Are environmental and geographic variables effective surrogates for genetic variation in conservation planning?

Jeffrey O. Hanson[1], *Jonathan R. Rhodes*[2], *Cynthia Riginos*[1], *Hugh P. Possingham*[1], *Richard A. Fuller*[1]

[1]*School of Biological Sciences, The University of Queensland, Brisbane, QLD, Australia*

[2]*School of Geography, Planning and Environmental Management, The University of Queensland, Brisbane, QLD, Australia*

Correspondance should be addressed to jeffrey.hanson@uqconnect.edu.au

20 June 2016

Abstract

Insert abstract here.

Short running title: Genetic surrogates in conservation

Word count: XXXXX

Number of references: XXXX

Number of figures: XXXX

Number of tables: XXXX

Number of text boxes: XXXX

Introduction

Protected areas spearhead conservation efforts (Sanderson *et al.* 2015). These places buffer species from anthropogenic impacts and provide places for species to persist. However, the resources available for conservation are limited. Thus protected areas need to be sited in places that achieve conservation goals for minimal cost. To ensure that the plans for protected areas (prioritizations) are cost-effective, they are often generated using decision support tools to identify near-optimal solutions. These prioritizations often contain plans for a network of multiple protected areas so they can preserve biological processes (eg. connectivity between populations). For instance, `Marxan` (Ball *et al.* 2009) uses amount-based targets to identify protected area networks that preserve a suitable amount of habitat for each species (eg. 100 km ^{2} of land occupied by the species needs to be secured in a reserve) and weights to penalize fragmented solutions (ie. the BLM parameter). However, conservation planning exercises typically assume that all individuals in the same species are equivalent.

Intra-specific genetic variation is important for long-term species persistence. As a consequence, there has been increasing interest in designing prioritizations that preserve and cultivate this variation (Moritz 1999; Crandall *et al.* 2000; Hendry *et al.* 2010). Although the strength of natural selection is a continuous force; broadly speaking, genetic variation can be classified as either adaptive or neutral (reviewed in Schoville *et al.* 2012). Adaptive genetic variation is associated with loci that (significantly) affect fitness. Typically, such "adaptive" loci are detected due to anomalous patterns among individuals (eg. Foll & Gaggiotti 2008; Duforet-Frebourg *et al.* 2014; Schoville *et al.* 2012; Whitlock 2015; but see Rellstab *et al.* 2015 for alternative methods). By preserving existing patterns of adaptive variation, protected areas can ensure that populations with particularly beneficial adaptations are not lost (Crandall *et al.* 2000). In contrast, neutral genetic variation is associated with loci that do not (significantly) affect fitness. Neutral variation often reflects the evolutionary history of different populations. By preserving existing patterns of neutral variation, protected areas can safeguard against the adverse effects of low genetic diversity (eg. inbreeding depression; Moritz 1999). Thus protected area networks should aim to preserve both adaptive and neutral patterns of genetic variation for species of conservation interest (Moritz 1999; Crandall *et*

al. 2000). Despite this, genetic data are not often used to inform conservation planning exercises because such data require considerable investment to obtain and analyze (Hendry *et al.* 2010; but see Moritz 1994 for developments on using genetic data in conservation planning)–especially since most large-scale exercises involve hundreds of species. Thus prioritizations may fail to secure a sufficient proportion of the intra-specific genetic variation of species, and ultimately fail to deliver long-term conservation outcomes.

Recently, conservation scientists have been investigating the use of surrogate data to generate prioritizations that adequately preserve intra-specific genetic variation–but without actually needing to utilize genetic data to acheive this. Since adaptive genetic variation is ultimately driven by extrinsic selection pressures, environmental variables have been used to guide reserve selection as surrogates for adaptive genetic variation (eg. environmental conditions; Pyke & Fischer 2005; Carvalho *et al.* 2011). By preserving a more representative sample of environmental conditions across the species' ranges, these researchers aimed to generate prioritizations that would preserve more adaptive variation. However, the effectiveness of this approach remains unverified. On the other hand, neutral genetic variation arises from a reduction in gene flow between individuals. Surrogates for neutral variation have been based on variables that predict the level of connectivity between planning units. For instance, the isolation by distance theory (IBD; Wright 1943) predicts that populations further apart will have less connectivity, exchange less genetic material, and in turn have less genetic variation in common. Based on this theory, geographic distance-based surrogates have been found to improve prioritizations in an island network (Ponce-Reyes *et al.* 2014). However this remains untested in spatially contiguous systems where connectivity is complicated by additional factors (eg. terrain or habitat type; Peterman *et al.* 2014a; Dudaniec *et al.* 2016)–the very systems commonly featured in real-world conservation planning exercises (eg. Cowling *et al.* 2003). Before conservation planners can reliably use surrogates for genetic data in conservation planning, the effectiveness of these surrogates need to be verified. Otherwise, conservation planners may waste precious resources on protecting places that provide little benefit to biodiversity.

