Effective assessments of the current status of biodiversity conservation are needed to support planning, policy and action, from local to global levels. Of particular use would be well documented, reproducible methods based on openly accessible data and tools. Such methods should provide an accurate estimate of the state of conservation of diversity, identifying gaps in current conservation systems, while providing a benchmark against which to measure success, including determining when conservation goals have been met. Here we introduce GapAnalysis, an open source R package developed to support the assessment of the state of conservation of taxa, both ex situ (in genebanks, botanic gardens and other repositories) and in situ (in protected natural areas). The eco-geographic variation maintained in conservation repositories and that present within protected natural areas is compared with the full extent of eco-geographic variation predicted within species’ native ranges. Eco-geographic conservation gaps are identified, and taxa are then prioritized for further action. The package enables users to quickly determine conservation status, gaps and priorities for as few as one to as many as thousands of taxa, for any wild flora or fauna for which occurrence data and species distribution models are available. To demonstrate the use of GapAnalysis, we provide case studies for wild plant taxa in the genera \textit{Capsicum} L. and \textit{Cucurbita} L, with an accompanying online R tutorial for three of the \textit{Cucurbita} species.

Keywords: biodiversity indicators, crop wild relatives, ex situ conservation, gap analysis, in situ conservation
Introduction

Ongoing global biodiversity loss represents the sixth major extinction event in Earth’s history, and the first to be driven primarily by human activities (Chapin et al. 2000), mainly due to increasing societal demand for food, water and natural resources (Millennium Ecosystem Assessment 2005). As an earth system process of critical concern, the current rate of decline in biodiversity is considered well outside the ‘safe operating space’ for humanity’s survival (Rockström et al. 2009).

As a result, specific goals and associated targets for biodiversity conservation are now included in major international agreements, including the United Nations Sustainable Development Goals (SDGs) (United Nations 2015); the Convention on Biological Diversity (CBD) Strategic Plan for Biodiversity 2011–2020, Aichi Biodiversity Targets (CBD 2010a), Global Strategy for Plant Conservation (GSPC) (CBD 2010b) and upcoming Post-2020 Biodiversity Framework (CBD 2020); the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) (CITES 2020); and the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA) (FAO 2002).

A wide range of indicators have been proposed, are under development, or have been deployed to measure progress toward these targets. While these are generally pragmatic and complementary (CBD 2018), additional metrics are needed to better understand the current status of conservation of taxa in the context of their overall extant diversity (Khoury et al. 2019b). Furthermore, indicators enabling a more comprehensive assessment of the degree of conservation of genetic diversity are required, a particular challenge to accomplish across multiple taxonomic groups due to a widespread lack of genetic data (Balmford et al. 2005, Hoban et al. 2020b).

To address persisting biodiversity conservation indicator needs, particularly with regard to efficiently assessing conservation of genetic diversity within and across taxa, Khoury et al. (2019b) offered a gap analysis methodology applied to both ex situ and in situ conservation systems. The method provided an approximation of the distribution of a wild plant species’ genetic diversity, using the extent of eco-geographic (i.e. geographic and ecological) variation in its predicted native range as a proxy, which has been shown to be an effective surrogate (Hanson et al. 2017, Hoban et al. 2018). The eco-geographic variation evident from an analysis of the locations of the ‘site of collection’ of samples safeguarded in conservation repositories (i.e. genebanks, seedbanks and botanic gardens) (ex situ), and that evident in the species’ range distributed within protected natural areas (in situ), was measured against the eco-geographic variation found within the species’ overall predicted native range. The process identified geographic and ecological gaps in current protection, which may represent focal points for future action. Taxa were then prioritized for further conservation efforts, and scores for multiple taxa combined to provide indicator metrics from local to national, regional and global scales. The methodology relied on openly accessible data and tools (Khoury et al. 2019a) and was reportable in easily digestible summary formats while also providing disaggregated taxon-specific information useful for conservation action. Furthermore, when applied repeatedly over time, results could be used to mark progress toward the goal of comprehensive conservation, including determining when that goal has been met.

The conservation gap analysis was built on methods developed over the past decade, first to measure the conservation status of taxa in repositories and to help guide further collecting efforts aimed at building more diverse ex situ collections (Ramírez-Villegas et al. 2010, Castañeda-Álvarez et al. 2016). Recently, the approach was adapted to measure representation within protected natural areas (Khoury et al. 2019b, c, d, 2020, Lebda et al. 2019, Mezghani et al. 2019, Myrans et al. 2020). Such studies have most often been conducted across a range of species within a genus, although they have also been applied at national (Norton et al. 2017, Khoury et al. 2020) and global (Castañeda-Álvarez et al. 2016, Khoury et al. 2019b) levels for specific groups of plants.

