RIKEN Integrated Sequence Analysis (RISA) System—384-Format Sequencing Pipeline with 384 Multicapillary Sequencer

Kazuhiro Shibata,[1,10] Masayoshi Itoh,[1,10] Katsunori Aizawa,[1] Sumiharu Nagaoka,[1,8] Nobuya Sasaki,[1,9] Piero Carninci,[1] Hideaki Konno,[1] Junichi Akiyama,[1] Katsuo Nishi,[1] Tokuji Kitsunai,[2] Hideo Tashiro,[2] Mari Itoh,[1] Noriko Sumi,[1] Yoshiyuki Ishii,[1] Shin Nakamura,[1] Makoto Hazama,[1] Tsutomu Nishine,[1] Akira Harada,[1] Rintaro Yamamoto,[1] Hiroyuki Matsumoto,[1] Sumito Sakaguchi,[1] Takashi Ikegami,[1] Katsuya Kashiwagi,[1] Syuji Fujiwake,[1] Kouji Inoue,[1] Yoshiyuki Togawa,[1] Masaki Izawa,[1,4] Eiji Ohara,[4] Masanori Watahiki,[4] Yuko Yoneda,[4] Tomokazu Ishikawa,[1,5] Kaori Ozawa,[5] Takumi Tanaka,[5] Shuji Matsuura,[5] Jun Kawai,[1] Yasushi Okazaki,[1] Masami Muramatsu,[1] Yorinao Inoue,[6] Akira Kira,[7] and Yoshihide Hayashizaki[1,3,11]

[1]Genome Exploration Research Group, RIKEN, Genomic Sciences Center (GSC) and Genome Science Laboratory, RIKEN, Tsukuba Institute, Core Research of Evolutional Science and Technology (CREST), Japan Science and Technology Corporation (JST), Tsukuba-shi, Ibaraki, 305–0074, Japan; [2]Advanced Engineering Center, RIKEN, Wako-shi, Saitama, 351–0198, Japan; [3]Medical School, Tsukuba University, Tsukuba-shi, Ibaraki 305–8575, Japan; [4]Research & Development, Nippon Gene Co., Ltd., Arakawa, Toyama, 930–0834, Japan; [5]Osaka Research Laboratories, Wako Pure Chemical Industries, Ltd., Amagasaki, Hyogo 661–0963, Japan; [6]RIKEN Harima Institute at SPring-8, Sayou, Hyougo 679–5148, Japan; [7]RIKEN, Wako-shi, Saitama, 351–0198, Japan

The RIKEN high-throughput 384-format sequencing pipeline (RISA system) including a 384-multicapillary sequencer (the so-called RISA sequencer) was developed for the RIKEN mouse encyclopedia project. The RISA system consists of colony picking, template preparation, sequencing reaction, and the sequencing process. A novel high-throughput 384-format capillary sequencer system (RISA sequencer system) was developed for the sequencing process. This system consists of a 384-multicapillary auto sequencer (RISA sequencer), a 384-multicapillary array assembler (CAS), and a 384-multicapillary casting device. The RISA sequencer can simultaneously analyze 384 independent sequencing products. The optical system is a scanning system chosen after careful comparison with an image detection system for the simultaneous detection of the 384-capillary array. This scanning system can be used with any fluorescent-labeled sequencing reaction (chain termination reaction), including transcriptional sequencing based on RNA polymerase, which was originally developed by us, and cycle sequencing based on thermostable DNA polymerase. For long-read sequencing, 380 out of 384 sequences (99.2%) were successfully analyzed and the average read length, with more than 99% accuracy, was 654.4 bp. A single RISA sequencer can analyze 216 kb with >99% accuracy in 2.7 h (90 kb/h). For short-read sequencing to cluster the 3′ end and 5′ end sequencing by reading 350 bp, 384 samples can be analyzed in 1.5 h. We have also developed a RISA inoculator, RISA filtrator and densitometer, RISA plasmid preparator which can handle throughput of 40,000 samples in 17.5 h, and a high-throughput RISA thermal cycler which has four 384-well sites. The combination of these technologies allowed us to construct the RISA system consisting of 16 RISA sequencers, which can process 50,000 DNA samples per day. One haploid genome shotgun sequence of a higher organism, such as human, mouse, rat, domestic animals, and plants, can be revealed by seven RISA systems within one month.

Present addresses: [8]Department of Applied Biology, Faculty of Textile Science, Kyoto Institute of Technology, Matsugasaki, Kyoto 606–8585, Japan; [9]Graduate School of Pharmaceutical Sciences, University of Tokyo, Bunkyo, Tokyo 113, Japan.
[10]These authors contributed equally to this work.
[11]Corresponding author.
E-MAIL rgscerg@rtc.riken.go.jp; FAX +81 298 36 9098.
Article and publication are at www.genome.org/cgi/doi/10.1101/gr.152600.

Various genome projects have produced huge amounts of sequences and determined the genome sequences of several organisms using the whole-genome shotgun approach or BAC shotgun method. The 96-format capillary sequence system has facilitated these projects greatly.

Since 1995, RIKEN has been conducting a mouse

encyclopedia project. The aims of this project are the collection of all full-length mouse cDNAs (Phase I) using a novel full-length cDNA cloning strategy (Carninci et al. 1996, 1997, 1998; Carninci and Hayashizaki 1999), determination of full sequences of those clones (Phase II), and mapping of those clones on the mouse genome (Phase III). The advantages of the full-length cDNA approach are the ability to directly identify primary structures of the proteins and to synthesize complete proteins from full-length cDNA clones. This full-length cDNA approach has requirements that differ from those for the whole-genome shotgun or BAC shotgun strategy. The strategy of Phase I is to cluster those clones with a 100-bp 3′ sequence whose length is calculated to offer the best performance for the clustering based on the frequency of the isoform gene in mouse and sequence accuracy (H. Konno, Y. Fukunishi, K. Shibata, M. Itoh, P. Carninci, Y. Sugahara, and Y. Hayashizaki, in prep.). To finish Phase I in three years would require a pipeline system designed to achieve a production rate of >40,000 samples/day.

The 384-format automated sample sequencing pipeline throughout the entire sequencing process is very advantageous as a large-scale sequencing system because it reduces the ID error rate and enables simple management of samples. ID error is the incorrect correspondence between 3′-end sequence of a clone in a well and a tag of the clone, that is, misidentification of clones. The basic 384-format RISA sequencing pipeline consists of colony picking, template preparation, reaction, and sequencing. For the colony-picking process, Q-bot (Genetix, UK) was employed to pick up randomly plated clones and inoculate them onto a 384-well plate. For the template preparation process, we developed a RISA inoculator to inoculate *Escherichia coli* and also developed a RISA filtrator to harvest these cells into the RISA filtration plate for the subsequent step to prepare plasmid DNA by RISA plasmid preparator (Itoh et al. 1999). All of these steps are fully automatic. These samples are subjected to the high-throughput thermal cycler (RISA thermal cycler, GS384, Sasaki et al. 1997) for the sequencing reaction, followed by the sequencing step that is achieved by the RISA sequencer system. These devices were originally developed by us and the RISA sequencer is an especially essential sequencing device at the downstream end of the high-throughput pipeline.

There are two different approaches to the development of high-throughput sequencing technology. One is to increase the readable length/clone (long-read sequencing) and the other is to increase the number of clones sequenced simultaneously (parallel multiplex sequencing). For long-read sequencing, remarkable progress was made to achieve >1300-bp read length, using the original linear polyacrylamide, optimized electrophoretic condition, adjusted sequencing condition, and a refined base-caller (Zhou et al. 2000). For parallel multiplex sequencing using multicapillary array, followed by pioneering work of 24-capillary array with scanning mechanism (Mathies and Huang 1992), a multiple-sheathflow analyzer (Kambara and Takahashi 1993, Zhang et al. 1999) were developed. Recently a scanner with rotary detection device that can detect up to 1000 capillaries was presented (Scherer et al. 1999).

Phase I of the RIKEN mouse encyclopedia project required a system based on the latter approach. To this end, a 384-multicapillary auto sequencer (RISA sequencer) was developed to sequence >40,000 samples/day based on the parallel multiplex concept. To detect the 384 capillaries simultaneously, the optical system and electrophoresis system needed to be optimized based on the signal-to-noise ratio, mobility, and electrophoresis time. To increase the throughput, offline cross-linked gel preparation was employed and the capillary array was replaced immediately after the previous run was finished, thus eliminating the lag time for matrix replacement. Under these conditions, the RISA sequencer could handle 3840 samples/day/machine, but because of the importance of the long-read sequence for Phase II, a long-read version was needed to enable an average read length of >650 bp.

In this paper we describe the RISA system, including the RISA 384-multicapillary array sequencer, for the full-length cDNA project from mouse house to data.

RESULTS AND DISCUSSION

Basic Guiding Principles of Development of the RISA System

The RISA system was developed to accomplish the RIKEN mouse encyclopedia project Phase I within 3 yr. The basic guiding principles of development were (1) a 384-format operation, (2) a linear process, (3) size reduction and miniaturization, (4) multi-parallel processing, (5) compatibility with other systems, and (6) cost reduction of consumables and labor.

Features of the RISA System

The RISA system consists of four processes, such as colony picking, template preparation, reaction, and sequencing as described in Figure 1.

