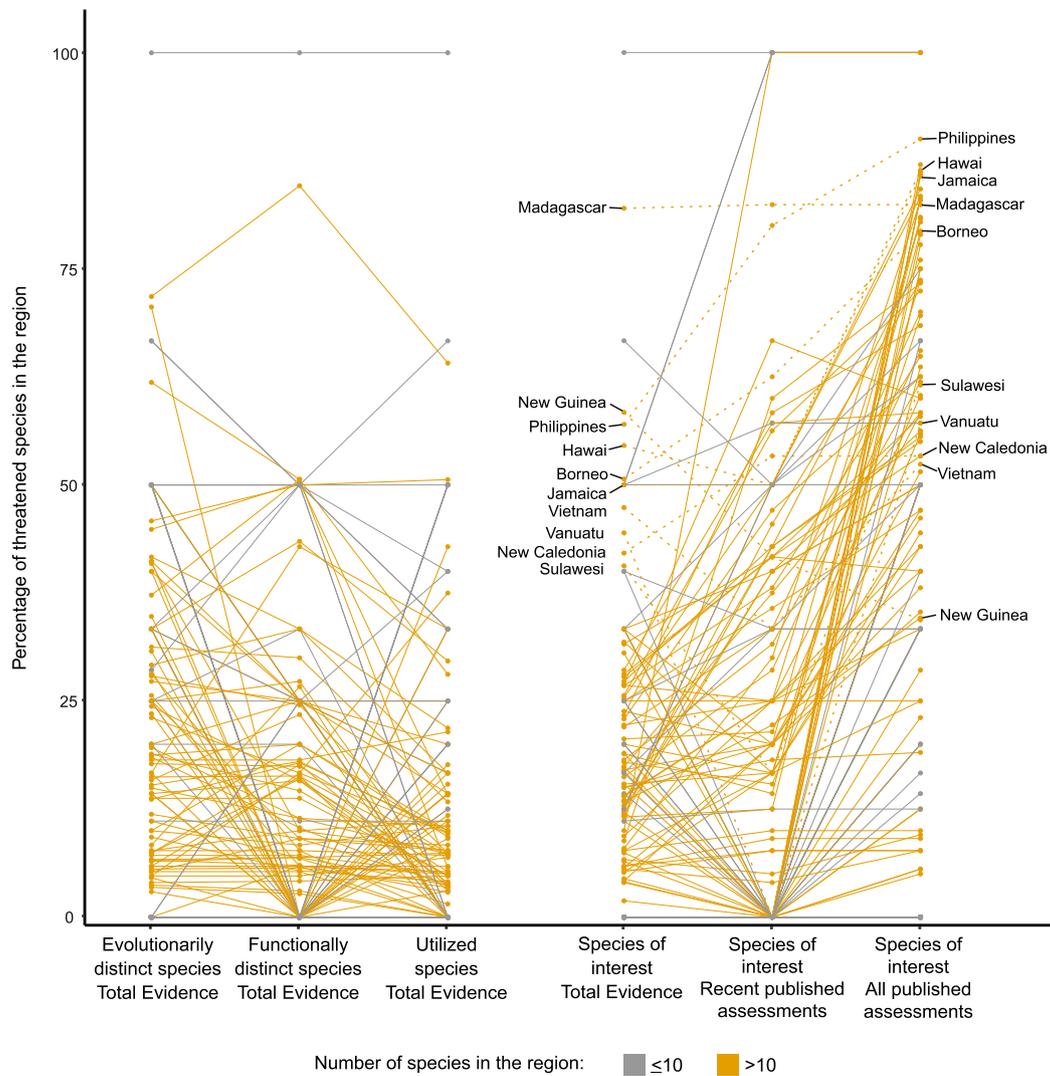
a



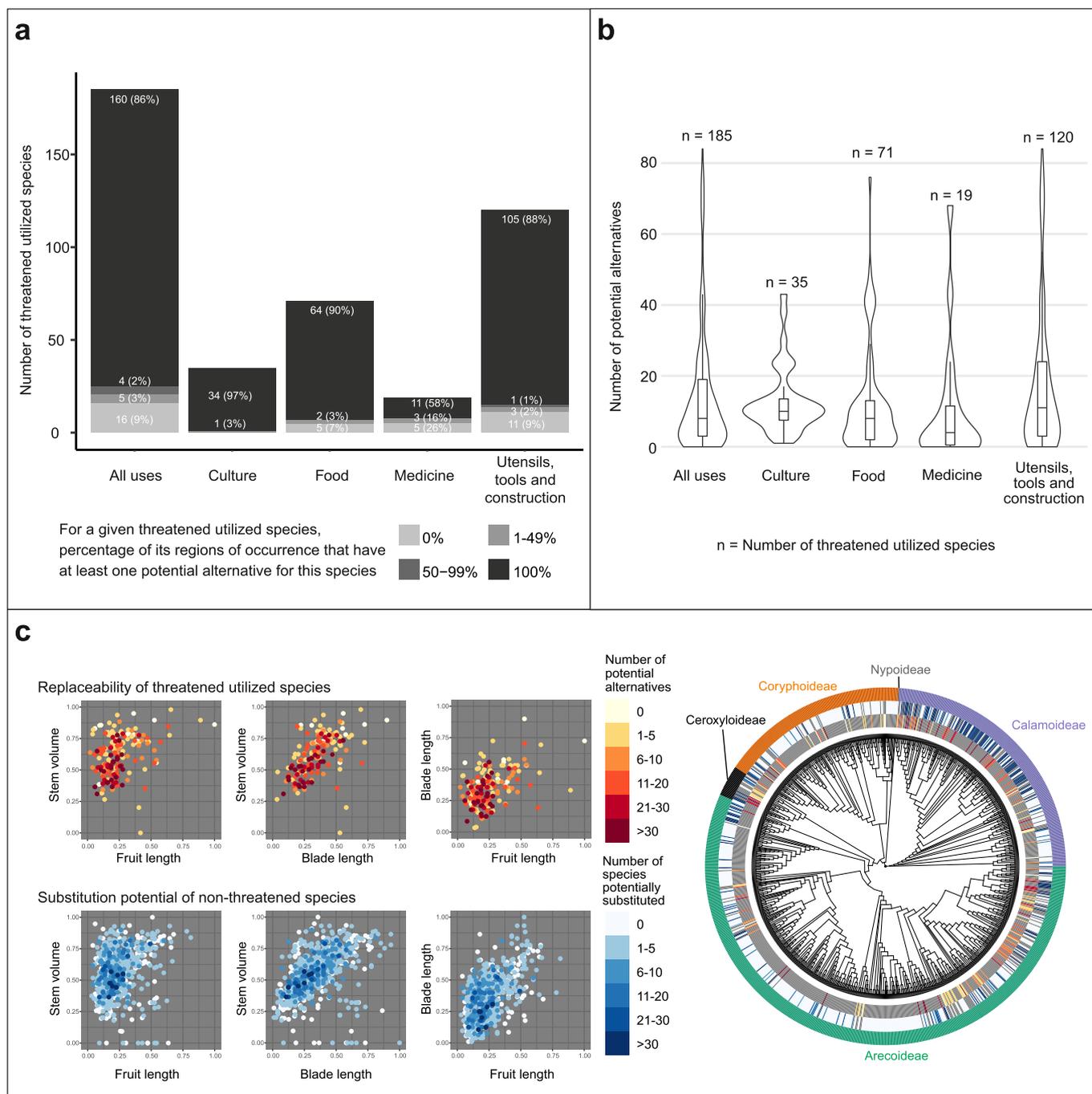
b



Extended Data Fig. 2 | Threatened palm species in different diversity and use categories. a. Comparison between global estimates obtained with or without ML predictions in different areas of the world. Triangles: Estimates obtained by combining recent published assessments and ML predictions from the most accurate model (total evidence approach). Circles: Extrapolations based on published assessments only. Grey: all published assessments; Black: Only recent (that is ≤ 10 -year-old) published assessments. America spans longitudes $\in [-180^\circ, -25^\circ]$, Africa / West Asia spans longitudes $\in [-25^\circ, 68^\circ]$, and East Asia / Pacific spans longitudes $\in [68^\circ, 180^\circ]$. **b. Relationships between regional percentage of threatened species and total species number in the region (top) and between regional percentage of threatened species in different diversity and use categories (bottom).** Percentages were obtained following the total evidence approach. Each dot represents a region with at least one palm species with data in the considered categories ($n = 225, 93, 132, 224, 218$ and 218 from top to bottom and left to right). The associations between the variables were measured using Pearson's product moment correlation coefficient in two-sided Pearson's correlation tests. r is Pearson's correlation coefficient and p is the associated p -value. The grey error band corresponds to the 95% confidence interval of the correlation coefficient.

