

Statistical Aspects of Genetic Mapping in  
Autopolyploids

Maria I. Ripol,<sup>1</sup> Gary A. Churchill,

Jorge A.G. da Silva,<sup>2</sup> Mark Sorrells

Department of Plant Breeding and Biometry

Cornell University, Ithaca, NY 14853

February 27, 1996

<sup>1</sup>present address:

<sup>2</sup>present address:

## Abstract

Many plant species of agriculture importance are polyploid, having more than two copies of each chromosome per cell. In this paper we describe statistical methods for genetic map construction in autopolyploid species. The first step is to determine the dosage of each DNA fragment (electrophoretic band) from its segregation ratio. Fragments present in single dose can be used to construct framework maps for individual chromosomes. Fragments present in multiple dose can often be used to link the single chromosome maps into homologous groups and provide additional ordering information. Marker phenotype probabilities were calculated for pairs of markers arranged in different configurations among the homologous chromosomes. These probabilities were used to compute a maximum likelihood estimator of the recombination fraction between pairs of markers. A likelihood ratio test for linkage of multidose markers was derived. The information provided by each configuration and power and sample size considerations are also discussed. A set of 294 RFLP markers scored on 90 plants of the species *Saccharum spontaneum* L. was used to illustrate the construction of an autopolyploid map. Previous studies conducted on the same data revealed that this species of sugar cane is an autooctaploid with 64 chromosomes arranged into eight homologous groups. The methodology described permitted consolidation of 54 linkage groups into ten homologous groups.

# 1 Introduction

Genetic linkage maps are powerful tools for the study of inheritance, facilitating gene localization and eventual cloning of genes causing various traits. Statistical methods for the construction of genetic maps in diploid species are relatively well developed (e.g., Lander and Green, 1987) and genetic maps have been constructed for several diploid crop species such as tomato, maize, lettuce and rice (Tanksley, 1989). However, a number of plant species of agricultural importance are polyploid. For some polyploid species, such as wheat and potato, genetic maps have been built based on closely related diploid species (O'Donoghue et al., 1989; Kam-Morgan et al., 1989). Mapping polyploid species, such as sugar cane, that do not have a close diploid relative, presents three major problems (Wu et al., 1992). First, the number of genotypes possible in polyploids is much greater than in diploids; second, the genotypes are not always readily identifiable by their phenotype; and third, the genomic constitution of the polyploid (auto vs allo) is unclear for many species.

In an allopolyploid species, such as wheat, there are specific meiotic pairings among the homologous chromosomes. Thus the genetics of allopolyploids is similar to diploids except for the multiple genomes. In autopolyploid species the chromosomes pair randomly at meiosis within homologous sets. The problem of constructing genetic maps in an autopolyploid is two-fold; loci must be ordered along individual chromosomes, and the chromosomes must be assigned to homologous groups. The latter problem is unique to autopolyploids. The construction of a genetic map in autopolyploids is accomplished in several steps. The dosage of each DNA fragment (electrophoretic band) is determined first; then pairs of loci are tested for linkage, according to their dosage and configuration. Distances between linked loci are calculated, and they are or-

dered along the chromosomes. Finally, the chromosomes are collected into homologous groups.

The rest of this paper is organized as follows. The BACKGROUND section discusses terminology, the relevant genetics of autopolyploid species, and the unique mapping problems they present. Further details on the genetics of autopolyploids can be found in da Silva and Sorrells (1996). In the RESULTS section we first discuss the expected number of higher dose fragments necessary to collect all of the chromosomes into homologous groups. The fragment phenotype probabilities were derived for different genotypic configurations of loci. These probabilities were used to compute a maximum likelihood estimator of the recombination fraction between two loci in a given configuration. A likelihood ratio test for linkage was derived, and significant linkages were transformed into map distances. The information provided by each configuration, and power and sample size considerations were also discussed. In the EXAMPLE section, a map of the sugar cane genome was used to illustrate the process of constructing a genetic map in an autopolyploid.

## 2 Background

### 2.1 Terminology

Numerous authors have reviewed linkage mapping using molecular markers and the associated terminology (e.g. Paterson et al.) and da Silva and Sorrells (1996) discussed specific terms used for polyploids. A restriction fragment length polymorphism (RFLP) that has been mapped is referred to as a marker and a DNA clone that has been labeled, as a probe. Restriction DNA fragments are often visualized as bands on autoradiographs of labeled membranes

from Southern blots but other detection technologies may also be used. A single dose restriction fragment (SDRF) is defined as a DNA fragment that is present in just one chromosome and can be identified by 1:1 (presence:absence) segregation regardless of the type of ploidy (We et al.). Fragments present in two and three copies (duplex and triplex loci) will be referred to as double- and triple-dose restriction fragments, (DDRF and TDRF, respectively) or multiple-dose DNA fragments.

## 2.2 Marker Genotype Configurations

In classical (diploid) genetic mapping with dominant alleles, when the two dominant alleles, A and B, are located on the same chromosome, and the two recessive alleles, a and b, are on the homologous chromosome, the alleles are said to be in a *coupling* configuration. When the dominant allele for one locus appears in a chromosome with the recessive allele for the other locus, the alleles are said to be in *repulsion*. The gametes that inherit the same arrangement present in the parental chromosomes are of *parental type*. The remaining gametes are *recombinant*.

For polyploids it is necessary to distinguish between the *haploid number*  $n$  (the number of chromosomes in a gamete) and the *monoploid number*  $x$  (the number of chromosomes in a basic set). The *basic set* of chromosomes for a species contains one chromosome from each homologous group. *Polyploids* are organisms with multiple copies of the basic chromosome set. The *ploidy number* of the species, denoted by  $m$ , represent the number of homologous chromosomes in each somatic cell. Diploids ( $m = 2$ ) have pairs of homologous chromosomes in each cell, while tetraploids ( $m = 4$ ) have four copies of each, and so on.

For all organisms, the total number of chromosomes in a somatic cell is given by  $2n = mx$ . In diploids,  $m = 2$  so  $n = x$ , that is, the number of chromosomes in a gamete equals the number of chromosomes in a basic set. In polyploids, each gamete will have half of the chromosomes in each homologous group, or  $n = mx/2$ . For example, humans are diploids with 46 chromosomes in each somatic cell, so  $n = 23, x = 23$  and  $m = 2$ , giving  $2n = 2x = 46$ . Wheat is hexaploid ( $m = 6$ ) with 42 chromosomes in each somatic cell, so  $n = 21$ , and  $x = 7$ , giving  $2n = 6x = 42$ .

