Genetic load, inbreeding depression, and hybrid vigor covary with population size:
An empirical evaluation of theoretical predictions

Jennifer N. Lohr1,2,3,4 and Christoph R. Haag1,2,5
1Department of Biology, Ecology and Evolution, University of Fribourg, 1700, Fribourg, Switzerland
2Zoological Institute and Zoological Museum, Biocenter Grindel, Hamburg University, Hamburg, Germany
3Tvärminne Zoological Station, FIN-10900, Hanko, Finland
4E-mail: jennifer.lohr@uni-hamburg.de
5Centre d’Ecologie Fonctionnelle et Evolutive – UMR 5175, 34293, Montpellier Cedex 5, France

Received May 12, 2015
Accepted October 14, 2015

Reduced population size is thought to have strong consequences for evolutionary processes as it enhances the strength of genetic drift. In its interaction with selection, this is predicted to increase the genetic load, reduce inbreeding depression, and increase hybrid vigor, and in turn affect phenotypic evolution. Several of these predictions have been tested, but comprehensive studies controlling for confounding factors are scarce. Here, we show that populations of Daphnia magna, which vary strongly in genetic diversity, also differ in genetic load, inbreeding depression, and hybrid vigor in a way that strongly supports theoretical predictions. Inbreeding depression is positively correlated with genetic diversity (a proxy for \( N_e \)), and genetic load and hybrid vigor are negatively correlated with genetic diversity. These patterns remain significant after accounting for potential confounding factors and indicate that, in small populations, a large proportion of the segregation load is converted into fixed load. Overall, the results suggest that the nature of genetic variation for fitness-related traits differs strongly between large and small populations. This has large consequences for evolutionary processes in natural populations, such as selection on dispersal, breeding systems, ageing, and local adaptation.

**KEY WORDS:** Daphnia, drift load, genetic drift, heterosis, segregation load.

Anthropogenic habitat loss and habitat fragmentation are the main causes of population and species extinction, particularly for already threatened species. However, even in undisturbed sites extinction may occur due to demographic or environmental stochasticity (Lande 1993; Beissinger and Westphal 1998; Hanski 1998; Willi et al. 2006). These factors have a greater impact on small compared to large populations, making population size a crucial factor for population persistence over time (Flather et al. 2011).

In addition to the ecological factors endangering small and isolated populations, there are also genetic risks associated with small population size, mainly due to genetic drift (Frankham 1995; Lande 1995; Lynch et al. 1995). Genetic drift, the random change in allele frequencies due to the sampling of a finite number of gametes each generation, is present in all finite populations, but its strength depends inversely on population size (\( 1/2N_e \) measures the strength of drift in diploid populations, where \( N_e \) is the effective population size; Nei and Tajima 1981). Genetic drift reduces the standing genetic variation, which may be important for adaptation; for instance upon environmental change (Lande 1993). It also limits selection in favor of beneficial alleles and against deleterious alleles, as alleles with a positive or negative selection coefficient \( s \) of substantially less than \( 1/2N_e \) are effectively neutral (Hartl and Clark 1997). Hence, weakly deleterious alleles can accumulate by drift in small populations. Finally, even for alleles
with higher selection coefficients, genetic drift increases the variance in allele frequency, which leads to increased homozygosity of some of these alleles. This is important because most deleterious alleles are recessive or partly recessive and hence increasing the variance in their frequencies will lead, on average, to a fitness decline (Wright 1977; Charlesworth and Charlesworth 1999). The fitness reduction due to accumulating weakly deleterious alleles and increased homozygosity is the “drift load” (i.e., one form of genetic load). It can be broken down into fixed load (fitness decrease due to weakly deleterious alleles that have become fixed by drift) and segregating load (fitness decrease due to segregating deleterious alleles; Whitlock et al. 2000).

Besides leading to genetic load and maladaptation (e.g., increased ageing) in small populations, genetic drift can also play a creative role by setting the stage for adaptations to limit its negative effects (Charlesworth and Charlesworth 1987; Lenormand et al. 2009). For instance, the evolution of high dispersal rates may be advantageous in populations that experience high drift load because offspring of matings between immigrants and local individuals benefit from hybrid vigor (Ingvarson and Whitlock 2000; Ebert et al. 2002; Ronce 2007). Similarly, the advantage of obligate outcrossing (i.e., limiting the negative effects of inbreeding) also varies with the level of genetic drift, and hence population size may be an important factor modulating the costs and benefits of particular breeding systems (Glémin and Ronfort 2013). Many of the evolutionary consequences of reduced population size are explained via the expected effects of genetic drift on the genetic load, inbreeding depression, and hybrid vigor. The theoretical predictions on these effects have been worked out in detail (Whitlock et al. 2000; Whitlock 2002; Glémin et al. 2003; Paland and Schmid 2003; Roze and Rousset 2004), but conclusive empirical studies are scant (Mattila et al. 2012; Willi et al. 2013). In this article, we test these predictions empirically in a system in which \( N_i \) is known to vary strongly among populations of a single species.

