Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 4129–4133, July 1984
Genetics

Selective isolation of cosmid clones by homologous recombination in *Escherichia coli*

(cosmid selection/cosmid vector/t complex cloning/R6K plasmid/*in vivo* packaging)

ANNEMARIE POUSTKA*, HANS-RICHARD RACKWITZ*, ANNA-MARIA FRISCHAUF*, BARBARA HOHN†, AND HANS LEHRACH*‡

*European Molecular Biology Laboratory, D-69 Heidelberg, Federal Republic of Germany; and †Friedrich-Miescher-Institut, CH-4002 Basel, Switzerland

Communicated by Paul Doty, March 15, 1984

ABSTRACT **A procedure for selection of specific cosmid clones by homologous recombination between cosmid clones from a library and sequences cloned into a plasmid has been developed. Cosmid libraries constructed in a *rec*$^-$ host strain are packaged *in vivo* into λ particles. Appropriate aliquots are then introduced into a *rec*$^+$ host containing the sequence used for selection cloned into a plasmid vector without sequence homology to the cosmid vector. After a short time for recombination, the cosmids are packaged *in vivo*. Cosmids that have taken up the plasmid by homologous recombination are isolated by plating under conditions selecting for the antibiotic resistance markers carried by both vectors. The recombined cosmids can lose the inserted sequence by another homologous recombination event and, after packaging *in vivo*, these revertants can be identified on appropriate indicator plates.**

Due to their large cloning capacity [typically 45 kilobase pairs (kbp) or more] cosmid vector systems offer major advantages in a number of applications; for example, the analysis of complex gene loci in mammals (1) or the cloning of intact and functional genes too large to be accommodated by λ replacement vectors (2).

As a step in the development of rapid "chromosome walking" techniques and to overcome some of the problems in the construction and isolation of cosmid clones, we have developed a protocol designed to simplify the construction and handling of cosmid libraries and to substitute genetic selection procedures for the otherwise work-intensive and difficult steps in the identification of cosmid clones.

The isolation of cosmid clones carrying sequence homologies to two clones derived from the proximal region of chromosome 17 of the mouse is described (3).

METHODS

Plasmids and Strains. DH1 (r_k^-, m_k^+, *recA*; *gyrA96*, *supE*) was provided by Doug Hanahan, W3110($r_k^- m_k^+$) was from Brian Seed, plasmid pRK419 was provided by Roberto Kolter, and pUC8 and pUC9 were from Joachim Messing. HL202, a suppressor-free derivative of DH1 was selected by using infection at high multiplicity with a mixture of *N* amber phages of λ and Φ80 host range. BHB3169 and BHB3175, respectively, were derived by lysogenizing W3110 $r_k^- m_k^+$ and HL202 with λ3169, a λ*imm*^{434}*cI*^{ts}*b2red3Sam7* phage. Lysogens were selected by challenge with λ*imm*^{434}*cIb2red3*. Construction and further characterization of the cosmid vectors pcos1EMBL and pcos2EMBL and the alternative selection plasmid ps1EMBL will be described elsewhere.

DNA Preparation. Plasmid and cosmid DNAs were prepared from saturated cultures by a modification of the alkaline lysis protocol (4). Vectors were further purified by two cycles of centrifugation on cesium chloride/ethidium bromide gradients.

Preparation of Plating Cells. For the recombination step, fresh single colonies of BHB3169 containing the probe plasmids were rechecked for temperature sensitivity. A heavy inoculum from an LB ampicillin plate was transferred to 500 ml of LB medium warmed to 30°C in a 2-liter Erlenmeyer flask. The cultures were grown at 30°C with vigorous aeration to an optical density of 0.3 (600 nm). The cells were harvested by centrifugation and resuspended in 25 ml of 10 mM $MgSO_4$. Plating cells for other steps in the protocol were either prepared analogously or by concentrating fresh overnight cultures 2-fold by centrifugation and resuspension in 10 mM $MgSO_4$.

