

# Genome sequencing on both strands: the Janus strategy

Valerie Burland, Donna L. Daniels, Guy Plunkett, III and Frederick R. Blattner  
Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

Received May 8, 1993; Revised and Accepted June 15, 1993

## ABSTRACT

The design of large scale DNA sequencing projects such as genome analysis demands a new approach to sequencing strategy, since neither a purely random nor a purely directed method is satisfactory. We have developed a strategy that combines these two methods in a way that preserves the advantages of both while avoiding their particular limitations. Computer simulations showed that a specific balance of random and directed sequencing was required for the most efficient strategy, termed the Janus strategy, which has been used in the *Escherichia coli* genome sequencing project. This approach depended on obtaining sequence easily from either strand of a cloned insert, and was facilitated by inversion of the insert in the engineered M13 vector Janus, by site-specific recombination. The inversion was accomplished simply by growth on the appropriate host strain, when the DNA strand incorporated into the new single stranded phage was complementary to that in the original phage, and was sequenced by the same simple protocol as the first strand.

## INTRODUCTION

Large scale DNA sequencing projects such as whole genome analysis demand special attention to strategy: which is more efficient, to sequence random clones from a library or to sequence specific clones from an ordered set in a directed manner? The scale of genome projects also demands consideration of technical simplicity (suitability for automation), informatics requirements and cost, all vital in designing the most efficient system. Construction of random libraries allows sequencing to proceed without prior finely detailed mapping or specific subcloning steps, but a high order of redundancy is necessary for a randomly sequenced project to be assembled with no specific 'finishing' steps. Different directed strategies, on the other hand, yield sequences which fit economically together in a predetermined fashion, but require a significant level of analysis or design in the form of mapping work to identify a set of overlapping template subclones, or in designing new primers for sequencing and DNA amplification in the case of primer 'walking'. Clearly both these strategies have valuable features. In the *Escherichia coli* genome project (1, 2) we have combined the particular benefits of each, using random sequencing to collect the bulk of the data initially, followed by a simplified directed strategy that uses no new clones or technical procedures, to complete the sequence. Computer simulations were used to investigate the

effects of different combinations of random and directed data collection on cost and efficiency, and the results indicated the optimum scheme. Seamless integration of the two strategies was facilitated by a bacteriophage M13 vector, Janus, engineered to allow sequencing of both strands from single stranded templates. A single cloning process yields Janus library clones whose inserts may be inverted simply by propagation on a host expressing a site specific recombinase, Int from phage lambda (3). The packaged genome then incorporates the second strand of the insert, and the same simple protocol may be used for template preparation. Sequencing random Janus clones followed by selection of specific clones for sequencing after inversion leads to a particularly cost-effective strategy striking a balance between random and directed approaches.

## MATERIALS AND METHODS

### Recombinant DNA techniques

Standard methods (4) were used for the construction of Janus with few modifications. Enzymes were used according to the manufacturers instructions and were obtained from Boehringer-Mannheim (restriction enzymes), United States Biochemical (Sequenase, Mung bean nuclease), New England Biolabs (T4 DNA ligase), and Stratagene (alkaline phosphatase). The *attP* fragment was obtained from pPH54 (5) provided by A. Landy. The fragment was isolated by gel electrophoresis through 2% Low Melting Point agarose (BRL) in Tris-acetate buffer, and was eluted using GeneClean (Bio101 Inc.). Bacterial strains were obtained from New England Biolabs (JM101) and Stratagene (XL1-blue).

### Oligonucleotides

The *attB* oligonucleotide strands (sequences are shown in Fig. 3A) and special sequencing primers were synthesised by the Beckman Research Institute of the City of Hope. A primer was designed for ease of sequencing across the *attB* part of the Janus candidates to confirm the structure of both the original and inverted forms, and also to analyse false positives in the libraries. This primer hybridised within the *lacZ* gene about 100 bases 3' of the cloning site; the primer sequence is 5' CCTCTTCGC-TATTACGCC 3'.

