Theor Appl Genet (1996) 93:1185–1192

C. Halldén · M. Hansen · N.-O. Nilsson · A. Hjerdin
T. Säll

Competition as a source of errors in RAPD analysis

Received: 21 November 1995 / Accepted: 23 February 1996

Abstract We have used artificial 1:1 DNA mixtures of all pairwise combinations of four doubled haploid *Brassica napus* lines to test the ability of RAPDs to function as reliable dominant genetic markers. In situations where a specific RAPD band is present in one homozygous line but absent in the other, the band is expected in the artificial heterozygote, i.e. in the 1:1 DNA mixture. In 84 of all 613 heterozygous situations analysed, the expected band failed to amplify in the RAPD reaction. Thus, RAPD markers will lead to an erroneous genetic interpretation in 14% of all cases. In contrast, the formation of non-parental heteroduplex bands was found at a frequency of only 0.2%. Analysis of 1:1 mixtures using (1) a different set of optimized reaction conditions and (2) a material with low genomic complexity (*Bacillus cereus*) gave identical results. Serial dilutions of one genome into another, in steps of 10%, showed that all of the polymorphic bands decreased in intensity as a linear function of their respective proportion in the mixture. In dilutions with water no differences in band intensity were detected. Thus, competition occurs in the amplification of all RAPD fragments and is a major source of genotyping errors in RAPD analysis.

Key words Random amplified polymorphic DNA · Competition · DNA mixtures

Introduction

Random amplification of polymorphic DNA (RAPD) analysis utilizes single arbitrary sequence 10-base oligonucleotides as primers to amplify discrete fragments of DNA in low-stringency polymerase chain reactions (PCR) (Williams et al. 1990; Welsh and McClelland 1990). The random amplified polymorphic DNA (RAPD) technique is fast, easy to perform and comparatively cheap. As it relies on universal sets of primers without need for prior sequence information, it is immediately applicable to the analysis of most organisms. This marker system has been used in many different applications involving the detection of DNA sequence polymorphisms: mapping in different types of populations (Carlson et al. 1991; Reiter et al. 1992), isolation of markers linked to various traits or specific targeted intervals (Giovannoni et al. 1991; Michelmore et al. 1991; Deloume et al. 1994) and applications like variety identification and analysis of parentage (Tinker et al. 1993; Mailer et al. 1994).

The RAPD technology, however, has some limitations. RAPD markers are in general dominant, thereby having a lower information content than codominant markers in linkage analysis of F_2 populations (Williams et al. 1990). Furthermore, the reliability of RAPD markers has been questioned in several studies (Weeden et al. 1992). As minor changes in reaction conditions can significantly alter the number and intensity of the amplification products, reproducibility can be difficult to maintain. Penner et al. (1993) reported difficulties in obtaining identical band patterns from the same set of primers and materials among different laboratories. In their study the type of thermocycler used for RAPD analysis seemed to be a key determinant of the reproducibility of band patterns. A number of different approaches have been adopted to account for artefacts affecting the reproducibility of the analysis. Obvious solutions to this problem are to carry out replicate runs and discard all non-reproducible bands (Hu and Quiros 1991; Reiter et al. 1992) or to use all bands and accept a certain error level (Stiles et al. 1993). If replicate runs are carried out, a correction for the effect of the artefacts can be made (Lamboy 1994). Another type of problem which has been reported is the occur-

Communicated by J. Beckmann

C. Halldén(✉) · M. Hansen · N.-O. Nilsson · A. Hjerdin
Hilleshög AB, Box 302, S-261 23 Landskrona, Sweden

T. Säll
Department of Genetics, Lund University, S-223 62 Lund, Sweden

rence of RAPD bands in progeny but not in their parental DNAs, a phenomenon explained as heteroduplex formation (Riedy et al. 1992; Hunt and Page 1992; Ayliffe et al. 1994). This can occur when allelic sequences of different lengths are amplified. However, this is not a major problem since codominance and the possibility of heteroduplex formation occurs only rarely in RAPD analysis.

It has been suggested that the outcome of a RAPD reaction is in part determined by a competition for priming sites in the genome (Williams et al. 1993). Amplification is probably initiated at many sites, but only a subset of all possible products are detected as visible bands after amplification. Several observations support this view. For example, in several mapping projects non-Mendelian inheritance for a significant fraction of all polymorphic bands has been detected, possibly indicating problems with reproducibility and with competition (Reiter et al. 1992; Echt et al. 1992). In addition, the number of bands amplified by a specific primer is approximately the same in small and large genomes. One possible explanation of this phenomenon is that a genome with low complexity shows on average a lower sequence complementarity between target and primer, whereas a genome with high complexity in general has a better sequence complementarity between target and primer (Williams et al. 1993). Finally, in their study of the inheritance of RAPD markers in F_1 hybrids of maize, Heun and Helentjaris (1993) detected several cases of apparent non-Mendelian inheritance. Taken together, these results indicate the existence of competition in RAPD assays.

