Reports

Direct sequencing of small genomes on the Pacific Biosciences RS without library preparation

Paul Coupland[1], Tamir Chandra[2], Mike Quail[1], Wolf Reik[2], and Harold Swerdlow[1]
[1]*Wellcome Trust Sanger Institute, Hinxton, UK,* [2]*Babraham Institute, Cambridge, UK*

BioTechniques 53:365-372 (December 2012) doi 10.2144/000113962
Keywords: Pacific Biosciences; PacBio; next-generation DNA Sequencing; single molecule real time; library prep

Supplementary material for this article is available at www.BioTechniques.com/article/113962

We have developed a sequencing method on the Pacific Biosciences RS sequencer (the PacBio) for small DNA molecules that avoids the need for a standard library preparation. To date this approach has been applied toward sequencing single-stranded and double-stranded viral genomes, bacterial plasmids, plasmid vector models for DNA-modification analysis, and linear DNA fragments covering an entire bacterial genome. Using direct sequencing it is possible to generate sequence data from as little as 1 ng of DNA, offering a significant advantage over current protocols which typically require 400–500 ng of sheared DNA for the library preparation.

Pacific Biosciences (Menlo Park, CA, USA) have developed a platform that will sequence a single molecule of DNA in real-time via the polymerization of that strand with a single polymerase (1–6). This technique has many benefits over multi-molecule (clonal) sequencing technologies (7,8); one such potential advantage is that it may not be absolutely necessary to make a library (i.e., create SMRT bells (9)) to generate sequence data. The only input (molecular) requirements to enable sequencing are a primed piece of DNA; both single-stranded and double-stranded molecules will work. The polymerase is necessarily highly processive starting with a location on the DNA at which it can bind, i.e., a free 3′-OH group. We decided to test whether any primed DNA molecules, lacking any other features of a PacBio SMRT bell, could be used directly in a sequencing reaction. The bound complex (DNA-primer-polymerase), although lacking PacBio adapter sequences, can still be sequenced on the PacBio platform. The present efficiency of this process, in terms of the numbers of reads generated and Mb yield per SMRT cell, is considerably less than that using standard libraries. With standard methods a typical SMRT cell will yield 35,000–50,000 reads and 100–160 Mb of mapped bases. The direct sequencing method described here has generated up to 3000 reads per SMRT cell and therefore its utility is limited to small genomes. However, this approach enables one to acquire sequence data from comparatively low amounts of DNA, even less than 1 ng of input, and within eight hours from receiving the sample. There is a slight time saving, compared with the 12 h required for standard library prep. This is not the main advantage, though it does now offer a route from sample to sequence within an average working day. This protocol may be of benefit to the direct sequencing of plasmids, single-standed or double-stranded viruses, mitochondrial DNA, and microbial pathogens in a clinical setting.

Materials and methods

M13mp18 viral DNA (both single-stranded and double-stranded; catalog no. N4040S and N4018S, respectively) and M13 forward (5′-GTTTTCCCAGTCACGAC-3′) and reverse sequencing primers (5′-AACAGC-TATGACCATG-3′) were from New England Biolabs (Hitchin, UK). Methicillin-resistant *Staphylococcus aureus* (MRSA) plasmids were purified from a solution prep of *S. aureus* TW20 using a Qiagen (Crawley, UK) Plasmid Midi Kit with Qiagen Genomic-tip 100/G following the manufacturer's "very low-copy plasmid/cosmid purification protocol" from a 500 ml culture. Plasmid Safe DNase (Epicentre Biotechnologies, Madison, WI, USA) was used to reduce the amount of linear single- and double-stranded molecules from the TW20 plasmid prep. Random hexamer primers from Roche (Welwyn Garden City, UK) were used, as provided in the Transcriptor First Strand cDNA Synthesis Kit. pET28a plasmid vectors encoding EcoDamI methyltransferase (Dam constructs) expressed in dam-/dcm *Escherichia coli* cells, were prepared in-house. Components from the DNA/Polymerase Binding Kit 2.0 from Pacific Biosciences were used during the annealing and binding reactions. The Annealing and Binding Calculator (version 1.3.1) provided by Pacific Biosciences was used to calculate the concentration of bound complex to be loaded onto the sample plate for the instrument. An MJ PTC-225 thermocycler from MJ Research (Watertown, MA, USA) was used for the annealing and binding reactions. The PacBio DNA Sequencing Kit 2.0 (8Rxn) and SMRT Cell 8Pac v2 (8 Cells) were used for sequencing. Sequence analysis was performed with SMRT portal, SMRT pipe, and SMRT View, version 1.3.1, and Motif Finder, version 0807, all from Pacific Biosciences.