The genetic map of an autopolyploid species has two components: linkage groups and homologous groups. The number of linkage groups or chromosomes ( $2n$ ), and the ploidy number ( $m$ ) are all particular to the species. The sugar cane species used here, *Saccharum spontaneum* L., is an autooctaploid ( $m = 8$ ) with  $2n = 64$  chromosomes in each cell, so a complete map will have 64 linkage groups organized into  $x = 8$  homologous groups of  $m = 8$  chromosomes each (da Silva, 1993).

We will make the following assumptions:

A1: homologous chromosomes segregate from each other at meiosis

A2: non-homologous chromosomes segregate independently

A3: no segregation distortion

A4: no crossover position interference

A5: no chromatid interference.

Assumption A1 ensures that all gametes will have exactly one of each kind of chromosome, avoiding any chromosome number mutations or aberrations. Together with A2 and A3 it gives a *segregation ratio* of 1:1 for two different

alleles at any locus. Assumptions A4 and A5 imply that the probability of recombination is the same throughout the genome.

The following additional assumptions, pertaining particularly to autopolyploids, will also be made here:

A6: each locus can belong to only one homologous group

A7: homologous chromosomes form random pairs at meiosis

A8: no double reduction - each chromosome will segregate from its pair at meiosis into a different gamete.

Assumption A6 can be violated when the particular DNA sequence that constitutes a locus is duplicated elsewhere in the genome. A7 assures that we are dealing with an autopolyploid, as opposed to an allopolyploid or something in between. There has been much discussion in the earliest literature on autopolyploids about double reduction (Haldane 1930, Mather 1936, Fisher 1943, Geiringer 1949). However, de Winton and Haldane (1932) note that double reduction in autotetraploids is “a rare or non-existent phenomenon”.

One of the complications when dealing with autopolyploid species is that a simple backcross design does not provide the necessary information to map the markers (Fisher, 1946; Bingham, et al., 1968; da Silva and Sorrells, 1995). A *modified backcross* was used by da Silva et al. (1993) to construct a map in *S. spontaneum*. From a highly heterozygous individual, SES208, a gamete was extracted and doubled to form the doubled haploid genotype ADP068, which was then crossed with SES208. The fragments or loci available for mapping will all be present in SES208 and absent in ADP068, so this resembles the conditions of a backcross.

A complete genetic map will have the required number of linkage groups,

each one assigned to a homologous group. Single dose fragments appear in one chromosome, so they can be used to establish a framework map of linkage groups. Under assumption A6, each locus belongs to only one homologous-group, so when a higher dose marker maps to more than one linkage groups, those chromosomes are assumed to belong to the same homologous group. Thus, higher dose fragments allow us to collect chromosomes into homologous sets.

### 3 Results

#### 3.1 Determination of Dosage

Given the ploidy level of an organism, we can estimate the dosage of a fragment by observing the proportion of plants in the progeny that exhibit the marker. For a marker A, we can classify progeny into two phenotypic classes A and  $\emptyset$ , where A represents the case when at least one copy of the fragment is present and  $\emptyset$  represents the case in which no copies are present. Out of  $m$  homologous chromosomes, the informative parent has  $k$  which contain the marker and  $m - k$  which do not, where  $k \in \{0, 1, 2, \dots, m/2\}$ . At meiosis, the  $m$  homologous chromosomes segregate randomly (under assumptions A1, A2, A7 and A8) so that each gamete receives half of the chromosomes. Assuming no segregation distortion (A3), the probability of a gamete having at least one copy of marker A. Thus the probability that a plant has phenotype A is

$$p(k) = 1 - \frac{\binom{k}{0} \binom{m-k}{m/2}}{\binom{m}{m/2}}, \quad k = 0, \dots, m/2. \quad (1)$$

Table 1 presents the expected proportion of plants with phenotype A for dosages ( $k$ ) and organisms of different ploidy levels ( $m$ ).

TABLE 1 GOES HERE

The observed segregation ratios for each of the 294 markers in the sugar cane study were calculated, and a histogram of their frequencies appears in Figure 1. The distribution of the frequencies appears to be bimodal, with a large group centered around 0.5, the expected segregation ratio for single dose fragments, and a smaller group centered around 0.78, where we would expect the double dose fragments. There is also one observation at 0.93, where a triple dose fragment should appear.

FIGURE 1 GOES HERE

The number of plants in the progeny that inherit at least one copy of a fragment, will be a binomial random variable, with sample size  $N$  equal to the number of progeny and the success probability  $p = p(k)$  as given by equation 1.

To classify fragments into dosage groups, we constructed acceptance regions for single, double, triple and quadruple dosage based on this binomial distribution. We assume  $N = 80$  as an approximate sample size for all markers in the sugar cane mapping data, allowing for some missing data. Slightly overlapping 99% acceptance regions for each dosage were constructed and used to classify each fragment according to its segregation ratio. The 294 polymorphic fragments were classified as: 244 singles, 41 doubles, 1 triple and 8 ambiguous markers. Of the ambiguous markers, 7 fall in the gap between the acceptance regions of single and double dose. The remaining marker is doubly classified as double and triple dose. More elaborate classification schemes were considered in Ripol (1994) but this method proved to be adequate for the problem of classifying this set of markers.

## 3.2 Connectedness

The problem of finding the number of higher dose fragments needed to assign all linkage groups into homologous sets can be formulated in graph theory terms (Chartrand and Lesniack, 1979). A *random graph* consists of a set of *vertices*  $V = V_i, i = 1, \dots, n$ , and a set of randomly selected *edges*  $E = (v_i, v_j)$ . Random edges can be placed on the graph by assigning a uniform probability to each vertex and selecting two of them without replacement. Visually, the vertices are represented by dots, and the edges by lines that connect the vertices that compose it. A *path* of length  $l$  between vertices  $i$  and  $j$  is an ordered set of edges  $\{(v_i, v_{k_1}), (v_{k_1}, v_{k_2}), (v_{k_2}, v_{k_3}), \dots, (v_{k_{l-1}}, v_{k_l}), (v_{k_l}, v_j)\}$  such that the first element of the first pair is vertex  $i$ , the last element of the last pair is vertex  $j$ , and for all other pairs, the first element of the pair is identical to the second element of the previous pair. Two vertices are *connected* if there exists a path between them and, by convention, every vertex is connected to itself. A *connected graph* is one in which every pair of vertices is connected.