Here we introduce GapAnalysis, an open source R package developed to support the assessment of the degree of protection of taxa in conservation systems, both ex situ and in situ, and to identify gaps in this conservation, based on eco-geographic methods and tools. The package represents the culmination of previous efforts to develop, refine and evaluate conservation gap analysis methods. The package provides a fully documented, tested and streamlined process enabling users to quickly determine conservation status and priorities for as few as one to as many as thousands of taxa, applicable to any wild species for which occurrence data and species distribution models are available (Hoban et al. 2020a). To demonstrate the use of GapAnalysis, we provide case studies for wild plant taxa in the genera Capsicum L. and Cucurbita L.

Features of the package

GapAnalysis was developed in R 3.6.2 (<www.r-project.org>), requiring (≥ 3.5.0). It is available on the Comprehensive R Archive Network (CRAN, <https://cran.r-project.org/package=GapAnalysis>) and GitHub (<https://github.com/CIAT-DAPA/GapAnalysis>). The package is composed of 12 functions placed within four families: pre-analysis, ex situ analysis, in situ analysis and summary evaluations (Table 1, Fig. 1). In short, the pre-analysis compiles necessary spatial layers required for the subsequent conservation gap analysis steps. The ex situ and in situ processes perform the respective conservation strategy gap analyses and produce both quantitative and spatial results. The summary evaluations merge the individual conservation assessments, compile the results across taxa, calculate the indicator and generate a summary html document for each taxon, which can be used to evaluate outputs and aid conservation planning. An additional three functions are internal and support the conservation gap analysis functions.
Both a table (data.frame) of occurrence data and a distribution model (RasterLayer) for each taxon are required inputs to GapAnalysis. The table of occurrence data is a data.frame with four columns: taxon, latitude, longitude and type. The column taxon is the name (often the scientific name) of the species; latitude and longitude are given in decimal degrees (or NA if unknown); and the column type refers to the type of record. The value ‘G’ is used for any ex situ samples (genebank, seedbank or botanical garden), whereas the value ‘H’ is used for any other reference samples (human observations, herbarium samples, etc.). The distribution model is a binary raster map where 0 indicates absence and 1 indicates presence of the taxon.

Pre-analysis

In the pre-analysis process, ecoregions (The Nature Conservancy Geospatial Conservation Atlas 2019) and

---

Figure 1. Diagram of the conservation gap analysis process for *Cucurbita palmata* S. Watson. Input data consist of occurrence data (blue dots) and species distribution models (with shades of orange indicating probabilities of occurrence). Conservation gap analysis results are provided for each metric, along with maps. The maps display areas within the distribution models of the species that are not currently conserved (green areas), as well as the conserved range (purple), the areas of no SDM-predicted occurrence (white) and the areas outside the known native range of the taxon (light grey). The final conservation score (FCS) for each strategy is the average across its three individual scores. The combined final conservation score (FCSc_mean) is the average of the FCSex and FCSin scores. These FCS scores are used to categorize taxa for further action: high priority (HP) (FCS < 25); medium priority (MP) (25 ≤ FCS < 50); low priority (LP) (50 ≤ FCS < 75); and sufficiently conserved (SC) (FCS ≥ 75). The indicator table calculates an overall level of conservation across all taxa, while the SummaryHTML process provides html outputs with interactive quantitative and spatial results per species.
protected areas (IUCN 2019) datasets are retrieved from an open access data repository (<https://dataverse.harvard.edu/dataverse/GapAnalysis>).

**Ex situ conservation gap analysis**

This analysis estimates the degree of representation of taxa, their populations and their underlying genetic diversity in ex situ conservation repositories. Three individual scores are calculated: the sampling representativeness score ex situ (SRSex), the geographical representativeness score ex situ (GRSex) and the ecological representativeness score ex situ (ERSex). The individual scores are averaged to provide a final conservation score ex situ (FCSex). All scores are bound between 0 and 100, with 0 representing an extremely poor state of conservation, and 100 representing comprehensive protection.

The SRSex process assesses the completeness of ex situ conservation collections, calculating the ratio of germplasm accessions (G) available in ex situ repositories to reference (H) records for each taxon, making use of all compiled records, regardless of whether they include coordinates, with an ideal (i.e. comprehensive) conservation ratio of 1:1 (Eq. 1). To calculate these counts, the process uses the internal function ‘OccurrenceCounts’. In this and all subsequent measurements, if no G or H records exist, taxa are automatically considered to be of high priority for further conservation action and assigned a value of 0. If there are more G than H records, SRSex is set to 100.

\[
SRSex = \left[ \frac{\text{number of G records}}{\text{number of H records}} \right] \times 100
\]  

(1)

The GRSex process provides a geographic measurement of the proportion of a species’ predicted range that can be considered to be conserved in ex situ repositories (Eq. 2). The GRSex uses buffers (default 50 km radius) created around each G occurrence with a valid latitude and longitude pair (using the internal function ‘Gbuffer’) to estimate geographic areas already well collected within the predicted habitat of each taxon. It then calculates the proportion of the distribution model covered by these buffers. During this process, an ex situ geographic gap map is also created for each species by mapping only the spatial areas within the distribution model of each taxon which are occupied by ecoregions not represented by G buffers.

