

Substrate Nucleotide-Determined Non-Templated Addition of Adenine by *Taq* DNA Polymerase: Implications for PCR-Based Genotyping and Cloning

BioTechniques 21:700-709 (October 1996)

V.L. Magnuson, D.S. Ally, S.J. Nylund, Z.E. Karanjawala, J.B. Rayman, J.I. Knapp, A.L. Lowe¹, S. Ghosh and F.S. Collins

National Center for Human Genome Research, National Institutes of Health, Bethesda, MD, and ¹Perkin-Elmer/Applied Biosystems Division, Foster City, CA, USA

ABSTRACT

The Applied Biosystems PRISM[®] fluorescence-based genotyping system as well as the Invitrogen TA Cloning[®] vector system are influenced by the tendency of *Taq* DNA polymerase to add an adenine nucleotide to the 3' end of PCR products after extension. Incomplete addition of adenine to a majority of PCR product strands creates problems in allele-calling during genotyping and potentially diminishes the cloning efficiency of such products. Experiments reported here show that certain terminal nucleotides can either inhibit or enhance adenine addition by *Taq* and that PCR primer design can be used to modulate this activity. The methods we propose can substantially improve allele-calling for problematic microsatellite markers when using GENOTYPER[®] software.

INTRODUCTION

Complex disease studies require a genome scan in order to search for the multiple genes involved. Multiplexing of fluorescently labeled microsatellite markers has been used successfully in both linkage and linkage disequilibrium mapping (2,3). The Applied Biosystems PRISM[™] fluorescence-based genotyping system (Perkin-Elmer/Applied Biosystems Division [PE/ABI], Foster City, CA, USA) offers a high throughput, automatable alternative to radioactive labeling of microsatellite markers. However, for a significant fraction of microsatellite markers, the ability to make unambiguous allele assignments is compromised by *Taq* stutter patterns and non-templated addition of adenine (1) ("allele plus A" or "allele + A"). The GENOTYPER[™] software package (PE/ABI) can detect and label the highest peak in a given pattern of bands including stutter bands and allele + A as long as a set of conditions are met (8). The labeling and filtering algorithm used in GENOTYPER is most successful for patterns generated when polymerase chain reactions (PCRs) result in, almost entirely, allele only or allele + A only.

Protocols that attempt to drive PCRs either to mostly allele or to mostly allele + A by using 2-step PCR and 3-step PCR plus a 90-min extension at 72°C (Table 1), respectively, have been used with success for many markers (9). However, the classic 3-step PCR plus 10-min extension at 72°C protocol (Table 1) has proven inadequate to drive many PCR products entirely to al-

lele + A (9). This can result in an undesirable peak profile where true allele and allele + A are nearly identical in height and GENOTYPER may label the allele peak in one individual whereas in another individual the allele + A peak will be labeled. This miscalling results in non-Mendelian inheritance and allelic binning problems, and therefore requires manual surveillance of every genotype.

During the course of product testing the ABI PRISM Linkage Mapping Set (PE/ABI), we tested 90 primer pairs from 7 marker panels. A survey was conducted to determine for each primer pair whether the 2-step PCR protocol or the 3-step + 90 min at 72°C protocol (Table 1) best amplified the given marker. Ideally, the 2-step protocol is optimal when products proceed to ≤30% allele + A, and the 3-step + 90 min at 72°C protocol is optimal when the reaction proceeds to ≥70% allele + A. Unfortunately, for 10% of the primer pairs, intermediate outcomes were still obtained, and therefore a more thorough understanding of the process was needed. For purposes of discussion, markers whose primer pairs amplified most consistently with a 2-step protocol will be referred to as "2-step markers" and those that amplified most consistently with the 3-step + 90 min at 72°C protocol will be called "3-step markers".

Because of the desire for a consistent PCR protocol across all markers, shorter cycling times and ease of storage, we investigated the possibility that there existed a sequence-specific component influencing the addition of

Table 1. PCR Protocols

2-Step PCR	3-Step PCR + 10-min Extension at 72°C	3-Step PCR + 90-min Extension at 72°C
95°C, 5 min	95°C, 5 min	95°C, 5 min
94°C, 15 s	94°C, 15 s	94°C, 15 s
55°C, 15 s	55°C, 15 s	55°C, 15 s
(10 cycles)	72°C, 30 s	72°C, 30 s
89°C, 15 s	(10 cycles)	(10 cycles)
55°C, 15 s	89°C, 15 s	89°C, 15 s
(23 cycles)	55°C, 15 s	55°C, 15 s
	72°C, 30 s	72°C, 30 s
	(20 cycles)	(20 cycles)
	72°C, 10 min	72°C, 90 min

adenine to the 3' end of PCR products by *Taq* DNA polymerase. Since the 5' end of the forward primer carries the fluorescent label, it is only this strand that is detected by the ABI Model 373A DNA Analysis System (PE/ABI). The sequence at the 3' end of this extended, labeled strand serves as a substrate for adenine addition by *Taq*. The terminal nucleotide of this strand is determined by the 5' end of the reverse primer used in generating the complementary unlabeled strand. Figure 1 illustrates the relationship between reverse primer sequence and terminal substrate nucleotide.

Examination of reverse primer sequences corresponding to markers that were considered to be "2-step markers" (30/90) revealed that 23/30 (76.7%) of the sequences began with "T" at the 5' end (the Y position in Figure 1). As a

result, the majority of the labeled PCR product strands had an "A" residue as the terminal nucleotide (the X position in Figure 1). On the other hand, the 5' end of the reverse primers corresponding to the larger group (60/90) of typically "3-step markers" contained only 6/60 (10%) "T"s (15/60 or 25% expected by random chance), and therefore mostly "non-A" terminal substrate nucleotides. This led to the hypothesis that, in the presence of an "A" residue as substrate nucleotide (the X position in Figure 1), *Taq* DNA polymerase may be inhibited from adding an additional "A" residue (allele + A). These observations are consistent with previous experiments of Hu (5) that showed that the "terminal extendase" activity of some prokaryotic and eukaryotic DNA polymerases is affected by the type of terminal nucleotides present. These

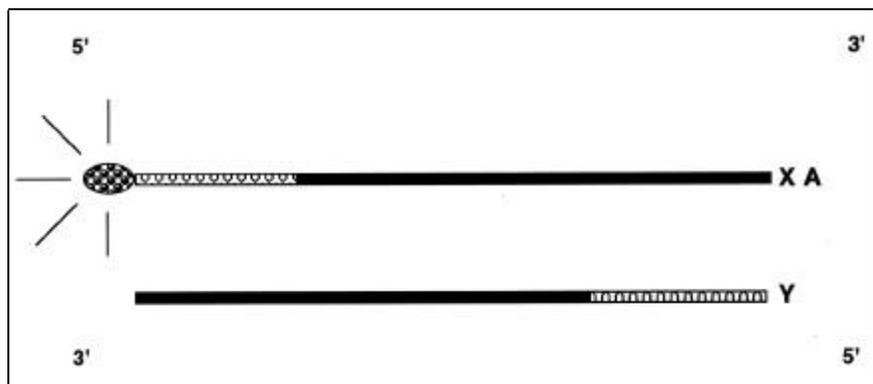


Figure 1. Terminal nucleotide (X) for *Taq* DNA polymerase non-templated addition of "A" is determined by 5'-end nucleotide (Y) of reverse primer. Spotted ellipse represents a fluorescent label on the labeled PCR product strand. Primers are indicated by open decorated boxes.

primer extension studies of annealed complementary oligos containing single-base trinucleotide overhangs (i.e., CCC, GGG, etc.) revealed that *Taq* DNA polymerase preferred to add an "A" nucleotide after extending "CCC" terminals and was least likely to add "A" after extending "AAA" termini (5).

In this study, we evaluated the effect of "A" vs. "non-A" substrates in the X position on "A" addition by *Taq*. Reverse primers were redesigned for "2-step markers" to promote the addition of "A" by *Taq*. This was done by replacing the 5'-end Y-position nucleotide (usually "T") of the reverse primer with "A". Additionally, reverse primers were redesigned for "3-step markers", to deter the addition of "A" by *Taq*, by replacing the 5'-end Y-position nucleotide (usually "non-T") of the reverse primer with "T".

MATERIALS AND METHODS

Reverse Primer Design and PCR Amplification

The fluorescently labeled dinucleotide repeat markers utilized in this study were from ABI PRISM Linkage Mapping Set market-ready Panels 1, 2, 11, 21, 22 and R&D Panels 8 and 28 (PE/ABI). Primer pairs for the following 39 microsatellite markers were used: DXS1106; D16S405; D16S415; D16S420; D16S423; D16S516; D16S520; D1S228; D1S234; D8S272; D1S235; D5S419; D5S422; D5S433; D5S644; D6S260; D6S309; D7S510; D7S513; D7S516; D7S669; D8S258; D8S556; D1S206; D1S218; D1S238; D1S498; DXS1001; DXS986; D15S117; D15S127; D15S130; D15S165; D16S401; D16S511; DXS1060; DXS1204; DXS1214; DXS1227. Twenty-seven reverse primers corresponding to "2-step markers" were redesigned by substituting the nucleotide on the 5' end with an "A" nucleotide (Y-position in Figure 1). Twelve reverse primers corresponding to "3-step markers" were redesigned by substituting a "T" for the 5' end nucleotide (Y-position in Figure 1).

DNAs were amplified in MicroAmp® 96-well plate systems (Perkin-Elmer, Norwalk, CT, USA). Each of 24

Research Reports

Table 2. The Effect of Reverse Primer Sequence Change in Forcing PCR Products Amplified by "2-Step" Primers from Allele to Allele + A

Marker	Original Reverse Primer				Modified Reverse Primer			
	X-nucleotide X	2-Step only	3-Step + 10-min ext. at 72°C	3-Step + 90-min ext. at 72°C	X-nucleotide X	2-Step only	3-Step + 10-min ext. at 72°C	3-Step + 90-min ext. at 72°C
DXS1106	...AAA	0.23	0.24	0.42	...AAT	0.49	0.64	0.85
D16S405	...TCA	0.23	0.32	0.62	...TCT	0.31	0.44	0.80
D16S415	...CTA	0.25	0.38	0.73	...CTT	0.41	0.70	0.82
D16S420	...CAA	0.23	0.44	0.68	...CAT	0.55	0.81	0.85
D16S423	...AAA	0.33	0.45	0.76	...AAT	0.69	0.80	0.80
D16S516	...CCA	0.33	0.47	0.62	...CCT	0.47	0.62	0.75
D16S520	...AGG	0.29	0.35	0.68	...AGT	0.54	0.76	0.85
D1S228	...CTA	0.28	0.64	0.75	...CTT	0.54	0.86	0.86
D1S234	...ACA	0.28	0.39	0.69	...ACT	0.46	0.65	0.74
D8S272	...CCA	0.21	0.35	0.66	...CCT	0.37	0.46	0.79
D1S235	...CTG	0.30	0.42	0.53	...CTT	0.43	0.66	0.73
D5S419	...GCA	0.23	0.24	0.53	...GCT	0.35	0.50	0.86
D5S422	...CTG	0.26	0.34	0.68	...CTT	0.33	0.49	0.71
D5S433	...AAA	0.19	0.52	0.77	...AAT	0.62	0.92	0.93
D5S644	...CAA	0.21	0.37	0.60	...CAT	0.58	0.81	0.86
D6S260	...GAA	0.29	0.51	0.49	...GAT	0.74	0.85	0.83
D6S309	...AAG	0.31	0.43	0.80	...AAT	0.57	0.83	0.88
D7S510	...GAA	0.28	0.72	0.77	...GAT	0.71	0.81	0.82
D7S513	...TGG	0.28	0.49	0.72	...TGT	0.48	0.79	0.73
D7S516	...AGA	0.32	0.41	0.66	...AGT	0.58	0.73	0.77
D7S669	...AAA	0.26	0.58	0.65	...AAT	0.64	0.77	0.82
D8S258	...CAA	0.32	0.31	0.46	...CAT	0.42	0.66	0.69
D8S556	...GCA	0.16	0.34	0.86	...GCT	0.40	0.78	0.90
D1S206	...TGG	0.24	0.53	0.71	...TGT	0.40	0.79	0.80
D1S218	...AAA	0.30	0.45	0.70	...AAT	0.37	0.49	0.60
D1S238	...ATG	0.38	0.48	0.62	...ATT	0.56	0.69	0.72
D1S498	...GCA	0.23	0.29	0.69	...GCT	0.37	0.71	0.86
Mean (27)		0.27	0.42	0.66		0.50	0.70	0.80

Terminal substrate nucleotides (the X position in Figure 1) for original and modified reverse primers are shown in bold type. Values are mean allele + A ratio $\{(\text{allele} + A)/([\text{allele}] + [\text{allele} + A])\}$ for PCR cycling condition. Means across 27 markers are presented at the end of the table and plotted in Figure 2.

marker primer pairs were amplified in the same 96-well plate for a given PCR cycling protocol (Table 1). Each marker was tested on three human genomic DNAs from CEPH family 884 (Bios Laboratories, New Haven, CT, USA) and CEPH individual 1347 02 (Bios Laboratories). All PCRs were performed using GeneAmp® PCR System 9600 (Perkin-Elmer) using Perkin-Elmer reagents according to the protocol described in ABI PRISM Panel inserts (7) with the 2-step, 3-step +

10-min or 3-step + 90-min extension at 72°C PCR protocols illustrated in Table 1. Briefly, the following reagents (PE/ABI) were combined to total a final volume of 15 µL: 1.0 µL forward and reverse primer mixture (5 µM); 1.2 µL human genomic DNA (50 ng/µL); 1.5 µL Perkin-Elmer 10× PCR Buffer II; 1.5 µL dNTP mixture (2.5 mM); 0.9 µL MgCl₂ (25 mM); 8.78 µL sterile deionized water; and 0.12 µL AmpliTaq® (5 U/µL; Perkin-Elmer).

Because of the quantitative nature of

the study, all PCR products were frozen immediately at -20°C after cycling was complete. This was to avoid any extra "A" addition that might occur after cycling. The samples remained frozen until pooled. Depending on dye, either 2, 4 or 6 µL of the PCR product samples (FAM:HEX:TET ratio equaling 1:2:1) were pooled together (maximum 12 per lane) in a final volume of 40 µL. A 1.5-µL aliquot of the pool was added to 3.5 µL of a 1:1:5 mixture of GS500 size standard, loading dye and formamide

(PE/ABI), denatured for 5 min at 95°C and loaded onto a 7% denaturing (8 M urea) polyacrylamide gel (Bio-Rad Laboratories, Hercules, CA, USA). The gel was 0.3 mm in thickness and the plates were 12 cm well-to-read. All gels were run on the ABI 373A “stretch” model sequencer (PE/ABI) for 6 h at 15 W (constant and limiting). The GENESCAN™ Analysis 672 Software Package Version 1.2 (PE/ABI) was utilized in conjunction with GENOTYPER Version 1.1b for PCR product analysis.

Analysis

Allele and allele + A peaks were determined by the peak height value (fluorescence units) given by the GENOTYPER program or by a graphical estimation of peak height 1 bp away from the appropriate labeled peak. Major peak heights (allele or allele + A) of less than 100 U were not used. Homozygote and distantly spaced heterozygote alleles were easiest to score for allele and allele + A. Heterozygote alleles spaced 2 bp apart were the most inaccurate due to overlapping of the area of the fluorescent tail from the first

allele peak into the stutter pattern of the second allele peak (see Results and Discussion). Obtaining the actual allele or allele + A components would require deconvolution of overlapping peak areas. Therefore, an overestimation of the amount of allele and/or allele + A present occurs in genotypes of this configuration.

The ratio of allele + A PCR products to the total number of PCR products $\{(\text{allele} + \text{A}) / ([\text{allele}] + [\text{allele} + \text{A}])\}$ was calculated for each allele of a given genotype. The mean allele + A ratio across 4 genotypes (maximum 8 alleles, 16 data points) was calculated for each marker primer pair (original and modified) at each PCR cycling condition (Tables 2 and 3). The mean across 27 “2-step marker” mean allele + A ratios (Table 2), the mean across 12 “3-step marker” mean allele + A ratios (Table 3) and the standard error of the mean (SEM) were calculated using the StatView 512+™ software package (BrainPower, Calabasas, CA, USA) and plotted (Figures 2 and 4) using the CricketGraph III™ software package (Computer Associates International, Islandia, NY, USA).

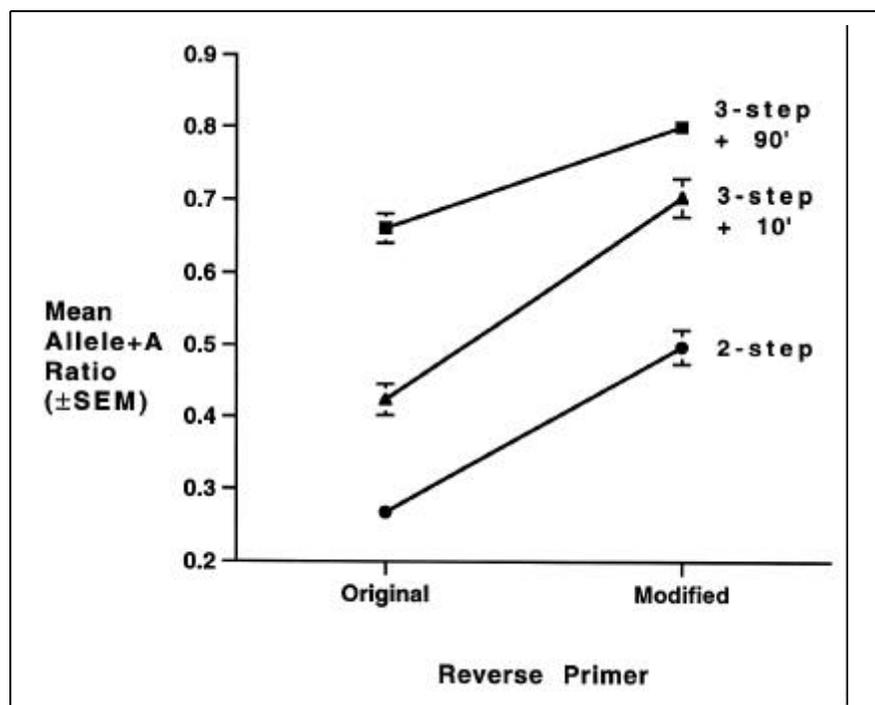


Figure 2. Changing reverse primer Y-nucleotide to “A” and terminal substrate X-nucleotide to “T” increases the proportion allele + A PCR products amplified under three different cycling protocols. Values are the mean of 27 mean allele + A ratios $\{(\text{allele} + \text{A}) / ([\text{allele}] + [\text{allele} + \text{A}])\}$ calculated in Table 2 \pm SEM. All error bars were plotted; those not visible are contained within plot symbols.

Research Reports

RESULTS AND DISCUSSION

Various parameters affect the non-templated addition of adenine to the PCR products by *Taq* DNA polymerase. For example, experiments have shown that changing the PCR cycling conditions (2-step vs. 3-step + 90 min at 72°C) can improve the performance of some primer pairs (9). Decreasing the amount of time that *Taq* is exposed to 72°C during cycling seems to decrease the opportunity for *Taq* to add on adenine (9). There are exceptions, however, and 2-step cycling may not be desirable for the generation of longer templates. For some PCR products, extended time at 4°C or room temperature is enough to allow the PCR to proceed further towards allele + A (5%–25% after 2 days at 4°C, unpublished observations). In fact, 2-step PCR products must be frozen if the allele is to remain the predominant product. Furthermore, our experience has shown that one PCR protocol (or even two) could not reproducibly extend all primer pairs to primarily the desired product allele or allele + A. Therefore, the rules governing *Taq*'s ability to add on adenine were investigated.

Conversion of “2-Step Markers” to “3-Step Markers”

In a survey of 90 primer pairs from the ABI PRISM Linkage Mapping Set, 30/90 were found to be “2-step markers”. Out of these 30, 23 of the reverse primers began with a “T” at the 5' end, resulting in an “A” as the substrate nucleotide on the 3' end of the labeled PCR product strand. The remaining 7/30 began with a “C”, resulting in a terminal “G” on the complementary strand. Experiments were designed to test the hypothesis that the presence of an adenine on the 3' end of a PCR product is inhibitory to adenine addition by *Taq* DNA polymerase. In this experiment, a 5' end “C” or “T” nucleotide was replaced with an “A” in the reverse primer for 27 of the “2-step markers”, resulting in a substrate nucleotide of “T” on the labeled PCR product strand. As the terminal substrate nucleotide, “T” was not expected to be inhibitory to allele + A addition by *Taq*. Thus, the change was expected to convert the “2-step markers” to “3-step markers”.

Table 2 shows a comparison between the mean allele + A ratio in PCR products amplified with the original reverse primer and the modified reverse primer across the three PCR protocols. The three terminal nucleotides encoded by the reverse primers, including the X-position nucleotide in Figure 1 (bold), are also shown. The removal of “A” (and “G”) as terminal substrate nucleotide markedly enhances the ability of the “2-step markers” to proceed to $\geq 70\%$ allele + A. While only 1/27 of the original reverse primers resulted in $\geq 70\%$ allele + A PCR products under the 3-step + 10 min at 72°C protocol, 16/27 of the modified reverse primers

resulted in $\geq 70\%$ allele + A PCR products under those same conditions. In addition, by using the 3-step + 90 min at 72°C protocol, 10/27 of the original reverse primers vs. 25/27 of the modified reverse primers resulted in $\geq 70\%$ allele + A PCR products. Even by using the 2-step protocol, 2/27 of the modified reverse primers yielded PCR products $\geq 70\%$ allele + A.

Figure 2 is a summary plot of the effect of changing the X-position nucleotide to “T” for “2-step markers”. The effect of the change in reverse primer sequence in increasing the proportion of allele + A PCR products is equivalent, on the average, to an

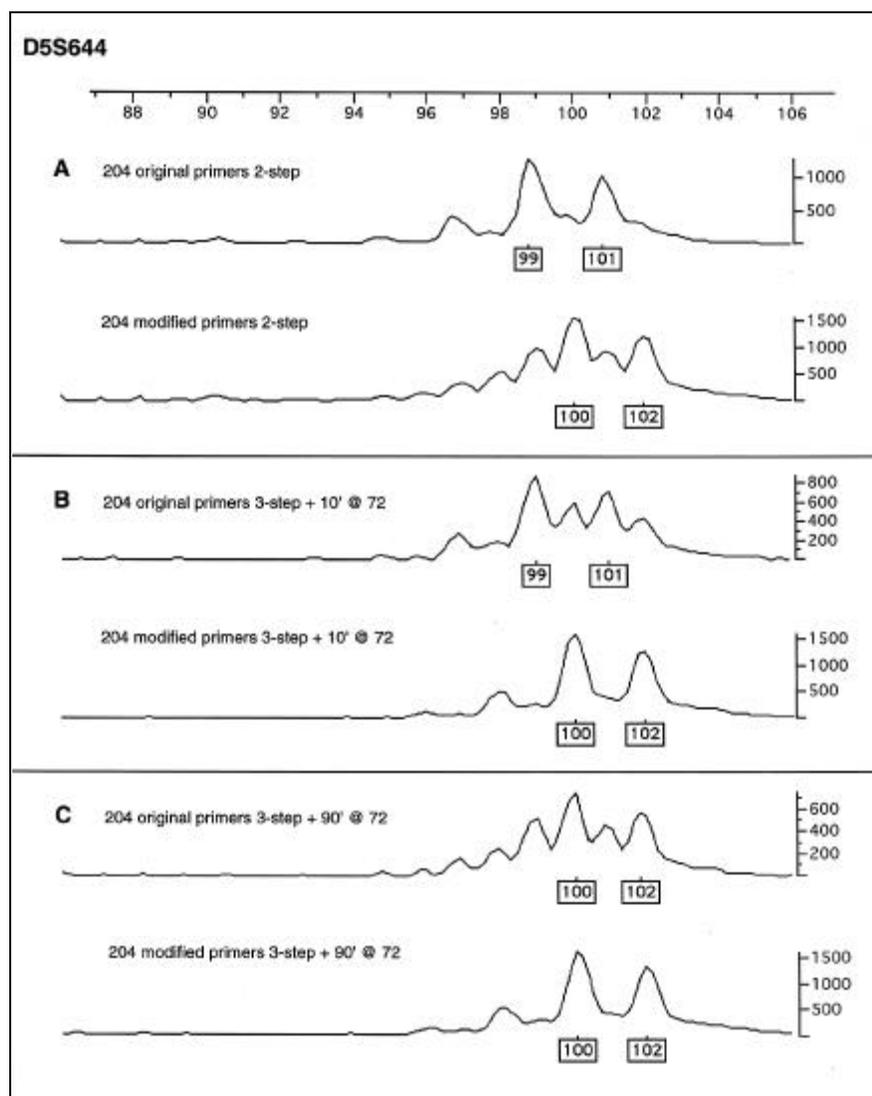


Figure 3. Electropherogram results generated by GENOTYPER displaying alleles for marker D5S644 from individual DNA No. 204 (CEPH 884-01) amplified with original vs. modified reverse primers under three PCR protocols. (A) 2-step, (B) 3-step + 10 min at 72°C, (C) 3-step + 90 min at 72°C.

Table 3. The Effect of Reverse Primer Sequence Change in Forcing PCR Products Amplified by "3-Step" Primers from Allele + A to Allele

Marker	Original Reverse Primer				Modified Reverse Primer			
	X-nucleotide X	2-Step Only	3-Step + 10-min ext. at 72°C	3-Step + 90-min ext. at 72°C	X-nucleotide X	2-Step Only	3-Step + 10-min ext. at 72°C	3-Step + 90-min ext. at 72°C
DXS1001	...AAC	0.80	0.83	0.83	...AAA	0.34	0.57	0.63
DXS986	...GCT	0.54	0.70	0.75	...GCA	0.31	0.57	0.78
D15S117	...GGG	0.44	0.64	0.68	...GGA	0.35	0.52	0.66
D15S127	...GTT	0.46	0.77	0.86	...GTA	0.26	0.45	0.82
D15S130	...TGC	0.77	0.80	0.79	...TGA	0.39	0.55	0.78
D15S165	...CCC	0.85	0.84	0.84	...CCA	0.57	0.81	0.82
D16S401	...AAT	0.81	0.84	0.85	...AAA	0.34	0.62	0.70
D16S511	...CTG	0.53	0.78	0.79	...CTA	0.45	0.75	0.74
DXS1060	...TAC	0.73	0.75	0.73	...TAA	0.43	0.66	0.71
DXS1204	...AAG	0.51	0.63	0.66	...AAA	0.44	0.56	0.67
DXS1214	...ACC	0.70	0.77	0.77	...ACA	0.40	0.46	0.74
DXS1227	...CTT	0.74	0.79	0.80	...CTA	0.43	0.75	0.74
Mean (12)		0.66	0.76	0.78		0.39	0.61	0.73

Terminal substrate nucleotides (the X position in Figure 1) for original and modified reverse primers are shown in bold type. Values are mean allele + A ratio $\{(allele + A)/([allele] + [allele + A])\}$ for PCR cycling condition. Means across 12 markers are presented at the end of the table and plotted in Figure 4.

increase in cycling time. For example, by using the original reverse primer, most of the "2-step markers", on average, cannot be forced to $\geq 70\%$ allele +

A, regardless of the cycling protocol. However, with the modified reverse primers, the same markers become usable with the 3-step + 10 min at 72°C

protocol (mean allele + A ratios \pm SEM = 0.704 ± 0.026). Overall, the greatest effect of the reverse primer modification on stimulating allele + A formation can be seen under the 2-step protocol, where alleviation of the sequence inhibition allows *Taq* to add on "A" even during the short extension time at 72°C.

Figure 3 illustrates the effect of changing the substrate X-nucleotide from "A" to "T" (reverse primer Y-nucleotide changed from "T" to "A") for chromosome 5 marker D5S644. Panels A, B and C show the difference in PCR product profiles generated by the original and the modified reverse primers using the protocols 2-step, 3-step + 10 min at 72°C and 3-step + 90 min at 72°C, respectively. As mentioned above (Materials and Methods), the heterozygote genotype that is 2 bp apart, like the one illustrated in Figure 3, is the hardest to score. For example, in Figure 3A (part 2, "204 modified primers 2-step"), the allele + A ratio for the first allele is overestimated because the allele + A peak (100 bp) also contains a component of the stutter pattern 2 bp downstream of the second allele + A peak (102 bp). The first allele peak (99 bp, not labeled by GENO-

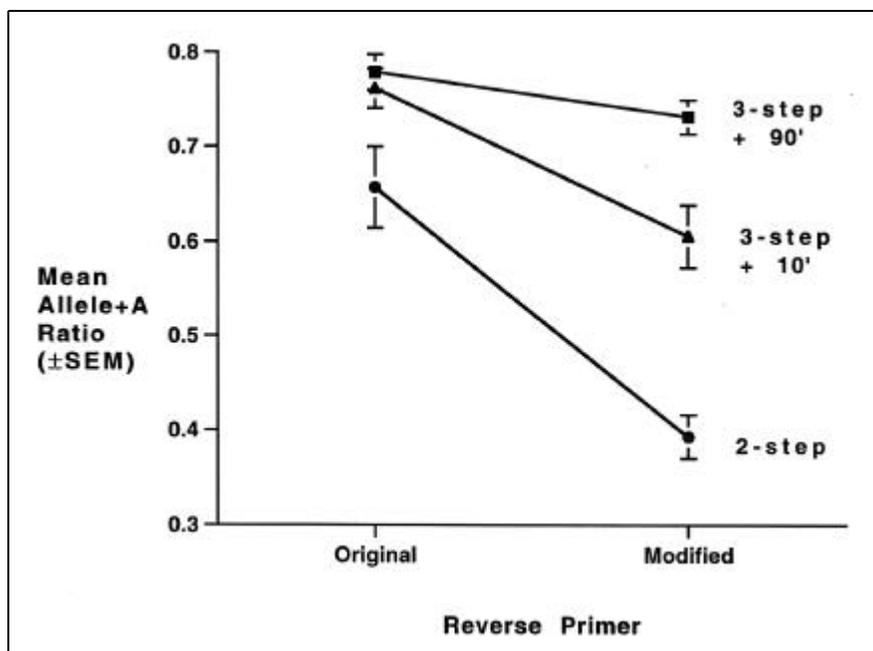


Figure 4. Changing the reverse primer Y-nucleotide to "T" and the terminal substrate X-nucleotide to "A" decreases the proportion of allele + A PCR products amplified under three different cycling protocols. Values are the mean of 12 mean allele + A ratios $\{(allele + A)/([allele] + [allele + A])\}$ calculated in Table 3 \pm SEM.

Research Reports

TYPYER) also contains a stutter component (-2 bp) from the second allele peak (101 bp, not labeled by GENOTYPER). If stuttering is equivalent on both allele and allele + A, then the allele + A ratio for the first allele in the genotype is proportional to the real value. However, the value for the second allele in the genotype (101 bp) will be overestimated even though it contains no stutter component, because it contains a component of the tail from the preceding allele + A peak (100 bp). The allele + A peak (102 bp) has no contaminating component, and therefore the allele + A ratio will be underestimated because of the overestimated allele (101 bp) value. Variation between individuals (and within individuals) in allele + A ratios for a given marker were due to allelic differences of this type rather than variations in PCR reproducibility.

Conversion of “3-Step Markers” to “2-Step Markers”

Twelve out of sixty “3-step markers” were selected to conduct the converse experiment, where the Y-nucleotide on the 5' end of the reverse primer was replaced with “T”. This would result in an “A” as a substrate (X-nucleotide) for the addition of “A”, which is predicted to be inhibitory. Table 3 shows a comparison between the mean allele + A ratio in PCR products amplified with the original reverse primer and the modified reverse primer across the three PCR protocols. The three terminal nucleotides encoded by reverse primers, including the X-position nucleotide in Figure 1 (bold), are also shown. Figure 4 is a summary plot of the effect of changing the X-position nucleotide to “A” for “3-step markers”.

As hypothesized, the presence of “A” as substrate nucleotide has a significant inhibitory effect on *Taq*'s addition of adenine to the PCR products generated by the modified reverse primers. While the PCR of 7/12 markers proceeds to $\geq 70\%$ allele + A with the 2-step protocol using the original primers, none of the modified primers gave this result (Table 3). Only 2/12 markers (D15S117 and DXS1204) failed to proceed to $\geq 70\%$ allele + A with the original primers under either

of the 3-step protocols. On the other hand, 9/12 and 3/12 of the modified reverse primers failed to proceed to $\geq 70\%$ allele + A with the 3-step + 10 min at 72°C and 3-step + 90 min at 72°C protocols, respectively.

A “catch-up” phenomenon appears to be at work during the 3-step + 90 min at 72°C protocol, where “A” is added to previously extended templates as opposed to being added during extension. This phenomenon is best illustrated in Figure 4, where on average the “3-step markers” proceed to $\geq 70\%$ allele + A whether the PCR cycling protocol is 3-step + 10 min (0.762 ± 0.021) or 3-step + 90 min (0.779 ± 0.019) at 72°C. When the reverse primer se-

quence produces an inhibitory terminal nucleotide, an extra 80 min extension time at 72°C can promote further “A” addition (0.732 ± 0.018). However, the equivalent result can be achieved with a “non-A” terminal nucleotide with the protocol 3-step + 10 min at 72°C (0.762 ± 0.021 vs. 0.732 ± 0.018). Figure 5 illustrates the effect of changing substrate X-nucleotide from “C” to “A” (reverse primer Y-nucleotide changed from “G” to “T”) on X chromosome marker DXS1001. Panels A, B and C show the difference in PCR product profiles generated by the original and the modified reverse primers using the protocols 2-step, 3-step + 10 min at 72°C and 3-step + 90 min at 72°C,