We can consider each homologous group of  $m$  chromosomes as a set of vertices, and higher dose fragments that map to two homologous chromosomes as the edges that connect them. Each one of the  $x$  homologous groups would then be a separate graph. Random edges assume that higher dose fragments occur randomly among the  $m$  homologous chromosomes.

The problem can now be stated in graph theory terms as follows: How many random edges are needed to simultaneously connect  $x$  graphs of  $m$  vertices each? A combinatorial approach to this problem seems possible but is fairly complicated. Therefore we decided to study the problem via a simulation. Edges were placed randomly and simultaneously in eight graphs of eight vertices each to simulate the sugar cane genome. Each simulation stopped

when all eight graphs were connected. This was repeated ten thousand times, and the number of edges needed to connect recorded for each one (see Figure 2). The mean of the empirical distribution was 152.34, and the variance 1387.89. More importantly, the percentiles indicated that 99% of the graphs will be connected with 250 edges, 95% with 222 edges, and 90% with 202 edges. In practical terms, more than 200 double dose fragments (or their equivalent in other higher dose fragments) may be needed to guarantee connectedness in most cases.

FIGURE 2 GOES HERE

### 3.3 Determination of Linkage

When considering two single dose fragments at a time, three possible situations arise; they could be linked in coupling, linked in repulsion, or unlinked. Linkage in coupling refers to the case in which the two fragments are physically located on the same chromosome. In repulsion the fragments are located on different chromosomes, but these belong to the same homologous group, so there is the possibility of recombination bringing the two fragments together into one chromosome. Unlinked fragments are located on non-homologous chromosomes so they segregate independently. Two single dose fragments linked in coupling segregate as in a diploid organism, so it is possible to use standard methods to test for linkage and to estimate the recombination distance between them.

When testing for linkage between two fragments of which at least one is present in dosage greater than one, it is necessary to consider the configurations of the fragments on the chromosomes of the informative parent. The set of possible configurations will depend on the dosage of the fragments involved. A single dose fragment will be present in one chromosome, a double dose fragment

will be present on two, and so on. However, it is important to consider the linkage phase of the fragments. The nine configurations that will be studied here are presented in Table 2. These include only single, double and triple dose fragments, and only some of the possible configurations for triple dose fragments.

For a given configuration, linkage can be detected by means of a likelihood ratio test as described below.

TABLE 2 GOES HERE

Each individual progeny can be classified into one of four phenotypic classes as determined by the presence or absence of each fragment. We denote the four phenotypic classes by AB, A, B and  $\emptyset$  accordingly. The number of the progeny in each phenotypic class will follow a multinomial distribution, where the phenotype probabilities depend on the configuration of the fragments and their recombination fraction. For each configuration, the four phenotype probabilities will be denoted by  $p_{AB}(r)$ ,  $p_A(r)$ ,  $p_B(r)$ ,  $p_{\emptyset}(r)$ . To compute these probabilities, we must consider:

1. all possible pairings of the  $m$  homologous chromosomes
2. crossing over events within each pairing
3. assortment of chromosomes into gametes
4. assignment of gametes to phenotypic classes.

Phenotype probabilities for autopolyploids of ploidy number  $m$  are listed in the appendix.

### 3.4 Estimation of Recombination Fractions

For each configuration, the phenotype probabilities are used to define the likelihood

$$L(\mathbf{p}(r)|\mathbf{x}) \propto (p_{AB}(r))^{X_{AB}} (p_A(r))^{X_A} (p_B(r))^{X_B} (p_\emptyset(r))^{X_\emptyset} \quad (2)$$

where  $X_{AB}$ ,  $X_A$ ,  $X_B$ , and  $X_\emptyset$  are the observed numbers of offspring in each phenotypic class. For a given configuration, the corresponding phenotype probabilities (presented in the appendix) are substituted into equation 2. A maximum likelihood estimator (MLE) is calculated by solving the likelihood equations, i.e. the equations obtained by setting the first derivative of the natural logarithm of the likelihood equal to zero. The likelihood equations are high degree polynomials with single roots in the interval  $(0, \frac{1}{2})$  and are easily solved by the Newton-Raphson algorithm.

Assuming that the configuration of the two fragments is known, we can calculate the MLE of the recombination fraction,  $\hat{r}$ , and test whether it is significantly different from 0.5 by means of a likelihood ratio test (Casella and Berger, 1990). We are effectively testing the null hypothesis that the fragments are unlinked vs the alternative that the fragments are linked in the given configuration. The likelihood ratio test statistic is given by:

$$\lambda(\mathbf{x}) = -2\log \frac{L(\mathbf{p}(\hat{r})|\mathbf{x})}{L(\mathbf{p}(0.5)|\mathbf{x})} \quad (3)$$

where  $L(\mathbf{p}(\hat{r})|\mathbf{x})$  is the likelihood evaluated at  $r = 0.5$ , the expected value under the null hypothesis. For a multinomial likelihood, this test statistic converges to a chi-square distribution with one degree of freedom, so we declare fragments linked at an  $\alpha$  level if  $\lambda(\mathbf{x}) > \chi_{1,\alpha}^2$ .

Given the dosages of the two fragments, several configurations are possible. In practice, one might test for linkage in each configuration that is consistent

with the dosages. To construct the *S. spontaneum* map, we tested only linkages in coupling for reasons discussed in the next section.

### 3.5 Power and Sample Size

To determine which configurations are the most informative we computed Fisher's information for each. These results for autooctaploids are summarized in Table 3, and the equations are plotted as a function of the recombination fraction in Figure 3. Not surprisingly, the information is generally highest for tightly linked loci and for configurations involving coupling linkages only. Coupling (c11) and double coupling (dc) provide, by far, the most information, followed by the asymmetric couplings (ac12, ac23 and ac13 in that order). This is because fragments linked in coupling can always undergo recombination whereas fragments linked in repulsion must first pair. Furthermore, fragments linked in repulsion segregate independently of other fragments and their presence can obscure recombination events that might have happened between the coupled fragments.

TABLE 3 GOES HERE

Two different scales were used in Figure 3 to illustrate how little information the configurations involving repulsion provide. Note that, in the case of coupling and repulsion (c&r), the information actually increases as the distance between the loci increases. Two loci that are closely linked will recombine very infrequently, so their alleles will appear together in most of the progeny. If we add an extra copy of each of the fragments (which will segregate independently of each other and of the linked fragments in most cases) then the fragments will appear together in the progeny just as often as they appear separately. The extra fragments will conceal the fact that the coupled markers are very

closely linked. As the distance between the two coupled fragments increases, the recombinations between them alter the equal segregation pattern enough for a slight increase in information to be detectable.

