The reduction of genetic diversity in threatened vertebrates and new recommendations regarding IUCN conservation rankings

Janna R. Willoughby^{1*}, Mekala Sundaram¹, Bhagya K. Wijayawardena², Steven J. A. Kimble¹, Yanzhu Ji¹, Nadia B. Fernandez^{1,2}, Jennifer D. Antonides¹, Maureen C. Lamb², Nicholas J. Marra³, and J. Andrew DeWoody^{1,2}

1. Department of Forestry and Natural Resources, Purdue University, West Lafayette, Indiana 47906 USA

2. Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47906 USA

3. Department of Population Medicine and Diagnostic Sciences, Cornell University, Ithaca, New York 14850

*corresponding author; Pfendler Hall G020, 715 W. State St., West Lafayette, IN 47906;
janna.willoughby@mac.com

Running Title: Conservation and genetic diversity

Abstract

Genetic diversity is necessary for evolutionary response to changing environmental conditions such as those facing many threatened and endangered species. To investigate the relationship between genetic diversity and conservation status, we conducted a systematic, quantitative review of vertebrate microsatellite data published since 1990: we screened 5165 previously

published articles and identified 1941 microsatellite datasets spanning 17,988 loci that characterized wild populations distributed among five vertebrate Classes. We analyzed these data in the context of conservation by comparing empirical estimates of heterozygosity and allelic richness between threatened and non-threatened species. We found that both heterozygosity and allelic richness are reduced in threatened species, suggesting that inbreeding and drift are both effective at removing genetic diversity in endangered populations. We then considered the criteria typically used to rank species of conservation concern (including declining population size, species range extent, and the number of mature individuals) to determine which of these criteria are most effective at identifying genetically depauperate species. However, we found that existing criteria failed to systematically identify populations with low genetic diversity. To rectify this, we suggest a novel approach for identifying species of conservation need by estimating the expected loss of genetic diversity. We then evaluated the efficacy of our new approach and found that it performs significantly better than existing methods for identifying species that merit conservation concern in part because of reduced genetic diversity.

Keywords: endangered species, IUCN criteria, microsatellite, range extent, population size, threatened

The IUCN (International Union for Conservation of Nature) is responsible for curating and maintaining the Red List, an accounting of species across the globe that require conservation efforts for long-term sustainability. This list is produced by assigning species to one of several hierarchical levels in order to convey relative conservation needs. The six IUCN conservation

ranks include 1) *Extinct in the wild*; 2) *Critically endangered*; 3) *Endangered*; 4) *Vulnerable*; 5) *Near threatened*; and 6) *Least concern*. Furthermore, IUCN identifies species as threatened, meaning that there is significant extinction risk, as those listed as *Critically endangered*, *Endangered*, and *Vulnerable* categories, whereas *Near threatened* and *Least concern* species face no immediate risk of extinction. In order to assign conservation designations to threatened taxa, IUCN uses several species-specific criteria. These factors include evidence of population decline, range extent and occupancy, and the estimated number of mature adults. Table 1 (criteria A-E) provides IUCN's specific cutoff values for each threatened category (IUCN 2013).

Estimates of a small census size derived via population decline, number of mature individuals, or other metrics, are likely to be good indicators of conservation concern for many species, but effective population size (N_e) is also a critical consideration and is often far less than the census size (Frankham 1995). N_e is the size of an idealized population that loses genetic diversity at the same rate as the focal population (Wright 1939). Conservation genetics theory predicts that populations with smaller N_e lose genetic diversity more rapidly than populations with a larger N_e (Allendorf & Luikart 2007). Molecular data can be used to estimate N_e , but biological variables such as sex ratio, effective number of breeders, and generation time cause nonlinear disparities between N_e and census size (Frankham 1995). For example, elephant seals (*Mirounga angustirostris*) have a unique mating system where a single male – the beachmaster – mates with a harem of females (Roelke et al. 1993). Because of the variation in reproductive success among males (Le Boeuf 1974), the N_e of elephant seals is much smaller than the census size (Allendorf & Luikart 2005; Bonnell & Selander 1974). Overall, these biological variables can result in a poor correlation between census size and N_e estimates based on genetic diversity.

Population genetics theory and data suggest that reductions in N_e result in a loss of genetic diversity, which can impact the long-term endurance of a population (Crow & Kimura 1970). Endangered species are typically characterized by many generations of reduced population size and often exhibit a pattern of low genetic diversity due to inbreeding and drift (Frankham 1996; Garner et al. 2005; Palstra & Ruzzante 2008; Spielman et al. 2004). Additionally, severe population bottlenecks can result in loss in genetic diversity (Allendorf 2005; England et al. 2003), the reduced population size presumably reducing N_e . Understanding these patterns of loss of diversity are important for conservation because the capacity for future evolution is based, at least in part, on the depth of the gene pool. If genetic diversity is low, the adaptive potential of a species is compromised (Markert et al. 2010). Therefore, the maintenance of genetic diversity is often important for the successful conservation of threatened and endangered species (Lacy 1997).

Conservation often requires focus on immediate demographic concerns, even though endangered species are negatively effected by reduced genetic diversity (Garner et al. 2005, Spielman et al. 2004) and, retaining genetic diversity in endangered species is important in promoting long-term sustainability (Lacy 1997). Herein, we examine the relationship between conservation status and genetic diversity by testing the hypothesis that vertebrate species of greater conservation concern have reduced microsatellite diversity compared to species of lesser conservation concern. We also consider the criteria typically used by IUCN to assign conservation rank – decline in population size, extent of species range, and number of mature adults (Table 1 criteria A-E) – in order to determine which IUCN criteria are most effective at identifying genetically depauperate species. We test the predictions that 1) declining populations, as categorized by IUCN, have less genetic diversity than increasing or stable

populations; 2) species with larger range sizes have more genetic diversity compared to species with smaller range sizes; and 3) species with fewer mature adults (as identified by IUCN) have less genetic diversity compared to species with a greater number of mature adults. We conclude by suggesting that IUCN explicitly consider genetic diversity as an aid in identifying species of conservation concern.

