

# Cell-free cloning using $\phi$ 29 DNA polymerase

Clyde A. Hutchison III\*, Hamilton O. Smith, Cynthia Pfannkoch, and J. Craig Venter

Synthetic Biology Group, The J. Craig Venter Institute, 9704 Medical Center Drive, Rockville, MD 20850

Contributed by Clyde A. Hutchison III, October 7, 2005

We describe conditions for rolling-circle amplification (RCA) of individual DNA molecules 5–7 kb in size by  $>10^9$ -fold, using  $\phi$ 29 DNA polymerase. The principal difficulty with amplification of small amounts of template by RCA using  $\phi$ 29 DNA polymerase is “background” DNA synthesis that usually occurs when template is omitted, or at low template concentrations. Reducing the reaction volume while keeping the amount of template fixed increases the template concentration, resulting in a suppression of background synthesis. Cell-free cloning of single circular molecules by using  $\phi$ 29 DNA polymerase was achieved by carrying out the amplification reactions in very small volumes, typically 600 nl. This procedure allows cell-free cloning of individual synthetic DNA molecules that cannot be cloned in *Escherichia coli*, for example synthetic phage genomes carrying lethal mutations. It also allows cell-free cloning of genomic DNA isolated from bacteria. This DNA can be sequenced directly from the  $\phi$ 29 DNA polymerase reaction without further amplification. In contrast to PCR amplification, RCA using  $\phi$ 29 DNA polymerase does not produce mutant jackpots, and the high processivity of the enzyme eliminates stuttering at homopolymer tracts. Cell-free cloning has many potential applications to both natural and synthetic DNA. These include environmental DNA samples that have proven difficult to clone and synthetic genes encoding toxic products. The method may also speed genome sequencing by eliminating the need for biological cloning.

DNA sequencing | rolling-circle amplification | synthetic DNA

Cloning foreign DNA sequences into vectors that can replicate in *Escherichia coli* or other host cells is arguably the defining technique of modern molecular genetics. However, this method has limitations, because certain sequences are difficult to propagate by using available vector/host systems. Some sequences are toxic to the host cell, whereas others are prone to replication errors. Cloning is also time-consuming. The isolation of a clone from a library and its preparation as a sequencing template typically takes longer than 2 days.

For these reasons, a cell-free enzymatic method for cloning and amplifying DNA sequences has long been appealing. The “polony” method for *in situ* PCR amplification of individual molecules (1) was designed to replace biological cloning by *in vitro* amplification with the aim of speeding large sequencing projects. However, PCR has certain characteristics that make it undesirable as the basis for a general cell-free cloning method. These include mutant “jackpots” that result when mutations arise during the early cycles of a PCR reaction and a lack of processivity leading to stuttering at sequences of low complexity, especially homopolymer tracts.

We describe here the use of  $\phi$ 29 DNA polymerase to clone individual molecules. This enzyme was attractive for cell-free cloning because it copies DNA with high fidelity (2), has a proofreading activity (3), and is highly processive (4). Our procedure is based on the rolling-circle amplification (RCA) reaction carried out isothermally by  $\phi$ 29 DNA polymerase at 30°C (5). Fig. 1 diagrams the stages of this reaction and is based on the work of Lasken and colleagues (5). Strands averaging 70 kb in length are copied by the enzyme (4). Therefore, products from small circular templates contain many tandem repeats of the starting molecule. These strands in turn serve as templates for new synthesis so that the reaction undergoes an exponential

phase in which the rate of initiation of new strands is proportional to the quantity of DNA already formed. In this reaction, synthesis is primed by random hexanucleotides rendered resistant to the exonuclease activity of the polymerase by the incorporation of phosphothioate linkages. Background synthesis is commonly observed when  $\phi$ 29 polymerase is primed with such random hexamers, if the reaction is carried out in the absence of added template. This background synthesis, which is believed to be templated by the random hexamer primers (for example, see ref. 6), has made the amplification of very small amounts of DNA difficult. We reasoned that reducing the volume of a  $\phi$ 29 polymerase reaction, while keeping the amount of template constant, would increase the template concentration and suppress background synthesis. This approach led to success in cell-free cloning of individual DNA molecules by using submicroliter reaction volumes.

## Materials and Methods

**DNA Preparations.**  $\phi$ X174 *am3cs70* single-stranded virion DNA and M13mp18 single-stranded virion DNA were obtained from NEB (Beverly, MA). The DNA referred to as syn $\phi$ X in the text was the ligated circular double-stranded synthetic  $\phi$ X174 DNA described by Smith *et al.* (7). The *Mycoplasma genitalium* genomic DNA library was prepared from DNA isolated from strain G37, obtained from the American Type Culture Collection in 1996. DNA sheared to 1.5–2.2 kb was cloned into a medium copy pBR322 derivative as described in ref. 8.

**Enzymes.**  $\phi$ 29 DNA polymerase was purchased from NEB in pure form at a concentration of 10,000 units/ml. Inorganic pyrophosphatase from United States Biochemical was supplied at 40 units/ml. Restriction enzymes were from NEB.

**$\phi$ 29 Polymerase Amplification Reactions.** Reactions contain 37 mM Tris·HCl (pH 7.5), 50 mM KCl, 10 mM MgCl<sub>2</sub>, 5 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 1 mM each of the four dNTPs, 1 mM DTT, 1× BSA (NEB), 0.2% Tween 20, 1 unit/ml yeast pyrophosphatase, 540 units/ml  $\phi$ 29 DNA polymerase, and 50  $\mu$ M exonuclease-resistant hexamer mix (Fidelity Systems, Gaithersburg, MD). All components except the enzymes and hexamers were made up as a 2× concentrated buffer (2xG-buffer). Immediately before setting up reactions, enzymes and H<sub>2</sub>O were added to 2xG-buffer, this was incubated for 10 min at room temperature, and the hexamers were added so that the concentration of all components was now 1.5× the final reaction conditions (this is  $\phi$ 29 Mix). Reactions were initiated by adding one volume of template DNA in TE buffer [10 mM Tris/1 mM EDTA (pH 8)] to two volumes of  $\phi$ 29 Mix. Reactions were carried out at 30°C. For our standard 600-nl reaction, 400 nl of  $\phi$ 29 Mix was added to a 0.5-ml PCR tube by using a 2- $\mu$ l Pipetman (Gilson, Middleton, WI) with ART 10 Reach tips (Molecular BioProducts, San Diego). Then, 200 nl of template dilution was added and the reaction was mixed by

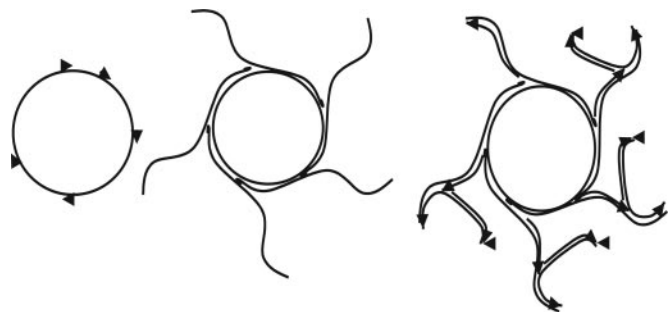
Conflict of interest statement: No conflicts declared.

Freely available online through the PNAS open access option.

Abbreviation: RCA, rolling-circle amplification.

\*To whom correspondence should be addressed. E-mail: chutchison@venterininstitute.org.

© 2005 by The National Academy of Sciences of the USA



**Fig. 1.** RCA by  $\phi 29$  DNA polymerase, after Dean *et al.* (5). Arrowheads represent random hexamer primers. This drawing is not to scale; circular templates are typically 5 kb, and product strands average 70 kb in length.

pipetting up and down  $\approx 20$  times. Reactions were overlaid with 10  $\mu\text{l}$  of bio-technology grade mineral oil (Bio-Rad) to prevent evaporation, then centrifuged briefly and checked visually to make sure the aqueous phase formed a small sphere at the bottom of the tube. Reactions were incubated in a thermocycler at 30°C for 6 h, then held at 4°C until analysis.

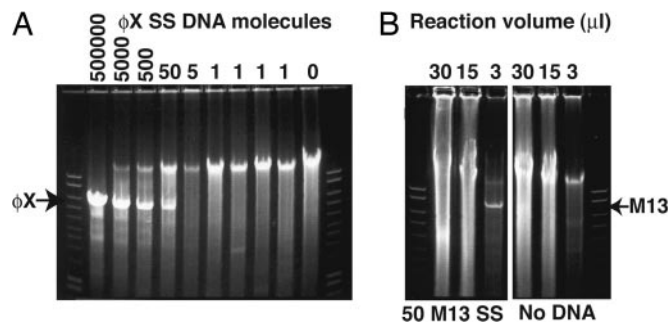
PCR was carried out in Advantage 2 buffer, using the Advantage HF 2 PCR kit (no. K1914-1, Clontech).

**Restriction Analysis of Amplification Reactions.** To digest the entire reaction, 20  $\mu\text{l}$  of buffer 2 plus BSA (1 $\times$ , NEB) containing 10 units of PstI was added to the reaction under mineral oil, and the tube was mixed briefly and incubated at 37°C for 30–60 min. The aqueous phase was removed and loaded onto an E-Gel (Invitrogen). Unless stated otherwise, 0.8% E-Gels were electrophoresed for 40 min at 60 V. DNA size markers were HyperLadder I (Biolone, Randolph, MA) with bands at 10 kb, 8 kb, 6 kb, 5 kb, 4 kb, 3 kb, 2.5 kb, 2 kb, 1.5 kb, 1 kb, 800 bp, 600 bp, 400 bp, and 200 bp. To analyze  $\phi 29$  polymerase reactions by restriction digestion and also by DNA sequencing, the reactions were diluted with 10  $\mu\text{l}$  of TE and then split for analysis.

**DNA Sequencing.** PCR and  $\phi 29$  polymerase reaction products were treated with shrimp alkaline phosphatase (SAP) and *E. coli* exonuclease I (Exo I) to remove primers and dNTPs before use as sequencing templates. SAP-Exo I master mix contained 0.5  $\mu\text{l}$  of SAP (Boehringer Mannheim, 1 unit/ $\mu\text{l}$ ), 0.1  $\mu\text{l}$  of Exo I (NEB, 20 units/ $\mu\text{l}$ ), 0.5 ml of 10 $\times$  SAP buffer, and 8.9  $\mu\text{l}$  of H<sub>2</sub>O per reaction. To sequence from a PCR reaction, 8  $\mu\text{l}$  of the reaction was added to 10  $\mu\text{l}$  of SAP-Exo I master mix, and the solution was incubated at 37°C for 60 min, incubated at 72°C for 15 min, and held at 4°C until use in sequencing. To sequence directly from a  $\phi 29$  polymerase reaction the 600-nl reaction was diluted with 10  $\mu\text{l}$  of TE. Four microliters of the reaction was added to 5  $\mu\text{l}$  of SAP-Exo I master mix, and the solution was incubated as for the PCR samples above. In both cases, 4  $\mu\text{l}$  of the treated templates were sequenced in 8- $\mu\text{l}$  reactions by using BigDye terminator chemistry (v3.1, Applied Biosystems) with 35 extension cycles of 4 min at 60°C. This high temperature prevents priming by any residual hexamers not degraded by Exo I treatment of  $\phi 29$  polymerase reactions. Reaction products were isopropanol-precipitated, rinsed with ethanol, dissolved in 20  $\mu\text{l}$  of H<sub>2</sub>O, and analyzed by using an Applied Biosystems capillary sequencer (Model 3100) equipped with 50-cm capillaries.

## Results

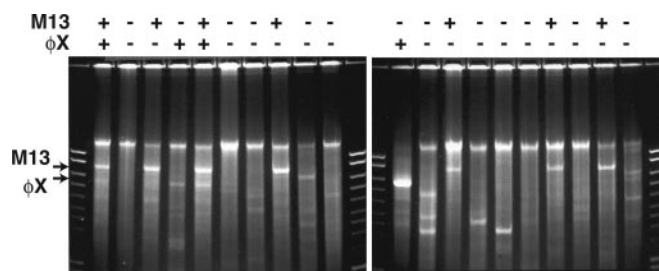
As the amount of template DNA in a  $\phi 29$  polymerase RCA reaction is decreased, the amount of background synthesis increases. If the template is circular  $\phi\text{X174}$  single-stranded DNA (Fig. 2A), then PstI cleaves authentic amplification product at the unique site to produce 5.4-kb linear DNA. The background



**Fig. 2.** Factors affecting background synthesis by  $\phi 29$  polymerase. (A) Background increases with decreasing template concentration. Three-microliter  $\phi 29$  polymerase reactions were primed by the indicated numbers of  $\phi\text{X}$  single-stranded DNA molecules, calculated on the basis of dilution factors from a stock of known concentration. Reactions were digested with PstI and analyzed as described above. (B) Background synthesis is suppressed by reducing  $\phi 29$  polymerase reaction volume.  $\phi 29$  polymerase reactions of 30, 15, and 3  $\mu\text{l}$  were assembled, each containing either 50 M13 single-stranded DNA molecules or no template DNA. The reaction products were digested with PstI, and one-half of each reaction was analyzed by gel electrophoresis as described above. The products of background synthesis, not cleaved by PstI, migrate as large DNA, slower than the largest marker DNA (10 kb).