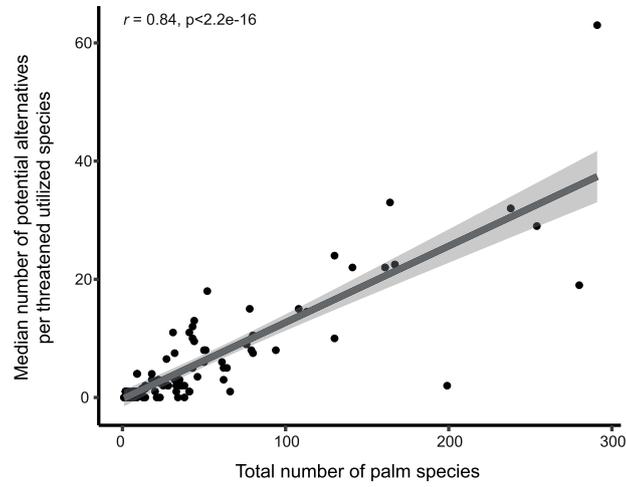


Extended Data Fig. 3 | Changes in regional percentage of threatened species and associated region rank depending on the diversity category and the extinction risk information used. 'total evidence' is the combination of recent published extinction risk assessments and most accurate machine learning predictions. Each line corresponds to a region, and lines are coloured by the total number of palm species in the region. Regions with $\geq 40\%$ threatened species of interest (that is top priorities) according to the total evidence approach are annotated, and the lines linking estimates from the total evidence approach and estimates based only on published assessments for these regions are thicker and dotted.

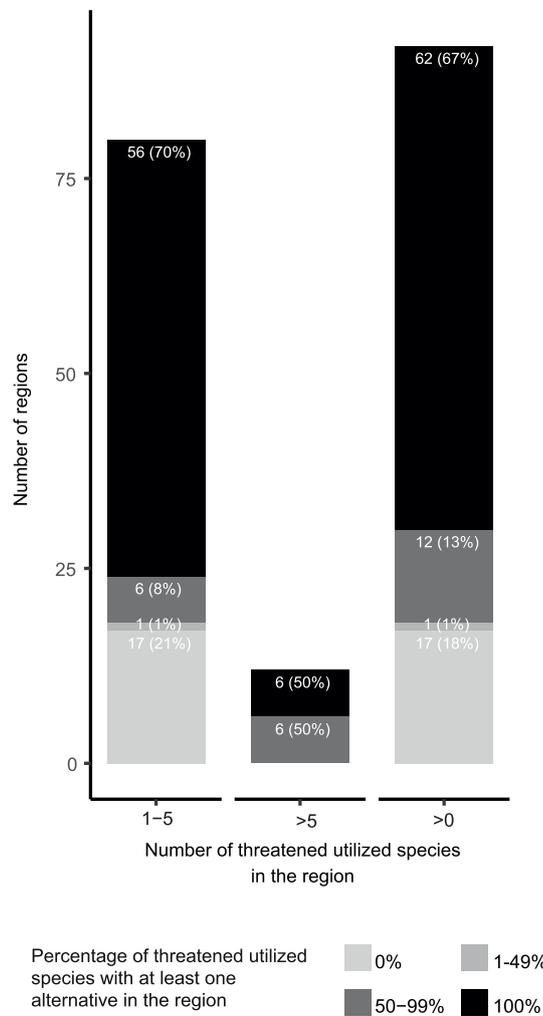


Extended Data Fig. 4 | Replaceability of threatened utilized species and potential for substitution of non-threatened species. The replaceability of a species is defined as the number of potential alternatives identified for that species. The potential for substitution of a species is defined as the number of threatened utilized species that may be substituted by that species. **a. Regional replaceability of threatened utilized species across use categories.** **b. Global replaceability of threatened utilized species across use categories.** The ‘violins’ represent the kernel density distributions. In a violin, the bold line represents the median value, the box spans values from the first to the third quartile, and the lines outside the box extend until the smallest and largest values, no further than 1.5 times the distance between the first and third quartiles. **c. Functional traits and phylogenetic distribution of species replaceability and potential for substitution, for all uses taken together.** The phylogenetic tree displayed was obtained by removing species without data from a maximum clade credibility consensus tree summarizing 750 palm phylogenetic trees estimated by Faurby et al.⁵⁶ (see Methods and Supplementary Data).