Here we investigate the effectiveness of environmental and geographic surrogates for adaptive and neutral genetic variation in the context of conservation planning. First, we aim to verify if environmental and geographic variables correlate with broad-scale patterns in adaptive and neutral

variation (respectively). Although these variables may correlate with genetic variation, they may not be effective surrogates if they have poor predictive ability or if these correlations are driven by spatial auto-correlation. Second, we aim to verify that environmental and geographic variables are effective surrogates for adaptive and neutral variation (respectively). We expect that prioritizations generated using surrogate-based targets will secure more genetic variation than prioritizations generated using only amount-based targets with the same number of planning units. Third, we aim to determine if the addition of surrogate-based targets to conservation planning exercises can result in more effective prioritizations. We expect that prioritizations generated using amount- and surrogate-based targets will sufficiently preserve the genetic variation of more species than those generated using only amount-based targets. Furthermore, we aim to test this hypothesis under several of different scenarios to see whether the additional constraints of real-world planning exercises exacerbate or nullify this effect. If these surrogates can improve the intra-specific genetic variation captured in a prioritization, then conservation planners may use these surrogates when genetic data is not available, and in turn deliver more effective protected areas.

Methods

Study system

We used species distribution and genomic data collected by the international IntraBioDiv project in the European Alps (Meirmans *et al.* 2011; see Gugerli *et al.* 2008 for further explanation of data collection methods). This data set contains information for 27 alpine plant species (Figure 1a). It was collected using a $20'$ longitude by $21'$ latitude grid (approx. 22.3 km $\times$ 25 km; Figure S1). Project members visited every second grid cell, and if a species was detected in a cell, samples were collected from three individuals. They genotyped samples using amplified fragment length polymorphisms (AFLP; Vos *et al.* 1995), and constructed matrices denoting the presence/absence of polymorphisms at loci for each species (mean 130.7 $\pm$ 54.9 SD markers genotyped per species; see Table 1 in Alvarez *et al.* (2009) for the number markers for each species). Thus the data set contains information describing the genomic properties of individuals for each species present in every second grid cell.

We chose this data set for several reasons. First, conservation planning exercises typically consider tens to hundreds of species with different evolutionary histories and life histories. To ensure that our results were relevant for conservation planners, we required a data set with a diverse set of species and a comparable number of species. The IntraBioDiv data set contains genetic data for 27 species with different evolutionary histories and life-history strategies. Second, we required genetic data for all species to be collected using a standardized sampling scheme, and this data set used a grid to collect samples. Third, we required the study area to contain broad-scale environmental variation to test the effectiveness of environmental variables. This data set spans across large attitudinal gradients. Although this data set contains AFLPs which not ideal as genomic representatives (cf. single nucleotide polymorphisms; SNPs), no multi-species genomic data set with similar properties currently exists. Thus this data set was well suited for our study.

This data set has previously been used to address a range of different questions. Many studies have used this data set to explore patterns of adaptive (Manel *et al.* 2010, 2012; Meirmans *et al.* 2011; Bothwell *et al.* 2013; Rolland *et al.* 2015) and neutral genetic variation (Alvarez *et al.* 2009; Meirmans *et al.* 2011; Taberlet *et al.* 2012). However, to our knowledge, only a single study has used this data set to explore the effectiveness of a surrogate in the context of conservation planning. Taberlet *et al.* (2012) investigated whether prioritizations generated using the spatial distribution of a comprehensive set species (> 300) could adequately secure the intra-specific genetic variation for 27 species. They found that the prioritization based on species distributions was substantially different to the prioritization based on genomic data. Their results suggest that prioritizations based on species distribution data alone will be unable to secure intra-specific genetic variation. However, whether or not environmental and geographic variables can augment prioritizations remains untested.

Landscape data

We examined the effectiveness of environmental and geographic surrogates for adaptive and neutral genetic variation (respectively). We utilized the sampling grid used by the IntraBioDiv project as planning units to develop prioritizations. We used only planning units that contained at least a single individual ($n = 149$) to reduce computational burden. To examine the potential effects of variation in cost (see below), we calculated the total human population density inside each

planning unit (obtained at 1 km^2 resolution from the Global Rural-Urban Mapping Project; `GRUMP V1`; CIESEN, Columbia University *et al.* 2011) and used this to represent acquisition cost (Figure 1b).

