Conservation genetics: where are we now?

Philip W. Hedrick

Genetic studies in endangered species have become widespread in the past decade, and with new information from various genome projects, new applications and insights are forthcoming. Generally, neutral variants are used for conservation applications, and when combined with highly variable loci and/or many more markers, these approaches should become much more informative. Conservation genetics is also concerned with detrimental and adaptive variation, which are more difficult to identify and characterize; however, the ability to predict the extent of such variation might become more successful and applied in future conservation efforts. Neutral variants might be used to identify adaptive variants, but the overlay of different mutational processes and selective regimes suggests that extreme caution should be used in making such identifications.

The recent extinction of many species, and continuing threats to many more has made conservation biology crucial in the 21st century. Although ecological, political, economic, and other forces might be primary concern for avoiding the extinction of most endangered species, for long-term persistence, genetics and related considerations has also been a focus of conservation effort. In particular, the application of new molecular techniques has made examination of genetics in endangered species feasible and genetic analysis has become widely used in conservation research.

Here, as an organizational theme for examining the status and possible future of conservation genetics, I consider the three major types of genetic variation — neutral, detrimental and adaptive (see Glossary) — and discuss how each is used in the conservation of endangered species. Of course, whether a particular variant is neutral, detrimental or adaptive depends on the environment, population size, genetic background and so on. For example, a particular allele that is adaptive, such as providing resistance to an infectious disease in one environment, could be detrimental when the pathogen is absent because of a pleiotropic cost associated with that allele. Alternatively, genetic variants that are neutral in one situation could be adaptive in another.

I also consider one of the great promises of neutral molecular variation, that of the use of the extent and pattern of neutral variation to predict the amount and significance of detrimental and adaptive variation. Although it has not proven compelling in some cases, with more informative loci and more genes, we might be able to utilize such observed associations better in the future.
It is often presumed that statistical significance between groups for neutral molecular markers indicates the presence of biologically important differences, or that populations have been separated long enough for biologically important differences to accumulate. Traditional molecular markers often provided inadequate statistical power to estimate the differences between groups in endangered species because they displayed too little variation. However, with the discovery of highly variable loci, such as microsatellites, or large numbers of independent markers, such as single nucleotide polymorphisms, the statistical power to differentiate between groups is now often very high.

Let us briefly examine the association of statistical significance based on molecular markers and biological meaningfulness of comparisons of groups in two basic forms: there might be no statistical significance when there are actual biologically meaningful differences between groups, or there can be a significant statistical difference between groups reflecting a meaningful biological difference (Table I). In both cases, statistics based on molecular markers result in an appropriate evaluation of the real biological situation.

Table I. Classification of the four general states of association of statistical significance of molecular markers and the biological importance of the difference between groups, with some possible examples

<table>
<thead>
<tr>
<th>Statistical significance</th>
<th>Biological importance</th>
<th>Possible examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>No</td>
<td>Recent divergence of groups</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>Long-term separation of groups</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>High gene flow but strong selective differences</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>High statistical power for molecular markers</td>
</tr>
</tbody>
</table>

However, problems result when statistical significance does not reflect biological meaningfulness, a conflict that can occur in two basic forms: there might be no statistical significance when there are actual biologically meaningful differences between groups, or there can be statistic significance between groups when there is no meaningful biological difference. In the first instance, there might be no significant difference based on molecular genetic markers but other, adaptively important loci, might be highly differentiated between populations. For example, in Scots pine Pinus sylvestris from Finland, several different molecular markers show little differentiation between northern and southern populations. However, many important adaptive quantitative traits show high levels of genetic differentiation between these populations in common experimental environments. In this case, the molecular data appear to be adequately reflecting the high level of gene flow in Scots pine. However, the selective forces between populations are so strong that they overcome the effects of gene flow and result in large adaptive genetic differences between populations. In this case, the error is not a typical ‘false negative’, because the result is correct for the neutral nuclear markers. The error results from not directly assaying, or being able to assay, the genes involved in adaptation.