\[
GRSex = \left[ \frac{\text{total area(km}^2) \text{ of G buffers}}{\text{total area(km}^2) \text{ of species distribution model}} \right] \times 100
\]  

(2)

The ERSex process provides an ecological measurement of the proportion of a species’ predicted range that can be considered to be conserved in ex situ repositories (Eq. 3). The ERSex calculates the proportion of terrestrial ecoregions (The Nature Conservancy Geospatial Conservation Atlas 2019) represented within the G buffered areas out of the total number of ecoregions occupied by the distribution model. During this process, an ex situ ecological gap map is also created for each species by mapping only the spatial areas within the distribution model of each taxon which are occupied by ecoregions not represented by G buffers.

\[
ERSex = \left[ \frac{\text{number of ecoregions represented within G buffers}}{\text{number of ecoregions represented within species distribution model}} \right] \times 100
\]  

(3)

The FCSex process calculates the average of the three ex situ conservation metrics (Eq. 4). It then assigns each taxon to a priority category based on the FCSex score, with high priority (HP) for further collecting for ex situ conservation assigned when FCSex < 25, medium priority (MP) where 25 ≤ FCSex < 50, low priority (LP) where 50 ≤ FCSex < 75 and sufficiently conserved (SC) for taxa whoseFCSex ≥ 75.

\[
FCSex = \left[ \frac{\text{SRSex} + \text{GRSex} + \text{ERSex}}{3} \right]
\]  

(4)

**In situ conservation gap analysis**

This analysis estimates the degree of representation of taxa, their populations and their underlying genetic diversity within protected natural areas listed in the World Database of Protected Areas (WDPA) (IUCN 2019), including terrestrial and coastal reserves marked as designated, inscribed or established. Three individual scores are calculated: the sampling representativeness score in situ (SRSin), the geographical representativeness score in situ (GRSex) and the ecological representativeness score in situ (ERSin). The scores are averaged to provide a final conservation score in situ (FCSin). All scores are bound between 0 and 100, with 0 representing an extremely poor state of conservation, and 100 representing comprehensive protection.

The SRSin process first removes occurrences that are outside the predicted distribution, and then calculates the proportion of remaining occurrences that fall within a protected area (Eq. 5).

\[
SRSin = \left[ \frac{\text{number of occurrences predicted as true positives falling within a protected area}}{\text{number of occurrences predicted as true positives}} \right] \times 100
\]  

(5)

The GRSex process provides a geographic measurement of the proportion of a species’ predicted range that can be considered
to be conserved in protected areas (Eq. 6). The GRSin compares the area (in km$^2$) of the distribution model located within protected areas versus the total area of the model, considering comprehensive conservation to have been accomplished only when the entire distribution occurs within protected areas. During this process, an in situ geographic gap map is also created for each species by subtracting the protected areas out of the distribution model of each taxon, revealing those areas in the model not currently in protected areas.

$$\text{GRS}_\text{in} = \left( \frac{\text{area (km}^2\text{)} \text{ of species distribution model located within protected areas}}{\text{total area (km}^2\text{)} \text{ of species distribution model}} \right) \times 100 \tag{6}$$

The ERSin process provides an ecological measurement of the proportion of a species' predicted range that can be considered to be conserved in protected areas (Eq. 7). The ERSin calculates the proportion of ecoregions encompassed within the range of the taxon located inside protected areas to the ecoregions encompassed within the total area of the distribution model, considering comprehensive conservation to have been accomplished only when every ecoregion potentially inhabited by a species is included within the distribution of the species located within a protected area. During this process, an in situ ecological gap map is also created for each taxon by mapping only the spatial areas within the distribution model of each taxon which are occupied by ecoregions not represented at all in protected areas.

$$\text{ERS}_\text{in} = \left( \frac{\text{number of ecoregions represented in species distribution model located inside of protected areas}}{\text{number of ecoregions represented within species distribution model}} \right) \times 100 \tag{7}$$

The FCSin process calculates the average of the three in situ conservation metrics (Eq. 8). It then assigns each taxon to a priority category based on the FCSin score, with high priority (HP) for further in situ conservation action assigned when FCSin ≤ 25, medium priority (MP) where 25 ≤ FCSin < 50, low priority (LP) where 50 ≤ FCSin < 75 and sufficiently conserved (SC) for taxa whose FCSin ≥ 75.

$$\text{FCS}_\text{in} = \left[ \frac{\text{SRS}_\text{in} + \text{GRS}_\text{in} + \text{ERS}_\text{in}}{3} \right] \tag{8}$$

Summary evaluations

The first process in summary evaluations compiles the ex situ and in situ conservation gap analysis data into a summary file and then identifies for each taxon among the FCSex and FSCin the lower (minimum) conservation score (FCS-min) and the higher (maximum) conservation score (FCS-max), along with producing a combined final conservation score (FCS-mean) by calculating the average of FCSex and FCSin values (Eq. 9). Each taxon is then categorized with regard to the min and max, and the combination score, with high priority (HP) for further conservation action assigned when FCS < 25, medium priority (MP) where 25 ≤ FCS < 50, low priority (LP) where 50 ≤ FCS < 75 and sufficiently conserved (SC) for taxa whose FCS ≥ 75.

