Theor Appl Genet (1985) 69:631–636

Estimating single gene effects on quantitative traits

2. Statistical properties of five experimental methods

D. G. Gilbert*

Department of Biophysics and Theoretical Biology, The University of Chicago, 920 East 58th Street, Chicago, Ill 60637, USA

Received July 30, 1984; Accepted August 5, 1984
Communicated by J. S. F. Barker

Summary. Experimental designs for measuring the effects of single loci on quantitative traits are compared for statistical properties. The designs tested are single population, combined strains, multiple strains, diallel of strains, and co-isogenic strains. Testing was done by simulating population genotypic and phenotypic arrays. Statistical properties measured are type I error, power, bias and efficiency. The relative ranking of designs is consistent for all properties and over eight conditions examined. The co-isogenic design is superior, followed closely by the single population method. The other three designs are similar in ability, with the diallel design somewhat superior. Based on its good statistical performance and wide feasibility, the single population method is recommended. The diallel method provides the most information on genetic components of variation.

Key words: Experimental design – Locus – Population genetics – Quantitative genetics – Statistics

Introduction

How can one determine if allelic variants at a given locus affect a given trait? One research paradigm developed to answer this question involves detailed and multilevel (genetic, biochemical, physiological, ecological) studies of a single locus (Clarke 1973; Koehn 1978). This paradigm has proved generally successful (Koehn 1983), but requires a large concentration of effort on single cases. There may be no better general method for answering the above question, due to the complex path from genes to traits. We may, however, find simpler methods suited to particular tasks. One of the major problems facing those who study locus effects is how certain are we that the locus of study, rather than other loci, is affecting the trait? This problem can be partly answered with a better knowledge of the statistical properties of the experimental design.

There are three general designs, with variants, in use now: combined strains, multiple strains, and co-isogenic strains. Two other designs are also considered: diallel of strains and single population. The breeding structure used in these designs is depicted in Table 1. The dissection, or partition, of trait variation that is possible with each design is also listed in Table 1.

The single population method (I) involves choosing individuals randomly from a population, and measuring them for trait value as well as locus genotype. Additive (*a*) and dominant (*d*) locus effects are then determined by comparing variation between classes of individuals with different locus types relative to variation within these classes. Examples related to this design are found in El-Kassaby (1982); Watt et al. (1983), and Zouros et al. (1980).

The combined strains method (II) involves selecting or producing several strains homozygous for a given allele at the locus of interest, then combining like homozygotes as a single strain with variable genetic background. Additive locus effects are determined by comparing trait responses of the different homozygous strains; dominant locus effects are determined with the F_1 hybrid of homozygous strains. These effects are measured relative to combined environmental (V_E) and background genetic (V_G) variation found within each strain. Examples of this method may be found in Aslund and Rasmuson (1976); Cavener and Clegg (1981); Fontdevila and Méndez (1979), and Van Delden et al. (1978).

The multiple strains (III) method is similar to II in that several strains homozygous for an allele are produced, but are not then combined. They may be highly inbred to eliminate within-strain background genetic variation. In this method, *a* effects are measured relative to additive genetic variation (V_A) found between the inbred strains of the same locus type; V_A is in turn measured relative to V_E within strains. In this design, *d* effects are not determinable. Examples of this method may be

* Current address: Biology Department, Indiana University, Bloomington, IN 47405, USA

Table 1. Diagrammatic representation of experimental designs. Letters designate alleles at a locus of a diploid individual, subscript numbers designate a genome type, and boxes enclose panmictic units. Variance components that can be partitioned are enclosed in parentheses. Genomic components (V): E, environmental; G, genetic; A, additive genetic; D, dominant genetic. Allelic components: a^2, additive locus; d^2, dominant locus

Design	Genomic diagram	Partitionable variance
I. Single population	: A_1A_2 A_3A_4 A_5B_6 : : A_7B_8 A_9B_{10} $B_{11}B_{12}$:	$(V_E+V_G+a^2)$, (V_E+V_G) & $(V_E+V_G+d^2)$, (V_E+V_G)
II. Combined strains	: A_1A_2 : : B_3B_4 : : A_5A_6 : × : B_7B_8 : ↓ : A_1B_4 : : A_6B_7 :	$(V_E+V_G+a^2)$, (V_E+V_G) & $(V_E+V_G+d^2)$, (V_E+V_G)
III. Multiple strains	: A_1A_1 : : A_2A_2 : : A_3A_3 : : B_4B_4 : : B_5B_5 : : B_6B_6 :	$(V_E+V_A+a^2)$, (V_E+V_A), (V_E)
IV. Diallel of strains	A_1 A_2 B_3 B_4 A_1 : 11 : 12 : 13 : 14 : A_2 : 21 : 22 : 23 : 24 : B_3 : 31 : 32 : 33 : 34 : B_4 : 41 : 42 : 43 : 44 :	$(V_E+V_A+a^2)$, (V_E+V_A), (V_E) & $(V_E+V_D+d^2)$, (V_E+V_D), (V_E)
V. Co-isogenic strains	: A_1A_1 : : B_1B_1 : : A_1A_1 : × : B_1B_1 : ↓ : A_1B_1 : : A_1B_1 :	(V_E+a^2), (V_E) & (V_E+d^2), (V_E)

