Introduction

Perhaps the most fundamental aim of conservation is to maximize the long-term persistence of biodiversity (McNeely 1994; Margules & Pressey 2000). To achieve this, conservation actions must preserve biodiversity patterns (eg. populations, species, ecosystems), but also crucially the processes that sustain them. One of the major tangible achievements of modern conservation has been the act of setting aside areas for preservation (Watson *et al.* 2014; Sanderson *et al.* 2015). Reserve networks buffer species from gross threatening processes and set the stage for enhanced management interventions (Gaston *et al.* 2008). Since the resources available for conservation action are limited, protected area networks must be sited in places that satisfy conservation objectives for minimal cost (Margules & Pressey 2000).

Today, the most widely used conservation planning tools focus on biodiversity patterns (`Zonation` and `Marxan`; Moilanen 2007; Ball *et al.* 2009). Decision makers can use these tools to obtain solutions that secure a proportion of the geographic range of each biological feature (populations, species, or ecosystems) of interest by setting targets. One method to incorporate data on the ecological and evolutionary processes that sustain biological features is to conserve a representative sample of the physical attributes that underpin these processes. To achieve this, current methods typically involve partitioning each feature into sub-groups based on an attribute variable that relates to a biodiversity process of interest (Cowling & Pressey 2001; Pyke & Fischer 2005; Klein *et al.* 2009). For instance, by dividing species distributions into sub-groups according to habitat discontinuities, and ensuring that each sub-group is represented in the solution, conservation planners can obtain prioritizations that promote adaptive processes (Carvalho *et al.* 2011). However, using this method is challenging because biodiversity often cannot be divided into operational groups for conservation planning without loss of information (Orians 1993; Pressey & Logan 1994; Faith & Walker 1996). This limitation has been known of quite some time, and dates back to some of the earliest reserve selection methods (Orians 1993).

To overcome this limitation, Faith and Walker (1993) developed a decision support tool (`DIVERSITY`) to identify prioritizations that secure a representative sample of the environmental variation found across a study area. Although this method was originally pitched as an alternative to species-based conservation planning (Faith 1994, 2003; Faith & Walker 1996), it can also identify solutions that conserve a representative sample of variation across a single species' range. Recent work has built on this reserve selection method by solving problems with more advanced optimization algorithms (Engelbrecht *et al.* 2016). However, existing `DIVERSITY`-based methods have limited utility for species-based conservation planning, in part, because they cannot accommodate multiple biological features and do not provide options to specify specific targets pertaining to the amount of habitat required to conserve species (unlike `Marxan` and `Zonation`). As a consequence, such methods are not useful for most conservation planning exercises (Margules & Pressey 2000).

Conservation planners lack a decision support tool that lets them set explicit targets to obtain solutions that (1) secure an adequate amount of habitat for each feature and (2) a representative sample of the variation in each feature. To begin to fill this gap, here we unite the ideas underpinning `DIVERSITY` and `Marxan` into new formulations of the reserve selection problem, and implement them in the `raptr` `R` package. This `R` package provides decision makers with the tools to generate prioritizations based on data that relate to biodiversity patterns and processes. Here, we aim to the provide an in-depth understanding of this `R` package and explore its functionality.

Methods

PROBLEM FORMULATION

Biodiversity features are defined as the entities that the prioritization is required to preserve (eg. species, ecosystems). Here, attributes are defined as the variation across the features' ranges that the prioritization is required to sample. These attributes can be intrinsic (eg. genetic or phenotypic) or extrinsic (eg. environmental

conditions) to the feature. There should, however, be a reasonable underlying hypothesis that relates the attribute to the biodiversity processes that needs to be conserved.

A set of attributes is conceptualized as a space (termed an "attribute space"). For example, a decision maker may require a prioritization that represents populations along climatic gradients. To achieve this, the decision maker might use a "climatic" attribute space with dimensions relating to mean annual temperature (°C) and precipitation (mm). Any given combination of temperature and precipitation may be conceived as a point in this climatic space. Although they exist as polygons, for simplicity, each planning unit may be thought to exist as a single point inside a given attribute space. By associating the planning units with climatic data, and calculating a descriptive statistic for each planning unit (eg. mean), they can be mapped from geographic space to this climatic attribute space. In addition to points representing planning units, attribute spaces also contain demand points.

Demand points (Faith & Walker 1993, 1996; Faith 2003) are designated by the decision maker to indicate regions of the attribute space that they wish to represent in the prioritization (see below for discussion on generating demand points for real-world data sets). For instance, by siting demand points throughout an attribute space, planners can obtain solutions that conserve a representative sample of an attribute space. Alternatively, by siting demand points in specific regions of an attribute space, planners can obtain solutions that target specific samples of an attribute space. For a given set of demand points, the shorter the distances between the demand points and the planning units selected for prioritization; the better a solution is at securing the desired variation in the attribute space. In any attribute space there may exist points that are impossible (eg. mean annual rainfall -5 mm), or do not occur in the study area (eg. mean annual temperature 30°C in Antarctica). Additionally, there may be some regions that are desirable for some features and undesirable for others (eg. conditions known to be outside the physiological tolerance of certain species). Thus a different set of demand points may be appropriate for different attribute spaces and features.

