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

Jeffrey O. Hanson, Jonathan R. Rhodes, Cynthia Riginos, Hugh P. Possingham, Richard A. Fuller  
School of Biological Sciences, The University of Queensland, Brisbane, QLD, Australia  
School of Geography, Planning and Environmental Management, The University of Queensland, Brisbane, QLD, Australia  
Correspondance should be addressed to [jeffrey.hanson@uqconnect.edu.au](mailto:jeffrey.hanson@uqconnect.edu.au)

24 January 2016

Insert abstract here.

## Introduction

Biodiversity is under threat. Over the last few decades, many species have gone extinct (Thomas *et al.* 2004; Barnosky *et al.* 2011). Today, thousands of species are listed as endangered or near threatened in the International Union for Conservation of Nature (IUCN) red list (<www.iucnredlist.org>). In the future, species are expected to face even more threats (eg. climate change; Thomas *et al.* 2004; IPCC 2007). In response to this crisis, the overarching aim of conservation is to maximize the long-term persistence of biodiversity (Margules and Pressey 2000; Moritz 1999).

The act of setting aside areas for preservation is the hallmark of modern conservation (Sanderson *et al.* 2015). Protected areas buffer species from anthropogenic impacts and provide places for biodiversity processes to persist (eg. nutrient cycling and water purification). Additionally, protected areas set the stage for direct management actions (eg. habitat rehabilitation, translocation, and invasive species management). Since the resources available for conservation are limited, the plans for protected areas (prioritizations) are often generated using spatially explicit biological data (eg. species distribution data) and decision support tools (eg. ; Ball *et al.* 2009; ; Moilanen 2007) that identify cost-effective solutions.

Protected areas must preserve intra-specific genetic variation to maximize species recovery (Moritz 1999; Crandall *et al.* 2000; Hendry *et al.* 2010). This variation can be classified as either adaptive or neutral (reviewed in **???**). Adaptive genetic variation is associated with loci currently under selection, whereas neutral genetic variation is associated with loci that are not. 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 variation reflects the evolutionary history of different populations. By preserving existing patterns of neutral variation, protected areas can secure different genetic lineages and avoid the adverse effects of low genetic diversity (eg. inbreeding depression; Moritz 1999). However, genetic data requires considerable investment to obtain and analyze, and as a consequence, conservation planners do not typically use genetic data to inform prioritizations (Hendry *et al.* 2010). Thus many prioritizations may fail to capture species’ intra-specific genetic variation, and ultimately fail to deliver conservation outcomes.

Recently, conservation scientists have begun investigating the use of surrogates for genetic data to generate more effective prioritizations without the overheads associated with genetic data. Since adaptive genetic variation is driven by selection pressures, environmental variables have been explored as potential surrogates for genetic variation (eg. Pyke and Fischer 2005; Carvalho *et al.* 2011). On the other hand, neutral variation arises from a reduction in gene flow, and so variables that predict reduction in gene flow between areas have been explored. For instance, the isolation by distance theory (IBD; Wright 1943) predicts that populations further apart will have lower gene flow, and in turn have greater neutral variation among them. The use of distance-based surrogates has been found to improve prioritizations in an island network (Ponce-Reyes *et al.* 2014). However the effectiveness of distance-based surrogates remains untested in systems where connectivity is complicated by additional factors (eg. elevation in terrestrial systems). Before conservation planners can reliably use surrogates for genetic data in conservation planning, the effectiveness of these surrogates needs to be verified. Otherwise, conservation planners may waste precious resources on securing places that do little to benefit biodiversity.

Our overarching aim is to investigate the effectiveness of environmental and geographic surrogates for adaptive neutral genetic variation in the context of conservation planning. First, we aim to test the effectiveness of these surrogates under the simplest of conditions by using a set of single-species prioritizations. Second, we aim to test the effectiveness of these surrogates in a more realistic multi-species context, and see how they compare to the single-species prioritizations. Third, we aim to understand the relationship between the proportion of surrogate variation secured in a multi-species prioritization and the proportion of genetic variation secured in a prioritization. If these surrogates can be shown to markedly improve the proportion of genetic variation secured 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

