EFFECTS OF POLLEN QUANTITY ON PROGENY VIGOR: EVIDENCE FROM THE
DESERt MUSTARD LESQUERELLA FENDLERI

RANDALL J. MITCHELL

Department of Biology, University of New Mexico, Albuquerque, New Mexico 87131

Key words.—Brassicaceae, gametophytic selection, pollen competition, pollination intensity, review.

Received September 3, 1996. Accepted May 16, 1997.

Intense competition among pollen grains for access to ovules often increases progeny vigor (Mulcahy 1971, 1979; Mulcahy and Mulcahy 1975; Mulcahy et al. 1975; McKenna 1986; Winsor et al. 1987; Bertin 1990; Schlichting et al. 1990; Richardson and Stephenson 1992; Janse and Verhaegh 1993; Quesada et al. 1993, 1996a,b; Palmer and Zimmerman 1994; but see Smith et al. 1990; Snow 1990, 1991). The most common explanation for this is that selection on gametophytes is an efficient means for removing deleterious alleles at loci important for both gametophytic and sporophytic life stages (Charlesworth and Charlesworth 1992; Hormaza and Herrero 1992; Stephenson et al. 1992). Gametophytic selection therefore has been proposed as an important factor in the success of angiosperms (Mulcahy 1979).

Studies of pollen competition and progeny vigor usually involve variation in the number of pollen grains applied to stigmas or of the distance pollen tubes must travel to reach ovules (utilizing natural or artificial variation style length). Studies of the latter type invariably show significant effects on progeny vigor (Mulcahy and Mulcahy 1975; Ter-Avanesian 1978; McKenna and Mulcahy 1983; Lee and Hartgerink 1986; McKenna 1986). However, effects of pollen load size are less consistent (e.g., Winsor et al. 1987; Bertin 1990; Smith et al. 1990; Snow 1990, 1991; Palmer and Zimmerman 1994). Furthermore, most studies to date involve cultivated plant species (but see Snow 1990, 1991; Richardson and Stephenson 1992), which may respond differently than wild plants because of their history of close inbreeding and hybridization.

To better understand the effects of pollen competition on progeny vigor, and why results vary among different studies, more empirical data are needed. Here I report results from a study of the effects of pollen load size on progeny vigor for a native, desert mustard (Lesquerella fendleri; Brassicaceae). I then review published studies of the effects of pollen load size on progeny vigor, considering why results may vary among methods and species.

METHODS

Lesquerella fendleri is a small perennial common in the deserts of the southwestern United States. The 15-mm diameter, yellow flowers are simultaneously hermaphroditic and are self-incompatible. Flowers in the field receive an average of 319 pollen grains (SE = 43, n = 17) during their one- to two-day existence, and produce 8.5 seeds per mature fruit (SE = 0.9, n = 60). Flowers typically contain about 20 ovules, and are pollinated by small bees and beetles (unpubl. data).

To assess the effects of pollen load size on progeny vigor, I artificially altered the amount of pollen applied to flowers on 13 potted plants in a greenhouse. All plants were collected from within 100 m of one another at the Sevilleta Long Term Ecological Research area (LTER), 80 km south of Albuquerque, New Mexico, in fall 1992. I applied three pollination treatments to flowers in randomized order within plants, using monofilament fishing line (Kearns and Inouye 1993). To do this, I placed anthers from at least eight plants not otherwise involved in the experiment in a small petri dish, and coated the fishing line with pollen by swirling it in the dish. Then, I either touched the line briefly to the stigma (low pollen treatment), swiped it lengthwise along the stigma once (medium treatment), or swiped it lengthwise along the stigma four or more times (high treatment), reloading the line with pollen after each touch or swipe. I intended these three treatments to provide a wide range of pollen availability, but still allow full or nearly full seed set (Mitchell 1997). I applied six replicates of these three treatments to each plant from June 8–13, 1993 (18 flowers on each of 13 plants, 234 flowers total). More than 72 hours after pollination I collected stigmas, squashed them on a microscope slide with basic fuchsin stain (Kearns and Inouye 1993), and counted pollen grains (Mitchell 1997). I analyzed results from all recovered fruits, including those with no seeds. Along with seed number and individual seed mass (total seed mass of a fruit/number of seeds in fruit), I also analyzed the proportion of pollen that successfully sired seeds (ratio of the number of seeds in a fruit to the number of pollen grains applied; seed/pollen ratio) as an index of the intensity of pollen selection. Unfortunately, I was unable to distinguish self and outcross pollen when counting pollen loads, and attempts to prevent self-pollen deposition were unsuccessful (see Mitchell 1997). Therefore, seed/pollen ratios include both the outcross pollen I applied and passively deposited self pollen (average self pollen loads on unpollinated flowers = 92 grains, n = 74). Trials indicated that prior arrival of self pollen does not decrease seed production (unpubl. data).

