ORIGINAL PAPER

M. Frisch · M. Quint · T. Lübberstedt ·
A. E. Melchinger

Duplicate marker loci can result in incorrect locus orders on linkage maps

Abstract Genetic linkage maps, constructed from multi-locus recombination data, are the basis for many applications of molecular markers. For the successful employment of a linkage map, it is essential that the linear order of loci on a chromosome is correct. The objectives of this theoretical study were to (1) investigate the occurrence of incorrect locus orders caused by duplicate marker loci, (2) develop a statistical test for the detection of duplicate markers, and (3) discuss the implications for practical applications of linkage maps. We derived conditions, under which incorrect locus orders do or do not occur with duplicate marker loci for the general case of n markers on a chromosome in a BC_1 mapping population. We further illustrated these conditions numerically for the special case of four markers. On the basis of the extent of segregation distortion, an exact test for the presence of duplicate marker loci was suggested and its power was investigated numerically. Incorrect locus orders caused by duplicate marker loci can (1) negatively affect the assignment of target genes to chromosome regions in a map-based cloning experiment, (2) hinder indirect selection for a favorable allele at a quantitative trait locus, and (3) decrease the efficiency of reducing the length of the chromosome segment attached to a target gene in marker-assisted backcrossing.

M. Frisch · M. Quint · T. Lübberstedt · A. E. Melchinger (✉)
Institute of Plant Breeding, Seed Science and Population Genetics,
University of Hohenheim,
70593 Stuttgart, Germany
e-mail: melchinger@uni-hohenheim.de
Fax: +49-711-4592343

Present address:
T. Lübberstedt, Danish Institute of Agricultural Sciences,
Research Centre Flakkebjerg,
4200 Slagelse, Denmark

Introduction

Genetic linkage maps are constructed from observed recombination frequencies between loci in experimental or natural populations with known pedigrees. They are an essential tool for practical applications such as marker-assisted selection, marker-assisted backcrossing, and map-based cloning of target genes. For these applications a correct linear order of loci within linkage groups is essential. Errors in locus order can seriously hamper the ability to map, isolate, or select for simple Mendelian and complex traits.

Duplications of chromosome regions occur frequently and seem to be an important mechanism of genome evolution (Ohno 1970). More than one third of a typical eukaryotic genome consists of duplicated genes and gene families. Such gene families can arise from polyploidization events such as those presumed to have preceded the origin of many plant species (Wendel 2000).

The portion of genes in the different model organisms concerning their presence as singletons or duplicate members of gene families is very variable. Tandem gene duplications appear to be ubiquitous in plant genomes (Acarkan et al. 2000; Tarchini et al. 2000). The complete genome sequence of *Arabidopsis* has revealed that an estimated 17% of the 25,000 genes is arranged in tandemly repeated segments (The Arabidopsis Genome Initiative 2000). For the monocot model organism rice, this portion of locally duplicated genes accounts for an estimated 22% of the approximately 30,000 genes available from the draft sequence (Goff et al. 2002). For both species, it has been recognized that 60% of the genome is contained within large duplicated segments (Blanc et al. 2000; Goff et al. 2002), with almost half of the *Arabidopsis* genes within the duplicated segments being conserved.

If a duplicate chromosome region contains a DNA sequence, which can be used as a molecular marker, the marker alleles at the two duplicate marker loci cannot be distinguished. Equal fragment length results in an equal banding pattern, and consequently, the alleles of duplicate

markers are scored in a mapping population as the alleles of one single marker. The recombination frequency between this non-existing 'ghost marker' and non-duplicated markers are different from those between the non-duplicated markers and the duplicate marker loci actually underlying the ghost marker. Since the locus order of linkage groups is determined on the basis of recombination frequencies between loci, incorrect recombination values for a linkage group can result in an incorrect locus order for the chromosome.

We encountered this phenomenon in a study with resistance gene analogues (RGAs, Quint et al. 2003). More generally, it can also be found for marker systems based on polymorphisms in short sequence stretches such as amplified fragment length polymorphism (AFLP) markers. Vuylsteke et al. (1999) conducted comprehensive AFLP mapping in two maize populations. From more than 1,000 markers mapped in the individual populations, 353 AFLP markers were in common, i.e., a given AFLP primer combination resulted in polymorphic AFLP fragments of identical size in both populations. 327 of these 353 AFLP markers (>92%) were considered as co-linear between both populations. However, the remaining 26 common AFLP markers (7.4%) mapped to different chromosomes in both populations. Thirteen of the respective AFLP fragments were sequenced. For three of these bands, sequences were (almost) identical, whereas for the other 10 bands sequence identity was restricted to restriction sites and selective nucleotides employed in the AFLP assay.

The term ghost marker was coined in analogy to the ghost QTL phenomenon (Martinez and Curnow 1992). The effect of such ghost markers on the construction of linkage maps and the consequences for marker-assisted selection as well as map-based cloning of target genes has not yet been investigated.

The objectives of our study were to (1) derive the recombination frequency between a marker and a ghost marker, (2) derive conditions under which duplicate markers result in an incorrect locus order of the respective linkage group, (3) investigate under which conditions the correct locus order is found, even if there are duplicate markers in the linkage group, (4) develop a test for detection of duplicate markers, and (5) discuss the consequences of duplicate markers for applications of linkage maps such as map-based cloning, marker-assisted selection and marker-assisted backcrossing.

Theory

Definitions

Assuming (1) a diploid species and (2) two duplicate marker loci carrying alleles which cannot be distinguished by the laboratory method used for the molecular marker analysis, the four alleles at the two duplicate markers are scored as the alleles of only one marker, which we call a 'ghost marker'. Segregation ratios, recombination frequencies with other loci, and the map position of a ghost marker are in general not identical with the corresponding parameters for the underlying duplicate markers.

Non-duplicate markers, of which the alleles can be distinguished by the laboratory method used for the molecular marker analysis, are referred to as 'distinguishable markers'. The term marker is used in the sense of distinguishable marker, when there is no further specification as a ghost marker or a duplicate marker.

An incorrect locus order is defined as an order which cannot be obtained by omitting loci from the correct locus order of all loci on the chromosome. In this study, we focus on incorrect locus orders for which a ghost marker maps to a chromosome interval, in which none of the underlying duplicate markers is located.

Notation

Consider a chromosome with n distinguishable marker loci at positions $k_1,..,k_n$. The positions are measured in map distance from the beginning of the chromosome, and

$k_u < k_{u+1}$ for $u \in \{1, \ldots, n-1\}$.

In addition we define the telomere map positions as k_0 and k_{n+1}.

We consider two duplicate markers, located at positions $i_1 < i_2$. The indices of the map positions k_u, which are located next to the duplicate markers and have a smaller map position than these, are denoted with x and y, respectively:

$x = \max(u|u \in \{0, \ldots, n\}, k_u < i_1)$

$y = \max(u|u \in \{0, \ldots, n\}, k_u < i_2)$

The map position of the ghost marker, resulting from linkage analysis involving the two duplicate loci i_1 and i_2, is denoted by i. The index of the map position k_u located next to the ghost marker and having a smaller map position than it is denoted with z:

$z = \max(u|u \in \{0, \ldots, n\}, k_u < i)$

Without loss of generality we assume $z \leq n-z$.

In this notation, a correct locus order is characterized by

$z = x$ or $z = y$,

whereas an incorrect locus order is characterized by

$z \neq x$ and $z \neq y$.

Assumptions and basic results

For our derivations, we assume no interference in crossover formation (Stam 1979). Under this assumption, crossover formation in adjacent marker intervals is stochastically independent, and the recombination frequency ρ between two loci is related to the respective map distance d by Haldane's (1919) mapping function

$\rho = (1 - e^{-2d})/2$.

