ARTICLES

Genome sequencing in microfabricated high-density picolitre reactors

Marcel Margulies[1]*, Michael Egholm[1]*, William E. Altman[1], Said Attiya[1], Joel S. Bader[1], Lisa A. Bemben[1], Jan Berka[1], Michael S. Braverman[1], Yi-Ju Chen[1], Zhoutao Chen[1], Scott B. Dewell[1], Lei Du[1], Joseph M. Fierro[1], Xavier V. Gomes[1], Brian C. Godwin[1], Wen He[1], Scott Helgesen[1], Chun He Ho[1], Gerard P. Irzyk[1], Szilveszter C. Jando[1], Maria L. I. Alenquer[1], Thomas P. Jarvie[1], Kshama B. Jirage[1], Jong-Bum Kim[1], James R. Knight[1], Janna R. Lanza[1], John H. Leamon[1], Steven M. Lefkowitz[1], Ming Lei[1], Jing Li[1], Kenton L. Lohman[1], Hong Lu[1], Vinod B. Makhijani[1], Keith E. McDade[1], Michael P. McKenna[1], Eugene W. Myers[2], Elizabeth Nickerson[1], John R. Nobile[1], Ramona Plant[1], Bernard P. Puc[1], Michael T. Ronan[1], George T. Roth[1], Gary J. Sarkis[1], Jan Fredrik Simons[1], John W. Simpson[1], Maithreyan Srinivasan[1], Karrie R. Tartaro[1], Alexander Tomasz[3], Kari A. Vogt[1], Greg A. Volkmer[1], Shally H. Wang[1], Yong Wang[1], Michael P. Weiner[4], Pengguang Yu[1], Richard F. Begley[1] & Jonathan M. Rothberg[1]

The proliferation of large-scale DNA-sequencing projects in recent years has driven a search for alternative methods to reduce time and cost. Here we describe a scalable, highly parallel sequencing system with raw throughput significantly greater than that of state-of-the-art capillary electrophoresis instruments. The apparatus uses a novel fibre-optic slide of individual wells and is able to sequence 25 million bases, at 99% or better accuracy, in one four-hour run. To achieve an approximately 100-fold increase in throughput over current Sanger sequencing technology, we have developed an emulsion method for DNA amplification and an instrument for sequencing by synthesis using a pyrosequencing protocol optimized for solid support and picolitre-scale volumes. Here we show the utility, throughput, accuracy and robustness of this system by shotgun sequencing and *de novo* assembly of the *Mycoplasma genitalium* genome with 96% coverage at 99.96% accuracy in one run of the machine.

DNA sequencing has markedly changed the nature of biomedical research and medicine. Reductions in the cost, complexity and time required to sequence large amounts of DNA, including improvements in the ability to sequence bacterial and eukaryotic genomes, will have significant scientific, economic and cultural impact. Large-scale sequencing projects, including whole-genome sequencing, have usually required the cloning of DNA fragments into bacterial vectors, amplification and purification of individual templates, followed by Sanger sequencing[1] using fluorescent chain-terminating nucleotide analogues[2] and either slab gel or capillary electrophoresis. Current estimates put the cost of sequencing a human genome between $10 million and $25 million[3]. Alternative sequencing methods have been described[4–8]; however, no technology has displaced the use of bacterial vectors and Sanger sequencing as the main generators of sequence information.

Here we describe an integrated system whose throughput routinely enables applications requiring millions of bases of sequence information, including whole-genome sequencing. Our focus has been on the co-development of an emulsion-based method[9–11] to isolate and amplify DNA fragments *in vitro*, and of a fabricated substrate and instrument that performs pyrophosphate-based sequencing (pyrosequencing[5,12]) in picolitre-sized wells.

In a typical run we generate over 25 million bases with a Phred quality score of 20 or better (predicted to have an accuracy of 99% or higher). Although this Phred 20 quality throughput is significantly higher than that of Sanger sequencing by capillary electrophoresis, it is currently at the cost of substantially shorter reads and lower average individual read accuracy. Sanger-based capillary electrophoresis sequencing systems produce up to 700 bases of sequence information from each of 96 DNA templates at an average read accuracy of 99.4% in 1 h, or 67,000 bases per hour, with substantially all of the bases having Phred 20 or better quality[23]. We further characterize the performance of the system and demonstrate that it is possible to assemble bacterial genomes *de novo* from relatively short reads by sequencing a known bacterial genome, *Mycoplasma genitalium* (580,069 bases), and comparing our shotgun sequencing and *de novo* assembly with the results originally obtained for this genome[13]. The results of shotgun sequencing and *de novo* assembly of a larger bacterial genome, that of *Streptococcus pneumoniae*[14] (2.1 megabases (Mb)), are presented in Supplementary Table 4.

