Am. J. Hum. Genet. 62:1189–1197, 1998

A Simple Method for Analyzing Microsatellite Allele Image Patterns Generated from DNA Pools and Its Application to Allelic Association Studies

Johanna Daniels,[1,2] Peter Holmans,[2] Nigel Williams,[1,2] Dragana Turic,[1,2] Peter McGuffin,[2] Robert Plomin,[3] and Michael J. Owen[1,2]

Divisions of [1]Psychological Medicine and [2]Medical Genetics, University of Wales College of Medicine, Cardiff; and [3]Centre for Social, Genetic and Developmental Psychiatry, Institute of Psychiatry, London

Summary

Allelic association studies provide the most powerful method for locating genes of small effect contributing to complex diseases and traits. However, in outbred populations, allelic association is usually maintained only over distances of $\leqslant 1$ cM. Therefore, systematic searches over large regions are costly. Here we present a method involving DNA pooling that can be used as a rapid preliminary screen for allelic association with the most common class of polymorphic markers, single-sequence repeats. Patient and control samples are pooled separately, and markers are typed in the two pools. By use of primers with fluorescent 5′ ends, PCR products can be analyzed on an automated sequencing apparatus. Allele image patterns (AIPs) produced for the two groups are overlaid and differences in pattern area between pools computed. From this, a ΔAIP statistic is calculated from the difference in areas between the two AIPs expressed as a fraction of the total shared and nonshared area. AIPs of a range of different-sized pools were generated by computer simulation for markers with a range of allele sizes and frequencies. ΔAIPs from pools and χ^2 values for individual genotypings were compared, with both simulated and real data from microsatellite markers. The results demonstrated a high correlation between ΔAIP and χ^2 values. ΔAIP analysis of real microsatellite data indicated the feasibility of using this method in systematic searches for allelic association and generated a small number of false positives but few false negatives. We conclude that ΔAIP analysis of DNA pools can be used effectively and efficiently as a rapid screen for allelic association in case-control studies.

Received March 19, 1997; accepted for publication March 10, 1998; electronically published April 17, 1998.

Address for correspondence and reprints: Dr. M. J. Owen, Neuropsychiatric Genetics Unit, Tenovus Building, University of Wales College of Medicine, Heath Park, Cardiff CF4 4XN. E-mail: Owenmj@cardiff.ac.uk

0002-9297/98/6205-0025$02.00

Introduction

Allelic association studies are a powerful way of locating genes for complex traits (Owen and McGuffin 1993; Risch and Merikangas 1996). They can be used both for fine mapping susceptibility loci as part of positional cloning strategies and for examining candidate genes. Large-scale systematic searches for allelic association have not been carried out, to date, because of the belief that they are not technically feasible, since dense marker maps and large numbers of subjects are required. An attractive solution may be to use sample pooling to reduce the amount of genotyping required (Daniels et al. 1995).

DNA pooling has been used successfully in quantitative trait locus (QTL) association studies in animals. Here it has been combined with "selective genotyping," whereby subjects are genotyped only if their scores on the trait under study lie at either extreme of the continuum. DNA from "high" and "low" groups are then pooled, which allows a considerable saving in genotyping (Asada et al. 1994; Darvasi and Soller 1994). DNA pooling has also been used to study recessive diseases in inbred human populations (Carmi et al. 1995). In both of these cases, a shift toward homozygosity of particular markers was sought in pooled DNAs. So far, DNA pooling has not been successfully applied to case-control association studies in outbred populations. There are several potential problems with the application of pooling to the commonest form of human genetic marker, single-sequence repeats (SSRs). Two of these result from PCR artifacts commonly associated with SSRs—namely, stutter banding and differential amplification—which confound attempts to estimate allele frequencies in pooled samples. In the present paper, we report that these difficulties can be circumvented by a simple method using commercially available software, which can be used to conduct case-control, or QTL, association studies in humans using di-, tri-, and tetranucleotide repeats.

Stutter bands are produced by the amplification of products one or two repeat units shorter than the cor-

rect-sized amplimere because of slippage of *Taq* polymerase on the repeated sequence. When pooled samples are studied, these bands summate with the correct-sized amplimeres of alleles one or two repeat units smaller. This confounds attempts to estimate the frequencies of different alleles directly from the amplified products of the pooled samples. Stutter is prevalent in dinucleotide repeats, the most common form of STS marker, but it occurs to a lesser extent in tri- and tetranucleotide repeats. Differential amplification is observed in heterozygotes and is caused by the preferential amplification of the smaller allele. This is observed in all classes of microsatellites and is believed to be caused by the larger alleles reannealing at a faster rate (since they contain more repeat units), which reduces the efficiency of PCR amplification (Demers et al. 1995).

Several groups using different approaches have attempted to estimate allele frequencies of microsatellite markers from pooled samples. Khatib et al. (1994) selected dinucleotide microsatellites that did not produce stutter bands and predicted allele frequencies and standard errors from pools amplified with the selected markers. Pacek et al. (1994) also chose markers, a tetranucleotide and 16-bp repeat, that did not produce stutter bands, to demonstrate that allele frequencies can be accurately estimated from pools. However, they observed differential amplification, which, in the resulting analysis of pooled genotypes, gave moderately inflated estimates of the smallest alleles and underestimates of the frequency of larger alleles.

LeDuc et al. (1995) developed a mathematical method to correct for stutter, based on measurement of the ratio between the amounts of true product and stutter bands for each marker. Reasonable estimates of allele frequencies were obtained for one marker, but comparison between groups was not attempted.

Perlin et al. (1995) described a more accurate method for the mathematical correction of stutter. The relative amount of product for the true amplimere and stutter bands was calculated individually for each allele of a marker and then used to compose a matrix. This was used to transform the results from a pool in which many alleles were represented. This method was more sophisticated than its predecessors because it took into account the observation that each allele of a particular marker has its own unique stutter pattern. Barcellos et al. (1997) have combined this method with simple mathematical correction for differential amplification and obtained promising data from four markers. However, each individual marker required a lengthy work-up, which would hinder large-scale studies for which many markers will be needed.

Our aim was to develop a simple method, which would not require lengthy analysis of each marker prior to genotyping, for the application of DNA pooling to large-scale association studies. The approach we have taken is to use pooling as a screening step to identify markers for individual genotyping. The first method we evaluated was to perform simple mathematical corrections for stutter and differential amplification. As we demonstrate, modifications of data by such imperfect methods cause both type I and type II errors to occur in association studies where statistical comparisons of allele frequencies between two samples are made. We therefore reasoned that, in these instances, in which the goal is not to estimate absolute allele frequencies for a single group, it might be better simply to compare the uncorrected products of pooled SSR amplification. We were encouraged by the observation that when SSRs are amplified in pooled DNA samples, the images they produce when analyzed by GENESCAN/GENOTYPER software on the ABI fluorescent system are highly stable and replicable. Here we describe a simple method for statistical comparisons of these images that allows the problems of stutter and differential amplification to be ignored and that can be readily used as a rapid screen in case-control association studies.

Material and Methods

Genotyping

High–molecular weight DNA was extracted from lymphocytes according to routine procedures. All samples were quantified on a Beckman spectrophotometer and diluted to 8 ng μl^{-1}. Pools consisted of 10 μl of each sample at 8 ng μl^{-1} and were constructed using a Kemble Instruments Guardian SPII sample processor. PCR was carried out in 96-well microtitre plates in a volume of 12 μl containing 200 μM dNTPs, 12 pmol of each primer (5′ fluorescently labeled for amplification of pools, 5′ radiolabeled with ^{33}P for genotyping of individual samples; termed "individual genotyping" hereafter), 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 48 ng of pooled DNA, and 2 U *Taq* polymerase. Cycling conditions were 94°C for 30 s, annealing temperature specific to the marker for 30 s, and 72°C for 30 s, for 28 cycles. For generation of allele image patterns, 0.15 μl of each PCR reaction was loaded onto a 6% denaturing gel and run at 13 W on an ABI 373 DNA sequencer. All gels were analyzed by use of GENESCAN software and were then imported into GENOTYPER. When individual genotyping was performed, allele sizes were determined by comparison to M13mp18 sequencing ladders on 6% denaturing polyacrylamide gels. All gels were scored by two independent raters.

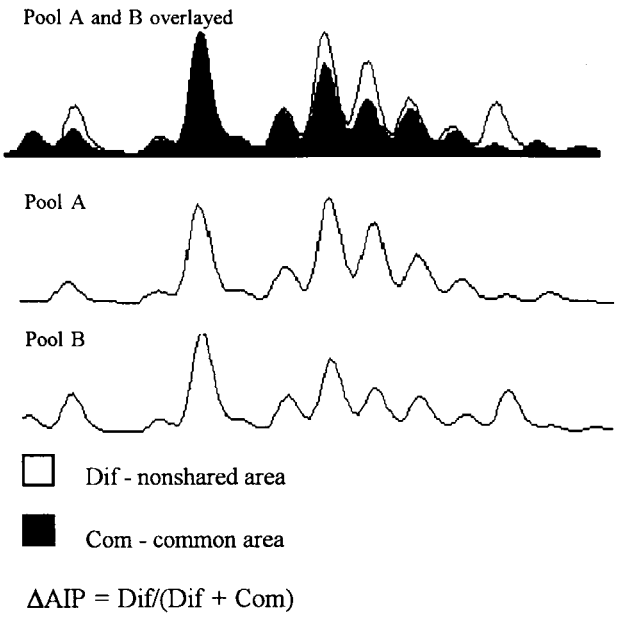


Figure 1 Overlaid AIPs of two different pools amplified with the marker D6S1666. Areas "Dif" and "Com" are the nonshared and common areas, respectively, between the two AIPs.

Mathematical Correction for Stutter and Differential Amplification

The area of each peak from an allele image pattern can be obtained from GENOTYPER. However, the relative peak areas are not accurate reflections of the true allele frequency because the tails of neighboring peaks overlap and because of two PCR artifacts, stutter banding and differential amplification. Stutter is greater for dinucleotides with secondary (α) and tertiary (β) bands ~50% and 25% the area of the main amplimere, although α and β are not constant across all alleles of a given marker (Perlin et al. 1995). In a pool consisting of a range of alleles, stutter bands will therefore contribute a large proportion of the area of peaks one or two repeat units smaller. An approximate correction for this artifact was attempted by estimating the relative areas of α and β bands from a number of homozygous individuals and by serially subtracting their contribution to smaller peaks, starting from the largest allele in the pool. No attempt was made to determine allele-specific stutter patterns.

Tri- and tetranucleotide repeats produce less stutter banding. However, they are still subject to differential amplification, often to a greater extent than dinucleotides, since they have a greater allele size range. To correct approximately for differential amplification, we measured its effect by analyzing a range of heterozygotes for each marker. The relative area of the larger peak compared with that of the smaller was then correlated with the difference in size between the two alleles, in base pairs. The area of every allele peak relative to the smallest was then calculated. Using this information, we determined an expression that corrects for the peak area of an allele dependent on its distance (in base pairs) from the smallest allele in the pool. To correct for both sources of artifact, first the correction for stutter and then that for differential amplification were applied to the area of each peak of the allele image pattern obtained from GENOTYPER.

Allele Image Analysis

Accuracy and reproducibility of pool construction and PCR were controlled for by constructing triplicate pools and duplicating the PCR of each pool. The six resulting pooled images were then overlaid in GENOTYPER, which scales all traces so that the height of the largest peak is equal in each. The baseline was determined automatically in GENOTYPER, and the images were not modified in any way, prior to analysis. The consensus image was taken to be representative of the allele frequencies in that pool. In >90% of marker/pool combinations, all six patterns aligned completely. In occasional cases, one—or, rarely, two—pool(s) misaligned. We attributed this to PCR failure rather than to errors in pool construction, since the same pools amplified perfectly for other markers. We have now aligned pools amplified with 200 different marker/pool combinations, and in only 11 cases did the pools misalign. When this occurred, the remaining four or five images that did align were taken as representative of the allele image pattern. The consensus images for the two pools to be compared were overlaid in GENOTYPER and were stored by use of the Apple Macintosh screen capture facility. The pooled allele image was isolated in SimpleText (Apple Computer) and then imported into the graphics manipulation software Debabelizer (Equilibrium) for analysis.

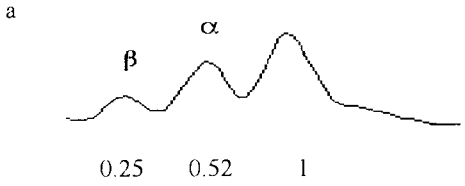
Calculation of ΔAIP

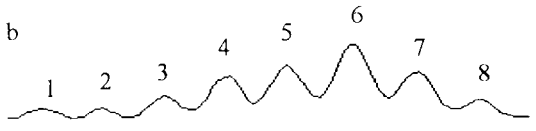
When comparing the results of pooled genotyping of two samples, we measured the total area that was not shared by the two superimposed consensus allele image patterns irrespective of how many times the curves from the two samples crossed. This was then expressed as a fraction of the total shared and nonshared area. This test statistic was called ΔAIP (allele image pattern difference). Calculation of ΔAIP is illustrated in figure 1. The area "Com" is common to both allele image patterns. The nonshared area is denoted "Dif." Areas Com and Dif were measured in Debabelizer (Equilibrium) by shading the shared and nonshared areas in different colors and calculating the pixel count for each. The ΔAIP test statistic was calculated from the expression Dif/(Dif + Com).

Simulation Procedure

It is not necessarily the case that the largest ΔAIP values will correspond to the most statistically significant differences. This is because the value of ΔAIP will be influenced by factors such as the number of marker alleles and the number of cases and controls in the sample. To investigate how ΔAIP depends on these factors, it was necessary to use computer simulation, since the large number of replicate samples necessary made actual genotyping impractical.

Allele image patterns from pooled samples were simulated in the following way. The contribution of each allele to the overall distribution was modeled as a normal distribution centered on the true size of the allele and with an SD of 0.5 bp. The effect of stutter bands was modeled by adding two further normal distributions with the same SD but 50% and 25% of the height, centered at 2 bp and 4 bp less than the true allele size. The overall allele-size distribution was obtained by summing the contributions of the alleles and their stutter bands. Although certainly an oversimplification, this method gave a reasonable approximation to real allele image patterns and was probably sufficient to allow the effect of variables such as sample size and allele number on ΔAIP to be determined. The allele-size distributions of case and control populations were normalized as in GENOTYPER by making the heights of the largest peaks equal. Dif and Com were measured by numerical integration, and the ΔAIP test statistic was calculated as above.





Allele	1	2	3	4	5	6	7	8
PA1	1600	1409	3047	5665	7636	10318	6846	2861
PA2	775	433	1379	1214	2624	7298	6131	2861

Figure 2 *a*, Stutter bands α and β produced for a homozygous individual amplified with the marker D6S309. α and β are, respectively, 0.52 and 0.25 times the relative area of the true allele. *b*, AIP generated by amplifying a pool of 100 individuals with D6S309. The table at the bottom of the figure shows the difference in peak area before ("PA1") and after ("PA2") mathematical correction for stutter.

