Genet. Res., Camb. (2003), **82**, *pp.* 41–53. *With 6 figures.* © 2003 Cambridge University Press
DOI: 10.1017/S0016672303006311 Printed in the United Kingdom

Detection of deleterious genotypes in multigenerational studies. III. Estimation of selection components in highly selfing populations

RENYI LIU[1], ALAN M. FERRENBERG[2], LAURA U. GILLILAND[1], RICHARD B. MEAGHER[1] AND MARJORIE A. ASMUSSEN[1]*

[1] *Department of Genetics, University of Georgia, Athens, GA 30602, USA*
[2] *University Computing and Networking Services, University of Georgia, Athens, GA 30602, USA*

(*Received 11 November 2002 and in revised form 21 March 2003*)

Summary

New paradigms in genetics have increased the chance of finding genes that appear redundant but in fact may have been preserved due to a small level of positive selection potential acting during each generation. Monitoring changes in genotypic frequencies within and between generations allows the dissection of the fertility, viability and meiotic drive selection components acting on such genes in natural and experimental populations. Here, a formal maximum likelihood procedure is developed to identify and estimate these selection components in highly selfing populations by fitting the time-dependent solutions for genotypic frequencies to observed multigenerational counts. With adult census alone, we can not simultaneously estimate all three selection components considered. In such cases, we instead consider a hierarchy of 11 models with either fewer selection components, complete dominance, or multiplicative meiotic drive with a single parameter. We identify the best-fitting of these models by applying likelihood ratio tests to nested models and Akaike's Information Criterion (*AIC*) and the Bayesian Information Criterion (*BIC*) to non-nested models. With seed census, fertility and viability selection are not distinguishable and thus can only be estimated jointly. A combination of joint seed and adult census data allows us to estimate all three selection components simultaneously. Simulated data validate the estimation procedure and provide some practical guidelines for experimental design. An application to *Arabidopsis* data establishes that viability selection is the major selective force acting on the *ACT2* actin gene in laboratory-grown *Arabidopsis* populations.

1. Introduction

In the last decade classical searches for mutant phenotypes in populations have given way to new sequence-based approaches of gene identification and characterization (McKinney *et al*., 1995; Arabidopsis Genome Initiative, 2000), with phenotypes to be determined later. This new paradigm in genetics has increased the chance of finding genes that appear redundant but in fact may have been preserved due to a small level of positive selection potential acting during each generation (Gilliland *et al*., 1998; Krakauer & Nowak, 1999; Martienssen & Irish, 1999; Tautz, 2000). However, measuring natural selection on genes is necessarily difficult when they are subject only to small incremental selection in each generation (Kimura, 1991; Ohta, 1992). Here, we formalize our early work (Asmussen *et al*., 1998; Gilliland *et al*., 1998) showing that multigenerational studies of genotypic frequencies may be the key to obtaining statistically significant measures of the selection potential of such genes.

There may be several selection components acting simultaneously on one genetic marker, such as viability, fertility and gametic selection. Accurately estimating the form and strength of these selection components allows us to understand this evolutionary process in detail. Several approaches have been used to estimate selection components in both animal and plant populations. One approach is to design a series of essentially separate experiments to estimate individual selection components (Prout, 1971; Bundgaard

* Corresponding author. Tel: +1 (706) 5421455. Fax: +1 (706) 5423910. e-mail: asmussen@uga.edu

https://doi.org/10.1017/S0016672303006311 Published online by Cambridge University Press

& Christiansen, 1972). A second method utilizes population structure information by taking mother–offspring combinations in conjunction with male population samples (Christiansen & Frydenberg, 1973). As a third approach, Clegg *et al.* (1978) estimated selection components based on multiple census data within each generation from a population.

Another approach, closely related to ours, was developed by DuMouchel & Anderson (1968), who set up a general maximum likelihood method for estimating selection parameters in experimental, random-mating populations. This procedure uses numerical methods to find a set of selection values that maximize likelihoods, which are functions of observed and expected genotypic frequencies over multiple, non-overlapping generations. The expected frequencies are calculated by iterating their recursion equations from the initial, known allele frequencies, under random mating. A chi-square criterion is used to test the goodness-of-fit of the model. This method was applied successfully to lethal loci in several *Drosophila melanogaster* populations (Anderson, 1969).

In this study, we develop a formal method for estimating multiple selection components in highly selfing populations by fitting a mathematical model with explicit time-dependent solutions to observed multigenerational genotypic counts. The underlying model for adult data was derived previously and provides the expected genotypic frequencies in adults at a diallelic locus in a selfing, diploid population under any combination of fertility, viability and gametic selection (Asmussen *et al.*, 1998). As a preliminary application, this model was applied to an experimental *Arabidopsis* population, which is ideal for this kind of study because of its small size, short life cycle, small genome, and negligible rate of outcrossing (Meyerowitz, 1989, 1994; Abbott & Games, 1989). Based on informal, *ad hoc* methods this model had a good fit to the first three generations of actin gene data, assuming either fertility or viability selection alone.

The present paper extends the previous study in five critical ways. First, a formal maximum likelihood procedure is developed and used to estimate selection parameters by fitting the explicit time-dependent solutions for genotypic frequencies to observed multigenerational data. Second, we use the simulated annealing algorithm to find the estimates for the parameters, and the profile likelihood method to obtain their confidence intervals. Third, two additional generations of data are used in the current estimation procedure for greater accuracy. Fourth, in addition to the adult census scheme considered before, we derive the genotypic frequency dynamics and estimation procedures for selection components under both a seed census and a joint adult/seed census. Fifth, simulated data are used to test the estimation procedure and deduce practical guidelines for experimental design.

Table 1. *Genotypic frequencies, counts and selection parameters*

	Genotype			
	A_1A_1	A_1A_2	A_2A_2	Total
Adult frequencies	u_{11}	u_{12}	u_{22}	1
Seed frequencies	s_{11}	s_{12}	s_{22}	1
Adult counts	N_{11}	N_{12}	N_{22}	N
Seed counts	n_{11}	n_{12}	n_{22}	n
Fertility parameters	f_{11}	f_{12}	f_{22}	N/A
Viability parameters	v_{11}	v_{12}	v_{22}	N/A
Meiotic drive parameters	m_{11}	m_{12}	m_{22}	1

2. The selfing model with adult census

The underlying selection model for an adult census was developed by Asmussen *et al.* (1998). This framework assumes two alleles, A_1 and A_2, at a diallelic, autosomal nuclear locus in a purely selfing, diploid population with discrete, non-overlapping generations, no mutation, gene flow, or seed dormancy, and large enough to preclude the effects of random genetic drift. The genotypes are subject to constant fertility selection, viability selection, and gametic selection via meiotic drive. Adult and seed frequencies and selection parameters for the three possible genotypes are as shown in Table 1. Following the notation of Asmussen *et al.* (1998), u_{ij} is the frequency and N_{ij} the count, of genotype A_iA_j in adults for $i \leqslant j = 1, 2$. The corresponding seed frequency and count are denoted by s_{ij} and n_{ij}, respectively. The fertility parameter f_{ij} represents the average number of offspring produced by an A_iA_j adult, whereas the viability parameter v_{ij} is the probability that an A_iA_j zygote survives to reproduce. The final selection parameter m_{ij}, for meiotic drive, is the average proportion of offspring that are A_iA_j when an A_1A_2 heterozygote selfs. The three meiotic drive parameters are non-negative numbers that sum to 1 ($m_{11}+m_{12}+m_{22}=1$).

