
Enzymatic Amplification of β -Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia

Randall K. Saiki, Stephen Scharf, Fred Faloona, Kary B. Mullis, Glenn T. Horn, Henry A. Erlich, Norman Arnheim

Two new methods were used to establish a rapid and highly sensitive prenatal diagnostic test for sickle cell anemia. The first involves the primer-mediated enzymatic amplification of specific β -globin target sequences in genomic DNA, resulting in the exponential increase (220,000 times) of target DNA copies. In the second technique, the presence of the β^A and β^S alleles is determined by restriction endonuclease digestion of an end-labeled oligonucleotide probe hybridized in solution to the amplified β -globin sequences. The β -globin genotype can be determined in less than 1 day on samples containing significantly less than 1 microgram of genomic DNA.

Recent advances in recombinant DNA technology have made possible the molecular analysis and prenatal diagnosis of several human genetic diseases. Fetal DNA obtained by amniocentesis or chorionic villus sampling can be analyzed by restriction enzyme digestion, with subsequent electrophoresis, Southern transfer, and specific hybridization to cloned gene or oligonucleotide probes. With polymorphic DNA markers linked genetically to a specific disease locus, segregation analysis must be carried out with restriction fragment length polymorphisms (RFLP's) found to be informative by examining DNA from family members (1, 2).

Many of the hemoglobinopathies, however, can be detected by more direct methods in which analysis of the fetus alone is sufficient for diagnosis. For example, the diagnosis of hydrops fetalis (homozygous α -thalassemia) can be made by documenting the absence of any α -globin genes by hybridization with an α -globin probe (3-5). Homozygosity for certain β -thalassemia al-

leles can be determined in Southern transfer experiments by using oligonucleotide probes that form stable duplexes with the normal β -globin gene sequence but form unstable hybrids with specific mutants (6, 7).

Sickle cell anemia can also be diagnosed by direct analysis of fetal DNA. This disease results from homozygosity of the sickle-cell allele (β^S) at the β -globin gene locus. The S allele differs from the wild-type allele (β^A) by substitution of an A in the wild-type to a T at the second position of the sixth codon of the β chain gene, resulting in the replacement of a glutamic acid by a valine in the expressed protein. For the prenatal diagnosis of sickle cell anemia, DNA obtained by amniocentesis or chorionic villus sampling can be treated with a restriction endonuclease (for example, Dde I and Mst II) that recognizes a sequence altered by the β^S mutation (8-11). This generates β^A - and β^S -specific restriction fragments that can be re-solved by Southern transfer and hybridization with a β -globin probe.

We have developed a procedure for the detection of the sickle cell mutation that is very rapid and is at least two orders of magnitude more sensitive than standard Southern blotting. There are two special features to this protocol. The first is a method for amplifying specific β -globin DNA sequences with the use of oligonucleotide primers and DNA

The authors are in the Department of Human Genetics, Cetus Corporation, 1400 Fifty-Third Street, Emeryville, California 94608. The present address for N.A. is Department of Biological Sciences, University of Southern California, Los Angeles 90089-0371.

QIAGEN Exhibit 1018.0001

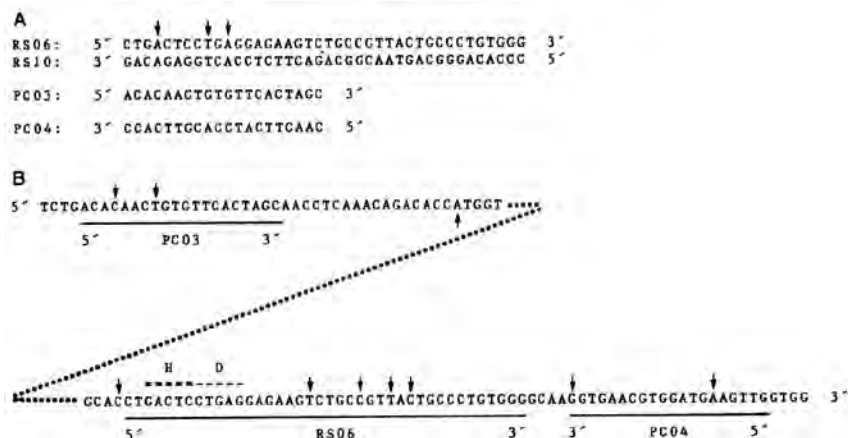


Fig. 1. Sequence of synthetic oligonucleotide primers and probe and their relation to the target β -globin region. (A) The primer PC03 is complementary to the (-)-strand and the primer PC04 is complementary to the (+)-strand of the β -globin gene. The probe RS06 is complementary to the (-)-strand of the wild-type (β^A) sequence of β -globin. RS10 is the "blocking oligomer", an oligomer complementary to the RS06 probe except for three nucleotides, indicated by the downward arrows. It is added before enzyme digestion to the OR reaction to anneal to the excess RS06 oligomer and prevent nonspecific cleavage products due to hybridization of RS06 to nontarget DNA (13). Because of the mismatches within the Dde I and Hinf I restriction sites, the RS06/RS10 duplex is not cleaved by Dde I and Hinf I digestion. (B) The relation between the primers, the probe, and the target β -globin sequence. The upward arrow indicates the β -globin initiation codon. The downward arrows indicate nucleotide differences between β - and δ -globin. The polymorphic Dde I site (CTCAG) is represented by a single horizontal dashed line (D), and the invariant Hinf I (GACTC) site is represented by double horizontal dashed lines (H).

polymerase (12). The second is the analysis of the β -globin genotype by solution hybridization of the amplified DNA with a specific oligonucleotide probe and subsequent digestion with a restriction endonuclease (13). These two techniques increase the speed and sensitivity, and lessen the complexity of prenatal diagnosis for sickle cell anemia; they may also be generally applicable to the diagnosis of other genetic diseases and in the use of DNA probes for infectious disease diagnosis.

Sequence amplification by polymerase chain reaction. We use a two-step procedure for determining the β -globin genotype of human genomic DNA samples. First, a small portion of the β -globin gene sequence spanning the polymorphic Dde I restriction site diagnostic of the β^A allele is amplified. Next, the presence or absence of the Dde I restriction site in the amplified DNA sample is determined by solution hybridization with an end-labeled complementary oligomer fol-

lowed by restriction endonuclease digestion, electrophoresis, and autoradiography.

The β -globin gene segment was amplified by the polymerase chain reaction (PCR) procedure of Mullis and Faloona (12) in which we used two 20-base oligonucleotide primers that flank the region to be amplified. One primer, PC04, is complementary to the (+)-strand and the other, PC03, is complementary to the (-)-strand (Fig. 1). The annealing of PC04 to the (+)-strand of denatured genomic DNA followed by extension with the Klenow fragment of *Escherichia coli* DNA polymerase I and deoxynucleotide triphosphates results in the synthesis of a (-)-strand fragment containing the target sequence. At the same time, a similar reaction occurs with PC03, creating a new (+)-strand. Since these newly synthesized DNA strands are themselves template for the PCR primers, repeated cycles of denaturation, primer annealing, and extension result in the exponential accumulation of the 110-base pair region defined

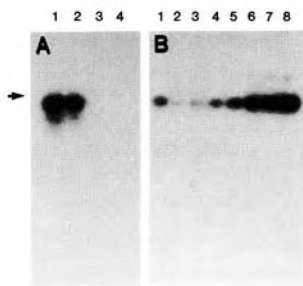


Fig. 2. Southern analysis of PCR amplified genomic DNA with the RS06 probe. (A) Samples (1 μ g) of genomic DNA were dispensed in microcentrifuge tubes and adjusted to 100 μ l in a buffer containing 10 mM tris, pH 7.5, 50 mM NaCl, 10 mM MgCl₂, 1.5 mM deoxynucleotide triphosphate [dNTP] each of all four was used, 1 μ M PC03, and 1 μ M PC04. After heating for 5 minutes at 95°C (to denature the genomic DNA), the tubes were centrifuged for 10 seconds in a microcentrifuge to remove the condensation. The samples were immediately transferred to a 30°C heat block for 2 minutes to allow the PC03 and PC04 primers to anneal to their target sequences. At the end of this period, 2 μ l of the Klenow fragment of *E. coli* DNA polymerase I (Biolabs, 0.5 unit/ μ l in 10 mM tris, pH 7.5, 50 mM NaCl, 10 mM

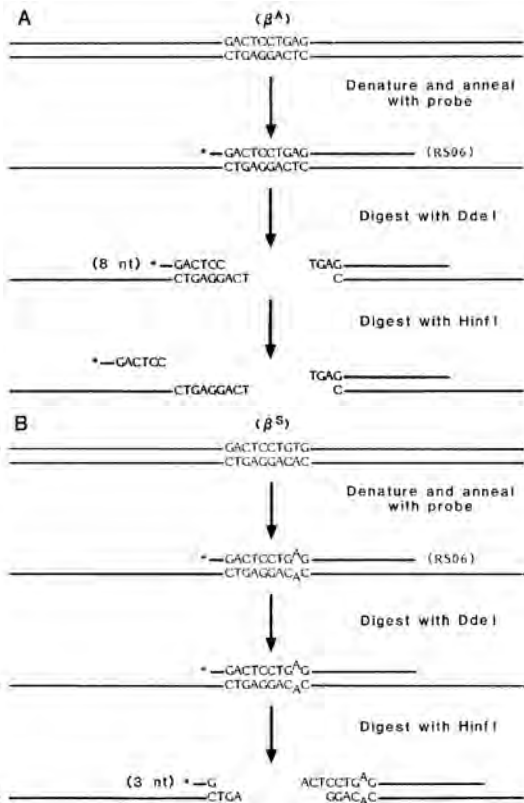
MgCl₂) was added, and the incubation was allowed to proceed for an additional 2 minutes at 30°C. This cycle—denaturation, centrifugation, hybridization, and extension—was repeated 19 more times, except that subsequent denaturations were done for 2 instead of 5 minutes. (The final volume after 20 cycles was 140 μ l.) Thirty-six nanograms of the amplified genomic DNA (5 μ l) were applied to a 4 percent Nusieve (FMC) alkaline agarose minigel and subjected to electrophoresis (50 V), for 2 hours until the bromcresol green dye front reached 4 cm. After neutralization and transfer to Genetrans nylon membrane (Plasco), the filter was “prehybridized” in 10 ml 3x SSPE (1 x SSPE is 0.18M NaCl, 10 mM NaH₂PO₄, 1 mM EDTA, pH 7.4), 5x DET (1x DET is 0.02 percent each polyvinylpyrrolidone, Ficoll, and bovine serum albumin; 0.2 mM tris, 0.2 mM EDTA, pH 8.0), 0.5 percent SDS, and 30 percent formamide for 4 hours at 42°C. Hybridization with 1.0 pmol of phosphorylated (with [γ -³²P]ATP) RS06 (~5 μ Ci/pmol) in 10 ml of the same buffer was carried out for 18 hours at 42°C. The filter was washed twice in 2x SSPE, 0.1 percent sodium dodecyl sulfate (SDS) at room temperature for 30 minutes, and autoradiographed at -70°C for 2 hours with a single intensification screen. (Lanes 1 to 3) DNA's isolated from the cell lines Molt4, SC01, and GM2064, respectively. Molt4 is homozygous for the normal, wild-type allele of β -globin ($\beta^A\beta^A$), SC-1 is homozygous for the sickle cell allele ($\beta^S\beta^S$), and GM2064 is a cell line in which the β - and δ -globin genes have been deleted ($\Delta\Delta$) (13). (Lane 4) Contains 36 ng of Molt4 DNA that was not PCR amplified. The horizontal arrow indicates the position of a 114-base marker fragment obtained by digestion of pBR328 with Nar I. (B) Thirty-six nanograms of 20-cycle amplified Molt4 DNA (see above) was loaded onto a Nusieve gel along with measured amounts of Hae III-Mae I digested pBR328: β^A (13) calculated to represent the molar increase in β -globin target sequences at PCR efficiencies of 70, 75, 80, 85, 90, 95, and 100 percent (lanes 2 to 8, respectively). DNA was transferred to Genetrans and hybridized with the labeled RS06 probe as described above. (Lane 1) Molt4 DNA (36 ng); (lanes 2 to 8) 7.3 $\times 10^{-4}$ pmol, 1.3 $\times 10^{-3}$ pmol, 2.3 $\times 10^{-3}$ pmol, 4.0 $\times 10^{-3}$ pmol, 6.8 $\times 10^{-3}$ pmol, 1.1 $\times 10^{-2}$ pmol, and 1.9 $\times 10^{-2}$ pmol of pBR328: β^A , respectively (20).

by the primers.

An example of the degree of specific gene amplification achieved by the PCR method is shown in Fig. 2A. Samples of DNA (1 μ g) were amplified for 20 cycles and a fraction of each sample, equivalent to 36 ng of the original DNA, was subjected to alkaline gel electrophoresis and transferred to a nylon filter. The filter was then hybridized with a ³²P-labeled 40-base oligonucleotide probe, RS06, which is complementary to the target sequence (Fig. 1A) but not to the PCR primers. The results, after a 2-hour autoradiographic exposure, show that a fragment hybridizing with the RS06 probe migrates at the position expected of the amplified target DNA segment (110 bases) (lanes 1 and 2). No hybridization with the RS06 probe could

be detected in unamplified DNA (lane 4). When PCR amplification was performed on a DNA sample derived from an individual with hereditary persistence of fetal hemoglobin in which both the δ - and β -globin genes are deleted (14), again no 110-base fragment was detected (lane 3). To estimate the yield and efficiency of 20 cycles of PCR amplification, we prepared a Southern blot that contained 36 ng of an amplified genomic DNA sample and a dilution series consisting of various amounts of cloned β -globin sequence. The efficiency was calculated according to the formula: $(1 + X)^n = Y$, where X is the mean efficiency per cycle, n is the number of PCR cycles, and Y is the extent of amplification (yield) after n cycles (for example, a 200,000-fold increase after 20 cycles). The

Fig. 3. Schematic diagram of oligomer restriction by sequential digestion to identify β^A - and β^S -specific cleavage products. The DNA sequences shown are the regions of the β -globin genomic DNA and the RS06 hybridization probe containing the invariant Hinf I site (GANTC, where N represents any nucleotide) and the polymorphic Dde I site (CTNAG). The remaining DNA sequences are represented as solid horizontal lines. The asterisk indicates the position of the radioactive ^{32}P label attached to the 5'-end of the RS06 probe with polynucleotide kinase. (A) Outline of the procedure and expected results when RS06 anneals to the normal β -globin gene (β^A). After denaturation of the genomic DNA and hybridization of the labeled RS06 probe to the complementary target sequence in the β^A gene, digestion of the probe-target hybrid with Dde I causes the release of a labeled (8-nt) cleavage product. Because of the relatively stringent conditions during Dde I digestion, the 8-nt cleavage product dissociates from the genomic DNA and the subsequent digestion with Hinf I has no effect. (B) Outline of Dde I and Hinf I digestion after hybridization of the RS06 probe to the sickle cell allele (β^S). As a consequence of the β^S mutation, the probe-target hybrid contains an A-A mismatch within the Dde I site and is not cleaved by the Dde I endonuclease. The Hinf I site, however, remains intact and digestion with that enzyme generates a labeled 3-nt product. Thus, the presence of the β^A allele is revealed by the release of a labeled 8-nt fragment, while the presence of β^S is indicated by a labeled 3-nt fragment.



amounts of cloned β -globin sequences used in this experiment were calculated to represent efficiencies of 70 to 100 percent.