Inbreeding depression refers to the reduced fitness of inbred offspring relative to offspring produced by random mating (Charlesworth and Charlesworth 1987; Keller and Waller 2002). Inbreeding depression is caused by the increased homozygosity of recessive and partially recessive deleterious alleles, as well as overdominant loci and certain forms of epistatic interactions (Charlesworth 1998; Charlesworth and Charlesworth 1999; Roff 2002; Carr and Dudash 2003). The fitness decrease upon inbreeding indicates that the majority of populations are polymorphic at fitness-determining loci. As a result inbreeding leads to increased homozygosity, demonstrating that homozygotes at these loci are on average less fit than heterozygotes. Theory predicts that the magnitude and genetic basis of inbreeding depression should vary systematically with population size and population structure (Glémin et al. 2003; Roze and Rousset 2004). In large populations, the frequency of deleterious mutations is expected to be close to mutation-selection balance, and allele frequencies at overdominant loci should be determined by the relative fitness of the different homozygotes (Hartl and Clark 1997). In small populations, partial purging of recessive-deleterious alleles may occur when they are exposed in homozygote form. This is especially true for lethal or semilethal alleles segregating in the population (Nei 1968; Wang et al. 1999; Kirkpatrick and Jarne 2000; Hedrick 2002; Glémin 2003). In addition, mildly deleterious alleles may become fixed by drift and variation at overdominant loci may be lost (Lynch et al. 1995; Bataillon and Kirkpatrick 2000; Glémin 2003). Because fixed alleles do not contribute to the fitness differences between inbred and randomly mated individuals within populations, theory therefore predicts low inbreeding depression in small populations. All these predictions use the local population as the reference for defining inbreeding, and inbreeding is measured as the excess of homozygotes compared to Hardy–Weinberg expectations within populations. Thus, even individuals from the smallest populations are considered noninbred as long as mating occurs at random. Such individuals are predicted to be homozygous at a larger number of loci than individuals from large populations due to drift rather than inbreeding. Defining the reference population precisely is important because in the literature all individuals from small populations are sometimes considered “inbred” with respect to the metapopulation, even under random mating (e.g., “remote inbreeding,” Hartl and Clark 1997).

Hybrid vigor (or heterosis) is the increase in fitness produced by outcrossing individuals from different populations, relative to individuals produced by random mating within populations. It occurs because outcrossing between populations results in increased heterozygosity (compared to parental populations). Recessive deleterious alleles become masked in a heterozygous state, which results in increased offspring fitness. It is not exactly the opposite of inbreeding depression, due to the strong contribution of fixed deleterious mutations (fixed within one but not in the other population used for outcrossing), and because it is, on average, caused by deleterious alleles of milder effect than inbreeding depression (Charlesworth and Charlesworth 1987; Whitlock et al. 2000). In addition, hybrid vigor may also be caused by alleles segregating at different frequencies in the two parental populations. Hybrid vigor, like inbreeding depression, is predicted to vary with population size and population structure (Whitlock et al. 2000; Glémin et al. 2003; Roze and Rousset 2004). Specifically, outcrossing between large populations is predicted to result in little hybrid vigor, whereas outcrossing between small populations (high variation in allele frequencies and high likelihood of fixation) is predicted to result in high amounts of hybrid vigor (Glémin et al. 2003). Theory also predicts a strong positive relationship between pairwise \( F_{ST} \) and hybrid vigor, as pairwise \( F_{ST} \) directly measures the difference in allele frequencies (including
Several empirical studies on an array of different organisms have attempted to test these or part of these theoretical predictions (Keller and Waller 2002; Paland and Schmid 2003; Reed 2005; Leimu et al. 2006; Willi et al. 2006; Escobar et al. 2008; Mattila et al. 2012; Mullarkey et al. 2013; Willi et al. 2013; Hoffman et al. 2014; Pekkala et al. 2014). While the results are generally in line with the predictions, only few studies have assessed inbreeding depression and hybrid vigor in parallel using direct comparisons of the parents with their inbred and outbred offspring. Such comprehensive studies are needed in order to gain a full understanding of the nature of the genetic load (magnitude, fixed load, segregating load; Paland and Schmid 2003). Moreover, in previous studies, the effects were often found to vary among traits (e.g., among different life-history traits or other traits expected to be correlated with fitness), and the contribution of trade-offs and local adaptation to these differences remains unclear; especially in studies where two habitat types (with differences in average population sizes) were compared. We therefore present results on the average values of life-history traits of parents along with estimates of inbreeding depression and hybrid vigor from eight populations of *Daphnia magna* that have been shown to vary along a gradient in neutral genetic diversity and effective population size (Walser and Haag 2012). The populations come from two types of habitats (small rock pools and larger ponds), but we have previously found no evidence for trade-offs between life history traits among these populations (Lohr et al. 2014; a result that was confirmed in the present study), and we also statistically corrected for population type in our analyses. Using populations along a gradient of effective population sizes has the advantage that not only small versus large populations can be compared, but rather, the theoretical predictions can be tested across a continuous range of populations. In addition, the *Daphnia* system is highly manipulable: we were able to create large numbers of inbred and outbred lines from the different *Daphnia* populations. This allowed us to directly compare parental clones with their outcrossed and inbred offspring and to estimate inbreeding depression, hybrid vigor, and genetic load using large sample sizes for many populations of a single species.