To describe the geographic location of each planning unit, we projected the grid into an equidistant coordinate system (Europe Equidistant Conic: `ESRI:102031`), calculated their centroids, and extracted their two-dimensional coordinates. Thus we constructed two-dimensional geographic space–wherein each planning unit occupies a single point–as a potential surrogate for neutral genetic variation. We expected planning units further apart from each other in this geographic space to contain individuals with increasingly different patterns in their neutral loci.

To describe the environmental characteristics of each planning unit, we used broad-scale climatic variation. We obtained data for 19 bioclimatic variables across the extent of the study area (at $1''$ resolution; Hijmans *et al.* 2005). We used this data set because it is commonly to map habitat suitability in conservation planning (Rodríguez *et al.* 2007). These layers and the planning units were projected into an equal-area coordinate system (Europe Lambert Conformal Conic: `ESRI:102014`). To reduce dimensionality, the 19 bioclimatic variables were clipped to the planning units and subjected to a principal components analysis (PCA; using `ArcMap` version 10.2.2; Table S1). The first three principal components (PCs) cumulatively explained 99.2 % of the variation and were used to generate new layers. We calculated the average value for each PC layer in each planning unit, and used these values to characterize climatic variation among the planning units (Figure S2). Thus we constructed a three-dimensional environmental space–wherein each planning unit occupies a single point–as a potential surrogate for adaptive variation. We expect planning units further apart from each other in this environmental space to contain individuals with increasingly different patterns in their adaptive loci.

Adaptive and neutral genetic data

To investigate the effectiveness of our supposed surrogates for adaptive and neutral genetic variation, we first needed to identify which of the sampled loci–if any–were adaptive. We used two outlier detection methods to achieve this (an individual- and a population-level method as recommended by Villemereuil *et al.* 2014). The basic premise underpinning such methods is that neutral loci are

expected to exhibit a specific level of variation among populations, and loci that deviate from this expectation should be under selection (reviewed in Schoville *et al.* 2012). The advantage of these methods–in contrast with environmental association analyses (Rellstab *et al.* 2015)–is that they do not use environmental data to identify loci under selection, which would have biased our analysis. Loci identified as outliers in both analyses were treated as adaptive and the rest were treated as neutral. To avoid false-positives, we omitted loci from both outlier detection methods where the global frequency of the minor allele was > 0.1 and assumed they were neutral.

The first outlier detection method we used involved fitting multinominal-Dirichlet models implemented in `BayeScan` (version 2.1; Foll & Gaggiotti 2008). We adopted a similar methodology to Bothwell *et al.* (2013) and applied it to each of the species in the data set. Following their methodology, we first grouped conspecifics into genetic lineages to avoid falsely classifying loci as adaptive. To complete this first step, for each species, we fit the admixture and correlated alleles model implemented in `Structure` [version 2.3.4; r@461; r@462] using the number of populations for each species previously determined by [Alvarez *et al.* (2009); `structure.params.LST[[MODE]]$numruns` replicates per K; 5000 admixture burnin iterations; 300000 burnin iterations; 100000 iterations used for inference]. We used `ClumPP` (version 1.1.2; Jakobsson & Rosenberg 2007) to combine replicate runs to group individuals into populations (greedy algorithm using the G' static; 1000 iterations). We then ran `Bayescan` (Foll & Gaggiotti 2008) for each species using these populations [1:1 prior odds; 20 pilot runs; 100000 burn-in iterations; 10000 post-burn-in iterations thinned by 10 iterations]. Additionally, we omitted individuals from the `Bayescan` analysis if their population membership was uncertain (maximum membership probability < 0.75). We ran 4 replicates per species using a suitable false discovery rate ($q \leq 0.1$; Benjamini & Hochberg 1995).

The second outlier detection method involved fitting PCAs to identify outlier loci (using methods implemented in the `pcadapt R` package; version 3.0; Duforet-Frebourg *et al.* 2014; Luu *et al.* 2016). To enable comparisons between the two outlier detection analyses, we used the same individuals in this analysis as used in the `BayeScan` analysis. For each of the 27 species, we first imputed missing data to permit further analysis. Specifically, the missing value for a given locus for a given individual was replaced with the average frequency of that locus among all conspecifics. Next, we ran a PCA and extracted the minimum number of components needed to secure the over-arching population

level variation among the loci (1 % of the variation in loci). We then used Mahalanobis distances (Mahalanobis 1936) to compute P values and then q-values (using the `qvalue R` package; Andrew J. Bass *et al.* 2015). Outlier loci were detected using a suitable false discovery rate ($q < 0.1$).

