Evolutionary Ecology Research, 2014, **16**: 91–99

Are recommendations for conservation informed by estimates of genetic divergence?

Heather R. Taft and Derek A. Roff

Department of Biology, University of California, Riverside, California, USA

ABSTRACT

Question: In studies on genetic divergence, does the divergence value co-vary with the recommended actions given in the paper's conclusion?

Data incorporated: Articles from the journals *Biological Conservation* and *Conservation Genetics* published between 2007 and 2011 were reviewed for studies assessing genetic divergence that gave conservation recommendations for wild populations.

Analytical methods: We used a general linear model with recommendation as the response variable to test this question. We then used a step procedure to eliminate variables that did not contribute significantly to the AIC value. The final model was compared with the constant-only model to test for significance.

Conclusion: The majority of recommendations given in studies that include an analysis of genetic variation are informed by the divergence value estimated. However, the type of marker used and sample size, used here as a surrogate for estimation error, are significant factors influencing the recommendation. There is a high correlation between divergence values of nuclear and non-nuclear markers but recommendations appear to depend primarily on the analysis of nuclear markers.

Keywords: conservation genetics, F_{ST}, genetic divergence, management recommendation, neutral markers, population genetics.

INTRODUCTION

Genetic diversity is now commonly estimated in populations of conservation concern. Genetic studies are important because fitness is often correlated with genetic diversity (Reed and 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 significant effects on small populations, and may be a direct cause of extinction (Frankham, 2003).

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

Correspondence: D.A. Roff, Department of Biology, University of California, Riverside, CA 92521, USA. e-mail: derek.roff@ucr.edu

whole. G_{ST}, an alternative way to calculate genetic divergence, was created by Nei (1973) to account for multiple alleles and is essentially the same as F_{ST}. The θ_{ST} statistic (Weir and 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 the analysis of variation in 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 data assuming that such loci 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 more recent and aim to eradicate problems associated with assessing only differences in heterozygosity in populations and not incorporating specific alleles in the equation. All these statistics are now widely used for assessing genetic divergence among populations (for further discussion, see Meirmans and Hedrick, 2011).

The numerous different molecular markers and divergence metrics may create a problem because, even though all metrics provide an index of divergence among populations, they are not mathematically identical and hence will not give the same value for the same data (Meirmans and Hedrick, 2011). Furthermore, differences among estimates will also occur owing to the particular type of marker used. For example, mitochondrial DNA (mtDNA) and chloroplast DNA (cpDNA) evolve at a faster rate than nuclear DNA, and hence we would expect divergence values estimated from the former types of marker to be generally higher than those estimated from the latter. Knowing the genetic structure within a population is useful for the creation of a conservation plan, but it is not clear if differences in divergence value due to the particular metric or molecular marker used will generate differences in conservation recommendations, or indeed, if recommendations are influenced in any way by the estimated divergence value. In this paper, we present an analysis comparing assessments of genetic divergence with the recommendations made for conservation at the conclusion of a genetic study. Our hypothesis is that the recommended action co-varies with the divergence value.

METHODS

Articles from the journals *Biological Conservation* and *Conservation Genetics* published between 2007 and 2011 were reviewed for studies assessing genetic divergence that gave conservation recommendations for wild populations (we used no captive or domestic populations; see evolutionary-ecology.com/data/2893Appendix.pdf). To avoid the problem of pseudo-replication, we used only studies focused on a single species. For the same reason, we used only studies reporting overall statistics, not pair-wise values, but only when divergence values for all types of genetic markers used in the study were provided. Because recommendations based on genetic divergence would be opposite to those for pure bred species (i.e. reduce connectivity to prevent further hybridization), studies on hybridization between wild and farm-raised species, or invasive species, were excluded. F_{ST}–Q_{ST} studies were also excluded, because our focus was 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, we used only the most recent data.

Altogether, 105 studies entered the analyses. The dataset included plants (29%) and animals (71%). The plant data consisted primarily of terrestrial perennial angiosperms,

with only one aquatic species, two gymnosperms, and a single bryophyte included in the dataset. Both vertebrates (84%) and invertebrates (16%) were represented in the animal taxa, as well as terrestrial (60%) and aquatic (40%) species.

