Statistical Genetics of an Annual Plant, *Impatiens capensis.*
I. Genetic Basis of Quantitative Variation

Thomas Mitchell-Olds* and Joy Bergelson†

*Division of Biological Sciences, University of Montana, Missoula, Montana 59812, and †Department of Zoology, University of Washington, Seattle, Washington 98195

Manuscript received May 2, 1989
Accepted for publication October 10, 1989

ABSTRACT

Analysis of quantitative genetics in natural populations has been hindered by computational and methodological problems in statistical analysis. We developed and validated a jackknife procedure to test for existence of broad sense heritabilities and dominance or maternal effects influencing quantitative characters in *Impatiens capensis.* Early life cycle characters showed evidence of dominance and/or maternal effects, while later characters exhibited predominantly environmental variation. Monte Carlo simulations demonstrate that these jackknife tests of variance components are extremely robust to heterogeneous error variances. Statistical methods from human genetics provide evidence for either a major locus influencing germination date, or genes that affect phenotypic variability per se. We urge explicit consideration of statistical behavior of estimation and testing procedures for proper biological interpretation of statistical results.

Understanding the potential for life history evolution in natural populations requires information on genetic and environmental determinants of life history variation, as well as knowledge of natural selection operating in the field (LANDE 1982). Applications of quantitative genetic methods have provided estimates of genetic variances and covariances in natural plant populations (e.g., VENABLE 1984; MITCHELL-OLDS 1986), which quantify constraints on rate and direction of evolutionary change. Despite recent extensions of quantitative genetic theory, analysis of empirical results often poses severe statistical problems (MITCHELL-OLDS and RUTLEDGE 1986; MITCHELL-OLDS and SHAW 1987). Choice of optimal statistical procedures may be difficult (e.g., SEARLE 1971; SWALLOW and MONAHAN 1984), and analysis of some data sets may be computationally infeasible (SHAW 1987). Nevertheless, understanding of underlying biological patterns and principles may be strongly influenced by choice of genetic models and analytical procedures. This paper illustrates the importance of considering statistical behavior of testing procedures in order to provide proper biological interpretation of genetic variation and natural selection operating in the wild.

Basic models of polygenic, additive genetic variance (BULMER 1980; FALCONER 1981) have been extended to consider nonadditive genetic variance, maternal effects, and major gene effects (OCKERHAM 1963; WILLHAM 1963; MORTON 1982). Nonadditive genetic variance is typically partitioned into dominance and epistatic effects. Dominance genetic variance is frequently found to be important in plant populations (THOMAS and GALE 1977; MATHER and JINKS 1982; SPRAGUE 1983; LANDE and SCHEMSKE 1985). Since directional dominance is a necessary condition for occurrence of inbreeding depression under many genetic models (CROW and KIMURA 1970), the widespread observation of inbreeding depression provides further evidence for the importance of dominance genetic variance. Epistatic variance is difficult to estimate with precision due to large statistical sampling error (HALLAUER and MIRANDA 1981), especially in field situations. Maternal effects, due to either cytoplasmic inheritance, maternal nuclear genes, or long term maternal environmental effects (FALCONER 1981) are well known in plants and animals (WILLHAM 1972; ROACH and WULFF 1987). Last, the number of loci influencing continuous characters in natural populations and the magnitude of their effects has been the topic of considerable discussion for many years, since the genetic architecture of quantitative variation affects the response to selection, and differs among competing models of evolutionary change (WRIGHT 1968; MAYO 1980; LANDE 1981, 1983; MATHER and JINKS 1982; GOTTLIEB 1984). Despite the importance of this question, the genetic basis of quantitative variation is poorly understood, since most statistical methods for examining possible major genes influencing continuous characters have statistical or empirical limitations that hinder their application to natural populations (MITCHELL-OLDS and RUTLEDGE 1986). In this paper, we show that most quantitative genetics exper-
iments involving sibling families contain statistical information that may provide weak evidence regarding patterns of genetic control of continuous characters. Previously MITCHELL-OLDS (1986) estimated levels of additive genetic variance in field populations of Impatiens capensis, using the sires component of variance from a nested design (North Carolina type I). Although additional information regarding determinants of phenotypic variation can be extracted from such experiments, more complete analysis of such data has been hindered by lack of statistically valid and numerically feasible methods (MITCHELL-OLDS and RUTLEDGE 1986; SHAW 1987). Here we develop and validate statistical methods for further analysis of such data, permitting significance testing of possible maternal and/or dominance genetic influences on phenotypic variation. We also show how methods from human statistical genetics may provide a crude test of whether continuous phenotypic variation is caused by many genes of small effect, or whether variation due to a locus with large differences between alleles (a "major gene" or "leading factor") (WRIGHT 1968) might be present in addition to polygenic variation. (Nothing in this paper should be construed as advocating a purely oligogenic model.)