In September 1993 I planted seeds into sterilized sand in 6-cm pots in a greenhouse at the University of New Mexico, and sprayed them with 1g/L gibberellic acid to break seed dormancy (Evans et al. 1996). Two maternal plants did not
produce enough seeds for planting, so I excluded them. I planted seeds from each treatment-plant combination in a randomized block design, with 42 blocks that each contained up to 33 seeds (three treatments for each of 11 plants). I recorded emergence of cotyledons (emergence date) to the nearest half-day, and measured seedling size (length and width of longest leaf, number of leaves) as each seedling reached 21 days postemergence.

To assess effects of pollen load size on progeny performance in the field, I transplanted some of the seedlings described above back to their site of origin. To do so, on November 21, 1993, I constructed 10 new blocks of seedlings. These blocks included as complete a sample of plant-treatment combinations as possible, achieved by randomly choosing plants from those alive in the original planting blocks. After cold-hardening these blocks for one week, I transplanted them to within 100 m of the origin of the parental plants at the Sevilleta LTER.

I used the remaining plants to construct new blocks for greenhouse measurements, in the same manner described for the field seedlings. Because unequal numbers of seedlings remained for each plant-treatment combination (I gave first priority to producing even replication in the field blocks), these blocks were strongly unbalanced. I recorded date of first flowering and date of death for each plant.

For analyses, I used mixed-model ANOVA (Proc GLM; SAS Institute 1989) and Type III sums of squares, considering maternal plant and replicate as random factors, and treatment as fixed. Therefore I tested treatment over treatment × plant (Sokal and Rohlf 1985).

**RESULTS**

As anticipated, pollen loads differed significantly among treatments, but seed set per fruit and proportion fruit set did not (Table 1). As a consequence, treatments differed significantly in seed:pollen ratios (Table 1). Overall, the seed:pollen ratio differed two-fold among treatments, with the fastest 6% of pollen grains siring seeds in the low treatment, and the fastest 3% in the high treatments, respectively. Thus, all treatments resulted in intense pollen selection (low seed:pollen ratios). Seed size did not vary among treatments. For brevity, Table 1 shows statistical tests only for treatment effects. For most traits, the full ANOVA also revealed significant maternal effects, and significant maternal plant × replicate interactions, which indicate that the magnitude of the differences among plants varied among replicates. In no cases were interactions of treatment with any other factor significant. Sequential Bonferroni adjustments to probability levels (Rice 1989) to account for nonindependence of traits did not change these conclusions.

I found no detectable effect of pollination treatment on any measure of early offspring quality. Pollen load size did not detectably affect date of emergence or proportion of seeds emerging (Table 2). These conclusions persisted when the data were analyzed using the accelerated failure time analysis advocated by Fox (1993; $\chi^2_1 = 1.1, P = 0.6$). Pollen load size also had no significant effect on any of the postgermination characters measured in the first 21 days (Table 2). There were no significant differences among maternal sibships for any traits except emergence date, proportion emerging, and number of leaves. Significant maternal plant × treatment interactions (for leaf length, leaf width, and leaf number) made it difficult to draw unambiguous conclusions about maternal plant and treatment effects in any case. Earlier emerging seedlings tended to grow larger by 21 days after emergence (the correlation between emergence date and the 21-day measures ranges from $-0.16$ to $-0.28$, all $P < 0.002$), but including emergence date as a covariate still did not reveal significant effects of treatment (unpubl. data). Furthermore, the lack of a detectable effect was not due to low statistical power. Power analysis (Cohen 1969) indicated that the observed magnitude of variation among treatments provided > 70% probability of detecting significance, even for the small treatment effects that occurred in this experiment. The sole exception to this was leaf number, for which power was 15%.

