Theor Appl Genet (1992) 83:294–300

The detection and estimation of linkage in polyploids using single-dose restriction fragments *

K.K. Wu [1,2,], W. Burnquist [1], M.E. Sorrells [1], T.L. Tew [2], P.H. Moore [2] and S.D. Tanksley [1]**

[1] Department of Plant Breeding and Biometry, 252 Emerson Hall, Cornell University, Ithaca, NY 14853, USA
[2] Hawaiian Sugar Planters' Association, 99-193 Aiea Heights Drive, Aiea, HI 96701, USA

Received May 1, 1990; Accepted April 30, 1991
Communicated by A.L. Kahler

Summary. Restriction fragment length polymorphism (RFLP) linkage maps have been constructed in several major diploid crops. However, construction of RFLP maps directly in polyploids has lagged behind for several reasons: (1) there are a large number of possible genotypes for each DNA probe expected in a segregating population, and these genotypes cannot always be identified readily by their banding phenotypes; and (2) the genome constitutions (allopolyploidy versus autopolyploidy) in many high polyploids are not clearly understood. We present here an analysis of these problems and propose a general method for mapping polyploids based on segregation of single-dose restriction fragments (SDRFS). SDRFs segregate 1:1 (presence:absence) in gametes of heterozygous plants. Hypothetical allopolyploid and autopolyploid species with four ploidy levels of 2n=4x, 6x, 8x, and 10x, are used to illustrate the procedures for identifying SDRFs, detecting linkages among SDRFs, and distinguishing allopolyploid versus autopolyploids from polyploids of unknown genome constitution. Family size required, probability of linkage, and attributes of different mapping populations are discussed. We estimate that a population size of 75 is required to identify SDRFs with 98% level of confidence for the four ploidy levels. This population size is also adequate for detecting and estimating linkages in the coupling phase for both allopolyploids and autopolyploids, but linkages in the repulsion phase can be estimated only in allopolyploids. For autopolyploids, it is impractical to estimate meaningful linkages in repulsion because very large family sizes (>750) are required. For high-level polyploids of unknown genome constitution, the ratio between the number of detected repulsion versus coupling linkages may provide a crude measurement of preferential chromosome pairing, which can be used to distinguish allopolyploidy from autopolyploidy. To create a mapping population, one parent (P_1) should have high heterozygosity to ensure a high frequency of SDRFs, and the second parent (P_2) should have a low level of heterozygosity to increase the probability of detecting polymorphic fragments. This condition could be satisfied by choosing outcrossed hybrids as one parental type and inbreds, haploids, or doubled haploids as the other parental type.

Key words: RFLP – Single-dose restriction fragment – Polyploids – Genetic mapping – Preferential chromosome pairing

Introduction

Linkage maps based on restriction fragment length polymorphic (RFLP) markers have been constructed in several major diploid crops, such as tomato (Bernatzky and Tanksley 1986), maize (Helentjaris et al. 1986; Burr et al. 1988; Hoisington et al. 1988), lettuce (Landry et al. 1987), and rice (McCouch et al. 1988). These maps and markers provide new plant breeding tools (Helentjaris et al. 1985; Nienhuis et al. 1987; Osborn et al. 1987; Paterson et al. 1988; Tanksley et al. 1989). Linkage maps of wheat (Kam-Morgan and Gill 1989) and potato (Bonierbale et al. 1988) are also under development, both utilizing diploid relatives in which mapping is more straightforward. However, for some polyploid species, such as

* Published as Paper No. 730 in the Journal Series of the Experiment Station, Hawaiian Sugar Planters' Association, and as Paper No. 792 in the Plant Breeding Series of Cornell University, Ithaca, NY 14853, USA

** To whom correspondence should be addressed

sugarcane (Sreenivasan et al. 1987), there are no closely related diploid species to use for mapping.

Construction of RFLP maps directly on polyploids is inherently more difficult than on diploids for several reasons. (1) A large number of genotypes for each DNA probe is expected in a segregating population; for example, hybridization of a single-copy DNA probe to restriction fragments from progeny plants of a selfed autotetraploid parent yields 19 possible genotypes if the parent had maximum heterozygosity (four alleles) per locus with bivalent chromosome pairings. (2) The genotypes of each probe-pair cannot always be identified through their RFLP phenotypes; this is especially true for higher polyploid species due to possible comigration of fragments on agarose gel electrophoresis. (3) The genome constitution (allopolyploid versus autopolyploid) of many polyploids is unclear, making it difficult to determine the patterns of inheritance. In fact, species with high ploidy levels may be mixtures of allopolyploid and autopolyploid genomes (de Wet 1980), for example, in sugarcane neither the basic chromosome number nor the genomic constitution is known with certainty (Sreenivasan et al. 1987)