To address the aims of the study, we utilized species distribution and genomic data collected by the IntraBioDiv project (Meirmans *et al.* 2011; see Gugerli *et al.* 2008 for methodological details). This dataset was well suited for this study, because it provides spatially explicit genomic data for a multitude of species across a broad geographic area and over large environmental gradients. As part of this project, data for 27 alpine plant species were collected across the European Alps according to a grid (22.3 km 25 km; Figure S1). Briefly, in every second cell, plant samples were collected from three individuals for each species present inside the cell. Cells that did not contain any plant samples were omitted from our analysis. Samples were genotyped using amplified fragment length polymorphisms (AFLP; Vos *et al.* 1995). Matrices denoting the presence/absence of polymorphisms at loci were constructed independently for each species. Thus for each second grid cell, for each of the species in the grid cell, the dataset contains information describing the genomic variation of individuals inside the grid cell. We used these grid cells as planning units to develop prioritizations to investigate the effectiveness of surrogates for the (AFLP) genetic variation. Most of our analyses described hereafter were performed using the R software package (version 3.2.3; R Core Team 2014).

### Surrogate data

We explored the effectiveness of climatic and geographic surrogates. To describe variation in the location of each grid cell, we projected the grid into an equi-distant coordinate system () and used the centroid of each cell. To describe the climatic variation among each cell, we obtained 21 bioclimatic variables (approx. 1 km resolution; Hijmans *et al.* 2005). These layers and the grid cells were projected into an equal-area coordinate system (). The bioclimatic variables were clipped to the extent of the grid and subject to a principle components analysis (PCA; Table S1). The first two principle components (PCs) cumulatively explained 97.2 % of the variation and were used to generate two new layers. We calculated the average value for each PC layer in each grid cell, and used these values to characterize the main themes of climatic variation among the grid cells (Figure S2).

### Outlier locus detection

To investigate the effectiveness of the surrogates for adaptive and neutral genetic variation, we first identified which of the sampled loci for each species were under selection. Specifically, we used outlier locus detection methods (evaluated in Pérez-Figueroa *et al.* 2010) to avoid circularity issues. The basic premise underpinning these analyses 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. In an earlier study, Bothwell *et al.* (2013) applied such outlier detection methods to one of the species in this dataset (*Getiana nivalis*). Here, we adopted a similar methodology and applied it to each of the 27 species in the dataset.

First, we grouped individuals into genetic lineages for each species. The purpose of this step was to avoid misclassifying loci as adaptive. We used discriminant analyses of principle components (DAPC; Jombart *et al.* 2010) to assign individuals to populations (implemented in R package; Jombart 2008). Briefly, for each species, we transformed the binary loci data using a principle components analysis (PCA) and obtained the principle components required to explain at least 90 % variation. We subject these components to multiple *k*-means clustering models with varying numbers of assumed populations (*k* = 1–10; 1000 iterations; 1000 random starts). We then applied Ward’s clustering method to the Bayesian Information Criterion (BIC) statistics associated with each model to partition the models into two groups, and selected the model before the first group switch (as recommended in the adegenet manual; Figure S4). The populations memberships associated with these models were used for subsequent analysis (Figure S5).

We identified loci under selection for each species (Table S2). For each species, we ran (version 2.1; Foll and Gaggiotti 2008) using the population memberships identified in the previous analysis (1:1 prior odds; 20 pilot runs; 50000 burn-in iterations; 5000 post-burn-in iterations thinned by 10 iterations). We omitted loci where the global frequency of the minor allele was to avoid false-positives (following guidelines in the manual). To ensure convergence, we ran 4 replicates per species using a suitable false discovery rate (FDR; ; Benjamini and Hochberg 1995).

After classifying loci as adaptive or neutral, we mapped the main gradients of the adaptive (if detected) and neutral genetic variation for each species (Figures S6–S32). For each species, for each type of genetic variation detected, we used Gower distances (Gower 1971) to express the genomic differences between individuals (using the R package to accommodate sparsity; Maechler *et al.* 2015). These distances were subject to multiple non-metric multi-dimensional scaling analyses (NMDS; implemented in the R package; Oksanen *et al.* 2015) using the minimum number of dimensions required to achieve an adequate stress value ( 0.1; 100 random starts for all models; Table S3). Since each grid cell had up to three samples per species, we used the average of the ordinated values associated with the samples in each grid cell to characterize their the genetic variation of individuals in the cells.

### Prioritizations

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 a proportion of the species’ range as well as a proportion of the species’ variation using an attribute space (*sensu* Hanson *et al.* 2016). We treated the genomic and surrogate variables as distinct attribute spaces. We set the demand points for each species and its attribute space using the values associated with the planning units. Although Hanson *et al.* (2016) recommend against using the same coordinates for demand points and planning units for real-word conservation planning exercises, we did this to provide a more robust test and reduce uncertainty in the analysis. 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. We solved all reserve selection problems to within 5 % of optimality using the R package (Hanson *et al.* 2016) and (Gurobi Optimization 2015).