Annealing reaction

Standard library preparation was omitted; the DNA templates were used directly in the annealing reaction. For each experiment, a quantity of DNA between 1 ng and 100 ng was annealed with suitable primers. With ssDNA, the annealing reaction used the standard PacBio protocol; i.e., 2 min at 80°C followed by cooling at 0.1°C/s to 25°C. With dsDNA, a different annealing protocol was used; the reaction was heated to 95°C for 5 min, then immediately snap-cooled on wet ice. As an example, when using ds M13mp18 DNA, 2.2 µL of DNA at 46 ng/µl (~100 ng), 0.9 µL PacBio Primer Buffer (10×), and both 0.9 µL forward primer (10 µM) and 0.9 µL

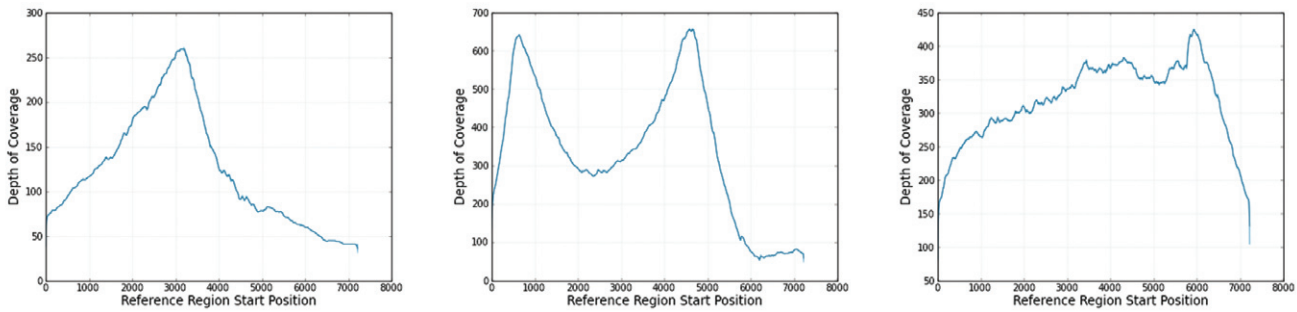

Figure 1. The coverage plot of M13mp18 viral genome after sequencing wih a single SMRT cell. (Left) M13mp18 ssDNA sequenced using the M13 forward sequencing primer. (Middle) M13mp18 dsDNA sequenced using the M13 forward and reverse sequencing primers. (Right) Coverage of M13mp18 dsDNA sequenced using random hexamer primers.

reverse primer (10 µM) were mixed in a final annealing reaction volume of 9 µL. The final concentration of DNA template was therefore ~2.5 nM, with 1000 nM primer (~400×). In order to use the PacBio Annealing and Binding calculator, we assumed that denatured M13mp18 DNA is comparable to a SMRT bell, with half the original double-stranded M13mp18 molecule's nominal length; i.e., one double-stranded 7.2-kb molecule, when denatured, becomes two 3.6 kb SMRT bells. A 2-fold dilution series of DNA was used to create additional annealing reactions in the range of 0.8-100 ng of DNA. There was a massive excess of forward and reverse primers at the lower concentrations of DNA in these reactions.