Emulsion-based sample preparation

We generate random libraries of DNA fragments by shearing an entire genome and isolating single DNA molecules by limiting dilution (Supplementary Methods). Specifically, we randomly fragment the entire genome, add specialized common adapters to the fragments, capture the individual fragments on their own beads and, within the droplets of an emulsion, clonally amplify the individual fragment (Fig. 1a, b). Unlike in current sequencing technology, our approach does not require subcloning in bacteria or the handling of

[1]454 Life Sciences Corp., 20 Commercial Street, Branford, Connecticut 06405, USA. [2]University of California, Berkeley, California 94720, USA. [3]Laboratory of Microbiology, The Rockefeller University, New York, New York 10021, USA. [4]The Rothberg Institute for Childhood Diseases, 530 Whitfield Street, Guilford, Connecticut 06437, USA.
*These authors contributed equally to this work.

IDT Exhibit 1070.0001

© 2005 Nature Publishing Group

individual clones; the templates are handled in bulk within the emulsions[9–11].

Sequencing in fabricated picolitre-sized reaction vessels

We perform sequencing by synthesis simultaneously in open wells of a fibre-optic slide using a modified pyrosequencing protocol that is designed to take advantage of the small scale of the wells. The fibre-optic slides are manufactured by slicing of a fibre-optic block that is obtained by repeated drawing and fusing of optic fibres. At each iteration, the diameters of the individual fibres decrease as they are hexagonally packed into bundles of increasing cross-sectional sizes. Each fibre-optic core is 44 μm in diameter and surrounded by 2–3 μm of cladding; etching of each core creates reaction wells approximately 55 μm in depth with a centre-to-centre distance of 50 μm (Fig. 1c), resulting in a calculated well size of 75 pl and a well density of 480 wells mm^{-2}. The slide, containing approximately 1.6 million wells[15], is loaded with beads and mounted in a flow chamber designed to create a 300-μm high channel, above the well openings, through which the sequencing reagents flow (Fig. 2a, b). The unetched base of the slide is in optical contact with a second fibre-optic imaging bundle bonded to a charge-coupled device (CCD) sensor, allowing the capture of emitted photons from the bottom of each individual well (Fig. 2c; see also Supplementary Methods).

We developed a three-bead system, and optimized the components to achieve high efficiency on solid support. The combination of picolitre-sized wells, enzyme loading uniformity allowed by the small beads and enhanced solid support chemistry enabled us to develop a method that extends the useful read length of sequencing-by-synthesis to 100 bases (Supplementary Methods).

In the flow chamber cyclically delivered reagents flow perpendicularly to the wells. This configuration allows simultaneous extension reactions on template-carrying beads within the open wells and relies on convective and diffusive transport to control the addition or removal of reagents and by-products. The timescale for diffusion into and out of the wells is on the order of 10 s in the current configuration and is dependent on well depth and flow channel height. The timescales for the signal-generating enzymatic reactions are on the order of 0.02–1.5 s (Supplementary Methods). The current reaction is dominated by mass transport effects, and improvements based on faster delivery of reagents are possible. Well depth was selected on the basis of a number of competing requirements: (1) wells need to be deep enough for the DNA-carrying beads to remain in the wells in the presence of convective transport past the wells; (2) they must be sufficiently deep to provide adequate isolation against diffusion of by-products from a well in which incorporation is taking place to a well where no incorporation is occurring; and (3) they must be shallow enough to allow rapid diffusion of nucleotides into the wells and rapid washing out of remaining nucleotides at the end of each flow cycle to enable high sequencing throughput and reduced reagent use. After the flow of each nucleotide, a wash containing apyrase is used to ensure that nucleotides do not remain in any well before the next nucleotide being introduced.

Base calling of individual reads

Nucleotide incorporation is detected by the associated release of inorganic pyrophosphate and the generation of photons[5,12]. Wells containing template-carrying beads are identified by detecting a known four-nucleotide 'key' sequence at the beginning of the read