The final classification could become a major concern in both evolutionary and conservation biology as large numbers of highly variable markers become available in many species. In general, statistical power for determining differentiation between groups is closely related to the number of independent alleles (S. Kalinowski, pers. commun.), so that, even for a few highly variable microsatellite loci, there can be extremely high statistical power. When there is such high statistical power, extremely small molecular genetic differences between groups become statistically significant. This is not a typical ‘false positive’ because the differences detected are real, but so small that they do not reflect biologically meaningful differences.

To determine a biologically meaningful difference, we need to define some measure or effect related to the likelihood of the accumulation of significant biological differences. One potential way to examine the relationship between biological and statistical significance is to evaluate the statistical power to detect a known biological effect. For example, the statistical power to detect a one-generation genetic bottleneck of different sizes can be compared to the ancestral population for different numbers of loci. In this case, assuming that each of ten loci had five alleles and the sample size is 40, the statistical power to detect a bottleneck of size 32 is >0.8. It is unlikely that a bottleneck of that size for one generation would have any biological effects, but there would be high statistical power to detect it with highly polymorphic loci. In general, we need to quantify the extent of the evolutionary effect that we are able to detect with highly variable molecular markers and evaluate whether this effect is likely to have important biological consequences.

References


e

Neutral variation

The extent and pattern of molecular variation within a population is generally consistent with neutral variation; that is, a balance predicted by a reduction in variation from genetic drift and an increase in variation from mutation. When the population is small, even if selection is acting on the variation at a given gene, genetic drift could have a greater effect than would selection on allele frequencies. In general, neutrality of genetic variants can be assumed when the selection coefficient $s$ (either the selective disadvantage of a detrimental allele or the advantage of an adaptive allele) is $<1/(2N_e)$, where $N_e$ is the effective population size. Therefore, because there are generally low effective population sizes in endangered species, genetic variants are more likely to be effectively neutral in endangered than in common species (the connection between single gene variation and quantitative genetic traits is discussed in Ref. 3).

http://tree.trends.com
Fig. 1. The relationship between heterozygosity from 29 microsatellite loci and the inbreeding coefficient from pedigree information for 29 individual gray wolves Canis lupus. The solid line represents the observed regression; the broken line represents the expected relationship assuming that the inbreeding coefficient of noninbred individuals is 0.75. $R^2 = 0.521$. Reproduced, with permission, from Ref. 20.

Most recent conservation genetics research has focused on the use of neutral molecular markers. These applications have been primarily used to identify species, evolutionarily significant units (ESUs), management units (MUs) and the origin of individuals\(^6\)\(^-\)\(^7\). Highly variable genetic markers, such as microsatellite loci, have allowed the quantification of patterns that are not apparent when using genetic markers with less variation. The use of large numbers of single-nucleotide polymorphisms (SNPs) might also provide high genetic resolution. However, the application of highly variable loci (or extremely large numbers of markers) must be used with some caution\(^8\) because statistically significant differences might not reflect biologically important differences (Box 1), or might give a different signal compared with those from other markers\(^9\).

Molecular genetic markers hold great promise for several other conservation applications, including approaches to measuring fundamental parameters important in conservation, such as effective population size\(^10\)\(^,\)\(^11\), past bottlenecks\(^12\)\(^,\)\(^13\), and sex-specific gene flow\(^14\) or founder contributions\(^15\). Molecular markers can also be used to infer the historical and geographical relationships between groups\(^16\)\(^,\)\(^17\). Although the power to infer such relationships is substantial, data from ancient specimens can now provide additional insight into the relationship of contemporary groups\(^18\).

In addition to identifying individuals within a particular group, highly variable loci or large numbers of markers might allow identification of the relationships between individuals or their individual levels of inbreeding\(^19\). For example, to examine the predictive value of microsatellite variants, the heterozygosity at 29 microsatellite loci in a captive population of gray wolves Canis lupus was compared with that from the known pedigree (Fig. 1)\(^20\). The predicted level of heterozygosity from the pedigree data and the observed heterozygosity from the microsatellites were almost identical.

