

**Genome Sequencing Technology:
Improvement of the Electrophoretic Sequencing Process and
Analysis of the Sequencing Tool Industry**

by

Kazunori Maruyama

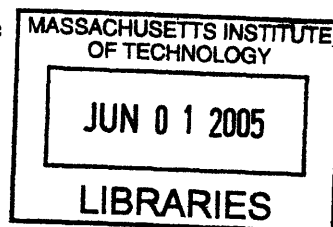
Ph.D. in Material Science, Mie University (2001)
M.S. in Macromolecular Science, Osaka University (1993)

Submitted to the Alfred P. Sloan School of Management and the Department of Chemical
Engineering in Partial Fulfillment of the Requirements for the Degrees of


**Master of Business Administration
and
Master of Science in Chemical Engineering**

In Conjunction with the Leaders for Manufacturing Program at the
Massachusetts Institute of Technology
June 2005

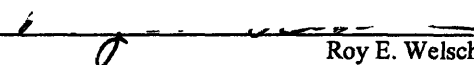
©2005 Massachusetts Institute of Technology. All rights reserved.



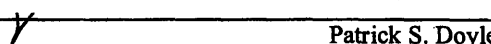
Signature of Author


Alfred P. Sloan School of Management
Department of Chemical Engineering
May 6, 2005

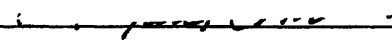
Certified by


Roy E. Welsch
Professor of Statistics and Management Science
Thesis Advisor

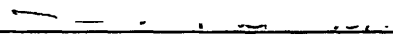
Certified by


Patrick S. Doyle
Assistant-Professor of Chemical Engineering
Thesis Advisor

Accepted by


Margaret Andrews
Alfred P. Sloan School of Management
Executive Director of Masters Program

Accepted by


Daniel Blankschtein
Department of Chemical Engineering
Graduate Committee Chairman

ARCHIVES

**Genome Sequencing Technology:
Improvement of the Electrophoretic Sequencing Process and
Analysis of the Sequencing Tool Industry**

by

Kazunori Maruyama

Ph.D. in Material Science, Mie University (2001)

M.S. in Macromolecular Science, Osaka University (1993)

Submitted to the Alfred P. Sloan School of Management and the Department of Chemical Engineering in Partial Fulfillment of the Requirements for the Degrees of

**Master of Business Administration
and
Master of Science in Chemical Engineering**

ABSTRACT

A primary bottleneck in DNA-sequencing operations is the capacity of the detection process. Although today's capillary electrophoresis DNA sequencers are faster, more sensitive, and more reliable than their precursors, high purchasing and running costs still make them a limiting factor in most laboratories like those of the Broad Institute. It is important to run those sequencers as efficiently as possible to reduce costs while producing robust assemblies. Polymer media for electrophoresis is the most important determinant for sequencing throughput. This thesis investigates the effect of polymer media on the performance of Applied Biosystems (ABI) 3730xl, the de-facto standard of DNA sequencers and develops analysis procedures for ABI3730xl system and its data.

Due to its use in the human genome project (HGP), ABI has established a monopolistic position in the DNA-sequencing tool industry. As the de-facto standard of DNA sequencers ABI3730xl is highly automated, well-optimized, and black-boxed, despite the importance of higher throughput sequencing for diagnostic applications, third parties have found it difficult to improve sequencing methods. This thesis also conducts an analysis of the DNA-sequencing tool industry to discuss how ABI has established current monopolistic status, what kind of business model would be attractive for ABI in the post-HGP period, how new companies can successfully enter this industry, and how they can keep improving DNA-sequencing throughput along the line of "Moore's law".

Thesis Supervisors:

Roy E. Welsch

Professor of Statistics and Management Science

Patrick S. Doyle

Assistant Professor of Chemical Engineering

Table of Contents

Acknowledgements	5
1. Introduction	6
2. Background	8
2.1 Broad Institute Overview	8
2.2 Genome-sequencing Operation Overview	11
2.2.1 Operation Process	11
2.2.2 Genome-sequencing Technology	13
3. Improvement of the Electrophoretic Genome-sequencing Process	16
3.1 Introduction	16
3.2 Research History of Polymer Media for DNA-sequencing	18
3.2.1 Linear Water-Soluble Polymers	18
3.2.2 Thermo-Sensitive Polymers and Polymers with Higher-order Structure	22
3.2.3 Summary of Past Polymer Research	24
3.3 Experimental Section	25
3.3.1 Selection of Polymers	25
3.3.2 Capillary Electrophoresis	25
3.3.3 Chemical Analysis	26
3.4 Results and Discussions	28
3.4.1 Establishment of Experimental Protocol for Polymer Research	28
3.4.2 POP7	35
3.4.3 DMA	41
3.4.4 LPA	44
3.4.5 Chemical Analysis of POP7	47
3.4.6 Polymer Mixtures	52
3.5 Possible Approaches for the Future	55
3.6 Summary	57

4.	Analysis of the Genome-Sequencing Tool Industry	58
4.1	Introduction	58
4.2	History of Genome-Sequencing Technology	58
4.3	Emergence of the Genome-Sequencing Tool Industry	63
4.4	Analysis of the Genome-Sequencing Tool Industry	65
4.4.1	Industry S-curves	65
4.4.2	Uniqueness and Complementary Assets to Create and Capture Value	66
4.4.3	Industry Structure – Present and Future	70
4.4.4	Value Chain and Boundaries	73
4.4.5	Strategy for the Future of ABI	74
4.5	Summary	75
5.	Conclusion	76

Acknowledgements

The author wishes to acknowledge the Leaders for Manufacturing Program for its support of this work.

He would also like to thank Robert Nicol, Patrick Cahill, James Meldrim, Karen Foley and the group of Technology Development at the Broad Institute.

Finally, he would like to thank his wife, Naoko, for her continuous support during his busy but fun days in Cambridge.

1. Introduction

Huge technological advancements in information technology and a deeper understanding of biomaterials in the past twenty years have allowed for the discovery of the entire genetic information of humankind. The Whitehead Institute Center for Genomic Research was a primary contributor to this achievement. In the 1990s, during the Human Genome Project, its sequencing operation played a key role in rapidly improving genome-sequencing throughput. The Whitehead Institute has recently combined research efforts with the Massachusetts Institute of Technology (MIT) and Harvard University to form the Broad Institute.

Even after the completion of the Human Genome Project in 2003, increased sequencing throughput at an affordable cost is critical to the realization of advanced and economical genomics-based medications. However, without tangible economical returns in the near future, research institutes and related industries in the genome-sequencing arena are evidently frustrated. Not only is the next incremental technological step necessary to move the field forward but also an appreciation for past and potential future balances of power in the arena would be enough to drive such a highly technology-driven industry toward future medical applications.

The Broad Institute has improved throughput of genome-sequencing operations with an eye toward future healthcare applications. A primary bottleneck of the current sequencing operation is the capacity of the sequence detection process using the “Genome Sequencer,” a key instrument in reading base code sequence for DNA. Many scientists and engineers in the Broad Institute are conducting research not only on multicapillary electrophoresis, the current sequencing technology, but also on next-generation technologies, to increase the capacity of the sequence detection process.

This thesis is based on a Leaders for Manufacturing (LFM) internship at the Broad Institute, where the author worked with the genome sequencer in terms of both technological and business aspects. The author investigated polymer media of the current genome sequencer to improve its operation efficiency. He also conducted technology and

business analyses for the genome-sequencing tool industry and made attempt to forecast the future of the industry.

The thesis proceeds as follows:

Chapter 2, “Background,” provides information on the Broad Institute as well as modern genome-sequencing technology. It concludes with a discussion of the detection phase of the genome-sequencing process, providing insight into the multicapillary electrophoresis procedure.

Chapter 3, “Improvement of Electrophoretic Genome-sequencing Process,” consists of the engineering portion of this thesis, investigating possibilities for improving the efficiency of current genome-sequencing. The discussion focuses on the role of polymer media in multicapillary electrophoresis technology.

Chapter 4, “Analysis of the Genome-sequencing Tool Industry,” examines the business perspective of genome-sequencing technology with an analysis of research history and the emergence of the genome-sequencing tool industry. The discussion includes how the monopoly status of Applied Biosystems, Inc. (ABI) has been successfully established, whether ABI’s dominance will continue in the future, and how the industry can be developed in the future.

Chapter 5, “Conclusion,” ties the discussion of the preceding chapters together and offers some perspectives on how conclusions can be applied outside the field of genomics.

2. Background

2.1 Broad Institute Overview

The Broad Institute is the collaborative research organization between MIT, Harvard University, and the Whitehead Institute; its official name is “The Eli and Edythe L. Broad Institute of MIT, Harvard University and its affiliated hospitals, and Whitehead Institute.” In June 2003, philanthropists Eli & Edythe Broad of Los Angeles gave \$100 million to create an institute with MIT, Harvard, and Whitehead to fulfill the genome's promise for medicine. The institute began its activity later that same year. Figure 1 shows the mission statement of the Broad Institute.

- Our scientific mission is:
 - ☛ To create tools for genomics medicine and make them broadly available to the scientific community
 - ☛ To apply these tools to propel the understanding and treatment of disease

- Our organizational mission is:
 - ☛ To enable collaborative projects that cannot be accomplished solely within the traditional setting of individual laboratories
 - ☛ To empower scientists through access to cutting-edge tools

Figure 1 Mission Statement of the Broad Institute¹

The Broad Institute consists of two kinds of organizations called “Programs” and “Platforms.” Programs are scientific research groups that focus on specific problems, led by researchers called Principal Investigators. Four Programs currently exist; they are Cell Components, Chemical Biology, Cancer Genomics, and Medical and Population Genomics. For example, the Cell Components Program, where Dr. Eric Lander is the Principal Investigator, focuses on identifying and monitoring functional components of the cell and applying this information to the basic understanding of human diseases.

¹ <http://www.broad.mit.edu/>

Platforms are responsible for the operation and development of Broad's common technology bases. "Program" research projects often include "platform" researchers and engineers to access Broad's shared technologies. Out of the four technological platforms, the author spent his internship at the Genome-sequencing and Analysis (GS&A) platform, which is one of the genome-sequencing centers with the world's highest throughput and its responsible for the implementation and technological development of Broad's genome-sequencing process.

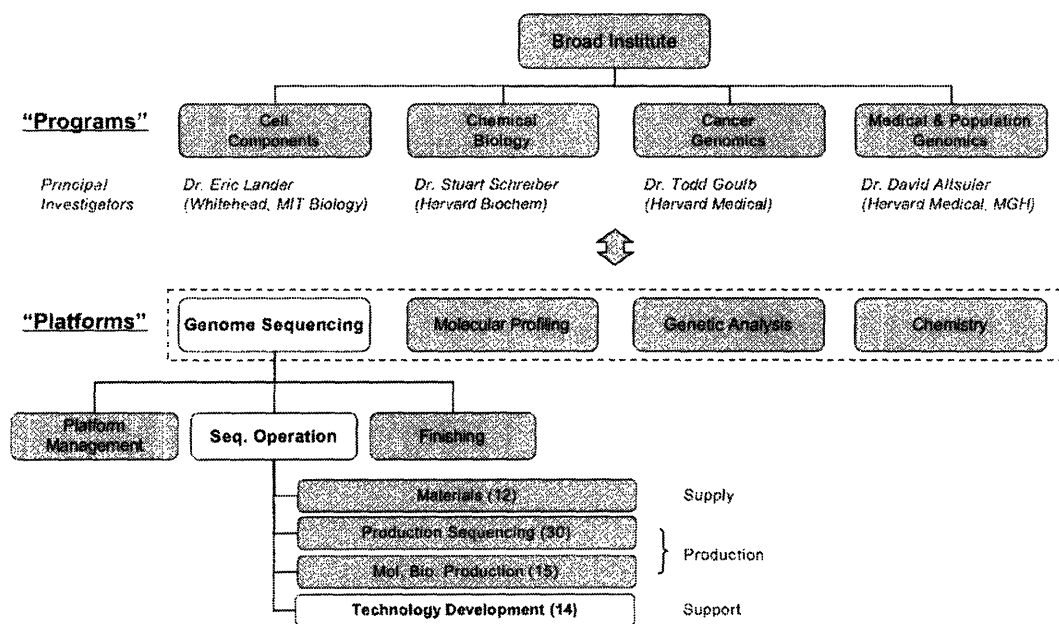


Figure 2 Organizational structure of the Broad Institute

The author conducted his internship at the Technology Development Group, or "Dev," under the Sequencing Operation Department of GS&A. Figure 2 shows the organizational structure of the Broad Institute. Dev is part of the technical support team for the Production Sequencing group and the Molecular Biology Production group. Dev does not conduct cutting-edge research but rather solves problems around the current DNA-sequencing process.

Improving the throughput of genome-sequencing and reducing the cost of genetic information are still critical to the future of genomics medication or tailored medication. As much as the Human Genome Project has dramatically reduced the cost of sequencing, the cost to sequencing one human’s DNA still remains around \$10 million, as shown in Figure 3. One of Dev’s key tasks is to reduce the sequencing cost so that it is affordable for individual medication service. Thus, Dev currently addresses a wide range of research projects in order to improve sequencing throughput – from the optimization of running conditions to the evaluation of next-generation technology developed by technology start-up companies.

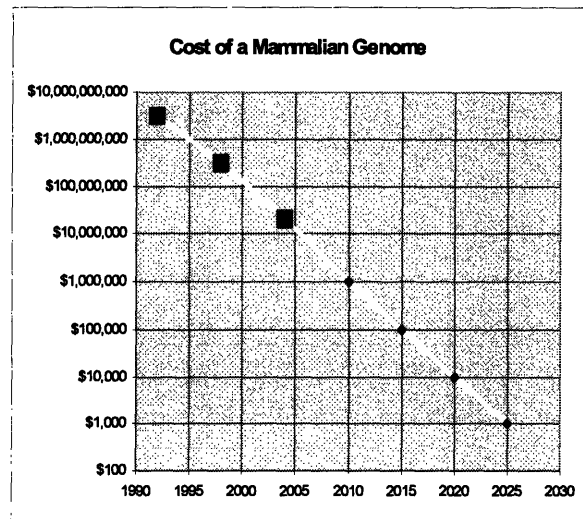


Figure 3 “Moore’s law” applied to genome-sequencing technology²

² Courtesy of Rob Nicol, Director of Sequencing Operations department, Broad Institute

2.2 Genome-sequencing Operation Overview

2.2.1 Operation Process

Figure 4 shows three major steps in the DNA-sequencing operation at the Broad Institute – molecular biology, core sequencing, and finishing. The Sequencing Operation department oversees the molecular biology process, handled by the Molecular Biology Production group, and core sequencing process, performed by Production Sequencing group, while the Finishing department is responsible for the finishing process. (See Figure 2 for the organizational structure of the Genome-sequencing platform.)

The molecular biology process consists of three sub-processes: DNA preparation, ligation, and transformation. In DNA preparation, a DNA sample, which typically includes three billion base pairs (bp) for a human genome, is enzymatically chopped up into around four-thousand bp fragments. These fragments, called “plasmids,” are embedded in the circular chromosome of *E. Coli* and allowed to proliferate on agar plates during ligation and transformation processes. The multiplied plasmids are then extracted from the *E.Coli* cell and purified to prepare a sample for the succeeding processes called sequencing reaction and detection, during which the DNA code is read by the electrophoretic technique. Schematic diagrams for the molecular biology process and the core sequencing process are shown in Figures 5 and 6 respectively.

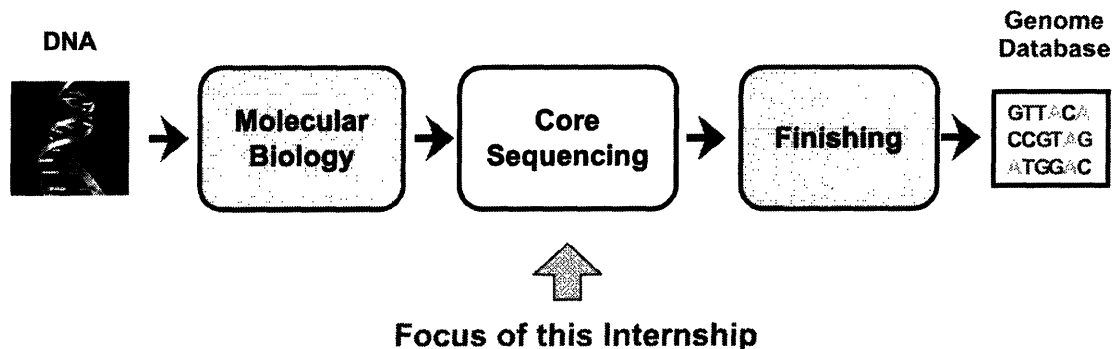


Figure 4 Three major processes in the DNA-sequencing operation

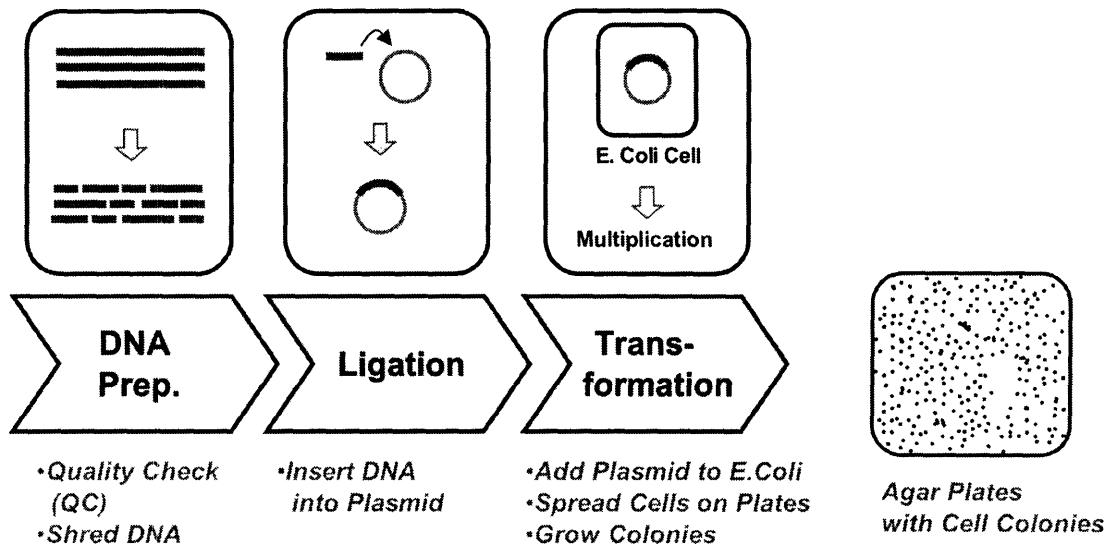


Figure 5 Schematic diagram of the molecular biology process

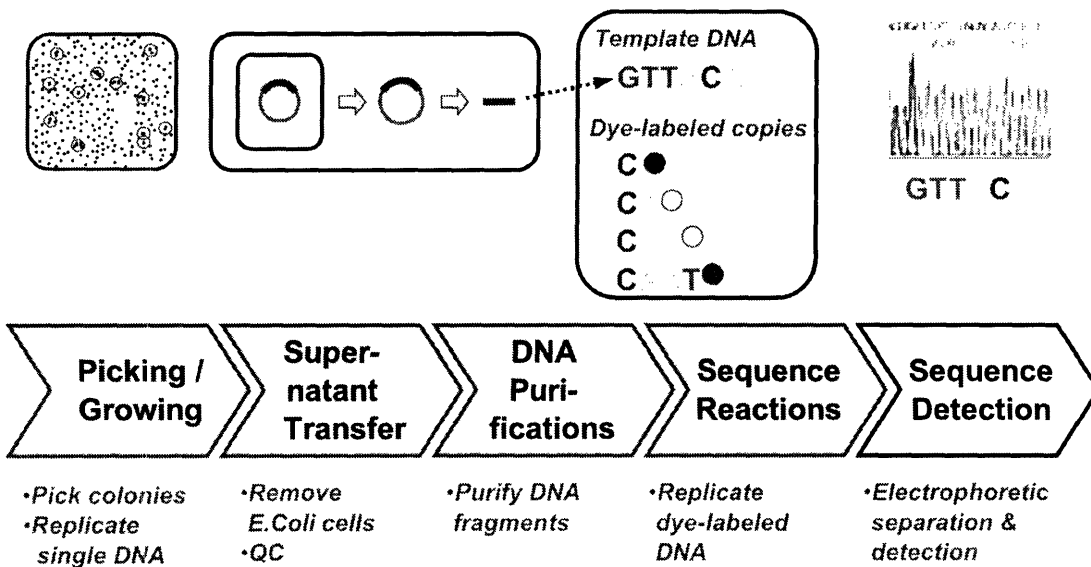


Figure 6 Schematic diagram of the core sequencing process

2.2.2 Genome-Sequencing Technology³

The currently-accepted genome-sequencing procedure is based on the Sanger method, which was developed between 1964 and 1977 by Professor Fred Sanger of the University of Cambridge. After some technological improvements, the modified Sanger method was established in late 1990s.⁴ Due to its use of reaction enzymes and Dideoxyribotriphosphoric acid (dd-NTP), the Sanger method is also called the “Enzyme method” or the “Dideoxy method.”

