3. S. Georgi, www.batteriesdigest.com/id380.htm (accessed June 2005).
4. R. M. Alexander, *J. Exp. Biol.* **160**, 55 (1991).
5. G. A. Cavagna, N. C. Heglund, C. R. Taylor, *Am. J. Physiol.* **233**, R243 (1977).
6. J. Drake, *Wired* **9**, 90 (2001).
7. S. Stanford, R. Pelrine, R. Kornbluh, Q. Pei, in *Proceedings of the 13th International Symposium on Unmanned Untethered Submersible Technology* (Autonomous Undersea Systems Institute, Lee, NH, 2003).
8. T. Starner, J. Paradiso, in *Low Power Electronics Design* (CRC Press, Boca Raton, FL, 2004), p. 45–1.
9. G. A. Cavagna, M. Kaneko, *J. Physiol.* **268**, 647 (1977).
10. S. A. Gard, S. C. Miff, A. D. Kuo, *Hum. Mov. Sci.* **22**, 597 (2004).
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 hour^{-1} with 20- and 29-kg loads and 4.8 km hour^{-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 hour^{-1} with 20- and 28-kg loads), electricity generation on the incline was actually substantially greater than that on the flat (table S1).
14. R. Margaria, *Biomechanics and Energetics of Muscular Exercise* (Clarendon, Oxford, 1976).
15. R. A. Ferguson *et al.*, *J. Physiol.* **536**, 261 (2001).
16. G. A. Cavagna, P. A. Willems, M. A. Legramandi, N. C. Heglund, *J. Exp. Biol.* **205**, 3413 (2002).
17. A. Grabowski, C. T. Farley, R. Kram, *J. Appl. Physiol.* **98**, 579 (2005).
18. J. M. Donelan, R. Kram, A. D. Kuo, *J. Exp. Biol.* **205**, 3717 (2002).
19. J. M. Donelan, R. Kram, A. D. Kuo, *J. Biomech.* **35**, 117 (2002).
20. J. S. Gottschall, R. Kram, *J. Appl. Physiol.* **94**, 1766 (2003).
21. Because this savings in metabolic energy represents only 6% of the net energetic cost of walking with the backpack (492 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*).
22. K. Schmidt-Nielsen, *Animal Physiology: Adaptation and Environment* (Cambridge Univ. Press, Cambridge, ed. 3, 1988).
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.
25. R. Kram, *J. Appl. Physiol.* **71**, 1119 (1991).
26. A. E. Minetti, *J. Exp. Biol.* **207**, 1265 (2004).
27. A. A. Biewener, C. T. Farley, T. J. Roberts, M. Temaner, *J. Appl. Physiol.* **97**, 2266 (2004).
28. T. M. Griffin, T. J. Roberts, R. Kram, *J. Appl. Physiol.* **95**, 172 (2003).

29. This work was supported by NIH grants AR46125 and AR38404. 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. Church[1]

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-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.

[1]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. [2]Department of Genetics, [3]Howard 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.)

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 IIs 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 in situ polonies (*6*) was easily applied to ePCR beads, resulting in a ~1.5-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 enyzmatic 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 × 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.

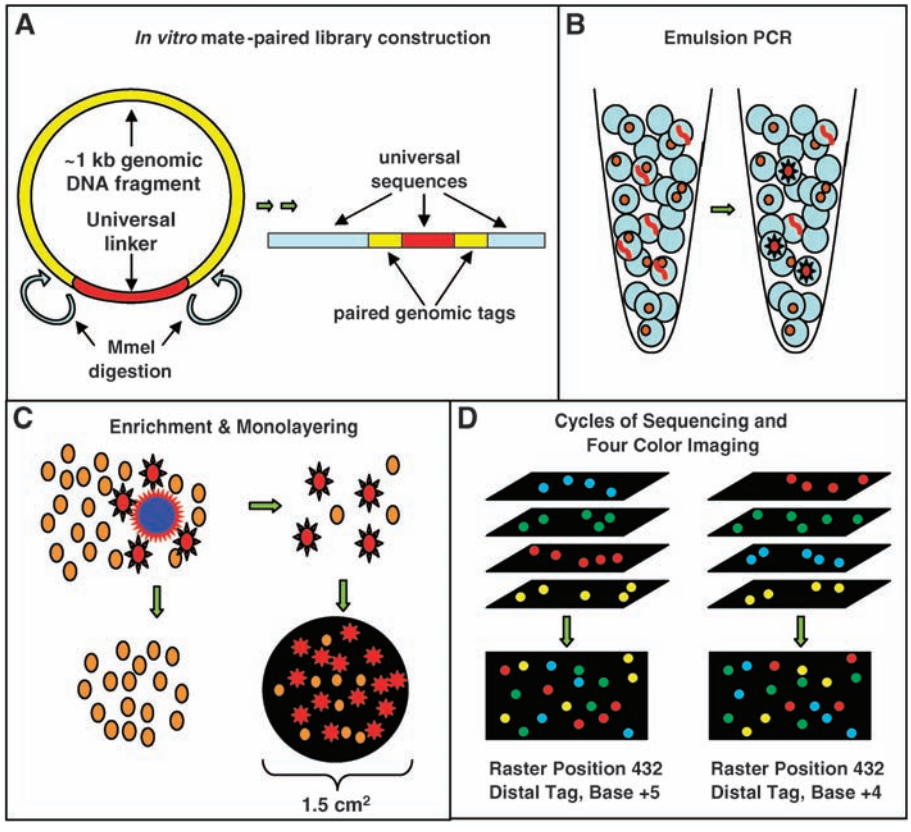

Fig. 1. A multiplex approach to genome sequencing. (**A**) Sheared, size-selected genomic fragments (yellow) are circularized with a linker (red) bearing Mme I recognition sites (Note S1). Subsequent steps, which include a rolling circle amplification, yield the 134- to 136-bp mate-paired library molecules 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 S6, 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_v^- 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 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 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 locations 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). Three of these six mutations represent heterogeneities in lambda Red or MG1655, or errors in the

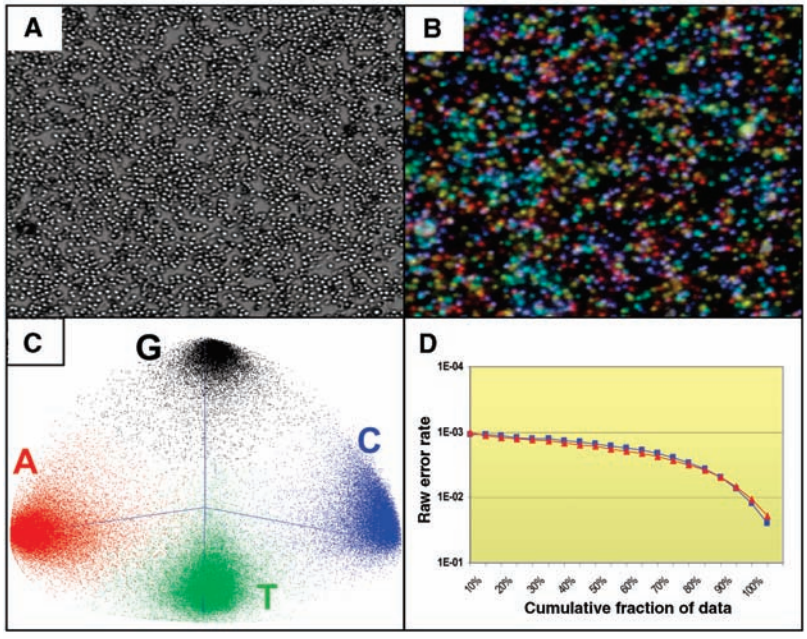

Fig. 2. Raw data acquisition and base calling. (**A**) Brightfield images (area shown corresponds to 0.01% of the total gel area) facilitate object segmentation by simple thresholding, allowing resolution even when multiple 1-μm beads are in contact. (**B**) False-color depiction of four fluorescence 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 $Q20 = 1 \times 10^{-2}$, $Q30 = 1 \times 10^{-3}$ {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× or more reads, and 86 of these were called correctly.

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

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 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*).