Linkage between two loci is measured by the linkage value (Schnell 1961)

$$\lambda = 1 - 2\rho = e^{-2d}. \tag{1}$$

Linkage values between distinguishable markers at positions k_u and k_v are denoted with $\lambda_{u,v}$, those between a distinguishable marker at position k_u and the duplicate markers with $\lambda_{u,i1}$ and $\lambda_{u,i2}$, respectively, and linkage between a distinguishable marker at position k_u and the ghost marker with $\lambda_{u,i}$. For sake of convenience in the subsequent derivations, in which linkage values are summed over marker intervals, we define the linkage between the telomere and the first locus next to the telomere to be zero:

$\lambda_{0,1} = 0$ if $x > 0$ $\lambda_{n,n+1} = 0$ if $y < n$

$\lambda_{0,i1} = 0$ if $x = 0$ $\lambda_{i2,n+1} = 0$ if $y = n$

$\lambda_{0,i} = 0$ if $z = 0$ $\lambda_{i,n+1} = 0$ if $z = n$

Fig. 1 Genotypes of a BC$_1$ population with respect to two duplicate markers i_1,i_2 and a marker k_u. For each multi-locus genotype, its frequency f and the bands scored for ghost marker i and marker k_u are given (assuming codominant inheritance). The genotypes which show recombination between the ghost marker i and marker k_u are listed as are the genotypes which were scored as heterozygous with respect to the ghost marker i
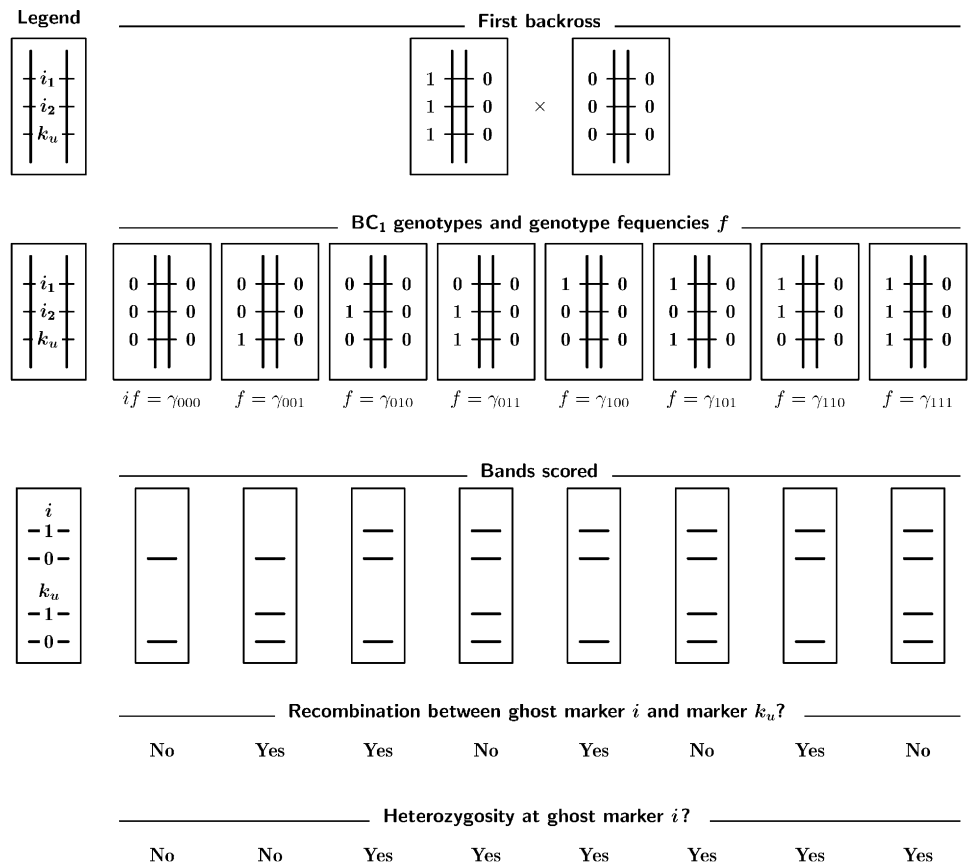
Using the stochastic independence of crossover formation in adjacent marker intervals delimited by the loci at positions $k_u<k_v<k_w$, it can be shown that

$$\lambda_{u,w} = \lambda_{u,v}\lambda_{v,w} \tag{2}$$

and, because $\lambda<1$,

$$\lambda_{u,w} < \lambda_{u,v}. \tag{3}$$

Another property used in the subsequent derivations is

$$\lambda_{u,v} = \lambda_{v,u}. \tag{4}$$

Linkage between a marker and a ghost marker

In this section we use the results of Schnell (1961) and therefore adopt his notation. A haplotype is denoted by a sequence of digits, where each digit corresponds to the origin of the allele at a certain locus. In this sequence, the digits 0 and 1 are used to denote that an allele is of maternal or paternal origin, respectively. The probability that an individual transmits a certain gamete to its progenies is denoted by γ, which is indexed with a sequence of digits describing its haplotype.

We consider three linked loci at positions i_1,i_2,k_u on a chromosome, where the loci at positions i_1 and i_2 are duplicate marker loci. We further assume a BC$_1$ mapping population $\frac{111}{000} \times \frac{000}{000}$ of indefinite population size and codominant markers. Since the alleles at i_1 and i_2 cannot be distinguished, the BC$_1$ genotypes $\frac{001}{000}, \frac{010}{000}, \frac{100}{000}, \frac{110}{000}$ with respect to i_1,i_2,k_u are scored as recombinant individuals with respect to the ghost marker at position i and the marker at position k_u (Fig. 1).

Applying Eq. 4 of Schnell (1961) yields the recombination frequency between the ghost marker at position i and the marker at position k_u as

$$\begin{aligned}
\rho_{i,u} &= \gamma_{001} + \gamma_{010} + \gamma_{100} + \gamma_{110} \\
&= \big(1 + \lambda_{i1,i2} - \lambda_{i1,u} - \lambda_{i2,u} \\
&\quad +1 - \lambda_{i1,i2} + \lambda_{i1,u} - \lambda_{i2,u} \\
&\quad +1 - \lambda_{i1,i2} - \lambda_{i1,u} + \lambda_{i2,u} \\
&\quad +1 + \lambda_{i1,i2} - \lambda_{i1,u} - \lambda_{i2,u}\big)/8 \\
&= 1/2 - 1/4\lambda_{i1,u} - 1/4\lambda_{i2,u} \\
&= \big(\rho_{i1,u} + \rho_{i2,u}\big)/2 \tag{5}
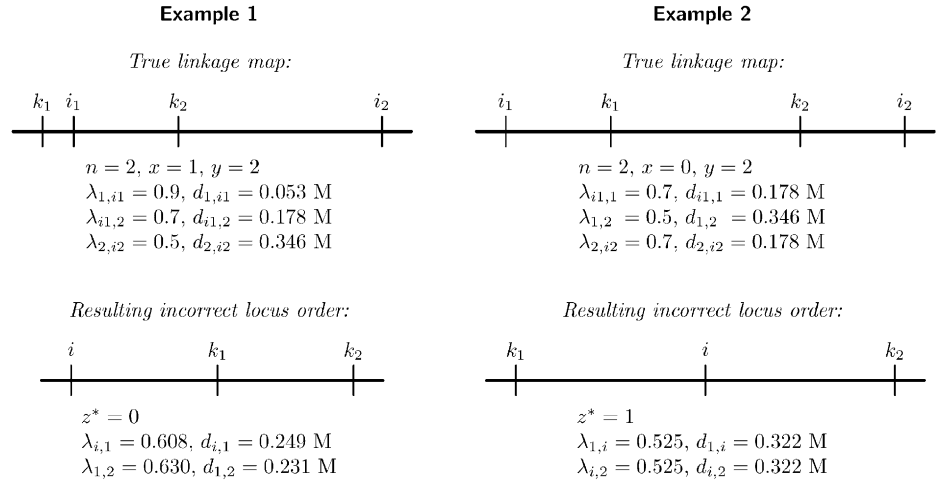\end{aligned}$$

and consequently (Eq. 1)

$$\lambda_{i,u} = \big(\lambda_{i1,u} + \lambda_{i2,u}\big)/2. \tag{6}$$

Note that Eq. 5 is valid for any linear order of the loci i_1,i_2,k_u on the chromosome, irrespective of the applied mapping function and additional loci on the chromosome (Schnell 1961).