To illustrate these concepts, consider the following example: we wish to develop a prioritization for a single species that has four populations. Since we can only afford to preserve three of the four populations, our objective is to conserve the most representative sample possible. To achieve this goal, we obtained annual rainfall (mm) and temperature (°C) data at the location of each population. We used this data to construct a two-dimensional climatic attribute space. Next, we generated demand points as equi-distant points inside this space. By comparing the distribution of the demand points to the distribution of the populations in the attribute space, we can identify the optimal solution (Fig. 1). We can see that a solution that prioritizes both populations *A* and *B*, effectively "doubles-up" on the same climatic characteristics and constitutes considerable redundancy if both selected. Instead, a more representative sample of the intra-specific variation could be captured by securing populations *A* (or *B*), *C*, and *D*. However, if the goal of the prioritization was to preserve populations living in warmer temperatures, then instead of siting demand points across the full range of conditions, we could site demand points in environmental conditions with temperatures over 30 °C (ie. the top two rows of demand points in Fig. 1). Given this new set of demand points, populations *A*, *B*, and *C* would be prioritized. Since demand points can be sited and weighted in any configuration, they provide a flexible means to guide the reserve selection process.

The `raptr R` package utilizes two novel formulations of reserve selection problem to generate prioritizations. These formulations are based on elements of the `Marxan`, `DIVERSITY`, and uncapacitated facility location problems (Cornuéjols *et al.* 1990). Since they are based on the unreliable and reliable facility location problems (Cui *et al.* 2010), the formulations are hereafter referred to as the "unreliable" and "reliable" formulations. The difference between the two formulations is that the reliable formulation explicitly considers the probability that planning units are occupied by features when calculating how well a given solution samples a feature's attribute space. On the other hand, the unreliable formulation assumes that features occupy all of the planning units within which they are found with 100 % certainty when performing these calculations. For brevity, we will define the simpler formulation–the unreliable formulation–below and define the more complex version–the reliable formulation–in Appendix S1. All mathematical terms defined hereafter are described in Table S1. For convenience, the cardinality of sets will be denoted using the same symbol used to denote the set.

Define F to be the set of features one wishes to conserve (indexed by f). Let J be a set of planning

units (indexed by j). Also, let A_j denote the area, and C_j denote the cost of preserving planning unit $j \in J$. To assess the extent to which each feature is secured in a given prioritization, let q_{fj} denote the probability of feature f occupying planning unit j. The level of fragmentation associated with a prioritization is parameterized as the total exposed boundary length (as in `Marxan`). Let the shared edges between each planning unit $j \in J$ and $k \in J$ be e_{jk}.

Let S denote a set of attribute spaces (indexed by s). Each $j \in J$ is associated with spatially explicit data that represent coordinates for each attribute space $s \in S$. Let I_{fsi} denote a set of demand points (indexed by i) for each feature $f \in F$ and each attribute space $s \in S$. Let λ_{fsi} denote the weighting for each demand point $i \in I$, $f \in F$ and $s \in S$. Let d_{fsij} denote the distance between each demand point $i \in I$ and each planning unit $j \in J$ for each feature $f \in F$ and attribute space $s \in S$. To describe the inherent variation in the distribution of demand points for feature f and space s, let δ_{fsi} denote the distance between each demand point $i \in I$ and the centroid of the demand points. Demand points with greater weight λ_{fsi} are more important, and the solutions that select planning units close to highly weighted demand points will be more favorable. As a consequence, the decision maker will need to choose an appropriate weighting for each demand point.

Targets are used to ensure that prioritizations sufficiently conserve each feature. Amount-based targets specify the minimum amount of habitat required for each feature to be adequately conserved (similar to those used in `Marxan`). Let T_f denote the amount of area or habitat that needs to be preserved for each feature $f \in F$. Space-based targets specify the minimum proportion of variation in the demand points that needs to be secured for each feature to be sufficiently conserved. Space-based targets are expressed as proportions–instead of a sum of weighted distances–by scaling the sum of weighted distances between the demand points and the selected planning units in a solution relative to the distances between demand points and the demand points' centroid. This scaling is conceptually similar to that used in calculating the R^2 statistic for K-means clustering analyses from the within sums of squares and total sums of squares [Greenacre & Primicerio (2014); page 106]. Let τ_{fs} denote the space-based targets for feature $f \in F$ and attribute space $s \in S$. The control variables for the unreliable formulation are the B, T_s, and τ_{fs} variables.

$$T_s = \text{amount target for feature } f \qquad\qquad \text{eqn 1a}$$
$$\tau_{fs} = \text{space target for feature } f \text{ in attribute space } s \qquad\qquad \text{eqn 1b}$$
$$B = \text{boundary length modifier (BLM) to penalize fragmented solutions} \qquad\qquad \text{eqn 1c}$$