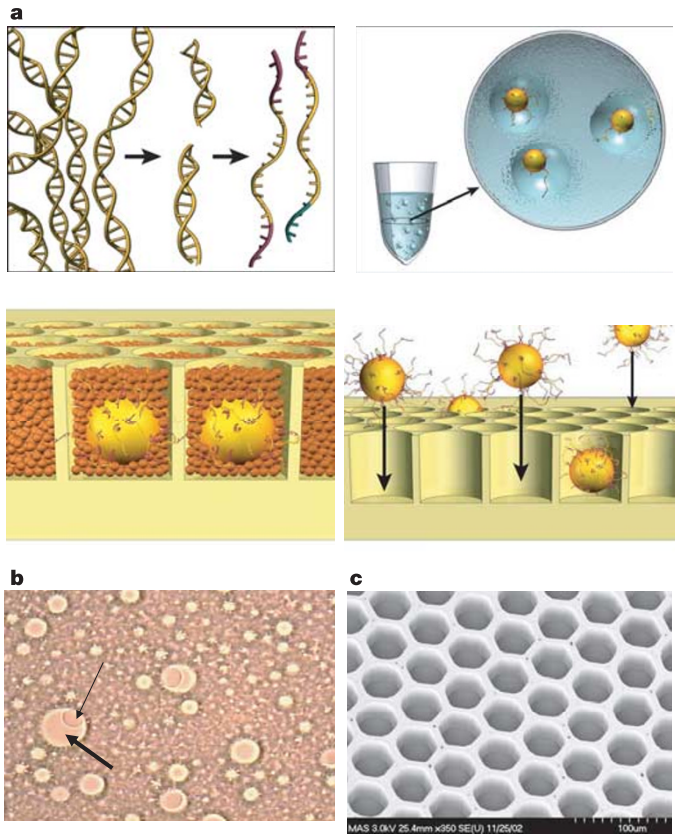


Figure 1 | Sample preparation. **a**, Genomic DNA is isolated, fragmented, ligated to adapters and separated into single strands (top left). Fragments are bound to beads under conditions that favour one fragment per bead, the beads are captured in the droplets of a PCR-reaction-mixture-in-oil emulsion and PCR amplification occurs within each droplet, resulting in beads each carrying ten million copies of a unique DNA template (top right). The emulsion is broken, the DNA strands are denatured, and beads carrying single-stranded DNA clones are deposited into wells of a fibre-optic slide (bottom right). Smaller beads carrying immobilized enzymes required for pyrophosphate sequencing are deposited into each well (bottom left). **b**, Microscope photograph of emulsion showing droplets containing a bead and empty droplets. The thin arrow points to a 28-μm bead; the thick arrow points to an approximately 100-μm droplet. **c**, Scanning electron micrograph of a portion of a fibre-optic slide, showing fibre-optic cladding and wells before bead deposition.

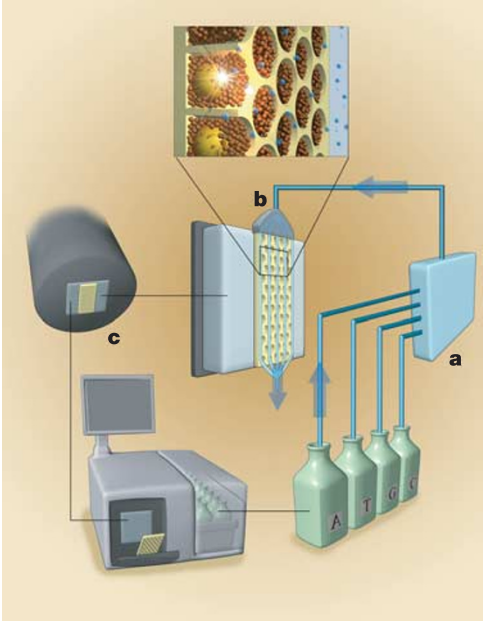


Figure 2 | Sequencing instrument. The sequencing instrument consists of the following major subsystems: a fluidic assembly (**a**), a flow chamber that includes the well-containing fibre-optic slide (**b**), a CCD camera-based imaging assembly (**c**), and a computer that provides the necessary user interface and instrument control.

IDT Exhibit 1070.0002

© 2005 Nature Publishing Group

(Supplementary Methods). Raw signals are background-subtracted, normalized and corrected. The normalized signal intensity at each nucleotide flow, for a particular well, indicates the number of nucleotides, if any, that were incorporated. This linearity in signal is preserved to at least homopolymers of length eight (Supplementary Fig. 6). In sequencing by synthesis a very small number of templates on each bead lose synchronism (that is, either get ahead of, or fall behind, all other templates in sequence[16]). The effect is primarily due to leftover nucleotides in a well (creating 'carry forward') or to incomplete extension. Typically, we observe a carry forward rate of 1–2% and an incomplete extension rate of 0.1–0.3%. Correction of these shifts is essential because the loss of synchronism is a cumulative effect that degrades the quality of sequencing at longer read lengths. We have developed algorithms, based on detailed models of the underlying physical phenomena, that allow us to determine, and correct for, the amounts of carry forward and incomplete extension occurring in individual wells (Supplementary Methods). Figure 3 shows the processed result, a 113-bases-long read generated in the *M. genitalium* run discussed below. To assess sequencing performance and the effectiveness of the correction algorithms, independently of artefacts introduced during the emulsion-based sample preparation, we created test fragments with difficult-to-sequence stretches of identical bases of increasing length (homopolymers) (Supplementary Methods and Supplementary Fig. 4). Using these test fragments, we have verified that at the individual read level we achieve base call accuracy of approximately 99.4%, at read lengths in excess of 100 bases (Table 1).

High-quality reads and consensus accuracy

Before base calling or aligning reads, we select high-quality reads without relying on a priori knowledge of the genome or template being sequenced (Supplementary Methods). This selection is based on the observation that poor-quality reads have a high proportion of signals that do not allow a clear distinction between a flow during which no nucleotide was incorporated and a flow during which one or more nucleotide was incorporated. When base calling individual reads, errors can occur because of signals that have ambiguous values (Supplementary Fig. 5). To improve the usability of our reads, we also developed a metric that allows us to estimate *ab initio* the quality (or probability of correct base call) of each base of a read, analogous to the Phred score[17] used by current Sanger sequencers (Supplementary Methods and Supplementary Fig. 8).