After classifying loci as adaptive or neutral, we mapped the main gradients of the adaptive (if detected) and neutral genetic variation for each species. For each species, we discarded the population groupings, and partitioned adaptive and neutral loci into separate matrices. We applied separate non-metric multi-dimensional scaling analyses to each matrix (NMDS; implemented in the `vegan R` package; Oksanen *et al.* 2015) using Gower distances (Gower 1971; using the `cluster R` package to accommodate sparsity; Maechler *et al.* 2015) to derive continuous variables that described the main gradients of adaptive and neutral genetic variation (separately) for each species (Table S2). To ensure that the ordinations described a sufficient amount of the genomic variation, we ran successive scaling analyses with increasing dimensionality until a sufficient stress value was obtained (maximum stress value ≤ 0.25; 100 random starts for each analysis). Since each grid cell had up to three samples per species, we used the average of the ordinated values associated with the individuals in each grid cell to express the typical genomic characteristics of individuals in the cells.

The previous set of analyses resulted in an adaptive (if detected) and neutral genetic space for each species. For a given species, each planning unit occupied by the species was associated with a multi-dimensional point in the species' neutral genetic space. Planning units that were closer together in this space are occupied by individuals with similar polymorphisms in their neutral loci. Similarly, if the species was associated with adaptive loci, each planning unit occupied by the species was associated with a multi-dimensional point in the species' adaptive genetic space. By spreading out conservation effort across a given genetic space, and in turn securing planning units occupied by individuals with increasingly different polymorphisms, prioritizations can secure more genetic variation (see Hanson *et al.* 2016 for discussion on attribute spaces).

Prioritization method

We used the unreliable representative and adequate prioritization (URAP) formulation of the reserve selection problem (Hanson *et al.* 2016). This formulation can identify the minimum number of

planning units required to preserve both a proportion of the species' range (using amount-based targets) and a proportion of variation found across the species' range (using space-based targets and attribute spaces; *sensu* Hanson *et al.* 2016). We treated the environmental and geographic surrogate variables as separate attribute spaces. We also treated the adaptive (if detected) and neutral variables as separate attribute spaces for each species. We set the demand points for each species and each attribute space using the values associated with the planning units. For every prioritization we generated, we calculated the proportion of the adaptive (if detected) and neutral genetic variation that every prioritization secured to assess its performance. Prioritizations associated with negative values–because they secured only a very small amount of genetic variation–were replaced with zeros to facilitate statistical analysis. We solved all reserve selection problems to within 10 % of optimality using the `rapr` R package (Hanson *et al.* 2016) and `Gurobi` (Gurobi Optimization 2015).

Statistical analyses

To address the first aim–to determine if the environmental and geographic variables correlate with spatial patterns of adaptive and neutral genetic variation–we computed dissimilarity matrices between the planning units, and tested for correlations between matrices based on surrogate and genetic data. For each species, we computed two dissimilarity matrices between the planning units occupied by the species in terms of their environmental conditions (their coordinates in the environmental space) and geographic position (their coordinates in the geographic space). We then computed another dissimilarity matrix between the planning units using their values in the species' neutral genetic space. If the species was associated with adaptive loci, we also computed another dissimilarity matrix between the planning units using the species' adaptive genetic space. All dissimilarity matrices calculated using Euclidean distances.

We then fit maximum likelihood population-effects (MLPE) models (Clarke *et al.* 2002; implemented in the `ResistanceGA R` package; Peterman *et al.* 2014b). These models explicitly accommodate the structure of dissimilarity matrices using random effects. For each species, we fit a MLPE model to correlate the species' dissimilarity matrices based on geographic position and neutral genetic variation. Additionally, if the species was associated with adaptive loci, we also fit a MLPE model to correlate the species' dissimilarity matrices calculated using environmental conditions and adaptive

genetic variation. All predictor variables were z-transformed prior to model fitting to improve convergence. To test if the surrogate variables explained a substantial amount of genetic variation, we conducted χ^2 tests between each model and its corresponding null model, and applied Bonferroni corrections.

To address the second aim–to determine if the environmental and geographic variables are effective surrogates for adaptive and neutral genetic variation–we generated a collection of single-species prioritizations and computed the proportion of intra-specific genetic variation they secured. For each species, we generated prioritizations by randomly selecting different combinations of planning units that were occupied by the species. We then computed the proportion of adaptive genetic variation and environmental variation secured in each prioritization. We then generated another random set of prioritizations, and computed the proportion of neutral genetic variation and geographic variation secured in each prioritization.