### Construction of random libraries

a) *Growth and preparation of DNA for cloning.* For large DNA preparations, Janus was grown in NZY medium on host JM101; phage were collected from 2 liter culture supernatants by precipitation with 4% PEG, 0.5MNaCl, and banded on cesium

chloride gradients. After dialysis to remove CsCl, DNA was released by phenol extraction and then dialysed extensively. The double stranded replicative form of Janus was prepared as a plasmid by alkaline lysis (4) and purified by cesium chloride density gradient centrifugation. Lambda clones were grown in 2 liters of NZC, using host strain LE392, and MOI= 0.01 (6). Phage were collected by PEG precipitation and purified by cesium chloride banding; DNA was prepared from the phage as for single strand Janus.

*b) Sonication and size fractionation of target DNA.* 5 $\mu$ g of lambda clone DNA were adjusted to 20 $\mu$ l with TE buffer in 0.5ml microfuge tubes. The samples were sonicated for 50 seconds using a cup-horn probe, with fresh ice cold water at the same level for each sample. Power was delivered by a Branson Sonifier (Cell Disruptor W-350) set at 17% of duty cycle and output control at 1, giving an output of 200W. These settings normally sheared all of the phage DNA, giving a size range from 2–300 hundred basepairs (bp) to about 10 kilobasepairs (kb), evenly distributed. The sonicated DNA was then repaired by digestion with 15 units of Mung bean nuclease, at room temperature for 10 minutes. The sample was then immediately loaded onto a 1% Seaplaque GTG agarose gel (FMC BioProducts) in Tris-acetate buffer and electrophoresed for 25 minutes at 25 volts. The parts of the gel containing the marker lanes were cut off and stained with ethidium bromide. The gel was then reassembled and the 0.7–2 kb sonicated DNA cut out and eluted by GeneClean (Bio101 Inc.), into 20  $\mu$ l TE buffer; the yield was 0.5–1.0  $\mu$ g. Samples in which sonication produced an uneven distribution of material in the 0.7–2.0 kb range were not used as they often contain an excess of very small fragments which clone more efficiently than the 1–2 kb pieces desired.

*c) Cloning with Janus.* In the *Escherichia coli* genome project (1, 2) Janus was used to make subclone libraries starting from *E. coli* clones in phage lambda (7). The following procedure was adapted from Bankier *et al.* (8). The vector was prepared by digestion with Sma I and treated with alkaline phosphatase to prevent recircularization. Target DNA was prepared from lambda clones by sonication to produce fragments ranging in size from 0.5 to several kb. Target DNA ends were repaired by digestion of single strand extensions with Mung Bean nuclease; other enzymes such as Klenow polymerase or T4 polymerase, or combinations of these, all give similar results. After end repair, the target DNA was size-fractionated by agarose gel electrophoresis, and fragments from 0.7 to 2kb were collected. Purified fragments were ligated into the prepared vector and transformed into XL1-blue competent cells prepared by the method of Hanahan (protocol 2 in reference 9). When these transformed bacteria were plated, JM101 was used to provide the bacterial lawn. This combination of host strains permitted the maximum efficiency of transformation and good growth of plaques on agar, both of which are vital for optimal recovery of library clones and fast throughput in a large project. Clones containing lambda DNA inserts were identified and discarded before sequencing.

To test the quality of prepared vector batches and monitor the cloning process, a preparation of lambda DNA was partially digested with Alu I. After fractionation to prepare the same size range as the sonicated targets, it was ligated into the vector sample. Since Alu I leaves blunt ends, no repair step is necessary

and the cloning efficiency is at least 10 fold higher than that of the end-repaired target.

The yield of clones (colorless plaques) was between 1000 and 3000 per 100ng prepared target DNA. In comparison with enzymatically digested DNA, mechanically sheared DNA clones inefficiently because many of the ends are unrepairable and fragments may also be internally damaged. However, all the restriction enzymes we tested singly or in combination (10) did not produce truly random libraries and mechanical methods are preferred. In the Janus libraries the ratio of clones to vector (colorless to blue plaques) was about 2 or 3 to 1, and the background of false positives (colorless plaques from a vector-only ligation) was 2 or 3 per 100 blue. These blue plaques derive from uncut or unphosphatased molecules and the few false positives result from loss of bases at the Sma I site.