Analysis of data from MITCHELL-OLDS (1986) will be used to address the following questions: 1) Is there evidence for maternal or dominance genetic variance? 2) Do alternative statistical estimators provide evidence for previously undetected genetic variation? 3) Is life history variation under polygenic control, or is there evidence for loci of large effect? 4) Are there genetic influences on levels of phenotypic variation?

**MATERIALS AND METHODS**

**Impatiens natural history:** Impatiens capensis Meerb., or jewelweed, is an erect summer annual native to northeastern North America, frequently found on streambanks and in floodplains that are partially disturbed. It is an obligate annual in that all viable seed germinate in the spring following their production (WALLER 1984). Their innate dormancy is broken by a period of cool moist stratification (LECK 1979). Given moist conditions, growth continues until late August, when their height may exceed 2 m. Jewelweed exhibits a mixed mating system as consequence of production of both cleistogamous (CL) and chasmogamous (CH) flowers. The CL flowers are obligately self-fertilized at an early stage in their development, while the CH flowers are predominantly outcrossed (MITCHELL-OLDS and WALLER 1985; KNIGHT and WALLER 1987). Plant size is the major component of maternal fitness in this strict annual: dry weight is highly correlated with total seed production ($r^2 = 0.92$) (WALLER 1979), while juvenile mortality is negligible in this study (MITCHELL-OLDS 1986).

**Experimental design:** Experimental design has been described in detail in MITCHELL-OLDS (1986). Briefly, parental plants were collected from near Madison and Milwaukee, Wisconsin, and controlled pollinations were conducted within each population in a nested sires/dams/progeny design (North Carolina design I) (COMSTOCK and ROBINSON 1948). Progeny were planted into a naturally seeded stand of I. capensis, and morphological and phenological characters were measured. Here we report on seed weight, germination date, plant size at monthly intervals in June, July, and August (estimated as height $\times \pi \times$ radius$^2$), and final dry weight, which is highly correlated with fecundity and will be taken as an indicator of fitness. We calculated growth rate between monthly size estimates as (final volume - initial volume)/time interval. Since quantitative genetic interpretation of selection gradients requires multivariate normality of predictor variables, all predictor variables were subject to a Box-Cox transformation (SOKAL and ROHLF 1981). Dry weight was divided by mean dry weight to provide an index of relative fitness (LANDE and ARNOLD 1983).

Quantitative genetic analysis and assumptions are reviewed in MITCHELL-OLDS (1986) and MITCHELL-OLDS and RUTLEDGE (1986). Proper randomization of crosses and planting location precluded most sources of bias. $F_{st}$ was estimated by isozyme analysis to be 0.47 in each population, permitting estimation of additive genetic variance from the covariance of siblings (COCKERHAM 1963; MITCHELL-OLDS 1986).

**Variance component analysis:** Alternative statistical procedures may differ in bias of estimates, ability to test particular hypotheses, statistical power of significance tests, robustness to violation of assumptions, and computational feasibility (e.g., SOKAL and ROHLF 1981; SWALLOW and MONAHAN 1984; SHAW 1987). Our inquiry required a test for existence of dominance and/or maternal effects, as well as the ability to pool sires and dams information regarding broad sense heritabilities. Although least squares ANOVA permits estimation of a variance component attributable to maternal and dominance effects using a linear combination of the observational components of variance (below), standard least squares methods do not provide a significance test for these effects when data are unbalanced. Maximum likelihood (SHAW 1987) proved computationally infeasible, and has unknown statistical performance with heterogeneous variances. Randomization tests of significance could not be applied since the appropriate resampling design for comparing components of variance due to sires and dams (at two different hierarchical levels) is unknown. SOKAL and ROHLF (1981) discuss how deletion of groups (e.g., sire families) is appropriate for testing of effects at lower levels (e.g., dams). However, methods for testing equality of sires and dams components of variance using randomization tests are unclear, since significance tests for a particular component require randomization at that level. Jackknife procedures have demonstrated good statistical behavior for estimation and testing of variance components in quantitative genetic applications (ARVESEN and SCHMITZ 1970; KNAPP, BRIDGES and YANG 1989), and are less computationally burdensome than bootstrap methods (EFRON 1982). For these reasons, we employed a nested ANOVA for unequal sample sizes (SOKAL and ROHLF 1981, pp. 293-300) using unweighted, delete-one-sire-family jackknife estimates (SOKAL and ROHLF 1981, p. 796) and approximate tests of significance (WU 1986). Our initial simulations found that the log transformation of ARVESEN and SCHMITZ (1970) was unnecessary, and that truncation of negative estimates of variance components caused rejection of the null hypothesis at undesirably high frequencies, so we report here jackknife results from an untransformed, unmatched variance component estimator (see also KNAPP, BRIDGES and YANG 1989).

