



Simple, robust methods for high-throughput nanoliter-scale DNA sequencing

Duane E. Smailus, Andre Marziali, Philip Dextras, et al.

Genome Res. 2005 15: 1447-1450

Access the most recent version at doi:[10.1101/gr.4221805](https://doi.org/10.1101/gr.4221805)

References

This article cites 12 articles, 3 of which can be accessed free at:
<http://genome.cshlp.org/content/15/10/1447.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

An advertisement banner with a teal background. On the left, the text reads "CRISPR and RNAi Genetic Screening. Your new superpower." In the center, there is a white box with the text "LEARN MORE". On the right, there is a photograph of a woman wearing a red and white superhero cape and mask, with the Cellecta logo (a green molecular structure) and the word "CELLECTA" below it.

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press

Methods

Simple, robust methods for high-throughput nanoliter-scale DNA sequencing

Duane E. Smailus,¹ Andre Marziali,² Philip Dextras,² Marco A. Marra,¹ and Robert A. Holt^{1,3}

¹Canada's Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, British Columbia, Canada V5Z 4S6; ²Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z1.

We have developed high-throughput DNA sequencing methods that generate high quality data from reactions as small as 400 nL, providing an approximate order of magnitude reduction in reagent use relative to standard protocols. Sequencing of clones from plasmid, fosmid, and BAC libraries yielded read lengths (PHRED20 bases) of 765 ± 172 ($n = 10,272$), 621 ± 201 ($n = 1824$), and 647 ± 189 ($n = 568$), respectively. Implementation of these procedures at high-throughput genome centers could have a substantial impact on the amount of data that can be generated per unit cost.

[Supplemental material is available online at www.genome.org.]

Between February 2004 and February 2005 the NCBI Trace Archive (<http://www.ncbi.nlm.nih.gov/Traces>) received 287,319,810 sequencing read electropherograms from various high-throughput DNA sequencing projects. A nominal cost of \$1 per read suggests that in its most strict definition, high-throughput DNA sequencing is presently at least a several hundred million dollar per year industry. Large-scale genomics efforts, particularly whole genome sequencing and polymorphism detection sequencing, continue to be cost-limited. Current efforts to increase data production per unit cost have focused on either (1) new and potentially revolutionary methods such as sequencing by synthesis (Brenner et al. 2000), polymerase colony sequencing (Mitra et al. 2003), and single molecule sequencing (Braslavsky et al. 2003; Levene et al. 2003); or (2) on evolutionary approaches that strive for volume reduction within the paradigm of Sanger sequencing (Sanger et al. 1977), four-color fluorescence (Smith et al. 1985, 1986), and capillary electrophoresis (Dovichi 1997). Here we focus on the latter approach. Typically, high-throughput genome centers use sequencing reaction volumes of several microliters, which yield quantities of products far in excess of what is required to generate high-quality data by capillary electrophoresis. The limiting factors for volume reduction have been reaction vessels (microtiter plates) that allow sample loss by seal leakage and condensation on unwetted inner well surfaces (see Supplemental material) and rigid adherence of laboratory liquid-dispensing robotics to these established microtiter plate formats.

Here we describe how testing and implementation of (1) a novel nanoliter-scale reaction vessel configuration, (2) submicroliter positive pressure microsolenoid-based liquid-dispensing robotics, and (3) optimized sequencing reaction chemistry, thermocycling conditions, and capillary electrophoresis injection parameters have allowed us to generate high-throughput sequence data of equal or greater quality to standard methods, with substantial reduction in reagent consumption. All materials we describe are commercially available, such that with appropriate attention to detail the process we describe can be easily implemented at other facilities.

³Corresponding author.

E-mail rholt@bcgsc.ca; fax (604) 877-6085.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.4221805>. Article published online before print in September 2005.

Results and Discussion

Validation tests were performed on 16 different libraries constructed using a variety of common low-, medium-, and high-copy number vectors (Table 1).