### Materials and Methods

#### ORIGIN AND COLLECTION OF EXPERIMENTAL ANIMALS

*Daphnia* are important zooplankton species in freshwater ponds and lakes around the world. They are cyclical parthenogens, meaning they reproduce mainly via asexual reproduction throughout the growing season. A switch to sexual reproduction is triggered environmentally, and sexual reproduction always results in rest-
representative and unbiased random sample of each population. In addition, only populations that undergo a minimum of one sexual phase per year were chosen, in order to avoid any effects of prolonged clonality. The collection of animals was done between 2007 and 2009, after which the clones were maintained as asexual lines in the laboratory until the beginning of the experiment in 2013. Cultures were reared in artificial Daphnia medium (ADaM; Klüttgen et al. 1994), fed daily with the unicellular algae Scenedesmus sp., and kept at 20°C with a 12L:12D photoperiod.

**ESTIMATION OF GENETIC DIVERSITY AND GENETIC DIFFERENTIATION**

Here, we use expected heterozygosity, a measure of genetic diversity, as a proxy for the effective population size. This is an estimate of relative (rather than absolute) long-term effective population size (i.e., relative to other populations included in the study), based on presumably neutral microsatellite markers. It includes the effects of both currently ongoing drift and drift during colonization bottlenecks. Smaller populations have increased drift both due to higher turnover (small bottlenecks, recent colonization) and reduced census size compared to large populations (Walser and Haag 2012). The microsatellite genotyping was done as described previously (Walser and Haag 2012; Lohr et al. 2014). Briefly, DNA was extracted from adult Daphnia using a hotspot protocol modified for zooplankton (Montero-Pau et al. 2008), and 5–10 individuals per population were scored at 32 microsatellite markers, divided into four multiplex sets to assess genetic diversity. We used expected heterozygosity ($H_E$) as a measure of genetic diversity and also as a correlate of effective population size ($N_E$), as it has been shown to be robust to small sample sizes (Nei 1978; Nei and Takahata 1993). We estimated $H_E$ and $F_{ST}$ (fixation index) using Arlequin version 3.5. The genotype data used in this study are the same as those in Lohr et al. (2014).

**GENERATION OF INBRED AND OUTBRED LINES**

To create outbred lines, clones from each of the eight populations (called hereafter focal populations) were outcrossed with clones from nearby populations of similar genetic diversity and size (hereafter termed outcross partners; Table 1). One population (AST) was used in two pairs (once as focal population, once as outcross partner); all other populations were used in one pair only. For each focal population, four clones were used as parental clones for outcrossing. Each clone was crossed to only one clone from the paired population (i.e., we had four pairs of clones per pair of populations). Furthermore, each of the parental clones from the focal populations was also used for inbreeding.

To generate outbred and inbred offspring clones, single clone cultures were established in the laboratory in 10 L plastic buckets. A single clone culture was generated for all focal population clones, as well as for all outcross population clones. These bucket cultures were reared for approximately 1 month, until they had reached high densities. Outcrosses were performed by combining the appropriate single clone cultures of focal and outcross clones. This was done by introducing ~100 females of each of the two clones together into a new bucket. Inbred offspring were generated using the original single clone cultures, which were left during two months to produce "self-fertilized" ephippia (by sexual mating between genetically identical males and females from the same clone; sex in Daphnia is environmentally induced, with no genetic differences between males and females). The cultures were fed unicellular algae daily, fresh medium was added to the buckets one time per week, and the buckets were kept at 20°C with a 16:8 light/dark photoperiod. The outcross cultures were treated in the same way, with sexual reproduction leading to inbred as well as outcrossed ephippia. The buckets were then left to dry out naturally, after which they were stored in the dark at 2°C for two months to ensure the ephippia had experienced the proper signal to exit diapause upon rehydration. Rehydration was done at 20°C by the addition of 8L of artificial Daphnia medium and the photoperiod was set to 16:8 light/dark. The buckets were checked regularly for hatchlings, which were then collected and placed individually in 50 ml tubes to establish clonal lines. Hatchling lines were verified as inbred or outbred using preselected informative microsatellite loci.

**LIFE-HISTORY TRAIT MEASUREMENT**

The experimental design was centered on the eight focal populations (Parental 1, P1) with their outcross partners (Parental 2, P2) and their respective inbred (I) and outbred (O) offspring clones. Per pair of populations (P1, P2), we used four clones from each of P1, P2, O, and I and 25 replicates per clone. This resulted in a total of 3200 individuals for which life-history traits were recorded. All treatments in this life-table were run concurrently, as one single experiment.

To begin, adult females were taken from the stock cultures and placed singly into 50 ml tubes. Each individual was fed 100 μL of unicellular algae (Scenedesmus sp.) per day and the medium was refreshed (i.e., the individual was changed into a new tube) three times per week (Mondays, Wednesdays, and Fridays). Throughout the experiment the Daphnia were kept in ADaM at 20°C with a 12L:12D photoperiod. Clonal offspring were discarded during these changes. The animals were passed through three generations under these standard experimental conditions in order to remove potential maternal and grand-maternal effects (Ebert 1993). Specifically, once a mother released her third clutch, one offspring from that clutch was chosen to start the next generation, and this procedure was repeated three times. Some replicates were lost during the three preexperimental conditions, but as this was anticipated, we had begun with double the number of
individuals required (i.e., 50 instead of 25 individuals per clone). Day zero of the experiment was when the fourth-generation offspring (third-clutch offspring of the third-generation females) were isolated into new tubes. Due to the excess of preexperimental replicates, it was possible to start the entire experiment (i.e., to have fourth-generation offspring of a sufficient number of mothers of each clone) during a time window of only 48 hours. These fourth-generation offspring were used as the experimental cohort (i.e., the life-history traits of these animals were recorded). The variation in birthing time was taken into account in the data analysis by subtracting one day from the age of animals born between 24 and 48 hours after the start of the experiment. In addition, only females were used in this experiment: The few fourth-generation male offspring were discarded and female offspring were used instead (environmentally induced male production was low in all clones under the experimental conditions).

