11 Using molecular methods to improve the genetic management of captive breeding programs for threatened species

Jamie A. Ivy and Robert C. Lacy

Captive breeding programs are powerful tools for the conservation of our natural resources. Both animal and plant biodiversity are in global decline (International Union for Conservation of Nature [IUCN] 2008), and the trend is for levels and rates of endangerment to continue increasing (Butchart et al. 2005; Chapter 1 by Honeycutt and colleagues). When in-situ conservation efforts are insufficient at stopping or reversing the decline of a species, captive populations often represent the only alternative for forestalling extinction. For example, it is currently estimated that hundreds of amphibian species will soon be extinct if emergency measures are not taken to protect populations in captivity until the threats to wild populations can be halted or overcome (Gascon et al. 2007). The success of numerous captive breeding programs has been well documented. As just a few of many examples, captive breeding programs have saved the black-footed ferret (*Mustela nigripes*), Przewalski's horse (*Equus caballus przewalskii*), and the California condor (*Gymnogyps californianus*) from final extinction (after the last wild populations were extirpated) and new wild populations of the golden lion tamarin (*Leontopithecus rosalia*), Arabian oryx (*Oryx leucoryx*), and whooping crane (*Grus americana*) have been successfully re-established from captive stocks.

Conservation breeding programs aim to maintain populations that are representative of their wild counterparts, to provide a reservoir for future reintroductions and recovery efforts (see Chapter 12 by Rhodes and Latch). Thus, the genetic goals of captive population management are to minimize genetic drift, retain genetic diversity, restrict inbreeding, and limit adaptation to captivity (Lacy 1994). The foundations of most captive breeding programs are pedigree analyses, which are used to manage both the demography and genetics of captive populations (Ballou & Foose 1996). Accurate pedigrees provide information on inbreeding, the kinships among individuals, and the distributions of individual founder contributions to a population. Collectively, this information is used to produce regular breeding recommendations intended to meet demographic and genetic goals. Although pedigree analyses are quite effective, they do have limitations. The most basic limitation of pedigree-based management is that it requires complete and accurate pedigrees to be effective. Thus, the genetic and demographic management of captive populations are hampered when missing or inaccurate parentage records produce incomplete pedigrees.

When a captive pedigree is inaccurate or incomplete, molecular data have the potential to improve breeding program management. Molecular markers can

P1: ICD Trim: 7in × 10in Top: 0.375in Gutter: 1in
CUUS999-11 cuus999/DeWoody ISBN: 978 0 521 51564 1 February 26, 2010 21:18

Table 11–1. Examples of studies that have incorporated molecular data into captive breeding programs

Contribution to breeding program	Species	Citation
Assessment of hybridization	Mexican gray wolf (*Canis lupus baileyi*)	Hedrick et al. 1997
	Lesser white-fronted goose (*Anser erythropus*)	Ruokonen et al. 2007
Gender identification	Old World vultures (*Gyps* species)	Reddy et al. 2007
Identification of geographic origin	Bearded vulture (*Gypaetus barbatus*)	Gautschi et al. 2003
	Galapagos tortoise (*Geochelone nigra*)	Russello et al. 2007
	Binturong (*Arctictis binturong*)	Cosson et al. 2007
Quantification of genetic differentiation	Baird's tapir (*Tapirus bairdii*)	Norton & Ashley 2004
	Chinese water deer (*Hydropotes inermis*)	Hu et al. 2007
	Scimitar-horned oryx (*Oryx dammah*)	Iyengar et al. 2007
Quantification of wild genetic diversity captured	Baird's tapir (*Tapirus bairdii*)	Norton & Ashley 2004
	Iberian wolf (*Canis lupus signatus*)	Ramirez et al. 2006
Species and subspecies identification	Chimpanzee (*Pan troglodytes*)	Ely et al. 2005
	Asian box turtles (*Cuora* species)	Spinks & Shaffer 2007

contribute to captive breeding programs in many ways, and numerous methods for incorporating molecular data into captive population management have been described (Table 11–1). The majority of widely applied methods, however, are focused on the accurate characterization of the individuals used to establish a captive stock, but are not related to the fundamental pedigree analyses that are used to select breeding pairs for the ongoing management and maintenance of captive breeding programs. For example, although it is invaluable to know that none of the captive Mexican wolf (*Canis lupus baileyi*) lineages contain any domestic dog (*C. lupus familaris*) or coyote (*Canis latrans*) ancestry (Hedrick et al. 1997), this information informs the selection of inbreeding pairs only so far as to justify the inclusion of all individuals in the pool of potential breeders. Rather than focusing on previously well-described methods for using genetic data to characterize populations genetically and taxonomically, this chapter discusses the methods and prospects for incorporating molecular data into the pedigree analyses that most captive breeding programs use for ongoing genetic and demographic management.

We start the chapter by providing a brief overview of common genetic terms used in pedigree and molecular analyses. We then discuss the goals of conservation breeding programs, and describe the methods typically used to genetically manage captive populations. The bulk of the chapter illustrates how molecular data can be incorporated into captive breeding programs to improve the effectiveness of pedigree-based population management. We discuss the benefits and limitations of different methods of incorporating molecular data into captive

P1: ICD Trim: 7in × 10in Top: 0.375in Gutter: 1in
CUUS999-11 cuus999/DeWoody ISBN: 978 0 521 51564 1 February 26, 2010 21:18

population management, and highlight issues that should be considered when different methods are employed. Finally, we discuss new prospects for the continuing incorporation of molecular markers into future captive breeding programs.

GENETIC TERMS AND THEIR USE IN CAPTIVE BREEDING PROGRAMS

Pedigree and molecular analyses share many analogous terms and concepts. Although some of the terms used in both types of analyses are quite similar, there are often subtle differences between their definitions. Furthermore, the relationships between pedigree-based and molecularly based concepts are not always readily apparent.

Kinship and mean kinship

The coefficient of kinship (or consanguinity) between two individuals (f) is the probability that two alleles at a given locus, one randomly drawn from each individual, are "identical by descent" – that is, they are copies of the same piece of deoxyribonucleic acid (DNA) descended from a common ancestor (Falconer & Mackay 1996). The concept is necessarily a relative one, in that all DNA can be traced back to a common source if the pedigree is extended far enough back in time. Therefore, in practice kinships are calculated relative to a baseline generation or source population in which all kinships are assigned to be 0, and alleles are all assumed to have independent origins. Subsequent to the base generation, if the parents of a diploid individual x are a and b, and the parents of individual y are c and d, then the kinship between individuals x and y is

$$f_{xy} = f_{yx} = \frac{1}{4}\left(f_{ac} + f_{ad} + f_{bc} + f_{bd}\right),$$

and the kinship of individual x to itself is

$$f_{xx} = \frac{1}{2}\left(1 + f_{ab}\right).$$

An individual's mean kinship ($\overline{f}$) is the average of pairwise fs between that individual and all living individuals in the population, including itself (Ballou & Lacy 1995; Lacy 1995). Under random mating, the $\overline{f}$ of an individual is the expected inbreeding coefficient of its offspring. Furthermore, an individual's $\overline{f}$ also provides a measure of its genetic representation in a population; individuals with high $\overline{f}$ have many living relatives and, thus, their alleles are well-represented, whereas individuals with low $\overline{f}$ are poorly represented because they have few living relatives.

Inbreeding coefficient

The inbreeding coefficient (F) of an individual is the probability that, at a given locus, both alleles are identical by descent (Falconer & Mackay 1996). Thus, an individual's F is equal to the f between its parents. If the parents of individual x

are a and b, then the inbreeding coefficient of individual x is

$$F_x = f_{ab}.$$

Given this equality, F_x also can be substituted into the equation for calculating an individual's kinship to itself:

$$f_{xx} = \frac{1}{2}(1 + F_x).$$

Coefficient of relationship

The coefficient of relatedness (r) between two individuals is the probability that, at a given locus, an allele sampled from one individual is identical by descent to at least one of the alleles at that locus in the second individual. (It should be noted that other authors have sometimes used the symbol r for the coefficient of kinship, f, as well as for other measures of relationship. We will use it here in the more consistent and precise sense of being the coefficient of relatedness as just defined.) In a noninbred, diploid population r is equal to $2f$; however, with inbreeding, the two measures of relatedness diverge, and r becomes less than $2f$. Numerous methods for estimating r from molecular data have been proposed for a variety of markers (e.g., Queller & Goodnight 1989; Lynch & Ritland 1999; Wang 2002). The performances of microsatellite-based estimators have received the most study, and have been shown to vary with the number of microsatellite loci employed, the numbers and frequency distributions of alleles at each locus, and the composition of relationship categories present in a population (Queller & Goodnight 1989; Ritland 1996; Lynch & Ritland 1999; Van de Casteele et al. 2001; Wang 2002; Milligan 2003). An individual's relationship to itself is always $r = 1$, and this measure of relatedness contains no information about the level of inbreeding. For that reason, and because r is not directly proportional to gene diversity (see next section), r is generally less useful for pedigree analyses and captive population management than is f.

Gene diversity

Gene diversity (G) is a common measure of genetic variation that can be calculated from both pedigrees and molecular data. G is defined as the heterozygosity expected in a random mating population, and it reflects both number of alleles and evenness of allele frequencies. When calculated from molecular data (e.g., microsatellites), the G of a single locus is

$$G = 1 - \sum p_i^2,$$

where p_i is the frequency of allele i, and the summation is over all alleles at a locus (Nei 1973). Multiple, single-locus estimates of G can be averaged to provide a measure of genome-wide variation. When calculated from a pedigree,

$$G = 1 - \overline{\overline{f}},$$

P1: ICD Trim: 7in × 10in Top: 0.375in Gutter: 1in
CUUS999-11 cuus999/DeWoody ISBN: 978 0 521 51564 1 February 26, 2010 21:18

in which $\overline{\overline{f}}$ is the average $\overline{f}$ in the population – that is, the mean of all pairwise kinships. An important distinction between molecular and pedigree-based measures of G is that molecular estimates represent the heterozygosity of alleles that differ by state (i.e., in molecular structure), and pedigree-based measures represent the probability that the two alleles at a locus are not identical by descent from a common ancestor in the pedigree.

In a randomly mating population, an individual's $\overline{f}$ is the expected inbreeding coefficient of that individual's offspring. By extension, $\overline{\overline{f}}$ is the expected mean inbreeding coefficient of all offspring. Thus, $\overline{\overline{f}}$ and $\sum p_i^2$ are conceptually equivalent because they both represent the average probability that an individual is homozygous at a given genetic locus. It is important to note, however, that the meanings of "homozygous" associated with the two concepts are not identical because one refers to identity by descent (pedigree-based measure) and one refers to identity of state (molecular measure). The difference in meaning between molecular and pedigree-based measures of G vanishes when diversity is expressed as a proportion of that in the defined baseline, or reference, population (e.g., the population founders). The proportional loss of molecular homozygosity due to accumulating kinship within a breeding population, relative to the reference population, is expected to be $\overline{\overline{f}}$. The G of the reference population in pedigree analyses is defined to be 1.0 because founders are assumed to be noninbred and to share no alleles that are identical by descent. Thus, the proportional change in heterozygosity (the decay through generations due to identity by descent) is expected to be the same whether measured by allele frequencies or pedigree-based kinships.