Methods

In order to investigate how the most commonly utilized IUCN ranking criteria relate to species genetic diversity, we required measures of genetic diversity from a wide array of vertebrate taxa. Although microsatellites are non-coding and presumed neutral, they are commonly used to provide estimates of population genetic diversity. For example, microsatellite heterozygosity is often positively, albeit weakly, correlated with fitness (Coltman and Slate 2003). To collect estimates of genetic diversity data, we used the ISI Web of Science to identify microsatellite studies published on wild vertebrate populations between 1990 and February 2014 using the topic search: microsatellite and (amphibia or avian or bird or fish or mammal or reptile). We then filtered studies by title to remove irrelevant medical and veterinary studies (terms: bipolar, cancer(s), carcinoma(s), carcinogenesis, cardio, cattle, child, childhood, clinical, diagnosis, diabetes(ic), farm, genome-wide, human genome, human microsatellite, laboratory, leukemia, patient, syndrome, tumor/tumour). We further limited the potential pool of studies to those published in the following journals: American Naturalist, Conservation Genetics, Conservation Genetics Resources, Evolution, Evolutionary Applications, Genetica, Genetics, Heredity, Journal of Evolutionary Biology, Journal of Heredity, Molecular Biology and Evolution, Molecular Ecology, Molecular Ecology Notes, Molecular Ecology Resources, Nature,

Proceedings of the National Academy of Sciences USA, Proceedings of the Royal Society-B, Science, and Trends in Ecology and Evolution.

We developed a standardized approach to screening and collecting data from manuscripts to help ensure consistency. First, we screened each paper to ensure that the studies contained data for at least one vertebrate wildlife population, used at least five microsatellite loci, and genotyped a minimum of ten individuals. Additionally, we required that observed heterozygosity and allelic diversity be reported so that we could calculate standard deviation across loci. Following the initial screening, we recorded the genus and specific epithet for the studied population, number of loci genotyped, and number of individuals genotyped for the sampling locality or population (as identified by the original authors) that had the most individuals genotyped. Additionally, we back traced primer citations and recorded incidences of cross-species (heterologous) primer use and cases where a mixture of cross-species and species-derived markers were used. We calculated the mean and standard deviation of a) observed heterozygosity across loci and b) the number of alleles per locus to serve as our response variables. In studies where multiple species were analyzed, we recorded information for one population for each species. When we found multiple population estimates from multiple studies for a single species, we utilized data from the population with the largest sample size, defined as the number of individuals multiplied by the number of loci genotyped.

After collecting microsatellite data from the literature, we used the IUCN Red List database (downloaded October 2014) to determine the proportion of all vertebrate taxa listed under each conservation rank and the criteria cited for the listing. We linked these data with the heterozygosity and allelic richness data gathered from the microsatellite studies. We also noted the IUCN assignment of species population size stability, which was ranked as declining,

increasing, or stable for many species. Finally, we estimated range size using species-specific range maps hosted by IUCN, for the subset of our species with available data (IUCN 2013). For each of these species, we transformed all maps into the cylindrical equal area projection using ArcMap10 (Environmental Systems Research Institute 2010) and then calculated the size of the area contained within each species map.

Statistical analyses

Before analyzing the connection between conservation status and microsatellite diversity, we considered the effect of shared phylogenetic history between species in our dataset, as well as the effects of differential sample size and the design source of the primers used. Because these nuisance variables are unlikely to be phylogenetically correlated, methods such as independent contrasts and phylogenetic least squares regression were inappropriate. Instead, we utilized a phylogenetic eigenvector regression (PVR), implemented in the R package PVR (Santos et al. 2013). To create the taxonomic variables used in the phylogenetic eigenvectors, we leveraged the Catalog of Life (<http://www.catalogueoflife.org>), which contains the accepted taxonomy and associated synonymous terms for >1.5 million species, and used the taxonomic levels of Class, Order, Family, Genus, and specific epithet. Additionally, we included the number of individuals and loci genotyped as well as the variable coded for primer origin (0=cross-species primers used; 1=both cross-species and species-derived primers used; 2=species-derived primers used) as additional predictor variables in our PVR. Finally, we used PVR to determine the amount of variation in the observed heterozygosity and allelic richness explained by the phylogeny and nuisance variables, as well as the amount of unexplained variation remaining in the dataset.

We examined our data for general trends regarding Class-specific genetic diversity and differences between threatened and non-threatened taxa within each class. To evaluate the

differences in genetic diversity observed between taxonomic Classes, we compared the 95% CI around each Class mean. Due to the observed differences in means, we analyzed each Class separately in further analyses. We also calculated the Pearson correlation coefficient between heterozygosity and allelic richness for each Class. We expected that if the correlation coefficient was high and positive, both estimates of diversity would provide effectively the same information. However, a poor or negative correlation would suggest differences in the signal conveyed by each metric.

We evaluated the differences between threatened and non-threatened species, impact of declining population size, species range extent, and small number of mature adults using a series of weighted linear models. To meet assumptions of normality, we transformed our heterozygosity and allelic richness estimates for each species. We calculated the log-odds of heterozygosity (*loh*) using the equation

$$loh = \log\left(\frac{H_O}{1-H_O}\right) \quad (1)$$

where H_O is the observed heterozygosity for a particular population. We interpreted the calculated odds ratio as the probability of a randomly drawn individual being homozygous at any particular locus. To transform allelic richness estimates, we took the square root of each value. After correcting for normality, we removed outliers since outliers have disproportionately high leverage in weighted regression. Within each Class, we ran three models using *loh* as our dependent variable and three models using square root of allelic richness as our dependent variable. In order to account for the variance within each species' genetic diversity estimate, we weighted our regressions using a measure of the relative standard deviation (*rsd*), which is defined as

$$rsd_i = \frac{x_i}{sd_i} \quad (2)$$

where the *rsd* of each study (*i*) is calculated from the mean and standard deviation of either heterozygosity or allelic richness. Next, we defined the weight (w_i) of each study as

$$w_i = \frac{1}{rsd_i^2} \quad (3)$$

for use in our regressions. We compared the 95% CI of the coefficient estimates within each taxonomic group (i.e. within each model) to look for significant differences among groups and for non-zero coefficient estimates.