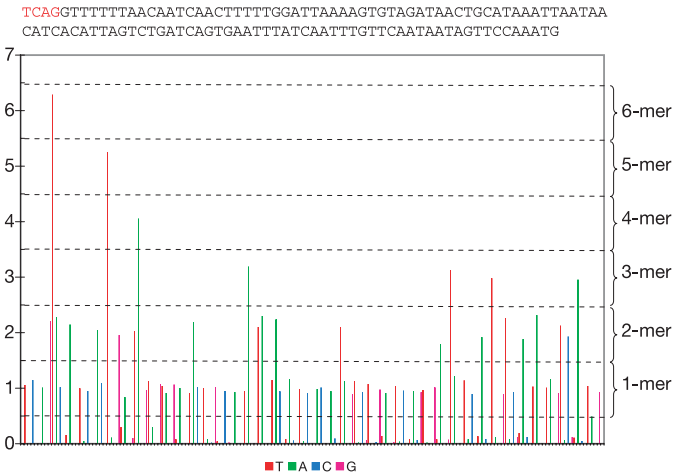


Figure 3 | Flowgram of a 113-bases read from an *M. genitalium* run. Nucleotides are flowed in the order T, A, C, G. The sequence is shown above the flowgram. The signal value intervals corresponding to the various homopolymers are indicated on the right. The first four bases (in red, above the flowgram) constitute the 'key' sequence, used to identify wells containing a DNA-carrying bead.

Higher quality sequence can be achieved by taking advantage of the high over sampling that our system affords and building a consensus sequence. Sequences are aligned to one another using the signal strengths at each nucleotide flow, rather than individual base calls, to determine optimal alignment (Supplementary Methods). The corresponding signals are then averaged, after which base calling is performed. This approach greatly improves the accuracy of the sequence (Supplementary Fig. 7) and provides an estimate of the quality of the consensus base. We refer to that quality measure as the *Z*-score—it is a measure of the spread of signals in all the reads at one location and the distance between the average signal and the closest base-calling threshold value. In both re-sequencing and *de novo* sequencing, as the minimum *Z*-score is raised the consensus accuracy increases, while coverage decreases; approximately half of the excluded bases, as the *Z*-score is increased, belong to homopolymers of length four and larger. Sanger sequencers usually require a depth of coverage at any base of three or more in order to achieve a consensus accuracy of 99.99%. To achieve a minimum of threefold coverage of 95% of the unique portions of a typical genome requires approximately seven- to eightfold over sampling. Owing to our higher error rate, we have observed that comparable consensus accuracies, over a similar fraction of a genome, are achieved with a depth of coverage of four or more, requiring approximately ten to twelve times over sampling.

Mycoplasma genitalium

Mycoplasma genomic DNA was fragmented and prepared into a sequencing library as described above. (This was accomplished by a single individual in 4 h.) After emulsion polymerase chain reaction (PCR) and bead deposition onto a $60 \times 60\,mm^2$ fibre-optic slide, a process which took one individual 6 h, 42 cycles of four nucleotides were flowed through the sequencing system in an automated 4-h run of the instrument. The results are summarized in Table 2. In order to measure the quality of individual reads, we aligned each high quality read to the reference genome at 70% stringency using flow-space mapping and criteria similar to those used previously in assessing the accuracy of other base callers[17]. When assessing sequencing quality, only reads that mapped to unique locations in the reference genome were included. Because this process excludes repeat regions (parts of the genome for which corresponding flowgrams are 70% similar to one another), the selected reads did not cover the genome completely. Figure 4a illustrates the distribution of read lengths for this run. The average read length was 110 bases, the resulting over sample 40-fold, and 84,011 reads (27.4%) were perfect. Figure 4b summarizes the average error as a function of base position. Coverage of non-repeat regions was consistent with the sample preparation and emulsion not being biased (Supplementary Fig. 8). At the individual read level, we observe an insertion and deletion error rate of approximately 3.3%; substitution errors have a much lower rate, on the order of 0.5%. When using these reads without any *Z*-score restriction, we covered 99.94% of the genome in ten contiguous regions with a consensus accuracy of 99.97%. The error rate in homopolymers is significantly reduced in the consensus sequence (Supplementary Fig. 7). Of the bases not covered by this consensus

Table 1 | Summary of sequencing statistics for test fragments

Size of fibre-optic slide	$60 \times 60\,mm^2$
Run time/number of cycles	243 min/42
Test fragment reads	497,893
Average read length (bases)	108
Number of bases in test fragments	53,705,267
Bases with a Phred score of 20 and above	47,181,792
Individual read insertion error rate	0.44%
Individual read deletion error rate	0.15%
Individual read substitution error rate	0.004%
All errors	0.60%

IDT Exhibit 1070.0003

© 2005 Nature Publishing Group

sequence (366 bases), all belonged to excluded repeat regions. Setting a minimum *Z*-score equal to 4, coverage was reduced to 98.1% of the genome, while consensus accuracy increased to 99.996%. We further demonstrated the reproducibility of the system by repeating the whole-genome sequencing of *M. genitalium* an additional eight times, achieving a 40-fold coverage of the genome in each of the eight separate instrument runs (Supplementary Table 3).

We assembled the *M. genitalium* reads from a single run into 25 contigs with an average length of 22.4 kb. One of these contigs was misassembled due to a collapsed tandem repeat region of 60 bases, and was corrected by hand. The original sequencing of *M. genitalium* resulted in 28 contigs before directed sequencing used for finishing the sequence[13]. Our assembly covered 96.54% of the genome and attained a consensus accuracy of 99.96%. Non-resolvable repeat regions amount to 3% of the genome: we therefore covered 99.5% of the unique portions of the genome. Sixteen of the breaks between contigs were due to non-resolvable repeat regions, two were due to missed overlapping reads (our read filter and trimmer are not perfect and the algorithms we use to perform the pattern matching of flowgrams occasionally miss valid overlaps) and the remainder to thin read coverage. Setting a minimum *Z*-score of 4, coverage was reduced to 95.27% of the genome (98.2% of the resolvable part of the genome) with the consensus accuracy increasing to 99.994%.

Discussion

We have demonstrated the simultaneous acquisition of hundreds of thousands of sequence reads, 80–120 bases long, at 96% average accuracy in a single run of the instrument using a newly developed *in vitro* sample preparation methodology and sequencing technology. With Phred 20 as a cutoff, we show that our instrument is able to produce over 47 million bases from test fragments and 25 million bases from genomic libraries. We used test fragments to de-couple our sample preparation methodology from our sequencing technology. The decrease in single-read accuracy from 99.4% for test fragments to 96% for genomic libraries is primarily due to a lack of clonality in a fraction of the genomic templates in the emulsion, and is not an inherent limitation of the sequencing technology. Most of the remaining errors result from a broadening of signal distributions, particularly for large homopolymers (seven or more), leading to ambiguous base calls. Recent work on the sequencing chemistry and algorithms that correct for crosstalk between wells suggests that the signal distributions will narrow, with an attendant reduction in errors and increase in read lengths. In preliminary experiments with genomic libraries that also include improvements in the emulsion protocol, we are able to achieve, using 84 cycles, read lengths of 200 bases with accuracies similar to those demonstrated here for 100 bases. On occasion, at 168 cycles, we have generated individual reads that are 100% accurate over greater then 400 bases.

Using *M. genitalium* we demonstrate that short fragments a priori do not prohibit the *de novo* assembly of bacterial genomes. In fact, the larger over sampling afforded by the throughput of our system resulted in a draft sequence having fewer contigs than with Sanger reads, with substantially less effort. By taking advantage of the over sampling, consensus accuracies greater then 99.96% were achieved for this genome. Further quality filtering of the assembly results in the selection of a consensus sequence with accuracy exceeding 99.99% while incurring only a minor loss of genome coverage. Comparable results were seen when we shotgun sequenced and *de novo* assembled the 2.1-Mb genome of *Streptococcus pneumoniae*[14] (Supplementary Table 4). The *de novo* assembly of genomes more complex than bacteria, including mammalian genomes, may require the development of methods similar to those developed for Sanger sequencing, to prepare and sequence paired end libraries that can span repeats in these genome. To facilitate the use of paired end libraries we have developed methods to sequence, in an individual well, from both ends of genomic template, and plan to add paired end read capabilities to our assembler (Supplementary Methods).

Future increases in throughput, and a concomitant reduction in cost per base, may come from the continued miniaturization of the fibre-optic reactors, allowing more sequence to be produced per unit area—a scaling characteristic similar to that which enabled the

Table 2 | Summary statistics for *M. genitalium*

Sequencing summary	
Number of instrument runs	1
Size of fibre-optic slide	60 × 60 mm^2
Run time/number of cycles	243 min/42
High quality reads	306,178
Average read length (bases)	110
Number of bases in high quality reads	33,655,553
Bases with a Phred score of 20 and above	26,753,540
Re-sequencing	
Reads mapped to single locations	238,066
Number of bases in mapped reads	27,687,747
Individual read insertion error rate	1.67%
Individual read deletion error rate	1.60%
Individual read substitution error rate	0.68%
Re-sequencing consensus	
Average over sampling	× 40
Coverage, all ($Z \geq 4$)	99.9% (98.2%)
Consensus accuracy, all ($Z \geq 4$)	99.97% (99.996%)
Consensus insertion error rate, all ($Z \geq 4$)	0.02% (0.003%)
Consensus deletion error rate, all ($Z \geq 4$)	0.01% (0.002%)
Consensus substitution error rate, all ($Z \geq 4$)	0.001% (0.0003%)
Number of contigs	10
De novo assembly	
Coverage, all ($Z \geq 4$)	96.54% (95.27%)
Consensus accuracy, all ($Z \geq 4$)	99.96% (99.994%)
Number of contigs	25
Average contig size (kb)	22.4

The individual read error rates are referenced to the total number of bases in mapped reads.

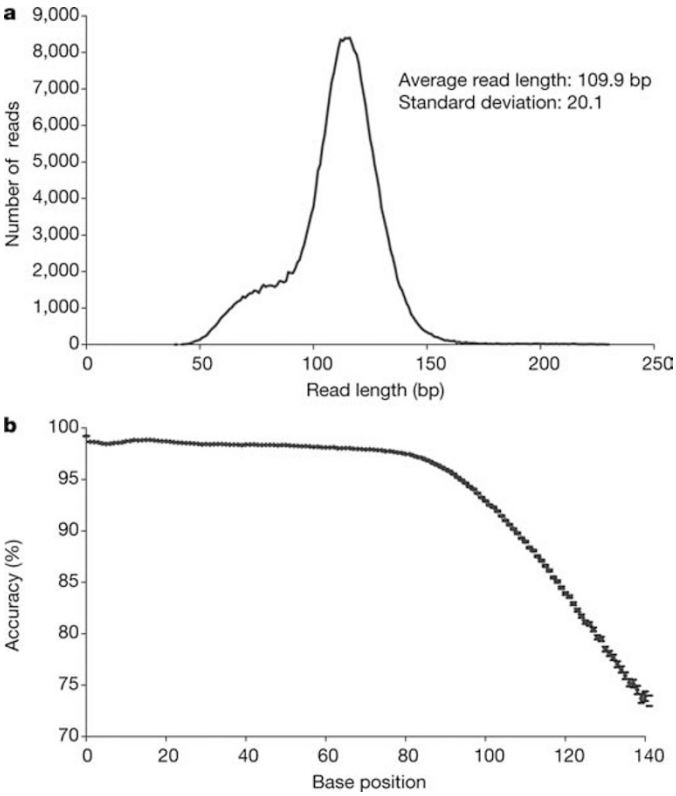


Figure 4 | *M. genitalium* data. a, Read length distribution for the 306,178 high-quality reads of the *M. genitalium* sequencing run. This distribution reflects the base composition of individual sequencing templates. **b**, Average read accuracy, at the single read level, as a function of base position for the 238,066 mapped reads of the same run.

© 2005 Nature Publishing Group

prediction of significant improvements in the integrated circuit at the start of its development cycle[18].

METHODS

Emulsion-based clonal amplification. The simultaneous amplification of fragments is achieved by isolating individual DNA-carrying beads in separate ~100-μm aqueous droplets (on the order of 2×10^6 ml^{-1}) made through the creation of a PCR-reaction-mixture-in-oil emulsion. (Fig. 1b; see also Supplementary Methods). The droplets act as separate microreactors in which parallel DNA amplifications are performed, yielding approximately 10^7 copies of a template per bead; 800 μl of emulsion containing 1.5 million beads are prepared in a standard 2-ml tube. Each emulsion is aliquoted into eight PCR tubes for amplification. After PCR, the emulsion is broken to release the beads, which include beads with amplified, immobilized DNA template and empty beads (Supplementary Methods). We then enrich for template-carrying beads (Supplementary Methods). Typically, about 30% of the beads will have DNA, producing 450,000 template-carrying beads per emulsion reaction. The number of emulsions prepared depends on the size of the genome and the expected number of runs required to achieve adequate over sampling. The 580-kb *M. genitalium* genome, sequenced on one 60×60 mm^2 fibre-optic slide, required 1.6 ml of emulsion. A human genome, over sampled ten times, would require approximately 3,000 ml of emulsion.

Bead loading into picolitre wells. The enriched template-carrying beads are deposited by centrifugation into open wells (Fig. 1c), arranged along one face of a 60×60 mm^2 fibre-optic slide. The beads (diameter ~28 μm) are sized to ensure that no more than one bead fits in most wells (we observed that 2–5% of filled wells contain more than one bead). Loading 450,000 beads (from one emulsion preparation) onto each half of a 60×60 mm^2 plate was experimentally found to limit bead occupancy to approximately 35% of all wells, thereby reducing chemical and optical crosstalk between wells. A mixture of smaller beads that carry immobilized ATP sulphurylase and luciferase necessary to generate light from free pyrophosphate are also loaded into the wells to create the individual sequencing reactors (Supplementary Methods).

