A simple note on how to save money in linkage analysis

Manfred Huehn · Hans-Peter Piepho

Abstract Genetic linkage maps are often based on maximum-likelihood estimates of recombination fractions which are converted into map units by mapping functions. This paper presents a cost analysis of linkage analysis for a segregating F_2 population with codominant or dominant molecular markers and a qualitative monogenic dominant–recessive trait. For illustration, a disease-resistance trait is considered, where the susceptible allele is recessive. Three sub-populations of the F_2 can be used for linkage analysis [susceptible (= recessive) individuals, resistant (= dominant) individuals, complete F_2]. While it is well-known that recessive individuals are more informative than dominant individuals, it is not obvious *a priori*, which of the three sub-populations should be preferred, when costs of phenotyping and genotyping are taken into consideration. A comparative economic analysis of alternative procedures of linkage detection based on these three sub-populations does exhibit a clear economic superiority of the sub-population of susceptible (= recessive) individuals, when costs of genotyping are high. This cost-effectiveness is due to the higher information content of this sub-population compared to the sub-population of dominant (= resistant) individuals and also compared to the complete F_2. Our final conclusion/recommendation is as follows: If the cost to genotype an individual is sufficiently large compared with the cost to phenotype an individual, then linkage analysis and genetic mapping should be only based on susceptible (= recessive) individuals. Conversely, if the cost of phenotyping exceeds that for genotyping, it may be preferable to genotype all plants. The exact conditions under which a strategy is preferable are described in the paper.

Keywords Cost analysis · Economics · Genetic mapping · Linkage analysis · Molecular marker · Recombination fraction

M. Huehn
Institute of Crop Science and Plant Breeding, University Kiel, Am Botanischen Garten 1-9, D-24118 Kiel, Germany

H.-P. Piepho (✉)
Institute for Crop Production and Grassland Research, University of Hohenheim, Fruwirthstrasse 23, D-70599 Stuttgart, Germany
e-mail: piepho@uni-hohenheim.de

Introduction

Recent advances in some fields (molecular biology, genomics, statistical genetics, linkage analysis, mapping techniques) provide novel research tools with numerous potential applications to practical plant breeding. DNA-based molecular marker technologies, for example, can be integrated into conventional breeding schemes to

replace certain steps of commonly used routine breeding methods which may be time-consuming, expensive, or both. One possible outcome may be an increase of efficiency of selection strategies for plant genetic improvement. Molecular marker-assisted methods, however, will not always be cost-effective. For certain applications, phenotypic screening may be easy, fast, and inexpensive, especially when conducted under optimal field conditions. For other applications, however, the contrary is true and phenotypic screening is difficult, time-consuming, and/or expensive. For such cases, molecular marker-assisted methods may provide significant advantages. An evident conclusion, therefore, is that the superiority of these molecular marker-assisted methods depends on the cost of genotypic relative to phenotypic screening. This cost-ratio, of course, varies among applications. In spite of the potential benefits of molecular marker-assisted selection strategies these methods have to be approved with regard to economic considerations. The potential benefits of marker-assisted methods may be quite different: improvement of selection efficiency, time savings, benefits associated with an accelerated release of improved plant breeding outputs, cost effectiveness and general superiority with regard to economics.

Little information is available on the costs of phenotypic and genotypic selection procedures. Only a few papers have addressed this topic (Ragot and Hoisington 1993; Moreau et al. 2000; Yu et al. 2000; Hoisington and Melchinger 2004; Kuchel et al. 2005). A comprehensive general discussion on cost relationships for procedures in maize breeding associated with conventional and molecular marker-assisted breeding methods has been recently published by Dreher et al. (2003) and Morris et al. (2003). But, empirical data on these economic relationships for other agricultural crops are rather rare. Therefore, it is difficult to make meaningful comparisons between the efficiency of marker-assisted relative to conventional plant breeding methodology without actual cost data. A detailed cost analysis of the field and laboratory procedures associated with conventional and molecular marker-assisted breeding strategies is an indispensable prerequisite for any reliable recommendation. The overall cost-effectiveness of

molecular marker-assisted methods will depend on four major parameters: (1) relative cost of phenotypic versus genotypic screening, (2) time savings achieved by the method, (3) benefits associated with an accelerated release of improved cultivars, and (4) amount of operating capital available to the breeding program.

