EVOLUTION

INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

PUBLISHED BY

THE SOCIETY FOR THE STUDY OF EVOLUTION

Vol. 56 **December 2002** **No. 12**

Evolution, 56(12), 2002, pp. 2347–2358

PERSPECTIVE: PURGING THE GENETIC LOAD: A REVIEW OF THE EXPERIMENTAL EVIDENCE

PETER CRNOKRAK AND SPENCER C. H. BARRETT[1]

Department of Botany, University of Toronto, 25 Willcocks Street, Toronto, Ontario, M5S 3B2, Canada

[1]*E-mail: barrett@botany.utoronto.ca*

Abstract.—Inbreeding depression, the reduction in fitness that accompanies inbreeding, is one of the most important topics of research in evolutionary and conservation genetics. In the recent literature, much attention has been paid to the possibility of purging the genetic load. If inbreeding depression is due to deleterious alleles, whose effect on fitness are negative when in a homozygous state, then successive generations of inbreeding may result in a rebound in fitness due to the selective decrease in frequency of deleterious alleles. Here we examine the experimental evidence for purging of the genetic load by collating empirical tests of rebounds in fitness-related traits with inbreeding in animals and plants. We gathered data from 28 studies including five mammal, three insect, one mollusc, and 13 plant species. We tested for purging by examining three measures of fitness-component variation with serial generations of inbreeding: (1) changes in inbreeding depression, (2) changes in fitness components of inbred lines relative to the original outbred line, and (3) purged population (outcrossed inbred lines) trait means as a function of ancestral outbred trait means. Frequent and substantial purging was found using all three measures, but was particularly pronounced when tracking changes in inbreeding depression. Despite this, we found little correspondence between the three measures of purging within individual studies, indicating that the manner in which a researcher chooses to estimate purging will affect interpretation of the results obtained. The discrepancy suggests an alternative hypothesis: rebounds in fitness with inbreeding may have resulted from adaptation to laboratory conditions and not to purging when using outcrossed inbred lines. However, the pronounced reduction in inbreeding depression for a number of studies provides evidence for purging, as the measure is likely less affected by selection for laboratory conditions. Unlike other taxon-specific reviews on this topic, our results provide support for the purging hypothesis, but firm predictions about the situations in which purging is likely or the magnitude of fitness rebound possible when populations are inbred remain difficult. Further research is required to resolve the discrepancy between the results obtained using different experimental approaches.

Key words.—Deleterious alleles, fitness, inbreeding depression, mating system, overdominance, purging.

Received February 25, 2002. Accepted September 3, 2002.

The deleterious consequences to fitness of matings among relatives has been known for over two centuries (Knight 1799). Darwin conducted some of the earliest experiments on the effects of selfing and outcrossing in over 50 plant taxa, and was one of the first to quantify the costs of inbreeding (Darwin 1868, 1876). More recently, inbreeding depression has become one of the most important areas of research in evolutionary biology and conservation genetics. Inbreeding depression is simply the reduction in fitness traits of inbred offspring in comparison to outbred offspring (Wright 1977; Charlesworth and Charlesworth 1987; Shields 1987). Most research on inbreeding depression during the past century has concentrated on domestic species and captive wild species (Ralls and Ballou 1986; Lacy et al. 1993). In addition, a disproportionate amount of research has been conducted on plants, as many species experience moderate to high levels of self-fertilization and inbreeding depression is of fundamental importance in mating-system evolution (Lande and Schemske 1985; Charlesworth and Charlesworth 1987; Barrett and Charlesworth 1991; Uyenoyama et al. 1993; Husband and Schemske 1995; Barrett and Harder 1996; Ritland 1996).

Most of the early literature on inbreeding depression focused on measuring its intensity when populations were inbred, and the relationship between the level of inbreeding (F) and inbreeding depression (Lynch and Walsh 1998). More recently, attention has focused on a phenomenon that Darwin first noted in his early *Ipomoea* experiments. Lineages of plants inbred for a number of generations initially experienced declines in components of fitness, but in subsequent generations these sometimes experienced a rebound in trait values. Darwin dubbed these plants ''heroes,'' as they were not only able to endure successive generations of inbreeding and recover fitness for certain traits, but also exhibited higher fitness than the original population (Darwin 1876). With no

understanding of Mendelian genetics, Darwin was at a loss to explain this phenomenon. Since then, researchers have coined the phrase ''purging the genetic load'' in reference to the fitness rebound that can occur in intensively inbred populations (Crow 1970).

The nature and degree of purging depends on the genetic basis of inbreeding depression. It is well known that inbreeding depression cannot occur if there are only additive gene effects for a trait (Falconer and MacKay 1996). If dominance is present, trait values decline linearly with the level of inbreeding (F) while epistatic allele interactions between loci can produce quadratic decreases in fitness (Crow and Kimura 1970; Charlesworth and Charlesworth 1999). There are two competing hypotheses to explain the decline in fitness with inbreeding: the partial dominance and overdominance hypotheses. The partial dominance hypothesis posits that inbreeding depression is the result of an increase in frequency of deleterious alleles. With an increase in mating among relatives, recessive deleterious alleles, once masked by dominance effects in the heterozygous form, become homozygous and express their effects on components of fitness (Davenport 1908; Bruce 1910; Keeble and Pellew 1910; Jones 1917). In contrast, the overdominance hypothesis posits that heterozygotes have higher fitness than homozygotes and that with inbreeding the frequency of homozygous loci increase, resulting in a decrease in fitness (East 1908; Shull 1908).

Evidence for both hypotheses exists in the literature, although it is commonly believed that the partial dominance hypothesis is more strongly supported (Charlesworth and Charlesworth 1999; Roff 2002). Recent research that has incorporated new experimental methods such as QTL mapping has lent support to both the partial dominance hypothesis (reviewed in Charlesworth and Charlesworth 1999; Crow 1999), and the overdominance hypothesis (Karkkainen et al. 1999; Li et al. 2001; Luo et al. 2001). Additionally, new advances in our understanding of gene regulation and enzyme biochemistry have provided some evidence for a mechanistic basis of overdominance (reviewed in Omholt et al. 2000 and de Vienne et al. 2001). If inbreeding depression is due primarily to strongly deleterious alleles, then purging the genetic load is a plausible mechanism by which populations could reduce the cost of inbreeding.

If inbreeding depression results primarily from deleterious alleles, another important consideration is the magnitude of the effect on fitness components when populations are inbred. With severe inbreeding depression, inbred individuals harboring deleterious alleles may die or not reproduce, effectively removing deleterious alleles from the population. Therefore, the magnitude of purging and the resulting rebound in trait values will be sensitive to the degree to which deleterious alleles are detrimental to fitness (Hedrick 1994; Wang et al. 1999; Willis 1999). Alleles of large effect, those that are lethal or semilethal when in homozygous form, will be relatively easily purged from the population (Lande and Schemske 1985; Charlesworth et al. 1990; Hedrick 1994; Schultz and Willis 1995; Wang et al. 1999). Alleles that are only partially deleterious will be more difficult to purge as inbred individuals carrying such alleles will have only slightly reduced fitness relative to individuals that are outbred (Hedrick 1994; Wang et al. 1999). These alleles are also commonly less recessive on average than strongly deleterious alleles (Simmons and Crow 1977; Crow and Simmons 1983). However, it seems most likely that inbreeding depression is due to deleterious alleles of *both* large and small effect: a mixed model system (Charlesworth and Charlesworth 1999; Wang et al. 1999). In such a case, successive generations of inbreeding will purge lethal alleles, while the genetic load resulting from mildly deleterious alleles will persist (Hedrick 1994; Wang et al. 1999). Although rarely discussed in the literature, it is also possible that inbreeding depression for any particular trait is due to both deleterious alleles of variable effect and fitness effects that result from the loss of heterozygosity alone. In such a case, purging the inbreeding load (deleterious alleles and overdominance) may be nonexistent or only partially successful, depending on the magnitude of the effect of deleterious alleles and the degree to which they affect fitness components compared to heterozygosity alone.