The study described here investigates the extent of competition in RAPD reactions and estimates the frequency of errors in the genetic interpretation of the band patterns. If competition is genotype-independent, i.e. the same subset of competing candidates is efficiently amplified in all genomic backgrounds, then competition will not generate errors in the genetic analysis. On the other hand, if competition is genotype-dependent, i.e. the genotype influences which subset of the potential products that will amplify, this will result in errors in the genetic interpretation. The reliability of dominant genetic markers can be tested by screening completely homozygous parents for polymorphisms, making crosses between the parents and subsequently investigating all F_1s for the presence of the predicted heterozygous bands. We have chosen a more general way to test the genetic reliability of RAPDs by using completely homozygous material (doubled haploids) and by producing artificial mixtures by mixing equal amounts of DNA of the different materials. Repeated RAPD analysis enables the discrimination between errors that are due to a lack of reproducibility, i.e. the phenomenon that different band patterns are produced when the same genotype is repeatedly assayed, and those due to competition, which is seen as the repeatable disappearance of an expected band in certain genotypes but not in others.

Materials and Methods

Material and DNA isolation

Four doubled haploid *Brassica napus* lines (Svalöf Weibull AB, Svalöv, Sweden) with known pedigrees were used in all experiments but one. The four lines were named A, B, C and D. DNA from all pairwise combinations of these lines were mixed in equal amounts in order to create artificial 1:1 mixtures. The mixtures were named AB, AC, AD, BC, BD and CD. To repeat the analysis in an organism with a different genomic complexity, we used four strains of *Bacillus cereus* (Swedish Diaries Association, Lund, Sweden) named E, F, G and H. Artificial 1:1 mixtures were made in the same way as for the *B. napus* samples and named EF, EG, EH, FG, FH and GH. Plant DNA was isolated from freeze-dried leaves by the CTAB-method as described by Hjerdin et al. (1994). Bacterial DNA was isolated as described by Ausubel et al. (1989). DNA concentrations were determined using fluorometry with Hoechst 33258 as binding dye.

RAPD analysis

PCR amplifications were performed in either Perkin-Elmer/Cetus 9600 or in MJ Research PTC-100 thermal cyclers. The reaction mixtures contained 80 ng of *B. napus* DNA, 0.4 μ*M* primer (Operon Technologies, Alameda, Calif.), 10 m*M* Tris pH 8.2, 50 m*M* KCl, 0.001% gelatine, 2.0 m*M* $MgCl_2$, 0.1 m*M* dNTP (Pharmacia) and 0.8 units of *Taq* DNA polymerase (Perkin-Elmer/Cetus) in a total volume of 50 μl. The thermal cyclers were programmed for 2 cycles of 60 s at 91°C, 7 s at 42°C and 70 s at 72°C, followed by 43 cycles of 1 s at 91°C, 7 s at 42°C and 70 s at 72°C. To repeat the analysis for a different set of PCR conditions, we used reaction mixtures containing 50 ng *B. napus* DNA, 0.3 μ*M* primer 10 m*M* Tris pH 8.2, 50 m*M* KCl, 0.001% gelatine, 2.5 m*M* $MgCl_2$, 0.1 m*M* dNTP and 1.25 units of *Taq* polymerase.

Amplification parameters were 2 cycles of 60 s at 94°C, 60 s at 36°C and 120 s at 72°C followed by 43 cycles of 30 s at 94°C, 30 s at 36°C and 60 s at 72°C. For DNA samples with low genomic complexity, reaction mixtures contained 12 ng of *B. cereus* DNA, 0.2 μ*M* primer, 10 m*M* Tris pH 8.2, 50 m*M* KCl, 0.001% gelatine, 2.0 m*M* $MgCl_2$, 0.1 m*M* dNTP and 1 unit *Taq* polymerase. Amplification parameters were 2 cycles of 60 s at 94°C, 60 s at 36°C and 120 s at 72°C followed by 43 cycles of 20 s at 94°C, 60 s at 36°C and 120 s at 72°C. DNA fragments were resolved by electrophoresis in 2% agarose gels (1:1 mixture of NuSieve and SeaKem agarose, FMC BioProducts) run at 4.0 V/cm in 1 X TAE for 4.5 h and visualized by ethidium bromide staining.

RFLP analysis

DNA samples from *B. napus* lines A and B were digested with the restriction enzymes *Eco*RI, *Hin*dIII and *Eco*RV, and the DNA fragments separated in 0.8% agarose gels. Southern blotting and hybridization were made as described by Hjerdin et al. (1994). All RAPD bands involved in erroneous situations were isolated from 2% agarose gels by excision and resolved in sterile water by heating for 5 min. These samples were used for reamplification of the band using the PCR conditions described previously. The reamplified bands were once again isolated, resolved in water and stored in a freezer. Radiolabelling of the isolated bands was performed in volumes of 12.5 μl containing 2.5 μl of the template DNA solution, 0.2 μ*M* primer, 10 m*M* Tris pH 8.2, 50 m*M* KCl, 0.001% gelatine, 2.0 m*M* $MgCl_2$, 0.05 m*M* each of dATP, dGTP and dTTP, 2.5 μl α-[^{32}P] dCTP (110 TBq/mmol, Amersham) and 1 unit *Taq* polymerase.