#### Screening for lambda in the Janus libraries

Library plaques were toothpicked into 100 $\mu$ l TE buffer in microtiter dish wells. Each dish was then replicated onto a lawn of bacteria (JM101) in top agar and incubated 18 hours at 37°, producing the microtiter dish array of phage samples as clear patches of lysis. The set of patches were lifted onto Nytran membranes (Schleicher and Schuell), denatured and neutralized, then baked at 68° for at least 2 hours. The membranes were hybridized to a lambda probe made by labelling Hind III-cut DNA by nick translation with <sup>32</sup>P. Standard procedures were followed (4). The membranes were exposed to Kodak XAR5 film and the hybridizing patches readily identified; these are marked in the microtiter dish wells with food color and the remaining samples transferred to a new dish. After they have been used to inoculate mini-cultures for sequencing, the phage samples are solidified by addition of 75  $\mu$ l of melted 0.4% agarose in TE buffer, cooled, sealed and stored at 4°.

#### DNA sequencing

For the large scale sequencing project, Janus library clones were grown and DNA template preparation carried out in microtiter dishes (11) for random sequencing. To analyse candidates for the Janus construct, templates prepared from 1ml cultures were extracted and purified in microfuge tubes (12). DNA sequencing was by the chain termination method of Sanger (13) and was performed by an automated system, using internal <sup>35</sup>S label, acrylamide gel electrophoresis and autoradiography (14). Reagents and Sequenase were obtained from United States Biochemicals.

#### Inversion of Janus inserts

The inverting host strain was constructed by transforming JM101 with the plasmid pHS3-1 provided by J. Gardner (15), containing the cloned lambda *int* gene fused to the P<sub>tac</sub> promoter. The transformed strain, called FB1898, was maintained on minimal agar with ampicillin (100 $\mu$ g/ml). For inversion of selected Janus clones on this host, bacteria were grown in NZY broth containing ampicillin, to mid-exponential phase, then 2ml portions were inoculated with the Janus samples. IPTG was added to 0.1mM and the cultures incubated at 37° with shaking for 6 hours to give titers similar to those obtained on JM101 without the plasmid. Phage was harvested and DNA templates prepared as for noninverted clones. In tests of Janus plaque color, spontaneous inversions and reversions were not detected; the frequencies of both were less than 1 in 10<sup>5</sup>.

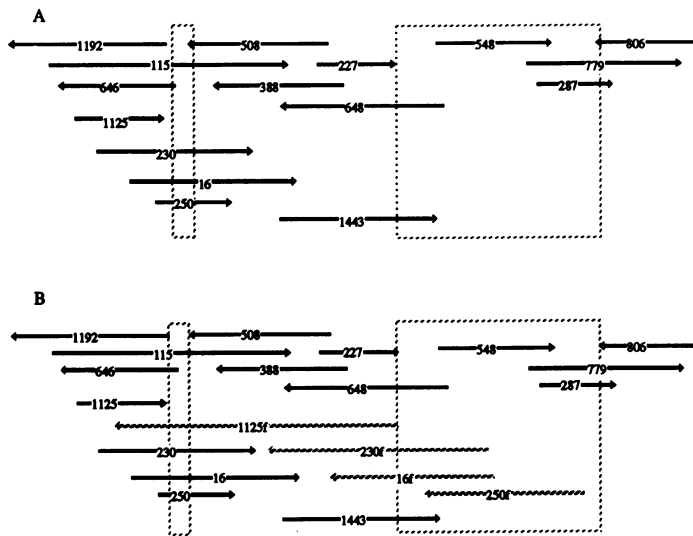
## Software

Computer programs for data collection, sequence assembly, and sequence analysis (1) were from DNASTAR packages. The program for the strategy simulations was specially written.

## RESULTS

### Strategy of sequencing with Janus

For each lambda clone, sequences were collected from the random Janus library until sufficient data were obtained to allow assembly (alignment) of sequences into contigs of several kb in length. For a library made from a 20 kb lambda clone, this number is about 500 sequences depending on the length of sequence readouts and data quality. These data assembled into 3–4 contigs having an average coverage of 6–8 $\times$ , with some areas of only 1 or 2 $\times$  coverage and other areas with data from only one strand. To address all these problem areas, specific Janus clones were selected by inspection of the alignment. Taking into account the average insert length, clones were chosen so that after inversion, new sequences collected from the opposite strand and the opposite end of the inverted clones would cover the problem areas. The new data extended the ends of contigs enabling them to merge, as well as improving coverage to meet the minimum standard of at least four determinations at every point and at least one on each strand, essential for the level of accuracy needed to locate authentic open reading frames and other features. This strategy has been used in the *E. coli* genome project and is illustrated in Fig. 1, a detailed example from one of the sequenced

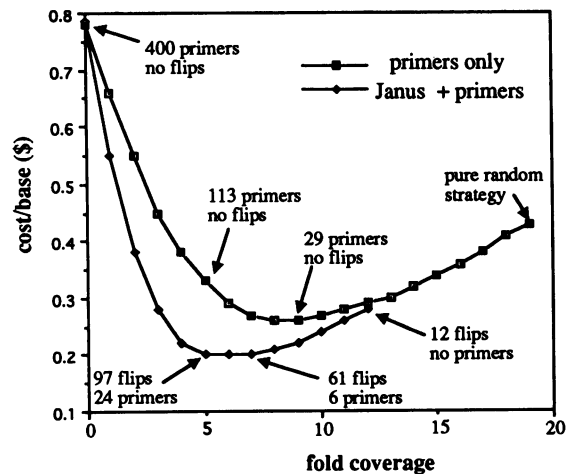


**Figure 1.** Sequencing strategy diagram generated by Seqman alignment computer program, edited for clarity. Data from the *E. coli* genome project (2); sequenced lambda clone Ec 27-236 (at 83.5 minutes on the physical map), a small portion of the lambda project is shown. Numbers represent Janus clone names. **A:** alignment of sequences collected in the random phase of the project, boxed areas are poorly covered or have data from one strand only. Clones 16, 230, 250 and 1125 were selected to provide additional data by inversion, and sequence was gathered from the opposite end for the complementary strand as described. **B:** alignment with the new sequences added (hatched arrows), showing improved coverage and second strand data for most of the problem areas. The two sequences from before and after inversion may overlap depending on the clone length and the length of the sequence. Data overlapping the flanks of the region shown were removed for clarity.

lambda clones (2) where sequencing inverted clones improved a poorly covered area.

### Simulations of combined strategies

To assess the optimum mix of random and inverted sequencing, a theoretical analysis was carried out. The effect of switching at different times from random data collection to selection of specific clones for inversion was analysed using a computer program to compare different finishing methods for simulated assemblies of random sequences. The simulation calculated the cost of finishing a 20 kb project by varied amounts of random sequencing followed by primer walking or Janus inversions to finish. Collection and assembly of raw data were determined by random number generation with assumed values for data quality, sequence readout length and cost of each operation. The minimum criteria for coverage require at least two determinations on each strand at every point. The number of finishing steps required (gaps or thin areas) was captured at different times during the simulated assembly. The curves in Fig. 2 show the cost of a sequencing project using different strategies: cost was plotted as a function of the amount of random sequence collected before finishing by a directed method. The upper curve represents finishing solely with primer walking steps, and the lower curve shows finishing with as many Janus inversions (flips) as possible followed by the much smaller number of primer walking steps needed. In this case, after collecting random sequences to each fold of coverage, the computer examined the assembly for places where Janus clones may be flipped and sequenced from the second strand to advantage. Finally a few primers were used to close the sequence. In either case the result was coverage at least twice on each strand at every point.



**Figure 2.** Comparison of costs of sequencing strategies with and without Janus inversions (flips) by computer simulation. Data points are averages of 25 simulated assemblies. Strategies using different combinations of primer walking and random sequence collection with no inversions (open squares), upper curve; with inversions, (filled diamonds), lower curve. Cost per finished base (\$) was plotted against average number of determinations at each point by random data collection (fold coverage). Cost per finished base included the fixed costs of labor and materials in the current project for template preparations, sequencing reactions, gels and autoradiography but did not include management, research or other overhead. Assumed values are: cost per raw base, 2.45 cents; total cost per gel (24 sequences) at 85% yield and 400 bases per readout, \$200; cost of primer for each walking step, \$30; primer spacing 200 bases; average length of insert in Janus clones, 1 kb.

The zero fold point (no random sequencing) is a pure primer walking strategy. At 5 fold coverage without flips, the number of primers is reduced more than 90% and the cost to finish more than halved. At 19 fold coverage, no primers are required and the strategy is pure random. The minimum cost for finishing by primer walking is at 8–9 fold random coverage, when 30–40 primers are needed. Using the Janus strategy, the minimal cost is between 5 and 7 fold random coverage. At 7 fold, 61 very cheap inversion steps reduce the need for new primers from 58 (without flips) to 6. These data are summarized in Table 1. Thus the point to switch to flipping is at 5 to 7 fold coverage for optimum efficiency, achieved with a 25% saving in cost. Above 10 fold the costs are equivalent and increase linearly as the burden of obtaining very large quantities of data outweighs the smaller number of finishing steps needed.

We also tested a number of other assumptions. A fourfold reduction in the cost of primers reduced the cost of the pure primer walking strategy by half, whereas a tenfold drop in primer cost produced a saving of two thirds. At 100 fold reduction in primer costs, such as might be obtained by the use of pentamer, hexamer or nonamer libraries (16, 17, 18), the cost per base for pure primer walking was reduced four fold. This does not take into account the high initial cost or the data management challenge. As primer costs become negligible there is little difference between the two mixed strategies by this simulation. At this point other considerations become more important.

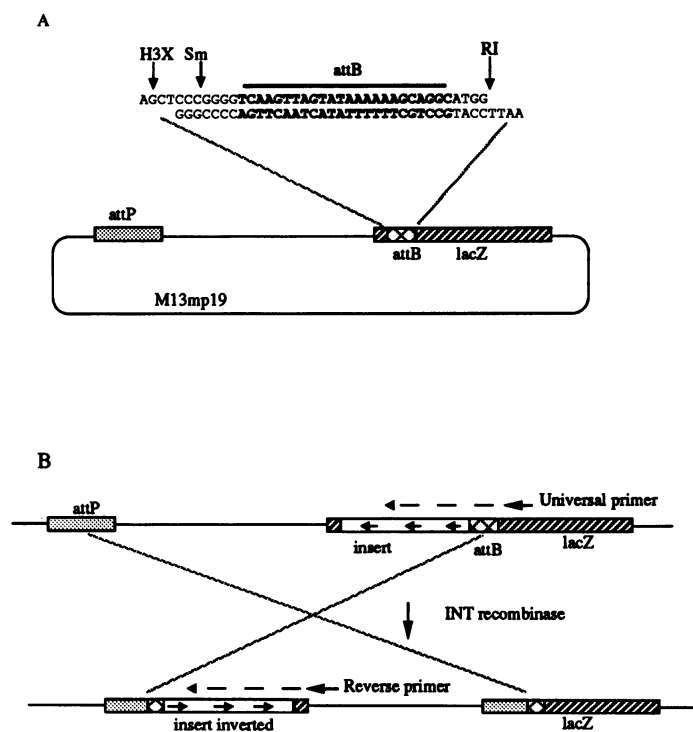
### Design and construction of Janus

M13 cloning vectors (19, 20) have the proven advantage of producing single stranded templates for primer extension (13) without competition from the complementary strand. Since M13 grows without lysing the host bacteria, DNA may be isolated easily by harvesting bacteriophage from culture supernatants, free from cell proteins, RNA or host genomic DNA. In addition the vector is distinguishable from phage containing cloned inserts by screening for plaque color (20). Janus was constructed from M13mp19 so that these features were preserved. Obtaining sequence from the second strand of the cloned insert would normally entail recloning in the opposite orientation or extensive random sequencing of shotgun clones to obtain both strands by chance. The second strand could be obtained by enzymatic synthesis on the first strand template (21), or by purification of the double stranded replicative form from cell lysates, procedures which lose the advantage of the simple preparation protocol for single strand DNA. By engineering a recombination system to invert the insert during growth, the second strand is made available for packaging and preparation as single strand template.

The site-specific recombination system for lambda integration into the *E. coli* genome was chosen for its efficiency and because

it is well understood. The lambda and *E. coli att* sites by which the phage integrates into the genome (*attP* and *attB* respectively) have been analyzed in detail (reviewed in 3) and the functional elements defined. Both *att* sites consists of a 15 basepair common core where the staggered cut sites for the Int recombinase are located. The *attB* site has a few bases on either side of the core which are essential for activity (22). The *attP* site is much larger, with essential flanking sequences of 160 and 82 basepairs, containing binding sites for the Int recombinase and IHF protein (5). Lambda integration is a reversible event but a second lambda-coded gene product (Xis) is required for excision. This gene is not present in the Janus construct, thus only one recombination event can occur. The hybrid *att* sites resulting from inversion are not recognised by Int so the inversion product is stable. The design of Janus is shown in Fig.3A.

To construct Janus, the two *att* sites were engineered into M13mp19 with a unique cloning site between them. Their relative orientations were designed to invert rather than delete the cloned segment in an intramolecular recombination. Inversion is a site-specific recombination event and acts upon the double stranded DNA of the replicative form of the phage. The whole region



**Figure 3.** Structure of Janus and inversion of the cloned insert. A) Diagram of the vector showing the relative positions of *attB* and *attP* segments and the single cloning site Sma I. Sequence of the synthetic segment is shown, with the *attB* core in boldface type; H3X indicates the 5' extension designed to ligate to a cut Hind III end but not regenerate the Hind III site. At the other end, RI indicates the Eco RI-complementary extension. The restriction site was regenerated but is not useful for cloning since it is on the wrong side of *attB* to obtain sequence by inversion. The dashed arrow represents sequence obtained from the Universal primer. B) Inversion of the cloned insert in the presence of Int recombinase. Before inversion sequence is obtained by extension of the Universal primer (dashed line above the upper section). The arrows in the insert box indicate the orientation of the insert with respect to the primer site. Upon inversion, recombination junctions are formed by the fused halves of *attB* and *attP* at each site. Sequence is now obtained by extension of the Reverse primer, indicated by the dashed arrow.

**Table 1.** Summary of cost comparisons for different strategies

Strategy	Cost <sup>a</sup>	Random coverage	Number of Primers	Total coverage
Primer walking only	0.78	0	400	4
Random only	0.43	19	0	19
Mixed <sup>b</sup>	0.26	8–9	29–42	9.2
Janus <sup>c</sup>	0.19	6–7	6–13	8.1

<sup>a</sup> \$ per finished base

<sup>b</sup> Random plus primer walking

<sup>c</sup> Random plus Janus inversions plus primers

between the *att* sites, including the cloned insert and the Reverse primer site, is inverted relative to the origin of replication (Fig.3B) so that the strand now packaged into phage heads is that complementary to the packaged strand of the original clone. The Reverse primer site is also on the packaged strand of the inverted DNA and is appropriately placed for sequencing the inverted insert (Fig.3B). Reverse sequencing primers are commercially available. Molecules not inverted (about 3 percent measured by plaque color of the inverted vector) do not interfere with the sequencing reactions since these do not contain the Reverse priming site.