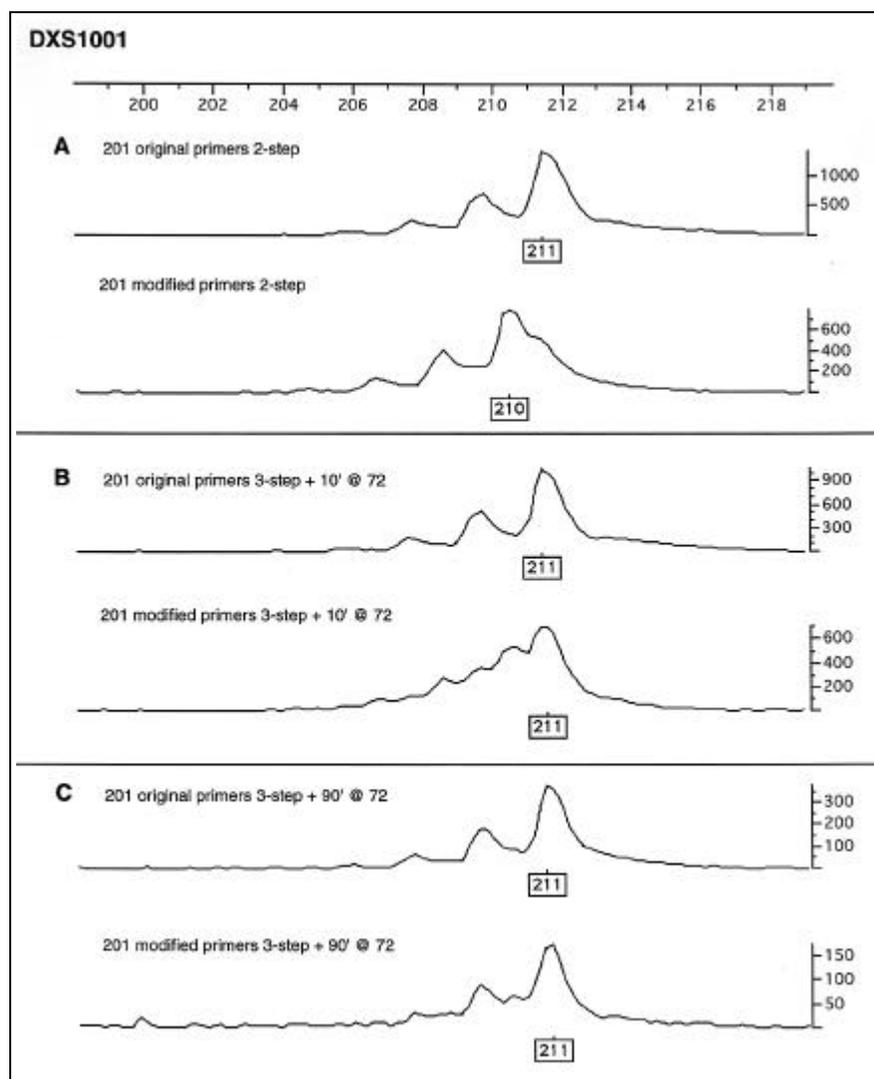


Figure 5. Electropherogram results generated by GENOTYPER displaying alleles for marker DXS1001 from individual DNA No. 201 (CEPH 884-16) amplified with original vs. modified reverse primers under three PCR protocols. (A) 2-step, (B) 3-step + 10 min at 72°C, (C) 3-step + 90 min at 72°C.

respectively. Note that a single base change between the original and the modified reverse primer creates an undesirable profile for this marker at 3-step + 10 min at 72°C (0.57 or 57% allele + A), rendering it much less useful for genotyping (Figure 5B).

Development of a Marker Rating System

In comparing the PCR behavior of the various markers studied, it became clear that some will proceed to allele + A at 2-step, while others will not, even at 3-step + 90 min at 72°C. For some markers, it seemed that sequence and cycling changes might be additive, for others interactive. In order to determine the effect of substrate nucleotides on “A” addition by *Taq*, a rating system was developed to monitor the change from “2-step” to “3-step markers” and vice versa. Table 4 illustrates the defin-

itions used to score the markers based on the behavior of their original reverse primers under the three PCR protocols. In addition, tallies of the number of markers with each score and the total of each terminal nucleotide A,G,T,C are given for each rating group. Table 4 shows that while “C” and “T” are good terminators for promoting “A” addition, “A” is clearly the inhibitory nucleotide, and “G” can also be inhibitory. It is noteworthy that “T” and “C” are only present as terminal nucleotides on markers of ratings of 5 or 6.

Table 4 shows that some of the markers studied here do not fall neatly into either “2-step” or “3-step” marker categories. Ideal “2-step markers” have ratings of 1 or 2. Markers of rating 3 give scorable results at 2-step or 3-step + 90 min at 72°C, but never do as well at 3-step + 10 min at 72°C (mean allele + A ratio \pm SEM = 0.484 \pm 0.035) as markers of ratings 5 or 6 (mean allele +

A ratio \pm SEM = 0.787 \pm 0.014). In practice, markers of intermediate ratings (4 and 7) are the most troublesome, and those tested here were equally inhibited by “A” and “G” terminal nucleotides. The 2/27 “2-step markers”, which scored a rating of 4 with the original reverse primers, scored ratings of 5 after modification of the reverse primers. Therefore, re-engineering their reverse primers made these markers readily usable, while avoiding an extra 80 min incubation at 72°C. Finally, 4/27 “2-step markers” scored a rating of 7; after re-engineering their reverse primers, 3 out of these 4 were converted to “3-step markers”.

The presence of an adenine as substrate X-nucleotide is highly inhibitory to the non-templated addition to PCR products of “A” by *Taq* DNA polymerase. In several cases illustrated, removing the inhibitory “A” (and “G”) allows the reaction to proceed to allele +

Research Reports

Table 4. Definition of Marker Rating System

Rating	Original primer mean allele + A ratio 2 Step	Original primer mean allele + A ratio 3 Step + 10 min at 72°C	Original primer mean allele + A ratio 3 Step + 90 min at 72°C	Number of Markers	Nucleotide X (Figure 1) A G T C
1	a	a	o	3	3 0 0 0
2	a	o	o	10	7 3 0 0
3	a	o	A	7	5 2 0 0
4	o	o	A	2	1 1 0 0
5	o	A	A	4	1 1 2 0
6	A	A	A	7	0 0 2 5
7	o	o	o	6	3 3 0 0

a = mean allele + A ratio ≤30%
A = mean allele + A ratio ≥70%
o = 30% < mean allele + A ratio <70%
Values are mean allele + A ratio $\{(\text{allele} + A)/([\text{allele}] + [\text{allele} + A])\}$ corresponding to original reverse primers listed in Table 2 and Table 3.

A, while simultaneously decreasing the cycling protocol from 3-step + 90 min at 72°C to 3-step + 10 min at 72°C. For high-throughput genotyping, a savings of 80 min is generated per PCR. (Alternatively, if an oven is used instead for prolonged 72°C incubations, this is less of a concern). The necessity of freezing 2-step PCR products means that the operator must be present at the termination of the run. When all PCR products go to allele + A, overnight PCR runs or storage at 4°C have no detrimental affects, since, if anything, this enhances the allele + A profile.

Our results are consistent with those seen in primer extension of single base trinucleotide overhangs (5). In the primer extension study, terminal "AAA" was most inhibitory to "A" addition, presumably due to a single "A" terminus. (There is actually some evidence in our study that multiple "A"s are more inhibitory than a single "A" [unpublished observations].) The terminal "GGG" was somewhat inhibitory to addition of "A", probably because *Taq* DNA polymerase prefers adding "G" rather than "A" onto "GGG" termini (5). The terminal "CCC" promoted "A" addition, but *Taq* also added "C" (although less frequently) onto the "CCC" terminus (5). Interestingly, a major difference between our study and the previous primer extension study (5), is that

we found terminal "T" to promote allele + A formation, while by primer extension a terminal (TT) "T" was excised and replaced by "A" (i.e., TTT became TTA), thereby showing no size change. We saw no evidence of this. In fact, all of our modified reverse primers for the conversion of "2-step markers" to "3-step markers" generated "T" termini and clearly promoted the formation of allele + A (Figure 2). The differing results obtained by the two studies may have been caused by differences in (i) the magnesium concentrations used (1.5 mM here vs. 2.25 mM in the primer extension study); or (ii) the extension times and temperatures used (30 s at 72°C during PCR vs. 1 min at 70°C, then reaction termination in the primer extension study). These parameters can affect both *Taq* DNA polymerase stuttering and allele + A formation (unpublished observations). Alternatively, because of the short fragment size and single-base trinucleotide end sequences used in the primer extension study, *Taq* DNA polymerase may have exhibited different extension enzymatics during PCR cycling.

The trends described here are significant. Given the number of primers studied, our results represent a consistent behavior of *Taq* DNA polymerase through multiple rounds of extension in multiple assays. These observations

can be used in primer design, depending on whether a product of allele or allele + A is desired. For example, since reverse primers of marker rating category 6 often generate PCR products containing "C" as terminal substrate (X-nucleotide), we recommend putting a "G" (Y-nucleotide) on the 5' end of all reverse primers to optimally promote allele + A PCR products. In fact, "G"-ended reverse primers ("C" in X-nucleotide position) are now routinely used by the FUSION (Finnish-United States Investigation of Non-Insulin Dependent Diabetes Mellitus) project, which is at present the largest intramural genotyping effort at the National Center for Human Genome Research. All PCRs are run using the 3-step + 10 min at 72°C protocol and stored at 4°C. An example of three markers used in the current study that have improved allele + A ratios after "G"-ended reverse primer modification are D16S405, D16S415 and D16S516. The mean allele + A ratios obtained for these three markers, using the 3-step + 10 min at 72°C protocol listed in Table 2, are 0.44, 0.70 and 0.62, respectively. The mean allele + A ratios obtained for these markers after "G"-ended reverse primer modification and PCR using the 3-step + 10 min at 72°C protocol are 0.81, 0.81 and 0.74, respectively (data not shown).

On the other hand, to avoid allele + A, putting a "T" (Y-nucleotide) on the 5' end of all reverse primers will inhibit allele + A formation under the right PCR conditions. Therefore, in the future, reverse primers designed for genotyping should correspond to desired cycling protocols and storage conditions to optimize for allele or allele + A as dictated by the needs of a given laboratory. These recommendations should also be useful in designing forward and reverse primers for generating PCR products intended to be cloned into TA cloning vectors (4,6), where the allele + A form is most desirable.

ACKNOWLEDGMENTS

We would like to thank Dr. Dennis Gilbert and the staff at Applied Biosystems Division of Perkin-Elmer for providing invaluable training and materials for this project.

REFERENCES

1. **Clark, J.** 1988. Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucl. Acids Res.* 16:9677-9686.
2. **Copeman, J., F. Cucca, C. Hearne, R. Cornall, P. Reed, K. Ronningen, D. Undlien, L. Nistico, R. Buzzetti, R. Tosi, F. Pociot, J. Nerup, F. Cornelis, A. Barnett, S. Bain and J. Todd.** 1995. Linkage disequilibrium mapping of a type 1 diabetes susceptibility gene (IDDM7) to chromosome 2q31-q33. *Nature Genet.* 9:80-85.
3. **Davies, J., Y. Kawaguchi, S. Bennett, J. Copeman, H. Cordell, L. Pritchard, P. Reed, S. Gough, S. Jenkins et al.** 1994. A genome-wide search for human type 1 diabetes susceptibility genes. *Nature* 371:130-136.
4. **Holton, T. and M. Graham.** 1991. A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. *Nucl. Acids Res.* 19:1156.
5. **Hu, G.** 1993. DNA polymerase-catalyzed addition of nontemplated extra nucleotides to the 3' end of a DNA fragment. *DNA Cell Biol.* 12:763-770.
6. **Marchuk, D., M. Drumm, A. Saulino and F. Collins.** 1991. Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucl. Acids Res.* 19:1154.
7. **Perkin-Elmer Applied Biosystems Division.** 1994. ABI PRISM™ Panel Inserts. Foster City, CA.
8. **Perkin-Elmer Applied Biosystems Division.** 1994. GENOTYPER™ DNA Fragment Analysis Software, User's Manual Set. Foster City, CA.
9. **Smith, J., J. Carpten, M. Brownstein, S. Ghosh, V. Magnuson, D. Gilbert, J. Trent and F. Collins.** 1995. An approach to genotyping errors caused by non-templated nucleotide addition by *Taq* DNA polymerase. *Genome Res.* 5:312-317.

Received 6 February 1996; accepted 5 June 1996.

Address correspondence to:

Victoria L. Magnuson
National Institutes of Health
National Center for Human Genome Research
Laboratory of Gene Transfer
49 Convent Drive, Room 3A-14
Bethesda, MD 20892, USA
Internet: magnuson@nchgr.nih.gov