found in McDonald et al. (1980); Starmer et al. (1977), and Tošić and Ayala (1981).

The diallel method (IV) is detailed in Gilbert (1985). Briefly, homozygous strains as per method III are crossed in all combinations, and the F_1 progeny are measured. Trait variation can then be partitioned into a, d, V_A, V_D and V_E components.

The co-isogenic method (V) involves producing strains identically homozygous, or nearly so, through the entire genome except at the locus of interest. Such co-isogenic strains retain genetic variation only at the one locus, thus within strain variation is V_E alone and between strain variation contains the additional a or d effects of the locus. This method is necessarily restricted to organisms where the genome can be manipulated. Examples of studies include Danford and Beardmore (1980) and Sheehan et al. (1979).

The use of these designs in hypothesis testing (if a locus affects a trait) and estimation (to what extent it affects a trait) is examined with computer-simulated populations. These populations consist of genomes with several polymorphic loci, and of phenotypes resulting from the combined a effects ($=V_A$) and d effects ($=V_D$) of the polymorphic loci, plus environmental effects. The basic model used for trait variation is

$$V_T = V_E + V_A + ka^2 + V_D + hd^2,$$

where T is total, E is environmental, A is additive genetic, D is dominant genetic, a^2 is additive locus, d^2 is dominant locus variation, and k, h are constants depending on locus allele frequencies. The effects of a given locus are then determined with each experimental design using the simulated populations. Table 2 shows the general flow of the simulation experiments and their relation to actual experiments. This basic trait model excludes factors of linkage disequilibria, epistasis and other interactions among loci and among traits and groups of loci.

Two hypothesis testing properties, type I error and power, and two estimation properties, bias and efficiency, are determined for the designs. Type I error is the probability of falsely rejecting the null hypothesis that the locus has no effect. Power is the probability of correctly rejecting the null hypothesis. Bias is a measure of the tendency of the sample estimate to deviate from the true value, and efficiency is a measure of the variance associated with the sample estimate of the locus effect.

Methods

The designs were tested by generating random population genotypic arrays, for each case, design and replicate population. Each genotypic array was changed to a phenotypic array

Table 2. Protocol for actual and simulated experiments

	Actual	Simulated
Population	Genotypes ↓ development Phenotypes	Genotypic array ↓ rules of transformation Phenotypic array
	↓	↓
Manipulation	Sampling, breeding and genetic manipulation according to design	Sampling according to design
	↓	↓
Observation	Measure traits of individuals	Accumulate sampled phenotypes
	↓	↓
Estimation	from observations and known manipulations	from observations and design model
		↓
		Compare estimate to actual value

Table 3. Population parameters for the cases studied

Case	No. of measured individuals	No. of unknown loci	Percent V_E of V_T	V_A per locus	
				Mean	s.d.
1	36	2	30	0	1
2	36	2	70	0	1
3	36	5	70	0	1
4	36	5	70	2	1
5	36	20	70	0	1
6	100	5	70	0	1
7	100	5	70	2	1
8	100	20	70	0	1

Constant for all cases: Test alpha: 0.050; V_D per locus: 0 mean, 1 s.d.; True additive effect (a): 0, 1, 2, 3, 4; True dominant effect (d): 0, a; No. of random populations: 10,000 per test of a effect; 5,000 per test of d effect

by rules of transformation detailed below. The phenotypic array was subject to an analysis of variance model appropriate to the design. Two basic results for each such test were calculated from the ANOVA: (1) the number of significant locus effects measured by F test in the ANOVA (the measured alpha), and (2) the sample estimated locus effect from the sums of squares. This simulation was performed with APL (Berry 1979) functions; these functions are available from the author on request.