Let the observational components of variance, $\sigma^2_i = \sigma^2_s + \sigma^2_b + \sigma^2_e$, where $\sigma^2_s$ is the sires component of variance, $\sigma^2_b$ is the dams component of variance, and $\sigma^2_e$ is the variance within full-sibships. In
random mating populations the causal components of variance are obtained from $\sigma_i^2 = 1/4V_i$, and $\sigma_O^2 = 1/4V_o + 1/4V_m + V_V = 1/4V_r + e_{OM}$, where $V_V =$ the additive genetic variance, $V_o =$ the dominance genetic variance, and $V_m =$ the maternal variance (which may be due to cytoplasmic, indirect genetic effects, and permanent maternal environmental effects) (Falconer 1981). Because $V_o$ and $V_m$ are confounded in this design, we use $\sigma_{OM}$ to indicate a combination of dominance and maternal effects.

Least squares estimates of the observational and causal components of variance were obtained by equating observed means to their expectations (Searle 1971; Falconer 1981; Sokal and Rohlf 1981). We estimated $\sigma_{OM}$ as $\sigma_o^2 - \sigma_d^2$ ("caret notation" denotes "an estimator of"), and a pooled, reduced variance estimate of the additive genetic variance can be obtained as $4\sigma_{OM}/(1 + \tilde{P}_M)$, where $\sigma_{OM} = 1/2(\sigma_o^2 + \sigma_d^2)$, provided that $\tilde{P}_M = 0$ (Cockerham 1963; Falconer 1981, p. 156). Finally, jackknife estimates and standard errors were obtained by an unweighted, delete-one-sire-family jackknife (Wu 1986) and compared to $t$-distribution. We also examined an alternative weighting scheme for $\sigma_{OM}$, weighting $\sigma_o^2$ and $\sigma_d^2$ in inverse proportion to their estimated variances (Searle 1971, p. 476), but this proved infeasible because negative estimates of $\text{var}(\sigma_o)$ and $\text{var}(\sigma_d)$ were obtained due to sampling error.

Since jackknife tests of significance are only approximate, we conducted Monte Carlo simulations to determine the statistical behavior of this test. Ideally, a statistical test will reject a true null hypothesis only 100% percent of the time. However, a statistical test may be too conservative (reject true $H_0$ fewer times than expected) or too liberal (reject true $H_0$ too many times) (Sokal and Rohlf 1981, p. 301). This can be tested by generating large numbers of pseudo-random data sets where $H_0$ is true, and determining whether the putative level of significance is actually provided by the test. In addition, this procedure can be used to examine the power of a test (how often $H_0$ is rejected when it is false) by generating simulated data where variance components have true values greater than zero. Biologically, such a power curve (see Figure 1) enables us to ask how frequently we attain significance for a particular heritability in an experiment. Finally, Monte Carlo simulations enable us to assess whether jackknife tests of significance are robust to possible heterogeneity of within-family variance.

For each assumed parametric value of $V_o$ and $\sigma_{OM}$ we generated 3000 pseudo-random data sets (detailed in Figure 1), then calculated jackknife estimates of variance components and approximate $t$-tests. We compared $t$-values with the appropriate critical value at $\alpha = 0.05$, and we computed the frequency of rejection of $H_0$. To examine the effect of heterogeneous within-family variances we randomly assigned 50% of simulated full-sibships to have within-family variance increased by a factor of two.

**Major gene analysis:** Little information exists on the genetic basis of quantitative variation (Falconer 1981; Barton and Turelli 1989). Although it is often assumed that there is a continuous distribution of allelic effects, with many genes of small effect and few genes of large effect, it is difficult to fit phenotypic data to these complex models. Instead, most segregation analyses of phenotypic data have considered simplified models. Purely polygenic models and oligogenic models represent opposite ends of a presumably continuous spectrum of gene effects. An intermediate (although equally arbitrary) approach considers a "mixed model" with a major locus and a segregating polygenic background (e.g., Morton 1982). Existing statistical methods for distinguishing between polygenic inheritance, oligogenic segregation, and mixed models are complex and have limited statistical power (reviewed in Mitchell-Olds and Ruddle 1986). One crude but simple method for detection of a major locus was suggested by Pearson (1994) and Fain (1978). They proposed that segregation of a major gene might be detected by graphical plots of family variance against family mean, since families homozygous at the major locus would have low variance and extreme mean values, while families that were segregating at the major locus would have larger within family variance and an intermediate mean phenotype. Consequently, as a test for the presence of major genes, Fain suggested a test for heterogeneity of within-family variances. The statistical power of this test is known to be limited (Fain 1978; Mayo, Hancock and Baghurst 1980). But it has the advantage of being usable in many other analyses from quantitative genetic studies. Furthermore, since most statistical tests assume homogeneity of error variances, this test should always be performed in routine statistical explorations. Fain's approach has been criticized because loss of developmental homeostasis might also cause heterogeneity of family variances (Mayo, Hancock and Baghurst 1980). However, we show below that graphical analysis easily discriminates between these two hypotheses.