For the duration of the experiment (i.e., from day zero up to the death of the last individual), the same experimental conditions as those outlined above were used. At each changing point during the experiment, the offspring produced by each individual were counted and discarded, as were any dead mothers. Recorded data were age at death (the day on which an individual was found dead), age at each reproduction event (the day on which clutches were released), and the number of offspring for each reproduction event. An earlier study had shown that no trade-off exists between the number and size of offspring across these populations (Lohr et al. 2014). In order to streamline the experimental procedures, we used falcon tubes in racks of 25, ordered in a way that they did not require individual labeling. On days between changes, we washed all dirty tubes from the previous day and prefilled them with ADaM. Then, on the days of changes, using a glass pipette we transferred *Daphnia* between racks of falcon tubes (to the identical position on the rack), which required only three seconds per *Daphnia*. Offspring were counted once all adults had been transferred to new racks, either on the day of changing or on the following morning, in parallel with starting the washing.

The following life-history traits were used in the analyses: age at death (in days), age at maturity (age at first reproduction in days), and reproductive output (total number of offspring produced). The intrinsic rate of increase, \( r \), was calculated from the life-table data by iteration of Newton’s approximation: \( 1 = \sum I_i m_i e^{-rt} \) (Lotka 1956; Desmaris and Tessier 1999).

**GENETIC LOAD, INBREEDING DEPRESSION, AND HYBRID VIGOR**

Inbreeding depression was estimated as \( \delta = 1 - W_{IN} / W_P \), where \( W_{IN} \) is the average trait value of the inbred individuals and \( W_P \) is the average trait value of the parents (Glémin et al. 2003). Hybrid vigor was estimated as \( H = 1 - (W_{P1} + W_{P2})/2W_{OUT} \), where \( W_{P1} \) is the average trait value of the focal parents, \( W_{P2} \) the average trait value of the outcross parents and \( W_{OUT} \) the average trait value of outcrossed (hybrid) individuals between these two parents. Finally, genetic load was calculated using \( L = 1 - W_P / W_{MAX} \), where \( W_{MAX} \) is the maximal trait value, usually defined as that of a hybrid of hybrid individuals. However, as hybrid-free genotypes are very unlikely to exist in nature (making absolute genetic load very difficult to assess empirically; Agrawal and Whitlock 2012), we used the maximal observed trait value. This assumes that the clone with the highest trait value (or lowest, in the case of age at maturity, see below) has a genetic load of zero. Hence, all our estimates of the genetic load are underestimates, but the relative load among different clones and populations should not be affected (Agrawal and Whitlock 2012). For age at maturity, a younger age was assumed to be positively correlated with fitness. Hence, inbreeding depression, hybrid vigor, and genetic load were estimated as:

\[
\delta = \frac{W_{IN} - W_P}{W_P} \quad \text{and} \quad H = \frac{(W_{P1} + W_{P2})}{2W_{OUT}} - 1, \quad L = \frac{W_P - W_{MAX}}{W_{MAX}} - 1, \quad \text{where} \quad W_{MAX} \quad \text{is the clone with the earliest age at maturity. This transformation could theoretically (but not in our data) produce estimates of \( \delta \), \( H \), and \( L > 1 \) (if one clone was more than twice as old at maturity than another clone).}
\]

However, the meaning remains the same: \( \delta = 0.2 \), for instance, means that the inbred offspring took 20% longer (not shorter) to maturity compared to their parents. All averages trait values were calculated across replicates of the same clone and one estimate of \( L \) was obtained for each parent clone, one estimate of \( \delta \) for each pair of parent and inbred offspring clones, and one estimate of \( H \) for each triplet of two parent and one hybrid offspring clone. Population-wide estimates of \( \delta \), \( H \), and \( L \) were then obtained by averaging these single estimates within populations (or pair of populations in the case of \( H \)).

Throughout our experiment, we used parents rather than random within-population crosses as the reference for assessing inbreeding depression, hybrid vigor, and genetic load. Including within-population crosses was technically not possible due to the size of the experiment and a dearth of discriminant markers (homologous for alternative alleles) to distinguish outcrossed versus inbred individuals in crosses within small populations (using mass cultures, i.e., the same methodology as for the crosses between populations). Such crosses would be needed to fully exclude the possibility that parts of our results could be explained by parents from small populations already being inbred because of a higher rate of deliberate within-clone mating (or other sources of inbreeding by nonrandom mating) in small populations. However, this idea is inconsistent with the reproductive biology of *Daphnia*. Mate choice in *Daphnia* is based mainly on random contact and, if anything, there is a low degree of inbreeding avoidance (De Meester and Vanoverbeke 1999; Winsor and Innes 2002). The sample sizes used for this experiment (five clones per population) were not large enough to estimate deviations from Hardy-Weinberg in any meaningful way, especially in small
populations in which these estimates would be based on only a few polymorphic loci. However, previous studies, including ones based on the same focal populations (Walser and Haag 2012), found no evidence for homozygote excess in populations of *Daphnia magna*, including in small rock pool populations (Haag et al. 2006). Hence, all evidence points to systematic inbreeding being rare or absent in *Daphnia*. Therefore we argue that, parents (sampled immediately after hatching from resting eggs) are a valid substitute for offspring of random within-population crosses, given the technical limitations to produce such offspring.