In our first linear model, we compared threatened species, which IUCN defines as those listed as *Critically endangered*, *Endangered*, and *Vulnerable*, to species listed as *Least concern*. (Sample size concerns did not permit robust modeling of differences between each of the IUCN conservation levels.) Because species of concern typically have smaller population sizes, we hypothesized that heterozygosity and allelic richness would be smaller in the threatened group compared to the *Least concern* species.

In our second model, we evaluated the impact of a declining population size by comparing microsatellite diversity between the species trend values (declining, increasing, stable) for species in the *Least concern* conservation group. We expected that, if the IUCN-assigned population trend *declining* was representative of decreasing N_e , heterozygosity and allelic richness would be smaller compared to the genetic diversity estimates identified by *increasing* and *stable* population sizes. Additionally, we compared the 95% CI estimates around our genetic diversity measurements for *declining* species in the *Least concern* group to species listed in the threatened categories. In this case, we expected that the threatened and *declining* species would have less genetic diversity compared to the *Least concern* and *declining* species if the IUCN criteria accurately identified populations that have declined enough to impact N_e . For this analysis, we limited our dataset by only considering species where the genetic data were

collected from a manuscript published within 5 years from the year in which the species was evaluated by IUCN.

The third model was designed to examine the relationship between genetic diversity and range size. The IUCN criteria consider small range sizes and evidence of fragmentation, continuing decline in available habitat, or fluctuations in range extent, but range size is the determining factor for listing species in any of the threatened categories. This implies that, in accordance with the IUCN criteria (Table 1), the typical species with smaller range sizes should have a smaller N_e and thus less genetic diversity compared to species with larger range sizes. Therefore, we used the range size estimates gathered from the *Least concern* group to predict our measures of genetic diversity within each Class.

Finally, we compared the mean heterozygosity and allelic richness of threatened species that were identified as having a small number of mature adults to the estimated mature population size. We assigned population size based on IUCN criteria that included a maximum number of mature adults fewer than 10,000 (criteria C and D; Table 1), for each species in which these criteria were cited in the species listing. For example, all critically endangered species listed under criteria C were coded as having a maximum of 250 mature individuals whereas endangered species listed under criteria C were also coded as 2500 individuals. Due to sample size constraints, as this criterion is not typically used in ranking species (Supplementary Information), all Classes were analyzed as a single group.

Results

We identified a total of 5285 studies that met our original search and journal criteria (Table A1). Of these, we gathered complete data sets (i.e. mean and standard deviation of

heterozygosity and allelic richness) for 3066 populations. After filtering to eliminate replicate species estimates by choosing populations with the largest sample size (sample size = number of individuals * number of loci), and removing taxonomic categories with small number of populations sampled (Cephalaspidomorphi, 6; Chondrichthyes, 33; Sarcopterygii, 1), 1371 populations from five vertebrate Classes (ray-finned fishes hereafter fishes, amphibians, birds, mammals, and reptiles) were used in our analyses (Table 2).

We found records for 7419 vertebrates listed as threatened on the IUCN Red List. These were comprised of 1520 critically endangered species, 2542 endangered species, and 3357 vulnerable species. Although all criteria utilized by the IUCN to list species were cited as a reason to list in at least one species per conservation group, the largest percentage of criteria used was small range extent or occupancy (Criterion B in Table 1; Figure A1). In particular, a small range extent was almost twice as likely to be cited than a small area of occupancy (Table A2). Of the four other criteria, quantitative modeling was cited in less than 1% of all threatened taxa, and small population size, declining population size, or a combination of small and declining populations were used in varying degrees among the threatened conservation ranks (Figure A1).

Our PVR analysis suggested that very little variability in our dataset could be explained by shared phylogeny or other nuisance variables. For the heterozygosity variable, we found that 1.8% of the variation was attributed to phylogenetic constraints and 0.7% was explained by the number of individuals, number of loci genotyped, and primer type (cross-species or species-cloned) used, meaning that 97.5% of the variation remained unexplained. The same general pattern was observed for allelic richness (93.9% unexplained, 5.7% attributed to phylogeny, and 0.4% attributed to other nuisance variables).

We compared the mean heterozygosity and allelic richness between taxonomic Classes and found that fishes and reptiles had more genetic diversity than amphibians, birds, and mammals. Specifically, fishes and reptiles had significantly higher mean heterozygosity than birds, and significantly higher allelic richness than amphibians, birds and mammals (Figure A2). Additionally, heterozygosity and allelic richness were correlated in all taxa ($r = 0.63-0.68$, $n = 93-474$; Table A3).

In our first regression model, we found that threatened birds, mammals, and reptiles had had reduced heterozygosity relative to *Least concern* birds, mammals, and reptiles. Fishes, birds, and reptiles had fewer alleles relative to *Least concern* species. We observed an increase in the number of alleles in threatened mammals compared to *Least concern* mammals and no difference between heterozygosity in threatened fishes compared to *Least concern* fishes. We also found no difference in terms of heterozygosity or alleles in *threatened* vs. *Least concern* amphibians (Fig. 1).