The SAR criterion

In the following, we use the 'sum of adjacent recombination frequencies' (SAR) criterion for locus ordering on multi-locus linkage maps and therefore briefly describe its properties. When applying the SAR criterion, the ordering of a multi-locus linkage map is done in two steps. First, the pairwise recombination frequencies between all loci of the linkage group are calculated. Second, the locus order is searched, which minimizes the sum of recombination frequencies ρ between adjacent loci on the linkage map. According to Eq. 1 this

Fig. 2 Two examples for incorrect locus orders resulting from duplicate marker loci i_1 and i_2. For descriptions of the variable names see text

Example 1

True linkage map:

$$k_1 \quad i_1 \qquad k_2 \qquad\qquad\qquad i_2$$

$n = 2,\ x = 1,\ y = 2$
$\lambda_{1,i1} = 0.9,\ d_{1,i1} = 0.053\ \text{M}$
$\lambda_{i1,2} = 0.7,\ d_{i1,2} = 0.178\ \text{M}$
$\lambda_{2,i2} = 0.5,\ d_{2,i2} = 0.346\ \text{M}$

Resulting incorrect locus order:

$$i \qquad\qquad k_1 \qquad\qquad k_2$$

$z^* = 0$
$\lambda_{i,1} = 0.608,\ d_{i,1} = 0.249\ \text{M}$
$\lambda_{1,2} = 0.630,\ d_{1,2} = 0.231\ \text{M}$

Example 2

True linkage map:

$$i_1 \qquad k_1 \qquad\qquad k_2 \qquad\qquad i_2$$

$n = 2,\ x = 0,\ y = 2$
$\lambda_{i1,1} = 0.7,\ d_{i1,1} = 0.178\ \text{M}$
$\lambda_{1,2} = 0.5,\ d_{1,2} = 0.346\ \text{M}$
$\lambda_{2,i2} = 0.7,\ d_{2,i2} = 0.178\ \text{M}$

Resulting incorrect locus order:

$$k_1 \qquad\qquad i \qquad\qquad k_2$$

$z^* = 1$
$\lambda_{1,i} = 0.525,\ d_{1,i} = 0.322\ \text{M}$
$\lambda_{i,2} = 0.525,\ d_{i,2} = 0.322\ \text{M}$

is mathematically equivalent to maximizing the sum of linkage values λ between adjacent loci, briefly referred to as sum of adjacent linkage values. This procedure is based on the proposition that only the correct locus order maximizes the sum of adjacent linkage values on a chromosome. We prove this proposition in the appendix to show that the SAR criterion is a valid method for constructing multi-locus linkage maps.

Because the SAR criterion is only minimized for the correct locus order, the locus order determined on the basis of the SAR criterion must be the same as the one found by any other valid method. Consequently, the results subsequently derived by using the SAR criterion also apply to any other valid locus ordering method.

Incorrect locus orders

The sum of adjacent linkage values for a locus order described by $n,x,y,$ and z is

$$L(n,x,y,z) = \sum_{0 \leqslant u < z} \lambda_{u,u+1} + \lambda_{z,i} + \lambda_{i,z+1} + \sum_{z < u \leqslant n} \lambda_{u,u+1}$$

$$= \sum_{0 \leqslant u < n} \lambda_{u,u+1} - \lambda_{z,z+1} + \lambda_{z,i} + \lambda_{i,z+1}. \tag{7}$$

Comparing the sum L for two alternative values z' and z'' and omitting equal terms yields

$$L(n,x,y,z') > \qquad L(n,x,y,z'')$$

$$\Leftrightarrow \lambda_{z',i} + \lambda_{i,z'+1} + \lambda_{z'',z''+1} > \lambda_{z'',i} + \lambda_{i,z''+1} + \lambda_{z',z'+1}. \tag{8}$$

All incorrect locus orders for a combination n,x,y can be described by their value of $z^* \in J = \{0,...,n\} \backslash \{x,y\}$. Mapping results in an incorrect locus order characterized by $z^* \in J$ if and only if

$$L(n,x,y,z^*) > L(n,x,y,x) \text{ and} \tag{9}$$

$$L(n,x,y,z^*) > L(n,x,y,y) \text{ and} \tag{10}$$

$$L(n,x,y,z^*) > \max_{z \in J}(L(n,x,y,z)). \tag{11}$$

Proposition (Case 1)

In a BC_1 population of infinite size, locus ordering according to the SAR criterion results in an incorrect locus order of type

$$z^* < x < y \text{ or } x < z^* < y \text{ or } x < y < z^*$$

with $z^* \in J$ if and only if

$$\lambda_{z^*,i} + \lambda_{i,z^*+1} + \lambda_{x,x+1} > \lambda_{x,i} + \lambda_{i,x+1} + \lambda_{z^*,z^*+1} \tag{12}$$

and

$$\lambda_{z^*,i} + \lambda_{i,z^*+1} + \lambda_{y,y+1} > \lambda_{y,i} + \lambda_{i,y+1} + \lambda_{z^*,z^*+1} \tag{13}$$

and Eq. 11 is true.

Proof (Case 1)

From Eq. 8 it follows that

$$L(n,x,y,z^*) > \qquad L(n,x,y,x)$$

$$\Leftrightarrow \lambda_{z^*,i} + \lambda_{i,z^*+1} + \lambda_{x,x+1} > \lambda_{x,i} + \lambda_{i,x+1} + \lambda_{z^*,z^*+1}$$

and

$$L(n,x,y,z^*) > \qquad L(n,x,y,y)$$

$$\Leftrightarrow \lambda_{z^*,i} + \lambda_{i,z^*+1} + \lambda_{y,y+1} > \lambda_{y,i} + \lambda_{i,y+1} + \lambda_{z^*,z^*+1}$$

which completes the proof.

Examples for Case 1

Two scenarios resulting in incorrect locus orders of type $z^* < x$ and $x < z^* < y$ for $n=2$ are shown in Fig. 2.

Proposition (Case 2)

In a BC_1 population of infinite size, locus ordering according to the SAR criterion results always in the correct locus order

$$z^* = x = y$$

if the two duplicate marker loci are located in the same marker interval, i.e., $x=y$.

Proof (Case 2)

We assume without loss of generality $z^* < x$ and obtain a contradiction. Because of Eq. 2

$$\lambda_{z^*,i} = \lambda_{z^*,z^*+1}\lambda_{z^*+1,i} < \lambda_{z^*,z^*+1}$$

$$\lambda_{i,z^*+1} = \lambda_{z^*+1,x}\lambda_{x,i} \leqslant \lambda_{x,i}$$

and because

$$\lambda_{x,x+1} = \lambda_{x,i1}\lambda_{x,i} < \lambda_{i1,x+1} < \lambda_{i2,x+1}$$

we have

$$\lambda_{x,x+1} < (\lambda_{i1,x+1} + \lambda_{i2,x+1})/2 = \lambda_{i,x+1}.$$

In consequence

$$\lambda_{z^*,i} + \lambda_{i,z^*+1} + \lambda_{x,x+1} < \lambda_{z^*,z^*+1} + \lambda_{x,i} + \lambda_{i,x+1}$$

or equivalently (using Eq. 8)

$$L(n,x,x,z^*) < L(n,x,x,x),$$

which completes the proof.

Proposition (Case 3)

In a BC_1 population of infinite size, locus ordering according to the SAR criterion does not result in an incorrect locus order of type

$$z^* < x \text{ or } z^* > y$$

if all markers are equally spaced with linkage value c

$$\lambda_{u,u+1} = c \qquad (14)$$

for $u \in \{1,...,n-1\}$ and $0 < c < 1$.