There is no single, nor simple, general answer as to when to apply molecular marker-assisted methods versus conventional methods in a breeding program. For an application of molecular marker-assisted methods in practical plant breeding linkage maps with a reasonable genome coverage are needed. Linkage maps of crop species are often constructed with segregating populations, for example F_2 populations or backcrosses. In this paper, some comments on a comparative economic analysis of alternative procedures for molecular marker-assisted linkage detection are presented. This study is restricted to an investigation of a qualitative monogenic dominant–recessive trait based on a segregating F_2 population. For illustration, a disease-resistance trait is considered, where the susceptible allele is recessive. This article presents selected results and conclusions from a theoretical study carried out to compare the cost-effectiveness of three approaches (based on different sub-populations of a segregating F_2) to linkage detection. The linkage to be determined is between a molecular marker (codominant or dominant) and a monogenic dominant disease-resistance trait. Analysis of two-point linkage is carried out by the traditional procedure of maximum likelihood estimation of recombination fractions. These estimates are converted into map units (by mapping functions). The present study assumes a set of molecular markers already to be available and mapped and known to be polymorphic for the F_2 population under study. It is further assumed that relative cost of phenotyping and genotyping can be quantified.

Problem and theory

Theoretical preliminaries

Assume a diploid segregating F_2 population co-segregating for a molecular marker (codominant or dominant) and a gene coding for a qualitative

trait as, for example, a disease resistance gene. The two alleles at the resistance gene locus are denoted by A (= resistant) and a (= susceptible) with A dominant over a. The marker alleles are B_1 and B_2 with a recombination value R between the marker and the disease resistance gene locus. An initial cross of homozygous parents AAB_1B_1 and aaB_2B_2 produces a double heterozygote F_1 generation AaB_1B_2 (with known genetic associations AB_1 and aB_2). Selfing or intercrossing the F_1 generation AaB_1B_2 creates a segregating F_2 population. Linkage analysis and genetic mapping is based on this F_2. Three sub-populations of the F_2 can be used for linkage analysis:

  I. Linkage analysis based on recessive (= susceptible) individuals.
 II. Linkage analysis based on dominant (= resistant) individuals.
III. Linkage analysis based on the complete F_2.

The recombination value is assumed to be equal in both sexes. The double heterozygote AaB_1B_2 produces the gametes AB_1, aB_1, AB_2, and aB_2 with frequencies $\frac{1}{2}(1-R), \frac{1}{2}R, \frac{1}{2}R$, and $\frac{1}{2}(1-R)$, respectively. The frequencies f_i of genotypes occurring in the segregating F_2 is given in Table 1 of Huehn and Piepho (2003). The costs associated with the use of sub-populations I, II, and III, respectively, may be quite different. The main objective of this paper is to compare the cost-effectiveness of approaches I–III. We denote:

$N_T^{(i)}$= total number of tested F_2-individuals (for population i)

Table 1 Lower bounds for u so that linkage analysis based on recessive (= susceptible) individuals is superior to linkage analysis based on the complete F_2

R	codominant marker	dominant marker
0.00	2.00	2.00
0.05	1.77	1.81
0.10	1.54	1.62
0.15	1.31	1.45
0.20	1.11	1.28
0.25	0.93	1.13
0.30	0.78	0.98
0.35	0.66	0.85
0.40	0.57	0.72
0.45	0.52	0.61
0.50	0.50	0.50

$N_R^{(i)}$ = number of dominant (= resistant) individuals (for population i)
$N_S^{(i)}$ = number of recessive (= susceptible) individuals (for population i)
$N_T^{(i)} = N_R^{(i)} + N_S^{(i)}$
a= cost per plant for testing resistance (= phenotyping cost)
b = cost per plant for marker determination (= genotyping cost)
$\hat{R}$ = maximum likelihood estimate of the recombination fraction R
$V(\hat{R})$ = variance of $\hat{R}$, which depends on R and on the sample size (number of individuals used)

For sufficiently large sample sizes, these variances $V(\hat{R})$ are well-known from general likelihood theory (Edwards 1972). They can be approximated by

$$V(\hat{R}) = \frac{1}{I(R) \cdot N}, \tag{1}$$

where $I(R)$ has been defined as the expected information content per observation and N is the sample size. One may compute the expected information $I(R)$ by the formula

$$I(R) = \sum_{i=1}^{s} \frac{1}{f_i}\left(\frac{df_i}{dR}\right)^2, \tag{2}$$

where f_i is the expected proportion in the ith of s classes and R is the theoretical (true) value of the parameter to be estimated. $I(R)$ depends on the sub-population (I, II, or III, respectively) used for linkage analysis and on the mode of inheritance at the marker locus (codominant or dominant expression of marker alleles).