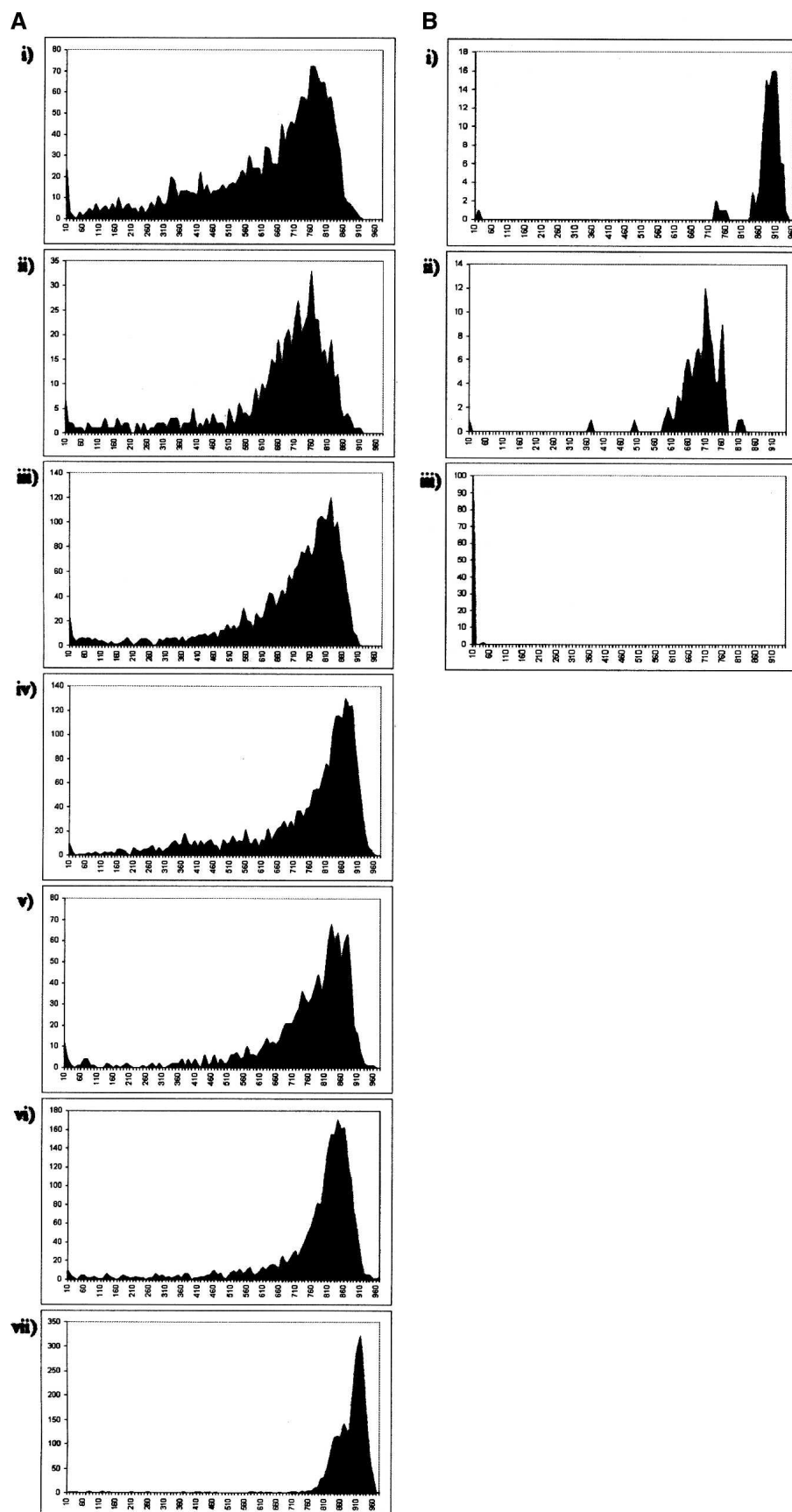
For plasmid libraries, mean PHRED20 read lengths in excess of 750 bp were achieved (Fig. 1A). For library 10790, which is a mock library in which every well contains the same human cDNA clone, the average read length approached 900 bp, and the longest read was 972 bp (Supplemental Fig. 1). These read lengths were comparable to those achieved with our standard DNA sequence production pipeline that uses 540 nL Big Dye Terminator Premix V3.1 (Applied Biosystems) (Table 1), and comparable to those typically submitted to the NCBI trace archive by other high-throughput centers. For all templates that were sequenced in this study, excluding the ~4% failed cultures, the average sequencing success rate (samples with a PHRED20 read length of at least 100 bp) was 97%. Although all data in the present study are single-end reads, when applied to paired-end sequencing this method would be expected to give a successful paired-end rate of 0.97×0.97 , or 94%. After sequencing these libraries using 31.25 nL of Big Dye terminators per well and obtaining adequate read length and signal strength (Table 1), we investigated the absolute lower limit for Big Dye terminator consumption. Ninety-six identical clones from library 10790 were sequenced using either 15.63 nL, 7.81 nL, or 3.13 nL of Big Dye terminators in a 400-nL total reaction that contained 20 nL of 15X reaction buffer; 1 nL, 0.5 nL, or 0.2 nL, respectively, of 100 μ M – 21 M13 forward primer; and an appropriate volume of Ultrapure water (Invitrogen). Mean PHRED20 read lengths obtained were 871 ± 97 , 684 ± 92 , and 0 bases, with maximum PHRED20 read lengths of 887, 703, and 0 bases, respectively (Fig. 1B), suggesting that the limiting volume of dye terminators in this system lies between 3.13 and 7.81 nL.