Results

RAPD genotyping errors in artificial DNA mixtures

It is generally believed that the majority of RAPD bands represent unique loci with two alleles, where the presence of a specific band represents the dominant allele and the absence of the same band represents the recessive allele. This model requires that a heterozygote always produces the band. In this experiment our aim is to investigate this aspect of RAPD markers. Instead of using sexual crosses to produce the heterozygotes, we have used DNA from four doubled haploid *Brassica napus* lines (A–D) to make artificial 1:1 mixtures representing all six pairwise combinations, i.e. AB, AC, AD, BC, BD and CD. In each case where a specific band is present in one line but absent in another, the 1:1 mixture of the two lines is the equivalent of a heterozygote. For all such 'heterozygous situations' we determined if the mixtures showed the expected parental bands. If the expected band was absent, this was scored as a 'genotyping error'.

A total of 120 randomly selected polymorphic primers were used to amplify RAPD bands from the pure lines and their mixtures. The primers were chosen from a screen of 466 10-mer primers, where 88 primers failed to amplify products of a sufficient quality for analysis. Of the remaining 378 primers, 167 showed identical patterns, and 211 showed polymorphisms among the four lines. The only criterion used when selecting the 120 primers was that each primer should detect at least 1 scorable polymorphism. The DNA samples were arranged such that each mixture was flanked by its corresponding two pure lines. This arrangement is illustrated in Fig. 1 where a number of polymorphic bands are shown. A total of 190 polymorphic bands were scored, representing 613 heterozygous situations. A number of errors were observed. Of the 120 primers, 40 showed at least one erroneous situation, i.e. absence of an expected band in a mixture. Of the 190 polymorphic bands, 48 showed absence in at least 1 of the mixtures and a total of 84 of the 613 heterozygous situations lacked a band amplified in one of the parental lines. The overall error rate per heterozygous situation was thus 14%. In Fig. 1b and c several genotyping errors are shown. It should be noted that the errors occur for RAPD bands of all intensity classes and can not obviously be explained as artefacts due to the poor reproducibility of faint bands.

The heterogeneity of genotyping errors among primers and among bands was investigated. The unconditional proportion of correct heterozygous situations is 86%. However, given a randomly chosen heterozygous situation that correctly shows an expected band, the probability that another heterozygous situation for the same primer is correct is 90%, whereas given an observed error the probability that another heterozygous situation is correct is only 62%. The corresponding probabilities at the band level is 93% and 44%, respectively, which means that the degree of dependence within bands is greater than the degree of dependence within primers. The overall conclusion to be drawn is that certain primers and bands are more liable to errors than others. We further investigated the pattern of heterogeneity by performing a correlation test. As primers vary with respect to the number of bands they amplify, we calculated the Spearman rank correlation coefficient between the proportion of errors per heterozygous situation and the total number of bands, the number of unpolymorphic bands and the number of polymorphic bands. Only the number of polymorphic bands showed

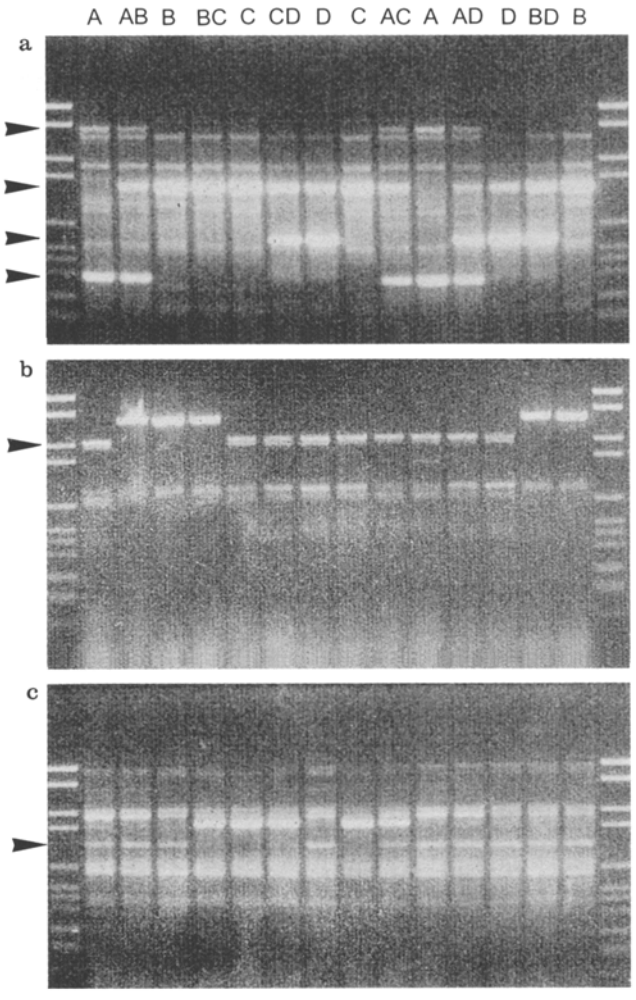

Fig. 1a–c Analysis of RAPD genotyping errors in the 1:1 DNA mixtures of all pairwise combinations of the four *Brassica napus* lines A, B, C, and D. **a** Four polymorphic bands (*arrows*) are successfully amplified with primer OP-G06 in all mixtures where the band is expected. **b** and **c** Examples of genotyping errors. In **b** a band (*arrow*) amplified with primer OP-B13 and originating from lines A, C and D fails to amplify in all mixtures with line B (AB, BC and BD). In **c** a band (*arrow*) amplified with primer OP-B11 and originating from lines A, B and D fails to amplify in mixtures BC and CD but is successfully amplified in mixture AC