The parameters which were varied for each case are number of individuals measured per test, number of 'unknown' loci affecting the trait, percentage of trait variance which is environmental in origin, and the mean and standard deviation of additive effects for the 'unknown' loci. The parameters held constant for all cases include the test alpha (the probability level which hypothesis tests where compared to), the dominant variance for 'unknown' loci, values of the true additive and dominant effects of the known locus, and the number of random populations examined. Genetic variation at the 'unknown' loci was parameterized by generating vectors containing values of a and d effects drawn from a random normal distribution with the specified means and standard deviations. Parameter values are listed in Table 3.

The designs were tested on an equal cost basis, using the assumption that the cost of a design is the total number of individuals measured (N). This assumption is not necessarily true in practice (see "Discussion"). On this equal cost basis, designs employing multiple strains (II, III, and IV) were based on S strains = $\sqrt{(N)}$ with $\sqrt{(N)}$ individuals per strain, so that the total N measured remained constant for all designs.

For each design, four index vectors of length N were generated, which indicate how alleles of the population genotypic array were to be selected. Indices I and J are the values [0...1] of the allele at the known locus, contributed by maternal and paternal genomes. Indices K and L are the identification numbers of the maternal and paternal genomes found in the following manner:

Design I. K and L are N gametes each chosen randomly, without replacement, from [1...N]. In other words, each haploid gamete is a random selection from the population genotypic array.

Design II. S genomes are randomly chosen from [1...N]. These S are split into two halves, half representing the strains with the 0 allele at the known locus, the other half the strains with the 1 allele at the known locus. The K and L maternal and paternal genome indices are then generated so that one-quarter of N represent individuals from a combined strain of 0/0 homozygotes at the known locus, one-quarter represent a combined strain of 1/1 homozygotes, and the remaining half represent the two reciprocal heterozygotes from the cross of 0/0 and 1/1 strains.

Design III. K and L are, identically, S gametes chosen randomly from [1...N]. In other words, each diploid gamete is homozygous, selected from a subset of S strains from the population genotypic array.

Design IV. S genomes are randomly chosen from [1...N]. These genomes are then permuted so that K and L represent the F_1 result of all combinations of matings between S maternal and paternal gametes.

Design V. One genome is randomly chosen from [1...N], and replicated to make N individuals with identical K and L gametes.

For each design, the appropriate analysis of variance model is selected. The model for each design is expressed as

combinations of six ANOVA contrasts: mean, additive locus, additive strain, dominant locus, dominant strain and total variance.

Testing of the *a* and *d* effects is done in the following manner:

(1) A population genotypic array (G) is generated as a random M×T×N array of [0...1], for M loci, T replicate trials, and N individuals. The population gene frequency per locus is randomly taken from (0.50...1.00).

(2) The genotypic array is transformed to a T×N phenotypic array (Y) according to additive and dominant effects of the 'known' (*a* and *d*) and 'unknown' loci (A and D), the locus (I, J) and gametic (K, L) indices, and random environmental variance (E). This transform is stated as:

$$Y = A \times P \leftarrow (G[K] \text{ and } G[L]) - A \times Q \leftarrow (G[K] \text{ nor } G[L])$$
$$+ D \times H \leftarrow (P \text{ nor } Q)$$
$$+ a \times p \leftarrow (I \text{ and } J) - a \times q \leftarrow (I \text{ nor } J)$$
$$+ d \times h \leftarrow (p \text{ nor } q)$$
$$+ \text{Random Normal } (0, \sigma_E).$$

(3) Sums of squares specified by contrasts for each design are calculated from the phenotypic array for the additive and dominant locus ANOVA models.

(4) The significance level for *a* and *d* effects from this ANOVA is compared with the test alpha. The measured alpha for each design and each value of *a* and *d* is computed as the proportion of significance levels at or below the test alpha, out of all the replicated populations tested. Type I error is found as the measured alpha when the true *a* or *d* value is 0.0. Power is calculated from the measured alpha when true *a* or *d* values are not equal to 0.0. The sample estimates of *a* and *d* are calculated from the Anova sums of squares, and their mean and variance are accumulated over the replicated population tests. Bias is determined from the average of sample estimates of *a* or *d*. Efficiency is found from the variance of sample estimates.