Following principle 1, a 384-format was employed for the sample plate, all enzymatic reaction systems, sample injection into capillary array of the sequencer, and sequence data analysis except for the template preparation process, which is directly compatible with the 384 format (see Methods). The minimum times of operation for sample parallel transfer without rearraying samples were employed for all processes. These developments contributed greatly to reducing ID error.
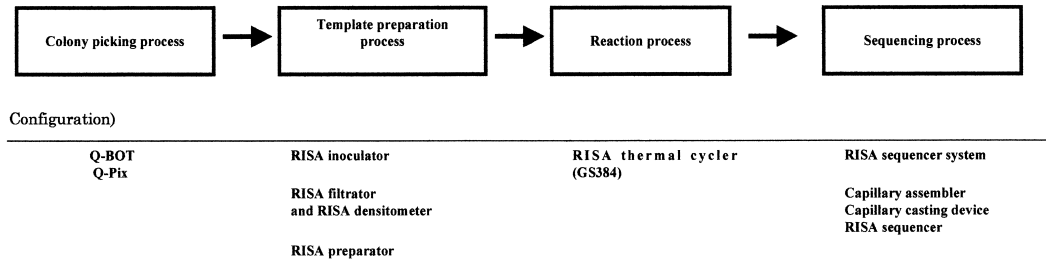
Figure 1 The RIKEN integrated sequence analysis (RISA) system. This system employs a 384-format linear process. Q-Bot and Q-pix are colony pickers that can inoculate *Escherichia coli* into the 384-format plate. A RISA inoculator, a RISA filtrator and densitometer, and a RISA plasmid preparator are used for plasmid DNA preparation. The RISA inoculator is a machine for dispensing medium in the 96-format deep-well plate and to inoculate *E. coli* from the 384-format plate to the 96-format deepwell. The RISA filtrator and densitometer harvest *E. coli* cells and measure the amount of cells. The RISA plasmid preparator prepares plasmid DNA by a filtration method. The RISA thermal cycler (GS384) is a four 384-well thermal cycler. The RISA sequencer system consists of the RISA capillary assembler (CAS), RISA casting device, and RISA sequencer that were originally developed. The RISA capillary assembler is used to make a 384-format capillary array unit. The RISA casting devise cleans capillaries and fills them with gelling solution. The RISA sequencer is a four-color 384-format capillary autosequencer.

The principle of linear processing with 384-format automatic operation does have a disadvantage. In a parallel transfer system, failures at each step, such as misspicking of *E. coli*, poor growth of *E. coli*, and low plasmid yield, accumulate to result in reduction of the final sequencing yield. If the DNA concentration of each well could be measured and if the wells with too low concentration of DNA could be eliminated, the sequencing output could be increased. Despite this disadvantage, we decided to employ the 384-format system throughout the process, from the picking up of colonies to data management, because the total production of data is increased enormously, without deteriorative sample ID error and not requiring the complicated plate and sample management.

Template Preparation Process

The RISA plasmid preparator, employing a 96-format plate, is a typical device designed from the concept of the linear processing (principle 2). This system was designed to avoid tube transfer of samples and use only a single filter column throughout the procedure. All reagents are injected from the upper side of column followed by suction downwards from the bottom side. In this system, samples are treated by a linear process like a belt conveyer system. This design concept was expanded to the entire procedure of the RISA system by employing 384-format sample transfer.

The direct sequencing process was designed so that the process was compatible with the plasmid preparation process. The RISA system was designed to cluster full-length cDNA. The length of the cDNA ranges from a few hundred bases to >10 kb. For those clones, template preparation with plasmid is effective because of the small size bias for the insert length. On the other hand, the shotgun strategy is employed for Phase II or BAC sequence. For short-insert clones of the shotgun library, direct sequencing is effective for template preparation because of shortening of the process time and reducing labor.

Reaction Process

Multi-parallel processing is also applied to the sequencing reaction. The RISA thermal cycler (GS384) was designed to have four 384-well thermal cycling or isothermal sites and to allow for independent operation of each site. Several sequencing reagents, DYEnamic ET terminator kit (Amersham), BigDyeTerminator kit (Perkin-Elmer), or transcriptional sequencing reagent (Sasaki et al. 1998a,b), originally developed by us, are available for reaction. This compatibility is convenient for experiments such as finishing.

Sequencing Process

To facilitate handling of the capillaries, the 384 capillaries should be assembled as one structure, also the samples on the 16 × 24 (384)-format plate should be directly injected and subjected to simultaneous detection. To realize this, the 384-capillary array was designed so that the 16 × 24 (384)-format alignment of capillaries on the injection side is the same alignment as that of the 384-format titer plate. This design allowed the direct sample injection to be operated using a 384-format plate on which the sequencing samples are prepared. The detection side of the capillary has a sheet shape for scanning detection. The detection window is made by burning out the outer coating material of the capillaries. As an injection device, a novel injection plate was fabricated to inject samples electrokinetically into capillary columns. This injection plate consists of a stainless steel plate with 384 pits as a cathode bottom plate and funnel-shaped insulators pressed into each pit. The sample can be directly transferred from the 384-format PCR plate to this 384-format injection plate using a commercially available dispenser. This type of injection plate showed the best result

among the various trials, including needle-like Pt wires, cathode electrodes, or stainless steel plate with 384 holes, which gave distorted band patterns. The bottom and the end of the capillaries must be separated for good band separation. Thus, direct injection from the 384-format injection plate enabled consistent sample flow from the picking up of colonies to the database based on the 384 format.

Peak Detection and Tracking on Focus Reference Plane

To avoid adjustment of the optical focus at the center of the capillary in each setting of the capillary array, we set the reference-ridged plate so that the capillaries on the plate were on the focal plane. When a capillary array is set on the machine, the capillaries are pushed against this plate. This mechanism works well, offering good reproducibility of the focusing. Before electrophoresis, a position of each capillary is determined by fitting data of sampling points around the peak to increase the dynamic range mentioned in the Methods section. During electrophoresis, the peak positions are tracked automatically.

Basic Design Concept of the Optical System of the 384-Capillary Sequencer

To select the best system to detect every signal from all capillaries of the 384-format capillary sheet, we examined the scanning system and the image system. The image system was examined under an assumption that all capillaries were excited uniformly at the same laser power as that of the scanning system. This is only an ideal condition for the image system and actually it is hard to excite uniformly a large number of capillaries. To evaluate these two systems, the data sampling rate and the conditions of electrophoresis, such as buffer, separation matrix, and electrophoresis voltage, should be fixed. The high quality of the sequencing data depends on the signal intensity and S/N ratio. The signal intensity and S/N ratio are discussed below for both the scanning and the image detection systems.

All conditions controlling the signal intensity are listed in Table 1. The signal intensity is reflected by the total detectable photon number from a single capillary during a single sampling interval (Dpn), which is given by integration of the detectable photon number/unit time (Dpnt) during the detection time. The detection time (Dt) is defined as the time during which the detector actually measures the fluorescent light during a single sampling interval (SI). The relationship of Dpn, Dpnt, and Dt is given by the formula Dpn = Dpnt × Dt.

The next step is to select a system that can maximize the Dpn to obtain higher signal intensity.

When the number of capillaries is one, we can employ a static optical system to detect one capillary. For this system, no scanning mechanism or focusing optical system is necessary. We define Dpnt in this system as Np. Hereafter the changes of Dpnt and Dt are discussed for the case of a greater number of capillaries (N). When the number of capillaries is >2, the scanning system or the image detection system is a candidate for the detection system.

In the scanning system, an objective lens can be set at a position close to a capillary to the mechanical limitation, under which the objective lens scans along the capillary array. The relative distance between the capillary of the multiple capillary sheet and the objective lens along the perpendicular axis to the capillary sheet is the same as the distance between the capillary and the objective lens in the above-mentioned single capillary system. Therefore, even if the capillary number increases to N, the maximum value of Dpnt does not change to Np when the objective lens is located in front of a capillary. Dpnt decreases as the distance of the objective lens increases from the front of a capillary. The decreasing ratio depends on the relative configuration between the objective lens and the capillary only. Thus, the effect of this decrease in the photon number is expressed as the product of Np by a constant (A), and we estimate Dpnt for N-capillary scanning system to be A × Np.

The Dt for a capillary in the N-capillary system is the time when fluorescent light from the capillary is

Table 1. Comparison of a Scanning System and An Image Detection System

	Detectable photon number/ unit time (Dpnt)	Detection time (Dt)	Total detectable photon number from a single capillary during a single sampling interval (Dpn)
Scanning system	$A \times Np$	$C \times SI/N$	$A \times C \times Np \times SI/N$
Image detection system	Np/N^2	$SI \times RT \times N$	$Np \times (SI/N^2 - RT/N)$
Ratio			$A \times C \times (1/N - RT/SI) \approx A \times C/N$

(A, C) Constants depending on the scanning optical and mechanical system; (Np) detectable photon number/ unit time for static single capillary detection system; (SI) sampling interval; (N) number of capillaries; (RT) reading time of total data of the CCD element.

detected during a single scan. The detector moves along the axis parallel to a capillary array. The sampling interval (SI) is the sum of the time when the scanning stage moves at a constant speed and the time when the scanning stage accelerates or decelerates. The latter is negligible in comparison with SI. The former is the sum of the time for detection (scanning time) and the time for return to the scanning starting point (returning time). The ratio between scanning time and returning time is a ratio between the velocity of the scanning stage at both scanning times. Thus, if we define the ratio (returning time/scanning time) as R, the time for detection is $SI/(1 + R)$, therefore the Dt is $SI/((1 + R) \times N)$. Here we define $1/(1 + R)$ as C, and Dt is expressed as $C \times SI/N$. Thus, we can present Dpn as $A \times C \times Np \times SI/N$ as shown in the top row of Table 1.

In an image detection system, the fluorescent light from the 384 capillaries should be focused on a focal plane at the same time. The detection system can have an objective lens of large diameter to collect as much light as possible from each capillary with a large view or a normal objective lens of ordinary diameter. The former system is difficult to design and set up with a wide objective lens for detecting 384 capillaries simultaneously. In the latter system, the distance between the focal point of objective lens and the capillary array to focus on N number of capillaries should be N-fold of the distance for single capillary detection. When N is large, the focal length is negligible in comparison with the distance between the objective lens and the capillary array. Thus, the distance is proportional to N and the amount of light is inversely proportional to the square of N. Therefore, in the case of the N-capillary system, $Dpnt = Np/N^2$.