We collected the following data 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_{ST}, etc.), the divergence value (the value of F_{ST}, G_{ST}, etc.), sample size, and the recommendations resulting from the study. We classified the latter 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 – studies that find low genetic divergence may not suggest a change in management, whereas studies that find large genetic divergence may suggest major management changes to maintain or decrease genetic divergence.
5. Translocation between populations – translocation is used to increase genetic diversity in populations by re-introducing alleles into a population lacking diversity.
6. Conserve diversity of the populations.
7. Protect populations – whether or not genetic divergence was found, managers might feel the populations require protection to ensure their survival.

Each recommendation above was used as a different categorical variable. If the recommendation was given in a paper it was scored as 1, otherwise zero. In total, seven divergence metrics were used together with nine types of molecular marker (Table 1). As noted in the Introduction, the divergence value will be a function of both the divergence metric and the type of molecular marker. The confidence in the divergence estimate may also be expected to influence any recommendations. Ideally, one would use the standard error (S.E.) as a measure of the reliability of the estimate. However, this was not given in all papers, so we used the sample size as a surrogate measure.

The statistical hypothesis of interest is that the recommended action co-varies with the divergence value. For any given recommendation, assuming no confounding variables, this hypothesis can be tested using a general linear model with recommendation as the response variable, divergence value as the model variable, and a binomial error term. We also considered the type of marker and sample size as potential additional model variables. After fitting this model, we used the step procedure in SPLUS to eliminate variables that did not contribute significantly to the AIC value. The final model was then compared with the constant-only model to test for significance.

Forty-five percent of the papers used more than a single marker/divergence metric combination: 33% ($n = 35$) reported two combinations, 8% ($n = 9$) reported three combinations, 3% ($n = 3$) reported four combinations, and 1% ($n = 1$) reported seven combinations. To account for possible pseudo-replication, we repeated our analyses using the mean divergence value with each marker category from each study.

Table 1. Distribution of divergence metrics and markers used in the present analysis

	Divergence metrics									Row statistics			
Marker	F_{ST}	F_{CT}	F_{SC}	D_{ST}	G'_{ST}	G_{ST}	Φ_{ST}	R_{ST}	θ_{ST}	Totals	Percent	Mean D	S.E.
Non-nuclear													
cpDNA	2	0	0	0	0	2	1	0	0	5	2.87	0.827	0.072
mtDNA	9	0	0	0	0	0	24	0	4	37	21.26	0.470	0.061
Nuclear													
ISSR	1	0	0	0	0	0	0	0	0	1	0.57	0.404	
AFLP	6	0	0	0	0	0	8	0	0	14	8.05	0.282	0.063
MHC	1	0	0	0	0	1	0	0	0	2	1.15	0.254	0.040
SNP	0	0	0	0	0	0	3	0	3	6	3.45	0.191	0.070
Allozymes	4	0	0	0	1	3	1	0	4	13	7.47	0.187	0.064
Microsatellites	14	2	1	1	1	3	5	15	53	95	54.60	0.166	0.018
RFLP	0	0	0	0	0	0	0	0	1	1	0.57	0.102	
Totals	37	2	1	1	2	9	42	15	65	174			
Percent	21.26	1.15	0.57	0.57	1.15	5.17	24.14	8.62	37.36				

Note: Statistical analysis showed no difference due to divergence metric but a significant difference among markers. Divergence values (D) grouped by marker are shown on the right with the markers ranked from high to low divergence values.

RESULTS

To maintain reasonable sample sizes, we divided markers into nuclear and non-nuclear categories. Studies using nuclear markers were generally more likely to make a management recommendation than those using non-nuclear markers (Table 2). As expected from the higher rates of evolution of mitochondrial and chloroplast DNA, the mean divergence value of non-nuclear markers (mean = 0.47, S.E. = 0.06, $n = 37$) was significantly greater than that of nuclear markers (mean = 0.21, S.E. = 0.02, $n = 137$; $F_{1,172} = 28.31$, $P < 0.0001$). Furthermore, there was no significant variation among markers within each category (non-nuclear: $F_{2,34} = 1.47$, $P = 0.24$; nuclear: $F_{8,128} = 1.60$, $P = 0.13$). This difference has a visually distinct impact on the relationship between the divergence value and the recommendations of the studies (see Fig. 1; for clarity of presentation, we have reversed the axes such that the response variable, 'Yes' or 'No', is on the x-axis).