Construction of Cosmid Libraries. High molecular weight DNA isolated from livers of inbred mice (129/Sv) was partially cleaved with *Sau*3A to an average length of 50–100 kbp, the enzyme was heat-inactivated, and the mixture was precipitated with ethanol and treated with calf intestine alkaline phosphatase (5). pcos1EMBL vector arms were prepared in adaptation of a protocol by Ish-Horowicz and Burke (6). Two separate aliquots of pcos1EMBL were cleaved with *Sal* I and *Sma* I respectively, treated with phosphatase, and recleaved with *Bam*HI. In analogy to a protocol developed for the cosmid vector tetracos (unpublished data), pcos2EMBL was prepared by cleavage with *Pvu* II, phosphatase treatment, and subsequent digestion with *Bam*HI. Vector arms generated by one of the two protocols were ligated to the partially cleaved phosphatase-treated mouse DNA, packaged *in vitro* as described by Scalenghe *et al.* (7), purified over a cesium chloride step gradient, and plated on DH1. After overnight growth, ≈2.5×10^5 kanamycin-resistant colonies were harvested by scraping the cells into LB medium with kanamycin (30 μg/ml) to an optical density of 0.05 at 600 nm. Cells were grown at 37°C to an optical density of 0.3 and superinfected at a multiplicity of 100 with λ3169 phage. After 3 hr at 37°C, the cells were harvested by centrifugation, resuspended in 0.05 vol of λ diluent (10 mM Tris·HCl, pH 7.6/10 mM $MgSO_4$/1 mM EDTA) and lysed by adding chloroform.

Library Amplification. Aliquots of 10^7 colony-forming units of the repackaged library were used to infect BHB3175 and grown on kanamycin plates at 30°C. Cells were harvested by scraping into 500 ml of LB medium with kanamycin (30 μg/ml) to a starting optical density of 0.05 at 600 nm. After growth at 30°C to OD 0.3, *in vivo* packaging was induced by transferring the culture to a 42°C waterbath for 20 min and then incubating at 37°C for 3 hr. Cells were harvested, concentrated, and lysed as described above. The stock of packaged cosmids was used in further experiments.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: bp, base pair(s); kbp, kilobase pairs; XGal, 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside.

‡To whom reprint requests should be addressed.

Selection of Cosmid Clones. The probe plasmids Tu66 and Tu169 (3) were transformed into BHB3169 and grown on ampicillin plates. Cells were plated as described above; 10^{10} cells were then preincubated for 15 min at 30°C with 10^9 colony-forming units of the packaged cosmid library, diluted with LB medium to 200 ml, and transferred to a shaking incubator at 30°C. After 1 hr, kanamycin was added to a concentration of 5 μg/ml. After 3 hr, *in vivo* packaging was induced as described above. Aliquots of the recombined packaged library were then used to infect DH1 or BHB3175. Cells were incubated in LB for 1 hr (to allow expression of the kanamycin resistance), pelleted, the supernatant was carefully removed, and the cells were resuspended in 0.1 vol of λ diluent and plated on nitrocellulose filters (Schleicher & Schuell) on L agar containing kanamycin (15 μg/ml) and ampicillin (30 μg/ml). Both centrifugation and plating on filters were necessary to avoid complications caused by ampicillinase activity. Colonies were picked and purified by another cycle of *in vivo* packaging in a 200-μl culture.

Selection of Revertants. Fresh colonies were used to inoculate 2 ml of LB medium containing kanamycin (30 μg/ml). After growth to OD ≈0.3 (600 nm), *in vivo* packaging was induced either by heat induction of a resident prophage or by superinfection with either λ3169 or λ3171 (λ*imm*^{434}*cI*^{ts}*Sam7*) phages. Packaged cosmids were used to infect BHB3175 or DH1 and aliquots were plated on kanamycin plates containing 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (20 μg/ml) (XGal). White colonies were picked, tested for ampicillin sensitivity, and analyzed further by restriction analysis.