In a recent review of the plant literature, Byers and Waller (1999) conclude that ''purging is an inconsistent force within populations'' (p. 479) and, particularly important, that the extent of fitness rebound is limited. Their review included a wide variety of studies that used different experimental and statistical methods to test for the incidence of purging. The heterogeneity of the data they surveyed could potentially give confounding results. For example, a frequently used method is to compare trait values between natural populations that are inferred to be historically inbred or outcrossed. Although such a method of comparison yields interesting results, the necessity of having to infer inbreeding as opposed to experimentally manipulating populations, makes the generality of the results using this method questionable. Although Byers and Waller (1999) found that experimental method had little influence in determining the frequency of purging, their sample size for the comparisons was limited (less than 10) and they did not test specifically for effects on the magnitude of purging. Interestingly, Byers and Waller (1999, p. 501) state, ''only F studies [studies that serially inbred populations] provided overall evidence for purging. . . '' suggesting that an experimental approach might be a more reliable method for detecting the effects of purging.

Ballou (1997) examined purging in 25 captive zoo mammals using pedigree records that allowed an assessment of the extent of historical inbreeding. He found that only one species showed significant rebound in components of fitness with successive generations of inbreeding. Despite this, 17 of the 25 species exhibited minor increases in fitness components with inbreeding, which constituted a significant, but relatively small (2% at $F = 0.25$) overall effect (Ballou 1997). It is important to note that the degree of purging in captive zoo species will be limited because management protocols equalize founder representation, and in turn inbreeding coefficients, across families. This makes it difficult for selection to eliminate deleterious alleles (Keller and Waller 2002). Purging will also be limited in captive conditions since inbreeding depression is often substantially lower in magnitude under such conditions compared to the wild (Dudash 1990; Crnokrak and Roff 1999; Keller and Waller 2002). Examples of purging in wild mammal (e.g., Visscher et al. 2001) and human populations (Rao and Inbaraj 1977) hint at the pos-

sibility that selection against the inbreeding load may be stronger in nature than in captivity.

Here we examine evidence for purging the genetic load by reviewing empirical tests of fitness-component rebound with inbreeding in animals and plants. In addition to the frequency that studies provided evidence for purging, we were particularly interested in the magnitude of trait rebound with inbreeding. For this reason, we restricted our criteria for inclusion to those studies where populations have been serially inbred for several generations, and where measured changes in fitness-related traits were documented. Thus we have not included studies where inbreeding was inferred from historical or ecological data (e.g., outcrossing rates in plants), since the data have been reviewed elsewhere (Byers and Waller 1999). Also this method is likely to be inferior to experimentally measuring F by producing populations with varying degrees of inbreeding (Waller 1986). We were also interested in the approach used to measure purging, as there is evidence to suggest that the method employed affects the estimated magnitude and frequency of rebound (Willis 1999). Lastly, we also examined whether different traits (closely or distantly related to fitness) and taxa (e.g., mammals, insects, plants) are affected to different degrees by purging, as there is some evidence to suggest that fitness variation in inbred populations may vary with the taxon or character investigated.

The Dataset

We restricted our survey to those studies that examined purging by serially inbreeding populations for at least two generations. This is because two is the minimum number of generations of inbreeding that are needed to detect a possible rebound in fitness components if purging occurs. Since different levels of inbreeding (F) will produce different magnitudes of inbreeding depression and subsequently affect the degree of purging, we took into consideration the maximal level of F for each study. This is important as the number of generations of inbreeding and the method used (e.g., full-sib mating or selfing) will affect the coefficient of inbreeding. F ranged from a low of 0.27 (butterfly species) to a high of 0.97 (various plants). Since our goal was to measure not only the incidence of purging but also the magnitude of trait rebound with purging, we were limited to experimental studies in which levels of inbreeding were manipulated to obtain accurate F estimates.

Animal studies primarily consisted of full-sib mated inbred lines, while all plant studies used selfing as the means of inbreeding. We found a total of 28 studies that examined purging in 22 species consisting of mice, fruit flies, house flies, a butterfly species, one snail species, and a variety of plants including terrestrial annuals, aquatic perennials, and a few tree species. Most studies examined a wide variety of traits, with the dataset primarily composed of life-history traits (72%): for example, offspring number, development time, seeds per fruit, and time to flowering. Morphological traits such as stem diameter and body weight (when not a function of development rate) were also represented, but more infrequently (22%) compared to life-history traits. A small minority of traits (6%) we categorized as physiological traits, as they pertained to biochemical pathways. The average number of traits measured in each study was 7.3. In total, we obtained 209 estimates of relative fitness for a variety of traits (for the complete dataset and references, see Electronic Appendix, currently available from the *Evolution* editorial office at: evolution@asu.edu).

Experimental Designs

Experiments investigating inbreeding depression typically involve artificially inbreeding a naturally outbred population. Although the manner in which inbreeding occurs differs among organisms based on their mating systems (few animals can self whereas many plant species can), most researchers establish 10 to hundreds of lines from the outbred population within which inbreeding is conducted. Serial inbreeding is maintained within these lines for a number of generations until the coefficient of inbreeding reaches a certain desired level. Fitness components of inbred lines (I1, I2, I3, etc.; Fig. 1) are then compared either to the original outbred population (the base population O1; Fig. 1), or with a concurrently maintained outbred control line (O2, O3, etc.; Fig. 1). Comparison of fitness components between mean inbred line and an outbred control allows an estimate of inbreeding depression for each generation of inbreeding: for example, inbreeding depression for first generation = 1 − I1/O1 (Fig. 1). Additionally, comparison of the fitness components of the mean inbred line (of last generation of inbreeding) and the original outbred line allows researchers to determine the degree to which inbreeding changes fitness components over time (1 − (I5 − O1)/O1; Fig. 1). This method also allows for the assessment of the degree of divergence in traits due to inbreeding.

Deviations from this general protocol do exist, the most common being the incorporation of an additional outcrossed generation after serial inbreeding (OC; Fig. 1). Here, the inbred lines are maximally outcrossed with each other to form a new outcrossed population with the purpose of restoring the original (base population) levels of heterozygosity that are lost within lines due to inbreeding. Traits are then compared with the original outbred population (OC/O1; Fig. 1). Willis (1999) recently advocated going one step further by inbreeding the newly formed outcrossed population (IOC; Fig. 1) and comparing inbreeding depression with that of the first generation of inbreeding (1 − I1/O1 vs. 1 − IOC/OC; Fig. 1). He argued that such a comparison would allow for an accurate assessment of the degree of purging, as the coefficient of inbreeding is then equal in both generations because the outcrossing event sets F back to zero (Falconer and MacKay 1996). Unfortunately, although this is clearly desirable, his own study is the only one in our survey that has employed this approach (see below).

Measures of Purging

We estimated the magnitude of purging using the three most commonly reported measures in the literature: (1) changes in inbreeding depression with successive generations of inbreeding, (2) relative changes in inbred line fitness components compared to original outbred fitness components, and (3) the ratio of purged population trait values as a function of ancestral outbred population trait values (formed by out-

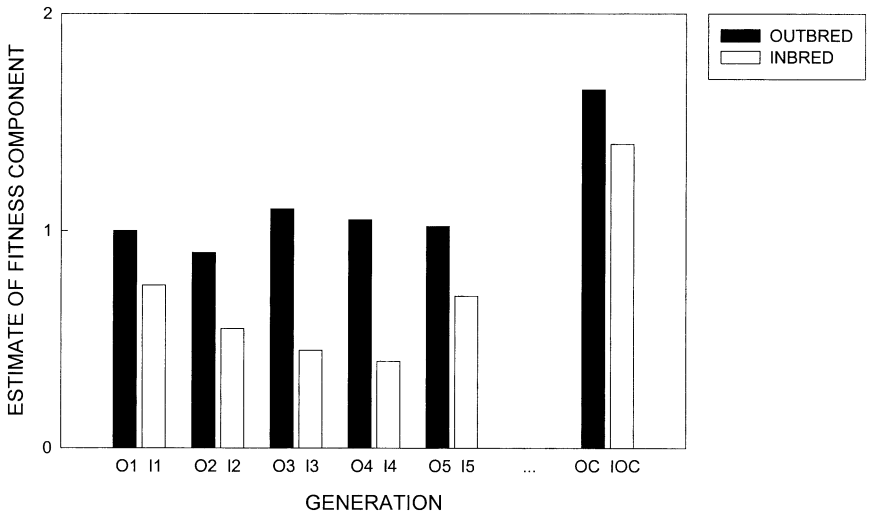

FIG. 1. Experimental protocols used to test for purging. O1, O2, O3. . . , estimate of fitness component value of outbred line for generation 1, 2, 3. . . etc.; I1, I2, I3. . . , estimate of fitness component value of inbred line for generation 1, 2, 3. . . etc.; OC, estimate of fitness component value of outcrossed inbred lines formed after several generations of inbreeding; IOC, estimate of fitness component value of inbred outcrossed lines formed by inbreeding OC.

crossing inbred lines). Each measure has its advantages and disadvantages when estimating the magnitude of purging. Because of this, we will define each measure and discuss the theoretical and practical issues involved with their use.