Binding reaction, loading, and sequencing

In the binding reaction, the ratio of polymerase to template DNA used was 3:1. First, 1.5 µL of polymerase (1600 nM) was combined with 25 µL of binding buffer giving a 90 nM polymerase solution. Four µl of a 1:1:1 DTT:dNTP:binding buffer mix (each from the PacBio Binding Kit) was added to the annealed template DNA and 1.5 µL of 90 nM polymerase was added. This was mixed gently by pipetting and then incubated at 30°C for 4 h.

The bound complex was loaded at 1 nM onto the instrument. Typically this is achieved by diluting the bound complexes with a mixture of 1:10 DTT:Complex Dilution Buffer. In this experiment, however, it was only possible to achieve a 1 nM loading concentration for the samples containing 100 ng and 50 ng input DNA. For the other samples in the 2-fold dilution series, the calculated concentration was <1 nM before dilution. The total volume of 14.6 µL of binding reaction was therefore loaded directly into the sample plate wells for each of these dilute samples.

Two × 45 min sequencing movies were acquired for each sample in this study. Mapping, de novo assembly, and modification analysis, were carried out with PacBio's SMRT Analysis pipeline run via the SMRT Portal interface. PacBio's Motif Finder was used in the final step of analysis for the pet28a plasmid vector to characterize the sequence specific motif at which base modifications were observed.

Table 1. Sequencing metrics for a dilution series of genomic input DNA.

Input DNA (ng)	# of Filtered Reads	Mean Mapped Read Length	% Bases Called	Coverage Depth (mean)	Consensus Accuracy (%)
100	1917	1559	100	329	100
50	2514	1484	100	412	100
25	2716	1378	100	384	100
12.5	2395	1294	100	297	100
6.3	1168	1058	100	103	100
3.1	613	1119	100	56	100
1.6	224	1046	99.9	18	99.8
0.8	74	674	91.4	3	93.4

M13mp18 DNA was sequenced on the PacBio RS using a single SMRT cell with 2 × 45 min movies for each sample. The input amount of DNA, used in the annealing and binding reaction, was reduced from 100 ng to <1 ng and the correlating sequence metrics are shown in each row. In terms of data yield, number of mapped reads, and % loaded ZMWs (filtered reads) on a SMRT cell, clearly the efficiency of this direct sequencing method is substantially less than that of the standard method, approximately 2%–5%, but the value attributed to the speed and ease of generating data from small amounts of DNA may be of greater importance to a specific application.

Results and discussion

At the outset of this study, an experiment using single-stranded M13mp18 viral DNA and the M13 forward sequencing primer (5′-GTTTTCCCAGTCACGAC-3′) showed that it was possible to generate sequence data directly from circular DNA molecules without library preparation; i.e., fragmentation, end repair, and adapter ligation. From 25 ng of ssDNA and 100-fold molar excess of primer, it was possible to map the data generated against the 7.2 kb M13mp18 reference sequence, calling 100% of the bases with 100% consensus accuracy. We next attempted to sequence double-stranded circular molecules of M13mp18 using both forward and reverse primers to obtain information from both strands in a single run. The sequencing of dsDNA molecules should have much wider application; for example, plasmids, phages, and ultimately larger genomes. Proving the ability to generate sequence data for both strands was therefore an important step in the development of this technique, especially considering the future application of PacBio for epigenetics (including hemi-methylation patterning); the ultimate goal was to sequence fragmented linear dsDNA, e.g., any sheared genomic sample, and generate enough useful data for future applications. We denatured the double-stranded DNA at 95°C for 5 min and snap-cooled (see Materials and Methods) in the presence of excess primer to successfully prime the two strands. This snap-cooling technique and the large primer concentration was utilized to give maximum opportunity for priming each strand while minimizing re-annealing of the genomic DNA. Alternative annealing conditions were tested as well: *(i)* following the standard PacBio recommended protocol of slowly cooling from 80°C to 25°C, *(ii)* snap-cooling from 95°C then raising the temperature to 45°C for 2 min, and *(iii)* cooling as quickly as possible on a thermocycler from 95°C to 45°C. In each of the latter three cases, far fewer reads were

Figure 2. SMRT View genome browser, showing sequence data mapped against M13mp18. A single >10 kb read, longer than the entire M13mp18 circular genome, is highlighted.

generated in the sequencing run. Snap-cooling on ice from 95°C was used subsequently for each dsDNA sample.

Figure 1 shows the difference in coverage profile of the M13mp18 genome when sequenced as ssDNA with the M13 forward sequencing primer, and the dsDNA sequenced with both the M13 forward and reverse sequencing primers (Figure 1, middle panel). The uneven coverage profile is due to the population of mapped reads (on each strand) having a distribution of read lengths, but most of the reads would start from approximately the same position on the genome. With PacBio sequencing at present there is a "dark-time" of several minutes prior to the start of movie acquisition, which is the time it takes from initiation of the sequencing reaction to alignment of the SMRT cell in the correct position. Although the priming sequence for a given strand is in the same position on each molecule, variation in the polymerization speed and longevity does account for some of the observed distribution. Additionally, the SMRT pipe software might have difficulty in mapping some reads, especially those that extend beyond the end of the linear FASTA reference. The DNA molecules sequenced were circular, but the reference used is a single linear sequence. Therefore, a number of the reads generated in these runs will, in fact, extend beyond the artificially imposed boundaries of the reference file. Some of the longest reads will also span the entire circular genome, further complicating the automatic analysis. The SMRT analysis software is not designed to deal with reads of this nature; although the initial filtering of the data are unaffected, as it's based on read quality metrics only. None of the reads have PacBio adapters that signal the end of a DNA template fragment so the standard re-sequencing (mapping) protocol in SMRT portal possibly contributes to the uneven coverage profiles generated (mapping thresholds were a maximum of one hit per read, 30% maximum divergence, and minimum anchor size of 12). Some reads were longer than the entire genome as evidenced by the maximum read lengths in the SMRT Portal raw read-length histogram (i.e., any reads >7.2 kb). These very long reads could be observed using PacBio's SMRT view software by concatenating two M13mp18 references in tandem into a single FASTA file (Figure 2).

The sequencing metrics for a 2-fold dilution series of M13mp18 input genomic DNA ranging from 100 ng to 0.8 ng is shown in Table 1. At the lowest input amount (0.8 ng) the run generated only 74 reads, but those reads gave enough data to call 91.4% of the genome with 93.4% consensus accuracy. With only 3.1 ng of starting material, mapped reads covered 100% of the bases at 100% consensus accuracy. Using the manufacturer's standard methods, one would be required to start with far more DNA, regardless of the target genome size; e.g., the standard protocol recommends a minimum of 500 ng for a 1-kb sequence. For a 7.2-kb sequence, the recommendation would be to use several µg.

To enhance this technique, we used random hexamer primers rather than primers specifically designed for the sample. In this case, no prior knowledge of the DNA sequence is necessary and the method, in principle, can be applied directly to a wide range of unknown samples e.g., in a clinical setting. Figure 1 (right panel) shows the coverage of ds M13mp18 sequenced with Roche's random hexamers. There is a more uniform coverage in comparison to the results obtained using the specific primers. The coverage distribution is still not ideal but having started with 50 ng of DNA, we generated similar data to that shown in Table 1 for 50 ng of input DNA (2392 mapped reads, 100% bases called, 100% consensus accuracy, 403 × mean coverage).