product, however, does not in general contain PstI sites and migrates as large DNA, making it easy to distinguish from the RCA product of  $\phi\text{X174}$  (Fig. 2A). As the reaction volume is decreased, keeping the amount of template DNA constant, background is dramatically decreased with a marked improvement in signal-to-noise ratio (Fig. 2B). Using 50 molecules of single-stranded M13 DNA (7.2 kb) as template, a PstI digest of a 3- $\mu\text{l}$   $\phi 29$  polymerase reaction displayed a distinct band ( $>10$  ng) of M13 linear. When template was omitted from the reaction, a similar amount of DNA was produced by background synthesis, but this DNA was not cleaved by PstI and migrated more slowly than the largest marker DNA (10 kb). Larger reactions (15 and 30  $\mu\text{l}$ ) produced so much background synthesis that no M13 product was visible.

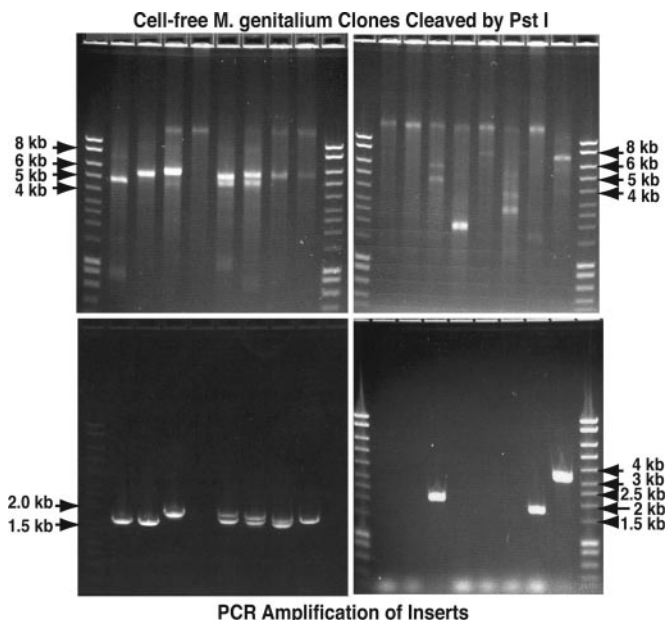
Using 600-nl  $\phi 29$  polymerase reactions, we have amplified single molecules more than  $10^9$ -fold to give 10 ng or more of product DNA, allowing easy visualization by ethidium bromide staining after gel electrophoresis. Fig. 3 displays gel electrophoresis of 20 duplicate 600-nl  $\phi 29$  polymerase reactions digested with PstI. The template for each reaction was 200 nl of the same limiting dilution of a mixture of  $\phi\text{X174}$  (5.4 kb) and M13 (7.2 kb) single-stranded DNAs. Each of these molecules contains a unique PstI site. Some lanes appear to contain neither  $\phi\text{X}$  nor



**Fig. 3.** Cell-free cloning of  $\phi\text{X174}$  and M13 from a mixture. Twenty duplicate 600-nl  $\phi 29$  polymerase reactions were made by using the same limiting dilution of a mixture of  $\phi\text{X174}$  and M13 DNAs (see text). The reaction products were digested with PstI and analyzed as described above. The + and – symbols above the gel indicate the presence or absence of bands with the molecular weights of linear  $\phi\text{X174}$  (5.4 kb) or M13 (7.2 kb) DNAs. The products of background synthesis, not cleaved by PstI, migrate as large DNA, slower than the largest marker DNA (10 kb).







**Fig. 5.** Cell-free cloning of *M. genitalium* genomic DNA. A library consisting of 1.5- to 2.5-kb fragments of *M. genitalium* DNA ligated to a pBR322 derivative was diluted 100-fold in TE from an initial DNA concentration of  $\approx 0.3$  ng/ $\mu$ l, heated at 95°C for 2 min, and quenched on ice. The DNA was diluted an additional 1,000-fold (Upper Left) or 2,000-fold (Upper Right) and used as template in two sets of 600-nl  $\phi$ 29 polymerase reactions (eight duplicate reactions per set), as described in *Materials and Methods*. One-half of each reaction was cleaved with PstI and analyzed by gel electrophoresis (Upper). Inserts were amplified by PCR from the remaining portion of each reaction, using M13 forward and reverse sequencing primers that flank the cloning site in the vector, and analyzed on the lower gels.