a significant correlation to the occurrence of errors. The rank correlation coefficient was 0.45 ($P < 0.0001$). To make a more detailed analysis of this effect we separated the data into three different subsets; primers with only 1 polymorphism, primers with 2 polymorphisms and primers with more than 2 polymorphisms. For each subset, the proportion of errors was calculated (Table 1). It is clearly seen that the occurrence of errors increases with the number of polymorphisms. A Friedman non-parametric test showed that the differences among the three groups were significant ($\chi^2 = 7.0$, $df = 2$, $P = 0.03$).

We also calculated the Spearman rank correlation between the error rate per mixture and the genetic distance calculated as band sharing. Band data from 154 primers, randomly selected among the 378 primers, were used to investigate the genetic distance between the lines. A total of 409 bands were scored, and genetic distances were calculated as $D = 1 -$ (the proportion of shared bands), as described in Halldén et al. (1994), which also confirmed the known relationships among the four lines. No significant correlation was found between error rate and genetic distance. Finally, we made an overall correlation between the size of a DNA fragment and the probability of error for the corresponding band level. The Spearman rank correlation coefficient gave a weak positive correlation (0.15) that was marginally significant ($P = 0.044$).

RAPD genotyping errors using a different set of reaction conditions

To investigate the influence of different PCR conditions on the frequency of errors, we repeated the test above using another protocol. This protocol differs from the one used in the previous experiment in the following way: the duration of the denaturation and primer annealing steps is longer, the template DNA concentration is lower and the primer, $MgCl_2$ and *Taq* DNA polymerase concentrations are higher. The amplification regime used in this experiment is similar to that described by Williams et al. (1990) and, thus, is similar to what is used in most RAPD studies. Three of the *B. napus* lines and their corresponding mixtures were screened for errors with 71 primers. The size of this experiment is consequently only about one-quarter that of the earlier experiment, but since we base our calculations on heterozygous situations the results are comparable. A total of 102 polymorphic bands were amplified. These generated 187 heterozygous situations of which 26 (14%) were erroneous. Thus, this result is very similar to the earlier result and indicates that competition resulting in the generation of errors is not a specific feature of a certain set of PCR conditions.

Table 1 The relative proportion of genotyping errors for different numbers of polymorphisms per primer and for each mixture. Averages are weighted by absolute numbers

Number of polymorphisms per primer	Artificial DNA mixture						
	AB	AC	AD	BC	BD	CD	Average
1	6.7	0.0	15.3	1.9	3.7	5.8	5.7
2	25.0	11.1	23.5	15.4	23.1	15.0	19.1
≥ 3	33.3	34.8	31.6	31.6	30.0	0.0	30.4
Average	17.5	10.3	20.5	11.5	12.2	7.6	13.7

RAPD genotyping errors using template DNA of low genomic complexity

To investigate if this type of error occurs also for a genome with different complexity, we used four strains of *Bacillus cereus* and their 1:1 mixtures and screened for errors as described above. The genome complexity of the haploid procaryote *B. cereus* is several magnitudes lower than that of *B. napus*, which is an allotetraploid eucaryote with a large genome containing many repeated DNA sequences. Using 87 polymorphic primers, we scored a total of 274 polymorphic bands, of which 68 (25%) were erroneous in at least 1 of the mixtures. The polymorphic bands generated 827 heterozygous situations of which 131 (16%) showed genotyping errors. This result is similar to the results obtained in the previous experiments. Thus, the error frequency seems to be independent of the type of template DNA. The degree of dependence within bands was calculated in the same way as for *B. napus*. Given a correct band the probability that another band is also correct is 93%, whereas given an erroneous band the probability is only 38%. The degree of dependence is thus somewhat stronger in *B. cereus* than in *B. napus*.

RAPD band intensity in dilution series

The relationship between template DNA concentration and amount of PCR product (seen as band intensity) was investigated in two experiments. In the first experiment the *B. napus* DNA samples were diluted in water in steps of 10%. No differences in band intensity were detected (Fig. 2a), indicating that a tenfold change in DNA concentration does not per se influence the final amount of PCR product under these amplification conditions. In the second experiment DNA from two lines were diluted into each other in steps of 10%, i.e. artificial DNA mixtures with the proportions 100:0, 90:10, 80:20, 70:30, 60:40, 50:50, 40:60, 30:70, 20:80, 10:90, and 0:100 were made. Thus, the series of DNA concentrations of each individual genome were identical to those in the dilution series with water. This experiment was performed using two combinations of pure lines, AC and CD, that was analysed with 129 and 126 primers, respectively. The result was very different from the experiment in which DNA was diluted in water. All polymorphic bands decreased in intensity as a linear function of their respective proportions in the artificial mixture