We used three additional sets of models, based on the criteria used to assign the conservation rank, to evaluate the genetic diversity in species of concern. In the first model, we evaluated the impact of a declining population size by analyzing genetic diversity between population trend categories of declining, increasing, or stable. We found that heterozygosity was lower for declining species relative to stable species in amphibians and mammals, and that allelic richness was not lower in declining species for any Class. However, declining birds had more alleles relative to stable bird species, and heterozygosity and allelic richness in increasing populations for all Classes generally had large variance estimates and were not significantly different from declining populations (Table 3). Additionally, we found that *Least concern* species with declining population sizes typically did not have significantly different

heterozygosity or allelic richness compared to threatened species with identified population declines. However, threatened mammals had significantly fewer alleles relative to mammals in the *Least concern* conservation group and threatened reptiles had significantly higher heterozygosity relative to *Least concern* reptiles (Table A4).

In our second set of models, we used range size estimates to predict heterozygosity and allelic richness for the *Least concern* species. We found no significant relationship between range size and heterozygosity in birds and mammals, whereas in fishes, amphibians, and reptiles range size was inversely related to heterozygosity. However, we found no relationship between range size and allelic richness for any Class (Table 4).

Our third set of models was designed to investigate the relationship between IUCN's estimated number of mature adults and estimates of genetic diversity. We found that the IUCN estimated adult population size was not a significant predictor of heterozygosity or allelic richness (Table 5).

Discussion

We tested the hypothesis that vertebrate populations of conservation concern (as defined by IUCN) had reduced genetic diversity compared to populations of lesser conservation concern, and found that this was generally true (Fig. 1). We also considered the criteria typically used by IUCN to assign conservation rank (Table 1 criteria A-E) and found that the IUCN categorization of declining population only predicted decreased genetic diversity in amphibians and mammals, and that a smaller species extent (i.e., range size) was never a predictor of decreased genetic diversity. Finally, we found that the number of mature individuals, as estimated by IUCN, did not relate to measures of genetic diversity (Tables 3-5). These combined results suggest that

species with reduced genetic diversity, a major conservation concern, may be overlooked with current IUCN methodology unless populations have undergone a drastic decrease in census size.

We found evidence to support the hypothesis that threatened species would be less genetically diverse than species of lesser conservation concern (Fig. 1). The hypothetical relationship between conservation status and heterozygosity is consistent with both theoretical expectations and with empirical population data (Evans and Sheldon 2008; Garner et al. 2005 Spielman et al. 2004). Habitat destruction, exploitation, pollution, and a variety of other factors ultimately lead to small population sizes that inherently lead to smaller N_e , which increases the likelihood that alleles will be lost to genetic drift and heterozygosity reduced via inbreeding (Vandewoestijne et al. 2008). Our inability to detect these trends within some taxa and measures of genetic diversity may be due to the timing of sampling in the original studies: sampled populations may have been large even if global species numbers were decreasing or the species may have been globally stable but have since suffered declines. Although we cannot eliminate temporal issues related to sampling, our results also suggest that the IUCN categorization may not accurately convey the rarity of some species (Toledo et al. 2014) since, at least for some taxa, listing as a threatened species does not necessarily indicate a smaller N_e . This could be a biological reality; some small populations with small census sizes do not have reduced N_e as measured via genetic diversity (e.g., Busch et al. 2007). Additionally, some reductions in demographic population size do not result in reductions in genetic diversity due to life history traits that limit the effects of inbreeding and drift (e.g. long generation time; Kimble et al. 2014, Marsack and Swanson 2009). Alternatively, the lack of difference between Least Concern and threatened species could be due to the use of insufficiently sensitive criteria to effectively identify species with reduced N_e . Based on our regression analysis, we suggest that IUCN

criteria do not sufficiently probe the necessary data to allow conservation ranking based on a decrease in genetic diversity and, therefore, in N_e .

For most of our comparisons between threatened and *Least concern* species, we found similar trends using heterozygosity (log-odds) and allelic richness (square root) data. Immediately following a bottleneck, the increase in frequency of a rare allele leads to an increase in heterozygosity (Leberg 1992). However, when population size decreases over many generations, heterozygosity changes more slowly relative to allelic richness (Cornuet and Luikart 1996). Since many of our species under consideration were threatened and declining, theory predicted that heterozygosity would have not been as sensitive to changes in N_e as would allelic richness. Additionally, inbreeding directly influences zygosity, whereas allelic richness is only indirectly affected by inbreeding and is directly impacted by drift (Crow & Kimura 1970). Inbreeding is typically thought to increase in populations with small N_e and lead to decreased fitness (e.g., the extinction vortex; Fagan & Holmes 2006), but our results suggest that contemporary, neutral genetic variation is shaped as much by drift as by inbreeding.

Test of current IUCN criteria

Our three sets of regression models that investigated IUCN criteria suggested that the criteria used by IUCN may not detect species with low genetic diversity. Our model using population size changes found that only IUCN-labeled *declining* amphibians and mammals have reduced genetic diversity compared to *stable* amphibians and mammals (Table 3). It is possible that this trend could be an artifact of the inherent discrepancies associated with nonstandardized taxonomic categories (i.e., splitting versus lumping; see Avise and Mitchell 2007) or with the variation in genetic diversity inherent among species where, for example, some *stable* species may have less standing genetic diversity than other *stable* species (Tessier and Louis 1999).

However, with the exception of heterozygosity in reptiles and allelic richness in mammals, we also found that *Least concern* species with *declining* population sizes did not have significantly different genetic diversity compared to threatened species with identified population declines (Table A4). This suggests that either 1) *declining*, threatened species do not have reduced genetic diversity or 2) some populations labeled as *declining* and in the *Least concern* group have reduced genetic diversity and require conservation action. Due to the abundance of work demonstrating loss of genetic diversity in declining and threatened species (Crnokrak and Roff. 1999; Ellis et al. 2006; Eldridge et al. 1999; Lacy 1997; Roelke et al. 1993; Spielman et al. 2004; Taylor et al. 1994; Westermeier et al. 1998), we suggest that some *declining* species in the *Least concern* category have reduced genetic diversity.