Under these conditions the time-dependent solutions for the adult genotypic frequencies in each generation $t \geqslant 1$ are:

$$u_{11}^{(t)} = \frac{x_1^{(t)}}{1+x_1^{(t)}+x_2^{(t)}}, \tag{1}$$

$$u_{12}^{(t)} = \frac{1}{1+x_1^{(t)}+x_2^{(t)}}, \tag{2}$$

$$u_{22}^{(t)} = \frac{x_2^{(t)}}{1+x_1^{(t)}+x_2^{(t)}}, \tag{3}$$

where for $i=1, 2$, $x_i^{(t)} = \dfrac{u_{ii}^{(t)}}{u_{12}^{(t)}}$ is the ratio of the frequency

https://doi.org/10.1017/S0016672303006311 Published online by Cambridge University Press

of A_iA_i to A_1A_2 adults in generation t, and

$$x_i^{(t)} = \begin{cases} (a_i)^t(x_i^{(0)} - x_i^*) + x_i^* & \textit{if } a_i \neq 1 \\ x_i^{(0)} + tb_i & \textit{if } a_i = 1, \end{cases} \tag{4}$$

$$a_i = \frac{f_{ii}v_{ii}}{m_{12}f_{12}v_{12}}, \tag{5}$$

$$b_i = \frac{m_{ii}v_{ii}}{m_{12}v_{12}}, \tag{6}$$

$$x_i^* = \frac{b_i}{1 - a_i} = \frac{f_{12}m_{ii}v_{ii}}{m_{12}f_{12}v_{12} - f_{ii}v_{ii}}. \tag{7}$$

The derivation details are shown in Asmussen *et al.* (1998), which also pointed out that this selfing model can also be applied with reasonable accuracy to partially random-mating populations in which the outcrossing rate is negligible.

3. Parameter estimation with adult census

(i) *The selection parameters to be estimated*

With the three selection components in our model, there are six independent parameters to be estimated: two for each component. This is because the three meiotic drive parameters sum to 1, and the frequency dynamics are identical whether the absolute or relative values are used for the fertility and viability components. Here, the relative fertility F_i and relative viability V_i for A_iA_i homozygotes are normalized with respect to a value of 1 for heterozygotes, i.e.

$$F_i = \frac{f_{ii}}{f_{12}} \tag{8}$$

and

$$V_i = \frac{v_{ii}}{v_{12}}, \tag{9}$$

for $i = 1, 2$.

Ideally, we could estimate all six independent selection parameters simultaneously. However, since the dynamics of the adult genotypic frequencies in equations (1)–(7) are totally determined by four composite parameters (a_1, a_2, b_1 and b_2), we can estimate at most four selection parameters at a time if only adults are censused. Subject to this constraint, we instead determine the best-fitting model among 11 cases with either fewer selection components, complete dominance, or multiplicative gametic selection with a single meiotic drive parameter (m) (Table 2).

(ii) *The log-likelihood function*

Under the assumption that the genotypic counts in each generation have a trinomial distribution (Weir, 1996) and the sampling for each generation is independent, the overall log-likelihood function L for g consecutive generations of sampling is the sum of the log-likelihoods for each generation, and can be written as

$$L = C + \sum_{t=1}^{g}\sum_{i=1}^{2}\sum_{j=i}^{2}(N_{ij}^{(t)} \ln u_{ij}^{(t)}), \tag{10}$$

where C is a constant depending on the observed counts, the genotypic frequencies $u_{ij}^{(t)}$ at time t are functions of the selection parameters and the initial genotypic frequencies given by equations (1)–(7), and $N_{ij}^{(t)}$ are the observed adult counts of type A_{ij}, for $i \leqslant j = 1, 2$ (Table 1). The maximum likelihood estimates are those selection parameters that maximize L.

(iii) *The estimation procedure*

The maximum likelihood estimation (MLE) procedure is applied after verifying that the multigenerational frequency data significantly deviate from selective neutrality with a G-test (Sokal & Rohlf, 1995). Since analytical solutions for the MLE estimates are not possible, we instead use numerical methods. For each model in Table 2, the estimation procedure consists of three steps.

First, the four composite parameters (a_1, a_2, b_1 and b_2) are calculated and used to express the log-likelihood function L in equation (10) in terms of the selection parameters to be estimated (Table 2, column 4).

Second, simulated annealing (SA) is used to find the set of parameter values that maximize L. SA is an algorithm that optimizes multidimensional functions efficiently. The basic idea is as follows. SA searches the whole parameter space for each parameter and tries to find the global optimum of the objective function. It moves both up and downhill as the optimization process proceeds, and it focuses on the most promising area of the parameter space. The computer code and full description of SA are provided by Goffe *et al.* (1994). The parameter spaces for F_i and V_i are both taken to be $(0,\ 1{\cdot}0 \times 10^{25})$, where the upper bound, $1{\cdot}0 \times 10^{25}$, is simply an arbitrarily large number chosen to include virtually all possible relative viabilities and fertilities for the two homozygotes; for m_{ij} and m we use $(0, 1)$.

Third, we use the profile likelihood method to obtain approximate 95% confidence intervals for the estimated parameters. The results are similar to those from a bootstrap with the advantage of being computationally more efficient (Lowther & Skalski, 1996).

(iv) *Identifying the best-fitting model*

As shown in Fig. 1, the 11 models considered (Table 2) have four levels of complexity, based on the

https://doi.org/10.1017/S0016672303006311 Published online by Cambridge University Press

https://doi.org/10.1017/S0016672303006311 Published online by Cambridge University Press

Table 2. *Selection parameters to be estimated in the 11 models with only adults censused*

Selection components[a]	Parameters estimated	Parameter conditions	Composite parameters[b]
Fertility and meiotic drive	F_1, F_2, m_{11}, m_{22}	$V_1=V_2=1$	$a_i=F_i/(1-m_{11}-m_{22})$, $b_i=m_{ii}/(1-m_{11}-m_{22})$, $i=1, 2$
Viability and meiotic drive	V_1, V_2, m_{11}, m_{12}	$F_1=F_2=1$	$a_i=V_i/(1-m_{11}-m_{22})$, $b_i=m_{ii}V_i/(1-m_{11}-m_{22})$, $i=1, 2$
Fertility and viability	F_1, F_2, V_1, V_2	$m_{11}=m_{22}=1/4$, $m_{12}=1/2$	$a_i=2F_iV_i$, $b_i=V_i/2$, $i=1, 2$
Fertility and multiplicative meiotic drive	F_1, F_2, m	$V_1=V_2=1$, $m_{11}=m^2$, $m_{12}=2m(1-m)$, $m_{22}=(1-m)^2$	$a_i=F_i/[2m(1-m)]$, $i=1, 2$ $b_1=m/[2(1-m)]$, $b_2=(1-m)/(2m)$
Viability and multiplicative meiotic drive	V_1, V_2, m	$F_1=F_2=1$, $m_{11}=m^2$, $m_{12}=2m(1-m)$, $m_{22}=(1-m)^2$	$a_i=V_i/[2m(1-m)]$, $i=1, 2$ $b_1=mV_1/[2(1-m)]$, $b_2=(1-m)V_2/(2m)$
Fertility	F_1, F_2	$V_1=V_2=1$, $m_{11}=m_{22}=1/4$, $m_{12}=1/2$	$a_i=2F_i$, $b_i=1/2$, $i=1, 2$
Viability	V_1, V_2	$F_1=F_2=1$, $m_{11}=m_{22}=1/4$, $m_{12}=1/2$	$a_i=2V_i$, $b_i=V_i/2$, $i=1, 2$
Meiotic drive	m_{11}, m_{12}	$F_1=F_2=1$, $V_1=V_2=1$	$a_i=1/(1-m_{11}-m_{22})$, $b_i=m_{ii}/(1-m_{11}-m_{22})$, $i=1, 2$
Multiplicative meiotic drive	m	$F_1=F_2=1$, $V_1=V_2=1$, $m_{11}=m^2$, $m_{12}=2m(1-m)$, $m_{22}=(1-m)^2$	$a_i=1/[2m(1-m)]$, $i=1, 2$ $b_1=m/[2(1-m)]$, $b_2=(1-m)/(2m)$
Fertility and viability with A_1 dominant and meiotic drive	f_{22}, v_{22}, m_{11}, m_{22}	$f_{11}=f_{12}=1$ $v_{11}=v_{12}=1$	$a_1=1/(1-m_{11}-m_{22})$, $a_2=f_{22}v_{22}/(1-m_{11}-m_{22})$ $b_1=m_{11}/(1-m_{11}-m_{22})$, $b_2=m_{22}v_{22}/(1-m_{11}-m_{22})$
Fertility and viability with A_2 dominant and meiotic drive	f_{11}, v_{11}, m_{11}, m_{22}	$f_{22}=f_{12}=1$ $v_{22}=v_{12}=1$	$a_1=f_{11}v_{11}/(1-m_{11}-m_{22})$, $a_2=1/(1-m_{11}-m_{22})$ $b_1=m_{11}v_{11}/(1-m_{11}-m_{22})$, $b_2=m_{22}/(1-m_{11}-m_{22})$

[a] Meiotic drive denotes the general case with arbitrary values for m_{11}, m_{12} and m_{22}; multiplicative meiotic drive denotes the special case where $m_{11}=m^2$, $m_{12}=2m(1-m)$ and $m_{22}=(1-m)^2$.