Gene diversity is more than just a convenient metric for quantifying the amount of genetic variation within a population. Gene diversity is proportional to the additive genetic variance in traits controlled by those loci, and therefore is proportional to the expected rate of response to selection (Falconer & Mackay 1996). Thus, the loss of gene diversity is both a measure of the accumulated inbreeding that can depress fitness of individuals and a measure of the loss of the population's potential for future adaptive evolution.

Applying concepts to incomplete pedigrees

Accurate, pedigree-based calculations of kinship, inbreeding, and gene diversity are possible only when pedigrees are completely known. Many captive populations, however, have missing or questionable parentage records that create ancestry gaps that result in incomplete pedigrees. To facilitate the calculation of genetic parameters in these situations, Ballou and Lacy (1995) developed algorithms for calculating f, $\overline{f}$, and $\overline{\overline{f}}$ from only the fully known lineages within each individual's pedigree. These algorithms use the proportion of an individual's genome that can be traced to known founders (k). Individuals with completely known ancestry have a k of 1 and individuals with no known ancestry (two unknown parents) have a k of 0. For any other individual,

$$k_x = \frac{(k_a + k_b)}{2},$$

where a and b are the parents of individual x.

P1: ICD Trim: 7in × 10in Top: 0.375in Gutter: 1in
CUUS999-11 cuus999/DeWoody ISBN: 978 0 521 51564 1 February 26, 2010 21:18

The kinship between individuals x and y (f'_{xy}) is the probability that an allele sampled from the known portion of y's genome is identical by descent to an allele sampled from among x's known maternal alleles, multiplied by the probability that a known allele sampled from x is maternally derived, plus the probability that the allele sampled from y is identical by descent to an allele sampled from among the known paternal alleles in x multiplied by the probability that a known allele sampled from x is paternally derived. Thus,

$$f'_{xy} = \left(f'_{my} \times \frac{k_m}{(k_m + k_p)} \right) + \left(f'_{py} \times \frac{k_p}{(k_m + k_p)} \right),$$

where m and p, respectively, refer to the dam and sire of individual x. The value of f_{xy} is undefined if individual x has no known ancestry. The kinship of an individual to itself is the probability that two alleles sampled with replacement from the known portion of the individual's genome are both the maternal allele, plus the probability that the two sampled alleles are both the paternal allele, plus the probability that one of the sampled alleles is maternal and the other is paternal multiplied by the probability that the maternal and paternal alleles are identical by descent. Thus,

$$f'_{xx} = \left[\frac{k_m}{k_m + k_p} \right]^2 + \left[\frac{k_p}{k_m + k_p} \right]^2 + \left[2 \times \left(\frac{k_m}{k_m + k_p} \times \frac{k_p}{k_m + k} \right) \times f'_{mp} \right].$$

To calculate $\overline{f}$ and $\overline{\overline{f}}$ from incomplete pedigrees,

$$\overline{f}'_x = \frac{\sum_{y=1}^{N} k_y \times f'_{xy}}{\sum_{y=1}^{N} k_y} \quad \text{and} \quad \overline{\overline{f}}' = \frac{\sum_{x=1}^{N} \sum_{y=1}^{N} \left(k_x \times k_y \times f'_{xy} \right)}{\sum_{x=1}^{N} \sum_{y=1}^{N} \left(k_y \times k_x \right)}.$$

Although these algorithms currently provide the only option for conducting pedigree analyses on incomplete pedigrees, short of either presuming that any undocumented parents were new, unrelated founders or excluding from analysis all animals with incompletely documented ancestries, their use is suboptimal because they produce values that can be either larger or smaller than the true values that would be calculated if the pedigree were completely known. Still, it is important to note that the algorithms are unbiased, as they assume that the probabilities of identity by descent are the same for alleles descended from unknown parts of the pedigree as they are for those alleles descended through traceable lineages.

AN INTRODUCTION TO CAPTIVE BREEDING PROGRAMS

The basic goal of most captive breeding programs is to maintain demographically self-sustaining populations that are genetically representative of their wild counterparts. Thus, in essence, captive breeding programs strive to prevent the evolution of captive populations away from the wild gene pool. This prevention is accomplished through careful genetic management that aims to retain genetic diversity, restrict inbreeding, and limit adaptation to captivity. Captive breeding

P1: ICD Trim: 7in × 10in Top: 0.375in Gutter: 1in
CUUS999-11 cuus999/DeWoody ISBN: 978 0 521 51564 1 February 26, 2010 21:18

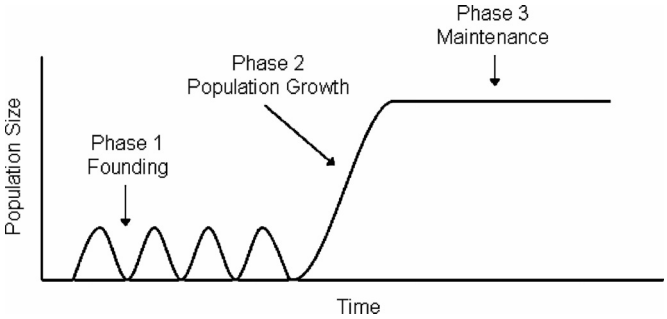


Figure 11–1. Three demographic phases experienced by a typical captive breeding program.

programs generally experience three demographic phases (Ballou & Foose 1996): founding, growth, and maintenance (Fig. 11–1).

The first phase of any captive breeding program is the founding of a captive population. Founders should genetically represent, as closely as possible, the species or population for which captive management is being initiated. Thus, the gene pool of the founders would ideally capture all possible alleles, at the same frequencies and linkage combinations observed in the wild. A common recommendation is that a captive population should have at least twenty unrelated founders, which would retain 97.5% of the gene diversity of the wild population from which they were randomly sampled (Fig. 11–2; Foose et al. 1986; Soule et al. 1986; Lacy 1994). For example, the recent population management guidelines for establishing captive assurance populations of threatened amphibians makes this explicit recommendation (Schad 2008), but then appropriately notes that usually more, and sometimes many more, than twenty individuals will need to be captured from the wild to assure that at least twenty successfully breed to establish the captive population.

Although founders are ideally unrelated, it is generally difficult, if not impossible, to ascertain relationships among wild-caught animals. Lacking such information, captive population founders are nearly always assumed to be unrelated. In other words, the founders are declared to be the baseline population for future kinship calculations. This means that captive breeding programs, at their best, retain the genetic variation that was present in the founders, with often considerable uncertainty regarding how well those founders represented the wild population(s) from which they came.

After captive populations are founded, the second phase of most captive breeding programs is population growth. The goal is to increase a newly founded population to a demographically self-sustaining size, while retaining both the allelic diversity and gene diversity (i.e., expected heterozygosity) captured by the founders. Captive populations should be grown as quickly as possible, because slower growth generally increases the likelihood that founders will die before contributing sufficient offspring to the breeding program. The probability that a given founder allele fails to be passed to a founder's offspring, and is therefore lost from a population, is equal to $(0.5)^n$, where n is the total number of offspring produced by the founder. Thus, for a given founder allele to be retained with a 99% probability, a founder must produce at least seven offspring (Fig. 11–3).

P1: ICD Trim: 7in × 10in Top: 0.375in Gutter: 1in
CUUS999-11 cuus999/DeWoody ISBN: 978 0 521 51564 1 February 26, 2010 21:18

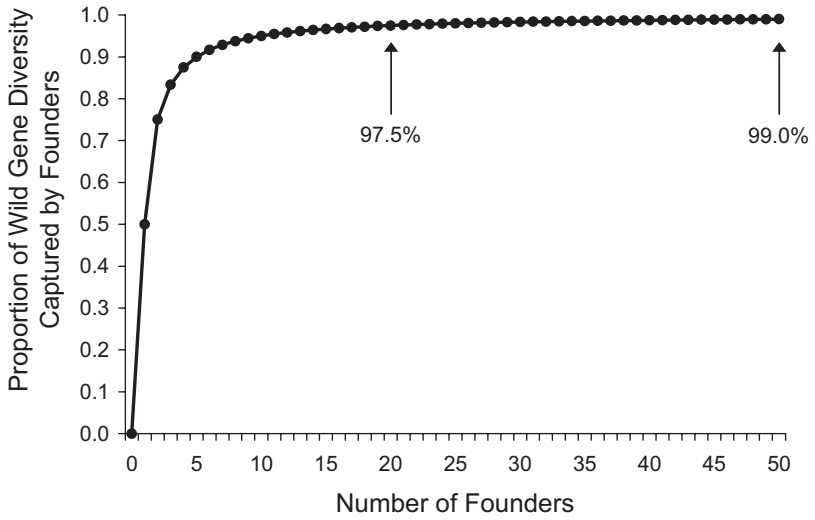


Figure 11–2. Proportion of wild gene diversity (expected heterozygosity) captured in population founders. Founders are assumed to be randomly selected from the wild population, and the proportion of gene diversity captured is calculated according to $H_f = H_w \times [1 - 1/(2N)]$, where H_f is the heterozygosity of the founders, H_w is the heterozygosity of the wild population, and N is the number of founders. It is commonly recommended that captive breeding programs secure at least twenty unrelated founders to retain approximately 97.5% of wild gene diversity.

Alleles are transmitted within linkage groups, rather than independently, however. Given the number of linkage groups in a typical vertebrate genome, a founder would need to produce at least twelve offspring to assure that all of its alleles were passed to its offspring with a 99% probability (Thompson 1995). Although the probability that a given founder's alleles will be retained increases as it produces more offspring, gene diversity will be maximized if all founders produce equal numbers of progeny. Thus, the two primary guidelines for captive population growth generally are 1) to grow the population as quickly as possible and 2) to have founders equally contribute to population growth.