Based on Eq. (2) and the expected proportion f_i in the different marker classes as reported in Huehn and Piepho (2003; Table 1) one obtains the following formulae for expected information content.

Codominant markers

$$[I(R)]_S = \frac{2}{R(1-R)} \tag{3}$$

for the sub-population of recessive (= susceptible) individuals,

$$[I(R)]_R = \frac{\frac{4}{3}}{1-R^2} + \frac{\frac{4}{3}}{2R-R^2} - \frac{2}{1-R+R^2} \qquad (4)$$

for the sub-population of dominant (= resistant) individuals, and

$$[I(R)]_T = \frac{1}{1-R^2} + \frac{1}{2R-R^2} + \frac{1}{2R-2R^2} - \frac{3}{2-2R+2R^2} \qquad (5)$$

for the complete F_2. These formulae have been given in Huehn (1995) and Huehn and Piepho (2003).

Dominant markers

$$[I(R)]_S = \frac{4}{2R-R^2} \qquad (6)$$

for the sub-population of recessive (= susceptible) individuals,

$$[I(R)]_R = \frac{\frac{4}{3}}{2R-R^2} - \frac{\frac{8}{3}}{R^2-2R+3} \qquad (7)$$

for the sub-population of dominant (= resistant) individuals, and

$$[I(R)]_T = \frac{2}{2R-R^2} - \frac{2}{R^2-2R+3} \qquad (8)$$

for the complete F_2.

Cost analysis

After the theoretical preliminaries, we proceed to the main objective of this paper: comparison of cost-effectiveness between approaches I, II, and III, respectively. While recessive individuals are more informative than dominant individuals, it does not immediately follow that approach I (use recessive individuals only) is generally the best strategy, when costs of phenotyping and genotyping are taken into consideration. Hence, a detailed cost analysis is in order.

The total costs for the three approaches can be calculated as follows:

$$C_S = N_T^{(i)} a + N_S^{(i)} b \qquad (9)$$

with C_S = total cost of linkage analysis for approach I (utilization of recessive (= susceptible) individuals) based on data of population i.

$$C_R = N_T^{(j)} a + N_R^{(j)} b \qquad (10)$$

with C_R = total cost of linkage analysis for approach II (utilization of dominant (= resistant) individuals) based on data of population j.

$$C_T = N_T^{(k)} a + N_T^{(k)} b \qquad (11)$$

with C_T = total cost of linkage analysis for approach III (utilization of recessive *and* dominant individuals) based on data of population k.

At first, we compare the costs associated with approaches I and III, i.e. linkage analysis based on recessive (= susceptible) individuals versus linkage analysis based on the complete F_2. This comparison can be expressed by the cost ratio

$$\text{cost ratio} = \frac{C_S}{C_T} = \frac{N_T^{(i)} a + N_S^{(i)} b}{N_T^{(k)} a + N_T^{(k)} b}. \qquad (12)$$

By using the expected segregation ratio $N_S^{(i)} : N_T^{(i)} = 1:4$, Eq. (12) can be rewritten as

$$\text{cost ratio} = \frac{u+4}{u+1} \cdot \frac{N_S^{(i)}}{N_T^{(k)}} \qquad (13)$$

with $u = b/a$.

To ensure an objective comparison between approaches I and III, we require that both approaches should provide results on linkage analysis with equal accuracy, i.e. equal variances of the estimates of recombination fractions. By Eq. (1), that means

$$[I(R)]_T \cdot N_T^{(k)} = [I(R)]_S \cdot N_S^{(i)}. \qquad (14)$$

Combination of Eq. (13) with Eq. (14) leads to

$$\text{cost ratio} = \frac{u+4}{u+1} \cdot \frac{[I(R)]_T}{[I(R)]_S}. \qquad (15)$$

The cost ratio depends (i) on the recombination fraction R and (ii) on the parameter u (genotyp-

ing cost relative to phenotyping cost). By simple algebraic manipulations, Eq. (15) can be expressed as

$$\text{cost ratio} = \frac{u+4}{u+1}\left(1 - \frac{1.5}{R^2 - 2R + 3}\right) \quad (16)$$

for dominant markers and

$$\text{cost ratio} = \frac{u+4}{u+1}\left(2 - \frac{0.5}{1+R} - \frac{0.5}{2-R} - \frac{0.75}{1-R+R^2}\right) \quad (17)$$

for codominant markers.