Table 2. Percentage of studies that provide recommendations following the use of non-nuclear and/or nuclear markers

	%Yes	
Recommendation[a]	Non-nuclear	Nuclear
1. Habitat	20.0	36.0
2. Connectivity	13.3	28.1
3. Genetic banking	0.00	20.2
4. Management	80.0	47.2
5. Translocation	6.7	20.2
6. Diversity	36.7	39.3
7. Protect	16.7	41.6

[a] Description of recommendations: (1) Habitat – 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.

Table 3. Summary of best fitting models, where recommendation was the response variable and possible model variables were divergence value (D), marker category (M), and sample size (N)

Recommendation[a]	Model 1[b]	*P*	Model 2[c]	*P*
1. Habitat	Full	0.0392	Full	0.0488
2. Connectivity	M	0.0162	M	0.0934
3. Genetic banking	Additive	< 0.0001	Additive	0.0007
4. Management	D + M + D*M	0.0018	M	0.0018
5. Translocation	D + M + D*M	0.0067	Full	0.0681
6. Diversity	Full	0.0037	Full	0.0191
7. Protect	Full	0.0063	M	0.0119

[a] See Table 2 for description of recommendations.
[b] Best fitting model using all data where 'Full' = D + M + D*M + N and 'Additive' = D + M + N.
[c] Best fitting model using means per study within each marker category.