The final phase of most captive breeding programs is population maintenance, during which a captive population is maintained at a carrying capacity that is usually dictated by the maximum size that can be supported by the resources (e.g., cage space) allocated by the institutions cooperating in the breeding program (Earnhardt et al. 2001). This allocation of resources is in turn often driven by the calculated number of breeders needed to maintain an acceptably low rate of gene diversity loss (Ballou & Foose 1996). For example, the target carrying capacity for okapis (*Okapia johnstoni*) in North American zoos is 200 animals (Petric & Long 2008). The target was set much higher, to 480 animals (Ballou & Mickelberg 2009), for golden lion tamarins (*L. rosalia*), because the species has a shorter generation length and therefore will lose genetic diversity more rapidly over time. At the capacity stage, after the desired population size has been reached, refined genetic management becomes a priority. The best methods for retaining genetic diversity, while still limiting inbreeding, are those that minimize the average kinship in a population (Ballou & Lacy 1995; Fernandez & Toro 1999; Sonesson & Meuwissen 2001). Methods for minimizing average kinship preferentially breed

P1: ICD Trim: 7in × 10in Top: 0.375in Gutter: 1in
CUUS999-11 cuus999/DeWoody ISBN: 978 0 521 51564 1 February 26, 2010 21:18

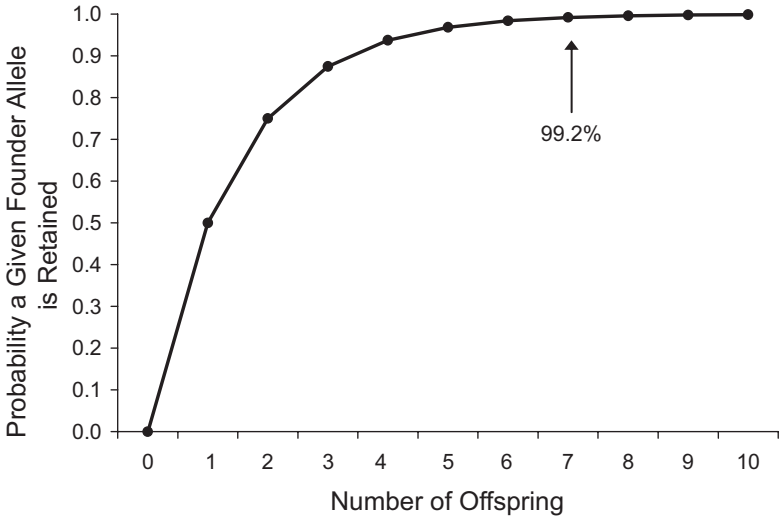


Figure 11–3. Probability that a unique founder allele is retained in a captive population as a function of the number of offspring the founder produces. For a given founder allele to be retained among the progeny with a 99% probability, a founder must produce at least seven offspring.

individuals that have the lowest $\overline{f}$, thereby equalizing founder contributions in the living population and retaining the genetic variation initially captured in the founders. By maintaining captive populations that genetically resemble their wild counterparts, captive breeding programs also aim to limit adaptation to captivity. Adaptation to captivity is a complex concept, however, because it can involve multiple genetic loci influencing multiple quantitative traits, which in turn can be subject to genotype-by-environment interactions. Furthermore, the manner in which a species may adapt to captivity is hard to predict, and the genes that may contribute to adaptation are difficult to identify.

Although the effectiveness would be difficult to evaluate, preferentially breeding those animals with the lowest $\overline{f}$ will slow adaptation to captivity. When breeding individuals are selected by $\overline{f}$, they are selected without regard for phenotype. Moreover, families with previously low survival or breeding success will be selected first for breeding, thereby working to equalize family sizes and countering, to the extent possible, natural selection for traits that might be adapted while in captivity. The equalization of family sizes is expected to halve the rate of adaptation to captivity, because selection is restricted to differential reproduction among siblings within families as selection between families is removed (Haldane 1924; King 1965; Frankham & Loebel 1992; Allendorf 1993; Lande 1995; Frankham et al. 2000). Alleles that prevent survival or reproduction in the captive environment will still be removed by selection, however, and family sizes will be difficult to equalize if reproduction and/or survival in captivity is poor. Breeding pairs often produce fewer or more offspring than desired, and culling excess to equalize family sizes is not always an acceptable option for endangered species and is not an efficient use of limited resources for population maintenance. Selection of breeding pairs by mean kinship is an efficient method to equalize family sizes to the extent possible, including preferentially breeding lineages that are

P1: ICD Trim: 7in × 10in Top: 0.375in Gutter: 1in
CUUS999-11 cuus999/DeWoody ISBN: 978 0 521 51564 1 February 26, 2010 21:18

under-represented due to poorer-than-desired breeding in prior generations, so that through the generations the representation of lineages is held as close to equal as possible. Although breeding pair selection schemes that use $\overline{f}$ likely reduce adaptation to captivity, little quantitative information is available regarding the extent to which real breeding programs can approximate an equalization of family sizes, the strength of remaining within-family selection for adaptations to the captive environment, or the rate of response to such selection pressures.

METHODS FOR INCORPORATING MOLECULAR DATA INTO CAPTIVE BREEDING PROGRAMS

Resolving pedigree gaps and identifying pedigree errors

Sources of pedigree errors and types of pedigree gaps

A fundamental way that molecular data can increase the effectiveness of captive breeding programs is to improve the accuracy and completeness of pedigrees. Captive breeding programs rely on pedigree analyses (f and $\overline{f}$ calculations) to select breeding pairs that will retain genetic diversity and limit inbreeding. Thus, the true effectiveness of genetic management is dependent on the accuracies of the pedigrees on which those analyses are based. Pedigree inaccuracies can arise through human error if, for example, an offspring's sire or dam is incorrectly recorded (e.g., Tzika et al. 2009). Errors also can be introduced into a pedigree by the behavior of the animals themselves. For example, the females of some bird species are known to "dump" eggs into nests of other females (Yom-Tov 1980; Petrie & Moller 1991; Zink 2000). If multiple breeding pairs of the same species are housed together and egg dumping occurs, a female may be erroneously assumed to be the mother of all the offspring in her nest. Identifying and resolving errors improves the accuracy of kinship calculations, which increases the true effectiveness of genetic management. Furthermore, the frequency of pedigree errors is currently rarely assessed, and reports of gene diversity and inbreeding levels in captive populations usually presume that pedigrees are recorded without error. This presumption is particularly problematic because those reports are often used to assess program success, so establishing pedigree accuracy also improves our ability to evaluate the status of captive breeding programs.

In addition to improving pedigree accuracy, molecular data also can be used to resolve pedigree gaps (e.g., Field et al. 1998; Zhang et al. 2005; Ivy et al. 2009). Gaps in a pedigree can arise for a variety of reasons. Before captive populations were recognized as important resources for conservation, historical parentage records were inconsistently maintained because few captive populations received regular genetic management. Thus, many contemporary captive populations have gaps deep in their pedigrees because they were initially founded during a time when accurate parentage records held little value. Captive breeding programs now strive to maintain complete pedigrees, but cases of unknown parentage still continue to arise in some situations. For example, many species are housed in groups that contain multiple males. If all males in a group are capable of siring offspring, the paternities of offspring born to the group are inherently

P1: ICD Trim: 7in × 10in Top: 0.375in Gutter: 1in
CUUS999-11 cuus999/DeWoody ISBN: 978 0 521 51564 1 February 26, 2010 21:18

uncertain. Pedigree gaps also can occur if a captive breeding program is composed of only a portion of the population maintained in captivity. In those situations, animals may move into and out of the managed population for a variety of reasons. If parentage records are not maintained for the portion of the population that does not belong to the captive breeding program, the ancestries of animals entering the captive breeding program from the unmanaged portion of the population are unclear. Those individuals cannot be accurately placed in the captive breeding program's pedigree, because neither their parentage nor their deeper ancestries are known.

Although gaps can be distributed throughout a pedigree, they often are characterized as being either historical or contemporary (Ivy et al. 2009). Contemporary gaps occur in the recent pedigree, and generally impact a limited number of individuals. For example, if a living animal with unknown parentage has not yet produced offspring, only the relationships between that individual and the rest of the population will be uncertain. Historical gaps occur deep in a pedigree, and can impact a significant proportion of the living population if the animal with unknown parentage has many descendants (Fig. 11–4). For example, consider a population in which 10% of living individuals are descended from an individual with unknown parentage. Even if the sires and dams of all of the living individuals are known, the relationships between all of those individuals and the rest of the population will be uncertain due to the unknown ancestry deeper in their pedigrees. This problem is compounded as the number of historical pedigree gaps increases and higher proportions of living descendants have unknown ancestry. In a sense, all captive populations have deep historical gaps because the founders of the pedigree were assumed to be unrelated. Thus, pedigree calculations yield

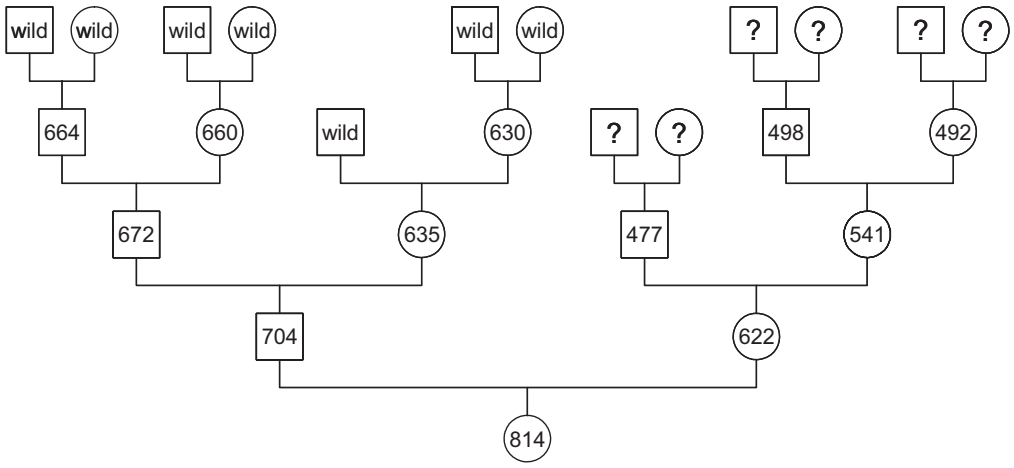


Figure 11–4. Examples of historical gaps in a captive parma wallaby (*Macropus parma*) pedigree. Squares in the pedigree represent sires, circles represent dams, individuals with "wild" parents were population founders, and individuals with "?" were born in captivity to undocumented parents. Individual 814 was living in the population as of January 1, 2007. Accurate kinships between that individual and all other individuals in the population could not be calculated because 814 had three ancestors (individuals 498, 492, and 477) with unknown parentage. Ancestor 498 had 45 additional living descendants in the population as of January 1, 2007, further hindering accurate kinship calculations for 29% of the total living population ($n = 157$).

P1: ICD Trim: 7in × 10in Top: 0.375in Gutter: 1in
CUUS999-11 cuus999/DeWoody ISBN: 978 0 521 51564 1 February 26, 2010 21:18

kinship and inbreeding estimates based solely on identity by descent from common ancestors within the captive population.

Using molecular data to identify errors and resolve gaps

Numerous types of molecular markers can be used to resolve gaps and ascertain the accuracy of pedigrees. Some common markers used for parentage analyses (see DeWoody 2005 for an overview) include restriction fragment length polymorphisms (RFLPs), mitochondrial DNA (mtDNA) haplotypes, and single nucleotide polymorphisms (SNPs). Each marker has benefits and drawbacks, and each marker type has unique characteristics that must be considered before it can be effectively applied to parentage analyses. One of the most common markers used to investigate parentage is single sequence repeats (SSRs). SSRs, or microsatellites, are biparentally inherited markers that are generally assumed to be selectively neutral. They are a popular choice for parentage analyses because, when compared to other molecular markers, microsatellite data are relatively easy to collect and only a few, suitably variable loci are usually required to ascertain parentage.