The decision variables are the X_j and Y_{fsij} variables.

$$X_j = \begin{cases} 1, & \text{if planning unit } j \text{ is selected for conservation action} \\ 0, & \text{otherwise} \end{cases} \qquad\qquad \text{eqn 2a}$$

$$Y_{fsij} = \begin{cases} 1, & \text{if demand point } i \text{ is used to represent planning unit } j \text{ for} \\ & \text{feature } f \text{ in space } s. \\ 0, & \text{otherwise} \end{cases} \qquad\qquad \text{eqn 2b}$$

Each demand point $i \in I$ for feature $f \in F$ and space $s \in S$ is represented by a single selected planning unit (ie. a $j \in J$ where $X_j = 1$). The degree to which a demand point i is represented by a planning unit j is determined by the distance between them (d_{fsij}). Generally–unless near zero space targets are used so that the problem is effectively unconstrained by the target–demand points are represented by their closest selected planning units. In poorer quality solutions, demand points will be represented by planning units that have larger distances between them. As a consequence, the sum of the weighted distances between all of the demand demand points and the planning units used to represent them will be larger, and so, the solution will capture less of the variation described by the demand points.

The unreliable formulation (URAP) is a defined as a multi-objective optimization problem.

$$\text{(URAP)} \quad \text{Min} \sum_{j=0}^{J-1} (X_j C_j) + \sum_{j=0}^{J-1} \sum_{k=j}^{J-1} X_j (1 - X_k) (Be_{jk}) + \qquad \text{eqn 3a}$$

$$\text{s.t.} \sum_{j=0}^{J-1} A_j q_{fj} \geq T_f \qquad \forall 0 \leq f \leq F - 1 \qquad \text{eqn 3b}$$

$$1 - \frac{\sum_{i=0}^{I-1} \sum_{j=0}^{J-1} \lambda_{fsi} d_{fsij}{}^2 Y_{fsij}}{\sum_{i=0}^{I-1} \lambda_{fsi} \delta_{fsi}{}^2} \geq \tau_{fs} \qquad \forall 0 \leq f \leq F - 1, \qquad \text{eqn 3c}$$
$$0 \leq s \leq S - 1$$

$$\sum_{j=0}^{J-1} Y_{fsij} = 1 \qquad \forall 0 \leq f \leq F - 1, \qquad \text{eqn 3d}$$
$$0 \leq s \leq S - 1,$$
$$0 \leq i \leq I - 1$$

$$Y_{fsij} \leq X_j \qquad \forall 0 \leq f \leq F - 1, \qquad \text{eqn 3e}$$
$$0 \leq s \leq S - 1,$$
$$0 \leq i \leq I - 1,$$
$$0 \leq j \leq J - 1$$

$$X_j, Y_{fsij} \in 0, 1 \qquad \forall 0 \leq f \leq F - 1, \qquad \text{eqn 3f}$$
$$0 \leq s \leq S - 1,$$
$$0 \leq i \leq I - 1$$

The objective function (eqn 3a) determines the utility of a given prioritization: a combination of the total cost of a prioritization and how fragmented it is. Constraints (eqn 3b) ensure that all the amount-based targets are met. Constraints (eqn 3c) ensure that all the space-based targets are met for each feature and each attribute space. For each feature and attribute space, the total weighted distance between the demand points and their closest selected planning units is calculated ($\sum_{i=0}^{I-1} \sum_{j=0}^{J-1} \lambda_{fsi} d_{fsij}{}^2 Y_{fsij}$). This total weighted distance is then scaled by the inherent variation in the demand points ($\sum_{i=0}^{I-1} \lambda_{fsi} \delta_{fsi}{}^2$). As previously mentioned, the resulting fraction is used to calculate a proportion conceptually similar to the R^2 statistic used in k-means clustering analysis. The constraints (eqn 3c) ensure that the proportion of variation in the demand points secured in the solution must be equal to or greater than the space-based target. Constraints (eqn 3d) ensure that only one planning unit is assigned to each demand point. Constraints (eqn 3e) ensure that demand points are only assigned to selected planning units. Constraints (eqn 3f) ensure that the X and Y variables are binary.

Optimization

Although the reserve selection problems presented here are non-linear (see Appendix S1 for reliable formulation), they can be linearized using methods described by Beyer *et al.* (2016) and Cui *et al.* (Cui *et al.* 2010). The `raptr` R package provides functions to express conservation planning data as linearized versions of the optimization problems and solve them using the commercial `Gurobi` software suite (www.gurobi.com). Presently, academics can obtain a license at no cost.

Examples

To showcase the behavior of the unreliable formulation, we conducted a simulation study and two case studies. To understand how long it would take to solve various sized problems, we also conducted a benchmark analysis (Appendix S2). We completed the analyses using `R` (version 3.3.2; R Core Team 2016) and solved all optimization problems to within 99 % of optimality.