One approach for avoiding the difficulty of mapping in polyploids is to analyze the segregation of each restriction fragment based in its presence or absence in the progeny. A fragment represented by a single dose (or one allele) in a heterozygous plant will segregate in a 1 : 1 ratio (presence : absence) in the gametes. Selfing of the same plant will produce progeny three-fourths of which will have the fragment and one-fourth of which will not; and a cross of the plant with another plant that does to have the fragment will produce progeny, only one-half of which have the fragment. A fragment that is present in a single dose and that segregates in a single-dose ratio (1 : 1) in the gametes of a plant is hereafter referred to as a **single-dose restriction fragment** (SDRF). Bonierbale et al. (1988) constructed an RFLP map by analyzing the segregation of SDRF markers in an interspecific hybrid population of diploid potato. In diploids, however, analysis of RFLPs as SDRF markers normally provides the same or less information than mapping based on identification of all genotypes (Tanksley et al. 1988a). However, in polyploids, analysis of SDRF markers can help solve the problems encountered in constructing RFLP maps. An SDRF is equivalent to a simplex allele in autopolyploids or to an allele at one heterozygous locus in a diploid genome in allopolyploids. The segregation of an SDRF is equivalent to the segregation of a simplex or a heterozygous allele in the gametes: half of the gametes will contain the DNA fragment and the other half will not. These gamete types can be visualized in the progeny from crosses in which one parent carries the fragment while the other parent does not. The gamete types of the SDRF markers can be analyzed for construction of linkage maps.

In this study, bivalent chromosome pairing among chromosomes in each linkage group is assumecd, and hypothetical of 2n = 4x, 6x, 8x, and 10x are used to illustrate the procedures for identifying SDRF markers, detecting linkages between them, and distinguishing allopolyploidy from autopolyploidy in species of unknown genome constitution. Required family sizes and attributes of different mapping populations are also discussed.

Identification of SDRFs

An SDRF in P_1 or in the progeny of $P_1 \times P_2$ can be identified according to two criteria: (1) the fragment should be present in P_1 yet absent in P_2, and (2) the fragment should segregate at a 1 : 1 ratio (presence : absence) in the progeny of $P_1 \times P_2$. These criteria are valid regardless of whether the species is allopolyploid, autopolyploid, or diploid. Presence versus absence reference to a fragment of a specific molecular weight does not require identification of the polymorphic fragment in the other parent. The methodology for the analysis is illustrated using hypothetical data in Fig. 1.

Ignoring those fragments that are present in both P_1 and P_2 or that do not segregate in a 1 : 1 ratio (determined by a χ^2 test) on the progeny, we identify fragment "A" of probe *i* and fragment "B" of probe *j* as meeting the criteria for classification as SDRFs (Fig. 1). In the progeny, five plants have both the A and B fragments (plants 1, 2, 8, 9, and 10), one plant contains A but not B, one plant contains B but not A, and five plants have neither A nor B (plants 3, 4, 5, 6, and 7). We thus designate the four genotype classes of the progeny as visualized gamete types of P_1 as AB, A, B, and null, respectively.

The conversion of genotype classes of the progeny to the gamete types of P_1 can be understood most easily in diploids. The cross of P_1 and P_2 could be a conventional diploid backcross, where P_1 is a hybrid F_1 plant of two inbred lines and P_2

DNA probe	Band no.	Parent		Progeny											
		P_1	P_2	1	2	3	4	5	6	7	8	9	10	11	12
i	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2(A)	-		-	-						-	-	-		-
	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	4	-		-	-	-	-	-	-	-	-	-			-
	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-
j	6	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	7(B)	-		-	-						-	-	-	-	
	8	-		-	-	-	-	-	-	-	-	-	-	-	-
	9		-	-	-	-	-	-	-	-	-	-	-	-	-

Fig. 1. Hypothetical fragment patterns of two DNA probes *i* and *j* of both parents and their progeny. For both parents, fragments 2, 4, 7, 8, and 9 are polymorphic, of which A and B are single-dose restriction fragments (=SDRFs) having a 1:1 ratio in the progeny

is either of the two inbred lines. The progeny of $P_1 \times P_2$ is thus equivalent to the first backcross progeny (Fig. 1). For probe *i* (or *j*), assuming that P_1 (= F_1) has two bands (heterozygous) and P_2 one band (homozygous) on an autoradiogram, the polymorphic band A (or B), ignoring the common bands between P_1 and P_2, should have a 1:1 ratio (presence:absence) in the backcross progeny. The 1:1 ratio reflects that the corresponding DNA fragment A (or B) detected by probe *i* (or *j*) is distributed in one-half of the gametes of P_1. The two DNA fragments A and B, because of independent assortment, will segregate into the same four gamete-types of P_1 as for polyploids.

The nature of P_1 and P_2 for maximizing the number of SDRFs in P_1 is discussed in detail in the mapping population section.