Image capture. A bead carrying 10 million copies of a template yields approximately 10,000 photons at the CCD sensor, per incorporated nucleotide. The generated light is transmitted through the base of the fibre-optic slide and detected by a large format CCD ($4,095 \times 4,096$ pixels). The images are processed to yield sequence information simultaneously for all wells containing template-carrying beads. The imaging system was designed to accommodate a large number of small wells and the large number of optical signals being generated from individual wells during each nucleotide flow. Once mounted, the fibre-optic slide's position does not shift; this makes it possible for the image analysis software to determine the location of each well (whether or not it contains a DNA-carrying bead), based on light generation during the flow of a pyrophosphate solution, which precedes each sequencing run. A single well is imaged by approximately nine 15 μm pixels. For each nucleotide flow, the light intensities collected by the pixels covering a particular well are summed to generate a signal for that particular well at that particular nucleotide flow. Each image captured by the CCD produces 32 megabytes of data. In order to perform all of the necessary signal processing in real time, the control computer is fitted with an accessory board (Supplementary Methods), hosting a 6 million gate Field Programmable Gate Array (FPGA)[19,20].

***De novo* shotgun sequence assembler.** A *de novo* flow-space assembler was developed to capture all of the information contained in the original flow-based signal trace. It also addresses the fact that existing assemblers are not optimized for 80–120-bases reads, particularly with respect to memory management due to the increased number of sequencing reads needed to achieve equivalent genome coverage. (A completely random genome covered with 100-bases reads requires approximately 50% more reads to yield the same number of contiguous regions (contigs) as achieved with 700-bases reads, assuming the need for a 30-bases overlap between reads[21].) This assembler consists of a series of modules: the Overlapper, which finds and creates overlaps between reads; the Unitigger, which constructs larger contigs of overlapping sequence reads; and the Multialigner, which generates consensus calls and quality scores for the bases within each contig (Supplementary Methods). (The names of the software modules are based on those performing related functions in other assemblers developed previously[22].)

Received 6 May; accepted 10 June 2005.
Published online 31 July 2005.

1. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA* **74**, 5463–5467 (1977).
2. Prober, J. M. *et al.* A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science* **238**, 336–341 (1987).
3. NIH News Release. NHGRI seeks next generation of sequencing technologies. 14 October 2004 ⟨http://www.genome.gov/12513210⟩.
4. Nyren, P., Pettersson, B. & Uhlen, M. Solid phase DNA minisequencing by an enzymatic luminometric inorganic pyrophosphate detection assay. *Anal. Biochem.* **208**, 171–175 (1993).
5. Ronaghi, M. *et al.* Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* **242**, 84–89 (1996).
6. Jacobson, K. B. *et al.* Applications of mass spectrometry to DNA sequencing. *GATA* **8**, 223–229 (1991).
7. Bains, W. & Smith, G. C. A novel method for nucleic acid sequence determination. *J. Theor. Biol.* **135**, 303–307 (1988).
8. Jett, J. H. *et al.* High-speed DNA sequencing: an approach based upon fluorescence detection of single molecules. *Biomol. Struct. Dynam.* **7**, 301–309 (1989).
9. Tawfik, D. S. & Griffiths, A. D. Man-made cell-like compartments for molecular evolution. *Nature Biotechnol.* **16**, 652–656 (1998).
10. Ghadessy, F. J., Ong, J. L. & Holliger, P. Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl Acad. Sci. USA* **98**, 4552–4557 (2001).
11. Dressman, D., Yan, H., Traverso, G., Kinzler, K. W. & Vogelstein, B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc. Natl Acad. Sci. USA* **100**, 8817–8822 (2003).
12. Ronaghi, M., Uhlen, M. & Nyren, P. A sequencing method based on real-time pyrophosphate. *Science* **281**, 363–365 (1998).
13. Fraser, C. M. *et al.* The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397–403 (1995).
14. Tettelin, H. *et al.* Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* **293**, 498–506 (2001).
15. Leamon, J. H. *et al.* A massively parallel PicoTiterPlate based platform for discrete picoliter-scale polymerase chain reactions. *Electrophoresis* **24**, 3769–3777 (2003).
16. Ronaghi, M. Pyrosequencing sheds light on DNA sequencing. *Genome Res.* **11**, 3–11 (2001).
17. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* **8**, 175–185 (1998).
18. Moore, G. E. Cramming more components onto integrated circuits. *Electronics* **38**(8) (1965).
19. Mehta, K., Rajesh, V. A. & Veeraswamy, S. FPGA implementation of VXIbus interface hardware. *Biomed. Sci. Instrum.* **29**, 507–513 (1993).
20. Fagin, B., Watt, J. G. & Gross, R. A special-purpose processor for gene sequence analysis. *Comput. Appl. Biosci.* **9**, 221–226 (1993).
21. Lander, E. S. & Waterman, M. S. Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* **2**, 231–239 (1988).
22. Myers, E. W. Toward simplifying and accurately formulating fragment assembly. *J Comput. Biol.* **2**, 275–290 (1995).
23. Ogawa, T. *et al. Increased Productivity For Core Labs Using One Polymer and One Array Length for Multiple Applications*. Poster P108-T. ABRF '05: Biomolecular Technologies: Discovery to Hypotheses (Savannah, Georgia, 5–8 February 2005); also available as Applied Biosystems 3730x/DNA Analyzer Specification Sheet (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge P. Dacey and the support of the Operations groups of 454 Life Sciences. This research was supported in part by the US Department of Health and Human Services under NIH grants.