A CCD element is a detection device of two-dimensional images. The CCD device is used as an image detection system of a multiple capillary sequencer. The figure of the capillaries should be focused on the CCD element to distinguish between two adjoining capillaries. As the number of capillaries increases, the number of pixels in the axis direction parallel to the capillary sheet should be increased in proportion to the number of capillaries. Thus, the time needed to read data on the CCD element is directly proportional to the number of capillaries. Therefore, the total reading time is the reading time (RT) for one pixel multiplied by the capillary number. In the case of the N-capillary system, Dt is $SI - RT \times N$ (see Table 1).

The Dpn of the scanning system is $A \times C \times Np \times SI/N$ and that of the image detection system is $Np \times (SI/N^2 - P \times RT/N)$. The ratio of Dpn of the scanning system and Dpn of the image detection system is $1/N - RT/SI$. Because the practically used SI value is 1 sec and that of RT is 1–0.1 msec, RT/SI is negligible. This gives the following relationship:

Dpn (image detection system) $\propto$ Dpn (scanning system)/N. Thus, when the total number of capillary increases, the scanning system offers a great advantage in terms of sensitivity (Dpn), when compared with an image detection system.

The noise of output of detector consists of three components: the noise generated by mechanical vibration, called "mechanical noise;" the noise originating from thermal electron fluctuation, named "thermal noise;" and the noise caused by statistical fluctuation originating from the physical process itself, or "quantum noise." The last two types are unavoidable in optical detection systems including scanning and image detection systems. However, mechanical noise arises only in scanning systems, which must employ mechanical movement to scan the capillary array. From the viewpoint of noise alone, an image detection system would be more advantageous than a scanning system.

However, our pilot study showed signal intensity to be a much more critical condition than noise. Even the signal intensity of simultaneous detection of 96 capillaries by an image detection system was almost at background level when we tried to detect the fluorescent light directly from the inside of the capillary. In the 384-capillary array, the fluorescent light is estimated to be 16-fold weaker than the light of 96-capillary array. Thus, signal intensity was more problematic than noise intensity in designing the 384-capillary image detection system. We therefore decided to employ a scanning system for the 384-capillary detection system. To do this, we had to overcome the noise problem that originated in PMT from mechanical vibration. To solve this problem, PMTs that are resistant to vibration noise were selected and used for the detection devices.

Performance of the RISA Sequencer System

Read Length and Reproducibility

For the evaluation test, samples were prepared as described in Methods using 250 ng/well of CsCl purified pGEM-3ZF(+) plasmid DNA. Injection and electrophoresis were done using a gel-filled capillary array as described in Methods. For the analyzed data (see Fig. 2), the horizontal axis represents the sampling number and the vertical axis gives the relative signal amplitude. Of 384 capillaries, 381 outputs (99.2%) were obtained and 380 outputs (99.0%) were base-called over 650 bp in 2.7 h. The mean read length exceeding 99% and 98% accuracy is 654.4 bp and 682.8 bp, respectively. A three-dimensional histogram of the read length at each accuracy from 98.0% to 99.5% is shown in Figure 3. A homology search was conducted using GENETIX-WIN software (Software Development Co., Ltd., Japan). An-
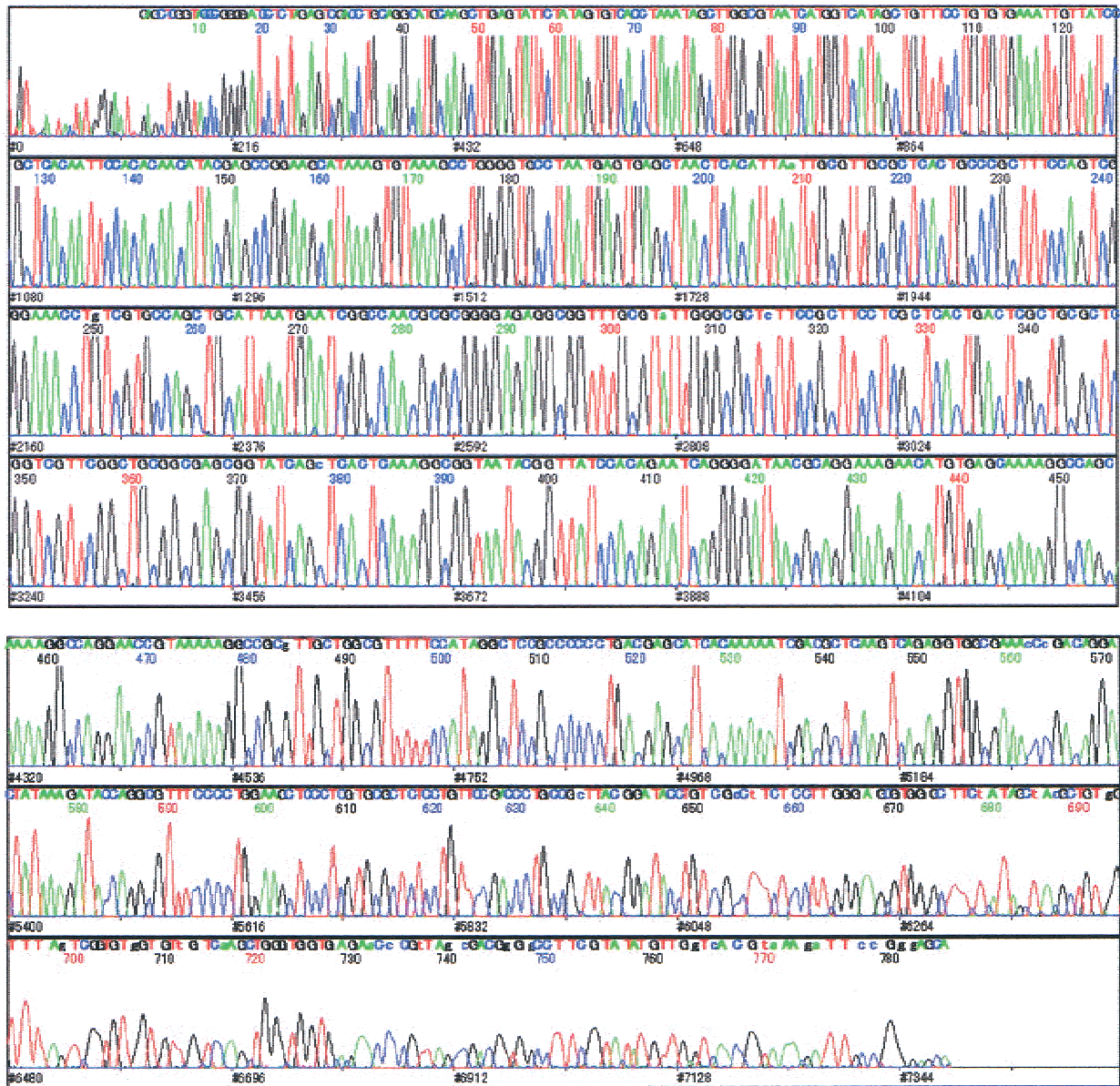
Figure 2 Analyzed and base-called electropherogram. The horizontal axis represents the scan number and the vertical axis gives the relative amplitude.

other experiment using the same samples was performed to confirm the reproducibility of the system. Of 384 capillaries, 379 outputs (98.7%) were obtained and 373 outputs had accuracy exceeding 99%. Of the 373 outputs, the mean read length exceeding 99% accuracy was 639.7 bp. These results indicate that the reproducibility is quite good. For short-read sequencing, electrophoresis for 1.5 h was achieved by changing the electrophoresis voltage to 158 V/cm and the read length of this condition was 350 bp, which is long enough for clustering full-length cDNA clones. Under these electrophoresis conditions 10 runs/day/machine (3,840 samples/day/machine) can be done.

Total Operating Cost for the RISA System

The greatest problem in mass sequencing is the operating cost. Because our capillary array is disposable, the electrophoresis cost may appear to be high. Calculation of the cost for consumables for the RISA sequencer gave the estimate of 20 cents (US$ 1 = 107 yen) for the analysis of one sequencing product, including not only the capillary itself but also all of the buffers, gels, basement plate rubbers, and other materials.

The total operating cost/sample from picking up the colonies to the output of raw data was calculated to be about 40 cents/clone excluding the cost of the sequencing kit. This figure includes deep-well plate, PCR
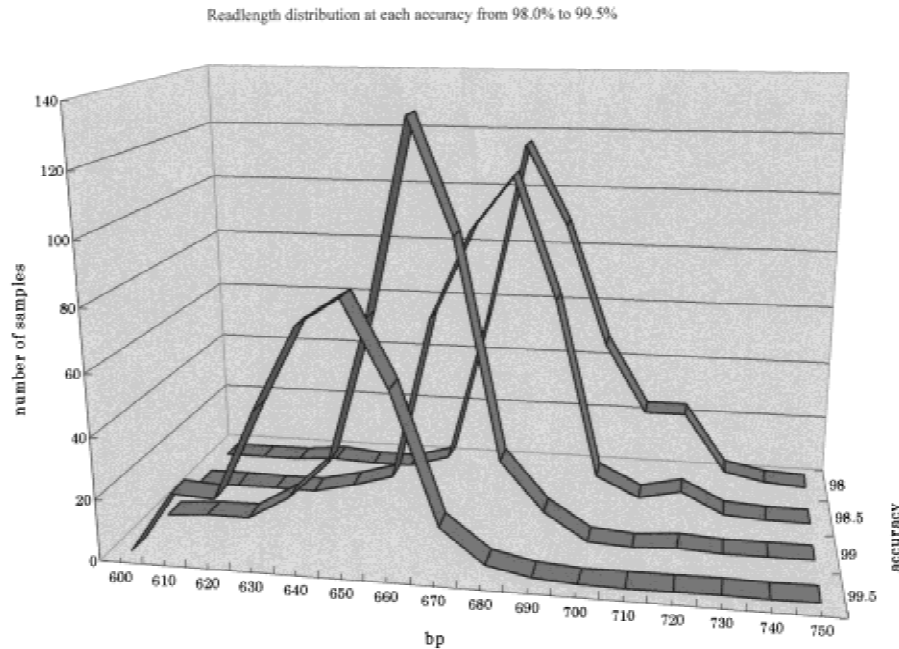
Figure 3 Three-dimensional histogram of read length exceeding 98.0, 98.5, 99.0, and 99.5% accuracy. The lines represent the histogram of each accuracy. Homology search was performed with GENETYX-WIN software (Software Development Co., Ltd, Japan).

plate, reagent, and other consumables. Thus, the total cost is enough low to put the mass sequencing into practice.