To understand how prioritizations generated using surrogate-based targets might perform under typical conservation planning scenarios, we generated a set of single-species and multi-species prioritizations. All of these prioritizations used 30 % amount-based targets to ensure an adequate proportion of habitat was secured for each species (Cowling *et al.* 2003; Margules and Pressey 2000; Klein *et al.* 2009). We generated a set of single-species prioritizations for each species: 1000 prioritizations using only the amount-based targets, a single prioritization using the amount-based target and 90 % surrogate-based targets, and a single prioritization using the amount-based target and 90 % genetic-based targets. We also generated three multi-species prioritizations using the three unique combinations of targets used to generate the single-species prioritizations. Note that we obtained multiple solutions for the amount-based single-species prioritizations because, unlike the other prioritizations, these targets yielded many optimal solutions with varying levels effectiveness.

To understand the relationship between the surrogate-based targets and the proportion of genetic variation secured, we generated two sets of multi-species prioritizations with increasing surrogate-based targets between 1 % and 100 % in increments of 11 %. The first set of prioritizations contained prioritizations with varying environmental surrogate-based targets with no specified geographic targets. The second contained prioritizations with varying geographic surrogate-based targets and no environmental targets. None of these prioritizations used amount-based targets.

## Statistical analysis

We tested for differences in the effectiveness of prioritizations representing typical scenarios using generalized linear mixed-effects models (GLMMs; using the R package; Bates *et al.* 2015). We fit two full models to the proportion of genomic variation secured in the single- and multi-species prioritizations using a logit link. Both models contained a predictor variable indicating if the prioritization was generated using amount, surrogate, or genetic targets, a second predictor variable indicating if the genetic variation was adaptive or neutral, and the interaction between these variables. Additionally, both models contained an intercept-only random effect to ensure that comparisons were made within species. Each full model was subject to a step-wise term deletion routine using likelihood ratio tests (LRT) to assess term significance (terms were retained when ). Finally, we conducted a post-hoc analysis using Tukey constrasts with a Bonferroni correction (using the R package; Hothorn *et al.* 2008).

We investigated the relationship between surrogate-based targets and the genetic variation secured in a prioritization using nonlinear mixed-effects models (NLMMs; using the R package). Two models were fit to the adaptive and neutral genetic variation secured using a negative exponential formulation to estimate the rate of the accumulation of genetic variation (; is the accumulate rate). Both models contained intercept and slope random effects to accommodate variation between species.

## Results

Surrogates-targets improved the performance of single-species prioritizations (Figure 1). There were significant differences among the performance of prioritizations depending on the targets used to generate them and if they were assessed according to adaptive or neutral genetic variation (, ). Specifically, the amount-based prioritization captured much less adaptive and neutral genetic variation than either the surrogate-based (, ; , ) or genetic-based (, ; , ) prioritizations. No differences were detected in the proportion of adaptive (, ) or neutral genetic variation (, ) secured by the the surrogate-based and genetic-based prioritizations. However, these patterns were not reflected in the multi-species prioritizations.

Surrogate-targets were not found to improve the performance of multi-species prioritizations (Figure 2; Figure 3). No significant differences in the effectiveness of prioritizations were found depending on whether they were generated using different targets (; ), if they were assessed using adaptive or neutral genetic variation (; ), nor the interaction between these variables (; ).

The surrogate variables showed a greater utility for neutral genetic variation than adaptive variation (Figure 4). The accumulation rate () was 1.83 0.18 S.E. for the proportion of adaptive variation captured in a prioritization using environmental surrogates. On the other hand, the rate () was 2.63 0.28 S.E. for neutral variation using geographic surrogates. Thus minor increases in the target for either surrogate results in major increases in the benefit of the prioritization. However, the coefficient was much larger for neutral genetic variation so minor increases in the target can acheive substantial gains.