was not perfectly homogeneous in size. The inserts were amplified by PCR using the M13 forward and reverse sequencing primer sites that flank the cloning sites in the vector (Fig. 5 Lower). Some lanes contain two bands of slightly different sizes, indicating that two molecules were amplified in those reactions. PCR products that appeared on the gel to be pure species were sequenced from both ends by using the M13 sequencing primers. Six of seven cloned molecules gave clearly readable sequence that matched the *M. genitalium* genome sequence perfectly throughout the readable range (out to  $\approx 650$  bp or greater; see the supporting information). One of the seven reactions appeared to contain a mixture of two sequences. The products of 600-nl  $\phi$ 29 polymerase reactions can also be sequenced directly, without PCR amplification. The chromatogram for such a sequence is shown in Fig. 4B, which displays a sequence that matches the *M. genitalium* genome accurately to beyond 750 bp (see the supporting information).

The high processivity of  $\phi$ 29 polymerase allows accurate amplification of homopolymer tracts from single molecules. DNA from an *M. genitalium* clone containing an A<sub>18</sub>/T<sub>18</sub> tract was diluted and used to prime single-molecule  $\phi$ 29 polymerase reactions. The DNA from these reactions was sequenced directly and portions of the chromatograms are shown in Fig. 4C. One strand shows a run of 18 A residues, and the complementary strand shows 18 T residues, as expected. The sequence is clearly readable beyond the homopolymer run on both strands and matches the *M. genitalium* genome sequence accurately for the readable length of the run ( $>700$  bp; see the supporting information).

### Discussion

Much recent work has focused on the use of  $\phi$ 29 polymerase to amplify small amounts of DNA. Sequencing templates can be

produced by amplification of plasmids from single bacterial colonies using RCA (9). Also, useful amounts of genomic DNA can be amplified from small numbers of cells by using the enzyme in the multiple displacement amplification (MDA) reaction (10). Cell-free cloning of small circular DNA involves even smaller amounts of template ( $<10^{-17}$  g for a 5-kb DNA molecule), and background synthesis, which occurs in the absence of added template when  $\phi$ 29 polymerase is primed by random hexamers, becomes a serious problem. Using submicroliter reaction volumes, we have amplified single circular DNA molecules to give amounts of DNA that can easily be visualized by gel electrophoresis and can be sequenced. The use of small reactions increases the template concentration and suppresses the amount of background synthesis. Reactants do not appear to be limiting in these reactions, which contain enough of the four dNTPs to produce  $\approx 1$   $\mu$ g of DNA per  $\mu$ l of reaction volume. It appears to be a fortunate coincidence that reduction of commonplace reaction volumes by only about one order of magnitude allows cell-free cloning, by using reaction volumes near the lower limit for conventional pipettes. In this article, we describe the use of 600-nl reactions that can be prepared by using conventional manual pipetting devices. However, preliminary results indicate that it is advantageous to reduce volumes even further (unpublished data). Robotic nanoliter pipetting devices [for example, Parallax 350 (Brooks Automation, Chelmsford, MA)] may yield improvements in the method and make its large-scale application more practical.

The only known requirement for a DNA to be cloned by  $\phi$ 29 polymerase is that it is circular, and of such a size that RCA can produce multiple tandem copies of the molecule. In principle, no vector is needed, because a molecule could be circularized in a blunt-end ligation reaction. Alternatively, special  $\phi$ 29 cloning vectors can be designed to facilitate circularization of the target DNA and subsequent use of the products for sequencing or other purposes (unpublished results). Such  $\phi$ 29 cloning vectors can be quite small (a few hundred base pairs or less) because they do not need to support replication in bacteria. The largest molecule we have cloned with  $\phi$ 29 polymerase is M13 (7.2 kb), but we have not tested larger molecules, and it seems likely that circular DNA 10 kb or larger might be cloned by this method. It also seems likely that the use of very small reaction volumes would be useful in the amplification of genomic DNA from single cells by the MDA reaction.

One application of the method described here is the cell-free cloning and sequencing of single synthetic DNA molecules without propagation in bacteria. This method can be used for molecules that are unclonable in *E. coli*. For example, synthetic  $\phi$ X174 genomes containing frameshifts in essential genes can be cloned by using  $\phi$ 29 polymerase (Fig. 4A). These genomes cannot be propagated as phages or, in general, propagated in *E. coli* by cloning in plasmid vectors either, because several of the  $\phi$ X174 gene products are toxic to *E. coli*. We therefore expect that cell-free cloning will be valuable for dealing with synthetic DNA that is difficult to clone by conventional methods.