**Detrimental variation**

Perhaps the most important early contribution of genetics to conservation was the recognition of the importance of inbreeding depression, which is thought to be due to increasing the homozygosity of detrimental alleles\(^21\). In addition to the expected decline in fitness with inbreeding, the mean population fitness can also decline over time, because detrimental mutations with a small selective disadvantage in a small population will become fixed, much as if they were neutral\(^22\). It is useful to distinguish between these effects\(^23\) and to define the genetic load as the reduction in mean population fitness from detrimental variation compared with that of a population without lowered fitness.

Inbreeding depression and genetic load have been of major concern when dealing with endangered species\(^24\) and inbreeding avoidance has become a priority in captive breeding programs.

**Levels of inbreeding depression and genetic load**

In a large population at equilibrium, substantial standing detrimental genetic variation is expected, as is a large reduction of fitness if inbreeding was to occur (Table 1). There is also a genetic load: because of the efficacy of selection in large populations, most of the detrimental variants are recessive and occur at a low frequency. If the population declines in size, purging of detrimental variation should take place, especially for alleles of large detrimental effect, thereby reducing inbreeding depression. Some detrimental variants might become fixed, particularly those of smaller effect, thus causing an increase in genetic load\(^22\). If the population remains small for an extended period, more detrimental variation could be purged, further reducing inbreeding depression, but more detrimental variants could be fixed, resulting in a higher genetic load. Such a population might show no lowered fitness upon inbreeding but, because of the fixation of detrimental loci, all individuals in the population might have a low fitness and the population might have a high genetic load.

**Table 1. Scenarios that produce generally different predicted levels of inbreeding depression and genetic load**

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Inbreeding depression</th>
<th>Genetic load</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large population at equilibrium</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Recently small population</td>
<td>Intermediate (purged some detrimental allele)</td>
<td>Intermediate (fixed some detrimental alleles of small effect)</td>
</tr>
<tr>
<td>Long-term small population</td>
<td>Low (purged detrimental allele of medium and large effect)</td>
<td>High (fixed many detrimental alleles of small effect)</td>
</tr>
</tbody>
</table>
Populations of some endangered species have become so small that they have lost genetic variation and appear to possess a high frequency of deleterious genetic variants, even to the level of fixation. To avoid extinction from this genetic deterioration, some populations might benefit from the introduction of individuals from related populations or even subspecies for genetic restoration, that is, elimination of deleterious variants and recovery to normal levels of genetic variation. An extreme example is the last remaining population of the Florida panther Puma concolor coryi, which has a suite of traits that suggest genetic drift has fixed (or nearly fixed) the population for previously rare and potentially deleterious traits. These traits, which are found in high frequency only in the Florida panther and are unusual in other puma subspecies, include high frequencies of cryptochordism (unilateral undescended testicles), kinked tail for the last five vertebra (Fig. 1; reproduced, with permission, from Ref. a), cowlick on the back, and the poorest semen quality recorded in any felid b. In addition, a large survey of microsatellite loci has shown that Florida panthers have much lower molecular variation than do other North American populations of pumas c.

Table I. The proportion of Florida panthers with a kinked tail and the distinctive cowlick before and after the introduction of Texas cougars d

<table>
<thead>
<tr>
<th></th>
<th>Before 1990</th>
<th>F1 and F2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion with kinked tail</td>
<td>0.88 (48)</td>
<td>0.00 (15)</td>
</tr>
<tr>
<td>Proportion with cowlick</td>
<td>0.93 (46)</td>
<td>0.14 (14)</td>
</tr>
</tbody>
</table>

The potential positive and negative genetic effects of introducing individuals from genetically diverse but geographically isolated populations into apparently inbred populations were theoretically evaluated before the introduction of Texas cougars P. c. stanleyana (recently, it has been proposed that all the North American subspecies be synonymized e).

Two caveats should be mentioned. First, during this process, some populations (or even species) might become extinct and the ones going extinct could be those with a higher genetic load. Consequently, the remaining populations might not have as high a genetic load as would be expected from the standing amount of detrimental genetic variation. Second, genetic load might be documented as a low estimate of fitness compared with other populations, or by comparing the fitness of progeny from between- and within-population crosses. However, carrying out such breeding experiments might not be possible, or the groups might differ in other characteristics f.

