4500–4504 *Nucleic Acids Research, 1997, Vol. 25, No. 22*

New dye-labeled terminators for improved DNA sequencing patterns

B. B. Rosenblum*, L. G. Lee, S. L. Spurgeon, S. H. Khan, S. M. Menchen, C. R. Heiner and S. M. Chen

PE Applied Biosystems, 850 Lincoln Centre Drive, Foster City, CA 94404, USA

Received August 14, 1997; Revised and Accepted October 3, 1997

ABSTRACT

We have used two new dye sets for automated dye-labeled terminator DNA sequencing. One set consists of four, 4,7-dichlororhodamine dyes (d-rhodamines). The second set consists of energy-transfer dyes that use the 5-carboxy-d-rhodamine dyes as acceptor dyes and the 5- or 6-carboxy isomers of 4′-aminomethylfluorescein as the donor dye. Both dye sets utilize a new linker between the dye and the nucleotide, and both provide more even peak heights in terminator sequencing than the dye-terminators consisting of unsubstituted rhodamine dyes. The unsubstituted rhodamine terminators produced electropherograms in which weak G peaks are observed after A peaks and occasionally C peaks. The number of weak G peaks has been reduced or eliminated with the new dye terminators. The general improvement in peak evenness improves accuracy for the automated base-calling software. The improved signal-to-noise ratio of the energy-transfer dye-labeled terminators combined with more even peak heights results in successful sequencing of high molecular weight DNA templates such as bacterial artificial chromosome DNA.

INTRODUCTION

Sanger dideoxy DNA sequencing is the most commonly used method for DNA sequencing, particularly in large scale genomic sequencing (1). Automated DNA sequencing uses fluorescent dyes for the detection of the electrophoretically resolved DNA fragments. Two variations of automated DNA sequencing have evolved: dye-labeled primer sequencing (2–4), in which the fluorescent dyes are attached to the 5′ end of the primer oligonucleotide, and dye-labeled terminator sequencing, in which the dyes are attached to the terminating dideoxynucleoside triphosphates (5,6). Each sequencing method has advantages and disadvantages.

Dye-labeled primer sequencing has benefited from the development of DNA polymerases which do not discriminate between deoxy- and dideoxynucleotides (7,8). These polymerases provide sequencing electropherograms with very even peak heights. Base-calling is easy and reliable, and the ability to call heterozygotes can be based on peak heights as well as the presence of two bases at a position (9,10). The major disadvantage of the dye primer method is the requirement for four separate extension reactions and four dye-labeled primers for each template.

The major advantages of dye-labeled terminator sequencing are convenience, since only a single extension reaction is required for each template, and the synthesis of a labeled primer is unnecessary, allowing the use of preferred hybridization sites. In addition false terminations, in which the DNA fragments are terminated by a deoxynucleotide rather than a dideoxynucleotide, are not observed as these products are unlabeled. The major disadvantage of dye-labeled terminators is that with every polymerase the pattern of termination with dye-labeled terminators has been found to be less even than for dye-labeled primers. The presence of very small or very large peaks can result in errors in automated base-calling.

We have evaluated the use of two new dye sets, 4,7-dichloro-substituted rhodamines (d-rhodamines), and a set of energy-transfer dyes that were previously described for use on dye-labeled primers (11). Use of the new terminators, d-rhodamine terminators and BigDye™ terminators, using the energy transfer dyes, required optimization of the linker attaching the dyes to the nucleotides, the dye isomer used, and the choice of each dye on a particular nucleotide. A major objective with both dye-sets was to obtain more even peak patterns compared with current DNA sequencing terminators. The energy-transfer dyes also offer the advantage of increased signal. The improved peak evenness found in both new dye sets allows greater accuracy in base-calling, longer reads and the ability to use dye-labeled terminators for heterozygote analysis.

MATERIALS AND METHODS

Dye-labeled terminators

The dye-labeled terminators were prepared by methods previously described (12). Briefly, the succinimidyl ester of each dye was mixed with the nucleoside triphosphates at pH 9. The products were purified on HPLC by anion-exchange chromatography to remove excess dye, followed by reverse-phase chromatography to separate unlabeled triphosphates and to separate dye-isomers.