[b] Defined in equations (5) and (6).

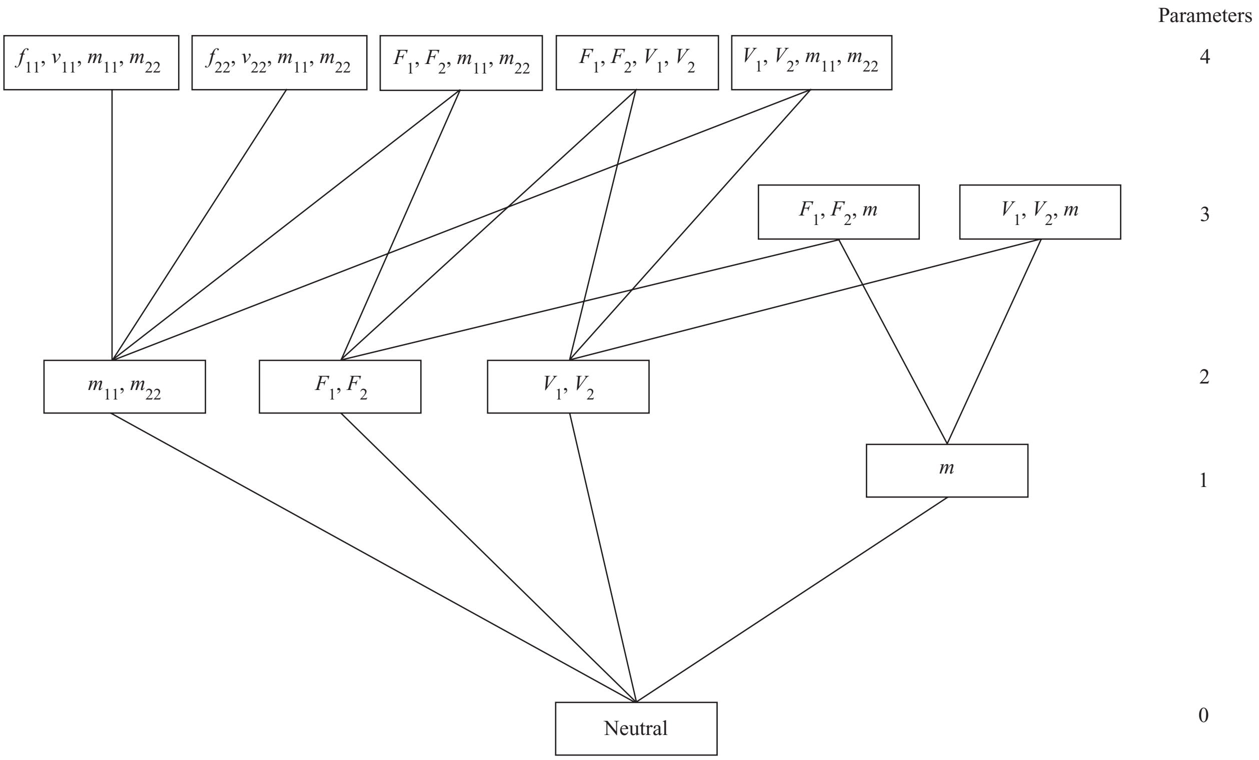
Parameters
$f_{11}, v_{11}, m_{11}, m_{22}$
$f_{22}, v_{22}, m_{11}, m_{22}$
F_1, F_2, m_{11}, m_{22}
F_1, F_2, V_1, V_2
V_1, V_2, m_{11}, m_{22}
4
F_1, F_2, m
V_1, V_2, m
3
m_{11}, m_{22}
F_1, F_2
V_1, V_2
2
m
1
Neutral
0

Fig. 1. Hierarchical display of the 11 models considered under an adult census with nested models connected by lines. Selection components are defined in Table 2 and equations (8) and (9). Selection components without parameters listed are assumed to be selectively neutral. The lowest tier with no parameters is the null model of selective neutrality.

number of parameters estimated and degrees of freedom, and have some nested structure (nested models are connected by lines). To identify the best-fitting model, we first use a *G*-test to eliminate the models which do not fit the data, i.e. those for which the observed multigenerational genotypic counts significantly deviate from those expected under their MLE selection parameter estimates. Here the *G*-statistic is

$$G = 2\sum_{t=1}^{g}\sum_{i=1}^{2}\sum_{j=i}^{2}\left[N_{ij}^{(t)}\ \ln\left(\frac{N_{ij}^{(t)}}{E_{ij}^{(t)}}\right)\right], \tag{11}$$

where $N_{ij}^{(t)}$ and $E_{ij}^{(t)}$ are the observed and expected genotypic counts of A_iA_j individuals in generation t, respectively. This test has a chi-square distribution with *2g-k* degrees of freedom, where g is the number of generations analysed from the population and k is the number of parameters that have been estimated.

Second, we find the better-fitting model within each nested pair in the remaining models. To compare two nested models M_1 and M_2 with $r_1 > r_2$ parameters, respectively, we use the likelihood ratio test (Wahrendorf *et al.*, 1987). The likelihood ratio statistic

$$LR(M_1, M_2) = 2[L(M_1) - L(M_2)], \tag{12}$$

here follows a central chi-square distribution with r_1–r_2 degrees of freedom under the null hypothesis that the fit of the two models is not significantly different, where $L(M_1)$ and $L(M_2)$ are the maximum log-likelihoods of the nested models M_1 and M_2, respectively. A significant *P*-value for this statistic indicates that model M_1 has a significantly improved fit to the observed data compared with the nested model M_2 and thus is the better-fitting model. Otherwise, M_2 is taken to be the better-fitting model, by parsimony. If a model is retained in one nested pairwise comparison but rejected in another, it is removed from further consideration.

Finally, we identify the overall best-fitting model from those still remaining via Akaike's Information Criterion (*AIC*) (Akaike, 1973) and the Bayesian Information Criterion (*BIC*) (Schwarz, 1978). The *AIC* and *BIC* for model M_i are

$$AIC_{M_i} = -2[L(M_i) - k_i] \tag{13}$$

and

$$BIC_{M_i} = -2L(M_i) + k_i\ \ln(N), \tag{14}$$

where $L(M_i)$ is the log-likelihood and k_i the number of parameters for model M_i, and N is the sample size. The only difference between the two indices is that *BIC* takes account of the sample size. The model that gives the minimum *AIC* and *BIC* is considered the best-fitting model. In most cases, especially when the sample size is large, the two indices will give a consistent result, i.e. favour the same model. Should an

https://doi.org/10.1017/S0016672303006311 Published online by Cambridge University Press

inconsistent result occur, some other method(s) of model selection, such as bootstrapping and cross-validation, could instead be applied (Forster, 2000). If no model is retained before the last step, it suggests that selection alone cannot account for the observed genotypic frequency dynamics and it is necessary to consider other possible evolutionary forces.

This best model identification process is fully automated with Perl scripts and is available from the authors upon request along with the estimation programs.