**RESULTS**

**Variance component analysis:** Components of variation for each character are presented in Table 1. The data differ in several respects from the analysis of Mitchell-Olds (1986). Slightly more plants are included in this broad sense analysis, since previous statistics were restricted to sires that were mated to more than one dam. The current approach also includes sires mated to only a single dam, which constitute full-sibships that confound additive genetic variance with dominance and maternal effects. Also, we analyzed a different set of predictor variables, each subject to Box-Cox transformation, while the previous analysis examined predictor variables on their original scale of measurement. Finally, these results utilize a different statistical estimation and testing procedure [jackknife estimates of ANOVA method variance components (Searle 1971) and jackknife tests of significance (Wu 1986), vs. ANOVA method variance components subject to randomization tests of significance]. Estimates of partitioning of variance (Table 1) are in general accord with Mitchell-Olds (1986). Seed weight and germination date show significant dominance and/or maternal effects in the Madison population, while analysis of the Milwaukee families is only suggestive of dominance or maternal effects. When families from the two populations are analyzed together, results generally follow this same pattern. There is one result in which new analyses are at odds with previous ones: estimates of heritability of fitness in the Milwaukee population are noticeably smaller and are no longer significantly different from zero.

**Statistical simulations of variance components:** The jackknife estimators of variance components were unbiased (data not shown), and provided useful
variances: simulations with heteroscedastic within-family tests on rejecting a true null hypothesis.

The jackknife test of significance was extremely robust (410 T. Mitchell-Olds and J. Bergelson).

Variance component analysis of Madison and Milwaukee populations, and pooled data from both populations. SW = seed weight, GD = germination date, JS = June size, EG = early growth rate, LG = late growth rate, and FT = relative fitness.  \( \hat{\sigma}_G, \hat{\sigma}_D, \hat{\sigma}_{BM} \), and \( \hat{\sigma}_D \) represent components of variance attributable to sires, dams, pooled sire plus dam effects, dominance and/or maternal effects, and within family variation, respectively. All variance components have been standardized so that total variance = 1.0. Negative estimates are due to sampling error, and may be considered to indicate zero variance attributable to a particular component. Estimates of variance components are from an unweighted, delete-one jackknife t-tests. * indicates P < 0.05, ** indicates P < 0.01, and *** indicates P < 0.001. Narrow sense or broad sense estimates of heritabilities may be estimated by multiplying \( \hat{\sigma}_D \) or \( \hat{\sigma}_D \) by 4/(1 + 1/\( n \)) = 2.72.

Significance tests on \( \hat{\sigma}_G, \hat{\sigma}_D, \) and \( \hat{\sigma}_{BM} \) were slightly conservative, rejecting a true null hypothesis 3.5%, 2.1% and 1.8% of the time at a nominal \( \alpha = 5\% \) level when \( n = 250 \). The jackknife test of significance was extremely robust to the presence of heterogeneous within-family variances: simulations with heteroscedastic within-family variances produced type I and II error rates virtually identical to those portrayed in Figure 1 (data not shown). In the absence of dominance and maternal variance, tests using \( \hat{\sigma}_{BM} \) were more likely to detect the existence of additive genetic variance than \( \hat{\sigma}_G \); the power of jackknife t-tests increased from \( \hat{\sigma}_G \) to \( \hat{\sigma}_{BM} \) to \( \hat{\sigma}_D \). With 250 individuals, a heritability of 40% was detectable only 33% of the time using \( \hat{\sigma}_G \), while \( \hat{\sigma}_D \) was found to be 'significant' 90% of the time at this heritability. However, due to the low statistical power of tests on \( \hat{\sigma}_{BM} \) (Figure 1), caution must be used when a strictly additive interpretation of \( \hat{\sigma}_D \) is being considered. At low levels of genetic variation it is unlikely that heritabilities could be shown to be significant: when \( h^2 = 10\% \) the null hypothesis was rejected in 11% and 18% of the trials using \( \hat{\sigma}_G \) and \( \hat{\sigma}_{BM} \), respectively. However, larger sample sizes (e.g., \( n = 500 \), Figure 1) gave substantially greater power for detection of moderate heritabilities.

**Major gene analysis:** Most characters showed strongly heterogeneous within-family variances (Table 2). Graphical and regression analysis (not shown) indicated that within-family variance increased with increasing trait values in both populations for all characters except germination date. This pattern of increasing variances presumably represents a scaling effect. However, in both populations germination date had maximum within-family variance in intermediate families. We tested this hypothesis against hypotheses that within-family variability increases or decreases monotonically, since maximum variability at extreme trait values might be an artifact of scale of measurement (DRAF and SMITH 1981). Using Equation 5 of MITCHELL-OLDS and SHAW (1987), the monotonic hypotheses are strongly rejected (Table 3, \( \hat{\sigma}_G \) and \( \hat{\sigma}_{BM} \) use \( \hat{\sigma}_D = 0 \). Thus FAI's (1978) predicted pattern of maximum