RESULTS

Strategy. The strategy of library amplification, clone selection, and clone manipulation (Fig. 1) relies on the repeated transfer of cosmid libraries and cosmid clones by *in vivo* encapsidation of cosmids into λ coats, allowing the rapid and efficient transfer of cosmid sequences between hosts of the appropriate genetic background.

(*i*) First, cosmid libraries constructed in a vector without sequence homology to a short probe plasmid are introduced into BHB3175, amplified on plates to minimize competition between clones, transferred to liquid culture for a short period, and packaged *in vivo*.

(*ii*) Aliquots of this library stored as a suspension of packaged cosmids are then used to infect BHB3169, a *recA*$^+$ *in vivo* packaging host carrying the selector plasmid. After a period of 1–3 hr, sufficiently long for recombination between homologous sequences to take place, the cosmids are packaged out and used to infect a *recA*$^-$ host strain (DH1 or BHB3175), minimizing the time of exposure to the *rec*$^+$ environment.

(*iii*) Cosmids that integrated the probe plasmid by homologous recombination are then selectively grown on plates containing antibiotics selecting for recipients of both probe plasmid and cosmid sequences.

(*iv*) The original structure of the selected cosmid is restored by reverting the recombination event (in the case of sequence identity between the homologous sequences). Recombined cosmids are packaged *in vivo*, introduced into a *recA*$^-$ *lac*$^+$ host, and plated on XGal plates, selecting for the resistance marker carried by the cosmid. Cosmid clones that have lost the probe plasmid carrying a *lac* operator sequence can be identified as white colonies.

In addition, further genetic manipulation of isolated cosmids by repeated integration and excision of special probe plasmids can be used to introduce additional genetic elements into cosmids or genes or in chromosome walking strategies.

Construction and Amplification of Cosmid Libraries. Cosmid libraries were constructed from 129 mouse DNA, using the R6K-derived cosmid cloning vectors pcos1EMBL or pcos2EMBL, a derivative carrying two cos sites to allow efficient construction of cosmid libraries (Fig. 2).

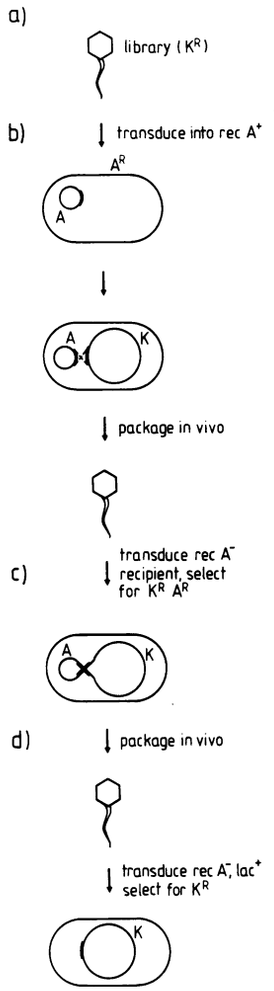

FIG. 1. Schematic drawing of steps in cosmid library amplification, the selection of specific cosmids, and excision of the selection plasmid. A^R, ampicillin resistance; K^R, kanamycin resistance.

Because of the variability of the cosmid cloning yields in using *in vivo* packaging strains as primary hosts, the libraries usually were constructed in DH1, packaged by superinfection with λ3169, and transferred to BHB3175 by infection. Phages used in *in vivo* packaging contain a temperature-sensitive immunity to allow induction of phage growth by temperature shift in the lysogenic strains, the b2 deletion decreases the efficiency of excision after induction, and the *Sam7* mutation allows packaging to proceed for long periods by blocking cell lysis and, therefore, increasing the efficiency of packaging. Except in experiments designed to revert recombination products by a second recombination event, *red*$^-$ phages were used.