To test the application of our direct sequencing method for the PacBio detection of modified bases, 6-kb vectors were sequenced with known positions of methylation. A

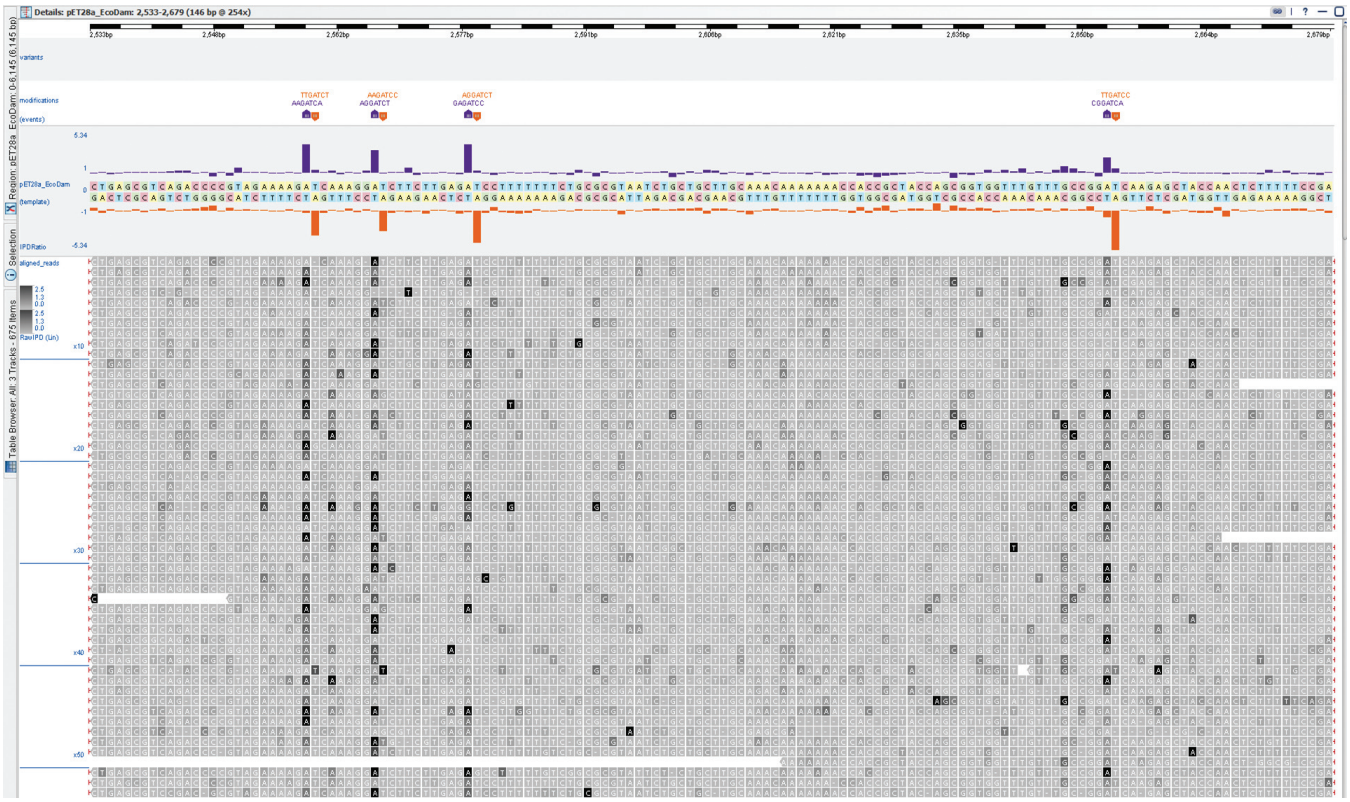

Figure 3. SMRT View genome browser, showing kinetic information elucidating modified bases. Four instances of m6A methylation in GATC motifs as detected with the PacBio base-modification analysis software (Motif Finder).

