

Thermo Sequenase™ DNA Polymerase and *T. acidophilum* Pyrophosphatase: New Thermostable Enzymes for DNA Sequencing

P.B. Vander Horn, M.C. Davis, J.J. Cunniff, C. Ruan, B.F. McArdle, S.B. Samols, J. Szasz, G. Hu, K.M. Hujer, S.T. Domke, S.R. Brummet, R.B. Moffett and C.W. Fuller
Amersham Life Science, Cleveland, OH, USA

BioTechniques 22:758-765 (April 1997)

ABSTRACT

A combination of thermostable enzymes has been developed that produces higher quality cycle sequences. Thermo Sequenase™ DNA polymerase is a thermostable enzyme engineered to catalyze the incorporation of ddNTPs with an efficiency several thousandfold better than other thermostable DNA polymerases. Since the enzyme also catalyzes pyrophosphorolysis at dideoxy termini, a thermostable inorganic pyrophosphatase is needed to remove the pyrophosphate produced during sequencing reactions. *Thermoplasma acidophilum* inorganic pyrophosphatase (TAP) is thermostable and effective for converting pyrophosphate to orthophosphate. The use of the combination of Thermo Sequenase polymerase and TAP for cycle sequencing yields sequence data with uniform band intensities, allowing the determination of longer, more accurate sequence reads. Uniform band intensities also facilitate interpretation of sequence anomalies and the presence of mixed templates. Sequencing PCR products of DNA amplified from heterozygous diploid individuals results in signals of equal intensity from each allele.

INTRODUCTION

When Sanger first used dideoxyribonucleoside triphosphates (ddNTPs) and the Klenow fragment of *E. coli* DNA polymerase for chain-termination sequencing (12), two properties of the enzyme became apparent. First, the rate of incorporation of ddNTPs is thousands of times slower than the rate for deoxyribonucleoside triphosphates (dNTPs). Second, the rate of incorporation of a ddNTP varies depending on the template sequence. The rate of ddNTP reaction and, hence, chain termination can be as much as 10-fold higher at some

bases in a sequence than others, resulting in variability in band intensity on electrophoresis gels.

Like the Klenow enzyme, most polymerases produce variable band intensities in sequencing. One exception is T7 DNA polymerase lacking exonuclease activity (T7 Sequenase™; Reference 13). This polymerase produces bands that are far less variable, particularly when reactions are run in the presence of Mn²⁺ (14). Under these conditions, T7 Sequenase catalyzes the incorporation of ddNTPs at the same rate as dNTPs. Tabor and Richardson have recently shown that the ability of a DNA polymerase to discriminate between deoxy and dideoxy nucleotides can be altered dramatically by substitution of a single amino acid in a conserved region of the polymerase (15). Changing a phenylalanine to a tyrosine in the polymerase domain of Klenow fragment led to decreased discrimination against ddNTPs.

Cycle sequencing has become a mainstay for fluorescent and some radioactive sequencing methods, since it helps ensure that sufficient labeled DNA will be generated for reliable sequencing (4,7). Cycle sequencing can be performed by adding polymerase after each thermal denaturation step (3,6), but it is more efficient and convenient to use a polymerase that is sufficiently stable to retain activity after cycling through 95°C. Thermostable DNA polymerases of the DNA polymerase I family (2) like *Thermus aquaticus* (*Taq*), as well as thermostable polymerases of the DNA polymerase II family such as *Pfu*, isolated from an Archaeon, yield variable sequence band intensities. These polymerases require high concentrations of ddNTPs, and *Taq* DNA polymerase has the additional disadvantage of 5' to 3' exonuclease activity.

We have engineered, by a combination of deletion and site-directed mutagenesis, a thermostable DNA polymerase

Table 1. Numbers are the Ratio of Reactivity of the Indicated dNTP to the Corresponding ddNTP

Enzyme	dNTP:ddNTP Reactivity			
	G	A	T	C
Δ Taq DNA Polymerase	200	2500	30 600	2400
Thermo Sequenase DNA Polymerase	0.41	0.71	0.93	0.72

Note that ddNTPs are somewhat better substrates than dNTPs for Thermo Sequenase.

that is exonuclease-free and does not discriminate against ddNTPs. We present the advantages of using Thermo Sequenase™ DNA Polymerase (Amersham, Cleveland, OH, USA) in combination with *Thermoplasma acidophilum* inorganic pyrophosphatase (TAP) for DNA sequencing.