We fit two full generalized linear models (GLM) with a logit links. The first model was fit to the proportion of adaptive genetic variation secured in a prioritization, and contained a continuous predictor variable measuring the proportion of environmental variation also secured in the prioritization, a categorical variable indicating the species for which the prioritization was generated, and an interaction term between these predictor variables. The second model was fit to the proportion of neutral genetic variation secured in a prioritization, and contained a continuous predictor variable assessing the proportion of geographic variation also secured in the prioritization. Similar to the previous model, this model also contained a predictor variable indicating the species for which the prioritizations were generated, and an interaction term. We subjected these models to a backwards step-wise term deletion routine to assess term significance. To gauge the performance of these surrogates for each species, we fit a suite of models correlating the proportion of adaptive and neutral genetic variation held in each prioritization and the environmental and geographic variation they also secured (respectively). We then computed the Cragg and Uhler's pseudo-R^2 value for each model (using the `pscl` R package; Jackman 2015).

To address the third aim–to determine if the use of surrogate-based targets can actually improve prioritizations–we investigated the effectiveness of surrogate-based targets under three scenarios with increasing levels of complexity. To assess the benefits of surrogate-based targets under the

simplest of circumstances, our first scenario involved generating single-species prioritizations for each species. However, most large-scale prioritizations involve a comprehensive set of species. By forcing solutions to secure individuals in multiple communities, this may result in prioritizations that secure a greater proportion of intra-specific genetic variation for each species. To test for this effect, our second scenario involved generating multi-species prioritizations. Finally, most large-scale prioritizations also explicitly consider opportunity cost to ensure that prioritizations are cost-effective. Our third scenario involved generating multi-species prioritizations that satisfied targets for minimal opportunity cost.

For each scenario, we generated three types of prioritizations using (1) only amount-based targets, (2) both amount- and surrogate-based targets, and (3) both amount- and genetic-based targets. The purely amount-based prioritizations served as our baseline–representing the prioritizations typically generated in conservation planning exercises (eg. by `Marxan`). We used 20 % amount-based targets in all prioritizations to ensure an adequate proportion of habitat was secured for each species. Based on the results from the analyses used to address our second aim, we set surrogate-based targets as 95 % and genetic-based targets as 80 % to ensure that most of the genetic variation was secured for each species. To accommodate the fact that there were a large number of optimal solutions for the single-species amount-based prioritizations, we generated 1000 replicate prioritizations. We computed the proportion of adaptive (if present) and neutral genetic variation secured for each species in the prioritizations.

We investigated the benefits of using surrogate-based targets in prioritizations using generalized linear mixed-effects models fit with logit liknks (GLMMs; using the `lme4 R` package; **???**). We fit a full model to the proportion of genetic variation secured in a given prioritization. This model contained categorical predictor variables to indicate the combination of targets used to generate the prioritizations (amount-only, amount and surrogate targets, and amount and genetic targets), the scenario that was used to generate the prioritization (single-species, multi-species, and multi-species with cost), the type of genetic variation measured (adaptive and neutral), and all interactions between these variables. This model accommodated prioritizations generated for the same species using a random intercept term. We subject the full model to a backwards step-wise term deletion routine to assess term significance. We conducted a post-hoc analysis on the minimal adequate

model using Tukey contrasts with a Bonferroni correction (using the `multcomp R` package; Hothorn *et al.* 2008).

Results

We successfully mapped the main spatial patterns of adaptive (if present) and neutral variation for each of the 26 species. The outlier detection methods detected adaptive genetic variation in twelve species. Generally, the two outlier detection methods gave fairly consistent results (mean 93.81 % loci per species ± 4.13 SD). Of the species that were associated with adaptive genetic variation, only a few loci were detected as under selection (mean 1.96 % loci per species ± 1.54 SD). After identifying the loci under selection, we used NMDS analyses to identify the main gradients of adaptive (if detected) and neutral variation for each species. For all species in which we detected adaptive loci, only a couple of continuous dimensions were needed to sufficiently describe their patterns of adaptive genetic variation (Table S2). On the other hand, a few additional continuous dimensions were occasionally needed to describe their patterns of neutral genetic variation (minimum $K = 2$; mean $K = 2.52$; max $K = 4$; Table S2). The ordinations resulting from the NMDS analyses were used to construct an adaptive (if present) and neutral genetic space for each species. The spatial distribution of these genetic spaces (Figures S29–S54) generally show patterns of spatial autocorrelation–especially the adaptive genetic spaces (eg. *Cerastium uniflorum* and *Dryas octopetala*). Additionally, several species show east-west patterns in their intra-specific genetic variation (eg. *Peucedanum ostruthium* and *Phyteuma betonicifolium*).

The dissimilarity matrices generated using the environmental and geographic variables correlated with those matrices based on the genetic characteristics for most species (Table 1). Specifically, environmental variation was found to significantly correlate with adaptive genetic variation for nine species (75 % associated with loci under selection; $P < 0.05$; mean 0.06 marginal R^2 ± 0.06 SD for significant models). Additionally, geographic distances between planning units were found to significantly correlate with neutral genetic variation among planning units for 26 species (96.3 % of all the species investigated; $P < 0.05$; mean 0.2 marginal R^2 ± 0.16 SD for significant models). These results suggest that prioritizations with different environmental characteristics tend to have individuals with different adaptive genetic characteristics. Similarly, planning units that are further

apart tend to have individuals with different neutral genetic characteristics. After verifying that the geographic and environmental variables generally correlated with genetic variation, our next step was to determine how effective these variables were as surrogates in the context of conservation planning.