Selection of Cosmid Clones Derived from the T Complex of the Mouse. To test the possibility of isolating specific cosmid clones from complex libraries by homologous recombination, Tu66 and Tu169, two different clones of *Eco*RI fragments derived by microdissection from the proximal part of mouse chromosome 17 in pUC9, were used (3). Clone Tu66 contains a 500-base-pair (bp) *Eco*RI fragment hybridizing to four different *Taq* I restriction fragments in the mouse genome. Tu169 contains two *Eco*RI fragments of 2.8 and 2 kbp. After recombination with the cosmid library, both clones gave rise to double-transducing particles at a frequency of ≈5×10^{-7}. Exact recombination frequencies vary with

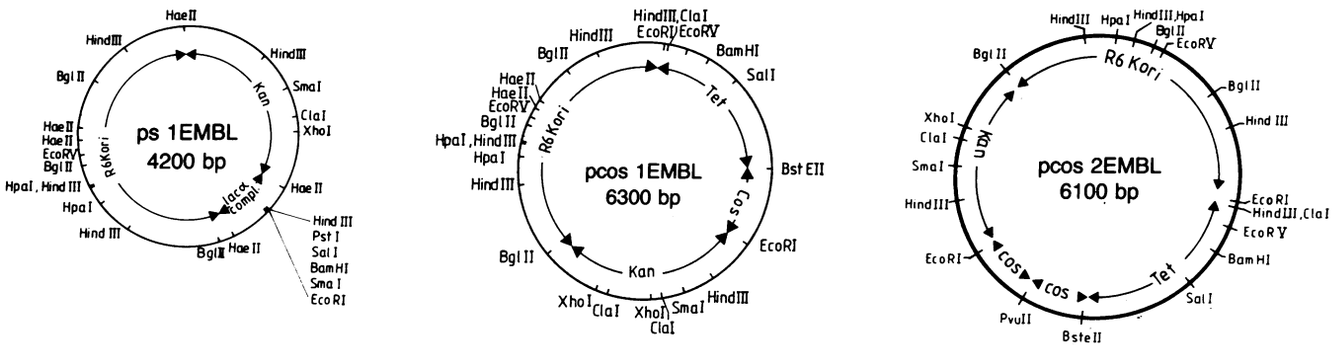

FIG. 2. Restriction maps of ps1EMBL, pcos1EMBL, and pcos2EMBL derived from sequence information and restriction digest data. R6K origin sequence and kanamycin-resistance (Kan) genes were derived from pRK419 (8); the tetracycline-resistance (Tet) gene and the *cos* sequence were from pHC79 (9). The *lac*-encoding sequence was derived from the plasmid pUC8 (10).

clone and library used, but in general have been found to lie between 10^{-6} and 10^{-8}. Typical titration results at the different steps are given in Table 1. In contrast, no double-transducing particles could be observed in control experiments using pUC9 without insert for selection (detection limits 10^{-9} and 10^{-10}).

In some experiments, a variable level of background colonies has, however, been observed, which in the reversion test only give rise to white colonies not resistant to ampicillin. Up to now, we have not been able to identify possible causes for this phenomenon. Clones recovered from a recombination experiment should, therefore, be checked in the reversion experiment to verify their ability to retransduce the ampicillin-resistance marker and the *lac* operator sequence.

Though both probe plasmids carry inserts of different lengths, the recombination frequencies were found to be rather similar. It can be expected that recombination over sequence homologies significantly shorter than the ones used here will proceed at decreased frequencies. Although in the experiments described here recombination was allowed to proceed for 3 hr before inducing the prophage, in general, recombination times of 1–2 hr were found to be sufficient.

Position and orientation of the recombination event has been verified by digestion of DNA isolated from recombinant cosmids with enzymes cleaving the polylinker sequence adjacent to the insert of the probe plasmid and hybridization with radioactively labeled pUC9 DNA and DNA of the probe plasmid used (Fig. 3). The resulting hybridization patterns verify the structure predicted for the product of a homologous recombination event over the probe sequence. As shown for clone 66.1, two different fragments hybridize with the probe sequence, only one of which also hybridizes with the plasmid probe. In addition, there is a 3.2-kbp band in all slots except slot 1. This band corresponds to linearized probe plasmid, which excised from the cosmid and replicated autonomously. *Eco*RI (slot 1) gives a pattern identical with the hybridizing bands of the cosmid. Analogous results were obtained for clone 169.1.

Table 1. Titration results

	Volume, ml	Titer
Step *b*		
Cell input	10	10^9/ml
Cosmid input	1	10^9/ml
Cosmid-infected cells (0 hr)	200	5×10^6/ml
Cosmid-infected cells (3 hr)	200	10^7 to 10^8/ml
Step *c*		
Packaged cosmids	20	10^8 to 10^9/ml
Packaging phage	20	10^7 to 10^8/ml
Fraction of recombinants ($Km^R Ap^R/Km^R$)		2×10^{-6} to 10^{-7}

Km^R, kanamycin resistance; Ap^R, ampicillin resistance.

Isolation of Revertant Clones. Often it is useful to isolate cosmids that have lost the integrated probe plasmid by a second homologous recombination event. To identify these revertants, cosmids were mobilized either by induction of a resident prophage or by superinfection with either red^+ (λ3171) or red^- (λ3169) phages carrying the *Sam*7 mutation required for efficient *in vivo* packaging. Using the red^+ phage (λ3171) providing λ recombination functions, reversion frequencies between 0.5% and 80% were reached. Unexpectedly, a low frequency of revertants could also be observed in experiments carried out using red^- phages (λ3169) in a $recA^-$ host environment. If available, products of this type of reversion were analyzed to further decrease the possibility of *red*- or *rec*-induced rearrangements of the cosmid clones.

DNA preparations from revertants show the expected changes in structure—a loss of the plasmid sequences, verified both by loss of the pUC9 band in an *Eco*RI digest and by

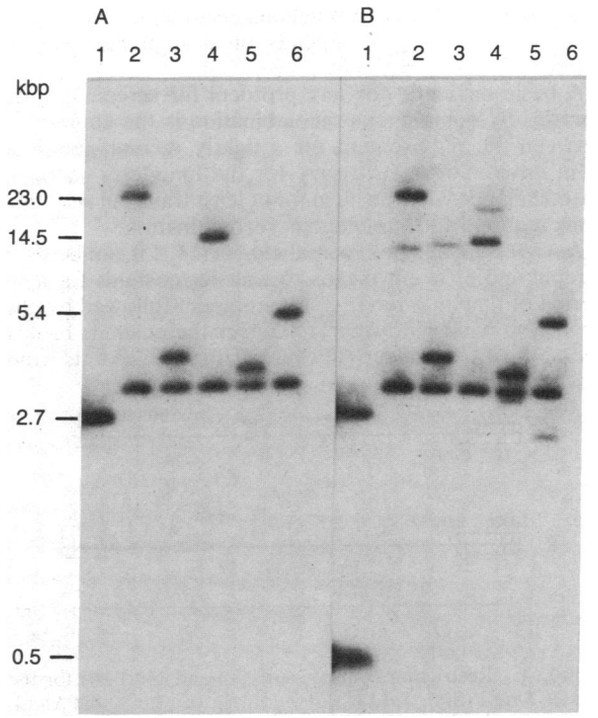

FIG. 3. *Eco*RI, *Sma* I, *Bam*HI, *Sal* I, *Pst* I, and *Hin*dIII (lanes 1–6, respectively) digests of clone 66.1 hybridized with pUC9 (*A*) and Tu66 (*B*).

Southern blot hybridization—as well as the expected decrease in staining and hybridization intensity to the *Eco*RI fragment used for recombination. To check further for potential instabilities, reversion products of clones were used in a second cycle of recombination and reversion. No changes in the *Eco*RI digestion pattern could be observed, indicating a high stability of these cosmid clones during the genetic manipulation steps.

Clone Characterization. To verify the expected structure of the clones, the restriction maps of revertants of clones 66.1 and 169.1 were determined by using standard procedures. In addition, information on the environment of the recombination site was obtained by measuring the sizes of the hybridizing fragments shown in Fig. 3. The resulting interpretation of the structures of recombinants and revertants is shown in Fig. 4.