For plants transplanted to the field, leaf number just before transplantation (54 days after planting) did not vary significantly among treatments (Table 2) although there was significant variation among maternal sibships ($P = 0.004$; the maternal plant × treatment interaction approached significance, $P = 0.06$). Unfortunately, all 283 plants unexpectedly died within two weeks of transplanting. The plants grown to maturity in the greenhouse showed no significant differences among treatments in date of flowering or date of death (Table 2).

**DISCUSSION**

Despite strong and significant variation among treatments in pollen load size and significant (though small) differences in seed:pollen ratios, I detected no significant effect of pollination treatment on any measure of offspring quality. Indeed, for only two of 10 measures did seeds from high pollen loads tend (nonsignificantly) to perform best. Perhaps any

---

### Table 1. Results from the pollination stage of the experiment. Means of plant means ± SE shown; $n = 12$ plants per entry. $F$, df, and $P$-values are from tests for a significant treatment effect in a mixed-model ANOVA in which maternal plant, treatment, replicate, and interactions were estimated; treatment effects are therefore tested over plant × treatment. Total $N = 216$ flowers or fruits.

<table>
<thead>
<tr>
<th>Pollen load</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
<th>$F$</th>
<th>df</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion fruit set</td>
<td>0.70 ± 0.08</td>
<td>0.69 ± 0.08</td>
<td>0.62 ± 0.07</td>
<td>0.9</td>
<td>2.22</td>
<td>0.4</td>
</tr>
<tr>
<td>Seed: pollen ratio</td>
<td>0.063 ± 0.008</td>
<td>0.040 ± 0.006</td>
<td>0.031 ± 0.005</td>
<td>12.1</td>
<td>2.20</td>
<td>0.004</td>
</tr>
<tr>
<td>Seeds</td>
<td>6.9 ± 1.0</td>
<td>6.9 ± 1.1</td>
<td>7.6 ± 1.7</td>
<td>1.1</td>
<td>2.20</td>
<td>0.3</td>
</tr>
<tr>
<td>Mass/seed (mg)</td>
<td>0.39 ± 0.03</td>
<td>0.40 ± 0.05</td>
<td>0.40 ± 0.04</td>
<td>0.9</td>
<td>2.20</td>
<td>0.4</td>
</tr>
</tbody>
</table>
effects of pollen load size on sporophytic quality would have been more noticeable under field conditions (Lee and Hartgerink 1986; Snow 1990) and after maternal effects dissipated, but the unexpected mortality of all transplants prevented that comparison here. However, other studies have shown effects at the seed and three- to four-week-old seedling stage (e.g., Bertin 1990; Björkman 1995), so the abbreviated time scale here is comparable with that in other studies that found positive results.

Why do my results and those of a few others (Smith et al. 1990; Snow 1990) differ from those for so many other species? Possible explanations include: inadequate statistical power; utilization of “small” pollen loads that were too large (see Schlichting et al. 1990); and study of a wild, self-incompatible plant species that might respond differently than self-compatible or cultivated species (see Snow 1990). The first possibility is readily dismissed as a general explanation, since my study had adequate power to detect reasonably strong effects had they existed (see Results).

To investigate the possibility that studies showing no effect involved “small” pollen loads that nonetheless were large enough to generate such intense pollen competition that no further gains in vigor were possible, I attempted to review all published studies of the effects of pollen load size on progeny vigor. I excluded from consideration any studies using other methods to vary pollen competition, such as cutting the style to different lengths (e.g., Mulcahy and Mulcahy 1975). All such studies showed a significant effect of pollen competition of progeny vigor. I also excluded any that did not report pollen load sizes and seed set per flower (e.g., Smith et al. 1990). For each study, I noted whether pollen load had a significant overall effect on any progeny character, and calculated seed:pollen ratio (as defined in the Methods) as an index of the intensity of pollen selection. When this ratio is small, few pollen grains successfully fertilize ovules, and competition is therefore more intense. I calculated this ratio separately for the lowest and highest pollen load treatments in each study. I used these values to compare studies finding no effect of pollen load size on any measure of progeny vigor with those finding a significant effect on any estimate of progeny vigor. Ideally, information such as pollen germination and attrition could be used to more accurately assess the fraction of potentially successful pollen that actually succeeded in siring seeds, but such information was not available for most studies.

