Rapid determination of nucleotide content and its application to the study of genome structure

Dan E. Krane, Daniel L. Hartl and Howard Ochman*

Department of Genetics, Box 8232, Washington University School of Medicine, St Louis, MO 63110, USA

Received July 26, 1991; Revised and Accepted September 4, 1991

ABSTRACT

We have developed a sensitive, reliable and accurate procedure for estimating the base composition of small samples of DNAs. This method has been applied to the analysis of genomic DNAs from several sources including large regions of human DNA cloned as yeast artificial chromosomes. To determine whether the human genome is compartmentalized into large segments of homogeneous base composition, we examined the GC content of a 1.2 megabase contig spanning the cystic fibrosis gene.

INTRODUCTION

The determination of the GC content of genomic DNA has long been used as a criterion for the classification and identification of bacterial species (1,2). Subsequently, the acquisition of nucleotide sequence data from a large number of genes and from diverse species, has demonstrated that the GC content of bacterial genomes is relatively uniform over the entire chromosome (3), validating the use of base composition as a taxonomic tool.

Examination of the base composition of eukaryotic genomes has revealed that, while overall %G+C can be used in the characterization of a species, the GC content within certain genomes appears to be very heterogeneous. Based on the analysis of buoyant density gradients, it has been suggested that the genomes of warm-blooded vertebrates are compartmentalized into long regions (>300 kilobases) of significantly different base composition termed 'isochores' (4,5). The discrete fractions apparent in density gradients contain large fragments of DNA whose total base composition is uniform within 1 or 2% G+C (6,7). Analyses of nucleotide sequences, particularly of the alpha and beta globin gene clusters, generally support the idea that mammalian genomes are structured into regions of varied GC content (8–10).

The predicted size of these large regions of homogeneous base composition does not make them amenable to analysis of their structure by conventional nucleotide sequencing. Moreover, the utility of buoyant density gradients is limited because fractions contain a mixture of many loci from the entire genome and studies of their chromosomal location and organization are difficult. In

contrast, DNA cloned as yeast artificial chromosomes (YACs) (11) corresponds closely to the size predicted for isochores, are relatively easy to manipulate and map, and are likely to be useful in studies of this level of organization within eukaryotic genomes.

Methods that have been traditionally utilized to determine the GC content of DNA do not easily lend themselves to the analysis of YACs. High performance liquid chromatography (HPLC) (12), thermal denaturation (T_m) (1,2) and buoyant density (13) typically require large amounts (a minimum of 200 ng and often >1 μ g) of starting material: commonly used preparations of YACs yield only nanogram (ng) quantities of purified clones. Mabuchi and Nishikawa (14) recently determined the relative base composition of restriction fragments based upon the selective binding of DNA fluorochromes. While this technique could be applied to small quantities of DNA, sequence specificities of the binding agents could bias the quantitation of base composition and may limit the range of GC content over which it is useful.

In this paper, we describe an accurate and universally applicable method for determining the base composition of small quantities of DNA. This technique, Base Ratio Analysis using Thin-layer chromatography (BRAT), involves the uniform radiolabeling of DNA followed by nuclease digestion and the separation of monophosphates by thin-layer chromatography (TLC). This method is suitable for the analysis of all DNA samples and allows accurate determination of GC content to within 0.5–1.0% in as little as one day after the DNA is purified. We have applied this procedure to a variety of bacterial and viral DNAs as well as large regions of the human genome cloned into YACs. We discuss the utility of this method in the characterization and mapping of isochores.

EXPERIMENTAL METHODS

DNA source and purification

Yeast artificial chromosomes were separated from yeast chromosomes on CHEF gels (15), excised, and purified by GeneClean (Bio 101 Inc.). Genomic DNA from all the bacterial species used were prepared as described by Marmur (16) and were provided by Jeffrey Lawrence. Viral DNAs were obtained from New England Biolabs as provided for size standards.