We expect that given the simplicity of the platform we describe here, and its foundation firmly within the time-tested Sanger sequencing paradigm, it will be easily implemented by any center engaged in moderate- to high-throughput DNA sequencing. The chemistry is robust for all vector types (plasmids, fosmids, BACs) typically utilized for high-throughput sequencing. While reagent, equipment, and labor costs will vary among high-throughput sequencing platforms at different institutions,

Table 1. Plasmid, BAC, and fosmid DNA libraries sequenced in a 400-nL reaction containing 31.25 nL of Big Dye terminators (V3.1, Applied Biosystems) as compared to standard 4000-nL reactions containing 540 nL of Big Dye terminators

Sequence type	Organism	Library name	Vector	Vector copy number	Mean insert size (kbp)	GC content	DNA conc. (mean \pm SD) (ng/ μ L)	Primer	31.25 nL Big Dye			540 nL Big Dye		
									Number of reads	Length (PHRED20 bp, mean \pm SD)	Mean signal strength	Number of reads	Length (PHRED20 bp, mean \pm SD)	Mean signal strength
5' EST	<i>H. sapiens</i>	10790 ^a	pOTB7	High	1.9	N/A	70.31 \pm 0.89	-21M13F	2304	881 \pm 72	1072	879 \pm 100	1053	
5' EST	<i>M. musculus</i>	LL005	pDNRLIB	High	2.3	51%	47.87 \pm 2.19	-21M13F, T7	2230	738 \pm 189	342	753 \pm 200	1987	
	<i>R. norvegicus</i>	LL012	pYX-Asc											
	<i>H. sapiens</i>	LL014	pOTB7											
	<i>X. tropicalis</i>	LL015	pCMV-SPORT6											
		LL016												
		LL017												
SAGE	<i>H. sapiens</i> (prostate cancer)	S1881	pZero-1	High	0.834	47%	30.42 \pm 0.82	T7	2281	764 \pm 157	381	786 \pm 156	1169	
WGS	<i>Cryptococcus neoformans</i>	CN23E	pBR194cE	Medium	10	47%	15.17 \pm 0.19	-21M13F	2304	715 \pm 171	104	825 \pm 142	593	
Transposon ^b	<i>Xenopus laevis</i>	TX060	pCMV-SPORT6	High	3.4 ^c	45%	72.38 \pm 2.63	SeqA2	1152	740 \pm 170	333	774 \pm 242	3190	
Fosmid end	<i>Chlamydia trachomatis</i>	CA000	pEpifos5	Low	30	42%	40.52 \pm 0.60	pEpifos5 Forward	1824	621 \pm 201	255	N/A	N/A	
		CD000												
		CE000												
		CL000												
BAC end	<i>Gasterosteus aculeatus</i>	GA000	pTarBAC2.1	Low	210	43%	50.76 \pm 0.87	T7	568	646.97 \pm 189	199	N/A	N/A	

^aLibrary 10790 is comprised of multiple copies of an identical clone.^bTransposons carrying unique priming sites are randomly inserted by recombination into pooled sets of plasmid clones (Butterfield et al. 2002).^cIncludes 2.3 kb of insert plus 1.1 kb transposon.



implementation of these methods at our center reduces the cost of sequencing reactions by ~90%, relative to the cost of our established 4- μ L reactions (Supplemental Fig. 3). We expect that cost savings realized from volume reduction can very rapidly offset the cost of the robotics and plasticware required for the process.

Methods

A flow diagram summarizing the basic process is presented in Supplemental Figure 2. Template DNA from multiple libraries constructed with different vector types (Table 1) was prepared as previously described (Yang et al. 2005). Briefly, for plasmid clones, 60 μ L of 2xYT liquid culture was grown for 18 h with shaking (350 rpm) in 384-deep well diamond plates (Axygen) covered with AirPore tape (Qiagen). Only sample wells that failed to show any cell growth (4%) were removed from analysis. To extract DNA, 60 μ L of lysis buffer (Qiagen) was added directly to the overnight culture. After 5 min of lysis, 60 μ L of neutralization buffer (Qiagen) was added. Plates were tape sealed (Edge Biosystems clear tape) and mixed by vortexing on a high-power multi-plate vortexer (VWR, model VX-2500) at maximum speed for 2 min prior to centrifugation at 4250g for 25 min. One hundred twenty microliters of cleared lysate was transferred from culture blocks into 240- μ L 384-deep well diamond plates containing 90 μ L of 100% isopropanol per well, mixed by inversion, and centrifuged at 2830g for 15 min. Isopropanol was decanted, and the DNA pellet was washed with 50 μ L of 80% ethanol and then air-dried.