In Drosophila melanogaster approximately half the effect of inbreeding depression is thought to be from nearly recessive lethals and half from detrimental alleles, with the standing amount of detrimental genetic variation of small effect but that have higher dominance g. However, D. melanogaster generally has an extremely large effective population size and the genetic architecture of their detrimental genetic variation probably reflects the first scenario in Table 1. Alternatively, for many endangered species, genetic
drift has been important, either because of a current small population size or a history of bottlenecks: the second scenario of Table 1 might therefore be more appropriate. As a result, endangered species might have a different genetic architecture with fewer variants of large detrimental effect, lower inbreeding depression, and perhaps higher genetic load, than do species with histories of larger population sizes.

If the genetic load increases in a population, the population might decrease in size, such that detrimental mutants of larger effect become effectively neutral and, subsequently, are more likely to become fixed. This feedback process has been named ‘mutation meltdown’ and, in theory, could result in the extinction of small populations27,28. However, it is not clear how significant mutation meltdown actually is because the extinction probability owing to other causes might be high in such small populations29 and other factors could also counter its effects30.

The magnitude and specific detrimental effects of alleles on fitness are highly variable because they might greatly depend on how these genotypes interact with the environment. Recent natural experiments are broadly consistent in reporting greater inbreeding depression in more stressful environments24. Inbreeding depression in captive populations or laboratory environments is generally thought to be underestimated compared with that in a natural environment. For example, the effect of inbreeding on male reproductive success in wild mice under laboratory conditions was minor, whereas in semi-natural conditions, inbred males had only ~20% of the success of outbred males31. In a comprehensive examination in Drosophila, the extent of inbreeding depression was greatly increased under stressful laboratory conditions32.

Specific deleterious variants

Populations of some endangered species possess an extremely high frequency of deleterious genetic variants33–35 and this genetic deterioration can result in extinction. Genetic restoration of populations can take place via the introduction of individuals from related populations or subspecies36 (Box 2). Gene flow from outside populations has appeared to restore the fitness of several populations that were suffering from a high frequency (or fixation) of detrimental alleles33–35.

Little progress has been made on the genetic characterization of genes with detrimental effects segregating in endangered species. However, the effort in the human and other genome projects to identify genes causing inherited disorders should provide information about homologous genes in endangered species. In addition, the ability to map genes affecting fitness-related traits portends imminent knowledge of the detailed architecture of genes affecting inbreeding depression, namely, the number and location of the genes, the distribution of their effects and their dominance, and the interaction (epistasis) of different genes.

There are documented examples of inherited disorders in captive populations of endangered species37,38, which would have important negative consequences for animals reintroduced to the wild. Although these recessive alleles are only present at a low frequency, many individuals in such a population could be heterozygous for them. Reducing the frequency of these alleles could be complicated and must be undertaken carefully to avoid jeopardizing the remaining genetic variation in these species36. For example, eliminating the recessive allele for chondrodystrophy, a form of dwarfism, from the California condor Gymnogyps californianus would require over half the population to be prevented from breeding38. Conversely, it appears that hereditary blindness in a captive population of gray wolves could be reduced without greatly influencing the genetic variation in the rest of the genome39. Disease alleles that have such severe negative effects are often easy to distinguish, but other detrimental alleles, which are present in every population, might not be as easy to identify39. Selection will also act to reduce the frequency of detrimental alleles, although if they are recessive the effect will be rather slow. If the population size is small, some detrimental alleles will be effectively neutral and purifying selection will not be effective. However, in a small population the expected frequency of recessive detrimental alleles of large effect is much lower than in an infinite population41.

Adaptive variation

The extent and pattern of adaptive (advantageous) variation is crucial to the long-term survival of endangered species. In particular, if there is no adaptive variation in a population that is facing a new environmental challenge, such as a new disease or competition from an introduced species, there is no potential for adaptive response except from new mutations or gene flow from other taxa. However, determining the extent and pattern of adaptive variation in the present or presumed future environments is rather difficult. Experimental tests of fitness and adaptiveness in a variety of environments could be carried out, but this is difficult, even in a model organism such as D. melanogaster, and practically impossible in an endangered species.