Next, we compare the costs associated with approaches II and III, i.e. linkage analysis based on dominant (= resistant) individuals versus linkage analysis based on the complete F_2. This comparison can be expressed by the cost ratio

$$\text{cost ratio} = \frac{C_R}{C_T} = \frac{N_T^{(j)}a + N_R^{(j)}b}{N_T^{(k)}a + N_T^{(k)}b}. \quad (18)$$

By using the segregation ratio $N_R^{(j)}:N_T^{(j)} = 3:4$, Eq. (18) can be rewritten as

$$\text{cost ratio} = \frac{u + \frac{4}{3}}{u+1} \cdot \frac{N_R^{(j)}}{N_T^{(k)}}. \quad (19)$$

The requirement of equal variances for $\hat{R}$ leads to

$$[I(R)]_R \cdot N_R^{(j)} = [I(R)]_T \cdot N_T^{(k)}. \quad (20)$$

Combination of Eq. (19) with Eq. (20) gives

$$\text{cost ratio} = \frac{u + \frac{4}{3}}{u+1} \cdot \frac{[I(R)]_T}{[I(R)]_R}. \quad (21)$$

If we insert in Eq. (21) the explicit formulae for expected information content one obtains

$$\text{cost ratio} = \frac{u + \frac{4}{3}}{u+1}\left(1 + \frac{0.5}{1 - 2R + R^2}\right) \quad (22)$$

for dominant markers and

$$\text{cost ratio} = \frac{1 + \frac{3}{4}u}{1+u}\left(1 + \frac{1}{7 - \frac{2}{1+R} - \frac{2}{2-R} - \frac{3}{1-R+R^2}}\right) \quad (23)$$

for codominant markers.

For each linkage analysis, the parameters R and u must be known to determine the cost-effectiveness of linkage analysis based on recessive (or dominant) individuals relative to the most commonly practiced linkage analysis based on the complete F_2.

In this paper, the cost ratios were calculated numerically for $0 \le R \le 0.50$ and a wide range of selected u-values.

Results

Comparison: approach I–approach III

For the comparison of the costs associated with approaches I and III (recessive individuals versus complete F_2) one obtains the following main results and conclusions (Figs. 1 and 2):

1. The cost ratio decreases with increasing u-values (for any given fixed recombination fraction R) (for dominant markers as well as for codominant markers). That means: The cost for approach I (linkage analysis based on recessive individuals) decreases with increasing genotyping cost relative to phenotyping cost (per plant) (Figs. 1 and 2).

2. The cost ratio decreases with increasing recombination fractions R (for any given fixed u-value) (for dominant markers as well as for codominant markers). The cost for approach I (linkage analysis based on recessive individuals) decreases with decreasing strength of linkage (Figs. 1 and 2).

3. For both marker classes (codominant and dominant) one obtains: The cost ratio exhibits a relatively weak dependence on the recombination fraction (for each fixed u), i.e. this ratio only slightly depends on R. But this ratio most strongly depends on the u-value (for each fixed R) (Figs. 1 and 2), i.e., on the ratio of genotyping cost relative to phenotyping cost. This ratio, however, may be quite different for different applications.

4. For each pair of parameter values R and u with $0 < R < 0.50$ one obtains: The cost ratio is smaller for codominant markers than for dominant markers. Consequently, if the
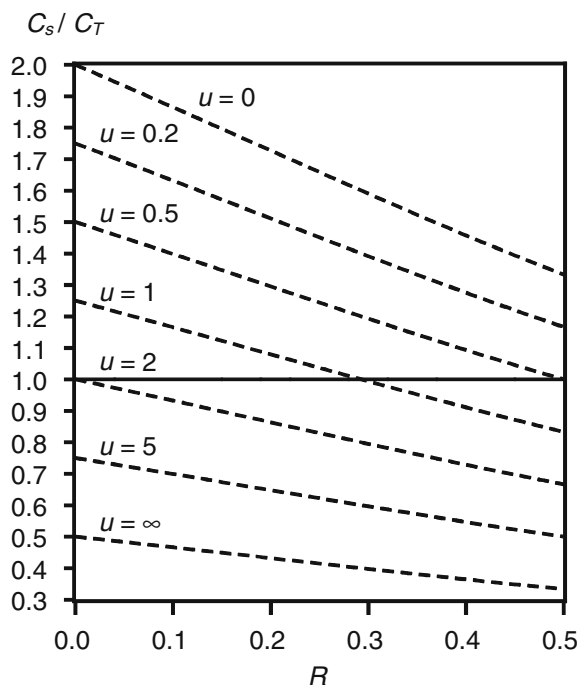
C_s/C_T