Figure 7 illustrates the basic idea behind the DNA-sequencing process that is based on the modified Sanger method. This process follows two steps:

- Prepare various lengths of complementary DNA (c-DNA) fragments using the original single-strand DNA as a template.
- Read out the base code of the replicated c-DNA to learn the base sequence of the original DNA sample.

In the “sequence reaction” process, polymerase chain reaction (PCR) is used to duplicate multiple c-DNA fragments.⁵ In the Sanger Method, a small amount of reaction terminators (dd-NTP) are used with reaction monomers (Deoxyribotriphosphoric acid, d-NTP) for PCR. Both monomers and terminators are comprised of nucleic acids with four different bases, i.e. adenine (A), cytosine (C), guanine (G), and thymine (T). Each type of dd-NTP is labeled with different color dyes through chemical modification. Polymerization continues while d-NTP is attached to the end of an active c-DNA chain. The reaction terminates once dd-NTP reacts instead of d-NTP. Finally, various lengths of c-DNA fragments, capped with either an A, C, G, or T-type of dd-NTP, are synthesized.

³ *Encyclopedia for Biotech and Genomics (in Japanese)*; Mitsubishi Research Institute; Mitsubishi Kagaku Institute of Life Science, Eds., Toyo Keizai Inc.: Tokyo, 2004.

⁴ More history of the development of the DNA-sequencing technique is discussed in Chapter 4.

⁵ This experimental method proliferates a particular region of a small amount of DNA molecules using a heat-stable enzyme, called DNA polymerase. K. B. Mullis, a researcher at the Cetus Corporation, invented PCR. He was awarded the Nobel Prize in Chemistry in 1993 for this achievement, only seven years after he first published his ideas.

“Sequencing detection” is the process by which the labeled c-DNA fragments are read out one by one, from short to long. The technique called electrophoresis is used for this procedure. Figure 8 shows the basic principle behind DNA electrophoresis. The multiplied c-DNA fragments are injected into a capillary filled with water-soluble polymers to be used as a sieving media. The c-DNA fragments are charged with electricity; under an electric field, they travel through the capillary toward the other end. Since shorter fragments experience less resistance from the polymer media, they travel through the capillary faster than the longer ones. At the end of the capillary, a laser beam excites the labeling dye and the luminescent intensity for each of the four colors is recorded as a function of time. In the current sequencing procedure, the instrument called “DNA sequencer” automatically carries out this process. Figure 9 shows the external appearance of the de facto standard of DNA sequencers, ABI3730xl.

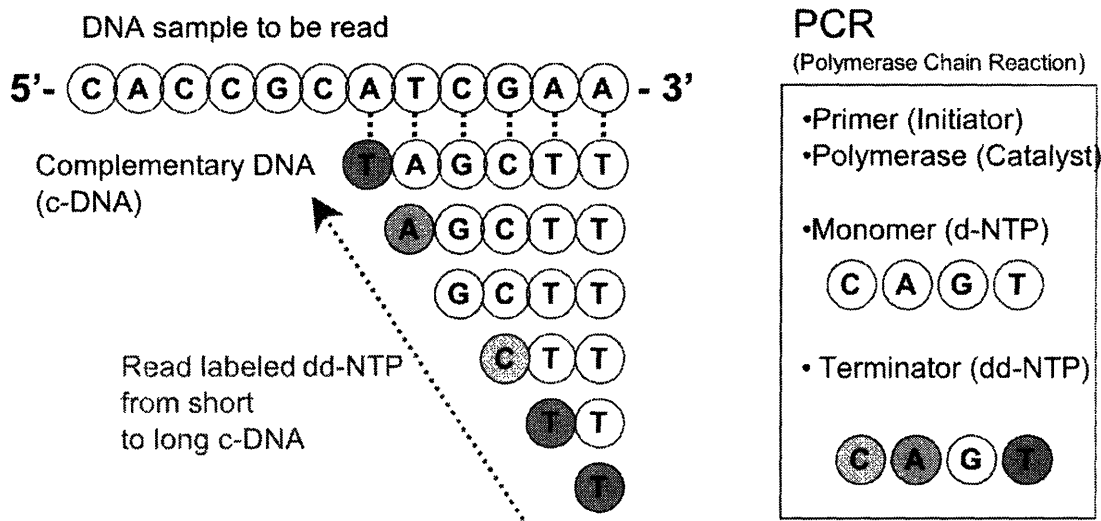


Figure 7 Outline of the DNA-sequencing procedure using Sanger’s method

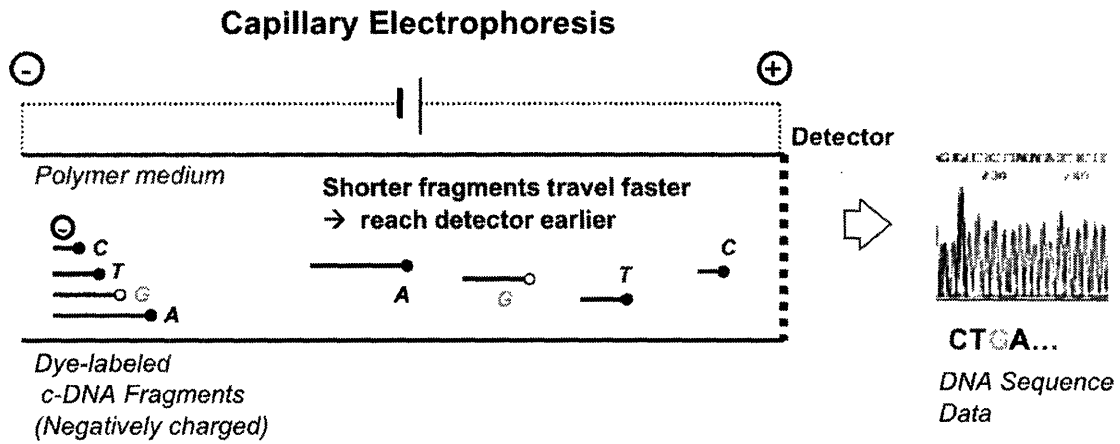


Figure 8 Outline of genome-sequencing with multicapillary electrophoresis

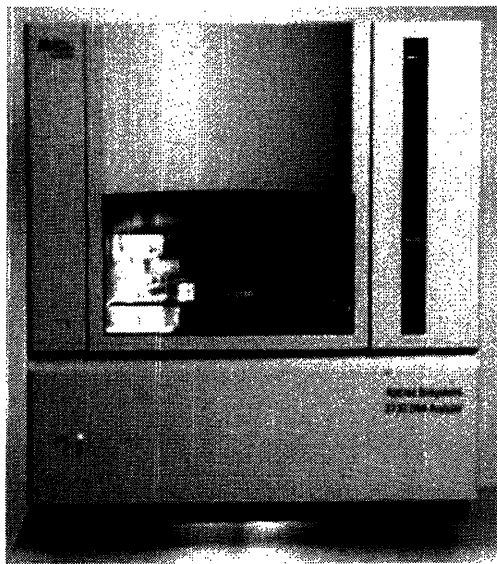


Figure 9 External appearance of ABI3730xl

3. Improvement of the Electrophoretic Genome-sequencing Process

3.1 Introduction

Throughput of DNA-sequencing operations has been vastly improved during the past fifteen years. It has become high enough to finish the human genome project a few years earlier than originally scheduled. Nevertheless, improvement should continue with an eye toward diagnostic and pharmaceutical applications.

The capacity of the sequence detection process is a primary obstacle to the sequencing operation. Although today's multicapillary electrophoresis DNA sequencers are faster, more sensitive, and more reliable than their precursors, high economic and operational costs still make them prohibitive for most laboratories like those of the Broad Institute's. It is important to run the sequencers as efficiently as possible in order to reduce costs while still producing robust assemblies. For instance, the cost to purchase the ABI3730xl is \$380,000 per unit. Designated consumables, such as polymers and buffers, drive operational costs even higher.⁶ The Broad Institute operates ninety-six units of ABI3730xl twenty-four hours a day throughout the year. Nevertheless, sequencing detection remains the bottleneck in the total operation process.

The main goal of this project is to improve the operational efficiency of ABI3730xl. Since this instrument employs capillary electrophoresis to sequence DNA, both data quality and operational efficiency depend completely on the performance of electrophoresis. Past research showed that there is a trade off in the relationship between migration speed (sequencing time) and data resolution. The faster the sequencing would run, the lower the quality in data resolution. Operational conditions, such as electric field and temperature, also affected this relationship. Although changing the polymer media is an expected and effective way to solve the poor ratio between speed and resolution, it has never been examined at Broad. Hence, the author tried to resolve possible further improvements to the sequencing efficiency by replacing POP7, the default polymer media of ABI3730xl.

⁶ For example, the Broad Institute annually spends more than \$400,000 to purchase the polymer, POP7, from ABI.

Before the examination of various polymers began, however, some issues required clarification. Since no one at Dev was familiar with previous research on polymer media in DNA electrophoresis, it was necessary to examine the research history. In order to use a different polymer on the 3730xl, the instrument's ability to work even with polymers quite different from POP7 had to be assured. The ingredients of POP7 also had to be identified for comparison. Thus, this project followed these four steps:

1. Examine past research efforts on polymer media
2. Establish a new experimental protocol to handle other polymers
3. Analyze the chemical composition of ABI's default polymer (POP7)
4. Run genome-sequencing using several polymer materials and study their performance

3.2 Research History of Polymer Media for DNA-sequencing

Much research on polymer media in DNA capillary electrophoresis was conducted in the late 1990s. Starting with linear polyacrylamide, researchers dealt with various water-soluble polymers as well as more complicated polymers.

3.2.1 Linear Water-Soluble Polymers⁷

To obtain a longer read length, the following three characteristics should be considered when choosing a polymer. First, high hydrophilicity allows polymer chains to expand sufficiently in water and to help forming a robust polymer network. Second, high molecular weight also allows polymer chains to form a polymer network robust enough to filter even long DNA fragments. Third, high polymer concentration is required to make the pore size of the network small enough and to give better peak resolution of shorter fragments. Figure 10 illustrates these three requirements for polymer media in DNA electrophoresis.

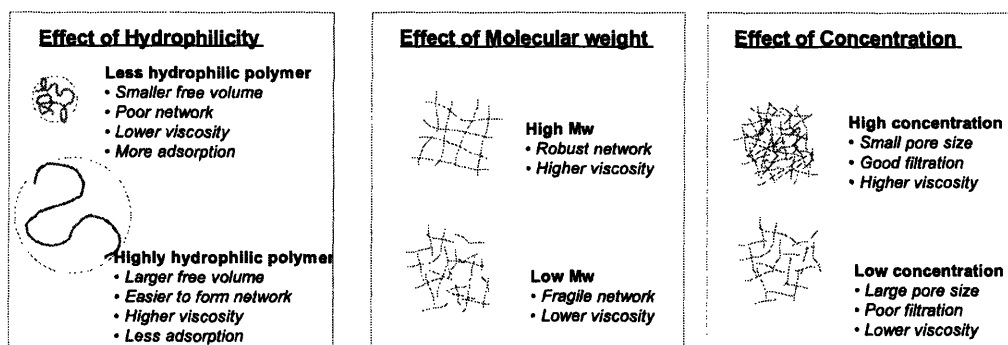


Figure 10 Relationship between polymer conformation and solution property

However, some trade-offs exist with these three features. Polymers that fulfill all three requirements make highly viscous solutions, which are difficult to inject into micro-capillaries. Sequencing throughput decreases with concentrated polymer solutions,

⁷ Albarghouthi, M.N.; Barron, A.E., *Electrophoresis*. 2000, 21, 4096-4111

since the migration speed of DNA fragments becomes slower. Additionally, polymer chain expansion due to high hydrophilicity does not have the benefit of being self-coating. Bare fused-silica capillaries cannot be used without pre-coating the interior with a polymer because arising electro-osmotic flow lowers the peak resolution. By adsorbing more onto the capillary interior and eliminating the cost-consuming pre-coating process, less hydrophilic polymer is desirable.

To solve for such trade-offs, various water-soluble polymers have been tested in DNA-sequencing. The choice of polymer was often arbitrary and empirical, primarily because the mechanism of DNA separation in uncross-linked polymer solutions is not fully understood. Table 1 lists the chemical structures of the major polymers that have been tested for this application.

Linear Polyacrylamide, or LPA, is considered one of the best performing polymers in terms of long read-length and short separation time due to its high water solubility and moderately attractive interaction with DNA molecules. However, LPA's high viscosity is a disadvantage in the injection process. In addition, since LPA does not have self-coating properties, the capillary manufacturing cost increases because of the need to pre-coat. Optimal results were obtained with a polymer media comprising 0.5% 270kDa LPA and 2% 17MDa LPA. This formulation delivered a read-length of 1,300bp at 98.5% accuracy in 2 hours at 70°C and 125V/cm.⁸ LPA is commercially used with the LongRead™ Matrix for Amersham Biosciences' MegaBACE™ sequencer.⁹

⁸ Zhou, H.; Miller, A.; Susic, Z.; Buchholz, B.; Barron, A. E.; Kotler, L.; Karger, B. L., *Anal. Chem.* **2000**, *72*, 1045-1052

⁹ Amersham Biosciences Website,
http://www4.amershambiosciences.com/aptrix/upp01077.nsf/Content/na_homepage

Table 1 Major water soluble polymers used for DNA-sequencing by capillary electrophoresis ⁷

Polymer	Chemical Structure
LPA	
PDMA	
PEO	
PVP	
PEG (End-capped)	$F_{2m+1}C_m-(O-CH_2-CH_2)_n-C_mF_{2m+1}$ <p style="text-align: right;">m = 6 or 8</p>
Poly(AAP)	
HEC	

Poly(N,N'-dimethylacrylamide), or DMA, is slightly more hydrophobic than LPA. Thus, DMA's general capacity as a capillary electrophoresis polymer is inferior to that of LPA. On the other hand, even bare silica capillaries can be used with DMA, since it can adsorb onto the inner wall of capillaries. Ramakrishna Madabhushi of ABI eagerly studied DMA as a DNA-sequencing matrix, focusing on this self-coating feature.¹⁰ ABI holds a basic patent on any water soluble polymers with self-coating features.¹¹ Since ABI declared that POP, the polymer matrices for its sequencing instruments, has self-coating features, it has been widely believed that POP consists of DMA. However ABI has never released any part of POP's ingredients. Excerpts from related review articles state, "The 3700TM instrument uses a separation matrix that is based on relatively low molar mass, linear polydimethylacrylamide, commercially known as POP,"⁷ and "PDMA ... is also widely used (POP4, 5, 6, Applied Biosystems)."¹²

The best results for DMA were up to 800 bases with a resolution limit of 0.5 (and 1,000 bases with a resolution limit of 0.3) and migration time of 96 minutes, which was achieved by using 2.5% w/v polymer, 150V/cm separation electric field, and 60 cm effective separation length at room temperature.¹³

Other water soluble polymers listed in Table 1 have electrophoresis performance inferior to both LPA and DMA. For instance, while PEO gives a relatively long read-length, seven hours were required to achieve the separation, which is very long compared to the one hour run time needed for LPA to achieve the same read-length.¹⁴ PVP and HEC only give read-lengths shorter than that of LPA.

3.2.2 Thermo-Sensitive Polymers and Polymers with Higher-order Structure^{15,12}

¹⁰ Madabhushi, R. S., *Electrophoresis*. **1998**, *19*, 224-230.

¹¹ Madabhushi, R. S.; Menchen, S. M.; Efcavitch, J. W.; Grossman, P. D., US Patent 5,567,292, 1996

¹² Barbier, V.; Viovy, J.-L.; *Current Opinion in Biotechnology*. **2003**, *14*, 51-57

¹³ Song, L.; Liang, D.; Fang, D.; Chu, B.; *Electrophoresis*. **2001**, *22*, 1987

¹⁴ Salas-Solano, O.; Carrilho, E.; Kotler, L.; Miller, A. W.; Goetzinger, E.- Susic, Z.; Karger, B. L.; *Anal. Chem.* **1998**, *70*, 3996-4003

¹⁵ Buchholz, B. A.; Shi, W.; Barron, A. E.; *Electrophoresis*. **2002**, *23*, 1398-1409

As described in the previous chapter, one problem with linear water-soluble polymers is the inverse relationship between peak resolution and solution viscosity. Raising hydrophilicity, molecular weight, or concentration to produce a higher resolution increases solution viscosity. Some attempts to change the viscosity between injecting the solution and running the electrophoresis have been tried.

One approach utilizes polymers whose water solubility and solution viscosity depend on temperature. For example, Poly(N,N'-isopropylacrylamide) (PNIPAM) and Polypropyleneoxide (PPO) show such a temperature dependency. PNIPAM has a transition temperature solubility (lower critical solution temperature or LCST) around 30°C. PNIPAM dissolves in water, and, thus, its solution viscosity increases below LCST, while the solution becomes less viscous above LCST. Therefore, the problem was expected to be solved by applying different temperatures at polymer injection and during electrophoresis. Several studies based on this idea were conducted. However, since hydrophilicity of PNIPAM is lower than DMA, even when below LCST, electrophoretic performance was much worse than for polymers like LPA and DMA.

The other approach uses highly structured copolymers, such as a block copolymer and a graft copolymer. For example, LPA-g-PNIPAM copolymer, which consists of an LPA main chain and a short PNIPAM graft chain, has interesting features, such as a solution-hydrogel transition driven by temperature.^{16,17} (See Figure 11B.) Utilizing this feature, it would be possible to inject a less viscous LPA solution in low temperature and then raise the temperature to form a hydrogel for electrophoresis. This polymer would also adsorb onto the capillary wall anchored by the PNIPAM graft polymer. This approach would be useful if the polymer structure could be freely controlled. However, good results have not been observed with this class of polymers, mainly because polymerization techniques have not yet been developed. For example, if high molecular weight water-soluble block and graft copolymers, such as LPA-DMA, PEO-PPO, and LPA-PPO, could be easily synthesized, they should be quite useful for this sort of use.

¹⁶ Barbier, V.; Buchholz, B. A.; Barron, A. E.; Viovy, J.-L. *Electrophoresis*. 2002, 23, 1441-1449

¹⁷ Below LCST, PNIPAM hydrophilicity is high enough to allow the polymer to dissolve, whereas above LCST, PNIPAM becomes hydrophobic enough to aggregate and cause gelation.