Proof (Case 3)

We assume without loss of generality $z^* < x$ and obtain a contradiction. Because of Eq. 2 we have

$$\lambda_{x-1,i1} = \lambda_{x-1,x}\lambda_{x,i1} < \lambda_{x-1,x} < \lambda_{i1,x+1}$$

$$\lambda_{x-1,i2} = \lambda_{x-1,x+1}\lambda_{x+1,i2} < \lambda_{i2,x+1}$$

from which follows (Eq. 6)

$$\lambda_{x-1,i} < \lambda_{i,x+1}. \qquad (15)$$

Moreover, for any $z < x$,

$$\lambda_{z-1,i} = \lambda_{z-1,z+1}\lambda_{z+1,i} < \lambda_{z+1,i}. \qquad (16)$$

From Eqs. 15 and 16 follows (using Eqs. 4 and 14)

$$\lambda_{x-1,i} + \lambda_{i,x} + \lambda_{x,x+1} < \lambda_{x-1,x} + \lambda_{x,i} + \lambda_{i,x+1} \text{ and}$$

$$\lambda_{z-1,i} + \lambda_{i,z} + \lambda_{z,z+1} < \lambda_{z-1,z} + \lambda_{z,i} + \lambda_{i,z+1}$$

or equivalently (using Eq. 8)

$$L(n,x,y,x-1) < L(n,x,y,x) \text{ and}$$

$$L(n,x,y,z-1) < L(n,x,y,z)$$

from which follows

$$L(n,x,y,z^*) < L(n,x,y,x)$$

which completes the proof.

Proposition (Case 4)

In a BC_1 population of infinite size, locus ordering according to the SAR criterion results in a correct locus order

$$z^* = x \text{ or } z^* = y$$

if all markers are equally spaced with linkage value c (Eq. 14) and the duplicate marker loci are located in the center between their flanking markers

$$\lambda_{x,i1} = \lambda_{i1,x+1} = \lambda_{y,i2} = \lambda_{i2,y+1} = \sqrt{c}. \qquad (17)$$

Proof (Case 4)

For $z^* < x$ and $y < z^*$ the proof corresponds to the proof for Case 3.

Because of Eq. 7

$$L(n,x,y,z) - L(n,x,y,z+1) = \lambda_{z,i} - \lambda_{z+2,i}$$
$$= \lambda_{z,i} - \lambda_{z+1,i} + \lambda_{z+1,i} - \lambda_{z+2,i} \qquad (18)$$

and for any $x < z < (y+x)/2$

$$\lambda_{z,i} > \lambda_{z+1,i} \qquad (19)$$

because (Eqs. 6, 17 and 14)

$$\lambda_{z,i1} + \lambda_{z,i2} > \lambda_{z+1,i1} + \lambda_{z+1,i2}$$
$$\Leftrightarrow \lambda_{i1,x+1}\lambda_{x+1,z} + \lambda_{z,y}\lambda_{y,i2} > \lambda_{i1,x+1}\lambda_{x+1,z+1} + \lambda_{z+1,y} + \lambda_{y,i2}$$
$$\Leftrightarrow \quad c^{z-x-1} + c^{y-z} > c^{z-x} + c^{y-z-1}$$
$$\Leftrightarrow \quad c^{z-x-1} - c^{z-x} > c^{y-z-1} - c^{y-z}$$
$$\Leftrightarrow \quad c^{z-x} > c^{y-z}$$
$$\Leftrightarrow \quad z - x < y - z$$
$$\Leftrightarrow \quad z < (y+x)/2.$$

For symmetry reasons, we have for $x+1+\delta \leq (x-y)/2$

$$L(n,x,y,x+1+\delta) = L(n,x,y,y-\delta), \qquad (20)$$

and as a special case

$$L(n,x,y,x+1) = L(n,x,y,y). \qquad (21)$$

From Eqs. 18, 19 and 20 it follows that for $x < z < (y+x)/2$

$$L(n,x,y,z) \geqslant L(n,x,y,z+1). \qquad (22)$$

We now assume without loss of generality $x+1 < z^* < (y+x)/2$ and obtain a contradiction, because from Eqs. 21 and 22 it follows that

$$L(n,x,y,y) = L(n,x,y,x+1) \geqslant L(n,x,y,z^*),$$

which completes the proof.

A test for detection of duplicate markers

Duplicate markers result in an excess of heterozygotes at the ghost locus in a BC_1 mapping population. Therefore, a statistical test of the null hypothesis that the frequency p of heterozygotes at the locus under consideration is 1/2 can help to identify them. While approximate χ^2 tests for segregation distortion are common (Weir 1996), we propose to use an exact test based on the binomial distribution because of its superior statistical properties.

Given the null hypothesis $H_0 : p = 1/2$ is true, the probability of obtaining more than m heterozygotes in a population of size s can be obtained from the probability function of the binomial distribution as

$$P_0(M > m) = 1 - \sum_{k=0}^{m} \binom{s}{k} 1/2^s.$$

We use the observed number m_b of heterozygotes in a population of size s as a test statistic, and determine the corresponding critical value m^* for testing H_0 for a given Type I error α by solving $P_0(M > m^*) \leq \alpha$. The null hypothesis is rejected if $m_b > m^*$.

The Type II error β of the test depends (1) on the recombination frequency $\rho_{i1,i2}$ between the duplicate loci, which determines the expected frequency of heterozygotes and (2) the size s of the mapping population. For two duplicate markers i_1 and i_2, the genotypes $\frac{10}{00}$, $\frac{01}{00}$, and $\frac{11}{00}$ are scored as heterozygous with respect to the ghost marker i (Fig. 1). Hence, the frequency of heterozygotes at the ghost marker can be determined in analogy to Eq. 5 from γ_{10},

Fig. 3 Occurrence of incorrect locus orders for the true locus order k_1, i_1, k_2, i_2. The *lines* are obtained by using $\lambda_{1,i1} = 0.60, 0.80, 0.90, 0.99$ in Eqs. 23 and 24 and solving for $\lambda_{2,i2}$. The *shaded areas* indicate parameter combinations of $\lambda_{1,i1}$, $\lambda_{i1,2}$, and $\lambda_{2,i2}$ for which incorrect locus orders do occur
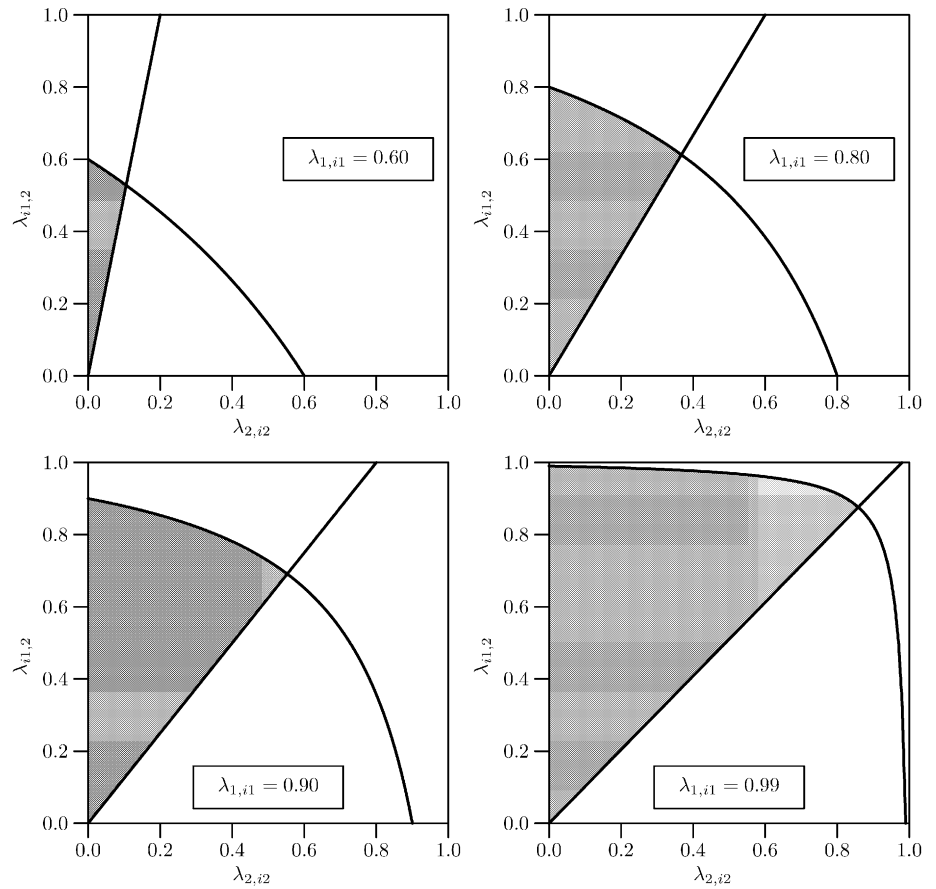
γ_{01}, and γ_{11} as $p = (1 + \rho_{i1,i2})/2$. Consequently, under the alternative hypothesis