Figure 1. (A) Distribution of read lengths (PHRED20 base count) for each library described in Table 1, sequenced with a 400-nL reaction that contained 31.3 nL of Big Dye terminators (v3.1, Applied Biosystems). (i) CA001, CD001, CE001, CL001: Fosmid end reads; (ii) GA000: BAC end reads; (iii) CN23E: medium copy-number plasmid whole genome shotgun reads; (iv) LL005-LL0017: high copy-number plasmid 5' EST reads; (v) TX060, TX067: high copy-number plasmid transposon-mediated shotgun reads; (vi) S1881: high copy-number plasmid SAGE library reads; (vii) 10790: high copy-number plasmid 5' EST reads from 2304 identical clones. (B) Distribution of read lengths (PHRED20 base count) for 96 identical clones from library 10790, sequenced with a 400-nL reaction that contained (i) 15.63 nL, (ii) 7.81 nL, or (iii) 3.13 nL of Big Dye terminators.

DNA pellets were resuspended in 10 mM Tris-HCl pH 8 containing 10 µg/mL RNase A (Qiagen). BAC DNA was prepared using a similar automated alkaline lysis procedure in 96-well format, as previously described (Schein et al. 2004). These are generally very crude DNA preparations, as no organic solvents, paramagnetic particles, membranes, or filters are used in the process. DNA was quantified using PicoGreen against a standard curve generated using known amounts of phage lambda DNA.

After careful consideration (see Supplemental material) we determined that the most practical method for delivering DNA template to a submicroliter reaction was to transfer a relatively large volume of dilute template using standard laboratory robotics, desiccate, and then resuspend in an appropriately small volume (200–400 nL) of sequencing reaction master mix. Using a Biomek FX (Beckman-Coulter), plasmid DNA was diluted 10-fold in Ultrapure water (Invitrogen), and 2 µL (~15–55 ng) was transferred to 384-well PCR cycle plates (ABGene), then completely desiccated in a drying oven for 10 min at 95°C. We find that it is important to restrict the volume of the initial transfer to ≤ 2 µL, as DNA that adheres to the inner well surface during desiccation will be unavailable for the sequencing reaction. High-throughput sequencing centers that serially process a large number of plates may wish to dry the plates overnight at room temperature in a laminar flow hood, as we have found this to work just as well as a drying oven.

Subsequent to template transfer, 400 nL of sequencing reaction mix containing 31.25 nL Big Dye Terminator Premix V3.1 (Applied Biosystems), 40 nL of a custom formulation of reaction buffer (25X Reaction Buffer [2M Trizma Base, 50 mM MgCl₂-6 H₂O] combined with an equal volume of 5X Big Dye Terminator Reaction Buffer V3.1 [Applied Biosystems]), 2 nL of 100 µM primer (Invitrogen), and 326.75 nL Ultrapure Water (Invitrogen) were added to each well using the Aurora Discoveries Flying Reagent Dispenser. This instrument was developed for high-throughput screening in the pharmaceutical industry and has not previously been applied to high-throughput sequencing applications (Supplemental material). Plates were sealed with SPRI Plug Low Volume lids (Agencourt Bioscience) to reduce residual air space in the wells and thereby reduce sample loss by evaporation. Again, the principal limiting factor for volume reduction of cycle sequencing has been loss of fluid from the sample due to elevated temperatures and rapid temperature changes (please refer to Supplemental Online Material for theoretical consideration of factors contributing to fluid loss and measurements of fluid loss using various types of plate seals). Agencourt SPRI lids recently became commercially available. Although the SPRI lids are an added cost to the process, they may be reused for up to 10 thermal cycling reactions, and their cost is offset from the savings achieved by reduced volume of sequence reaction mix.