Table 1. The expected frequencies of both SDRF A and B linked in coupling or repulsion phase in both allopolyploids and autopolyploids. SDRF = single-dose restriction fragment; r_1 = coupling recombinant fraction; r_2 = repulsion recombinant fraction; $w = 1/(h-1)$; where h = no. of homologous chromosomes

Gamete type	Coupling	Repulsion	No. observed
AB	$0.5\,(1-r_1)$	$0.25(1-w)+0.5\,wr_2$	*a*
A	$0.5\,r_1$	$0.25(1-w)+0.5\,w(1-r_2)$	*b*
B	$0.5\,r_1$	$0.25(1-w)+0.5\,w(1-r_2)$	*c*
null	$0.5\,(1-r_1)$	$0.25(1-w)+0.5\,wr_2$	*d*

Detection of linkage

At meiosis, fragment A will be transmitted to one-half of the P_1 gametes. The same is true for B. If the association of A and B is random, the four classes of gamete types will occur in equal frequency. Let *a, b, c, d* be the observed number of plants in the four classes in the progeny of $P_1 \times P_2$. The ratio of nonrecombinant type versus recombinant type is expected to be 1:1. If the observed plants of the nonrecombinant type deviate significantly from those of the recombinant type, fragments A and B are expected to be linked either on the same chromosome (coupling phase) or on two homologous chromosomes (repulsion phase). To detect the linkage, the following equation (Mather 1951) is used:

$$\chi^2[1] = (a-b-c+d)^2/(a+b+c+d).$$

The value of χ^2 [1] is compared to a χ^2 table value with 1 *df*. A significant result indicates that A and B are linked in either the coupling or repulsion phase.

Estimation of linkage

The expected frequencies of the four gamete types are listed in Table 1. The maximum likelihood estimators of the recombination fraction (*r*) for both the coupling phase and repulsion phase of linkages are:

coupling: $r_1 = (b+c)/n$
repulsion: $r_2 = [(h\text{h}1)(a+d) - 0.5(h-2)n]/n,$

where h = the number of homologous chromosomes and $n = a+b+c+d$. For allopolyploids, $h = 2$, and for autopolyploids, h equals the ploidy level that is 4, 6, 8, or 10 in this report.

In the repulsion phase, chromosomes bearing DNA fragments A and B have a $1/(h-1)$ chance of being paired and partitioned into different gametes, or have a $1-(1/(h-1))$ chance of pairing not with each other but each pairing randomly with one of the other $(h-2)$ homologous chromosomes. [Note: For *h* homologous chromosomes, there are $y = (h-1)(h-3)(h-5)\ldots(h-(h-1))$ ways of random pairing, of which two particular chromosomes paired in $z = (h-3)(h-5)\ldots(h-(h-1))$ ways. The probability for A and B paired in meiosis is $(z/y) = 1/(h-1)$.] Of the n observed plants, $n' = n/(h-1)$ plants are responsible for estimating r. If the number of plants in the four gamete types are a', b', c', and d', then $r_3 = (a'+d')/n' = (h-1)(a'+d')/n$ and a χ^2-test should be $\chi^2[1'] = (a'-b'-c'+d')^2/n'$. The number of plants having random pairing of homologous chromosomes is expected to be $(1/4([1-1/(h-1)]n$ in each of the four types. Substituting $a'-(1/4)[1-1/(h-1)]n$ and d' with $d-(1/4)[1-1/(h-1)]n$ in $r_3 = (h-1)(a'+d')/n$, r_3 becomes $[(h-1)(a+d) - 0.5(h-2)n]/n$, which equals the maximum likelihood estimator of r_2. Since a', b', c', and d' are usually unknown, it is difficult to use r_3 and $\chi^2[1']$.

Negative values of r_2 could arise and indicate false assumptions about the value of h. For example, with $h = 8$ for autooctroploids, r_2 is -1.6 when $a,d = 10$ for $n = 100$. The calculation for $r_2 = -1.6$ is correct; however, the number of recombinants ($a, d = 10$) is questionable in this case. If $h = 8$ is true, the chance for $a, d = 10$ for $n = 100$ is almost nil, because a and d are each expected to be equal to $\{r_2/(8-1) + /1/4)[1-1/(8-1]\}100$. For $r_2 \geq 0$, a, d will be ≥ 21.5. On the other hand if $a, d = 10$ is actually observed for $n = 100$, one would conclude that $h = 2$ for allooctoploids.

The variances, V_1 and V_2 of the recombination fractions r_1 and r_2, are:

coupling: $V_1(r_1) = [r_1(1-r_1)]/n,$
repulsion: $V_2(r_2) = [2r_2w + (1-w)][2(1-r_2)w + (1-w)]/(4nw^2),$

where $w = 1/(h-1)$.

For the segregation of three SDRFs (A, B, and C) linked on the same chromosome, the expected frequencies of gamete types are listed in Table 2. The maximum likelihood estimators of r_{AB} (recombination fraction between A and B) and r_{BC} (recombination fraction between B and C) are:

$$r_{AB} = (c+d+e+f)/n$$
$$r_{BC} = (b+c+f+g)/n,$$

where $n = (a+b+c+d+e+f+g)$.