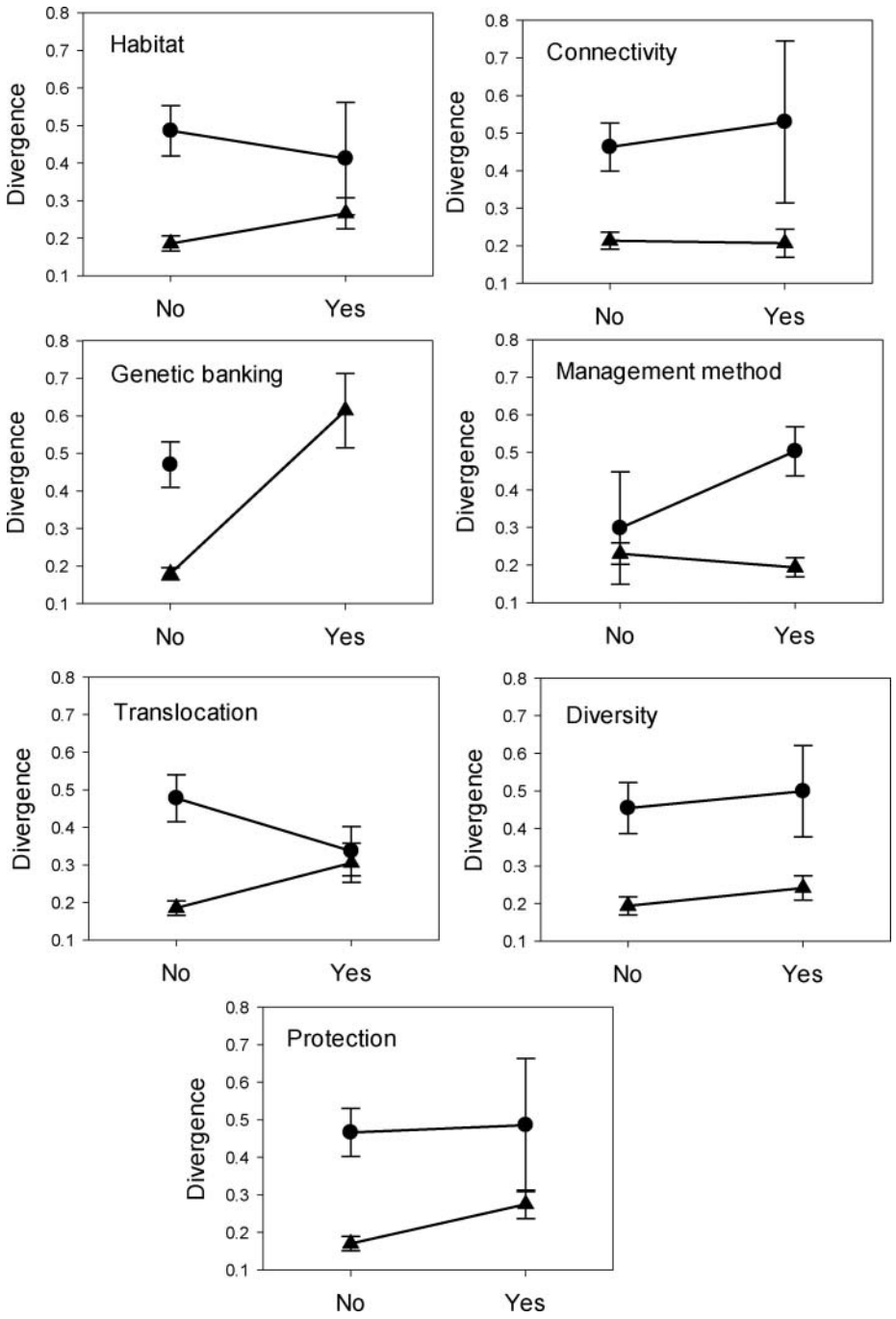

Fig. 1. Mean divergence values plotted as functions of the type of molecular marker (● = non nuclear, ▲ = nuclear, ± 1 standard error) and whether the specified recommendation was given (No = not given, Yes = recommended in paper).

Table 4. Summary of best fitting models using only nuclear markers

Recommendation[a]	Model 1[b]	*P*	Model 2[c]	*P*
1. Habitat	D + N	0.0363	D + N	0.1223
2. Connectivity	NS	NS	NS	NS
3. Genetic banking	D	< 0.0001	D	0.0013
4. Management	NS	NS	NS	NS
5. Translocation	D	0.0155	D	0.1634
6. Diversity	D + N	0.0037	N	0.0329
7. Protect	D	0.0094	D	0.1510

[a] See Table 2 for description of recommendations.
[b] Best fitting model using all data.
[c] Best fitting model using means per study within each marker category. If the best fitting model was not significant, we show the results for the model with the variables retained in the analysis using all data for comparison.

Owing to the apparent interaction in some of the plots between marker category and divergence value, we used the starting model

$$\text{Recommendation (Yes = 1, No = 0)} = D + M + D*M + N,$$

where D is the divergence value, M is the marker category, and N is the sample size (for clarity of presentation, the coefficients are omitted from the model). Because sample size, N, is a surrogate for the estimation error of the divergence value, we expect that N will positively co-vary with recommendation. In general, the results using the entire dataset were congruent with those using the study means within marker category (Table 3). Recommendations co-varied with divergence values in six of the seven categories when the full data set was used and in four categories when means were used. It is clear from the plotted data that recommendations are more likely to be influenced by analyses based on nuclear markers than non-nuclear markers (Fig. 1). To better appreciate the effect of marker category, we repeated the analysis separately for each of the two marker categories. None of the recommendations co-varied with divergence value obtained using non-nuclear markers. In only one case, habitat, did the recommendation using non-nuclear markers depend on sample size ($F_{35,36} = 11.23$, $P < 0.0001$) and in this case, as predicted, a recommendation was more likely with an increase in sample size. Recommendations based on the analysis of nuclear markers typically co-varied with divergence value and/or sample size, with an increase in divergence value and/or sample size generating a recommendation (Fig. 1, Table 4). Sixteen of the studies used both nuclear and non-nuclear markers: there is a highly significant correlation between the mean divergence values from these studies ($r = 0.862$, $F_{1,14} < 0.0001$) but it is not clear to what extent recommendations were based on one or both types of analysis.