* To whom correspondence should be addressed. Tel: +1 650 638 5566; Fax: +1 650 638 6666; Email: rosenbbb@perkin-elmer.com

Table 1. Linkers, fluorophores and final terminator concentration for each nucleotide of the three terminator sets: rhodamine, d-rhodamine and BigDye

Terminators	Rhodamine			d-Rhodamine			BigDye			
Nucleotide	Linker[a]	Dye	Final conc. (μM)	Linker	Dye	Final conc. (μM)	Linker	Donor dye	Acceptor dye	Final conc. (μM)
ddATP	PA	5-R6G	0.02	PA	dR6G-2	0.02	PA	6-FAM	dR6G-2	0.11
ddCTP	PA	6-ROX	0.13	EO	dTAMRA-2	0.12	EO	6-FAM	dROX-2	0.16
ddGTP	PA	5-R110	0.01	EO	dR110-2	0.01	EO	5-FAM	dR110-2	0.10
ddTTP	PA	6-TAMRA	0.23	EO	dROX-1	0.18	EO	6-FAM	dTAMRA-2	1.12

[a]The linkers are PA, propargylamino and EO, propargyl ethoxyamino.

DNA sequencing

Dye-labeled terminator cycle sequencing with d-rhodamine and BigDye terminators, using the energy-transfer dyes, was performed using AmpliTaq DNA polymerase, FS, according to the ABI PRISM™ sequencing manual (PE Applied Biosystems, Foster City, CA). With the BigDye terminators, dUTP was substituted for dTTP at the same concentration in the dNTP mix. The concentrations of the dNTPs in the reactions were 100 μM for dATP, dCTP and dTTP (or dUTP) and 500 μM for dITP. The concentrations of the terminators used in the reactions were determined by titrating each terminator in single-color terminator reactions and selecting the concentration which maximized the signal of the 700th nucleotide (12). The concentrations were adjusted according to the relative brightness of each dye in order to obtain approximately equivalent signal for each color in the four-color reaction (Table 1). The d-rhodamine and BigDye terminator sets used 10 nm virtual filters on the CCD camera of the ABI PRISM 310 and 377 centered at 540 or 545 (ABI PRISM 310), 570, 595 and 625 nm.

Bacterial artificial chromosomal (BAC) DNA was purified by an alkaline lysis protocol (13) and was sequenced with BigDye terminators with a slight modification of the terminator protocol. Table 2 shows the reagents used per reaction for BAC sequencing. BAC samples were cycled in a Perkin-Elmer 9600 thermocycler according to the following protocol: initial denaturation at 95°C for 5 min, followed by 30 cycles of 95°C for 30 s, 55°C for 20 s, and 60°C for 4 min. Excess terminators were removed using Centri-Sep spin columns (Princeton). Samples were vacuum dried and then resuspended in 2 or 4 μl of formamide, heated to 95°C for 2 min and 2 μl of the sample loaded on the ABI PRISM 377. The BAC clone, bWXD342, used in these studies contains an insert, 169 kb in length, from the human X chromosome, locus Xq21.3.

Table 2. Reagent mix used for BAC sequencing

Reagent	Per reaction
Terminator mix	8.0 μl
5X sequencing buffer	4.0 μl
Amplitaq FS for dye terminators	0.5 μl
Primer	4.0 pmol
BAC DNA	~400 ng
Distilled water	to volume
Final volume	40.0 μl

Single color analysis of the d-rhodamine and BigDye terminators was performed using the ABI PRISM 310 Genetic Analysis system. Single color sequencing reactions were prepared as described earlier, except that in all cases excess terminators were removed using Centri-Sep spin columns (Princeton). Samples were resuspended in Template Suppression Reagent (TSR; PE Applied Biosystems, Foster City, CA) and heated to 95°C for 2 min. A computer program has been developed to determine peak heights and to calculate the mean, standard deviation and relative error, where the relative error is the ratio of the standard deviation to the mean.

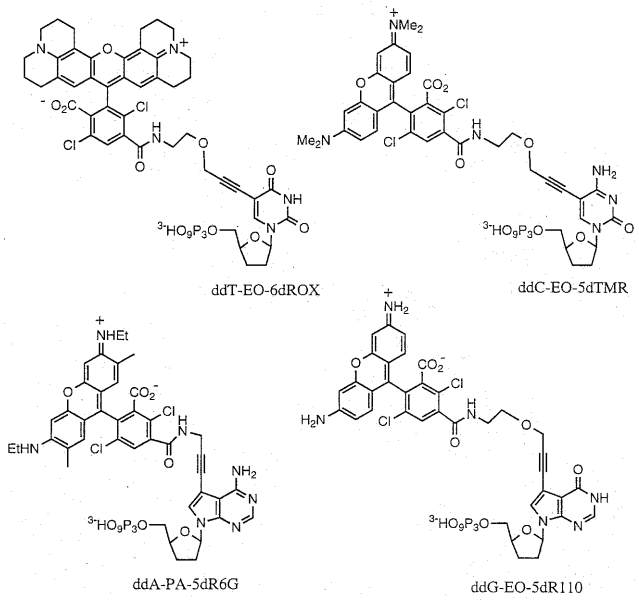

Figure 1. Structure of the d-rhodamine dye-labeled terminators.

RESULTS AND DISCUSSION

Dichlororhodamine terminators

In order to optimize the performance of the new d-rhodamine terminators, we synthesized and tested 39 out of a possible 64 combinations of the four dichlororhodamine dyes (both 5- and 6-carboxy isomers), propargylamino (PA) or propargyl ethoxyamino (EO) linker, and nucleotide terminators ($8 \times 2 \times 4 = 64$). The structure of both the dye and the linker between the nucleotide and the dye affected the pattern of termination (manuscript in preparation). We chose the dye set which maximized the evenness of the peaks in the sequencing pattern. The final d-rhodamine terminator set had mobility shifts within the half base requirement for minimal artifacts. (The set of all 32 possible rhodamine dyes with the EO linker was tested but a 4-dye set was not found that had acceptable mobility characteristics.) The structures of the d-rhodamine dye-labeled terminators are shown in Figure 1. Only the A-terminator retains the propargylamino linker of the original terminators (Table 1).

BigDye terminators

We have previously reported on a set of energy-transfer dyes for dye-labeled primer sequencing that uses the 5-carboxy-dichloro-rhodamine dyes as acceptor dyes and the 5- or 6-carboxy isomers of 4′-aminomethylfluorescein as the donor dye. These dyes show both improved spectral resolution and improved brightness compared with the standard dyes used for dye-primer sequencing (11). Here, we have investigated the use of the energy-transfer dyes on dye-labeled terminators.

We synthesized and tested 18 out of a possible 64 combinations of energy-transfer dyes (four 5-carboxy-d-rhodamines with both 5- and 6-carboxy isomers of 4′-aminomethylfluorescein), propargyl-amino (PA) or propargyl ethoxyamino (EO) linker, and four nucleotide terminators ($8 \times 2 \times 4 = 64$). Again, we chose the dye set which maximized the evenness of the peaks in the sequencing pattern and minimized the dye-related mobility effects. The final BigDye terminator set had mobility shifts within the half base requirement for minimal artifacts. The structure of one of the four BigDye terminators, the ddT-EO-6CFB-dTMR, is shown in Figure 2. In addition to varying the terminator structure, we found that varying the structure of the dNTPs also affected the pattern of termination. By substituting dUTP for dTTP the termination pattern for energy-transfer dye-labeled ddT terminators was improved for each of the seven different energy-transfer dye/linker/ddT compounds tested.

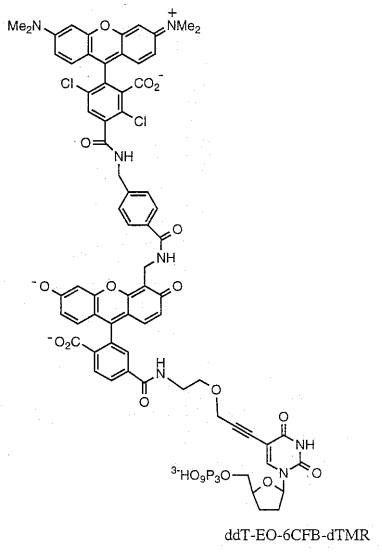

Figure 2. Structure of the ddT-BigDye terminator.

A feature of the BigDye terminator set is that the total signal is increased compared with either the rhodamine or d-rhodamine dye sets. BigDye terminators can be used with advantage in cases when template molar equivalents are limited. BAC DNA fragments typically have very high molecular weights and cannot be sequenced with single dyes. Energy-transfer dye-labeled primers (9,14–17) have proven useful for these templates (13). Here we demonstrate the signal advantage of energy-transfer dyes combined with the convenience of dye-labeled terminators. Figure 3 shows a sequencing pattern for a bacterial artificial chromosome template (clone:bWXD342; sequencing reaction Table 2) using the BigDye terminators.