4. Estimation with seed or joint adult/seed census

Our selfing model and estimation procedure assume only adults are censused. However, in some cases, a seed census may be taken, either alone or jointly with an adult census. With seed census alone, the dynamics of the genotypic frequencies (Appendix) are again determined by four composite parameters (a_1, a_2, c_1 and c_2), where a_i is a function of the joint fertility/viability effect $J_i = F_i V_i$, and c_i only involves the meiotic drive parameters. Therefore, the model has only four parameters (J_1, J_2, m_{11} and m_{22}), which can be estimated simultaneously. The log-likelihood function is analogous to equation (10), with adult frequencies $u_{ij}^{(t)}$ and counts $N_{ij}^{(t)}$ replaced by seed frequencies $s_{ij}^{(t)}$ and counts $n_{ij}^{(t)}$, respectively.

The ultimate goal of estimating all six selection parameters (and all three selection components) simultaneously can be achieved with joint adult/seed censuses, however, because the joint dynamics of seed and adult genotypic frequencies are determined by six independent composite parameters (a_1, a_2, b_1, b_2, c_1 and c_2). The log-likelihood function for g consecutive generations of joint adult/seed data is

$$L' = C' + \sum_{t=1}^{g} \sum_{i=1}^{2} \sum_{j=i}^{2} (N_{ij}^{(t)} \ln u_{ij}^{(t)}) + \sum_{t=1}^{g} \sum_{i=1}^{2} \sum_{j=i}^{2} (n_{ij}^{(t)} \ln s_{ij}^{(t)}), \tag{15}$$

where C' is a constant, $N_{ij}^{(t)}$ and $u_{ij}^{(t)}$ are the genotypic count and frequency for A_iA_j adults, respectively, and $n_{ij}^{(t)}$ and $s_{ij}^{(t)}$ are the corresponding values in seeds in each generation t (Table 1).

This same formulation can be used with joint adult/seed data whichever life stage is censused first. If the census starts from an adult census with frequencies $u_{ij}^{(0)}$, then the seeds in generation $t \geqslant 1$ are produced by the adults in generation $t-1$. Under this convention, adult frequencies in each subsequent generation $t \geqslant 1$, $u_{ij}^{(t)}$, can be calculated as before from equations (1)–(7), and the seed frequencies can be calculated from equations (A4)–(A6), using the relation

$$y_i^{(t)} = \left(\frac{f_{ii}}{m_{12} f_{12}}\right) x_i^{(t-1)} + \frac{m_{ii}}{m_{12}}, \tag{16}$$

for $i = 1$, 2. Alternatively, if censusing starts from a seed census, the seed frequencies $s_{ij}^{(t)}$ in each generation can be calculated from the initial seed frequencies, $s_{ij}^{(0)}$, using equations (A4)–(A10). Assuming the adults in generation $t \geqslant 1$ are those which survive from the seeds in generation $t-1$, the subsequent adult frequencies can be calculated via equations (1)–(3), using the relation

$$x_i^{(t)} = \frac{v_{ii}}{v_{12}} y_i^{(t-1)}, \tag{17}$$

for $i = 1$, 2.

5. Simulation study

We tested this estimation procedure with simulated data, using the initial conditions and parameter values from the experimental *Arabidopsis* study below wherever possible. Expected genotypic frequencies were calculated under specified selection parameters for five consecutive generations and the population initiated with only heterozygotes at generation 0. The multigenerational counts for the three genotypes in consecutive generations were generated as random samples from these expected frequencies with sample sizes of 50, 75, 100, 300 and 500 each generation. The performance of the method was evaluated by the average normalized deviation of each estimated parameter $\hat{z}$ from its true value z,

$$\Delta z = \left| \frac{\hat{z} - z}{z} \right|, \tag{18}$$

and the average width of the confidence intervals with 100 runs.

With adult census, we considered two cases where populations are subject to either fertility ($F_1 = 1{\cdot}2$, $F_2 = 0{\cdot}8$) or viability selection ($V_1 = 1{\cdot}2$, $V_2 = 0{\cdot}8$). The selection form appears to have a significant effect on the accuracy of estimation. For example, the probability that the best-fitting model correctly identifies the selection form improves much faster with increasing sample size under viability selection. To achieve a probability of 80 %, 100 individuals need be sampled in each generation in the viability case, whereas more than 300 are needed when genotypes differ in fertility (Fig. 2). As shown in Fig. 3, the fertility case also produces larger normalized deviations and wider confidence intervals, which were calculated over the runs with the selection form correctly identified. A similar simulation study validated the estimation procedure with seed and joint adult/seed census under two parameter sets ($J_1 = 1{\cdot}2$, $J_2 = 0{\cdot}8$, $m_{11} = 0{\cdot}2$, $m_{22} = 0{\cdot}3$) and ($F_1 = 1{\cdot}0$, $F_2 = 0{\cdot}8$, $V_1 = 0{\cdot}7$, $V_2 = 1{\cdot}3$, $m_{11} = 0{\cdot}2$, $m_{22} = 0{\cdot}3$) (data not shown).

https://doi.org/10.1017/S0016672303006311 Published online by Cambridge University Press

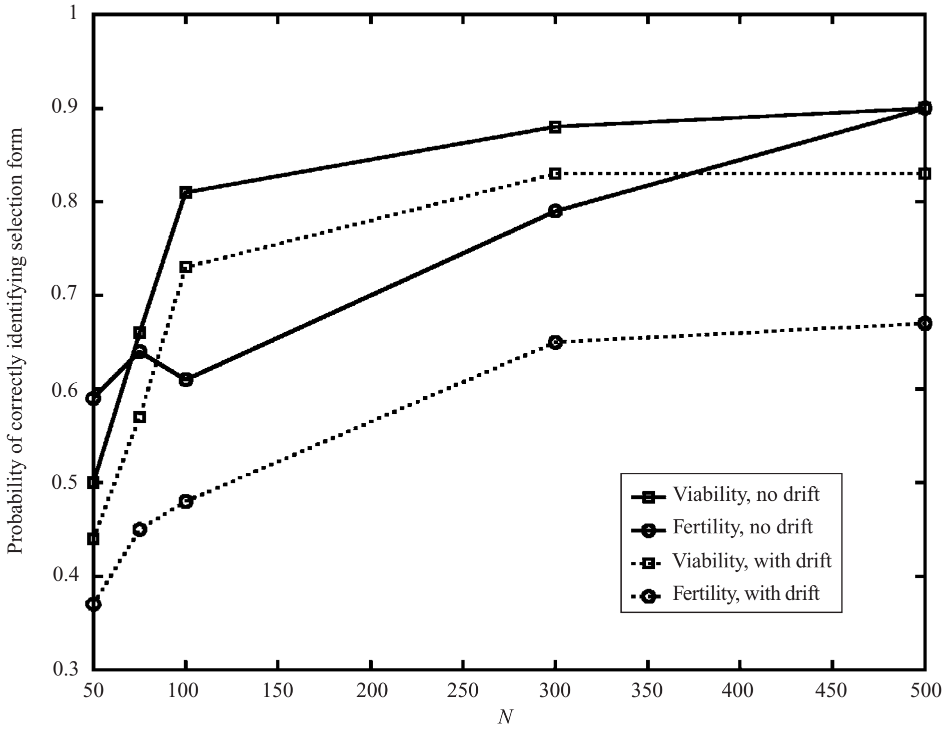

Fig. 2. The probability of identifying the correct form of selection with adult census as a function of sample size N with (dotted lines) and without (continuous lines) genetic drift. Random samples were drawn from populations in five consecutive generations with genotypic frequencies based on viability ($V_1=1{\cdot}2$, $V_2=0{\cdot}8$) or fertility selection ($F_1=1{\cdot}2$, $F_2=0{\cdot}8$) alone. With genetic drift, it is assumed that adult population size is the sample size and each adult produces an effectively infinite number of seeds.