C_s/C_T

Fig. 1 Cost ratios C_S/C_T [= cost of linkage analysis based on recessive (= susceptible) individuals relative to the cost of linkage analysis based on the complete F_2] with dominant markers for different recombination fractions (R) and some selected ratios of genotyping to phenotyping costs (u)
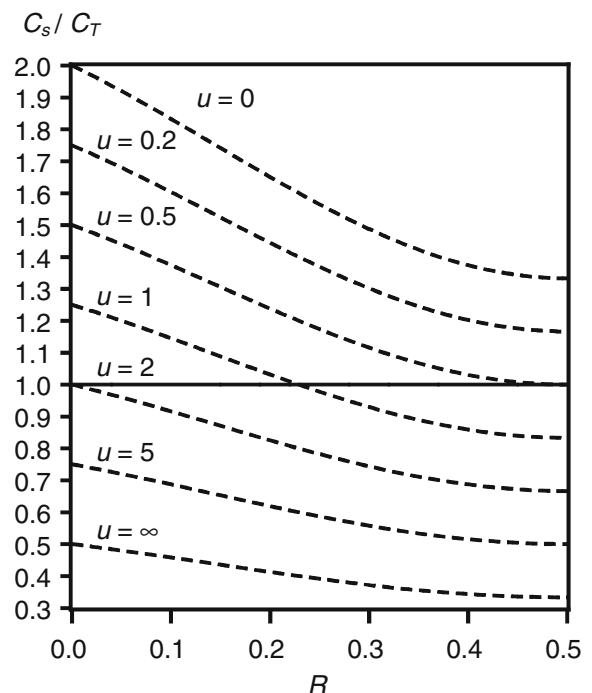
Fig. 2 Cost ratios C_S/C_T [= cost of linkage analysis based on recessive (= susceptible) individuals relative to the cost of linkage analysis based on the complete F_2] with codominant markers for different recombination fractions (R) and some selected ratios of genotyping to phenotyping costs (u)

genotyping costs for dominant and codominant markers are equal, then linkage analysis based on codominant markers is less expensive than linkage analysis based on dominant markers (Figs. 1 and 2).

5. For $0 < R < 0.50$ one obtains: $0.33 <$ cost ratio < 2. Thus, the total cost of linkage analysis for approach I is always smaller than twice the total cost of linkage analysis for approach III and it is always larger than one third of the total cost for approach III (for both marker classes: codominant and dominant) (Figs. 1 and 2).

6. The differences of C_S (expressed in percent of C_T, i.e. $C_T \cong 100\%$) between 'dominant markers' and 'codominant markers' are presented in Fig. 3. The numerical values of these differences are between 0% and 10.34%. This maximum is attained for $R=0.316$ and $u=0$. The differences of C_S between 'dominant markers' and 'codominant markers' decrease with increasing

u-values (for each recombination fraction R with $0 < R < 0.50$). This implies that the difference of linkage analysis costs between both marker classes decreases with increasing genotyping cost relative to phenotyping cost. For any given fixed u-value, these differences between both marker classes exhibit maxima which are located at intermediate recombination fractions (Fig. 3).