FIGURE 3 GOES HERE

When the null hypothesis that the two loci are unlinked is false, the likelihood ratio test statistic  $\lambda(\mathbf{x})$  has an approximate large-sample noncentral chi-square distribution with noncentrality parameter equal to:

$$c = N \sum \frac{[p_i(0.5) - p_i(\hat{r})]^2}{p_i(\hat{r})} \quad (4)$$

where  $N$  is the multinomial sample size (Agresti, 1990). Values of central and noncentral chi-square distributions were used to compute the power of our  $\alpha = 0.01$  tests, with  $N = 80$ , for different configurations, as determined by  $P(\lambda(\mathbf{x}) > \chi_{1,0.01}^2)$ . These power curves appear in Figure 4. The power to detect linkages drops as the recombination fraction increases. This drop in power is slower for coupling (c11) double coupling (dc) and asymmetric coupling 1-2 (ac12), the same configurations that give us the most information. The power never exceeds 35% for any configuration involving repulsion, not even with very tight linkages. For asymmetric coupling 2-3 and 1-3, the power is reasonably good for small recombination fractions but drops rapidly as  $r$  increases.

FIGURE 4 GOES HERE

Similar computations can be carried out to determine the sample sizes needed to detect linkage with power 80% for different configurations. Results are summarized in Figure 5.

FIGURE 5 GOES HERE

## 4 The Sugar Cane Map

To construct the sugar cane map, the 287 markers for which dosage was determined were tested for pairwise linkage. Based on the information and power finding in the previous section, configurations involving repulsion were ignored. Thus, every pair of single dose markers was tested for coupling (c11), every pair of doubles tested for double coupling (dc), single markers were tested against doubles for linkage in asymmetric coupling 1-2 and against the triple in asymmetric coupling 1-3, and doubles against the triple in asymmetric coupling 2-3.

**Linkage Groups** A framework map of 233 single dose markers was provided by da Silva (1993). Single dose markers in coupling behave the same in diploids and autopolyploids. Thus it is possible to map them using, for example, MAPMAKER multipoint linkage analyses (Lander et al., 1987). The eleven remaining single dose markers were added to the framework map using the linkages and distances suggested by the likelihood ratio test approach. The resulting 54 linkage groups provide a framework map into which we can now map the higher dose markers. Higher dose markers are expected to provide connections between linkage groups to form homologous groups.

**Homologous Groups** Markers are expected to map to as many linkage groups as their dosage suggests. Linkage groups containing the same multiple dose markers are assumed to be homologous (see assumption A6). To illustrate the process of mapping a new marker into the framework map, consider the double dose marker UMC058a. Analysis in asymmetric coupling 1-2 reports it as being linked to markers SG029b, SG212a, UMC128b, SG155a, CDSR078a, and CDSB009b. The first three all map to homologous group III, two of them

into linkage group 8 and the other one into linkage group 17. The last three markers, however, map into homologous group V, IV, and VIII respectively. No linkages were reported in double coupling or asymmetric coupling 2-3. Based on this information, UMC058 was assigned to homologous group III, linkage groups 8 and 17. The other three linkages are considered type I errors, or the result of some violation of the model assumptions.

**Highly Polymorphic RFLP Markers** Another method of finding homologous chromosomes is by mapping multiple RFLP fragments from the same probe (da Silva, 1993). In diploids, RFLP bands can be taken in pairs and analyzed as co-dominant markers. There are two homologous chromosomes of each kind, so two alleles can be visualized in the banding pattern. In polyploids, more homologous chromosomes are present, so there are more opportunities for variability in the alleles. Because it is hard to be certain of the relationship among the bands, each one was scored separately as a dominant marker. It is very possible, though, that all or most of the bands represent different alleles at the same locus that are present in different homologs.

These highly polymorphic probes are easily identifiable by the same library name and probe number, with an additional letter at the end to signify the fragment within the probe. Connections between linkage groups were made on the basis of these highly polymorphic probes (da Silva 1993). A total of 41 of these markers resulted in 29 different connections, some of which were redundant. These connections are not as reliable as those provided by multidoses fragments, so they should be verified with higher dose fragments whenever possible. For example, fragments a, b and c identified by probe CDSR065 map to chromosomes 40, 9 and 29 respectively, all of which belong to homologous group I, while fragment d of the same probe maps to chromosome

4 in homologous group III. The first three homologies are confirmed by double dose fragments of other highly polymorphic probes. Fragment d is firmly established in homologous group III, as three other highly polymorphic probes connect chromosome 4 to others in this group.

The map of the sugar cane genome is presented in Figure 6. Using the higher dose markers and the highly polymorphic probes, the 54 linkage groups were assembled into 10 homologous groups, some of them containing less than 8 linkage groups. Several markers remain unlinked: 7 of ambiguous dose, 23 single dose, 11 double dose that mapped into only one linkage group, and 1 triple dose that mapped into 2 linkage groups.

FIGURE 6 GOES HERE

## 5 Conclusions

The statistical methodology described here for genetic mapping in autopolyploids is fully extendable to organisms of other ploidy levels. Phenotype probabilities for general autopolyploids are listed in the appendix. Given the dosages and configuration, these probabilities can be used in a multinomial model to find maximum likelihood estimates of the recombination fractions between two loci and to test for linkage with an asymptotic likelihood ratio test.

Configurations involving only linkage in coupling proved to be much more informative than those involving linkages in repulsion. Power and sample size calculations support this conclusion. Linkage maps in autopolyploids, then, will most likely be based in linkages in coupling, unless thousands of offspring are available.

The construction of the sugar cane map reveals several areas in which difficulties may be encountered. The dosage of each fragment must be determined

accurately, since the subsequent analysis is based on the dosage classification. Expected segregation ratios for higher dosages that are very similar to one another, random deviations from the expected segregation ratios, and segregation distortion, might preclude the classification of all markers. Markers of ambiguous dosage are best left out of any linkage analysis. The different approaches discussed here to determine dosage attempt to corroborate the dosages of the markers, and are used to eliminate from the analysis any markers that do not show consistent results.

In constructing the map itself, several thousand tests are performed. Even when using a strict  $\alpha$  level such as 0.01, many type I errors will undoubtedly occur. Some of these are discovered through inconsistencies in the map, and these are considered on an individual basis to determine the best placement of each marker involved. Others might have gone unnoticed, but will probably be detected as more markers are added to this map.