Sealed plates were thermocycled using Tetrad peltier thermal cyclers (MJ Research). Thermal cycling conditions for the reaction format presented here were 50 × (96°C, 10 sec; 43°C, 5 sec; 60°C, 240 sec) with ramping rates of 1°C/sec. The optimal number of cycles was not evaluated, and fewer than 50 cycles may be adequate. Unincorporated nucleotides were removed from the sequence reactions by ethanol/EDTA precipitation as described in Yang et al. (2005), with the exception that 6.6 µL of 38 mM EDTA pH 8, rather than 2 µL of 125 mM EDTA pH 8, was added to the sequencing reaction products prior to the addition of 18 µL of 95% EtOH.

Purified sequence reaction products were resuspended overnight at 4°C (to maximize recovery of purified reaction products) in 10 µL Ultrapure Water (Invitrogen) and sequenced on one of seven 3730xl DNA Analyzers (Applied Biosystems) using 50-cm capillaries and POP-7 polymer (Applied Biosystems). Because the sequence reaction products from our nanoliter-scale reactions

were resuspended in the same volume as our standard 4-µL reactions, the labeled DNA was at lower concentration. As such, it was necessary to optimize the electrokinetic injection parameters such that a sufficient amount of labeled reaction products was injected in each capillary electrophoresis run. For the standard Applied Biosystems 3730xl run module, it was empirically determined that doubling the injection time (to 30 sec) but keeping the injection voltage, run time, and run voltage the same (1.5 kV, 5640 sec, and 8.5 kV, respectively) loaded sufficient material into the capillaries of the sequencers for generation of equivalent read lengths to our standard 4-µL reactions. The PHRED software package (v 0.020425.C) (Ewing and Green 1998) was used for base-calling and quality score assignments. For each read, the reported length is a count of the total number of PHRED20 bases.

Acknowledgments

We thank Jim Kronstad, Pieter J. de Jong, Marian Sadar, Robert Brunham, and Vancouver, B.C., Canada and IMAGE Consortium for DNA libraries used in this study. We thank members of the DNA Sequencing and Informatics Groups at Canada's Michael Smith Genome Sciences Centre for their technical assistance. In particular, we thank Ranibar Guin and Joseph Ray Santos for help with data processing. M.A.M. and R.A.H. are Michael Smith Foundation for Health Research scholars.

References

- Braslavsky, I., Hebert, B., Kartalov, E., and Quake, S.R. 2003. Sequence information can be obtained from single DNA molecules. *Proc. Natl. Acad. Sci.* **100**: 3960–3964.
- Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., et al. 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* **18**: 630–634.
- Butterfield, Y.S., Marra, M.A., Asano, J.K., Chan, S.Y., Guin, R., Krzywinski, M.I., Lee, S.S., MacDonald, K.W., Mathewson, C.A., Olson, T.E., et al. 2002. An efficient strategy for largescale high-throughput transposon-mediated sequencing of cDNA clones. *Nucleic Acids Res.* **30**: 2460–2468.
- Dovichi, N.J. 1997. DNA sequencing by capillary electrophoresis. *Electrophoresis* **18**: 2393–2399.
- Ewing, B. and Green, P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**: 186–194.
- Levene, M.J., Korlach, J., Turner, S.W., Foquet, M., Craighead, H.G., and Webb, W.W. 2003. Zero-mode waveguides for single-molecule analysis at high concentrations. *Science* **299**: 682–686.
- Mitra, R.D., Shendure, J., Olejnik, J., Edyta Krzymanska, O., and Church, G.M. 2003. Fluorescent in situ sequencing on polymerase colonies. *Anal. Biochem.* **320**: 55–65.
- Sanger, F., Nicklen, S., and Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74**: 5463–5467.
- Schein, J., Kucaba, T., Sekhon, M., Smailus, D., Waterston, R., and Marra, M. 2004. High-throughput BAC fingerprinting. *Methods Mol. Biol.* **255**: 143–156.
- Smith, L.M., Fung, S., Hunkapiller, M.W., Hunkapiller, T.J., and Hood, L.E. 1985. The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: Synthesis of fluorescent DNA primers for use in DNA sequence analysis. *Nucleic Acids Res.* **13**: 2399–2412.
- Smith, L.M., Sanders, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connell, C.R., Heimer, C., Kent, S.B., and Hood, L.E. 1986. Fluorescence detection in automated DNA sequence analysis. *Nature* **321**: 674–679.
- Yang, G.S., Stott, J.M., Smailus, D.E., Barber, S.A., Balasundaram, M., Marra, M.A., and Holt, R.A. 2005. High-throughput sequencing: A failure mode analysis. *BMC Genomics* **6**: 2.

Web site references

<http://www.ncbi.nlm.nih.gov/Traces>; National Center for Biotechnology Information (NCBI) Trace Archive

Received June 1, 2005; accepted in revised form August 1, 2005.