## Discussion

### Introduction

* aim
  + we aimed to investigate the benefits of using surrogates for genetic variation in conservation planning.
* innovation
  + first to test the effectiveness of environmental and geographic surrogates in a conservation planning context
* key findings
  + surrogate variables are beneficial in single-species contexts.
  + but not in multi-species contexts when a comprehensive set of species is used.
* consequences
  + highlights the benefit of using a comprehensive set of species
  + supports literature on using comprehensive set of species

### Single + multi-species analysis

### Correlation analysis

### Limitations/Future directions

* spatial/surrogate data
  + isolation by resistance: least-cost paths and circuit theory distances are becoming more population for modeling connectivity, and may offer better surrogates than linear geographic distances. However, the scaling for these variables requires an understanding of the often non-linear relationship, which is not available to conservation planners
  + cost data: here we used no cost data to provide a rigorous test for the benefits of surrogates in conservation planning. however in real conservation planning situations, prioritizations are largely driven by cost. future studies should look at including cost data to see if surrogates provide any benefit when including realistic cost data.
* genetic data
  + AFLP data: next gen data would be better however the aim of a prioritization is to capture the main gradients of genetic variation–not all of the variation.
  + adegenet: whilst other software packages are more popular (eg. Structure), we used adegenet because information theory is more reliable than an ad-hoc second derivative process for comparing models. Additionally, it is computationally infeasible to analyze run 27 species in Structure in a suitable period of time.

### Concluding remarks

* take-home messages
  + surrogates are effective for single-species prioritizations in conservation planning
  + however, surrogates don’t provide any noticeable benefits when using a comprehensive set of species in a multi-species prioritization
  + conservation planners are encouraged to use comprehensive sets of species

## 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

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 A. Moilanen, K. A. Wilson, & H. Possingham) pp. 185–189 Oxford University Press, Oxford, UK.

Barnosky, A. D., Matzke, N., Tomiya, S., Wogan, G. O. U., Swartz, B., Quental, T. B., Marshall, C., McGuire, J. L., Lindsey, E. L., Maguire, K. C., Mersey, B., Ferrer, E. A. (2011) Has the earth’s sixth mass extinction already arrived? *Nature*. **471**, 51–57.

Bates, D., Mächler, M., Bolker, B., Walker, S. (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*. **67**, 1–48.

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.

Bothwell, H., Bisbing, S., Therkildsen, N., Crawford, L., Alvarez, N., Holderegger, R., Manel, S. (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.

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.

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.

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.

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.

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

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.

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

Hanson, J. O., Rhodes, J. R., Possingham, H. P., Fuller, R. A. (2016) RAPR: representative and adequate prioritisations in R. *Unpublished*. **X**, XX–XX.

Hendry, A. P., Lohmann, L. G., Conti, E., Cracraft, J., Crandall, K. A., Faith, D. P., Hauser, C., Joly, C. A., Kogure, K., Larigauderie, A., Magallon, S., Moritz, C., Tillier, S., Zardoya, R., Prieur-Richard, A. H., Walther, B. A., Yahara, T., Donoghue, M. J. (2010) Evolutionary biology in biodiversity science, conservation, and policy: A call to action. *Evolution*. **64**, 1517–1528.

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.

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

IPCC (2007) *Climate change 2007: The physical science basis. contribution of working group i to the fourth assessment report of the intergovernmental panel on climate change*. Cambridge University Press, Cambridge, UK.

Jombart, T. (2008) adegenet: a r package for the multivariate analysis of genetic markers. *Bioinformatics* doi: [10.1093/bioinformatics/btn129](http://dx.doi.org/10.1093/bioinformatics/btn129).

Jombart, T., Devillard, S., Balloux, F. (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*. **11**, 1–15.

Klein, C. J., Wilson, K. A., Watts, M., Stein, J., Carwardine, J., Mackey, B., Possingham, H. P. (2009) Spatial conservation prioritization inclusive of wilderness quality: A case study of australia’s biodiversity. *Biological Conservation*. **142**, 1282–1290.

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

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

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.

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

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*.

Pérez-Figueroa, A., García-Pereira, M. J., Saura, M., Rolán-Álvarez, E., Caballero, A. (2010) Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology*. **23**, 2267–2276.

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.

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 (2014) R: A language and environment for statistical computing.

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.

Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., Erasmus, B. F. N., Siqueira, M. F. de, Grainger, A., Hannah, L., Hughes, L., Huntley, B., Jaarsveld, A. S. van, Midgley, G. F., Miles, L., Ortega-Huerta, M. A., Townsend Peterson, A., Phillips, O. L., Williams, S. E. (2004) Extinction risk from climate change. *Nature*. **427**, 145–148.

Thomas, J. A., Telfer, M. G., Roy, D. B., Preston, C. D., Greenwood, J. J. D., Asher, J., Fox, R., Clarke, R. T., Lawton, J. H. (2004) Comparative losses of british butterflies, birds, and plants and the global extinction crisis. *Science*. **303**, 1879–1881.

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.

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