7. In the field of practical applications, the most interesting question is: Are there situations where linkage analysis only based on recessive (= susceptible) individuals is superior (i.e. less expensive) to the commonly practiced linkage analysis based on the complete F_2? This superiority is equivalent to "cost ratio < 1". By Eqs. (16) and (17) one obtains the conditions
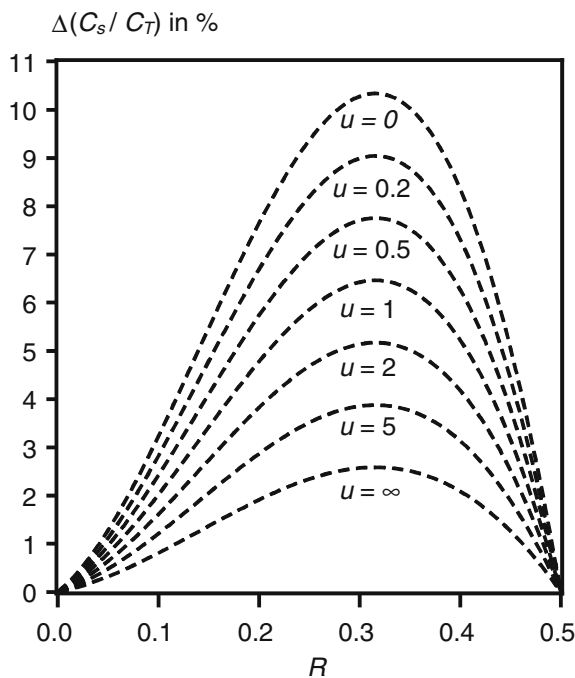
$$u > 2(1 - R)^2 \tag{24}$$

for dominant markers and

$\Delta(C_s / C_T)$ in %

Fig. 3 Differences of C_S (ΔC_S; expressed in percent of C_T, i.e. $C_T \cong 100\%$) between 'dominant markers' and 'codominant markers' for different recombination fractions (R) and some selected ratios of genotyping to phenotyping costs (u)

$$u > \frac{12}{\frac{2}{1+R} + \frac{2}{2-R} + \frac{3}{1-R+R^2} - 4} - 4 \qquad (25)$$

for codominant markers.

Some numerical values for this bound are presented in Table 1. A necessary condition for a superiority of approach I compared to approach III is $u > 0.50$ or $a < 2b$. In other words, the cost (per plant) for testing resistance (= phenotyping cost) must be smaller than twice the cost (per plant) for marker determination (= genotyping cost). For recombination fractions R with $0.30 \leq R \leq 0.50$, i.e. weak linkage, the bounds for u are between 0.8 and 0.5 (for codominant markers) and between 1 and 0.5 (for dominant markers). For stronger linkages, i.e. smaller R-values, these lower bounds for u are much larger (up to 2) (Table 1).

An economic superiority of approach I compared to approach III, therefore, can be only expected for sufficiently large u-values, that is for sufficiently large genotyping cost relative to phenotyping cost. This superiority increases with increasing u-values and with increasing R.

Some numerical results for cost reduction (in %) of approach I relative to approach III are presented in Table 2—for different recombination fractions (R) and some ratios of genotyping to phenotyping costs (u) for codominant as well as for dominant markers. This cost reduction increases (i) with increasing u-values (for each fixed R) and (ii) with increasing R (for each fixed u) (Table 2). For $0 < R < 0.50$, the cost reduction for codominant markers is always larger than the cost reduction for dominant markers. A maximum cost reduction of 66.7% is attained at $R = 0.50$ and u = infinite.

These conclusions are only valid for sufficiently large genotyping cost relative to phenotyping cost. For many interesting traits in the field of plant improvement, however, the reverse is true and the phenotyping cost exceeds the genotyping cost by far. The previous conclusions on economic superiority of approach I are, therefore, of differentiated relevance for applied plant breeding depending on the trait of interest.

Comparison: approach II–approach III

For the comparison of the costs associated with approaches II and III (dominant individuals versus complete F_2) one obtains: All cost ratios are larger than 1 (for dominant markers as well as for codominant markers) with numerical values between 1.5 (for $R = 0$ and $u = \infty$) and 4.0 (for $R = 0.50$ and $u = 0$). Thus, no superiority of approach II (linkage analysis based on dominant individuals) compared to approach III (linkage analysis based on the complete F_2) can be observed. The numerical results of the cost ratios are, therefore, not presented here. They can be obtained from the first author upon request.