It is clear that several studies will be needed to be combined to achieve a complete map of sugar cane. Of the 64 expected chromosomes arranged into 8 homologous groups of 8 chromosomes each, the data provided 54 chromosomes arranged into 10 homologous groups, some containing less than eight chromosomes. The results from the Monte Carlo studies on completeness of the map suggest that approximately 200 connections might be necessary to complete the map of sugar cane. In the data set used here we found 30 connections provided by double dose fragments, 1 provided by the triple dose fragment, and 41 provided by highly polymorphic probes. As expected, many of these connections are redundant. A complete map of sugar cane can be achieved by mapping additional markers in this population and by combining the results of this study with those of independent studies on the same or related varieties of sugar cane.

## REFERENCES

- Agresti, A. (1990). *Categorical Data Analysis*. New York: Wiley.
- Bingham, E.T., Burnham, C.R. and Gates, C.E. (1968). Double and single backcross linkage estimates in autotetraploid maize. *Genetics* **59**:399.
- Bonierbale, M.W., R.L. Plaisted, and S.D. Tanksley (1988). RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. *Genetics* **120**: 1095-1103.
- Casella, G. and Berger, R. (1990). *Statistical Inference*. Pacific Grove, California: Wadsworth and Brooks.
- Chartrand, G., and Lesniack, L. (1979). *Graphs and Diagraphs*. Pacific Grove, California: Wadsworth and Brooks.
- da Silva, J.A.G. (1993). *A Methodology for Genome Mapping of Autopolyploids and its Application to Sugarcane (Saccharum Spp.)* Ph.D. thesis, Plant Breeding Dept., Cornell University.
- da Silva, J. and M.E. Sorrells (1996). Linkage analysis in polyploids using molecular markers. *In*: Jauhar, P. (ed.) *Methods of Genome Analysis in Plants: Their Merits and Pitfalls*. CRC Press Inc.
- da Silva, J.A.G. and Sorrells, M.E. (1995) "Linkage Analysis in Polyploids Using Molecular Markers"

da Silva, J., M.E. Sorrells, W.L. Burnquist, and S.D. Tanksley (1993). RFLP linkage map and genome analysis of *Saccharum spontaneum*. *Genome* **36**: 782-791.

de Winton, D., and Haldane, J.B.S. (1932). "Linkage in the Tetraploid *Primula Sinensis*". *Journal of Genetics* **24**: 121-144.

Fisher, R.A. (1943). "Allowance for Double Reduction in the Calculation of Genotype Frequencies with Polysomic Inheritance". *Philosophical Transactions of the Royal Statistical Society, Series B*, **233**: 55-87.

Geiringer, H. (1949). "Chromatid Segregation of Tetraploids and Hexaploids". *Genetics* **34**: 665-684.

Griffiths, A.J.F., Miller, J.H., Suzuki, D.T., Lewontin, R.C. and Gelbart, W.M. (1993). *An Introduction to Genetic Analysis, 5th edition*. New York: W.H. Freeman and Company.

Haldane, J.B.S. (1919). "The Combination of Linkage Values and the Calculation of Distances Between Loci of Linked Factors". *Journal of Genetics* **8**: 299-309.

Haldane, J.B.S. (1930). "Theoretical Genetics in Autopolyploids". *Journal of Genetics* **22**: 359-372.

Haldane, J.B.S., and Waddington, C.H. (1930). "Inbreeding and Linkage".

*Genetics* **16**: 357-374.

Kam-Morgan, L.N.W., B.S. Gill, and S. Muthukrishnan (1989). DNA restriction fragment length polymorphisms: a strategy for genetic mapping of D genome of wheat. *Genome* **32**: 724-732.

Lander, E.S., and Green, P. (1987). "Construction of Multi-Locus Genetic Linkage Maps in Humans". *Proceedings of the National Academy of Science USA*. **84**: 2363-2367.

Lander, E.S., Green, P., Abrahamson, A., Barlow, M., Daley, M., Lincoln, S., Newburg, L. (1987). "MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations". *Genomics* **1**:174-181.

Mather, K. (1936). "Segregation and Linkage in Autotetraploids". *Journal of Genetics* **32**: 287-314.

O'Donoghue, L.S. et al. (1992). An RFLP-based map of oat on a cross between two diploid taxa (*Avena atlantica* x *A. hirtula*). *Genome* **35**: 765-771.

Ripol, M.I. (1994) Statistical Aspects of Genetic Mapping in Autopolyploids. Masters Thesis, Cornell University.

Wu, K.K., Burnquist, W., Sorrells, M.E., Tew, T.L., Moore, P.H. and Tanksley, S.D. (1992). "The Detection and Estimation of Linkage in Polyploids Using Single-Dose Restriction Fragments". *Theoretical and Applied Genetics* **83**:

294-300.

# Appendix 1

## COUPLING

Phenotype Probabilities

$AB$	$\frac{1}{2}(l-r)$
$A$	$\frac{1}{2}r$
$B$	$\frac{1}{2}r$
$\emptyset$	$\frac{1}{2}(l-r)$

## REPULSION

Phenotype Probabilities

$AB$	$\frac{1}{4}\left(\frac{m}{m-1}\right) + \frac{1}{2}\left(\frac{1}{m-1}\right)r$
$A$	$\frac{1}{4}\left(\frac{m}{m-1}\right) - \frac{1}{2}\left(\frac{1}{m-1}\right)r$
$B$	$\frac{1}{4}\left(\frac{m}{m-1}\right) - \frac{1}{2}\left(\frac{1}{m-1}\right)r$
$\emptyset$	$\frac{1}{4}\left(\frac{m}{m-1}\right) + \frac{1}{2}\left(\frac{1}{m-1}\right)r$

## ASYMMETRIC COUPLING 1-2

Phenotype Probabilities

$AB$	$\frac{1}{2} - \frac{1}{4}\left(\frac{m-2}{m-1}\right)r$
$A$	$\frac{1}{4}\left(\frac{m-2}{m-1}\right)r$
$B$	$\frac{1}{4}\left(\frac{m}{m-1}\right) + \frac{1}{4}\left(\frac{m-2}{m-1}\right)r$
$\emptyset$	$\frac{1}{4}\left(\frac{m-2}{m-1}\right) - \frac{1}{4}\left(\frac{1-2}{m-1}\right)r$