Changes in inbreeding depression with inbreeding

One of the most intuitive measures used to estimate purging is to track the change in inbreeding depression (δ) with serial inbreeding ($1 - I1/O1$ vs. $1 - I5/O5$; Fig. 1). If purging has occurred, after an initial increase, δ should decrease with successive generations of inbreeding. Since δ is a function of outbred fitness components, any changes in fitness of the inbred lines not due to the purging of deleterious alleles (e.g., environmental changes across generations) should be controlled for. Despite this, decreases in δ can occur without purging. With serial inbreeding, selection against deleterious alleles will be strong, but as effective population size (N_e) declines due to demographic factors or inbreeding, a large portion of deleterious alleles can become effectively neutral to selection (Crow and Kimura 1970; Wright 1977). With time, deleterious alleles can become fixed in lines due to drift. Additionally, deleterious alleles can become fixed in inbred lines due to background selection or selective sweeps (Charlesworth et al. 1997; Charlesworth and Charlesworth 1999; Wang et al. 1999). Because of these factors, although strongly deleterious alleles such as lethals may be purged from a population with inbreeding, a certain portion of the genetic load due to detrimentals can become fixed and permanently decrease the relative fitness of inbred, and in certain cases, control lines (Hedrick 1994; Charlesworth and Charlesworth 1999; Wang et al. 1999). Such an effect will obscure differences between the control and inbred lines, leading to the spurious conclusion that δ has been reduced. In situations such as this, between-population comparisons involving populations with varying degrees of inbreeding can be used to effectively measure inbreeding depression, as opposed to the within-population estimates reported here (see Ritland 1990). Additionally, increases in viability in inbred lines due to environmental factors that do not affect the outbred lines, have also been demonstrated (Kalinowski et al. 2000; Kalinowski and Hedrick 2001), but are most likely infrequent.

Since δ will increase with F (Lynch and Walsh 1998), we calculated F standardized estimates of inbreeding depression, b (the cost of inbreeding for a given F). Although standardizing for variation in F accounts for the major difference between generations of inbreeding, it cannot account for all variation. Variation in the number of generations of inbreeding may affect levels of purging as selection against the inbreeding load has longer to act in intensively inbred populations. We standardized initial and final δ by dividing by, respectively, initial and final coefficients of F (for derivation of equations, see Crnokrak and Roff 1999). Initial F is simply the level of inbreeding for the first generation of sib mating in animals, or selfing in plants, while final F is a function of the number of generations and manner of inbreeding and thus represents the maximal level of F for that study. The only investigation that involves an exception to this relationship is the *Mimulus* study by Willis (1999) where final F is equal to initial F. This is because final δ was measured in the outcrossed inbred lines (formed after five generations of selfing) where $F = 0.5$, because the outcrossing event sets F back to zero (for advantages to this protocol, see Discussion). Since Willis (1999) measured δ in the ''purged'' population as opposed to the last generation of inbreeding, as was done in the 12 other studies, we also ran an analysis excluding his study. We found 65 paired estimates of δ reported for 13 studies (see Electronic Appendix for details).

Although purging can also be assessed by tracking changes in the inbreeding load, B (Crow 1970), we were unable to obtain an adequate sample size for this measure because a regression of log trait value on F for a number of generations of inbreeding is required. A number of studies included in this review inbred populations for less than three generations, thus preventing us from calculating the necessary regressions.

Tracking relative inbred line fitness components with inbreeding

Most studies reviewed here provided data such that purging could be examined by comparing changes in fitness components in inbred lines during inbreeding with the original outbred population (1 − (I5 − O1)/O1; Fig. 1). Some of these studies also measured the relative fitness components of inbred lines compared to an outbred control line over time. If purging occurs with successive generations of inbreeding, then the relative trait values of inbred lines, after an initial drop, should start to rebound to levels comparable to the original outbred level (see Lacy and Ballou 1998). This would result in a nonlinear, decelerating decline in inbred line trait values with increasing *F*. It assumes that the resultant increase in fitness components with inbreeding truly results from the selective elimination of lines (between line selection) that express deleterious mutations, and is not due to environmental effects or adaptation to laboratory growth conditions. Purging of the deleterious load can also occur by selection within lines (Charlesworth and Charlesworth 1999), but its effects are likely to be more evident when selection acts between lines (Wang et al. 1999). Comparing inbred trait values with a concurrently running control line would account for some effects due to adaptation, but not all.

Our second measure of purging is inbred relative fitness (IRF). If purging has occurred, we should see comparable or significantly higher inbred trait values over time compared with the original outbred trait values (IRF estimate significantly greater than one; Fig. 2). Lack of a significant difference between final inbred and initial outbred trait values also indicates that complete purging has occurred. This is because the null expectation is that subsequent inbreeding should decrease the fitness components of inbred lines (Fig. 2). But no significant difference between outbred and inbred lines may also be due to inadequate sample sizes as much as a real lack of difference. A low value for IRF indicates that initial outbred fitness is significantly larger than final inbred fitness and therefore no purging has occurred (Fig. 2). An alternative explanation is that serial inbreeding decreases genetic variation, and subsequently, fitness components, despite significant purging. Additionally, nonlinear decreases in trait values with increasing *F* as shown in Figure 2 (partial and complete purging lines) can also result from epistasis rather than purging (Willis 1993; Lynch and Walsh 1998). Also, the arguments concerning the conditional nature of δ (see previous section) also apply to IRF. Nearly all of the traits examined could be scored for IRF, so that the sample size for this measure of purging was the largest of the three (205).

Ratio of purged population fitness components in relation to ancestral fitness

If purging has occurred with serial generations of inbreeding, inbred lines should harbor fewer deleterious alleles. In addition, inbreeding (particularly in plants, where it can be severe due to selfing) will result in a genetic divergence among replicate lines due to drift. If inbred lines are fully outcrossed after a number of generations of inbreeding, the subsequently formed ''purged'' population (outcrossed inbred lines; OC; Fig.1) will have levels of heterozygosity near or equivalent to the ancestral outbred population but will have a substantially reduced genetic load (East and Jones 1919). In such a case, the ratio of purged population trait means as a function of ancestral outbred population trait means (OC/O1; Fig. 1) should be significantly larger than one, as the reduced genetic load should result in an increase in values of the fitness components. Values not significantly greater than one indicate little to no purging, while values significantly less than one indicate reduced fitness component values, due presumably to the loss of genetic variation and/or fixation of deleterious alleles. We call this measure of purging RATIO, as it is the ratio of purged population as a function of ancestral population trait values. We found 56 RATIO estimates of purging for 10 studies split between mammals and plants. Interestingly, the RATIO measure can be used both to test for purging (East and Jones 1919) and to make inferences about the genetic basis of inbreeding depression (see Roff 2002). This is because a definitive measure of purging implies that inbreeding depression is at least partially due to deleterious alleles as opposed to the loss of heterozygosity. Although the RATIO measure has been advocated as an accurate measure of purging, it is not free of problems arising from the effects of adaptation to laboratory conditions (Willis 1999; see Discussion).

Statistical Analyses

Since all purging measures were expressed as proportions or ratios, we used nonparametric analyses to test for purging. In addition, we also used meta-analysis to test for differences in inbreeding depression and relative fitness components with inbreeding.