MATERIALS AND METHODS

Cloning and In Vitro Mutagenesis

In vitro manipulations of DNA were performed using standard protocols (11). Manipulations of the *Taq* DNA polymerase gene were based upon the sequence in GenBank® D32013. TAP was cloned from *T. acidophilum* (strain 25905; ATCC, Rockville, MD, USA) using the polymerase chain reaction (PCR) and the DNA sequence (10). The DNA sequence of all constructs was verified. The Thermo Sequenase and TAP proteins were expressed using plasmid pRE2 (8). TAP was isolated from *E. coli* by a procedure based on that of Richter and Schäfer (9).

Determining Relative Reaction Rates of ddNTPs and dNTPs

Modified DNA sequencing reactions were performed for each ddNTP. The ratio of the concentration of dNTPs to ddNTPs was adjusted so that 10–50 chain-termination bands were observed. The universal cycle primer was labeled by extension in the presence of 3 pmol of [α -³³P]dATP, dCTP and dGTP using M13mp18 single-stranded DNA for 5 min at 45°C. Termination reactions were initiated by transferring aliquots of the primer-labeling reaction into appropriate mixtures of dNTPs and one ddNTP and incubating for 10 min at 72°C. For Δ Taq™ DNA Polymerase (Amersham), the ratios of ddNTPs to dNTPs were 400:1 for ddATP, 200:1 for ddCTP, 30:1 for ddGTP and 600:1 for ddTTP. For Thermo Sequenase DNA polymerase, the concentration ratios of ddNTPs to dNTPs were all 1:5.

Autoradiograms were scanned using a SciScan™ 5000 Scanning Densitometer (Amersham). The intensities of 10–20 bands in each lane were fit to an exponential function $i = ae^{-kn}$, where i is band intensity and n is the number of the band. Parameters a and k are characteristic constants. The ratio of intensity of each band to its predecessor is given by $r = e^{-k}$. From r , the ratio of reaction rate for ddNTPs to the reaction rate for dNTPs under the conditions of the experiment is calculated as $r/(1-r)$. Multiplication by the ratio of concentration of ddNTPs to dNTPs gives the intrinsic ratio of reactivities shown in Table 1.

Dye Primer DNA Sequencing

Dye primer cycle sequencing reactions were performed in 52 mM Tris-HCl, pH 8.9, 13 mM MgCl₂, with 18 U of Δ Taq or 18 U of Thermo Sequenase and 0.005 U of TAP. The

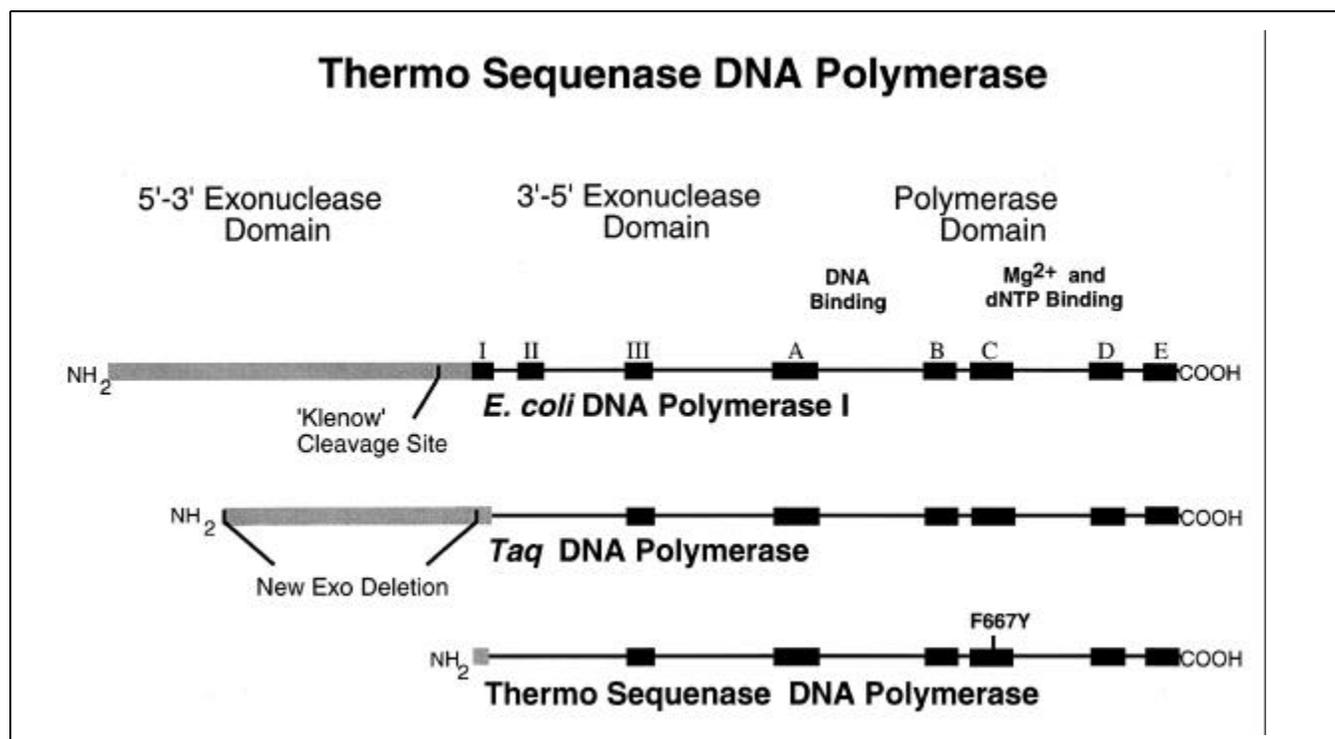


Figure 1. Comparative structures of type I DNA polymerases. Activity domains and areas of sequence conservation are depicted for *E. coli*, *Taq* and Thermo Sequenase DNA polymerases. *Taq* and Thermo Sequenase do not have 3' to 5' exonuclease activity. Thermo Sequenase has a deletion of the 5' to 3' exonuclease domain, and amino acid 667 (relative to *Taq* DNA polymerase) has been changed from phenylalanine to tyrosine.

amount of M13mp18 DNA varied as indicated. Thermo Sequenase reactions used 0.2 pmol of JOE- and FAM-labeled primers (A and C termination reactions, respectively) and 0.8 pmol of TAMRA- and ROX-labeled primers (G and T termination reactions, respectively) (PE Applied Biosystems, Foster City, CA, USA). Nucleotide termination reactions contained 60 μ M dNTPs and 0.2 μ M of either ddATP, ddCTP, ddGTP or ddTTP. Δ Taq reactions used 0.4 pmol of JOE- and FAM-labeled primers (A and C termination reactions, respectively) and 0.8 pmol of TAMRA- and ROX-labeled primers

(G and T termination reactions, respectively). The nucleotide termination reactions contain 7.5 μ M dNTPs with either 30 μ M ddATP, 15 μ M ddCTP, 2.25 μ M ddGTP or 45 μ M ddTTP. All reactions were cycled 60 times as follows: 95°C for 30 s, 55°C for 30 s, then 72°C for 60 s. Reactions were combined, ethanol-precipitated, dissolved in formamide containing 10 mM EDTA and then heat-denatured before loading onto a sequencing gel. PCR products were treated with exonuclease I (exo I) and shrimp alkaline phosphatase (SAP) before sequencing (5). Reactions were run on a Model 373 DNA