**TABLE 1**

<table>
<thead>
<tr>
<th>Population</th>
<th>( \hat{\sigma}_G )</th>
<th>( \hat{\sigma}_D )</th>
<th>( \hat{\sigma}_{BM} )</th>
<th>( \hat{\sigma}_D )</th>
<th>( \hat{\sigma}_D )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Madison:</td>
<td>SW ~0.251 0.769*** 0.259*** 1.019*** 0.482</td>
<td>GD ~0.107 0.925* 0.109*** 0.482 0.782</td>
<td>JS ~0.057 0.052 ~0.001 0.105 1.002</td>
<td>EG ~0.002 0.029 0.015 0.028 0.969</td>
<td>LG ~0.068 0.052 ~0.008 0.120 1.016</td>
</tr>
<tr>
<td>Milwaukee:</td>
<td>SW 0.098 0.361* 0.229*** 0.264 0.541</td>
<td>GD 0.147* ~0.153 ~0.003 ~0.300 1.006</td>
<td>JS 0.070 ~0.012 0.029 ~0.082 0.942</td>
<td>EG ~0.008 0.035 0.131 0.041 0.975</td>
<td>LG ~0.040 ~0.045 ~0.003 ~0.085 1.005</td>
</tr>
<tr>
<td>Pooled:</td>
<td>SW ~0.109 0.607*** 0.249*** 0.716*** 0.502</td>
<td>GD ~0.051 0.203* 0.076*** 0.254 0.848</td>
<td>JS ~0.016 0.052 0.181 0.069 0.964</td>
<td>EG ~0.014 0.063 0.025 0.077 0.951</td>
<td>LG ~0.071 0.131 0.030 0.202 0.939</td>
</tr>
</tbody>
</table>

*Significance tests on \( \hat{\sigma}_G, \hat{\sigma}_D, \) and \( \hat{\sigma}_{BM} \) are slightly conservative, rejecting a true null hypothesis 3.5%, 2.1% and 1.8% of the time at a nominal \( \alpha = 5\% \) level when \( n = 250 \). The jackknife test of significance was extremely robust to the presence of heterogeneous within-family variances: simulations with heteroscedastic within-family variances produced type I and II error rates virtually identical to those portrayed in Figure 1 (data not shown). In the absence of dominance and maternal variance, tests using \( \hat{\sigma}_{BM} \) were more likely to detect the existence of additive genetic variance than \( \hat{\sigma}_G \); the power of jackknife t-tests increased from \( \hat{\sigma}_G \) to \( \hat{\sigma}_{BM} \) to \( \hat{\sigma}_D \). With 250 individuals, a heritability of 40% was detectable only 33% of the time using \( \hat{\sigma}_G \), while \( \hat{\sigma}_D \) was found to be 'significant' 90% of the time at this heritability. However, due to the low statistical power of tests on \( \hat{\sigma}_{BM} \) (Figure 1), caution must be used when a strictly additive interpretation of \( \hat{\sigma}_D \) is being considered. At low levels of genetic variation it is unlikely that heritabilities could be shown to be significant: when \( h^2 = 10\% \) the null hypothesis was rejected in 11% and 18% of the trials using \( \hat{\sigma}_G \) and \( \hat{\sigma}_{BM} \), respectively. However, larger sample sizes (e.g., \( n = 500 \), Figure 1) gave substantially greater power for detection of moderate heritabilities.

**Major gene analysis:** Most characters showed strongly heterogeneous within-family variances (Table 2). Graphical and regression analysis (not shown) indicated that within-family variance increased with increasing trait values in both populations for all characters except germination date. This pattern of increasing variances presumably represents a scaling effect. However, in both populations germination date had maximum within-family variance in intermediate families. We tested this hypothesis against hypotheses that within-family variability increases or decreases monotonically, since maximum variability at extreme trait values might be an artifact of scale of measurement (DRAF and SMITH 1981). Using Equation 5 of MITCHELL-OLDS and SHAW (1987), the monotonic hypotheses are strongly rejected (Table 3, \( \hat{\sigma}_G \) and \( \hat{\sigma}_{BM} \) use \( \hat{\sigma}_D = 0 \). Thus FAI's (1978) predicted pattern of maximum

**FIGURE 1.**—Power curves for variance components. Percent probability of rejection of null hypothesis that a particular variance component equals zero plotted against percent of total variance attributable to that particular component of variance. From bottom to top, the first three curves show results with \( n = 250 \), representing \( \hat{\sigma}_G, \hat{\sigma}_{BM} \), and \( \hat{\sigma}_D \), respectively. The topmost curve indicates \( \hat{\sigma}_D \) when \( n = 500 \). When percent of variance equals zero then the null hypothesis is true, so values at intersection with left vertical axis are type I error rates. When percent of variance attributable to a particular component is greater than zero then each curve indicates the power of the test (i.e., the probability of rejecting the null hypothesis when it is false, or one minus the type II error rate). Simulations involving \( \hat{\sigma}_G \) and \( \hat{\sigma}_{BM} \) assume \( \hat{\sigma}_D = 0 \), and simulations involving \( \hat{\sigma}_D \) and \( \hat{\sigma}_{BM} \) assume \( \hat{\sigma}_D = 0 \).
variability in intermediate families is observed for germination date in both populations.