$$H_{A,\rho} : p = \frac{1 + \rho_{i1,i2}}{2},$$

the probability of obtaining m heterozygotes is

$$P_{A,\rho}(M = m) = \binom{s}{m} \left(\frac{1 + \rho_{i1,i2}}{2}\right)^m \left(\frac{1 - \rho_{i1,i2}}{2}\right)^{s-m}$$

and the power $1 - \beta$ of the test can be obtained from

$$\beta = \sum_{k=0}^{m^*} P_{A,\rho}(M = k).$$

Numerical results

Occurrence of incorrect locus orders

In the previous section, we derived for the general case of n markers the conditions under which incorrect locus orders occur if there are duplicate marker loci on a chromosome. Here, we illustrate the typical properties of situations for which incorrect locus orders occur with numerical examples for the four locus case.

The locus order k_1, i_1, i_2, k_2 is characterized by $n=2$ and $x=y$. As shown in the theory section (Case 2), no incorrect locus order can occur for this situation.

The locus order k_1, i_1, k_2, i_2 is characterized by $n=2$, $x=1$, $y=2$. For these parameters the locus orders k_1, i, k_2 ($z^*=x=1$) and k_1, k_2, i ($z^*=y=2$) are correct. Simple combinatorical considerations show that the only incorrect locus order is i, k_1, k_2 ($z^*=x-1=0$). Applying Eqs. 12 and 13 yields the conditions under which this incorrect locus order occurs:

$$\lambda_{1,i1} \lambda_{i1,2} > \left(\lambda_{i1,2} + \lambda_{2,i2}\right)/2 \text{ and} \tag{23}$$

$$\lambda_{1,i1} + \lambda_{1,i1} \lambda_{i1,2} \lambda_{2,i2} > \lambda_{i1,2} + \lambda_{2,i2}. \tag{24}$$

Setting $\lambda_{1,i1} = 0.60, 0.80, 0.90, 0.99$ and solving both inequalities for $\lambda_{i1,2}$ results in the graphs in Fig. 3, showing combinations of linkage values $\lambda_{1,i1}$, $\lambda_{i1,2}$, $\lambda_{2,i2}$ for which incorrect locus orders do occur. A prerequisite for an incorrect map order is that $\lambda_{i1,2}$ is greater than $\lambda_{2,i2}$. The set of parameter combinations, for which mapping results in an incorrect locus order, increases with increasing linkage between k_1 and i_1: For $\lambda_{1,i1} < 0.5$ only correct map orders are found, as can be seen from Eq. 23 using simple arithmetics. In contrast, for very tight linkage ($\lambda_{1,i1} = 0.99$) incorrect map orders occur for a broad range of parameter settings, including $0.8 > \lambda_{i1,2} > \lambda_{2,i2}$. Summarizing, incorrect map orders occur if (1) k_1 and i_1 are tightly linked and (2) linkage between i_1 and k_2 is greater than linkage between k_2 and i_2 (Fig. 3).

The locus order i_1, k_1, k_2, i_2 is characterized by $n=2$, $x=0$, and $y=2$. For these parameters, correct locus orders are

Fig. 4 Occurence of incorrect locus orders for the true locus order i_1,k_1,k_2,i_2. The *lines* are obtained by using $\lambda_{1,2}$=0.1,0.5 in Eqs. 25 and 26 and solving for $\lambda_{2,i2}$. The *shaded areas* indicate parameter combinations of $\lambda_{i1,1}$, $\lambda_{1,2}$, and $\lambda_{2,i2}$ for which incorrect locus orders do occur
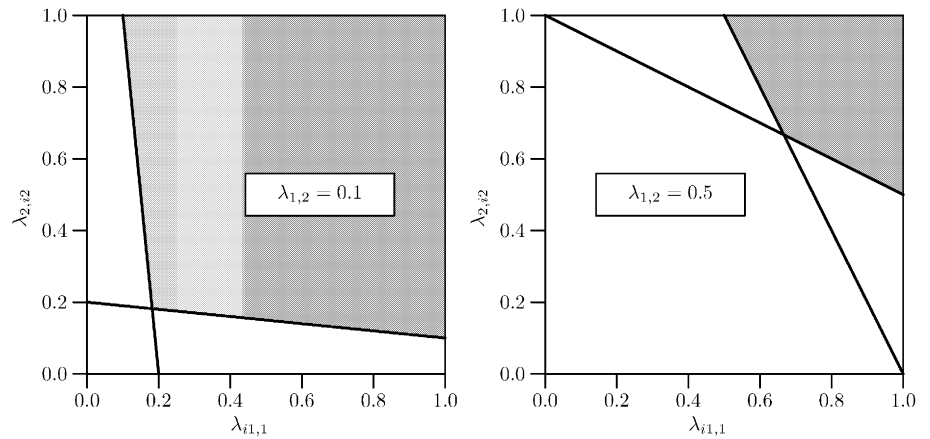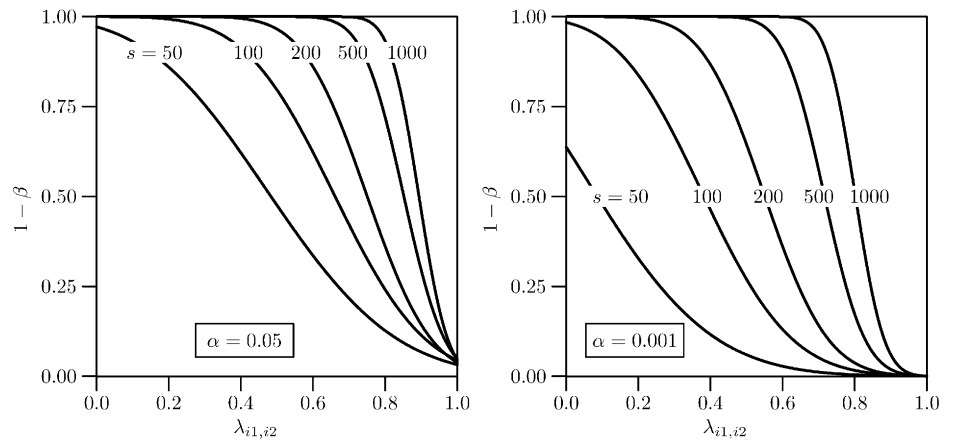

Fig. 5 Power $1-\beta$ of the exact test for segregation distortion in a BC$_1$ population for Type I errors α=0.05 and 0.001, depending on population size s and linkage $\lambda_{i1,i2}$ between two duplicate marker loci

i,k_1,k_2 (z^*=x=0) and k_1,k_2,i (z^*=y=2). The only incorrect locus order is k_1,i,k_2 (z^*=x+1=1). Applying Eqs. 12 and 13 yields the conditions under which this incorrect locus order occurs:

$$\left(\lambda_{i1,1}\lambda_{1,2} + \lambda_{i2,2}\right)/2 > \lambda_{1,2} \text{ and} \tag{25}$$

$$\left(\lambda_{i1,1} + \lambda_{1,2}\lambda_{i2,2}\right)/2 > \lambda_{1,2}. \tag{26}$$

Setting $\lambda_{1,2}$=0.1,0.5 and solving both inequalities for $\lambda_{2,i2}$ results in the graphs in Fig. 4, showing combinations of linkage values $\lambda_{i1,1}$, $\lambda_{1,2}$, $\lambda_{2,i2}$, for which incorrect locus orders do occur. For loose linkage between k_1 and k_2 ($\lambda_{1,2}$=0.1), the set of parameter combinations resulting in incorrect locus orders is quite large, a prerequisite is that neither $\lambda_{i1,1}$ nor $\lambda_{2,i2}$ is smaller than 0.2. With increasing linkage of k_1 and k_2, the set of parameter combinations decreases, for which incorrect locus orders do occur. However, even for very tight linkage, incorrect map orders do occur, if the linkage values $\lambda_{i1,1}$ and $\lambda_{2,i2}$ are large and have approximately the same value. Summarizing, incorrect locus orders occur for loose linkage of k_1 and k_2, when linkage between i_1 and k_1 as well as between k_2 and i_2 is almost equal and each value is at least twice as large as linkage between k_1 and k_2 (Fig. 4).