* To whom correspondence should be addressed

Nick translation and nuclease digestion

Nick translations (17) were performed in a final volume of 20 μ l using a final concentration of 0.5 mM dTTP and [α - 32 P]-labeled dGTP, dATP, and dCTP (also 0.5 mM, with a specific activity of >400 Ci/mmol). One hundred ng of bacterial and phage DNAs, or 1 to 5 ng of YAC DNA, were uniformly radiolabeled. Typically, 20 reactions were performed simultaneously by combining all reagents in a master mix before dispensing to tubes containing DNA to be tested. After incubation for one hour at 15°C, nick translation reactions were terminated by the addition of 2 μ l 0.5M EDTA (pH 8.0) and heating to 65°C for 10 minutes. Unincorporated radionucleotides were removed by ethanol precipitation with 1/4 volume 10M NH₄OAc. Samples were resuspended in 6.9 μ l of 50mM NaOAc (pH 5.3), denatured at 90°C for 5 minutes and flash cooled by placing on ice. After cooling, 1 μ l (3 units) of a 1:10 dilution of nuclease P1 (BRL) was added and complete digestion of the DNA to monophosphates was achieved by incubation at 37°C for 30 minutes.

TLC separation and quantification

Nuclease digested samples containing dGMP, dAMP and dCMP (each [α - 32 P]-labeled) and unlabeled dNMPs were dried to a volume of 3 μ l under vacuum before they were spotted on polyester polyethyleneimine cellulose TLC plates (T-6765, Sigma). An initial thin layer chromatographic separation was achieved using 0.1M K₂HPO₄ at 4°C. Further separation was obtained by redeveloping the plates in the same solvent system a second time (Figure 1).

The relative amounts of radiolabeled guanine, adenine and cytosine in each of the samples were determined by scanning the developed plates with a Betascope 603 Blot Analyzer (Betagen) for a minimum of one hour. Counts from each nucleotide were determined by demarcating each with a rectangle of minimum size using the scanner's manual quantitation feature while viewing the autoradiographic image at 90–99% contrast.

Figure 1. Autoradiographic image of radiolabeled monophosphates separated by thin layer chromatography. Lanes contain replicates of nick-translated, P1-digested lambda DNA. Relative positions of each of three radiolabeled monophosphates are shown. Areas from which counts were collected for base ratio analysis are demarcated by rectangles.

Only nucleotides yielding 50,000 to 500,000 counts/hour were used for base ratio calculations. Undigested polynucleotides (accounting for <1% of the total counts applied to each spot) do not migrate from the origin and the limited mobility of residual unincorporated triphosphates does not interfere with the counting of decays originating from the radiolabeled monophosphates released by nuclease digestion. Similar results may be obtained by scintillation counting of the spots or by densitometric analysis of autoradiograms.

Calculations

The GC content of each sample was determined by comparing the ratio of radiolabeled nucleotides (dGMP, dAMP or dCMP) to those incorporated in a standard of known base composition, usually lambda DNA. Since lambda is 49.8% G+C (18), the ratios of counts in the dGMP to dAMP comparison and in the dCMP to dAMP comparison should both be 0.992. Variability in the specific activity of each of the three radionucleotides results in deviations from the expected ratio. To correct for these differences, the averages of the observed G:A and C:A ratios for standards were used to create a correction factor which was applied to the G:A and C:A ratios of unknowns as in the following equation:

$$R_c = R_u \left(\frac{0.992}{\lambda} \right) \quad (1)$$

where R_c = the corrected ratio of either the G:A or C:A counts, R_u = the uncorrected ratio of G:A or C:A counts observed, and λ = the observed ratio of counts for those nucleotides seen for lambda bacteriophage DNA. Corrected G:A and C:A ratios were then converted to %G+C estimates with the equation:

$$\%G+C = 100 \left(\frac{R_c}{R_c + 1} \right) \quad (2)$$

If the %G+C values obtained from the G:A and C:A ratios differed by >2.5% G+C, the experiment was repeated. The base composition of each sample was determined by averaging the results of at least two replicates.

RESULTS**Establishing standards**

Initial experiments were performed on eight replicates of 100 ng of lambda DNA using aliquots of the same reagents. The base ratio estimates were all found to be in agreement to within 0.3% of the average and had a standard deviation of 0.14% (Table 1). To determine the accuracy and reliability of BRAT, subsequent

Table 1. Estimates of base composition of lambda bacteriophage DNA as determined by BRAT.

Sample	dCMP cts/hr	dAMP cts/hr	dGMP cts/hr	Actual G:A	Actual C:A	Corrected G:A	Corrected C:A	%G+C from G:A	%G+C from C:A	Average GC content*
1	101565	125520	114199	0.910	0.809	0.995	0.993	49.9	49.8	49.9
2	89251	111366	100297	0.901	0.801	0.985	0.984	49.6	49.6	49.6
3	96652	120412	108722	0.903	0.803	0.987	0.985	49.7	49.6	49.7
4	109092	134475	123086	0.915	0.811	1.001	0.996	50.0	49.9	50.0
5	116919	145099	131652	0.907	0.806	0.992	0.989	49.8	49.7	49.8
6	104317	128932	116012	0.900	0.809	0.984	0.993	49.6	49.8	49.7
7	112385	138326	125409	0.907	0.812	0.992	0.997	49.8	49.9	49.9
8	126074	154731	141406	0.914	0.815	0.999	1.000	50.0	50.0	50.0

*variance = 0.017, standard deviation = 0.140

experiments applied these procedures to estimate the base composition of a virus and two plasmids whose base compositions are known to range from 44.8 and 50.7%. In each case, base composition determined by BRAT agreed to within 0.5% of the values determined by complete sequencing of the viruses and plasmids (Table 2).

Analyses of genomic DNA

To test BRAT on more complex source DNA, we applied this technique to a variety of bacterial DNAs, human genomic DNA and yeast artificial chromosomes. Estimates of GC content for the six bacterial species range from 44% to 61% as determined by other methods including thermal denaturation, HPLC and buoyant density. GC contents obtained for these genomic DNAs using BRAT corresponded closely with previously reported base compositions (Table 3). Four additional bacterial samples, *Escherichia fergusonii*, *Serratia odorifera* and two strains of *Escherichia vulneris*, of unknown base composition were analyzed by BRAT and estimates of their base compositions are also reported in Table 3. Estimates of the base composition of total human genomic DNA range from 38.1 to 43.4% G+C (17) and we obtained a value of 41.7% G+C by BRAT.

The base composition of yeast artificial chromosomes isolated from two separate loci in the human genome: the hypoxanthine

phosphoribosyltransferase (HPRT) gene, and a series of five clones from the cystic fibrosis gene covering almost 1.2 megabases, were determined by BRAT (Table 4). Approximately 5 ng of YAC DNA was used in each BRAT analysis and the %G+C estimates in Figure 2 are the averages of at least three separate runs for each sample (Table 4). Smaller quantities of starting material isolated for the YACs resulted in slightly greater variability (~1%) in the base composition estimates among replicates than those obtained for lambda standards or bacterial genomic DNA. Across 1.2 megabases of the cystic fibrosis locus, the average base composition of individual YACs are within 0.5% G+C. The base composition of 41.3% G+C for the yHPRT clone is similar to the 40.3% G+C determined from 56 kb of contiguous sequence from the HPRT gene (20).

DISCUSSION

Variation in the base composition of long regions of the genomes of warm blooded vertebrates has been suggested to play important roles in the regulation of gene expression (4,21), chromatin structure (22), codon usage (6), and rates of evolution in the genome (8,9,23). New cloning systems, such as yeast artificial

Table 2. Comparisons of base compositions determined by nucleotide sequencing and BRAT.

DNA sample	Replicate	Corrected GA	Corrected CA	%G+C from GA	%G+C from CA	Average GC content	Sequence Determination
ϕX 174	1	0.830	0.821	45.4	45.1	45.2	44.8 (31)
	2	0.810	0.792	44.8	44.2	44.5	
BlueScript	1	1.004	1.064	50.1	51.6	50.8	50.3 (32)
	2	1.015	1.070	50.4	50.5	50.4	
pUC19	1	1.015	1.070	50.4	51.7	51.0	50.7 (33)
	2	1.002	1.063	50.0	51.5	50.8	

Table 3. Genomic DNA base compositions determined by BRAT.

Organism	Replicate	Corrected GA	Corrected CA	%G+C from GA	%G+C from CA	Average GC content	previous estimates(19)
<i>Homo sapiens</i>	1	0.696	0.744	41.0	42.7	41.8	38.1 – 43.4
	2	0.693	0.739	40.9	42.5	41.7	
	3	0.688	0.740	40.8	42.5	41.6	
<i>B. subtilis</i>	1	0.763	0.846	43.5	45.8	44.7	43.9 – 44.9
	2	0.775	0.843	43.7	45.7	44.7	
<i>S. dysenteriae</i>	1	0.955	1.050	48.9	51.3	50.1	50.0 – 53.4
	2	1.003	1.101	50.1	52.4	51.2	
<i>E. coli</i> B	1	0.974	1.015	49.4	50.4	49.9	50.0 – 52.5
	2	0.993	1.028	49.8	50.7	50.2	
<i>E. coli</i> K12	1	1.061	1.051	51.5	51.3	51.4	49.8 – 51.5
	2	1.038	1.042	50.9	51.0	51.0	
<i>E. fergusonii</i>	1	1.083	1.068	51.8	51.5	51.7	n.a.
	2	1.001	1.082	50.0	52.0	51.0	
<i>S. typhimurium</i>	1	1.181	1.206	54.2	54.5	54.3	51.5 – 54.0
	2	1.123	1.181	52.9	54.2	53.5	
<i>E. vulneris</i> 1	1	1.290	1.252	56.3	55.6	56.0	n.a.
	2	1.170	1.209	53.9	55.7	54.8	
<i>E. vulneris</i> 2	1	1.371	1.401	57.6	58.2	57.9	n.a.
	2	1.290	1.345	56.3	57.4	56.8	
<i>S. odorifera</i>	1	1.353	1.352	57.6	57.6	57.6	n.a.
	2	1.263	1.354	55.8	57.5	56.7	
<i>S. marsecensis</i>	1	1.60	1.57	61.6	61.0	61.3	57.5 – 60.9
	2	1.435	1.634	58.9	62.0	60.5	

E. vulneris 1 = ATCC 29943, *E. vulneris* 2 = ATCC 33821, *S. marsecensis* = ATCC 13880, *S. odorifera* = ATCC 3307, *S. dysenteriae* = ATCC 13313, *E. fergusonii* = ATCC 35469, *S. typhimurium* = LT2. n.a. = not available.

Table 4. Base composition of yeast artificial chromosomes containing of human DNA inserts.