$$\text{FCS}_\text{mean} = \frac{\text{FCS}_\text{ex} + \text{FCS}_\text{in}}{2} \tag{9}$$

Here we also provide code for the calculation of an indicator across assessed taxa, which can be applied at national, regional, global or any other scale (Khoury et al. 2019b) (Eq. 10). The indicator is provided separately with regard to ex situ, in situ, min, max and combined conservation metrics, by deriving the proportion of species categorized as SC or LP out of all assessed taxa (Khoury et al. 2019b).

$$\text{Indicator} = \left[ \frac{\text{number of taxa whose FCS is} \geq 50}{\text{Total number of taxa}} \right] \times 100 \tag{10}$$

As a final step, a Rmarkdown process (function ‘SummaryHTML’) compiles the quantitative results and the spatial outputs for each taxon, producing interactive html documents.

**Internal GapAnalysis functions**

GapAnalysis includes three internal functions to support the conservation analysis. First, function ‘OccurrenceCounts’ produces counts per type of occurrence (G: ex situ sample and H: reference observations), generating a ‘data.frame’ as output. Second, function ‘Gbuffer’ generates a circular buffer of user-defined size (default 50 km radius) around each G point for each taxon. This buffer (stored in memory as a ‘RasterLayer’ object) represents the geographic areas already considered to have been sufficiently collected for ex situ conservation. Third, ‘ParamTest’ checks if occurrence data and distribution models exist for each species and sets the conservation scores to zero if there are no occurrences with coordinates or no model.

**Case studies**

We provide GapAnalysis case studies for wild plants within two different economically important genera with contrasting conservation status and metrics, namely, 1) for 37 taxa that are the wild relatives of chile peppers (Capsicum L.); and 2) 16 wild relatives of pumpkins, squashes, zucchini and gourds (Cucurbita L.) (Fig. 2). The methods for compiling the occurrence data, performing species distribution modeling and conducting GapAnalysis for these taxa are described in full in Khoury et al. (2019c, d). In short, occurrence data were drawn from biodiversity and conservation repository...
databases and from the authors’ own botanical explorations. Duplicates and non-wild records were removed, taxonomic names were standardized and records were classified as H (reference) or G (ex situ sample) as per the definitions described in this paper (see Input data). Species distributions were produced with the MaxEnt algorithm (Phillips et al. 2017) using 26 climatic and topographic predictors (Jarvis et al. 2008, Fick and Hijmans 2017) at a spatial resolution of 2.5 arc-minutes, employing a subset of variables as well as number and location of pseudo-absences specific to each taxon. The occurrence data, models and full results are available at Khoury et al. (2019c, d). A GapAnalysis R tutorial for three of the Cucurbita species is provided in the usage section of the GapAnalysis GitHub repository (<https://github.com/CIAT-DAPA/GapAnalysis>) and included as an example in the R package itself. Note that the three-species R package tutorial uses a lower spatial resolution for efficiency (10 arc-min versus 2.5 arc-min), and hence the GapAnalysis results differ from those described below and in Khoury et al. (2019c, d).

The 37 assessed wild Capsicum taxa range from the southern United States to northern Argentina, and include a few widely distributed members (C. annuum var. glabriusculum and C. rhomboideum in Central and South America and C. baccatum var. baccatum and C. chacoense in South America). Most taxa, however, are endemic species restricted to specific environments, particularly in coastal Brazil, the Galapagos, mainland Ecuador and Peru. Roughly half (18) of the species were assessed as high priority for conservation based on the final conservation score (FCSc-mean), with ex situ conservation gaps being more severe than in situ (94.6% of the taxa categorized as high priority based on FCSex alone) (Fig. 2). From the rest of the taxa, 17 were classified as medium priority, two as low priority and none considered sufficiently conserved. The two relatively well conserved taxa, assessed as low priority for further action based on their combined conservation scores, were both narrow-range endemic species, C. cardenasii (FCSc-mean = 61.5) and C. galapagoense (FCSc mean = 53.7). Their conservation status differed, however, by strategy, with C. cardenasii well conserved ex situ but less well in situ, and C. galapagoense the opposite. Note these and all other GapAnalysis scores were bound between 0 and 100, with 0 representing an extremely poor state of conservation, and 100 representing comprehensive protection. The indicator score across Capsicum, based on the proportion of taxa with a final combined conservation score ≥ 50 (Eq. 10, function ‘indicator’ in the R package), was 5.4 on the scale of 0–100.