## REPULSION 1-2

Phenotype	Probabilities
$AB$	$\frac{1}{8} \left( \frac{3m-4}{m-1} \right) - \frac{1}{2} \left( \frac{1}{m-1} \right) r$
$A$	$\frac{1}{8} \left( \frac{m}{m-1} \right) - \frac{1}{2} \left( \frac{1}{m-1} \right) r$
$B$	$\frac{1}{8} \left( \frac{3m}{m-1} \right) + \frac{1}{2} \left( \frac{1}{m-1} \right) r$
$\emptyset$	$\frac{1}{8} \left( \frac{m-4}{m-1} \right) + \frac{1}{2} \left( \frac{1}{m-1} \right) r$

## DOUBLE COUPLING

Phenotype	Probabilities
$AB$	$\frac{1}{4} \left( \frac{3m-2}{m-1} \right) - \frac{1}{2} \left( \frac{m-2}{m-1} \right) r + \frac{1}{4} \left( \frac{m-2}{m-1} \right) r^2$
$A$	$\frac{1}{2} \left( \frac{m-2}{m-1} \right) r - \frac{1}{4} \left( \frac{m-2}{m-1} \right) r^2$
$B$	$\frac{1}{2} \left( \frac{m-2}{m-1} \right) r - \frac{1}{4} \left( \frac{m-2}{m-1} \right) r^2$
$\emptyset$	$\frac{1}{4} \left( \frac{m-2}{m-1} \right) - \frac{1}{2} \left( \frac{m-2}{m-1} \right) r + \frac{1}{4} \left( \frac{m-2}{m-1} \right) r^2$

## COUPLING AND REPULSION

Phenotype	Probabilities
$AB$	$\frac{1}{8} \left( \frac{5m-4}{m-1} \right) - \frac{1}{8} \left( \frac{m-6}{m-1} \right) r - \frac{1}{4} \left( \frac{1}{m-1} \right) r^2$
$A$	$\frac{1}{8} \left( \frac{m}{m-1} \right) + \frac{1}{8} \left( \frac{m-6}{m-1} \right) r + \frac{1}{4} \left( \frac{1}{m-1} \right) r^2$
$B$	$\frac{1}{8} \left( \frac{m}{m-1} \right) + \frac{1}{8} \left( \frac{m-6}{m-1} \right) r + \frac{1}{4} \left( \frac{1}{m-1} \right) r^2$
$\emptyset$	$\frac{1}{8} \left( \frac{m-4}{m-1} \right) - \frac{1}{8} \left( \frac{m-6}{m-1} \right) r - \frac{1}{4} \left( \frac{1}{m-1} \right) r^2$

## DOUBLE REPULSION

Phenotype	Probabilities
$AB$	$\frac{1}{16} \left( \frac{9m^2-34m+24}{(m-1)(m-3)} \right) + \frac{1}{2} \left( \frac{m-4}{(m-1)(m-3)} \right) r + \frac{1}{2} \left( \frac{1}{(m-1)(m-3)} \right) r^2$
$A$	$\frac{1}{16} \left( \frac{3m^2-10m}{(m-1)(m-3)} \right) - \frac{1}{2} \left( \frac{m-4}{(m-1)(m-3)} \right) r - \frac{1}{2} \left( \frac{1}{(m-1)(m-3)} \right) r^2$
$B$	$\frac{1}{16} \left( \frac{3m^2-10m}{(m-1)(m-3)} \right) - \frac{1}{2} \left( \frac{m-4}{(m-1)(m-3)} \right) r - \frac{1}{2} \left( \frac{1}{(m-1)(m-3)} \right) r^2$
$\emptyset$	$\frac{1}{16} \left( \frac{m^2-10m+24}{(m-1)(m-3)} \right) + \frac{1}{2} \left( \frac{m-4}{(m-1)(m-3)} \right) r + \frac{1}{2} \left( \frac{1}{(m-1)(m-3)} \right) r^2$

## ASYMMETRIC COUPLING 1-3

Phenotype	Probabilities
$AB$	$\frac{1}{2} - \frac{1}{8} \left( \frac{m-4}{m-1} \right) r$
$A$	$\frac{1}{8} \left( \frac{m-4}{m-1} \right) r$
$B$	$\frac{3}{8} \left( \frac{m}{m-1} \right) + \frac{1}{8} \left( \frac{m-4}{m-1} \right) r$
$\emptyset$	$\frac{1}{8} \left( \frac{m-4}{m-1} \right) - \frac{1}{8} \left( \frac{m-4}{m-1} \right) r$

### ASYMMETRIC COUPLING 2-3

Phenotype	Probabilities
$AB$	$\frac{1}{4} \left( \frac{3m-2}{m-1} - \frac{1}{4} \right) r + \frac{1}{8} \left( \frac{m-4}{m-1} \right) r^2$
$A$	$\frac{1}{4} \left( \frac{m-4}{m-1} \right) r - \frac{1}{8} \left( \frac{m-4}{m-1} \right) r^2$
$B$	$\frac{1}{8} \left( \frac{m}{m-1} \right) + \frac{1}{4} \left( \frac{m-4}{m-1} \right) r - \frac{1}{8} \left( \frac{m-4}{m-1} \right) r^2$
$\emptyset$	$\frac{1}{8} \left( \frac{m-4}{m-1} \right) - \frac{1}{4} \left( \frac{m-4}{m-1} \right) r + \frac{1}{8} \left( \frac{m-4}{m-1} \right) r^2$

Table 1: Segregation Ratios in Autopolyploids

	diploid	tetraploid	hexaploid	octaploid	dexaploid
k	m=2	m=4	m=6	m=8	m=10
1	1/2	1/2	1/2	1/2	1/2
2	1	5/6	4/5	11/14	7/9
3		1	19/20	13/14	11/12
4		1	1	69/70	41/42
5			1	1	251/252

Table 2: Genotypic Configurations

name		dose(A)	dose(B)	notation
Coupling	c11	1	1	$(AB)(ab)_{m-1}$
Repulsion	r11	1	1	$(Ab)(aB)(ab)_{m-2}$
Asymmetric Coupling 1-2	ac12	1	2	$(AB)(aB)(ab)_{m-3}$
Repulsion 1-2	r12	1	2	$(Ab)(aB)_2(ab)_{m-3}$
Double Coupling	dc	2	2	$(AB)_2(ab)_{m-2}$
Coupling and Repulsion	c&r	2	2	$(AB)(Ab)(aB)(ab)_{m-3}$
Double Repulsion	dr	2	2	$(Ab)_2(aB)_2(ab)_{m-4}$
Asymmetric Coupling 1-3	ac13	1	3	$(AB)(aB)_2(ab)_{m-3}$
Asymmetric Coupling 2-3	ac23	2	3	$(AB)_2(aB)(ab)_{m-3}$