Meta-analysis is particularly appropriate in reviews of this sort as the inherent variation within and between studies will confound the results obtained from conventional tests of dif-

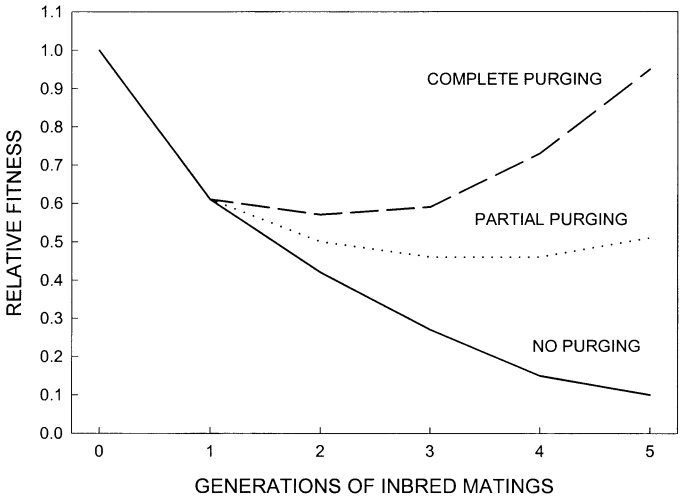

FIG. 2. Three possible scenarios for purging the inbreeding load used for comparative purposes with the inbred relative fitness measure. No purging: fitness component of inbred lines continues to decrease linearly with successive generations of inbreeding. Partial purging: inbred line fitness component decreases initially, then rebounds to approximately 50% of outbred level. Complete purging: after an initial drop, inbred lines recover fitness component value to levels comparable to original outbred level.

ferences between means such as *t*-tests (Hedges and Olkin 1985). Variability between studies can also be accounted for statistically by weighting effects for different sample sizes and variance components (Arnqvist and Wooster 1995). Any interpretation of mean effect size must be done in light of the inevitably large variance component of effect size. The estimate of the variance of effect size can also be tested against a null hypothesis of zero (no heteroscedasticity) to determine if our studies have significant heterogeneity.

Meta-analysis requires the calculation of the mean effect size $\bar{\Delta}$, the variance of $\bar{\Delta}$: $\sigma^2(\Delta)$, and the *Q* statistic, which allows for testing of $\sigma^2(\Delta)=0$. The statistical methods of meta-analysis are laid out in detail in Hedges and Olkin (1985) and will not be discussed here except what is necessary for the present analysis. We used MetaWin 2.0 (Rosenberg et al. 2000) for all meta-analyses. Finally, we determined the degree to which the file drawer problem (i.e., what is the possibility that the number of unpublished null results invalidate the results of the meta-analysis by bringing the significance down to $P = 0.05$?) affects our statistical analysis by estimating the ''fail-safe'' number: that value of the number of filed studies (those studies not published) required to bring the probability of a type-I error (rejecting null hypothesis when true) to the desired level of significance (Hedges and Olkin 1985). A small fail-safe number (in reference to the total sample size) means that the findings of the meta-analysis are not resilient to the file drawer problem and we must then reconsider our conclusions.

For all statistical tests, we used a hierarchical method of analysis: first, we analyzed the entire dataset (all estimates reported per study used), followed by an analysis of study means (mean values calculated per species/study combination; i.e., the Lacy and Ballou study constitutes three separate data points since three different species were examined). The number of estimates available for each measure of purging varied among studies and ranged from a low of one to a high of 32 (median = 6). Since measuring a large number of traits in a single species/study does not constitute independent data (due to potential correlations between traits within a species), we also calculated mean estimates of purging (averaged across traits for each measure of purging) per study. We then ran separate analyses for different trait types, as life-history traits are known to exhibit higher inbreeding depression than morphological traits (DeRose and Roff 1999), which in turn may affect the degree of purging (Hedrick 1994). Comparing the entire dataset with the study means allowed us to infer the degree of bias introduced from the overrepresentation of certain traits within studies. Meta-analysis is not subject to this problem, as the method uses mean values per study as raw data.

There was substantial variation between studies for the maximal level of *F* reached after successive generations of inbreeding (range 0.27–0.97). This is the product of variation in the number of generations of inbreeding (average: 4.5) and the different method of inbreeding employed in each study (full-sib or selfing). As most plant studies conducted inbreeding by selfing, the mean maximal level of *F* was significantly higher in plants compared to mammals and insects ($F_{2,23} = 7.51$, $P = 0.004$; plants: 0.81 ± 0.03; mammals: 0.62 ± 0.06; insects: 0.55 ± 0.07). Naturally outbred organisms that experience more intense inbreeding are expected to exhibit higher levels of inbreeding depression (Lynch and Walsh 1998), which may affect the magnitude of purging (Hedrick 1994; Charlesworth and Charlesworth 1999; Wang et al. 1999). We therefore corrected estimates of δ for variation in *F* (*b* estimate) since, in most cases, the comparison between initial and final δ constituted a significantly larger difference in *F* than comparisons across taxa. Additionally, when the data permitted, we conducted analyses for the entire dataset within taxonomic groupings (in which variation in *F* is lowest) to avoid potential confounding effects.

Results

Question 1: Does inbreeding depression decrease with successive generations of inbreeding?

To test if final *b* is lower than initial *b*, we used nonparametric Wilcoxon-signed ranks test (to determine if initial − final $b = 0$; Sokal and Rohlf 2000) and meta-analysis. Since we have no a priori reason to believe that differences between initial and final *b* will be constant across studies, we used a random-effects model as opposed to the more common fixed-effects model for the meta-analysis (Hedges and Olkin 1985).

Since the dataset is composed of a wide variety of organisms, there may be inherent differences between taxa in the frequency and magnitude of purging, due to differences in mating systems (Byers and Waller 1999). Such variation may obscure general patterns of purging. Because of this, we used Kruskal-Wallis to test for effects of taxonomic type (mammals, insects, molluscs, and plants) on the mean difference in *b* (initial − final *b*). Additionally, we tested for between taxon heterogeneity in variance components for the meta-analysis. We pooled data from different taxa when results revealed no significant effect of taxonomic group.

Results: significant decrease in b with inbreeding.—When standardized for variation in *F*, 72% of traits showed a decrease in *b* with successive generations of inbreeding. Mean final *b* was significantly lower than mean initial *b* for mammals and plants and nonsignificant for insects and molluscs (Fig. 3; significant differences between taxonomic groups: $\chi^2 = 18.79$, df = 3, $P = 0.0003$). The drop in *b* for mammals was primarily due to a single mouse study and therefore we caution any interpretation of the results as evidence for purging for this subset of the data. The percent drop in *b* for plants was moderate at 16% (whole dataset). The analysis of study means revealed no significant differences between taxonomic groupings (4 taxa separate: $\chi^2 = 4.04$, df = 3, $P = 0.26$; animals versus plants: $\chi^2 = 2.91$, df = 1, $P = 0.09$) and an overall significant decrease in *b* by 48% (mean initial $b = 0.73 \pm 0.18$, mean final $b = 0.38 \pm 0.08$; $z = 2.25$, df = 12, $P = 0.01$). This result represents a substantial decrease in the cost of inbreeding. The study means analysis for plants alone revealed a 15% drop in the cost of inbreeding, which was not significant, presumably due to the small sample size of eight ($z = 1.32$, df = 7, $P = 0.19$). Removing the Willis (1999) study did not change appreciably the results for plants: 15% drop in *b* ($z = 0.93$, df = 6, $P = 0.36$).