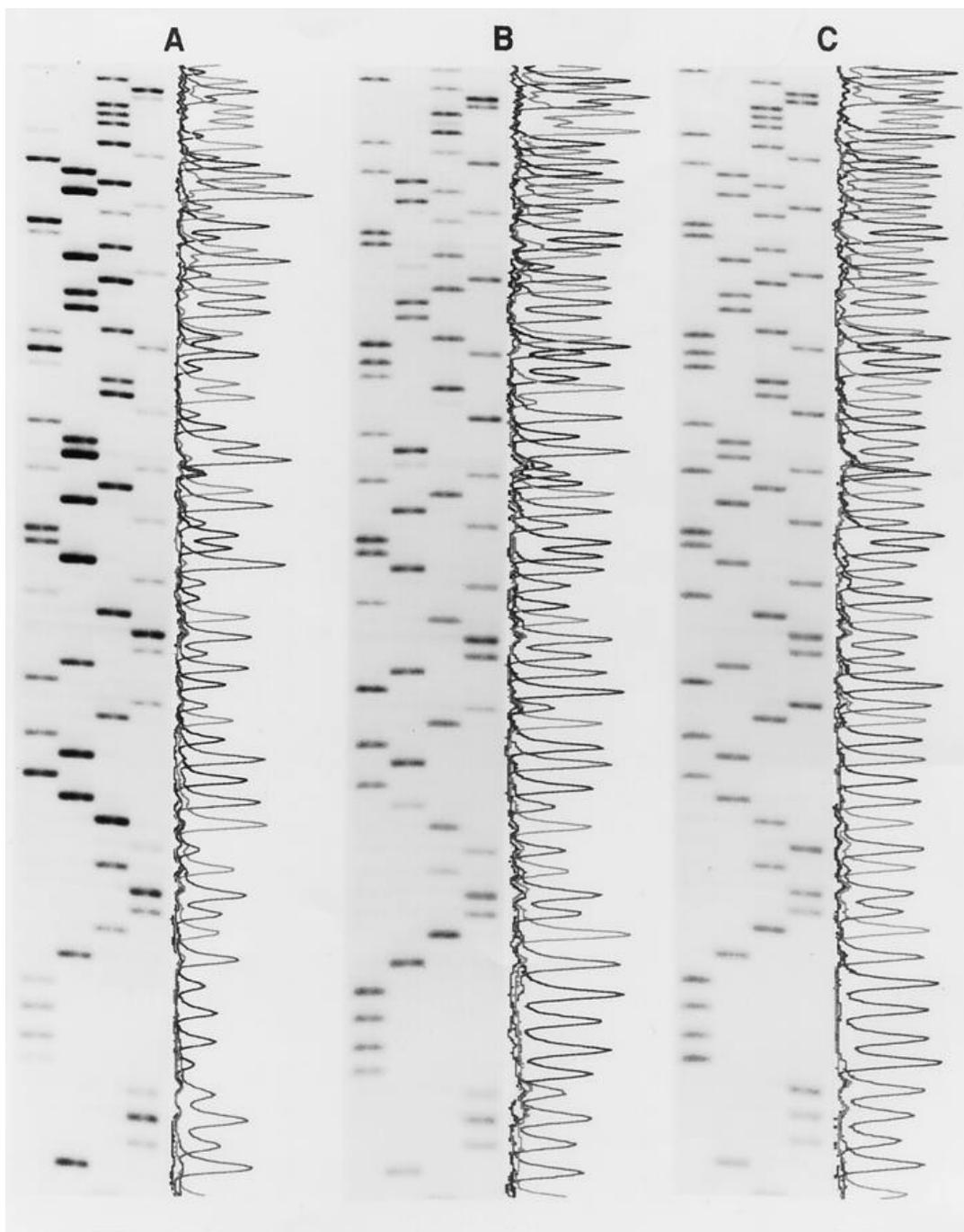


Figure 2. Band variability and pyrophosphorolysis. Cycle sequencing autoradiogram and densitometric analysis of sequencing reactions using Δ Taq (A), Thermo Sequenase (B) and Thermo Sequenase with TAP (C). Lanes are loaded G, A, T and C. Densitometric results are indicated as follows: black, G; green, A; red, T; and blue, C.

Sequencer (PE Applied Biosystems) and analyzed using software version 1.2.1.

Radioactive DNA Sequencing

Cycle sequencing with ΔTaq DNA Polymerase was performed with the ΔTaq Cycle Sequencing Kit (Amersham) using the two-step cycle-labeling protocol. Thermo Sequenase sequencing reactions were done as above except for the following modifications: 8 U of Thermo Sequenase (with or without 0.015 U of TAP) were used, and the termination reactions contained 100 μ M dNTPs and 1 μ M ddATP, ddCTP, ddGTP or ddTTP.

RESULTS AND DISCUSSION

Construction of Thermo Sequenase DNA Polymerase

Functional and conserved domains of three type I DNA polymerases are shown in Figure 1. The phenylalanine-to-tyrosine change, which affects discrimination against ddNTPs, occurs in domain C, and PCR was used to introduce this change into the *Taq* DNA polymerase gene. A new start site for the polymerase was engineered by deleting the first 272 amino acids and adding a methionine residue for translation initiation. This deleted the 5' to 3' exonuclease domain, while allowing abundant expression of a stable and homogeneous enzyme. Starts at other positions such as amino acid 236, found in ΔTaq (1), lead to heterogeneous populations of proteins caused by limited proteolysis near the amino terminus of the protein. The half-life of polymerase activity at 95°C is 50 min.

TAP

Inorganic pyrophosphate (PP_i) and the 3'-terminal dideoxynucleoside monophosphate (ddNMP) can react to form ddNTP in the reversal of the polymerase reaction, pyrophosphorolysis. Thermo Sequenase, like T7 Sequenase polymerase, catalyzes the removal of the 3'-ddNMP, which results in some weak or absent sequencing bands (Figure 2B). If PP_i is hydrolyzed to orthophosphate, pyrophosphorolysis is prevented (14).

A thermostable inorganic pyrophosphatase such as TAP suppresses pyrophosphorolysis in DNA sequencing reactions with Thermo Sequenase polymerase (Figure 2C). The combination yields lower variability in band intensity than either ΔTaq or Thermo Sequenase polymerase alone. TAP has no effect upon ΔTaq sequences because pyrophosphorolysis is not a problem for polymerases with low ddNTP reactivity. TAP suppresses pyrophosphorolysis artifacts in all fluorescent and radioactive sequencing applications that we have tested. No loss of TAP activity is detected after 2 h incubation at 95°C under sequencing conditions.

Relative Reactivity of ddNTPs and dNTPs

The rate of incorporation of each of the ddNTPs relative to the rate for the corresponding dNTP was measured by running modified sequencing reactions. The values obtained with ΔTaq polymerase (Table 1) are similar to those reported for *Taq* DNA polymerase (15). Rates with dNTPs range from 200-fold to 30 000-fold greater than the rates with the corresponding ddNTP, confirming that ddNTPs are relatively poor substrates for this polymerase. The values for Thermo Sequenase

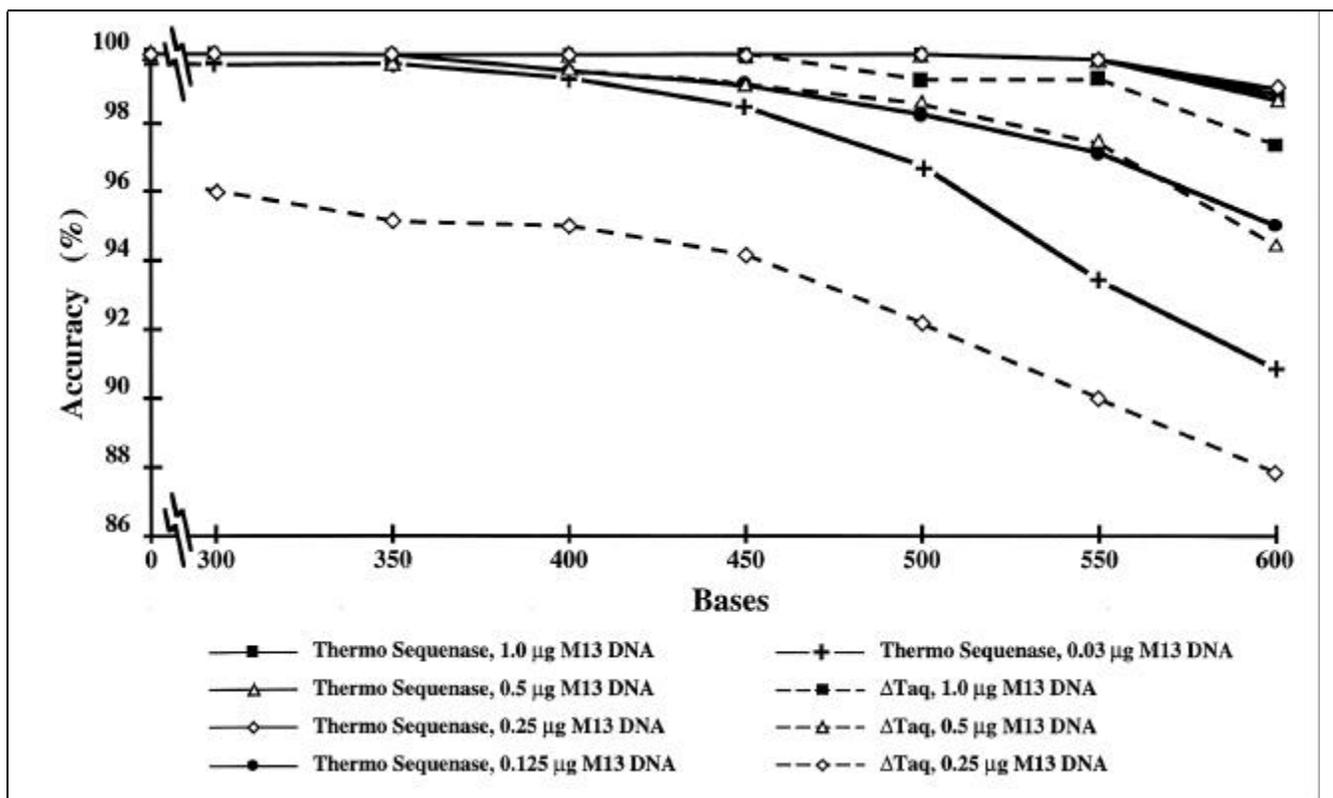


Figure 3. Effect of template amount on sequence accuracy. The graph compares the base-calling accuracy of sequence data derived using Thermo Sequenase and ΔTaq polymerases.

nase polymerase are all less than one, indicating that ddNTPs are better substrates than dNTPs for this polymerase.

Concentration ratios of ddNTP to dNTP in *Taq* sequencing reactions are approximately 30:1 rather than the ratios of 0.01:1 needed with Thermo Sequenase. As a consequence, radioactive sequencing reactions with *Taq* DNA polymerase are often run with 30 pmol of each dNTP (4 μ M), along with up to 1800 pmol (240 μ M) of the appropriate ddNTP. This amount of dNTP is only sufficient to synthesize 0.25 pmol of DNA, barely sufficient for overnight exposure. Increasing the amounts of nucleotides at these ddNTP:dNTP ratios is prohibitively expensive. Thus, a large fraction of the available nucleotides are consumed at concentrations close to the K_m values of the DNA polymerases used. In the case of Thermo Sequenase, which uses ddNTPs more efficiently, the amount of expensive ddNTP is decreased to 4 pmol and the amount of dNTP is increased to 400 pmol each (50 μ M). The 10-fold increase in dNTP concentration improves the efficiency of

DNA synthesis, eliminating a possible source of background signal resulting from nonspecific terminations during inefficient polymerization.

Uniform Band Intensities Improve Sequence Read-Length and Accuracy

Quantitative measurements of band intensities indicate that more than 80% of the bands in Thermo Sequenase sequences differ in intensity from their neighbors by less than 20%, whereas neighboring bands for *Taq* DNA polymerase vary by 10-fold or more (Figure 2). To demonstrate the effect of band variability on sequence accuracy, we compared Thermo Sequenase to Δ *Taq* using varying amounts of template DNA (Figure 3). With Δ *Taq*, sequence accuracy was seriously compromised with 0.25 μ g of DNA and insufficient signal was obtained with 0.125 μ g or less. Using Thermo Sequenase, accurate sequence information was obtained with as little as 0.03 μ g of template.