Table 3. Comparison of number of sequencing accuracy, read length and total signal for different templates with the three dye terminator chemistries

Template	%GC	Errors to 720 bases			Read length at 98.0% accuracy			Signal strength		
		Rhod	dRhod	BigDye	Rhod	dRhod	BigDye	Rhod	dRhod	BigDye
349, -21M13	65.3	50	14	16	394	704	683	1453	732	1422[a]
349, -21M13		45	15	34	543	688	613	2812	1182	2993
4009, -21M13	71.4	44	18	26	586	692	646	1098	333	2056[a]
4009, Reverse M13		28	14	10	616	701	741	693	397	2692[a]
ABD 114, Reverse M13	59	16	7	7	691	781	733	2830	1413	3698[a]
pGEM, -21M13	50.8	11	4	8	718	796	723	3478	1447	4865
pGEM, -21M13		21	6	6	654	784	776	3548	1374	3834[a]
pGEM, Reverse M13		12	9	0	785	727	831	3202	1374	4268[a]
pcDNA, Reverse M13	33	4	4	2	790	790	843	3929	1366	5184[a]
DJ2, Reverse M13	30	7	3	2	743	798	790	1874	724	4434[a]
ABD 116, -21M13	37.3	12	6	1	744	791	810	2268	948	2750
ABD 116, -21M13		25	11	12	612	736	769	1974	502	1346[a]
ABD 116, Reverse M13		6	2	2	788	788	776	2397	1090	4385
ABD 116, Reverse M13		3	6	5	807	740	807	2102	650	2384[a]
ABD 100, -21M13	48	26	8	2	605	785	813	877	375	1118[a]
ABD 100, Reverse M13		21	12	6	659	727	758	1336	498	1606[a]
ABD 90, -21M13	36.3	33	18	4	364	671	835	1172	442	1884[a]
ABD 90, Reverse M13		3	8	12	834	773	720	1064	602	1552[a]
MEAN		20	9	9	663	748	759	2117	858	2915

[a]BigDye terminator separations were performed with half the sample load compared with rhodamine separations. Signal value has been normalized by doubling the signal value for the BigDye terminators for these separations.