In applications of cell-free cloning to synthetic DNA it is important that the amplified DNA has exactly the sequence of the starting molecule. We demonstrated that molecules cloned by  $\phi$ 29 polymerase give accurate DNA sequencing results. However, sequencing determines the consensus sequence and does not detect random mutations that occur during amplification. It is therefore of interest to try to estimate the mutation frequency in the  $\phi$ 29 product amplified from a single molecule. However, this analysis is not straightforward because RCA is not a simple doubling process. Consequently, the usual equation, which relates polymerase error rate, mutation frequency, and number of doublings, does not apply. Because many copies are made directly from the original input molecule, jackpots of mutants should not be a problem because they are with muta-



tions arising during the first cycles of a PCR reaction. The  $\phi 29$  polymerase appears to have an accuracy comparable to other polymerases with a 3' proofreading exonuclease (2), which are generally in the range of  $10^{-6}$  to  $10^{-8}$  (11). Calculations based on published estimates of the  $\phi 29$  polymerase error rate (2, 9, 11), and reasonable models of the reaction, indicate that the majority of molecules resulting from  $10^9$ -fold amplification of a 5-kb circular DNA by  $\phi 29$  polymerase should have exactly the same sequence as the parental molecule (see the supporting information).

Cell-free cloning may also be applied to facilitate DNA sequencing projects. It should be useful in situations where cloning in *E. coli* is difficult or impossible. For example, some environmental DNA samples are difficult to clone in *E. coli* (for example, see ref. 12). Cloning using  $\phi 29$  DNA polymerase should be useful in the sequence analysis of such samples. In conventional genome sequencing projects the method could

substantially reduce the time required for preparation of sequencing templates and also reduce problems with unclonable sequences. Mate-paired sequence reads from the ends of  $\phi 29$  polymerase cell-free clones should also facilitate the assembly of data generated by a new generation of sequencers such as the Genome Sequencer 20 (454 Life Sciences/Roche Applied Science) (13, 14). Such devices have extremely high throughput and do not rely on *E. coli* cloning but currently produce only short read lengths of lower quality than conventional methods. Combining such "454 data" with  $\phi 29$  polymerase cell-free clone data could potentially result in an extremely rapid and accurate genome sequencing method.

We thank members of the Synthetic Biology group for helpful discussions, John Glass for help with real-time PCR, and Mahir Maruf for preparation of plasmid DNA. This work was supported by grants from the U.S. Department of Energy and the J. Craig Venter Science Foundation.

1. Mitra, R. D. & Church, G. M. (1999) *Nucleic Acids Res.* **27**, e34.
2. Esteban, J. A., Salas, M. & Blanco, L. (1993) *J. Biol. Chem.* **268**, 2719–2726.
3. Garmendia, C., Bernad, A., Esteban, J. A., Blanco, L. & Salas, M. (1992) *J. Biol. Chem.* **267**, 2594–2599.
4. Blanco, L., Bernad, A., Lazaro, J. M., Martin, G., Garmendia, C. & Salas, M. (1989) *J. Biol. Chem.* **264**, 8935–8940.
5. Dean, F. B., Nelson, J. R., Giesler, T. L. & Lasken, R. S. (2001) *Genome Res.* **11**, 1095–1099.
6. Holbrook, J. F., Stabley, D. & Sol-Church, K. (2005) *J. Biomol. Tech.* **16**, 125–133.
7. Smith, H. O., Hutchison, C. A., III, Pfankoch, C. & Venter, J. C. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 15440–15445.
8. Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., *et al.* (2000) *Science* **287**, 2185–2195.
9. Nelson, J. R., Cai, Y. C., Giesler, T. L., Farchaus, J. W., Sundaram, S. T., Ortiz-Rivera, M., Hosta, L. P., Hewitt, P. L., Mamone, J. A., Palaniappan, C. & Fuller, C. W. (2002) *BioTechniques* **32**, Suppl., 44–47.
10. Sato, M., Ohtsuka, M. & Ohmi, Y. (2005) *Biomol. Eng.* **22**, 129–132.
11. Kunkel, T. A. (2004) *J. Biol. Chem.* **279**, 16895–16898.
12. Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J. M., Segall, A. M., Mead, D., Azam, F. & Rohwer, F. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 14250–14255.
13. Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z., *et al.* (2005) *Nature* **437**, 376–380.
14. Rogers, Y. H. & Venter, J. C. (2005) *Nature* **437**, 326–327.