In another set of models, we considered species range sizes and found no significant relationship with heterozygosity in birds and mammals, a negative relationship with heterozygosity in fishes, amphibians, and reptiles, and no relationship with allelic richness in all Classes (Table 4). This is counter to the pattern laid out by the IUCN criteria, where decreasing range extent leads to a higher conservation ranking (IUCN 2013; Table 1). While this criterion certainly identifies rare species (Yu & Dobson 2000), it will not systematically identify species with small N_e or reduced genetic diversity since, depending on the Class, the relationship between N_e and range size either does not exist or is in the opposite direction of what was anticipated by the IUCN criteria.

Finally, we compared the IUCN estimated number of mature adults to heterozygosity and allelic richness and found no significant relationship (Table 6). This is somewhat counterintuitive, as N_e is limited by the number of breeding adults found in a population. However, disparate estimates of population size relative to estimates of N_e are not uncommon

(Busch et al. 2007; reviewed in Frankham 1995). While our analysis of this criterion could have been influenced by including all Classes as a single group, the same regression without fishes and reptiles (significantly different Classes; Table A2) resulted in the same qualitative model output. Because of the relatively small samples size presumably due to the relative scarcity of available data in which to assess this criterion, it is possible that the number of mature individuals could accurately represent N_e . This is unlikely to be the case, however, as reviews of estimates covering the number of adults and estimated N_e suggest that the number of adults is not a consistent predictor of N_e (Frankham 1995).

Although we observed differences between the genetic diversity of threatened species compared to the *Least concern* species, these patterns were not related to the factors used by IUCN to identify and rank species of conservation concern. This suggests that the decrease we observed is due to an overall smaller N_e in threatened species, which is to be expected (Frankham 1996), but further suggests that species with small N_e will not necessarily be identified by IUCN criteria. This means that species with atypical mating systems or other atypical biologic characteristics may be systematically overlooked by the IUCN conservation ranking criteria. Therefore, we suggest that additional criteria are needed that specifically consider estimates of genetic diversity to systematically identify species with low or rapidly decreasing genetic diversity.

Proposed IUCN criteria - incorporating N_e

The current IUCN criteria effectively identify species with small demographic sizes, which is important in highlighting conservation need in some cases. However, genetic diversity is important in ensuring species survival (Frankham 1996) and, as such, we suggest that IUCN revise their formal evaluations to also include criteria capable of identifying species with low

genetic diversity. The concept of the effective population size (N_e) provides a mathematical framework in which the loss of genetic diversity can be modeled (Crow & Kimura 1970).

Provided N_e is known, the idealized rate of loss can be estimated using the equation

$$H_t = H_o \left(1 - \frac{1}{2N_e}\right)^t \quad (4)$$

where H_o is the heterozygosity at some measured point in time and H_t is the predicted amount of heterozygosity in the population after t generations, based on N_e (Crow & Kimura 1970).

However, N_e is generally unknown and is difficult to precisely estimate (Palstra & Ruzzante 2008). For the purposes of conservation ranking, however, we suggest approximating the census size based on known biological data (e.g. niche requirements, available habitat, previous census estimates) and inferring an estimate of N_e . For example, estimating population size as closest to 100, 500, 5000, or 10,000 is feasible for many species. For typical species, N_e could then be calculated using a conservative estimate of the N_e to census size ratio (0.14; Palstra & Ruzzante 2008). N_e could also be estimated from the various other equations that adjust for variation in reproductive success, skewed sex ratio, or population fluctuations (Crow & Kimura 1970) or utilize other types of data (Wang 2015) if these data are known or can be estimated. However if it is estimated, a rough estimate of N_e is sufficient to project the rate of loss of genetic diversity. Using available heterozygosity estimates and the estimated N_e , the number of generations (t) until heterozygosity drops below the 25% quantile of heterozygosity for all species within our database (25% quantile = 0.54) can be estimated. Finally, we suggest applying a cut off value to the estimated number of generations and using the estimate t value as an additional criterion (i.e. critically endangered $t \leq 10$; endangered $t \leq 50$; vulnerable $t \leq 100$) to identify threatened species (Fig. 2).

We have tested our novel approach using empirical data collected from the literature and compiled within our extensive microsatellite database. First, we identified species within the database for which IUCN criteria estimated population size (criteria C, D; Table 1) and for which multiple heterozygosity estimates were present. If three or more records were found, we randomly selected two records. Next, we used the IUCN estimated population maximums to estimate N_e using the N_e to census size ratio (0.14; Palstra & Ruzzante 2008). We then calculated t , the number of generations until heterozygosity is expected to reach the 25% quantile or below, using equation 4. Finally, we randomly assigned each species record, within the pair of records, to the *heterozygosity estimate* group or the *N_e - t estimate* (Table A5) and correlated the values using the Spearman rank method. We computed the p-value by permuting t 10,000 times (resampling using the ‘sample’ function with replacement in the R base package) and computing a null distribution of 10,000 *rho* estimates. We found that the number of generations (t) to the 25% quantile was well correlated with the independently obtained estimate of heterozygosity ($\rho=0.75$, $p\text{-value}=0.000$, $n=23$), meaning that a higher t indicates a more rapid rate of genetic diversity loss and, therefore, a greater conservation need compared to lower t estimates. Compared to the correlation between mean heterozygosity and the unaltered IUCN estimated population maximum ($\rho=0.21$, $p\text{-value}=0.325$, $n=23$), our t estimate is a much better identifier of low genetic diversity.