The variances for r_{AB} and r_{BC} are $[r_{AB}(1-r_{AB})]/n$ and $[r_{BC}(1-r_{BC})/n$, respectively. These are analogous to standard expectations of backcross (Mather 1951) and doubled haploid lines (Snape 1988).

Table 2. The expected frequency of gamete types of three SDRFs, linked in coupling phase. SDRF = single-dose restriction fragment; r_{AB} = recombination fraction between A and B; r_{BC} = recombination fraction between B and C

Gamete type	Expected frequency	No. observed
ABC	$0.5\,(1-r_{AB})\,(1-r_{BC})$	*a*
AB	$0.5\,(1-r_{AB})\,r_{BC}$	*b*
AC	$0.5\,r_{AB}\,r_{BC}$	*c*
A	$0.5\,r_{AB}\,(1-r_{BC})$	*d*
BC	$0.5\,r_{AB}\,(1-r_{BC})$	*e*
B	$0.5\,r_{AB}\,r_{BC}$	*f*
C	$0.5\,(1-r_{AB})\,r_{BC}$	*g*
null	$0.5\,(1-r_{AB})\,(1-r_{BC})$	*h*

Family size for detecting SDRFs

A restriction fragment on the P_1 autoradiogram may be present due to a single allele (single dose), two alleles (double dose), etc. If it is single dose, the fragment in P_1 will segregate 1:1 (presence:absence) in the gametes of P_1 or the progeny of $P_1 \times P_2$, if P_2 does not have the same corresponding fragment for both allopolyploids and autopolyploids. For allopolyploids, a double-dose fragment will not segregate or have a 3:1 ratio in the progeny, depending on whether the two fragments are on homologous or homoeologous chromosomes, respectively. For autopolyploids, a double-dose fragment will have a $(3h-2):(h-2)$ ratio. For example, the ratios are 5:1, 4:1, 3.7:1, and 3.5:1 for autotetraploids, autohexaploids, autooctoploids, and autodecaploids, respectively. The ratio approaches 3:1 as h becomes large. If the fragment on the autoradiogram is multiple dose (because of three or more alleles), it will have various ratios all higher than 3:1 in the progeny for both allopolyploids and autopolyploids. Therefore, it is important that we distinguish a 1:1 ratio from a ratio of 3:1 or greater with a high level of confidence, in order to distinguish SDRFs (or 1:1 ratio) from other categories (or $\geq 3:1$) in both allopolyploids and autopolyploids.

To test the null hypothesis, Ho: (presence:absence) = (1:1) against the alternative hypothesis, Ha: (presence:absence) $\geq (3:1)$, a χ^2 test can be used. Le p = the number of plants in the progeny of $P_1 \times P_2$ with the presence of a polymorphic band, e.g., band F, and q = the number of plants in the progeny of $P_1 \times P_2$ without F; then the χ^2 test is

$$\chi^2[2] = (p-q)^2/(p+q).$$

A nonsignificant test will lead to the rejection of Ha and the acceptance of Ho: the polymorphic band F is a SDRF. Since the nonsignificant test is critical, both type I ($=\alpha_1$) and type II ($=\alpha_2$) error rates should be considered. For $\alpha_1 = \alpha_2 = (0.025, 0.010, 0.005)$, the family sizes (Mather 1951) are 54, 75, 92. For a nonsignificant $\chi^2[2]$ test to accepted 1:1 ratio in the progeny and to determine polymorphic fragments that are also single-dose fragments in P_1, family sizes of 54, 75, or 92 will have respectively 95, 98, or 99% confidence level calculated by $(1-\alpha_1-\alpha_2)100\%$.

Probes that produce SDRFs need to be identified and selected from a DNA clone library. If 75 plants are used, the $\chi^2[2]$ test can be achieved in two steps: (1) use 37 plants (about half of 75) to collect data from the DNA library for the first selection of probes that are most likely to have the desired markers (=SDRFs) showing ratios nonsignificantly different from 1:1; and (2) use an additional 38 plants to collect additional data from the selected probes and later combine both data to make a final $\chi^2[2]$ test on 75 plants. The first selection of probes with 37 plants will result in making 25% of type II error and 1% of type I error. The second selection with the combined data collected from 75 plants, which would reduce the type II error from 25 to 1%, will result in making a total error of 2%, or giving a 98% confidence that the probes detect SDRFs. [Note: type II error can be further reduced by increasing the level of type I error without changing family size. For $n=75$, increasing α from 1 to 5% will result in reducing α_2 from 1% to less than 0.1%.]