**DISCUSSION**

**Genetic and maternal variation:** Quantitative geneticists use broad sense heritabilities as a rough estimate of the overall level of genetic variation for a trait. Broad sense estimates may be easier to obtain than narrow sense estimates, but they are likely to be upwardly biased by nonadditive genetic variance (FALCONER 1981). Comparison between broad sense and narrow sense estimators typically involve a tradeoff between unbiased estimators of low precision (e.g., \( \sigma^2_g \)) and biased estimators with greater statistical power (\( \sigma^2_M \)). In partially inbred populations, the regression of offspring on parent includes part of the dominance variance, so choice of optimal experimental design procedures can be especially difficult in natural plant populations (COCKERHAM 1963; MITCHELL-OLDS and RUTLEDGE 1986). We chose to investigate broad sense estimators in order to take advantage of their generally greater statistical precision (lower mean square error) and hopefully detect family effects on life history variation that were not detectable in the narrow sense heritability analyses of MITCHELL-OLDS (1986).

Yet, we found that most estimates of broad sense heritabilities do not differ significantly from zero in either the Madison or Milwaukee populations, or when data from both populations are pooled.

These analyses also permit examination of evidence for dominance genetic and maternal effects for life history traits. In the Madison population both seed weight and germination date show significant dominance and/or maternal variance. This agrees with many studies on the genetics of plant and animal populations (FALCONER 1981; ATCHLEY 1984; ROACH and WULFF 1987), in which maternal effects are often found to be most important in early stages of the life cycle. The Milwaukee population shows additive genetic variance for germination date, and a significant broad sense heritability for seed weight. This broad sense variation apparently is not due to dominance or maternal effects, although the low statistical power of the test on \( \sigma^2_M \) must be considered.

Although we had anticipated that broad sense estimates would reveal statistically significant heritabilities for more characters than the narrow sense estimates calculated previously (MITCHELL-OLDS 1986), none of the mid- or late-life cycle characters had detectable genetic or maternal variances. This indicates that virtually all observed phenotypic variation is due to environmental differences. In contrast to previous results from the Milwaukee population (MITCHELL-OLDS 1986), plant dry weight, a major indicator of individual fitness, was not significant in these new analyses (Table 1). Although the jackknife tests of significance were somewhat conservative (Figure 1), this does not explain the failure to attain statistical significance with these estimators. The jackknife estimator of broad sense heritability, \( H^2_j \), was substantially lower than the previous least squares estimate of narrow sense heritability, \( H^2_L = 0.29 \). This difference is due primarily to use of a larger set of families than before (including families where sire and dam effects are confounded), as well as the negative correlation expected between estimates of the sires and dams components of variance (Figure 2).

What biological conclusion can be drawn from contradictory results from two statistical tests for heritability of fitness? Although each estimation and testing procedure is statistically valid (SEARLE 1971; SOKAL and ROHLF 1981; WC 1986), our Monte Carlo simulations provide information on the weight that may

---

**TABLE 2**

<table>
<thead>
<tr>
<th>Character</th>
<th>Madison</th>
<th>Milwaukee</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( V )</td>
<td>( F )</td>
</tr>
<tr>
<td>Seed weight</td>
<td>21.3</td>
<td>2.035***</td>
</tr>
<tr>
<td>Germination date</td>
<td>11.0</td>
<td>1.473*</td>
</tr>
<tr>
<td>June size</td>
<td>10.5</td>
<td>1.447*</td>
</tr>
<tr>
<td>Early growth</td>
<td>17.8</td>
<td>1.831***</td>
</tr>
<tr>
<td>Late growth</td>
<td>15.0</td>
<td>1.674**</td>
</tr>
<tr>
<td>Fitness</td>
<td>33.0</td>
<td>2.884***</td>
</tr>
</tbody>
</table>

Levene’s test (LEVENE 1960) for heterogeneity of within full-sib family variances. The dependent variable is the absolute value of individual deviation from full-sibship mean, \( |z_i - \bar{z}| \). One way analyses of variance of absolute deviations give \( V \), the percent of differences in variability attributable to full-sib family, and \( F \), the F-ratio from Levene’s test ANOVA. Madison df = 59, 70. Milwaukee d.f. = 44, 157. Significance levels as in Table 1.