Table 3: Fisher's Information for  $r$  in Autooctaploids

c11	$\frac{1}{r-r^2}$
r11	$\frac{1}{12+r-r^2}$
ac12	$\frac{3(2+3r-3r^2)}{28r-19r^2-18r^3+9r^4}$
r12	$\frac{8+r-r^2}{(r-6)(r-2)(r+1)(r+5)}$
dc	$\frac{12(11-14r+7r^2)}{7r(22-23r+12r^2-3r^3)}$
c&r	$\frac{(1+2r^2)(38-7r-7r^2)}{7(-18+r+r^2)(-2+r+r^2)(4+r+r^2)}$
dr	$\frac{4(r+2)^2(7r^2+28r+67)}{7(14-4r-r^2)(1+4r+r^2)(41+4r+r^2)}$
ac13	$\frac{3+r-r^2}{42r-41r^2-2r^3+r^4}$
ac23	$\frac{4(11+22r-39r^2+28r^3-7r^4)}{7r(44+14r-44r^2+21r^3-6r^4+r^5)}$

**Figure 1** Observed segregation ratios for the 294 fragments used in the *S. spontaneum* study.

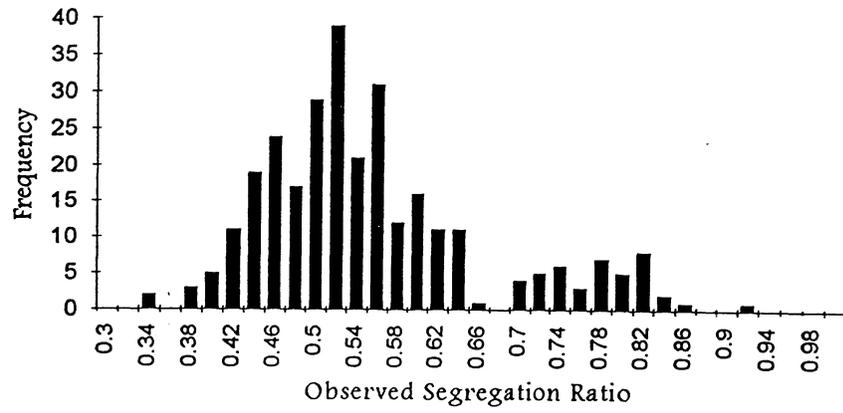
**Figure 2** Numbers of edges needed to connect eight graphs of eight vertices each based on 100 simulations. In eight of the simulations, the number of edges exceeded 350.

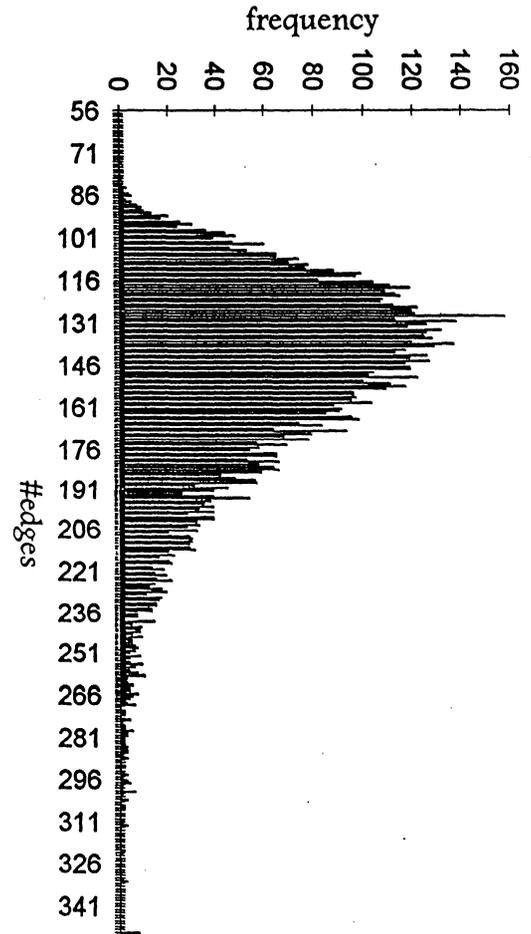
**Figure 3** Fisher's information for linkage under different configurations (c11-coupling, dc-double coupling, ac12-asymmetric coupling 1-2, ac23, asymmetric coupling 2-3, ac13-asymmetric coupling 1-3, dr-double repulsion, r12-repulsion 1-2, r11-repulsion, c&r-coupling and repulsion) shown as a function of the recombination fraction.