Molecular data can be used to establish parentage through either exclusion or assignment. For parentage exclusion, the genetic profiles of putative parents and offspring are compared to identify incompatibilities (Fig. 11–5). A putative parent is rejected as the sire or dam of an offspring if, at a given locus, the putative parent and offspring share no alleles. The probability that a given set of molecular loci will correctly exclude a potential parent is based on both the number of alleles and allele frequencies exhibited by the assayed loci (Selvin 1980). Exclusion probabilities rise as numbers of loci increase, numbers of alleles per locus increase, and allele frequencies become more uniform. Many captive populations exhibit low genetic variation because they were founded by small numbers of individuals. This problem is often exacerbated in breeding programs for endangered species (i.e., those of greatest conservation concern), because the small, wild populations from which founders were obtained often also had low genetic variability, fewer founders were initially available to establish the captive breeding program, and it is difficult or impossible to add more founders to augment genetic variation as time goes on. Thus, on average, a greater number of molecular markers may be needed for effective parentage exclusion in captive populations than in wild populations. Although parentage exclusion can be effective at ascertaining parentage, it does have limitations. For parentage exclusion to be maximally effective, genetic profiles must be available for all potential parents. Furthermore, suites of molecular loci with low exclusion probabilities may exclude only some of all possible parents.

If parentage exclusion identifies two or more genetically plausible parents when all possible parents have been sampled, molecular data can be used to assign the most likely parent to an offspring. A number of different likelihood-based methods of parentage assignment have been proposed, but categorical allocation is usually the most appropriate for captive populations because it assigns an entire offspring to a single sire or dam (fractional allocation assigns a portion of an offspring to each potential parent; Jones & Ardren 2003). Parentage assignment can be quite effective in some situations, but there is always some uncertainty associated with assigned parentage. Parentage assignment identifies the most

P1: ICD Trim: 7in × 10in Top: 0.375in Gutter: 1in
CUUS999-11 cuus999/DeWoody ISBN: 978 0 521 51564 1 February 26, 2010 21:18

		MICROSATELLITE LOCI													
ID	SEX	MP03		Pa595		Me14		Me17		Y175		G31-1		G19-1	
840	M	200	240	340	346	156	176	139	139	289	289	118	118	195	195
771	M	200	240	320	346	156	176	133	139	275	289	118	118	195	201
774	F	200	200	340	346	156	176	139	139	289	291	118	118	195	201
772	M	200	208	346	346	156	176	139	139	**275**	**287**	118	120	**169**	**201**
773	F	200	236	320	346	156	156	133	139	275	289	118	122	**171**	**201**
775	F	200	236	340	346	156	156	139	139	289	289	**122**	**122**	195	201
776	F	**204**	**204**	346	346	156	176	139	139	**275**	**291**	118	118	**169**	**197**

Figure 11–5. Example of parentage exclusion using microsatellites (data from Ivy et al. 2009). The parents of offspring 840 are 771 and 774; all other potential parents have been excluded at one or more loci. Microsatellite genotypes that are inconsistent with the parentage of 840 are highlighted in gray.

likely parent of an offspring, but any potential parent that cannot be genetically excluded has some probability of being the actual parent of the offspring. Thus, if all possible parents have been sampled, and parentage exclusion identifies two or more genetically plausible parents, captive breeding programs should attempt to increase their exclusion probabilities by adding additional marker loci before resorting to parentage assignment.

Regardless of the type of analysis employed, the effectiveness with which molecular data are able to ascertain parentage is dependent on sample availability. Genetic samples are rarely available for animals that are no longer living in a captive population, because the banking of samples from all animals is not routine. Still, programs like the San Diego Zoo's "Frozen Zoo" (Benirschke 1984) provide access to many more genetic samples than were available in the past. As the number of genetic samples that are banked continues to increase, pedigree gaps and uncertain parentage will become easier to investigate. Computer simulations have suggested that, for captive breeding programs that aim to minimize average kinship, recent ancestry has a greater influence than deeper ancestry on the retention of genetic diversity and the accumulation of inbreeding (Rudnick & Lacy 2008). Consequently, if a captive breeding program has a significant number of gaps spread throughout its pedigree, resolving even only the contemporary gaps can still significantly improve genetic management.

Estimating relatedness between individuals with unknown ancestry

Improving the completeness and accuracy of pedigrees can positively impact the management of captive breeding programs, but some captive pedigrees are riddled with historic gaps that can not be directly resolved. If a significant proportion of a living population has uncertain ancestry due to irresolvable pedigree gaps, molecular data can be used to quantify relationships by estimating r for pairs of individuals. Captive breeding programs select breeding pairs based on $\overline{f}$, which are calculated from a matrix that contains the f for each possible pair of individuals living in a population (Fig. 11–6). When pedigree gaps are present, each f is calculated from only the known portions of pedigrees (Ballou & Lacy 1995). The accuracy and value of these estimates decrease, however, as less of an animal's ancestry is available for the calculations, and no estimate is possible if neither parent is known. Molecular estimates of r can be used to modify or replace

ID	1	2	3	4	5	$\overline{f}$
1	0.50	0.03	0.02	0.00	0.25	0.16
2	0.03	0.50	0.03	0.00	0.03	0.12
3	0.02	0.03	0.53	0.00	0.05	0.13
4	0.00	0.00	0.00	0.50	0.00	0.10
5	0.25	0.03	0.05	0.00	0.50	0.17

Figure 11–6. Kinship (*f*) matrix for five individuals, with resultant mean kinship values ($\overline{f}$) for each individual listed on the right. Individual identification numbers (IDs) are given across the top and down the left side of the matrix. An individual's *f* to itself is shaded in gray, which helps to illustrate that the matrix is symmetric. Individual 4 is not related to any other individuals in the matrix and, thus, has the lowest $\overline{f}$. Individual 3 is slightly inbred, as its *f* to itself is greater than 0.5 (i.e., the value expected for noninbred individuals). Individual 5 has the highest average relationship to the rest of the individuals in the matrix, with $\overline{f} = 0.17$.

f in calculations, but genetic management is improved only if the incorporation of molecular estimates of relatedness increases the accuracy of those calculations.

Directly replacing f *with an estimate of* r *derived from molecular data*

Both *r* and *f* are measures of the ancestry shared by two individuals; thus, estimates of *r* could be used to replace estimates of *f* in the matrix used for mean kinship calculations. In the absence of inbreeding, the resulting matrix of *r* would be twice the matrix of *f*, so assessments of the proportional value of breeders would be the same. The most significant challenge to directly using molecular estimates of *r* in these calculations is that the estimates are notoriously inaccurate and suffer from large sampling variances (Csillery et al. 2006 and references therein). For example, unrelated individuals are expected to exhibit an *r* of 0.0000 in a noninbred, diploid population. When using a suite of ten microsatellite loci that each exhibit ten alleles at equal frequencies, however, the mean and standard deviation of *r* for 10,000 simulated pairs of unrelated individuals is 0.0002 ± 0.1069 (Fig. 11–7A). The *r* distributions of full-siblings and half-siblings also overlap in this simulation (Fig. 11–7A), further demonstrating the inaccuracy of *r* estimates that are based on the modest number of markers that are typically readily available for endangered species.

The accuracies of *r* estimators have been shown to be affected by the number of microsatellite loci employed, the numbers and frequency distributions of alleles at each locus, the composition of relationship categories present in a population, and the reference population from which allele frequencies are calculated (Queller & Goodnight 1989; Ritland 1996; Lynch & Ritland 1999; Van de Casteele et al. 2001; Wang 2002; Milligan 2003; Bink et al. 2008). Standard suites of microsatellites generally range from five to twenty loci, which are highly unlikely to produce estimates of *r* that are precise enough to produce calculated values of mean kinships that are sufficient to guide selection of the best breeding pairs. In the event that a suite of microsatellites is determined to produce sufficiently precise estimates of *r*, values of *r* and *f* still can not be directly substituted for each other due to differences in their definitions. In a noninbred, diploid population, *f* can be replaced by 1/2, but this equality fails to hold as inbreeding accumulates in a population. Additionally, whereas calculations of *f* from pedigrees range

P1: ICD Trim: 7in × 10in Top: 0.375in Gutter: 1in
CUUS999-11 cuus999/DeWoody ISBN: 978 0 521 51564 1 February 26, 2010 21:18

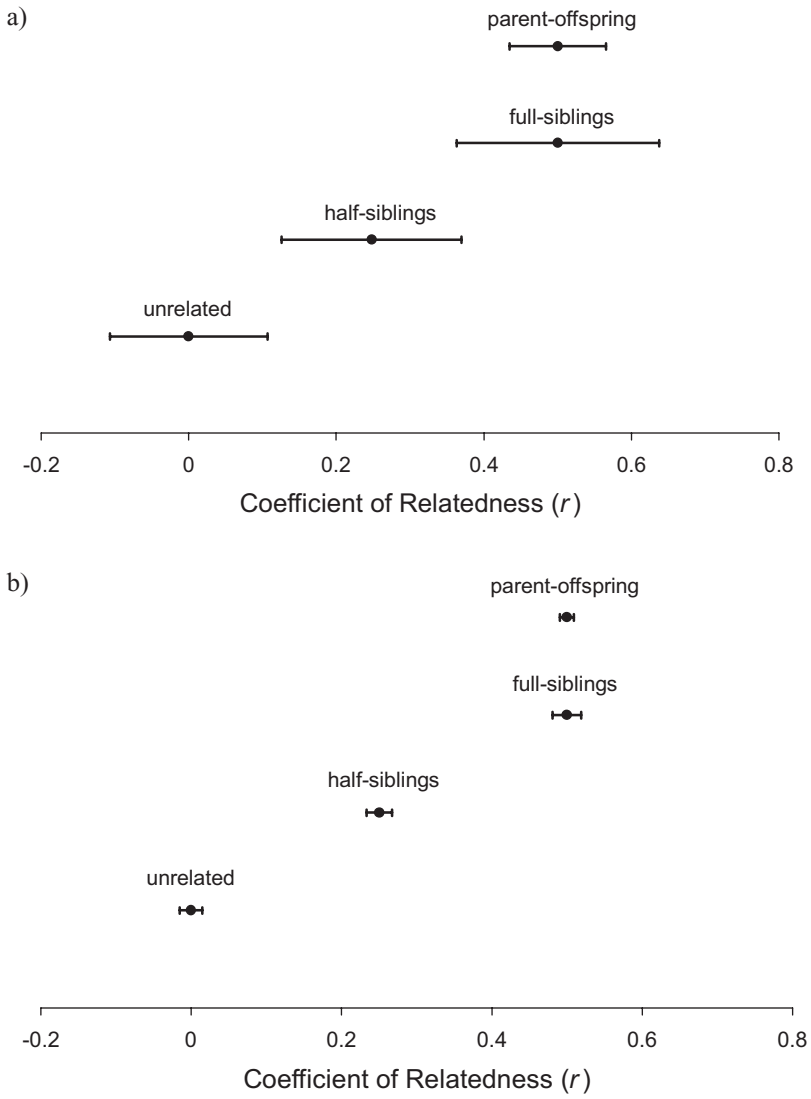


Figure 11–7. Means and standard deviations for four relationship categories, calculated from A) 10 microsatellite loci and B) 500 microsatellite loci. For each relationship category, 10,000 simulated pairs were used to calculate means and standard deviations. The microsatellite genotypes for all simulated individuals were drawn from loci that exhibited ten alleles at equal frequencies.

from 0 to 1, estimates of r based on molecular data can be negative because a pair of individuals can share fewer alleles even than what would be expected by chance if they were unrelated ($r = 0$). Yet, in the context of pedigree relationships, individuals can not be less related than having no common ancestors. Computationally, negative values could be incorporated into $\overline{f}$ calculations, but it can be difficult to interpret negative r values in the context of f.

