UNIVERSITY OF CALIFORNIA
RIVERSIDE

Neutral Markers, Quantitative Genetics and the Use of Statistics to Inform Conservation

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Evolution, Ecology, and Organismal Biology

by

Heather R. Taft

December 2012

Dissertation Committee:
Dr. Derek A. Roff, Chairperson
Dr. Helen M. Regan
Dr. Norman C. Ellstrand

The Dissertation of Heather R. Taft is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

Acknowledgements

The text of this dissertation is, in part, a reprint of the material as it appears in *Conservation Genetics* (2012) and appears here with kind permission from Springer Science and Business Media.

ABSTRACT OF THE DISSERTATION

Neutral Markers, Quantitative Genetics and the
Use of Statistics to Inform Conservation

by

Heather R. Taft

Doctor of Philosophy, Graduate Program in Evolution, Ecology and Organismal Biology
University of California, Riverside, December 2012
Dr. Derek A. Roff, Chairperson

Conservation genetics is a booming field focused on assessing the genetic structure and diversity among subpopulations of different species. However, the utilization of genetic analyses in management plans remains unclear because it is not known how often they are considered during creation of plans. Chapter one considers the question of how closely correlated the results from a genetic assessment on population structure are with the recommendations given at the conclusion of a study. Since conservation tends to have limited financial resources, it is imperative that the money spent on genetic studies is providing beneficial information for conservation. This analysis shows that genetic divergence is correlated with the recommendations, but different genetic markers (i.e. microsatellites) and divergence metrics (i.e. F_{st}) show different relationships between the recommendations and genetic divergence, possibly due to differences in the sample size associated with different markers.

Small populations, such as those of conservation concern, that inbreed may lose alleles over time. This can be problematic if the alleles lost are beneficial, but making genetic assessments based on neutral genetic markers gives no information on the fitness of populations. Ideally, conservation programs should use quantitative genetics to assess population fitness. Chapter two looks at the change in additive genetic variance in populations following a dramatic decrease in population size (bottleneck). On average, populations experiencing a bottleneck showed an increase in additive genetic variance for populations that had levels of inbreeding equal to or less than that of sibling mating (0.25); above that a decrease in genetic variation was observed.

Since genetic assessments using neutral markers can be costly, and the use of quantitative genetics may be impractical for informing conservation recommendations, chapter three looks at the use of statistical models to assess International Union for the Conservation of Nature Red list status using ecological characteristics. Here characteristics shared among species that have been assigned a threat status and those that have not been assigned a status, or those needing an update, are used in logistic regressions, regression tree analyses and discriminant function analyses to predict threat status and identify those species in most immediate need of attention. We found that logistic regression and discriminant function analysis are very good at predicting threat status. When resources are limited, using data on previously assessed species may be a good alternative to inform conservation recommendations.

Table of Contents

List of Figures

List of Tables

Introduction

Conservation genetics focuses on combining the use of genetic analysis with conservation biology and the preservation of species. There are a number of different uses for conservation genetics, including assessing human impacts on populations, identifying management units, helping to define taxonomy, and reducing the level of inbreeding in populations. To an extent these are interrelated, but this dissertation is primarily focused on the last of these uses - the use of genetic analysis to study inbreeding in populations, and suggests an alternate way to assess threat status when genetic data are unavailable. Inbreeding can be a serious problem (Hedrick 1995), particularly if a population is small (Bryant et al. 1999), as many populations in need of conservation tend to be. With inbreeding comes reduced genetic diversity, the backbone for adaptation and evolution in a changing environment, causing a reduction in fitness (Reed and Frankham 2003), and possibly even population extinction (Frankham 2003).

Habitat fragmentation has impacted population structure and inbreeding in numerous species ranging from bighorn sheep (Epps et al. 2005), bobcats, and coyotes (Riley et al. 2006) to voles (Gerlach & Musolf 2000), salamanders (Marsh et al. 2008), frogs (Arens et al. 2007), beetles (Keller & Largiader 2003), birds (Bush et al. 2011), plants (Gong et al. 2010), and even aquatic species (Miyake et al. 2011). Genetic studies

used to assess the level of isolation among populations that formerly exchanged genes freely via migration has become a popular method to identify populations in need of conservation efforts. Unfortunately, there is a lack of communication between the scientists performing the genetic assessments and those individuals in charge of monitoring the populations and designing and implementing conservation plans (Sutherland et al. 2004; Sutherland et al. 2006). Are genetic studies making important recommendations for conservation that may be overlooked in a conservation plan? Chapter one looks at the relationship between the level of genetic divergence found among subpopulations and the recommendations for conservation given at the conclusion of a paper. Leaving genetics out of a conservation plan could result in greater expenses later to overcome problems that may have been avoided if genetics were considered initially. However, not all genetic studies are informative for conservation, so scientists should consult conservation managers as well, to ensure their efforts are useful (Mace & Purvis 2008).

One problem that all these genetic assessments overlook is that neutral markers (such as microsatellites) are indirect measures of fitness and do not necessarily indicate divergence in ecologically important traits (Bekessy et al. 2003). Quantitative genetic assessments are better for assessing diversity in ecologically important traits. But how are quantitative genetic measurements affected by inbreeding? Chapter two tackles the long standing debate of whether a decrease in population size followed by a population expansion, also known as a bottleneck, can increase additive genetic variance. An increase in variance with inbreeding is opposite to that expected using neutral markers,

where inbreeding would cause a loss of alleles and thus a loss of variance. Increased variance suggests that bottlenecked populations may be able to adapt and evolve in a changing environment faster than a non-bottlenecked population. However, the use of bottlenecks for conservation is *highly* risky for several reasons discussed in chapter two.

Overall, it appears that neutral markers may not collect the data ideal for conservation managers, quantitative genetic assessments are time consuming and not practical for small populations of conservation concern, and any genetic assessment requires money, a resource often lacking in conservation. So, how can managers assess the threat status of species when genetic data is lacking, or when they simply have incomplete data sets? Chapter three compares the use of three common statistical analyses to determine how well ecological characteristics, such as an individual's size, can be used to predict threat status as assigned by the International Union for the Conservation of Nature. The ability to use data previously gathered on an unlisted species in comparison to data on listed species can identify species that may be threatened and in need of closer monitoring and help direct the use of resources for optimal conservation success. This is also a method of threat assessment that managers with access to statistical software can perform on their own, without calling in the aid of others, which may require more money on tight a budget.

References

Arens P, van der Sluis T, van't Westende WPC, Vosman B, Vos CC, Smulders MJM (2007) Genetic population differentiation and connectivity among fragmented Moor frog (Rana arvalis) populations in The Netherlands. *Landscape Ecology*, **22**, 1489-1500.

Bekessy SA, Ennos RA, Burgman MA, Newton AC, Ades PK (2003) Neutral DNA markers fail to detect genetic divergence in an ecologically important trait. *Biological Conservation*, **110**, 267-275.

Bryant EH, Backus VL, Clark ME, Reed DH (1999) Experimental tests of captive breeding for endangered species. *Conservation Biology*, **13**, 1487-1496.

Bush KL, Dyte CK, Moynahan BJ, Aldridge CL, Sauls HS, Battazzo AM, Walker BL, Doherty KE, Tack J, Carlson J, Eslinger D, Nicholson J, Boyce MS, Naugle DE, Paszkowski CA, Coltman DW (2011) Population structure and genetic diversity of greater sage-grouse (Centrocercus urophasianus) in fragmented landscapes at the northern edge of their range. *Conservation Genetics*, **12**, 527-542.

Epps CW, Palsboll PJ, Wehausen JD, Roderick GK, Ramey RR, McCullough DR (2005) Highways block gene flow and cause a rapid decline in genetic diversity of desert bighorn sheep. *Ecology Letters*, **8**, 1029-1038.

Frankham R (2003) Genetics and conservation biology. *Comptes Rendus Biologies*, **326**, S22-S29.

Gerlach G, Musolf K (2000) Fragmentation of landscape as a cause for genetic subdivision in bank voles. *Conservation Biology*, **14**, 1066-1074.

Gong W, Gu L, Zhang DX (2010) Low genetic diversity and high genetic divergence caused by inbreeding and geographical isolation in the populations of endangered species Loropetalum subcordatum (Hamamelidaceae) endemic to China. *Conservation Genetics*, **11**, 2281-2288.

Hedrick PW (1995) Gene Flow and Genetic Restoration - the Florida Panther as a Case-Study. *Conservation Biology*, **9**, 996-1007.

Keller I, Largiader CR (2003) Recent habitat fragmentation caused by major roads leads to reduction of gene flow and loss of genetic variability in ground beetles. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **270**, 417-423.

Mace GM, Purvis A (2008) Evolutionary biology and practical conservation: bridging a widening gap. *Molecular Ecology*, **17**, 9-19.

Marsh DM, Page RB, Hanlon TJ, Corritone R, Little EC, Seifert DE, Cabe PR (2008) Effects of roads on patterns of genetic differentiation in red-backed salamanders, *Plethodon cinereus*. *Conservation Genetics*, **9**, 603-613.

Miyake T, Nakajima J, Onikura N, Ikemoto S, Iguchi K, Komaru A, Kawamura K (2011) The genetic status of two subspecies of Rhodeus atremius, an endangered bitterling in Japan. *Conservation Genetics*, **12**, 383-400.

Riley SPD, Pollinger JP, Sauvajot RM, York EC, Bromley C, Fuller TK, Wayne RK (2006) A southern California freeway is a physical and social barrier to gene flow in carnivores. *Molecular Ecology*, **15**, 1733-1741.

Sutherland WJ, Armstrong-Brown S, Armsworth PR, Brereton T, Brickland J, Campbell CD, Chamberlain DE, Cooke AI, Dulvy NK, Dusic NR, Fitton M, Freckleton RP, Godfray HCJ, Grout N, Harvey HJ, Hedley C, Hopkins JJ, Kift NB, Kirby J, Kunin WE, MacDonald DW, Marker B, Naura M, Neale AR, Oliver T, Osborn D, Pullin AS, Shardlow MEA, Showler DA, Smith PL, Smithers RJ, Solandt J-L, Spencer J, Spray CJ, Thomas CD, Thompson J, Webb SE, Yalden DW, Watkinson AR (2006) The Identification of 100 Ecological Questions of High Policy Relevance in the UK. *Journal of Applied Ecology*, **43**, 617-627.

Sutherland WJ, Pullin AS, Dolman PM, Knight TM (2004) The need for evidence-based conservation. *Trends in Ecology & Evolution*, **19**, 305-308.

Chapter 1

Is genetic divergence correlated with recommendations for conservation?

Heather R. Taft and Derek A. Roff

Department of Biology, University of California, Riverside, California 92521

Abstract

Genetic analyses are increasingly being used on populations of conservation concern, but whether the analyses are being used to inform conservation recommendations is unclear. Here we performed a meta-analysis, to determine whether recommendations for conservation are correlated with measures of genetic divergence. Specifically we examined the relationship between the results of a genetic study and the conservation recommendations given at the conclusion of a paper, along with the justifications discussed for making those recommendations. Not all DNA markers (i.e. microsatellites, mitochondrial DNA, etc.) and divergence metrics (i.e. θ, Φ, etc.) can be used together because some metrics, such as R_{ST}, are specifically designed for use with certain markers, whereas other marker and metric combinations are used together regularly. We found that recommendations are correlated with the results of the genetic assessment for particular marker and metric combinations. However, different marker

and metric combinations are correlated with different recommendations. Our results suggest that the difference between the metrics is due to differences in sample size. Studies using mitochondrial DNA tended to have smaller per subpopulation sample sizes and suggested less extreme recommendations, such as habitat maintenance, despite having rather large divergence values. The use of larger sample sizes indicating smaller but statistically significant divergence in microsatellite studies appears to have led to smaller divergence values being used to define subpopulations as divergent. Just because statistical analyses indicate a significant divergence value does not mean the divergence value is biologically significant. This is an important point for consideration when conducting genetic analyses.

Introduction

In the past there has been a dichotomy between conservation biologists in the academic arena and individuals in the field who create and institute recovery plans. The academics were often tackling scientific questions that were impractical from a management perspective (Whitten et al. 2001), and few recovery plans were created using scientific research to inform management decisions (Pullin et al. 2004; Pullin & Knight 2005). Practitioners creating recovery plans face time constraints preventing adequate reviews of scientific literature and most recovery plans end up being based on previous plans (Pullin et al. 2004; Sutherland et al. 2004; Pullin & Knight 2005). Emphasis has been placed on bridging this gap by creating more communication between scientists and conservation managers (Schemske et al. 1994; Sutherland et al. 2004;

Sutherland et al. 2006; Shaw et al. 2010). It has been suggested that scientific literature should be made more accessible to individuals creating recovery plans (Sutherland et al. 2004; Sunderland et al. 2009), but it can take four or more years for an article to be published (Fazey et al. 2005). For a population in immediate need of assistance even a single year to analyze, write and publish may be too long. Academic scientists have also been urged to get more involved in management (Whitten et al. 2001). Conservation managers could be consulted early in the planning process to prevent unnecessary, costly actions from being incorporated into management and to ensure future modifications based on known issues are not necessary (Mace & Purvis 2008). On the other side, scientists should consult managers prior to performing a study to ensure they are addressing an important, informative question (Mace & Purvis 2008; Milner-Gulland et al. 2012).

Along with direct communication between academics and managers, it is important for research papers to include management recommendations based on their conclusions. Fortunately, recommendations have become more common in scientific literature (Flaspohler et al. 2000). As genetic studies are increasing in conservation, recommendations based on the results are critical to interpret genetic studies for individuals not fluent in the field. Genetic studies have not been used very much in recovery plans in the past (Schemske et al. 1994; Stinchcombe et al. 2002; Moyle et al. 2003) but when genetic management is suggested as a recovery strategy it is often given high priority (Moyle et al. 2003). Genetic studies are important because fitness is often correlated with genetic diversity (Reed & Frankham 2003) and thus small, fragmented

populations may have increased inbreeding or reduced adaptive potential (Bakker et al. 2010). Changes in genetic structure caused by issues such as fragmentation can have devastating consequences for small populations, eventually leading to their extinction (Frankham 2003).

To assess loss of genetic diversity Wright (1951) developed the statistic, F_{ST}, to compare the genetic diversity in subpopulations with respect to the entire population. If F_{ST} is greater than zero the subpopulation has reduced diversity compared to the overall population. F_{ST} was originally defined as

$$F_{ST} = \frac{\sigma_x^2}{\bar{x}(1-\bar{x})},$$

where $\bar{x}$ is the mean frequency of allele *x*, and σ_x^2 is its variance (Wright 1943, 1951). F_{ST} can also be calculated using the migration rate (*m*) and the population size using Wright's (1951) commonly seen equation:

$$F_{ST} = \frac{1}{4Nm+1}$$

if *m* is small. An alternate way to calculate genetic divergence was created by Nei (1973) to account for multiple alleles and is essentially the same as F_{ST}:

$$G_{ST} = \frac{H_T - H_S}{H_T}$$

where H_T and H_S represent the heterozygosity of the total population and a subpopulation, respectively. The statistic θ_{ST} (Weir & Cockerham 1984) is also used to estimate genetic divergence using analysis of variance to partition subpopulation variance and overall variance in allele frequencies. Another statistic, Φ_{ST} (Excoffier et al. 1992), is similar to θ_{ST},

but is used for data such as microsatellites and single nucleotide polymorphisms using what the authors call analysis of molecular variance. R_{ST} was created by Slatkin (1995) to be used with microsatellite loci assuming they accumulate mutations in a step-wise manner, so alleles that have fewer repeat differences are assumed to be more closely related. Jost's *D* (2008), Hedrick's G'_{ST} (2005) and Meirmans and Hedrick's G''_{ST} (2011) are much newer. They aim to get rid of problems associated with only assessing differences in heterozygosity in populations and not incorporating specific alleles in the equation. All these statistics are becoming more and more common for assessing genetic divergence among populations (see Meirmans and Hedrick [2011] for a further discussion on these statistics).

Knowing the genetic structure within a population is useful for conservation purposes, but it is unclear whether the information is being used to inform conservation. The price of performing a genetic study continues to decrease, but for conservation programs, where money is scarce, they are still expensive. If the results are not being usefully applied to conservation the money may be better utilized elsewhere, though the results could be extremely valuable, even critical, to save a species from extinction when applied properly. In this paper we report on a meta-analysis comparing assessments of genetic divergence to the recommendations made for conservation at the conclusion of a genetic study. Specifically we ask whether genetic divergence is correlated with the recommendations. Then, using the justifications for the recommendations discussed in the conclusion, we assess whether other population characteristics (genetic and non-genetic) significantly impact the recommendations.

Methods

Articles from the journals *Conservation Genetics* and *Biological Conservation* published between 2007 and 2011 were reviewed for studies assessing genetic divergence that gave conservation recommendations for wild populations (no captive or domestic populations were used; see Appendix A for a list of studies). To avoid the problem of pseudo-replication, only studies focused on a single species were used. Only studies reporting overall statistics, not pair-wise values were used for the same reason, but only when divergence values for all types of markers used in the study were provided. Studies on hybridization between wild and farm raised species, or invasive species, were excluded because recommendations based on genetic divergence would be opposite to those for pure bred species (i.e. reduce connectivity to prevent further hybridization). F_{ST} - Q_{ST} studies were excluded because our focus is on the use of F_{ST} and not the relationship between F_{ST} and the quantitative genetic metric Q_{ST}. Finally, when studies assessed divergence over multiple years, only the most recent data were used.

The following data were collected from each study: the neutral Markers used to assess genetic divergence (mtDNA, microsatellites, etc.), the Divergence Metric used to assess genetic divergence (F_{ST}, G_{ST}, θ, etc.), the Divergence Value (the value of F_{ST}, G_{ST}, etc.), sample size, the Recommendations resulting from the study, which were classified into seven categories:

1. Habitat maintenance - enhance habitat, increase area protected
2. Maintain/enhance connectivity between populations
3. Start/change method of genetic banking, ex situ conservation methods suggested

4. Change method of management/management units suggested

5. Translocation between populations

6. Conserve diversity

7. Protect populations

and the Justifications given for making those Recommendations, classified into the eight population characteristics (1-4 genetic and 5-8 non-genetic):

1. Genetic differentiation - whether the authors concluded that genetic divergence was present among populations and thus little or no connectivity remained

2. No genetic differentiation - whether the authors concluded that there was no genetic divergence and thus populations were connected (1 and 2 are not mutually exclusive because in some studies neither Justification was used)

3. Genetic diversity - the populations were genetically diverse

4. Low genetic diversity - the populations had minimal or no diversity (3 and 4 are not mutually exclusive either because in some studies neither Justification was used)

5. Few individuals/population declining

6. Unresolved disturbances - disturbances such as continued hunting were still present and affecting the population

7. Habitat loss/fragmentation

8. Possible ecological differences in locations to which the populations were adapted

Some studies used multiple Markers and different Divergence Metrics to assess genetic divergence. To better understand how specific Markers and Divergence Metrics

affect the genetic Divergence Value, subsets of the data that included a single type of Marker and a single type of Divergence Metric were also used in the analyses.

Several studies generally referred to genetic divergence as F_{ST} without reference to a specific method of calculation, such as Weir and Cockerham's (1984) θ. These Divergence Values were analyzed as F_{ST}, but when references were made to a specific method of calculation, such as θ, they were entered into the analysis as such. The categories for the Recommendations and the Justifications for making the Recommendations were scored binomially: 1 if the Recommendation was suggested or the Justification discussed and 0 if the Recommendation or Justification was not suggested or discussed. Studies generally did not discourage a particular action, so the analyses are based on the Recommendations being made or not made.

Unlike the rest of the studies, Castellano et al. (2009) recommended managing populations as a single unit, not multiple units. Because Recommendations are binomial in our dataset, and their Recommendation is opposite to all the other studies, this study was excluded from all analyses using the Recommendation "Management Units."

To determine if Marker and Divergence Metric could be used as two categories, instead of having to use each type of Marker and Divergence Metric in analyses, we used an analysis of variance (ANOVA) with Divergence Value set as the response variable and Marker and Divergence Metric as the predictor variables.

As all the response variables used in these analyses were binomial, we used a general linear model (GLM) with a binomial error to analyze the data. To assess the impact genetic studies have on Recommendations we analyzed the data in two ways. First, to

determine if there was a significant association between the Divergence Values and the Recommendations, indicating that the Divergence Values influenced the conclusions of the study, we used just the Divergence Values from the genetic studies as independent variables regardless of the Divergence Metric. Following this the same analyses were performed on a subset of data with specific Markers and Divergence Metrics. Finally, to determine which, if any, factors contributed to the Recommendations we used the Justifications for the Recommendations as independent variables, which included genetic and non-genetic population characteristics (see above). For this analysis we included all the justifications (#1 – 8), just the justifications based on genetic divergence (#1 – 2), all genetic considerations (#1 - 4), and the justifications excluding any genetic considerations (#5 – 8, as outlined above).

Analyzing the variables from the genetic studies

Many of the studies used multiple Markers and/or multiple Divergence Metrics to calculate genetic divergence. To minimize problems of pseudo-replication we analyzed each Recommendation separately using two different formats for the data.

1. To avoid issues with multiple Divergence Values from each study, we used the median Divergence Value in analyses with each Recommendation

   $$Recommendation\ (0,1) = a + bDivergence\ Value + \varepsilon \quad (1)$$

   where a and b are constants and ε is a binomial error term. In total, 7 analyses were performed, one for each Recommendation (e.g. Recommendation 1 vs. median Divergence Value). We also ran the tests with the mean and the highest

Divergence Value from each study, which gave the same statistical conclusions so the results are not reported here.

2. The second set of analyses used subsets from a data set that included every Marker and Divergence Metric from each study (i.e. a study was represented once if only one F_{ST} value was calculated, but multiple times if different Markers or Divergence Metrics were used). Subsets using specific Markers and Divergence Metrics, where enough (>10) data points were available, were used to determine whether specific Marker and Divergence Metric combinations were correlated with the Recommendations. For example, only the Divergence Values obtained using microsatellites and calculated using F_{ST} were analyzed in a GLM using Equation 1. Four groupings of Markers and Divergence Metrics had a large enough sample size (N) to perform these analyses: microsatellites with θ (N=51), F_{ST} (N=14), R_{ST} (N=14), and mtDNA with Φ (N=21).