a



b



Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Regional use resilience for all uses taken together. a. Correlation between median number of alternatives per threatened utilized species and total palm species richness in the region. Each dot represents a region with at least one threatened utilized palm species ($n=92$). The association between the variables was measured using Pearson's product moment correlation coefficient in a two-sided Pearson's correlation test. r is Pearson's correlation coefficient and p is the associated p -value. The grey error band corresponds to the 95% confidence interval of the correlation coefficient. **b. Regional percentages of threatened utilized species with alternatives.**

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- | | | |
|-------------------------------------|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | A description of all covariates tested |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> | Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	Data collection was performed in RStudio, using many different R packages and custom scripts. All R packages and their reference are cited in the Methods section, and the version numbers are cited in the text. All custom scripts are available on Zenodo (doi: https://doi.org/10.5281/zenodo.6678122)
Data analysis	Data analysis was performed in RStudio, using many different R packages and custom scripts. All R packages and their reference are cited in the Methods section, and the version numbers are cited in the text. All custom scripts are available on Zenodo (doi: https://doi.org/10.5281/zenodo.6678122)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All data necessary to perform the analyses are provided as Supplementary Tables or on Zenodo (doi:<https://doi.org/10.5281/zenodo.6678122>)

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	The study consists in 2 main parts: 1) using existing published extinction risk assessments to train machine learning models to predict the extinction risk of non assessed species. 48 models are trained, which use different parameter values accounting for data unbalance (described in methods). The model with the highest accuracy is chosen to make final predictions. These predictions are then used to assess how evolutionary, functional and use diversity are threatened and to identify regions with higher threat levels. 2) The potential of a non-threatened species to replace a threatened utilized species in a given region is assessed. This approach relies on plant use, phylogenetic, functional and biome datasets previously published. This is then used to identify regions of lower resilience, i.e. where threatened utilized species are unlikely to be replaced.
Research sample	The study focused on all palm species (Arecaceae) and machine learning was used on species for which we had at least one spatial occurrence data point (1889 species). Datasets are described in Supplementary Figure 1 and sampled species are listed in Supplementary Table 5
Sampling strategy	The main aspects of the study do not rely on statistical tests where sample size matters. The study focused on as many species as possible (i.e. all species for which we had data, see above). Statistical tests were used in 3 occasions for method design and method robustness assessment: <ul style="list-style-type: none"> - we tested if extinction risk predictors were correlated to each other in order to use models including or excluding correlated predictors. This was done on all 1889 species, so the sampling size was sufficient. - we tested if the number of occurrences available for a species was correlated with the prediction success. This was done on test species only. Sample size ranged from 36 to 75 depending on the test (see Supplementary Note). This is sufficient for basic correlation and logistic regression tests and was only used to get an indication of how our predictions may be influenced by number of spatial occurrences, but not to draw any interpretations regarding palms in general. - we looked at the phylogenetic signal in palm use data using Fritz & Purvis D statistic. This was done at the scale of the family, and sample size corresponded to all species available in the tree we used, which is a tree including most palm species (ca. 2400 species after taxonomic reconciliation with our dataset). Sample size is sufficient according to the original publication describing the D statistic, and it could not be higher as it already includes all species.
Data collection	Data were not newly collected but compiled from existing datasets (Global Biodiversity Information Facility, PalmTraits database previously published by Kissling et al., 2019, palm use data from the World Checklist of Useful Plant Species of Diazgranados et al., 2020, IUCN Red List and ThreatSearch databases). Data were mostly compiled by authors S. Bellot and Y. Lu.
Timing and spatial scale	Spatial occurrence data and extinction risk assessment data were retrieved in spring 2018, when the study started. Other data were compiled in 2020.
Data exclusions	Some analyses were performed with and without >10-year old extinction risk assessments, to explore the effect of keeping/excluding assessments considered as outdated by the IUCN. Spatial occurrence data identified as erroneous (see Methods) were excluded.
Reproducibility	There were no experiments as defined here. The code used is described and provided in Github, so that the analysis can be reproduced.
Randomization	There were no statistics requiring comparisons between experimentally-defined groups.
Blinding	Blinding was not relevant for our study as data were not newly collected.
Did the study involve field work?	<input type="checkbox"/> Yes <input checked="" type="checkbox"/> No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- | n/a | Involvement in the study |
|-------------------------------------|--|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Human research participants |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |

Methods

- | n/a | Involvement in the study |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |