Efficient large-scale sequencing of the *Escherichia coli* genome: implementation of a transposon- and PCR-based strategy for the analysis of ordered λ phage clones

H.Kasai¹, S.Isono², M.Kitakawa², J.Mineno³, H.Akiyama⁴, D.M.Kurnit⁵, D.E.Berg⁶ and K.Isono^{1,2,*}

¹Postgraduate School and ²Department of Biology, Faculty of Science, Kobe University, Rokkodai, Kobe 657, ³Takara Shuzo, Ohtsu, ⁴Toray, Kamakura, Japan, ⁵Howard Hughes Medical Institute, University of Michigan Medical Center, Ann Arbor, 48109-0650 and ⁶Department of Molecular Microbiology and Genetics, Washington University Medical School, St Louis, MO 63130, USA

Received September 29, 1992; Accepted November 10, 1992

ABSTRACT

We have developed a strategy for efficient sequence analysis of the genome of *E.coli* K-12 using insertions of a Tn5-derived mini-transposon into overlapping ordered λ phage clones to provide universal primer-binding sites, and PCR amplification of DNA segments adjacent to the insertions. Transposon-containing clones were selected by blue plaque formation on a *dnaB*_{amber} *lacZ*_{amber} *E.coli* strain. Insertion points every 0.5–1 kb were identified by 'analytical PCR' and segments between the transposon inserts and phage arms were amplified by 'preparative PCR' using one biotinylated and one non-biotinylated primer. Single strands of amplified DNA fragments were coupled to Streptoavidin-coated paramagnetic beads (Dynabeads M280) through their biotin tails, purified magnetically, and used as templates for fluorescence-based automatic nucleotide sequencing.

INTRODUCTION

We have joined in an international effort to determine the complete nucleotide sequence of the genome of *Escherichia coli* K-12 in the next few years. The modest size of the *E.coli* genome, 4.7×10^6 bp (1), and the availability of an ordered library of λ phage clones constructed by Kohara *et al.* (2) make this project highly feasible. Nearly a half century of molecular genetic and physiological analyses of this model organism have provided an extraordinary base of knowledge that will aid in interpretation of the sequence data, which in turn will contribute to further analyses of *E.coli* and to the solution of various problems of general biological importance. The sequence data along with the ease of further genetic analysis of each region of interest would greatly augment the value of genetic, molecular and physiological information already available with this bacterium (3–6). In addition, the genome analysis of *E.coli* may reveal new principles of the organization of genomic structure and should enhance our understanding of how so many genes of diverse functions have arisen and evolved.

The ordered λ phage clone bank (2) harbors more than 1,600 well characterized members and covers more than 95% of the genome with five-fold redundancy. A 476-member mini-set selected from the bank has been distributed world-wide and has served as a resource for genetic and physical mapping, the search for *E.coli* homologs of genes characterized in other bacteria (e.g. refs. 7–10; see also the 'Physical Map of *E.coli*' section in *J. Bacteriol.*), and the development of efficient methods for reverse genetics and allelic replacement (11, 12). Traditionally, DNA sequencing using λ clones such as these has entailed shotgun subcloning into M13-type vectors, highly redundant sequencing of the subclones, and alignment of their sequence data. Recent applications of such strategies to the genome analysis of *E.coli* produced 111.4 kb of sequence containing 54 previously known and 30 candidate genes in one collaborative project work (13) and 92.4 kb containing 47 previously known and 43 candidate genes in another work (14). Extrapolation from these data suggests that the *E.coli* genome should contain 3,500–4,500 genes. Experience from these and other large-scale sequencing has shown, however, that shotgun subcloning approaches are quite wasteful of reagents, and of human and computer resources, problems that become particularly acute in the sequencing of the larger genomes of humans and eukaryotic model organisms. An alternative approach can be provided by systematic walking across cloned DNA by repeatedly synthesizing new primers and extending sequences outwards. However, this is costly when scaled up and, like the shotgun subcloning strategy, it is not very useful in regions of repetitive sequence, such as are found in most eukaryotes.