Analyzing the Justifications for the Recommendations

The Justifications for the Recommendations had eight binomial variables, four of which were genetic (genetic differentiation, no differentiation, genetic diversity, and low genetic diversity), which were based solely on whether or not the authors said divergence or diversity were or were not found. In these analyses the Divergence Value was not used because the significance of the actual value was investigated in the previous analyses and we assumed that the Divergence Value was taken as the basis for the genetic Justifications given for the Recommendation.

1. The first analysis examined the importance of divergence as a Justification for making Recommendations (See Table 1.1). A GLM with the binomial error was performed using all eight justification variables (x) for each of the seven Recommendations (seven models total).

$$Recommendation\ (0,1) = a + b_1x_1 + b_2x_2 + b_3x_3 + \ldots + b_8x_8 + \varepsilon \quad (2)$$

Each of the models was compared to a model that contained the six Justification variables excluding the two genetic divergence Justifications.

Table 1.1. Interpretation of the model comparisons in the analyses assessing the importance of the Justifications for the Recommendations.

Comparison	Df[a,b]	How to interpret a significant comparison
1. All Justifications vs No divergence Justifications	2,97	Divergence is important
2. All Justifications vs No genetic Justifications[c]	4,97	Genetic variables in general are important
3. All vs Divergence Justifications	6,97	Variables other than divergence are important (includes non-genetic and diversity variables)
4. All vs Genetic Justifications	4,97	Non-genetic variables are important
5. All vs Null model	8,97	At least one of the Justification variables is important

a: Degrees of freedom for the Justification analyses using all the data.

b: Denominator df for microsatellites and θ: 44; for mtDNA and Φ:15.

c: The genetic Justifications include: genetic differentiation, no differentiation, genetic diversity, and low genetic diversity.

2. Next we assessed whether or not any of the genetic Justifications for Recommendations helped explain the Recommendations. Here we compared a model with all eight Justifications to a model without the four genetic Justifications.
3. This was followed by analyses to assess how important the other Justification variables were for making Recommendations. We compared a model with all eight Justifications to a model with only the two divergence Justifications, divergent and not divergent.
4. Next, we compared a model with all eight Justifications to one with the four genetic Justifications alone.
5. Finally, a model with all eight Justification variables was compared to a null model with no Justification variables to determine the significance of the full model.

These same model comparisons used to assess the Justifications for Recommendations were performed again on subsets of the data where Marker and Divergence Metric were held constant. Only two subsets had enough studies to perform these analyses: microsatellites and θ (N = 51), and mtDNA and Φ (N=21). This is different from the previous analyses with subsets of the data using single Markers and Divergence Metrics because here we have more variables to compare, requiring a larger sample size.

Results

One hundred and six studies were used in the analysis. The dataset included plants (29%) and animals (71%). The plants were dominated by terrestrial perennial angiosperms, with only one aquatic species, two gymnosperms and a bryophyte included in the dataset. Both vertebrates (84%) and invertebrates (16%) were represented in the animals, as well as terrestrial (60%) and aquatic (40%) species.

There was a significant difference between Divergence Values using the Marker as the predictor variable (ANOVA, $F_{(8,165)} = 9.234$, $P < 0.0001$), indicating that combining all Markers into a single category is inappropriate. Similarly, there was a significant difference between Divergence Values ($F_{(8,165)} = 4.289$, $P = 0.0001$) using the Divergence Metric as the predictor variable, so it may be more appropriate to analyze each Marker and Divergence Metric combination separately. Including both Marker and Divergence Metric as independent variables in an ANOVA, along with the interaction term, only produced a significant result for the Marker (Marker: $F_{(8,146)} = 4.736$, $P < 0.0001$, Divergence Metric: $F_{(8,146)} = 0.998$, $P = 0.4402$, Interaction: $F_{(11,146)} = 0.582$, $P = 0.8416$).

Are the Divergence Values correlated with a particular Recommendation?

To avoid issues of pseudo-replication with multiple Divergence Values from each study, only a single value from each study was used; because Divergence Values within a study may not be normally distributed we ran the tests using the median Divergence Value as the response variable and each of the seven Recommendations as the predictor variable. Because of the number of tests (7) we did a Bonferroni correction and set the

significance level for the Divergence Value at 0.05/7=0.007. In only one case did the correlation between Divergence Value and Recommendation ("Genetic banking") approach statistical significance (Table 1.2).

Table 1.2. Probabilities from analyzing the importance of the median Divergence Value from a study for making specific Recommendations. *P*-values less than 0.05 are indicated in bold.

	Median Divergence Value	
	df = 1, 104*	
	F	***P* - value**
Connectivity	2.8310	0.0955
Protect Population	0.3593	0.5502
Habitat Maintenance	0.6552	0.4201
Management Units	2.4078	0.1238
Genetic Banking	**7.2741**	**0.0082**
Conserve Diversity	0.7436	0.3905
Translocation	0.3196	0.5731

* df for "Management Units" was 1, 103

The above tests combined all types of Markers (microsatellites, mtDNA, etc.) and Divergence Metrics (F_{ST}, θ, etc.). To assess the influence of individual combinations of Markers and Divergence Metrics (see columns in Table 1.3 for specific combinations) we reran the above tests for those combinations that had more than 10 observations. Given that there were four combinations and seven Recommendations we set the

Table 1.3. Probabilities from analyzing the importance of Divergence Values for making specific Recommendations using subsets of the data with specific Markers and Divergence Metrics. *P*-values less than 0.05 are indicated in bold.

	Microsatellites and θ		Microsatellites and F_{ST}		Microsatellites and R_{ST}		mtDNA and Φ	
	df = 1, 51[a]		df = 1, 12		df = 1, 13		df = 1, 22	
	F	*P* - value	F	*P* - value	F	*P* - value	F	*P* - value
Connectivity	0.0313	0.8604	0.0869	0.7732	1.7839	0.2046	0.0692	0.7950
Protect Populations	0.6807	0.4132	1.0061	0.3356	2.5879	0.1317	0.0609	0.8073
Habitat Maintenance	1.9748	0.1660	0.7862	0.3927	0.7466	0.4032	1.1121	0.3031
Management Units	0.2155	0.6445	0.5572	0.4698	0.0594	0.8113	1.3018	0.2662
Genetic Banking	3.0599	0.0863	NA[b]	NA	6.9198	**0.0208**	NA[b]	NA
Conserve Diversity	4.3221	**0.0427**	0.3902	0.5439	2.7807	0.1193	0.3242	0.5749
Translocation	0.0070	0.9339	0.0449	0.8357	0.0392	0.8462	32.4131	**<0.0001**

a: df for "Managements Units" was 1, 50.

b: No studies recommended "Genetic banking" using the data subsets with microsatellites and F_{ST} or mtDNA and Φ.

significance level at 0.05/28 = 0.0018. In only three cases is $P<0.05$ and after correcting for multiple tests, the only significant correlation between a particular Recommendation and statistic was the Recommendation "Translocation" using the combination of mtDNA and Φ (Table 1.3).

Are the Justifications for the Recommendations significantly associated with the Recommendations?

At the conclusion of each paper one or more of the seven Recommendations (listed in the methods section) were given based upon one or more of the eight population characteristics justifying the Recommendations (Justifications, see methods), four of which were derived from the genetic analysis.

Using all the data produced 10 tests in which $P<0.05$, which is significantly different from the expected 5% by chance alone (binomial test, $P<0.0001$, Table 1.4). In no test were "Habitat maintenance," "Genetic banking," or "Translocation" significant. Because of the large number of tests (35) identifying the "significant" (i.e. $P<0.05$) tests is difficult as none of the values are less than the Bonferroni-corrected P value of 0.05/35=0.0014. However, six of the "significant" tests implicate genetic variables, although divergence per se appears not to have played a significant role in the Recommendations (Test 1, Table 1.4).

When the tests were applied to the two subsets for which enough data were available a somewhat different picture emerges. In both cases there were significantly more "significant" tests than expected by chance (binomial test, $P=0.003$, $P<0.00001$, respectively, Table 1.4). The data subset using microsatellites and θ consistently

Table 1.4. Probabilities from the ANOVAs used to compare general linear models analyzing the Justifications for making conservation Recommendations. See Table 1 for the interpretation of the tests. *P*-values less than 0.05 are indicated in bold.

	Recommendations						
Test for significance of	Maintain Connectivity	Protect Pops	Habitat Maintenance	Management Units	Genetic Banking	Conserve Diversity	Translocation
All Data							
1. Divergence	0.3103	0.1241	0.8250	0.0525	0.5640	0.3020	0.3816
2. Genetic variables	0.2757	**0.0041**	0.7115	**0.0130**	0.2984	**0.0231**	0.0881
3. Variables other than divergence	**0.0348**	**0.0191**	0.3741	0.2607	0.4938	**0.0172**	0.1016
4. Non-genetic variables	**0.0471**	0.1530	0.2546	0.4151	0.8570	0.0976	0.1456
5. One or more predictor variables	0.0690	**0.0108**	0.5110	**0.0374**	0.5743	**0.0259**	0.0727
Subset using microsatellites and θ (not enough df for R_{ST} and F_{ST})							
1. Divergence	0.5027	0.6579	0.2806	0.7815	**0.0015**	0.9294	0.5964
2. Genetic variables	0.6501	0.1877	0.5799	0.6706	**< 0.0001**	0.8890	0.7315
3. Variables other than divergence	0.2678	**0.0039**	0.3483	0.5621	**0.0007**	0.3496	0.9891
4. Non-genetic variables	0.2166	**0.0065**	0.1764	0.4370	**0.0369**	0.3342	0.9914
5. One or more predictor variables	0.3553	**0.0079**	0.4432	0.6639	**0.0002**	0.4848	0.9664
Subset using mtDNA and Φ							
1. Divergence	**< 0.0001**	0.1750	**0.0119**	0.9374	NA*	0.3938	1.0000
2. Genetic variables	**< 0.0001**	**0.0157**	**0.0046**	0.3890	NA	0.2690	0.7543
3. Variables other than divergence	**< 0.0001**	**0.0001**	**0.0103**	0.2661	NA	0.0900	0.1391
4. Non-genetic variables	**< 0.0001**	**< 0.0001**	0.0993	0.4373	NA	0.1487	0.0835
5. One or more predictor variables	**< 0.0001**	**0.0001**	**0.0096**	0.3843	NA	0.1177	0.1580

* No study that used mtDNA and Φ recommended genetic banking.

produced significant tests based on the Recommendations "Protect populations" and "Genetic banking," only one of which was correlated with genetic divergence though. In contrast, the subset using mtDNA and Φ consistently produced significant tests based on the Recommendations "Maintain connectivity," "Protect populations," and "Habitat maintenance," all of which were correlated with genetic divergence and/or diversity. Chi-squared analyses were performed to investigate the reason for the discrepancy between the two metrics. First we used the type of organism subdivided four different ways: 1. vertebrate or invertebrate, 2. terrestrial or aquatic, 3. plant or animal, and 4. mammals, herps, fish, birds, invertebrates, and plants. Second, the rationale for performing the study was investigated by looking at how many of the studies used each of the following rationales: assess genetic diversity, assess genetic structure, assess the impact of a translocation or assess the impact of a bottleneck. The ratio of the total sample size to the number of subpopulations (total sample size divided by the number of subpopulations assessed) was also considered. Not all studies give the subpopulation sample size, so this ratio approximates a per subpopulation sample size and was used to see if sample size may have impacted the results. The ratios of sample size to the number of subpopulations were divided into eight categories: 1-10, 11-20, 21-30, 31-40, 41- 50, 51-100, 101-150, and >151. Finally, the year of publication was considered: 2007, 2008, 2009, 2010, and 2011. The six studies that used both mtDNA and microsatellites were excluded from the analyses because of overlapping data, so any distinctions between the markers could be observed. There was only one significant difference found in these

analyses, the ratio of the sample size to the number of subpopulations assessed (χ^2 = 17.274, df = 7, *P* = 0.0157; Table 1.5). In general, mtDNA had smaller ratios than microsatellites, thus authors may be more hesitant to suggest more dramatic Recommendations, such as translocation, when they have smaller sample sizes per subpopulation.

Table 1.5. Results from the χ^2 analysis used to identify the reason for the difference in the Recommendations associated with microsatellites and mtDNA.

χ^2 factors	χ^2	df*	*P*-value
Organism			
1. vertebrate, invertebrate	0.0036	1	0.9522
2. terrestrial, aquatic	0.0287	1	0.8655
3. plant, animal	2.0628	1	0.1509
4. mammals, herps, fish, birds, invertebrates, and plants	5.1978	5	0.3922
Rationale for performing the study	4.0282	3	0.2584
Ratio of sample size to number of subpopulations	17.2743	7	**0.0157**
Year published	4.0708	4	0.3965

* Degrees of freedom

Discussion

Assessments of genetic divergence are often performed without a clear understanding of how the results will be applied to conservation. In fact, there is no threshold of genetic divergence used to define divergent populations, so genetic analyses are entirely reliant upon those performing the genetic assessments to determine whether populations are divergent and what actions should be taken to conserve the populations. We analyzed the recommendations given at the end of genetic analyses to determine whether the results of genetic studies are actually related to the recommendations given. The median genetic Divergence Values in the analyses using just the variables from the genetic studies indicated that divergence only impacted one of the seven Recommendations often discussed at the end of a genetic study. In the Justification analyses again genetic divergence was only found to be important for a few of the Recommendations using subsets of the data. Since the choice of using different Markers (microsatellites, mtDNA, etc.) and different Divergence Metrics (θ, Φ, R_{ST}, etc.) would lead to different Divergence Values, grouping the data by Marker and Divergence Metric would make the relationship between the Divergence Value and the Recommendation more apparent by excluding the noise of other Markers and Metrics. The analyses using microsatellites and θ indicated only two Recommendations were correlated with the Justifications: "Protect populations," and "Genetic banking." "Genetic banking" was the only one of these Recommendations that had a genetic basis (Table 1.4). There were few studies that used mtDNA and Φ; however, here three of the Recommendations using this subset of the data were significantly related to the Justifications: "Maintain connectivity,"

"Protect populations," and "Habitat maintenance." Studies using mtDNA and Φ may be more likely to give those Recommendations over studies using other Marker and Divergence Metric combinations because, as the dataset used here indicated, studies using mtDNA had a smaller per subpopulation sample size. Authors may be more hesitant to recommend more extreme management options such as "Translocation" and "Management units" without larger sample sizes per subpopulation.

The Markers and Divergence Metrics used in these analyses are not directly interchangeable. However, because of the use of multiple Markers or multiple Divergence Metrics each study often had multiple Divergence Values: as a consequence, dividing the data up by different Markers and Divergence Metrics left information out that potentially contributed to the Recommendations made in each study.

Eighty-five of the 106 studies in this analysis assessed the statistical significance of their genetic divergence values. Nine studies stated that they had significant divergence and thus the populations were divergent when the divergence value was very small (~ 0.05). A divergence value this small may not be biologically significant. Sample size would impact the significance value and populations with a large sample size may show a significant divergence for very small divergence values (Caujape-Castells 2010). The problem with small divergence values is that they may simply be an issue of autocorrelation due to poorly chosen sampling sites if the sites were inadvertently located far enough apart within a single population for divergence to be observed (Schwartz & McKelvey 2009; Caujape-Castells 2010). Populations where mating occurs among individuals residing close to each other would be an ideal situation to observe this issue

(Schwartz & McKelvey 2009). Unfortunately, autocorrelation is a problem that even spatial structure programs, such as STRUCTURE, cannot overcome (Schwartz & McKelvey 2009).

This issue with small divergence values highlights a problem of inconsistency in defining populations as divergent and raises questions about whether the recommendations are biologically necessary. In conservation biology there is also an issue of uncertainty about what a non-divergent result means. Latta (2008) pointed out that the finding of genetic divergence could mean there is local adaptation, but not finding genetic divergence does not mean there is no local adaptation. Thus recommendations for management would be the same regardless of whether or not the populations were divergent. Studies on genetic divergence may not always be necessary if the conclusion of no genetic divergence may be interpreted as inconclusive leading to the same recommendation regardless of the result.

Unfortunately, neutral Marker divergence does not always indicate significant divergence in ecologically important traits (Bekessy et al. 2003). In fact, neutral genetic divergence can be observed in selfing populations, isolated populations, and those with high apomixis without being detrimental. If genetic divergence is based on markers that are actually neutral, any amount of genetic divergence may not indicate a cause for concern. Quantitative genetic divergence or loss of diversity in a quantitative trait would be more indicative of a cause for conservation concern, but performing quantitative genetic assessments on small wild populations would be costly and potentially inappropriate due to inadequate sampling. In fact, managing for high genetic variation

may not always be the best method for conservation (McKay et al. 2005). If a population is in the process of adapting to a new location, conserving diversity that is no longer beneficial may reduce the population's fitness and slow the rate of evolution (Stockwell et al. 2003).

Besides assessments of genetic divergence there are many other uses for genetics in conservation including defining taxonomy (Frankham 2003), assessing kinship among individuals to reduce inbreeding (Avise 2010), identifying invasive species, assessing presence or abundance of small populations or dangerous animals using non-invasive DNA collection, such as hair samples, and diagnosing and tracking diseases (Schwartz et al. 2007). Financial resources are scarce when dealing with species of conservation concern. Assessments of genetic structure may be important information for conservation of species on the verge of extinction, where adequate money is available. However, it is important to ensure that money is only used on genetic studies that will impact conservation management (Howes et al. 2009). Other species that are insufficiently funded may benefit more from directly using money for management purposes based on known information instead of performing a genetic assessment.

Overall, it does appear that genetic studies contribute to management recommendations. However, the specific Markers and Divergence Metrics influence recommendations differently. The use of more Markers and Divergence Metrics in a study will help paint a better picture of the genetic structure within the subpopulations. Moritz (1994) suggested using both mtDNA and nuclear markers, such as microsatellites, to determine whether or not different management units should be utilized. In general,

the use of multiple Markers should be expanded to other recommendations such as translocation, where major changes in management are being made that will directly affect populations when financial resources are available to perform an extensive genetic analysis.

References

Avise JC (2010) Perspective: conservation genetics enters the genomics era. *Conservation Genetics*, **11**, 665-669.

Bakker J, van Rijswijk MEC, Weissing FJ, Bijlsma R (2010) Consequences of fragmentation for the ability to adapt to novel environments in experimental Drosophila metapopulations. *Conservation Genetics*, **11**, 435-448.

Bekessy SA, Ennos RA, Burgman MA, Newton AC, Ades PK (2003) Neutral DNA markers fail to detect genetic divergence in an ecologically important trait. *Biological Conservation*, **110**, 267-275.

Caujape-Castells J (2010) General G (ST) and theta inflation due to biased intra-population sampling, and its consequences for the conservation of the Canarian Flora. *Conservation Genetics*, **11**, 709-720.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes - application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479-491.

Fazey I, Fischer J, Lindenmayer DB (2005) Who does all the research in conservation biology? *Biodiversity and Conservation*, **14**, 917-934.

Flaspohler DJ, Bub BR, Kaplin BA (2000) Application of Conservation Biology Research to Management. *Conservation Biology*, **14**, 1898-1902.

Frankham R (2003) Genetics and conservation biology. *Comptes Rendus Biologies*, **326**, S22-S29.

Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633-1638.

Howes BJ, Pither R, Prior KA (2009) Conservation implications should guide the application of conservation genetics research. *Endangered Species Research*, **8**, 193-199.

Jost L (2008) Gst and its relatives do not measure differentiation. *Molecular Ecology*, **17**, 4015-4026.

Latta RG (2008) Conservation genetics as applied evolution: from genetic pattern to evolutionary process. *Evolutionary Applications*, **1**, 84-94.

Mace GM, Purvis A (2008) Evolutionary biology and practical conservation: bridging a widening gap. *Molecular Ecology*, **17**, 9-19.

McKay JK, Christian CE, Harrison S, Rice KJ (2005) "How local is local?" - A review of practical and conceptual issues in the genetics of restoration. *Restoration Ecology*, **13**, 432-440.

Meirmans PG, Hedrick PW (2011) Assessing population structure: F-ST and related measures. *Molecular Ecology Resources*, **11**, 5-18.

Milner-Gulland EJ, Barlow J, Cadotte MW, Hulme PE, Kerby G, Whittingham MJ (2012) Ensuring applied ecology has impact. *Journal of Applied Ecology*, **49**, 1-5.

Moritz C (1994) Applications of mitochondrial DNA analysis in conservation - a critical review. *Molecular Ecology*, **3**, 401-411.

Moyle LC, Stinchcombe JR, Hudgens BR, Morris WF (2003) Conservation genetics in the recovery of endangered animal species: A review of US endangered species recovery plans (1977-1998). *Animal Biodiversity and Conservation*, **26**, 85-95.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, **70**, 3321-3323.

Pullin AS, Knight TM (2005) Assessing conservation management's evidence base: a survey of management-plan compilers in the United Kingdom and Australia. *Conservation Biology*, **19**, 1989-1996.

Pullin AS, Knight TM, Stone DA, Charman K (2004) Do conservation managers use scientific evidence to support their decision-making? *Biological Conservation*, **119**, 245-252.

Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. *Conservation Biology*, **17**, 230-237.

Schemske DW, Husband BC, Ruckelshaus MH, Goodwillie C, Parker IM, Bishop JG (1994) Evaluating Approaches to the Conservation of Rare and Endangered Plants. *Ecology*, **75**, 585-606.

Schwartz MK, Luikart G, Waples RS (2007) Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology and Evolution*, **22**, 25-33.

Schwartz MK, McKelvey KS (2009) Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics*, **10**, 441-452.

Shaw JD, Wilson JRU, Richardson DM (2010) Initiating dialogue between scientists and managers of biological invasions. *Biological Invasions*, **12**, 4077-4083.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457-462.