SIMULATION STUDY

Methods

We simulated a hypothetical study area with square planning units arranged in a 10×10 grid (Fig. 2). We then simulated three species to inhabit this study area. Firstly, we simulated a hyper-generalist species (hereafter referred to as the "uniformly distributed species"). It occupied all planning units with equal probability (Fig. 2a; eqn 4a). Secondly, we simulated a species with simple habitat requirements (hereafter referred to as the "normally distributed species"; Fig. 2b). This species was most likely to be found in planning units nearest to the center of the study area. It was simulated using the density function of a multivariate normal distribution (represented by $\mathcal{N}$; eqn 4b). Thirdly, we simulated a species two distinct populations (hereafter referred to as the "bimodally distributed species"; Fig. 2c). It was simulated using the maximum density of two multivariate normal distributions (eqn 4c). For a given species, planning unit occupancy was calculated using the (X, Y) coordinates of the units' centroids and the relevant equation. We used a geographic attribute space to provide an intuitive visualization of the solutions. Demand points were set as the planning units' centroids and were weighted according to the units' probability of occupancy.

$$P\left(\text{uniformly distributed species}|\,(x,y)\right) = 0.5 \qquad \text{eqn 4a}$$

$$P\left(\text{normally distributed species}|\,(x,y)\right) = \frac{\mathcal{N}\left(\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]\left[\begin{smallmatrix} 0.0 \\ 0.0 \end{smallmatrix}\right]\left[\begin{smallmatrix} 12.58 & 0 \\ 0 & 12.5 \end{smallmatrix}\right]\right)}{2} \qquad \text{eqn 4b}$$

$$P\left(\text{bimodally distributed species}|\,(x,y)\right) = \text{Max} \begin{cases} \mathcal{N}\left(\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]\left[\begin{smallmatrix} -3.75 \\ -3.75 \end{smallmatrix}\right]\left[\begin{smallmatrix} 10 & 0 \\ 0 & 10 \end{smallmatrix}\right]\right), \\ \frac{\mathcal{N}\left(\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]\left[\begin{smallmatrix} 3.75 \\ 3.75 \end{smallmatrix}\right]\left[\begin{smallmatrix} 8 & 0 \\ 0 & 8 \end{smallmatrix}\right]\right)}{2} \end{cases} \qquad \text{eqn 4c}$$

We generated four solutions for each species. Firstly, to represent solutions generated using conventional methods (eg. `Marxan`), we generated solutions using 20 % amount targets. Secondly, to show how the addition of space targets can affect solutions, we generated solutions using 20 % amount targets and 90 % space targets to capture the geographic spread of each species (hereafter referred to as geographic spread targets). Thirdly, to represent solutions generated using conventional planning methods that penalize for fragmentation, we generated solutions using 20 % amount targets and a boundary length modifier of 2.5. Fourthly, to illustrate the combined effects of using amount and space targets as well as penalties for fragmentation, we generated solutions using 20 % amount targets, 90 % geographic spread targets, and boundary length modifiers of 2.5.

Uniformly distributed species

The solution generated for the uniformly distributed species using an amount target prioritized 20 planning units near the southern end of study area (Fig. 2a). For this species, all solutions containing this number of planning units are optimal when using a 20 % amount target. We obtained this particular solution due to artifacts in the solver (eg. seed for the random number generator, order of variables in the problem matrix). This solution captured a small amount of the geographic spread of the species (-23.64 % sampled). In fact,

the coverage was so poor that the proportion was negative because the solution captured a less representative sample than a solution containing one planning unit in the center of the species' distribution.

By explicitly targeting representativeness, we obtained a solution that captured the geographic spread of the uniform species (93.39 % sampled; Fig. 3b). Although this solution secured an adequate amount of habitat and a representative sample of the species' geographic spread, this solution was highly fragmented. However, by penalizing fragmentation using a boundary length modifier, we were able to obtain a well connected solution that met all of the objectives (Fig. 3d). This solution prioritized the same number of planning units as the previous solutions even though it is far superior. As we can see, under the simplest of circumstances, reserve selection methods may not yield solutions that secure a representative sample of features unless constraints are used to guarantee this property.

Normally distributed species

The solution generated for the normally distributed species using just an amount target prioritized planning units in the species' core range area (Fig. 3e). While this strategy is cost-effective for protecting habitat, it was ineffective for securing a representative sample of the species' geographic spread (-6.14 % sampled) because it did not protect any planning units along the periphery of the species' distribution. By using amount and space targets, we were able to conserve an adequate amount of habitat and also secure a representative sample of the species' geographic spread (99.06 % sampled; Fig. 3f). Similar to the solutions for the uniformly distributed species, this solution was highly fragmented and we were able to obtain a better connected solution by specifying fragmentation penalties (Fig. 3h). However, unlike the solutions for the uniformly distributed species, the solution generated using an amount target, a space target, and fragmentation penalties required more planning units than the other solutions. These results suggest that solutions may need to prioritize more planning units to meet additional conservation objectives.