I located 12 studies that met my criteria, five of which involved Cucurbita pepo (Table 3). To avoid oversampling that species, I only report results from one study on C. pepo, and from one on crosses between C. pepo and C. texana. My conclusions were unaltered by this conservative step. Because pollen and seed values for some studies varied or were ambiguous, Table 3 reports both high and low estimates of seed:pollen ratios. For simplicity, I report results based on the lowest values for each study. The trends reported below were unaffected by which estimates were used, although statistical significance of the differences disappeared when the largest estimates were used. Seed:pollen ratios were lower overall in the three studies that found no significant effect of pollen load size on progeny vigor. The mean value for low pollen loads was 0.082 for studies showing no effect, and 0.266 for those showing an effect (Table 3). The difference between means for these groups was biologically large, and marginally significant (one-tailed t-test on arcsine, square-root transformed values, \( P = 0.04 \)). Because of the small sample size, only very strong differences were likely to be significant, and I took this to indicate that pollen selection tended to be very intense in the studies, showing no effect, even in the low pollen treatment. The mean seed:pollen ratio for high pollen loads was 0.021 for studies showing no effect, and 0.076 for those showing an effect (Table 3). This difference approaches significance (one-tailed t-test as described above, \( P = 0.10 \)). Together, these results are consistent with
the proposition (see also Schlichting et al. 1990) that the low treatments in my study and in Snow's (1990, 1991) were not low enough to produce a meaningful difference between treatments in pollen competition. In other words, although the low treatment received less pollen, it may nonetheless have caused such intense selection (in my study only 6% of pollen grains successfully sired seeds) that no additional gains in vigor were possible when the intensity of competition was doubled so that only 3% of *L. fendleri* pollen tubes were successful.

Figure 1 represents this argument (see also Snow 1986; Schlichting et al. 1990; Richardson and Stephenson 1991; Mitchell 1997). In this hypothetical example, as pollen load increases, pollen competition and progeny vigor both increase nonlinearly, but at different rates. In the region labeled “filling,” pollen loads are small enough that only a fraction of ovules are fertilized. This means there is little or no pollen competition, and that more pollen increases seed set, but not seedling vigor. In the region labeled “sorting,” all ovules are fertilized (or more ovules are fertilized than can be matured), and seed set per fruit does not increase with pollen load. However, seed:pollen ratio decreases, and an even-smaller fraction of the available pollen tubes is successful. This improves mean offspring quality. In the region labeled “surplus,” more pollen does not increase progeny vigor because all available genetic variation for offspring quality has been exploited.

In terms of Figure 1, studies finding no effect may have used pollen loads that were so large that both high and low pollination treatments were in or close to the “surplus” region. Also, there may have been no effect of pollen load on vigor if both treatments were in the “filling” region, but this would be signified by large differences in seed set per fruit among treatments. To detect a significant effect of pollen competition or progeny vigor, one pollination treatment must be in a different region from the others. Studies showing a significant effect probably had the low treatment in the “filling” or “sorting” regions, and the high treatment in the “surplus” region. However, because seeds in few-seeded
fruits frequently are larger than those in many-seeded fruits, comparison of vigor for seedlings from the “filling” region with any other region may be suspect (Charlesworth 1988; Stephenson et al. 1988), requiring more complicated experimental designs to remove effects of seed size on vigor (e.g., Schlichting et al. 1990; Quesada et al. 1993, 1996a,b).