YAC	Replicate	Corrected GA	Corrected CA	%G+C from GA	%G+C from CA	Average GC content	Std. dev.
yKM19-3 (180 kb)	1	0.737	0.799	42.4	44.4	43.4	0.58
	2	0.703	0.771	41.3	43.5	42.4	
	3	0.707	0.766	41.4	43.4	42.4	
yCF-4 (330 kb)	1	0.724	0.766	42.0	43.4	42.7	0.45
	2	0.708	0.731	41.4	42.2	41.8	
	3	0.701	0.759	41.2	43.2	42.2	
yW30-5 (300 kb)	1	0.722	0.747	41.9	42.7	42.3	0.47
	2	0.744	0.765	42.7	43.3	43.0	
	3	0.713	0.739	41.6	42.5	42.1	
yJ311-2 (230 kb)	1	0.725	0.747	42.0	42.8	42.4	0.10
	2	0.728	0.757	42.1	42.4	42.6	
	3	0.728	0.736	42.1	42.4	42.5	
yJ311-5 (200 kb)	1	0.695	0.726	41.0	42.1	41.5	0.70
	2	0.746	0.756	42.7	43.0	42.9	
	3	0.731	0.733	42.2	42.3	42.3	
yHPRT (680 kb)	1	0.718	0.744	41.8	42.7	42.2	0.46
	2	0.718	0.725	41.8	42.0	41.9	
	3	0.694	0.713	41.0	41.6	41.3

Figure 2. GC content of yeast artificial chromosomes spanning the chromosomal region containing the gene encoding cystic fibrosis transmembrane conductance regulator. Designation of YAC clones follows that of Green and Olson (23). %G+C and the relative position and size of each clone is shown.

chromosomes, allow the manipulation of clones approaching the size of isochores and should be useful for further study at this level of compartmentalization within eukaryotic genomes. We describe a widely applicable procedure that allows the rapid and accurate determination of the base composition of DNA samples using nanogram quantities of starting material, enabling the use of YACs for such studies.

This new method, base ratio analysis using thin-layer chromatography (BRAT), provides reproducible estimates of GC content accurate to within 0.5% G+C when using 100 ng quantities, and consistent to within 1.0% using 5 ng quantities, of sample DNA. Base composition estimates made with this technique are in close agreement over a broad range of GC content with those determined for bacterial, viral and plasmid DNA by sequencing, thermal denaturation, HPLC, or buoyant density experiments (Tables 2 and 3).

We have applied this method to determine the base composition of genomic DNAs and large fragments of cloned human DNA. BRAT experiments yielded base content estimates using nanogram quantities of YAC clones and lambda DNA that are consistent over subsequent runs. The slight increase in the variability of BRAT results seen using YAC DNA relative to genomic and density gradient purified viral and plasmid DNAs may be caused by three factors. First, an increase in the relative amount of random error with the decrease in the amount of starting material could be responsible for the increased variability. Similarly, a reduction in the number of counts scanned by the betascope due to a decrease in the incorporation of [α - 32 P] dNTPs in the nick translation process may reduce the technique's resolving power. Finally, minor contamination from yeast chromosomes migrating to a similar mobility as the YACs on CHEF gels may occur, although an increase in the variability seen using decreasing amounts of lambda DNA purified by CsCl gradients is also seen. Since the 9.9 kb YAC vector contributes <5% of the total DNA in each YAC clone and its base composition is 43.7% G+C (23), it does not have an appreciable effect on the estimates of %G+C for each insert.

Base ratio analysis of YACs from the cystic fibrosis locus in humans, one of the longest contiguously cloned regions (25), reveals that the base composition of this region is remarkably uniform over almost 1.2 megabases (Figure 2). Recent sequence analysis confirms that the coding regions of the cystic fibrosis transmembrane conductance regulator gene (26) are relatively uniform in their base compositions across the entire gene. Moreover, the G+C content at the third positions of codons—which is generally indicative of overall base composition of a region (4)—is 43.1% in agreement with the values of 42.4 to 42.7% obtained by BRAT from the YACs of that region. The GC contents of the five YACs containing the cystic fibrosis gene indicate that it would likely lie in the L2 fraction defined by buoyant density gradients (27). The L2 fraction has been estimated to be 41% G+C, contain over 30% of all non-satellite DNA in the human genome and is enriched in tissue specific genes relative to regions with higher GC content (4). However, estimates of GC contents based on fractionation by density gradients often have poor resolution due to the pooling of replicate gradient fractions from different preparations (28) and can contain fragments differing by as much as 2% G+C. Data obtained by BRAT for the 680 kb YAC clone yHPRT which contains the human hypoxanthine phosphoribosyltransferase gene (29) yielded an estimated GC content of 41.3% and indicates that it would be found in the same gradient fraction as the cystic fibrosis gene.

Given that very little difference in base composition was observed over the entire region, it is reasonable to assume that the cystic fibrosis gene does not span two large adjoining regions of substantially differing GC content. Buoyant density experiments have predicted that isochores are much longer than 300 kb in length (7) and it is possible that the entire cystic fibrosis region lies within one such region. The homogeneity of base composition over the almost 1.2 Mb cloned in YACs from the cystic fibrosis locus indicates that isochores within the human genome could indeed be much longer than 300 kb and that they may approach the length of 1–2 Mb required for them to account for the banding pattern of metaphase chromosomes (28).

Base composition of genomic DNA has long been used as a criterion in the classification of bacteria, and the GC contents of several hundred bacterial isolates have been determined (19). The ten bacterial species analyzed in this study include four bacterial lineages—*E. fergusonii*, *S. odorifera*, and two strains classified as *E. vulneris*—whose base compositions have not been previously determined. *E. fergusonii* is very closely related to *E. coli* (30) and it is not surprising that its GC content of 51.4% is close to the range (50.1 to 51.2%) determined by BRAT for two isolates of *E. coli*. Despite the original classification of the two isolates of *Escherichia vulneris* into a single species, these isolates are very distantly related based on the nucleotide sequences of two protein-coding regions (30). Their base compositions estimated by BRAT (56.0% and 57.9% G+C) also reflect this genetic dissimilarity and further argues that these strains are members of different species. Among enteric bacteria, species of *Serratia* are typically characterized by high GC contents and the genomes of *S. odorifera* and *S. marcescens* were estimated by BRAT as being 57.2% and 60.9% G+C, respectively.

Base ratio analysis using thin-layer chromatography or BRAT, is a sensitive and reliable alternative to traditional methods of determining base composition. Its ability to reproducibly provide estimates of GC content accurate to within 1% using as little as 5 ng of DNA make it especially well suited for the analysis of the genome organization of eukaryotes using the large clones which can be obtained with the yeast artificial chromosome cloning system. The use of BRAT on YACs from homologous regions of different species may also give useful insights into the way in which isochores have evolved. As the number of available YAC contigs increases BRAT should allow the rapid determination of the size of isochores in the human genome as well as the characteristics of their boundaries. And, as more detailed physical maps of genomes made from clones in YAC libraries become available, BRAT should also allow the construction of detailed compositional map of eukaryotic genomes. BRAT analyses of YACs from diverse species could be used to conduct a rapid and sensitive survey that would reveal the extent of compartmentalization in the genomes of eukaryotic species.

ACKNOWLEDGEMENTS

The authors wish to thank the Isaac Wong for assistance with TLC, Clare Huxley and Eric Green for supplying YACs, Jeffrey Lawrence for bacterial DNAs, and John Carulli and Jeffrey Lawrence for comments of the manuscript. This work was supported by grants from the National Institutes of Health, National Science Foundation and the Lucille P. Markey Charitable Trust.