Stinchcombe J, Moyle LC, Hudgens BR, Bloch PL, Chinnadurai S, Morris WF (2002) The influence of the academic conservation biology literature on endangered species recovery planning. *Conservation Ecology*, **6**.

Stockwell CA, Hendry AP, Kinnison MT (2003) Contemporary evolution meets conservation biology. *Trends in Ecology & Evolution*, **18**, 94-101.

Sunderland T, Sunderland-Groves J, Shanley P, Campbell B (2009) Bridging the Gap: How Can Information Access and Exchange Between Conservation Biologists and Field Practitioners be Improved for Better Conservation Outcomes? *Biotropica*, **41**, 549-554.

Sutherland WJ, Armstrong-Brown S, Armsworth PR, Brereton T, Brickland J, Campbell CD, Chamberlain DE, Cooke AI, Dulvy NK, Dusic NR, Fitton M, Freckleton RP, Godfray HCJ, Grout N, Harvey HJ, Hedley C, Hopkins JJ, Kift NB, Kirby J, Kunin WE, MacDonald DW, Marker B, Naura M, Neale AR, Oliver T, Osborn D, Pullin AS, Shardlow MEA, Showler DA, Smith PL, Smithers RJ, Solandt J-L, Spencer J, Spray CJ, Thomas CD, Thompson J, Webb SE, Yalden DW, Watkinson AR (2006) The Identification of 100 Ecological Questions of High Policy Relevance in the UK. *Journal of Applied Ecology*, **43**, 617-627.

Sutherland WJ, Pullin AS, Dolman PM, Knight TM (2004) The need for evidence-based conservation. *Trends in Ecology & Evolution*, **19**, 305-308.

Weir BS, Cockerham CC (1984) Estimating f-statistics for the analysis of population-structure. *Evolution*, **38**, 1358-1370.

Whitten T, Holmes D, MacKinnon K (2001) Conservation biology: a displacement behavior for academia? *Conservation Biology*, **15**, 1-3.

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

Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323-354.

Chapter 2

Do Bottlenecks Increase Additive Genetic Variance?

Heather R. Taft and Derek A. Roff

Department of Biology, University of California, Riverside, California 92521

Abstract

Because of anthropogenic factors many populations have been at least temporarily reduced to a very small population size. Such reductions could potentially decrease genetic variation and increase the probability of extinction. Analysis of molecular markers has shown a decrease in genetic variation but in many cases this has not reduced the ability of the population to recover from the bottleneck. This apparent paradox is resolved by a consideration of how population bottlenecks can affect additive genetic variance, the relevant measure of ability to respond to selective factors. A bottleneck has the potential to increase additive genetic variance in a population. This may result in an increase in fitness, particularly in populations of conservation concern that are small and lack genetic variation. Here we present a meta-analysis of experimental tests of this prediction using models designed to fit data that is strictly additive and data that has non-additive components. This analysis shows that additive genetic variance in a dataset

dominated by morphological traits increases, on average, after a bottleneck event when the inbreeding coefficient is less than 0.3, but neither of the theoretical models alone can adequately explain this result. Because of our inability at present to predict the results of a population bottleneck in a specific case and the probability of extinction associated with small population size we caution against using bottlenecks to increase genetic variance, and thus the fitness, of endangered populations.

Introduction

Because of anthropogenic effects many populations have been reduced at least temporarily to very low numbers. The possibility of an increased probability of extinction because of decreased genetic variance is a significant concern of conservation biology as evidenced by the number of papers focused on this phenomenon in journals such as Conservation Genetics, Conservation Biology and Animal Conservation: a search using the term "bottleneck*" under the division of Biodiversity Conservation in web of science produced 402 papers, 342 of which were published in the last 10 years. Analysis of genetic variation using molecular markers in populations passing through such population bottlenecks has shown significant loss of genetic variation: examples include eutherian mammals (Bonnell & Selander 1974; Neumann et al. 2004; Culver et al. 2008; Durrant et al. 2009; Haanes et al. 2010; Corti et al. 2011; Fenderson et al. 2011; Ricanova et al. 2011; Sastre et al. 2011), marsupials (Sinclair et al. 2002; Cardoso et al. 2009), birds (Bellinger et al. 2003; Munoz-Fuentes et al. 2005; Funk et al. 2010; Kuro-o et al. 2010), amphibians (Schoville et al. 2011), fish (Consuegra et al. 2005; Earl et al. 2010;

Swatdipong et al. 2010) and plants (Jacquemyn et al. 2010). Despite the loss of molecular variation after passing through a bottleneck numerous species (but not all – see, for example, Heber & Briskie 2010) have shown few ill effects and their populations have expanded and persisted: for example, the Northern elephant seal (Hoelzel 1999; Weber et al. 2000; Hoelzel et al. 2002), the kakerori, an endemic bird of the Cook Islands (Chan et al. 2011), the Seychelles kestrel (Groombridge et al. 2009), the skink, *Oligosoma suteri* (Miller et al. 2011), the butterfly, *Parnassius Apollo* (Habel et al. 2009) and the stingless bee *Melipona scutellus* (Alves et al. 2011). For other examples see Reed (2010). Similarly, many highly successful invasions have started from very small introduced populations: the Norway rat on Moturemu Island in New Zealand (Russell et al. 2009), the house finch in eastern United States (Hawley et al. 2008), lake trout in the Rocky Mountains (Kalinowski et al. 2010), the dwarf honey bee in the near East (Moritz et al. 2010), and barbed goatgrass in California (Meimberg et al. 2006). In their review Puillandre et al. (2008) found reduced genetic variability based on molecular markers in the invasive population relative to the native in 80.5% of the 62 cases for which such data were available.

The persistence and population expansion of populations passing through a bottleneck appears paradoxical in the light of decreased molecular genetic variation. This apparent paradox is resolved by a consideration of how additive genetic variance, the genetic factor determining response to selection, is changed by passage through a population bottleneck. Individual and population fitness is a function of life history components such as fecundity, development time and survival (Roff 1997). Other types

of traits such as those relating to behavior, morphology or physiology may also have a significant impact on fitness. Variation in these components is measured not by neutral molecular variation but by quantitative genetic variation. There may be little or no correspondence between these two components of genetic variation (Merila & Crnokrak 2001; Palo et al. 2003; Leinonen et al. 2008) and thus if we wish to understand the consequences of bottlenecks on fitness and the ability of a population to respond to selective factors following the bottleneck we must examine changes in quantitative genetic variation, specifically additive genetic variance, rather than that in neutral molecular markers (Dlugosch & Parker 2008).

An increase in additive genetic variance, V_A, is possible if there is dominance or epistatic variance present in the population (Robertson 1952; Goodnight 1987; Cockerham & Tachida 1988; Goodnight 1988; Tachida & Cockerham 1989a, b; Willis & Orr 1993; Cheverud & Routman 1996; Barton & Turelli 2004). Passage through a bottleneck will change the allele frequencies thereby changing the relative magnitude of the component genetic variances, leading in some cases to an increase in V_A (reviewed in Roff 1997, p 299-305). In general, dominance rather than epistatic effects are more likely to generate such increases (Turelli & Barton 2006), and could be beneficial by increasing the ability of populations to evolve. Under a strictly additive model, however, V_A should decrease with inbreeding as alleles are lost (Wright 1951; Lande 1980) and bottlenecked populations could suffer a decrease in their ability to evolve under new conditions. How a population responds to a bottleneck may be specific to the population itself, but if there is a general trend across species it may help explain why some

populations, such as the Northern elephant seal, are able to persist for long periods of time despite experiencing dramatic drops in their population size.

The change in V_A depends upon the genetic architecture of specific traits, which is a phenomenon that can only be answered by empirical study. However, certain classes of traits, such as those associated directly with fitness (life history traits), show significant inbreeding depression indicative of directional dominance and possibly epistasis (DeRose & Roff 1999). For such traits we might expect a relatively high likelihood of an increase in V_A following a bottleneck. On the other hand, morphological traits typically do not exhibit much, if any, inbreeding depression and are assumed to be determined in large measure by V_A (Roff 1998; DeRose & Roff 1999; Wright et al. 2008). A meta-analysis of components of genetic variance from line cross analyses demonstrated that life history traits display significantly higher dominance and epistatic effects than morphological traits but both epistatic and dominance effects are found in 67% of morphological traits examined (Roff & Emerson 2006). Thus, given the presence of dominance and epistatic effects, which imply the existence of some non-additive genetic variance in both life history and morphological traits, there is the possibility that V_A will increase after a bottleneck, and the assumption that the additive genetic variance in morphological traits decreases when the population passes through a bottleneck is premature.

At this time there are a number of studies assessing the change in V_A following a bottleneck using different species, traits, inbreeding coefficients and bottleneck sizes so a meta-analysis of the overall effect of bottlenecks on additive genetic variance is now possible. Here we address the general question of whether or not there is a consistent

change in additive genetic variance following a bottleneck. Additionally, we investigate whether there are significant differences due to trait type (morphological, life history, behavioral and physiological), species, study (referred to as "paper"), inbreeding coefficient, and bottleneck size. The last two factors are particularly important. Bottleneck size (N) and the inbreeding coefficient (f) are directly related by the equation $f = 1/2N$ for a single generation of inbreeding among unrelated individuals, so the inbreeding coefficient will be large if the bottleneck size is small. However, for several reasons, this relationship may not be directly applicable to all experiments, and, in general, the inbreeding coefficient may be a better metric with which to assess the effect of a bottleneck among different experiments than the bottleneck size per se. First, the individuals chosen for the bottleneck may, unbeknownst to the researcher, be related, leading to an inbreeding coefficient that is larger than expected after a bottleneck of a particular size. Second, if the bottleneck was maintained for several generations to obtain a certain inbreeding coefficient as done in Kristensen et al. (2005), the above relationship must be modified to account for the multiple generations of inbreeding. Third, experiments are not consistent in the number of generations of flushing (where population size increases using only the bottlenecked individuals and their offspring as parents) that take place prior to assessing the change in additive genetic variance (Table 2.1). Thus one of the important questions we address is the extent to which bottleneck size versus the inbreeding coefficient can account for changes in additive genetic variance following a bottleneck.

Table 2.1. Bottleneck size, inbreeding coefficient, generations the bottleneck was maintained and the generations after a bottleneck before additive genetic variance was assessed. Due to the number of generations before additive genetic variance was assessed after a bottleneck the relationship between bottleneck size and the inbreeding coefficient will not be exact.

Paper	Bottleneck size	*f*	Gens bottleneck maintained	Gens after bottleneck before analysis
Andersson et al. 2010	1	0.535	1	3
van Heerwaarden et al. 2008	2	0.276	1	3
Whitlock and Fowler 1999	2	0.32	1	1
Kristensen et al. 2005	8	0.67	18	
	2	0.67	5	
Saccheri et al. 2001	2	0.27	1	3
	6	0.1	1	
	20	0.03	1	
Briggs and Goldman 2006	2	0.25	1	
Fernandez et al. 1995	2	0.25	1	
Fernandez et al. 2003	2	0.25	1	
Cheverud et al. 1999	4	0.39	3	
Lopez-Fanjul and Villaverde 1989	2	0.25	1	
Lopez-Fanjul et al. 1989	2	0.5	1	3
Garcia et al. 1994	2	0.25	1	
	2	0.5	3	
	2	0.73	6	
Bryant and Meffert 1995	4	0.125	1	
Bryant and Meffert 1996	4	0.125	1	
Bryant and Meffert 1993	2	0.25	1	
	8	0.0625	1	
	32	0.015625	1	
Bryant et al. 1986	2	0.25	1	
	8	0.0625	1	
	32	0.015625	1	
Meffert 1995	4	0.125	1	

Material and Methods

We performed an extensive literature search of empirical studies examining the effect of population bottlenecks on V_A and recorded data on study characteristics (species, trait, bottleneck size, inbreeding coefficient, number of bottlenecks) and V_A, of control and experimental (i.e. bottlenecked) lines. Any study that reported V_A on populations that experienced a bottleneck in comparison to a control population was included, except for Lints and Bourgois (1984) which did not contain enough information to determine the inbreeding coefficient of the bottlenecked population. Seventeen studies are included in the analysis, 14 of which used invertebrates. Eight different species are included, most of which are dipterans (flies); five studies used *Musca domestica* (house fly), six used *Drosophila melanogaster* (fruit fly) and one used *D. bunnanda*. Among the studies many different traits were measured, with different bottleneck sizes experienced prior to assessing V_A (see Appendix B for the data set). Data from 107 experiments were used in the present analysis, 86 of which used invertebrates, dominated by 43 experiments using *Musca domestica* and 24 using the lepidopteran *Bicyclus anynana*. While many different traits were measured, the majority (85) can be classified as morphological traits with only one experiment measuring physiological traits, 11 measuring behavioral traits, and 10 measuring life history traits. One study (Bryant & Meffert 1993) reported the average of eight different morphological traits. Here these data are assessed as a composite trait in three experiments. All other traits had enough data reported to be assessed as single traits. Several studies performed the same experiment multiple times (same bottleneck sizes/inbreeding coefficients and traits used). When this occurred we used the average as

a single datum in the meta-analysis instead of the median for consistency since several studies only reported averages, and paper is included in the analysis to avoid issues with pseudo-replication.

Seven different bottleneck sizes were used. Here we consider a bottleneck as a reduction in population size that is followed by a population flush. For two studies the effect of consecutive bottlenecks (where a bottleneck occurs and is followed by a population flush, then the bottleneck and flush are repeated in the same population several generations later) was assessed on the same population by recording data after each flushing event (Lopez-Fanjul et al. 1989; Bryant & Meffert 1993). Only the data from the first bottleneck was used in the present analysis. Because no studies used multiple bottlenecks on separate experimental lines, we did not consider this variable in our analyses.

Additive genetic variance itself cannot be used as a variable because it depends upon the scale of measurement. One solution is to standardize the variance to the control, i.e. use the ratio of the variance in the experimental to the variance in the control line (Van Buskirk & Willi 2006). However, ratios are frequently highly skewed in their distribution and may show considerable heteroscedasticity. As the latter was clearly evident in the present data set we removed 15 data points with negative variances and used the log transformed ratio of V_A for the bottleneck population relative to the control, $\log\left(\frac{V_{A(bottleneck)}}{V_{A(control)}}\right)$ (referred to as log$R$ throughout the paper).

In the absence of non-additive effects, such as might be expected in many morphological traits, additive genetic variance is expected to decrease with the inbreeding coefficient according to the relationship (Crow & Kimura 1970)

$$V_{A(bottleneck)} = (1 - f)V_{A(control)} \quad 1$$

where f is the inbreeding coefficient of the bottlenecked population. Standardizing by dividing by $V_{A(control)}$ and taking logs gives the expected additive model

$$\log\left(\frac{V_{A(bottleneck)}}{V_{A(control)}}\right) = \log R = \log(1 - f) \quad 2$$

This model can be tested using a generalized version of the additive model

$$\log\left(\frac{V_{A(bottleneck)}}{V_{A(control)}}\right) = \log R = a + b\log(1 - f) \quad 3$$

where, with reference to equation 2, it can be seen that if only additive genetic variance is present the intercept, *a*, has an expected value of 0 and the slope, *b*, an expected value of 1.

If dominance and/or epistatic variance are present, which may be more prevalent in life history traits, non-additive effects must be incorporated. The analysis by Barton and Turelli (2004) indicates that with dominance and epistasis V_A is expected to increase until about $f = 0.4$ and thereafter decline. Van Buskirk and Willi (2006) suggested using the approximate non-additive formula for the additive genetic variance in relation to the inbreeding coefficient given by Barton and Turelli (2004)

$$V_{A(bottleneck)} = (1 - f)(V_{A(control)} + 2fV_D + 4fV_{AA} + \text{higher order}) \quad 4$$

where V_D is the dominance variance and V_{AA} is the additive by additive epistatic variance. Standardizing as above and taking logs gives

$$\log\left(\frac{V_{A(bottleneck)}}{V_{A(control)}}\right) = \log R = \log\left\{(1-f)\left(1+\frac{fc}{V_{A(control)}}\right)\right\} \qquad 5$$

where, because the dominance variances, epistatic variances and higher order terms could not be estimated from the published data, $c = 2V_D + 4V_{AA}$ + higher order. When $c = 0$ equation 5 reduces to equation 2. Statistical analysis of this model is hampered by the fact that $V_{A(control)}$ occurs on both sides of the equation, and thus a basic assumption of linear regression analysis is violated. With this caveat we can approximately test this model by fitting the generalized non-additive model

$$\log\left(\frac{V_{A(bottleneck)}}{V_{A(control)}}\right) = \log R = a + b\log\left\{(1-f)\left(1+\frac{fc}{V_{A(control)}}\right)\right\} \qquad 6$$

where *a*, *b* and *c* are fitted constants. Parameter estimates were obtained by maximum likelihood assuming a normally distributed error term. In addition to this model we also tested the general curvilinear model

$$\log\left(\frac{V_{A(bottleneck)}}{V_{A(control)}}\right) = \log R = a + bf + cf^2 \qquad 7$$

Because, as we shall show, the data fitted the generalized additive model we compared the relative fit of model 6 with model 3, which is the same as model 6 when *c* = 0, to determine if the incorporation of additive effects significantly improved the fit of the model, by constructing the F statistic, $F_{1,n-3} = \frac{SS.2 - SS.3}{SS.3/(n-3)}$. SS.2 is the residual sums

of squares of the generalized additive (2 parameter) model, SS.3 is the residual sums of squares for the generalized non-additive (3 parameter) model and *n* is the sample size (Dobson 1983). Generalized coding in SPLUS for this type of analysis is given in Roff (2006).

The above approach does not include the possible confounding effects of trait type, species, etc. which may impact the change in V_A. Therefore, we modified the generalized additive model by using stepwise regression with predictor variables paper, species, bottleneck size, trait type, and log(1-*f*). Because of sample size restrictions, interaction terms were omitted:

$$\log\left(\frac{V_{A(bottleneck)}}{V_{A(control)}}\right) = \log R = a + b\log(1-f) + c_1x_1 + c_2x_2 + c_3x_3 + c_4x_4 \qquad 8$$

where x_1,..x_4 are the four added predictor variables and c_1,…c_4 the fitted constants.

Results

Linear regression of log*R* on log(1-*f*) (i.e. the generalized additive model) is highly significant ($F_{1,105} = 7.29$, $P = 0.0081$, Fig. 2.1). The slope (1.04±0.38SE) is not significantly different from the predicted value of 1 ($t = 0.10$, df = 105, $P = 0.9201$), but the intercept (0.17±0.06) is significantly different from 0 ($t = 2.63$, df = 105, $P = 0.0098$,

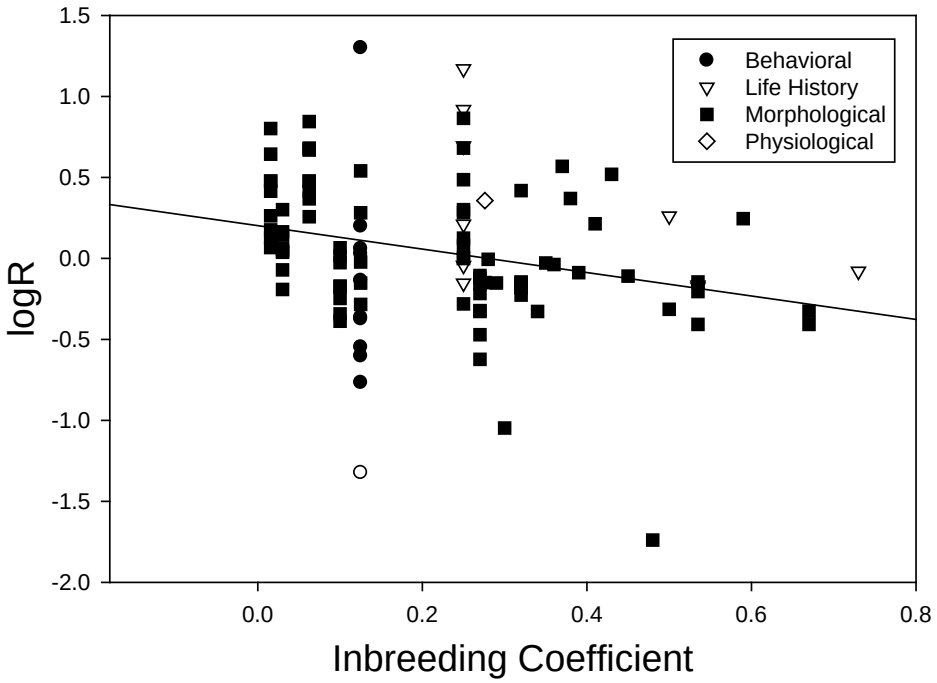

Figure 2.1. Plot of log*R* on the inbreeding coefficient, *f*, for each trait type (morphological, life history, behavioral, and physiological). The figure is plotted using the inbreeding coefficient instead of log(1-*f*) for ease of interpretation. Note that in this case the expected intercept is 0 and the expected slope is -1.

Fig. 2.1). This is not what we expect in fitting the linear model if only additive effects are present. Using only the morphological data there is still a significant fit to the generalized additive model ($F_{1,83}$ = 15.75, P = 0.0002), but again the intercept (0.23±0.06) is significantly different from 0 (t = 3.68, df = 83, P = 0.0004), and the slope (1.53±0.39) is not significantly different from 1 (t = 1.37, df = 83, P = 0.1739). There is no correlation between the life history data and the generalized additive model ($F_{1,8}$ = 2.57, P = 0.1473), but this could be due to lack of power in the linear regression analysis

caused by the small sample size for life history traits. An intercept significantly greater than zero, the expected value if there is no increase in V_A, indicates that over a range of f up to about 0.3 (as shown in Fig. 2.1) there is an increase in V_A following a bottleneck. Given that the data fitted the generalized additive model we compared the fit of the data to this model with that of the generalized non-additive model (equation 6).