The extensive molecular data available today might provide new ways to determine whether adaptive selection has operated in the past on a given gene and, therefore, whether it might operate in the future. For example, rather than being able to measure the impact of selection in a single or a few generations by determining differential viability or reproduction, the cumulative effect over many generations could be observed by analyzing DNA variation. Two such approaches are the higher rate of nonsynonymous (amino acid changing) to synonymous substitutions and the pattern of sequence variation within and between species42,43.
sequences are placed in a phylogenetic tree with sequences from related species, they are widely distributed and the allelic lineages are maintained even across species, unlike the prediction for neutral loci. These findings suggest that long-term balancing selection has occurred and imply that the different variants have contributed to pathogen resistance.

The best general approach for maintaining genetic variation in a captive population is to minimize mean kinship51,52. A breeding scheme to select for any specific putatively adaptive allele might result in a faster loss of variation at most other loci in the genome. Taken to the extreme, if the contribution of one individual with a rare allele is greatly increased, then the pedigree will overly reflect the ancestral contributions of that individual. Selecting for a rare adaptive allele in captive pedigrees appears to be worse than the mean kinship approach, but the magnitude of the effect depends on the depth and complexity of the pedigree and the population size53.

**Neutral variation as an indicator of detrimental and adaptive variation**

The extent and pattern of neutral genetic variation has been used as a guide to the amount and pattern of detrimental and adaptive variation. If this association were always positive and strong, it could be a good predictor of how much detrimental variation might be present that could lower fitness and how much adaptive variation might be present to deal with future challenges. With greater numbers of more variable loci, estimates of the extent and pattern of neutral genetic variation will presumably become more accurate. However, different evolutionary scenarios might be responsible for different amounts of neutral variation (Table 2) and these scenarios might in turn result in different amounts or patterns of detrimental23 or adaptive variation. Positive associations might occur when the population size has been large or small for a long time or where stochastic effects dominate the extent of genetic variation. For example, there is a high correlation of (neutral) microsatellite and (adaptive) MHC loci variation in Gila topminnow Poeciliopsis o. occidentalis54 and desert bighorn sheep Ovis canadensis55,56 populations, two species in which stochastic factors appear to be important in the present spatial pattern of variation.

However, there might be situations in which such an inference based on a positive association is unfounded. For example, some time after a bottleneck, there might be a negative association between neutral and detrimental variation, because the amount of detrimental variation could recover faster as a result of a higher mutation rate. This might not be the case for all neutral variation because some genes, such as microsatellite loci, have mutation rates that are as high as those for detrimental variation. A negative association could also occur when separately bottlenecked populations become
mixed, resulting in high neutral variation but low detrimental variation. Finally, in a metapopulation, there might be a positive association between neutral and detrimental variation, but depending on the level of extinction, recolonization, and gene flow, the levels of variation might be either extremely low or high.

In some specific cases, it appears that there might be little genetic variation for some markers and substantial variation for others. For example, the cheetah Acinonyx jubatus was documented to have low genetic variation for allozymes and other genetic markers\textsuperscript{57}, and great concern was raised about its long-term survival prospects. However, further studies showed that cheetahs have substantial genetic variation for microsatellites\textsuperscript{58}, and captive cheetahs appear to exhibit inbreeding depression for juvenile survival\textsuperscript{59}, indicating variation for detrimental alleles. These differences have been explained with scenarios that include a bottleneck\textsuperscript{58} or a metapopulation with a small effective population size\textsuperscript{59} and different mutation rates for different markers.