pET28a vector encoding EcoDam methyltransferase, which generates N6-methyladenine (m6A) methylation in GATC motifs, was directly sequenced with subsequent kinetic analysis using PacBio's software to identify base modification. Random hexamer primers were used and the experimental procedure was as described previously for M13mp18, starting with 25 ng of DNA. Figure 3 shows the SMRT View genome browser depiction of a portion of the sequence data for one of the vectors sequenced using this direct sequencing technique; four instances of GATC methylated sites are evident as peaks in the purple trace (+ strand) and orange trace (- strand). As the GATC sequence is a palindrome, there is an m6A base on both strands, and by observing the inter-pulse distance (IPD) ratio reported by the PacBio software, it is exceptionally easy to see these base modifications. Other studies have used the PacBio RS for base modification analysis on similar plasmid/methyltransferase models (10,11) and entire bacterial genomes (12) but to our knowledge these studies follow standard library preparation protocols and required far greater amounts of DNA than in the technique described here. The data we generated were analyzed with Motif Finder, an application provided by PacBio, for mining polymerization kinetics for motifs associated with base modifications. In this vector, 50 instances of GATC methylated at the A position were identified; there are 25 GATC sites in the sequence and wild type EcoDam was expected to methylate each one of them.

Direct sequencing was then tested using a DNA extract of *Staphylococcus aureus* TW20, a MRSA strain and well-known nosocomial infection (13). The plasmids of this bacterial sample were of interest as an example of the application of the PacBio RS to infectious disease identification through sequencing. Antibiotic resistance genes are often carried on plasmids (14,15) and can spread very quickly in heterogeneous bacterial communities (16–19). DNA was extracted from a solution culture of TW20 and digested with Plasmid Safe DNase to reduce the amount of linear fragments and effectively increase the concentration of plasmids in the sample. An electropherogram of the final sample showed that the DNA preparation also contained a smear of linear double stranded fragments ranging from 100 bp to >25 kbp, with a peak at approximately 20 kb (Supplementary Figure 1). The two plasmids in TW20 are double-stranded and circular, with lengths of 3 kb and 30 kb. We generated sequence data using random hexamer primers in the annealing reaction. Four reactions containing 50 ng of the *S. aureus* DNA preparation with various amounts of hexamer primers, from 10-fold to 600-fold, i.e., 500 ng, 1 µg, 10 µg, and 30 µg per annealing reaction, were performed in 9 µL reaction volumes. A single SMRT cell was sequenced for each reaction and the trend observed across these four reactions showed fewer mapped reads as the amount of random hexamer primers increased. This is perhaps because of the proximity of annealed primers on the DNA strand at higher concentrations, leading to polymerases colliding with one another, or simply the reduction of signal to noise as two fluorescent signals could be observed concurrently. The annealing reaction with 10-fold primers generated 3240 mapped reads, 20-fold generated 3085 mapped reads, 200-fold generated 2911 mapped reads, and 600-fold generated 2011 mapped reads, all with a mean mapped read length of approximately 500 bp. There was also a difference in coverage depth between the two plasmids; the mean coverage for the 3-kb plasmid was 35×, but only 5× coverage was obtained for the 30-kb plasmid, which is due mostly to the difference in plasmid length. There is a loading inefficiency of larger molecules because of their lower diffusion coefficient, as well as the disparity between the molecule's hydrodynamic radius and the very small zero-mode waveguide (ZMW). Future upgrades to the loading mechanism on the PacBio instrument (MagBead loading) which should eliminate this problem are very close to release. The combined sequence data from these four SMRT cells produced 13,724 reads; 479 reads mapped to the plasmids and 11,247 to the genome (5.3 Mb mapped providing a mean 1.6× coverage), an overall mapping rate

of 85% which is not dissimilar to standard mapping rates of SMRT bell libraries we have made (from a recent single SMRT cell of *S. aureus* TW20 1 kb SMRT bell library 39,478 reads were mapped from 47,465 filtered reads, a mapping rate of 83%).