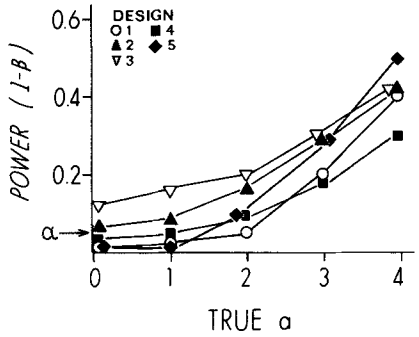

Fig. 1. Statistical power for detecting true additive locus effects (*a*), case 4. Designs: *1* single population; *2* combined strains; *3* multiple strains; *4* diallel; *5* co-isogenic

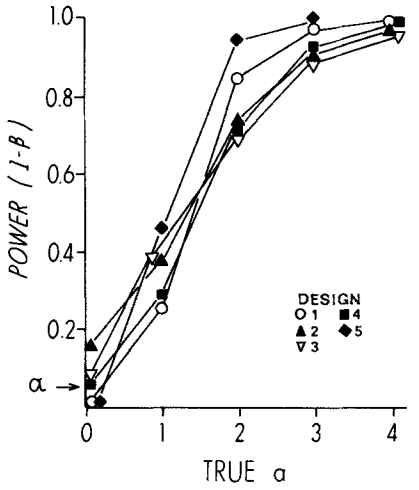

Fig. 2. Statistical power for detecting true additive locus effects (*a*), case 6. Designs: *1* single population; *2* combined strains; *3* multiple strains; *4* diallel; *5* co-isogenic

Results and discussion

The parameters which were varied between cases (Table 3) have, as expected, strong effects on the overall ability of the designs to detect locus effects: decreasing the number of measured individuals, increasing the number of unknown loci involved, increasing the average additive effects of unknown loci, and increasing the percentage of environmental variance all decrease the power and efficiency of the designs. Type I error and bias are not noticeably affected by these changes. Figures 1 and 2 illustrate the type I error and power curves for cases 4 and 6. There are only minor changes in the relative ranking of designs between the cases, so that we can summarize the design abilities as the average over the eight cases. These results are given in Table 4, expressed relative to the best design (V). Absolute average properties for design V are also given.

Type I error was measured as the proportion of significance levels less than or equal to the test alpha of 0.05, for true effects of 0.0. Design V is conservative for *a* effects; the measured alpha is 10 times less than the test alpha. All comparisons with design V show a higher error rate, with the extreme in design II. Power, expressed as logit (1/beta), declines relative to design V to 36% for design III. Bias, expressed as the maximum deviation of average sample estimates from the true values of *a* or *d*, increases to 400% of design V in design II. Efficiency, the variance of sample estimates, increases to about 388% of design V in design II.

These comparisons permit a qualitative ranking of the designs. The co-isogenic design (V) is best for all statistical properties, and the single population method (I) is second best for all properties. The other three are fairly similar, but the diallel of strains design (IV) appears somewhat better than the other two.

Applications of these results to experimental work depends on the importance of the various statistical properties as well as other considerations. These results may, with caution, be used in selecting appropriate designs. Existing studies (see "Introduction") for detecting single locus effects have been primarily aimed at hypo-

Table 4. Ranking of experimental designs for estimating single gene effects on quantitative traits. Average value of properties from 8 cases, expressed as percent of design V (absolute average of design V in parentheses). θ: a, additive locus effect; d, dominant locus effect. Type I error: measured alpha. Power: logit $(1/\beta)$. Bias: max $(\hat{\mu} - \theta)$. Efficiency: $\hat{\sigma}^2$

Property	θ	Design				
		I Single popln.	II Combined strains	III Multiple strains	IV Diallel strains	V Co-isogenic
Type I error	a	150	2,400	2,175	1,275	100 (0.004)
	d	100	175	–	78	100 (0.055)
Power	a	70	41	36	42	100 (2.874)
	d	75	66	–	72	100 (1.566)
Bias	a	175	400	300	360	100 (0.123)
	d	140	140	–	140	100 (0.266)
Efficiency	a	132	388	213	247	100 (0.0864)
	d	160	607	–	356	100 (0.0671)
Estimable variance component		a, d, V_E	a, d, V_E	a, V_A, V_E	a, d, V_A, V_D, V_E	a, d, V_E
Feasibility		wide	medium	medium	medium	limited

thesis testing, with type I error being the major concern. On this basis, designs I and V are notably superior to the others, and IV is intermediate. The power characteristics can provide information on sample sizes necessary for detecting locus effects for various genetic architectures. For example (Fig. 2), with a sample of 100 individuals, designs I and V have a 90% or better chance of detecting a locus additive effect that is 2 s.d. greater than the average background effect, compared to about 70% for designs II, III and IV.

Designs may also be chosen for what they can measure: all measure *a* effects, but design III does not measure *d* effects. Only designs III and IV permit estimation of the background genetic variance independent of environmental variance. This may be important additional information to use in interpreting the statistical significance or non-significance of a locus effect (see, for example, the discussion in Gilbert 1985).