Screening for plaque color in M13 depends on expression of the  $\beta$ -galactosidase  $\alpha$ -complementing segment coded by part of the *lacZ* gene. Insertion of foreign DNA into the cloning site located in the exon disrupts the gene and the enzyme is no longer produced (19). An *attB* segment was designed to read in frame after insertion into the *lacZ* gene. The segment was synthesized with an adjacent Sma I site for library cloning (Fig.3A). Single strand extensions of Hind III and Eco RI restriction sites were used to ligate the segment between these sites in M13mp19. The *attP* site was obtained as a 400 basepair Hind III-Bam HI fragment from plasmid pPH54, a subclone of the *att* region from lambda (5). The ends of the purified fragment were filled in using Klenow polymerase and ligated into the Ava II site (similarly filled in) in M13mp19. Thus the two *att* sites were introduced into M13mp19 with the unique Sma I site between them. The construct's ability to produce blue plaques on plates containing X-gal was confirmed. Since the *attP* fragment was ligated by a blunt end reaction, it was important to determine the orientation of *attP* within the construct. Each *att* insertion was identified by hybridization of candidate plaques to <sup>32</sup>P- end labelled or nick translated *att* probes. Candidates which formed blue plaques and hybridized with both probes were tested for the ability to invert by a functional assay. Colorless and blue plaques were counted after growth in the presence of induced Int recombinase; inversion dissociates the *lacZ* promoter from the coding sequence and prevents expression of  $\beta$ -galactosidase. Correctly oriented constructs gave 100% blue plaques before inversion, and yielded more than 95% colorless plaques when plated 2 hours after Int induction in liquid culture. The structure of Janus was then confirmed by restriction site mapping and by sequencing across the *attB* segment in the uninverted and across the 3' recombination junction in the inverted form (Fig.3B).

## DISCUSSION

Tackling the analysis of entire genomes has called for new ways of combining technically simple random sequencing with some level of directed approach for the finishing steps. In the *E.coli* project, the key to integrating random and directed sequencing strategies was the ability to invert the insert within the vector. Obtaining sequence from the second strand of selected clones without having to make double stranded template was a significant advantage and enabled the mixed strategy to be developed. Now project finishing is possible with a relatively small number of easy directed steps and no further mapping or subcloning, while the order of redundancy needed is much reduced.

Purely directed strategies have the advantage that a much smaller amount of data is collected than in random projects, and that the fit of each new sequence to the existing data is tailored. The success of most directed strategies however, depends on precise identification and analysis of the template clones before

sequencing: for example, nested deletions (23) and sequencing from mapped transposons (24, 25) entail prior characterization of the clones. Although there are strategies that avoid the work of subcloning and mapping, such as 'multiplex' sequencing (26) or primer 'walking', the design of oligo primers or probes from sequence at the end of a runout is a potentially limiting factor. Since new oligonucleotides must be designed and synthesized for each sequencing reaction, good design depends on having good data at the extreme end of the readout and long readouts are necessary to keep the number of primers to a minimum—both of course possible but difficult to sustain in a high throughput project. Failed reactions must always be repeated since all the sequences in the strategy plan are necessary for completion. In addition many starting points must be used simultaneously for closure in a reasonable time, creating significant costs of sample tracking and record keeping. In walking strategies, multiple hexamer priming (16, 17, 18) may reduce primer costs significantly. Such methods may eventually be able to use genomic DNA directly as a template, though this has yet to be demonstrated even by 'cycle' sequencing with a thermostable polymerase (reusing the template in thermal cycles) (27). Thus either some subcloning (28) is required, or the target DNA must be amplified from the genome by the Polymerase Chain Reaction (29) before sequencing. Finally, if these methods are used to produce sequence with only one determination at each point they become vulnerable to mutations in the template and to sequencing errors, not all of which are easily detected.

The advantages of random sequencing are that no analysis of the template clones is necessary, that only one process is needed to collect the data, and that success does not depend on the success of any individual sequence reaction, but depends only the total amount of data collected. Informatics requirements are therefore modest. These factors become particularly important when a large-scale project is undertaken. However, purely random strategies demand prohibitively large amounts of data for completion and accuracy. For example, to obtain coverage of each base at least twice on each strand, average coverage must approach 20 $\times$ , or 1760 clones for a 20 kb segment compared to fewer than 600 by the Janus strategy or 400 by primer walking only. Janus takes advantage of the conceptual simplicity of the purely random strategy but reduces the amount of effort involved; at the same time the finishing steps are acquired at nearly the same cost as random data.

