

HIERARCHICAL ANALYSIS OF INBREEDING DEPRESSION IN *PEROMYSCUS POLIONOTUS*

ROBERT C. LACY,¹ GLEN ALAKS, AND ALLISON WALSH

Department of Conservation Biology, Daniel F. and Ada L. Rice Center, Brookfield Zoo, Brookfield, Illinois 60513

¹E-mail: rlacy@ix.netcom.com

Abstract.—The severity of inbreeding depression appears to vary among taxa, but few ecological or other patterns have been identified that predict accurately which taxa are most sensitive to inbreeding. To examine the causes of heterogeneity in inbreeding depression, the effects of inbreeding on reproduction, survival, and growth were measured in three replicate experimental stocks for each of three subspecies of *Peromyscus polionotus* mice. Inbreeding of the dam reduced the probability of breeding, the probability of producing a second litter, and litter size. Inbreeding of the litter caused depression of litter size, juvenile viability, and mass at weaning, and caused an increase in the within-litter variance in mass. In spite of differences between the subspecies in natural population sizes, genetic variation, and mean rates of reproduction and survival, all variation observed between experimental populations in their responses to inbreeding could be attributed to random founder effects. The genetic load of deleterious alleles in each replicate was unequally partitioned among its founder pairs, and different founders contributed to the load affecting different fitness components. Thus, inbreeding depression for any one fitness component, in our experimental environment, must be due to relatively few deleterious alleles with major effects. Genetic loads so comprised would be expected to diverge among natural populations due to both random drift and selective removal of recessive deleterious alleles during population bottlenecks. The near universality of inbreeding depression would be maintained, however, if different alleles contribute to inbreeding depression of different fitness components and in different environments.

Key words.—Founder effect, genetic load, inbreeding depression, *Peromyscus polionotus*.

Received July 14, 1995. Accepted June 28, 1996.

Inbreeding has been reported to cause reduction in a wide range of fitness and performance measures, including fertility, growth rates, stability of development, survival, disease resistance, intraspecific competitive ability, and tolerance of environmental stresses (Darwin 1868, 1876; Wright 1977; Falconer 1989). Significant inbreeding depression has been described in most species of sexually reproducing animals that have been studied with sufficient sample sizes to allow detection of moderate inbreeding depression. An absence of inbreeding depression would indicate either that there was no genetic variation for fitness or that alleles causing lower fitness were not, on average, recessive. Among mammals, most reported cases of no significant inbreeding depression (e.g., chacma baboons studied by Bulger and Hamilton [1988]; some taxa in Ralls et al. [1988]; prairie dogs studied by Hoogland [1992]) are based on sample sizes too small or inbreeding levels too low to allow statistical detection of any but abnormally strong inbreeding depression. Although depression of fitness resulting from inbreeding appears to be nearly universal, the severity of the inbreeding depression and the fitness components that are impacted seem to be variable (Lacy 1993). For example, the review by Ralls et al. (1988) of juvenile mortality in 40 mammal taxa found the cost of inbreeding, quantified by the number of “lethal equivalents” (Morton et al. 1956), to range from -0.68 (weak, but nonsignificant, benefit from inbreeding) to 15. In a study of eight populations of *Peromyscus*, Brewer et al. (1990) found some populations to show inbreeding depression in litter size, others to show depression in survival, others to show depression in growth, and one to show no significant inbreeding depression in any of these fitness components.

Any population with genetic variation in fitness carries, by definition, a “genetic load”—the reduction in mean fitness relative to the most adaptive genotype. Recessive deleterious

alleles can accumulate in large, outbreeding populations (constituting a “mutational load”) because of recurrent mutation and protection from natural selection in heterozygotes. At loci with heterozygote superiority, balancing selection will maintain a “segregational load” of genetic variation (Crow 1993). Most studies suggest that inbreeding depression is predominantly caused by the expression of the mutational load of deleterious recessive alleles (the “dominance hypothesis,” Crow 1948, 1952; Charlesworth and Charlesworth 1987; Johnston and Schoen 1995), but exposure of the segregational load could also contribute to inbreeding depression (the “overdominance hypothesis,” Crow 1948, 1952; Mitton 1993).