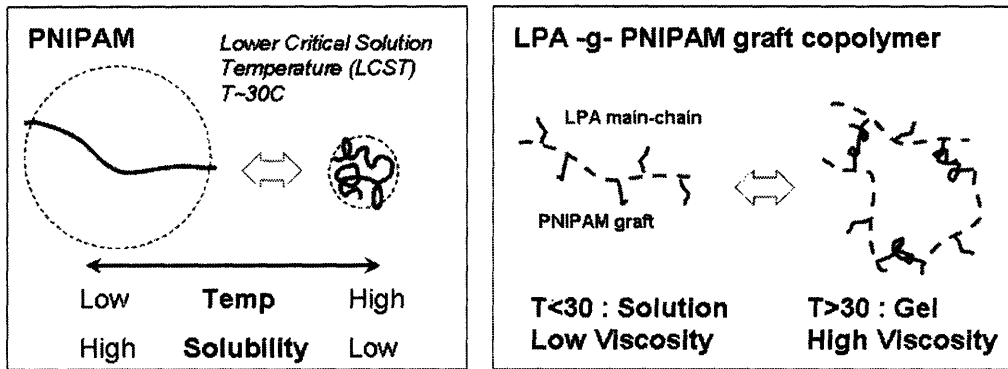


Figure 11 Application of thermosensitive polymers. A, Thermosensitive polymer with LCST; B, LPA-g-PNIPAM graft copolymer

3.2.3 Summary of the Past Polymer Research

LPA has achieved the highest separation capability, even though the cost consuming pre-coating process is required. DMA [Poly(N,N'-dimethyl acrylamide)] has good separation capability compared to LPA, and it does not require pre-coating. According to past journal articles and reviews, these two polymers are currently in commercial use. LongRead™ Matrix for Amersham Biosciences MegaBACE™ is mostly to be LPA, while POP™ matrix for Applied Biosystem ABI3700™ series is believed to be DMA.

Thermo-sensitive polymers and polymers with higher-order structure have the possibility to reconcile deficiencies between high separation capability, short migration time, solution viscosity and self-coating property, which LPA lacks. However any polymer exceeding LPA has yet been developed. New polymerization techniques that can synthesize block and graft copolymers using various water-soluble polymers and thermo-sensitive polymers need to be developed for more advanced research.

3.3 Experimental Section

3.3.1 Materials

Due to time and facility constraints, polymers were not synthesized in this project but purchased from other laboratories. Table 2 shows the list of polymers used for this study.

Table 2 Polymer samples used for this experiment

Name	Mw ¹⁾	Appearance	Vendor
POP7	n/a	Solution in buffer	ABI (Default polymer for ABI3730xl)
LPA 800k	600-1000k*	10% aq. solution	Polysciences, Inc ²⁾
LPA 5000k	~5000k	1% aq. solution	Polysciences, Inc
DMA 100k	125k	Powder	Polymer Sources, Inc ³⁾
DMA 1000k	1010k	Powder	Polymer Sources, Inc
LPA-DMA random copolymer	350k	Powder	Polymer Sources, Inc

1) Weight averaged molecular weight, 2) <http://www.polysciences.com/shop/>, 3) <http://www.polymersource.com/>

3.3.2 Capillary Electrophoresis

Separations were studied using ABI3730xl instruments and performed with a standard DNA sample, called BigDye™ ver.3.1 (Applied Biosystems, Foster City, CA, USA). Before injection, DNA was suspended in formamide, heated at 60°C for 2 min for denaturation, and then cooled on ice. Standard, thirty-six centimeters, uncoated fused-silica capillaries were used. Polymer solutions were prepared in 1X running buffer solution for ABI3730xl (Applied Biosystems). Each solution was mixed overnight in order to be dissolved. Electrophoresis experiments were performed at 50, 60, and 70°C of the separation temperature. The running voltage was changed among 8.5, 10.0, and 13.2kV, which correspond to 236, 278, and 367 V/cm of the electric field, respectively. Since the length of the capillary is 36cm. In most cases, the separation temperature was 60°C and the electric field was 278V/cm, except as otherwise noted. The electropherograms were analyzed by fitting a Gaussian shape to the individual peak profiles using Fityk Software.¹⁸

¹⁸ <http://www.unipress.waw.pl/fityk/>

3.3.3 Chemical Analysis

3.3.3.1 Liquid Chromatography – Mass Spectrum (LC/MS) and Gas Chromatography – Mass Spectrum (GC/MS) Measurement

Sample Preparation POP7 (0.4g) was precipitated in 15mL of methanol. Low molecular weight additives were extracted from the precipitate by supersonication for 30 min. Supernatant of 1 and 2 were merged, filtrated and dried. The residual was dissolved in water, filtrate with a 0.45 μ m filter before being used for measurement. Measurement conditions are summarized in Table 3 and 4.

Table 3 LC/MS Measurement Conditions

Parameters	Descriptions
Instrument (LC)	Waters2690
Column	Waters X-Terra MS C18 2.0 mm ϕ ×15 mm×2.5 μ m
Mobile Phase	Water/Methanol = 1/1
Flow Rate	0.3 mL/min
Injection Amount	10 μ L
Instrument (MS)	Micromass Q-ToF
Ionization method	ESI Positive and Negative Mode
Measured Mass	40~2000 amu
Capillary Voltage	3,000 V(Positive), 2,500 V(Negative)
Cone Voltage	50 V
Source Block Temp	80°C
Desolvation Temp	350°C
Scanning Time	2 sec
UV Wavelength	280 nm

Table 4 GC/MS Measurement Conditions

Parameters	Descriptions
GC/MS Instrument	HP5973MSD
Ionization	EI
Column	DB-1 0.25mm ϕ ×30m
Membrane thickness	0.25 μ m
Oven Temp	50°C (1min)→10°C /min→300°C (15min)
Injection Temp	300°C
Column Flow Rate	1mL/min
Injection Amount	1 μ L

3.3.3.2 Nuclear Magnetic Resonance (NMR) Measurement

^1H -NMR spectra and ^{13}C -NMR spectra were measured with the Unity400 spectrometer (Varian, Inc).

Qualitative Analysis Solvent was evaporated from the sample solution. Then D_2O was added to the residual to prepare the NMR sample.

Quantitative Analysis ^1H -NMR was measured with a double-tube (5mm in diameter). The outside and inside tubes were filled with the sample solution as provided and deuterated chloroform (NMR lock solvent), respectively. Due to the high proton concentration, a single pulse was used to excite. Two runs were performed with 60 sec and 120 sec of waiting time, in consideration of the polymer sample's long relaxation time.

Confirmation of the SEC sample's chemical composition A part of the precipitate in methanol was dissolved in water and used for the NMR sample. ^1H -NMR was measured with a double tube as the quantitative analysis.

3.3.3.3. Size Exclusion Chromatography (SEC) Measurement

A part of the precipitate in methanol was used for the SEC measurement. A sample solution of 0.1wt% was prepared with 0.1M KH_2PO_4 pH=7, and then filtered with a 0.45 μm PTFE filter. The measurement conditions are summarized in Table 5.

Table 5 SEC Measurement Conditions

Parameters	Descriptions
Instrument	Tosoh HLC-8220 GPC(L)
Detector	RI (Incorporated)
Mobile Phase	0.1M KH_2PO_4 pH=7
Flow Rate	1.0mL/min
Injection	0.1wt% x 100uL
Column	TSKgel GMPWXL (30cmx2)
Column Temp	30°C
Calibration Sample	Single Dispersion PEG/PEO
Calibration Method	Standard PEG/PEO
Calibration Curve	Cubic function

3.4 Results and Discussions

3.4.1 Establishment of Experimental Protocol

In the beginning of the project, there was very little information about the polymer media used with ABI3730xl, since no projects at the Broad Institute have addressed this issue in the past. Moreover, ABI makes money not only from instruments but also from consumables including polymer media and buffer solutions. Information on the proprietary consumables has been kept secret from users like Broad.

Another problem was the highly-automated sequencing process at the Broad Institute. As shown in Figure 12, output from this process emerges as a computer data file in ABI's proprietary format, which can be handled only by the special software provided by the company. The ABI format data includes the following information:

- Experimental conditions
- Raw data on the luminescent intensity of the four bases as a function of time
- The base code sequence as determined by the software named "Base Caller"
- Quality of base code sequencing accuracy (Q-value)

There are two problems with the ABI format data.

- Base Caller software is designed to label base code sequence and calculate Q-values assuming standard running conditions (temperature, electric field, polymer, etc) of ABI3730xl. When different polymer media are used, the base calling and Q-value calculations do not make any sense.
- Raw data cannot be extracted and handled with common peak-fitting programs, and the data are difficult to analyze manually.

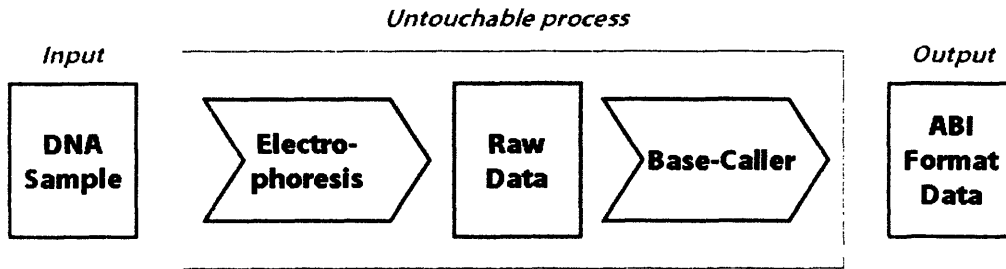


Figure 12 Schematic diagram of the ABI3730xl operational procedure

A: Base-calling Data

bp	Nu	Q value	bp	Nu	Q value	bp	Nu	Q value
15	T	5	41	G	51	71	G	50
16	C	3	42	T	34	72	C	54
17	G	5	43	T	28	73	T	54
18	T	3	44	A	21	74	G	54
19	C	3	45	A	19	75	C	50
20	C	6	46	T	26	76	T	49
21	T	5	47	G	35	77	G	46
22	G	10	48	T	25	78	C	41
23	C	20	49	T	46	79	T	60
24	A	8	50	G	42	80	T	49
25	G	5	51	C	43	81	C	54
26	C	9	52	T	42	82	T	54
27	T	6	53	A	30	83	C	60
28	G	23	54	C	52	84	C	57
29	G	32	55	T	38	85	T	57
30	T	27	56	A	42	86	C	52
31	T	17	57	C	49	87	A	38
32	A	9	58	T	52	88	C	47
33	T	13	59	G	52	89	T	52
34	G	24	60	C	40	90	G	52
35	A	17	61	T	49	91	T	42
36	T	10	62	G	46	92	C	60
37	T	27	63	A	42	93	T	57
38	A	21	64	C	30	94	C	60
39	C	28	65	A	27	95	C	48
40	T	37	66	A	42	96	A	28
			67	T	50	97	C	52
			68	G	54	98	T	57
			69	C	39	99	T	57
			70	T	50	100	C	57

B: Visualized Data

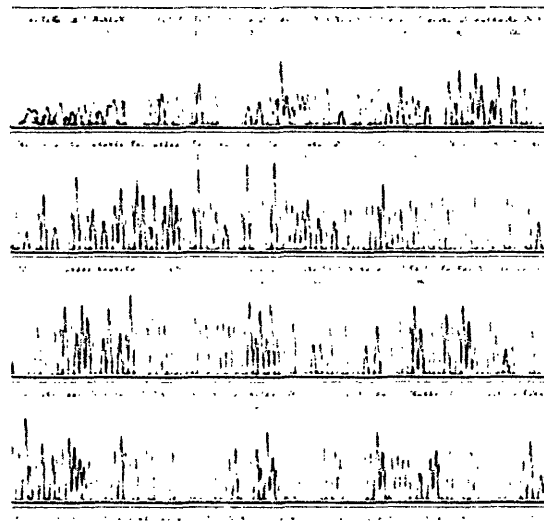


Figure 13 Sample data after base calling. A: Base-calling results and Q-value; B: Visualization. Base calling data correspond to the data shown in Figure 15, 16, and 17, while the visualized data do not.

The experimental project began by examining how to extract raw data from the ABI format data and analyze them manually. With raw data, the resolution for certain peaks can be calculated and the data quality among different experimental conditions can be compared.¹⁹ Research discovered the “org.biojava.bio.chromatogram” free computer software program designed to handle raw data in the ABI format data file offered by BioJava, an open-source project dedicated to providing a Java framework for processing biological data.²⁰ Sample raw data extracted using this software package is shown in Figure 15.

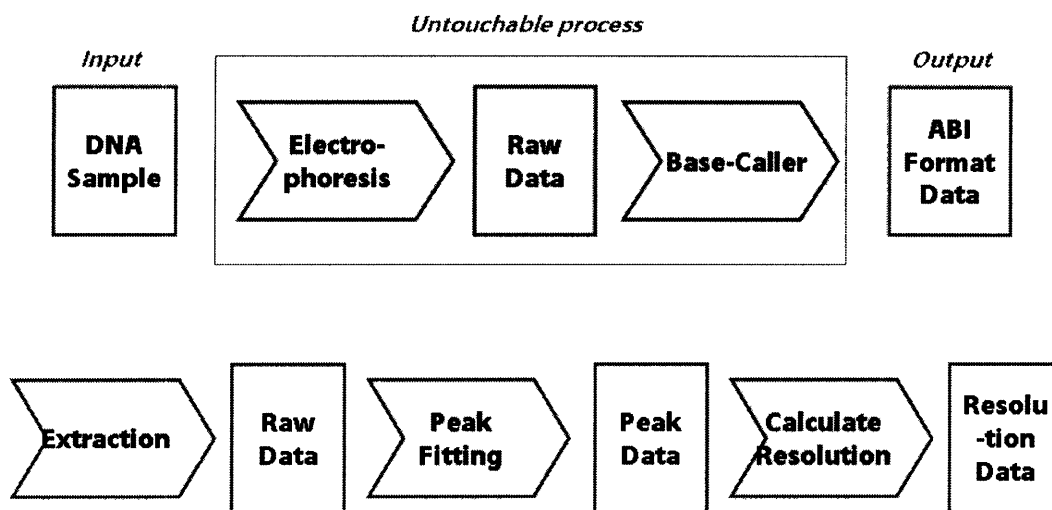


Figure 14 Schematic diagram of the operational procedure used for this research

¹⁹ In most studies introduced in the previous chapters, researchers used old models of DNA sequencers to acquire raw data. It is much easier to retrieve raw data from such old models.

²⁰ <http://www.biojava.org>

<http://www.biojava.org/docs/api/org/biojava/bio/chromatogram/package-summary.html>

ABI's standard DNA sample, "BigDye ver 3.1," was used throughout this research. Data of Adenine fluorescent intensity were used to calculate resolution through these properties because of strong fluorescent intensity, evenly distributed peak position, and constant data quality resulting from the labeled dye's high durability. In order to elucidate the effects of polymer media on the separation, selected peaks in the electropherogram raw data were fitted using the nonlinear curve-fitting software, Fityk.²¹ Figure 16 shows BigDye Adenine fluorescent data under standard conditions. The fluorescent peaks that correspond to the following bases are used for fitting and calculating resolution: 53, 56, 105, 106, 108, 109, 160, 162, 263, 268, 318, 320, 423, 426, 567, 570, 640, 649, 724, and 730. These peaks are marked by yellow circles in Figure 16.

Resolution was calculated using the following equation:¹⁶

$$R = \frac{\sqrt{5.5} (tm_2 - tm_1) tm_2}{2 (tm_2 + tm_1) \sigma \Delta N} \cdot 1$$

Where R, tm_1 , tm_2 , σ , and ΔN are resolution, migration time of peak 1, migration time of peak 2, average peak width for peaks 1 and 2, and number of bases separating peaks 1 and 2, respectively. Using the best base-callers available, in optimal conditions, sequences can typically be read at an accuracy of 98.5 to 99%, down to a resolution of 0.3.⁸

²¹ <http://www.unipress.waw.pl/~wojdyr/fityk/>

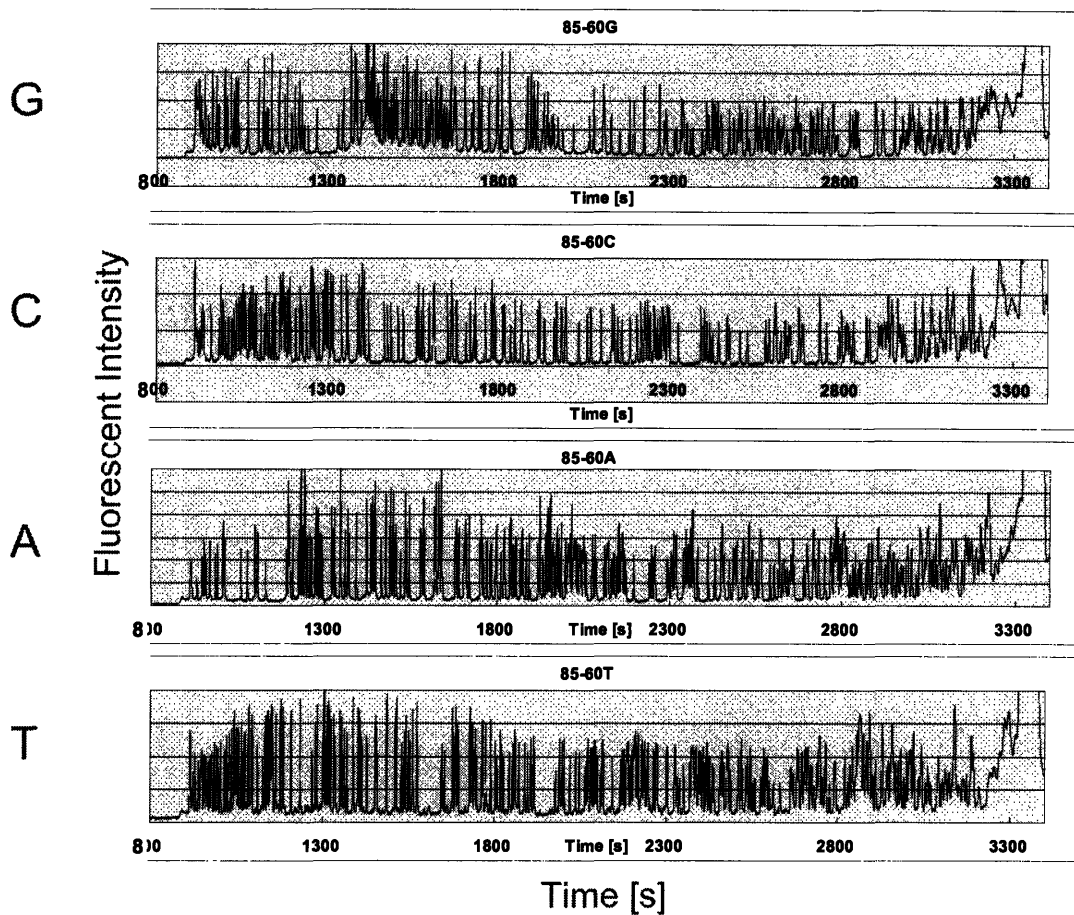


Figure 15 Extracted raw electrophoresis data using ABI3730xl. Emission strength (arbitrary unit) versus time for Guanine, Cytosine, Adenine, and Thymine from top to bottom. The data were obtained using ABI's standard DNA sample, "BigDye ver 3.1," under standard running conditions (polymer media: POP7, electric field: 236V/cm, running temperature: 60°C)

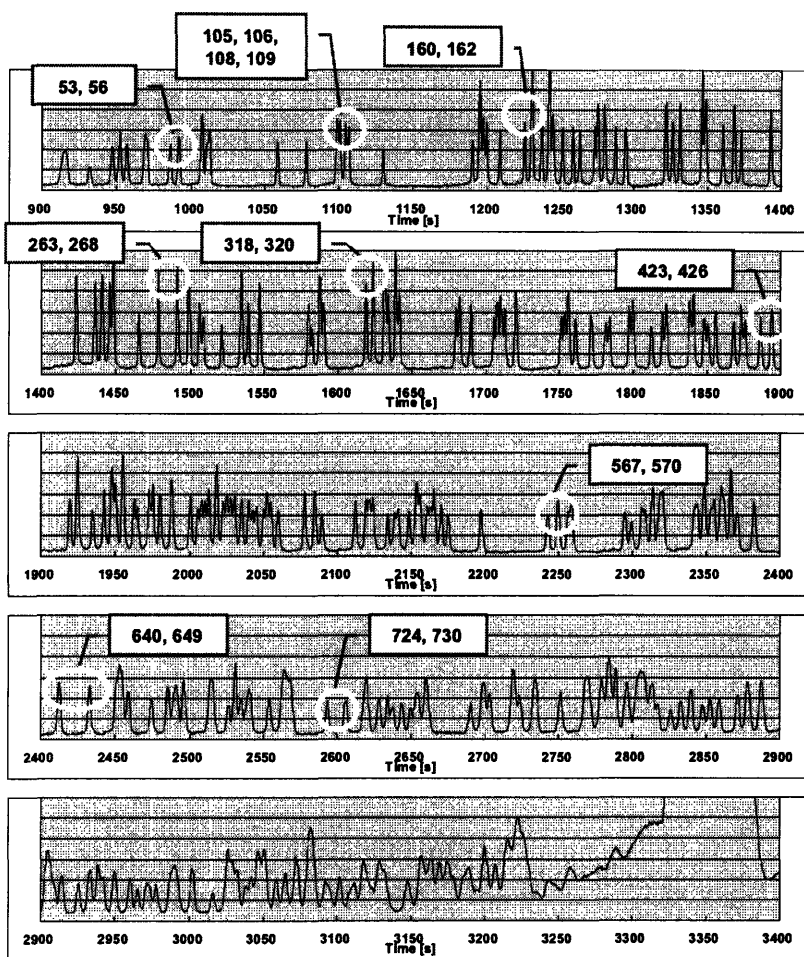


Figure 16 Fluorescent data of “BigDye ver.3.1” for Adenine. Peaks used for fitting and calculating resolution are shown in the yellow circles. (polymer media, POP7; electric field, 236V/cm; running temperature, 60°C).

For typical DNA-sequencing operations, the quality of base-calling accuracy is expressed by a Q-value.²² Defined by the following equation, Q-values are calculated by the base-caller.

$$Q = -10 \log_{10} (p),$$

where p is the estimated error probability for the base-call. When a base-call has a probability of 1/1000 of being incorrect, the Q-value is assigned as 30. The Broad Institute secures 99% of base-calling accuracy or Q20 level of base calling quality.

Figure 17-A is the Q-value plot for the BigDye at Broad's standard running condition (POP7, 236V/cm, 60°C) Under this condition, a base call below 700bp can clear a Q20 quality level. Figure 17-B shows the plot of resolution data manually calculated from the same raw data. Comparing both plots, the author decided to consider $R=0.4$ the same level of quality as Q20. It should be noted that although a rough relationship exists between Q20 and the resolution used in this project, the relationship is not linear and the degree of correlation is fairly low. (See Figure 18.) The resolution defined in this research is useful in examining data obtained under sub-standard conditions; however, to be used in in-depth comparisons with Q-value, a more sophisticated way to calculate the resolution would be required.

²² Ewing, B.; Green, P. *Genome Research*. 1998, 8, 186-194

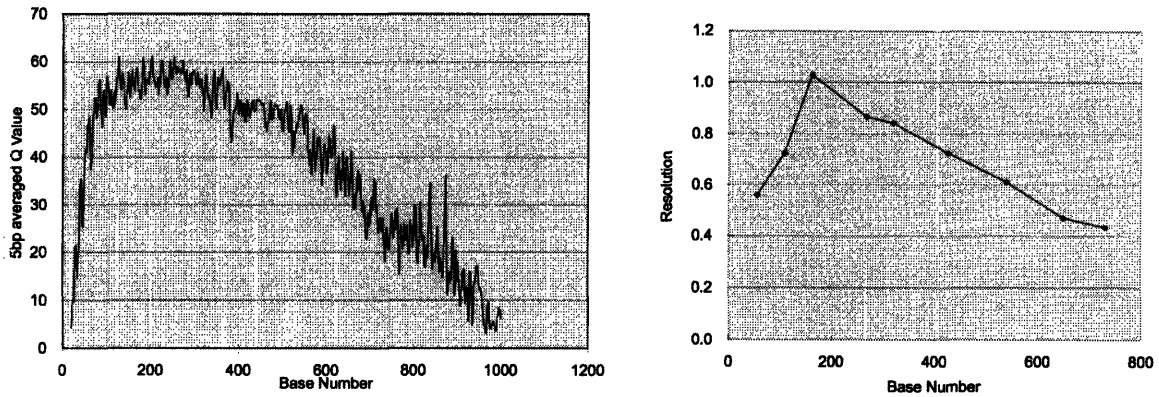


Figure 17 Comparison between Q-value (left) and Resolution (right). Both are calculated from the raw data shown in Figure 15.

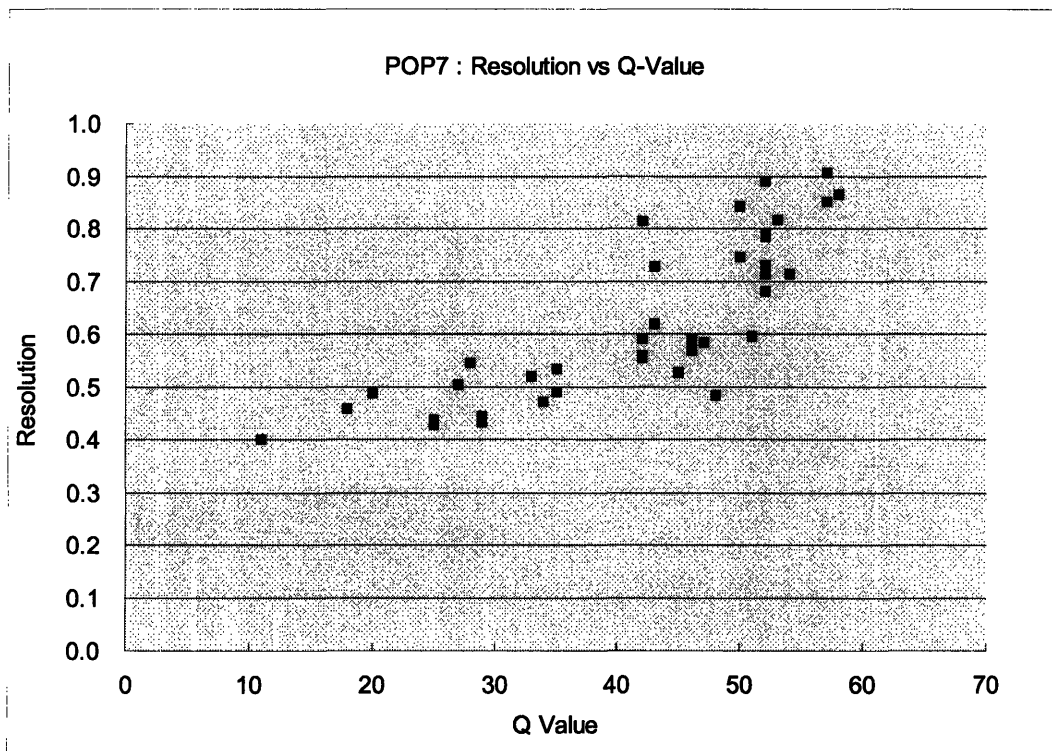


Figure 18 Correlation between Resolution and Q-value. (DNA sample: BigDye ver.3.1; polymer media: POP7; running temperature: 50°C, 60°C and 70°C; electric field: 236V/cm and 278V/cm)

3.4.2 POP7

ABI's default polymer, Performance Optimized Polymer ver.7, POP7, was chosen for the first experiment of the project in order to confirm if the new data analysis method would work with ABI3730xl. This experiment would also supply standard data, which could be compared with data for other polymers. Data quality's dependence on temperature and electric field was also investigated.

Figure 19 shows the sequencing data quality's dependence on temperature for POP7. Resolution and Q-value were plotted versus base pair number (bp). Temperature ranged from 50°C to 70°C, ABI's recommendation for sequencing operation; electric field was fixed at 236V/cm. Within this temperature range, POP7 worked well and exhibited only slight temperature dependence. Resolution begins to increase at around 0.6 at 50bp and reaches its maximum of $R \sim 1.0$ at 200bp. It then decreases to its minimum of $R = 0.4$ at the base pair that is slightly longer than 700. The shape of the Q-value curve resembles the shape of the resolution curve. Thus, the resolution analysis could be used instead of the Q-value analysis. Figure 20 presents the resolution data plot as a function of migration time. Migration speed also exhibits only a slight dependence on running temperature.

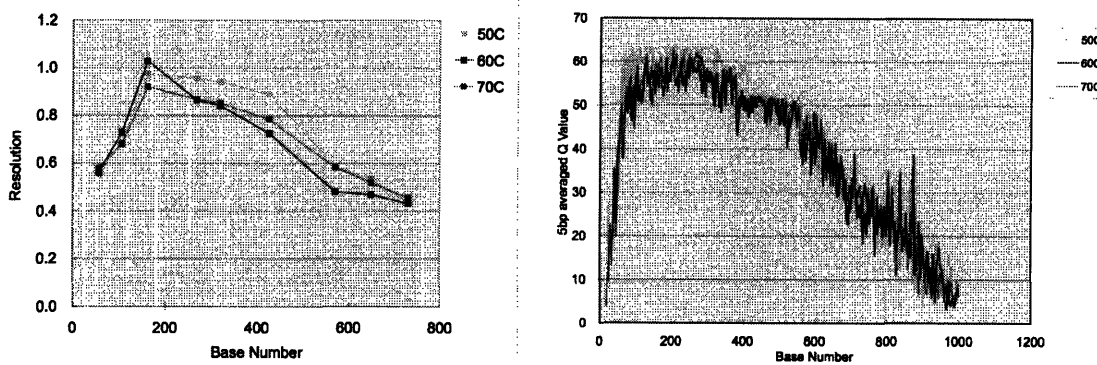


Figure 19 Temperature dependence of Resolution (left) and Q-value (right) versus base number. (Polymer: POP7, electric field: 236V/cm)

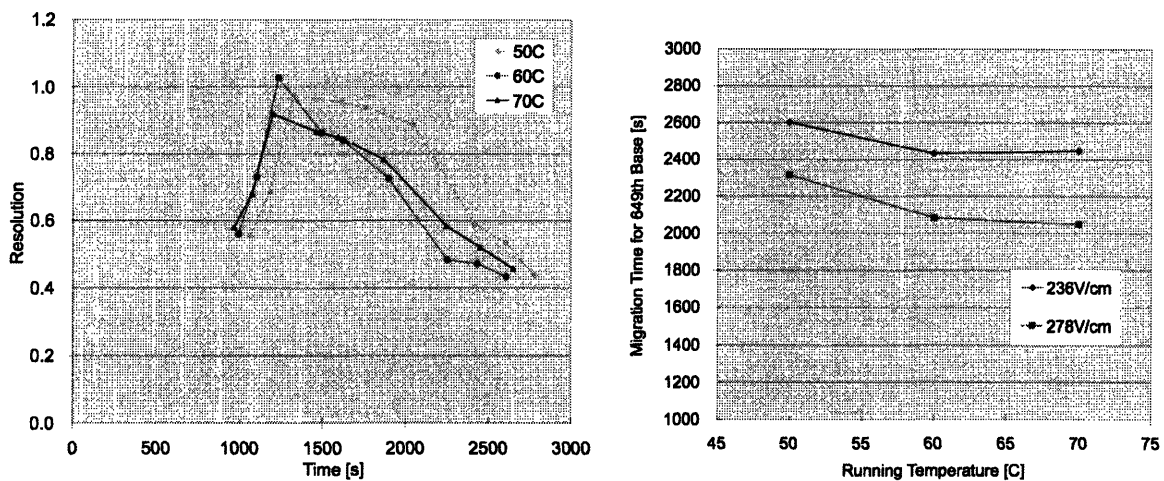


Figure 20 Temperature dependence of resolution for each base. Resolution for each base number is plotted versus migration time (left) and temperature dependence of migration time for the 649th base (right). (Polymer: POP7, electric field: 236V/cm)

Figure 21 shows the relationship between electric field and sequencing data quality for POP7. Resolution was calculated for electric field of 236V/cm, 278V/cm, and 367V/cm, while temperature was fixed at 60°C. When comparing the resolution data at 236V/cm and 367V/cm, the data quality at the higher electric field seems inferior to the data quality at the lower electric field. For example, the base number at R=0.4 is 730bp at 236V/cm, while only 650bp at 367V/cm. On the Q-value plot, the read length at 367V/cm is obviously shorter than those at 236V/cm and 278V/cm. (See Figure 21B.) Results from both analyses are fairly consistent.

Figure 22 shows the effect of electric field on the migration time of electrophoresis. As electric field increases, migration time quickly decreases. Since a very high electric field deteriorates data quality, as shown in Figure 21, the electric field obviously affects the relationship between data quality and migration time. Figure 23 demonstrates relationship between the highest base number at a Q20 level of data quality and the maximum number of runs per day.²³ Assuming the number of bases at a Q20 or higher quality per day is expressed by the product of the Q20 base number and the maximum number of runs per day, it would be calculated as 20,000bp for 236V/cm, 23,000bp for 278V/cm, and 30,000bp for 367V/cm. In other words, even when quality deterioration is taken into account, a higher electric field like 367V/cm would still improve the operation efficiency, compared to the current setting of 278V/cm. The higher electric field could cause potential disadvantages, such as heavier damage on the instruments, so more detailed studies are required before this can be applied to practical operation.

No statistical treatments were conducted on the resolution analysis in this report. Since the developed resolution analysis was carried out manually, a statistical approach would unrealistically increase the workload. It should be noted that while this manual method is useful for analysis of experimental data under extreme running conditions, it is not suitable for analysis requiring high accuracy, such as running condition optimization.

²³ This result is taken from the unpublished study conducted by James Meldrim in the Technology Development group, Sequencing Operation Platform of the Broad Institute.

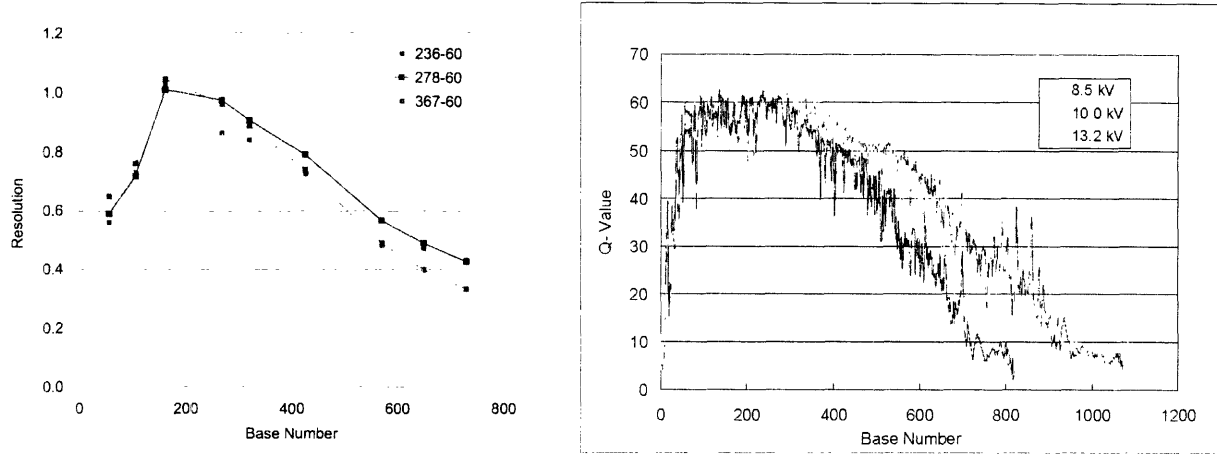


Figure 21 Dependence of Resolution (left) and Q-value (right) on electric field versus base number. (Polymer: POP7, running temperature: 60°C.)

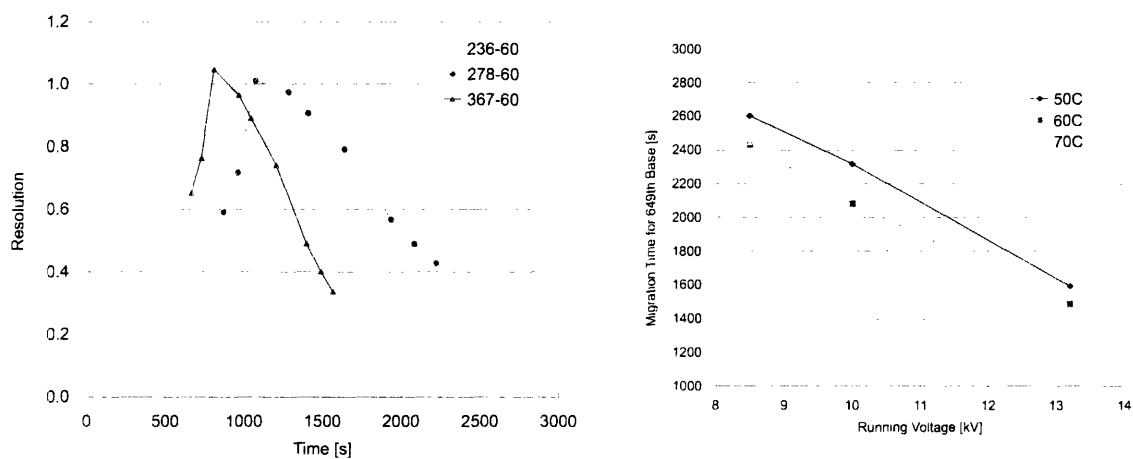


Figure 22 Dependence of resolution for on electric field for each base. Resolution for each base number is plotted versus migration time (left) and temperature dependence of migration time for 649th base (right). (Polymer: POP7, Running temperature: 60°C.)

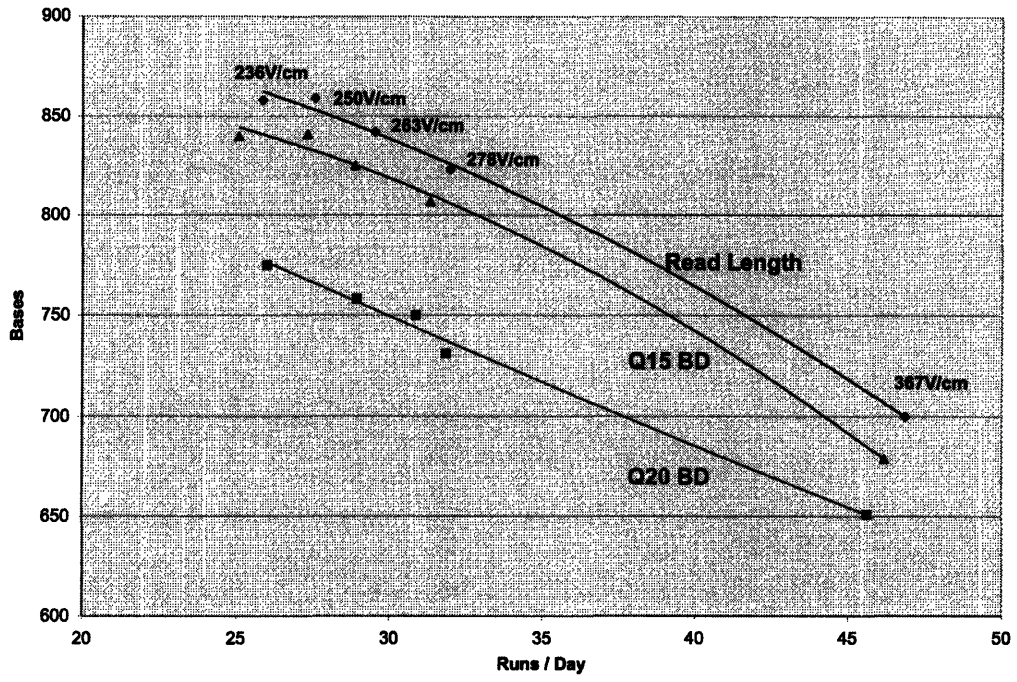


Figure 23 Effect of electric field on Q20 bases and available runs per day.

3.4.3 DMA

DMA [poly(N,N'-dimethylacrylamide)] was chosen as the second polymer after POP7 because it is widely believed to be the main ingredient of POP7, as mentioned in section 3.2.1.

Figure 24 shows the raw electropherogram data for BigDye measured with DMA100k.²⁴ Peak separation was much poorer compared with POP7. Since most peaks could not be determined, resolution could not be calculated. The solution viscosity of DMA100k was also much lower than that of POP7. These results suggest that a molecular weight of 100k is too low for this polymer. Electropherogram and resolution data for DMA with a higher molecular weight (1000k) are shown in Figure 25 and 26, respectively. The peak resolution is better than DMA100k, yet still worse than POP7, both for 5.0% and 7.5% concentration.

These results do not support the general belief that the main ingredient in POP7 is DMA. Even for a 7.5% solution of DMA1000k, the peak resolution was still inferior to POP7. ABI's researcher Dr. Madabhushi found an experimental result that a 6.5% solution of DMA (molecular weight 98k) showed 600bp of read length in 125 minutes (resolution > 0.59).¹⁰ Since he noted that "the increase in DMA molecular mass beyond 98kDa slightly improved the resolution but adversely increased the viscosity of the solution," the results of this project would not be improved even if a higher molecular weight for DMA was employed. It can be concluded that DMA is not the best polymer for DNA electrophoresis.

²⁴ DMA100k is DMA polymer with an average molecular weight of 100kDa. See Table 2.

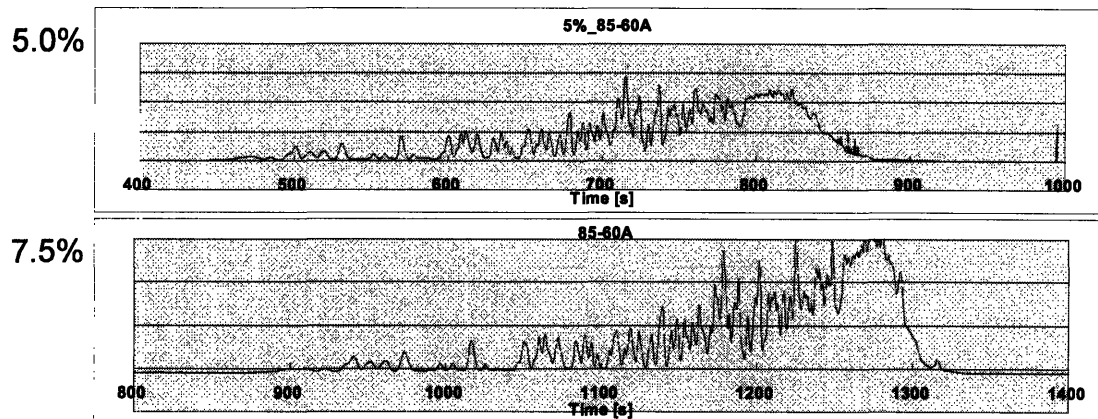


Figure 24 Raw electrophoresis data for DMA100k

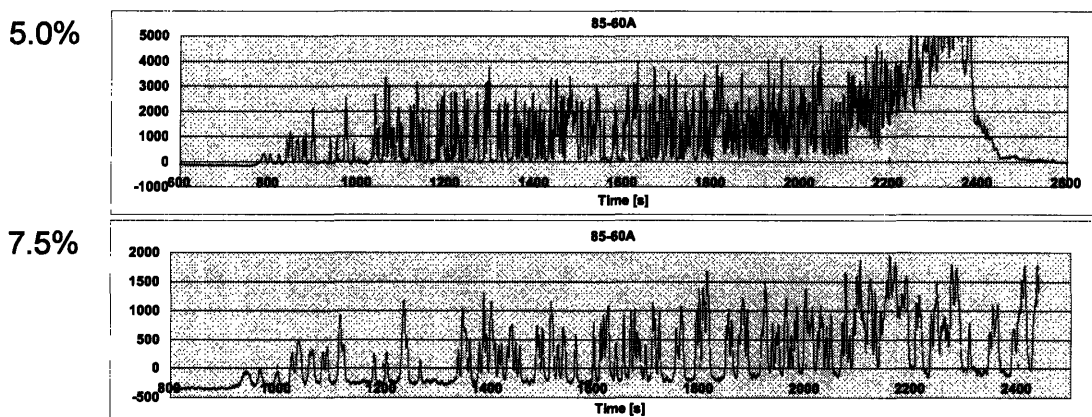


Figure 25 Raw electrophoresis data for DMA1000k

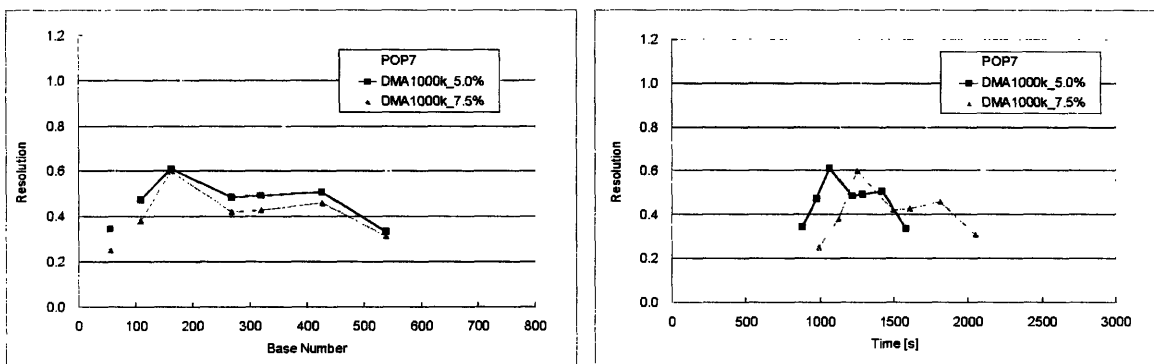


Figure 26 Resolution for 5.0% and 7.5% of DMA1000k as a function of base number (left) and migration time (right).

3.4.4 LPA

LPA (linear polyacrylamide) is one of the most widely used separation media in DNA-sequencing due to its excellent performance in terms of read length and separation time. Since LPA does not have self-coating capabilities, it is necessary to periodically pre-coat the capillaries. Amersham's MegaBACE uses the high molecular weight, LPA-based LongRead Matrix. In this section, the effects of polymer concentration, temperature and electric field on the peak resolution of electropherograms are investigated using a commercially available LPA with a molecular weight of 800k.

Figure 28 shows electropherograms measured with various concentrations of LPA800k solution. In general, the peak resolution was better than with DMA. The relationship between resolution and base number is shown in Figure 29. The best resolution was obtained with 3.0wt% solution; resolution decreases with both higher and lower concentrations. LPA800k demonstrates a fairly good peak separation capability, but it is still inferior to POP7, especially in longer bp regions. The resolution plot crosses the $R=0.4$ and $R=0.3$ lines at 500bp and 700bp, respectively. In the low bp region, meanwhile, the resolution was even better than with POP7. The resolution versus time plot is shown in Figure 29. The separation time for LPA800k is shorter than that of POP7, whereas the separation time increases for higher concentration solutions.

The best explanation for these observations is that POP7 is not DMA but rather LPA, perhaps with a slightly higher molecular weight than LPA800k. In that case, LPA could potentially be used as a polymer medium for ABI3730xl after adjusting molecular weight and other physical properties.

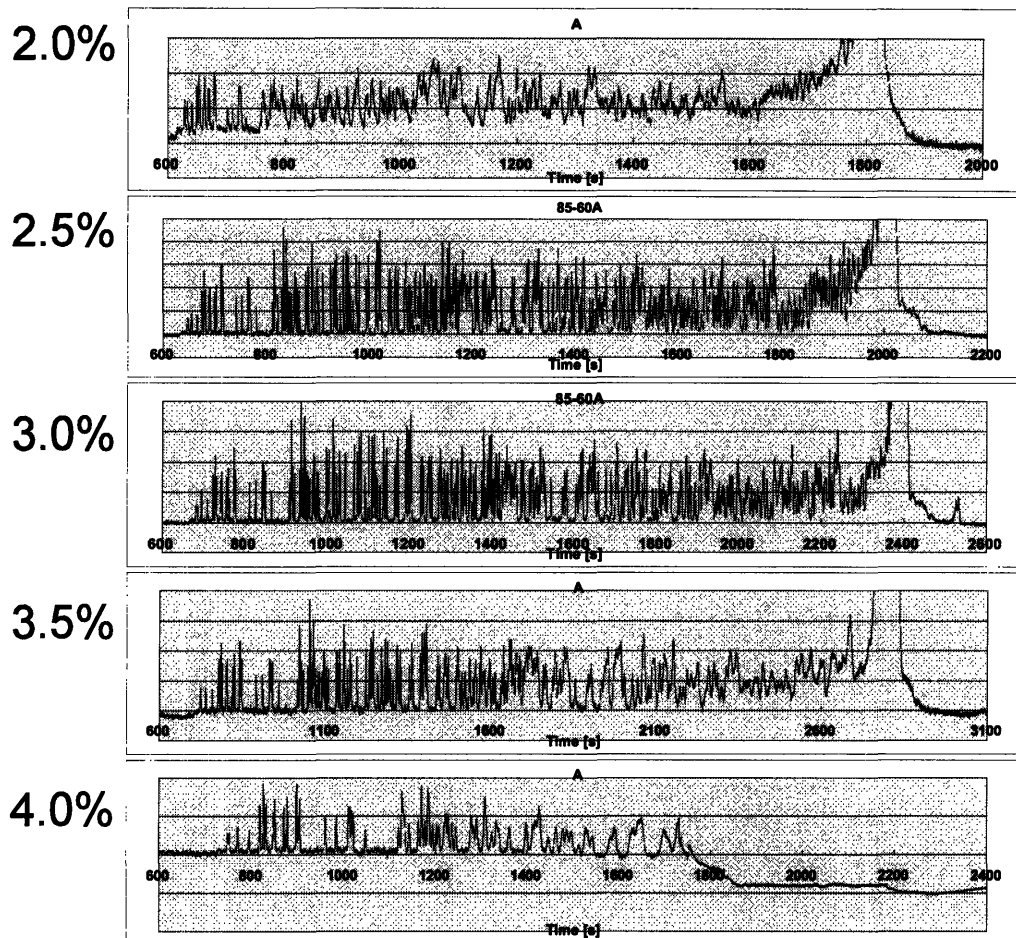


Figure 27 Raw electrophoresis data for LPA800k. From top to bottom, polymer concentration is 2.0, 2.5, 3.0, 3.5, 4.0 wt%. (Electric field 236V/cm, running Temperature 60°C.)

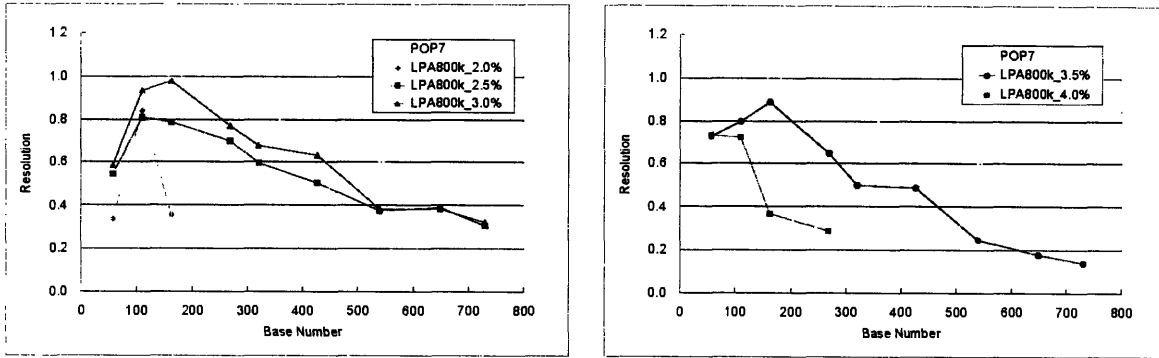


Figure 28 Resolution of electrophoretic peaks measured with various concentrations of LPA800k. (Electric field 236V/cm, running temperature 60°C.)

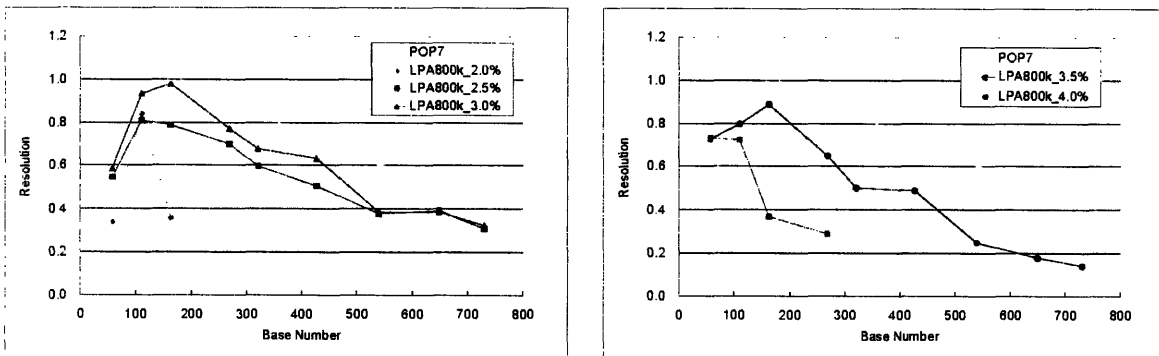


Figure 29 Relationship between electrophoretic peak resolution and migration time measured with various concentrations of LPA800k. (Electric field 236V/cm, running temperature 60°C.)

3.4.5 Chemical Analysis of POP7

Observations from previous sections suggest that a main ingredient in POP7 could be LPA. To clarify this, chemical analysis was performed on POP7. Low molecular weight additives were examined by Liquid Chromatography-Mass Spectroscopy (LC/MS) and Gas Chromatography-Mass Spectroscopy (GC/MS). Nucleic Magnetic Resonance Spectroscopy ($^1\text{H-NMR}$ and $^{13}\text{C-NMR}$) was used for qualitative analysis of polymeric materials and quantitative analysis for all chemical species. Finally, Size Exclusion Chromatography (SEC) was conducted to confirm the molecular weight distribution of polymeric materials.

LC/MS A broad peak was observed at 0.99 minutes on MS-TIC (Total Ion Chromatogram). Figure 30 shows the mass spectrum of this peak. The observed $[\text{M}+\text{Na}]^+$ ion proved the existence of a chemical species with $\text{M}=243$. This chemical species is supposed N-tris (hydroxymethyl)methyl-3-amino propanesulfonic acid (TAPS).

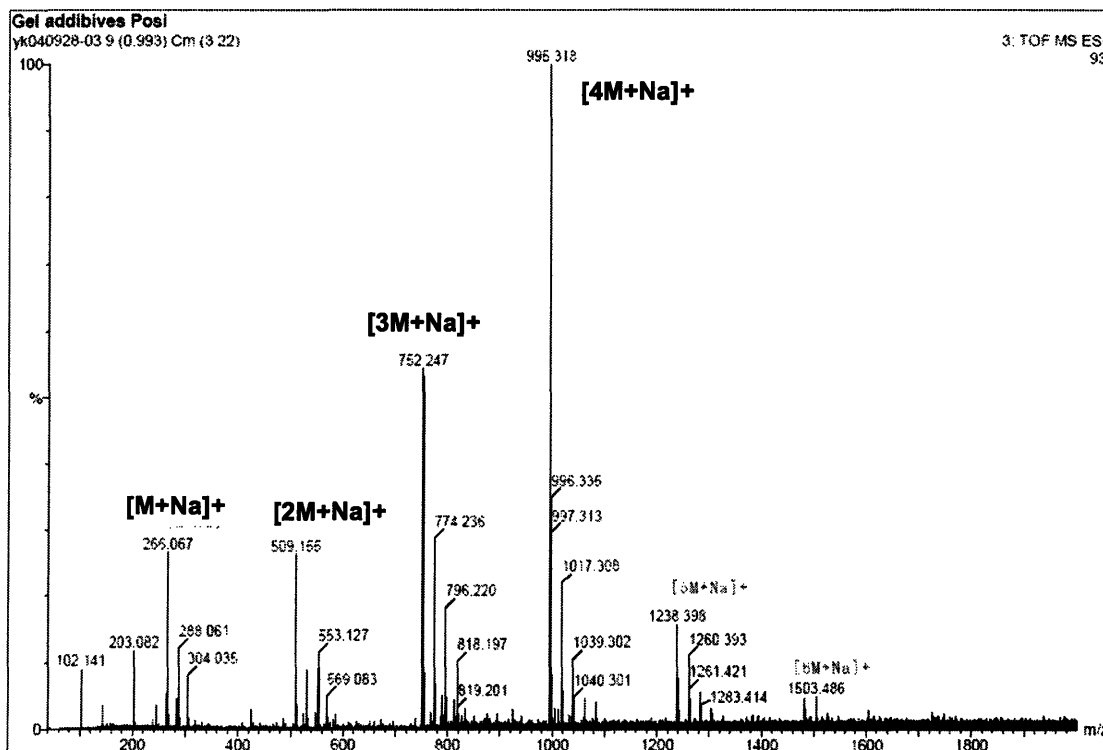


Figure 30 MS for the peak at 0.99 minutes of retention time on MS-TIC

GC/MS Figure 31 shows the GC/MS chromatogram of the pre-treated sample. The large peak A is observed at 6 to 10 minutes on the GC spectrum. The lower part of Figure 36 shows the mass spectrum of peak A. Using spectrum library analysis, MS peak with M=44 was concluded to be Urea.

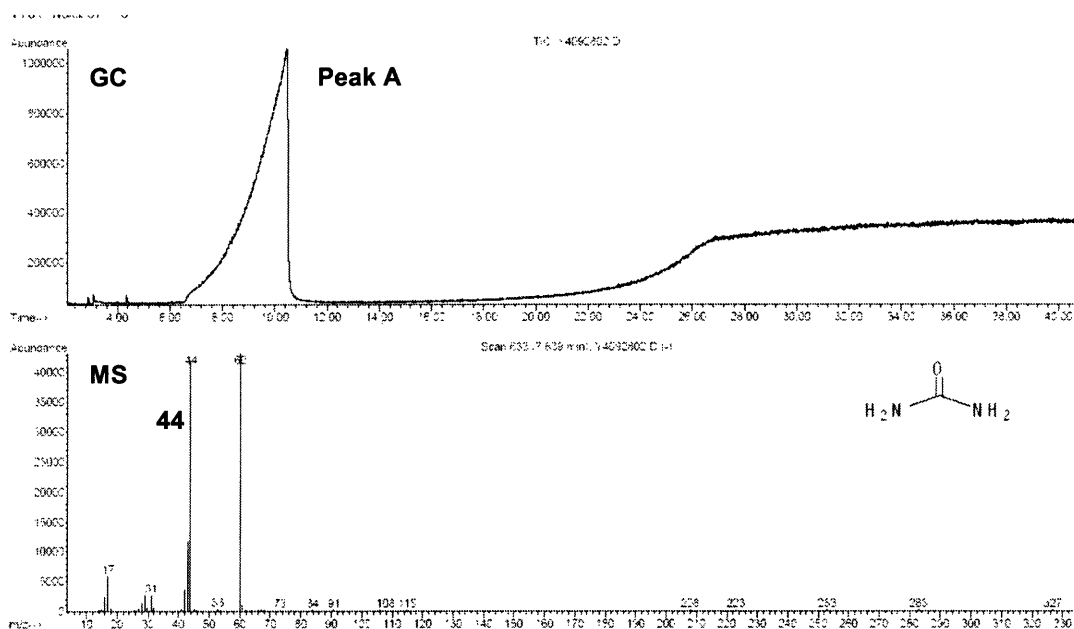


Figure 31 GC for pre-treated sample of POP7 and MS for the Peak A (GC/MS)

NMR Figures 32 and 33 show ^1H and ^{13}C -NMR spectra of POP7, respectively. Taking both MS and NMR analyses into consideration, it was determined that Urea and TAPS are involved in POP7 as low molecular weight ingredients. Peak assignments are drawn in each Figure. From the analyses of the chemical shift and peak area, the main polymeric ingredient of POP7 was determined to be LPA. Peaks at 2.95ppm in ^1H -NMR as well as 175ppm and 36.5ppm in ^{13}C -NMR seemed to be derived by the secondary polymer component, DMA. Table 6 shows the component ratio of POP7 estimated from qualitative ^1H -NMR analysis.