In contrast to Capsicum, the 16 wild Cucurbita taxa, which are distributed from the central, southwestern and far southeastern USA south to Central America, with two additional species in South America (C. ecuadorensis and C. maxima subsp. andreana), were generally determined to be considerably better conserved ex situ, and slightly less well represented in protected areas in situ. As a result, based on the combined conservation assessment, no taxa were categorized as high priority for further action, 13 taxa were assessed as medium priority, two taxa as low priority and one as sufficiently conserved (Fig. 2). The indicator score across the Cucurbita species, based on the proportion of taxa with a final combined conservation score (FCSc mean) ≥ 50, was therefore 18.8 on the scale of 0–100. While most taxa were medium priority (25 ≤ FCSc mean < 50), there was wide variation in the FCS values across species for both conservation strategies. Current ex situ conservation status varied from extremely low for taxa with very few ex situ samples (e.g. C. cordata, only 3 G samples, FCSex = 12.44), to relatively high for taxa with many and highly geographically and environmentally representative ex situ samples including endemics (e.g. C. pepo subsp. fraterna, FCSex = 89.6), and

![Figure 2. GapAnalysis results for Capsicum (37 taxa) and Cucurbita (16 taxa) based on Khoury et al. (2019c, d). HP: high priority; MP: medium priority; LP: low priority; and SC: sufficiently conserved. The barplot (left) shows the number of species in each conservation category, whereas the numbers (right) are genus-level averages of GapAnalysis scores.](image-url)
more widespread species (e.g. *Cucurbita pepo* subsp. *ovifera* var. *ozarkana*, $FCS_{in} = 81.1$). Regarding in situ conservation, three taxa showed fairly high degrees of habitat protection – *C. cordata* ($FCS_{in} = 75.4$), *C. pepo* subsp. *fraternalis* ($FCS_{in} = 69.0$) and *C. palmata* ($FCS_{in} = 52.4$). On the contrary, $FCS_{in}$ scores ranged as low as 20.8 for *C. ecuadorensis*, and 22.9 for *C. pepo* subsp. *ovifera* var. *texana*.

**Discussion**

GapAnalysis facilitates a standardized estimation of the conservation status (and gaps therein) of taxa, their populations and their underlying genetic diversity, which can be applied from the single taxon to global (multi-species) levels, depending on available data. The method is useful both for information on conservation needs for specific species, as well as the prioritization of needs across taxa. The method is designed to use continually updated and openly available data and tools to enable its sustainability over time and its adaptability for diverse research purposes. In creating quantitative and spatial targets of comprehensive conservation against which to measure the current status of taxa, it represents a more holistic approach to estimating the state of conservation of diversity than do those based simply on counts of accessions held in genebanks, or on categorization of threats to species and their habitats in situ. The operationalization of the method in this package therefore usefully complements existing indicators for international and national conservation goals (Khoury et al. 2019b, Hoban et al. 2020a).

GapAnalysis is intended for the assessment of the conservation status of taxa. Hence, we have deliberately refrained from including functions pertaining to the downloading and cleaning of occurrence data, or for the construction of species distributions models. For downloading, pertinent packages include ‘rgbif’ (Chamberlain and Oldoni 2020), ‘genesyrs’ (Obreza 2019) and ‘spocc’ (Chamberlain 2020). Cleaning and coordinate quality assurance can be performed with ‘CoordinateCleaner’ (Zizka et al. 2019), whereas ‘taxize’ (Chamberlain and Szécs 2013) is useful for taxonomic naming. Species distribution modelling packages include ‘sdm’ (Naimi and Araújo 2016), ‘dismo’ (Hijmans et al. 2017), ‘maxnet’ (Phillips 2017), ‘wallace’ (Kass et al. 2018) and ‘BiodiversityR’ (Kindt 2019). We recommend that users build their conservation assessment workflows taking into account four major steps: 1) gathering occurrence data; 2) cleaning occurrence data; 3) mapping species distributions; and 4) conservation gap analysis.

We further note that GapAnalysis can complement conservation threat assessments, for example those in R package ‘RedListR’ (Lee et al. 2019). For the R tutorial example *Cucurbita* taxa, the preliminary threat assessments produced by ‘RedListR’ indicated that all three species were likely of Least Concern based on their extents of occurrence and areas of occupancy. Current Red Listings for the taxa assign *C. digitata* as of Least Concern, while *C. cordata* and *C. palmata* are categorized as Data Deficient (Khoury et al. 2019d). GapAnalysis revealed rather wide variation among these taxa regarding their degree of representation in protected areas, with *C. digitata* particularly under-protected. All three species require further collecting for ex situ conservation, with *C. cordata* most urgently prioritized for action.

With regard to the ex situ conservation analyses, the method is vulnerable to deficiencies in the completeness and availability of genebank, seedbank and botanic garden data. Openly available online conservation repository databases are not fully representative of all such institutions worldwide, those institutions that are listed may not report all holdings, and locality information (most importantly, coordinates) is lacking for many records that are available (Khoury et al. 2020). The SRSex function is uniquely useful as it is able to utilize all available G data, regardless of whether they possess coordinates. This said, due to these data gaps, it is possible that the method may underestimate the true degree of ex situ conservation of taxa.