We do not suggest that genetic diversity should be the primary criteria for IUCN ranking, but given the molecular data and technology available today it seems reasonable to include N_e as an additional criteria for IUCN conservation ranking. The estimation of N_e using demographic data can be difficult due to a lack of required life-history data, and estimating N_e via genetic data is difficult due to underlying population structure and the variance among N_e estimators (Neel et

al. 2013; Wang 2015). However, we suggest that estimating N_e based on available data is generally viable and important, particularly when other species-specific analyses (e.g. quantitative models of species viability) are absent. We suggest that our method of estimating N_e and quantifying genetic diversity has a strong theoretical justification and is empirically tractable as a criterion for listing species on the IUCN Red List. Our estimates of t suggest that our cutoff values for listing species into threatened categories are reasonable (Fig. 2; Table A5), but obviously these estimates should be tested across a wide variety of taxa prior to adoption by IUCN.

Conclusions

The genetic diversity of vertebrates, as estimated via heterozygosity and allelic richness, is generally reduced in threatened species relative to species of lesser conservation concern. However, the criteria utilized by IUCN to identify species in need of conservation action do not systematically identify genetically depauperate species. Therefore, we propose that IUCN incorporates an additional criteria that addresses effective population size and genetic diversity for the use in ranking conservation priorities.

Acknowledgments

We thank members of the DeWoody lab for comments and feedback on the manuscript. This work was supported by Purdue via funding to J.A.D. through the University Faculty Scholar program and by a National Science Foundation doctoral dissertation improvement grant (DEB-1407021 to J.R.W. and J.A.D.). The funding bodies had no role in study design, collection and analysis of data, writing the manuscript, or the decision to submit the manuscript for publication.

Literature Cited

- Allendorf F. 2005. Genetic drift and the loss of alleles versus heterozygosity. *Zoo Biology* **5**:181-190.
- Allendorf F, Luikart G. 2007. *Conservation and the Genetics of Populations*. Blackwell Publishing, Oxford, UK.
- Avise JC, Mitchell D. 2007. Time to standardize taxonomies. *Systematic Biology* **56**:130-133.
- Bonnell ML, Selander RK. 1974. Elephant seals: genetic variation and near extinction. *Science* **184**:908-909.
- Busch JD, Waser PM, DeWoody JA. 2007. Recent demographic bottlenecks are not accompanied by a genetic signature in banner-tailed kangaroo rats (*Dipodomys spectabilis*). *Molecular Ecology* **16**:2450-2462.
- Cornuet J-M, Luikart G. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**:2001–2014.
- Crnokrak P, Roff DA. 1999. Inbreeding depression in the wild. *Heredity* **83**:260-270.
- Crow JF, Kimura M. 1970. *Introduction to population genetics theory*. Harper & Row, Publishers, New York.
- Eldridge MDB, King JM, Loupis AK, Spencer PBS, Taylor AC, Pope LC, Hall GP. 1999. Unprecedented low levels of genetic variation and inbreeding depression in an island population of the black-footed rock-wallaby. *Conservation Biology* **13**:531-541.
- Ellis JS, Knight ME, Darvill B, Goulson D. 2006. Extremely low effective population sizes, genetic structuring and reduced genetic diversity in a threatened bumblebee species, *Bombus sylvarum* (Hymenoptera: Apidae). *Molecular Ecology* **15**:4375-4386.

- England PR, Osler GHR, Woodworth LM, Montgomery ME, Briscoe DA, Frankham R. Effects of intense versus diffuse population bottlenecks on microsatellite genetic diversity and evolutionary potential. *Conservation Genetics* **4**:595-604.
- Environmental Systems Research Institute. 2010. ArcMap. Version 10. Environmental System Research Institute, Inc., Redlands, California.
- Evans SR, Sheldon BC. 2008. Interspecific patterns of genetic diversity in birds: correlations with extinction risk. *Conservation Biology* **22**:1016-25.
- Fagan WF, Holmes EE. 2006. Quantifying the extinction vortex. *Ecology Letters* **9**:51-60.
- Frankham R. 1995. Effective population size/adult population size ratios in wildlife: a review. *Genetical Research* **66**:95-107.
- Frankham R. 1996. Relationship of genetic variation to population size in wildlife. *Conservation Biology* **10**:1500-1508.
- Garner A, Rachlow JL, Hicks JF. 2005. Patterns of genetic diversity and its loss in mammalian populations. *Conservation Biology* **19**:1215-1221.
- International Union for the Conservation of Nature (IUCN). 2013. IUCN Red List of Threatened Species. <www.iucnredlist.org>.
- Kimble JA, Rhodes OE, Williams RN. 2014. Unexpectedly low rangewide population genetic structure of the imperiled Eastern Box Turtle *Terrapene c. carolina*. *PLoS One* **9**:e92274.
- Lacy RC. 1997. Importance of genetic variation to the viability of mammalian populations. *Journal of Mammalogy* **78**:329-335.
- Le Boeuf B. 1974. Male-male competition and reproductive success in elephant seals. *American Zoologist* **14**:163-176.

- Leberg PL. 1992. Effects of population bottlenecks on genetic diversity as measured by allozyme electrophoresis. *Evolution* **46**:477-494.
- Markert JA, Champlin DM, Gutjahr-Gobell R, Grear JS, Kuhn A, McGreevy TJ. 2010. Population genetic diversity and fitness in multiple environments. *BMC Evolutionary Biology* **10**:205.
- Marsack K, Swanson BJ. 2009. A genetic analysis of the impact of generation time and road-based habitat fragmentation on Eastern box turtles (*Terrapene c. carolina*). *Copeia* **4**:647–652.
- Neel MC, McKelvey K, Ryman N, Lloyd MW, Short Bull R, Allendorf FW, Schwartz MK, Waples RS. 2013. Estimation of effective population size in continuously distributed populations: there goes the neighborhood. *Heredity* **111**:189-199.
- Palstra FP, Ruzzante DE. 2008. Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? *Molecular Ecology* **17**:3428-3447.
- Reiter J, Panken KJ, LeBoeuf BJ. 1981. Female competition and reproductive success in northern elephant seals. *Animal Behavior* **29**:670-687.
- Roelke ME, Martenson JS, O'Brien SJ. 1993. The consequences of demographic reduction and genetic depletion in the endangered Florida panther. *Current Biology* **3**:340-350.
- Santos T, Diniz-Filho JA, eLuis Mauricio Bini TR. 2013. PVR: Computes phylogenetic eigenvectors regression (PVR) and phylogenetic signal-representation curve (PSR) (with null and Brownian expectations). R package version 0.2.1. <<http://CRAN.R-project.org/package=PVR>>