Family size for detecting linkages in coupling phase

For a pair of linked DNA fragments, A and B, the number of recombinant plants increases as the distance between A and B on a chromosome increases. Consequently, the observed ratio of plants for nonrecombinant type versus recombinant type approaches 1:1. A significant $\chi^2[1]$ test is desired for rejecting the null hypothesis (Ho) that a 1:1 ratio is true (or there is no linkage) and accepting the alternative hypothesis (Ha) that the two fragments are linked. The relationship between family sizes required and the maximum detectable recombination fraction ($=\max r_1$) values with one tail significant probability level $\alpha_1 = 0.01$ is shown in Fig. 2. It is established by $\max r_1 = \{0.5 - z(\alpha_1)\sqrt{V_1(r_1)}\}$, where z is the variable of a standard normal distribution, with $z(\alpha_1) = 2.3264$ and $V_1(0.5) = 0.5(1-0.5)/n$. For any calculated recombination value, r_1 with $r_1 \leq \max r_1$, the $\chi^2[1]$ test will reject the Ho and accept Ha that r_1 is significantly less than 0.5.

The type II error rate (α_2) was not specified in the equation of $\max r_1$. Two reasons for not using α_2 in $\max r_1$ are: (1) there is no type II error once the result of the test is proved significant, and (2) a greater range of r_1 can be detected for a given family size; for example, for $n=75$ with $\alpha_1 = 0.01$, the range is from 0 to 0.37 ($=\max r_1$), whereas with $\alpha_1 = \alpha_2 = 0.01$, the range will be from 0 to 0.25.

One has to choose the range of linkages to decide the proper family size for estimating r_1. For most practical purposes in plant breeding, the linkages in the range of 0.1 to 0.25 are useful in developing tags for genes of interest (Tanksley 1988 b). A sample size of 75 can be chosen because it is adequate to detect linkages of

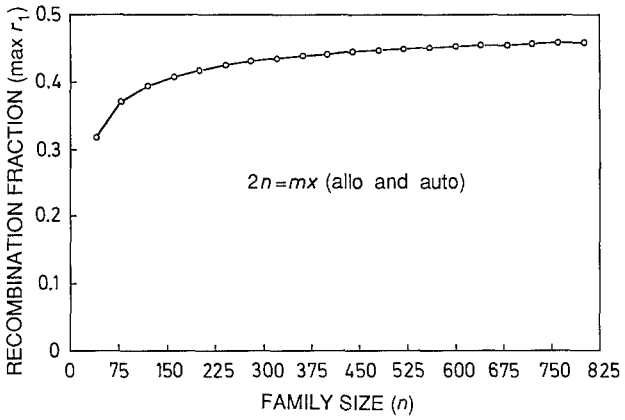

Fig. 2. The maximum detectable recombination fractions (maxr_1) for linkage in the coupling phase for both allopolyploids and autopolyploids with 2n=mx, where m=level of ploidy. The curve is established on the basis of maxr_1=0.5 (1−2.3264 $\sqrt{1/n}$) with $P \leq 0.01$

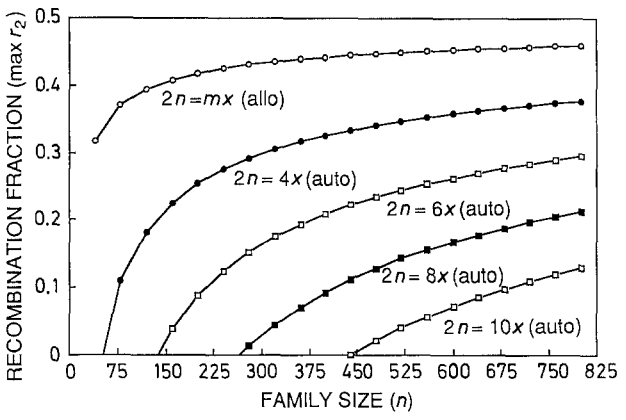

Fig. 3. The maximum detectable recombination fractions (maxr_2) for linkage in repulsion phase for both allopolyploids and autopolyploids with 2n=mx, where m=level of ploidy. The curves are established on the basis of maxr_2=0.5 (1−2.3264 $(h-1)\sqrt{1/n}$) with $P \leq 0.01$

$r_1 \leq 0.25$ with a confidence level ≥98%, since at maxr_1=0.25, the type I and II error rates are each equal to 0.01.

This is the same family size required for detecting SDRFs. Following the same example as in detecting SDRFs, we can use the data collected from the 37 plants to first select possible linked probe pairs. Since this is a preselection with a smaller sample size, the significant probability level should be higher than 0.01 (0.10, for example) to ensure that more linkage pairs are selected. After selection, the combined data of 75 plants are then tested for linkages using 0.01 ($=a_1$) as the significant probability level.

Family size for detecting linkages in repulsion phase

The family size for detecting maximum repulsion linkages in P_1 of allopolyploids is the top curve (2n=mx) as shown in Fig. 3. It is expected to be the same curve (=maxr_1) as that for detecting coupling linkages. Because in allopolyploids the number of homologous chromosomes is $h=2$, the likelihood estimator of $r_2=[(h-1)(a+d)-0{,}5(h-2)n]/n$ reduces to $r_2=(a+d)/n$, which is similar to $r_1=(b+c)/n$ for coupling linkages.