The meta-analysis also revealed no significant heterogeneity between taxonomic groupings (animals vs. plants; Table 1), a similarly large mean effect size of 0.42, and a 95%

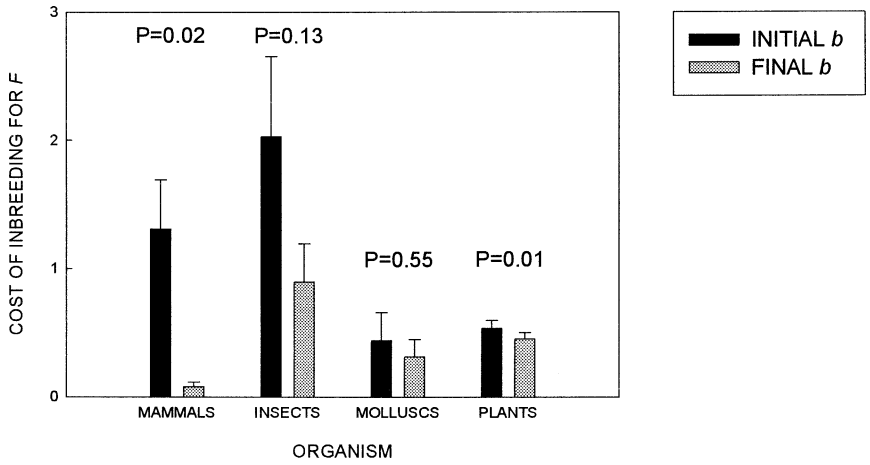

FIG. 3. Mean (± SE) initial and final inbreeding depression standardized for *F* (*b*), for the four taxonomic groupings: mammals, insects, molluscs, and plants. Mean values are for the entire dataset. *P*-values above bars are significance levels for differences between initial and final *b* for each group using Wilcoxon-signed ranks test.

confidence interval that does not include zero (see Table 1). The low variance of the mean effect size and the moderate (compared to the sample size) fail-safe number, indicates that the results are robust to heterogeneity between studies and publication bias (Table 1). Thus we can conclude that the cost of inbreeding substantially decreased with generations of inbreeding for the studies we examined.

Question 2: Is there a significant increase in relative inbred fitness components with successive generations of inbreeding?

To determine if purging had occurred, we tested whether relative inbred trait means rebounded after several generations of inbreeding. If complete purging had occurred, then IRF should be significantly greater than one (Fig. 1) or there should be no significant difference between outbred and inbred trait values (see Fig. 2). No purging would mean that the IRF estimate is less than one and very low (see Fig. 2). Partial purging is much more difficult to measure statistically as it is a relative measure between complete and no purging. We used the Wilcoxon signed-ranks test to test for significant deviations from one (no difference between outbred and inbred fitness components). Since the IRF measure contained multiple estimates for each study, we only used mean estimates per study and did not conduct an analysis on the whole dataset.

Results: moderate trait rebound in inbred lines.—Using IRF as a measure of purging, we found that 53% of the reported traits revealed evidence of complete purging with inbreeding. Only 8% of the traits revealed substantial purging where inbred line trait values were significantly larger than outbred values after several generations of inbreeding. Roughly half of the reported traits (47%) revealed no purging. However, this was not due to a lack of effect of inbreeding on traits, as mean initial δ was moderately high at 0.21 ± 0.04 (as was final δ, 0.24 ± 0.06) and significant greater than zero, indicating significant levels of inbreeding depression. We caution against accepting the interpretation that the lack of a significant difference between inbred line and outbred fitness components is evidence for purging, since the lack of a difference could also be due to inadequate sample sizes.

When we divided the dataset according to trait type we found evidence of purging for morphological traits in mammals, but only weak evidence for purging in plants (Table 2). On the other hand, life-history traits showed only moderate evidence for purging in both mammals and plants (Table 2). Insects showed substantial rebounds for life-history traits, with the mean IRF estimate of 1.07 (inbred fitness on average larger than outcross fitness).

Question 3: Is purged population fitness as a function of ancestral population fitness significantly greater than one?

To test if the RATIO measure of purging is significantly greater than one, we used the Wilcoxon signed-ranks test.

TABLE 1. Meta-analysis results for the test statistics initial − final *b* and RATIO. Δ and CI (95%), mean effect size and 95% confidence interval; *Q*, equivalent chi-square test statistic to test whether $\sigma^2(\Delta) = 0$ (no *Q*-values were larger than the 100 $(1 - \alpha)$ percentile point, and therefore we cannot reject the null hypothesis); Q_{TOTAL}, total heterogeneity; $Q_{BETWEEN}$, heterogeneity between taxonomic groupings; $\sigma^2(\Delta)$, variance of Δ; *N*, number of studies; N_n, fail-safe study number (to test file drawer problem).

Analysis	$\bar{\Delta}$ and CI (95%)	$\sigma^2(\Delta)$	Q_{TOTAL}	$Q_{BETWEEN}$	*N*	N_n
b	0.12 ≤ 0.42 ≤ 0.82	0.264	12.61	2.28	13	26
RATIO	1.04 ≤ 1.17 ≤ 132	0.042	8.00	NA	10	9

TABLE 2. Mean (±SE) estimates of inbred relative fitness (IRF) for morphology/life-history traits and among different taxonomic groups. IRF = 1 − (I5 − O1)/O1; see Figure 1: also equivalent to relative fitness measure in Figure 2. Significant values less than one indicate no or partial purging, and values not significantly different from one indicate complete purging (see Fig. 1 for details). Significance was tested against one: no significant difference between inbred line and outbred fitness component estimates.

Trait	Mammals	Insects	Plants
Morphology	0.95 ± 0.059	NA	0.78 ± 0.060*
Life history	0.81 ± 0.057*	1.07 ± 0.242	0.70 ± 0.047*

* $P < 0.001$.

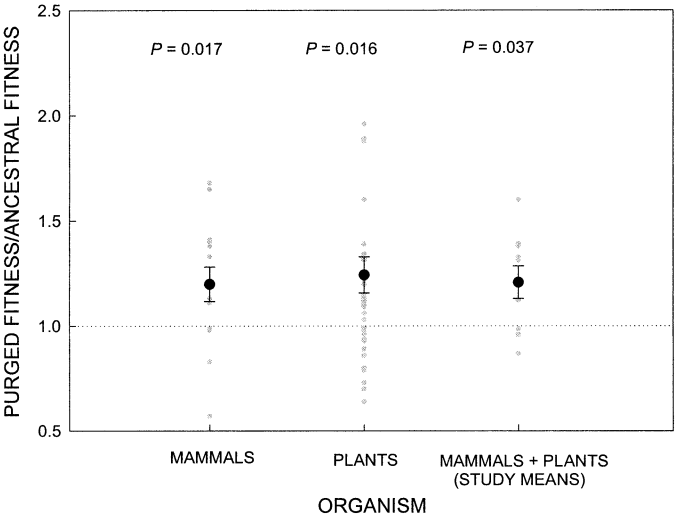

FIG. 4. Mean RATIO estimates (± SE) for mammals and plants for the entire dataset and study means (mammals and plants combined). *P*-values represent significance levels for Wilcoxon signed-ranks test for RATIO estimates greater than one. Dashed line represents RATIO of one. RATIO, purged population (outcrossed inbred lines) trait mean as a function of ancestral outbred trait mean.

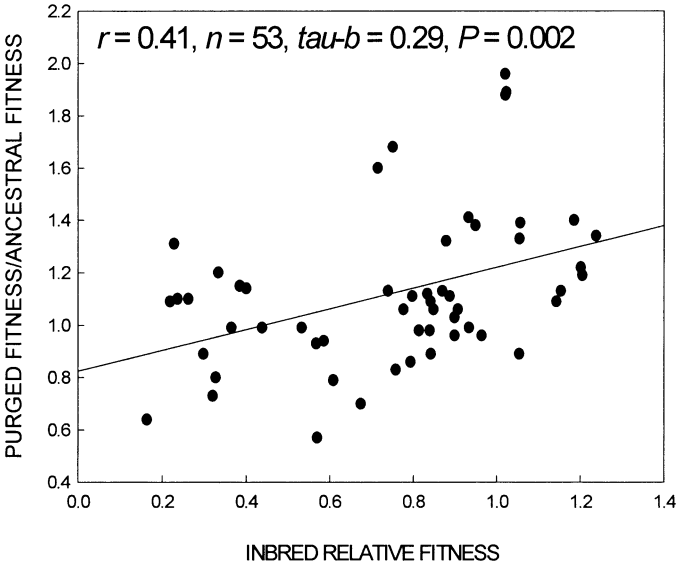

FIG. 5. Nonparametric Kendall's *tau-b* correlation for the measures of purging: RATIO versus inbred relative fitness (IRF). IRF, mean inbred line trait values after serial generations of inbreeding as a function of original outbred trait values; RATIO, purged population (outcrossed inbred lines) fitness component values as a function of ancestral outbred fitness component values.