**STATISTICAL ANALYSES**

The relationships between genetic diversity and the traits age at death, reproductive output, age at maturity, the intrinsic rate of increase, as well as the inbreeding depression, hybrid vigor, and genetic load of these traits were evaluated using linear-mixed models in R, using the nlm package (R Development core team 2013). Two separate models were run: one to investigate inbreeding depression, thus comparing inbred with parental clones (only focal populations), and one to investigate hybrid vigor, thus comparing outbred with all parental clones.

In the first model, we included genetic diversity of the focal populations and breeding type (parental P1 and inbred) as fixed factors, and population and clone as random factors. To account for data structure, each specific parent clone and its inbred offspring clone was assigned a “pair_of_clones” label, and pair_of_clones was used as an additional random factor. The basic form of the model was: trait ~ genetic diversity*breeding type, random = ~ (1+breed|pair_of_clones)+(1|pop)+(1|clone).

In this model, the interaction between breeding type and genetic diversity tests whether the relative performance of the parents and inbreds (i.e., inbreeding depression) changes with genetic diversity.

In the second model, breeding type (parental, hybrid) was used without distinction between P1 and P2 and tested for an interaction with average genetic diversity between the two parental populations. In addition, each specific focal population and its corresponding outcross partner was assigned a “triplet_of_pops” label, with the third “population” of the triplet being the hybrids between the two parental populations. Furthermore, “triplet_of_clones” labels were assigned to each triplet consisting of two parental clones and their outcrossed offspring clone. The basic form of the model was: trait ~ genetic diversity*breed, random = ~ (1 + breed|triplet_of_pops) + (1 + breed|triplet_of_clones) + (1|pop) + (1|clone).

In this model, the interaction between breeding type and genetic diversity tests whether the relative performance of the average of the parents (across both parental populations) and the hybrids (i.e., hybrid vigor) changes with genetic diversity.

Because only one estimate of the intrinsic growth rate \( r \) was obtained per clone, the random term clone was dropped from the models investigating \( r \). Similarly, one estimate of genetic load was obtained for each clone, one estimate of inbreeding depression for each pair_of_clones, and one estimate of hybrid vigor for each triplet_of_clones. In addition, the effect of breeding type was already included in the estimate. Thus, we used the three simplified models to investigate these variables, including only one fixed factor (genetic diversity) and one random factor ((1|pop) for inbreeding depression and genetic load, and (1|triplet_of_pops) for hybrid vigor). In the case of hybrid vigor, the average genetic diversity between the two parental populations (focal, outcross partner) was used.

To make sure that any observed correlations with genetic diversity were not only due to a difference between the four small populations (from small rock pools in Northern Europe) and the four large populations (from larger ponds in central Europe), we added a fixed factor “pond size class” to our standard models and tested whether the effect of genetic diversity remained significant. Note that this test is necessarily more conservative because it tests for a correlation with genetic diversity within the two groups after the difference between groups has been removed. For each trait we inspected plots of the residuals to ensure no skew or pattern. Plots of the response variables were inspected to ensure the relationship was linear, and finally, quantile–quantile plots per population were inspected to ensure each was reasonably normally distributed.

**Results**

**PARENTAL POPULATIONS**

As already found in our previous work, individuals from small, genetically less diverse populations had substantially (up to 50%) reduced average lifespans compared to those from large, more diverse populations (Lohr et al. 2014; Fig. 1). Across populations, there was a positive relationship between lifespan and genetic diversity. We also found a weak positive linear relationship with genetic diversity for average daily reproduction and a negative linear relationship for age at maturity. Furthermore, there was a strong positive relationship between genetic diversity and the intrinsic rate of increase (Lohr et al. 2014; Fig. 1).

**EFFECT OF INBREEDING AND OUTBREEDING**

In contrast to the parental clones, the life-history traits of the inbred and outbred clones showed little relationship with genetic diversity (Fig. 1, Table 2). Instead there were large differences between the breeding types (parental, inbred, outbred), and strong interactions between genetic diversity and breeding type. Specifically, in the smaller populations there was very little inbreeding depression but a large amount of hybrid vigor, whereas in the large
populations the opposite was true: strong inbreeding depression and low hybrid vigor (Fig. 2; Table 2). In fact, the large populations experienced even some outbreeding depression, with the average hybrid offspring values being slightly lower than those of the parents (Figs. 1 and 2). The interactions between breeding type and genetic diversity remained significant also in the model including pond size class as a factor (Table 2).

These fitness differences translated into estimates of inbreeding depression, hybrid vigor, and the genetic load that did strongly covary with genetic diversity (Figs. 3 and 4, Table 3). Overall, we found positive linear relationships between genetic diversity and inbreeding depression and negative linear relationships between genetic diversity and hybrid vigor and between genetic diversity and genetic load (Table 3). We also found a strong positive relationship between the pairwise genetic differentiation ($F_{ST}$) and hybrid vigor (linear-mixed model results: age at death: $t = 4.50, df = 6,23, P = 0.004$; age at maturity: $t = 3.51, df = 6,23, P = 0.013$; reproductive output: $t = 11.28, df = 6,23, P < 0.001$; intrinsic growth rate: $t = 7.46, df = 6,23, P < 0.001$, Fig. 5).