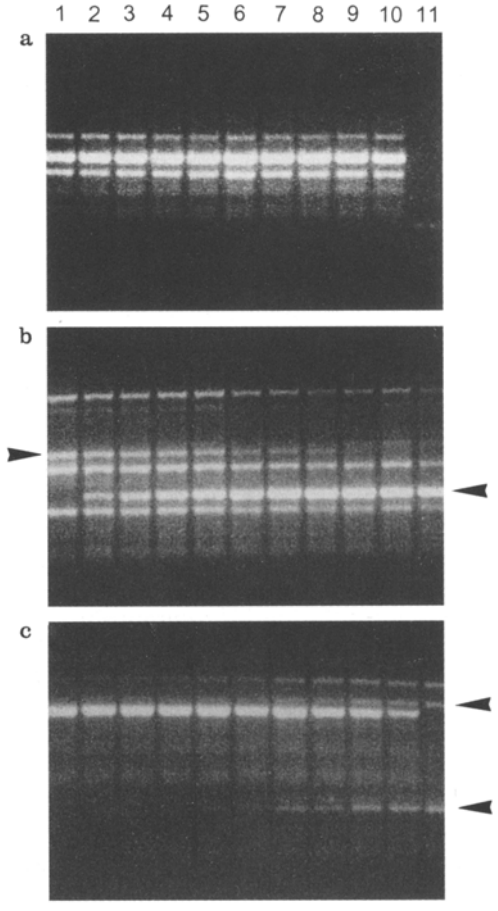

Fig. 2a–c Analysis of RAPD band intensity in dilution series. **a** Dilution of *B. napus* DNA in water. The DNA (1.6 ng/µl) was diluted into water in steps of 10%, from 100% in *lane 1* to 0% in *lane 11*, and amplified with primer OP-N19. No change in RAPD band intensity can be seen in the inverval 100% to 10% DNA. **b** DNA from lines A and C diluted into each other, in the A:C ratios 100:0, 90:10, 80:20, 70:30, 60:40, 50:50, 40:60, 30:70, 20:80, 10:90 and 0:100, and amplified with primer OP-H06. A decrease in RAPD band intensity can be seen for the two polymorphic bands (*arrows*). Both bands are successfully amplified in the 1:1 mixture, thus resulting in a correct genetic interpretation. **c** DNA from lines C and D diluted into each other, as described above, and amplified with primer OP-F20. Two of the polymorphic bands (*arrows*) originating from line D are not amplified in the 1:1 mixture, thus resulting in genotyping errors

(Fig. 2b). The rate of the decrease varied among bands so that some bands disappeared after a small dilution whereas others were visible as strong bands even in the 10:90 mixture. Figure 3 summarizes the information on band disappearance as a function of DNA proportion in the artificial mixture. A number of polymorphic bands were very poorly amplified and failed to amplify the expected product in the 1:1 (50:50) mixture, thus giving rise to a genotyping error. In correspondence with the previous experiment, approximately 15% of all heterozygous situations produced such errors, shown in Fig. 3. This experiment also constitutes a replicate of the initial test for errors. By comparing the heterozygous situations in the two datasets we found that : 85.3% repeatedly showed the expected band in the mixture, 6.2% showed the band in one of the experiments but not in the other and 8.5% did not show an expected band in either experiment.

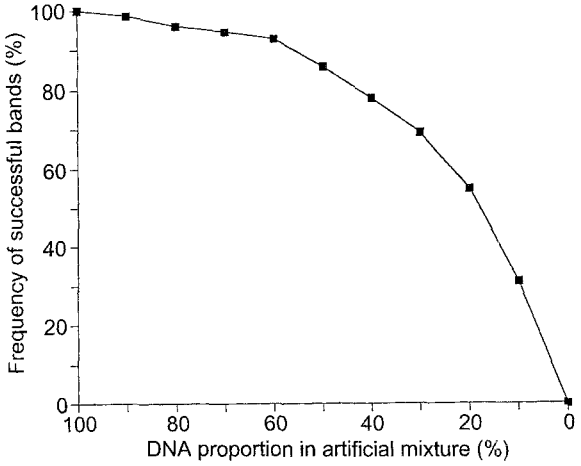

Fig. 3 Frequency of successful RAPD bands as a function of DNA proportion in artificial mixtures. Reduced proportion of DNA in the artificial mixture leads to reduced frequency of successfully amplified RAPD bands (in contrast to dilution of DNA in water). The frequency of bands that disappear in the 1:1 mixtures is approximately 14%

RFLP analysis of RAPD bands

To estimate the genomic copy number of the target DNA sequences, we isolated polymorphic bands involved in error situations and used them as probes in RFLP analysis. The selection of bands included bands that failed to amplify in 1:1 mixtures ('suppressed bands') and bands that appeared to suppress such bands ('successful bands'), i.e. bands that were successfully amplifed while the suppressed bands were not. We also included 11 cases where a single band failed to amplify in a monomorphic background.