Various possible causes of variation in the effect of inbreeding have been hypothesized. The mutational load could be purged by natural selection when populations experienced episodes of inbreeding, perhaps during bottlenecks of local populations (Lande and Schemske 1985; Charlesworth and Charlesworth 1987; Hedrick and Miller 1992). The segregational load could also be partly reduced during population bottlenecks, because genetic drift in small, inbreeding populations can result in the fixation of alleles and permanent loss of heterozygote advantage. In spite of some support from data on human populations (Rao and Inbaraj 1977) and plants (e.g., Barrett and Charlesworth 1991; Nason and Ellstrand 1995), however, taxonomic surveys have not supported well the expectation that populations that have experienced bottlenecks would be less affected by inbreeding. For example, the more endangered mammal taxa (e.g., golden lion tamarins, *Leontopithecus rosalia*; Pere David’s deer, *Elaphurus davidianus*; Eld’s deer, *Cervus eldi thamin*; scimitar-horned oryx, *Oryx dammah*; Speke’s gazelle, *Gazella spekei*) listed in Ralls et al. (1988) showed as much inbreeding depression in juvenile survival as did the common species. Ballou (1995)

found that the cost of inbreeding in some mammal taxa could be only partly reduced by generations of inbreeding in captivity. Small insular populations of *Peromyscus* mice studied by Brewer et al. (1990) expressed genetic loads under inbreeding at least as great as did the larger central populations. In birds, an island population of great tit did show lesser effects of inbreeding on overall fitness than did larger populations (van Noordwijk and Scharloo 1981).

In plants, Levin (1984) found peripheral populations of a phlox to have smaller genetic loads than central populations, but taxa of limited distribution have been reported to show no lesser effects of inbreeding than widespread taxa (Karron 1989; Widén 1993; Bijlsma et al. 1994). Nason and Ellstrand (1995) studied inbreeding in wild radish, an introduced, self-incompatible annual that likely went through population bottlenecks at the time of colonization and during subsequent range expansion. They found inbreeding depression in maternal fecundity, but little evidence that highly deleterious alleles depressed seed germination, seedling viability, plant biomass, or other components of fitness, suggesting that strongly deleterious alleles had been purged during genetic bottlenecks. However, even species that often self can suffer as much depression in fitness when inbred as do plants that normally outcross (Barrett and Kohn 1991).

The efficiency with which natural selection can reduce the mutational load depends on the underlying nature of that load. Fully recessive lethal alleles are purged most readily, but alleles with lesser impacts on fitness may often be fixed during inbreeding rather than purged (Hedrick 1994). Experiments with *Drosophila* indicate that about half of the genetic load with respect to viability results from recessive lethal alleles (Simmons and Crow 1977). Johnston and Schoen (1995) found that most of the inbreeding depression in total fitness of two self-fertilizing plants was caused by new, incompletely recessive, deleterious mutations. It is not known whether the numbers and kinds of genes contributing to inbreeding depression are consistent across fitness components (fecundity and viability) or taxa. To date, few taxonomic or ecological trends have emerged from empirical data that might indicate the determinants of variation in inbreeding depression among natural populations.

Variation in the responses to inbreeding can result simply from the action of random drift on genetic variation among individuals. In a theoretical analysis, Schultz and Willis (1995) showed that random Poisson variation of deleterious alleles among individuals can be the major cause of variation in inbreeding depression. Experimental studies on crop plants, laboratory rodents, and *Drosophila* (reviewed by Wright 1977), and *Tribolium* (Pray et al. 1994; Pray and Goodnight 1995) have demonstrated that lines started from randomly selected founder pairs can show highly variable responses to inbreeding. However, all of these previous experimental studies used domesticated or laboratory stocks that had already undergone many generations of strong artificial selection, including, in most cases, selection for survival and fecundity under inbreeding. Random genetic variation will contribute to variation in inbreeding depression whenever experimental or natural populations are founded with relatively few individuals, but the extent to which such founder effects might cause the variation in inbreeding de-

pression reported among taxa and natural populations has not been determined.