There is another reason to exercise caution when using molecular estimates of r to guide captive breeding programs. Conceptually, there is a subtle difference in the genetic goals being pursued by pedigree-based and molecularly based breeding strategies; pedigree-based breeding strategies (i.e., mean kinship

P1: ICD Trim: 7in × 10in Top: 0.375in Gutter: 1in
CUUS999-11 cuus999/DeWoody ISBN: 978 0 521 51564 1 February 26, 2010 21:18

breeding schemes) aim to keep the genetic contributions of all founders as constant as possible through subsequent generations, whereas breeding strategies that use empirical estimates of relatedness as the basis for selecting pairs preferentially breed those individuals with the most distinctive array of alleles at the sampled loci (Jones et al. 2002). In the former case, the breeding program will seek to minimize homozygosity due to common descent (autozygosity) and maintain allele frequencies across all loci as closely as possible to the frequency distribution in the wild-caught founders. In the latter case, the breeding scheme will seek to minimize homozygosity due to both 1) identity by descent and 2) identity of state from independent origins (allozygosity), and to maintain allele frequencies at the sampled loci as evenly as possible. If large numbers of unlinked, fully neutral loci are used for the estimation of relatedness, however, then the extent of allozygosity among pair-wise comparisons should be relatively constant and estimates of *r* should be proportional to the true, but perhaps not otherwise knowable, pedigree relationships.

Estimates of *r* from molecular markers are often strongly influenced by the alleles that are present in only one or a few individuals, and individuals carrying rare alleles will often be accorded preference in breeding, either because of their lower estimated kinships to the population or even specifically as a means to retain rare alleles that otherwise would be at high risk of being lost from the population. It is important to note, however, that the conservation goal of captive breeding programs is rarely to preserve the specific alleles present at the sampled loci, but rather to preserve diversity across the entire genome to retain, as closely as possible, the characteristics of the wild population in the captive stock. Moreover, if one or more of the sampled loci are linked to (or are themselves) loci under selection, then individuals with unique alleles might be carrying rare, deleterious alleles. Preferentially breeding these individuals will increase heterozygosity at the sampled loci, but it will do so by shifting allele frequencies away from those of the founders, possibly increasing the frequency of alleles that were under negative selection in the wild population. More work is needed on the likelihood of linkage of commonly used markers (e.g., microsatellites) to loci under selection before we can know whether rarer alleles are simply more informative about relationships, as is commonly assumed, or also sometimes reflect the consequences of past natural selection.

As microsatellite characterization technologies improve and microsatellite data become less costly to collect, it may one day be possible to use hundreds of microsatellites to calculate *r* for some species. Even with a suite of 500 unlinked microsatellite loci that each exhibit 10 equally frequent alleles, *r* estimates for a range of relatedness categories still exhibit standard deviations near 0.01, however (Fig. 11–7B). It is currently unclear what level of accuracy is needed for *r* estimates to be beneficially incorporated into $\overline{f}$ calculations. Thus, additional research is necessary to identify the most likely scenarios for which estimates of *r* could be directly used to improve the genetic management of captive populations.

Using molecular estimates of r *to identify relationship categories*

When typical suites of molecular markers are employed, relatedness estimators have been shown to be sufficiently accurate for distinguishing first-order

P1: ICD Trim: 7in × 10in Top: 0.375in Gutter: 1in
CUUS999-11 cuus999/DeWoody ISBN: 978 0 521 51564 1 February 26, 2010 21:18

relatives from unrelated individuals (Piper & Rabenold 1992; Blouin et al. 1996; Glaubitz et al. 2003). Thus, rather than using estimates of *r* directly in mean kinship calculations, a more useful way of incorporating *r* into captive population management might be to use them for identifying degrees of relationship (Ivy et al. 2009). If pairs of individuals with little or no known ancestry could be assigned to relationship categories, inaccurate estimates of *f* in the matrix used for $\overline{f}$ calculations could be replaced with new values to represent specific degrees of relationship. For example, if an *r* estimate suggested that two individuals were first-order relatives, their *f* might be set to 0.25 (the theoretical value expected for that relationship).

For a given suite of microsatellites, empirically determined allele frequencies can be used to simulate pairs of individuals that represent various relationship categories. Means and standard deviations can then be calculated across hundreds or thousands of pairs to describe the expected distributions of *r* values representing each type of relationship (Fig. 11–7), and each pair of individuals can be assigned a degree of relationship based on the distribution into which their *r* falls. When the distributions of two relationship categories overlap, the midpoint between the means can be used as a cutoff for assigning a pair to one category or the other (Blouin et al. 1996). For example, consider the overlapping full-sibling and half-sibling *r* distributions illustrated in Fig. 11–7A. The midpoint between the distribution means is 0.3741. Thus, for pairs that exhibit *r* values between the distribution means, pairs with *r* values greater than the cutoff would be classified as full-siblings and pairs with *r* values less than the cutoff would be assigned as half-siblings or to lesser degrees of relatedness.

Estimates of *r* theoretically can be used to discriminate among numerous relationship categories. The sampling variances associated with a given suite of molecular markers, however, limit the level of discrimination possible. As the number of potential relationship categories to be considered increases, more distribution overlap will be observed, and it will be harder to conclusively place a pair of individuals into a given relationship category. Furthermore, relationship categories of the same order often have similar means and distributions (e.g., full-siblings and parent–offspring pairs both exhibit a mean *r* of 0.5; Fig. 11–7). Because $\overline{f}$ calculations use *f* values rather than specific relationship categories, it is more important to correctly identify the order of relationship than the actual type of relationship that exists between two individuals. In other words, if *r* suggests that two individuals are either siblings or a parent and offspring, identifying the correct relationship category is less important than identifying that the individuals are first-order relatives with an expected *r* of 0.5. Standard suites of ten to twenty microsatellites should often be sufficient for distinguishing among first-order relatives, second-order relatives, and unrelated individuals with acceptable confidence, but differentiating among more distant relationship categories is likely beyond the capabilities of most molecular studies.

The suitability of a given suite of molecular markers for distinguishing among relationship categories can be evaluated by a number of methods. One option is to calculate misclassification rates for a range of relationship categories (Blouin et al. 1996). After *r* distributions have been generated, misclassification rates can be calculated by simulating individuals of a given relationship category and

P1: ICD Trim: 7in × 10in Top: 0.375in Gutter: 1in
CUUS999-11 cuus999/DeWoody ISBN: 978 0 521 51564 1 February 26, 2010 21:18

quantifying the proportion of those individuals that are incorrectly assigned to an alternate category. A second option for assessing the suitability of molecular markers for distinguishing among relationship categories is to estimate the proportion of variance explained in the marker-based relatedness by true relatedness (Van de Casteele et al. 2001). For a given relationship category, the proportion of variance explained by true relatedness can be calculated from the sampling variance associated with the distribution of *r* values for that category. It is currently unclear what misclassification rates and proportions of variance explained by true relatedness are needed for *r* estimates to be beneficially incorporated into mean kinship calculations. It is likely, however, that different levels of accuracy of estimated *r* will be acceptable for different pedigrees that exhibit variable amounts of unknown ancestry. For example, more accurate estimates of *r* may be required if little unknown ancestry is present and the pedigree-based values of *f* are already fairly accurate, whereas less accurate estimates of *r* might be acceptable if the amount of unknown ancestry in a pedigree is so large as to prohibit meaningful kinship calculations from pedigree information alone.

Molecular markers can be powerful tools for inferring relationship categories, but inconsistencies in category assignment can arise when multiple pairs of individuals are considered simultaneously. For example, it is possible for *r* estimates to suggest that an individual is closely related to one of a known sibling pair but unrelated to the other sibling. Although there are no mathematical reasons why these relationship category assignments can not be incorporated into $\overline{f}$ calculations, results such as these are difficult to interpret in the context of a pedigree. Because these types of situations can arise for empirical estimates of *r*, it is important that the assignments of relationship categories be carefully considered both against each other and in conjunction with any ancillary pedigree information that might be available. Even if one or both parents of an individual are unknown, there may still be information available about the birth date, birth location, and possible parents. This information can be used to either support or refute relationships suggested by molecular estimates of *r* (Gautschi et al. 2003; Ivy et al. 2009).

Although the focus of the discussion has been on using molecular data to improve captive breeding programs for endangered species, in which the genetic goal is to minimize genetic change through generations, the methods we discuss for using molecular data also apply to programs that use artificial selection to improve aspects of performance of economic gain. The Chapter 11 Box describes the use of molecular data to identify categories of kinship in a population of trees, for which direct tracking of pedigree relationships through generations would be difficult and require many years.

Using molecular data to manage breeding programs for organisms living in groups

Pedigree analyses and the genetic management of breeding programs have historically been focused on species that can be placed in pairs for breeding, so that pedigree records on parentage can be maintained and pairings can be controlled. As described earlier in this chapter, pedigree information is often incomplete

P1: ICD Trim: 7in × 10in Top: 0.375in Gutter: 1in
CUUS999-11 cuus999/DeWoody ISBN: 978 0 521 51564 1 February 26, 2010 21:18

for a variety of reasons. Thus, molecular data can be a powerful tool for providing missing relationship information so that captive breeding programs can be more effective at minimizing genetic change (see Chapter 10 by Waples and colleagues). For many species that require captive breeding programs for their conservation, however, genetic management problems go beyond simple gaps in the pedigree or the uncertain ancestry of some individuals. The requirements of a species and constraints on breeding facilities often necessitate that breeding populations be maintained in large groups with multiple males and females with no possibility for controlling pairings, often little opportunity to observe parentage, and sometimes limited ability to identify and monitor individuals through their lifetimes. Examples of species that are maintained and bred within groups include antelope species which are adapted to living in herds, bats, penguins, flamingos, and other species that breed within colonies and often need the stimulation or protection of a colony before they will breed, and small species such as many frogs, insects, snails, and fishes for which it is usually impractical to manage as individuals or pairs maintained in separate enclosures. Increasingly, it is being recognized that species that are difficult or impossible to manage with traditional pedigree-based methods include many that are threatened with extinction. For example, some of the highest levels of endangerment and extinction are among the freshwater fishes, terrestrial and freshwater mollusks, coral reef inhabitants, and amphibians (Baillie et al. 2004; IUCN 2008).