A total of 51 polymorphic bands, representing a majority of the previously detected erroneous heterozygous situations, were isolated. The bands were radioactively labelled and probed on membranes containing DNA from lines A and B, and then digested with the restriction enzymes *Eco*RI, *Eco*RV and *Hin*dIII. The DNA sequence copy number was categorized as either high-, low- or single-copy. The results are shown in Table 2. Our a priori expectation was that suppressed bands would show a lower DNA sequence copy number than competing bands, i.e. the success rate of amplification of the band was expected to be positively correlated to the sequence copy number. No such correlation between the DNA sequence copy number and the success of amplification was found. Thus, the actual DNA sequence rather than the sequence copy number seems to be the important determinant for successful amplification.

Table 2 Determination of DNA sequence copy number of successful or suppressed bands

DNA sequence copy number	Band type			
	Successful band	Suppressed band	Suppressed band in monomorphic background	Sum
High copy	4	10	2	16
Low copy	3	4	2	9
Single copy	13	6	7	26
Sum	20	20	11	51

In the screening for errors among the six artificial mixtures, the primer OP-P11 produced an unexpected result (Fig. 4a). This primer amplified a band in the F_1 mixture of lines B and C but not in either of the pure lines. This pattern of inheritance is equivalent to that reported by Hunt and Page (1992) and Ayliffe et al. (1994), where heteroduplex formaton between different allelic RAPD products was shown to occur. To confirm that heteroduplex formation was operating also in this case, we isolated the bands that were specific for the two parental lines and the corresponding bands in the F_1 mixture in addition to the presumed heteroduplex band. RFLP analysis as described above showed exactly the same band patterns for all 5 bands (Fig. 4b), indicating that this non-parental band is indeed created by heteroduplex formation.

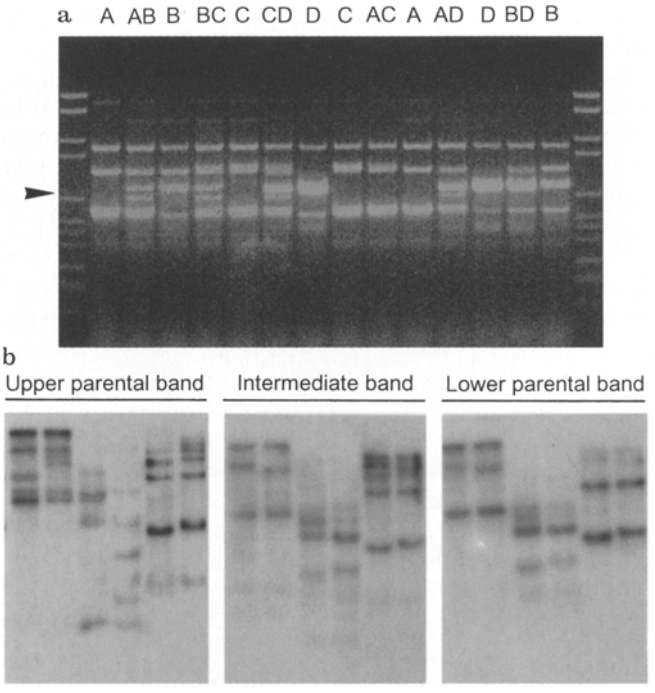

Fig. 4a,b Heteroduplex formation. **a** In mixtures AB, BC, CD and AD a band is present (*arrow*) that does not originate from either of the parental lines but is intermediate in size compared to two polymorphic parental bands. **b** This band showed the same restriction pattern as the two bands in the parental lines, when they were used as probes in RFLP analysis with DNA from lines A and B digested with *Eco*RI, *Hin*dIII and *Eco*RV

Discussion

To estimate the proportion of erroneous heterozygous bands the homozygous lines and their 1:1 mixtures were compared. The observed errors represent both reproducible errors due to strong competition effects and errors due to the poor reproducibility of certain bands generated by less strong competition effects and sample-to-sample differences in reaction conditions. The overall estimate of this error rate is approximately 15% in our study. The consistency of the error rate observed for such diverse genomes as *Brassica napus* and *Bacillus cereus* and for the rather different amplification protocols used in this study indicates that this phenomenon is a general feature of RAPD reactions. In contrast, the formation of heteroduplex molecules as seen by the existence of 1:1 mixture-specific non-parental bands was detected at a frequency of only 0.2%. Other studies have indicated similar low-frequency estimates for heteroduplex formation (Ayliffe et al. 1994; Hunt and Page 1992).

The RAPD protocol described in the first experiment used very fast transitions between the denaturation, annealing and synthesis steps and has been used to facilitate the screening of many samples. The other two protocols are similar to the protocol of Williams et al. (1990) and use amplification regimes of longer duration. All three protocols have been optimized with respect to template, primer, $MgCl_2$, and *Taq* DNA polymerase concentrations. The following criteria were used to specify the conditions of each experiment. The band patterns should not show high-molecular-weight smears and should be invariant in tenfold dilutions with water. The assays were prepared using master mixes and careful manual or robotic multipipetting in order to eliminate as much sample-to-sample variation as possible. Since one of the real benefits of RAPD analysis is that the markers can be used as truly universal markers without prerequisites such as optimization of each primer, and since it is highly impractical to optimize the use of the hundreds of primers involved in the average RAPD

study, we have not optimized the reaction conditions for each individual primer.