We also performed a meta-analysis for the RATIO measure, using the most commonly reported trait for all 10 studies (reproductive output, examples of which include litter size in mice and number of flowers/seeds in plants). We used only one estimate per study (and its corresponding standard deviation) as opposed to a mean estimate per study, since a few studies only reported single estimates (Hedges and Olkin 1985). To analyze the mean effect size for this measure, we used Hedge's ln response ratio (Hedges et al. 1999), which allows for the testing of nonparametric ratio estimates with favorable sampling distributions (Hedges et al. 1999).

Results: frequent and substantial evidence for purging found.—Unlike the previous measure of purging that showed moderate fitness rebound with inbreeding, the RATIO measure revealed evidence for frequent and substantial purging in a variety of traits. Purged population trait values as a function of ancestral population trait values (OC/O1; Fig. 1) were significantly greater than one, 48% of the time. Figure 4 indicates that very little difference exists between the mean RATIO measures of purging between mammals and plants (no data available for insects and molluscs). Additionally, mean RATIO estimates were significantly greater than one using a Wilcoxon signed-ranks test for both organism types (mammals: $z = 2.40$, df = 13, $P = 0.017$; plants: $z = 2.42$, df = 41, $P = 0.016$). The study means analysis revealed similar effects to that for the whole dataset (Fig. 4): the ratio of purged fitness components to ancestral fitness components was significantly greater than one ($z = 2.09$, df = 9, $P = 0.037$). The average increase in trait values for the purged population compared to the ancestral population was 20% in mammals and 24% in plants. Both these values represent significant increases in fitness components with inbreeding and therefore can be interpreted as evidence for purging of the genetic load.

The results of our meta-analysis revealed that the ratio of purged-population fitness components to ancestral-population fitness components was significantly greater than one (Table 1). Additionally, the estimated pooled variance was low, indicating little between-study heterogeneity (see Table 1). Thus significant and substantial purging is frequently measured when purged-population trait values are reported as a function of ancestral-population trait values. Despite this, the fail-safe number of nine indicated that more studies are needed to fully resolve this issue. The low fail-safe number is most likely due to the moderate mean effect response ratio of 1.17. Although moderate for the purpose of meta-analysis, a relative fitness ratio of 1.17 represents a substantial rebound in fitness for important life history traits.

Question 4: Do the three different measures of purging return consistent results for the same trait?

To find out whether there was a significant effect of measure type on the magnitude of purging, we determined the degree of correspondence between measures of purging for all traits within each study that reported two or more measures of purging. We found that only six studies involving 33 traits reported all three measures of purging. We used Kendall's *tau-b* nonparametric paired measure of association (Sokal and Rohlf 2000) to determine whether paired measures of purging give similar results. For this analysis we compared the difference in *b* (initial − final *b*) with IRF and with the RATIO measure of purging.

Results: limited correspondence between the different measures of purging.—We found a significant correlation between the IRF and RATIO measures of purging ($r = 0.41$, *tau-b* = 0.29, $P = 0.002$): traits that showed substantial and significant purging for the RATIO measure, also showed substantial and significant purging for the IRF measure (Fig. 5).

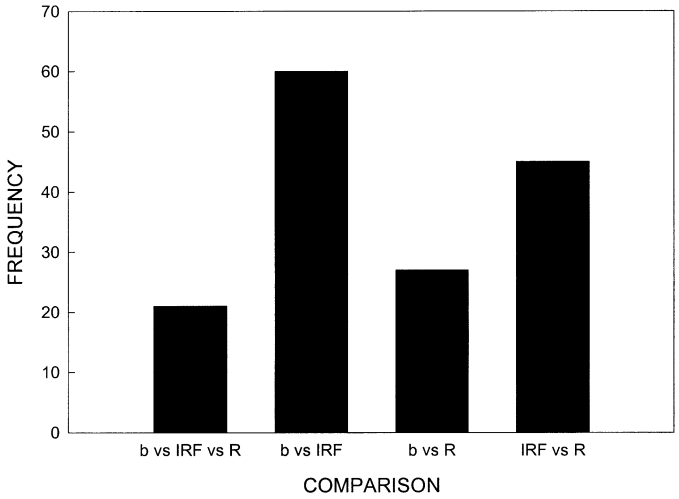

FIG. 6. Comparisons among the three different measures of purging: initial − final *b* (b) inbred relative fitness (IRF), and RATIO (R). Frequency values for different purging measure combinations represent the degree to which different measures return similar results (data scored as significant purging/no-purging) for each trait out of all available traits. Initial − final *b*, difference in inbreeding depression before and after serial generations of inbreeding (standardized for *F*); IRF, mean inbred line trait values after serial generations of inbreeding as a function of original outbred trait values; RATIO, purged population (outcrossed inbred lines) fitness component value as a function of ancestral outbred fitness component value.

We found no correspondence between IRF and initial − final *b* ($r = 0.21$, *tau-b* = 0.004, $P = 0.97$). Additionally, we found no correspondence between initial − final *b* and RATIO ($r = 0.14$, *tau-b* = 0.01, $P = 0.92$).

Although one comparison between measure types revealed some level of correspondence (Fig. 5), the correlation coefficient is only moderately high (0.41), indicating that the information obtained from different measures of purging often did not correspond. Additionally, the frequency of substantial purging (using the above dataset) between the three measures differed: initial − final *b* = 0.10, IRF = 0.06, RATIO = 0.27. This pattern is similar to that for the whole dataset; initial − final *b* = 0.06, IRF = 0.08, RATIO = 0.38. Figure 6 reveals that the frequency of correspondence between the different measures (data scored as significant purging/no-purging) is highest for initial − final *b* versus IRF and IRF versus RATIO measures, and it is lowest for the two measures we predicted might represent purging: initial − final *b* versus RATIO.

In summary, purging was most likely to be measured where reports included fitness components of the purged population as a function of the ancestral population (RATIO measure) and changes in inbreeding depression were standardized for *F* (*b*). Thus, the manner in which a researcher chooses to measure the magnitude and frequency of purging has an important influence on the results obtained.

DISCUSSION

Is Purging Common When Populations Are Inbred?

Although the idea that the deleterious load can be effectively purged with inbreeding has been frequently discussed in the literature, two recent reviews have concluded that only limited evidence exists for this phenomenon (Ballou 1997; Byers and Waller 1999). Our review of purging revealed two major findings. First, with some exceptions, our results using different measures lend support to the purging hypothesis for most traits. This is particularly true for the results obtained from the measure we believe is most powerful in detecting purging: changes in inbreeding depression. Second, we found that the probability and degree of purging detected is conditional on the manner in which it is measured. Only moderate rebound was reported if researchers tracked changes in relative inbred line fitness components. In contrast, greater levels of purging were evident when researchers tracked changes in the cost of inbreeding standardized for *F*, and when outcrossing inbred lines and comparing fitness components to the ancestral population. This last method reveals that the mean relative increase in fitness components with purging is substantial (20% in mammals and 24% in plants), but ranges from a low of 0% to a high of 96%. Although the decrease in the cost of inbreeding for a given *F* was substantial (average decrease in *b* of 42–48% between initial and final inbreeding), we found only partial correspondence between this and the RATIO measure of purging. For traits in which both measures were available, approximately 73% of the cases showed no correspondence between *b* and RATIO. Thus we are left with having to account for the discrepancies between the results obtained with different methods. Our discussion begins by examining the factors that could contribute to this discrepancy.

Adaptation to Laboratory Conditions or Purging?

The lack of correspondence between the two measures we ranked highest in terms of power to detect purging (*b* and RATIO) suggests an alternative to the purging hypothesis when characterizing purging using the RATIO measure. The substantial trait value increase for outcrossed inbred lines may be due to inadvertent adaptation to growing conditions (see Willis 1999). Such a scenario is possible based on what we know of how experiments are often conducted. For example, plants that mature early under laboratory conditions may be favored due to time constraints as researchers inadvertently reduce the intergeneration interval between serial generations of inbreeding. Additionally, selection within and among lines for general conditions in the laboratory or glasshouse may result in adaptive evolution to those conditions. Successive generations of selection may result in rapid changes in mean maturation time, and in conjunction with other correlated traits, ultimately in higher fitness in the inbred lines without any marked decrease in the genetic load. If the inbred lines are then outcrossed to restore original heterozygosity, and fitness components are compared with the ancestral population (our RATIO measure), one should observe a marked relative increase in trait values in the ''purged'' population. This is to be expected since the ancestral population has not experienced similar selection pressures. A rebound of this sort could easily be mistaken as evidence for purging.