## ACKNOWLEDGEMENTS

This is paper 3348 from the Laboratory of Genetics. Development, construction and verification of Janus was supported by the Romnes Fund through the University of Wisconsin Graduate School. The *E.coli* Genome Project is supported by NIH award HG00301. We thank M.Schwid, B.Fritz and E.Sommers for technical assistance, J.Gardner and A.Landy for kindly providing plasmids and N.Peterson for administration. We also thank the members of the *E.coli* Genome project teams.

## REFERENCES

1. Daniels,D.L., Plunkett III,G., Burland, V. and Blattner,F.R. (1992) *Science*, **257**, 771–778.
2. Burland,V., Plunkett III,G., Daniels,D.L. and Blattner,F.R. (1993) *Genomics*, **16**, 551–561.
3. Landy,A. (1989) *Annu. Rev. Biochem.*, **58**, 913–949.

4. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor University Press, Cold Spring Harbor.
5. Hsu, P.-L., Ross, W. and Landy, A. (1980) *Nature*, **285**, 85–91.
6. Dunn, I.S. and Blattner, F.R. (1987) *Nucleic Acids Res.*, **15**, 2677–2698
7. Daniels, D.L. (1990) In Drlica, K. and Riley, M. (eds.) *The Bacterial Chromosome*. American Society for Microbiology, Washington, D.C., pp. 43–52.
8. Bankier, A.T., Weston, K.M. and Barrell, B.G. (1987) *Methods in Enzymology*, **155**, 51–93.
9. Hanahan, D. (1985) In Glover, D.M. (ed.) *DNA cloning—A Practical Approach*. IRL Press, Oxford. Vol.1, pp. 109–135.
10. D.L. Daniels and V. Burland, unpublished data.
11. Olson, C.H., Blattner, F.R. and Daniels, D.L. (1991) *Methods*, **3**, 27–32.
12. United States Biochemical Corporation, 1990. Sequenase protocol booklet.
13. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
14. Daniels, D.L., Marr, L., Brumley, R.L. and Blattner, F.R. (1990) In Sarma, R.H., and Sarma, M.H. (eds.), *Structure and Methods*. Adenine Press, Guilderland, NY, Vol.1, pp. 29–35.
15. Lee, C.L., Gumpert, R.I. and Gardner, J.F. (1990) *J. Bacteriol.* **172**, 1529–1538.
16. Kieleczawa, J., Dunn, J.J. and Studier, F.W. (1992) *Science*, **258**, 1787–1791.
17. Kotler, L.V., Zevin-Sonkin, D., Sobolev, I.A., Beskin, A.D. and Ulanovsky, L.E. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 4241–4245.
18. Siemieniak, D.R., and Slightom, J.L. (1990) *Gene*, **96**, 121–124.
19. Messing, J., Gronenborn, B., Muller-Hill, B. and Hofschneider, P.H. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 3642–3646.
20. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) *Gene*, **33**, 103–119.
21. Smith, V., and Chee, M. (1991) *Nucleic Acids Res.*, **19**, 6957.
22. Mizuuchi, M., and Mizuuchi, K. (1985) *Nucleic Acids Res.*, **13**, 1193–1208.
23. Barnes, W.M., Bevan, M. and Son, P.H. (1983) *Methods in Enzymology* **101**, 98–122.
24. Adachi, T., Mizuuchi, M., Robinson, E.A., Appella, E., O'Day, M., Gellert, M. and Mizuuchi, K. (1987) *Nucleic Acids Res.*, **15**, 771–784.
25. Liu, L., Whalen, W., Das, A. and Berg, C.M. (1987) *Nucleic Acids Res.*, **15**, 9461–9469.
26. Ohara, O., Dorit, R.L., and Gilbert, W. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 6883–6887.
27. Innis, M.A., Myambo, K.B., Gelfand, D.H. and Brow, M.A.D. (1988) *Proc. Natl. Acad. Sci. USA*, **85**: 9436–9440.
28. Shyamala, V., and Ames, G.F. (1989) *Gene*, **84**, 1–8.
29. Strauss, E.C., Kobori, J.A., Siu, G. and Hood, L.E. (1986) *Analytical Biochem.*, **154**, 353–360.