Overall, the positive correlation between neutral and adaptive variation might not be particularly high. However, high neutral variation might indicate the potential for significant adaptive variation. Low neutral variation could indicate low adaptive variation, but the present population might either be well adapted or poorly adapted to its environment. Furthermore, gene flow between populations could result in low differentiation for neutral markers between populations, but there might still be strong adaptive differences in the populations. The extent and pattern of neutral variation is the result of nonselective forces and can potentially be used to identify the past importance of finite size, bottlenecks and population structure. However, adaptive variants might differ in mutation rate from neutral markers and selection might result in rather different patterns of variation within and between populations. In other words, neutral variants might be used as a guide to understanding nonselective effects, but the overlay of different mutational processes and selective regimes suggests that great caution should be used in making such predictions.

### Prospects

Genetics will probably play an even more important role in conservation in the future than it does now. The utilization of information from the human and other genome projects will provide substantially more background understanding and the appropriate use of these data should be of great benefit to conservation. In particular, techniques to screen and analyze large amounts of data, whether it is variation at marker loci or DNA sequence data, will be used to determine specific groups and individuals. However, to complement the high statistical power from these data, an evolutionary perspective is required to evaluate their biological importance. In addition, these data will allow an understanding of past evolutionary events in endangered species, whether they are nonselective, such as bottlenecks or gene

### Glossary

**Adaptive variation**: genetic variation that produces an advantage in fitness. A population might be fixed for an adaptive variant or polymorphic because the variant is newly arisen or is maintained by balancing selection.

**Detrimental variation**: variation that has a negative effect on fitness, often brought into the population by mutation and sometimes increased by genetic drift.

**Genetic load**: the reduction in mean fitness resulting from detrimental variation for a population compared to a population without lowered fitness.

**Genetic restoration**: the elimination or reduction of deleterious variants and recovery to normal levels of genetic variation for a population in poor genetic health by the introduction of individuals from another genetically healthy population.

**Inbreeding depression**: the decline in progeny fitness that occurs as the result of inbreeding.

**Neutral variation**: variation that has a small selective coefficient(s) relative to the population size, such that \( s < 1/(2N_e) \), where \( s \) is either the selective disadvantage of a detrimental allele or the selective advantage of an adaptive allele and \( N_e \) is the effective population size.

### Table 2. The amount of observed neutral variation and general predictions of the amounts of detrimental and adaptive variation under different evolutionary scenarios

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Observed Neutral variation</th>
<th>Predicted Detrimental variation</th>
<th>Adaptive variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equilibrium</td>
<td></td>
<td>High</td>
<td>High, Low to medium (retention more than neutrality)</td>
</tr>
<tr>
<td>Small population</td>
<td></td>
<td>Low</td>
<td></td>
</tr>
<tr>
<td>Bottleneck</td>
<td></td>
<td>Low (loss of alleles more than heterozygosity)</td>
<td>Low (loss of lethals more than detrimentals)</td>
</tr>
<tr>
<td>Shortly after</td>
<td></td>
<td>Low to high (depending upon mutation rate of variants)</td>
<td>High (assuming high mutation rate)</td>
</tr>
<tr>
<td>Some time after</td>
<td></td>
<td>Low to high (depending upon mutation rate of variants)</td>
<td>Low (assuming high mutation rate)</td>
</tr>
<tr>
<td>Metapopulation</td>
<td></td>
<td>High (owing to fixing of different alleles in different populations)</td>
<td>Low (owing to purging of variants within each population)</td>
</tr>
<tr>
<td>Mixture of separately bottlenecked populations</td>
<td>Low (owing to low effective population size)</td>
<td>Low (owing to purging)</td>
<td></td>
</tr>
<tr>
<td>With extinction, recolonization, and low gene flow</td>
<td>High (as if one large population)</td>
<td>High (as if one large population)</td>
<td></td>
</tr>
<tr>
<td>No extinction and more gene flow</td>
<td></td>
<td></td>
<td>High (as if one large population)</td>
</tr>
</tbody>
</table>

http://tree.trends.com
flow, or selective, such as detrimental or adaptive mutations. Examination of the pattern of variation at tightly linked loci might allow the determination of whether alleles have recently undergone selection to increase their frequency. This could be particularly important because the present-day small population size of some endangered species might make variants effectively neutral, but, if a population recovers to larger numbers these variants could become detrimental or adaptive.

References