Results

Mathematical Correction for Stutter and Differential Amplification

Figure 2 shows typical shadow bands observed for the marker D6S309. The areas of the secondary and tertiary peaks are 0.52 and 0.25 of the area of the true allele. Curve *b* shows the allele pattern for a pool of 100 individuals, amplified with D6S309. The peak areas of each allele are shown before (PA1) and after (PA2) correction for stutter. The peak areas of alleles have been reduced by various amounts. Alleles 4 and 5 show marked reductions in area, which indicates that most of the original total was due to stutter bands from peaks 6 and 7. After correction, allele 3 has a greater peak area than does allele 4.

Table 1 shows the allele frequencies calculated by applying both the stutter and the differential amplification corrections to five different pools. The pools were constructed from 100 individuals of known genotypes, for the marker D20S112. Statistical differences between true allele frequencies (individually genotyped) and those estimated from pooled samples (corrected) were calculated by use of the program CLUMP (Sham and Curtis 1995) and are also shown in table 1. Discrepancies between true and estimated allele frequencies are dependent on the composition of individual pools. Correction for differential amplification and stutter has worked well for pool 3 but not for pools 1 or 2; estimated allele frequencies in pool 2 differ by $\leqslant 55\%$ from the true values obtained from individual genotyping. The estimated allele frequencies and those obtained from individual genotyping were then compared between pools. Comparisons of the estimated data suggest that the allele frequencies in pools 1 and 2 do not differ significantly ($P = .959$), when in fact the true difference is significant at the 1% level ($P = .01$). Conversely, there is a significant difference in estimated genotypes between pools 2 and 3 ($P = .008$) that is not present on individual genotyping ($P = .392$). On the basis of these and other examples (data not shown), we conclude that the simple methods we have used to correct for stutter and differential amplification do not provide sufficiently accurate estimates of allele frequencies for use in allelic association studies: statistical comparisons of allele frequencies estimated from pooled genotypes lead to unacceptably high type I and II error rates.

Table 1

Estimated and True Allele Frequencies in Five Pools of Known Composition for Marker D20S112

	Pool 1		Pool 2		Pool 3		Pool 4		Pool 5	
	Estimated	True	Estimated	True	Estimated	True	Estimated	True	Estimated	True
Allele 1	15	20	12	10	10	10	0	0	0	0
Allele 2	43	30	41	30	32	30	22	20	24	20
Allele 3	10	20	11	10	9	10	20	20	16	10
Allele 4	6	10	9	20	30	30	41	40	44	50
Allele 5	9	10	10	20	6	10	5	10	3	10
Allele 6	17	10	16	10	13	10	13	10	14	10
Estimated vs. true	$P = .087$		$P = .053$		$P = .933$		$P = .689$		$P = .144$	

NOTE.—Allele frequencies estimated from pooled samples by mathematical correction compared with true allele frequencies for five different pools. Values of *P* refer to statistical differences between true and estimated allele frequencies calculated by use of the program CLUMP (Sham and Curtis 1995). See text for further details.

Pool Reproducibility

A striking finding was the stability and reproducibility of the allele image patterns for each pool. Pools were constructed in triplicate and amplified in duplicate, so that six images were generated for each marker. Figure 3 shows the six allele image patterns overlaid for three dinucleotide repeat markers. It is evident that there is almost 100% alignment between the six images produced for each marker, and this clearly demonstrates the reproducibility of both the pooling process and of PCR amplification of pooled DNA. When we followed this procedure for 11 randomly chosen markers, we found that a mean of 94% (range 88%–99%) of the area under the curves was common to all six traces.

Simulation Results

Test criteria were calculated by simulating 5,000 pairs of case and control allele patterns from the same population. The criterion for a test of significance level 5% was estimated by the 250th largest of the ΔAIP statistics. This was done for markers encompassing a wide range of numbers of alleles and allele frequencies and for case and control samples containing 25, 50, 100, 200, and 400 individuals. It was found that, in general, test criteria depended on the informativity of the marker (measured here by the PIC): the higher the PIC, the higher the criteria. In addition, markers with larger numbers of alleles required higher criteria. As expected, test criteria were reduced for larger samples. The results are not shown in detail. However, for a marker with PIC of 0.5, it was found that a ΔAIP of 0.2 gave an ~5% test for samples containing 50 cases and 50 controls, a ΔAIP of 0.15 was required for 100 cases and controls, and a ΔAIP of 0.1 for 200 cases and controls. For a marker with PIC of 0.7, the corresponding ΔAIP values are 0.25, 0.2, and 0.15, and for a marker with PIC of 0.9, they are 0.35, 0.27, and 0.2. Simulation studies also showed that, for a given relative risk, both ΔAIP and Pearson χ^2 analysis of individual genotypes are generally less powerful when the associated allele is of a very high or very low frequency but that the two methods did not differ greatly in this respect.