The reconstructions were prepared by digesting the β -globin plasmid, pBR328: β^A , with the restriction enzymes Hae III and Mae I. Both of these enzymes cleave the β -globin gene within or very near to the 20 base regions that hybridize to the PCR primers, generating a 103-base pair (bp) fragment that is almost identical in size and composition to the 110-bp segment created by PCR amplification. After hybridization with the RS06 probe and autoradiography, the amplified genomic sample was compared with the known standards, and the result indicated an overall

efficiency of approximately 85 percent (Fig. 2B), representing an amplification of about 220,000 times (1.85^{20}).

Distinguishing the β^A and β^S alleles by the oligomer restriction method. We have previously described a rapid solution hybridization method that can indicate whether a genomic DNA sample contains a specific restriction enzyme site at, in principle, any chromosomal location (13). This method, called oligomer restriction (OR), involves the stringent hybridization of a ^{32}P end-labeled oligonucleotide probe to the specific segment of the denatured genomic DNA which spans the target restriction site. The ability of a mismatch within the restriction

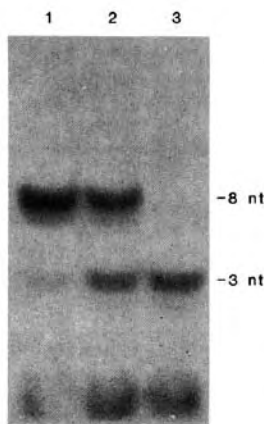


Fig. 4. Demonstration of OR sequential digestion with cloned β -globin genes. The sequential digestion strategy was demonstrated by annealing the RS06 probe to the β -globin plasmids pBR328:: β^A and pBR328:: β^S (73). The methods were similar to those described (73). Cloned β -globin DNA (45 ng; 0.01 pmol) was placed in a microcentrifuge tube, adjusted to 30 μ l with TE buffer (10 mM Tris, 0.1 mM EDTA, pH 8.0), overlaid with 0.1 ml of mineral oil. The DNA was denatured by heating for 5 to 10 minutes at 95°C. Ten microliters of 0.6M NaCl containing 0.02 pmol of phosphorylated (with [γ -32P]ATP) RS06 probe oligomer (~5 μ Ci/pmol) was added and annealed for 60 minutes at 56°C. Unlabeled RS10 blocking oligomer (4 μ l; 200 pmol/ml) (Fig. 1) (73) was then added, and the hybridization was continued for 5 to 10 minutes. Next, 5 μ l of 100 mM $MgCl_2$ and 1 μ l of Dde I (Biolabs, 10 units) was added and incubated for 20 minutes at 56°C; 1 μ l of Hinf I (Biolabs, 10 units) was added and digestion was continued for 20 minutes at the same temperature. The reaction was terminated by the addition of 4 μ l of 100 mM EDTA and 6 μ l of tracking dye to a final volume of 61 μ l; a portion (8 μ l) (6 ng, 0.0013 pmol) was applied to a 0.75-mm thick, 30 percent polyacrylamide minigel (19 acrylamide: 1 bis) in a Hoefer SE200 apparatus and subjected to electrophoresis (300 V) for 1 hour until the bromophenol blue dye front reached 3 cm. The top 1.5 cm of the gel, containing intact RS06, was cut off and discarded. The remaining gel was autoradiographed with a single intensification screen for 18 hours at -70°C. (Lane 1) six nanograms of pBR328:: β^A ; (lane 2) 3 ng of pBR328:: β^A plus 3 ng of pBR328:: β^S ; and (lane 3) 6 ng of pBR328:: β^S .

site to prevent cleavage of the duplex formed between the probe and the target genomic sequence is the basis for detecting allelic variants. The presence of the restriction site in the target DNA is revealed by the appearance of a specific labeled fragment generated by cleavage of the probe.

For the diagnosis of sickle cell anemia, the probe was designed to be complementary to a region of the β -globin gene locus surrounding the sixth codon. In the β^A allele, the nucleotide (nt) sequence at this position contains a Dde I restriction site, but due to the single base mutation, this site is absent in the β^S allele. Our strategy for generating specific probe cleavage products for each allele is shown in Fig. 3. It is based on the presence of an invariant Hinf I restriction enzyme site immediately adjacent to the polymorphic Dde I restriction site. Resolution of the labeled oligomer cleavage products produced by sequential digestion with Dde I and Hinf I allows us to distinguish between the two alleles. In an individual homozygous for the wild-type β -globin allele AA, Dde I digestion will produce a labeled octamer (8 nt) from the probe. Because of its short length, the 8-nt cleavage product will dissociate from the genomic target DNA and the subsequent digestion with Hinf I has no effect. In

the case of SS homozygotes, however, Dde I digestion does not cleave the probe since a base pair mismatch exists in the recognition sequence formed between the probe and target DNA. The invariant Hinf I site will then be cleaved during Hinf I digestion, releasing a labeled trimer (3 nt). In an AS heterozygote, both a trimer and an octamer would be detected. The resolution of the intact 40-base probe, the 8-nt and the 3-nt cleavage products is achieved by polyacrylamide gel electrophoresis. Experiments testing the sequential digestion strategy with plasmids carrying the β^A and β^S alleles show that, in each case, the expected probe cleavage products were produced (Fig. 4).

Analysis of genomic DNA samples by PCR and OR. Eleven DNA samples derived from lymphoblastoid cell lines or white blood cells were analyzed for their β -globin genotype by standard Southern blotting and hybridization of the Mst II RFLP (10), identifying the genotypes of the samples as either AA, AS, or SS. Six of these samples (and one additional one) were then amplified by PCR for 20 cycles starting with 1 μ g of DNA each. An aliquot of the amplified DNA sample (one-fourteenth of the original 1- μ g sample) was hybridized to the RS06 probe and digested with Dde I and then Hinf I. A

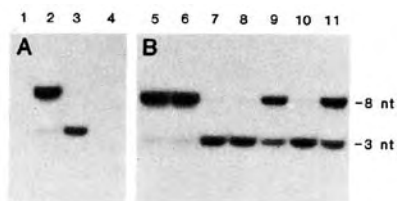


Fig. 5. Determination of the β -globin genotype in human genomic DNA with PCR-OR. Samples (1 μ g) of human genomic DNA were amplified for 20 cycles (as described in Fig. 2A). The amplified DNA's (71 ng) were hybridized to the RS06 probe and serially digested with Dde I and Hinf I (as described in Fig. 4). Each sample (6 μ l) was analyzed by 30 percent polyacrylamide gel electrophoresis and autoradiographed for 6 hours at -70°C with one intensification screen. Each lane contains 7 ng of genomic DNA. (Lane 1) Unamplified Molt4 DNA (negative control); (lane 2) amplified Molt4 ($\beta^A\beta^A$); (lane 3) SC-1 ($\beta^S\beta^S$); (lane 4) GM2064 ($\Delta\Delta$); (lanes 5 to 11) clinical samples CH1 ($\beta^A\beta^A$), CH2 ($\beta^A\beta^A$), CH3 ($\beta^S\beta^S$), CH4 ($\beta^S\beta^S$), CH7 ($\beta^A\beta^S$), CH8 ($\beta^S\beta^S$), and CH12 ($\beta^A\beta^S$), respectively.

portion (one-tenth) of this oligomer restriction reaction was analyzed on a polyacrylamide gel to resolve the cleavage products, and the results obtained after 6 hours of autoradiography are shown Fig. 5. The high sensitivity achieved with the PCR and OR method is demonstrated by the strength of the autoradiographic signal derived from only 1/140 of the original 1- μ g sample (7 ng). Each sample determined to be AA by RFLP analysis showed a strong 8-nt fragment while those typed as SS showed a strong 3-nt fragment. Analysis of the known AS samples revealed both cleavage products.

In the analysis of the AA samples, a faint 3-nt could be detected in addition to the primary 8-nt signal. The reasons for this band remain unclear, although incomplete Dde I cleavage or the occasional failure of the 8-nt fragment to disassociate from the target DNA may contribute to the nonspecific 3-nt product generated by Hinf I digestion. In the analysis of the SS samples, a very faint 8-nt band was also observed in addition to the expected 3-nt signal. We have determined that the background 8-nt product detected in SS samples can be attributed to the δ -globin gene, which is highly homologous

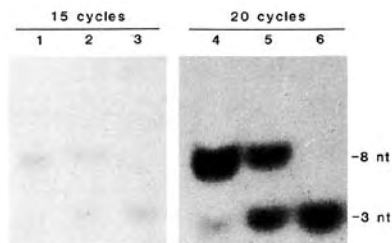


Fig. 6. Effect of cycle number on signal strength. Genomic DNA (1 μ g) from the clinical samples CH2 ($\beta^A\beta^A$), CH12 ($\beta^A\beta^S$), and CH5 ($\beta^S\beta^S$) were amplified for 15 and 20 cycles and equivalent amounts of genomic DNA (80 ng) were analyzed by oligomer restriction. (Lanes 1 to 3) DNA (20 ng) from CH2, CH12, and CH5, respectively, amplified for 15 cycles; (lanes 4 to 6) DNA (20ng) from CH2, CH12, and CH5, respectively, amplified for 15 cycles; (lanes 4 to 6) DNA (20 ng) from CH2, CH12, and CH5, respectively, amplified for 20 cycles. Autoradiographic exposure was for 2.5 hours at -70°C with one intensification screen.

to β -globin. The nucleotide sequence of the two β -globin primers used for amplification is shown in Fig. 1. The downward pointing arrows indicate the differences between the β - and δ -globin genes. We hypothesized that the faint 8-nt signal observed in the SS samples was due to some amplification of the δ -globin gene by these primers and the subsequent cross-hybridization of the amplified δ sequences with the RS06 probe used in the OR procedure. δ -Globin has the same Dde I site as normal β -globin, and the duplex formed between an amplified δ gene segment and the RS06 probe would be expected to yield an 8-nt fragment on Dde I digestion even though there are sequence differences (four mismatch out of 40 bases) between RS06 and δ -globin. It is likely that δ -globin sequences may be amplified to some extent and detected weakly with the RS06 probe in all DNA samples, but that its contribution to the total signal is very small and detectable only when the sample is SS and no 8-nt fragment from the β -globin gene is expected. We tested this hypothesis by treating an SS DNA sample before amplification with the enzyme Mbo I. Since there is a recognition site for this enzyme in the target DNA of the

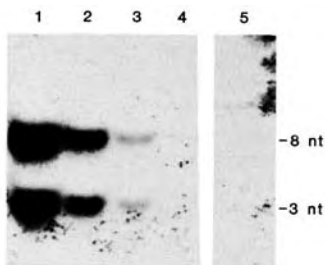


Fig. 7. Detection threshold for PCR-OR. Fivefold serial dilutions of genomic DNA (500, 100, 20, and 4 ng) from the sample CH12 ($\beta^A\beta^S$) were amplified by 20 cycles of PCR and one-tenth each reaction (50, 10, 2, and 0.4 ng) was analyzed by OR. The gel continued (lane 1) genomic DNA; (12.5 ng); (lane 2) 2.5 ng; (lane 3) 0.5 ng; (lane 4) 0.1 ng; (lane 5) 12.5 ng genomic DNA from the globin deletion cell line GM2064. Autoradiographic exposure was for 20 hours at -70°C with an intensification screen.

δ - but not the β -globin gene, cleavage of the δ gene between the regions that hybridize to the PCR primers should prevent its subsequent amplification (but not of β -globin). Our results showed that an SS DNA sample, first digested with Mbo I, gave only the 3-nt product but not the 8-nt product, this is consistent with the hypothesis of δ -globin amplification.

Effect of PCR cycle number on detection threshold. The strength of the autoradiograph signal detected by OR as a function of PCR cycle number and autoradiographic exposure was examined. The signal intensity after 20 cycles is at least 20 times as strong as that for 15 cycles and the determination of the β -globin genotype can be made with an autoradiographic exposure for only 2 hours (Fig. 6). The observed increase of ≥ 20 -fold is consistent with our estimates of 85 percent efficiency per cycle, calculated from the data in Fig. 2B ($1.85^5 = 21.7$). Coupled with the time that it takes to actually carry out the PCR and OR procedures, a 20-cycle PCR allows a diagnosis to be made in less than 10 hours with a DNA sample of $1 \mu\text{g}$.

Since all of the previous PCR experiments were done with $1 \mu\text{g}$ of genomic DNA, we explored the effect of using significantly smaller amounts of DNA as template for PCR amplification. The results obtained with

20 cycles of PCR amplification on 500, 100, 20, and 4 ng of DNA from an AS individual are shown in Fig. 7. After analysis of 1/40 of each sample by the OR procedure and a 20-hour autoradiographic exposure, the β -globin genotype could be easily determined on DNA samples of 20 ng or about 100 times less than is needed for a typical Southern transfer and hybridization experiment. In this experiment, only a small fraction (1/40) of the starting material was placed on the gel; therefore it should be possible to analyze samples of less than 20 ng of genomic DNA (20 ng is equivalent to approximately 6000 haploid genomes) if a larger proportion of the material was utilized in the OR and gel electrophoresis steps.

Diagnostic applications of the PCR-OR system. When currently available methods are used, the completion of a prenatal diagnosis for sickle cell anemia takes a period of several days after the DNA is isolated. With $1 \mu\text{g}$ of genomic DNA, the β -globin genotype can be determined by the PCR-OR method in less than 10 hours; 20 cycles of amplification requires about 2 hours (each full cycle takes 6 to 7 minutes in our protocol), the oligomer restriction procedure involving liquid hybridization and enzyme digestions require an additional 2 hours, and the electrophoresis takes about an hour. Autoradiographic exposure for 4 hours is sufficient to generate a strong signal. Because this method includes a liquid hybridization protocol and involves the serial addition of reagents to a single tube, it is simpler to perform than the standard Southern transfer and hybridization procedure. Prior to electrophoresis, all of the reactions can be done in two small microcentrifuge tubes and could readily be automated.

The sensitivity, as well as the speed and simplicity, of this procedure is also important for clinical applications. Twenty nanograms of starting material can provide an easily detectable result in an overnight autoradiographic exposure. This sensitivity makes the PCR-OR method particularly valuable in cases where poor DNA yields are obtained from prenatal samples. In addition, DNA samples of poor quality (very low average

molecular weight) can give excellent results in the PCR-OR protocol.