- Spielman D, Brook BW, Frankham R. 2004. Most species are not driven to extinction before genetic factors impact them. *Proceedings of the National Academy of Science* **101**:15261-15264.
- Taylor AC, Sherwin WB, Wayne RK. 1994. Genetic variation of microsatellite loci in a bottlenecked species: the northern hairy-nosed wombat *Lasiorhinus krefftii*. *Molecular Ecology* **3**:277-290.
- Tessier N, Bernatchez L. 1999. Stability of population structure and genetic diversity across generations assessed by microsatellites among sympatric populations of landlocked Atlantic salmon (*Salmo salar* L.). *Molecular Ecology* **8**:169-179.
- Toledo LF, Becker GG, Haddad CFB, Zamudio KR. 2014. Rarity as an indicator of endangerment in neotropical frogs. *Biological Conservation* **179**:52-62.
- Vandewoestijne S, Schtickzelle N, Baguette M. 2008. Positive correlation between genetic diversity and fitness in a large, well-connected metapopulation. *BMC Biology* **6**:46.
- Wang J. 2015. Estimation of effective population sizes from data on genetic markers. *Philosophical Transactions of the Royal Society B* **360**:1395-1409.
- Westermeyer RL, Brawn JD, Simpson SA, Esker TL, Jansen RW, Walk JW, Kershner EL, Bouzat JL, Paige KN. 1998. Tracking the long-term decline and recovery of an isolated population. *Science* **282**:1695-1698.
- Wright S. 1939. Size of population and breeding structure in relation to evolution. *Science* **87**:430-431.
- Yu J, Dobson FS. 2000. Seven forms of rarity in mammals. *Journal of Biogeography* **27**:131-139.

Figure 1. Linear regression coefficients as predictors of heterozygosity (log-odds heterozygosity; panels A, C, E, G, and I) and allelic diversity (square root of the number of alleles; panels B, D, F, H, and J) in fishes (A and B), amphibians (C and D), birds (E and F), mammals (G and H), and reptiles (I and J) using the model $Y_i = \beta_1 * \text{conservation status}_i + \epsilon_i$. Error bars are 95% CIs calculated from the standard error estimates. Significant relationships (95% CIs that did not overlap within each Class) in are noted with *, and model statistics are also shown (F, F-statistic; DF, degrees of freedom; r^2).

Figure 2. Proposed IUCN criterion (a) and example of usage (b). We have proposed estimating census size and effective population size (N_e) in order to determine the number of generations (t) until the species loses heterozygosity below the 25% of values (H_t) gathered in our database of microsatellite studies. The number of generations can then be used as an indicator of conservation need, based on the loss and projected future loss of genetic diversity.

- 1 Table 1. Description of criteria and cutoffs used by IUCN to determine conservation rank of vertebrate species; see IUCN technical
2 documentation, available at <http://www.iucnredlist.org/>, for more information.

Criteria	Critically endangered	Endangered	Vulnerable
A. Reduction in population size			
1.	$\geq 90\%$ over the last 10 years or three generations	$\geq 70\%$ over the last 10 years or three generations	$\geq 50\%$ over the last 10 years or three generations
2.	$\geq 80\%$ over the last 10 years or three generations	$\geq 50\%$ over the last 10 years or three generations	$\geq 30\%$ over the last 10 years or three generations
3.	$\geq 80\%$, projected or suspected to be met within the next 10 years or three generations	$\geq 50\%$, projected or suspected to be met within the next 10 years or three generations	$\geq 30\%$, projected or suspected to be met within the next 10 years or three generations
4.	$\geq 80\%$, projected or suspected to be met within the next 10 years or three generations and where cause is persistent	$\geq 50\%$, projected or suspected to be met within the next 10 years or three generations and where cause is persistent	$\geq 30\%$, projected or suspected to be met within the next 10 years or three generations and where cause is persistent
B. Geographic range			
1.	extent of occurrence $\leq 100 \text{ km}^2$ *	extent of occurrence $\leq 5000 \text{ km}^2$ *	extent of occurrence $\leq 20,000 \text{ km}^2$ *
2.	area of occupancy $\leq 10 \text{ km}^2$ *	area of occupancy $\leq 500 \text{ km}^2$ *	area of occupancy $\leq 2000 \text{ km}^2$ *
C. Number mature adults & population decline			
	≤ 250 mature adults and either:	≤ 2500 mature adults and either:	$\leq 10,000$ mature adults and either:
1.	continuing decline $\geq 25\%$ within three years or one generation	continuing decline $\geq 20\%$ within five years or two generations	continuing decline $\geq 10\%$ within ten years or three generations
2.	continuing decline, observed, projected, or inferred, in numbers of mature individuals *	continuing decline, observed, projected, or inferred, in numbers of mature individuals *	continuing decline, observed, projected, or inferred, in numbers of mature individuals *

D. Number mature adults	≤ 50 mature adults	≤ 250 mature adults	$\leq 1,000$ mature adults or very restricted area of occupancy or number of locations
E. Quantitative analysis	probability of extinction in the wild $\geq 50\%$ within 10 years or three generations	probability of extinction in the wild $\geq 20\%$ within 20 years or five generations	probability of extinction in the wild $\geq 10\%$ within 100 years or five generations
Proposed criterion:			
F. Neutral genetic diversity **	number of generations until heterozygosity decreases below the 25% quantile ≤ 10	number of generations until heterozygosity decreases below the 25% quantile ≤ 50	number of generations until heterozygosity decreases below the 25% quantile ≤ 100

*category has sub-requirements not listed

** criterion proposed within current paper, number of generations (t) calculated from equation 4

4 Table 2. Summary of mean and standard deviation of heterozygosity and allelic richness
 5 collected for analysis, screened from 5916 populations.