Table 6 Estimated ratio of each component for POP7 (wt%)

Ingredients	wt%
Urea	37.6
TAPS*	1.7
LPA	1.7
DMA	0.04
H ₂ O	58.9

* TAPS is assumed as NaOH salt for calculation.

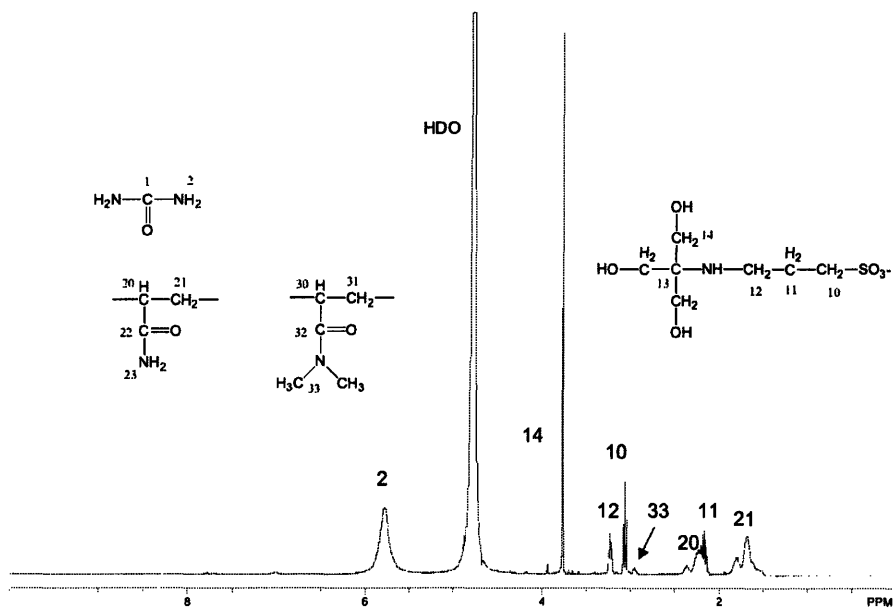


Figure 32 ^1H -NMR spectrum for POP7 (diluted by D_2O)

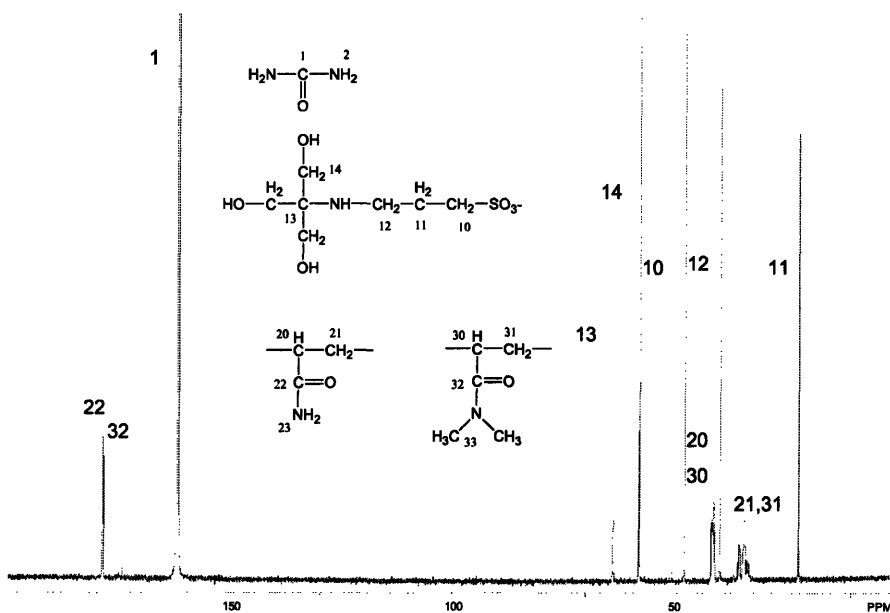


Figure 33 ^{13}C -NMR spectrum for POP7, magnified (diluted by D_2O)

SEC The molecular weight distribution of POP7 and LPA800k are measured as shown in Table 7. Although the average molecular weights of these two LPA polymers were almost the same, the molecular weight distributions were quite different. The molecular weight distribution of POP7 is very small with only 1.8 on the polydispersity index (PDI), while LPA800k has a rather broad distribution of 14 on PDI. LPA800k includes larger amounts of low-molecular-weight polymer chains, which contribute to the resolution inferior to that of POP7.

Table 7 Average molecular weight and polydispersity index for POP7 and LPA800k.

Polymer	Mw	Mn	PDI
POP7	8.29×10^5	4.4×10^5	1.8
LPA800k	2.86×10^6	2.1×10^5	14

Identity of POP7 The chemical analysis results support the hypothesis from the previous sections that POP7 consists mainly of LPA, although trace amounts of DMA are also present. The DMA portion could be mixed or co-polymerized with LPA to give POP7 self-coating capabilities. While the average molecular weight was similar between POP7 and LPA800k, POP7's molecular weight distribution was much narrower, explaining POP7's better resolution. POP7 include very common ingredients, like Urea and TAPS.

POP7 does not seem to be a complex polymer mixture, but rather a simple LPA polymer. Hence, POP7 could be replaced by a low-cost, general-purpose LPA. Annual savings at the Broad Institute were calculated \$383,000 per year, if POP7 could be completely replaced by a low-cost LPA800k. This corresponds to a 94.5% of reduction in polymer cost.²⁵

²⁵ Suppose that the polymer concentration of POP7 is 3%. Since the price of POP7 solution is \$1,690.50 for one case (= 6 boxes = 30 bottles = 840 mL), 1mL of POP7 solution would cost \$2.01. Hence, unit cost of POP7 (solid) would be \$67.08 per gram. Meanwhile, since LPA800k (10% aqueous solution, Polysciences, Inc.) costs \$92.75 for one bottle, or 250g, the unit cost of LPA800k (solid) is calculated to be \$3.71 per gram. As Broad institute consumes 240 cases (\$405,720) of POP7 every year, possible saving producing polymers in-house would be \$383,000 per year.

3.4.6 Polymer Mixtures

POP7 is a mixture, or copolymer, of LPA and a trace amount of DMA. This composition benefits from traits of both polymer, LPA and DMA, i.e. excellent peak separation and self-coating capability. Thus, polymer mixture is appealing because it could satisfy various requirements without the problems found with other polymers.

However, different polymers are usually immiscible. In fact, only one study has ever been reported for DNA electrophoresis with a polymer mixture.²⁶ Examining different combinations of LPA and DMA only a mixture of LPA with a trace amount of low molecular weight DMA (0.2%, 8k Da) overcome the incompatibility of these two polymers. In this section, DNA-sequencing with the LPA and DMA mixture is studied. It would be expected that LPA/DMA block copolymer would work as a compatibilizer; however, it has never been synthesized due to a lack of appropriate synthesis methods. In this experiment, LPA/DMA random copolymer with 350k of molecular weight (COP350k) was tested as a compatibilizer instead.

Figure 34 shows phase diagrams for LPA800k–DMA100k (A) and LPA800k–DMA1000k (B). Only a low concentration of DMA100k alleviated incompatibility. The mixtures of LPA800k (2.5%) and DMA100k (0.5%) does not show a significant difference from sole LPA800k (2.5%) alone on the peak resolution of the electropherogram (Figure 35) . Since the frequency of “unstable current errors,” which are often observed when LPA is used alone, was reduced to some degree, DMA may work as self-coating reagent.

Figure 36 is the phase diagram of LPA800k and COP350k. These two polymers phase separate very easily. For example, 2.5% of LPA800k and 0.5% of COP350k temporarily formed a pseudo-miscible solution, but then eventually gelled. Thus, COP350k did not work as a compatibilizer for LPA and DMA. Figure 37 shows the electrophoretic resolution for a mixture of 2.5% of LPA800k and 0.5% of COP350k; however, the resolution is obviously worse than LPA800k (2.5%) alone. Poor compatibility of two polymers further decreases resolution, as observed in previous research.

²⁶ Song, L.; Liang, D.; Chen, Z.; Fang, D.; Chu, B. *J. Chromatogr. A*. **2001**, *915*, 231-239

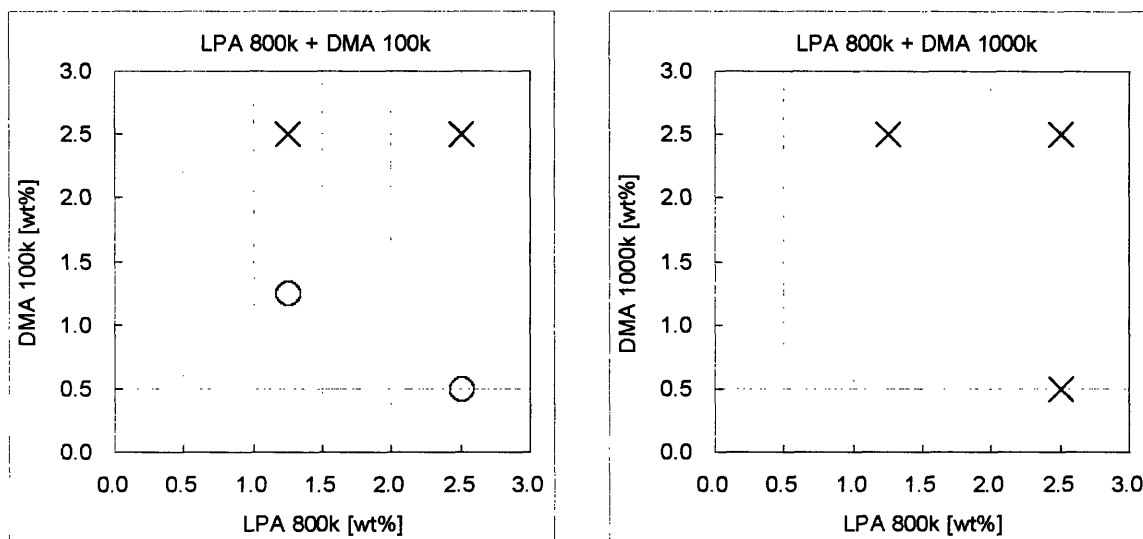


Figure 34 Phase diagrams for mixtures of LPA800k and DMA100k (A), LPA800k and DMA 1000k (B). O, miscible; Δ , pseudo-miscible; \times , immiscible.

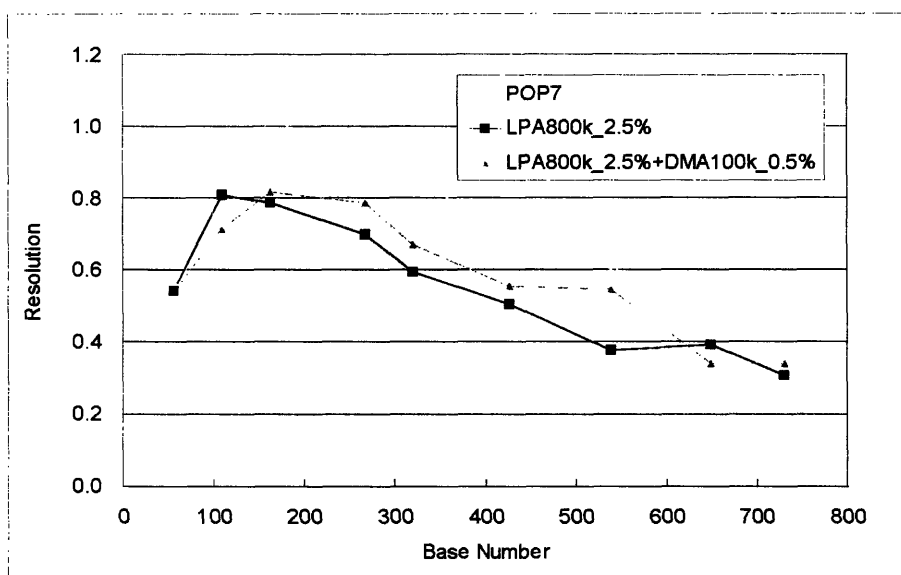


Figure 35 Resolution of electrophoretic peaks measured with a mixture of LPA800k and DMA100k. (Electric field 236V/cm, running temperature 60°C.)

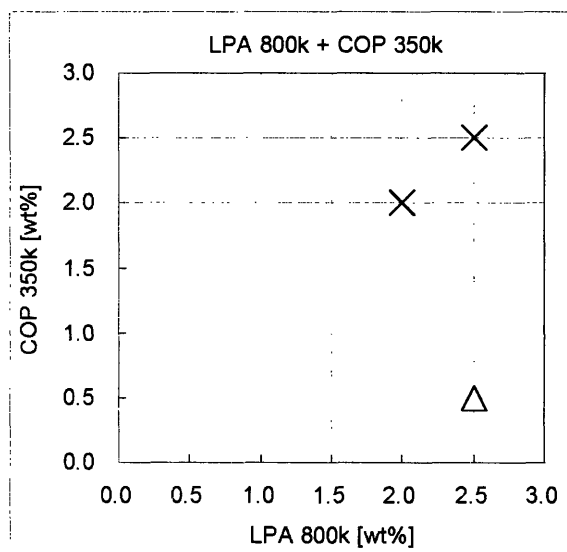


Figure 36 Phase diagrams for LPA 800k and COP350k. O, miscible; Δ, pseudo-miscible; ×, immiscible.

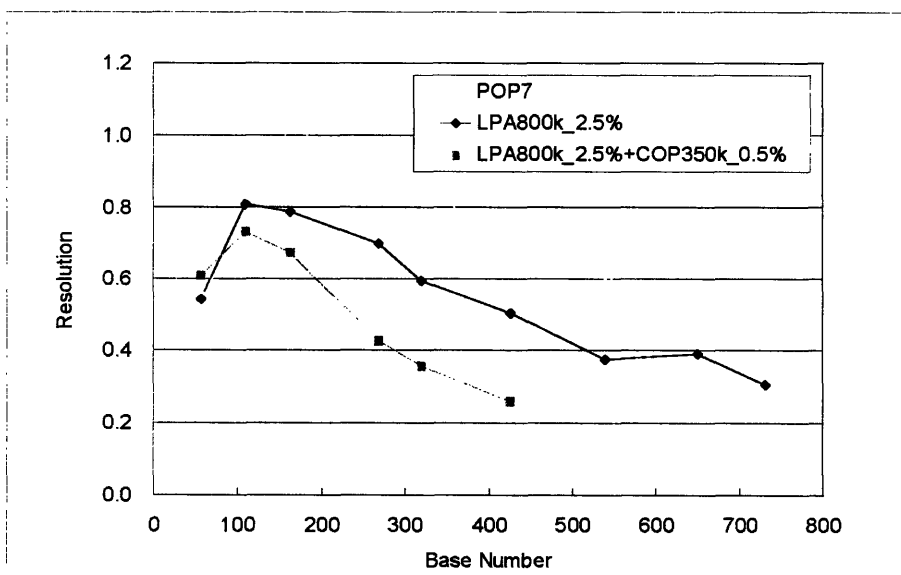


Figure 37 Resolution of electrophoretic peaks measured with mixture of LPA800k and COP350k. (Electric field 236V/cm, running temperature 60°C.)

3.5 Possible Approaches for the Future

In this project, focus was limited to the investigation of default polymers and several basic polymers. However, the primary objective of this research at Broad should be trying to test really cutting edge materials beyond the basic polymers. If Broad continues this study, what approach would be favorable for further investigation?

One possible interesting approach would be the use of a thermosensitive polymer microgel. A microgel is a hydrous polymer gel particle with a diameter ranging from tens of nanometer to one micrometer. Only one research attempt utilizing polyacrylamide microgel is known to date. However, it has not performed better than simple LPA.²⁷ As explained in section 3.2.2, some polymers have LCST or UCST and change their hydrophilicity with temperature. Microgel particles made of such thermosensitive polymers²⁸ show phase transition between emulsion (with low solubility) and gel particle (with high solubility), depending on temperature change. Taking advantage of this behavior, it may be possible to inject a polymer as a low viscosity emulsion into a capillary, while electrophoresis could be conducted through a robust gel network. In addition, if core-shell particles consisting of a thermosensitive core and highly water-soluble shell were prepared, electrophoretic performance could be even better than that of sole thermosensitive microgel particles. Such microparticles have never been applied to DNA electrophoresis.

PNIPAM microgel²⁹ and Poly(N-ethylacrylamide) microgel³⁰ are the only thermosensitive microgels that have been prepared to date. Since both of them have LCST and their phase transition occurs around room temperature, they are not suitable for application in DNA electrophoresis, which is typically performed at around 50 to 70°C. A problem with synthesizing the UCST type of microgel is the need to develop a new polymerization method. Microgel of LCST polymer like PNIPAM can be prepared by basic emulsion polymerization or suspension polymerization because these polymers cannot dissolve in water and they form an emulsion at polymerization temperature (typically more than 50°C).

²⁷ Doherty, E. A. S.; Kan, C. W.; Barron, A. E.; *Electrophoresis* **2003**, *24*, 4170-4180

²⁸ Pelton, R. *Adv. Col. Interf. Sci.* **2000**, *85*, 1-33

²⁹ For example, Jones, C. D.; Lyon, L. A. *Macromolecules* **2000**, *33*, 8301-8306

³⁰ Lowe, J. S.; Chowdhry, B. Z.; Parsonage, J. R.; Snowden, M. J. *Polymer* **1998**, *39*, 1207-1212

Most UCST polymers, such as PPE, cannot be polymerized by these polymerization method and, thus, a new method should be considered first. This approach requires close collaboration with synthetic polymer scientists.

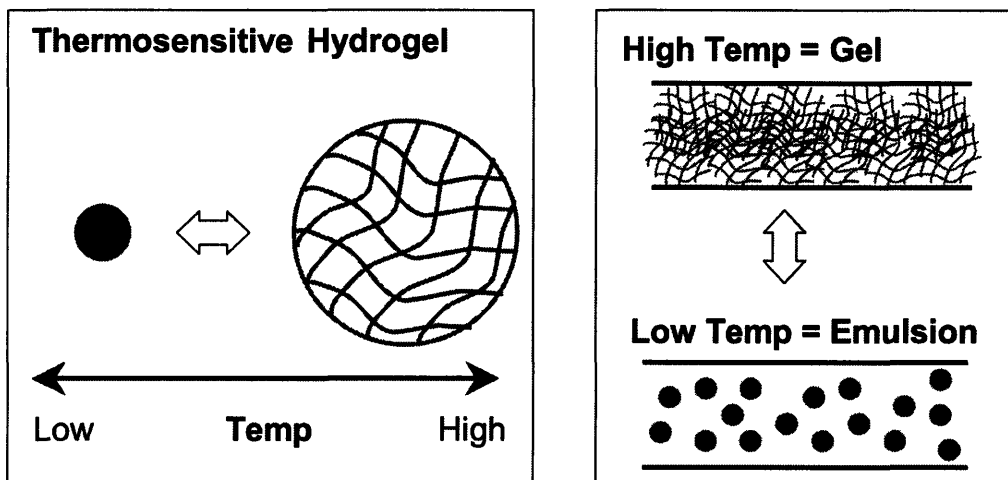


Figure 38 Thermosensitive microgel

3.6 Summary

In an effort to improve the operational efficiency of ABI3730xl, a DNA sequencer based on multicapillary electrophoresis, effect of polymer media on electrophoresis performance was investigated. The ultimate goal of this study is creating new polymer materials, which allow the current genome-sequencing process higher throughput. In this research, as a first step, various water-soluble polymers including POP7, LPA, and DMA were examined as polymer media for the ABI3730xl.