Family sizes of autopolyploids required for detecting linkages in the repulsion phase, however, are quite different from those for the coupling phase. Figure 3 shows the relationship between family sizes required and the maximum detectable r_2(=maxr_2) values with one tail significant probability $\alpha_1=0.01$. The maximum values of r_2 are calculated by assigning different values of h and n to equation $\max r_2=\{0.5-z(\alpha_1)\sqrt{V_2(r_2)}\}$, with $z(0.01)=2.3264$ and $V_2(0.5)=0.25(h-1)^2/n$. [Note: "$r_2 \leq \max r_2$ "⇔"$(0.5-r_2)/\sqrt{0.25(h-1)^2/n} \geq z(\alpha_1)$" can be converted to $\chi^2[1] \geq z^2(\alpha_1)$. Therefore, $\chi^2[1]$ also can be used to detect linkages in the repulsion phase.] At higher ploidy levels, larger family sizes are required to detect repulsion linkages.

A family size of 75 can detect tight linkages of autotetraploids but with large error variance; for example, if $r_2=0.1$, the standard error would be 0.167 with a 95% confidence interval of 0.1 ± 0.327, which makes the value of 0.1 practically meaningless. Seventy-five plants is not a sufficient number to cannot detect any linkages in repulsion phase for autopolyploids with ploidy levels equal to or greater than six (Fig. 3).

The relative efficiency (E) of estimating recombination fractions between linkages in coupling and in repulsion phases is $E(r_1, r_2)=(1/V_1)/(1/V_2)$. Taking $r_1=r_2=0.2$ with the same family size of n, for example, $E=7.4$, 2.5, 1.3, and 0.8% for $h=4$, 6, 8, and 10, respectively. To have $r_2=0.2$ with a similar confidence level of $r_1=0.2$, the family size for estimating r_2 needs to increase 13.5 times of n for autotetraploids. This low efficiency makes estimation of r_2 for autopolyploids impractical.

Estimation of recombination fraction and its confidence interval

A significant $\chi^2[1]$ for detecting linkages indicates that the recombinant fraction ($r=r_1$ or r_2) is significantly less than 0.5 and can be calculated by using one of the maximum likelihood estimators. To estimate the confidence interval of r, different family sizes may be required, depending on the desired width of the estimated intervals and the calculated r values. Intervals estimated by $r \pm 1.96$ SE (SE=standard error of r) will provide 95% confidence that true r is in this range. Larger family sizes are required to reduce the SE or to maintain a constant SE for increasing r values. For instance, for a family size of 75 and an estimated r of 0.25, the estimated standard error of r is 0.05. Considering one SE, the true value of r may be any value between 0.20 and 0.30 with ca. 68% confidence. For two SE, the interval is from 0.15 to 0.35

with approx. 95% confidence, but the width of the interval is 0.2. In order to keep the width to 0.1 and with a 95% confidence interval, the SE must be reduced to 0.025 from 0.05 by increasing the family size from 75 to 300. However, when the estimated r value decreases, its SE also decreases, reducing the ranges of the confidence interval. For example, for $r=0.107$ and $n=75$, the interval is 0.037 to 0.177, which is less than that from 0.15 to 0.35 for $r=0.25$.

At low values of r, the binomial distribution for the number of recombinants becomes more asymmetrical. The symmetrical interval calculated by $(r \pm zSE)$ underestimates the true interval and often includes undefined negative values (Silver 1985). The equations for estimating the lower and upper limits of the confidence interval of observed r values in this paper are listed in the Appendix. Take $n=40$, for example: the limits of 95% confidence interval for $r=0.1$ are [0.007, 0.193] versus [0.028, 0.237], calculated respectively from ± 1.96 SE and from binomial distribution; for $r=0.2$, they are [0.076, 0.324] versus [0.094, 0.356]; and for $r=0.3$, they are [0.158, 0.442] versus [0.166, 0.466].

Detection of allopolyploid or autopolyploid

Genome constitutions of many high-level polyploid species are unknown, (Simmonds 1979). However, most of them are assumed to be allopolyploids based on bivalent pairing on chromosomes in meiosis. Analysis of SDRF markers can potentially aid in distinguishing autopolyploidy from allopolyploidy.

Comparing the maximum detectable r values ($\max r_1$) in Fig. 2 with top curve of $\max r_2$ in Fig. 3, for allopolyploids with a segregating population size of $n=75$, the proportion of linkages that can be detected in the coupling phase is expected to be the same as that in the repulsion phase. Again, comparing curves in both figures for an autopolyploid (2n=4x), the proportion of linkages that can be detected in the repulsion phase is about one-fourth of those that can be detected in the coupling phase; for 2n=6x, 8x, and 10x, all linkages that can be detected are expected to be in the coupling phase and none in the repulsion phase.