Bimodally distributed species

The solution generated for the bimodally distributed species using an amount target only conserved individuals belonging to one of the two populations (Fig. 3i). As a consequence, this solution did not secure a representative sample of the species' geographic spread (8.09 % sampled). The addition of fragmentation penalties exacerbated this issue, and resulted in a solution that sampled even less of the species' geographic spread (8.09 % sampled; Fig. 3k). However, this issue was resolved by using a space target to obtain a solution that secured both populations (99.39 % geographic spread secured; Fig. 3l). This finding suggests that species with large intra-specific variation could benefit the most from prioritizations generated using space-based targets.

CASE STUDY 1

Methods

We investigated how space-based targets can be used in a multi-species planning context to generate a prioritization that sufficiently preserves the species' realized niches. By preserving the populations in suitable habitats with different environmental conditions, conservation planners can preserve the species' adaptive landscape and foster resilience to environmental change (Moritz 2002). We selected Queensland, Australia as the study area. We obtained data for 19 bioclimatic variables across the region (at $30''$ resolution from www.worldclim.org; Hijmans *et al.* 2005) and subjected them to a principal components analysis (using ArcMap 10.3.1). We used the scores from the first two principal components to characterize the environmental variation across the study area (explaining 99.5 % of the total variation; Fig. 4).

We used four bird species in this case study: blue-winged kookaburra (*Dacelo leachii*), brown-backed honeyeater (*Ramsayornis modestus*), brown falcon (*Falco berigora*), and pale headed rosella (*Platycercus adscitus*). These species span a range of different habitat requirements. We mapped the extent of occurrence

for each species (Fig. 5). To do this, we obtained occurrence records from the Atlas of Living Australia across the whole of Australia (using the `ALA4R R` package; Raymond *et al.* 2015), spatially thinned the data to omit points within 10 km^2 of each other to ameliorate the effects of sampling bias (using the developmental version of the `spThin R` package; https://github.com/mlammens/spThin; Aiello-Lammens *et al.* 2015), and fit 85 % minimum convex polygons (using the `adehabitatHR R` package; Calenge 2006). We used this method because it is entirely reproducible using freely available data.

We generated 20 demand points for each species (Fig. S2). To ensure that the demand points reflected the core parts of the species' realized niches, we used the following method. For each species, we generated random points inside the species' geographic range and at each point extracted the principal component values at that location. We then fit hyperbox kernels to the distribution of principal component values to characterize the realized niche of each species using a manually chosen bandwidth of 0.2 and a 0.5 quantile to map the core parts of the species' niches (implemented in the `hypervolume R` package; Blonder *et al.* 2014). We then generated uniformly distributed points inside the species' distribution in environmental space, and extracted the kernel density at their locations. These uniformly distributed points and associated density estimations were used as demand point coordinates and weights (respectively).

We generated two multi-species prioritizations. The first solution was generated using 20 % amount targets for each species. The second solution was generated the same amount targets with an additional 10^{6} % niche-based target for each species.

Results

Generally, the solution generated using just amount targets preserved a representative sample of each the four bird species' niches. (Figure S1). This solution secured over 10^{6} % of the realized niche of blue-winged kookaburra (84.51 %), brown falcon (93.51 %), brown-backed honeyeater (-29.33 %), and the pale-headed rosella (64.8 %). Yet it failed to achieve this for the . This result demonstrates that although conventional methods may yield solutions that conserve a representative sample of some features, only through the use of explicit targets can planners obtain cost-effective solutions that secure a representative sample of all features.

CASE STUDY 2

Methods

Here we used space-based targets to generate a prioritization securing a representative sample of a species' intra-specific genetic variation. We used species occurrence and genetic data collected by the international IntraBioDiv project in the European Alps (Gugerli *et al.* 2008; Alvarez *et al.* 2009; Meirmans *et al.* 2011). Although this dataset contains multiple species, we used data for the betony-leaved rampion (*Phyteuma betonicifolium*) as a simple case study, because exhibits significant inter-population genetic structure (Meirmans *et al.* 2011). Members of the IntraBioDiv project collected data using a 20′ longitude by 21′ latitude grid (approx. 22.3 km × 25 km; Fig. 7a). They visited every second grid cell, and if the species was detected in a cell, samples were collected from three individuals. Samples were genotyped using amplified fragment length polymorphisms (AFLP; Vos *et al.* 1995), and used to construct matrices denoting the presence/absence of polymorphisms at loci. In total, 131 individuals were genotyped at 138 markers.

We used non-metric multi-dimensional scaling (NMDS; using Gower distances to accommodate sparsity; Gower 1971; implemented using the `cluster R` package; Maechler *et al.* 2015) to ordinate the presence (or absence) of locus-specific alleles within individuals into two continuous variables (implemented in the `vegan R` package; Oksanen *et al.* 2015). These continuous variables described the main axes of genetic variation within the species. We averaged together values corresponding to samples collected in the same grid cell, and used them to create a genetic attribute space (Figs. 7c and 7d). To assess spatial auto-correlation, we calculated Moran's I auto-correlation index for each NMDS axis using scaled great circle distances between the grid cells' centroids (using the `ape R` package; Paradis *et al.* 2004).