**Figure 1.** Age at death, age at maturity, reproductive output, and intrinsic growth rate of parental (■), outbred (♦), and inbred (○) clones. Panels A–D compare the parentals of each focal population with their inbred progeny, while panels E–H compare the averages of each pair of parental populations with their outbred progeny. Error bars represent 1 standard error of the clonal means.
Table 2. Results of the linear-mixed effect models testing for a relationship between genetic diversity and breeding type ("breed") with the life-history traits age at death, age at maturity, reproductive output, and intrinsic growth rate.

<table>
<thead>
<tr>
<th>trait</th>
<th>Inbreeding depression</th>
<th>Hybrid vigor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Parent versus inbred</td>
<td>Parents versus outbred</td>
</tr>
<tr>
<td></td>
<td>$t$</td>
<td>df</td>
</tr>
<tr>
<td>Age at death</td>
<td>1257</td>
<td>1927</td>
</tr>
<tr>
<td>Genetic diversity</td>
<td>1.11</td>
<td>6</td>
</tr>
<tr>
<td>Breed</td>
<td>-2.55</td>
<td>1223</td>
</tr>
<tr>
<td>Genetic diversity * breed</td>
<td>9.94</td>
<td>1223</td>
</tr>
<tr>
<td>Age at maturity</td>
<td>1257</td>
<td>1927</td>
</tr>
<tr>
<td>Genetic diversity</td>
<td>-1.34</td>
<td>6</td>
</tr>
<tr>
<td>Breed</td>
<td>-0.33</td>
<td>1223</td>
</tr>
<tr>
<td>Genetic diversity * breed</td>
<td>-5.02</td>
<td>1223</td>
</tr>
<tr>
<td>Reproductive output</td>
<td>1257</td>
<td>1927</td>
</tr>
<tr>
<td>Genetic diversity</td>
<td>1.40</td>
<td>6</td>
</tr>
<tr>
<td>Breed</td>
<td>-7.46</td>
<td>1223</td>
</tr>
<tr>
<td>Genetic diversity * breed</td>
<td>20.91</td>
<td>1223</td>
</tr>
<tr>
<td>Intrinsic growth rate</td>
<td>64</td>
<td>95</td>
</tr>
<tr>
<td>Genetic diversity</td>
<td>2.95</td>
<td>6</td>
</tr>
<tr>
<td>Breed</td>
<td>-1.20</td>
<td>30</td>
</tr>
<tr>
<td>Genetic diversity * breed</td>
<td>4.75</td>
<td>30</td>
</tr>
</tbody>
</table>

Shown are the sample sizes $N$, $t$-ratios, degrees of freedom (df), and associated $P$-values for each trait. A star indicates that the factor remained significant ($P < 0.05$), even when including pond size class and the interaction between breeding type and pond size class in the model. Note that the df refer to those used in the $t$-test, that is, the residual df of the model. $N$ refers to the number of individuals tested for the traits age at death, age at maturity, and reproductive output. However, in the case of the intrinsic growth rate, $N$ refers to the total number of clones as the intrinsic growth rate is estimated per clone from a cohort of individuals.

Discussion

Our results are in clear support of theoretical models based on recurrent mutation to unconditionally deleterious alleles on the effects of population size on inbreeding depression, hybrid vigor, and genetic load. This study is the first to find such clear and unequivocal evidence for all of the predicted effects. The following points make our results robust and should allow the findings presented here to be applied far beyond the Daphnia system: (1) we have investigated a system with particularly strong and continuous variation in effective population sizes within a single species; (2) Daphnia are clonal and hence we were able to use large sample sizes with independent statistical replicates of each genotype; (3) we could remove or control for the majority of the potential confounding factors normally encountered in such studies; such as environmental heterogeneity and differences in evolutionary (e.g., colonization) history among populations or species. Our “small” populations all occurred in small rock pools in Northern Europe and our “large” populations in larger ponds in central Europe. However, the effects remained significant even after accounting for pond size class (and therefore implicitly for rock pool versus nonrock pool and for Southern versus Northern latitudes) in the statistical models; (4) our experimental design addressed the alternative hypothesis of local adaptation coupled with trade-offs for the differences in life-history traits among populations. Indeed, interpreting life-history differences among populations in terms of genetic load is difficult because certain traits may show local adaptation (e.g., shortened lifespan in small populations as a response to increased external mortality or greater environmental stochasticity). Other traits may then show differences among populations due to trade-offs, which are well documented in a variety of systems (Keller and Genoud 1997; Abrams 2004; Reznick et al. 2004; Chen and Maklakov 2012). In our study, there were no indications of trade-offs among life-history traits: Individuals from small populations started to reproduce later, had a lower rate of increase, lower fecundity, and a shorter life span than individuals from large populations. Moreover, if any of these trait differences were caused by local adaptation, we would expect that outcrossed progeny between two small populations should retain the phenotype of the parents. Instead, here we found strong hybrid vigor in all traits, which
strongly supports the idea that individuals from these small populations did not show locally adapted life-history traits, but instead were carrying a higher mutational load accumulated by genetic drift compared to individuals from large populations. It is notable that inbreeding depression, genetic load, and hybrid vigor, all varied in a continuous fashion with genetic diversity and were not just due to differences between small and large populations. This also suggests that there is no qualitative difference between different sources of genetic drift (reduced census population size, bottlenecks during recent colonization events) with respect to their effects on inbreeding depression, genetic load, and hybrid vigor.