Power of detecting duplicate loci

Testing for segregation distortion is important to detect duplicate marker loci and, hence, avoid inappropriate application of incorrect linkage maps. Here, we investigate the power of the exact test for segregation distortion in a BC$_1$ population, depending on the size of the mapping population and the linkage value between the duplicate marker loci.

For a Type I error α=0.05 of incorrectly assuming the presence of a ghost marker, using a mapping population of size s=50, the power of detecting a ghost marker is greater than 0.9 only if the linkage between the duplicate loci is greater than 0.2 (which corresponds to a map distance of approximately 50 cM) (Fig. 5). Mapping populations with size s=500 or even 1,000 are required to detect with a high probability ghost loci, resulting from duplicate markers with linkage values between 0.8 and 1.0 (corresponding to map distances of about 10 and 0 cM).

For a smaller Type I error α=0.001, the minimum population size required to detect duplicate loci with a high probability is s=100, if the linkage value is lower than 0.2 (Fig. 5). For linkage values greater 0.8, populations larger than s=1,000 individuals are required.

Discussion

Genetic model

For our derivations we used the assumption of no interference (Stam 1979) underlying Haldane's (1919) mapping function. This is a simplified mathematical model and there exist more sophisticated models of crossover formation in meiosis, which fit experimental data better (McPeek and Speed 1995). Briefly, the assumption of no interference has (1) the advantage of mathematical simplicity, yielding equations which can be easily evaluated and (2) the results can be applied without knowing the exact amount of interference in the chromosome region under consideration. For a more detailed discussion concerning the use of the assumption of no interference see Frisch and Melchinger (2001). Note that Eq. 6, defining linkage between a ghost locus and a distinguishable marker, holds true for arbitrary mapping functions. How-ever, the results for locus ordering may be affected when applying a mapping function different from Haldane's.

The definition of an incorrect locus order in the theory section considers a locus order as correct if the ghost marker maps to one of the two intervals in which the duplicate markers are located. This is appropriate for two situations: (1) The marker itself is part of the gene (e.g., RGA or EST markers), and the target gene is duplicated, or (2) the marker is tightly linked to a target gene and the complete region containing marker and target gene is duplicated.

In contrast, if only the marker is duplicated, but not the target gene, and only one of the two duplicate marker loci is tightly linked with the target gene, then this definition of incorrect locus orders is not appropriate. In such a case, linkage analysis does not identify the ghost marker as being tightly linked to the target gene, because recombination between the ghost marker and a target is the mean recombination frequency between the target and the two duplicate loci (Eq. 5). This situation may also negatively affect construction of linkage maps, but is not the subject of the present study.

Ghost QTL and ghost markers

Ghost QTL and ghost markers share the properties that (1) biometrical analysis maps a locus to an incorrect position on the linkage map, and (2) this is caused by the fact that not a single locus but two indistinguishable loci are underlying the observed differences between individuals.

However, there are also fundamental differences between the two phenomena:

1. Ghost markers occur in the initial construction of a linkage map, whereas ghost QTL are detected in QTL analysis conducted after having a linkage map available.

2. Ghost markers can map outside the interval of the duplicated markers, whereas ghost QTL are located between the underlying QTL.
3. Ghost markers result from duplicated DNA sequences, whereas ghost QTL may occur from two loci having entirely different DNA sequences but affecting the same phenotypic trait.

Summarizing, the ghost marker phenomenon has similarities to the ghost QTL phenomenon, but from the differences mentioned above, the implications for practical applications are different.

Segregation distortion caused by zygotic selection

If segregation distortion is detected at a marker locus, this may not only be due to duplicate markers, but also due to various other reasons, one of which is zygotic section. In this case, an excess of heterozygotes follows from a reduced fitness of homozygotes. To distinguish both situations, the following considerations can be made: For duplicate markers, segregation distortion occurs only at the ghost locus. In contrast, for zygotic selection, segregation distortion occurs not only at the locus which is affected by selection, but also at closely linked loci.

This is illustrated by a numerical example: Consider a 2 M chromosome, carrying 21 equally spaced markers and a BC_1 mapping population consisting of $s=100$ individuals. If two duplicate marker loci are located at map positions 0.87 and 1.13 ($\lambda_{i1,i2}=0.6$), then the test for segregation distortion ($\alpha=0.05$) detects segregation distortion with a probability $1-\beta\approx0.7$ at map position 1.0 (Fig. 6). At all other loci, segregation distortion is only detected with the probability of the Type I error $\alpha=0.05$. If there is zygotic selection at the locus at position 1.0 such that from the homozygotes only 50% survive, then with a comparable probability of about 0.7, segregation distortion is detected by the test at map position 1.0. However, in this case linked markers adjacent to the locus at map position 1.0 also display segregation distortion with a high probability (Fig. 6).

Consequently, if segregation distortion is detected only at one locus, chances are high that duplicate markers are the reason, whereas if segregation distortion is detected at several closely linked loci, this can be taken as an indicator for zygotic selection.

Effects of sampling and locus ordering method

The proofs in our theoretical investigation assume an indefinite population size resulting in exact linkage values λ. However, in a mapping study, linkage is estimated from a finite sample of a population, and a considerable estimation error may occur depending on the sample size. Therefore, we used a simulation study to investigate whether the theoretically expected results for indefinite populations, known linkage values, and the SAR criterion

Fig. 6 Probability of detecting segregation distortion (α=0.05) at loci equally distributed on a 2 M chromosome with a BC_1 population of size s=100. *Left diagram* Duplicate loci at map positions 0.87 and 1.13 ($\lambda_{i1,i2}$=0.6). *Right diagram* Zygotic selection with a survival rate of homozygotes of 0.5
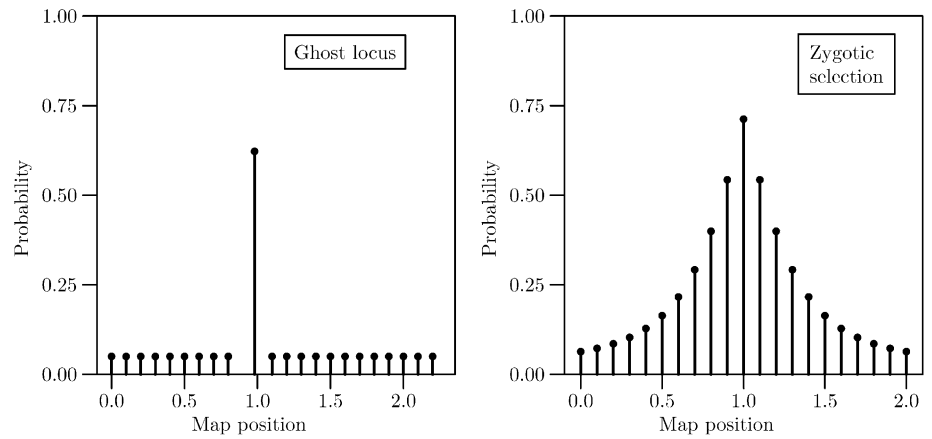
are obtained when applying different mapping programs to finite populations.

The simulation program Plabsim (Frisch et al. 2000) was used to generate the datasets and the mapping programs GMendel (Liu and Knapp 1990; Holloway and Knapp 1993), Mapmaker (Lander et al. 1987), and Joinmap (Stam 1993) were applied for linkage analysis. The GMendel software performs locus ordering with the SAR criterion, while Joinmap uses a modification of the SAR criterion and Mapmaker applies a maximum likelihood approach.

