Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome
Jay Shendure, et al.
Science 309, 1728 (2005);
DOI: 10.1126/science.1117389

The following resources related to this article are available online at www.sciencemag.org (this information is current as of October 31, 2008):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/cgi/content/full/309/5741/1728

Supporting Online Material can be found at:
http://www.sciencemag.org/cgi/content/full/1117389/DC1

A list of selected additional articles on the Science Web sites related to this article can be found at:
http://www.sciencemag.org/cgi/content/full/309/5741/1728#related-content

This article cites 23 articles, 11 of which can be accessed for free:
http://www.sciencemag.org/cgi/content/full/309/5741/1728#otherarticles

This article has been cited by 170 article(s) on the ISI Web of Science.

This article has been cited by 39 articles hosted by HighWire Press; see:
http://www.sciencemag.org/cgi/content/full/309/5741/1728#otherarticles

Information about obtaining reprints of this article or about obtaining permission to reproduce this article in whole or in part can be found at:
http://www.sciencemag.org/about/permissions.dtl
Reports

11. Supporting material is available on Science Online.
12. Because it is a prototype, there has been no attempt to reduce the weight of the backpack—indeed, it is substantially "overdesigned." Further, the 5.6 kg includes the weight of six load cells and one 25-cm-long transducer, each with accompanying brackets and cables, as well as other components that will not be present on a typical pack. In future prototypes, we estimate that the weight will exceed that of a normal backpack by no more than 1 to 1.5 kg.
13. Under high-power conditions (5.6 km hr-1 with 20- and 29-kg loads and 4.8 km hr-1 with a 38-kg load), power generation on the incline was the same as on the flat. Under low-power conditions (4.8 km hr-1 with 20- and 28-kg loads), electricity generation on the incline was actually substantially greater than that on the flat (table S1).
21. Because this savings in metabolic energy represents only 6% of the net energetic cost of walking with the backpack [462 W] (table S3) [17, 18], accurate determinations of the position and movements of the center of mass, as well as the direction and magnitude of the ground reaction forces, are essential to discern the mechanism. This will require twin-force-platform single-leg measurements, as well as a complete kinematics and mechanical energy analysis (19, 20). The energy analysis is made more complex because the position of the load with respect to the backpack frame and the amount of energy stored in the backpack springs vary during the gait cycle. Finally, electromyogram measurements are also important to test whether a change in effective muscle moment arms may have caused a change in the volume of activated muscle and hence a change in metabolic cost (20, 27, 28).
23. This assumes that electronic devices are being powered in real time. If there were a power loss of 50% associated with storage (such as in batteries) and recovery of electrical energy, then these factors would be halved.
24. When not walking, the rack can be disengaged and the generator cranked by hand or by foot. Electrical powers of ~3 W are achievable by hand, and higher wattage can be achieved by using the leg to power it.
29. This work was supported by NIH grants AR46125 and AR83404. Some aspects of the project were supported by Office of Naval Research grant N000140310568 and a grant from the University of Pennsylvania Research Foundation. The authors thank Q. Zhang, H. Hofmann, W. Megill, and A. Dunham for helpful discussions; R. Sprague, E. Maxwell, R. Essner, L. Gazit, M. Yuhas, and J. Milligan for helping with the experimentation; and F. Letterio for machining the backpacks.

Supporting Online Material
www.sciencemag.org/cgi/content/full/309/5741/1725/DC1
Materials and Methods
SOM Text
Figs. S1 and S2
Tables S1 to S4

References

14 February 2005; accepted 25 July 2005
10.1126/science.1111063

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 electro

†To whom correspondence should be addressed.
E-mail: shendure@alumni.princeton.edu (J.S.), gregory_porreca@student.hms.harvard.edu (G.J.P.);
‡These authors contributed equally to this work.

1Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. 2Department of Genetics.
3Howard Hughes Medical Institute, Washington University, St. Louis, MO 63110, USA.

1 These works were supported by NIH grants AR46125 and AR83404. Some aspects of the project were supported by Office of Naval Research grant N000140310568 and a grant from the University of Pennsylvania Research Foundation. The authors thank Q. Zhang, H. Hofmann, W. Megill, and A. Dunham for helpful discussions; R. Sprague, E. Maxwell, R. Essner, L. Gazit, M. Yuhas, and J. Milligan for helping with the experimentation; and F. Letterio for machining the backpacks.

14 February 2005; accepted 25 July 2005
10.1126/science.1111063

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-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.

1728
9 SEPTEMBER 2005 VOL 309 SCIENCE www.sciencemag.org

Downloaded from www.sciencemag.org on October 31, 2008
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 noninteracting amplifications within a milliliter-scale volume (15–17). Amplification products of individual compartments are captured via inclusion of 1-μm 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

![Fig. 1](https://www.sciencemag.org/content/sci/309/5738/1729/F1.large.jpg)

**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-μm beads (Note S2). (C) Hybridization to nonmagnetic, 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 S5, and fig. S7.
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× 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 × 1000) were collected. Cycle 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 reads 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 distance for a given raw base call to the median centroid of the nearest cluster.

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 (Table 1). These six mutations represent heterogeneities in lambda Red or MG1655, or errors in the

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). Three of these six mutations represent heterogeneities in lambda Red or MG1655, or errors in the

<table>
<thead>
<tr>
<th>Coverage</th>
<th>Percent of genome</th>
<th>Correctly called mock substitutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1× or greater</td>
<td>91.4%</td>
<td>86 of 87</td>
</tr>
<tr>
<td>2× or greater</td>
<td>83.3%</td>
<td>88 of 90</td>
</tr>
<tr>
<td>3× or greater</td>
<td>74.9%</td>
<td>78 of 78</td>
</tr>
<tr>
<td>4× or greater</td>
<td>66.9%</td>
<td>75 of 76</td>
</tr>
<tr>
<td>5× or greater</td>
<td>58.6%</td>
<td>67 of 67</td>
</tr>
</tbody>
</table>

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× or more reads, and 86 of these were called correctly.
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× coverage 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 Gaussian 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.

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 concatenates, 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 ligation (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 electro- phoretic 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 flowcell 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.

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.

<table>
<thead>
<tr>
<th>Position</th>
<th>Type</th>
<th>Gene</th>
<th>Context</th>
<th>Confirmation</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>986,328</td>
<td>T → G</td>
<td>ompF</td>
<td>10 region</td>
<td>Yes</td>
<td>Evolved strain only</td>
</tr>
<tr>
<td>931,955</td>
<td>8-bp deletion</td>
<td>lrp</td>
<td>Frameshift</td>
<td>Yes</td>
<td>Evolved strain only</td>
</tr>
<tr>
<td>985,791</td>
<td>T → G</td>
<td>ompF</td>
<td>Glu → Ala</td>
<td>Yes</td>
<td>Evolved strain only</td>
</tr>
<tr>
<td>1,976,527 − 1,977,302</td>
<td>776-bp deletion</td>
<td>flhD</td>
<td>Promoter</td>
<td>Yes</td>
<td>MG1655 heterogeneity</td>
</tr>
<tr>
<td>3,957,960</td>
<td>C → T</td>
<td>ppiC</td>
<td>5’ UTR</td>
<td>Yes</td>
<td>MG1655 heterogeneity</td>
</tr>
<tr>
<td>λ-red, 3274</td>
<td>T → C</td>
<td>ORF61</td>
<td>Lys → Gly</td>
<td>Yes</td>
<td>λ-red heterogeneity</td>
</tr>
<tr>
<td>λ-red, 9846</td>
<td>T → C</td>
<td>cl</td>
<td>Glu → Glu</td>
<td>Yes</td>
<td>λ-red heterogeneity</td>
</tr>
</tbody>
</table>
We demonstrate low costs of sequencing, mate-paired reads, high multiplexities, 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-μm beads can potentially be fit in the area of a standard microscope slide (fig. S6). We achieve raw data acquisition rates of ~400 bps, 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 information per bit of 10,000 bits collected is a benchmark that can be expected. Nevertheless, PUMA (p53–up-regulated modulator of apoptosis) (8–10), which targets mitochondria, caspases are activated, and apoptosis rapidly ensues (11). Thus, p53 possesses a proapoptotic function that is independent of its transcriptional activity (12–15).

If p53 directly engages MOMP in cooperation with BAX, no further requirement for additional proapoptotic Bcl-2 proteins would be expected. Nevertheless, PUMA (p53–upregulated modulator of apoptosis), a proapoptotic BH3-only protein, is a direct transcriptional target of p53. Furthermore, mice deficient in Puma are resistant to p53-dependent, DNA damage–induced apoptosis even though p53 is stabilized and accumulates in the cytoplasm (6, 16–18). A better understanding of the distinct nuclear and cytoplasmic proapoptotic functions of p53 may reveal strategies for the prevention and treatment of cancer.

**PUMA Coupled the Nuclear and Cytoplasmic Proapoptotic Function of p53**

**Jerry E. Chipuk,1* Lisa Bouchier-Hayes,1 Tomomi Kuwana,1,2 Donald D. Newmeyer,1 Douglas R. Green1‡†**

The Tp53 tumor suppressor gene product (p53) functions in the nucleus to regulate proapoptotic genes, whereas cytoplasmic p53 directly activates proapoptotic Bcl-2 proteins to permeabilize mitochondria and initiate apoptosis. Here, we demonstrate that a tripartite nexus between Bcl-xl, cytoplasmic p53, and PUMA coordinates these distinct p53 functions. After genotoxic stress, Bcl-xl sequestered cytoplasmic p53. Nuclear p53 caused expression of PUMA, which then displaced p53 from Bcl-xl, allowing p53 to induce mitochondrial permeabilization. Mutant Bcl-xl that bound p53, but not PUMA, rendered cells resistant to p53-induced apoptosis irrespective of PUMA expression. Thus, PUMA couples the nuclear and cytoplasmic proapoptotic functions of p53.

Numerous genes are regulated by p53, such as Adam-12, which encode death receptors [for example, FAS (CD95)] and proapoptotic Bcl-2 proteins (for example, BAX, BID, Noxa, and PUMA) (2–7). In parallel, p53 also accumulates in the cytoplasm, where it directly activates the proapoptotic protein BAX to promote mitochondrial outer-membrane permeabilization (MOMP) (8–10). Once MOMP occurs, proapoptotic factors (for example, cytochrome c) are released from mitochondria, caspases are activated, and apoptosis rapidly ensues (11). Thus, p53 possesses a proapoptotic function that is independent of its transcriptional activity (12–15).

If p53 directly engages MOMP in cooperation with BAX, no further requirement for p53-dependent transcriptional regulation of additional proapoptotic Bcl-2 proteins would be expected. Nevertheless, PUMA (p53–upregulated modulator of apoptosis), a proapoptotic BH3-only protein, is a direct transcriptional target of p53. Furthermore, mice deficient in Puma are resistant to p53-dependent, DNA damage–induced apoptosis even though p53 is stabilized and accumulates in the cytoplasm (6, 16–18). A better understanding of the distinct nuclear and cytoplasmic proapoptotic functions of p53 may reveal strategies for the prevention and treatment of cancer.

**Fig. 1. DNA damage–induced p53 Bcl-xl and PUMA Bcl-xl complexes.** (A) Proteins from cytosolic extracts prepared from wild-type or puma−/− MEFs treated with 5 μm/cm² UV were immunoprecipitated with an agarose-conjugated antibody to Bcl-xl, eluted, subjected to SDS-PAGE, and visualized by silver staining. Bands were excised and subjected to tryptic digestion and mass spectrometry. The asterisk (*) indicates a fragment of Bcl-xl or p53. (B) Cytosolic extracts were treated as in (A), but protein complexes were analyzed by Western blot. mIgG (mouse immunoglobulin G) is a control antibody.

---

*Division of Cellular Immunology, La Jolla Institute for Allergy and Immunology, 10555 Science Center Drive, San Diego, CA 92121, USA. †University of Iowa, Carver College of Medicine, Department of Pathology, Iowa City, IA 52242, USA.

*Present address: Department of Immunology, St. Jude Children’s Research Hospital, 332 North Lauderdale Street, Memphis, TN 38105, USA.

†To whom correspondence should be addressed. E-mail: dgreen5240@uic.edu

---

**References and Notes**

10. http://arpep.med.harvard.edu/Polonator/Polonator.htm
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. Elbing, J. Ichikawa, J. Malek, P. McEwan, K. McKernan, A. Sheridan, and D. Smith (Argencourt); S. Skiena (SUNY–Stony Brook) C. Felts (RP); R. Fincher (Alcott); D. Focht (Biotech); and M. Hofstelder 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.sciencemag.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.