Depression of fitness under inbreeding is therefore nearly ubiquitous, yet perhaps highly variable. As a step toward understanding the causes of variation in inbreeding depression, we have undertaken examination of the impacts of inbreeding on several components of fitness in populations of *Peromyscus* mice, with the focus on the predictability of inbreeding depression among taxa, among local populations, and among lineages within populations. In previous papers, Brewer and colleagues (Brewer et al. 1990; Lacy 1992) presented analyses of inbreeding depression in eight populations of *Peromyscus* mice (in five subspecies of *P. polionotus* and two subspecies of *P. leucopus*). High variability but no trends were identified at the subspecies level. In this paper, we present an analysis of the predictability of inbreeding depression at a finer grain—variation among lab populations created by replicate sampling of three subspecies of *Peromyscus polionotus*, and variation among founder lineages. The patterns reported here suggest that much of the variation among studies in the severity of inbreeding depression might have methodological rather than biological causes. However, the mechanisms of inbreeding depression suggested by our findings would be expected to lead to strong and predictable divergence in genetic loads among populations with different histories of population size, in contrast to the almost universality of the phenomenon of inbreeding depression and the lack of clear patterns among natural populations in the cost of inbreeding.

MATERIALS AND METHODS

Subjects comprised laboratory stocks of three subspecies of the beach mouse or old-field mouse, *Peromyscus polionotus*. Mice were trapped in Sherman live-traps and returned to the research animal facilities of the Brookfield Zoo (Brookfield, Illinois). *Peromyscus polionotus subgriseus* were collected in April, May, June, and December 1990 from the Ocala National Forest in Marion County, north-central Florida. The trapping sites were within a 10 x 5 km area in a mosaic of pine forest and clear cuts in the Hopkins Prairie Management Unit. The old-field mice were collected in open clear-cut habitat and along the sides of the dirt roads connecting the clear-cut areas. Of 128 *P. p. subgriseus* collected, 66 were bred in the laboratory, and the progeny of 52 of these were randomly assigned to three replicate laboratory stocks. *Peromyscus polionotus rhoadsi* were collected in April, May, June, and December 1990 and May 1991 from a contiguous 4 km² area of open pine-scrub habitat 5 km south of the center of the town of Lake Placid, Highlands County, Florida, and along a roadside 2 km north of the center of the town. Of 109 *P. p. rhoadsi* collected, 80 were bred in the laboratory, and the progeny of 52 of these were assigned to three replicate laboratory stocks. *Peromyscus polionotus leucocephalus* were collected in June 1990 and June and July 1991 from a section of Eglin Air Force Base on Santa Rosa Island, Florida. Trapping was conducted along a continuous 9 km stretch of dunes on the south side of the island. Of 83 *P. p. leucocephalus* collected, 52 were bred in the laboratory, and the progeny of 50 of these were assigned to three replicate

laboratory stocks. Thus, each replicate stock descended from a discrete, random set of wild-caught founders.

Laboratory stocks were housed in a common room, maintained at a mean temperature of 21°C (range 20–24°C), under a 12:12 L:D photoperiod. Mice were housed in 18 cm × 28.5 cm × 12.5 cm deep plexiglas cages, with corn cob bedding and ad libitum water and Agway Prolab mouse food. Mice selected for experimental breeding were virgins, in adult pelage, and usually between 60 and 160 d of age at the time of pairing. Cages were checked daily for new litters. On the day of parturition, the pups were counted, and pairs were provided with cotton for constructing nests. Litters were weaned at 20 d of age, at which time they were weighed (to 0.1 g), ear-punched with identifying codes, and housed with same-sex littermates. At the time of weaning of the first litter produced by a pair, sires and dams were separated. Most pairs had bred following the birth of the first litter, producing a second litter 3–9 d after weaning the first litter. Thus, each experimental pair was used to produce at most two litters. Pairs not having produced a litter were separated 63 d after pairing.

For each replicate lab stock, in each of eight to 10 generations, some mice of low inbreeding coefficients ($f < 0.10$) were paired to produce similarly noninbred progeny; some noninbred but related mice were paired to produce inbred litters; some inbred but unrelated mice were paired to produce noninbred litters; and some inbred and related mice were paired to produce more highly inbred progeny. Inbred pairings often consisted of full-sib or half-sib matings, but littermates were never paired for breeding. The breeding design was not regular: we chose to model the sort of complex pedigree that could develop in small natural populations (albeit with frequent close inbreeding). As a result, the highest inbreeding coefficients achieved ($f = 0.375$ to $f = 0.594$, depending on the replicate) were much lower than would have been produced from line-breeding of first-order relatives. Another difference between our breeding program and a line-breeding design is that natural selection against deleterious alleles would occur through the loss of individuals in a constantly recombining gene pool, rather than primarily through the loss of family lines that do not exchange genes with other lineages.