Bacterial transposons have allowed the development of a third approach that we anticipate will soon be competitive with shotgun subcloning and primer walking in many applications. Transposons are specialized DNA segments that can hop quasi-randomly to target DNAs, and can be engineered to serve as mobile binding sites for optimal universal sequencing primers. They can be used to gain access to all regions of cloned target DNA without ever disrupting the linkage among its component parts (15). The efficiency of transposon-based sequencing strategies is enhanced

*To whom correspondence should be addressed at: Department of Biology, Faculty of Science, Kobe University, Rokkodai, Kobe 657, Japan

by using insertions that have been mapped, so that repetitive sequencing can be avoided. Although a PCR mapping strategy has been taken through preliminary development (16), its full deployment has been hampered by the difficulty of reproducibly amplifying segments longer than about 6 kb, and thus of finding transposon insertions near the center of the 15–20 kb-long segments typically cloned in large capacity lambda phage vectors.

We have developed and implemented a transposon- and PCR-based sequencing strategy that avoids the pitfalls of conventional shotgun subcloning and primer walking approaches, and the limitations of the initial transposon- and PCR-based method. We exploit the extensive overlapping among λ phage clones in the five-fold redundant library of λ phage clones (2), and focus only on transposon insertions that are close enough to each end of the cloned DNA segment to be easily mapped by PCR (up to 5 kb). To enhance the efficiency of this strategy, we incorporated better binding sites for PCR and sequencing primers into the *Tn5supF* transposon, and engineered two point mutations into its transposase gene that increase transposition frequency (17). Furthermore, we used biotinylated PCR primers to adopt the single-strand solid-phase sequencing procedure developed by Uhlen and coworkers (18, 19) and FITC (fluorescein isothiocyanate)-labelled sequencing primers to perform sequencing in an automated DNA sequencer (20). The approaches developed here are well suited for the automation that will become increasingly characteristic of genome projects in coming years.

MATERIALS AND METHODS

Bacterial strains, phages and preparation of DNA

DB4496 F⁻ *sup⁺dam* (15) harbors plasmids pBRG1310 (donor of the original *Tn5supF*) and p3 (a *kan tet⁺amber amp⁺amber* plasmid for the selection of *supF*). It was used for the transposition of mini-Tn5 into phage clones. DK21 *dnaB⁺lacZ⁺amber* (21) was used for the screening of resultant λ phage clones carrying mini-Tn5. KR3 was derived from DB4496 by curing its plasmids, and KRG1 was obtained by introducing plasmid pKRG1 into strain KR3 (see below). The ordered λ phage clones (2) used in this study were 1A7, 4E3, 9D7, 5H7, 7F12, 12C2, 3C9, 24D6, 5G6, 1A6, 1F11, 6H11, 10D6, 4F6, 5F11, 8B7, 10C10, 6D4, 1G6 and 1G9, which carry the chromosomal segment from 31 min through 33 min of the genetic map (3). Plasmid pKRG1 was constructed by introducing a derivative of *Tn5supF* into a mini F cosmid vector pRE435 (22) as described below. The oligonucleotide primers used are listed in Table 1.

Modification of *Tn5supF* and construction of plasmid pKRG1

The transposon donor plasmid pBRG1310 contains a *Tn5supF* mini-transposon and its transposase gene (15). The mini-transposon consists of *supF* for selection and each of the 19-bp O- and I-end segments of an IS50 (from Tn5) which constitute sites of transposase action. This transposon was modified for improved PCR amplification and sequencing and for enhanced transposition. The modification was accomplished in two steps by PCR amplification using the '5'-add on' primers listed in Table 1. Primers were synthesized on an automated DNA synthesizer from Applied Biosystems (model 380A). First, we introduced M3 and M4 sequencing primers of M13 phage just interior to the O- and I-ends, respectively (Figure 1). For this purpose, a DNA fragment containing sequences of M4, P, *supF*, I2 and M3 was prepared by PCR using M4-P and M3-I2 primers and plasmid pBRG1310 as a template. Second, I- and O-ends were

Table 1. Oligonucleotide primers used

I. For <i>Tn5supF</i> -M3:M4 construction	
M4-P :	TGTCGACGTC GTGACTGGGA AAACCTGGC GGGGATCCT AATCTTTCT C
M3-I2 :	GGCCGTCGTT TTACAACGTC GTTTCGGACT TTTGAAAG
I-M4 :	GGGAATTCCT GTCTCTTGAT CAGATCTAGT ACTGTCTGA CTGGGAAAC
O-M3 :	GGGAATTCCT GACTCTTATA CACAAGTACT GGCCGTCGTT TTAC
II. For hyper-active transposase construction	
tnp-5' :	GGGAATTCGG GTCCTGAAC GGAACC
tnp-m1 :	GGTAGTGAA GCCGCCAGG AAGGCGCTT
tnp-m2 :	CCAGATTATC CGGCTTCTCC ATGCGTTGCC
tnp-m3 :	AAGCGCTTC CTGGGCGGCT TCACATACC
tnp-m4 :	GGCAACGCAT GGAGAAGCCG GATAATCTGG
tnp-3' :	GGTCGACTG ATCAGATCTT GATCC
III. Biotinylated primers for coupling beads	
P :	GAGAAAGAAT TAGGATCCCC
I2 :	TCSAAGGTTT GAATCCTTCC
L4 :	TGCTTCTCAT AGATCTTTCG
R2 :	GGTTCATTAC TGAACACTCG
IV. FITC-labeled primers for nucleotide sequencing	
FM3-3 :	CGACGTTGTA AAACGACGGC CAGT
FE3 :	CGCCAGGGTT TTCCAGTCA CGAC

generated by PCR amplification of the product of the first step, using the I-M4 primer and the O-M3 primers, each of which contained a terminal *EcoRI* site. The modified transposon thus constructed was 303 bp long and was incorporated into the *EcoRI* site of the mini-F plasmid pRE435 as depicted in Figure 1. The transposase gene was similarly amplified from pBRG1310 by PCR, using primers tnp-5' and tnp-3', which contained terminal *EcoRI* and *SalI* sites, respectively. Two amino acid substitutions, Met⁵⁶ to Ala and Glu³⁴⁵ to Lys, resulting in changes from the wild type to a hyperactive form of transposase as described by Wiegand and Rezinikoff (17) were then introduced into the transposase gene by PCR. This mutated gene was cloned into pUC118, sequenced for confirmation of its structure, and then incorporated next to the improved *Tn5supF* element in the above-mentioned pRE435 plasmid (Figure 1). The resultant transposon donor plasmid, pKRG1, was used to transform strain KR3 to ampicillin resistance and to λ S7⁺ sensitivity. One of the transformants was named KRG1 and used in subsequent experiments.

Tn5supF transposition and analysis of insertion sites by 'analytical PCR'

Transposition of the modified *Tn5supF* mentioned above to individual λ phage clones was achieved by infecting an exponential culture of strain KRG1 at a multiplicity of five phages per cell and incubating the infected cells for 90–120 min at 37°C. Aliquots of the resultant lysates were plated on a lawn of DK21 in molten agar containing IPTG (isopropyl- β -D-thiogalactopyranoside) and X-gal (5-bromo-4-chloro-3-indolyl- β -D-galactoside). Only *Tn5supF*-containing λ phages form plaques under these conditions. These are blue due to suppression of the *lacZ⁺amber* mutation of DK21. Individual blue plaques were isolated with toothpicks and suspended in 50 μ l of SM buffer (100 mM NaCl, 8 mM MgSO₄·7H₂O, 50 mM Tris-HCl pH 7.5, 0.01% gelatin). Ten μ l aliquots of each suspension were spotted on a lawn of strain DK21, incubated overnight at 37°C, and phages were collected in 500 μ l of SM buffer. Aliquots (1 μ l) of each phage suspension were mixed with 1 μ l of 10 $\times$ PCR buffer (300 mM Tricine-NaOH, pH 8.4 at 20°C, 20 mM MgCl₂, 50 mM β -mercaptoethanol, 0.1% gelatin, 1% Tween 20) and boiled for 5 min as described by Ponce and Micol (23). They were then subjected to 'analytical PCR' in a total volume

Figure 1. Construction of plasmid pKRG1. The original *Tn5supF* transposon (15) shown at the top left was modified by incorporating the M3 and M4 primer sequences of M13 phage mp18 as described in Materials and Methods. An *EcoRI* linker was then attached to both ends of the final construct. The transposase gene (top right) was modified by PCR by substituting the codon ATG corresponding to Met at position 56 (M56) with GCC for Ala, and the codon GAG corresponding to Glu at position 345 (E345) with AAG for Lys. This was achieved by performing three steps of PCR using the primers depicted. In the first step, three PCR fragments were amplified using *tnp-5'* and *tnp-m3*, *tnp-m1* and *tnp-m4*, and *tnp-m2* and *tnp-3'* as primers and pBRG1310 as a template. In the second step, the three PCR fragments obtained in the first step were combined into two by cross-over PCR using *tnp-5'* and *tnp-m4*, and *tnp-m1* and *tnp-3'* as primers. The two fragments were then combined into one in the third step by cross-over PCR using *tnp-5'* and *tnp-3'* as primers. The modified transposon and transposase gene were incorporated into a mini-F cosmid, pRE435.

of 10 μ l containing 250 μ M dNTPs, 0.1 unit of Taq DNA polymerase (Kurabo, Japan) and 4.5 pmoles of each primer, either P and R2, or I2 and L4 (see Table 1), for 30 cycles of reaction at 94°C for 1 min, 53°C for 1 min, and 72°C for 5 min in a model TC1 thermal cycler from Perkin-Elmer Cetus. Each reaction mixture was then electrophoresed and the sizes of PCR products estimated to locate *Tn5supF* insertions.

'Preparative PCR' amplification of sequencing templates

For sequencing, phage clones carrying *Tn5supF* at appropriate intervals (ideally 400 to 500 bp) were selected and their DNAs were amplified by 'preparative PCR'. Reactions were performed in a total volume of 50 μ l of the mixture used for 'analytical PCR'. When the intervals between the two nearest transposon insertion points exceeded 1 kb, crossover PCR (16) was used with a mixture of the two phages as templates and P and I2 primers (Figure 1, Table 1). The reaction mixture in this case was subjected first to two cycles of 94°C for 1 min, 53°C for 30 sec and 72°C for 30 sec, then to 34 cycles of 94°C for 20 sec, 53°C for 30 sec, and 72°C for 5 min, and finally to one cycle of 94°C for 20 sec, 53°C for 30 sec, and 72°C for 10 min.

In the 'preparative PCR', one of the primers used was either mono-biotinylated at the 5'-end with DMT-biotin-C6-PA from Cambridge Research Biochemicals, or di- or tri-biotinylated at the 5' end as well as one or two internal positions with Biotin-ON™ phosphoramidite from Clontech. After the PCR reaction, a 40 μ l portion from each reaction was mixed with 40 μ l of Dynabeads M280-Streptavidin (5 μ g/ μ l) washed with B&W buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 2.0 M NaCl) prior to use, and incubated for 30 min at room temperature with

occasional agitation. The bead-coupled PCR products were collected using a magnetic particle concentrator from Dynal, model MPC-E, and the supernatant was discarded. The bead-DNA complexes were washed with 40 μ l of B&W buffer and then incubated with 10 μ l of 0.1 M NaOH for 10 min at room temperature to denature the DNA strands. One DNA strand remained attached to the beads which were washed once with 50 μ l of 0.1 M NaOH, once with 40 μ l of B&W buffer, and finally with 50 μ l of TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA). The volume of the bead-DNA solution was adjusted to 6.5 μ l with distilled water.

Nucleotide sequencing

Sequencing reactions were performed using the two-step protocol for Bst™ polymerase (BioRad). Each bead-coupled DNA preparation was mixed with 1.5 μ l of the annealing buffer (100 mM Tris-HCl pH 8.5, 100 mM MgCl₂) and 1 pmol of either M3 or M4 sequencing primer labeled with FITC at multiple sites (developed and supplied by Takara Shuzo, Kyoto). The mixtures were incubated at 65°C for 15 min and then cooled down to room temperature for 15 min. One unit of Bst™ polymerase and 3 μ l of distilled water were added to each mixture for a final volume of 13 μ l. Each mixture was then divided into four 3 μ l portions, mixed with 2.5 μ l each of dA, dC, dG and dT elongation/termination mixtures, and incubated at 65°C for 10 min. The DNA was collected by the magnetic particle concentrator to which 4 μ l of stop solution (Takara Shuzo) were added. The mixtures were boiled for 1 min and a 2 μ l portion of each was analyzed in an automatic DNA sequencer from Hitachi, model SQ3000.

Figure 2. Distribution of transposon in the ordered clones from the *terC* region (min 31–33) of the *E. coli* chromosome. Twenty clones were used to apply the method developed in this work. Horizontal bars indicate the location and extent of the cloned DNA (the λ arms are not shown) with filled bars indicating the terminal 4 kb and stippled bars the central part of the cloned segment. Downward open arrows indicate the positions of transposon insertions as estimated by analytical PCR. Clone names (2) are given underneath the right end (defined by proximity to the right vector arm) of the chromosomal segment in each clone. Numbers in parentheses after clone names are the 'mini-set' clone numbers (2). Note that transposon insertions in the central (stippled) region of each clone were generally not extensively sought (see text).

RESULTS AND DISCUSSION

Overview of strategy and selection of optimal primers for PCR and nucleotide sequencing

We devised a strategy for efficiently sequencing large segments of the *E. coli* genome by focusing on just the terminal few kb of cloned DNA segments in overlapping phage clones (2). In this strategy, PCR was used in two steps, first analytical and then preparative, as described in Materials and Methods. The advantage of this procedure is that the position of the *Tn5supF* insertion can be determined quite accurately in the analytical PCR step, thereby minimizing the need for redundant sequencing and subsequent elaborate alignment of sequence data. Preliminary experiments confirmed that most (>95%) *Tn5supF* insertions were in the same orientation relative to the λ vector arms, as reported (24). This simplified our protocol, in allowing use of just two, instead of four, pairs of primers for the initial analytical PCR mapping of insertions. To fully implement this strategy, however, it was necessary to have primers that were well suited for the two PCR steps as well as for nucleotide sequencing of the resultant PCR products. Various oligonucleotides of 17 to 25 bases in length and specific to regions of *Tn5supF*, and to the right or left vector arms adjacent to the cloning site were tested for efficiency in PCR amplification, and for sequencing the PCR products. Two pairs of primers, P and R2, and I2 and L4 (Table 1), resulted in high efficiency PCR amplification of chromosomal segments from plaques of transposon-containing phage, and were used in subsequent experiments. For analytical PCR the primers were used as such, while for preparative PCR one to three biotin moieties were attached to R2 and L4 (λ arm) primers. This allowed the coupling of one strand of a preparative PCR product to Dynabeads and its magnetic purification for use as a sequencing template (see below).