The grid cells were used as planning units for generating prioritizations. Since data were not collected in every grid cell, and the planning units were arranged in a checkerboard pattern. The grid-cell averaged ordinations were used to describe the typical genetic characteristics of individuals in the planning unit. Since the number of planning units was relatively small, we used the same grid-cell averaged ordinations as demand points. To ensure that the solutions did not prioritize particularly costly areas, we obtained population density data (1 km resolution from the Global Rural-Urban Mapping Project; `GRUMP V1`; Center for International Earth Science Information Network (CIESIN) *et al.* 2011) and estimated the total population density inside each grid cell. We used these values to represent opportunity cost (Fig. 7b).

We generated two solutions. The first solution was generated using an 10 % amount-based target. The second solution was generated using the same amount-based target and an additional 85 % genetic-based target.

Results

A two-dimensional attribute genetic space was able to capture most of the variation between individuals (stress = 0.17; Figs. 7c–7d). Planning units located near each other contained individuals with similar genetic characteristics (Moran's I: NMDS axis 1, $I = 0.4$, $P < 0.001$; Moran's I: NMDS axis 2, $I = 0.32$, $P < 0.001$). In terms of the average genetic characteristics of individuals in the planning units, they tended to cluster into two main groups, with evidence of within-group structure inside the larger group (Fig 8c). This analysis supports previous work by Alvarez *et al.* (2009) who also found evidence of genetic structure within this species.

The solution generated using just the amount target failed to preserve a representative portion of the species' genetic variation (15.52 % sampled; Figs. 8a and 8c). This solution only conserved individuals in one of the two main genetic groups. In fact, this solution is just a collection of the cheapest planning units occupied by the species needed to fullfil the target. Whereas, the solution generated using both amount and genetic targets conserved individuals in each of the two main genetic groups (99.96 % sampled; Figs. 8b and 8d). Although the solutions only differ by a single planning unit, by securing individuals in both main genetic groups the second solution was able to secure a much more representative sample of the species' genetic variation for only a minor increase in cost (920.08 compared to 99.41 total cost).

Implications and future directions

The `raptr` `R` package provides a unified approach to reserve selection. Conservation planners can use this `R` package to generate prioritizations that secure intra- and inter-specific biodiversity patterns. Both the simulations and case studies show that explicitly targeting a representative sample of each feature in a conservation planning exercise can substantially alter prioritizations. Additionally, we found that prioritizations often needed to secure more habitat to conserve a representative sample of each feature. Although it may not be practical to conserve substantially more habitat than the amount specified using amount-based targets (eg. 10 % geographic range; Margules & Pressey 2000), we did find that even small increases in reserve size could yield large gains.

One of the key advantages of this package is that it is general enough to incorporate almost any spatially explicit variable. This package can accommodate intrinsic or extrinsic variation in the feature(s). For example, adaptation processes could be secured using environmental variation (eg. Pyke & Fischer 2005), and trophic processes could be conserved by capturing the overlapping distributions of predator and prey species (eg. Rayfield *et al.* 2009; Chernomor *et al.* 2015). Note that maximizing one process can result in trading off against another (eg. geographic spread vs. connectivity). Additionally, space targets can be used to achieve representation for reasons that are not related to biodiversity processes, such as obtaining a geographically representative set of reserves so there is equitable access to parks by people (eg. Moilanen *et al.* 2013). As long as the variation can be described using Euclidean distances–which could be achieved through transformation or dimension reduction–the `R` package can be used to obtain a representative sample.

The degree to which a prioritization truly secures a representative sample of a feature depends on (1) the attribute space(s), (2) the distribution of demand points chosen by the conservation planner, and (3) the space target used. The optimal solution will not effectively represent the decision maker's objectives if an inappropriate set of spatial variables or demand points are used to represent the attribute space. Demand points should be distributed across the full range of variation in an attribute space to obtain solutions that secure a representative sample of the attribute space (see the `make.DemandPoints` function for statistical routines to achieve this). As with conventional amount-based targets, planners will have to ensure that space targets are high enough to fulfill conservation objectives (eg. when setting targets to conserve intra-specific genetic variation; Frankham *et al.* 2014). We encourage conservation planners to consult with experts to identify suitable targets.

Attribute data need to be spatially comprehensive to map planning units to attribute spaces. However, many real-world data sets are patchy. For example, due to constrained resources, genetic data may only be available for some planning units. To use such patchy data, planners could omit the planning units that are missing data, or estimate the missing data using spatially explicit models (eg. krigging or generalized dissimilarity models; Oliver & Webster 1990; Ferrier *et al.* 2007). This approach has been successfully applied to a range of biological data sets (eg. Thomassen *et al.* 2010). We caution that inaccuracy and uncertainty in data can negatively impact prioritizations (Wilson *et al.* 2005), and recommend that planners ensure that data are sufficiently reliable or use the reliable formulation (described in Appendix S1) to accommodate uncertainty into the reserve selection process.