Strong hybrid vigor has been documented previously for populations of *Daphnia magna* (Ebert et al. 2002; Haag et al. 2002), but mainly in crosses between small populations. Here, we show that the magnitude of hybrid vigor is strongly influenced by population size. The strong hybrid vigor observed in small populations strongly supports the view that such populations accumulate a genetic load of mildly deleterious and partly or fully recessive mutations (Charlesworth and Charlesworth 1999; Eyre-Walker and Keightley 2007) in a largely independent manner. Furthermore, a substantial part of these alleles occur in homozygous form (either because they are fixed or segregating at high frequencies), so that outcrossing between small populations results in the masking of a substantial number of deleterious populations results in inbreeding depression, whereas the difference between the outbred and parental lines gives the hybrid vigor (Paland and Schmid 2003).

Hence, hybrids between small populations are not mutation-free, but rather carry deleterious mutations in heterozygous form. The same is true for parents from large populations, as is evident from the strong inbreeding depression observed in these populations. Mutations contributing to inbreeding depression are predicted to be, on average, more strongly deleterious than those contributing to hybrid vigor (Charlesworth and Charlesworth 1987; Whitlock et al. 2000). Hence, the fact that outcrossing between small populations resulted in trait values similar to those observed in parents from large populations suggests either a high degree of recessivity of deleterious mutations (both those heterozygous in hybrids as well as those in parents) or a substantial genetic load even in large populations. Because mildly deleterious mutations (the ones predicted to contribute to hybrid vigor) are unlikely to be strongly recessive (Charlesworth and Charlesworth 1999; Eyre-Walker and Keightley 2007), the latter interpretation appears more likely, and hence all our estimates of genetic load are likely underestimates with respect to absolute load (Agrawal and Whitlock 2012). What is evident, however, is that the load shifts in a continuous fashion from a...
predominance of fixed load in the smallest populations to a predominance of segregating load the largest populations. The latter is likely due mainly to alleles segregating at low frequency (Kimura and Crow 1964), which would explain why we did not see a strong masking effect and hence no strong hybrid vigor in the crosses between large populations. Alternatively, the effects of outbreeding depression may have outweighed any remaining benefits of outcrossing.

The observation of outbreeding depression in F1 outcroses between large populations (with F1 and both parents tested in the same environment) suggests that individuals from the large populations have some intrinsic coadaptation, which is disrupted upon mating with individuals from other populations (e.g., Lynch 1991; Orr 1995; Edmands and Deimler 2004). The breakdown of intrinsic coadaptation in the F1’s may either be explained by underdominance or (more likely) epistatic interactions.
Inbreeding depression | Hybrid vigor | Genetic load
---|---|---
Age at death | 3.08 | 6 | 0.022 | -3.61 | 6 | 0.011 | -4.17 | 6 | 0.006
Age at maturity | 3.17 | 6 | 0.019 | -2.81 | 6 | 0.031 | -2.38 | 6 | 0.055
Reproductive output | 10.16 | 6 | <0.001 | -7.30 | 6 | 0.001 | -9.36 | 6 | <0.001
Intrinsic growth rate | 3.31 | 6 | 0.016 | -3.66 | 6 | 0.011 | -5.90 | 6 | 0.001

Shown are t-ratios, degrees of freedom (df) and associated P-values for each trait. Note that the df refer to those used in the t-test, that is the residual df of the model.

Intrinsic coadaptation is a form of local adaptation, where specific alleles perform better with other specific alleles found in the same population either at the same locus (underdominance) or at different loci (epistasis).

That populations of *D. magna* exhibit strong inbreeding depression has also been shown previously (e.g., De Meester 1993), but these tests included only large populations. We found that inbreeding depression decreased with population size, in a neat spectrum from large to small population size. This suggests that small populations had a strongly decreased load of segregating deleterious alleles, either due to fixation of such alleles (recall that homozygous deleterious alleles do not contribute to inbreeding depression) or due to purging. The results on hybrid vigor are clear evidence for the contribution of fixed load to decreased inbreeding depression in small populations. However, whether or not purging also contributed remains less clear. Purging is most efficient for strongly deleterious (e.g., lethal or sterility inducing) alleles (Glémin 2003). Such alleles cannot become fixed in a population and thus do not contribute to fixed load and hybrid vigor. Lethal and sterility alleles are known to be present in some populations of *D. magna* (Routtu et al. 2014), and hence a contribution of purging appears likely. However, the finding of increased genetic load in small populations clearly indicates that, while purging by drift may contribute to lower inbreeding depression in small populations, it does not increase mean fitness (Glémin 2003). In other words, the possible reduction in the frequency of recessive and strongly deleterious alleles in small populations due to purging does not counterbalance the negative fitness effects of drift load (i.e., the fraction of the segregating mutation load that is fixed by genetic drift in finite populations; Lynch et al. 1995). Indeed, genetic load increased with decreasing population size in a nearly linear manner, suggesting that even the
Figure 5. Hybrid vigor ($H$) in relation to the pairwise genetic differentiation ($F_{ST}$) between each focal population and its outcross partner for age at death, age at maturity, reproductive output, and intrinsic growth rate. Error bars represent 1 standard error of the clonal means.