Comparison: approach I–approach II

For the comparison of the costs associated with approaches I and II (recessive individuals versus dominant individuals) one obtains: All cost ratios are smaller than 1 (for dominant markers as well as for codominant markers) with numerical values

Table 2 Cost reduction (in %) of approach I relative to approach III for different recombination fractions (R) and some ratios of genotyping to phenotyping costs (u) for codominant and dominant markers**Numbers in parentheses are the values for dominant markers

R	u				
	1	2	5	10	infinite
0.00	– (–)	0.00 (0.00)	25.00 (25.00)	36.36 (36.36)	50.00 (50.00)
0.05	– (–)	4.00 (3.36)	28.00 (27.52)	38.91 (38.50)	52.00 (51.68)
0.10	– (–)	8.38 (6.76)	31.28 (30.07)	41.69 (40.67)	54.19 (53.38)
0.15	– (–)	12.93 (10.19)	34.70 (32.64)	44.59 (42.85)	56.47 (55.10)
0.20	– (–)	17.46 (13.64)	38.10 (35.23)	47.47 (45.04)	58.73 (56.82)
0.25	2.20 (–)	21.76 (17.07)	41.32 (37.80)	50.21 (47.23)	60.88 (58.54)
0.30	7.02 (0.60)	25.62 (20.48)	44.21 (40.36)	52.67 (49.40)	62.81 (60.24)
0.35	11.07 (4.80)	28.85 (23.84)	46.64 (42.88)	54.73 (51.53)	64.43 (61.92)
0.40	14.12 (8.90)	31.30 (27.12)	48.47 (45.34)	56.28 (53.62)	65.65 (63.56)
0.45	16.02 (12.87)	32.82 (30.29)	49.61 (47.72)	57.25 (55.64)	66.41 (65.15)
0.50	16.67 (16.67)	33.33 (33.33)	50.00 (50.00)	57.58 (57.58)	66.67 (66.67)

between 0.1 (for R=0.50 and u=∞) and 1.0 (for R=0 and u=0). This shows that approach I (linkage analysis based on recessive individuals) is always superior to approach II (linkage analysis based on dominant individuals). The numerical results of the cost ratios are, therefore, not presented here. They can be obtained from the first author upon request.

For larger u-values, the sub-population of recessive (= susceptible) individuals exhibits some essential advantages for linkage analysis and genetic mapping. The reason for this superiority of recessive plants is based on the well-known fact that these plants have the highest information content compared to the sub-population of dominant (= resistant) plants and also compared to the complete F_2. For codominant markers, recessive plants have about twice the information content of dominant plants (Huehn and Piepho 2003). Among susceptible plants there are only three genotypes which can be distinguished phenotypically at the marker level. But, there are six different resistant genotypes with only three distinct phenotypes at the marker level (Huehn and Piepho, 2003; Table 1).

For dominant markers, there are three different genotypes with only two distinct phenotypes at the marker level (for susceptible plants) and six different genotypes with only two distinct phenotypes at the marker level (for resistant plants) (Huehn and Piepho, 2003; Table 1). From these facts, a higher information content of susceptible plants must be expected.

Discussion

The preceding results and conclusions are mainly determined by the cost per plant for marker determination (b) and by the cost per plant for testing resistance (a). The costs of initially developing molecular markers linked to the trait of interest were not considered. The analysis assumed that suitable molecular markers were already available. Although large numbers of molecular markers are in fact available for most major crops of importance for agriculture, continued development of new marker systems is necessary to reduce the cost associated with the use of molecular markers. The cost of developing new markers is usually absorbed by the public sector or by academic research. In this study, therefore, only costs of using the markers are included while costs of developing new markers are not considered.

Our results show that if the cost to genotype a plant is sufficiently large compared with the cost to phenotype the plant, then linkage analysis and genetic mapping should be only based on susceptible (= recessive) plants. Conversely, if the cost of phenotyping exceeds that for genotyping, it may be preferable to genotype all plants. These findings will not be surprising, yet it is useful to be able to make a precise assessment as to when it is preferable to genotype only susceptible plants.

In this paper we have addressed the question of whether it is preferable to genotype only a sub-

population of phenotyped genotypes or all of them for the purpose of accurately mapping a resistance gene. This is perhaps the first question to be answered when designing a mapping study, but not the only one. Once the optimal genotyping strategy has been identified, two further questions are: (1) What sample size is needed to detect linkage between the putative gene and one or more of its neighboring markers with sufficient probability? (2) What sample size is needed to achieve sufficient accuracy of the estimated position of the resistance gene?