Pairings were selected from matrices of kinship coefficients to produce a range of inbreeding levels at each generation. Some progeny from each pairing were used as breeders for the subsequent generation, and most families contributed mice to inbred and to noninbred pairings. An attempt was made to set up 40–50 pairs per replicate each generation, but poorer production by *P. p. leucocephalus* reduced sample sizes of those stocks relative to the other two subspecies. Matings of wild-caught mice and additional matings of non-virgin mice for stock maintenance were not included in the data analyses.

Six aspects of reproductive performance were monitored: the proportion of pairs producing litters conceived within the up to 63 d that each pair was kept together (P[breed]); the proportion of breeding pairs (those that produced at least one litter and were housed together through the rearing of the first litter) that rebred during the postpartum estrous and produced a second litter (P[2nd litter]); the number of pups born in each litter (litter size); the survival of pups from birth to

weaning (viability); the mass of pups at weaning (mass); and the within-litter variance in mass at weaning (for those litters with two or more offspring weaned), expressed as a standard deviation (SD[mass]). The fates of pups within litters are not independent. Therefore, each litter rather than each offspring was considered as an independent data point for analysis. Viability was analyzed as a categorical trait, and was assigned a value of 1 if more than half of the progeny of a litter survived, 0 otherwise. Mass at weaning was averaged across the surviving progeny per litter. To evaluate the combined effects of fertility, fecundity, infant survival, and growth, overall reproductive success (RS) was assessed as the total mass of offspring weaned for the zero, one, or two litters produced by each pair.

In addition to the inbreeding coefficients of the litter and the dam, independent variables tested as potential predictors of reproductive performance were the age of the dam, the parity of the dam (first versus second litter), and the sex ratio of the litter. Sires were present during the rearing of the first litter for each pair, but were not present while the second litter was being reared. Therefore, the effect of “parity” includes both the effect due to a litter being the first versus second litter of a dam, and any effect due to the presence vs. absence of the sire while the litter is reared.

The measures of reproductive performance take on various distributions and, therefore, require different statistical approaches. The probability of pairs producing at least one litter (P[breed]), the probability of those breeding pairs producing a second litter (P[2nd litter]), and viability were scored as 0/1 categorical responses and are presumed to result from binomial processes. (When viability is examined as a proportion of pups surviving, the distribution is strongly bimodal. In the majority of litters either all pups survived or all died.) The responses of these variables to inbreeding of the litter (f -litter) and the dam (f -dam), and to parity of the dam, were assessed with logistic multiple regression models, fitted by maximum likelihood estimation. Logistic regression is appropriate for predicting response by a categorical (binomial or multinomial) dependent variable to changes in continuous or categorical predictor variables (Hosmer and Lemeshow 1989). Logistic regression fits data to models of the form: $P = e^{a + bx} / (1 + e^{a + bx})$, in which P is the probability or expectation of the dependent variable (e.g., the probability of survival of a litter), x is the array of independent variables (f -litter, f -dam, and parity), and a and b are fitted regression coefficients. Significance of predictors is determined by a chi-square test on -2 times the difference in the log-likelihood ratios for the model with and the model without the effect.

The distributions of the other components of reproductive performance (litter size, mass, and SD[mass]) were continuous and approximately normal. The distribution of the total mass of offspring weaned per pair (RS) deviated considerably from normality. However, the effects of inbreeding on RS were strong and are biologically important. The responses of these continuous variables to inbreeding and to other predictors were assessed by least-squares linear multiple regressions. For both logistic and linear regression models, we first examined full models with all main effects and interactions. Across all models, only two of 45 two-way inter-

action effects (and no three-way interactions) between parity, f -litter, and f -dam were significant at the $P < 0.05$ level. Therefore, these possibly spurious interaction effects were omitted from the analyses.

Differences among subspecies were examined by including subspecies as a categorical predictive variable in the models, and testing (via least-squares linear regression or maximum likelihood logistic regression) for significance of including subspecies as a main effect (for differences in elevation of regression lines) or as an interaction with other effects (differences in slope among replicates). Differences among replicates within subspecies were similarly examined, by testing the significance of replicate nested within subspecies as a categorical main effect and in interactions with other predictors.