Due to the difficulties inherent in trying to manage group-living species as individuals, some breeding programs rely on breeding schemes that manage movements of individuals among breeding groups, rather than on trying to control each pairing, to maximize retention of genetic diversity within and among groups. Methods such as maximum avoidance of inbreeding (Kimura & Crow 1963; Princée 1995) rotate one sex or the other through breeding groups to delay inbreeding and equalize group contributions to future generations. Such regular systems of breeding can be difficult to sustain for many species of wildlife, however, because successful reproduction is often not sufficiently reliable to prevent the failure of some groups. Still, if the survival and reproduction of groups and the transfers among groups can be carefully monitored, then the within- and between-group gene diversity, and pair-wise estimates of genetic divergence, can be used to determine the optimal transfers that retain overall genetic variation while avoiding excessive within-group inbreeding (Wang 2003, 2004).

When species must be managed as groups, molecular data can be used to monitor and enhance breeding schemes that are based on the transfer of individuals (e.g., maximum avoidance of inbreeding). The rate of gene diversity loss from groups and divergence among groups can be empirically measured to determine if breeding programs are performing as desired. This information can be particularly informative because, given the lack of control over the pattern of breeding within most groups (e.g., all the breeding could be done by one pair per group or could be spread among multiple pairs), the actual structure of genetic variation among groups could diverge substantially from that expected under theoretical models.

Molecular estimates of gene diversity also could be used in addition to or in place of models based on group histories to guide animal transfers among groups. The observed gene diversity of an overall population can be partitioned

P1: ICD Trim: 7in × 10in Top: 0.375in Gutter: 1in
CUUS999-11 cuus999/DeWoody ISBN: 978 0 521 51564 1 February 26, 2010 21:18

into between- (G_b) and within- (G_w) subpopulation components (Nei 1973), and the concept of kinship, defined as the probability that alleles sampled from two entities are autozygous, can be applied to between-group relationships (G_b) as well as relationships among pairs of individuals (f and r). Mathematically and conceptually, the gene diversity between entities (when scaled so that complete lack of alleles shared by common descent is assigned $G = 1$) is the same as the converse of kinship ($1 - f$) between those entities, whether the entities are individuals, breeding groups, or larger populations. Thus, in direct parallel to traditional pedigree management focused on individuals, crossing between groups with lowest kinship (highest G_b) will result in a minimization of inbreeding ($1 - G_w$), and preferentially propagating the groups that have lowest mean kinship (highest mean G_b to the array of groups) will maximally retain overall gene diversity and slow the response to natural selection by countering the between-group component of selection.

Just as the precisions of r estimates between individuals are strongly dependent on the number of loci sampled, the precision of the parallel estimates of relatedness between groups is dependent on the number of individuals (and the number of loci) sampled. Because the genotypes of multiple individuals from each group are used to estimate the relatedness between groups, those estimates will be more precise than those of relatedness between diploid individuals. Therefore, whereas the precision of estimates of genetic measures based on partial pedigree information and the usefulness of management based on pedigree calculations degrades as the level of management moves from individuals to groups, the precision of empirical estimates of population structure and the potential usefulness of molecular data to guide program management increases. Work is needed to explore the relative costs and benefits (in terms of reaching genetic goals as well as resource costs) of more intensive individual and pedigree-based management versus coarser, group level, and empirically guided management. We are unaware of any breeding programs for group-living endangered species that are currently being guided by a molecular characterization of genetic structure each generation, but such methods may become more useful and efficient as it becomes easier, in terms of both time and money, to score large numbers of marker loci for a wider variety of species.

Identifying individuals with rare or important alleles

It has been suggested that molecular data could be used to augment captive breeding programs by characterizing genotypes at genetic loci known to be important to fitness, then managing breeding programs to maximize the retention of valuable alleles. There are two variants on this approach that are worth considering. First, we could select for animals carrying alleles believed to be especially important, such as variants at the major histocompatibility complex loci (Hedrick 2002; Hughes 1991). Second, we could measure variation at random loci, and then preferentially breed those animals that appear to carry the rarest alleles. In theory, this approach could produce a population with even more gene diversity than was present in the wild population, by creating more equal allele frequencies than existed in the source population.

P1: ICD Trim: 7in × 10in Top: 0.375in Gutter: 1in
CUUS999-11 cuus999/DeWoody ISBN: 978 0 521 51564 1 February 26, 2010 21:18

Although these ideas deserve more evaluation, we would caution, as have others (Vrijenhoek & Leberg 1991), that there are some potential drawbacks. First, we know only few of the many loci that might be critical to individual fitness and population viability. If we select breeding pairs on the basis of those few loci about which we do know something, we are likely to cause rapid depletion of genetic variability at other loci that may be just as important (Lacy 2000; Hedrick 2001). Because the alleles that are advantageous will depend on the environment in which the animals live, many alleles that encode adaptations important in natural environments may be neutral or even deleterious in a specific captive environment.

A strategy of preferentially breeding animals that have the rarest alleles, without trying to prejudge which alleles will be most advantageous, has perhaps more merit than attempts to select the animal with superior alleles. Even this strategy has risks, however. Initially rare alleles may have been rare for a good reason. Selecting for them may increase frequencies of mutations that were deleterious in natural populations. We may be on safer ground if we use strategies that attempt to minimize the rate at which the populations in captivity diverge genetically from the genotypic composition of the wild populations from which they came, rather than trying to improve upon the results of the prior evolution in wild populations. If the divergence at sampled loci is indicative of genetic divergence across the genome as a whole, identifying individuals that carry unique or rare alleles at presumably neutral loci (e.g., microsatellites) also could be used as a means of selecting priority breeders that may be more likely to carry rare alleles at other loci throughout the genome. This selection of breeders is exactly the aim of incorporating molecular estimates of relatedness into pedigree analyses previously discussed, so there is little to be gained by otherwise or further targeting rare alleles for augmentation in a conservation breeding program.

Unresolved needs and unexploited opportunities

This chapter has focused on plausible methods for incorporating empirical estimates of relatedness into captive breeding programs. Areas that require further research have been mentioned throughout, but here we summarize the most critical research needs. There has been a moderate amount of work on how well genetic goals can be achieved by using various schemes for selecting priority breeding pairs (e.g., Willis 1993; Haig et al. 1994; Ballou & Lacy 1995; Willis 2001; Jones et al. 2002; Ivy et al. 2009). Much more work is needed, however, to evaluate the effectiveness of alternative methods for preserving genetic variation in captive populations, especially under various conditions of species biology, varying information availability, and rigor of management. At this time, managers of captive populations of endangered species have little guidance on which strategies to employ, particularly when it comes to incorporating molecular data into genetic management.

Incorporating molecular estimates of r *into genetic management*

The accuracies of *r* estimates need to be further explored, to identify how exact *r* estimates must be to improve, rather than degrade, genetic management. Thus,

P1: ICD Trim: 7in × 10in Top: 0.375in Gutter: 1in
CUUS999-11 cuus999/DeWoody ISBN: 978 0 521 51564 1 February 26, 2010 21:18

before estimates of *r* are incorporated into mean kinship calculations, appropriate precisions for those estimates must be defined. If estimates of *r* are to be used to place pairs of individuals into relationship categories, acceptable misclassification rates also must be identified. It is likely that different levels of *r*-estimate accuracy will be acceptable for different pedigrees, so a range of pedigrees that exhibit variable amounts of unknown ancestry should be considered when investigating these questions.

Assessing the role of selection

Almost all methods used to manage captive breeding programs were developed from genetic theory that presumes that genetic variation is neutral. The variation that is important to population persistence is, by definition, under selection (at least during some times), however. It is possible that calculations based on neutral theory provide accurate enough projections of genetic changes in captive populations, especially if the captive environment removes many of the selective pressures that sculpted genomes in the wild. Models that include plausible selective forces need to be examined, however. More needs to be known about the role of natural selection in captive populations (Frankham et al. 1986), with respect to both useful adaptation and removal of deleterious alleles (Arnold 1995; Lacy & Ballou 1998) and possibly harmful consequences for long-term population fitness and preservation of species characteristics (e.g., Frankham & Loebel 1992; Bryant & Reed 1999; Frankham et al 2000). Montgomery and coworkers (2009) recently reported that diversity at microsatellite loci is lost from managed captive populations more rapidly than expected based on neutral theory, presumably due to linkage to loci under selection. We need more information on the strength of natural selection in captive populations, the nature of the alleles favored or eliminated, and the rate of response under different breeding conditions.

Resources for quantitative genetic analysis

There is a largely untapped potential for captive populations of wild species to serve as a resource for studies of the genetic control and evolution of quantitative traits (Arnold 1995). Quantitative genetic partitioning of variance in traits has long been an important tool in identifying the genetic variation available for manipulation in domesticated species and in elucidating the past evolutionary forces in wild species (Falconer & Mackay 1996; Roff 1997; Lynch & Walsh 1998). Traditionally, such methods depended on specific crosses or correlations among classes of relatives, so most work has involved domesticated species or laboratory strains. It has been assumed that studies on such experimental populations that are removed from their wild ancestors by many generations of selective breeding will still reveal information about the structure of genetic variation in quantitative traits. Yet, thousands of species of wildlife are propagated in zoological parks, and some have pedigrees extending back five or more generations. Therefore, the raw material exists for quantitative genetic studies not just on guinea pigs, *Mus*, and *Drosophila*, but also on many other species representing most orders of mammals, birds, reptiles, and amphibians.

The pedigrees from most wildlife captive breeding programs are complex and, after the first generation or two, contain few cases of simple relationships (such

P1: ICD Trim: 7in × 10in Top: 0.375in Gutter: 1in
CUUS999-11 cuus999/DeWoody ISBN: 978 0 521 51564 1 February 26, 2010 21:18

as full-siblings and half-siblings) that are not also confounded by other paths of common ancestry. Therefore, many pedigrees from conservation breeding programs could provide only limited data for methods that depend on analyses of resemblance among discrete classes of relatives. With the development, however, of estimation procedures for using complex pedigrees to estimate variance components (Kruuk & Hill 2008), the pedigrees of zoo populations become a potentially vast source of data for such studies (Pelletier et al. 2009). These methods have been applied increasingly to wild populations, but the depth of pedigrees in such studies is rarely to the extent available from many captive populations. When such analyses depend on and presume accurate pedigrees, molecular genetic tools provide the means to fill in gaps, resolve uncertainties, and correct errors. As discussed earlier in text, however, the power of molecular data to reconstruct pedigrees with confidence is often limited (Ivy et al. 2009). The mostly complete pedigrees of many captive populations may provide a more useful starting point for such work than the largely unknown pedigrees of wild populations, which often must be constructed entirely from molecular data.

"Animal models" that use matrices of relatedness estimates but do not require reconstructed pedigrees are also now available (Frentiu et al. 2008), and molecular genetic data on samples from captive populations could be a valuable resource for such work. When populations with complete pedigrees are also extensively sampled for molecular genetic characterization, they would provide the opportunity to directly compare pedigree-based and pedigree-free animal models for estimating components of genetic variance (see Chapter 6 by Nichols & Neale).