The reason for the difference can be seen in the analyzed results of Δ *Taq* and Thermo Sequenase reactions (Figure 4). In the region shown, the signal-to-noise ratio for the Δ *Taq* sequence is lower than that of the Thermo Sequenase sequence. Overly strong "T" peaks (arrows) interfere with the recognition of weaker "A" and "C" peaks following them. Also, around base 563 of the Δ *Taq* sequence, a background "T" signal (asterisk) interferes with the interpretation of a G base, resulting in the calling of an "N". In the segment shown, a total of six base-calling errors were made using Δ *Taq*, largely because of variability in band intensities, whereas no errors occurred with Thermo Sequenase polymerase.

Sequencing Heterozygotes

Uniform peak heights facilitate the detection of mixed template sequences such as DNA from heterozygous individuals. PCR products of human DNA, heterozygous within the MSH2 locus, were sequenced using Thermo Sequenase polymerase. Coincident peaks of half intensity were obtained where the allele sequences differ (Figure 5). With sequencing methods producing bands that vary by more than a factor of 2, such sites are more difficult to identify. Other examples of mixed-sequence templates include DNA amplified from environmental samples of populations such as viruses and mixed tissue samples isolated from tumors.

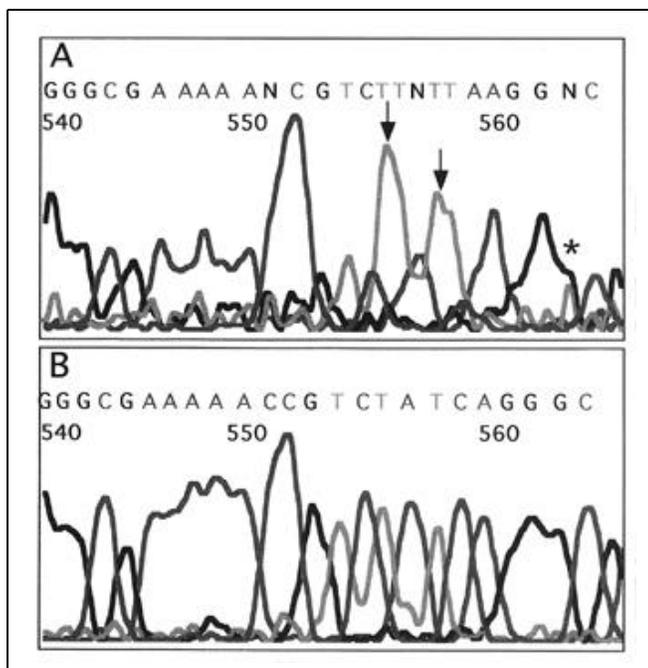


Figure 4. Band variability affects sequence accuracy. (A) Δ *Taq*; (B) Thermo Sequenase DNA polymerase. Both panels show sequences from Figure 3 obtained with 1 μ g of template.

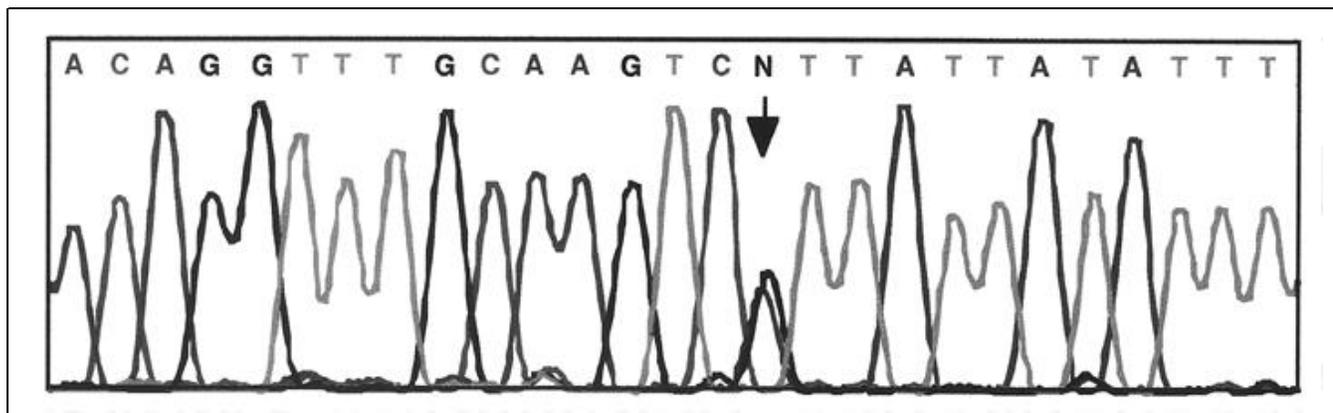


Figure 5. Detection of heterozygotes by direct sequencing of PCR products. DNAs of test individuals were sequenced following amplification of chromosomal DNA by PCR. Shown is a polymorphism in an intron of the human *MSH-2* gene.