DISCUSSION

We describe an efficient and easy protocol for the genetic selection and manipulation of cosmids containing eukaryotic sequences in *Escherichia coli*. Similar to other genetic selection techniques, our procedure relies on techniques to bring sequences cloned into different vectors together in an environment allowing homologous recombination to take place, followed by selection of the (usually rare) recombination products. Different implementations of this basic concept, using either phage–phage (11) or plasmid–phage recombination (12) have been described.

For the selection of specific clones from libraries encompassing more than 10^5 different sequences, a number of problems have to be addressed. To allow the specific detection of homologies between the insert sequences, pairs of vectors without sequence homologies must be used. We have, therefore, developed sets of vectors (cosmids and a probe plasmid vector) by modification of pRK419 (8), a plasmid containing the β and γ origins of the naturally occurring multicopy plasmid R6K (13) and a Tn*903*-derived kanamycin resistance gene (Fig. 2). Since these vectors have no sequence homology to commonly used ColE1-based plasmid and cosmid vectors, homologous recombination using appropriate vector pairs depends on a sequence homology within the cloned segments.

A basic difficulty for any protocol for screening cosmid libraries by homologous recombination is the contradiction between the requirement for a tightly recombination-deficient environment necessary for the long-term stability of cosmid clones with the need for at least transient rec^+ conditions required for homologous recombination.

In vivo packaging of cosmid clones (14, 15) can be used as a rapid and efficient means to transfer cosmids for a short period of time into the rec^+ environment followed by another *in vivo* packaging step to transfer the cosmids back to a $recA^-$ host. By selectively mobilizing the cosmid component, this second *in vivo* packaging step allows a tight selection for recombinant molecules that carry the probe plasmid *cis* to the *cos* sequence. This is combined with the powerful, but by itself not *cis*-selective, antibiotic-resistance selection and, therefore, allows the isolation of recombinants that occur at frequencies of 10^{-7} to 10^{-8}, levels far below the expected reversion rates of point mutations. The expected frequency of recombinants (10^{-6} to 10^{-8}) is the product of the frequency of unique gene clones in the library ($\approx 10^{-5}$) with the expected recombination frequency (10^{-1} to 10^{-3}). This illustrates the need to use highly efficient procedures for the transfer and selection of cosmid clones.

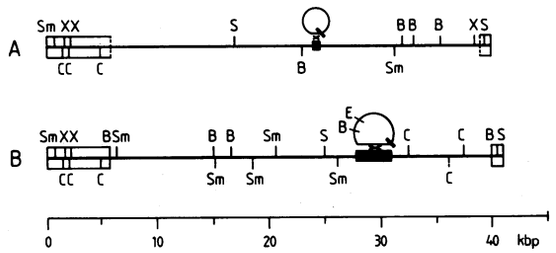

FIG. 4. Restriction maps of 66.1 (*A*) and 169.1 (*B*) for the enzymes *Sma* I (Sm), *Sal* I (S), *Bam*HI (B), *Cla* I (C), and *Xho* I (X). Position and orientation of homologies to probe plasmids are indicated by black bar. Boxed region represents the cosmid vector. Line drawn in the selection plasmid adjacent to the position of sequence homology shows location of polylinker sequence.

The use of the λ packaging system in the selection of the recombinants introduces a size selection that could decrease the possibility of recovering very large cosmid clones by this protocol. This difficulty can be decreased by the use of small probe plasmids (≈2.7 kbp for pUC9; ≈4 kbp for ps1EMBL). Since characterized cosmids from the library showed an average size of ≈45 kbp, we expect this size constraint not to be of major importance in the selection of cosmids. In addition, the size limits for the one-step packaging of cosmid sequences should be less stringent than the repetitive discrimination occurring during phage growth. Selection for essentially full-length cosmid clones could, however, contribute to discrimination against products of deletion events occurring in the rec^+ environment.