The PCR method is likely to be generally applicable for specific gene amplification since a fragment encoding a portion of the HLA-DQ α locus has recently been amplified with this procedure (15). We have carried out PCR amplification on a 110-bp β -globin sequence with an overall efficiency per cycle of about 85 percent. We have also amplified longer β -globin fragments (up to 267 bp), but the yield was lower under our standard conditions. Efficient amplification of a 267-bp fragment required some variation in the PCR procedure. In principle, increasing the number of PCR cycles should yield even greater amplification than that reported here (~220,000-fold after 20 cycles).

Our method for the diagnosis of sickle cell anemia involves the coupling of the PCR procedure with that of oligomer restriction. It was designed to distinguish between two alleles that differ by a polymorphic restriction site. The PCR-OR method is applicable as well to the diagnosis of other diseases where the lesion directly affects a restriction enzyme site or where the polymorphic site is in strong linkage disequilibrium with the disease causing locus. If the polymorphism is in linkage equilibrium with the disease, PCR-OR requires family studies to follow the inheritance of the disease locus.

In the case of the β -globin locus, the presence of the invariant Hinf I restriction site adjacent to the polymorphic Dde I site allows a sequential digestion procedure to identify both the β^A and β^S alleles. In principle, this approach does not require that the two sites be immediately adjacent but only that the cleavage product generated by digestion at the polymorphic site dissociate from the target to prevent cutting at the invariant site. Since the restriction enzyme digestion conditions used here are fairly stringent for hybridization, we estimate that the polymorphic and invariant sites could perhaps be separated by as much as 20 bp.

The application of the PCR method to prenatal diagnosis does not necessarily depend on a polymorphic restriction site or on the use of radioactive probes. In fact, the

significant amplification of target sequences achieved by the PCR method allows the use of nonisotopically labeled probes (16). Amplified target sequences could also be analyzed by a number of other procedures including those involving the hybridization of small labeled oligomers which will form stable duplexes only if perfectly matched (6, 7, 17, 18) and the recently reported method based on the electrophoretic shifts of duplexes with base pair mismatches (19). The ability of PCR procedure to amplify a target DNA segment in genomic DNA raises the possibility that its use may extend beyond that of prenatal diagnosis to other areas of molecular biology.

References and Notes

1. Y. W. Kan and A. M. Dozy, *Proc. Natl. Acad. U.S.A.* **75**, 5631 (1978).
2. K. E. Davies and D. P. Ellis, *Biochem. J.* **226**, 1 (1985).
3. Y. W. Kan, M. S. Golbus, A. M. Dozy, *N. Eng. J. Med.* **295**, 1165 (1976).
4. A. M. Dozy et al., *J. Am. Med. Assoc.* **241**, 1610 (1979).
5. E. M. Rubin and Y. W. Kan, *Lancet* **1985-I**, 75 (1985).
6. M. Pirastu et al., *N. Eng. J. Med.* **309**, 284 (1983).
7. S. H. Orkin, A. F. Markham, H. H. Kazazian, Jr., *J. Clin. Invest.* **71**, 775 (1983).
8. R. F. Geever et al., *Proc. Natl. Acad. Sci. U.S.A.* **78**, 5081 (1981).
9. J. C. Chang and Y. W. Kan, *N. Eng. J. Med.* **307**, 30 (1982).
10. S. H. Orkin et al., *ibid.*, p. 32.
11. J. T. Wilson et al., *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3628 (1982).
12. F. Faloona and K. Mullis, in preparation.
13. R. Saiki, N. Arnheim, H. Erlich, *Biotechnology* **3**, 1008 (1985).
14. D. Tuan et al., *Proc. Natl. Acad. Sci. U.S.A.* **80**, 6937 (1983).
15. S. Scharf, unpublished data.
16. R. Saiki et al., unpublished data.
17. B. J. Conner et al., *Proc. Natl. Acad. Sci. U.S.A.* **80**, 278 (1983).
18. V. J. Kidd, R. B. Wallace, K. Itakura, S. L. C. Woo, *Nature (London)* **304**, 230 (1983).
19. R. M. Myers, N. Lumelsky, L. S. Lerman, T. Maniatis, *Nature (London)* **313**, 495 (1985).
20. Calculation of the amounts of p β R328:: β^A needed as standards to estimate PCR efficiency was done in the following way. If we assume that a human haploid genome size is 3×10^9 bp, 36 ng of DNA is equivalent to 1.8×10^{-8} pmol. The extent of amplification after 20 cycles at, for example, 85 percent efficiency is obtained with $(1.8 \times 10^{-8}) (1.85^{20})$ or 4.0×10^{-3} pmol of plasmid DNA (Fig. 2B, lane 5).

20 September 1985; accepted 15 November 1985

The Molecule of the Year

Ruth Levy Guyer and Daniel E. Koshland, Jr.

Science HAS SELECTED THE POLYMERASE CHAIN REACTION AS the major scientific development of 1989 and has chosen for its first "Molecule of the Year" the DNA polymerase molecule that drives the reaction. The list from which the polymerase chain reaction (PCR) was chosen included an impressive array of accomplishments in many areas of science and technology; additional kudos are therefore conferred to 17 of the other big "stories" that made 1989 an exciting year for scientists and for followers and beneficiaries of science. Although the PCR procedure was introduced several years ago, use of the technique truly burgeoned in 1989; in much the same way, the full potentials of many of the interesting "runner-up" scientific achievements of this year are likely to be realized sometime in the years to come.

The first PCR papers were published in 1985. Since that time PCR has grown into an increasingly powerful, versatile, and useful technique. The PCR "explosion" of 1989 can be seen as the result of a combination of improvements in and optimization of the methodology, introduction of new variations on the basic PCR theme, and growing awareness by scientists of what PCR has to offer. With PCR, tiny bits of embedded, often hidden, genetic information can be amplified into large quantities of accessible, identifiable, and analyzable material. A single cell provides enough material for analysis; a single hair can be used to identify an individual.

The basic PCR reaction. The starting material for PCR, the "target sequence," is a gene or segment of DNA. In a matter of hours, this target sequence can be amplified a millionfold. How this is accomplished is shown in the accompanying figure. The complementary strands of a double-stranded molecule of DNA are separated by heating. Two small pieces of synthetic DNA, each complementing a specific sequence at one

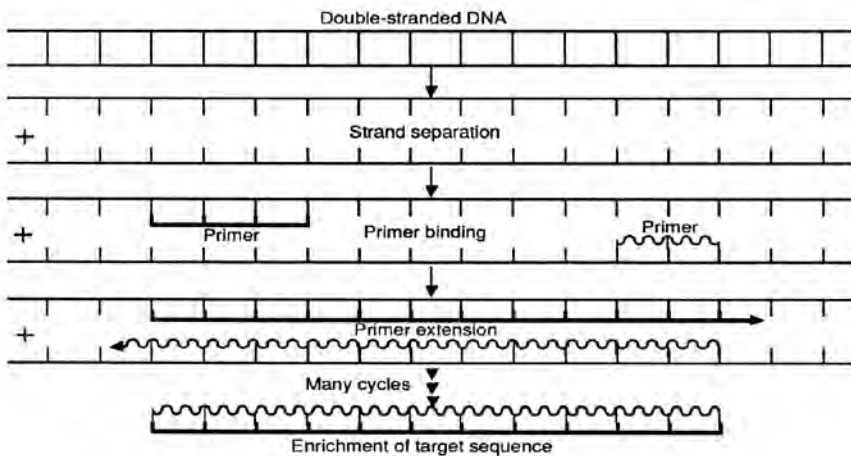
end of the target sequence, serve as primers. Each primer binds to its complementary sequence. Polymerases start at each primer and copy the sequence of that strand. Within a short time, exact replicas of the target sequence have been produced. In subsequent cycles, double-stranded molecules of both the original DNA and the copies are separated; primers bind again to complementary sequences and the polymerase replicates them. At the end of many cycles, the pool is greatly enriched in the small pieces of DNA that have the target sequences, and this amplified genetic information is then available for further analysis.

Evolving PCR. Many improvements on the original PCR method have been made. One of the first was the substitution of a heat-stable enzyme for the original DNA polymerase, which was heat-labile and had to be replenished after each cycle. The stable "Taq polymerase," which comes from bacteria that live in hot springs, continues working almost indefinitely despite the heating steps. Taq polymerase improved the yield, generated more specific and longer products, and facilitated automation.

New strategies have also been devised for flanking unknown sequences with defined primer sites. For standard PCR, the sequences at both ends of a target sequence have to be known. "Inverse" PCR provides a way of sequencing DNA outside the primer sites rather than between two primer sites. Primer molecules are synthesized with their sequences reversed. The target DNA is cut and circularized, and, when the polymerase extends the primer, it does so around the circle in the direction opposite that which would have been taken by standard PCR primers. "Anchored" PCR was developed for studying genes that encode proteins for which partial sequences are known. For anchored PCR, only one defined primer sequence is needed, not two.

The implications of inverse and anchored

QIAGEN Exhibit 1018.0009



PCR for DNA sequencing are astounding: enormous stretches of DNA can be sequenced once a tiny bit of sequence is known. Both techniques make it possible to proceed along the DNA, continually redefining “ends” to which synthetic primers can be bound and then extended.

Applications of PCR. The basic PCR procedure has been valuable in disease diagnosis because specific DNA sequences can be amplified enormously (the needle in the haystack). One of the first uses led to improved diagnosis of a genetic disease (sickle cell anemia), because the PCR technique depended on much less clinical material than standard procedures. (Because PCR is exquisitely sensitive, unusual care is taken to avoid the amplification of contaminants.) PCR can also be used to amplify trace amounts of genetic material of infectious agents in blood, cells, water, food, and other clinical and environmental samples. PCR-based tests are especially valuable for detecting pathogens that are difficult or impossible to culture, such as the agents of Lyme disease and AIDS. For cancer diagnosis and cancer research, PCR can indicate what genes are expressed or turned off, because the messenger RNA molecules associated with such genes can be converted into complementary DNA sequences that then can be amplified.

DNA samples in trace materials (semen,

blood, hairs) found at the scene of a crime have been compared with DNA samples from crime suspects; both acquittals and convictions have resulted from such comparisons. Missing persons have also been positively identified through PCR-based comparisons. The resolution of paternity cases has been aided by comparing DNA from a child with that of the alleged father. And matches of transplant donors and recipients are facilitated with PCR. “Universal” primers are being used to determine the extent of homology in the sequences of conserved genes from different samples. Such comparisons, which help to establish evolutionary relations among organisms, can even include extinct organisms, because DNA samples extracted from mummies, bones, and other archival materials can be used.

PCR may soon replace gene cloning as the amplification method of choice for gene sequencing, for which large amounts of DNA are needed. PCR is also providing new options in molecular genetics studies for adding genetic information to target materials or for altering what is already there.

The rate at which new PCR-based techniques have been developed suggests that this technology is proliferating as rapidly as its Taq polymerase molecules replicate target sequences.

A Technique Whose Time Has Come

Nigel J. Walker

Developed in the mid 1990s for the analysis and quantification of nucleic acids, real-time PCR is a molecular biological technique gaining rapidly in popularity. It is based on the technique of the polymerase chain reaction (PCR) that was first envisioned by Kary Mullis almost 20 years ago, during a moonlit drive through the redwood hills of California (1). The technology of PCR (2) has become one of the most influential discoveries of the molecular biology revolution and one for which Mullis received the Nobel Prize in 1993. Because of the impact of PCR and the thermostable *Taq* DNA polymerase (the enzyme responsible for the PCR revolution), the pair was named as the first "Molecule of the Year" by *Science* in 1989 (3). In many ways, the recent development of real-time PCR seems set to change the general use of PCR.

The advancement provided by the real-time version of PCR is due to its unique ability to monitor the complete DNA amplification process. During conventional PCR, the two strands of a DNA molecule are subjected to a series of heating and cooling cycles that result in DNA strand separation, oligonucleotide primer annealing, and thermostable *Taq* DNA polymerase-directed primer extension, ultimately generating two identical daughter strands. Iterative cycling of the process exponentially amplifies the number of original DNA molecules, hence the term PCR (4). After completion of the PCR reaction, amplification products are analyzed by size-fractionation of the amplified sample with the use of gel electrophoresis.

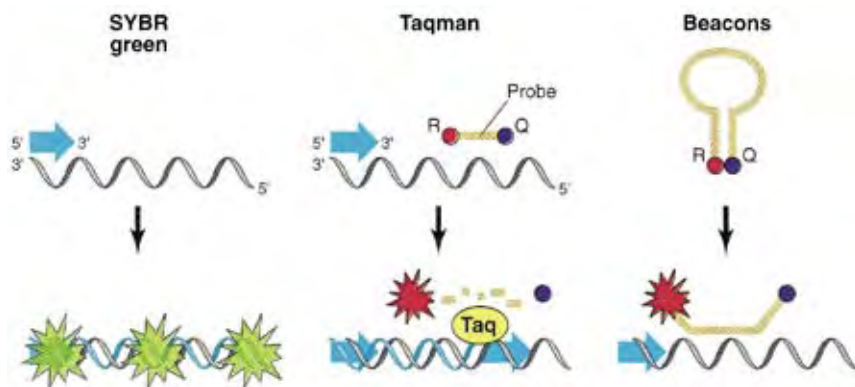
In the mid 1990s, researchers showed that the 5' nuclease activity of the *Taq* DNA

polymerase could be exploited as a method to indirectly assess the level of DNA amplification with the use of specific fluorescent probes (5), eliminating the need for electrophoresis. Around the same time, researchers showed that real-time monitoring of the DNA amplification within the PCR reaction tube during the PCR could be achieved by using fluorescent DNA binding dyes, which is known as kinetic PCR (6). The coupling of these two processes (7, 8) led to today's technology of fluorescence detection real-time PCR.