Since the significance of a ΔAIP value depends on the number of alleles, and their frequencies, a more accurate estimate of the *P* value may be obtained by simulating case and control samples from a population with allele frequencies estimated from the peak heights of the control sample (a program for this purpose is available from P.H. on request). Such "simulated *P* values" were obtained for all the analyses of real marker data presented in this article. However, since we recommend that the

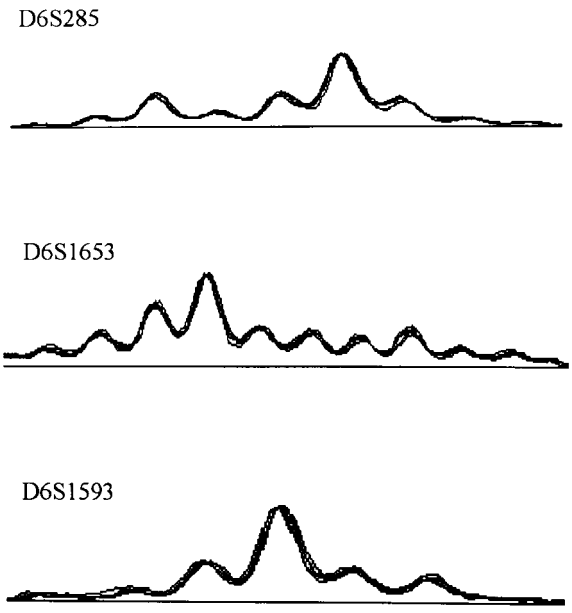


Figure 3 DNA from 100 individuals was pooled in triplicate. Each of the three pools underwent PCR amplification in duplicate. Data from three microsatellite markers (D6S285, D6S1653, and D6S1539) are shown. This process generated six AIPs for each marker (3 pools × 2 PCRs). Shown are the results of overlaying the six AIPs for each marker.

Table 2

Comparison of Individual and Pooled Genotyping

Marker	No. of Alleles	ΔAIP	ΔAIP *P* Value	χ^2 *P* Value
DRD5	13	0.43	<0.001	0.024
D6S260	20	0.49	<0.001	0.163[a]
D6S309	13	0.22	0.110	0.107
D6S296	20	0.14	0.924	0.197
D6S1279	10	0.13	0.133	0.193

NOTE.—For each marker, two sample groups each of $n = 100$ (except D6S260, for which $n = 50$ for both groups) were compared by use of both ΔAIP comparison of pooled samples and following individual genotyping. Values of *P* for ΔAIP were obtained by simulating case and control samples from a population with allele frequencies estimated from the peak heights of the control sample (a program for this purpose is available from P.H. on request). χ^2 values for individual genotypes were calculated by use of CLUMP (Sham and Curtis 1995).

[a] A *P* value of .036 was obtained by comparison of the 162-bp allele with the remaining alleles combined.

ΔAIP method be used simply as an initial screening procedure prior to individual genotyping, precise significance levels will not usually be required.

Comparisons between Individual and Pooled Genotyping

To illustrate use of the ΔAIP statistic, we report data from five markers, one tetranucleotide (D6S1279) and four dinucleotide repeats (D6S309, D6S296, D6S260, and DRD5), using pools made from real groups of experimental subjects. D6S1279, D6S309, D6S296, and DRD5 were amplified from the same patient and control pools, each containing 100 individuals. D6S260 was amplified from a second pair of pools, each containing 50 individuals. The ΔAIP test statistic was applied to overlaid allele images of patient and control pools for each of the markers. These markers were also genotyped individually in all patient and control samples. The statistical difference in allele frequencies obtained by individual genotyping between patient and control groups was calculated by use of CLUMP (Sham and Curtis 1995). Table 2 shows the corresponding ΔAIP, its *P* value, and *P* values from CLUMP analysis of individual genotypes obtained for each marker. Significant differences between cases and controls were detected by ΔAIP analysis of pools and were confirmed with individual genotyping for DRD5 and D6S260. ΔAIP analysis of the remaining three markers failed to show significant differences between the two groups, and this was also confirmed by individual genotyping.

In a second series of experiments, we constructed 13 pools from 100 individuals who had been typed with the marker D6S1279. The pools were made such that the smallest, middle, and largest alleles of the pool were represented at different frequencies (10%–50% at 10% intervals). Table 3 shows ΔAIP values calculated for differences among 22 pairs of pools and the corresponding χ^2 values calculated from statistical comparisons of the individual genotypes. It can be seen that the ΔAIP method successfully detected all the differences among pools. In four instances, differences were detected that were not present on individual genotyping. These data are shown graphically in figure 4, which also shows the correlation between ΔAIP and χ^2 values. The value of 0.69 is within the confidence intervals obtained from simulation studies.

We developed the ΔAIP method for use in systematic searches for linkage disequilibrium. Therefore, to test this method in a more realistic setting, we conducted a study to determine whether it would detect the known linkage disequilibrium between hemochromatosis and markers at 6p21.3 (Raha-Chowdhury et al. 1995). Forty closely spaced microsatellite markers spanning ~41 cM of chromosome 6p24-21.1 were genotyped in pools constructed from 30 patients with hemachromatosis and from 30 ethnically matched controls. All markers were dinucleotide repeats, apart from D6S1034, D6S1263, and D6S1006 (trinucleotides), and D6S1955 and D6S1279 (tetranucleotides). (DNA from hemachroma-