To maximize the long-term persistence of biodiversity, decision makers need to identify prioritizations that preserve existing patterns of biodiversity and the processes that support them. It is becoming increasingly apparent that simply protecting a large area will fail to achieve this (Barnes 2015). Here, we developed the `raptr` R package to provide conservation planners with the tools to deliver cost-effective prioritizations that secure an adequate amount of a representative sample of biodiversity features. By exploring the functionality of this software package, we found that explicitly targeting a representative sample of each feature can result in substantially different solutions.

Acknowledgements

Data accessibility

The `raptr` R package can be downloaded from The Comprehensive R Archive Network (https://CRAN.R-project.org/package=raptr). All data, code, and results are stored in an online repository (https://github.com/jeffreyhanson/raptr-manuscript) to permit replication and validation of this study.

References

Aiello-Lammens, M.E., Boria, R.A., Radosavljevic, A., Vilela, B. & Anderson, R.P. (2015). spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, **38**, 541–545.

Alvarez, N., Thiel-Egenter, C., Tribsch, A., Holderegger, R., Manel, S., Schönswetter, P., Taberlet, P., Brodbeck, S., Gaudeul, M., Gielly, L., Küpfer, P., Mansion, G., Negrini, R., Paun, O., Pellecchia, M., Rioux, D., Schüpfer, F., Van Loo, M., Winkler, M., Gugerli, F. & IntraBioDiv Consortium. (2009). History or

ecology? Substrate type as a major driver of patial genetic structure in alpine plants. *Ecology Letters*, **12**, 632–640.

Ball, I., Possingham, H. & Watts, M.E. (2009). Marxan and relatives: software for spatial conservation prioritisation. *Spatial conservation prioritisation: Quantitative methods & computational tools* (eds A. Moilanen, K.A. Wilson & H. Possingham), pp. 185–189. Oxford University Press, Oxford, UK.

Barnes, M. (2015). Aichi targets: Protect biodiversity, not just area. *Nature*, **526**, 195–195.

Beyer, H.L., Dujardin, Y., Watts, M.E. & Possingham, H.P. (2016). Solving conservation planning problems with integer linear programming. *Ecological Modelling*, **328**, 14–22.

Blonder, B., Lamanna, C., Violle, C. & Enquist, B.J. (2014). The *n*-dimensional hypervolume. *Global Ecology and Biogeography*, **23**, 595–609.

Calenge, C. (2006). The package adehabitat for the R software: tool for the analysis of space and habitat use by animals. *Ecological Modelling*, **197**, 1035.

Carvalho, S.B., Brito, J.C., Crespo, E.J. & Possingham, H.P. (2011). Incorporating evolutionary processes into conservation planning using species distribution data: A case study with the western mediterranean herpetofauna. *Diversity & Distributions*, **17**, 408–421.

Center for International Earth Science Information Network (CIESIN), International Food Policy Research Institute (IFPRI), The World Bank & Centro Internacional de Agricultura Tropical. (2011). *Global rural-urban mapping project, Version 1 (GRUMP V1): Urban extents grid*. Socioeconomic Data and Applications Center (SEDAC), Palisades, NY, http://dx.doi.org/10.7927/H4GH9FVG.

Chernomor, O., Minh, B.Q., Forest, F., Klaere, S., Ingram, T., Henzinger, M. & Haeseler, A. von. (2015). Split diversity in constrained conservation prioritization using integer linear programming. *Methods in Ecology and Evolution*, **6**, 83–91.

Cornuéjols, G., Nemhauser, G.L. & Wolsey, L.A. (1990). The uncapacitated facility location problem. *Discrete location theory* (eds P.B. Mirchandani & R.L. Francis), pp. 119–171. Wiley, New York.

Cowling, R.M. & Pressey, R.L. (2001). Rapid plant diversification: Planning for an evolutionary future. *Proceedings of the National Academy of Sciences*, **98**, 5452–5457.

Cui, T.T., Ouyang, Y.F. & Shen, Z.J.M. (2010). Reliable facility location design under the risk of disruptions. *Operations Research*, **58**, 998–1011.

Engelbrecht, I., Robertson, M., Stoltz, M. & Joubert, J.W. (2016). Reconsidering environmental diversity (ED) as a biodiversity surrogacy strategy. *Biological Conservation*, **197**, 171–179.

Faith, D.P. (2003). Environmental diversity (ED) as surrogate information for species-level biodiversity. *Ecography*, **26**, 374–379.

Faith, D.P. (1994). Phylogenetic pattern and the quantification of organismal biodiversity. *Philosophical Transactions: Biological Sciences*, **345**, 45–58.

Faith, D.P. & Walker, P.A. (1993). DIVERSITY: A software package for sampling phylogenetic and environmental diversity. Reference and user's guide (version 1.0), http://australianmuseum.com/uploads/documents/faith%20diversity.pdf.

Faith, D.P. & Walker, P.A. (1996). Environmental diversity: On the best-possible use of surrogate data for assessing the relative biodiversity of sets of areas. *Biodiversity & Conservation*, **5**, 399–415.

Ferrier, S., Manion, G., Elith, J. & Richardson, K. (2007). Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions*, **13**, 252–264.