---

**TABLE 3**

<table>
<thead>
<tr>
<th>Population</th>
<th>Variance vs. mean</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Madison</td>
<td>Increasing</td>
<td>5.510</td>
<td>0.001</td>
</tr>
<tr>
<td>Madison</td>
<td>Decreasing</td>
<td>5.096</td>
<td>0.001</td>
</tr>
<tr>
<td>Milwaukee</td>
<td>Increasing</td>
<td>6.737</td>
<td>0.001</td>
</tr>
<tr>
<td>Milwaukee</td>
<td>Decreasing</td>
<td>4.302</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Hypothesis test of whether within full-sib family variance is a monotonic increasing or decreasing function of family mean, employing the jackknife test of significance developed by MITCHELL-OLDS and SHAW (1987). Equation 5. The statistical test is similar to traditional tests of a general linear hypothesis, employing a second order polynomial regression, and asking whether the maximum occurs at an intermediate value, as indicated by least squares regression, or whether the maximum can be forced to occur at the most extreme (minimum or maximum) observed value of family mean, corresponding to monotonic decreasing or increasing functions, respectively.
be given to the alternative conclusions. The pooled estimator, \( \hat{\sigma}_{db}^2 \), provides greater statistical precision (48% smaller standard error of the estimate in Monte Carlo simulations, data not shown) and generally greater statistical power (Figure 1) than \( \hat{\sigma}^2 \). Consequently, we are more inclined to believe the results from the present analysis, indicating that there is little or no additive genetic variation for plant size (and therefore level of seed production) in these populations of *L. capensis*. Nevertheless, other studies on this species (WALLER 1984; MITCHELL-OLDS and WALLER 1985) have found inbreeding depression for major components of fitness, which implies the existence of dominance genetic variance in these populations (GROW and KIMURA 1970). Failure to detect significant broad sense heritabilities for several characters in this study is likely due to the low statistical power of variance component analysis (Figure 1).

**Major genes and phenotypic variance:** We found heterogeneity of within-family variances for nearly all characters in both populations. In most cases, variance increased monotonically with family mean, indicating a simple effect of scale of measurement. The one exception to this pattern was germination date, which had maximum within-family variance at intermediate trait values in both populations. There are several possible explanations for heterogeneous patterns of within-family variation. Heterogeneity of family variances might be due to 1) artifactual effects of scale of measurement, 2) differences in developmental homeostasis, 3) mortality selection against extreme phenotypes, 4) variation in levels of inbreeding between sibships, 5) segregation of major genes, or 6) genes for variability *per se*.

Several of these hypotheses can be eliminated as explanations of our results. Each will be considered in turn: 1) If heterogeneity of variances is an effect of scale of measurement then we would expect greater variability at larger values of the trait (SOKAL and ROHLF 1981). This pattern is observed for most characters in both populations, but not for germination date. 2) Differences in developmental homeostasis are expected to cause high variability in families that deviate from the mean or optimum, but lower variance near the center of the distribution (LERNER 1954; MITTEN and GRANT 1984). This is reverse to the observed pattern (Figure 3). In related work, MITCHELL-OLDS and WALLER (1985) tested for differences in developmental homeostasis in several leaf characters, but found no evidence for such canalization in those traits. 3) Mortality selection against extreme individuals might reduce variance in extreme families. However, levels of mortality in this experiment were extremely low (MITCHELL-OLDS 1986), so this factor cannot explain these observations. 4) Variability in level of inbreeding among families might cause heterogeneity of family variances (COCKERHAM 1963; MCCALL, MITCHELL-OLDS and WALLER 1989). We avoided this problem by performing controlled crosses among unrelated individuals to create outbred progenies. Note that this factor could not be eliminated had open pollinated maternal sibships simply been collected in the field.

The last two hypotheses cannot be distinguished with our data: 5) FAIN (1978) suggested that larger variance is expected in phenotypically intermediate families if they are segregating at a locus of large effect, while families at both extremes would be homozygous, and consequently have less segregating genetic variance. In the Madison population there is a significant dominance and/or maternal influence on germination date, so that differences in full-sibship variability might not be of genetic origin. However, in the Milwaukee population there is no trace of maternal influence on germination date. Consequently, different levels of variability among full-sib families must be due to genetic differences. Therefore, we may be seeing effects of major genes for quantitative characters, or 6) effects of genes that directly influence variability of phenotypic expression, *per se*. While these data cannot distinguish between these possibilities, they are both of considerable evolutionary interest.

The relative importance of polygenic variation vs. loci of large effect (major genes or "leading factors")
FIGURE 3.—Scatter plot of absolute deviation of each individual from its full-sib family mean, |z - \bar{z}_i|, against full-sib family mean value in Madison (A) and Milwaukee (B). Data have been Box-Cox transformed from their original germination cohort values. Only families with two or more progeny are shown. Families have significantly heterogeneous within family variances by Levene’s test (Table 2), and maximum variability occurs at intermediate germination dates (Table 3). This is exactly the pattern predicted by FAIN (1978) as an indicator of segregation of major genes. This pattern is expected if intermediate families are segregating at a locus of large effect, while families at both extremes are homozygous and consequently have less segregating genetic variance.