Order	Description	Total number of populations ¹	Number of unique species	Number of threatened species ²	Number of loci analyzed
Actinopterygii	fishes	728	474	64	7003
Amphibia	amphibians	118	93	27	1113
Aves	birds	380	284	61	4004
Mammalia	mammals	520	339	80	4315
Reptilia	reptiles	195	131	41	1553

6 1 number of collected data points

7 2 species listed as *critically endangered*, *endangered*, or *vulnerable*

8 Table 3. The influence of population size trend on genetic diversity; linear regression coefficients
 9 as predictors of heterozygosity (log odds H) and allelic diversity (square root of allelic richness)
 10 using the model $Y_i = \beta_1 * \text{population trend}_i + \epsilon_i$.

Model	Class	Population trend	Coefficient estimate ¹	F ²	df ³	P-value	r ²
Heterozygosity							
	fish			100.6	3 and 45	< 0.001	0.87
		declining	0.910, 1.895				
		increasing	0.347, 1.851				
		stable	1.396, 1.781				
	amphibian			13.92	2 and 36	< 0.001	0.44
		declining	-0.273, 0.120				
		increasing	--				
		stable	0.668, 1.470				
	bird			105.3	3 and 113	< 0.001	0.74
		declining	0.921, 1.305				
		increasing	0.649, 1.205				
		stable	1.016, 1.413				
	mammal			99.99	3 and 118	< 0.001	0.70
		declining	0.147, 0.364				
		increasing	0.854, 1.476				
		stable	0.940, 1.244				
	reptile			96.60	2 and 29	< 0.001	0.87

	declining	-0.792, 1.156				
	increasing	--				
	stable	1.104, 1.466				
Allelic richness						
fish			386.7	3 and 45	< 0.001	0.96
	declining	2.935, 4.168				
	increasing	1.722, 3.329				
	stable	3.249, 3.680				
amphibian			363.7	2 and 36	< 0.001	0.95
	declining	2.193, 2.940				
	increasing	--				
	stable	2.484, 2.938				
bird			824.5	3 and 113	< 0.001	0.96
	declining	2.783, 3.093				
	increasing	2.493, 3.029				
	stable	2.281, 2.650				
mammal			666.0	3 and 118	< 0.001	0.94
	declining	2.406, 2.918				
	increasing	2.378, 2.975				
	stable	2.742, 3.061				
reptile			2862	2 and 29	< 0.001	0.99
	declining	0.595, 4.089				
	increasing	--				

stable 4.200, 4.424

-
- 11 1 Significant relationships, 95% CIs that did not overlap within each Class, are bolded
- 12 2 F-statistic
- 13 3 degrees of freedom;

Table 4. The influence of range size on genetic diversity across taxa; linear regression coefficients as predictors of heterozygosity (log odds H) and allelic diversity (square root of allelic richness) using the model $Y_i = \beta_1 + \beta_2 * \text{range size}_i + \epsilon_i$.

Model	Taxa	Intercept	Slope ¹	F ²	df ³	P-value	r ²
Heterozygosity							
	fish	0.993, 1.354	-0.477, -0.043	5.532	1 and 42	0.023	0.12
	amphibian	1.120, 1.789	-4.881, -2.810	52.98	1 and 48	< 0.001	0.52
	bird	1.039, 1.280	-0.035, 0.052	0.152	1 and 188	0.696	0.00
	mammal	0.859, 1.126	-0.031, 0.020	0.187	1 and 79	0.666	0.00
	reptile	1.226, 1.608	-3.045, -0.303	5.731	1 and 36	0.022	0.14
Allelic richness							
	fish	2.972, 3.556	-0.593, 0.037	2.986	1 and 42	0.091	0.07
	amphibian	2.593, 3.105	-0.618, 1.039	0.248	1 and 48	0.621	0.00
	bird	2.724, 2.950	-0.007, 0.044	2.114	1 and 188	0.148	0.01
	mammal	2.632, 2.933	-0.036, 0.057	0.205	1 and 79	0.652	0.00
	reptile	3.779, 4.610	-1.895, 2.933	0.178	1 and 36	0.676	0.00

¹ Significant relationships, 95% CIs that did not overlap within each Class, are bolded

² F-statistic

³ degrees of freedom;

21 Table 5. The influence of number of individuals on genetic diversity; linear regression
 22 coefficients as predictors of heterozygosity (log odds heterozygosity) and allelic richness (square
 23 root of allelic richness) using the model $Y_i = \beta_1 + \beta_2 * \text{number of mature individuals}_i + \varepsilon_i$.

Model	Intercept	Slope ¹	F ²	df ³	P-value	r ²
Heterozygosity	0.465, 0.890	-0.012, -0.008	2.176	1 and 37	0.1486	0.06
Allelic richness	2.120, 2.288	-0.004, 0.001	0.030	1 and 37	0.865	0.00

24 1 Significant relationships, 95% CIs that did not overlap, are bolded

25 2 F-statistic

26 3 degrees of freedom;

27

Model: $y_i = \beta_1 * \text{conservation rank} + \varepsilon_i$
Figure 1

Least concern Threatened

a. Conceptual outline of proposed IUCN criterion

b. Examples of proposed IUCN criterion usage

Species	N	N_e	H_o	H_t	t	Conservation status
<i>Telespiza cantans</i> Laysan finch ¹	10,000	1,400	0.54	0.54	0	Critically endangered
<i>Equus ferus</i> Wild horse ²	250	35	0.73	0.54	21	Endangered
<i>Ailuropoda melanoleuca</i> Giant panda ³	2,500	350	0.60	0.54	68	Vulnerable

Figure 2