The DNA amplified by preparative PCR using either of the

two primers mentioned above was then used to identify optimal sequencing primers. Various transposon-specific sequences, including the P and I2 PCR primers, were tested, but none gave good sequence readings of more than 300 bases. Moreover, the band intensity was so weak that the combined products from five PCR amplification tubes were needed for a single set of sequencing reactions. In contrast, reactions using either the M4 or M3 primer and M13mp18 phage DNA as a template gave highly readable and accurate sequence readings of up to 600 bases per single run under the same conditions. These experiments showed that the proper choice of primer, and the use of single-strand DNA templates were crucial for efficient long distance sequencing. M3 and M4 primer binding sites from M13 vectors were therefore added near each end of a modified *Tn5supF* element, just exterior to the P and I2 PCR primer binding sites that immediately flank its *supF* marker (Figure 1). This arrangement ensured that any PCR products of rare mis-priming by the P and I2 primers would not yield sequence ladders and complicate subsequent analyses.

Establishment of a new PCR-based sequencing procedure

Our PCR mapping showed that the distribution of insertions within 5 kb of the ends of cloned segments was somewhat non-random, with some rather prominent 'cold spot' regions containing few if any insertions (Figure 2). Insertion in the same regions in overlapping phage clones was quite random, however (see phage 3C9 and 4E3). This result indicated that the observed 'cold spots' were probably due to a long range context effect, perhaps reflecting selection for an appropriate, non-toxic level of *supF* function, as described by Kersulyte *et al.* (24). These cold spots are, in any case, not a reflection of a special sequence feature in the target region itself. We also found transposition to be as much as ten-fold less frequent than normal in certain phage clones. To overcome this problem, we replaced the wild

Figure 3. Trial sequencing run to demonstrate effects of number of biotin moieties on the recovery of PCR-amplified template DNA strands using Dynabeads. PCR primers (I2) containing either one (a) or two biotin molecules (b) were compared with each other with respect to their efficiency in the recovery of amplified DNA fragments (see Materials and Methods). In each case, the same amount of PCR product (monitored by agarose gel) from the same amplification was adsorbed to Dynabeads M280, denatured and washed to remove non-biotinylated complementary strands, before using the bead-bound DNA as a sequencing template.

type transposase gene with a hyperactive allele as described in Materials and Methods and in Figure 1. This alteration allowed 30 to 100-fold increase in the yield of transposon insertions in any phage clone.

The DNAs of these transposon-carrying phage clones were then subjected to preparative PCR and the products were collected magnetically (Materials and Methods). The length and nucleotide sequence of PCR-amplified fragments were found to strongly affect product recovery, and in some cases, rather large amounts (as much as 80%) of amplified fragments remained in the supernatant fraction after collecting the Dynabead-DNA complexes with the magnetic concentrator (data not shown). As a result, sequence readings were distinctly poorer than with M13 phage DNA. Several further improvements were therefore developed. First, the efficiency of DNA coupling with Dynabeads was reinforced by primers with two or three biotin groups, rather than one, in preparative PCR. This gave us much better sequence readings as shown in Figure 3. Second, DNA polymerases other than Klenow fragments and Sequenase were examined. We found that BstTM polymerase from BioRad (25) gave better results. With this enzyme, the overall sequence reaction was more efficient and the heights of individual peaks were more uniform than those with other enzymes. This increased the accuracy in base assignment by the computer program. Third, we increased the fluorescence of the sequencing reaction products by using sequencing primers containing two or three, rather than one, FITC groups, such as FE3 (a two FITC-containing primer for M4) and FM3-3 (a three FITC-containing primer for M3). Collectively, the use of two or three biotins in the primer for template strand preparation, BstTM polymerase, and sequencing

Figure 4. Nucleotide sequencing with an M4 primer containing one (a) or two (b) FITC (FE3) fluorescence groups.

primers with multiple FITC labels resulted in excellent sequence readings, typically out to 600 bp as shown in Figure 4b.

Use of 'Cross-over PCR' to fill sequencing gaps

Although the transposon- and PCR-based method described above seemed well suited for large scale analysis, the task of isolating transposon insertions every 0.5 kb was somewhat complicated by the non-randomness of insertion (see Figure 2). This problem was largely overcome at the nucleotide sequencing stage, however, using crossover PCR (16), a method in which DNAs from two Tn5^{supF}-containing phages are mixed and the segment between their insertion sites is amplified using transposon-specific primers. Since transposon insertions were recovered preferentially in one orientation relative to the λ arms (noted above), crossover PCR using P and I2 primers (Table 1) enabled us to isolate segments between most pairs of neighboring transposons as fragments containing P and M4 primer binding sites at one end and I2 and M3 binding sites at the other: the P or the I2 PCR primer was biotinylated, and single strands from these crossover PCR products were recovered on Dynabeads M280, and more than 500 bp of sequence was determined on each strand (data not shown). By using crossover PCR products in this way, we found it sufficient to have Tn5 insertions every 1 kb, instead of 0.5 kb, to achieve complete coverage of any target region.

In these experiments we found that the efficiency of recovery of preparative PCR products on Dynabeads was inversely related to their size, especially for products of more than 2 kb. This could have complicated the sequencing of regions for which relatively few phage clones were available, since lower density coverage in the clone library meant that longer stretches of sequence had to be recovered from each phage. This potential difficulty was also overcome using crossover PCR products as sequencing templates, however. Typically our crossover PCR products were 1–2 kb long, and thus could be recovered with ease, even when used to access DNA segments four or five kb in from the end of the cloned segment, i.e. from one of the λ phage vector arms.

Figure 5. Typical nucleotide sequencing data. DNA was amplified in preparative PCR using P and R2 (with two biotin moieties) primers and sequenced using an M4 primer containing two FITC groups (FE3).

Implementation of transposon- and PCR-based strategy for large scale genome analysis of *E. coli*

The rapid and systematic sequencing method described above was used in an *E. coli* genome analysis project using 20 ordered phage clones covering the *E. coli terC* region. To enhance the speed of sequencing, DNA segments from clones carrying appropriate transposon insertions were sequenced with only two- to three-fold redundancy. In most cases this provided sequence data of 400 to 500 bp per analysis, as shown in Figure 5. The accuracy of automatic base assignment, i.e. the computer-based conversion of wave form data from DNA sequencer into bases, was between 95 and 99%. This accuracy level would be acceptable for first pass analysis, provided that the number of bases in the sequenced region is determined very accurately. To achieve this, we added to each sequence run a control reaction in which M13 phage DNA was used as template. At the same time, the computer program for assigning bases was modified to cross-reference the data between control and sample runs. In this way the number of bases within the first 400 bp of the sample were counted with an accuracy of at least 99.5%, as determined in control experiments using λ phage clones from already well characterized chromosomal regions.

We have constructed an '*E. coli* nucleotide sequence database' by assembling the existing sequence data from the EMBL, GenBank and DDBJ databases and developing a system for efficient retrieval of the information. The sequence data we are currently generating will be integrated into this *E. coli* database, which will be updated and released at frequent intervals. Isolation and characterization of mutants that directly test the possible contributions of open reading frames, putative regulatory sites

or other sequence features are feasible and particularly important in the use of *E. coli* as a model organism. We have developed a simple λ phage infection protocol for efficiently recombining insertion mutant alleles that are generated in phage clones into homologous loci in the *E. coli* chromosome (12). Accordingly, the large number of phages with precisely located transposon insertions, that we find so useful for high-efficiency sequencing, are being saved for complementary functional analyses.

In summary, we propose that the use of an ordered library of λ phage clones and transposon- and PCR-based methods for preparing single stranded DNA sequencing templates from them, as described here, should greatly increase the ease and efficiency of large scale genome sequencing efforts. In the case of the present multinational *E. coli* effort, we can expect that data from our laboratory and those of collaborators to provide new insights into principles of genome organization, evolution and function, to help guide further genetic and physiological studies, and to further enhance the value of *E. coli* as a research tool for modern biology.