Table 3

Detection of Differences among Pools

ΔAIP	χ^2	Simulated ΔAIP *P* Value	χ^2 *P* Value
.098	4.00	.528	.261
.468	12.86	<.001	<.001
.471	24.93	<.001	<.001
.683	40.00	<.001	<.001
.412	2.77	<.001	.428
.407	10.00	<.001	.019
.584	21.04	<.001	<.001
.042	2.34	.917	.505
.359	9.00	<.001	.029
.392	2.22	<.001	.528
.560	4.00	<.001	.261
.720	24.93	<.001	<.001
.840	41.49	<.001	<.001
.350	22.52	<.001	<.001
.200	3.61	.034	.306
.010	4.00	.996	.261
.430	24.93	<.001	<.001
.400	4.00	<.001	<.001
.420	10.00	<.001	.019
.590	21.04	<.001	<.001
.160	2.22	.179	.528

NOTE.—Thirteen pools were constructed from 100 individuals and amplified with the marker D6S1279, such that the smallest, middle, and largest alleles of the pool were represented at different frequencies (10%–50% at 10% intervals). ΔAIP values are calculated for differences between 22 pairs of pools, and the corresponding χ^2 values are calculated from statistical comparisons of the individuals' genotypes by use of CLUMP (Sham and Curtis 1995). Values of *P* for ΔAIP were obtained as described in the note to table 2.

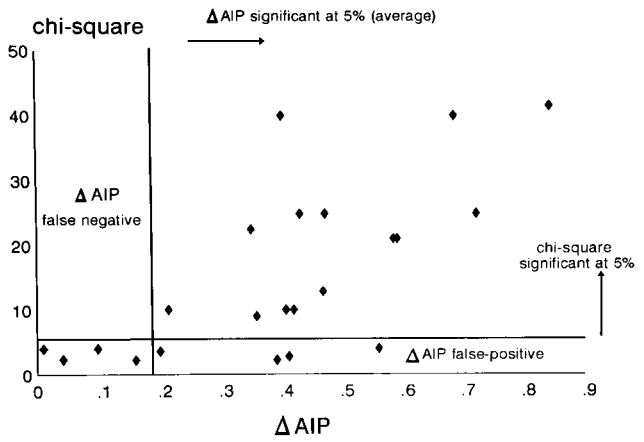


Figure 4 Data from table 3 displayed graphically. Shown is the correlation between ΔAIP and χ^2 values for 22 comparisons between pools of 100 individuals, with marker D6S1279. Correlation coefficient = 0.69 (0.40, 0.86). Criteria for ΔAIP vary slightly between pools; the average is shown.

tosis patients was kindly supplied by M. Worwood). Pools were made and the ΔAIP method applied according to the methods described above. Individual genotyping of hemochromatosis case and control individuals was undertaken with the seven markers that yielded a ΔAIP simulated P value of <.05. The results are shown in table 4.

It can be seen that the ΔAIP pooling method successfully detected the previously observed linkage disequilibrium between hemochromatosis and the 122-bp allele of marker D6S265 (Raha-Chowdhury et al. 1995), as well as association with the closely adjacent marker D6S1558. Interestingly, D6S1691, which lies only 1.6 cM the other side of D6S265, also gave a ΔAIP simulated P value of <.05, although the results of individual genotyping failed to reach statistical significance. This, together with the finding that individual genotyping of D6S265 just failed to reach statistical significance at the .05 level, confirms the results of simulation studies (data not shown) and suggests that, under some circumstances, ΔAIP may be more powerful than χ^2 analysis of individual genotypes. The remaining five markers, for which ΔAIP analysis of pooled genotypes gave a $P <$ .05, did not show significant differences between cases and controls on individual genotyping.

These results show that ΔAIP analysis is sufficiently sensitive to detect moderate associations and that, although its false-positive rate may be slightly higher than its nominal value, it is not excessively so (here, five false positives out of 40 tested loci).

Discussion

We initially addressed the problem of PCR artifact by attempting to estimate allele frequencies in pooled DNA samples after correction for both stutter and differential amplification (Daniels et al. 1995). LeDuc et al. (1995) independently applied a similar method to correct for stutter (but not for differential amplification) to estimate allele frequency of microsatellite markers in DNA pools. Our data suggest that simple correction for PCR artifact is liable to cause errors (both type I and type II) when estimated allele frequencies are compared between two groups. It should be noted that the simple methods used by ourselves and by LeDuc et al. (1995) did not take into account the fact that the relative heights of stutter bands vary across alleles of a given marker—nor did we correct for the fact that the shape of allele peaks in GENOTYPER is not normal with the result that the tails of peaks overlap with their neighbors. It is therefore possible that more sophisticated mathematical correction might result in accurate estimation of allele frequencies from analysis of pools by GENOTYPER (Perlin et al. 1995; Barcellos et al. 1997). However, this is likely to require extensive data collection, to allow determination of allele-specific stutter patterns and the degree of differential amplification for each marker prior to the analysis of pooled samples.

The novel approach that we have taken is effectively to ignore the problems of stutter and differential amplification by directly comparing the allele image patterns for pooled cases and pooled controls. This is possible because the allele image patterns obtained for pools are highly reproducible. We have developed the ΔAIP test statistic, which describes the differences in allele image patterns between pools. This is highly correlated with χ^2 values that are obtained by genotyping all individuals. Simulations have demonstrated that the power of ΔAIP to detect allele differences between pools is comparable with that of χ^2 analysis of individual genotypes. However, care must be taken with the interpretation of the test criteria because simulations were based on realistic approximations to stutter bands and differential amplification, but it is possible that different values of these could lead to different results. Therefore the test criteria given above are intended as a rough indication only.