The discordance that we found between our different measures of purging (notably changes in inbreeding depression

and RATIO) leads us to conclude that in these cases, the frequently measured increase in trait values when inbred lines are outcrossed may be due to adaptation to growth conditions and not purging. Both purging and adaptation to laboratory conditions are factors that could be responsible for the rebounds in fitness components that we found for traits using the RATIO measure. However, changes in inbreeding depression are not likely to be affected to the same degree by adaptation to laboratory conditions since a concurrently maintained control line is used as the outbred reference. This assumes that inbred lines do not experience faster responses to selection compared to the outbred control (for exceptions see Willis 1999). Therefore, the decrease in the cost of inbreeding for a given *F* that we observed (particularly for the plant data) would appear to provide evidence for purging.

Testing the Genetic Basis of Inbreeding Depression

The potential occurrence of adaptation to laboratory conditions has important implications for researchers interested in the genetic basis of inbreeding depression. This is because comparing trait values of outcrossed inbred lines with ancestral population trait values (RATIO measure) is a frequently employed experimental method to test between the partial and overdominance theories. This method assumes that inbreeding can purge the genetic load to some measurable degree and that trait rebound with inbreeding is not due to causes other than purging. Essentially, traits that show some level of purging indicate that inbreeding depression is at least partially due to deleterious alleles and not to a loss of heterozygosity. This is therefore used as support for the partial dominance hypothesis (Roff 2002). Conversely, a lack of purging (traits rebound to ancestral levels) could indicate that inbreeding depression is due to a loss of heterozygosity and has been used to support to the overdominance hypothesis (Roff 2002).

Using our RATIO measure, we found equivocal support for both the partial dominance and overdominance hypotheses for 56 traits. However, due to the lack of correspondence between the different measures of purging (notably *b* and RATIO), we caution against interpreting this result as evidence that overdominance contributes to inbreeding depression. The lack of correspondence suggests that rebounds in traits may be the result of adaptation to growing conditions and therefore casts doubt on the utility of the RATIO measure to infer the genetic basis of inbreeding depression. Further, theoretical considerations alone have shown that deleterious alleles of small effect are difficult to purge (Hedrick 1994; Wang et al. 1999). This would lead to the erroneous conclusion when using the RATIO measure that inbreeding depression is due to overdominance since mildly deleterious recessives should act like symmetrical overdominant loci (J. Willis, pers. comm. 2002).

Experimental Protocols

If rebounds in trait values with inbreeding are not due to purging but are instead a result of adaptation to laboratory conditions, what can be done experimentally to tease apart these two confounding effects? Although molecular methods that measure changes in deleterious gene frequency (measures that are not subject to adaptation to laboratory conditions) can be used to verify the results obtained from quantitative methods discussed here, they are available only for the few species that have been intensively studied at the molecular level (Charlesworth and Charlesworth 1999). For most species, measuring quantitative fitness component variation with inbreeding is one of the few available approaches that can be used to detect purging.

The problem of distinguishing between purging and adaptation to growth conditions rests in part on the use/misuse of an appropriate outbred control group for the purpose of comparing fitness components. An appropriate control group would be one for which the inadvertent selective regime experienced by the inbred lines would also apply. This would require knowing what effect maintaining and choosing individuals to propagate the inbred lines has on the traits of interest. Since this selection is inadvertent, knowing its effects and being able to replicate them in the outbred lines is impossible.

One potential way of getting around this problem for those experiments that compare purged population fitness components with the ancestral population (OC vs O1; Fig. 1) is to inbreed the outcrossed inbred lines (purged population) and measure inbreeding depression ($1 - IOC/OC$; Fig. 1; Willis 1999). If purging has occurred, then inbreeding depression should be significantly reduced ($1 - IOC/OC < 1 - I1/O1$; Fig. 1) or nonexistent. If no purging has occurred, then inbreeding depression should be no different than in the first generation of inbreeding ($1 - IOC/OC = 1 - I1/O1$; Fig. 1). With the exception of the Willis (1999) study, all comparisons of changes in inbreeding depression reported here are between the first and last generation of inbreeding ($1 - I1/O1$ and $1 - I5/O5$ respectively; Fig. 1) and not between the first inbred generation and the inbred purged population ($1 - I1/O1$ and $1 - IOC/OC$ respectively; Fig. 1). Additionally, comparing inbreeding depression between the first inbred and purged population reduces the confounding effect of linkage disequilibrium (which can produce transient correlations between traits) that accumulates with generations of inbreeding (Falconer and MacKay 1996). We therefore strongly encourage future studies to incorporate the method first advocated by Willis (1999) to measure purging in experimentally inbred populations.

The other potential solution is among-population comparisons of inbreeding depression where varying inbreeding history can be accurately inferred. Knowing the inbreeding history of a population is critical since a population that has previously experienced severe inbreeding (due to a population bottleneck) is expected (with some qualifications) to have a lower genetic load (Lynch and Walsh 1998; Keller and Waller 2002). Between-population comparisons are most useful as an adjunct to experimental studies, as fixation of deleterious alleles is believed to be substantially slower in naturally inbred populations than experimentally manipulated ones. However, this approach possesses its own problems and it is questionable to what degree of accuracy one can infer the history of inbreeding for any given naturally inbred population (Keller and Waller 2002).

Conclusions

Unlike previous reviews on purging, results for the measures we calculated in this study provide evidence to suggest that purging occurs. In many cases the extent of rebounds in trait values can be substantial in experimentally inbred populations. However, this result must be viewed in light of certain qualifications. The frequency and magnitude of purging with inbreeding appears to depend upon the manner in which trait fitness components are experimentally assessed. Additionally, we found little correspondence between different measures of purging, most notably between changes in inbreeding depression and the use of outcrossed inbred lines. The discrepancy among the results we obtained warrants caution in making any general conclusions concerning the magnitude and frequency of purging in animal and plant populations, particularly when using the RATIO measure of purging. Although adaptation to laboratory conditions may account for rebounds in fitness, there is evidence for purging in a significant number of the studies reviewed here. Despite the large body of information accumulated during the past century pertaining to inbreeding depression, substantially more research is needed to understand the dynamics of purging and particularly to resolve the fundamental issue of the genetic basis of inbreeding depression.

Acknowledgments

This research was supported by a postdoctoral scholarship from Le Fonds Québécois de la Recherche sur la Nature et les Technologies to PC and a research grant from the Natural Sciences and Engineering Research Council of Canada to SCHB. We wish to thank J. Willis for insightful comments on an earlier version of this manuscript and two anonymous reviewers for their comments.

Literature Cited

Arnqvist, G., and D. Wooster. 1995. Meta-analysis: synthesising research findings in ecology and evolution. Trends Ecol. Evol. 10:236–240.

Ballou, J. D. 1997. Ancestral inbreeding only minimally affects inbreeding depression in mammalian populations. J. Hered. 88: 169–178.

Barrett, S. C. H., and D. Charlesworth. 1991. Effects of a change in the level of inbreeding on the genetic load. Nature 352: 522–524.

Barrett, S. C. H., and L. D. Harder. 1996. Ecology and evolution of plant mating. Trends Ecol. Evol. 11:73–79.

Bruce, A. B. 1910. The Mendelian theory of heredity and the augmentation of vigor. Science 32:627–628.

Byers, D. L., and D. M. Waller. 1999. Do plant populations purge their genetic load? Effects of population size and mating history on inbreeding depression. Annu. Rev. Ecol. Syst. 30:479–513.

Charlesworth, B., and D. Charlesworth. 1999. The genetic basis of inbreeding depression. Genet. Res. Camb. 74:329–340.

Charlesworth, B., M. Nordborg and D. Charlesworth. 1997. The effect of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet. Res. Camb. 70:155–174.

Charlesworth, D., and B. Charlesworth. 1987. Inbreeding depression and its evolutionary consequences. Annu. Rev. Ecol. Syst. 18:237–268.

Charlesworth, D., M. T. Morgan, and B. Charlesworth. 1990. Inbreeding depression, genetic load and the evolution of outcrossing rates in a multi-locus system with no linkage. Evolution 44: 1469–1489.