The environmental and geographic variables were moderately effective surrogates for genetic variation in most species (Figure 2). The relationship between the proportion of adaptive genetic variation secured in a prioritization and the proportion of environment variation it also secured was found to depend on the species for which the prioritization was generated (${\chi^2}_{11} = 1.283587 \times 10^4; P > 0.99$). The slope for each species were all significantly different to zero (minimum $z = 0.15$; maximum $P > 0.99$; Table S3). The explanatory power of the environmental surrogates were also generally quite high for most species (mean 0 $R^2 \pm 0$ SD; Figure S56). The nature of these relationships suggest that most of the adaptive variation for a given species can be secured in a prioritization using moderately high environmental-based targets (> 80 %). Similarly, the proportion of neutral genetic variation in a prioritization was affected by the interaction between species and the proportion of neutral variation secured in the prioritization (${\chi^2}_{26} = 1.780429 \times 10^4$; $P > 0.99$). The slope for each species were also significantly different to zero (minimum $z = 0.02$; maximum $P > 0.99$; Table S3). Generally, the explanatory power of geographic variables were higher than that of environmental variables (mean 0 $R^2 \pm 0$ SD; Figure S57). Compared to the environmental variables, the slopes of these relationships suggest that very high geographic-based targets are required to generate prioritizations that secure a sufficient proportion of neutral genetic variation for most species (> 95 %). Additionally, all species were associated with prioritizations that secured a large proportion of environmental or geographic variation (> 70 %), yet preserved only a very small proportion of their intra-specific genetic variation (< 10 %; Figures S55 and S56; such as *Gypsophila repens* and *Luzula alpinopilosa*). These results suggest while prioritizations generated using environmental and geographic targets tend to secure more genetic variation, prioritizations generated using only these targets can sometimes secure only a small fraction of intra-specific genetic variation.

To determine if the addition of environmental and geographic targets could improve the effectiveness of prioritizations in a more realistic context, we generated prioritizations according to three scenarios using different combinations of targets (Figure 3). We then computed the proportion of adaptive

and neutral genetic variation secured in each prioritization (Figure 4). The proportion of genetic variation secured in a prioritization was found to depend on the combination of targets used to generate the prioritization (${\chi^2}_2 = 16.35$; $P < 0.001$). Specifically, the amount-based prioritizations tended to secure less variation than the prioritizations generated using additional surrogate (100 % mean difference ± 100 SE ; $Z_1 = 0.28$; $P > 0.99$) or genetic variables (86.93 % mean difference ± 74.05 SE; '$Z_1 = 1.81$; P0.21). No evidence was found to suggest that the prioritizations that included additional surrogate- or genetic-based targets secured a different proportion of genetic variation (43.89 % mean difference ± 100 SE; $Z_1 = -0.25$; $P =>$ 0.99). We also found that the prioritizations tended to secure more adaptive than neutral genetic variation (10.87 % mean difference ± 52.63 SE; ${\chi^2}_1 = 4.46$; P0.035). We also found significant differences in the average proportion of genetic variation secured by prioritizations in each of the scenarios (${\chi^2}_3 = 5.35$; P 0.15). Specifically, both the multi-species (50 % mean difference ± 51.38 SE; $Z_1 = 0$; $P =>$ 0.99) and multi-species with cost (68.55 % mean difference ± 70.38 SE; $Z_1 = 0.9$; $P > 0.99$) scenarios were found to secure more genetic variation than the single species scenario. Neither the multi-species scenarios were found to secure a different proportion of genetic variation (89.75 % mean difference ± 79.3 SE; $Z_1 = 1.61$; P0.64). These results suggest that surrogate-based targets can substantially improve the proportion of intra-specific genetic variation secured in a prioritization for most species.

Discussion

- introduction We aimed to investigate the effectiveness of broad-scale environmental and geographic variables as surrogates for adaptive and neutral genetic variation in conservation planning. Our study is the first to address this hypothesis using moderately realistic planning scenarios. Overall, we found that the spatial patterns of environmental and geographic variables correlate with those of adaptive and neutral patterns for most species. Furthermore, we found that the addition of environmental and geographic targets to the planning process resulted in prioritizations that generally secured a greater proportion of adaptive and neutral genetic variation.