(WRIGHT 1968) has been the subject of continuing debate for many years (WRIGHT 1968; LANDE 1983; GOTTLIEB 1984, 1985; COYNE and LANDE 1985; SHRIMPTON and ROBERTSON 1988a, b; SING et al. 1988; LANDER and BOTSTEIN 1989) and may influence strongly the response to selection. Quantitative geneticists have argued convincingly that most metrical characters are under polygenic control (FALCONER 1981; LANDE 1981; BARTON and TURELLI 1989), although WRIGHT (1968, p. 411) has suggested that “It is probable that a leading factor or factors could be isolated much more often than has been the case.” Although existing statistical approaches provide limited insight regarding this problem (MAYO, HANCOCK and BAGHURST 1980; Cockerham 1986), maximum likelihood analysis of segregation in pedigrees permits examination of discrete or metrical characters with flexible models incorporating major loci and polygenic variance (E. A. THOMPSON and MITCHELL-OLDS, unpublished results). With the advent of extensive genetic maps employing large numbers of codominant molecular markers, the potential exists for resolution of this question in the near future (MICHELMORE and HULBERT 1987; CHANG et al. 1988; PATERNSON et al. 1989; MICHELMORE and SHAW 1988; LANDER and BOTSTEIN 1989).

Genetic influences on levels of phenotypic variability have been analyzed in plants and animals (MITTON and GRANT 1984; MITCHELL-OLDS and WALLER 1985; SCHLICTING 1986; LEARY and ALLENDORF 1989). Genetic differences in fluctuating asymmetry have been detected in many studies (e.g., VRIJENHOEK and LERMAN 1982; LEAMY 1984; LEARY et al. 1985), although MITCHELL-OLDS and WALLER (1985) found that selfed and outcrossed plants do not differ in their levels of phenotypic variance for leaf shape in *I. capensis*. The pattern of genetic changes in variability observed (Figure 3) are exactly opposite to the predictions of standard theories of developmental homeostasis. Better understanding of the evolutionary implications of these observations would require detailed genetic analysis of these populations. However, regardless of the genetic basis of these heterogeneous variances, they have important statistical consequences. Most parametric tests of statistical significance assume homogeneous error variances (SEARLE 1971; SOKAL and ROHLF 1981), and violation of this assumption may invalidate statistical and biological conclusions. Accordingly, examination of family variances is an essential aspect of quantitative genetic analyses which influences genetic and statistical interpretation.

Statistical methodology: Analysis of quantitative genetics in natural populations has been hindered by computational and methodological problems in statistical analysis (MITCHELL-OLDS and RUTLEDGE 1986; SHAW 1987). Maximum likelihood offers many statistical advantages but proved computationally infeasible.
in this instance. We have shown that jackknife analysis of variance components provides a flexible method for estimation and testing of biologically important questions regarding influences on life history variation. For example, classical least square methods do not allow tests of dominance/maternal effects. Furthermore, these jackknife tests of significance proved to be extremely robust to heterogeneity of within family variances.

When statistical procedures, such as maximum likelihood or resampling methods, rely on asymptotic properties, then Monte Carlo simulations are essential to provide full confidence in estimates and tests. These simulations permit evaluation of rates of type I and type II errors, bias and standard error of an estimator, performance at finite sample sizes, and robustness to assumption violations. It is essential to understand the statistical behavior of one's chosen method for proper biological interpretation of statistical results. Failure to detect significant levels of genetic variance has little meaning if statistical power is low, but can be taken as strong evidence for lack of genetic variation when statistical power is high. For example, in the pooled analysis of the Madison and Milwaukee populations together (n = 454), we can conclude with strong statistical confidence thatheritabilities for most characters are less than 0.20 (Figure 1). We urge explicit consideration of statistical characteristics of estimators, and in particular of statistical power, in quantitative genetic analyses.

We are grateful to F. Alfeldorf, C. Denniston, R. Nakamura, F. Thompson, D. Wiernasz, S. Via, D. Waller, B. Weir, and three anonymous reviewers for comments and discussion. T.M.O. was supported by National Science Foundation grants BSR-83 11817, BSR-8412172, and U.S. Department of Agriculture Competitive Grant 88-37151-3958. J.M.B. was supported by a National Science Foundation predoctoral fellowship and a National Science Foundation dissertation improvement grant. This contribution number 102 from the University of Montana Herbarium and Center for Plant Diversity.

**LITERATURE CITED**


Efron, B., 1982 The jackknife, the bootstrap, and other resampling plans. Society for Industrial and Applied Mathematics, Philadelphia.


Lande, R., 1981 The minimum number of genes contributing to quantitative genetic variation between and within populations. Genetics 99: 541–553.


Mayo, O., T. W. Hancock and P. A. Baghurst, 1980 Influence


Communicating editor: B. S. Weir