Unfortunately, no polymers investigated in this research exhibited better capabilities than the default polymer, POP7. However, it was confirmed that simple and inexpensive LPA could replace POP7 with some optimization. The best results for lead length with such alternative polymers were 500bp ($R=0.4$) or 700bp ($R=0.3$), while POP7 is 700bp ($R=0.4$).

New research procedures consisting of raw electropherogram data extraction, peak fitting, and the calculation of peak resolution were established. Chemical analysis of POP7 was conducted to show that the main ingredient of POP7 is not widely-believed DMA, but LPA with a narrow molecular weight distribution. Possible future approach using thermo-sensitive water-soluble polymer is proposed.

4. Analysis of the Genome-Sequencing Tool Industry

4.1 Introduction

Through his internship at the Broad Institute, the author worked intimately with ABI3730xl, and gained an appreciation for ABI's huge presence in the sequencing-tool industry. ABI has maintained a monopoly over the industry since its emergence, and its products have been the de facto standard in the genome-sequencing field. Interested in ABI's dominant status in this relatively new business sector, the author conducted an analysis of the genome-sequencing tool industry. Important questions raised include how ABI's monopoly status has been established, whether its dominance will continue in the future, and how the industry can develop in the future.

4.2 History of Genome-sequencing Technology^{4,31}

The widely-accepted genome-sequencing technique, the Sanger method, was established by Professor Fred Sanger of the University of Cambridge, who sequenced the complete 5,386-letter genetic code of a virus called *PhiX174* in 1977. For this effort, he won his second Nobel Prize in Chemistry in 1980, sharing it with Walter Gilbert of Harvard University and Paul Berg of Stanford University. For a more detailed description of the current DNA-sequencing process, refer to Section 2.2.2.

Figure 38 illustrates the original Sanger method's procedure. Sanger originally employed a radio isotope (RI) such as ³²P and ³⁵S to label dd-NTP reaction terminators. This method requires four distinct PCR reactions to read out the base code of a single DNA. In the first reaction, only A-type ddNTP was labeled by RI, then C, T, and G-types of dd-NTP were RI-labeled for each subsequent reaction. These four reaction solutions were electrophoresed in separated lanes on an acrylamide slab-gel. RI-labeled DNA fragments were then detected with autoradiography.

³¹ Shreeve, J.; *The Genome War – How Craig Venter Tried to Capture the Code of Life and Save the World*; Alfred A. Knopf: New York, 2004

A problem with this method was the very low throughput to read bases. Only a limited number of DNA could be analyzed simultaneously. In addition, each band had to be read manually by a researcher. Thus the cycle time was almost one day, and operation throughput was only hundred base pairs (bp) per day. Supposing this method were applied to the human genome with three billion bp, it would take a hundred thousand years to read all the base codes.

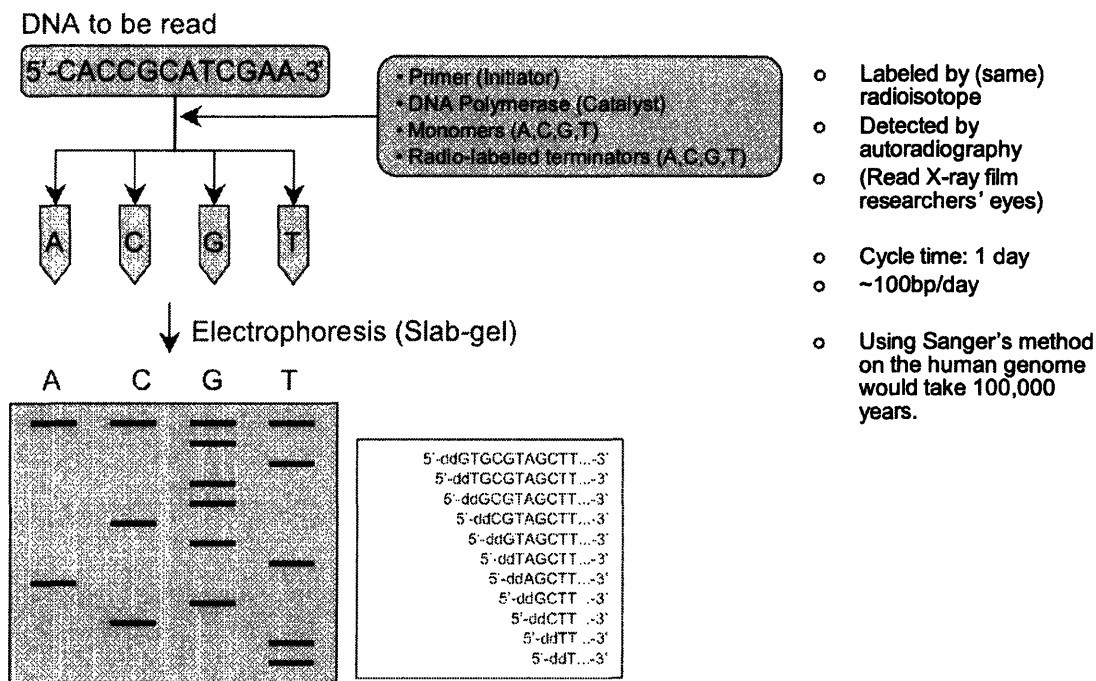


Figure 38 Outline of the original Sanger method

In the wake of Sanger's innovation around 1980, a lot of researchers tried to automate the Sanger method. For instance, five years prior to the start of the International Human Genome Project, Professor Akiyoshi Wada of the University of Tokyo launched a project among Japanese research institutes and companies in 1981.³² The project proposed a prototype for an automatic DNA analyzer but was eventually dismissed after just three

³² In the Wada Project, Hitachi invented high-throughput detection technology, which was eventually incorporated into ABI's products.

years.³³ Some American companies, like DuPont, also made some efforts to design an automatic DNA sequencer, as did the American Department of Energy (DOE).

At the time, Applied Biosystems Inc., or ABI, established in 1981 by Professor Leroy Hood and Dr. Mike Hunkapiller of California Institute of Technology, also worked on the automation of the DNA sequencer. In 1986, Dr. Hunkapiller's younger brother Tim invented the "four-color dye method," making the epoch that allowed ABI to successfully launch the world's first automatic DNA sequencer, ABI Prism370A.

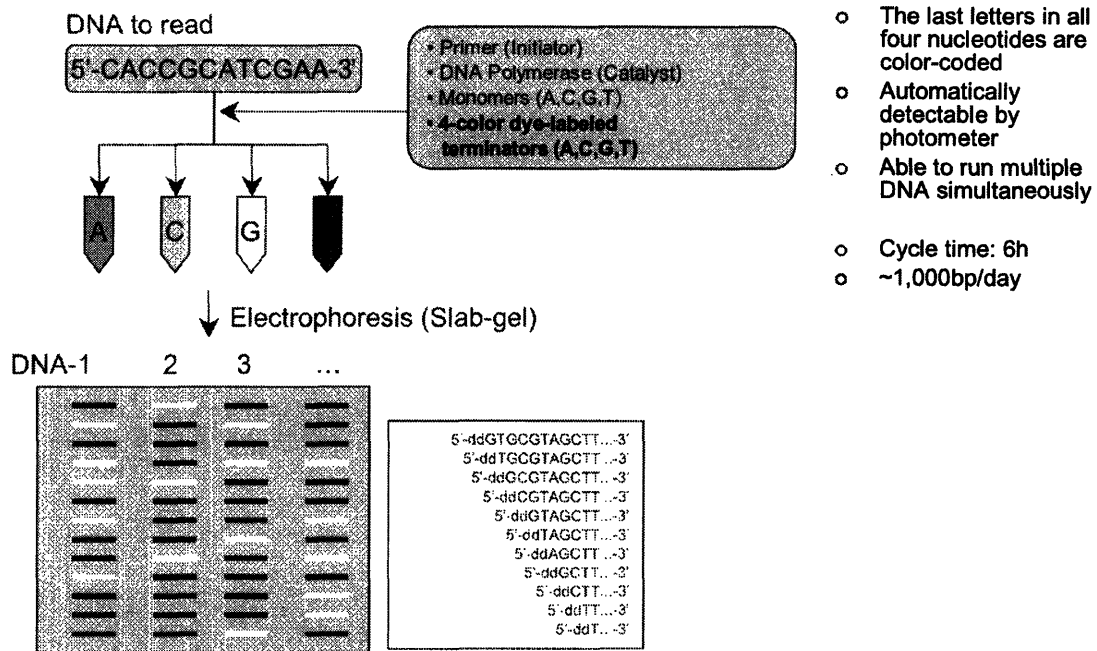
Figure 39 illustrates the "four color dye method." This method is characterized by the use of four different colors of electron transfer dyes, instead of radioisotope, to label ddNTP. Excited by a certain laser beam wavelength, these dyes illuminate four different colors corresponding to the four different base nucleotides; i.e. they are green for Adenine, blue for Cytosine, yellow for Guanine, and red for Thymine. The four-color method requires only one PCR reaction and one lane of electrophoresis per single DNA sample. In addition, since each band is automatically detected by a spectrometer, the operation is dramatically more efficient than the original Sanger method. The cycle time of the sequencing operation is less than six hours, while the daily throughput is 100bp – almost ten times shorter than the original Sanger method.

After the "four-color dye method" was invented, several chemical and biological analysis instrument companies conducted research on applying the multi-capillary electrophoresis to DNA sequencers, instead of using the inefficient slab-gel electrophoresis. ABI again led in the intense competition and finally released the world's first automatic, multicapillary DNA sequencer ABI Prism3700 in 1997. DNA separation was carried out in a fused-silica capillary (25 to 100 μm in diameter) filled with an aqueous solution of linear polymer, instead of the acrylamide slab-gel. ABI Prism3700 was equipped with 96 capillaries and a CCD detector, which allowed it to analyze 96 DNA samples simultaneously.³⁴ Figure 40

³³ Segawa, S., "Koshiba Project and Wada Project"(in Japanese), <http://www.mainichi-msn.co.jp/kagaku/rieki/column/news/20040531org00m040002000c.html>, Mainichi Shimbun, May 31, 2004,

³⁴ ABI established an alliance with Hitachi for access to its basic patent for detecting mechanisms as well as precise control technology, including robotics and mechatronics, used for automating operations. ABI has also contracted this Japanese electronics giant to manufacture the ABI3730xl.

outlines ABI Prism3700's detection mechanism. Argon laser beam comes from lateral side of the capillaries, while a perpendicularly-held CCD camera simultaneously detects fluorescence from each capillary. In addition, the cycle time was drastically reduced, since a higher electric field can be applied. For example, the cycle time of ABI's latest model ABI3730xl is only one or two hours, and the sequencing throughput reaches almost 600,000bp per day. The third-generation sequencing technique really opened the door for the deep investigation of human genome. Without the multi-capillary DNA sequencer, neither the International Human Genome Project nor Celera Genomics would have successfully finished by the 2001 deadline, or maybe they would have never started.



- The last letters in all four nucleotides are color-coded
- Automatically detectable by photometer
- Able to run multiple DNA simultaneously
- Cycle time: 6h
- ~1,000bp/day

Figure 39 Outline of the improved “four-color dye method”

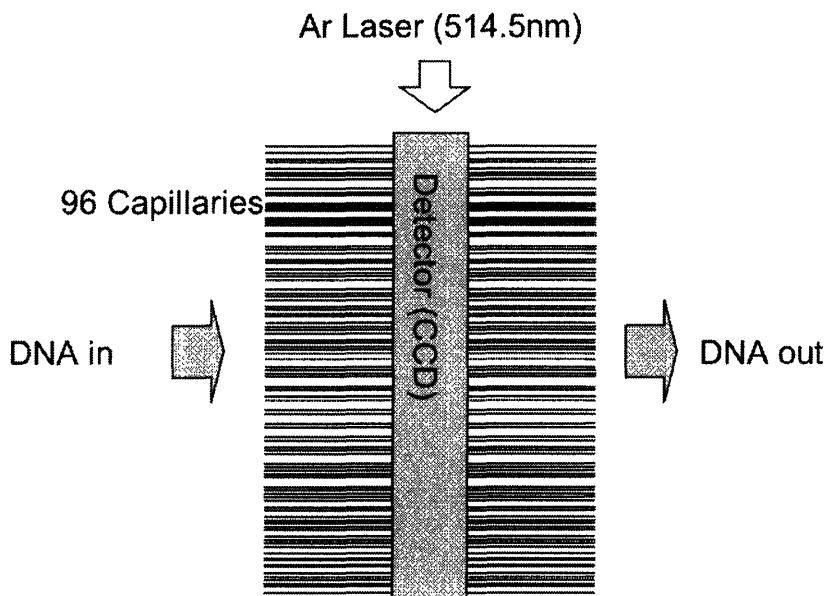


Figure 40 Outline sketch of the multicapillary electrophoresis detection mechanism

4.3 Emergence of the Genome-Sequencing Tool Industry

Figure 41 illustrates the historical facts regarding genome-sequencing technology development and the genome-sequencing tool industry. After the release of ABI370A, the first automatic DNA sequencer powered by its proprietary four-color dye technology, ABI successfully attained a monopolistic market position in the sequencing tool industry. Even though a couple of competitors, like Molecular Dynamics (now Amersham Bioscience), existed, ABI claimed almost 90% of the market share by 1995. ABI had already been declared the dominating force in the genome research arena. "While the market was specialized, Hunkapiller's enterprise was exploiting its dominant position so freely that its customers grumbled that the abbreviation stood for 'Arrogance Beyond Imagination'."³¹ In 1993, Perkin Elmer Corporation, a manufacturer of analytical and optical instruments, had acquired ABI.

In 1997, when the high-throughput, multicapillary DNA sequencer ABI3700 was released, the CEO of Perkin Elmer, Tony White, decided to bet the firm's future on this cutting edge genome analysis technology. He spun off ABI from Perkin Elmer and established Celera Genomics to start a new genetic information business powered by its high-throughput DNA sequencer.³⁵ ABI provided the ABI3700 sequencer not only to Celera but also to the International Human Genome Project. Due to its excellent data quality, high throughput, highly automated system, and wide technical support, ABI3700 quickly established a dominant position in the sequencing tool industry.

³⁵ The Perkin Elmer trade name and analytical and optical instrument business were sold off in 1999.

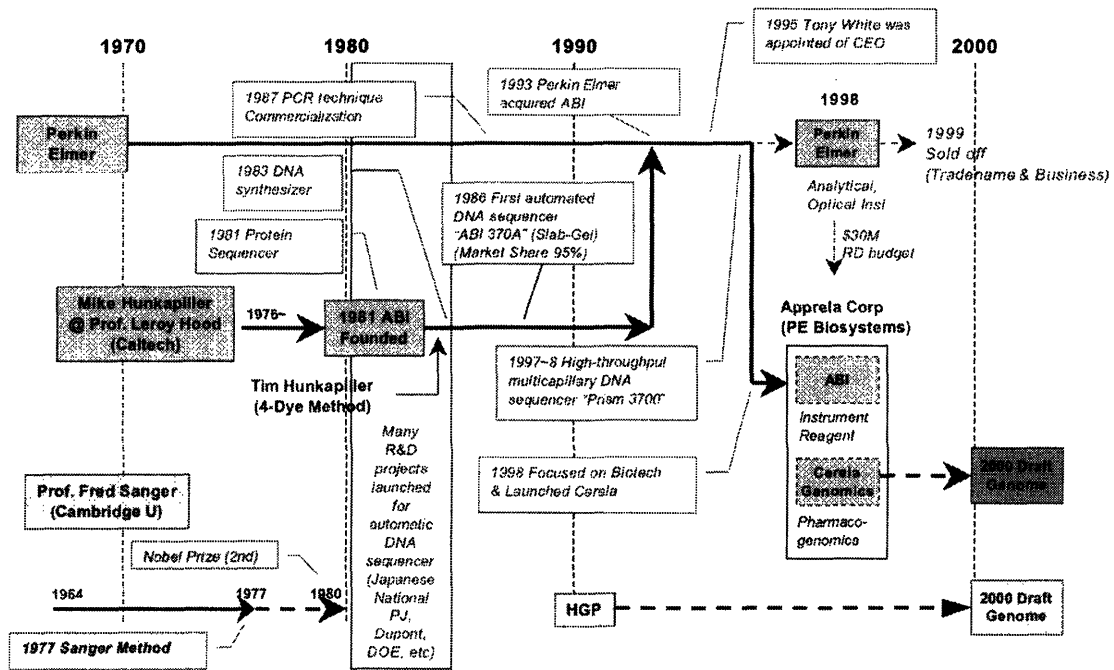


Figure 41 Chronological diagram of the history of genome-sequencing technology and the genome-sequencing tool industry.

4.4 Analysis of the Genome-Sequencing Tool Industry

4.4.1 Industry S-curves

Figure 42 shows the S-curves for the genome-sequencing tool industry illustrated in terms of daily throughput. Three S-curves correspond to the original Sanger method (first generation), the four-color dye method (second generation), and the fully-automated, multicapillary DNA sequencer (third generation), respectively. The first generation of technology achieved only 100 bases per day, while the second and third generation technologies achieved 1,000 and 600,000 bases of daily output. Thus, the four-color dye method and the automatic multicapillary system can be considered “disruptive” technologies to previous technologies. Note that Applied Biosystems developed both disruptive technologies.

The industry currently seems subject to “natural technological limits.” The current multicapillary electrophoresis system has matured enough technologically, having gone through eight years and one revision after the first product release in 1997. Considering that the four-color dye method (2G) and the multicapillary system (3G) were released at six and eleven years respectively after the release of the previous technology, the next-generation technology could be released anytime now. In fact, various next-generation sequencing technologies have already been studied in many organizations.³⁶ Such “fourth-generation” technologies include both straight extensions of the current Sanger-type technology (e.g. Microelectrophoretic sequencing by the Whitehead BioMEMS laboratory) and new sequencing methods based on alternative technologies (e.g. Hybridization sequencing by Affymetrix and Perlegen, and Pyrosequencing by 454 Corporation). The Dev department at the Broad Institute has long worked with some of these research groups and companies to assess when such technologies can be applied to its high-throughput genome-sequencing production facility. Interestingly, ABI has not played an important role the next-generation of sequencing technology. This means that ABI may not be able to maintain its current monopolistic status in this industry, unless it is conducting in-house research behind-the-scenes or it acquires a company that is conducting research on the next-generation technologies.

³⁶ Shendure, J., Mitra, R. D., Varma, C., and Church, G. M., *Nature Reviews*, 2004,5, 335-344

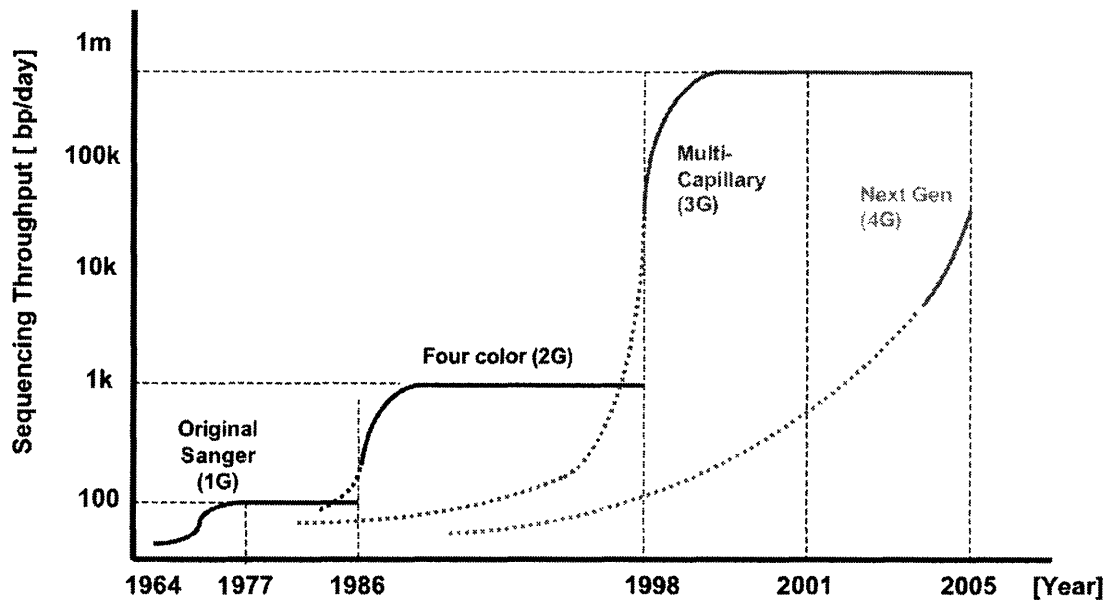


Figure 42 S-curves for the genome-sequencing tool industry

4.4.2 Uniqueness and Complementary Assets to Create and Capture Value

The first and largest value created in this industry is the invention of the Sanger method by Professor Sanger of the University of Cambridge. Although he received the Nobel Prize for this achievement, his invention went public and he himself has not captured any profit value from the value creation.