Performance of the RISA System

The total requirement of the RISA system for the RIKEN mouse encyclopedia project is the treatment of >40,000 samples/day for clustering full-length cDNA. To fulfill this requirement, the RISA system provides small number of machines, requiring a small amount of space. The RISA system consisting of two Q-pixs, a RISA inoculator, a RISA filtrator and densitometer, two RISA preparators, four GS384s, four CASs, six capillary casting devices, and 16 RISA sequencers handles 50,000 samples/day for a maximum output for clustering cDNA (at present because of a space limitation, the capacity is 30,000 samples/day with 20 RISA sequencers).

We conducted sequencing with the RISA system using pGEM-3ZF(+) colonies. Ten 384-format plates were prepared in the colony picking process and processed in the downstream sequencing line. The success rates for the colony picking, the plasmid preparation, and the reaction process are 99.95%, 98.09%, and 82.04%, respectively. The success rate in the colony picking process is the ratio of the number of wells showing good growth of *E. coli* to the total number of wells. The success rate in the plasmid preparation process is the ratio of the number of wells showing good growth in the cultivation step to the number of suc-

cessful wells in the colony picking process. The success rate in the reaction and electrophoresis process is the ratio of the number of base-called wells to the number of wells showing successful previous process. The final success rate was 80.47%. The major cause of the difference between this result and the performance of RISA sequencer itself (99.2%) is the distribution of the amount of templates well by well.

The error rate of the pilot sequence at each read length is shown in Figure 4. Homology search was performed using FASTA ver. 3.2t09. The numbers of mismatches, insertions, and deletions were counted base by base. The error rate at a certain base position was calculated by dividing the sum of the number of mismatches, insertions, and deletions at each position located closer to the primer than the base position by the total base-called base number from the primer to the base position. The error rate increases rapidly >620 bp at which the error rate is 2%. The discrepancy between the value and 682-bp read length using CsCl purified plasmid DNA is caused by the quality of the plasmid DNA.

From these data, the RISA system in the long-read version can be used to achieve 18.4 Mb/day as the total output. The sequencing capacity of the RISA system is so enormous that seven RISA systems can produce one-fold of the human genome shotgun sequence within one month.

In conclusion, we have developed a 384-format sequence pipeline, the RISA system, which can produce >40,000 samples/day for clustering cDNA. In our mouse full-length cDNA project, we have tagged (sequenced from 3′ end) 929,827 mouse full-length cDNA clones with the RISA system. Those tags (3′ end sequences) are clustered at 128,597 species of full-length cDNA (Carninci et al., in prep.). The advantage of employing the 384-format throughout the entire process is the reduction of ID error. In actual operation, the ID error of the samples is 3.7%. A subset of the identified and highly normalized clones is printed on DNA chip for large-scale expression analysis (Miki et al., in prep.). Thus, the RISA system has an enormously powerful capacity for achieving large-scale full-length cDNA analysis. Also, the RISA system can produce sequences at 18.4 Mb/day, and one haploid genome shotgun se-
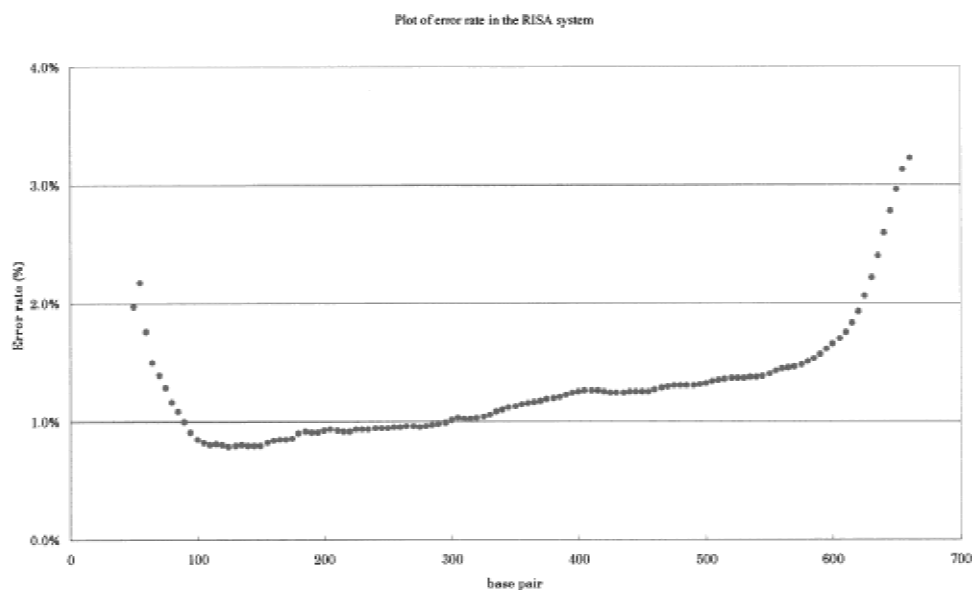
Figure 4 Error rate distribution of sequences using the RISA system. The template was pGEM–3ZF(+) plasmid DNA. A homology search was performed using FASTA ver. 3.2t09. The error rate at each position of the sequenced DNA is plotted. The vertical axis represents the percentage of misreading of the DNA fragment from primer to the position.

quence of a higher organism, such as human, mouse, rat, domestic animals, and plants, can be produced by seven RISA systems within one month.

METHODS

Library Construction

The cDNA libraries were prepared by the cap-trapper method (Carninci et al. 1996, 1997, 1998; Carninci and Hayashizaki 1999) and ligated into λFLCI (Carninci et al., in prep.).

Colony Picking and Storage

The appropriate aliquots of library solution in storage tubes were spread on LB plate containing 400 µg/mL of ampicillin in 22-cm square dishes by SuperGyro (Wakenyaku, Japan) with glass beads. The plates were incubated at 30 °C for 18 h. The colonies that appeared were picked by Q-bot or Q-pix (Genetix) and inoculated into 384-well plates whose wells were filled with 40 µL of LB medium containing 7% glycerol and 100 µg/mL of ampicillin. The inoculated plates, designated as original plates, were incubated at 30°C for 18 h in closed boxes with wet paper towels. After cultivation, the plates were replicated with 384 pins and incubated as above, designated as replica plates. To check for contamination of the plates, the clones in both plates were inoculated on LB plates with ampicillin using 384 pins, incubated at 30°C overnight, and visually checked. Both plates were stored at −80°C. When cultivating the clones for the following plasmid preparation or cell PCR, replica plates were used.

384 Format-Based Sample Handling

All procedures of the RISA system were designed based on the 384 format. As mentioned in Results and Discussion, only the plasmid preparation step is carried out a 96-format plate. The plasmid preparation step requires a sufficient amount of *E. coli* cells for preparing plasmid DNA for sequencing. The 384-

well plate has too small an area and too small a capacity to harvest a sufficient number of cells and to filtrate them without stacking of the cells. For these reasons, the 96-format was employed only for the step from cultivation of *E. coli* to recover the purified plasmid DNA. However, this step was also designed to be compatible with the 384-format with connection from upstream 384-format plate in which *E. coli* cells are stocked and to connect to downstream 384-format PCR plate for cycle sequencing or transcriptional sequencing without changing the relative positional relation of samples.

In our system, the row and column of the 384-format plate are identified by alphabets for the rows and by numbers in double figures for columns, such as A-01, A-02, and so on. Samples in 384-format wells can be transferred to 96-format plates using a 96-pin array in four steps. As the first step, a 96-subset in the 384-format is transferred to the first 96-format plate by a 96-pin array so that the upper left pin of the 96-pin array picks up the A-01 well of the 384-format plate. We call this 96-subset the A-subset. The other three 96-subsets of the 384-format plate are transferred by the same procedure in the second, third, and fourth rounds so that A-02, B-01, and B-02 of the 384-format plate are picked up by the upper left pin of the 96-pin array to make the second, third, and fourth 96-format plates. We call these B-subset, C-subset, and D-subset, respectively.

RISA Inoculator

Figure 5(A,B) shows the photograph of the equipment as a whole and its schematic plan, respectively. This machine can dispense various volumes of the medium in 96-format deep-well plates and inoculate clones from a 384-well plate with 96-pin arrays in a fully automatic manner.