We investigated a chromosome with $(k_1, i_1, k_2, i_2) = (0.0, 0.1, 0.3, 1.0)$ and generated with Plabsim 100 BC_1 populations $\left(\frac{111}{111} \times \frac{000}{000}\right) \times \frac{000}{000}$ for each population size s=50,100,250,100, and 5,000. The populations were evaluated for the genotype at loci k_1 and k_2. The genotypes $\frac{10}{00}$, $\frac{01}{00}$, and $\frac{11}{00}$ with respect to loci i_1 and i_2 were scored as $\frac{1}{0}$ with respect to the ghost marker i, $\frac{00}{00}$ was scored as $\frac{0}{0}$.

The resulting datasets were analyzed with GMendel. For a population size of s=50, the incorrect locus order i, k_1, k_2 was found in 23% (Table 1). With increasing population size, the percentage of incorrect maps increased and reached 93% for s=5,000. In consequence, for small populations the estimation error of the recombination frequencies resulted in frequent findings of a correct locus order or no linkage at all. However, for large populations the incorrect locus order was observed in most cases, as expected from theory.

One population of size n=5,000 was analyzed with GMendel, Joinmap, and Mapmaker. All three programs yielded the incorrect locus order i, k_1, k_2. The programs

GMendel and Mapmaker estimated the map distances $\left(\hat{d}_{i,1}, \hat{d}_{1,2}\right) = (0.372, 0.297)$, which were close to those expected from Eq. 6 $\left(\hat{d}_{i,1}, \hat{d}_{1,2}\right) = (0.370, 0.300)$. However, Joinmap estimated $\left(\hat{d}_{i,1}, \hat{d}_{1,2}\right) = (0.289, 0.508)$, which is surprising because according to theory, the occurrence of duplicate markers should not influence the map distance $d_{1,2}$ between distinguishable markers k_1 and k_2. Consequently, the incorrect locus order was observed irrespective of the locus ordering method implemented in these programs, as expected from theoretical considerations.

Type of the mapping population

Throughout our study we focused on BC_1 mapping populations, because determining the fractions of recombinant gametes and heterozygotes is simple in BC_1 (Fig. 1). However, all results obtained for locus ordering using the SAR criterion depend only on the known recombination frequencies between loci. How and from which type of population they are obtained is irrelevant for the derivations. For a different type of mapping population, e.g., an F_2 population, the procedure of obtaining recombination frequencies between loci differs, but the locus ordering procedure based on the SAR criterion does not. Therefore, the presented results are valid for any type of mapping population in which heterozygous individuals occur.

In recombinant inbred lines or doubled haploids, both homologues of a chromosome are identical copies and, therefore, all loci are homozygous. However, two duplicate marker loci may carry two different alleles and therefore will be scored as heterozygous with respect to the ghost marker. In consequence, markers which are scored as heterozygous in these two types of mapping populations may be ghost markers.

Implications for applying linkage maps

The incorrect map position of a ghost marker affects the application of linkage maps, for which not only tight linkage of a target gene and adjacent markers is required,

Table 1 Locus orders resulting from applying GMendel to simulated BC_1 datasets of size s =50, 100, 250, 1,000 and 5,000. The underlying linkage map was (k_1, i_1, k_2, i_2)=(0,0.1,0.3,1.0)

Estimated locus order	Population size s				
	50	100	250	1,000	5,000
	[%]				
No linkage	55	6	0	0	0
k_1, k_2, i	17	38	38	38	7
k_1, i, k_2	5	7	3	0	0
i, k_1, k_2	23	49	59	62	93

but the correct location of the target gene with respect to flanking markers is important. Examples are map-based cloning, marker-assisted backcrossing, and marker-assisted selection.

In map-based cloning, the chromosome region where a gene is located, is first determined with a low-density linkage map. Then, the region of the target gene is analyzed with a marker density higher than 1 marker per cM, in order to fine-map the gene and to locate a marker interval, to be used for genomic library screening.

In the fine mapping step, no problems from incorrect locus orders are expected. First, incorrect locus orders do not occur for equal marker spacing (Proposition 4); second, numerical evaluation of Eqs. 12 to 13 shows that for marker distances smaller than 1 cM the correct locus order is always found. However, in the first stage with low density linkage maps, duplicate markers can result in mapping the target gene into an incorrect chromosome region, such that none of the high-resolution markers investigated in the second step are tightly linked to the target gene.

For marker-assisted selection, a QTL is mapped to a chromosome interval, and subsequently, the markers flanking the chromosome interval are used for indirect selection for the presence of the favorable allele at the QTL. Estimated locations of the QTL are usually not precise point estimates, but the QTL is assumed to be located in a so-called support interval, which often covers large chromosome segments of up to 90 cM (Visscher et al. 1996).

If a chromosome region is duplicated, and contains a marker and a QTL, then tight linkage between the marker and the QTL is detected, irrespective of the duplication. Because of the large marker distances required to select for a QTL in a support interval, both the marker and QTL may map into an incorrect flanking marker interval. Selection for the markers incorrectly assumed to be flanking the target region may not be an indirect selection for the chromosome region which carries the QTL. This can greatly reduce the efficiency of marker-assisted selection.

In marker-assisted backcrossing for introgression of a target gene from a donor parent into the genetic background of a recipient parent, markers can be used for two purposes: (1) to select for the presence of a tightly linked target gene (foreground selection) and (2) to select against the genetic background of the recipient parent (background selection) (Tanksley et al. 1989). Marker-assisted backcrossing is routinely applied, e.g., in maize breeding to introgress transgenes in inbred lines used for production of commercial hybrids.

For foreground selection, the linkage between marker and target gene needs to be very tight. If a chromosome region is duplicated, and contains the target gene and a tightly linked marker, no negative effects of the duplication with respect to foreground selection are expected.

In marker-assisted background selection, a primary goal is to reduce the length of the donor chromosome segment around the target gene (Stam and Zeven 1981; Young and Tanksley 1989; Frisch et al. 1999). This is achieved by selecting for the allele of the recipient at markers flanking the target gene. In backcross programs, population size is usually restricted by the reproduction coefficient of the species and practical constraints. In order to observe recombination between the target gene and the flanking markers with a high probability in a finite population, linkage between the marker and target gene should not be extremely tight (Frisch et al. 1999). This implies the use of more distant marker brackets for background selection, which can result in incorrect locus orders.

In consequence, if the target gene maps into an incorrect chromosome interval, selection for markers incorrectly assumed to flank the target gene does not reduce the donor chromosome segment attached to the target gene. This can greatly reduce the efficiency of fast recovery of the recurrent parent genome.

Conclusions and further research needs

Pointing out the extent of duplicated sequences as well as the evolutionary formation of large gene families by duplication events in eukaryotic genomes, the ghost marker phenomenon was presumably overlooked so far. The existence of ghost markers in linkage maps is very likely, and many of them remain undetected because of very close linkage of the underlying duplicate loci and the insufficient size of mapping populations. Furthermore, the ghost marker phenomenon has the potential to provide new explanations for distorted segregation at numerous loci in existing and emerging linkage maps.

The application of duplicated sequences as molecular markers is not restricted to gene-derived markers like ESTs or RGAs. Single bands of several other molecular marker types like AFLPs and SSRs are also known to frequently represent multiple sequences resulting in the same complications for the construction of linkage maps. Because correct linkage maps are essential for important applications in genetics and breeding, many interesting questions concerning this subject warrant further research: What are the consequences, if only the marker but not the target gene is duplicated? How do duplicate markers affect map distances between observable markers in multipoint estimation of recombination frequencies? Can the presented approach be extended to more sophisticated crossover formation models? How does the mode of inheritance (codominance vs. dominance) affect the ghost marker phenomenon? Furthermore, besides the implications of ghost markers on map-based cloning, MAS, and MAB, it is also necessary to investigate their influence on known discrepancies between genetic and physical maps regarding locus order and contig assembly.

Appendix

To show that the SAR criterion is suitable for locus ordering we consider a chromosome consisting of n loci at positions $k_1,...,k_n$ and prove by mathematical induction:

Proposition (SAR)

The SAR for the correct locus order is smaller than the SAR for any other order of the n loci on the chromosome.