The fit of the data to the generalized non-additive model produces the following equation

$$\log R = -0.07 + 0.13 \log\left\{(1-f)\left(1+\frac{f(0.95)}{V_{A(control)}}\right)\right\}$$

The non-additive model is approaching a significantly better fit than the fit of the generalized additive model ($F_{1,104} = 3.84$, $P = 0.0527$). If, as we hypothesized in the introduction, we expect to see an increase in V_A then our test is one-tailed and we would conclude the fit to be significant ($P = 0.0263$). The results suggest an increase in V_A following a bottleneck for some range of f.

When the curvilinear model (equation 7) is fit to the data, it produces results similar to the additive model. The curvilinear model gives the following equation $\log R = 0.24 - 1.15 f + 0.73 f^2$ and is a significant fit to the data ($F_{2,104} = 4.19$, $P = 0.0178$), but the intercept (0.24±0.10) is significantly different from 0 ($t = 2.50$, df = 104, $P = 0.0141$). However, the curvature is opposite to that expected. Using only the morphological data there is still a significant fit to the curvilinear model ($F_{2,82} = 9.06$, $P = 0.0003$), but again the intercept (0.31±0.09) was significantly different from 0 ($t = 3.64$,

df = 82, $P = 0.0005$). There was no correlation between the life history data and the curvilinear model ($F_{2,7} = 1.42$, $P = 0.3032$).

Stepwise multiple regression with predictor variables paper, species, bottleneck size, trait type, and log(1-*f*) gave the final model $\log R = 0.22 + 1.32\log(1-f) + c(paper)$, where *c* ranged from 0.06 to 0.94 depending on the paper ($F_{17,89} = 3.16$, $P = 0.0002$, r^2=0.38). This result indicates that the generalized additive model with intercepts varying among papers (study) is an adequate predictor of log*R*. While the standard errors attached to the partial regression coefficients do not permit a definitive statement concerning individual studies, it is noteworthy that only two of the 17 studies had a predicted intercept less than zero; that by Saccheri et al. (2001) on morphological variation in the lepidopteran *Bicyclus anynana* and that by Meffert (1995) on behavioral variation in the dipteran, *Musca domestica*. From the multiple regression model the intercept value of the regression $\log R = a + b\log(1-f)$ for any given paper is given by $0.22+c_i$, where c_i is the coefficient associated with the *i*th paper: 14 of the 16 intercepts are greater than zero supporting the previous analysis that the variance after the bottleneck is greater than before for some lower range of *f*.

In summary, we interpret these results to indicate that in general, even for morphological traits, an increase in additive genetic variance after passage through a population bottleneck is likely to occur for some values of *f*.

Discussion

The multiple regression analysis indicates that changes in additive genetic variance in bottleneck experiments are better explained by the inbreeding coefficient than the actual bottleneck size since log(1-*f*) was maintained in the stepwise regression and bottleneck size was not. This is not unexpected given that such experiments have used individuals of varying relatedness and in some cases extended bottlenecks for multiple generations prior to flushing, making the inbreeding coefficient a more appropriate indicator of the level of inbreeding than bottleneck size. If for example, a bottleneck consists of two individuals that are full sibs as opposed to two unrelated individuals then, even though the bottleneck size is two in both cases, the inbreeding coefficient at the time of the bottleneck is 0.25 in the first case and zero in the second. Because of the interaction between bottleneck size and inbreeding coefficient, studies to assess the effect of bottlenecks should maintain a constant inbreeding coefficient across experiments, such as $f = 0$ or 0.25, regardless of the number of individuals used to explicitly look at the effect of bottlenecks. The actual value of *f* in these studies is also not always measured at the time of the bottleneck, but is measured at the time V_A is assessed. As shown in Table 2.1, this could be several generations after the bottleneck event leading to *f* values that are greater than what would be expected based on the bottleneck size.

The majority of the traits assessed in the present analysis can be classified as morphological traits and are thus expected to exhibit lower magnitudes of dominance and epistatic variance than life history traits and thus would be expected to exhibit relatively small increases in V_A after passing through a bottleneck (Crnokrak & Roff 1995; Roff &

Emerson 2006). However, Roff and Emerson (2006) found that dominance effects were present in 95% of traits assessed in their study. They also found that epistatic effects were present in 67% of morphological traits. Even if the magnitude of change is expected to be small for morphological traits, given the ubiquity of non-additive genetic variance, an increase in V_A with inbreeding or passing through a bottleneck is not unexpected in morphological traits.

The fit to the generalized additive model presents an apparent paradox, because at f=0 we would expect no change in additive genetic variance. A solution to this is suggested by the two observations that the generalized non-additive model is statistically very close to being a better fit than the generalized additive model (and is so using a one-tailed test) and that paper (study) significantly increases the variance explained. We suggest that the data consists of a heterogeneous set of relationships, some of which follow the strictly additive expectation (equation 1) and others that follow the non-additive expectation (equation 4). Given that equation 4 is only an approximation and does not include all the epistatic variance, some cases may involve functions different from equation 4 though following the general pattern of an initial increase with f followed by a decline. The overall effect of this heterogeneity is to produce a general linear relationship with f and an intercept greater than zero (Fig. 2.2).

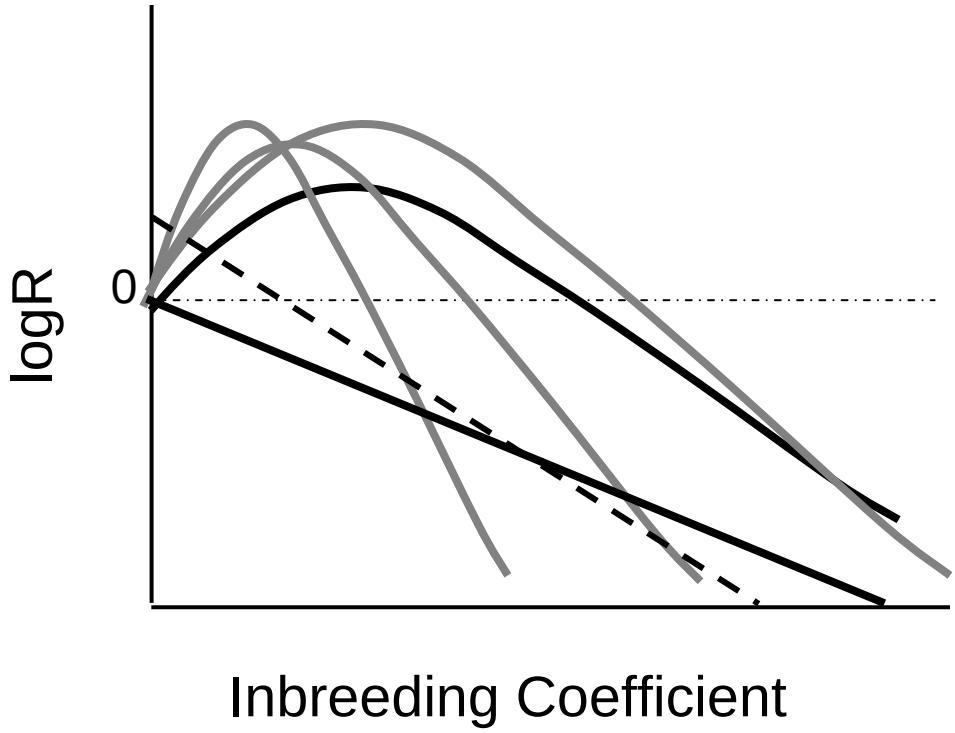

Figure 2.2. Schematic illustration of the hypothesis that the fit of the generalized additive model is a consequence of heterogeneity among papers. Cases in which there is no significant non-additive genetic variance will follow the solid linear line, whereas those in which non-additive genetic variance causes an increase in variances for some lower range in *f* are shown by the gray curves, each separate study being characterized by a particular value of *c*. When combined these gray curves average out to the solid black curve. In some cases the approximate formula for the effect of non-additive variance (equation 5) may not be appropriate but we still expect a rise and fall in log*R*. The heavy dotted line represents the overall change observed in the analysis here.

Our results are in contradiction to Van Buskirk and Willi (2006) who concluded that changes in variance of morphological traits with *f* were consistent with the strictly additive model. The difference in conclusions lies in the fact that the analysis of Van Buskirk and Willi did not consider whether the additive model was itself an adequate model, they simply compared the fit to the non-additive model. Our analysis shows that the additive model is not a sufficient explanation for the relationship between the change in additive genetic variance for morphological traits and *f* because the intercept is significantly different from the expected value of zero (on a log scale). Van Buskirk and Willi's (2006) analysis considered studies that measured both *f* and the change in additive genetic variance (and heritability) and included studies other than those examining the effect of bottlenecks. This increased the number of studies involving life history traits: however, their conclusion that the data fit the non-additive model is somewhat flawed because they did not consider alternate models, such as the null model of no fit to the data. Our data are insufficient to address the question of changes in variance in life history traits, though given the results for morphological traits we expect that stronger effects will be found in general in life history traits.

The increase in additive genetic variance following a bottleneck is a consequence of the presence of dominance variance which will also lead to inbreeding depression in the initially small bottlenecked population. Thus the probability of the population responding rapidly to selection due to an increased heritability will be mitigated to some extent by the decline in trait values. Our analysis suggests that values of *f* less than about 0.3 will give rise to increased heritabilities. Estimates of inbreeding depression for a

variety of fitness-related traits for an inbreeding coefficient of 0.25 range from 0 to 91% with a mean of 17% (SE=4.1%) and median of 8% (data from table 8.9 in Roff 1997). This could be an underestimate as many laboratory studies are included here and inbreeding depression is lower in benign environments (Armbruster & Reed 2005; Fox & Reed 2011). Fox and Reed (2011) found a 32% average decrease in fitness, not just fitness-related traits, when f=0.25 in benign and stressful environments. With such a decrease the increase in heritability would be just sufficient to counterbalance the deleterious effect of inbreeding. These calculations point to the need for direct empirical estimates of the consequences of response to selection on fitness following a bottleneck event. Using the multiple regression equation gives predicted values of *R* ranging from 0.6 to 1.6, depending on the study. Given that increases for life history traits are likely to be larger than this a twofold increase in V_A is a reasonable possibility. Heritabilities of life history traits average about 0.25 (see table 2.8 in Roff 1997) which would give a heritability after the bottleneck of R/(R+3): thus a twofold increase in V_A would lead to a 1.6 times increase, resulting in a heritability of 0.4. Such an increase could be enough to offset the decline in trait value and produce an increased response to selection. Importantly, the increase in potential response may be a significant factor in overcoming the deleterious effects of inbreeding in a population passing through a bottleneck.

Our analysis has addressed the immediate consequence of the effect of a population passing through a bottleneck. What happens in the longer term is a different question and one that will depend upon the types of changes in genetic architecture that occur as a result of the bottleneck and the subsequent effects of mutation and selection.

If only allele frequencies change then, given the same selection regime, we might expect the allele frequencies to return to their pre-bottleneck frequencies (Roff, 1997; Barton and Turelli, 2004). If alleles are lost then it is possible that a new optimum might be favored. Additive genetic variance could contribute to an increase in the evolutionary rate of bottlenecked populations. This could be very important for populations of conservation concern if the population is not extirpated by the event causing the bottleneck. The ability to adapt to changing conditions could, in principle, be enhanced by passage through a population bottleneck. However, the use of bottlenecks to achieve such a goal is not a recommended conservation approach for four reasons. First, the parameter space over which additive genetic variance increases is small and over much of the space a decrease in additive genetic variance is more likely. Second, there is no present means of predicting whether in a specific case additive genetic variance will increase even in the likely region of *f*. Third, inbreeding depression is very likely (Keller & Waller 2002) even if additive genetic variance is increased and could eliminate or reverse any effects of increased additive genetic variance (Frankham 1998; Frankham et al. 1999; Willi et al. 2006). Finally small populations have an elevated risk of extinction because of demographic stochasticity (Lande et al. 2003; Griffen & Drake 2008).

The analysis of molecular markers is important to determine population structure, effective population size and demographic history but its low or lack of correlation with quantitative genetic variation makes it a poor method for assessing the evolutionary potential of a population. For this purpose we must turn to quantitative genetics to determine the potential for response to novel conditions in wild populations (Kruuk

2004). Unfortunately, at this time use of these models is predicated on the assumption that genetic variances and covariances remain constant which may not be valid for populations passing through population bottlenecks. Further empirical and theoretical study of this problem is required to better understand its implications for conservation biology.

References

Alves DA, Imperatriz-Fonseca VL, Francoy TM, Santos-Filho PS, Billen J, Wenseleers T (2011) Successful maintenance of a stingless bee population despite a severe genetic bottleneck. *Conservation Genetics*, **12**, 647-658.

Andersson S, Ellmer M, Jorgensen TH, Palme A (2010) Quantitative genetic effects of bottlenecks: experimental evidence from a wild plant species, *Nigella degenii*. *Journal of Heredity*, **101**, 298-307.

Armbruster P, Reed DH (2005) Inbreeding depression in benign and stressful environments. *Heredity*, **95**, 235-242.

Barton NH, Turelli M (2004) Effects of genetic drift on variance components under a general model of epistasis. *Evolution*, **58**, 2111-2132.

Bellinger MR, Johnson JA, Toepfer J, Dunn P (2003) Loss of genetic variation in Greater Prairie Chickens following a population bottleneck in wisconsin, USA. *Conservation Biology*, **17**, 717-724.

Bonnell ML, Selander RK (1974) Elephant seals: genetic variation and near extinction. *Science*, **184**, 908-909.

Briggs WH, Goldman IL (2006) Genetic variation and selection response in model breeding populations of *Brassica rapa* following a diversity bottleneck. *Genetics*, **172**, 457-465.

Bryant EH, McCommas SA, Combs LM (1986) The effect of an experimental bottleneck upon quantitative genetic-variation in the housefly. *Genetics*, **114**, 1191-1211.

Bryant EH, Meffert LM (1993) The effect of serial founder-flush cycles on quantitative genetic-variation in the housefly. *Heredity*, **70**, 122-129.

Bryant EH, Meffert LM (1995) An analysis of selectional response in relation to a population bottleneck. *Evolution*, **49**, 626-634.

Bryant EH, Meffert LM (1996) Nonadditive genetic structuring of morphometric variation in relation to a population bottleneck. *Heredity*, **77**, 168-176.

Cardoso MJ, Eldridge MDB, Oakwood M, Rankmore B, Sherwin WB, Firestone KB (2009) Effects of founder events on the genetic variation of translocated island populations: implications for conservation management of the northern quoll. *Conservation Genetics*, **10**, 1719-1733.

Chan C-H, Robertson HA, Saul EK, Nia LV, Vy Phuong L, Kong X, Zhao Y, Chambers GK (2011) Genetic variation in the kakerori (Pomarea dimidiata), an endangered endemic bird successfully recovering in the Cook Islands. *Conservation Genetics*, **12**, 441-447.

Cheverud JM, Routman EJ (1996) Epistasis as a source of increased additive genetic variance at population bottlenecks. *Evolution*, **50**, 1042-1051.

Cheverud JM, Vaughn TT, Pletscher LS, King-Ellison K, Bailiff J, Adams E, Erickson C, Bonislawski A (1999) Epistasis and the evolution of additive genetic variance in populations that pass through a bottleneck. *Evolution*, **53**, 1009-1018.

Cockerham CC, Tachida H (1988) Permanency of response to selection for quantitative characters in finite populations. *Proceedings of the National Academy of Sciences of the United States of America*, **85**, 1563-1565.

Consuegra S, Verspoor E, Knox D, de Leaniz CG (2005) Asymmetric gene flow and the evolutionary maintenance of genetic diversity in small, peripheral Atlantic salmon populations. *Conservation Genetics*, **6**, 823-842.

Corti P, Shafer ABA, Coltman DW, Festa-Bianchet M (2011) Past bottlenecks and current population fragmentation of endangered huemul deer (Hippocamelus bisulcus): implications for preservation of genetic diversity. *Conservation Genetics*, **12**, 119-128.

Crnokrak P, Roff DA (1995) Dominance variance - associations with selection and fitness. *Heredity*, **75**, 530-540.

Crow JF, Kimura M (1970) *Introduction to Theoretical Population Genetics*. Harper & Row, New York.

Culver M, Hedrick PW, Murphy K, O'Brien S, Hornocker MG (2008) Estimation of the bottleneck size in Florida panthers. *Animal Conservation*, **11**, 104-110.

DeRose MA, Roff DA (1999) A comparison of inbreeding depression in life-history and morphological traits in animals. *Evolution*, **53**, 1288-1292.

Dlugosch KM, Parker IM (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, **17**, 431-449.

Dobson AJ (1983) *An Introduction to Statistical Modelling*. Chapman and Hall, London.

Durrant CJ, Beebee TJC, Greenaway F, Hill DA (2009) Evidence of recent population bottlenecks and inbreeding in British populations of Bechstein's bat, Myotis bechsteinii. *Conservation Genetics*, **10**, 489-496.

Earl DA, Louie KD, Bardeleben C, Swift CC, Jacobs DK (2010) Rangewide microsatellite phylogeography of the endangered tidewater goby, Eucyclogobius newberryi (Teleostei: Gobiidae), a genetically subdivided coastal fish with limited marine dispersal. *Conservation Genetics*, **11**, 103-114.

Fenderson LE, Kovach AI, Litvaitis JA, Litvaitis MK (2011) Population genetic structure and history of fragmented remnant populations of the New England cottontail (Sylvilagus transitionalis). *Conservation Genetics*, **12**, 943-958.

Fernandez A, Toro MA, Lopez-Fanjul C (1995) The effect of inbreeding on the redistribution of genetic variance of fecundity and viability in *Tribolium castaneum*. *Heredity*, **75**, 376-381.

Fernandez J, Rodriguez-Ramilo ST, Perez-Figueroa A, Lopez-Fanjul C, Caballero A (2003) Lack of nonadditive genetic effects on early fecundity in *Drosophila melanogaster*. *Evolution*, **57**, 558-565.

Fox CW, Reed DH (2011) Inbreeding depression increases with environmental stress: an experimental study and meta-analysis. *Evolution*, **65**, 246-258.

Frankham R (1998) Inbreeding and extinction: Island populations. *Conservation Biology*, **12**, 665-675.

Frankham R, Lees K, Montgomery ME, England PR, Lowe EH, Briscoe DA (1999) Do population size bottlenecks reduce evolutionary potential? *Animal Conservation*, **2**, 255-260.

Funk WC, Forsman ED, Johnson M, Mullins TD, Haig SM (2010) Evidence for recent population bottlenecks in northern spotted owls (Strix occidentalis caurina). *Conservation Genetics*, **11**, 1013-1021.

Garcia N, Lopez-Fanjul C, Garcia-Dorado A (1994) The genetics of viability in *Drosophila melanogaster* - effects of inbreeding and artificial selection. *Evolution*, **48**, 1277-1285.

Goodnight CJ (1987) On the effect of founder events on epistatic genetic variance. *Evolution*, **41**, 80-91.

Goodnight CJ (1988) Epistasis and the effect of founder events on the additive genetic variance. *Evolution*, **42**, 441-454.

Griffen BD, Drake JM (2008) A review of extinction in experimental populations. *Journal of Animal Ecology*, **77**, 1274-1287.

Groombridge JJ, Dawson DA, Burke T, Prys-Jones R, Brooke MdL, Shah N (2009) Evaluating the demographic history of the Seychelles kestrel (Falco araea): Genetic evidence for recovery from a population bottleneck following minimal conservation management. *Biological Conservation*, **142**, 2250-2257.

Haanes H, Roed KH, Flagstad O, Rosef O (2010) Genetic structure in an expanding cervid population after population reduction. *Conservation Genetics*, **11**, 11-20.

Habel JC, Zachos FE, Finger A, Meyer M, Louy D, Assmann T, Schmitt T (2009) Unprecedented long-term genetic monomorphism in an endangered relict butterfly species. *Conservation Genetics*, **10**, 1659-1665.

Hawley DM, Briggs J, Dhondt AA, Lovette IJ (2008) Reconciling molecular signatures across markers: mitochondrial DNA confirms founder effect in invasive North American house finches (Carpodacus mexicanus). *Conservation Genetics*, **9**, 637-643.

Heber S, Briskie JV (2010) Population Bottlenecks and Increased Hatching Failure in Endangered Birds. *Conservation Biology*, **24**, 1674-1678.

Hoelzel AR (1999) Impact of population bottlenecks on genetic variation and the importance of life-history; a case study of the northern elephant seal. *Biological Journal of the Linnean Society*, **68**, 23-39.

Hoelzel AR, Fleischer RC, Campagna C, Le Boeuf BJ, Alvord G (2002) Impact of a population bottleneck on symmetry and genetic diversity in the northern elephant seal. *Journal of Evolutionary Biology*, **15**, 567-575.

Jacquemyn H, Roldan-Ruiz I, Honnay O (2010) Evidence for demographic bottlenecks and limited gene flow leading to low genetic diversity in a rare thistle. *Conservation Genetics*, **11**, 1979-1987.

Kalinowski ST, Muhlfeld CC, Guy CS, Cox B (2010) Founding population size of an aquatic invasive species. *Conservation Genetics*, **11**, 2049-2053.

Keller LF, Waller DM (2002) Inbreeding effects in wild populations. *Trends in Ecology & Evolution*, **17**, 230-241.

Kristensen TN, Sorensen AC, Sorensen D, Pedersen KS, Sorensen JG, Loeschcke V (2005) A test of quantitative genetic theory using *Drosophila* - effects of

inbreeding and rate of inbreeding on heritabilities and variance components. *Journal of Evolutionary Biology*, **18**, 763-770.

Kruuk LEB (2004) Estimating genetic parameters in natural populations using the 'animal model'. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, **359**, 873-890.

Kuro-o M, Yonekawa H, Saito S, Eda M, Higuchi H, Koike H, Hasegawa H (2010) Unexpectedly high genetic diversity of mtDNA control region through severe bottleneck in vulnerable albatross Phoebastria albatrus. *Conservation Genetics*, **11**, 127-137.

Lande R (1980) Genetic variation and phenotypic evolution during allopatric speciation. *American Naturalist*, **116**, 463-479.

Lande R, Engen S, Saether B (2003) *Stochastic Population Dynamics in Ecology and Conservation*. Oxford University Press, New York.