This machine consists of two lines named DL and SL (Fig. 5B). DL conveys the trays carrying the eight 96-format deep-well plates, and SL conveys the trays with eight 384-format replica plates with *E. coli*. As the first step, the tray carrying
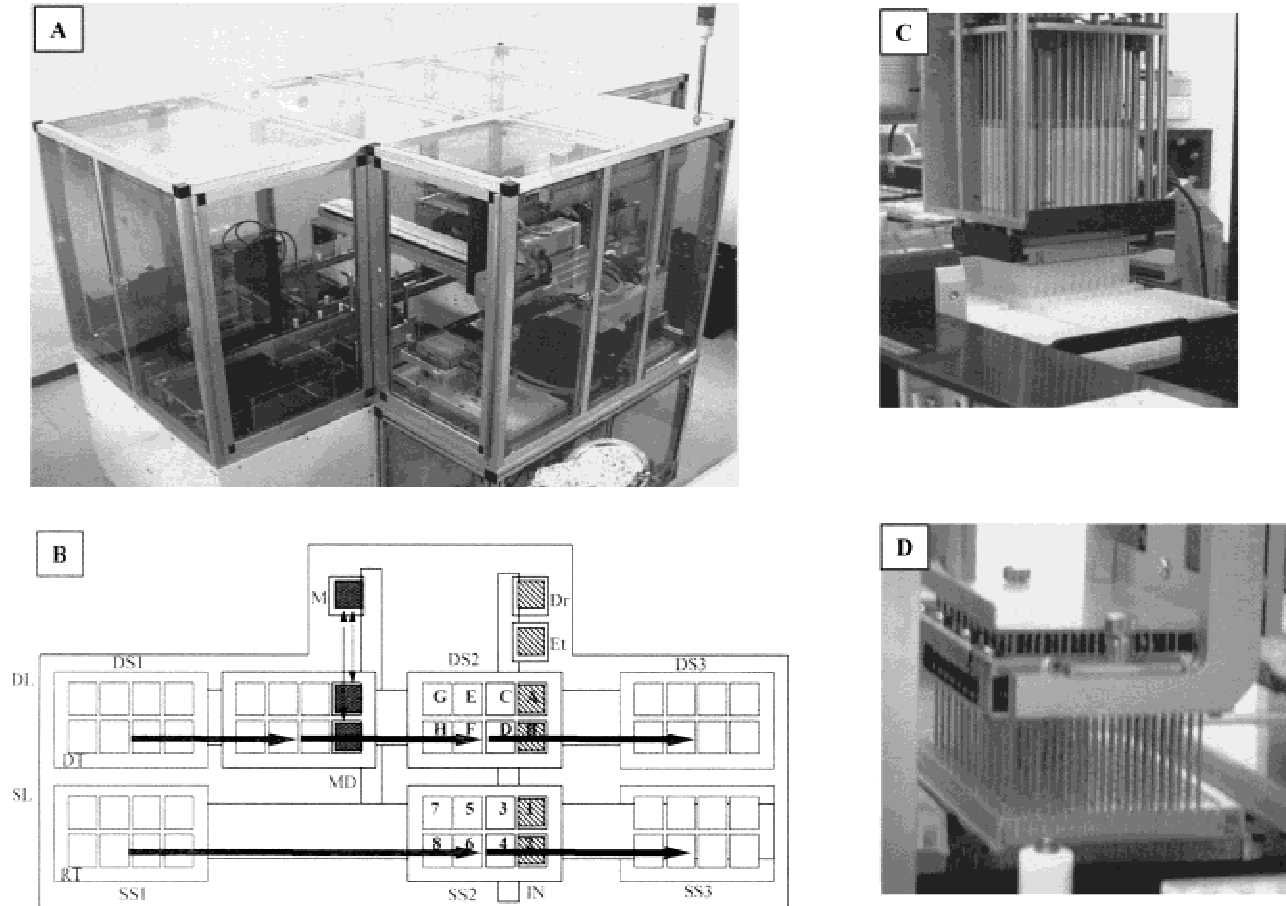
Figure 5 The RISA inoculator. (*A*) Photograph of RISA inoculator. (*B*) Schematic representation of RISA inoculator. (*C*) Movable 96-channel auto-dispenser at medium-dispensing site. (*D*) Movable 96 pins for clone inoculation at inoculation site. In *B*, 384-well plates and deep-well plates are supplied on 384-well plate trays (RT) and 96-format deep-well plate trays (DT) at the respective stackers [(SS1) RT stacker; (DS1) DT stacker], and conveyed on the respective lines [(SL) RT line; (DL) DT line]. A total of 8 DTs and 10 RTs can be piled up on SS1 and DS1, respectively. Medium is dispensed at the medium-dispensing site (MD) by a movable 96-channel autodispenser (C). The autodispenser transfers an appropriate volume of medium at the medium bath (M) to the deep-well plate. For the inoculation, the movable 96-pin array (D) is first washed in ethanol bath (Et), and dried by an air dryer (Dr). The 96-pin array then moves to SL of the inoculation site (IN), picks clones from the 384-well plate, and inoculates in deep-well plate at DL of IN. In subset manipulation, the A-subset clones (for the definition of A, B, C, and D subsets, see Methods) of the 384-well plate at position 1 on the tray at the SS2 site are inoculated in the deep-well plate at position A on the tray at DS2. In the same way, B, C, and D subsets of the same 384-well plate are inoculated in the deep-well plates at positions B, C, and D, respectively. The subsets of the next well plate are at position 2; these subsets are inoculated into the plates at positions E, F, G, and H. In this way, all well plates are manipulated sequentially, being transported by belt conveyer. The manipulated trays are then piled up at DS3 and SS3. At positions IN of SL and DL, barcode readers detect the 384-well plate and deep-well plate IDs.

the empty 96-format deep-well plates and those carrying the eight 384-format replica plates with *E. coli* are placed at sites named DS1 and SS1, respectively. As the next step, the tray with empty 96-format deep-well plates are transferred onto the DL, and at the site named MD, the deep-well plates are filled with medium carried from site M using the dispenser shown in Figure 5C. Subsequently, the trays with the deep-well plate and those with the 384-replica plate are transferred to DS2 and SS2, respectively. At this site, the 96-pin array shown in Figure 5D inoculates *E. coli* from the 384-format replica plate on line SL to the 96-format deep wells on line DL. The exchanging format from 384 to 96 is described above; four 96-format plates are inoculated from one 384-format replica plate. Finally, the trays with inoculated deep-well plates are transferred to the site named DS3. A total of 8 and 10 trays

are stacked at the DS3 and SS3 sites, respectively. Medium dispensing and inoculation of one deep-well plate can be done in 45 sec.

RISA Filtrator and RISA Densitometer

Figure 6(A,B) shows the photograph of these instruments and their schematic plan, respectively. The RISA filtrator transfers the cultivated broth in the 96-format plate to the filter plate in which *E. coli* cells are trapped on the glass filter for the subsequent plasmid preparation step. This machine consists of two lines [cultivated deepwell line (CDL) and filter plate line (FL)] and transferring unit (TSU). A plate stacker (ST) is also shown in the photograph. This machine can transfer cultivated *E. coli* from the 96-format deep-well plate to the 96-
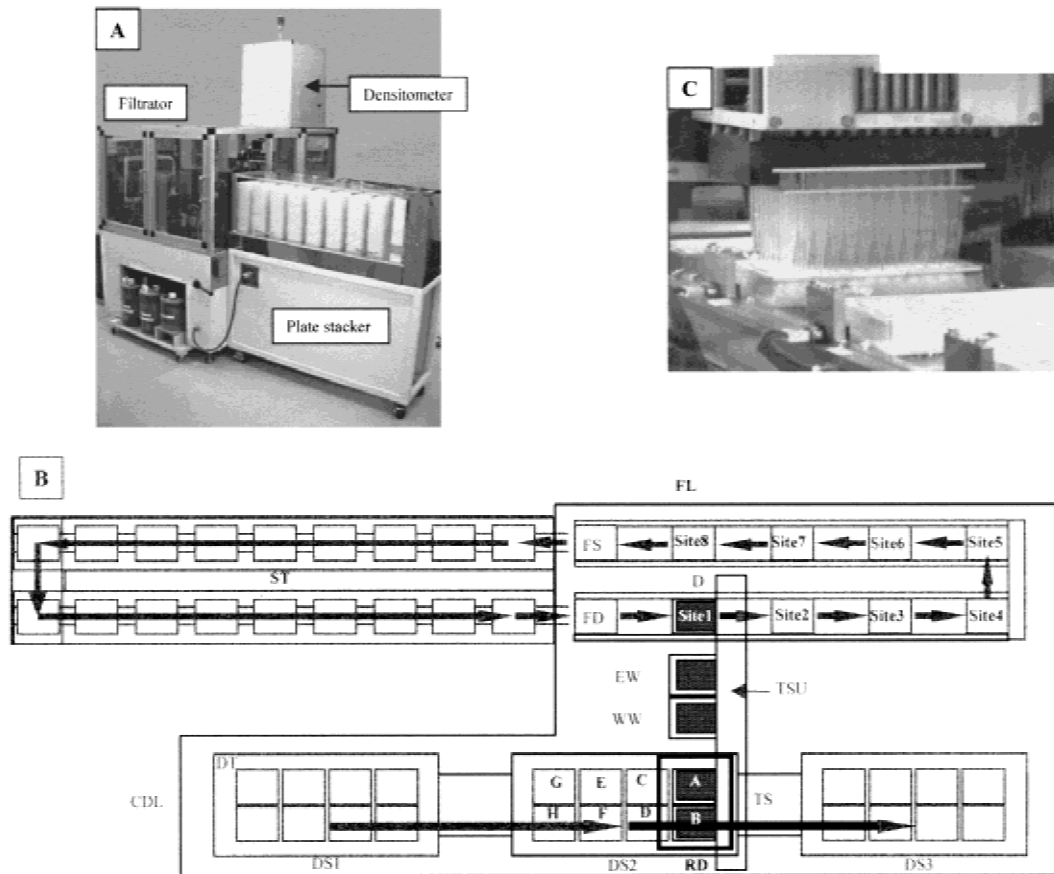
Figure 6 The RISA filtrator and densitometer. (*A*) Photograph of RISA filtrator and densitometer. (*B*) Schematic representation of RISA filtrator and densitometer. (*C*) Vacuum filtration sites and conveyer of RISA filtrator. In *B*, deep-well plates on DT are supplied at position DS1 on CDL. Eight DTs can be piled up on DS1. Filter plates for plasmid preparation are piled up in plate stacker (ST) and supplied to the plate separation apparatus (FD). Each filter plate is carried to the dispensing site (D). DT is conveyed to the culture transfer site (DS2), and all deep-well culture plates are scanned by RISA densitometer at the densitometery site (RD). After scanning, the next step is to obtain the appropriate volume of culture liquid from the deep-well plate at position A of DS2 on CDL by suction using the movable 96-channel autodispenser. The dispenser transfers samples to the filter plate at site 1 on FL. The filter plate is subjected to suction under vacuum at stage 1, then it conveyed to the next stage in a 30-sec interval. After transferring of the samples, the autodispenser is washed in an ethanol bath (EW) and water bath (WW). Twenty filter plates are piled up at plate-stacking apparatus (FS), and a set of 12 of the filter plates is conveyed to ST step by step. At the RD, site 1, and site 8, barcode readers detect the deep-well plate and filter plate IDs.