Question (1) was addressed in Huehn and Piepho (2003), where it is shown how to choose sample size to achieve a desired power (probability to detect linkage). The answer to question (2) depends on marker spacing. Clearly, the higher the marker density, the less crucial it is to place the putative gene exactly between its flanking markers. It is more crucial to be able to place the gene at its accurate position. Mapping accuracy can easily be assessed by looking at the accuracy of the estimate of R for the closest marker, as quantified by its variance or its standard error. As a typical value for R, one might use the mean recombination fraction between adjacent markers. The variances are given in our paper (Eqs. 3–8), and it is a simple matter to determine the sample size needed to achieve a desired accuracy of the estimate of R, as quantified by its standard error. The sample size for a given strategy (dominant or co-dominant marker, genotyping of sub-population or the whole population) is then given by

$$N = \frac{1}{I(R)(SE)^2} \qquad (26)$$

where SE is the desired standard error and $I(R)$ is the information as given in Eqs. (3–8). Which of the possible strategies is preferable may be decided based on the relative costs of genotyping as shown in the present paper.

Placing the resistance gene correctly between its flanking markers is desirable, e.g., for marker-assisted selection (MAS). The least that one would require is that the resistance gene and two flanking markers considered for MAS are put in correct order. A more refined analysis could

involve using a multi-point approach to mapping the resistance gene (Ott 1991). Based on such an analysis, a confidence interval for map position could be computed. This approach was not considered in our paper. It should be stressed that with very dense maps correct ordering of all markers and the resistance gene will be difficult. Standard procedures for likelihood-based multi-point analysis rely on the assumption that markers have been ordered correctly, which is a strong assumption (Frisch et al. 2004). With very dense maps it is quite unlikely that a perfectly correct ordering of markers can be achieved by linkage analysis (two-point or multi-point). A detailed analysis of these issues is beyond the scope of this paper, which was mainly focussed on linkage detection rather than mapping.

As mentioned earlier, our development requires prior knowledge on a number of items. The genetic basis of the trait under consideration must be known (number of genes, dominant versus additive) and a map must be available. When this information is not yet available, additional costs will need to be considered in designing an overall optimal mapping strategy. The cost for doing the linkage analysis then only is one building block in a full economic assessment of alternative strategies.

References

Dreher K, Khairallah M, Ribaut JM, Morris M (2003) Money matters (I): costs of field and laboratory procedures associated with conventional and marker-assisted maize breeding at CIMMYT. Mol Breed 11:221–234

Edwards AWF (1972) Likelihood. Cambridge University Press, Cambridge

Frisch M, Quint M, Lübberstedt T, Melchinger AE (2004) Duplicate marker loci can result in incorrect locus orders on linkage maps. Theor Appl Genet 109:305–316

Hoisington DA, Melchinger AE (2004) From theory to practice: Marker-assisted selection in maize. In: Lörz H, Wenzel G (eds) Molecular marker systems in plant breeding and crop improvement (Biotechnology in Agriculture and Forestry, Vol. 55). Springer-Verlag, Berlin, Heidelberg, pp 335–352

Huehn M (1995) Determining the linkage of disease-resistance genes to molecular markers: the LOD SCORE method revisited with regard to necessary sample sizes. Theor Appl Genet 90:841–846

Huehn M, Piepho HP (2003) Determining the sample size for co-dominant molecular marker-assisted linkage detection for a monogenic qualitative trait by controlling the type-I and type-II errors in a segregating F2 population. Theor Appl Genet 106:840–845

Kuchel H, Ye G, Fox R, Jefferies S (2005) Genetic and economic analysis of a targeted marker-assisted wheat breeding strategy. Mol Breed 16:67–78

Moreau L, Lemarie S, Charcosset A, Gallais A (2000) Economic efficiency on one cycle of marker-assisted selection. Crop Sci 40:329–337

Morris M, Dreher K, Ribaut JM, Khairallah M (2003) Money matters (II): costs of maize inbred line conversion schemes at CIMMYT using conventional and marker-assisted selection. Mol Breed 11:235–247

Ott J (1991) Analysis of human genetic linkage. The Johns Hopkins University Press, Baltimore

Ragot M, Hoisington DA (1993) Molecular markers for plant breeding: comparisons of RFLP and RAPD genotyping costs. Theor Appl Genet 86:975–984

Yu K, Park SJ, Poysa V (2000) Marker-assisted selection of common beans for resistance to common bacterial blight: efficacy and economics. Plant Breed 119:411–415