REFERENCES

1. Marmur, J. and Doty, P. (1962) *J. Mol. Biol.*, **5**, 109–118.
2. Schildkraut, C.L., Marmur, J. and Doty, P. (1962) *J. Mol. Biol.*, **4**, 430–443.
3. Muto, A. and Osawa, S. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 166–169.
4. Bernardi, G. (1989) *Ann. Rev. Genet.* **23**, 637–661.
5. Ikemura, T., Wada, K.-W. and Aota, S.-I. (1990) *Genomics* **8**, 207–216.
6. Bernardi, G. and Bernardi, G. (1985) *J. Mol. Evol.* **22**, 363–365.
7. Bernardi, G., Olofsson, B., Filipinski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M. and Rodier, F. (1985) *Science* **228**, 953–958.
8. Hardison, R., Krane, D., Vandenberg, D., Cheng, J.-F., Mansberger, J., Taddie, J., Schwartz, S., Huang, X. and Miller, W. (1991) *J. Mol. Biol.* in press.
9. Bernardi, G. and Bernardi, G. (1991) *J. Mol. Evol.* **33**, 57–67.
10. Aissani, B., D'Onofrio, G., Mouchiroud, D., Gardiner, K. and Bernardi, G. (1991) *J. Mol. Evol.* **33**, 493–503.
11. Burke, D.T., Carle, G.F. and Olson, M.V. (1987) *Science* **236**, 806–811.
12. Gehrke, C.W., McCune, R.A., Gama-Sosa, M.A., Ehrlich, M. and Kuo K.C. (1984) *J. Chrom.* **301**, 199–219.
13. Schildkraut, C.L. and Maio, J.J. (1969) *J. Mol. Biol.* **46**, 305–312.
14. Mabuchi, T. and Nishikawa, S. (1990) *Nucleic Acids Res.* **18**, 7461–7462.
15. Chu, G., Vollrath, D. and Davis, R. (1986) *Science* **234**, 1582.
16. Marmur, J. (1961) *J. Mol. Biol.* **3**, 208–221.
17. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) *J. Molec. Biol.* **113**, 237.
18. Daniels, D.R., Schroeder, J.L., Szybalski, W., Sanger, F., Coulson, A.R., Hong, G.F., Hill, D.F., Petersen, G.B., and Blattner, F.B. (1983) In Hendrix, R.W., Roberts, J.W., Stahl, F.W. and Weisberg, R.A. (eds.) *Lambda II*, Cold Spring Harbor Laboratory, NY, pp. 519–674.
19. Normore, W.M. (1976) In Fasman, G.D. (ed.) *CRC Handbook of Biochemistry and Molecular Biology*, vol. II, 3rd edition, CRC Press, Cleveland. p. 65.
20. Edwards, A., Voss, H., Rice, P., Sivitello, A., Stegemann, J., Schwager, C., Zimmermann, J., Erfle, H., Caskey, C.T. and Ansorge, W. (1990) *Genomics* **6**, 593–608.
21. Bird, A. (1986) *Nature* **321**, 209–213.
22. Tazi, J. and Bird, A. (1990) *Cell* **60**, 909–920.
23. Wolfe, K.H., Sharp, P.M. and Li, W.-H. (1989) *Nature* **337**, 283–285.
24. Burke, D.T., Carle, G.F. and Olson, M.V. (1987) *Science* **807**, 806–811.
25. Green, E.D. and Olson, M.V. (1990) *Science* **250**, 94–98.
26. Zielenski, J., Rozmahel, R., Bozon, D., Kerem, B.-S., Grezelczak, Z., Riordan, J.R., Rommens, J. and Tsui, L.-C. (1991) *Genomics* **10**, 214–228.
27. Zerial, M., Salinas, J., Filipinski, J. and Bernardi, G. (1986) *Eur. J. Biochem.* **160**, 479–485.
28. Gardiner, K., Aissani, B. and Bernardi, G. (1990) *EMBO J.* **9**, 1853–1858.
29. Huxley, C., Hagino, Y., Schelssinger, D. and Olson, M.V. (1991) *Genomics* **9**, 742–750.
30. Lawrence, J.G., Ochman, H. and Hartl, D.L. (1991) *J. Gen. Microbiol.* in press.
31. Sanger, F., Coulson, A.R., Friedmann, T., Air, G.M., Borrell, B.G., Slocombe, P.M., and Smith, M. (1978) *J. Mol. Biol.* **125**, 225–246.
32. Stratagene, La Jolla, CA.
33. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) *Gene* **33**, 103–119.