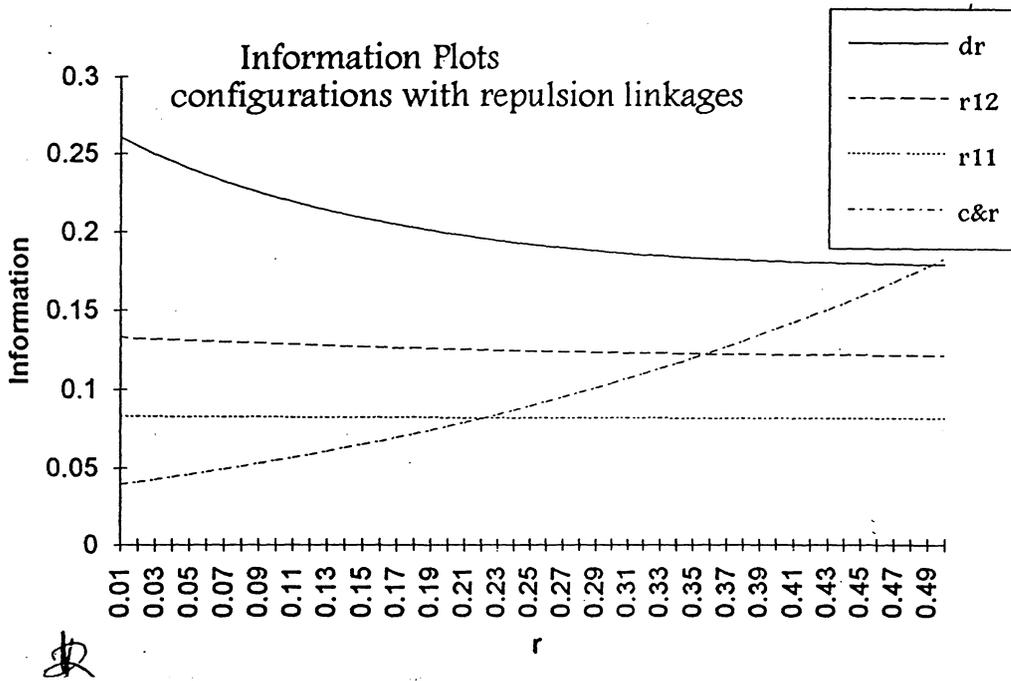
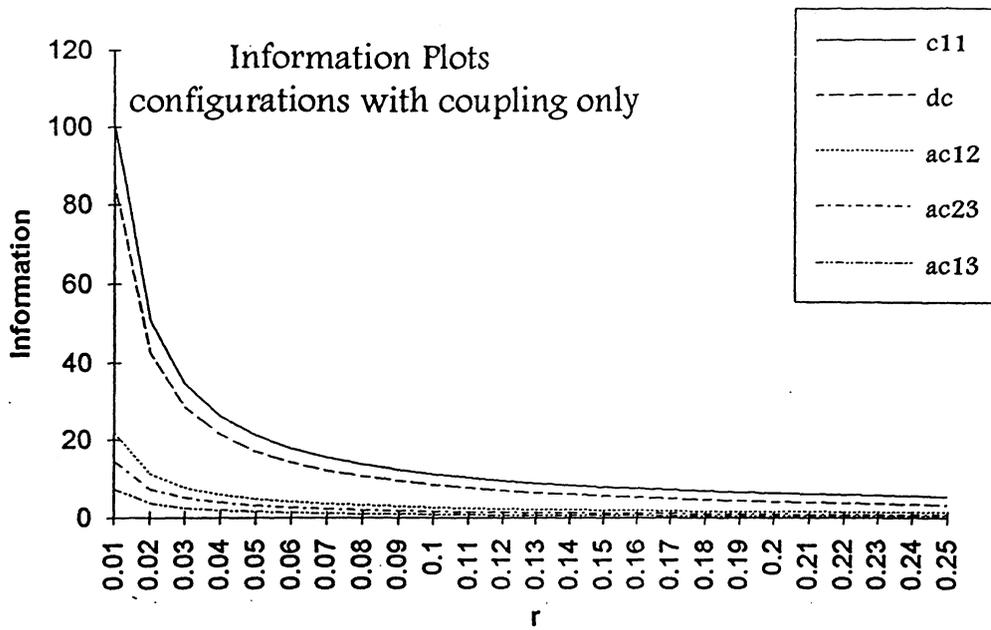
**Figure 4** Power to detect linkage with a sample size of 80 progeny for different configurations shown as a function of the recombination fraction.

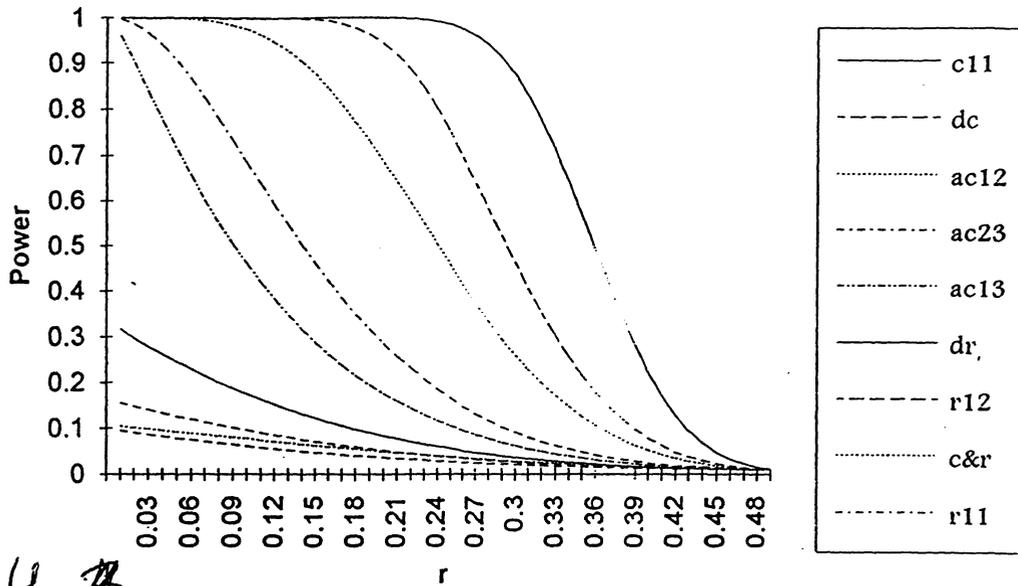
**Figure 5** Sample sizes needed to achieve 80% power for linkage detection for different configurations shown as a function of the recombination fraction.

**Figure 6** A linkage map of *S. spontaneum*.

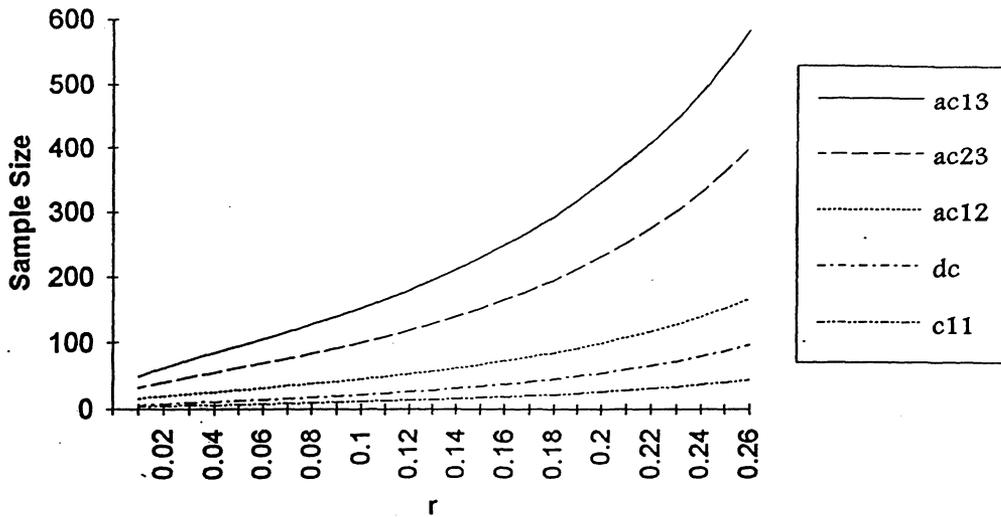








Sample Size Needed for 80% Power configurations involving coupling only



Sample Size needed for 80% Power configurations involving repulsion