- consequences for conservation planning

- outlier detection methods
- mapping genetic variation
- limitations/future directions We wish to alert our readers to several limitations associated with our analysis. First, the size of the planning units we used (23 × 25 km) is much larger than typically used in conservation planning scenarios (between 1–10 kmˆ2). We used this resolution because the genetic data were collected at this scale. Whilst we could have interpolated the genetic data to a finer resolution and used smaller planning units, this would have biased our analysis because this process would have artificially introduced additional spatial auto-correlation into the dataset. Second, we used geographic distances as surrogates for connectivity between planning units. Although connectivity between areas can be modeled using isolation by resistance measures (McRae 2006) derived from landscape data (eg. Peterman *et al.* 2014a), they are unlikely to be of use in broad-scale conservation exercises since they need to be tuned using species-specific parameters. Third, we used AFLP data to characterize the main gradients of genetic variation. Whilst next-generation sequencing provides higher resolution genetic information than AFLP data (reviewed in Schoville *et al.* 2012), our methodology would still have utilized only the main gradients of the genetic variation in order to generate prioritizations in a feasible amount of time. Overall, we are confident that the limitations associated with our study are relatively minor.
- Concluding remarks The overarching aim of our study was to investigate the effectiveness of surrogates in conservation planning. We found that environmental and geographic are generally effective surrogates for most species. Although prioritizations generated using amount-only targets tended to secure most of the adaptive variation for most species, the addition of surrogate targets to the planning process substantially improved the proportion of intra-specific neutral genetic variation. Conservation planners are therefore urged to include these surrogates in conservation planning exercises.

Acknowledgements

JOH is funded by an Australian Postgraduate Award (APA) scholarship. RAF has an Australian Research Council Future Fellowship. This work was supported by the Centre of Excellence for Environmental Decisions (CEED) and the Landscape Ecology and Conservation Group (LEC) at The University of Queensland.

References

1. Alvarez, N., Thiel-Egenter, C., Tribsch, A., Holderegger, R., Manel, S., & Schönswetter, P. *et al.* (2009). History or ecology? Substrate type as a major driver of patial genetic structure in alpine plants. *Ecology Letters*, 12, 632–640.

2. Andrew J. Bass, J. D. S. with contributions from, Dabney, A., & Robinson, D. (2015). *Qvalue: Q-value estimation for false discovery rate control.*

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

4. Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57, 289–300.

5. Bonin, A., Nicole, F., Pompanon, F., Miaud, C., & Taberlet, P. (2007). Population adaptive index: A new method to help measure intraspecific genetic diversity and prioritize populations for conservation. *Conservation Biology*, 21, 697–708.

6. Bothwell, H., Bisbing, S., Therkildsen, N., Crawford, L., Alvarez, N., & Holderegger, R. *et al.* (2013). Identifying genetic signatures of selection in a non-model species, alpine gentian (gentiana nivalis l.), using a landscape genetic approach. *Conservation Genetics*, 14, 467–481.

7. 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.

8. CIESEN, Columbia University, International Food Policy Research Institute, The World Bank, & Centro Internacional de Agricultura Tropical. (2011). Global rural-urban mapping project, version 1 (gRUMP v1): Urban extents grid.

9. Clarke, R. T., Rothery, P., & Raybould, A. F. (2002). Confidence limits for regression relationships between distance matrices: Estimating gene flow with distance. *Journal of Agricultural, Biological, and Environmental Statistics*, 7, 361–372.

10. Cowling, R. M., Pressey, R. L., Rouget, M., & Lombard, A. T. (2003). A conservation plan for a global biodiversity hotspot - the cape floristic region, south africa. *Biological Conservation*, 112, 191–216.

11. Crandall, K. A., Bininda-Emonds, O. R. P., Mace, G. M., & Wayne, R. K. (2000). Considering evolutionary processes in conservation biology. *Trends in Ecology & Evolution*, 15, 290–295.

12. Dudaniec, R. Y., Worthington Wilmer, J., Hanson, J. O., Warren, M., Bell, S., & Rhodes, J. R. (2016). Dealing with uncertainty in landscape genetic resistance models: A case of three co-occurring marsupials. *Molecular Ecology*, 25, 470–486.

13. Duforet-Frebourg, N., Bazin, E., & Blum, M. G. (2014). Genome scans for detecting footprints of local adaptation using a bayesian factor model. *Molecular Biology and Evolution*, 31, 2483–2495.

14. Foll, M. & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A bayesian perspective. *Genetics*, 180, 977–993.

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

16. Gugerli, F., Englisch, T., Niklfeld, H., Tribsch, A., Mirek, Z., & Ronikier, M. *et al.* (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.