Frankham, R., Bradshaw, C.J. & Brook, B.W. (2014). Genetics in conservation management: Revised recommendations for the 50/500 rules, red list criteria and population viability analyses. *Biological Conservation*,

170, 56–63.

Gaston, K.J., Jackson, S.E., Cantu-Salazar, L. & Cruz-Pinon, G. (2008). The ecological performance of protected areas. *Annual Review of Ecology, Evolution and Systematics*, **39**, 93–113.

Gower, J.C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, **27**, 857–871.

Greenacre, M. & Primicerio, R. (2014). *Multivariate analysis of ecological data*. BBVA Foundation, Madrid, Spain.

Gugerli, F., Englisch, T., Niklfeld, H., Tribsch, A., Mirek, Z., Ronikier, M., Zimmermann, N., Holderegger, R. & Taberlet, P. (2008). Relationships among levels of biodiversity and the relevance of intraspecific diversity in conservation - a project synopsis. *Perspectives in Plant Ecology, Evolution and Systematics*, **10**, 259–281.

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

Klein, C., Wilson, K., Watts, M., Stein, J., Berry, S., Carwardine, J., Smith, M.S., Mackey, B. & Possingham, H. (2009). Incorporating ecological and evolutionary processes into continental-scale conservation planning. *Ecological applications*, **19**, 206–217.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. & Hornik, K. (2015). Cluster: Cluster analysis basics and extensions. R package version 2.0.3, http://cran.r-project.org/package=cluster.

Margules, C.R. & Pressey, R.L. (2000). Systematic conservation planning. *Nature*, **405**, 243–253.

McNeely, J.A. (1994). Protected areas for the 21st-century: Working to provide benefits to society. *Biodiversity and Conservation*, **3**, 390–405.

Meirmans, P., Goudet, J., IntraBioDiv Consortium & Gaggiotti, O. (2011). Ecology and life history affect different aspects of the population structure of 27 high-alpine plants. *Molecular Ecology*, **20**, 3144–3155.

Moilanen, A. (2007). Landscape Zonation, benefit functions and target-based planning: unifying reserve selection strategies. *Biological Conservation*, **134**, 571–579.

Moilanen, A., Anderson, B.J., Arponen, A., Pouzols, F.M. & Thomas, C.D. (2013). Edge artefacts and lost performance in national versus continental conservation priority areas. *Diversity and Distributions*, **19**, 171–183.

Moritz, C. (2002). Strategies to protect biological diversity and the evolutionary processes that sustain it. *Systematic Biology*, **51**, 238–254.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H. & Wagner, H. (2015). vegan: community ecology package. R package version 2.3-2, http://cran.r-project.org/package=vegan.

Oliver, M.A. & Webster, R. (1990). Kriging: A method of interpolation for geographical information systems. *International Journal of Geographical Information Systems*, **4**, 313–332.

Orians, G.H. (1993). Endangered at what level? *Ecological Applications*, **3**, 206–208.

Paradis, E., Claude, J. & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in r language. *Bioinformatics*, **20**, 289–290.

Pressey, R.L. & Logan, V.S. (1994). Level of geographical subdivision and its effects on assessments of reserve coverage: A review of regional studies. *Conservation Biology*, **8**, 1037–1046.

Pyke, C.R. & Fischer, D.T. (2005). Selection of bioclimatically representative biological reserve systems under climate change. *Biological Conservation*, **121**, 429–441.

R Core Team. (2016). R: A language and environment for statistical computing. http://www.R-project.org.

Rayfield, B., Moilanen, A. & Fortin, M.-J. (2009). Incorporating consumer-resource spatial interactions in

reserve design. *Ecological Modelling*, **220**, 725–733.

Raymond, B., VanDerWal, J. & Belbin, L. (2015). ALA4R: Atlas of Living Australia (ALA) Data and Resources in R. R package version 1.25, https://github.com/AtlasOfLivingAustralia/ALA4R.

Sanderson, E.W., Segan, D.B. & Watson, J.E.M. (2015). Global status of and prospects for protection of terrestrial geophysical diversity. *Conservation Biology*, **29**, 649–656.

Thomassen, H.A., Buermann, W., Milá, B., Graham, C.H., Cameron, S.E., Schneider, C.J., Pollinger, J.P., Saatchi, S., Wayne, R.K. & Smith, T.B. (2010). Modeling environmentally associated morphological and genetic variation in a rainforest bird, and its application to conservation prioritization. *Evolutionary Applications*, **3**, 1–16.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T. van de, Hornes, M., Friters, A., Pot, J., Paleman, J., Kuiper, M. & Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.

Watson, J.E., Dudley, N., Segan, D.B. & Hockings, M. (2014). The performance and potential of protected areas. *Nature*, **515**, 67–73.

Wilson, K.A., Westphal, M.I., Possingham, H.P. & Elith, J. (2005). Sensitivity of conservation planning to different approaches to using predicted species distribution data. *Biological Conservation*, **122**, 99–112.