Testing of a novel ultra-sensitive DNA sequencing instrument based on single photon detection.

O. Bilenko\textsuperscript{1}, D. Gavrilov, B. Gorbovitski\textsuperscript{1}, M. Gouzman, G. Gudkov, O. Kosobokova, N. Lifshitz\textsuperscript{1}, S. Luryi, A. Stepoukhovitch, M. Tcherevishnick, G. Tyshko, V. Gorfinkel

\textit{Department of Electrical and Computer Engineering, SUNY SB, Stony Brook, NY}

V. Bajii, A. Bahret, R. Shah, L. King, L. Speigel, W.R. McCombie

\textit{Cold Spring Harbor Laboratory, Woodbury Genome Center, Cold Spring Harbor, NY}

\textbf{Key words:} DNA sequencing instrument, single photon detection,

To whom correspondence should be addressed:
Prof. Vera Gorfinkel
Phone: (631) 751-87-38
Fax: (631) 751-24-57
E-mail: vera@ece.sunysb.edu

\textsuperscript{1} BioPhotonics Corporation, NY
Abstract

A novel DNA ultra-sensitive DNA sequencing instrument based on single photon fluorescence detection has been developed at SUNY Stony Brook. Twenty single capillary prototypes were built and field-tested in collaboration with the Cold Spring Harbor Laboratory Genome Center. Sensitivity of the instrument was tested using standard production samples and serial dilutions in water prepared at the CSHL Genome Center. The quality and reliability of the trace data was analyzed by sequencing a cDNA clone from a canine EST library. The sequence of the entire gene was completed with primer walking using samples diluted so that they contained only 1/400 of the amount of Dye Terminator Sequencing mix recommended by the provider.

1. Introduction

In the present paper we describe results of testing a highly sensitive DNA sequencing instrument of unique design developed at SUNY Stony Brook and focus primarily on demonstrating detection of very low amounts of DNA sequencing product.

A major increase in the sensitivity of DNA sequencing instruments would provide several advantages. First, it would allow a reduction in the amount of sequencing reagents used, which constitute a significant portion of the cost at any large-scale sequencing center. Another advantage of a highly sensitive sequencer would potentially be to allow direct sequencing from very high molecular weight clones, minimizing subcloning.

Furthermore, it is likely that the percentage of sequencing cost attributed to reagents will increase with time as automation and robotics continues to reduce the cost of labor at sequencing centers. Most centers now use about 1/24th to 1/32nd the volume of terminator mix recommended by the vendor (typically Applied Biosystems). Reducing this by more than an order of magnitude would make reagent cost a negligible factor in overall sequencing costs.

There are two ways to reduce the amount of reagent used, which are not mutually exclusive. One can either reduce the amount of the terminator mix added to a fixed volume reaction resulting in a dilution or reduce the overall reaction volume or both.

One of the difficulties inherent to reducing the amount of reagent used is robust processing of very small volumes. Volumes of typical reactions at genome centers are 5-10 ul. However, several groups have already demonstrated sub-microliter reactions [He et all, 2001]. The second difficulty is detecting and analyzing very small amounts of DNA product. Improved detection of small amounts of DNA can be obtained using single photon counting technique for fluorescence detection.

The extremely high sensitivity of single photon detectors is based on a combination of their very high internal gain and extremely low dark noise. For example, a typical single photon sensitive photo-multiplying tube (PMT) has gain of ~10^8 electrons per one registered photon, and a dark count of ~100-200 counts/second at room temperature. A key difference between single photon counters and more commonly used linear photodetectors (found in commercial sequencers) is in the character of their response to light radiation. While detectors produce electric response proportional to the intensity of measured light, the output signal of single photon counters is in the form of short, binary electric pulses (one pulse per one photon). The amplitude of these pulses is large enough to be processed directly by digital electronic circuits. Therefore, single photon counters allow the use of entirely digital circuitry for signal recording and processing (unlike linear photodetectors which need signal digitization prior to any signal processing). The digital character of the output signal is a major advantage of single photon
detectors, since the associated digital circuitry does not add noise to the acquired signal. Thus, single photon counters have inherently high signal-to-noise ratio. The only unavoidable noise in single photon detector is the noise associated with the stochastic nature of the measured signal. This noise is proportional to the square root of the measured signal. For example, if the intensity of the signal is $10^6$ counts per second (c/s), the expected dispersion of this value is of the order of 1,000 c/s, providing a signal-to-noise (S/N) ratio of $\sim 1,000$. The minimum noise of a single photon detector is the noise of 10 to 20 c/s associated with its dark count.