For each linkage group, the observed ratio of detectable SDRF pairs linked in repulsion versus coupling phases is also expected to be 1:1 for allopolyploids and is 0.25:1 or 0:1 for autopolyploids for $m \geq 4$. This ratio can be used as an index to measure the preferential chromosome pairing among chromosomes in each linkage group. For high-level polyploids ($m \geq 4$) with unknown chromosome pairing behavior or unknown genome constitution (allopolyploids versus autopolyploids), with $n=75$, a nonsignificant χ^2 test for a 1:1 ratio of detected linkage pairs within a linkage group will indicate preferential bivalent chromosome pairing in the linkage group, and for a 0:1 ratio will indicate random bivalent association of homologous chromosomes. A χ^2 test significantly different from either a 1:1 or a 0:1 ratio will indicate that chromosomes in the linkage group may have partial preferential or multivalent chromosomal pairings.

Once a large number of SDRF pairs is detected and grouped into linkage groups, the results of a χ^2 test to compare ratios among and within linkage groups can be analyzed. If most or all the detected linkage groups have 1:1 (or 0:1) ratios, the species should be an allopolyploid (or autopolyploid) species.

Mapping populations

A haploid population derived from a highly heterozygous plant (P_1) is the most efficient mapping population for SDRFs because of DNA fragments in the "P_2" can be considered null. Therefore, every band in P_1 is polymorphic between P_1, and "P_2" and can be tested for a 1:1 ratio in the haploid population. High levels of heterozygosity in P_1 will maximize the chance of identifying SDRFs.

A hybrid population produced from a cross between a heterozygous (outcrossed) parent (P_1) and a haploid or a homozygous (inbred or doubled haploid) parent (P_2) is the second best mapping population. The two parents can be of the same or different species, as long as polymorphism between them is high and a large number of hybrids can be produced.

For progeny derived from selfing P_1, SDRF markers A and B can still each be identified by a 3:1 ratio in the progeny. The four classes (AB, A, B, and null) in F_2, however, have an expected ratio of 9:3:3:1. The likelihood estimation equation for A and B linked in the coupling phase is $[2(r-1)a/(3-2r+r^2)]+[2(1-r)(b+c)/(2r-r^2)]+[2d/(r-1)]=0$ (Allard 1956), with $V(r)=[r(2-r)(3-2r+r^2)]/[2(3-4r+2r^2)n]$ (Mather 1951). Because of the low efficiency compared to other crosses (Mather 1951; Allard 1956), selfing is least desirable.

Conclusion

There are many types of polyploids besides the two typical types, allopolyploids or autopolyploids. Also, polyploidy is a state, not a process or an event. Through evolution, an autopolyploid species may eventually become diploidized (de Wet 1980). The property of chromosome association of many autopolyploid species is therefore likely to be a combination of random and preferential chromosome association.

In polyploids of 2n=mx, there are x chromosome linkage groups. Usually, chromosomes in most of the x linkage groups are paired in bivalents, with few multivalent chromosome pairing groups. Bivalent chromosome pairing groups were used to develop the expected frequencies in this paper. The preferential chromosome pairing index can be used to verify the bivalent assump-

tions for chromosomes in each linkage group with a χ^2 test.

Once a large number of SDRF pairs is detected and placed into linkage groups from a polyploid species of unknown genome constitution, the ratio (index) of the observed number of repulsion versus coupling linkages for each linkage group can be tested against the expected ratio of either 1:1 (index = 1) for preferential bivalent chromosome pairings or 0:1 (index = 0) for nonpreferential bivalent chromosome pairings. An index with a value significantly different from 1 or 0 may indicate partial preferential bivalent chromosome pairings or multivalent chromosome pairings.

If the observed index values are 1's for all linkage groups, there is an allopolyploid species; if they are 0's, an autopolyploid; and if the average index is between 0 and 1, an autopolyploid but in the process of diploidization (Kimber 1984).

Acknowledgements. We would like to thank S. J. Schwager, M. W. Bonierbale, J. Anderson, and J. da Silva of Cornell University, and B. C. Wells and S. Kumar of the Hawaiian Sugar Planters' Association for reading the manuscript and providing valuable suggestions.