When this method was applied to real data, we found that, in general, significant ΔAIP values occurred when χ^2 obtained from individual genotyping were significant. We found a number of cases in which the ΔAIP was significant and the χ^2 was not; these may be regarded as false positives. This confirms our view that the most appropriate use of the ΔAIP DNA pooling method in association studies is as part of a rapid initial screen in which a liberal test criterion is employed (say, α of <.05 or .10) to identify markers for individual genotyping. It does not appear that false positives can be identified from the magnitude of the ΔAIP statistic (table 4 and unpublished data), but we are currently working on var-

Table 4

Linkage Disequilibrium Mapping of Hemochromatosis

Marker	cM[a]	ΔAIP	ΔAIP *P* Value	χ^2
D6S1674	...	.26	.08	
D6S309	.0	.13	.78	
D6S277	.6	.28	.15	
D6S1034	4.1	.25	.20	
D6S1263	3.1	.20	.06	
D6S1955	1.0	.14	.42	
D6S1006	2.1	.09	.96	
D6S1279	1.0	.19	.27	
D6S1593	1.9	.16	.58	
D6S1653	.2	.15	.29	
D6S259	1.2	.20	.52	
D6S1605	3.6	.20	.46	
D6S1584	1.7	.18	.51	
D6S1700	1.8	.16	.61	
D6S285	.0	.09	.39	
D6S422	1.5	.15	.69	
D6S1665	.7	.29	.09	
D6S1597	1.6	.11	.60	
D6S1588	.1	.35	.08	
D6S1686	1.2	.19	.14	
D6S1660	.9	.23	.13	
D6S1554	1.8	.23	.08	
D6S1691	.7	.39	.01	.19
D6S265	1.6	.37	.01	.06[b]
D6S1558	.0	.28	.02	.02
D6S1666	.5	.24	.25	
D6S1568	2.2	.31	.04	.88
D6S439	.6	.29	.07	
D6S1629	.2	.26	.07	
D6S291	1.4	.10	.48	
D6S1602	2.6	.07	0.95	
D6S1548	.2	.29	.03	.84
D6S1641	1.5	.27	.04	.33
D6S1562	4.1	.22	.13	
D6S1616	1.4	.15	.79	
D6S426	1.1	.44	.01	.65
D6S1549	1.6	.17	.28	
D6S1582	2.9	.44	.72	
D6S282	1.2	.03	1.00	

NOTE.—Forty microsatellite markers spanning ~41 cM of chromosome 6p24-21.1 were genotyped in pools constructed from 30 patients with hemachromatosis and from 30 ethnically matched controls.

[a] Approximate distance between markers (ftp://ftp.Genethon.fr/pub/gmap/nature-1995 and http://cedar.genetics.soton.ac.uk/public_html/index.html). Pools were made and the ΔAIP method applied according to the methods described above. Individual genotyping of hemochromatosis case and control individuals was undertaken with the eight markers that yielded a ΔAIP simulated *P* value of <.05. χ^2 values for comparison of individual genotypes from patients and control subjects were calculated by use of CLUMP (Sham and Curtis 1995).

[b] 122 bp allele vs. rest, $P = .016$. Marker D6S265 is the approximate location of the hemachromatosis gene.

ious technical improvements, such as the elimination of nontemplated nucleotide addition, to reduce the false-positive rate. However, in large-scale studies in which the burden of individual genotyping resulting from false positives is likely to be high, we suggest the introduction of a second stage in which markers that are positive in the first pooling stage undergo pooled genotyping in a second sample of patients and controls. Only markers for which ΔAIP is significant in stage 2 are then subjected to individual genotyping. We have noted that, rarely, false negatives may occur when the associated allele is large, because of the effects of differential amplification. However, we believe that this problem can be avoided if the overlaid AIPs are inspected at the time of ΔAIP calculation. If differences are apparent in the portion of the AIPs deriving from larger alleles, then individual genotyping should be undertaken as a precaution.

Putative linkages for complex traits often span large genetic distances. ΔAIP facilitates a rapid screen of regions of interest for allelic association. In the situation in which a region of interest spans 20 cM, $\geqslant$21 markers would have to be typed to cover the interval at a spacing of 1 cM, a reasonable marker density for identifying allelic association. With the ΔAIP method, only 12 PCRs (six each for patient and control pools) are needed for each marker. Eight markers can be run per lane on an ABI 373/377 (because of the different sizes of the amplimers and the differential labeling of the forward primers with the three fluorophores available); hence, all markers can be run on a typical 36-lane ABI gel. Therefore, the screen for the whole 20-cM region can be run and analyzed on one gel. In comparison, standard genotyping of 100 patient and control individuals with 21 markers would require running $\geqslant$84 gels. Analysis is also far simpler for the ΔAIP method, since allele calling (which, in our hands, has an error rate of ~1%) is not required and the ΔAIP test statistic is calculated from unmanipulated raw data.