Before these designs can be considered on the above bases, a researcher must determine what is feasible with the organisms of interest. The experiments referenced for design I all involve species with no backlog of genetic manipulation (butterflies, oysters, Douglas fir). It is noteworthy that this design, which involves no manipulation of genomes but relies on randomization by natural segregation and assortment, acts nearly as well as the genetically precise design V. Designs II, III and IV can be applied to any species which can be artificially bred; their feasibility varies with feasibility of breeding. Design V is limited to a few species, beyond those which clone naturally, where stocks of highly inbred strains or special genotypes exist.

Another consideration in comparing utility of the designs is cost. A common protocol with *Drosophila* is to produce strains from single (gravid) females collected from nature, combine the strains homozygous for the same allele at a locus, then use these combined strains in several design II experiments. With such a protocol, locus types need be determined only once and groups of individuals can be measured per strain, substantially reducing the cost of experiments per tested individual. The power of designs II, III and IV is proportional to the number S of strains used. With this consideration, an initial investment to collect many strains can make these designs comparable to I and V.

Acknowledgements. I am indebted to W. T. Starmer for providing the environment for this work, and to NSF DEB-81-08772 grant to W. T. Starmer. Rollin Richmond, Dan Sulzbach, Tom Starmer, and anonymous reviewers aided substantially in clarifying my ideas on this subject. Indiana University and Syracuse University provided computing facilities.

References

Aslund SE, Rasmuson M (1976) Mating behavior as a fitness component in maintaining allozyme polymorphisms in *Drosophila melanogaster*. Hereditas 82:175–178

Berry P (1979) Sharp APL reference manual. IP Sharp, Toronto

Cavener DR, Clegg MT (1981) Evidence for biochemical and physiological differences between enzyme genotypes of *Drosophila melanogaster*. Proc Natl Acad Sci USA 78: 4444–4447

Clarke B (1973) The contribution of ecological genetics to evolutionary theory: detecting the direct effects of natural selection on particular polymophic loci. Genetics 79: 101–113

Danford ND, Beardmore JA (1980) Selection at the esterase-6 locus in *Drosophila melanogster* by added enzyme substrates in the culture medium. Genetica 50:171–178

El-Kassaby YA (1982) Associations between allozyme genotypes and quantitative traits in Douglas-fir [*Pseudotsuga menziesii* (Mirb.) Franco]. Genetics 101:103–115

Fontdevila A, Méndez J (1979) Frequency dependent mating in a modified allozyme locus of *Drosophila pseudoobscura.* Evolution 33:634–640

Gilbert DG (1985) Estimating single gene effects on quantitative traits. 1. A diallel method applied to *Est 6* in *D. melanogaster.* Theor Appl Genet 69:625–629

Koehn RK (1978) Physiology and biochemistry of enzyme variation. The interface of ecology and population genetics. In: Brussard PF (ed) Ecological genetics: the interface. Springer, Berlin Heidelberg New York, pp 51–72

Koehn RK, Zera AJ, Hall JG (1983) Enzyme polymorphism and natural selection. In: Nei M, Koehn RK (eds) Evolution of genes and proteins. Sinauer Sunderland Mass, pp 115–136

McDonald JF, Anderson SM, Santos M (1980) Biochemical differences between the products of the *Adh* locus in *Drosophila.* Genetics 95:1013–1022

Sheehan KB, Richmond RC, Cochrane BJ (1979) Studies of esterase 6 in *Drosophila melanogaster.* 3. The developmental pattern and tissue distribution. Insect Biochem 9: 443–450

Starmer WT, Heed WB, Rockwood-Sluss ES (1977) Extension of longevity in *Drosophila mojavensis* by environmental ethanol: differences between subraces. Proc Natl Acad Sci USA 74:387–391

Tošić M, Ayala FJ (1981) Density- and frequency- dependent selection at the *Mdh-2* locus in *Drosophila pseudoobscura.* Genetics 97:679–701

Van Delden W, Boerema AC, Kamping A (1978) The alcohol dehydrogenase polymorphism in populations of *Drosophila melanogaster.* 1. Selection in different environments. Genetics 90:161–191

Watt WB, Cassin RC, Swan MS (1983) Adaptation at specific loci. 3. Field behavior and survivorship differences among Colias Pgi genotypes are predictable from in vitro biochemistry. Genetics 103:725–739

Zouros E, Singh SM, Miles HE (1980) Growth rate in oysters: an overdominant phenotype and its possible explanations. Evolution 34:856–867