In general, analysis of amplification during real-time PCR has been achieved by detecting the fluorescence that is either directly or indirectly associated with the accumulation of the newly amplified DNA (see figure, page 15). The detection system that is almost synonymous with real-time PCR is the "Taqman" system (8), which uses a fluorescence resonance energy transfer (FRET) probe as a reporter system. A FRET probe is a short oligonucleotide that is complementary to one of the strands. The probe contains a "reporter" and a "quencher" fluorescent molecule at the 5' and 3' end of the probe, respectively. This probe is included in the real-time PCR reaction along with the required forward and reverse PCR primers. The quencher fluorochrome on the probe, because it is in such close proximity to the reporter, is able to quench the fluorescence of the reporter. As the *Taq* DNA polymerase enzyme replicates the new strand of DNA, the nuclease activity degrades the FRET probe at the 5' end, which is bound to template DNA strand, in a manner much like the PacMan video game character. This degradation releases the reporter fluorochrome from its proximity to the quencher, resulting in fluorescence of the reporter. Accumulation of fluorescent reporter molecules, as a result of amplification of the

The author is in the Environmental Toxicology Program, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709, USA. E-mail: walker3@niehs.nih.gov

QIAGEN Exhibit 1018.0011

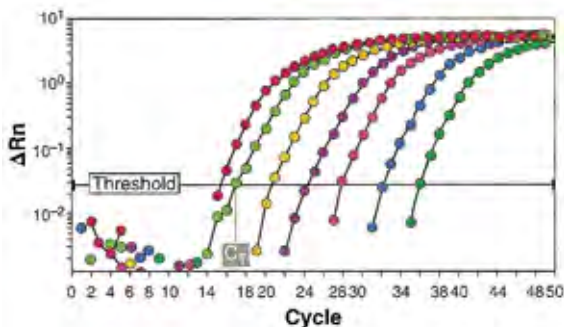


Detection systems. Fluorescent detection of amplification can be achieved using double-stranded DNA binding dyes like SYBR green (green starbursts) or with FRET-based probes such as Taqman 5' nuclease-sensitive probes or DNA binding probes (red starbursts).

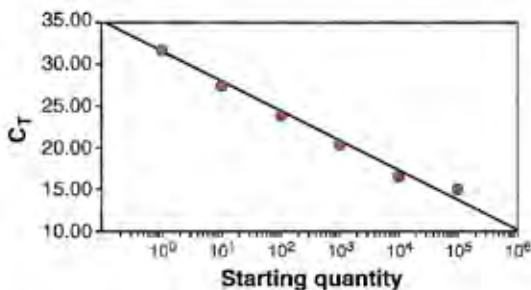
target, can then be detected by an appropriate optical sensing system such as the Taqman, an “indirect” system that detects the accumulation of fluorescence rather than the amplified DNA itself. In contrast, a commonly used “direct” method uses a fluorescent DNA (SYBR green) that binds nonspecifically to double-stranded DNA, and the accumulation of the fluorescence bound to the amplified DNA target is measured. The “Molecular Beacon” technology is another direct approach that uses FRET-based fluorescent probes to bind the amplified DNA. In the unbound state, the quencher and reporter fluorochromes are maintained in close proximity via a hairpin loop designed into the sequence of the probe. Binding of the probe at a target sequence-specific region to its complementary strand on the amplified target DNA separates the two fluorochromes, thereby alleviating the FRET interference and allowing the reporter to fluoresce. Related systems that use FRET-based PCR primers incorporated into the amplified DNA have also been developed (9). In these systems, the reporter and quencher fluorochromes are maintained in a hairpin loop structure via a sequence that is added to the 5' end of one of the PCR primers. Disruption of the hairpin loop structure during incorporation of the primer into the amplified DNA product results in loss of the FRET interference, leading to fluores-

cence of the reporter molecule. Choosing a detection system is a major consideration in developing a real-time PCR assay, and each type of system has its pros and cons. The decision is often a compromise between desired specificity, assay development time, and cost per assay.

An inherent property of PCR that is exploited in real-time PCR is that the more copies of nucleic acid one starts with, the fewer cycles of template amplification it takes to make a specific number of products. Therefore, the number of cycles needed for the amplification-associated fluorescence to reach a specific threshold level of detection (the C_T value) is inversely correlated to the amount of nucleic acid that was in the original sample (see figure, top page 16). Because the progress of amplification is monitored throughout the PCR process in real-time PCR, a C_T value can be determined during the exponential phase of a PCR reaction, when amplification is most efficient and least affected by reaction-limiting conditions. The quantity of DNA in the sample can then be obtained by interpolation of its C_T value versus a linear standard curve of C_T values obtained from a serially diluted standard solution (see figure, middle page 16). In practice, such curves are linear over more than five orders of magnitude, a dynamic range that is unsurpassed by other methods



Amplification time for fluorescence detection. Curves representing the cycle-dependent fluorescence associated with amplification of a specific gene product from 0.1 to 100 ng total RNA are shown. The C_T value for each sample is determined from each curve as the cycle at which the fluorescence achieves a specific threshold value. ΔR_n , normalized fluorescence.



Dynamic range of real-time PCR. Fluorescent detection of the amplification of the gene product is linear over five orders of magnitude. Values were calculated on the basis of C_T values from amplification figure. Slope, -3.538 ; y intercept, 31.806 ; correlation coefficient, 0.991

for quantitative DNA analysis. It is also possible to quantify differences in nucleic acid levels without such curves. For example, during the exponential phase of PCR, the number of DNA strands theoretically doubles during each cycle (assuming amplification is 100% efficient). Consequently, a sample that has twice the number of starting copies compared with another sample would require one less cycle of amplification to generate an equivalent number of product strands. By using the difference in the C_T values for two samples, therefore, one can mathematically determine the relative difference in the level of the nucle-

ic acid of interest of different samples (10).

Instrumentation systems for real-time PCR have undergone extensive changes already. Initially they were large machines that took up almost an entire lab bench and were incapable of a true analysis in real time. As the technology improved, systems became available that allow the PCR reaction to be monitored as it occurs. The size and cost of these systems has been reduced so dramatically that they are now reasonably priced units with very small footprints. Given the added capabilities of the real-time technology, it is likely that in the near future such systems will become the standard PCR platform in the general laboratory, in much the same way that the 96- and 384-microwell PCR machines have superseded initial large-tube format units.

In the laboratory, real-time reverse transcriptase-PCR (RT-PCR) has become the method of choice for the rapid and quantitative examination of the expression of specific genes. Users of the method can rapidly, reproducibly, and statistically determine even small (twofold) changes in the

expression of hundreds of samples per day. Such analyses would be very difficult and laborious with the use of traditional hybridization techniques, such as Northern blotting or RNase protection assays. The low RNA quantities required (nanograms) make this assay more suitable for the analysis of samples obtained from laser microdissected tissue. In addition, realtime approaches are perfectly adaptable for high-throughput and quantitative gene expression studies, especially with the recent development of high-throughput 384-well real-time PCR instruments. These advances are timely, given the

completion of the Human Genome Project. Recent advances in microtechnologies, such as DNA chip technology, coupled with proteomics and metabonomics (11) are establishing the new science of systems biology (12). This division of biology seeks to use information from the integrated analysis of genes or proteins and metabolism to develop computational models of cellular function and physiology. Such models are enhanced through the availability of quantitative information on the expression of genes, proteins, or metabolites (13). Quantitative real-time PCR analysis of gene expression is likely to play a key role in this burgeoning field.

Though PCR has influenced drastically the way in which molecular biological research is conducted, its impact on our everyday lives has yet to be fully realized. Because of its reduced detection times and simplification of quantitation, real-time PCR systems are likely to have the greatest impact on the general public in environmental monitoring and nucleic acid diagnostics. Most uses of PCR-based diagnostics still require specialized laboratory services, but several companies are developing devices that will allow such rapid PCR analysis to be performed in the near future at the point the sample is collected instead of in a far-removed laboratory. Devices coupling microfluid technology with realtime PCR analysis of nucleic acid detection are set to make a huge impact in many facets of our lives as we enter the genomic information era (14). Such devices would incorporate nucleic acid extraction and PCR detection in small, disposable sys-

tems that are readable through attachments to a personal computer. This technology could potentially enable inhome testing for nucleic acids from bacterial or viral pathogens. As methods for the quantitative analysis of gene expression and DNA levels continue to evolve in sophistication, devices that incorporate the concepts of real-time PCR will likely herald the era of individualized genomics and genetic testing.

References

1. K. B. Mullis, *Sci. Am.* **262**, 56 (1990).
2. R. K. Saiki *et al.*, *Science* **230**, 1350 (1985); R. K. Saiki *et al.*, *Nature* **324**, 163 (1986).
3. R. L. Guyer, D. E. Koshland Jr., *Science* **246**, 1543 (1989).
4. B. A. White, Ed., *PCR Protocols: Current Methods and Applications* (Humana Press, Totowa, NJ, 1993), vol. 15.
5. P. M. Holland, R. D. Abramson, R. Watson, D. H. Gelfand, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7276 (1991).
6. R. Higuchi, C. Fockler, G. Dollinger, R. Watson, *Biotechnology* **11**, 1026 (1993); R. Higuchi, G. Dollinger, P. S. Walsh, R. Griffith, *Biotechnology* **10**, 413 (1992).
7. U. E. Gibson, C. A. Heid, P. M. Williams, *Genome Res.* **6**, 995 (1996).
8. C. A. Heid, J. Stevens, K. J. Livak, P. M. Williams, *Genome Res.* **6**, 986 (1996).
9. N. J. Walker, *J. Biochem. Mol. Toxicol.* **15**, 121 (2001).
10. M. W. Pfaffl, *Nucleic Acids Res.* **29**, E45 (2001).
11. E. K. Lobenhofer, P. R. Bushel, C. A. Afshari, H. K. Hamadeh, *Environ. Health Perspect.* **109**, 881 (2001); J. K. Nicholson, J. C. Lindon, E. Holmes, *Xenobiotica* **29**, 1181 (1999).
12. W. W. Gibbs, *Sci. Am.* **285**, 52 (2001); T. Ideker, T. Galitski, L. Hood, *Annu. rev. Genomics Hum. Genet.* **2**, 343 (2001).
13. T. Ideker *et al.*, *Science* **292**, 929 (2001).
14. P. Mitchell, *Nature Biotechnol.* **19**, 717 (2001).

Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome

Jay Shendure,^{1††} Gregory J. Porreca,^{1††} Nikos B. Reppas,¹ Xiaoxia Lin,¹ John P. McCutcheon,^{2,3} Abraham M. Rosenbaum,¹ Michael D. Wang,¹ Kun Zhang,¹ Robi D. Mitra,² George M. Church¹

We describe a DNA sequencing technology in which a commonly available, inexpensive epifluorescence microscope is converted to rapid nonelectrophoretic DNA sequencing automation. We apply this technology to resequence an evolved strain of *Escherichia coli* at less than one error per million consensus bases. A cell-free, mate-paired library provided single DNA molecules that were amplified in parallel to 1-micrometer beads by emulsion polymerase chain reaction. Millions of beads were immobilized in a polyacrylamide gel and subjected to automated cycles of sequencing by ligation and four-color imaging. Cost per base was roughly one-ninth as much as that of conventional sequencing. Our protocols were implemented with off-the-shelf instrumentation and reagents.

The ubiquity and longevity of Sanger sequencing (1) are remarkable. Analogous to semiconductors, measures of cost and production have followed exponential trends (2). High-throughput centers generate data at a speed of 20 raw bases per instrument-second and a cost of \$1.00 per raw kilobase. Nonetheless, optimizations of electrophoretic methods may be reaching their limits. Meeting the challenge of the \$1000 human genome requires a paradigm shift in our underlying approach to the DNA polymer (3).

Cyclic array methods, an attractive class of alternative technologies, are “multiplex” in that they leverage a single reagent volume to enzymatically manipulate thousands to millions of immobilized DNA features in parallel. Reads are built up over successive cycles of imaging-based data acquisition. Beyond this common thread, these technologies diversify in a panoply of ways: single-molecule versus multimolecule features, ordered versus disordered arrays, sequencing biochemistry, scale of miniaturization, etc.

(3). Innovative proof-of-concept experiments have been reported, but are generally limited in terms of throughput, feature density, and library complexity (4–9). A range of practical and technical hurdles separate these test systems from competing with conventional sequencing on genomic-scale applications.

Our approach to developing a more mature alternative was guided by several considerations. (i) An integrated sequencing pipeline includes library construction, template amplification, and DNA sequencing. We therefore sought compatible protocols that multiplexed each step to an equivalent order of magnitude. (ii) As more genomes are sequenced de novo, demand will likely shift toward genomic resequencing; e.g., to look at variation between individuals. For resequencing, consensus accuracy increases in importance relative to read length because a read need only be long enough to correctly position it on a reference genome. However, a consensus accuracy of 99.99%, i.e., the Bermuda standard, would still result in hundreds of errors in a microbial genome and hundreds of thousands of errors in a mammalian genome. To avoid unacceptable numbers of false-positives, a consensus error rate of 1×10^{-6} is a more reasonable standard for which to aim. (iii) We sought to develop sequencing chemistries compatible with conventional epifluorescence imaging. Diffraction

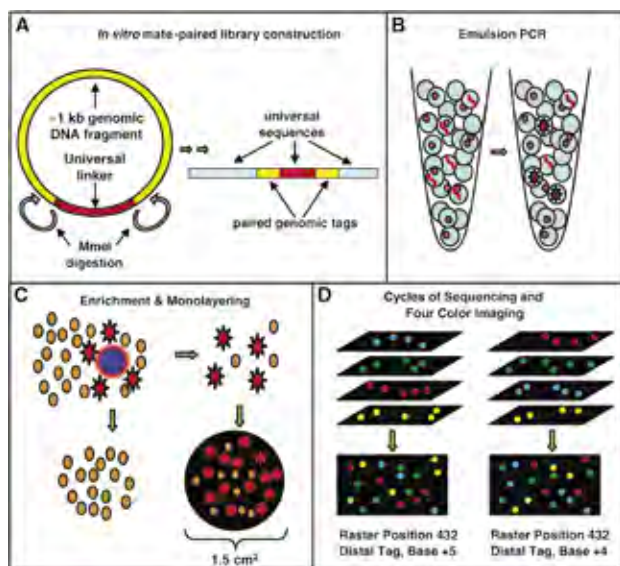
¹Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. ²Department of Genetics,

³Howard Hughes Medical Institute, Washington University, St. Louis, MO 63110, USA.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: shendure@alumni.princeton.edu (J.S.), gregory_porreca@student.hms.harvard.edu (G.J.P.)

Fig. 1. A multiplex approach to genome sequencing. (A) Sheared, size-selected genomic fragments (yellow) are circularized with a linker (red) bearing Mme I recognition sites (Note S1). Subsequent steps, which include a rolling circle amplification, yield the 134- to 136-bp mate-paired library molecules shown at right. (B) ePCR (14) yields clonal template amplification on 1- μ m beads (Note S2). (C) Hybridization to non-magnetic, low-density "capture beads" (dark blue) permits enrichment of the amplified fraction (red) of magnetic ePCR beads by centrifugation (Note S3). Beads are immobilized and mounted in a flowcell for automated sequencing (Note S4). (D) At each sequencing cycle, four-color imaging is performed across several hundred raster positions to determine the sequence of each amplified bead at a specific position in one of the tags. The structure of each sequencing cycle is discussed in the text, Note S6, and fig. S7.



tion-limited optics with charge-coupled device detection achieves an excellent balance because it not only provides submicrometer resolution and high sensitivity for rapid data acquisition, but is also inexpensive and easily implemented.

Conventional shotgun libraries are constructed by cloning fragmented genomic DNA of a defined size range into an *Escherichia coli* vector. Sequencing reads derived from opposite ends of each fragment are termed "mate-pairs." To avoid bottlenecks imposed by *E. coli* transformation, we developed a multiplexed, cell-free library construction protocol. Our strategy (Fig. 1A) uses a type II restriction endonuclease to bring sequences separated on the genome by ~1 kb into proximity. Each ~135-base pair (bp) library molecule contains two mate-paired 17- to 18-bp tags of unique genomic sequence, flanked and separated by universal sequences that are complementary to amplification or sequencing primers used in subsequent steps. The in vitro protocol (Note S1) results in a library with a complexity of ~1 million

unique, mate-paired species.

Conventionally, template amplification has been performed by bacterial colonies that must be individually picked. Polymerase colony, or polony, technologies perform multiplex amplification while maintaining spatial clustering of identical amplicons (10). These include in situ polonies (11), in situ rolling circle amplification (RCA) (12), bridge polymerase chain reaction (PCR) (13), picotiter PCR (9), and emulsion PCR (14). In emulsion PCR (ePCR), a water-in-oil emulsion permits millions of non-interacting amplifications within a milliliter-scale volume (15–17). Amplification products of individual compartments are captured via inclusion of 1- μ m paramagnetic beads bearing one of the PCR primers (14). Any single bead bears thousands of single-stranded copies of the same PCR product, whereas different beads bear the products of different compartmentalized PCR reactions (Fig. 1B). The beads generated by ePCR have highly desirable characteristics: high signal density, geometric uniformity, strong feature separation, and a size that is small but still resolvable by

inexpensive optics.

Provided that the template molecules are sufficiently short (fig. S1), an optimized version of the ePCR protocol described by Dressman *et al.* (14) robustly and reproducibly amplifies our complex libraries (Note S2). In practice, ePCR yields empty, clonal, and nonclonal beads, which arise from emulsion compartments that initially have zero, one, or multiple template molecules, respectively. Increasing template concentration in an ePCR reaction boosts the fraction of amplified beads at the cost of greater nonclonality (14). To generate populations in which a high fraction of beads was both amplified and clonal, we developed a hybridization-based in vitro enrichment method (Fig. 1C). The protocol is capable of a fivefold enrichment of amplified beads (Note S3).

Iterative interrogation of ePCR beads (Fig. 1D) requires immobilization in a format compatible with enzymatic manipulation and epifluorescence imaging. We found that a simple acrylamide-based gel system developed for in situ colonies (6) was easily applied to ePCR beads, resulting in a $\sim 1.5\text{-cm}^2$ array of disordered, monolayered, immobilized beads (Note S4, Fig. 2A).

With few exceptions (18), sequencing biochemistries rely on the discriminatory capacities of polymerases and ligases (1, 6, 8, 19–22). We evaluated a variety of sequencing protocols in our system. A four-color sequencing by ligation scheme (“degenerate ligation”) yielded the most promising results (Fig. 2, B and C). A detailed graphical description of this method is shown in fig. S7. We begin by hybridizing an “anchor primer” to one of four positions (immediately 5′ or 3′ to one of the two tags). We then perform an enzymatic ligation reaction of the anchor primer to a population of degenerate nonamers that are labeled with fluorescent dyes. At any given cycle, the population of nonamers that is used is structured such that the identity of one of its positions is correlated with the identity of the fluorophore attached to that nonamer. To the extent that the ligase discriminates for complementarity at that queried position, the fluorescent signal allows us to infer the identity of that

base (Fig. 2, B and C). After performing the ligation and four-color imaging, the anchor primer:nonamer complexes are stripped and a new cycle is begun. With T4 DNA ligase, we can obtain accurate sequence when the query position is as far as six bases from the ligation junction while ligating in the 5′→3′ direction, and seven bases from the ligation junction in the 3′→5′ direction. This allows us to access 13 bp per tag (a hexamer and heptamer separated by a 4- to 5-bp gap) and 26 bp per amplicon (2 tags $\times$ 13 bp) (fig. S7).

Although the sequencing method presented here can be performed manually, we benefited from fully automating the procedure (fig. S3). Our integrated liquid-handling and microscopy setup can be replicated with off-the-shelf components at a cost of about \$140,000. A detailed description of instrumentation and software is provided in Notes S5 and S7.

As a genomic-scale challenge, we sought a microbial genome that was expected, relative to a reference sequence, to contain a modest number of both expected and unexpected differences. We selected a derivative of *E. coli* MG1655, engineered for deficiencies in tryptophan biosynthesis and evolved for ~ 200 generations under conditions of syntrophic symbiosis via coculture with a tyrosine biosynthesis-deficient strain (23). Specific phenotypes emerged during the laboratory evolution, leading to the expectation of genetic changes in addition to intentionally engineered differences.

An in vitro mate-paired library was constructed from genomic DNA derived from a single clone of the evolved Trp^- strain. To sequence this library, we performed successive instrument runs with progressively higher bead densities. In an experiment ultimately yielding 30.1 Mb of sequence, 26 cycles of sequencing were performed on an array containing amplified, enriched ePCR beads. At each cycle, data were acquired for four wavelengths at 20 $\times$ optical magnification by rastering across each of 516 fields of view on the array (Fig. 1D). A detailed description of the structure of each sequencing cycle is provided in Note S6. In total, 54,696 images (14 bit, 1000 $\times$ 1000) were collected. Cycle

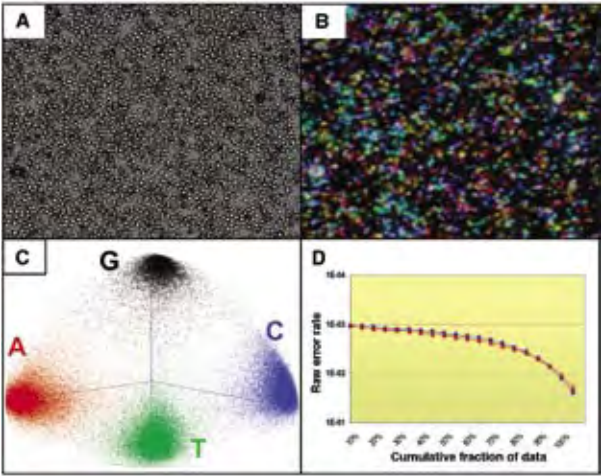


Fig. 2. Raw data acquisition and base calling. (A) Brightfield images (area shown corresponds to 0.01% of the total gel area) facilitate object segmentation by simple thresholding, allowing resolution even when multiple 1- μ m beads are in contact. (B) False-color depiction of four fluorescence images acquired at this location from a single ligation cycle. A, gold; G, red; C, light blue; T, purple. (C) Four-color data from each cycle can be visualized in tetrahedral space, where each point represents a single bead, and

the four clusters correspond to the four possible base calls. Shown is the sequencing data from position (–1) of the proximal tag of a complex *E. coli*–derived library. (D) Cumulative distribution of raw error as a function of rank-ordered quality for two independent experiments (red triangles, 18.1-Mb run; blue squares, 30.1-Mb run). The x axis indicates percentile bins of beads, sorted on the basis of a confidence metric. The y axis (logarithmic scale) indicates the raw base-calling accuracy of each cumulative bin. Equivalent Phred scores are Q20 = 1×10^{-2} ; Q30 = 1×10^{-3} [Phred score = $-10[\log_{10}(\text{raw per-base error})]$]. Cumulative distribution of raw error with sequencing by ligation cycles considered independently is shown in fig. S8.

Table 1. Genome Coverage and SNC prediction. Bases with consistent consensus coverage were used to make mutation predictions. To assess power, the outcome of consensus calling for the mock SNC positions with various levels of coverage was determined. Data from two independent sets of mock SNCs are shown. “86 of 87,” for example, means that 87 of the 100 mock SNCs were present in the sequence that was covered with $1\times$ or more reads, and 86 of these were called correctly.

Coverage	Percent of genome	Correctly called mock substitutions
1 $\times$ or greater	91.4%	86 of 87 88 of 90
2 $\times$ or greater	83.3%	78 of 78 75 of 76
3 $\times$ or greater	74.9%	67 of 67 68 of 68
4 $\times$ or greater	66.9%	58 of 58 62 of 62

times averaged 135 min per base (~90 min for reactions and ~45 min for imaging), for a total of ~60 hours per instrument run.

Image processing and base calling algorithms are detailed in Note S7. In brief, all images taken at a given raster position were aligned. Two additional image sets were acquired: brightfield images to robustly identify bead locations (Fig. 2A) and fluorescent primer images to identify amplified beads. Our algorithms detected 14 million

objects within the set of brightfield images. On the basis of size, fluorescence, and overall signal coherence over the course of the sequencing run, we determined 1.6 million to be well-amplified, clonal beads (~11%). For each cycle, mean intensities for amplified beads were extracted and normalized to a 4D unit vector (Fig. 2, B and C). The Euclidean distance of the unit vector for a given raw base call to the median centroid of the nearest cluster serves as a natural metric of the

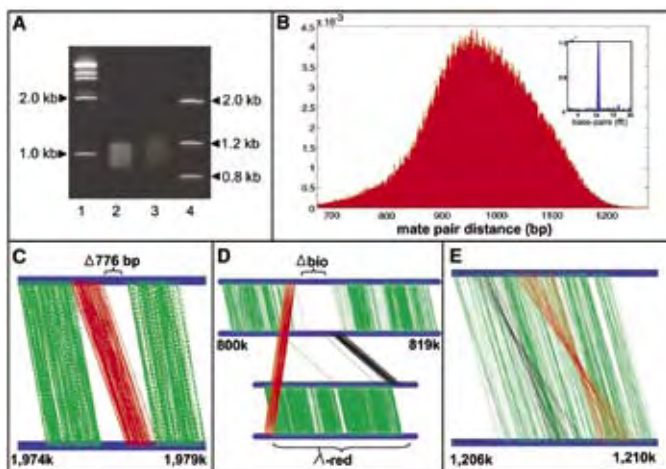


Fig. 3. Mate-paired tags and rearrangement discovery. (A) Diagnostic 6% polyacrylamide gel of the sheared, size-selected genomic DNA from which the library was constructed. Lanes 1 and 4 are molecular size markers. Lane 2 represents the material used in the library sequenced to generate the paired-tag mappings in (B), and lane 3 represents genomic DNA for a different library. (B) Histogram of distances between ~1 million mapped mate-pair sequences. The probability of circularization favors integrals of the helical pitch of DNA, such that the Fourier transform of the distribution (inset) yields a peak at 10.6 bp (27). (C to E). Consistent, aberrant mapping of unplaced mate-pairs to distal sequences revealed information about underlying rearrangements. Top and bottom blue bars indicate genomic positions for proximal and distal tags, respectively. Green connections indicate mate-pairings that fall within expected distance constraints, whereas red and black connections indicate aberrant connections (red indicates connections between the same strand, and black, connections between opposite strands). (C) Detection of a 776-bp deletion in the *flhD* promoter (24). (D) Detection of the replacement of the *bio* locus with the lambda red construct. (E) Detection of the P-region inversion (25). Detection of the inversion on a background of normally mate-paired reads indicates that the inversion is heterogeneously present.

quality of that call.

The reference genome consisted of the *E. coli* MG1655 genome (GenBank accession code U00096.2) appended with sequences corresponding to the *cat* gene and the lambda Red prophage, which had been engineered into the sequenced strain to replace the *trp* and *bio* operons, respectively. To systematically assess our power to detect single-base substitutions, we introduced a set of 100 random single-nucleotide changes into the reference sequence at randomly selected positions (“mock SNCs”) (Table 1).

An algorithm was developed to place the discontinuous reads onto the reference sequence (Note S7). The matching criteria required the paired tags to be appropriately oriented and located within 700 to 1200 bp of one another, allowing for substitutions if exact matches were not found. Of the 1.6 million reads, we were able to confidently place ~1.16 million (~72%) to specific loca-

tions on the reference genome, resulting in ~30.1 million bases of resequencing data at a median raw accuracy of 99.7%. At this stage of the analysis, the data were combined with reads from a previous instrument run that contributed an additional ~18.1 million bases of equivalent quality (Fig. 2D). In this latter experiment, ~1.8 million reads were generated from ~7.6 million objects (~24%), of which ~0.8 million were confidently placed (~40%).

High-confidence consensus calls were determined for 70.5% of the *E. coli* genome for which sufficient and consistent coverage was available (3,289,465 bp; generally positions with ~4× or greater coverage). There were six positions within this set that did not agree with the reference sequence, and thus were targeted for confirmation by Sanger sequencing. All six were correct, although in one case we detected the edge of an 8-bp deletion rather than a substitution (Table 2).

Table 2. Polymorphism discovery. Predictions for mutated positions were tested and verified as correct by Sanger sequencing. We found three mutations unique to the evolved strain—two in *ompF*, a porin, and one in *lrp*, a global regulator.

Position	Type	Gene	Context	Confirmation	Comments
986,328	T → G	<i>ompF</i>	–10 region	Yes	Evolved strain only
931,955	8-bp deletion	<i>lrp</i>	Frameshift	Yes	Evolved strain only
985,791	T → G	<i>ompF</i>	Glu → Ala	Yes	Evolved strain only
1,976,527–1,977,302	776-bp deletion	<i>flhD</i>	Promoter	Yes	MG1655 heterogeneity
3,957,960	C → T	<i>ppiC</i>	5' UTR	Yes	MG1655 heterogeneity
λ-red, 3274	T → C	<i>ORF61</i>	Lys → Gly	Yes	λ-red heterogeneity
λ-red, 9846	T → C	<i>cl</i>	Glu → Glu	Yes	λ-red heterogeneity

Three of these six mutations represent heterogeneities in lambda Red or MG1655, or errors in the reference sequence; three were only present in the evolved variant (Table 2). Of the 100 mock SNCs, 53 were at positions called with high confidence. All of these were correctly called as substitutions of the expected nucleotide (59 of 59 on a second set of mock SNCs). The absence of substitution errors in ~3.3 Mb of reference sequence positions called with high confidence suggests that we are achieving consensus accuracies sufficient for resequencing applications. Percentage of the genome covered and mock SNC discovery at various levels of coverage are shown in Table 1.

Despite 10× overage in terms of raw base pairs, only ~91.4% of the genome had at least 1× coverage (fig. S4). Substantial fluctuations in coverage were observed owing to the stochasticity of the RCA step of library construction. We are currently generating libraries that are more complex and more evenly distributed.

A Gaussoid distribution of distances between mate-paired tags was observed, consistent with the size selection during library construction (Fig. 3, A and B). Notably, the helical pitch of DNA (~10.6 bp per turn) is evident in the local statistics of ~1 million circularization events (Fig. 3B). As a function of the number of bases sequenced, we generated over an order of magnitude more mate-pairing data points than an equivalent amount of conventional sequencing. To detect genomic rearrangements, we mined the unplaced mate-pairs for consistent links between genomic regions that did not fall within the expected distance constraints. In

addition to detecting the expected replacements of the *trp* and *bio* operons with *cat* and lambda Red prophage (Fig. 3D), we detected and confirmed the absence of a 776-bp IS1 transposon (Fig. 3C), a previously described heterogeneity in MG1655 strains (24). We also detected and confirmed a ~1.8-kb region that was heterogeneously inverted in the genomic DNA used to construct the library (Fig. 3E), owing to activity of *pin* on the invertible P region (25).

We observe error rates of ~0.001 for the better half of our raw base calls (Fig. 2D). Although high consensus accuracies are still achieved with relatively low coverage, our best raw accuracies are notably one to two orders of magnitude less accurate than most raw bases in a conventional Sanger sequencing trace. The PCR amplifications before sequencing are potentially introducing errors at a rate that imposes a ceiling on the accuracies achievable by the sequencing method itself. One potential solution is to create a library directly from the genomic material to be sequenced, such that the library molecules are linear RCA amplicons. Such concatemers, where each copy is independently derived from the original template, would theoretically provide a form of error correction during ePCR.

Our algorithms were focused on detection of point substitutions and rearrangements. Increasing read lengths, currently totaling only 26 bp per amplicon, will be critical to detecting a wider spectrum of mutation. A higher fidelity ligase (20) or sequential nonamer ligations (20, 21) may enable completion of each 17- to 18-bp tag. Eco P15 I, which generates ~27-bp tags, would allow even

longer read lengths while retaining the same mate-pairing scheme (26).

We estimate a cost of \$0.11 per raw kilobase of sequence generated (Note S8), roughly one-ninth as much as the best costs for electrophoretic sequencing. Raw data in all sequencing methods are generally combined to form a consensus. Even though costs are generally defined in terms of raw bases, the critical metric to compare technologies is consensus accuracy for a given cost. There is thus a need to devise appropriate cost metrics for specific levels of consensus accuracy.

If library construction costs are not included, the estimated cost drops to \$0.08 per raw kilobase. Higher densities of amplified beads are expected to boost the number of bases sequenced per experiment. While imaging, data were collected at a rate of ~400 bp/s. Although enzymatic steps slowed our overall throughput to ~140 bp/s, a dual flow-cell instrument (such that the microscope is always imaging) will allow us to achieve continuous data acquisition. Enzymatic reagents, which dominate our cost equation, can be produced in-house at a fraction of the commercial price.

References and Notes

1. F. Sanger *et al.*, *Nature* **265**, 687 (1977).
2. F. S. Collins, M. Morgan, A. Patrino, *Science* **300**, 286 (2003).
3. J. Shendure *et al.*, *Nat. Rev. Genet.* **5**, 335 (2004).
4. I. Braslavsky, B. Hebert, E. Kartalov, S. R. Quake, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 3960 (2003).
5. T. S. Seo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 5926 (2005).
6. R. D. Mitra, J. Shendure, J. Olejnik, O. Edyta Krzymanska, G. M. Church, *Anal. Biochem.* **320**, 55 (2003).
7. M. J. Levene *et al.*, *Science* **299**, 682 (2003).
8. M. Ronaghi, S. Karamohamed, B. Pettersson, M. Uhlen, P. Nyren, *Anal. Biochem.* **242**, 84 (1996).
9. J. H. Leamon *et al.*, *Electrophoresis* **24**, 3769 (2003).
10. <http://arep.med.harvard.edu/Polonator/Plone.htm>
11. R. D. Mitra, G. M. Church, *Nucleic Acids Res.* **27**, e34 (1999).
12. P. M. Lizardi *et al.*, *Nat. Genet.* **19**, 225 (1998).
13. C. P. Adams, S. J. Kron, U.S. Patent 5,641,658 (1997).
14. D. Dressman, H. Yan, G. Traverso, K. W. Kinzler, B. Vogelstein, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8817 (2003).
15. D. S. Tawfik, A. D. Griffiths, *Nat. Biotechnol.* **16**, 652 (1998).
16. F. J. Ghadessy, J. L. Ong, P. Holliger, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4552 (2001).
17. M. Nakano *et al.*, *J. Biotechnol.* **102**, 117 (2003).
18. A. M. Maxam, W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977).
19. F. Barany, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 189 (1991).
20. J. N. Housby, E. M. Southern, *Nucleic Acids Res.* **26**, 4259 (1998).

We demonstrate low costs of sequencing, mate-paired reads, high multiplicities, and high consensus accuracies. These enable applications including BAC (bacterial artificial chromosome) and bacterial genome resequencing, as well as SAGE (serial analysis of gene expression) tag and barcode sequencing. Simulations suggest that the current mate-paired libraries are compatible with human genome resequencing, provided that the read length can be increased to cover the full 17- to 18-bp tag (fig. S5).

What are the limits of this approach? As many as 1 billion 1- μ m beads can potentially be fit in the area of a standard microscope slide (fig. S6). We achieve raw data acquisition rates of ~400 bp/s, more than an order of magnitude faster than conventional sequencing. From another point of view, we collected ~786 gigabits of image data from which we gleaned only ~60 megabits of sequence. This sparsity—one useful bit of information per 10,000 bits collected—is a ripe avenue for improvement. The natural limit of this direction is single-pixel sequencing, in which the commonplace analogy between bytes and bases will be at its most manifest.

21. S. C. Macevicz, U.S. Patent 5,750,341 (1998).
22. S. Brenner *et al.*, *Nat. Biotechnol.* **18**, 630 (2000).
23. N. B. Reppas, X. Lin, in preparation.
24. C. S. Barker, B. M. Pruss, P. Matsumura, *J. Bacteriol.* **186**, 7529 (2004).
25. R. H. Plasterk, P. van de Putte, *EMBO J.* **4**, 237 (1985).
26. M. Mucke, S. Reich, E. Moncke-Buchner, M. Reuter, D. H. Kruger, *J. Mol. Biol.* **312**, 687 (2001).
27. D. Shore, R. L. Baldwin, *J. Mol. Biol.* **170**, 957 (1983).
28. For advice, encouragement, and technical assistance, we are deeply indebted to J. Zhu, S. Douglas, J. Chou, J. Aach, M. Nikku, A. Lee, N. Novikov, and M. Wright (Church Lab); A. Blanchard, G. Costa, H. Ebling, J. Ichikawa, J. Malek, P. McEwan, K. McKernan, A. Sheridan, and D. Smith (Agencourt); S. Skiena (SUNY-Stony Brook); C. Felts (RPI); R. Fincher (Alcott); D. Focht (Bioprotech); and M. Hotfelder and J. Feng (Washington University). We thank B. Vogelstein, J. Edwards, and their groups for assistance with emulsion PCR. This work was supported by the National Human Genome Research Institute—Centers of Excellence in Genomic Science and U.S. Department of Energy—Genomes to Life grants.

Supporting Online Material

www.science.org/cgi/content/full/1117389/DC1
SOM Text
Figs. S1 to S8

14 July 2005; accepted 27 July 2005

Published online 4 August 2005;

10.1126/science.1117389

Include this information when citing this paper.

QIAGEN Exhibit 1018.0021

Microfluidic Digital PCR Enables Multigene Analysis of Individual Environmental Bacteria

Elizabeth A. Ottesen,¹ Jong Wook Hong,² Stephen R. Quake,³ Jared R. Leadbetter^{4*}

Gene inventory and metagenomic techniques have allowed rapid exploration of bacterial diversity and the potential physiologies present within microbial communities. However, it remains nontrivial to discover the identities of environmental bacteria carrying two or more genes of interest. We have used microfluidic digital polymerase chain reaction (PCR) to amplify and analyze multiple, different genes obtained from single bacterial cells harvested from nature. A gene encoding a key enzyme involved in the mutualistic symbiosis occurring between termites and their gut microbiota was used as an experimental hook to discover the previously unknown ribosomal RNA-based species identity of several symbionts. The ability to systematically identify bacteria carrying a particular gene and to link any two or more genes of interest to single species residing in complex ecosystems opens up new opportunities for research on the environment.

A major challenge of environmental science is the identification of microbial species capable of catalyzing important activities *in situ* (1). PCR-based techniques that use single genes as proxies for organisms or key microbial activities continue to provide valuable insights into microbial community diversity (2–4). However, it has been difficult to interrelate gene inventories to derive correspondences between any two or more specific genes of interest, or to determine the phylogenetic species identity of organisms carrying particular genetic capabilities. Metagenomic (5) analyses of complex communities are dominated by genome “shrapnel”; unless the microbial community is dominated by one or a few species (6, 7), resident genomes are not reliably reconstructed via computation (8, 9). A gene of interest can be attributed to a specific organism only if it is linked to an un-

ambiguous phylogenetic marker (i.e., on the same genome fragment) (5, 10). Both PCR and metagenomic studies are typically carried out on homogenized, whole-community genomic DNA preparations. Thus, the cell as a distinct informational entity is almost entirely lost.

Outside of traditional culture-based isolation, few approaches can attribute multiple genes to a single species or cell type. Microautoradiography (11) and stable isotope probing (12) allow detection of cells or retrieval of genetic material from organisms that use a substrate of interest, but these techniques require active cellular incorporation of that substrate. Microscopy-based *in situ* hybridization techniques [fluorescence *in situ* hybridization (FISH) and variants (13, 14)] allow colocalization of sequences through probe hybridization, but these methods require that both genes be actively transcribed, that their sequences be known in advance, and that their difference from related, nontarget genes is sufficient to enable effective probe design and implementation. Single-cell whole-genome amplification has recently been reported for a highly abundant, culturable marine microbial species, but has not yet been shown to be scalable to interrogating multitudes of diverse, co-resident mi-

¹Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA. ²Materials Research and Education Center, Samuel Ginn College of Engineering, Auburn University, Auburn, AL 36849, USA. ³Department of Bioengineering and Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA. ⁴Environmental Science and Engineering Program, California Institute of Technology, Pasadena, CA 91125, USA.

*To whom correspondence should be addressed. E-mail: jleadbetter@caltech.edu

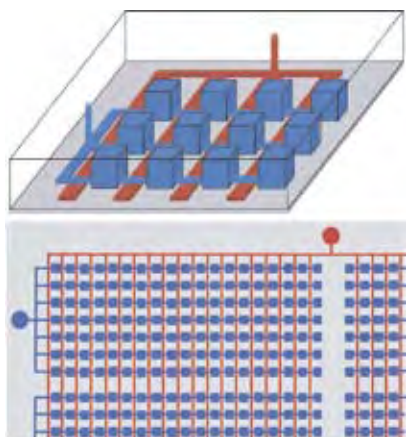


Fig. 1. Microfluidic digital PCR chip. Top: Schematic diagram showing many parallel chambers (blue) connected by channels to a single input. When pneumatic or hydraulic pressure is applied to the control channel network (red), the membranes between the red and blue channels are deflected upward, creating micromechanical valves. When the valves are closed, the continuous blue network is partitioned into independent PCR reactors. Bottom: Schematic showing how a single valve connection can be used to partition thousands of chambers. In the device used, each experimental sample could be partitioned into 1176 chambers, and each device contained 12 such sample panels.

crobes (15). Here, we applied commercially available microfluidic devices to perform a variant of “digital PCR” (16), separating and interrogating hundreds of individual environmental bacteria in parallel.

Microfluidic devices allow control and manipulation of small volumes of liquid (17, 18), in this case allowing for rapid separation and partitioning of single cells from a complex parent sample. Single, partitioned cells served as templates for individual multiplex PCR reactions using primers and probes for simultaneous amplification of both small-subunit ribosomal RNA (rRNA) and metabolic genes of interest. Primers and probes with broad target specificities were used, with subsequent resolution of exact gene sequences after successful amplification and retrieval. This technique operates independent of gene expression, position on the genome,

or physiological state of the cell at the time of harvest. The result was rapid colocalization of two genes (encoding 16S rRNA and a key metabolic enzyme) to single-genome templates, along with the determination of the fraction of cells within the community that encoded them. Subsequent retrieval of PCR products from individual chambers allowed sequence analysis of both genes.

Phylogenetic analysis of the rRNA gene allows classification of the host bacterium, and the metabolic gene is sequenced to confirm that the cell carried the genotype of interest. Additionally, because microfluidic digital PCR yields fluorescent signal upon amplification of a gene regardless of the number of copies present in the cell, this approach can yield estimates of the fraction represented by a given species within the general microbial community. The number of *rrn* operons present in a genome can vary widely, ranging from 1 [e.g., *Rickettsia prowazekii* (19)] to 15 [e.g., *Clostridium paradoxum* (20)], confounding the interpretation of traditional environmental gene inventories. Moreover, the use of single-cell PCR to prepare clone libraries avoids complications and PCR artifacts such as amplification biases and unresolvable chimeric products (21).

We used this technique to examine a complex, species-rich environment: the lignocellulose-decomposing microbial community resident in the hindguts of wood-feeding termites. Therein, the bacterial metabolism known as CO₂-reductive homoacetogenesis is one of the major sources of the bacterial fermentation product acetate (22). Acetogenic bacteria must compete for hydrogen with *Archaea* that generate methane, a potent greenhouse gas for which termites are considered a small yet significant source. Because of their high rates of bacterially mediated homoacetogenesis, many termites contribute less to the global methane budget than they might otherwise (23). Additionally, acetate serves as the insect host’s major carbon and energy source, literally fueling a large proportion of this mutualistic symbiosis (22, 24, 25). A key gene of the homoacetogenesis pathway encodes formyl-tetrahydrofolate synthetase (FTHFS) (26). Although a diverse inventory



Fig. 2. Multiplex microfluidic digital PCR of single cells in environmental samples. Six panels from a representative experiment show microfluidic digital PCR on diluted hindgut contents harvested from a single *Z. nevadensis* individual. Left: Multiplex PCR using “all-bacterial” 16S rRNA gene (red fluorescence) and “clone H group” (27) FTHFS gene (green fluorescence primers and probes. Reaction chambers that contained both genes in 1/500,000 dilutions from this and other on-chip experiments were sampled and the PCR products were analyzed (see Fig. 3). Right: The same, except that 16S rRNA primers specifically targeted members of the “termite cluster” (33) of the spirochetal genus *Treponema*.

of termite hindgut community FTHFS variants already existed (27), the identities of the organisms dominating homoacetogenesis in termites had remained uncertain. Here, with the use of microfluidics, we discovered the identities of a multitude of FTHFS-encoding organisms by determining their specific 16S rRNA gene sequences.