format RIKEN filter plate or Millipore filter plate and harvests *E. coli* cells. Each well of the 96-format RIKEN filter plate has a glass filter on membrane filter. The funnel is designed under the membrane filter at the bottom, through which reagents can be removed by suction (Itoh et al. 1999). In the first step, empty filter plates are stacked in ST connected with the RISA filtrator as shown in Figure 6A. Empty filter plates are carried individually in the filtrator. On the other hand, trays on which 96-format deep-well plates with cultivated *E. coli* are set at the DS1 site on CDL. Each tray is transported to the DS2 site. Each deep well is scanned by the cell density detection system using CCD cameras located at RD site, named the RISA densitometer, as shown in Figure 6A. The cell density data is input as pixel data and calculated as cell density for each well.

The cultivated *E. coli* is transferred from the 96-format deepwell plate to 96-format filter plate at site 1 on FL with 96-multidispenser (Fig. 6C), the cultivated culture is removed by suction from the bottom funnel of the filter plate in 30 sec,

and *E. coli* cells become trapped on the filter plate. The filter plate is transported from one site to the next position, finally to be recovered by ST. After all filter plates return to ST, ST is removed from the RISA filtrator and is connected to the RISA plasmid preparator for the subsequent plasmid preparation step.

RISA Plasmid Preparator

Plasmid DNA for the reaction template is prepared by the filtration method (Itoh et al. 1997). This method, based on the principle of silica surface absorption of DNA in the presence of guanidine–HCl, is mechanized using an RISA inoculator, RISA filtrator, RISA densitometer, and RISA plasmid preparator. This system is fully automatic and the RISA plasmid preparator can produce 40,000 samples in 17 h as reported elsewhere (Itoh et al. 1999). The component machines can prepare plasmid DNA with either the originally developed

RIKEN filter plate or Millipore filter plate. To recover the plasmid DNA, 100 µL of 1 mM Tris-HCl solution is added to each column of the filter plate and the filter plate is centrifuged to elute the plasmid DNA samples onto the 96-format plate. The eluted samples are dried up at 65°C overnight.

Reaction

In the reaction process, dried plasmid DNA samples are dissolved in 20 µL of distilled water and the solutions of 7 µL are transferred in a 384-PCR plate with a 96-format dispenser (EDR384, Biotech, Japan).

For the sequencing reaction, cycle sequencing or transcriptional sequencing is employed. Bandpass filters of the RISA sequencer are selected based on the type of fluorescent dye. The cycle sequencing reaction was conducted using a DYEnamic ET terminator kit (Amersham Pharmacia Biotech) with 2.0 pmole -21M13 oligo (5'-TGTAAAACGACGGCCAGT-3') or 1233REV oligo (5'-AGCGGATAACAATTTCACA-CAGGA-3') as a sequencing primer. The enzymatic dideoxy sequencing reaction was conducted following the manufacturer's protocol. The total reaction volume was 7 µL. The temperature-cycling protocol for sequencing chemistry was performed on a RISA thermal cycler (GS384). The thermal cycling set of 95°C for 20 sec, 50°C for 15 sec, and 60°C for 60 sec was performed 25 times. Reaction products were purified with a high-throughput ethanol precipitation procedure, as described elsewhere (Aizawa et al., in prep.). The ABI Prism Big-Dye teminator kit (Perkin-Elmer) can be also used following a similar protocol to that described above.

An alternative is the transcriptional sequencing (TS) reaction that we developed. Direct sequencing reactions are carried out in 20 µL containing 10 ng of unpurified PCR products, 40 mM Tris-HCl (pH 8.0), 8 mM $MgCl_2$, 5 mM DTT, 2 mM spermidine-$(HCl)_3$, 2 mM 1,8 diamino-octane, 25 mM NaCl, 0.4 mM $MnCl_2$, and a transcriptional sequencing kit (TS kit) based on mutant T7 RNA Polymerase (Izawa et al. 1998; Sasaki et al. 1998a,b; the TS kit is being prepared for marketing). These reaction mixtures were incubated at 37°C for 60 min. Next, the unincorporated dye terminators were separated from the sequencing products by ethanol precipitation, dried, and suspended with 10 µL of loading buffer [60% formamide, 4 mM Tris-HCl (pH 8.0), 0.4 mM EDTA].

Capillary Array Assembler

The photograph of a capillary array assembler is shown in Figure 7A. To assemble the 384-capillary array, 16-capillary drums and a basement plate should be prepared. The basement plate consists of two metal bottom plates with 16×24 (384) holes and silicon rubber that is sandwiched between these two metal bottom plates (Fig. 7B). Sixteen capillaries (SGE, Australia) of 100 µm inner diameter, 300 µm outer diameter with polyacrylic derivative coating resin unrolled
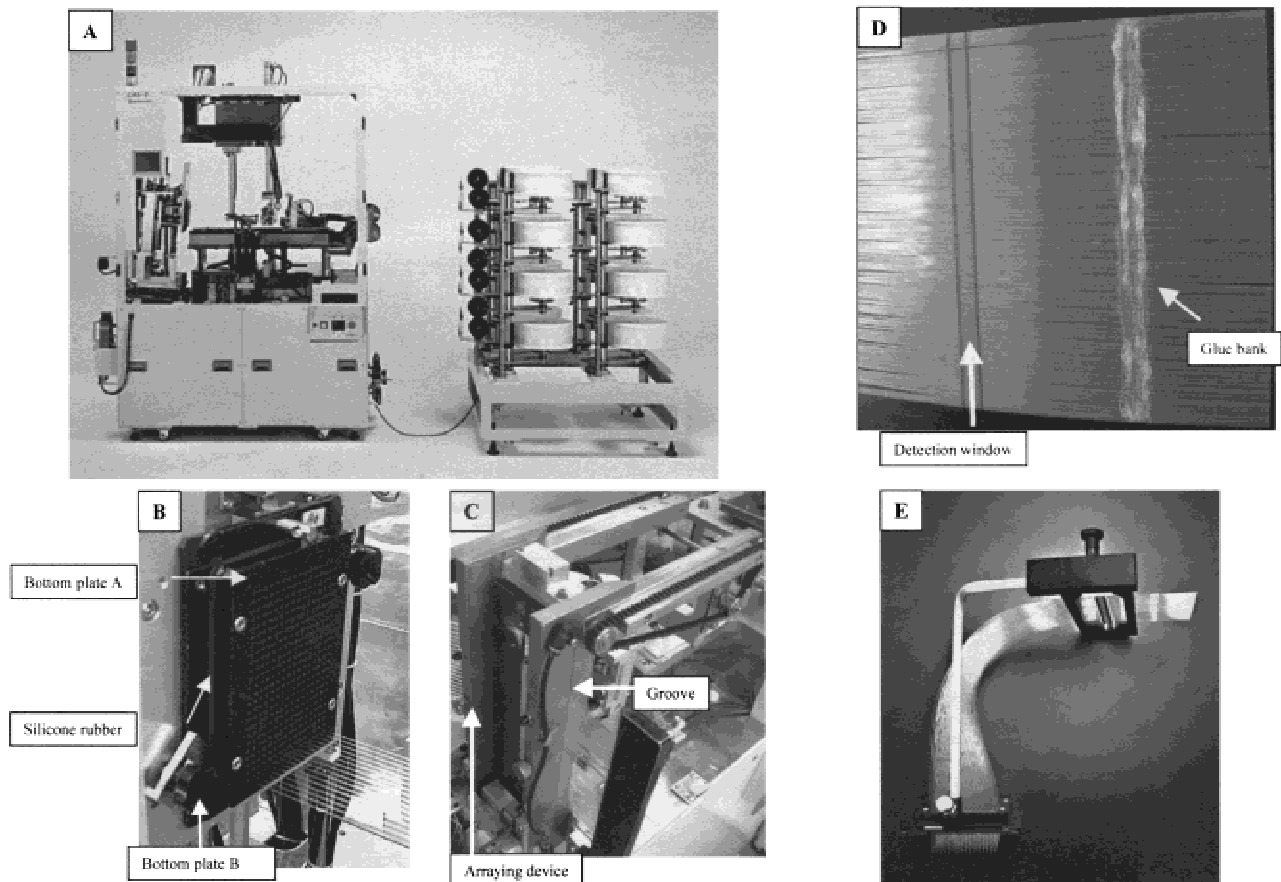

Figure 7 Photographs of the RISA capillary array assembler (CAS). (*A*) The RISA capillary array assembler. (*B*) Sixteen capillaries penetrating the silicone plate. (*C*) The arraying device. (*D*) Capillary sheet. (*E*) The 384-capillary cassette.

from capillary bobbins penetrate the basement plate by the guidance of the 16-needle array (Fig. 7B). A set of 16 capillaries is trimmed at 48 cm, transported to the groove of the arraying device (Fig. 7C), and piled up onto the previous set. As the next step, the position of the basement plate is changed to prepare the next operation described above.

Thus, the 16 × 24 (384)-format array at the injection side is aligned as a single 384-capillary sheet at the detection site (Fig. 7D) resulting in the final 384-format RISA capillary array (Fig. 7E) in the same format as the 384-well titer plate. This is a very convenient structure for injecting the 384 DNA samples on a 16 × 24 (384)-format plate to the capillary array in a single step. By burning out the outer coating with a heated Nichrome wire, a detection window is made 10 cm from the end of the capillary, which is then fixed by resolidifying coating material (Fig. 7D). The burning process is done twice to eliminate laser scattering caused by ashes from the coating materials produced by the first burning operation. A glue bank (Fig. 7D) is made on the capillary sheet of the injection side to dam up the anode buffer that streams up on the outer wall of the capillaries. The detection window is flushed with clean air and covered with a plastic cover to avoid dust contamination.

RISA Casting Device

The RISA casting device shown in Figure 8 was developed to clean capillaries and to fill each one with the gel solution. Each machine has four airtight chambers so that the basement plate of the capillary arrays functions as a lid. All of the washing reagents and gels in the reservoir are automatically injected into capillaries or squeezed out from capillaries.

Capillary Cleanup and Gel Casting

Capillary cleanup and gel casting are carried out by the RISA capillary casting device. The bare inner surface of the capillaries was washed sequentially with 0.1 M $NaHCO_3$ solution for 10 min, Milli-Q SP water for 3 min, 1 M HCl for 3 min, and again with Milli-Q SP water for 3 min, then dried with a N_2 gas stream for several minutes. This novel washing method
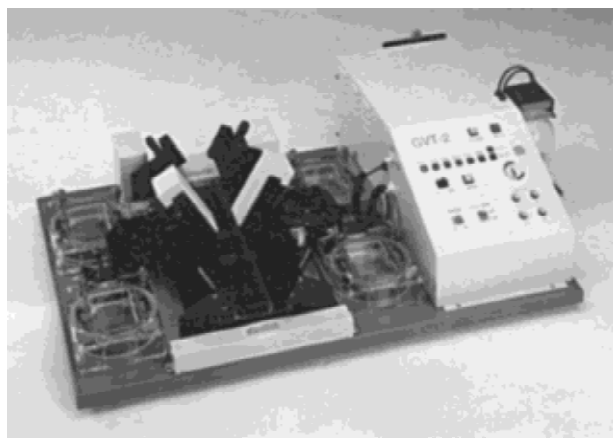

Figure 8 Photograph of the RISA capillary casting device. It has four chambers that are pressurized or depressurized to suck out or suck in the alkaline solution, acid solution, or water. Reservoirs for washing solutions are set on a movable stage that is controlled so that the washing step proceeds automatically.

greatly reduces the bubble formation in capillaries. This washing method is described in detail elsewhere (K. Aizawa, K. Shibata, M. Muramatsu, and Y. Hayashiza, in prep.).

The acrylamide solutions were prepared using 10.0 mL of PagePlus (AMRESCO), 10.0 mL of ×10 TBE (890 mM Tris, 890 mM boric acid, 25 mM EDTA–Na$_2$), Milli-Q SP water, and 42.0 g of urea (Bio-Rad). After adding Milli-Q SP water so that the total volume was 100.0 mL, the solutions were carefully degassed at 0°C for 45 min. Small particles were eliminated by passing the solution through a 0.2-μm membrane filter (Millipore). For polymerization, followed by adding 0.055 ml of N, N, N′, N′-tetramethylethylenediamine (Bio-Rad), 0.5 mL of a 10% solution of ammonium persulfate (Bio-Rad) was added. The resulting solutions were injected into the capillaries using a RISA casting device. Gelling PagePlus matrix in capillaries were left for 3–8 h at room temperature before electrophoresis.

RISA Sequencer

RISA conducts 384-format four-color auto-sequencing. Figure 9A is a photograph of the RISA 384-multicapillary autosequencer. The capillary array is set in a temperature-controlled chamber using a PID-controlled heater (Fig. 9B). The 3-cm capillary inlet is outside the chamber and dipped in a cathode buffer during electrophoresis. The anode of the capillary array is dipped in an above buffer reservoir. The cathode buffer reservoir is located on a computer-controlled movable stage. A special injection plate (Fig. 10A) is used for simultaneously injecting 384 samples electrokinetically into each capillary (Fig. 9C). A high-voltage power supply (HAR-15P10-SS200, Matsusada Precision, Japan) is used for injection and electrophoresis.

A schematic view of the optical system is presented in Figure 11. This optical system can excite the fluorescence-labeled chain termination reaction product, collect fluorescence, and detect spectral light with the band pass filter and photomultiplier. An Ar laser (LH, Model 183-D0230, Spectra-Physics Lasers) is used as an excitation light source. The 65-mW laser beam is guided to a scanning objective lens through a tunnel mirror and is focused on the capillary array. The fluorescent light from the fluorescent-labeled sample is collected with the objective lens and passed through the cut-off rejection filter to eliminate laser-scattering light. Collected fluorescent light is focused and passed through an alias to cut off defocusing rays and to increase the spatial resolution of the optical system. Through a set of four convex lenses, a ray is divided into four rays. Each ray passes through a different band-pass filter (OMEGA Optical), and is detected with Photomultiplier Tube (PMT, R1635–09, Hamamatsu Photonics, Japan).

PMT current is converted to voltage, which is then converted to numerals using a 12-bit analog/digital converter at 100 kHz. In a single scan, the capillary sheet is detected as 384 peaks, and each peak corresponds to respective capillaries. The component of detected light is the scattering laser light by the capillary itself and gel matrix. During electrophoresis, the detected signal is so strong that the amplitude is exceed a dynamic range of 12 bits, which is the range of data acquisition front-end circuit. Therefore, the detected peak height is saturated and the correct amplitude cannot be measured. To estimate the peak height at the saturated peak, the peak height is determined by estimation using polynomial fitting with data around the peak position. Thus, the dynamic range
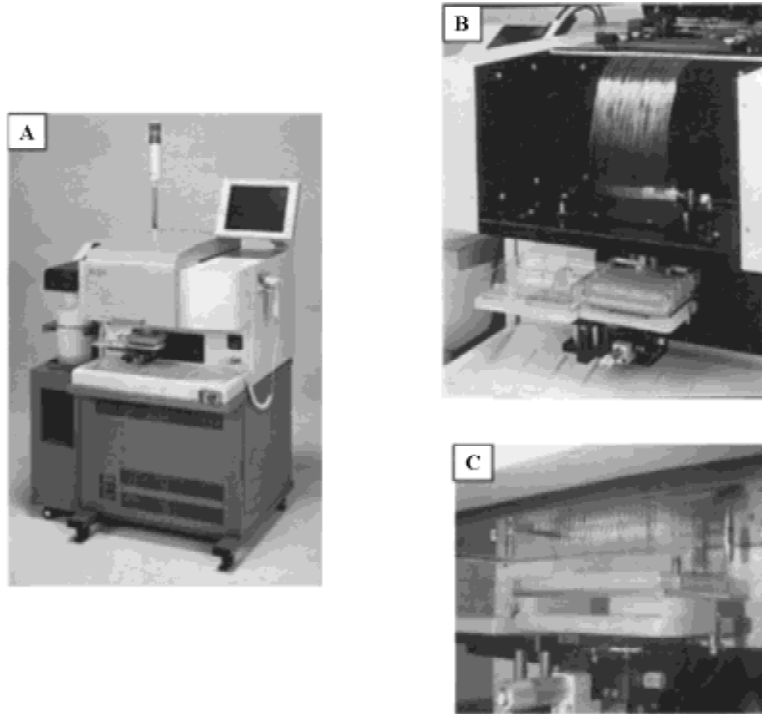
Figure 9 Photographs of the RISA sequencer. (*A*) Overall view. (*B*) A capillary array in the temperature-controlled chamber. In *A* the LCD screen displays the machine status or electropherogram and enables control of the machine from the touch screen with a pen. The colored lamp atop the pole shows the current status of the machine. The unit with the white bottle is a drainage device for used buffer. *B* shows the temperature-controlled chamber in which the 384-capillary array is located under the lid with a handle. *C* shows 384 capillaries inserted into the injection plate at the sample injection operation.

data storage, and the other is for a graphic user interface (GUI). The VxWorks real-time operating system is used for the machine control board and a Windows NT operating system for the GUI board. The boards have independent IP addresses and communicate via TCP/IP protocol. When a control button is pressed on the screen, the GUI board sends a message to the machine control board to operate the hardware. These boards communicate with other computers. When a machine is booted, the control board pulls up the machine control program from a server. When the data collection is completed, raw data is sent from the control board to an analysis server. Because of the expandability of network, a single RISA system can be easily expanded to a multi-RISA system.

To simplify RISA sequencer operation, the RISA operation protocol can be displayed on the computer screen for easy learning. A set of parameters including injection voltage, injection time, the voltage for electrophoresis, data-collection time, and temperature is stored in a server and downloaded to each machine. Up to five sets of these parameters can be input, which are one default set and four optional sets. A barcode reader on the RISA sequencer reads the Run ID displayed on the barcode on the injection plate. When the data collection is completed, these data are transferred to an analysis server and basecalled.