The development of this new enzyme system for cycle sequencing should increase the speed of genome sequencing. The accurate detection of multiple sequences from heterozygous individuals will facilitate genotyping.

ACKNOWLEDGMENTS

P.B.V.H. and M.C.D. contributed equally to the work described in this paper. Primer sequences for Figure 5 courtesy of Dr. Richard Kolodner, Harvard Medical School. We thank Dr. Charles C. Richardson for his careful reading of this manuscript. This work was supported in part by Grant 70NANB4H1523 from the Advanced Technology Program of the U.S. Department of Commerce.

REFERENCES

1. **Barnes, W.M.** 1992. The fidelity of Taq polymerase catalyzing PCR is improved by an N-terminal deletion. *Gene* 112:29-35.
2. **Braithwaite, D.K. and J. Ito.** 1993. Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Res.* 21:787-802.
3. **Huibregtse, J.M. and D.R. Engelke.** 1988. DNA Sequencing from complex double-stranded templates. *DNA Prot. Eng. Techniques* 1:39-41.
4. **Koop, B.F., L. Rowan, W.-Q. Chen, P. Deshpande, H. Lee and L. Hood.** 1993. Sequence length and error analysis of Sequenase® and automated Taq cycle sequencing methods. *BioTechniques* 14:442-447.
5. **Mamone, J.T., S.B. Samols and C.W. Fuller.** 1994. Use of the Sequenase PCR product sequencing kit for analysis of 16S rRNA genes from novel microorganisms. *Comments* 21:34-36.
6. **McMahon, G., E. Davis and G.N. Wogan.** 1987. Characterization of c-Ki-ras oncogene alleles by direct sequencing of enzymatically amplified DNA from carcinogen-induced tumors. *Proc. Natl. Acad. Sci. USA* 84:4974-4978.
7. **Murray, V.** 1989. Improved double-stranded DNA sequencing using the linear polymerase chain reaction. *Nucleic Acids Res.* 17:8889.
8. **Reddy, P., A. Peterkofsky and K. McKenney.** 1989. Hyperexpression and purification of *Escherichia coli* adenylate cyclase using a vector designed for expression of lethal gene products. *Nucleic Acids Res.* 17:10473-10488.
9. **Richter, O.-M.H. and G. Schäfer.** 1992. Purification and enzymatic characterization of the cytoplasmic pyrophosphatase from the thermoacidophilic archaeobacterium *Thermoplasma acidophilum*. *Eur. J. Biochem.* 209:343-349.
10. **Richter, O.-M.H. and G. Schäfer.** 1992. Cloning and sequencing of the gene for the cytoplasmic inorganic pyrophosphatase from the thermoacidophilic archaeobacterium *Thermoplasma acidophilum*. *Eur. J. Biochem.* 209:351-355.
11. **Sambrook, J., E.F. Fritsch and T. Maniatis.** 1989. *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
12. **Sanger, F., S. Nicklen and A.R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463-5467.
13. **Tabar, S. and C.C. Richardson.** 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA* 84:4767-4771.
14. **Tabar, S. and C.C. Richardson.** 1990. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase: effect of pyrophosphorolysis and metal ions. *J. Biol. Chem.* 265:8322-8328.
15. **Tabar, S. and C.C. Richardson.** 1995. A single hydroxyl moiety in Pol I-type DNA polymerases is responsible for distinguishing between deoxy- and dideoxyribonucleotides. *Proc. Natl. Acad. Sci. USA* 92:6339-6343.

Address correspondence to Carl W. Fuller, Amersham Life Science, Inc., 26111 Miles Road, Cleveland, OH 44128, USA. Internet: carl.fuller@amersham.com