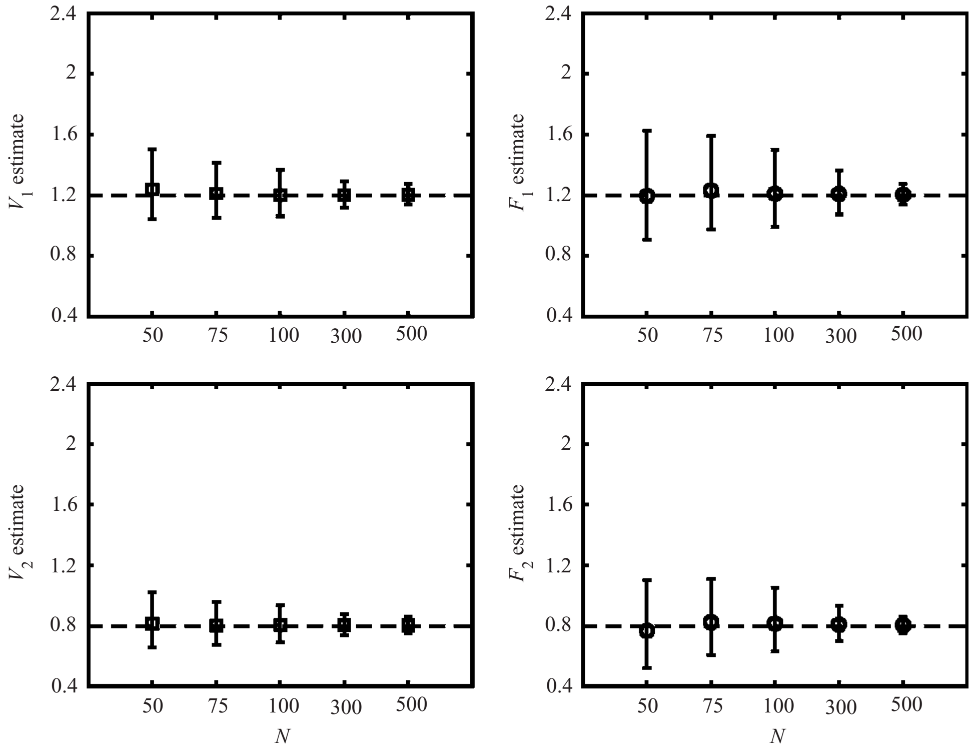

Fig. 3. Estimates and 95% confidence intervals with adult census and viability or fertility selection alone, assuming no genetic drift, as a function of sample size N. Dashed lines indicate the true values ($V_1=1{\cdot}2$, $V_2=0{\cdot}8$ or $F_1=1{\cdot}2$, $F_2=0{\cdot}8$).

https://doi.org/10.1017/S0016672303006311 Published online by Cambridge University Press

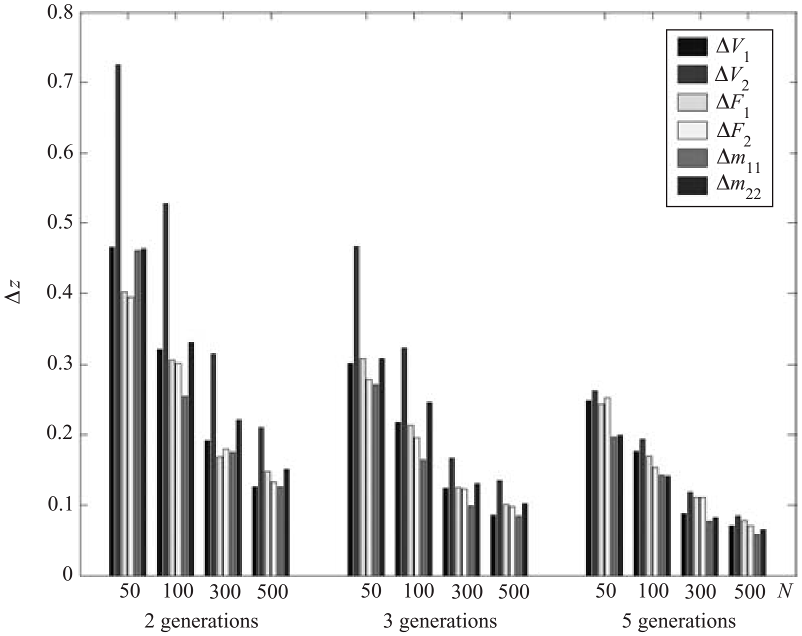

Fig. 4. Normalized deviation of estimates from the true values (Δz, defined in equation 18) with a joint adult/seed census for 2, 3 and 5 generations of data, and sample size N of 50, 100, 300 and 500 each generation. The true values are $F_1=1{\cdot}0$, $F_2=0{\cdot}8$, $V_1=0{\cdot}7$, $V_2=1{\cdot}3$, $m_{11}=0{\cdot}2$, $m_{22}=0{\cdot}3$.

Further simulations with joint adult/seed census revealed that the estimation accuracy improves quickly with the number of sample generations g and sample size per generation N (Fig. 4). An interesting observation is that with a fixed total sample size, a greater accuracy is achieved by sampling more generations (g) rather than more individuals per generation (N). For example, the two sample schemes with 100 individuals sampled in each of two generations ($g=2$, $N=100$) versus 50 sampled in each of three generations ($g=3$, $N=50$) produce similar accuracy, but this required more individuals to be sampled in the former case (400 vs 300).

Our model and estimation procedure assume that the population is sufficiently large for the effect of genetic drift to be ignored. However, in some empirical studies the population size could be limited. One example is our case study, where adult population size is limited to 100 although each adult produces a large number of progeny (more than 10 000). In such cases, the population dynamics could be skewed by the limited adult population size. We simulated this scenario by using the genotypic frequencies in our adult samples as the basis to predict the genotypic frequencies in each seed population, which is assumed to be of unlimited size. As shown in Figs 2 and 5, genetic drift makes it more difficult to identify the selection form and decreases the estimation accuracy. With a sample size of 100 from a population subject only to fertility selection ($F_1=1{\cdot}2$, $F_2=0{\cdot}8$), the probability of correctly identifying the form of selection is 48 % and the normalized deviation is around 0·25; these two numbers are 73 % and 0·13, respectively, if selection is on viability ($V_1=1{\cdot}2$, $V_2=0{\cdot}8$). We can thus apply our procedure to our *Arabidopsis* case study below and achieve a reasonable accuracy without considering genetic drift.

6. Case study

Here we illustrate our estimation procedure by using it to estimate selection components from actin gene data from an experimental *Arabidopsis* population. *Arabidopsis* has five ancient subclasses of actin with distinct patterns of spatial and temporal expression, which are involved in many cytoskeletal processes affecting plant development. As a test of our method for estimating selection components from multigenerational data, we focused on the *ACT2* genotypic frequencies. *ACT2* is one of three vegetative actin genes that are all highly expressed in root, shoots, stems and most floral organs, but *ACT2* is not expressed in pollen or ovules (An *et al*., 1996). The *act2-1* mutant allele contains a large T-DNA insertion near the start codon and has the potential to be a null allele (McKinney *et al*., 1995).