Leinonen T, O'Hara RB, Cano JM, Merila J (2008) Comparative studies of quantitative trait and neutral marker divergence: a meta-analysis. *Journal of Evolutionary Biology*, **21**, 1-17.

Lints FA, Bourgois M (1984) Population crash, population flush and genetic-variability in cage populations of *Drosophila melanogaster*. *Genetics Selection Evolution*, **16**, 45-55.

Lopez-Fanjul C, Guerra J, Garcia A (1989) Changes in the distribution of the genetic variance of a quantitative trait in small populations of *Drosophila melanogaster*. *Genetics Selection Evolution*, **21**, 159-168.

Lopez-Fanjul C, Villaverde A (1989) Inbreeding increases genetic variance for viability in *Drosophila melanogaster*. *Evolution*, **43**, 1800-1804.

Meffert LM (1995) Bottleneck effects on genetic variance for courtship repertoire *Genetics*, **139**, 365-374.

Meimberg H, Hammond JI, Jorgensen CM, Park TW, Gerlach JD, Rice KJ, McKay JK (2006) Molecular evidence for an extreme genetic bottleneck during introduction of an invading grass to California. *Biological Invasions*, **8**, 1355-1366.

Merila J, Crnokrak P (2001) Comparison of genetic differentiation at marker loci and quantitative traits. *Journal of Evolutionary Biology*, **14**, 892-903.

Miller KA, Towns DR, Allendorf FW, Ritchie PA, Nelson NJ (2011) Genetic structure and individual performance following a recent founding event in a small lizard. *Conservation Genetics*, **12**, 461-473.

Moritz RFA, Haddad N, Bataieneh A, Shalmon B, Hefetz A (2010) Invasion of the dwarf honeybee Apis florea into the near East. *Biological Invasions*, **12**, 1093-1099.

Munoz-Fuentes V, Green AJ, Negro JJ, Sorenson MD (2005) Population structure and loss of genetic diversity in the endangered white-headed duck, Oxyura leucocephala. *Conservation Genetics*, **6**, 999-1015.

Neumann K, Jansman H, Kayser A, Maak S, Gattermann R (2004) Multiple bottlenecks in threatened western European populations of the common hamster Cricetus cricetus (L.). *Conservation Genetics*, **5**, 181-193.

Palo JU, O'Hara RB, Laugen AT, Laurila A, Primmer CR, Merila J (2003) Latitudinal divergence of common frog (Rana temporaria) life history traits by natural

selection: evidence from a comparison of molecular and quantitative genetic data. *Molecular Ecology*, **12**, 1963-1978.

Puillandre N, Dupas S, Dangles O, Zeddam JL, Capdevielle-Dulac C, Barbin K, Torres-Leguizamon M, Silvain JF (2008) Genetic bottleneck in invasive species: the potato tuber moth adds to the list. *Biological Invasions*, **10**, 319-333.

Reed DH (2010) Albatrosses, eagles and newts, Oh My!: exceptions to the prevailing paradigm concerning genetic diversity and population viability? *Animal Conservation*, **13**, 448-457.

Ricanova S, Bryja J, Cosson J-F, Csongor G, Choleva L, Ambros M, Sedlacek F (2011) Depleted genetic variation of the European ground squirrel in Central Europe in both microsatellites and the major histocompatibility complex gene: implications for conservation. *Conservation Genetics*, **12**, 1115-1129.

Robertson A (1952) The effect of inbreeding on the variation due to recessive genes. *Genetics*, **37**, 188-207.

Roff DA (1997) *Evolutionary Quantitative Genetics*. Chapman Hall, New York.

Roff DA (1998) Effects of inbreeding on morphological and life history traits of the sand cricket, *Gryllus firmus*. *Heredity*, **81**, 28-37.

Roff DA (2006) *Introduction to Computer-Intensive Methods of Data Analysis in Biology*. Cambridge University Press, New York.

Roff DA, Emerson K (2006) Epistasis and dominance: Evidence for differential effects in life-history versus morphological traits. *Evolution*, **60**, 1981-1990.

Russell JC, Abdelkrim J, Fewster RM (2009) Early colonisation population structure of a Norway rat island invasion. *Biological Invasions*, **11**, 1557-1567.

Saccheri IJ, Nichols RA, Brakefield PM (2001) Effects of bottlenecks on quantitative genetic variation in the butterfly *Bicyclus anynana*. *Genetical Research*, **77**, 167-181.

Sastre N, Vila C, Salinas M, Bologov VV, Urios V, Sanchez A, Francino O, Ramirez O (2011) Signatures of demographic bottlenecks in European wolf populations. *Conservation Genetics*, **12**, 701-712.

Schoville SD, Tustall TS, Vredenburg VT, Backlin AR, Gallegos E, Wood DA, Fisher RN (2011) Conservation genetics of evolutionary lineages of the endangered mountain yellow-legged frog, Rana muscosa (Amphibia: Ranidae), in southern California. *Biological Conservation*, **144**, 2031-2040.

Sinclair EA, Costello B, Courtenay JM, Crandall KA (2002) Detecting a genetic bottleneck in Gilbert's Potoroo (Potorous gilbertii) (Marsupialia : Potoroidae), inferred from microsatellite and mitochondrial DNA sequence data. *Conservation Genetics*, **3**, 191-196.

Swatdipong A, Primmer CR, Vasemagi A (2010) Historical and recent genetic bottlenecks in European grayling, Thymallus thymallus. *Conservation Genetics*, **11**, 279-292.

Tachida H, Cockerham CC (1989a) A building block model for quantitative genetics. *Genetics*, **121**, 839-844.

Tachida H, Cockerham CC (1989b) Effects of identity disequilibrium and linkage on quantitative variation in finite populations. *Genetical Research*, **53**, 63-70.

Turelli M, Barton NH (2006) Will population bottlenecks and multilocus epistasis increase additive genetic variance? *Evolution*, **60**, 1763-1776.

Van Buskirk J, Willi Y (2006) The change in quantitative genetic variation with inbreeding. *Evolution*, **60**, 2428-2434.

van Heerwaarden B, Willi Y, Kristensen TN, Hoffmann AA (2008) Population bottlenecks increase additive genetic variance but do not break a selection limit in rain forest *Drosophila*. *Genetics*, **179**, 2135-2146.

Weber D, Stewart BS, Garza JC, Lehman N (2000) An empirical genetic assessment of the severity of the northern elephant seal population bottleneck. *Current Biology*, **10**, 1287-1290.

Whitlock MC, Fowler K (1999) The changes in genetic and environmental variance with inbreeding in *Drosophila melanogaster*. *Genetics*, **152**, 345-353.

Willi Y, Van Buskirk J, Hoffmann AA (2006) Limits to the adaptive potential of small populations. *Annual Review of Ecology Evolution and Systematics*, **37**, 433-458.

Willis JH, Orr HA (1993) Increased heritable variation following population bottlenecks - the role of dominance. *Evolution*, **47**, 949-957.

Wright LI, Tregenza T, Hosken DJ (2008) Inbreeding, inbreeding depression and extinction. *Conservation Genetics*, **9**, 833-843.

Wright S (1951) The Genetical Structure of Populations. *Annals of Eugenics*, **15**, 323-354.

Chapter 3

Statistical Methods for Analyzing Threat Status

Heather R. Taft[a], Derek A. Roff[a], Atte Komonen[b] and Janne S. Kotiaho[b]

a: Department of Biology, University of California, Riverside, California 92521 USA

b: Department of Biological and Environmental Science, University of Jyväskylä, P.O. Box 35, FI-40014, Finland

Abstract

If species with shared ecological or life history characteristic also tend to share their risk of extinction, then ecological characteristics may be used to predict which species may be at risk, although not yet classified as such, by the International Union for Conservation of Nature (IUCN). To determine whether there are non-threatened or unclassified species that share the characteristics of threatened species statistical models can be used, but which model is best and whether multiple models should be used is unknown. Here three types of statistical models were compared: regression trees, logistic regression, and discriminant function analysis. The utility of these methods was compared using data on the ecological characteristics of Finnish lepidopterans (butterflies and moths). Overall, logistic regression performed slightly better than discriminant

function analysis, and both were better than regression trees. Because of uncertainty in classification we recommend that multiple analyses be performed and special attention paid to those species for which the methods disagree.

Introduction

Assessing the risk of extinction is important to determine which species are most prone to extinction and may be in need of human intervention. Categories have been defined by the International Union for Conservation of Nature (IUCN) to classify a species threat status on the basis of their risk of extinction. Classifications largely rely on quantitative information, but in practice there is strong involvement of expert opinion. Many different methods have been used to assess extinction risk, including population models (McCarthy et al. 2004), species-area correlations (Grelle 2005), and genetic analyses (Dunham et al. 1999). Obtaining the type of data needed to assess extinction risk can be problematic. For example, the assessment of population size and changes in population size can be difficult, particularly for small populations whose individuals are not easily located. Genetic data, which may provide information about the degree of inbreeding and gene flow among populations, is costly, especially if multiple species are assessed, and may also be hard to obtain from small populations. However, data on simple ecological characteristics can be obtained from existing knowledge of the natural history of species, or less extensive population monitoring, which relies heavily on volunteer work, can provide a viable alternative to assess extinction risk.

Among the studies that use ecological characteristics to assess extinction risk, several different types of analyses have been used, including multiple regression (Purvis et al. 2000; Krüger & Radford 2008), regression tree analysis (Boyer 2008; Davidson et al. 2009; Boyer 2010), logistic regression (Mattila et al. 2006; Franzen & Johannesson 2007; Mattila et al. 2008), and risk ranking (Kotiaho et al. 2005). Several of these studies used multiple tests to assess extinction risk, often initially analyzing single variables, followed by one of the statistical tests mentioned above to analyze the complete data set. Bielby et al. (2010) compared decision trees and phylogenetic comparative methods. Here we compare three statistical approaches: regression tree analysis, logistic regression, and discriminant function analysis. Our aim was to determine if one or a combination of these statistical methods can be used on data sets to classify previously unclassified species into a threat category, threatened or non-threatened, highlighting unclassified species that may need immediate attention. Reclassification of a formerly non-threatened species to a threatened status is also possible and may indicate that a species is in more immediate need of attention. The specific situation we envision is one in which most species within a particular taxon have been classified as threatened or non-threatened. If it can be shown that this classification is well predicted by one or more of the three statistical approaches using general ecological and life history parameters then unclassified species of the same taxon that share characteristics with threatened species can be identified to help prioritize further assessments of threat status.

We explore this approach using data on Finnish lepidopterans previously analyzed by Komonen et al. (2004), Kotiaho et al. (2005), Mattila et al. (2006), and Mattila et al.

(2008). Komonen et al. (2004) used analysis of variance and linear regressions on subsets of variables to assess butterfly mobility, but did not relate this to IUCN threat status. Mattila et al. (2006) and Mattila et al. (2008) used similar analyses, individually running logistic regressions on each variable with IUCN threat status as the dependent variable and then using a multinomial logistic regression to determine the ability to classify species into their correct IUCN threat status. Kotiaho et al. (2005) primarily used t-tests and logistic regressions on single variables to compare threatened and non-threatened species. Four of the variables found to be significantly related to IUCN threat status were used to create a risk ranking of all the species by ranking the species within each variable and then summing the ranks of the species across the variables. This ranking was used in a logistic regression and compared to the actual IUCN threat status of the species.

In the present analysis we address the question of whether the ecological variables in a data set consisting of threatened and non-threatened species may be used directly en masse to predict the probability of unclassified species being threatened or non-threatened and identify non-threatened species that may need reclassification of their threat status.

Methods

The Kotiaho et al. (2005) butterfly data set consisted of 94 species and 13 predictor variables: family, genus, species, abundance, distribution, distribution change, resource distribution, extent of range, larval specificity, female size, length of flight

period, mobility and habitat breadth (see Komonen et al. (2004) and Kotiaho et al. (2005) for variable definitions). Because the primary criterion for listing these species according to IUCN threat status is a function of three "distribution" variables (distribution, distribution change, and extent of range), we included these variables as predictor variables to assess whether any of the other variables were better at predicting IUCN threat status. After initial analysis these distribution variables were excluded from subsequent analyses to determine whether any more easily accessible variables (i.e. those obtainable from published natural history descriptions) could be used to predict threat status. One species, *Glaucopsyche alexis*, was listed using only abundance as the criterion and so this species was not used in the analyses. Thirteen other species were excluded due to missing data. The analyzed data consisted of 18 threatened and 62 non-threatened species. The rest of the variables, excluding resource distribution due to lack of data, were used to predict IUCN threat status as given in the 2000 Finnish Red List (Rassi et al. 2001). As noted above, one of the principal aims of the analysis was to investigate the ability of variables that are readily available from published data on the natural history of a species to determine threat status: therefore, we ran the analyses with and without the variable abundance, which might typically be difficult to estimate and not available in many cases. However, due to the similarity of the results, only the analyses excluding abundance are reported here (see Appendix A for a table of the analyses including abundance).

Two data sets on moths were used, one on Noctuids (Mattila et al. 2006) and one on Geometrids (Mattila et al. 2008). These two data sets consisted of 284 and 306

species, respectively, and each had the same eight predictor variables: genus, species, male size, length of flight period, larval specificity, resource distribution, overwintering stage, and distribution change. After analyses were run with each data set they were combined into a single data set with the added variable "taxonomic family" to increase the power of the analyses. Forty species in total from these data sets were excluded due to missing data. The analyzed data consisted of 68 threatened and 482 non-threatened species. Distribution change was the only distribution variable for these data sets and after initial analysis, was, as before, excluded to determine which other variables may be important for predicting IUCN threat status. The response variable for all data sets was binomial, threatened or not threatened. It included all species listed as near threatened or higher according to the IUCN threat status as threatened and the remainder as non-threatened.

Three types of statistical analyses were investigated: regression trees, logistic regression, and discriminant function analysis. Logistic regression and discriminant function analysis are used to create models that can predict group membership on the basis of given variables. The optimal models for predicting group membership will typically comprise a subset of the variables. Regression trees also create models that can be used to predict group membership on the basis of given variables, but allow for complex interaction among variables at different levels, and can utilize the same variables multiple times at different levels within the model, potentially leading to an increase in explained variance.

An important consideration is the probability of assigning a species to the correct category, threatened or non-threatened, by chance alone. To determine this we used simulations (coding in R given in Appendix B). First, we generated a vector, **V,** of length *N*, where *N* is the total number of species in the sample with ones in the first n_1 rows and zeros in the remaining N-n_1 rows, the former being the number of threatened species and the latter the number non-threatened species. These zeros and ones were rearranged at random in the vector. The number of correct assignments in the threatened category, N_1, is given by $N_1 = \sum_{i=1}^{n_1} V_i$, that is, the number of ones in the first n_1 rows, and the number of correct assignments in the non-threatened category, N_0 is $N_0 = N - n_1 - \sum_{i=n_1+1}^{N} V_i$, that is the number of zeros in the remaining N-n_1 rows. The total number of correct assignments, N_T, is thus $N_T = N_1 + N_0$. These three numbers were stored and the process repeated to generate three new samples. The whole process was replicated 10,000 times generating a matrix with three columns (total correctly assigned, correctly assigned to threatened, correctly assigned to non-threatened) and 10,000 rows. From this matrix we calculated the distribution of correct assignments. For each column we determined the probability of correctly assigning n_j species (j=0 to N, j=0 to n_1, j =0 to N-n_1) as S_J/10000, where S_J is the number of times n_j appeared in the relevant column.

Regression trees

Roff and Roff (2003) initially suggested regression tree analysis to determine factors contributing to the risk of extinction. Several studies have since used this

approach (Watson-Jones et al. 2006; Boyer 2008; Davidson et al. 2009; Boyer 2010). Regression trees use both categorical and continuous variables. The trees are created by splitting the data from the best predictor variable at a point called a node by using the deviance (such as least squares) to optimize the placement of the split along a range of values for the predictor variable. This creates two branches, one branch that is assessed as being more prone to extinction and one that is less prone to extinction. Splitting continues by re-analyzing the data at each node to again pick the best variable and the point within it to predict extinction risk. In regression trees the same variable may be used multiple times allowing for interactions among variables at different levels within the tree. The splitting can continue until each species has a single terminal node. The next step in the analysis determines the level of splits that have significant statistical support. This optimal, pruned, tree is first determined by cross-validation (Roff 2006, p.189), which is then verified by significance testing based on a randomization procedure with 5,000 randomizations. Roff (2006) provides detailed instructions on the implementation of the above approach with SPLUS coding available at http://www.biology.ucr.edu/people/faculty/Roff.html. The coding can readily be modified to run in R with the tree function *rpart*.

Logistic regression

Logistic regression is used to predict the placement of an observation in a discrete category, here threatened or non-threatened, based on a set of predictor variables. As with all multiple regression models, interactions can be added to the model, though the number of possible interactions rises steeply with the number of predictor variables, a

disadvantage not suffered by the regression tree approach. The optimal model is chosen using a stepwise procedure. Various metrics can be used to assess the adequacy of the model fit including Nagelkerke's R^2, which is an approximation of the R^2 used in multiple regression, and Akaike's Information Criterion (AIC). Different models can be obtained depending on both the metric that is used to fit the model and the particular type of stepwise procedure. Forward stepwise appears to generally give a simpler model than either backwards or both-ways, with the latter two typically giving the same model (Roff 2006). Which of the resulting models is the best predictor can be determined using cross-validation (see sample coding in section C.6.1 in Roff [2006], coding also available at URL given in section 2.2). In the present analysis our aim was to determine the accuracy of models created using logistic regression at predicting the correct assignment of a species as threatened or non-threatened. To use the model as a mechanism for placement of a species into the threatened or non-threatened category, we required a threshold value, e.g. 0.5, above which species were placed into one category and below which they were placed into the other. Alternatively, one could assign two thresholds, such as 0.25 and 0.75, species below the lower threshold being placed in one category, species above the upper threshold being placed in the second category and species lying between the two thresholds being classified as "uncertain." We explored the consequences of these two approaches.

An important point to note in this method of analysis is that whereas the stopping point for the stepwise regression is defined by a metric such as AIC, the adequacy of the model in the present context is measured by the assignment to the two categories:

because of this, it is possible for the "best" model to contain more or fewer variables than that specified by the "best" stepwise logistic model.

Discriminant function analysis

Discriminant function analysis predicts group membership in a categorical response variable by using several predictor variables. A model is created using a linear function of the predictor variables to determine the placement of an observation, species in this case, into a category (threatened/non-threatened). As with most statistical methods that assume univariate or multivariate normality, statistical power can be greatly decreased by outliers in the data. Problems also arise with non-normally distributed data and heterogeneity of the variance-covariance matrix, particularly when the sample size is small (Tabachnick & Fidell 2007). There are several covariance structures that can be identified when performing a discriminant function analysis: homoscedastic, spherical, proportional, group spherical, equal correlation, and heteroscedastic. The heteroscedastic structure is the most general moving up to the homoscedastic being one of the most specific with fewer parameters estimated. Principal components can also be specified for analysis in a discriminant function analysis. Once a model is created the accuracy of the model can be assessed using a predictive classification table or cross-validation. In the analyses here the covariance structure that correctly assigned the most threatened species was used for comparison among statistical tests.

Determining the preferred method

As indicated below, regression tree analysis was not satisfactory for either of the two data sets (butterflies or moths) and therefore, our further analysis focused upon

logistic regression versus discrimination function analysis. We compared the ability of these two methods to correctly classify species into the threatened or non-threatened category with a χ^2 analysis. Of particular interest are those species which were incorrectly classified according to one or both methods: we plotted the predicted values from the logistic regression against the predicted values from the discriminant function analysis to see whether the species that were classified differently fell near the 0.5 cutoff. It is safer to classify non-threatened species as threatened than it is to classify threatened species as non-threatened because in the former case a species will receive attention, but in the latter case a threatened species that needs attention will be overlooked. Therefore, we used the number of correctly classified threatened species in a final comparison of methods to determine which method is to be preferred, at least for the data sets assessed here.

Results

Among the three distribution variables used in the butterfly data set (distribution, range position, and distribution change), only distribution and range position were highly correlated (r = 0.64; Table 3.1). Correlations between any of the distribution variables and the other variables was highest for distribution and mobility (r = 0.74) and distribution and length of flight (r = 0.64), though these values were not high enough to cause problems with collinearity since they are less than 0.90 (Tabachnick & Fidell 2007, p. 89-90). None of the correlations among variables for the Geometrid or Noctuid data sets exceeded an absolute value of 0.26 and thus did not pose problems with collinearity.

Table 3.1. Correlations among the continuous variables used in the analyses of the butterfly data (sample size = 18 threatened species and 62 non-threatened species).

	Abundance	Distribution	Range position	Mobility	Length of flight	Female size
Distribution	-0.42**					
Range position	-0.48***	0.64***				
Mobility	-0.55***	0.74***	0.42***			
Length of flight	-0.24*	0.64***	0.29*	0.57***		
Female size	-0.10	0.07	-0.11	0.31**	0.19	
Distribution change	-0.33	0.32**	0.17	0.35***	0.21*	-0.10

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, probabilities not corrected for multiple test.

Regression trees

A significant regression tree, with two nodes was obtained ($p = 0.0002$) using only the distribution variables from the butterfly data set. This split in the regression tree was based on distribution and correctly classified 94% of threatened species and 95% of non-threatened species. When distribution change was used in a regression tree analysis on the Geometrid data the pruned regression tree was significant ($p = 0.0006$) and had two nodes but all threatened species were misclassified as non-threatened. It appears that the threatened species did not have any defining range of distribution change to use for dividing the data. When just distribution change was used in a regression tree analysis on the Noctuid data set the pruned regression tree was significant ($p = 0.0002$) and had five

nodes with a misclassification rate of 11%. Using distribution change in a regression tree analysis on the combined moth data set, the pruned regression tree was significant (p = 0.0002) and had four nodes but all threatened species were misclassified again indicating no defined range for the threatened species.