Induction basis

From Eq. 1 it follows that the locus order with a minimum SAR has a maximum sum of adjacent linkage values. For a subset of three loci at positions $k_{u1} < k_{u2} < k_{u3}$ with $u_1, u_2, u_3 \in \{1,...,n\}$ the correct locus order has the greatest sum of adjacent linkage values because

$$L(k_{u1}, k_{u2}, k_{u3}) > L(k_{u2}, k_{u1}, k_{u3})$$
$$\Leftrightarrow \quad \lambda_{u1,u2} + \lambda_{u2,u3} > \lambda_{u2,u1} + \lambda_{u1,u3}$$
$$\Leftrightarrow \quad \lambda_{u2,u3} > \lambda_{u1,u2}\lambda_{u2,u3}$$

and

$$L(k_{u1}, k_{u2}, k_{u3}) > L(k_{u2}, k_{u3}, k_{u1})$$
$$\Leftrightarrow \quad \lambda_{u1,u2} + \lambda_{u2,u3} > \lambda_{u2,u3} + \lambda_{u3,u1}$$
$$\Leftrightarrow \quad \lambda_{u1,u2} > \lambda_{u1,u2}\lambda_{u2,u3}$$

Induction hypothesis

Consider a subset of $3 < m < n$ loci at map positions $k_{u1} < \cdots < k_{um}$ with $u_1,...,u_m \in \{1,...,n\}$ and denote $U = \{u_1,...,u_m\}$. We assume that for the subset of m loci, the correct locus order has the largest sum of adjacent linkage values.

Induction step

Consider an $m+1$ th locus at map position k_v ($v \in \{1,...,n\}$ and $vrlap/ \in U$). We further define

$$w = \max (s|s \in \mathcal{U}, k_s < k_v)$$

The correct locus order for the $m+1$ loci has the sum of adjacent linkage values

$$L = \sum_{\mu \in \mathcal{U}, \mu < w} \lambda_{\mu,\mu+1} + \lambda_{w,v} + \lambda_{v,w+1} + \sum_{\mu \in \mathcal{U}, w < \mu} \lambda_{\mu,\mu+1},$$

whereas an incorrect locus order has the sum of adjacent linkage values

$$L^* = \sum_{\mu \in \mathcal{U}, \mu < w-\varepsilon} \lambda_{\mu,\mu+1} + \lambda_{w-\varepsilon,v} + \lambda_{v,w-\varepsilon+1}$$
$$+ \sum_{\mu \in \mathcal{U}, w-\varepsilon < \mu} \lambda_{\mu,\mu+1}$$

where $w - \varepsilon \in \mathcal{U}$ and without loss of generality $0 < \epsilon$.

Omitting equal terms in L and L^* yields

$$L^* < L$$
$$\Leftrightarrow \lambda_{w-\varepsilon,v} + \lambda_{v,w-\varepsilon+1} + \lambda_{w,w+1} < \lambda_{w-\varepsilon,w-\varepsilon+1} + \lambda_{w,v} + \lambda_{v,w+1},$$

which can be shown to be true using the induction assumption and Eq. 3:

$$\lambda_{w-\varepsilon,v} \quad = \lambda_{w-\varepsilon,w-\varepsilon+1}\lambda_{w-\varepsilon+1,v} < \lambda_{w-\varepsilon,w-\varepsilon+1}$$
$$\lambda_{v,w-\varepsilon+1} = \lambda_{w-\varepsilon+1,w}\lambda_{w,v} \quad\quad < \lambda_{w,v}$$
$$\lambda_{w,w+1} \quad = \lambda_{w,v}\lambda_{v,w+1} \quad\quad\quad < \lambda_{v,w+1}.$$

Hence, if the SAR criterion is valid for m loci, it is also valid for $m+1$ loci. The induction basis shows that the SAR criterion is valid for $m=3$, which completes the proof.

Note that the SAR criterion is used e.g., in software GMendel (Liu and Knapp 1990, Holloway and Knapp 1993), which is widely used by plant geneticists. However, to our knowledge a proof of its optimum properties has not yet been published. Moreover, the above proof is constructive in that it describes an algorithm which finds the correct locus order for n loci with $n^2/2 - 4$ comparisions. To our knowledge, available mapping software applies either a search of the entire parameter space, which requires $n!/2$ comparisons to find the correct locus order, or it applies numerical optimization methods, which may converge only to local maxima of the SAR instead of the global maximum.

References

Acarkan A, Rossberg M, Koch M, Schmidt R (2000) Comparative genome analysis revealed extensive conservation of genome organziation among *Arabidopsis thaliana* and *Capsella rubella*. Plant J 23:233–243

Blanc G, Barakat A, Guyot R, Cook R, Delseny M (2000) Extensive duplication and reshuffling in the Arabidopsis genome. Plant Cell 12:1093–1101

Frisch M, Melchinger AE (2001) The length of the intact chromosome segment around a target gene in marker-assisted backcrossing. Genetics 157:1343–1356

Frisch M, Bohn M, Melchinger AE (1999) Minimum sample size and optimal positioning of flanking markers in marker-assisted backcrossing for transfer of a target gene. Crop Sci 39:967–975 (erratum: Crop Sci 39:1913)

Frisch M, Bohn M, Melchinger AE (2000) Plabsim: software for simulation of marker-assisted backcrossing. J Hered 91:86–87

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, Glazebrook J, Sessions A, Oeller P, Varma H (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). Science 296:92–100

Haldane JBS (1919) The combination of linkage values and the calculation of distance between the loci of linkage factors. J Genet 8:299–309

Holloway JL, Knapp SJ (1993) GMendel 3.0: software for the analysis of genetic markers and maps. Oregon State University, Corvallis

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) Mapmaker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Liu BH, Knapp SJ (1990) GMendel: a program for Mendelian segregation and linkage analysis of individual or multiple progeny populations using likelihood ratios. J Hered 81:407

Martinez O, Curnow RN (1992) Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theor Appl Genet 85:480–488

McPeek MS, Speed TP (1995) Modeling interference in genetic recombination. Genetics 139:1031–1044

Ohno S (1970) Evolution by gene duplication. Springer, Berlin, Heidelberg, New York

Quint M, Dußle CM, Melchinger AE, Lübberstedt T (2003) Identification of genetically linked RGAs by BAC screening in maize and implications for gene cloning, mapping, and MAS. Theor Appl Genet (in press)

Schnell FW (1961) Some general formulations of linkage effects in inbreeding. Genetics 46:947–957

Stam P (1979) Interference in genetic crossing over and chromosome mapping. Genetics 92:573–594

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: Joinmap. Plant J 3:739–744

Stam P, Zeven AC (1981) The theoretical proportion of the donor genome in near-isogenic lines of self-fertilizers bred by backcrossing. Euphytica 30:227–238

Tanksley SD, Young ND, Patterson AH, Bonierbale MW (1989) RFLP mapping in plant breeding: new tools for an old science. Biotechnology 7:257–263

Tarchini R, Biddle P, Wineland R, Tingey S, Rafalski A (2000) The complete sequence of 340 kb of DNA around the rice Adh1-Adh2 region reveals interrupted co-linearity with maize chromosome 4. Plant Cell 12:381–391

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

Visscher PM, Thompson R, Haley CS (1996) Confidence intervals in QTL mapping by bootstrapping. Genetics 143:1013–1020

Vuylsteke M, Mank R, Antonise R, Bastiaans E, Senior ML, Stuber CW, Melchinger AE, Lubbersted T, Xia XC, Stam P, Zabeau M, Kuiper M. (1999) Two high-density AFLP linkage maps of *Zea mays* L.: analysis of distribution of AFLP markers. Theor Appl Genet 99:921–935

Weir B (1996) Genetic data analysis. Sinauer, Sunderland, Mass., p 27

Wendel JF (2000) Genome evolution in polyploids. Plant Mol Biol 42:225–49

Young ND, Tanksley SD (1989) RFLP analysis of the size of chromosomal segments retained around the *Tm*-2 locus of tomato during backcross breeding. Theor Appl Genet 77:353–359