Crnokrak, P., and D. A. Roff. 1999. Inbreeding depression in the wild. Heredity 83:260–270.

Crow, J. F. 1970. Genetic loads and the cost of natural selection. Pp. 128–177 *in* K. I Kojima, ed. Mathematical topics in population genetics. Springer, Berlin.

———. 1999. The rise and fall of overdominance. Plant Breed. Rev. 17:225–257.

Crow, J. F., and M. Kimura. 1970. An introduction to population genetics theory. Harper and Row, New York.

Crow, J. F., and M. J. Simmons. 1983. The mutation load in *Drosophila*. Pp. 1–35 *in* H. L. C. M. Ashburner and J. N. Thomson, eds. The genetics and biology of *Drosophila*. Academic Press, London.

Darwin, C. 1868. The variation of animals and plants under domestication. Vol. 1. Appleton, New York.

———. 1876. The effects of cross and self-fertilization in the vegetable kingdom. John Murray, London.

Davenport, C. B. 1908. Degeneration, albinism and inbreeding. Science 28:454–455.

DeRose, M. A., and D. A. Roff. 1999. A comparison of inbreeding depression in life-history and morphological traits in animals. Evolution 53:1288–1292.

de Vienne, D., B. Bost, J. Fievet, M. Zivy, and C. Dillmann. 2001. Genetic variability of proteome expression and metabolic control. Plant Physiol. Biochem. 39:271–283.

Dudash, M. R. 1990. Relative fitness of selfed and outcrossed progeny in a self-compatible, protandrous species, *Sabatia angularis* L. (Gentianaceae): a comparison in three environments. Evolution 44:1129–1139.

East, E. M. 1908. Inbreeding in corn. Rep. Conn. Agric. Exp. Stn. 1907:419–429.

East, E. M., and D. F. Jones. 1919. Inbreeding and outbreeding: their genetic and sociological significance 147. Lippincott, Philadelphia, PA.

Falconer, D. S., and T. MacKay. 1996. Introduction to quantitative genetics. 4th ed. John Wiley and Sons, New York.

Hedges, L. V., and I. Olkin. 1985. Statistical methods for meta-analysis. Academic Press, Orlando, FL.

Hedges, L. V., J. Gurevitch, and P. S. Curtis. 1999. The meta-analysis of response ratios in experimental ecology. Ecology 80: 1150–1156.

Hedrick, P. W. 1994. Purging inbreeding depression and the probability of extinction: full-sib mating. Heredity 73:363–372.

Husband, B. C., and D. W. Schemske. 1995. Evolution of the magnitude and timing of inbreeding depression in plants. Evolution 50:54–70.

Jones, D. F. 1917. Dominance of linked factors as a means of accounting for heterosis. Genetics 46:465–479.

Kalinowski, S. T., and P. W. Hedrick. 2001. Inbreeding depression in captive bighorn sheep. Anim. Conserv. 4:319–324.

Kalinowski, S. T., P. W. Hedrick, and P. S. Miller. 2000. Inbreeding depression in the Speke's gazelle captive breeding program. Conserv. Biol. 14:1375–1384.

Karkkainen, K., H. Kuittinen, R. van Treuren, C. Vogl, S. Oikarinen, and O. Savolainen. 1999. Genetic basis of inbreeding depression in *Arabis petraea*. Evolution 53:1354–1365.

Keeble, F., and C. Pellew. 1910. The mode of inheritance of stature and of time of flowering in peas (*Pisum sativum*). J. Genet. 1: 47–56.

Keller, L., and D. M. Waller. 2002. Inbreeding effects in wild populations. Trends Ecol. Evol. 17:230–241.

Knight, T. 1799. An account of some experiments on the fecundation of vegetables. Philos. Trans. R. Soc. Lond. 89:195–204.

Lacy, R. C., A. Petric, and M. Warneke. 1993. Inbreeding and outbreeding in captive populations of wild animals. Pp. 352–374 *in* N.W. Thornhill, ed. The natural history of inbreeding and outbreeding: theoretical and empirical perspectives. Univ. of Chicago Press, Chicago, IL.

Lacy, R. C., and J. D. Ballou. 1998. Effectiveness of selection in reducing the genetic load in populations of *Peromyscus polionotus* during generations of inbreeding. Evolution 52:900–909.

Lande, R., and D. W. Schemske. 1985. The evolution of self-fer-

tilization and inbreeding depression in plants. I. Genetic models. Evolution 39:24–40.
Li, Z.-K., L. J. Luo, H. W. Mei, D. L. Wang, Q. Y. Shu, R. Tabien, D. B. Zhong, C. S. Ying, J. W. Stansel, G. S. Khush, and A. H. Paterson. 2001. Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield. Genetics 158:1737–1753.
Luo, L. J., Z.-K. Li, H. W. Mei, Q. Y. Shu, R. Tabien, D. B. Zhong, C. S. Ying, J. W. Stansel, G. S. Khush, and A. H. Paterson. 2001. Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. II. Grain yield components. Genetics 158:1755–1771.
Lynch, M., and B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer, Sunderland, MA.
Omholt, S. W., E. Plahte, L. Oyehaug, and K. Xiang. 2000. Gene regulatory networks generating the phenomena of additivity, dominance and epistasis. Genetics 155:969–980.
Ralls, K., and J. Ballou. 1986. Captive breeding programs for populations with a small number of founders. Trends Ecol. Evol. 1: 19–22.
Rao, P. S. S., and S. G. Inbaraj. 1977. Inbreeding effects on human reproduction in Tamil Nadu of South India. Ann. Hum. Genet. Lond. 41:87–98.
Ritland, K. 1990. Inferences about inbreeding depression based on changes of the inbreeding coefficient. Evolution 44:1230–1241.
———. 1996. Inferring the genetic basis of inbreeding depression in plants. Genome 39:1–8.
Roff, D. A. 2002. Inbreeding depression: tests of the overdominance and partial dominance hypotheses. Evolution 56:768–775.
Rosenberg, M. S., D. C. Adams, and J. Gurevitch. 2000. MetaWin. Statistical software for conducting meta-analysis: fixed effect models, mixed effect models, and resampling tests. Ver. 2.0. Sinauer, Sunderland, MA.
Schultz, S. T., and J. H. Willis. 1995. Individual variation in inbreeding depression: the roles of inbreeding history and mutation. Genetics 141:1209–1223.
Shields, W. M. 1987. Dispersal and mating systems: investigating their causal connections. Pp. 156–213 *in* B. D. Chepko-Sade and Z. T. Halpin, eds. Mammalian dispersal patterns: the effects of social structure on population genetics. Univ. of Chicago Press, Chicago, IL.
Shull, G. H. 1908. The composition of a field of maize. Ann. Breed. Assoc. 4:296–301.
Simmons, M. J., and J. F. Crow. 1977. Mutations effecting fitness in *Drosophila* populations. Annu. Rev. Genet. 11:49–78.
Sokal, R. R., and F. J. Rohlf. 2000. Biometry. 3rd ed. W.H. Freeman and Company, New York.
Uyenoyama, M. K., K. E. Holsinger, and D. M. Waller. 1993. Ecological and genetic factors directing the evolution of self-fertilization. Oxf. Surv. Evol. Biol. 9:327–381.
Visscher, P. M., D. Smith, S. J. G. Hall, and J. L. Williams. 2001. A viable herd of genetically uniform cattle. Nature 409:303.
Waller, D. M. 1986. Is there disruptive selection for self-fertilization? Am. Nat. 128:421–426.
Wang, J., W. G. Hill, D. Charlesworth, and B. Charlesworth. 1999. Dynamics of inbreeding depression due to deleterious mutations in small populations: mutation parameters and inbreeding rate. Genet. Res. Camb. 74:165–178.
Willis, J. H. 1993. Effects of different levels of inbreeding on fitness components in *Mimulus guttatus*. Evolution 47:864–876.
———. 1999. The role of genes of large effect on inbreeding depression in *Mimulus guttatus*. Evolution 53:1678–1691.
Wright, S. 1977. Evolution and the genetics of populations. Vol. 3. Experimental results and evolutionary deductions. Univ. of Chicago Press, Chicago, IL.

Corresponding Editor: J. Mitton