Using variables from the butterfly data set not explicitly used for determining IUCN threat status (family, mobility, larval specificity, habitat breadth, female size, and flight length), a pruned tree could not be created for the model excluding the distribution variables and abundance because only one terminal node was produced during the cross-validation. Thus, in this case, regression tree analysis could not discriminate threatened from non-threatened species.

When all the variables (male size, length of flight period, larval specificity, and overwintering stage) except distribution change were used from the Geometrid data set, a non-significant pruned tree with two nodes resulted (p = 0.2826), but no threatened species were correctly classified because both nodes were classified as non-threatened in the regression tree. Thus, breaking the data down into these two nodes based on length of flight period did not allow for enough subdivision of the data to correctly assign IUCN threat status. When all the variables from the Noctuid data set or the combined data set (male size, length of flight period, larval specificity, and overwintering stage) except distribution change were used, a pruned tree could not be created because only one terminal node was produced during the cross-validation.

We conclude that for these data sets regression tree analysis did not result in a satisfactory prediction of status.

Logistic regression

A stepwise logistic regression on the distribution variables from the butterfly data set retained the variables distribution and distribution change and was able to correctly classify 94% of threatened species and 95% of non-threatened species. The logistic regression on distribution change in the Geometrid data set correctly classified 7% of threatened species and 99% of non-threatened species. Using the Noctuid data set or the combined data set the logistic regression was unable to correctly classify any threatened species.

Because of limited data in the butterfly data set, we restricted the logistic regression analysis to additive models only (i.e. interactions were excluded). Incorporating all the variables except the distribution variables and abundance produced a model that correctly assigned 67% of threatened and 95% of non-threatened species when the fitted value cutoff point was 0.50 (Table 3.2). This cutoff point (0.5) for the fitted values produced the highest correct assignment for both threatened and non-threatened species. The alternate criterion of 0.25 as the cutoff point for the non-threatened species and anything greater than 0.75 as a cutoff point for the threatened species gave correct assignments of 39% for the threatened species and 79% for the non-threatened species, with 21 ambiguous species. A similar response to the cutoff values was seen for the moth data (results not shown), so all reported classification assignments for the logistic regressions use the 0.5 cutoff.

Table 3.2. Classification of threatened and non-threatened species by logistic regression and discriminant function analysis[a] for the butterfly data when the distribution variables and abundance are excluded from the analysis.[b]

	Threatened		Non-threatened	
	Predicted	P^c	Predicted	p
Logistic regression - 0.5 cutoff	12/18	<0.0001	59/62	<0.0001
Stepwise logistic regression	10/18	0.0005	57/62	<0.0001
Discriminant function analysis - homoscedastic	10/18	0.0005	58/62	<0.0001

a: Regression tree analysis was excluded because a pruned tree could not be created.

b: When multiple analyses were performed, such as when different structures were used for discriminant analysis, only the analysis with the best result is given.

c: Probability of correctly predicting by chance alone at least as many as observed by a given method.

The stepwise logistic regression on the Geometrid data set using all variables except distribution change included interactions due to a larger sample size and retained all variables and their interaction terms except the four-way interaction. This regression correctly assigned 47% of the threatened species, and was able to correctly classify all but one of the non-threatened species ($F_{47,263} = 2.84$, $p < 0.001$; AIC = 168.11; Table 3.3). The stepwise logistic regression using all the variables from the Noctuid data set except

distribution change retained all the variables and their interaction terms, did poorly at correctly assigning threatened species, but well at assigning non-threatened species (21% and 96% respectively; $F_{61,244} < 0.001$, $p = 1$; AIC = 2887.41; Table 3.3). The stepwise logistic regression on all variables from the combined data set except distribution change retained all the variables and their four-way interaction terms except the interaction involving family, male size, larval specificity, and overwintering stage, correctly assigned one third of threatened species (29%), and was able to correctly assign almost all the non-threatened species (96%; $F_{112,477} < 0.001$, $p = 1$; AIC = 4687.5; Table 3.3).

Discriminant function analysis

Discriminant function analysis produced results that were not quite as good as the logistic regression at classifying threatened and non-threatened species (88% and 89% respectively) using only the distribution variables from the butterfly data set. For the Geometrid, Noctuid and combined data sets 90%, 100% and 96% of threatened species respectively were correctly classified and 52%, 46% and 53% of non-threatened species were correctly classified using just distribution change.

The discriminant function analysis on the butterfly data set including all variables except the distribution variables and abundance was significant (Table 3.2), correctly classifying 56% of threatened species and 90% of non-threatened species. Weighting was highest for family and habitat breadth when classifying threatened species.

The discriminant function analysis for the Geometrid data set including all the variables except distribution change performed best using the principal components

Table 3.3. Classification of threatened and non-threatened species by regression tree analysis, logistic regression, and discriminant function analysis for the moth data when distribution change was excluded from the analysis.[a]

	Geometrids				Noctuids				All Moth Data			
	Threatened		Non-threatened		Threatened		Non-threatened		Threatened		Non-threatened	
	Predicted	*p*	Predicted	*p*	Predicted	*p*	Predicted	*p*	Predicted	*p*	Predicted	*p*
Regression tree	0/30	1	238/238	<0.0001								
Logistic regression - 0.5 cutoff	14/30	<0.0001	237/238	<0.0001	8/38	0.1166	234/244	<0.0001	27/68	<0.0001	404/482	1
Stepwise logistic regression	14/30	<0.0001	237/238	<0.0001	8/38	0.1166	234/244	<0.0001	20/68	0.0001	465/482	<0.0001
Discriminant function analysis	8/30	0.0087	226/238	<0.0001	12/38	0.0013	214/244	0.1166	13/68	0.0566	455/482	<0.0001

a: Data for the regression tree analysis was excluded for the Noctuid and combined data sets because a pruned tree could not be created.

model and correctly classified 27% of the threatened species and 95% of the non-threatened species (Table 3.3). The best discriminant analysis including all the variables from the Noctuid data set except distribution change used the equal correlation model and correctly assigned 88% of non-threatened species (88%) and 32% of threatened species. The best discriminant analysis on the combined data set including all variables except distribution change used the equal correlation model also, and assigned 94% of non-threatened species, but only 19% of threatened species.

Which method is best?

Logistic regression and discriminant function analysis were able to correctly classify a significant number of threatened and non-threatened species for most analyses. Using the butterfly data set the logistic regression and the discriminant function analysis did not differ in the number of correctly classified threatened and non-threatened species ($\chi^2 = 0.1257$, df = 1, $p = 0.723$). However, using the combined moth data the logistic regression and the discriminant function analysis did differ significantly in the number of correctly classified threatened and non-threatened species ($\chi^2 = 6.416$, df = 1, $p = 0.0113$).

Table 3.4 indicates that logistic regression and discriminant function analysis agreed on the classification, rightly or wrongly, of all but two of the threatened butterfly species with the logistic regression classifying the species “correctly” according to the published IUCN red list. Of particular interest is the fact that the incorrectly classified butterfly species do not cluster about the intersection of the 0.5 cutoff (vertical and horizontal lines in Fig. 3.1) demarking the transition from threatened to non-threatened

Table 3.4. Comparison of classifications by logistic regression and discriminant function analysis for threatened butterfly and moth species.

Logistic regression classification	Discriminant function analysis classification	Number of threatened butterfly species classified as indicated	Number of threatened moth species classified as indicated
Threatened	Threatened	10	8
Non-threatened	Threatened	0	5
Threatened	Non-threatened	2	19
Non-threatened	Non-threatened	6	36

species for each analysis (the moth data could not be plotted this way because the fitted values for the logistic regression were all 0s and 1s). The two species that were correctly classified by the logistic regression but not the discriminant function analysis are not notable outliers: in one case the species lies close to the "decision" boundary for the discriminant function analysis while in the other case the species lies close to the "decision" boundary for the logistic regression. For such species the discrepancy in the analyses suggests a closer inspection. While the logistic correctly classified more species than the discriminant function analysis the difference overall is relatively minor and we recommend that both methods be used with additional attention being paid to those species classified differently.

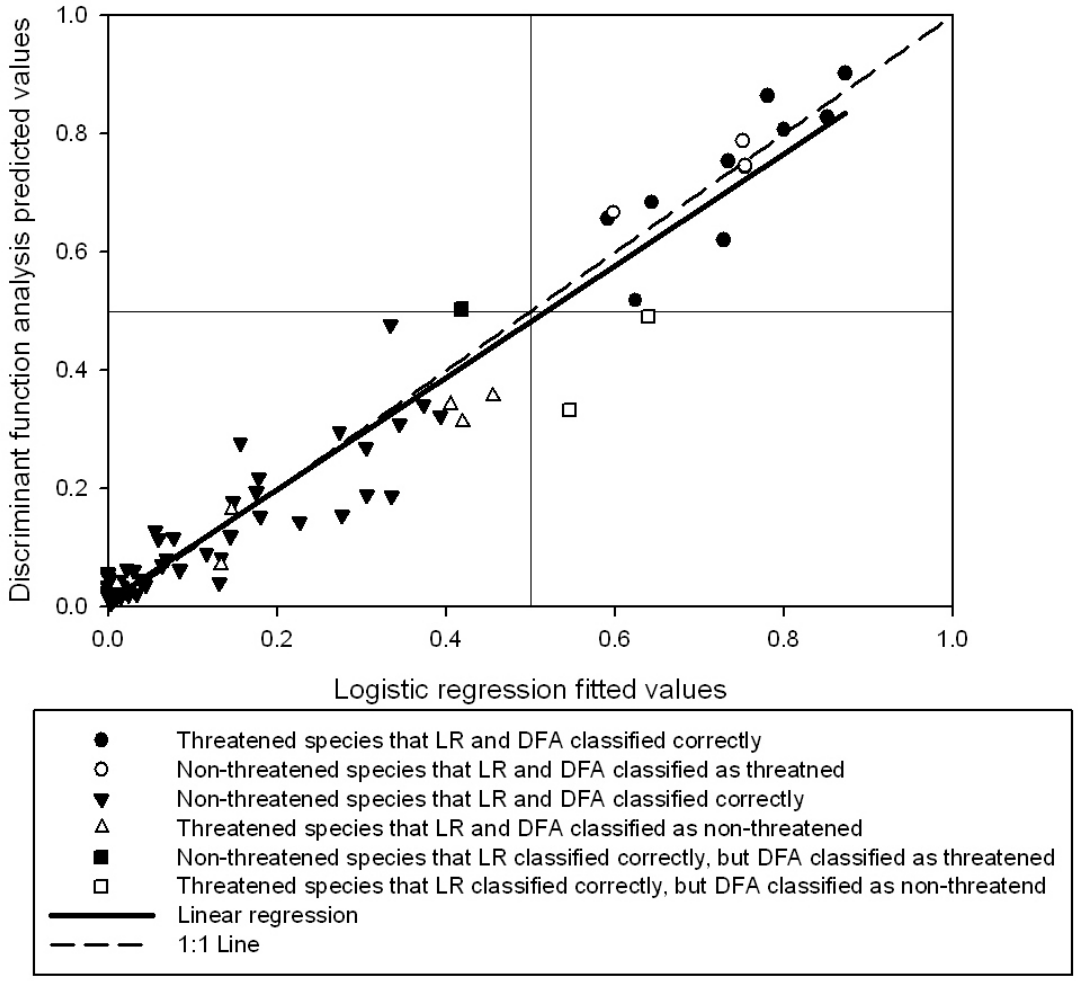

Figure 3.1. Plot of the logistic regression (LR) fitted values versus the discriminant function analysis (DFA) predicted values indicating the two methods have approximately a 1:1 relationship in their prediction of threat status.

Discussion

Regression trees, logistic regression, and discriminant function analysis were used to determine which analysis was able to more accurately predict IUCN threat status in a set of data on butterfly species and two sets of data on moth species in Finland. All of the analyses on the butterfly data and most of the analyses on the moth data sets were able to correctly assign a significant number of threatened and non-threatened species. Regression tree analysis was not very helpful in classifying species in the two data sets although previous analyses have shown promise in its use for such a purpose (Roff & Roff 2003; Watson-Jones et al. 2006; Boyer 2008; Davidson et al. 2009; Boyer 2010). Importantly, the approach here did not fail because it incorrectly classified species. For the present data sets, it failed because no trees could be produced. Thus we recommend that this approach still be tried for other data sets.

There was a decrease in the percentage of correctly classified species when a stepwise logistic regression was used on the butterfly data set and on several moth regressions. As noted above, this is not unexpected as the criterion for the best fitting model is not the same as the criterion for the stopping point in the stepwise regression. Thus for logistic regression analysis a second step, namely a comparison of the models predicting correct assignment is called for, which can be done using simulation (see Appendix B). Since we are primarily focused on the category the species fall within, and not necessarily how good individual variables are at classifying the species, a case can be made for leaving all the variables in the model, because including variables will not increase the misclassification rate. Comparing the full model and the stepwise model can

then be useful to determine the ecological characteristics that have the strongest correlation with IUCN threat status, and which variables increase the number of correctly assigned species even if the difference is not significant.

An important point to note is that using the overall correct classification rate can be highly misleading. For example, suppose that 90% of species were classified as non-threatened and the statistical analysis classified all species as non-threatened, then the overall correct classification rate is 90%, which appears to be very good but is actually of little use. This issue was particularly evident in the moth data sets and illustrates the importance of analyzing the ability of the statistical analysis to classify species into each category, as done here. When methods of comparison such as these are used classification to both categories should always be reported.

Overall, the logistic regression gave the best results followed by the discriminant function analysis. All variables can be used in these analyses, so deciding among variables is not an issue. Given the ease with which the analyses here can be performed, we suggest that multiple analyses be undertaken to identify species that may not be consistently classified as threatened.

Suggested reassessments based on the current analysis

The butterfly species *Boloria frigga, B. freija,* and *B. thore,* were classified as threatened by logistic regression and discriminant function analysis, though they are not threatened according to the IUCN red list. These species may be at an increased risk of extinction. In Kotiaho et al. (2005) these species had an Ecological Risk Rank of 4, 11, and 15 respectively. Our results concur with Kotiaho et al. (2005), that a reassessment of

the IUCN threat status of these species may be necessary. Using just discriminant function analysis *Pyrgus centaureae* was also classified as threatened. In Kotiaho et al. (2005) this species ranked eighth and may deserve a reassessment as well, but the three species mentioned previously should be the priority.

For the moth data sets five non-threatened species were classified as threatened by both logistic regression and discriminant function analysis. All the misclassified moth species fall in the family Noctuidae. *Cucullia gnaphalii*, *Dryobotodes eremita*, and *Orthosia populeti* were classified as threatened when the Noctuid data set was used. *D. eremita*, *Abrostola triplasia*, and *A. tripartita* were classified as threatened when the combined data set was used. Re-evaluation and assigning an IUCN threat status could enhance the species' chance for recovery by increasing the attention and potentially the management efforts that would otherwise not be received.

This analysis has shown that a variety of statistical analyses can produce useful assessments of threat status and that data readily available on ecological characteristics can be useful for identifying species in need of reassessment of their threat status. Extending the analyses beyond the geographical area of interest should be done cautiously as important variables may change in different locations and at different scales (Nylin & Bergstrom 2009). In particular, if abundance is to be used as a variable, understanding its relationship to the species being assessed will be important to determine whether a positive or negative relationship to distribution can be generalized (See the following articles for a debate on positive and negative density-distribution relationships Paivinen et al. 2005; Blackburn et al. 2006; Blackburn & Gaston 2009; Komonen et al.

2009; Kotiaho et al. 2009; Selonen & Helos 2010; Komonen et al. 2011). These caveats notwithstanding, the present results suggest that "off-the-shelf" statistical methods such as logistic regression and discriminant function analysis can be extremely valuable in determining the IUCN threat status of a species in areas for which only limited abundance data are available.

References

Bielby J, Cardillo M, Cooper N, Purvis A (2010) Modelling extinction risk in multispecies data sets: phylogenetically independent contrasts versus decision trees. *Biodiversity and Conservation*, **19**, 113-127.

Blackburn TM, Cassey P, Gaston KJ (2006) Variations on a theme: sources of heterogeneity in the form of the interspecific relationship between abundance and distribution. *Journal of Animal Ecology*, **75**, 1426-1439.

Blackburn TM, Gaston KJ (2009) Sometimes the obvious answer is the right one: a response to 'Missing the rarest: is the positive interspecific abundance-distribution relationship a truly general macroecological pattern?' *Biology Letters*, **5**, 777-778.

Boyer AG (2008) Extinction patterns in the avifauna of the Hawaiian islands. *Diversity and Distributions*, **14**, 509-517.

Boyer AG (2010) Consistent ecological selectivity through time in pacific island avian extinctions. *Conservation Biology*, **24**, 511-519.

Davidson AD, Hamilton MJ, Boyer AG, Brown JH, Ceballos G (2009) Multiple ecological pathways to extinction in mammals. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 10702-10705.

Dunham J, Peacock M, Tracy CR, Nielsen J, Vinyard G (1999) Assessing extinction risk: Integrating genetic information. *Conservation Ecology*, **3**.

Franzen M, Johannesson M (2007) Predicting extinction risk of butterflies and moths (Macrolepidoptera) from distribution patterns and species characteristics. *Journal of Insect Conservation*, **11**, 367-390.

Grelle CEV (2005) Predicting extinction of mammals in the Brazilian Amazon. *Oryx*, **39**, 347-350.

Komonen A, Grapputo A, Kaitala V, Kotiaho JS, Paivinen J (2004) The role of niche breadth, resource availability and range position on the life history of butterflies. *Oikos*, **105**, 41-54.

Komonen A, Paivinen J, Kotiaho JS (2009) Missing the rarest: is the positive interspecific abundance-distribution relationship a truly general macroecological pattern? *Biology Letters*, **5**, 492-494.

Komonen A, Paivinen J, Kotiaho JS (2011) Varying definitions of abundance and incomplete assemblages challenge the generality of the interspecific abundance-distribution relationships. *Annales Zoologici Fennici*, **48**, 161-166.

Kotiaho JS, Kaitala V, Komonen A, Paivinen J (2005) Predicting the risk of extinction from shared ecological characteristics. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 1963-1967.

Kotiaho JS, Komonen A, Paivinen J (2009) On the obvious positive interspecific relationship between abundance and distribution: a reply to Blackburn and Gaston. *Biology Letters*, **5**, 779-780.

Krüger O, Radford AN (2008) Doomed to die? Predicting extinction risk in the true hawks Accipitridae. *Animal Conservation*, **11**, 83-91.

Mattila M, Kotiaho JS, Kaitala V, Komonen A (2008) The use of ecological traits in extinction risk assessments: A case study on geometrid moths. *Biological Conservation*, **141**, 2322-2328.

Mattila N, Kaitala V, Komonen A, Kotiaho JS, Paivinen J (2006) Ecological determinants of distribution decline and risk of extinction in moths. *Conservation Biology*, **20**, 1161-1168.

McCarthy MA, Keith D, Tietjen J, Burgman MA, Maunder M, Master L, Brook BW, Mace G, Possingham HP, Medellin R, Andelman S, Regan H, Regan T, Ruckelshaus M (2004) Comparing predictions of extinction risk using models and subjective judgement. *Acta Oecologica-International Journal of Ecology*, **26**, 67-74.

Nylin S, Bergstrom A (2009) Threat status in butterflies and its ecological correlates: how far can we generalize? *Biodiversity and Conservation*, **18**, 3243-3267.

Paivinen J, Grapputo A, Kaitala V, Komonen A, Kotiaho J, Saarinen K, Wahlberg N (2005) Negative density-distribution relationship in butterflies. *BMC Biology*, **3**, 5.

Purvis A, Gittleman JL, Cowlishaw G, Mace GM (2000) Predicting extinction risk in declining species. *Proceedings of the Royal Society B-Biological Sciences*, **267**, 1947-1952.

Rassi P, Alanen A, Kanerva T, Mannerkoski I (2001) The Red List of Finnish Species. Ministry of the Environment and Finnish Environment Institute, Helsinki.

Roff DA (2006) *Introduction to Computer-Intensive Methods of Data Analysis in Biology*. Cambridge University Press, New York.

Roff DA, Roff RJ (2003) Of rats and Maoris: a novel method for the analysis of patterns of extinction in the New Zealand avifauna before European contact. *Evolutionary Ecology Research*, **5**, 759-779.

Selonen V, Helos I (2010) Positive density-distribution relationship in Finnish butterflies. *Annales Zoologici Fennici*, **47**, 347-350.

Tabachnick BG, Fidell LS (2007) *Using Multivariate Statistics*, Fifth edn. Pearson Education, Inc., Boston.

Watson-Jones SJ, Maxted N, Ford-Lloyd BV (2006) Population baseline data for monitoring genetic diversity loss for 2010: A case study for Brassica species in the UK. *Biological Conservation*, **132**, 490-499.

Appendix A

Studies used in the meta-analysis assessing the correlation between genetic divergence and recommendations for conservation.

Anderson FE (2007) Population genetics of the carinate pillsnail, Euchemotrema hubrichti: genetic structure on a small spatial scale. *Conservation Genetics*, **8**, 965-975.

Angienda PO, Lee HJ, Elmer KR, Abila R, Waindi EN, Meyer A (2011) Genetic structure and gene flow in an endangered native tilapia fish (Oreochromis esculentus) compared to invasive Nile tilapia (Oreochromis niloticus) in Yala swamp, East Africa. *Conservation Genetics*, **12**, 243-255.

Aragundi S, Hamrick JL, Parker KC (2011) Genetic insights into the historical distribution of Polylepis pauta (Rosaceae) in the northeastern Cordillera Oriental of Ecuador. *Conservation Genetics*, **12**, 607-618.

Aspi J, Roininen E, Kiiskila J, Ruokonen M, Kojola I, Bljudnik L, Danilov P, Heikkinen S, Pulliainen E (2009) Genetic structure of the northwestern Russian wolf populations and gene flow between Russia and Finland. *Conservation Genetics*, **10**, 815-826.

Attard CRM, Beheregaray LB, Jenner C, Gill P, Jenner M, Morrice M, Bannister J, LeDuc R, Moller L (2010) Genetic diversity and structure of blue whales

(Balaenoptera musculus) in Australian feeding aggregations. *Conservation Genetics*, **11**, 2437-2441.