ACKNOWLEDGEMENTS

We are grateful to Dynal A.S. (Oslo) and Dynal Japan (Tokyo) for providing us with Dynabeads M280. This work was supported in part by grant-in-aids for scientific research Nos. 01655005, 01880029, 02237105 and 03221105 from the Ministry of Education of Japan (to K.I.), by grants DEFGO-89ER60862 from the US Department of Energy and HG-000563 from the US Public Health Service (to D.E.B.) and from the Howard Hughes Medical Institute (to D.M.K.).

REFERENCES

1. Smith, C.L., Econome, J.G., Schutt, A., Klco, S. and Cantor, C. (1987) *Science* **236**, 1448–1453.
2. Kohara, Y., Akiyama, K. and Isono, K. (1987) *Cell* **50**, 495–508.
3. Bachmann, B.J. (1990) *Microbiol. Rev.* **54**, 130–197.
4. VanBogelen, R.A. and Neidhardt, F.C. (1991) *Electrophoresis* **12**, 955–994.
5. Rudd, K.E. (1992) In Miller, J. (ed.), *A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 2.3–2.43.
6. Kröger, M., Wahl, R., Schachtel, G. and Rice, P. (1992) *Nucleic Acids Res.* **20** (suppl.), 2119–2144.
7. Umeda, M. and Ohtsubo, E. (1989) *J. Mol. Biol.* **208**, 601–614.
8. Umeda, M. and Ohtsubo, E. (1990) *J. Mol. Biol.* **213**, 229–237.
9. Singer, M., Baker, T.A., Schnitzler, G., Deischel, S.M., Goel, M., Dove, W., Jaacks, K.J., Grossman, A.D., Erickson, J.W. and Gross, C.A. (1989) *Microbiol. Rev.* **53**, 1–24.
10. Komine, Y., Adachi, T., Inokuchi, H. and Ozeki, H. (1990) *J. Mol. Biol.* **212**, 579–598.
11. Kulakauskas, S., Wikström, P.M. and Berg, D.E. (1991) *J. Bacteriol.* **173**, 2633–2638.
12. Phadnis, S.H., Kulakauskas, S., Krishnan, B.R., Hiemstra, J. and Berg, D.E. (1991) *J. Bacteriol.* **173**, 896–899.
13. Yura, T., Mori, H., Nagai, H., Nagata, T., Ishihama, A., Fujita, N., Isono, K., Mizobuchi, K. and Nakata, A. (1992) *Nucleic Acids Res.* **20**, 3305–3308.
14. Daniels, D.L., Plunkett, G., III, Burland, V.D. and Blattner, F.R. (1992) *Science* **257**, 771–778.
15. Phadnis, S.H., Huang, H.V. and Berg, D.E. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5908–5912.
16. Krishnan, B.R., Kersulyte, D., Brikun, I., Berg, C.M. and Berg, D.E. (1991) *Nucleic Acids Res.* **19**, 6177–6182.
17. Wiegand, T.W. and Rezinikoff, W.S. (1992) *J. Bacteriol.* **174**, 1229–1239.
18. Hultman, T., Ståhl, S., Homes, E. and Uhlén, M. (1989) *Nucleic Acids Res.* **17**, 4937–4946.
19. Hultman, T., Bergh, S., Moks, T. and Uhlén, M. (1991) *BioTechnics* **10**, 84–93.
20. Kambara, H., Nishikawa, T., Katayama, Y. and Yamaguchi, T. (1988) *Biotechnology* **6**, 816–821.
21. Kurnit, D.M. and Seed, B. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 3166–3169.
22. Eichenlaub, R. (1985) In Becker, Y. (ed.), *Recombinant DNA Res. and Virus*. Martinus Nijhoff Publishing, Boston, pp. 39–57.
23. Ponce, M.R. and Micol, J.L. (1992) *Nucleic Acids Res.* **20**, 623–623.
24. Kersulyte, D., Krishnan, B.R. and Berg, D.E. (1992) *Gene* **114**, 91–96.
25. Ye, S.Y. and Hong, G.F. (1987) *Scientia Sinica* **30**, 503–506.