In our initial study (Gilliland *et al*., 1998), an experimental population was initialized with a single A_1A_2 heterozygote, where A_1 represents the wild-type allele *ACT2*, and A_2 the mutant allele *act2-1*. Approximately 100 adult plants were then randomly sampled and genotyped in each of the next three consecutive generations. It was found that although adult plants homozygous for the *act2-1* mutant allele

https://doi.org/10.1017/S0016672303006311 Published online by Cambridge University Press

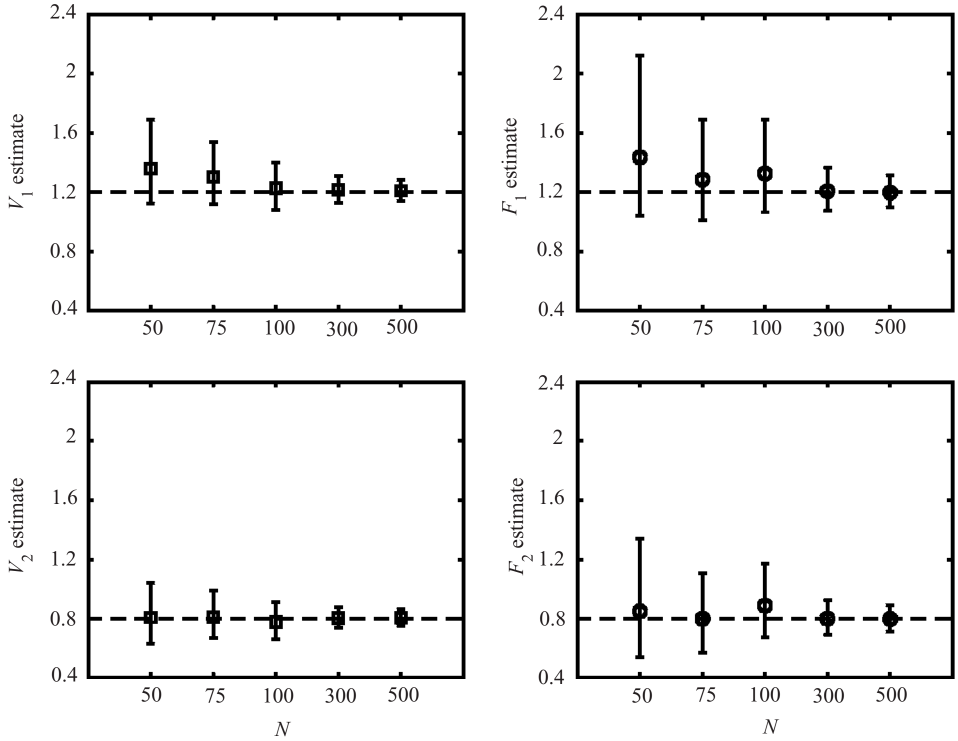

Fig. 5. Estimates and 95 % confidence intervals with adult census and viability or fertility alone, with genetic drift, as a function of sample size N. Dashed lines indicate the true values ($V_1 = 1{\cdot}2$, $V_2 = 0{\cdot}8$ or $F_1 = 1{\cdot}2$, $F_2 = 0{\cdot}8$).

appeared to be robust, morphologically normal on soil, and fully fertile, the mutant allele was at extremely low frequencies relative to the wild-type in the F_2 and F_3 generations. Our preliminary *ad hoc* analysis suggested that the *act2-1* allele is under strong negative selection potential relative to *ACT2* due to loss of either fertility or viability (Asmussen *et al.*, 1998). Here, to refine these initial results, formally estimate selection parameters, and test the robustness of our mathematical projections, this experimental population was continued for an additional two generations.

The first step in our analysis revealed a significant overall deviation from the genotypic frequencies expected in the absence of selection ($G = 78{\cdot}917$, $df = 9$, $P \ll 0{\cdot}001$); the null hypothesis of no selection is thus rejected. The second level of analysis shows that none of the 11 models considered deviates significantly from the observed data at the 0·05 significance level ($P = 0{\cdot}083$–$0{\cdot}848$). The results from the pairwise comparisons of all the nested pairs of models are shown in Table 3. This leaves us four models as the remaining candidates for the overall best-fitting model: fertility selection alone (F_1, F_2), viability selection alone (V_1, V_2), fertility and viability selection with A_1 dominant and meiotic drive (f_{22}, v_{22}, m_{11}, m_{22}), and fertility and viability selection with A_2 dominant and meiotic drive (f_{11}, v_{11}, m_{11}, m_{22}).

Table 3. *Pairwise likelihood ratio* (*LR*) *test for nested models for the* Arabidopsis *data*

Model (M_1)	Nested model (M_2)	LR	*df*	*P* value
$\boldsymbol{f_{11}}$, $\boldsymbol{v_{11}}$, $\boldsymbol{m_{11}}$, $\boldsymbol{m_{22}}$	m_{11}, m_{22}	8·666	2	0·013*
$\boldsymbol{f_{22}}$, $\boldsymbol{v_{22}}$, $\boldsymbol{m_{11}}$, $\boldsymbol{m_{22}}$	m_{11}, m_{22}	8·666	2	0·013*
$\boldsymbol{F_1}$, $\boldsymbol{F_2}$, $\boldsymbol{m_{11}}$, $\boldsymbol{m_{22}}$	m_{11}, m_{22}	8·666	2	0·013*
$\boldsymbol{V_1}$, $\boldsymbol{V_2}$, $\boldsymbol{m_{11}}$, $\boldsymbol{m_{22}}$	m_{11}, m_{22}	8·666	2	0·013*
F_1, F_2, m_{11}, m_{22}	$\boldsymbol{F_1}$, $\boldsymbol{F_2}$	2·648	2	0·161
F_1, F_2, V_1, V_2	$\boldsymbol{F_1}$, $\boldsymbol{F_2}$	2·648	2	0·161
F_1, F_2, m	$\boldsymbol{F_1}$, $\boldsymbol{F_2}$	2·394	1	0·122
F_1, F_2, V_1, V_2	$\boldsymbol{V_1}$, $\boldsymbol{V_2}$	0·326	2	0·850
V_1, V_2, m_{11}, m_{22}	$\boldsymbol{V_1}$, $\boldsymbol{V_2}$	0·326	2	0·850
V_1, V_2, m	$\boldsymbol{V_1}$, $\boldsymbol{V_2}$	0·000	1	1·000
$\boldsymbol{F_1}$, $\boldsymbol{F_2}$, $\boldsymbol{m}$	m	8·63	2	0·013*
$\boldsymbol{V_1}$, $\boldsymbol{V_2}$, $\boldsymbol{m}$	m	8·558	2	0·014*

In each pairwise comparison, the model in bold font is considered the better-fitting model and retained for further analysis. If a model is retained in one comparison, but rejected in another, it is removed from further consideration. For example, (F_1, F_2, m_{11}, m_{22}) is retained when compared with (m_{11}, m_{22}), but rejected when compared with (F_1, F_2); it is thus removed. This leaves four models – (f_{11}, v_{11}, m_{11}, m_{22}), (f_{22}, v_{22}, m_{11}, m_{22}), (F_1, F_2) and (V_1, V_2) – as candidates for the best-fitting model.
* Significant at the 0·05 level.

Finally, as shown in Table 4, the model with viability selection alone (V_1, V_2) has the least value for both *AIC* and *BIC*, and is thus considered the best-fitting model. The best-fitting estimates for the two parameters are $\hat{V}_1 = 1{\cdot}176$ and $\hat{V}_2 = 0{\cdot}861$, with the corresponding confidence intervals (1·039, 1·343) and

https://doi.org/10.1017/S0016672303006311 Published online by Cambridge University Press

Table 4. *Candidates for the best-fitting model for the* Arabidopsis *data, with their AIC and BIC indices, maximum log-likelihood values* (*L*), *and parameter estimates with their 95% confidence intervals* (*CI*)

	Fertility selection alone (F_1, F_2)	Viability selection alone (V_1, V_2)	Fertility and viability selection with A_1 dominant and meiotic drive (f_{22}, v_{22}, m_{11}, m_{22})	Fertility and viability selection with A_2 dominant and meiotic drive (f_{11}, v_{11}, m_{11}, m_{22})
AIC	821·1	818·779	822·452	822·452
BIC	829·418	827·097	839·088	839·088
L	−408·550	−407·389	−407·225	−407·225
Parameter 1	$\hat{F}_1=1{\cdot}288$	$\hat{V}_1=1{\cdot}176$	$\hat{f}_{22}=1{\cdot}926$	$\hat{f}_{11}=1{\cdot}254$
(CI)	(1·050, 1·608)	(1·039, 1·343)	(0·197, 5·688)	(0·000, 4·413)
Parameter 2	$\hat{F}_2=0{\cdot}827$	$\hat{V}_2=0{\cdot}861$	$\hat{v}_{22}=0{\cdot}387$	$\hat{v}_{11}=1{\cdot}068$
(CI)	(0·633, 1·075)	(0·744, 0·999)	(0·139, 3·855)	(0·294, >1000)
Parameter 3	N/A	N/A	$\hat{m}_{11}=0{\cdot}212$	$\hat{m}_{11}=0{\cdot}266$
(CI)			(0·109, 0·408)	(0·000, 0·523)
Parameter 4	N/A	N/A	$\hat{m}_{22}=0{\cdot}393$	$\hat{m}_{22}=0{\cdot}204$
(CI)			(0·067, 0·597)	(0·198, 0·248)