Another important characteristic of single photon detection is its dynamic range. A large dynamic range allows simultaneous detection of very weak and very strong multi-color fluorescence signals. In principle, the dynamic range of a single photon counting systems is intrinsically very wide because electric pulses produced by such a counter in response to single photons are very short (less than $10^{-9}$s). Therefore the maximum count rate, which can be measured by a single photon PMT detector, is of the order of $10^8$ count/s.

The tremendous advantages offered by single photon counting became a strong motivating factor for the group at SUNY SB to design and build an efficient and inexpensive DNA sequencer based on this technology.

In order to test the sequencing performance of such an instrument, twenty single capillary sequencers (SBS-2000) were built and deployed at the SUNY SB Sensor CAT. This facility was used to carry out a number of sequencing tests in collaboration with the Cold Spring Harbor Genome Center. Below we describe the results of this testing using a clone from a canine EST library as a model system.

### 2. Materials and Methods

#### 2.1. DNA sequencer SBS-2000

Schematics and photographs of the SBS2000 instrument are shown in Fig.1.

![Fig. 1. Schematics and photograph of Stony Brook sequencer SBS-2000](image)

The instrument consists of three main modules: a fiberized laser illumination module, a single capillary separation/reading module (right photo) and a photoreceiving module. All modules are aligned at the assembly stage and then connected via standard optical fibers and electric cables.
The Laser illumination module is equipped with a fiberized laser working in continuous wave (CW) regime. Since the reading head has a fiber input for the excitation light, and achromatic optics, the laser wavelength can be selected and changed according to the user’s need without any misalignment of the reading head. In order to find the most efficient excitation wavelength for BigDye labeled DNA samples, different single-laser sources emitting at 488, 514 and 532 nm at the output power levels varied between 200 µW to 40mW were tested. It was found that a Nd-YAG laser (λ=532 nm) provides more efficient excitation than Ar-ion lasers (λ=488nm and 514 nm) at the same output power. The obtained transfer matrix for BigDye labeled samples illuminated by Nd-YAG laser was: \{A(0.22 0.83 0.58 1), G(0.07 0.44 1 0.36), T(0.38 1 0.005 0.1), C(1 0.3 0 0)\}.

A labeled DNA sample undergoes separation in a single-capillary, fiberized separation/reading module. This module comprises a miniature high voltage supply (up to 15kV) with a built in voltmeter and micro-ampermeter, tube-changer carousels carrying tubes with DNA samples and running buffers for capillary inlet and outlet, capillary holder, and the fiberized reading head.

The electrokinetic injection is carried out at 0.5 - 3 KV, and then the running voltage (8-15 KV) is applied to the capillary inserted into the fiberized reading head. Excitation light is delivered to the capillary via 62µm illumination fiber from a miniature fiberized 532nm Nd YAG laser. Excited fluorescence is collected by 200µm fiber-receiver and delivered to the photoreceiving module.

In the photoreceiving module the fluorescent signal purified from the laser light with either a notch or step filter undergoes time-division multiplexing, with a specially designed rotating filter wheel with four 10-20 nm, 3-4OD band-pass filters (OMEGA Optics) corresponding to the emission maxima of the four used dyes. After filtering, fluorescence signal is detected by a single photon detector (Hamamatsu H7464, H6240). An FPGA based photon counting circuit board is used to count and integrate data from the single photon detector while the light is passing through each of four filters, thus obtaining the signal amplitude in the four channels. These four values are transferred to the computer during each revolution of the wheel (~ 10-25 revolutions a second). During the separation process, sequencing traces are displayed in real time on the computer screen. After completion of the sequencing run, sequencing data processing can be performed either on the same computer or on any computer connected to the lab network.

2.2. Sequencing conditions

In all experiments described below BigDye sequencing reagents (BigDye™ Terminator v2.0 Cycle Sequencing Ready Reaction Kit, Part no. 4390253) were used. All samples were electrokinetically injected (3kV, 30 seconds) into 55cm, 75µm ID capillaries from PolyMicro coated according to the protocol described below. All runs were carried out at 8.5kV during ~2 hours at room temperature. Beckman CEQ separation polymer (CEQ TM Separation Gel - LPA I) was used.