SUMMARY

The most significant limitation of using pedigree-based analyses to demographically and genetically manage captive populations is that management can be severely hindered by inaccurate or incomplete pedigrees. Molecular data can improve the genetic management of captive breeding programs when pedigree-based management is ineffective, by helping to meet the standard genetic goals of retaining gene diversity and limiting the accumulation of inbreeding. The most basic use for molecular data in captive breeding programs is to resolve pedigree errors and gaps. For pedigrees with deep, unknown ancestries, empirical estimates of r also can be used to calculate kinships or to identify relationship categories for pairs of living individuals. Molecular data also could be used to guide the genetic management of group-living species through the empirical characterization of within- and between-group genetic structure, although we are unaware of any breeding programs that have yet attempted to do this. Although it is clear that there are many methods for incorporating molecular data into the management of captive breeding programs, additional research is still needed in a variety of areas to further clarify which strategies for incorporation are the most effective. Furthermore, it also should be noted that, for the many species that can be maintained and propagated in tightly controlled paired breeding, the recording of parentage information, pedigree calculations, and identification of optimal breeding pairs is simple, precise, and inexpensive, so it is unlikely that it will be

effective and efficient in those cases to replace traditional pedigree methods with programs guided by molecular analysis.

REFERENCES

Allendorf FW (1993) Delay of adaptation to captive breeding by equalizing family size. *Conservation Biology*, **7**, 416–419.

Arnold SJ (1995) Monitoring quantitative genetic variation and evolution in captive populations. In: *Population Management for Survival and Recovery: Analytical Methods and Strategies in Small Population Conservation* (eds. Ballou JD, Gilpin M, Foose TJ), pp. 293–317. Columbia University Press, New York.

Baillie JEM, Hilton-Taylor C, Stuart SN, editors (2004) *IUCN Red List of Threatened Species. A Global Species Assessment.* IUCN, Gland, Switzerland and Cambridge, UK.

Ballou JD, Foose TJ (1996) Demographic and genetic management of captive populations. In: *Wild Mammals in Captivity* (eds. Kleiman DG, Lumpkin S, Allen M, Harris H, Thompson K), pp. 263–283. University of Chicago Press, Chicago.

Ballou JD, Lacy RC (1995) Identifying genetically important individuals for management of genetic diversity in pedigreed populations. In: *Population Management for Survival and Recovery: Analytical Methods and Strategies in Small Population Conservation* (eds. Ballou JD, Gilpin M, Foose TJ), pp. 76–111. Columbia University Press, New York.

Ballou JD, Mickelberg J (2009) *International Population Management Plan for Ex-situ Golden Lion Tamarins*. National Zoological Park, Washington, DC.

Benirschke K (1984) The frozen zoo concept. *Zoo Biology*, **3**, 325–328.

Bink MCAM, Anderson AD, van de Weg WE, Thompson EA (2008) Comparison of marker-based pairwise relatedness estimators on a pedigreed plant population. *Theoretical and Applied Genetics*, **117**, 843–855.

Blouin MS, Parsons M, Lacille V, Lotz S (1996) Use of microsatellite loci to classify individuals by relatedness. *Molecular Ecology*, **5**, 393–401.

Bryant EH, Reed DH (1999) Fitness decline under relaxed selection in captive populations. *Conservation Biology*, **13**, 665–669.

Butchart SHM, Stattersfield AJ, Baillie J et al. (2005) Using Red List Indices to measure progress towards the 2010 target and beyond. *Philosophical Transactions of the Royal Society of London-Series B: Biological Sciences*, **360**, 255–268.

Cosson L, Grassman LL Jr, Zubaid A et al. (2007) Genetic diversity of captive binturongs (Arctictis binturong, Viverridae, Carnivora): implications for conservation. *Journal of Zoology*, **271**, 386–395.

Csillery K, Johnson T, Beraldi D et al. (2006) Performance of marker-based relatedness estimators in natural populations of outbred vertebrates. *Genetics*, **173**, 2091–2101.

DeWoody JA (2005) Molecular approaches to the study of parentage, relatedness, and fitness: practical applications for wild animals. *Journal of Wildlife Management*, **69**, 1400–1418.

Earnhardt JM, Thompson SD, Marhevsky E (2001) Interactions of target population size, population parameters, and program management of viability of captive populations. *Zoo Biology*, **20**, 169–183.

Ely JJ, Dye B, Frels WI et al. (2005) Subspecies composition and founder contribution of the captive U.S. chimpanzee (Pan troglodytes) population. *American Journal of Primatology*, **67**, 223–241.

Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics*, 4th edn. Pearson/Prentice, Harlow, UK.

Fernandez J, Toro MA (1999) The use of mathematical programming to control inbreeding in selection schemes. *Journal of Animal Breeding and Genetics*, **116**, 447–466.

Field D, Chemnick L, Robbins M, Garner K, Ryder OA (1998) Paternity determination in captive lowland gorillas and orangutans and wild mountain gorillas by microsatellite analysis. *Primates*, **39**, 199–209.

Foose TJ, Lande R, Flesness NR, Rabb G, Read B (1986) Propagation plans. *Zoo Biology*, **5**, 139–146.

P1: ICD Trim: 7in × 10in Top: 0.375in Gutter: 1in
CUUS999-11 cuus999/DeWoody ISBN: 978 0 521 51564 1 February 26, 2010 21:18

Frankham R, Hemmer H, Ryder OA et al. (1986) Selection in captive environments. *Zoo Biology*, **5**, 127–138.

Frankham R, Loebel DA (1992) Modeling problems in conservation genetics using captive Drosophila populations: rapid genetic adaptation to captivity. *Zoo Biology*, **11**, 333–342.

Frankham R, Manning H, Margan SH, Briscoe DA (2000) Does equalisation of family sizes reduce genetic adaptation to captivity?. *Animal Conservation*, **3**, 357–363.

Frentiu FD, Clegg SM, Chittock J et al. (2008) Pedigree-free animal models: the relatedness matrix reloaded. *Proceedings of the Royal Society of London-Series B: Biological Sciences*, **275**, 639–648.

Gascon C, Collins JP, Moore RD et al. (2007) *Amphibian Conservation Action Plan*. IUCN/SSC Amphibian Specialist Group. Gland, Switzerland.

Gautschi B, Jacob G, Negro JJ et al. (2003) Analysis of relatedness and determination of the source of founders in the captive bearded vulture, Gypaetus barbatus, population. *Conservation Genetics*, **4**, 479–490.

Glaubitz JC, Rhodes EO Jr, DeWoody JA (2003) Prospects for inferring pairwise relationships with single nucleotide polymorphisms. *Molecular Ecology*, **12**, 1039–1047.

Haig SM, Ballou JD, Casna NJ (1994) Identification of kin structure among Guam rail founders: a comparison of pedigrees and DNA profiles. *Molecular Ecology*, **3**, 109–119.

Haldane JBS (1924) A mathematical theory of natural and artificial selection. Part I. *Transactions of the Cambridge Philosophical Society*, **23**, 10–41.

Hedrick PW (2001) Conservation genetics: where are we now?*Trends in Ecology and Evolution*, **16**, 629–636.

Hedrick PW (2002) The importance of the major histocompatibility complex in declining populations. In: *Reproduction and Integrated Conservation Science* (eds. Wildt DE, Holt B). Cambridge University Press, Cambridge.

Hedrick PW, Miller PS, Geffen E, Wayne R (1997) Genetic evaluation of the three captive Mexican wolf lineages. *Zoo Biology*, **16**, 47–69.

Hu J, Pan H, Wan Q, Fang S (2007) Nuclear DNA microsatellite analysis of genetic diversity in captive populations of Chinese water deer. *Small Ruminant Research*, **67**, 252–256.

Hughes AL (1991) MHC polymorphism and the design of captive breeding programs. *Conservation Biology*, **5**, 249–251.

International Union for Conservation of Nature (IUCN) (2008) http://www.iucnredlist.org.

Ivy JA, Miller A, Lacy RC, DeWoody JA (2009) Methods and prospects for using molecular data in captive breeding programs: an empirical example using parma wallabies (Macropus parma). *Journal of Heredity*, **100**, 441–454.

Iyengar A, Gilbert T, Woodfine T et al. (2007) Remnants of ancient genetic diversity preserved within captive groups of scimitar-horned oryx (Oryx dammah). *Molecular Ecology*, **16**, 2436–2449.

Jones AG, Ardren WR (2003) Methods of parentage analysis in natural populations. *Molecular Ecology*, **12**, 2511–2523.

Jones KL, Glenn TC, Lacy RC et al. (2002) Refining the whooping crane studbook by incorporating microsatellite DNA and leg banding analyses. *Conservation Biology*, **16**, 789–799.

Kimura M, Crow JF (1963) On the maximum avoidance of inbreeding. *Genetical Research*, **4**, 399–415.

King JL (1965) The effect of litter culling–or family planning–on the rate of natural selection. *Genetics*, **51**, 425–429.

Kruuk LEB, Hill WG (2008) Introduction. Evolutionary dynamics of wild populations: the use of long-term pedigree data. *Proceedings of the Royal Society of London-Series B: Biological Sciences*, **275**, 593–596.

Lacy RC (1994) Managing genetic diversity in captive populations of animals. In: *Restoration of Endangered Species* (eds. Bowles ML, Whelan CJ), pp. 63–89. Cambridge University Press, Cambridge.

Lacy RC (1995) Clarification of genetic terms and their use in the management of captive populations. *Zoo Biology*, **14**, 565–577.

P1: ICD Trim: 7in × 10in Top: 0.375in Gutter: 1in
CUUS999-11 cuus999/DeWoody ISBN: 978 0 521 51564 1 February 26, 2010 21:18

Lacy RC (2000) Should we select genetic alleles in our conservation breeding programs? *Zoo Biology*, **19**, 279–282.

Lacy RC, Ballou JD (1998) Effectiveness of selection in reducing the genetic load in populations of *Peromyscus polionotus* during generations of inbreeding. *Evolution*, **52**, 900–909.

Lande R (1995) Breeding plans for small populations based on the dynamics of quantitative genetic variance. In: *Population Management for Survival and Recovery: Analytical Methods and Strategies in Small Population Conservation* (eds. Ballou JD, Gilpin M, Foose TJ), pp. 318–340. Columbia University Press, New York.

Lynch M, Ritland K (1999) Estimation of pairwise relatedness with molecular markers. *Genetics*, **152**, 1753–1766.

Lynch M, Walsh B (1998) *Genetic analysis of quantitative traits*. Sinauer, Sunderland, MA.

Milligan BG (2003) Maximum-likelihood estimation of relatedness. *Genetics*, **163**, 1153–1167.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences USA*, **70**, 3321–3323.

Norton JE, Ashley MV (2004) Genetic variability and population differentiation in captive Baird's tapirs (*Tapirus bairdii*). *Zoo Biology*, **23**, 521–531.

Pelletier F, Réale D, Watters J, Boakes EH, Garant D (2009) Value of captive populations for quantitative genetics research. *Trends in Ecology and Evolution*, **24**, 263–270.

Petric A, Long S (2008) *Population Analysis and Breeding plan. Okapi (Okapia johnstoni) Species Survival Plan*. Association of Zoos and Aquariums, Silver Spring, MD.

Petrie M, Moller AP (1991) Laying eggs in others' nests: intraspecific brood parasitism in birds. *Trends in Ecology and Evolution*, **6**, 315–320.