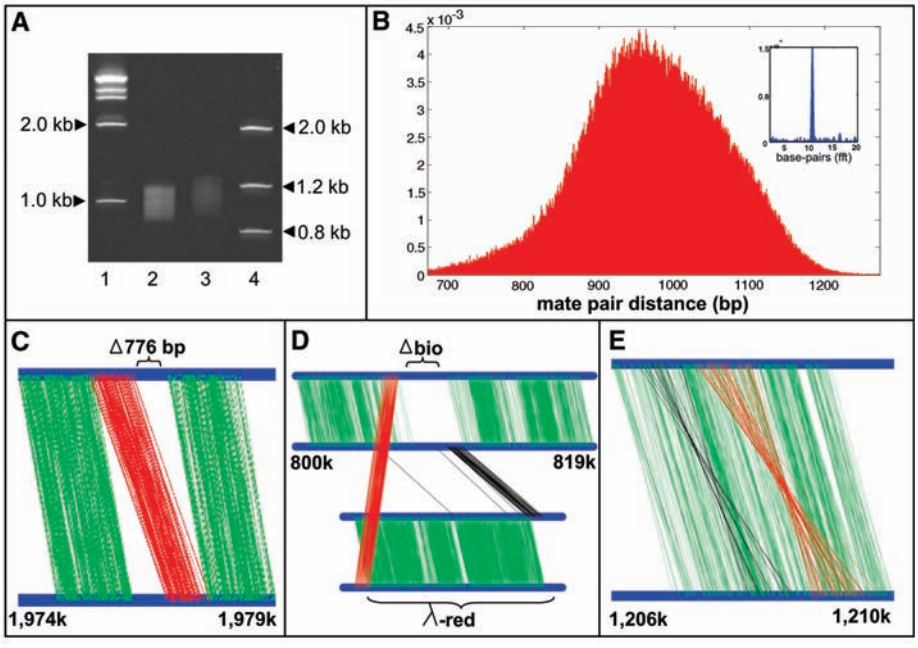

Fig. 3. Mate-paired tags and rearrangement discovery. (**A**) Diagnostic 6% polyacrylamide gel of the sheared, size-selected genomic DNA from which the library was constructed. Lanes 1 and 4 are molecular size markers. Lane 2 represents the material used in the library sequenced to generate the paired-tag 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.

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	*ompF*	−10 region	Yes	Evolved strain only
931,955	8-bp deletion	*lrp*	Frameshift	Yes	Evolved strain only
985,791	T → G	*ompF*	Glu → Ala	Yes	Evolved strain only
1,976,527−1,977,302	776-bp deletion	*flhD*	Promoter	Yes	MG1655 heterogeneity
3,957,960	C → T	*ppiC*	5′ UTR	Yes	MG1655 heterogeneity
λ-red, 3274	T → C	*ORF61*	Lys → Gly	Yes	λ-red heterogeneity
λ-red, 9846	T → C	*cI*	Glu → Glu	Yes	λ-red heterogeneity

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 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.

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-μm 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.

References and Notes

1. F. Sanger *et al.*, *Nature* **265**, 687 (1977).
2. F. S. Collins, M. Morgan, A. Patrinos, *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).
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 (Bioptechs); 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.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.

PUMA Couples 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. Green[1]*†

The *Trp53* 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.

The antineoplastic function of p53 occurs primarily through the induction of apoptosis (*1*). p53 undergoes posttranslational modification in response to oncogene-activated signaling pathways or to genotoxic stress; this allows stabilization of p53, which accumulates in the nucleus and regulates target gene expression. Numerous genes are regulated by p53, such as those encoding 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, proapoptogenic 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 modifier 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.

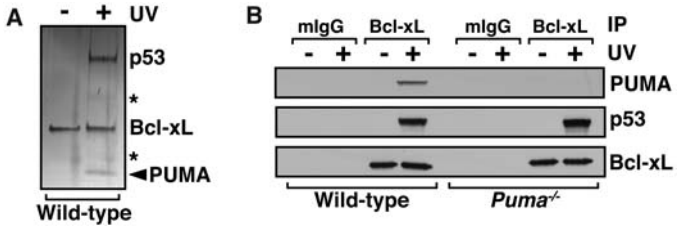

Fig. 1. DNA damage–induced p53·Bcl-xL and PUMA·Bcl-xL complexes. (**A**) Proteins from cytosolic extracts prepared from wild-type or *Puma*$^{-/-}$ MEFs treated with 5 mJ/cm^2 UV were immunoprecipitated with an agarose-conjugated antibody to Bcl-xL, eluted, subjected to SDS-PAGE, and visualized by silver 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.

[1]Division of Cellular Immunology, La Jolla Institute for Allergy and Immunology, 10355 Science Center Drive, San Diego, CA 92121, USA. [2]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@aol.com