The “clone H group” of FTHFS genotypes corresponds to a large fraction of the sequences collected during an inventory of FTHFS genes present in the termite hindgut environment (27). We designed a specific primer set and a fluorescein-labeled probe capable of on-chip detection and amplification of the genotypes comprising this FTHFS group. We also redesigned broad-specificity “all-bacterial” 16S rRNA gene primers and used a previously published probe (28) to amplify and detect bacterial rRNA genes. Both the all-bacterial 16S rRNA gene and clone H group FTHFS primer-probe sets showed single-molecule sensitivity in multiplex on-chip reactions using purified plasmid or termite gut community DNA. The observed success rate for the amplification of individual genes from single-molecule templates was 48% (fig. S1) (29); thus, the

success rate for coamplification of two genes from single-molecule templates is estimated to be about 1 in 5.

Freshly collected termite hindgut luminal contents were suspended in a PCR reaction buffer and loaded into a microfluidic device (29). Each microfluidic panel uses micromechanical valves to randomly partition a single PCR mixture into 1176 independent 6.25-nl reaction chambers (Fig. 1). We considered single-cell separation to be achieved when fewer than one-third of chambers showed multiple cells or cell aggregates (30). PCR was carried out on a conventional flat-block thermocycler. Amplification was monitored using 5′ nuclease probes to generate a fluorescent signal detected with a modified microarray scanner.

Multiplex PCR amplifications from single cells or cell aggregates were successfully performed using diluted gut contents that had been partitioned on-chip (Fig. 2, left). We found global averages of $1.2 (\pm 0.8) \times 10^8$ total bacterial 16S rRNA gene encoding units and $1.5 (\pm 1.0) \times 10^6$ total clone H

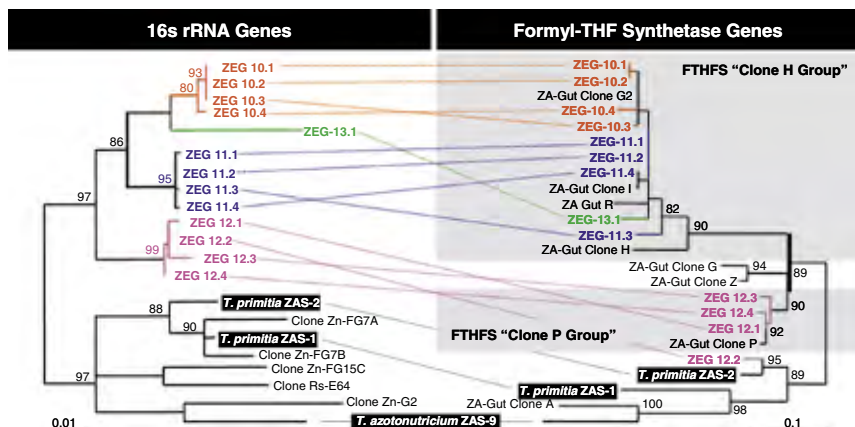


Fig. 3. "Clone H group" and "clone P group" FTHFS genes are encoded by not yet cultivated termite gut treponemes. Left: Phylogenetic tree of 16S rRNA genes cloned from cultivated strain isolates (orange) and from hindgut community microbiota. Right: Phylogenetic tree of FTHFS genes from the termite hindgut. Dotted lines connect genes believed to originate from the same genome. Incongruent gene phylogenies implicate acquisition of FTHFS genes via lateral gene transfer and can be observed in both isolated species (*Treponema primitia* ZAS-1) and proposed "environmental genomovars" (ZEG 12.2). Scale bars represent substitutions per alignment position. The trees were constructed using TreePuzzle (39); 630 (16S rDNA) and 249 (FTHFS) nucleotide positions were used. Citations for all sequences are listed in table S1.

group FTHFS gene encoding units per *Zootermopsis nevadensis* termite (31). This suggests that, in *Z. nevadensis*, these particular FTHFS genes are carried by a minority population representing ~1% of gut symbionts. The observed variability of these measurements was not surprising, as the *Z. nevadensis* specimens examined were collected from different colonies and locations and had been maintained in captivity for varying periods of time.

Amplification products were retrieved from reaction chambers via syringe needle and were reamplified, cloned, sequenced, and analyzed using standard methods. Twenty randomly selected chambers that had amplified only a 16S rRNA gene (and not FTHFS) yielded a diversity of *Endomicrobia*, *Firmicutes*, *Bacteroidetes*, *Proteobacteria*, and *Spirochaetes* ribotypes, as expected on the basis of prior 16S rRNA gene clone libraries (32) (figs. S2 and S3). Two-thirds of chambers positive for FTHFS genes did not amplify 16S rRNA genes when either all-bacterial or termite treponeme-specific rRNA gene primers were used. This amplification

success rate is comparable to that observed when purified, single-molecule templates were used (e.g., fig. S1) and remains a target for refinement and improvement in the future.

PCR products were retrieved and analyzed from 28 reaction chambers that coamplified both FTHFS and 16S rRNA genes. In 10 of those reactions, sequence analyses revealed that the FTHFS gene had coamplified with a clade of closely related 16S rRNA gene sequences affiliating with the "termite spirochete cluster" (33) of the genus *Treponema*. Members of this novel clade were never observed in chambers that lacked FTHFS gene amplification. An additional three chambers contained a single FTHFS type and multiple 16S rRNA genotypes, one of which in each affiliated with the above-mentioned group [*Zootermopsis* environmental genomovar (ZEG) 11.4, 10.2, and 10.1]. These latter reactions also contained two additional other *Spirochaetes* (Zn-FG7A and B in Fig. 3) in one chamber, a single *γ-Proteobacterium* sequence (Zn-FG12) in the second, and a *Firmicutes* sequence (Zn-FG1) in the third.

The remaining 15 chambers analyzed (which coamplified FTHFS and rRNA genes) yielded 16S rDNA sequences in proportions that corresponded well with the ribotype diversity encountered in the general non-FTHFS encoding population. On the basis of this evidence, we conclude that the unique cluster of termite gut treponeme rRNA gene sequences that were repeatedly identified in FTHFS-containing chambers represents the ribotype of the FTHFS-encoding cells. We attribute the instances of FTHFS colocalization with other rRNA gene sequences to cell-cell aggregations. The latter is not to be unexpected in a complex, wood particle-filled, sticky environment such as the termite hindgut (34, 35). Such aggregations appear to be largely random, although there may be a slight enrichment of proteobacterial sequences relative to the general population (figs. S2 and S3). Our results show that FTHFS sequences present in ~1% of bacterial cells were, in 13 of 28 trials, found in association with a 16S rRNA sequence type not identified in 20 random samplings of the bacterial population (16S rRNA-only chambers) at large. The probability of a 16S rRNA gene sequence type that is present in less than 5% of the population randomly colocalizing with FTHFS in 13 of 28 trials is low, on the order of 10^{-10} (36).

Refined phylogenetic analysis of 16S rRNA gene sequences that were repeatedly isolated from FTHFS-containing reaction chambers revealed that all such 16S rRNA gene sequences affiliated within the termite gut treponeme cluster of *Spirochaetes*. These 16S rRNA genes group into four distinct ribotype clusters (Fig. 3). These four sequence types share within-group sequence identity of >99% and between-group identities of 95 to 99%. We propose the term “environmental genomovar” (genome variant) to describe not-yet-cultivated organisms shown to encode two or more known genes of interest. Here, we label the four 16S ribotypes identified as ZEG 10 through 13; three genomovars (ZEG 10, 11, and 13) encode clone H group FTHFS sequences, whereas one genomovar (ZEG 12) encodes a clone P group FTHFS sequence. Previously, nine termite gut trepo-

nemes had been isolated and assigned the strain epithet ZAS (*Zootermopsis* acetogenic spirochete) 1 through 9 (37, 38).

To build additional support for a spirochetal origin of clone H group FTHFS genotypes, we designed and used a termite treponeme-specific 16S rRNA gene primer set and gene probe, with the aim of reducing nonspirochetal background (Fig. 2, right). The frequency with which clone H group FTHFS genes were recovered increased from 1 in 175 cells of the general bacterial population, to 1 in 16 treponemal cells [several termite gut treponemes are already known or suspected to encode FTHFS genotypic variants that would not amplify with the clone H group FTHFS primer and probe set (27) (fig. S1)]. Similar to the amplification success rates observed in experiments using the “all-bacterial” 16S rRNA gene primers (Fig. 2, left) and those using the clone H primers against purified single-molecule templates (fig. S1), about one-third of FTHFS-positive reaction chambers also amplified detectable levels of 16S rRNA genes. Treponemal cells were deduced to constitute 10 to 12% of the bacterial community of *Z. nevadensis* (comparing amplification frequencies in the left and right panels of Fig. 3).

Our results show that specific, not yet cultivated *Treponema* species encode variants of a key gene underlying the dominant bacterial metabolism known to affect the energy needs of their termite hosts. The microfluidic, multiplex digital PCR approach taken here can be extended to expand our understanding of the genetic capacities of not-yet-cultivated species, and to collect and collate genetic information in a manner that builds conceptual genomovars that directly represent the organisms catalyzing important activities in various environments of global relevance.

References and Notes

1. N. D. Gray, I. M. Head, *Environ. Microbiol.* **3**, 481 (2001).
2. E. Zuckerkandl, L. Pauling, *J. Theor. Biol.* **8**, 357 (1965).
3. P. Hugenholtz, B. M. Goebel, N. R. Pace, *J. Bacteriol.* **180**, 4765 (1998).
4. S. J. Sogin, M. L. Sogin, C. R. Woese, *J. Mol. Evol.* **1**, 173 (1971).
5. C. S. Riesenfeld, P. D. Schloss, J. Handelsman, *Annu. Rev. Genet.* **38**, 525 (2004).
6. G. W. Tyson *et al.*, *Nature* **428**, 37 (2004).

7. M. Strous *et al.*, *Nature* **440**, 790 (2006).
8. S. G. Tringe *et al.*, *Science* **308**, 554 (2005).
9. J. C. Venter *et al.*, *Science* **304**, 66 (2004); published online 4 March 2004 (10.1126/science.1093857).
10. O. Béjà *et al.*, *Science* **289**, 1902 (2000).
11. J. L. Nielsen, D. Christensen, M. Kloppenborg, P. H. Nielsen, *Environ. Microbiol.* **5**, 202 (2003).
12. M. Manefield, A. S. Whiteley, R. I. Griffiths, M. J. Bailey, *Appl. Environ. Microbiol.* **68**, 2002 (2002).
13. R. I. Amann, W. Ludwig, K. H. Schleifer, *Microbiol. Rev.* **59**, 143 (1995).
14. K. Zwirgmaier, W. Ludwig, K. H. Schleifer, *Mol. Microbiol.* **51**, 89 (2004).
15. K. Zhang *et al.*, *Nat. Biotechnol.* **24**, 680 (2006).
16. B. Vogelstein, K. W. Kinzler, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9236 (1999).
17. T. Thorsen, S. J. Maerkl, S. R. Quake, *Science* **298**, 580 (2002); published online 26 September 2002 (10.1126/science.1076996).
18. J. W. Hong, S. R. Quake, *Nat. Biotechnol.* **21**, 1179 (2003).
19. H. Pang, H. H. Winkler, *J. Bacteriol.* **175**, 3893 (1993).
20. F. A. Rainey, N. L. Ward-Rainey, P. H. Janssen, H. Hippe, E. Stackebrandt, *Microbiology* **142**, 2087 (1996).
21. S. G. Acinas, R. Sarma-Rupavtarm, V. Klepac-Ceraj, M. F. Polz, *Appl. Environ. Microbiol.* **71**, 8966 (2005).
22. J. A. Breznak, J. M. Switzer, *Appl. Environ. Microbiol.* **52**, 623 (1986).
23. A. Brauman, M. D. Kane, M. Labat, J. A. Breznak, *Science* **257**, 1384 (1992).
24. D. A. Odelson, J. A. Breznak, *Appl. Environ. Microbiol.* **45**, 1602 (1983).
25. A. Tholen, A. Brune, *Environ. Microbiol.* **2**, 436 (2000).
26. L. G. Ljungdahl, *Annu. Rev. Microbiol.* **40**, 415 (1986).
27. T. M. Salmassi, J. R. Leadbetter, *Microbiology* **149**, 2529 (2003).
28. M. T. Suzuki, L. T. Taylor, E. F. DeLong, *Appl. Environ. Microbiol.* **66**, 4605 (2000).
29. See supporting material on Science Online.
30. Assuming a Poisson distribution, if 67% of chambers are empty, then the expected number of cells per chamber is $-\ln 0.67$ or 0.40. The probability that a chamber contains more than one cell is $1 - 0.67 - ((e^{-0.40}) \times (0.40^1)) / (1!) = 0.22$ or 22%.
31. Value ± 1 standard deviation; 13 termites served as source of cells for $n = 32$ sample panels. All sample panels that met our conditions for single-cell separation and contained at least one FTHFS-positive chamber were used in calculation of gut bacterial loads.
32. M. Ohkuma, T. Kudo, *Appl. Environ. Microbiol.* **62**, 461 (1996).
33. T. G. Lilburn, T. M. Schmidt, J. A. Breznak, *Environ. Microbiol.* **1**, 331 (1999).
34. J. A. Breznak, H. S. Pankratz, *Appl. Environ. Microbiol.* **33**, 406 (1977).
35. J. R. Leadbetter, J. A. Breznak, *Appl. Environ. Microbiol.* **62**, 3620 (1996).
36. The binomial distribution function was used to calculate the probability that, in 13 of 28 trials, a sequence that is present in 5% of chambers (0 of 20 16S-only chambers) would randomly colocalize with clone H group FTHFS sequences.
37. J. R. Leadbetter, T. M. Schmidt, J. R. Graber, J. A. Breznak, *Science* **283**, 686 (1999).
38. T. G. Lilburn *et al.*, *Science* **292**, 2495 (2001).
39. H. A. Schmidt, K. Strimmer, M. Vingron, A. von Haeseler, *Bioinformatics* **18**, 502 (2002).
40. We thank M. Unger, A. Daridon, and L. Warren for advice and discussions. Supported by NIH grant 1R01 HG002644-01A1, NIH National Research Service Award grant 5 T32 GM07616, an NIH Director's Pioneer Award, and NSF grant DEB-0321753. S.R.Q. is a founder, shareholder, and consultant for Fluidigm Corporation.

16 June 2006; accepted 30 October 2006
 10.1126/science.1131370

HotStart-IT™: A Novel Hot Start PCR Method Based on Primer Sequestration

Introduction

A common problem with the polymerase chain reaction is the formation of non-specific products, especially primer-dimers. These unwanted products interfere with generating the desired amplicons. A major cause for this lack of specificity occurs during reaction set-up and before the initial denaturation step. At these lower temperatures before denaturation, primers can bind to non-target sequences with incomplete homology, such as other primers⁽¹⁾. These non-specific hybrids can be extended by the polymerase, even at lower temperatures, which creates competing targets during subsequent PCR cycling (e.g. primer-dimers). Hot start PCR methods provide a solution to this lack of specificity by reducing or eliminating non-specific product formation before high-temperature cycling⁽²⁾.