A homemade basecaller has been developed (unpubl.). Raw data is smoothed using a wavelet function. The baseline is fitted with polynomial approximation and subtracted from the raw data. After transformation using a predefined matrix value, each band is shifted to a proper position so the bands do not overlap. The signal's high-frequency component is removed using a fast Fourier transformation (FFT) and a filter, and band patterns are decombolved. The base is defined as the signal having maximum amplitude.
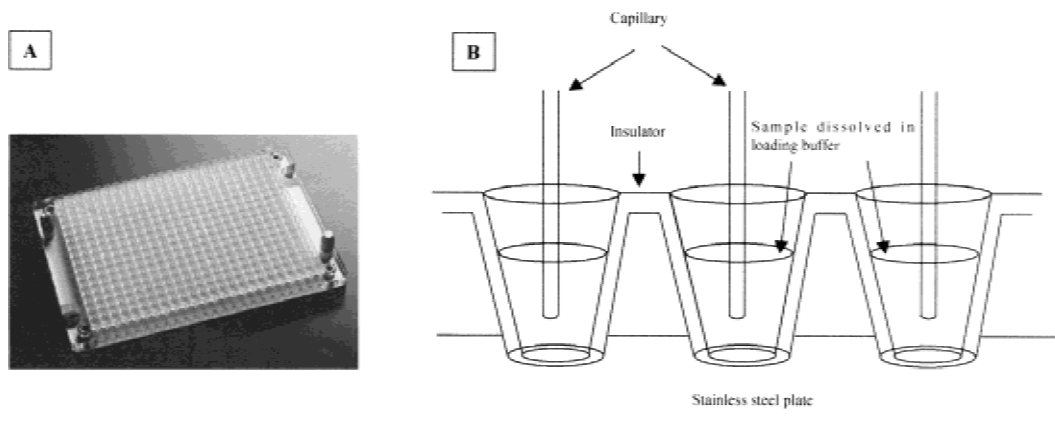
is expanded from 12 bits to 14 bits. Each capillary is tracked by tracking program during electrophoresis to correct the position shift caused by capillary heat expansion.

The RISA sequencer uses two computer boards. One is for machine control, data acquisition, preliminary analysis, and

Figure 10 An injection plate (*A*) and its structure (*B*). This plate has 24 × 16 (384) wells for sample injection and two poles for adjusting the position to the capillary array. The bottom is a stainless steel plate and the wall is an insulator. The bottom plate is connected to a ground line at injection. Sequence products are dissolved in 14 μL of loading buffer.
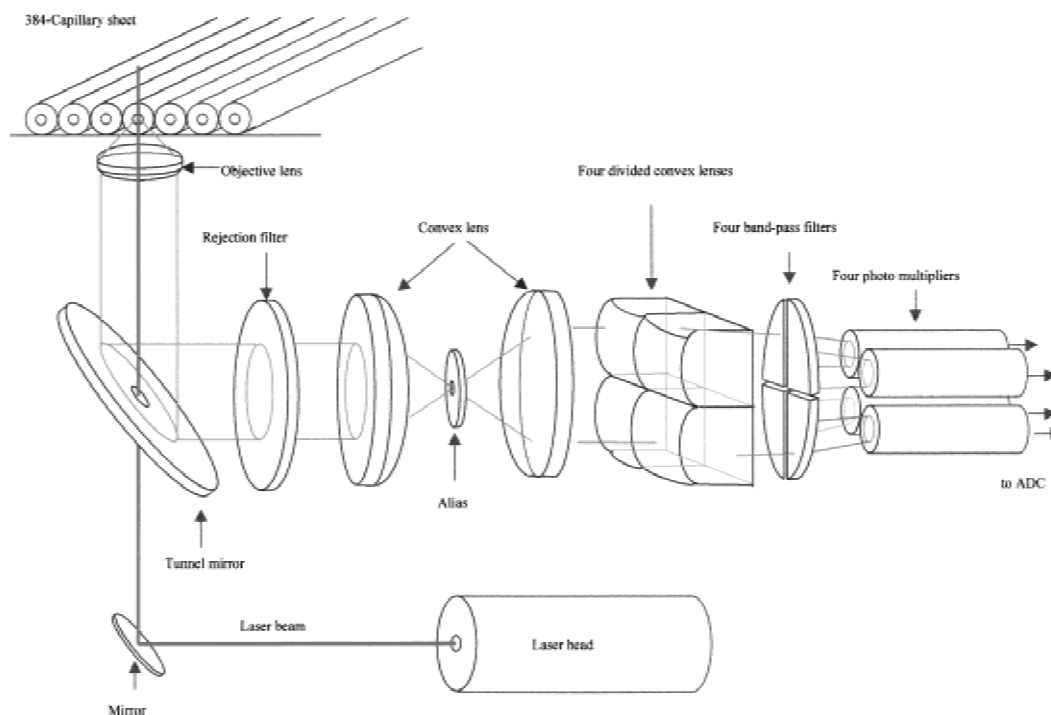
Figure 11 Schematic view of the optical system. A laser beam emitted from the laser head is used as an excitation light source. The laser beam is guided to a scanning objective lens through a tunnel mirror and is focused on capillary array. The fluorescent light from the dye combined with a reaction product is collected with the objective lens and is passed through a cut-off rejection filter to eliminate laser scattering light. The collected fluorescent light is focused and passed through an alias to cut off a defocusing rays. Four convex lens divide the ray four. Each ray passes through a different band-pass filter and is detected with a photo multiplier. These optical and detection devices without a laser head are scanned during data taking time.

Injection and Electrophoresis

Samples are electrokinetically loaded using a field of 73 V/cm for 60 sec at room temperature. Electrophoresis conditions for separating cycle sequencing products are as follows: electric field = 100 V/cm, effective capillary length = 38 cm (distance between injection inlet and detection zone), total length = 48 cm, and data sampling rate = 0.752 Hz. The 30-cm portion of the effective capillary length is heated at 50°C.

ACKNOWLEDGMENTS

We thank Saiko Takaku, Yuji Sogabe, Hiromi Takano, Toshio Hiraoka, and Junko Ishikawa for their assistance in preparing the data for this paper and Life Tech, Ltd. for their technical assistance. We thank all members of the Genome Exploration Research Group of the Genomic Sciences Center in RIKEN and the Genome Science Laboratory, RIKEN Tsukuba Institute. This study has been supported by Special Coordination Funds and a Research Grant for the RIKEN Genome Exploration Research Project, and Research and Development for Applying Advanced Computational Science and Technology (ACT–JST) of Japan Science and Technology Corporation (JST). The development of the RISA sequencer was especially supported by CREST (Core Research for Evolutional Science and Technology). All of these research grants are the fund from Science Technology Agency in Japanese Government to Y.H. This work was also supported by a Grant-in-Aid for Scientific Research on Priority Areas and Human Genome Program, from the Ministry of Education, Science and Culture, and by a Grant-in-Aid for a Second Term Comprehensive 10-Year Strategy for Cancer Control from the Ministry of Health and Welfare to Y.H.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Carninci, P. and Hayashizaki, Y. 1999. High-efficiency full-length cDNA cloning. *Methods Enzymol.* **303:** 19–44.
Carninci, P., Kvam, C., Kitamura, A., Ohsumi, T., Okazaki, Y., Itoh, M., Kamiya, M., Shibata, K., Sasaki, N., Izawa, M., et al. 1996. High-efficiency full-length cDNA cloning by biotinylated CAP trapper. *Genomics* **37:** 327–336.
Carninci, P., Westover, A., Nishiyama, Y., Ohsumi, T., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M., Schneider, C., and Hayashizaki, Y. 1997. High efficiency selection of full-length cDNA by improved biotinylated cap trapper. *DNA Res.* **4:** 61–66.
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. 1998. Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. *Proc. Natl. Acad. Sci.* **95:** 520–524.
Itoh, M., Carninci, P., Nagaoka, S., Sasaki, N., Okazaki, Y., Ohsumi, T., Muramatsu, M., and Hayashizaki, Y. 1997. Simple and rapid preparation of plasmid template by a filtration method using microtiter filter plates. *Nucleic Acids Res.* **25:** 1315–1316.
Itoh, M., Kitsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,

1770 Genome Research
www.genome.org

Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., et al. 1999. Automated filtration-based high-throughput plasmid preparation system. *Genome Res*. **9:** 463–470.

Izawa, M., Sasaki, N., Watahiki, M., Ohara, E., Yoneda, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y. 1998. Recognition sites of 3'-OH group by T7 RNA polymerase and its application to transcriptional sequencing. *J. Biol. Chem.* **273:** 14242–14246.

Kambara, H. and Takahashi, S. 1993. Multiple-sheathflow capillary array DNA analyser. *Nature* **361:** 565–566.

Mathies, R.A. and Huang, X.C. 1992. Capillary array electrophoresis: An approach to high-speed, high-throughput DNA sequencing. *Nature* **539:** 167–168.

Sasaki, N., Izawa, M., Shimojo, M., Shibata, K., Akiyama, J., Itoh, M., Nagaoka, S., Carninci, P., Okazaki, Y., Moriuchi, T., et al. 1997. A novel control system for polymerase chain reaction using a RIKEN GS384 thermalcycler. *DNA Res.* **4:** 387–391.

Sasaki, N., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsuura, S., Carninci, P., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y. 1998a. Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. *Proc. Natl. Acad. Sci.* **95:** 3455–3460.

Sasaki, N., Izawa, M., Sugahara, Y., Tanaka, T., Watahiki, M., Ozawa, K., Ohara, E., Funaki, H., Yoneda, Y., Matsuura, S., et al. 1998b. Identification of stable RNA hairpins causing band compression in transcriptional sequencing and their elimination by use of inosine triphosphate. *Gene* **222:** 17–23.

Scherer, J.R., Kheterpal, I., Radhakrishnan, A., Ja, W.W., and Mathies, R.A. 1999. Ultra-high throughput rotary capillary array electrophoresis scanner for fluorescent DNA sequencing and analysis. *Electrophoresis* **20:** 1508–1517.

Zhang, J., Voss, K.O., Shaw, D.F., Roos, K.P., Lewis, D.F., Yan, J., Jiang, R., Ren, H., Hou, J.Y., Fang, Y., et al. 1999. A multiple-capillary electrophoresis system for small-scale DNA sequencing and analysis. *Nucleic Acids Res.* **27:** e36.

Zhou, H., Miller, A.W., Sosic, Z., Buchholz, B., Barron, A.E., Kotler, L., and Karger, B.L. 2000. DNA sequencing up to 1300 bases in two hours by capillary electrophoresis with mixed replaceable linear polyacrylamide solutions. *Anal. Chem.* **72:** 1045–1052.