Finally, the technique was used to sequence linear molecules of *Candidatus* Phytoplasma mali, a plant-pathogenic mycoplasma with a small genome of ~600 kb that is 21.4% GC and characterized by large terminal inverted repeats and covalently closed hairpin ends (20). The DNA was sheared to approximately 3-kb fragments and a 25-ng aliquot was sequenced using random hexamers in a similar manner to that described previously. From a single SMRT cell with 2 × 45 min movies, only 870 post-filter reads were generated of which 63 reads mapped, with a mean consensus accuracy of 84.4%. The mean mapped read-length was 817 bp and the coverage only 0.08%. The poor mapping rate is most likely due to a greater percentage of low-quality reads from this particular sample. Although the yield is poor, direct sequencing of these linear DNA molecules shows some promise too. A blastn (21) search using the NCBI server against the refseq_genomic database called out *Candidatus* Phytoplasma mali as the most likely taxonomic hit (Supplementary Table 1). This suggests it is possible to obtain enough information from very few mapped reads to begin to identify the genomes present in a sample. However, comparing the difference in data yield between the *S. aureus* and *Ca*. Phytoplasma mali, it is clear that further optimization of the method is required to improve the number of reads that can be mapped when sequencing linear molecules from a variety of genomic samples.

The method described here utilizes the PacBio RS platform for direct sequencing, enabling the generation of sequence data from small single- and double-stranded DNA genomes. Potentially this technique also could be applied to circularized molecules, e.g., amplicons or sheared fragments that have been circularized. However, the additional circularization step and clean up would mean relatively minor time and DNA savings compared with current PacBio protocols. The direct sequencing technique could allow the identification of plasmids present in a bacterial sample in an extremely straightforward and fast manner. Although there is an indication that different genomes may be more or less accessible with this method, we have demonstrated its application to sequencing ssDNA and dsDNA viruses, plasmid vector models for methylation studies, antibiotic resistance gene-carrying plasmids, and the entire genome of a clinically relevant microbial pathogen. All of these were performed without the need for library preparation, and it is possible to generate sequence data within 8 h from <1 ng of DNA without a PCR amplification step. The fact that our method can be performed without a priori knowledge of any sequence and with no organism-specific reagents, coupled with its simplicity and speed, makes it particularly well suited for use in acute disease and infectious outbreak scenarios.

Acknowledgments

We would like to thank NEB for providing the M13 primers which were not currently available for purchase. Theresa Feltwell of The Wellcome Trust Sanger Institute for culturing *S. aureus* TW20 and performing the plasmid prep. Albert Jeltsch and Tomasz Jurkowski for providing the original Dam constructs. This work was supported by the Wellcome Trust grant 098051 (PC, MQ, HS) and The Cambridge Cancer Center (TC, WR).

Competing interests

The authors declare no competing interests.