Is it possible to determine the boundaries between these different regions a priori? I know of no reliable way to predict the upper level of the “sorting” region. However, the upper limit for the “filling” region can be determined from dose-response relationships, since it should occur at the point where seed production no longer increases with pollen load. The dose-response relationship for _L. fendleri_ (Mitchell 1997) indicates that the upper end of the “filling” region should be ~120 pollen grains. This means that my low treatment was larger than the upper limit of the “filling” region by 20 or so pollen grains, and suggests that all of my three treatments were probably in the “sorting” and “surplus” zones. In retrospect, using a pollen load smaller than the breakpoint of the dose-response curve (~50) might have better allowed me to detect any effect of pollen competition on progeny vigor. Note that dose-response curves and this breakpoint may vary among individual plants (Mitchell 1997), further complicating the issue.

Are there other differences between those studies finding a significant effect of pollen load size on progeny vigor and those that do not? Self-compatibility is not associated with a particular outcome (Table 3), but all three studies showing no effect involved wild species. This is not consistent with the hypothesis that cultivated, self-compatible plants (which are required before these can be tested definitively). However, many other hypotheses are consistent with those that do not. Self-compatibility is not associated with any other region may be suspect (Charlesworth 1992; Charlesworth, D., and B. Charlesworth. 1992. The effects of selection in the gametophyte stage on mutational load. Evolution 46:703-720.


ACKNOWLEDGMENTS

J. Avritt, B. Cabin, A. Evans, D. Furcht, P. Giegick, D. Marshall, E. Mitchell, K. Mitchell, and C. O’Dear provided help, support, and advice. Special thanks to N. Waser for suggesting inclusion of Figure 1, and to A. Stephenson, A. Snow, and an anonymous reviewer for many helpful comments on an earlier version of the manuscript. The Sevilleta LTER and the Sevilleta National Wildlife Refuge provided permission to collect and transplant plants. R. Bertin, A. Snow, and A. Stephenson provided unpublished information on pollen and seed production in their studies. This research was supported by National Science Foundation grant DEB 92-03203. This is publication 95 of the Sevilleta LTER.
ADVANTAGES OF MULTIPLE MATINGS TO FEMALES: A TEST OF THE INFERTILITY HYPOTHESIS USING LIZARDS

MATS OLSSON1,2 AND RICHARD SHINE2

1University of Göteborg, Department of Zoology, Medicinaregatan 18, 413 90 Göteborg, Sweden
2University of Sydney, School of Biological Sciences A08, New South Wales 2006, Australia

Key words.—Female fitness, infertility hypothesis, Lacerta agilis, multiple matings, sand lizard.

Recent studies have reported high frequency of multiple mating by females in many animal species (see e.g., Smith [1984] and references therein; Birkhead and Møller 1992). However, the mechanism by which multiple mating enhances female fitness remains obscure in most cases. One potential advantage of female multiple matings would be to reduce the risk of mating singly with a sterile male. In the sand lizard, Lacerta agilis, unfertilized eggs are easily identified, and we exploit this phenomenon to estimate the frequency of male sterility in a natural population, and look for increased fertility with increased number of female matings and partners.

Broadly, the potential advantages of multiple matings to females can be assigned to one of four categories: (1) transfer of nutrients in the seminal fluids (Leopold 1976); (2) genetic benefits to offspring quality (e.g., via greater genetic diversity in the offspring, enhanced sperm competition, or increased opportunity for female choice among sperm from different males; Bellis and Baker 1990; Madsen et al. 1992; see also Capula and Luiselli 1994; Olsson et al. 1994a,c); (3) Fisherian (“run-away”) selection for sperm performance (Keller and Reeves 1995); and (4) as “insurance” against male infertility (i.e., ensuring that sufficient viable sperm are available to fertilize the eggs; e.g., Petrie et al. 1992). Although the fourth of these hypotheses is the simplest, it has attracted relatively little attention, perhaps primarily because data on incidence of male sterility in natural populations are exceptionally rare. We exploit such a dataset on sand lizards, and set out to examine how common male sterility is in our study population and to what degree male sterility may select for multiple matings in females.

Clearly, the first step in quantifying the incidence of male sterility, or testing for its effects, is to be able to distinguish between fertilized and unfertilized eggs. Although this seems obvious, in practice it is a difficult task in many types of animals. Incompatibility between male and female haplotypes may become evident very early in embryogenesis, so that the challenge is to distinguish between an egg that has never been fertilized compared with one that has been fertilized but has undergone very little development prior to embryo death. For example, most mammals are viviparous...