Furthermore, the standardized size (default 50 km) of the spatial buffer drawn around G points to estimate areas already represented ex situ generalizes variation across wild plants with regard to genetic diversity within and among populations, which is influenced by reproductive strategy, including degree of outcrossing versus inbreeding, and pollinator and disperser behavior and patterns; ecological interactions (e.g. competition with other species and degree of obligate association with other species); and topographic factors (i.e. physical, climatic or ecological barriers to movement), among others (Hoban and Strand 2015). The GRSex and ERSex functions permit modifications of the standard buffer size, if taxon-specific knowledge is available to better inform the likely spatial coverage drawn around a collected population. We also note that the methodology assumes that the existence of a G point indicates that a plant population at that location has been adequately sampled across its individuals. In reality, field collecting for ex situ conservation may not have comprehensively sampled populations at the resolution needed for all conservation needs, thus further collecting may also be warranted within populations with preexisting G points, particularly to capture rare alleles (Hoban et al. 2020b).

With regard to the in situ conservation analyses, the method depends on the quality of the WDPA dataset. Our review of the most current version (IUCN 2019) leads us to consider that the dataset is of very high quality with regard to the number of protected areas and their spatial extents. Our default setting included all terrestrial and coastal reserves marked as designated, inscribed or established. Nevertheless, the user could refine these settings if interested in more stringent conservation standards. We also note that there are other open space lands, for example those managed by the U.S. Forest Service or Bureau of Land Management in the USA (aside from Wilderness Areas and a few other select designations), which are not listed in the WDPA due to being multi-use and not primarily focused on biodiversity conservation (Khoury et al. 2020). The WDPA may thus already be considered as relatively stringent about its definitions of protected areas (UNEP-WCMC 2019) and might
be seen by some land managers as underestimating habitat conservation. The incorporation of other effective area-based conservation measures (UNEP-WCMC 2020) into the protected areas analysis would provide a more generous interpretation of potential in situ conservation. On the other hand, we note that plants in general receive a disproportionately small percentage of conservation funding compared to fauna (Frances et al. 2018), and following from this, wild plant populations even in WDPA lands may not be adequately surveyed or managed. Robust long-term protection of these plants in these areas likely require active taxon- and population-specific management plans (Khoury et al. 2019c).

The methods implemented in GapAnalysis have only been applied to wild plant species thus far, in particular crop wild relatives and other wild flora of socioeconomic and cultural value, for two reasons. First, modeling distributions of domesticated plants and of animals (wild or domesticated) present unique challenges, for crop varieties because their ranges are determined by cultural as well as ecological factors (Ramirez-Villegas et al. 2020), and for livestock because of data gaps (FAO 2015; McGowan et al. 2018). Second, measuring in situ conservation of domesticated biodiversity necessitates a methodology based on on-farm and other forms of agricultural or rangeland conservation, rather than WDPA protected areas (Altieri and Merrick 1987, Bellon 2004). We plan in future versions of GapAnalysis to include tools useful for measuring conservation of these forms of biodiversity, toward the overall goal of enabling assessments of the conservation status of the full range of taxa targeted in international conservation agreements (Khoury et al. 2019b, Hoban et al. 2020a). Moreover, as genetic data across populations of more taxa becomes more readily available, we will look to adapt GapAnalysis to enable a more direct assessment of the state of conservation of species’ genetic diversity.

Finally, we pose that because GapAnalysis uses predicted species distributions, the approach offers potential to study how modifications of these distributions resulting from land use, climate or other changes, can alter conservation scores and targets. Likewise, GapAnalysis can be used to test the effect on species’ conservation status due to changes to the protected areas in a given country, region or globally.

To cite GapAnalysis or acknowledge its use, cite this Software note as follows, substituting the version of the application that you used for ‘version 1.0’:

Carver, D. et al. 2020. GapAnalysis: an R package to calculate conservation indicators using spatial information. (ver. 1.0).

Acknowledgments – The authors thank the myriad botanists, taxonomists, plant collectors, geospatial scientists and genetic resource professionals who have compiled and made available through open-access repositories the occurrence and spatial information that can be employed in GapAnalysis. We thank Kauê de Sousa for substantial code inputs.

Funding – Work contributing to this method was supported by the Government of Norway; the Biodiversity Indicators Partnership, an initiative supported by UN Environment, the European Commission and the Swiss Federal Office for the Environment; the United States Department of Agriculture; the CGIAR Genebank Platform; and the Global Crop Diversity Trust. C.K.K. was supported by grant no. 2019-67012-29733/project accession no. 1019405 from the USDA National Institute of Food and Agriculture. The funders played no role in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the article for publication. The USDA is an equal opportunity employer and provider. The data used to construct Fig. 2 are available in Dryad Digital Repository: <https://doi.org/10.5061/dryad.t4b8gtj1x>.

Conflicts of interest – The authors declare no conflicts of interest.