The experiment involving dilution series shows that competition affects all bands, even bands that are not involved in situations with genotyping errors. Thus, the effect of competition is quantitative and not qualitative. Instead, the degree of competition experienced by a band in a specific genotype determines the expected copy number of amplified fragments in the final product. It follows that the two types of errors identified above are not independent. If the expected number of amplified fragments is high the band will be reproducibly visible; if the expected number is low the band will be reproducibly invisible; and if the expected number is intermediate the band will be poorly reproducible. In the dilution series such intermediate bands should be the bands that disappear in mixtures with approximately equal amounts of the parental genomes. Dependent on the precission of the labwork, the variation in copy numbers realized for a specific product in repeated assays will be large or small. If the variation is small, then the range of expected copy numbers resulting in unreliable bands will be narrow. From our repeated analysis it was seen that, 6.2% showed the band in one of the experiments but not in the other, whereas 8.5% showed the absence of an expected band in both experiments. This indicates that reproducible errors constitute a significant proportion of the errors that occur in a RAPD reaction. The observation in the dilution series that all bands are affected by competition, in addition to the constancy of the error rate for different genomes and methods, indicates that the phenomenon as such is universal. To our knowledge, the only other experiment where the issue of errors due to competition has been addressed is the study by Heun and Helentjaris (1993). In their study of maize they found, in correspondence with the results of this study, polymorphic bands that were absent in F_1 individuals. However, their results differed quantitatively from ours in that they found a lower frequency of such errors. In addition, Heun and Helentjaris (1993) reported that certain bands are often involved in erroneous situations whereas others appear as expected in all F_1s. This result is consistent with our observation that there is a strong positive dependence within primers and bands for the probability of errors.

RAPD markers are used in many different applications. The consequences of errors due to competition and poor reproducibility depends on the type of application but also on the genetic structure of the material. In general, the problems are less serious in the investigation of haploids or completely homozygous genotypes, whereas heterozygous materials are more problematic. In the following we discuss the effects of errors due to competition in sexual diploids. In linkage mapping, genotyping errors can be partly eliminated by discarding all polymorphisms where the expected RAPD band is absent in F_1's or artificial 1:1 mixtures. Additional errors due to competition may furthermore be identified in the segregating population from distorted segregations. The possibility of doing this depends on the degree of distortion and the number of individuals scored. For example, consider two loci with the alleles *A/a* and *B/b* amplified by the same primer, where the *A* band is successful in competition and the *B* band is suppressed. The strongest effect of competition on the *B* band will occur for genotype *AABb*. The weakest effect of competition will occur for *AaBB* whereas the genotypes with an equal number of alleles for the band, i.e. *AABB* and *AaBb*, ought to be intermediate. Table 3 shows the expected segregation ratios for the *B* band for four different competition scenarios in an F_2 population (e.g. after selfing an *AaBb* individual). The segregation ratios vary widely among the different cases and are also dependent on the recombination frequency, *r*, between the two loci. In all cases, except with genotyping errors in *AABb* genotypes alone, the number of offspring required for a statistical power of 95% in a test of a 3:1 segregation of the band, at a significance level of 5%, is less than 20. Such skewed segregation ratios are identified in all mapping projects. Whenever a locus with a skewed segregation ratio is not identified and discarded, it will cause a bias in the recombination frequency estimates (cf. Säll and Nilsson 1994). In a species with a reasonable number of linkage groups, however, any two randomly chosen loci are most likely unlinked. We therefore believe that situations in which competition passes unnoticed and leads to biased recombination estimates are relatively rare.

One of the most common uses of RAPDs is in bulked segregant analysis (BSA), where a created linkage dis-

Table 3 Expected segregation ratios obtained in different competition scenarios and for different recombination frequencies, *r*, between *A* and *B*. The number of offspring required to obtain a power of 95% in testing 3:1 segregation at a significance level of 5% is shown for each scenario

Genotypes not showing the *B*-band	Expected segregation of *B*-band			Number of offspring for 95% power		
	$r = 0$, coupling	$r = 0$, repulsion	$r = 0.5$ (unlinked)	$r = 0$, coupling	$r = 0$, repulsion	$r = 0.5$ (unlinked)
AABb	3:3	3:1	5:3	–	–	173
A-Bb	1:3	1:3	3:5	9	9	19
A-Bb, AABB	0:1	1:3[a]	5:3	3	9	14
A-B-	0:1	1:3	3:13	3	9	8

[a] The segregation ratio closest to 3:1 for any $0 \leq r \leq 0.5$ occurs for $r = 1/3$ in this case, and is equal to 1:2 (in all other cases the segregation ratio closest to 3:1 for any $0 \leq r \leq 0.5$ is among the values presented)

equilibrium is utilized to accumulate markers for a selected genomic region. This technique is probably quite robust for errors due to competition. Given the terminology defined above, the greatest risk of an error occurs when the *A* band is linked to the region and the *B* band is unlinked. In this case there is a clear risk of observing a false linkage of the *B* band to the region. The two bands will then appear as alleles at a codominant locus. Incidentally, in a BSA project in sugar beet we observed some cases where two apparently codominant bands seemed to be linked to the target locus. In each case, one of the bands was found to be unlinked after confirmation (unpublished results).