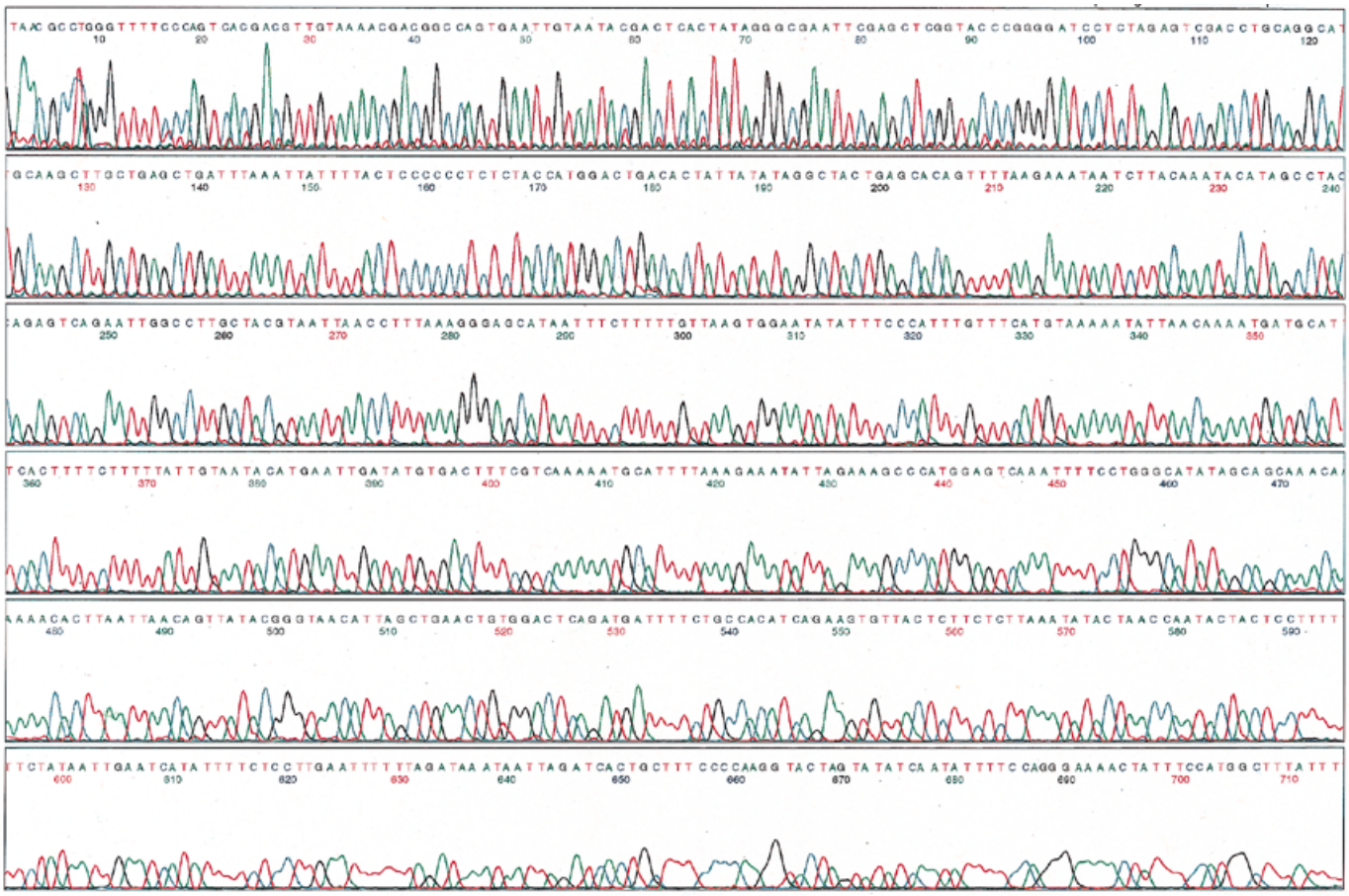

Figure 3. Sequencing of BAC DNA (clone:bWXD342) with BigDye terminators.

Comparison of the three terminator sets

Table 1 shows the matching of linker and fluorophore for the original rhodamine, the d-rhodamine, and the BigDye terminators. Table 3 compares the sequencing accuracy, read-lengths and signal for the three terminator sets for different templates. As a result of the better peak evenness of the d-rhodamine and BigDye terminator sets both base-calling accuracy and read-lengths are improved. Although the total signal for the d-rhodamine terminators is reduced compared with the rhodamines (equal amounts of template were loaded on the sequencing gels), the multicomponent noise is also reduced due to the better spectral resolution of the d-rhodamine dye set. The signal strength for the BigDye terminators with most of the templates is higher than the signal strength for the rhodamine terminators as expected based on the brightness of the energy-transfer dyes (11). The 2-fold reduction in multicomponent noise for the d-rhodamine and BigDye terminators compared with the rhodamine terminators is not reflected in the signal number (11).

We have analyzed a series of different templates with three sets of dye-labeled terminators, rhodamine, d-rhodamine and BigDye, to compare peak evenness and to identify sequence context effects. The sequencing patterns of a portion of template DJ2 using the three sets of dye-terminators are shown in Figure 4. In the rhodamine set, very weak G peaks after A peaks are observed, with some weak G peaks after C peaks. There are also some very strong peaks that result in the smallest peaks being >10-fold smaller than the largest peaks, with a small peak frequently appearing just before or after these large peaks. In both the d-rhodamine and the BigDye terminator patterns, the peaks are of more even heights, so that in general the adjacent peaks are <5-fold different in size. The small G peaks after A peaks or C peaks has also improved, with the BigDye terminators showing no weak G peaks and the d-rhodamines showing a few weak G peaks after A peaks or C peaks. These small G peaks are reliably called by the automated software due to the better balance in the peak heights. The BigDye terminators show weaker T peaks after G peaks, but again, because the overall pattern is more balanced, these T peaks are called by the automated software.