Author contributions

Daniel Carver: Data curation (equal); Formal analysis (lead); Investigation (supporting); Methodology (supporting); Software (lead); Writing – original draft (lead); Writing – review and editing (supporting). Chrystian Sosa: Data curation (lead); Formal analysis (lead); Investigation (supporting); Methodology (supporting); Software (lead); Writing – original draft (supporting); Writing – review and editing (supporting). Colin Khoury: Conceptualization (lead); Data curation (supporting); Investigation (lead); Methodology (lead); Software (supporting); Supervision (lead); Visualization (supporting); Writing – original draft (lead); Writing – review and editing (lead). Harold Achicanoy: Investigation (supporting); Methodology (supporting); Software (supporting); Writing – review and editing (supporting). Maria Victoria Diaz: Investigation (supporting); Methodology (supporting); Software (supporting); Writing – review and editing (supporting). Steven Sotelo: Investigation (supporting); Methodology (supporting); Software (supporting); Writing – review and editing (supporting). Nora Castaneda-Alvarez: Investigation (supporting); Methodology (supporting); Writing – review and editing (supporting). Julian Ramirez-Villegas: Conceptualization (lead); Funding acquisition (lead); Investigation (lead); Methodology (lead); Project administration (lead); Software (supporting); Supervision (lead); Visualization (supporting); Writing – original draft (supporting); Writing – review and editing (lead).

Transparent Peer Review

The peer review history for this article is available at <https://publons.com/publon/10.1111/ecog.05430>.

Data availability statement

The data used to construct Fig. 2 are available in Dryad Digital Repository: <https://doi.org/10.5061/dryad.t4b8gtj1x>.