DISCUSSION

There is no generally accepted 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 analysed the recommendations given at the conclusion of genetic analyses to determine whether the results of genetic studies are actually

related to the recommendations. Our analysis indicates that the majority of recommendations given in studies that include a genetic analysis of population variation using nuclear markers are informed by the divergence value estimated. Studies using non-nuclear markers showed inconsistent results.

Eighty-five of the 105 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 studies with a large sample size may show a significant genetic divergence among populations for very small divergence values (Caujape-Castells, 2010). In this regard, it is noteworthy that sample size, a measure of the error of estimation in the divergence value, often contributed to the recommendation. The problem with small divergence values is that they may simply be a consequence 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. Unfortunately, autocorrelation is a problem that even spatial structure programs, such as STRUCTURE, cannot overcome (Schwartz and McKelvey, 2009).

The issue of inconsistency in defining populations with small divergence values as divergent or not raises questions about the biological justification of the recommendations to ensure the continued exchange of individuals between populations and prevent the erosion of genetic diversity. As pointed out by Latta (2008), finding genetic divergence could mean there is local adaptation, but not finding genetic divergence does not mean there is no local adaptation because neutral genetic markers may not be affected during adaptation. Thus recommendations for management could be the same regardless of whether the populations were divergent.

Overall, our analysis shows that genetic studies using nuclear markers do contribute to management recommendations, though the present analysis cannot determine if those recommendations are appropriate or not. However, as the specific genetic markers influence recommendations, the use of more genetic markers in a study may help paint a better picture of the genetic structure within the subpopulations and aid the ability to manage populations properly. Moritz (1994) suggested using both mtDNA and nuclear genetic markers, such as microsatellites, to determine whether different management units should be utilized. Given that there is a high correlation between divergence values in studies that used the two types of markers, our analysis suggests that recommendations are most likely driven by the analysis of nuclear markers.

ACKNOWLEDGEMENTS

We thank Norman Ellstrand for helpful comments on the manuscript.

REFERENCES

Bakker, J., van Rijswijk, M.E.C., Weissing, F.J. and Bijlsma, R. 2010. Consequences of fragmentation for the ability to adapt to novel environments in experimental *Drosophila* metapopulations. *Conserv. Genet.*, **11**: 435–448.

Caujape-Castells, J. 2010. General G (ST) and θ inflation due to biased intra-population sampling, and its consequences for the conservation of the Canarian flora. *Conserv. Genet.*, **11**: 709–720.

Excoffier, L., Smouse, P.E. and Quattro, J.M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes – application to human mitochondrial DNA restriction data. *Genetics*, **131**: 479–491.

Frankham, R. 2003. Genetics and conservation biology. *Comptes Rendus Biol.*, **326**: S22–S29.

Hedrick, P.W. 2005. A standardized genetic differentiation measure. *Evolution*, **59**: 1633–1638.

Jost, L. 2008. G_{ST} and its relatives do not measure differentiation. *Mol. Ecol.*, **17**: 4015–4026.

Latta, R.G. 2008. Conservation genetics as applied evolution: from genetic pattern to evolutionary process. *Evol. Appl.*, **1**: 84–94.

Meirmans, P.G. and Hedrick, P.W. 2011. Assessing population structure: F_{ST} and related measures. *Mol. Ecol. Resources*, **11**: 5–18.

Moritz, C. 1994. Applications of mitochondrial DNA analysis in conservation – a critical review. *Mol. Ecol.*, **3**: 401–411.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA*, **70**: 3321–3323.

Reed, D.H. and Frankham, R. 2003. Correlation between fitness and genetic diversity. *Conserv. Biol.*, **17**: 230–237.

Schwartz, M.K. and McKelvey, K.S. 2009. Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conserv. Genet.*, **10**: 441–452.

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

Weir, B.S. and Cockerham, C.C. 1984. Estimating *F*-statistics for the analysis of population-structure. *Evolution*, **38**: 1358–1370.

Wright, S. 1951. The genetical structure of populations. *Ann. Eugenics*, **15**: 323–354.