Piper WH, Rabenold PP (1992) Use of fragment-sharing estimates from DNA fingerprinting to determine relatedness in a tropical wren. *Molecular Ecology*, **1**, 69–78.

Princée FPG (1995) Overcoming the constraints of social structure and incomplete pedigree data through low-intensity genetic management. In: *Population Management for Survival and Recovery: Analytical Methods and Strategies in Small Population Conservation* (eds. Ballou JD, Gilpin M, Foose TJ), pp. 124–154. Columbia University Press, New York.

Queller DC, Goodnight KF (1989) Estimating relatedness using genetic markers. *Evolution*, **43**, 258–275.

Ramirez O, Altet L, Ensenat C et al. (2006) Genetic assessment of the Iberian wolf *Canis lupus signatus* captive breeding program. *Conservation Genetics*, **7**, 861–878.

Reddy A, Prakash V, Shivaji S (2007) A rapid, non-invasive, PCR-based method for identification of sex of the endangered Old World vultures (white-backed and long-billed vultures) – implications for captive breeding programmes. *Current Science*, **92**, 659–662.

Ritland K (1996) Estimators for pairwise relatedness and individual inbreeding coefficients. *Genetical Research*, **67**, 175–185.

Roff DA (1997) *Evolutionary quantitative genetics*. Chapman & Hall, New York.

Rudnick JA, Lacy RC (2008) The impact of assumptions about founder relationships on the effectiveness of captive breeding strategies. *Conservation Genetics*, **9**, 1439–1450.

Ruokonen M, Andersson A, Tegelstrom H (2007) Using historical captive stocks in conservation: the case of the lesser white-fronted goose. *Conservation Genetics*, **8**, 197–207.

Russello MA, Hyseni C, Gibbs JP et al. (2007) Lineage identification of Galapagos tortoises in captivity worldwide. *Animal Conservation*, **10**, 304–311.

Schad K, editor (2008) *Amphibian Population Management Guidelines*. Amphibian Ark Amphibian Population Management Workshop; December 10–11, 2007; San Diego, CA. Amphibian Ark, www.amphibianark.org.

Selvin S (1980) Probability of nonpaternity determined by multiple allele codominant systems. *American Journal of Human Genetics*, **32**, 276–278.

Sonesson AK, Meuwissen THE (2001) Minimization of rate of inbreeding for small populations with overlapping generations. *Genetical Research*, **77**, 285–292.

Soule M, Gilpin M, Conway W, Foose T (1986) The millennium ark: how long a voyage, how many staterooms, how many passengers? *Zoo Biology*, **5**, 101–113.

Spinks PQ, Shaffer HB (2007) Conservation phylogenetics of the Asian box turtles (Geoemydidae, Cuora): mitochondrial introgression, numts, and inferences from multiple nuclear loci. *Conservation Genetics*, **8**, 641–657.

Thompson EA (1995) Genetic importance and genomic descent. In: *Population Management for Survival and Recovery: Analytical Methods and Strategies in Small Population Conservation* (eds. Ballou JD, Gilpin M, Foose TJ), pp. 76–111. Columbia University Press, New York.

Tzika AC, Remy C, Gibson R, Milinkovitch MC (2009) Molecular genetic analysis of a captive-breeding program: the vulnerable endemic Jamaican yellow boa. *Conservation Genetics*, **10**, 69–77.

Van de Casteele T, Galbusera P, Matthysen E (2001) A comparison of microsatellite-based pairwise relatedness estimators. *Molecular Ecology*, **10**, 1539–1549.

Vrijenhoek RC, Leberg PL (1991) Let's not throw out the baby with the bathwater: a comment on management for MHC diversity in captive populations. *Conservation Biology*, **5**, 252–254.

Wang J (2003) Maximum likelihood estimation of admixture proportions from genetic data. *Genetics*, **164**, 747–765.

Wang, J (2004) Monitoring and managing genetic variation in group breeding populations without individual pedigrees. *Conservation Genetics*, **5**, 813–825.

Wang JL (2002) An estimator for pairwise relatedness using molecular markers. *Genetics*, **160**, 1203–1215.

Willis K (1993) Use of animals with unknown ancestries in scientifically managed breeding programs. *Zoo Biology*, **12**, 161–172.

Willis K (2001) Unpedigreed populations and worst-case scenarios. *Zoo Biology*, **20**, 305–314.

Yom-Tov Y (1980) Intraspecific nest parasitism in birds. *Biological Reviews of the Cambridge Philosophical Society*, **55**, 93–108.

Zhang YP, Ryder OA, Zhao QG et al. (2005) Non-invasive giant panda paternity exclusion. *Zoo Biology*, **13**, 569–573.

Zink AG (2000) The evolution of intraspecific brood parasitism in birds and insects. *American Naturalist*, **155**, 395–405.

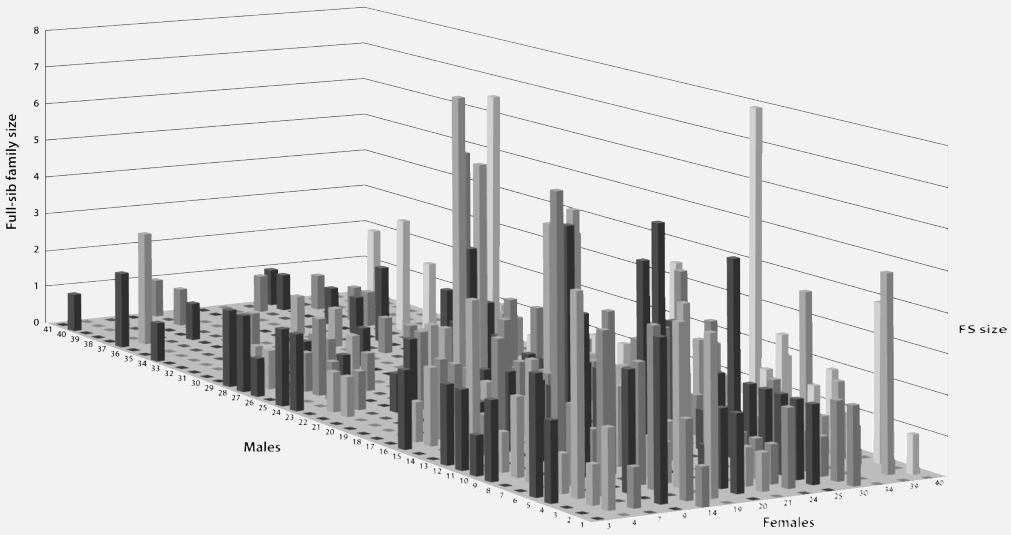
CHAPTER 11 BOX: PEDIGREE RECONSTRUCTION: AN ALTERNATIVE TO SYSTEMATIC BREEDING

Yousry A. El-Kassaby

Problem

Forest tree breeding programs follow the classical recurrent selection scheme starting with phenotypic selection, followed by breeding and testing. Ultimately the cycle is completed by genotypic selection of elite individuals for either starting a new breeding cycle or establishing seed orchards for the production of genetically improved seed for reforestation. Trees' long life cycle, delayed reproductive maturity, and geographically extensive and protracted testing make breeding a daunting task. Increasing programs' efficiency by either eliminating the breeding phase or simplifying testing would be of immense value. Here we illustrate the combined use of DNA fingerprinting and pedigree reconstruction to assemble mating designs (half- and full-sibling [HS and FS] families) from natural pollination and to demonstrate the method's utility in simplifying progeny testing.

P1: ICD Trim: 7in × 10in Top: 0.375in Gutter: 1in
CUUS999-11 cuus999/DeWoody ISBN: 978 0 521 51564 1 February 26, 2010 21:18



Box Figure 11–1–1. Three-dimensional diagram showing the formation of FS within HS families.

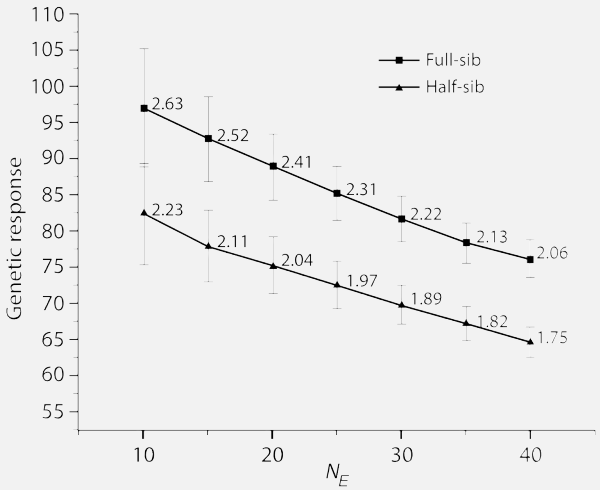
Case Study

The combined use of DNA fingerprinting and partial pedigree reconstruction is employed to assemble a mating design using 551 wind-pollinated western larch (*Larix occidentalis* Nutt) seedlings representing fourteen seed-donors from a forty-one-parent seed orchard (Funda et al. 2008). The assembled mating design encompassed 221 FS families nested within the maternal (mean: 15.8; range: 12–23) and paternal (mean: 5.4; range: 0–14) HS families (Box Fig. 11–1–1). This mating design, with its multiple crosses across the mating landscape, offers better genetic sampling than those produced from traditional mating designs, which restrict crossing within independent subsets of parents.

Progeny testing can be simplified if it is reduced to HS families using wind-pollinated seed and if FS families are assembled using DNA fingerprinting and pedigree reconstruction (Box Fig. 11–1–1). A conventional Douglas-fir (*Pseudotsuga menziesii* (Mirb.) Franco) FS progeny test (60 parents producing 150 FS and 50 HS families) was treated as HS, and pedigree reconstruction was implemented to assemble FS and genetic gain from the two approaches (conventional FS vs. reconstructed FS from HS) was compared. The small genetic gain difference between the FS (100%) and HS (85%) testing (Box Fig. 11–1–2) highlights the efficiency of the proposed approach, specifically at effective population sizes similar to those used in seed-orchard establishment (Box Fig. 11–1–2; $> N_E = 20$) (El-Kassaby & Lstiburek 2008).

This example demonstrates the benefits of integrating molecular marker technology with advanced pedigree-reconstruction models and modern quantitative genetics methods in transforming traditional forest tree breeding to molecular and genomics-based breeding. Additionally, pedigree reconstruction has important implications in understanding the extent of genetic diversity and relatedness among members of natural and domesticated populations. This understanding is an essential component to their effective utilization and conservation.

P1: ICD Trim: 7in × 10in Top: 0.375in Gutter: 1in
CUUS999-11 cuus999/DeWoody ISBN: 978 0 521 51564 1 February 26, 2010 21:18



Box Figure 11–1–2. Rate of genetic response with its corresponding effective population size for FS and HS selection methods.

REFERENCES

El-Kassaby YA, Lstibůrek M (2008) Breeding without breeding. *Genetics Research*, **91**, 111–120.

Funda T, Chen C, Liewlaksaneeyanawin C, Kenawy AMA, El-Kassaby YA (2008) Pedigree and mating system analyses in a western larch (Larix occidentalis Nutt) experimental population. *Annals of Forest Science*, **65**, 750.