Table 4. Relative errors[a] of peak heights for bp 10–315 in pGEM with the three terminator sets and dye primers

	Dye-primer	Rhodamine	d-Rhodamine	BigDye dT	dU
A	0.27	0.83	0.49	0.23	0.23
C	0.26	0.60	0.33	0.28	0.31
G	0.31	0.83	0.42	0.37	0.31
T	0.22	0.52	0.36	0.55	0.40
Average	0.26	0.69	0.40	0.36	0.32

The dye-primer reactions use c^7dGTP in the dNTP mix, while the terminator reactions use dITP in the dNTP mix instead of dGTP.

[a]The relative error is the ratio of the standard deviation of the peak heights to the mean of the peak heights for a selected group of peaks.

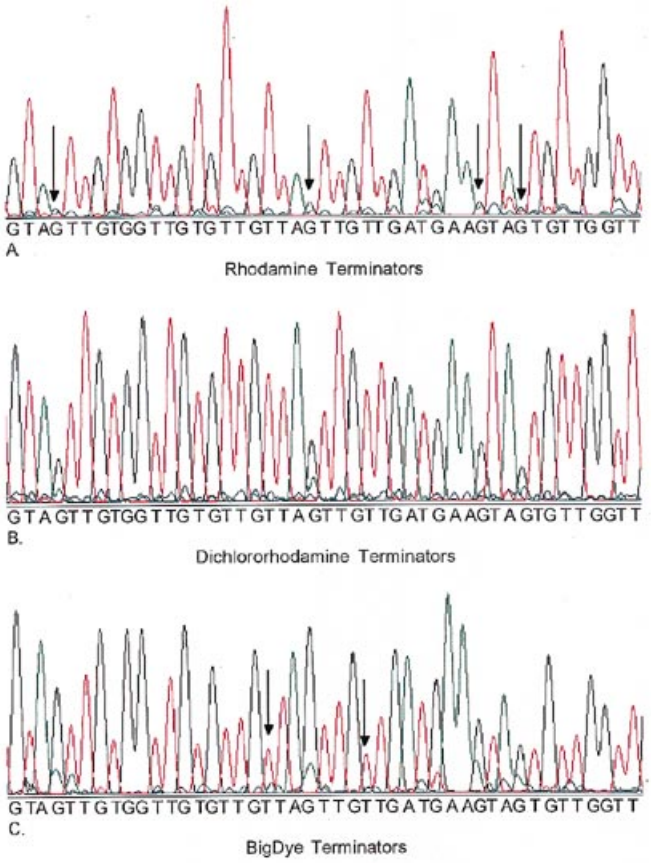

Figure 4. Comparison of sequencing patterns of rhodamine, d-rhodamine and BigDye terminators on an AT rich template, DJ2. Arrows in (**A**) are G peaks with weak signal. Many of these G peaks are similar in size to noise peaks under adjacent peaks. In (**B**) and (**C**), all of the G peaks are much larger than any noise peaks. In (C), the arrows indicate weaker T peaks following G peaks. These T peaks are larger than any noise peaks and are called by the automated software.

The peak patterns can be quantitatively evaluated by measuring the peak heights of a given sequence and calculating an average and standard deviation. The data are normalized by defining the relative error as the ratio of the standard deviation to the average peak height. A completely uniform series of peaks would yield a relative error value of 0. This is unlikely to occur for any type of Sanger sequencing except over a very short group of peaks, because of the exponential decay of the terminal events with increasing fragment size (12), results in decreasing peak height with increasing fragment length over a large group of bases. Thus, for a group of >100 bases, a value of 0.15–0.3 would be expected for dye primer sequencing (Table 4). Table 4 shows the results of peak height evaluation for the rhodamine, d-rhodamine and BigDye terminator sets, along with the relative error values for dye-labeled primer sequencing for the same region.

CONCLUSIONS

We have developed two new dye-terminator sets that are both improvements over previous dye-terminators. The peak patterns of these chemistries are nearly as even as dye-labeled primer sequencing patterns. The genome sequencing community requires data with a high confidence of base-calling, and data generated by two different sequencing approaches in areas with single-orientation coverage (18–22). In the past this has meant that the bulk of the data were generated using dye-primers, supplemented by dye-terminators only when necessary. These requirements may now possibly be met with the two new dye-terminator chemistries, eliminating the need for dye primers. Sequencing for heterozygote analysis may be able to be performed with these new dye-terminators rather than dye-primers. Future work in enzyme engineering and dye synthesis to further enhance the performance of dye-labeled terminators will likely render obsolete traditional dye-labeled primer sequencing.