In general, we found the fidelity of the selection and reversion steps to be remarkably high, although more experience with a larger sample of different cosmid clones clearly will be required to put more stringent limits on the level of occurrence of cosmid instabilities. Clones have been tested for rearrangement by different techniques. The comparison of the cosmid restriction maps with the results of genomic Southern blots show no evidence for rearrangement during cloning, amplification, selection, and reversion steps (data not shown). This is in agreement with model experiments carried out using the inverse selection system, in which recombinants between ColE1 origin H-2 cosmids and H-2 fragments cloned into ps1EMBL (Fig. 2) were selected. Again, no changes in the restriction pattern beyond those predicted by the integration were observed (data not shown).

Changes in the primary structure of the gene sequences recovered after reversion can be predicted for recombination events occurring between slightly mismatched sequences, a case expected in recombination between different members of gene families as well as in recombination between genomic and cDNA sequences. Both repair mechanisms acting on transiently mismatched sequences and nonequivalent crossover points during forward and backward recombination will lead to unpredictable changes in the structure of the recovered clone. In some cases, these effects can be used as tools for mutagenesis, the construction of hybrid genes, or the removal of intron sequences.

Influences of length and sequence divergence of the homologous sequences have been studied in detail for phage–plasmid recombination systems (12). Although only few results are available for plasmid-cosmid recombination, similar effects can be expected to be observed. Especially interesting for many applications could be the higher stringency of homologous recombination in discriminating between slightly diverged sequences, useful both in the selective isolation of specific members of multi-gene families and in chromosome-walking experiments.

We thank Roberto Kolter, Noreen Murray, and Doug Hanahan for providing plasmids and strains; Michael Steinmetz for H-2 cosmid clones; Brian Seed for making information available before publication; Tamotsu Hashimoto-Goto for exchange of information and vectors; Guenther Zehetner for computer programs used in the restriction analysis; Hajo Delius for a few crucial electron micro-

graphs; and Wendy Moses for preparing the manuscript. We would also like to thank Bernhard Herrmann for contributing the genomic Southern blots and ideas on all steps of the protocol. Elliot Ehrich, John Bedbrook, and Thoralf Sundt have contributed to earlier stages of the work. H.-R.R. was supported by Training Grant GBI-046-D of the Biomolecular Engineering Programme of the Commission of the European Communities.

1. Steinmetz, M., Winoto, A., Minard, K. & Hood, L. (1982) *Cell* **28,** 489–498.
2. de Saint Vincent, B. R., Delbrück, S., Eckhart, W., Meinkoth, J., Vitto, L. & Wahl, G. (1981) *Cell* **27,** 267–277.
3. Röhme, D., Fox, H., Herrmann, B., Frischauf, A.-M., Edström, J.-E., Mains, P., Silver, L. M. & Lehrach, H. (1984) *Cell* **36,** 783–788.
4. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7,** 1513–1523.
5. Frischauf, A.-M., Lehrach, H., Poustka, A. & Murray, N. (1983) *J. Mol. Biol.* **170,** 827–842.
6. Ish-Horowicz, D. & Burke, J. F. (1981) *Nucleic Acids Res.* **9,** 2989–3014.
7. Scalenghe, F., Turco, E., Edström, J.-E., Pirrotta, V. & Melli, M. (1981) *Chromosoma* **82,** 205–216.
8. Kolter, R. & Helinski, D. R. (1978) *Plasmid* **1,** 571–580.
9. Hohn, B. & Collins, J. (1980) *Gene* **11,** 291–298.
10. Vieira, J. & Messing, J. (1982) *Gene* **19,** 259–268.
11. Carroll, D. & Ajioka, R. S. (1980) *Gene* **10,** 273–281.
12. Seed, B. (1983) *Nucleic Acids Res.* **11,** 2427–2445.
13. Kontomichalou, P., Mitani, M. & Clowes, R. C. (1970) *J. Bacteriol.* **104,** 34–44.
14. Vollenweider, H. J., Fiandt, M., Rosenvold, E. C. & Szybalski, W. (1980) *Gene* **9,** 171–174.
15. Lindenmaier, W., Hauser, H., Greiser de Wilke, I. & Schütz, G. (1982) *Nucleic Acids Res.* **10,** 1243–1256.