Bester-van der Merwe AE, Roodt-Wilding R, Volckaert FAM, D'Amato ME (2011) Historical isolation and hydrodynamically constrained gene flow in declining populations of the South-African abalone, Haliotis midae. *Conservation Genetics*, **12**, 543-555.

Boessenkool S, Taylor SS, Tepolt CK, Komdeur J, Jamieson IG (2007) Large mainland populations of South Island robins retain greater genetic diversity than offshore island refuges. *Conservation Genetics*, **8**, 705-714.

Bush KL, Dyte CK, Moynahan BJ, Aldridge CL, Sauls HS, Battazzo AM, Walker BL, Doherty KE, Tack J, Carlson J, Eslinger D, Nicholson J, Boyce MS, Naugle DE, Paszkowski CA, Coltman DW (2011) Population structure and genetic diversity of greater sage-grouse (Centrocercus urophasianus) in fragmented landscapes at the northern edge of their range. *Conservation Genetics*, **12**, 527-542.

Butcher PA, McNee SA, Krauss SL (2009) Genetic impacts of habitat loss on the rare ironstone endemic Tetratheca paynterae subsp paynterae. *Conservation Genetics*, **10**, 1735-1746.

Campbell S, Guay PJ, Mitrovski PJ, Mulder R (2009) Genetic differentiation among populations of a specialist fishing bat suggests lack of suitable habitat connectivity. *Biological Conservation*, **142**, 2657-2664.

Castellano CM, Behler JL, Amato G (2009) Genetic diversity and population genetic structure of the Wood Turtle (Glyptemys insculpta) at Delaware Water Gap National Recreation Area, USA. *Conservation Genetics*, **10**, 1783-1788.

Chaves-Campos J, Johnson SG, de Leon FJG, Hulsey CD (2011) Phylogeography, genetic structure, and gene flow in the endemic freshwater shrimp Palaemonetes suttkusi from Cuatro Ci,negas, Mexico. *Conservation Genetics*, **12**, 557-567.

Coleman RA, Pettigrove V, Raadik TA, Hoffmann AA, Miller AD, Carew ME (2010) Microsatellite markers and mtDNA data indicate two distinct groups in dwarf galaxias, Galaxiella pusilla (Mack) (Pisces: Galaxiidae), a threatened freshwater fish from south-eastern Australia. *Conservation Genetics*, **11**, 1911-1928.

Cook BD, Bunn SE, Hughes JM (2007) Molecular genetic and stable isotope signatures reveal complementary patterns of population connectivity in the regionally vulnerable southern pygmy perch (Nannoperca australis). *Biological Conservation*, **138**, 60-72.

Crawford LA, Desjardins S, Keyghobadi N (2011) Fine-scale genetic structure of an endangered population of the Mormon metalmark butterfly (Apodemia mormo) revealed using AFLPs. *Conservation Genetics*, **12**, 991-1001.

Dawe KL, Shafer ABA, Herman TB, Stewart DT (2009) Diffusion of nuclear and mitochondrial genes across a zone of secondary contact in the maritime shrew, Sorex maritimensis: implications for the conservation of a Canadian endemic mammal. *Conservation Genetics*, **10**, 851-857.

de Oliveira LR, Hoffman JI, Hingst-Zaher E, Majluf P, Muelbert MMC, Morgante JS, Amos W (2008) Morphological and genetic evidence for two evolutionarily significant units (ESUs) in the South American fur seal, Arctocephalus gazella. *Conservation Genetics*, **9**, 1451-1466.

Dellicour S, Frantz AC, Colyn M, Bertouille S, Chaumont F, Flamand MC (2011) Population structure and genetic diversity of red deer (Cervus elaphus) in forest fragments in north-western France. *Conservation Genetics*, **12**, 1287-1297.

Dong L, Zhang J, Sun Y, Liu Y, Zhang YY, Zheng GM (2010) Phylogeographic patterns and conservation units of a vulnerable species, Cabot's tragopan (Tragopan caboti), endemic to southeast China. *Conservation Genetics*, **11**, 2231-2242.

Dostalek T, Munzbergova Z, Plackova I (2010) Genetic diversity and its effect on fitness in an endangered plant species, Dracocephalum austriacum L. *Conservation Genetics*, **11**, 773-783.

Drees C, Zumstein P, Buck-Dobrick T, Hardtle W, Matern A, Meyer H, von Oheimb G, Assmann T (2011) Genetic erosion in habitat specialist shows need to protect large peat bogs. *Conservation Genetics*, **12**, 1651-1656.

Echelle AA, Hackler JC, Lack JB, Ballard SR, Roman J, Fox SF, Leslie DM, Van Den Bussche RA (2010) Conservation genetics of the alligator snapping turtle: cytonuclear evidence of range-wide bottleneck effects and unusually pronounced geographic structure. *Conservation Genetics*, **11**, 1375-1387.

Escalona T, Engstrom TN, Hernandez OE, Bock BC, Vogt RC, Valenzuela N (2009) Population genetics of the endangered South American freshwater turtle,

Podocnemis unifilis, inferred from microsatellite DNA data. *Conservation Genetics*, **10**, 1683-1696.

Fady B, Lefevre F, Vendramin GG, Ambert A, Regnier C, Bariteau M (2008) Genetic consequences of past climate and human impact on eastern Mediterranean Cedrus libani forests. Implications for their conservation. *Conservation Genetics*, **9**, 85-95.

Faulks LK, Gilligan DM, Beheregaray LB (2010) Evolution and maintenance of divergent lineages in an endangered freshwater fish, Macquaria australasica. *Conservation Genetics*, **11**, 921-934.

Fenderson LE, Kovach AI, Litvaitis JA, Litvaitis MK (2011) Population genetic structure and history of fragmented remnant populations of the New England cottontail (Sylvilagus transitionalis). *Conservation Genetics*, **12**, 943-958.

Furches MS, Wallace LE, Helenurm K (2009) High genetic divergence characterizes populations of the endemic plant Lithophragma maximum (Saxifragaceae) on San Clemente Island. *Conservation Genetics*, **10**, 115-126.

Furmankiewicz J, Altringham J (2007) Genetic structure in a swarming brown long-eared bat (Plecotus auritus) population: evidence for mating at swarming sites. *Conservation Genetics*, **8**, 913-923.

Gaspari S, Airoldi S, Hoelzel AR (2007) Risso's dolphins (Grampus griseus) in UK waters are differentiated from a population in the Mediterranean Sea and genetically less diverse. *Conservation Genetics*, **8**, 727-732.

Gong W, Gu L, Zhang DX (2010) Low genetic diversity and high genetic divergence caused by inbreeding and geographical isolation in the populations of endangered species Loropetalum subcordatum (Hamamelidaceae) endemic to China. *Conservation Genetics*, **11**, 2281-2288.

Goulson D, Kaden JC, Lepais O, Lye GC, Darvill B (2011) Population structure, dispersal and colonization history of the garden bumblebee Bombus hortorum in the Western Isles of Scotland. *Conservation Genetics*, **12**, 867-879.

Griffiths AM, Sims DW, Johnson A, Lynghammar A, McHugh M, Bakken T, Genner MJ (2011) Levels of connectivity between longnose skate (Dipturus oxyrinchus) in the Mediterranean Sea and the north-eastern Atlantic Ocean. *Conservation Genetics*, **12**, 577-582.

Guay PJ, Chesser RT, Mulder RA, Afton AD, Paton DC, McCracken KG (2010) East-west genetic differentiation in Musk Ducks (Biziura lobata) of Australia suggests late Pleistocene divergence at the Nullarbor Plain. *Conservation Genetics*, **11**, 2105-2120.

Guschanski K, Olivieri G, Funk SM, Radespiel U (2007) MtDNA reveals strong genetic differentiation among geographically isolated populations of the golden brown mouse lemur, Microcebus ravelobensis. *Conservation Genetics*, **8**, 809-821.

Haanes H, Roed KH, Flagstad O, Rosef O (2010) Genetic structure in an expanding cervid population after population reduction. *Conservation Genetics*, **11**, 11-20.

Haanes H, Roed KH, Solberg EJ, Herfindal I, Saether BE (2011) Genetic discontinuities in a continuously distributed and highly mobile ungulate, the Norwegian moose. *Conservation Genetics*, **12**, 1131-1143.

Hammer MP, Unmack PJ, Adams M, Johnson JB, Walker KF (2010) Phylogeographic structure in the threatened Yarra pygmy perch Nannoperca obscura (Teleostei: Percichthyidae) has major implications for declining populations. *Conservation Genetics*, **11**, 213-223.

Hargreaves S, Maxted N, Hirano R, Abberton M, Skot L, Ford-Lloyd BV (2010) Islands as refugia of Trifolium repens genetic diversity. *Conservation Genetics*, **11**, 1317-1326.

Hefti-Gautschi B, Pfunder M, Jenni L, Keller V, Ellegren H (2009) Identification of conservation units in the European Mergus merganser based on nuclear and mitochondrial DNA markers. *Conservation Genetics*, **10**, 87-99.

Herr CA, Sykes PW, Klicka J (2011) Phylogeography of a vanishing North American songbird: the Painted Bunting (Passerina ciris). *Conservation Genetics*, **12**, 1395-1410.

Hmeljevski KV, Reis A, Montagna T, dos Reis MS (2011) Genetic diversity, genetic drift and mixed mating system in small subpopulations of Dyckia ibiramensis, a rare endemic bromeliad from Southern Brazil. *Conservation Genetics*, **12**, 761-769.

Honnay O, Adriaens D, Coart E, Jacquemyn H, Roldan-Ruiz I (2007) Genetic diversity within and between remnant populations of the endangered calcareous grassland plant Globularia bisnagarica L. *Conservation Genetics*, **8**, 293-303.

Hu LJ, Uchiyama K, Shen HL, Ide Y (2010) Multiple-scaled spatial genetic structures of Fraxinus mandshurica over a riparian-mountain landscape in Northeast China. *Conservation Genetics*, **11**, 77-87.

Hull JM, Anderson R, Bradbury M, Estep JA, Ernest HB (2008) Population structure and genetic diversity in Swainson's Hawks (Buteo swainsoni): implications for conservation. *Conservation Genetics*, **9**, 305-316.

Izawa T, Kawahara T, Takahashi H (2007) Genetic diversity of an endangered plant, Cypripedium macranthos var. rebunense (Orchidaceae): background genetic research for future conservation. *Conservation Genetics*, **8**, 1369-1376.

Jacquemyn H, Brys R, Adriaens D, Honnay O, Roldan-Ruiz I (2009) Effects of population size and forest management on genetic diversity and structure of the tuberous orchid Orchis mascula. *Conservation Genetics*, **10**, 161-168.

Jacquemyn H, Roldan-Ruiz I, Honnay O (2010) Evidence for demographic bottlenecks and limited gene flow leading to low genetic diversity in a rare thistle. *Conservation Genetics*, **11**, 1979-1987.

Jones JM, Gibson JP (2011) Population genetic diversity and structure within and among disjunct populations of Alnus maritima (seaside alder) using microsatellites. *Conservation Genetics*, **12**, 1003-1013.

Juan Borrell Y, Bernardo D, Blanco G, Vazquez E, Sanchez JA (2008) Spatial and temporal variation of genetic diversity and estimation of effective population sizes in Atlantic salmon (Salmo salar, L.) populations from Asturias (Northern Spain) using microsatellites. *Conservation Genetics*, **9**, 807-819.

Kanda N, Goto M, Kato H, McPhee MV, Pastene LA (2007) Population genetic structure of Bryde's whales (Balaenoptera brydei) at the inter-oceanic and trans-equatorial levels. *Conservation Genetics*, **8**, 853-864.

Kang M, Wang J, Huang HW (2008) Demographic bottlenecks and low gene flow in remnant populations of the critically endangered Berchemiella wilsonii var. pubipetiolata (Rhamnaceae) inferred from microsatellite markers. *Conservation Genetics*, **9**, 191-199.

Keller D, Brodbeck S, Floss I, Vonwil G, Holderegger R (2010) Ecological and genetic measurements of dispersal in a threatened dragonfly. *Biological Conservation*, **143**, 2658-2663.

Knittweis L, Kraemer WE, Timm J, Kochzius M (2009) Genetic structure of Heliofungia actiniformis (Scleractinia: Fungiidae) populations in the Indo-Malay Archipelago: implications for live coral trade management efforts. *Conservation Genetics*, **10**, 241-249.

Lambeets K, Breyne P, Bonte D (2010) Spatial genetic variation of a riparian wolf spider Pardosa agricola (Thorell, 1856) on lowland river banks: The importance of functional connectivity in linear spatial systems. *Biological Conservation*, **143**, 660-668.

Lara-Ruiz P, Chiarello AG, Santos FR (2008) Extreme population divergence and conservation implications for the rare endangered Atlantic Forest sloth, Bradypus torquatus (Pilosa : Bradypodidae). *Biological Conservation*, **141**, 1332-1342.

Lehtonen PK, Tonteri A, Sendek D, Titov S, Primmer CR (2009) Spatio-temporal genetic structuring of brown trout (Salmo trutta L.) populations within the River Luga, northwest Russia. *Conservation Genetics*, **10**, 281-289.

Lesobre L, Lacroix F, Caizergues A, Hingrat Y, Chalah T, Saint Jalme M (2010) Conservation genetics of Houbara Bustard (Chlamydotis undulata undulata): population structure and its implications for the reinforcement of wild populations. *Conservation Genetics*, **11**, 1489-1497.

Lloyd MW, Burnett RK, Engelhardt KAM, Neel MC (2011) The structure of population genetic diversity in Vallisneria americana in the Chesapeake Bay: implications for restoration. *Conservation Genetics*, **12**, 1269-1285.

Lopez-Pujol J, Font J, Simon J, Blanche C (2007) Can the preservation of historical relicts permit the conservation of endangered plant species? The case of Silene sennenii (Caryophyllaceae). *Conservation Genetics*, **8**, 903-912.

Lopez A, Vera M, Otero-Ferrer F, Pardo BG, Martinez P, Molina L, Bouza C (2010) Species identification and genetic structure of threatened seahorses in Gran Canaria Island (Spain) using mitochondrial and microsatellite markers. *Conservation Genetics*, **11**, 2431-2436.

Luikart G, Amish SJ, Winnie J, Beja-Pereira A, Godinho R, Allendorf FW, Harris RB (2011) High connectivity among argali sheep from Afghanistan and adjacent countries: Inferences from neutral and candidate gene microsatellites. *Conservation Genetics*, **12**, 921-931.

Massa-Gallucci A, Coscia I, O'Grady M, Kelly-Quinn M, Mariani S (2010) Patterns of genetic structuring in a brown trout (Salmo trutta L.) metapopulation. *Conservation Genetics*, **11**, 1689-1699.

May-Itza WD, Quezada-Euan JJG, Medina LAM, Enriquez E, De la Rua P (2010) Morphometric and genetic differentiation in isolated populations of the endangered Mesoamerican stingless bee Melipona yucatanica (Hymenoptera: Apoidea) suggest the existence of a two species complex. *Conservation Genetics*, **11**, 2079-2084.

May S, Beebee TJC (2010) Recent introduction or ancient ancestry? Use of genetic evidence to investigate the origins of range edge populations in natterjack toads (Bufo calamita). *Conservation Genetics*, **11**, 293-300.

Measey GJ, Tolley KA (2011) Investigating the cause of the disjunct distribution of Amietophrynus pantherinus, the Endangered South African western leopard toad. *Conservation Genetics*, **12**, 61-70.

Mendez M, Rosenbatan HC, Bordino P (2008) Conservation genetics of the franciscana dolphin in Northern Argentina: population structure, by-catch impacts, and management implications. *Conservation Genetics*, **9**, 419-435.

Miyake T, Nakajima J, Onikura N, Ikemoto S, Iguchi K, Komaru A, Kawamura K (2011) The genetic status of two subspecies of Rhodeus atremius, an endangered bitterling in Japan. *Conservation Genetics*, **12**, 383-400.

Moreira RG, McCauley RA, Cortes-Palomec AC, Fernandes GW, Oyama K (2010) Spatial genetic structure of Coccoloba cereifera (Polygonaceae), a critically

endangered microendemic species of Brazilian rupestrian fields. *Conservation Genetics*, **11**, 1247-1255.

Morrison CL, Ross SW, Nizinski MS, Brooke S, Jarnegren J, Waller RG, Johnson RL, King TL (2011) Genetic discontinuity among regional populations of Lophelia pertusa in the North Atlantic Ocean. *Conservation Genetics*, **12**, 713-729.

Mylecraine KA, Bulgin NL, Gibbs HL, Vickery PD, Perkins DW (2008) Limited genetic structure and evidence for dispersal among populations of the endangered Florida grasshopper sparrow, Ammodramus savannarum floridanus. *Conservation Genetics*, **9**, 1633-1638.

Natoli A, Peddemors VM, Hoelzel AR (2008) Population structure of bottlenose dolphins (Tursiops aduncus) impacted by bycatch along the east coast of South Africa. *Conservation Genetics*, **9**, 627-636.

Neel MC (2008) Patch connectivity and genetic diversity conservation in the federally endangered and narrowly endemnoc plant species Astragalms albens (Fabaceae). *Biological Conservation*, **141**, 938-955.

Ottewell KM, Bickerton D, Lowe AJ (2011) Can a seed bank provide demographic and genetic rescue in a declining population of the endangered shrub Acacia pinguifolia? *Conservation Genetics*, **12**, 669-678.

Overeem RL, Peucker AJ, Austin CM, Dann P, Burridge CP (2008) Contrasting genetic structuring between colonies of the World's smallest penguin, Eudyptula minor (Aves : Spheniscidae). *Conservation Genetics*, **9**, 893-905.

Ozerov MY, Veselov AJ, Lumme J, Primmer CR (2010) Genetic structure of freshwater Atlantic salmon (Salmo salar L.) populations from the lakes Onega and Ladoga of northwest Russia and implications for conservation. *Conservation Genetics*, **11**, 1711-1724.

Pavlova A, Walker FM, van der Ree R, Cesarini S, Taylor AC (2010) Threatened populations of the Australian squirrel glider (Petaurus norfolcensis) show evidence of evolutionary distinctiveness on a Late Pleistocene timescale. *Conservation Genetics*, **11**, 2393-2407.

Perez T, Vazquez F, Naves J, Fernandez A, Corao A, Albornoz J, Dominguez A (2009) Non-invasive genetic study of the endangered Cantabrian brown bear (Ursus arctos). *Conservation Genetics*, **10**, 291-301.

Peters MB, Turner TF (2008) Genetic variation of the major histocompatibility complex (MHC class II beta gene) in the threatened Gila trout, Oncorhynchus gilae gilae. *Conservation Genetics*, **9**, 257-270.

Peters MD, Xiang QY, Thomas DT, Stucky J, Whiteman NK (2009) Genetic analyses of the federally endangered Echinacea laevigata using amplified fragment length polymorphisms (AFLP)-Inferences in population genetic structure and mating system. *Conservation Genetics*, **10**, 1-14.

Pimper LE, Baker CS, Goodall RNP, Olavarria C, Remis MI (2010) Mitochondrial DNA variation and population structure of Commerson's dolphins (Cephalorhynchus commersonii) in their southernmost distribution. *Conservation Genetics*, **11**, 2157-2168.

Pohjamo M, Korpelainen H, Kalinauskaite N (2008) Restricted gene flow in the clonal hepatic Trichocolea tomentella in fragmented landscapes. *Biological Conservation*, **141**, 1204-1217.

Pujolar JM, Bevacqua D, Capoccioni F, Ciccotti E, De Leo GA, Zane L (2011) No apparent genetic bottleneck in the demographically declining European eel using molecular genetics and forward-time simulations. *Conservation Genetics*, **12**, 813-825.

Rasmussen KK, Kollmann J (2008) Low genetic diversity in small peripheral populations of a rare European tree (Sorbus torminalis) dominated by clonal reproduction. *Conservation Genetics*, **9**, 1533-1539.

Rayner MJ, Carraher CJF, Clout MN, Hauber ME (2010) Mitochondrial DNA analysis reveals genetic structure in two New Zealand Cook's petrel (Pterodroma cookii) populations. *Conservation Genetics*, **11**, 2073-2077.

Reid SM, Wilson CC, Mandrak NE, Carl LM (2008) Population structure and genetic diversity of black redhorse (Moxostoma duquesnei) in a highly fragmented watershed. *Conservation Genetics*, **9**, 531-546.

Ronka A, Kvist L, Karvonen J, Koivula K, Pakanen VM, Schamel D, Tracy DM (2008) Population genetic structure in the Temminck's stint Calidris temminckii, with an emphasis on Fennoscandian populations. *Conservation Genetics*, **9**, 29-37.

Russello MA, Stahala C, Lalonde D, Schmidt KL, Amato G (2010) Cryptic diversity and conservation units in the Bahama parrot. *Conservation Genetics*, **11**, 1809-1821.

Saarman NP, Louie KD, Hamilton H (2010) Genetic differentiation across eastern Pacific oceanographic barriers in the threatened seahorse Hippocampus ingens. *Conservation Genetics*, **11**, 1989-2000.

Salinas M, Altet L, Clavel C, Almela RM, Bayon A, Burguete I, Sanchez A (2011) Genetic assessment, illegal trafficking and management of the Mediterranean spur-thighed tortoise in Southern Spain and Northern Africa. *Conservation Genetics*, **12**, 1-13.

Sanllorente O, Hammond RL, Ruano F, Keller L, Tinaut A (2010) Extreme population differentiation in a vulnerable slavemaking ant with a fragmented distribution. *Conservation Genetics*, **11**, 1701-1710.

Sastre N, Vila C, Salinas M, Bologov VV, Urios V, Sanchez A, Francino O, Ramirez O (2011) Signatures of demographic bottlenecks in European wolf populations. *Conservation Genetics*, **12**, 701-712.

Shapcott A, Dowe J, Ford H (2009) Low genetic diversity and recovery implications of the vulnerable Bankoual, Palm Livistona carinensis (Arecaceae), from North-eastern Africa and the Southern Arabian Peninsula. *Conservation Genetics*, **10**, 317-327.