Based on Sanger's ground-breaking technology, ABI has claimed virtually all the technological achievement for both electrophoresis technologies and automation operation systems. The three largest technological advancements made by ABI and its affiliates are the four-color dye method, the self-coating polymer, the multicapillary detection system, and the precisely controlled automation system. These technologies are patented by ABI

and its affiliate, Hitachi. ABI created the second and third generation of sequencing machines, based on these unique technologies.

ABI also built many sets of “complementally assets”. ABI’s business model goes beyond simply selling instruments to customers. After selling the sequencer, ABI continues to provide consumables, such as polymer solution, running buffer, standard DNA samples, etc. These consumables are totally black-boxed. Yet, the quality of sequenced data obtained with this package of expensive consumables is guaranteed by ABI. On-site customer support is available for large customers. For example, ABI’s on-site technicians at the Broad are on call seven days per week. Other complementary assets include production capability. Manufacturing of DNA sequencers are outsourced to Hitachi, a Japanese leading electronics manufacturer. Working with Celera genomics and HGP for long time, ABI also maintains a well-recognized brand in the biotech industry and research communities. Table 8 summarizes ABI’s technological uniqueness and complementary assets.

Table 8 ABI’s uniqueness and complementary assets

Uniqueness	Complementary Asset
Proprietary Technology on Electrophoresis	Customer support (7d/w)
- Four-color dyes labeling	Total packaging including consumables
- Self-coating polymer	Quality assurance
- Multicapillary detection system	Manufacturing capability (Hitachi)
Automation technology (Hitachi)	Brand

ABI not only created the most value for the second-, and third-generation technologies through outstanding technological uniqueness, but it also captured this value by taking advantage of strong complementary assets. Its collaboration with Hitachi plays a very important role both for creating and capturing value. ABI has taken charge of research and prototyping, while Hitachi has worked on system development and manufacturing. As an original member of the Japanese national project, Hitachi has individually created value on the automation of the DNA sequencer. Among ABI’s technological uniqueness, the multicapillary detection system and the precisely-controlled, automatic operation system are Hitachi’s proprietary technologies. Hitachi’s rich experience in the production of

precisely-controlled, automatic mechatronics systems has also been considered an important complementary asset.

Figure 43 shows the Uniqueness–Complementary Assets matrix for the genome-sequencing tool industry. Since uniqueness is fairly easy to maintain through registering patents and tightly holding complementary assets, this industry can be placed in “The Big Win” category. This means that the industry allows a limited number of companies to attain large-scale success within their realms, even allowing companies, like ABI, to establish easily monopolistic status.

Although the industry has so far created much value, it may not have sufficiently captured it yet. ABI seems to successfully capture most present potential industry earning (PIE), however the current market is still limited to research. It is unclear if ABI can continue to capture the most value when the market extends to the medical and diagnostics markets in the future. Then, the next-generation of technology could dominate the industry.

		Complementary Assets	
		Available	Tightly Held
Uniqueness	Easy to Maintain	The Inventor's Dream <i>Qualcom in CDMA. Biotech?</i>	The Big Win <i>Nylon, Viagra, Kodak, Windows</i> ABI
	Hard to Maintain	Few Returns to Innovation <i>Linbeck, Int. Wind. Wiper</i>	Comp. Assets Play <i>Dell, P&G, Caldon. M.Lynch</i>

Figure 43 Henderson’s Uniqueness vs. Complementary Assets Matrix

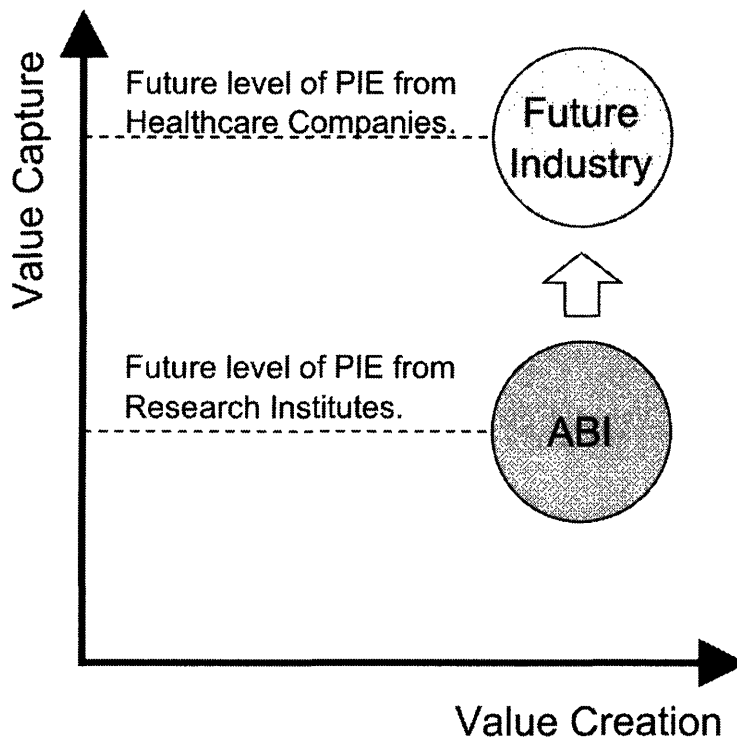


Figure 44 Value Creation vs. Value Capture

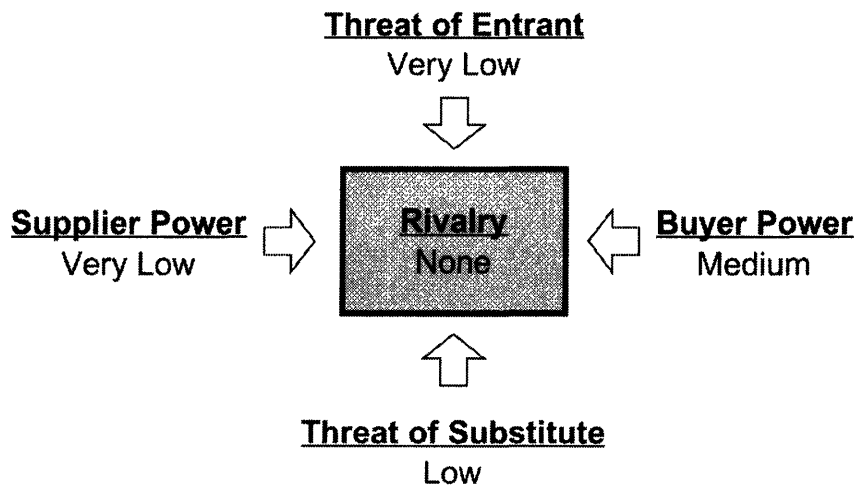


Figure 45 Five forces analysis of the genome-sequencing tool industry

4.4.3 Industry Structure – Present and Future

4.4.3.1 Five Forces Analysis for Current Industry

The five forces analysis for the current genome-sequencing tool industry is presented in Figure 45.

Rivalry among existing firms: None

The industry is monopolized by ABI. This position is supported by both technological uniqueness and complementary assets as described in the previous section.

Supplier's power: Very low

From ABI-Hitachi's viewpoint, suppliers are merely vendors of general components and materials, such as CCDs, lasers, pumps, electric power units, plastic plates, polymeric materials, etc. Since the number of sequencers produced is very small, ABI-Hitachi cannot afford to customize any components, and by necessity, they use a typical grade of components. Hence, the only difference among suppliers is price.

Buyer's power: Medium

Currently only public and private research institutions are customers. PIE still remains small and is even shrinking after the completion of the Human Genome Project due to reducing government funds for genome-sequencing research. (Many research institutions rely on government funds to purchase expensive DNA sequencers.) Requests for price discounts are fairly strong, given the strong competition among genome centers to obtain government funding. On the other hand, there are virtually no alternative products and buyers cannot help but buy ABI's products eventually. Most of the DNA sequencers are so automated and black-boxed that buyers must rely on the companies in the industry for maintenance, etc.

Threat of new entrants: Very low

The market is relatively small and now shrinking. The current technology has matured, and

the leaning curve effect for incumbents cannot be neglected. Both ABI's technological uniqueness and complementary assets are very strong, and the entrance barrier is very high. Entrance would be easier with more inexpensive technology.

Threat of substitute products or services: Low

At this point, the threat of substitute products is fairly low. Cheap alternative technologies are, however, currently under development by some research groups, such as the Whitehead BioMEMS laboratory, Affymetrix, Illumina, 454 Corporation, etc., whereas ABI does not seem to be working on such next-generation technologies. In the future, the fourth-generation sequencing technology may replace ABI's market presence, which is based on third generation technology.

4.4.3.2 Industry Evolution

In the near future, two radical transformations are expected in this industry. One is a change of customers, and the other is the emerging fourth-generation technologies. Although the current customer base is limited to public and private research institutes and PIE is still small, companies in the healthcare industry, such as pharmaceutical companies and hospitals, have long been expected to be the new customer base of this industry. This huge market forecast had actually been the persuading argument to bring much government and private funding into this industry during the HGP era. Recent research shows, however, that will take longer than predicted for the healthcare industry to become a new customer.

In the healthcare industry, genome-sequencing technology will be utilized for the single nucleotide polymorphisms-based diagnostics (SNP, pronounced "snip"). "SNP" is the technical jargon for a single-letter mutation or coding error on the genomic code. SNPs are subtle and seemingly random variations among the smallest components of genes (nucleotides). Some inherited SNP variations are believed to cause genetically-based diseases and others are responsible for genetically-inherited characteristics like eye color. The SNP discovery challenge is to find the individual genetic building blocks that correspond to various illnesses; this information may then lead to cures through gene

therapy.³⁷

Figure 46 shows an example of SNPs-based diagnostics. Assume that there are three SNPs related to Alzheimer's disease and persons with SNP1, SNP2, and SNP3 have a 20%, 10% and 5% higher risk of Alzheimer's disease. The person who has all three of the SNPs may have a 100% higher risk of the disease. Since 99.9% of human genes are common, SNPs are just 0.1% of human gene, but it still corresponds to three million of base pairs. Many researchers are now pursuing the challenge of discovering all important SNPs, but it will take quite a while before enough information, including interactions among each SNPs, is collected.

The genome-sequencing tools market for diagnostic purposes will open when SNP-based diagnostics have been sufficiently realized to be useful for at least one disease. The industry will have to wait quite a bit to open this huge market. It is not clear if ABI can maintain its current advantage, both in technological uniqueness and complementary assets, until then, given upcoming fourth-generation technology.

Reference (Normal)	A T C G A T T A C G A T T A
SNP1 (20% Higher Risk)	A T C G (T) T T A C G A T T A
SNP2 (10% Higher Risk)	A T C G A T T A (G) G A T T A
SNP3 (5% Higher Risk)	A T C G A T T A C G A (C) T A
SNP1+2+3 (100% Higher Risk)	A T C G (T) T T A (G) G A (C) T A

Figure 46 Schematic illustration of SNPs-based diagnostics

³⁷ Wolff, G., *The Biotech Investor's Bible*, New York, John Wiley & Sons, Inc., 2001

4.4.4 Value Chain and Boundaries

Figure 47 shows the value chain for the genome-sequencing tool industry. In addition to its original strength in basic research and product packaging, ABI has established a strong sales capability, including customer support and consumable sales. In addition, through its collaboration with Hitachi, ABI has also integrated an assembly (manufacturing) capability. Through this vertical integration, ABI has successfully captured most of the value which had been historically created in the industry. On the other hand, other players in the industry are typically small research organizations or start-up companies, which tend to remain in basic research and product packaging.

As discussed in the previous section, the structure of the value chain currently does not seem very dynamic. Only with a disruptive technology, could a new player be able to step into ABI's current position, although it will not be so easy to defeat ABI by only technological uniqueness. To appeal to customers who feel comfortable with ABI's integrated service offering, the new company would need to establish a set of complementary assets, perhaps through another vertical integration.

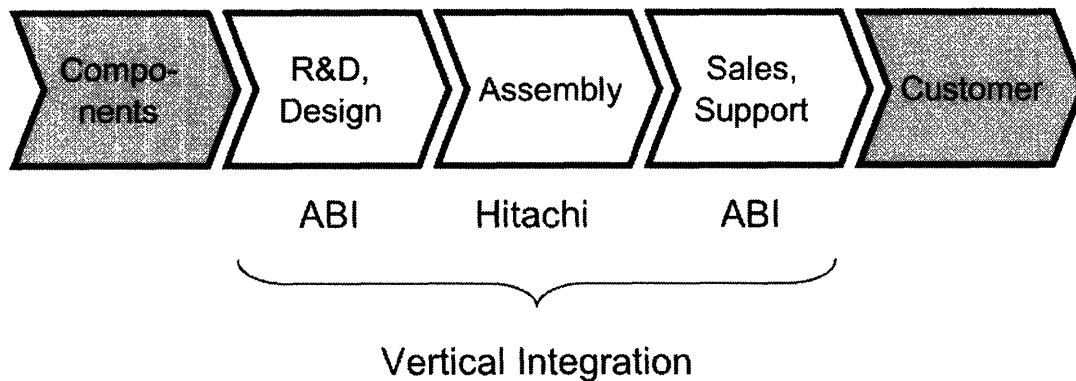


Figure 47 Value chain for the genome-sequencing tool industry

4.4.5 Strategy for the Future of ABI

After the completion of the “Genome War” – the Human Genome Project competition, Craig Venter, the former CEO of Celera Genomics, was fired. Celera and ABI are now managed under the same umbrella by Applera Corporation, the stock-holding company for two of the most outstanding companies in the genomics arena. Current management is focusing on genome-based diagnostics with synergy from Celera and ABI. Celera Diagnostics, a joint venture between Celera and ABI is now making every effort to develop new SNP-based genomic diagnostics, competing among world leading research institutes and start-up companies. However, this strategy only aids the decline of ABI and Celera’s stock prices.

Due to recent serious financial concerns, ABI reduced its workforce by 3.5%, or 145 employees in the first fiscal quarter of 2005. Many of the eliminated positions were in research and development (R&D), but the sales and manufacturing departments were also impacted. ABI had made even more drastic workforce reductions in December 2002, laying off five hundred employees, however, those cost reductions had little effect on relative spending as a percentage of revenues, and little impact on earning growth. Business conditions remain challenging, and lack of revenue growth still remains ABI’s number one dilemma.

Considering ABI’s outstanding technological uniqueness in the genome-sequencing technology and its outstanding complementary assets acquired through vertical integration, the author recommends that ABI should try to remain the number one provider of genome-sequencing tools even when next-generation technology becomes available. ABI should invest more on next-generation sequencing technology with much higher throughput and greater cost advantage, instead of waiting for somebody else to discover whole SNPs information, opening the door to new genomic diagnostics. ABI may want to concentrate more on in-house basic research or invest in hopeful start-up companies, such as 454 Corporation, which are doing research on the different technologies within ABI’s technological competency.

ABI should also work more with potential customers, such as healthcare management organizations (HMOs), health insurance companies, and pharmaceutical companies. Using already known SNPs information, ABI can work with those companies to estimate disease

occurrence risk more accurately. Although the number of known SNPs is still few, such activities would help the company establish a bridge with the future genomic diagnostic market, which could be another complementary asset of ABI.

4.5 Summary

In the genome-sequencing tool industry, there were three technological innovations to date. The first one was made by a Nobel laureate, and the next two were both derived by ABI. Based not only on outstanding technological uniqueness, but also strong complementary assets, ABI established monopolistic status in the industry. To build complementary assets, ABI has vertically integrated several components in the value chain, including manufacturing capability and technological support.

After the accomplishment of the Human Genome Project, however, ABI is struggling with low profitability and a diminished impetus for future growth. The reason is, contrary to early optimistic forecasts, the market's expansion to the whole healthcare industry will be delayed. Based on this industry analysis, the author recommends that ABI should focus more on next-generation genome-sequencing technology to maintain its current monopolistic position even in the upcoming genomic diagnostics period.

5. Conclusion

Recent advances in information technology, bioengineering techniques, and data-processing methodology have opened the door to modern research approaches to genomics and have inspired ambitious efforts, like the International Human Genome Project. Playing a main role in the HGP, the Broad Institute established the genome-sequencing capability with the world's highest level of throughput. Even after the completion of the HGP, the Broad Institute continues to address improvements in sequencing throughput with an eye toward future genomics-based medication.

In the pursuit of operational efficiency, the author attempted to find a way to improve the efficiency of the genome sequencer, an instrument that reads a DNA base code through capillary electrophoresis technology, which plays a central role despite certain obstacles in the whole genome-sequencing process.

Chapter 3 examined polymer media, a determining factor in sequencing throughput, that is used for the electrophoresis, to Figure out how it could be modified from the current standard. This was the first attempt of this sort at the Broad Institute. The methodology employed here involved the following stages.

1. Study past research efforts about polymer media
2. Establish a new experimental protocol to handle other polymers
3. Analyze the chemical composition of ABI's default polymer (POP7)
4. Run genome-sequencing on several polymer materials to learn how they work

Conducting research on past studies, the author learned that LPA and DMA are known as the best polymers for DNA electrophoresis, and that many water-soluble polymers demonstrate an inferior performance compared to these basic polymers. Although structured polymers have the possibility to work better than simple acrylamides, the lack of polymerization technique still remains a hurdle for implementation.

The author introduced a new method to extract raw data from the sequencer's highly black-boxed internal process, which allows analyzing electrophoretic performance under various running conditions. This new analysis makes a wide variety of further investigations on ABI3730xl possible beyond the new polymer research.

A chemical analysis of the current default polymer media, POP7, showed that it consists mainly of LPA, although a trace amount of DMA was detected. This DMA could be mixed or co-polymerized with LPA for self-coating purposes. The average molecular weight of POP7 was around 800k with relatively narrow distribution. Thus, POP7 could be replaced by a low-cost, general purpose LPA. Assuming POP7 were completely replaced with LPA800k, the annual cost reduction at the Broad Institute was calculated to be possibly \$383,000 per year, which corresponds to a 94.5% reduction in the current polymer cost.

Chapter 4 revised the genome-sequencing technology and analyzed the genome-sequencing tool industry to learn how ABI established its current monopolistic status. ABI achieved two of the three technological innovations. By succeeding on building strong complementary assets based on outstanding technological uniqueness, ABI currently enjoys monopolistic status in the industry. In order to build complementary assets, ABI vertically integrated some components in the value chain, including manufacturing capability and technological support.

After the completion of the Human Genome Project, however, ABI seems to be struggling with low profitability and little technological advantage for future growth. This is due to the fact that their customer base still remains only in research institutes, and that market expansion to the healthcare industry will take longer than predicted. Based on the industry analysis, the author recommends that ABI should focus more on the next-generation of genome-sequencing technology to allow the company to maintain its current monopolistic position in the future genomic medication market.