To date,, we have worked with pools of $\leqslant$100 individuals, but there is no reason why our method should not work well with pools made from a larger number of subjects, since the sample-size–limits of DNA pooling are constrained only by the total number of DNA molecules in the PCR reaction (~10,000 in the present study). However, the optimal number of individuals for pool construction will depend not only on technical constraints but also on other factors, such as the power required and the relative need to constrain type 1 versus type 2 errors.

The power of ΔAIP seems to be less affected than that of Pearson χ^2 by a large number of alleles. When the number of alleles is large, however, a method such as CLUMP (Sham and Curtis 1995), applied to the individual genotypes, may well be more powerful than the Pearson χ^2 statistic used here. A limitation of the ΔAIP method is that, although it can predict overall differences, it does not indicate which alleles are associated with the difference. However, this can be ascertained when following up with individual genotyping.

A biological approach to reducing stutter and differ-

ential amplification would make the allele image patterns easier to analyze and the ΔAIP test statistic more accurate. Stutter can be reduced to some extent by higher-fidelity *Taq* polymerase that is less prone to slippage on the repeat units. Preferential amplification of the smaller allele can be reduced by use of peptide nucleic acid (PNA) oligomers that can sufficiently block DNA template-inhibiting reassociation while allowing primer extension. PNA has been used with great efficiency in the amplification of a 16-bp repeat microsatellite, D1S80, amplifying the larger allele with 4.2 times greater efficiency (Demers et al. 1995). This is not, however, a perfect solution to the problem of differential amplification. The concentration of PNA added to the PCR would have to be optimized for every marker, and the degree of enhancement of different alleles of the same marker is variable. Therefore, until mathematical correction methods are sophisticated enough to deal with the complexities of amplification of many alleles in pools, the ΔAIP test statistic will be the simplest and most powerful approach to screening for frequency differences between pools.

The development of this method means that we can now look forward to genomewide allelic association studies that until now have not been considered feasible (Risch and Merikangas 1996). Assuming a sex-averaged genome size of 3,700 cM, ~3,500 markers will be required for a genome scan at 1-cM intervals. Conventional genotyping of 400 patient and control samples would require 2,800,000 genotypes. This is reduced to just 7,000 by use of the ΔAIP method. For larger sample sizes, the relative efficiency of DNA pooling is even greater.

The ΔAIP DNA pooling method facilitates a rapid screen for allelic association over large genetic distances and has the capacity to handle large sample sizes while reducing costs in terms of time and consumables. Its application should greatly reduce the next rate-limiting step in the dissection of complex traits, the fine mapping of putative loci in large critical regions.

Acknowledgments

This work was supported by U.S. National Institute of Child Health and Human Development grant HD-27964 and by U.K. Medical Research Council grant G9309834.

References

Asada Y, Varnum DS, Frankel WN, Nadeau JH (1994) A mutation in the Ter gene causing increased susceptibility to testicular teratomas maps to mouse chromosome 18. Nat Genet 6:363–368

Barcellos LF, Klitz W, Field L, Tobias R, Bowcock AM, Wilson R, Nelson MP, et al (1997) Association mapping of disease loci by use of a pooled DNA genomic screen. Am J Hum Genet 61:734–747

Carmi R, Rokhlina T, Kwitek-Black AE, Elbedour K, Nishimura D, Stone ES, Sheffield VC (1995) Use of human DNA pooling strategy to identify a human obesity syndrome locus on chromosome 15. Hum Mol Genet 4:9–13

Daniels JK, Williams NM, Plomin R, McGuffin P, Owen MJ (1995) Selective DNA pooling using automated technology: a rapid technique for allelic association studies in quantitative traits and complex diseases. Psychiatr Genet Suppl 5: S39

Darvasi A, Soller M (1994) Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. Genetics 138:1365–1373

Demers DB, Curry ET, Egholm M, Sozer AC (1995) Enhanced PCR amplification of VNTR locus D1S80 using peptide nucleic acid (PNA). Nucleic Acids Res 23:3050–3055

Khatib H, Darvasi A, Plotski Y, Soller M (1994) Determining relative microsatellite allele frequencies in pooled DNA samples. PCR Methods Appl 4:13–18

LeDuc C, Miller P, Lichter J, Parry P (1995) Batched analysis of genotypes. PCR Methods Appl 4:331–336

Owen MJ, McGuffin P (1993) Association and linkage: Complementary strategies for complex disorders. J Med Genet 30:638–639

Pacek P, Sajantila A, Syvanne A-C (1993) Determination of allele frequencies at loci with length polymorphism by quantitative analysis of DNA amplified from pooled samples. PCR Methods Appl 2:313–317

Perlin MW, Lancia G, Ng S-K (1995) Toward fully automated genotyping: genotyping microsatellite markers by Deconvolution. Am J Hum Genet 57:1199–1210

Raha-Chowdhury R, Bowen DJ, Stone C, Pointon JJ, Terwilliger JD, Shearman JD, Robson KJH, et al (1995) New polymorphic microsatellite markers place the haemochromatosis gene telomeric to D6S105. Hum Mol Genet 4:1869–1874

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516–1517

Sham P, Curtis D (1995) Monte Carlo test for association between disease and alleles at highly polymorphic loci. Ann Hum Genet 59:97–105