References

1. **Eid, J., A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, et al.** 2009. Real-Time DNA Sequencing from Single Polymerase Molecules. Science *323*:133-138.
2. **Korlach, J., A. Bibillo, J. Wegener, P. Peluso, T.T. Pham, I. Park, S. Clark, G.A. Otto, and S.W. Turner.** 2008. Long, processive enzymatic DNA synthesis using 100% dye-labeled terminal phosphate-linked nucleotides. Nucleosides Nucleotides Nucleic Acids *27*:1072-1083.
3. **Korlach, J., K.P. Bjornson, B.P. Chaudhuri, R.L. Cicero, B.A. Flusberg, J.J. Gray, D. Holden, R. Saxena, et al.** 2010. Real-time DNA sequencing from single polymerase molecules. Methods Enzymol. *472*:431-455.
4. **Levene, M.J., J. Korlach, S.W. Turner, M. Foquet, H.G. Craighead, and W.W. Webb.** 2003. Zero-mode waveguides for single-molecule analysis at high concentrations. Science *299*:682-686.
5. **McCarthy, A.** 2010. Third generation DNA sequencing: pacific biosciences' single molecule real time technology. Chem. Biol. *17*:675-676.
6. **Schadt, E.E., S. Turner, and A. Kasarskis.** 2010. A window into third-generation sequencing. Hum. Mol. Genet. *19*:R227-R240.
7. **Carneiro, M.O., C. Russ, M.G. Ross, S.B. Gabriel, C. Nusbaum, and M.A. Depristo.** 2012. Pacific biosciences sequencing technology for genotyping and variation discovery in human data. BMC Genomics *13*:375.
8. **Rasko, D.A., D.R. Webster, J.W. Sahl, A. Bashir, N. Boisen, F. Scheutz, E.E. Paxinos, R. Sebra, et al.** 2011. Origins of the E. coli strain causing an outbreak of hemolytic-uremic syndrome in Germany. N. Engl. J. Med. *365*:709-717.
9. **Travers, K.J., C.S. Chin, D.R. Rank, J.S. Eid, and S.W. Turner.** 2010. A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res. *38*:e159.
10. **Clark, T.A., I.A. Murray, R.D. Morgan, A.O. Kislyuk, K.E. Spittle, M. Boitano, A. Fomenkov, R.J. Roberts, and J. Korlach.** 2012. Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing. Nucleic Acids Res. *40*:e29.
11. **Song, C.X., T.A. Clark, X.Y. Lu, A. Kislyuk, Q. Dai, S.W. Turner, C. He, and J. Korlach.** 2012. Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. Nat. Methods *9*:75-77.
12. **Murray, I.A., T.A. Clark, R.D. Morgan, M. Boitano, B.P. Anton, K. Luong, A. Fomenkov, S.W. Turner, et al.** 2012. The methylomes of six bacteria. Nucleic Acids Res.
13. **Holden, M.T., J.A. Lindsay, C. Corton, M.A. Quail, J.D. Cockfield, S. Pathak, R. Batra, J. Parkhill, et al.** 2010. Genome sequence of a recently emerged, highly transmissible, multi-antibiotic- and antiseptic-resistant variant of methicillin-resistant Staphylococcus aureus, sequence type 239 (TW). J. Bacteriol. *192*:888-892.
14. **Robicsek, A., G.A. Jacoby, and D.C. Hooper.** 2006. The worldwide emergence of plasmid-mediated quinolone resistance. Lancet Infect. Dis. *6*:629-640.
15. **Svara, F. and D.J. Rankin.** 2011. The evolution of plasmid-carried antibiotic resistance. BMC Evol. Biol. *11*:130.
16. **Haenni, M., E. Saras, V. Metayer, B. Doublet, A. Cloeckaert, and J.Y. Madec.** 2012. Spread of the blaTEM-52 gene is mainly ensured by IncI1/ST36 plasmids in Escherichia coli isolated from cattle in France. J Antimicrob Chemother.
17. **Miró, E., C. Segura, F. Navarro, L. Sorli, P. Coll, J.P. Horcajada, F. Alvarez-Lerma, and M. Salvadó.** 2010. Spread of plasmids containing the bla(VIM-1) and bla(CTX-M) genes and the qnr determinant in Enterobacter cloacae, Klebsiella pneumoniae and Klebsiella oxytoca isolates. J. Antimicrob. Chemother. *65*:661-665.
18. **Valverde, A., R. Canton, M.P. Garcillan-Barcia, A. Novais, J.C. Galan, A. Alvarado, F. de la Cruz, F. Baquero, and T.M. Coque.** 2009. Spread of bla(CTX-M-14) is driven mainly by IncK plasmids disseminated among Escherichia coli phylogroups A, B1, and D in Spain. Antimicrob. Agents Chemother. *53*:5204-5212.
19. **Dionisio, F., I. Matic, M. Radman, O.R. Rodrigues, and F. Taddei.** 2002. Plasmids spread very fast in heterogeneous bacterial communities. Genetics *162*:1525-1532.
20. **Kube, M., B. Schneider, H. Kuhl, T. Dandekar, K. Heitmann, A.M. Migdoll, R. Reinhardt, and E. Seemuller.** 2008. The linear chromosome of the plant-pathogenic mycoplasma 'Candidatus Phytoplasma mali'. BMC Genomics *9*:306.
21. **Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*:3389-3402.

Received 21 September 2012; accepted 12 November 2012.

Address correspondence to Paul Coupland, Wellcome Trust Sanger Institute, Genome Campus, Hinxton, Cambridgeshire, UK. Email: pc10@sanger.ac.uk

To purchase reprints of this article, contact: biotechniques@fosterprinting.com