Current hot start methods target the polymerase by muting its activity before the initial denaturation step, most commonly with a blocking antibody or chemical modification. Problems with these methods are two-fold: 1) the antibody is from hybridoma cells which can contaminate reactions with mammalian DNA; and 2) removal of the chemical blocking group on the polymerase typically requires initial denaturation times of greater than 10 minutes which causes heat-damage to DNA samples. An ideal solution is a hot start method that eliminates the risk of contaminating mammalian nucleic acids while also allowing shorter initial denaturation times.

To address these issues, USB Corporation has developed an alternative hot start strategy called HotStart-IT™. This

novel method uses a single-stranded DNA binding protein to sequester primers at lower temperatures, making them unavailable for extension by the polymerase. Following the initial denaturation step, the binding protein is inactivated and the primers are free to participate in the amplification reaction (Fig. 1). This technique solves the problems that arise from other hot start methods because the binding protein is produced in *E. coli* and is heat-labile so that initial denaturation times are 2 minutes or less. Thus, there is no chance of mammalian genomic DNA contamination and less chance of heat-induced DNA damage. Another advantage is that because the method targets the primers and not the polymerase, it is portable to a variety of thermostable polymerases. USB has demonstrated that the "primer-sequestration" technique effectively blocks non-specific product formation before thermal-cycling and enhances end-point and real-time PCR reactions.

Method Validation

To validate the USB HotStart-IT™ method, a 306 base pair target from the single-copy numb gene was amplified from human genomic DNA with USB Taq PCR Master Mix (without hot start) or with HotStart-IT™ Taq PCR Master Mix. The standard Taq Master Mix does not contain any binding protein and, thus, has no hot start feature. Primers were designed with 3 bases of complete homology at their 3'-ends to favor primer-dimer formation. To provide the most extreme testing conditions, the reactions were set-up

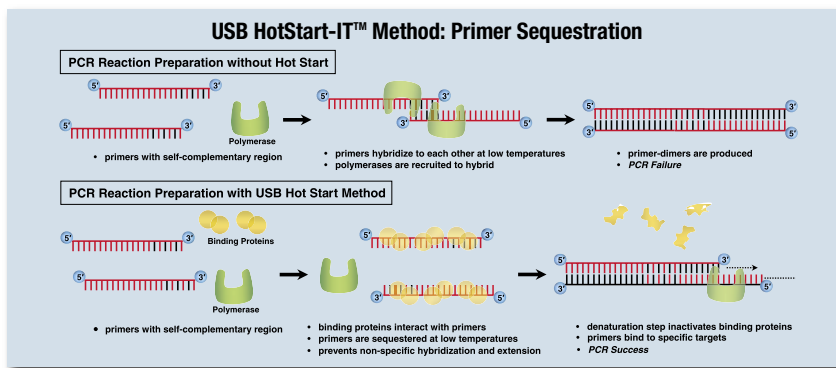


Fig. 1. HotStart-IT™ method. Top Panel: Non-specific products can be generated at low temperatures which causes PCR reaction failure. Bottom Panel: HotStart-IT™ binding protein blocks non-specific product formation at low temperatures which results in successful PCR reactions.

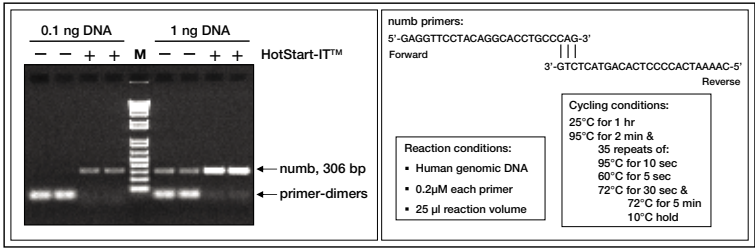


Fig. 2. HotStart-IT™ validation test.
Left Panel: USB hot start method results in a shift from primer-dimers to the desired target.
Right Panel: Description of reaction conditions for the experiment.

at room temperature and were also pre-incubated at 25°C for one hour prior to thermal-cycling to facilitate primer binding and low-temperature synthesis. As seen in Figure 2, the results of this “failure-by-design” experiment demonstrated that when no hot start method was used, the primer-dimer created in the pre-incubation step competed with the diluted template for preferential amplification. This problem intensified as less template was used. When HotStart-IT™ was used, a shift from mainly primer-dimers to the desired product occurred, which enhanced the overall specificity and yield of the reaction.

To demonstrate the functional equivalence of USB HotStart-IT™ compared to current hot start methods, the previous assay was performed versus two alternative products using 1 nanogram of human genomic DNA as template. As seen in Figure 3, all three hot start methods caused a shift from primer-dimers to the specific target compared to a reaction which used no hot start method. These data confirm that the USB hot start technique is as effective as methods that use a blocking antibody (Platinum® Taq) and chemical modification (HotStarTaq™) without the risk of mammalian DNA contamination or DNA damage.

Applications

In addition to standard end-point PCR, the HotStart-IT™ method has been extended to other applications. HotStart-IT™ has been added to USB FidelityTaq™, which is a mixture of

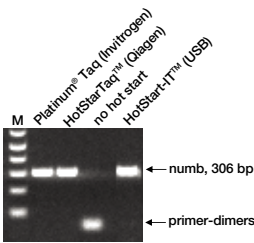


Fig. 3. HotStart-IT™ method comparison.
 “Failure-by-design” experiment with 1 nanogram of human genomic DNA as template. Results demonstrate that the USB hot start method is comparable to both antibody and chemical methods.

Taq DNA polymerase and a minor amount of a proofreading polymerase. This long and accurate enzyme combination allows up to 6-fold higher fidelity than Taq alone and extremely long amplification of targets in excess of 20 kb.

The HotStart-IT™ technology has also been combined into a variety of real-time PCR master mixes. For SYBR® Green I and probe-based (e.g. TaqMan®) detection methods, master mixes are available with or without dUTP and a heat-labile uracil-DNA glycosylase (UDG) for carry-over contamination prevention. These universal mixes come with separate tubes of the passive references dyes ROX™ and fluorescein and are compatible on most PCR instruments. Both the SYBR® Green I (Fig. 4) and probe-based (Fig. 5) mixes have superb sensitivity as they typically detect fewer than 10 target copies and are linearly quantitative over 7 to 8 orders of magnitude with correlation coefficients in excess of 0.95.

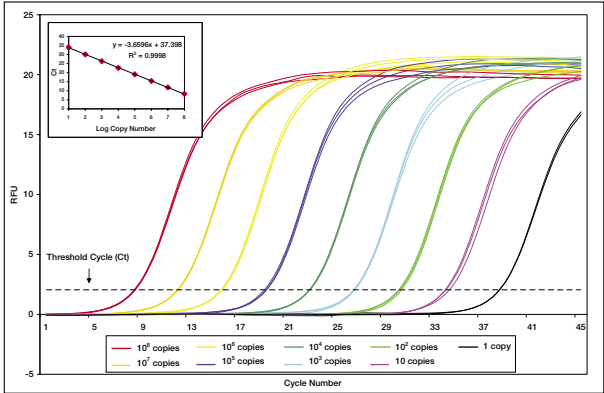


Fig. 4. GAPDH assay using HotStart-IT™ SYBR® Green qPCR Master Mix with dUTP and UDG (PN 75760).
 Triplicate reactions were performed on a cloned region of the human GAPDH gene as template using an ABI 7500 Fast instrument. The non-specific dsDNA binding dye, SYBR® Green I, was used to detect a 122 bp amplicon. ROX was used as the passive reference dye. The amplification process was linear over eight orders of magnitude (see inset) and a single copy of the target could be reliably detected. The no template control reaction generated no measurable fluorescence (not shown).

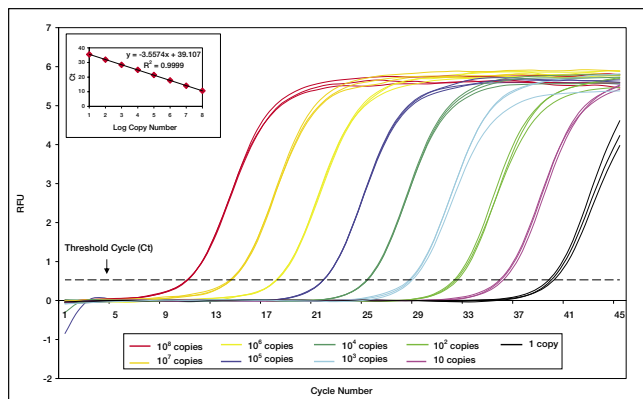


Fig. 5. GAPDH assay using HotStart-IT™ Probe qPCR Master Mix with dUTP and UDG (PN 75764). Triplicate reactions were performed on a cloned region of the human GAPDH gene as template using an ABI 7500 Fast instrument. A TaqMan® probe with FAM as the reporter fluorophore and BHQ-1® as the quencher was used to detect the 122 bp amplicon. ROX was used as the passive reference dye. The amplification process was linear over eight orders of magnitude (see inset) and a single copy of the target could be reliably detected. The no template control reaction generated no measurable fluorescence (not shown).

Master mixes are also available for real-time, one-step RT-PCR (reverse transcription-PCR) using either SYBR® Green or probes as the detection chemistries. One-step RT-PCR simplifies RNA quantification by combining conversion of RNA to cDNA by MMLV-RT and amplification with Taq DNA Polymerase in a single, sequential reaction.

For custom applications in which a certain thermostable polymerase is preferred or required, the HotStart-IT™ binding protein may be used separately by adding it directly to the polymerase to provide hot start capability. This binding protein, which is the active component of the HotStart-IT™ method, is also useful in multiplex situations with four or more primer sets. Since one microgram of binding protein will sequester about 5 pmol of primers, users may tailor the amount of protein based upon their multiplexing demands to achieve the best results.

Summary

Hot start PCR methods provide a solution to the problem of non-specific product formation that can occur before high-temperature cycling. The USB HotStart-IT™ method is a novel approach that uses a binding protein which traps or sequesters primers before thermal cycling and releases them after heat-inactivation. Like current methods, the

advantages of using hot start methods are: 1) room temperature reaction assembly; 2) high specificity; and 3) high sensitivity. Unlike current methods, the advantages of the USB hot start method are: 1) the binding protein is produced in *E. coli* which eliminates potential mammalian DNA contamination; and 2) no extensive initial denaturation step is necessary which could damage DNA samples. In addition, because the method is directed toward the primers and not the polymerase, it is portable to any thermostable polymerase.

USB offers the HotStart-IT™ technology in a wide variety of products ranging from standard or end-point PCR, to high fidelity PCR, to real-time or quantitative PCR. Each product line is available in a 2X Master Mix format which provides for maximum convenience, ease-of-use, and reproducibility. These master mixes require only template, primers and water and are designed to be compatible on most PCR instruments. It is recommended to use USB HotStart-IT™ products for all demanding PCR applications.

References

1. Gillam, S. *et al.*, (1975) *Nucleic Acids Research* **2**(5), 625–634.
2. Chou, Q. *et al.*, (1992) *Nucleic Acids Research* **20**(7), 1717–1723.



USB Corporation | 26111 Miles Road | Cleveland, OH 44128 216.765.5000 | www.usbweb.com
USB Europe GmbH | Hauptstrasse 1 | 79219 Staufen, Germany +49 (0)76 33 933 40 0 | www.usbweb.de

©2007 USB Corporation.

USB and the logo design are registered trademarks of USB Corporation. HotStart-IT and the phrase Fueling Innovation in Life Science are trademarks of USB Corporation. HotStart-IT Taq DNA Polymerase – Patent pending.

Taq DNA Polymerase – sold under licensing arrangements with Applied Biosystems. Purchase is accompanied by a limited license to use it in the Polymerase Chain Reaction (PCR) process in conjunction with a thermal cycler whose use in the automated performance of the PCR process is covered by the up-front license fee, either by payment to Perkin-Elmer or as purchased, i.e., an authorized thermal cycler. The Polymerase Chain Reaction (PCR) is covered by patents owned by Roche Molecular Systems and F. Hoffmann-La Roche Ltd. Uracil-DNA Glycosylase (UDG) – Purchase of this product is accompanied by a limited license under U.S. Patent Nos. 5,035,996; 5,683,896; 5,945,313; 6,518,026 and 6,287,823 and corresponding foreign patents. SYBR is a registered trademark of Molecular Probes, Inc. and is provided under an agreement with Molecular Probes, Inc. TaqMan is a registered trademark of Roche Molecular Systems, Inc. Platinum is a registered trademark of Invitrogen. HotStarTaq is a trademark of Qiagen. ROX is a trademark of Applied Biosystems or its subsidiaries in the U.S. and certain other countries. BHQ-1 is a registered trademark of Biosearch Technologies, Inc.

QIAGEN Exhibit 1018.0030

HotStart-IT.TM

Less Heat. No DNA Damage. We couldn't be more specific.

USB's novel hot start PCR method requires no extensive heat denaturation step. The result is no damage to precious samples with increased specificity and yield. HotStart-ITTM: highly sensitive to all your critical PCR applications.

HotStart-ITTM Taq DNA Polymerase

- Based on primer sequestration—a novel method developed by USB scientists
- Unique protein binds primers to prevent mispriming
- Primers are released during heat denaturation

Benefits

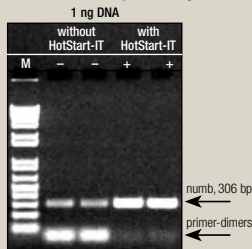
High Sensitivity

- Detects <10 target copies
- Linear dynamic range of 7–8 orders of magnitude, $R^2 \geq 0.95$

Ease of Use

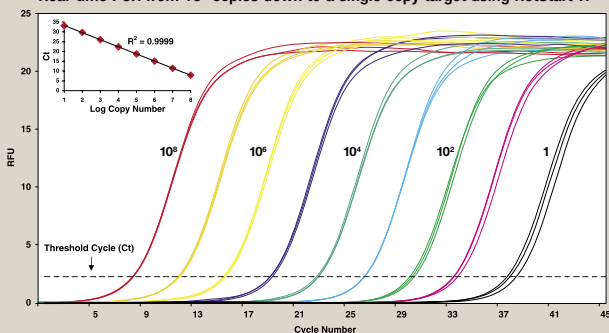
- Multiple instrument compatibility

Increased specificity of HotStart-ITTM Taq DNA Polymerase



Results clearly demonstrate a shift from predominantly primer-dimers to the specific target when HotStart-ITTM is included in the reactions.

Real-time PCR from 10^8 copies down to a single copy target using HotStart-ITTM



For more information on HotStart-ITTM

call 800.321.9322 or visit www.usbweb.com

In Europe: +49(0)76 33-933 40 0 or visit www.usbweb.de/hotstart

 **usb**
Innovation
in Life Science