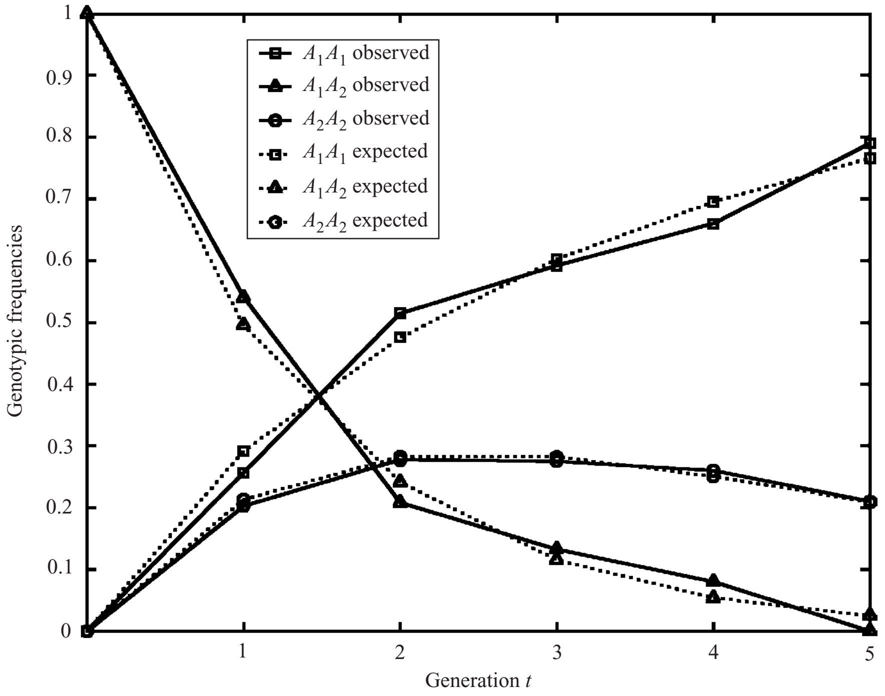

Fig. 6. Viability selection is the major evolutionary force acting on the *ACT2* gene in an experimental *Arabidopsis* population. The observed (continuous curves) and expected (dotted curves) genotypic frequencies are plotted as a function of generation (t) for five generations. The expected frequencies are calculated under the best-fitting model, viability selection alone, with $\hat{V}_1=1{\cdot}176$, $\hat{V}_2=0{\cdot}861$, $F_1=F_2=1$, $m_{11}=m_{22}=0{\cdot}25$ and $m_{12}=0{\cdot}5$.

(0·744, 0·999). This represents a case of directional selection against the *act2-1* allele, with heterozygous carriers (A_1A_2) being 85% (1/1·176) as viable, and homozygotes for the *act2-1* allele (A_2A_2) being 73% (0·861/1·176) as viable as wild-type (A_1A_1). The excellent fit between the observed and expected genotypic frequencies under this best-fitting model is shown in Fig. 6.

7. Discussion

We have developed a formal statistical procedure to estimate selection components for a diallelic autosomal locus in a diploid, selfing population by fitting time-dependent solutions for genotypic frequencies to observed multigenerational genotypic counts. The full model includes three selection components

https://doi.org/10.1017/S0016672303006311 Published online by Cambridge University Press

(fertility, viability and gametic selection), each of which has two independent parameters after the fertilities and viabilities are normalized with respect to those of heterozygotes. With adult census, we can estimate at most four of the six independent selection parameters simultaneously, because the dynamics of the genotypic frequencies are determined by four composite parameters (a_1, a_2, b_1 and b_2). Due to this limitation, a procedure was developed for finding the best-fitting model out of a hierarchy of 11 models with either fewer selection components, complete dominance, or multiplicative gametic selection with a single meiotic drive parameter. With seed census, the effects of fertility and viability selection are not distinguishable and thus can only be estimated jointly, as their product, along with meiotic drive. Estimation of all three selection components simultaneously can be achieved with a joint adult/seed census, which is thus to be used if at all possible.

Although our theory has so far assumed the data come from a series of consecutive generations, in fact this is not a necessity, as long as there are enough degrees of freedom for estimation. Each adult census or seed census contributes two degrees of freedom, corresponding to the two independent genotypic frequencies. With either adult or seed census alone, we need to sample at least two generations after the initial generation 0 to estimate up to four independent selection parameters, and one more census point to test the goodness-of-fit of the models and identify the best-fitting model. Here a census point means one generation of either adult or seed census. With joint adult and seed census, we need only census each life stage at least once, and at least three census points altogether, to have enough degrees of freedom to estimate all six independent selection parameters. We thus have enough degrees of freedom to estimate all three selection components in a highly selfing population, and test their goodness-of-fit, as long as there are at least four census points, with adults and seeds both sampled at least once.

We validated the applicability of our procedure with simulated data. The results also provide two other useful guidelines for the experimental design of such selection component estimation studies. First, with adult census, a larger sample size is needed when the population is subject to viability rather than fertility selection to achieve similar accuracy. Second, with joint adult/seed census, estimation accuracy is best improved by increasing the number of generations sampled rather than the number of individuals sampled per generation. More precise information on the optimal sampling scheme and the applicability of the procedure can be deduced by performing such simulations for the case at hand.

We illustrated our estimation procedure by applying it to multigenerational genotypic frequency data from the *act2-1* actin mutant and *ACT2* wild-type alleles segregating in an experimental *Arabidopsis* population. Our finding that viability selection alone is the best-fitting model is consistent with the previous *ad hoc* result (Asmussen *et al*., 1998) and suggests that the *ACT2* actin allele is preserved in laboratory-grown *Arabidopsis* populations over the *act2-1* allele, because it is under strong viability selection but not subject to significant selection via fertility or meiotic drive. This result is consistent with gene and protein expression data suggesting that *ACT2* is primarily expressed in vegetative tissues of *Arabidopsis*.

The possibility that some or all of the frequency change may be due to selection on linked genes in linkage disequilibrium with the *ACT2* locus is unlikely based on the following reasons: (1) The ecotype WS *Arabidopsis thaliana* that was used to create the *act2-1* mutant by T-DNA insertion was collected in 1943 and has gone through a large number of generations of inbreeding, and should be highly homozygous at any loci linked to *ACT2*. (2) After T-DNA insertion, measures were taken to ensure the T-DNA insertion on the *ACT2* locus was the only disruption in the selected plant (Gilliland *et al*., 1998). (3) Our multigenerational study was initiated with a single plant heterozygous at the *ACT2* locus, so that all linked loci should remain genetically homogeneous in subsequent generations. (4) A follow-up study shows that the root hair elongation defects of the *act2-1* mutant can be fully rescued by an *ACT2* genomic transgene, and impairment of root hair functions such as nutrient mining, water uptake and physical anchoring are the likely cause of the reduced fitness seen for *act2-1* mutants in this multigenerational study (Gilliland *et al*., 2002).

Our model and estimation procedure assume that the population is sufficiently large for genetic drift effect to be ignored. This assumption could be violated in empirical studies. One example is our *Arabidopsis* case study, where adult population size is limited to 100 although each adult produces a large number of progeny. In other cases, the seed population could also be small. To estimate selection components in small (e.g. <30) populations, a Monte Carlo simulation-based method similar to the one developed by Keightley *et al*. (1996) can be used.

In conclusion, monitoring changes in genotypic frequencies over multiple generations allows the formal dissection of the fertility, viability, and meiotic drive selection components present in highly selfing populations such as those of the model plant *Arabidopsis thaliana*. Dynamical data from multiple time points in this and other biological contexts represent a valuable and under-utilized tool for studying natural and experimental populations.