Capillary Coating Protocol

In all experiments fused silica capillaries from PolyMicro with 75µm inner diameter were used. Capillaries were coated as below protocol:

Rinse the capillary with methyl alcohol (5 ml)
Rinse the capillary with distilled water (5 ml)
Rinse the capillary with 1M sodium hydroxide solution (5 ml) and leave the solution inside for 1 hr
Rinse the capillary with distilled water (5 ml) and leave the solution inside for 1 hr
Mix 4 µL MPS and 1 mL of 6M acetic acid. Fill the capillary with the solution and leave the solution inside for 1 hr
Rinse the capillary with distilled water (5 ml)

Prepare the coating solution:

Acrylamide 4%
TEMED 0.1%
APS 0.2%
Water 95.7%

Fill the capillary with this solution immediately and leave the solution inside for 1 hr

Rinse the capillary with distilled water (5 ml)

Fill the capillary with Beckman-Coulter polymer.

2.3. Dilution of sequencing standards

Big Dye Terminator Long Read Standards (ABI, part # 4307049) were purchased. The contents of a single tube was then resuspended in 500 ul of HiDye Formamide (ABI, part # 4311320)) as per the instructions provided. A portion of this was set aside and considered as full strength sample. The rest was then diluted in water to final dilution of 1:10, 1:25, 1:100, and 1/1000. Separate 10 ul aliquot of each sample was the denatured at 95° C for 5 min immediately before separation on the SB sequencer, ABI 3700, and the Amersham Pharmacia MegaBACE 1000.

2.4. Preparation and dilution of sequencing products

Several clones originating from the CSHL Genome Center were chosen at random. These clones were part of a BAC DNA shotgun library produced from 1.5-3Kb inserts ligated into pZERO-2 vector (Invitrogen). The insert was ligated into the multiple cloning site, flanked by M13 –21 forward and reverse primer binding sites. Plasmid DNA was isolated from these clones using a variation of an alkaline lysis procedure followed by a modified SPRI method (Hawkins et al, 1994). Approximately 200 ng of plasmid was used in a cycle sequencing reaction which contained 2 ul of Big Dye Terminator reagent (1/4 reaction), 5 pmoles of M13 –21 forward primer (Life Technologies), and 1 ul of 5X sequencing buffer (400 mM Tris-Cl, 80mM MgCl2) in a total volume of 10 ul. Reactions were cycled according to the manufacturer’s protocol. Sequencing products were purified by ethanol precipitation and re-suspended in 20 ul of HiDye Formamide . Samples were split into two 10 ul aliquots. One of these aliquots was used to dilute the sequencing products to a final dilution of 1:10, 1:50 and 1:100. Since the original reaction represented ¼ of the amount of reaction mix suggested by ABI these dilutions of the reaction represent an amount of material that is 1/40, 1/100, and 1/400 respectively, what is recommended by ABI. The undiluted sample as well 10 ul of each of the dilutions were denatured at 95° C for 5 min and immediately run on the SBS-2000 Sequencer.

2.5. Sequencing of a gene from a canine EST library

A canine EST library was previously constructed at CSHL (Greg Hannon, personal communication). The canine cDNAs were directionally cloned into pBluescript II KS(+) vector (Stratagene). The cDNA sequence was flanked by M13 –21 forward and reverse primer binding sites. A random clone was chosen from this library. Plasmid DNA was isolated as described above. The 5' and 3' ends of the cDNA insert were sequenced using the M13 –21 forward and reverse primers using 2 ul of Big Dye Terminator reagent (1/4 reactions), 5 pmoles of primer and 1 ul of 5X sequencing buffer (ABI) in a total volume of 10 ul. The sequencing products were precipitated and resuspended in 20 ul of HiDye Formamide. The sample was then diluted 1:10, 1:50 and 1:100 with water, denatured for 5 min at 95° C and then run on the SB sequencer. The trace data were then deconvoluted and base-called. A
A cantilever primer walking strategy was utilized. Primers were picked manually from both the forward and reverse traces to extend into the cDNA sequence from either end. Subsequent primer sites were chosen approximately 350 bases away from the beginning of the trace. The new sets of primers were resuspended in water and 5 pmoles were used in the next round of sequencing reactions. The process was then repeated twice more until the reads overlapped sufficiently.