Author Information Sequences for *M. genitalium* and *S. pneumoniae* are deposited at DDBJ/EMBL/GenBank under accession numbers AAGX01000000 and AAGY01000000, respectively. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Correspondence and requests for materials should be addressed to J.M.R. (jrothberg@454.com).

IDT Exhibit 1070.0005

© 2005 Nature Publishing Group

ERRATUM

doi:10.1038/nature04774

Significant primordial star formation at redshifts $z \approx 3$–4

Raul Jimenez & Zoltan Haiman

Nature 440, 501–504 (2006)

In this Letter, the received and accepted dates were incorrect. They should read: 'Received 6 September 2005; accepted 4 January 2006.'

CORRIGENDUM

doi:10.1038/nature04726

Genome sequencing in microfabricated high-density picolitre reactors

Marcel Margulies, Michael Egholm, William E. Altman, Said Attiya, Joel S. Bader, Lisa A. Bemben, Jan Berka, Michael S. Braverman, Yi-Ju Chen, Zhoutao Chen, Scott B. Dewell, Alex de Winter, James Drake, Lei Du, Joseph M. Fierro, Robin Forte, Xavier V. Gomes, Brian C. Godwin, Wen He, Scott Helgesen, Chun Heen Ho, Stephen K. Hutchison, Gerard P. Irzyk, Szilveszter C. Jando, Maria L. I. Alenquer, Thomas P. Jarvie, Kshama B. Jirage, Jong-Bum Kim, James R. Knight, Janna R. Lanza, John H. Leamon, William L. Lee, Steven M. Lefkowitz, Ming Lei, Jing Li, Kenton L. Lohman, Hong Lu, Vinod B. Makhijani, Keith E. McDade, Michael P. McKenna, Eugene W. Myers, Elizabeth Nickerson, John R. Nobile, Ramona Plant, Bernard P. Puc, Michael Reifler, Michael T. Ronan, George T. Roth, Gary J. Sarkis, Jan Fredrik Simons, John W. Simpson, Maithreyan Srinivasan, Karrie R. Tartaro, Alexander Tomasz, Kari A. Vogt, Greg A. Volkmer, Shally H. Wang, Yong Wang, Michael P. Weiner, David A. Willoughby, Pengguang Yu, Richard F. Begley & Jonathan M. Rothberg

Nature 437, 376–380 (2005)

In this paper, the second name of Chun Heen Ho was misspelled as He. In the earlier Corrigendum (*Nature* **439,** 502; 2006) the surnames of Stephen K. Hutchison and Brian C. Godwin were misspelled as Hutchinson and Goodwin. In addition, in the Supplementary Information of the original paper, there were errors in the sequences of the DNA capture and enrichment primers, and in the formulation of emulsion oil. The Supplementary Information was updated on 4 May 2006.

CORRIGENDUM

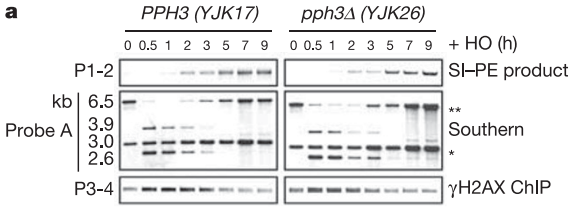
doi:10.1038/nature04772

A phosphatase complex that dephosphorylates γH2AX regulates DNA damage checkpoint recovery

Michael-Christopher Keogh, Jung-Ae Kim, Michael Downey, Jeffrey Fillingham, Dipanjan Chowdhury, Jacob C. Harrison, Megumi Onishi, Nira Datta, Sarah Galicia, Andrew Emili, Judy Lieberman, Xuetong Shen, Stephen Buratowski, James E. Haber, Daniel Durocher, Jack F. Greenblatt & Nevan J. Krogan

Nature 439, 497–501 (2006)

Figure 3a of this Letter contains an inadvertently duplicated panel set: those referring to '*pph3*Δ (YJK26)' are identical to '*PPH3* (YJK17)'. The corrected panels are shown here. Note that the quantification shown in Fig. 3b refers to the correct panels. Our results and conclusions are unaffected by this oversight.



CORRIGENDUM

doi:10.1038/nature04773

A high-mobility electron gas at the $LaAlO_3/SrTiO_3$ heterointerface

A. Ohtomo & H. Y. Hwang

Nature 427, 423–426 (2004)

In Fig. 2a and b of this Letter, the *y* axis should be labelled in units of ohms per square, not milli-ohms per square. This error does not occur elsewhere in the paper, and does not change any of our results or conclusions.

IDT Exhibit 1070.0006

© 2006 Nature Publishing Group