ACKNOWLEDGEMENTS

We thank Jim Bowlby for developing software for peak height evaluations and Bill Efcavitch for helpful discussions. We are grateful to Mike Hunkapiller for expediting the dye-terminator effort, to Krishna Upadhya, Tony Constantinescu, Ron Graham, Paul Kenney, Brian Evans, Scott Benson, Pete Thiesen and Mary Fong for performing the dye syntheses and purifications, to Pavel Cotofana and James Liang for synthesis of dye terminators and to Gilbert Amparo for analysis of base-calling accuracy. The clone DJ2 was a gift from Drs Simon Plyte and Jim Woodget. The purified BAC clone was prepared by Dr B.H. Brownstein at Washington University Medical School.

REFERENCES

1 Sanger, F., Nicklen, S., and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
2 Smith, L.M., Fung, S., Hunkapiller, M.W. Hunkapiller, T.J., and Hood, L.E. (1985) *Nucleic Acids Res.*, **13**, 2399–2412.
3 Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heiner, C., Kent, S.B.H., and Hood, L.E. (1986) *Nature*, **321**, 674–679.
4 Voss, H., Schwager, C., Wirkner, U., Sproat, B., Zimmermann, J., Rosenthal, A., Erfle, H., Stegemann, J., and Ansorge, W. (1989) *Nucleic Acids Res.*, **17**, 2517–2527.
5 Prober, J.M., Trainor, G.L., Dam, R.J., Hobbs, F.W., Robertson, C.W., Zagursky, R.J., Cocuzza, A.J., Jensen, M.A., and Baumeister, K. (1987) *Science*, **238**, 336–341.
6 Bergot, B.J., Chakerian, V., Connell, C.R., Eadie, J.S., Fung, S., Hershey, N.D., Lee, L.G., Menchen, S.M., and Woo, S.L. (1994) U.S. Patent 5,366,860.
7 Tabor, S., and Richardson, C.C. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 6339–6343.
8 Tabor, S., and Richardson, C.C. (1990) *J. Biol. Chem.*, **265**, 8322–8328.
9 Metzger, M.L., Lu, J., and Gibbs, R.A. (1996) *Science*, **271**, 1420–1422.
10 Gibbs, R.A., Nguyen, P.N., Edwards, A., Civitello, A.B., and Caskey, C.T. (1990) *Genomics*, **7**, 235–244.
11 Lee, L.G., Spurgeon, S.L., Heiner, C.R., Benson, S.C., Rosenblum, B.B., Menchen, S.M., Graham, R.J., Constantinescu, A., Upadhya, K.G., and Cassel, J.M. (1997) *Nucleic Acids Res.*, **25**, 2816–2822.
12 Lee, L.G., Connell, C.R., Woo, S.L., Cheng, R.D., McArdle, B.F., Fuller, C.W., Halloran, N.D., and Wilson, R.K. (1992) *Nucleic Acids Res.*, **20**, 2471–2483.
13 Marra, M., Weinstock, L.A., and Mardis, E.R. (1996) *Genome Res.*, **6**, 1118–1122.
14 Ju, J., Ruan, C., Fuller, C.W., Glazer, A.N., and Mathies, R.A. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 4347–4351.
15 Ju, J., Kheterpal, I., Scherer, J.R., Ruan, C., Fuller, C.W., Glazer, A.N., and Mathies, R.A. (1995) *Anal. Biochem.*, **231**, 131–140.
16 Ju, J., Glazer, A.N., and Mathies, R.A. (1996) *Nucleic Acids Res.*, **24**, 1144–1148.
17 Hung, S.-C., Ju, J., Mathies, R.A., and Glazer, A.N. (1996) *Anal. Biochem.*, **243**, 15–27.
18 Hunkapiller, T., Kaiser, R.J., Koop, B.F., and Hood, L. (1991) *Science*, **254**, 59–67.
19 Hawkins, T.L., McKernan, K.J., Jacotot, L.B., MacKenzie, J.B., Richardson, P.M., and Lander, E.S. (1997) *Science*, **276**, 1887–1889.
20 Bult, C. J., *et al*. (1996) *Science*, **273**, 1058–1073.
21 Fleischmann, R.D., *et al*. (1995) *Science*, **269**, 496–512.
22 Wilson, R., *et al*. (1994) *Nature*, **368**, 32–38.