17. Gurobi Optimization, I. (2015). Gurobi optimizer reference manual.

18. Hanson, J. O., Rhodes, J. R., Possingham, H. P., & Fuller, R. A. (2016). RAPR: representative

and adequate prioritisations in R. *Unpublished*, X, XX–XX.

19. Hendry, A. P., Lohmann, L. G., Conti, E., Cracraft, J., Crandall, K. A., & Faith, D. P. *et al.* (2010). Evolutionary biology in biodiversity science, conservation, and policy: A call to action. *Evolution*, 64, 1517–1528.

20. 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.

21. Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50, 346–363.

22. Jackman, S. (2015). *pscl: Classes and methods for R developed in the political science computational laboratory, stanford university.* Department of Political Science, Stanford University, Stanford, California.

23. Jakobsson, M. & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23, 1801–1806.

24. Luu, K., Blum, M. G., & Duforet-Frebourg, N. (2016). *Pcadapt: Fast principal component analysis for outlier detection.*

25. Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2015). *Cluster: Cluster analysis basics and extensions.*

26. Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Sciences, Calcutta*, 2, 49–55.

27. Manel, S., Gugerli, F., Thuiller, W., Alvarez, N., Legendre, P., & Holderegger, R. *et al.* (2012). Broad-scale adaptive genetic variation in alpine plants is driven by temperature and precipitation. *Molecular Ecology*, 21, 3729–3738.

28. Manel, S., Poncet, B. N., Legendre, P., Gugerli, F., & Holderegger, R. (2010). Common factors drive adaptive genetic variation at different spatial scales in arabis alpina. *Molecular Ecology*, 19,

3824–3835.

29. McRae, B. H. (2006). Isolation by resistance. *Evolution*, 60, 1551–1561.

30. 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.

31. Moritz, C. (1994). Defining evolutionarily significant units for conservation. *Trends in Ecology & Evolution*, 9, 373–375.

32. Moritz, C. (1999). Conservation units and translocations: Strategies for conserving evolutionary processes. *Hereditas*, 130, 217–228.

33. Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., & O'Hara, R. B. *et al.* (2015). *Vegan: Community ecology package*.

34. Peterman, W. E., Anderson, T. L., Ousterhout, B. H., Drake, D. L., Semlitsch, R. D., & Eggert, L. S. (2014a). Differential dispersal shapes population structure and patterns of genetic differentiation in two sympatric pond breeding salamanders. *Conservation Genetics*, 16, 59–69.

35. Peterman, W. E., Connette, G. M., Semlitsch, R. D., & Eggert, L. S. (2014b). Ecological resistance surfaces predict fine-scale genetic differentiation in a terrestrial woodland salamander. *Molecular Ecology*, 23, 2402–2413.

36. Ponce-Reyes, R., Clegg, S. M., Carvalho, S. B., McDonald-Madden, E., & Possingham, H. P. (2014). Geographical surrogates of genetic variation for selecting island populations for conservation. *Diversity & Distributions*, 20, 640–651.

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

38. Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, 24, 4348–4370.

39. Rodríguez, J. P., Brotons, L., Bustamante, J., & Seoane, J. (2007). The application of predictive modelling of species distribution to biodiversity conservation. *Diversity and Distributions*, 13,

243–251.

40. Rolland, J., Lavergne, S., & Manel, S. (2015). Combining niche modelling and landscape genetics to study local adaptation: A novel approach illustrated using alpine plants. *Perspectives in Plant Ecology, Evolution and Systematics*, 17, 491–499.

41. 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.

42. Schoville, S. D., Bonin, A., Francois, O., Lobreaux, S., Melodelima, C., & Manel, S. (2012). Adaptive genetic variation on the landscape: Methods and cases. In: *Annaul review of ecology, evolution, and systematics*, Annual review of ecology evolution and systematics (ed. Futuyma, D.). pp. 23–43.

43. Taberlet, P., Zimmermann, N. E., Englisch, T., Tribsch, A., Holderegger, R., & Alvarez, N. *et al.* (2012). Genetic diversity in widespread species is not congruent with species richness in alpine plant communities. *Ecology Letters*, 15, 1439–1448.

44. Villemereuil, P. de, Frichot, É., Bazin, É., François, O., & Gaggiotti, O. E. (2014). Genome scan methods against more complex models: When and how much should we trust them? *Molecular Ecology*, 23, 2006–2019.

45. Vos, P., Hogers, R., Bleeker, M., Reijans, M., Lee, T. van de, & Hornes, M. *et al.* (1995). AFLP: A new technique for dNA fingerprinting. *Nucleic Acids Research*, 23, 4407–4414.

46. Whitlock, M. C. (2015). Modern approaches to local adaptation. *The American Naturalist*, 186, S1–S4.

47. Wright, S. (1943). Isolation by distance. *Genetics*, 28, 114–138.