Appendix. Time-dependent solutions for seed genotypic frequencies

The recursions for seed frequencies are

$$s'_{11} = \frac{(f_{11}v_{11})s_{11} + (f_{12}m_{11}v_{12})s_{12}}{\bar{w}}, \tag{A1}$$

$$s'_{12} = \frac{(f_{12}m_{12}v_{12})s_{12}}{\bar{w}}, \tag{A2}$$

$$s'_{22} = \frac{(f_{22}v_{22})s_{22} + (f_{12}m_{22}v_{12})s_{12}}{\bar{w}}, \tag{A3}$$

where $\bar{w} = f_{11}v_{11}s_{11} + f_{12}v_{12}s_{12} + f_{22}v_{22}s_{22}$ is the normalization factor, corresponding to the net mean fitness in the population.

A direct adaptation of the procedure and technique used in Asmussen *et al.* (1998) for adults shows that the time-dependent solutions for the seed genotypic frequencies in each generation $t \geqslant 1$ are

$$s_{11}^{(t)} = \frac{y_1^{(t)}}{1 + y_1^{(t)} + y_2^{(t)}}, \tag{A4}$$

$$s_{12}^{(t)} = \frac{1}{1 + y_1^{(t)} + y_2^{(t)}}, \tag{A5}$$

$$s_{22}^{(t)} = \frac{y_2^{(t)}}{1 + y_1^{(t)} + y_2^{(t)}}, \tag{A6}$$

where for $i = 1, 2$, $y_i^{(t)} = \frac{s_{ii}^{(t)}}{s_{12}^{(t)}}$ is the ratio of the frequency of A_iA_i to A_1A_2 seeds in generation t, and

$$y_i^{(t)} = \begin{cases} (a_i)^t (y_i^{(0)} - y_i^*) + y_i^* & \textit{if } a_i \neq 1 \\ y_i^{(0)} + tc_i & \textit{if } a_i = 1, \end{cases} \tag{A7}$$

$$a_i = \frac{f_{ii}v_{ii}}{m_{12}f_{12}v_{12}}, \tag{A8}$$

$$c_i = \frac{m_{ii}}{m_{12}}, \tag{A9}$$

$$y_i^* = \frac{c_i}{1 - a_i} = \frac{m_{ii}f_{12}v_{12}}{m_{12}f_{12}v_{12} - f_{ii}v_{ii}}. \tag{A10}$$

We thank M. Goodisman, R. Cartwright and R. Mauricio for helpful discussions, and we thank W. G. Hill and two anonymous reviewers for many valuable comments, including suggesting our use of the profile likelihood method and tests with simulated data. This work was funded by grant support from the National Institutes of Health (NIH), GM-36397 (R.B.M.) and GM-48528 (M.A.A.).

References

Abbott, R. J. & Games, M. F. (1989). Population genetic structure and outcrossing rate of *Arabidopsis thaliana* (L.) Heynh. *Heredity* **62**, 411–418.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (ed. B. N. Petrov & F. Csaki), pp. 267–281. Budapest: Academiai Kiado.

An, Y. Q., McDowell, J. M., Huang, S., McKinney, E. C., Chambliss, S. & Meagher, R. B. (1996). Strong, constitutive expression of the *Arabidopsis* ACT2/ACT8 actin subclass in vegetative tissues. *The Plant Journal: for Cell and Molecular Biology* **10**, 107–121.

Anderson, W. W. (1969). Selection in experimental populations. I. Lethal genes. *Genetics* **62**, 653–672.

Arabidopsis Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815.

Asmussen, M. A., Gilliland, L. U. & Meagher, R. B. (1998). Detection of deleterious genotypes in multigenerational studies. II. Theoretical and experimental dynamics with selfing and selection. *Genetics* **149**, 727–737.

Bundgaard, J. & Christiansen, F. B. (1972). Dynamics of polymorphisms. I. Selection components in an experimental population of *Drosophila melanogaster*. *Genetics* **71**, 439–460.

Christiansen, F. B. & Frydenberg, O. (1973). Selection component analysis of natural polymorphisms using population samples including mother–offspring combinations. *Theoretical Population Biology* **4**, 425–445.

Clegg, M. T., Kahler, A. L. & Allard, R. W. (1978). Estimation of life cycle components of selection in an experimental plant population. *Genetics* **89**, 765–792.

DuMouchel, W. H. & Anderson, W. W. (1968). The analysis of selection in experimental populations. *Genetics* **58**, 435–449.

Forster, M. R. (2000). Key concepts in model selection: performance and generalizability. *Journal of Mathematical Psychology* **44**, 205–231.

Gilliland, L. U., Asmussen, M. A., McKinney, E. C. & Meagher, R. B. (1998). Detection of deleterious genotypes in multigenerational studies. I. Disruptions in individual *Arabidopsis* actin genes. *Genetics* **149**, 717–725.

Gilliland, L. U., Kandasamy, M. K., Pawloski, L. C. & Meagher, R. B. (2002). Both vegetative and reproductive actin isovariants complement the stunted root hair phenotype of the *Arabidopsis act2-1* mutation. *Plant Physiology* **130**, 2199–2209.

Goffe, W. L., Ferrier, G. D. & Rogers, J. (1994). Global optimization of statistical functions with simulated annealing. *Journal of Econometrics* **60**, 65–99.

Keightley, P. D., Hardge, T., May, L. & Bulfield, G. (1996). A genetic map of quantitative trait loci for body weight in the mouse. *Genetics* **142**, 227–235.

Kimura, M. (1991). Recent development of the neutral theory viewed from the Wrightian tradition of theoretical population genetics. *Proceedings of the National Academy of Sciences of the USA* **88**, 5969–5973.

Krakauer, D. C. & Nowak, M. A. (1999). Evolutionary preservation of redundant duplicated genes. *Seminars in Cell and Developmental Biology* **10**, 555–559.

Lowther, A. B. & Skalski, J. R. (1996). *Design and Analysis of Salmonid Tagging Studies in the Columbia Basin,* vol. VII*: Monte-Carlo Comparison of Confidence Interval Procedures for Estimating Survival in a Release–Recapture Study, with Application to Snake River Salmonids*. Technical report DOE/BP-02341-5.

Martienssen, R. & Irish, V. (1999). Copying out our ABCs: the role of gene redundancy in interpreting genetic hierarchies. *Trends in Genetics* **15**, 435–437.

McKinney, E. C., Ali, N., Traut, A., Feldmann, K. A., Belostotsky, D. A., McDowell, J. M. & Meagher, R. B. (1995). Sequence-based identification of T-DNA insertion mutations in *Arabidopsis*: actin mutants *act2-1* and *act4-1*. *The Plant Journal: for Cell and Molecular Biology* **8**, 613–622.

Meyerowitz, E. M. (1989). *Arabidopsis*, a useful weed. *Cell* **56**, 263–269.

Meyerowitz, E. M. (1994). Structure and organization of the *Arabidopsis thaliana* nuclear genome. In *Arabidopsis* (ed. E. M. Meyerowitz & C. R. Summerville), pp. 21–37. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Ohta, T. (1992). The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics* **23**, 263–286.

Prout, T. (1971). The relation between fitness components and population prediction in *Drosophila*. I. The estimation of fitness components. *Genetics* **68**, 127–149.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6**, 461–464.

Sokal, R. R. & Rohlf, F. J. (1995). *Biometry*, 3rd edn. San Francisco: W. H. Freeman.

Tautz, D. (2000). A genetic uncertainty problem. *Trends in Genetics* **16**, 475–477.

Wahrendorf, J., Becher, H. & Brown, C. C. (1987). Bootstrap comparison of non-nested generalized linear models: applications in survival analysis and epidemiology. *Applied Statistics* **36**, 72–81.

Weir, B. S. (1996). *Genetic Data Analysis II*. Sunderland, MA: Sinauer Associates.

https://doi.org/10.1017/S0016672303006311 Published online by Cambridge University Press