References

Altieri, M. A. and Merrick, L. 1987. In situ conservation of crop genetic resources through maintenance of traditional farming systems. – Econ. Bot. 41: 86–96.
Balmford, A. et al. 2005. The 2010 challenge: data availability, information needs and extraterrestrial insights. – Phil. Trans. R. Soc. B 360: 221–228.
Bellon, M. 2004. Conceptualizing interventions to support on-farm genetic resource conservation. – World Dev. 32: 159–172.
Castañeda-Álvarez, N. P. et al. 2016. Global conservation priorities for crop wild relatives. – Nat. Plants 2: 16022.
Chamberlain, S. 2020. spocc: interface to species occurrence data sources. R package ver. 1.0.8. – <https://cran.r-project.org/web/packages/spocc/index.html>.
Chamberlain, S. A. and Oldoni, D. 2020. rgbif: interface to the global biodiversity information facility API. R package ver. 2.2.0. – <https://cran.r-project.org/web/packages/rgbif/index.html>.
Chapin, F. S. et al. 2000. Consequences of changing biotic diversity. – Nature 405: 234–242.
Convention on Biological Diversity (CBD) 2010a. Aichi biodiversity targets. – <https://www.cbd.int/sp/targets/>, accessed 22 April 2020.
Convention on Biological Diversity (CBD) 2010b. Global strategy for plant conservation. The Targets, 2011–2020. – <https://www.cbd.int/gspc/targets.shtml>, accessed 22 April 2020.
Convention on Biological Diversity (CBD) 2018. Strategic plan indicators. – <https://www.cbd.int/sp/indicators/>, accessed 22 April 2020.
Convention on Biological Diversity (CBD) 2020. Preparations for the post-2020 biodiversity framework. – <https://www.cbd.int/conferences/post2020>, accessed 22 April 2020.
Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). 2020. – <https://www.cites.org/eng/documents/StrategicVision>, accessed 22 April 2020.
Fick, S. E. and Hijmans, R. J. 2017. Worldclim 2: new 1-km spatial resolution climate surfaces for global land areas. – Int. J. Climatol. 37: 4302–4315. Data at: <http://worldclim.org/version2>.
Food and Agricultural Organization of the United Nations (FAO). 2002. The International Treaty on Plant Genetic Resources for Food and Agriculture. – Food and Agriculture Organization of the United Nations, Rome, Italy, <http://www.fao.org/3/j40510e.pdf>.
Food and Agricultural Organization of the United Nations (FAO). 2015. The second global assessment of animal genetic resources. – Food and Agriculture Organization of the United Nations, Rome, Italy, <http://www.fao.org/publications/sowangr/en/>.
Frances, A. L. et al. 2018. Conservation status and threat assessments for North American crop wild relatives. – In: Greene, S. L. et al. (eds), North American crop wild relatives, vol. 1: Conservation strategies. Springer International Publishing AG, pp. 189–208.
Hanson, J. O. et al. 2017. Environmental and geographic variables are effective surrogates for genetic variation in conservation planning. – Proc. Natl Acad. Sci. USA 114: 12755–12760.
Hijmans, R. J. et al. 2017. dismo: species distribution modeling. R package ver. 1.1-4. – <https://CRAN.R-project.org/package=dismo>.
Hoban, S. and Strand, A. 2015. Ex situ seed collections will benefit from considering spatial sampling design and species’ reproductive biology. – Biol. Conserv. 187: 182–191.
Hoban, S. et al. 2018. Implementing a new approach to effective conservation of genetic diversity, with ash Fraxinus excelsior in the UK as a case study. – Biol. Conserv. 225: 10–21.
Hoban, S. et al. 2020a. Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. – Biol. Conserv. 248: 108654.
Hoban, S. et al. 2020b. Taxonomic similarity does not predict necessary sample size for ex situ conservation: a comparison among five genera. – Proc. R. Soc. B 287: 20200102.
International Union for Conservation of Nature (IUCN) 2019. World database on protected areas. – <https://protectedplanet.net/>, accessed 19 April 2019.
Jarvis, A. et al. 2008. Hole-filled seamless SRTM data V4. – International Center for Tropical Agriculture (CIAT), <http://srtm.csi.cgiar.org>.
Kass, J. M. et al. 2018. wallace: a flexible platform for reproducible modeling of species niches and distributions built for community expansion. – Methods Ecol. 9: 1151–1156.
Khoury, C. K. et al. 2019a. Data for the calculation of an indicator of the comprehensiveness of conservation of useful wild plants. – Data Brief 22: 90–97.
Khoury, C. K. et al. 2019b. Comprehensiveness of conservation of useful wild plants: an operational indicator for biodiversity and sustainable development targets. – Ecol. Indicators 98: 420–429.
Khoury, C. K. et al. 2019c. Modeled distributions and conservation status of the wild relatives of chile peppers (Capsicum L.). – Divers. Distr. 26: 209–225.
Khoury, C. K. et al. 2019d. Distributions, conservation status and abiotic stress tolerance potential of wild cucurbits (Cucurbita L.). – Plants People Planet 2: 269–283.
Khoury, C. K. et al. 2020. Crop wild relatives of the United States require urgent conservation action. – Proc. Natl Acad. Sci. USA 117: 33351–33357.
Kindt, R. 2019. BiodiversityR: package for community ecology and suitability analysis. R package ver. 2.11-3. – <https://cran.r-project.org/web/packages/BiodiversityR/index.html>.
Lebeda, A. et al. 2019. Research gaps and challenges in the conservation and use of North American wild lettuce germplasm. – Crop Sci. 59: 2337–2356.
Lee, C. K. F. et al. 2019. RedListr: tools for the IUCN Red Lists of ecosystems and threatened species in R. – Ecosphere 42: 1050–1055.
McGowan, P. J. K. et al. 2018. Tracking trends in the extinction risk of wild relatives of domesticated species to assess progress against global biodiversity targets. – Conserv. Lett. 12: e12588.
Mezghani, N. et al. 2019. Distributions and conservation status of carrot wild relatives in Tunisia: a case study in the western Mediterranean basin. – Crop Sci. 59: 1–12.
Millennium Ecosystem Assessment. 2005. Ecosystems and human well being: synthesis. – Island Press, Washington DC.
Myrans, H. et al. 2020 Modelled distributions and conservation priorities of wild sorghums (Sorghum Moench). – Divers. Distr. 26: 1727–1740.
Naimi, B. and Araújo, M. B. 2016. sdm: a reproducible and extensible R platform for species distribution modelling. – Ecoscience 39: 368–375.
Norton, S. L. et al. 2017. Priorities for enhancing the ex situ conservation and use of Australian crop wild relatives. – Aust. J. Bot. 65: 638–645.
Obreza, M. 2019. genesysr: genesys PGR Client. R package ver. 1.0.0. – <https://cran.r-project.org/web/packages/genesysr/index.html>.
Phillips, S. 2017. maxnet: fitting 'maxent' species distribution models with 'glmnet'. R package ver. 0.1.2. – <https://cran.r-project.org/web/packages/maxnet/index.html>.
Phillips, S. J. et al. 2017. Opening the black box: an open-source release of Maxent. – Ecoscience 40: 887–893.
Ramírez-Villegas, J. et al. 2010. A gap analysis methodology for collecting crop gene pools: a case study with Phaseolus beans. – PLoS One 5: e13497.

Ramírez-Villegas, J. et al. 2020. A gap analysis modeling framework to prioritize collecting for ex situ conservation of crop landraces. – Divers. Distrib. 26: 730–742.

Rockström, J. et al. 2009. A safe operating space for humanity. – Nature 461: 472–475.

The Nature Conservancy Geospatial Conservation Atlas. 2019. Terrestrial ecoregions. – <https://geospatial.tnc.org/datasets/7b7fb9d945544d41b3e7a91494e42930_0>.

UNEP-WCMC. 2019. User manual for the world database on protected areas and world database on other effective area-based conservation measures: 1.6. – UNEP-WCMC, Cambridge, UK, <http://wcmc.io/WDPA_Manual>, accessed 14 July 2020.

UNEP-WCMC. 2020. The world database on other effective area-based conservation measures. – UNEP-WCMC, Cambridge, UK, <https://www.protectedplanet.net/c/other-effective-area-based-conservation-measures>, accessed 14 July 2020.

United Nations. 2015. Sustainable development goals. – <https://www.un.org/sustainabledevelopment/sustainable-development-goals>, accessed 14 May 2020.

Zizka, A. et al. 2019. coordinatecleaner: standardized cleaning of occurrence records from biological collection databases. – Methods Ecol. Evol. 10: 744–751.