A slight increase in fitness at intermediate population sizes, which has been predicted by models of purging (Glémin et al. 2003; Roze and Rousset 2004; Theodorou and Couvet 2006; Haag and Roze 2007) was not observed in our study, possibly because any such effect was simply overwhelmed by drift load.

Throughout our study, we have assumed that parents are valid substitutes for offspring of random within-population crosses. We have motivated this with a large body of literature showing no systematic inbreeding in *Daphnia* (see methods). Furthermore, the results of our life-table experiments are inconsistent with prior systematic inbreeding in small populations: One round of experimental within-clone mating should still increase the inbreeding coefficient of offspring by 50%, which should not lead to offspring with roughly equal trait values to those of the parents (as seen in our results) unless there was very little segregating variation. The absence of segregating variation due to prior inbreeding rather than drift would require fully inbred parents (i.e., several rounds of prior within-clone mating). Yet clones from all populations were heterozygous, at least at some microsatellite loci. In addition, the existence of faithfully inbreeding lines within populations of *Daphnia* is inconsistent with all previous knowledge on the breeding system of the species (Winsor and Innes 2002).

While the strong genetic load in small populations suggests that individuals from these populations perform poorly when compared to outcrossed individuals or immigrants from larger populations, it remains unknown whether this also affects population persistence (i.e., whether selection is hard or soft in *Daphnia* populations). So far, there is no evidence that extinctions in the metapopulations are driven by genetic factors, and even very small populations persist over many years and appear to be doing fine. Large fitness differences are, however, observed upon immigration and hybridization (Ebert et al. 2002). These results have important implications for habitat loss and fragmentation. Within populations, new deleterious mutations may have one of two fates: they may become fixed via genetic drift and then contribute to the genetic load, or they may be removed by selection (part of this may be via purging). The overall outcome in small populations will be decreased fitness over time due to reduced genetic diversity and the accumulation of deleterious mutations. These populations will not only have decreased fitness, but as they have lowered standing genetic diversity they are less capable of responding to natural selection, thus increasing their odds of extinction (Whitlock 2002). Taken to the extremes we can speak about possible mutational meltdown in small populations (Lynch et al. 1995), which may be a potential problem for very remote *Daphnia* populations with nearly no migration events, as mutational meltdown is thought to occur when there is an abundance of deleterious recessives of small effect (Glémin et al. 2003).

In conclusion, our results provide strong empirical support for the theoretical prediction that the nature of genetic load differs strongly between large and small populations: In large populations, genetic load is mostly due to deleterious alleles segregating at low frequencies (and hence most deleterious alleles occur in heterozygous form). As a result, inbreeding depression is strong, but mean population fitness and the fitness of outcrossed individuals are not strongly affected. In contrast, in small populations the genetic load occurs mostly in its fixed form. Hence inbreeding
depression is low, but mean population fitness is strongly decreased and outcrossing results in large hybrid vigor. These differences in the nature of the genetic load imply that also many evolutionary processes in Daphnia vary strongly depending on population size. The strong hybrid vigor in small populations implies that immigrants into these populations have a selective advantage (Ebert et al. 2002) and hence experience positive selection for dispersal (Ronce 2007). In addition, the genetic differentiation among populations is likely driven by drift during colonization and founder events; but also by ongoing drift whereby clonal selection may further reduce the effective population size. Local adaptation, on the other hand is likely only a weak driver of population differentiation in these populations. This is due to the inefficacy of selection (as evident through the high fixation load), whereby local adaptation is greatly reduced and, if present, mostly due to alleles of large effects. In contrast, in large populations, local adaptation may be a much stronger factor, as also suggested by our finding of local coadaptation (outbreeding depression). Furthermore, the evidence for only weak effects of genetic drift in large populations suggests improved efficacy of selection, which should favor local adaptation also with respect to different environmental conditions. Together the different forms of local adaptation may select against dispersal and favor the long-term maintenance of initial allele-frequency differences in large populations, as envisaged by the monopolization hypothesis (De Meester et al. 2002). Finally, selection for the promotion of outbreeding via breeding system may also depend on population size. Such a system is known in Daphnia, where nonmale producing (NMP) clones are obligate outcrossers. Our results suggest that the cost of within-clone mating is strongest in the largest populations, but that this cost has to be modulated by the frequency in which it occurs under natural circumstances. In very large populations, the likelihood of within-clone mating even at the end of the season after clonal selection may be very low, so that the cost may almost never be incurred. However, in populations of intermediate size, inbreeding depression is still strong enough, and the likelihood of within-clone mating may be substantial enough for obligate outcrossing to be favored.

ACKNOWLEDGMENTS
We thank D. Slodowics, S. Diserens, C. Reisser, A. Marcelino, and R. Schleberger for help with running the life-table experiment, and B. Walser and Y. Galimov for providing Daphnia clones. We also thank S. Edmands and three anonymous reviewers for their constructive comments on the manuscript. This work was supported by the Novartis Foundation for Medical and Biological Research (Grant 11A12), the Swiss National Science Foundation, and by the University of Fribourg.

LITERATURE CITED


Associate Editor: S. Edmans
Handling Editor: J. Conner