The same clone was also sequenced in the finishing group at the CSHL Genome Center by experienced sequencers using standard procedures for high quality results (coverage rules, expert manual editing and quality checking, etc.). This group used standard ABI instruments for sequencing. The overlaps and trace data from this sequencing are available as supplemental material.

2.6. Processing of SBS-2000 traces using PHRED software

Algorithms used in PHRED for base-calling and quality values assignment are discussed in detail in [B. Ewing et all, 1998]. Below we describe preprocessing procedures used for preparation of SBS-2000 data to be accepted by PHRED.

The SBS-2000 produces 10 to 25 data samples per second that results in 50 to 125 sampling points per 5 second width peak. Data preprocessing includes noise filtering, peak smoothing, base-line removal, cross-talk filtering, mobility shift correction, data decimation and conversion to the SCF format.

Recorded data set is first processed using a simple low-pass MA filter to reduce noise. The least squares approximation of data by polynomial of the 3rd degree [A. Savitzky and M.J.E. Golay, 1964] is used for peak smoothing.

For base-line removal and cross-talk filtering we use the techniques described in [Z. Yin et all, 1996]. The base-line is removed from each trace separately. The cross-talk matrix is calculated for each dye/machine combination and stored in the configuration file.

Mobility shift correction used for SBS-2000 data processing is based on equalization of spacing between adjacent peaks. The method is similar to the algorithm described in [M.C. Giddings et all, 1998] and uses non-linear scaling of traces in spatial domain to compensate differences in dye mobility. The algorithm does not require prior calibration or manual fitting of shift functions. Mobility shift correction procedure independently scales traces trying to equalize peak spacing keeping instant average peak spacing intact. The procedure scales traces gradually in a number of steps starting from the middle of the run where effect of mobility shift is minor and resolution of sequencing is still good, and moves toward the beginning of the sequencing run where mobility shift is the most severe. Applied to relatively small shifts on each step, the strategy of gradual spatial scaling allows compensation of a substantial mobility shift over the entire sequencing run.

Further processing steps include data decimation to sampling rate of approximately 10 data points per peak. Resulting file is converted to SCF v2.00 format.

Since in our sequencing experiments we used ABI sequencing chemistry, we carried out base calling using PHRED software trained for the ABI-3700. In order to check our data processing procedure we applied it to multiple sequencing runs with known sequence and found fairly good agreement between known sequences and PHRED base-calling. We also compared quality factors calculated by the PHRED software with behavior of peak width/peak spacing curves. We found that the quality of the base-calling rapidly decreases when the peak width approaches peak spacing and drops below 98.5% when peak width and peak spacing become equal. The same behavior was consistently observed for all processed data files.
3. Results and Discussion

3.1. Detection of diluted sequencing standards

The separation and resolution capability of the SBS-2000 Sequencer was initially tested using commercially available sequencing standards. Dilutions were made as described in the Materials and Methods section. The tests were carried out on using 20mW, 532 nm Nd-YAG laser. Identical regions of three sequencing runs of Big Dye labeled DNA sequence standard are shown in Fig.2.

![Electropherogram](image)

Fig. 2. Fragments of sequencing runs of DNA Controls/Standards (Big DyeTM Terminator Sequencing Standard, part # 4304154).

Full strength sample (upper panel), 1:100 dilution (middle panel), and 1:1000 sample(lower panel) diluted in water. Run conditions: separation medium – Beckman CEQ separation polymer, 50cm, 75µm ID capillary from PolyMicro coated according to the protocol described in the previous section, room temperature, 30 seconds injection at 3 kV, run voltage 8 kV. Data was processed with PHRED base calling software.

Note, that the amplitudes of the peaks in the electropherograms do not scale with the dilution. While the content of the labeled material in the sample tubes differed by a factor of 1000, the amplitudes differed only by a factor of about sixty. We believe that this is due to the enhancement of the injection efficiency caused by an increase of the electric field in the injection tube when highly resistive water is added. Our preliminary data indicate that the separation quality of diluted samples can be improved by optimizing the sample injection conditions. Moreover, we often observed significant improvement of the sequencing quality in highly diluted samples, probably due to impurity dilution by the addition of a relatively large volume of distilled, de-ionized water.