To examine whether the alleles contributing to inbreeding effects were distributed uniformly across founder genomes or were descended from specific founders, the inbreeding coefficient of each pairing was partitioned into components due to each founder of the lab stocks. Each of these partial inbreeding coefficients (designated f_i , for the inbreeding attributed to founder i) measures the probability that an individual is homozygous (identical by descent) for an allele descended from the specified founder. The sum, across all founders, of the partial inbreeding coefficients for a descendant is equal to the overall inbreeding coefficient for that individual. The partial inbreeding coefficients, f_i , for all descendants of founder i can be readily calculated by using the additive matrix method for calculating inbreeding coefficients (Ballou 1983) but assigning 0 to the kinship of each other founder to all individuals in the pedigree.

Most founders were bred as monogamous pairs, thereby linking partial inbreeding coefficients of the paired founders throughout the pedigree. Partial inbreeding coefficients of linked founders were combined to yield metrics giving the probability of identity by descent of alleles from the founder pair. In addition, a few founders were repaired during the production of laboratory stocks, when a founder was captured pregnant (so that her first captive-born litter had a different sire than all subsequent litters), or when a founder in a pair died early and was replaced by another wild-caught mate. These pairing changes led to partial linkage of the descendants of a few groups of three founders (e.g., a female and her several mates). Partial inbreeding coefficients were combined for analysis whenever the correlation of the partial inbreeding coefficients exceeded $r = 0.50$ in the descendant pedigree. Correlations of founder ancestry in the descendants was always less than 0.42 (in which cases the f_i effects were analyzed individually) or greater than 0.64 (f_i s grouped for analysis).

For each measure of reproductive performance, regression models were examined in which the set of partial inbreeding coefficients replaced f -litter as predictors. (That is, models of the form $Y = a + bf$ were expanded to $Y = a + b_1f_1 + b_2f_2 + \dots + b_8f_8$, with $f = \sum f_i$.) Significance of differences in inbreeding effects among founder lineages was tested by comparing the additional variance explained with this partitioning of the inbreeding coefficient to the residual error variance (for linear regression models), or by the increase in the log-likelihood ratio when inbreeding was partitioned (for

logistic regression models). The effects of partitioning inbreeding coefficients among founder groups was tested within each replicate, and then pooled across replicates and subspecies for an overall test of the heterogeneity of genetic loads across founders.

Statistical analyses were conducted with the SYSTAT (Wilkinson 1990) computer package, using the STATS module for examination of distributions, the GLM module for parametric multiple regressions, and the LOGIT module for logistic regressions. The multiple regressions used Type III sums of squares (SS), in which the effect attributed to each factor is partitioned from the residuals after all other factors have been included in the model. This prevents possible misattribution of an effect to one factor when it is caused by another factor in the model with which the first is correlated. The disadvantage of Type III SS is that when two factors contribute interchangeably to an effect, then neither factor may be recognized as significant because the effect is removed by adjustment for the other factor before each is tested (Shaw and Mitchell-Olds 1993). Because sample sizes varied among replicates and subspecies, the Satterthwaite approximation (Sokal and Rohlf 1981, pp. 293–308) was used to adjust the denominator mean square and df of nested hierarchical F -tests when subspecies level variation was tested against variation among replicates within subspecies. Levels of significance reported in the tables are for individual tests. In each table, some of the effects with single-test significance levels of $0.01 < P < 0.05$ are nonsignificant when sequential Bonferroni inequalities (Rice 1989) are used to adjust P -values for the multiple factors examined in each regression model. With two exceptions (tests of variation among subspecies in P[2nd litter] and SD[mass]), all tests individually significant at $P < 0.01$ remained significant when P -values were adjusted.

RESULTS

Sample Sizes and Distributions of Fitness Measures

Table 1 gives sample sizes for each of the three replicate laboratory stocks of each subspecies. Although an attempt was made to initiate each replicate with eight pairs of founders, founders that died after producing only one or two litters were replaced as breeders with another wild-caught founder. With one exception, due to the failure of many potential founder pairs of *P. p. leucocephalus* to breed, each replicate stock descended primarily from eight pairs of founders, with one or a few other founders contributing a small number of descendants.

Distributions of some of the measures of reproductive performance, growth, and survival deviated from normality in one or more of the populations. The distribution of litter size was approximately normal in each subspecies (neither skewness nor kurtosis was significant), even though litter size is a discrete variable (ranging from 1 to 9 in these data). Mass was leptokurtic in each subspecies. Log transformation, often applied to mass measurements, increased the kurtosis, and introduced skew. SD[mass] was both leptokurtic and positively skewed in each subspecies. RS was bimodally distributed because many pairs produced no progeny. Because the distributions of SD[mass] and RS deviated from normality