RAPD markers have also been extensively used in the analysis of genetic variation. The presence of errors in such applications will generally cause the accessions to appear more variable than really is the case. Estimates of genetic variation will thus be biased upwards and closely related individuals will appear to be more distantly related. The most serious effect of this type of error occurs when RAPDs are used to deduce parentage. It is easy to see that the described effects lead to erroneous exclusion of parentage. A particularly unpleasant property of the competition effect is tllllhat the risk of an incorrect exclusion increases with the number of investigated bands. Moreover if the error is due only to competition, replicating the tests will do nothing to improve the situation.

Acknowledgements We gratefully acknowledge Carina Åkesson and Helena Persson for excellent technical assistance, Bo Gertsson for providing the plant material, Dr. J. Nilsson for providing bacterial strains and Dr. M. Höglund for critical reading of the manuscript. This work was supported by the Erik Philip-Sörensen Foundation.

References

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (1989) Current protocols in molecular biology, vol 1. Green Publishing Assoc & Wiley Interscience, New York

Ayliffe AM, Lawrence GJ, Ellis JG, Pryor AJ (1994) Heteroduplex molecules formed between allelic sequences cause nonparental RAPD bands. Nucleic Acids Res 22:1632–1636

Carlson JE, Tulsieram LK, Glaubitz JC, Luk VWK, Kauffeldt C, Rutledge R (1991) Segregation of random amplified DNA markers in F_1 progeny of conifers. Theor Appl Genet 83:194–200

Delourme R, Bouchereau A, Hubert N, Renard M, Landry BS (1994) Identification of RAPD markers linked to a fertility restorer gene for the *Ogura* radish cytoplasmic male sterlity of rapeseed (*Brassica napus* L.) Theor Appl Genet 88:741–748

Echt CS, Erdahl LA, McCoy TJ (1992) Genetic segregation of random amplified polymorphic DNA in diploid cultivated alfalfa. Genome 35:84–87

Giovannoni JJ, Wing RA, Ganal MW, Tanksley S (1991) Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. Nucleic Acids Res 19:6553–6558

Halldén C, Nilsson NO, Rading IM, Säll T (1994) Evaluation of RFLP and RAPD markers in a comparison of *Brassica napus* breeding lines. Theor Appl Genet 88:123–128

Heun M, Helentjaris T (1993) Inheritance of RAPDs in F_1 hybrids of corn. Theor Appl Genet 85:961–968

Hjerdin A, Säll T, Tuvesson S, Halldén C (1994) RFLP markers in the genus *Beta*: characterization of DNA sequences from a *Beta vulgaris* library. Genetica 92:91–99

Hu J, Quiros CF (1991) Identification of broccoli and cauliflower cultivars with RAPD markers. Plant Cell Rep 10:505–511

Hunt GJ, Page RE (1992) Patterns of inheritance with RAPD molecular markers reveal novel types of polymorphism in the honey bee. Theor Appl Genet 85:15–20

Lamboy WF (1994) Computing genetic similarity coefficients from RAPD data: correcting for the effects of PCR artefacts caused by variation in experimental conditions. PCR Methods Appl 4:38–43

Mailer RJ, Scarth R, Fristenski B (1994) Discrimination among cultivars of rapeseed (*Brassica napus* L.) using DNA polymorphisms amplified from arbitrary primers. Theor Appl Genet 87:697–704

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA 88:9828–9832

Penner GA, Bush A, Wise R, Kim W, Domier L, Kasha K, Laroche A, Scoles G, Molnar SJ, Fedak G (1993) Reproducibility of random amplified polymorphic DNA (RAPD) analysis among laboratories. PCR Methods Appl 2:341–345

Reiter RS, Williams JGK, Feldmann KA, Rafalski JA, Tingey SV, Scolnik PA (1992) Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs. Proc Natl Acad Sci USA 89: 1477–1481

Riedy MF, Hamilton WJ, Aquadro FC (1992) Excess of non-parental bands in offspring from known primate pedigrees assayed using RAPD PCR. Nucleic Acids Res 20:918

Stiles JI, Lemme C, Sondur S, Morshidi MB, Manshardt R (1993) Using randomly amplified polymorphic DNA for evaluating genetic relationships among papaya cultivars. Theor Appl Genet 85:697–701

Säll T, Nilsson NO (1994) The robustness of recombination frequency estimates in intercrosses with dominant markers. Genetics 137:589–596

Tinker NA, Fortin MG, Mather DE (1993) Random amplified polymorphic DNA and pedigree relationships in spring barley. Theor Appl Genet 85:976–984

Weeden NF, Timmerman GM, Hemmat M, Kneen BE, Lodhi MA (1992) Inheritance and reliability of RAPD markers. In: Hoisington D, McNab A (eds) Proc Symp Appl RAPD Technol Plant Breed, Crop Sci Soc Am, Minneapolis, pp 12–17

Welsh J, McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res 18:7213–7218

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535

Williams JGK, Hanafey MK, Rafalski JA, Tingey SV (1993) Genetic analysis using random amplified polymorphic DNA markers. Methods Enzymol. 218:705–740