![Quality factor plot](image)

Fig. 3. Quality factor calculated by PHRED software for full strength and diluted DNA Controls/Standards (Big DyeTM Terminator Sequencing Standard, part # 4304154)

Data read length depended on the degree of the sample dilution. At 98.5% accuracy the readlength varied from 550bases for full strength sample to ~500 bases for 1:1000 diluted samples (see Fig.3).
Dilutions higher than 1:40, when run on the ABI 3700 and the MegaBACE 1000 failed to produce any signal above baseline (data not shown).

3.2. Sequencing of a gene from a canine EST library.

A canine EST clone was isolated and the 5’ and 3’ regions sequenced using the –21 M13 forward and reverse universal primers. The signal strength was of such magnitude that serial dilutions of the sequencing reactions were made. A 1:100 dilution of the forward and reverse reactions were separated on the SB Sequencer.

The traces were analyzed using a base-caller in development at Stony Brook and confirmed manually. From these base-called traces, primers were picked to both walk further into the sequence. These new primers were used in a second round of sequencing reactions, again at 1:100 dilution. Another set of primers were picked from these traces and another round of sequencing reactions were performed. These sequences were aligned to confirm that there was an overlap of sequences originating from the 5’ and 3’ end reads and to completely cover the gene (Fig.4).

Fig. 4. Alignment of 1:100 diluted reactions that span the cytochrome C oxidase, subunit I gene. Dotted lines indicate the demarcation between the separate reads
The FASTA files generated from these separate traces were also aligned manually to re-create the gene sequence. The same clone was sequenced using 1/2 BDT reactions on the ABI 377 to high accuracy by the CSHL Genome Center finishing group. The sequence was then used in a BLAST search at the NCBI site. The traces obtained from the both the ABI instrument and the SB sequencer showed 99% identity to a canine cytochrome oxidase subunit I gene. We then compared the ABI 377 and the SBS-2000 Sequencer traces to each other. When the ABI and the SB sequencer FASTA files were aligned it was observed that there were two extra bases in the SB sequencer read. These two areas were then located on the actual traces and it was found that these extra bases were erroneously identified as bases by the base-calling software which is still in development. After, these bases were removed from the FASTA file the ABI 377 sequence and the SB Sequencer sequence showed 100% identity. Two additional canine EST clones were sequenced in this manner by the SB sequencer to ensure applicability of the instrument (data not shown).

Thus, we accurately sequenced a small gene using the SBS-2000 sequencer with as little as 0.25% of the fluorescent labelled product that is recommended by ABI.

4. Concluding remarks

We have described the operational principles, design and performance of a novel DNA sequencing instrument based on the single photon counting fluorescence detection technique. The instrument has a modular architecture. It comprises several robust and compact components connected with standard electronic cables and optical fibers. The modular architecture provides for flexibility in design and assembly of instruments for different applications, which may vary considerably in their sensitivity, throughput, size and cost.

Using BigDye sequencing chemistry and various DNA samples we demonstrated significantly higher sensitivity of the SBS 2000, compared to those offered by commercial instruments. For highly diluted sequencing samples we demonstrated read lengths exceeding 500bp at 98.5% accuracy at room temperature using commercial sequencing chemistry. Visually, we were able to distinguish peaks up to fragment size of ~650-700 bases. These results also demonstrate the theoretical reduction of DNA template may parallel the reduction of sequencing reagent. Typically at the Cold Spring Harbor Genome Center, a ¼ Big Dye Reaction contains 200 ng of template DNA. If this reaction were diluted 1/100 it would allow for sequencing with as little as 2 ng of template DNA.

The detection of such small amounts of sequencing product implies the successful application of the SBS 2000 on traditionally difficult templates which contain very low copies of primer binding sites (e.g. BAC, PAC and genomic DNA). The sensitivity of the SBS 2000 to discriminate between background and signal also implies that it would be an ideal platform to detect SNPs (single nucleotide polymorphisms) with a high degree of accuracy. We are curently testing both the above applications on the instrument.

Our detection technology and the instrument design are ideally suited to the implementation of automated high-throughput sequencing machines and systems. Several challenges remain in the development of this instrument. The most challenging task will likely be the implementation of robust base calling software, which we are actively pursuing. Another goal is the implementation of a 32-capillary automated DNA sequencer. This next generation device will also incorporate automated temperature control, sample injection and polymer replacement capabilities. We believe that the high sensitivity and dynamic range of our instruments in conjunction with emerging techniques for low
volume DNA sample preparation will lead to a very significant reduction of the sequencing cost and open possible horizons in template variety that can be sequenced using fluorescent sequencers.

References