Smith RJ, Waldren S (2010) Patterns of genetic variation in Colchicum autumnale L. and its conservation status in Ireland: a broader perspective on local plant conservation. *Conservation Genetics*, **11**, 1351-1361.

Snoj A, Razpet A, TomIjanovic T, Treer T, Susnik S (2007) Genetic composition of the Jadro softmouth trout following translocation into a new habitat. *Conservation Genetics*, **8**, 1213-1217.

Spradling TA, Tamplin JW, Dow SS, Meyer KJ (2010) Conservation genetics of a peripherally isolated population of the wood turtle (Glyptemys insculpta) in Iowa. *Conservation Genetics*, **11**, 1667-1677.

Su ZH, Zhang ML, Sanderson SC (2011) Chloroplast phylogeography of Helianthemum songaricum (Cistaceae) from northwestern China: implications for preservation of genetic diversity. *Conservation Genetics*, **12**, 1525-1537.

Takaki Y, Kawahara T, Kitamura H, Endo K, Kudo T (2009) Genetic diversity and genetic structure of Northern Goshawk (Accipiter gentilis) populations in eastern Japan and Central Asia. *Conservation Genetics*, **10**, 269-279.

Tarazi R, Mantovani A, dos Reis MS (2010) Fine-scale spatial genetic structure and allozymic diversity in natural populations of Ocotea catharinensis Mez. (Lauraceae). *Conservation Genetics*, **11**, 965-976.

Van Looy K, Jacquemyn H, Breyne P, Honnay O (2009) Effects of flood events on the genetic structure of riparian populations of the grassland plant Origanum vulgare. *Biological Conservation*, **142**, 870-878.

Van Rossum F (2008) Conservation of long-lived perennial forest herbs in an urban context: Primula elatior as study case. *Conservation Genetics*, **9**, 119-128.

Vander Mijnsbrugge K, De Cock K, Cox K, Breyne P (2010) Conservation measures for Rosa arvensis Huds. in Flanders (Belgium) based on congruent genetic and phenotypic population differentiation. *Conservation Genetics*, **11**, 2243-2253.

Velo-Anton G, Garcia-Paris M, Rivera AC (2008) Patterns of nuclear and mitochondrial DNA variation in Iberian populations of Emys orbicularis (Emydidae): conservation implications. *Conservation Genetics*, **9**, 1263-1274.

Viaud-Martinez KA, Brownell RL, Komnenou A, Bohonak AJ (2008) Genetic isolation and morphological divergence of Black Sea bottlenose dolphins. *Biological Conservation*, **141**, 1600-1611.

Vilas R, Bouza C, Castro J, Lopez A, Martinez P (2010) Management units of brown trout from Galicia (NW: Spain) based on spatial genetic structure analysis. *Conservation Genetics*, **11**, 897-906.

Warnock WG, Rasmussen JB, Taylor EB (2010) Genetic clustering methods reveal bull trout (Salvelinus confluentus) fine-scale population structure as a spatially nested hierarchy. *Conservation Genetics*, **11**, 1421-1433.

Zhan QQ, Wang JF, Gong X, Peng H (2011) Patterns of chloroplast DNA variation in Cycas debaoensis (Cycadaceae): conservation implications. *Conservation Genetics*, **12**, 959-970.

Appendix B

Data used for the analysis. $V_{A\ control}$ and $V_{A\ bottleneck}$ are the mean additive genetic variance values for control and experimental lines. LogR is the log of the ratio of the experimental to the control lines.

Paper	Species	Trait	Bottleneck Size	Inbreeding Coefficient	$V_{A\ control}$	$V_{A\ bottleneck}$	LogR
van Heerwaarden et al. 2008	*Drosophila bunnanda*	DesicRes	2	0.276	0.1393	0.3168	0.3567
van Heerwaarden et al. 2008	*Drosophila bunnanda*	SPBristle	2	0.276	0.1970	0.1397	-0.1494
Whitlock and Fowler 1999	*Drosophila melanogaster*	WingArea	2	0.32	0.0012	0.0007	-0.2262
Whitlock and Fowler 1999	*Drosophila melanogaster*	Wing Angle 5-7-4	2	0.32	0.0009	0.0006	-0.1592
Whitlock and Fowler 1999	*Drosophila melanogaster*	Wing Angle 8-7-6	2	0.32	0.0022	0.0016	-0.1459

Paper	Species	Trait	Bottleneck Size	Inbreeding Coefficient	$V_{A\ control}$	$V_{A\ bottleneck}$	LogR
Whitlock and Fowler 1999	*Drosophila melanogaster*	Wing Angle 2-9-3	2	0.32	0.0001	0.0001	-0.1481
Whitlock and Fowler 1999	*Drosophila melanogaster*	Wing Angle 2-1-5	2	0.32	0.0002	0.0001	-0.1805
Whitlock and Fowler 1999	*Drosophila melanogaster*	Wing Angle 2-3-5	2	0.32	0.0006	0.0004	-0.1806
Kristensen et al. 2005	*Drosophila melanogaster*	SPBristle	8	0.67	0.8020	0.3140	-0.4072
Kristensen et al. 2005	*Drosophila melanogaster*	SPBristle	2	0.67	0.8020	0.3790	-0.3255
Saccheri et al. 2001	*Bicyclus anynana*	WhiteArea	2	0.27	0.0004	0.0002	-0.3274
Saccheri et al. 2001	*Bicyclus anynana*	Contrast	2	0.27	4.2000	1.0000	-0.6232
Saccheri et al. 2001	*Bicyclus anynana*	BlackArea	2	0.27	0.1650	0.1150	-0.1568

Paper	Species	Trait	Bottleneck Size	Inbreeding Coefficient	$V_{A\ control}$	$V_{A\ bottleneck}$	LogR
Saccheri et al. 2001	*Bicyclus anynana*	Colour	2	0.27	37.0000	29.0000	-0.1058
Saccheri et al. 2001	*Bicyclus anynana*	GoldArea	2	0.27	0.1050	0.0500	-0.3222
Saccheri et al. 2001	*Bicyclus anynana*	WingArea	2	0.27	74.0000	25.0000	-0.4713
Saccheri et al. 2001	*Bicyclus anynana*	ExtraRing	2	0.27	14.0000	8.5000	-0.2167
Saccheri et al. 2001	*Bicyclus anynana*	BandIndex	2	0.27	0.0133	0.0100	-0.1222
Saccheri et al. 2001	*Bicyclus anynana*	WhiteArea	6	0.1	0.0004	0.0004	-0.0263
Saccheri et al. 2001	*Bicyclus anynana*	Contrast	6	0.1	4.2000	2.8000	-0.1761
Saccheri et al. 2001	*Bicyclus anynana*	BlackArea	6	0.1	0.1650	0.0750	-0.3424

Paper	Species	Trait	Bottleneck Size	Inbreeding Coefficient	$V_{A\ control}$	$V_{A\ bottleneck}$	LogR
Saccheri et al. 2001	*Bicyclus anynana*	Colour	6	0.1	37.0000	43.0000	0.0653
Saccheri et al. 2001	*Bicyclus anynana*	GoldArea	6	0.1	0.1050	0.0430	-0.3877
Saccheri et al. 2001	*Bicyclus anynana*	WingArea	6	0.1	74.0000	50.0000	-0.1703
Saccheri et al. 2001	*Bicyclus anynana*	ExtraRing	6	0.1	14.0000	14.5000	0.0152
Saccheri et al. 2001	*Bicyclus anynana*	BandIndex	6	0.1	0.0133	0.0075	-0.2472
Saccheri et al. 2001	*Bicyclus anynana*	WhiteArea	20	0.03	0.0004	0.0009	0.3010
Saccheri et al. 2001	*Bicyclus anynana*	Contrast	20	0.03	4.2000	2.7000	-0.1919
Saccheri et al. 2001	*Bicyclus anynana*	BlackArea	20	0.03	0.1650	0.1800	0.0378

Paper	Species	Trait	Bottleneck Size	Inbreeding Coefficient	$V_{A\ control}$	$V_{A\ bottleneck}$	LogR
Saccheri et al. 2001	*Bicyclus anynana*	Colour	20	0.03	37.0000	54.0000	0.1642
Saccheri et al. 2001	*Bicyclus anynana*	GoldArea	20	0.03	0.1050	0.1170	0.0470
Saccheri et al. 2001	*Bicyclus anynana*	WingArea	20	0.03	74.0000	85.0000	0.0602
Saccheri et al. 2001	*Bicyclus anynana*	ExtraRing	20	0.03	14.0000	19.5000	0.1439
Saccheri et al. 2001	*Bicyclus anynana*	BandIndex	20	0.03	0.0133	0.0113	-0.0711
Briggs and Goldman 2006	*Brassica rapa*	CotyledonSize	2	0.25	0.0210	0.0643	0.4862
Fernandez et al. 1995	*Tribolium castaneum*	Fecundity	2	0.25	20.8000	18.8000	-0.0439
Fernandez et al. 1995	*Tribolium castaneum*	EggPupaViability	2	0.25	2.9000	42.9000	1.1701

Paper	Species	Trait	Bottleneck Size	Inbreeding Coefficient	$V_{A\ control}$	$V_{A\ bottleneck}$	LogR
Fernandez et al. 1995	*Tribolium castaneum*	LateViability	2	0.25	13.0000	107.70	0.9183
Fernandez et al. 2003	*Drosophila melanogaster*	EarlyFemaleFecundity	2	0.25	30.0000	21.0714	-0.1534
Cheverud et al. 1999	*Mus musculus*	AdultBodyWt	4	0.28	8.5000	8.3950	-0.0054
Cheverud et al. 1999	*Mus musculus*	AdultBodyWt	4	0.29	8.5000	5.99	-0.1520
Cheverud et al. 1999	*Mus musculus*	AdultBodyWt	4	0.3	8.5000	0.7625	-1.0472
Cheverud et al. 1999	*Mus musculus*	AdultBodyWt	4	0.32	8.5000	22.2800	0.4185
Cheverud et al. 1999	*Mus musculus*	AdultBodyWt	4	0.34	8.5000	3.9967	-0.3277
Cheverud et al. 1999	*Mus musculus*	AdultBodyWt	4	0.35	8.5000	7.9600	-0.0285

Paper	Species	Trait	Bottleneck Size	Inbreeding Coefficient	$V_{A\ control}$	$V_{A\ bottleneck}$	LogR
Cheverud et al. 1999	*Mus musculus*	AdultBodyWt	4	0.36	8.5000	7.7900	-0.0379
Cheverud et al. 1999	*Mus musculus*	AdultBodyWt	4	0.37	8.5000	31.47	0.5685
Cheverud et al. 1999	*Mus musculus*	AdultBodyWt	4	0.38	8.5000	19.91	0.3697
Cheverud et al. 1999	*Mus musculus*	AdultBodyWt	4	0.39	8.5000	6.9450	-0.0877
Cheverud et al. 1999	*Mus musculus*	AdultBodyWt	4	0.41	8.5000	13.9100	0.2139
Cheverud et al. 1999	*Mus musculus*	AdultBodyWt	4	0.43	8.5000	28.1100	0.5194
Cheverud et al. 1999	*Mus musculus*	AdultBodyWt	4	0.45	8.5000	6.61	-0.1092
Cheverud et al. 1999	*Mus musculus*	AdultBodyWt	4	0.48	8.5000	0.1550	-1.7391

Paper	Species	Trait	Bottleneck Size	Inbreeding Coefficient	$V_{A\ control}$	$V_{A\ bottleneck}$	LogR
Cheverud et al. 1999	*Mus musculus*	AdultBodyWt	4	0.59	8.5000	14.9500	0.2452
Lopez-Fanjul and Villaverde 1989	*Drosophila melanogaster*	EggPupaViability	2	0.25	19.8438	98.0000	0.6936
Lopez-Fanjul et al. 1989	*Drosophila melanogaster*	BristleNumber56	2	0.5	5.1500	2.5000	-0.3139
Garcia et al. 1994	*Drosophila melanogaster*	EggPupaViability	2	0.25	82.6000	134.18	0.2107
Garcia et al. 1994	*Drosophila melanogaster*	EggPupaViability	2	0.5	82.6000	150.72	0.2612
Garcia et al. 1994	*Drosophila melanogaster*	EggPupaViability	2	0.73	82.6000	68.77	-0.0796
Bryant and Meffert 1995	*Musca domestica*	WLTW	4	0.125	0.0006	0.0006	-0.0220
Bryant and Meffert 1996	*Musca domestica*	WingLength	4	0.125	0.0007	0.0005	-0.1525

Paper	Species	Trait	Bottleneck Size	Inbreeding Coefficient	$V_{A\ control}$	$V_{A\ bottleneck}$	LogR
Bryant and Meffert 1996	*Musca domestica*	InnerEyeSep	4	0.125	0.0011	0.0006	-0.2839
Bryant and Meffert 1996	*Musca domestica*	ScutellumLengthMid	4	0.125	0.0001	0.0005	0.5406
Bryant and Meffert 1996	*Musca domestica*	MetatibiaLength	4	0.125	0.0001	0.0003	0.2816
Bryant and Meffert 1993	*Musca domestica*	composite	2	0.25	0.0006	0.0012	0.2825
Bryant and Meffert 1993	*Musca domestica*	composite	8	0.0625	0.0006	0.0018	0.4771
Bryant and Meffert 1993	*Musca domestica*	composite	32	0.015625	0.0006	0.0011	0.2632
Bryant et al.1986	*Musca domestica*	WingLength	2	0.25	0.0011	0.0006	-0.2808
Bryant et al.1986	*Musca domestica*	WingWidth	2	0.25	0.0005	0.0006	0.0792

Paper	Species	Trait	Bottleneck Size	Inbreeding Coefficient	$V_{A\ control}$	$V_{A\ bottleneck}$	LogR
Bryant et al.1986	*Musca domestica*	HeadWidth	2	0.25	0.0003	0.0004	0.1249
Bryant et al.1986	*Musca domestica*	InnerEyeSep	2	0.25	0.0003	0.0022	0.8653
Bryant et al.1986	*Musca domestica*	ScutellumLengthMid	2	0.25	0.0006	0.0012	0.3010
Bryant et al.1986	*Musca domestica*	ScutellumWidthBase	2	0.25	0.0005	0.0024	0.6812
Bryant et al.1986	*Musca domestica*	LengthThoracicSuture	2	0.25	0.0004	0.0004	0.0000
Bryant et al.1986	*Musca domestica*	LengthMetafemur	2	0.25	0.0006	0.0006	0.0000
Bryant et al.1986	*Musca domestica*	WingLength	8	0.0625	0.0011	0.0019	0.2576
Bryant et al.1986	*Musca domestica*	WingWidth	8	0.0625	0.0005	0.0013	0.4150

Paper	Species	Trait	Bottleneck Size	Inbreeding Coefficient	$V_{A\ control}$	$V_{A\ bottleneck}$	LogR
Bryant et al.1986	*Musca domestica*	HeadWidth	8	0.0625	0.0003	0.0014	0.6690
Bryant et al.1986	*Musca domestica*	InnerEyeSep	8	0.0625	0.0003	0.0021	0.8451
Bryant et al.1986	*Musca domestica*	ScutellumLengthMid	8	0.0625	0.0006	0.0018	0.4771
Bryant et al.1986	*Musca domestica*	ScutellumWidthBase	8	0.0625	0.0005	0.0024	0.6812
Bryant et al.1986	*Musca domestica*	LengthThoracicSuture	8	0.0625	0.0004	0.0019	0.6767
Bryant et al.1986	*Musca domestica*	LengthMetafemur	8	0.0625	0.0006	0.0014	0.3680
Bryant et al.1986	*Musca domestica*	WingLength	32	0.015625	0.0011	0.0014	0.1249
Bryant et al.1986	*Musca domestica*	WingWidth	32	0.015625	0.0005	0.0013	0.4150

Paper	Species	Trait	Bottleneck Size	Inbreeding Coefficient	$V_{A\ control}$	$V_{A\ bottleneck}$	LogR
Bryant et al.1986	*Musca domestica*	HeadWidth	32	0.015625	0.0003	0.0009	0.4771
Bryant et al.1986	*Musca domestica*	InnerEyeSep	32	0.015625	0.0003	0.0019	0.8016
Bryant et al.1986	*Musca domestica*	ScutellumLengthMid	32	0.015625	0.0006	0.0009	0.1761
Bryant et al.1986	*Musca domestica*	ScutellumWidthBase	32	0.015625	0.0005	0.0022	0.6435
Bryant et al.1986	*Musca domestica*	LengthThoracicSuture	32	0.015625	0.0004	0.0012	0.4771
Bryant et al.1986	*Musca domestica*	LengthMetafemur	32	0.015625	0.0006	0.0007	0.0669
Andersson et al 2010	*Nigella degenii*	Flowering Date	1	0.535	29.52	19.62	-0.1774
Andersson et al 2010	*Nigella degenii*	Flower Number	1	0.535	1.2300	0.8400	-0.1656

Paper	Species	Trait	Bottleneck Size	Inbreeding Coefficient	$V_{A\ control}$	$V_{A\ bottleneck}$	LogR
Andersson et al 2010	*Nigella degenii*	Plant height	1	0.535	965.8200	602.7600	-0.2048
Andersson et al 2010	*Nigella degenii*	Leaf length	1	0.535	1.9300	1.3800	-0.1457
Andersson et al 2010	*Nigella degenii*	Sepal length	1	0.535	0.4600	0.1800	-0.4075
Meffert 1995	*Musca domestica*	Buzz	4	0.125	0.3500	0.0167	-1.3222
Meffert 1995	*Musca domestica*	Creep	4	0.125	0.6750	0.1917	-0.5468
Meffert 1995	*Musca domestica*	Hold	4	0.125	0.1750	0.0300	-0.7659
Meffert 1995	*Musca domestica*	Lift	4	0.125	0.0050	0.1000	1.3010
Meffert 1995	*Musca domestica*	Buzz	4	0.125	0.3500	0.3750	0.0300

Paper	Species	Trait	Bottleneck Size	Inbreeding Coefficient	$V_{A\ control}$	$V_{A\ bottleneck}$	LogR
Meffert 1995	*Musca domestica*	Close	4	0.125	0.1100	0.0467	-0.3724
Meffert 1995	*Musca domestica*	Hold	4	0.125	0.0550	0.0633	0.0613
Meffert 1995	*Musca domestica*	Lift	4	0.125	0.0600	0.0150	-0.6021
Meffert 1995	*Musca domestica*	Lunge	4	0.125	0.1500	0.0650	-0.3632
Meffert 1995	*Musca domestica*	Touch	4	0.125	0.4000	0.6333	0.1996
Meffert 1995	*Musca domestica*	Wingout	4	0.125	1.1500	0.8417	-0.1356

Appendix C

Classification of threatened and non-threatened species by regression tree analysis, logistic regression, and discriminant analysis for the butterfly data when abundance is included in the analysis.[a]

	Threatened		Non-threatened	
	Predicted	P^b	Predicted	p
Regression tree	14/18	<0.0001	60/62	<0.0001
Logistic regression - 0.5 cutoff	13/18	<0.0001	61/62	<0.0001
Stepwise logistic regression	9/18	0.0027	61/62	<0.0001
Discriminant function analysis - spherical	13/18	<0.0001	58/62	<0.0001

a: When multiple analyses were performed, such as when different structures were used for discriminant analysis, only the analysis with the best result is given.

b: Probability of correctly predicting by chance alone at least as many as observed by a given method.

Appendix D

R code for assessing the chance that a given number of correctly assigned species or more could occur randomly to determine whether the analyses presented here are able to correctly classify threatened species more accurately than random assignment.

```
set.seed(20)
Max.rep <- 10000
N <- matrix(0,Max.rep,3)
for ( i in 1:Max.rep)
{
    X           <- matrix(0,80)   # First put zeros (= non-threatened)
    X[1:17]     <- 1              # Set Threatened =1 Assuming that 17 rows are threatened
    Y           <- sample(X)      # Pick random sample of length X without replacement
    Z.total     <- Y
    Z.threatened <- Z.total[1:17]  # 1 to 17 are those threatened 1s are correct zeros are not
    Z.not       <- Z.total[18:80]  # 18 to 80 are those not threated 0s are correct 1s are not
    N[i,2:3]    <- c(length(Z.threatened[Z.threatened==1]), length(Z.not[Z.not==0]))
# Number correct
    N[i,1] <- sum(N[i,2:3])              # The sum of correct assignments
}
print(N)

# Now find the number in each category
N.total     <- unique(N[,1])      # Find total number correctly assigned
n           <- length(N.total)
P.total     <- matrix(0,n,2)
P.total[,1] <- N.total
for ( i in 1:n){    P.total[i,2] <-length( N[N[,1]==N.total[i],1])/Max.rep}
P.total
sum(P.total[,2])

# Threatened correctly assigned
N.threatened     <- unique(N[,2])  # Distribution of correctly assigned threatened
n                <- length(N.threatened)
P.threatened     <- matrix(0,n,2)
P.threatened[,1] <- N.threatened
for ( i in 1:n){    P.threatened[i,2] <-length( N[N[,2]==N.threatened[i],1])/Max.rep}
P.threatened
sum(P.threatened[,2])

# Not threatened correctly assigned
N.not       <- unique(N[,3])
n           <- length(N.not)
P.not       <- matrix(0,n,2)
```

```
P.not[,1] <- N.not
for ( i in 1:n){       P.not[i,2] <-length( N[N[,3]==N.not[i],1])/Max.rep}
P.not
sum(P.not[,2])

# What is the probabilty of correctly assigning at least N.threatened correctly?
# Suppose observed N.threatened = 11 Have to pick all cases greater than 11
# But suppose that the largest in output is 10 (as it is here).  This means that probability is less than 1 in
      10,000
# Suppose the observed values is 5.  Then have to pick all values greater than 5
 N.threatened <- 5
 P.more.than.obs <- P.threatened[P.threatened[,1]>=N.threatened,]
 P <- sum(P.more.than.obs[,2])
```