Appendix

To estimate the lower (r_{low}) and upper (r_{high}) limits of a confidence interval for the observed recombination fraction ($=i/n$) for allopolyploids (both coupling and repulsion linkage) and autopolyploids (coupling linkage only), the following two equations need to be solved:

(1) for r_{low}

$$\alpha_1 = \sum_{x=i}^{n} \frac{n!}{x!\,(n-x)!} (r_{low})^x (1-r_{low})^{(n-x)}$$

(2) for r_{high}:

$$\alpha_2 = \sum_{x=0}^{i} \frac{n!}{x!\,(n-x)!} (r_{high})^x (1-r_{high})^{(n-x)},$$

where x = the number of recombinant plants and has a binomial distribution $b(x; n, nr)$; i = the number of observed recombinant plants $P\{r_{low} \le i/n \le r_{high}\} = 1-\alpha_1-\alpha_2$; $\alpha_1 = P\{x \ge i;\ x \sim b(x; n, nr_{low})$ and $(0 \le r_{low} \le i/n)\}$ and $\alpha_2 = P\{x \le i;\ x \sim b(x; n, nr_{high})$ and $(i/n \le r_{high} \le 0.5)\}$; $\alpha_1 = \alpha_2 = (1/2\alpha$, if $0 < r_{low}$ and $r_{high} < 0.5$ or $\alpha_1 + \alpha_2 = \alpha$, if $r_{low} = 0$ or $r_{high} = 0.5$; and α = the probability of error, while $(1-\alpha)\,100\%$ = percent confidence of the estimated interval, e.g., $(1-0.05)\,100\% = 95\%$ confidence level.

References

Allard RW (1956) Formulas and tables to facilitate the calculation of recombination values in Heredity. Hilgardia 24:235–278

Bernatzky R, Tanksley SD (1986) Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences. Genetics 112:887–898

Bonierbale MW, Plaisted RL, Tanksley SD (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics 120:1095–1103

Burr B, Burr FA, Thompson KH, Albertson MC, Stuber CW (1988) Gene mapping with recombinant inbreds in maize. Genetics 118:519–526

Helentjaris T, King G, Slocum M, Siedenstrang C, Wegman S (1985) Restriction fragment length polymorphisms as probes for plant diversity and their development as tools for applied plant breeding. Plant Mol Bol 5:109–118

Helentjaris T, Slocum M, Wright S, Schaefer A, Nienhuis J (1986) Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphism. Theor Appl Genet 72:761–769

Hoisington DA, Coe EH, Neuffer MG (1988) Gene list and linkage map of maize (*Zea mays* L.). Maize Genet Coop Newslett 62:125

Kam-Morgan LNW, Gill BS (1989) DNA restriction fragment length polymorphisms: a strategy for genetic mapping of D genome of wheat. Genome 32:724–732

Kimber G (1984) Evolutionary relationships and their influence on plant breeding. In: Gustafson JP (ed) Gene manipulation in plant improvement. Plenum Press, New York, pp 281–293

Landry BS, Kesseli RV, Farrara B, Michelmore RW (1987) A genetic map of lettuce (*Lactuca sativa* L.) with restriction fragment length polymorphism, isozyme, disease resistance and morphological markers. Genetics 116:331–337

Mather K (1951) The measurement of linkage in heredity. Methuen, London, p 149

McCouch SR, Kochert G, Yu ZH, Wang ZY, Khush GS, Coffman WR, Tanksley SD (1988) Molecular mapping of rice chromosomes. Theor Appl Genet 76:815–829

Nienhuis J, Helentjaris T, Slocum M, Ruggero B, Schaefer A (1987) Restriction fragment length polymorphism analysis of loci associated with insect resistance in tomato. Crop Sci 27:797–803

Osborn TD, Alexander DC, Fobes JF (1987) Identification of restriction fragment length polymorphisms linked to genes controlling soluble solids content in tomato fruit. Theor Appl Genet 73:350–356

Paterson AH, Lander ES, Hewitt JD, Paterson S, Lincoln SE, Tanksley SD (1988) Resolution of quantitative traits into Mendelian factors by using a complete RFLP linkage map. Nature 335:721–726

Simmonds NW (1979) Principles of crop improvement. Longman, New York, p 408

Silver J (1985) Confidence limits for estimates of gene linkage based on analysis of recombinant inbred strains. J Hered 76:436–440

Snape JW (1988) The detection and estimation of linkage using doubled haploid or single seed descent populations. Theor Appl Genet 76:125–128

Sreenivasan TV, Ahloowalia BS, Heinz DJ (1987) Cytogenetics. In: Heinz DJ (ed) Sugarcane improvement through breeding. Elsevier, New York, pp 211–253

Tanksley SD, Miller J, Paterson A, Bernasky R (1988a) Molecular mapping of plant chromosomes. In: Gustafson JP, Appels R (eds) Chromosome structure and function. Plenum Press, New York, pp 157–173

Tanksley SD, Bernatzky R, Lapitan NL, Prince JP (1988b) Conservation of gene repertoire but not gene order in pepper and tomato. Proc Natl Acad Sci USA 85:6419–6423

Tanksley SD, Young ND, Paterson AH, Bonierbale MW (1989) RFLP mapping in plant breeding: new tools for an old science. Bio/Technol 7:257–264

Wet JMJ de (1980) Origins of polyploids. In: Lewis WH (ed) Polyploidy biological relevance. Plenum Press, New York, pp 3–7